Predicting Telcom Customer Churn

Submitted by: Aditi Gupta

Predicting Telcom Customer Churn

- What is customer churn?

- Problem Statement

- Context

- Exploratory Data Analysis

- Preprocessing and Training

- Modelling and Model Evaluation

- Appendix

What is customer churn?

Customer churn is defined as when customers or subscribers discontinue doing business with a firm or service.

Problem Statement

JB link telco company is encountering a problem of a high 27% customer loss leading to a 12% drop in our customer numbers. And urgently need to forecast which customers are prone to churn and recommend tailored strategies to retain customers.

Context

Customer churn is a critical challenge faced by the telecom industry. As customers switch from one service provider to another, telecom companies experience revenue loss and increased customer acquisition costs. To address this issue, we embarked on a project to develop machine learning models that can predict the likelihood of customer churn.

- Gather insights from the data to understand what is driving the high customer churn rate.
- Develop a Machine Learning model that can accurately predict the customers that are more likely to churn.
- Prescribe customized actions that could be taken to retain each of those customers.

Exploratory Data Analysis

Contract Type

- Customers with month-to-month contracts have a significantly higher churn rate. Approximately 54% of customers with month-to-month contracts churned, compared to only 11% with one-year contracts and 3% with two-year contracts.
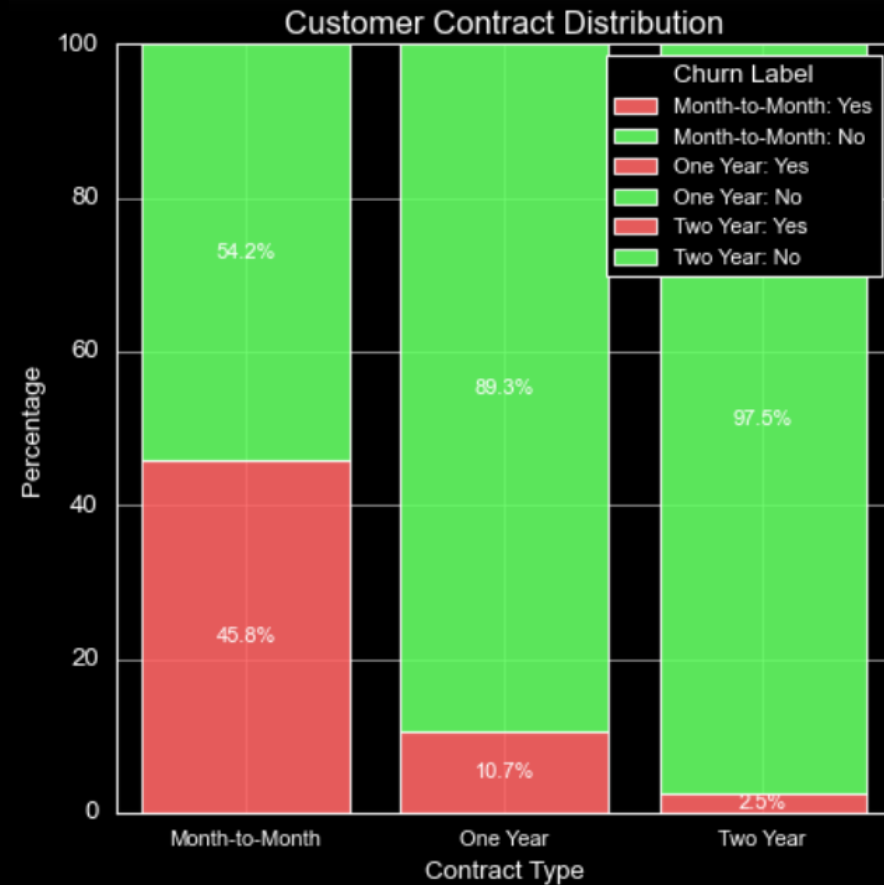
Customer Satisfaction

- Customers with satisfaction scores below 3 are more likely to churn. Visualization of customer satisfaction distribution highlighted that lower scores correlate with higher churn.

Customer Satisfaction Distribution with Churn Value

Internet Type

- Customers with cable or DSL services show higher churn rates compared to those with fiber optic services. This insight suggests that service quality differences impact customer retention.

Customer Internet Service Distribution
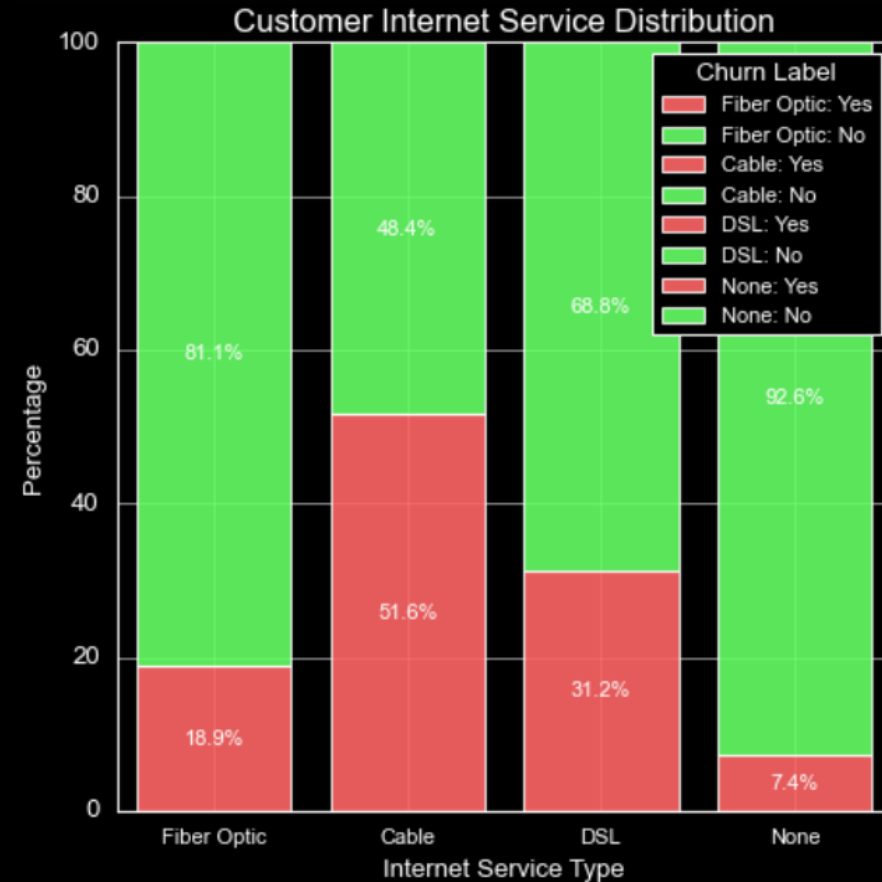
Geographic Analysis

- Los Angeles has the highest number of customers, but San Diego has the highest number of churned customers, with many citing better offers from competitors. This indicates a regional disparity in competitive pressure and customer retention.

Top 10 Cities with Churn Label: Yes and Internet Type (Sorted by Total Customer Count)
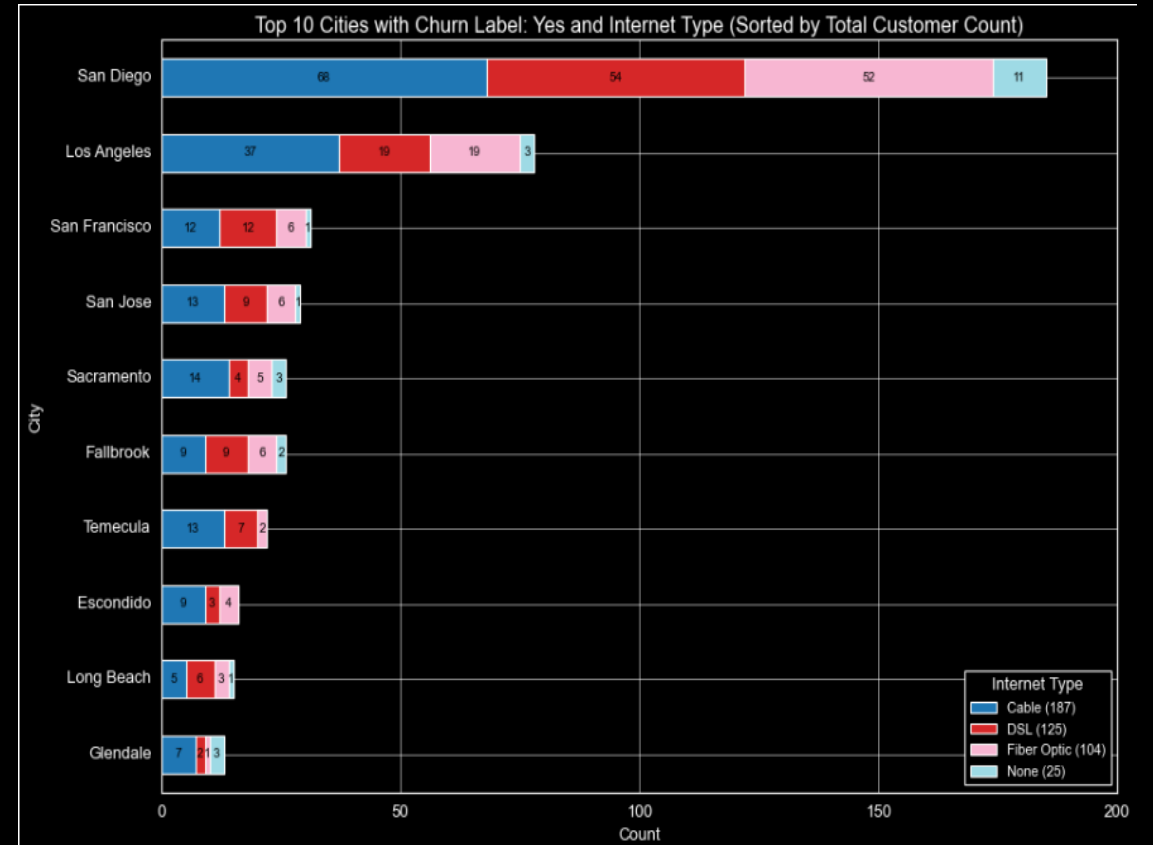
PREPROCESSING AND TRAINING

Customer Statistics

- Feature Selection: Removed unnecessary columns and retained features with significant correlation to churn. Created a correlation matrix to identify these features, setting a threshold of 0.2.

- Data Transformation: Converted categorical variables into numerical ones using one-hot encoding.

- Data Splitting: Split the data into training (80%) and testing (20%) sets to prepare for model training and evaluation.

-

Correlation of Features with Churn Value
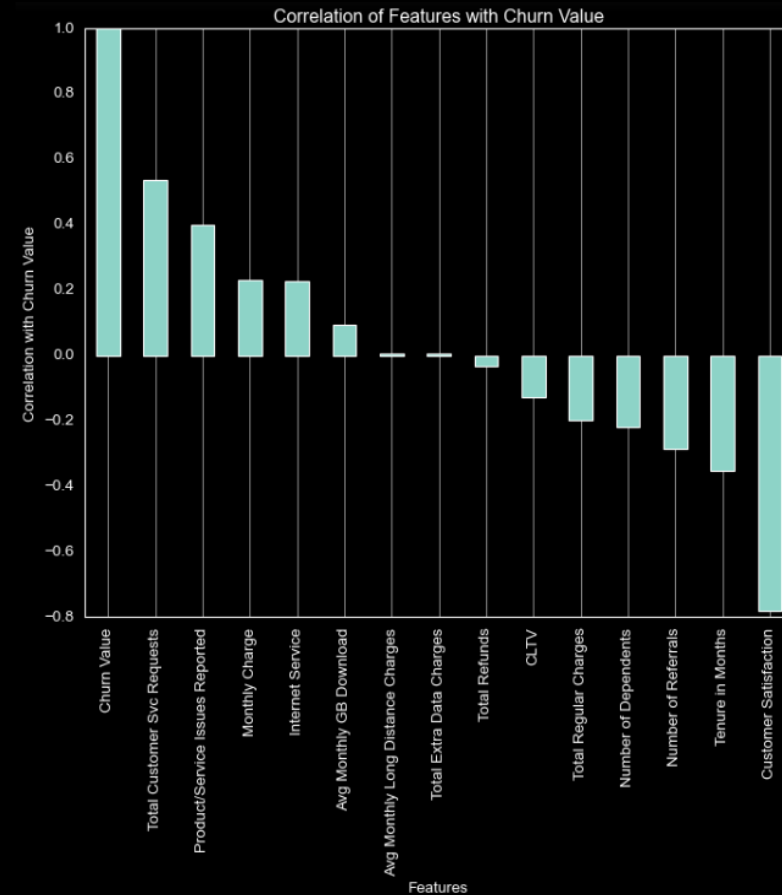
Modelling and Model Evaluation

# Potential customers who are most likely to churn

- Goal: Main goal was to identify the potential customers who are most likely to churn.

- I evaluated three models and after comparing Logistic Regression, Random Forest Classifier and HistGradientBoostingClassifier models, I select the best-model based on a good balance between all the measuring.

```
                             Model  Accuracy  F1 Score  Precision  Recall \
0         Random Forest Classifier  0.898510  0.805970   0.881306  0.7425
1              Logistic Regression  0.886444  0.776536   0.879747  0.6950
2  HistGradientBoostingClassifier  0.916253  0.840541   0.914706  0.7775

       Confusion Matrix
0  [[969, 40], [103, 297]]
1  [[971, 38], [122, 278]]
2   [[980, 29], [89, 311]]
```

Confusion Matrix for ENT

- High True Positives and True Negatives: The model correctly identified a large number of both churned and non-churned customers.

- Low False Negatives: Only 29 false negatives, indicating that the model missed few actual churn cases.

- Moderate False Positives: 89 false positives, which is manageable but indicates some over-prediction of churn.

Confusion Matrix for HistGradientBoostingClassifier

	Predicted No Churn	Predicted Churn
Actual No Churn	980	29
Actual Churn	89	311
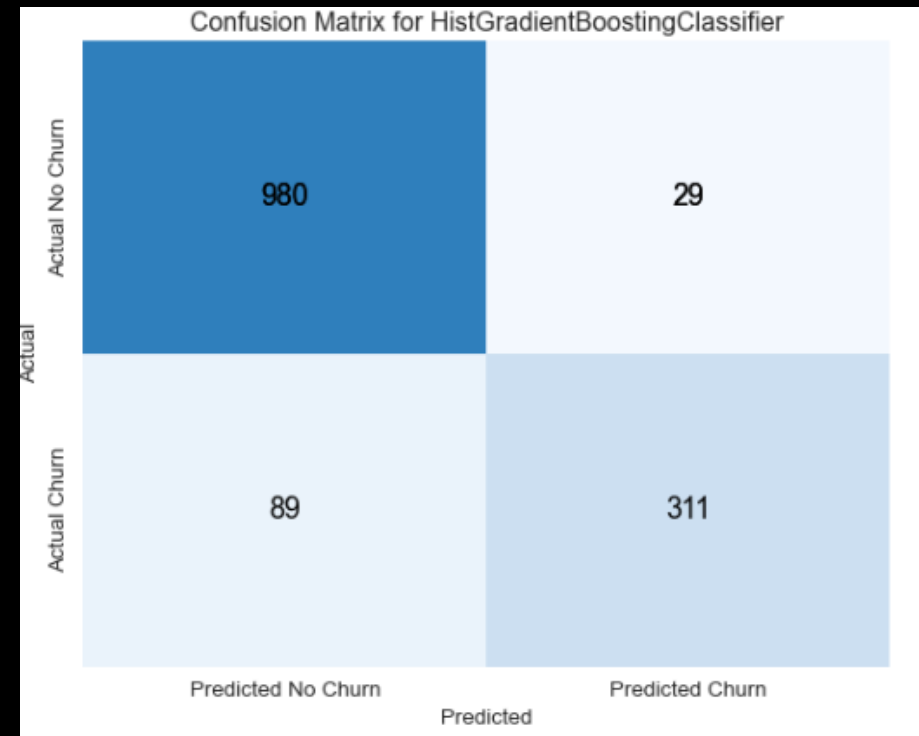
Appendix

Column Name	Description
Churn Value	1 = the customer left the company this quarter. 0 = the customer remained with the company
Customer ID	A unique ID that identifies each customer
Referred a Friend	Indicates if the customer has ever referred a friend or family member to this company
Number of Referrals	Indicates the number of referrals to date that the customer has made
Tenure in Months	Indicates the total amount of months that the customer has been with the company by the end of the quarter specified
Offer	Identifies the last marketing offer that the customer accepted, if applicable
Phone Service	Indicates if the customer subscribes to home phone service with the company
Avg Monthly Long-Distance Charges	Indicates the customer's average long-distance charges, calculated to the end of the quarter
Multiple Lines	Indicates if the customer subscribes to multiple telephone lines with the company
Internet Service	Indicates if the customer subscribes to Internet service with the company
Internet Type	Indicates the type of Internet service the customer subscribes

Column Name	Description
Avg Monthly GB Download	Indicates the customer's average download volume in gigabytes, calculated to the end of the quarter
Online Security	Indicates if the customer subscribes to an additional online security service provided by the company
Online Backup	Indicates if the customer subscribes to an additional online backup service provided by the company
Device Protection Plan	Indicates if the customer subscribes to an additional device protection plan for their Internet equipment
Premium Tech Support	Indicates if the customer subscribes to an additional technical support plan from the company with reduced
Streaming TV	Indicates if the customer uses their Internet service to stream television programing from a third-party provider
Streaming Movies	Indicates if the customer uses their Internet service to stream movies from a third-party provider
Streaming Music	Indicates if the customer uses their Internet service to stream music from a third-party provider
Unlimited Data	Indicates if the customer has paid an additional monthly fee to have unlimited data downloads/uploads
Contract	Indicates the customer's current contract type
Paperless Billing	Indicates if the customer has chosen paperless billing
Payment Method	Indicates how the customer pays their bill

Column Name	Description
Monthly Charge	Indicates the customer's current total monthly charge for all their services from the company
Total Regular Charges	Indicates the customer's total regular charges, excluding additional charges
Total Refunds	Indicates the customer's total refunds
Total Extra Data Charges	Indicates the customer's total charges for extra data downloads above those specified in their plan
Total Long Distance Charges	Indicates the customer's total charges for long distance above those specified in their plan
Gender	The customer's gender
Age	The customer's current age
Under 30	Indicates if the customer is under 30 years old
Senior Citizen	Indicates if the customer is 65 or older
Married	Indicates if the customer is married
Dependents	Indicates if the customer lives with any dependents: Yes, No. Dependents could be children, parents, grandparents, etc.
Number of Dependents	Indicates the number of dependents that live with the customer

Column Name	Description
City	The city of the customer's primary residence
Zip Code	The zip code of the customer's primary residence
Latitude	The latitude of the customer's primary residence
Longitude	The longitude of the customer's primary residence
Population	A current population estimate for the entire Zip Code area
CLTV	Customer Lifetime Value. A predicted CLTV is calculated using corporate formulas and existing data. The higher the value, the more valuable the customer
Churn Category	A high-level category for the customer's reason for churning
Churn Reason	A customer's specific reason for leaving the company
Total Customer Svc Requests	Number of times the customer contacted customer service in the past quarter
Product/Service Issues Reported	Number of times the customer reported an issue with a product or service in the past quarter
Customer Satisfaction	A customer's overall satisfaction rating of the company from 1 (Very Unsatisfied) to 5 (Very Satisfied) collected on customer service requests

Thank you!