## CUSTOMER CHURN PREDICTION FOR TELCO COMPANY

**Team Members:**

Silvana Ruiz Illas

Jericho Matulin

Richard (Colses) Edmunds

Ugochukwu Ibecheozor

Rodrigo Salomon

### Business Problem

Telco Company X, a prominent player in this market, faces a critical challenge –customer churn. In a highly competitive industry, customers consider various factors beyond premiums when choosing their service provider. Since acquiring a new customer costs 5-10 times more than retaining an existing one, customer retention has become even more important than customer acquisition. Hence, we need to understand which customers may churn so retention strategies can be devised accordingly.

### Business Objective

In this project, our business objective is to analyse your firm's customer-level data, build predictive models to identify customers at high risk of churn and identify the leading indicators. Thus, our focus would be on.

1. Retaining highly profitable customers.
2. Predicting which customers are at high risk of churn (to devise customer retention strategies accordingly) and optimise engagement with them.
3. Improve overall customer satisfaction and loyalty.

### Methodological Approach

**Data Cleaning & Pre-Processing**

* Check for Missing Values
* Check for Duplicates
* Type Casting Attributes
* Value Replacement
* Transformation of Critical Attributes

**Exploratory Data Analysis**

* Feature Analysis: Correlation Matrix and Box Plots
* Uni-variate Analysis: Histograms and Bar Graphs
* Bi-variate Analysis: Using *Group By* Clause
* Label Encoding of Categorical Variables

**Visualizations**

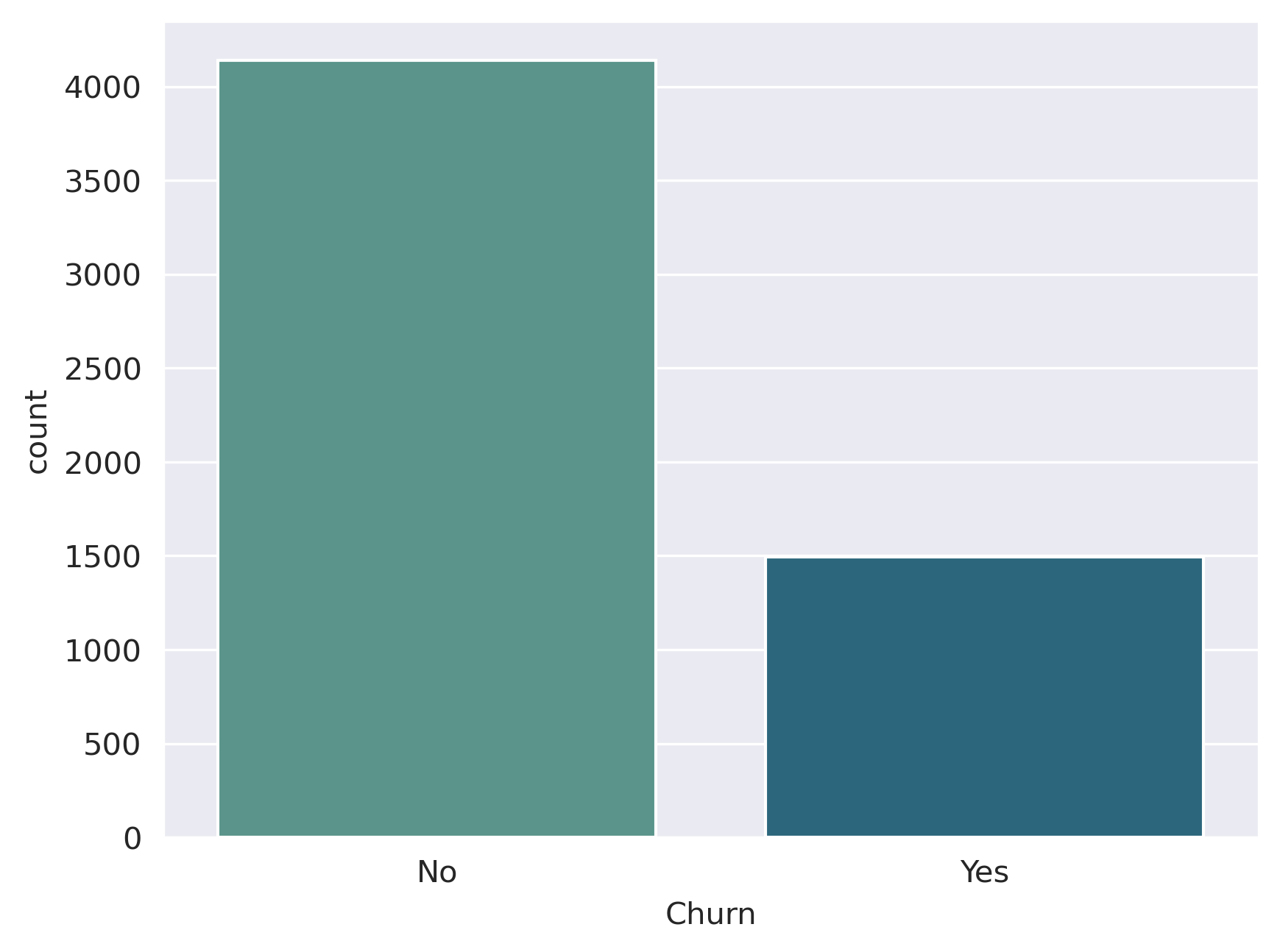
****

Fig 1: Original Churn distribution

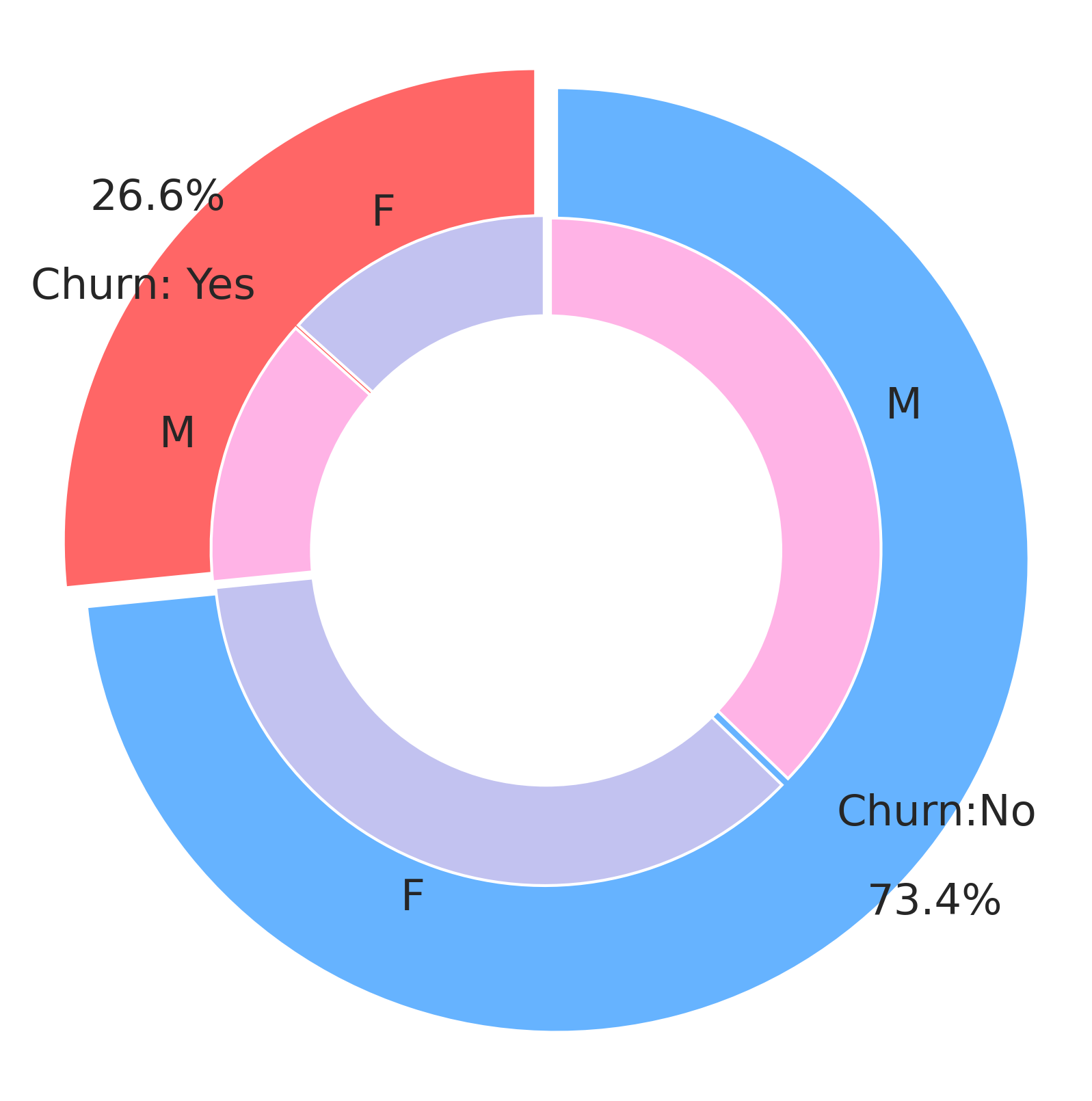
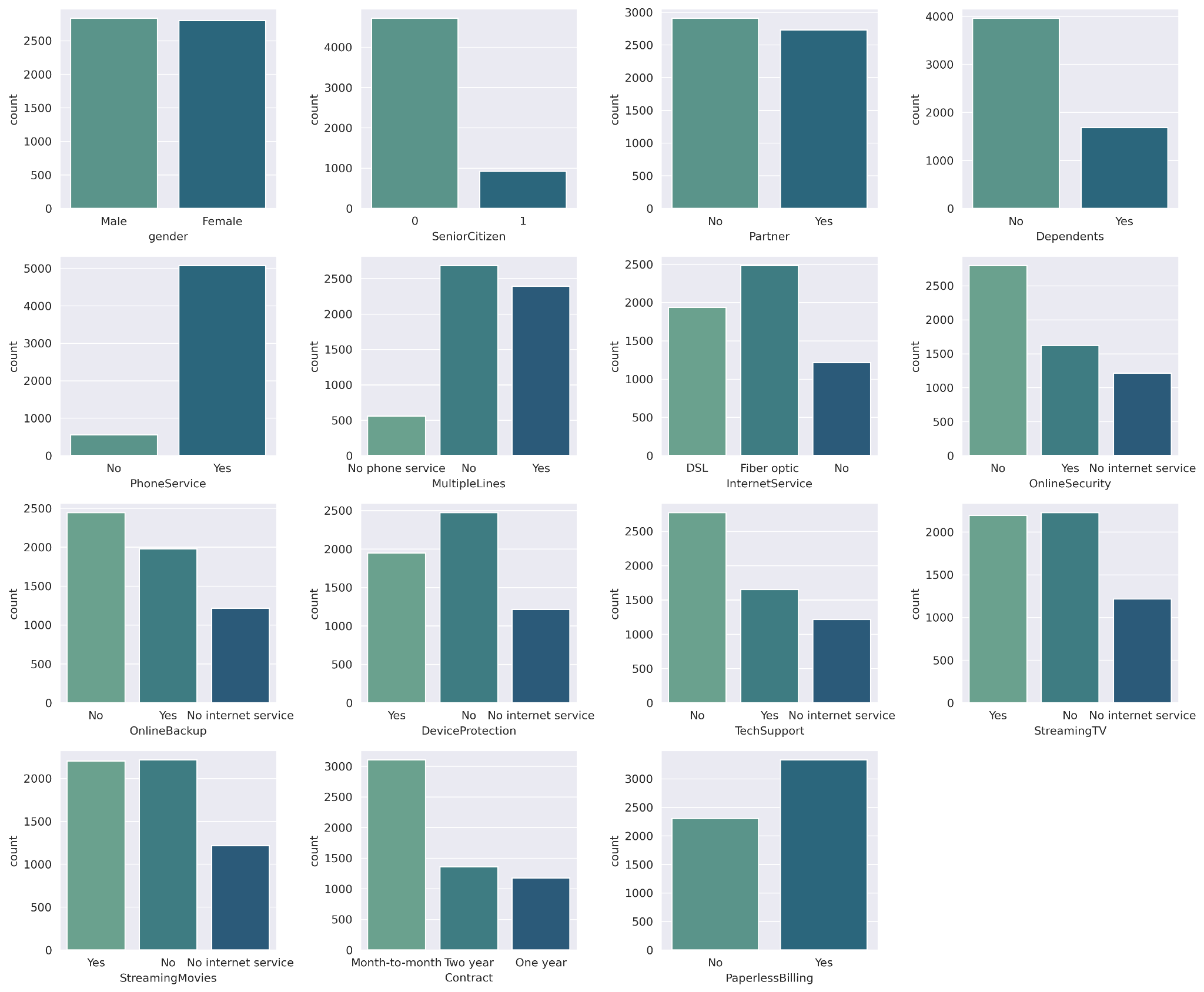
****

Fig 2: Churn distribution with respect to Gender - Male (M), Female (F)



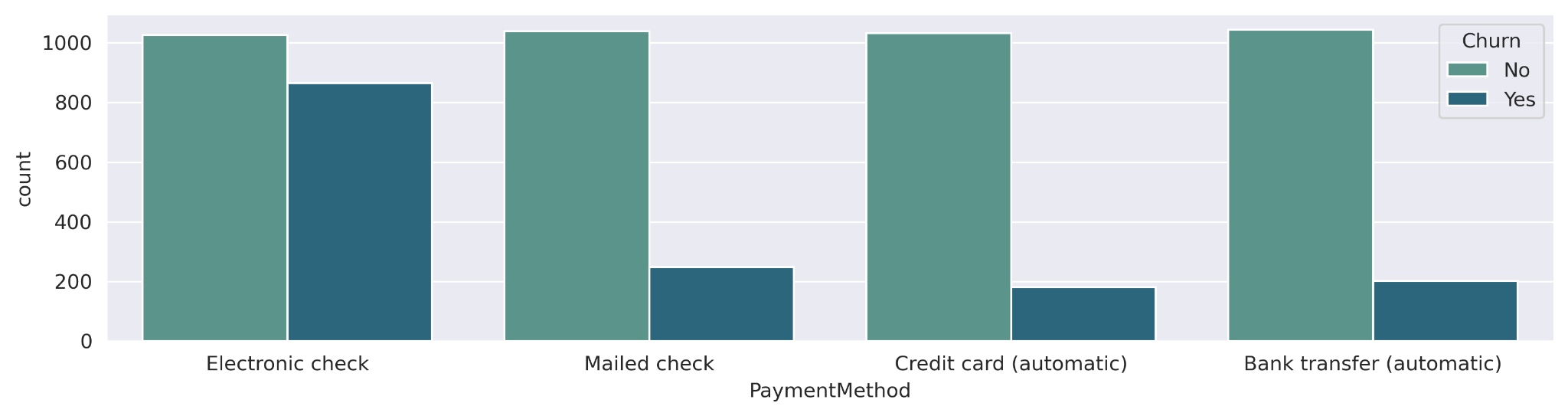


Fig 3: Categorical features count plot

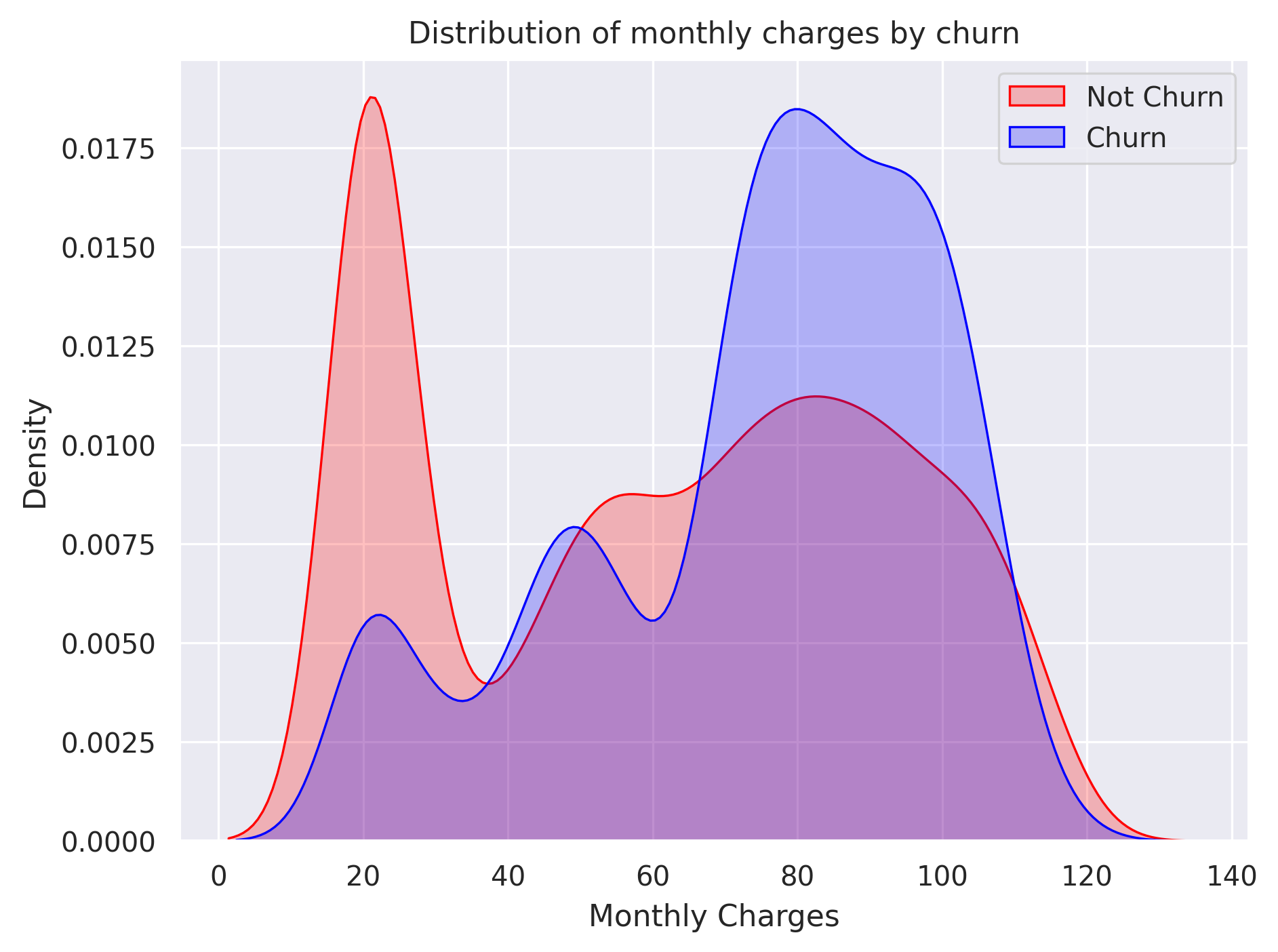


Fig 4: Distribution of monthly charges by churn

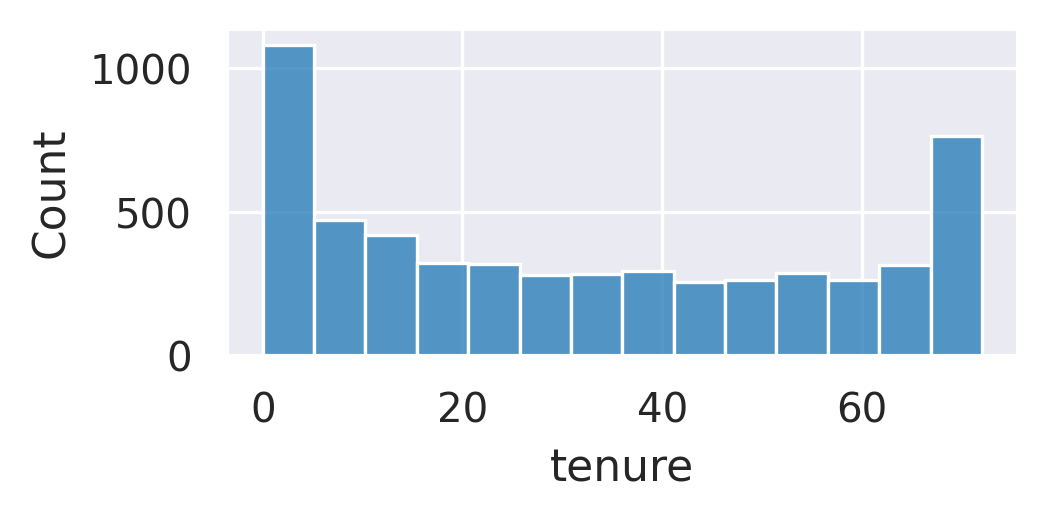


Fig 5: Distribution of tenure by churn

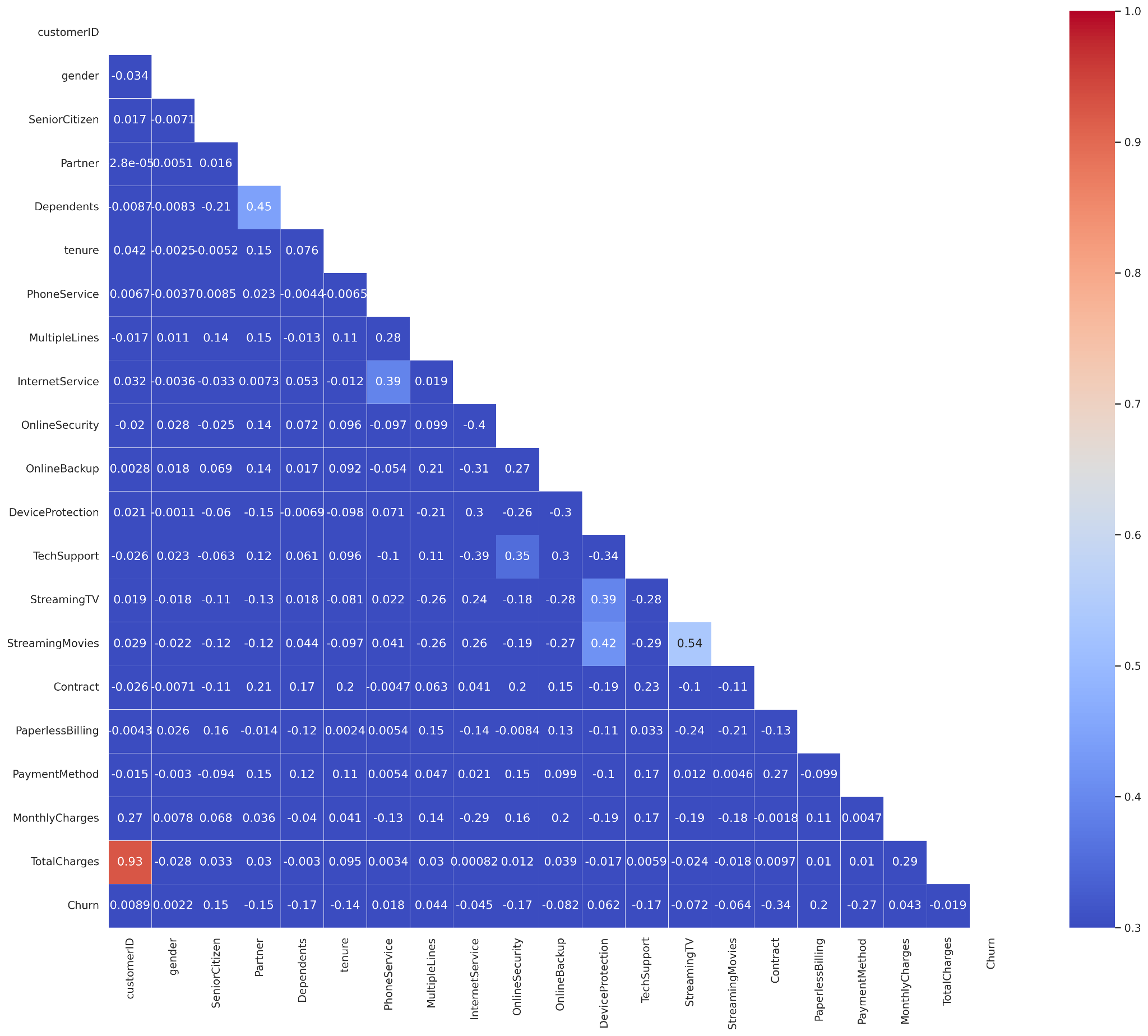


Fig 6: Correlation matrix

The correlation between the numeric features (tenure, MonthlyCharges, TotalCharges, and SeniorCitizen) is relatively low, which means these variables are not highly collinear, making them all potentially valuable for predictive modeling.

**Feature Engineering**

From the above visualizations, we noticed that new customers are more likely to churn and this indicates that customer retention efforts might be most effective early in the customer lifecycle.

* Binning the tenure feature into six (6) ranges.
* High Monthly Charges (>= 75th percentile): Binary indicator for high-paying customers
* Contract Length Indicator: Binary for month-to-month contract vs longer contracts
* Auto-pay Feature: Binary for automatic payment methods

**Handling Data Imbalance**

As we have learned in class, data imbalance affects machine learning models by tending only to predict the majority class and ignoring the minority class. Hence, the minority class is significantly misclassified compared to the majority class. Thus, we use techniques to balance class distribution in the data.

Compared to fig.1, we now managed to have a balanced data set.

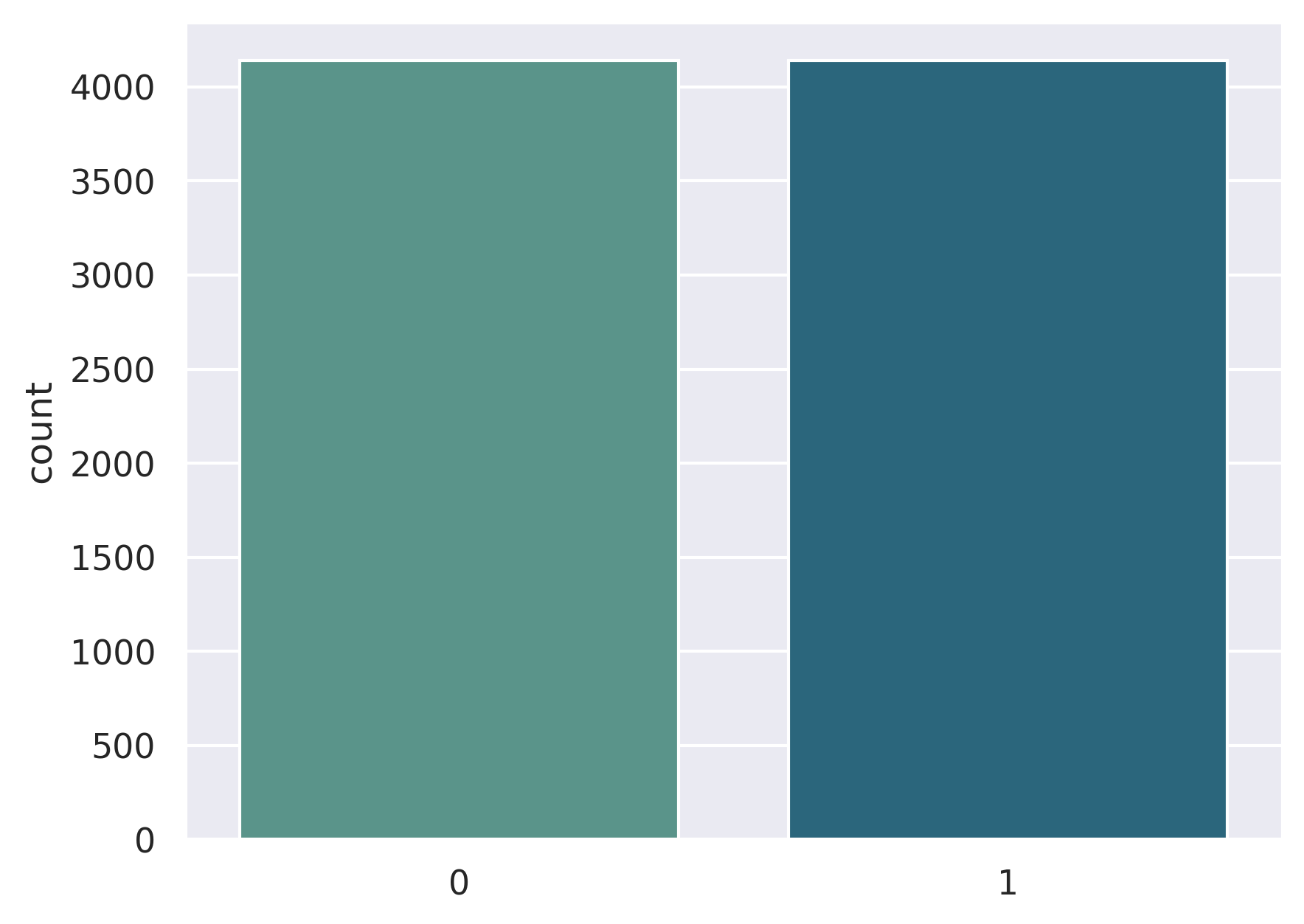


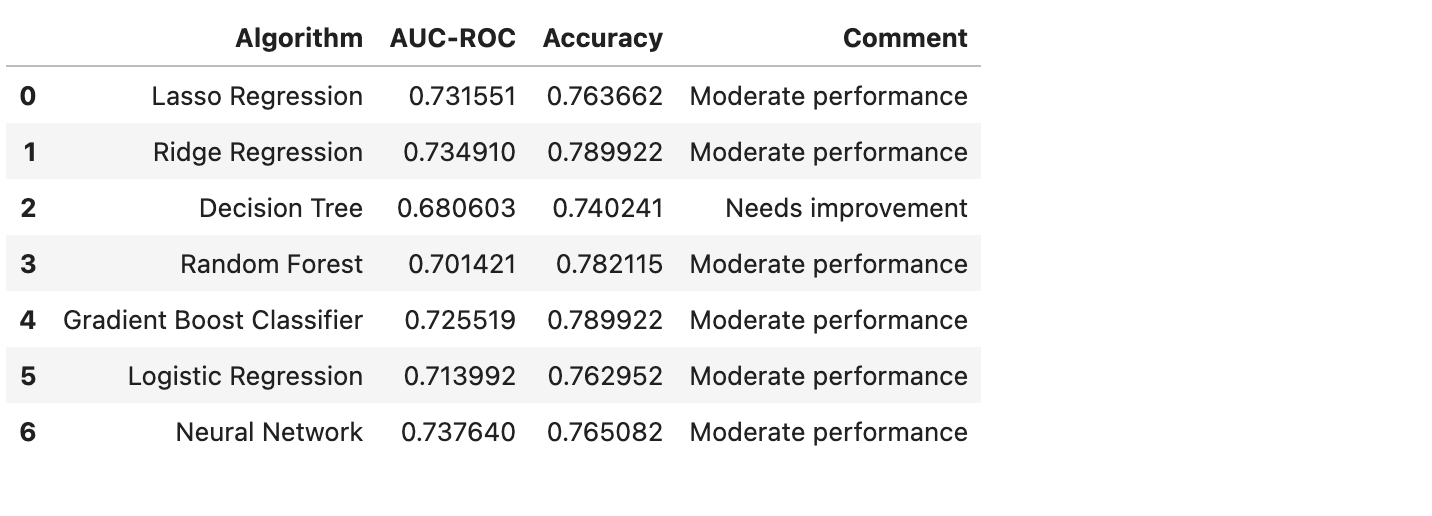
Fig 7: Handled Imbalanced Data

### Model Training

We have used the following ML & DL models to train our dataset.

* Lasso Regression
* Ridge Regression
* Decision Tree
* Random Forest
* Rndom Forest
* Gradient Boost Classifier
* Logistic Regression
* Neural Network

Below is the summary output of all the models used in the process.



The AUC-ROC (Area Under the Receiver Operating Characteristic Curve) score is a performance measurement for classification models, and it tells us how well the model distinguishes between churners and non-churners.

1. **Lasso Regression (0.731)**:
   * This model has a **moderate ability** to differentiate between customers who churn and those who don’t. It is performing well but can still be improved. About **73% of the time**, this model correctly ranks a churner higher than a non-churner.
2. **Ridge Regression (0.735)**:
   * Similar to Lasso, Ridge regression is **moderately effective** in predicting churn. It performs slightly better than Lasso, with **73.5% accuracy in ranking churners higher than non-churners**.
3. **Decision Tree (0.682)**:
   * The decision tree model is less effective, with an AUC-ROC of **68.2%**, meaning it struggles more to distinguish between churners and non-churners compared to other models. It is still better than random guessing, but there's room for improvement.
4. **Random Forest (0.704)**:
   * The random forest performs better than the decision tree but is still **moderately effective**, with a **70.4% ability to rank churners higher** than non-churners. It's a more reliable model than the decision tree but not as strong as Ridge or Neural Networks.
5. **Gradient Boosting Classifier (0.725)**:
   * The gradient boosting model performs well, with an AUC-ROC of **72.5%**, indicating that it is **quite good at distinguishing churners**, and even stronger than the Random Forest and Logistic Regression model.
6. **Logistic Regression (0.713)**:
   * Logistic regression performs similarly to gradient boosting, with an AUC-ROC of **71.3%**. It is **moderately effective** at predicting churn and does slightly better than random forest and gradient boost.
7. **Neural Network (0.738)**:
   * The neural network is the **best performing model** here, with an AUC-ROC of **73.8%**, indicating that it is the **most accurate in distinguishing churners** from non-churners. It has a higher chance of correctly identifying at-risk customers.

**AUC-ROC around 0.73** (e.g., Lasso, Ridge, Gradient Booster, Neural Network): These models are **good at predicting churn**. They perform better than others and can be trusted for reasonably accurate churn predictions.

**AUC-ROC between 0.68 and 0.72** (e.g., Decision Tree, Random Forest, Logistic Regression): These models are **moderately effective** but may miss more churners compared to the stronger models.

### Recommendations

* **Improved Customer Retention:** Identified high-value customers for targeted promotional offers and loyalty programs, leading to increased customer retention and reduced churn rates.
* **Enhanced Marketing Effectiveness:** Enabled data-driven allocation of marketing resources towards high-value customer segments, maximizing return on Investment.
* **Data-driven Decision Making:** Empowered stakeholders with CLV insights and interactive visualizations, facilitating informed decisions regarding customer acquisition, retention, and overall business strategy.
* **Enhanced User Experience:** Provided a user-friendly Q&A interface for easy access to information, promoting data democratization and knowledge sharing within the organization reducing the time required for developing efficient SQL queries.

Based on the churn analysis, we identified that customers with shorter tenure are most likely to leave the service. To reduce churn, we recommend the company focus on

1. Focus on Retaining New Customers

* **Insight**: Customers with shorter tenures are more likely to churn.
* **Recommendation**: Develop targeted retention strategies for new customers, such as offering incentives like discounts, additional services, or loyalty programs during the early months of their contracts to increase engagement and satisfaction.

2. Promote Long-Term Contracts

* **Insight**: Customers on month-to-month contracts are more likely to churn compared to those on annual contracts.
* **Recommendation**: Offer attractive discounts or incentives for customers to switch from month-to-month contracts to annual or long-term plans. This helps increase customer commitment and reduces churn.

4. Enhance Online Security and Support Services

* **Insight**: Customers who do not subscribe to services like online security, technical support, and backup solutions are more likely to churn.
* **Recommendation**: Bundle essential services (e.g., online security, technical support) into standard service plans or offer them at discounted rates to increase customer value perception and reduce the likelihood of churn.

5. Improve Customer Experience in High-Charge Segments

* **Insight**: Customers paying higher monthly charges may have higher expectations, and churn may be higher in these segments if expectations aren't met.
* **Recommendation**: Focus on improving the experience for high-paying customers by providing premium support, faster internet options, or exclusive service offerings. Identifying and addressing pain points for these customers can prevent them from seeking alternative providers.