Customer Churn Prediction Project Documentation

Introduction

The objective of this project is to predict customer churn for a telecom company using machine learning and data analytics. Proactively identifying customers likely to churn empowers the business to take action and improve retention rates.

Dataset Overview

- Source: Public Telco Customer Churn dataset
- Samples: ~7,000 customer records
- Features: Customer demographics, account info, subscribed services, MonthlyCharges, TotalCharges, tenure, and churn label.

Data Preparation & Cleaning

- Dropped non-informative columns (e.g., customerID).
- Converted 'TotalCharges' from string to numeric format and imputed missing values with the median.
- Encoded categorical variables using label encoding.
- Target variable 'Churn' mapped as 0 (No churn) and 1 (Churn).

Exploratory Data Analysis (EDA)

- Churn Distribution: The majority of customers did not churn; about a quarter did. The bar plot clearly shows class imbalance.
- Correlation Heatmap: Strongest negative correlation with churn: tenure (-0.35) and contract type (-0.39), indicating long-term customers and those with long contracts are less likely to churn. Positive correlations include MonthlyCharges (+0.19) and TotalCharges (+0.20), suggesting higher charges relate to more churn.

Model Building & Evaluation

- Classifier Used: Logistic Regression (baseline model)
- Train/Test Split: 70/30 ratio.
- Results:
  - Accuracy: 81%
  - Confusion Matrix: Good at predicting non-churners (class 0), reasonable recall for churners (class 1).
  - Precision/Recall: Precision for churners = 0.68, recall = 0.57; better performance on non-churners.
  - F1 Scores: Overall balanced, indicating solid baseline modeling.

Feature Importance & Interpretation

- Top Factors increasing churn:
    - Higher MonthlyCharges, TotalCharges, PaperlessBilling, InternetService.
- Top Protective Factors:
    - Longer tenure, having a contract, PhoneService.

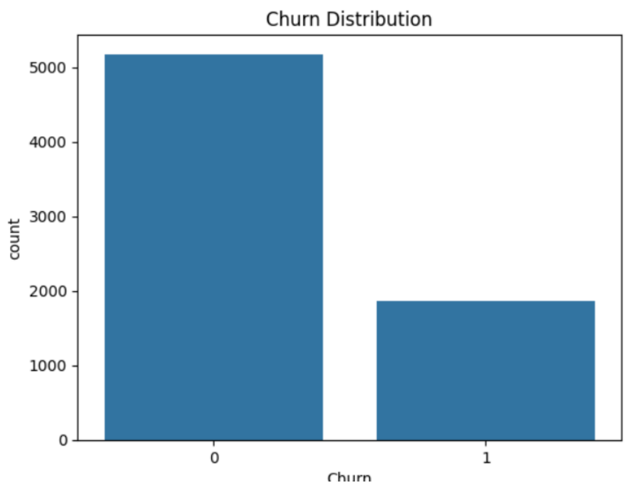
Business Insights & Recommendations

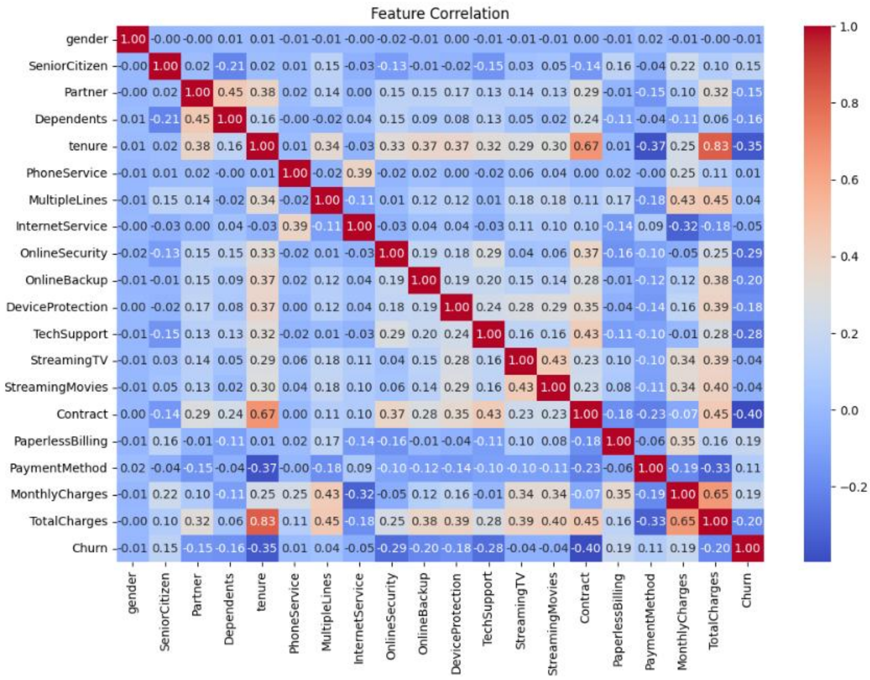
- Customers with short tenure or month-to-month contracts are significantly more likely to churn.
- Higher monthly bills increase the probability of churn, suggesting the need to review value offerings for high-bill customers.
- Strategies should target customers early in their tenure and those with flexible contracts to reduce churn rate.

Conclusion

The project demonstrates an end-to-end solution for churn prediction: from data wrangling and EDA to modeling and business interpretation. The logistic regression model yields solid results and pinpoints actionable drivers of churn, providing a foundation for business decision-making and further analytics work.

Results






```
Accuracy: 0.8106956933270232
Confusion Matrix:
 [[1386  153]
 [ 247  327]]
Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.90      0.87      1539
           1       0.68      0.57      0.62       574

    accuracy                           0.81      2113
   macro avg       0.76      0.74      0.75      2113
weighted avg       0.80      0.81      0.81      2113
```