The Champion of Images: Understanding the role of images in the decision-making process of online hotel bookings.

Gijs Overgoor
University of Amsterdam
g.overgoor@uva.nl

William Rand
NC State University
wmrand@ncsu.edu

Willemijn van Dolen
University of Amsterdam
w.m.vandolen@uva.nl

Abstract

Images are vitally important in interesting consumers and helping them to make decisions. Images of a hotel are particularly important and were used to sell hotels even before the Internet, when travel agencies would often have brochures about hotel properties that they used to entice travelers. On many online travel agency (OTA) websites, the hotel's image can take up 33% of the space on the hotel property page, but the importance of this image in the decision-making process has yet to be studied. For many OTAs, there are currently no quantitative analytic methods that help determine which image to display in this critical location. In this research, we use deep learning to extract information directly from hotel images and we apply image analytics to understand the importance of this information in the online hotel booking process. To provide managerial insights, we will combine a prediction model, with the t-distributed Stochastic Neighbor Embedding (t-SNE) to classify and understand the types of images hotels generally use as their thumbnail or "champion" image and what aspects of these images elicit consumers to consider and book a hotel.

1. Introduction

Millions of travelers world-wide visit online travel agencies (OTAs) to fulfill their travel needs. OTAs aid the traveler by helping them decide what flight to take, what hotel to book, and where to go for vacation. OTAs work with the hotels and airlines to provide the consumer with a plethora of travel options and information to consider when deciding which travel option to choose. This fast-growing marketplace constituted 600B USD in online travel sales in 2016 and is projected to grow to more than 800B USD in 2020 (Statista 2016). Online hotel bookings are a major part of this industry and currently capture 39% of the US online digital booking market (Travel Trends 2017). With such a large consumer demand and stiff competition in the hotel market, if a hotel property or an OTA on which it is hosted wants to be successful, it is essential to attract consumers quickly and to provide them with information and imagery that will facilitate their reservation. The image(s) of a hotel have always been part of the hotel listing, even before the Internet, when travel agencies would often have brochures about hotel properties that they used to entice travelers. On many OTA websites, the hotel's image can take up to 33% of the space on the hotel property page, but the importance of this image in the decision-making process has yet to be studied.

Previous research has shown that there are several key attributes that consumers consider for their purchase decisions of hotels including: price, room availability, hotel category, brand, amenities, location, and customer reviews [1, 2]. Researchers have also shown that the Internet enhances a consumer's efficiency in search and evaluation [3, 4]. Since consumers are scanning more and more properties in an efficient manner, it is important for any firm that wants to capture consumer attention to make sure that their images effectively do so. Due to advanced filtering capabilities, the online traveler is able to quickly evaluate potential candidate hotels through various parameters [5, 6]. Additionally, tracking consumer clicks, browsing, searches, and purchase behavior has shown to be useful in understanding the decision-making process [7]. Combining these various consumer signals can improve click-through rates (CTR) by optimizing search results [8]. Moreover, presenting advertisement banners that match an online consumers personality has been shown to increase conversions [9]. In a similar way, presenting consumers with the right image should increase clickthrough rates. Previous research has explored what aspects are important to online travelers when making a travel purchase decision and has examined how tracking customer behavior can be leveraged to optimize search results and presentation of information, but very little past research has explored

URI: https://hdl.handle.net/10125/64240
978-0-9981331-3-3

the role of images in travel choice selection.

When an online traveler is searching for a hotel, one of the first pieces of information they are presented with in the search results is the hotel image, a thumbnail that appears next to the information presented for the hotel. Given the power of visual information [10], it seems clear that this can have an important effect on the customer decision-making process. A study on the relationships between firm- and user-generated content on social media [11] shows that firm-generated content has the strongest effect on the consideration and purchase intent of consumers. This is exactly the point when the image of a hotel will play a vital role. The user is searching, which means that they are already aware of travel opportunities, so firm-generated content has the greatest possibility of affecting the consumer's decision to move from mere awareness to including the firm in the user's consideration set. A user's click on a hotel listing in the search result to find out more information about a hotel could be viewed as an inclusion of this hotel into the consideration set. An important piece of firm-generated content in this case is the hotel image and it potentially impacts whether or not a hotel will be included in the consideration set.

To date, there is very little knowledge about the impact of imagery in the online travel industry. A notable exception is a paper by Zhang et al. (2017) [12], in which the authors show that photography quality positively influences the demand of properties on AirBnB. Our work is very different than this previous work since we: (1) look at more traditional hotel listings, (2) examine the effect of the image on the consideration decision as opposed to the purchase decision, and (3) we explore not just the general quality of the image, but the exact underlying aspects that motivate a click by the consumer. This gives us the ability to not just make managerial suggestions about the role of images, but also to provide insight into what images managers should use to maximize clickthrough. To summarize, in this paper we will explore the role of the image in the transition from search result to hotel information page and we will investigate what aspects of images and what types of images generally perform the best for hotels in an online environment.

In OTA terminology, the image that is presented as a thumbnail next to the hotel information on the search results page and the first image displayed after clicking on the listing, is called the champion image. For many OTAs, there are currently no quantitative analytic methods that help determine which image to display in this location. In this paper we plan to use machine learning techniques to identify whether it is possible to better understand which images, and specifically which concepts present in those images, are more likely to generate a higher CTR. We will do this by developing an Image Score, based on image analytics, for each image that will predict how likely that image is to generate user interest. In addition, we use an advanced embedding method that maps the high-dimensional features from our neural networks onto a two dimensional space that allows a quick evaluation of what types of images in different locations or contexts. This enables the development of managerially relevant explanations for why some images perform better than others.

Our study makes several important contributions:

- This is the first paper that investigates the impact of images on the consumer decision-making process for online hotel bookings at a prominent OTA.
- We use advanced visual analytic methods to explore the aspects of images that drive the inclusion of hotels into the consideration set of consumers.
- Our method is not just a black-box, unstructured prediction, but instead provides interpretable information that hotel managers can use to decide what images to use as their "champion" image. We provide a managerially relevant argument as to why certain images do better than others.
- We provide a structured, unified method for mapping of images that are generally used by hotels, which also shows how well these images perform. This unsupervised clustering system captures the heterogeneity across images and provides insights into what general features of images do better than other images, and can take the context of the hotel into account.

2. Framework

Our framework is based around the use of image analytics. Image analytics is the automatic extraction of structured data from images using computational methods [13]. We extract several features automatically from the images of the hotels. These features cover a wide range of aspects of images, from the visual properties to the semantic content. Using convolutional neural network (CNN) architectures, we will extract visual features and then classify the hotel images on the basis of those features. Specifically we use two deep neural network structures to extract these features: (1) ResNet50 to classify 1000 ImageNet object categories that are present in the image [14] and (2)

Places365-ResNet to classify 365 scene categories in the images [15].

2.1. Image Classification

Recent advances 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 [10, 18, 19]. In this paper, we use CNNs to extract conceptual information from the hotel images that we can then build upon to find the relationship between concepts in those images and consumer engagement. 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.

Part of the dramatic increase in image processing capability is derived from the convolutional aspect of CNNs. The first CNN, LeNet5, was developed by LeCun et al. [20]. 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 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.

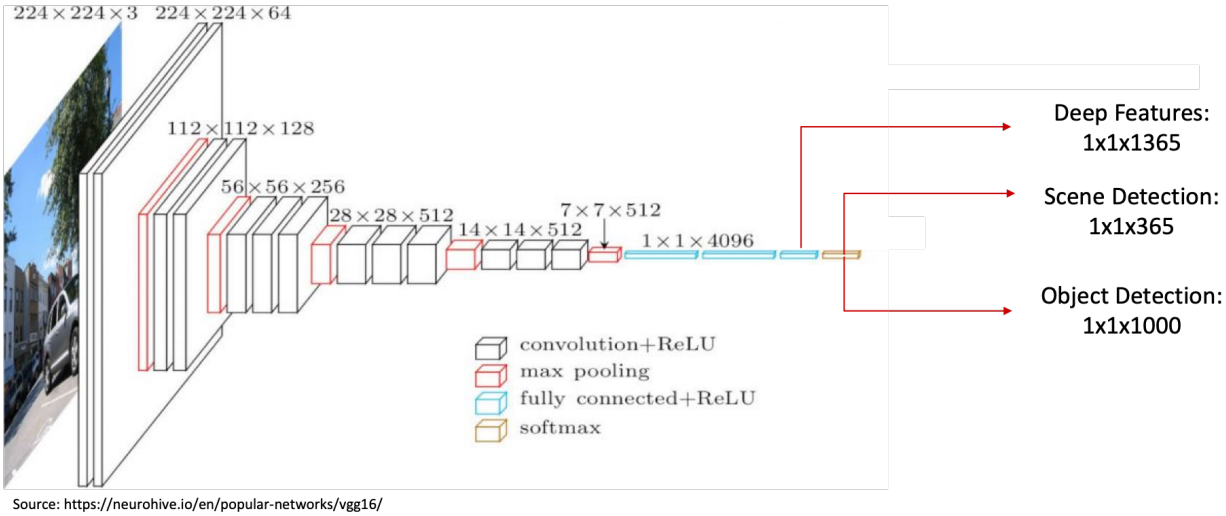
For our application, we use two pre-trained CNNs to identify objects and scenes. In addition, we use a novel hybrid of the two networks to extract our deep features. A visual representation of the VGG16 architecture and the feature dimensions can be found in Figure 1.

The automatic identification of objects in images has received considerable academic research attention since the start of the ImageNet Large Scale Visual Recognition Challenge [17]. 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 make use of one of the CNNs that did particularly well in the ImageNet challenege. In particular, we use the pre-trained CNN proposed and developed by He et al. [14], which won the ImageNet challenge. The 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[1]. 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 in this research, The final result is a distributional representation of all of the objects present in every hotel image in our dataset.

For the scene classification we use a deep neural structure trained on previous images of different locations, called the Places Database [15]. 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 the presence of each of the 365 scene categories in the image. The final result is a distributional representation of the presence of scenes for every hotel image in our dataset.

Previous work in social media analysis [10, 19] show that the activation output of deeper layers of a CNN are useful for popularity prediction. The way this is done is by creating a model that relates this output layer to the popularity of a set of known social media images. The output layer represents raw image information that has not (yet) been translated into a meaningful prediction, yet it has been structurally processed. We use the last fully connected layer of a novel hybrid model trained to recognize both objects and scenes (see Figure 1). 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

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



Source: https://neurohive.io/en/popular-networks/vgg16/

Figure 1. The VGG16 architecture [16] 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 [17] and places 365 [15].

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 clickthrough rates. 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 clickthrough rates than predictions made directly from the final concept classifications. In this paper, we use theses "deep features" to make predictions of clickthrough rate. 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 was 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.

3. Empirical Application

3.1. Data

Our data consist of a very 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 search request. Following the request, which includes several parameters (e.g., destination city, travel dates, number of travelers), the website presents the consumer with an ordering of available hotels in the city on a search results page. Every hotel listing on the search result page consists of the name of hotel, thumbnail or "champion" image, price (with potential discount), number of stars, average reviews. In addition to the standard information that is the provided for every hotel, there is information about deals or specialties unique for that particular hotel or search result. These items of special information, such as "free breakfast", "reserve now pay later", or "only 1 left at this price", are often colorfully presented with visual cues such as banners or highlighted text. After obtaining the default set of results, consumers can click on a hotel on that page, continue to the next page of results, or use the sort/filter functionality to refine their results based on hotel characteristics.

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. This resulted in 3.4M queries. These queries resulted in a search result page, where the hotels are listed, and potentially, a hotel info page, where one particular hotel and all of its information are presented. Not all consumers necessarily clickthrough to an underlying hotel, but the consumer can get to the hotel info page from the search result page by clicking on the image or the name of the hotel. This is 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.

At the consumer level we can observe the search criteria, search results, the clicks, view duration and other behavioral data on the website. After scanning

through all of the data, we observe that an online traveler clicks 4-5 times on average and that 0.7% of the online visits result in an actual booking. Another important observation that highlights the importance of images to the consumer decision process is that about 35% of the visitors perform some kind of more detailed action directly related to an interest in the hotel images, which includes clicking directly on the thumbnail image or browsing the photo gallery.

At the hotel level we obtain all of the characteristics described above, such as the price, rating, and features, the champion image, other images of the hotel, and the overall performance of a hotel listing in terms of clicks.

3.2. Method

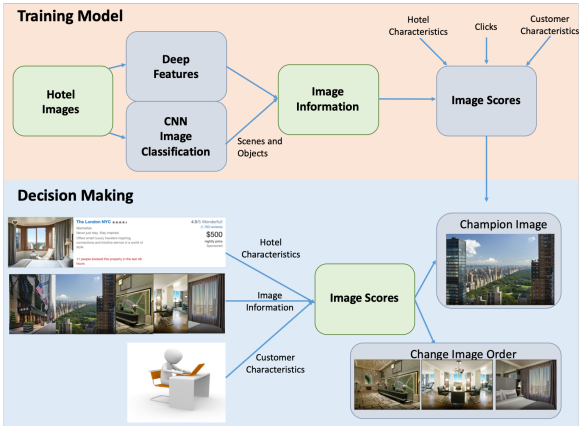


Figure 2. Image Scoring Method.

One goal of this research is the creation of an image score for each image in the hotel database. To create this score, we will combine the features described in the previous section. This image score can then be used for future images to help decide which images should be shown to consumers searching for a hotel to maximize CTR. After training and optimizing our model we can then use it to construct image scores for new images. Subsequently, we can rank the new images for each hotel based on their image score; choose the champion image based on this information, and then display the images with the highest scores first. Figure 2 illustrates the process of using historic information to train a model and use the image information, hotel- and customer characteristics to score the images, choose a champion image and re-order the other hotel images. We will test this new image scoring method on historical data to see how accurately the image score predicts which hotels users click on. We use the extracted features as an input for a support vector regression to make this prediction.

3.3. Predictive 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. 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 to features extracted using: (1) the *Places365* neural architecture, (2) the *Objects* detection neural architecture, (3) our *Deep Features* neural architecture that uses the 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:

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

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:

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

which can be formulated as

$$\arg\max_w \sum_{i=1}^{e} \log p(y_i|F_i, w) + \lambda||w_k||^2 \tag{3}$$

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 [21]. 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.

3.4. Explanatory Model

Although 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, it does not really allow for interpretation of the underlying structure of the data and does not control for differences between locations. These 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. Traditionally, a popular method for dimension reduction is Principle Component Analysis [22], but this linear method is not able to handle complex non-linear data. In particular, PCA and other linear techniques are not capable of retaining the local structure of the data while also preserving some of the information from the global structure of the data in a single map.

For this reason, we use an embedding algorithm that maps high-dimensional data onto a two dimensional space, called the t-distributed Stochastic Neighbor Embedding (t-SNE), which is popularly used to analyze image data in a constructive manner [23]. 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. Each of the clusters of images in the map will show the different types of images that are used, such as pool, view, lobby or hotel room. It could also reveal other underlying aspects that make these image similar, such as shared color or brightness. This visualization enables the capture of commonly occurring elements as well as the heterogeneity across locations, i.e., it may well be the case that pictures of pools do better in Miami, while pictures of skyscrapers do better in New York City. We can then take the output of t-SNE to graphically represent the images on a two-dimensional space while highlighting the best performing images, and controlling for location. This will help us to better understand what aspects of the images do better than others.

4. Results

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.

Table 1. Experiment 1 Results

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

The results in Table 1 show the prediction accuracy of using the extracted image information from the pretrained 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 fact that this model does poorly compared to the other models, suggests that in future work it might be useful to build a model that was used to classify concepts in images related to hotels only. This might improve the accuracy of all of these models. The places model and the Objects model would be improved since they would be trained on relevant images, and the deep features model would potentially provide better structural information. Moreover, the combination model would be improved since it is composed of these three other models.

Now that we have a model that can predict the performance hotels based on just the image information we want to know how and why certain images are doing well. In order to evaluate the correlation of objects with the performance of the hotel images, 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 clickthrough rates. The weights of the support vector regression with the distributional representation of the scenes as input shows 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.

This provides us with key insights into the best performing champion images in several ways. First, it shows what works well in general seems to be either the hotel room or the front of the hotel. Given that users are probably concerned with the appearance of the hotel and the room that they will be staying in, it is not a surprise that generally buildings or a view work well. 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.

To further investigate what types of images are used by hotels and what works well, especially for the different locations we use a t-SNE mapping that includes a performance indicator. Figure 3 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 3 shows that there are a few types of images that are generally used by hotels already: 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, as discussed above this could potentially be because we are aggregating across all five locations, and, in fact, when we map the images separately across the destinations we see that there are definitely types of images that work best depending on the particular city where they are used. Figure 4 shows the mapping for the 5 different cities in our dataset: Boston, Miami, New York City, San Francisco and Seattle.

We observe in Figure 4 that for each of the 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. It is interesting that for these two destinations 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. For Miami (upper middle), we observe from the mapping of the images that it is much more common to use the pool as the thumbnail image than for the other locations. The best performing images are a bit more scattered here than for the other destinations, but we observe that the pool, the bar, and the outside of the hotel are all included in the best performing images. Lastly, for Seattle (bottom right) we observe that the images containing a view from the hotel room and the hotel room itself work best. In addition, the indoor pools of the hotels are used quite frequently as well. Potentially because Seattle is known for its rainy weather and so indoor activities are important, but it is also know for its beautiful views so the views from the rooms are important as well.

5. Discussion

In this project, we use image analytics and artificial intelligence to understand the role of images in the decision-making process of consumers booking hotels online. The image scoring framework that we have developed can be used for future firm-generated images to help OTAs decide which images they should show to consumers. Using our framework we are able to explore not just which images do particularly well at maximizing clickthrough rates, but also which concepts and scenes are also vital to engage the consumer. This gives us the ability to make recommendations to managers about how to design future images for their property. We were then able to draw relationships between the different aspects of these 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 that we provided 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 generally work well in Boston and San Francisco.

Though we have done the best to examine the relationship between champion 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 analyzing the resultant model and through the use of t-SNE we have developed new hypotheses as to why particular images do well in maximizing clickthrough rates. These hypotheses could be tested in future causal work.

In fact, in future work, we hope to carry out experiments, both in labs and in the field, that can help to substantiate the causal relationship between these concepts and clicks. This could be done for instance, by using A/B tests to show different users on an OTA website different champion images. We would argue that images that maximize our image score metric will receive more clicks and thus show that our model can be used to develop a prescriptive analytic that moves beyond telling firms which images are doing well, and instead tells them which images will do well in the future. Moreover, we hope to use additional image analytics, such as visual complexity [24], to further explore the relationship between image attributes and the decision by consumers to click.

In addition, as mentioned above, we have used off-the-shelf neural net architectures, even if we have combined them in novel and interesting ways. In the future, we hope to develop image analytic architectures that are specific to the travel industry, which should increase the accuracy of our models. Even with our current models we are able to discuss how different factors affect the consumer's decision-making process, and we have explored some of those factors, but additional work is needed to determine exactly what concepts are driving these decisions, and to look at how the individual city context affects this process. For instance, we present some hypotheses about how Boston is different from Seattle and so different images perform better in each case. One possible explanation for this is self-congruity [25], where images do better because they are what the consumer expects of that location; we will explore this theoretical structure and others in the future to help inform these relationships.

In general, we have presented one of the first image analytic frameworks that can be used at large-scale to help online travel firms decide which images to show consumers, and our initial results show that this framework could result in a substantial increase in clickthrough to the hotel page, providing firms with an advantage in an fast-growing and highly competitive market.

References

[1] K. Kaldis and E. Kaldis, "Emmantina and palmira beach hotels: Distribution for independent hotels," *EggerR. BuhalisD.(Eds.), eTourism: Case studies*, pp. 65–73, 2008.

[2] M. D. Musante, D. C. Bojanic, and J. Zhang, "An evaluation of hotel website attribute utilization and effectiveness by hotel class," *Journal of Vacation Marketing*, vol. 15, no. 3, pp. 203–215, 2009.

[3] E. Christou and P. Kassianidis, "Consumer's perceptions and adoption of online buying for travel products," *Journal of Travel & Tourism Marketing*, vol. 12, no. 4, pp. 93–107, 2002.

[4] S.-I. A. So and A. M. Morrison, "Destination marketing organizations'web site users and nonusers: A comparison of actual visits and revisit intentions," *Information Technology & Tourism*, vol. 6, no. 2, pp. 129–139, 2003.

[5] S. H. Jun, C. A. Vogt, and K. J. MacKay, "Online information search strategies: A focus on flights and accommodations," *Journal of Travel & Tourism Marketing*, vol. 27, no. 6, pp. 579–595, 2010.

[6] E. Varkaris and B. Neuhofer, "The influence of social media on the consumers hotel decision journey," *Journal of Hospitality and Tourism Technology*, vol. 8, no. 1, pp. 101–118, 2017.

[7] A. Ghose, P. G. Ipeirotis, and B. Li, "Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content," *Marketing Science*, vol. 31, no. 3, pp. 493–520, 2012.

[8] B. De los Santos and S. Koulayev, "Optimizing click-through in online rankings with endogenous search refinement," *Marketing Science*, vol. 36, no. 4, pp. 542–564, 2017.

[9] G. L. Urban, G. Liberali, E. MacDonald, R. Bordley, and J. R. Hauser, "Morphing banner advertising," *Marketing Science*, vol. 33, no. 1, pp. 27–46, 2013.

[10] A. Khosla, A. Das Sarma, and R. Hamid, "What makes an image popular?," in *Proceedings of the 23rd international conference on World wide web*, pp. 867–876, ACM, 2014.

[11] A. Colicev, A. Kumar, and P. O'Connor, "Modeling the relationship between firm and user generated content and the stages of the marketing funnel," *International Journal of Research in Marketing*, vol. 36, no. 1, pp. 100–116, 2019.

[12] S. Zhang, D. Lee, P. V. Singh, and K. Srinivasan, "How much is an image worth? airbnb property demand estimation leveraging large scale image analytics," 2017.

[13] C. Solomon and T. Breckon, *Fundamentals of Digital Image Processing: A practical approach with examples in Matlab*. John Wiley & Sons, 2011.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[15] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 6, pp. 1452–1464, 2017.

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[17] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[18] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, pp. 844–848, IEEE, 2014.

[19] M. Mazloom, R. Rietveld, S. Rudinac, M. Worring, and W. Van Dolen, "Multimodal popularity prediction of brand-related social media posts," in *Proceedings of the 24th ACM international conference on Multimedia*, pp. 197–201, ACM, 2016.

[20] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, *et al.*, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[21] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *JMLR*, vol. 9, pp. 1871–1874, 2008.

[22] K. Pearson, "Liii. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[23] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[24] P. Machado, J. Romero, M. Nadal, A. Santos, J. Correia, and A. Carballal, "Computerized measures of visual complexity," *Acta psychologica*, vol. 160, pp. 43–57, 2015.

[25] M. J. Sirgy and C. Su, "Destination image, self-congruity, and travel behavior: Toward an integrative model," *Journal of Travel Research*, vol. 38, no. 4, pp. 340–352, 2000.

Figure 3. 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.

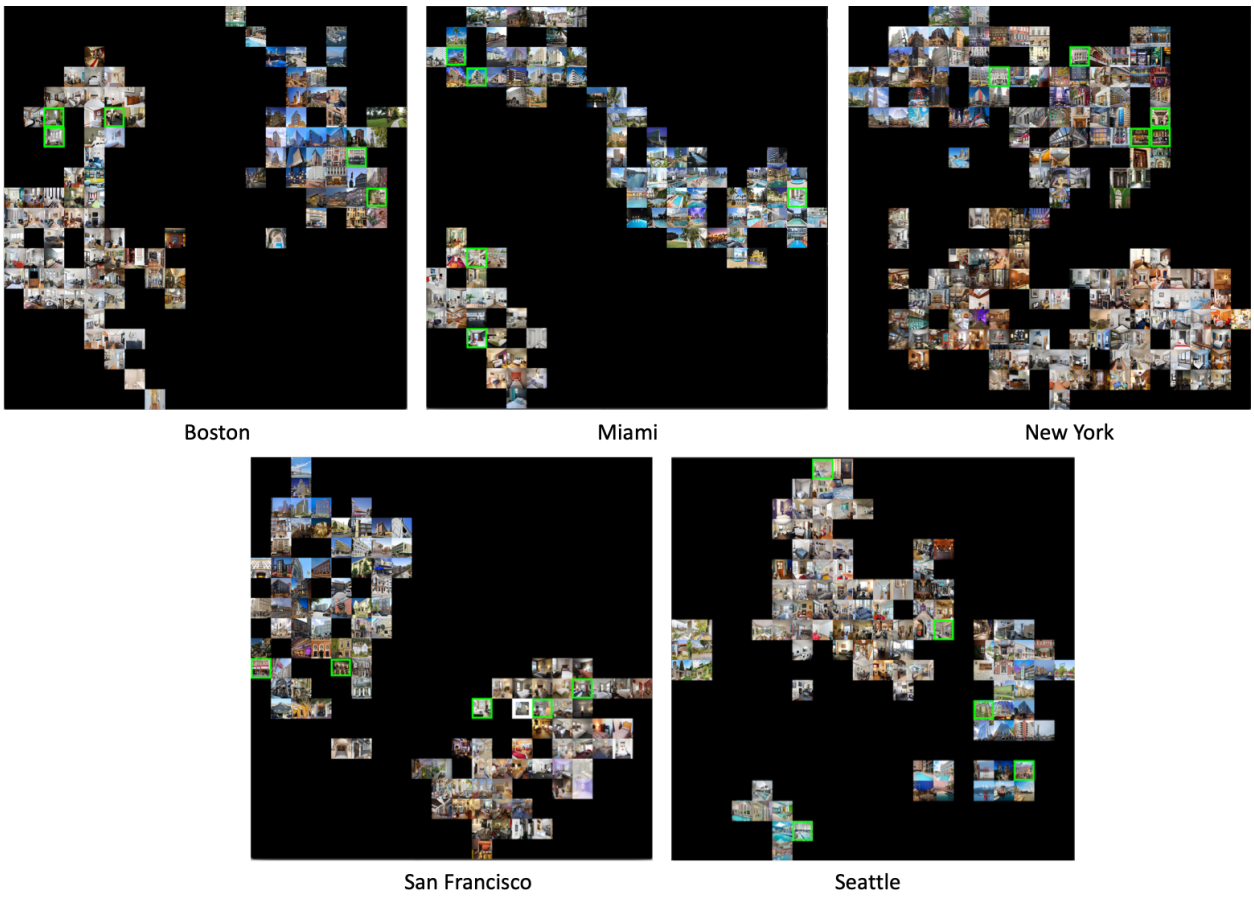


Figure 4. The t-SNE visualization of the hotel images in the 5 different locations. The green squares indicate the five best performing images of each location.