Phase 3 Project

Submitted by Troye Gilbert.

TOC

Overview

Reducing Customer Churn is a major objective for any firm. An additional possible source of income for every firm is the ability to predict customer churn, commonly referred to as customer attrition. The cost to the firm is impacted by customer churn. Increased customer churn results in revenue loss and higher marketing expenses because it takes longer to acquire new consumers.

Problems to solve

1. Detecting early indicators of churn, such as changes in customer behavior, reduced engagement, or declining usage.

2. Understanding different customer segments and their propensity to churn. Different strategies might be needed for each segment.

3. Monitoring and benchmarking against competitors to understand how competitive offerings and marketing strategies may impact customer churn.

4. Ensuring pricing aligns with customer value perception and market conditions. Revising pricing structures or offering more value-based pricing can influence customer retention.

Business Objective

To analyze the past data and predict whether the customer will churn or not in the next 6 months. This would help the bank to have the right engagement with customers at the right time.

Data Understanding

Data Understanding

The train and test data was obtained from Kaggle. The data was quite clean so no much data cleaning was done apart from handling the outliers and dealing with the unbalanceness of the data. The data dictionary of the dataset is as follows:

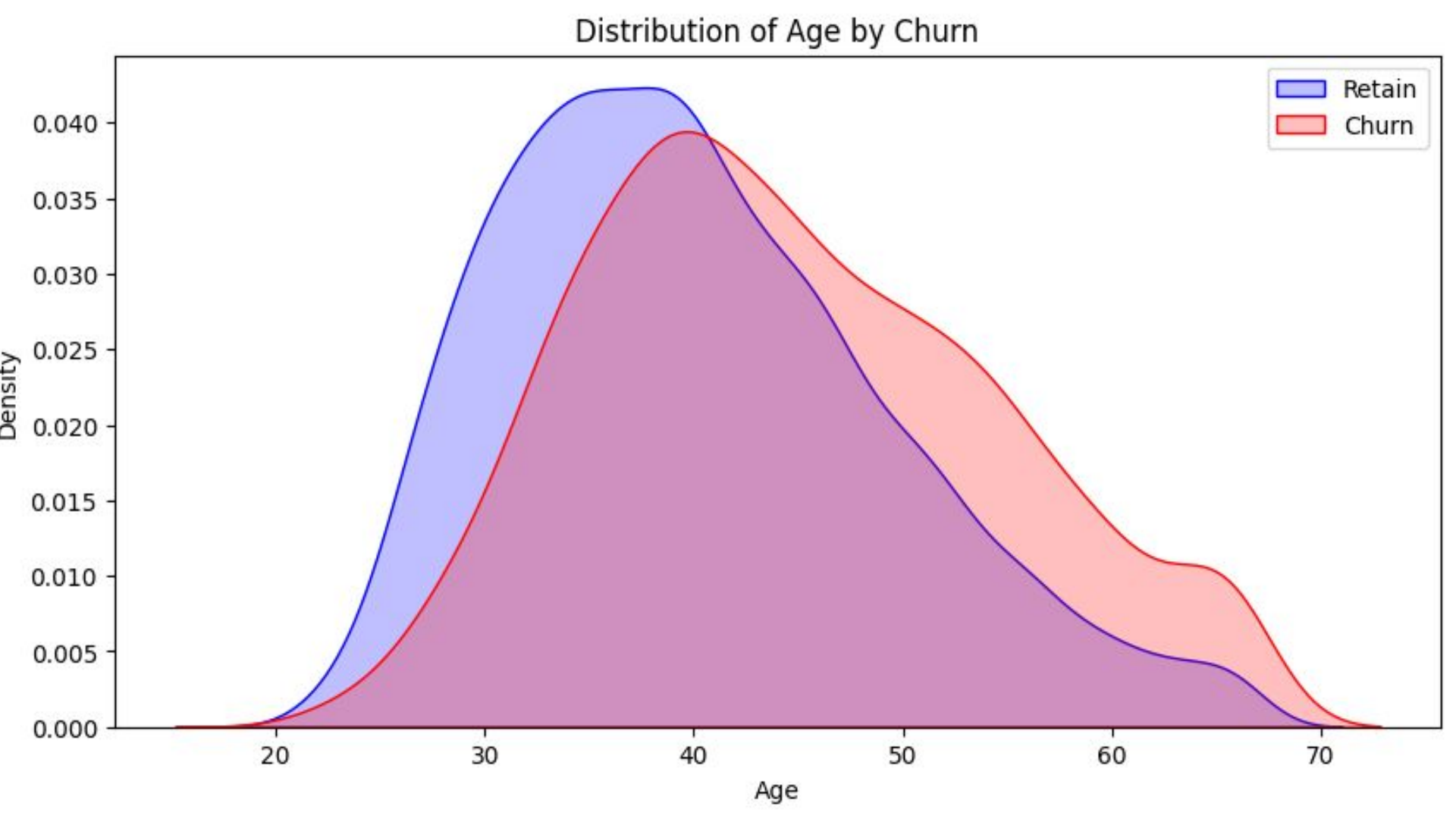
- **Unique Identifier:** A unique identifier for each customer row.
- **Age:** Age of the customer, which can impact banking behavior.
- **Gender:** Customer's gender, contributing to demographic analysis.
- **Income:** Yearly income, a key metric to evaluate customer spending patterns.
- **Balance:** Average quarterly balance in the bank, indicative of customer activity.
- **Vintage No.:** Number of years the customer has been associated with the bank.
- **Transaction Status:** Indicates whether the customer has conducted transactions in the last three months.
- **Product Holdings:** Number of products the customer holds with the bank.
- **Credit Card:** Whether the customer holds a credit card or not, reflecting financial engagement.
- **Credit Category:** Classification based on credit score, offering insights into financial stability.
- **Is_Churn (Target Variable):** Indicates if the customer is likely to churn in the next 6 months, crucial for retention strategies.

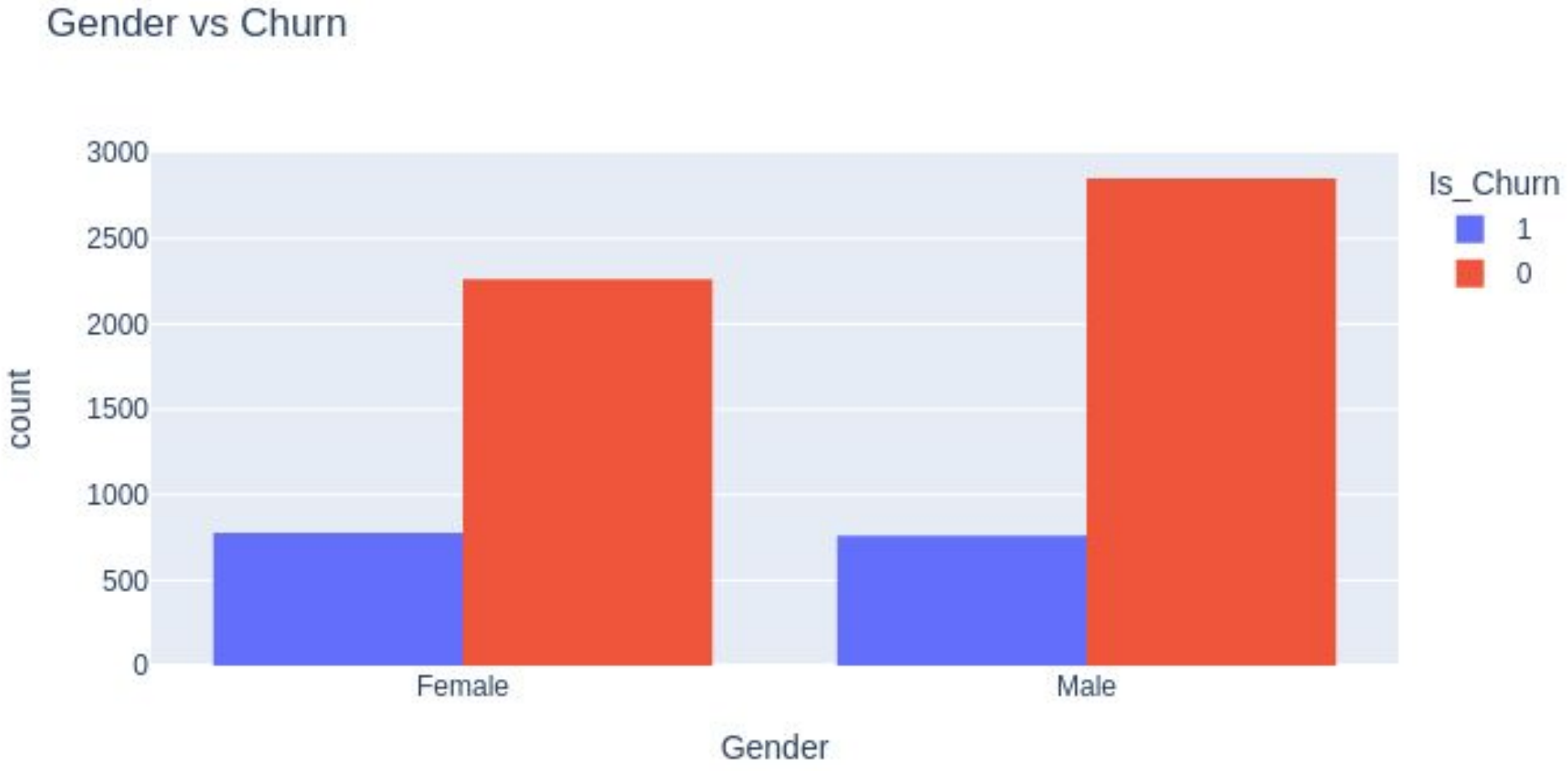
EXPLORATORY DATA ANALYSIS

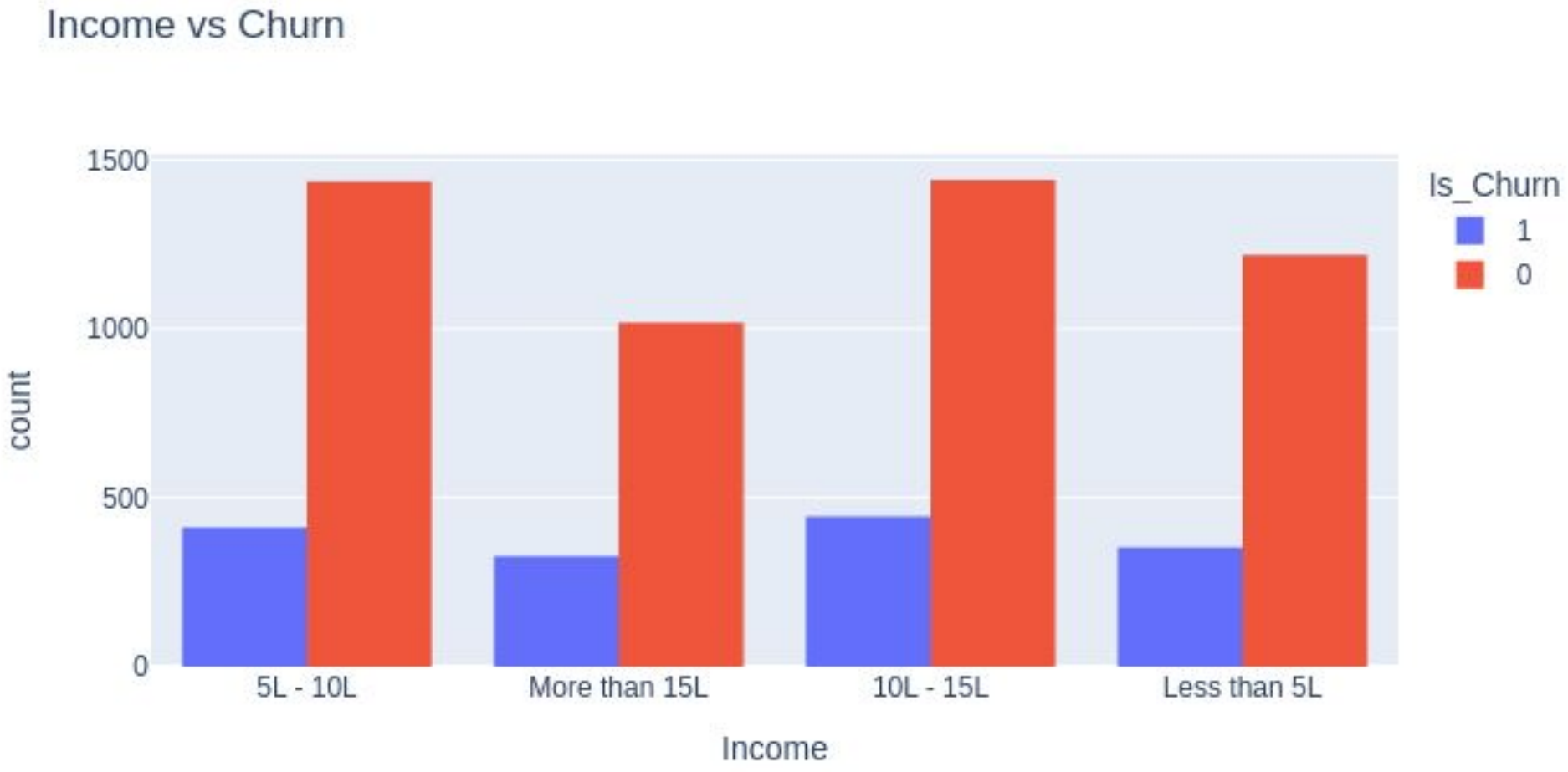
EDA is a crucial step in the data analysis process. It involves examining and summarizing the main characteristics of a dataset. EDA helps identify patterns, trends, relationships, and potential anomalies.

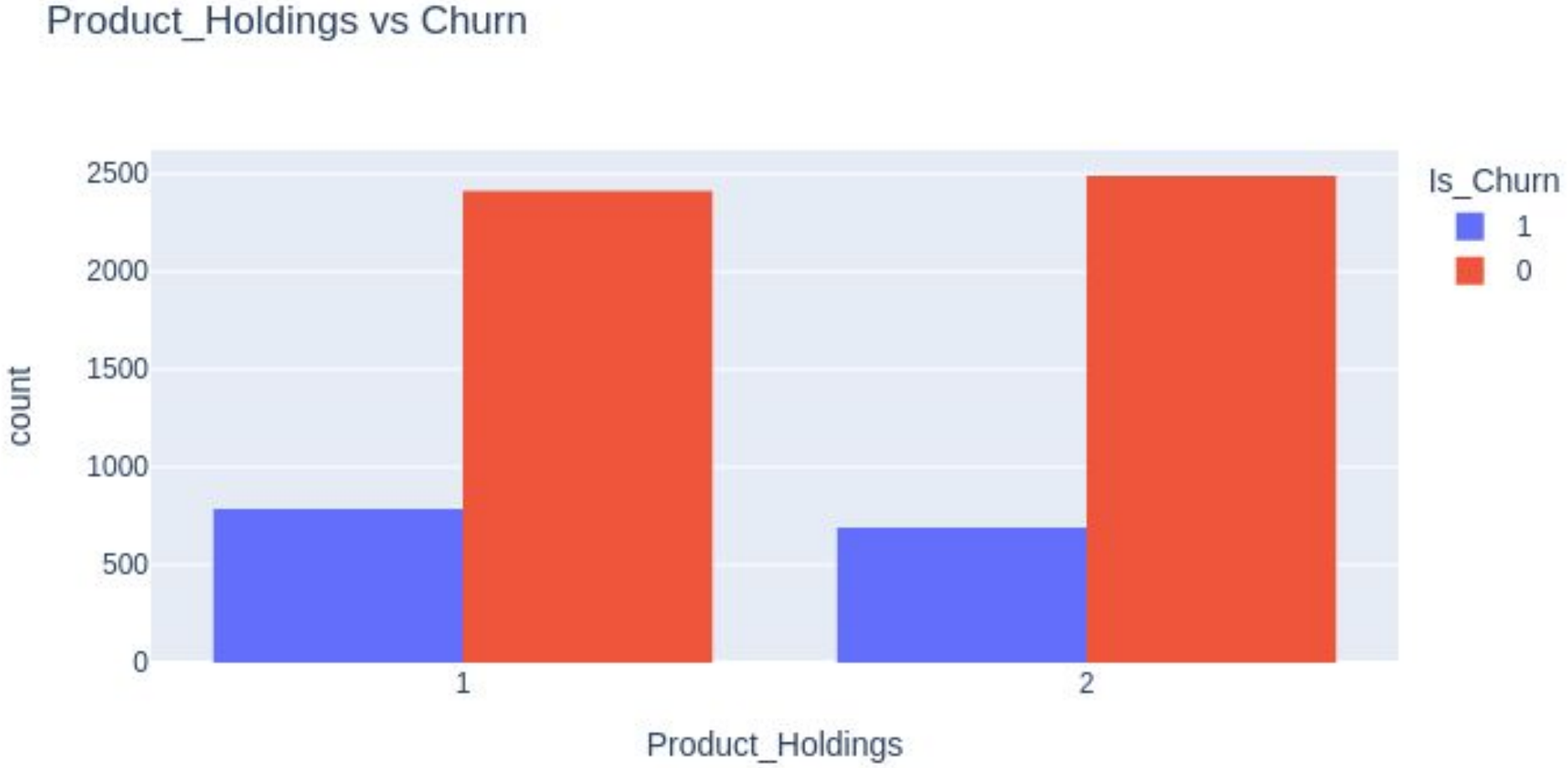
Techniques Used in EDA:

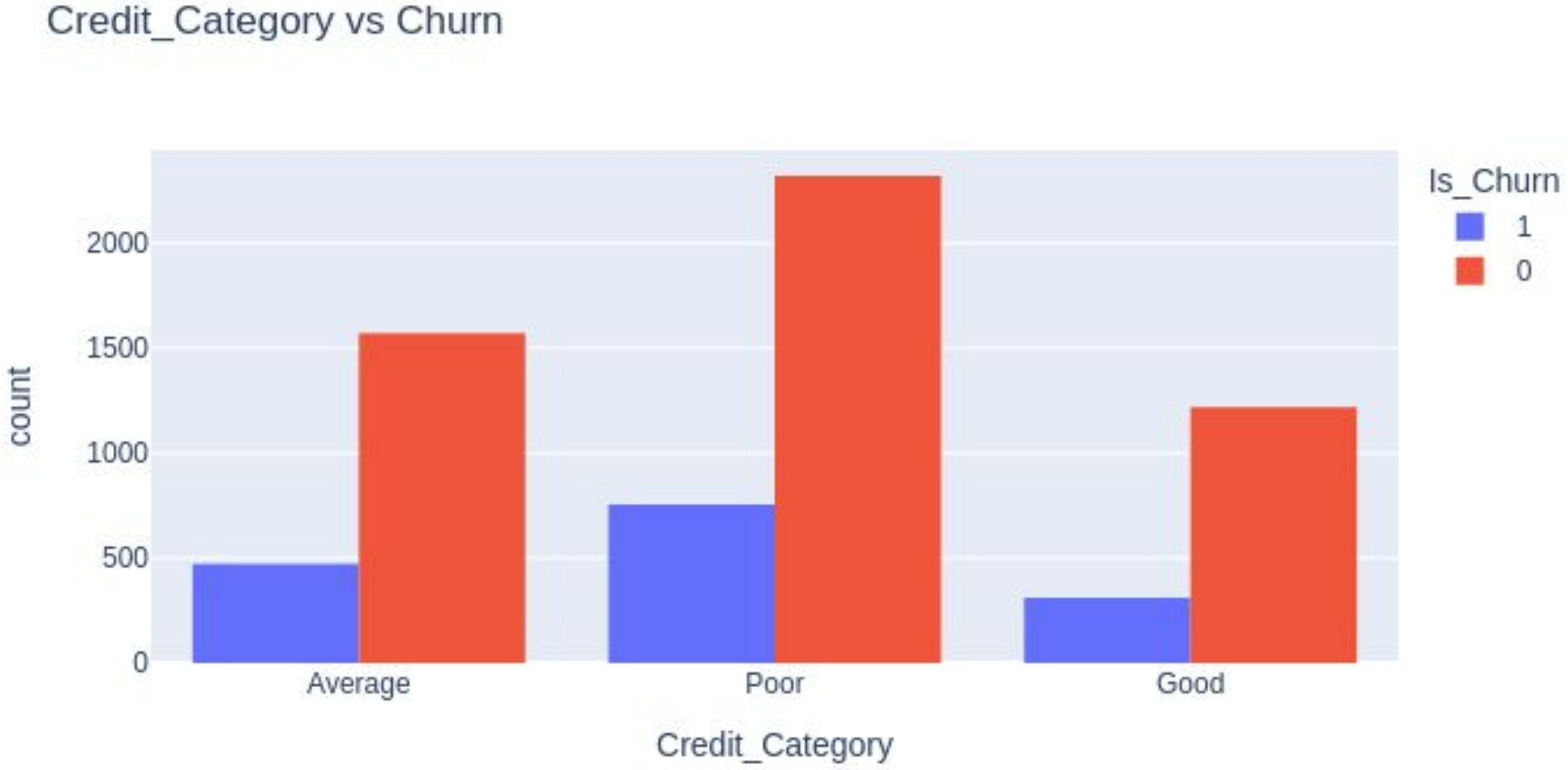
1. **Descriptive Statistics:**
   - Provide an overview of the central tendency, dispersion, and shape of data distribution.
2. **Data Visualization:**
   - Histograms, box plots, scatter plots, and pair plots help visualize data relationships and distributions.
3. **Correlation Analysis:**
   - Examine relationships between numeric variables using correlation matrices or heatmaps.


Distribution of Age by Churn
Retain
Churn
Density
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
20
30
40
50
60
70
Age


Gender vs Churn
3000
2500
2000
1500
1000
500
0
count
Female
Male
Gender
Is_Churn
1
0


Income vs Churn
1500
1000
500
0
count
5L - 10L
More than 15L
10L - 15L
Less than 5L
Income
Is_Churn
1
0


Product_Holdings vs Churn
2500
2000
1500
1000
500
0
count
1
2
Product_Holdings
Is_Churn
1
0


Credit_Category vs Churn
Is_Churn
1
0
count
2000
1500
1000
500
0
Average
Poor
Good
Credit_Category

Deductions

Gender and Churn:

- A greater percentage of females are likely to churn compared to males.
- This observation can provide insights into gender-specific behaviors or satisfaction levels that impact churn.

Income and Churn:

- Higher-income customers with earnings greater than 10 Lakhs are more likely to churn.
- This may suggest that affluent customers have higher expectations or alternative options available, leading to higher churn rates.

Product Holdings and Churn:

- Almost equal number of customers have 1 to 2 product holdings.
- There is no clear distinction in churn rates based on the number of product holdings.
- This indicates that the number of product holdings may not be a strong determinant of churn.

Credit Rating and Churn:

- Customers with poor credit ratings dominate the dataset and are also more likely to exit.
- This suggests that credit rating might influence customer behavior or satisfaction levels, contributing to churn.

Data Preparation

- We will process, transform, and manipulate to create the final dataset for modeling. Features may be selected, and engineering new features might occur to improve the performance of the models.
- We will alter the dataset to make it model ready. We are going to perform the following operations:

1. Encode Categorical Features.

2. Feature Scaling.

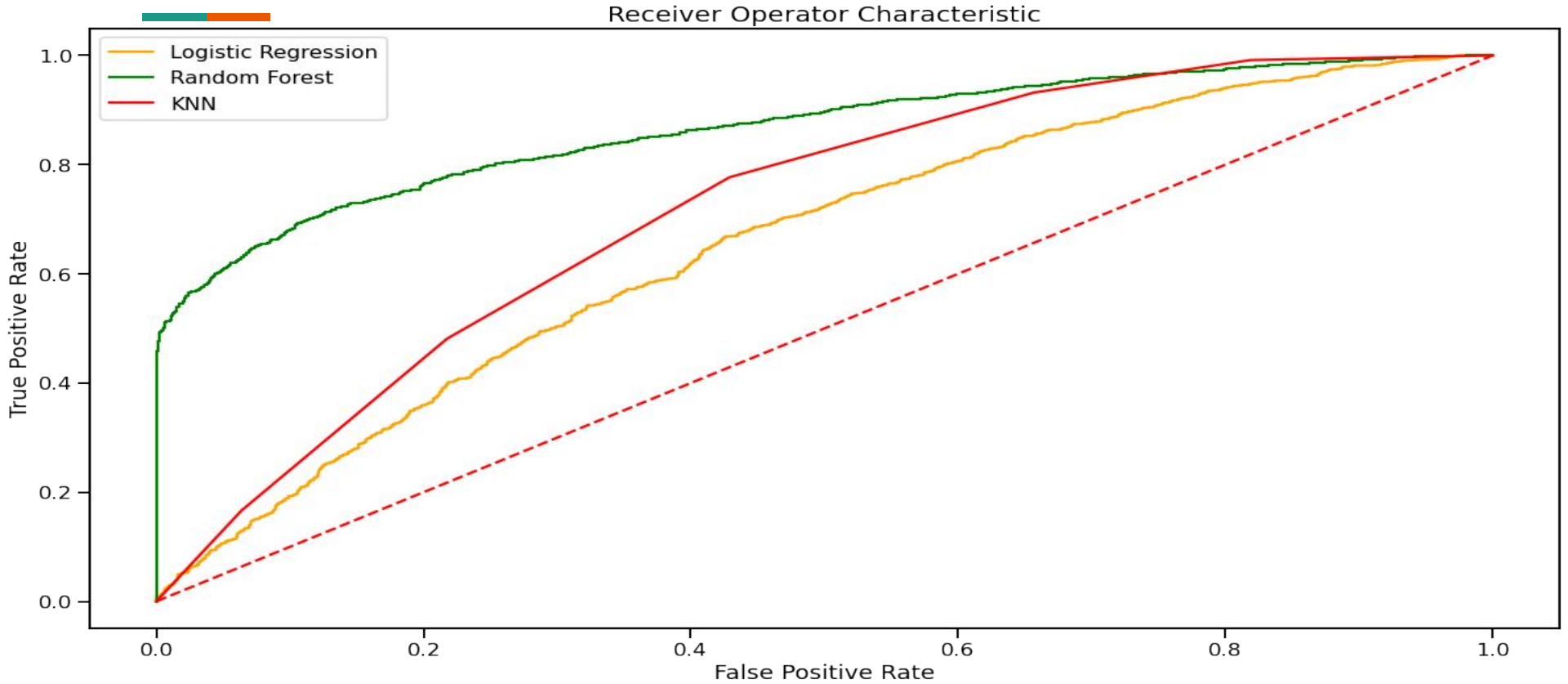
3. Imbalanced Dataset.

4. Feature Selection.

Modelling

Involves selecting the optimal classification model by training and testing multiple algorithms, such as Logistic Regression, Random Forest, and Decision Trees. Evaluate each model using metrics like accuracy, precision, recall, and ROC-AUC to determine its suitability for the business objectives. Fine-tune the models' hyperparameters and analyze feature importance to enhance performance.

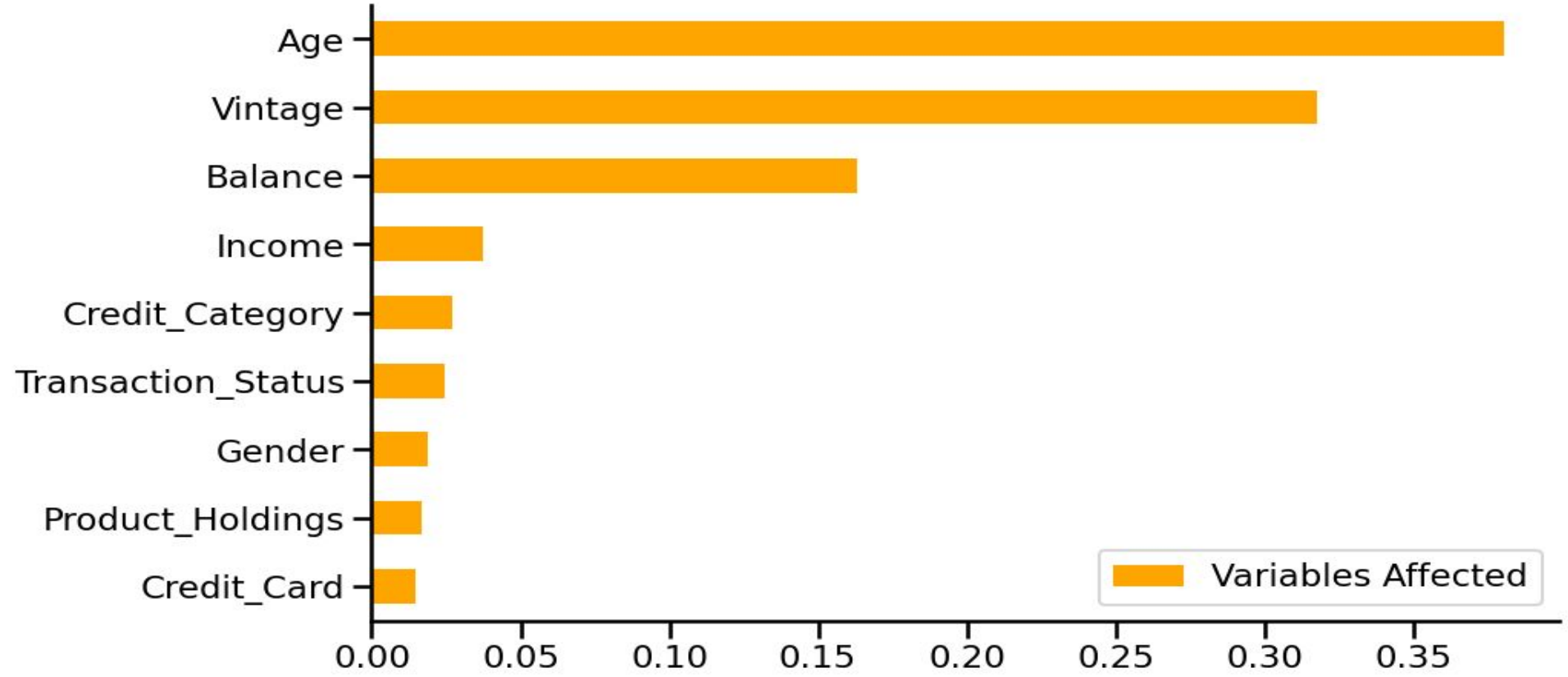
Model Comparison



Deductions

- ROC-AUC curve shows the model performance by plotting the false positive rate to true positive rate.
- More the skewness of the curve towards the upper left corner higher is the area under the roc curve and better is the model performance.
- The Random Forest showed a tremendous performance but it is likely to overfit.
- Most of the models have achieved around 65% accuracy.
- From the ROC-AUC curve it is clear that Logistic Regression performed poorly.

Feature Importance



Conclusions

- From the above representation Age and Vintage feature are the most effective features. The bank should focus on these features to reduce the churn rate. Also, the number of product holdings and whether a customer has a credit card contributes the least in target prediction.
- The number of product holdings and whether a customer has a credit card contribute the least to the target prediction of churn. This suggests that these features might not have a strong impact on the likelihood of churn.
- In addition, new customers are more likely to churn and also customers with high income are hard to retain. The Bank can focus on these 2 groups to bring down the churn rate.

Recommendations

- ➔ **Enhance Customer Onboarding:** Improve the onboarding process for new customers, focusing on personalized offers and engagement to reduce churn among this demographic.
- ➔ **Segmented Retention Strategies:** Implement targeted retention campaigns for high-income customers who are harder to retain, offering personalized incentives or exclusive benefits.
- ➔ **Streamline Product Offerings:** Simplify the range of product offerings for customers with fewer product holdings to make them more accessible and enhance engagement.
- ➔ **Enhanced Customer Support:** Strengthen customer support to address concerns and feedback more promptly, ensuring satisfaction and loyalty.
- ➔ **Predictive Analytics:** Utilize advanced analytics and machine learning models to identify potential churn risks early and proactively engage with those customers.

Thank you.