Customer Churn Prediction Project — Executive Summary & README

Project Overview

This project predicts customer churn using the Telco Customer Churn dataset and frames it as a financial retention problem. It demonstrates an end-to-end data science workflow — from EDA and feature engineering to modeling, evaluation, and SHAP explainability — with a focus on interpretability and business value.

Objective

Predict which customers are likely to churn (close accounts or stop using services) and explain why, enabling proactive retention strategies.

Tools & Stack

Python • Pandas • NumPy • Scikit-Learn • XGBoost • SHAP • Matplotlib • Seaborn • Tableau • Jupyter Notebook

Workflow Summary

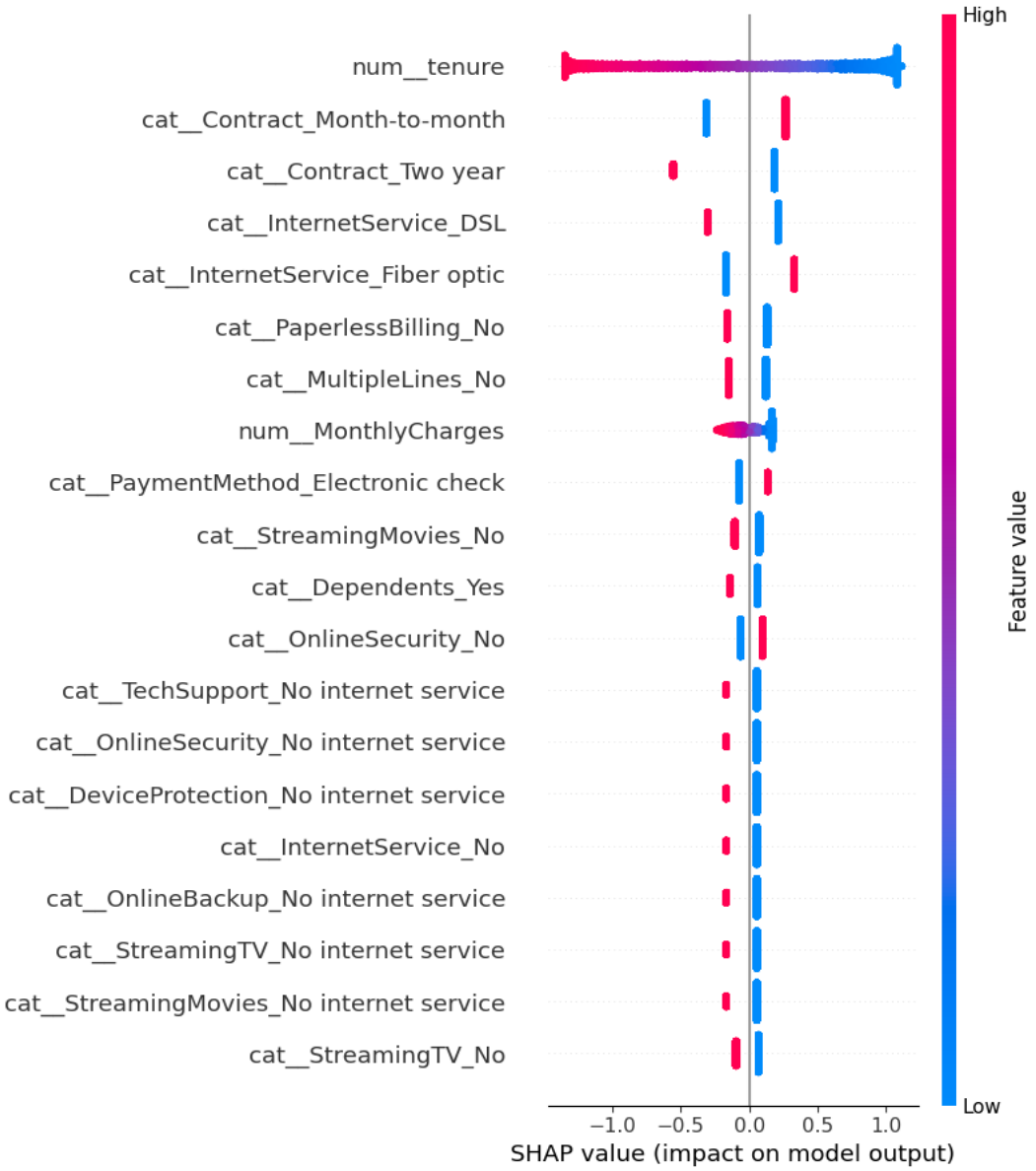
Step 1: Data Understanding & Cleaning — 7,043 rows × 21 columns. Cleaned missing TotalCharges, dropped customerID, encoded categoricals.
Step 2: Exploratory Data Analysis — Churn rate 26.5%. Higher churn for month-to-month contracts, high monthly charges, and no add-on services.
Step 3: Modeling — Logistic Regression, XGBoost, Random Forest compared using Accuracy, Recall, Precision, F1, and AUC.
Step 4: Explainability — Used SHAP to identify most impactful churn features.
Step 5: Insights & Actions — Translated model findings into retention strategies.

Modeling Results

Model	Accuracy	Recall	Precision	F1	ROC-AUC	Notes
Logistic Regression	0.803	0.548	0.654	0.596	0.845	Best balance & interpretability
XGBoost	0.785	0.521	0.611	0.562	0.823	Slightly lower AUC, more flexibility
Random Forest	0.788	0.454	0.640	0.531	0.821	Overfitting tendency

Explainability — SHAP Summary

Top positive churn drivers: short tenure, month-to-month contracts, high monthly charges, electronic check payment, and lack of online services (security/support).
Top negative drivers: long tenure, two-year contracts, lower charges, and bundled services.


num__tenure
cat__Contract_Month-to-month
cat__Contract_Two year
cat__InternetService_DSL
cat__InternetService_Fiber optic
cat__PaperlessBilling_No
cat__MultipleLines_No
num__MonthlyCharges
cat__PaymentMethod_Electronic check
cat__StreamingMovies_No
cat__Dependents_Yes
cat__OnlineSecurity_No
cat__TechSupport_No internet service
cat__OnlineSecurity_No internet service
cat__DeviceProtection_No internet service
cat__InternetService_No
cat__OnlineBackup_No internet service
cat__StreamingTV_No internet service
cat__StreamingMovies_No internet service
cat__StreamingTV_No
−1.0
−0.5
0.0
0.5
1.0
SHAP value (impact on model output)
High
Low
Feature value

Business Insights & Recommended Actions

Insight	Action
New customers churn more	Implement early onboarding and retention offers
Month-to-month contracts churn highest	Incentivize annual or multi-year plans
High monthly charges increase churn	Introduce tiered pricing or loyalty discounts
Customers lacking add-ons churn more	Bundle services (security, support, backup)
Electronic check users churn more	Promote digital autopay and paperless billing

Next Steps

- Tune XGBoost hyperparameters (learning rate, max depth)
- Adjust classification threshold to boost recall
- Develop Streamlit app for live churn scoring
- Track experiments using MLflow