Telco - Final Report

Problem statement

Telco, a leading provider of telephone, internet, and streaming TV services in California, reached a customer base of 7,073 during the third quarter, thanks to attractive promotions. However, the company faces significant challenges in customer retention, with a churn rate of 27% during this period. This high turnover rate not only destabilizes the customer base but also poses a risk to the company's long-term profitability.

This report presents a machine learning-based solution to forecast customer churn. By leveraging advanced data analysis and machine learning techniques, Telco can identify customers at risk of leaving and implement targeted interventions to enhance retention and loyalty.

Datasets

The dataset used for this study was downloaded from Kaggle . It includes key variables relevant to customer churn, such as demographic, service usage, and billing information.

The initial database is composed of the following columns with their meaning:

CustomerID : A unique ID that identifies each customer.

Count : A value used in reporting/dashboarding to sum up the number of customers in a filtered set.

Country : The country of the customer's primary residence.

State : The state of the customer's primary residence.

City : The city of the customer's primary residence.

Zip Code : The zip code of the customer's primary residence.

Lat Long : The combined latitude and longitude of the customer's primary residence.

Latitude : The latitude of the customer's primary residence.

Longitude : The longitude of the customer's primary residence.

Gender : The customer's gender: Male, Female

Senior Citizen : Indicates if the customer is 65 or older: Yes, No

Partner : Indicate if the customer has a partner: Yes, No

Dependents : Indicates if the customer lives with any dependents: Yes, No. Dependents could be children, parents, grandparents, etc.

Tenure Months : Indicates the total amount of months that the customer has been with the company by the end of the quarter specified above.

Phone Service : Indicates if the customer subscribes to home phone service with the company: Yes, No

Multiple Lines : Indicates if the customer subscribes to multiple telephone lines with the company: Yes, No

Internet Service : Indicates if the customer subscribes to Internet service with the company: No, DSL, Fiber Optic, Cable.

Online Security : Indicates if the customer subscribes to an additional online security service provided by the company: Yes, No

Online Backup : Indicates if the customer subscribes to an additional online backup service provided by the company: Yes, No

Device Protection : Indicates if the customer subscribes to an additional device protection plan for their Internet equipment provided by the company: Yes, No

Tech Support : Indicates if the customer subscribes to an additional technical support plan from the company with reduced wait times: Yes, No

Streaming TV : Indicates if the customer uses their Internet service to stream television programing from a third party provider: Yes, No. The company does not charge an additional fee for this service.

Streaming Movies : Indicates if the customer uses their Internet service to stream movies from a third party provider: Yes, No. The company does not charge an additional fee for this service.

Contract : Indicates the customer's current contract type: Month-to-Month, One Year, Two Year.

Paperless Billing : Indicates if the customer has chosen paperless billing: Yes, No

Payment Method : Indicates how the customer pays their bill: Bank Withdrawal, Credit Card, Mailed Check

Monthly Charge : Indicates the customer's current total monthly charge for all their services from the company.

Total Charges : Indicates the customer's total charges, calculated to the end of the quarter specified above.

Churn Label : Yes = the customer left the company this quarter. No = the customer remained with the company. Directly related to Churn Value.

Churn Value : 1 = the customer left the company this quarter. 0 = the customer remained with the company. Directly related to Churn Label.

Churn Score : A value from 0-100 that is calculated using the predictive tool IBM SPSS Modeler. The model incorporates multiple factors known to cause churn. The higher the score, the more likely the customer will churn.

CLTV : Customer Lifetime Value. A predicted CLTV is calculated using corporate formulas and existing data. The higher the value, the more valuable the customer. High value customers should be monitored for churn.

Churn Reason : A customer's specific reason for leaving the company. Directly related to Churn Category.

Data Cleaning and Data Wrangling

To prepare the data for analysis, several cleaning steps were undertaken:

- The conversion of some columns from the type 'string' to that of 'numeric' :

- The conversion of some columns from the type 'string' to that of 'boolean':

- The elimination of some columns because:

- presenting more than 50% of missing information: Total Charges, Churn Reason

- presenting no interest in the development of the model: CustomerID and Count

- presenting a strong collinearity with other columns : Churn Value

- The elimination of three (3) lines of information because presenting missing values for the column 'Churn Label'

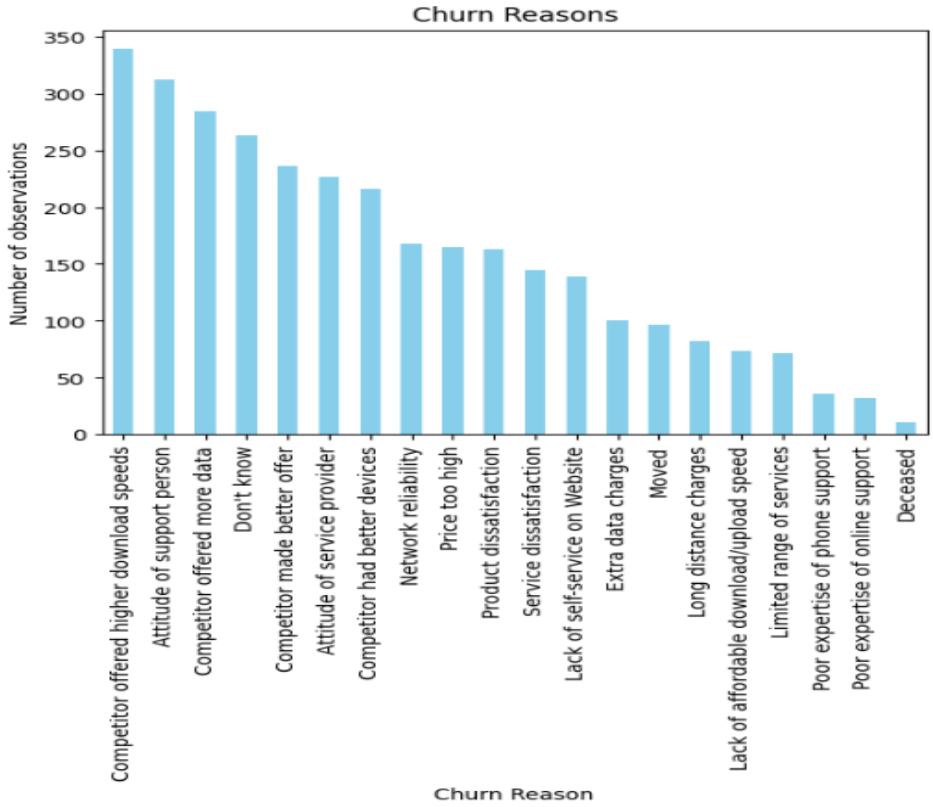
Exploratory Data Analysis

Key findings from the exploratory analysis include:

1. **Reasons for churn are predominantly related to service quality, customer service, and pricing (18.44%):**

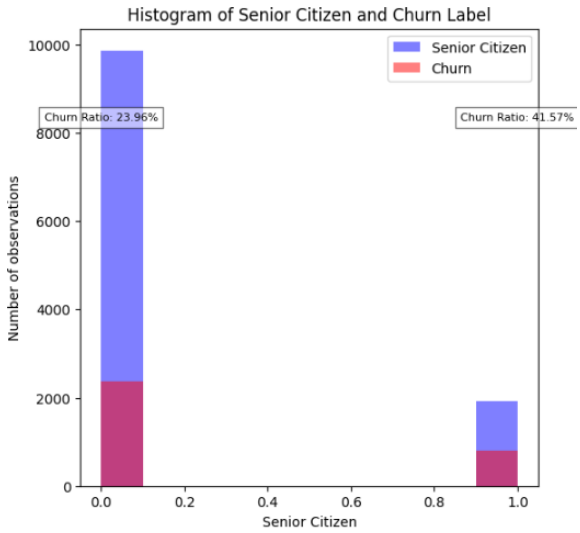
Reason	%
Quality of products and services offered to customers	43.93
Customer service	26.04
Pricing	18.44
Other: Deceased, Moved, Don't know	11.59
Total	**100.00**

Expressed in terms of frequency, the reasons for unsubscribing are as follows:



Telco - Final Report

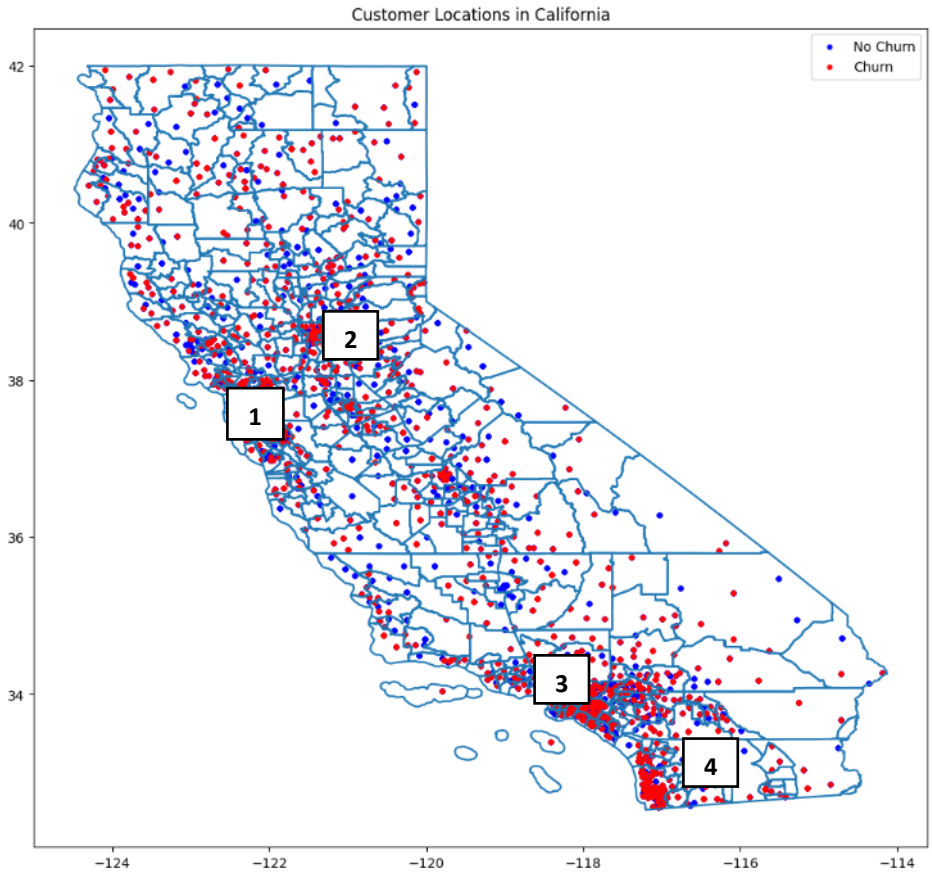
2.- Senior Citizens are the most likely to churn with a percentage of 41.57% in this segment of customers vs



3.- Gender does not significantly influence churn:

- Male (26.19%),
- Female (27.46%).

4.- Geographic churn clusters indicate churn is not evenly distributed across California



Telco - Final Report

Method

To predict churn, three machine learning models were tested using the Python library scikit-learn.

The approach adopted was to find the optimal parameters for these models using two different methods. This led to the following results:

	Model	Recall_train	Recall_test	Accuracy
0	KNN	0.979384	0.983179	0.959107
1	Random Forest	0.995207	0.995360	0.991589
2	Gradient Boosting	0.993182	0.995940	0.991299

Gradient Boosting emerged as the most effective model due to its high recall and precision. It is recommended for deployment

Recommendations and Next Steps

The model is robust, balanced and efficient. The following steps are suggested to enhance customer retention:

- Improve data collection on churn reasons for targeted interventions.
- Automate the calculation of Total Charges to enable profitability-based churn analysis.

The next steps should be:

- Validate of the model by executive stakeholders.
- Train the sales team to use the model effectively.
- Develop a proactive customer retention policy based on model insights.
- Regularly evaluate and refine the model post-deployment.

https://www.kaggle.com/datasets/yeanzc/telco-customer-churn-ibm-dataset