NOVA IMS Information Management School

MDSAA

Master Degree Program in

Data Science and Advanced Analytics

DATA-DRIVEN MARKETING FOR THE E-COMMERCE OF BRANDS

Ema Mandura

Work Project

presented as a partial requirement for obtaining the Master's Degree Program in Data Science and Advanced Analytics

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

DATA-DRIVEN MARKETING FOR THE E-COMMERCE OF BRANDS

by

Ema Mandura

Project Work report presented as a partial requirement for obtaining the Master's degree in Advanced Analytics, with a Specialization in Business Analytics

Co Supervisors*:* Nuno Antonio and Paulo Rita

February 2023

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism, any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honour from the NOVA Information Management School.

27 February 2023

DEDICATION

To my mum, for forcing me to finish my thesis. And to my dad, for telling me I do not have to.

ABSTRACT

The topic of this project is a data-driven marketing strategy for an e-commerce shoe brand Lovidovi Shoes. The company wanted to improve their digital marketing performance, and an improvement was defined as an increase in sales. While general best practices for successful Facebook Business advertising have been researched, each business is unique and optimal results are achieved through internal research. Data collected over the course of seven years was first centralized, and then analysed, using tools such as Power BI and Python, in order to determine the best audience and ad settings. The findings made on base data were re-evaluated and fine-tuned through testing. The final result showed that the best performing ads target an audience of women of all ages. Bosnia and Herzegovina makes the brands most profitable market, with the biggest number of sales, and the lowest cost per purchase. The feed placement on Facebook and Instagram get the best reaction, and the bestsellers are products in the white sneakers category. The ads created as part of this project showed significantly better performance by the company's standards, and average performance by the industry's standards. This project designed a simple guide on how to start making a shift towards data-driven marketing approaches with a limited budget, and has given the company motivation to utilize its data more and better.

KEYWORDS

Digital Marketing; Data Analysis; E-commerce; Customer-centric;

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

B2B	Business to business
B2C	Business to consumer
C2C	Consumer to consumer
CPR	Cost per result
CTR	Click through rate
ETL	Extract, transform, and load
GDPR	General Data Protection Regulation
SEO	Search engine optimization

1. INTRODUCTION

With vast amounts of data now available, companies in almost every industry, like banking and insurance, focus on exploiting data for competitive advantage against other companies. We live in the era of Big Data and the volume and variety of data have far outstripped the capacity of manual analysis, and in some cases have exceeded the capacity of conventional databases, requiring each time more processing power (Foster Provost, 2013). At the same time, computers have become far more powerful, networking is ubiquitous, and algorithms have been developed to connect datasets to enable broader and deeper analyses (Foster Provost, 2013) leading companies to turn their heads to Data Science and its unlimited potentialities.

As digitalization takes over society, more and more data is produced every day, and thanks to developing technologies, more significant percentages of the same data get appropriately stored. The increasing availability of data has inspired businesses across all industries to invest in data analysis and focus on exploiting data for competitive advantage. As companies look for ways to utilize the data, they simultaneously face issues regarding change acceptance within their organizations (Andrejevic, 2014). As new approaches are introduced, the need for training and recruitment of new professionals in the data science field arises.

As most successful companies shift from traditional to more data-driven decision making, younger companies are gaining advantage from building digital systems, rather than redesigning existing ones. One of those companies is Lovidovi Shoes, an e-commerce brand founded in 2015. The brand is a sustainable, slow fashion manufacturer and seller, specialized in leather footwear and bags. Lovidovi products retail price ranges between 100 and 200 euros, with a part of revenue made with every sale donated to the SOS Children's Villages of Bosnia and Herzegovina.

With the increased popularity of online shopping, traditional marketing methods are being replaced with customer-centric digital marketing. While this digital transformation allows for a better customer experience, it also challenges everything we know about good marketing practices (Cicerchia, 2021).

Lovidovi Shoes uses four different e-commerce shops and social media messaging as its sales channels. The four main markets, geographically, are Bosnia and Herzegovina, Croatia, Serbia and the European Union, focusing on Germany. The company's marketing is limited to digital marketing, with Facebook and Instagram paid advertising being its primary medium. In addition, e-mail marketing, influencer marketing and native Tik Tok marketing is also used.

As all conversions occur online, Lovidovi Shoes has a good channel for collecting accurate customer data, which has accumulated over seven years. The task of utilizing the available

data to improve the marketing practices was given to the company's marketing manager. The six month remodelling process will be reviewed through this project report.

The project is roughly divided into three phases, similarly to Figure 1.

1. Data analysis – in this phase, the existing data was analysed for patterns and valuable information.
2. Trial and error – in the second phase, different marketing campaigns were carried out, based on the results from phase one. Additional relevant data was collected. The data was analysed both scientifically and empirically conclude marketing approaches.
3. Testing – in the final stage, the winning approaches from the trial and error phase are applied again to prove or disprove their validity.

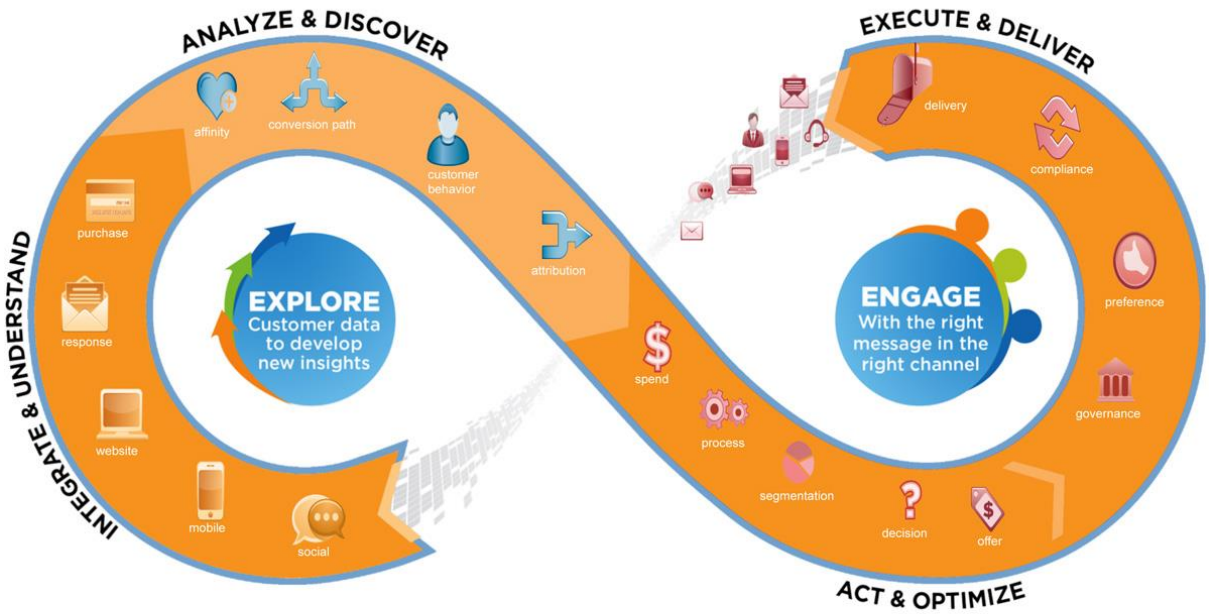

Figure 1 Data Driven Approach to Marketing (Sankaran, 2016)

The project's goal was to find the most profitable strategy for the brand, in terms of its target audience and marketing content. The results should be increased satisfaction of existing customers and acquisition of new customers.

As the company's marketing strategies were primarily based on gut feeling, and successful campaigns were achieved through trial and error, it was the most convenient time to start implementing data-driven decision making in marketing. With seven years of operation over four different markets, Lovidovi Shoes has collected enough data to analyse its customer network and patterns.

While much research is done on how to best the best utilizing digital marketing and data-driven techniques, it is hardly a one-size-fits-all situation (Khorev, 2022). As new small businesses are founded daily, brand-specific solutions are always needed. Data-driven approach enables that – tailoring digital marketing based on the business data.

The idea behind the project was to use existing data to understand customer behaviour better, then use the data to make better marketing decisions. Within the project's scope, different campaigns were planned to run, based on the insights from the existing data. The best performing campaigns were to be further analysed from these campaigns as their concept was to be used again in the future. From the top campaigns, a hypothesis was drawn about what kind of campaign is the most successful, regarding digital channels, products, and audiences and content. A new round of campaigns was to be ran to test the hypothesis. This project aimed to improve the overall performance of digital marketing campaigns, with the number of sales as the main metric. The cost per purchase is also considered, with the best result having to fit within the profitability price range set as part of the project.

However, another business problem was identified through this project. Due to the company's structural organization, data collection was not done strategically. The available data is decentralized and automatically through different business processes. Decentralized data caused the creation of data silos – disconnected data collected from different sources and stored in different formats. To approach this project efficiently, data integration was necessary first.

2. LITERATURE REVIEW

2.1. E-COMMERCE OF BRANDS

E-commerce, by its original definition, is buying and selling products through electronic measures. However, when e-commerce is talked about, it is usually with transactions over the internet in mind (Patel, 2017).

The three types of e-commerce are business-to-business (B2B), business-to-consumer (B2C), and consumer-to-consumer (C2C). B2B stands for transactions between business, where the end consumer is not a part of it. It can refer to either wholesalers buying from manufacturers, or retailers buying from wholesalers, but it almost always involves purchases of larger quantities meant for selling and not personal use (Chen, 2022). The B2C model explaines shopping as we know it – customers buying from businesses. This model tradionally referred to storefront shopping, but with the intorduction of e-commerce, it definiton became more tied to internet shopping. One of the advantages of online B2C businesses is the exlusion of the middleman. Businesses, either manufacturers or wholesalers can now sell directly to the consumer, ensuring better prices for the consumer (Kenton, 2023). C2C represents the most direct business model, in which consumers sell to other consumers directly over third-party online platforms. It can be looked at as a garage sale that no longer needs a physical location, but can be done over e-commerce platforms such as eBay, Etsy, and Craigslist (Tarver, 2022).

E-commerce has gained popularity due to its convinience. While every brick-and-mortar store has limits to what and how much of it can be sold, the internet can hold great amounts of different products. Even though they might be shopping on different online shops, customers still get the feeling of getting everything in just one store. Another perk of buying online is that the internet has no business hours. As long as there is an internet connection, the consumer can make a purchase at any time of day. The bloom of e-commerce businesses is also a strong indicator of globalization, as there is now a way to sell worldwide without opening stores or registering companies in different countries (Patel, 2017). It helps both sellers by lowering their costs of operation and buyers by allowing them to shop from home.

2.2. DIGITAL MARKETING

Digital marketing, in broader terms, refers to all forms of advertising delivered over digital channels. The digital channels are search engines, websites, social media, email and mobile apps (Dholakia, 2015). As more businesses move online, the growth in popularity of digital marketing becomes obvious, but it is almost equally relevant for bricks and mortar businesses, as most of humanity relies on the Internet for product and service research, if not its purchase.

Modern digital marketing consist of a network of different channels and marketers are challenged to find the best strategies in order to achieve engagement from their target audiences. The importance of omnichannel solid strategies is viewed best through a visible increase in retention. Companies with solid omnichannel presence have an average 89% retention rate, compared to other companies, which average at 33% (Invesp, 2016).

Daily, the average customer jumps over various digital channels, leaving a trail of breadcrumbs for the marketer to follow in their search for optimized marketing strategies. The most

common advertising approaches, explained below - are paid search, search engine optimization, content marketing, social media marketing, e-mail, mobile and digital marketing.

Paid search is advertising on top or side of a search engine results page. As 93% of all online interactions start through a search engine, this is a powerful medium (Laurinavicius, 2022). The advertiser is charged per click fabricated by the ad, and the ad is shown to the relevant audience based on an individual's online behaviour and what each person is searching for. This ensures that the person looking for socks will not be shown an ad for children diapers, which decreases the possibility of creating negative brand awareness.

Search engine optimization (SEO) is a tool that allows a brand to achieve the results of paid search without paying. Using relevant keywords and phrases, companies can improve their website's rank in search results, which increases organic site traffic. According to (ZLW, 2021), 67% of clicks go to the first five search results – meaning that the higher up a website is on the search list, the less likely a customer is to go to their competition.

Content marketing is based on building customer trust through relevant content sharing. If a company provides information the customer is looking for, without promoting their product, the customer is more likely to react its advertising efforts promoting its product. This type of marketing generates three times the leads gathered from paid search advertising (ZLW, 2021).

Social media marketing is an ever-changing channel, as it moves through different social media apps as they gain and lose popularity. It is mostly used for content distribution, giving most options for peer-to-peer sharing. Social media content sharing is our era's word of mouth, the most reliable marketing tactic traditionally (Hubspot, 2022).

E-mail marketing is the best channel for contacting customers and prospective customers directly. In order to be successful, such marketing must be engaging, relevant, informative and entertaining; otherwise, it loses its value in the receiver's junk mail (Hubspot, 2022).

Mobile marketing is the most intimate form of advertising, and must be approached cautiously. Advertising through SMS can almost guarantee that the message will reach the customer, but the customer will not be pleased if they feel like their personal space was invaded to deliver irrelevant information (Kenton, 2022).

Digital marketing will expand as new apps, websites and devices are introduced. The appeal of digital marketing in its beginnings was the low cost compared to traditional mediums. As more marketers joined the network, both demand and prices were raised. However, with an estimated 378 billion U.S. dollars spent on digital advertising worldwide in 2020, digital is becoming the norm of marketing (Statista, 2021). With the introduction of data-driven marketing, digital advertising is again redefined, and questions are raised over for how much longer it can serve marketers rather than the customers.

2.3. DATA-DRIVEN MARKETING

Data-driven marketing is the biggest game changer in the marketing industry. It is the whole process of understanding existing and potential data and how to organize, analyse, and apply it to improve marketing outcomes. In such an approach, marketers build strategies based on the insights gained from data analysis, intending to predict customer behaviour from their previous engagement (Higuera, 2021). Data-driven marketing solutions help companies deliver the right content to the right people, at the right time, through a suitable medium, thus significantly improving the chances of forming a beneficial relationship with the customer (Lennon, 2017).

A significant shift is made with the introduction of data into marketing decision-making. The focus of marketing now changes from the product to the customer (Camilleri, 2019). What is implied product-based marketing should first be revisited to explain customer-centric marketing better. Product-based marketing is the main principle of traditional marketing, and it is founded on a simple logic – the marketer identifies a customer's need and then attempts to sell it to them. If a customer is looking for a product, they will be targeted with ads for that and similar products (Camilleri, 2019). In a product-based approach, every customer looking for the same product will get the same offer. When it comes to customer-based strategies, more questions are to be considered. Has the customer looking for the product made a purchase before? Have they looked for the product already and did not proceed with the purchase? Are they long-time loyal customers? With questions like these in mind, should all those customers be given the same offer, or should it be communicated to them the same way? Enhancing and personalizing the customer experience is one of customer-centric marketing's main premises. To provide the best personalized experience, marketers should first understand the relationship between the customer, product and the brand (Camilleri, 2019). As one of the most common approaches, customer segments are created based on customer profiles derived from their demographic and behavioural data, and different segments are targeted with different promotional content (Stelman, 2020). The premise of customer centred model is visualized in Figure 2.

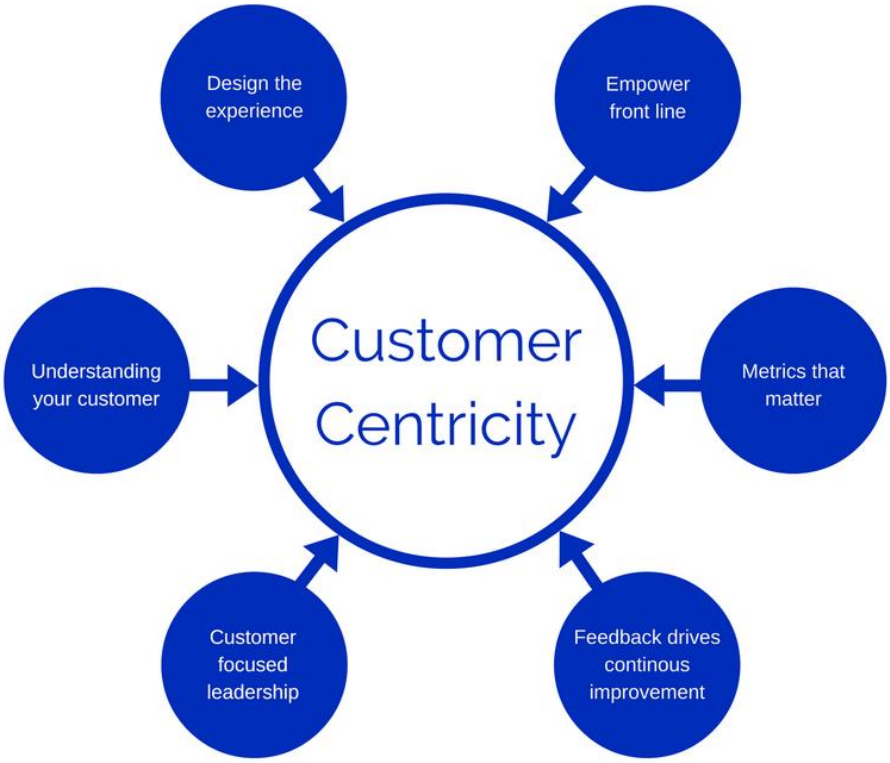

Figure 2 The Customer Edge Customer Centricity Model (L2M Rail, 2021)

For example, a group of customers who added products to their cart but abandoned the checkout might all get a special discount to encourage them to complete a purchase. On the other hand, a group of customers who have not purchased in a long time might get an offer of the company's bestsellers, in order to remind them why they made purchases in the past and drive them to make a new purchase. Companies often utilize their understanding of human nature and knowledge about their customers' behaviour to decide on their tactics, whether it comes to setting prices, drafting contracts, minimizing risk perceptions or extracting additional information (Kiron, 2012).

However, it is not only businesses that benefit from these advertising techniques. Using data-driven marketing generates profit for the company, but also serves the customer by providing timely and relevant offers, and the society by creating new work positions within the Data Science field (Milheiro, 2019). With the increased popularity of these technologies, the consumers are also becoming aware of their role in the chain. They might prevent marketers from showing them irrelevant ads by blocking unwanted content. On the other hand, they might notice the correlation between their actions and the offers they get shown and take advantage of that knowledge to get the best deals for themselves. For example, a customer might become aware that after they leave a product in their cart for some time, they will get a discount by e-mail for the same product (Camilleri, 2019).

Data-driven marketing aims to convert the collected data into revenue and acquire loyal customers. Technology used in these strategies is advanced enough to allow business to reach more people with more offers. However, it must be used reasonably, as it might result in a contra effect. Consumers get bombarded daily with countless promotions from countless advertisers, and feeding them more unnecessary content might result in them losing interest in the brand. Therefore, trained professionals should tactically use the available for the best results (Camilleri, 2019).

2.3.1. General Data Protection Regulation

It is common, in our ever-developing society, for technology to advance too fast for law to follow, often leaving temporary loopholes in the legislation. Consumer protection laws are stretched outside their limits with user data collection, use, and distribution. With some countries taking the potential danger for data privacy seriously, and others unaware of their role in everyday lives, the regulations to control how digital customer data is handled vary geographically.

The General Data Protection Regulation (GDPR) was finalized and published for the European Union in 2016. GDPR consists of six main principles, stating that personal data should be:

1. Processed lawfully, fairly and transparently;
2. Collected for specific legitimate purposes only;
3. Adequate, relevant and limited to what is necessary;
4. Accurate and kept up to date;
5. Stored only as long as it is necessary;
6. Protected with appropriate security measures, ensuring its integrity and confidentiality.

The GDPR applies to the personal data of all EU residents at all locations, meaning it is irrelevant whether data is processed outside the EU. Personal data is any information that can be used to identify an individual. How does the GDPR protect EU citizens? Every EU citizen has the right to, at any time, require a copy of all the data a company and any third-party has about them and demand complete deletion of their data from a company's system. The citizens also need to give the company full consent for data portability – meaning companies cannot exchange customer data without the customer's knowledge. In case of a breach of GDPR, whether internally or externally, the company must notify the citizens in less than 72 hours. The cost of breaching the GDPR goes up to 20 million euros or 4% of the annual turnover – if the latter is greater than 20 million (Milheiro, 2019).

How does this help when users still accept long Terms and Conditions without reading them? The EDPR has also changed how companies are required to display online consent and consent withdrawal forms. Those are no longer allowed to be hidden between the lines of long, illegible texts and must be visible and understandable for the user (Camilleri, 2019).

Finally, the customer decides how much privacy they will sacrifice for comfort. For example, most of us often accept all cookies on a website, in order to get to the wanted content as soon as possible. Sometimes, accepting cookies is a requirement for accessing the content. Other times, we could just as quickly uncheck all boxes with a few extra clicks and still end up on the same web page.

With all this in mind, nobody can take sole responsibility of data protection – not the law regulators, not the end users, and neither the companies handling the data. It is a joint effort to use available data in the most responsibly, without breaking government nor ethical laws.

2.3.2. Ethics of Data-Driven Marketing

As consumers become more aware of how much of their data is being collected, the public opinion about the importance of data protection splits. The phenomenon most talked about is how after we discuss a topic near our phone, we start getting ads about that topic. The river of conspiracy theories flows from this spring. Is Facebook eavesdropping on us? Has the government planted bugs in our home? Some insist they never talked about the topic, but only thought of it – so they must also be reading our minds.

What is happening behind this is relatively straightforward. No, Facebook is not eavesdropping on us. However, it is tracking us. Nevertheless, only because we gave it permission to. Facebook tracks our online behaviour, what we search for, and what we like and dislike – but this is old news. However, it also uses advanced location tracking techniques to follow our offline behaviour (Selman, 2021).

For example, Person A and Person B have a conversation, and Person A talks about having dinner at Place C, which Person B has never heard of. Later in the day, Person B sees an ad for Place C. Why is that? Like we said, Facebook knows what both persons do online. It also know they were at the same place today, thanks to location tracking. They are also friends on Facebook. Person A was at Place C yesterday. Since they are friends, Facebook thinks Person B might also like Place C. The fact that Person A mentioned any of this out loud had no impact on Facebook. Person B might have gotten, and probably has, an ad for a different place Person A was the other day, but as they have not explicitly mentioned it, Person B did not register it as unusual, or because of Facebook eavesdropping (Selman, 2021).

Is the user given the option to opt out of tracking? Of course. If the user plays around with permission and privacy settings, they can easily limit how much private data Facebook and similar apps are allowed to use. However, for as long as social media is used to any extent, the activity across the app will be enough for the app owner's to learn user behaviour. To conclude, the only way to make sure social media apps are not doing any tracking at all is not to use social media apps at all. However, this is a two way street, and by not letting apps use their data, the user is also giving up the right to use the data they are providing (Selman, 2021).

As eager as companies are to use customer data to gain an advantage, there are still multiple challenges along the data journey. The main issue concerns the potential risk and harm of using consumer data that goes both unrecognized by the consumers and ignored by the companies. However, data is vital for beneficial customer and company relationships. Companies exist to satisfy customer needs, and as technology develops, customers get greedier, due to the infinite number of choices they are offered (Solomon, 2016). Being bombarded with information daily, customers quickly block our irrelevant news and offers, leaving companies to fight for their attention. Companies must turn to data-driven techniques to properly deliver value, understand their customers' problems, and provide the best customer-oriented solutions (Kotler, 2016).

The lack of information about how data is collected and used puts the customer in a disadvantageous position. The scarce available information is often low level for a layman to understand, and as the idea of data remains abstract to the average person, the possible consequences of information sharing stay overlooked (Strycharz, 2019).

Companies often base their data storage and security solutions on the profitability factor. With this approach, privacy often suffers. It is common for companies to share their customer data with third parties if it promises to bring them profit, directly or indirectly. The risks of unauthorized data sharing involve decreased customer trust, drop in brand value, and even legal consequences (Schneider, 2017). Still, companies consider the trade-off and often make the less ethical decision. Moreover, as customers are widely uninformed and unaware of their rights, they are not very likely to notice the breaches in their data privacy, nor recognize its consequences.

Other than unethical practices, data-driven marketing has also influenced company transparency. The customer has a right to know when something is being advertised to them – we do not give the same importance to an ad we see on TV and a recommendation from our friend. Therefore, all paid advertisement should always be marked as such. Thanks to the sophistication of marketing tools, advertisement is becoming increasingly hard to tell apart from organic content – sometimes even with clearly stated disclaimers. This leads to decreased transparency (Bivins, 2009).

As effective as they seem, algorithms are, similarly to humans, not always to be trusted. Machine learning predictions may not work as expected, and lead to the delivery of wrong content to inappropriate people. Consumers can get exposed to content that they find irrelevant and sometimes offensive. In these situations, customer data ends up being used against both the user and the marketer (Walker, 2019).

According to a survey included in (Raappana, 2020), the news about various data leakages tied to big companies have not affected the majority of respondent's behaviour on social media. Still, a significant amount of almost 40% have reduced their use of the services included in the data leakages. Regarding brand trust, most respondents appreciated being able to control

how their data is stored and distributed. While most people approached giving out personal data for company use for free, they have also shown interest in selling their behavioural data for a fee. As they also feel mostly comfortable sharing the data for scientific research, the question is raised about the ethics of the user side of data-driven marketing. Are customers apprehensive about their data privacy, or do they just dislike the idea of companies profiting from them? Within the scope of the research, respondents were also asked about their opinion about adding a Fair Data label to digital services – which means services would have to earn the right promote themselves as a company that handles data fairly. Most respondents were for introducing such a label and believed that the requirements for its carriers should be security, transparency, fairness, trustworthiness, and consent.

In conclusion, ethics in data-driven marketing are equally hard to maintain as in any other aspect of life. Expecting completely ethical behaviour from organizations hungry for profit is idealistic, but the tightening of regulations should be an improvement. Consumers should be given more control of their personal data, clear and easy to understand policies, and options to opt out of tracking that is not necessary for the primary functions of the service they are using.

3. METHODOLOGY

3.1. PROJECT SCOPE

New competencies and skills are added to job requirement lists as technologies develop. Marketing has always been a field that requires a wide variety of skills. Marketers were always expected to have vast product knowledge, exceptional communication skills, presentation abilities and a good understanding of human psychology. With the switch to digital marketing, various computer and web skills were added to the requirements and the ability to adapt and learn continuously as the practices rapidly change (LinkedIn, 2020).

With the Lovidovi Shoes business located mostly online, it is company's policy to resort solely to digital marketing campaigns. In the first years after the founding, digital marketing was handled by the company founder personally, with quite generic content and automated settings. With the growth of the company, there was a need to establish a marketing department. Before the start of the project, the company's marketing was performing with a turnover. However, after the expenses of materials, manufacture, shipping, exchanges and returns were considered, more was left to be desired regarding the company's overall profit.

3.2. DATA PREPARATION AND ANALYSIS

3.2.1. Data Integration

Lovidovi Shoes is an e-commerce company, selling exclusively online. As its primary point-of-sale system, it uses Shopify. Shopify is a proprietary platform that offers sellers services necessary for operating an online store, including payments, marketing, shipping, and customer engagement tools (Shopify, 2022). However, Lovidovi does not utilize all of Shopify's features, and additional tools are used. Social media marketing is done through the Meta Business suite, Omnisend is used for e-mail marketing, Loox for product reviews and different financial apps are used dependent on the country. Each tool collects and stores data in different ways, making it harder to combine and connect the data. The company's marketing manager having access to all the separate dataset is an upside for this project. The downside is that no budget is allocated for data integration; therefore, no paid tools can be used, and no additional personnel can be assigned. As the most researched and used methods for data integration are cost- and time-consuming, a specific approach needs to be designed for the purpose of this project.

Most research related to data silos is done assuming that the silos are created throughout different departments of big companies. The ideal solution repeatedly suggested is creating a unified data management systems. Within the scope of this project, this is not possible. The most popular technique for data integration is ETL – extract, transform, load. This is a three-step process in which data is extracted from the source, then cleaned and loaded into the target system. This process is usually done through specialized software, but can be done

manually (Ralph, Kimball, 2004). For the manual executed, multiple system operators and teams are usually used. This process must be carried out strategically and carefully to make it a one-man job.

3.2.1.1. Data silos

An issue that occurred in this project is related to siloed data. This means that data is fragmented due to its accumulation through different channels. For example, user information is collected through the order system and subscriptions, and stored separately. A records form one dataset might be a duplicate of a record from the other, but this is harder to detect between different datasets within the same one (Stedman, 2021). The challenge of siloed data is often present in data-driven practices, and has been researched accordingly.

The causes for the creation of data silos can be technical, organizational or even cultural. In bigger companies, they are usually caused by different operational units with the inefficient data flow. For smaller business, data silos are likely to occur due to lack of strategic planning and different prioritization (Stedman, 2021). In most cases, data silos result from how a business is organized. In companies with individually managed departments, data silos form because of different priorities and managing principles. If a company with multiple departments gives them freedom to choose IT technologies which work best for the specific department, rather than the company, there will be multiple different applications and programs across the company. With a setup like this, data silos are hard to avoid. Even with company-level choice of programs, various available platforms can be the cause of decentralized IT operations, and consequentially decentralized data. Sometimes, the company culture alone is the reason behind data silos. If departments see data as their hard-earned assets, they might view data sharing as negative rather than positive thing. Another common cause of siloed data is business growth. With expansions, companies often need quick and immediate solutions, and are less likely to consider the permanent damage of not handling their data correctly from the beginning.

As a result of data silos, part of the data can have limited availability, meaning that not all users have access to all data. This can lead to incomplete data sets, which can yield incomplete conclusions. If the marketing team cannot access the sales team's data, they might miss out on helpful information in the dataset they cannot use. For the same reason, data inconsistency issues can arise. Data quality is threatened if separate operational units store and format data differently. As previously mentioned, duplicate data is often collected without data transparency, which is a waste of resources.

Breaking down data silos helps connect data and ensure its better usability. The most common technique is data integration – extracting data from the source system, consolidating it and loading it into a target system. As straight forward as it sounds, it is a complicated process. Multiple available tools for data integrations are all costly and time-consuming (Patel, 2019). Data warehouses and data lakes for storing data that can be used for reporting applications

and as enterprise data management systems, are common solutions for data silos. In addition to these, a cultural change in the company is also needed, if the prevention of creating new data silos is needed as well as a breakdown of the existing ones.

There is much left to be desired in the existing research regarding the breakdown of data silos, as most of the solutions only apply to bigger companies. With digitalization, smaller business are at an increasingly higher risk of data silos, as the search for the most affordable tools leaves their organization spread across different platforms. All-inclusive platform that can be used to manage a whole organization are often out of budget for small businesses, and free to use programs rarely meet all needs, so additional tools are needed. This causes data decentralization. Naturally, data silos form in this environment. The available approaches to break them down are all costly and unapproachable to small business.

Small companies often lack both human and financial resources to tackle those issues, which leads to them being ignored. Other business processes are prioritized, and those companies often miss out on the benefits they could get from fixing this issue, as they are not immediately apparent. A for more efficient data centralization in smaller businesses with limited budget solution is yet to be found.

3.2.1.2. ETL Process

The Extract, transform, load (ETL) process typically consists of additional sub-steps. Once the process begins, the first step would be to build reference data. Reference data is usually static data used to categorize other data (Loshin, 2014). After reference data is set, the other data can be extracted from the sources. Extraction would mean getting the data into a readable format, preferably for all sources. Next step is data validation ensuring that the data is clean and satisfies the data quality requirements. Then, the data can be transformed. This is the main step of the process, where data is cleaned further, business rules are applied, and all significant alternations are done on the data. Data is stored in staging tables – temporary storage between the source and target when transformed. An audit report is needed after this stage to help further use the new data. Furthermore, the data is loaded to the target dataset and saved.

Different datasets were used and created in this step of the process, varying in relevance and usability. For the principal analysis, three datasets were most important – Orders, Subscribers and Ads. The Orders dataset was used to understand the company's overall performance. Subscribers dataset is an extension of the customer dataset and is crucial for understanding user behaviour. Lastly, Ads dataset gives insight into how well previous marketing attempts worked, and how they can be improved. The tools for handling the data were predominantly Power BI, a data visualization software used for Business Intelligence, and Jupyter Notebook, a computing platform used for coding in Python.

3.2.2. Orders Dataset

The data for this project was of high volume, as I had access to the complete company database. However, for the purpose of easier processing, 10 thousand orders, corresponding to 10 thousand data rows, were used for the first dataset. The orders chosen were the most recent, as they were of the highest quality.

The data preparation process started after the extraction of the 10 thousand rows. The first step was dropping of irrelevant columns. In this process, columns with more than 50% empty values were dropped. The columns dropped with this condition mainly were a cause of the system offering the options to store values that the company is not using in their sales process – such as *Duties Value*, *Shipping Province*, and *Tags*. The next step was looking for duplicates, but first, a duplicate had to be defined for this context. The system generates unique order IDs, so there is not duplicates by the order number. If the same person ordered the same thing on the same day – would that make it a duplicate? While it is unlikely that the same person wants the same pair of shoes in the same size, it is still not impossible, and cannot be ruled out as a duplicate. This assumption ruled out the possibility of duplicates existing within this dataset. For further data cleaning, data types were checked. There were no issues here as the dataset was automatically generated with pre-set data types. Next, empty values were analysed. The only empty values in the dataset turned out to be related to the shipping address. However, they all shared the same shipping address – some version of "showroom", and had missing shipping zip code and country. All of them were manually entered, meaning a salesperson and not a customer made them. As the showroom is the only location where pick up is possible, sometimes it is just quickly entered without further details. It was easy to fill in the empty values with the zip code and country of where the showroom is located – 72230, Bosnia and Herzegovina. With this, the data preparation was done. At the end of data preparation, 54 columns were left in the data set. Some of them were *ID, Name, Email, Financial Status, Paid at, Fulfillment Status, Fulfilled at, Currency, Subtotal, Shipping, Taxes, Total, Discount Code, Discount Amount, Shipping Method, Created at,* columns describing the line item, as well as columns with billing and shipping*, Notes, Note Attributes, Cancelled at, Payment Method, Payment Reference, Refunded Amount, Risk Level, Source, Phone.*

The company started using a centralized order management system in the last two years, meaning all orders are registered in a uniform way. This has significantly increased the data quality. When it comes to five measures of data quality, the given dataset performs as follows:

- Accuracy – data is correct with minimized possibility of human error. A customer or a sales assistant can enter an order into the system. The input form checks the validity of entries that can be checked, and the product and price are always taken from the system, which makes them always correct. Other data inputs are vulnerable to human mistakes but are also less relevant for the analysis.
- Completeness – the data is complete, as an order cannot be entered into the system without all required fields.

- Reliability – the personal information about the customers in this dataset might not be aligned with their records elsewhere, but other order details only exist in this dataset and therefore are reliable.
- Relevance – this dataset is relevant, as its columns were taken from a larger dataset, while the irrelevant ones were dropped.
- Timeliness – the information is up to date when the data preparation started. Updates made after the start of this project were not taken into consideration.

Overall, the dataset meets the data quality requirements and is suitable for use in such a project. (Sarfin, 2022)

3.2.3. Subscribers Dataset

Separately from the orders data, customers' data is also available for the use in this project. However, the customer data is extracted from an order, and in addition to what is already known from the order, there is not much useful information within this dataset.

However, there is also a separate dataset including subscriber data, which can better help understand the audience –customers and subscribers. To subscribe, users need to enter their e-mail and date of birth. There are additional optional fields, such as phone number, and the shipping information is automatically added for users that made an order. For the purpose of this project, only six relevant columns were kept for the 7410 subscribers. All columns involving personal information that can be used to identify an individual, such as e-mail, name, or address have been dropped for privacy reasons, and replaced with an auto generated Subscriber ID. The remaining columns are *Country Code*, *Zip*, *Total Spent*, *Total Orders* and *Birthday.*

In data preparation, the dataset was first checked for empty values. The column with a significant number of null values is *Zip.* This is because the zip code exists only for the subscribers who have already ordered. The country code is not missing in the same manner, in that is is automatically added based on the IP address rather than entered by the user. As the empty values are really missing, it was decided to drop the column due to the inability to fill them appropriately and the lack of usefulness of the column.

Then, the age was extracted from the birthday column in Python, because age is easier to analyse and use than birthday is. At this point it became interesting to look at the gender of the subscriber as well. However, this information is not collected during subscription. According to the company, this information was irrelevant, as the websites only sell women's shoes, so the gender is assumed to be female. This is probably true regarding who the shoes are for, but it is not a rule that a woman will always buy them. From a marketer's perspective would be valuable to have this information, because it can help decide whether it is profitable to advertise to men as well. Looking at the boxplot of the Age column in Figure 7, most subscribers are aged 30-45. This complies well with the brand's assumption that their target audience is 25-50.

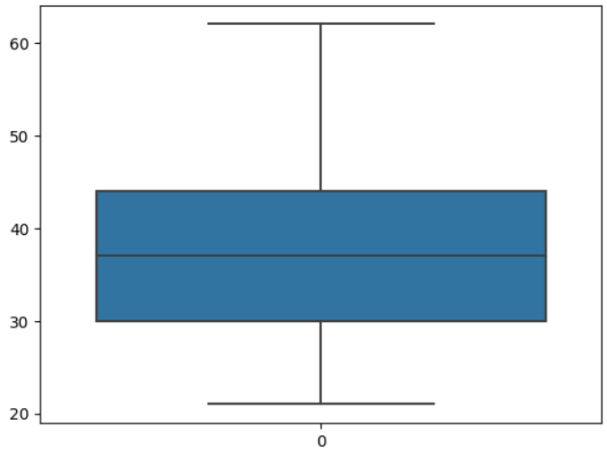

Figure 3 Distribution of Customer Age

Also, the number of subscribers per country corresponds well to the number of orders per country, with most subscribers being from Bosnia and Herzegovina, then Germany, Croatia and Serbia. The number of orders per subscribers shows that around 50% of subscribers have not purchased yet. Also, only 20% of the people who placed an order made more than one purchase. This means there is 20% returning customers. The average retention rate for the appareal industry is 26%, which makes this a satisfactory number (Teneva, 2022). Still, it leaves 80% of customers who have only made one purchase, so increasing customer satisfaction could be a meaningful goal. Also, these percentages among subscribers show an interesting advertising opportunity – turning subscribers into customers and getting customers to make another purchase is one of the least costly and most efficient marketing approaches. Nevertheless, this is not the primary focus of this project, but will be considered for future use.

3.2.4. Advertising Dataset

Another dataset was needed for deeper analysis of the connection between sales and marketing. As most digital marketing is done through the Facebook Business Ads Manager, this is the data source chosen for the third dataset.

As the data available through this channel is very high in volume, it was necessary to find a way to extract only the most helpful information. After an attempt at exporting various different tables, none appeared to be very usable in the context of this project. A decision was made first to create a custom report and then extract further data based on the results of its analysis.

The most valuable information for this project was easy to define – *Which ad campaigns were the most successful in the past?* While *most successful* can be defined in many ways, in terms of the company's business rules, most successful equals most sales. To find the top campaigns, a report was made from all four ad accounts used for Facebook advertising, including the following columns – *Ad Set Name*, *Purchases, Cost per purchase*, *Clicks*, *Frequency*, and *Reach.*

As the data was exported directly from where it was created all the data quality requirements were assumed to be met. This datasets only purpose was to get the top 10 most successful ad sets throughout the years, so even if there were faulty data in the other rows, it would not affect the analysis.

Using Python, the table was sorted by the values of the *Purchases* column's values, as seen in Table 1. This column has the number of purchases that were completed through the link in the ad set. From these columns alone, there is not much to conclude about those ad sets, but one thing does catch a digital marketer's eye. The mean of *Frequency* for the top 10 rows is 5.15. This column represents the average number of times each person saw an ad. In advertising, most people would agree that it is not desirable to have this number be over 2, because it means the ad has already reached the limit for reach, and is now showing the ad again to the same people – which might result in them being bored or even irritated by it. However, looking at the finding that most successful ads all have a frequency over 2, it seems that sometimes it is valuable to show the same ad to the same person multiple times. The empirical conclusion drawn from this data is that is not necessary to stop an ad based on the frequency, as people might take a few looks before completing a purchase through an ad.

Ad Set Name	Purchases	Currency	Cost per purchase	Clicks (all)	Frequency	Reach	Reporting starts	Reporting ends
New Ad Set	112	EUR	1.973	10,888	2.906	139,358	24.9.2019	24.10.2022
New Ad Set	67	EUR	3.030	10,877	3.901	93,318	24.9.2019	24.10.2022
Sneakherz 2.0 - PRESALE Ad set	56	EUR	3.334	4,902	3.987	67,452	24.9.2019	24.10.2022
FEED	35	EUR	7.100	5,150	4.364	72,506	24.9.2019	24.10.2022
FEED AIR	33	EUR	9.874	8,117	4.339	149,516	24.9.2019	24.10.2022
New Ad Set	33	EUR	5.882	4,476	3.462	89,625	24.9.2019	24.10.2022
ost Engagement (Last 14 Days)	30	EUR	75.611	2,047	12.742	12,050	24.9.2019	24.10.2022
Feed	25	EUR	13.019	4,908	6.880	52,253	24.9.2019	24.10.2022
Women - engaged shoppers	24	EUR	8.331	6,106	4.197	100,167	24.9.2019	24.10.2022
Feed - All Women	22	EUR	9.015	2,887	4.761	51,824	24.9.2019	24.10.2022

Table 1 Top 10 Campaigns

Additional columns were manually added to the dataset to increase this data's usefulness further. The dataset was cut to just the top 10 rows, and then more columns were added. *Location*, *Age group*, *Gender*, *Interests*, *Placement*, *Daily budget*, *Number of ads*, and *Products*

advertised were chosen as the most relevant ad set characteristics that can help nbetter understand the demographics and settings.

With only ten rows of this dataset, any technical analysis is redundant. Just looking at the data, it is easy to draw the basic conclusions. The most successful audience is women, ages 16-65+, in Bosnia and Herzegovina, without any narrowing by interests. This indicates that the broader the audience, the better the performance – at least for this market. Further on, the best ad placement is on Facebook and Instagram feeds, followed by automatic placement. The number of ads in an ad set is best limited to one – this is assumed to be because the ad set budget is then aimed at the single ad rather than the algorithm having to test which ad should be pushed the most. Regarding products – the ads with most sales are promote either just sneakers or multiple different models.

3.3. TRIAL AND ERROR

Data analysis was expected to answer the questions about the target audience age, gender, and country, as well as the best ad placement and the bestselling products. The conclusions made in the data analysis step will be applied in a real-life scenario, to best evaluate which ones have the most value. Each idea from the data analysis results will be cross-tested with the remaining ones. This is the best way to properly learn which factor had the most effect on the results. First, a number of campaigns are going to be run. After two weeks, the results of all the campaigns are going to be collected and analysed.

3.3.1. Campaign Setup

The following structure was decided on to best cross-test the chosen ad set settings.

Campaign *Women* will contain the following ad sets:

- *Women/25-45/Feed*,
- *Women/25-45/Automatic placement,*
- *Women/20-50/Feed,*
- *Women/20-50/Automatic placement,*
- *Women/18-65+/Feed,* and
- *Women/18-65+/Automatic placement*.

This campaign will be aimed only at women, with each ad set combining the different age ranges and placements that were decided through the analysis.

Following the first example, two other campaigns will differ only in gender. One campaign will target only men, and another all genders. Each ad set will only have one ad to avoid an uneven distribution of funds for the same ad across different ad sets, which would make the results harder to interpret.

As stated before, the ads will not be tested based on the budget; therefore all ad sets are given the same budget. The daily budget chosen for each ad sets 13 USD, the average daily budget of the ten most successful campaigns from the past.

For the ad creative, a single image is used. The same image is used across all ad sets, with the same text and call to action. The product advertised is Sneakherz 2.0, which are white sneakers. The ads are set to run for twi weeks, from April 4th to April 18th. This period is chosen based on peak selling time being May. The trial campaigns are set in April, which is a raising period, and after they end, there will be a month long break before launching the final campaign. The final campaign is planned to start May 16th, which is predicted to be the most prolific period for sales. It is crucial to leave time between the end of the first and start of the second campaign, as continuous imposition of content may have a negative impact on customer interest.

Once the first group of campaigns start running, there will be no continuous assessment. The ads will be left to run for two weeks before any analysis. This is to ensure that the algorithm has passed the learning stage and that the results are complete and final at the time of analysis.

After the campaigns finished the running cycle, a report was created in Facebook Ads Manager including the same metrics as the one used to identify the most successful campaigns of all time. The report contains 18 rows, each corresponding to one ad set. Other than the descriptive name*, Reach, Impressions, Purchases, Cost per purchase, Frequency,* and *Clicks* are available for each ad set.

In order to improve data usefulness, three new columns were added first. Columns *Gender*, *Age* and *Placement* are columns relating to the Ad set settings are derived from the Ad set name. The string of the Gender/Age/Placement format used as the ad set name was split on '/' and used for the separate columns. For a more formal record, *women* and *men* are replaced with *female* and *male*. *AP* in Placement was substituted for *Automatic.* These columns were added to the group data data.

Additionally, two more columns are created for the purpose of measuring the results. Column *Percentage clicked* is created as a quotient of the columns *Clicks* and *Reach.* Column *Clicks* contains the number of unique clicks, corresponding to the number of users that have clicked on an ad. Column *Reach* shows number of people who saw the ads at least once. *Reach* is different from impressions, which may include multiple views of ads by the same people. The *Percentage clicked* column shows the percentage of people who have clicked on the ad, out of all the people that have seen the ad. The metric *Percentage bought* is equivalent to the quotient of *Purchases* and *Clicks.* This column shows the percentage of people who have completed a purchase out of all who clicked on the ad. In this case, the clicks value is more valuable than the reach value, as it better represents the audience interested in the ad. People

who have scrolled past the ad can hardly be considered potential buyers, while the ones that visited the page can be tracked further and turned into customers.

3.4. TESTING

For the final stage of the methodology, the results reached through previous stages were tested. Only one ad set will be launched in this phase, using the settings that are deemed to be the most favourable. An ad set aimed at women ages 18+ in Bosnia and Herzegovina, with only one ad in the feed placement will run for two weeks at peak season – May 16th to May 30th. Again, the product advertised here is a white sneaker, this time Chichak White.

However, to get the best results, following the guidelines set throughout this project is not enough. During the analysis of the new campaigns, one percentage that should be improved was noted. For the best performing new ads, the percentage of users clicking is higher than in old campaigns. However, the percentage of the people who made a purchase is lower. While increasing the number of people clicking on the ad could potentially improve sales, it is more useful to focus on the people the people who have visited the website through the ad, because they mostly have some level of purchase intent. Leading a user to checkout after they land on the product page is no longer the job of the advertiser or the Facebook algorithm, but rather the web designer.

According to (Keskin, 2022), the product page is an advertising opportunity. Some clear best practices are in place and following them can help decrease the possibility of losing a customer between and ad click and checkout. In collaboration with the web developer, the product page was re-designed to comply with the suggested rules to improve the performance of the upcoming ad campaign.

The one thing visitors are sure to look at are product images. These must be of good quality, and represent the product well. There should be no ambiguity about which product from the image is being sold. As the product images are the first thing a potential customer looks at, they can be used as a medium to get a company value or a product feature across. For this project, an image of a flyer about the collaboration between the brand and SOS Children Village's was added as the last image in the set, to ensure the visitors do not miss the humanitarian background of the brand.

Word of mouth is one of the most efficient marketing mediums. One way to incorporate this into an e-commerce shop is to use pictures and names of actual customers when discussing customer satisfaction. This was done by adding a snippet that says ''*Sanja, Ena and 6,500+ others are satisfied Lovidovi customers*.'' accompanied by image icons of the named customers. Phrased like this, this sentence has a more positive effect on visitors than a pure statistic of ''6,500+ nameless people are satisfied Lovidovi customers''.

Photo product reviews, of course, further contribute to the word of mouth community feeling. Visitors are more likely to believe review images and customer comments about the product

than anything the product page says, so adding customer reviews is crucial to getting customer trust.

The product description should always be detailed and helpful, but also not redundant. Most customers do not want to read a whole story before finding out the product's material, so clearly stating the products main features in simple format is the best approach.

Another important measure stated in (Keskin, 2022) is about keeping the user on the product page – the roads from the product page should only lead to checkout. Therefore, additional information, such as the size guide, should be offered on the same page.

With all the basics covered, the main product and brand values should be reinforced through the product page. Lovidovi is a brand that makes handmade, sustainable and socially responsible leather shoes. This should be highlighted early on the product page, setting the brand apart from the competition.

Adding the delivery and payment terms to the product page also improves user experience, as it does not require going through multiple checkout steps before getting the information crucial to the purchase.

The product page's overall colour and shape design was adjusted to look clean and calming, in order to not stress the visitor into leaving the page. The only bright colour on the page is the red used for the price. According to (Rego, 2022), this colour of price is associated with sale, so the visitors looking at it immediately think of a good deal, even when the price is not reduced from the regular one.

A part of the re-designed product page can be seen in Figure 14.

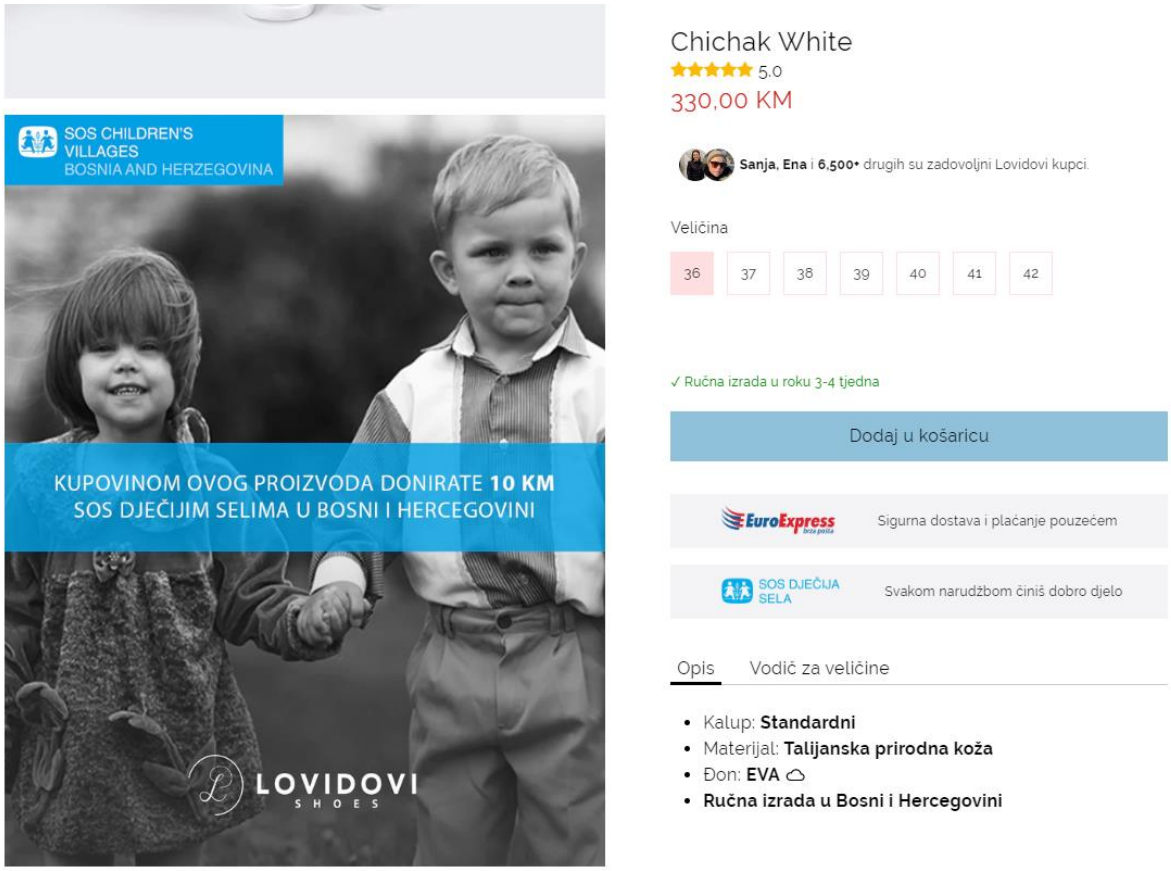

Figure 4 Screenshot of Product Page

While seemingly not a marketing issue, redesigning the product page has the most potential to turn a visitor into a customer. Clicking on a digital ad is the online equivalent of walking into a store after seeing the window. The product page is then the salesmen who can sell something to the visitor.

After the product page is adjusted to increase the percentage of visitor purchases, the Facebook Ads campaign can start running.

4. RESULTS AND DISCUSSION

Having analysed datasets from different data sources, some significant conclusions were drawn. The process aimed to create a clear blueprint for creating ad campaigns. To have the most relevant information, a set of questions to be answered.

What age is the target audience? – The company aims to make shoes for the 25-50 age group. Looking at the customer base findings, the majority of Lovidovi customers are between 30 and 45 years old. However, advertising data shows that the most successful campaigns were without age restrictions. None of the above conclusions contradict, so testing the different age ranges, from the broadest to the narrowest found in the data, made sense.

What gender is the target audience? – There is a lack of gender indicators in the data but looking at the customer names proves that the audience is not 100% female. An all-female, an all-male and a mixed audience were tested, in order to bring more clarity.

Which country is the target audience? – Bosnia and Herzegovina has showed to be the most profitable market overall. For this project, in agreement with the company management, it was decided to focus on this country, rather than explore new ones. The goal was to make the most of this market's potential.

Which placements work best? – As for the ad placement within the Facebook marketing channel, feeds have the best performance, followed by automatic placement. As Facebook has recently introduced new features, including Advantage+ automatic placement, an improved version of the latter placement, it was also tested.

Which products are the best-sellers? – According to the data from the target period (April-June) 2021, white sneakers were the most sold model. Out of ten best performing ads overall, five ads were also used to promote white sneakers. Other than that, only two places were held by single models, while the remaining ads were multi-model ones, promoting the web shop in general. Overall, the top best-selling products were mostly white sneakers, and all the models of white sneakers take up about 30% of overall sales. The best product to sell in this period was chosen without an alternative, and it is a new model of white sneakers.

Visual analysis of the Orders dataset was done using Power BI. The first looked at is the distribution of orders over countries. As seen in Figure 3, the top three markets are Bosnia and Herzegovina, Germany and Croatia. Next, Austria and Serbia also make a significant percentage, while all the other countries fall under 1%, making them irrelevant. This information is interesting from the marketer's perspective, as decision must be made. Should the focus be on the three main markets and increasing revenue within them, or should an attempt be made at improving presence and sales in other, still unestablished markets.

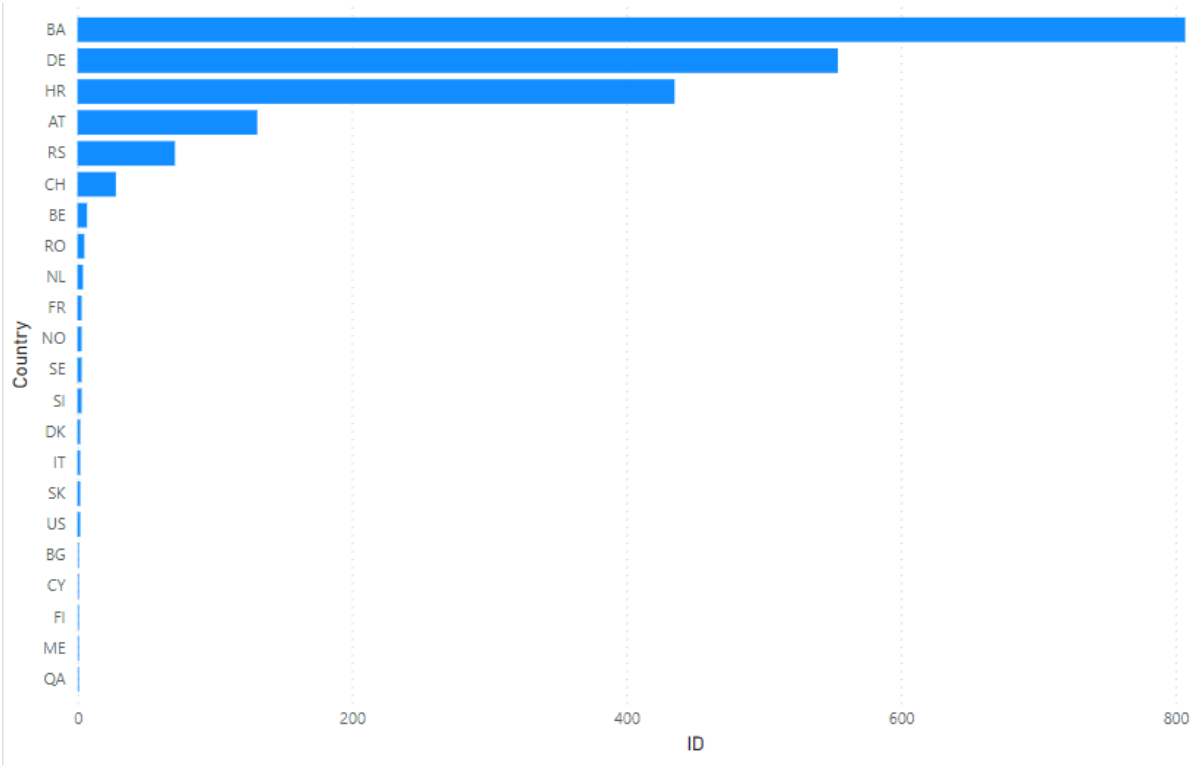

Figure 5 Orders per Country

This raises another question – if there many sales in a market, does that make it a good, profitable market? Figure 4 tells us that no, it does not. Among all the orders that were returned, Germany holds the biggest share. This makes it an unfavourable market, as the company does not charge for shipping or returns, so a return costs money. In this case no sale at all makes for a better business deal than a sale that results in a return. Exploring new markets looks more optimistic than focusing on a market prone to returns.

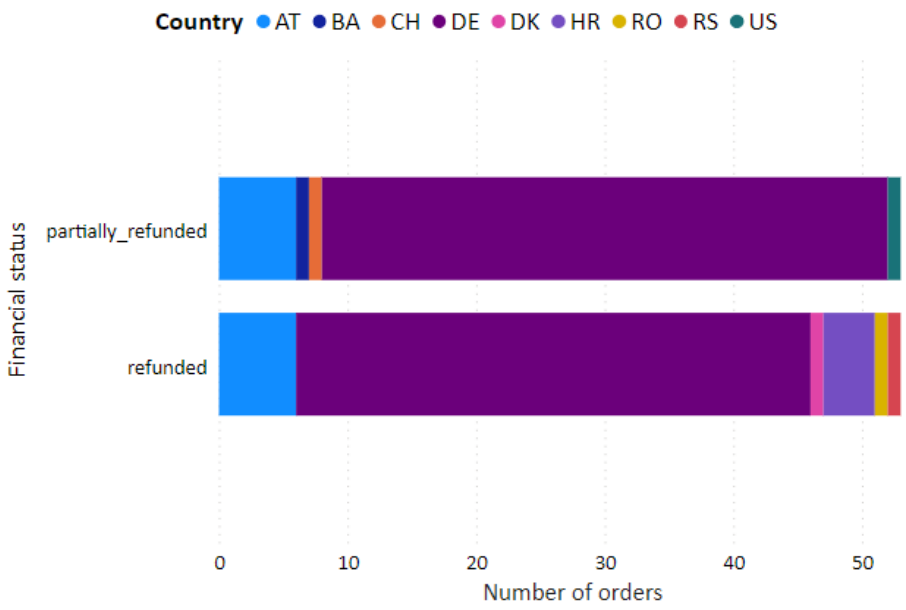

Figure 6 Returns per Country

Figure 5 shows a line chart of sales through months for the previous year. This this line chart is essential because it shows how sales are seasonal, with lows in February and July and peaks in May, September and November. To get the best results for this project, it is important to conduct it during some of the peak times. Even with November being the most prolific month, with the projects deadline in mind, choosing the spring peak months for the experiment part of the project seems like the best option.

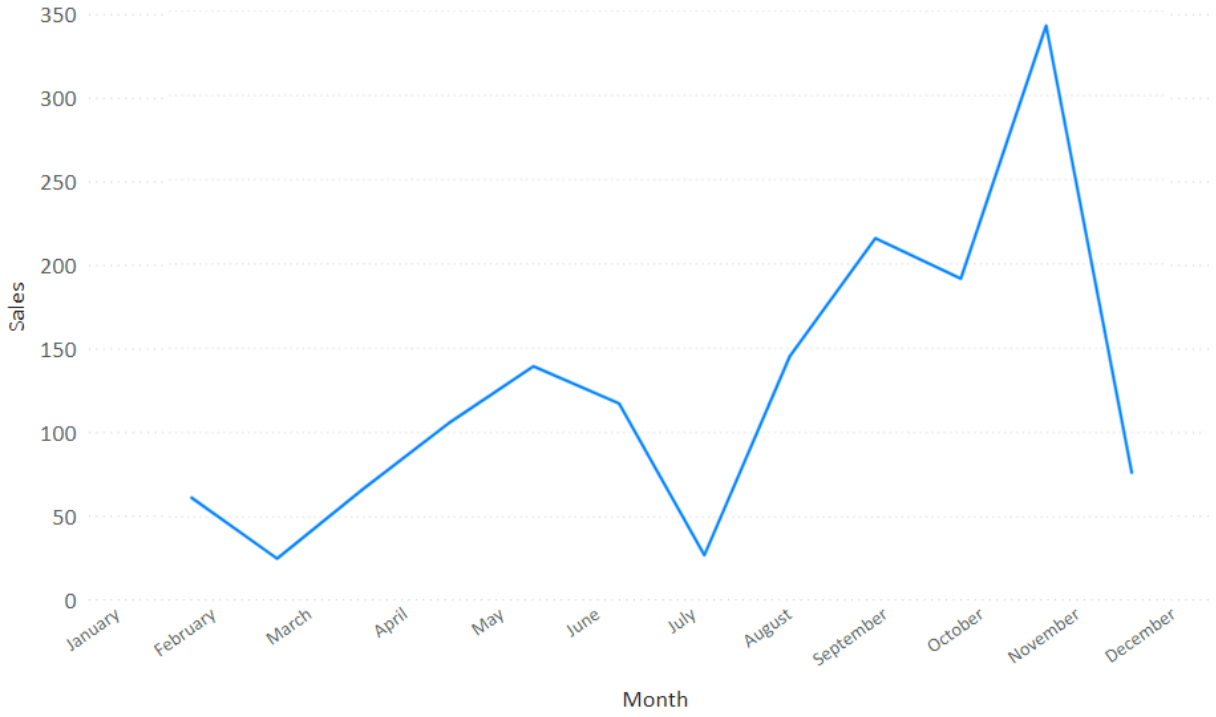

Figure 7 Sales over Time

When focusing on the April, May and June 3-month-period, it is also important to see which type of products are the most popular. As seen on the pie chart in Figure 6, which shows the top 10 products by sales for this period in the previous year, 2021, the most purchased product is Sneakherz White-Beige. This product is a classic, basic sneaker model in white, with small beige detail. The two products following it are an all-black oxford shoe and an airy model with a light sole and breathable leather in white. Knowing the products, it is easy to conclude that the best-selling models are in basic, easy to style colours. They are also every-day shoes that people wear warmer weather. This information is important when choosing which products to advertise during this period, even if there is a new collection.

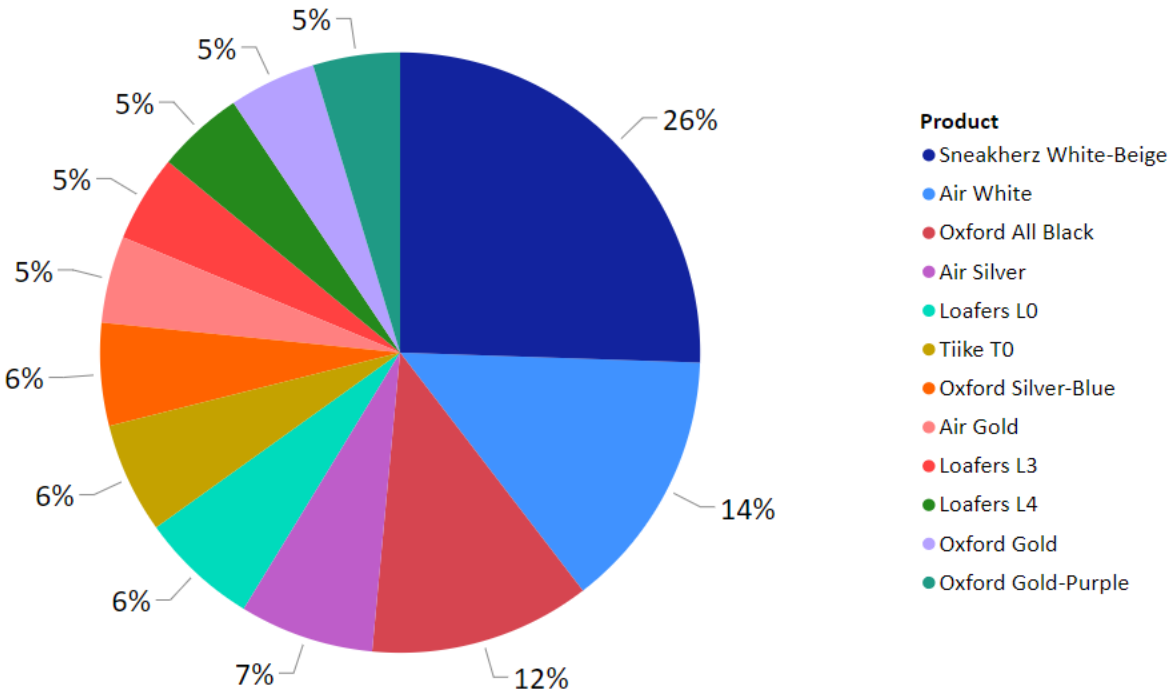

Figure 8 Besteselling products in April, May and June

Table 2 shows the top 3 best performing ad sets from the trial, based on the number of purchases. The immediate connection between the top 3 ad sets is that they all include women in the audience and have a narrower age range. Based on this, it could be concluded that targeting the brand's most specific audience is more effective than expanding it. Looking at the following tables, we can better determine if this conclusion has significant sustention.

Campaign name	Ad Set Name	Placement	Gender	Age	Reach	Impressions	Purchases	Cost per purchase	Frequency	Clicks (all)	Percentage Clicked	Percentage Bought
WOMEN	WOMEN/25-45/FEED	Feed	Female	25-45	106,182	630,197	82	2.219	5.93562	10,432	0.098246	0.00786
ALL	ALL/25-45/FEED	Feed	All	25-45	100,727	378,302	46	3.956	3.755712	9,313	0.092458	0.004939
ALL	ALL/20-50/FEED	Feed	All	20-50	93,357	197,211	40	4.550	2.112435	9,225	0.098814	0.004336

Table 2 Top 3 Trial Campaigns

Table 3 shows the percentage metrics when the entire dataset is grouped by Age, rather than just the top 3 results. The mean of all rows in that group represents the percentage. The differences between the means are all in less than 0.001, which is insignificant. It can be assumed that expanding or narrowing the age range does not make much difference, probably because all the ranges fall into the most extensive range. Results would probably differ more, if the age ranges excluded one-another, but that is not useful for this project, as the target audience age range is already narrow enough and binning it further would not have much impact. On the other hand, testing out different successive age bins would go against the point of this project, which is to make decisions based on data, and data suggests that 35-45 five is the most prolific age group. However, looking at the results from this stage, it can be said that the assumption that targeting a more specific age group rather than covering all ages is in fact wrong, as there is not much proof of it. This part of the analysis concludes that using the maximum age range is the best practice for further marketing.

Age	Percentage Clicked	Percentage Bought
18-65	0.069180	0.005573
20-50	0.067433	0.005068
25-45	0.068287	0.005723

Table 3 Click and Bought Percentages by Age

In contrast, Table 4 shows significant differences between different gender groups. When grouped by gender, the mean Percentage Clicked is highest among women and significantly higher, over six times higher in women than men. It can be assumed that the proximity in percentages between Female and All is thanks to the Female *in* All.

Gender	Percentage Clicked	Percentage Bought
All	0.094453	0.003749
Female	0.095448	0.004000
Male	0.014999	0.008616

Table 4 Click and Bought Percentages by Gender

However, the results in Percentage Bought are interesting because while All and Female are still similar, Male has a twice as high mean. This does not mean that more man decided to purchase after seeing the ad, nor does it mean that they are more likely to buy. Because Percentage Bought shows the percentage of people who have completed a purchase out of the people who clicked on the ad – this means that once they interacted with an ad, men are most likely to purchase. In other words, men only click on the ad if they intend to purchase, while women click just to explore. This is interesting, but not very relevant to the nature of our analysis. A new column is added to get a better comparison for this case. The column *Percentage Bought Overall* is introduced to show the percentage of people who purchased out of all who have seen the ad. In Table 5, it is visible that this metric shows how men that are shown the ad are, after all, less than half as likely to make a purchase.

Gender	Percentage Clicked	Percentage Bought Overall
All	0.094453	0.000355
Female	0.095448	0.000385
Male	0.014999	0.000128

Table 5 Click and Bought Overall Percentages by Gender

Also, purchases made by men made up around 15% of the total sales. Due to great similarity between All and Female, with a drastic difference of both to Male, the assumption that it is the female population was targeted mainly by the algorithm in the All group remains. For this project, men can be ruled out as a significant target audience under regular conditions. More work can be done on men-specific advertising, and how to form an audience that will include a higher percentage of man who will interact with the ad.

Looking at the same metrics for different placement, that is Feed and Automatic placement, – the percentages are very similar, with Feed performing slightly better. According to (Breen, 2022), the best placement option is defined for each brand, and automatic placement is best used to determine which placement works best. The breakdown of the automatic placement ad sets by placement shows that over 50% of sales are still made through feed. This leads to the conclusion that Feed is the best placement for the brand. It is also the easiest for users to interact in, as it is the most permanent, no-time-limit placement.

Women and the 18-65 age group seem to be the most successful combinations in this trial. Nevertheless, it is essential to consider how these ad sets performed compared to the ones used in the data analysis. The metrics Percentage Clicked and Percentage Bought have been added to the original top 10 best performing ads dataset to make a simple comparison.

Table 6 shows the mean values for the top 10 ad sets of all time, the mean for the 18 ad sets from the trial. It is visible that the trial ad sets are not performing very well in comparison to the original ones. Although, this is not a fair or even a useful comparison. What makes most

sense is to look at the results for the top performing ad sets from the trial. Since the trial only has 18 ad sets, only top 3, by the number of purchases will be considered.

	Top 10	**New 18**
Percentage clicked	0.081048	0.068300
Percentage bought	0.008267	0.005455

Table 6 Click and Bought Percentages Old vs. New Campaigns

Table 7 shows the means for these three ad sets. While the percentage of people interacting with the ads is higher than of the original ad sets, there is a lower percentage of purchases. Once again, this does not mean less people are buying – just that more people are leaving the website without purchasing.

	New top 3
Percentage clicked	0.096506
Percentage bought	0.005712

Table 7 Click and Bought Percentages New Top 3 Campaigns

It is hard to judge the performance based solely on these metrics, so to make a more concrete evaluation, we can look at the most obvious metric – number of purchases. The total number of purchases for the original dataset is 437. The trial ad sets add up to 503 purchases. As the first dataset has 10, and the second 18 ad sets, this is not proof that the trial brought better results. However, it is enough proof to show that the trial ad sets have the potential to reach the success level of the top campaign. Comparing the ad sets separately shows that some of the trial campaign already made it to replace some of the top 10 campaigns of all time.

Table 8 shows the new ranking list of campaigns, with the trial campaigns added to the dataset. Six out of eighteen new campaigns have made it into the top. Based on this alone, the experiment can be considered successful.

Ad Set Name	Purchases	Cost per purchase	Clicks (all)	Frequency	Reach
New Ad Set	112	1.973	10,888	2.906342	139,358
WOMEN/25-45/FEED	82	2.219	10,432	5.935062	106,182
New Ad Set	67	3.030	10,877	3.901305	93,318
Sneakherz 2.0 - PRESALE Ad set	56	3.334	4,902	3.987739	67,452
ALL/25-45/FEED	46	3.956	9,313	3.755712	100,727
ALL/20-50/FEED	40	4.550	9,225	2.112435	93,357
WOMEN/18-65+/FEED	37	4.918	9,643	3.597804	97,524
ALL/18-65+/FEED	36	5.055	9,368	2.383094	95,184
WOMEN/25-45/AP	35	5.200	10,523	4.728366	10,6979
FEED	35	7.100	5,150	4.364866	72,506

Table 8 Overall Campaign Ranking

After running for two weeks, the final campaign exceeded expectations regarding the number of purchases. Because two weeks were used in the trial campaigns, a report about the results was generated while the campaign was left to run.

The five questions above are answered again with the updated information to explain the results best. Regarding the target audience, it is recommended to use a broad age audience, involving everybody over 18. This is because using a narrower audience had no proven benefit, and it is better not to cut off potential buyers by age. When it comes to gender, women are the better performing gender, and for this project, the purchases made by men are not of great significance when trying to achieve the best overall performance. Bosnia and Herzegovina is clearly the brand's best market and the cheapest in terms of cost per purchase, making it the best location for increasing revenue. The feed placement on Instagram and Facebook works best with the brand and gets the best reaction from the audience. In terms of products, white sneakers are the brand's bread-winning model, as not only the top selling products fall into this category, but also no model in this category has had insufficient sales.

Table 9 shows the results of the campaign after 14 days. The immediate indicator of its success is the number of purchases, which is higher than in the best campaign so. Another interesting improvement is in the percentages of people clicking the ad and buying the product.

Campaign name	Purchases	Cost per result	Reach	Clicks (all)	Frequency	Percentage Clicked	Percentage Bought	Percentage Bought Overall
Chichak/Women/Feed	115	3.03	93318	10,877	3.901305	0.116558	0.010573	0.001232

Table 9 Top Campaign Performance

The increase of the clicked percentage can be attributed to good targeting – meaning that more people shown the ad were actually interested. The percentage of the people who have reached checkout after clicking on the ad has doubled, which means the attempts to improve this number were useful.

The average cost per result (CPR) for the apparel industry is 10.32 euros (Irvine, 2022). With that in mind, the winning campaign had an over three times better CPR. For optimal performance, a frequency anywhere between 1.8 and 4 is expected (Harper, 2018). Most of the ads created in this project met that criteria. Facebook calculates the click through rate (CTR) as Clicks divided by Impressions times 100. This equals to the Percentage Clicked value times 100. As 1.24% is the average CTR for the apparel industry, this project has brought the ad's CTR closer to the industry standard.

In Table 10, we can see the breakdown of the ad performance based on the age group. Obviously most sales were still within the 25-45 age group, but there were significant sales in the younger and older age groups as well. This confirms that expanding the data range for the ad set was a good assumption, as over 20% of the sales came from users outside of the brand's target age group. The percentage metric shows the most interesting irregularity for the 18-24 age group. This group has high percentage of clicks, but the purchase percentage is lower among the lower ones. Based on the brand's previous observations and experience, this age group does not typically fit the brand's pricing range. This could be why purchases are low for this age group, even though interest is shown.

Ages	Purchases	Reach	Clicks (all)	Percentage Clicked	Percentage Bought	Percentage Bought Overall
18-24	8	24,397	2,455	0.100627	0.003259	0.000328
25-34	52	21,587	3,866	0.179089	0.013451	0.002409
35-44	36	19,826	2,447	0.123424	0.014712	0.001816
45-54	16	16,642	1,482	0.089052	0.010796	0.000961
55-64	3	8,226	612	0.074398	0.004902	0.000365
65+	0	2,640	15	0.005682	0.000000	0.000000

Table 10 Ad Performance by Age Group

At the beginning of the project, the most successful campaign of all time had 112 purchases. At the end of this project, the most successful campaign has 115 purchases. Looking at Table 11, only a small increase in the bought and clicked percentages can be seen. However, these numbers cannot be observed without context.

	Original top campaign	**Final top campaign**
Purchases	112	115
Percentage Bought	0.010287	0.010573
Percentage Clicked	0.07813	0.116558

Table 11 Top Original and Final Campaign Comparison

In this comparison, the best performing ad ever was used – this is in no way an indicator of how an average ad performed. Even though this campaign had 112 purchases, the average number of purchases for the 450 campaigns was 6.5. After adding only 18 new campaigns, created based on data analysis, this average has gone up to 8.4. The aim of this project was not to out-do the top campaign, but rather to create a campaign blueprint that would perform as well as the top campaign.

5. CONCLUSION

The problem presented as the motivation for this project was the company's dissatisfaction with the marketing performance. The objective set was improving the marketing approach, measured explicitly through increased sales per ad. In order to achieve this, a switch from intuition based to data-driven marketing was made

Throughout the project, company data collected over the course of seven years was used. Before significant analysis, data integration was done by temporarily breaking down data silos and centralizing the data. As the budget created limitations, a more great-scale, long-term solution to siloed data was not possible. Instead, a simple step-by-step process was designed to centralized data when needed for a specific project.

The data was then analysed for a better understanding of the company's customer base and an overview of the current digital marketing performance. Trial campaigns were then designed, following the best practices defined through the data analysis.

With the guidelines setup with this project, the company should be able to run ads that perform better than the ones in the past. While the algorithm and changes in trends can never guarantee constant success, this one-time testing can significantly reduce the need for multiple tests usually done with each campaign run. A/B testing with minor adjustments is suggested occasionally, in order to keep the findings up to date.

The main factor that made the realization of this project possible is unlimited access to the company's data. The limitations on what can be done with the data were imposed because it is a part of a master's thesis, making it a one-person project. Also, the company's goals and ambitions for the given period largely impacted the topic and focus of the project. The company dictated that this research should focus on the main marketing channel and its improvement rather than exploring different or new mediums. The best general solution that can be reused continuously was also required, rather than allowing for a more specific solution.

The major contribution of this project to the company is its impact on how data-driven marketing is viewed. With the success of the trial campaigns, the company has gained insight about the potential of using data and further investing in its collection, management and analysis. The company now understands that if a small project like this already impacted the increase in sales, it is worth planning a larger scope data-driven marketing project.

From the marketing point of view, this project helped set up a pipeline from data to value. The always available data might seemed scattered, and it was hard to decide where to start and what to look for. While every brand has its unique structure, this project can be used as a

foundation for resolving the issues related to data storage and data analysis when an insufficient budget is available.

5.1. Future Work

Through this research, some information was uncovered that was not used within this project. For example, a significant percentage of subscribers that are not customers yet allows e-mail marketing campaigns, which are un-costly advertising mediums with high probability of success.

Another recommendation for future work is exploring the male audience as a target audience. This research has shown that men have the highest percentage of clicks turned into percentages, but the lowest percentage of ad views turned into clicks. This much space for research about targeting the right audience within the male audience. Occasion campaigns have a lot of potential to work well, with products being advertised as gifts during the holiday seasons. The Croatian market has exponential growth in sales with potential to have an even greater share of overall sales. Some future work should consider exploring this market separately and improving it.

The last proposal for future work is more extensive involvement with other advertising channels. So far, Facebook Business has showed to be the most helpful advertising channel, but this is a biased conclusion because of the fact that it is the most used channel with most work put into it. Hopefully this project can stabilize the Facebook channel and leave time and money for exploring other mediums.

REFERENCES

Andrejevic, M. (2013, April 6). The Big Data Divide. International Journal of Communication, 8.

Bivins, T. H. (2017). Mixed media: Moral distinctions in advertising, public relations, and journalism (3rd ed.). New York, NY: Routledge.

Camilleri, M. A. (2020). The use of data-driven technologies for customer-centric marketing. International Journal of Big Data Management, 1(1), 50.

Chen, J. (2022, November 03). Business-to-business (B2B): What it is and how it's used. Retrieved February 25, 2023, from https://www.investopedia.com/terms/b/btob.asp

Cicerchia, S. (2021). The Digital Transformation in the Sneaker Market: Nike's In-Store Customer Experience (Unpublished master's thesis). Luiss Guido Carli.

David Kiron, P. (2012, September 18). Innovating with analytics. Retrieved August 14, 2022, from https://sloanreview.mit.edu/article/innovating-with-analytics/

Dholakia, S. (2015). Designing a Marketing Organization for the Digital Age (Rep.). Harvard Business School Publishing.

Facebook ad placements: How to choose the best one. (2021, November 17). Retrieved September 6, 2022, from https://klientboost.com/facebook/facebook-ad-placements/

Harper, B. (2018, September 18). Facebook ad frequency: How high is too high? Retrieved January 27, 2023, from https://www.socialmediatoday.com/news/facebook-ad-frequency-how-high-is-too-high/532559/

Higuera, V. (2016, October 26). Definition of data driven market research. Retrieved April 27, 2022, from https://smallbusiness.chron.com/definition-data-driven-market-research-38715.html

HubSpot. (n.d.). 2022 marketing statistics, Trends & Data - the ultimate list of Digital Marketing stats. Retrieved June 20, 2022, from https://www.hubspot.com/marketing-statistics

Irvine, M. (2022, December 14). Facebook ad benchmarks for your industry. Retrieved February 17, 2023, from https://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks

Kenton, W. (2022, October 29). Mobile marketing: Definition, how it works, and examples. Retrieved June 15, 2022, from https://www.investopedia.com/terms/m/mobile-marketing.asp

Kenton, W. (2023, February 14). B2C: How business-to-consumer sales works, 5 types and examples. Retrieved February 25, 2023, from https://www.investopedia.com/terms/b/btoc.asp

Keskin, S. (2022, December 19). 13 of the best product page examples we've seen (and why they work). Retrieved December 25, 2022, from https://www.drip.com/blog/product-page-examples

Khorev, M., & Mike Khorev I'm a growth marketing consultant who helps B2B. (2022, February 04). 12 effective digital marketing tactics and strategies in 2022 and beyond. Retrieved December 27, 2022, from https://mikekhorev.com/12-effective-digital-marketing-tactics-strategies

Kimball, R., & Caserta, J. (2004). The data warehouse ETL toolkit practical techniques for extracting, cleaning, conforming, and delivering data. Indianapolis, United States: Wiley.

Kotler, P., Kartajaya, H., & Setiawan, I. (2016). Marketing 4.0: Moving from traditional to digital (1st ed.). Hoboken, NJ, United States: Wiley.

L2M Rail. (2021). L2M Rail's path to Customer-Centric Research (CCR) and Customer-Centric Innovation (CCI) [Digital image]. Retrieved May 2, 2022, from https://l2mrail.com/wp-content/uploads/2021/05/image-1.png

Lau, K. (2017, May 11). What is data-driven marketing? - definition, examples and case studies. Retrieved June 3, 2022, from https://www.makethunder.com/data-driven-marketing-definition-examples/

Loshin, D. (2014, July 29). What is reference data? Retrieved November 27, 2022, from https://blogs.sas.com/content/datamanagement/2014/07/29/what-is-reference-data/

Marketing job description template: Linkedin Talent Solutions. (n.d.). Retrieved August 8, 2022, from https://business.linkedin.com/talent-solutions/resources/talent-engagement/job-descriptions/marketing

Milheiro, J. L. (2019). Next Best Action – a Data-Driven Marketing Approach (Unpublished master's thesis). NOVA Information Management School.

Patel, J. (2019). Bridging Data Silos using big data integration. International Journal of Database Management Systems, 11(3), 01-06.

Patel, K. (2017, October 27). What is e-commerce?: Types of e-commerce: E-commerce websites. Retrieved February 24, 2023, from https://www.internetconsultancy.pro/blog/what-is-ecommerce/

Powerful Digital Marketing Statistics and facts for 2023. (n.d.). Retrieved February 18, 2023, from https://bestwriting.com/blog/digital-marketing-statistics

Provost, F., & Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking. Sebastopol, United States: O'Reilly.

Raappana, H. (2020). Ethics in Data-Driven Marketing (Unpublished master's thesis). University of Vaasa.

Rego, B. (2022, September 08). Ux writing for e-commerce prices. Retrieved December 28, 2022, from https://uxcontent.com/ux-writing-e-commerce/

Sankaran, G. (2016). Data driven approach to account based marketing campaigns [Digital image]. Retrieved May 2, 2022, from https://www.latentview.com/wp-content/uploads/2016/07/Data-driven-approach-to-account-based-marketing-campaigns.jpg

Sarfin, R. (2022, November 03). 5 characteristics of data quality - see why each matters to your business. Retrieved May 10, 2022, from https://www.precisely.com/blog/data-quality/5-characteristics-of-data-quality/

Schneider, M. J., Jagpal, S., Gupta, S., Li, S., & Yu, Y. (2017). Protecting customer privacy when marketing with second-party data. International Journal of Research in Marketing, 34(3), 593-603.

Selman, H. (2021, January 25). Why we see Digital ads after talking about something. Retrieved August 20, 2022, from https://www.mcnuttpartners.com/why-we-see-digital-ads-after-talking-about-something/

Shopify (2022). All about shopify. Retrieved August 2o, 2022, from https://www.shopify.com/about

Solomon, M. R., Bamossy, G. J., Askegaard, S., & Hogg, M. K. (2016). Consumer behaviour: A European perspective. New York: Pearson.

Statista. (2021, July). Global Internet Advertising Revenue 2020-2025. Retrieved July 13, 2022, from https://www.statista.com/statistics/237800/global-internet-advertising-revenue/#:~:text=In%202021%2C%20global%20internet%20advertising,to%20488%20billion%20by%202025.&text=Internet%20or%20online%20advertising%20encompasses,media%2C%20display%20and%20mobile%20advertising.

Status. (2022, February 07). 4 best ways to breakdown data silos. Retrieved January 2, 2023, from https://status.net/articles/data-silos-information-silos/

Stedman, C., & Fredsall, A. (2021, October 04). What are data silos and what problems do they cause? Retrieved January 2, 2023, from https://www.techtarget.com/searchdatamanagement/definition/data-silo

Stelman, T. (2020, February 27). Customer-Driven Marketing VS Product-Driven Marketing. Where Do You Stand? [Web log post]. Retrieved May 4, 2022, from https://www.optimove.com/blog/customer-driven-marketing-vs-product-driven-marketing-where-do-you-stand

Strycharz, J., Van Noort, G., Smit, E., & Helberger, N. (2019). Protective behavior against personalized ads: Motivation to turn personalization off. Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 13(2).

Tarver, E. (2022, October 09). What is c2c? how does the customer-to-customer model work? Retrieved February 25, 2023, from https://www.investopedia.com/terms/c/ctoc.

Teneva, D. (2022, August 19). Repeat purchase rate for ecommerce brands. Retrieved February 22, 2023, from https://www.metrilo.com/blog/repeat-purchase-rate

Walker, K. L., & Moran, N. (2018). Consumer information for data-driven decision making: Teaching socially responsible use of data. Journal of Marketing Education, 41(2), 109-126. doi:10.1177/0273475318813176

ZLW. (n.d.). Organic seo vs PPC in 2021: CTR Results & Best Practices. Retrieved September 17, 2022, from https://www.zerolimitweb.com/organic-vs-ppc-2021-ctr-results-best-practices/?gclid=gclid&&gclid=gclid

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa