Telecom Customer Churn Prediction Project Report

# Title Page

Project Title: Telecom Customer Churn Prediction

Author: Raushan Ranjan

# Abstract

Customer churn is a significant concern in the telecom industry. This project focuses on understanding the reasons behind customer churn and predicting which customers are likely to leave. Using exploratory data analysis (EDA), data preprocessing, and machine learning models such as decision trees and logistic regression, this project analyzes a customer dataset to gain insights and make predictive models.

# Introduction

Churn refers to when a customer discontinues using a service. Reducing churn is essential for maintaining revenue and growth. This project aims to explore the dataset, perform preprocessing, conduct EDA, and apply machine learning models to predict churn.

# Dataset Description

Source: Telecom customer dataset

Total Records: 7043

Attributes: 38 attributes including demographic details, service subscriptions, and revenue metrics.

Example columns:

- Gender, Age, Married

- Internet Service, Contract, Payment Method

- Monthly Charge, Total Charges, Churn Category

# Data Preprocessing

- Removed columns with >50% missing values: Offer, Churn Reason, Churn Category

- Converted categorical variables to numerical

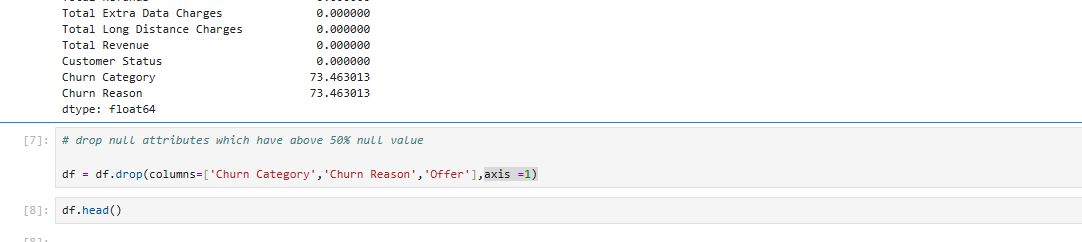
- Scaled continuous features using MinMaxScaler

- Dropped unimportant features like Zip Code, Latitude, Longitude, City

Data Cleaning Result:

- Final attributes: 32

- Converted all 'Yes'/'No' to 1/0



# Exploratory Data Analysis (EDA)

1. Churn Count: 26.54% customers have churned

2. Churn by Gender: Both genders show a similar churn trend

3. Churn by Marital Status: Higher churn observed among married individuals

4. Tenure Analysis: Customers with short tenure (1-2 months) have higher churn

5. Payment Method: Customers paying via bank withdrawal show higher churn

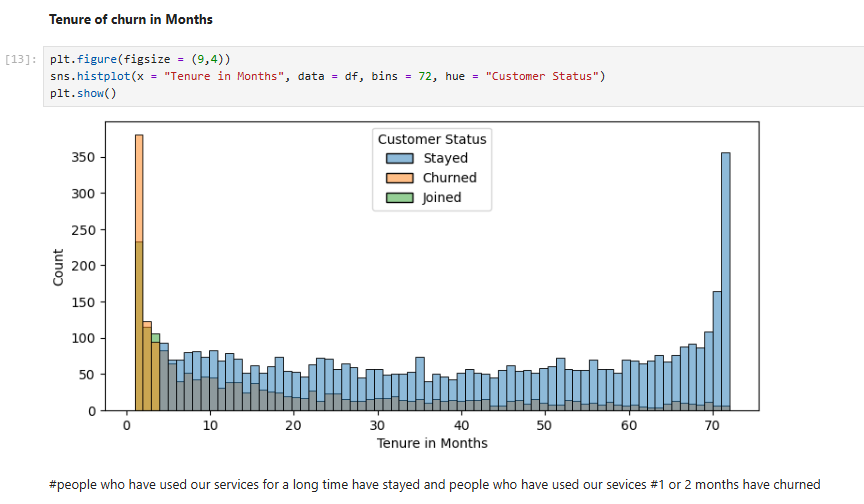
6. Internet Service: Churn is higher among those with internet services

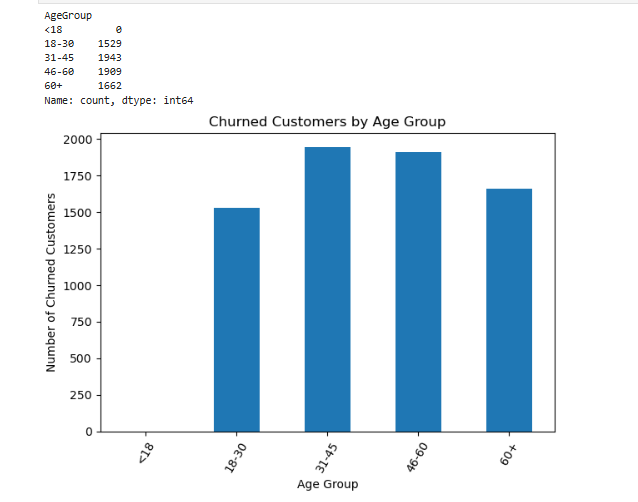
7. Contract Type: Month-to-month contracts see the highest churn

8. Internet Type: Fiber optic users show higher churn

9. Online Services: Lack of online security and backup increases churn

10. Churn by Age Group: Highest churn in age groups 31-45 and 46-60



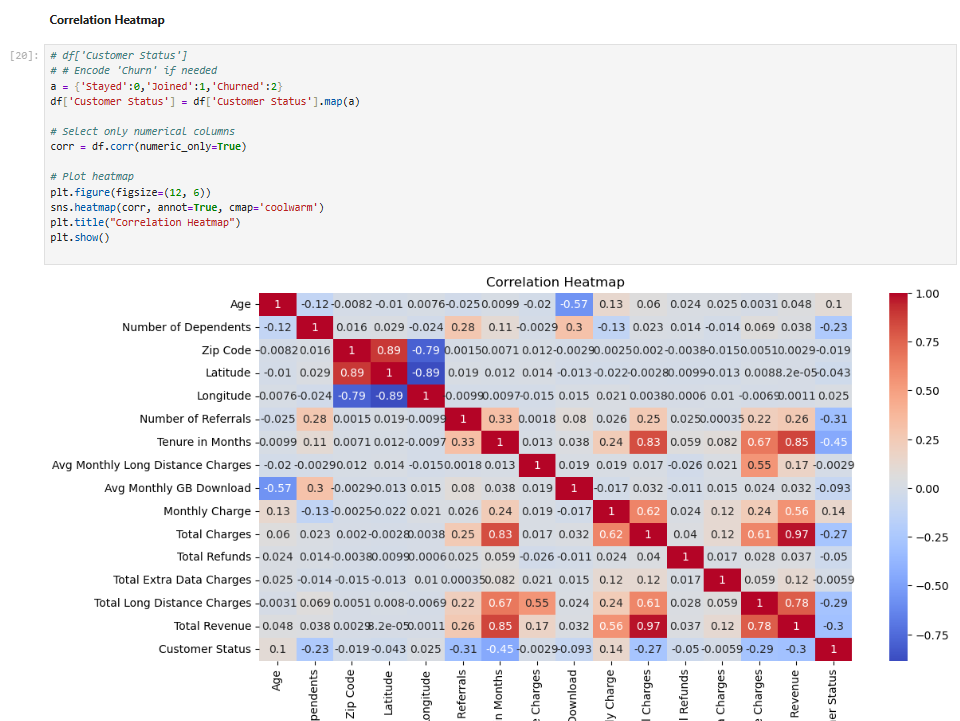


# Correlation Analysis

- Strong correlation between:

- Total Charges & Total Revenue

- Tenure & Revenue

- Weak correlation between age and churnMachine Learning Models

Decision Tree Classifier:

- Train Accuracy: 99.97%

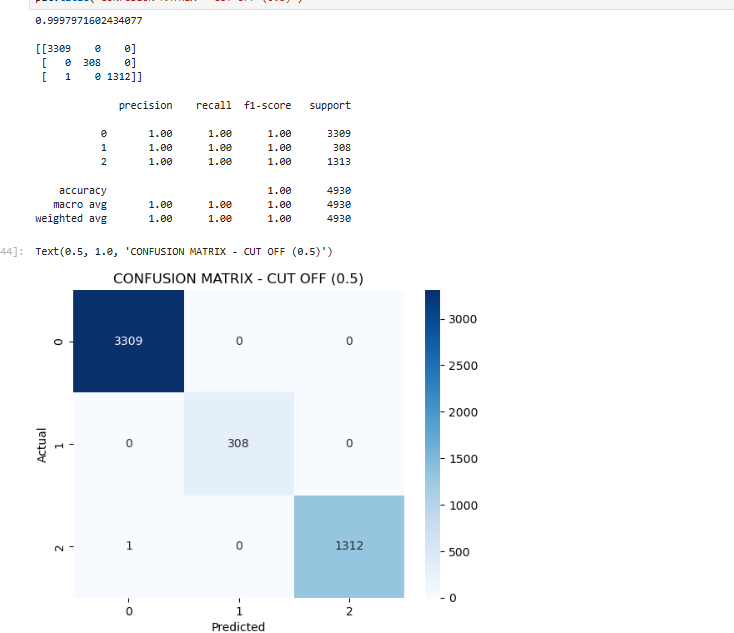
- Test Accuracy: 78.65%

- Test Confusion Matrix:

- Stayed: Precision 0.88, Recall 0.87

- Joined: Precision 0.68, Recall 0.66

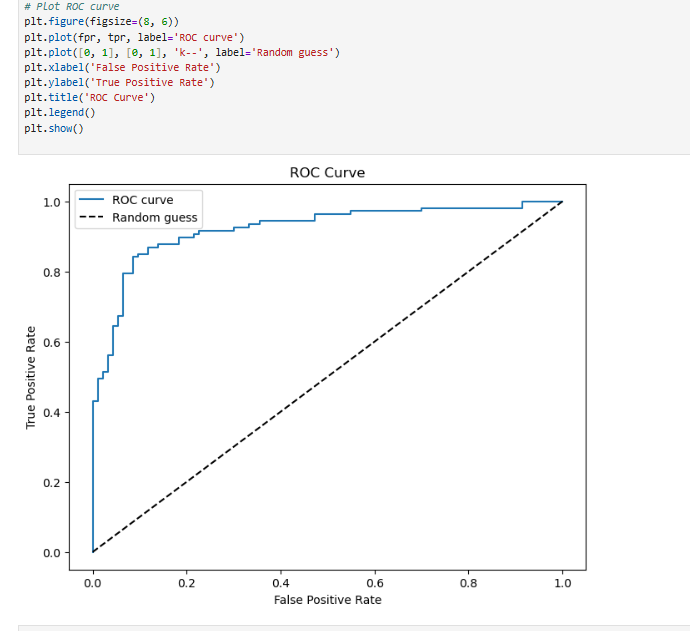
- Churned: Precision 0.59, Recall 0.62



Logistic Regression & ROC Curve:

- Built on sample classification data

- Plotted ROC Curve



# Model Evaluation

- Overfitting observed in Decision Tree (very high train accuracy)

- Logistic Regression gave good probability scores for ROC

- F1-scores suggest moderate success in classifying churned customers

# Conclusion & Recommendations

- Month-to-month contracts, lack of online services, and bank withdrawal payments are strong churn indicators.

- Customers aged 31-60 are most prone to churn.

- Providing better device offers, security, and backup may reduce churn.

- Consider using more robust models like Random Forest or XGBoost.

# References

- Customer Churn Dataset (CSV)

- Python Libraries: pandas, seaborn, matplotlib, scikit-learn