Bridging the Realms of Machine Learning Predictions and   
Customer Retention

Keith Hammond

A Thesis Submitted in Partial Fulfilment

of the requirements for the

Degree of

Master of Science in Data Analytics

Diagram

Description automatically generated with medium confidence

September 2023

Supervisor: Marina Iantorno

**ACKNOWLEDGEMENTS**

Id like to express my deepest appreciate to my Mother, whom this year has shown unbelievable strength and resilience as she battled cancer.  
Her Constant support has aided as inspiration through my academic journey. I’d like to state my appreciation for the sacrifices she has committed to throughout my life to see me prosper.  
Id also like to express gratitude to my Father, who’s advice, unwavering support and trust in me has been a constant in being able to achieve success in my academic journey.  
Their love, sacrifices, and encouragement have been my foundation, and I dedicate this thesis to them wholeheartedly.

I’d also like to state my thanks and gratitude to my supervisor Marina, for her guidance and support throughout this study. Along with all the staff and lecturers within CCT Dublin, without their teachings and direction this would not be achievable.

Finally, I would like to dedicate this thesis to the memory of Christopher Bell,  
who was taken from us far too soon. A cherished friend whose enthusiasm and zest for life continues to influence through the light he shone upon this world.

**ABSTRACT**

Data analytics is progressively thought as a key component for companies in search of not only trying to continue within the market but prosper. An important factor in this realm is that of client retention, which is a critical factor of company achievements within this era.

This thesis delves into how machine learning has resulted in undiscovered knowledge for implementing client retention strategies.

Beginning by conducting a thorough assessment of data analytics technologies available. Examining the performance of multiple analytics technologies, highlighting their efficacy in improving customer retention through a series of results. Our findings establish the framework for identifying the most effective methods.

The next part of the research will entail digging into the customer's voice, attempting to comprehend the undiscovered territory of customer feedback and satisfaction statistics. Uncovering the hidden relationships between data-driven improvements and customer loyalty by using a strategy that includes reviews. Acquiring vital insights into client views and feelings, offering a deeper picture of their preferences and expectations.

Finally, the research investigation finishes by linking the worlds of advanced machine learning and primary research. By combining machine learning results with the rich tapestry of primary research findings gained from interviews, a route towards practical suggestions for businesses seeking to improve customer retention through data analytics is forged.

Table of Contents

[1. Introduction 1](#_Toc146205350)

[Research Objectives: 2](#_Toc146205351)

[2. Research Design 5](#_Toc146205352)

[3. Literature Review 11](#_Toc146205353)

[Machine Learning 14](#_Toc146205354)

[Predictive Analytics 15](#_Toc146205355)

[Gaps In Research: 16](#_Toc146205356)

[4. Methodology 19](#_Toc146205357)

[Machine Learning: 19](#_Toc146205358)

[Predictive Analysis: 21](#_Toc146205359)

[Logistic Regression 23](#_Toc146205360)

[Decision Tree 24](#_Toc146205361)

[Random Forest 25](#_Toc146205362)

[K Nearest Neighbour 26](#_Toc146205363)

[Support Vector Machine 27](#_Toc146205364)

[Confusion Matrix 28](#_Toc146205365)

[ROC Chart 30](#_Toc146205366)

[5. Implementation 31](#_Toc146205367)

[6. Results 41](#_Toc146205368)

[7. Discussion 47](#_Toc146205369)

[Interpretation of Results: 47](#_Toc146205370)

[Comparison with Literature: 49](#_Toc146205371)

[Inconsistencies and New Insights: 50](#_Toc146205372)

[Limitations: 51](#_Toc146205373)

[8. Conclusion 52](#_Toc146205374)

[REFERENCES 55](#_Toc146205375)

[APPENDICES- A – Workflow 69](#_Toc146205376)

[APPENDICES- B - Interview Transcripts 78](#_Toc146205377)

[APPENDICES- C - Interview Consent 102](#_Toc146205378)

1. **Introduction**The title the author has chosen for this data analytic project is “Bridging the Realms of Machine Learning and Customer Retention: Implementation Recommendations through Data Analytics”

Data Analytics has transformed the way everyday business conducts its operations, from the collection of big data to the utilization of it to align with company goals. The corporate environment has seen an unparalleled boom in data collection and consumption in the twenty-first century, leading to the growth of data analytics as a critical tool for generating insights and driving strategic choices. (Saha, 2021) Among the multiple applications that data analytics has transformed, one that stands out is client retention. Customer retention, or keeping current customers engaged and loyal, is critical to long-term business development and profitability. One of the focal points is not just being able to attract new business but from within the data at a company’s disposal being able to use the full capability of the available information. In this new data driven era, companies have vast quantities of big data for customer information, establishing a huge capability in the use of data analytics implementation in improve customer retention. (Mcafee and Brynjolfsson, 2014) This data can include purchase history, browsing trends, comments, and interactions, which may give important insights into customer preferences and behaviours when used appropriately. Delving into the branches of information that can be extracted from data to better understand company goals and the wants of the people that pay for their services, in doing so the experience on both sides can be tailored more effectively to benefit both parties to ultimate efficiency while removing the noise to produce clean results. As a result, incorporating data analytics approaches into client retention tactics has become a strategic requirement for organisations looking to gain a competitive advantage in today's volatile industry.

**Background and Context**  
The Authors interest in investigating into this area of research of how data analytics in enhancing customer retention strategies stems from the Authors own experience of working in business development and key account management roles over the last decade , seeing and working with the transition in multiple fields from being that of non-data orientated business to the ever expanding implementation of data analytics used in driving goals and results, particularly that of customer retention in the highly competitive markets we have in today’s economy.

The Author was interested in this topic as an individual that has witnessed the transition hands on in their own professional career. Comparing the experience to that of other professionals in close proximity the Author showed an interesting insight on the subject matter at 1st glance.  
  
Despite rising acknowledgement of data analytics' potential benefits in improving customer retention, there is still a gap in understanding how data analytics are implemented and how they convert into improved customer retention rates. (Damsten, 2023)  
Seeing how a company can invest in the area of data analytics in such a manner that they can improve overall business goals and in turn improve their own pipeline in gaining and retaining that of their current customers, along with that of their potential new customers and, the influence of customer relationship management forming this data analytical approach can be beneficial , be it the before and after of implementing the models has an effect on the company’s performance in this topic area and how the data is analytically processed in the method.

**Research Objectives:**  
In the ever-changing world of company operations, client retention has emerged as a critical component of organisational success. Companies are increasingly concentrating on harnessing data analytics to improve customer interaction, create loyalty, and, as a result, raise retention rates as digitalization and globalisation transform market dynamics. (Ross, 2023) While the promise of data analytics is clear, a thorough assessment of its actual impact on customer retention is required.

**Problem Hypothesis:**  
Businesses in the 21st century is more than ever introducing the use of data analytics to improve client retention. Businesses look to enhance customer retention through gathering and analysing customer data, assessing data analytics tools, and acting upon the results of the analysis.

The research question, goals, scope, and assumptions are all related to the study's problem hypothesis, which is the implementation of data analytics to improve client retention. The study topic is concerned with the function of data analytics, and the aims are intended to give a thorough knowledge of how companies may utilize data analytics to boost customer retention. The study's scope defines what will be included in the analysis, and the assumptions set the stage for the study's conclusions. The problem hypothesis statement explains the study's core point.

In this case, the statement is that businesses can use data analytics to enhance customer retention. The statement also highlights the three main objectives of the study: collecting and analysing customer.

**Research Problem:**  
The primary problem addressed by this research is the challenge of client retention. Many firms struggle to retain consumers, resulting in lower revenue and market share. Businesses now have the chance to obtain significant insights into customer behaviour and preferences thanks to the advancement of data analytics. (Marx, 2021)  
Many businesses, however, fail to employ data analytics effectively to boost client retention. Consequently, the following is the study's research problem: How might data analytics be used to boost client retention strategies. Determining the precise process, resources, and techniques that businesses employ to effectively utilise customer data is challenging. The relationship between data analytics activities and customer retention results is fully understood empirically, notwithstanding the anecdotal evidence that data analytics has a positive impact on customer retention strategies. The fundamental issue addressed in this study is deconstructing and grasping the complex interaction between data analytics and client retention tactics. This study tries to unravel the approaches, obstacles, and accomplishments of incorporating data analytics into customer retention practises. By investigating this interaction, the study aims to give a more in-depth knowledge of how data analytics is changing the customer retention environment.

**Research Question:**  
The following research questions are what the Author has highlighted in being effective to address the research problem:

1. Does using Data Analytics enhance retention of customers?
2. What data analytics tools are used in enhancing customer retention in businesses?
3. Combining machine learning results with primary research what recommendations for companies can improve customer retention through data analytics implementation?

**Objectives:**The Following objectives are what the Author has highlighted as being the primary aim of the project in investigating how data analytics used by companies improve customer retention. The are as follows:

1. Through case studies, compare the performance of various machine learning tools and determine the most successful ones for customer retention.

2. Collect and analyse customer data to uncover links between data-driven improvements and customer loyalty. Individual predictive customer churn rates to be measured through analysis of data.

3. Combine machine learning results with primary research findings (interviews) to develop practical suggestions for firms looking to improve customer retention through data analytics. Identifying distinct recommendations for focused retention strategies.

The study's goals provides a path for answering the research issue. The initial goal is to assess the efficacy of data analytics strategies in enhancing client retention .The second goal is to gather and evaluate client information. This is a critical step in comprehending how firms may utilize data analytics to improve client retention. Both will entail a comparison analysing to previous research to see which strategies have previously been successful. The final goal is to use data analytics to create actionable suggestions for organizations to maximize customer retention. This will include combining the findings of analysis with research findings and producing suggestions for businesses to employ in order to retain consumers in a data-driven world.

These strategic study objectives are in line with the research questions and give a framework for investigating the function of data analytics in improving client retention in the twenty-first century. They include data analysis, tool evaluation, and the use of machine learning findings to make practical business suggestions.

2. **Research Design**

**Data Collection Method:**A well-rounded primary data gathering technique was used to explore the influence of data analytics on customer retention. The basic data collecting process consisted of two major components: gathering reputable datasets from multiple websites and conducting in-depth interviews with industry specialists.

The first stage was to collect a dataset from a varied group of internet sources to then use data analytics to improve client retention strategy. Customer behaviour patterns including contract histories, user review engagement indicators were all included in the datasets. The data gathering procedure adheres rigorously to ethical concerns.

Simultaneously, three industry experts with extensive expertise in both data analytics and client retention methods were interviewed in-depth. The participants for these interviews were chosen via judgement sampling, which picked individuals with broad expertise and perspectives on the research issue. This approach ensured that the interviews produced useful insights and opinions

**Judgement Sampling:**  
The use of judgement sampling for participant selection was based on the assumption that the experiences and views of these experts would give a thorough knowledge of the complexities and subtleties of data analytics in the context of customer retention. The sample technique sought to include individuals with a wealth of knowledge from various industries in order to capture a varied variety of opinions and approaches.

Judgement sampling was utilized to find the right population. This is a non-probability sampling strategy that allowed the researcher to choose participants based on criteria. The participants' skill and experience in CRM operations, data analytics, and client retention tactics was the criterion in this scenario. The objective was to choose individuals who have a thorough grasp of the issue and can contribute significant insights to the research. (Fleetwood, 2023)   
The selected applicants had then been invited to engage in in-depth interviews in the second round. In-depth interviews are a valuable approach of data collection for this study since they allow for a strategic analysis of the participants' experiences and viewpoints. The interviews focused on CRM operations, data analytics methods, and customer retention strategies from the participants' experiences.

The sample size was determined by the number of participants who fulfil the population of interest's eligibility requirements. It was of satisfactory quantity to conduct 3 interviews with professionals totalling over 30 years within CRM environments, this enabled the research to produce the richness and depth of data needed to satisfy the criteria but confined enough to enable satisfactory management and data analysis. The population of interest was correctly represented through the sample.

**Data Collection Instruments:**  
This procedure entailed obtaining essential datasets from the identified targets ensuring strict ethical norms. To address privacy concerns, personally identifiable information was cleaned from the primary research and the acquired data did not contain personal information, guaranteeing compliance with data protection standards. The interviews were carried out utilising a semi-structured interview methodology, which allowed for a more balanced approach to data collecting. (Sarib and Mashuri, 2022) The interview questions were carefully developed to cover a wide range of topics, including the respondents' responsibilities in using data analytics, problems encountered, effective tactics employed, and opinions of the influence of data analytics on customer retention. The semi-structured format of the interviews allowed for open and frank replies while still ensuring that the primary study objectives were met.

**Scope and Limitations:**   
The scope of the research will look at conventional organizations in the twenty-first century and how they may utilize data analytics to improve client retention. The study will cross a range of industries and sectors, encompassing both B2B and B2C contexts, with the goal of distribution of a broad oversight giving industries a comprehensive understanding of the methods of data analytics' application in the realm of customer retention, which included the gathering and analysing of customer data, assessing machine learning methodologies used to gain insight into data analytical ability to enhance customer retention, and at the end producing strategic suggestions for organizations to use data analytics to optimize customer retention. In order to highlight the diverse machine learning methodologies and predictive analytics.  
Academic insight was secured through the path of existing literature and observed research, meaning an inherent dependency on accessible artefacts and already mapped paths, restricting the scope of beginning on an adventure of original data gathering and collecting. Second, because technological and business paradigms are unpredictable and incessantly variable, the resonance of the findings may be subjected to the steady procession of time, potentially necessitating periodic revisits and updates to the findings' expanse and scope. Finally, due to practical and logistical limits inherent in the study environment, the thesis may not thoroughly cover every conceivable industry or sector in the variety of constraints and concerns.   
This detailed investigation covers the research's fundamental components, laying the groundwork for the following: to go beyond hypothesis and anecdotal evidence to achieve a profound knowledge supported by observed analysis and intellectual investigation. The dissertation seeks to reveal the transformational influence of data analytics on the modern dynamics of customer retention tactics.

**Assumptions:**  
This conduction of this research analysis requires consumer data is precise and   
to provide results relevant to the defined queries. It also presupposes that the data analytics technologies employed are current and effective in data analysis. Furthermore, the study believes that the business environment is significant to the findings and suggestions.

1. The study assumes that typical organizations gather and use consumer data to inform client retention tactics.

2. The study implies that data analytics may be a useful technique for improving client retention in traditional firms.

3. The study expects that the suggested recommendations will be applicable to conventional companies in a variety of industries.

As has been previously stated in the above research objectives and hypothesis section within the Introduction of the Authors Research paper the exploration of Machine Learning data analytics in customer retention outlines the main objectives of gathering and the analysis of consumer data, critical evaluation of the data analytics used in pursuing greater customer retention and at the end of the article rich suggestions to then improve this goal.

To meet the research objectives and offer practical suggestions for organizations to enhance their client retention tactics, the study will collect both primary and secondary data. The study's scope includes organizations in various industries in the United States, and assumptions include access to consumer data, the usefulness of data analytics tools, and the applicability of suggestions across industries and sizes.

**Validity**   
The 2 types of validation that where applied in this research project where that of relevant and accuracy.  
The relevancy is done through referencing the data collected being directly compatible to the problem identified in the topic area hypothesis , assuring that the validation in the process being confirmed when the Author asks the Primary Research question with respect to the primary research.  
As an overview of the data collection the process and means of which the collection and analysis is conducted in a manner that is an area that companies need a conduction in and the goal of the exercise being able to report requirements enabling help for them to understand the objectives from the point of view of other professionals in serving the data analytics of business practice for growing a company’s customer retention, to fit a company’s intellectual strategy and how it could improve the business customer retention and goals in this era.

The accuracy conducted through the data analytic test and training models to assure that the catalogue of data within the process can be conducted in such a fashion to demonstrate that the data is indeed accurate to the process.  
To reduce possible biases and confounding variables, rigorous techniques were used in both the dataset collecting and the interviews. Stringent quality control processes were used for the dataset to discover and correct abnormalities, assuring the data's integrity. The semi-structured style of the interviews enabled consistent data collection while allowing for spontaneous ideas. with a balanced viewpoint, recognising the limits of a qualitative technique and the restricted number of interviews done.   
Through these forms of validation, the Author believes the will be able to give credible insight into the data analytics throughout the process and guide businesses to optimal performance in gaining and retraining their customer base.

**Ethical Considerations**  
The surge of Data Analytical manipulation also increases the potential variety of ethical considerations one must take when collecting, analysing, and using the information in the appropriate manner. In correspondence to conduct a data analysis report there is a number of ethical considerations that the author is going must be aware of to ensure the ethics in the reporting of this project are conducted in a responsible manner. The following are the ethical considerations the author has highlighted for this Research project:  
  
**Informed Consent:**  
Each participant's informed consent was carefully acquired before the interviews began. The objectives, procedures, and potential ramifications of the research were thoroughly explained to the interview subjects.  
  
**Data Privacy:**  
The first key area to be highlighted is that of data privacy for ethical consideration in the Data Analysis Report. Regardless of field all companies must understand the customers rights in their data to remain private and not to be used in any manner that is unsuitable to the consent that was given,  
There are a several means to address any ethical concerns in regards to this matter:  
  
**Limited Access:**   
Introducing and adhering to controlled access to consumer data , companies enforce that only a limited quantity of authorized individuals have access to the systems containing customer data, in turn reducing the potential mishandling of an individual’s information  
  
Along with these measures for assuring the privacy of peoples data gaining consent the individual gives permission to use their data it may not always be clear as to the extent of which a company can use that data, Leaders should inform any persons providing data, when they are collecting that data, of to all means that it can precisely how their data is going to be used, preferably in a form of explanation that would ease subjects into providing more accurate information for further analysis. This explanation of intent should in turn give the data givers an understanding of use and consent to use data given in the means of which the company wants to use shared data without any misconceptions that would arise further down the line if the person was not informed and consented to this form of use.(Lukic, 2015)

**Comply with GDPR regulations:**   
Just as the company abides by laws set upon them from the country, they are established they must also respect the operations and rules of all nations that prospective data providers are given that confidential data from.   
An example of these known regulations is the General Data Protection Regulation, (GDPR).   
GDPR came live on the 25th of May 2018, and affects all Business’ within the EU.  
GDPR gives people the right to know how their information is controlled, that their personal data is stored properly, and can request such information at any point. Personal data is that can identify a person by itself or together with information. The data subjects involved with GDPR is everyone to whom the data belongs to.  
It is of upmost that the Author ensures that these GDPR regulations are abided to as the penalties to Companies if they didn’t abide by the regulations, they face a potential fine of 4% of overall company worldwide turnover. This would be a devastating loss to a company and connects to the previous subject of transparency between provider and user as this would eliminate this possibility of damages. (EU Commission, 2023)  
  
Demonstrating to those within the report that the collection and conduction of the data analytical report that their ethics are being taken into account to put them at ease that their rights are being a heard to and their interest is in mind, implementing these it can create an established base to ensure the ethical use of data used within the report is being done so in an ethical manner to eliminate fear of privacy risks.

3. **Literature Review**Customer retention has become a major goal for firms in the contemporary business period owing to the enormous influence that it may have on the company's profitability and sustainability. Companies can do this through the use of Data Analytics to enhance customer retention by executing Customer Relationship analytical models. Assembling and Analysis on consumer information to pinpoint patterns within the data to tailor the experience to company key performance indicators.

In the Authors Literature review the importance of using Data Analytics in enhancing customer retention through the use of customer relationship management models will be investigated.

**Arguments for the Assumption that Data Analytics Enhances Customer Retention:**There are various reasons to believe that data analytics improves client retention in 21st-century company. According to one perspective, data analytics enables organizations to obtain insights into client behaviour and preferences. Businesses may use this to increase customer interactions and retention rates. (Raquib, 2023) Data analytics, for example, may be used to examine consumer input to uncover prevalent pain issues. Businesses may then utilize this data to enhance their products and services while decreasing customer turnover.

Another argument is that data analytics enables organizations to tailor targeted marketing strategies to specific groups of clients. This can boost consumer engagement and retention. Data analytics can be implemented in the evaluation of consumer data to create a clearer understanding of potential consumers whom are more perceptive to a particular marketing campaign. Once identified companies can then use these as their target market in the creation of personalized marketing to increase customer retention. (Dang, 2023)

In addition, data analytics enables firms to maximize consumer interactions across different channels such as sales, marketing, and customer support. This can boost client happiness and retention. Data analytics, for example, may be used to study client interactions with a firm and find areas for development. Businesses, for example, can utilize data analytics to analyse customer support interactions and identify the most prevalent problems that consumers encounter. Businesses may enhance customer happiness and retention by addressing these challenges.

Furthermore, data analytics helps businesses to maximize consumer interactions across different departments of a company, including sales, marketing, and customer support. This can increase customer satisfaction and retention. Data analytics used to examine customer interactions with a company to improve on areas like customer support, pinpointing problems that customers face and use it resolve the problems.

Another reason to believe that data analytics improves customer retention is that it helps firms to track consumer feedback and engagement levels. Businesses that analyse customer data can immediately discover unfavourable feedback and remedy it before it leads to client attrition. Furthermore, by tracking customer engagement levels, firms may detect consumers who are losing interest and take proactive steps to keep them.

Finally, data analytics enables companies to improve their client retention tactics by testing and iterating on various ways. Businesses may try alternative retention methods and analyse their efficacy using data analytics. This enables them to discover and optimize the most effective techniques over time, resulting in higher client retention rates.

**Arguments Against the Assumption that Data Analytics Enhances Customer Retention:**  
Even though there is a vast quantity of research that supports the concept that data analytics enhances customer retention, there is also an argument for the opposite view. One being that using data analytics can be very time intensive. Customer data analysis necessitates considerable resources, such as specialized tools, qualified staff, and infrastructure. As a result, data analytics may become unavailable to small and medium-sized firms with insufficient resources.

Another objection to the premise is that data analytics can be intrusive and may jeopardize client privacy. Customer data collection and analysis might cause privacy issues, especially if the data is sensitive or personal. This can breed mistrust and harm consumer relationships, resulting in greater customer turnover.

Overreliance on data-driven decision making can result from data analytics. As can be seen through the research that data analytics give meaningful insight on consumer data but it shouldn’t be the only driving force of the decisions being implemented. (Nolis, 2020)

Data analytics is susceptible to biases and inaccuracies. To give useful insights, data analytics relies on reliable and impartial data. Biases and inaccuracies, on the other hand, can emerge at numerous phases of the data analytics process, resulting in erroneous or misleading conclusions.  
Richard, J., Thirkell, P. and Huff, S. (2007) explored the impact of CRM for customer retention in a business-to-business (B2B) environment. The research resulted in findings that using data analytics in conjunction with CRM has a substantial influence on customer retention in B2B.

However, not all research agree that data analytics improves client retention. Hennig-Thurau, T., Langer, M.F. and Hansen, U. (2001) investigated the influence of customer education on trust and relationship quality in a field investigation. The study discovered that customer education has a long-term beneficial influence on trust and relationship quality. Instead of depending simply on data analytics, the authors suggested that firms should focus on educating their consumers in order to boost trust and relationship quality. These findings imply that data analytics may not be the only way to improve client retention.

Marwa et al. (2019) did a CRM model literature review. The research discovered a dearth of empirical data to support the premise that CRM models improve client retention.

Soltani , Z. and Navimipour, N.J. (2016) did an investigation on using data analytics to improve CRM models for customer retention. Data analytics, according to the assessment, may improve CRM by offering insights into consumer behaviour, preferences, and demands. According to the authors, data analytics is an excellent method for enhancing client retention. These findings provide credence to the notion that data analytics improves client retention.

Akter, S. and Wamba, S.F. (2016) investigated the influence of data analytics on e-commerce client retention. The study discovered that data analytics has a considerable impact on client retention in e-commerce. To enhance customer retention, it was suggested that companies should have a dedicated data analytics section to gain key findings into consumer trends.

In this literature review the following data analytics models are used in Academic studies researched in enhancing Customer Retention:

1. **Machine Learning Models:** Machine learning models are a set of algorithms and statistical models that use historical data to make predictions and identify patterns through the use of being programmed to achieve targeted results. ML is used throughout the papers in targeted marketing offers. Golbayani, P., Florescu, I. and Chatterjee, R. (2020) all used machine learning in using neural networks, SVM’s and decision trees in evaluation of customer data for customer retention.
2. **Predictive Analytics:** Predictive analytics is a technique that analyses past data and predicts future events using statistical models and machine learning algorithms. Perianez, A.P. et al. (2017) employed predictive analytics to determine the elements most likely to cause consumer turnover in mobile gaming. The authors employed machine learning algorithms to forecast the possibility of customer turnover after analysing data on consumer behaviour and use trends. This enabled the organization to provide targeted offers and services to consumers who are most likely to churn. A multitude of papers used predictive analytics to construct predictive models for consumer churning identification

For the continuity of the readers experience in the literature review, the data analytics techniques used have been segmented into their own headings for the stated techniques by the author in the order they appeared in the above list. They are as follows:

**Machine Learning**Xiahou, X. and Harada, Y. (2022) used ML models to inspect the impact customer satisfaction has on customer retention in e-commerce. Using customer happiness information, the use of neural networks, support vector machines and random forests were implemented in predicting customer retention. The results showed that it could accurately predict customer retention on satisfaction and suggested that companies use the models in the creation of customer retention campaigns.

Garg et al. (2020) examined the influence of customer involvement on customer retention in the retail business using media analytics. To quantify client involvement and loyalty, the authors collected data from social media networks and applied network analysis techniques. According to the report, customer involvement has a considerable beneficial influence on customer retention, and businesses may utilise social media analytics to identify and target highly involved consumers with personalised retention offers and incentives.

Jain and Pamula (2020) examined the influence of consumer sentiment on customer retention in the hospitality and tourism sector using ML models. Based on consumer sentiment data, the authors employed several ML algorithms, including decision trees, logistic regression, and k-nearest neighbours, to predict client retention. According to the report, ML models can reliably predict customer retention and may be used by businesses to design personalised retention tactics that target consumer pain areas. The process was also able to filter out fake reviews from the samples for more accurate data analytics in use for customer retention

Shah, S.S. (2020) examined the influence of customer lifetime value (CLV) on customer retention in the Telecoms sector using ML models. To forecast customer retention based on CLV data, the utilization of a variety of ML methods, including k-means clustering, decision tree and neural networks. According to the report, ML models can reliably predict customer retention and may be used by businesses to design personalised retention strategies that target high-value consumers.

**Predictive Analytics**Drachen, A. *et al.* (2016) investigated the impact of data analytics on customer retention in mobile gaming using predictive analytics. In the research paper workings, the creation of a predictive model using machine learning to predict the probability of customer retention decreasing. The validation in the model was in showing that customer turnover forecast in showcasing the elements that impacted customer retention. Based on the findings, it was advised that mobile gaming companies use predictive analytics to identify and target players at high risk of churn that occurs at the start of free to play games.

Wassouf, W.N. *et al.* (2020) examined the influence of customer satisfaction on customer retention in the telecoms business using predictive analytics. Using customer satisfaction levels as the focal point the creation of a predictive model was implemented in the probability of customer attrition.  
It was seen that customer happiness was a strong indicator through the model on impacting customer retention, it also through the results was able to effectively identify features that impacted on customer retention. The results for was a correlation between the classified categories and features to maintain customer retention in offering offers and services to targeted customers.

Hapsari, R., Clems, M. and Dean, D. (2016) examined the influence of service quality on customer retention in the airline industry using predictive analytics. The creation of a predictive model using service quality scores to predict customer retention. The report found that service quality had a significant effect on customer retention. The model was effective in the identification of features that influence service quality for customer retention. The results of the report suggested for airline companies to focus on increasing customers service quality using predictive analysis in being able to target customers at risk of churning with personalized offers to maintain customer retention.

WU,S. et al (2021) used logistic regression and random forest in the creation a churn prediction model. The research inspected the features that drive customer turnover, such as demographics, use habits, and service quality, using data from a telecom company. The research in this instance resulted in the discovery that using random forest would outperform logistic regression in forecasting customer attrition.

Tariq, M. *et al.* (2021) used a deep neural network to create a customer churn prediction model. The literature analysed customer behaviour and predicted customer attrition using data from e-commerce. The research revealed that deep neural networks outperformed such as logistic regression and decision trees in forecasting customer attrition.

AMUDA,K. and ADEYEMO,A.(2020) used a Multilayer Perceptron Artificial Neural Network architecture in creating a customer churn prediction model for financial institutions. The research analysed customer behaviour and predicted customer attrition using data from an online education platform. The gradient boosting decision tree resulted in being more accurate in forecasting customer turnover than logistic regression and random forest.

**Gaps In Research:**Despite the advantages highlighted throughout the literature review for the combination of Data Analytics combined with Customer Relationship Management the Author has noticed multiples gaps in the Academic research.:

1. **Scarcity of experimental studies:**  
While there are many conceptual papers on the use of data analytics and CRM to improve customer retention, actual studies that demonstrate the effectiveness of these techniques are wanting. The conduction of further research is needed to expand on the influence of Data Analytics in the enhancement of customer retention in multiple industries.

2. **Majority of research focused on Bigger Companies:**  
The majority of existing research on data analytics and CRM for customer retention is geared toward large corporations. SMEs, on the other hand, are an essential component of the economy and confront distinct issues in terms of client retention. More study is required to understand how data analytics may be effectively employed in SMEs to improve client retention.

3. **Lack of consideration for the ethical consideration when using data:**The ethical considerations that are discussed in the Authors paper have noticeably been non-existent in the research papers that author read in the compilation of this literature review and the difficulties that come with. Moving forward research papers should address these difficulties in regards to using data analytics in customer retention to give greater understanding to the field on the use of consumer data.

4. **Poor consideration of human elements in customer retention:**  
While data analytics may give useful insights into consumer behaviour and preferences, the significance of human variables such as customer emotions and attitudes is sometimes disregarded. More study is required to understand how data analytics may be used in conjunction with human insights to improve customer retention.  
  
The 1st noticeable gap in the lack of research on how data analytics has impacted on customer retention in a multitude of industries, A lot of the studies focused on very particular industries on a multitude of occasions, mainly that of banking. There is a need for data analytics and crim integration to improve customer retention in a plethora of other industries , e-commerce, retail, there is a lack of academic reports on these areas and how to implement the features to impact customer retention in these sectors.

Along with this there is also a very limited amount done of the last impact of data analytics for customer retention in emerging markets available to the researchers. A lot of untapped potential in the possibilities of DA+CRM in ever evolving but current under developed societies, there is need for a conduction of more research in these areas to see if there is a possibility to increase customer retention in these sectors moving forward.

An article that exposes some of these limitations is "Customer Relation Management, Smart Information Systems and Ethics" by Kevin, M. and Ana, F. (2019).  
This academic research paper showed that in relation to CRM there is a disregard to the ethical use of consumer data, and the concerns regarding such are ignored on a frequent basis. Al-Tit, A(2020) underlines the possibility of using data analytics for customer retention but highlights that once again that more investigation into the topic area in factoring in the variable of human involvement in effectively proceeding customer retention In SME’s, Ethical use and distribution of the data involved in the Data Analytics and the considerations that need to be addressed while using customer information for CRM. While there’s noteworthy benefits in using Data Analytics and CRM to enhance customer retention methods and results, the collection, manipulation of that information and use of the customers data does raise multiple ethical concerns- will the data stay private? Is the data safe? Will the data be used responsibly. Companies must ensure that the customers data is used in the most ethical way, as using the data provided by the customer for other means that they have not consented to can and should lead to irreparable damage to the companies’ public image. Such ethical considerations need in-depth research to build a greater universal management for ethical data management in business.

Within the literature reviewed on Data Analytics for enhancing customer retention is the lack of including the importance of consumer happiness. While customer happiness is frequently addressed in conjunction with customer retention, few studies expressly investigate the link between customer satisfaction, data analytics, and CRM. " *Impact of CRM factors on customer satisfaction and Loyalty*." Long, C.L.S. *et al.* (2020) highlights the potential of data analytics for improving customer satisfaction, but more research on how data analytics can be used in conjunction with CRM to improve customer satisfaction and retention is required.

Another gap in the literature on data analytics and CRM for client retention is the neglect of cultural issues. Cultural norms and values impact customer behaviour and preferences, which differ between areas and nations. The paper " Issues and Perspectives in Global Customer Relationship Management " by Pancras , J. *et al.* (2006) emphasizes the importance of taking cultural differences into account when designing and implementing CRM strategies, but more research is needed to determine how data analytics can be used to identify cultural differences and tailor retention strategies accordingly as through the paper the generalization of global practice rather than tailored cultures.

More study on the utilization of developing technologies in data analytics and CRM for customer retention is required. As more and more technologies continue to implement machine learning and Artificial intelligence (AI) into their applications there is an incredible opportunity to enhance customer retention. As these are currently progressing in the field the research into them is absent, the research article " INVESTIGATING THE EFFECT OF ARTIFICIAL INTELLIGENCE ON CUSTOMER RELATIONSHIP MANAGEMENT PERFORMANCE IN E-COMMERCE ENTERPRISES " by Li, L. *et al.* (2022) observes how AI is used in enhancing customer retention, yet, more research is desirable to states how the emerging technologies in unition with data analytics and CRM in enhancing customer retention.

As can be seen in the report that can through the selection of using Judgement sampling that in the selecting participants for the report, it is equally important to understand the limitations that come through the assessment. Judgement sampling can aid in the recruitment of people with relevant expertise and experience with CRM operations and data analytics. In-depth interviews, which allow for a deep analysis of the participants' experiences and opinions, may be an acceptable data gathering strategy for this study. The analysis and interpretation of the findings entail arranging and summarizing the acquired data as well as making sense of the results in order to make conclusions based on the research question and the data collected.

**4. Methodology**    
This section describes the detailed research methodology, data gathering methodologies, and data analysis techniques utilised in the study "How Data Analytics Has Improved Customer Retention in the 21st Century."  
  
**The following section includes an introduction of each approach, how they work, and their distinct responsibilities in developing accuracy for client retention via these models.**

**Machine Learning:**Machine Learning (ML) emerges as a light of revolution in the dynamic field of data analytics, were information reigns supreme. This technology enables computers to learn from data and make educated predictions or conclusions. This section delves into the core of Machine Learning, revealing how it works and its critical role in improving customer retention accuracy through the development of advanced prediction models. Machine Learning, a subtype of artificial intelligence, is a driving force in data analysis and interpretation. (Insight, 2023) It enables systems to understand patterns, detect abnormalities, and provide insights without the need for explicit programming.

**How it Works:** When exposed to fresh data, these algorithms learn from it and use the insights obtained to produce predictions or classifications.

Machine Learning works in a cyclical manner of data analysis and model refinement:

1. **Data Collection and Preparation:** The procedure starts with gathering relevant data, which will be used to train the machine learning model. This data is cleaned, processed, and organised so that it may be analysed.
2. **Feature Extraction:** From the data, relevant features or qualities are extracted. These characteristics are used as input variables by the model to produce predictions or classifications.
3. **Model Selection**: A appropriate machine learning technique is chosen based on the situation at hand. The complexity and use of algorithms range from simple regression to complicated neural networks.
4. **Model Training:** The chosen algorithm is trained using previous data to recognise patterns and correlations in the data. To minimise prediction errors, the model iteratively modifies its internal parameters.
5. **Validation and Testing:** To verify accuracy and generalisation to multiple contexts, the trained model is tested using new, previously unseen data. Testing assists in identifying any overfitting or underfitting concerns.
6. **Prediction and Optimisation:** Once confirmed, the model is used to fresh data to produce predictions or classifications. The model's performance is constantly optimised in response to comments and fresh data.

**Role in Customer Retention Accuracy:** Machine Learning is an essential component of customer retention accuracy. Machine learning models may predict future behaviour, such as customer turnover, by using previous customer data, purchase history, interactions, and demographic information. These predictive models enable firms to take proactive efforts to increase customer pleasure and loyalty, such as targeted retention programmes.

Machine Learning emerges as a powerful tool in the complex world of customer retention, enabling organisations to make proactive decisions based on data-driven insights. Machine Learning has several applications, including improving client retention accuracy through diverse strategies:

1. **Predictive Churn Analysis:** Machine Learning models analyse previous customer data to anticipate who is likely to churn in the future. These algorithms predict possible churn risks by detecting trends in consumer behaviour, such as decreased engagement or diminishing sales. (Shobana , 2023)
2. **Personalised Interventions:** With projections in hand, organisations may adapt interventions to each individual client. If a model detects a high-risk consumer, the company can re-engage them with personalised incentives, unique offers, or targeted messages.
3. **Customer Segmentation:** Machine Learning organises customers into groups based on their behaviours, preferences, and attributes. This segmentation enables organisations to develop diverse retention strategies for different client groups, each with its own set of demands.
4. **Behavioural Pattern Recognition:** Machine Learning finds churn-preventive behavioural patterns. These trends, such as customer service contact, operate as early warning signs, allowing firms to respond before customers lose interest.
5. **Product suggestions:** Machine Learning can provide personalised product suggestions by analysing historical purchase histories and client preferences. This improves the shopping experience and increases client loyalty.
6. **Optimal Communication Timing:** Machine Learning examines client interactions to discover the most effective times to communicate. Customer engagement and retention are increased through timely contacts, whether via email or targeted marketing. (Leachmann and Scheibenreif, 2023)

Machine Learning converts data into predictive capability. It enables organisations to make educated decisions that directly affect client retention by combing through enormous information and uncovering hidden trends. Its incorporation into predictive models improves the accuracy and precision of client retention efforts, propelling them beyond traditional approaches and into a domain where insights drive action.

Businesses are seeing the value of data-driven decision-making as the digital landscape advances. The significance of machine learning in customer retention is more than just prediction; it is also about proactive engagement, personalised interactions, and the development of long-term connections. Machine Learning emerges as the catalyst that transforms data into foresight, tactics into success, and encounters into lasting relationships in the process of fostering loyalty and increasing client retention.

**Predictive Analysis:**PredictiveAnalysis is a data-driven strategy that employs historical data, statistical algorithms, and machine learning approaches to forecast future events based on patterns and trends identified in the data. (IBM, 2023) By projecting potential situations, it goes beyond descriptive analysis and gives actionable insights. Predictive Analysis is a dynamic process that employs historical data and complex statistical algorithms to produce accurate predictions about future outcomes. Far beyond typical analytics, this technique uses data patterns and trends to predict future occurrences, directing proactive decision-making

**How it Works: How Predictive Analysis Operates**

The Predictive Analysis is based on the synthesis of historical data and statistical models. It is divided into numerous stages:

1. **Data Collection and Preparation:** The first stage is to collect relevant historical data, which will serve as the foundation for prediction. The data must be cleansed, organised, and translated into an analysis-ready format.
2. **Feature Selection:** Feature selection is determining the features that have the greatest influence on the desired outcome. These characteristics are used as input for the prediction model.
3. **Model Development:** The heart of Predictive Analysis is the development of predictive models. To find patterns and correlations in data, statistical methods like as regression, decision trees, and neural networks are used. These algorithms use historical data to build models that represent the underlying dynamics.
4. **Model Training and Validation:** The model is then trained and verified using historical data. The performance of the model is tested to guarantee its accuracy and dependability in forecasting outcomes.
5. **Prediction and Interpretation:** With a verified model in hand, additional data inputs are fed into the model to make future predictions. These projections' insights help decision-makers devise strategies and solutions.

**Role in Customer Retention Accuracy:**Predictive Analysis is critical in identifying consumers who are likely to churn in the future in the field of customer retention. Predictive models may identify clients at danger of churning by analysing previous customer data such as purchase behaviour, interaction frequency, and engagement patterns. This helps firms to enhance client retention accuracy by taking focused activities such as delivering personalised incentives or interventions. (Olson and Writer, 2023)  
Predictive Analysis is extremely useful in improving client retention accuracy. Predictive models can estimate which consumers are at danger of churning in the future by analysing previous customer data such as purchase history, engagement patterns, and interaction frequency. This predictive capacity is used to develop tailored retention tactics that reduce churn risks and increase client loyalty.

1. **Predictive Models for Customer turnover:** Predictive models can discover early signs of customer turnover. These algorithms anticipate whether current consumers are demonstrating similar behaviour by identifying patterns linked with customers who have already churned. This foresight helps organisations to strategically invest resources in order to retain high-risk clients.
2. **Personalised Interventions:** Using predictive model insights, firms may personalise interventions depending on client profiles. For example, if a model predicts that a certain client is likely to churn, the organisation may re-engage the consumer and avoid churn by offering personalised incentives, unique discounts, or targeted messages.
3. **Resource Allocation:** Predictive Analysis aids in resource allocation by concentrating retention efforts on customers who are most likely to churn. This avoids wasting money on customers who are less likely to turnover, maximising the value of retention measures.
4. **Improved Customer Experience:** Predictive models can uncover churn-causing issues such as bad customer service or a lack of interest. Businesses may then address these issues, improving overall customer experience and loyalty.
5. **Continuous Learning:** As fresh data becomes available, predictive models develop over time. This repeated learning process improves forecast accuracy, allowing organisations to change their tactics in response to changing client behaviours.

Predictive Analysis serves as a lighthouse in the world of data-driven decision-making. Because of its capacity to predict future outcomes based on existing data, it enables firms to take proactive efforts to improve client retention accuracy. Predictive Analysis takes client retention tactics to new heights by detecting possible churn risks, personalising interventions, optimising resource allocation, and enhancing customer experiences. (Qualtrics, 2022) s the digital world evolves, this practise becomes increasingly important in the attempt to develop long-term customer connections and support corporate success.

In summary, each algorithm through Machine Learning and Predictive Analysis, plays a unique yet interrelated role in improving customer retention accuracy: Predictive Analysis anticipates future behaviour for targeted interventions.

When these tactics are carefully combined, they enable organisations to use the power of data to optimise their customer retention efforts, resulting in increased accuracy and long-term client connections.

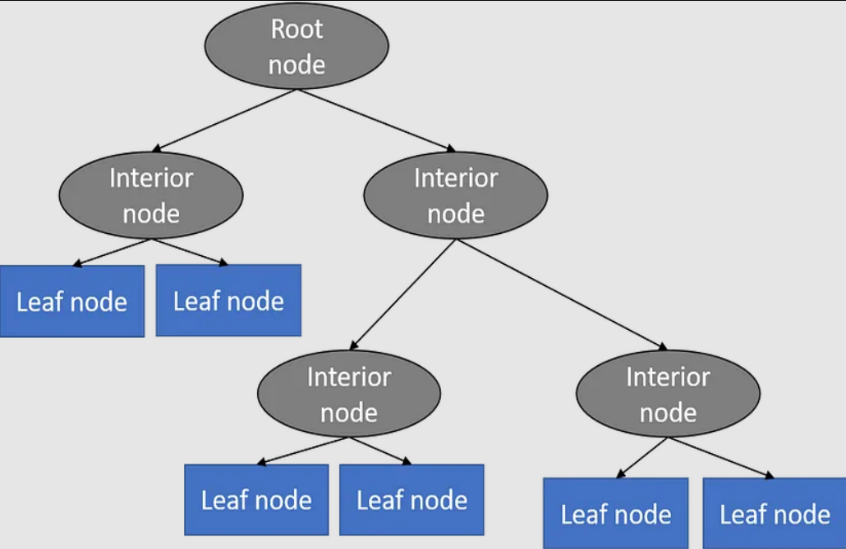
**The following section includes an overview of each algorithm used in the study:**

**Logistic Regression**  
Logistic Regression, is a methodology for measuring and interpreting the connection between two binary variables. Logistic regression contrasting to that of linear regression, predicts a result categorizing into 1 of 2 groups. The operational focus of logistic regression is to forecast the likelihood of a twofold event, such as will a customer will purchase (1) or not (0), based on a variety of independent variables. It’s conducted by using the sigmoid function, to represent the relationship between a set of input characteristics and a two-fold output.  
It assists in estimating the possibility of an event occurring, providing insights into decision-making, risk assessment, and categorization tasks involving two alternative outcomes. This probability value may then be utilised to construct binary predictions by selecting a decision threshold.

  
Fig 1 – Logistic Regression Formula (Tibco, 2023)

* P: This represents the probability that the dependent variable equals 1 (in binary logistic regression, where there are two possible outcomes, typically coded as 0 and 1).
* e: This is the base of the natural logarithm.
* B0​,B1​, B2…: These are the coefficients associated with each independent variable
* X1​,X2​,…,​. The predictor variables or traits are those that are utilised to create predictions. These variables might be numerical or category. (Tibco, 2023)

**Decision Tree**  
A decision tree is a common machine learning approach for regression and classification applications. It is a decision-making system that uses a tree-like structure to create judgements based on the characteristics of the incoming data. Because of its tree-like decision mechanism, decision trees are relatively easy to visualise. Furthermore, while optimising the decision tree model, it will acquire greater accuracy in forecasting possible clients who will quit.

  
Fig 2- Decision Tree Diagram (K, G. 2020)

It is composed of nodes. This method works on the root, internal, and leaf points. The root node symbolises the whole population or dataset, and the internal nodes represent a data aspect or attribute. The decision rules or circumstances that are utilised to create decisions are represented by the edges, and the outcomes or decisions are represented by the leaf nodes. The leaf node of a decision tree provides a class label (decision), and each inner node indicates a choice. A "test," such as whether a coin will land heads or tails, is a quality. The routes that connect the leaves and roots show categorization rules. (Vaidya, 2023)

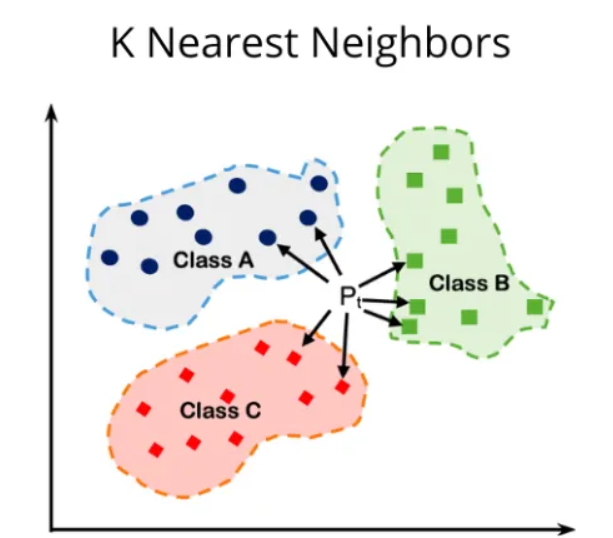
**Random Forest**  
It is a well-known ensemble learning approach that incorporates both regression and classification. It is a decision tree extension that improves accuracy while avoiding overfitting. A random forest is an ordered collection of decision trees, each trained on a distinct sample of data and characteristics. The random forest output is compiled from the average or a substantial majority of the individual tree outputs.



Fig 3- Random Forest diagram (IBM, 2023)

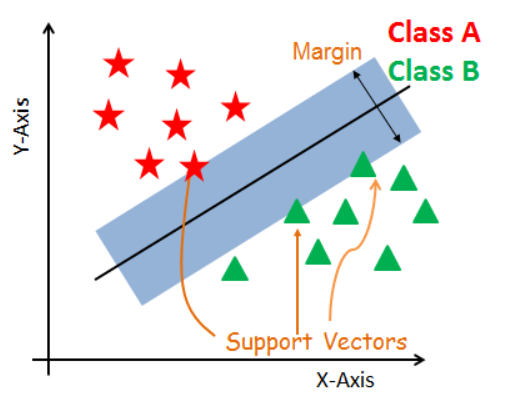
Over decision trees, random forests have various benefits. They are less prone to overfit because they decrease model variation by averaging goods and services from numerous trees. In addition, these sensors can handle with high-dimensional data and nonlinear correlations between characteristics and the target variable. Random forest is a supervised learning method that may be used for classification and regression. Similarly, the random forest algorithm generates decision trees from data samples, then receives predictions from each of them, and eventually votes to choose the best option. (Vivekanandan, 2023)

**K Nearest Neighbour**  
KNN falls within the area of instance-based learning or memory-based learning, in which new instances are labelled based on previously stored examples in memory. In this group of approaches, KNN is the most extensively employed. (Hachcham, 2023) KNN is also non-parametric, which means it does not make data assumptions and hence is more suited to real-world issues. It is also known as a lazy algorithm since all of the data points are used during the testing phase. (Joby, 2023)

  
Fig 4-K Nearest Neighbour diagram (Siddique, 2023)

KNN classifies records by calculating the distances between data points. Distance is calculated in feature space using multidimensional vectors. The length of a straight line between two locations (insert ref) is commonly used for measurement in KNN. Other distance metrics, such as Manhattan, Murkowski, and hamming distances, are also employed. When categorising objects, the feature vector is compared to the training data, and the class that is closest to it is chosen. The "K" represents the number of training examples closest to the new position. (Logunova, 2022)

**Support Vector Machine**  
Support vector machine (SVM) is a supervised learning model that uses association algorithms to recognize patterns in the data and can be used for classification and regression. The RBF kernel is often used, which maximizes the margin around the hyperplane to get the best classification.

  
Fig 5- SVM diagram (Navlani, 2019)

SVM works by generating an N-dimensional hyperplane that divides data points into two categories to represent observations in a high-dimensional space. The aim is to identify a hyperplane that splits the data points ideally, so that one category is on one side of the hyperplane and the other is on the other. (Brownlee, 2020) A kernel function is used to map the border between classes on each data instance, which is subsequently mapped into higher dimensional feature space. A kernel is basically a method for computing the dot product of two vectors, x and y. Because the kernel has a significant influence on SVM generalisation performance

The SVM formula is as follows:

H(x) = sign(w·x + b)

Where:

* H(x) is the decision function that forecasts the class label given an input vector x.
* The 'sign' function gives back +1 if the argument is positive, -1 if negative, and 0 if neutral.
* w is the weight vector that specifies the hyperplane's orientation.
* · is the dot product of the weight vector w and the input vector x.
* The bias component, b, specifies the hyperplane's offset from the origin.Top of Form

The decision boundary (the hyperplane) in this equation is defined by the values of the weight vector w and the bias term b. The SVM algorithm's purpose is to identify w and b values that maximise the margin between the two classes while minimising classification error. (Cornell, 2023)

**Confusion Matrix**

The accuracy, precision, recall, and F1 of the model are quantified and visualised in the confusion matrix. A confusion matrix is a way for summarising a classification algorithm's performance.  
The 3 parameters of performance are defined as followed:

Accuracy is defined as the proportion of accurately predicted examples to total examples.

Precision is the positive predictive value, which is defined as the ratio of true positive predictions to total anticipated positives.

The F1 Score is a function of accuracy and recall. F1 Score is required while attempting to strike a balance between Precision and Recall.



Fig 6- Confusion Matrix Labels (Raschka, 2023)

In relation to the confusion matrix the following explain what the labels of the matrix mean:

P (for Positive): The observation is favourable.  
N (for Negative): The observation is negative.  
  
True Negative (TN): A client did not churn (false) and will not churn (false).   
False Positive (FP): A client did not churn (false), yet turnover is projected (true).   
False Negative (FN): A customer churned (true), although no churn is projected (false).   
TP (TP): A client churned (true) and is anticipated to churn (true).  
  
The following are the most often used descriptive metrics for model evaluation using the confusion matrix:

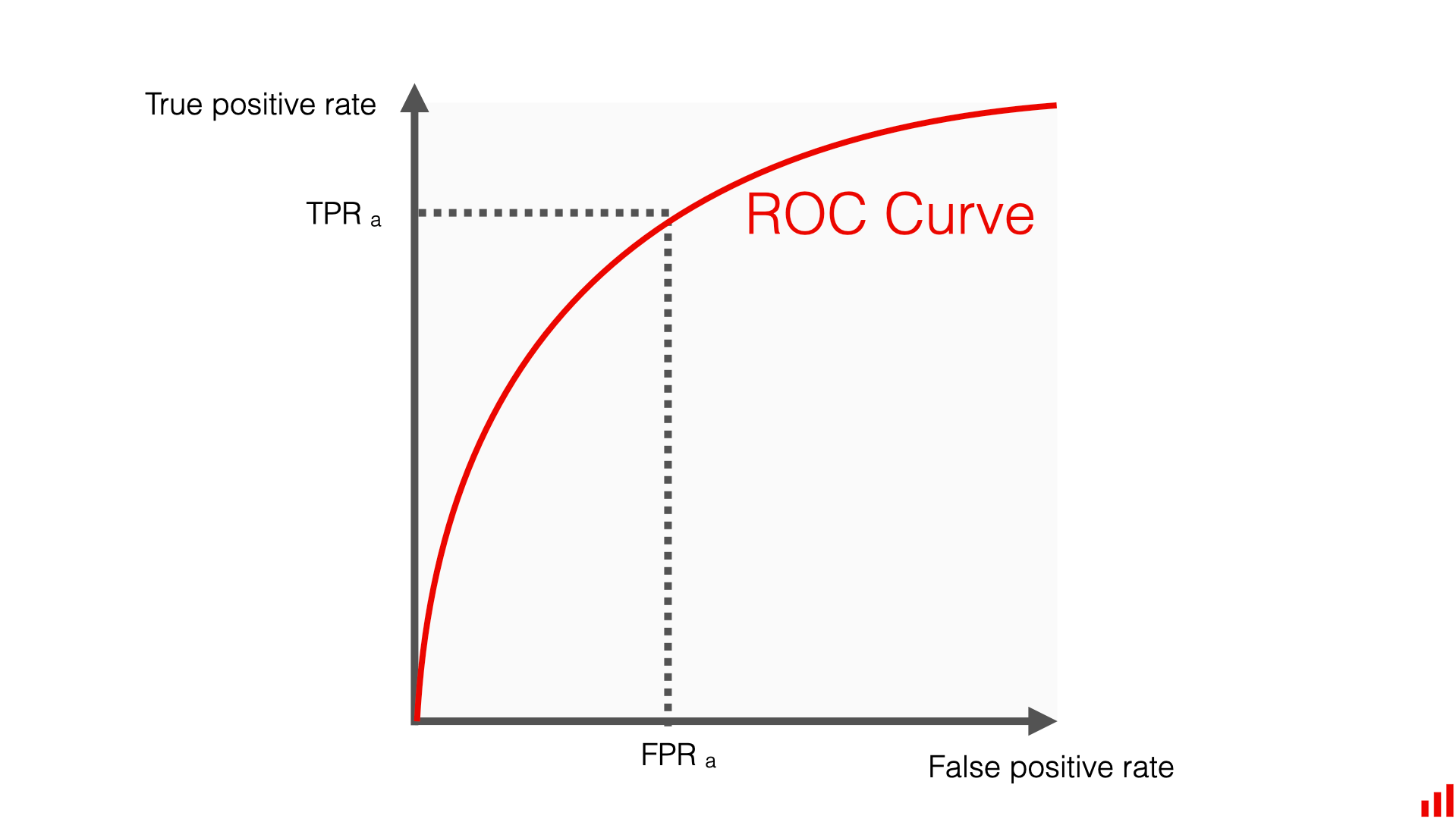
• Precision = (TP+TN) / Total   
How many of the classes (both positive and negative) have been accurately predicted

• Precision = (TP+FP) / TP   
How many of the classes we expected to be positive are really positive

• Recall = TP / (TP+FN)   
How many positive classes were accurately predicted.

• F1-Score = (Recall+ Precision)/(Recall+ Precision)  
 F1 Score is useful for measuring both recall and precision at the same time.  
(Narkhede, 2021)

**ROC Chart**  
A Receiver Operating on a ROC Chart The signature (ROC) chart depicts the model's binary classification efficiency at various classification thresholds. On various threshold values, it divides the True Positive Rate (TPR) by the number of false positives (FPR). The ROC chart, like the gain and lift charts, may be used to compare different classification methods. The ROC curve may be constructed at various classification limitations by plotting the real positive rate (responsiveness or review) vs the misleading positive rate (1 - explicitness). The classification performance of a model is often seen to be better when its ROC curve is closer to the plot's top left corner than when it is closer to the diagonal (random guessing) line. The ROC curve is a widely used device in AI and data mining applications for comparing and contrasting the outcomes of various order calculations. (Glen,2019)

  
Fig 7- ROC Curve Diagram (Evidently,2023)

5. **Implementation**

**I**n this section the Author goes into the practical components of the research, explaining the complete measures taken to explore the influence of data analytics on client retention. Research Framework, preparation, exploratory data analysis, feature engineering, model selection and development, validation, and analytics solution deployment were all part of the implementation phase. The Crisp DM Research Model was utilized om this instance.

Create a predictive model in which will cover all parts of the data science lifecycle such as data cleaning, exploratory data analysis, and the development of models such as k nearest neighbours, support vector machines, logistic regression decision trees, random forests, and so on. Evaluate a model shortlist, a model that provides the best accuracy, and finally, the likelihood of customer turnover. Providing a visual representation of wherever it is needed.

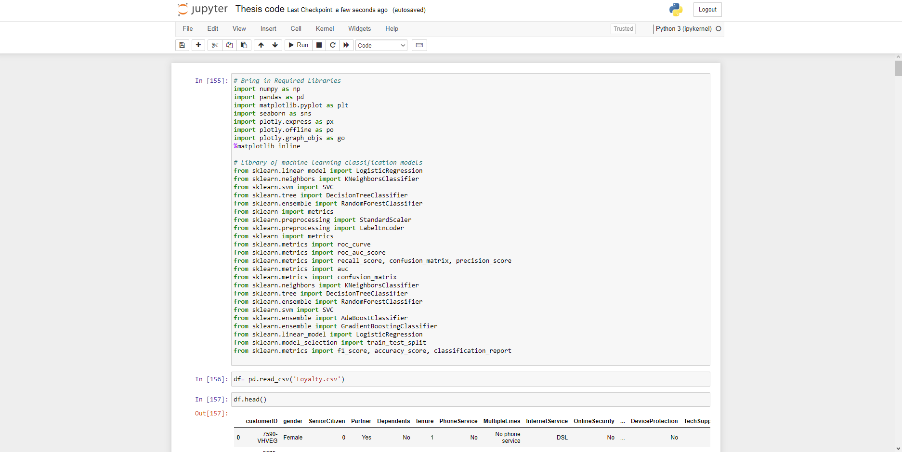
**What is Crisp Dm**?  
Before we can grasp the components of our research framework, we must first define Crisp-DM. Crisp-DM is an acronym that stands for Cross Industry Standard Process for Data Mining.

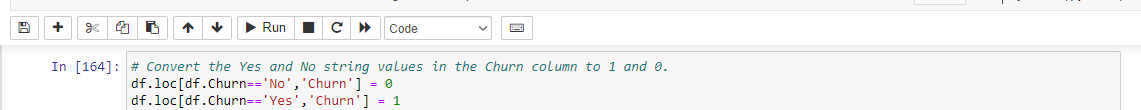
Many Analytic studies entail gathering data, preparing it for our model, creating and processing the data through the models, and then analysing the outcomes. Crisp-DM is a popular analytical model of progress for a business-oriented approach. It is a defined technique for the data mining process and its expected consequences. It is beneficial in an analytical setting to foresee or solve business challenges by categorising them into six sections, which are as follows:  
A- Business Understanding, B- Data Understanding, C- Data Preparation, D- Modelling, E- Evaluation, F- Deployment  
  
Advantages of utilising CRISP-DM as the research framework:

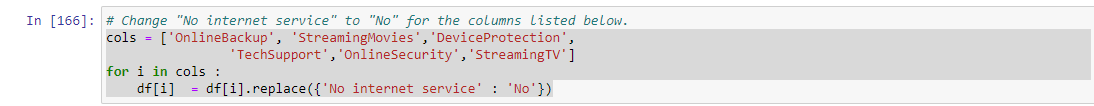
1. While it was designed with data mining in mind, one of the developers claims that from a commercial standpoint, implementing the Crisp-DM algorithm has advanced direction.
2. A project management model aids in the implementation of our process while keeping it schooled in its approach to answering the topic at hand.
3. Its model description may be utilised across many industries without causing confusion or ambiguity.
4. Knowing the company venture's emphasis from the start of the data query helps focus the workings inside the data to the precise goal that's seeking to be attained., highlighting customers at risk of churning
5. Once the deployment is released, the workings may be used to re-distribute the information and variables in order to build a more expansive understanding of opportunity in its knowledge.

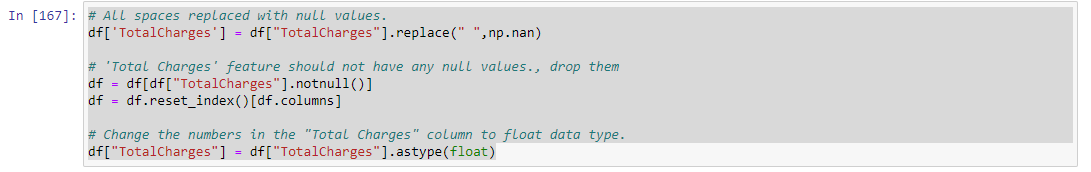
As stated in the previous point, the flexibility to reap a plethora of benefits from the algorithm's processes and workings and gain more knowledge on the relationships between the data and the question, recycling through the process reveals more possibilities than just the one being asked at the start of its cycle.

**Business Understanding**In this study, business understanding is clearly identifying consumers at danger of departing, which is universal from industry to industry, the sphere of assessing customers, and approaches to produce more retention decisions that can benefit a company. The first step in generating value from the CRISP-DM model is to understand the business task. The goal of this Data Analytical study is to gain insight into which model best targets customers for the business's goals. In the process, we will discover some influential outcome factors in the models and correctly model our data to highlight the risk of each customer leaving.  
The following report is created using the Crisp MD methodology to help generate insights as to accurately target customers at risk of leaving.

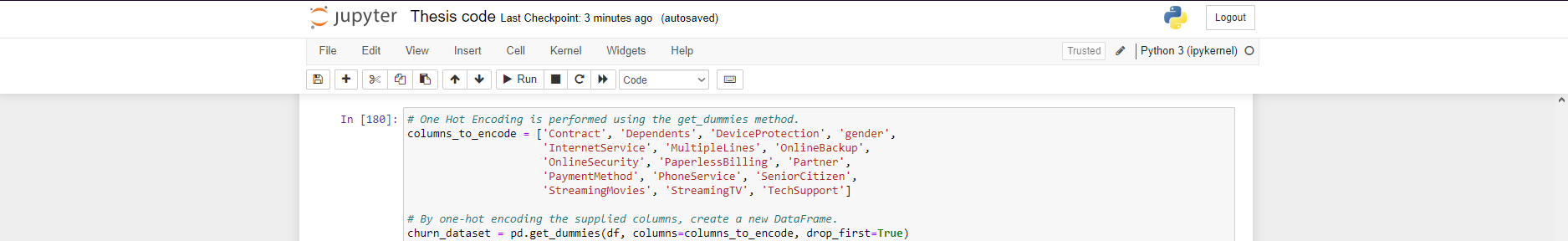
**Data Understanding**The dataset presented comes from IBM collection on Gihub, and it confirms the quality of the dataset in the author's collection. Given that the Author did not conduct primary research for the purpose of the project, it should be mentioned that the sources of the data, methods of gathering variables, and any probable scenario of encountered difficulties are not addressed in the report. This would be useful knowledge for prospects who utilise the process deployment and result readings to replicate the stages of data collecting. This data collection has 7043 records. It also has 21 columns or features. As you can see, there are elements such as customer ID, gender, senior citizen partner dependents, and so on. So there are a total of 21 such features on which we will base our prediction model. The dataset is presented in tabular format and contains both numerical and category data.  
The author will develop, model, and assess the consumers who best correlate to departing based on the examination of these acquired data frames. While performing EDA (Exploratory Data Analysis) on the data for pre-cleansing, it is clear that there are no missing variables and no duplications.  
   
 Fig 8 - Dependencies

**Data Preparation**  
Importing required libraries such as NumPy, pandas, matplotlib, and Seaborn   
This data collection has a total of 21 variables, a total of 7043 observations, and no missing values. There are also two numerical variables, as shown on the right side, therefore there are two numerical variables. 12 categorical variables and 6 Boolean variables.  
  
Performing data manipulation on the goal variable, churn. If you observe, it contains two string values: yes and no. Yes indicates that the client will leave, while no indicates that they will remain. We need to transform these string values to numeric values since our machine learning model only accepts input in numeric form, therefore in this column, I'm converting value no to 0 and value yes to 1.  
  
 Fig 9- convert string values 1 and 0 churn  
There are six columns, namely- online backup, streaming movies, device protection, tech assistance, online security, and streaming TV, with values ranging from no and yes to no internet service. Because value no and no internet service are the same thing, in order to be consistent, the author turned the value no internet service to the value ‘no’.  
  
So in the next cell, first create a list of columns that have these values, which we call Cols, and then we use a floor loop to pick each column one by one from the list called cols, and then we use a method called replace to replace no internet service with no.

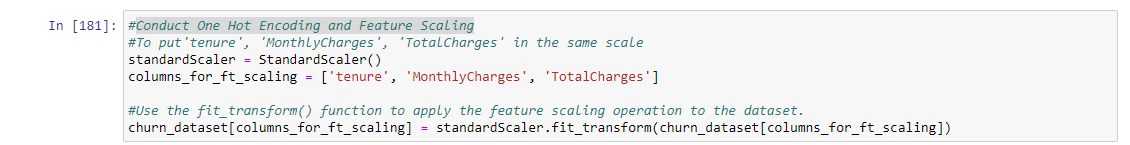
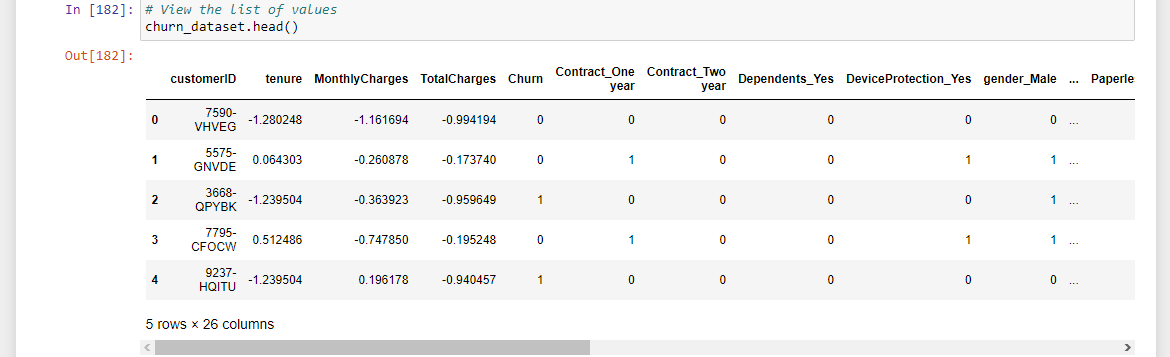
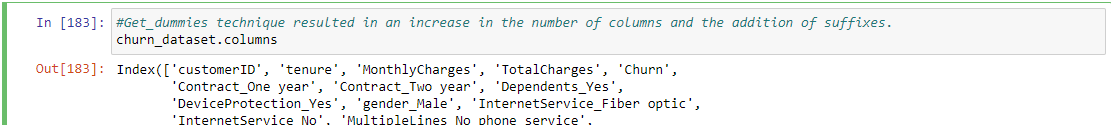
  
 Fig 10- convert no internet service to NO

The replace method, which has two parameters, is then used to replace all of the spaces with null values. The first argument is a space in inverted commas, while the second is nan or not a number denoting a null value, which is dropped in the next line. Then, in the next line, convert these categorical values to floats by using the as type method and supplying float as an argument within this function.  
 Fig 11- replace all spaces with null values   
The "Total Charges" column is presented as categorical, yet it contains some float values. This column's data type required to be changed to float. This is required to have data in the correct format with consistent data type.

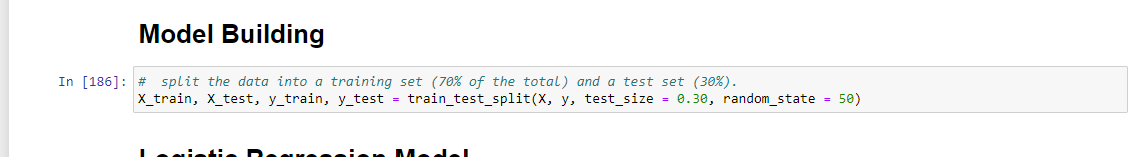
After completing the necessary data modification, the Author undertook some visualisation to investigate the data. So, in the following cell, have overall customer churn pie chart, we can see that 26.6% of consumers were churned away, while the remaining 73.4% remained loyal to the organisation.  
  
There are certain category and Boolean variables in the data set that require one hot encoding so that our machine learning model can interpret these values optimally. As previously stated, machine learning models will be unable to infer these category characteristics as feasible inputs for model training until they are converted into dummy variables.

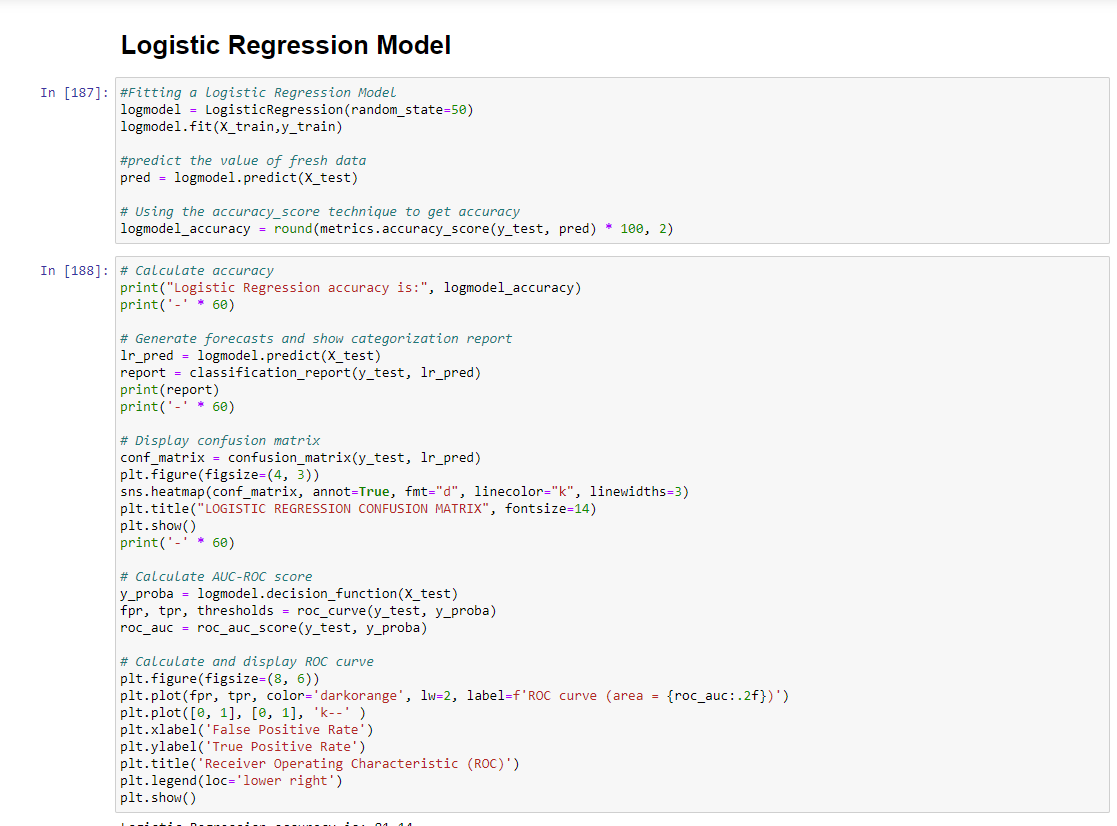
To convert categorical features into dummy variables it was done using pandas' dummy variable’s function. There are three parameters to this procedure. Number one is the dataset holding the names of the columns on which we wish to execute one hot encoding.  
The second input is a list of category and Boolean column names to be converted.   
The third input was drop\_first = true; by maintaining these parameters set to true, we dropped the first column since one column tells the value of the other column. So, if you delete male, the value of female will suffice, which is why we leave drop\_first as true here.   
 Fig 12- One Hot Encoding   
  
Please bear in mind that you cannot maintain all of the dummy variables in the model because this would result in multicollinearity. So, if you are developing a model and have made certain dummy variables, always leave out one of them. If there are three dummy variables, remove one while keeping the others.

Because the columns tenure, monthly charges, and total charges are not on the same scale, using feature scaling. Take data from these columns for training, we may not obtain optimum predictions, therefore we need to bring them on the same scale. Using a StandardScaler class from the sklearn package, and I'm creating an object of this class called StandardScalar. Then made a list of the columns that will be subjected to feature scaling. That list now has columns for feature or ft scaling.   
In the next line, I utilised the fit transform method, which is coupled with a StandardScalar object. Gave the dataset and a column list to this procedure in order to apply the scaling operation.

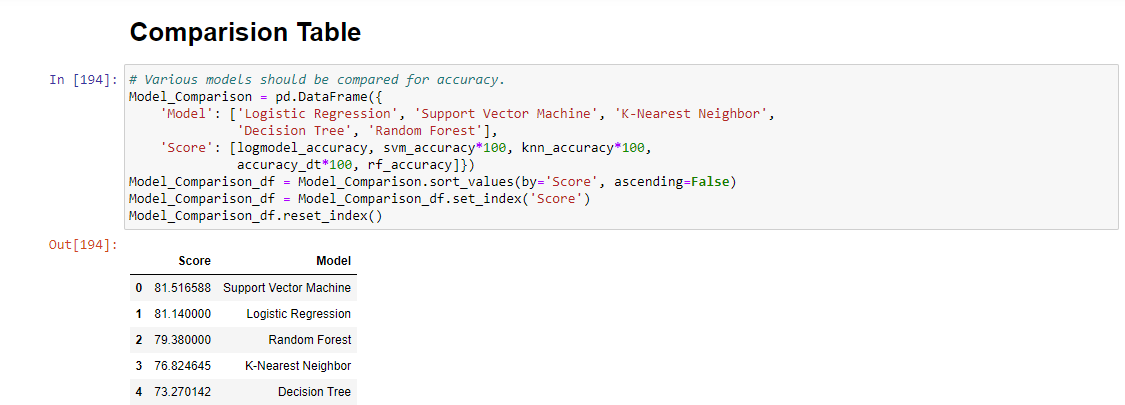
  
 Fig 13- Feature Scaling  
  
In the next line, the author only wanted to see the subset of data after one hot encoding and feature scaling. So you can see that the three columns after the customer ID column are now on the same scale. And its added additional columns with suffixes, for example.   
  
 Fig 14- New Dataset Head  
Column contract, for example, had three values: one year, two years, and month to month. The absence of a third column named contract month to month is due to the fact that the author retained the parameter drop underscore first equals true, and therefore a column with the suffix contract month to month was not added here. So referring to this specific option. Drop\_first = true, hence the first column, contract month to month, was removed from here, leaving only two columns visible.  
The next cell displays all of the columns with suffixes that were added as part of the obtain underscore dummy procedure. As you can see, all other category variables are similarly affected. As a result, all of these columns have suffixes that correspond to particular column values.  
  
 Fig 15- Suffixes and Added Columns

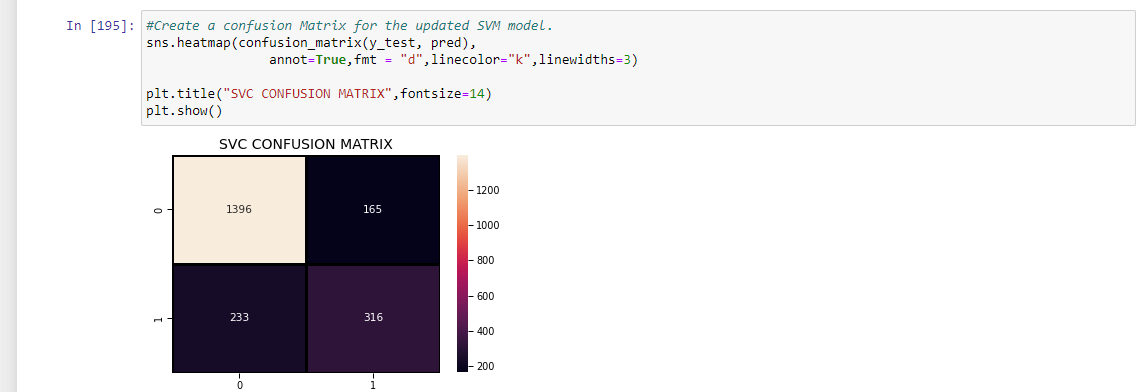
Following that, all of the feature columns are assigned to a variable called x, and the goal variable is assigned to y. So x comprises all variables except the goal variable named churn and the customer ID column, which has no importance in this model training and is therefore removed, y will only contain the target variable churn after this action.

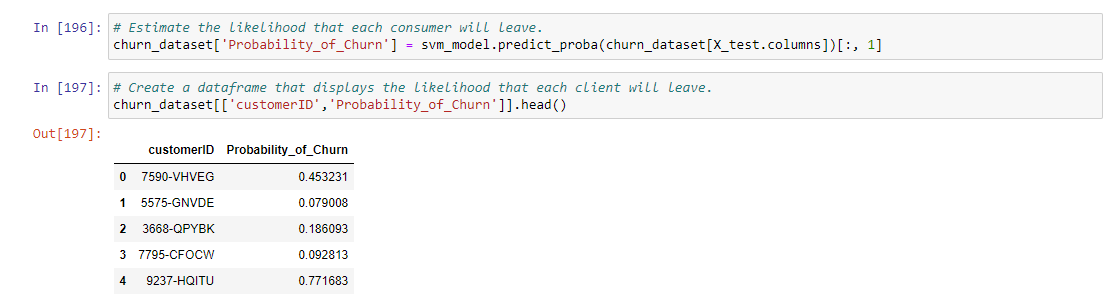
**Modelling**  
In the next cell, the author divide the entire dataset into training and test data sets. So, in the following line, used the class named train underscore test and To score is split from the sklearn.model package and invoke the test train split. This method's parameters are feature variable x, target variable y, test size 0.3 or 30%, and random state variable 50. We will use a training data set to train our model and a test data set to validate it.  
 Fig 16- Model Building  
  
The next cell is for training the logistic regression model. Make an object called logmodel of logistic regression, which will be our classifier that we'll fit to the training set we made previously. We are sending an input containing a random state equal to 50.

You can see that the input arguments to this function are x train and y train when using a fit method on logmodel to fit the model on training data set. As a result, the model will discover the relationship between the X and Y trains. And, if those correlations are established, the model will be able to forecast the retention value for fresh and previously unreported observations..   
  
 Fig 16- Logistic Regression code

That is why the Author has established a second test set to evaluate the model using previously uncovered observations or data. In the next line, we forecast the outcome of previously unknown data. To forecast the churn outcome for test observations, using the predict method on the classifier named logmodel. That is why the Author passed the X test.  
To calculate the accuracy of this logistic regression model, an argument within this predict method carries the predicted outcomes for all the observations in an excellent accuracy score technique connected with the metrics package is being employed. In this method, passing two parameters: y test, which contains actual results, and lr­\_pred, which contains predicted values. Then using a round function on this operation to determine the accuracy value to two decimal places, and trained the support vector machine, K nearest neighbour, decision tree, and random forest models in the same way. You can see that the Author has trained these models and the steps are more or less same for all of them within the Appendices.  
  
For comparison the Development of a Panda's data frame that shows a machine learning model name and matching accuracies side by side was created using two columns: number 1 is model, which includes a list of model names; and number 2 is accuracies, which contains the accuracies. The second column is called score, and it gives accuracy scores for each of them. So, in the next line, all of the row values got sorted in descending order by score.

**Evaluation**  
  
 Fig 18- Comparison Table of Machine Learning Predictive Models  
Using column score to create the index, in the final line, the utilisation of the reset function was implemented to bring all columns to the same level. Otherwise, it will display columns in the Pivot Table firm. That is why this specific strategy was deployed. So, when running the cell, you can see that it received the table below. So, there is a column named score that displays the accuracy scores, followed by the model’s name that corresponds to the model name. As you can see, the Support Vector Model has the best accuracy.

Inside this procedure, I'm supplying two arguments named confusion\_ metrics. The first parameter is y\_test, which includes the actual values, followed by pred, which contains forecasted values. I received the following result when I ran the cell.   
 Fig 19- New SVC Confusion Matrix

1396 and 316 are the two numbers. Are valid predictions, but the numbers 233 and 165 are inaccurate. So, we have around 1712 right predictions and 398 incorrect guesses, which is why I'm constructing a SVM model as this algorithm has the best accuracy of all the models, thus should use it to deploy it in production to receive the forecasts. It's usually a good idea to train various models and determine which one provides the highest accuracy before using that model as the final model.  
Adding a new column named probability of churn in the data set. To determine the likelihood of churning out for each individual client, The Author used predict proba. In the following cell using the columns customer ID and probability of churn to compare the likelihood of churn for each individual customer. As you can see, the first client has a 45.3% chance of churning, the second has a 7.9% chance, the third has a 18.6% chance, and the last has a 77.1% chance of churning. This indicates that the first or final customer will leave at any moment in the future. So, there is an opportunity to engage with them, interact with them, and if possible, give them new offers, and incentives associated with new products you have, so that they can stay longer with you make sure you have enough data about them so that you know their individual needs and accordingly, you can offer them relevant product and services, you can also try to get their feedback on various products and services before they leave. So that you have some good data points about why people are leaving then take corrective measures accordingly.   
 Fig 20- Customers Churn Probability

**Deployment**The efforts in constructing an effective churn prediction model utilising Support Vector Machine (SVM). The deployment phase is critical because it focuses on the practical application of the research study to improve data-driven decision-making and client retention methods inside our organisation.

**Model Inclusion**:  
Incorporate the SVM churn prediction model into our operational systems in this stage. This usually entails collaborating closely with the IT department to achieve a smooth transition. Once incorporated, the model becomes a component of our customer relationship management (CRM) system, allowing for real-time forecasts and suggestions.

**Monitoring and Evaluation:**Performance of the model must be regularly tracked and assessed. We established methods for constant monitoring to guarantee that the model's predictions match real consumer behaviour. If we find differences or changes in model performance, we conduct model retraining to preserve accuracy.

**Training and Documentation:**Provide training to relevant teams, such as marketing, customer support, and management, to ensure they understand how to interpret and act on the model's predictions and suggestions. For your convenience, comprehensive documentation is supplied.

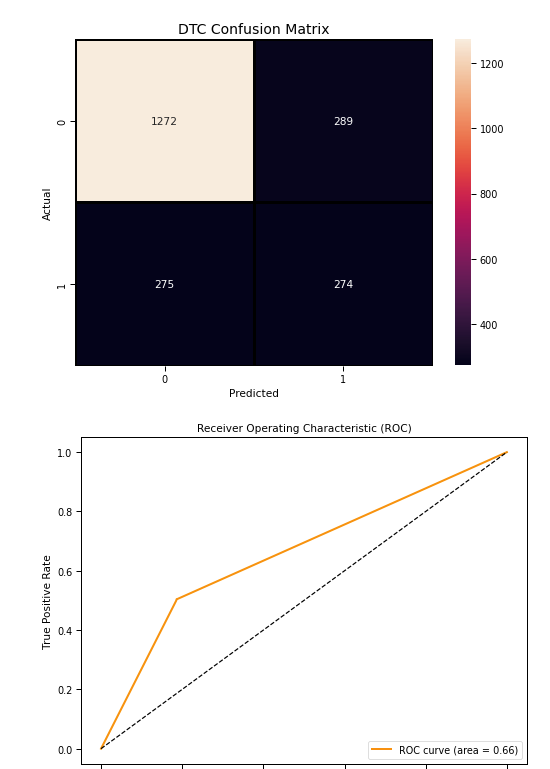
**Deployment Plan and Rollout:**A complete deployment plan describing the timing, responsibilities, and major milestones is carried out. Implement the SVM model progressively, beginning with a trial phase to detect any possible concerns before going full-scale.

**Feedback Loop:**Create a feedback loop between the teams that will be utilising the model and the data scientists who created it. This allows us to gather feedback on the model's usability and efficacy and make any necessary changes or modifications.

6. **Results**

**Results of Analysis:**

**Machine Learning:** Machine learning applied to past customer data provided remarkable insights into retention prediction. Variables such as contract length engagement frequency, and demographic characteristics were all taken into account. The support vector classifier (SVC) method emerged victorious, with a 81.5% accuracy in predicting client retention.   
   
 Fig 21- SVM Model Accuracy  
  
The study emphasised the importance of ongoing customer involvement and intelligent targeting in increasing retention rates. Visualisations such as precision-recall and correlation heat maps shed light on the algorithm's discriminative abilities, allowing for more informed decision-making. According to this concept, customers who made interacted with the company and had longer contracts were more likely to remain loyal

**Predictive Analysis:** Predictive analysis is a powerful technique for forecasting client behaviour trends across time. A rigorous analysis was performed in this work utilising machine learning methods such as logistic regression, decision trees, k-nearest neighbour, and random forests. As stated above the predicted accuracy of these models was assessed over a certain time period. The support vector classifier was most accurate with an accuracy of 81.5% in forecasting client retention. To highlight the performance measures of these models, visual representations in the form of confusion matrices and ROC curves were used, emphasising the predictive power of the algorithms.   
   
 Fig 22- Decision Tree Roc Curve and Confusion Matrix  
  
As can be seen In this study’s results, the author explored multiple data analytic approaches to establish their utility in anticipating client retention. While all approaches provided valuable data, the support vector classifier emerged as the most accurate predictors of client retention. Being able to design focused strategies for enhancing customer happiness and retention rates by combining these techniques and exploiting the insights gained by machine learning and data visualisation.

**Data Interpretation:**

**Machine Learning for Customer Insights:** The success of the SVM algorithm emphasises the critical importance of consumer behaviour patterns in retention prediction. The algorithm's ability to distinguish between retained and non-retained clients emphasises the importance of ongoing interaction and focused marketing efforts.  
It underlines the need of consistent client interaction and tailoring of promotions to the demographics, emphasising that data-driven tactics may be helpful in developing consumer loyalty.  
The machine learning insights provide strategic insights that guide organisations towards informed decisions.

**Predictive Analysis Implications:** The SVM’s forecasting power enhances its potential as a strategic tool for enterprises. The algorithm's high accuracy in projecting client retention suggests that it might be used in business strategic planning, its flexibility to complicated data linkages gives it usefulness with real-world application, allowing organisations to painstakingly estimate retention rates—a vital component in future planning, providing firms with a critical tool for accurately anticipating.

Individual customer churn ratings may change between models such as SVM and logistic regression due to variations in how models make predictions and award probability values.  
Here are some of the main reasons why the SVM model resulted in having better accuracy in predicting customer churn:

**Decision Boundary and Model Complexity:**SVM seeks the hyperplane with the greatest margin between classes. It may capture complicated, non-linear decision boundaries using techniques such as kernel functions.

**Noise and Outlier Robustness:**SVM is relatively resilient to noisy data and outliers. It focuses on determining the margin that maximises class separation, which can lead to improved generalisation and performance, particularly when the dataset contains noisy or outlier data points**.**

**Optimal Margin Separation:**The goal of SVM is to discover the decision boundary (hyperplane) that maximises the margin between classes. Because it focuses on identifying the most representative border, margin maximisation can lead to improved generalisation.

**Effective Class Separation in Imbalanced Datasets**If your churn dataset contains a class imbalance (i.e., disproportionately more non-churners than churners), SVM can handle this issue effectively. Its margin-based strategy can help to keep the model from being biased towards the majority class, which is a prevalent problem in unbalanced datasets.

**Feature Selection**SVM is sensitive to feature scaling and typically benefits from having features scaled to the same range. If feature scaling is done correctly, it can improve SVM performance.

**Interpret Results**Although SVM is less interpretable than decision trees or logistic regression, it can deliver greater prediction results. When selecting a model, the trade-off between interpretability and performance is frequently taken into account.

The superior performance of SVM in predicting customer churn in your dataset over logistic regression, decision trees, and random forests could be attributed to its ability to handle complex, non-linear relationships, robustness to noise, effective handling of class imbalance, and optimal margin separation. To derive conclusive conclusions regarding model performance, you must examine the individual characteristics of your dataset as well as the depth of hyperparameter adjustment.

**Results of Initial Research Goals:**Through the research the goals and objectives were effectively solved, the study’s goals and objectives described in this thesis introduction through rigorous analysis using different machine learning models.

**Research Question 1: Does using Data Analytics enhance retention of customers?**The study has shown that using data analytics improves client retention dramatically. To anticipate customer churn and measure the success of data analytics in keeping consumers, we used a variety of machine learning models, including Logistic Regression, Support Vector Machine (SVM), Decision Trees, and Random Forest. These models repeatedly shown that data analytics, when used correctly, leads to higher customer retention rates.  
  
**Research Question 2: What data analytics tools are used in enhancing customer retention in businesses?**The analysis discovered the exact data analytics tools and approaches used by organisations to improve customer retention. Through the study it was shown that crucial tools such as SVM, Random Forest, and Logistic Regression through case study research and the analysis of machine learning models. These techniques have been proven to be accurate in helping increase customer retention and their usefulness has been tested in a variety of sectors and business environments.

**Research Question 3: Combining machine learning results with primary research what recommendations for companies can improve customer retention through data analytics implementation?**  
Adopting a vision through the conducted Interviews and case studies in the literature review, the efficacy of various machine learning techniques, highlighting the most effective ones for client retention. Reviews, interviews, consumer feedback and satisfaction data, demonstrate a strong correlation between data-driven changes and greater customer loyalty. Through combined insights from machine learning results with primary research findings it has been provided actionable advice for businesses looking to improve customer retention using data analytics. They are as follows:  
  
**Personalised Customer Experiences:**Use customer data analytics to develop personalised customer experiences. Individual preferences and behaviour patterns uncovered through data analysis may be used to tailor product suggestions, marketing messaging, and offers.

**Loyalty Prediction:**Implement a proactive prediction system that uses machine learning models such as Support Vector Machines (SVM) to identify clients at risk of leaving. Contact these clients with customised retention incentives or interventions to resolve their problems as soon as possible. **Feedback Analysis**Collect and analyse consumer feedback on a continuous basis through surveys, reviews, and social media analysis. To improve overall satisfaction, identify areas for improvement and resolve consumer issues.

**Valuable Customer Prediction**Using predictive analytics, Focus your resources and marketing efforts on high-value clients who are likely to earn long-term income.

**Omnichannel Engagement**Use an omnichannel approach to communicate with customers across several touchpoints, including as email, social media, mobile applications, and in-store experiences. Ensure that the consumer journey is consistent and smooth across all channels.

**Retention Incentives**Create retention incentives that are based on data-driven insights. Customers who display behaviours signalling future churn, such as decreased purchase frequency, should be offered incentives, discounts, or loyalty programmes.

**Customer Support improvement**Improve customer service by leveraging previous data to forecast and handle the sorts of difficulties that clients may experience.

**Continuous Monitoring and Adaptation:**Use real-time monitoring of consumer behaviour and feedback. Maintain agility and be prepared to adjust plans as client preferences and market conditions change.

**Competitive Analysis:**Use competitive intelligence tools to continuously analyse your rivals' strategy and client retention efforts. To remain competitive in the industry, modify your own strategies.  
  
These results are intended to address diverse retention methods depending on each company's unique goals and issues. The research not only solved the research questions given, but it also met the aims outlined in this thesis. It gives actual proof of data analytics' usefulness in enhancing customer retention, defines the tools and approaches utilised for this goal, and offers practical advice to help firms implement data-driven strategies to promote customer loyalty.

7. **Discussion**

In this chapter, it is shown to dig into the complexities of the results acquired through a thorough examination of the data analytics techniques. The next sections give a comprehensive assessment of these findings, a comparison with current research, insights gained from discrepancies, and an acknowledgement of the study's limitations. Before delving into this section lets take a moment to revisit the research objectives. The investigation of how machine learning data analytics impacts client retention was the study's foundation. The findings being discussed meticulously reflect the anticipated importance of data analytics approaches in accomplishing this critical goal.

**Interpretation of Results:**

**Machine Learning**   
The results of machine learning produce intriguing findings. The dense work of impacts on client retention is revealed by the comparison of algorithms and models. This result supports the concept that machine learning's power rests in its capacity to reveal hidden complexities, possibly changing the landscape of retention tactics.

Continuing the analytical adventure, immersed in the world of Machine Learning (ML), a cornerstone of modern data analytics. This section of the study will navigate the complexities of ML results, uncovering hidden patterns and insights inside the maze of consumer data. This investigation not only aligns with the study aims, but it also illuminates new vistas that rethink the orchestration of client retention methods.

**Alignment with Research Objectives and Initial Hypotheses**  
The outlines of the ML discoveries merge nicely with the larger study goals. The study finds refuge in the of insights originating from ML, which is anchored in the aim of understanding how data analytics catalyses client retention. The basic idea, that ML has the ability to uncover hidden complexities driving client behaviour, is in sync with the numerous insights unveiled by this technology.

The correlation between ML findings and study objectives highlights the power of algorithms in uncovering patterns that influence consumer behaviour. The complicated network of associations revealed by ML analysis lends empirical credence to the concept that ML may extract subtle impacts from the interplay of variables.

**Unveiling Hidden Patterns**  
Beyond alignment, venture into machine learning reveals the underlying cadences inside consumer data. Like skilled conductors, the computers uncover patterns from seeming randomness. The ability of ML to detect even the smallest signals in the middle of noise adds a new level to understanding of customer behaviour dynamics.

For example, when ML models pinpoint the link between customer contracts and purchase payment to loyalty, an interesting surprise emerges. This realisation leads to the realisation that this customer involvement may be used as a strategic tool to influence decisions and promote long-term loyalty. The discovery of these complicated linkages paves the door for the strategic orchestration of client encounters and retention efforts.

**Predictive Analysis**  
The predictive analytic landscape emerges as a focus point. A complex interplay of algorithms and models produces predictions that weave the fabric of client behaviour. This result confirms the primary idea, heightening expectations for predictive analytics as a foundation for developing novel client retention initiatives.

Predictive analysis, a pillar of modern data analytics, emerges like a tapestry woven with numerous threads of algorithms and models, each precisely calibrated to foresee client behaviour. During this round of discussion, it is shown to go into the intricacies of predictive analytic findings, encompassing both predicted alignments with original ideas and the discovery of unknown territory.

**Alignment with Research Objectives and Initial Hypotheses**  
The predictive analysis results are consistent with the research goals outlined in this study. The accuracy of predictive models provides an echo into the quest to comprehend how data analytics catalyses client retention. The study envisioned predictive analysis as a useful tool, predicting insights into client behaviours that are crucial to retention tactics.

The alignment of predictive models with the specified assumptions demonstrates the power of these approaches in finding patterns, trends, and underlying dynamics that support customer retention. The models demonstrate data analytics' predictive capacity in decoding the mysterious language of customer behaviour, so substantiating the premise that predictive analysis is a cornerstone in the architecture of customer retention enhancement.

**Exploration of Uncharted Territories**  
In addition to the expected alignment with study objectives, the results of predictive analysis have taken us to unexplored territory, illuminating the dynamics of client behaviour in novel ways. They have discovered correlations that were previously hidden inside the maze of data thanks to their complicated algorithms and analytical abilities.

One notable result is the impact of secondary factors that, while not immediately obvious, have a significant impact on client retention. For example, predictive models revealed that consumer contracts payment method had a critical impact. This conclusion has significant implications for the design of retention strategies, implying that improving customer support experiences may offer higher retention returns.

Finally, engagement in predictive analytic findings exposes valuable insights. The similarity with research aims is reminiscent of predictive models' foresight in the field of customer retention. Uncharted territory inquiry reveals unseen effects and factors that expand the symphony of insights. The likeness with current literature increases the potential universality of predictive modelling. Among these insights that reframe the view of the temporal dynamics of consumer behaviour. As the digestion of these discoveries, keep the limits that frame insights in mind, conscious of the complex interaction between data, models, and the wider reality they strive to portray.

**Comparison with Literature:**

**Machine Learning**   
Machine learning's capacity to uncover hidden complexities. This study, on the other hand, adds a novel dimension by immediately grafting machine learning onto the canvas of client retention, exposing its practical usefulness in real-world circumstances.

ML findings are consistent with previous research themes, notably the work of Xiahou, X. and Harada, Y. (2022) Their emphasis on machine learning's ability to reveal hidden insights is similar to the authors study's investigation of client retention dynamics. In a way, the work replicates their feeling inside the particular melody of customer connections.

The findings are consistent with Shah, S.S. (2020) research, which highlighted ML's predictive power. The work adds to the picture by not only proving its predictive capability but also demonstrating its practical relevance on developing retention strategies. This dynamic movement emphasises the path from prediction to strategic action, as demonstrated by the incorporation of ML insights into retention orchestration.

**Predictive Analysis**   
The correlation between findings and past research Wassouf, W.N. *et al.* (2020) is impressive, reinforcing predictive analysis's ability to forecast customer behaviour. The research goes beyond this resonance by highlighting the path from forecast to strategic execution in the field of client retention.

Predictive analysis findings are consistent with recent study by Drachen, A. *et al.* (2016) which lauded the usefulness of predictive analysis in projecting consumer behaviour. However, The research goes beyond these models' predictive capabilities by directly tying the forecasts to retention methods, establishing a pipeline from insight to action.

This expansion of alignment resonates with Hapsari, R., Clems, M. and Dean, D. (2016) work, which demonstrated machine learning's capacity to identify latent patterns. The work offers evidence that not only verifies their findings but also expands their scope to the area of customer retention by using machine learning to consumer behaviour data. The harmonic convergence of findings and these established bodies of study emphasises the potential of predictive modelling to transcend specific sectors or areas.

**Inconsistencies and New Insights:**

**Machines Learning**Discordant notes are included among the resonant melodies of ML's results. An unexpected discrepancy emerges as ML models occasionally fail to forecast short-term swings in client behaviour. This difference encourages a more in-depth investigation of the temporal dynamics of customer encounters, as well as a more nuanced understanding of the relationship between immediate events and long-term patterns.

**Predictive analysis**   
The efficacy of the models is dependent on the quality and breadth of the available data. Incomplete or skewed data may have a subtle impact on model accuracy, refracting on conclusions through an imprecise point of view. Furthermore, the deployment of the models may face difficulties in industries or segments where data patterns differ considerably from the training data, limiting the generalizability of the findings.

**Limitations:**

**Machine Learning**Knowing the constraints that throw shadows on the discoveries. Biases included in training data might gradually alter the conclusions, lowering the quality of insights. Furthermore, the generalisation of ML-generated models to sectors or segments with dissimilar data patterns requires caution due to the threat of contextual specificity  
The journey into ML produces an oversight vibrating with alignments, insights, and reflections. The alignment with research goals highlights ML's revolutionary significance in exposing underlying patterns that guide client retention. The discovery of hidden connections adds to the insights, recalibrating the knowledge of the intricate interaction of client behaviours. The alignment with previous literature strengthens ML's significance in the larger data-driven narrative. The discordant notes of irregularity stimulate inquiry into the temporal symphony of consumer interactions among various harmonies. Cognizant of the constraints that define the findings, as well as the biases that run through algorithms and the contextual aspects that shape the retention strategies.  
  
 **Predictive Analysis**  
The efficacy of the models is dependent on the quality and breadth of the available data. Incomplete or skewed data may have a subtle impact on model accuracy, refracting the conclusions through an imprecise prism. Furthermore, the deployment of the models may face difficulties in industries or segments where data patterns differ considerably from the training data, limiting the generalizability of the findings.The absorption in predictive analytic findings exposes a symphony of alignments, discoveries, and thoughts. The correspondence with research aims is reminiscent of predictive models' foresight in the field of customer retention. Uncharted territory inquiry reveals unseen effects and factors that expand the insights. The correspondence with current literature increases the potential universality of predictive modelling. The conflicts give insights that reframe the view of the temporal dynamics of consumer behaviour. Take on board these discoveries, keep the limits that frame the insights in mind, conscious of the complex interaction between data, models, and the wider reality they strive to portray.  
8. **Conclusion**   
The findings of this Research study demonstrate the importance of data analytics in enhancing customer retention strategies. Through a thorough examination of numerous Machine learning methodologies and predictive analysis the author got significant insights into how organisations may successfully engage and maintain their consumer base.

The study in the field of predictive analysis has revealed that utilising historical data and advanced modelling methodologies may result in reliable projections of customer behaviour. This allows organisations to anticipate client demands, adjust their services, and reduce leaving rates. The ability to evaluate consumer feedback in real time enables businesses to respond quickly to concerns, improve customer experiences, and develop deeper connections.

Machine learning algorithms have demonstrated their ability to personalise interactions, allowing businesses to make personalised suggestions and offers. This level of personalisation increases not just consumer happiness but also loyalty and engagement. Furthermore, the incorporation of data visualisation tools has exposed patterns and trends within complicated information, allowing decision-makers to find actionable insights that lead strategic objectives.

This study's contributions are both theoretical and practical. Theoretically, this study supports the rising importance of data analytics in today's corporate settings. It emphasises the need of firms embracing data-driven decision-making and using the possibilities of various analytical tools. In practise, the findings have important implications for companies looking to boost client retention. Businesses may redesign their customer interaction tactics, increase operational efficiency, and ultimately improve their bottom line by implementing the approaches described in this research.

Finally, this study highlights the crucial relevance of data analytics in building client retention.   
The combination of the methods researched has resulted in actionable insights that enable organisations to create closer connections with their consumers. As the sphere of data analytics continues to steer the data-driven economy, these insights provide a road map for organisations to succeed in an increasingly competitive climate. Companies that embrace the promise of data analytics may not only retain consumers, but also set the road for future innovation and sustainable growth.

The world has experienced a paradigm change in this era of fast technical innovation and digital transformation. Businesses are no longer controlled by old approaches; instead, they are leveraging the influence of data analytics to get valuable insights into customer behaviour, preferences, and trends. This is essential in client retention, as businesses try to build longer relationships with their consumers. This study digs into the diverse environment of data analytics and its substantial influence on 21st-century consumer retention methods. Wanting to shed light on how these strategies together contribute to improving customer retention by undertaking a thorough study that combines predictive analysis through machine learning.

**Personalization and Engagement using Machine Learning**Through personalised interactions, machine learning has emerged as a powerful tool for increasing client retention. Businesses may analyse large databases using machine learning algorithms to provide personalised suggestions and services. This personalised approach not only delights customers but also fosters attachment to the company. In an era where consumer tastes are diverse and dynamic, the ability to create personalised experiences may make or break a company's ability to retain consumers.

**Anticipating Customer Behaviour Using Predictive Analysis**Predictive analysis, a process that uses previous data to project future trends and client behaviour, is one of the investigation's pillars. The study was able to precisely forecast client preferences and actions by using advanced modelling methods. This enables organisations to anticipate their consumers' demands, resulting in higher customer satisfaction and lower churn rates. Companies may adjust their offers to meet individual tastes by employing predictive research, resulting in a more personalised interaction with customers.

**Uncover Patterns and Insights Through Data Analytics**Discovering hidden patterns, connections, and trends by visualising data using graphs, charts, and interactive dashboards. Decision-makers can thereby glean useful insights from otherwise overwhelming data. This enables organisations to make educated decisions that meet the demands and expectations of their customers.

**Synthesis of Findings and Contributions**  
The outcomes of the study highlight the enormous potential of data analytics in designing client retention efforts. The combination of predictive analysis through machine learning provides a comprehensive framework for organisations to engage with consumers in a comprehensive manner. The research has shown that when these strategies are used in tandem, they may offer crucial insights that help businesses retain consumers more successfully.

**Implications for Practise and Theory**The research's consequences go beyond the immediate findings. Practically, firms can gain from using the approaches investigated in this study. They may refine consumer interaction tactics, optimise resource allocation, and improve overall operational efficiency by embracing data analytics. The theoretical implications are similarly important; the research adds to the expanding body of knowledge that highlights the importance of data analytics in current business settings.

In a larger sense, the study provides useful insights for a variety of businesses, ranging from retail to banking. Companies may build personalised experiences that drive customer loyalty and retention by leveraging the capabilities of predictive analysis through machine learning. Furthermore, the finding gives up new areas for further investigation. This involves optimising algorithmic models, investigating novel technologies like as artificial intelligence and blockchain, and investigating cross-industry applications.

**A Closing Note on Significance**

This study magnifies the pivotal role that data analytics occupies in moulding customer retention dynamics in the 21st century. The incorporation of predictive analysis through machine learning has concluded in a comprehensive understanding of customer behaviour and preferences. By embracing data-driven decision-making, companies can not only retain their customer base but also lay the foundation growing their relationship amongst their customers for longer retentions.

The insights collected from this research serve as a light of direction as firms negotiate the complexity of a developing economy. They emphasise the need of taking a comprehensive approach to data analytics since it leads the way for stronger customer interactions, better customer experiences, and long-term company success. This study's significance reverberates in the wide tapestry of data analytics, underlining the potential for dramatic change in the larger area of data-driven decision-making.

**REFERENCES**

Alam, M.M.D. and Noor, N.L.M. (2020) “*The Relationship Between Service Quality, Corporate Image, and Customer Loyalty of Generation Y: An Application of S-O-R Paradigm in the Context of Superstores in Bangladesh*,” Accessed at: https://journals.sagepub.com/doi/pdf/10.1177/2158244020924405.  
Accessed on: 8/5/2023

Andiran,B. et al (2022) “*Sentiment Analysis on Customer Satisfaction of Digital Banking in Indonesia*”  
Accessed at: <https://www.researchgate.net/profile/Bramanthyo-Andrian/publication/359684655_Sentiment_Analysis_on_Customer_Satisfaction_of_Digital_Banking_in_Indonesia/links/6247cad68068956f3c61a229/Sentiment-Analysis-on-Customer-Satisfaction-of-Digital-Banking-in-Indonesia.pdf>  
Accessed on: 9/5/2023

Akbar,M (2013 )“*Three Competing Models on Customer Loyalty in the Context of Mobile Subscribers*”  
Accessed from: https://pdfs.semanticscholar.org/204b/ec1a15db5295dd3cbdaa7f38609ccccac68a.pdf

Accessed on: 11/5/2023

Akter, S. and Wamba, S.F. (2016) “*Big Data Analytics in e-commerce: A systematic review and agenda for future research”*, *SpringerLink*. Available at: https://link.springer.com/article/10.1007/s12525-016-0219-0   
Accessed on: 4/5/2023

Al-Tit, A(2020) “*E-commerce drivers and barriers and their impact on e-customer loyalty in small and medium-sized enterprises (SMES)”*  
Accessed at: <https://www.econstor.eu/bitstream/10419/248015/1/1700196405.pdf>  
Accessed on 12/5/2023

AMUDA,K. and ADEYEMO,A.(2020) “*Customers Churn Prediction in Financial Institution Using Artificial Neural Network”* accessed from: <https://arxiv.org/ftp/arxiv/papers/1912/1912.11346.pdf>  
Accessed on: 5/5/2023  
  
Anderson, J., Jolly, L. and Fairhurst, A. (2007) “*Customer relationship management in retailing: A content analysis of retail trade journals, Journal of Retailing and Consumer Services*” Available at: https://www.sciencedirect.com/science/article/abs/pii/S0969698907000161   
Accessed: 11/5/2023

Arik, K. and Gezer,M. (2022) “*The study of indicators affecting customer churn in MMORPG games   
with machine learning models*”  
Accessed at: https://www.researchgate.net/profile/Kaan-Arik/publication/366714541\_The\_study\_of\_indicators\_affecting\_customer\_churn\_in\_MMORPG\_games\_with\_machine\_learning\_models/links/63b00ac9a03100368a415f64/The-study-of-indicators-affecting-customer-churn-in-MMORPG-games-with-machine-learning-models.pdf  
Accessed on: 9/5/2023

Barbier, G. and Liu, H. (2011) *“Data Mining in Social Media”*  
Accessed at: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=8a60b082aa758c317e9677beed7e7776acde5e4c>  
Accessed on: 9/5/23

Barbu, C.M. and Ziegler, J. (2018) “*Designing interactive visualizations of personalized review data*.” “”Available at: https://www.researchgate.net/publication/328334662\_Designing\_Interactive\_Visualizations\_of\_Personalized\_Review\_Data\_for\_a\_Hotel\_Recommender\_System   
Accessed: 7/5/2023

Ballester,M. Grau-carles,P. Sainz, J.(2017)  
“*Customer segmentation in e-commerce: Applications to the cashback business model”*  
Accessed at:  
<https://fardapaper.ir/mohavaha/uploads/2018/07/Fardapaper-Customer-segmentation-in-e-commerce-Applications-to-the-cashback-business-model.pdf>  
Accessed on: 11/5/2023

Brownlee, J. 2020 ‘*Framework for Data Preparation Techniques in Machine Learning*’ Machine Learning Mastery [Online] Accessed: <https://machinelearningmastery.com/framework-for-data-preparation-for-machine-learning/>  
Accessed on: 9/5/2023

Brownlee, J. (2020) “*Support Vector Machines for machine learning”*, Available at: https://machinelearningmastery.com/support-vector-machines-for-machine-learning/ Accessed on: 11/9/2023

Caruana,A. (2003)” *The impact of switching costs on customer loyalty: A study among corporate customers of mobile telephony*”   
Accessed at:https://link.springer.com/content/pdf/10.1057/palgrave.jt.5740113.pdf

Accessed on: 8/5/2023  
  
Chen,A.(2019) “*Steam Games: Factors that Drive Customer Retention*”  
Accessed at: <https://mmss.wcas.northwestern.edu/thesis/articles/get/1010/Chen.Athena.MMSS%20Thesis.pdf>  
Accessed on:7/5/2023  
  
Chitturi,R. (2009) “*Emotions by Design: A Consumer Perspective*”  
Accessed at:https://www.researchgate.net/profile/Ravindra-Chitturi/publication/275952326\_Emotions\_by\_Design\_A\_Consumer\_Perspective/links/554a6f360cf29752ee7c19ce/Emotions-by-Design-A-Consumer-Perspective.pdf  
Accessed on:8/5/2023

Cornell (2023) “*SVM*, *Lecture 9: SVM*. “Available at: https://www.cs.cornell.edu/courses/cs4780/2018fa/lectures/lecturenote09.html   
Accessed on: 11/9/2023

Dai, Y. and Huang, J.-H. (2021) “*A Sales Prediction Method Based on LSTM with Hyper-Parameter Search,*” Available at: <https://www.researchgate.net/publication/349111600_A_Sales_Prediction_Method_Based_on_LSTM_with_Hyper-Parameter_Search>  
Accessed on 10/5/2023

Dam,N. Dinh,T. and Menvielle,W.(2019) “*Marketing Intelligence from Data Mining Perspective — A Literature Review*”  
Accessed at: <http://www.ijimt.org/vol10/859-MK033.pdf>  
Accessed on:9/5/2023

Damsten, E. (2023) “*The role of data analytics in effective CRM*”, *Omnitas Consulting*.   
Available at: https://www.omnitas.se/the-role-of-data-analytics-in-effective-crm/   
Accessed on: 25/7/2023

Dang, T. (2023) “*The role of Data Science in Digital Marketing*, *Top Software Outsourcing Company in Vietnam” - Orient Software*.   
Available at: https://www.orientsoftware.com/blog/data-science-in-digital-marketing/   
Accessed on: 10/7/2023  
  
Dehghanpouri, H., Rostamzadeh, R.R. and Soltani, Z. (2020) “The impact of trust, privacy and quality of service on the success of E-CRM: the mediating role of customer satisfaction.” Available at: https://www.researchgate.net/profile/Reza-Rostamzadeh-2/publication/340840032\_The\_impact\_of\_trust\_privacy\_and\_quality\_of\_service\_on\_the\_success\_of\_E-CRM\_the\_mediating\_role\_of\_customer\_satisfaction/links/618c08423068c54fa5ca2396/The-impact-of-trust-privacy-and-quality-of-service-on-the-success-of-E-CRM-the-mediating-role-of-customer-satisfaction.pdf   
Accessed: 13/5/2023

Deloitte (2023) ‘*Data-Driven Change Management Using Transformation Intelligence,* Deloitte Online  
Available at: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/human-capital/us-data-driven-change-management-using-transformation-intelligence.pdf   
Accessed on: 20/3/ 2023

*Dilini, Kulawansa and Perera (2018) ‘A review of Big Data Analytics for customer relationship management ‘ online* (no date). Available at: https://ieeexplore.ieee.org/abstract/document/8736131   
Accessed on: 20/3/2023

Drachen, A. *et al.* (2016) “*Rapid prediction of player retention in free-to-play mobile games”*. Available at: https://arxiv.org/pdf/1607.03202.pdf   
Accessed on: 5/5/2023

Evidently (2023) “*How to explain the ROC AUC score and Roc Curve?*”Available at: https://www.evidentlyai.com/classification-metrics/explain-roc-curve   
Accessed on: 11/9/2023

EU Commission (2023) “*New standard contractual clauses - questions and Answers Overview*, *European Commission*.” Accessed at: https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/new-standard-contractual-clauses-questions-and-answers-overview\_en   
Accessed on: 13/5/2023

Farabi, A.A. (2020) ‘*CRISP-DM Methodology in technical view’*, *Data Science Blog*. [online] Accessed: https://www.data-science-blog.com/wp-content/uploads/2016/09/data-science-blog-logo.png.   
Accessed on: 9/5/2023   
  
Fleetwood, D. (2023) *Non-probability sampling: Types, examples, & advantages*, *QuestionPro*. Available at: https://www.questionpro.com/blog/non-probability-sampling/   
Accessed on: 3/7/2023

Gao,Y. Liu,H. and Sun,Y (2023) *” Research on customer lifetime value based on machine learning algorithms and customer relationship management analysis model*”  
Accessed at:https://www.cell.com/heliyon/pdf/S2405-8440(23)00591-1.pdf  
Accessed on: 9/5/2023

Garg et al. (2020) “*Examining the relationship between social media analytics practices and business performance in the Indian retail industry: The mediating role of customer engagement*”   
Available from :https://pdf.sciencedirectassets.com/271677/1-s2.0-S0268401220X00028/1-s2.0-S0268401219305493/am.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEKH%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJHMEUCIQDMmBKkLMZbcpc7yXI7jXMy0YCAo06fkbPGsLjj7EC7VAIgPG16c%2Biz73KtqFq%2FHEkBOF15jOazD1EXXEqh2EX%2FwWQqvAUIuv%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FARAFGgwwNTkwMDM1NDY4NjUiDMHnrMAeILdfAP6rYSqQBay41%2FeGRgEUK%2ByB2cLxdlINS5YFhhJs0yX1MwXOWTRruTYFxBACkB1%2BAjppnZFVf8mPgGqYQwEbCb1PGJwH%2FsTJH%2BPud88ObZ9wyQy%2BcMgPMVBs3IlbQw5nGUEDkFZuAUrZh9mHS2ib84hyS3NNlJ8EdEAi5gqJK4GYope3BZfbNkZB2pSDrjhMvViNaerpZ6v3YweQ5lam0cJNM7uDQlFJufM%2BU0ZHVsZmPHNrl2JtuBAliecOrl%2BoFr2Q2bBp%2Bqpe5vgJoSOIjwSmI2b2BpybsFAbS%2BKQinu49W7uqCRgtw8MZCGFsJWgivWVqMjXs4L6T46or9%2FiWKEX9RYBmX%2BaSBTV%2FLgUpC37qdW7rkpGI7nSy%2ByM9ZeGzBeuZLpK1X6iRCC7EBV6U5rUMpoHV8HtfvLdWay9tYafDzyj6Hy89vj7G1LEy%2FIyWROoqUQyiz81wF6dLyreYqQ8vcIgNkBwSJ1WDys59kmOkln9tbvlJSxRMquprEtwGOCwnSTCZuqFng1XZzNAVOJMfRSKk5bVY4cJvgJtRlBPMurIQupGqeoLSqv%2FKVt3VDG%2F8gKCDvZleUCuPxzbSM34k6cwsCLvP6A8BtFjDQP9U82hb0nkWOPH4vMOdv1mQTECEkVvqgHbL5%2BDVf8MkQiTPxAtGUCutlXRFEirS3hpSijs9VunKzaZXaHGPmEN%2BzHjuuF4JdB6Kwg4Jnnet%2BbnLoksmyfqo6ROweNyzQEs%2BiHx26LeXRHCJ5JGPt%2Fx7FEO4SXpx9Z23TVlD21M1wrEnLD1%2FgnuTG%2B0llz3wc6zLFJaPjpWosrPUEptlIU525j%2BrmgZsitHqe7S62HiNw5wZx1ZdI8qephptmEkhcWojieHExmHMND%2F96IGOrEBqXqTdtdv0Z0UoYnOjcr0sbli52ex9c9Sjc6snhzVqDhzYJ3TYeuGAaOOZYt8BsxFCfNQh4Xp2MtNHM9MW7T8BMK7tKaSEl%2FJzQiaU6DSGGfQp%2F9wxdayjwwXsdUS3jgTUD8J8QSa7%2FCD6Oydwqv4Rd9y2jTRPJgHYXeDS6uwelHdfPt%2BtbujC7gR9j9lOngTbmDQl8VxzK1yVnqsCmN4lnvIFu4YhKw9nD2I1oSiVbrU&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20230512T093123Z&X-Amz-SignedHeaders=host&X-Amz-Expires=300&X-Amz-Credential=ASIAQ3PHCVTYVVPGBHEY%2F20230512%2Fus-east-1%2Fs3%2Faws4\_request&X-Amz-Signature=881300f592382dc9484b49ad5821c00656d152a6bca35ea479286dd8d93b62df&hash=98815dc6480df829138420e62280751891a16a68c611bf15952960d18f8bfa65&host=68042c943591013ac2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S0268401219305493&tid=pdf-918216fd-5b08-4e73-8cf8-442eee9fb8f6&sid=0693944158fca74d1c6b062359aa8b428a62gxrqb&type=client  
  
Accessed on : 6/5/2023  
  
Glen, S. (2019) “*Comparing model evaluation techniques part 2: Classification and clustering”*, *Data Science Central*. Available at: https://www.datasciencecentral.com/comparing-model-evaluation-techniques-part-2/   
Accessed on: 11/9/2023  
  
Golbayani, P., Florescu, I. and Chatterjee, R. (2020) “*A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees*”, arXiv.org. Available at: https://arxiv.org/abs/2007.06617   
Accessed on: 4/5/2023

Goutam, D (2020) “*CUSTOMER LOYALTY DEVELOPMENT IN ONLINE SHOPPING: AN INTEGRATION OF E-SERVICE QUALITY MODEL AND THE COMMITMENT-TRUST THEORY*”  
Accessed at: <https://idr.nitk.ac.in/jspui/bitstream/123456789/16885/2/158041SM15F08.pdf>  
Accessed on: 11/5/2023  
  
Great Learning (2022) ‘*Why using CRISP-DM will make you a better data scientist?*’  
Accessed: https://www.mygreatlearning.com/blog/why-using-crisp-dm-will-make-you-a-better-data-scientist/ Accessed on: 9/5/2023  
  
Gupta, S. (2018) *Sentiment analysis: Concept, analysis and applications*, *Medium*. Available at: https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17   
Accessed on: 11/7/2023

Hachcham, A. (2023) “*The KNN algorithm - explanation, opportunities, limitations”*, *neptune.ai*. Available at: https://neptune.ai/blog/knn-algorithm-explanation-opportunities-limitations   
Accessed on: 11/9/2023  
  
Hapsari, R., Clems, M. and Dean, D. (2016) “*The mediating role of perceived value on the relationship between service quality and customer satisfaction: Evidence from Indonesian airline passengers, Procedia Economics and Finance*. “Available at: https://www.sciencedirect.com/science/article/pii/S2212567116000484?ref=pdf\_download&fr=RR-2&rr=7c6326284b197535   
Accessed on:5/5/2023

Harter, J. and Schmidt, F. (2002) “*Business-Unit-Level Relationship Between Employee Satisfaction, Employee Engagement, and Business Outcomes: A Meta-Analysis*.” Available at: https://members.bestbusinesscoach.ca/wp-content/uploads/2022/11/Business-Unit-Level-Relationship-Between-Employee-Satisfaction-Employee-Engagement-and-Business-Outcomes-A-Meta-Analysis-1.pdf   
Accessed: 12/5/2023

Hennig-Thurau, T., Langer, M.F. and Hansen, U. (2001) “*Modeling and managing student loyalty “*. Available at: https://www.marketingcenter.de/sites/mcm/files/downloads/research/lmm/literature/hennig-thurau\_et\_al.\_2001\_jsr\_modeling\_and\_managing\_student\_loyalty\_an\_approach\_based\_on\_the\_concept\_of\_relationship\_quality.pdf   
Accessed on: 11/5/2023

Hidayat,A. Saifullah,M. and Ishak,A. (2016) “*Determinants of Satisfaction, Trust, and Loyalty of Indonesian E-Commerce Customer”*  
Accessed at: http://www.ijem.upm.edu.my/vol10noS/11)-VOL\_10(S)2016\_Anas%20Hidayat(Determinants%20of%20Satisfaction...)Pages%20151-166.pdf  
Accessed on:8/9/2023

Ho, R.C., Withanage, M.S. and Khong, K.W. (2020) “*Sentiment drivers of hotel customers: A hybrid approach using unstructured data from online reviews.”* Available at: https://expert.taylors.edu.my/file/rems/publication/100069\_7543\_2.pdf   
Accessed on: 12/5/2023

Hotz, N. (2022) ‘*What is CRISP DM?*’, Data Science Process Alliance.

Accessed: <https://www.datascience-pm.com/crisp-dm-2/>

Accessed on: 9/5/2023

Hussain,R. Al-Nasser,A. and Hussain,Y (2014) “*Service quality and customer satisfaction of a UAE-based airline: An empirical investigation”* Accessed at: https://www.researchgate.net/profile/Rahim-Hussain-3/publication/267454805\_Service\_quality\_and\_customer\_satisfaction\_of\_a\_UAE-based\_airline\_An\_empirical\_investigation/links/59e7162aaca2721fc2304e3e/Service-quality-and-customer-satisfaction-of-a-UAE-based-airline-An-empirical-investigation.pdf  
Accessed on: 7/5/2023

Hwang, J., Park, S.-G. and Kim, I. (2020) *“Understanding motivated consumer innovativeness in the context of a robotic restaurant: The moderating role of product knowledge*,” Accessed at: <https://www.sciencedirect.com/science/article/abs/pii/S144767702030156X>  
Accessed on 5/5/2023

IBM (2023) *What is predictive analytics?*, *IBM*.   
Available at: https://www.ibm.com/topics/predictive-analytics   
Accessed on: 15/7/2023

IBM (2023) “*What is Random Forest?”*Available at: https://www.ibm.com/topics/random-forest   
Accessed on: 11/9/2023

Insight. (2023) *How big data and Artificial Intelligence Are Transforming Business*, *Insight*. Available at: https://www.insight.com/en\_US/content-and-resources/2021/how-big-data-and-artificial-intelligence-are-transforming-business.html   
Accessed on: 23/7/2023

Jahanzeb, S., Fatima, T. and Khan, M.B. (2011) “*An empirical analysis of customer loyalty in Pakistan’s telecommunication industry,*”  
Accessed at: https://link.springer.com/article/10.1057/dbm.2011.2  
Accessed on:8/5/2023

Jain, H., Khunteta, A. and Srivastava, S. (2020) “*Churn Prediction in Telecommunication using Logistic Regression and Logit Boost,*”  
Accessed at: https://www.sciencedirect.com/science/article/pii/S1877050920306529?ref=pdf\_download&fr=RR-2&rr=7c6465632ddd413c  
Accessed on:9/5/2023

Jain, P.K. and Pamula, R. (2020) “*A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews*,” arXiv.org. Available at: https://arxiv.org/abs/2008.10282   
Accessed on: 6/5/2023

Jiang,H and Zhang,Y (2016) “*An investigation of service quality, customer satisfaction and loyalty in China's airline market*”  
Accessed at: <https://isidl.com/wp-content/uploads/2017/06/E4077-ISIDL.pdf>  
Accessed on: 7/5/2023

Jiang,L. Jun,M. and Yang, Z. (2015) “*Customer-perceived value and loyalty: how do key service quality dimensions matter in the context of B2C e-commerce*?” Accessed at : [https://www.researchgate.net/profile/Zhilin-Yang-9/publication/276398835\_Customer- perceived\_value\_and\_loyalty\_how\_do\_key\_service\_quality\_dimensions\_matter\_in\_the\_context\_of\_B2C\_e-commerce/links/56a1dd3f08ae27f7de26e415/Customer-perceived-value-and-loyalty-how-do-key-service-quality-dimensions-matter-in-the-context-of-B2C-e-commerce.pdf?\_sg%5B0%5D=started\_experiment\_milestone&\_sg%5B1%5D=started\_experiment\_milestone&origin=journalDetail](https://www.researchgate.net/profile/Zhilin-Yang-9/publication/276398835_Customer-%20%20perceived_value_and_loyalty_how_do_key_service_quality_dimensions_matter_in_the_context_of_B2C_e-commerce/links/56a1dd3f08ae27f7de26e415/Customer-perceived-value-and-loyalty-how-do-key-service-quality-dimensions-matter-in-the-context-of-B2C-e-commerce.pdf?_sg%5B0%5D=started_experiment_milestone&_sg%5B1%5D=started_experiment_milestone&origin=journalDetail)

Accessed on: 11/5/23  
  
  
Joby, A. (2023)” *K nearest neighbor or KNN algorithm and it’s essence in ML*. “Available at: <https://learn.g2.com/k-nearest-neighbor>  
Accessed on: 11/9/2023

K, G. (2020) “*Machine learning basics: Decision tree regression”*, *Medium*. Available at: https://towardsdatascience.com/machine-learning-basics-decision-tree-regression-1d73ea003fda Accessed on: 11/9/2023

Khandelwal, R. (2020) “*Quick and easy explanation of logistics regression”*, *Medium*.  
Available at: https://towardsdatascience.com/quick-and-easy-explanation-of-logistics-regression-709df5cc3f1e   
Accessed on: 11/9/2023

Kaur, H. and Soch, H. (2012) “*Validating Antecedents of Customer Loyalty for Indian Cell Phone Users*,” Accessed at: <https://journals.sagepub.com/doi/pdf/10.1177/0256090920120404>  
Accessed on:8/5/2023

Kevin,M and Ana , F(2019 )“*Customer Relation Management, Smart Information Systems and Ethics*”

Accessed on: https://pdf.sciencedirectassets.com/778421/1-s2.0-S2515856219X51007/1-s2.0-S251585622030016X/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjELr%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJHMEUCIQD6E0R4%2FK5YUKk%2F2jfBYJvLw7M4%2BtaxvR2%2BoR5RYN%2Bd1wIgROTZ3LfrwR5%2BoRcHsrqh5lL6vRHJnxvDrwqAHWy%2B2HwqvAUI0%2F%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FARAFGgwwNTkwMDM1NDY4NjUiDMC5eFOXPR6x210V8SqQBeSEy3%2FSDNn2%2B%2B3njLBENWFNd%2F18bTUmZ2umLmz3eIoRfP4%2BWEejzTiA1RAvlsw0eN7dNDhsv2g0xXXRB9XbZoXOKJBpR%2F%2FPdF%2FyUl%2BK91BR4s9OTyBt8qEY%2FIUZnjeZs88joBvsUCKkHSWeguIQxBdHstRHTq9eJQDCNr%2FxrRYXiqbsrkT9sFOFL4BQweAVAqe5h3zfPyCp7YVe2S3OBdgCtVrqTzg3d1um7PgqJsh%2FlxZkiJOV84tspmGJJb59Umgc%2B6OC%2B6tENl09J3T0UyA8CCpiyrGJFCOtFrqpgU6ZR9K%2BQnHvcu8QwV7ceozyrOdx2KmLr%2FfUv2oySPJnzYAuUSLFl6Kn1jd9DhR5cbRuncLknUmrL525yBP%2F3sbKA2NvQV%2BnEUXUpvSmuynaIlZ1DKkt27KbSbOMuEtIZQT5%2FfDaGytZE3Z9tQViV4LTp4N4Kk5eZoWPx42Xq5wuQXrm5LRPcw9h3zJvC%2FXpbijZLb%2FLts73egqnJvKOGe0Y2xstZrnBeU8%2FC8N6miZM2HKqP1ydOZx5groETUAPrBuJgUpnJcc0W7F%2BfGDb6jsN0oIdCy%2F3CvFdBJRP4Uttt8vjGpT2%2BIzLm2ARqSCgeS%2Bu7wFpCLWaMXj8I8mTt4EF7%2FdRJOhugYoS8x9QxRR%2BXoNwdGjgtKBM40di03y6abhrbV%2FSe14ZYXKtrBze2NjQNCTBIN9BlU6K55rL9XtWAH8mIfcqnjb%2FcRtBN8xVx9GImqOKv9PXZ9DBsXbRpR9V%2BOGqZCoGMPbe878humpy%2FDhLjIokiWN%2Bm0LGxppI%2B4ctKcJq7reFc2mRiPCk3awd9txohOFglvzU4dxLduWpUK39apHQ5kTrFY3AUa9k3U5CMK%2B%2B%2FaIGOrEB6%2FL8b64wLjUw%2FHaBpSAtkwaQNAepe4mREQvf0UWHT5D1rzFAK3Opvaa5kXV%2BzfhpLhIipkKMgebS6i9eR9Sc%2FrkxNNaGKf%2B6kdZEEtN%2B6us%2B85UG0JIPVfcIt%2FMWS7aOw3EyrvH57RflImu5bKrt6HpK0xU3Wq1yi791b5F0O%2FXDL1qIlEAWhtlMHQXW3OIN5tOUa5qy2N0uUDfF12owa4G%2F7nkxHhNtHHCr3vdEwK6E&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20230513T103610Z&X-Amz-SignedHeaders=host&X-Amz-Expires=300&X-Amz-Credential=ASIAQ3PHCVTYRL3UFEJS%2F20230513%2Fus-east-1%2Fs3%2Faws4\_request&X-Amz-Signature=445b0102d4534a5155002c25cc19c512910d4d2d57d421d3f61d33204e23ee49&hash=4dc85e675622a0b4becd144a06c0546cc6ecdbee96bd459cb32718b5b04a7741&host=68042c943591013ac2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S251585622030016X&tid=spdf-6f5990c5-1f9e-4bae-926b-005c8deed6e2&sid=0693944158fca74d1c6b062359aa8b428a62gxrqb&type=client&tsoh=d3d3LnNjaWVuY2VkaXJlY3QuY29t&ua=0a02550152045a5655&rr=7c6a42471d5c7539&cc=ie  
Accessed on: 12/5/2023

Khrais, L.T. (2020) “*Role of artificial intelligence in shaping consumer demand in e-commerce*”, MDPI. Accessed at: https://www.mdpi.com/1999-5903/12/12/226   
Accessed on: 4/5/2023

Kim,J. and Hastak, M (2018)” So*cial network analysis: Characteristics of online social networks after a disaste*r” Accessed At: <https://e-tarjome.com/storage/panel/fileuploads/2019-03-02/1551517800_E11862-e-tarjome.pdf>  
Accessed on:9/5/2023

Kim, W.H., Lim, H.J. and Brymer, R.A. (2015) “*The effectiveness of managing social media on hotel performance*,” *International Journal of Hospitality Management*, 44, pp. 165–171. Accessed at: <https://www.sciencedirect.com/science/article/abs/pii/S0278431914001704>  
Accessed on: 5/5/2023

Kim, S.-L. and Lee, H. (2022) “*Customer Churn Prediction in Influencer Commerce: An Application of Decision Trees*,”. Available at: <https://www.sciencedirect.com/science/article/pii/S1877050922001703?ref=pdf_download&fr=RR-2&rr=7c6465633de2413c>  
Accessed on 9/5/2023

Kitapchi,O et al. (2013)“*The paths from service quality dimensions to customer loyalty: An application on supermarket customers*”  
Accessed at: <https://www.researchgate.net/profile/Ibrahim-Dortyol/publication/263039992_The_paths_from_service_quality_dimensions_to_customer_loyalty_An_application_on_supermarket_customers/links/55dc286108aed6a199ac7a86/The-paths-from-service-quality-dimensions-to-customer-loyalty-An-application-on-supermarket-customers.pdf>  
Accessed on:11/5/2023

Kohavi, R. et al.(2004) “*Lessons and Challenges from Mining Retail E-Commerce Data*”  
Accessed at: <http://mail.im.tku.edu.tw/~cjou/bi2009/DM-lessons.pdf>  
Accessed on: 11/5/2023

Leachmann, L. and Scheibenreif, D. (2023) *Using technology to create a better customer experience*, *Harvard Business Review*. Available at: https://hbr.org/2023/03/using-technology-to-create-a-better-customer-experience   
Accessed on: 2/9/2023

Li, L. *et al.* (2022) “ *INVESTIGATING THE EFFECT OF ARTIFICIAL INTELLIGENCE ON CUSTOMER RELATIONSHIP MANAGEMENT PERFORMANCE IN E-COMMERCE ENTERPRISES”*. Available at: http://www.jecr.org/sites/default/files/2023vol24no1\_Paper5.pdf   
Accessed: 13/5/2023  
  
Liddle, J. (2023) “*The Real Big Data Gold Mine is unstructured”*, *Nasuni*.   
Available at: https://www.nasuni.com/blog/the-real-big-data-gold-mine-is-unstructured/  
Accessed on: 7/7/23

Logunova, I. (2022) “*K-Nearest Neighbors (KNN) algorithm for Machine Learning”*, *Serokell Software Development Company*. Available at: https://serokell.io/blog/knn-algorithm-in-ml   
Accessed on: 11/9/2023

Long, C.L.S. *et al.* (2020) “*Impact of CRM factors on customer satisfaction and Loyalty*.”  
 Available at: https://pdfs.semanticscholar.org/387f/a6baa9b571e8ba49a7de3aa5d224c7700300.pdf   
Accessed on: 13/5/2023

Lukic,J (2015) ‘*Leadership challenges for the big data era’* Research gate [online]  
Accessible from: https://www.researchgate.net/publication/317174920\_Leadership\_Challenges\_for\_the\_Big\_Data\_Era  
Accessed on: 12/5/2023

Mahajan, V., Misra, R. and Mahajan, R. (2017) “*Review on factors affecting customer churn in telecom sector,”* Accessed at: <https://www.researchgate.net/publication/319023470_Review_on_factors_affecting_customer_churn_in_telecom_sector>  
Accessed on: 8/5/2023

Marwa *et al.* (2019) “*The effect of customer relationship management practices on airline customer loyalty*” Available at: https://www.econstor.eu/bitstream/10419/215629/1/5-2-2.pdf   
Accessed: 4/5/2023

Marx, A. (2021) “*Customer retention analytics: 5 strategies to Reduce Churn*, *Thematic*.”   
Available at: https://getthematic.com/insights/5-ways-data-and-text-analytics-improve-customer-retention/   
Accessed on: 4/7/2023

Mcafee, A. and Brynjolfsson, E. (2014) “*Big Data: The management revolution*”, *Harvard Business Review*.   
Available at: https://hbr.org/2012/10/big-data-the-management-revolution   
Accessed on: 25/7/2023

Mirzaei, T. and Iyer, L. (2014) “*Application of predictive analytics in Customer Relationship Management: A literature review and classification”*, *AIS Electronic Library (AISeL)*.   
Available at: https://aisel.aisnet.org/sais2014/23/   
Accessed on: 20/3/2023

Misra,R. Mahajan,R. Mahajan,V (2017)“*Review on factors affecting customer churn in telecom sector”*Accessed at:https://www.researchgate.net/profile/Richa-Misra-2/publication/319023470\_Review\_on\_factors\_affecting\_customer\_churn\_in\_telecom\_sector/links/5dc535da92851c818036faf7/Review-on-factors-affecting-customer-churn-in-telecom-sector.pdf  
Accessed on:7/5/2023

Narkhede, S. (2021) “*Understanding confusion matrix”*, *Medium*. Available at: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62   
Accessed on: 11/9/2023

Navlani, A. (2019) “*Scikit-learn SVM tutorial with Python (Support Vector Machines)”*. Available at: <https://www.datacamp.com/tutorial/svm-classification-scikit-learn-python>  
Accessed on: 11/9/2023

Nolis, J. (2020) “*You’re relying on Data Too Much*”.*Medium*.   
Available at: https://towardsdatascience.com/youre-relying-on-data-too-much-250d4edc70c3   
Accessed on: 27/8/2023

Olson, S. and Writer, S. (2023) *Customer retention basics, 8 strategies, and metrics*, *Zendesk*. Available at: https://www.zendesk.com/blog/customer-retention/   
Accessed on: 28/8/2023

Ostrowski, P., O’Brien, T.P. and Gordon, G.J. (1993) “*Service Quality and Customer Loyalty in the Commercial Airline Industry,*” Accessed at: <https://journals.sagepub.com/doi/abs/10.1177/004728759303200203?journalCode=jtrb>  
Accessed on:7/5/2023

Pai, A. (2023) *What is tokenization in NLP? here’s All you need to know*, *Analytics Vidhya*. Available at: https://www.analyticsvidhya.com/blog/2020/05/what-is-tokenization-nlp/   
Accessed on: 8/4/2023

Pancras , J. *et al.* (2006) “*Issues and Perspectives in Global Customer Relationship Management*.” Available at: https://www.researchgate.net/profile/Joseph-Pancras/publication/42428087\_Issues\_and\_Perspectives\_in\_Global\_Customer\_Relationship\_Management/links/5bfed90292851c63caafb01d/Issues-and-Perspectives-in-Global-Customer-Relationship-Management.pdf   
Accessed: 13/5/2023

Peng, J., Quan, J. and Zhang, S. (2013) “*Mobile phone customer retention strategies and Chinese e-commerce,”*Accessed from: https://www.sciencedirect.com/science/article/abs/pii/S1567422313000409  
Accessed on:8/5/2023

Perianez, A. et al. (2017) “*Churn prediction in Mobile Social Games: Towards a complete assessment “* Available at: https://arxiv.org/pdf/1710.02264   
Accessed on: 11/5/2023

Qualtrics (2022) *Customer retention: Strategic plans & tactics to improve*, *Qualtrics*. Available at: https://www.qualtrics.com/uk/experience-management/customer/customer-retention-programmes/   
Accessed on: 11/7/2023

Qualtrics (2023) *What is customer sentiment and how do you measure it?Qualtrics*, *Qualtrics*. Available at: https://www.qualtrics.com/experience-management/customer/customer-sentiment/   
Accessed on: 11/7/2023

Raquib, S. (2023) “*Tracking customer behaviour and preferences with Data Analytics*, *ARS Analytics*”  
Available at: https://www.arsanalytics.com/post/tracking-customer-behavior-and-preferences-with-data-analytics   
Accessed on: 10/8/2023

Raschka, S. (2023) “*Confusion Matrix: Creating a confusion matrix for model evaluation*, *creating a confusion matrix for model evaluation*.” Available at: https://rasbt.github.io/mlxtend/user\_guide/evaluate/confusion\_matrix/   
Accessed on: 11/9/2023

Razin et al.(2021)“A *Long Short-Term Memory (LSTM) Model for Business Sentiment Analysis Based on Recurrent Neural Network*”  
Accessed at: <https://www.researchgate.net/profile/Bapayya-Kommula/publication/348748584_An_Efficient_Energy_Management_of_Hybrid_Renewable_Energy_Sources_Based_Smart-Grid_System_Using_an_IEPC_Technique/links/60586be8a6fdccbfeaf8b25e/An-Efficient-Energy-Management-of-Hybrid-Renewable-Energy-Sources-Based-Smart-Grid-System-Using-an-IEPC-Technique.pdf#page=18>  
Accessed on:8/5/2023

Richard, J., Thirkell, P. and Huff, S. (2021) “*An examination of Customer Relationship Management (CRM) technology”* . Available at: https://www.researchgate.net/profile/James-Richard-3/publication/232863119\_An\_Examination\_of\_Customer\_Relationship\_Management\_CRM\_Technology\_Adoption\_and\_its\_Impact\_on\_Business-to-Business\_Customer\_Relationships/links/55dbc6cc08aed6a199ac6924/An-Examination-of-Customer-Relationship-Management-CRM-Technology-Adoption-and-its-Impact-on-Business-to-Business-Customer-Relationships.pdf   
Accessed on: 4/5/2023

Ross, B. (2023) “*How customer expectations are changing digital strategies*” *Linchpin SEO*.   
Available at: https://linchpinseo.com/how-customer-expectations-are-changing-digital-strategies/   
Accessed on: 1/7/2023

Sabbeh,s (2018)|”*Machine-Learning Techniques for Customer Retention: A Comparative Study*”  
Accessed at: <https://pdfs.semanticscholar.org/2a9f/505e1ab148aa3d91810f509ee133272be554.pdf>  
Accessed on 5/5/2023

Saha, D. (2021) “*Google cloud brandvoice: How the world became data-driven, and what’s next*” *Forbes*. Available at: https://www.forbes.com/sites/googlecloud/2020/05/20/how-the-world-became-data-driven-and-whats-next/   
Accessed on: 17//5/2023

Saha, L. *et al.* (2021) “Amalgamation of Customer Relationship Management and data analytics in different business sectors-A systematic literature review,” *MDPI*.. Available at: <https://www.mdpi.com/2071-1050/13/9/5279>  
Accessed on: 20/3/ 2023

Salon, D.S. (2023) *Using NLP to analyze customer feedback*, *DSS Roundtable*. Available at: https://roundtable.datascience.salon/using-nlp-to-analyze-customer-feedback   
Accessed on: 15/8/2023

Sarib, M and Mashuri, S. (2022)  *“Semi-structured Interview: A Methodological Reflection on the Development of a Qualitative Research Instrument in Educational Studies”*. Available at: https://www.researchgate.net/publication/358893176\_Semi-structured\_Interview\_A\_Methodological\_Reflection\_on\_the\_Development\_of\_a\_Qualitative\_Research\_Instrument\_in\_Educational\_Studies   
Accessed on: 23/7/2023

Sayad, S. (2023) “*Logistic Regression*“Available at: https://saedsayad.com/logistic\_regression.htm Accessed on: 11/9/2023

Shah, S.S. (2020) “Developing promotional model using customer lifetime value score “  
Available at: https://norma.ncirl.ie/4417/1/shreysanjayshah.pdf   
Accessed on: 6/5/2023

Sharma, N. (2022) “*Logistic regression “*, *Brainalyst*. Available at: https://learning.brainalyst.in/logistic-regression/   
Accessed on: 11/9/2023

Shobana, J. (2023) *E-commerce customer churn prevention using machine learning-based business intelligence strategy*, *Measurement: Sensors*. Available at: https://www.sciencedirect.com/science/article/pii/S2665917423000648   
Accessed on: 6/9/2023.

Shobana, J.V. and Murali, M. (2021) “*An efficient sentiment analysis methodology based on long short-term memory networks*,”, Available at: <https://link.springer.com/article/10.1007/s40747-021-00436-4>  
Accessed on: 10/5/2023

Shrimal,D and Patil, H(2020)”*A Qualitative Approach to Customer Segmentation and Customer Churn Application”* Accessed from: <https://ijaem.net/issue_dcp/A%20Qualitative%20Approach%20to%20Customer%20Segmentation%20and%20Customer%20Churn%20Application.pdf>  
Accessed on:7/5/2023

Siddique, A. (2023) “*Exploring KNN with Different Distance Metrics*.” Available at: https://blog.devgenius.io/exploring-knn-with-different-distance-metrics-85aea1e8299   
Accessed on: 11/9/2023

Soltani , Z. and Navimipour, N.J. (2016) “*C*omputers in human behavior , Customer relationship management mechanisms: A systematic review of the state of the art literature and recommendations for future research.” Available at: http://modir3-3.ir/article-english/ISI/isi1-9785.pdf   
Accessed on: 11/5/2023

Stieglitz,S. and Dang-Xuan,L (2012) “*Social media and political communication: a social media analytics framework”*Accessed at: <http://cloud.politala.ac.id/politala/Jurnal/JurnalTI/Jurnal%2020/10.1007_s13278-012-0079-3.pdf>  
Accessed on:9/5/2023

Syaqirah, N. (2014) “*Managing customer retention of hotel industry in Malaysia, Managing Customer Retention of Hotel Industry in Malaysia*.” Available at: https://www.sciencedirect.com/science/article/pii/S1877042814029553?ref=pdf\_download&amp;fr=RR-2&amp;rr=7c5d1f412a810752   
Accessed on: 4/5/2023

Tariq, M. *et al.* (2021) “*Distributed model for customer churn prediction using convolutional neural network*.” Available at: https://www.researchgate.net/profile/Muhammad-Usman-Tariq-2/publication/351813838\_Distributed\_model\_for\_customer\_churn\_prediction\_using\_convolutional\_neural\_network/links/60c082baa6fdcc512816d582/Distributed-model-for-customer-churn-prediction-using-convolutional-neural-network.pdf   
Accessed on: 5/5/2023

Tibco (2023) “*Probability calculation using logistic regression”*, *TIBCO* Available at: https://docs.tibco.com/pub/sfire-dsc/6.5.0/doc/html/TIB\_sfire-dsc\_user-guide/GUID-C4D05ED0-3392-4407-B62A-7D29B26DC566.html   
Accessed on: 11/9/2023

Toivonen, T. *et al.* (2019) “*Social media data for conservation science: A methodological overview*,”   
Accessed At: <https://www.sciencedirect.com/science/article/pii/S0006320718317609>  
Accessed on:9/5/2023

Ullah,I et al. (2019) “A *Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector*”  
Accessed at: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8706988.  
Accessed on:9/5/2023

Vaidya, D. (2023) “*Decision tree - what is it, uses, examples, vs Random Forest”*.   
Available at: https://www.wallstreetmojo.com/decision-tree/  
Accessed on: 11/9/2023

Vivekanandan, M. (2023) “*Random Forest - a multitude of decision trees”*, *LinkedIn*. Available at: https://www.linkedin.com/pulse/random-forest-multitude-decision-trees-madhavan-vivekanandan Accessed on: 11/9/2023

Wang,A. and Marikannan, Booma. (2020) “*Optimising e-commerce customer satisfaction with machine learning”*Accessed at: <https://iopscience.iop.org/article/10.1088/1742-6596/1712/1/012044/pdf>  
Accessed on: 9/5/2023

Wassouf, W.N. *et al.* (2020) “*Predictive analytics using big data for increased customer loyalty: Syriatel Telecom Company Case Study - Journal of Big Data”*, *SpringerLink*. Springer International Publishing. Available at: https://link.springer.com/article/10.1186/s40537-020-00290-0#Sec29   
Accessed on: 20/3/2023.

Weinstein, A. (2001) “*Customer retention: A usage segmentation and customer value approach”.* Available at: https://link.springer.com/content/pdf/10.1057/palgrave.jt.5740051.pdf   
Accessed: 12/5/2023  
  
WU,S. et al (2021) “*Integrated Churn Prediction and Customer Segmentation Framework for Telco Business”* Accessed from: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9406002  
Accessed on: 5/5/2023

Xhema,J. and Metin,H. (2018) “*Switching-Costs, Corporate Image and Product Quality effect on Customer Loyalty: Kosovo Retail Market*”  
Accessed at: <https://www.researchgate.net/profile/Hasan-Metin/publication/329149878_Switching-Costs_Corporate_Image_and_Product_Quality_effect_on_Customer_Loyalty_Kosovo_Retail_Market/links/5e05f94b4585159aa49e4ddd/Switching-Costs-Corporate-Image-and-Product-Quality-effect-on-Customer-Loyalty-Kosovo-Retail-Market.pdf>  
Accessed on:11/5/2023

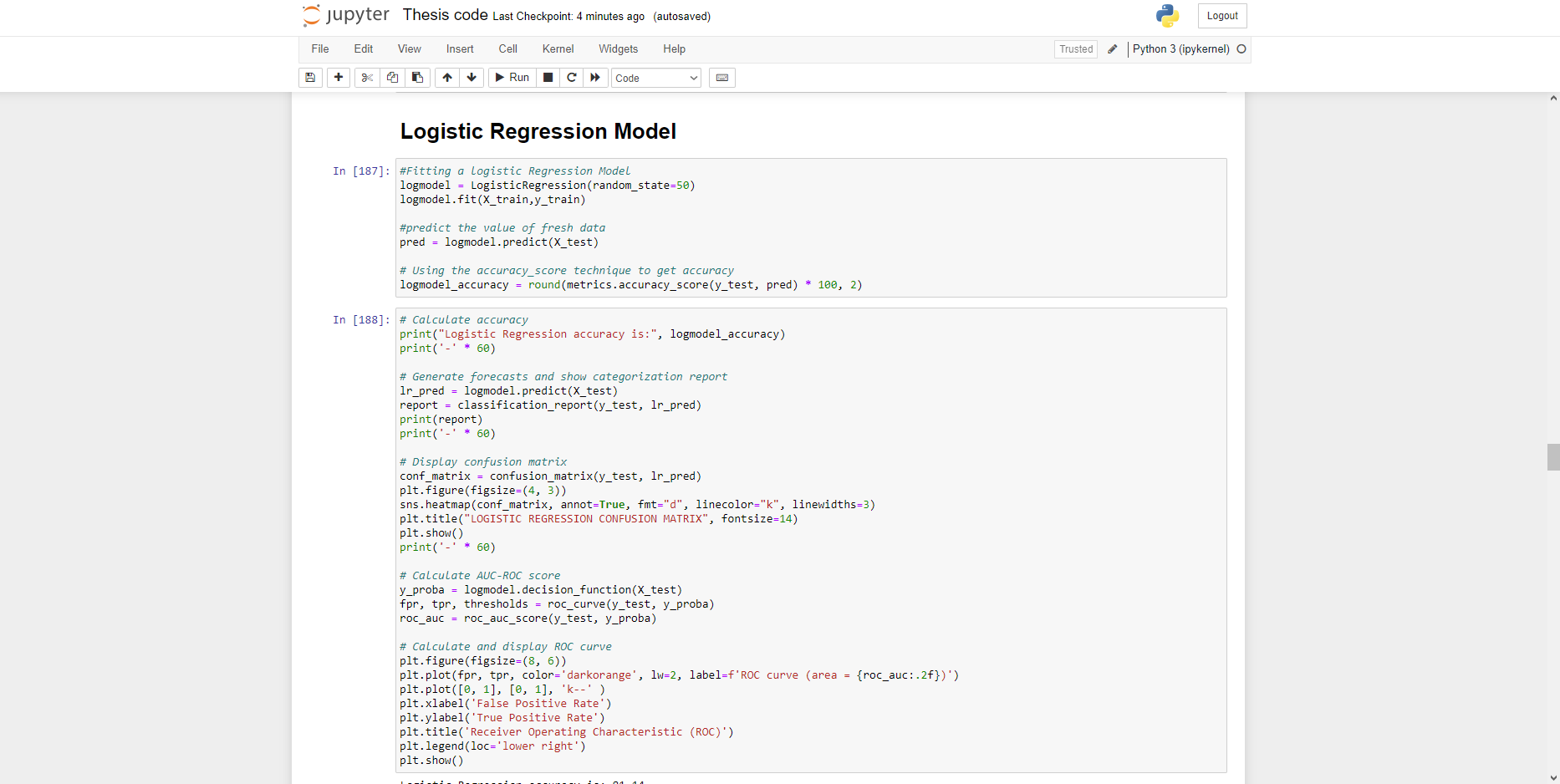
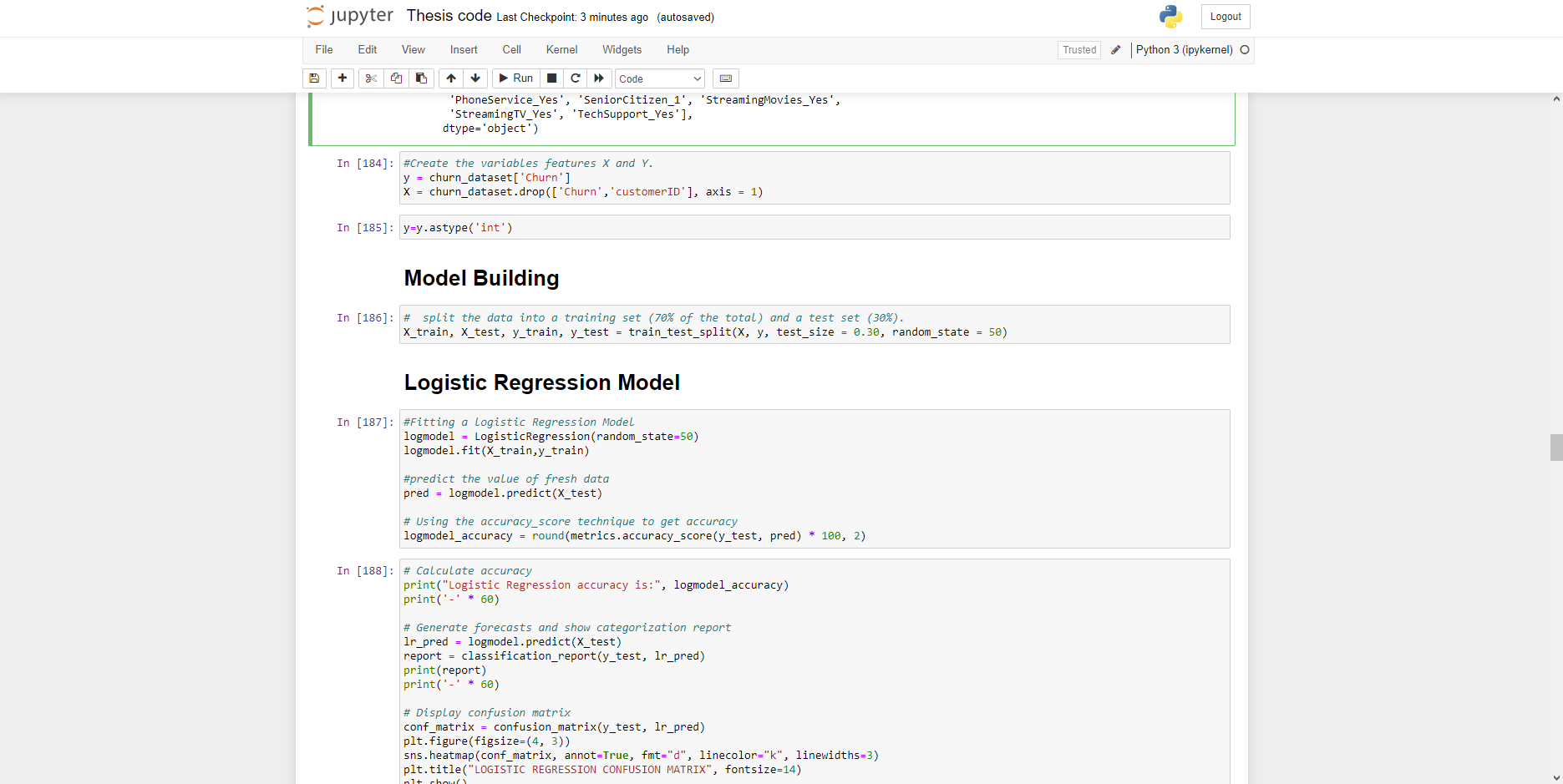
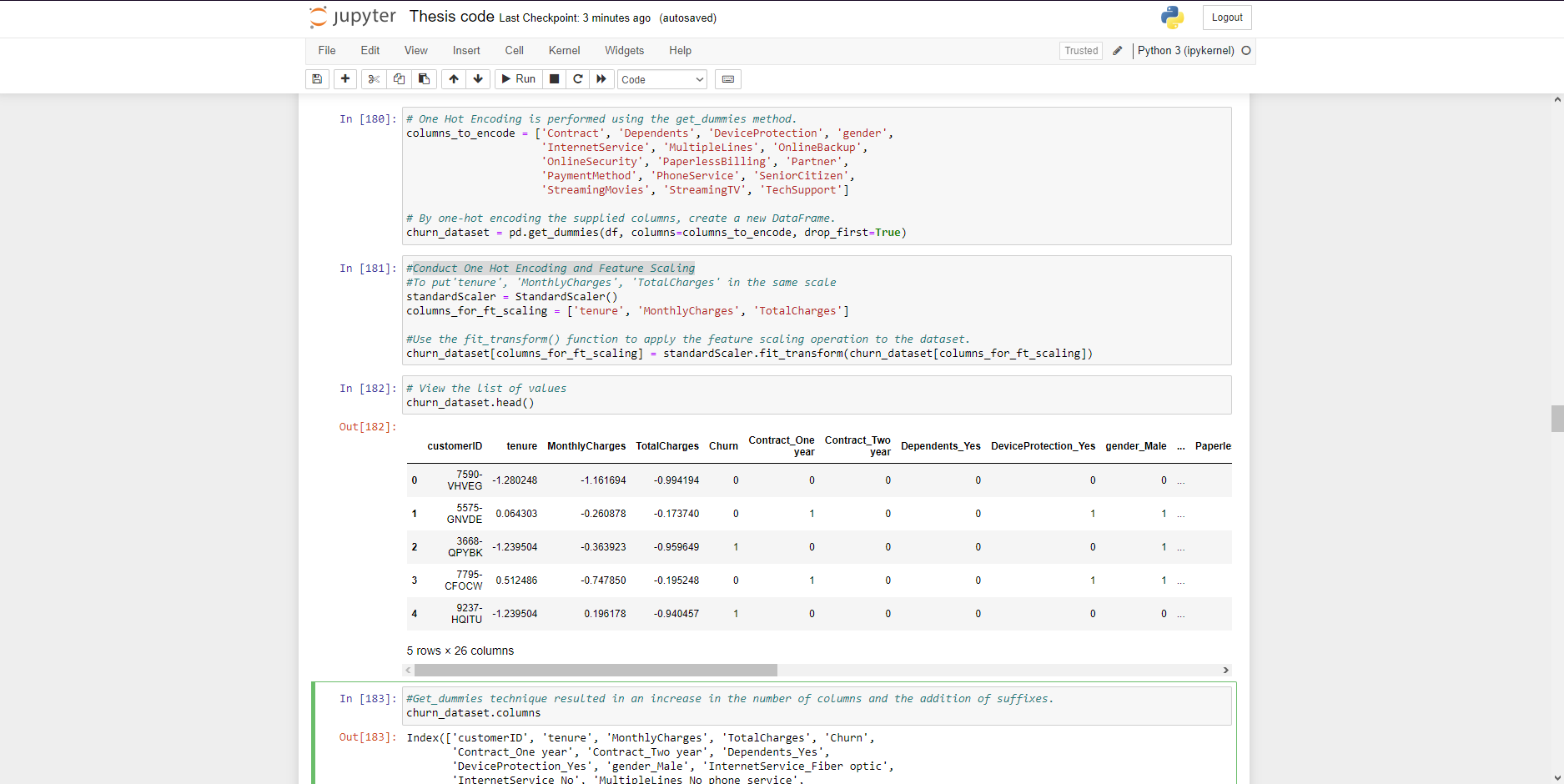
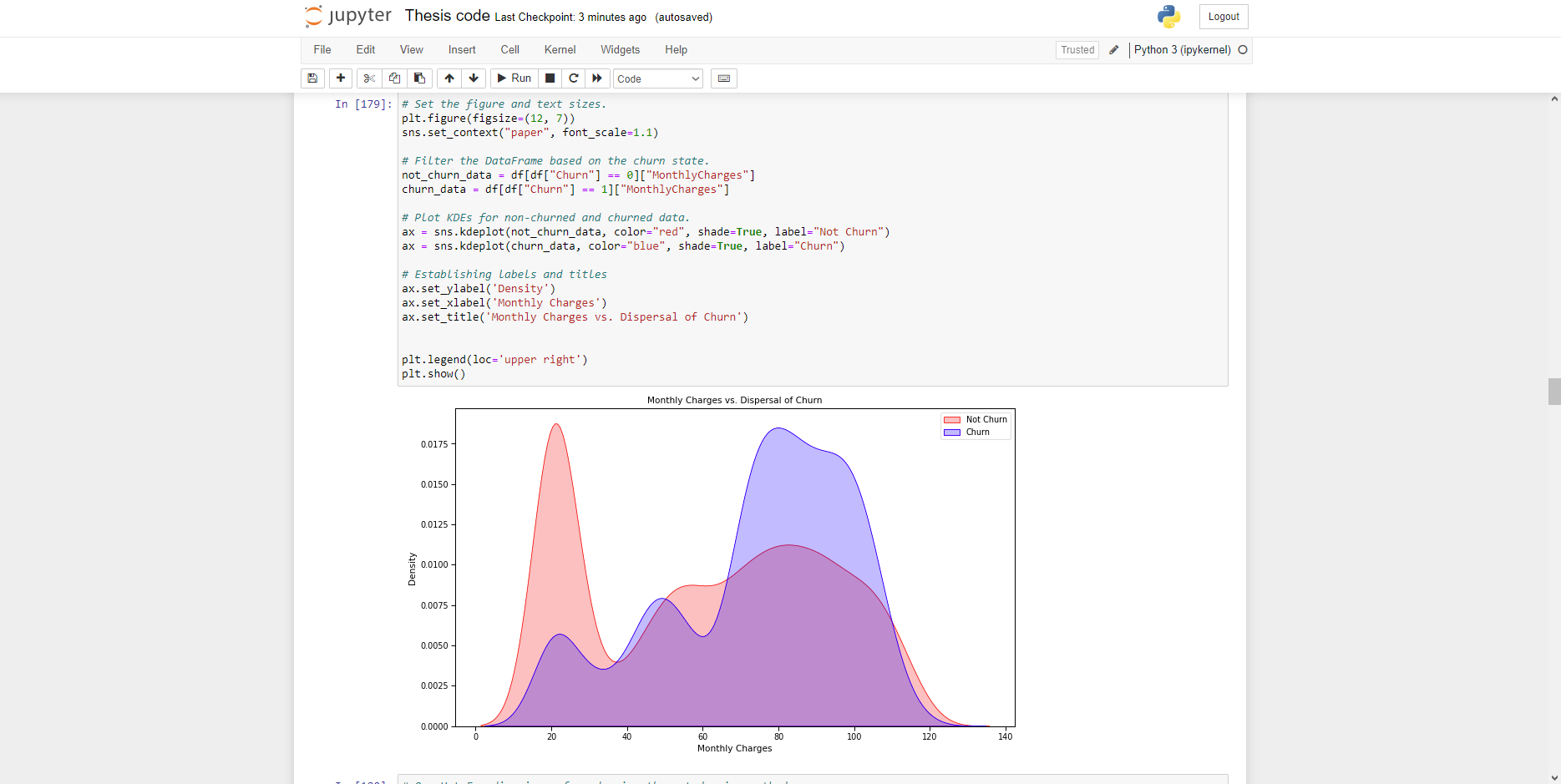
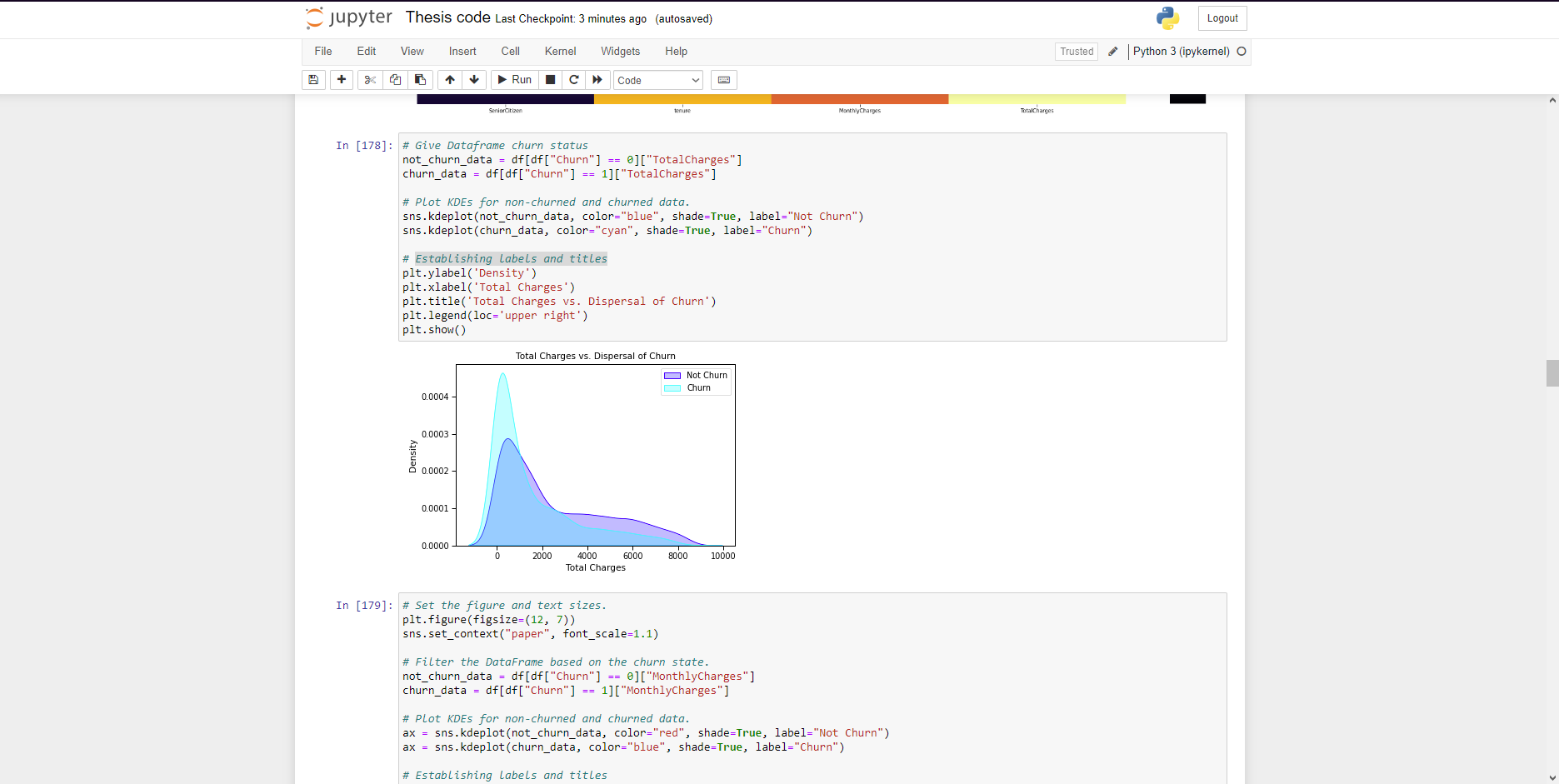
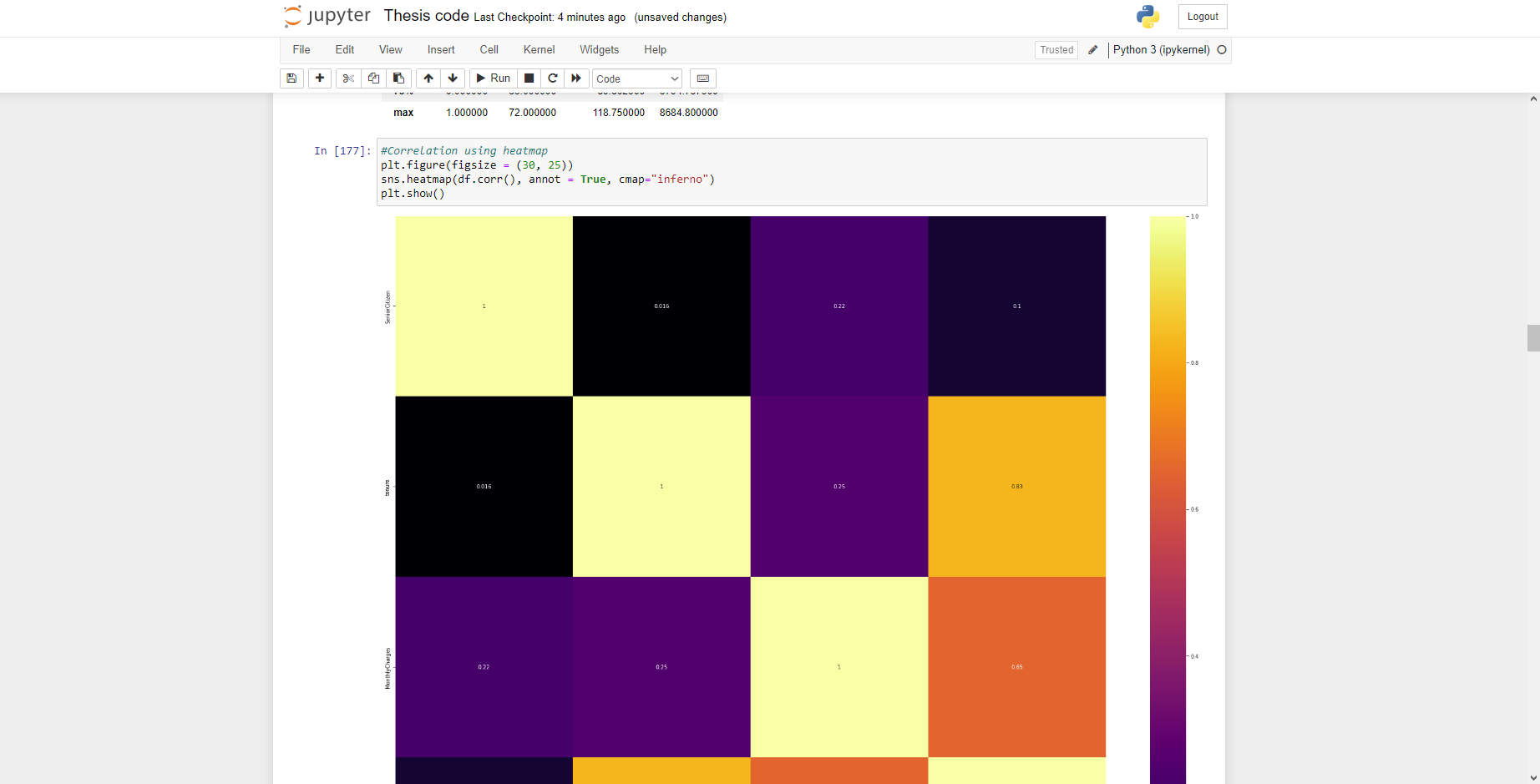
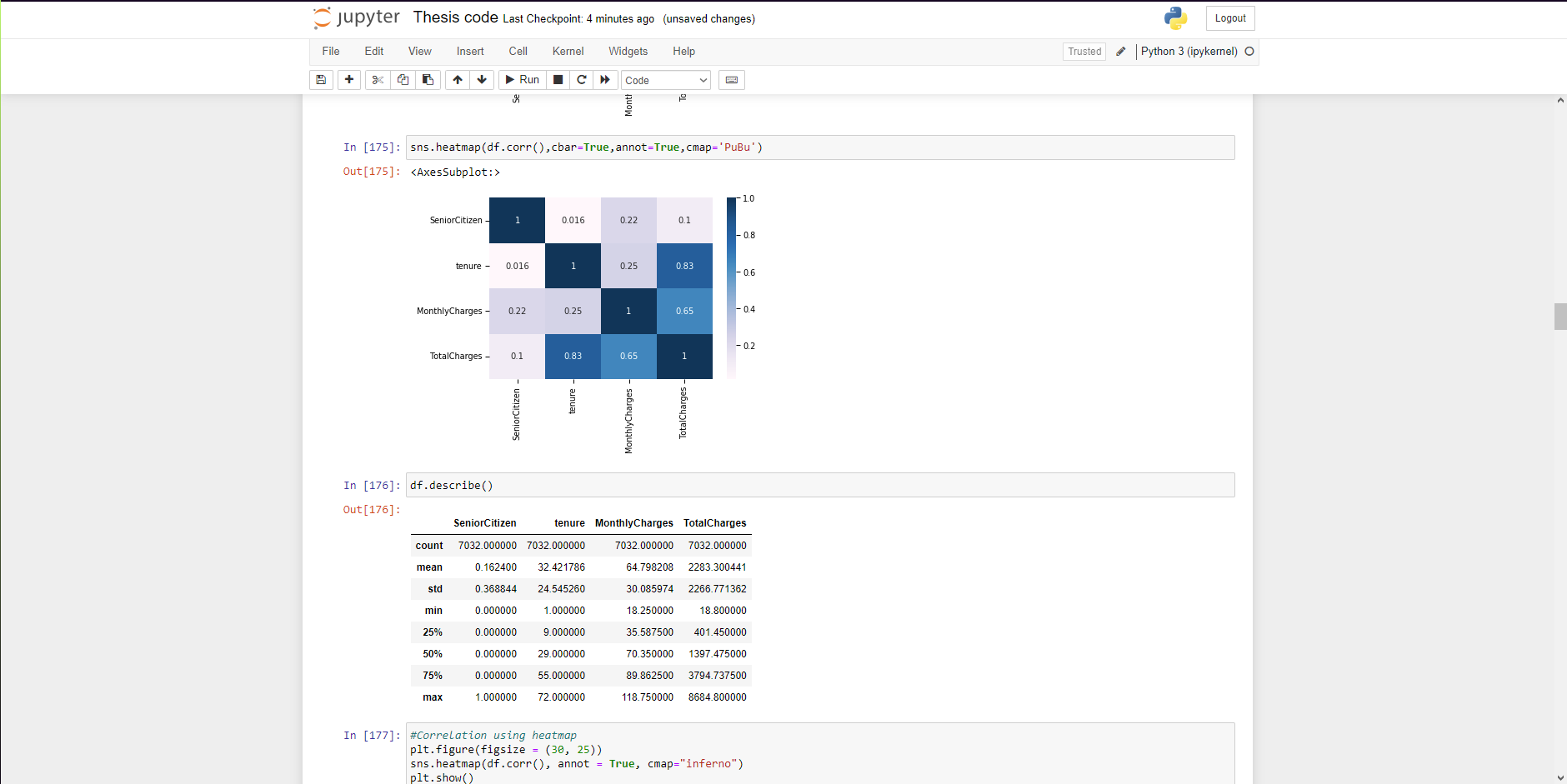
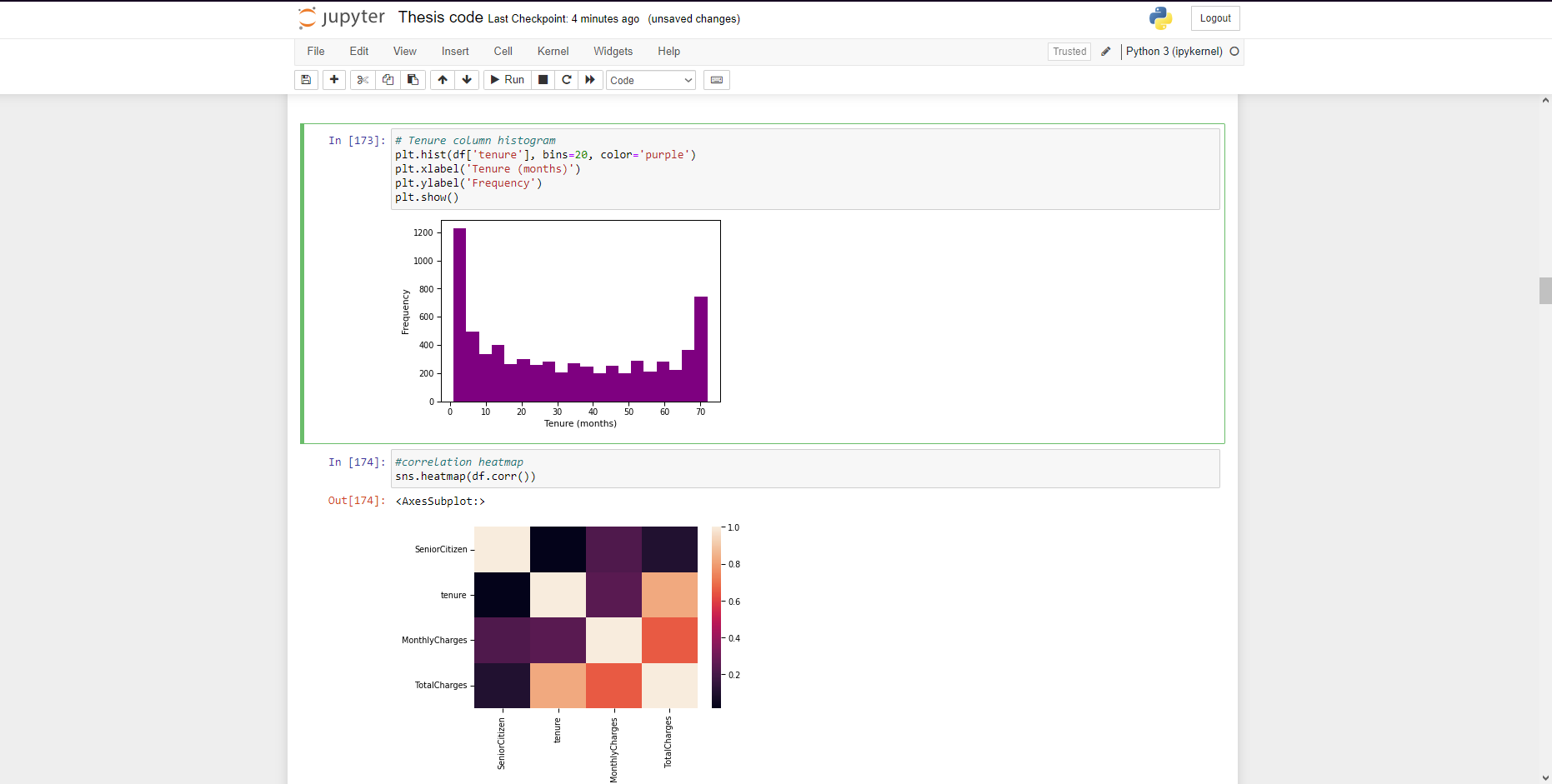
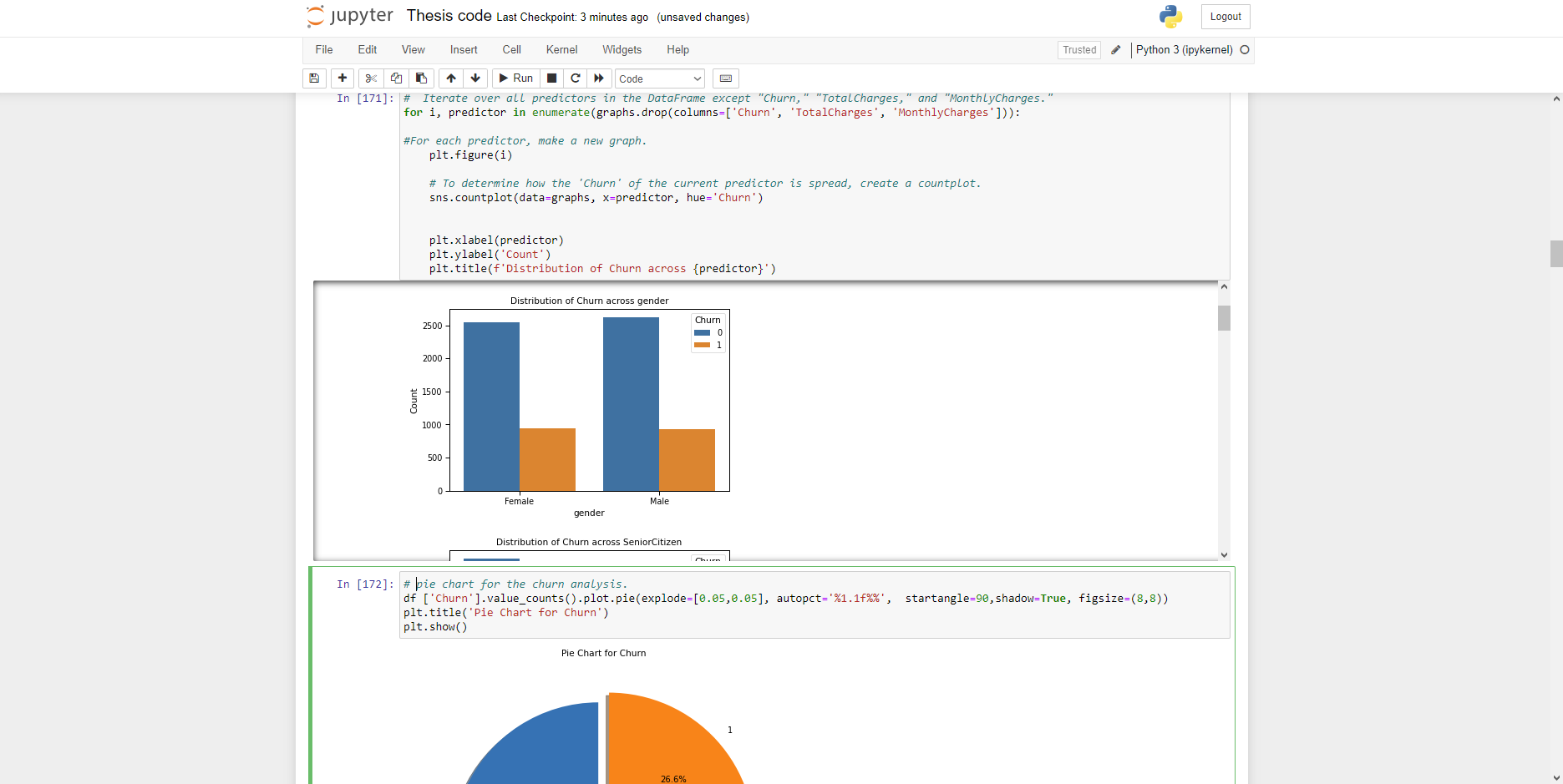
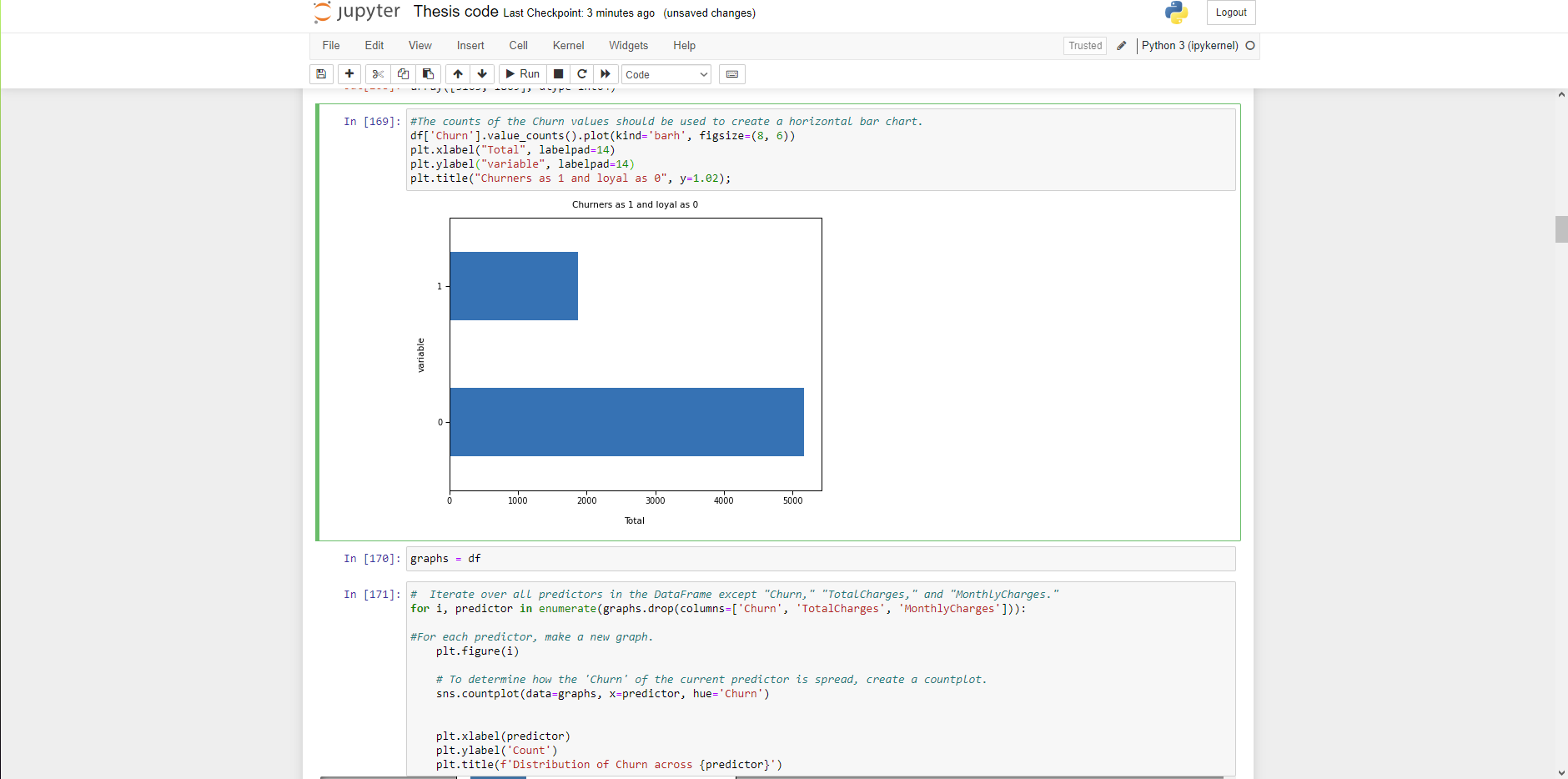
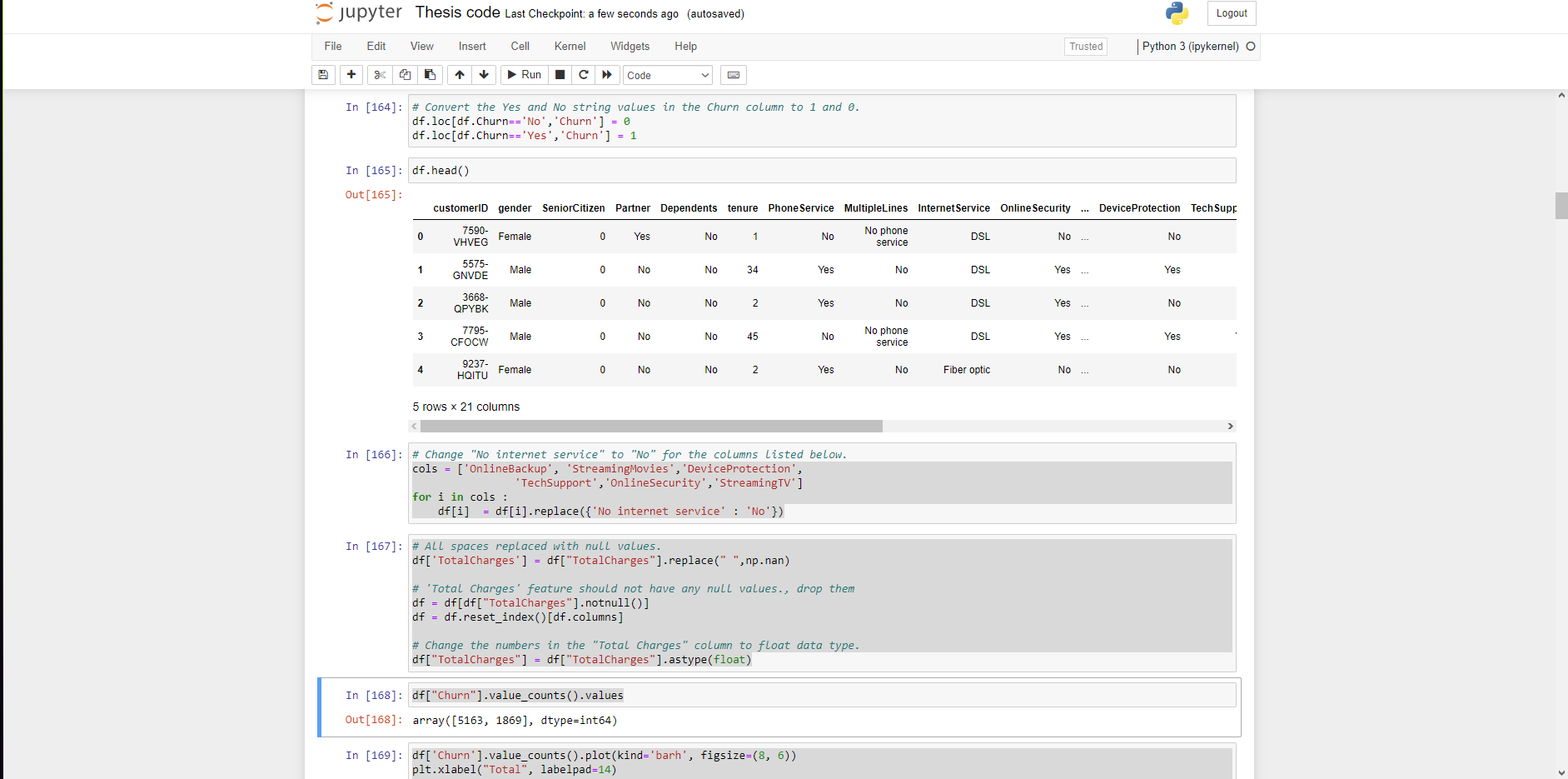
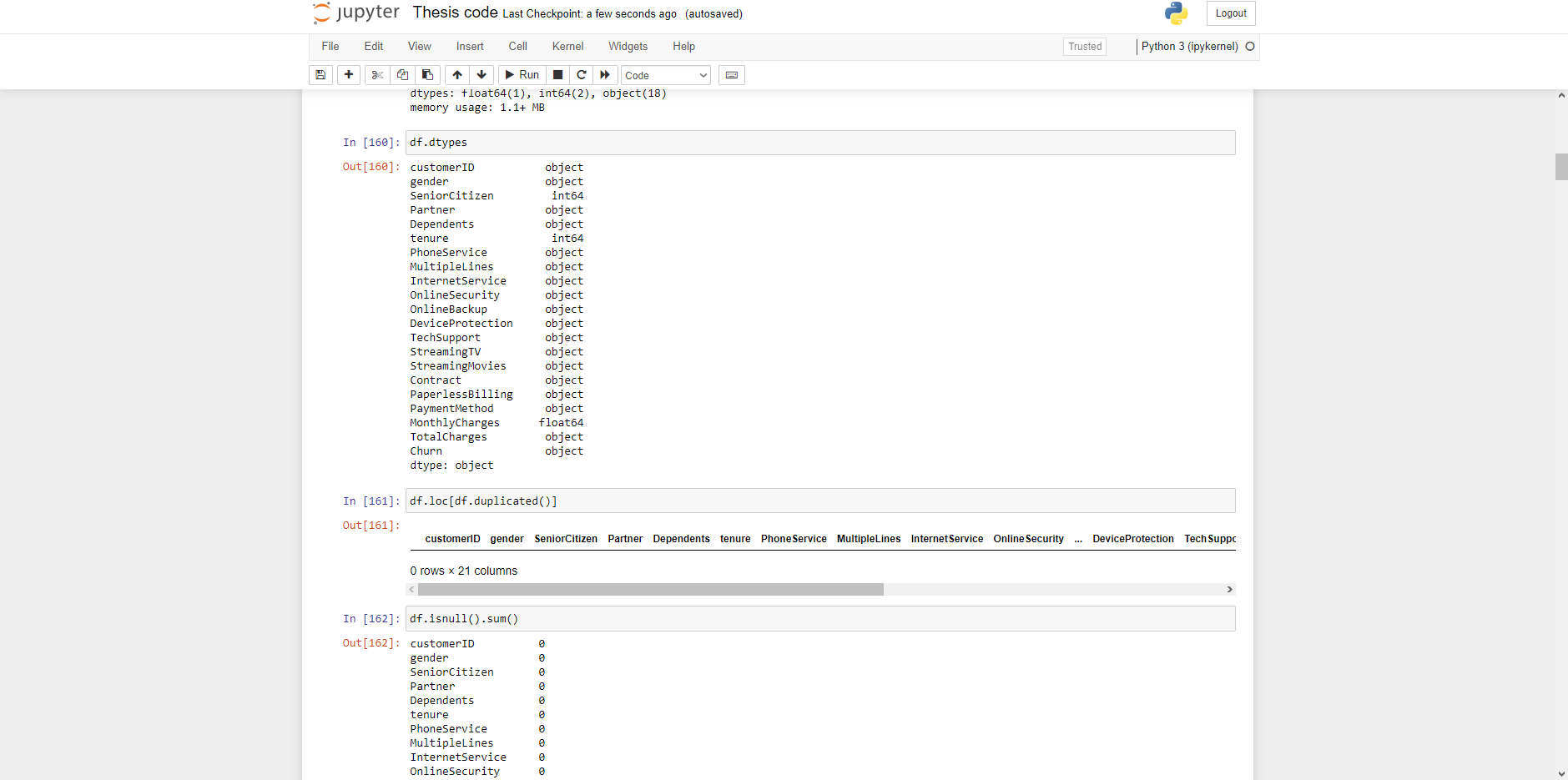
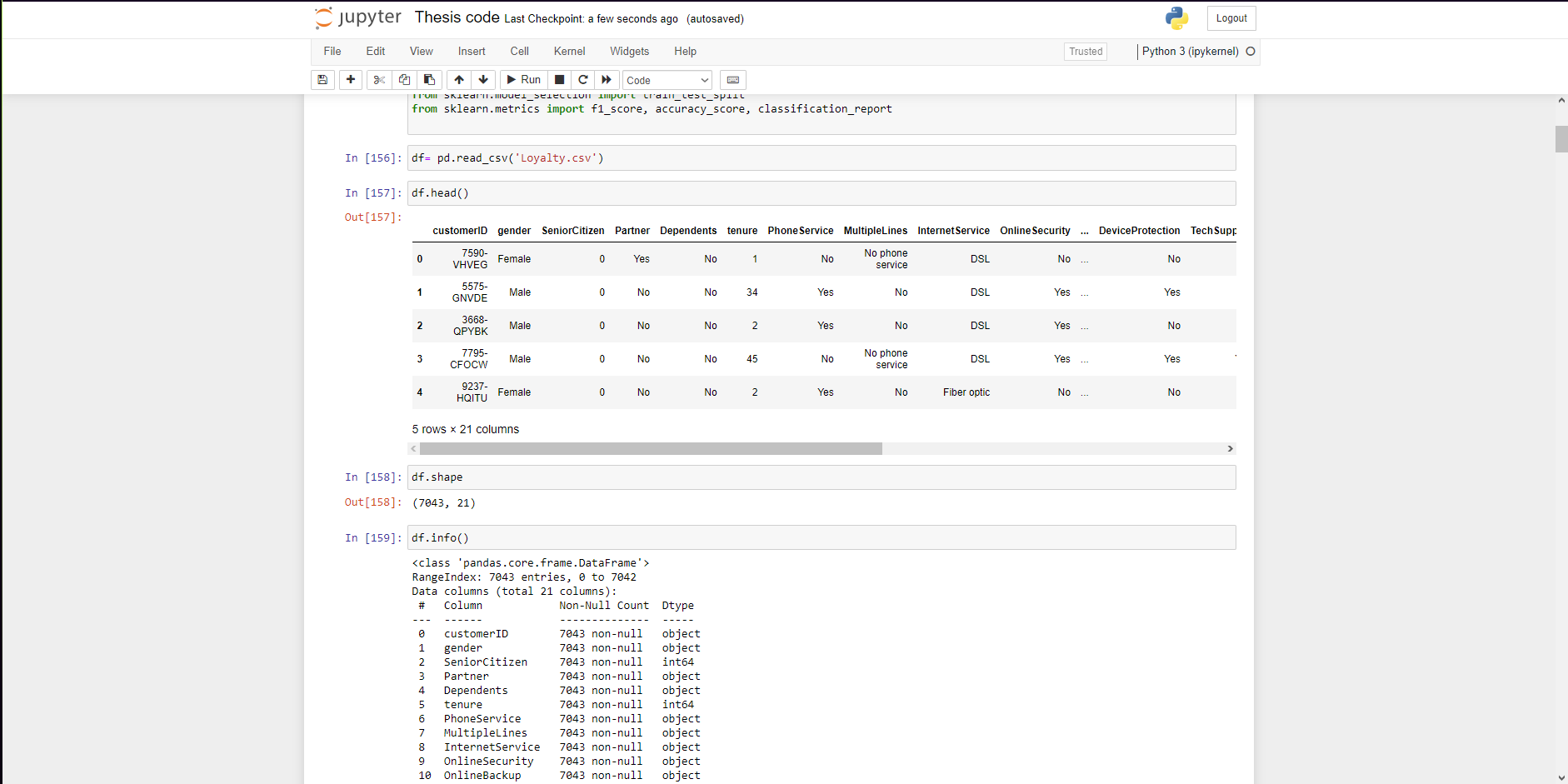
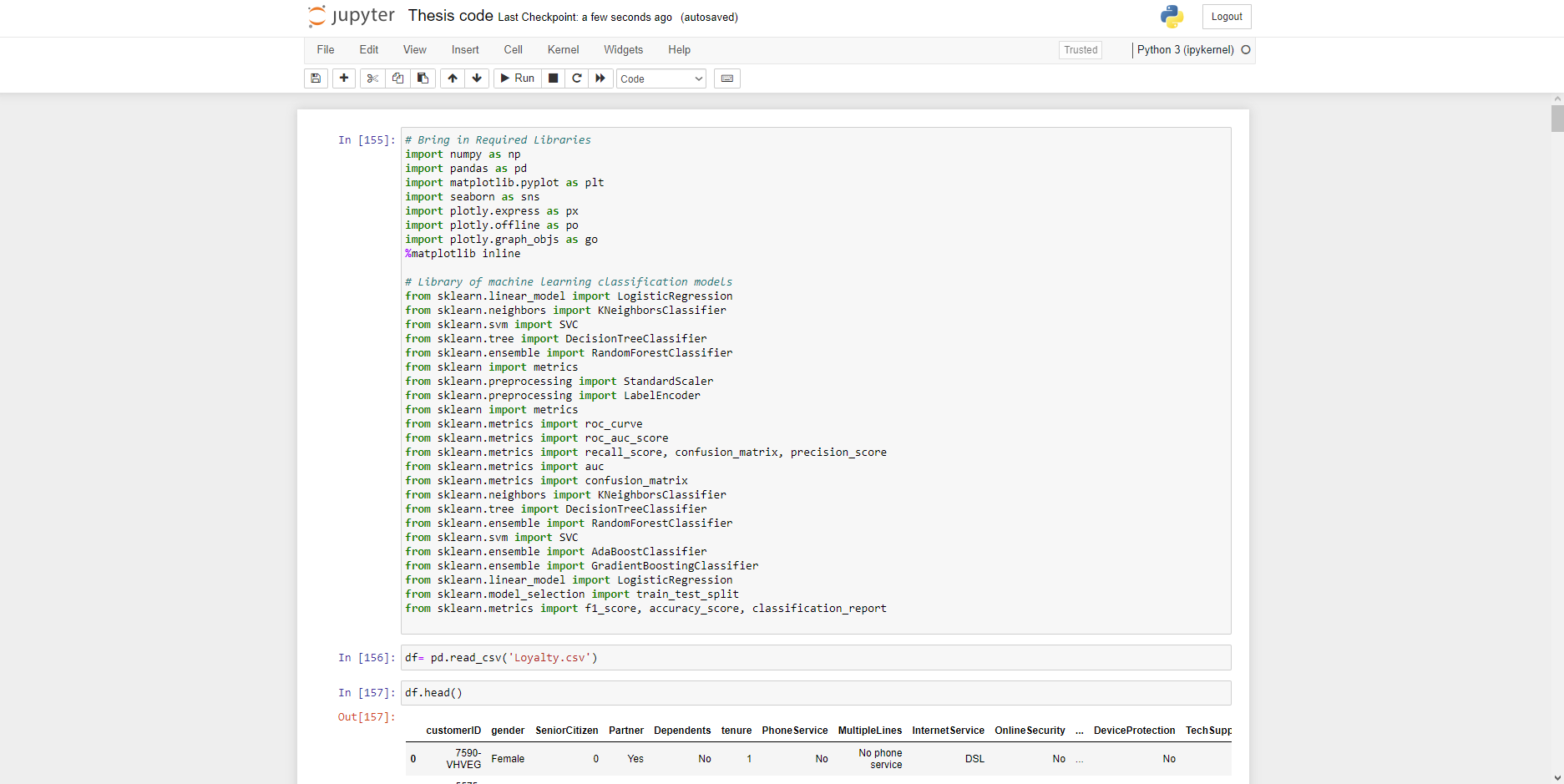
Xiahou, X. and Harada, Y. (2022) “*B2C e-commerce customer churn prediction based on K-means and SVM”,* MDPI. Available at: https://www.mdpi.com/0718-1876/17/2/24   
Accessed on: 6/5/2023

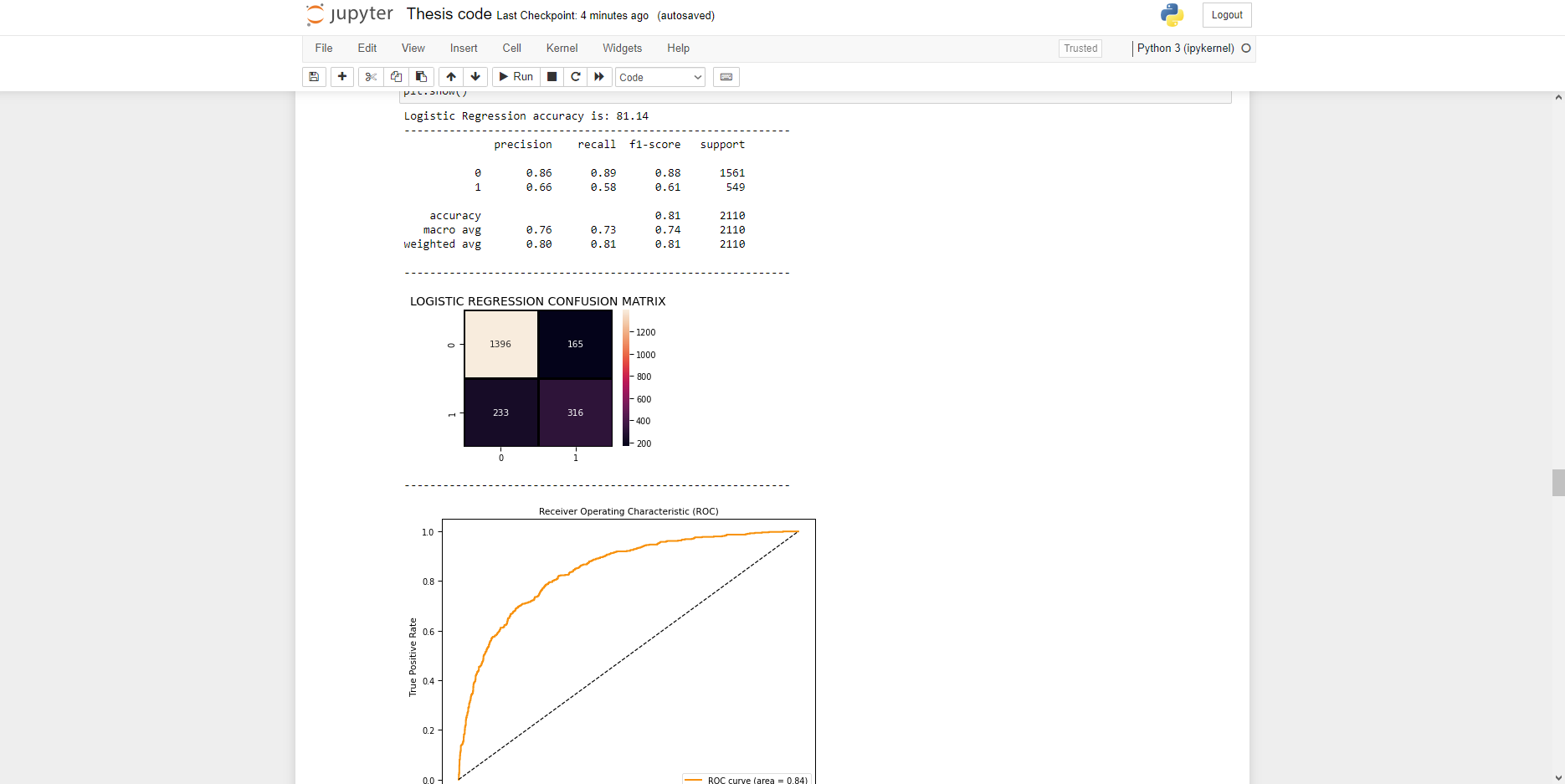
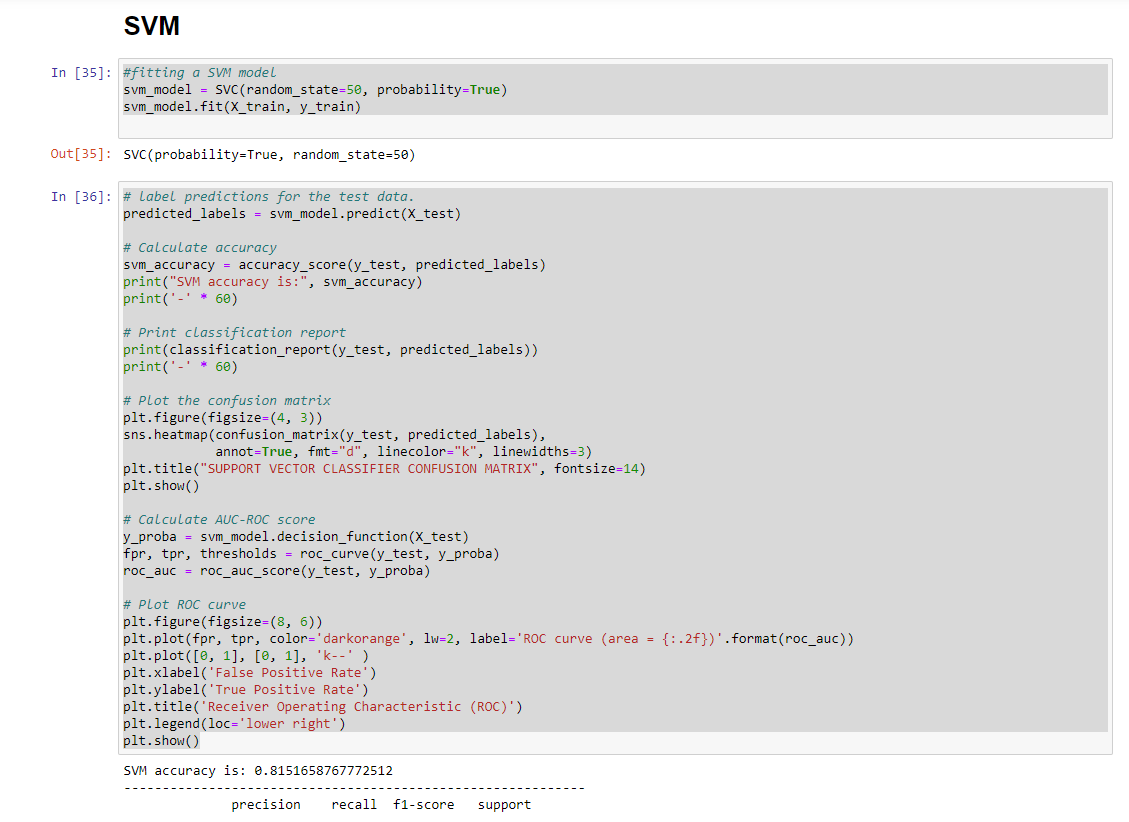
Yang,K. Kim ,J. and Yu,B. (2020) “*On Analyzing Churn Prediction in Mobile Games*”  
Accessed at: <https://arxiv.org/ftp/arxiv/papers/2104/2104.05554.pdf>  
Accessed on: 10/5/2023

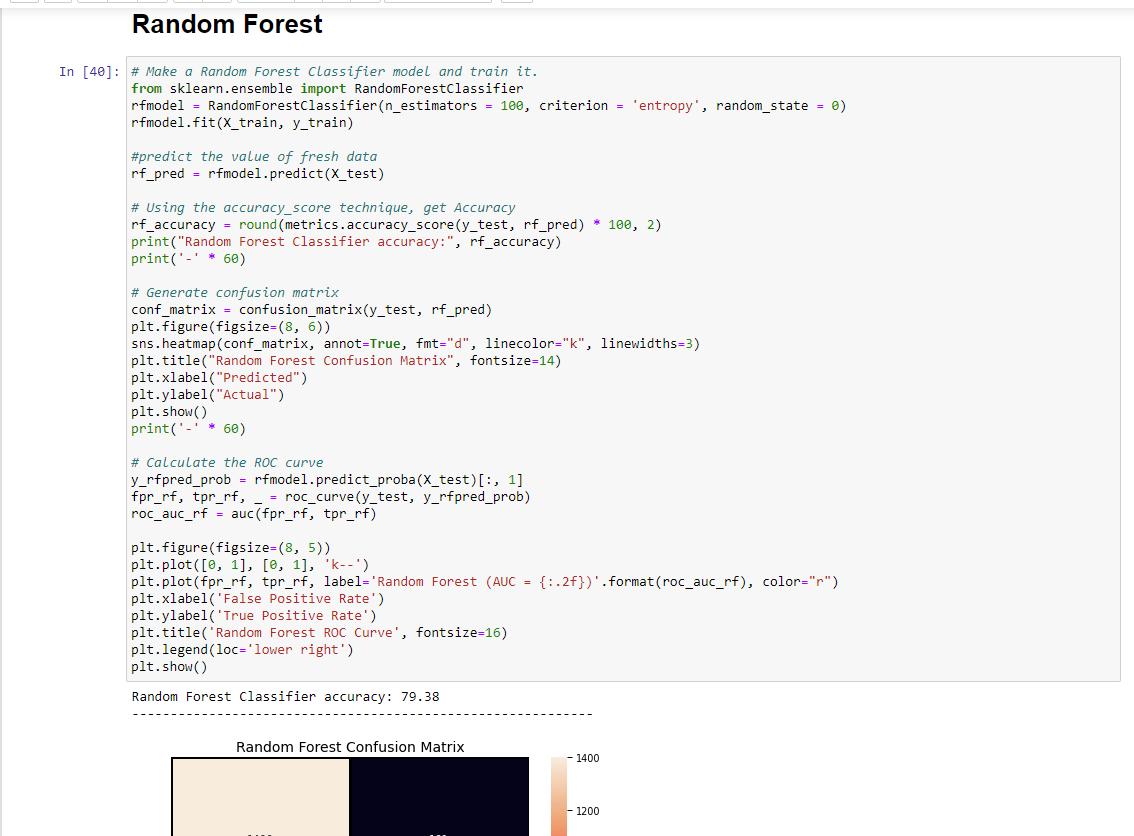
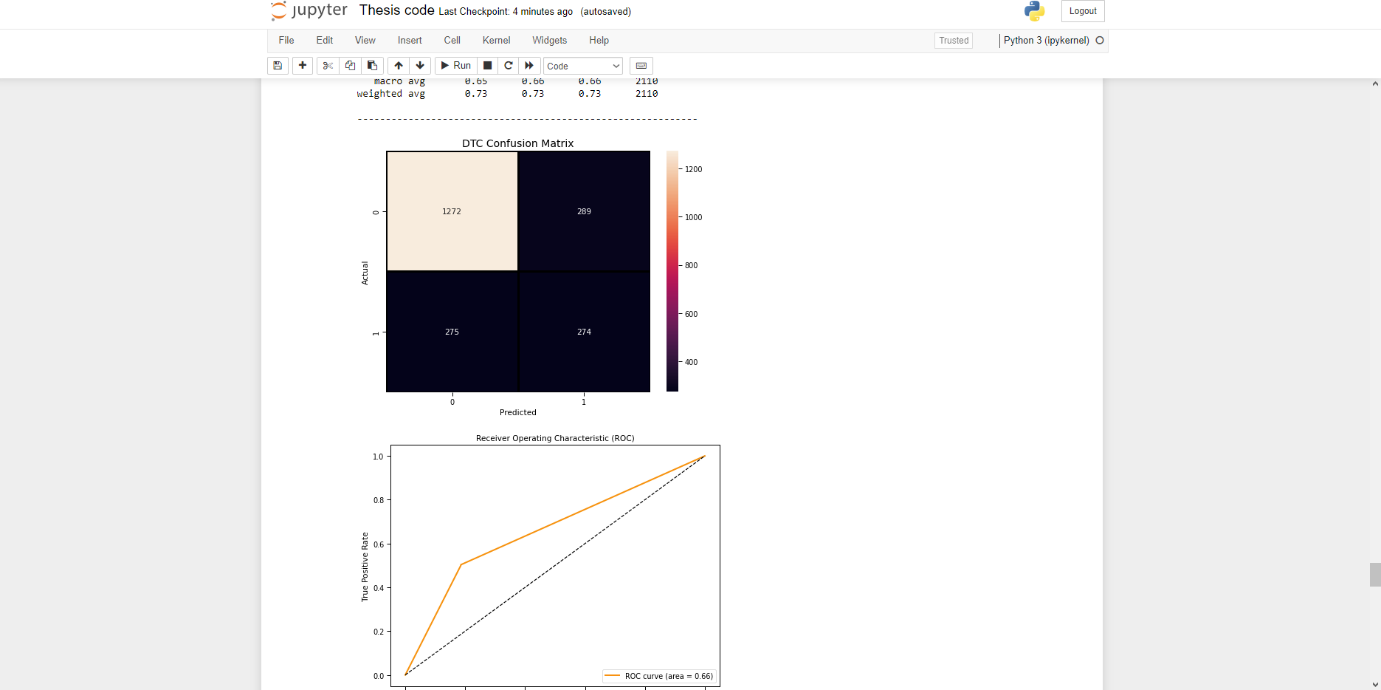
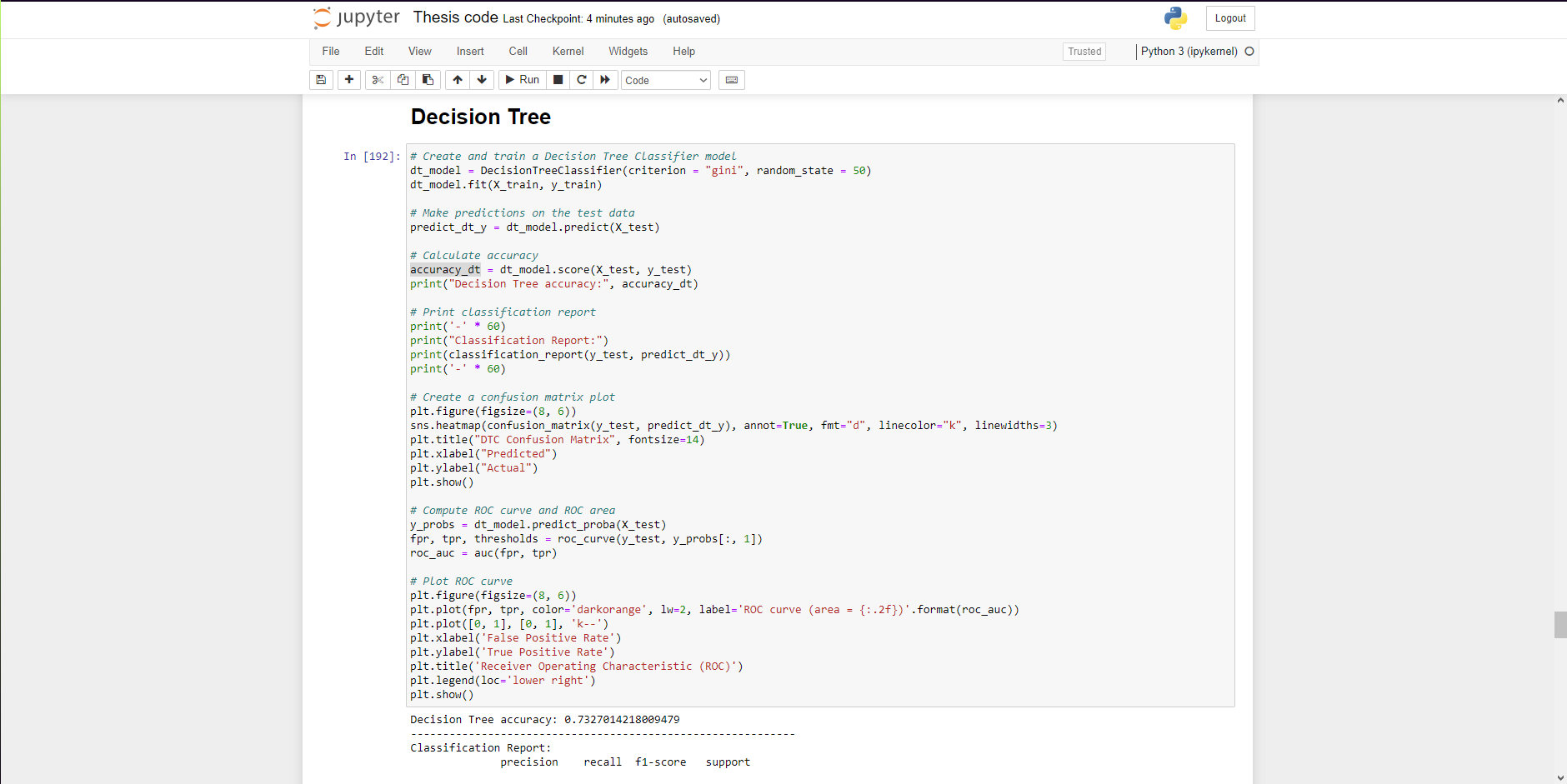
Zahran,H. (2022 ) ”*Graph-based knowledge modeling and analytics for capturing and prediction of customer behaviou*r”  
Accessed at: <https://curve.carleton.ca/system/files/etd/c55f1a9f-8dca-4b42-bcac-13e7a6f526d6/etd_pdf/b2999586fe931e47723ca99c7199bce9/zahran-graphbasedknowledgemodelingandanalyticsfor.pdf>  
Accessed on:9/5/2023

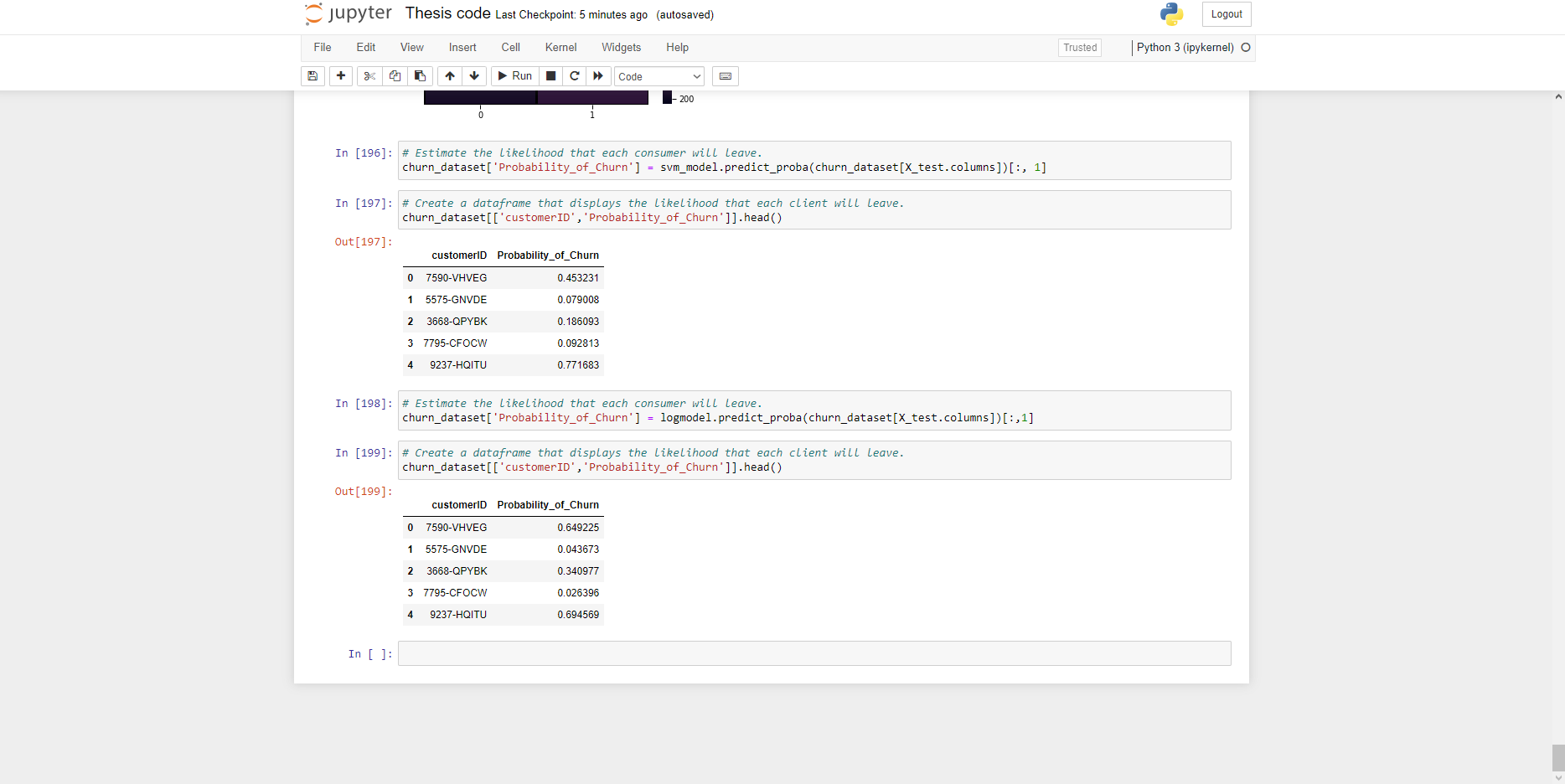
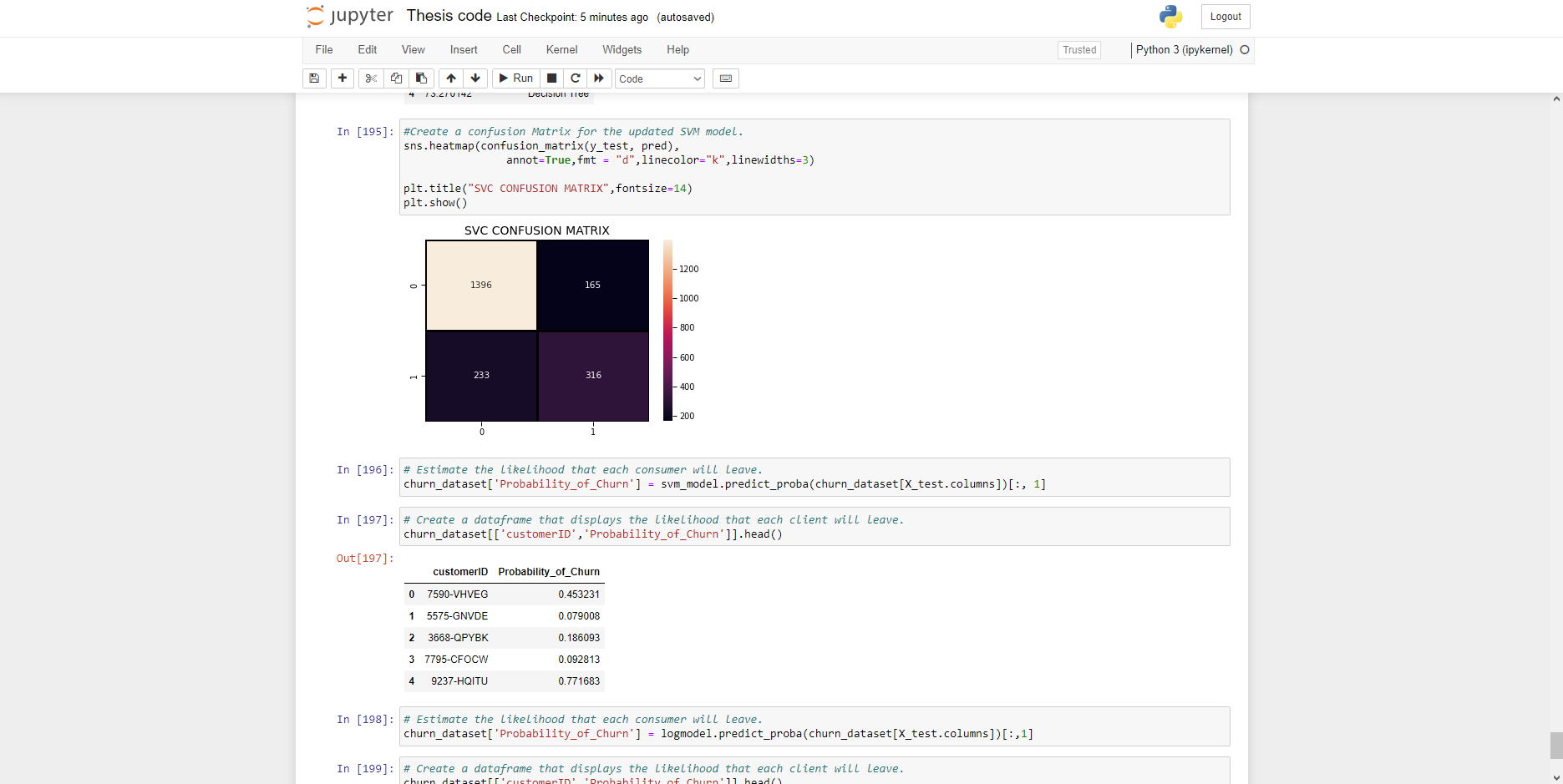
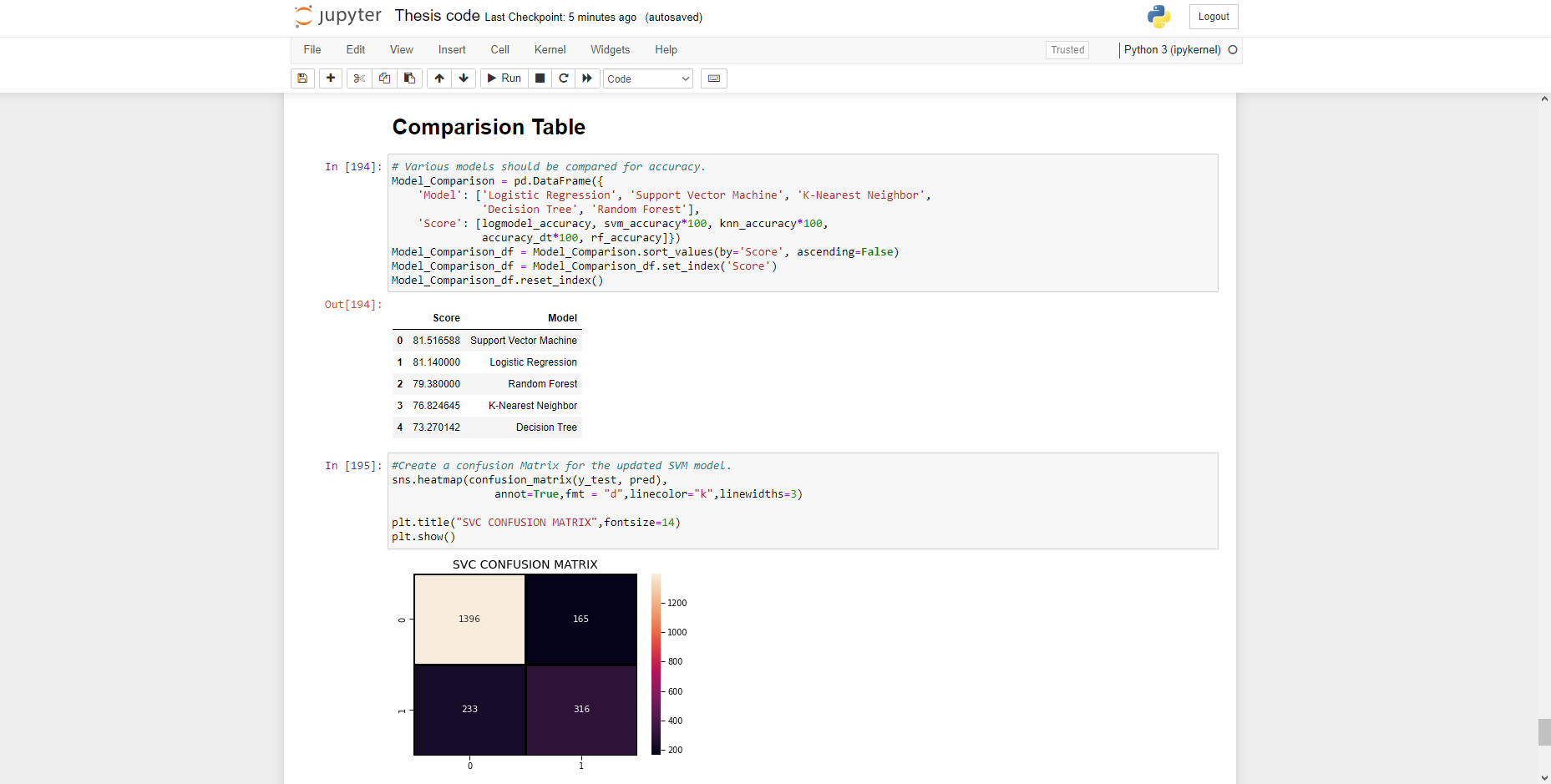
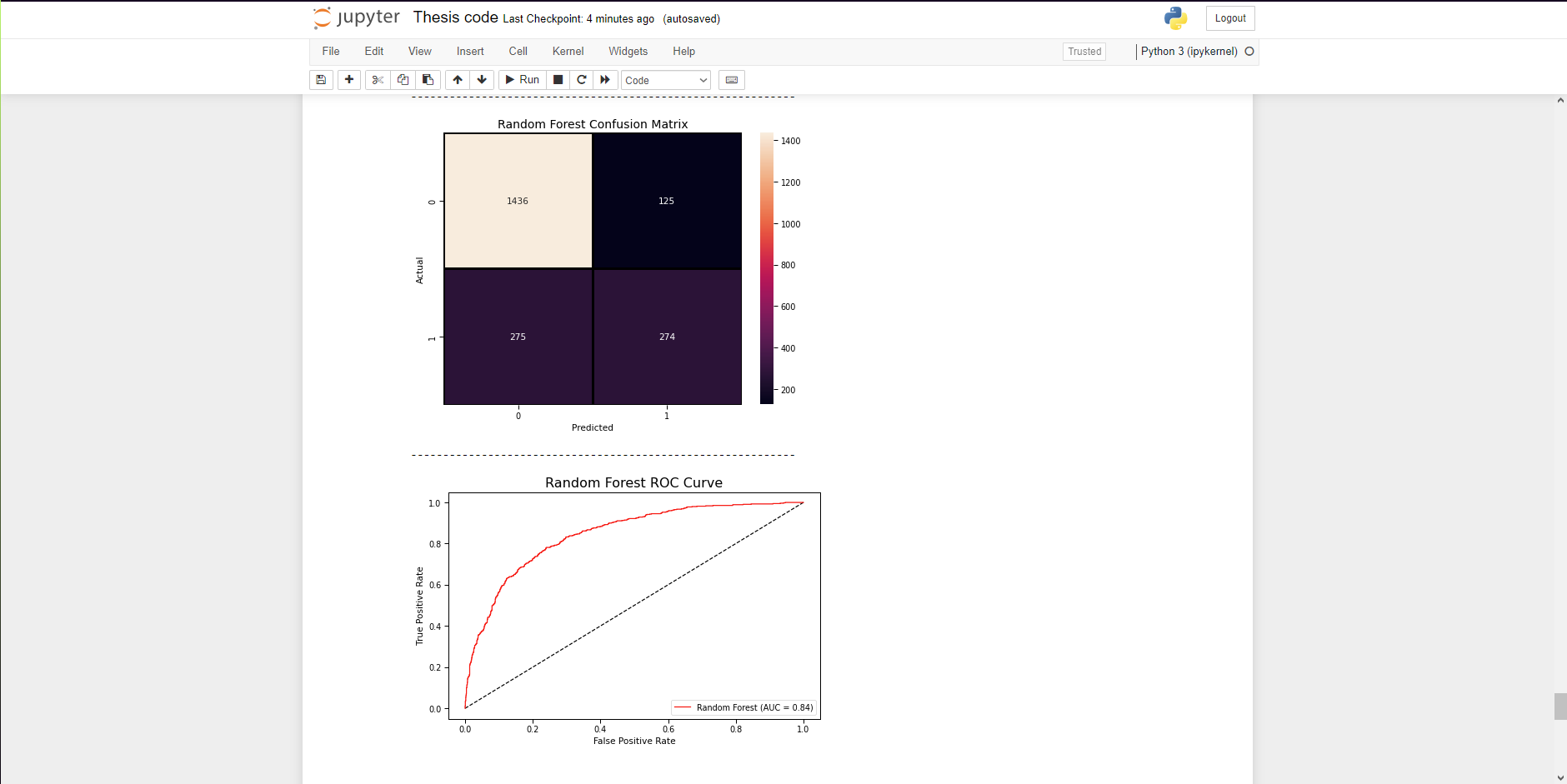
Zhao,Y.(2018) ”*Employee Turnover Prediction with Machine Learning: A Reliable Approach*”  
Accessed at: <https://www.researchgate.net/profile/Yue-Zhao-70/publication/328772915_Employee_Turnover_Prediction_with_Machine_Learning_A_Reliable_Approach/links/5dec49f34585159aa46ba991/Employee-Turnover-Prediction-with-Machine-Learning-A-Reliable-Approach.pdf>  
Accessed on:9/5/2023

**APPENDICES- A – Workflow**Authors Gihub Link: https://github.com/sba22421/ThesisSource Data set available from <https://github.com/orgs/IBM/repositories?q=churn&type=all&language=&sort=>







**APPENDICES- B - Interview Transcripts**For the sake of keeping previous places of work anonymous, when mentioned in the transcript the place of work has been replaced with “company” in all interviews. This has been implemented to ensure the privacy of places of work.  
  
**Interview-1**Interviewer  
To the very first question , is going to be, could you introduce yourself and your background in data analytics on CRM?

Speaker 1   
Yep. Cool. I have worked in data analytics, across supply chain CRM ,customer certification, which is it's kind of related CRM, and now customer support. So you have kind of all the user spectrum, because then in terms of the customer lifecycle,

Interviewer  
Very good. And with regards to data analytics, what drew you into working in the realm of data analytics?

Speaker 1   
So my background was in software engineering. But I didn't want to work with Java, essentially just working on giant mainframe problems. So instead, I took a couple years out, went into business. And I discovered that working with data was kind of an unavoidable part of it. And yeah, just gradually got more and more interested and find more more ways to apply it. And here we are.

Interviewer  
Now the next couple of questions are going to be specifically on the impact of CRM and data analytics. So the first question I'll put to you is, from your experience, what is the function of CRM in corporate operations.

Speaker 1   
So in my experience, CRM is essentially a way to pull people down the acquisition funnel, and keep them. So I mean, let's say you have someone who browses a web page, you have the cookie, you potentially have your email address dependent login, and OS, that gives you opportunity to engage with them, I'm putting down from acquisition to I forgot to bring down to actually make the purchase. And then once you have to make the purchase, then they become a customer, and then you try to get more value out of them. So essentially, having someone not know you, to get to know you to keep you. And then for you to become a brand ambassador, that's kind of the overall goal in my experience.

Interviewer  
Now, that's a that's a lovely way of describing the pipeline of the customer lifecycle. And so moving into that, how has the use of data analytics from your experience, impacted CRM practices.

Speaker 1   
So I worked in “company”, a couple of years. And “company”, it's obviously it's high end, luxury, retail. But a lot of the historical CRM, if you will, has been having people sending someone a letter saying, here's the discount, bring them into the store, and actually schmoozing them and talking to them for a while to get them to come in and spend time in the store. And then walking around actually, literally holding their hand helping to shop as a personal shopper, and learning more about them that way. Once you start to look at data, then you start to identify people who you may not have identified using clienteling or are using that kind of approach. So immediately broadens your scope in creating different segments that you can target segments differently. So we did a project where we looked at about 900,000 different customers over the course of 14 years. And we segmented them into about 14 different categories. We use those to drive marketing campaigns, we used to overlap that with website usage to present discounts or present quotes when people come on to the website. So yeah, essentially, it gives you much more opportunity to engage with a vast amount of people and a much more personalized and specific and direct way.

Interviewer  
Yep, lovely, your answer there actually went into the next question I have, which is providing a real world scenario in which you've seen analytics influence CRM operations and outcomes. And you mentioned how “company”, bring people in and be able to tailor their you know, their experience while they're there, Do you have another example of how Analytics has influenced the operation and the outcome?

Speaker 1   
Yeah, so you know what's Salesforce?, Our role was to try to figure out how to get people to recertify MBSR technologies, then, just pretty much doing pretty basic calculations on when certifications will lapse. And, you know, due to the volume of people who will be affected and when to be affected. We went back to both departments of the admin teams, the customer success teams, and back to the advocate, but the guys who went and created training plans for customers. And we gave them very specific dates about when these people are going to lapse. So even rudimentary calculations and looking at data in a different way, then people have previously looked at it. It can drive kind of interactions with multiple different teams and stakeholders, and have like real concrete effects on how the customers, first of all experienced the product, but also feel like they're valued, because you're telling them Oh, we know that we're, we're aware of what you're doing. We think we can add value here by pointing you in this direction. Here you go. Here's some suggestions on here's how you can leverage them to make your business better. So yeah, that's kind of another fairly recent example.

Interviewer  
That's perfect. I'm going to talk about the analytical side of say, like, the different software's that you have used. And on the CRM section with the next question, what have you seen, through your experience, the most prevalent analytical approaches, or procedures in the CRM setting?

Speaker 1   
In what sense, so you're doing like, which tools people use the most or kind of

Interviewer  
Yes, if you can talk about the tools, or if you would, like if you'd like to talk about, say, if one case, if it was predictive modelling, if you had use any sentiment analysis, to obtain customer insight in the CRM sphere?

Speaker 1   
Well, okay, so sounds good analysis, we kind of use it right now, my current position. So in “Company”, we've launched a new bot called ‘fin’. It's like the, it's probably the most advanced support tool known to man, just. So yeah, we use GPT for a lot. So what we do is we pull out the ticket information, and we pipe it through GPT. And then it helps us to summarize conversations with people. And we can really drill down into figuring out where people's issues are, where their pain points are. For example, we had a spike in volume over the last few weeks, we did this exercise that we pulled out that a lot of people are coming in through a specific kind of workflow them that we use kind of automations. And we figured out that we can actually improve their experience by giving them a slightly different kind of outcome. So let's say they come in and they have an issue with how the semester separate website, instead of saying, oh, here comes you go into a general inbox, we can ask them about things to more different questions to figure out where not to do with latency or where it's due with them setting up some of the settings in indirect channel messenger. And then it gets them to specific person, I guess the response faster, I guess, logic conversation faster. So what's one example using some kind of sentiment analysis now, in general, CRM, at least in the roles I have been in has been less concerned with using fancy tech, I'm more concerned with using effective tech. So a lot of AV testing a lot of kind of small scale, kind of iterative experimentation. We just brought up Thomas we did a lot of work with website, just doing T testing to figure out which groups react better to eat or getting through shopping baskets, or how to day we react to carousels and on the website, Premier Clinton products. Same deal with actually working with in store customers looking at us which customers have lapsed or what their timeframe has been since last purchase. And then again, just setting up experiments to say what works better, giving them a discount given them a cheap pre product, giving them a free product, giving them a buyback. So when they come back next time to get discount, but it's all hypothesis testing and experimentation. Fancy tech. I mean you can use it for segmentation. We did segmentation model was built in our that was kind of it was a big project. It was gross predictive modelling. Not a whole lot really, I mean a lot of work with association rules, so building out kind of typical products that you would expect people to use, but that was used to inform experimentation more so then, kind of driving anything into website, I think we use Salesforce a lot. And a lot, the Einstein analytics tools are built in there which do a lot of primary product recommendation. So Salesforce, it's a big, clunky, horrible thing. But it's quite powerful in terms of actual day to day leveraging of analytics that is more or less just testing.

Interviewer  
Perfect, and from all those different varied technologies and software such as ‘Fin’, as you mentioned, or sales force and your new bot that's going to take over the world. Have you discovered? Hopefully, we're using all these have you discovered any issues or limits when using these analytical methods on data?

Speaker 1   
Probably the most salient is in trying to determine going to churn rates trying to figure out a way to predict who's going to leave you. A lot of I don't accept marketers, but a lot of people that I've worked with, they've been on the assumption that we can use the data we have in our little silo to predict where people leave, but you're a little square, or you're a couple of tables in a database only to help you so much. And customers have a lot broader scope than just your use your your platform, for example, we're trying to figure out, does the amount of support tickets someone opens? does that relate to how they churn? And you know, I mean, you can make the argument, yeah, I mean, if you're happy with the product, you'll open less tickets, or if you're engaged product more, but at the end of the day, if a company is making money doesn't matter how many sports tickets to open or close your account. So there are a lot more kind of macro effects that you need to take into account. Maybe this departure factor analysis are going to be a bit deeper with customer interviews. But quantitative data is not enough to really understand the full picture of what a customer's situation is, that's probably been the most probably the hardest one to explain to people as well, because everyone thinks their data is gold, and they want to really leverage it. And I mean, for the most part, you can, but you need to talk to people as well.

Interviewer  
100% true my use of the CRM systems when I was working with “company”, I was contacting in one instance, 400 people in one week. And what I found a lot of the time was, while the, the CRM system that “company”, had created was basically just going by what their previous car model was, through talking to a lot of people, there was information that wasn't in the database, from the say, if they changed area, do they need to drive more, drive less if they've increased family, you know, if they have kids come on, along the way, they were taken about all those different parts that weren't actually down on paper. So, 100% agree with everything that you've said there.

Speaker 1   
So, that that's a part of the segmentation process. So if you have clever modelling, you can figure out which kind of people are most similar to norm and which ones are most different than then they can have, you can zone in on them for more personal touch.

Interviewer  
I'm just going to talk about customer retention strategies now for the next couple of questions. And from your experience, how have you seen if so, how good data analysis helps design and strengthen client retention strategies.

Speaker 1   
So not sure if you have been involved with like email marketing before, but there's kind of a general expression, which is spray and pray. So you take your whole database and you send an email to everyone. And you see what sticks. It works sometimes. It's difficult to get analytics because everything's so general, it's hard to drill down and find out who worked best for. So probably the most effective way has been in more specific segmentation. And in when you're designing kind of whatever campaign you're going to use, like some kind of activation campaign. You don't look at everyone what you do is you look at people who are either buying a different brand you want to reactivate into a brand, people who were buying like to say you're buying Foundation, what is the average? short distance between purchase a foundation. You can also support this by quality by go and talk to people in the store to find out what did how long it is, but how long a bottle of financial would last and use that designing but the segmentation and avoiding Senator everyone is the most effective way to retain people because you make it specific to them. You're not just going to spam your email address or their inbox.

Interviewer  
Yep. during my time working at ‘company’ , the good portion of the casino we found ended up not responding to emails, and were actually more effective if we sent out letters. And we found that they much prefer that to we would dare within the letter, we would have a promotional piece. And they would be more susceptible and open to returning via that method, then an email.

Speaker 1   
broke down was your high value people, you give them something they can hold in your hand, and they come back. Great.

Interviewer  
next one is can you provide instances of effective customer retention that was established or improved? Using data analytic insights?

Speaker 1   
Yeah, so go back to the example of from “company” do is in that case, we looked at like I look at works your person A, you certify as developer for “company”. Generally, if you certify, as a developer, we did about two years, you let it lapse and you look back because you only want certification that then you get your job, you don't care anymore. What we noticed was that for partners like your “company” utilize that kind of stuff, they were letting people lose to certifications, which is not ideal, because we want to make sure make sure that we have high certification volumes. So we can prove usage and kind of show that we're the most using the second heard of the vertical. By looking at the amount of time people had certified during engagement with self paced learning, with classroom learning, we were able to go to these things, run four or 5000 different accounts and say, Come here, we want you to recertify here this time right now. So working with again, work cross functionally, because they're the best projects, we worked with partner “company”, to work in sales and the portfolio team to extend these people's certification limits manually by about eight months. And then in doing that, we all signed them up to use free training tokens. So we gave the tokens, but then they actually went back to learning and they booked two certifications, I don't know how it ended up because I left after that period of time. But by all accounts, in the time, where we gave tokens, the uptake was, like 60- 70%, because people were again, it was getting back to the personal touch. So when coming to you and saying, Here you go, here's an opportunity for you. We know you pay us a lot of money, we want to make sure you get value. So really delivering the message that to people by using data, very specific data on who was losing it and where and divide that we can drive for that customer by having the certification command. So I think it's, it's kind of used it to deliver the story to people.

Interviewer  
As you say, the more personalized touch leads into the next question and how we kind of have to tread lightly upon that. In regards to ethics, from your, from your experience, what are some ethical concerns to consider when using customer data and retention strategies?

Speaker 1   
It's interesting question. I'm not sure if I classed as ethical, but I'll let you kind of make that determination because maybe two examples one was most of my CRM comes from our domicile, I can see it from there. First one is. So “company”, it's this customer base is kind of has a wide demographic. And what we've tended to happen, at least with mailing campaigns and going to form campaigns is that people who would have been previous customers may have passed in time sensitive last been contacted. So again, once GDPR came in, and people started realizing this data shouldn't be there, there was a lot of issues, whilst a lot of issues, there was numerous issues where people were being contacted. And do his emails and letters come and say in “company”, combine this and that person, sadly, no longer with us. And it led to kind of stress for families. So in that case, the comments, we're going to need to do it, do it, do it, delete it, or we're going to purged. But in that case, it's kind of difficult to automatically know this. So be aware that this is a factor as part of any retention campaigns you're running. And making sure that when you do get any feedback that it is taken on these accounts are destroyed. That's obviously on top of any kind of standard procedures you have for keeping data. Last thing we kept nothing less than three years, I believe. Or at least anything beyond three years, we anonymized, if I remember correctly, but anyway, that's one the other one was that, again, “company”, luxury, high end retail. And some of the customers spend outrageous amounts of money, like numerous years of my salary per year. How would that information I've been able to send that around and having people access to that information. Or people who are very much in the public sphere, like certain MMA people, or people from TV, having that information behind closed doors, or only like, X amount of people can access that with certain privileges is very important. Like, we could go in and see anyone's personal home address, because we are working to data, people in the store could not, which is important because you never know who to come in and say, Okay, here's 10 grand give me X person's email or actual home address, so I can go and find them. So yeah, more, maybe more probably long blend security, but obviously has some ethical ramifications also.

Interviewer  
100%. I touched on that a couple of times in my workings, is the fact that who has access to information is very much an ethical consideration for departments, as you say, you can't just have the people in the store have the information and then someone comes in off the street after seeing such applicant go in and say Here I'll pay it to find out like where it is because that could handle to burglaries and all different sorts of issues. So for the last couple of questions, we're going to be focusing more on kind of future or trends, weight analytics. So I'm going to start off by asking you, How do you envision the relationship between the analytics and customer retention changing in the next couple of years?

Speaker 1   
What I'm seeing, at least from our point of view, in my current position is how easy it is to get information about people. Does that tie into other information like not sure how familiar you are with like Google Analytics or like how website links works, but in some cases If someone comes in, and they start a ticket, for example, but you're logged into different accounts, we can then link that account to someone. Once we did, the accounts are linked, we potentially then go on buy information and go wider. So what I see coming is lots more data sources and sources of information. I mean, it will tie people back from work accounts to social media accounts to let's say, you have your “company”, has agreements with lots of different companies that they have CRM data, they can pull it in, they give you awards based on working here, they're everywhere, our test code where they work with certain or they work with different companies that you can earn points. Bigger, bigger, wider data, I think is going to be where everything comes and ridiculously specific recommendations. It's already here. But I believe it's going to get more and more ridiculous, like I've read recently that Google now has records of where advertisements are placed in roadsides, them, like stands, they can track where you have been, where you're driven, and they can surface knowledge relating to that letter to you today. So just the fact that advertising is now is never off, you're on its Minority Report, you're going to be the person to wherever you go, you're going to see personalized advertisements for you. That's what I see happening. How do I achieve it? Probably some, like step forward and transformer technology are making neural networks a lot more efficient, or do something with like, really clever embeddings in super large language model type things. But that's ultimately, I think, where it's going to go. Yeah.

Interviewer  
that's very interesting. I hadn't heard about the the billboard one. But as you say, if they're being able to track, say, from one location to another, which is probably linked with, say, you use Google Maps, and you go on a journey, and they're able to see you travelled from certain points, but they'd be able to tailor different ads, it's, it's definitely going to be an interesting time. In regards to that, and you're kind of touching on a near the end of your answer there. What advice would you provide to businesses, looking to use analytics to improve their client retention efforts?

Speaker 1   
In my experience, people tend to go big, they go big, because the ones kind of do one on one hand, and they wanted to everything worked perfectly. My experience doesn't work. Everything has to be tailored, smaller, specific. So yeah, small scale experimentation that can be generalized. But starting small and building upwards, don't start from the very, very top and try and go wide straightaway. Because it, it doesn't work. And it makes people unhappy, at least in my experience.

Interviewer  
That goes back into the whole implementing it. But you know, the specifics of the CRM to each individual, it's a tailored effort for their experience, instead of as you say, the just the wider scale of a will put this plan in operation and hopefully catches many.

Speaker 1   
Yeah, I mean, it goes when you have to assume that the company is smaller. I mean, it depends on what segment you're in. If you're Google, you're not going to do small, you're going to go broad, you're going to a big, if you're a mom and pop restaurants down in the side of the road, and you just happen to have like a free instance of HubSpot. Anyway, a little bit of CRM. Keep a tailor to keep specific

Interviewer  
100%. Last thing on the future trends, What recommendations would you give to businesses wishing to implement more data driven operations?

Speaker 1   
I would say invest in people. technology comes and goes, but people come and go as well. But what I have learned over the last couple of years is that in learning technology, it's essentially trivial. I mean, anyone can learn to use spreadsheets, anyone can learn to do a bit of modelling with RMD. You have a data frame and you run an algorithm overnight to give it something close to the truth. The difficulty is getting people to tie the analytics to business outcome, trying to figure out what this work is actually going to drive in decision making and what problem you're trying to solve with it. You can do cool things with data, but you can also spend a lot of time wasted just chasing that rabbit holes. Be specific about what you want to achieve, and be able to tie it to what is going to happen because of it. That'd be probably the best piece of advice that I can give.

Interviewer  
Lovely. And to finish off, finally, is there anything further, you'd like to say or discuss about the influence of data analytics on customer retention?

Speaker 1   
I probably say that, that analytics is it's a tool. I mean, ultimately, it comes down to I mean, I, I have a personal pet peeve when people say I want to be data driven. I don't like to say their job is to say is be data informed. Ultimately, people are going to use your intuition at various different levels, given them the data to kind of inform and give the extra nuance and subtlety to that intuition is what really helps. So yeah, I mean, ultimately got to get them out to people and data analytics is a tool so use it carefully.

**Interview-2**

Interviewer   
Perfect. Thank you for joining today. We'll start with a couple questions on the introduction and background. So for First off, could you please introduce yourself your background, your background in data analytics on CRM?

Speaker 2   
Yeah, so i've been working with CRM for about 10 years and I on that involved, analytics, both data and business. And I started off in California and the wineries and that was looking at financial analytics, tax information, crush reports. And so it was a lot of yield and operational data that I would have been processing, then I moved back home to go kind of sabbatical off and went into the hospitality industry. And it was just one of those jobs to, you know, not have to think about. Then I came back into the, the industry or business as you would, and through the e cigarette company. So I was working with a lot of operational and data analysis in terms of chemical breakdown, and toxicity, stuff like that. And then I moved up to Kildare to work in Dublin. And I was with a company called “company”, that manufacturer paints, so I took over their supervisor and their CRM division. But alongside of that, I was doing production analysis, and hazardous goods data, export import analysis, financial analysis, sales analysis. So there's a lot of different data points to work with. And I am with “company”, I'm their CRM executive. So I deal with the full CRM database. We incorporates analysis of the not just the database, but the interactions between the clients and the partners and solicitors and matters in terms of the financial end of it, I do reporting and analysis on that based on frequency distribution maps, geo locations, some heat mapping. And most recently, we been rolling out an integration project, which will bring CRM and financial all together as oil as some with an automated mailing scraper. So it'll pull all of that contact data and frequency each charts and things like that some of the kind of more simpler aspects of it. Oh, and I did work for a little while for nonprofits. And then analysis of public education in the UK, Ireland. Then a bit of marketing on the side as well, SEO.

Interviewer   
It's an amazing experience to have like the repertoire where needlessly it speak for itself, and the longevity in the fields, which brings me on to what drew you to work in the realm of data analytics?

Speaker 2   
And I kind of fell into it. So when I started off in the States, I was college doing night classes, and I needed a job. So I applied to a winery and it was an Office admin job. What they didn't tell me was it was it was turning into a not a reconstruction job. But basically, the job was reconstructing income tax and large volumes of data over it was a four year period I was there. But it was a 15 year period that the data existed on it was transforming that from handwritten notes, to digital, and then analyzing reporting on those trends, etc. So I kind of fell into it. And then it's just been part of every job I've had since then it just kind of comes with the territory. But what drives me to stay in it is it's it's a puzzle that needs to be solved. But it's an ongoing puzzles, that never ends. So you don't get the end of it. You just keep going and there's new bits and pieces. So it's kind of like a big treasure hunt.

Interviewer   
That's a perfect analogy for the next couple of questions that I'm going to ask is going to be on the topic of the impact of CRM on data analytics. So I'll start off with what is the function of CRM in current corporate operations.

Speaker 2   
CRM is very misunderstood in corporate. So you hear CRM and people think, oh, it's customer service. There's a little bit of that. But CRM can be categorized into two external and internal, internal CRM is about dealing with your internal stakeholders, ensuring that they have the data to make the business decisions that drives progress. And that your data and your information is up to date. And obviously queries external is dealing with customers or other businesses, external third party businesses. And, again, it's about keeping those people happy, and updating your information, all of that stuff. And data analytics, again, incorporates they hear that and they think, almost like accountancy in terms of data points, they're the they assume that you're just looking at sales numbers and volumes, and tracking that. But in reality, sir, CRM it has, it's a multifaceted job. And, for instance, in my last position, I was CRM supervisor, I dealt with internal customers, external customers, suppliers, third party auditors, and my, the data I had to analyze was everything from in house manufacturing, to weather trends, and logistical routes based on you know, is there a strike in France, are the ports going to be closed in Denmark due to a storm that's coming in this week. So it's a global look at everything the business is doing. And then using the information you've gathered, to actually bring it together and produce something that's usable for the business to drive forward. And if I know that a product is under selling in a region, based on the information, I can then get my sales reps to hit that region in a specific way that will round it out. The world of CRM is not just oh, this is your job. And this is all you do in the world to CRM is you have to be able to do everything and put your hand to everything. For people that come into this kind of business, if they're not. If they're not multitasking and multi thinking and problem orientated, they won't last too long. Because it gets frustrating.

Interviewer   
You've touched on the next couple of questions, with your answer there. So I'm just going to split them into two. How as I say you've you've given a couple of examples there, on another one, how has the use of data analytics impacted CRM practices in your experience?

Speaker 2

Well, it's greatly enhanced it. And when I started out, as I said, it was paper, pen and paper. And it was notes that somebody had kept in a drawer. And all the information was in the business lead or sales or whoever it was in their head. It wasn't in a document that everybody could see. There was a very big, there's a very big disconnect in a lot of organizations where this department does this. But that department doesn't know that it crosses over. So with data analytics, and CRM, now we can say, Yeah, your department focuses on x. But these three departments are touching on your area, and have an add a new lead into that area. Do you want to maybe have a chat with them communicate over? Or I can say, okay, well, you know, this region isn't doing so well, why don't we take a look at what's happening there, and I can give that to business leads. And it also advise us to demonstrate our day service, okay. So traditionally, companies or management will look at financial figures, and they'd say, Oh, we're up 10%. This quarter, that's great. But by looking at the data, we can say yes, or, or our profits are up. But our costs are also up. So, we're not hitting those or there's these hidden costs that you're not seeing in your cogs because those are those are being offset. So a lot of companies won't count temps from agencies as a headcount. So, they could have 10 permanent staff and 50 temps but in their returns that 10 per month. and staff. So, they count that as a one-off service. They count them as service charges. But it has to go into the cogs, but it usually set off against another business center. So you're able to pull it all all that information together and say, yes, your profits are up. But the service charges are killing your margins. So, you're going to have to either reduce your margin, you're reduce your margins, sorry, increase your margins, reduce your your costs, or you're going to have to transform it a little more, expand your range, things like that. And within the service industry for data, data use in particular, where we're not offering any particular product, but an actual service. And it's less about pricing and more about utilization of people and departments, how can we, you know, have two departments work as a cross in a cross matrix rather than, you know, a stagnant form. And so it does open up a whole world of understanding. And I think one of the things for CRM, it's fine, because we're learning as we go. But for the business side of it, they need to be educated, to be able to understand what we're seeing. And because obviously, we speak different languages.

Interviewer   
You are a leader, it's great. You know, as I say, for someone that has so much experience across so many different spheres within these systems. Because as you keep leading into what I'm going to ask next. I don't want you to I don't expect to use up all your wealth of experience on these. Otherwise, we'll be here all day. And I'd want to be very particular on your time. So next up, could you provide a real world scenario in which data analytics had a substantial influence on CRM operations and those outcomes?

Speaker 2  
I'm trying to think of one. Okay, so I was working for a company, previous company, the paint manufacturer, their structure is a bit different than most in terms of the manufacturing industries. So they have a large, what they call internal manufacturers in Sweden and one in Italy, or Europe. And then the plants and say, Ireland, they buy the material from them. And so we obviously have that internal cost. Now, it's all under one company. So it's, it's just moving the funds about to pay off those costings. And a few years ago, we were doing an implementation on an Oracle project at Oracle, our 12. Basically, we had to break down every single product we had in the company, this corporation is global. 95 billion turnover a year one of the biggest manufacturers of paint in the world, if not the biggest, contracts with everything from military to civilian, you name it. So at six sites that I was managing on the implementation, what we had to do was basically recode every single product. And but during which we find that there were several project products that we were getting from these other sites that were overpriced. And our finance team had put that down to a currency discrepancy. Well, I went in and I wrote a sequel, program script, to track the currency fluctuations over the last five years. And then I took the data that they had sent on the pricing and sent put thought into it and ran it against each other while it was 40% higher than the current function fluctuations would allow for so what was happening was, they were meant to call charge those costs plus 5%. They were charging this cost plus 40%. And then the fluctuations were coming in on top of it. Now these are not for profit centers. These are manufacturing internal call centers so they shouldn't be doing that. So I highlighted it and brought the finance director for Europe the whole lot and basically what had happened is then it opened a can of worms on the head to price restructure pricing across your all plants and on it See, there was multiple questions. Why was this? but the data was there the data showed, this is what happens when there's so much cost for them to produce and your or, and seek this so much we are charged, this is transport, but we had all the information, we're able to transfer it all over. Basically, it was a legacy issue. And before a restructure back in 2005. Their plant managers had been on margins onto the products. And then when the structure came in, they were low data, but then obviously, it was already in the price. So the plant manager just didn't bother removing it. So the price just kept going up every year, because the cost of goods went up every year. So it was 40%. And top of all the costs, good increases, and etc, etc. And the interesting point, though, about that oracle system was, so I was covering six sites across Europe, it was fairly good, big team. And so we recoated nine and a half 1000 products, with six different variants of each products. And on the six different sites, that was six different languages, six different pricing structures, across 10,000 customers, and six months to recode. That and it was it was down to the historical data, we were able to centralize into a system, that data lotus, like I said, we will SQL scripts and once we had those, it was able to transform everything. But the data, the data and it allowed us to not just identify an issue with the pricing but alias to and then produce a solid database or post to our knowledge. So that whenever we implemented there was a point 01 deviation on the data. So it was fairly solid. I again, you know, yourself with data, it takes a long time. And it's the cleaning, that's the most important part. But you know, it is it was very important. And we had really good results. So we're very happy with that.

Interviewer   
In such a short turnaround, or six months is remarkable feature that really stands out of what you said there with the quantity and to be able to find such a problem within it. And being able to highlight it and create that system in six months is remarkable. You mentioned SQL, going to now go into the analytic methods in CRM, based on that. First one is going to ask is what are some of the most prevalent data analytic approaches or procedures that you've used and CRM settings?

Speaker 2   
Yeah. So, in terms of relating the techniques, it's a bit hard, but I would do depend on the situation. And I work a lot in Excel and SQL and Power BI and stuff like that. And so in terms of tools, Excel is always my base model. I use that for differentiating data, removing any kind of junk as they call it, cleaning the data. And then I'd use SQL for my loading tables into the systems that I'm working with at the moment that I'm working with LexisNexis at the minute, so I can do a direct edit that way. And but I haven't done any of the kind of deep dive data analysis, a lot of it is on the fly custom or bespoke analysis that's needed, just due to the nature so for instance, utilization of equipment in the last position, so we had X number of machines, we pumped out X number of gallons per our which Probert died per minute. And then I did a cost analysis based on a head kind of utility bills product called you know, etc, and broke down in cost per minute. And then we got a utilization ratio from that. And at the minute, the system I'm using is more about networks and contacts. So I do things like touch base contact analysis. So how often is such an this partner and touch with X number of clients per month? What's a bill biller mind. What's her data house or data health is based on an algorithm that was written by the company. So we're sitting at 83. We're doing you know, so it's a very, there's no approach, one approach fits all. It depends on what's been asked. And then within the job I can be asked to do to run analysis on attendees for events in the last year. No problem that's very simple. By doing analysis on attendees and events, people that have read these emails, people that have contacted people that have given us work, you know, so goes on and on. And then you have to kind of design around that. So in terms of theoretical methods, I wouldn't be overly up on that. But it's more of the practical side and, okay, working on the fly, if that makes sense.

Interviewer   
Being on so many different projects, have you seen or been able to explain some machine learning techniques such as predictive modelling, sentiment analysis, being used in those areas to obtain customer insight into CRM scenarios,

Speaker 2   
It's been more crm based software, it was always something that would have encompassed those so trend analysis customer, like, what you said predictive customer trends or behaviour, and sales analysis. And there may have been like, with the logistical side, it was weather patterns and road conditions. And then we threw in some cultural stuff as well. And there, there was the visualization aspect of it, but a lot of it is. A lot of it's kind of depart from the theory. I mean, you do things, your simple stuff, like your root cause analysis and all that for problems rather than your problems or whatever. And so why is why is it costing us x amount to do this service, when it costs us half the amount for the same service in a different location, things like that, that's fine. And trend analysis in terms of sales or services, or anything like that. I've done those before for clients. And we had so I had a client, maybe seven years ago, it was a natural gas car manufacturer out of Oregon. And they were looking at the trends for the use of natural gas. And I was able to produce a decent report based on the Madrid Metro lines and some other uses of it. But where I am now, in the last job as well, we did it wasn't really something that was needed, because the software was already there, that would pull that information straight out for yo And then because the last position was segmented, you had finance and then you had it was a multitier corporation where, you know, you had managers of managers and managers of managers and meetings about meetings about meetings. It did take a while to get through any of it, but I wouldn't have used any real deep, deep learning stuff.

Interviewer   
that's 100%. When it comes to data analytics, have you discovered any issues or limits when using data analytics on data?

Speaker 2   
I have come to some, some limits in terms of maybe the size of the data sets. And or maybe if well, you'll find for a lot of the databases similar to what I'm working on. There's missing information. So there's always going to be some kind of miss and missing information. And then one of our aspects of our software also is that People, the Public can sign up for, say, events, but they don't always fill everything in. So you have to go and manually find out what's going on. So updating that. And then with GDPR, as well, you're very restrained, you're very constricted in terms of what contacts you can make and how you use that data. So I could have 500 contacts or 500, data point contacts today. 200. But I do not consent anymore. Into the preference centre, all of a sudden, you've lost 200 of your data points, and you can't hold them, you have to remove those. So that is a big problem, because then you've lost all of that historical data.

Interviewer   
Yes. Do you think about it? GDPR has really changed the game when it when it comes to data. So one during my time with “company”, and the finance and insurance, and it was, it was all still very much keeping all the customer information and all the documents on folders. And then as soon as GDPR came in, everything had to be torched. And because the way that they had used the was anytime they'd bring someone back in for planned, it was like, Oh, don't worry, you put the customer at ease, we've got the information here, we've just got it all through again. But it was torched all that and then bred the talent, the customer, they need to bring it in, like fresh every time. So I can understand the limitations when it comes to GDPR.

Speaker 2   
And the problem with GDPR is because it was a one fit all solution, it was accustomed to industries. I mean, you go to your barber, and your barber knows what your haircut is, and knows your phone numbers and has that on records. Now you have to bring that in with that, or I'll give the marketing permission to hold that. You know, that industry doesn't need GDPR industries where your mass marketing Yeah, that needs it, no problem. But, again, it wasn't suitable for all industries. And like you said, the automotive especially where you've been with the same place for years. And all of a sudden, oh, no, we don't know who you are, we need to get all this. It knocks customers off. So it can be it can be very restraining, especially if you're working with historical data that you need his key points for your analysis, but they're gone. And then you have to speak with the business management team going, we did have this last month, but it's completely different this month, because you've lost all these people. And we lost these people because you didn't get consent. Or they're in Germany, and they needed double verification consent. So not only do they have to physically sign up, you have to send them an email to verify that they signed up. Where in Ireland, it's an uptight, it's not, you know, it gets a bit crazy when you're looking at different countries as well. So and that's the other thing that geographical restrictions on the data stuff, or the privacy stuff can get a bit a bit confusing as well. So I think in terms of restrictions or limitations on data, trying to run I mean, a simple like, I run Power BI off a lot of different our data, our sql base and Excel to all this stuff, even down to the technology end of it, the amount of technology or processing part it takes to run something like that. You know, it's called it's cost a bit of money to run them. I mean, we were looking at upgrading one of the SQL servers, and I think they were looking 20 grand upgraded. You know, so your crawl cost prohibitive, the limit of data, being able to process, and it's simple CSVs. And you can run that fairly straightforward, but it was more complex. If it has, you know, live links and geo locations and whatever other data you need. It can it takes time. It's not as straightforward. Oh, yeah, I'll have that in two minutes for you.

Interviewer   
The conversation we just had there perfectly runs into the next action on customer retention strategies. And in your opinion, how can good data analysis help design and strengthen client retention strategies?

Speaker 2   
Well, I mean, I've always been a big believer that your attention for customers is a personalized thing. You have to be a gardener with customers. You have to stand and take care of them and make sure that you can spot any kind of RSA that's coming in or fungus, that being competitors. With analytics, we can take a look at you know the purchase and trend and we can take a look at the marketplace that they're actually in as well. So not only looking at their what that customer has done, but look at what their competitors in the area has done on bespoke those reports. And we can even say listen to sales or business reps or whoever it is, we can say, Here's a report for your, you know, customer that brings in X million a year, here's a competitor analysis for them. So we're doing the work for them. We're giving them the information, and it's freely available, but it may be we are seeing something different from what they're seeing Am I did have in the market, and I keep going back to manufacturing plus, because I can't get too much into where I am now. The manufacturing place, one of the analysis we had to do was colour trends in terms of what the marketplace was calling for, on colour, paint colours, and kitchen colours. And then we had to do a spectral analysis based on I think it was a 64 data point rating, infrared rating of the colours. And the very deviate the variations between batches. So we did that for a customer across three year period. And we find that the deviation was point two, five per batch, which wasn't too bad, the industry standard was point seven, five. So this is as close as we could get. And they were very happy with that, because what was happening was up to a certain point, they didn't have the machinery to read those variations. And then after that the calibration of their system was different from ours. So, by doing that, I was able then to say to their sales guy or account manager, you need to get somebody in to calibrate their system to our system. So they're seeing what we're seeing. So we can prove that that information is there. And the customer appreciates that, you know, it's an extra step, the data allows us to see these things that they may not, not see but may not be obvious. And then again, you know, speaking purely from the sales side of it, or the customer, the CRM sales side of it, if I can see a customer is buying X amount of a specific product a month, and I can see the cost of goods have gone down, I can offer them a deal straightaway. And I have access to that straightaway. It were my sales guys might. And so you know, it allows us to see that. But in terms of retention, the more information we have on the customer, the better trends we can see, or the better information we can see on them. And then we can adopt our, our approach to retain them on to provide a service that is beyond what they're actually used to doing CRM for external bodies. I spoke to clients every day in the last job, and I could tell the mood of the client By the way they said hello, you know, so I can tell what I can do for them or what I can't do for them. By the data I had before the call is what allows us to prep for everything as well.

Interviewer   
That's a wonderful answer. And actually, you as it's been the common theme throughout this call so far, I ran into my next question, so I don't even need to ask that one. And we had we had touched on GDPR. So which brings me to the ethics side, what are some ethical concerns to consider when using customer data to update retention strategies.

Speaker 2   
I try to keep it , it’s a an overall retention strategy or information data. I tried to keep it as anonymous as possible. So obviously names and values stuff like that set aside because I know the business development, they want to know how many cases or how many, how many products we sold that week. And where are they sold to? I can say this chain bought this or this? You know, without giving the specifics. If it's for specific clients, obviously, there's a high requirement to be very cautious with the data. So we're not sharing anything that doesn't need to be shared. We're not sending documents we're sending links with expires on non-downloadable. Everything is cloud based it's not on this, the actual computers, or laptops. It's all we have in our in-house servers that we keep our most sensitive things on that are securely locked down in a locks server room. out in premises. So, it's not cloud based for that stuff. So, we're very hypervigilance, that we're I am now, because it's a legal practice very, very hyper vigilant about data privacy. We have some of the leading data privacy attorneys globally in this firm. So everything gets passed by them in terms of policy on what we can what we can share. On they have to sign off on it. And it does, though, again, it's, it's kind of inhibited in terms of, we can't go oh, here, look at all the information we've gathered and look at all these good numbers that we have, because we can't share that with specific areas. But in terms of the ethical, the only thing I'd say, again, it would be hard, it's hard to have to remove X number of data points from a system. Because you know, you need that data, but you can't actually do anything with it. So one of the things I do that I will do a report a complete anonymous report, and Joe on just numbers, no date, no personal details, nothing in it, it'll be X number of clients did this. And that's all it is, I'll have that before I delete information off the system. And then when it's deleted, it's completely it's shredded and cleansed. But it's the closest thing I can do to keep some kind of data from it.

Interviewer   
Yep, perfect. I'm moving on to future trends. And how do you envision the relationship between CRM and data analytics changing in the next few years?

Speaker 2   
I think it's going to become a bigger part. As I said, before, CRM has misunderstood job or position. A lot of people believe it's it's just customer service under a different name, I think what's going to happen is you're going to see a lot more data analysts, and business analysts going straight into CRM, and CRM becoming its own system. And it's not only it's not only data analysts, it's going to be your business development within the CRM, your retention specialists. And you may even see some of the HR people transfer into CRM for internal CRMs. Because that's a huge thing. At the minute, there's a big trend there, it's, you know, it's about, obviously off of COVID. Everybody's realizing now, they can work from home. And they can have a bit of a better life and a better balance, and HR driving that for a lot of companies. So, you'll start to see that I think. And I also think there be, there is enormous push to outsource a lot of this stuff to other countries. But similar to what you've seen where the production circle or cycle during the 80s. The countries that are being this has been pushed to are going to be over capacitated. On the hill. Basically, what happened during the 80s was manufacturing went to Africa. There was too much there wasn't enough people, it went to China, there wasn't enough people, it went to Eastern Europe, it was enough people came back. And we ended up you know, it was being made here anyway, so it was a full circle. So you'll see that being cycled through and then there'll be another big Porsche, but that that's a normal cycle. One of the things that you'd have to be worried about, though, is the sophistication of the scams that are occurring at the moment on the development of those, especially now that AI has kind of become easily accessible to the public. And your scripts are getting more intelligence. And large data sets are being uploaded to that and databases are being hacked from that. So that's another kind of area that's to be watched is how we interact. I mean, I heard I know that Microsoft are to bring out an AI, Oracle are bringing out one , Alexa snake, like most of the CRM companies are bringing out AI to accompany their systems, and it will ease a bit of the heavy lifting. But we're still going to have to make sure everything's 100% or as accurate as we can get because with the hallucinations or the AI are having at the moment it's, it's a bit worrisome.

Interviewer   
Again, which was what advice would you provide to businesses looking to use data analytics to improve client retention?

Speaker 2   
Just be careful in terms of what you're looking for. And if you're looking to know the history of your clients for the last 10 years, what the most popular products are and services, all of that stuff, that's fine, that's easy to do. No problem, I can give you that. I can, yeah, all of this. But if you're looking to know what the next big thing is, well, whoever's analyzing your data, if they have anything to do with your sales side of it, they're going to be influenced, whoever's presenting the data, they're going to be influenced. So what you want is you want to roll point, you want a non-expert , expert in the room. And what I mean by that is, take somebody from, you know, the warehouse. So, they're shipping every single day, they know what that's they know what's going on, right? They don't need to see the data to tell you they've shipped X number of pallets of stuff in the last three weeks. Take somebody there and present, give it a presentation, tell them and say to them, what do you think, Is this accurate to you, because they'll be able to give you a bit more insight. And on a great example of that was the Disneyland and Tokyo, where they brought a janitor in when they were designing the restaurants and said draw a restaurant. And the chefs and the executives all drew a standard, you know, slope, slope, roof, etc, well, the janitor drew sushi, and said, This is what a restaurant in Japan is. It's, it's sushi. So that's how they designed it. So, it's an known expert expert that you need to look at the data. And but in terms of use an AI, yeah, it's the idea behind it of being able to process such huge amount of information and produce those analysis is great. But you have to be very careful on what you're asking it. If you ask it to give you a trend analysis on this customer over these number of years, based on this product and purchase history. That's what it'll give you but it won't give you you know, any information moving forward. All AI is taking the information you're giving and it's re jigging it to present it. It doesn't have intuition. So, if you know that I'm trying to get. Yeah, so if you know that you're solvent base materials are your best sellers. But because of the push for environmental controls, water base is what you're going to be selling more next year. Ai doesn't know this, they only know the numbers you've given it. So, you have to be very, very, very careful with using any of the intelligence tools. And but in terms of human interpretations, there's a better you've a better reliability on that, I think, for intuition. And I think with the way the trends are going at the moment, and data analytics, and you're going to see deep learning AI, they're all going to be merged into one system. Which, you know, obviously, your machine learning is one level, your deep learning is another and then you'll have your AI to interpret everything that you've learned from those systems. And but I'd say just watch the space and maybe take a cautious approach to using any of those tools.

Interviewer   
Perfect. Finally, on the future trends aspect, What recommendations would you provide to organizations wishing to implement more data driven CRM operations?

Speaker 2   
Get the right people in place, get people that are naturally inquisitive people that can and are not, don't bring in people that are yes, people. Yes, people will say yes, that's no problem. Yeah, I agree with your data. Yes, no, but we'll get people that will actually look at something and say, oh, there's something not right here. Something needs to be tweaked. Or they'll actually look in deeper. You want an inquisitive mind behind data analytics, you want somebody that is looking at the whole puzzle, but looks at the fine details as well as your people or your base. So have them in place. First, have the right people there. Next, invest in the technology, invest in the crack software. If the software is not good, if the technology is not good, your people can't do their jobs. And then the third point, be patient with the data , clean, training, data cleaning is 90% of the job, and 10% is processing. Because we all know that most of the time the data is not clean, and it's out of date as soon as it goes in the system. So that to gather with knowing what you're looking for, it may take time to understand what you need. But go in knowing what you want. And then develop work with your team and develop at odds. Because at the end of the day, your analysts, your CRMs, they'll give you whatever you need, or sorry, give you whatever you want. But they don't know what you need until you tell them what you need. So there's a bit of a, there's a bit of a gap there where you may get a director or whatever business lead saying, oh, I need x. Grant, CRM data analysts gives him that Oh, no, that's not what I need. I want this Oh, that's not what you asked me for. So, it's all about we're going to communicating with CRM, because at the end of the day, that's where your information is coming from

Interviewer   
100%. And I recently had an interview for a data analyst position, and in that interview process, They asked me, well, what would I do? And the first question I asked them was, what are you looking for? It's like I can do all these different processes that from feature segmentation, you know, processing everything, unless you're going to tell me what specifically you're looking for. While I could do a time series or sentiment, but it's not going to lead towards the answer that you would like. And as you say, once you get that information about what they're actually specifically looking for, and being able to work from there, that is such an important feature of it.

Speaker 2   
Well, that's it, I mean, you can make it look as fancy as possible and apply all the techniques in the world. And, but at the end of the day, it doesn't result in what they're looking for, they may want a little pie chart showing the differences between three segments, no problem. But if they tell you, you want an analysis of three different segments across this time period, and you, you're going to produce that when all they want to do is see the percentages of sales or whatever it is, you know, so it's all about the communication and understanding what they want, understand what they want or need, and then explaining it to you. And then you can go and do whatever you need to do. There's not a big not a big difference in the vocabulary. I mean, you can talk about data analytics in the simplest form possible. That's no problem. Because at the end of the day, it's a filing cabinet. Each file is a folder, each folder contains spreadsheets. But in that spreadsheet, it's all your data. How do you want me to show you that on the board like this, this is kind of part of it? And what would you want me to present to you and give you the information you need. And it can be a bit frustrating for anybody on the business side when they're looking this information, because obviously there's always urgency, etc, etc. That's another thing that companies need to watch out for is they don't need to be overly patient, but they need to understand that it's not a click of a button to produce the results that they want. It's not a phone call, it takes hours and hours of prep and processing to get the data correct. And if they put the work, you know, the company or agency puts the work in, then they're going to get the results they want. And they're going to get them very clean very quick once everything's in place. But until it's in place, there's a lot of barriers there for the teams.

Interviewer   
Finally, as we reach the end of the questions, I'd like to ask is there anything further you'd like to say or discuss about the influence of data analytics on client retention?

Speaker 2  
And I think it's underestimated in terms of what you can achieve with data analytics with the data points on retention. And not just on sales, but on personal personalities, etc. I mean, data while you have to be inquisitive, and well you have to be able to problem solve. You have to be intuitive as well. You have to know say for instance, you have I'm a big CEO, that is splitting their time, 15 minutes a meeting a day. That's all they can do. Well, their data analysts can look at. And this sounds very strange, but can do a social analysis of that CEO that you're approaching with this pitch, that pitch could be worth 100 million quid on the information you're going to get from a data analyst, is going to be very important. Because does that CEO prefer? Coffee? Tea? Does that CEO prefer to come in through the west side or the east side of the building? Does that CEO refer you to finish a 1430 or 1439. And that's not easily accessible, but the data is there in, you know, in this sphere of information, that we live in that kind of social engineering for data analytic analysts is very easy. As long as we have the information, we can break that down. And you see that with a lot of intelligence organizations, and that's what they do, they take all this data, give it to analysts, who then break down a social trend, to approach new, you know, new people, or whatever they do. But in terms of retention, I think it's good to see, you know, if you're approaching large company, or you have large clients, large company, and you're able to say, right, I need to see how many people are letting go or I need to see how much of the market these people actually have? Well, you're able to offer that client, you know, a better deal based on their marketing position. So if they have 30% of the market, and you know that they're going off to the next 10%, you're able to get in the ground floor, and the data analysts can give you them the market positioning, and then give your competitor analysis and the demographics of the area and, you know, words, the most likely location for their next site based on sales, and you know, all the rest. It brings in spherical plan, and it brings in a multitude of different things. But I would say that for retention, don't only look at the sales numbers, look at the whole picture. Look at what they've done, who they've spoke to in the newspapers. You know, if you had last year, if you had somebody that was from Twitter approaching Musk about the deal? Well, I am 100% certain that he had analysts working on everything to do with Twitter. And that's why he kept saying No, I don't want to buy, I didn't want to buy, because his analyst was saying, Yeah, reject, reject, reject, oh, go for it. It's, you know, it's all about the numbers. It's all statistics at the end of the day. But it can be used in so many other ways that and I don't think businesses see it, to appreciate it or kind of understand what it can actually be done or what can be done for it. Because a whole profile can be built around clients. And it's it doesn't even it's not just the business end of it. It's the public appearances. It's everything else as well. And that might be outside of the scope of it, but that's just kind of my opinion on what can be done.

Interviewer   
exactly, that's exactly what that question was for was opinion because it's such a vast amount of experience behind the opinion is priceless. We've come to the conclusion of all the questions.

**Interview-3**

Interviewer   
Thank you. And the last thing i I'm sure you're aware from looking over these masters that I will give the college the email address, which they obviously already have, just because they're going to pick someone not random to say that these interviews weren't fabricated.

Speaker 3  
You know, and I'm happy to support you if it's necessary. So yeah.

Interviewer   
Perfect. Thank you. Well, thanks for joining me today, it really is appreciated. We'll start off with just a quick introduction and background. So could you please introduce yourself and your background in data analytics? And CRM?

Speaker 3  
Yeah, so I am actually an engineer, I work in a consulting company called “company”, and my background is mainly software development DevOps. But lately, I'm leaning towards data engineering and data analytics. So right now I'm working on an ESG information system for investment funds and investment banking, like money management, finance, and investment banking. So I'm doing a lot of data transformation and data analysis for them. So

Interviewer   
What drew you to work in the realm of analytics?

Speaker 3  
Actually was almost like natural, like there was a gap in my team, and we didn't have any budget to hire anybody else. So I kind of like ended up doing the work myself.

Interviewer   
Okay. The next set of questions are going to be the, on the impact of customer relationship management and data analytics. The first one is, what is the function of CRM and current corporate operations?

Speaker 3  
Yeah. Well, at the moment, I'm not working with any CRM. But I can tell you, I've been a consultant for many years. Like one post I've noticed is that all the companies that were operating open an Excel spreadsheet, they have stopped doing so. And now they're operating over a CRM. The reason behind that is that integration with social media email integration with you know, direct customer contact, like things like Hubspot, and the others, which actually allows you to keep the track, set reminders and things like that. It's slowly winning the market. So I'd say right now, I don't think anybody will set up a company which deals with customers. We have the CRM and enterprise here and whereas five years ago, it was totally doable.

Interviewer   
From your experience, how has the use of data analytics impacted customer retention practices?

Speaker 3  
Well, pretty much very much, because, again, you know, since Tableau and all the other started coming up, like I can see companies like, you know, analysing, what's the profile of the customer? What's the profile of the customers that dropped from the system, even feeding that into machine learning algorithms to predict what's the dropout rate of a customer, purely dicing, you know, tickets on support, like, our customer has an issue, if that customer is already on the category you have, he's going to leave within the next three to six months, you know, like, has a lower priority, but then if you have a customer, which is on the category of A is going to be here for at least another two years, because higher priority? And I think that analytics, it's making them, so to say the revenue streams more efficient? In my opinion, so.

Interviewer   
could you provide a real world scenario, from what you worked at, in which data analytics had a substantial influence on customer retention operations or outcomes?

Speaker 3  
Well, I can, yes, but it's a bit finicky. So we, I used to work for a company called “company”, in the past. And CRM, actually, we want to avoid that we didn't have enough capacity. So, it's basically like, you know, serve the customers we had in the pipeline. So, what we did is we pull out a record of polling traded companies in, in, like, you know, it's full of information on what revenue they have. And we actually need some calculation which shouldn't speak to them, or which customers was going to cut a major or massive impact in our finances. And then we prioritise them. And what we did is we redirected some of the customers to kind of like friendly companies, where we know that we're going to be looked after well, basically, on the fact that we wouldn't be able to attend them as we thought it was going to be possible. And that's coming from a CRM crunch with data from companies, he becomes like growth, you know, like spending and things like that.

Interviewer   
Okay, the next couple of questions are going to be on data analytic methods in customer retention. Could you what are some of the most prevalent data analytic approaches or procedures use in a customer retention setting?

Speaker 3  
Well, I mean, it's something close to Crisp DM. But it's very bespoke, to be honest, like, most of this work was done by marketing, and was done by people who are not into data science. And, like, we supervise it in a way that, you know, like, we, we made sure that the numbers made sense to us. But at the end of the day, it was pretty much you know, bespoke, and it had a bit of a good feeling. But I will say, Crisp DM to me, is possibly the most accurate method. Now, we are talking about big customers, we are talking about accounts, which are millions, rather than hundreds of euros. Like if you were, like dealing with for example, I work for another company, which was about forex, if you were dealing with, with other companies, like, I mean, I don't see the cost benefit of using for that, like, if your customers only deal like a few 100 euros on their lifetime, you probably segment them. And then once you have been segmented, try to maximise the groups that you can stretch the most like, you know, there's always a segment of customers which stay 1,2, 3 months and leave segments of customers that they are there forever, you don't need to worry about this job. But then you have customers which just stay with you two or three years and then go to the competitor or yes go that's where you need to focus the efforts and that's where your own marketing campaigns and you know, free games or things like that. So that's how I think you know, like it will go for segmenting customers, you know, seeing what are the common pain points on the ones that leave after a couple of years or a few years’ time and motivating them to stay.

Interviewer   
What you said there Linked in with another answer I got from one of my other interviewees who stated that like they were looking to invest more in the area and it was going to cost 20,000 just for the use of the systems. Look to the bigger side of companies compared to your smaller set your local bakery down the road. So there what you've said there is 100% collaborative, with other answers that I've gotten. Could you explain how machine learning techniques such as predictive modelling, have been used to obtain customer insights?

Speaker 3  
Well, in a way, it's not really so much about customer insights, but it's more about kind of customer satisfaction or improving that path. So, I used to work for a company that ran a massive dormant campaign where they will give you something like 10 euros, if you trade something within the next 30 days. So, at some point, what happened is the Know Your Customer process was very clumsy. And you need to send two pictures, a passport photo proof of address, you know, like a few things, and getting the customers through the system. We know your customer system. It's really painful, like it's really slow. Like one agent could only process something like 20 or 30 customers a day. I don't remember the numbers. But, you know, so what we did is we introduced machine learning to process things like passport, does dispatch forecasts anywhere, does this part, passport has a phone photo of one person, what's the address is passport. You know, things like that. So we will reduce the noise about 60, 70%. Because what's the point on somebody reading a customer and then finding on the last photo, that it has a glare or it's good. So, you can't accept that as I know your customer. So, we cannot lose the customer there. That actually helped us because the competence at the time was very slow to process, the onboarding. And we could onboard customers on the same day. So that actually gave us insights on customers want to pretty much log into the system, overload the photos and start using it straight away. So that's, that's an example. That was done by machine learning. And it was done by Google Cloud Platform with all the vision API's and all the sort of like composing API can so yeah,

Interviewer   
Have you discovered any issues or limits when using analytical methodologies to data?

Speaker 3  
Yes, the data sets never hold truth 100%, the complete market hypothesis in finance, it's true only if you have all the data, if you don't have all the data, it's not true. And then the most glaring thing here is, you don't have all the data ever, you only have the data of your customers, you don't have the data of the people who are not your customers. So, if you make any decision you maximise for people who are already your customers, but you don't know what people are not working with you are looking for. So that's where I think that analytics, even though it's a very powerful ally, it doesn't really give you the full picture. Does that make sense?

Interviewer   
yes. 100%, as you say, like you're working with your own segmentation of the market and not by what other Companies would have.

Speaker 3  
Send that blend to the wire, and then you come back with the plane. And it's completely, you know, completely wasted with bullets and things like that. And you basically say, you know, like, from all the planes, they got back, where do I need to put more steel to reinforce them, and you put the steel where the bullets are, but you don't have the information of which planes fell into the ocean, because they, they just destroy them, which are the ones that you want to protect, not the ones that came back, the ones that came back are fine. Same with customs.

Interviewer   
It's a very good example. The next three questions are going to be on retention strategies. And I'll start with in your opinion, how can good data analysis help design and strengthen client retention strategies?

Speaker 3  
What I think by loading the data, it's what gives you retention. And let me explain that. So, imagine that you are a bank, and a bank or insurance company or similar, and you start capturing clicks on the website and you start capturing motion of the mouse and things like that. And then there you can work out patterns of this user has been trying to get a quote for the next 20 minutes. And the pattern is you know, 10 minutes getting a quote, 10 minutes of radio silence and then 10 minutes back, I just in the quote. So, things like that indicates to me that this customer is looking for the best price possible. So, you have the information of the customer, you have the information or the pricing of the world. So that's what I think the date analytics of all these data sets will help immensely companies to have retention. So, we get to a point where you know, you know that this customer got car insurance for 600 euros, and then he probably got sick in other websites, what's the best price we can do to them, which will still make money. And the next best action is called the customer as a we can give you another 50% discount or a percent discount or 50 euros, it's gone. If you come with us now, that probably is going to be the tipping point for the customer. And I think this real time analytics where you can see real time which customers are caught in your website, or your systems, it's the future.

Interviewer   
Your answer has actually led into the next question where I was going to say can you provide instances of effective customer retention strategies that were established or improved use and thought analytic insights?

Speaker 3  
For example, anyone, like you send a customer to renew an email for the insurance, and then what happens is they never open the email, or they open the email, and then never pay for the renewal. So that gives you information on what is this customer missing, to retain, you know, to be retained. So, you can do the same reach out to the customer. Hey, we saw you open the email. Is there any problem with the quote, is it too expensive. And then the interesting point is that there might be a multiple amount of reasons. So, you know, that's what I think offering these columns and offering, you know, like, kind of modelling the price of your insurance, with this analytics in mind, will give you some margin to actually provide extra discounts. Like people tend to buy things more on the spot when they're talking to a human rather than a system, like when you talk to a system is very cold, like the gas, you know, get that as price. It doesn't matter if the battery is blue or green, I just want the best price. But when you're talking to a person, and you see reassurance in there, and they're you know, there's a close contact that can help you to maintain the deals. So that's that's one example.

Interviewer   
On the ethical side, what are some of the ethical concerns to consider when using customer data to update retention strategies?

Speaker 3  
Well, depending on the industry, I mean, like, for example, I used to work for a place where the insurance was cheaper for women than for men. So, they wouldn't bother too much about women, because at the end of the day, the competition was fierce. And the mandate was, you know, be pushy on men, but don't worry too much about woman, because the margins are very small. So even if you win the deal, basically, what you are doing is losing money, because the amount of financial effort required to follow up to win a deal is not worth what the money is coming in the company. So that to me, raise ethical concerns. So, you know, if you're working with personal data, medical records, things like that. It's problematic. But yeah, in general, I think there should be the same way that in banking, there is a compliance officer, there shall be a data compliance officer in place on companies, which actually have you know, deals with, with customer data, which is of sensitive nature, you know, like, the car insurance thing actually went away, because it was banned, the discrimination by gender, like you have to provide the same price being man or woman. But I can see holes, clustering customers, even risk factors can lead into ethical issues, like for example, somebody disabled with, you know, like minor vision impairment, which is fit for driving will get a hard time getting an insurance quarter the moment. So, yeah, it's, that's where I think the ethical concerns arise, if that makes sense.

Interviewer  
The next couple of questions are going to be about future trends. And how do you envision the relationship between customer relationship management and data analytics changing in the next few years?

Speaker 3  
I think it's going to be embedded with next gen intelligence. I mean, all these language models, like I don't see how, in the next five years, a company is going to survive. If you're going to start embracing language models to follow up with the customer retention strategies and CRMs and you know, all these customer management, like I just don't see how I mean, it's, I'm doing some work on the side about that about machine learning models and let me tell you, it's scary what you can achieve with very little. So, we are back into the FinTech revolution 2.0. So, when FinTech started growing it just cover some bugs. And they didn't even cut down to realise, I'd say right now, we are in a point that somebody with very low investment could build a CRM, which is driven by AI and gives you insights about the customer from AI, and could take over HubSpot, and the others very quickly. Watching, you know, imagine an agent, which is getting like 200 emails a day, and you don't have time to read them. And that's when you get like, we're going to get back to you in the next week. Something like that. Like, can you imagine instead of reading every single email, getting a prioritisation of which emails are hottest, and two-line resume for every email, like, you know, this guy's not very happy. This person feels like it could wait for a couple of days. I think that I will say this is the next big thing.

Interviewer   
Yeah, it's like integrating a bit of sentiment analysis within the machine learning to be able to tell from the responses. Who do you need to contact first? Or who? Who is happy from the reply, and you don't have to really worry about  
The next question is what advice would you provide to businesses looking to use that analytics to improve their customer retention efforts?

Speaker 3  
Well, I think in general, the advice I would give is model your current process, but the decisions you make make them uninformed data, at the moment. Before data analytics, many companies will actually just do the customer retention problems. As of you know, like, when you are fishing with an IT, like, you know, throw the net into the water, hold off, and then oh, I got 20 fish here. But 200 metres down the line, I got like 2000 fish. So obviously, I'm going to go 20 metres down the line. But then somebody with a fishing rod in that 20 fish per catch up and throat area can be making a huge amount of profits, because they have a more much more finer data analytics process. So, what I recommend is use data analytics to increase the efficiency of your business, but not to do what you always do. So, yep.

Interviewer   
What recommendations would you provide to organisations or businesses wishing to implement more data driven operations,

Speaker 3  
Take care of the quality of your data, the quality of the data is super important. And it's the concern number one, the quality of the data it was drives your noise. So, you know, if you will feature in the TV, you will get a lot of traffic in your website, which really are in sales, and adding converting, but if you use that, it will be troublesome. So, you just need to filter by posts of those customers converted, but you still will have a black swan event because you're featured on TV and more people will buy on the spot, and things like that. So, make sure that your data is curated. It's a concern number one and that by the way is very expensive.

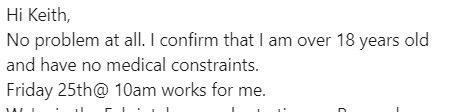
Interviewer   
Alot of the answers have said that in relation to the data quality and then the EDA involved is as you say, not only expensive but the importance and the time consuming more solid tonight and and it takes up 80 to 90% of the priority of the data.   
Just to finish now, we've come to the very final part of this interview David finally is raising for her to say or discuss about the influence of data analytics on client retention.

Speaker 3  
Um I will say that data analytics in my opinion on client retention eats what property agents are to the property market and market dynamics there's so because we data analytics you can reach more suitable customers and you can reach you know more suitable deals, it makes everything more way more efficient, and it makes everything more, you know, dense. So, it raises the bar of the entry level. So, you know, if you are HubSpot, you are everywhere. Like if I want to set up a company and I use a CRM, I will prefer hubspot, because that's what everybody does. But broadly, there are like another 200 SAS startups which do exactly what I need, but because they don't know them. It's hard for me to reach them. But because HubSpot has a huge amount of market and has like a huge amount of data analytics, like they can be what I see them, so I will be more inclined to work with them and any other small startup. So, as I say, data analytics, I think its dynamics in the market and making it very polarised. Is it for the big shots? Harder for the smallest? That's my feeling at the moment.

Interviewer   
Perfect, we have reached the end of the questions I'm going to stop the recording.

**APPENDICES- C - Interview Consent**This section shows the permission of consent given to the researcher, including the confirmation that volunteers are above 18 years of age and have no medical constraints restricting them from partaking in the interview process. The 1st two were confirmed in direct messages and the 3rd confirm at the start of the Interview.

Interviewee 1 consent



Interviewee 2 consent



Interviewee 3 consent

