Customer Churn Analysis Report

Executive Summary

This comprehensive exploratory data analysis examines customer churn patterns using a telecommunications dataset containing **7,043 customer records** across **21 features**. The analysis successfully identifies key churn drivers through systematic data preprocessing, univariate and bivariate analysis, revealing critical insights for customer retention strategies.churn.ipynb

Dataset Overview

The dataset encompasses customer demographics, service subscriptions, account information, and billing details. Key features include gender, senior citizen status, partner/dependent information, tenure duration, contract type, internet and phone services, additional support services (online security, device protection, tech support), entertainment options (streaming TV/movies), payment methods, monthly charges, total charges, and the target variable - churn status.churn.ipynb

Data Characteristics

Numerical variables show the following distributions: average customer tenure is 32.4 months (ranging 0-72 months), mean monthly charges are $64.76 ($18.25-$118.75), and approximately 16.2% of customers are senior citizens.churn.ipynb

Data Preprocessing

The analysis implemented rigorous data cleaning procedures including handling missing values in the TotalCharges column by replacing blanks with zero and converting the data type from object to float for proper numerical analysis. Data quality checks confirmed no duplicate records exist in the dataset, ensuring analysis integrity. The SeniorCitizen feature was transformed from binary encoding (0/1) to categorical labels (Yes/No) for improved interpretability.churn.ipynb

Analytical Framework

The exploratory analysis employed multiple visualization techniques to uncover churn patterns across different customer segments:

Overall Churn Distribution - Count plots and pie charts revealed the proportion of churned versus retained customers, establishing baseline churn rates.churn.ipynb

Demographic Analysis examined churn rates segmented by gender and senior citizen status, identifying vulnerable customer populations.churn.ipynb

Tenure Analysis utilized histogram distributions to explore the relationship between customer longevity and churn probability, revealing critical retention timeframes.churn.ipynb

Contract Type Analysis investigated how contract duration (month-to-month, one-year, two-year) influences customer retention patterns.churn.ipynb

Service Subscription Analysis created comprehensive visualizations across multiple service dimensions including phone service, multiple lines, internet service type, online security, online backup, device protection, tech support, streaming TV, and streaming movies to identify service combinations associated with higher retention.churn.ipynb

Payment Analysis examined churn patterns across different payment methods (electronic check, mailed check, bank transfer, credit card) to identify payment behaviors correlated with churn.churn.ipynb

Financial Analysis explored the relationship between monthly payment amounts and churn likelihood through bar plot comparisons.churn.ipynb

Key Technical Implementations

The analysis leveraged Python's data science ecosystem including **NumPy** for numerical computations, **Pandas** for data manipulation and preprocessing, **Matplotlib** for foundational plotting capabilities, and **Seaborn** for advanced statistical visualizations. Visualization techniques included count plots with value labels for categorical distributions, pie charts for percentage breakdowns, histograms with hue separation for continuous variable analysis, and bar plots for comparative analysis.churn.ipynb

Analytical Insights Structure

The systematic approach progressed from univariate analysis (understanding individual feature distributions) to bivariate analysis (examining relationships between features and churn), enabling identification of multi-dimensional churn risk factors. The multi-subplot approach for

service features allowed efficient comparison across nine different service categories simultaneously, revealing cross-service patterns.churn.ipynb

Portfolio Recommendations

This analysis demonstrates strong **data cleaning**, **exploratory data analysis**, **statistical visualization**, and **business insight generation** capabilities. For portfolio presentation, consider adding: a summary dashboard with key churn rate metrics, correlation heatmap identifying feature relationships, customer segmentation profiles based on high/medium/low churn risk, actionable business recommendations derived from findings, and potential next steps including predictive modeling approaches (logistic regression, random forest, XGBoost) for churn prediction.churn.ipynb

The comprehensive nature of this EDA covering data quality assessment, demographic analysis, service utilization patterns, financial behavior, and contract analysis positions it as a strong foundation for advanced predictive modeling and strategic business decision-making in customer retention initiatives.churn.ipynb