MASTER'S INTERNSHIP REPORT

PERSONALIZATION THROUGH A PROACTIVE LIVE CHAT IN AN E-COMMERCE: THE CASE OF BYSIDE'S CLIENT, A MULTINATIONAL RETAIL COMPANY

AUTHOR: Ricardo Nuno Teixeira Pimenta Marvão Mendes

SUPERVISOR: Moe Bagheri

PORTO'S SCHOOL, SEPTEMBER, 2021

PERSONALIZATION THROUGH A PROACTIVE LIVE CHAT IN AN E-COMMERCE: THE CASE OF BYSIDE'S CLIENT, A MULTINATIONAL RETAIL COMPANY

AUTHOR: Ricardo Nuno Teixeira Pimenta Marvão Mendes

Internship report presented to IPAM, to fulfill the requirements needed to obtain the Master's Degree in Global Marketing, developed under the scientific supervision of Professor Moe Bagheri.

ACKNOWLEDGMENTS

Writing a thesis is a challenging and demanding task, one that would have been almost impossible to accomplish without the constant support, motivation, and supervision of my thesis supervisor, Professor Moe Bagheri. Professor Moe, I would like to thank you for your hard work and patience in guiding me through this academic milestone, you were dedicated, motivational and a big support in this long journey.

I would also like to thank my superiors and colleagues at BySide who helped me get through this important chapter in my academic life. I would like to specially thank the ones that have been in closer contact with me regarding the development of this thesis.

Nevertheless, I extend my appreciation to my family and friends that were always by my side during all the steps, obstacles, and achievements. To my parents, siblings, and to my closest friends, who provided interesting discussions as well as happy and healthy distractions to rest my mind from this demanding research.

ABSTRACT

Retail e-commerce companies currently struggle in managing and optimizing the performance of a proactive live chat software application. It is assumed by companies present in the sector that providing personalized assistance to online visitors brings positive outcomes, however, there is no scientific evidence in this field to prove this assumption. This research aims to bring new insights into the contribution personalization can have regarding the performance of this app. Specifically, it investigates whether increasing personalization on the provided assistance to the online visitor has an impact on the number and value of influenced checkouts.

To test the hypothesis that providing more personalized assistance to the online visitor through this application leads to increased sales, the performance results of this app in the Croatian market of a multinational retail client were analyzed. Two five-month periods were observed, one providing non-personalized assistance and the other with personalized assistance for online visitors, the results of both periods were analyzed using three independent samples t-tests. The outcomes showed a statistically significant positive effect of the personalized assistance in the application performance results.

These results suggest that online visitors who received personalized assistance are more likely to proceed to the checkout funnel and complete the purchase and to perform checkouts with a higher value. On this basis,

personalization should be considered when managing or optimizing proactive live chat campaigns in retail e-commerce.

The thesis is finalized by outlining its limitations and proposing new avenues of research.

RESUMO

Atualmente, empresas no setor do retalho com lojas online têm alguma dificuldade em gerir e otimizar a *performance* de um *live chat* proativo, uma aplicação de *software*. É assumido, pelas empresas presentes no setor, que prestar assistência personalizada ao visitante online traz resultados positivos, contudo, não há evidências científicas nesta área que corroborem este pressuposto. Este estudo visa trazer novos conhecimentos sobre o contributo que a personalização pode ter relativamente à performance desta app. Especificamente, investiga se o aumento da personalização na assistência prestada ao visitante online tem um impacto no número e valor dos *checkouts* influenciados.

Os resultados da performance desta *app* foram analisados no mercado croata de um cliente multinacional de retalho, de modo a testar a hipótese de que oferecer assistência personalizada ao visitante online através desta aplicação resulta num aumento em vendas. Dois períodos de cinco meses foram observados, um prestando assistência não personalizada e o outro prestando assistência personalizada ao visitante online, os resultados de ambos os períodos foram analisados usando três t-tests para amostras independentes. Os testes realizados mostraram uma diferença positiva estatisticamente significativa do efeito da assistência personalizada nos resultados da *performance* desta aplicação.

Estes resultados sugerem que os visitantes online que receberam assistência personalizada têm maior probabilidade de procederem para o funil de checkout e completarem a compra e ainda de efetuarem checkouts de valor mais elevado. Tendo isto em conta, a personalização deve ser considerada quando se gerem ou otimizam campanhas de um *live chat* proativo em lojas online de retalho.

A tese é finalizada destacando as limitações desta e propondo novos caminhos para estudos futuros.

INDEX

List of Figures

List of Tables

List of Graphs

List of Appendices

LIST OF ABBREVIATIONS

1M€ - One million euros

ACEPI – Associação do Comércio Eletrónico e da Publicidade Interativa

APP – Software Application

B2B – Business to Business

BI – Business Intelligence

C2C – Click-to-Call

CCC – Cloud Contact Center

CSC – Customer Support Center

CDP – Customer Data Platform

CRM – Customer Relationship Management

CRO – Conversion Rate Optimization

E-commerce – Electronic commerce

eNPS – Employee Net Promoter Score

GDPR – General Data Protection Regulation

IT – Information Technology

NPS – Net Promoter Score

MQLs – Marketing Qualified Leads

ROI – Return on Investment

SD – Standard Deviation

SMEs – Small and Medium-size Enterprises

UX – User experience

GLOSSARY

Application – An application (app) is a type of software that bundles together certain features in a way that is accessible to a user. Software applications can be mobile apps or applications present on a website.

Banner – Banners, also known as banner ads or display ads, are digital advertisements displayed in the header, footer, or sidebars of websites.

Campaign – A digital marketing campaign is an online marketing action put forward by a company to drive engagement, conversions, traffic, or revenue.

Click-to-Call (C2C) – A digital communication solution on the website that allows the visitor to click a button or text to contact a Call Center line, to receive assistance during one's journey on the website. It provides a tailored experience for the website visitor, enhances mobile usage, and increases conversions.

Cloud Contact Center (CCC) – A CCC is an internet-based Contact Center that handles all inbound and outbound customer communications for a company.

Customer Data Platform (CDP) – A CDP is a collection of software that creates a persistent, unified customer database that is accessible to other systems. Data is pulled from multiple sources, cleaned, and combined to create a single customer profile.

Database – A database is an organized collection of data stored as multiple datasets.

Dataset – Structured collection of data generally associated with a unique body of work.

Data mining – Process of extracting and discovering patterns in large datasets involving methods at the intersection of machine learning, statistics, and database systems.

Digital Transformation – Digital transformation is the integration of digital technology into all areas of a business, fundamentally changing how you operate and deliver value to customers.

eNPS – Stands for "Employee Net Promoter Score", an employee satisfaction measurement scale.

Inbound Marketing – Inbound marketing is a marketing strategy that attracts customers by creating valuable content and experiences tailored to them, consists of entirely organic leads.

Influenced Checkout – (BySide's terminology) Checkouts made by online customers that were influenced by any application before the visitor made a purchase on the website.

Influenced Conversion – (BySide's terminology) Conversions made by online customers that were influenced by any application before the visitor made a purchase on the website.

Lead – A "lead" or "prospect", is a qualified visitor, a visitor already determined to be a prospective customer.

Lead generation – The marketing process of stimulating and capturing interest in a product or service to develop a sales pipeline, allows companies to nurture targets until they're ready to buy.

Lead activation – (BySide's terminology) Performance-based marketing approach that aims to reduce the percentage of website visitors that do not result in leads, through continuous efforts to target the right customers with the right message at the right time.

Live chat – Live chat is used by sales, marketing, and customer support staff to answer questions from online customers that seek assistance during their website sessions. It is a software application that provides live support to online visitors.

Marketing Cloud – Marketing Cloud is a digital marketing toolbox that unites customer data with customer behaviors in real-time to create more intelligent communications and interactions that respond to and anticipate customer needs.

NPS – Stands for "Net Promoter Score", a customer satisfaction measurement scale.

Omnichannel Strategy – Omnichannel marketing strategy is the integration and cooperation of the various channels that organizations use to interact with consumers, intending to create a consistent brand experience.

Outbound Marketing – Outbound marketing or interruption marketing is a marketing strategy that consists of the promotion of a product through continued advertising, promotions, public relations, and sales.

Software – Software is all the functional aspects of a computer that do not refer to its physical components (hardware). Scripts, applications, programs and a set of instructions are all terms often used to describe different types of software.

Trigger – A marketing technique that triggers the automatic sending of a message to potential or actual customers during their customer journey, for instance after a specific event.

1. INTRODUCTION

This chapter aims to provide a clear and concise contextualization of the entire thesis. The fundamental questions such as what, why, who, and how are answered in the following subchapters.

To start with, the question "What is this research about?" is responded by presenting an overview of the company itself and its client along with introducing the research problem and research question. "Why is this research worthwhile?" and "Who will benefit from the research?" are answered on the third subcategory by recognizing the significance of the present study and its limitations. Lastly, "How is the research conducted?" is explained with the thesis structure.

1.1. BACKGROUND

In the 20th century, technology revolutionized enterprises in the way they reach, interact, and manage their audience, particularly in the last decade (Ariguzo et al., 2006). This transformation came unannounced, suddenly companies needed a website, to master several tools, platforms, processes, and so on. Competition became fiercer and therefore the desire to stand out, to do more, and to do better became the main priority.

A few decades ago, with the emergence of the first e-commerce, large companies believed that to build a successful online store the entire customer journey should be 100% online, something that BySide has not believed in since the early days (Magalhães, 2020). To this day, the company believes that enterprises should aim to combine the online channel with other channels, having an omnichannel strategy.

BySide, appeared in 2006, being officially founded in January 2007, disrupting online banking markets with a simple product, a digital solution that would improve online UX, named click-to-call, firstly adopted in Portugal by Cetelem, a commercial brand specialized in consumer credit, owned by BNP Paribas. The CEO of BySide, Vitor Magalhães (2020) claims that at that time, this was a new revolutionary product that probably set BySide as the first company in Europe to place a click-to-call on a banner of a website.

Due to the huge success of the click-to-call solution, the company started to gradually grow during the following years, from financial terms, achieving rapidly 1M€ of annual revenue, to internationalization, having clients all over the world that provide more than 50% of the revenue compared to national clients. Since then, BySide has been distinguished with several awards, among them:

- Referenced twice in Deloitte’s Fast 500 list 2012 & 2013
- Gartner’s Cool Vendor 2014 & Gartner’s Magic Quadrant reference 2015
- NOS Innovation prize 2015
- EDP Partners, Winner on the Client Relationship category
- Won twice ACEPI’s award in the E-commerce category

During the past 15 years, BySide has expanded and opened offices in Madrid and São Paulo and grew its manpower from 3 to more than 80 employees. Four years ago, it also founded Bytalk, which is a brand specialized in Digital Transformation, building SME's digital marketing foundations.

Nevertheless, BySide continues to strive for greatness. Some of its long-term goals can be found at "BySide Vivid Vision 2022", which is a document developed in 2019 to forecast the year 2022. By 2022, the company projects to be at the forefront of Digital Optimization for corporate brands in Telecommunications, Financial Markets, and Retail, reach 20M€ of annual revenue, grow the team to 140 members, strengthen the presence outside Europe, particularly in Asia and South America, and aims to attain an annual eNPS of above 80% and an NPS of over 70%.

All things considered, even though the current internship is being held in BySide, this case portraits a business problem and its solution of a BySide's client, a multinational company present in the retail sector. The name of the client will not be mentioned in this research to be GDPR compliant.

1.2. Research Problem and Research Question

BySide has managed to position itself as a respected, reliable, and successful "lead activation company", focused on increasing conversions through CRO marketing, and providing strategic guidance to corporate enterprises.

The concepts of lead generation and lead activation are often confused. Lead generation, the most known terminology, is the process of converting web users into visitors and then into leads, by transforming a web user that is simply browsing the web to a website visitor and then to a more probable client, one that engaged, or merely indicated interest in any way (lead) with the company's blog post, live event or any other piece of company's content. Lead generation marketing is the process of generating Marketing Qualified Leads (MQLs), legitimate and sincere clients that intend to buy (Świeczak & Łukowski, 2016).

On the other hand, lead activation is where BySide steps in and presents to the market an original terminology that has gained notoriety over the years. Since its inception, BySide defines lead activation as a more performance-based marketing approach that aims to reduce the percentage of website visitors that do not result in leads, through continuous efforts to target the right customers with the right message at the right time (BySide, 2016). It is the process of engaging with visitors who have your ideal client profile before they've made a purchase, purposefully engaging more efficiently to increase their probability of becoming a high-quality sales opportunity (Imagine, 2018). The goal is to maximize revenue originated from website visitors, to focus on lead quality and optimizing ROI rather than simply converting more visitors into leads.

To do so, BySide divides its service (lead activation) into two main branches: Marketing Cloud and Cloud Contact Center (CCC).

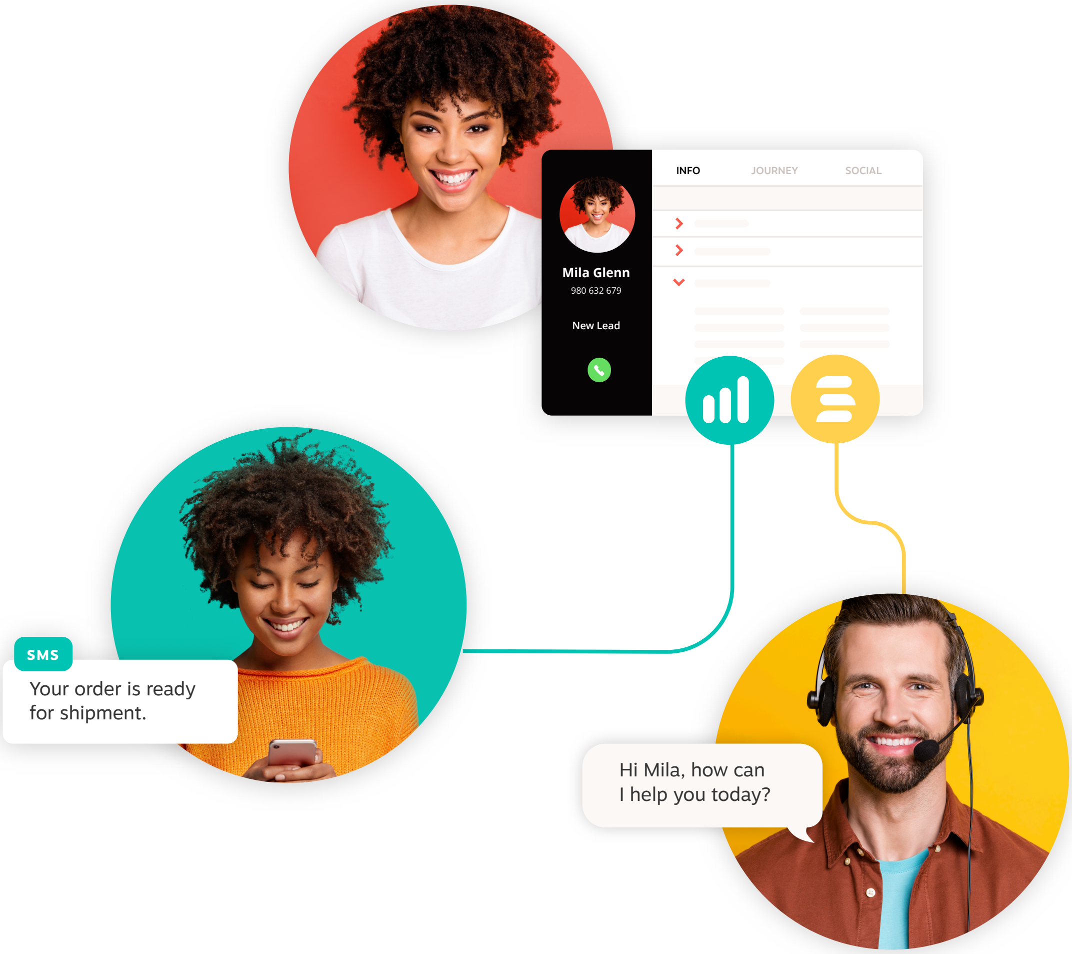


Figure 1: Lead Activation with Marketing Cloud and Cloud Contact Center

Source: (BySide, 2020)

BySide's Marketing Cloud is a real-time integrated suite, a digital marketing toolbox of multiple marketing tech solutions, where it is possible to score and qualify leads, build audiences, personalize, and optimize experiences on the web and mobile applications (Apps). It also collects customer data from any device, centralizes and arranges it on BySide's Customer Data Platform (CDP), a platform that aggregates and arranges raw data so it can be then transformed into usable data. As the Customer Data Platform Institute (CDPI) puts it, a CDP is “a marketer-managed system that builds a unified, persistent customer database that is accessible to other systems”. CDPs collect data from different sources to provide a more detailed overview of the customer (Earley, 2018). This aggregated data can be acted upon in real-time, a great way to personalize user interactions, while staying GDPR compliant.

On the other hand, the other main branch of the Lead Activation service provided by the company, CCC, is the central point where all inbound and outbound customer communications are managed from and to different channels. With Marketing Cloud and CCC, BySide allows their clients to collect information from multiple platforms, control their online business actions in just one place and even share their personalized reports wherever they desire.

BySide looks at its clients as partners working towards the same goals. It does not simply produce; the company has a strong market experience that enables it to choose the clients that want to steer in the same direction. With BySide, enterprises can have a strategic data-driven approach outlining their clients’ online performance goals.

This thesis is developed based on the objectives of one of BySide’s clients to take its e-commerce to the next level by providing more and better customer support in their main online channel. Following a careful evaluation of its platform, BySide’s solution was requested and in 2017 the partnership was made, launching an alliance with a subsidiary of a multinational retail company by providing a live chat app, a real-time customer support chat connecting online customers with customer support representatives live, in their e-commerce (Elmorshidy, 2013).

Considering that this retail company is present in many countries and diverse markets, BySide needs to constantly evaluate each market's performance, forecast possible increase, or decrease in online traffic, and optimize the proactive live chat app to its full capacity. Taking that into account, the general **research question** for this thesis is:

- "*How can a retail e-commerce increase the number of chat-influenced checkouts?*".

Important to highlight that chat-influenced checkouts are considered, by BySide's client and in this research, checkouts made by the e-commerce customers that were influenced by BySide's proactive live chat app in the last 7 days before the visitor made a purchase on the website. In other words, the customers that contacted the Contact Center using the chat app and afterward made a purchase, BySide then matches the two events (contact request and checkout).

1.3. SIGNIFICANCE OF THE STUDY AND ITS LIMITATIONS

There is some existing research regarding e-commerce, fewer studies focus on online assistance and no research was found that deep dived solely into how personalization can improve the live chat application performance. In this thesis, live chat personalization will be its cornerstone.

The main goal of this thesis is to explore how a proactive live chat, an online assistance tool, should be utilized, managed, and improved. The aim is to not only provide a valuable contribution to the scientific world and unveil some results that have never been explored but also to help companies in the retail sector prosper by improving their online customer support management and enabling them to optimize their live chat app to create a better, evolved UX for their customers. It is assumed that these improvements will also have positive repercussions on customer happiness, satisfaction, loyalty, engagement, brand awareness (due to referral marketing), brand intimacy, and so on.

Brand intimacy is considered to be one component of the customer-firm relationship, referring to the customer-brand connection (Yim et al., 2008). The concept of brand intimacy emerged with Sternberg's interpersonal triangular theory of love in an interpersonal context, in 1986. Since then, some researchers regard brand intimacy as a construct of love that "transforms the interaction from an instantaneous, transactional exchange to a strong and enduring relationship" (Yim et al., 2008, p. 742).

Nevertheless, every research has its limitations and this one is not an exception. This study recognizes that in 2020 and 2021, due to the global Covid-19 pandemic, there are indisputable changes in the online environment that directly influences the research results. According to Statista (2020), a 7% increase occurred in online traffic worldwide in October 2020 compared to January 2020. Concerning the retail sector, there was an increase of 36% in June 2020 of the average monthly visits on retail e-commerce website traffic worldwide compared to the same month in 2019 (Statista, 2020). Taking that into account, it can be assumed that customers tend to use more online shopping to follow the imposed restrictions by the governments, which naturally increased the website traffic of the online stores resulting in a need to upgrade companies' e-commerce.

On one hand, this increase in online traffic became inconsistent, since during the past and present year (2020-2021) the lockdown in different countries was asynchronous and government-dependent, making them unpredictable. This turns out to be an additional challenge to evaluate different periods within a country.

On the other hand, this unfortunate virus caused an acceleration in companies' digital transformation/evolution, as KPMG (2020, p. 3) puts it in their Global retail trends 2020 report "Interestingly, COVID-19 has accelerated key fundamental trends that were already influencing the sector — rather than stop these trends in their tracks." Provoking a huge opportunity to evaluate the new consumer trends, patterns, and habits appearing within the market and making it even more imperative to improve online customer support.

Furthermore, this is a specific case of a company present in the retail sector; therefore, the findings of the developed research should not be interpreted as an industry standard.

1.4. THESIS STRUCTURE

This document has a quantitative methodological approach and is composed of six chapters and their respective subsections. As with any academic thesis, these chapters structure the introduction, development, and conclusion of the thesis. These are the established chapters:

- Chapter one is the Introduction. It introduces the case framework, presents its background providing context to the reader, states the general research question, its significance, and limitations.

- Chapter two provides the literature review that is composed of three main subchapters: E-commerce, Business Intelligence, Marketing Automation. These linked study fields were selected to address this company's case research more easily, providing information regarding what was already researched and where is the research gap.

- Chapter three presents the Research Methodology and Methods. This chapter outlines the methodological approach and the roadmap of how the research is carried out, including the primary data collection techniques and the secondary data collection process.

- Chapter four describes the analysis and data presentation. It presents the research findings after an in-depth analysis of the collected data.

- Chapter five is the Discussions part. It explains the main outcomes of the findings, also providing recommendations for companies present in the same sector.

- Chapter six presents the main conclusions of this thesis. It provides recommendations for future research opportunities.

2. LITERATURE REVIEW

To achieve the objective of the current thesis, the following three main research themes are selected to take a deep dive into the topic under investigation.

E-commerce aims to collect every significant research paper approaching online commerce and its specificities, provide a more detailed overview on online customers, online assistance, and what has been explored so far by academic researchers in these topics.

Secondly, **Business Intelligence** is a mandatory theme to be explored in this thesis since it will analyze the outcome of collected data from BySide's client, which will be further explained in the Methodology chapter.

Finally, to present a complete and valuable Literature Review for the present study it was of best interest to explore **Marketing Automation** scientific research papers.

2.1. E-COMMERCE

Electronic commerce or e-business, from now on referred as e-commerce, refers to the online commerce, a transaction of goods and services through electronic communications (Tian and Steward, 2008).

According to Ariguzo, Mallach and White (2006), in the early 90s the World Wide Web gained notoriety and changed the world forever. While it is open for debate where it all started, one thing is certain, the first documented secure online retail transaction occurred in 1994, transforming people's view on commerce. Since that day, it was possible to shop online, people could simply buy products without leaving their homes, just a click away.

At the turn of the century, business transactions made online through this new channel, e-commerce, had an exponential growth, and since then it is remembered as "the most dramatic communications growth period in history" (Esch, 2002). In the following years, e-commerce became a global sensation, the business opportunities in it were indisputable. Afterwards, research studies started to appear addressing the most various themes within e-commerce, its consumers, the digital transformation required for the physical companies and so on.

The year 2020 came and taught humanity some valuable lessons on how to never doubt the resilience of humankind, their ability to overcome adversities and the capability of transforming digitally entire industries to keep moving forward.

In June 2019, Statista released some data of retail e-commerce sales worldwide from 2014 with projections until 2023. In that market study, it demonstrated the exponential growth of e-commerce sales and projected, in 2023, a growth of more than 85% of electronic commerce compared to 2019, almost doubling online sales attributed to e-commerce worldwide.

However, Covid-19 stroke in the first quarter of 2020 throughout the world forcing most countries to declare national lockdowns. In the UK, in May 2020, internet retail sales represented roughly 33% of total retail sales, an approximately 14% increase compared to May 2019, when internet retail sales represented “only” 19% of the total retail sales (Statista, 2020). Yet, the global retail e-commerce sales decelerated due to the global pandemic in 16,5% of the annual growth rate, most markets will register a strong growth, however India and China will experience a substantially deceleration compared to previous years (emarketer, 2020).

Digital adoption grew, especially in European countries where it went from 81% to 95% (considering McKinsey Digital survey to more than 20k Europeans consumers from April 28 to May 20, 2020), mainly due to Covid-19 pandemic, a rise that would have taken otherwise two to three years to happen (Fernandez et al., 2020). At the same time, a portion of retail companies were caught unprepared for the increased number of visitors they were receiving, pushing them to reinforce their e-commerce customer support.

Thus, the need of online chat assistance rose. Although, there were some requirements regarding this feature. First, it is important to bear in mind that there are multiple types of customer support chats that will be further described, and second Contact Centers (CCs) have a limited number of customer support/sales agents, an obstacle to assist more visitors with the chat solution but also, an opportunity for other technological solutions, such as the chatbot.

2.1.1.Online consumer behavior during Covid-19 Pandemic

The year 2020 was a peculiar year, it started off with the Covid-19 pandemic, a result from the coronavirus, a contagious disease first detected in Wuhan, China, in November 2019 (Live Science, 2020) which spread like wildfire throughout the globe due also to the ever-evolving globalization, reaching the seven continents, contaminating, and killing millions of people. This Corona virus SARS-CoV-2 burst out a global pandemic that began in March 2020, forcing countries to implement restrictions to reduce the spread of Covid-19. As previously mentioned in the E-commerce chapter there was a significant increase in online traffic thanks mostly to the national lockdowns some countries declared, forcing people to stay at home and shop online instead of going to the local markets and shopping malls. This subchapter will collect and present some important articles that cover the registered changes in consumer behavior during this global pandemic to better understand the consumers further analyzed in this study.

According to Sayyida, Hartini, Gunawan and Husin (2021) in "*The Impact of the Covid-19 Pandemic on Retail Consumer Behavior. Aptisi Transactions on Management*" there are four consumer behaviors according to the types and needs of the consumers in shopping: pure offline shopping; showrooming; webrooming and pure online shopping.

Pure offline shopping is the traditional process of physical shopping, contrasting with the **Pure online shopping** that are consumers with an entire online customer journey. **Showrooming** and **Webrooming** combine both physical and online shopping, showrooming is the consumer behavior of someone who seeks information on services or products offline and later makes the purchases, subscriptions and so on in the online channel. Webrooming is the exact opposite of showrooming, it combines both channels, but it refers to a consumer who search for every possible information of a product/service online just to then make the purchase on a physical store (Sayyida et al., 2021).

Nevertheless, these four types of consumer behavior were influenced due mostly to the global pandemic in 2020, pure offline shopping, showrooming and even webrooming had to stop in countries where national lockdowns were declared, no one could gather information from products/services physically or buy anything since all stores were closed and entire countries populations were confined into their homes. United Kingdom, Germany, and France, third, fifth and sixth largest e-commerce markets worldwide respectively, increased their online retail sales from total retail sales in 2020 comparing to the previous year, by 3,8%, 2,3% and 3% respectively, reaching 27,5%, 11,2% and 12,2% (Sayyida et al., 2021).

These factors only reinforced the need to invest in an improved User Experience (UX) to the consumer to increase the online retail sales, as Safara (2020, p. 2-3) puts it "in a successful e-commerce application, understanding consumer behavior requires identifying reasons that encourage consumers to buy from the web-sites, however, identifying consumer behavior and encouraging factors is difficult (Agrawal et al., 2018)". It is important to listen to what the online visitor has to say and start to understand their pains and desires to better serve them in the future.

In that line of thought, it is imperative to acknowledge that there are different online shoppers with different needs and doubts. Ideally, as presented in the article "*Ideal types of online shoppers: a qualitative analysis of online shopping behavior*" there are five types of online shoppers: Conservative shoppers; Rational Shoppers; Hedonistic Shoppers; Spontaneous Shoppers and Vanguard Shoppers (Kettunen et al, 2018).

Conservative shoppers are people who want to see and feel the product before purchasing it, physical stores are their main source of purchases, online shopping is only an option when the price is lower online, for example. Like the consumer habit previously mentioned "showrooming", these shoppers have low online shopping activity, however, when they are active, they have more probability in purchasing the items. They want an to browse the website easily and in a simple way, if the website is unclear or seems suspicious in any way the purchase will probably not be made, simplicity and clarity are key for this type of shoppers (Kettunen et al, 2018).

Rational shoppers are active online shoppers, although as the name suggests, are rational buyers, that take some time to decide what to buy after careful consideration and analysis of the product/service. They are not people who buy on the first website visit, it might happen but even after finding a suitable product with an appropriate price they may take some time to decide and go through with the purchase. "Rational shoppers tend to give and take feedback regarding online shopping, but only if the service or the product has been really good or really bad. They also appreciate having an easy and suitable way of giving feedback to the online shop." (Kettunen et al, 2018, p. 8-9).

On the contrary, **Spontaneous shoppers** are very active impulsive shoppers, they are emotions driven, for them, online shopping is easy, simple, and mostly the fastest option to shop whenever they feel like it. "Personalized messaging is seen as a service and an opportunity rather than being annoying." (Kettunen et al, 2018, p. 10). Spontaneous shoppers appreciate clear websites, easy to navigate to find the information they seek without spending effort searching for it (Kettunen et al, 2018).

The **Hedonistic shoppers** value the remote experience online shopping enables them to have, these people appreciate easiness, they often shop in online shops that offer easy return policy, free delivery, and flexible payments methods, they prefer familiar shops to them since it reduces the risk level. Online shopping brings pleasure for itself to these shoppers, even if they do not complete a purchase the experience of searching is already satisfying. Online shop membership in shops that meet their needs and treat them accordingly is something valuable for these loyal customers (Kettunen et al, 2018).

Finally, the **Vanguard shoppers** are comfortable in online shopping, they value the better product selection, and the better platform online shopping has for information search and comparison. These shoppers return to the same shops due to previous good experiences, the products quality, and the high level of website applications. Also, online recommendations regarding the shop service and applications are viewed as important, they are considered to be experienced shoppers that trust their judgements and instincts however they are willing to follow online trend setters as fashion bloggers in order to stay up to date (Kettunen et al, 2018).

Considering different online shoppers and their distinct behaviors in their customer journey, to have personalized online assistance is a most, to address the customer in the more accurate way possible, topic further approached in the next subchapter.

2.1.2. Proactive Live Chat

The chat application evolved over the last two decades and now there are different types of online chat solutions and even new ones that emerged over time, such as the chatbot, that will be further explained and defined in this section.

To start with, there are reactive and proactive chats. On one hand, **reactive chat** is a chat that needs to be responded urgently to users' traffic requests as soon as they arrive to the Customer Support Center (CSC), causing outages (Bastug et al., 2014). On the other hand, **proactive chat** is a chat that proactively offers assistance based on some predetermined rules (**triggers**) (Maloney & Kemp, 2015).

According to Estela Viñarás (2018), in "*What is Trigger Marketing and when can we apply it?*" Cyberclick blogpost, **trigger marketing** is a marketing technique which activates automatic messages to be delivered to a customer, or a potential customer, in a specific moment of his/her customer journey. This marketing technique is driven by "**events**", a user action (subscription, user goes to a specific webpage, comments a blogpost and other possible actions) or even a specific day (Christmas, Easter and so on). The ultimate goal is to maximize the personalization and to nail the perfect timing with the **NBO** (next best offer), in order to increase the conversion rate. NBOs are highly personalized offers by a company with the goal to redirect the visitor to the right product, service or information at the right time, with the right approach (using a convenient channel) and deal (sharing a promotion) as Bhagyeshwari Chauhan (2017) explains, in "*Next Best Offer: Predicting Your Customer's Wants Before They Do*" Datahut blogpost.

In 2014, John Carroll University developed and set a proactive commercial chat software product that boosted chat usage for the business sector. The software was implemented to increase chat usage in the library website and the outcome was remarkable, it resulted in an increase from 3% to 21% of the answered chats rate, over a six-month period (Zhang & Mayer, 2014).

Live Chat is another distinct online type of service from customer support services such as telephones and emails (Elmorshidy, 2013). It is a medium available for customers in some online stores to contact directly and in real-time sellers with their questions, doubts, or intentions to purchase a product or service (Kang et al., 2015).

Nowadays, customers are more demanding and prefer to resolve their issues immediately rather than waiting for an answer, something that **live** chat came to fulfill (Elmorshidy, 2013). Online shoppers transitioned from passive consumers to proactively engaging consumers wanting to converse with a customer support agent in real-time (Kang et al., 2015). Live chat solution allows customer support agents to satisfy consumer needs by providing valuable and adequate information, erasing products skepticism, or shortening the perceived distance between the agent and the customer (Lv et al., 2018).

2.2. Business Intelligence

In the past several decades, enterprises have evolved at a faster pace than ever, driven by globalization, innovation and, of course, technology. Taking that into account it has also grown within, the competitivity in the market and the difficulty to stand out from the crowd for a corporation.

Business Intelligence, also known and from now on referred as BI, emerges as an outcome of companies wish to be better, faster, and bigger. This field of study deals with strategic thinking, decision-making and guidance when it comes to steer an enterprise.

Nowadays BI is defined as <u>neither a product nor a system</u> (Moss & Atre, 2003), it is a process used for the collection, integration, analysis, and presentation of business insights through multiple data tools enabling data collection and visualization, explained in greater detail further. "An important IT framework that can help organizations managing, developing and communicating their intangible assets such as information and knowledge" (Alnoukari et al., 2012, p. 1). Above all, BI eases the decision-support applications used by enterprises such as data mining, business analysis, cross-functional decision-support databases and predictable/statistical analysis (Moss & Atre, 2003). A software that integrates data from multiple data sources of the company, analyzes them and presents timely verdicts in the form of reports and dashboards (Cristescu, 2017).

The goal is to produce valuable information to the manager or board member to take clearer, data-driven decisions in how to achieve a certain goal, milestone, or position in the market (Cristescu, 2017). The author concludes by claiming that the decisions that outcome from these reports are infinite, it can regard competition, customers, markets characteristics or even the company figures, such as total revenue.

2.3. Marketing Automation

The term "Marketing Automation" sounds somehow futuristic; however, it has been around since 2001 thanks to John D.C. Little, who first introduced the terminology at the 5th Invitational Choice Symposium UC Berkeley 2001 (Little, 2001).

Marketing automation is the use of software to automate marketing activities and processes such as customer data integration and campaign management (Todor, 2016). It is used to increase efficiency within the marketing department activities, to allow better decision-making with data-driven decisions and to provide a more personalized experience to their customers resulting in an increase of customer satisfaction and loyalty (Bucklin et al., 1998). The main idea behind this concept is to adaptively react to customer choices on the web (Heimbach et al., 2015). Ultimately, marketing automation merges software and strategy, it allows companies to nurture prospects with highly personalized, valuable content that helps increase customer interest, in order to transform an interested visitor into a delighted customer (HubSpot, 2021).

The core of the concept is the automatic personalization of marketing activities with its foundations in the business-to-business (B2B) area (Heimbach et al., 2015). In B2B businesses, marketing and sales departments manage their customers or accounts by addressing their customers with a one-to-one approach (sometimes using CRMs, a concept further explained), providing proximity and nurturing their interest in personalized offers, as BySide does. However, these personalized offers are usually made in corporate environments (B2B), since in business-to-consumer (B2C) the company is selling a product(s) or a service(s) to a wider audience, as BySide's client, present in the retail sector, does. So, to provide a personalized customer experience to their clients the company cannot have a one-to-one approach, it would not be viable due to this costly personalized communication (Heimbach et al., 2015). To go around the topic, B2C businesses can apply marketing automation to their campaigns, for example, to address customers with more personalized content/offers, increasing their interest and involvement with the brand since they are now receiving more tailored, relevant, and valuable information (Dijkstra, 2008). Therefore, companies that use marketing automation probably enhance their retention rate, conversion rate and are even able to increase their cross and up-selling sales (Heimbach et al., 2015).

A Customer Relationship Management platform, from now on referred as CRM, is an application for businesses to gather and organize customers data to assist sales, marketing, and customer service departments to better understand their customers desires and needs (Todor, 2016). CRM enables companies to deliver the right message at the right time. Marketers can use CRMs to identify the most profitable customer relationships, optimize communications from the customer's point of view to maximize customer value and increase customer loyalty (Allen et al., 2001).

Marketing Automation appears to help and improve direct marketing, email marketing, CRMs, SMS marketing and so on. It can be used to increase efficiency in limitless marketing activities, areas, and tools by allowing companies to communicate in real time to their customers in every desired device (website, social media account, app, etc.) (Heimbach et al., 2015).

This information technology (IT) tool is attracting more and more attention in recent years with vendors of these software (for example: Eloqua, HubSpot, BySide) alleygating that enables companies to align their marketing and sales system interfaces to level-up their lead qualification processes via "lead scoring and nurturing", for example (Järvinen & Taiminen, 2015). **Lead scoring** is the process of prioritizing which customer leads should be preferably targeted, depending on several activities performed by potential customers (leads), this can include their journey within the company's website or other characteristics that place the lead as being more "valuable" to the company (Nygård & Mezei, 2020).

Basically, without a marketing automation tool, there is manual lead scoring, consisting in the process of scoring each activity taken by the lead or characteristic an importance score, the leads with highest overall scores will be then approached by salespeople (Nygård & Mezei, 2020). Furthermore, machine learning appears to be a viable alternative to this manual process, transforming lead scoring into a fully automated process present in some marketing automation software (Nygård & Mezei, 2020).

Machine learning (ML) is a research field within computer science field umbrella, a type of Artificial Intelligence (AI) which provides machines the ability to learn without explicit programming from a human being (Mahdavinejad et al., 2018). ML can also be defined as computational methods which uses experiences to improve the machine performance and provide more accurate predictions in the future (Mohri et al., 2018). All in all, "Machine learning consists of designing efficient and accurate prediction *algorithms*" (Mohri et al., 2018, p. 1). The most used ML algorithms are association rule, genetic algorithm, decision tree, neural networks, K-nearest neighbor and linear as well as logistic regression (Nygård & Mezei, 2020).

In Nygård and Mezei (2020) research, there were no comparisons between lead scoring using ML and manual lead scoring, so they were not able to affirm with certainty which one is better. However, with machine learning-based lead scoring models they presented a viable alternative to the previous process, allowing the creation of enhanced CRMs. Hence, taking a step further into a more data-driven and automatized Marketing era.

3. RESEARCH METHODOLOGY

This chapter is divided into four sections.

Firstly, Methodological Approach section aims to provide an overview of the entire research, present the research problem, research question and the research hypotheses.

Secondly, Data Collection Methods displays the type of data collected and further analyzed, and which sampling method was used.

Thirdly, Data Analysis describes how the researcher processed and analyzed the collected data, how the data is prepared, which software was used, and which statistical tests were conducted to test the hypotheses presented in the first section of this chapter.

Lastly, Justification and Contributions of the research explains why this specific methodological approach was selected, this research contributions and avenues for further scientific research.

3.1. METHODOLOGICAL APPROACH

To start with, it is essential to review some definitions to clarify and define the methodology of this thesis. **Research** literally means "to search again" and is defined as one of the main tools to provide new knowledge to the world, the market, the economy, or another area of uncertainty, that covers a long range of phenomena (Zikmund et al., 2009). However, business research is more specific, it is considered to be an essential tool for problem-solving and decision-making activities in management (Zikmund et al., 2009). The present research is a **descriptive research**, a specific type of scientific research with the main objective of describing, as the name says, characteristics of environments, organizations, objects, groups, or people (Zikmund et al., 2009).

Methodology is a concept enhanced over the years (Saunders et al., 2009). Mark Saunders, a PhD Professor of Business Research Methods in the School of Management at the University of Surrey, and a famous author of several scientific articles on research methods, is probably one of the few people comfortable in defining it. To him, methodology refers to the theoretical plan of how the research is handled (Saunders et al., 2009).

This research will be developed on a secondary data basis, it will collect data that has been gathered for other purposes and analyze it. It will have a **quantitative methodological approach**, meaning that it will be based on numbers and statistics rather than words and interpretations. A quantitative research tests objective theory by evaluating the relationship among variables (Creswell, 2009).

BySide, among other services and solutions, provides personalized proactive live chats to various clients including e-commerce platforms. These solutions are utilized to logistically optimize their usage, achieve a higher assisted customer satisfaction rate, influenced checkouts rate, among others. Since the retail clients have some challenges in minimizing the checkout abandonment rates of their e-commerce and improve UX with customer support along with the insufficient scientific articles in this area, the **research problem** can be defined as "*The effects of providing personalized assistance to e-commerce visitors through a proactive live chat application have not been closely studied*".

The **research gap** discovered on the reviewed literature is about *"How companies can optimize their proactive live chat tool to increase sales through personalization?"*. This resulted in the formulation of the **general research question** that will be further investigated "*How can a retail e-commerce increase the number of chat-influenced checkouts?*".

To better structure this study, **research hypotheses** were established to be accepted or refuted latter on:

⇒ **H_1**: *When an e-commerce business in the retail sector provides a more personalized approach to the visitor, through the proactive live chat application, it will result in more influenced checkouts.*

⇒ **H_2**: *With a more personalized approach to the online visitor the conversion rate of influenced checkouts over the number of contacts made will increase.*

⇒ **H_3**: *With a more personalized approach to the online visitor the average value of the influenced checkout will increase.*

3.2. Data Collection Methods

Quantitative methods, such as the ones used in this research, are the traditional research approaches which collects and analyses data, makes interpretations, and presents the results of the research in a consistent way. Contrary to qualitative methods, which collects open-ended data, analysis images and text, and presents personal interpretations of the findings (Creswell, 2009).

In this quantitative research, the entire population that contacted the Contact Center through the chat app will be analyzed. **Sampling methods** will not be used since the data gathered regards the entire population under scope and not a sampled group from it.

The data of the selected case will be drawn from one of BySide's platforms, internally entitled "Backoffice", that gathers detailed information from each of the multiple markets of this e-commerce client. The present study will evaluate if there is a significant difference between two periods of 5 months within one market of this e-commerce client, the Croatian market, where it has been implemented a more personalized approach to the online visitors in this app. This market was selected considering the consistency of the implemented live chat campaigns throughout each period.

In the first five-month period, 13th of March to 12th of August of 2020, there was only one live chat campaign active, the *Time elapsed (2s)*, that was present in several webpages of the e-commerce. This campaign activates the chat in a 2 second timeframe, the lower time possible taking into account that the service needs to be loaded on the visitor's webpage, meaning that there was no *personalized* approach to the online visitor at the time, the chat was simply popping up to the visitor in selected pages.

Between the two time periods, this market implemented more chat campaigns, one by one, providing a more personalized approach to the online visitor and allowing the CC to better evaluate which skill (specialized group of operators within a CC team), would suit best for each contact request. At BySide and in the current study, a **contact request** happens when the online visitor responds to the first standard message or, as BySide calls it, the Welcome message (ex: Hello, is there anything we can assist you with?) that is presented to them in the chat window.

On the second five-month period that we will be considering in this study, 1st of November to 31st of March 2021, it was added, comparing it to the first five-month period, three more chat campaigns (*Leaving the Checkout*; *Delivery Cost*; *Cart value)* with distinct behavioral triggers to activate the chat to better filter the question/request that would be made by the visitor. According to Phrasee blogpost "*What is: behavior triggered email marketing*" (2017), behavioral trigger email marketing refers to emails automatically sent to each lead based on their behavior or even their lack of behavior sometimes. In the present study, behavioral trigger chat marketing prompts or offers a chat automatically considering the lead behavior. With these three new chat campaigns, more tailored to each lead, this five-month period is considered to have a more personalized approach to the e-commerce visitor, targeting different costumers in different times of their customer journey.

The collected data will be further analyzed comparing the two time periods of the Croatian market and their evolution throughout the total observed time. Cross-country observation will not be made, since only one country will be evaluated and each country has their own demographics characteristics, such as total population, shopping habits, and purchasing power.

3.3. DATA ANALYSIS

Data Analysis section is where it presents the type of statistical analysis that will be undertaken, which software is going to be used and which statistical test will be performed.

The case of BySide's client will have the investigation based on numbers and as a part of the data preparation process for more in-depth analysis, the information gathered will contain the outliers present in the populations and will be double-checked for missing data. The software that will be used to conduct the analysis is **SPSS**, giving the ability to perform **t-tests** and to test the hypotheses.

Apart and before conducting t-tests, the data will be described in a detail manner to understand the results further.

3.4. Justification and Contribution of the Research

The methodological approach chosen was a result of three main reasons. First, it would not be viable to observe more markets where the client is present since thousands of people will be already analyzed and other markets did not have consistent activations and deactivation of chat campaigns as this market had. Secondly, it is not feasible to retrieve smaller samples from the population since there is no demographic data from the observed leads. Thirdly, the presented case is GDPR compliant, so it does not compromise any information from the company or from its e-commerce customers.

Furthermore, quantitative approach is selected because the client's data is already available in the forms of numbers and values that justifies the chosen research method. It was not possible to collect data from the client's customer through interview and surveys. Hence, qualitative methods were not applicable to this research.

To sum up, the present research will provide **new knowledge to the scientific world** since there is no published research investigating solely personalization within the live chat app of an e-commerce and its advantages and disadvantages for the management of this tool. Additionally, this thesis will be a valuable source of gathered information regarding this topic, since secondary data from this specific multinational company in the retail sector will be collected, compiled, and presented. All in all, it will **help companies in the retail sector on how to manage their proactive live chat apps and improve the UX they provide to their online customers**.

4. FINDINGS

This chapter will cover the conducted analysis, firstly it will present a descriptive analysis of the collected data and secondly the implemented test and its results to unveil the findings of this study. To structure the following steps this chapter will be divided into two main parts: **Descriptive analysis** and **Results**.

The first part, descriptive analysis, will not present demographic information regarding the observed populations since this study, as explained previously, is GDPR complaint. In other words, each lead or chat request in the e-commerce client will not be described regarding his/her personal data. This subchapter will describe some important analysis taken between different periods of times that will be considered in the Croatian market of this retail client.

Then, the second part will present the study findings through three T-tests performed with SPSS, the chosen software to conduct the statistical tests.

4.1. DESCRIPTIVE ANALYSIS

Descriptive analysis is, as Zikmund, Babin, Carr and Griffin (Zikmund et al., 2009, p. 481) describe, "the elementary transformation of data in a way that describes the basic characteristics such as central tendency, distribution, and variability." In this chapter, the analyzed data will be presented and described.

This study compares two distinct time periods that used two different strategies in managing these proactive live chat applications, furthermore these periods will be described in a more detail manner.

The first five-month period, from 13th of March to 12th of August 2020, had solely one active chat campaign, the *Time elapsed (2s)*, this chat campaign activates the chat without considering any behavioral characteristic of the visitor. Basically, it would activate the chat as soon as the online visitor enters a webpage (2 seconds is the minimal default time possible since BySide service needs to be loaded on the visitor end) of the group pages present in this campaign. Furthermore, this campaign performs two validations before offering a chat, as any other campaign further analyzed in this study. These validations were made to ensure that the offered chat would be made within operational hours of the Contact Center and that it would have an available operator on the CC end to answer the chat request, significantly decreasing the abandonments or the non-answered chat requests.

Between the two observed timespans, new campaigns were implemented gradually, and new group pages were added to the existing *Time elapsed (2s)* chat campaign, these actions were made throughout the two months and a half asynchronously.

The second observed five-month period, from 1st of November to 31st of March 2021, had already 4 live chat campaigns undergoing, the one previously mentioned with more webpages being impacted by this campaign and three new chat campaigns that activate the chat depending on behavioral characteristics of the online visitor.

Leaving the Checkout, one of the implemented chat campaigns that use behavioral triggers, is a chat campaign that activates the chat when a visitor enters the checkout funnel and, for some reason, leaves the checkout funnel steps to continuing browsing the website or even to leave the website. In each of the cases, if the visitor, after the campaign being triggered and a chat being offered, requests a chat (by typing in a message to the operator in the chat window) the specialized agent from the CC will be already aware that this visitor entered the checkout funnel but did not complete his/her purchase. This information will allow the agent to have a more personalized approach to the visitor.

Delivery Cost, other implemented chat campaign in the second observed period, activates the chat dependent on the shipping value the client has. This value, set as equal to or higher than 59 HRK in this market (8€ rounded up), is calculated within a checkout funnel step. So logically, this campaign is activated on that specific checkout step. This campaign helps operators to be already prepared for the possible doubts, concerns the customer may have since sometimes the shipping value is high due to the far location inserted by the customer or the number of purchased items or if the visitor is aiming to acquire a fragile item or even an item with a big dimension that requires a special kind of transportation. All these mentioned scenarios require special attention by a specialized operator, something that this chat campaign already allows to have, a personalized approach.

Finally, the *Cart value* was another chat campaign implemented during this five-month period that considers behavioral features of the online visitor. This chat campaign triggers the chat when they reach a certain amount of value in their shopping cart. This market set this value as 900 HRK (120€ rounded down), which means that a customer to have an offered chat of this campaign needs to add to the shopping cart an item that costs 900 HRK or more or add multiple items that together reach 900 HRK or higher.

Nevertheless, it is important to highlight that Covid-19 pandemic influenced the test results since it was registered an increase in online traffic at the beginning of May 2020 hitting the 787k page requests on the 3rd of May (Sunday), the day in 2020 with more page requests, as presented below. In these five months it was also possible to verify that there were 67M page requests, almost 6,6M visits to the e-commerce, from those visits 1,9M were unique visitors, 1,9M bounces and the average session was 7 minutes and 22 seconds.

Figure 2: Online traffic from 13th of March 2020 until 12th of August 2020

Source: Author

It is possible to associate the increase in online traffic in the beginning of May 2020 (Figure 2) to the Covid-19 lockdown period since Croatia went into lockdown on 23rd of March till the 4th of May, two weeks after Italy, the first European country to declare a national lockdown due to the spread of coronavirus (Jason Horowitz, article in New York Times magazine, 2020). With all physical stores closed at the time, it was predictable that the online traffic would increase since people would still have the need to purchase items, even if it was not possible to do so physically.

These values would only be surpassed in the end of January 2021 achieving the maximum of 919k page requests on the 24th of January, the more page requests registered in the first half of 2021, within the second observed period, as presented below. Additionally, it is possible to see the two big drops in online traffic in December 2020, presented in more detail in the below image, these two drops were due to national festivities, 24th (Christmas's Eve) and 31st (New Year's Eve), with 184k and 191k page requests, respectively.

Figure 3: Online traffic from 1st of November 2020 until 31st of March 2021

Source: Author

In the second observed period (Figure 3), regarding total values there was also 8,9M visits, 2,6M unique visitors, 89M page requests, 2,6M bounces and the average session of the online shopper was set at 10 minutes and 11 seconds.

Comparing both periods in terms of online traffic, it is easily noticed the substantial increase in the second period, with an increase of 2,3M online visits, 639k of unique visitors, 21M of page requests, 678k bounces and the average session of the visitor increased almost 3 minutes. The retail e-commerce was able to maintain the Bounce Rate at 29% in both periods of times, increase 32% in page requests and unique visitors, and a 35% in website visits. Also, slightly increased the average sessions per visit from 3,32 to 3,38 sessions per visit and even increased the average session time, in other words, there were more online clients browsing the website, more often and spent more time in it.

Furthermore, this fluctuations in online traffic are just one of many events that influenced the results of this study. Additionally, it is important to state that with more online visitors the more critical it is to provide a good UX on the website.

Considering the total recorded values of the first period, there were 1 093 432 chats offered, in other words, there were more than 1 million occasions within Contact Center working hours where an online visitor fulfilled the chat campaign trigger while there was an operator available and consequently the chat campaign was activated. From these chats offered, some were accepted by the visitor, resulting in approximately 27k chats requests, a 2,5% of Chats requests/Offered chats, as it is possible to observe in the below figure. Then, 89% of those were answered by the CC and only 7,3% from the requested resulted in an influenced checkout, achieving 1990 influenced checkouts in these five months of 2020.

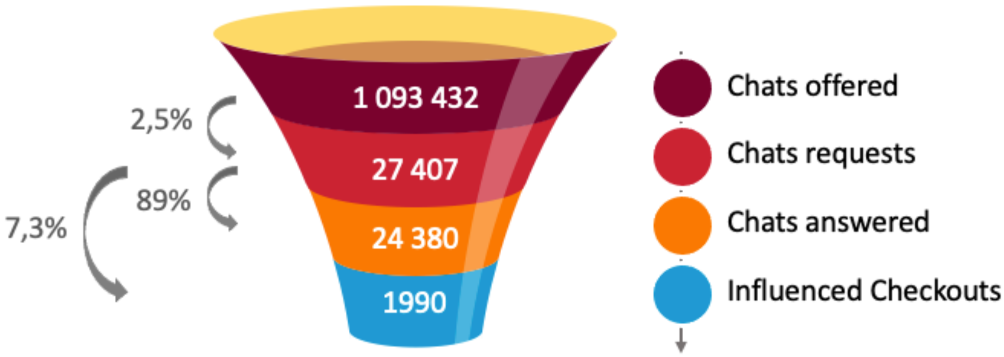


Figure 4: First period results
Source: Author

In the second period, as it is presented below, there were some differences, the Chats requests/Offered chats percentage dropped to 0,8%, resulting in less than half of the chats requests registered in the other period (27k to 11k chats requests). Other points that are important to highlight is the fact that although there were fewer chat requests the CC efficiency increased and it could answer 95,8% of the chat requests. Alongside the CC efficiency in answering chats, the newly implemented live chat campaigns had positive results, increasing the conversion rate between Requested chats and Influenced Checkouts from 7,3% to an astonishing 17,9%, more than doubling the conversion rate from the previous period.

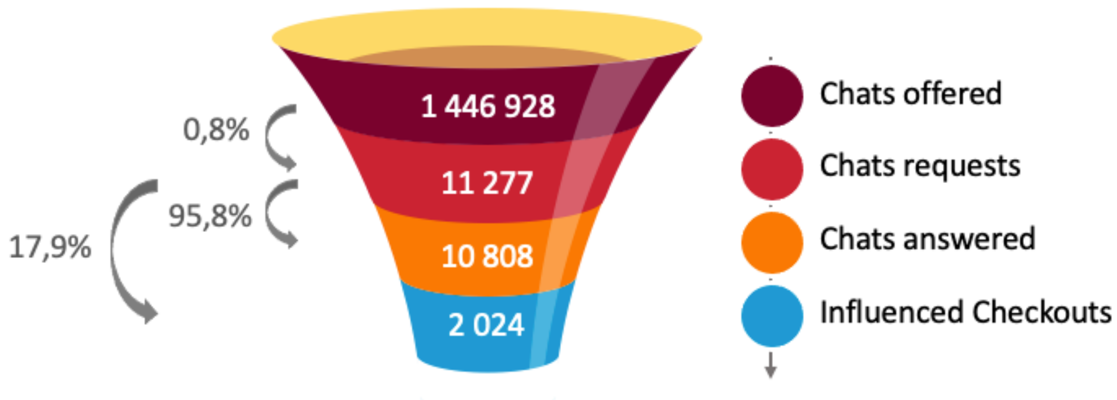


Figure 5: Second period results

Source: Author

In addition, it is imperative to highlight that this company (the multinational retail client) has an attribution model set specifically for them. Since sales are not made on the contact/chat window there is no truthful way to consider a purchase made by the assistance provided to the visitor, that is why the terminology *Influenced Checkouts* appeared. In these scenarios, an *influenced checkout* is considered a checkout that occurred till seven days after the contact request was made, so if a customer contacts the Contact Center on day 10th and makes a purchase on the 12th it will be considered an influenced checkout.

When a checkout occurs, BySide's system will match, through the visitor id and considering the past 7 days contact center data (present day and 7 previous days) the two events (contact request made to the contact center and the conversion) to present the *Influenced Checkouts*. Multiple checkouts can be attributed to one contact but only the last contact made will be linked to the registered checkout.

To conduct statistical tests with SPSS program there is a need to acquire raw data in a tabular manner and not just total values. For that, the required data was retrieved from two different databases to conduct this study. The contact center database that gathers information related to the Contact Center performance/activity contains the contact requests that are further analyzed. The other data source this study needed was the e-commerce database. This dataset has all kinds of data regarding the e-commerce registered activity, among them: number of visits, online traffic and so on. The required data in this case was the number of checkouts of each of the observed period. It should be noted that to retrieve the contact requests data the periods previously mentioned were considered. However, to gather the conversions data it had to be considered the previously mentioned periods and 7 additional days to cover and collect possible conversions that happened days afterwards. Taking that into account, the checkouts gathered for this research were the ones registered between the two following periods: from 12th of March 2020 to 19th of August of 2020 and from 1st of November 2020 to 7th of April 2021.

After retrieving raw data regarding the contacts and the conversions it was imperative to link the two datasets to obtain the same total values in a tabular manner as the settled attribution model. Hence, the function VLOOKUP was used in the program Microsoft Excel to match the conversions with the contact requests made through the visitor id. Nevertheless, it was important to filter out the matches that had more than 7 days length or less than 0 days in the cases the conversion happened previously to the contact request made. Taking that into account, the function DAYS was used to calculate the difference between the date of the conversion and the date of the contact to afterwards only consider the matches that had 0 up to 7 days of difference.

Finally, with the dataset already arranged it was time to perform the statical tests in SPSS.

4.2. Results

Results is the second subchapter of the **Findings** chapter which presents the results of the research after conducting the statical tests performed in SPSS, as previously mentioned.

Prior to presenting the conducted T-tests results, it is important to visualize the differences between the distributions between the two observed periods alongside with representing these differences visually. In that line of thought, it was concluded that to better structure this part this subchapter should be divided into two main parts: **Descriptive Statistics Analyses** and **Independent Samples T-tests**.

4.2.1. Descriptive Statistical Analysis

To start off this subchapter, Table 1 presents the Statistical table of the distribution results of the first period. After covering every aspect of it, a histogram will be presented to display in a more visual manner the distribution of the first period, considering the value of the checkouts (**Checkouts_value**) as the X axis and the Y axis as the **frequency** of purchases with those values.

Statistics

Checkouts_value

N	Valid	1990
	Missing	0
Mean		2330.7218
Std. Error of Mean		81.51764
Median		1000.4500
Mode		249.00
Std. Deviation		3636.45439
Variance		13223800.5
Skewness		4.062
Std. Error of Skewness		.055
Kurtosis		23.822
Std. Error of Kurtosis		.110
Range		40315.20
Minimum		9.90
Maximum		40325.10
Sum		4638136.44
Percentiles	25	499.0000
	50	1000.4500
	75	2498.2500

Table 1: Statistical Analysis of the first period results

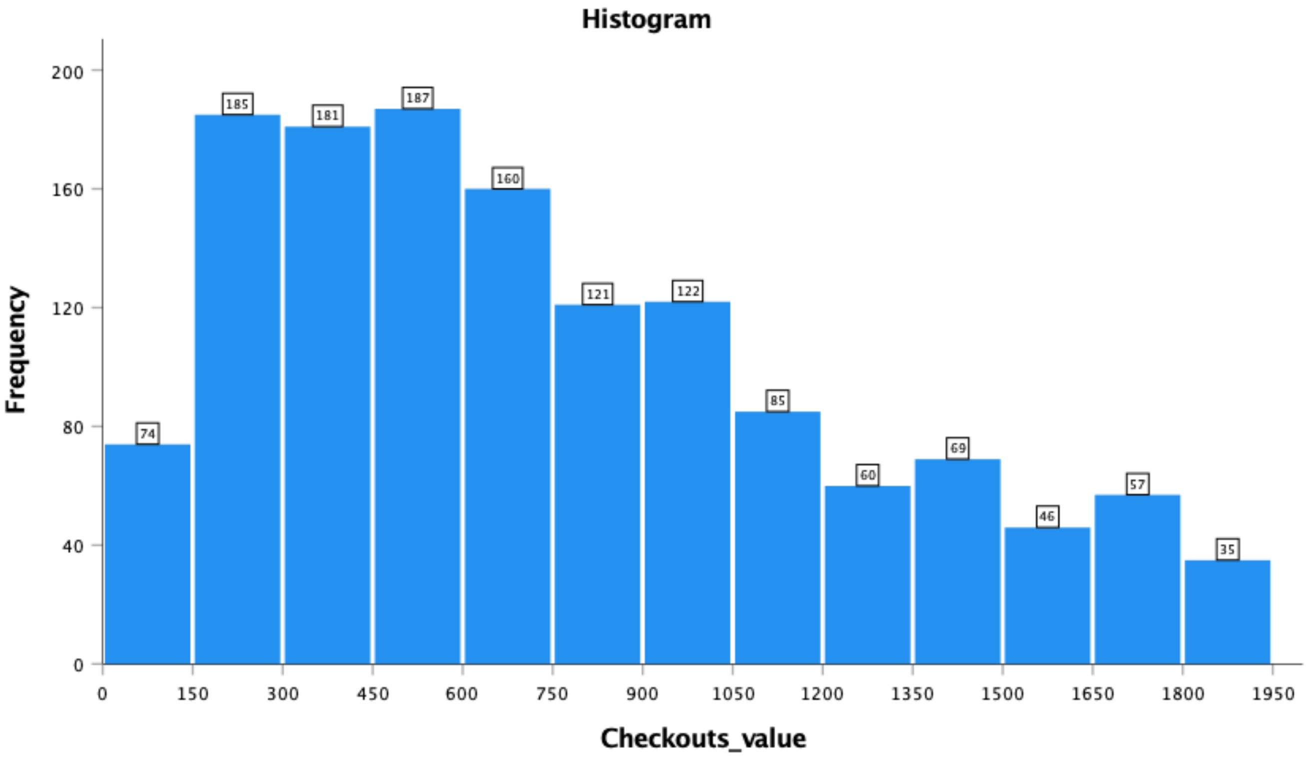
Source: Author

Firstly, from Table 1 one can verify that there were 1990 influenced checkouts as previously stated, from those it was possible to acquire the mean of this population, being that approximately 2330,7 HRK (310€). The total value of influenced checkouts in this period was 4 638 136,44 HRK (around 616 872€), the most frequent purchase by value was set at 249 HRK (33€) registering 16 checkouts with this value and the maximum value of a checkout made during this period was a 40 325,1 HRK (5 363€) purchase.

Secondly, an important point that should be noticed in this population is the fact that 50% of the registered checkouts values were situated between the percentile 25 and the percentile 75, from 499 HRK to 2 498,25 HRK or, with a 1 to 0,133 exchange rate from HRK to euros, from 66,37€ to 332,27€. Also, the standard deviation in this population is higher than the mean at approximately 3636,45 HRK (483,65€) due to presence of outliers in the dataset, mild and extreme outliers further mentioned and analyzed.

Thirdly, as for **Skewness** and **Kurtosis**, since the value of the skewness is greater than 1 (4.062) it is easily affirmed that this distribution is right skewed and since the distribution as a positive kurtosis value greater than 1 (23.822) is it possible to state that it is leptokurtik, meaning that the distribution is taller than normal.

On the next page, it is possible to consult the Histogram (Graph 1) created from the checkout values of this period (X axis) and the purchase frequency (Y axis), in other words, how many times there was a purchase within that range of price. From the three most frequent ranges with a value situated between 150 and 600 HRK (19,95€ and 79,80€) there were 553 checkouts made, approximately 28% of the total checkouts registered.



Graph 1: Histogram of the first period

Source: Author

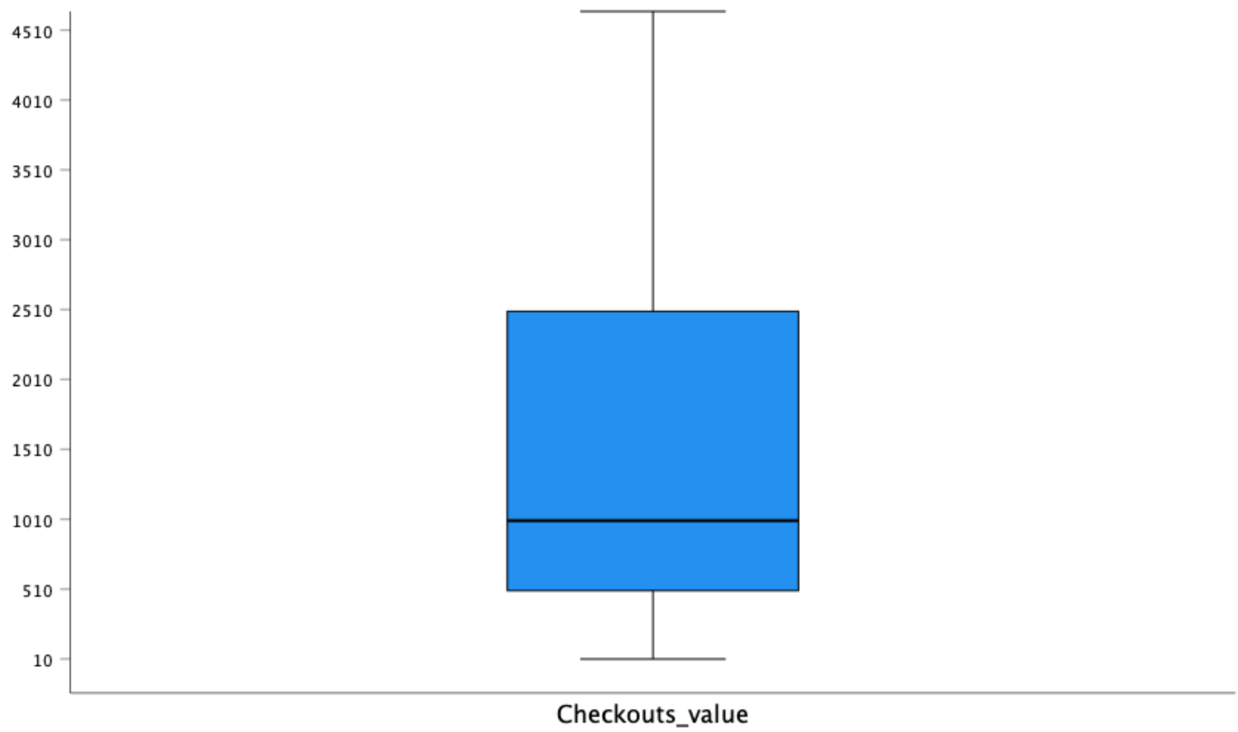
While consulting the table **Statistics** (Table 1) previously presented of this period, it is possible to calculate the interquartile range (**IQR**), presented in the below boxplot, using the formula Q3-Q1, in this specific case it would be 1999,25 HRK (2498,25–499). Then, with the IQR it is possible to reach to the value from where values above it shall be considered outliers, there are mild outliers and extreme outliers, represented by a circle and a star respectively in SPSS boxplot graphs respectively. To calculate a **mild outlier** and an **extreme outlier** over the range in the boxplot graph it should be considered the following formulas:

Mild outlier should be values considered above Q3 + 1,5 x IQR

Extreme outlier should be values considered above Q3 + 3 x IQR

Considering the referred formulas, it was established as a mild outlier checkouts with a value above 5 497,13 HRK (2498,25 + 1,5 x 1999,25) or 731,12€ and extreme outliers with a value over 8 496 HRK (2498,25 + 3 x 1999,25) or 1 129,97€.

Keeping this in mind, there are 224 outliers in this dataset, from those 110 are mild outliers and 114 are extreme outliers. It was not possible to retrieve the outliers of this dataset considering that they were registered checkouts and not simply errors so, to present the boxplot in a more visual manner it is presented below (Graph 2) one boxplot graph of the three built, presenting the percentiles and the minimum and maximum margins without the outliers. The other two boxplots built that present the boxplot containing all the outliers of the dataset and another one highlighting the outlier's margins are in Appendix IV and V.



Graph 2: Boxplot without outliers of the first period

Source: Author

Referring to Graph 2 that illustrates the boxplot, one can see the IQR, the median (1 000,45 HRK) represented by the line in the middle of the IQR and the minimum (9,9 HRK) and the maximum (5 463 HRK), represented by the lower and the higher "whisker", respectively. The minimum and the maximum stands for the minimal value and the highest value in the present dataset without considering the outliers, that in this case are only above the maximum.

In the second observed population, there were significant differences considering different strategies used to manage this proactive live chat app. In Table 2 below, one can notice every difference in a more detailed manner.

Statistics

Checkouts_value

N	Valid	2024
	Missing	0
Mean		2814.7749
Std. Error of Mean		132.46934
Median		1572.8500
Mode		1299.00[a]
Std. Deviation		5959.64834
Variance		35517408.3
Skewness		23.896
Std. Error of Skewness		.054
Kurtosis		822.853
Std. Error of Kurtosis		.109
Range		216510.00
Minimum		20.00
Maximum		216530.00
Sum		5697104.36
Percentiles	25	800.1000
	50	1572.8500
	75	3240.2750

a. Multiple modes exist. The smallest value is shown

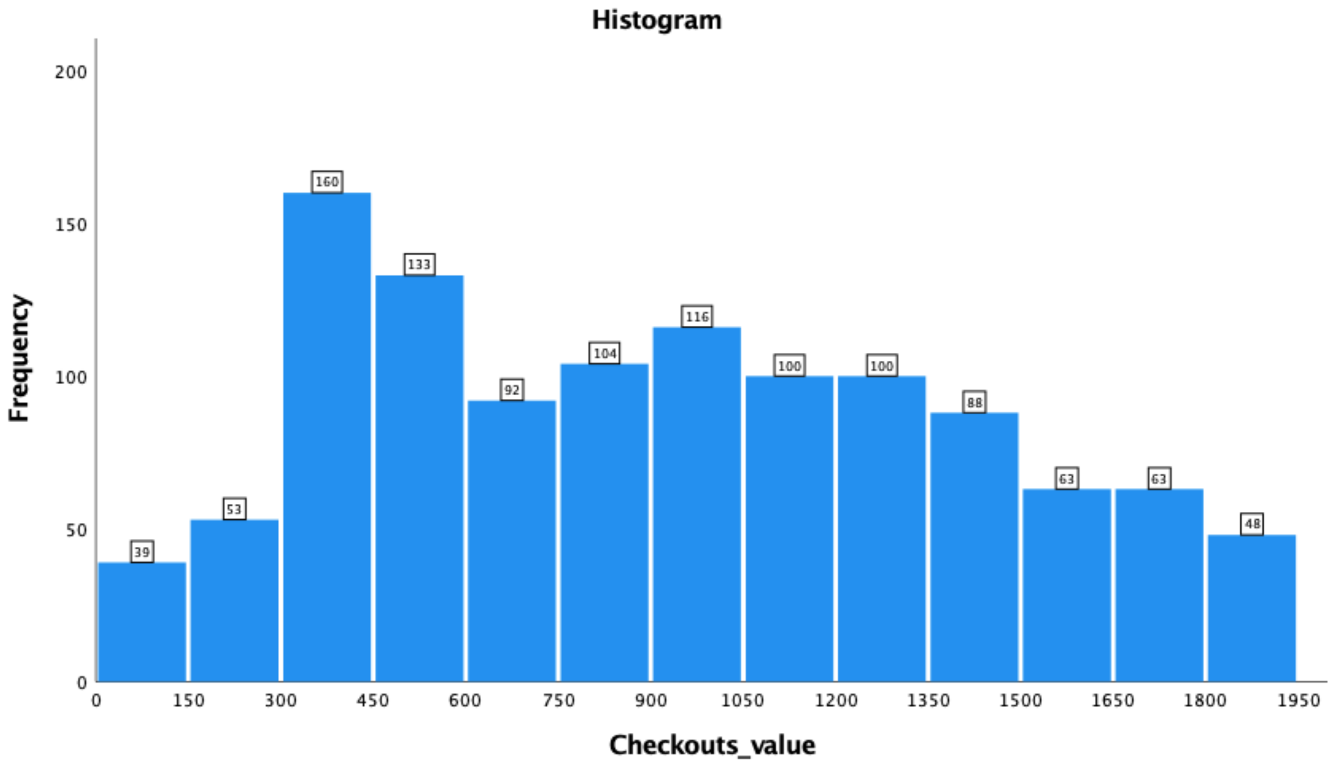
Table 2: Statistical Analysis of the second period results

Source: Author

As it can be seen in Table 2, there were more 34 influenced checkouts than the previous period, reaching 2024 influenced checkouts, the mean rose approximately 64,4€ from 2330,7 HRK (310€) to 2814,77 HRK (374,4€). The mode in this population has multiple values, in the above table (Table 2) it is referred the smallest value that is 1299 HRK (172,8€) and had a frequency of 10, meaning that 10 checkouts were registered with that specific value, however as the caption states, there were also 10 checkouts made in this period with a different value, 1499 HRK (199,4€) to be more exact. Considering the small increase in terms of number of influenced checkouts and the significant positive difference in the average order value of influenced checkouts, one can see that there was an increase in terms of total value of influenced checkouts, from 4 638 136,44 HRK to 5 697 104 HRK (616 872€ to 757 715€), an increase of around 140 843€.

Additionally, on the 20th of March 2021 was registered a checkout of an astonishing 216 530 HRK (28 798,49€), making it the maximum purchase in terms of value anyone made on this e-commerce during these 5 observed months. Regarding the percentiles, 50% (difference between percentile 25 and 75 or in other words, difference between Q1 and Q3) of the persons who converted after they were influenced by this app made a checkout with a value between 800,1 HRK and 3240,3 HRK (106,41€ and 430,96€), presenting a significant increase comparing to the other period that had these values set at from 66,37€ to 332,27€. Also, the standard deviation in this population is again high, at approximately 5959,65 HRK (792,63€), due to the presence of outliers in the dataset, as the previous observed population had.

Moreover, as in the first period, the distribution of this dataset has a positive Skewness value greater than 1 (23.896) and a Kurtosis value also greater than 1 (822.853), which means the distribution is right skewed and taller than normal. Comparing to the previous observed dataset, one can see that this one is significantly more right skewed and taller.



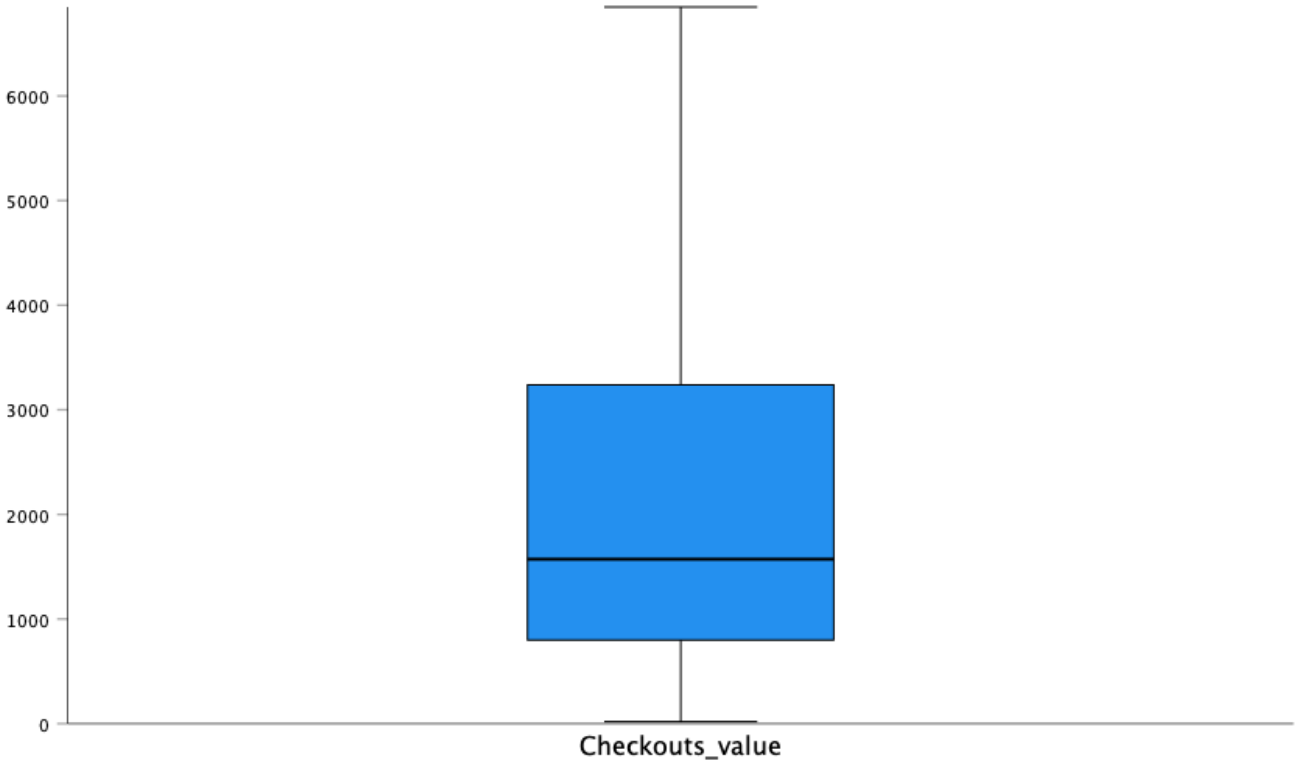
Graph 3: Histogram of the second period

Source: Author

Graph 3 presents the Histogram created from the checkout values (X axis) and the frequency of purchases with the respective values ranges (Y axis). From the two most frequent ranges with a value situated between 300 and 600 HRK (39,90€ and 79,80€) there were 293 checkouts made, roughly 14,5% of the total checkouts registered within this period.

While observing the **Statistics** table (Table 2) previously presented of this period, one can calculate the interquartile range (**IQR**), presented in the boxplot in Graph 4, resulting in 2440,17 HRK (3240,27–800,1). Afterwards, as explained in the previous observation over the other period, it should be calculated the values from where mild and the extreme outliers shall be considered, in this dataset these values are set at 6 900,53 HRK (3240,27 + 1,5 x 2440,17) or 917,77€ and 10 560,78 HRK (3240,27 + 3 x 2440,17) or 1404,58€, respectively.

There are 168 outliers in this dataset, from those 95 are mild outliers and 73 are extreme outliers. Again, it was not possible to retrieve the outliers of this dataset considering that they were registered checkouts so, to present the boxplot in a more visual manner it is presented below one boxplot graph (Graph 4) that does not contain the outliers of the dataset, attached is shared the other two boxplot graphs created, one containing the outliers (Appendix VI) and another with the outlier's margins (Appendix VII). Since the outliers are essential elements of the dataset it is important to consider them.



Graph 4: Boxplot without outliers of the second period

Source: Author

On the boxplot in Graph 4, one can notice the IQR, the median (1 572,85 HRK) represented by the line in the middle of the IQR and the minimum (20 HRK) and the maximum (6 875 HRK), represented by the lower and the higher "whisker", respectively. The minimum and the maximum stands for the minimal value and the highest value in the dataset without considering the outliers, that in this case are only above the maximum.

Comparing both periods IQR ranges, the significant increase from 1999,25 HRK to 2440,17 HRK (265,9€ to 324,54€) is noticeable, mainly due to the increase in standard deviation (*SD*) of the 2nd population observed, passing from

3 636,454 to 5 959,648, a measurement of spread, meaning that in the second period the observed dataset had values more distant from the mean.

4.2.2. Hypotheses testing

To test the three hypotheses, it was fundamental to conduct three independent samples t-tests considering the independent populations.

Therefore, on one hand, an independent samples t-test was performed to validate or refute the hypothesis that providing a more personalized approach to the visitor through the proactive live chat app will result in more online sales (H_1). This test, from now on referred as *first hypothesis t-test*, compared the mean difference of the number of influenced checkouts per month between both observed periods.

On the other hand, it was imperative to conduct a second t-test, from now on referred as *second hypothesis t-test*, which considered different monthly conversion rates that considers not only the influenced checkouts per month but also the number of chat requests per month that originated those influenced checkouts (H_2). This statistical test was conducted with the sole purpose to evaluate if there were significant differences between the mean of the conversion rates of both observed periods.

Additionally, a third t-test was performed, henceforth referred as *third hypothesis t-test*, to evaluate if there are significant statistical differences between the means of the average influenced checkout value, to conclude if having a more personalized approach to the visitor would result in an increase over the average online influenced purchase (*H_3*).

Firstly, before performing the ***first hypothesis t-test*** it was necessary to state the first hypothesis and define the null hypothesis and the alternative hypothesis, presented below**:**

⇒ **H_1**: *When an e-commerce business in the retail sector provides a more personalized approach to the visitor, through the proactive live chat application, it will result in more influenced checkouts.*

$$H_1: \mu_1 < \mu_2$$

⇒ **Null hypothesis**: the mean of one population is equal to the mean of the other population. $H_0: \mu_1 = \mu_2$

⇒ **Alternative hypothesis**: the mean of one population is not equal to the mean of the other population. $H_a: \mu_1 \neq \mu_2$

Secondly, in the results of this statistical test there are two tables: the **Group Statistics** and the **Independent Samples Test**, presented below.

Group Statistics (H_1)					
	Period	N	Mean	Std. Deviation	Std. Error Mean
Nr_influenced_checkouts	1	5	390.00	404.709	180.991
	2	5	404.80	41.572	18.591

Table 3: Group Statistics table of the first hypothesis t-test

Source: Author

Independent Samples Test (H_1)										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Nr_influenced_ checkouts	Equal variances assumed	23.459	.001	-.037	8	.971	-6.800	181.944	-426.363	412.763
	Equal variances not assumed			-.037	4.084	.972	-6.800	181.944	-507.867	494.267

Table 4: Independent Samples Test table of the first hypothesis t-test

Source: Author

From these results presented in Table 3 and Table 4, one can see that there is a difference between both periods means of 6.8 influenced checkouts. However, to see if this difference is statistically significant it is necessary to firstly interpret the results of *Levene's Test for Equality of Variances.*

From the first part of the second table, it is noticeable that the *p* (column Sig.) is .001. The decision rule in the Levene's states that if *p* is less or equal to .05 the variances are significantly different so the *bottom row* of results should be considered for *t*.

Subsequently, interpreting the results of the bottom row of the *t-test for Equality of Means*, is important to recall the decision rule for assessing if the test is significant or not (α= .05). If *p* is higher than .05, in these results it is .972, as it is possible to consult on column Sig. (2-tailed), the test is **not** significant, so the null hypothesis (H_0) should be accepted. In other words, the means of both periods of the number of influenced checkouts per month does **not** differ significantly, t (4.084) = - .04, p = .972.

Next, to analyze the significance of the second hypothesis through the ***second hypothesis t-test***, it was also outlined the null and the alternative hypotheses that should be considered in this t-test:

$\Rightarrow$ **H_2**: *With a more personalized approach to the online visitor the conversion rate of influenced checkouts over the number of contacts made will increase.* $H_2: \mu_1 < \mu_2$

$\Rightarrow$ **Null hypothesis**: the mean of one population is equal to the mean of the other population. $H_0: \mu_1 = \mu_2$

⇒ **Alternative hypothesis**: the mean of one population is not equal to the mean of the other population.

$$H_a: \mu_1 \neq \mu_2$$

The results of the conducted statistical test are presented below in Table 5 and Table 6, which are interpreted afterwards.

Group Statistics (H_2)					
	Period	N	Mean	Std. Deviation	Std. Error Mean
Influenced_checkouts_rates	1	5	734.60	277.69732	124.19002
	2	5	1838.00	381.58682	170.65081

Table 5: Group Statistics table of the second hypothesis t-test

Source: Author

Independent Samples Test (H_2)										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Influenced_ checkouts_ rates	Equal variances assumed	.855	.382	-5.228	8	.001	-1103.40	211.05653	-1590.0972	-616.70276
	Equal variances not assumed			-5.288	7.309	.001	-1103.40	211.05653	-1590.2265	-608.57353

Table 6: Independent Samples Test table of the second hypothesis t-test

Source: Author

From these results presented in Table 5 and Table 6, one can see that there is a difference between both periods means (-1103.40 or 11.03%), the influenced checkouts rate passed from 7.35% to 18.38%.

Moreover, to see if this difference is statistically significant it is necessary to start with interpreting the results of *Levene's Test for Equality of Variances*. There, it is noticeable that the *p* is .382, when the *p* is higher then .05 the variances are not significantly different so the *top row* of results should be considered for *t*.

While interpreting the results of the top row of the *t-test for Equality of Means*, it is important to highlight that in this test the *p-value* is lower than .05 (it is .001), meaning that the test **is significant**, so the null hypothesis (H_0) should be rejected, and the alternative hypothesis accepted (H_a). In other words, the means of both periods of influenced checkouts rates per month does differ significantly, t (8) = -5.23, p = .001. Additionally, since μ_1 is lower than μ_2 (7,346% < 18,38%) the H_2 should be accepted.

Finally, to analyze the significance of the third hypothesis through the ***third hypothesis t-test***, it was also stated the third hypothesis, the null and the alternative hypotheses that should be considered in this t-test and presented the respective results tables:

⇒ **H_3**: *With a more personalized approach to the online visitor the average value of the influenced checkout will increase.*

$$H_3: \mu_1 < \mu_2$$

⇒ **Null hypothesis**: the mean of one population is equal to the mean of the other population.

$$H_0: \mu_1 = \mu_2$$

⇒ **Alternative hypothesis**: the mean of one population is not equal to the mean of the other population.

$$H_a: \mu_1 \neq \mu_2$$

Group Statistics (H_3)					
	Period	N	Mean	Std. Deviation	Std. Error Mean
Checkouts_value	1	1990	2330.72	3636.454	81.518
	2	2024	2814.77	5959.648	132.469

Table 7: Group Statistics table of the third hypothesis t-test

Source: Author

Independent Samples Test (H_3)										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Checkouts_value	Equal variances assumed	.469	.493	-3.100	4012	.002	-484.053	156.146	-790.185	-177.921
	Equal variances not assumed			-3.112	3355.793	.002	-484.053	155.542	-789.019	-179.087

Table 8: Independent Samples Test table of the third hypothesis t-test

Source: Author

After conducting the test, referring to Table 7 and Table 8 there is a difference between both periods means (-484.053 HRK or -64,38€), the average value of the influenced checkouts increased from 2 330,72 HRK to 2 814,77 HRK or 309,99€ to 374,36€.

Nevertheless, it is essential to interpret the results of *Levene's Test for Equality of Variances* to verify if this difference is statistically significant. Therefore, in the mentioned section of the Table 8 the p is .493, having a p that is higher then .05 means that the variances are not significantly different so the *top row* of results should be considered for t.

Thereafter, in this t-test the *p-value* is lower than .05 (it is .002), meaning that the test **is significant**, so the null hypothesis (H_0) is rejected, and the alternative hypothesis is accepted (H_a). Thus, the means of both periods average value of influenced checkouts does differ significantly, t (4012) = -3.10, p = .002. Also, since μ_1 is lower than μ_2 (2 330,72 HRK < 2814,77 HRK) the H_3 is accepted.

5. DISCUSSION

The Discussion chapter aims to explain, evaluate, and summarize the main takeaways from the previously presented results. Then, it will link the new findings with the conducted Literature Review of this study and the establish research question.

This is the study section where the data gathered and presented will be properly transformed into usable information, from where the second subchapter Recommendations will emerge, compiling new insights from which retail companies can afterwards act upon.

5.1. KEY FINDINGS AND INTERPRETATIONS

According to the previously presented chapter **Results**, having a more personalized approach over the management of the proactive live chat application has positive outcomes.

The initial **research problem** "*The effects of providing personalized assistance to e-commerce visitors through a proactive live chat application have not been closely studied*" states in a concise manner that there is a lack of scientific research in this topic, something tackled with the present research. In the Results part, three statistical tests were performed to closely study this theme and evaluate the repercussions of increasing the personalization in the approach taken over the end user, the online visitor, through this application.

On one hand, the results indicate that increasing the personalization in the proactive live chat app does not directly translate into more influenced checkouts. On the other hand, it will improve the influenced checkouts rates, if it considers the number of influenced checkouts over the number of contacts made to the CC, and it will increase the average value of an influenced checkouts.

In other words, providing a more personalized assistance through the analyzed application to the lead will result in higher changes of her to convert, make a purchase, and will increase the final checkout value, making online visitors more likely to buy more items or items with higher values. The statistically significant difference between the influenced checkout rates also shows the path to optimize the proactive live chat campaigns and Contact Center's team performance since with the same amount of chat requests it will lead to more influenced conversions, in this case, influenced checkouts.

Considering that the first hypothesis (H_1) was rejected "When *an e-commerce business in the retail sector provides a more personalized approach to the visitor, through the proactive live chat application, it will result in more influenced checkouts."* and the second hypothesis (H_2) was accepted "*With a more personalized approach to the online visitor the conversion rate of influenced checkouts over the number of contacts made will increase."* It is also assumed that the first hypothesis failed to be accepted since it was not taken into account the number of chat requests between both periods, only the influenced checkouts. Consequently, presenting a small and non-significant increase in the total number of influenced checkouts. If the number of chat requests is taken into consideration, then there will be a considerable difference between the performance of both periods, as it is possible to confirm from the validation of the second hypothesis.

Moreover, there is a strong possibility that the difference between both periods in terms of conversion rates could have been also related to the doubts online customers had about Click & Collect or Home Delivery, during the first months of the pandemic in 2020. It can be assumed the online visitors became more and more informed over these points during the past year, by getting better in consulting this information online, since the governments communications became more frequent, or even due to the possibility of the retail client presenting this information the online visitor sought in a more user-friendly way on their website. Also, one aspect that was not approached during the research was *returning visitors*, it is assumed that in the second observed period there were probably more returning visitors who were already well aware of how everything regarding Click & Collect and Home Deliveries was functioning, reducing the number of customers interacting with this application with mere doubts over these points and consequently increasing the conversion rate of influenced checkouts over the number of contacts made.

Additionally, one aspect that is important to highlight while evaluating the findings, is that BySide always strives to deliver the best service possible, that is why tests were made within both periods, to ensure the operations were running smoothly as they were supposed to. Therefore, it is worth pointing out that among the observed periods data there were 43 chat requests on the first observed period and 11 chat requests on the second period that were tests. All this chat requests were answered, and no conversion was made so there was no influenced checkout linked to these contacts. Although it was technically possible to retrieve the 54 tests from the populations it was not feasible for this study to do so, considering the time and the load of work it would require.

5.2. Recommendations

This subchapter provides clear and straightforward recommendations for companies present in the retail sector to act upon while managing this application in their e-commerce.

The recommendations that should be considered managing this tool goes beyond the personalization aspect and should be interpreted as mere suggestions, as the title of this subchapter implies.

First of all, to optimize the Contact Center activity it is advised that the implemented proactive live chat campaigns perform the two validations previously mentioned before offering chats to online visitors, validating the CC operational hours and operator availability. These two validations will prevent the Contact Center of being overflooded with contact requests, in this case chat requests, and at the same time enable them to answer to the most chat requests possible, reducing the number of visitors that requested assistance and unfortunately were not assisted. In other words, performing these validations to present an offered chat means that if there is no operator available, or the visitor fulfilled the campaign requirements outside of CC operational working hours, chat assistance will not be offered to the lead in order to not generate chat requests that will most likely not be answered. In that sense, no queue is created, and the assistance provided by the operator will be as quick as possible.

However, in rare scenarios, a "temporary queue" will be created, this will only occur in specific situations such as when there have been four chats offered when there was only one operator available and the first lead to type something will create the chat request that will fulfil that available slot. If one of the other three leads then types a message in the chat window, it will be placed on a "temporary queue" that will assign the chat request as soon as there are any available slots, in the meantime zero chats will be offered, to prevent having more leads on this "temporary queue".

In this line of thought, to optimize the CC activity it is also possible, in BySide's Console, to have more than 1 chat request being delivered to an already occupied operator, this definition is called "simultaneous chats" and it defines the maximum of chat requests operators within a skill can receive. Meaning that an operator can be answering to two, three or even four chats simultaneously, optimizing its performance, this will also enable your team to boost the efficiency and answer even more chat requests at the end of the day.

Finally, getting back to the cornerstone of this study (personalization), it is now evident that it has positive outcomes considering the results previously exposed and interpreted in the first subchapter of the present chapter.

All things considered, companies can now focus on where/when the online visitor should be approached, to assist him through his customer journey on their e-commerce (customer support) or to perform up-sell or cross-sell to boost online sales (assist sales), rather than if it is worth to have a proactive live chat instead of a reactive live one. The implemented personalized chat campaigns can consider several aspects of the leads to be as spot on as possible and save both (client and Contact Center) time. These campaigns can consider, among other aspects: Shopping cart value; Delivery cost of the purchase; Time spent by the visitor in a webpage; Visitor's path in his customer journey; Specific to the area of the website; Specific to the geolocation the lead is in and so on. The goal, as always, in personalizing proactive live chat campaigns, is to target the right customers with the right message at the right time.

6. CONCLUSIONS

This is the final part of this thesis, it aims to summarize and reflect upon the developed research, acknowledge the study limitations, present in a concise and clear manner the new knowledge this research has contributed to the scientific world, and make recommendations for future work under this research topic.

The following four subchapters were created to better organize this chapter.

6.1. CONTRIBUTIONS

This subchapter goal is to state the contributions this study has made to the scientific world to better understand the topic under investigation.

Going back to the research gap discovered on the reviewed literature, *"How companies can optimize their proactive live chat tool to increase sales through personalization?"*, few scientific articles were addressing the personalization feature in such a developed and evolved app and not one deep dived into the topic as this research did. This study focused on that specific aspect to better assist companies in optimizing their online customer support and improve their visitor's UX.

The presented data contributes to a clearer understanding of the management of a proactive live chat application and its specificities, so retail companies can be better prepared to manage such evolved tool.

Furthermore, the results indicate that increasing personalization within the proactive live chat campaigns has positive outcomes, it will increase the influenced checkouts rate and the average influenced checkout value. These two validated hypotheses were not thoroughly studied before, and they are the main contributions of the present study. From these conclusions, researchers can now develop more studies in this field considering these new insights from the retail sector.

6.2. Limitations and Future Research

The Limitations and Future Research subchapter of the Conclusions part aims to extend the limitations already mentioned in the Introduction chapter, providing a better understanding of the boundaries of the present research and brief guidelines for future researchers to better study the topic under investigation in this thesis.

In the subchapter *Significance of the study and its limitations,* of the Introduction section, is referred the inconsistency over the online visitor's behavior and traffic considering the pandemic situation and the governmental restrictions that were changed from time to time over these two years, 2020 and 2021. Furthermore, the digitalization acceleration that the COVID-19 pandemic provoked was highlighted.

While conducting the data analysis, it was possible to notice that interesting insights would probably emerge over undertaking the present analysis with solely the chat requests that were answered (*chats answered*) and not with all the chat requests registered over the observed periods. Although technically possible, it was not feasible to conduct this analysis over chats answered due to the limited timeframe for conducting this thesis, something future researchers can study upon. Future research can analyze the outcomes of increasing personalization on a proactive live chat application considering chats answered instead of the chats requested. It would be a great breakthrough in understanding better how it is possible to generate more conversions/checkouts with this online customer support tool.

Additionally, other aspects were not thoroughly approached due to the aim of this research but can be developed in future research such as analyzing *first-time visitors* or *returning visitor*s and their differences while using such applications. This gives room for future research to deep dive into this particular topic since the results of addressing a visitor who is on the website for the first time may differ from the results of addressing returning visitors.

To better understand the presented results, future studies could also analyze on a campaign level, comparing the results between different proactive live chat campaigns, and evaluating which has the higher conversion rate. Analysis can be conducted over the area of the website that the campaigns are being activated, comparing for example different product categories, upper and lower funnel, and so on. There is still a lot of managerial actions and strategies that were not evaluated in an application that is present nowadays in most retail e-commerce websites.

6.3. Final Reflection

Finally, the last subchapter of the final chapter, this section describes how effective the undertaken methodology was in answering the research question, whether any new questions or unexpected insights arose in the process, and to conclude, presents the closing remarks of the study.

In short, as stated previously, the quantitative methodological approach was selected considering this study was developed on a secondary data basis, data already collected for other purposes (to manage this application and all marketing actions taken on this software). Bearing that in mind, this thesis was developed under this approach considering the data retrieved from the population was already presented in the form of numbers. Business assumptions were made (hypotheses) to afterward be accepted or rejected after the statistical tests being performed.

All in all, two of the three established hypotheses were accepted, providing ground-breaking insights over proactive live chat app management. Referring to the **research question** of this study "*How can a retail e-commerce increase the number of chat-influenced checkouts?*", it is now safe to say that one can boost the number of influenced checkouts by increasing the personalization of the provided online support to better assist the e-commerce visitor with the right message at the right time and subsequently augment the conversion probability.

In retrospect, there is still a substantial path to be taken to study and analyze deeply this topic, to better understand how to manage such app in e-commerce, as it was presented in the subchapter Limitations and Future Research. However, the present research can be considered as a stepping-stone for future studies.

On a final note, it is vital to bet on personalization in the online customer support provided to the visitors since digitalization is not the future but the reality. Companies need to optimize and enhance this type of application to better serve their online customers in the most efficient approach possible.

REFERENCES

Agrawal, S., Das, A., Gaikwad, A., & Dhage, S. (2018). Customer churn prediction modelling based on behavioural patterns analysis using deep learning. In *2018 International conference on smart com- puting and electronic enterprise (ICSCEE)* (pp. 1–6). IEEE.

Allen, C., Kania, D., & Yaeckel, B. (2001). *One-to-One Web Marketing*. 1–16.

Alnoukari, M., Alhawasli, H. A., Alnafea, H. A., & Zamreek, A. J. (2012). Business Intelligence and Agile Methodologies for Knowledge-Based Organizations: Cross-Disciplinary Applications. *Business Science Reference (an Imprint of IGI Global)*, *XII*(3), 56–59. http://www.cepis.org/upgrade/index.jsp?p=2100&n=2102

Ariguzo, G. C., Mallach, E. G., & White, S. D. (2006). The first decade of e-commerce. *International Journal of Business Information Systems*, *1*(3), 239–255. https://doi.org/10.1504/IJBIS.2006.008598

Bastug, E., Bennis, M., & Debbah, M. (2014). Living on the edge: The role of proactive caching in 5G wireless networks. *IEEE Communications Magazine*, *52*(8), 82–89. https://doi.org/10.1109/MCOM.2014.6871674

Bucklin, R. E., Lehmann, D. R., & Little, J. D. C. (1998). From decision support to decision automation: A 2020 vision. *Marketing Letters*, *9*(3), 235–246. https://doi.org/10.1023/A:1008047504898

Creswell, J. W. (2009). Research design: Qualitative, quantitative, and mixed methods approaches. In *Libr* (Third). SAGE Publications.

Cristescu, M. P. (2017). Business Intelligence Integrated Solutions. *Revista Academiei Fortelor Terestre*, *4*, 270–277.

Dijkstra, A. (2008). *The Psychology of Tailoring-Ingredients in Computer-Tailored Persuasion*. *2*, 765–784.

Earley, S. (2018). The Role of a Customer Data Platform. *IEEE Computer Society*, *20*(1), 69–76. https://doi.org/10.1109/MITP.2018.011301803

Elmorshidy, A. (2013). Applying The Technology Acceptance And Service Quality Models To Live Customer Support Chat For E-Commerce Websites. In *The Journal of Applied Business Research* (Vol. 29, Issue 2). http://www.cluteinstitute.com/

Esch, Jr., A. (2002) 'Internet II: the adventure continues', *Information Management Journal*, Vol. 36, No. 2, pp.31–37.

Fernandez, S., Jenkins, P., & Vieira, B. (2020). Europe's digital migration during COVID-19: Getting past the broad trends and averages. In *McKinsey & Company*.

Heimbach, I., Kostyra, D. S., & Hinz, O. (2015). Marketing Automation. *Business and Information Systems Engineering*, *57*(2), 129–133. https://doi.org/10.1007/s12599-015-0370-8

Järvinen, J., & Taiminen, H. (2015). Harnessing marketing automation for B2B content marketing. *Industrial Marketing Management*, *54*, 164–175. https://doi.org/10.1016/j.indmarman.2015.07.002

Kang, L., Wang, X., Tan, C. H., & Zhao, J. L. (2015). Understanding the Antecedents and Consequences of Live Chat Use in Electronic Markets. *Journal of Organizational Computing and Electronic Commerce*, *25*(2), 117–139. https://doi.org/10.1080/10919392.2015.1033935

Kettunen, E., Kemppainen, T., Lievonen, M., Makkonen, M., Frank, L., & Kari, T. (2018). Ideal types of online shoppers: a qualitative analysis of online shopping behavior. In MCIS 2018: 12th Mediterranean Conference on Information Systems (pp. 1-16). MCIS.

Larissa T. Moss & Atre S (2003). Business Intelligence Roadmap (Boston:Addison-Wesley)

Little JD (2001). Marketing automation on the internet. In: UC Berkeley 5th invit choice symp, Monterey

Lv, Z., Jin, Y., & Huang, J. (2018). How do sellers use live chat to influence consumer purchase decision in China? *Electronic Commerce Research and Applications*, *28*, 102–113. https://doi.org/10.1016/j.elerap.2018.01.003

Mahdavinejad, M. S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., & Sheth, A. P. (2018). Machine learning for internet of things data analysis: a survey. *Digital Communications and Networks*, *4*(3), 161–175. https://doi.org/10.1016/j.dcan.2017.10.002

Maloney, K., & Kemp, J. H. (2015). Changes in reference question complexity following the implementation of a proactive chat system: Implications for practice. *College and Research Libraries*, *76*(7), 959–974. https://doi.org/10.5860/crl.76.7.959

Mohri, M., Rostamizadeh, A.,, Talwalkar, A. (2018). *Foundations of Machine Learning*. Cambridge, MA: MIT Press. ISBN: 978-0-262-03940-6

Nygård, R., & Mezei, J. (2020). Automating Lead Scoring with Machine Learning: An Experimental Study. *Proceedings of the 53rd Hawaii International*

Conference on System Sciences, 1439–1448. https://doi.org/10.24251/hicss.2020.177

Safara, F. (2020). A Computational Model to Predict Consumer Behaviour During COVID-19 Pandemic. *Computational Economics*, *0123456789*. https://doi.org/10.1007/s10614-020-10069-3

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students* (fifth). Pitman Publishing. www.pearsoned.co.uk

Sayyida, S., Hartini, S., Gunawan, S., & Husin, S. N. (2021). The Impact of the Covid-19 Pandemic on Retail Consumer Behavior. *APTISI Transactions on Management (ATM)*, *5*(1), 79–88. https://doi.org/10.33050/atm.v5i1.1497

Świeczak, W., & Łukowski, W. (2016). Lead Generation Strategy as a Multichannel Mechanism of Growth of a Modern Enterprise. *Marketing of Scientific and Research Organisations*, *21*(3), 105–140. https://doi.org/10.14611/minib.21.09.2016.11

Tian, Y., & Stewart, C. (2008). History of E-Commerce. *Electronic Commerce*, *January 2007*. https://doi.org/10.4018/9781599049434.ch001

Todor, R. D. (2016). Marketing Automation. *Customer Relationship Management*, *9*(2), 87–94. https://doi.org/10.4324/9781315764597-12

Yim, C. K. B., Tse, D. K., & Chan, K. W. (2008). Strengthening Customer Loyalty through Intimacy and Passion: Roles of Customer–Firm Affection and Customer–Staff Relations in Services. *Journal of Marketing Research*, *45*(852), 1–756.

Zhang, J., & Mayer, N. (2014). Proactive chat reference: Getting in the users' space. *College and Research Libraries News*, *75*(4), 202–205. https://doi.org/10.5860/crln.75.4.9107

Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2009). *Business Research Methods* (Eight).

APPENDICES

Appendix I. Interview Guide of BySide's CEO & Founder, Vitor Magalhães

1. What was the need behind the rebranding of "Made to Work" to "BySide"?
2. What were the initial products/services of BySide?

2.1. There is any service or product BySide stopped to produce? If yes, why it did?

2.2. Which service was added posteriorly to BySide value preposition?

3. What are the most important milestones BySide has achieved?
4. Do you consider BySide to be a born global company?

4.1. SMEs (Small and medium-sized enterprises) were never in BySide's plans?

5. What are BySide top 5 competitors?
6. How did Covid-19 influence BySide?

6.1. Did it accelerate BySide growth or had the opposite result?

6.2. Did it provoke an increase or decrease in the number of clients?

6.3. What are the main clients' concerns during the pandemic?

7. What does the future hold for BySide?

7.1. What does the company aim to achieve in 5 years?

7.2. What are the medium-long term goals?

Appendix II. Interview Guide of BySide's Sales Director, Pedro Cruz

1. What is the unique value preposition BySide presents nowadays to clients?

1.1. Why clients go for BySide solution and not any of their competitors?

1.2. What is the conversion rate of possible clients (trial phase) to clients?

2. What are the company's segments? (<u>Presence in which sectors</u>)

2.1. How does BySide gain more clients? Does the majority present their problems to BySide or does BySide approach them first? (<u>Referral)</u>

2.2. When approaching a client, does BySide presents its services or does it look for the client's needs and sees where it can help the enterprise to prosper?

3. How did Covid-19 influence BySide?

3.1. Did it accelerate BySide growth or had the opposite result?

3.2. Did it provoke an increase or decrease in the number of clients?

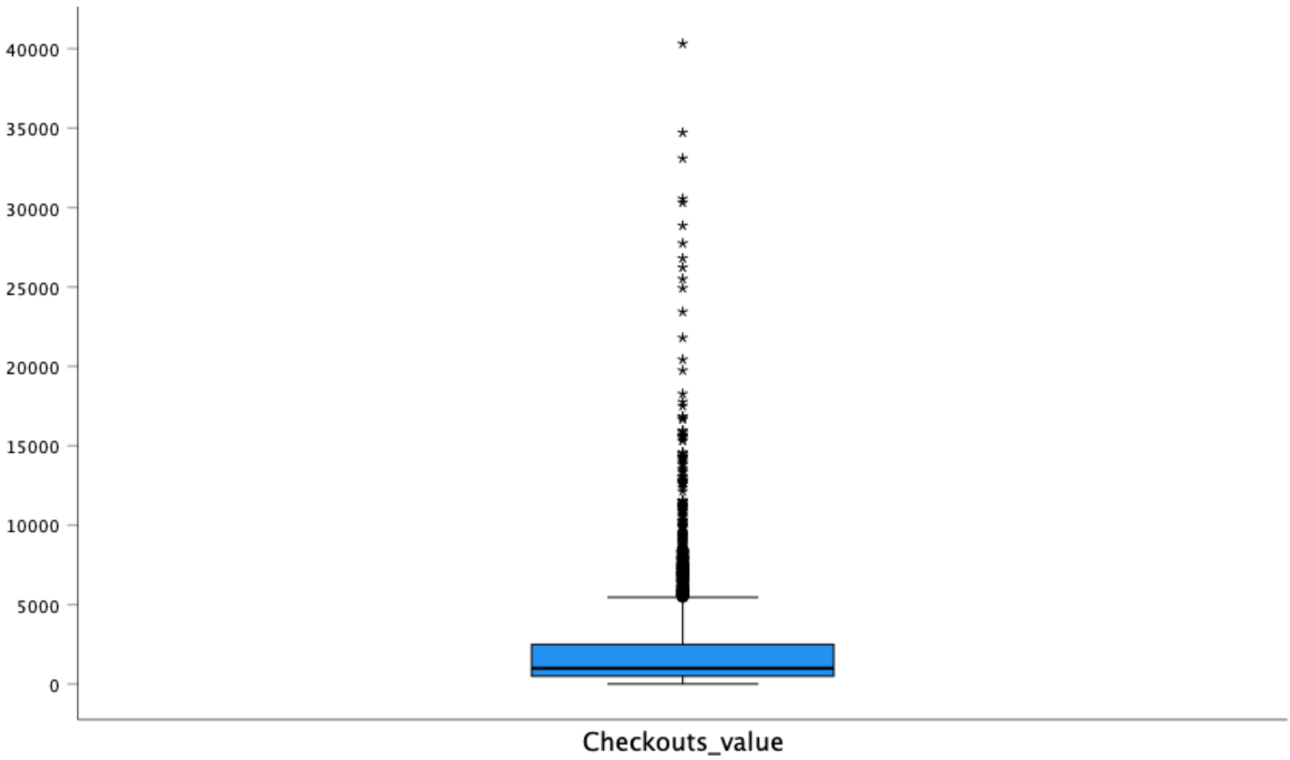
3.3. What are the main clients' concerns during the pandemic?

Appendix III. Interview Guide of BySide's client Account Manager, Rita Vaz

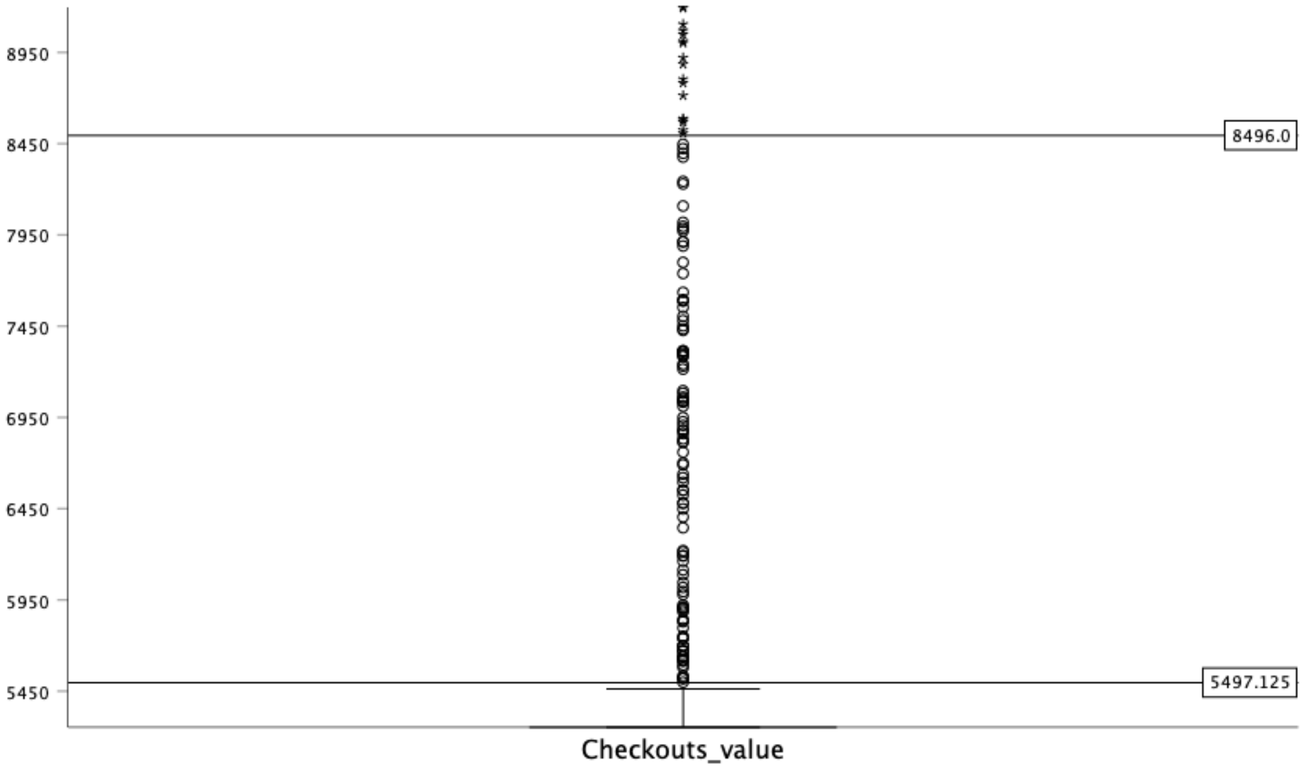
1. In what way did Covid-19 pandemic influence the client presented in this research?

1.1. How did the client adapt to the new market behaviour?

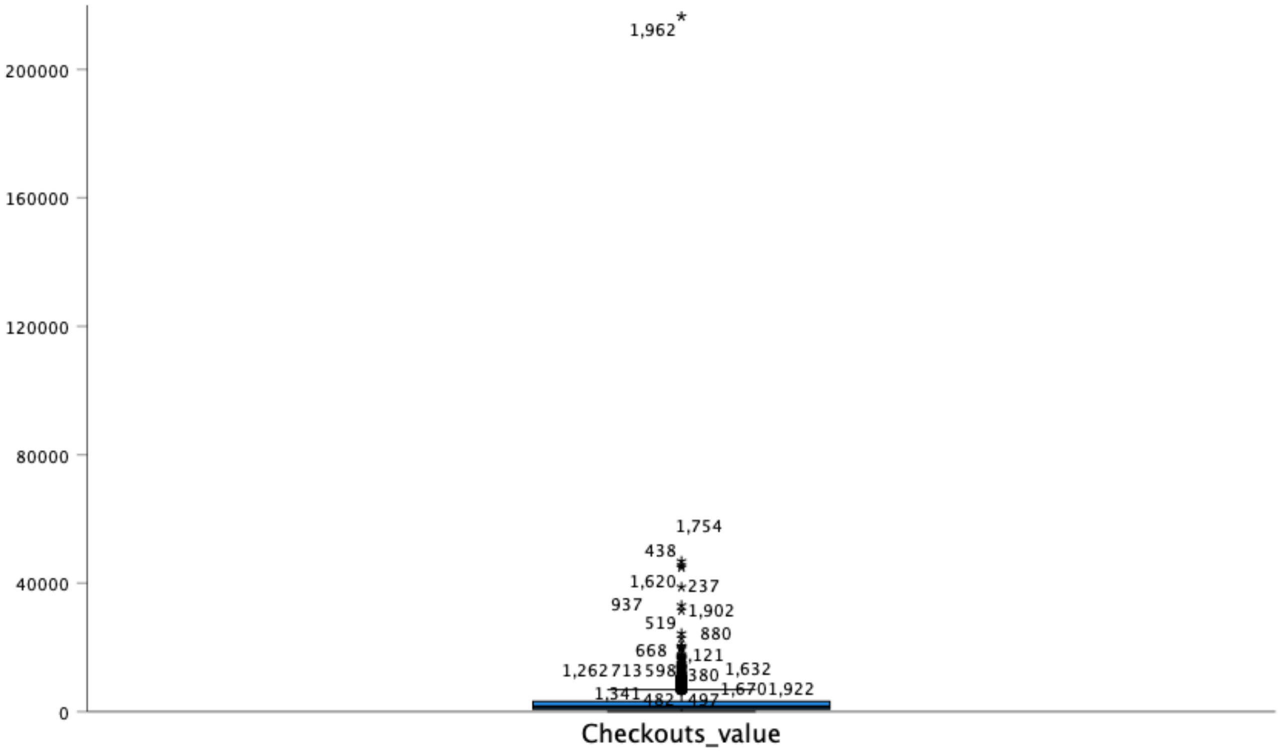
Appendix IV. Boxplot of the first observed period containing outliers



Appendix V. Boxplot of the outlier's margins of the first period



Appendix VI. Boxplot of the second observed period containing outliers



Appendix VII. Boxplot of the outlier's margins of the second period