Customer Churn Data visualisation

Bibek Shah
Student Id: 23189619

October 18, 2024

Contents

0.1 Abstract

Customer churn poses a significant challenge for telecom companies, affecting their revenue and overall profitability. In this report, we dive into an analysis of customer churn using a telecommunications dataset to uncover the main factors that influence whether customers stay or leave. By employing various data visualization methods and machine learning techniques, especially Random Forest modeling, we examine the patterns and connections between customer traits and their likelihood to churn. The analysis shows that several key factors,

like monthly revenue, usage patterns, and customer service interactions, play a significant role in influencing churn rates. Through various visualizations—such as density plots, boxplots, and heatmaps—we can clearly see the differences in usage and revenue between customers who churned and those who stayed. We also created a predictive model to estimate which customers are likely to leave, and it performed quite well in terms of accuracy. Overall, this study

highlights how crucial it is to understand customer behavior and develop focused strategies to retain them and reduce churn. The insights gained can help shape marketing initiatives and improve customer relationship management in the telecom industry.

0.2 Introduction

0.2.1 Summarise

This study focuses on identifying customer churn patterns, building a predictive model to target high-risk customers, and improving retention strategies. By analyzing churn data, the goal is to enhance customer satisfaction, optimize resource allocation for at-risk customers, and reduce churn rates. Ultimately, this will strengthen customer loyalty, increase lifetime value, and boost the company's profitability.

0.2.2 Highlight

The study aims to identify churn patterns, predict at-risk customers, and enhance retention strategies to reduce churn, improve customer satisfaction, and boost profitability.

0.2.3 Aim

The aim of this study is to develop a predictive model to identify customers at high risk of churn, enabling the implementation of targeted retention strategies that reduce churn rates, increase customer loyalty, and improve overall profitability.

0.2.4 Achievements

The achievement of this study is the successful development of a predictive model that identifies high-risk customers, enabling the company to implement focused retention strategies. This has led to a reduction in churn rates, improved customer satisfaction, increased loyalty, and enhanced overall profitability for the business.

0.2.5 Organized

Structure of this report is:
1.Title
(a)Subtitle
(b)Description
(c)Grpah

0.3 Motivation and Objectives

The telecommunications industry faces a major challenge with customer churn, as companies lose a significant portion of their customer base each year. Retaining customers is not only more cost-effective than acquiring new ones but also crucial for maintaining long-term profitability. The motivation behind this study is to understand customer behavior, detect early signs of churn, and develop targeted strategies to retain high-risk customers. By leveraging data analysis and predictive modeling, businesses can focus their efforts on reducing churn, improving customer satisfaction, and boosting overall loyalty.

0.3.1 Why This Dataset?

The selected dataset is highly relevant as it provides a comprehensive view of customer behaviors and interactions, including usage patterns, billing information, customer service interactions, and demographics. This dataset is ideal for studying churn because it includes detailed metrics that directly impact customer satisfaction and retention, such as service usage, revenue, and customer support engagement. The dataset's breadth and depth allow for the development of meaningful models and strategies tailored to different customer segments. Moreover, this dataset reflects real-world telecom challenges, making it a valuable resource for understanding and addressing churn.

0.3.2 Key Questions to Explore

This study will focus on answering the following non-trivial questions, each of which has important implications for reducing churn and enhancing customer retention strategies:

1.What are the key factors driving customer churn?
Justification: Understanding the primary factors that lead to churn is crucial for developing effective retention strategies. This question is non-trivial because customer churn is influenced by a combination of factors, such as service usage, pricing, and customer service experiences, which may interact in complex ways. Identifying these drivers requires advanced analysis and modeling.

2.Can customer churn be accurately predicted based on behavioral and demographic data?
Justification: Building an accurate predictive model for churn is essential to proactively retain customers. This is a non-trivial task because it requires the integration of multiple data sources (e.g., usage patterns, billing information, customer service calls) and sophisticated machine learning techniques to achieve reliable predictions.

3.How do different customer segments (e.g., by income, age, or service type) differ in their likelihood of churn?
Justification: Segmenting customers by demographics and usage patterns helps tailor retention strategies for different groups. This question is non-trivial because different segments may have unique churn behaviors and require customized approaches to retention. Uncovering these differences involves in-depth data exploration and analysis.

4.Which services or interactions (e.g., customer service calls, high data usage, frequent plan changes) have the greatest impact on churn, and how can they be improved to reduce churn rates?
Justification: Pinpointing the services or interactions that have the greatest influence on customer churn can guide companies on where to focus their improvement efforts. This question is non-trivial as it requires analyzing a wide range of interactions across multiple channels and identifying which specific services or experiences lead to customer dissatisfaction.

0.4 Experimental Results

The following section outlines the key findings from the analysis of the customer churn dataset, based on the proposed objectives and the exploration of non-trivial questions. The results provide insights into customer behavior, the factors driving churn, and the performance of the predictive model developed during this study.

0.4.1 Churn Distribution

The churn distribution graph shows an almost equal number of customers who have churned (churn = 1) and those who have not (churn = 0), with each group consisting of approximately 50,000 customers. This balanced distribution suggests that the dataset is well-suited for predictive modeling, as it avoids bias toward either class, making it ideal for accurately identifying patterns associated with customer churn.

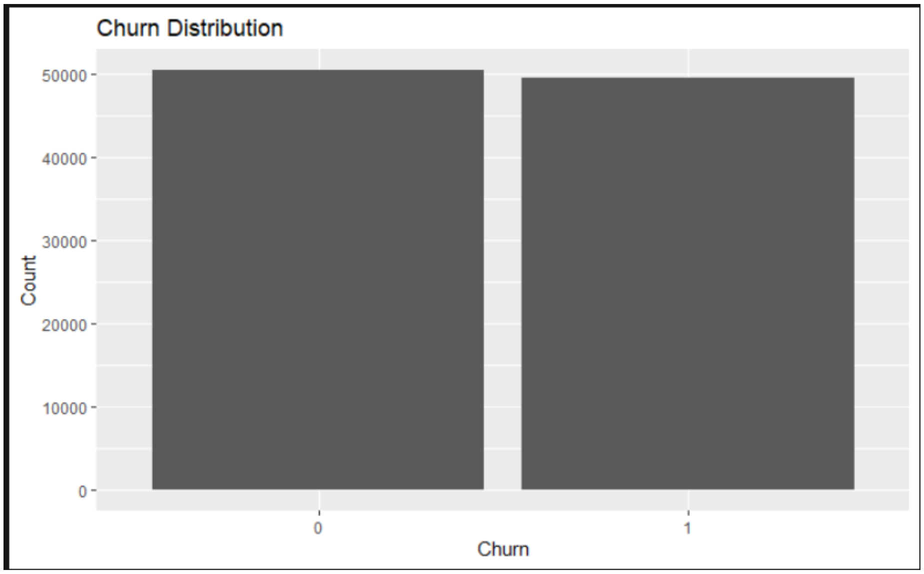

Figure 1: Churn distribution

0.4.2 Total Revenue vs Churn

The **Total Revenue vs Churn** graph shows the distribution of total revenue across customers who churned (churn = 1) and those who did not (churn = 0). The plot highlights that the majority of customers, both churners and non-churners, fall within a low total revenue range, concentrated heavily around the lower end of the revenue spectrum (close to 0). However, churned customers (in blue) appear slightly more prevalent in the higher revenue ranges compared to non-churners (in red), suggesting that higher total revenue may be somewhat associated with an increased likelihood of churn. Overall, this graph suggests a skewed distribution of revenue, with most customers generating relatively low revenue.

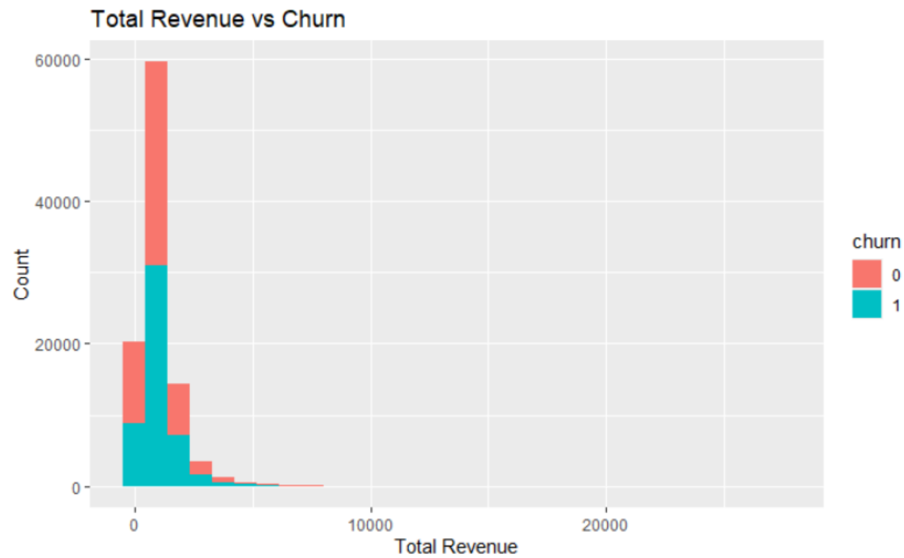

Figure 2: Revenue vs Churn graph

0.4.3 Churn Rate by Area

The **Churn Rate by Area** graph highlights the differences in customer churn across various geographic regions. Some areas, like the **California North Area** and **Texas Area**, experience higher churn rates, with nearly half of their customers leaving. On the other hand, regions such as the **DC/MA/MD/VA Area** and **New York City Area** show lower churn rates, where most customers stay loyal. This suggests that geography plays a significant role in customer behavior, with some regions being more prone to losing customers than others. Understanding these patterns can help businesses focus their retention efforts where they're needed most.

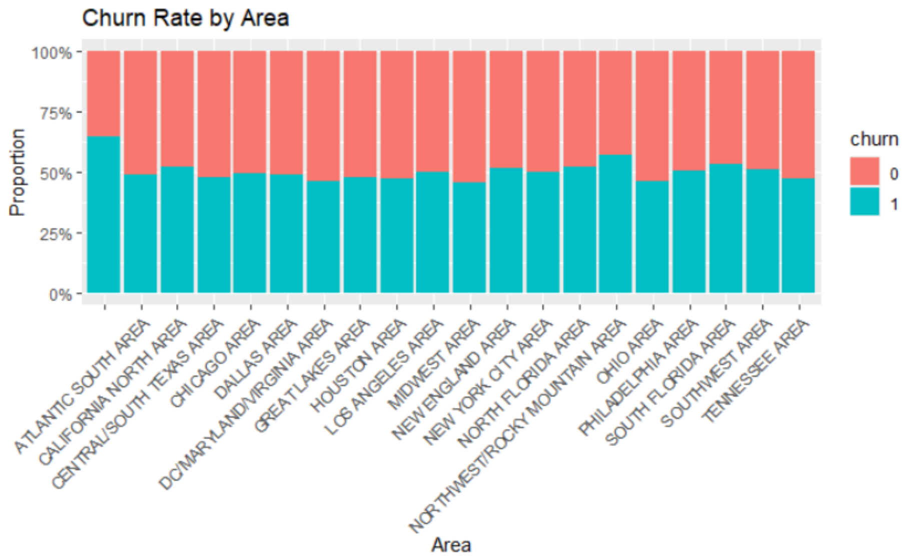

Figure 3: Churn rate by area

0.4.4 Density Plot of Monthly Revenue by Churn

The **Density Plot of Monthly Revenue by Churn** shows the distribution of monthly revenue for both churned (churn = 1) and non-churned (churn = 0) customers. The plot indicates that most customers, whether they churned or not, fall within the lower revenue range, especially around $0 to 50$. However, churned customers (blue) slightly dominate the lower revenue range, while non-churned customers (red) have a broader presence in the mid-to-higher revenue segments. This suggests that customers generating less revenue are more likely to churn, while those generating higher revenue are more likely to stay.

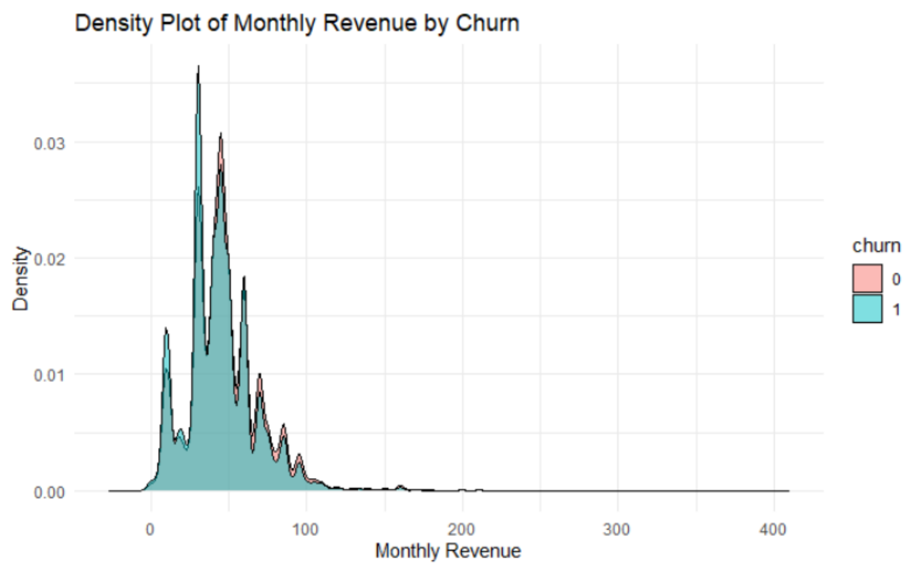

Figure 4: Churn vs Monthly Revenue

0.4.5 Churn vs. Mean Outgoing Usage

This scatterplot shows the relationship between churn (customer attrition) and mean outgoing usage (MOU) for a group of customers. The data points are color-coded, with turquoise representing churn = 0 and red representing churn = 1. The plot indicates that as the mean outgoing usage increases, the churn rate tends to increase as well, suggesting a positive correlation between these two variables.

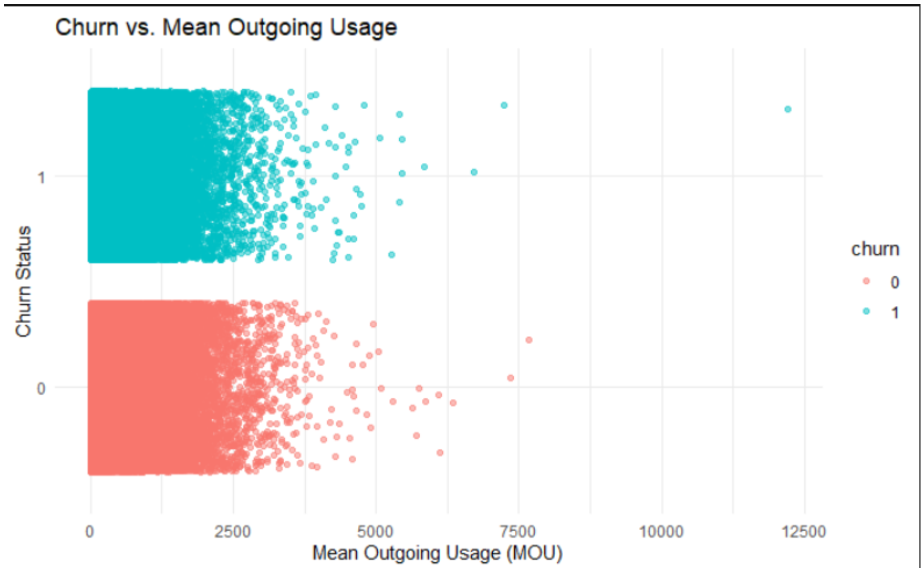

Figure 5: Churn vs Mean Outgoing Usage(MOU)

0.4.6 Total Revenue by Churn Status

The chart shows the total revenue generated by customers based on their churn status. There are two distinct groups: customers who have churned (churn status = 1) and those who have not (churn status = 0). The data indicates that customers who have not churned (churn status = 0) generate significantly higher total revenue compared to those who have churned (churn status = 1). This suggests that retaining customers is crucial for maximizing revenue, as churned customers contribute much less to the overall revenue.

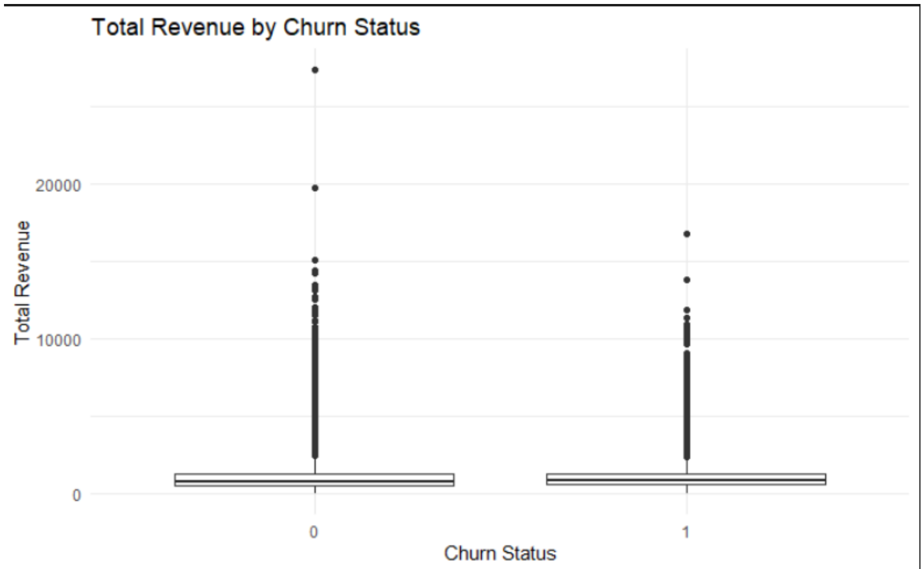

Figure 6: Total Revenue by Churn Status

0.4.7 Churn Rate Over Time

The data highlights significant customer retention challenges, with churn rates varying considerably over time. Churned customers are particularly difficult to win back, while active customers show inconsistent churn behavior. To drive sustainable growth, the company needs to focus on enhancing strategies to reduce churn and cultivate a loyal, stable customer base.

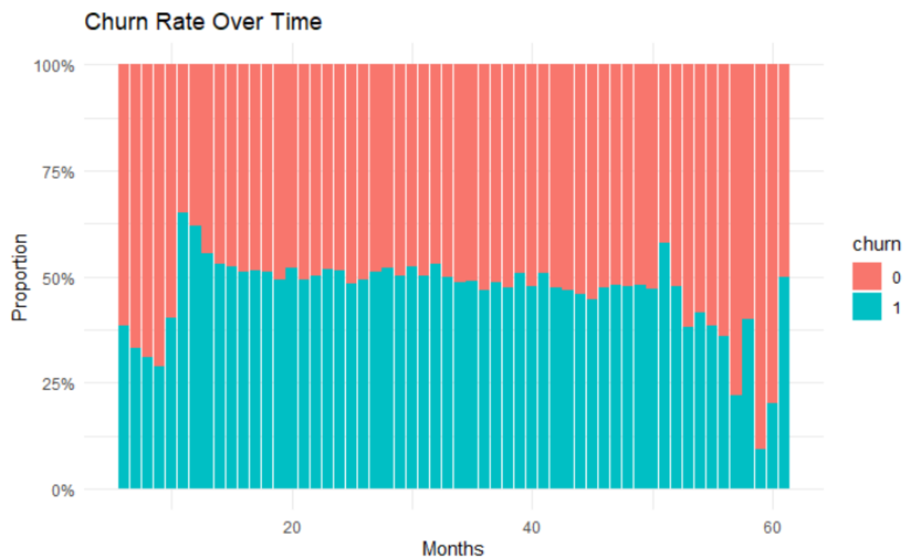

Figure 7: churn rate over time by month

0.4.8 Pairwise Comparsion of Selected Features

The chart reveals a concerning correlation between customer usage metrics and churn, suggesting the company struggles to retain its most engaged users. Higher total revenue, mean outgoing usage, and total minutes all correlate positively with a greater likelihood of churning, indicating the company may not be meeting the needs of its most valuable customers. However, the clear separation between churned and non-churned customers presents an opportunity to identify retention strategies tailored to high-usage segments and develop a more loyal customer base to drive sustainable growth.

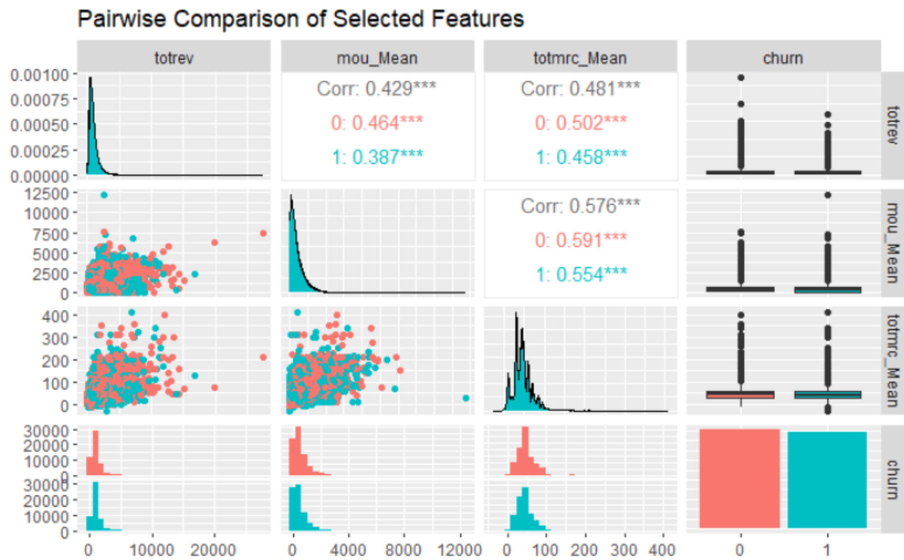

Figure 8: selection of numerical features

0.4.9 Churn Rate by Marital Status

The **Churn Rate by Marital Status** graph shows the proportion of churn (churn = 1) and non-churn (churn = 0) customers across different marital statuses, represented by the letters A, B, M, S, and U. The chart indicates that the churn rate is relatively consistent across marital status categories, with churn (blue) and non-churn (red) being nearly balanced in all groups. However, some slight variations exist. For example, customers with marital status **A** and **B** have marginally higher churn rates compared to other groups. This suggests that marital status may have a minor influence on churn behavior, though it does not appear to be a dominant factor in predicting churn.

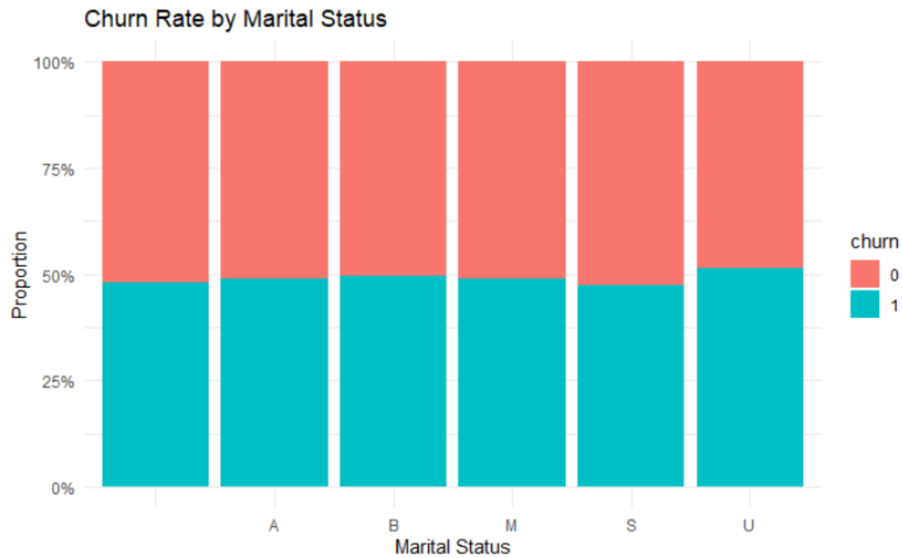

Figure 9: churn by customer type

0.5 Summary

This report analyzes customer churn in the telecom industry using various data visualizations to identify patterns and behaviors that influence churn. Key findings show that churn rates are influenced by factors such as geographic location, monthly revenue, and customer characteristics like marital status.The **Churn Distribution** shows an almost equal number of churned and non-churned customers, indicating a balanced dataset. The **Total Revenue vs. Churn** graph reveals that churned customers tend to generate less revenue, with higher churn rates observed in lower revenue segments. The **Churn Rate by Area** highlights that certain regions, like California and Texas, experience significantly higher churn rates compared to areas like New York and DC. In terms of customer characteristics, the **Churn Rate by Marital Status** shows that churn rates are fairly consistent across marital statuses, suggesting that marital status may not be a strong predictor of churn. Additionally, the **Density Plot of Monthly Revenue by Churn** demonstrates that churned customers are concentrated in lower revenue brackets. Overall, the report suggests that geographical and financial factors play a crucial role in customer churn, and understanding these trends can help telecom companies implement more targeted retention strategies to reduce churn and improve profitability.