Visual analytics and artificial intelligence for marketing

Overgoor, G.

Publication date
2021

Document Version
Final published version

Link to publication

Citation for published version (APA):
Overgoor, G. (2021). *Visual analytics and artificial intelligence for marketing*.

VISUAL ANALYTICS AND ARTIFICIAL INTELLIGENCE FOR MARKETING

GIJS OVERGOOR

Visual Analytics and Artificial Intelligence for Marketing

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op woensdag 30 juni 2021, te 10.00 uur

door Gijs Overgoor
geboren te Enschede

Promotiecommissie

Promotor:	prof. dr. W.M. van Dolen	Universiteit van Amsterdam
Copromotor:	dr. W.M. Rand	North Carolina State University
Overige leden:	prof. dr. E.O. Postma	Tilburg University
	prof. dr. M. Salomon	Universiteit van Amsterdam
	dr. J.Y. Guyt	Universiteit van Amsterdam
	dr. K. Pak	Universiteit van Amsterdam
	prof. dr. M. Worring	Universiteit van Amsterdam
	prof. dr. N. Helberger	Universiteit van Amsterdam

Faculteit Economie en Bedrijfskunde

Contents

Chapter 1

Introduction

"We are in a knowledge economy, its currency is information, and the performance indicator is increasing returns to knowledge." (Kumar et al., 2019, p. 16) It is crucial for companies to effectively store, process, and analyze data. In marketing, a company's ability to translate data into knowledge can greatly improve how effective decision-makers and content creators are in marketing their brand and product. This is especially relevant for online environments, such as social media platforms and e-commerce websites, where consumers are overloaded with information and firms are competing for their attention. Company websites and social media platforms are important components of companies' marketing strategies. For example, social media is used to achieve a variety of key marketing objectives, from creating awareness to facilitating sales (Batra and Keller, 2016; Kumar et al., 2013; Kumar et al., 2016; Colicev et al., 2018; Luo, Zhang, and Duan, 2013). By effectively using firm-generated content, that is "liked" by consumers, marketers can positively influence their consumers' purchase behavior (Beukeboom, Kerkhof, and Vries, 2015; Mochon et al., 2017). In a similar way, e-commerce platforms are becoming ever more important to consumers as they spend more and more time online. At the time of writing, in the middle of a pandemic, stay-at-home orders and social distancing have accelerated the importance of e-commerce platforms even more (OECD, 2020). To help the consumer in their search for information and during the deliberation phase, it is important to understand which information (photos and text) are enticing and/or likeable. Finding the drivers behind these likes and clicks can help (1) understand how consumers interact with the information that is presented to them and (2) improve marketing content.

Digital marketing content is mostly textual or visual in nature. Recent studies have studied the drivers behind the effectiveness of textual content (Berger and Milkman, 2012; De Vries, Gensler, and Leeflang, 2012; Hewett et al., 2016; Stephen, Sciandra, and Inman, 2015), but there is very little research on the effectiveness of visual content. This is remarkable, because most online platforms are full of visual content and it is well known that visual content is generally more engaging than textual content (Li and Xie, 2017); it grabs consumers attention (Pieters and Wedel, 2004); it is memorable; and it is much faster for consumers to process than it is to process text (Blanco, Sarasa, and Sanclemente, 2010). There are also several studies that highlight that visual content eases information processing (Bettman, Luce, and Payne, 1998; Pan, Zhang, and Law, 2013; Zhang et al., 2019). In addition, we know that consumers use visual content in their online purchase decisions (Dzyabura, El Kihal, and Ibragimov, 2018; Zhang et al., 2019). It is clear that visual content is important. However, we have limited knowledge about what makes it likeable or clickable. We can tackle this problem by studying the impact of visual content in two ways: 1) theory-driven investigation and 2) data-driven exploration.

First, theory-driven investigation takes an established theoretical framework that can be used to examine the visual content through a theoretical lens. Visual complexity theory (Attneave, 1954; Donderi, 2006) offers such a theoretical lens through which we can explore the likeability of visual content. In fact, we observe several contrasting views in the advertising literature related to visual complexity: simplicity vs. complexity. Some studies suggest simplicity to enhance liking (Aitchison, 2012; Book and Schick, 1997) and others suggest that ads with a higher complexity work best (Nelson, 1994; Putrevu, Tan, and Lord, 2004). Pieters, Wedel, and Batra, 2010, in turn, divide the visual complexity into two main types (e.g. feature complexity and design complexity) and show that its relationship with attitude towards ads is more nuanced. Low complexity, or simplicity, works best for feature complexity, but high complexity works best for the design complexity. And most recently, (Shin et al., 2019), find the exact opposite for visual content on social media. These opposing findings suggest the relationship between visual complexity to be even more nuanced, and perhaps to be non-linear. In addition, individual components of the overarching categories of feature- and design complexity may even have differential effects on liking behavior. Visual complexity theory provides us with an interesting perspective on the impact of visual content on social media likes. Therefore, theory-driven investigation is suitable to study the relationship between visual content and likes.

Second, data-driven exploration does not require established theory, but instead it drives the exploration of new theory and new hypotheses based on findings in the data. This is useful, because knowledge about consumer product search and click behavior with respect to visual content is limited, beyond what has already been stated above. It is known that consumers use the visual content provided by e-commerce platforms for their decisions (Pan and Zhang, 2016; Noone and Robson, 2016). Based on these works, I hypothesize that the visual content presented to consumers during their search for products influences click decisions as a stand-alone attribute and/or through interaction with textual and numerical information. However, little is known about the aspects of this content that drive the clicks. Therefore, data-driven exploration is suitable to study the relationship between visual content and clicks.

The goal of this dissertation is to learn more about why consumers like and click on visual content online. I will employ both theory-driven investigation and data-driven exploration to get closer to that objective. Regardless of the option best suited for the problem at hand, I need automated frameworks that can objectively score visual content. This information can then be related, at scale, to consumers and their decisions. The methods and tools I need to study the impact of visual content bring us to an important subgoal of this dissertation: The development of visual analytics.

1.1 Visual Analytics

It is known that the growth of new disruptive technologies has resulted in an exponential growth in data. In fact, the estimates highlight that 90% of today's data has been generated in the last two years (Forbes, 2018). What's more important, and something that is not as well known, is that up to 90% of this data is unstructured (CIO, 2019). Unstructured data is data that can't be captured or stored in a traditional column-row type table or spreadsheet. Instead, it represents data that is not organized in a pre-defined manner (Wikipedia, 2017). Examples of unstructured data include, text, sound, images, and videos. In marketing, unstructured data presents a rich source of information, but at the same time, it is an information source that remains largely untapped (Balducci and Marinova, 2018). Indeed, 95% of businesses cite the need to manage unstructured data as a problem for their business (Forbes, 2019). Previous research states that this is likely due to technical difficulties in translating the unstructured data into structured information (Ordenes and Zhang, 2019; Ma and Li, 2019). The solution to this problem can be found in Artificial Intelligence (AI) and machine learning. Businesses recognize this, and 98.8% of them invest in initiatives on Big Data and AI (Vantage, 2020). A large part of this investment goes into technologies based on the use of machine learning. AI and machine learning are certainly changing the business landscape and the question of how we can leverage their power for marketing is important and timely. By designing the methods necessary to study the impact of visual content at scale, this dissertation answers that question and it aims to provide the tools to turn unstructured data into information, and information into insights.

It is clear that unstructured data presents both opportunities and challenges to marketers (Balducci and Marinova, 2018; Wedel and Kannan, 2016). On the one hand, unstructured data is a valuable source of information for firms and consumers, but on the other hand its unstructured nature makes it quite difficult to process. In the past decade, many text mining methods have been developed that demonstrate the value of mining textual information for marketing (Ordenes and Zhang, 2019; Berger et al., 2020). Visual content, however, hadn't received quite as much attention, because of the increased difficulty behind extracting information from imagery. For this reason, the main focus of this dissertation is to use visual analytics or image mining methods to study the impact of online visual content in various settings, such as social media, e-commerce platforms, and comparison shopping websites.

Before we dive into why this needs to be explored, it is helpful to describe why we are now able to translate visual content into information. Recent developments in computer vision, and the study of deep learning, give us the ability to automatically extract valuable information from visual content. Part of the dramatic increase in image processing capability comes from the use of Convolutional Neural Networks (CNNs). The first CNN, LeNet5, was developed by LeCun et al. (LeCun et al., 1998). Neural nets have existed for a long time, but LeCun et al. developed convolutions to break up an image into different areas that focus on processing one particular part of the image. The LeNet5 architecture showed that convolutions are effective at extracting image features. Because each convolution is a type of filter that is applied multiple times to different parts of the image, the CNN uses only a small set of parameters that need to be estimated to detect similar features in multiple locations in an image. Nowadays, we can use large datasets with labeled images and the increasingly cheap nature of computer power to learn the parameters in convolutions at a large scale. The CNN architecture builds up a large amount and variety of information from the image and combines all of these different types of information to enable identification of complex concepts in the image. By scanning over a large number of pre-labeled images and adjusting weights, the CNN can "learn" how to recognize the labeled information in the images. After a CNN is trained on millions of labeled images, we can utilize it to extract useful information from visual marketing stimuli.

The information extracted using CNNs generally represents semantic information, but we can also use basic image processing to extract low-level image information based on visual complexity theory (Corchs et al., 2016; Pieters, Wedel, and Batra, 2010; Shin et al., 2019). Low-level image information represents information on the pixel level, which includes basic features such as colors, luminance, and edges. Research has established how measures related to these basic image aspects influence fixation, attention, and attitude towards visual content (Gorn et al., 1997; Gorn et al., 2004; Lichtlé, 2007; Machado et al., 2015; Heaps and Handel, 1999; Peracchio and Meyers-Levy, 2005). In particular, the variation in these basic concepts has shown to impact attitude and liking behavior as well (Pieters, Wedel, and Batra, 2010; Shin et al., 2019). Visual complexity captures this variation and it can be measured and approximated in many different ways (Corchs et al., 2016; Nagle and Lavie, 2020). In this dissertation, we combine existing methods with new ways to extract visual complexity measures from imagery.

Using both the semantic and basic-level visual information extracted from images, in combination with success measures - likes on social media, clicks on e-commerce websites - we can study the consumer interaction with visual content online and quantify its impact. Studying consumer interaction with visual content is relevant in two ways: 1) Substantively, because there is little research in marketing that studies the impact of visual content, simply due to a lack of methods to study it at scale. As a result, there are a lot of untouched research questions in this space. 2) Methodologically, it is necessary to develop new ways of using visual analytics in marketing.

We have recently seen an adoption of visual analytics methods for marketing research (Ordenes and Zhang, 2019). Nanne et al. (2020) use various computer vision model to analyze user-generated imagery and show how computer vision can be utilized to analyze it, what pre-trained methods are most accurate and how visualization tools (such as t-SNE) in combination with these methods can help understand what is posted about brands. Zhang and Luo (2018) use deep learning model to predict the chances of restaurant survival based on reviews with images on yelp. Hartmann et al. (2019) use CNN and text mining tools to classify twitter data and find that in particular brand selfies (invisible consumers holding a brand product) lead to high levels of engagement. Dzyabura, El Kihal, and Ibragimov, 2018 leverage machine learning and computer vision to predict product success prior to launch. They show that using the image information they can predict how likely it is for a fashion product to be returned. Zhang et al. (2017) and Zhang et al. (2019) use visual analytics methods to establish the short- and long-term impact of the quality of images of AirBnb listings. Burnap and Hauser (2018) and Burnap, Hauser, and Timoshenko (2019) leverage image mining methods to predict design gaps, and in addition they use these methods to automatically generate new product designs. Finally, Liu, Dzyabura, and Mizik (2020) demonstrate how marketers can use visual analytics to "visually" listen to what users say about brands on social media through their imagery. The works mentioned in this paragraph are either in progress or they were published during the write up of this dissertation, highlighting the recency of this stream of research. Most of them call for more research on visual analytics in marketing.

In this dissertation, I make theoretical and methodological contributions to visual marketing. The theoretical contributions can come from both theory-driven investigation and the data-driven exploration. Choosing the right approach (i.e., theory-driven and/or data-driven) helps us determine how visual analytics can be used to tackle the problems at hand. The methodological contributions to the field of visual analytics in marketing then come from the development of new methods required to answer these substantive questions. I use a combination of methods, such as CNNs and basic image processing to extract information from imagery. Subsequently, I use machine learning and econometrics to relate this information to likes and clicks to establish the impact that visual content has on consumer decisions. Importantly, I don't just quantify the impact of images online, instead I design interpretable methods to establish the types of images or the aspects of certain images that drive their success. The proposed frameworks are scalable and adaptable to any dataset or domain where a consumer is presented with visual content. In the next section, I present an outline of each empirical study that constitutes this dissertation. As the studies were done in collaboration with my supervisors and co-authors, I will sometimes use "we" instead of "I" when discussing what was done in these studies.

1.2 Outline

1.2.1 Chapter 2

AI has proven to be useful in many applications from automating cars to providing customer service responses. However, although many firms want to take advantage of AI to improve marketing, they lack a process by which to execute a Marketing AI project. In chapter 2, I define what Marketing AI entails and I discuss the use of AI to provide support for marketing decisions. Based on the established CRISP-DM framework (Chapman et al., 2000), we create a process for managers to use when executing a Marketing AI project and discuss issues that might arise. We explore how this framework was used to develop three cutting edge Marketing AI applications. We highlight that, within marketing AI projects, most time is spent in the data understanding, data preparation, and modelling steps. Most of the data used in these projects is unstructured, and therefore requires extra processing and modelling before it can be understood and used to generate insights. This is one of the main differences between Marketing AI projects and traditional marketing projects. Chapter 2 highlights the managerial importance of the methods proposed in the subsequent chapters and it provides the reader with the definitions and problem understanding necessary to appreciate the methods and tools presented in the rest of this dissertation.

1.2.2 Chapter 3

Social media channels are becoming increasingly important marketing channels, and recently these channels are becoming dominated by content that is not textual, but visual in nature. Relating textual content to sales and conversions is difficult enough, but visual content is even more difficult to analyze. In chapter 3, I explore how consumers engage with visual content. Specifically, we investigate the role of the complexity of images in creating consumer liking. To carry this out, we mainly use the theory-driven investigation approach by extracting a number of different features of the images posted on Instagram by brands and relating these features to likes on the images. We introduce a framework that uses a combination of basic image processing and deep learning to automatically extract theory-driven visual complexity measures from images at a large scale. In an experiment we establish that the automated measures we construct accurately reflect the perceived visual complexity. These visual complexity measures are then used as variables in an econometric model that explores their relationship to the liking of posts on Instagram. The dataset we use consists of over 150,000 FGC from 650 brands on Instagram. The results show that there is a non-linear relationship between the complexity of images and the amount of likes that they generate from consumers. In general, the framework provides a holistic view of the relationship between the unique aspects of visual complexity and consumer liking of firm-generated imagery on social media. We then provide insights into how this knowledge can be used to generate and choose better social media images.

1.2.3 Chapter 4

On many e-commerce websites, the product image can take up a large space of the search result listing, but the importance of this image in the decision-making process has yet to be studied. In this chapter, I use deep learning to extract information directly from product images and I apply visual analytics to understand the importance of this information during consideration set formation. The framework proposed in chapter 3 uses theory-driven investigation and feature engineering based on visual complexity theory. In chapter 4, I mainly use the data-driven exploration approach, by utilizing transfer learning on pre-trained CNNs and ex-post interpretation using visualization tools and other machine learning methods instead. The proposed framework is applied to an extensive dataset of consumer search for hotels on the website of a global online travel agency. We predict hotel-level click-through rates using image information that we extract with CNNs and find that we are able to accurately predict what hotel will be more likely to be clicked on. We complement these findings using LambdaMART (Burges, 2010) to predict consumer clicks during search and find that on average there is a 10% improvement when we incorporate image information as compared to just the textual and numerical features. In addition, we find that the imagery affects the importance of other attributes such as price, with a decrease in the importance of price by 70% in some locations. Finally, in a neuroscience experiment we show that our results can be explained by the fact that the human brain processes high click-through rate images differently than low click-through rate images. Overall, I present one of the first visual analytic frameworks that can be used at a large-scale to help understand the impact of imagery on e-commerce websites. This framework is adaptable to any online consumer search setting where a consumer is presented with different types of information.

1.2.4 Chapter 5

Chapter 5 serves as a look into the future of visual analytics for marketing. I present another useful application using visual analytics for mining consumer opinions and perceptions directly from image data using an unsupervised learning approach. Unsupervised or self-supervised learning is one of the main research areas in computer vision at this time, and this has important implications for marketing as well. Unsupervised and self-supervised learning methods do not require labels to be provided by humans, which makes them broadly applicable. These methods are excellent resources for theory exploration and new hypothesis generation (Daviet, 2020). Chapter 5 provides an excellent example of exploring new territories related to unstructured data in marketing. After focusing on firm-generated visual content, in chapters 3 and 4, I switch focus to user-generated visual content. Specifically, I look at the visual component of reviews on travel comparison platform TripAdvisor. Mining opinions from online reviews has been shown to be extremely valuable to businesses (Wang, Chaudhry, and Pazgal, 2019; Chakraborty, Kim, and Sudhir, 2019). There has been a surge of research focused on understanding consumer brand perceptions from the textual content of online reviews using text mining methods. With the increase in smartphone usage and ease of posting images, these reviews now often contain visual content. The problem that mining opinions communicated by users through their UGI can be quite challenging. We tackled this problem by proposing an unsupervised cluster method to understand imagery generated by users in their online reviews in the travel industry. Using the deep embedded clustering model (Xie, Girshick, and Farhadi, 2016), we group together similar UGI and examine the average review ratings of these clusters to identify imagery associated with positive and negative reviews.

After training the method on the entire dataset, we map out individual hotels and their corresponding UGI to show how hotel managers can use the method to understand their performance in particular areas of their service offerings based on UGI. The performance in a cluster relative to the population can be a clear indicator of areas that need improvement or areas that should be highlighted in the hotel's marketing efforts. For example, our results clearly show 10 main types of imagery that are generally posted by users. In general, we observe that dissatisfied customers post zoomed in pictures of tangibles in the hotel, such as style features, or furnishing, cleanliness and damages. The satisfied customer, for our New York City data, tends to share images of the Empire State Building. The clustering method can be updated or fine-tuned when new data is acquired or it can automatically assign new images into one of the clusters to identify certain marketing stimuli.

1.2.5 Chapter 6

Finally, in the sixth and final chapter, I synthesize the results of the studies presented in the previous chapters, reflect on the contributions of my work, and suggest directions for future research.

Chapter 2

Letting The Computers Take Over: Using AI to Solve Marketing Problems

Authors: Gijs Overgoor, Manuel Chica, William Rand, Anthony Weishampel

This paper has been published in the *California Management Review.* Gijs Overgoor was the leading author for this study. Manuel Chica was mainly responsible for example 1. Anthony Weishampel was mainly responsible for example 3. William Rand fulfilled a supervisory role for the paper.

2.1 Abstract

Artificial Intelligence (AI) has proven to be useful in many applications from automating cars to providing customer service responses. However, though many firms want to take advantage of AI to improve marketing, they lack a process by which to execute a Marketing AI project. This paper will discuss the use of AI to provide support for marketing decisions. Based on the established CRISP-DM framework, we create a process for managers to use when executing a Marketing AI project and discuss issues that might arise. We will explore how this framework was used to develop three cutting edge Marketing AI applications.

2.2 Introduction

AI is one of the most popular buzzwords in business today, but that is for a very good reason; AI has shown to be a very powerful tool for many marketing applications. AI has been around for decades, but its recent popularity is due to three major factors: (1) the growth of "Big Data", (2) cheap, scalable, computational power, and (3) the development of new AI techniques. In the past, one of the problems with many AI methods was that they required a lot of data in order to train, but before the advent of the big data revolution that data was often hard to come by. Moreover, even when large-scale data was available it would often take way too long to actually train AI models on this data. The development of new high-performance computing systems that can parallelize this process has made that cheaper and faster than ever before. Finally, new AI methods, such as deep learning, have been developed that can take advantage of both large-scale data and cheap computational power at the same time (Darwiche, 2018).

We have already seen the potential impact of AI on marketing, as illustrated by the power of Amazon's recommender systems for product purchases and "anticipatory" 1-hour shipping or Google's ability to automatically pair advertising with content (Conick, 2016). In the near future, AI is expected to make marketing more efficient by speeding up the decision-making process and providing marketing managers with information and insights that they could not develop in any other way.

There has been good academic research into examples of how AI can facilitate marketing. For instance, AI has shown to help out marketing by the use of text mining to help understand online WOM (Netzer et al., 2012; Tirunillai and Tellis, 2014), modeling direct marketing responses using evolutionary programming (Cui, Wong, and Lui, 2006), predicting churn using classification trees (Lemmens and Croux, 2006), and adapting websites automatically to better fulfill customer needs (Hauser, Liberali, and Urban, 2014; Hauser et al., 2009) among many other applications. Yet, there is still a need for more research into how AI can help solve marketing problems (Chintagunta, Hanssens, and Hauser, 2016). For Marketing AI to be truly successful, managers need to be better equipped to understand how to implement a Marketing AI solution, and that is one area that has not been well researched yet.

The goal of this paper is to explain how a manager should tackle applying AI to a Marketing problem. This is not a how-to paper that will talk about every single AI model that is available and what the benefits of any particular approach are. Instead, the goal of this paper is to present a framework, based on the popular CRISP-DM framework (Shearer, 2000), that lays out the steps to take when using AI to help solve a marketing problem. It can provide a frame of reference for people with substantial

AI knowledge, and it can be a great gateway to the topic of AI for people trying to implement AI in their business. Although seasoned AI experts are not the intended audience for the paper the framework might help them to structure their projects and explain the steps in the process to business leaders or colleagues that lack AI experience. The framework will be explored in three state-of-the-art Marketing AI examples, and the lessons learned from applying this framework will be discussed.

2.3 AI, Machine Learning, Data Mining, and Analytics

It is first important to discuss some basic terms used throughout this paper. Starting with the most obvious question, "What is artificial intelligence?" One classic textbook in the field by Russell and Norvig 2016 defines *artificial intelligence* as the study of the "general principles of rational agents and on components for constructing them." Agents in this context refers to any system which can perceive the world around it in some way and take action on the basis of those perceptions. A rational agent, in Russell and Norvig's explanation, is "one that acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome (Russell and Norvig, 2016)." Russell and Norvig deliberately steer away from using humans as the measure for AI since that can be very hard to define, while rationality is much easier to assess scientifically. Moreover, in many ways this definition is more useful from a marketing perspective, since it emphasizes making the best decision possible under the given information. *Marketing AI* can now be defined as: the development of artificial agents which given the information they have about consumers, competitors, and the focal company, suggest and / or take marketing actions to achieve the best marketing outcome. Some examples of Marketing AI that meet this definition would include: chatbots for customer service, tools that model the potential outcomes of a new marketing campaign, recommender systems that help managers choose content for online marketing, or models that identify latent characteristics of consumers that are predictive of future interactions with the company.

AI has recently become popular because it provides a cheap way to make predictions about complex problems based on examples in historical data that a company might already have. Machines are often able to predict better than humans and they can do it much faster. Especially with significant improvements in computational power and data availability in the last decade the cost of prediction has dropped significantly, leading to a dramatic increase in the popularity of AI (Agrawal, Gans, and Goldfarb, 2018b). A formal definition of machine learning from one of the classic textbooks in the field states: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E" (Mitchell et al., 1997). In other words, if a computer program can improve performance of a task based on an experience with that task and does so because of that experience, then it has learned. In practice, *Machine learning* is "a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty.." (Murphy, 2012). Though AI is broader than just machine learning, most of the applications of Marketing AI are in the machine learning space.

Although they are often used together, AI and machine learning differ from the "classical" modeling approach, often statistical modeling, that is traditionally used in Marketing in several ways: First, machine learning and classical models have different goals. Classical models emphasize causal and explanatory relationships, whereas machine learning focuses on operational effectiveness and prediction accuracy (Agrawal, Gans, and Goldfarb, 2018b). Second, classical models require the modeler to understand the relation and implementation that the variable has on an equation in an effort to best estimate the function output to a certain error, generally based on theory. Econometrics, for example, is defined as "the interaction of economic theory, observed data and statistical methods" (Verbeek, 2008). In contrast, besides providing a cheap method for prediction, machine learning can learn from data without relying on assumptions or rules-based programming and it can often model much more complex interactions between variables (Varian, 2014). This paper will show that Marketing AI solutions are often a combination of AI and classical methods, where AI is used to make predictions or automate processes and traditional methods are used to create an understanding of underlying relationships and mechanisms. For example, in many cases machine learning is used to make predictions about complex data sources and this information is then used as input for an econometric model.

Another term that is often used in related projects is data mining, and this term is important for this discussion because the framework that we are going to use to develop our Marketing AI process approach is actually drawn from data mining. *Data mining* is sometimes referred to as knowledge discovery from databases, and it is the study of identifying patterns in data (Shmueli et al., 2017). To this extent, data mining is often part of machine learning since it needs the patterns identified by data mining in order to create rules to predict future behavior. One particularly exciting example of data mining in marketing is Amazon's retail forecasting methods where they use AI to anticipate product demand far enough in advance that they can make sure the product is stored nearby before the consumer decides to buy, allowing them to ensure 1-hour delivery of their products (Selyukh, 2018). This kind of customer tracking is one example of discovering purchase patterns through data mining.

Finally, how does this all relate to marketing analytics? Schmueli, Bruce and Patel define *business analytics* as "..the practice and art of bringing quantitative data to bear on decision making (Shmueli et al., 2017)." So *marketing analytics* is bringing quantitative data to bear on marketing decision making. AI, machine learning, and data mining are all techniques that can help in making better decisions using data, but there are other related methods, such as classical models that could also be used to develop marketing analytics; thus, though these terms, AI, machine learning, data mining, and marketing analytics, all overlap and relate to each other, they are distinct in their own right.

2.3.1 Methods and Toolkits

This section is mostly interesting for readers with limited knowledge of AI, because we will quickly survey some of the models, tools and platforms that make use of AI or are used for AI. In particular, we will examine the different types of AI models that exist, and the software packages that exist to support the use of those models.

Machines are now better, cheaper and faster at making predictions and there is more data available than ever. Generally there are three ways a machine learns:

- **Supervised Learning**: In this setup, there is data with labeled responses available and the machine learns how to recognize the labels based on the data. The data is split up into training and testing data. The machine learns on the training data and is then evaluated by comparing the predicted labels and the true labels. After the machine is trained and evaluated it can then be deployed and predict the labels based on the new examples that it has not seen before. For example, if you know the historical customer lifetime value (CLV) of a group of customers and a set of characteristics about those customers. The machine can learn a model that relates the characteristics of the customers to CLV. Then, in test time, the machine is presented with customer characteristics and predicts what the CLV of the customer will be. The effectiveness of the model is measured by how well it does on the testing data. Essentially, supervised learning is where you teach the machine by showing examples. Examples of supervised learning include classification, support vector machine and decision trees.

- **Unsupervised Learning**: In these problems, there is unlabeled data available and the machine has no information on what the data represents. The machine will then learn to recognize patterns and similarities in the data. Consequently, it can group certain observations or recognize patterns. In the case of unsupervised learning the machine learns without a teacher. For example, if a firm wanted to cluster their customers, then unsupervised learning could be used to automatically identify the clusters that have the most in common. A manager would still have to figure out what the clusters represent, but this approach can be quite powerful. Examples of unsupervised learning include clustering and anomaly detection.

- **Reinforcement Learning**: Reinforcement learning is similar to unsupervised learning except now the machine learns by getting some feedback after taking actions. The machine takes actions based on a predicted reward structure and learns by adjusting strategy based on the difference between the predicted and realized output of an action. Reinforcement learning means the machine learns by trial and error, but the trials that it attempts are guided by the model and the algorithm. An example would be trying to learn what order of ads to show to a customer in order to encourage them to make a purchase. Eventually, at each time point the model has to make a decision about what ad to show based on the interest the consumer has shown so far, but it does not know if it has made the right decision until the customer makes an actual purchase. Examples include Q-learning and adverserial networks

Machine learning methods do not necessarily only learn in one of these three ways. It is possible for a machine learning method to be used for supervised, unsupervised, and reinforcement learning. For instance, a method in machine learning that has recently become very popular due to the increase of data availability and computer power is deep learning. Deep learning methods have shown to work well for all three categories of machine learning. There are various deep neural network structures that apply to different types of data that work particularly well because of the architecture of the network. For example, convolutional neural networks work well for image classification, whereas recurrent neural networks are used for sequence models, such as time series(Goodfellow, Bengio, and Courville, 2016).

Another important aspect of machine learning is whether the learning occurs online or offline. The main difference between online and offline machine learning is that with online machine learning the model learns based on one incoming observation at a time whereas offline learning uses all the available data at once to learn a model across all of the data. Overall, online machine learning can be faster and more efficient, however the accuracy compared to offline learning is often lower (Burlutskiy et al., 2016). For example, it is possible to do online deep learning, but deep learning generally requires a lot of offline training time and data to get the desired accuracy. The examples presented in this research mostly concern the use of offline machine learning. The machines learn offline (i.e., they are pre-trained on offline data) and are then deployed to make predictions on new data. Even though they make predictions on newly incoming online data the learning has occurred offline. Ideally, the end result of the examples discussed will be that after deployment the machines are able to learn online and in real-time as well, or the offline models are updated on a regular basis using the new data. The different categories of learning can occur online as well as offline, but some categories lend themselves better for offline learning, such as supervised learning, and others generally work well online, such as reinforcement learning.

Complementary to machine learning, statistical models (described earlier as one of the "classical" models) play an important role in Marketing AI projects. Statistical models require a more theoretical understanding and intuition behind methods, the variables and their relationship to the business problem. Examples of statistical models include econometric methods as simple as regression analysis, but also include methods, such as support vector machines and causal state modeling. Regression analysis is a group of mathematical procedures for studying the relationship between an outcome variable and a set of explanatory variables (Verbeek, 2008). Support vector machines are statistical models that are used to analyze data for classification and also regression (that is, support vector regression) (Cristianini, Shawe-Taylor, et al., 2000). Support vector machines are an example of a statistical model that uses machine learning to classify by optimizing their parameters to allow the maximum separation between the classes, being also able to solve regression problems. We will explore causal state modeling in more depth in one of the examples below, but causal state modeling builds a statistical model, similar to a Markov model, from historical sequence data.

In addition to statistical models and machine learning, another tool that is often used as a component of AI is computational modeling, often agent-based modeling (ABM) in particular. *Agent-based modeling* is a model where rules of behavior are written for autonomous agents that have their own properties and behaviors, and then those agents are embedded in a computational environment where they can interact (Rand and Rust, 2011; Wilensky and Rand, 2015). ABM by itself is not AI. ABM is a modeling framework, just as game theory is a modeling framework. ABM is a way that you can describe a system, but does not necessarily need to use any form of AI. However, ABM and other computational modeling frameworks can use AI. Machine learning, for example, can be used to optimize and calibrate agent actions and interactions with each other and the system in an agent-based model, which we will explore in one of the case studies below. Because both computational models and AI are written in computer code and often used in Marketing to understand human behavior they are sometimes conflated. They are distinct, but related methods.

In the end, building any Marketing AI requires a great understanding of the business, the data and the methods, which makes human judgment an ever important part of AI (Agrawal, Gans, and Goldfarb, 2018a). Judgment is necessary to determine the trade-offs of certain actions and to determine what makes an accurate prediction. Our framework provides the right questions to ask at the different stages of a Marketing AI project for effective judgment.

We will go into more detail about several of these types of AI in the case studies described below. Now that we have discussed some of the AI methods and techniques we will discuss implementation. There are three ways AI can be implemented:

- **Writing code from scratch**, using a programming language that works well for the task at hand. Popular examples are R and Python.

- **Using prebuilt packages or libraries** in programming languages such as R, Python, MATLAB and SAS. This is similar to writing from scratch, except you make use of pre-coded functions and scripts for a family of AI methods that have been previously written and distributed. Examples of these packages are:

  Scikit-learn (Pedregosa et al., 2011): a machine learning package for python that has several methods and libraries built in such as: classification, regression, clustering, model selection and data pre-processing.

  mlr (Bischl et al., 2016): the scikit-learn equivalent for R.

  Rpart (Therneau, Atkinson, and Ripley, 2015): an R package for recursive partitioning and regression trees.

  Dplyr (Wickham et al., 2016): an R package focused on data wrangling. This is mostly used for data processing and structuring.

  TensorFlow (Martin Abadi et al., 2015): an open-source machine learning and deep learning framework that runs on python.

  Keras (Chollet et al., 2015): A high-level neural networks API, written in python and capable of running on top of TensorFlow. Keras makes it easy to build deep neural network architectures by using pre-trained models or standardized layers of the networks.

  SciPy (Jones, Oliphant, Peterson, et al., 2001–): A library that combines several AI related packages for python. It includes an interactive console called iPython.

Machine Learning Toolbox : A toolbox for Machine Learning in MATLAB.

SAS: SAS, one of the main business analytics framework in the industry, has machine learning and AI related methods built in.

Tableau: A visualization software that is very convenient for making raw data understandable and visual.

- **Use "plug and play" software** that provide a user friendly tool to implement the methods described.

  Weka (Hall et al., 2009) and **Orange** (Demšar et al., 2013): These open source software tools, both developed in Java, are two of the standard tools in academia for machine learning and data mining. They have a comprehensive set of algorithms for supervised and unsupervised data mining tasks as well as pre-processing, post-processing, and visualization. Additionally, the user can use its visual work-flow to design the experiments and manage their own datasets. An API can be used to link the algorithms from the user's own code.

  Knowledge Extraction based on Evolutionary Learning (KEEL) (Alcalá-Fdez et al., 2008): KEEL is an open source platform developed in Java that makes use of AI methods (i.e., evolutionary algorithms and fuzzy sets theory) to evolve machine learning algorithms such as classification, regression, clustering, or association rules.

In general, coding your own models and tools from scratch allows you the most flexibility for tailoring your model, while using the plug and play approach gives you the least flexibility, with the pre-built tool kits approach being somewhere in the middle. However, the trade-off is development time. The plug and play tools can be used on a new dataset in very little time, while writing your own code from scratch can take a considerable amount of time. There are many additional tools that are available for the implementation and creation of Marketing AI tools, but this list is a good introduction to many of the most powerful tools. In the next section we will move back to discussing the overall framework that can assist managers in creating and carrying out Marketing AI projects.

2.4 The CRISP-DM Framework

The Cross-Industry Standard Process for Data Mining (CRISP-DM) (Shearer, 2000) was not originally developed for applying AI methods to business decision-making processes, but it provides a strong basis for such a framework. In this paper, we will adapt this framework to help decide when and how to use AI and machine learning to solve marketing problems. CRISP-DM was created by a consortium of companies working together in 1996, and though there are other frameworks for the development of data mining solutions (Shmueli et al., 2017), CRISP-DM is the most widely taught and used (Shafique and Qaiser, 2014; Brown, 2015; Onwubolu, 2009). The goal of the CRISP-DM project was to create an open process model to describe a standard approach to use during data mining and analytics. In this way, the CRISP-DM was envisioned to be a *best practices* model of how to conduct data mining work. CRISP-DM was developed as a hierarchical description, so every phase can be unpacked to additional phases and so on, all the way down to the actual implementation of the project, but in this paper, we will only describe the high-level phases, the aspects that relate to those phases, and how they can be used for Marketing AI (Chapman et al., 2000).

CRISP-DM has been used before in many different contexts from manufacturing (Harding, Shahbaz, Kusiak, et al., 2006) to bank fraud (Rocha and Sousa Junior, 2010), and even in marketing contexts (Moro, Laureano, and Cortez, 2011; Gersten, Wirth, and Arndt, 2000), but to our knowledge, it has not been explored substantially in the academic marketing literature, despite calls that marketing researchers themselves employ a more rigorous method of data mining (Garver, 2002). As part of this project, we had originally considered creating a new process model for Marketing AI, but the steps of CRISP-DM are well-defined and well-accepted. We will however adapt the details, and descriptions of the CRISP-DM framework to a Marketing AI context. The major phases of the CRISP-DM process are the following:

1. **Business Understanding**

   The first goal when considering whether to employ Marketing AI in any context, is to *determine the marketing objectives.* What is this marketing action or decision trying to achieve? For instance, in the image selection project for the online travel agency the goal was to increase click-through rate (CTR) for hotel listings, but often the answer is to increase sales. At this point is often useful to *assess the situation.* What is currently being done to achieve the marketing goals? In many cases in marketing, the answer will be that humans are currently making the decisions that we either want the computer to make, or no one is making the decision in any structured way currently, and the computer can help to make decisions in this space. Once this has been done then it is possible to *determine the Marketing AI goals.* How will the success of the project be determined? Once these questions have been answered then it is possible to start to *produce a Marketing AI project plan.* This involves scoping out the rest of the steps described below.

2. **Data Understanding**

   AI, in general, is highly reliant on data. In fact, there is some evidence in the AI world that data is more important than the model (Halevy, Norvig, and Pereira, 2009). Regardless, understanding the data will be critical to any Marketing AI project. The first part of this phase will be to *collect the initial data*. Identify which data is relevant to the project and then *describe the data* in detail, preferably using a *data dictionary*, which is essentially a formalized description of all of the data that can be used to discuss the data among team members who may have different backgrounds.

   Once the data has been collected and described, then it is important to *explore the data*. Often this task is guided by the marketing objectives, so the process revolves around trying to identify what factors of the data are associated with the objectives. For instance, how many conversions are we making per day and what is the average value of those conversions? Usually this exploration is best handled by *visualizing the data* to illustrate and explore patterns. At this time, it is also useful to *verify data quality*. This involves checking to make sure that there is not any missing data, or that the data actually makes sense.

3. **Data Preparation**

   Data preparation is where most time is spent on a Marketing AI project; even more than on the modeling efforts. The first part of this phase is to *select the data*. Which means choosing exactly which data needs to be incorporated into the Marketing AI solution both for development and testing. The data will often need to be *cleaned* at this point. Cleaning involves making sure that all of the data look similar in structure. This could involve removing data which is missing values, or normalizing the data to enable easy comparison between different types of data. Besides cleaning the data it may also be necessary to add to it. In some cases the raw data is inappropriate for modeling and it is better to *construct new data*, which is often done by constructing derived values from the raw data. For instance, taking textual data and tokenizing it (Feldman, Sanger, et al., 2007).

   It may also be necessary to *integrate data*. For instance if the data is spread over many files with different columns, it may be easier and in fact provide new insights to bring all that data together into one table or repository. After that data has been properly integrated the final step in data preparation involves *formatting the data* appropriately. If using an off-the-shelf AI tool, such as Keras or other tools, then it is often the case that the data needs to be formatted in a specific way (Witten et al., 2016).

4. **Modeling**

When many people think of creating Marketing AI, this is the phase that they are actually thinking of. How do we build the model that will help us make a decision automatically? The first aspect will be to *identify the modeling technique or techniques*. This involves figuring out which approach from neural nets to decision trees to agent-based modeling to linear regression best solve the problem being examined. In some cases, the answer is to explore multiple modeling methods, and then assess which one performed the best. Once the techniques have been chosen, the next step is to *generate the testing criteria*. How will the model be assessed? Often this involves taking a set of data and holding it out, and identifying a metric of performance that will be used to assess the model. Unlike statistical models it is often the case that Marketing AI models are built on one set of data and then tested on another. This additional dataset is sometimes called a hold-out set, or a test set. The idea is that if the model generalizes from another data set, often called the training dataset, to the testing dataset then it is more likely that it will also turn out to be useful in data that has not been seen (Murphy, 2012). A dataset is either split in to two datasets, a training and testing set, or into three datasets, a training, validation and testing set. There is no general rule on the best way to split up the data, because of several factors that influence the performance such as the sample size, signal-to-noise ratio, the number of hyper parameters and the general complexity of a model (Friedman, Hastie, and Tibshirani, 2001). We often see a split of about 50-90% to training data, and about 10-50% to testing or validation and testing data. Given enough data, one approach is to use a learning curve, which plots the size of the training set against the accuracy of the model to help determine a suitable training set size (Beleites et al., 2013). The next step is to *build the model* using the training set, and some cases to fine-tune it on the validation set. Often, we are searching for the optimal bias-variance composition (Friedman, Hastie, and Tibshirani, 2001). The bias-variance tradeoff is about finding the right fit to achieve the best prediction accuracy. At times, a model might be too simple when it has very few parameters and it might have a low variance, but the predictions are systematically off of the true value (i.e. there is a high bias and the model underfits the data). A solution is to increase the complexity of the model, the predictions might now be more accurate on average, however the predictions are much more spread out (i.e. there is a high variance and the model overfits the data), because along with the underlying pattern it follows the noise, or outliers, in the data. An optimal bias-variance tradeoff aims for a high prediction accuracy, that doesn't overfit or underfit the data. Once the model has been built it can then be assessed by examining its performance on the testing set. This is the standard by which the Marketing AI solution will be *assessed*. It is important to note that this is an iterative process. After assessment the model can be adjusted and re-trained for improvement and then re-assessed again until the results are satisfactory.

It is also common for Marketing AI projects to go through the steps of data understanding, data preparation and modeling multiple times, because sometimes machine learning can help to to understand or prepare the data, even before a full model is built to help solve the business problem. For example, a first-level machine learning model can be pre-trained on offline data, sometimes unrelated to the project, for it to learn to recognize patterns or make predictions about the complex data formats that the final model will be examining. Subsequently, this pre-trained model is then used to prepare data or make predictions on new data that will be used in a second-level model that helps to solve the business problem. For instance, a neural network could be used for identifying concepts in images, and then those concepts can be fed to another model that relates those concepts to some final outcome, such as engagement with the image. Throughout this paper, we will explore how it is necessary to specialize the original CRISP-DM framework for Marketing AI.

5. **Evaluation**

   Now that we have a model and we have assessed its performance, its time to *evaluate the results*. Have we in fact met the goals that we laid out in the first step of this project? As part of this phase it is often useful to *review the process* that was used to arrive at this model and determine if all the data is still available and can be made available in a way to facilitate deployment of the model. Finally, *determine next steps*. Is there more to be done or is this it? If we have met our goals then it is time to move on to actually deploying the solution.

6. **Deployment**

   The final step is to deploy the Marketing AI solution in a way that will actually increase business value. As part of this it is necessary to *plan deployment* to understand exactly when and how the tool will be implemented. An important aspect of any major change is also to *plan how to monitor and maintain* the tool. If this process has been carried out correctly, then the tool should be well-designed at this stage, but it may become less accurate over time and that needs to be monitored and assessed on a regular basis. A *final report* should also be put together and the whole process should be *reviewed* and takeaways for future similar projects should be discussed and actions should be taken to make sure the process improves every time.

One of the things that is very different about Marketing AI, as opposed to traditional data mining, is that when deployed a Marketing AI can be set up to continually update itself using new data. In the example below of the word-of-mouth DSS the simulation model could be improved automatically after each marketing campaign to reflect the newest results; while in the example of image selection the Marketing AI could continually update what aspects of an image are important.

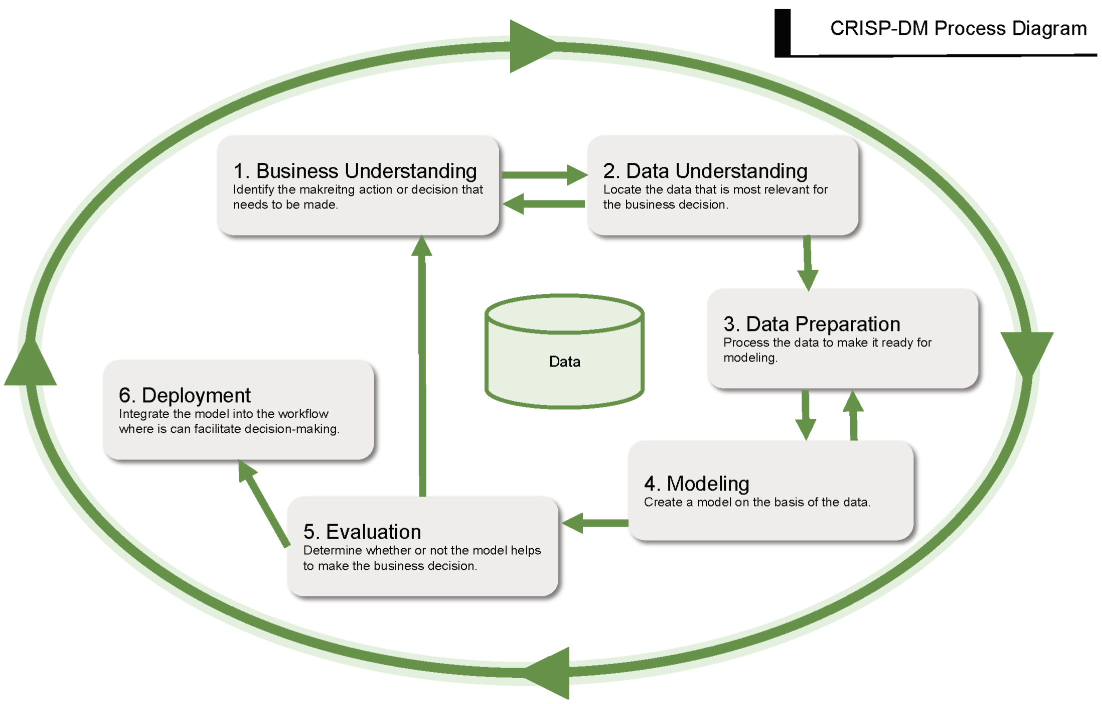


Figure 2.1: Illustration of the CRISP-DM process.

In a common visualization of CRISP-DM these phases are featured as flowing into each other, but as you can see from 2.1 there are some backwards arrows as well.

For instance, early on it is often necessary to iterate back and forth between business understanding and data understanding. As more (or less) data is found than the modelers had anticipated, then it may be useful to revise the scope of the project up or down. Another backwards arrow exists between data preparation and modeling, since it may be necessary to extract additional features from the data to facilitate modeling. Finally, once the model has been built and is being evaluated, it is necessary to make sure it fulfills the goals laid out in the business understanding phase. In many cases even after the whole process is finished that is not the end, instead the whole process will start back over with a new Marketing AI problem, building on the learning from the last project.

These arrows indicate the standard flow for the CRISP-DM process, but at times it is necessary to move and back and forth between different stages for other reasons. We have highlighted a few of those in the examples below, and illustrated them in the accompanying diagrams with dashed arrows.

2.5 The Examples

Now that we have described how to use the CRISP-DM framework to implement a Marketing AI project at an abstract level, we will provide some examples of how this process can be carried out using three distinct, real projects that are on the leading edge of AI applications in marketing.

2.5.1 A Decision Support System for Word-Of-Mouth Programs

Word-of-mouth (WOM) is a powerful force for marketing (Trusov, Bucklin, and Pauwels, 2009) but managers and marketers in an organization need to know how to design and implement their marketing policies using a WOM program to achieve their business target objectives. Some examples of WOM programs and decisions include: (1) balancing WOM with traditional marketing investment (Libai, Muller, and Peres, 2013), (2) designing influencer strategies on Instagram or Twitter (Golkar Amnieh and Kaedi, 2015), and (3) harnessing the positive effects of promotions and incentivization campaigns (Schmitt, Skiera, and Van den Bulte, 2011).

It is difficult to know, in advance, whether a WOM campaign will actually result in increased WOM and viral effects. This is why many companies have started to use decision support systems (DSS) (Lilien, 2011) to help marketers test out WOM programs before the program launches. These DSS can use simulation to provide feedback to marketers and guide them on how best to deploy their marketing strategy.

By using a DSS, marketers can explore and test a wide variety of WOM programs and marketing campaigns, observe their impact in a simulated market, and have more knowledge about the market before starting program implementation.

In this case study we were approached by a massive online freemium app that was interested in knowing what would be the best way to incentivize customer conversions using a WOM campaign (Chica and Rand, 2017). In this section, we will discuss how we used the CRISP-DM framework to use AI to create a DSS for this system. The full example application of the CRISP-DM framework is illustrated in Figure 2.2.

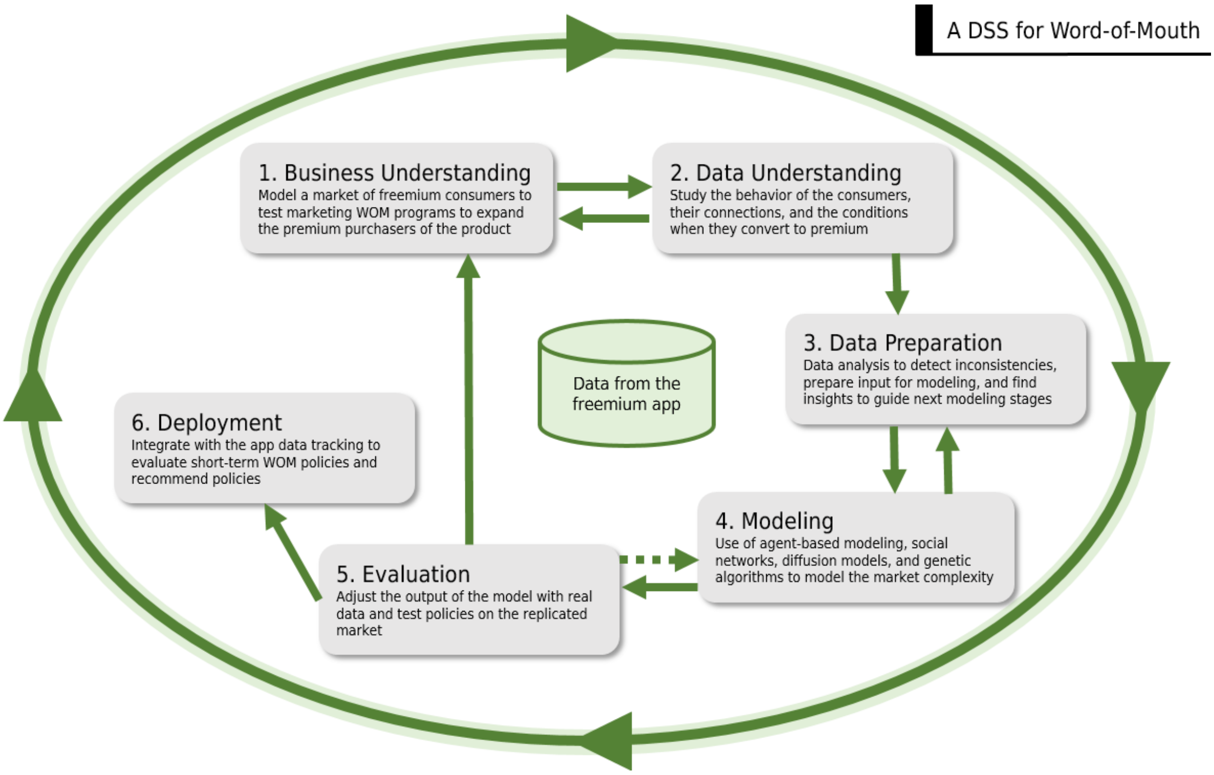


Figure 2.2: Diagram showing the main steps for the WOM DSS Example. Analyze the data, design the model, evaluate, and deploy it in a cyclic process. This is a constant cycle since the deployment of each campaign provides better data for improving the model. Dashed arrow shows an additional link between steps 4 and 5 in this case with respect to the standard CRISP-DM cycle.

Business Understanding

As background, the online game's main revenue stream is from premium conversions of users who can buy items and extra functionality online (Chica and Rand, 2017). Basic users can freely access and play the game and interact with other users, but premium users receive additional benefits such as weekly in-game currency allowances, the ability to adopt virtual pets, access to all the avatars, and premium-only adventures.

The managers of the organization wanted to know whether WOM played a significant role in the adoption of premium services by freemium users, and based on this knowledge, potentially design a reward-based marketing campaign to maximize the spread of positive WOM about premium membership, which might, in turn, increase overall adoption rates and revenue. These marketing campaigns comprise WOM incentivization of users by rewarding policies (e.g., bonus features for the app or gifts). The DSS would need to provide answers about those app users to target: the most likely to convert to premium, just random users, or those who are already premium.

Involving stakeholders of the organization at this stage is important and the modeler should spend enough time to understand the business questions. Otherwise, the next steps of the CRISP-DM process will not meet business objectives and decisions might have to be revisited. This project will be judged as successful if:

1. The model replicates, using the real data from the company, the network of consumers of the app and their behaviors (including the diffusion models of premium adoption).

2. If the overall system is able to evaluate incentivization policies and their effect on premium conversions.

Data Understanding

When all the goals are defined it is important to understand the availability and meaning of the relevant data. The organization (in this case, the company that created the online freemium app) provided information about: (1) some WOM programs they are interested in exploring, (2) data about the actual consumers and their behaviours, (3) data about consumer relationships within the game, and (4) the global key performance indicators of their business.

Specifically, the data provided by the company had three different dimensions. First, the app use behaviour by the users such as daily logins, time from a freemium subscription to acquiring premium subscription, and activity time log. Second, information about the friends (contacts) of the app users and activity between these network connections. And finally, a historical time series of premium conversions as well as new user's registry. The initial data consisted of a data-set of 1.4 million users, with almost 10 million connections between the users (the game supports the idea of "friending" another user). A user was only allowed to have at maximum 100 connections. The data-set was from June 2010 to 2012, and 6.32% of them were premium or became premium during the time of the study.

Data Preparation

In the next stage, we carried out exploratory analysis to understand the data. The output of this stage will be used to create a DSS that better represents the market reality. The main steps to prepare the available data are the following:

- First, we calculated the conversion rate from freemium to premium. Those active accounts (i.e., users playing for more than 10 days) had a premium conversion rate of 16%.
- We tracked the weekly use of the app and we distinguished two kind of days: weekdays and weekends. During weekdays, users are not as active as during the weekend because the app is a game for kids and they have more time during the weekend to play with. Also, we observed seasonality during holidays, but the overall trend was stable.
- The average number of friends of the users (average degree of the social network) is 11.8. Premium users have twice as many friends as free users do. Premium users are also more likely to be friends with other premium users.
- The degree distribution of the social network of users is heavily bi-modal. There is a group of users who have very few friends, and another group clustered under the upper limit of friends, with fewer users in between these extremes.
- Finally, as the DSS will be used for planning short-term campaigns, we extracted periods of time of 2-3 months from the total number of weeks to create and validate the behaviour of the DSS in the next phases instead of considering large periods of time (e.g., one or two years of tracked data).

Modeling

The data preparation of the previous stage is closely related to the modeling stage where mathematical and computational tools are employed to create the DSS. As said, the process is cyclic and there is a feedback between them: modeling requires a specific data preparation while the output of the analyzed data also conditions the modelling techniques to use. The modelling techniques we use in this stage can be grouped into:

- An agent-based model (ABM) framework (Macal and North, 2005) that generates artificial agents to be the real app users of the app. Each agent has a set of behavioural rules (Wilensky and Rand, 2015) that control their activity with the app, or subscription state.

- A social network (Newman, Barabási, and Watts, 2006) that provides the environment in which the agents operate. It was important to make sure that the network replicated the degree distribution of the real app. To do this, we employed an algorithm that randomly generates links between the agents of the ABM framework until the artificial social network has a given degree distribution (Viger and Latapy, 2005). In this case, the algorithm generated a bimodal distribution, as found through the data analysis on the real pool of users of the app. This bimodal distribution was most likely a product of the friend upper limit that the app imposed.

- We include a diffusion mechanism to define and simulate the adoption of premium contents by the users of the app. Given the importance of the social dimension for premium conversions observed in the data analysis, this diffusion mechanism is integrated, together with the activity rules of the agents, as part of the reasoning of the agents. Concretely, two mechanisms of diffusion were modeled: the agent-based Bass model (Rand and Rust, 2011) and a complex contagion (Centola and Macy, 2007).

- The framework is also enriched with an automated calibration using genetic algorithms (Chica et al., 2017), which is an AI-based optimization procedure. This calibration is also related to the evaluation of the model (next CRISP-DM stage) because it searches for the best set of values for the parameters of the model to fit the key performance indicator that outputs the model with the real historical data (conversions from freemium to premium). Additionally, it also helps with the sensitivity analysis of the model and behavioral tests, needed for its validation.

Evaluation

It is necessary to evaluate and show the system goodness to guarantee its business success and acceptance by stakeholders. In our case we examined whether the simulation model captures the reality of the market. It is also important that the system can generate realistic outcomes of a WOM program as the goal is to identify campaigns which minimize the cost of the campaigns and maximize their revenue by means of new premium adopters. A difficulty found when evaluating the deployed system was how to compare the obtained results after applying the policies with respect to the application of other strategies. This is clearly a common problem and therefore we need to rely on simulations' results and past real system behaviors to assess the success of the applied marketing policies.

By using the model built in the previous stage, we evaluate how, by calibrating the parameters with the automated calibration, the output of the model fits with the reality. Deviation measures such as the mean absolute percentage error (MAPE) or the root mean squared error (RMSE) are used to quantify the distance of the model with respect to real premium conversions. We follow a train-test approach when calibrating the system: the automated calibration uses around 80% of the period data and leaves 20% for testing the generalization of the model for the data which were not used during the calibration. For instance, to calibrate the model 60 days of historical premium daily conversions was used as a training set and then 31 days was used for the test data set.

However, automated calibration is not enough to evaluate the behaviour of the system, and techniques such as sensitivity analysis and validation tests are also carried out to study the output of the system. The modeler needs to use automated calibration methods judiciously and in iterative and controlled way in order to manually filter the different alternatives (Chica et al., 2017). Otherwise, if modelers blindly accept the calibrated parameters without studying them, these values will be forced to match the historical behaviour, with the subsequent risk of treating the model as a black box.

Automated calibration is only a step within model validation and should be considered as part of the model building and validation process. Other useful steps to consider to ensure empirical validation are stress tests and sensitivity analysis (i.e., quantifying how "sensitive" the model is with respect to its input parameters). The latter techniques help us explore parameters that are not working properly, or missing features of the modelling. We also employed case studies of incentivization campaigns for users that were premium in order to see how they spread the positive WOM and the implications for increasing the pool of premium users in the artificial market. Sometimes, it became clear that the model needed to be revised and modified (as shown in Figure 2.2 by the additional dashed arrow between evaluation and modeling). If this was the case then a previous stage of the CRISP-DM process has to be revisited to change the model and evaluate it again.

Additionally, the evaluation and first use of the system may also create create new questions for the users and modelers about their marketing campaigns (that is, the business understanding step). In this case, again the ongoing cycle of the CRISP-DM approach will lead to the modification of the modeling, analysis of new data, or creation of new models or sub-models for the DSS. Thus, it is possible to view the CRISP-DM process not as a one-shot process, but rather an ongoing cycle that will consider new needs and objectives that are generated every time the process is carried out.

Deployment

Once the DSS is validated, it could be integrated with the marketing department and a periodic tracking of the obtained premium conversions, new users, and their activity is incorporated to the DSS to continuously update and calibrate the system. Managers can then ask questions about the WOM programs and apply them to the system, collecting their outputs and comparing them with the results in the real market (i.e., an in-market validation). These questions were related to the number of targeted users by incentivization policies and how to select them (the most likely to convert to premium, random uers etc.). Having a DSS with on-the-fly recommendations for managing WOM decisions was a real achievement for the managers of the app because they can anticipate and test their marketing ideas with minimal risk.

A manager's questions can be explored by changing or incorporating new initial conditions to the model (e.g., increasing the incentivization of the influences to spread positive WOM), running the models included in the DSS, and comparing with the baseline strategy (not carrying out the campaign or using the standard one). One important finding during this stage was that increasing social influence between users by rewarding users who adopt premium content has a positive nonlinear impact on increasing the number of premium adopters of the app (i.e., referrals can be quite successful). We also observed how the lift in additional premium members does not demonstrate linear behavior when the social influence is increased by rewarding users at the time of adoption. This observation facilitated the managers' understanding of the dynamics of the premium conversions and will guide future directions for testing and applying their app marketing policies.

2.5.2 Automatic Scoring Images for Digital Marketing

In this project, we used visual analytics and AI to understand the role of images in the decision-making process of consumers booking hotels online. The hotel images are important tools to achieve marketing purposes such as creating brand awareness on social media platforms or to facilitate sales. Currently in many companies, the image selection is done on the basis of 'expertise' or a 'gut' feeling decision. Brand managers often determine based on their creativity and experience what image to select. Recently, a large global online travel agency was interested in automating this process with AI. So, we applied a combination of multiple convolutional neural networks and a support vector regression to score hotel images based on their potential to be clicked on and our algorithm automatically selects the image with the highest potential. In addition, we used the information from our regression model to understand the role of the image in the consumer decision-making process. The main goal for this project

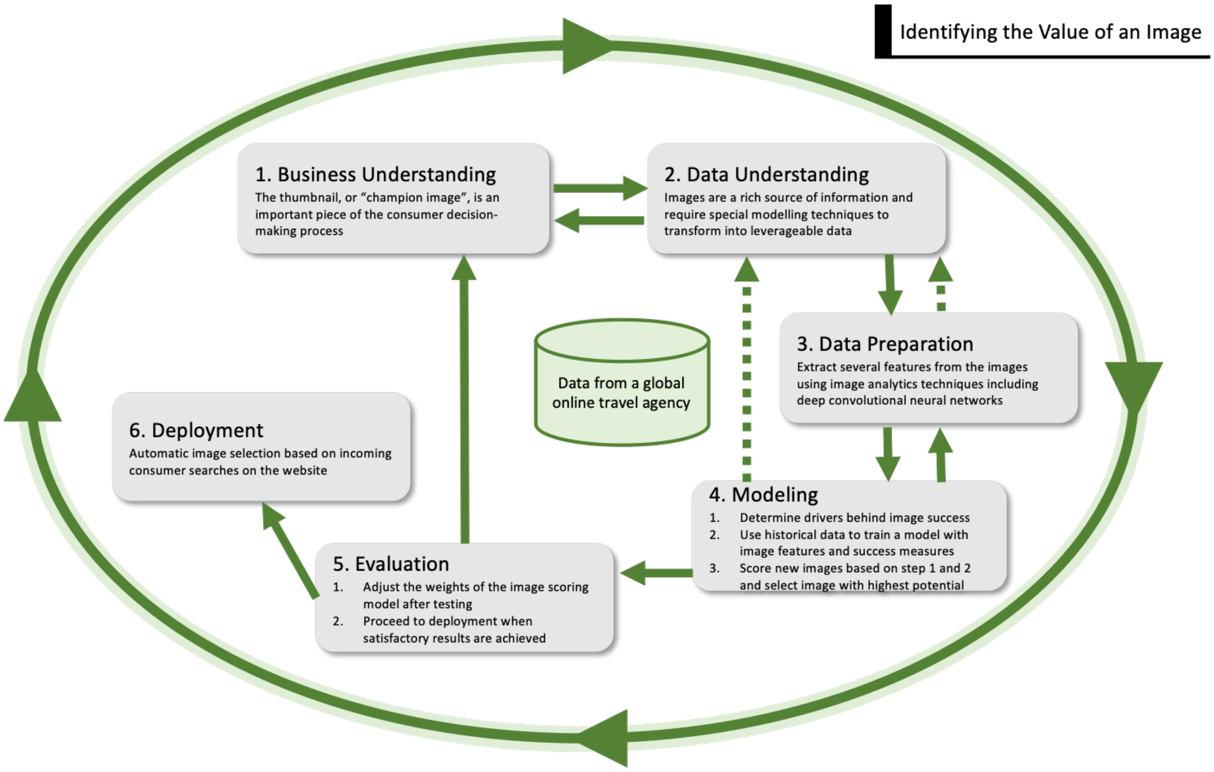


Figure 2.3: Diagram showing the main steps for the Image Scoring Example. Analyze the data, design the model, evaluate and deploy it in a cyclic process. Data understanding, data preparation, modeling happen twice in the process, first for extraction of image information and then image selection. This is illustrated by the additional dashed arrows.

was improving the CTR, but it turns out it is not limited to an increase in CTR. First, by automating the process we saved the marketing managers time. Second, as a consequence of modeling the impact of images on consumer decisions, we now understand what aspects of an image drive engagement online; Third, image scoring and selection algorithm can also be offered as a service to hotels. This example will explore the use of CRISP-DM for this Marketing AI project and is illustrated in 2.3.

Business understanding

The thumbnail, or “champion image”, is the first image of a hotel that a consumer encounters, which makes it an important piece of the consumer decision-making process. The “champion image” is currently still randomly chosen and the online travel agency is interested in automating its selection. The technologies from computer science make it easier to extract information from images at a large scale. The information extracted can then be related to the images’ success measures. The agency wanted to use these technologies to create an image scoring tool that scores images based on their predicted success in engaging consumers. In this section, we will illustrate the process of creating such a tool using the CRISP-DM framework.

As part of this project, the goal was to develop an image scoring tool to increase the CTR for hotels on the website of the online travel agency. We aimed to create a framework that can be used to automatically extract information from images at a large scale, and relate the information extracted to the CTR of the hotel listings. This would then be used to create an image score to automatically select the image with the highest chance of success.

The project will be judged as successful if:

1. We can determine the aspects of an image that make it successful in a way that is reusable for the creation and promotion of future images.

2. The CTR related to the images increases after selecting images based on the image score.

As mentioned, the data is provided by a global online travel agency, so there is plenty of click-stream data and hotel data available. The most important aspect is to have access to the images and to be able to relate them to the consumers' decision-making process for hotel bookings. Since, as is often the case, these data-sets were stored in different locations, we had to work with the company's database specialists to obtain the data from different sources.

Data Understanding

Images are very different from traditional relational data. They are a rich source of information, that can not easily be broken up and stored in a table. It requires special modeling techniques and algorithms to turn images into variables used for analysis. Images in this example consist of 480x480 pixels described by an R, G and B channel with numbers ranging between 0-255, that is 480x480x3 = 691,200 numbers representing a single image. After processing we are able to map this data to a series of features that describe the image.

The CRISP-DM framework requires exploration of the data to create an understanding. Images require complicated modeling techniques to extract useful information, which makes data exploration before preparation difficult. We use pre-trained convolutional neural networks (CNNs) to extract a rich feature set from the images. Essentially, we use these CNNs to make a prediction about what is portrayed in the image and use that prediction as input for our models. In this Marketing AI setting the Data Understanding and Data Preparation are a combined process and even in the early stages different modeling techniques are evaluated to make the most accurate prediction about the images. First we follow the data understanding, data preparation, modeling steps to understand the images and then we follow the same three steps again for the image selection process. This process is illustrated in 2.3 with the additional dashed arrows between steps 2, 3, and 4. These dashed arrows reflect necessary deviations from the standard CRISP-DM framework.

We gathered the “champion” images for all hotels. The “champion” image is the thumbnail that is presented next to the hotel information on the search result page after searching for a destination. We also gathered other hotel information, including aggregate clicks from search result page to hotel page, price of the hotel, number of stars, average customer review, distance to downtown, and related fixed effects. In addition, we gathered customer level click-stream data that consists of individual searches of consumers on the website. A customer can decide to click on a hotel listing or use one of the actions to sort or filter the search result to set preferences. Naturally, we follow the data exploration step here as well to try to gain an understanding of the behavior of consumers on the web-page.

Data Preparation

We extract two types of information from the images: visual complexity and semantic information. Visual complexity captures the complexity of an image, which reflects the overall variation of several aspects of an image. Semantic information covers what is depicted in an image.

The visual complexity of an image is measured by examining the variation at a pixel-level. For example, we look at how much variation there is between every pixel in the image for the colors and brightness. The amount of detail in an image is captured by finding the edge density, and the visual clutter of an image (Rosenholtz, Li, and Nakano, 2007). The last element of visual complexity that we consider is the number of objects. We use a CNN, this is a type of deep learning structure that best learns image information, to detect the objects and then we simply count how many there are in the image.

For the semantic information of images we classify images by using two CNNs. The first one (He et al., 2016) returns a distributional representation of 1000 common objects detected in the image, such as cars, people, and animals, while the second one (Zhou et al., 2018) identifies scene categories, such as beach, hotel room, and library. Once this is done, we have all the information we need from the images ready for analysis and it is time to connect the information from the images with the related metadata. We want to understand the incoming online travelers and understand what information they come across, but most importantly what images they see and what makes them click on hotel listings. We connect the weblogs, hotel information and the images to construct the final data-set.

Modeling

Data preparation and modeling go hand in hand as some essential steps in understanding the success of images requires extracting the information from the image. The next step is understanding the aspect of images that make them successful and to use this information to select the "best" image. There are three stages in the modeling process: First is to model the historic data that we have already gathered. We use this historic data to understand what drives the clicks after the search result for hotel bookings. Second, we use the historic data to train our model to recognize the image with the highest potential for success. Third, we use the trained model to score new incoming images and predict their success.

A regression model allows us to understand the relationship between the CTR and the images. We control for other factors by adding the other variables in our data-set that also impact CTR for online travel search. The second step requires us to create image scores. We can use the results from the regression model and/or a separate machine learning model to assign weights to different aspect of images to determine their importance. The machine learning algorithm will then determine the optimal images to show for an incoming search to maximize CTR.

Evaluation

To test the effect of image score on a particular hotel and as a general model we need to carry out an experiment. This can be done in an experimental setting where for a certain location or a certain type of booking we randomly split incoming searches into two possibilities using an A/B test: (A) control group with no changes, and (B) treatment group where we show the optimal image based on the image score and keep the rest the same. We are then able to determine if there is an increase in the CTR for this particular hotel. An increase in CTR means the image scoring method works.

This study highlights some of the aspects of images that make them successful. For instance, are more colors better? Depending on the destination, do consumers like images of beaches or pools better? This information can be used for image design purposes and can help create more effective images. As the images that are used get better and data is collected about CTR on those images, then that data can be used to improve the model even more, creating a model which improves itself over time. We constantly make predictions about what images and what aspects of images work best for an incoming consumer and with the immediate feedback (clicked or not clicked) we could continuously update the model.

Deployment

A successful deployment of the data mining results requires integration of the model onto the website of the global travel agency and this requires the right flow of data. First, it requires the web developers to integrate the image scoring model into the website, such that for every consumer coming in the right images for the right hotels are shown. It is also essential to have continuous improvement of the model and the weights that determine the importance of the aspects of images to ensure the image selection stays optimal and up-to-date. The deployment should lead to a significant improvement in CTR.

A successful implementation of this automatic image selection would also mean that after going through the different steps of the CRISP-DM framework, the system itself becomes a self sustainable mechanism. In other words, a true AI image selection tool should be able to go through the data, modeling and evaluation steps with little human guidance. It would automatically optimize the system and it should make for easy human judgment to complement all the predictions. That is in many ways the true goal of Marketing AI.

2.5.3 Prioritizing Customer Service on Social Media

Social Media is an important marketing channel for companies to directly engage with consumers, and is serving an increasingly important role in customer service (Rohm, Kaltcheva, and Milne, 2013; Woodcock et al., 2011; Ma and Li, 2019). Rather than merely listening, many firms directly engage with the consumers through social media platforms. This has led to an increase in direct interactions between the firms and individual consumers that has revolutionized Customer Relationship Management (CRM), creating a type of Social CRM (Woodcock et al., 2011). Firms are deploying resources to help respond to these concerns, but most firms do not have the resources to respond to every customer service comment that comes in over social media, so it is important to prioritize these issues. To effectively determine which users to respond to on social media, we need to determine which features are important and whether we can accurately detect these characteristics.

Ma et al. 2018 showed that it was useful to respond to customers on social media, but that responding to them may encourage future negative word-of-mouth . Thus being able to identify which users should be responded to and allocate the resources to engage with these consumers is very important. Building a tool that can identify these

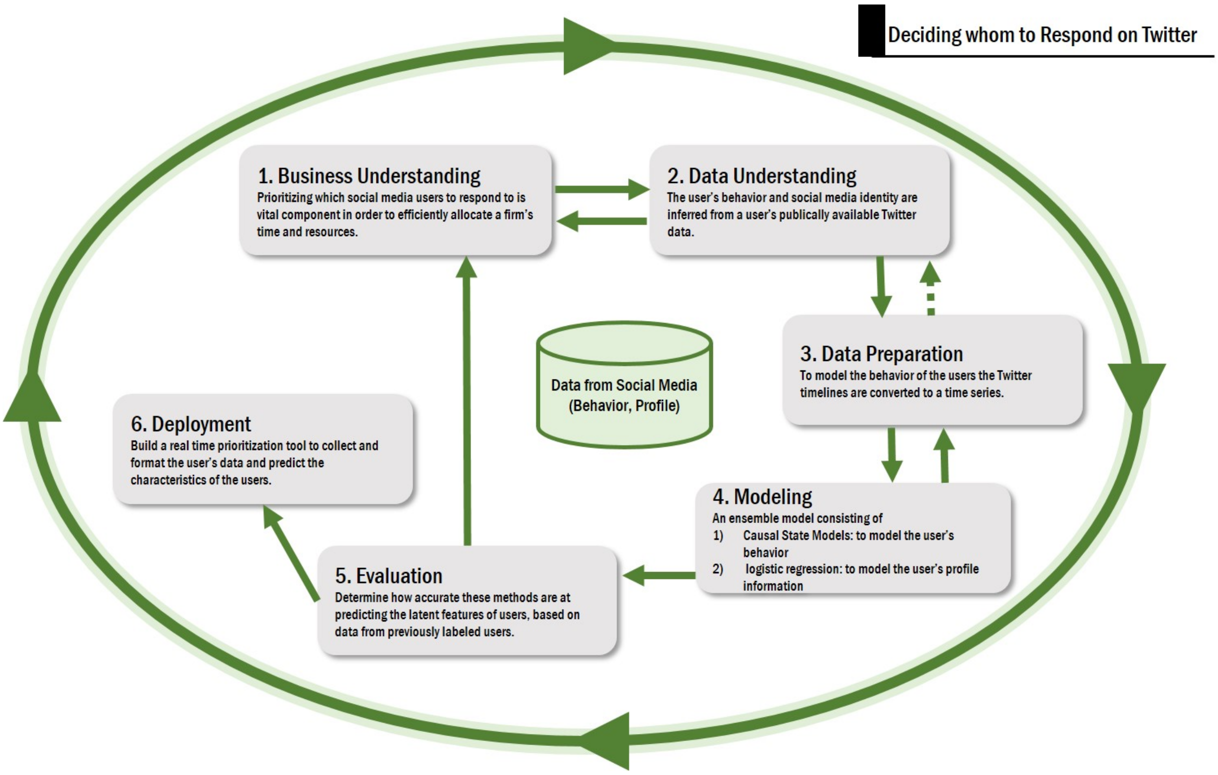


Figure 2.4: Diagram showing the main steps for the Social Media Customer Service Example. Analyze the data, design the model, evaluate and deploy it in a cyclic process. Data understanding, data preparation, modeling and evaluation happen twice in the process, first for collecting and formatting the Twitter data, then applying the model to characterize and prioritize the users.

users in near real-time and provide a more efficient allocation of time and resources. The general goal of this project was to develop a Marketing AI tool to help prioritize customer service requests on social media and is illustrated in 2.4. This marketing AI tool consists of analyzing and summarizing the social media users into a markovian like model and building a supervised learning model to classify and prioritize the users. Unlike the previous examples, this project was undertaken without a focal firm, but instead as a general tool development project.

Business understanding

To break down the overall perspective of prioritizing social media users, it was first necessary to consider what aspects of a user would mean that they are a higher priority. After examining the business objectives at hand, it appeared that there were three features of the users that could be ascertained from social media that could help determine their value to the firm.

- Are they geographically relevant customers, e.g., are they in a market where the firm's product / service is sold?

- Do they have a potentially high customer lifetime value (CLV)?
- Will they provide an overall increase in positive word-of-mouth about the brand?

The first characteristic reflects whether the user can even be a customer of the firm. If a user can not be a customer due to geographical constraints, then the only potential value is related more to brand building than actual sales (*geography*). The second characteristic focuses on the user's potential to be a high CLV consumer (*CLV*). A wealthy user has the possibility of being a valuable customer. The third feature is an estimate of the user's effect on the firm's digital word-of-mouth (*WOM*). It has been well documented that the digital word-of-mouth affects the firm's marketing effects and sales (Chevalier and Mayzlin, 2006; Duan, Gu, and Whinston, 2008; Ma et al., 2018) and is an important feature to include when prioritizing users.

The project will be judged as successful if:

1. We can accurately classify social media users into their appropriate groups based on the user's profile information and behavior.
2. We can use that classification to prioritize social media customer service requests in a manner that optimizes a company's resources.

Data Understanding

This Marketing AI application focuses on Twitter. Twitter was chosen because it is one of the most active platforms where consumers voice their opinions about brands. Firms are able to directly respond to users complaints and concerns on Twitter, and many users take advantage of this capability (Einwiller and Steilen, 2015). For each one of these users there are many variables to consider that can be extracted from social media. The approach presented in this example uses only publicly available social media content. Using only publicly accessible data ensures that any firm can collect the data and implement these methods.

For the initial development of the project we decided to focus on the two types of data that seemed to have the best potential for helping with classification: profile information and behavioral data. Profile data is data available on the profile of Twitter public accounts, and includes information such as the number of tweets, number of followers, number of Twitter accounts who follow the user, etc. The behavioral data for a user is their activity on Twitter over time. The benefit of using these data as the independent variables to determining the classification of a user, is that these variables can be collected and formatted for analysis at close to real time, which was important for the business context explained above. The ability to collect this data at near real time allows for the customer service to quickly collect the data, apply the model, and prioritize users.

Data Preparation

To prepare the data, both the profile information and the Twitter timeline of the users need to be formatted. Using the Twitter API these data sets can be collected and formatted relatively quickly minimizing the time between the initial customer service interaction and the firm's response.

The profile information includes six variables of interest. These variables are: (1) *number of tweets* by the user, (2) *number of followers*, (3) *number of followees*, i.e., the number of Twitter accounts that follow the user, (4) *number of statuses favorited* by the user, (5) whether the user is *verified*, i.e., a user that Twitter has confirmed that the user is who they claim to be, and (6) *the number of times the user is listed*, i.e., appears in a public Twitter list. These profile variables are not only publicly accessible but they are also required for all users. Thus all of the users being investigated will contain these variables. These variables can be used as the independent variables to build models to classify the users into the latent groups: CLV, Location, and WOM.

The API gives companies the ability to quickly collect the timeline of the last 3,200 tweets from a user. Using these tweets, we format the data into time series, consisting of whether or not a user tweeted during a two hour window. Once all of these data are formatted, the next step is to develop the model.

Modeling

Once the data are formatted, we can apply the AI methods to classify the users based on the features of interest. In the end, the decision was made to use two separate learning algorithms for the two types of data, profile and behavior data, and then combine the results in to an ensemble model to classify the user. Combing the two types of data improved the performance of the methods compared to a single model.

For the *profile* information it was decided to use a traditional logistic model since there was a discrete class output and static independent variables. For the *behavior* data, a relatively new technique known as Causal State Models (CSM) was used. CSMs have been selected to model the time series as opposed to other time series models such as ARIMA models, because CSMs provide a clear easy interpretable model. The resulting CSM provides an exact model where the researcher can determine the current state of the user based on the previous events. The CSMs for each user are constructed through the Causal State Splitting and Reconstruction algorithm

(Shalizi and Shalizi, 2004), which produces a CSM that is minimally complex while maintaining the maximum predictive power (Darmon et al., 2013). Once the CSMs are constructed for a user, we apply K-Nearest Neighbor method to classify the user into the appropriate level. Using a distance metric based on the differences between the probabilities of events occurring between the two models, we are able to find the K nearest CSMs from the training set of users. The user is then classified by the majority of these K nearest neighbors.

The final classification of the user, which accounts for both of these types of model is a weighted average of the results of these two models. This weighted average improves classification accuracy of the users while accounting for the two different types of data.

Evaluation

To evaluate the proposed methods, 4,776 users, who reached out to a company on a specific day were investigated. For each one of these users, their profile information and Twitter timelines were collected. These users were selected because each one of these users directly tweeted at a firm using the official Twitter handle of that firm (called *mentions* and *replies* on Twitter), on a specific date. The collection of these user's timelines and profile information were collected on the same day in order to ensure the same research time frame for each user. These data were collected seven months after the initial tweets to the firms. This seven-month delay was necessary so we could determine what the activity of the user would be after their initial contact with the firm. To assess the performance of the models in all cases, the dataset was divided into a training set and a testing set with $\frac{5}{6}$ of the users in training set and the remaining $\frac{1}{6}$ users being in the testing set. The models were built on the training set and the testing set was used to determine the performance of the models.

Each of these users were labeled into their appropriate classes for each of the three characteristics of interests. As stated earlier, the three prioritizing latent characteristics investigated by this study are the location of the user, the customer's lifetime value and the brand WOM of the user. The success of the classification method is dependent on which feature was being predicted. The model had the most success in predicting the location of the user and the internet brand WOM of the user. The model struggled with predicting the customer's lifetime value. Despite the lack of success in predicting the customer's lifetime value, being able to detect the user's brand WOM and location of the user provides the firm with useful information when deciding which users to respond to on social media.

Deployment

A successful deployment of this system would involve integrating the model into customer service systems. The brand's Twitter account would have to collect he user's content whenever multiple users reach out to the firm on Twitter. The model will then help classify the user according to the latent characteristics described above. Based on these latent classes and company information, a final tool could be developed that would present customer service representatives with a list of users assigned to them for response prioritized by this information. Each firm might prioritize different aspects of the classes, e.g., one firm might want to prioritize high CLV over WOM and geography. Similar methods could also be developed for other classes of interest besides CLV, Geography, and WOM, if the firm was interested in those classes. The exact same models would not be used, but new models could easily be generated using the same basic framework.

2.6 Conclusion

In this paper, we have shown how the CRISP-DM framework can be utilized to develop AI solutions to marketing problems. We have illustrated this idea with three novel and interesting case studies that are significant advances in Marketing AI themselves, and have elucidated a number of principles and concerns that marketing managers should address when carrying out Marketing AI projects.

There are a few lessons that we would like to highlight for future efforts in Marketing AI. We will break them down by the relevant phase:

- *Business Understanding:* Involving stakeholders of the organization is important and the modeler should spend enough time to understand the business questions to answer. There should be a clear understanding of what the business problem is, why it is an important problem to solve and what a good solution will look like. Otherwise, the next steps of the CRISP-DM process will not be adequate and previous decisions and steps might need to be revisited again.

- *Data Understanding:* Our understanding of data has become more complex as we have developed ways to deal with new forms of data, such as images, text, and video. Often these data forms require going through some of the CRISP-DM steps twice. First, we follow the data understanding, data preparation, modeling and evaluation steps to make predictions about the unstructured data and turn this into features and representations that we can use in our models and then we follow the same steps again using these new features to actually build our Marketing AI solution.

- *Data Preparation:* One of the most important aspects of data preparation is constructing and enhancing the data. Many modeling forms work better when the raw data has been transformed in some way. For instance, causal state models require discretized data, and the logistic models of the image selection example needed some aspects of the images to be transformed into arrays of boolean variables to work properly. If this process is carried out correctly, it is possible to transform data that was not usable into something that can help increase the success of marketing outcomes.

- *Modeling:* The Marketing AI process is cyclic and there is a feedback between the phases. Modeling requires a specific data preparation while the output of the analyzed data also determines which modeling technique is best to use. In addition, modeling is often required to make predictions about complex data formats which in turn is used as input again into another model.

- *Evaluation:* Calibration is only a step within model validation and should be considered as part of the model building and validation process. Other useful steps to consider to ensure empirical validation are stress tests, case studies, and sensitivity analysis.

- *Deployment:* The whole CRISP-DM process might be thought of as a never-ending cycle. Each iteration creates new questions and new possibilities which can be addressed in the next cycle. The goal of Marketing AI is to automate as much as possible the continual refinement of the previous models, so that even as the models themselves answer more and more questions, they are also answering them better and better over time.

In future work, we hope to go beyond the reference model and examples for Marketing AI described in this publication, and instead, develop a full user guide for how to carry out the tasks of Marketing AI, similar to the user manual that is available for CRISP-DM (Chapman et al., 2000). However, we feel that even the level of detail explored in this paper starts to put Marketing AI on a more firm ground. It is quite possible that the tenets and issues explained in this document will need to be reevaluated in the future, but for now we feel that the proposed framework provides a valuable way to approach the development of Marketing AI.

Chapter 3

Simplicity is not Key: Understanding Firm-Generated Social Media Images and Consumer Liking

Authors: Gijs Overgoor, William Rand, Willemijn van Dolen, Masoud Mazloom

This paper is currently under 3rd round review at the *International Journal of Research in Marketing*. Gijs Overgoor was the leading author for this study. Willemijn van Dolen and William Rand fulfilled a supervisory role for the paper. Masoud Mazloom was mainly involved with data collection and initial analysis.

3.1 Abstract

Social media channels are becoming increasingly important marketing channels, and recently these channels are becoming dominated by content that is not textual, but visual in nature. Relating textual content to sales and conversions is difficult enough, but visual content is even more difficult to analyze. In this paper, we explore how consumers engage with visual content. Specifically, we explore the role of the complexity of images in creating consumer liking. To carry this out, we use a number of different features of the images posted on Instagram by brands and relate these features to likes on the images. We use image mining methods, including convolutional neural networks, that can automatically extract interpretable visual complexity measures from images. We show that there is a non-linear relationship between the complexity of images and the amount of likes that they generate from consumers. Our framework provides a holistic view of the relationship between the unique aspects of visual complexity and consumer liking of firm-generated imagery on social media.

3.2 Introduction

Social media platforms are becoming some of the main channels for achieving a variety of key marketing objectives, from creating awareness to facilitating sales (Batra and Keller, 2016; Kumar et al., 2013; Kumar et al., 2016; Colicev et al., 2018; Luo, Zhang, and Duan, 2013). More and more firms actively participate on social media; they create fan pages on various platforms and generate content to improve their social media marketing strategies. According to a study by media agency Havas (2017), 60% of the content generated by brands is declared as "poor and irrelevant or it fails to deliver". As the amount of online firm-generated content (FGC), such as Instagram posts and brand tweets, continues to increase and the overall amount of content pushed to consumers explodes, it becomes more and more challenging to attain and hold the consumer's attention. To create content that is appealing to consumers requires insight in the popularity of FGC. In this paper, we will focus on an important indicator of consumer engagement, namely 'liking' behavior. 'Liking' behavior is when users on social media platforms state that they 'like' a particular piece of content. Liking behavior has been shown to have positive effects on brand evaluations (Beukeboom, Kerkhof, and Vries, 2015) and it can cause positive change in customers' offline behavior (Mochon et al., 2017). Finding the drivers behind the liking of FGC will improve the understanding of consumer interests and behavior. In addition, Kumar et al. (2016) show that liking can further improve brand evaluations and FGC in general positively impacts consumer spending and overall profitability. By effectively using FGC, marketers can also positively influence their consumers' purchase behavior (Goh, Heng, and Lin, 2013; Scholz et al., 2018).

Although recent studies shed some light on the determinants of the liking and engagement with textual content in social media (Berger and Milkman, 2012; De Vries, Gensler, and Leeflang, 2012; Hewett et al., 2016; Stephen, Sciandra, and Inman, 2015), there is very little research on the liking of predominantly visual content. This is remarkable given the growth of visual social media on platforms such as Instagram. It poses a new dimension to the challenges of the marketing manager, whose key concern is to create content that stops the consumer when scrolling through their social media content. Studies on how people perceive scenes (i.e. information that can flow from a physical environment into a perceptual system, such as images through the human eye) show that observers understand and comprehend the visual information of a scene within 100 milliseconds (Potter, 1976; Oliva, 2005). In addition, a study on the relationships between FGC and user-generated content (UGC) on social media (Colicev, Kumar, and O'Connor, 2019) shows that visual FGC on social media has a significant effect on the consideration and purchase intent of consumers and that the effect of vividness, in particular, is even stronger than some of the dimensions of UGC. They also show that FGC positively influences the engagement and helps to build a brand following. So, it is crucial that marketing managers create visual content that is likeable by the observer at the first look. Therefore, there is a need for empirical investigation of what aspects of visual content generate liking to help firms to be more effective with their visuals on social media.

In the literature, mostly in the context of advertising, we see two opposing views on how to best attract attention to visual content; on the one hand it is suggested to create simplicity (Aitchison, 2012; Book and Schick, 1997) and on the other hand, there is an emphasis on complexity (Nelson, 1994; Putrevu, Tan, and Lord, 2004). Visual Complexity Theory (Attneave, 1954; Donderi, 2006) forms the basis of a more in-depth research in the debate between simplicity and complexity and its effect on attitude towards offline advertisements (Pieters, Wedel, and Batra, 2010). The authors show a positive as well as a negative impact for different visual complexity measures. That is, higher visual complexity in terms of basic perceptual features ("feature complexity") decreases consumers' attitude towards the ad and higher visual complexity in terms of design ("design complexity") increases consumers' attitude towards the ad. A recent study by Shin et al. (2019), investigates the impact of different image content features, including visual complexity, on social media engagement. This study finds the exact opposite, where there is a positive relationship between engagement and pixel-level complexity (i.e. "feature complexity") and a negative relationship between engagement and object complexity (i.e. "design complexity").

Inspired by these studies, their contrasting views, the divide we observe in the advertising literature and recent advances in computer science, we aim to empirically explain the effect of visual complexity on the liking of firm-generated imagery (FGI) on Instagram. Based on previous findings, and the notion that visual complexity is not a monolithic construct (Corchs et al., 2016; Nagle and Lavie, 2020), we argue that: 1) the relationship between visual complexity and liking is most likely non-linear of nature, as opposed to the linear effects found in previous studies (Pieters, Wedel, and Batra, 2010; Shin et al., 2019). 2) By dividing the visual complexity into several individual measures we can provide a holistic view and a better understanding of the relationship between visual complexity and consumer liking.

We construct a framework to automatically extract several measures of feature- and design complexity from the images. We use image processing and deep neural networks to obtain complexity scores from the image and then relate these features to consumer liking of the image on social media. After investigating the effects of the individual complexity measures for feature- and design complexity on liking, we will provide managerial implications on how to control and/or develop the FGI to maximize liking. Building on our feature complexity results, we will examine guidelines for modification of the imagery with the use of Instagram filters or photo editing software. Using the design complexity results, we will develop instructions for the composition of a photograph.

Our study makes several contributions. First, by expanding, automating, validating and implementing the visual complexity framework as proposed by Pieters, Wedel, and Batra (2010), we improve current knowledge in visual marketing literature by showing that, when examining the components that constitute visual complexity individually, there are non-linear relationships between the different types of visual complexity and consumer liking and that it is optimal to stay in the mid-levels of feature complexity, while choosing either end of the spectrum of design complexity works best.

Second, our methods for extracting information from visual content on social media create rich sources of information. Since our model provides automated insights into what content is present in different images, it provides brand managers with information on why pictures will be liked by consumers. The image analysis framework that we present can also be informative for future studies on imagery or studies that try to leverage image information. In addition, we show that the individual aspects of visual complexity influence consumer liking above and beyond a wide range of content characteristics, such as photography attributes (Zhang et al., 2017; Zhou et al., 2018), specific types of images (extracted using multiple pre-trained convolutional neural networks), or faces.

Third, we contribute to the need for exploration of the impact of FGI on social media using a visual complexity framework on a large-scale social media dataset, (Hewett et al., 2016). Moreover, rather than stating that particular individual images are popular, we build design and feature insights about why those images are popular.

In the next section, we present our conceptual framework. After outlining the model, methodological approach and variable operationalization, we summarize the results. We conclude with the theoretical contributions of our research and the managerial implications.

3.3 Conceptual Framework

Complexity of images has been studied extensively in different research fields such as psychology, computer science and advertising. Visual complexity has been defined in many different ways, and there is no standardized set of visual complexity measurement. Palumbo et al. (2014, p. 4) best summarize visual complexity as follows: *"Visual complexity is broadly defined as the level of detail or intricacy contained within* *an image (Snodgrass and Vanderwart, 1980). It has been suggested that perceived complexity correlates positively with the amount of variety in a picture (Heylighen, 1997) and that it corresponds to the degree of difficulty people show when describing a visual stimulus (Heaps and Handel, 1999)".* In other words, complexity depends on a variety of image features ranging from basic, unstructured variation to semantic, structured variation.

Measures have been created to help understand images using Visual Complexity Theory (Attneave, 1954; Donderi, 2006). These measures are based on the notion that there is a close relationship between the complexity of a visual stimulus and the size of the file used to store the image (Donderi, 2006). Most images nowadays are stored as computer files using a compression algorithms, such as JPEG or GIF, and the hypothesis is that if there is little complexity in the image, these compression algorithms work better resulting in a smaller file size, and vice versa for images with higher visual complexity. This has ultimately led to the application of Algorithmic Information Theory (Chaitin, 1977; Solomonoff, 1964) to define and measure visual complexity.

Several studies have investigated ways of measuring visual complexity. A popular method for determining the visual complexity of an image has been to derive scores by asking participants to provide ratings of complexity based on a number of scales (Bonin et al., 2003; Snodgrass and Vanderwart, 1980). Palumbo et al. (2014) identify that people's ratings can be confounded and that this way of measuring is only useful for images that have already been produced, and does not provide insight into how to produce images with certain complexity characteristics. The researchers recommend using algorithms as they represent a more accurate and practical solution.

In this study, we use algorithms to automatically extract image information related to visual complexity and some additional information about the content of imagery. Table 3.1 lists all the automated, and interpretable[1], measures included in our framework. We categorize these variables into Feature Complexity or Design Complexity based on the way they are measured. Measures categorized as Feature Complexity measure pixel-level variation of an image, whereas measures categorized as Design Complexity measure variation related to the design and arrangement of objects in the image. This set of variables was derived from a more comprehensive list, Table 3.8 in the Appendix, of all different ways visual complexity has been approximated in

[1] All our main variables of interest are interpretable, hence their selection. The DC- and FC-Control variables are included because these two variables have been related to complexity in the past and can not be measured or approximated by any of the other variables. They are controls, because they are not interpretable or controllable by the marketer/firm. The Content Control variables control for the semantic content of the image, but are not visual complexity measures themselves.

the past. Many of these measures have been used in different studies to approximate perceived visual complexity in past research, but they have not been individually related to liking before and have not been explored in a comprehensive way. In this research, we provide a *holistic* view of complexity and its relationship with liking.

Table 3.1: List of image statistics used in this paper - Feature Complexity (FC), Design Complexity (DC) and Controls. The FC and DC are controls, because they are more difficult to interpret.

Type	**Measure**	**Reference**
FC	Color Complexity	Artese (2014), Corchs (2016), Hasler (2003)
FC	Edge Density	Cavalcante (2014), Corchs (2016), Rosenholtz (2007)
FC	Luminance Entropy	Cavalcante (2014)
DC	Object Count	Oliva (2001), Pieters (2010)
DC	Object Arrangement asg	Oliva (2001), Pieters (2010)
DC	Object Arrangement Irregularity	Pieters (2010)
DC - Control	Region Count	Comaniciu (2002)
FC - Control	Frequency Factor	Corchs (2013), Corchs (2016)
Content Control	Photography measures	Zhang (2017), Zhang (2018)
Content Control	Face Detection	Parkhi (2015)
Content Control	Adjective-Noun Pairs	Borth (2013)
Content Control	Scenes	Zhou (2014)

Cardy and Dobbins (1986, p. 672) refer to liking as "a self-referent evaluative response to a stimulus". This translates into users on social media evaluating a piece of content and expressing their own identification with the content. Research on exposure effects (Moreland and Zajonc, 1977) indicates that liking or affect can precede recognition and can appear without conscious reaction. Additionally, a study on web design showed that people decide whether they like or dislike what they see in 50 ms (Lindgaard et al., 2006). This study also suggests that the liking of imagery may be closely related to overall impressions of design layout, color etc. For websites, the visual complexity of the page impacts a user's initial impression through emotional responses (Deng and Poole, 2010). People's perceptions, preferences and attitudes with regards to visual objects, scenes and displays are also influenced by visual complexity (Machado et al., 2015). Some of the visual complexity measures in our study capture the image features that evoke primary reactions, while other perceptual features that contribute to visual complexity evoke cognitive and affective responses (Palumbo et al., 2014; Pecchinenda et al., 2014; Chatterjee, 2004). Our framework investigates the relationship between these different types of visual complexity and the liking of FGI.

Visual complexity and its effect on the liking of imagery or visual content in marketing has barely been studied, but there are two notable exceptions: Pieters, Wedel, and Batra (2010) and Shin et al. (2019). Pieters, Wedel, and Batra (2010) focuses on visual complexity as a characteristic of an ad and its influence on attitudes towards ads and brands. Shin et al. (2019) focuses on the effect of image characteristics, including visual complexity, on liking and reblogging behavior on social media platform

Tumblr. These studies measure the visual complexity using the JPEG file size, and an additive measure to approximate the complexity related to the objects. They then study the linear relationship between these aggregate measures with the attitude of advertising and liking of Tumblr posts, respectively. Interestingly, these studies have contrasting findings, which can not be explained solely by the fact that attitude toward ads is different than liking behavior. Especially, since both these studies use the same mechanisms to hypothesize the effects. The focal priority of this study is related to these works, however instead of using aggregate measures we investigate the individual aspects of visual complexity and their impact on liking of imagery. We split up the visual complexity into several automated and interpretable measures for the following reasons: 1) Visual complexity is not a monolithic construct (Corchs et al., 2016; Nagle and Lavie, 2020). Instead, it is constituted by many different factors. Previous studies have shown that there are many different ways to approximate the perceived visual complexity (Olivia et al., 2004; Artese, Ciocca, and Gagliardi, 2014; Corchs et al., 2016; Nagle and Lavie, 2020). These investigations highlight that each type or measure contributes to the perceived visual complexity in a different way. 2) Visual complexity as a single construct is difficult to interpret and control. In addition, its impact on consumer liking can be difficult to disentangle. It is not clear how a manager can increase, or decrease, visual complexity as a whole. This is easier to control when visual complexity is split up into individual measures. 3) The overarching categories, feature complexity and design complexity, for the automated and interpretable measures have been shown to influence attitude and liking (Pieters, Wedel, and Batra, 2010; Shin et al., 2019). However, the linear approximations using aggregate constructs limit the implications of the findings, especially considering the contrasting findings of these two studies. We posit that instead of using aggregate constructs, splitting visual complexity up into individual, interpretable measures, and exploring non-linear relationships, we can get a better understanding of the relationship between visual complexity and consumer liking of FGI.

3.3.1 Feature Complexity and Design Complexity

Derived from past research (Pieters, Wedel, and Batra, 2010), we distinguish two categories of visual complexity: Feature complexity and design complexity. These categories relate to the gist of an image (Oliva, 2005). The gist of an image can be defined as the phenomenon that an observer can comprehend a variety of perceptual and semantic information from a view of a real-world environment with just a glance. In other words, the brain quickly makes sense of what we see. Oliva (2005) distinguishes perceptual and conceptual gist, where the former describes the basic image properties the brain uses to provide a structural representation of an image (feature complexity) and the latter includes the semantic information that is inferred while viewing an image or shortly after (design complexity). Furthermore, from a managerial perspective, we view feature complexity as the type of complexity that arises at the moment a picture is taken. It is a set of basic image features that can be modified using programs such as photo editing software or by simply using a filter on Instagram. On the other hand, design complexity is something in direct control of the photographer. For example, the photographer can decide to zoom in or out, or arrange objects or people to his/her preference.

The distinction between feature- and design complexity becomes even more apparent when we study the mechanism that connects them to consumer liking. Feature complexity evokes low-level visual processes and activates early layers in the visual processing system of the brain (Groen et al., 2013). Feature complexity is hypothesized to impact liking behavior through the peripheral route of persuasion (Shin et al., 2019), based on the elaboration likelihood model (Petty and Cacioppo, 1986). Design complexity, in contrast, evokes mid-level visual processes and activates later layers in the visual processing system (Groen et al., 2013). Design complexity is hypothesized to influence the liking through the central route of persuasion (Shin et al., 2019).

3.3.2 Feature Complexity

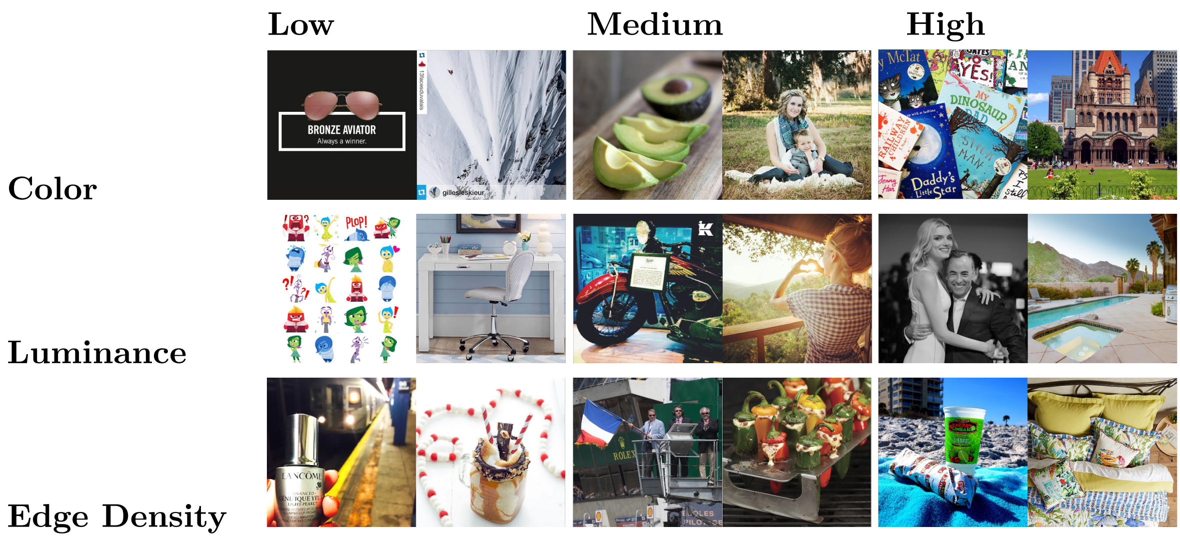


Figure 3.1: Sample images of low, medium and high complexity for each individual measure of feature complexity.

The feature complexity measures are based on low-level visual processes in the primary visual cortex (Palmer, 1999). Feature complexity represents pixel-level variation and unprocessed or unstructured image information without regards to the meaning of the image. An image is perceived more complex when there is more detail and higher variation in (a) color, (b) luminance, and (c) edges of an image. Feature complexity is determined when an image is taken and the basic pixel-level characteristics are encoded.

If an image has a feature complexity that is too high, it can be hard to understand the content of an image, so it is expected to negatively influence attitudes at high levels (Pieters, Wedel, and Batra, 2010). On the other hand, feature complexity can be experienced as a positive peripheral cue that increases physiological arousal and enhance memory (Shin et al., 2019), since images that are too simple may be boring and forgettable. As a result, we expect that there is a non-linear relationship between feature complexity and liking. A certain level of complexity is needed to provide positive peripheral cues, however too much complexity might make it too difficult to process. This negative effect of high complexity may be heightened in an environment such as social media, where a user is scrolling through content quickly, and may not have enough time to process a complex image. For these reasons, we propose the following hypothesis, with respect to feature complexity:

H1: Feature Complexity, composed of (a) color, (b) luminance, and (c) edge density, has an inverted u-shaped relationship with the liking of FGI.

Feature complexity is based on the variation in pixels in an image and can be automatically extracted from the underlying data in the image. Based on algorithmic information theory, research has shown that the minimum length of the code required to describe a visual image constitutes a good measure of its complexity (Simon, 1972; Leeuwenberg, 1969). Image compression techniques (Shapiro and Stockman, 2001) reduce the amount of memory needed relative to the original image by stripping an image of its redundancies, this is standard for image compression (Wallace, 1992). More detail in the basic visual features means more computer memory is needed to store an image. Pieters, Wedel, and Batra (2010) and Shin et al. (2019) use the amount of computer memory (i.e., JPEG file size) as their measure for feature complexity[2]. Although convenient, the file size of an image does not provide information as to what specific visual feature contributed to the complexity or consumer interaction with the image. Additionally, it does not provide content managers with the information to control or manipulate feature complexity to maximize liking.

[2]In Shin et al. (2019), this is called pixel-level variation

Therefore, we will examine the three basic visual features individually that together constitute the main components determining feature complexity of the image: Color, Luminance and Edge Density. We propose measures for the complexity within each of these features to develop a better understanding of their individual effects on liking. See Figure 3.1 for sample images of low, medium and high complexity. The complexity scores will allow content creators to see what basic visual features are helping (or harming) the liking of their content. This will give managers the ability to manipulate images in such a way that they can neutralize the harmful effects of one and enhance the beneficial effects of another. For each visual feature, we describe its relevance with regards to liking and the way its complexity will be measured. In section 3, we show the validation for these measures in an experiment.

Color

The effects caused by colors have been studied extensively throughout the past century in a range of research fields. Color has been shown to have an effect on people's perception (Bevan and Dukes, 1953; Tom et al., 1987), beliefs (Bellizzi, Crowley, and Hasty, 1983) and psychological reactions (Bellizzi and Hite, 1992). Particularly, research has shown that color causes emotions that consequently affect a consumer's attitude towards advertisements (Burke and Edell, 1989). According to Gorn et al. (1997, p. 1): "... an important goal of an advertiser would be to select colors that maximize attention, provide a more realistic and appealing portrayal of the product or service, and arouse appropriate feelings." Gorn et al. (1997) discuss the crowded marketplace in which it can be important for a print advertisement to stand out. In the world of social media, the importance of standing out and not being "crowded out" has become even more relevant considering the vast amount of content that is posted.

Color can be used to generate good feelings and increase the persuasiveness of advertising (Tucker and Smith, 1987). From a managerial perspective, color is manipulated to improve liking of content. Generally, this is done intuitively. Lichtlé (2007) show that color affects emotions and attitude towards the ad and they motivate managers to tweak it to favor attitude towards their ads. However, they also show that colors do not affect all consumers identically and some factors limit the effectiveness of an ad. As opposed to individual colors, Meyers-Levy and Peracchio (1995) give insight into how usage of color variation can influence consumer's attitude towards advertisement. Since most images shared on social media are composed of many different colors, this research seems more relevant to the question at hand. For these reasons we do not investigate individual colors and their effects, but instead investigate the relationship between overall color variation and liking on social media. Considering the rather limitless nature of color variation in an image, we do not expect more color variation to always be better, since too much color variation could make the image hard to process. However, we do expect that there is a certain level of color complexity needed to obtain the attention of the consumer.

Luminance

Luminance is the luminous intensity projected on a given area in a given direction. Luminance in an image is often referred to as the brightness of an image. It is the degree of lightness or darkness of a color (Gorn et al., 1997; Valdez and Mehrabian, 1994) and constitutes a continuous dimension where low-brightness colors appear "darkish" due to black mixed into the pigment and high-brightness colors appear "whitish" and pastel-like in appearance. The difference between luminance and brightness is that luminance can be measured objectively whereas brightness is a human perception. Generally, luminance is included as one of the aspects that constitute the visual complexity of an image.

Brands or products within ads are harder to identify when the differences in luminance are small (Palmer, 1999). Schindler (1986) hypothesized that a majority of firms would employ ads with high contrast in luminance to increase audience attention. Low brightness has also been used to invoke greater trustworthiness (Kim and Moon, 1998). In the space of visual attention, the influence of luminance on fixation duration (i.e., the time the eyes remain fixated on a given location of an image) has been investigated Henderson, Nuthmann, and Luke (2013). This work shows that decreasing overall luminance increases fixation duration. Nuthmann (2017) and Nuthmann and Einhäuser (2015) investigate the factors that influence fixation and show that luminance, among others, significantly impacts fixation and fixation duration. We posit that fixation is closely related to the liking of social media imagery, in the sense that when consumers are scrolling through the content, they construct quick judgments of the gist of the scene to decide whether or not to stop and take a closer look. At this point, the consumer has then already decided whether or not the content is likeable (Lindgaard et al., 2006). Both steps in the process of evaluation by the consumer are influenced by where the eyes are looking, which is directly impacted by the luminance. We follow the authors of these works in constructing a luminance measure to investigate its influence on liking. We expect that there is a certain level of luminance complexity needed for the consumer to fixate on the posted visual content.

Edge Density

The last basic image feature that we will explore is edge density. An edge in an image is a boundary or contour at which a significant change occurs in some physical aspect of an image. Heaps and Handel (1999) had participants rank texture images along several perceptual dimensions including complexity, connectedness, depth, orientation, repetitiveness, and structure. Berlyne (1958) proposed that the number, amount of detail, irregularities of objects, and the irregularity of their arrangement in the image increase complexity. The variation in these perceptual dimensions is reflected by the edge density.

To date, there has been no research investigating the effect of the edge density on the liking of or attitude toward advertisement. However, prior research shows that the effectiveness of an ad depends on the cognitive effort that is needed for the consumer to process an advertisement (Peracchio and Meyers-Levy, 2005). Considering that people automatically try to reduce cognitive effort, by making use of peripheral cues (Petty and Cacioppo, 1986), too much detail can hurt the attention. A high edge density reflects a high level of detail in an image and would therefore require more effort of the consumer to process. Pieters, Wedel, and Batra (2010) show that feature complexity harms both attention and attitude towards advertisement, which can be (partly) due to the edge density within the advertisement. Based on these observations, we do expect the edge density to have an impact, that is not captured by the other features, on the consumer liking. An image needs a certain amount of detail to be likeable, but it needs to stay within the cognitive resources of the consumer (Peracchio and Meyers-Levy, 2005).

3.3.3 Design Complexity

Design complexity of an image captures the complexity of the semantic information of the scene in an image. Design complexity evokes mid-level visual processes based on objects and pattern recognition (Palmer, 1999; Pieters, Wedel, and Batra, 2010). Images with a higher variation in terms of patterns and objects present are more complex. Design complexity may impact pleasure and arousal when viewing an image that directly influences the formation of a first impression (Tuch et al., 2009). Therefore, we expect that design complexity directly impacts the liking of FGI on social media. Higher complexity in design has shown to improve attitude towards advertisements, due to the collative properties of an image (Palmer, 1999; Pieters, Wedel, and Batra, 2010). Other research, however, has found negative (Shin et al., 2019), or non-linear (Geissler, Zinkhan, and Watson, 2006; Deng and Poole, 2010), effects. The main

Figure 3.2: Sample images of low, medium and high complexity for each individual measure of design complexity. Number Objects, Irregularity of Object Arrangement, Asymmetry of Object Arrangement

explanation for these negative findings is that a higher design complexity requires more cognitive effort to understand the "story" of an image. Therefore, we expect that an image either needs to be simple so that the story is easy to comprehend, or be more complex in the design, and thus enhancing the collative properties and enhanced arousal that can make it appealing and likeable (Palmer, 1999; Tuch et al., 2009). We hypothesize the following with respect to the design complexity:

H2: Design Complexity, composed of (a) the number of objects, (b) irregularity of object arrangement, and (c) asymmetry of object arrangement, has a u-shaped relationship with the liking of FGI.

Pieters, Wedel, and Batra (2010) identify six general principles of the design complexity of ads: quantity of objects, irregularity of objects, dissimilarity of objects, detail of objects, asymmetry of object arrangement, irregularity of object arrangement. Subsequently, they add all these up into a single measure for design complexity. Design complexity is calculated from scoring the images manually on these six general principles. Shin et al. (2019) create an automated measure of the total design complexity score, related only to the number of objects, by using the output of a convolutional neural network. Instead, we propose individual measures that capture the key elements of the design complexity of an image and investigate their relationship with liking separately. Although, Pieters, Wedel, and Batra (2010), identify six principles of design complexity, the irregularity of objects and the detail of objects that they discuss are more reflective of pixel-level variation (Feature Complexity), and are closely related to the edge density. For this reason, we do not include these in the design complexity category. Additionally, we find that it is better to combine the quantity of objects and the dissimilarity of objects into a single variable that measures the number of unique objects. This results in three automated, interpretable, measures for design complexity: (a) Number of Unique Objects, (b) Irregularity of the Object Arrangement, and (c) the Asymmetry of the Object Arrangement. See Figure 3.2 for sample images with low, medium and high design complexity in the three features. In Section 3, we show the validation for these measures in an experiment.

Objects

Olivia et al. (2004) show that visual complexity depends (in part) on the quantity and variety of objects. Berlyne (1958) proposed that the number of elements or objects in an image increases the complexity. Design complexity is higher when there are more objects in an image, and when these objects are dissimilar. Objects impact pleasure and arousal when viewing an image that in turn influences the formation of a first impression (Tuch et al., 2009). Object information has shown to be particularly useful for social media popularity prediction (Overgoor et al., 2017; Khosla, Das Sarma, and Hamid, 2014; Mazloom et al., 2016). This research does not emphasize what specific objects are popular, but they use the distributional representation of the prediction for all possible objects as input for the prediction. In other words, what matters are combinations or variations of objects present in an image. A single, or small number of, object(s) clearly conveys what the image is about and can therefore be effective, while a large number of unique objects is expected to benefit the likeability of an image as well (Pieters, Wedel, and Batra, 2010; Shin et al., 2019).

Irregularity of the Object Arrangement

The irregularity of the object arrangement has to do with the spatial arrangement of the objects. In other words, where the objects are in the image. High irregularity reflects a more random distribution of objects across the image (Berlyne, 1958; Pieters, Wedel, and Batra, 2010). Low irregularity represents a regular or organized pattern of the object placement. This means the appearance of one object is easily predicted from its neighbors. Examples of low, medium and high irregularity of object arrangement are shown in Figure 3.2. A regular pattern comes across more organized and therefore less cluttered (Rosenholtz, Li, and Nakano, 2007), and thus positively influence the liking (Pieters, Wedel, and Batra, 2010). The other end of the spectrum can instead represent a very creative design and therefore be appealing to users as well (Rosenholtz, Li, and Nakano, 2007).

Asymmetry of Object Arrangement

An image that is more asymmetrical in its object arrangement is perceived to be more complex. The asymmetry of the object arrangement is reflective of the layout of objects, similar to the irregularity, rather than the individual objects themselves (Pieters, Wedel, and Batra, 2010). The expectations are very similar to the irregularity of the object arrangement, however, we do expect that it can be hard to determine the mid-levels of asymmetry, apparent in the visualization in Figure 3.2. Users are more likely to be drawn to images that have high symmetry or high asymmetry; thus, the ends of the spectrum are expected to be most positively correlated to the liking.

3.4 Empirical Application

To test our visual complexity framework, we have gathered a rich visual social media dataset from Instagram. Before we go into the details of the data collection process and operationalization of the variables we explain our reasoning for choosing Instagram, and at the same time, we will give a quick introduction of some of Instagram's main features and how those features are used in our framework.

3.4.1 Instagram

Instagram allows users to generate content and share this content with other users across the platform. Unlike text-based or mixed media social media platforms, Instagram is considered a visual social media platform meaning that its main focus is visual content - imagery in particular. A user shares (posts) an image with a short description (caption) on their Instagram page. Users can choose to 'follow' other users, in which case new photos from a user they follow will automatically show up in their feed. Typically, users follow dozens, hundreds, or even thousands of other users or brands that are (actively) generating content. The followers can show appreciation of the content posted by 'liking' it, which they do by clicking on a heart-shaped icon, or double tapping on the image. Users can also comment on other users' photos.

After taking a photo, a user has several ways to quickly edit it before sharing it on Instagram. One of Instagram's most popular features is the possibility of adding a filter to a photo. These filters add a certain visual effect to the photo, for example turning the photo into a black and white photo or intensifying shadows and brightening highlights. On Quora[3], Instagram CEO and founder Kevin Systrom describes filters as follows: *"Our filters are a combination of effects - curve profiles, blending modes, color hues, etc. In fact, I usually create them in Photoshop before creating the algorithms to do them on the phone"*. Instagram allows users to take, edit and share a photo within seconds. We perceive that these filters will be relevant for manipulating the feature complexity of the imagery.

Additionally, users can make use of hashtags (a topic marker starting with a '#' character, such as #selfie or #nature) in their description of the photo which allows the specific posts to be found by other users and brands can use it to target a specific audience. This is similar to the way hashtags are used on Twitter to mark the topic of a tweet. Additionally, users can tag other users in the image or in the description, which means that they will get a notification that they have been tagged in that post. For example, if an image is a group shot with multiple people in it, it is common practice to tag those people who have an Instagram account. This means that the post is now not only visible on the page of the user that generated the posts, but it is also visible on the page of the tagged users.

Instagram is one of today's most popular social media platforms with over 800 million active monthly users (Mathison, 2018). Its users have shared over 50 billion photos to date and share an average of 95 million photos and videos per day. They "like" about 4.2 billion posts each day. It has also shown to be a particularly interesting platform for brands. In 2016, almost 50% of US brands were using Instagram for social media marketing and this has risen to over 70% recently (Osman, 2018). A social media study conducted by Forrester (Elliott, 2014) reviewed how the top 50 global brands market on social networks. Forrester evaluated 11.8 million user interactions on 2,489 posts made by 249 branded profiles and collected data on how many top brands use each social network, how many followers they've collected, how often they post, and how often users interact with their posts. They found that the average number of Instagram followers for a top brand in 2016 was already over 1 million. The next section describes the Instagram dataset we have created.

3.4.2 Data

Before collecting the data from Instagram, we selected which brands we would analyze based on their Gartner L2 Digital IQ index.[4] We have selected the top 1000 highest ranked brands based on this index. Subsequently, we have collected all posts

[3]Retrieved from: https://www.quora.com/What-do-the-different-image-filters-on-Path-Instagram-Oink-etc-actually-do/answer/Kevin-Systrom

[4]Retrieved from: https://www.l2inc.com/about/l2-digital-iq-index

of these brands over a 1-year period, starting on 05/01/2015 and ending 04/30/2016. To ensure an equal comparison between brands we have decided that out of the 1000 brands we only include brands that post at least once a week over the focal period, resulting in approximately 150,000 posts corresponding to 633 brands across 27 different industries. This selection was driven by the fact that we intend to analyze overall impact of the complexity measures across all industries. We want to understand the image aspects that drive the liking of images regardless of the brand posting it or the audience it is aimed for.

The posts considered for this study are FGI only, which means they are generated and posted on the accounts owned by the brand. The visual complexity measures in this study and their impact on liking lead to managerial implications, particularly relevant to content managers. Therefore, we have not included user-generated content or online word-of-mouth as these lead to separate implications.

It is important to note that this might be one of the last comprehensive datasets coming from Instagram. After the issues regarding Cambridge Analytica (Wagner, 2018), which happened shortly after our data collection, the API has become rather limited. Currently, to gather such a large variety of posts from different brands in a method that abides by Instagram's Terms of Service would require permission from every single brand being examined (and the users tagged in the post). This also limits the depth of our analysis, since we are unable to go back and gather additional data.

3.4.3 Model Development

We model the liking of FGI by gathering the number of likes for each post and applying a model suitable for count data: negative-binomial (NBD) regression. Here is the model specification:

$$\begin{aligned} log(y_i) = \alpha + \beta_1 Color_i + \beta_2 Color_i^2 + \beta_3 Luminance_i + \beta_4 Luminance_i^2 \\ + \beta_5 EdgeDensity_i + \beta_6 EdgeDensity_i^2 + \beta_7 FrequencyFactor_i + \beta_8 FrequencyFactor_i^2 \\ + \beta_9 Objects_i + \beta_{10} Objects_i^2 + \beta_{11} IrregularityOA_i + \beta_{12} IrregularityOA_i^2 \\ + \beta_{13} AsymmetryOA_i + \beta_{14} AsymmetryOA_i^2 + \beta_{15} NumRegions_i + \beta_{16} NumRegions_i^2 \\ \gamma_1 log(Followers_b) + \gamma_2 TextPositive_i + \gamma_3 TextNegative_i + \gamma_4 BrandControls \\ + \gamma_5 TemporalControls + \gamma_6 PhotographyControls + \gamma_7 ContentControls \end{aligned} \tag{3.1}$$

where the i subscript indicates a particular post. The liking of posts is a non-negative integer with a high variance. It appears to follow a near power-law distribution, something that has been observed in many cases of social media prediction research (Gelli et al., 2015; Khosla, Das Sarma, and Hamid, 2014; Mazloom et al., 2016). The majority of posts receive very few likes whereas a few posts receive up to a million likes. Therefore, we expect over-dispersion in the data. $Color_i$, $Luminance_i$, and $EdgeDensity_i$ are the main feature complexity components extracted from post i, $FrequencyFactor_i$ is added as a control variable that also measures feature complexity. $Objects_i$, $IrregularityOA_i$, and $AsymmetryOA_i$ correspond to the design complexity components.

γ_1 and γ_4 capture the brand-level fixed-effects to control for the variation due to the brands. The number of followers captures the size of the audience and the activity and hashtags measures the frequency of posting and hashtags used. We used specific measures instead of brand-level fixed effects in the form of dummies, because we want to attribute the variation in brand to observed variables.

McParlane, Moshfeghi, and Jose (2014) show how the time of posting affects image popularity on social media. We follow their approach by including three time-dependent dummy variables to control for time of posting - time of day, day of week and the season. Textual information is included as control variables as it is complementary to visual information for popularity prediction (Overgoor et al., 2017). Specifically, we include the positive and negative sentiment scores extracted from the image caption. Finally, we have operationalized 34 content control variables, related to photography, type of image and presence of a humane face. These are extracting using multiple pre-trained convolutional neural networks. The full model as shown in Equation 3.1 achieves our highest observed adjusted R-squared.

3.4.4 Robustness

We compare the negative-binomial model to several options suitable for count data - Poisson, zero-inflated and hurdle regression.

Poisson: A Poisson distribution is parameterized by λ, which is both the mean and the variance (equidispersion). The equidispersion assumption is often violated, because a distribution of counts usually has a variance that is not equal to its mean - especially for social media popularity counts. Performing Poisson regression on count data that exhibits this behavior results in a model that does not fit well. Unlike the Poisson distribution, for a negative-binomial distribution the variance and the mean are not equal. This suggests it might serve as a useful approximation for modeling counts with variability different from its mean. The variance of a negative binomial distribution is a function of its mean and has an additional parameter, θ, called the dispersion parameter. The variance of the NBD can be described by $Var(Y) = \lambda + \lambda^2/\theta$. As the dispersion parameter gets larger and larger, the variance converges to the same value as the mean, and the negative binomial turns into a Poisson distribution. To

test for the most appropriate model we perform a Likelihood Ratio (LR) test between both models. In the presence of Poisson overdispersion the LR test will reject the null hypothesis of theta being equal to infinity. Previous research (Lovett, Peres, and Shachar, 2013; Rooderkerk and Pauwels, 2016) has utilized the negative binomial to model post popularity as well. In our empirical evaluation of the model we observe that the negative-binomial model is indeed a better choice.
Zero-Inflated and Hurdle models: We will not consider hurdle type models, because hurdle models treat zeros as if they come from a separate data generating process. In our case zeros or the positive number of likes come from the same process. In the case of an excessive number of zeros, a zero-inflated model, which is a mixture of Bernoulli probabilities and a count model, would be more appropriate. To test whether there is an excess of zeros, we perform a Vuong test (Vuong, 1989) after modeling both a regular negative-binomial regression and a zero-inflated negative binomial regression. The number of occurrences of zero likes is generally quite low and it turns out that in the empirical application we do not observe excess zeros, a zero-inflated version of our model is not needed.

No re-specification of the model is needed and we proceed to use the negative-binomial regression as described by equation 1. In the next section we describe the empirical application of our visual complexity framework on an Instagram dataset.

3.4.5 Variable Operationalization

Dependent Variable

As discussed above, for this paper we will focus on the consumer liking of imagery by examining how many likes an image receives on Instagram. Likes indicate that the image engaged the user enough that they took the time to click on the image to indicate their interest and subsequently press the like button. It reflects well the first impression and affection consumers have with the image. The dependent variable consists of the total number of likes the image received.

Feature Complexity - Color

Color variation increases the feature complexity of content. We measure the color complexity of an image by describing the richness of the color constitution. We follow (Corchs et al., 2016; Hasler and Suesstrunk, 2003) in their construction of colorfulness of an image. It consists of a linear combination of the mean and standard deviation of the pixel cloud in the color plane. In (Hasler and Suesstrunk, 2003), the authors test a large variety of similar ways to measure colorfulness. We have taken the most accurate representation from this study.[5] First, we transform the image from RGB space to CIELab colorspace. We then calculate the μ_C, σ_a, and σ_b, that represent the mean Chroma, standard deviation along the a axis, and the standard deviation along the b axis respectively. From there we can best estimate the colorfulness of image i as follows:

$$Color_i = 0.94 * \mu_C + \sqrt{\sigma_a^2 + \sigma_b^2} \tag{3.2}$$

Feature Complexity - Luminance

We will construct an entropy measure for luminance variety. First, we extract the luminance by transforming the RGB color space to YUV from which we can calculate the luma value (Y) per pixel. We use the luminance value of each individual pixel to find all unique levels of luminance in the image. Then, we count the number of pixels that contain these levels of luminance to construct the luminance variety entropy measure. The formula describing this looks very similar to the color variety measure:

$$Luminance_i = -\sum_{j=1}^{T} n_j log(\frac{n_j}{N}) \tag{3.3}$$

where T is the total number of unique luminance levels. nj is the count of pixels that contain unique luminance level j. N is the number of total pixels.

Feature Complexity - Edge Density

To detect edges in the image we use the Canny edge detector (Canny, 1987). Every pixel in the image will be classified as either 0 (not on an edge) or 1 (on an edge). As a result, the edge density measure is the total number of pixels on an edge divided by the total number of pixels in an image. The edge density is denoted by the formula:

$$EdgeDensity_i = \frac{e_i}{N} \tag{3.4}$$

where e is the result of the binary classification of pixel i. N is the total number of pixels.

Design Complexity - Objects

Recent advances in computer science have provided us the ability to automatically extract conceptual information from a large number of images. CNNs have been very effective in classifying images (He et al., 2016). More recently, object localization (i.e.

[5] After our validation experiment in the next section this turns out the most accurate measure for FGI on Instagram as well.

detecting and localizing multiple objects within an image, instead of classifying an entire image) has become more accurate. He et al. (2017) proposed Mask R-CNN, using Region-Based CNNs (Girshick et al., 2014) to classify regions of interests within images, to accurately detect objects within an image. In (Nagle and Lavie, 2020), the authors show that this is in fact the most effective individual predictor of visual complexity. Using a pre-trained Mask R-CNN, trained to recognize 81 different types of objects, we are able to count the total number of (unique) objects within an image.[6]

Design Complexity - Irregularity of Object Arrangement

The Feature Congestion measure of visual clutter, proposed by Rosenholtz, Li, and Nakano (2007), measure does not explicitly find objects, but it incorporates certain aspects of perceptual organization, such as grouping by proximity and similarity. When the appearance of one object is easily predicted from its neighbors, then there is a regular or structured arrangement of objects present. For this reason, we find that the orientation clutter reflects the irregularity of object arrangement. Using the code[7] provided by Rosenholtz, Li, and Nakano (2007), we compute oriented opponent energy (Bergen and Landy, 1991), which returns a bi-vector: $(kcos(2\theta), ksin(2\theta))$, at each image location and scale. θ is the local orientation and k is related to the extent to which there is a single strong orientation at the given scale and location. Orientation clutter is computed as the volume or area of an orientation distribution ellipsoid, which is the determinant of the covariance matrix of the bi-vector. The irregularity of the object arrangement is then calculated by averaging over the entire image.

Design Complexity - Asymmetry of Object Arrangement

Using the same feature congestion map, with respect to the orientation, we can calculate the vertical and horizontal asymmetry. Inspired by Zhang et al. (2017), we divide the image into two planes (top and bottom, and left and right, for vertical and horizontal respectively) and compare opposite arrangement irregularity differences. Each pixel is compared to its vertical (horizontal) counterpart. A larger difference represents a larger asymmetry of object arrangement.

[6] Interestingly, in our validation experiment (next section) we find that when asking participants to rate the complexity of images in terms of the number of *unique* objects, the total number of objects, instead of the total number of unique objects, better reflects the perceived complexity. For this reason, we use the total number of objects in our regression analysis.

[7] The authors have provided the MATLAB code at http://dspace.mit.edu/handle/1721.1/37593

Control Variables

Feature Complexity Control - Frequency Factor: The ratio between the frequency corresponding to the 99% of the image energy and the Nyquist frequency (Corchs et al., 2016).
Design Complexity Control - Number of Regions: Calculated using the mean shift algorithm (Comaniciu and Meer, 2002).
Brand Followers: The size of the audience is reflected by the number of followers of the brand posting the images. Upon inspection, we observe that the number of followers is highly correlated with the number of likes. The number of followers of the brand will be included as a brand-level fixed effects. We perceive the number of followers of a brand shows a brand's social media ability and it captures part of a brand's overall popularity. To reduce the variance of the number of variables we have log transformed it for analysis.
Brand Activity:We measure brand activity by using the number of posts that the brand has created on Instagram during the measurement time period. We expect that when brands are actively working on engaging with their consumers it can increase their overall image popularity. Brand activity will be included into the model as part of the brand-level fixed effects.
Time of day / day of week / season: We also incorporate time of day, day of the week, and season of the year control variables. For time of day, we record if images are either posted in the morning (06:00 am to 11:59 am), afternoon (12:00 pm to 06:00 pm), evening (06:00 pm to 11:59 pm) or night (12:00 am to 05:59 am). For day of week, we record if Images are either posted during the weekend (Fri-Sun) or on a weekday (Mon-Thu). Finally, for season of the year we examine if the images are posted during one of the four seasons in winter, spring, summer or autumn.

Additionally, we have information about the post itself that is time-independent and does not directly relate to the brand, but it can help us control for others aspects of the post.
Number of image tags: An image tag is a reference to some other user (person or brand) within the caption or image itself. We do not have information on the user that is tagged, but we do know the number of users that are tagged in the image. The tag itself might lead to an increase in the number of views, because it now includes the audience of the tagged users on top of the brand followers that were already going to be exposed to the content.
Textual Sentiment: Visual and textual information are complementary for popularity prediction (Overgoor et al., 2017). Therefore, we include the positive and negative sentiment scores extracted from the image caption. We use Sentistrength (Thelwall

et al., 2010) to calculate positive and negative scores ranging from 1 (neutral) to 5 (very high valence).

Content Controls: To control for the content characteristics of the image, we have operationalized a large set of image features. First, we constructed 13 photography attributes, using some state of the art image mining methods, as proposed in Zhang et al. (2017) and Zhang and Luo (2018), and added this to the regression. Second, we construct a set of most frequent types of images, and the presence of faces (humans). We operationalized this set of variables using three separate pre-trained CNNs. 1) we detected Adjective-Noun Pairs using the MVSO model (Borth et al., 2013), also used by Rietveld et al. (2020) in their study on User-Generated Imagery on Instagram. From the classifications, we created binary indicators for the top 10 most frequently occurring Adjective-Noun Pairs in our dataset. 2) we detected scenes using a pre-trained CNN trained to recognize 365 scenes/places (Zhou et al., 2018). We created binary indicators for the top 10 most frequently occurring scenes/places in our dataset. 3) we utilized a pre-trained VGG16 architecture CNN (Parkhi, Vedaldi, and Zisserman, 2015) to detect faces in images, creating a binary indicator for the presence of faces. For a full list and description of all these control variables we refer you to the Appendix.

3.4.6 Model Estimation

As in traditional negative binomial regression analyses, we estimated Equation 3.1 by maximizing the log-likelihood function. We normalized all the explanatory variables in the final model, such that their beta coefficients can be compared. We checked for multicollinearity, the variance inflation factors revealed that there is no issue.

3.5 Validation Experiment

Visual complexity has been studied, tested and validated extensively in the studies presented in Table 3.8, but the main focus of these studies has been to estimate the correlation between image features and visual complexity as a single construct. These studies have asked participants to rate imagery based on the perceived visual complexity. Subsequently, the correlation between the image features and visual complexity was tested and combinations of features to best estimate the complexity have been proposed. However, as stated before, visual complexity is not a monolithic construct and seen from the evidence in these studies it is often a combination of features that can also be non-linear in nature. For this reason, and for interpretability purposes, we construct individual measures of visual complexity. The validation of these specific measures has not been done before, and we need to ensure that what we measure is indeed what is perceived and, perhaps more importantly, whether these types of complexity are indeed interpretable. The goal of this experiment is to validate these individual measures by asking participants to rate the imagery on the specific types of complexity.

We follow (Shin et al., 2019) in their assumption that it can be difficult for humans to objectively judge and rate abstract concepts about images. In this sense, it can be difficult to judge an image on our proposed complexity measures. It is more intuitive instead, to view this as a ranking problem. This way we can ask participants to judge pairs of images and select the image that feels most complex instead of asking them to rate an image based on the perceived complexity. For each complexity dimension, the participants can have a reference point to compare against, instead of scoring individual images.

In the validation, we test if the rankings of our automated measures correspond to those of the participants. We have sampled 900 image pairs for each complexity measure. We have sampled 300 images from between the 10th and 35th percentiles, 300 images from between the 40th and 60th percentile, and 300 images from between 65th and 90th percentile. As we are testing non-linear relationships between our visual complexity measures and liking, we want to validate comparisons of low-medium, low-high and medium-high complexity imagery. This way we can not only compare if our measures are accurate in ranking the images, but we can also distinguish if the differences are perceived more easily between different regions of the distribution.

The validation experiment was performed with 289 undergraduate students. For each participant in the survey we randomly drew 35 pairs of images out of the 900 image pairs, for each of the 6 complexity measures. Each image pair was, on average, rated by 10+ participants, to ensure validity of the results. The Cronbach Alpha of our measures was .74, exceeding the commonly accepted threshold of .7 (Nunnally, 1978). The image out of the image pair that receives the majority vote is considered the image that is perceived to be most complex by the participants. We then compare the number of times our selected option agrees with the selected option by the participants as a percentage of the total image pairs. In addition, we investigate the success percentage in the case of unanimous votes. The higher the percentage the better the automated objective measure reflects the perceived complexity per dimension.

The main results of the validation experiment are summarized in Figure 3.3. Overall, we observe that all automated measures are in over 60% agreement with the majority vote of the participants, which is comparable to findings in other studies (Shin et al., 2019). The results highlight that our measures accurately reflect the perceived

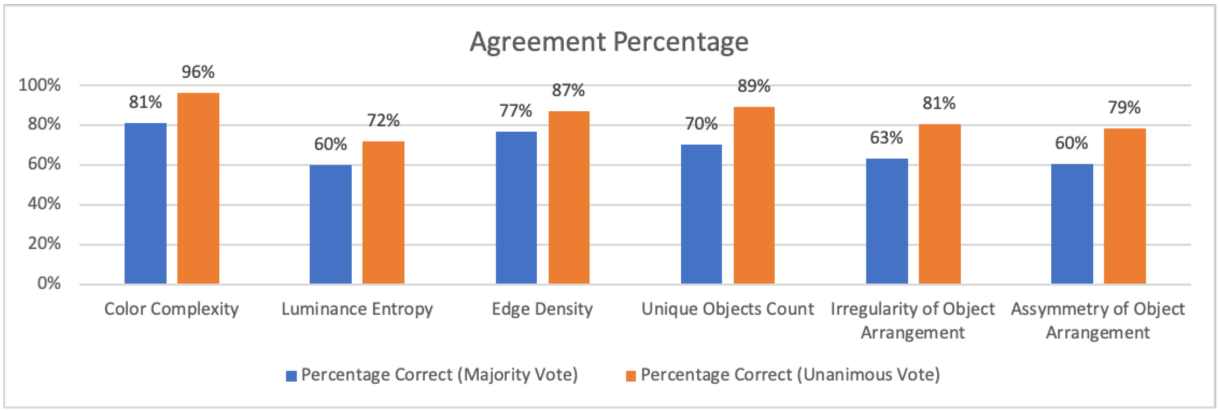


Figure 3.3: Agreement percentages between the predicted scores and the participants' votes. The blue bars represent the agreement between the majority vote and the automated measures. The orange bars represent the agreement between images that received a unanimous vote and the automated measures.

complexity and what they are claimed to measure. The results also highlight that measures such as color complexity and edge density are easier to interpret and detect by participants, than luminance or the asymmetry of the object arrangement. In addition, we observe that the images that received a unanimous vote from the participants on average lead to an increase in agreement percentage of between 10% and 19%, reaching up to 96% agreement for color complexity. This means that our measure can accurately predict the most complex image when the participants of the survey are in agreement.

Table 3.2: Agreement percentages when sampling from different ranges of the distribution: Low, medium and high. In bold are the highest percentages per row.

Type	**Low - Medium**	**Medium - High**	**Low - High**
Color Complexity (Majority Vote)	82%	72%	**89%**
Color Complexity (Unanimous)	96%	91%	**99%**
Luminance Entropy (Majority Vote)	56%	**63%**	61%
Luminance Entropy (Unanimous)	**83%**	67%	69%
Edge Density (Majority Vote)	83%	61%	**86%**
Edge Density (Unanimous)	83%	63%	**92%**
Unique Objects Count (Majority Vote)	67%	71%	**73%**
Unique Objects Count (Unanimous)	86%	87%	**91%**
Irregularity of Object Arrangement (Majority Vote)	62%	56%	**72%**
Irregularity of Object Arrangement (Unanimous)	79%	75%	**87%**
Assymmetry of Object Arrangement (Majority Vote)	47%	66%	**68%**
Assymmetry of Object Arrangement (Unanimous)	64%	73%	**94%**

In Table 3.2 we present the agreement percentages for images pairs where we sampled from different ranges of the distribution, which is relevant for our study as we are investigating non-linear relationships with the liking. We observe that for 5 out of 6 measures the agreement percentage was highest for the low vs. high comparison. This means that when there is a larger difference between the automated complexity scores it is generally easier to judge by the participants. For the luminance entropy, we observe that it was harder for participants to distinguish between images sampled from low and medium ranges. In addition, we observe that for the asymmetry of the object arrangement, the low and medium range images were hard to distinguish, whereas the low vs. high, and medium vs. high resulted in 68% agreement, increasing to 94% in case of unanimous vote. For the rest of the measures the agreement percentages are to be expected, with the highest scores for low vs. high and still large percentage even when comparing the low and high range images to medium range images. Overall, we can conclude that the measures accurately reflect the perceived complexity, and that a bigger difference between the measures makes it easier to distinguish between images. Only low and medium measures asymmetry of object arrangement are not distinguishable, which needs to be taken into consideration in the analysis.

3.6 Results

Table 3.3, shows the descriptive statistics of the variables in the model. The number of likes shows a power-law distribution where the majority of posts receive very few likes and a small number of posts receive a large number of likes. The color complexity ranges from 0 to 18.41, a mean of 3.00 with a large tale on the upper end. The luminance complexity entropy measures ranges from .02 to 2.85 where the majority of posts lie between 2.5 and the maximum. The number of objects detected in the images ranges from 0 to 100. On average there are approximately 17 objects in an image. The irregularity of the object arrangement ranges from .00 to .23 with a mean of .06, whereas the asymmetry of object arrangement ranges from .00 to .92, with a mean of .24. Figure 3.4 visualizes the correlations between our main variables of

Table 3.3: Descriptive Statistics

Variable	**mean**	**sd**	**min**	**max**
Likes	4,138	30,694	0	935,690
Color	3.00	1.75	.00	18.41
Luminance	2.48	.49	.02	2.85
Edge Density	.09	.03	.00	.35
Frequency Factor	.42	.04	.00	.49
Objects	17.38	20.51	0	100
Irregularity of OA	.06	.02	.00	.23
Asymmetry of OA	.24	.11	.00	.92
Region Count	56.51	30.11	0	1,320
Followers	168,751	1,679,190	15	46,098,258
Text Sentiment Positive	1.74	.89	1	5
Text Sentiment Negative	1.23	.56	1	5

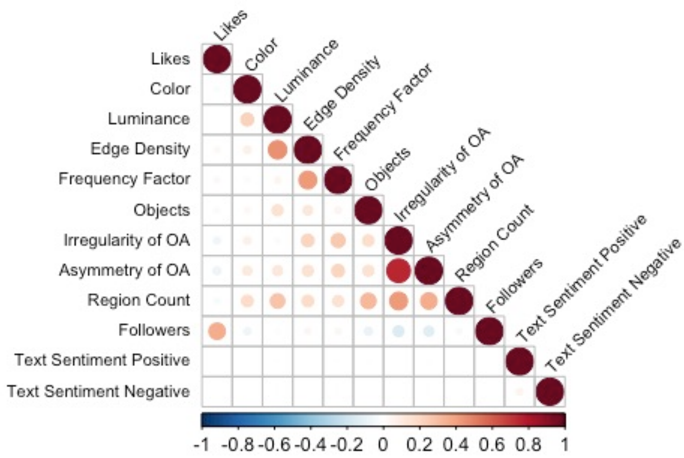


Figure 3.4: Correlations between main variables of interest

interest. The correlations between the visual complexity measures are modest, only the irregularity and the asymmetry of object arrangement are moderately correlated. We have tested for multicollinearity using variance inflator factors and we observe no issues.

In Table 3.4, the results of four different regression analysis are listed. The "Linear-Aggregate" model is to replicate the findings of (Shin et al., 2019). The "Linear-Individual" model highlights that splitting up feature- and design complexity into individual measures offers a more nuanced view of these findings. The "Quadratic-Aggregate" results support hypotheses H1 and H2, where we find that when we use the aggregate measures, as used by Pieters, Wedel, and Batra, 2010 and Shin et al., 2019, the effects are in fact non-linear. The results show that there is an inverted u-shape relationship ($\beta = 1.954, p < .01$ and $\beta = -2.467, p < .01$ for feature complexity and feature complexity squared respectively) between feature complexity and liking, thus we accept H1. On the other hand, we find that design complexity has a u-shape relationship ($\beta = -1.400, p < .01$ and $\beta = 1.716, p < .01$ for =design complexity and design complexity squared respectively) with liking, accepting H2. The effects turn out to be more nuanced, however. The "Linear-Individual" model shows that for the linear findings of previous studies the effects are more nuanced. The full model "Quadratic-Individual" provides a holistic view of the relationship between visual complexity and consumer liking of FGI. The individual effects for feature complexity variables highlight that H1 is indeed fully supported and each individual variable has an inverted u-shape relationship with the liking. That is, we find positive main effects of the color ($\beta = .222, p < .05$) the luminance ($\beta = .221, p < .01$), edge density ($\beta = 1.730, p < .01$) and the feature complexity control variable frequency factor ($\beta = 1.502, p < .01$) and a negative effect of their square terms ($\beta = -.742$, $\beta = -.121$, $\beta = -2.013$, and $\beta = -.512$, with $p < .05$). Initially, the relationship between liking and these measures of feature complexity are positive, but when they increase, it prompts decreasing returns for the liking. Thus, we accept H1a (color), H2b (luminance), and H2c (edge density), there is an inverted u-shape relationship between the individual components of feature complexity and liking of social media imagery.

Table 3.4: Negative binomial for 4 different specifications of visual complexity. The first two columns are linear estimations using a linear combination and individual specification, respectively. The second two columns are quadratic estimations using the same two specifications.

	Linear-Aggregate	Linear-Individual	Quadratic-Aggregate	Quadratic-Individual
Feature Complexity	$.590^{***}$ (.023)		1.954^{***} (.062)	
Edge Density		$.461^{***}$ (.028)		1.730^{***} (.085)
Luminance		$.139^{***}$ (.015)		$.221^{***}$ (.073)
Color		$-.179^{***}$ (.043)		$.222^{**}$ (.089)
Frequency Factor		$.707^{***}$ (.031)		1.502^{***} (.245)
Feature Complexity^2			-2.467^{***} (.103)	
Edge Density^2				-2.013^{***} (.126)
Luminance^2				$-.121^{**}$ (.050)
Color^2				$-.742^{***}$ (.153)
Frequency Factor^2				$-.512^{***}$ (.155)
Design Complexity	$-.413^{***}$ (.025)		-1.400^{***} (.090)	
Objects		$-.070^{***}$ (.010)		$-.177^{***}$ (.029)
Irregularity of OA		$.147^{***}$ (.044)		-1.143^{***} (.179)
Asymmetry of OA		$-.702^{***}$ (.039)		$-.304^{***}$ (.108)
Region Count		.026 (.102)		.178 (.141)
Design Complexity^2			1.716^{***} (.155)	
Objects^2				$.111^{***}$ (.032)
Irregularity of OA^2				1.789^{***} (.236)
Asymmetry of OA^2				$-.589^{***}$ (.199)
Region Count ^{2}				$-.339$ (.437)
Log(Followers)	$.922^{***}$ (.001)	$.920^{***}$ (.001)	$.921^{***}$ (.001)	$.921^{***}$ (.001)
Text Sentiment Positive	$.004^{**}$ (.002)	$.004^{**}$ (.002)	$.004^{**}$ (.002)	$.004^{**}$ (.002)
Text Sentiment Negative	$-.003$ (.003)	$-.003$ (.003)	$-.003$ (.003)	$-.003$ (.003)
(Intercept)	-1.502^{***} (.036)	-2.006^{***} (.049)	-1.532^{***} (.039)	-2.373^{***} (.101)
Photography Controls	✓	✓	✓	✓
Content Controls	✓	✓	✓	✓
Temporal Controls	✓	✓	✓	✓
Brand Controls	✓	✓	✓	✓
Observations	147,963	147,963	147,963	147,963
Adjusted R-Squared	.438	.452	.446	.453
Note:				*p<.1; **p<.05; ***p<.01

The top row in Figure 3.5 visualizes the effects for the feature complexity measures. We observe clear inverted u-shape relationships for each of these variables with consumer liking. Each variable was normalized using a min-max normalization to be within 0 and 1. The global maxima for each of these functions lies within the domain of each variables, with global maxima of .15, .91 and .43 for color, luminance, and edge density respectively.[8] We observe a gradual drop off for increasing values of color, and a gradual drop off for decreasing values of luminance. The edge density has a steep drop-off on either side of the maximum.

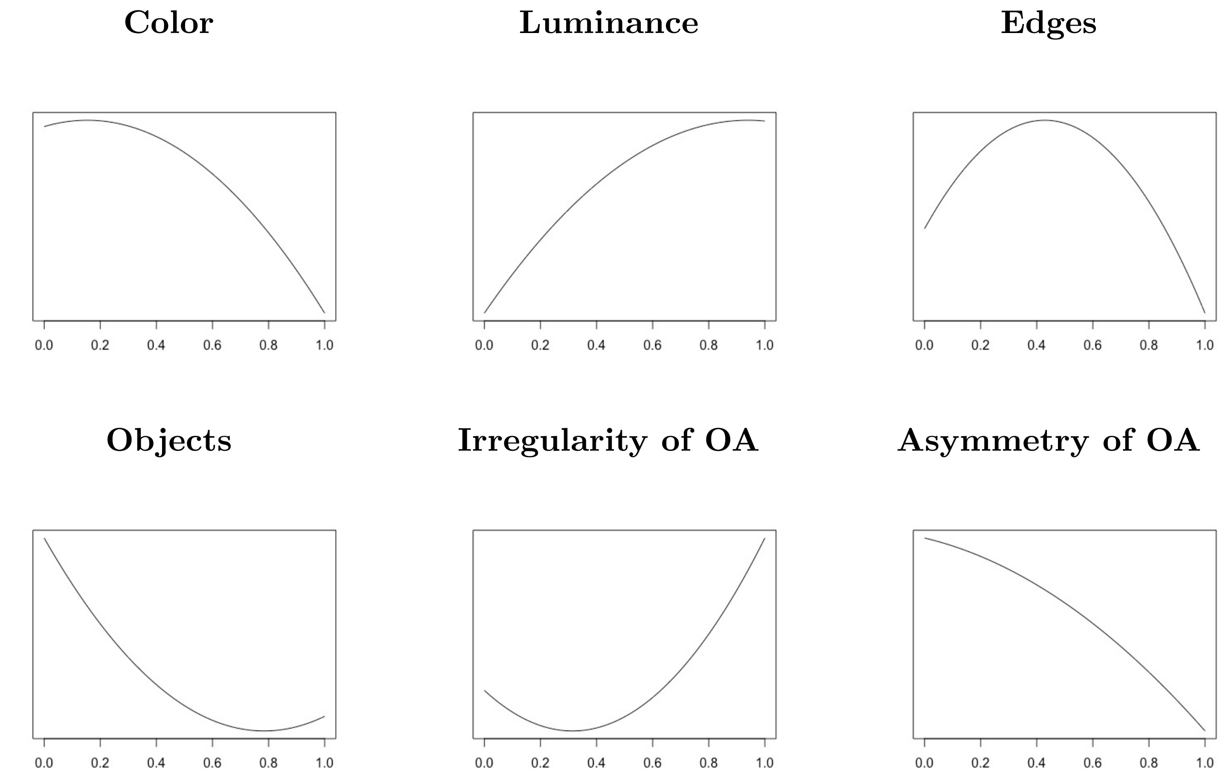


Figure 3.5: Visualization of the effects for each individual measure. The variables in the regression analysis are normalized to be between 0 and 1. The y-axis represents the estimated number of likes, all else being equal.

For the individual design complexity measures we find only partial support for H2. The results show negative main effects for objects ($\beta = -.177, p < .05$) the irregularity of object arrangement ($\beta = -1.143, p < .01$), and asymmetry of object arrangement ($\beta = -.304, p < .01$), but no significant effect for the design complexity control region count ($\beta = .178, p > .1$). The squared terms have positive effects for objects ($\beta = -.111, p < .01$) the irregularity of object arrangement ($\beta = -1.1789, p < .01$), thus supporting H2a and H2b. However, we find a negative effect for squared term of the asymmetry of OA ($\beta = -.589, p < .01$), not supporting H2c. In addition, we find no effect for the design complexity control variable region count ($\beta = -.339, p > .1$).

The bottom row in Figure 3.5 visualizes the effects for the design complexity measures. We observe clear u-shape relationships for objects and irregularity of the object arrangement. Each variable was normalized using a min-max normalization

[8]These maxima lie close to the means of each normalized variable

to be within 0 and 1. The global minima for two of these functions lies within the domain of each variable, with global minima of .80 and .32 for objects and irregularity of object arrangement.[9] We observe an inverted u-shape relationship for the asymmetry of object arrangement with a global maximum of -.26 for asymmetry of object arrangement. This is outside of the domain for this variable. The estimated number of likes is monotonically decreasing for the entire domain of the asymmetry of object arrangement, so the relationship with liking is, in fact, negative.

To summarize, the results show all aspects of feature complexity influence liking in an inverted u-shape type of relationship, fully supporting H1 including a, b, and c. Design complexity as a whole has a u-shape relationship with liking, supporting H2, but we only find support for H2a (objects) and H2b (irregularity of the object arrangement) individually. Combining the estimated coefficients with a plot, we find a negative relationship between the asymmetry and liking of consumers, which suggests that a symmetrical design for images is most strongly related with consumer liking, not supporting H2c.

Table 3.5: Stepwise regression, by introducing more controls in each step to highlight the robustness of our results.

	Stepwise Regression			
	Incl. Brand	**Incl. Temporal**	**Incl. Photography**	**All**
Feature Complexity				
Edge Density	2.036^{***} (.084)	2.036^{***} (.084)	1.728^{***} (.085)	1.730^{***} (.085)
Edge Density^2	-2.316^{***} (.125)	-2.316^{***} (.125)	-1.938^{***} (.125)	-2.013^{***} (.126)
Luminance	$.311^{***}$ (.071)	$.311^{***}$ (.071)	$.175^{**}$ (.073)	$.221^{***}$ (.073)
Luminance^2	$-.181^{***}$ (.047)	$-.181^{***}$ (.047)	$-.103^{**}$ (.050)	$-.121^{**}$ (.050)
Color	$.240^{***}$ (.065)	$.240^{***}$ (.065)	$.265^{***}$ (.089)	$.222^{**}$ (.089)
Color^2	$-.796^{***}$ (.137)	$-.796^{***}$ (.137)	$-.717^{***}$ (.152)	$-.742^{***}$ (.153)
Frequency Factor	1.236^{***} (.241)	1.236^{***} (.241)	1.360^{***} (.245)	1.502^{***} (.245)
Frequency Factor^2	$-.356^{**}$ (.152)	$-.356^{**}$ (.152)	$-.382^{**}$ (.154)	$-.512^{***}$ (.155)
Design Complexity				
Objects	$-.140^{***}$ (.028)	$-.140^{***}$ (.028)	$-.169^{***}$ (.029)	$-.177^{***}$ (.029)
Objects^2	$.111^{***}$ (.032)	$.111^{***}$ (.032)	$.109^{***}$ (.032)	$.111^{***}$ (.032)
Irregularity of OA	-1.348^{***} (.170)	-1.348^{***} (.170)	-1.177^{***} (.178)	-1.143^{***} (.179)
Irregularity of OA^2	2.037^{***} (.228)	2.037^{***} (.228)	1.849^{***} (.236)	1.789^{***} (.236)
Asymmetry of OA	$-.328^{***}$ (.099)	$-.328^{***}$ (.099)	$-.342^{***}$ (.108)	$-.304^{***}$ (.108)
Asymmetry of OA^2	$-.477^{**}$ (.194)	$-.477^{**}$ (.194)	$-.569^{***}$ (.199)	$-.589^{***}$ (.199)
Region Count	$.248^{*}$ (.140)	$.248^{*}$ (.140)	.184 (.141)	.178 (.141)
Region Count^2	$-.392$ (.437)	$-.392$ (.437)	$-.290$ (.437)	$-.339$ (.437)
Log(Followers)	$.923^{***}$ (.001)	$.923^{***}$ (.001)	$.922^{***}$ (.001)	$.921^{***}$ (.001)
Text Sentiment Positive	$.004^{**}$ (.002)	$.004^{**}$ (.002)	$.004^{**}$ (.002)	$.004^{**}$ (.002)
Text Sentiment Negative	$-.003$ (.003)	$-.003$ (.003)	$-.003$ (.003)	$-.003$ (.003)
(Intercept)	-2.507^{***} (.099)	-2.507^{***} (.099)	-2.379^{***} (.101)	-2.371^{***} (.101)
Brand Controls	✓	✓	✓	✓
Temporal Controls		✓	✓	✓
Photography Controls			✓	✓
Content Controls				✓
Observations	147,963	147,963	147,963	147,963
Adjusted R-Squared	.445	.446	.446	.453
Note:				*p<.1; **p<.05; ***p<.01

[9]These minima do not lie close to the means of each normalized variable

3.6.1 Robustness Checks

To test the robustness of our model, we also investigated the robustness of inclusion of content control variables, the size of the effects of our visual complexity variable, confirmed the predictive validity of our model, and examined a generic brand-level fixed effects model. Our results show strong correlations and these additional analyses simply strengthen those findings. The results neither confirm nor disconfirm a direct causal effect.[10]

Robustness against inclusion of controls variables: First, we included a set of temporal controls related to time of posting, day of posting and season of posting. As observed in Table 3.5, there is no significant change in the results. Second, we added the photography controls to the regression. In Table 3.5, we observe that the absolute numbers change slightly, though the direction of the effects remain the same. Finally, we added a set of most frequent types of images, and the presence of faces (humans) to the model. Table 3.5 shows that our results are robust to the inclusion of these 34 content control variables. The visual complexity effects are present above and beyond the types of images and photography attributes. The full length table with coefficients for the content control variables can be found in Table 3.10.

Table 3.6: Overview of model fit of Poisson vs. Negative Binomial Regression

Model	**Log Likelihood**	**AIC**	**BIC**
Negative Binomial	-952,709	1,905,545	1,906,179
Poisson	-62,437,559	124,875,244	124,875,868

Model Fit: As observed in Table 3.6, the Negative Binomial Regression fits the data better than a Poisson regression. This was expected due to the overdispersion that we observe in the number of likes variable. There is a very long tail in the distribution: Few posts obtain a large volume of likes, whereas the vast majority of posts obtain few likes.

Predictive Validity: To assess the predictive validity of our model, we split the data in 20% test set and 80% training set. We predict the liking for the test dataset using our trained model, the results in Table 3.7 are the average for 5-fold cross validation. We investigate the predictive power of our negative binomial regression using the visual complexity measures we propose, and compare them to three benchmark models. We use Spearman Rank Correlation (SRC) to measure the rank correlation

[10]We have also performed analyses at the industry level and observed some minor changes only.

Table 3.7: RMSE and Spearman Rank Correlation for out of sample prediction using our method compared to Pieters et al., Shin et al. and Corchs et al. benchmarks

	RMSE	**SRC**
Corchs	15,057	.9318
Pieters	15,225	.9309
Shin	15,275	.9307
This paper	**14,913**	**.9319**

between observed rank and predicted rank of the liking and the Root Mean Squared Error (RMSE). As observed in Table 3.7, the rank correlation is .9319, which is very high. The measure indicates a high level of predictive validity. The RMSE is quite high as well, which shows that it is much easier to predict the relative ranking of certain posts than it is to predict the exact number of likes. Especially posts with an extremely high number of likes are difficult to predict and this increases the RMSE. Most importantly, we observe that compared to the benchmark our model performs better. Compared to benchmarks Pieters, Wedel, and Batra, 2010 and Shin et al., 2019, that use the JPEG file size for the feature complexity and an additive measure for design complexity, we perform about 2 % better on the RMSE and slightly better on the SRC. Compared to a model using the 11 complexity measures by Corchs et al., 2016 we perform about 1% better on the RMSE and equal on the SRC. The SRC for all models was already quite high, most of which is driven by the number of followers, so this shows little improvement. On the RMSE, however, the improvement is substantial enough to make a difference. The predictive validity combined with the interpretability of our method over the benchmarks highlights the importance of our framework.

Size of effects: We observe an inverted u-shape relationship between the measures of feature complexity and the liking of social media imagery. These results suggest that we would be able to find the optimum for both these measures that would lead to the highest number of expected likes when keeping all other factors the same. As an examination of the effect size we explored what this optimization effect would be for choosing the optimal image over a non-optimal image and we observe that improving feature complexity to its theoretical optimum would increase the expected number of likes by 19%. Given that the average likes on an image in our brand set is 4138, this would result in an increased number of likes of 786 on average. A quick and easy way for brands to improve the feature complexity of an image is to apply a filter based on the complexity scores. We explored the effect of choosing the right filter and we observe that just by choosing the right filter would improve the expected number of likes by 3%, see Appendix C and Figure 3.6 for an illustration. That means this would result in an increased number of likes of 125 on average. It is important to note that applying a filter takes less than a second since it involves simply clicking on the appropriate filter. This means that the ROI from either minor uses of our model is high.

Brand-level fixed effects: We have chosen to include specific brand-level fixed effects to account for the fact that brands have very different social media capabilities from each other. The specific brand-level fixed effects that we include are post frequency and the number of followers. However, we also examined a negative binomial model with fixed-effects for the brands, but it does not lead to a change in our conclusions. For the estimation of this model we allowed fixed effects for each individual brand. This estimation takes a lot of computing effort (over 600 extra variables for regression) without improving the results. We observe highly similar results for all our variables.

3.7 Discussion

In this paper, we have sought to understand the influence of different measures of visual complexity on the liking of social media imagery. Therefore, we have expanded, automated, and scaled up the existing visual complexity framework, as proposed by Pieters, Wedel, and Batra, 2010, and we have created automated measures for measuring feature- and design complexity. Subsequently, we have investigated the influence of each individual measure on the liking of social media imagery. We observed an inverted u-shape relationship of feature complexity, including its individual components, with liking, fully supporting Hypotheses 1,1a,1b, and 1c. There are optimal levels of color, luminance and edge density for which liking is highest, visualized in Figure 3.5. In contrast, we observe a u-shape relationship of design complexity, including two out of three of its individual components, with liking, thus supporting Hypotheses 2, 2a, and 2b. Design complexity in terms of unique objects and irregularity of the object arrangement needs to either be simple or complex, with mid-levels performing worse. For design complexity in terms of asymmetry of object arrangement, we observe a negative relationship with liking, with symmetrical images most strongly correlated with liking, not supporting Hypothesis 2c.

3.7.1 Theoretical Implications

By introducing, implementing and validating an automated visual complexity framework, partially derived from Pieters, Wedel, and Batra, 2010, we make two major contributions. First, we find that the relationship between the two overarching categories of visual complexity, feature complexity and design complexity, and consumer liking is non-linear. We find an inverted u-shape relationship for feature complexity and a u-shape relationship for design complexity. Previous theory has established that feature complexity negatively influences attitudes (Pieters, Wedel, and Batra, 2010), while being able to provide positive peripheral cues (Shin et al., 2019). Our results suggest truth in both findings, such that a certain level of complexity is needed to provide the necessary positive peripheral cues, while too much complexity might make it too difficult to process or to hard to recognize what is on the image. The optimum level of feature complexity is somewhere in the mid-regions, depending on specific aspects such as color, luminance or edge density.

The same two studies are also contrasting in their findings related to design complexity. Higher complexity in the design has collative properties that can increase liking (Palmer, 1999; Pieters, Wedel, and Batra, 2010), while simplicity in terms of design makes an image much easier to understand and therefore increases liking (Shin et al., 2019). We find that either end of the design complexity spectrum is positively related to liking. Simple designs indeed make an image easy to process cognitively and therefore easy to understand, while an elaborate and creative design that is very complex has those collative properties that make it likeable.

Second, instead of using aggregated measures for feature and design complexity, we developed and validated a set of measures that provides us with a more nuanced and interpretable view of the relationship between liking and visual complexity. We observe that all three aspects of feature complexity (i.e. color, luminance, and edge density) influence the liking of FGI uniquely. The results show that a certain level of complexity and variety in color and luminance is needed to capture the users' attention and is important for directing the eyes of the consumer to the content. However, the inverted u-shape relationship shows that it only helps the liking to a certain extent. Stuffing an image with variation hurts liking. The results show a similar pattern for the edge density. The edge density measures the amount of detail in an image. An image low in edge density is not appealing enough for the consumer to like and an image that is high in edge density contains too much detail and may prevent easy comprehension by the consumer. Too much information and variation has also been shown to negatively impact users' interest in home pages of websites (Geissler, Zinkhan, and Watson, 2006).

For design complexity, we observed a u-shape relationship with liking for the number of objects and the irregularity of the object arrangement. Simple designs activate object and pattern recognition processes plus they are easier to comprehend and process. For these reasons, they are more engaging (Palmer, 1999). For example, an image with a single object is likeable, because it clearly reflects what the image is about. The same holds for a regular object arrangement. The asymmetry of the object arrangement has a strictly negative relationship with liking, such that asymmetrical images are correlated to fewer likes than symmetrical ones, which is in line with these same expectations. On the other hand, an image with a higher number of unique objects and/or an irregular arrangement has aesthetic qualities and can therefore be more engaging Berlyne, 1958.

Overall, we can conclude that visual complexity is clearly not a linear, monolithic construct and can therefore not be captured by a single additive measure. In addition, it's relationship with consumer liking is not linear and can be interpreted more easily using its individual components.

3.7.2 Methodological Implication

We have developed a framework that enables researchers to study image-based social media in a similar manner as they are currently studying text-based social media. The automated measures have been validated in an experiment to ensure that they accurately represent how visual complexity with respect to its individual components are perceived. From here, we have identified the aspects of social media imagery that lead to liking rather than the particular images that are liked. This gives us theoretical principles about how to design image-based social media that advertisers and marketing managers can benefit from. We observe that aspects of visual complexity, derived from Visual Complexity Theory (Attneave, 1954; Donderi, 2006), influence liking of social media imagery. It is important to note that these are basic aspects of photos that influence liking regardless of what is depicted in the image. We have confirmed this by showing that the relationships don't change after including a wide variety of content characteristics as control variables in our regression. The visual complexity measures in our study capture the image features that evoke primary reactions and we find that those features seem to directly influence the liking. It may be the case that users are not always aware of why they like an image (Moreland and Zajonc, 1977). This means that a user mentally likes an image viewed on Instagram sometimes before realizing what is shown on the image. It is important to realize that visual complexity needs to be incorporated in the creation of FGI on social media. Methods such as our approach are necessary to identify what aspects of an image make it likeable.

3.7.3 Managerial Implications

The combination of understanding how different aspects of feature- and design complexity influence the liking and the automated extraction of this information directly from images makes a powerful tool for content managers. Using this information, they can effectively improve their marketing content on a large scale to better connect with their customers. In turn, this will strengthen customer-brand relationships (Kumar et al., 2016).

The results show that it is important for marketing managers to stay in the middle, or the 'Goldilocks' region of feature complexity. Managers should make sure there is enough variation and complexity in the image to evoke a positive first impression from consumers, but not make it too complex in this regard since that will hurt consumers' liking of social media imagery. We conducted an additional analysis to investigate the effects of using filters on Instagram. Our filter analysis showed that out of this 19% theoretical improvement, approximately 3% can be attributed to choosing the right filter based on the feature complexity scores. Remember, a filter in Instagram is very easy to apply to an image since it is a coherent part of the posting process. This means that content managers are able to improve liking by 3% with just a few additional clicks. Figure 3.6 and Appendix C illustrate what the filter guide could look like. Finally, optimizing feature complexity measures can even lead to a theoretical increase of 19% in likes. Based on those findings, and the open-source code, one could design a tool or dashboard that automatically rates, and potentially optimizes, newly produced FGI to achieve an ever higher gain than the illustrated filter guide.

Unlike feature complexity, design complexity is not something that simply arises after the photo is taken. Design complexity needs to be considered before the photo is taken. It is interesting to note that either a simple or complex design can work well. Our recommendation is to use a regular and simple design, using a unique object or a regular arrangement of multiple objects in the image, but to be aware of their symmetrical arrangement and orientation. Make sure that they tell a coherent story and form a well-designed, uncluttered image as a whole. On the other hand, creatively designed images that use a lot of objects and an irregular arrangement can certainly also work well. Using our framework the content creator can quickly determine where an image lies on the spectrum and decide if it is located in a sub-optimal region. In addition, a marketer can design a photograph using many different combinations and use a dashboard, as mentioned in the previous paragraph, to choose optimal content automatically.

3.7.4 Limitations and Future Research

Although our study offers key insights into the impact of visual complexity on the liking of FGI and contributes to both theory and practice, it has several limitations that need to be acknowledged.

Our single dependent variable, liking, has some limitations that could be addressed in future research. One issue with liking on Instagram, especially with the new regulations regarding privacy, is that we can't know who has liked the post. This might be relevant, because if a person that has a lot of followers likes a particular post, this post might show up on the feed of its followers as well. Therefore, the "image journey" has not been investigated. Some likes hold more value than others, which means that just the count of likes may lack some depth. This so-called "image journey" is a problem that requires a much richer dataset, so we can know exactly how much extra exposure a singe like has generated. With the new regulations this is fairly impossible as the consumer data is mostly restricted, but would be a very interesting future research direction.

In future work, we could also examine other measures of consumer engagement, besides pure liking behavior. For instance, comments come from a different motivation than relevant for the type of engagement we are interested in. In many cases, comments are replies to what has been said in other comments and this does not really reflect brand awareness. Additionally, similar to the likes we only know the volume of comments per post, but we don't have any insight into the content of the comments. Therefore, we do not know the motivation behind the comments, which makes it even less relevant for this study. For future research it would be interesting to take a look at the drivers behind the comments.

It would also be interesting to explore the moderating factors that influence the liking of content. Moderators such as brand strength or brand familiarity could potentially moderate the relationship between visual complexity and liking. When users are familiar with a brand, the impact of image complexity on liking might be different. The filter guide and the optimal filter to improve liking might be dependent on the brand and what they have posted before.

Our complexity framework is the first for automated extraction of complexity measures of images. However, the accuracy and quality of information can be improved. Especially with the development of deep learning models and more data. Instead of using the orientation clutter measures to estimate the irregularity and asymmetry of the object arrangement, one could utilize the output of the MaskRCNN object detection to determine these directly. In addition, the pre-trained Mask RCNN model that we used is trained to recognize only 81 unique objects. Future research could explore training it to be able to recognize a larger variety of objects. To have a model that covers all of this would require more training data and a lot more computing power, which is something that will be possible in future research.

Our complexity framework opens up possibilities for a wide range of applications. Managers, policy makers and marketing professionals alike can directly extract large amounts of information from images and use this information to better understand their consumers and optimize their content accordingly, and hopefully use it for the "good" of the consumer. Image analytics at scale can offer key insights in understanding the diffusion of online FGI and we encourage future exploration of the possible applications.

3.8 Appendix

3.8.1 Visual Complexity

Table 3.8: Comprehensive list of visual complexity approximations and measures. The automated and interpretable visual complexity measures used in the paper are derived/chosen from this list of complexity measures on the basis of interpretability and correlation.

Category	**Complexity Measure**	**Reference**	**Interpretability**
Feature	Color Count	Artese (2014), Corchs (2016)	Yes
Feature	Color Entropy	Artese (2014), Corchs (2016)	Yes
Feature	Color Harmony	Artese (2014), Corchs (2016)	Yes
Feature	Colorfulness	Corchs (2016), Hasler (2003)	Yes
Feature	Contrast (G)	Cavalcante (2014), Corchs (2016), Haralick (1973)	No
Feature	Correlation (G)	Corchs (2016), Haralick (1973)	No
Feature	Edge Density	Corchs (2016), Rosenholtz (2007)	Yes
Feature	Energy (G)	Corchs (2016), Haralick (1973)	No
Feature	Frequency Factor	Corchs (2016), Corchs (2013)	No
Feature	Homogeneity (G)	Corchs (2016), Haralick (1973)	No
Feature	JPEG File Size	Corchs (2016), Corchs (2013), Forsythe (2011)	No
Feature	Luminance Entropy	Cavalcante (2014)	Yes
Design	Object arrangement	Pieters (2010)	Yes
Design	Object count	Oliva (2004)	Yes
Design	Object irregularity	Pieters (2010)	Yes
Design	Region Count	Comaniciu (2002)	No
Feature/Design	Clutter	Rosenholtz (2007)	Yes
Feature/Design	Deep Neural Network	Nagle (2020), Machado (2015)	No

3.8.2 Descriptives and Regression

Table 3.9

Variable	mean	sd	min	max
Dependent Variable				
Likes	4,138	30,694	0	935,690
Feature Complexity				
Color	3.00	1.75	.00	18.41
Luminance	2.48	.49	.02	2.85
Edge Density	.09	0.03	.00	.35
Frequency Factor	.42	.04	.00	.49
Design Complexity				
Objects	17.38	20.51	0	100
Irregularity of OA	.06	.02	.00	.23
Asymmetry of OA	.24	.11	.00	.92
Region Count	56.51	30.11	0	1,320
Textual Sentiment				
Text Sentiment Positive	1.74	.89	1	5
Text Sentiment Negative	1.23	.56	1	5
Hashtags	4.31	5.24	0	39
Brand Specific				
Followers	168,751	1,679,190	15	46,098,258
Posts	232	155	52	990
Photography				
Diagonal Dominance	.69	.24	0	1
Rule of Thirds	.59	.12	0	1
Physical Dominance (Vertical)	.83	.14	0	1
Physical Dominance (Horizontal)	.85	.13	0	1
Color Balance (Vertical)	.79	.08	0	1
Color Balance (Horizontal)	.75	.09	0	1
Figure-Ground Size Diff	.39	.35	0	1
Figure-Ground Color Diff	.32	.20	0	1
Figure-Ground Texture Diff	.15	.11	0	1
Saturation	.30	.17	0	1
Contrast	.45	.16	0	1
Clarity	.49	.25	0	1
Warmth	.25	.22	0	1
Content Controls (Binary)				
Crazy car	.02	.13	0	1
Classic castle	.03	.16	0	1
Hot girls	.01	.11	0	1
Outdoor party	.01	.09	0	1
Busy office	.01	.08	0	1
Amazing food	.01	.09	0	1
Hot cup	.01	.11	0	1
Cute animals	.01	.08	0	1
Outdoor wedding	.02	.13	0	1
Favorite team	.004	.06	0	1
Art studio	.004	.06	0	1
Bakery/shop	.03	.17	0	1
Beach	.01	.10	0	1
Clean room	.04	.20	0	1
Coffee shop	.01	.11	0	1
Desert/sand	.01	.10	0	1
Museum/indoor	.02	.14	0	1
Nursery	.02	.14	0	1
Ocean	.01	.07	0	1
Playroom	.03	.16	0	1
Face	.25	.43	0	1
Temporal Controls (Binary)				
Afternoon	.30	.46	0	1
Evening	.39	.49	0	1
Night	.18	.39	0	1
Weekend	.19	.40	0	1
Spring	.10	.30	0	1
Summer	.31	.46	0	1
Fall	.33	.47	0	1

Table 3.10: Full set of results with coefficients for all control variables. Full length of Table 3.4

	Quadratic-Individual
Feature Complexity	
Edge Density	1.730^{***} (.085)
Luminance	$.221^{***}$ (.073)
Color	$.222^{**}$ (.089)
Frequency Factor	1.502^{***} (.245)
Feature Complexity^2	
Edge Density^2	-2.013^{***} (.126)
Luminance^2	$-.121^{**}$ (.050)
Color^2	$-.742^{***}$ (.153)
Frequency Factor^2	$-.512^{***}$ (.155)
Design Complexity	
Objects	$-.177^{***}$ (.029)
Irregularity of OA	-1.143^{***} (.179)
Asymmetry of OA	$-.304^{***}$ (.108)
Region Count	.178 (.141)
Design Complexity^2	
Objects^2	$.111^{***}$ (.032)
Irregularity of OA^2	1.789^{***} (.236)
Asymmetry of OA^2	$-.589^{***}$ (.199)
Region Count ^{2}	$-.339$ (.437)
Brand Specific	
Log(Followers)	$.921^{***}$ (.001)
Posts	$-.287^{***}$ (.003)
Text	
Text Sentiment Positive	$.004^{**}$ (.002)
Text Sentiment Negative	$-.003$ (.003)
Hashtags	$.024^{***}$ (.0004)
Temporal Controls	
Afternoon	$-.011^{*}$ (.006)
Evening	$.022^{***}$ (.006)
Night	$.074^{***}$ (.007)
Weekend	$.047^{***}$ (.005)
Spring	$-.217^{***}$ (.007)
Summer	$-.230^{***}$ (.005)
Fall	$-.197^{***}$ (.005)
Photography Controls	
Diagonal Dominance	$-.016^{**}$ (.008)
Rule of Thirds	$.116^{***}$ (.018)
Vertical Physical Dominance	.018 (.014)
Horizontal Physical Dominance	.006 (.016)
Horizontal Color Balance	$-.415^{***}$ (.027)
Vertical Color Balance	$.108^{***}$ (.033)
FG Size Difference	$.077^{***}$ (.007)
FG Color Difference	$.115^{***}$ (.012)
FG Texture Difference	$-.010$ (.020)
Saturation	.002 (.024)
Contrast	$-.008$ (.018)
Clarity	$-.123^{***}$ (.010)
Warmth	$-.086^{***}$ (.009)
Content Controls - ANP	
Crazy car	$.043^{***}$ (.015)
classic castle	$.053^{***}$ (.013)
Hot girls	$.052^{***}$ (.017)
Outdoor party	.013 (.021)
Busy office	$-.042^{*}$ (.023)
Amazing food	$-.061^{***}$ (.020)
Hot cup	$-.038^{**}$ (.018)
Cute animals	.021 (.023)
Outdoor wedding	$.042^{***}$ (.014)
Favorite team	$-.034$ (.032)
Art studio	$-.005$ (.031)
Bakery/shop	$-.045^{***}$ (.012)
Content Controls - Places365	
Beach	$.108^{***}$ (.019)
Clean room	$-.088^{***}$ (.010)
Coffee shop	$-.024$ (.018)
Desert/sand	$.047^{**}$ (.018)
Museum indoor	$.051^{***}$ (.014)
Nursery	$-.047^{***}$ (.014)
Ocean	$.106^{***}$ (.026)
Playroom	$.091^{***}$ (.012)
Content Controls - Face VGG16	
Face	$-.055^{***}$ (.005)
(Intercept)	-2.401^{***} (.099)
Observations	147,963
Adjusted R-Squared	.453
Overdispersion θ	1.96^{***} (.007)
Note:	*p<0.1; **p<0.05; ***p<0.01

3.8.3 Filter Guide

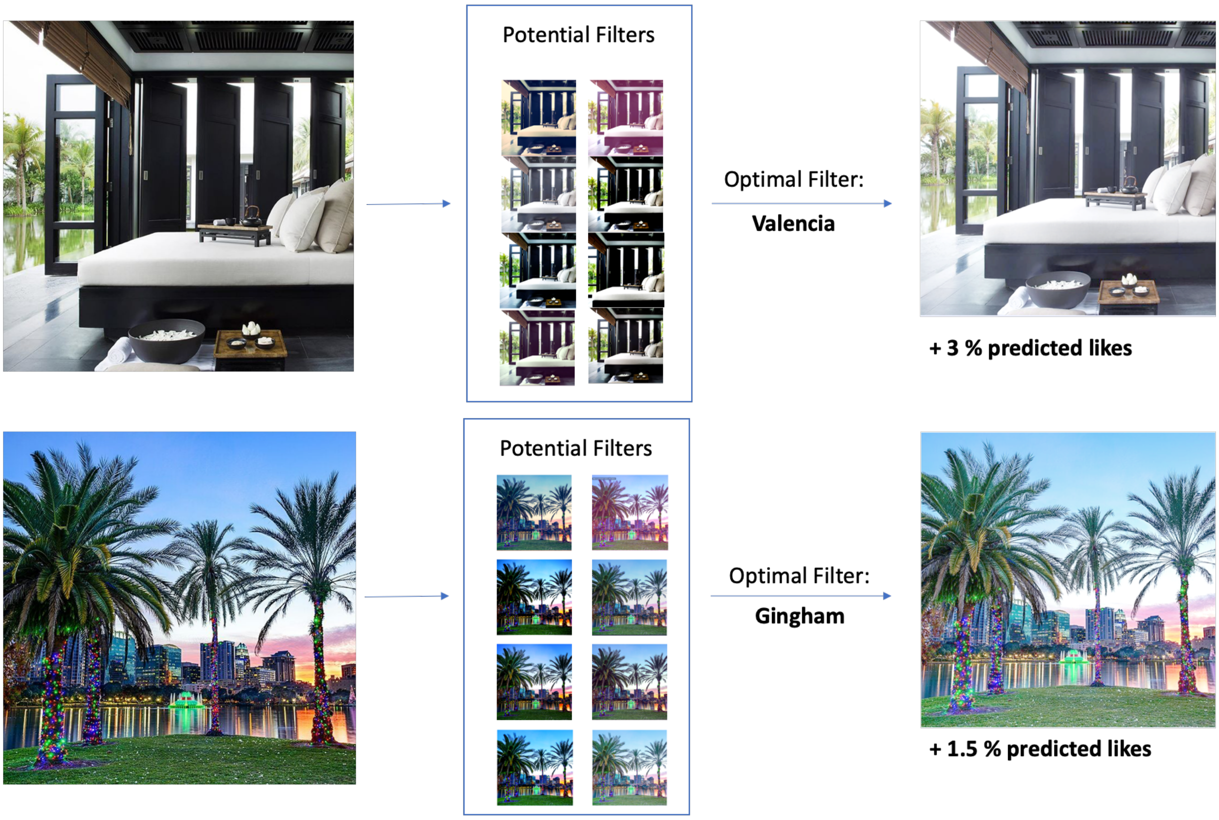


Figure 3.6: Visualization of the filter guide. Given a picture and it's complexity score, we can apply potential filters and analyze the new complexity scores. From there, we can select the optimal filter, (no filter included as option), applying the filter leads to a predicted increase of 3% for the top picture and 1.5% for the bottom picture. We picked low and high colorfulness pictures for our illustration. The predicted likes increase is out-of-sample prediction, and we used these posts' actual values for all other variables.

A quick and easy way for brands to improve the feature complexity of an image is to apply a filter based on the complexity scores. We explored the effect of choosing the right filter and we observe that just by choosing the right filter would improve the expected number of likes by 3%. That means this would result in an increased number of likes of 125 on average. Figure 3.6, illustrates the process of a filter guide. Based on the feature complexity scores of these images, we explore a set of potential filters. Then, we analyze what the new complexity scores would be after applying these potential filters. The filter (no filter as part of the options), that brings us closest to the optimal values for each of the individual would provide us with the highest predicted number of likes. We can then select the optimal filter based on the predicted scores. In the examples, that gets us to an increase of 3 % (1.5 %) for the top (bottom) image in predicted likes, all non-image characteristics being equal. We used actual posts and their corresponding scores for all variables. It is important to note that applying a filter takes less than a second since it involves simply clicking on the appropriate filter. The calculation of the feature complexity scores for all potential filters also takes less than a second. An automated tool, therefore, would quickly be able to apply the best filter based on the visual complexity of the image.

3.8.4 Convolutional Neural Networks and Content Controls

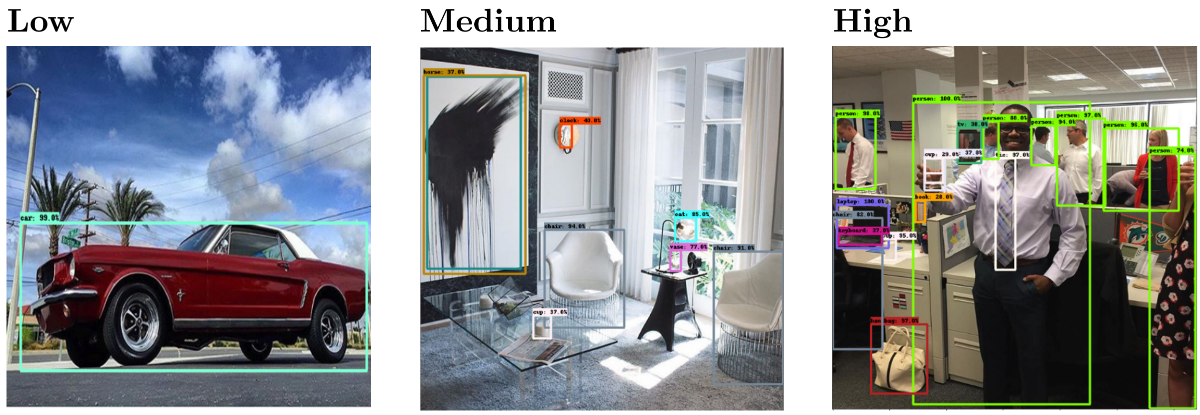


Figure 3.7: Visualization of Mask RCNN object detection for low, medium and high number of objects.

In the last decade, researchers in computer science have developed the ability to automatically extract conceptual information from a large number of images. This information has shown to be particularly useful in a number of research fields. Recently, we have also seen an adoption of these methods for marketing research, especially in online settings where image data is often used. In this paper, we use CNNs to extract the object complexity and the content control variables. CNNs are powerful deep learning networks developed primarily for image recognition. CNNs have been successful in identifying objects in images, such as faces, humans and animals, or scenes such as park, coffee shop, beach etc. Convolutions are effective at extracting image features, because they are a type of filter that is applied multiple times to different parts of the image. The CNN uses only a small set of parameters that need to be estimated to detect similar features in multiple locations in an image. Nowadays, we can use large datasets with labeled images and the increasingly cheap nature of computer power to learn the parameters in convolutions at a large scale. The CNNs have several types of layers (mathematical manipulations) to extract different types of information from an image. The CNN architecture builds up a large amount and variety of information from the image and combines all of these different types of information to enable identification of complex concepts in the image. By scanning over a large number of pre-labeled images and adjusting weights the CNN can “learn” how to recognize the labeled information in the images of the training set. We use four different pre-trained CNNs to extract our content information. Three of these are pre-trained classifiers that classify an image as belonging to a certain class, the fourth is an object localization classification. Instead of classyfing an image as a whole, it first determines regions of interest that are then classified to be of a certain class. First, we will explain the mask RCNN architecture and how we use it to extract the number of objects. Then, we'll discuss the three pre-trained image classifiers that we use to create our binary content indicators.

More recently, object localization (i.e. detecting and localizing multiple objects within an image, instead of classifying an entire image) has become more accurate. He et al., 2017 proposed Mask R-CNN, using Region-Based CNNs Girshick et al., 2014 to classify regions of interests within images, to accurately detect objects within an image. In (Nagle and Lavie, 2020), the authors show that this is in fact the most effective individual predictor of visual complexity. Using a pre-trained Mask R-CNN, trained to recognize 81 different types of objects, we are able to count the total number of (unique) objects within an image. Figure 3.7, visualizes the output of object detection using Mask RCNN. We use the latest MaskRCNN architecture, Inception ResNet V2 Mask RCNN trained on the coco dataset. As shown, it does not always detect all objects, nor does it classify them perfectly. However, our validation experiment does show that it accurately reflects the perception when we simply count the number of detection boxes from the classifier. In future research, once these models become faster and more accurate, we expect that the object complexity score will be more accurate. In addition, one could use the distribution of the detected objects for the irregularity of object arrangement and asymmetry of object arrangement as well.

For the construction of the content indicators, we use three CNNs trained to recognize, scenes, adjective-noun pairs and faces. For the places/scenes classification we use a deep neural structure trained on previous images of different locations, called the Places Database (Zhou et al., 2018). The Places Database consists of 10 million scene photographs, all labeled with scene semantic categories. It comprises a diverse list of types of environment encountered in the world. For instance, scenes include: Lobby, Jacuzzi, Dorm Room, and Building Facade. The deep learning model accurately identifies 365 scene categories depicted in images. Similar to object detection, the pre-trained CNN returns a probability score for each of the 365 scene categories in the image. The final result is a distributional representation of the identification of scenes for every hotel image in our dataset. We detected adjective-noun pairs using the MVSO model (Borth et al., 2013). The model accurately identifies 1200 adjective-noun pairs. The binary indicators for both these variables were constructed by simply selecting the top 10 most frequent classes from both of these pre-trained CNNs applied to our dataset. From there, an indicator would indicate 1 if the image was classified as being one of these top 10 most frequent scenes or adjective-noun pairs. Lastly, for the face detection, we used a CNN pre-trained to recognize faces (Parkhi, Vedaldi, and Zisserman, 2015). In case the model detects a face in the image, we assign a 1 to the face indicator.

3.8.5 Alternative Complexity

These 11 measures evaluate visual complexity measures proposed in Corchs et al., 2016. The measures found in other papers, are either highly similar or create a combination of filters and compression. We are already dealing with compressed imagery, so we can only use the file size as a measure for compression. They find a correlation of r=0.81 with perceived complexity of participants from a linear combination of these measures. M7 and M9 correspond to the edge density and color as we use them. M5 and M8 are visual complexity controls that we incorporate in our paper. M11 was not used, because this is closely related to the photography controls. Finally, M6 is what both the other benchmark papers (Pieters, Wedel, and Batra, 2010; Shin et al., 2019) use as their measure for the feature complexity.

- **M1**: Contrast; it measures the intensity contrast between a pixel and its neighbors over the whole image.
- **M2**: Correlation; it measures how correlated a pixel is to its neighbors over the whole image.
- **M3**: Energy; it is the sum of squared elements in the GLCM.
- **M4**: Homogeneity, it measures the closeness of the distribution of elements in the GLCM with respect to the GLCM diagonal.
- **M5**: Frequency Factor, it is the ratio between the frequency corresponding to the 99% of the image energy and the Nyquist frequency (highest possible frequency in an image).
- **M6**: Compression Ratio, which is the JPEG file size.
- **M7**: Edge Density, same as the edge density measure in our study.
- **M8**: Number of regions, computed with the superpixel-based fast fuzzy C-means image segmenation as proposed by Lei et al., 2018 (more advanced method than proposed in (Corchs et al., 2016).
- **M9**: Colorfulness; it consists in a linear combination of the mean and standard deviation of the pixel cloud in the color plane (Artese, Ciocca, and Gagliardi, 2014).
- **M10**: Number of Colors; measures the number of distinct color in the RGB image.

- **M11**: Color Harmony, based on the perceived harmony of color combinations. It is composed of three parts: the chromatic effect, the luminance effect, and the hue effect. The image is split up into 10 segments, based on (Lei et al., 2018), each with their average color. We then take the minimum of the harmony of each segment compared to all others.

3.8.6 Photography Attributes

As control variables for our study we compute the photography attributes used in Zhang et al., 2017 and Zhang and Luo, 2018. The attributes are split up into three main categories: Color, Composition and Figure-Ground Relationship.

Composition

First, we compute a saliency map of the image, assigning a saliency score to every pixel in the image. Then, we use the superpixel algorithm to segment the image into 10 main regions.[11] The salient region in the image is the segment with the highest average saliency score.

- **Diagonal dominance** We calculate the distance between the center of the salient region to each of the two diagonals of a photo. The diagonal dominance is the negative of the minimum of these two distances.

- **Rule of thirds** We calculate the distance from the center of the salient region to each of the four intersections of the two horizontal lines and the two vertical lines that evenly divide the photo into nine parts. The rule of thirds score is the negative of the minimum of these distances.

- **Physical visual balance** We calculated two physical visual balance measures: vertical and horizontal. We calculated the weighted saliency centroid from a weighted average centroid. We weigh the centroid of each of the 10 segments by the average saliency score to find the weighted center of the image. The vertical (horizontal) physical visual balance is than the distance from that center to the horizontal (vertical) line splitting the image into two halves.

- **Color visual balance** We calculated two scores for color visual balance: vertical and horizontal color visual balances. Each pixel is compared to its vertical (horizontal) counterpart. The score is the average euclidean distance of each pixel pair.

[11] we chose to do the segmentation like (Zhang and Luo, 2018) to match previous work, instead of the superpixel-based fast fcm used above

Figure-ground relationship

Figure refers to the foreground, and ground refers to the background, of a photo. For the first three figureground relationship features, we first use the Grabcut algorithm (Rother, Kolmogorov, and Blake, 2004) to identify the figure and background of each photo. In the following, we explain how we extract each attribute for figureground relationship.

- **Size difference** We take the difference between the number of pixels of the figure and that of the background, normalized by the total number of pixels of the photo.
- **Color difference** We first calculate the average RGB vectors for figure and ground. Then the color difference is the Euclidean distance between the two RGB vectors.
- **Texture difference** Difference in edge density between the figure and the ground.

Color

- **Brightness** is the average of the value dimension of HSV across pixels (Datta et al. 2006).
- Saturation is the average of saturation cross pixels.
- **Contrast of brightness** was calculated as the standard deviation of the value dimension of HSV cross pixels.
- **Clarity** A pixel is defined to be of enough clarity when the Value of the HSV is more than 0.7.
- **Warm** hue the warm hue level for the photo is the proportion of warm hue (i.e., red, orange, yellow) pixels in a photo.
- **Colorfulness** - We already have this measure as part of our visual complexity framework.

Chapter 4

The Champion of Images: Understanding the Role of Images in the Decision-Making Process of Online Consumers.

Authors: Gijs Overgoor, William Rand, Willemijn van Dolen, H. Steven Scholte

This paper has been invited for resubmission at *Marketing Science*. Gijs Overgoor was the leading author for this study. Willemijn van Dolen and William Rand fulfilled a supervisory role for the paper. H. Steven Scholte was mainly responsible for carrying out the fMRI experiment and corresponding analysis.

4.1 Abstract

We propose a visual analytics framework to study the importance of the product image during consideration set formation on e-commerce websites. We apply our proposed framework to an extensive dataset of consumer search for hotels on the website of a global online travel agency. We predict product-level click-through rates using image information that we extract with convolutional neural networks and find that we are able to accurately predict what hotel will be more likely to be clicked on. We complement these findings using LambdaMART to predict consumer clicks during search and find that on average there is a 10% improvement when we incorporate image information as compared to just the textual and numerical features. In addition, we find that the imagery affects the importance of other attributes such as price, with a decrease in the importance of price by 70% in some locations. Finally, in a neuroscience experiment we show that our results can be explained by the fact that the human brain processes high click-through rate images differently than low click-through rate images. Overall, we present one of the first visual analytic frameworks that can be used at a large-scale to help understand the impact of imagery online.

4.2 Introduction

Images dominate most mainstream products and services pages. In fact, the image is one of the first pieces of information a consumer searching for products observes, usually in the form of a thumbnail that appears next to the product description and pricing information. Compelling images can be more important to entice online consumers than the product description or consumer ratings and some surveys claim that up to 93 % of consumers consider the visual appearance as the key deciding factor for purchase decisions.[1] As E-commerce businesses have made it easy for consumers to quickly compare a large number of products, the product image can be an important tool for companies to use to grab the attention of the searching consumer and convince them to consider their products. If marketers want to use these images to their advantage, then developing methods to understand the role of the image in the decision-making process of online consumers is essential.

For this development, there is a nascent body of literature starting to form around understanding visual information that can be built upon. Academic researchers and practitioners have established the power of extracting visual information using image mining methods and visual analytics. For example, Khosla, Das Sarma, and Hamid (2014) show that information extracted from images by use of a Convolutional Neural Network (CNN) is effective in predicting likes on social media. More recently, marketing researchers have adopted similar image mining methods to study user-generated social media imagery (Hartmann et al., 2019; Rietveld et al., 2020; Liu, Dzyabura, and Mizik, 2020). In the e-commerce space, researchers show that image information extracted with CNNs is an effective predictor of product return rates in the fashion industry (Dzyabura, El Kihal, and Ibragimov, 2018) or restaurant survival (Zhang and Luo, 2018). This information is also be useful to predict design gaps (Burnap and Hauser, 2018). These studies highlight the importance of using visual analytics to study the impact of visual content online. In this study, we propose a visual analytics framework to extract information from product imagery and relate this information to click-stream data.

To date, there is still very little knowledge about the impact of product imagery on consumers searching for information online. However, there have been some studies recently that investigate images and their relationship with the demand of properties on Airbnb. Zhang et al. (2017) and Zhang et al. (2019) show that photography quality positively influences the demand of properties on Airbnb in the short term, but that

[1] https://www.mdgadvertising.com/marketing-insights/infographics/its-all-about-the-images-infographic/ *(accessed on september 4, 2020)* & https://www.justuno.com/blog/65-e-commerce-statistics-about-consumer-psychology/ *(accessed on september 4, 2020)*

listings with high quality images might negatively influence the long term demand through unrealistic expectations. Another study shows that the lay-out of images and selecting the right image as the main image in the lay-out leads to an increase in demand as well (Li et al., 2019). The images of properties on Airbnb clearly influence the demand of listings, but little is known about the impact of product imagery when a consumer is presented with a large list of alternatives when searching for products.

In this research, we examine the effect of the product image on the consideration set formation of consumers searching for hotels on the website of a global online travel agency (OTA). We view a consumer's click on a listing in the search result page to find out more information about a hotel as an inclusion of this hotel into the consideration set. We focus on this stage of the search process, because we can link the product image in the listing directly to the consumer search and the quick decisions that a consumer makes when presented with a list of products.

We study the impact of product images on consumer consideration set formation in several stages: First, we perform a click-through rate prediction using just the information extracted from the product image. This highlights how much information a product image conveys to consumers. In this aggregate-level study we also visualize the types of images that generally work well for specific locations; second, we draw from the Learning-to-Rank literature and use the LambdaMart (Burges, 2010) model to predict individual consideration set decisions. We establish the importance of product images and we show that the images impact the evaluation of other attributes, such as price, presented to the consumer; third, we conduct an fMRI experiment to confirm the results and expectations from the prediction methods. The application of the prediction methods combined with the fMRI experiment not only leads to important insights about the role of images for experience products and the travel industry, but also to provide general insights into the relationship between product images and other attributes (i.e., the textual and numerical information displayed in product listings) during search. Though we have presented these methods in the context of hotel search, the methodology presented in the paper is adaptable and can be applied to any online consumer search setting that involves visual content.

Our study makes several important contributions:

- Our proposed framework provides a holistic view of all relevant factors for predicting individual clicks conditional on consumer search requests. This method not only shows the importance of including image features for prediction, but it highlights changes in the importance of features such as price as a consequence of incorporating image features.

- We connect image processing in the human brain to consumer search and consideration set formation. The combination of aggregate- and individual-level prediction methods with a fMRI experiment provides multiple methods that confirm the fact that images influence consumers' propensity to click on certain products.

- We use advanced visual analytic methods to explore the aspects of images that drive the inclusion of products into the consideration set of consumers. This is one of the first papers that investigates the impact of images on the consumer decision-making process for online product search at scale.

- Our method is not just a black-box, unstructured prediction, but instead provides interpretable information that managers can use to decide what images to use as their "champion" image.

The rest of the paper is organized as follows. First, we discuss the related literature on consumer search and choice and the importance of imagery from a marketing and neuroscience perspective. Then, we introduce the concepts from image mining and the learning-to-rank literature that form the basis of our framework. Next, we discuss the methodology and results of the three stages: hotel-level prediction, consumer-level prediction and the fMRI experiment. We conclude with a general discussion of the findings and we list limitations and directions for future research.

4.3 Background

This research focuses on the first stage of the consumer choice process: The consideration set formation. During search, the consumer often tries to accomplish multiple goals, including maximizing choice accuracy and minimizing cognitive effort required (Bettman, Luce, and Payne, 1998). By creating a more narrow consideration set from all possible choices, consumers reduce the cognitive load and they can more efficiently evaluate alternatives (Hauser, 2014). In today's e-commerce environment it is necessary and rational to use heuristics to quickly evaluate the vast number of possible options. During the formation of the consideration set the consumer quickly processes presented information, generally using simple decision rules and using few attributes, and identifies some alternatives to keep around for a deeper examination while discarding others (Moe, 2006). The consumer tries to maximize utility by making trade-offs between these goals. The consideration set formation and the process behind it have been studied extensively in the literature (Andrews and Srinivasan, 1995; Bronnenberg and Vanhonacker, 1996; Draganska and Klapper, 2011; Honka and Chintagunta, 2016). This research has established that the inclusion of a product in the consideration set is critical for the purchase decision.

One of the first pieces of information that a consumer encounters when searching for a product is the thumbnail image of a product and for most e-commerce websites it can take up a large part of the product listing. However, most studies that model consumer search and consideration set heuristics focus on the numerical and textual information that is presented to the consumer, and neglect this large visual stimulus. Few studies have explored the impact of imagery on the consideration set formation. This is most likely due to technical challenges involved with translating unstructured data, such as images, into more structured information that can be analyzed systematically (Ma et al., 2018). We provide a framework that allows researchers to utilize the information conveyed by product images.

It is particularly interesting to understand the importance of images during the consideration set formation and how they influence the process for several reasons. First, visual stimuli are effective in capturing consumers' attention (Pieters and Wedel, 2004; Pieters, Wedel, and Batra, 2010; Wang, Tsai, and Tang, 2018). Compelling visual stimuli can help entice customers to take a closer look at a certain product during consideration set formation. Second, when the consumer is presented with a search result list, the thumbnail image and the product description are listed side-by-side, which means that the consumer can quickly process both types of information per product for several brands/firms (Pieters, Erdem, and Martinovici, 2019). It is in this stage where firms are competing to grab a consumer's attention, where these images may have the greatest effect. In addition, we can investigate the interaction between the images and the textual/numerical product information, such as price and ratings. The link between images, product information and consumer decisions is much harder to establish in a later stage of consumer search. Third, images offer a way for consumers to visualize and imagine the sensory experience associated with the product or service (Sparks and Wang, 2014; Baek and Ok, 2017; Blanco, Sarasa, and Sanclemente, 2010). They can aid online consumers in creating a mental image about the product(Jeong and Choi, 2005), which can make up for the lack of a physical inspection. Fourth, consumers are interested in information that will make them feel good about their choice when making a purchase (Bilgihan, Okumus, and Nusair, 2013). Images can quickly convey this type of information. Finally, images not only help consumers process information to make choices, but it also improves the experience of decision-making and can sometimes enable consumers to consider products that they would not have considered based on other criteria (Pan, Zhang, and Law, 2013). Especially for experiential products, the decision-making process is extensive and risky, because it holds many unknowns and high costs (Sirakaya and Woodside, 2005). To illustrate, Park, Yin, and Son (2019), show that review score and hotel images were the most sought-after type of information by consumers searching for hotels. Our analysis also contributes to this literature by demonstrating that the image conveys important information to consumers, and that it interacts with the other product information that is presented to consumers.

We hypothesize that the product image is an important attribute that consumers consider when making decisions online. However, images are not easily quantifiable in a systematic way for analysis. Recently, we have also seen an adoption of image mining methods for marketing research, especially in online settings where image data is often used. Table 4.8, in the appendix, provides an overview of recent research in Marketing. Our work adds to this line of research by using a combination of methods, such as pre-trained CNNs and a hybrid VGG16 model, and newer methods, such as LambdaMART click prediction method, to explain the relationship between images and consideration set formation.

It is known that the image recognition algorithms are largely based on the biological visual system. In fact, several studies show that CNNs mimic the human visual system such that early layers in a CNN match early visual areas in humans, whereas higher level layers in a CNN match later visual areas in humans (Güçlü and Gerven, 2015; Eickenberg et al., 2017). The visual system has multiple pathways used to solve different tasks and it optimizes these tasks based on multiple cost functions similar to the way deep learning models learn (Scholte et al., 2018). In addition, the effect of visual stimuli on the human brain has received quite some interest in the neuromarketing literature (Ariely and Berns, 2010; Knutson et al., 2007; Ambler, Ioannides, and Rose, 2000; Stoll, Baecke, and Kenning, 2008; Deppe et al., 2005; Erk et al., 2002; Plassmann, Ramsøy, and Milosavljevic, 2012). These studies highlight that the human brain responds differently to varying visual stimuli. The brain regions that are activated by these stimuli help us understand the impact they have on consumers during search for products or choices between brands. A suitable method for studying activation of different brain regions when processing stimuli is fMRI (Couwenberg et al., 2017). This motivates the use of an fMRI experiment to confirm results and expectations that we find from the prediction model based on CNNs. It advances our understanding of what makes images effective and it also advances our knowledge of deep learning methods for (neuro) marketing purposes.

Images are expected to influence the process of consumer search and consideration set formation, either by attracting attention and increasing CTR directly or by influencing the evaluation of other attributes the consumer considers. Either way, multiple researchers are calling for the investigation of the impact of imagery online from both a (visual) marketing (Blanco, Sarasa, and Sanclemente, 2010; Ordenes and Zhang, 2019; Hauser, 2014; Kirillova and Chan, 2018; Liu, Dzyabura, and Mizik, 2020) and neuroscience perspective (Reimann et al., 2010; Jai, O'Boyle, and Fang, 2014). In this study, we answer this call using visual analytics methods and deep learning as well as an fMRI experiment to understand neural responses. In the next section we will describe relevant image mining methods and a learning to rank approach. We will describe how we use these methods to further understand the impact of imagery during consideration set formation of consumers searching for hotels online.

4.4 Framework

To study the impact of images on the consideration set formation of online consumers we apply our method to data from a global online travel agency. There are several studies, both quantitative and qualitative, that have studied the impact of imagery on travel decisions using a variety of methods in different settings (see Table 4.8 in the Appendix for an overview). In two consecutive studies most closely related to our research, Zhang et al. (2017; 2019), show the importance of property images on property demand on Airbnb. Besides explaining the relationship between image quality and short- and long term demand, they hypothesize that the property image displayed on the search result may impact whether or not a property is incorporated in the consideration set. The online travel industry is an appropriate outlet to test our method, because imagery aids the consumer in imagining what they will experience after purchase and this helps reduce the perceived risk (Kim and Mattila, 2011) and assists in creating a mental image of what the experience of staying at the hotel might feel like (Baek and Ok, 2017). In a qualitative study and survey, Noone and Robson (2016) find that hotel images affect the consumer choices during the information search phase and to a lesser extent also during the deliberation phase of the consumer decision making process. It also facilitates consumers developing a view of the destination and inferring what kind of travelers a hotel attracts (Sparks and Wang, 2014; Jeong and Choi, 2005; Noone and Robson, 2016). A positive experience envisioned through the hotel image, therefore has a positive impact on the propensity of a hotel to be considered. And finally, hotel images are an important information source during the early stages of decision-making (Park, Yin, and Son, 2019). The rest of this section is dedicated to the data description, explanation of the method for the extraction of the image information and the learning-to-rank framework.

4.4.1 Data

Our data consist of a large set of consumer searches and the results of those searches for hotels on the website of a prominent global online travel agency. In this dataset, a search starts with a request. Following the request, which includes several parameters (e.g., destination city, travel dates, number of travelers), the website presents the consumer with a list of available hotels in the city on a search results page. Every hotel listing on the search result page consists of the name of the hotel, a thumbnail image, called the "champion" image, price (with potential discount), and the number of stars. After obtaining the default set of results, consumers can click on a hotel on that page, continue to the next page of results or end their search.

Table 4.1: Descriptive Statistics

	Observations	Mean	Median	Sd	Min	Max
Fixed Hotel Data						
Appearances in search results	1,414	1,864	393	4,929	10	52,063
Clicks	1,414	74.57	17	225.87	0	4618
Number of stars	1,414	3.18	3.5	1.17	1	5
Chain	1,414	0.47	0	0.50	0	1
Type of hotel (hotel vs. other)	1,414	0.60	1	0.50	0	1
Variable Hotel Data						
Price	2,612,397	271.35	233.36	268.68	14.25	9,999
Discount percentage	2,612,397	0.19	0.15	0.15	0	0.97
Free cancellation	2,636,000	0.32	0	0.47	0	1
Sponsored	2,636,000	0.33	0	0.47	0	1
Consumer Search Data						
Days in advance	40,553	41.19	16	63.08	0	500
Length of stay	40,553	2.66	2	2.37	1	28
Number of adults	40,553	2.09	2	0.80	1	7
Number of kids	40,553	0.22	0	0.62	0	5
Number of rooms	40,553	1.06	1	0.29	1	8
Number of clicks	40,553	2.60	1	3.15	0	218

The descriptive statistics in Table 4.1 show the hotel information that is fixed (i.e., hotel information that is the same for every consumer search), the variable hotel information (i.e., hotel information that can be different per consumer search query) and consumer search information (i.e., the parameters specified by the consumer at the start of the search). The fixed and variable hotel information are the data that we were able to collect that we have referred to as the other attributes earlier in the paper. The thumbnail image is a fixed piece of hotel information, because it does not change between consumer search sessions.

We collected every consumer search for five major destinations in the United States (Boston, Miami, New York, San Francisco and Seattle) for the month of July, 2019. Each query resulted in a search result page, where 25 hotels are listed [2], and if they clicked on a listing, then a hotel info page, where one particular hotel and all of its information are presented. Not all consumers necessarily click through to an underlying hotel, but a large number do click through. These clickthroughs are the focal priority of this study, because we want to investigate the impact of the thumbnail or "champion" image on the decision to click on a certain hotel. What the consumer observes afterward and the impact that this information has on the second stage of search is beyond the scope of this research, since we do not observe actual purchases.

We cleaned the data for outliers and non-US travelers. [3] In addition, we removed

[2] Every search result page in our dataset consists of 25 hotels.

[3] We decided to delete consumers from outside of the US, because observed prices were inconsistently saved. About 72 % of the search sessions were from US locations.

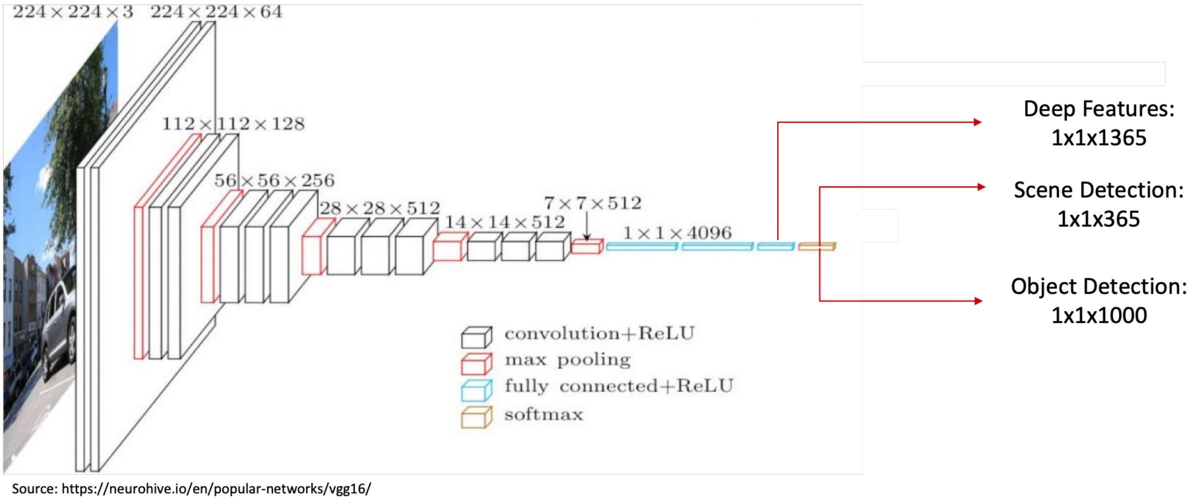


Figure 4.1: The VGG16 architecture (Simonyan and Zisserman, 2014) and an illustration of the three types of features. The deep features come from the output of the last fully connected layer of a hybrid network trained on both Imagenet (Russakovsky et al., 2015) and places 365 (Zhou et al., 2017).

sessions with incomplete hotel information, meaning that we did not include any session or search results for which we did not have fixed hotel information and the hotel images available. We observe that an online traveler clicks 2-3 times on average, and that the majority of search sessions consist of a single click. We can not track particular customers after a session, so we do not observe consumers that come back to continue the search. In this case, we have treated them as a new session.

4.4.2 Feature Extraction

For our application, we use two pre-trained CNNs that identify objects and scenes. In addition, we use a novel hybrid of the two networks to extract our deep features. A visual representation of the VGG16 architecture and the feature dimensions can be found in Figure 5.1.

The automatic identification of objects in images has received considerable academic research attention since the start of the ImageNet Large Scale Visual Recognition Challenge (Russakovsky et al., 2015). The challenge evaluates algorithms for object detection and image classification at large scale. As part of the challenge, a dataset is provided with millions of label images on which CNNs, or any machine learning model, can be trained. For the identification of objects in hotel images we use the CNN that won the ImageNet challenge in 2016 (He et al., 2016). This CNN returns a distributional representation of 1000 common objects detected in the image. For instance, objects in the ImageNet challenge include: armchair, trundle bed, desktop computer, and doormat[4]. In other words, for each of the 1000 ImageNet objects that were labeled in the training set, the He et al. CNN returns a probability score of the particular object being present in the image. When we apply this CNN to the hotel data, the final result is a distributional representation of all of the objects present in every hotel image.

For scene classification we use a deep neural structure trained on previous images of different locations, called the Places Database (Zhou et al., 2017). The Places Database consists of 10 million scene photographs, all labeled with scene semantic categories. It comprises a diverse list of types of environment encountered in the world. For instance, scenes include: Lobby, Jacuzzi, Dorm Room, and Building Facade. The deep learning model accurately identifies 365 scene categories depicted in images. Similar to object detection, the pre-trained CNN returns a probability score for each of the 365 scene categories in the image. The final result is a distributional representation of the identification of scenes for every hotel image in our dataset. A deep neural net can be disassembled into a series of layers that encode different amounts of information (see Figure 5.1). The final layer of the network is the one used to make traditional predictions, but the layers before the final layer, i.e., the deep layers have been shown in previous research in social media analysis to also contain information that is useful for prediction (Khosla, Das Sarma, and Hamid, 2014; Mazloom et al., 2016). The way this is done is by creating a model that relates the output of the last fully connected layer to the popularity of a set of known social media images. The output of the last fully connected layer represents raw image information that has not (yet) been translated into a meaningful prediction, yet it has been structurally processed. In the pre-trained CNNs the output of this last fully connected layer is the input of the softmax layer where it is translated into probabilities. We use the last fully connected layer of a novel hybrid model trained to recognize both objects and scenes. The main reason these deep features work well at prediction is because essentially it is a transformation of the information from the pixel-level information to structured information about the image. The closer the layer of a CNN is to the final layer, the more this information is structurally related to what the model is trained to recognize. Generally, one of the last layers, called the softmax layer, turns these deep features into the classification probability. In cases where the only interest is in classifying the image contents, this last step helps us to understand and interpret what is depicted in the image, but it does not necessarily help in cases where the goal is predicting popularity or CTR. By using the output before the softmax layer we essentially have access to more of the image information and this can create better predictions of CTR than predictions made directly from

[4] `http://www.image-net.org/challenges/LSVRC/2010/browse-synsets`

the final concept classifications. In this paper, we use these "deep features" to make predictions of CTR. We do this by applying the concept of transfer learning. Transfer learning is when a researcher uses a neural net or machine learning model that is trained for one task to perform an entirely novel task. In this case, that means going from classifying which objects or scenes are in the image to directly predicting the ctr of the image. This helps explain more of the variability in the performance of hotel images by constructing a customized model.

In addition, to the visual information we can extract using CNNs, we use visual complexity measures. Visual complexity influences the attitude towards ads and brands (Pieters, Wedel, and Batra, 2010). In addition, Scholte et al. (2009) show that contrast values in natural (i.e., non-artificial) images follow a Weibull distribution and the parameters beta and gamma of this distribution explain much of the early responses in the visual system. Groen et al. (2013) slightly modify these values to obtain the Contrast Energy (CE) and Spatial Coherence (SC) of images. We use the CE and SC to model the visual complexity of our imagery.

4.4.3 Learning-To-Rank

We are interested in understanding the impact of images on consumer consideration while searching online. The Learning-To-Rank literature from computer science studies this from the approach of ranking search results based on relevance to the consumer or the query (Chapelle, Chang, and Liu, 2010). In essence, information retrieval and consumer search are both ranking problems (Yoganarasimhan, 2020). The most relevant product to the consumer should be the one that the consumer clicks on. Therefore, the search engine should present the products that are most relevant first and less relevant ones later. In the case of understanding the impact of images, our approach is also ranking, yet from a different angle. Our aim is to rank the hotels in our data set, based on hotel-level aggregate CTR or based on consumer-level clicks within a query, because we are interested in learning what makes a hotel more relevant than another. Therefore, if we can predict the ranking of hotels relative to the others we can infer what aspects made this hotel clicked on over others. Building on this we can try to create prediction models based on a lists of hotels (search queries) and clicks, and learn a ranking model to understand what hotels are most relevant and clicked on. The attributes that are used for the decision to click can then be highlighted using such a model. We perform two ranking studies to investigate the importance of images. First, we use features extracted using CNNs to predict ctr on the hotel level with Support Vector Regression (Fan et al., 2008) to highlight the importance of images for hotels. Second, we utilize LambdaMart (Wu et al., 2010;

Chapelle, Chang, and Liu, 2010; Burges, 2010) to predict individual clicks of hotels within a search result page, to understand how important images are for consideration and how they compare to other attributes.

4.5 Study 1 - Hotel-Level Prediction

In Study 1, we predict a hotel's CTR based on the image information alone. The goal is to see if we can make accurate predictions using just the image information, so that we can highlight the importance of the image as a hotel attribute and to see if the image reflects inherent qualities of a hotel. In turn, we also use a visualization method to offer some explanation and to highlight the main aspects behind effective images. In Study 1, we don't use the other hotel data, because we will investigate this in more detail in Study 2.

The CTR is measured by the number of times a hotel appears on a search result page divided by the number of times a hotel is clicked on by the consumer.

4.5.1 Prediction Model

We use 70% per cent of our data for training and the remaining 30% for testing. Suppose $H = H_{TR} \cup H_{TE}$ is a set of m hotels where $H_{TR} = \{(H_1, y_i), (H_2, y_2), ...,$ $(H_k, H_k)\}$ is a training set consisting of k hotels and $H_{TE} = \{(H_{e+1}, y_{e+1}), ..., (H_m, y_m)\}$ is a test set consisting of the other hotels in B. By dividing the hotels into a training and testing set we define H_{TR} and H_{TE} as two matrix representations of all hotels and the extracted image information represented by a feature set F, where each row is a hotel and each column is a feature of that hotel. We then apply a support vector regression four times using four sets of features extracted from the images. Specifically, we apply a support vector regression of the extracted features on the CTR of hotels using: (1) the output of the *Places365* neural architecture, (2) the output of the *Objects* detection neural architecture, (3) our *Deep Features* neural architecture that uses the last fully connected layer before the softmax layer of both neural net architectures, and (4) a *Combination* architecture that uses the features of all three of the previous architectures:

$$\begin{aligned} H_{TR} &= [F_1, F_2, ..., F_e] \\ H_{TE} &= [F_{e+1}, F_{e+2}, ..., F_m] \end{aligned} \tag{4.1}$$

Each row of H_{TR} and H_{TE} represents a hotel. We train a model on H_{TR} and report the result of prediction on testing set H_{TE}.

Let F_i be a set of features extracted from hotel i and y_i the corresponding CTR of this hotel. The idea is to optimize w, parameter vector of function $f_w()$, on H_{TR}

to minimize the error between y_i and $f_w(H_{TR})) = w^T h(H_{TR})$. We optimize the following objective function:

$$\sum_{i=1}^{e}(y_i - f_w(F_i)) + \lambda * ||w_k||^2 \tag{4.2}$$

which can be formulated as

$$\arg\max_w \sum_{i=1}^{e} \log p(y_i|F_i, w) + \lambda||w_k||^2 \tag{4.3}$$

where $\log p(y_i|F_i, w) = \frac{1}{1+\epsilon^{-w^T F_i}}$.

To find the optimal value of w we use $L2$ regularized loss Support Vector Regression from the LIBLINEAR package (Fan et al., 2008). After training the model and finding the optimum value of w on H_{TR}, we use it for prediction of CTR on H_{TE}. We report the Spearman Rank Correlation between the predicted CTR and the actual CTR to measure the performance. We use a rank correlation, because we are interested in the ability to predict the relative performance of a hotel as compared to another hotel over another hotel, rather than predicting the exact CTR.

4.5.2 Results

Table 4.2: Spearman Rank Correlation between true CTR and predicted CTR

Image Features	Rank Correlation
Places365	0.4195
Objects	0.3046
Deep Features	0.5530
Combination	**0.5650**

As mentioned in the previous section, we use an $L2$ regularized loss Support Vector Regression to predict the CTR of a hotel based on the image information extracted from the thumbnail image for all four sets of architectures that we describe. After predicting the CTR of each hotel at test time we compute the Spearman's rank correlation between the prediction and ground truth, the actual CTR of the hotels. Spearman's rank returns a value between [-1, 1], where a value of 1 corresponds to perfect correlation.

The results in Table 4.2 show the prediction accuracy of using the extracted image information from the pre-trained CNNs to predict hotel clicks. The best performing model is the Combination model, which uses the combined set of the Place365 net, the Objects Net, and the hybrid Deep Features model to predict the ctr. This model results in a correlation of 0.5650, which shows that using the image information we are able to predict relatively accurately the number of times a hotel is clicked on.

When looking at the individual models and not the combination model, the best performance is achieved by the deep features and not by the semantic information of hotels, i.e., the objects and scenes. This object model and the scene model are generic models that were trained on general images and not on hotel images. The performance of the prediction model based on the deep features nearly matches the performance of the combination model. For this reason we use the deep features as input for the visualization and for the consumer-level prediction.

The deep features are much more accurate for prediction, but these features have not been translated into semantic information. To highlight some correlations between the images and the CTR, we can use the scenes/places classification. The pre-trained CNN returns a probability distribution of the image depicting a certain scene, which means that for all 365 scenes that the model was trained to recognize it returns a prediction probability of the image depicting this scene. The SVR, with the scene vector, then estimates weights that correspond to each of these scene scores. We compute the mean of the SVR weights across the 10 train/test splits of the data for cross-validation, and sort them. The goal of this process is to identify which scenes / places are present in the image that are most likely to be associated with high CTR. The weights of the support vector regression with the distributional representation of the scenes as input show us the following correlations:

- **High Positive Impact:** Hotel/Outdoor, Building Facade, Hotel Room.
- **Low Positive Impact:** Skyline, Lobby.
- **Low Negative Impact:** Jacuzzi, Window/indoor.
- **High Negative Impact:** Jail Cell, Dorm Room.

It's important to note that there are no jail cells or dorm rooms in our dataset. Instead it shows that when the thumbnail of a hotel receives a high probability score for either of these scenes from the pre-trained CNN this is not a good sign. Most likely, the image will look like a jail cell or a dorm room to the consumer as well. This is relevant managerial information, since the classifications can highlight whether or not an image should be included in an online product display.

4.5.3 Explanatory Model

The support vector regression model gives us the ability to predict the potential CTR for images and gives some insight into the aspects of images that work well overall, but it does not easily allow for interpretation of the underlying structure of the data and does not control for differences between locations. Images are extremely high dimensional and so identifying one particular aspect that explains why some images do better than others is very difficult. Thus, we rely on dimension reduction to manage the high dimensions of the CNN features. A commonly used method in image research to analyze image data in a constructive manner, is an embedding algorithm that maps high-dimensional data onto a two dimensional space, called the t-distributed Stochastic Neighbor Embedding (t-SNE) (Maaten and Hinton, 2008). The t-SNE algorithm is very effective in visualizing high-dimensional data by assigning each datapoint a location on a two-dimensional map. It maps images based on their similarities which enables a quick examination of what is generally used as the "champion" or thumbnail image by hotels. We take the output of t-SNE to graphically represent the images that are generally used by hotels on a two-dimensional space while highlighting the best performing images, and controlling for location. This visualization enables the capture of commonly occurring elements as well as the heterogeneity across locations and will help us to better understand what aspects of the images do better than others.

Figure 4.2 visualizes all the thumbnail images that are used by all the hotels in our dataset. We can observe that it accurately maps the images that are similar to each other close together in the two-dimensional space. The green squares indicate the five best performing images in the space. Figure 4.2 shows that hotels tend to prefer a few specific types of images: pools (indoor and outdoor) dominating the left side of the figure, the front of a hotel on the upper side of the figure and the hotel rooms, which cluster on the bottom. The rest of the space is filled with images that have the lobby, additional interior images and skyline views. In the overall image that combines all of the cities, the best performing images are scattered across the space, meaning that overall the exact type of image does not seem to matter in terms of obtaining the highest possible CTR.

However, we observe in Figure 4.3 that for each of the five destinations the images that are used can be clustered by similarities into 2 or 3 major clusters. It also shows that the high performance images are clustered as well, which indicates that there are particular types of images that work well for each of these locations. For example, it is apparent that for New York City (top right) the images that work best are the images that show the front of the hotel. Specifically, these are the images with the entrance of a hotel that match the urban style of New York. As for Boston (upper left) and San Francisco (bottom left), it is more common that hotel images that portray the hotel room do well. These results seem consistent with the travel marketing literature discussed below.

Figure 4.2: The t-SNE visualization of all the thumbnail images used by all hotels across destinations in our database. The green squares indicate the 5 best performing hotel images.

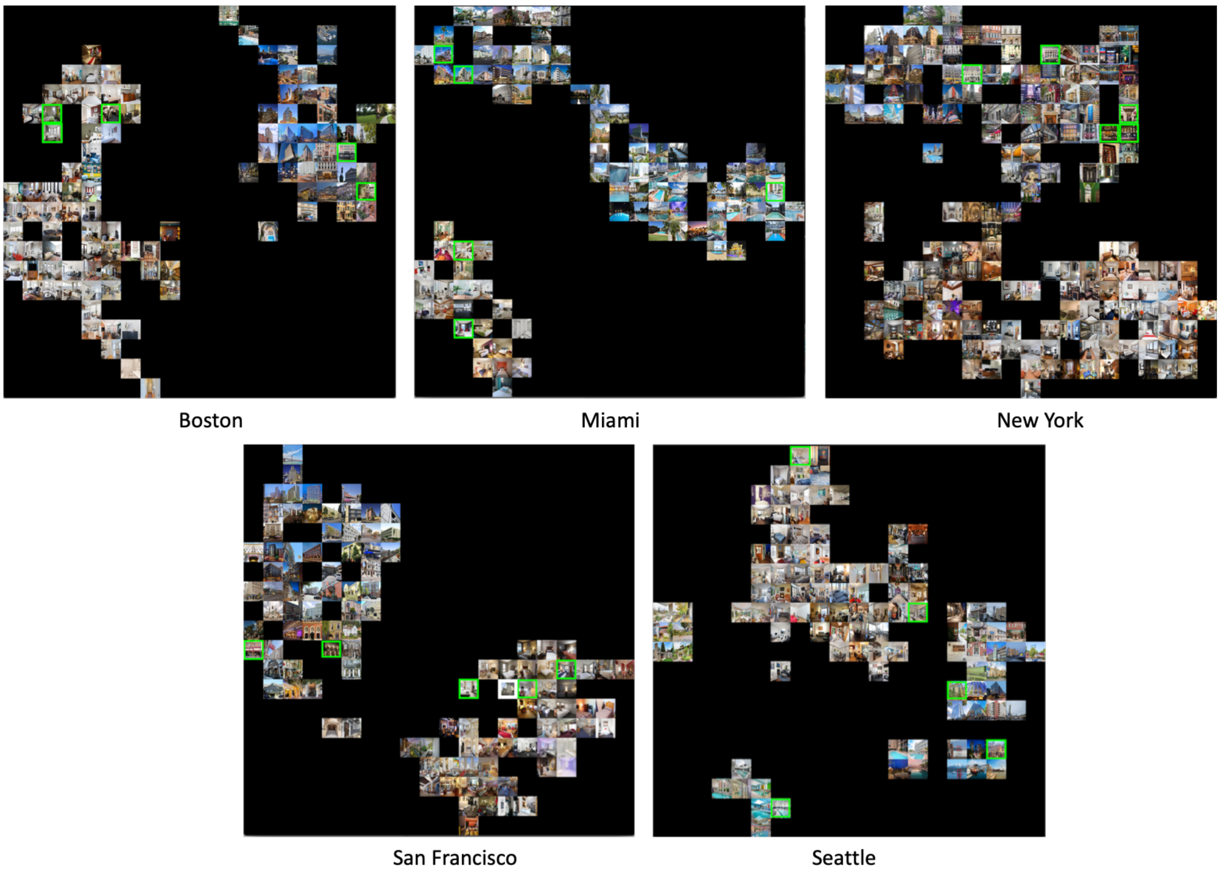


Figure 4.3: The t-SNE visualization of the hotel images in the 5 different locations. The green squares indicate the five best performing images of each location.

4.5.4 Discussion

The support vector regression model predicts CTR fairly well based on image-based data.[5] There are two potential explanations for the prediction accuracy of CTR using just the image features: First, images are important to consider because consumers use images to help determine their consideration set. We know that images help increase decision-making efficiency, because consumers can detect and process image features more easily than text (Blanco, Sarasa, and Sanclemente, 2010; Zhang et al., 2019). They can help consumers quickly imagine what it would be like to stay at a hotel (Jeong and Choi, 2005; Noone and Robson, 2016). Second, images often depict the other attributes of a hotel. For example, one can think of a situation where an online traveler can quickly, approximately infer hotel class, price, or certain amenities from the images. This would be in line with previous research that shows that consumers sometimes use images as a substitute or complement to examinations of other attributes (Pan, Zhang, and Law, 2013; Kirillova and Chan, 2018). From the SVR weights we can derive some key insights into the best performing champion images in several ways. First, it shows that either the hotel room or the front of the hotel work well in general. Given that users are probably concerned with the appearance of the hotel and the room that they will be staying in, this is not a surprise. Second, images classified as a jail cell or dorm room generally do not elicit customers to click on hotel listings. There are no actual jail cells or dorm rooms in our dataset, but the fact that the hotel images appear somewhat similar to these concepts is not a good sign. Managers could use this information to provide objective insights into what works and what does not work when generating images for OTAs. Using the t-SNE we were able to draw relationships between the different aspects of images and the decision by the consumer to include the hotel into their consideration set. Our method is not just an unstructured prediction, but instead provides interpretable information that hotel managers can use to decide what images to use as their "champion" image. The mapping shows that there is quite a variety of images that are used by hotels and that in general people like to see where they will be sleeping or what the hotel building looks like. By using the unsupervised clustering system based on t-SNE we capture the heterogeneity across images in different locations. The location maps clearly show that each location has a set of images that perform best. For instance, New York City hotels do best with an image of the entrance of the hotel, whereas in Seattle people are mostly interested in seeing the view from where they are staying; while, Miami hotels aim to entice customers with their pools and the hotel rooms

[5] As a reference, in social media image popularity prediction studies we observe prediction-ground-truth rank correlations of 0.4-0.5 for Flickr (Khosla, Das Sarma, and Hamid, 2014) and up to 0.5 for Instagram (Mazloom et al., 2016)

generally work well in Boston and San Francisco.

Sparks and Wang (2014) show that water-based features enhance liking. Our results confirm this and it makes sense that in an exotic and sunny destination such as Miami these images elicit attention for hotels. As for the views of the hotel room, the findings of an eye-tracking study showed that a nature-based servicescape is the key to attracting customers' visual attention (Wang, Tsai, and Tang, 2018). New York City and the front of the hotel could suggest that the urban feel and architectural style that is aesthetically pleasing are important (Kirillova and Chan, 2018). Noone and Robson (2016) show that some participants favored hotels that had a particular style. It is also interesting that for Boston and San Francisco people seem to care more about the room that they will be staying in. This could be because these destinations are dominated by business travelers more than the other destinations in our dataset, and that business travelers care more about having a good location to work from. In study 2 we will reinforce the findings of study 1 and we will further explore the importance of image attributes and their relation to other hotel attributes.

4.6 Study 2 - Consumer-Level Prediction

Now that we have established the importance of images for the prediction of CTR for hotels on the OTA website, we will build a prediction model that uses individual consumer search and clicks. Before, we were interested in modeling the number of times a hotel was clicked on in the search result page relative to the number of times the hotel appeared in the search result (e.g. CTR). In this study, we examine the individual consumers, with certain search parameters (e.g., length of stay, number of adults and kids, advance booking time), and predict the probability that a certain hotel in the list will be clicked on. We approach this again as a ranking problem, in the sense that a consumer click provides a piece of relevancy feedback for a hotel in a list. This means that, at training time, we can model pair-wise preferences between pairs of hotels on the search result page based on hotel characteristics and we use consumer clicks as indicators of relevance to the consumer. At test time, we can then predict relevancy scores for each hotel on the search result page and predict the conditional probability that a hotel is clicked on by the consumer. The approach comes from re-ranking documents in search results based on their relevancy to the query, which translates well to modeling clicks for consumer search when we assume that a consumer clicks on the hotel that is most relevant to her at that time. Essentially, we predict the probability of inclusion of a hotel in the consideration set of a consumer.

We choose the LambdaMART algorithm (Burges, 2010) to solve this ranking problem. LambdaMART is the leading Learning-To-Rank algorithm, because of its flexibility, approximation ability and interpretability.[6] LambdaMART consists of two main parts and functionalities that need explaining: 1) LambdaRANK, this is the pairwise learning-to-rank algorithm to model pairwise preferences between documents, and 2) Multiple Additive Regression Tree (MART), which is a Boosted Regression Tree algorithm that estimates an ensemble of regression trees. Essentially, LambdaMART is a pair-wise ranking loss function ("Lambda") attached to a gradient boosting machine ("MART"). We first explain briefly how we go from the LambdaRank specification to LambdaMART. For a full step-by-step explanation of the construction of the LambdaMART algorithm we refer you to Appendix B, and Burges, 2010.

4.6.1 LambdaMART

In the learning-to-rank literature the most commonly used loss function is the Normalized Discounted Cumulative Gain (NDCG), which is also used for the personalized search ranking mechanism by Yoganarasimhan (2020)[7]. We use this metric to train our model and to update the weights of the scoring function. The LambdaRank model is based on the idea that we only need the gradient of the costs with respect to the model scores and not the costs themselves Burges, Ragno, and Le, 2007, which means that the NDCG is used to directly update the weights.

The main goal of this class of learning to rank models is to design a scoring function that scores documents/products (or in this example: Hotels) based on relevance to the search query, and thus the consumer. The updating of the weights in the scoring function in Lambdarank is done by gradient descent using pairwise preferences within a list. Instead of specifying the gradient of the cost with respect to the model scores, we can model the gradient with respect to the scores directly using gradient lambdas.

These gradient lambdas are essentially forces that move hotels up or down predicted rankings. If hotel i turns out to be more relevant (i.e. when it is clicked) than hotel j then hotel i will receive a push of λ_{ij} up the list and hotel j a push of the same magnitude down the list. The magnitude is the change in NDCG that can be obtained by swapping the hotel i and hotel j in ranking. This magnitude is 0 when hotel i was already predicted to be the most relevant, whereas this magnitude gets larger the lower the ranking of hotel i. The weights of the model are then updated based on these

[6] For an excellent use case of the LambdaMart algorithm we refer you to Yoganarasimhan (2020). The author of this paper shows that by personalizing the search result rankings using LambdaMART the clicks to the top position improves by 3.5% and the average error in rank of a click reduces by 9.43%.

[7] see section 5.2 in their paper for a more detailed explanation of this metric

forces. In our setting, each time a hotel is clicked on in a list, this hotel will receive a push up the ranking of λ_{ij} relative to every other hotel that ranked as preferred over hotel i and all these hotels will receive one push downward of the same magnitude. The weights of the function f will then be updated with $\delta w = -\eta \sum_i \lambda_i \frac{\partial s_i}{\partial w}$. Where λ_i is the sum of all positive and negative forces a particular hotel.

For instance, in a search result list of 25 hotels, 1 hotel is clicked on. Let this hotel be hotel i, then the lambda for this hotel is the sum of all λ_{ij}'s based on scores s_i and s_j and the lambda for all other hotels equals -λ_{ij}. Note that all other documents are un-clicked, so the lambda of these pairs equal zero, because they are considered equally (ir)relevant and the model "ignores" these. The weights of the scoring function can then be updated based on the notion that hotel i is more relevant than all others. The scoring model is thus trained on these pairwise preferences.

The lambda's can be defined by (please refer to the Appendix to see how to arrive at this specification):

$$\lambda_{ij} = \frac{\partial C(s_i - s_j)}{\partial s_i} = \frac{-\sigma |\Delta NDCG_{ij}|}{1 + e^{\sigma(s_i - s_j)}} \tag{4.4}$$

The specification of the gradients of LambdaRank, are the same as the gradients of the cost with respect to the model scores in MART (Friedman, 2001). MART improves its predictions by training a new tree to predict the errors (using the gradients of the loss function) of the trees which came before. MART models derivatives and LambdaRank works by specifying the derivatives at any point during the training. LambdaMART is the result of combining these two algorithms. As suggested by Yoganarasimhan (2020), we implement LambdaMART using a Gradient Boosting Machine that utilizes newton boosting steps. The newton step[8] is used to determine the optimal step size that minimizes the loss of the regression tree, using least squares to compute the splits of each node. In LambdaMART each tree computes the lambda per hotel across the entire dataset.

Here we are maximizing the NDCG (Burges, 2010), so we update the weights by:

$$w \rightarrow w + \eta \frac{\partial C}{\partial w} \tag{4.5}$$

Recall that λ_i is the sum of all positive forces minus the sum of all negative forces, constituted by each time hotel i is clicked on for the former and each time hotel i is not clicked on for the latter, compared to the predicted ranking of the other hotels such that we can define:

[8] for a detailed explanation of the newton step to calculate the optimal step size, we refer you to (Friedman, 2001) or (Burges, 2010)

$$\lambda_i = \sum_{j:\{i,j\}\in I} \lambda_{ij} - \sum_{i:\{j,i\}\in I} \lambda_{ij} \tag{4.6}$$

Where I contains a set of pair indices for which hotel i was clicked on and all hotels j were displayed on the search result page. Therefore the first summation represents every time hotel i was clicked on and all hotels j were displayed, whereas the second summation represents every time hotel i was displayed, while another hotel j was clicked. Burges (2010) simplify this notation by

$$\sum_{\{i,j\}\rightleftharpoons I} \lambda_{ij} = \sum_{j:\{i,j\}\in I} \lambda_{ij} - \sum_{i:\{j,i\}\in I} \lambda_{ij} \tag{4.7}$$

For any given state of the model and a given hotel i we can write down a utility for which λ_i is the derivative of that utility.[9]

$$C = \sum_{\{i,j\}\rightleftharpoons I} |\Delta NDCG_{ij}| log(1 - e^{\sigma(s_i - s_j)}) \tag{4.8}$$

such that,

$$\frac{\partial C}{\partial s_i} = \sum_{\{i,j\}\rightleftharpoons I} \frac{-\sigma|\Delta NDCG_{ij}|}{1 + e^{\sigma(s_i - s_j)}} = \sum_{\{i,j\}\rightleftharpoons I} -\sigma|\Delta NDCG_{ij}|\rho_{ij} \tag{4.9}$$

where $\rho_{ij} = \frac{-\lambda_{ij}}{\sigma|NDCG_{ij}|}$ and

$$\frac{\partial^2 C}{\partial s_i^2} = \sum_{\{i,j\}\rightleftharpoons I} \frac{-\sigma|\Delta NDCG_{ij}|}{1 + e^{\sigma(s_i - s_j)}} = \sum_{\{i,j\}\rightleftharpoons I} \sigma^2|\Delta NDCG_{ij}|\rho_{ij}(1 - \rho_{ij}) \tag{4.10}$$

Therefore the newton step for the k^{th} leaf of the m^{th} tree is given by:

$$\gamma_{km} = \frac{\sum_{x_i\in R_{km}} \frac{\partial C}{\partial s_i}}{\sum_{x_i\in R_{km}} \frac{\partial^2 C}{\partial s_i^2}} = \frac{\sum_{x_i\in R_{km}} \sum_{\{i,j\}\rightleftharpoons I} |\Delta NDCG_{ij}|\rho_{ij}}{\sum_{x_i\in R_{km}} \sum_{\{i,j\}\rightleftharpoons I} |\Delta NDCG_{ij}|\sigma\rho_{ij}(1 - \rho_{ij})} \tag{4.11}$$

$x_i \in R_{km}$ represents the data points that land in the k^{th} leafnode of the m^{th} tree.

Implementation

We train and test a model for each of the five locations in our dataset. We use 70% of the queries for training and 30% of the data for click prediction in the testing set. As described above the metric that we are using to optimize during training

[9] We are maximizing the NDCG metric, so we have a utility instead of a cost

is NDCG. We also evaluate the click prediction using the NDCG. Yoganarasimhan (2020) show that results are robust to any metric used, but that NDCG offers the best performance. There are several hyperparameters that affect performance of the model. We use cross-validation to find the optimal number of trees. We then select the number of trees with the highest performance for each location. We experimented with several hyperparameters and use 0.1 for the learning rate and 255 leaves per tree for optimal results. We use LightGBM in Python to train our model.

4.6.2 Features

Prior work has identified five factors that have a substantial impact on hotel considerations: price and discount, hotel category, brand, location, and number of stars (Kaldis and Kaldis, 2008; Musante, Bojanic, and Zhang, 2009). And the focal priority of this study: hotel images. In addition, there are consumer characteristics and trip characteristics that influence the decision: number of adults, number of kids, number of rooms, length of stay, booking in advance etc. We distinguish between 3 types of information: Fixed hotel information, variable hotel information, customer search information.

Fixed Hotel Information

This hotel information is constant throughout the data collection period. Out of all the clicks that we gathered and the search results, we obtained information for 1,414 hotels across our five locations. We are able to observe the following fixed hotel information: ***Hotel Images - Deep Features*** - we extract information from the image as described in the image classification section. Here we use Principal Component Analysis to shrink the dimensions of the deep features from 1365 to the number of principal components that explain 75% of the variation in the hotel images for a particular location. This speeds up the training process and it will provides us with some interpretability as well. ***Hotel Images - Visual Complexity*** - we extract two types of visual complexity measures that have been shown to impact early visual processes in the brain: CE and SC (Groen et al., 2013). ***Number of Stars*** - The number of stars of the hotels ranging from 1.0 to 5.0.

Variable Hotel Information

This hotel information is variable for every search, depending on time of day, availability and type of search by the consumer. We obtained the following variable hotel information: ***Price*** - this is the price that is displayed to the consumer. ***Discount*** - in this case there is an “original” price that is struck through, the displayed price is the discounted price. It is both an indicator of a strike through and a percentage discount. ***Free Cancellation*** - there are several special indicators for our OTA, but this is the only one we can observe in our dataset. ***Sponsored*** - this is an indicator for a sponsored listing. This is indicated to the consumer. Generally these are on the top of the page, but we have also seen instances for lower ranked hotels.

Consumer Search Information

The consumer search information differs per session and we have indicators for certain actions taken during the session. We are able to observe the following information: ***City*** - One out of 5 major destinations in the USA: Boston, Miami, New York, San Francisco or Seattle. ***Entrypage*** - there are 5 ways a customer has entered the search process: homepage, hotel search result, hotel information page, referral (such as Yelp or Tripadvisor), or other. ***OTA Brand*** - Our OTA has several daughter brands through which the search could have happened. The information presentation (apart from brand colors) is consistent across brands. ***Length of Stay*** - How many nights is the consumer trying to book. ***Days in Advance*** - How many days in advance is the consumer searching for hotels. ***Number of adults*** - How many adults are indicated by the consumer for this search. ***Number of children*** - How many children are in the consumer's party. ***Number of rooms*** - How many rooms is the consumer searching for.

The total number of constructed features is 39 excluding images and 59 including the principle components and the visual complexity from the images. [10]

4.6.3 Results

We use the method described above to learn a ranking system that learns relevance scores for hotels in a search result. Instead of reranking the hotels based on these scores and presenting these to consumers, we use the ranking scores from the output of the model to predict what hotels the consumer is most likely to click on. The NDCG score in test time reflects how the predicted ranking based on the model scores matches the hotel that was clicked. For example, the score would be 1 if the hotel with the highest prediction score was actually clicked on and it would be at the lower bound of 0.2 if the clicked hotel received the lowest relative prediction score. A score of 0.5 would mean that, on average, the clicked hotel would be among the top 5 predicted scores. What we are mainly interested in is the improvement in the NDCG

[10] The number of features varies per hotel location, because some of the categorical features have empty categories for particular hotels and the number of components that explain 75% of the variance differs per location, because of the number of hotels in those locations.

performance when we include the image features as compared to the model where we use only the other features and later we compare the importance of all features to the model. Note that this is robust to any type of accuracy measure Yoganarasimhan, 2020.

Prediction Accuracy

The results for click prediction in the testing data using the LambdaMART model are presented in Table 4.3. We compare the results for: 1) A random prediction model, in which we assign random scores to each hotel presented on the search result page. 2) The LambdaMART model using all features except the images. 3) The LambdaMART model using all features including the image features. We observe fairly similar prediction accuracy across locations for the non-image features, ranging from an NDCG of 0.4380 in New York City to 0.4914 in Miami, with a much better performance than a random prediction model. After including the image features to the model, we observe improved prediction across locations, with NDCG scores ranging from 0.4726 in Seattle to 0.5717 in Miami. Table 4.4 shows the improvement of the prediction after inclusion of the image features. The improvement is lowest for Seattle with minimal improvement of 3.0%, then San Francisco with an improved prediction of 6.7% as compared to a model without image features. Boston, Miami and New York City all see a big improvement prediction accuracy, improving prediction with 13.1 %, 16.3 % and 12.5 % respectively.

Table 4.3: Normalized Discounted Cumulative Gain for click prediction in test per location. Comparison between a random model, a model with all features except image features, a model with all features including image features.

	Boston	**Miami**	**New York City**	**San Francisco**	**Seattle**
Random	0.3249	0.3252	0.3250	0.3254	0.3247
Non-Image	0.4584	0.4914	0.4380	0.4533	0.4587
All Features	**0.5186**	**0.5717**	**0.4926**	**0.4837**	**0.4726**

Table 4.4: Improvement in model performance due to inclusion of image features

Location	**Improvement**
Miami	16.3%
Boston	13.1%
New York City	12.5%
San Francisco	6.7%
Seattle	3.0%

Feature Importance and Change

From regression trees it is easy to determine the importance of certain variables or features to the target variable of interest, in our case clicks. The feature importance is a score that indicates the relative importance of a feature when making a prediction from a regression. The feature importance is based on the number of times a feature is selected for splitting, weighted by the squared improvement to the model as a result of each split and averaged over all trees (Elith, Leathwick, and Hastie, 2008). Figure 4.4 shows the feature importance for the top 10 features for each location. The orange bars represent the relative importance of features for prediction using the LambdaMART model without using the image features and the blue bars represent the relative importance of features for prediction with the same model including the image features. What is noticeable, from the orange bars, is that the most important feature for all locations is the price. When we include the image features, however, the importance of price decreases significantly. The blue bars highlight the relative importance of image features, which are the principle components and the complexity features. We observe that the image features make up a large part of the top 10 most important features and in for 4 out of 5 locations they are in the top 3. Interestingly, free cancellation in Miami, became increasingly important for prediction. Table 4.5 shows the change in price importance for each location. The table has the normalized change in feature importance as compared to the other variables. The image features combined take up between 32 % (Boston) and 68% (Miami) of the relative importance. Based on this normalization, we observe a decrease in importance of 79.3% in Miami, about 3% for New York City and slight increases for San Francisco, Boston and Seattle. Especially in Miami, the image features seem to completely take over importance for prediction over price.

Table 4.5: Normalized Change in relative importance for the hotel price feature.

Location	Change in Relative Importance of Price
Miami	-79.3%
New York	-2.9%
San Francisco	+.3%
Seattle	+.6%
Boston	+3.5%

4.6.4 Discussion

The LambdaMART click prediction results confirm the importance of images during consideration set formation found in the hotel-level prediction of study 1. The use of image features improves prediction accuracy with more than 10 % across locations,

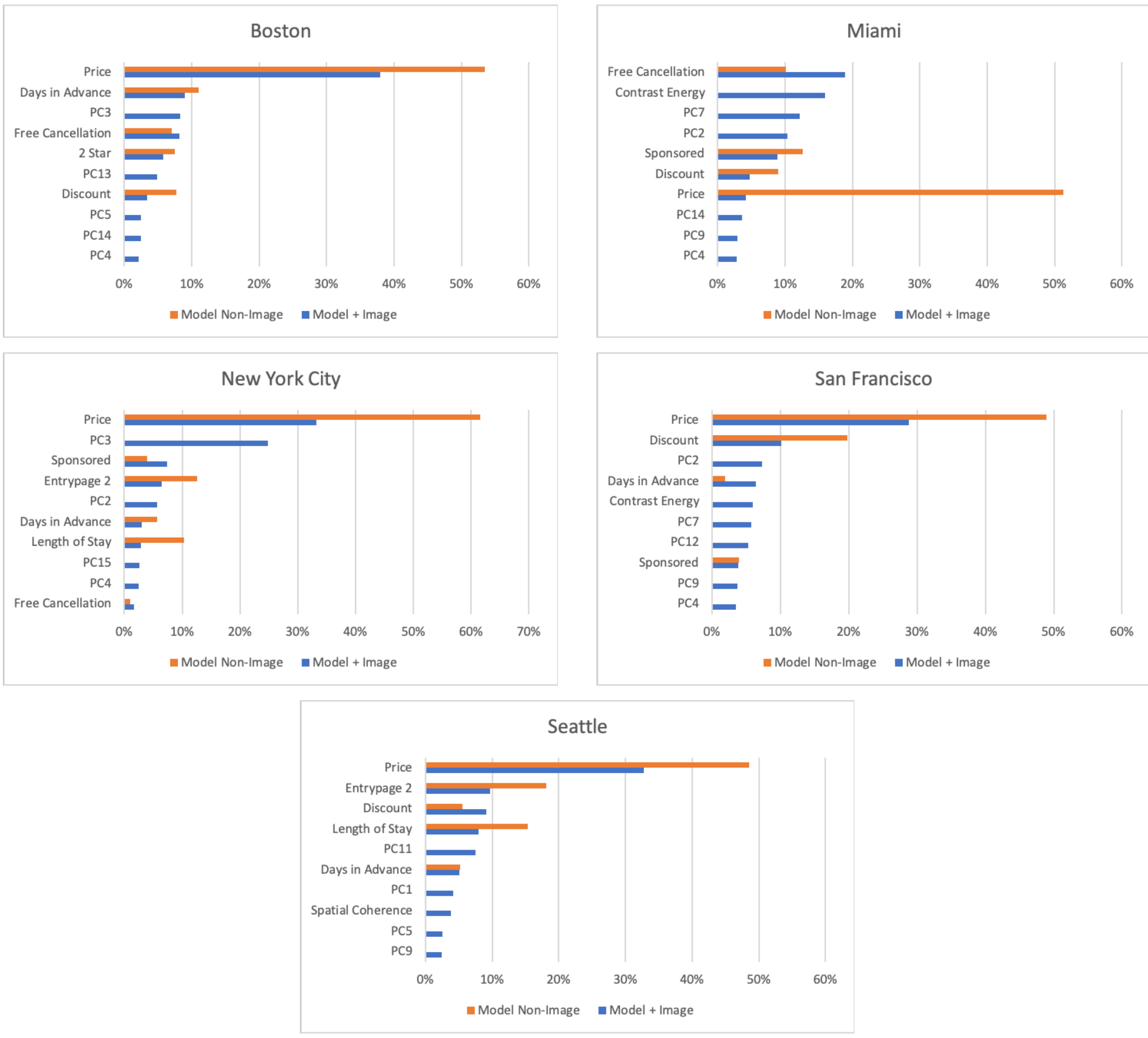


Figure 4.4: Feature importance per location for the top 10 most important variables for prediction using the model that includes the image features. In orange the relative importance of the features for the LambdaMART prediction excluding image features and in blue the feature importance for the same LambdaMART prediction including image features.

indicating that images are important attributes that consumers consider while narrowing down a list of options on a search result page. We included both low-level image statistics and the principal components of high-level images features to cover all aspects of the images that can influence decisions. The results also show the heterogeneity per location, and that the importance of the image is context dependent. For Miami, the image features were much more important than for Seattle. Research has shown that natural images and pretty scenarios can entice potential customers, and one could reason that for an exotic location as Miami, hotel images can portray the appeal of such a location (Sparks and Wang, 2014; Wang, Tsai, and Tang, 2018). We may expect that there are certain locations for which it is easier for a hotel image to stand out than for others.

The results of the feature importance highlight the importance of image features over other attributes. The inclusion of images in the prediction model leads to an overall shift in importance of attributes, where the relative importance of some attributes decreases, but increases for other attributes. These results suggest that images are an important attribute that consumers consider, and that they may even be used as a heuristic, compensating for other hotel attributes (Pan, Zhang, and Law, 2013; Kirillova and Chan, 2018).

In addition, the images contain information that is also reflected by the other attributes and could therefore reduce the relative importance of those attributes, whereas attributes that can not be reflected by the image (such as free cancellation or sponsoring) increase in importance for the same reason. Lastly, previous research has shown the importance of price. Price is an attribute that is mostly considered in the first stage (Moe, 2006) (i.e., consideration set formation) of the search process and we observed that price is one of the features that sees a significant decrease in relative importance when including the image features in some locations, suggesting that consumers might use the image to infer price related information to make quick decisions. For example, there might simply be collinearity between the image and the price, but at the same the image can potentially impact a consumer's price sensitivity as well. Overall, we can assume that images hold more (important) information than their textual/numerical counterparts, and considering that images are processed much easier and faster than text (Blanco, Sarasa, and Sanclemente, 2010; Zhang et al., 2019), we can conclude that consumers use images in their consideration set heuristics (Glöckner and Betsch, 2008; Hauser, 2014).

4.7 Study 3 - Consumer Neuroscience study

In this final study we measure brain activity, using BOLD-MRI, to compare the brain's response to images of hotels with high CTR and low CTR. The goal is to understand if, and if so how, a consumer's brain responds differently to images that are frequently clicked vs. images that are not clicked as often. This experiment can confirm and explain findings of the two preceding studies. Specifically these results will help use to see to what degree the differences in brain responses are limited to the posterior part of the cortex, more involved in perception, or also involve subcortical areas, and the anterior part of the brain that are involved in decision making. We are looking solely at the neural response to viewing the images, not the choice of hotels, and it does not include the rest of the hotel descriptions. This way we can investigate whether the differences are mainly in the regions responsible for visual processing and perception (i.e. posterior part of the cortex) or if there are specific images that trigger decision-making behavior already.

A selection of 102 images where shown in a shuffled order and presented, three times in different runs, to 22 participants in a fMRI scanner.[11] For each image the participants were asked 1 out of 6 agree/disagree questions, to which subjects responded as quickly as possible. The questions served two purposes: to keep subjects engaged, but also to make it possible to perform an analysis, not only on the basis of high vs low CTR's, but also on the basis of the subjects own preferences.

4.7.1 Data analysis

Analysis was performed using FEAT (FMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl) and custom Matlab code. The functional data were motion- and slice-time corrected (Jenkinson et al., 2002). A temporal median filter was applied to remove low frequencies, after which the data was spatially smoothed with Gaussian kernel at a 5 mm FWHM. The preprocessed scans were subjected to voxel-wise event-related GLM analysis using FILM (Woolrich et al., 2001) by convolving the onset times of each trial with a double gamma function to model the hemodynamic response function. We generated, for two different analyses, explanatory variables (EVs). For the first we used the 6 categories described in the stimulus section (13 stimuli per category, per run). We contrasted, per run the high and low CTR categories with each other. The resulting maps were first pooled across runs (fixed effects) and then across subjects (mixed effects using FLAME2, Woolrich, 2008). Results were corrected for multiple comparisons using

[11] see Appendix for more details on the procedure

cluster correction implemented in FSL (z = 2.3, p ¡ 0.05, Worsley, 2001). [12]

4.7.2 BOLD-MRI: results

Table 4.6: Neural correlates related to click-through rates. *Abbrevations in table - COG: Center of Gravity, SMG: Supramarginal Gyrus, PCG: Postcentral Gyrus, STG: Superior Temporal Gyrus, IPL: Inferior Parietal Lobule, PrCG: Precental Gyrus, LOC: Lateral Occiital Cortex, ITG: Inferior Temporal Gyrus*

		High vs. Low CTR			
nVoxels	pVal	Zmax	COG X	COG Y	COG Z
2935	$7.75e-07$	3.81	-58.2	-22.4	13.4
		Areas: SMG Left, PCG Left, STG Left, IPL			
1076	0.00507	3.41	63.2	-13.7	34.7
		Areas: SMG Right, PCG Right, PrCG Right			
		Low vs. High CTR			
nVoxels	pVal	Zmax	COG X	COG Y	COG Z
4796	$7.66e-10$	4.23	25.1	-78.2	10.4
		Areas: V1 Right, V2 Right, LOC Right, ITG Right			
1149	0.00336	3.99	-31.5	-87.2	21.3
		Areas: V1 Left, V2 Left, LOC Left			

When comparing the neural correlates between high and low CTR images we observe that they activate different areas of the brain. Hence, high and low CTR images are processed differently by the human brain. From Table 4.6 and Figure 4.5 we observe that areas that show a difference for this contrast are limited to the posterior part of the neocortex. We see the involvement of early visual cortex (V1,V2) involved in the processing of low-level image properties, and later in time also involved in the integration of visual information, but also the cortical pathway involved in object perception (Lamme and Roelfsema, 2000).

Activity in this pathway also shows a particularly clear correlation with 'activity' in deep neural networks. The early layers of DNN's like Alexnet, VGG and Resnet18 show similar patterns of activation as early visual cortex while the higher visual areas like ITG match activity in the top layers of networks like AlexNet, VGG and Resnet18 (Güçlü and Gerven, 2015; Eickenberg et al., 2017; Scholte et al., 2018). Furthermore, we observe differential activity in the lateral parietal cortex, with the involvement of areas like the inferior parietal lobulus (IPL) and the supramarginal gyrus associated with attention and the perception of space.

[12] Note that we left out of this current analysis, per run, 24 stimuli that can be used for future confirmatory analysis.

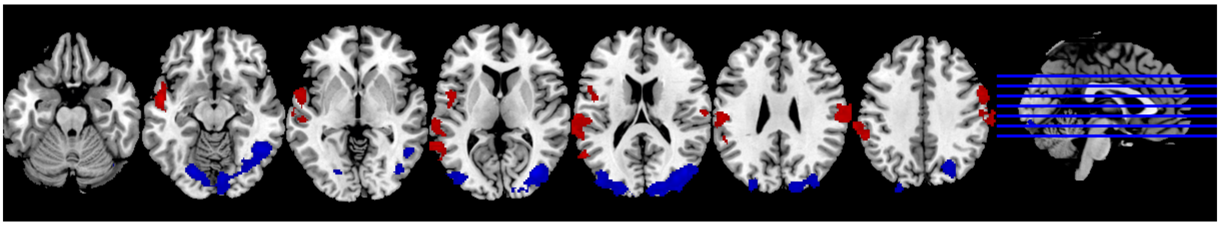


Figure 4.5: Whole brain FMRI results. It visualizes the statistically significant activated regions associated with the High CTR images in red and Low CTR images in blue.

4.7.3 Discussion

The results of the fMRI experiment are relevant for several reasons. There is a link between the way a CNN processes visual information and how the human brain processes visual information (Güçlü and Gerven, 2015; Eickenberg et al., 2017; Scholte et al., 2018). We have established that we can make predictions about CTR using the visual features from a CNN and the results of our experiment demonstrate that there is a link between the visual system and the ctr, which confirms why we are able to make accurate predictions using a CNN. There is increased activity in the visual cortex for low CTR images, which occurs when there is extra visual analysis happening, while this is not observed for high CTR images. A potential explanation could be that the high CTR images lead to immediate (automatic/unconscious) click-through behavior, or feedforward processing (Potter et al., 2014), whereas low CTR images need extra visual inspection. This extra inspection is implemented by feedback processing from top, to lower-tier areas (Groen et al., 2018). An alternative explanation could be that low and high CTR images differ in terms of low-level image features. However, we controlled for this by pre-selecting the images in such a way that they are comparable in this aspect. Viewing the high CTR images activates regions involved in spatial representations and the perception of space more than low CTR images. This could indicate that the high CTR images induce the subjects to go beyond direct visual analysis of the images and engage in visual reconstructions of the scenes. This is relevant for experiential products, because sensory-enabling presentation (i.e. using images) influence the decision process of consumers shopping online by activating the visual perception (Jai, O'Boyle, and Fang, 2014). Importantly we see no involvement of regions of the brain that are classically involved in decision making and buying behavior. The results indicate that high CTR images are processed differently visually, and more extensively in higher-level areas, but do not, by themselves, induce 'buying' behavior. Referring back to the product-level study, we can conclude there are certain types of images that more effectively activate the visual perception. It is important for managers to be aware of the types of images that are most effective at this. As shown, effectiveness is very destination specific. The consumer-level prediction scores could be a starting point for selecting the most effective image. Future research could investigate such an image scoring mechanism in more detail. The fMRI experiment highlights that the activated brain regions are effective predictors of image effectiveness (Couwenberg et al., 2017), which, in turn, can be approximated using CNNs (Eickenberg et al., 2017).

4.8 Overall Discussion

In this project, we used visual analytics and artificial intelligence to understand the role of images during consideration set formation of consumers searching for hotels online. Using different methods we investigated the importance of images and how they impact other hotel attributes, such as price and number of stars. In study 1, we used a hotel-level CTR prediction model to show that we can make accurate predictions about a hotel's performance on the website of an OTA. In addition, we used t-SNE, a dimension-reduction and visualization method (Maaten and Hinton, 2008), to highlight what images hotels generally use as their thumbnail image and what works well across locations. In study 2, we used a consumer-level click prediction model to highlight the importance of images in predicting individual clicks during consumer search. Drawing from methods in the field of learning to rank, we used the LambdaMART model (Wu et al., 2010; Burges, 2010) to predict clicks and understand hotel relevance to consumers. We established that on average the click prediction accuracy using image features is 10% higher than not using the image features. In addition, we used the change in relative feature importance to show that certain features, such as price, become significantly less important for predicting clicks, while others, free cancellation or sponsoring, become more important. Study 2 confirmed that images are indeed important and that they interact with the other hotel attributes. Finally, in study 3, we perform a neuroscience experiment to understand neural responses to the hotel images in our data. The fMRI results showed that consumers respond differently to high CTR vs. low CTR images. We observed that high CTR images activated areas in the parietal cortex more than low CTR images. These areas are involved in the deployment of attention, but also in the representation of space. Low CTR images activated regions in the early occipital and temporal lobe, involved in visual perception and object identification. These same areas have also been shown to correlate well with CNNs and therefore provide us with (some) causal explanation as to why we are able to predict hotel performance online using just the hotel images and CNNs and why images are important predictors of clicks.

4.8.1 Theoretical Implications

In line with expectations and previous research in different settings, we have established the importance of imagery during consideration set formation in e-commerce. We have extended previous theory about product / service search to help understand the interplay between images and other attributes of a product. Prior investigations have shown that visual stimuli and imagery impact decisions in various ways, ranging from engagement on social media (Li and Xie, 2017; Rietveld et al., 2020) to capturing attention (Pieters and Wedel, 2004; Pieters, Wedel, and Batra, 2010; Wang, Tsai, and Tang, 2018). Additionally, visual stimuli influence consumer decisions through the neural responses they elicit (Stoll, Baecke, and Kenning, 2008; Knutson et al., 2007; Couwenberg et al., 2017). Lastly, specific to the field of online travel, we know that image quality impacts the online demand (Zhang et al., 2017; Zhang et al., 2019) and several qualitative studies (Noone and Robson, 2016; Kirillova and Chan, 2018) and eye-tracking studies (Pan, Zhang, and Law, 2013; Pan and Zhang, 2016) have established that images are a much sought-after attribute by online travelers. Our research connects knowledge from various fields and highlights the importance of product images in an empirical setting as well as a neuroscientific experiment. Images are important predictors of consumer decisions.

Our work contributes to the need for the investigation of the impact of imagery online from both a (visual) marketing (Blanco, Sarasa, and Sanclemente, 2010; Ordenes and Zhang, 2019; Hauser, 2014; Kirillova and Chan, 2018; Liu, Dzyabura, and Mizik, 2020) and neuromarketing (Reimann et al., 2010; Jai, O'Boyle, and Fang, 2014) perspective. The fMRI experiment showed that high CTR images activate different brain regions as compared to low CTR images, which means they impact quick-processing and consideration, but they are not directly related to decision-making or purchase behavior. Previous research has shown that the visual processing system in the brain is closely related to layers in CNNs (Güçlü and Gerven, 2015; Eickenberg et al., 2017). The results of our experiment demonstrate that there is a link between the visual system and the CTR of hotels, which confirms why we are able to make accurate click predictions using the output of a CNN. This is an important contribution, because it means that, to a certain extent, we can use computer vision methods, and CNNs in particular, to mimic neural responses to visual stimuli for marketing research (Kamitani and Tong, 2005). Using this knowledge we connect our prediction models to consumer search as well.

Our findings advance the understanding of consumer search. As both studies 1 and 2 show, there is an interaction between visual and textual information that is presented to consumers on e-commerce websites. Consumers use heuristics to narrow down large lists of information (Glöckner and Betsch, 2008; Hauser, 2014) during the first stage of consumer search (Moe, 2006). The search results on the website of an OTA consist of textual and visual information, and images can make up a large part of the listing.[13] In our comparison of prediction with and without image features we observe a shift in feature importance of certain hotel attributes. When we factor in the results of the fMRI experiment and the relationship between the visual system of the brain and our visual analytics framework, all three studies point in the direction that qualities can be inferred from images that are also (in part) represented by textual features and that the consumer uses both types of information to make decisions. This highlights that images are not just stand-alone attributes that are used for consideration set heuristics (Park, Yin, and Son, 2019), but they are also a substitutes/complements for other attributes (Pan, Zhang, and Law, 2013). This is both an important implication and an area for future research. When modeling consumer search and consideration set heuristics, relying only on textual information that is presented overlooks important information in the visual component of a product listing, as well as the interaction between the textual and visual components. Focusing only on text could also lead to overestimation of the impact of attributes that are represented by images as well.

4.8.2 Methodological Implications

We demonstrate how to leverage unstructured image-based information. We show that methods from deep learning and learning to rank allow marketers to understand online consumers and their interaction with information (Zhang et al., 2017; Yoganarasimhan, 2020; Dzyabura, El Kihal, and Ibragimov, 2018). This opens up a wide range of possibilities for marketers to study the impact of (visual) stimuli online. CNNs, which extract information from images, and a LambdaMART ranking model or other boosted regression tree approach can be easily implemented using open-source software and packages in widely used programming languages such as Python and R. We have shown how marketers and researchers can utilize and combine these methods to approach interesting marketing problems. Additionally, the results highlight how the importance of images and what aspects to display are highly location-dependent. Our method is easily adapted to the different locations and can be extended to other contexts as well. However, the results also show that researchers need to be careful about the curation of data and generalizability of results when working with a single data source. Most importantly, our framework is applicable to *any* setting in which a consumer is presented with visual and textual information.

[13] In fact, between the time we collected the data and now, the image on the search result page has taken up increasingly more space in comparison to the textual information on the website of the OTA that we studied

4.8.3 Managerial Implications

We avoid the problem that most deep learning approaches face where they can not provide insight into the explanation for their ("black-box") prediction, by using tools such as LambdaMART and t-SNE to provide a managerially relevant argument as to why certain images do better than others. Managers could leverage the LambdaMART model scores for a particular search result and calculate the scores for every available image, which can then be used to select the most relevant image to present to the consumer. Li et al. (2019) show that, with a similar methodology, optimizing the photo lay out leads to an increase in demand and annual revenue for a property on Airbnb. Our approach allows managers the ability to easily examine millions of images and determine the role that these images are playing in online transactions.

Lastly, our study highlights that images play an important role, so managers should consider competing on images, like they compete on pricing or other features. Designing better images can make your hotel stand out and put you ahead of the competitors. Consumers are influenced by the product presentation, such that it can compensate for other attributes that have shortcomings (Kirillova and Chan, 2018). Managers should be aware what aspects of their hotel make them unique and portray these well in their thumbnail image. After all, this is their "champion" image, so it should be their best image. This depends on the hotel and the location, but from our results a suggestion could be that for a very stylistic hotel in NYC a picture of the front of the hotel is chosen as the thumbnail, whereas a hotel in Miami has a picture of their pool and bar instead. A future research avenue in this space would also be to make the shown thumbnail image conditional upon the incoming consumer.

4.8.4 Limitations and Future Research

Though we have done the best to examine the relationship between images and the decision by a consumer to include them in their consideration set, there still exist several limitations to our work. First of all, we have not explored the decision to actually make a purchase. It might very well be the case that the features that drive a consumer to consider a particular hotel are different from those visual features that result in an actual booking. However, even if that is true, the hotel must first be included in the consideration set before a purchase can be made, so this study complements any work that examines actual purchase behavior. Moreover, we have carried out a predictive study that shows that our model can do a good job at predicting which features result in a click, but our model is not a causal model, and so does not necessarily show that their is a causal relationship between these concepts and the decision to click by a consumer. However, by complementing our prediction methods with an fMRI experiment and visualization tools, we have developed new hypotheses as to why particular images do well in maximizing CTR and how images relate to other hotel attributes.

Based on our findings and limitations we identify a number of future research directions. First, as mentioned briefly in the theoretical implications section, we view investigation of the dynamics between textual and visual features online as an important future research area. Textual content has been investigated extensively in the past decade (Berger et al., 2020), and more recently new methods have allowed researchers to do similar investigations into the impact of imagery (Zhang et al., 2017; Zhang and Luo, 2018; Dzyabura, El Kihal, and Ibragimov, 2018). However we have not seen research in marketing that specifically focuses on the dynamics between textual content and imagery. Our results highlight that there is a shift in feature importance after including image features into the model, but we do not map the interactions between them, nor do we provide an exact explanation of their relationship in this work. Second, we have shown that the preference for images depends on location, but one could imagine that the impact of images is context dependent in general. The LambdaMART method models the context in the way that it calculates relevance scores for hotels listed on the search result page, based on hotel attributes and consumer (search) attributes, and it allows for interaction between them. However, we do not provide a causal explanation. In future work, it would be useful to look into what aspects of images or what types of images work well for specific search results, taking other hotel attributes into consideration. For example, it might make sense to show different images to a family of four looking for a hotel in Miami (e.g., a pool) than a business traveler looking for hotel for just themselves in Boston (e.g., the fitness center), but this relationship between the search query and the image needs to be investigated. Our t-SNE visualization hints at some of these contexts. In addition, for a specific hotel the content manager might consider posting an image of the hotels' most prized or unique amenity, e.g., a rotating restaurant, to set itself apart from other hotels on the search result page, rather than posting the standard front of the hotel image that is most standard right now. An investigation into the context could help create a method that can present the right image in the right context, preferably with clear managerial explanations.

4.9 Appendix

4.9.1 Lit Review

Study	**Models & Methods**	**Goal**
Zhang et al. 2019	VGG16 and Dynamic Game Model	Quantifying the impact of image quality on (long-term) demand on Airbnb
Zhang et al. 2017	ResNet50 and Diff-in-Diff	Estimating the impact of image quality on (short-term) demand on Airbnb
Li et al. 2019	ResNet50 and PCM	Estimating the impact of image layout on demand on Airbnb
Liu et al. 2020	VGG16, SVM and Multidim. Scaling	Understanding how brands are portrayed by users on social media
Klostermann et al. 2018	Google Vision API and clustering	Understanding user generated imagery and brands on social media
Wulf et al. 2019	ResNet50 and PCA	Improving car design by using user-generated images on online product forums
Zhang et al. 2018	Clarifai API and XGBoost	Predicting restaurant survival using user generated images on yelp
Hartmann et al. 2019	VGG16 and GradCAM	Measuring the impact of brand selfies on twitter.
Dzyabura et al. 2018	VGG16 and GBRT	Predicting product success prior to launch
Nanne et al. 2020	YoloV2, Google Vision, Clarifai and t-SNE	Generating marketing insights from images through computer vision
Rietveld et al. 2020	MVSO and Google Vision	Explaining consumer engagement with posts on Instagram.
This paper	Hybrid VGG16, SVR, t-SNE and LambdaMART	Understanding the role of images during consideration set formation.

Table 4.7: Summary of Image Mining Studies in Marketing.

Study	Method	Setting	Main Finding
Zhang et al. 2019	Image mining	Airbnb	High-quality images on Airbnb might lead to unrealistic expectations which hurts the future demand for a property.
Zhang et al. 2017	Image mining	Airbnb	The use of professional quality photos increases the demand for Airbnb properties.
Sparks et al. 2014	Survey	Advertising	Consumers prefer (and recall) natural over constructed tourism activity images.
Kirillova et al. 2018	Survey	Advertising	Hotels that are represented online in a more aesthetically pleasing manner are perceived to provide better service.
Wang et al. 2018	Eye-tracking	Advertising	Images of beautiful nature or images of performing arts are more effective in eliciting consumer attention than constructed images.
Pan et al. 2013	Eye-tracking	OTA	Hotel images improve the process and experience of decision-making and makes hotels considerable even if they were not to be considered based on other criteria.
Baek et al. 2017	Survey	Hotel Website	The design displayed by images of the physical environment of a hotel, arouse consumer emotions and shape the expected quality.
Pan et al. 2016	Eye-tracking	OTA	Hotel images reduces the cognitive load during the information processing of consumers searching for hotels online.
Jeong et al. 2005	Content Analysis	OTA	Using images of the hotel, the service and its personnel a hotel website can improve consumer attitude towards the hotel.
Noone et al. 2016	Eye-tracking + Interview	OTA	Hotel images affect the consumer choices in both the browsing phase and the deliberation phase of the consumer decision making process.
Li et al. 2019	Image mining	Airbnb	The cover photo (thumbnail image) has a bigger impact than the other photos and choosing a high quality photo of the bedroom as a cover photo leads to the biggest demand.

Table 4.8: Summary of Research on Imagery in Online Travel.

4.9.2 LambdaMART

Before we can describe the LambdaMart algorithm we give a brief overview of the evolution from RankNet, a precursor to LambdaRank (Burges et al., 2005), to LambdaMART following the overview provided by (Burges, 2010). This helps understanding the mechanisms of our ranking method and it will help in explaining its relevance for click prediction and the relative improvement of the model when we include image features.

We first describe RankNet and transition in to LambdaRank and end with the LambdaMart model specification. The main goal for RankNet, and the other learning to rank models described in this section, is to optimize the ranking of a list of documents by modelling pairwise preferences. The cost function, specified below, is minimized by minimizing the number of times two documents in a list have to be swapped in order to get to the optimal ranking. In the case of a hotel search result page, the goal of RankNet would be to estimate a score function that scores hotels based on relevance to the consumer and the weights of this function are updated with stochastic gradient descent based off of the cost function. The observed clicks are used as relevance indicators, and the cost function is minimized when the hotel that has the highest relevance score is the hotel that is clicked on by the consumer. Any time this isn't the case some pairs of hotels need to be swapped to reach the optimal ranking. In LambdaRank the updating of the weights based on the pairwise preferences is then done for all hotels in a list at the same time, using lambdas, because we don't need to specify the costs, we just need to specify the gradient with respect to the cost. LambdaMart then combines this method with multiple additive regression trees, which provides with easy feature importance measures.

RankNet

The training data is partitioned by hotel search and at any given point during the training the input feature vector $x \in \mathbf{R}^F$ is mapped to a score f(x). For a given hotel search query, each pair of hotels H_i and H_j with feature vectors x_i and x_j have calculated scores $s_i = f(x_i)$ and $s_j = f(x_j)$, where w is a set of parameters that define function f. When H_i is preferred over H_j (for example when H_i is clicked and H_j is not), then the modeled probability through a sigmoid function for $H_i \triangleright H_j$ is given by:

$$P_{ij} = P(H_i \triangleright H_j) = \frac{1}{1 + e^{-\sigma(s_i - s_j)}} \tag{4.12}$$

with this probability we can define the log-likelihood function or cost entropy by taking the deviations of the modeled output probabilities from the desired probabilities by.

$$C = -\bar{P_{ij}} log P_{ij} - (1 - \bar{P_{ij}}) log(1 - P_{ij}) \tag{4.13}$$

For a given search query, let $S_{ij} \in \{0, \pm 1\}$; such that S_{ij} equals 1 if H_i is preferred over H_j, -1 for the opposite, and 0 for equal preference. With the assumption that the desired ranking is deterministically known such that $\bar{P_{ij}} = \frac{1}{2}(1 + S_{ij})$, which gives

$$C = \frac{1}{2}(1 - S_{ij})\sigma(s_i - s_j) + log(1 + e^{-\sigma(s_i - s_j)}) \tag{4.14}$$

the cost entropy is symmetric, and therefore $C = log(1 + e^{-\sigma(s_i - s_j)})$ for $S_{ij} = 1$, $C = log(1 + e^{-\sigma(s_i - s_j)})$ for $S_{ij} = -1$ and $C = log2$ when $S_{ij} = 0$. The gradient of this cross-entropy is then

$$\frac{\partial C}{\partial s_i} = -\frac{\partial C}{\partial s_j} = \sigma(\frac{1}{2}(1 - S_{ij}) - \frac{1}{1 + e^{-\sigma(s_j - s_i)}}) \tag{4.15}$$

Using stochastic gradient descent, with learning parameter η, the gradient is used to update the weights w of model f by:

$$w \rightarrow w - \eta \frac{\partial C}{\partial w} = w - \eta(\frac{\partial C}{\partial s_i}\frac{\partial s_i}{\partial w} + \frac{\partial C}{\partial s_j}\frac{\partial s_j}{\partial w})) \tag{4.16}$$

By specifying the gradient as in Equation 4.16 we can introduce the "Lambda" in this model. Combining Equations 4.15 and 4.16 we can define $\frac{\partial C}{\partial w} = \lambda_{ij}(\frac{\partial s_i}{\partial w} - \frac{\partial s_j}{\partial w})$ where

$$\lambda_{ij} = \frac{\partial C(s_i - s_j)}{\partial s_i} \tag{4.17}$$

This specification has two main advantages: 1) Instead of working directly with the derivative of the cost function with respect to the weights of the model, we can instead use the derivative of the cost with respect to the model scores, and we don't need the costs themselves. This is similar to gradient formulation used by boosted regression trees such as MART (Friedman, 2001), which turned out useful when combining the ranker with boosted regression trees. 2) Since we do not need to explicitly specify the cost function we can use more sophisticated loss functions and bypass optimization difficulties related to loss functions that are discrete or non-continuous (Burges, Ragno, and Le, 2007).

The gradient lambdas are essentially forces that move hotels up or down predicted rankings. If hotel i is more relevant than hotel j then hotel i will receive a push of λ_{ij} up the list and hotel j a push of the same magnitude in the same ranking. The weights of the model are then updated based on these forces. In our setting, each time a hotel is clicked on in a list, this hotel will receive a push up the ranking of λ_{ij} relative to every other hotel that ranked as preferred over hotel i and all these hotels will receive one push downward of the same magnitude. The weights of the function f will then be updated with $\delta w = -\eta \sum_i \lambda_i \frac{\partial s_i}{\partial w}$. Where λ_i is the sum of all positive and negative forces for hotel i.

For instance, in a search result list of 25 hotels, 1 hotel is clicked on. Let this hotel be hotel i, then the lambda for this hotel is the sum of all λ_{ij}'s based on scores s_i and s_j and the lambda for all other hotels equals -λ_{ij}. Note that all other documents are un-clicked, so the lambda of these pairs equal zero, because they are considered equally (ir)relevant and the model "ignores" these.

LambdaRank

In the learning-to-rank literature the most commonly used loss function is the Normalized Discounted Cumulative Gain (NDCG), which is also used for the personalized search ranking mechanism by Yoganarasimhan (2020)[14]. The LambdaRank model is based on the idea that we only need the gradient of the costs with respect to the model scores and not the costs themselves Burges, Ragno, and Le, 2007. Experiments have shown that instead we can modify Equation 4.17 by multiplying by the change of NDCG (or any ranking metric that we are interested in optimizing) that is gained from swapping hotel i and hotel j in the ranking. This is useful, because we can directly use the model scores that we are interested in optimizing. The lambda's can be defined by:

$$\lambda_{ij} = \frac{\partial C(s_i - s_j)}{\partial s_i} = \frac{-\sigma |\Delta NDCG_{ij}|}{1 + e^{\sigma(s_i - s_j)}} \tag{4.18}$$

Here we are maximizing the NDCG (Burges, 2010), hence we replace the gradient descent in Equation 4.16 by:

$$w \rightarrow w + \eta \frac{\partial C}{\partial w} \tag{4.19}$$

LambdaMART

The specification of the gradients of LambdaRank, are the same as the gradients of the cost with respect to the model scores in MART (Friedman, 2001). MART improves its predictions by training a new tree to predict the errors (using the gradients of the loss function) of the trees which came before. MART models derivatives and LambdaRank works by specifying the derivatives at any point during the training. LambdaMART is the result of combining these two algorithms. As suggested by Yoganarasimhan (2020), we implement LambdaMART using a Gradient Boosting Machine that utilizes newton boosting steps. The newton step [15] is used to determine the optimal step size

[14] see section 5.2 in their paper for a more detailed explanation of this metric

[15] for a detailed explanation of the newton step to calculate the optimal step size, see (Friedman, 2001) or (Burges, 2010)

that minimizes the loss of the regression tree, using least squares to compute the splits of each node. In LambdaMART each tree computes the lambda per hotel across the entire dataset. Recall that λ_i is the sum of all positive forces minus the sum of all negative forces, constituted by each time hotel i is clicked on for the former and each time hotel i is not clicked on for the latter, compared to the predicted ranking of the other hotels such that we can define:

$$\lambda_i = \sum_{j:\{i,j\}\in I} \lambda_{ij} - \sum_{i:\{j,i\}\in I} \lambda_{ij} \tag{4.20}$$

Where I contains a set of pair indices for which hotel i was clicked on and all hotels j were displayed on the search result page. Therefore the first summation represents every time hotel i was clicked on and all hotels j were displayed, whereas the second summation represents every time hotel i was displayed, while another hotel j was clicked. Burges (2010) simplify this notation by

$$\sum_{\{i,j\}\rightleftharpoons I} \lambda_{ij} = \sum_{j:\{i,j\}\in I} \lambda_{ij} - \sum_{i:\{j,i\}\in I} \lambda_{ij} \tag{4.21}$$

For any given state of the model and a given hotel i we can write down a utility for which λ_i is the derivative of that utility [16].

$$C = \sum_{\{i,j\}\rightleftharpoons I} |\Delta NDCG_{ij}| log(1 - e^{\sigma(s_i - s_j)}) \tag{4.22}$$

such that,

$$\frac{\partial C}{\partial s_i} = \sum_{\{i,j\}\rightleftharpoons I} \frac{-\sigma|\Delta NDCG_{ij}|}{1 + e^{\sigma(s_i - s_j)}} = \sum_{\{i,j\}\rightleftharpoons I} -\sigma|\Delta NDCG_{ij}|\rho_{ij} \tag{4.23}$$

where $\rho_{ij} = \frac{-\lambda_{ij}}{\sigma|NDCG_{ij}|}$ and

$$\frac{\partial^2 C}{\partial s_i^2} = \sum_{\{i,j\}\rightleftharpoons I} \frac{-\sigma|\Delta NDCG_{ij}|}{1 + e^{\sigma(s_i - s_j)}} = \sum_{\{i,j\}\rightleftharpoons I} \sigma^2|\Delta NDCG_{ij}|\rho_{ij}(1 - \rho_{ij}) \tag{4.24}$$

Therefore the newton step for the k^{th} leaf of the m^{th} tree is given by:

$$\gamma_{km} = \frac{\sum_{x_i \in R_{km}} \frac{\partial C}{\partial s_i}}{\sum_{x_i \in R_{km}} \frac{\partial^2 C}{\partial s_i^2}} = \frac{\sum_{x_i \in R_{km}} \sum_{\{i,j\}\rightleftharpoons I} |\Delta NDCG_{ij}|\rho_{ij}}{\sum_{x_i \in R_{km}} \sum_{\{i,j\}\rightleftharpoons I} |\Delta NDCG_{ij}|\sigma\rho_{ij}(1 - \rho_{ij})} \tag{4.25}$$

$x_i \in R_{km}$ represents the data points that land in the k^{th} leafnode of the m^{th} tree.

[16] We are maximizing the NDCG metric, so we have a utility instead of a cost

4.9.3 fMRI Setup

Participants

We recruited and measured 22 participants, that had booked a hotel via a hotel booking site (Booking.com, Expedia) in the last 2 years). Two subjects where excluded from analysis for excessive movement. We selected a heterogeneous audience for this experiment: our group of participants varied in gender (9 male, 11 female), age (range: 18–44 years; M = 28) and in educational background (highest qualification: lower vocational education and high school = 40%, higher vocational education = 40%, university (graduate level) = 20%).

Stimuli

To investigate the neural response to hotel images, we selected a set of 102 images that we divided into 6 categories. The first distinction is dependent on the hotels CTR and we separated the hotels into high CTR and low CTR. To do this we selected images from the 75-th percentile and up as high CTR and 25-th percentile and down as low CTR images, to ensure a clear distinction between the two categories. From there, we divided each set of images into 3 categories based on the visual complexity characteristics, Contrast Energy and Spatial Coherence, described above. These two measures are already known to elicit neural responses (Groen et al., 2013; Scholte et al., 2009), so we divided the stimuli into high, medium and low groups of combined CE and SC to effectively investigate the impact of the images for high CTR vs. low CTR. In the end, there were 17 images for each group of high/low CTR and high/med/low visual complexity. To familiarize subjects with stimulus presentation, 30 practice images (never used in the main experiment) were presented without feedback before the first session.

Experimental Procedure and MRI recording

All 102 images were shown, in each of 3 runs that where presented while recording BOLD-MRI using a Philips Achieva XT 3T scanner with a 32-channel head-coil located at the University of Amsterdam, The Netherlands. In each run 744 volumes (T2*-weighted GE-EPI, multi-band recordings (TR = 1000 ms, TE = 27.6 ms, FA = 60°, SENSE = 1.5, MB = 4, FOV = 216 mm2, matrix size = 80x80, 45 slices, slice thickness 2.7 mm, slice gap = 0.27 mm) were recorded. At the start of the session a 3D-T1 weighted scan (TE = 3.8 ms, TR = 8.2 ms, FA = 8°, FOV = 2562, matrix size = 2562, 160 sagittal slices, slice thickness = 1mm) was acquired which was used to register the functional volumes of each run to the structural brain, after which they were registered to standard MNI (Montreal Neurological Institute) space. Each of the 102 images was shown once in each of the 3 runs, repeated once. The stimuli were presented on a 32 inch BoldScreen (Cambridge Research Systems) that was viewed through a mirror attached to the head coil, using Presentation software (Neurobehavioral Systems, Albany, CA, USA). The participants were asked agree/disagree questions to keep them engaged and to match results of high- and low CTR with image preferences. Responses were collected with two button boxes, one for each index finger, response mappings were counterbalanced across subjects. There were 6 questions that appeared randomly: 1) This is a pretty hotel, 2) I would recommend this hotel to others, 3) I would book a room in this hotel, 4) This is an ugly hotel, 5) I would not recommend this hotel to others, 6) I would never book a room in this hotel.

Chapter 5

In the Eye of the Reviewer: An Application of Unsupervised Clustering to User Generated Imagery in Online Reviews

Authors: Gijs Overgoor, Rohan Mestri, William Rand

Finalist for the best paper award at the *Hawaiian International Conference on System Sciences (HICSS) 2021*. Gijs Overgoor was the leading author for this study. Rohan Mestri was involved with the method development. William Rand fulfilled a supervisory role for the paper.

5.1 Abstract

Mining opinions from online reviews has been shown to be extremely valuable in the past decades. There has been a surge of research focused on understanding consumer brand perceptions from the textual content of online reviews using text mining methods. With the increase in smartphone usage and ease of posting images, these reviews now often contain visual content. We propose an unsupervised cluster method to understand the user-generated imagery (UGI) of online reviews in the travel industry. Using the deep embedded clustering model we group together similar UGI and examine the average review ratings of these clusters to identify imagery associated with positive and negative reviews. After training the method on the entire dataset, we map out individual hotels and their corresponding UGI to show how hotel managers can use the method to understand their performance in particular areas of customer service based on UGI. The performance in a cluster relative to the population can be a clear indicator of areas that need improvement or areas that should be highlighted in the hotel's marketing efforts. Overall, we present a useful application using visual analytics for mining consumer opinions and perceptions directly from image data.

5.2 Introduction

72% of consumers always or frequently read reviews before deciding where to visit, eat or stay, and on average those users read nine reviews before making a decision to book a hotel or a restaurant.[1] Online reviews have become an integral part of the online travel industry. Consumers search out reviews when making decisions and these reviews have an important impact on a consumer's decision to book a hotel or to visit a restaurant (Park, Yin, and Son, 2019). In turn, for businesses they are a valuable source of feedback to understand customer experiences and perceptions (Chakraborty, Kim, and Sudhir, 2019). Businesses can improve their services and facilities based on these insights (Wang, Chaudhry, and Pazgal, 2019). Clearly, it's crucial for a business to understand what the consumer is saying about their business. Generally, there are 3 types of information presented to a consumer in the form of reviews. Review score (numerical), review text (textual) and review imagery (visual). The numerical and textual parts of the review have been covered extensively in the past, but the visual component of reviews has not received sufficient attention.

Smartphones and the ease of capturing and sharing photos online have led to a

[1] https://tripadvisor.mediaroom.com/2019-07-16-Online-Reviews
-Remain-a-Trusted-Source-of-Information-When-Booking-Trips-
Reveals-New-Research

great increase of UGI online. More often reviews are now accompanied with UGI. For example, there are over 160M photos generated by travelers on TripAdvisor.[2] Given the evidence that images are more engaging and hold more information than text (Li and Xie, 2017), one can imagine that the UGI in online reviews holds valuable information for businesses. For instance, Zhang et al. 2018 show that UGI on Yelp provides information about a restaurant's survival potential. Ma et al. 2018 are one of the very few studies that investigate UGI in the online review context. They indicate that the main reason it is an understudied area is likely due to technical difficulties in translating the images into structured information.

The extraction of information from images, or image mining, has been shown to be useful in recent research. There are several marketing studies that now utilize these methods to connect imagery to interesting marketing problems (Zhang and Luo, 2018; Zhang et al., 2017; Liu, Dzyabura, and Mizik, 2018; Overgoor, Rand, and Van Dolen, 2020). As for online reviews, Ma et al 2018 establish that UGI contains useful information for predicting the helpfulness of reviews. Despite these developments, there is no research that has shown to be effective in opinion mining from UGI in online reviews.

We propose an unsupervised clustering method, based on the deep embedded clustering method (Xie, Girshick, and Farhadi, 2016), to cluster UGI based on visual similarity. This method is completely unsupervised, which means that we do not need to teach or train our model to recognize or predict specific labels. Instead our method automatically detects UGI clusters with similar visual properties. Our method provides several directly managerially relevant outputs: (1) we identify and cluster UGI that users generally post with their reviews across different hotels, (2) by highlighting the distribution of review scores across these clusters we show what users generally post when they are satisfied vs. dissatisfied, (3) when looking at individual hotels we highlight high and low performance areas making it easy to identify places for improvement.

5.3 Background

Previous research has established the impact of online reviews on the online travel industry using a variety of methods (Berger et al., 2020). On comparison websites such as TripAdvisor and Yelp the review information is generally presented to consumers by the average score review score, scores on aspects such as cleanliness and location and by individual consumer reviews (Goh, Heng, and Lin, 2013). The average scores

[2]http://ir.tripadvisor.com/static-files/6d4c71fd-3310-48c4-b4c5-d5ec04e69d5d

can easily be used in statistical models to understand their impact on demand or consumer choices, but the individual reviews consist of unstructured data such as text and images that need processing techniques to transform into useful information (Ordenes and Zhang, 2019). A survey study about information sources and their importance for online hotel bookings showed that in terms of the types of information, consumer review scores are perceived as the most important source, followed by hotel images and descriptive (textual) information about hotels (Park, Yin, and Son, 2019).

On the most popular platforms there is an abundance of reviews available and it is impossible for consumers or even businesses to process all this information manually. As an example, in 2014 there were on average 165,000 new reviews each day on Tripadvisor alone.[3] As mentioned earlier, a user only reads nine reviews on average, which means that users rely on review summaries and platform filtering to examine the information. For these reasons, we rely on machine learning methods to examine this unstructured information. The textual component of reviews has received quite some attention in the past decade. Using NLP, previous work has examined the sentiment of text and the over-arching topic of a review and then used that information to make predictions about the review score or the review helpfulness (Tsai et al., 2020).

Text mining methods have been used in a variety of applications, either to understand the person or group of people generating and/or receiving textual content (Berger et al., 2020). These methods are very helpful to generate marketing insights from textual content to understand consumers and their perception of brands. Several (recent) studies have investigated the use of text mining to summarize a large number of reviews online, mostly with the intent of providing consumers searching for information with the most helpful content. Applications of text mining to reviews range from exploring customer satisfaction (Berezina et al., 2016), identifying the most informative reviews and sentences on TripAdvisor (Hu, Chen, and Chou, 2017) to mining consumer opinions and sentiments (Abdi et al., 2018). For example, Tsai et al. 2020 first classify reviews as helpful vs. non-helpful and then highlight the hotel features in the helpful reviews. Based on these features and their helpfulness scores platforms could enhance search functions to allow consumers to filter or group based on what they are interested in. In general, research using text mining methods has highlighted how to extract information from reviews at scale and how this information can be used to improve platforms or identify performance issues or highlights for businesses. A major limitation is that these studies look only at the textual and/or numerical component of reviews and overlook the visual component that is often a part of consumers' reviews. Although, the textual component plays a prominent role in the review, in recent years the image has become increasingly helpful in reviews

[3] https://www.telegraph.co.uk/travel/lists/TripAdvisor-in-numbers/

(Ma et al., 2018).

Images are now increasingly available, because of smartphones and online platforms that make it easy to upload an image along side text. People process visual information more easily than text (Morin, 2011) and we have seen ample evidence that content online that has a visual component is much more engaging (Li and Xie, 2017). For these reasons, it is crucial to understand what is presented in these images. With the refinement of image mining methods in the past decade, we are becoming increasingly capable of extracting useful information from imagery online. Images provide managers with a way to visually listen, i.e., better understand, to what consumers are posting about brands on social media (Liu, Dzyabura, and Mizik, 2018). Studies in the online travel industry also established the impact of hotel or AirBnB property images on consideration set formation and demand using image mining methods (Zhang et al., 2017; Zhang et al., 2019; Overgoor, Rand, and Van Dolen, 2020). As for online reviews, Ma et al 2018, propose a deep learning approach to predict the helpfulness of reviews and establish that UGI provides additional information about customer experiences and improves helpfulness prediction. There are currently no studies that offer a summary or overview of what the imagery in online reviews represent.

Most of the online review studies mentioned previously focus on providing a summary of useful information that is presented in text-based reviews at a large scale. The main reason behind this focus on summaries is that there is simply too much information for consumers to process (Tsai et al., 2020). Another reason for the summarization is that it can detect underlying opinions in these reviews (Hu, Chen, and Chou, 2017). As a result, the methods turn out to be valuable to both the consumer and the business (Chakraborty, Kim, and Sudhir, 2019). Images could play a similar role. A problem, however, is that images are very rich sources of information. A standard User-Generated Image on TripAdvisor has 224 by 224 pixels, or about 50k pixels each consisting of a Red, Green, and Blue channel. As a result, it is difficult to summarize what is portrayed by an image, let alone a set of images. For this reason, we turn to a very common method used for grouping data based on similarities: Clustering.

Clustering is a method of unsupervised learning (i.e., we do not need to give the machine feedback on what it is learning), which is a popular technique in machine learning to get a better understanding of our data. When we apply a clustering algorithm we group data points into groups and these groups can provide us with higher-level information about what the data looks like. For example, k-means clustering (MacQueen et al., 1967), can be used for customer segmentation based on purchase history, interests and demographics. In the online travel industry, Ahani et al. 2019, show that by clustering consumers based on textual reviews and ratings they can understand why different travelers select certain spa hotels and how these hotels can use this information to better service their (potential) customers. The goal of our research is to understand consumers through the images they post with their online reviews.

Images are a very high-dimensional information source that needs additional processing. First, we need to translate the pixels of an image using a Convolutional Neural Network (CNN) architecture (LeCun and Bengio, 1995). A CNN uses convolutions to examine pictures through "filters". These filters are responsible for detecting certain patterns, where early layers detect simple image information (e.g., colors, lines, or edges) and later layers detect more complex image information (e.g., complex shapes, buildings, faces, etc.). Generally, the height and width dimension of the convolutional layers in an architecture such as the VGG16 (Simonyan and Zisserman, 2014) increases as the information is processed by new layers and the number of filters increases. Basically, this means that what the net is detecting gets more complex. Eventually, using other layers such as flattening and fully connected layers, an image is embedded onto a vector space. The image is now translated from unstructured image information into structured information, or features, that can be input into a model. Our method uses the Deep Embedded Clustering (Xie, Girshick, and Farhadi, 2016), that is shown to be much more effective for clustering imagery than standard clustering methods, such as k-means or hierarchical clustering. General clustering methods fail at clustering images, because of the high dimensions of the data, even if they were to be embedded on to a vector space before they are used in these methods. In the next section, we discuss the clustering method in more detail and then we highlight how we can use it to understand what consumers usually capture in their UGI and what we can do with this information.

5.4 Method

In unsupervised clustering, traditional algorithms like the k-Means algorithm generally fail on higher dimensional data and they are too computationally expensive. Deep Neural Network-based clustering methods have risen to prominence in the past few years to solve both these issues. We use the Deep Embedding Clustering model, first proposed in (Xie, Girshick, and Farhadi, 2016), to perform unsupervised clustering over high dimensional data. One of the primary advantages of this method is that, after training, we can still feed unseen samples into the model and the model maps the unseen samples to their most probable cluster. This is helpful, because we can cluster all reviews of a single hotel and then highlight the clusters associated with high or low review scores, to identify performance areas. At the same time, we can

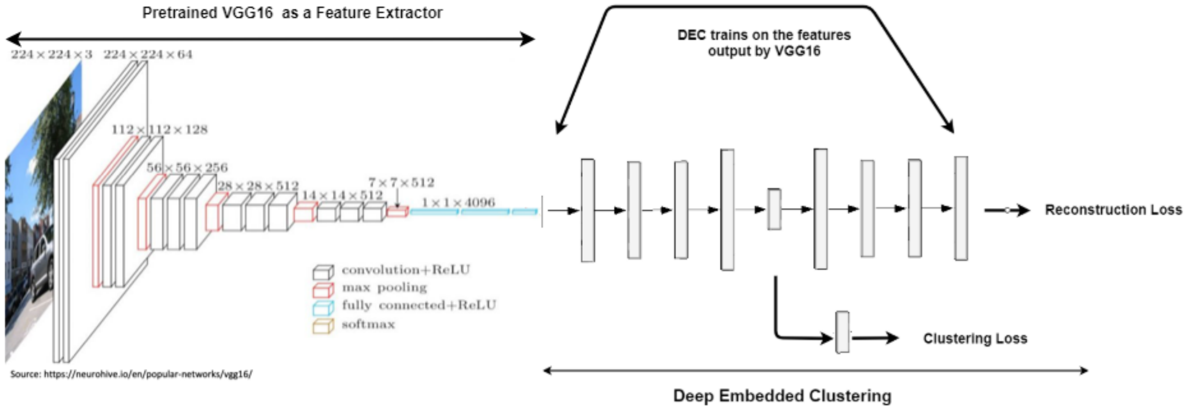


Figure 5.1: The above diagram shows the combined model of the Deep Embedded Clustering along with the VGG16 model at the preprocessing stage. The DEC model clearly shows the stacked denoising autoencoder with each of the vertical bars representing fully connected layers of proportional node size. The DEC model carries out the process of clustering these features into a discrete set. These clusters can be obtained from the latent space after detaching the decoder.

also directly label a new image as belonging to a high or low performance cluster, which is useful for managers to identify potentially effective marketing assets.

5.4.1 Deep Embedded Clustering

Stochastic Neighbor Embedding (SNE) (Hinton and Roweis, 2003) is a method proposed to reduce the dimensionality of a set of data points from a higher dimensional space to a lower dimensional space, while maintaining the neighborhood similarities in the lower dimensional space as observed in the higher dimensions. The neighborhood similarities are measured by a Gaussian Similarity Kernel function, both in the high dimensional space as well as in the lower dimensional space, and the difference between these similarities is minimized using the Kullback−Leibler (KL) divergence metric. In order to solve the crowding problem (i.e. limited space for "neighbors" in high-dimensional data when forced onto a 2-dimensional plane) observed in the SNE method, van der Maaten et al. 2008 proposed to use the t-distribution based similarity kernel in the lower dimensional space and thus, formed the t-Distribution Stochastic Neighbor Embedding (tSNE).

Motivated by the tSNE, the Deep Embedded Clustering (DEC) model was developed using an autoencoder framework for dimensionality reduction. A Stacked Denoising Autoencoder learns the distribution of the input data, by training it to recreate the input data itself. The input X is fed to the encoder part of the autoencoder, which tries to "embed" the higher dimensional data X into the lower dimensional data z, by shrinking the data to pass through a bottleneck. The z-points in the latent space are then passed through the decoder, which then tries to reconstruct the initial input. This autoencoding is useful, because it forces the model to preserve as much information in the dimension reduction, otherwise the decoder wouldn't be able to reconstruct the input. This makes the latent space as informative of the imagery as possible, which makes the cluster assignment more effective as a result.

After training the autoencoder, we obtain a latent space from which we can obtain reasonable estimates of the initial cluster centres in the data distribution. The k-means algorithm then processes the latent space to find the initial cluster centers of the N_c clusters. Once we have these cluster centers, we find the t-distribution based similarity from each embedded datapoint z to these N_c cluster centres. These similarities are computed in a probabilistic manner, indicating the probability that the datapoint z_i will lie in a cluster with mean μ_j (with degrees of freedom df and k iterating over each cluster) and is given as:

$$q_{ij} = \frac{(1+\frac{||z_i-\mu_j||^2}{df})^{\frac{-df+1}{2}}}{\sum_k (1+\frac{||z_i-\mu_k||)^2}{df})^{\frac{-df+1}{2}}}$$

In the finetuning process, this probability distribution is then self-trained to follow a target distribution, and the difference of these distributions is minimized using the KL Divergence metric. The target distribution is chosen in a way that it "sharpens" the probabilities of membership into a particular cluster, thus refining these clusters close to a single convergence point. This target distribution is given as:

$$p_{ij} = \frac{\frac{q_{ij}^2}{\sum_i q_{ij}}}{\sum_j \frac{q_{ij}^2}{\sum_i q_{ij}}}$$

Guo et al. (Guo et al., 2017b) proposed an improved version of the DEC which also minimizes the reconstruction loss along with the clustering loss, as a dual loss function. This is shown to maintain the local structure preservation property. Similarly, in our paper, we have used this dual loss function for training the model. The autoencoder part of the DEC model tries to reconstruct the features given by the VGG-16 model. The reconstruction loss is a mean squared error loss between the actual features and the predicted features.

5.4.2 Transfer Learning

Guo et al. 2017 explored the possibility of using convolutional layers in both the encoder as well as a decoder to reconstruct the image. This is essential as compared to using the traditional fully connected autoencoder as the latter trains on direct pixel values and hence fails to capture the features provided by the convolutional layers. Similarly, we train on features as opposed to direct pixel values, but instead of using convolutional layers in the autoencoder, we use a pretrained CNN to transform the images into features.

We are using the VGG-16 model (Kalliatakis, 2017) which is pretrained on the Places dataset with 365 scene categories (Kalliatakis, 2017). This network is effective at detecting common places, such as hotel rooms, pools, or parks among others. Previous research on the impact of imagery on online hotel bookings show that this pretrained CNN is an effective method for extracting features that can be used for click-through rate prediction (Overgoor, Rand, and Van Dolen, 2020). Each image is fed into this pretrained model first and it outputs a vector of size 365, with each of the 365 nodes consisting of a probability of belonging to one particular class. The magnitude of the remaining nodes is largely diminished due to the usage of a softmax activation function and we can not use this 365-length vector as features. Hence, we replace the last layer of the pretrained model with a sigmoid activation function, which serves as reasonable estimates of the features. These features are then fed to the encoder. The decoder is used to reconstruct these features from the latent space and the reconstruction loss is used to train the encoder so it preserves as much of the information while shrinking the dimensions.

5.5 Results

In this section we describe the application of the DEC to a set of online reviews with UGI scraped from TripAdvisor. First, we describe the data, then we show the results of the clustering method to the entire dataset to understand the distribution of UGI across clusters. We then highlight three example hotels and what we can learn from the clustering of the UGI. And finally, we discuss how we can use the method to identify useful marketing assets.

5.5.1 Data

We test our method on a collection of reviews with images from a group of New York City hotels from TripAdvisor. In total, we collected 5499 online reviews resulting in 9155 UGI. The average review rating is 4.48/5 with a standard deviation of 0.92. About 60% of the reviews have a 5 star rating. This is expected as TripAdvisor highlights that about 87% of customers write a review about a positive experience.[4] About 75% of the reviews with images have only a single image and more than 95% of these reviews have less than 5 images. The maximum number of images belonging to

[4]https://tripadvisor.mediaroom.com/2019-07-16-Online-Reviews
-Remain-a-Trusted-Source-of-Information-When-Booking-Trips-
Reveals-New-Research

a single review is 28. We do not observe a significant correlation between the number of images and the review rating.

5.5.2 Overall clustering

We start with a clustering of all images across all hotels. We feed the entire data into our clustering model without attaching any labels to these data samples. After obtaining the clusters, we associate each embedded datapoint with its numerical rating, since some reviews feature multiple images that means that every image associated with that review receives the same numerical rating.[5] Given that δ_{ij} is the binary membership of data sample i in the cluster j and τ_i is the rating given by the user, to which the sample belonged. The aggregate rating of the cluster j can be calculated as $\rho_j = \frac{\sum_{i=1}^{N} \delta_{ij}\tau_j}{\sum_{i=1}^{N} \delta_{ij}}$.

After some exploration, we have set the number of clusters and the latent space dimension to 10. The resulting clusters are both specific and substantial, which is the most important trade-off to consider here. In the near feature, we aim to do a more extensive hyperparameter tuning.

In Table 5.1, we have listed the ten (10) clusters with four (4) sample images that best represent each cluster. Note that the method is completely unsupervised, therefore it does not get any feedback on similarities between images from labels. Instead it is based solely on the imagery themselves. We observe that the images within each cluster are remarkably similar in terms of the content they depict. From the sample imagery we observe the following about the clusters:

- Cluster 1 - zoomed in images of specific details
- Cluster 2 - lobby, bar or general hotel area
- Cluster 3 - Seating areas within the hotels
- Cluster 4 - Views from the hotels
- Cluster 5 - Front or outside of the hotels
- Cluster 6 - Food and Drinks
- Cluster 7 - Hotel rooms
- Cluster 8 - Style-details
- Cluster 9 - Bathroom

[5]The results are robust to a weighting based on the number of UGI per review, where a single image receives a higher weighting than images that belong to reviews with multiple images.

- Cluster 10 - Empire State Building

This is the UGI that travelers to New York City share in their online reviews about hotels. When we look at the mean and standard deviation in particular, we observe that Cluster 1 represents UGI most related to dissatisfaction (lowest average score) with the hotels, whereas Cluster 10 represents UGI most closely related to satisfaction (highest average score). The cluster with the lowest average score (1) shows UGI portraying aspects of the hotel experience that are subpar, such as close ups of a bed, a closet or a bathtub/shower. The highest rated cluster (10), in contrast, mainly shows the Empire State Building, which could be thought of as UGI portraying the experience. We also observe that the standard deviation for Cluster 1 is largest, meaning that there is a larger spread in satisfaction across this cluster. We also observe this for the bathroom cluster (9). The other 8 clusters have fairly comparable average ratings and standard deviations, even though they portray very different UGI. In the individual level analysis we can think of these clusters as performance areas, which encapsulate more than just the physical areas, but also the features, experiences, and stylistic elements of individual hotels.

5.5.3 Individual Hotels and Distribution across clusters

After training the model with the entire set of data, we then segment the data and bin them into groups, with each group corresponding to images belonging to one particular hotel. We are able to observe the cluster distribution per hotel. We can then compare the means and standard deviation of individual hotels to those of the population. This provides indicators for high or low performance areas for individual hotels. To illustrate this application we have selected three hotels. Hotel A, whose average ratings are below average, Hotel B, whose average ratings are similar to the population average, and Hotel C, whose average ratings are above average. Figure 5.2, shows the performance of the hotels in each of the 10 performance areas, represented by the clusters, in comparison to the population average for these performance areas. The green rectangles indicate where the hotels over perform (i.e., a hotel average rating that is significantly higher than the average of the other hotels) and the red box indicates where the hotel under performs (i.e., a hotel average rating that is significantly lower than the average of all hotels). We observe that the Hotel A under performs in three clusters: Cluster 2, Cluster 8 and Cluster 10. Recall, that these clusters represent the lobby/bar areas (2), style details (8), and Empire State Building pictures (10) respectively. This hotel is generally under performing, as reflected by a below average rating, but performs especially poorly in these three areas. Although it performs poorly overall, it does seem to have little issue in Cluster 1, which means that in general it has very little problems with specific tangibles, such as cleanliness, messiness, or damage, which is what Cluster 1 typically identifies. This is also reflected by a very low standard deviation in this cluster as compared to the population. The manager of this hotel might not be able to do much about the view it has of the Empire State Building, but it could potentially address the lobby / bar issues and the style details. On the other hand, Hotel B, is an average hotel, which is reflected by most performance areas, though it does perform below average for the first cluster. A clear insight for the manager of Hotel B is to examine the images related to Cluster 1 and see what can be done to improve performance in this area. Hotel B over performs in Cluster 5, which generally represents the outside or front of the hotel. These might be architectural style indicators, which seem to be appreciated by the visitors of Hotel B, and the manager could highlight these in their marketing efforts. Finally, Hotel C is an above average hotel. We can observe that this hotel over performs consistently (also reflected by low standard deviation across clusters), and does exceptionally well in areas related to Clusters 1, 8, and 9. All these are related to the hotels interior. Hotel C performs very well overall, but especially the bathroom and style details are very well received by its patrons. A manager of this hotel could play up these details in marketing materials and on their website.

5.6 Discussion

Methods for understanding large scale unstructured data such as image mining or visual analytics are becoming increasingly important and useful. In the past decade we have seen a surge of studies on online reviews focused on opinion mining and summarization of textual content, but little research on visual content. Effectively summarizing UGI is difficult because of the high-dimensional nature of the images. In this research, we present an image mining method to understand what consumers portray in their UGI in online reviews. We leverage the information portrayed by UGI using Deep Embedded Clustering, a high-dimensional clustering method that is much more effective than traditional clustering methods, such as k-means. We apply transfer learning to a CNN originally trained to recognize 365 places to embed imagery onto a vector space and use these features to effectively cluster UGI from TripAdvisor reviews of hotels New York City. It is important to emphasize that the method is unsupervised, so it does not need any human feedback for training. The system automatically identifies the 10 clusters. The method can then stay "up-to-date" every time it is fine-tuned on new data, but it can also directly distribute new imagery across the clusters to identify marketing stimuli.

In addition to the real-time managerial insights such a system generates, it also helps us develop a deeper understanding of consumer behavior with respect to UGI. Our results clearly show 10 main types of imagery that are generally posted by users. Using the distribution of UGI and corresponding reviews across these clusters we highlight that, in general, dissatisfied customers post zoomed in pictures of tangibles in the hotel, such as style features, or furnishing, cleanliness and damages. The satisfied customer, for our New York City data, tends to share the Empire State Building. This is in line with previous research on textual reviews (Berezina et al., 2016). The standard deviations of the clusters also provide us with information on the volatility of certain areas as compared to others. In general, the clusters portray some clear performance areas for hotels, not reflected by the ratings generally offered by website such as Yelp and TripAdvisor. This highlights a clear advantage of our method, but also lends itself for an interesting study on the performance of individual hotels or competition between hotels.

The application of our method on the three individual hotels show how managers can easily identify areas in which the hotel is over or under performing. We saw, for example, that Hotel A was performing very well in Cluster 1, even though it was under performing overall. It also highlighted some important areas where it was underachieving. Using this information a hotel manager can identify areas that might need work, but it can also use UGI from high performance clusters for their marketing assets. We saw that Hotel B, an average hotel in terms of rating, was clearly under performing in a basic service area such as Cluster 1. At the same time it showed that customers generally liked the look of the hotel. As for the high performance hotel, we observed that it was performing very consistently across the board, but even then we were able to highlight some areas the hotel might want to focus on in its marketing materials.

There are a number of areas where this method can be improved. We have limited ourselves to a single city and a group of hotels within that city. Though we have examined thousands of images, the method scales well to millions of images, in which case the results would be much more robust. At the same time, we plan on comparing multiple locations. UGI would look very different for a less touristy destination than New York City. It will be interesting to observe differences across different cities and locations. We can also take the performance of individual hotels to the next level to explore hotel competition.

Another future direction that needs to be explored is the comparison of these results to a similar clustering for the textual component of the reviews. In fact, we plan on conducting the same analysis for text, as well as the other information that is presented to users of these platforms to investigate what information a consumer uses from different data sources. A comparison should be made to the visual content generated by the hotels as well. In a broader scope, this could be extremely important, because most methods for consumer search and information diffusion focus on a single modality, generally with few variables. We know that users look at both the text and images at the same time when making decisions, so understanding these interactions is imperative.

In general, we presented a useful visual analytics framework to process imagery at a large scale. The unstructured nature of imagery makes it difficult to mine opinions or consumer perceptions. We have shown how marketers can use the deep embedded clustering method to discover performance areas within UGI that are part of the consumer reviews and how to use this to investigate where they are performing well and in what areas they could improve in comparison to competition. Other studies that utilize visual analytics to solve business problems already highlighted the usefulness of these methods, but they generally need some kind of human feedback or coding to be effective. This unsupervised method is easily scalable and adaptable to larger datasets and different domains, without requiring assumptions about the distribution or underlying mechanisms.

In conclusion, visual analytics has many promising applications. Ordenes and Zhang 2019 highlight several exciting directions and marketing applications such as shopping in Amazon's cashless stores (Amazon Go) or Zillow's pricing algorithms. Here visual analytics is used by Amazon to detect and automatically check out the items consumers put in their bags and by Zillow to automatically adjust prices based on detected objects in imagery. For electronic marketing specifically, research and marketers alike can use visual analytics, such as the method presented in this research, to understand consumer perceptions and opinions expressed through UGI. This is not limited to online travel or review platforms, but can also be applied to restaurants or other products and it can be used on platforms such as eBay, Amazon or Social Media. Another area that we envision to be worth exploring in this space is the automatic generation of effective visual marketing stimuli based on these new insights.

Table 5.1: Overall UGI Clustering. 10 Clusters, with 4 sample images, mean, standard deviation and number of samples per cluster.

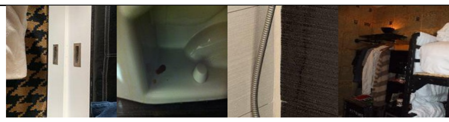
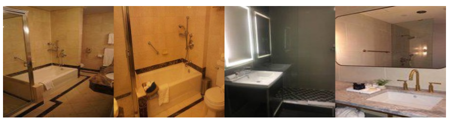
Cluster	Label	Samples	Mean	SD	#UGI
1	Zoom		3.93	1.25	450
2	Bar/Lobby		4.54	0.69	959
3	Seating Areas		4.54	0.77	793
4	Views		4.61	0.63	847
5	Hotel Front		4.53	0.69	841
6	Food/Drinks		4.57	0.71	737
7	Rooms		4.49	0.74	1700
8	Style Details		4.46	0.79	838
9	Bathrooms		4.36	0.86	1100
10	Empire State		4.68	0.57	768

Figure 5.2: Average rating per cluster for 3 hotels. The selection includes a hotel with a below average rating (top), average rating (middle) and above average rating (bottom). The shaded blue bars represent the population average rating of the cluster and orange the average rating of the cluster for the hotel. The green (red) boxes indicate statistically significant positive (negative) differences from the population average.

Chapter 6

Conclusion

The main goal of this dissertation was to understand the impact of visual content online. As a means to reach that goal, I studied how we can leverage AI and machine learning for marketing, using theory-driven investigation and data-driven exploration approaches. This resulted in interesting managerial and methodological implications. The importance of studying the impact of visual content is illustrated by calls from researchers to understand the impact of (online) visual content (Blanco, Sarasa, and Sanclemente, 2010; Kirillova and Chan, 2018; Liu, Dzyabura, and Mizik, 2018), calls from scholars to further advance visual analytics methods for marketing research (Ordenes and Zhang, 2019; Zhang and Luo, 2018; Zhang et al., 2019), and large scale investments in AI and machine learning by businesses.

The four chapters presented in this dissertation each contribute to the main research goal individually. Chapter 2 focused on managerial guidance to the implementation of marketing AI projects. Through the lens of the data mining framework, CRISP-DM (Chapman et al., 2000), I explored the phases of marketing AI projects and questions managers should ask during each. I demonstrated that the success of a marketing AI project depends heavily on collaboration and communication between stakeholders, understanding the data (including how it is generated and where it is stored) and how to process it effectively. I dug deeper into the latter in the following chapters. In chapter 3, I presented a framework to extract visual complexity from social media images at scale, using a theory-driven investigation approach. I then combined this framework with an econometric model to study the non-linear relationship of visual complexity measures with liking of firm-generated imagery. In chapter 4, I used an approach that leans more towards data-driven exploration. I introduced an end-to-end machine learning framework that relates the information from thumbnail images on e-commerce platforms to consumer clicks and consideration set formation. First, I quantified the impact of the product image. Second, I demonstrated how to use interpretable machine learning methods to understand what types of images work for certain locations and how the image information interacts with the textual and numerical information that is presented to consumers during search. Finally, in chapter 5, I provided a quick look into the future of visual analytics for marketing. An unsupervised clustering method was used to mine consumer opinions from the visual content they share. Clustering this content revealed interesting performance areas for hotels, and combined with ratings, it highlighted over- or under performance. It demonstrated the potential of unsupervised learning techniques for marketing research. Combined, these chapters present theoretical, methodological, and managerial advancements to the marketing domain, which I will discuss in the remainder for this chapter.

6.1 Theoretical Implications

The findings of the empirical work presented in this dissertation contribute to a better understanding of how consumers interact with visual content online. The theoretical contributions of each individual empirical study are outlined in the corresponding chapters. In this section, I discuss the most important overarching contributions on a broader level. First, we have come closer to understanding what aspects of firm-generated imagery makes them more effective. For visual complexity, findings of two previous works were contrasting. With respect to the feature complexity, Pieters, Wedel, and Batra (2010) found a negative linear relationship with the attitude towards ads, whereas Shin et al. (2019) found a positive linear relationship with the engagement of social media engagement. The exact opposite was found for the design complexity by these same papers. In chapter 3, we find truth in both studies by means of a non-linear relationship. We find an inverted u-shape relationship with consumer liking for the feature complexity and a regular u-shape relationship between consumer liking and design complexity. So, the findings by Pieters, Wedel, and Batra (2010) hold for the middle of the spectrum towards the end of the spectrum of visual complexity (i.e., negative for feature complexity and positive for design complexity), whereas the findings of Shin et al. (2019) corresponding with the beginning of the spectrum towards the middle of the spectrum of visual complexity (i.e., positive for feature complexity and negative for design complexity). In addition, and perhaps most importantly, we find that the impact is more nuanced and there are differential relationships between the individual aspects that constitute the two types of visual complexity and consumer liking. It is also important to note, that we controlled for a wide-range of other image attributes, such as content and photography, and the visual complexity measures impact the liking of social media posts above and beyond these other image aspects. In chapter 4, we don't explore theory-driven feature engineering, but instead we aim for ex-post interpretation (i.e. data-driven exploration). The results show that different aspects and types of images work for different locations, and possible other circumstances. Showing a certain type of image can positively influence the perception of service quality and other hotel qualities (Wang, Tsai, and Tang, 2018; Kirillova and Chan, 2018). In addition, in a fMRI experiment we demonstrate that high-CTR images activate different brain regions than low-CTR images, similar to what was found with respect to attractive product packaging (Stoll, Baecke, and Kenning, 2008; Hubert et al., 2013). In conclusion, the most important contribution of this dissertation to visual marketing is that certain image aspects, such as visual complexity, can universally make an image more effective, but that the "perfect" image is also highly context dependent. Second, the findings in chapter 3 and 4 advance our understanding on the impact of visual content on consumer decisions. In chapter 3, I highlight the differential impact of individual visual complexity measures on consumer liking. In chapter 4, I quantify the impact of the product image on consumer search and click decisions. First, we establish that images convey enough information to accurately predict the relative click-through rates of hotels, based on the image information alone. Second, we confirm these findings by comparing individual consumer-level click predictions and show that we are over 10% more accurate in predicting consumer clicks. This highlights that the information conveyed to consumers through the product image is part of a consumer's decision to take a closer look at a hotel. Based on previous research it was expected that imagery indeed influences consumer decisions (Zhang et al., 2017; Zhang et al., 2019; Pieters, Wedel, and Batra, 2010; Shin et al., 2019), but I have now quantified their effect at scale, also in settings that had remained under explored.

Third, imagery are important factors that consumers consider in their decision, but they also influence the importance of other information presented to consumers through textual data. The consumer-level prediction methods highlight that factors, such as price or free cancellation, change after inclusion of the image features. This is important, because it could mean that either the information presented by the image correlates strongly with the information presented by text and numbers, or that the image information influences the consumers' perception of the other information (Wang, Tsai, and Tang, 2018). Based on expectations of previous research (Pan, Zhang, and Law, 2013; Park, Yin, and Son, 2019; Zhang et al., 2019), I expect it to be a combination of both. Either way, this is an important contribution and an area for future research, because research in consumer search and choice has not (yet) incorporated image information in their methods. Ignoring the information presented to consumers through firm-generated imagery can lead to a bias in the estimation of the impact of certain variables on consumer decisions.

6.2 Methodological Implications

Another focal area of this dissertation was the development of new methods and frameworks to translate unstructured data, such as imagery, into insights. In this section, I discuss the main methodological contributions and developments to visual marketing. In chapter 3, I have expanded and automated measures of visual complexity proposed by (Pieters, Wedel, and Batra, 2010) and Shin et al., 2019. The measures translate images into theoretically relevant information that accurately reflects human perception. The proposed visual complexity measures have a differential impact on consumer liking, as shown in our study, but there are many settings in which visual complexity can influence consumer perception and attitude (Donderi, 2006; Palumbo et al., 2014; Machado et al., 2015). Our framework can aid future research on the interaction between online users and imagery. In fact, the code is made available, which enables researchers to automatically extract visual complexity measures from their imagery. In chapter 4, I presented an end-to-end machine learning framework that can be used to study consumers decisions in any setting in which they are presented with multiple modalities of information. We showed that a package of methods from deep learning and learning to rank allow marketers to study online consumers and their interaction with information (Zhang et al., 2017; Yoganarasimhan, 2020; Dzyabura, El Kihal, and Ibragimov, 2018). This opens up a wide range of possibilities for marketers to explore the impact of (visual) stimuli online. The (pre-trained) CNNs and the LambdaMART ranking model (or another boosted regression tree approach) can be easily implemented using open-source software and packages in widely used programming languages such as Python and R. I have shown how marketers and researchers can utilize and combine these methods to approach substantive marketing problems. Finally, in chapter 5, I demonstrated another method for processing and understanding large scale unstructured data. I leverage a Deep Embedded Clustering, a high-dimensional clustering method that is completely unsupervised. Unsupervised learning is crucial as it overcomes the problem of manually labelling or coding content, which is very costly. Unsupervised learning, such as the clustering method presented in this chapter, enables researchers to automatically label visual marketing stimuli, learn effective representations of visual marketing stimuli that can be used as input for statistical methods, or to summarize user-generated imagery - which we show in this chapter. I believe that unsupervised and/or self-supervised learning will be an important future direction of machine learning in marketing, especially for visual content, such as imagery and video. In general, in this dissertation and corresponding empirical studies, I have introduced and demonstrated effective and promising tools, methods, and frameworks that allow researchers and marketers alike to process a large number of visual content and study their relationship with consumers.

6.3 Managerial Implications

Chapter 2 highlighted the managerial importance of AI and its use for marketing. Most of the managerial implications of my entire dissertation can be placed directly into the steps of the framework presented of chapter 2 - Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.

Business Understanding: In chapter 2, I showed the importance of involving stakeholders of the organization. In addition, there should be a clear understanding of what the business problem is, why it is an important problem to solve and what a good solution will look like. In chapters 3 and 4 we have learned about the impact of visual content on consumer decisions. For social media, using this information, businesses can effectively improve their marketing content on a large scale to better connect with their customers. In turn, this will strengthen customer-brand relationships (Kumar et al., 2016). For product imagery on e-commerce websites, managers should compete on images, like they compete on pricing or other features. Consumers are influenced by the product presentation, such that it can compensate for other attributes that have shortcomings (Kirillova and Chan, 2018). Designing better images, or showing the right images, can make your product stand out and put you ahead of the competitors. The combination of understanding how different aspects of imagery influence attitudes and decisions and the automated extraction of this information directly from images makes a powerful tool for content managers.

Data Understanding, Data Preparation, Modeling, and Evaluation: Unstructured data requires going through these steps multiple times. First, it is necessary to follow the data understanding, data preparation, modeling and evaluation steps to make predictions about the unstructured data and turn this into features and representations that can be used in other methods. Chapters 3,4, and 5 all demonstrate this. In each of these chapters the images are first translated into visual complexity scores (e.g., Chapter 3) or feature vectors (e.g., Chapters 4 and 5). Second, we follow the same steps again using these new features to gain insights into the interaction of the consumer with the visual content. The translated unstructured information combined with the structured information that was already there can then be investigated further. For example, in Chapter 3, we combine all information sources in a regression model to study liking behavior. In Chapter 4, we use the feature vectors for the hotel-level CTR prediction and the corresponding visualizations, and in the LambdaMART model for consumer-level prediction we combine the feature vectors with the numerical and textual data that consumers also use in their click decisions. Chapter 5 utilizes the feature vectors of all images for the clustering algorithm. The Marketing AI process is often cyclic and there is a feedback between the phases. Modeling requires a specific data preparation while the output of the analyzed data also determines which modeling technique is best to use. In addition, modeling is often required to make predictions about unstructured data which in turn is used as input again into another model.

Deployment: The methods proposed in chapters 3 and 4 would ideally be deployed to improve the design and curation of new and existing visual marketing stimuli. In chapter 3, our analysis on Instagram filters showed an easy to achieve 3% increase of likes based on the feature complexity scores. This means, that content managers are able to improve liking by 3% with just a few additional clicks when deployed properly. In the Appendix of chapter 3 we have demonstrated such a filter guide. With these findings and those with respect to the design complexity, in combination with the open-source code, one could design a tool or dashboard that automatically rates, and potentially optimizes, newly produced FGI that maximizes liking, or other outcomes of interest. In chapter 4, managers could leverage the LambdaMART model scores for a particular search result and calculate the scores for every available image, which can then be used to select the most relevant image to present to the consumer. Li et al. (2019) show, with a similar methodology, how optimizing the photo lay out leads to an increase in demand and annual revenue for a property on Airbnb. Our approach allows managers the ability to easily examine millions of images and determine the role that these images are playing in online transactions. Finally, in chapter 5, we discussed how to use the unsupervised method to summarize user-generated imagery. The application demonstrated how managers can easily identify performance areas.

6.4 Future Research

This dissertation provides valuable avenues for further research in the area of visual marketing. In each chapter we reflected on the limitations of the research and provided some directions to future research. In this section, I highlight two avenues of future research that I believe are important and timely with respect to consumers and their interaction with unstructured data, and I end with a brief opinion on the future of visual analytics for marketing.

First, as mentioned briefly in the theoretical implications sections, we view the investigation of the dynamics between textual and visual content online as an important future research area. Textual content has been investigated extensively in the past decade (Berger et al., 2020), and in this dissertation, and other recent works (Zhang et al., 2017; Zhang and Luo, 2018; Dzyabura, El Kihal, and Ibragimov, 2018), we have demonstrated similar investigations into the impact of imagery. However, we have not seen research in marketing that specifically focuses on the dynamics between textual- and visual content. The results in chapter 4 showed that the importance of some of the textual and numerical attributes change after including the image information. This means that our consumer prediction method uses the numerical and textual information differently when the image information is present. As discussed, this can mean two things: either there is significant correlation between the information conveyed through the different modalities, or the information from the visual content influences the evaluation and perception of the textual and numerical content. Either way, studying this interaction seems imperative to understand consumer decisions online. Recent research suggests that instead of focusing on predicting human behavior, a task that AI under performs at, we should use AI and machine learning methods to understand how the information environment affects consumer choices (Gal and Simonson, n.d.). I envision that methods such as eye-tracking analysis, conjoint analysis, and (interpretable) multi-modal deep learning methods, possibly in combination with existing search and choice models will provide us with the insights needed to unfold these dynamics.

Second, this dissertation focused on static visual content, which means that the natural next step is moving visual content (e.g., video). Since 1941, television advertising has been a crucial form of communication between firms and consumers. In addition, online video platforms, such as YouTube, have become important marketing channels. Moreover, most social media platforms now accommodate videos and new platforms such as TikTok exclusively use videos. In short, videos are relevant for marketing and there is a need to understand their impact. However, beyond some investigations into the impact of TV advertising, research in this space has been limited, which can again be explained by a lack of methods and/or availability of resources. Schwenzow et al. (2020) recognized this and provided us with an overview of current state-of-the-art methods readily available to use for business research. Many of the methods of this dissertation will be, at least somewhat, relevant for studying the impact of videos. Overall, the development of methodology in this space is necessary, and studying the impact of videos is an exciting new area of research in marketing.

Third, the methods presented in chapters 3, 4, and 5, aid our understanding of consumer interactions with visual content and chapter 2 helps businesses and policy makers use AI for their marketing endeavours. The framework and the methods introduced in this dissertation warrant responsibilities. As marketers, we have a responsibility to ensure the integrity of the information presented to the consumers; to prevent manipulation of the content consumers use to construct their choices (Gal and Simonson, n.d.); and to be weary of personalization, targeting, and content optimization for the sole purpose of profit maximization. In fact, we even need to take this one step further and use these methods to address and solve some of the grand challenges of society (Academy of Marketing Science, 2021) by ensuring that the visual analytics methods are not only about the consumers, but for the consumers. In addition, our understanding of the impact of visual content can aid marketing communication efforts about sustainable consumption - to nudge people to serve the common good. This goal, or awareness in general, can and should be incorporated into the different stages of the marketing AI framework presented in chapter 2.

Finally, I want to briefly discuss my view on using visual analytics for marketing. Its use for marketing has been established, both through this dissertation and through its recent attention by the field. I have introduced the two main ways to study visual marketing using visual analytics: theory-driven investigation and data-driven exploration. An example of the former I have demonstrated in chapter 3, in which I took the concept of visual complexity, and an existing theoretical framework by Pieters, Wedel, and Batra, 2010, to construct automated measures that were expected to influence liking. An example of the latter I have demonstrated in chapter 4, in which I theoretically establish the potential importance of imagery, but instead of extracting specific theory-driven features I used methods - visualization, feature importance and a fMRI experiment - to explain why images impact decisions and, perhaps more importantly, explore new theory. Traditional marketing prefers the theory-driven feature engineering approach, because this is the way research has been done for decades, and I certainly see merit in this approach. It allows for a solid theory-driven examination of images, often with years of knowledge to explain the drivers behind the impact the content has and how it is perceived by consumers. It does, however, require making a trade-off between interpretability and predictive ability (Rajaram and Manchanda, 2020), because you limit viewing the visual content through a single theoretical lens as you reduce the thousands of numbers that represent one image to a single score. In addition, it requires existing theory, which in the study of visual marketing stimuli, or unstructured data for that matter, is often lacking. This is where I believe the data-driven feature interpretation approach can offer new ways of exploring theory that could lead to exciting new hypotheses. Given the exponential increase of (unstructured) data and because humans simply can't label data fast enough, I envision that unsupervised and self-supervised learning methods and the data-driven exploration approach will form the basis for many future visual marketing studies.

Bibliography

Abdi, Asad et al. (2018). “Machine learning-based multi-documents sentiment-oriented summarization using linguistic treatment”. In: *Expert Systems with Applications* 109, pp. 66–85.

Academy of Marketing Science, Journal of the (2021). *Reimagining Marketing Strategy: Driving the Debate on Grand Challenges*. URL: `https://www.springer.com/journal/11747/updates/18087648`.

Agrawal, Ajay K, Joshua S Gans, and Avi Goldfarb (2018a). *Prediction, Judgment and Complexity: A Theory of Decision Making and Artificial Intelligence*. Tech. rep. National Bureau of Economic Research.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb (2018b). *Prediction Machines: The simple economics of artificial intelligence*. Harvard Business Press.

Ahani, Ali et al. (2019). “Market segmentation and travel choice prediction in Spa hotels through TripAdvisor’s online reviews”. In: *International Journal of Hospitality Management* 80, pp. 52–77.

Aitchison, Jim (2012). *Cutting edge advertising: How to create the world’s best print for brands in the 21st century*. FT Press.

Alcalá-Fdez, J et al. (2008). *Knowledge extraction based on evolutionary learning*.

Ambler, Tim, Andreas Ioannides, and Steven Rose (2000). “Brands on the brain: neuro-images of advertising”. In: *Business Strategy Review* 11.3, pp. 17–30.

Andrews, Rick L and TC Srinivasan (1995). “Studying consideration effects in empirical choice models using scanner panel data”. In: *Journal of Marketing Research* 32.1, pp. 30–41.

Ariely, Dan and Gregory S Berns (2010). “Neuromarketing: the hope and hype of neuroimaging in business”. In: *Nature reviews neuroscience* 11.4, pp. 284–292.

Artese, Maria Teresa, Gianluigi Ciocca, and Isabella Gagliardi (2014). “Good 50x70 project: a portal for cultural and social campaigns”. In: *Archiving Conference*. Vol. 2014. 1. Society for Imaging Science and Technology, pp. 213–218.

Attneave, Fred (1954). “Some informational aspects of visual perception.” In: *Psychological review* 61.3, p. 183.

Baek, Jooa and Chihyung Michael Ok (2017). “The power of design: How does design affect consumers’ online hotel booking?” In: *International Journal of Hospitality Management* 65, pp. 1–10.

Balducci, Bitty and Detelina Marinova (2018). “Unstructured data in marketing”. In: *Journal of the Academy of Marketing Science* 46.4, pp. 557–590.

Batra, Rajeev and Kevin Lane Keller (2016). “Integrating Marketing Communications: New findings, new lessons, and new ideas”. In: *Journal of marketing* 80.6, pp. 122–145.

Beleites, Claudia et al. (2013). “Sample size planning for classification models”. In: *Analytica chimica acta* 760, pp. 25–33.

Bellizzi, Joseph A, Ayn E Crowley, and Ronald W Hasty (1983). “The effects of color in store design.” In: *Journal of retailing*.

Bellizzi, Joseph A and Robert E Hite (1992). “Environmental color, consumer feelings, and purchase likelihood”. In: *Psychology & marketing* 9.5, pp. 347–363.

Berezina, Katerina et al. (2016). “Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews”. In: *Journal of Hospitality Marketing & Management* 25.1, pp. 1–24.

Bergen, James R and Michael S Landy (1991). “Computational modeling of visual texture segregation”. In: *Computational models of visual processing* 17, pp. 253–271.

Berger, Jonah and Katherine L Milkman (2012). “What makes online content viral?” In: *Journal of marketing research* 49.2, pp. 192–205.

Berger, Jonah et al. (2020). “Uniting the tribes: Using text for marketing insight”. In: *Journal of Marketing* 84.1, pp. 1–25.

Berlyne, Daniel E (1958). “The influence of complexity and novelty in visual figures on orienting responses.” In: *Journal of experimental psychology* 55.3, p. 289.

Bettman, James R, Mary Frances Luce, and John W Payne (1998). “Constructive consumer choice processes”. In: *Journal of consumer research* 25.3, pp. 187–217.

Beukeboom, Camiel J, Peter Kerkhof, and Metten de Vries (2015). “Does a virtual like cause actual liking? How following a brand’s Facebook updates enhances brand evaluations and purchase intention”. In: *Journal of interactive marketing* 32, pp. 26–36.

Bevan, William and William F Dukes (1953). “Color as a variable in the judgment of size”. In: *The American journal of psychology* 66.2, pp. 283–288.

Bilgihan, Anil, Fevzi Okumus, and Khaldoon Nusair (2013). "Online Hotel Booking Experience: Flow Theory, Measuring Online Customer Experience and Managerial Implications". In:

Bischl, Bernd et al. (2016). "mlr: Machine Learning in R". In: *Journal of Machine Learning Research* 17.170, pp. 1–5. URL: `http://jmlr.org/papers/v17/15-066.html`.

Blanco, Carlos Flavián, Raquel Gurrea Sarasa, and Carlos Orús Sanclemente (2010). "Effects of visual and textual information in online product presentations: looking for the best combination in website design". In: *European Journal of Information Systems* 19.6, pp. 668–686.

Bonin, Patrick et al. (2003). "A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies". In: *Behavior Research Methods, Instruments, & Computers* 35.1, pp. 158–167.

Book, Albert C and C Dennis Schick (1997). *Fundamentals of copy & layout*. McGraw Hill Professional.

Borth, Damian et al. (2013). "Large-scale visual sentiment ontology and detectors using adjective noun pairs". In: *MM*.

Bronnenberg, Bart J and Wilfried R Vanhonacker (1996). "Limited choice sets, local price response, and implied measures of price competition". In: *Journal of Marketing Research* 33.2, pp. 163–173.

Brown, Meta S. (July 2015). "What IT Needs To Know About The Data Mining Process". In: *Forbes*. URL: `https://www.forbes.com/sites/metabrown/2015/07/29/what-it-needs-to-know-about-the-data-mining-process/#63bf06dd515f`.

Burges, Christopher J, Robert Ragno, and Quoc V Le (2007). "Learning to rank with nonsmooth cost functions". In: *Advances in neural information processing systems*, pp. 193–200.

Burges, Christopher JC (2010). "From ranknet to lambdarank to lambdamart: An overview". In: *Learning* 11.23-581, p. 81.

Burges, Chris et al. (2005). "Learning to rank using gradient descent". In: *Proceedings of the 22nd international conference on Machine learning*, pp. 89–96.

Burke, Marian Chapman and Julie A Edell (1989). "The impact of feelings on ad-based affect and cognition". In: *Journal of marketing research*, pp. 69–83.

Burlutskiy, Nikolay et al. (2016). "An investigation on online versus batch learning in predicting user behaviour". In: *International Conference on Innovative Techniques and Applications of Artificial Intelligence*. Springer, pp. 135–149.

Burnap, Alex and John Hauser (2018). "Predicting" Design Gaps" in the Market: Deep Consumer Choice Models under Probabilistic Design Constraints". In: *arXiv preprint arXiv:1812.11067*.

Burnap, Alex, John R Hauser, and Artem Timoshenko (2019). "Design and evaluation of product aesthetics: a human-machine hybrid approach". In: *Available at SSRN 3421771*.

Canny, John (1987). "A computational approach to edge detection". In: *Readings in computer vision*. Elsevier, pp. 184–203.

Cardy, Robert L and Gregory H Dobbins (1986). "Affect and appraisal accuracy: Liking as an integral dimension in evaluating performance." In: *Journal of applied psychology* 71.4, p. 672.

Centola, Damon and Michael Macy (2007). "Complex contagions and the weakness of long ties". In: *American Journal of Sociology* 113.3, pp. 702–734.

Chaitin, Gregory J (1977). "Algorithmic information theory". In: *IBM journal of research and development* 21.4, pp. 350–359.

Chakraborty, Ishita, Minkyung Kim, and K Sudhir (2019). "Attribute Sentiment Scoring with Online Text Reviews: Accounting for Language Structure and Attribute Self-Selection". In: *Available at SSRN 3395012*.

Chapelle, O, Y Chang, and TY Liu (2010). *The Yahoo! learning to rank challenge*.

Chapman, Pete et al. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*. Tech. rep. SPSS.

Chatterjee, Anjan (2004). "Prospects for a cognitive neuroscience of visual aesthetics". In: *Bulletin of psychology and the arts* 4.2, pp. 56–60.

Chevalier, J. and D. Mayzlin (2006). "The Effect of Word of Mouth on Sales: Online Book Reviews". In: *Journal of Marking Research* 43(3), pp. 345–354.

Chica, Manuel and William Rand (Oct. 2017). "Building agent-based decision support systems for word-of-mouth programs. A freemium application". In: *Journal of Marketing Research* 54, pp. 752–767. DOI: `http://dx.doi.org/10.1509/jmr.15.0443`.

Chica, Manuel et al. (2017). "Multimodal optimization: an effective framework for model calibration". In: *Information Sciences* 375, pp. 79–97.

Chintagunta, Pradeep, Dominique M Hanssens, and John R Hauser (2016). *Marketing science and big data*.

Chollet, François et al. (2015). *Keras*. `https://github.com/fchollet/keras`.

CIO (2019). *AI unleashes the power of unstructured data*. URL: `https://www.cio.com/article/3406806/ai-unleashes-the-power-of-unstructured-data.html`.

Colicev, Anatoli, Ashish Kumar, and Peter O'Connor (2019). "Modeling the relationship between firm and user generated content and the stages of the marketing funnel". In: *International Journal of Research in Marketing* 36.1, pp. 100–116.

Colicev, Anatoli et al. (2018). "Improving Consumer Mindset Metrics and Shareholder Value Through Social Media: The Different Roles of Owned and Earned Media". In: *Journal of marketing* 82.1, pp. 37–56.

Comaniciu, Dorin and Peter Meer (2002). "Mean shift: A robust approach toward feature space analysis". In: *IEEE Transactions on pattern analysis and machine intelligence* 24.5, pp. 603–619.

Conick, Hal (2016). "The Past, Present and Future of AI in Marketing". In: *Marketing News, December* 29.

Corchs, Silvia Elena et al. (2016). "Predicting complexity perception of real world images". In: *PloS one* 11.6, e0157986.

Couwenberg, Linda E et al. (2017). "Neural responses to functional and experiential ad appeals: Explaining ad effectiveness". In: *International Journal of Research in Marketing* 34.2, pp. 355–366.

Cristianini, Nello, John Shawe-Taylor, et al. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.

Cui, Geng, Man Leung Wong, and Hon-Kwong Lui (2006). "Machine learning for direct marketing response models: Bayesian networks with evolutionary programming". In: *Management Science* 52.4, pp. 597–612.

Darmon, D. et al. (2013). "Understanding the Predictive Power of Computational Mechanics and Echo State Networks in Social Media". In: *arXiv:1306.6111*.

Darwiche, Adnan (Sept. 2018). "Human-level Intelligence or Animal-like Abilities?" In: *Commun. ACM* 61.10, pp. 56–67. ISSN: 0001-0782. DOI: `10.1145/3271625`. URL: `http://doi.acm.org/10.1145/3271625`.

Daviet, Remi (2020). "Bayesian Deep Learning for Small Datasets: Leveraging Information from Product Pictures". In: *Work in progress*.

De Vries, Lisette, Sonja Gensler, and Peter SH Leeflang (2012). "Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing". In: *Journal of interactive marketing* 26.2, pp. 83–91.

Demšar, Janez et al. (2013). "Orange: Data Mining Toolbox in Python". In: *Journal of Machine Learning Research* 14, pp. 2349–2353. URL: `http://jmlr.org/papers/v14/demsar13a.html`.

Deng, Liqiong and Marshall Scott Poole (2010). "Affect in web interfaces: a study of the impacts of web page visual complexity and order". In: *Mis Quarterly*, pp. 711–730.

Deppe, Michael et al. (2005). "Nonlinear responses within the medial prefrontal cortex reveal when specific implicit information influences economic decision making". In: *Journal of neuroimaging* 15.2, pp. 171–182.

Donderi, Don C (2006). "Visual complexity: a review." In: *Psychological bulletin* 132.1, p. 73.

Draganska, Michaela and Daniel Klapper (2011). "Choice set heterogeneity and the role of advertising: An analysis with micro and macro data". In: *Journal of Marketing Research* 48.4, pp. 653–669.

Duan, W., B. Gu, and A. Whinston (2008). "Do online reviews matter? – An empirical investigation of panel data." In: *Decision Support Systems* 45, pp. 1007–1016.

Dzyabura, Daria, Siham El Kihal, and Marat Ibragimov (2018). "Leveraging the power of images in predicting product return rates". In: *Available at SSRN 3209307*.

Eickenberg, Michael et al. (2017). "Seeing it all: Convolutional network layers map the function of the human visual system". In: *NeuroImage* 152, pp. 184–194.

Einwiller, S. and S. Steilen (2015). "Handling complaints on social network sites - An analysis of complaints and complaint responses on Facebook and Twitter pages of large US companies". In: *Public Relations Review* 41(2), pp. 195–204.

Elith, Jane, John R Leathwick, and Trevor Hastie (2008). "A working guide to boosted regression trees". In: *Journal of Animal Ecology* 77.4, pp. 802–813.

Elliott, Nate (2014). *Instagram Is The King Of Social Engagement*. URL: `https://go.forrester.com/blogs/14-04-29-instagram-is-the-king%20%5C%5C%20-of-social-engagement`.

Erk, Susanne et al. (2002). "Cultural objects modulate reward circuitry". In: *Neuroreport* 13.18, pp. 2499–2503.

Fan, Rong-En et al. (2008). "LIBLINEAR: A Library for Large Linear Classification". In: *JMLR* 9, pp. 1871–1874.

Feldman, Ronen, James Sanger, et al. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press.

Forbes (2018). *How much data do we create every day*. URL: `https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/`.

— (2019). *Big data goes big*. URL: `https://www.forbes.com/sites/rkulkarni/2019/02/07/big-data-goes-big/?sh=78a02fd20d7b`.

Friedman, Jerome H (2001). "Greedy function approximation: a gradient boosting machine". In: *Annals of statistics*, pp. 1189–1232.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (2001). *The elements of statistical learning*. Vol. 1. 10. Springer series in statistics New York.

Gal, David and Itamar Simonson (n.d.). "Predicting consumers' choices in the age of the internet, AI, and almost perfect tracking: Some things change, the key challenges do not". In: *Consumer Psychology Review* ().

Garver, Michael S (2002). "Using data mining for customer satisfaction research". In: *Marketing Research* 14.1, p. 8.

Geissler, Gary L, George M Zinkhan, and Richard T Watson (2006). "The influence of home page complexity on consumer attention, attitudes, and purchase intent". In: *Journal of advertising* 35.2, pp. 69–80.

Gelli, Francesco et al. (2015). "Image popularity prediction in social media using sentiment and context features". In: *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, pp. 907–910.

Gersten, Wendy, Rüdiger Wirth, and Dirk Arndt (2000). "Predictive modeling in automotive direct marketing: tools, experiences and open issues". In: *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 398–406.

Girshick, Ross et al. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587.

Glöckner, Andreas and Tilmann Betsch (2008). "Multiple-reason decision making based on automatic processing." In: *Journal of experimental psychology: Learning, memory, and cognition* 34.5, p. 1055.

Goh, Khim-Yong, Cheng-Suang Heng, and Zhijie Lin (2013). "Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content". In: *Information systems research* 24.1, pp. 88–107.

Golkar Amnieh, Iman and Marjan Kaedi (2015). "Using estimated personality of social network members for finding influential nodes in viral marketing". In: *Cybernetics and Systems* 46.5, pp. 355–378.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep learning*. MIT press.

Gorn, Gerald J et al. (1997). "Effects of color as an executional cue in advertising: They're in the shade". In: *Management science* 43.10, pp. 1387–1400.

Gorn, Gerald J et al. (2004). "Waiting for the web: how screen color affects time perception". In: *Journal of marketing research* 41.2, pp. 215–225.

Groen, Iris IA et al. (2013). "From image statistics to scene gist: evoked neural activity reveals transition from low-level natural image structure to scene category". In: *Journal of Neuroscience* 33.48, pp. 18814–18824.

Groen, Iris IA et al. (2018). “Scene complexity modulates degree of feedback activity during object detection in natural scenes”. In: *PLoS computational biology* 14.12, e1006690.

Güçlü, Umut and Marcel AJ van Gerven (2015). “Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream”. In: *Journal of Neuroscience* 35.27, pp. 10005–10014.

Guo, Xifeng et al. (2017a). “Deep clustering with convolutional autoencoders”. In: *International conference on neural information processing*.

Guo, Xifeng et al. (2017b). “Improved deep embedded clustering with local structure preservation”. In: *International Joint Conference on Artificial Intelligence*.

Halevy, Alon, Peter Norvig, and Fernando Pereira (2009). “The unreasonable effectiveness of data”. In: *IEEE Intelligent Systems* 24.2, pp. 8–12.

Hall, Mark et al. (2009). “The WEKA data mining software: an update”. In: *ACM SIGKDD explorations newsletter* 11.1, pp. 10–18.

Harding, JA, M Shahbaz, A Kusiak, et al. (2006). “Data mining in manufacturing: a review”. In: *Journal of Manufacturing Science and Engineering* 128.4, pp. 969–976.

Hartmann, Jochen et al. (2019). “The power of brand selfies in consumer-generated brand images”. In: *Available at SSRN*.

Hasler, David and Sabine E Suesstrunk (2003). “Measuring colorfulness in natural images”. In: *Human vision and electronic imaging VIII*. Vol. 5007. International Society for Optics and Photonics, pp. 87–95.

Hauser, John R (2014). “Consideration-set heuristics”. In: *Journal of Business Research* 67.8, pp. 1688–1699.

Hauser, John R, Guilherme Liberali, and Glen L Urban (2014). “Website morphing 2.0: Switching costs, partial exposure, random exit, and when to morph”. In: *Management science* 60.6, pp. 1594–1616.

Hauser, John R et al. (2009). “Website morphing”. In: *Marketing Science* 28.2, pp. 202–223.

Havas (2017). *Welcome to Meaningful Brands 2017*. URL: `https://dk.havas.com/wp-content/uploads/sites/37/2017/02/mb17%20%5C%5C%20-brochure-final-web.pdf`.

He, Kaiming et al. (2016). “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

He, Kaiming et al. (2017). “Mask r-cnn”. In: *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969.

Heaps, Christopher and Stephen Handel (1999). "Similarity and features of natural textures." In: *Journal of experimental psychology: Human perception and performance* 25.2, p. 299.

Henderson, John M, Antje Nuthmann, and Steven G Luke (2013). "Eye movement control during scene viewing: Immediate effects of scene luminance on fixation durations." In: *Journal of experimental psychology: Human perception and performance* 39.2, p. 318.

Hewett, Kelly et al. (2016). "Brand buzz in the echoverse". In: *Journal of marketing* 80.3, pp. 1–24.

Heylighen, Francis (1997). "Objective, subjective and intersubjective selectors of knowledge". In: *Evolution and cCognition* 3.1, pp. 63–67.

Hinton, Geoffrey E and Sam T Roweis (2003). "Stochastic neighbor embedding". In: *Advances in neural information processing systems.*

Honka, Elisabeth and Pradeep Chintagunta (2016). "Simultaneous or sequential? search strategies in the us auto insurance industry". In: *Marketing Science* 36.1, pp. 21–42.

Hu, Ya-Han, Yen-Liang Chen, and Hui-Ling Chou (2017). "Opinion mining from online hotel reviews–a text summarization approach". In: *Information Processing & Management* 53.2, pp. 436–449.

Hubert, Marco et al. (2013). "Neural correlates of impulsive buying tendencies during perception of product packaging". In: *Psychology & Marketing* 30.10, pp. 861–873.

Jai, Tun-Min, Michael W O'Boyle, and Dan Fang (2014). "Neural correlates of sensory-enabling presentation: An fMRI study of image zooming and rotation video effects on online apparel shopping". In: *Journal of Consumer Behaviour* 13.5, pp. 342–350.

Jenkinson, Mark et al. (2002). "Improved optimization for the robust and accurate linear registration and motion correction of brain images". In: *Neuroimage* 17.2, pp. 825–841.

Jeong, Miyoung and Jiyoung Choi (2005). "Effects of picture presentations on customers' behavioral intentions on the web". In: *Journal of Travel & Tourism Marketing* 17.2-3, pp. 193–204.

Jones, Eric, Travis Oliphant, Pearu Peterson, et al. (2001–). *SciPy: Open source scientific tools for Python.* [Online; accessed ¡today¿]. URL: `http://www.scipy.org/`.

Kaldis, Konstantina and Emmanouil Kaldis (2008). "Emmantina and Palmira beach hotels: Distribution for independent hotels". In: *EggerR. BuhalisD.(Eds.), eTourism: Case studies*, pp. 65–73.

Kalliatakis, Grigorios (2017). *Keras-VGG16-Places365.* `https://github.com/GKalliatakis/Keras-VGG16-places365`.

Kamitani, Yukiyasu and Frank Tong (2005). "Decoding the visual and subjective contents of the human brain". In: *Nature neuroscience* 8.5, pp. 679–685.

Khosla, Aditya, Atish Das Sarma, and Raffay Hamid (2014). "What makes an image popular?" In: *Proceedings of the 23rd international conference on World wide web.* ACM, pp. 867–876.

Kim, Jinwoo and Jae Yun Moon (1998). "Designing towards emotional usability in customer interfaces trustworthiness of cyber-banking system interfaces". In: *Interacting with Computers* 10, pp. 1–29.

Kim, Sungsoo and Anna S Mattila (2011). "An examination of electronic video clips in the context of hotel Websites". In: *International Journal of Hospitality Management* 30.3, pp. 612–618.

Kirillova, Ksenia and Janelle Chan (2018). ""What is beautiful we book": hotel visual appeal and expected service quality". In: *International Journal of Contemporary Hospitality Management* 30.3, pp. 1788–1807.

Klostermann, Jan et al. (2018). "Extracting brand information from social networks: Integrating image, text, and social tagging data". In: *International Journal of Research in Marketing* 35.4, pp. 538–556.

Knutson, Brian et al. (2007). "Neural predictors of purchases". In: *Neuron* 53.1, pp. 147–156.

Kumar, Ashish et al. (2016). "From social to sale: The effects of firm-generated content in social media on customer behavior". In: *Journal of marketing* 80.1, pp. 7–25.

Kumar, V et al. (2013). "Practice prize winner—creating a measurable social media marketing strategy: increasing the value and ROI of intangibles and tangibles for hokey pokey". In: *Marketing Science* 32.2, pp. 194–212.

Kumar, V et al. (2019). "Understanding the role of artificial intelligence in personalized engagement marketing". In: *California Management Review* 61.4, pp. 135–155.

Lamme, Victor AF and Pieter R Roelfsema (2000). "The distinct modes of vision offered by feedforward and recurrent processing". In: *Trends in neurosciences* 23.11, pp. 571–579.

LeCun, Yann et al. (1998). "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11, pp. 2278–2324.

LeCun, Yann and Yoshua Bengio (1995). "Convolutional networks for images, speech, and time series". In: *The handbook of brain theory and neural networks* 3361.10.

Leeuwenberg, EL (1969). "Quantitative specification of information in sequential patterns." In: *Psychological review* 76.2, p. 216.

Lei, Tao et al. (2018). "Superpixel-based fast fuzzy C-means clustering for color image segmentation". In: *IEEE Transactions on Fuzzy Systems* 27.9, pp. 1753–1766.

Lemmens, Aurélie and Christophe Croux (2006). "Bagging and boosting classification trees to predict churn". In: *Journal of Marketing Research* 43.2, pp. 276–286.

Li, Hanwei et al. (2019). "Estimating and Exploiting the Impact of Photo Layout in Sharing Economy". In: *Available at SSRN.*

Li, Yiyi and Ying Xie (2017). "Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement". In: *Journal of Marketing Research*, p. 0022243719881113.

Libai, Barak, Eitan Muller, and Renana Peres (2013). "Decomposing the value of word-of-mouth seeding programs: Acceleration versus expansion". In: *Journal of Marketing Research* 50.2, pp. 161–176.

Lichtlé, Marie-Christine (2007). "The effect of an advertisement's colour on emotions evoked by attitude towards the ad: The moderating role of the optimal stimulation level". In: *International journal of advertising* 26.1, pp. 37–62.

Lilien, Gary L (2011). "Bridging the academic-practitioner divide in marketing decision models". In: *Journal of Marketing* 75.4, pp. 196–210.

Lindgaard, Gitte et al. (2006). "Attention web designers: You have 50 milliseconds to make a good first impression!" In: *Behaviour & information technology* 25.2, pp. 115–126.

Liu, Liu, Daria Dzyabura, and Natalie Mizik (2018). "Visual listening in: Extracting brand image portrayed on social media". In: *Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence.*

— (2020). "Visual listening in: Extracting brand image portrayed on social media". In: *Marketing Science* 39.4, pp. 669–686.

Lovett, Mitchell J, Renana Peres, and Ron Shachar (2013). "On brands and word of mouth". In: *Journal of marketing research* 50.4, pp. 427–444.

Luo, Xueming, Jie Zhang, and Wenjing Duan (2013). "Social media and firm equity value". In: *Information systems research* 24.1, pp. 146–163.

Ma, Yufeng et al. (2018). "Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep leaning". In: *International Journal of Hospitality Management* 71, pp. 120–131.

Ma, Yun and Qing Li (2019). "A weakly-supervised extractive framework for sentiment-preserving document summarization". In: *World Wide Web* 22.4, pp. 1401–1425.

Maaten, Laurens van der and Geoffrey Hinton (2008). "Visualizing data using t-SNE". In: *Journal of machine learning research* 9.Nov, pp. 2579–2605.

— (2008). "Visualizing data using t-SNE". In: *The Journal of Machine Learning Research.*

Macal, C. M. and M. J North (2005). "Tutorial on agent-based modeling and simulation". In: *Proceedings of the 37th conference on Winter simulation*. ACM, pp. 2–15.

Machado, Penousal et al. (2015). "Computerized measures of visual complexity". In: *Acta psychologica* 160, pp. 43–57.

MacQueen, James et al. (1967). "Some methods for classification and analysis of multivariate observations". In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Oakland, CA, USA, pp. 281–297.

Martin Abadi et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. URL: `http://tensorflow.org/`.

Mathison, Rob (2018). *23+ Useful Instagram Statistics for Social Media Marketers*. URL: `https://blog.hootsuite.com/instagram-statistics/`.

Mazloom, Masoud et al. (2016). "Multimodal popularity prediction of brand-related social media posts". In: *Proceedings of the 24th ACM international conference on Multimedia*. ACM, pp. 197–201.

McParlane, Philip J, Yashar Moshfeghi, and Joemon M Jose (2014). "Nobody comes here anymore, it's too crowded; Predicting Image Popularity on Flickr". In: *Proceedings of International Conference on Multimedia Retrieval*. ACM, p. 385.

Meyers-Levy, Joan and Laura A Peracchio (1995). "Understanding the effects of color: How the correspondence between available and required resources affects attitudes". In: *Journal of consumer research* 22.2, pp. 121–138.

Mitchell, Tom M et al. (1997). *Machine learning*. McGraw-Hill Boston, MA:

Mochon, Daniel et al. (2017). "What are likes worth? A Facebook page field experiment". In: *Journal of marketing research* 54.2, pp. 306–317.

Moe, Wendy W (2006). "An empirical two-stage choice model with varying decision rules applied to internet clickstream data". In: *Journal of Marketing Research* 43.4, pp. 680–692.

Moreland, Richard L and Robert B Zajonc (1977). "Is stimulus recognition a necessary condition for the occurrence of exposure effects?" In: *Journal of personality and social psychology* 35.4, p. 191.

Morin, Christophe (2011). "Neuromarketing: the new science of consumer behavior". In: *Society* 48.2, pp. 131–135.

Moro, Sergio, Raul Laureano, and Paulo Cortez (2011). "Using data mining for bank direct marketing: An application of the crisp-dm methodology". In: *Proceedings of European Simulation and Modelling Conference-ESM'2011*. EUROSIS-ETI, pp. 117–121.

Murphy, Kevin P. (2012). *Machine Learning: A Probabilistic Perspective.* The MIT Press.

Musante, Michael D, David C Bojanic, and Jian Zhang (2009). "An evaluation of hotel website attribute utilization and effectiveness by hotel class". In: *Journal of Vacation Marketing* 15.3, pp. 203–215.

Nagle, Fintan and Nilli Lavie (2020). "Predicting human complexity perception of real-world scenes". In: *Royal Society Open Science* 7.5, p. 191487.

Nanne, Annemarie J et al. (2020). "The use of computer vision to analyze brand-related user generated image content". In: *Journal of Interactive Marketing.*

Nelson, Roy Paul (1994). *The design of advertising.* Brown & Benchmark Pub.

Netzer, Oded et al. (2012). "Mine your own business: Market-structure surveillance through text mining". In: *Marketing Science* 31.3, pp. 521–543.

Newman, Mark, Albert-László Barabási, and Duncan J Watts (2006). *The structure and dynamics of networks.* Princeton University Press.

Noone, Breffni M and Stephani KA Robson (2016). "Understanding consumers' inferences from price and nonprice information in the online lodging purchase decision". In: *Service Science* 8.2, pp. 108–123.

Nunnally, Jum C (1978). *Psychometric Theory: 2d Ed.* McGraw-Hill.

Nuthmann, Antje (2017). "Fixation durations in scene viewing: Modeling the effects of local image features, oculomotor parameters, and task". In: *Psychonomic bulletin & review* 24.2, pp. 370–392.

Nuthmann, Antje and Wolfgang Einhäuser (2015). "A new approach to modeling the influence of image features on fixation selection in scenes". In: *Annals of the New York Academy of Sciences* 1339.1, pp. 82–96.

OECD (2020). *E-commerce in the times of COVID-19.* URL: `http://www.oecd.org/coronavirus/policy-responses/e-commerce-in-the-time-of-covid-19-3a2b78e8/`.

Oliva, Aude (2005). "Gist of the scene". In: *Neurobiology of attention.* Elsevier, pp. 251–256.

Olivia, Aude et al. (2004). "Identifying the perceptual dimensions of visual complexity of scenes". In: *Proceedings of the Annual Meeting of the Cognitive Science Society.* Vol. 26.

Onwubolu, Godfrey (2009). "An inductive data mining system framework". In: *Proceedings of the International Workshop on Inductive Modeling (IWIM'09).* Citeseer, pp. 108–113.

Ordenes, Francisco Villarroel and Shunyuan Zhang (2019). "From words to pixels: text and image mining methods for service research". In: *Journal of Service Management.*

Osman, Maddy (2018). *18 Instagram Stats Every Marketer Should Know for 2018.* URL: https://sproutsocial.com/insights/instagram-stats/.

Overgoor, Gijs, William Rand, and Willemijn Van Dolen (2020). "The Champion of Images: Understanding the role of images in the decision-making process of online hotel bookings". In: *Proceedings of the 53rd Hawaii International Conference on System Sciences.*

Overgoor, Gijs et al. (2017). "A spatio-temporal category representation for brand popularity prediction". In: *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval.* ACM, pp. 233–241.

Palmer, Stephen E (1999). *Vision science: Photons to phenomenology.* MIT press.

Palumbo, Letizia et al. (2014). "Examining visual complexity and its influence on perceived duration". In: *Journal of vision* 14.14, pp. 3–3.

Pan, Bing and Lixuan Zhang (2016). "An eyetracking study on online hotel decision making: The effects of images and umber of options". In:

Pan, Bing, Lixuan Zhang, and Rob Law (2013). "The complex matter of online hotel choice". In: *Cornell Hospitality Quarterly* 54.1, pp. 74–83.

Park, Sangwon, Yizhen Yin, and Byung-Gak Son (2019). "Understanding of online hotel booking process: A multiple method approach". In: *Journal of Vacation Marketing* 25.3, pp. 334–348.

Parkhi, Omkar M, Andrea Vedaldi, and Andrew Zisserman (2015). "Deep face recognition". In:

Pecchinenda, Anna et al. (2014). "The pleasantness of visual symmetry: Always, never or sometimes". In: *PloS one* 9.3, e92685.

Pedregosa, F. et al. (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Peracchio, Laura A and Joan Meyers-Levy (2005). "Using stylistic properties of ad pictures to communicate with consumers". In: *Journal of consumer research* 32.1, pp. 29–40.

Petty, Richard E and John T Cacioppo (1986). "The elaboration likelihood model of persuasion". In: *Communication and persuasion.* Springer, pp. 1–24.

Pieters, FGM, Tulin Erdem, and Ana Martinovici (2019). "Eye-Movements, Attention, and Utility Accumulation During Brand Choice". In: *Tulin and Martinovici, Ana, Eye-Movements, Attention, and Utility Accumulation During Brand Choice (February 4, 2019).*

Pieters, Rik and Michel Wedel (2004). "Attention capture and transfer in advertising: Brand, pictorial, and text-size effects". In: *Journal of Marketing* 68.2, pp. 36–50.

Pieters, Rik, Michel Wedel, and Rajeev Batra (2010). "The stopping power of advertising: Measures and effects of visual complexity". In: *Journal of marketing* 74.5, pp. 48–60.

Plassmann, Hilke, Thomas Zoëga Ramsøy, and Milica Milosavljevic (2012). "Branding the brain: A critical review and outlook". In: *Journal of consumer psychology* 22.1, pp. 18–36.

Potter, Mary C (1976). "Short-term conceptual memory for pictures." In: *Journal of experimental psychology: Human learning and memory* 2.5, p. 509.

Potter, Mary C et al. (2014). "Detecting meaning in RSVP at 13 ms per picture". In: *Attention, Perception, & Psychophysics* 76.2, pp. 270–279.

Putrevu, Sanjay, Joni Tan, and Kenneth R Lord (2004). "Consumer responses to complex advertisements: The moderating role of need for cognition, knowledge, and gender". In: *Journal of current issues & research in advertising* 26.1, pp. 9–24.

Rajaram, Prashant and Puneet Manchanda (2020). "VIDEO INFLUENCERS: UNBOXING THE MYSTIQUE". In: *Work in progress.*

Rand, William and Roland T Rust (2011). "Agent-based modeling in marketing: Guidelines for rigor". In: *International Journal of Research in Marketing* 28.3, pp. 181–193.

Reimann, Martin et al. (2010). "Aesthetic package design: A behavioral, neural, and psychological investigation". In: *Journal of Consumer Psychology* 20.4, pp. 431–441.

Rietveld, Robert et al. (2020). "What You Feel, Is What You Like Influence of Message Appeals on Customer Engagement on Instagram". In: *Journal of Interactive Marketing* 49, pp. 20–53.

Rocha, B Carneiro da and R Timoteo de Sousa Junior (2010). "Identifying bank frauds using CRISP-DM and decision trees". In: *International Journal of Computer Science and Information Technology* 2.5, pp. 162–169.

Rohm, A., V. Kaltcheva, and G. Milne (2013). "A mixed-method approach to examining brand-consumer interactions driven by social media". In: *Journal of Research in Interactive Marketing* 7(4), pp. 295–311.

Rooderkerk, Robert P and Koen H Pauwels (2016). "No comment?! The drivers of reactions to online posts in professional groups". In: *Journal of interactive marketing* 35, pp. 1–15.

Rosenholtz, Ruth, Yuanzhen Li, and Lisa Nakano (2007). "Measuring visual clutter". In: *Journal of vision* 7.2, pp. 17–17.

Rother, Carsten, Vladimir Kolmogorov, and Andrew Blake (2004). “” GrabCut” interactive foreground extraction using iterated graph cuts”. In: *ACM transactions on graphics (TOG)* 23.3, pp. 309–314.

Russakovsky, Olga et al. (2015). “Imagenet large scale visual recognition challenge”. In: *International journal of computer vision* 115.3, pp. 211–252.

Russell, Stuart J and Peter Norvig (2016). *Artificial intelligence: a modern approach.* Malaysia; Pearson Education Limited,

Schindler, Pamela S (1986). “Color and contrast in magazine advertising”. In: *Psychology & Marketing* 3.2, pp. 69–78.

Schmitt, Philipp, Bernd Skiera, and Christophe Van den Bulte (2011). “Referral programs and customer value”. In: *Journal of Marketing* 75.1, pp. 46–59.

Scholte, H Steven et al. (2009). “Brain responses strongly correlate with Weibull image statistics when processing natural images”. In: *Journal of Vision* 9.4, pp. 29–29.

Scholte, H Steven et al. (2018). “Visual pathways from the perspective of cost functions and multi-task deep neural networks”. In: *Cortex* 98, pp. 249–261.

Scholz, Michael et al. (2018). “Dynamic effects of user-and marketer-generated content on consumer purchase behavior: Modeling the hierarchical structure of social media websites”. In: *Decision support systems* 113, pp. 43–55.

Schwenzow, Jasper et al. (2020). “Understanding videos at scale: How to extract insights for business research”. In: *Journal of Business Research* 123, pp. 367–379.

Selyukh, Alina (2018). *Optimized Prime: How AI And Anticipation Power Amazon’s 1-Hour Deliveries.* URL: `https://www.npr.org/2018/11/21/660168325/optimized-prime-how-ai-and-anticipation-power-amazons-1-hour-deliveries`.

Shafique, Umair and Haseeb Qaiser (2014). “A comparative study of data mining process models (KDD, CRISP-DM and SEMMA)”. In: *International Journal of Innovation and Scientific Research* 12.1, pp. 217–222.

Shalizi, C. and K. Shalizi (2004). “Blind Construction of Optimal Nonlinear Recursive Predictors for Discrete Sequences”. In: *CoRR* cs.LG/0406011. URL: `http://arxiv.org/abs/cs.LG/0406011`.

Shapiro, Linda and George C Stockman (2001). “Computer vision. 2001”. In: *ed: Prentice Hall.*

Shearer, Colin (2000). “The CRISP-DM model: the new blueprint for data mining”. In: *Journal of data warehousing* 5.4, pp. 13–22.

Shin, Donghyuk et al. (2019). “Enhancing Social Media Analysis with Visual Data Analytics: A Deep Learning Approach”. In: *Forthcoming at MIS Quarterly.*

Shmueli, Galit et al. (2017). *Data mining for business analytics: concepts, techniques, and applications in R*. John Wiley & Sons.

Simon, Herbert A (1972). "Complexity and the representation of patterned sequences of symbols." In: *Psychological review* 79.5, p. 369.

Simonyan, Karen and Andrew Zisserman (2014). "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556*.

Sirakaya, Ercan and Arch G Woodside (2005). "Building and testing theories of decision making by travellers". In: *Tourism management* 26.6, pp. 815–832.

Snodgrass, Joan G and Mary Vanderwart (1980). "A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity." In: *Journal of experimental psychology: Human learning and memory* 6.2, p. 174.

Solomonoff, Ray J (1964). "A formal theory of inductive inference. Part I". In: *Information and control* 7.1, pp. 1–22.

Sparks, Beverley A and Ying Wang (2014). "Natural and built photographic images: Preference, complexity, and recall". In: *Journal of Travel & Tourism Marketing* 31.7, pp. 868–883.

Stephen, Andrew T, Michael Sciandra, and Jeffrey Inman (2015). "Is it what you say or how you say it? How content characteristics affect consumer engagement with brands on Facebook". In:

Stoll, Marco, Sebastian Baecke, and Peter Kenning (2008). "What they see is what they get? An fMRI-study on neural correlates of attractive packaging". In: *Journal of Consumer Behaviour: An International Research Review* 7.4-5, pp. 342–359.

Thelwall, Mike et al. (2010). "Sentiment strength detection in short informal text". In: *J. Am. Soc. Inf. Sci. Technol.* 61.12, pp. 2544–2558.

Therneau, Terry, Beth Atkinson, and Brian Ripley (2015). *rpart: Recursive Partitioning and Regression Trees. R package version 4.1–10*.

Tirunillai, Seshadri and Gerard J Tellis (2014). "Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation". In: *Journal of Marketing Research* 51.4, pp. 463–479.

Tom, Gail et al. (1987). "Cueing the consumer: The role of salient cues in consumer perception". In: *Journal of consumer marketing* 4.2, pp. 23–27.

Trusov, Michael, Randolph E Bucklin, and Koen Pauwels (2009). "Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site". In: *Journal of Marketing* 73.5, pp. 90–102.

Tsai, Chih-Fong et al. (2020). "Improving text summarization of online hotel reviews with review helpfulness and sentiment". In: *Tourism Management* 80, p. 104122.

Tuch, Alexandre N et al. (2009). "Visual complexity of websites: Effects on users' experience, physiology, performance, and memory". In: *International journal of human-computer studies* 67.9, pp. 703–715.

Tucker, Gregory K and Mickey C Smith (1987). "Direct-to-consumer advertising: Effects of different formats of warning information disclosure on cognitive reactions of adults". In: *Journal of pharmaceutical marketing & Management* 2.1, pp. 27–41.

Valdez, Patricia and Albert Mehrabian (1994). "Effects of color on emotions." In: *Journal of experimental psychology: General* 123.4, p. 394.

Vantage, New (2020). *Big Data and AI executive survey*. URL: `http://newvantage.com/wp-content/uploads/2020/01/NewVantage-Partners-Big-Data-and-AI-Executive-Survey-2020-1.pdf`.

Varian, Hal R (2014). "Big data: New tricks for econometrics". In: *Journal of Economic Perspectives* 28.2, pp. 3–28.

Verbeek, Marno (2008). *A guide to modern econometrics*. John Wiley & Sons.

Viger, Fabien and Matthieu Latapy (2005). "Efficient and simple generation of random simple connected graphs with prescribed degree sequence". In: *Lecture Notes in Computer Science. Computing and Combinatorics*. Vol. 3595. Springer, pp. 440–449.

Vuong, Quang H (1989). "Likelihood ratio tests for model selection and non-nested hypotheses". In: *Econometrica: Journal of the econometric society*, pp. 307–333.

Wagner, Kurt (2018). *Instagram is limiting how much data some developers can collect from its API — and cutting off others altogether*. URL: `https://www.recode.net/2018/4/2/17189512/instagram-api%5C%5C-facebook-cambridge-analytica`.

Wallace, Gregory K (1992). "The JPEG still picture compression standard". In: *IEEE transactions on consumer electronics* 38.1, pp. xviii–xxxiv.

Wang, Tsai, Chia Tsai, and Ta Tang (2018). "Exploring Advertising Effectiveness of Tourist Hotels' Marketing Images Containing Nature and Performing Arts: An Eye-Tracking Analysis". In: *Sustainability* 10.9, p. 3038.

Wang, Yang, Alexander Chaudhry, and Amit Pazgal (2019). "Do Online Reviews Improve Product Quality? Evidence from Hotel Reviews on Travel Sites." In: *Evidence from Hotel Reviews on Travel Sites.(January 22, 2019)*.

Wedel, Michel and PK Kannan (2016). "Marketing analytics for data-rich environments". In: *Journal of Marketing* 80.6, pp. 97–121.

Wickham, Hadley et al. (2016). *dplyr: A Grammar of Data Manipulation. R package version 0.5. 0*.

Wilensky, Uri and William Rand (2015). *Introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo*. MIT Press.

Witten, Ian H et al. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Woodcock, N. et al. (2011). "The evolving data architecture of social customer relationship management". In: *Journal of Direct, Data and Digital Marketing Practice* 12(3), pp. 249–266.

Woolrich, Mark (2008). "Robust group analysis using outlier inference". In: *Neuroimage* 41.2, pp. 286–301.

Woolrich, Mark W et al. (2001). "Temporal autocorrelation in univariate linear modeling of FMRI data". In: *Neuroimage* 14.6, pp. 1370–1386.

Worsley, Keith J (2001). "Statistical analysis of activation images". In: *Functional MRI: An introduction to methods* 14.1, pp. 251–70.

Wu, Qiang et al. (2010). "Adapting boosting for information retrieval measures". In: *Information Retrieval* 13.3, pp. 254–270.

Wulf, Jochen et al. (2019). "A computational visual analysis of image design in social media car model communities". In:

Xie, Junyuan, Ross Girshick, and Ali Farhadi (2016). "Unsupervised deep embedding for clustering analysis". In: *International conference on machine learning*, pp. 478–487.

Yoganarasimhan, Hema (2020). "Search personalization using machine learning". In: *Management Science* 66.3, pp. 1045–1070.

Zhang, Mengxia and Lan Luo (2018). "Can user generated content predict restaurant survival: deep learning of yelp photos and reviews". In: *Available at SSRN 3108288*.

Zhang, Shunyuan et al. (2017). "How much is an image worth? Airbnb property demand estimation leveraging large scale image analytics". In: *Airbnb Property Demand Estimation Leveraging Large Scale Image Analytics (May 25, 2017)*.

Zhang, Shunyuan et al. (2019). *Can Lower-quality Images Lead to Greater Demand on AirBnB?* Tech. rep. Working Paper, Carnegie Mellon University.

Zhou, Bolei et al. (2017). "Places: A 10 million image database for scene recognition". In: *IEEE transactions on pattern analysis and machine intelligence* 40.6, pp. 1452–1464.

— (2018). "Places: A 10 million image database for scene recognition". In: *IEEE transactions on pattern analysis and machine intelligence* 40.6, pp. 1452–1464.

Summary

In today's online environments, such as social media platforms and e-commerce websites, consumers are overloaded with information and firms are competing for their attention. Most of the data on these platforms comes in the form of text, images, or other unstructured data sources. It is important to understand which information on company websites and social media platforms are enticing and/or likeable by consumers. The impact of online visual content, in particular, remains largely unknown. Finding the drivers behind likes and clicks can help (1) understand how consumers interact with the information that is presented to them and (2) leverage this knowledge to improve marketing content. The main goal of this dissertation is to learn more about why consumers like and click on visual content online. To reach this goal visual analytics are used for automatic extraction of relevant information from visual content. This information can then be related, at scale, to consumer and their decisions.

The results of four empirical studies are presented. The first empirical chapter highlights the managerial importance of visual analytics and AI. In addition, it provides the reader with the definitions and problem understanding necessary to appreciate the methods and tools presented in the rest of this dissertation. The next chapter consists of a theory-driven investigation of the relationship between the visual complexity of firm-generated imagery and consumer liking on social media. The third chapter utilizes a data-driven exploration approach to study the impact of product images on consumers decisions on e-commerce websites. The final empirical chapter serves as a look into the future of visual analytics for marketing.

Samenvatting

In hedendaagse online omgevingen, zoals social media platformen en e-commerce websites, worden consumenten overladen met informatie en concurreren bedrijven voor hun aandacht. Het meeste van de data op deze websites komt in de vorm van tekst, plaatjes, en andere ongestructureerde databronnen. Het is belangrijk om te begrijpen welke informatie op websites en social media platformen aantrekkelijk en "likeable" zijn voor consumenten. De impact van, met name visuele, online content is tot op heden vrijwel onbekend. Het uitvinden wat deze likes en kliks teweegbrengt helpt met (1) het begrijpen hoe consumenten omgaan met de informatie die aan hen gepresenteerd wordt en (2) hoe we deze kennis kunnen gebruiken om marketing content te verbeteren. Het doel van dit proefschrift is het begrijpen waarom consumenten visuele content liken en waarom ze er op klikken. Om dit doel te bereiken wordt gebruik gemaakt van visual analytics om automatisch de relevante informatie uit visuele content te halen. Deze informatie kan dan, op grote schaal, gerelateerd worden aan consumenten en hun beslisgedrag.

Het eerste hoofdstuk beschrijft het belang van visual analytics en AI voor managers. De definities en probleemstellingen die noodzakelijk zijn om de in dit proefschrift gebruikte methodes te begrijpen en te waarderen worden hier ook beschreven. Het volgende hoofdstuk bevat een theorie-gedreven onderzoek naar de relatie tussen de visuele complexiteit van fotos en likes door consumenten op social media. Het derde hoofdstuk gebruikt een data-gedreven benadering om de impact van online productfoto's op het beslisgedrag van consumenten op te bepalen. Het laatste hoofdstuk dient als een blik in de toekomst van visual analaytics voor marketing.

Acknowledgements

I would not have been able to do this PhD or write this dissertation without all the amazing people in my life. One of those amazing people is you, reader. Thank you for opening up this book and reading through it. I want you to know that you are an absolute legend! Now, to the other important people that made this journey of knowledge happen.

I see the start of my academic career somewhere half-way through the first semester of my master's in econometrics. In particular, during the machine learning course. During this course I discovered my passion for machine learning and its (endless) applications to business. This newly found passion led to the grade that allowed me to write my master's thesis on this topic. I wrote my first ever research project under the expert supervision of Prof. Marcel Worring. Marcel, thank you for introducing me to machine learning and inviting me to be part of an ongoing research project with Masoud Mazloom. It really marks the start of my academic career and it excites me that you are part of the dissertation committee as well. During the thesis, I spent most of my time working directly with Masoud. I am forever grateful to you, Masoud. You sparked my interest in doing research and very early in the thesis process you recommended I'd do a PhD. In fact, it was so early that it was almost weird, but here we are!

From the thesis, I almost "seemlessly" progressed into my PhD, which brings me to my next academic mentor, Willemijn. First of all, thank you for going on this adventure with me. This was by no means an ordinary PhD, because you not only created a PhD position for me, but you also gave me the freedom to do this from the United States. It certainly wasn't without challenges, but I think we can now conclude that we made it work, right? Thank you for your guidance and belief in me and thank you for all the meetings, both virtual and in-person. They always left me inspired and hopeful for what was to come!

My American adventure did not only turn out to be fruitful for my relationship with Barbara, but it also brought me Bill Rand. Bill, I don't think I can express how grateful I am for all that you have done for me. I guess the fact that, besides Barbara, you were the only person that was there for my engagement, wedding, and honeymoon about says it all. From our first lunch together in Raleigh (the one during which I somehow convinced you enough to sponsor my visa), until now you have been an inspiration. If people ever ask me what I want to be when I grow up, I'll say that I want to be just like Bill!

Doing research and writing papers can be lonely at times, especially when working remotely for most of it, but because of my fellow students and colleagues I never feel alone. Anthony, you were my first PhD buddy here in the US. I still remember the first time we met when you asked me from across the hall if I was Dr. Rand's new student. It immediately made me feel welcome, and it feels like the next day I was already crashing on your couch. I will always be grateful for your hospitability. You really pushed it when you had breakfast, coffee, and a snack ready for me to conquer the day. The last person that made me feel that way must have been my mom. I am also grateful for your excellent navigation skills, without them I would have never seen all those amazing sights in Hong Kong. The other PhD buddy I am thankful for is Leo. We followed the same path with our bachelor's and master's degrees in econometrics, but it wasn't until the PhD that we became good friends. I guess it took some time for us to realize that we had some shared interests. Hospitability seems like a quality that I seek out in my friends (or maybe I just look for couches and find friends as a result..), because that certainly applies to you as well. I truly enjoyed all the coffee, lunch, and beer breaks every time I came back to Amsterdam. It also means a lot to me that you visited us twice in the US, once in Charlotte to shoot a bunch of guns and once in Beaufort for the wedding. I look forward to celebrating our finishing of the PhDs next time I see you. The last PhD buddy I am grateful for is Bob. Especially during the beginning of my PhD you showed me the way and I always enjoyed our conversations and research discussions.

I am also thankful for everyone at the CADS lab. Collaborating with y'all was truly a pleasure and I'm excited for what the future brings for all you bright minds! Thank you to all my colleagues at the University of Amsterdam. Even though I only spent a few months out of the year at ABS you always made me feel part of the group. I am especially grateful for Jonne, Andrea, Ruta, and all others that helped me during the job market process. I appreciate the time and effort you all put into helping me prepare my packet and my research talks. Thank you to all co-authors I have worked with or I am still working with - Meg, Steven, Manuel, Rohan, Koen, Yakov, Anatoli, and Kwong. I'm learning so much from all of you. Finally, I want to express my gratitude to Mary Gros for inviting me to be part of the Academic community at Teradata and Teradata Universe. You have connected me with some of the brightest minds in both academics and industry, thank you!

To all my friends, thank you. I'm not going to name all of you, because you know who you are. I want you to know that I appreciate you. I'm humbled by all the love and support I receive from all of you. I am extremely lucky to have such an amazing group of friends. Thank you for the times you texted, facetimed, and hung out with me. A special thanks to all the people that provided me with a place to stay when visiting! Then, to Brendan, Christa, and the rest of Crossfit Meck, I want you to know that you have been imperative to my physical, and mostly my mental health. I probably talked your ears off before, during, and after every workout. Many of the roadblocks in my research I have conquered right before or after going to the gym. I will miss you guys and I can only hope to find such a great group of people to work out with in Rochester.

Finally, I want to express gratitude to my family. Sjoerd, thank you for just being the best friend I could ever wish for. Maaike, I'm lucky to have such an amazing big sister. Thank you for sharing with me how proud you are of me, and also for keeping me company in the car on my drive to the university. Thanks to Lisanne, Jessie, and Mila, for always welcoming me with open arms into your home. You are the best! Thank you to Robbin for always being there to offer support (or to ask me statistics questions). I will forever be indebted to all of you for letting me stay for weeks on end! Of course, eternal gratitude and thanks to my parents, Marty and Jolanda. Thank you for always believing in me and for instilling the confidence that I am capable of doing whatever I want! Mama, thank you for teaching me to be curious and to ask big questions, a useful quality for an academic. Papa, thank you for teaching me discipline. You have shown me the importance of showing up, every workday, every practice, every commitment - another useful quality for an academic. I guess it makes sense you both taught me these things since I expressed that I wanted to be a professor at the age of two. And guess what? I made it! Ok, I do still have to make it through the professor ranks, and yes I probably didn't have Marketing Professor in mind when I said it, but it still counts right?! Before I get to the last person I want to thank, I have to thank her parents. Manuel and Alida, thank you for all you do for us. I know you have both wondered what it really is that I'm doing and when I would finally get a real job and actually stay here, so here we are! Now, to the very last, but also the most important person to thank. My soulmate, Barbara. We made it! This PhD was intertwined with the adventure that you and I have been on. From dating long distance, to figuring out when I could finally get a visa to come live with you in the US, to visa extensions, to long trips away, to new visas and to this exciting new opportunity that is now finally in sight. You have been the best support system I could have ever wished for. Thank you for always pushing me to do the best I can and for going along with my weird habits and ridiculous schedule. Thank you for unconditional love and support. No matter what happens during my day, I can always expect the same hug at the end of it to forget (or celebrate) it all! I'm eternally grateful for this amazing adventure we are on and I'm excited for what the future holds for us. Next up, Rochester NY. I can only try to be as much of a support system to you now that you transition back into student life yourself. I promise I will do my best!

Humans are not the only important beings in life. Here's to you Tyrone. Thank you for being the best dog in the world!