Customer Churn Analysis for Telecom Industry

# Introduction

Customer churn refers to the percentage of subscribers who discontinue their services over a given period. In the highly competitive telecom sector, retaining customers is more cost-effective than acquiring new ones. This project focuses on analyzing factors that drive churn and building a predictive model to identify customers at risk.

# Abstract

The objective of this project was to analyze telecom customer data to identify patterns in churn behavior and predict the likelihood of a customer leaving. Using a Random Forest Classifier, we modeled churn based on features such as contract type, monthly charges, and internet service. The model achieved an overall accuracy of 80%, with detailed EDA and SHAP visualizations aiding interpretability. Actionable insights were derived for reducing churn and improving customer retention.

# Tools Used

- Python: pandas, seaborn, matplotlib, scikit-learn, SHAP  
- IDE: Google Colab  
- Libraries: SHAP for explainability, seaborn/matplotlib for visualization  
- Dataset: Telco Customer Churn Dataset (Kaggle) [LINK](https://drive.google.com/file/d/1lxlqeuc1XsfNJCMlUlGwzuVAgyiit-xt/view?usp=sharing)

# Steps Involved

1. Data Preprocessing  
- Converted 'TotalCharges' to numeric and filled missing values with median.  
- Encoded categorical features using LabelEncoder.  
- Removed irrelevant column 'customerID'.  
  
2. Exploratory Data Analysis  
- Churn Distribution: ~27% customers churned.  
- Contract Type: Customers on month-to-month contracts had significantly higher churn.  
- Monthly Charges: Higher charges correlated with increased churn.  
- Internet Service: Fiber optic users showed greater churn risk.  
- Correlation Matrix: Tenure and contract length negatively correlated with churn.

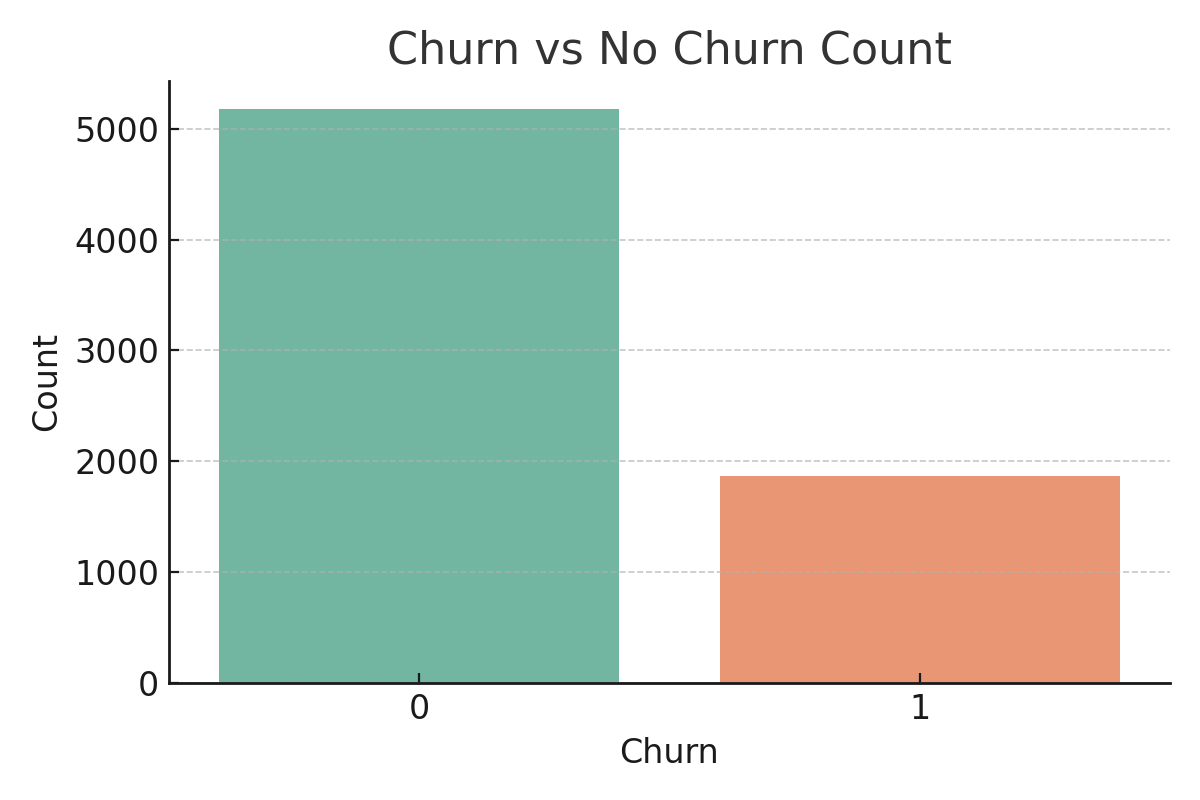
3. Model Building  
- Model: Random Forest Classifier  
- Test Accuracy: 80%  
- Classification Report:  
 - Precision: 83% (Class 0), 67% (Class 1)  
 - Recall: 92% (Class 0), 48% (Class 1)  
 - F1-score: 87% (Class 0), 56% (Class 1)  
  
4. Model Explainability  
- Used SHAP summary plots to identify feature importance.  
- Most influential features: tenure, Contract, MonthlyCharges, and InternetService.  
  
5. Customer Segmentation  
- Created rule-based segments: 'At Risk', 'Loyal', 'Dormant', and 'Others'.  
- Segmentation based on churn probability, tenure, and monthly charges.  
- Visualized segment counts and average churn probabilities.

# Final Recommendations

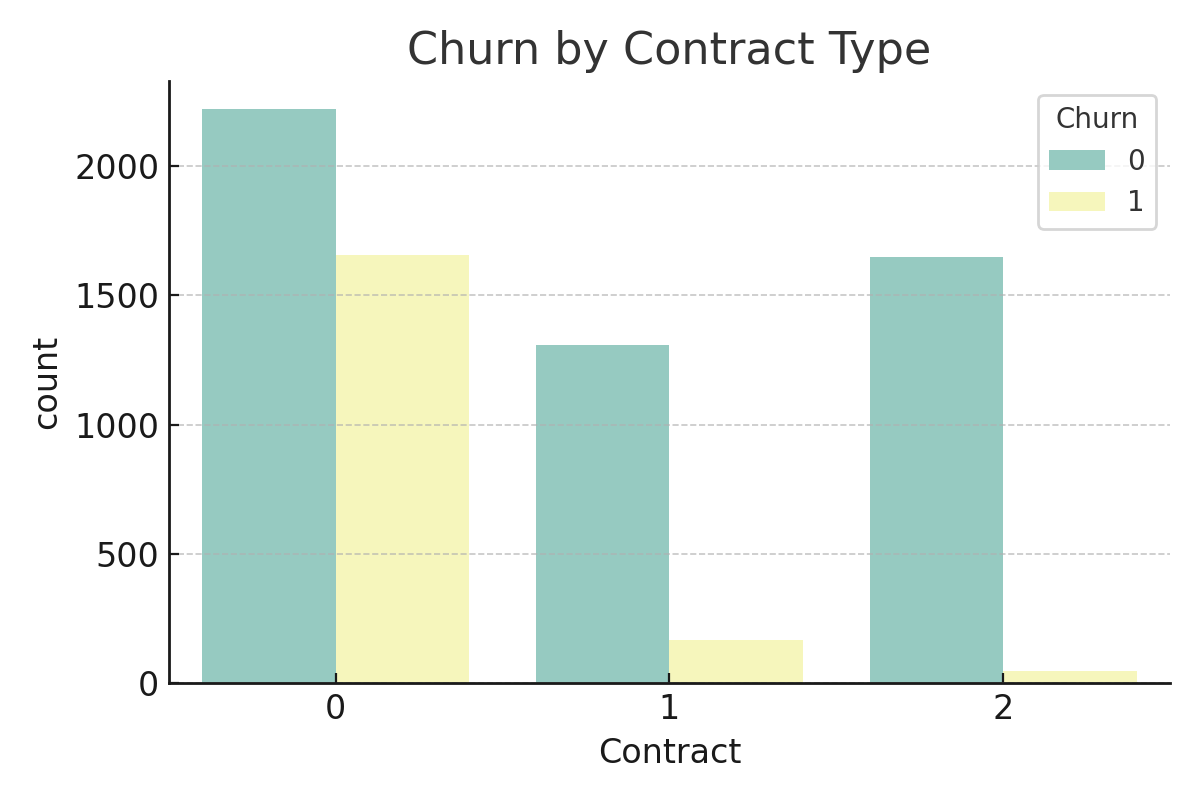
- Encourage customers to shift from month-to-month to annual contracts through loyalty discounts.  
- Target 'At Risk' customers (high churn probability + short tenure) with personalized retention offers.  
- Monitor customers with high charges and low engagement as potential churn triggers.  
- Improve service quality and support for fiber optic users, as they show higher churn.  
- Use churn probability and customer segmentation in CRM systems to automate follow-ups and offers.

# Key Visualizations

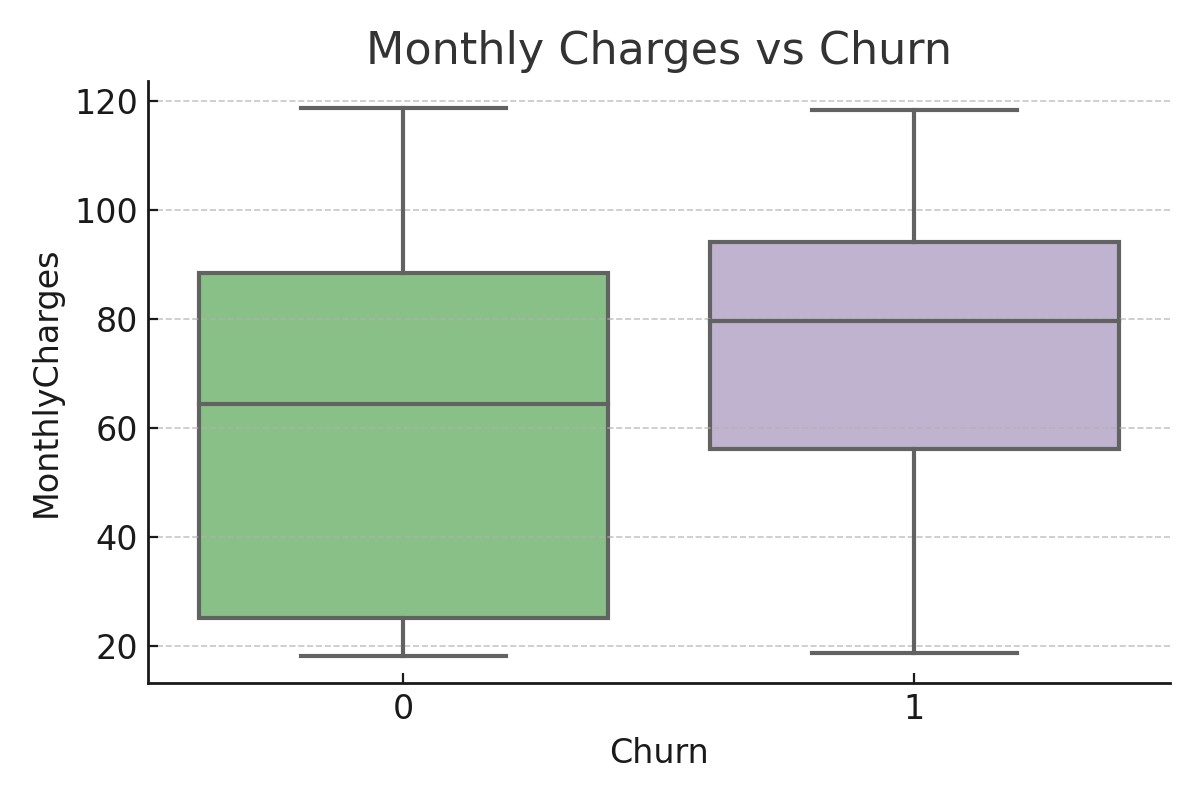
**1. Churn vs No Churn Count**



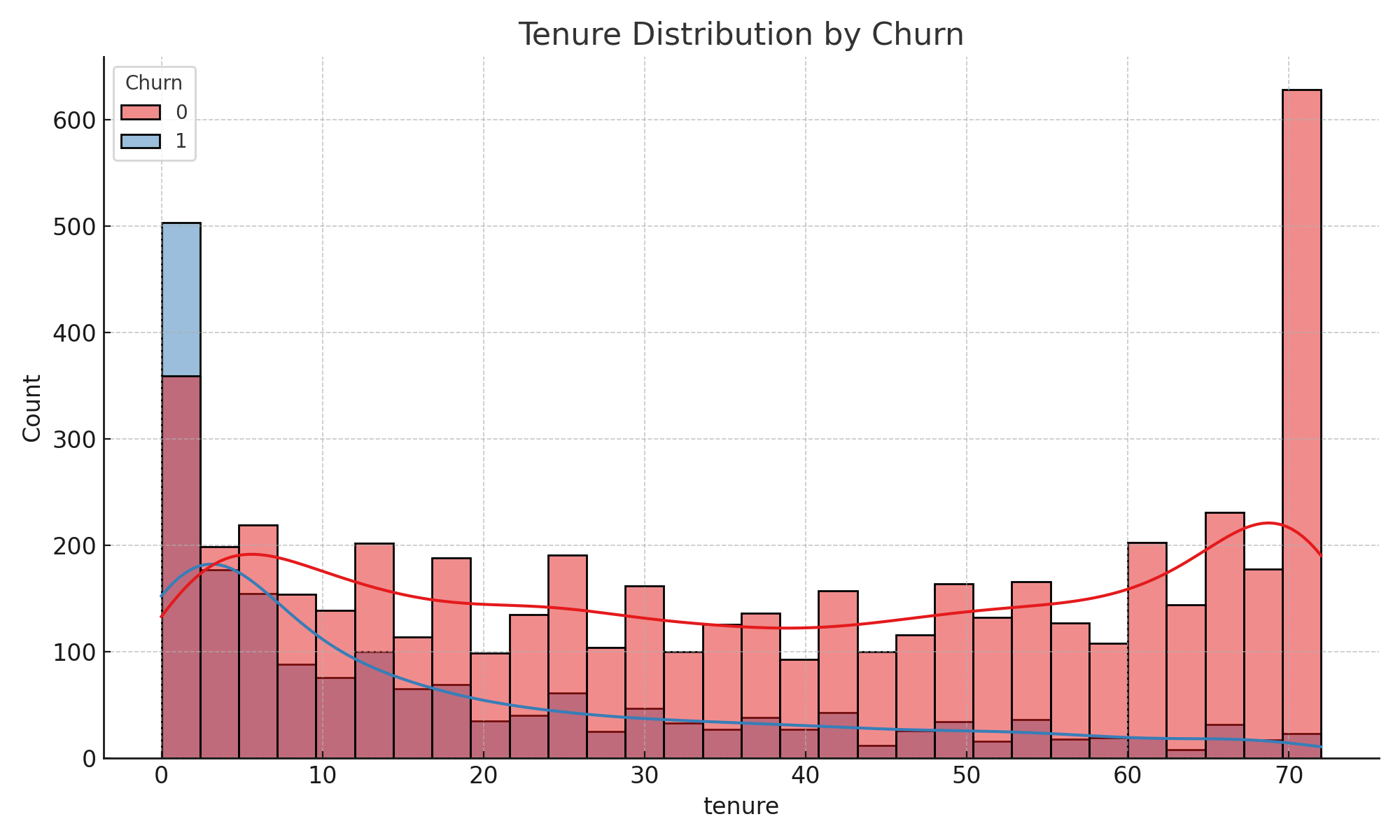
**2. Churn by Contract Type**



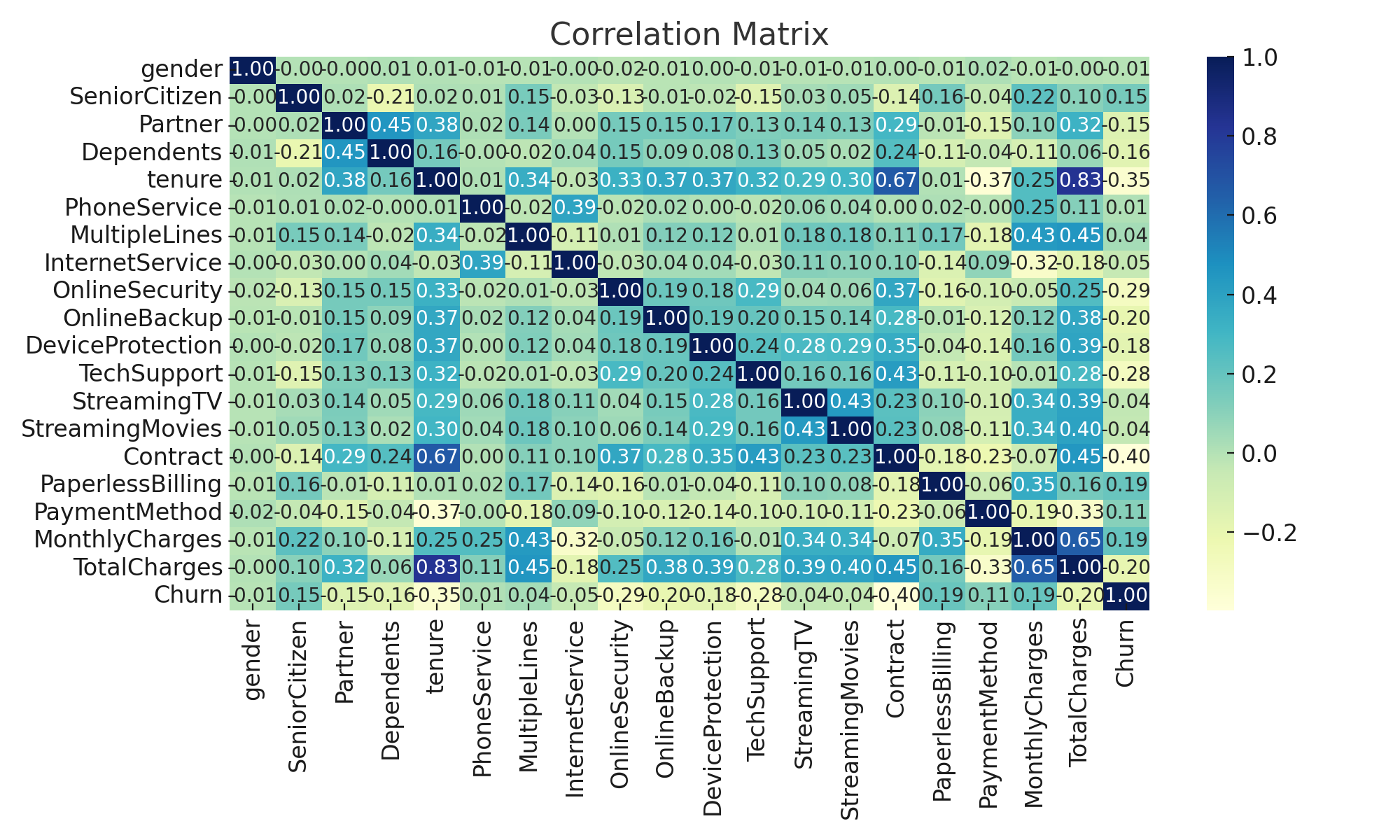
**3. Monthly Charges vs Churn**



**4. Tenure Distribution by Churn**



**5. Correlation Matrix**



# Conclusion

This project provided deep insights into the factors contributing to customer churn in the telecom sector. The Random Forest model, supported by visual analytics and SHAP explanations, demonstrated that shorter contract types, higher monthly charges, and specific internet services significantly increase churn risk. Telecom companies can reduce churn by promoting long-term contracts, offering targeted discounts, and improving service quality for at-risk groups.