UNIVERSITY OF ECONOMICS AND LAW

FACULTY OF INFORMATION SYSTEM

FINAL PROJECT

DIGITAL MARKETING ANALYSIS

231MI4901

TOPIC: ANALYZE THE CUSTOMER CHURN RATE OF TELECOMMUNICATIONS COMPANIES

Lecturer: Le Hoanh Su, PhD.

Group: JM

1. K214162149 – Le Tuan Nguyen
2. K214162159 - Le Thi Bich Van
3. K214162151- Hoang Thi Nhat Quyen
4. K214160986 - Phan Thuy Anh

Ho Chi Minh City, November 30th, 2023

JM GROUP

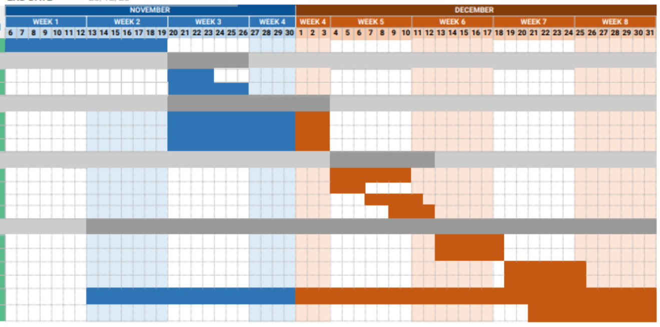
No	Full Name	Student ID	Contribution
1	Le Tuan Nguyen	K214160993	25%
2	Phan Thuy Anh	K214160986	25%
3	Hoang Thi Nhat Quyen	K214162151	25%
4	Le Thi Bich Van	K214162159	25%

GANTT CHART

GANTT CHART

PROJECT	Analyze the customer churn rate of telecommunications companies	STAR DATE	6/11/2023
GROUP	JM GROUP	END DATE	25/12/23

NO	JOB TITLE	PIC	STAR DATE	END DATE	TIME PERIODS	PERCENT COMPLETION
0	Picking data modules, Define problems	All	06/11/23	19/11/23	13	100%
1	**Project overview**					
1.1	Reason for choosing the topic	Nhat Quyen	20/11/23	23/11/23	3	100%
1.2	Objectives, Target, and Scope of the Project	Thuy Anh	17/11/23	26/11/23	9	100%
2	**Theoretical basis**				0	
2.1	Customer Analytics	Nhat Quyen	20/11/23	03/12/23	13	100%
2.2	Methods and Tools	Thuy Anh	20/11/23	03/12/23	13	100%
2.3	Measurement Metrics	Bich Van	20/11/23	03/12/23	13	100%
3	**Data preparation**					
3.1	Collection data	Thuy Anh	04/12/23	10/12/23	6	100%
3.2	Exploratory Data Analysis (EDA)	Tuan Nguyen	04/12/23	06/12/23	2	100%
		Bich Van	07/12/23	11/12/23	4	
3.3	Bivariate Analysis	Tuan Nguyen	09/12/23	12/12/23	3	100%
4	**Telco Churn Analytics**					
4.1	Model Prediction	Tuan Nguyen	13/12/23	18/12/23	5	100%
4.2	Visualize with Power BI	Bich Van	13/12/23	18/12/23	5	100%
5	**Solution**	Nhat Quyen	19/12/23	25/12/23	6	100%
6	**Conclusion**	Thuy Anh	19/12/23	25/12/23	6	100%
7	**Slide**	Thuy Anh	13/12/23	31/12/23	18	100%
8	**Word (Final)**	All	21/12/23	31/12/23	10	100%



COMMITMENT

The project "ANALYZE THE CUSTOMER CHURN RATE OF TELECOMMUNICATIONS COMPANIES" is our research work, conducted under the guidance and enthusiastic support of Dr. Le Hoanh Su.

The dataset is provided by IBM, and the simulation results and solutions presented in this project are the outcomes of our analysis, based on the knowledge acquired from the course.

All reference sources used in this thesis are clearly published in reputable scientific journals, cited, and properly acknowledged.

Group JM

ACKNOWLEDGMENT

We would like to extend our deepest gratitude to Ph.D Le Hoanh Su, our esteemed supervisor, for his continuous guidance and unwavering support throughout the enriching two-month learning journey. This project stands as a testament to his dedicated mentorship and profound wisdom, which have been instrumental in our success.

Our sincere appreciation goes to Mr. Su for generously imparting invaluable knowledge and insights that not only facilitated our improvement but also played a pivotal role in the successful completion of this project. Your teachings have provided us with a thorough and comprehensive understanding of the subject matter, enriching our overall learning experience.

Group JM

Ho Chi Minh city, 30/12/2023

TABLE OF CONTENTS

TABLE OF FIGURE

CHAPTER 1: PROJECT OVERVIEW

1.1. Reason for choosing the topic

Choosing the topic "Telco Customer Churn" is based on the actual challenges that the telecommunications industry is facing. In a competitive business environment, customer retention becomes a key factor determining the success or failure of a telecommunications business. The loss of customers not only affects revenue but also poses major challenges in terms of maintaining a business's position and building its reputation.

Research not only helps predict customer behavior but also provides cost optimization opportunities. By analyzing and understanding the factors that lead to churn, businesses can develop flexible retention and loss prevention strategies, thereby reducing the costs associated with attracting new customers.

Faced with the increasing diversity of customer data in the industry, from transaction information, and service usage behavior to feedback from social networks, the topic becomes attractive with comprehensive analysis capabilities. A better understanding of customer behavior can help businesses adapt and apply flexible retention measures while optimizing business strategies to retain and attract new customers.

The "Telco Customer Churn" study not only focuses on data analysis but also uses advanced technology such as artificial intelligence and machine learning to build more accurate prediction models. This combination of telecommunications and new technology not only opens up new opportunities but also creates added value.

The results of the research are not only theoretical but also have practical applications for telecommunications businesses. The information and retention strategies developed from the topic can be directly integrated into daily business strategies, helping businesses optimize performance and strengthen relationships with customers. In total, choosing the topic "Telco Customer Churn" not only meets research needs in this field but also brings high practical value to businesses, promising to open up new opportunities in relationship management. customer system and comprehensive business strategy.

1.2. Objectives, Target, and Scope of the Project

1.2.1. Objectives

Objective 1: Develop a Predictive Model for Churn Behavior

The main goal of the study is to build and develop an accurate predictive model for customer churn behavior in the telecommunications industry. By utilizing advanced data analysis and machine learning methods, this research aims to create a predictive tool capable of identifying customers with a high churn risk, providing crucial information for businesses to formulate effective retention strategies.

Objective 2: Analyze Factors Influencing Churn Decisions

The study will conduct in-depth analysis of the factors influencing customer churn decisions in the telecommunications industry. From transactional information and service usage behavior to cultural and social factors, the objective is to gain a better understanding of which factors significantly contribute to customers' decisions to leave the service.

Objective 3: Propose Optimal Retention Strategies

This objective focuses on proposing purposeful retention strategies based on the analysis and predictions. The research will construct a framework for implementing retention campaigns, assisting businesses in optimizing flexibility and responding quickly to changes in customer behavior. The ultimate goal is to help businesses prevent losses and maintain strong relationships with existing customers.

1.2.2. Subjects and Scope of the Study

- ***Research Subjects:*** The project is conducted on a dataset consisting of 4,709 entries related to the list of current and churned customers of the telecommunications company.

- ***Scope of the Study:*** The project is conducted on the Telco_Churn_Data.csv dataset.

CHAPTER 2: THEORETICAL BASIS

2.1. Overview of Customer Analytics

2.1.1. Customer Analytics concept

Customer analytics is an important area of modern business strategy, placing customers at the center of every decision. It is not just about collecting and processing data, but also about a deep analysis process to understand customer behavior and needs. From small transactions on websites to large interactions on social networks, customer analytics collects and combines all sources of information to create a complete picture of the customer. (Murphy & Kiwak, 2023)

More importantly, customer analytics not only helps businesses identify current customers but also predict future trends. By using predictive models and advanced analytics tools, businesses can adapt their marketing strategies and serve customers flexibly and effectively. With this ability, customer analytics is not only a data analysis tool but also a breakthrough in building close and sustainable relationships with customers, opening up new opportunities for growth and development. business success.

2.1.2. Importance

In the age of digitalization and fierce competition, understanding and responding flexibly to customer needs is the key to maintaining and growing a business. This has fueled the growth of the field of customer analytics, where companies not only collect data but also apply smart strategies to optimize every interaction with customers.

One of the main reasons companies put so much effort into customer analytics is its ability to provide a detailed and multi-dimensional view of customer behavior and desires. From online shopping data to social media interactions, every customer touchpoint generates valuable data, helping businesses build detailed and multidimensional customer profiles. This not only helps identify personal characteristics and preferences but is also the basis for creating a personalized experience for each customer. (William & Mary, 2023)

Not only stopping understanding, customer analysis also provides tools to predict future customer behavior. Using predictive models and machine learning algorithms, businesses can come up with potential scenarios of customer interactions. This helps them optimize marketing strategies, services, and even products to meet expected customer needs, creating a flexible and responsive business environment.

Another important benefit of customer analytics is the ability to enhance customer engagement and relationships. By understanding their desires and responses, businesses can create high-performing messaging and engagement strategies. This not only creates a positive customer experience but also drives loyalty and strengthens brand positioning.

Finally, customer analysis is an indispensable tool for shaping a business's long-term strategy. Continuously monitoring and reflecting on customer data helps businesses adapt quickly to changes in the market and maintain the flexibility needed to survive and grow in today's dynamic business environment...

In short, customer analytics is not just a data analysis tool, but also the key that opens the door to deep understanding and flexible interaction with customers, creating endless opportunities for customers. business development and success in today's digital age.

2.1.3. Operational process

Customer Analytics is a multidimensional process that not only helps businesses understand current customers but also predict and interact with them flexibly. Here are five key steps in the process:

2.1.3.1. Collecting and Processing Customer Data

The customer analysis process starts with data collection and processing. For this step, businesses need to collect information from many different sources, such as transaction systems, websites, and social networks. This helps create a rich and diverse data set about customer behavior. Next, data processing is extremely important to remove noise and standardize information, ensuring accuracy and consistency. Using efficient cloud storage and database technology helps ensure safety and easy access to customer data. (*Customer Behavior Analysis: What You Need to Know | Binarbase*, 2023)

2.1.3.2. Data Analysis

Multivariate data analysis and clustering play an important role in finding important relationships and trends in the data. Multivariate analysis helps businesses better understand the relationship between variables, while cluster analysis can classify customers into segments based on common characteristics. Predictive modeling and the use of machine learning algorithms provide the ability to predict future customer behavior, opening up opportunities to adjust and optimize business strategies. (*Customer Behavior Analysis: What You Need to Know | Binarbase*, 2023)

2.1.3.3. Building Customer Profile

Data analysis helps build detailed profiles of customers, from preferences, and demographic characteristics to shopping patterns. By understanding this characteristic, businesses can create personalized experiences, providing products and services suitable for each customer. Customer profiles are an important tool for customizing your marketing strategy and serving your customers flexibly and efficiently. (Oates, n.d.)

2.1.3.4. Prediction and Optimization

Data analysis not only helps understand current customers but also predicts future behavior. Using predictive modeling and machine learning algorithms, businesses can tailor marketing and service strategies to meet expected customer desires. Strategic optimization helps businesses become flexible and responsive to changes in the market.

2.1.3.5. Interaction and Feedback

Ultimately, customer analytics is not just about understanding but also about interacting with them. Create smart engagement strategies based on information from analytics to help maintain and strengthen customer relationships. Continuously monitoring and improving strategies based on new data and customer feedback helps businesses remain agile and ready to respond quickly to changes in the market.

Customer analytics is not just a tool but a comprehensive strategy that helps businesses build and maintain strong and lasting relationships with customers. In today's digital age, a deep understanding of customers is not only the key to success but also the deciding factor for a business's survival in today's market.

2.2. Methods and Tools

2.2.1. Logistic Regression

Logistic regression is a widely used statistical method in predictive analysis. In the context of customer analysis, logistic regression is often used to predict the probability of an event occurring based on one or more independent variables. For example, it can be used to predict whether a customer will purchase a product, whether a customer will continue to choose a business, etc. This information can then be used to formulate strategies and policies to retain customers, calculate the risk of their departure, predict the potential customer growth in the future, or classify customers into specific groups.

To build a logistic regression model, it is necessary to prepare the required tools, specifically by installing basic libraries such as NumPy, Pandas, scikit-learn, matplotlib, and seaborn in the Python programming language. Python is a flexible and popular language in the field of data analysis and machine learning, providing many powerful libraries for building and evaluating logistic regression models. Below are some ways you can use logistic regression in customer analysis.

2.2.2. Predicting Churn Probability

This method can be used to predict the probability that a customer will leave the service or stop using a product. This helps in implementing measures to retain customers. The process of using logistic regression to predict the probability of churn is modeled as follows:

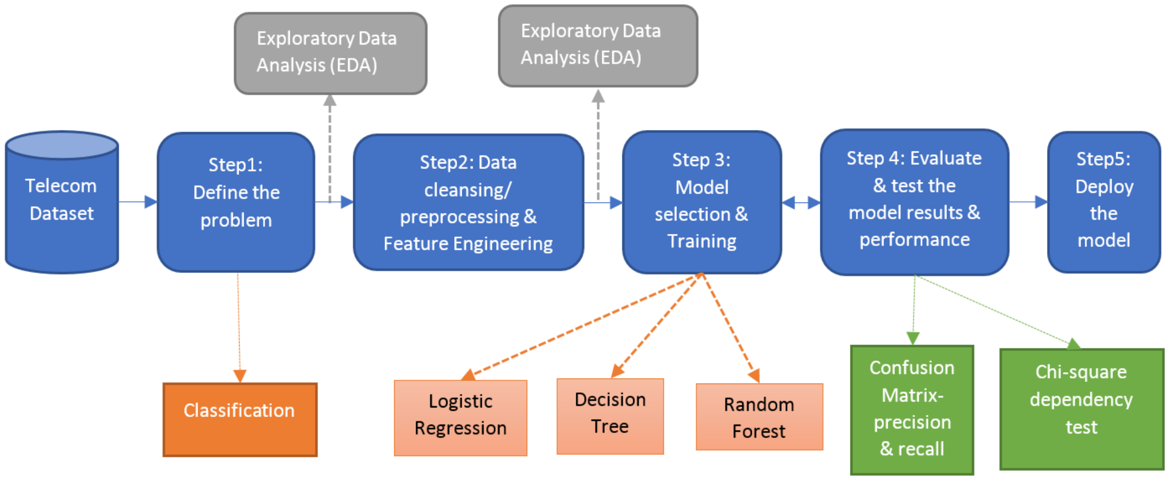


Figure 1: Steps to building a Logistic Regression Model of Churn (Customer Churn – Logistic Regression With R - DataScienceCentral.com, 2017)

After completing the first step, which is importing the necessary libraries, proceed to split the data into two sets: training (train) and testing (test). Additionally, the data preparation for both train and test includes data preprocessing and data normalization steps to ensure the model's construction. Next, you simply need to call the LogisticRegression() method to create a regression object and use "fit" to train the model on the training data. In the model performance evaluation step, accuracy_score will be used to calculate accuracy, and classification_report will be utilized to display other metrics such as Accuracy (Accuracy: Percentage of correct predictions out of total predictions), precision (High precision: Percentage of correct positive predictions (true positives) out of the total positive predictions (true positives + false positives)), recall (using the same formula as precision but differing in focus, recall concentrates on capturing as many positive emotions as possible, reducing the risk of missing positive emotions), and F1-score (Combining precision and recall into a single index, the harmonic mean of both, suitable when both sensitivity and accuracy are important) on the test set.

2.2.3. Analysis of Factor Influence

Logistic regression models can help determine the level of influence of specific factors (e.g., price, product quality) on customer purchasing decisions. Similarly to the process of building the regression model as analyzed above, to understand the influence of each factor on the customer's product selection decision, after evaluating the model, an additional step is needed: extracting the coefficients of each independent variable and printing them to gain the most comprehensive overview. Through the final results, one can examine the coefficients of each variable to understand their impact. Positive coefficients indicate a positive and significant influence on purchasing decisions, while negative coefficients suggest a negative impact on consumer retention and decision-making.

Thanks to its powerful and prominent capabilities, logistic regression has become a valuable tool in customer analysis and predicting their behavior, particularly in the fields of customer relationship management and marketing. Specifically, the benefits of applying this method to customer analysis include.

2.2.4. Benefits

Probability Prediction: Logistic regression allows predicting the probability of an event occurring, such as customer churn or growth. This provides detailed and flexible information about the likelihood of events in the future.

ROC Curve and AUC: This model is often used alongside the ROC (Receiver Operating Characteristic) curve and AUC (Area Under the Curve) to evaluate the model's quality. This helps assess both the accuracy and sensitivity of the model.

Impact Analysis: The coefficients of variables indicate their impact on the probability of an event. This aids in analyzing and understanding the effects of factors on customer behavior.

Handling Independent Variables: Logistic regression can handle both continuous independent variables and categorical variables flexibly. This facilitates the integration of various types of data conveniently.

Understandable and Visual: The output results of this method are often easy to understand and visualize, especially when applied to classify customers into groups or predict specific probabilities.

Adjustable Decision Threshold: The decision threshold can be adjusted to meet specific business requirements. This is crucial when balancing between accuracy and sensitivity considerations.

Easy Implementation and Integration: Logistic regression can be easily implemented in various programming languages and integrated into different data analysis systems.

Handling Overfitting: Logistic regression models tend to experience less overfitting compared to more complex models. This helps reduce the risk of the model not generalizing well to new data.

2.2.5. Limitations

While logistic regression is a powerful tool for customer analysis, it also has limitations that need to be considered when applied in specific situations. Here are some key limitations of logistic regression:

Linear Assumption: Logistic regression assumes a linear relationship between independent and dependent variables. This means the model may not be effective if the relationship is not truly linear.

Sensitivity to Noise and Outliers: Logistic regression can be sensitive to noisy data and outliers, which can significantly impact the model's coefficients and predictions.

Multicollinearity: If independent variables in the model have high levels of correlation (multicollinearity), estimating coefficients may become uncertain and challenging to interpret.

Potential for Overfitting: Logistic regression models can still be prone to overfitting, especially when there are many independent variables relative to the sample size.

Inability to Handle Nonlinear Phi Relationships: Logistic regression cannot directly handle nonlinear relationships. If the relationship between independent variables and probability is not linear, the model may not accurately reflect that relationship.

Inappropriateness for Modeling Complex Relationships: In some cases, the relationship between independent variables and probability may be too complex to be accurately described by a simple logistic regression model.

Threshold Decision Requirement: To predict the final class, a decision threshold must be chosen, and changes in the threshold can impact the accuracy of the model.

2.3. Measurement Metrics

2.3.1. Customer Lifetime Value (CLV)

2.3.1.1. Definition

Customer lifetime refers to the period from when a customer starts purchasing or using a company's products/services until they completely discontinue. The Customer Lifetime Value (CLV) is the value that a customer contributes to your company's products throughout their lifetime. (KC Karnes, 2023) Long-term, sustainable profit for businesses comes from loyal customers because the value of customer lifetime is high. Therefore, for a business to thrive, it needs many loyal customers as they bring in long-term and sustainable profits.

2.3.1.2. Cohort Analysis

Cohort Analysis is a behavioral analysis method used to examine the effectiveness of marketing strategies for a specific target segment related to marketing costs and customer lifetime value. The dataset is divided into correlated groups, known as "cohorts," before conducting the analysis. Alongside Retention rate and Churn rate, Cohort analysis provides insights into the customer's loyalty to the product over time, thereby accurately calculating CLV. (XPON Technologies Group, 2018)

Example: In digital marketing, a marketing team can use cohort analysis to track the effectiveness of various campaigns or marketing channels. They can group users by the month they first interacted with the brand and track their behavior over time. This can help the team identify trends in interaction or conversion and optimize their marketing efforts accordingly.

Acquisition Date	Users	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
Jan 25	1,098	100%	33.9%	23.5%	18.7%	15.9%	16.3%	14.2%	14.5%	13.3%	13.0%	12.1%
Jan 26	1,358	100%	31.1%	18.6%	14.3%	16.0%	14.9%	13.2%	12.9%	14.5%	11.3%	
Jan 27	1,257	100%	27.2%	19.6%	14.5%	12.9%	13.4%	13.0%	10.8%	11.4%		
Jan 28	1,587	100%	26.6%	17.9%	14.6%	14.8%	14.9%	13.7%	11.9%			
Jan 29	1,758	100%	26.2%	20.4%	16.9%	14.3%	12.7%	12.5%				
Jan 30	1,624	100%	26.4%	18.1%	13.7%	15.4%	11.8%					
Jan 31	1,541	100%	23.9%	19.6%	15.0%	14.8%						
Feb 01	868	100%	24.7%	16.9%	15.8%							
Feb 02	1,143	100%	25.8%	18.5%								
Feb 03	1,253	100%	24.1%									
All Users	**13,487**	**100%**	**27.0%**	**19.2%**	**15.4%**	**14.9%**	**14.0%**	**13.3%**	**12.5%**	**13.1%**	**12.2%**	**12.1%**

Figure 2: Cohort Analysis: Beginners Guide to Improving Retention (Cohort Analysis: Beginners Guide to Improving Retention, 2023)

Perform Cohort Analysis for CLV

Step 1: Set up Cohort Groups: Identify Cohort groups based on factors such as the date of the first interaction, acquisition channel, or device used.

Step 2: Collect Data: After establishing Cohort groups, track various metrics for each group. These metrics may include interaction-related indicators (session count, screen views), conversion metrics (goal completion, commercial transactions), as well as demographic and behavioral information (age, gender, interests).

Step 3: Data Analysis: Monitor changes in metrics over time, identify trends, and patterns in the data for each Cohort group.

Step 4: Optimize Strategy: Use detailed information from the cohort analysis to adjust marketing strategies and attract customers. Adjust marketing spending, test different channels or tactics, modify pricing, or offer products to optimize the performance of the strategy.

Cohort analysis allows businesses to gain insights into customer behavior and preferences over time, enabling strategic adjustments to improve overall performance and customer lifetime value.

2.3.1.3. CLV Calculation Formula in the Traditional Approach

$$CLV = \$M\left[\frac{r}{1+d-r}\right]$$

Where

- M: Annual profit rate per customer (revenue minus all costs divided by the total number of customers in the year).
- r: Annual retention rate, which is the number of customers at the end of the year minus the number of customers acquired in the year, all divided by the number of customers at the beginning of the year.
- d: Discount rate is the discount rate used to bring CLV to present value.

Limitations of the CLV Model

Does not measure changes in customer revenue and costs over time, making it unsuitable for a company pursuing customer share marketing goals.

Assumes that the retention rate (loyalty rate) is stable and does not change over time, which may not be the case as companies pursuing marketing share goals may experience increasing loyalty.

Does not apply discount rates to future customer revenue and costs, leading to an exaggerated lifetime value for the customer to the company.

2.3.1.4. CLV Calculation Formula Based on Initial Profit

$$CLV_{alternative} = \$M\left[\frac{1+d}{1+d-r}\right]$$

This formula incorporates an initial cash flow M, which is the certain amount received at the beginning of a customer relationship. Therefore, this replacement formula always has a higher value than the original formula. It is used to assess how much a business is willing to spend to acquire these customers. This modified formula allows businesses to consider the initial profit obtained from customers when calculating their lifetime value.

2.3.1.5. The Relationship Between Retention Rate and Customer Lifetime Value

Customer retention is a crucial driver for a company's financial success, reducing employee turnover can lead to increased profitability ranging from 25% to 85%. To measure the economic benefits of increasing the retention rate, there are three approaches:

Spreadsheet Model: Build a profit and cash flow calculation model based on the variation of the retention rate. The model will analyze "what if" scenarios to assess the benefits of increasing the retention rate. This helps the company measure the impact of retention on its finances.

Cash Flow Model: Develop a cash flow prediction model based on customer relationships and calculate CLV. The model calculates sensitivity assumptions to evaluate the economic benefits of increasing the retention rate. From the model, businesses can assess the impact of retention on customer and company value.

Direct Assumption Analysis on CLV Formula: Use the CLV formula with assumptions about profit rates and constant retention rates. Evaluate the economic benefits of increasing the retention rate through direct analysis. This provides a detailed insight into how changes in the retention rate affect the long-term value of customers.

2.3.2. Churn Rate

2.3.2.1. Definition

Churn Rate is the percentage of customers who stop using your company's products or services within a specified period. (Dataflo, n.d.)Therefore, if the Churn Rate is high, evaluating and improving the acquisition of new customers is crucial.

Figure 3: Churn Rate

The Churn Rate is calculated using the formula (*How to Easily Calculate Churn Model for Your Mobile App*, n.d.):

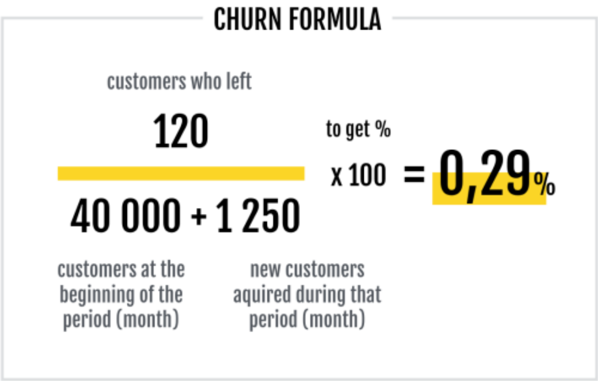


Figure 4: Churn Formula

Where:

- Total customers churned are those who stopped using the company's products or services within a specific month/quarter/year.
- Total customers represent the entire number of customers in that time period.

2.3.2.2. Role of Churn Rate

Evaluate Growth and Profitability: Churn Rate allows businesses to compare the number of new customers and those leaving. If the Churn Rate is high, it may lead to profit loss and negative growth.

Customer Behavior Analysis: Based on the Churn Rate, businesses can analyze why customers are leaving or canceling services. This provides valuable insights to improve products or services.

Identify Key Customers: Churn Rate helps assess the importance of each customer, determining key customers that the business needs to retain.

Business Overview: Churn Rate trends reflect the overall business situation and market dynamics. This helps businesses devise appropriate development strategies.

Evaluate Marketing Campaigns: Churn Rate can help assess the performance of advertising and marketing campaigns. If the churn rate decreases after a campaign, it may be considered an effective strategy.

CHAPTER 3: DATA PREPARATION

3.1. Collection data

Overview:

The Telco customer churn dataset encapsulates a snapshot of a fictitious telecommunications company's operations during the third quarter in California, engaging with a substantial customer base of 7043 individuals. This dataset intricately details customer demographics, shedding light on factors such as age, gender, income, and marital status, providing a nuanced understanding of the diverse clientele.

A pivotal aspect illuminated by this dataset is the phenomenon of customer churn, denoting instances where customers discontinue or terminate their association with the telco company. The dataset further introduces key metrics crucial for gauging customer interactions. The Satisfaction Score serves as a measure of customer contentment, likely derived from surveys or feedback mechanisms, providing insights into the overall customer experience.

Additionally, the Churn Score emerges as a predictive metric, assigning numerical values to customers based on various factors to anticipate the likelihood of churn. This aids the telco company in proactively identifying high-risk customers and implementing targeted retention strategies.

A notable inclusion is the Customer Lifetime Value (CLTV) Index, a metric estimating the total revenue expected from a customer throughout their entire association with the company. This index serves as a relative measure, offering valuable insights into the varying degrees of customer importance to the telco business.

In essence, this dataset encapsulates a rich tapestry of customer-related information, encompassing their profiles, satisfaction levels, likelihood of churning, and their estimated lifetime value to the telco company. Leveraging this dataset for analysis equips the telco company with the tools to make informed decisions regarding customer retention, satisfaction enhancement, and overall business strategy.

Size: 977.5 kB

Type: CSV.

Column: 21 columns (features). The "Churn" column is target.

Rows: Each row represents a customer, each column contains customer's attributes described on the column Metadata. The raw data contains 7043 rows (customers).

3.2. Exploratory Data Analysis (EDA)

Import the necessary libraries:

First, you need to install the required libraries such as Pandas, NumPy, Seaborn, and Matplotlib...

```
#import the required libraries
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.ticker as mtick
import matplotlib.pyplot as plt
%matplotlib inline
```

3.2.1. Statistic Analysis

```
[ ] telco_base_data.shape

    (7043, 21)

[ ] telco_base_data.columns.values

    array(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
           'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
           'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',
           'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract',
           'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges',
           'TotalCharges', 'Churn'], dtype=object)
```

From the above code, this dataframe has 7043 rows and 21 columns.

```
[ ] telco_base_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customerID        7043 non-null   object
 1   gender            7043 non-null   object
 2   SeniorCitizen     7043 non-null   int64
 3   Partner           7043 non-null   object
 4   Dependents        7043 non-null   object
 5   tenure            7043 non-null   int64
 6   PhoneService      7043 non-null   object
 7   MultipleLines     7043 non-null   object
 8   InternetService   7043 non-null   object
 9   OnlineSecurity    7043 non-null   object
 10  OnlineBackup      7043 non-null   object
 11  DeviceProtection  7043 non-null   object
 12  TechSupport       7043 non-null   object
 13  StreamingTV       7043 non-null   object
 14  StreamingMovies   7043 non-null   object
 15  Contract          7043 non-null   object
 16  PaperlessBilling  7043 non-null   object
 17  PaymentMethod     7043 non-null   object
 18  MonthlyCharges    7043 non-null   float64
 19  TotalCharges      7043 non-null   object
 20  Churn             7043 non-null   object
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
```

The columns are listed in the table with the name of each column, the number of non-null values for each column is displayed, and Dtype is the data type of each column. From the above result, we can see that some columns have missing data. Most columns are of the object type and have no null values.

```
# Check the descriptive statistics of numeric variables
telco_base_data.describe()
```

	SeniorCitizen	tenure	MonthlyCharges
count	7043.000000	7043.000000	7043.000000
mean	0.162147	32.371149	64.761692
std	0.368612	24.559481	30.090047
min	0.000000	0.000000	18.250000
25%	0.000000	9.000000	35.500000
50%	0.000000	29.000000	70.350000
75%	0.000000	55.000000	89.850000
max	1.000000	72.000000	118.750000

The analysis of numerical columns reveals that SeniorCitizen is a categorical variable, so the distribution of 25% - 50% - 75% is not applicable. 75% of customers have a tenure of less than 55 months. The average monthly fee is $64.76, while 25% of customers pay more than $89.85 per month.

The describe() function provides a concise summary of the dataset:

- Count: Displays the total number of rows.

- Mean: Displays the mean value.
- Std: Shows the standard deviation.
- Min: Displays the minimum value.
- 25%: First quartile or 25th percentile.
- 50%: Median or 50th percentile.
- 75%: Third quartile or 75th percentile.
- Max: Displays the maximum value.

In addition to the describe() function mentioned earlier, the group created a summary table that provides statistical information about columns with object type data (string or non-numeric) in the DataFrame.

```
telco_base_data.describe(include='object')
```

	customerID	gender	Partner	Dependents	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup
count	7043	7043	7043	7043	7043	7043	7043	7043	7043
unique	7043	2	2	2	2	3	3	3	3
top	7590-VHVEG	Male	No	No	Yes	No	Fiber optic	No	No
freq	1	3555	3641	4933	6361	3390	3096	3498	3088

Now, to view the data types of the columns, we can use dtypes.

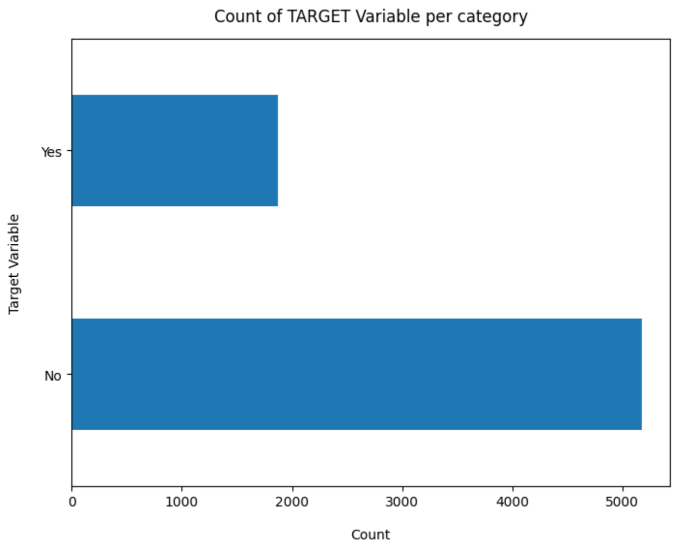


Figure 5: Count of Target Variable per category

Reviewing the variable of the "Churn" column, it can be observed that the no-churn rate is high, indicating that the business performance of Telco is quite good.

```
100*telco_base_data['Churn'].value_counts()/len(telco_base_data['Churn'])
```

```
No     73.463013
Yes    26.536987
Name: Churn, dtype: float64
```

```
telco_base_data['Churn'].value_counts()
```

```
No     5174
Yes    1869
Name: Churn, dtype: int64
```

The data is highly imbalanced, with a ratio of 73:27. Therefore, the team will analyze the data with separate target values to obtain detailed information.

3.2.2. Data Cleaning

The initial step involves checking for null values in the data file. The result returned indicates that there are no recorded null cases.

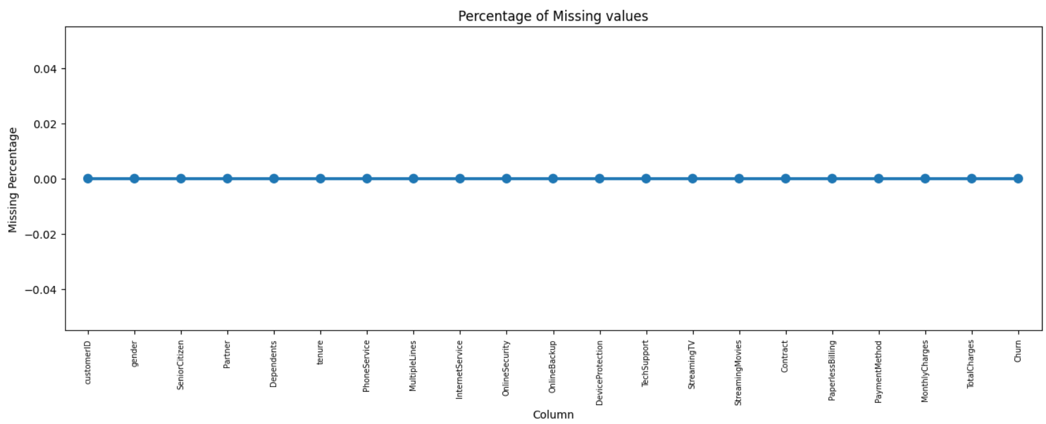


Figure 6: Percentage of Missing values

```
# show the column with null values
telco_base_data.isnull().sum()
```

```
customerID          0
gender              0
SeniorCitizen       0
Partner             0
Dependents          0
tenure              0
PhoneService        0
MultipleLines       0
InternetService     0
OnlineSecurity      0
OnlineBackup        0
DeviceProtection    0
TechSupport         0
StreamingTV         0
StreamingMovies     0
Contract            0
PaperlessBilling    0
PaymentMethod       0
MonthlyCharges      0
TotalCharges        0
Churn               0
dtype: int64
```

As a general rule, for features with fewer missing values, regression can be used to predict missing values or fill in the average value of existing values, depending on the feature. For features with a very high number of missing values, it is better to drop those columns as they provide very little detailed information for analysis.

The TotalCharges column should represent an amount of money. Converting it to a numeric type reveals 11 null values. Following the missing value handling rules, because the percentage of these records compared to the total dataset is very low, they can be safely ignored for further processing.

```
telco_data.TotalCharges = pd.to_numeric(telco_data.TotalCharges, errors='coerce')
telco_data.isnull().sum()
```

```
customerID          0
gender              0
SeniorCitizen       0
Partner             0
Dependents          0
tenure              0
PhoneService        0
MultipleLines       0
InternetService     0
OnlineSecurity      0
OnlineBackup        0
DeviceProtection    0
TechSupport         0
StreamingTV         0
StreamingMovies     0
Contract            0
PaperlessBilling    0
PaymentMethod       0
MonthlyCharges      0
TotalCharges       11
Churn               0
dtype: int64
```

Customers are segmented based on tenure (the duration of the customer's relationship). For example: tenure < 12 months (1-12), tenure 1 to 2 years (13-24), etc. The grouping includes 72 months and is divided into 6 segments.

```
# Group the tenure in bins of 12 months
labels = ["{0} - {1}".format(i, i + 11) for i in range(1, 72, 12)]
telco_data['tenure_group'] = pd.cut(telco_data.tenure, range(1, 80, 12), right=False, labels=labels)
telco_data['tenure_group'].value_counts()
```

```
1 - 12     2175
61 - 72    1407
13 - 24    1024
25 - 36     832
49 - 60     832
37 - 48     762
Name: tenure_group, dtype: int64
```

Finally, delete unnecessary columns to proceed with processing the data, including 'customerID','tenure'.

```
#drop column customerID and tenure
telco_data.drop(columns= ['customerID','tenure'], axis=1, inplace=True)
telco_data.head()
```

3.2.3. Data Exploration

3.2.3.1. Distribution plot of individual predictor factors based on Churn

Based on gender, the difference rate between the two is almost negligible, indicating that gender does not significantly influence churn outcomes.
In terms of age, the noticeable difference is evident between the elderly and the young. The no-churn rate is quite high for the younger age group compared to the older age group. The number of older individuals participating in the business's product usage is also relatively low.

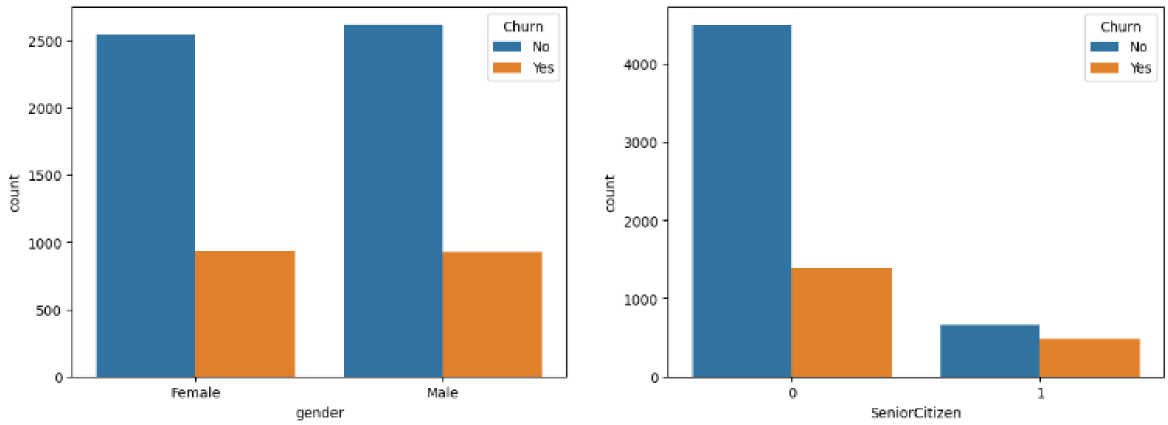


Figure 7: Distribution plot of Gender and Citizen based on Churn

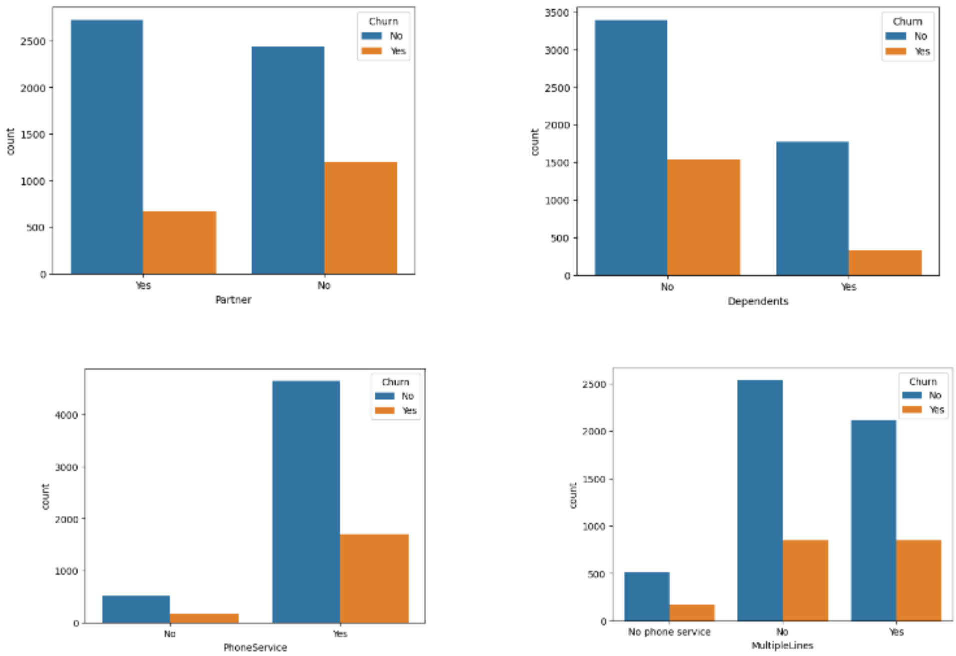


Figure 8. Distribution plot of Partner, Dependents, PhoneServive, MultipleLines based on Churn

The next four values indicate the influence of the variables "Dependents," "PhoneService," and "MultipleLines," which show significant differences. Meanwhile, "Partner" does not exhibit such differences. To further explore the correlation between these value columns, it will be discussed in the correlation section later on.

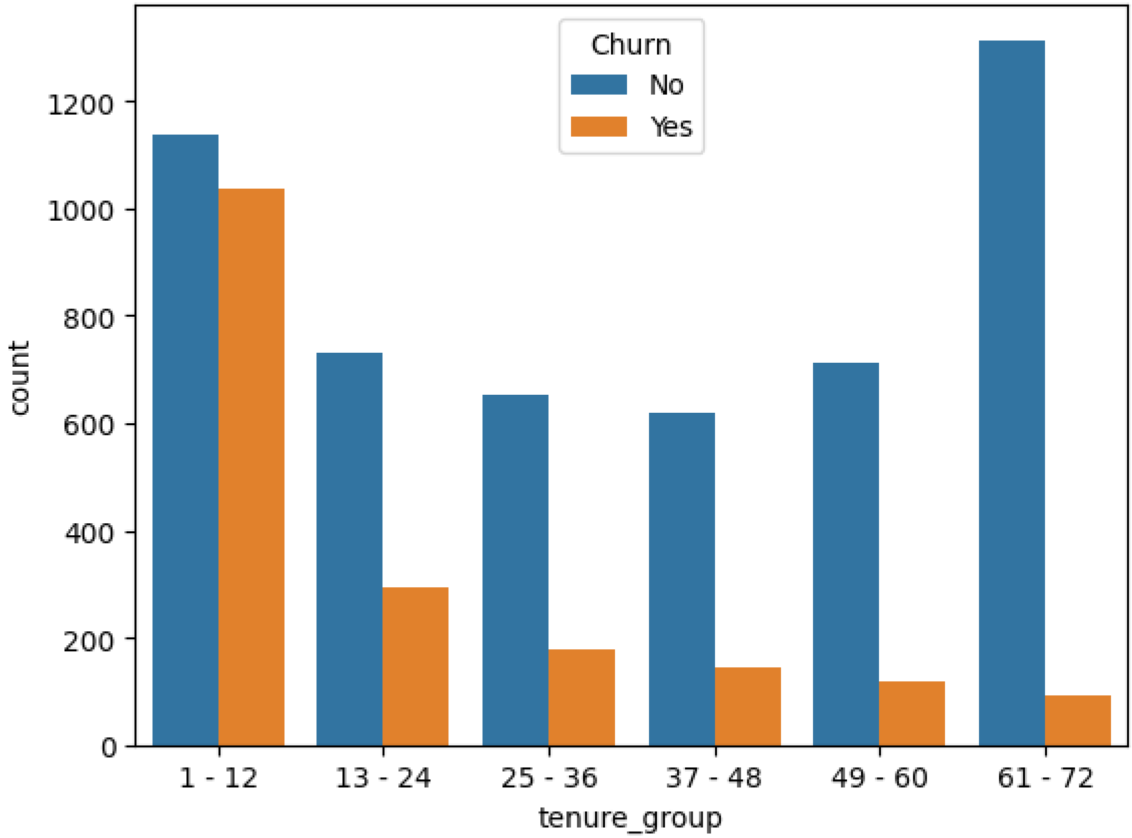


Figure 9. Distribution plot of Tenure group based on Churn

At this point, the variable "tenure" shows a clear trend, indicating that the longer the tenure, the lower the likelihood of customer churn, and the retention rate increases. New customer groups also have a relatively high retention rate, but they also exhibit a high churn rate.

3.2.3.2. The Relationship Between Monthly Charges and Total Charges

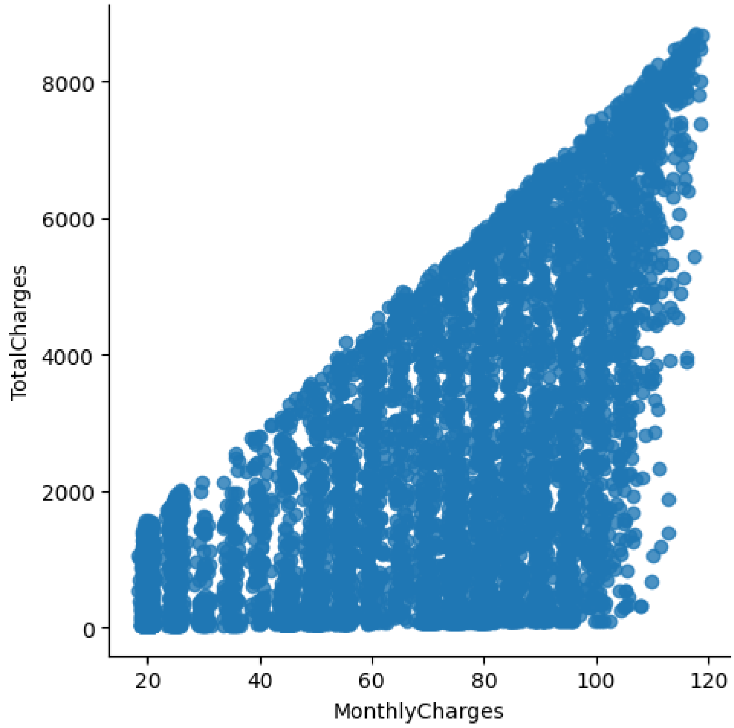


Figure 10. Scatter plot Between Monthly Charges and Total Charges

As predicted, the total charges increase with the monthly charges. To delve deeper into customer churn, the group calculates the Churn Rate based on Monthly Charges and Total Charges.

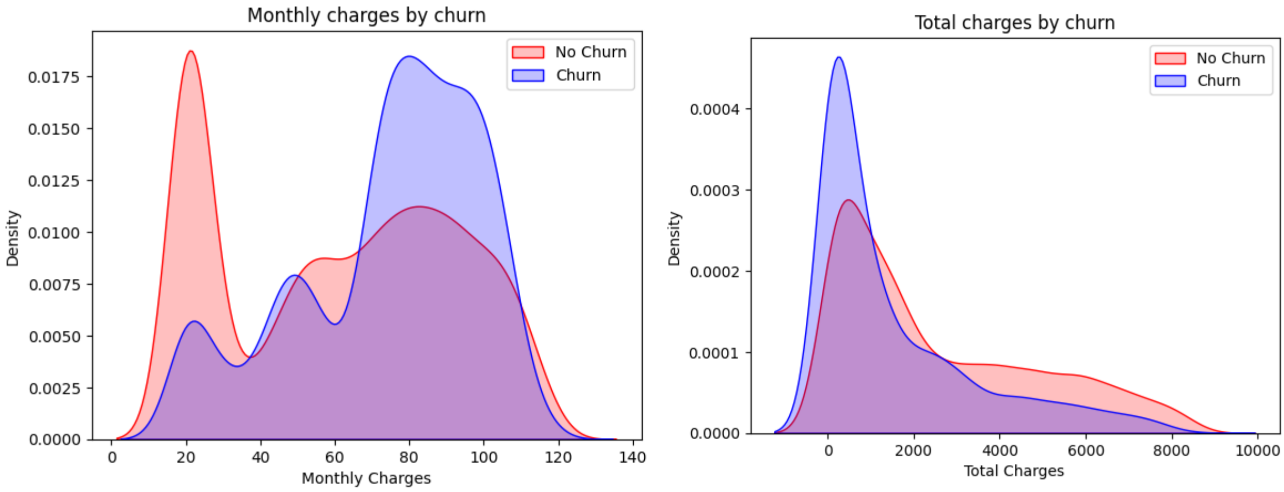


Figure 11. Relationship Between Monthly Charges and Total Charges

Churn is higher with lower Total Charges. However, when combining detailed information about the three parameters Tenure, Monthly Charges, and Total Charges, it can be observed that the monthly charges are higher for new customers (lower tenure), leading to lower total charges. Therefore, all three factors demonstrate that the monthly charges for investing in a new customer are high, and their recorded total charges will be low due to their relatively high churn rate.

3.2.3.3. Correlation

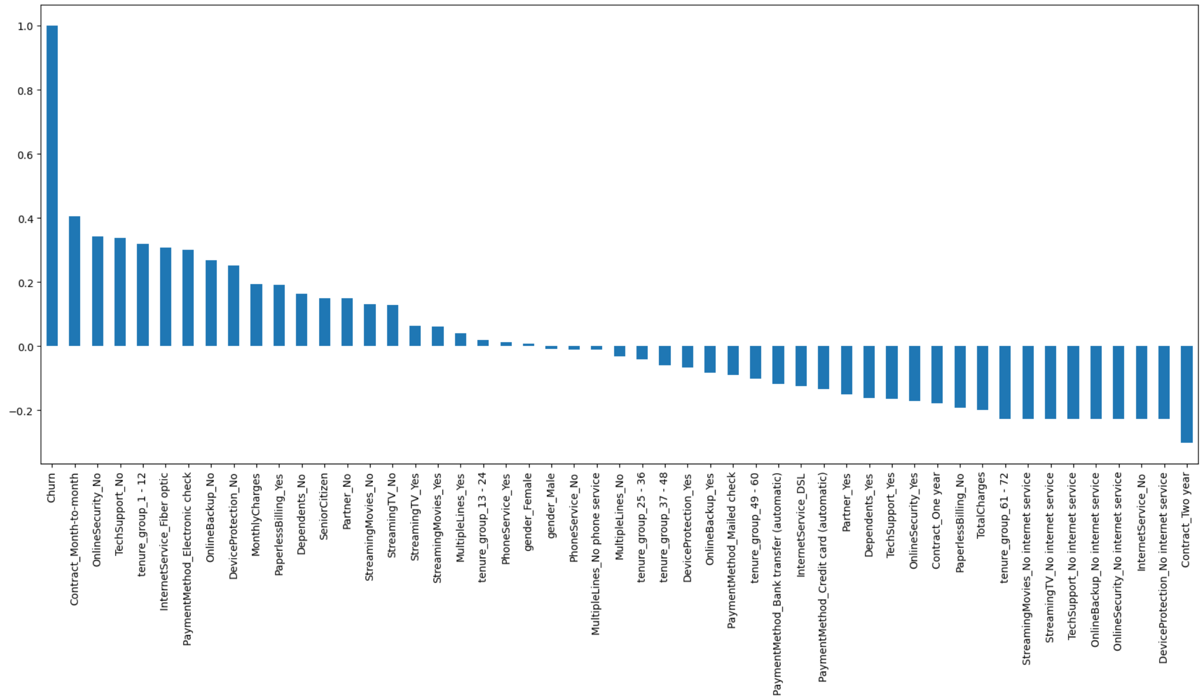


Figure 12. Relationships between all predictive factors and 'Churn'

It can be observed that:

- High Churn: Month-to-Month contract, No online security, No tech Support, First-year subscription, Fiber Optics Internet. Some reasons for customer churn include those who subscribe to month-to-month contracts, lack online security, lack technical support, are in their first year of subscription, and use Fiber Optics Internet.
- Low Churn: The low churn rate is associated with long-term contracts, not subscribing to internet services, and customers who have been with the company for more than 5 years.

Factors such as Gender, Phone Service Availability, and MultipleLines do not have a significant impact on Churn.

This is also clearly illustrated in the heatmap below.

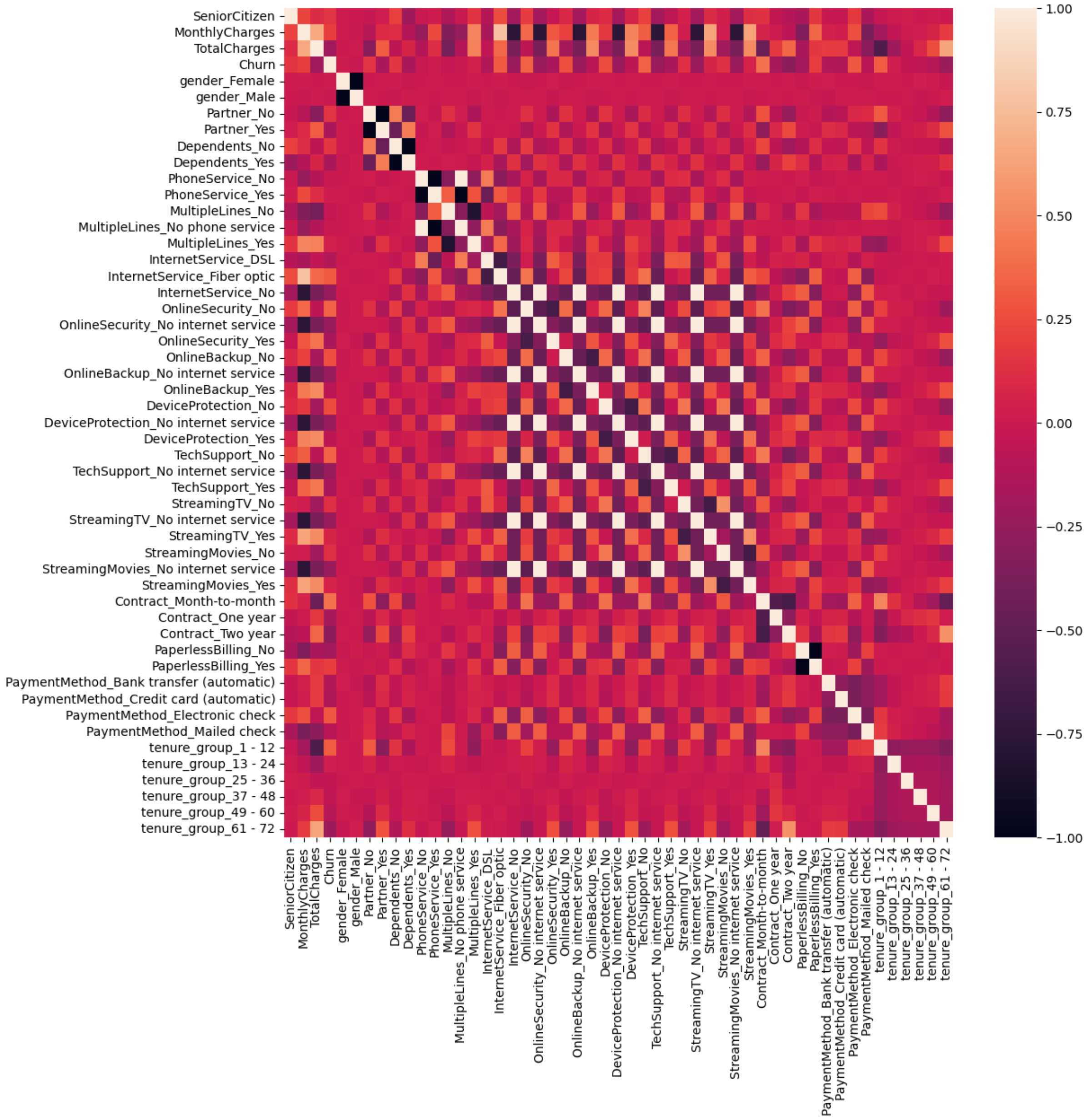
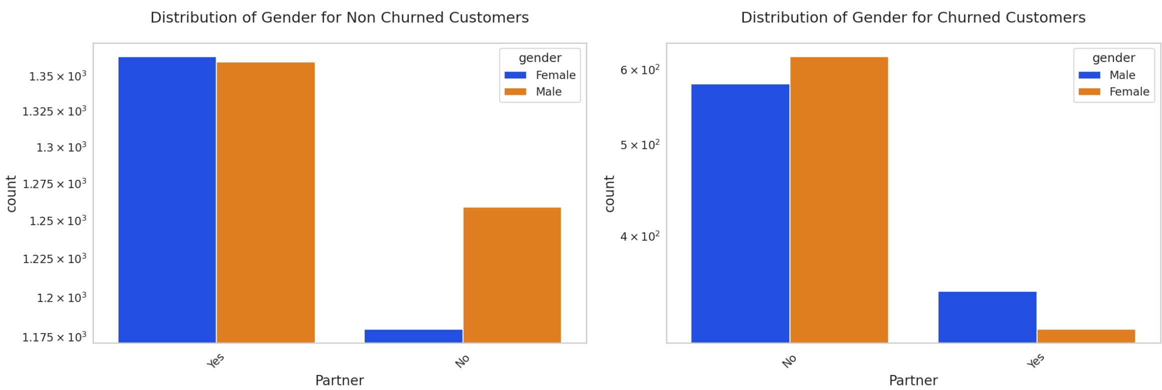


Figure 13. Heatmap relationships between all predictive factors and 'Churn'

3.3. Bivariate Analysis



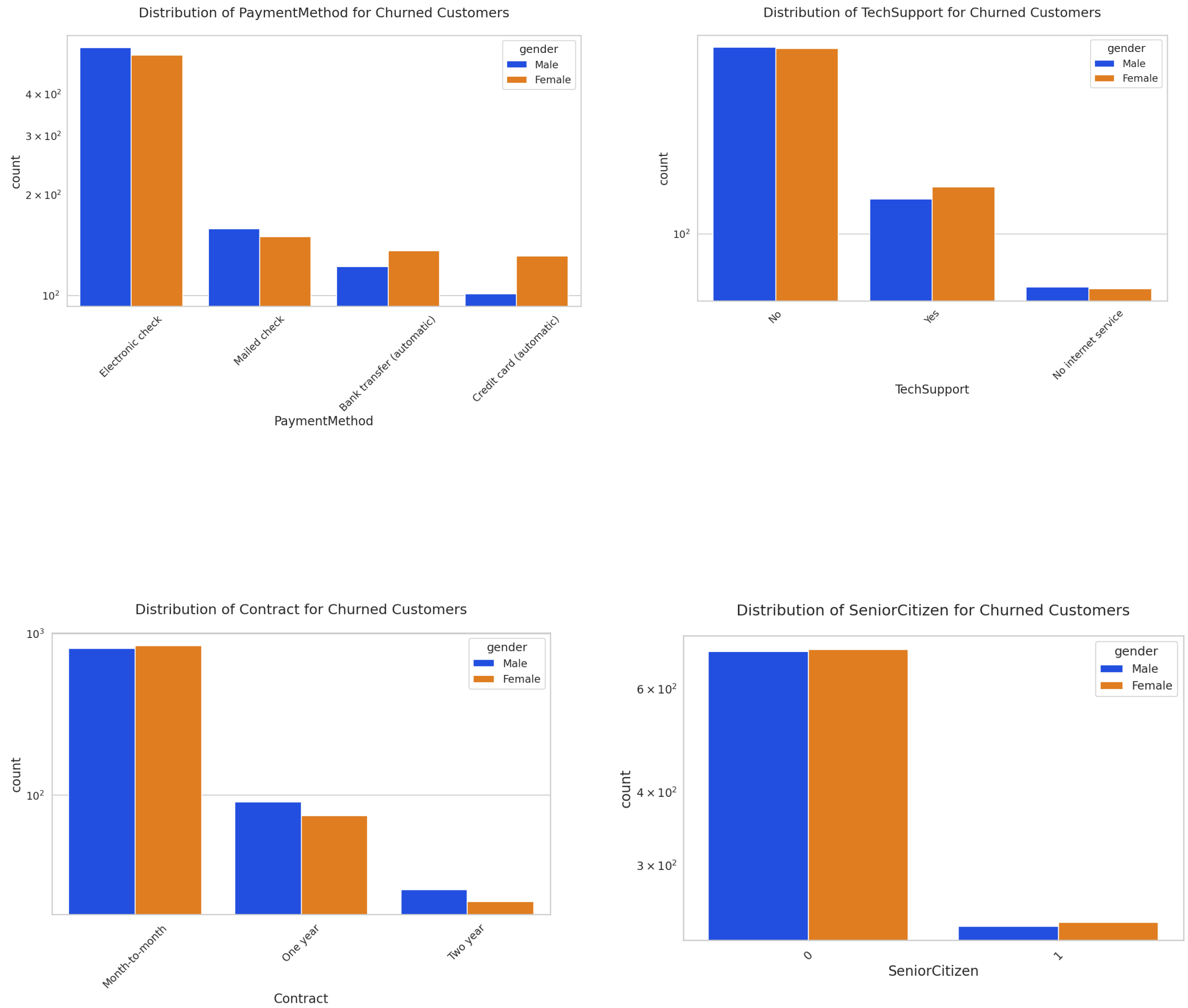


Figure 16 . Bivariate Visualization Chart base on Gender

The team has some insights into this dataset, including objects with a high likelihood of churning:

- Electronic check is the payment method with the highest churn rate. The business may consider eliminating this payment method.
- In terms of contract type, customers with month-to-month contracts are more likely to churn, possibly because they have no contract terms and can easily switch providers.
- Customers without online security and technical support are more likely to churn. The company should enhance security features in subscription packages and provide better technical support.
- Senior citizens are also more likely to churn.

CHAPTER 4: CUSTOMER ANALYTICS

4.1. Model Prediction

4.1.1. Decision Tree Classifier

Step 1: Create a Decision Tree Classifier model (Imbalanced Data)

Step 2: Train the Decision Tree model on the original training data

Step 3: Predict the target variable for the test set

Step 4: Calculate the accuracy of the model on the test set

```
model_dt.score(x_test,y_test)
```

```
0.7803837953091685
```

Step 5: Summarize model evaluations on the imbalanced test set

```
print(classification_report(y_test, y_pred, labels=[0,1]))
```

```
              precision    recall  f1-score   support

           0       0.85      0.86      0.85      1034
           1       0.59      0.57      0.58       373

    accuracy                           0.78      1407
   macro avg       0.72      0.71      0.72      1407
weighted avg       0.78      0.78      0.78      1407
```

We can see that the accuracy is quite low because it is an imbalanced dataset. We should not consider Accuracy as our sole metric to measure the model, as Accuracy will be low in imbalanced datasets. Therefore, we need to check precision, recall, and F1 score for the minority class. Clearly, precision, recall, and F1 score are too low for class 1, i.e., churn customers.

To address this, let's apply the SMOTEENN (UpSampling + ENN) technique.

Step 6: Apply SMOTEENN to handle the imbalanced dataset

Step 7: Split the data into training and testing sets from the balanced data

Step 8: Create a new Decision Tree model for the balanced data

Step 9: Run the new model

```
model_dt_smote.fit(xr_train,yr_train)
yr_predict = model_dt_smote.predict(xr_test)
model_score_r = model_dt_smote.score(xr_test, yr_test)
print(model_score_r)
print(metrics.classification_report(yr_test, yr_predict))
```

```
0.9443016281062554
              precision    recall  f1-score   support

           0       0.93      0.95      0.94       507
           1       0.96      0.94      0.95       660

    accuracy                           0.94      1167
   macro avg       0.94      0.94      0.94      1167
weighted avg       0.94      0.94      0.94      1167
```

Step 10: Display the confusion matrix of the model on the balanced test set.

Now we can see better results, with an accuracy of 94%, and recall, precision, and F1 score showing good performance for the minority class.

Reference results: DecisionTreeVisualization.png

4.1.2. Random Forest Classifier

Similarly to the Decision Tree Classifier, the Random Forest is implemented following these steps:

Step 1: Create and Train the Random Forest Classifier Model

```
model_rf=RandomForestClassifier(n_estimators=100, criterion='gini', random_state = 100,max_depth=6, min_samples_leaf=8)
model_rf.fit(x_train,y_train)
```

```
RandomForestClassifier
RandomForestClassifier(max_depth=6, min_samples_leaf=8, random_state=100)
```

Step 2: Predict and Evaluate the Random Forest Classifier Model

```
y_pred=model_rf.predict(x_test)
model_rf.score(x_test,y_test)
```

```
0.8038379530916845
```

```
print(classification_report(y_test, y_pred, labels=[0,1]))
```

```
              precision    recall  f1-score   support

           0       0.83      0.92      0.87      1030
           1       0.70      0.47      0.56       377

    accuracy                           0.80      1407
   macro avg       0.76      0.70      0.72      1407
weighted avg       0.79      0.80      0.79      1407
```

Step 3: Use SMOTEENN for Imbalanced Data

Step 4: Train the Random Forest Model on the Augmented Data

Step 5: Run the New Model

```
print(model_score_r1)
print(metrics.classification_report(yr_test1, yr_predict1))
```

```
0.9316823228010248
              precision    recall  f1-score   support

           0       0.95      0.89      0.92       508
           1       0.92      0.97      0.94       663

    accuracy                           0.93      1171
   macro avg       0.93      0.93      0.93      1171
weighted avg       0.93      0.93      0.93      1171
```

With the Random Forest Classifier, the results are quite good, actually better than the Decision Tree.

Reference results: randomforest.png

4.2. Visualize with Power BI

To visualize data, in this section the team used the Power BI tool to build a dashboard. In this dashboard there will be 5 pages, including 1 dashboard containing a number of interactive elements to analyze customer churn data. There are "Summary", "Churn Reasons", "Customer Detail" and "Ask a question" tabs represented by stylized buttons on a digital waveform background with raised dots connected by link lines navigate to the report pages, giving it a look and feel like network.

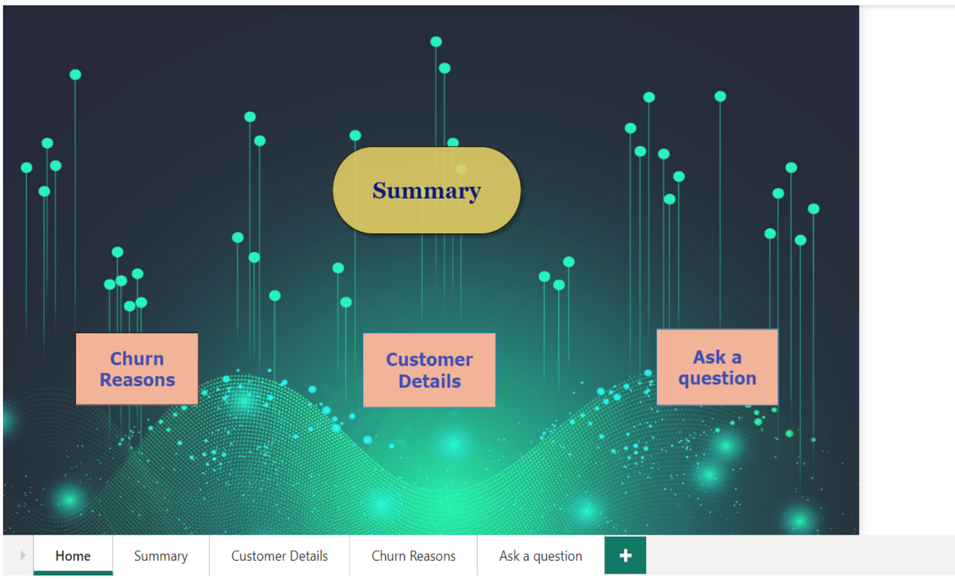


Figure 17. Main screen overview of reports

Next is a dashboard containing statistical data used to analyze customer risk, specifically identifying customers who are likely to churn (cancel their services). Here's a breakdown of the statistics presented:

Looking at the dashboard in the Client Profile section, this dataset has 7043 total of individuals represented. This number is the starting point for assessing risk and making data-driven decisions.

- Demographics:
    - The pie chart indicates a balanced gender distribution, with almost equal percentage of men and women: 49% male (approx. 3.45K) and 50% female (approx. 3.56K).
    - The bar graph shows the distribution of time customers spend with the service. New customers (0-20 months) and long-time customers (41-60 months) account for a large number, which may reflect the company's marketing and customer retention strategy.
    - The number of senior citizen is small at 1,142 people. Along with the number of customers who are partners of the service, there is an even distribution with about 3.6 thousand customers who are not partners and 3.4 thousand who are partners using the company's services.
- Phone Service:
    - Most customers use telephone services. This metric provides an overview of the most basic services the company offers.
    - Cyber Security and Technical Support Services have lower subscription rates, possibly due to cost or customer perception of the value of these services.
    - With Internet Service, a significant number of churners had fiber optic service (0.46K), fewer had DSL, and some had no internet.
- Contract Terms:

- There is a clear trend of customers choosing monthly contracts over long-term commitments. This may be due to flexibility or uncertainty about continued use of the service.
- Average monthly charges are $64.76, while average total charges are $2,283.

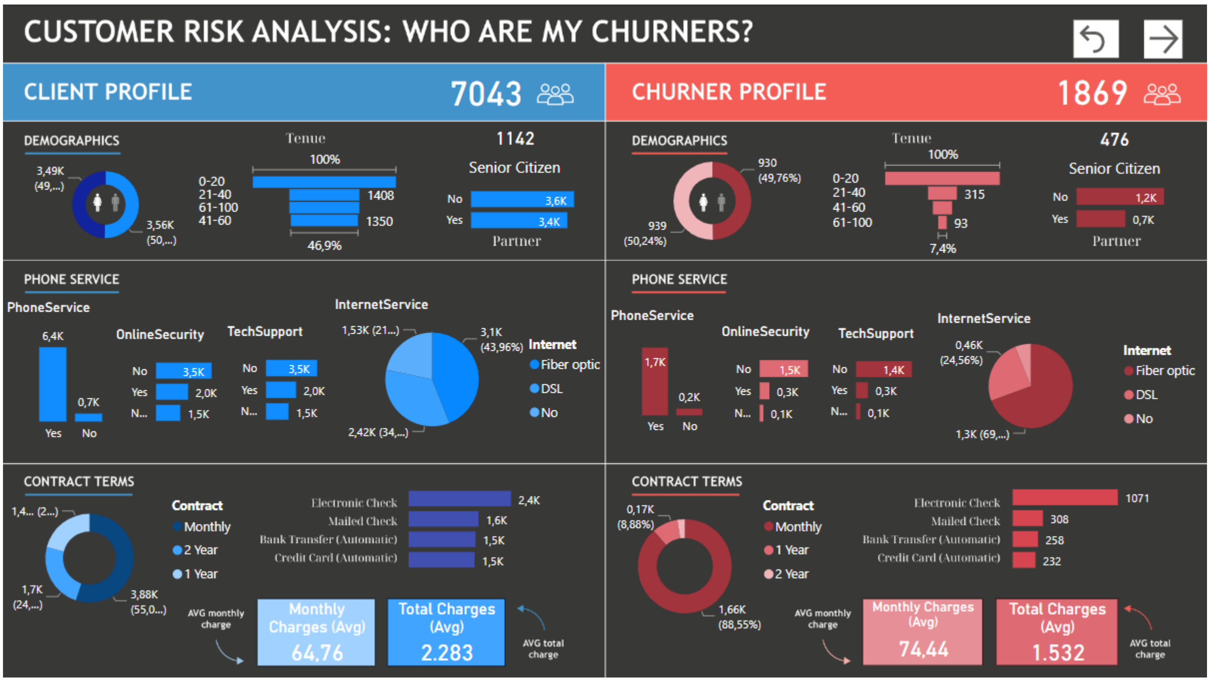


Figure 18. Analyze customer churn rate

In the Churner Profile section, the total number of churners are 1869 individuals who have discontinued service. This number can be an indicator of how satisfied or dissatisfied customers are with products or services.

- Demographics:
  - Gender split is nearly even: 49.76% male (930) and 50.24% female (939). The gender breakdown in this group is similar to the overall customer base, but there is a higher proportion of the elderly and

those with partners, possibly because this customer group has different needs and priorities.

- The tenure distribution indicates a higher churn rate among newer customers, particularly in the 0-20 and 21-40 month categories. This could imply that customers are leaving early in their lifecycle, possibly due to dissatisfaction or competitive offers.
- The senior citizen and partner indicators show that fewer senior citizens and individuals with partners are churning. This might reflect higher service satisfaction or less inclination to change among these groups, or possibly better retention strategies aimed at these segments.

- Phone Service:
  - The majority of churners (1.7K) had a phone service, which suggests that having a phone service alone is not enough to retain customers. The company might need to look into the quality of the service or the competitiveness of the pricing.
  - A notable number of churners did not subscribe to online security and tech support. This could indicate that customers who do not perceive additional value from these services might be more likely to leave.
  - With Internet Service, a significant portion of churners had fiber optic internet service. If fiber optic service is associated with higher costs, this could suggest a price sensitivity issue among churners. Alternatively, it might indicate issues with the fiber optic service itself, such as reliability or speed, leading to dissatisfaction.
- Contract Terms:
  - The overwhelming majority of churners were on monthly contracts. This is significant because it implies a possible correlation between non-commitment and churn. Customers on

short-term contracts may feel less tied to the service and can leave more easily.

- The average monthly charges are higher for churners at $74.44, compared to the average customer profile. This could be a crucial insight, suggesting that higher prices might be driving customers away.
- The average total charges for churners are lower than the average total charges for all customers, which could mean that churners are leaving before accruing the same level of charges as longer-term customers.

In Customer Detail, we can see specific information about each customer's metrics. The interface allows interaction via Customer ID to change information options for each customer. In Personal Details there will be information such as customer ID, gender, age. The Other Details section indicates whether the customer is a senior citizen or not and the time since the customer started using the service. The Phone Service section will show information about each specific service type that the customer has registered to use. Finally, there is the service registration contract term as well as payment methods. On the left are 3 statistical indicators showing the customer's likelihood of leaving, the total amount of fees paid and assessing the customer's churn risk status.

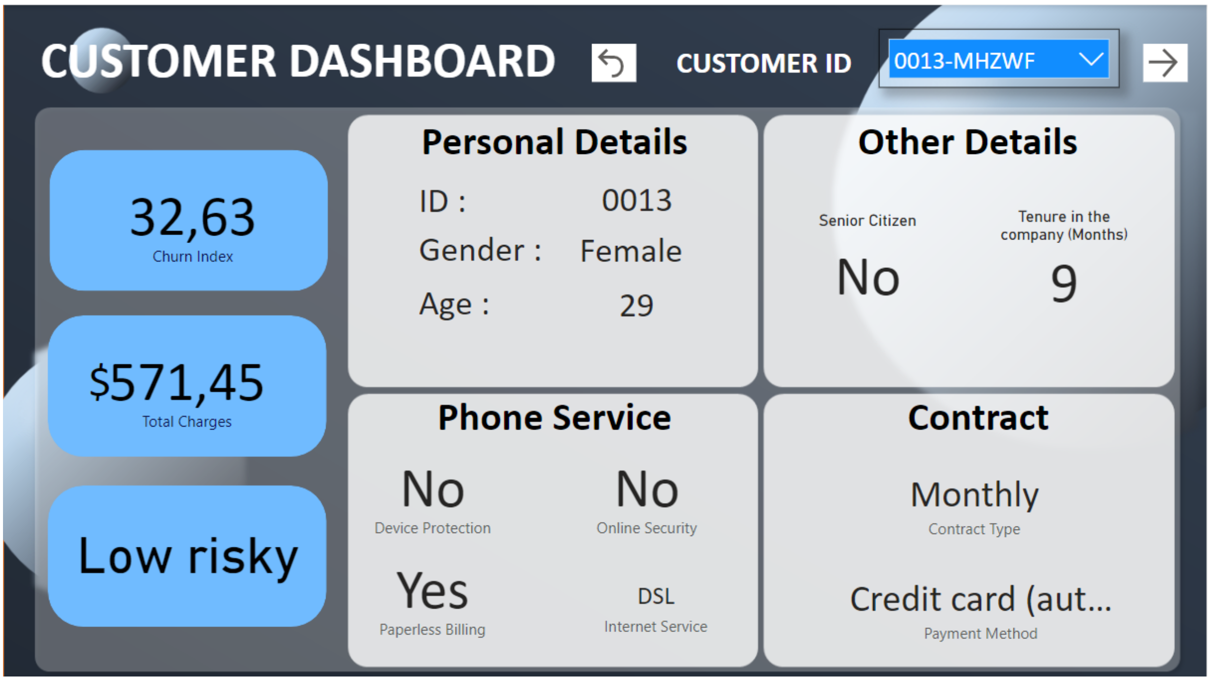


Figure 19 . Information and specific assessment of each customer's situation

Churn Reason's overall statistics table includes customer churn risk indicators. The Risky Customers section represents the customers who are considered at risk of churning. Average Risk Score likely represents an average of several risk metrics calculated for each customer, with higher scores possibly indicating a higher likelihood of churn. Total Charges of Risky Customers is the sum of charges from customers who are considered risky, which indicates the potential revenue at risk if these customers churn. Total Charges shows the total revenue from all customers, providing a context for the amount attributed to risky customers.

The visualizations also provide additional insights about Prediction by Risk Group. These categories help to segment the customer base by perceived risk levels, which can be crucial for targeted retention strategies. Average Total Charges by Risk Group shows the average revenue contribution by each risk group, highlighting the financial impact of each segment.

From these metrics, we can infer that a significant portion of the revenue is associated with customers who are at some level of churn risk. The company might prioritize retention efforts starting with the 'High risky' and 'Risky' groups to mitigate potential revenue loss.

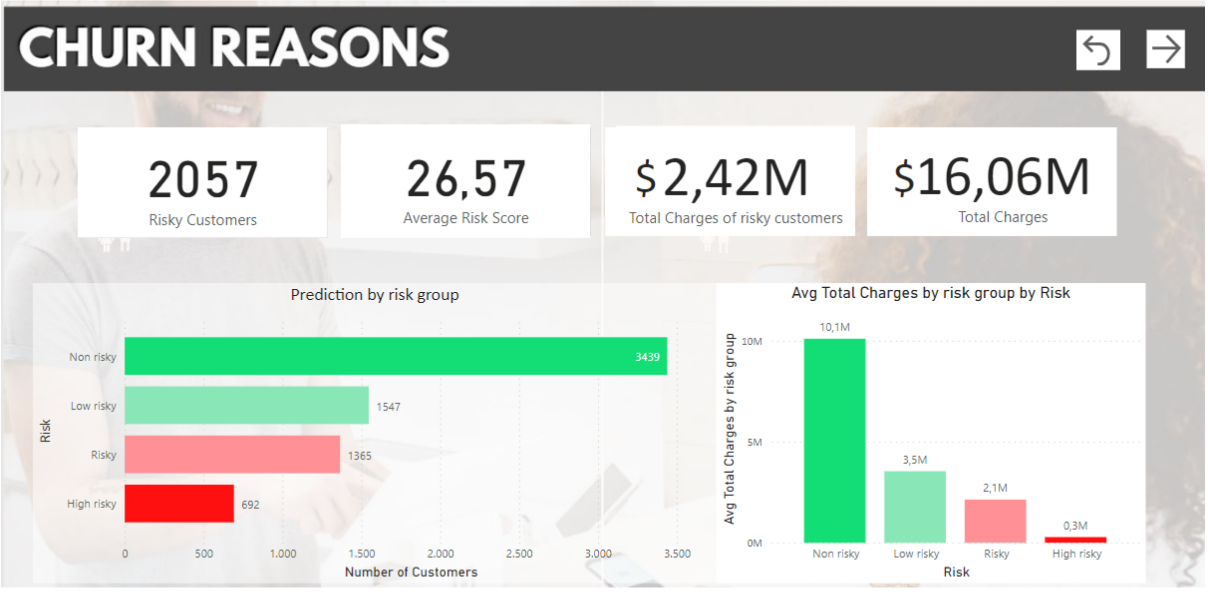


Figure 20. Overview of customer churn reasons

The final page is a simple user interface with various suggested queries related to customer churn data analysis. The queries are designed to help a user interact with a dataset to extract specific insights about customer churn.

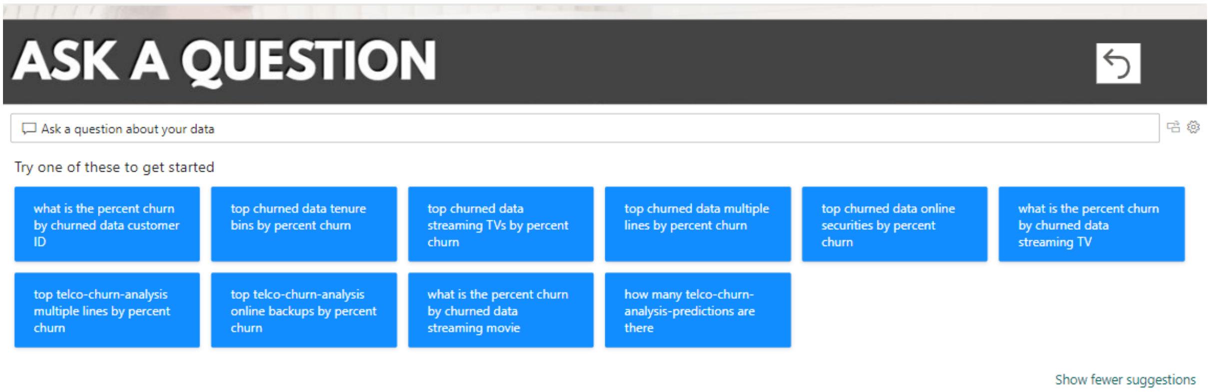


Figure 21. Statistics on customer churn issues

These queries aim to identify patterns and trends in churn data, such as churn rates associated with different customer IDs, contract types, lease terms, and types of services provided. usage, including device protection and internet services. The queries also suggest focusing on identifying the top factors or services associated with the highest churn rates. Here is an example of the suggested queries.

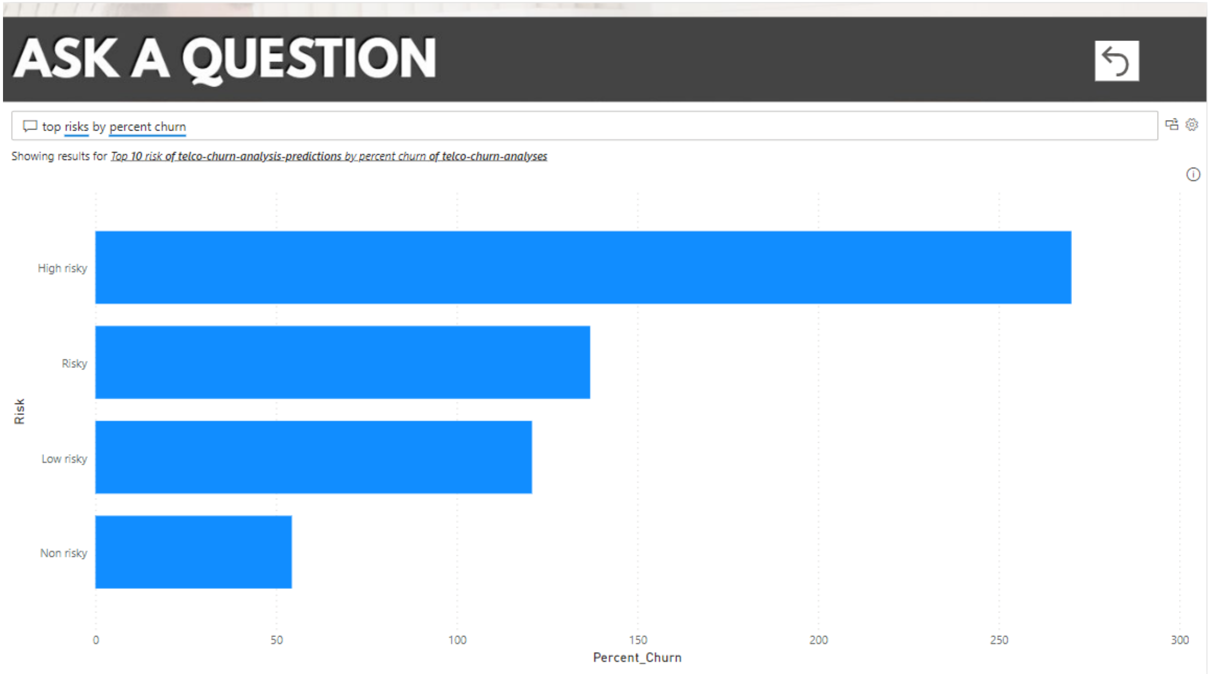


Figure 22. Top risk categories by percent churn

CHAPTER 5: SOLUTION

The telecommunications industry is marked by cutthroat competition and ever-evolving consumer expectations, necessitating a deep understanding of customer churn and effective strategies for customer retention. This essay delves into four key strategies based on a comprehensive analysis of a telco customer dataset. The findings reveal distinctive patterns among churners, underscoring the need for targeted approaches to retain customers.

5.1. Segmentation and Personalization:

One clear trend from the data is the higher churn rate among new customers, especially those in the 0-20 and 21-40 months tenure groups. To counter this, telco companies can implement segmentation strategies. For instance, create personalized marketing campaigns targeting new customers, offering exclusive welcome packages or discounts. Tailoring communication channels based on customer demographics and preferences can significantly enhance the customer experience and foster a sense of individualized care.

Practical Suggestion: Establish a customer segmentation framework that considers factors such as tenure, age, and usage patterns. Develop targeted promotional offers for each segment, ensuring that new customers feel valued and engaged.

5.2. Enhanced Online Security and Support Services:

A significant number of churners did not sign up for online security and technology support. To address this, telco companies should invest in robust online security measures and provide comprehensive technology support. Develop educational campaigns to highlight the importance of online security, and offer user-friendly resources for customers to easily access technology support when needed.

Practical Suggestion: Launch an awareness campaign about the benefits of online security and technology support. Provide step-by-step guides, webinars, and 24/7 customer support to assist customers in utilizing these services effectively.

5.3.Contract Term Optimization:

The data indicates that a majority of customers leaving the telco service were on month-to-month contracts. To optimize contract terms, telco companies can incentivize longer commitments. Offer discounts, exclusive perks, or additional services for customers opting for longer-term contracts. Clearly communicate the advantages of extended contracts, emphasizing cost savings and added value.

Practical Suggestion: Introduce loyalty programs that offer escalating benefits for customers with longer contract durations. Provide incentives such as free upgrades, priority customer service, or loyalty points that can be redeemed for exclusive rewards.

5.4.Value-Driven Pricing Models:

The analysis reveals that the average monthly fee for a churner is higher at $74.44 than the average customer profile. To address this, telco companies should reassess their pricing models to ensure they align with customer expectations and offer compelling value propositions. Implement flexible pricing options, bundled services, and periodic promotions to create a perception of value for money.

Practical Suggestion: Conduct customer surveys to understand their perception of service value and pricing. Based on feedback, refine pricing models to offer tiered plans, allowing customers to choose plans that align with their usage patterns and budget constraints.

In conclusion, the telco industry stands to benefit significantly from a nuanced understanding of customer churn patterns and the implementation of tailored retention strategies. By adopting segmentation and personalization, enhancing online security and support services, optimizing contract terms, and embracing value-driven pricing models, telco companies can create a customer-centric approach that builds lasting loyalty.

Practical implementation of these strategies involves not only a careful analysis of customer data but also ongoing communication and adaptation to evolving customer needs. As the telco landscape continues to evolve, companies that prioritize customer retention through strategic initiatives will be better positioned for sustained success and growth in a highly competitive market.

CHAPTER 6: CONCLUSION

This study on "Telco Customer Churn" has yielded dramatical insights into the intricated dynamics of customer retention in the telecommunications industry. Through the development and application of a predictive model utilizing advanced data analysis and machine learning techniques, we have successfully identified customers at high risk of churn. While this project report has made substantial contributions to the understanding of customer churn in the telecommunications industry, it is not without restrictions. This research was conducted on a specific dataset, and as such, the findings may not be universally applicable to all telecom sectors or regions. Future research could expand on this work by considering a more diverse range of data sources, cultural contexts, and industry environments. In addition, as the telecommunications industry and customer behaviors continue to evolve, particularly with the advancement of technology and changing social dynamics, ongoing research will be essential. Future studies should aim to continuously refine the predictive models, incorporate real-time data, and consider emerging trends and technologies. In conclusion, this study has not only advanced our understanding of the factors driving customer churn but has also provided actionable strategies for telecommunications businesses. The insights gained from this research promise to help businesses not only mitigate the risks associated with customer churn but also enhance their overall customer relationship management and strategic positioning in the competitive market landscape.

REFERENCES

1. *Cohort Analysis: Beginners Guide to Improving Retention*. (2023, November 29). CleverTap. Retrieved December 30, 2023, from https://clevertap.com/blog/cohort-analysis/
2. *Customer behavior analysis: What you need to know | Binarbase*. (2023, September 12). Binary Base. Retrieved December 30, 2023, from https://www.binarbase.com/article/customer-behavior-analysis-what-you-need-to-know
3. *Customer Churn – Logistic Regression with R - DataScienceCentral.com*. (2017, April 13). Data Science Central. Retrieved December 30, 2023, from https://www.datasciencecentral.com/customer-churn-logistic-regression-with-r/
4. Dataflo. (n.d.). *What is Churn Rate and how to calculate it?* Dataflo. Retrieved December 30, 2023, from https://www.dataflo.io/metricbase/churn-rate
5. *How to easily calculate churn models for your mobile app*. (n.d.). AppAgent. Retrieved December 30, 2023, from https://appagent.com/blog/how-to-easily-calculate-churn-model-for-your/
6. KC Karnes. (2023, November 29). *What Is It and How to Calculate - Customer Lifetime Value: What Is It and How to Calculate*. CleverTap. Retrieved December 30, 2023, from https://clevertap.com/blog/customer-lifetime-value/

7. Murphy, T., & Kiwak, K. (2023, 5 18). *customer analytics (customer data analytics)*. Business Analytics. https://www.techtarget.com/searchbusinessanalytics/definition/customer-analytics
8. Oates, D. (n.d.). *How to build a customer profile for effective marketing*. Experian. Retrieved December 30, 2023, from https://www.experian.co.uk/blogs/latest-thinking/marketing-solutions/profiling-customers/
9. William & Mary. (2023, March 7). *The Importance of Analytics in Digital Marketing*. Online Business Programs | W&M Raymond A. Mason. Retrieved December 30, 2023, from https://online.mason.wm.edu/blog/data-analytics-in-digital-marketing
10. XPON Technologies Group. (2018, March 29). *Measuring Customer Lifetime Value*. XPON Technologies. Retrieved December 30, 2023, from https://xpon.ai/resources/how-to-use-cohort-analysis-to-measure-customer-lifetime-value/