Data Analysis Report of Telecom Customer Churn

Prepared By :

GURUR AŞER

gururaser@hotmail.com

19th August 2023

Introduction:

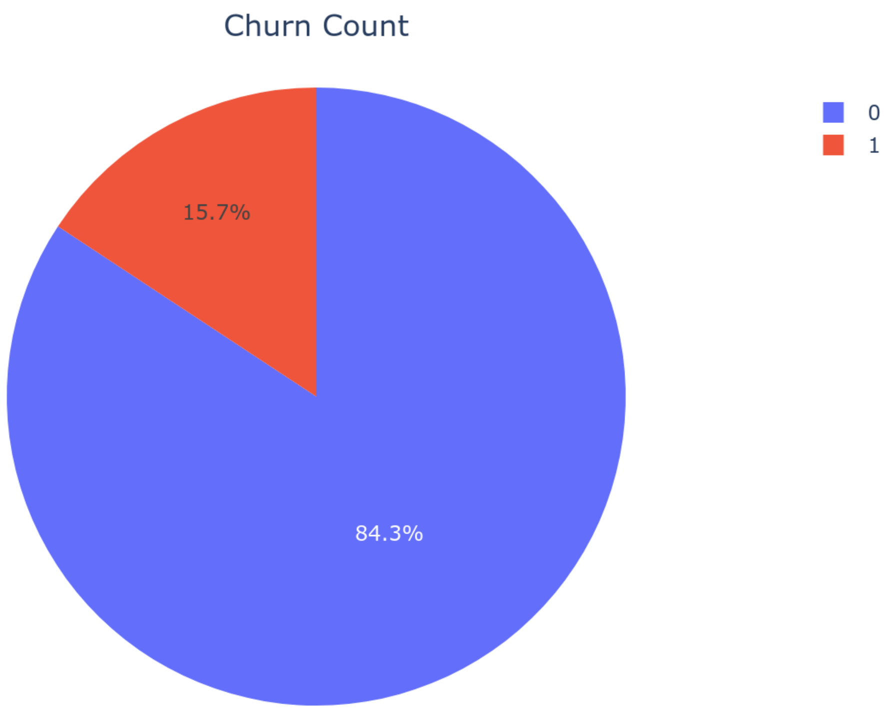
This report presents the results of an analysis of data collected over a 12-month period from an Iranian telecom company's database. The dataset includes information on 3150 customers, with each row representing a unique customer and 13 columns containing various attributes, such as call failures, frequency of SMS, number of complaints, subscription length, age group, charge amount, type of service, seconds of use, status, frequency of use, and customer value. The aim of the analysis was to investigate the factors that contribute to customer churn in the telecom sector and to predict churn.

Methodology:

The dataset used in this report was obtained from data sources called Kaggle. The data analysis tools used included the Python programming language, DataCamp Workspace, and Python and visualization libraries such as pandas, Plotly, NumPy, Seaborn, and Matplotlib. Python code was written to clean, process, visualize, and analyze the data.

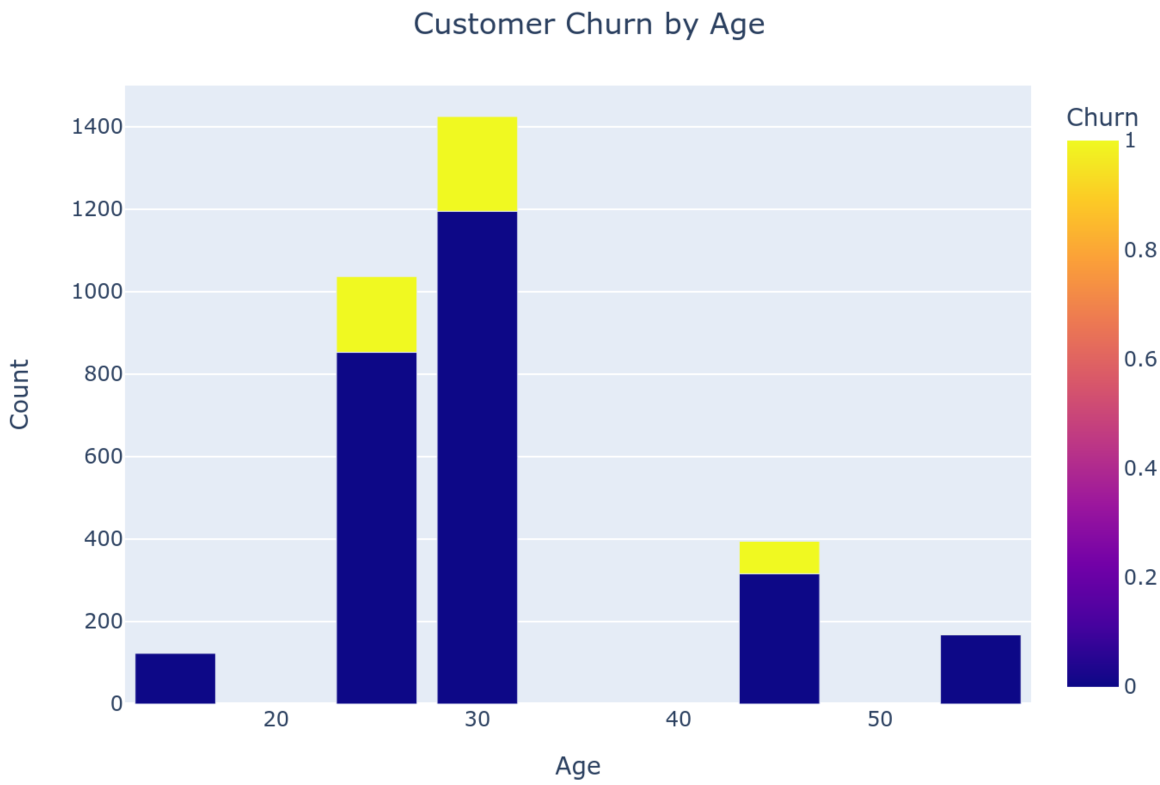
Findings and Analysis:

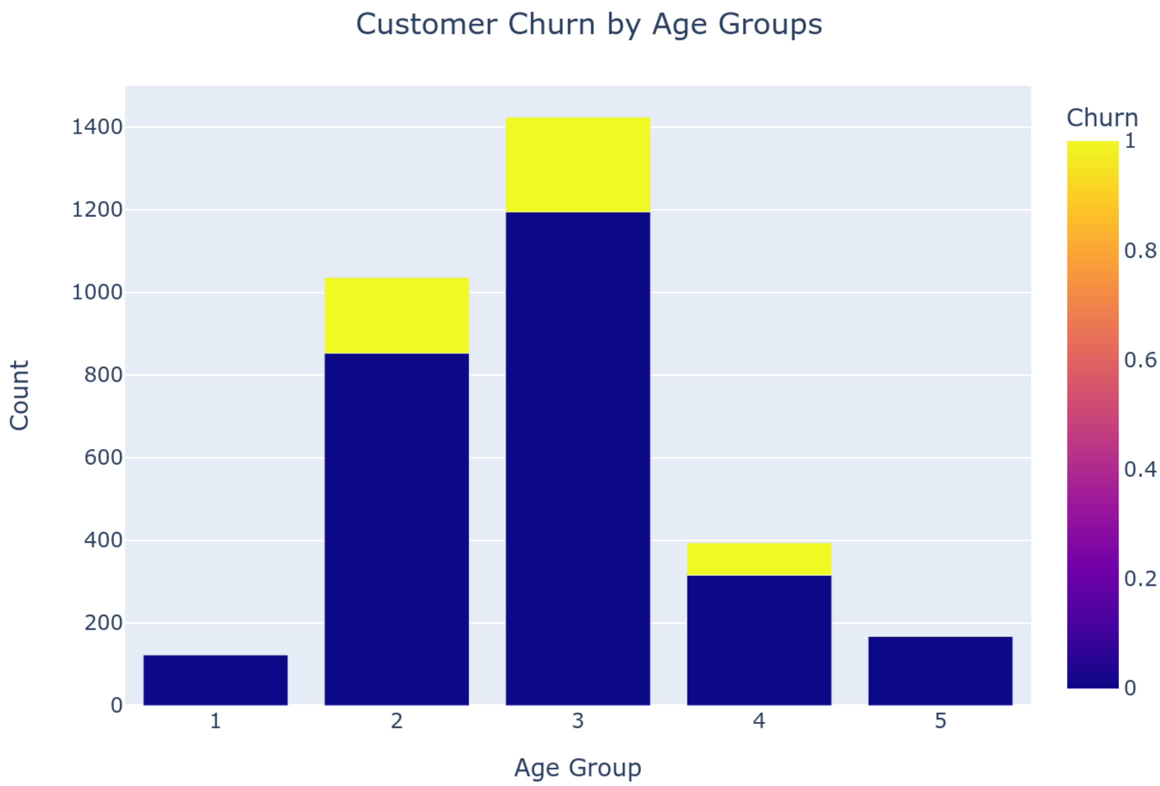
1.Percentile of Customer Churn

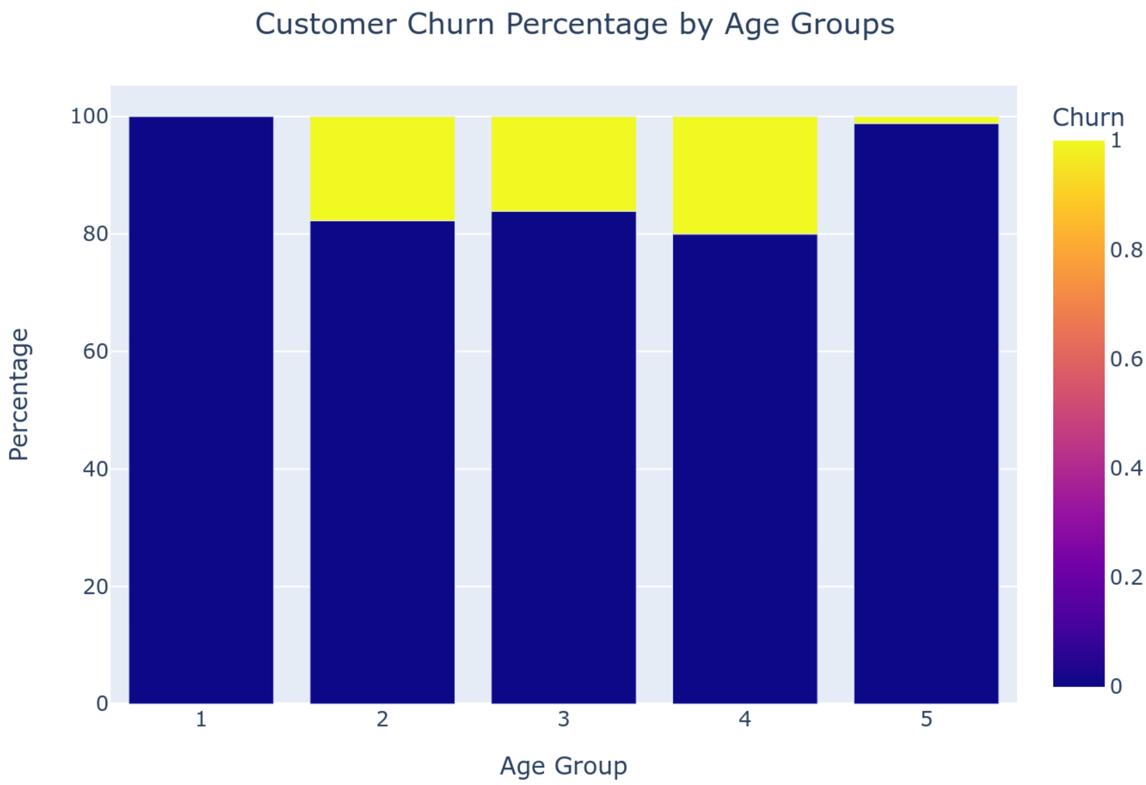


Out of the customers in the dataset, 84.3% (2655 customers) have not canceled their membership, while 15.7% (495 customers) have terminated their membership.

2. Which age groups are the customers who canceled their membership?


Customer Churn by Age
Count
1400
1200
1000
800
600
400
200
0
20
30
40
50
Age
Churn
1
0.8
0.6
0.4
0.2
0

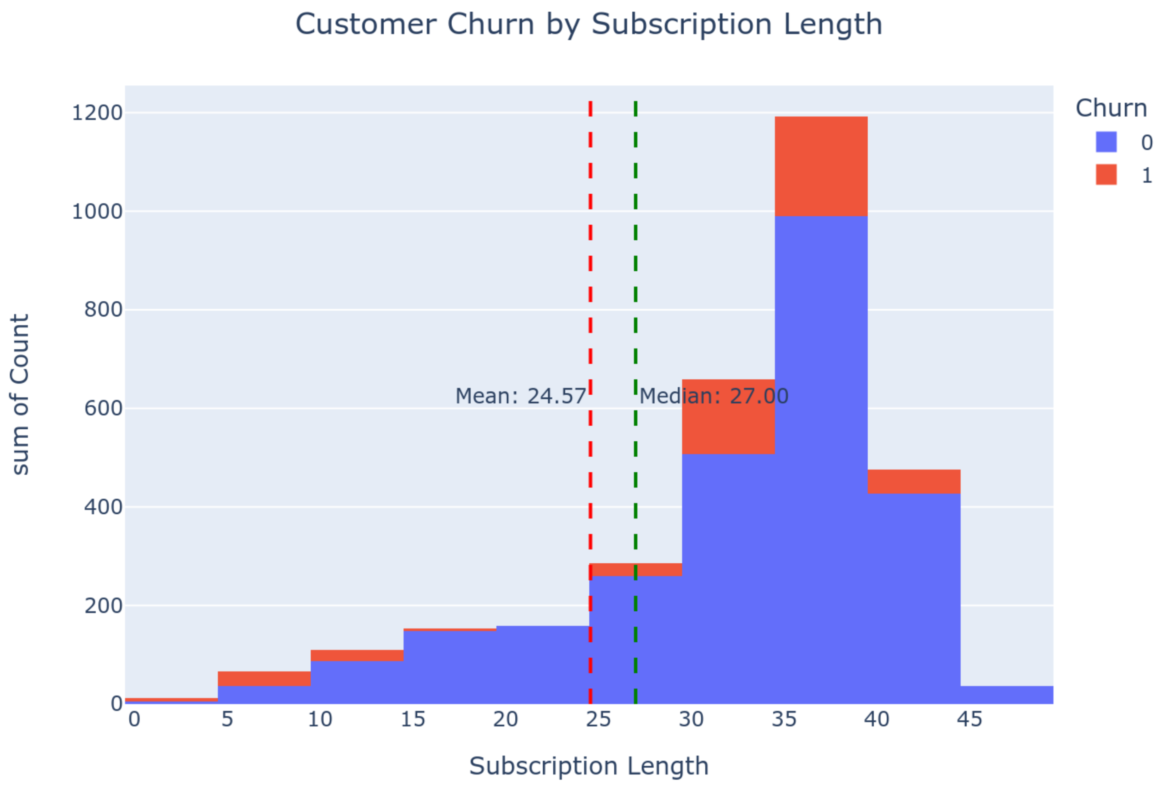

Customer Churn by Age Groups
Count
1400
1200
1000
800
600
400
200
0
1
2
3
4
5
Age Group
Churn
1
0.8
0.6
0.4
0.2
0



Based on the graphs above, we can make the following observations:

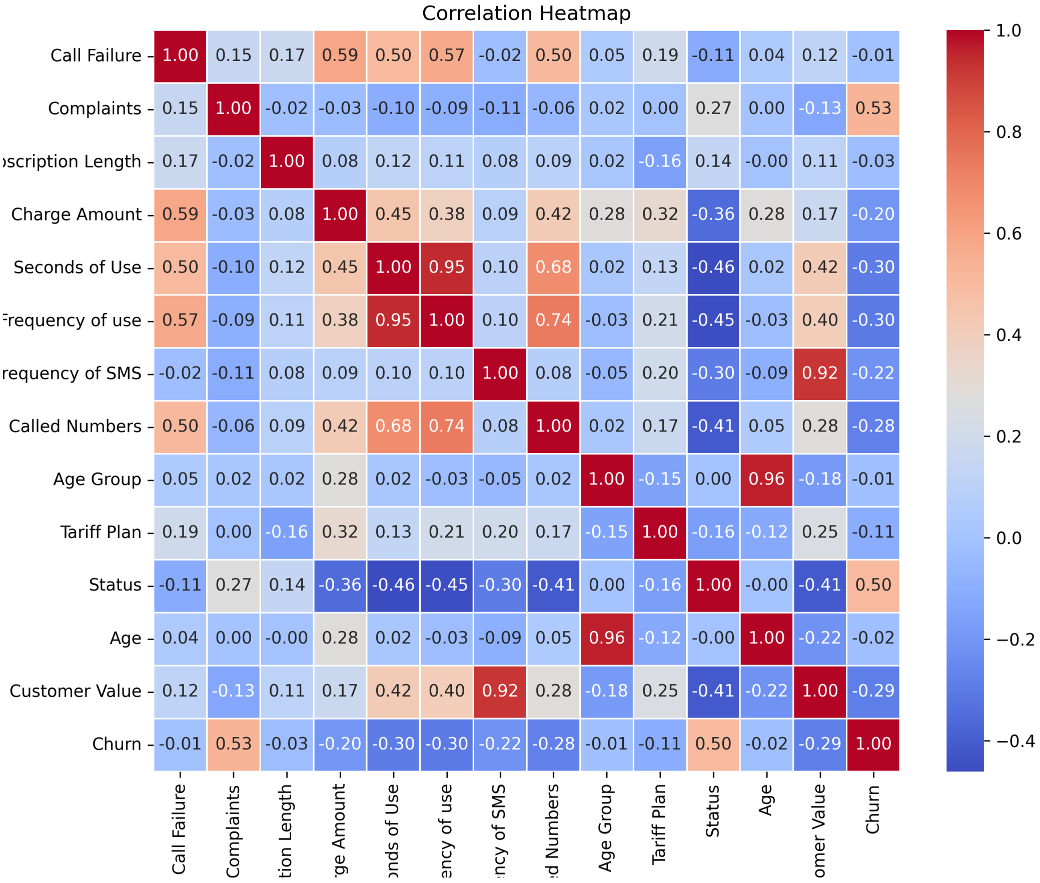
- The data set includes 5 different age groups, ranging from 1-15 years old to 55 years old.
- None of the users in the 15 year old group cancelled their membership.
- 184 users (18%) in the 25 age group cancelled their membership.
- 230 users (16%) in the 30 age group cancelled their membership.
- 79 users (20%) in the 45 age group cancelled their membership.
- Only 2 users (1%) in the 55 age group cancelled their membership.

2. How long was the subscription length of customers who canceled their membership?



According to our data, customers who have ended their subscriptions had an average subscription length of 24.57 months and a median subscription length of 27.00 months. We have noticed that most customers (203 customers) who ended their memberships had a membership period of 35-39 months at most.

3. Which features may be associated with customer churns?



A correlation coefficient of 0.53 between complaints and churn indicates that there is a moderate positive correlation between the two variables. Therefore, as the number of complaints increases, the likelihood of churn also increases. In other words, customers who have filed complaints are more likely to churn than those who have not.

A correlation coefficient of 0.50 between status and churn also indicates a moderate positive correlation between the two variables. This means that as the status of a customer changes from active to inactive, the probability of churn also increases. In other words, inactive customers are more likely to churn than active customers.

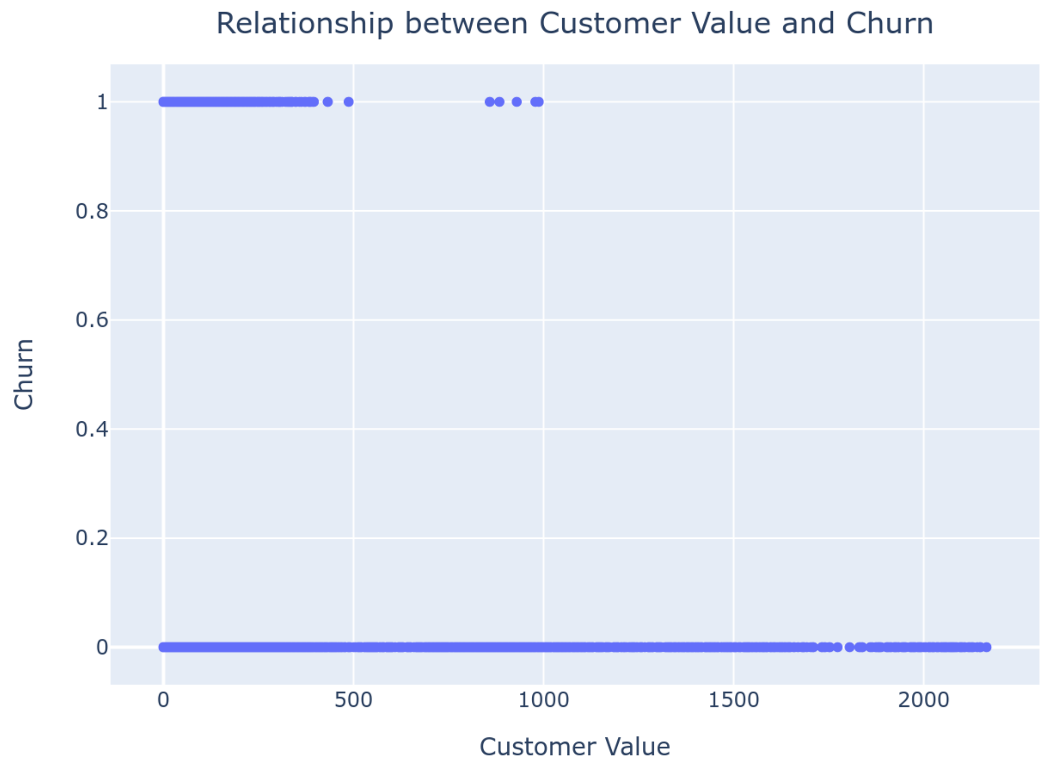
A correlation coefficient of -0.30 between seconds of use and churn indicates a weak negative correlation between the two variables. This means that as the number of seconds of use increases, the probability of churn decreases. In other words, customers who use their phones more are less likely to churn than those who use their phones less.

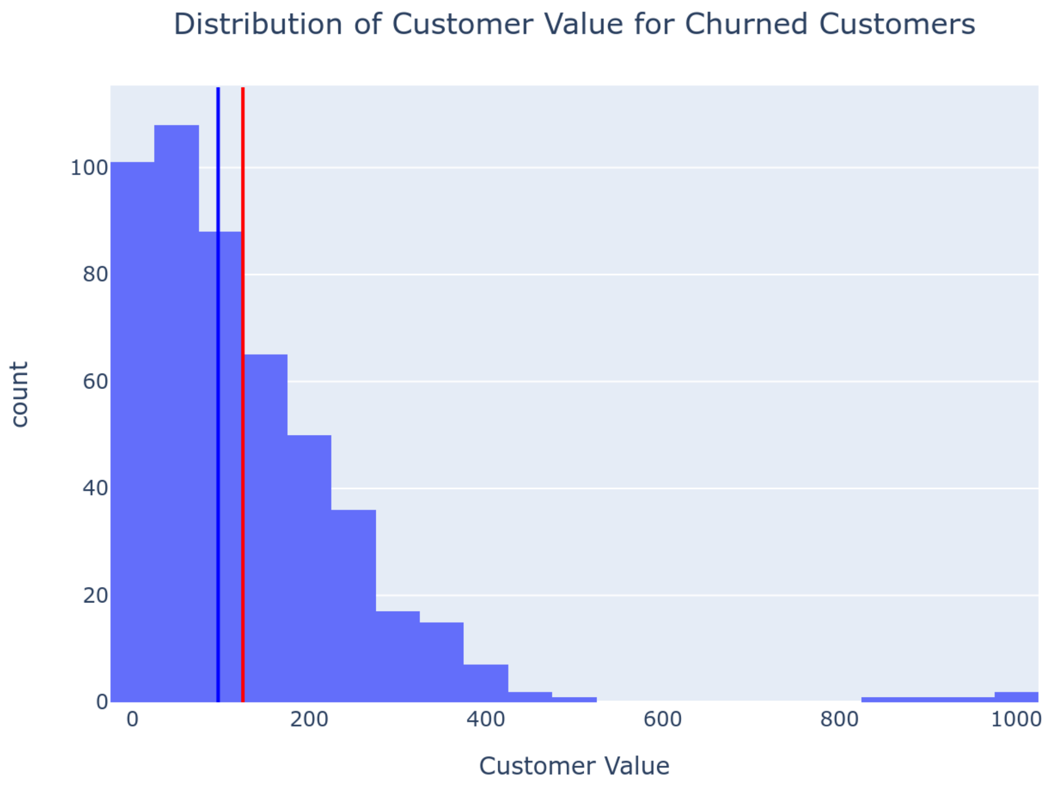
A correlation coefficient of -0.30 between frequency of use and churn also suggests a weak negative correlation between the two variables. This means that as the frequency of use increases, the probability of churn decreases. In other words, customers who use their phones more often are less likely to churn than those who use their phones less often.

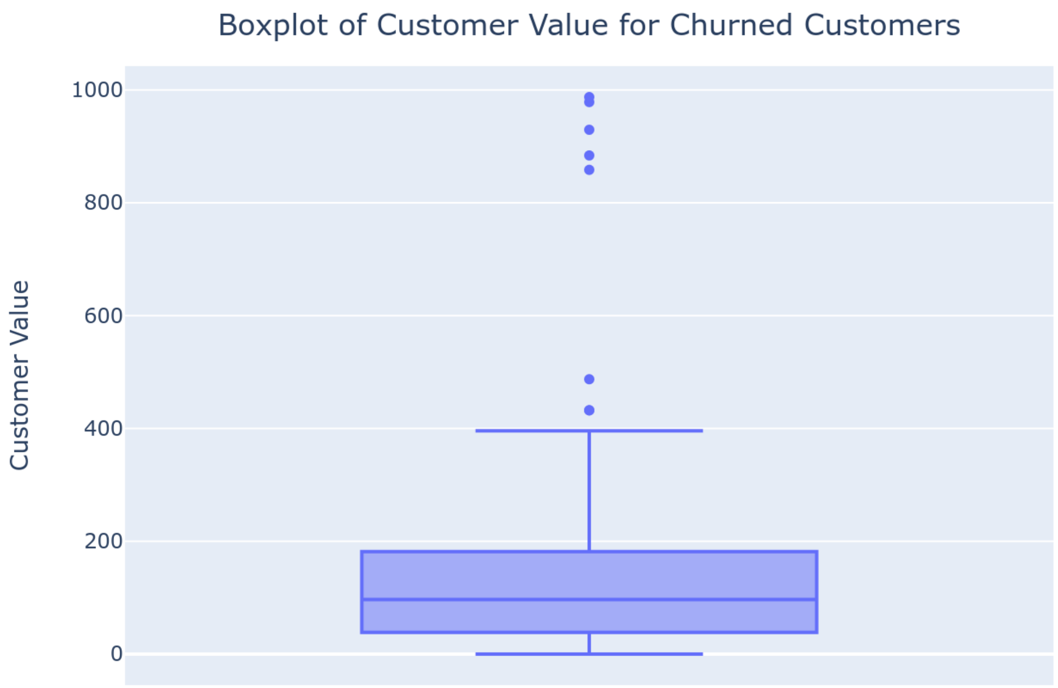
Overall, the correlation coefficients indicate that the following factors are most likely to increase the likelihood of customer churn:

- Filing complaints
- Changing status from active to inactive
- Using the phone less frequently

4. Relationship between 'Customer Value' and 'Churn'




Distribution of Customer Value for Churned Customers
count
100
80
60
40
20
0
0
200
400
600
800
1000
Customer Value

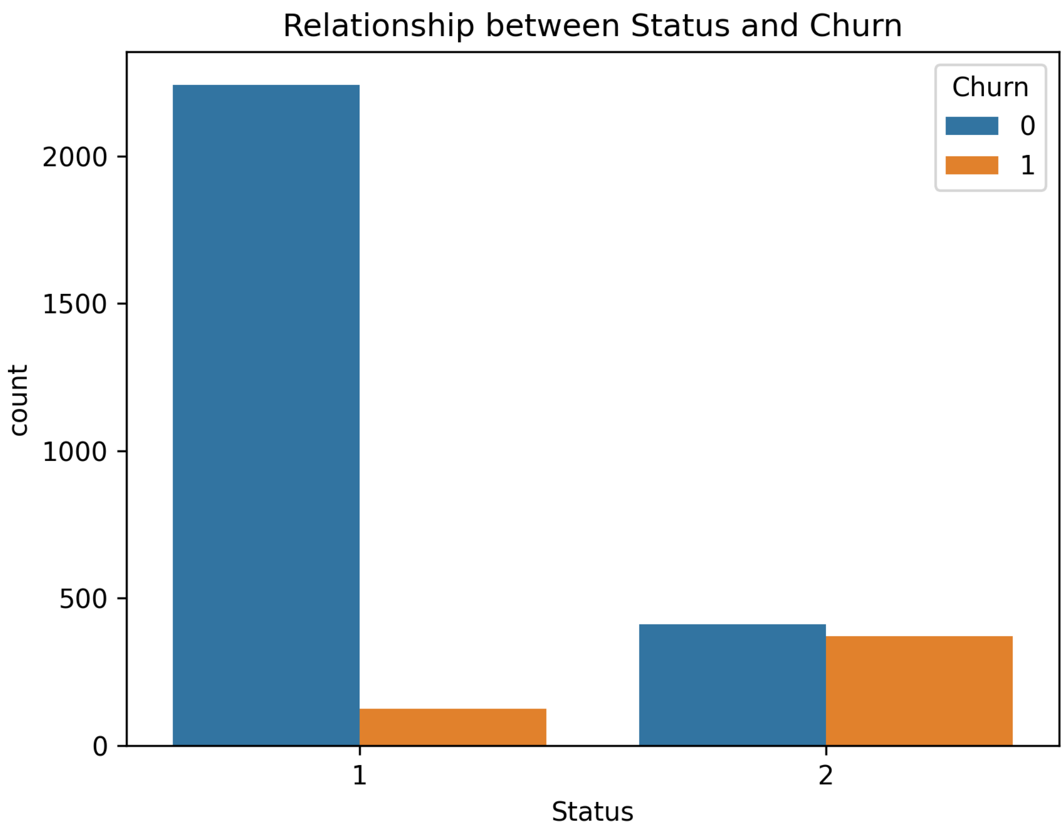


Based on our analysis of the graphs provided, we have come to the following conclusions:

- It is evident from the graphs that customers with a higher Customer Value tend to stay with the company, while those with a lower Customer Value tend to churn. In fact, our analysis shows that customers with a Customer Value greater than 1000 did not churn, indicating that the company needs to focus on attracting and retaining high-value customers.
- On the other hand, the majority of churned customers had a customer value between 0-400. This suggests that the company needs to improve its services or marketing strategies to retain these customers. Perhaps offering personalized deals or discounts based on their usage patterns could help retain such customers.
- Moreover, we found that the average user value of churned customers was 124.81, with a median value of 96.84. This indicates that even though the majority of churned customers had a lower customer value, there were a significant number of customers with higher customer value who also churned.

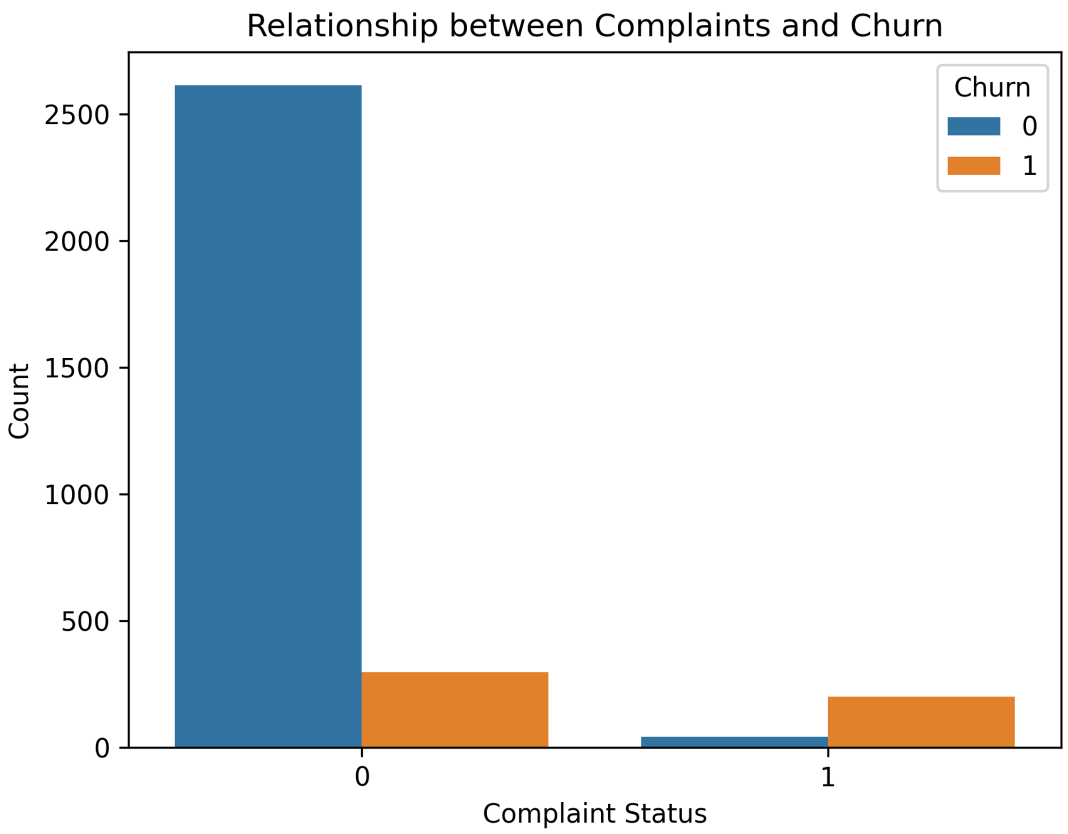
The company needs to identify the reasons behind the churn of these high-value customers and take appropriate measures to retain them.

5. Relationship between ‘Status’ and ‘Churn’



The percentage of customers who have churned is considerably lower among active customers (1), while the percentage of customers who have churned among inactive customers (2) is nearly the same as that of customers who have not churned.

6. Relationship between ‘Complaints’ and ‘Churn’



The percentage of customers who have churned service is significantly lower for those who have not made any complaints (0), in contrast to those who have made complaints (1), where the percentage of customers who have churned is notably higher.

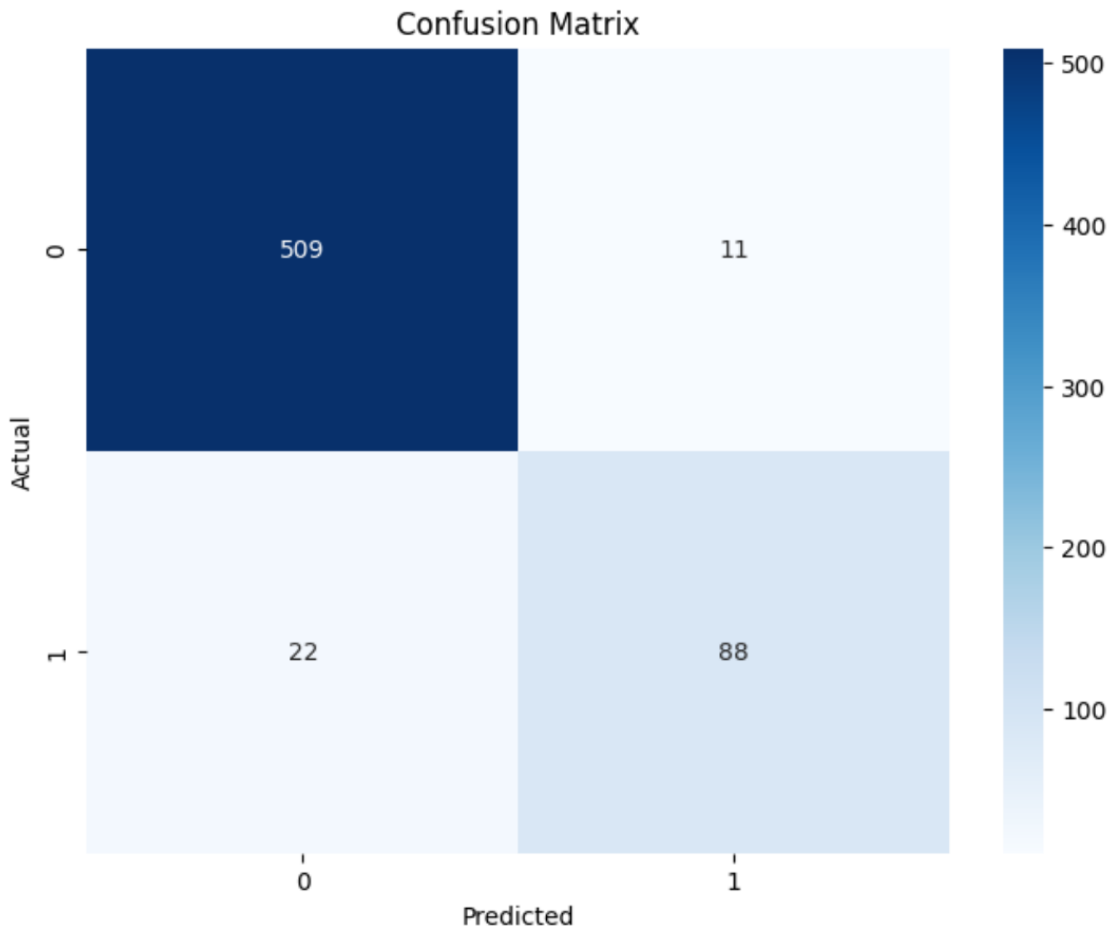
Machine Learning Process

We utilized the LazyPredict package to determine the most suitable Machine Learning Algorithm and ultimately decided on the LGBMClassifier algorithm from the available options. Following some necessary adjustments, we achieved a 99% success rate in predicting customer membership terminations in the training set and a 95% success rate in the test set.

Here are some of the scores our model achieved:

- Accuracy: 0.95
- Misclassification Rate (Error Rate): 0.05
- Sensitivity (Recall): 0.80
- False Positive Rate: 0.02
- Specificity: 0.98
- Precision: 0.89
- Prevalence: 0.17

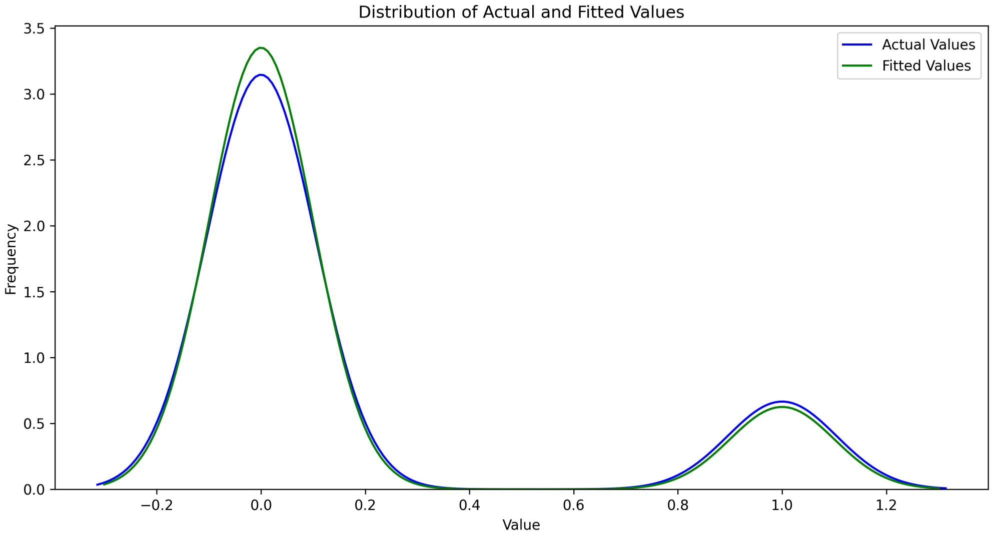
Confusion Matrix



Rates from Confusion Matrix

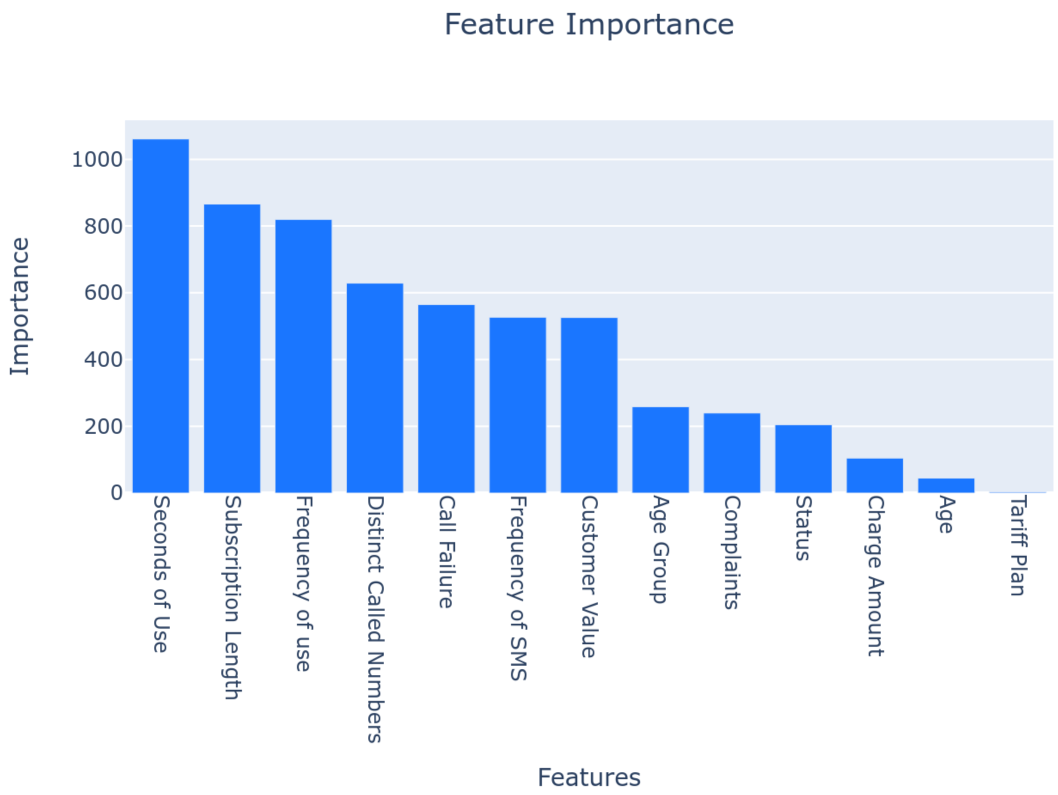
- Accuracy: Overall, how often is the classifier correct?
  - (TP + TN) / Total
- Misclassification Rate: Overall, how often is it wrong?
  - (FP + FN) / Total
  - equivalent to 1 minus Accuracy
  - also known as "Error Rate"
- True Positive Rate: When it's actually yes, how often does it predict yes?
  - TP/actual yes
  - also known as "Sensitivity" or "Recall"
- False Positive Rate: When it's actually no, how often does it predict yes?
  - FP/actual no
- True Negative Rate: When it's actually no, how often does it predict no?
  - TN/actual no
  - equivalent to 1 minus False Positive Rate
  - also known as "Specificity"
- Precision: When it predicts yes, how often is it correct?
  - TP/predicted yes
- Prevalence: How often does the yes condition actually occur in our sample?
  - actual yes/total

Distribution Plot



The more the two colors overlap, the more accurate the model

Feature Importance



Upon examining the graph above, it becomes evident that the duration in seconds of customer calls is the most significant parameter in training the model.

Conclusion

Overall, this data analysis report provides valuable insights into the factors that contribute to customer churn in the telecom sector. Through a combination of data visualization and machine learning techniques, we were able to identify key correlations between customer behavior and churn. Specifically, we found that customers who file complaints, become inactive, or use their phones less frequently are more likely to churn. Additionally, we found that customers with higher Customer Value tend to stay with the company, while those with a lower Customer Value tend to churn.

By applying machine learning algorithms and feature engineering techniques, we were also able to develop a predictive model that can accurately identify customers who are likely to churn. With a 95% success rate in predicting customer membership terminations, this model has the potential to help telecom companies proactively retain customers and reduce churn rates.

Overall, this report provides a valuable resource for telecom companies looking to better understand the factors that contribute to customer churn and develop effective strategies to retain customers.