Exploratory Data Analysis (EDA) on Customer Churn

Prepared by: Awais Aslam
Position: Data Science Intern, EcodeCamp
Date: 11/8/2024

Table of Contents

1. Introduction

This report presents the findings from an exploratory data analysis (EDA) conducted on a customer churn dataset. The primary objective of this analysis is to uncover patterns and insights that may indicate the factors influencing customer churn.

2. Data Collection

Dataset Overview

- Source: Kaggle
- Dataset Name: Customer Churn Dataset
- Features: CreditScore, Age, Tenure, Geography Balance, NumOfProducts, HasCrCard, IsActiveMember, EstimatedSalary, Exited.
- Target Variable: Exited (indicates whether the customer has churned or not)

The dataset consists of 10000 records and 14 features. Each feature provides information relevant to customer behavior and demographic details.

3. Data Preprocessing

Handling Missing Values

There is no missing values in the dataset

Handling Outliers

There is a outliers in the dataset but I don’t remove them cozz there is too many outliers, if I remove them maybe important features of the dataset can be removed, so I just leave the outliers as it.

Data Normalization/Standardization

Data was standardized using:

I do Standardization by using sklearn.preprocessing after the train_test_split. I also mentioned the effect of standardization by visualization.

Effect of Scaling:

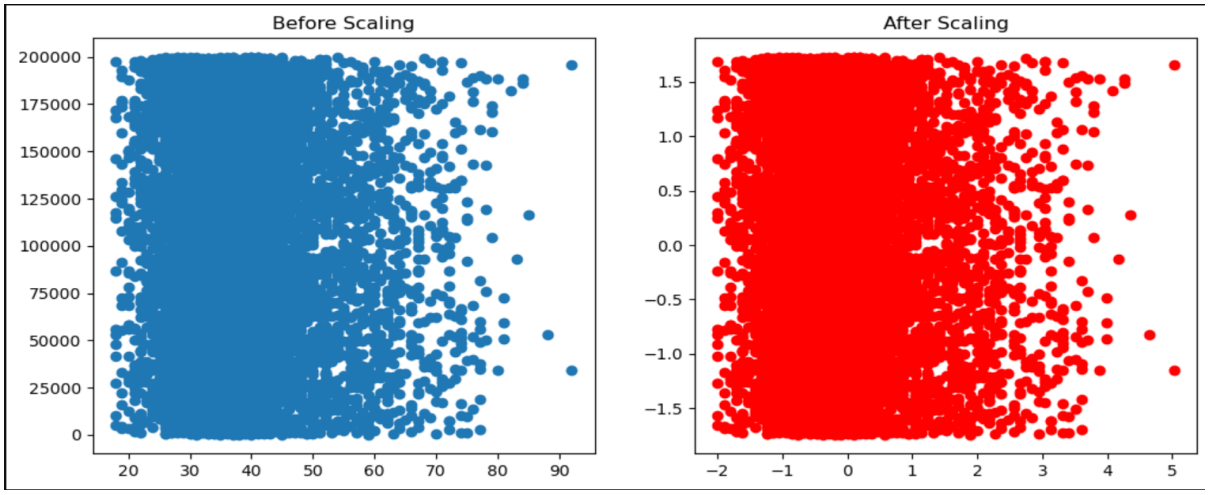

Figure 1

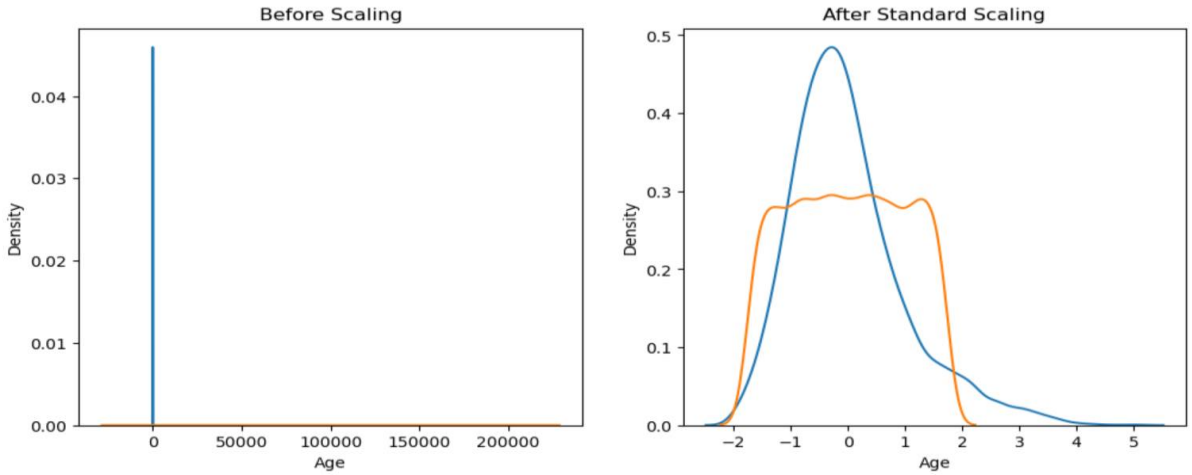

Figure 2

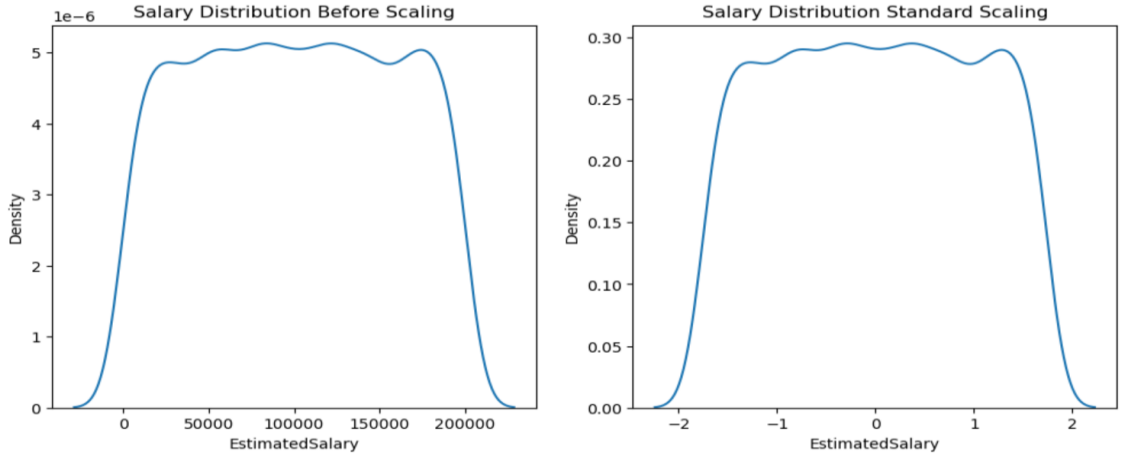

Figure 3

4. Exploratory Analysis

Statistical Analysis

Summary Statistics

The following summary statistics provide insights into the distribution of key features:

- **CreditScore:** Mean = 650.528800, Median = 652.000000, Std Dev = 96.653299
- **Age:** Mean = 38.921800, Median = 37.000000, Std Dev = 10.487806
- **Tenure:** Mean = 5.012800, Median = 5.000000, Std Dev = 2.892174

Balance: Mean = 76485.889288, Median = 97198.540000, Std Dev = 62397.4052

EstimatedSalary: Mean = 100090.239881, Median = 100193.915000, Std Dev = 57510.4

Visualizations

Histograms

- **Age Distribution:**

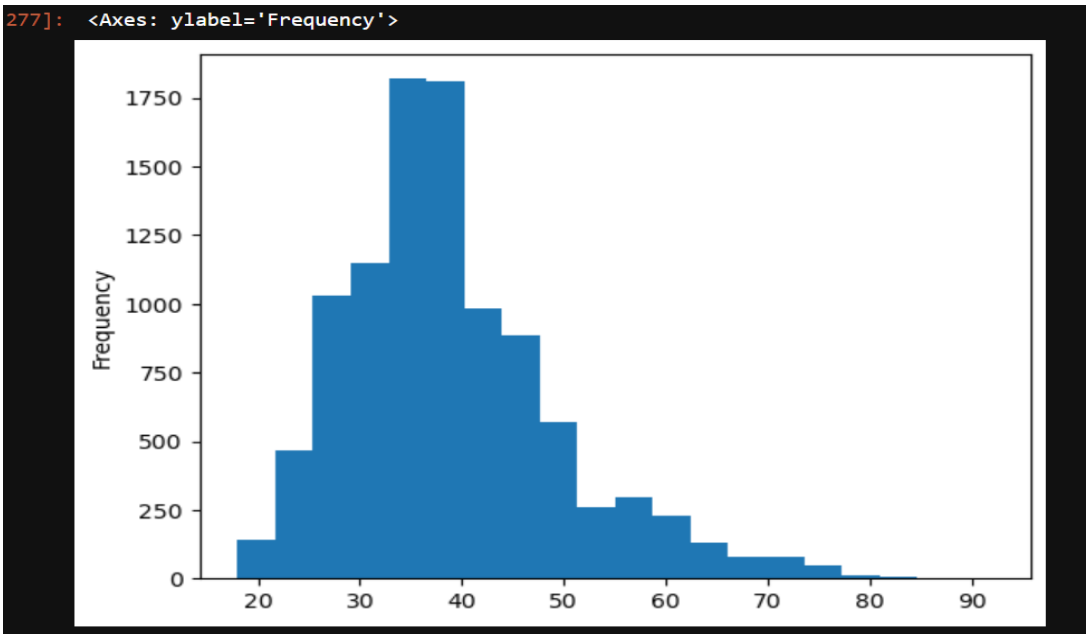

Figure 4

Barplot

- **Exited Distribution:**

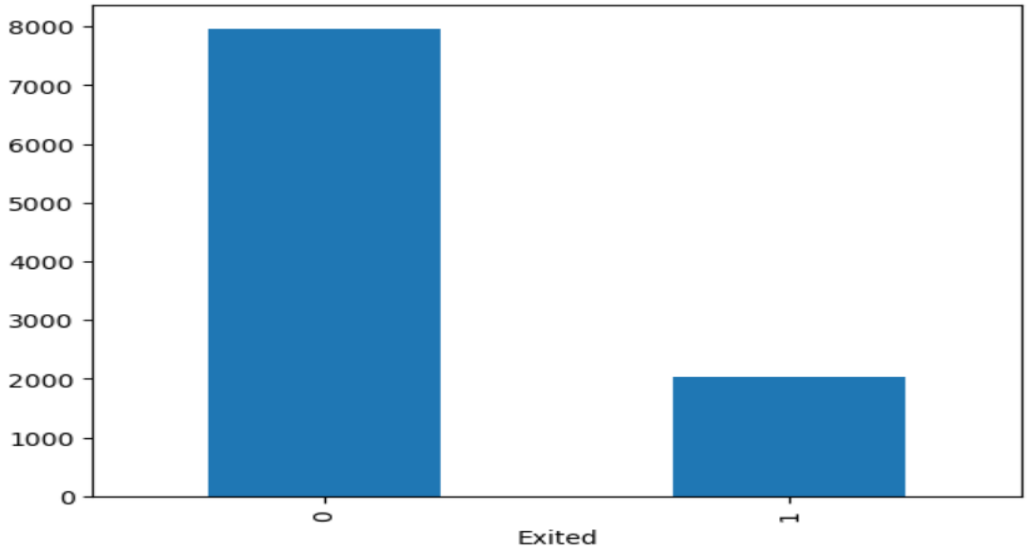

Figure 5

Pieplot

- **Exited Percentage:**

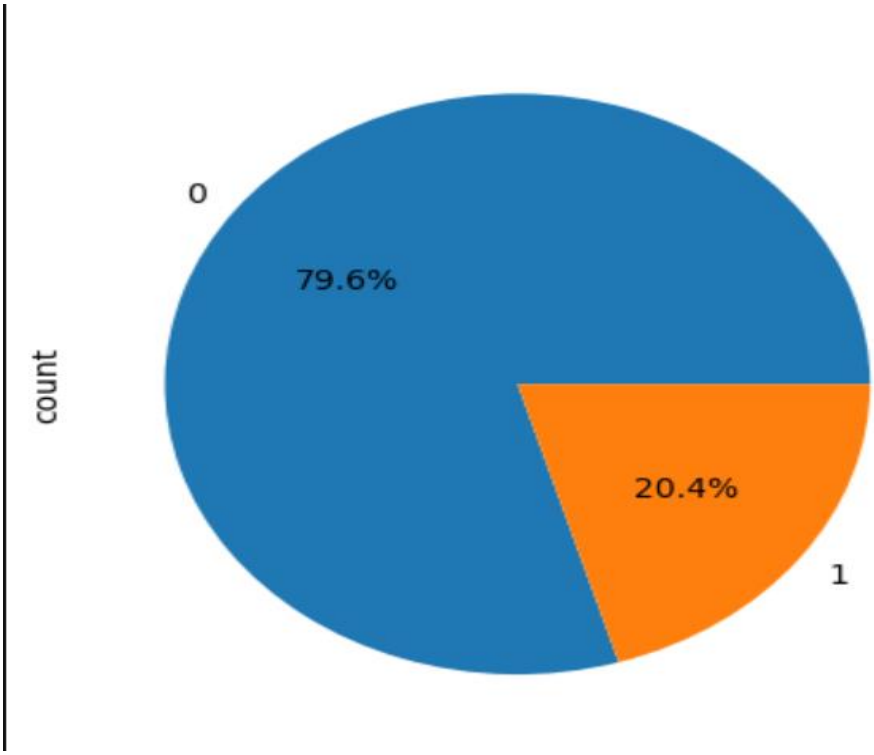

Figure 6

Kdeplot

- **Age Distribution by Gender:**

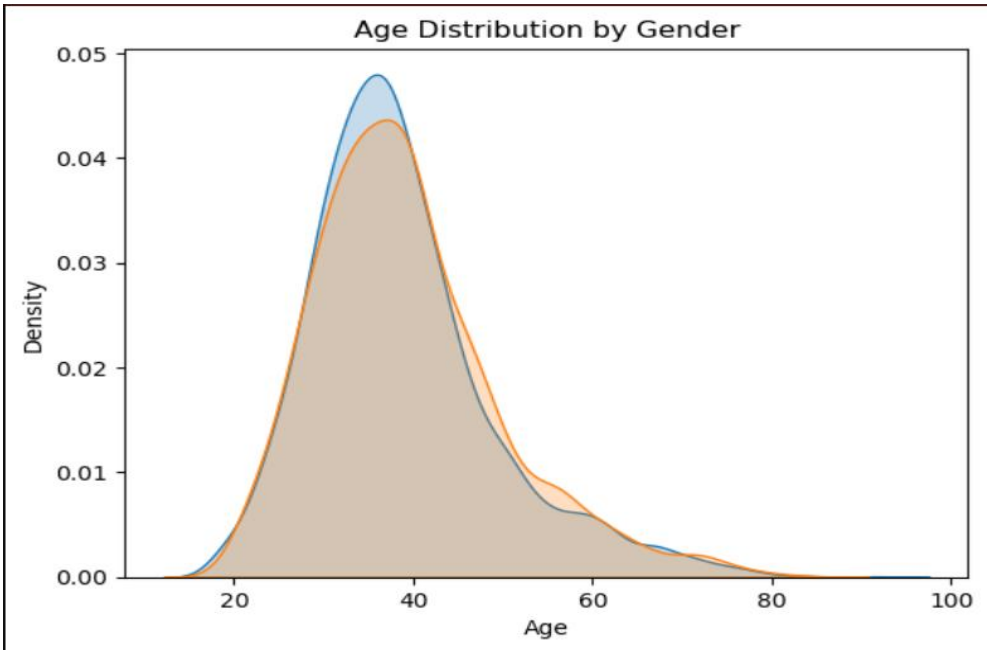

Figure 7

Distplot

- **Age vs Exited:**

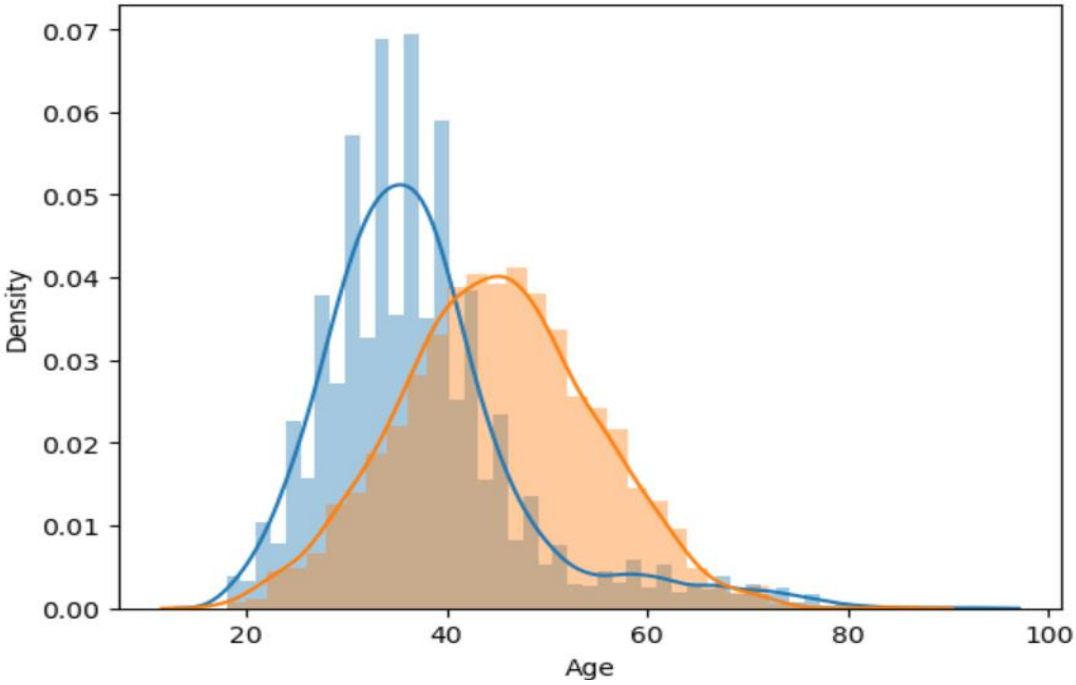

Figure 8

Histograms

- **Estimated Salary Distribution:**

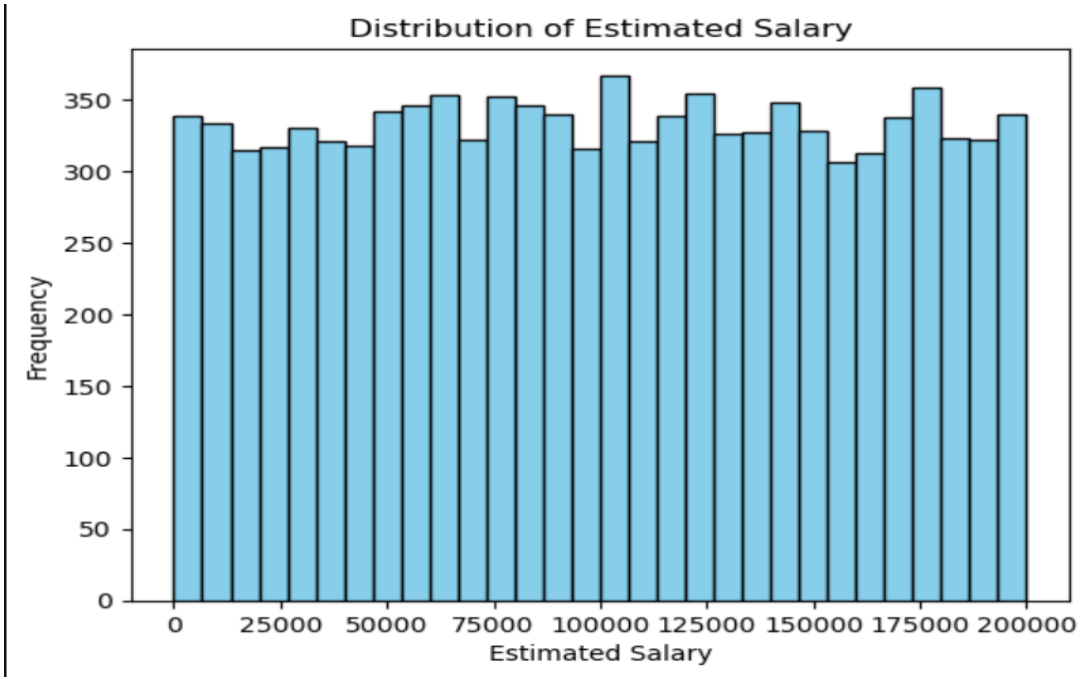

Figure 9

Scatter Plots

- **Age vs. Estimated Salary:**

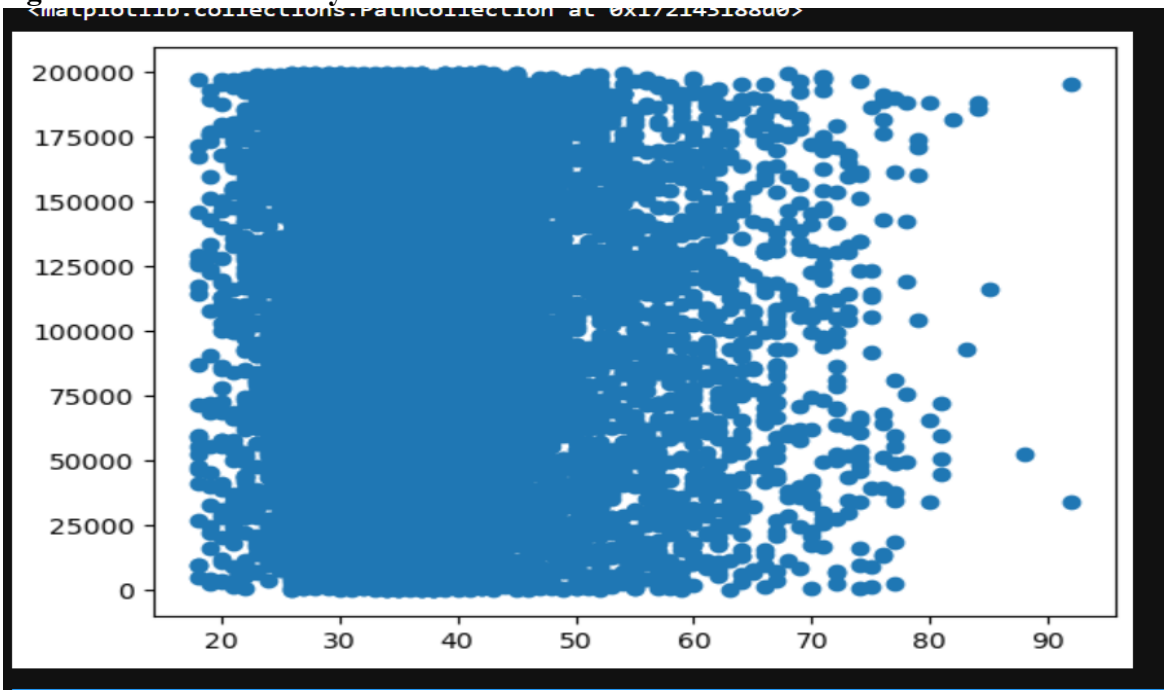

Figure 10

Barplot

- **Geography Distribution:**

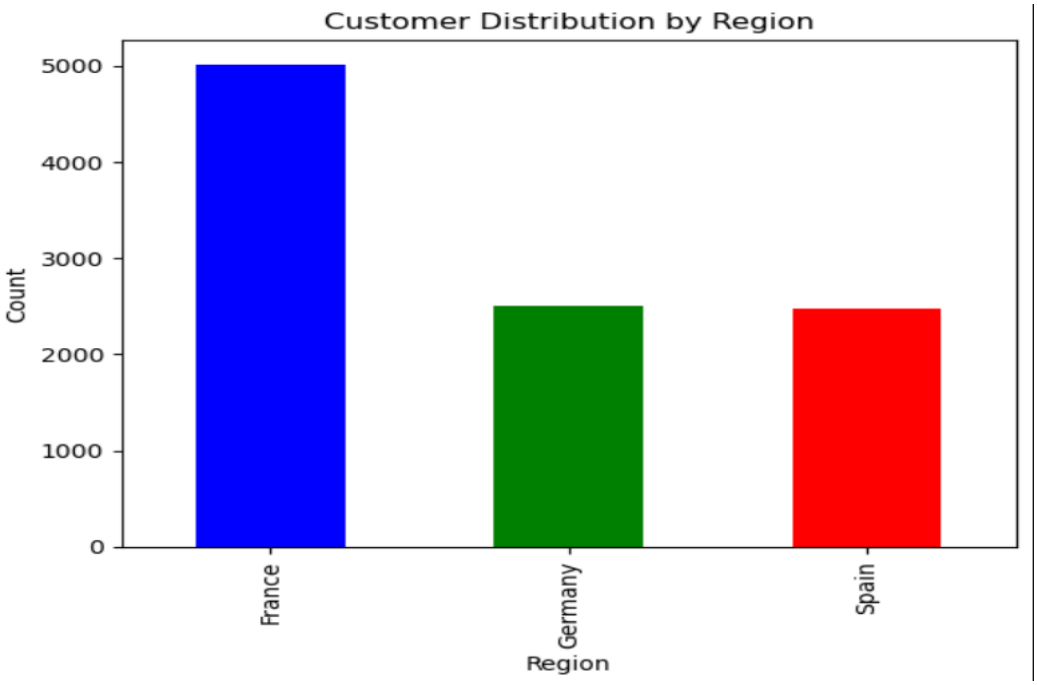

Figure 11

Pieplot

- **Geography Percentage:**

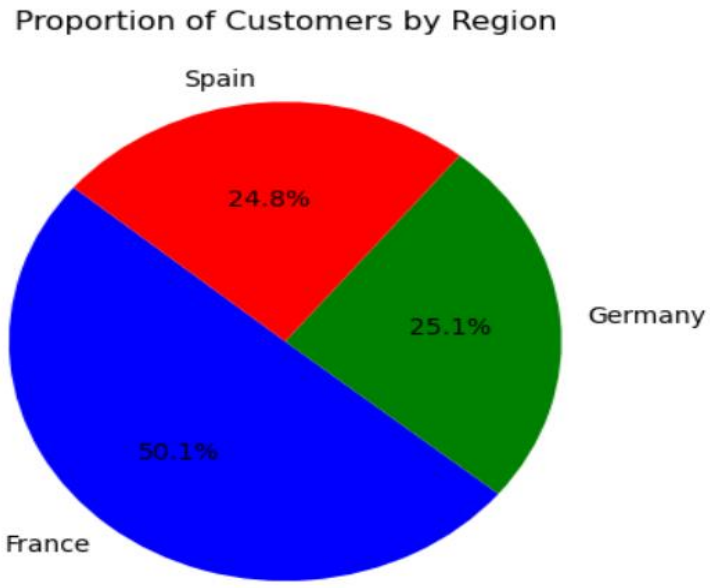

Figure 12

Barplot

- **Gender Distribution:**

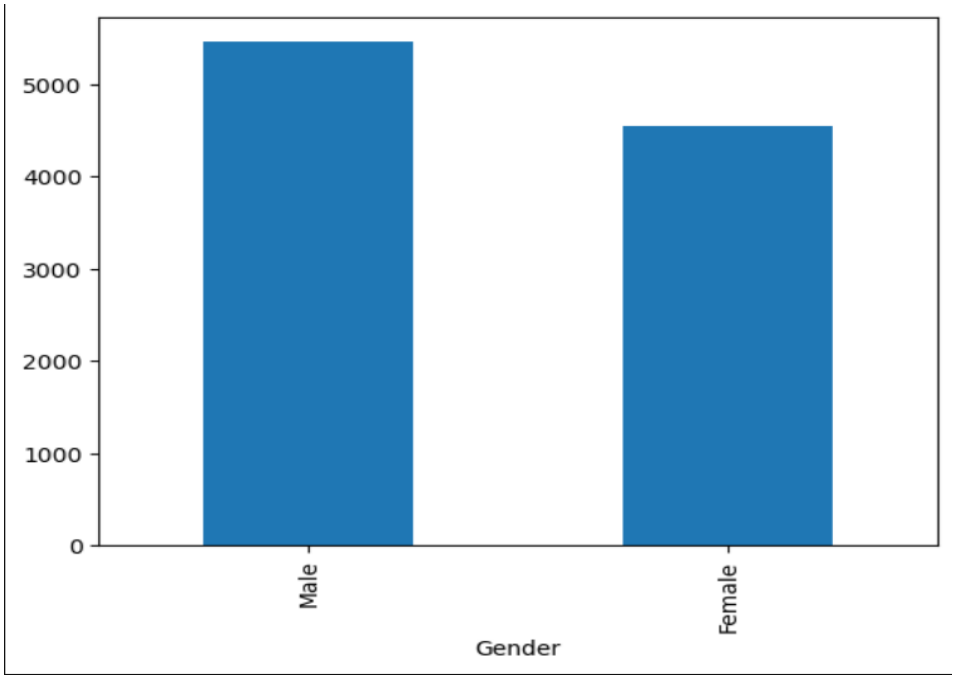

Figure 13

Pieplot

- **Gender Percentage:**

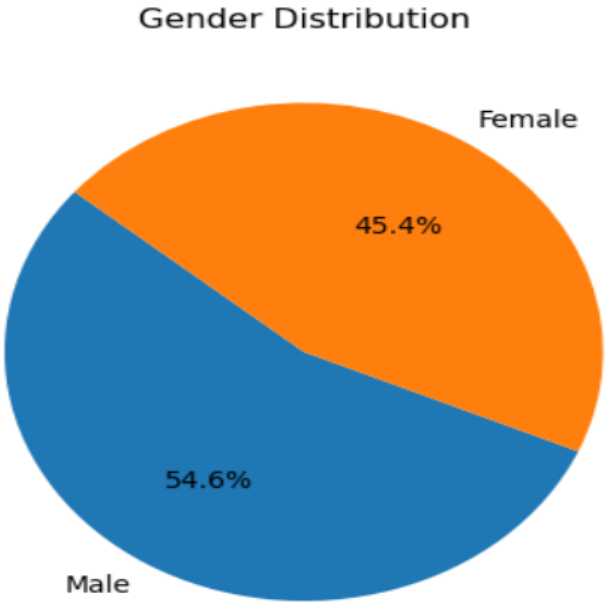

Figure 14

Barplot

- **Gender vs Exited**

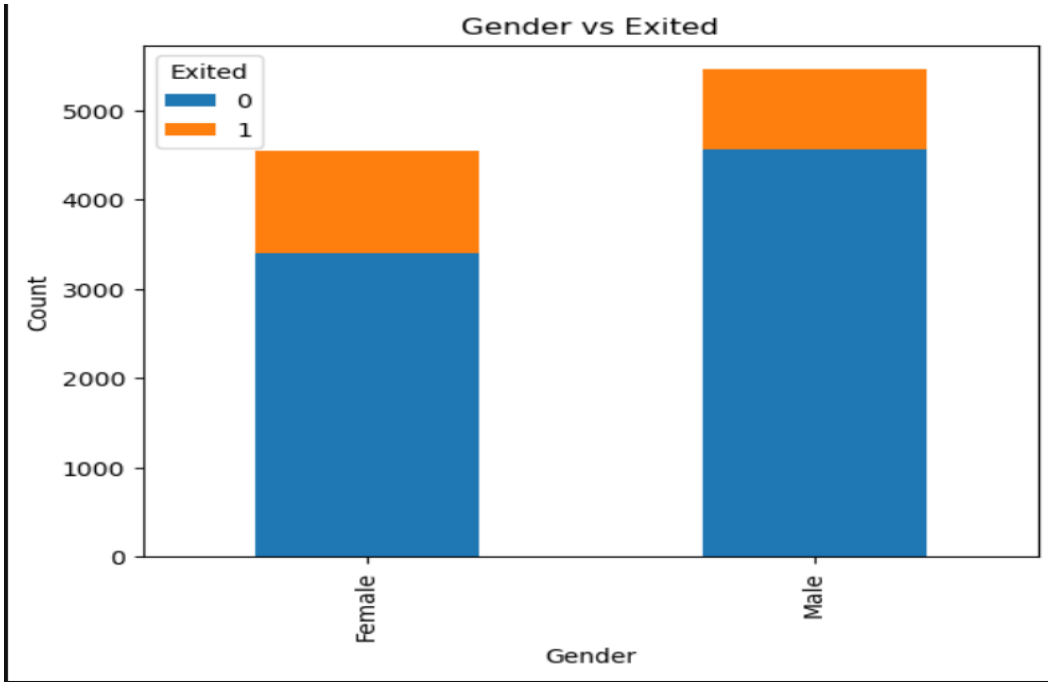

Figure 15

Heatmap

- **Correlation Gender vs Exited:**

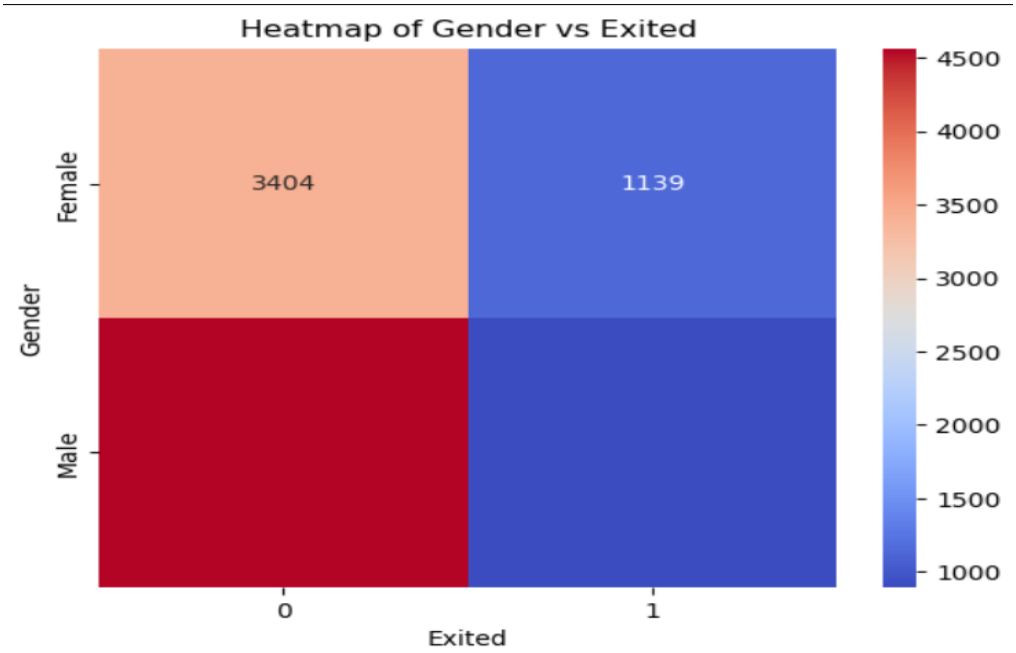

Figure 16

Barplot

- **Distribution of Customers with Credit Card:**

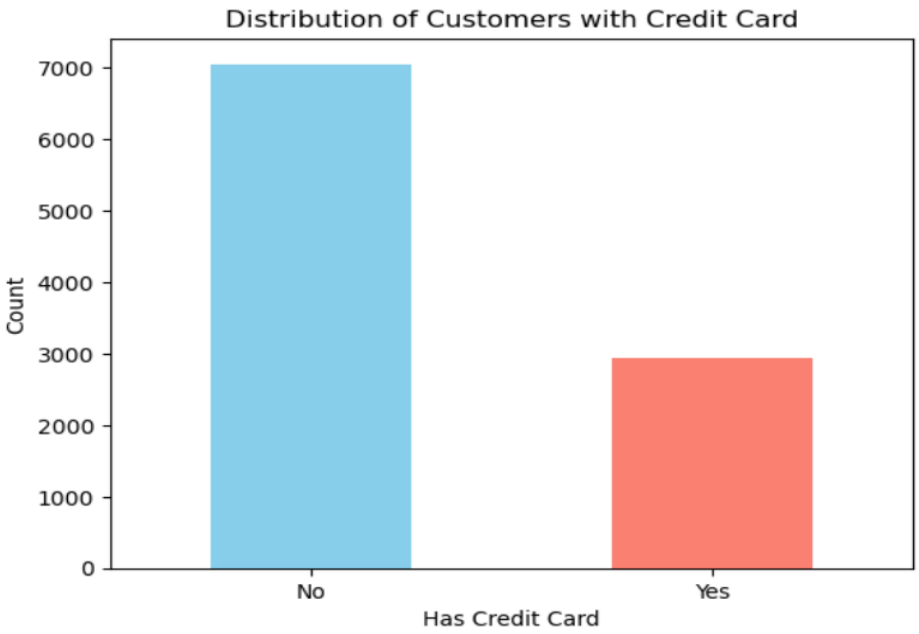

Figure 17

Pieplot

- **Proportion of Customers with Credit Card:**

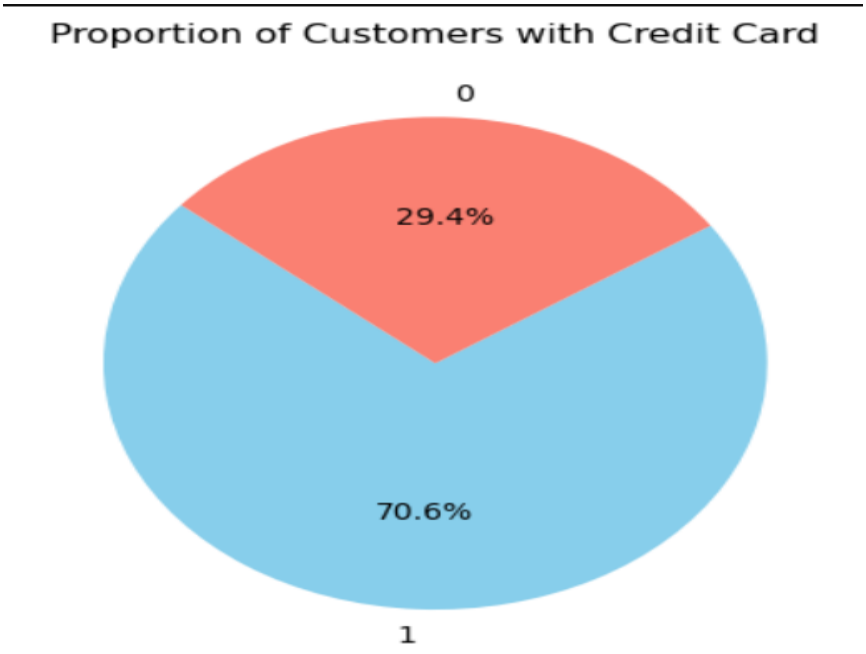

Figure 18

Barplot

- **Distribution of Active vs Inactive Members:**

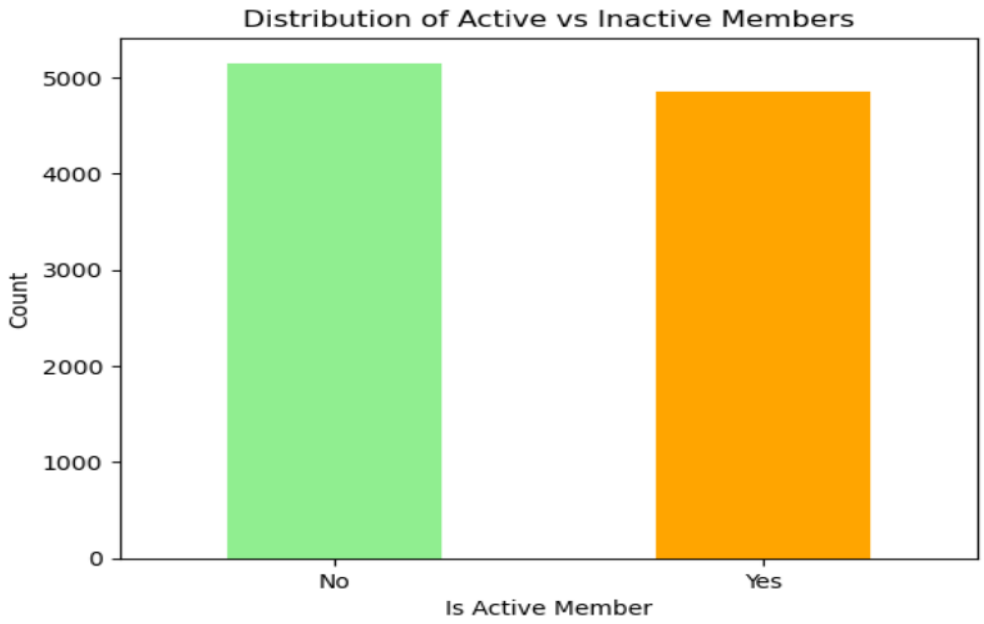

Figure 19

Correlation Matrix

- **Correlation Heatmap:**

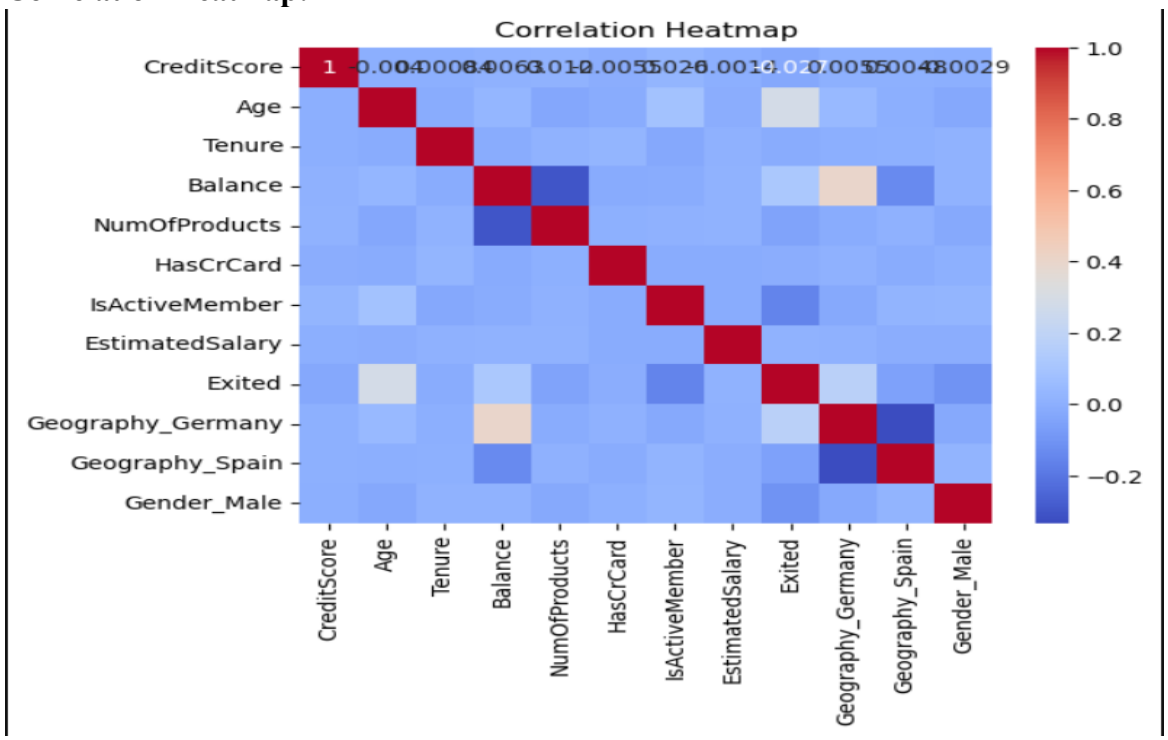

Figure 20

Side-by-side histograms:

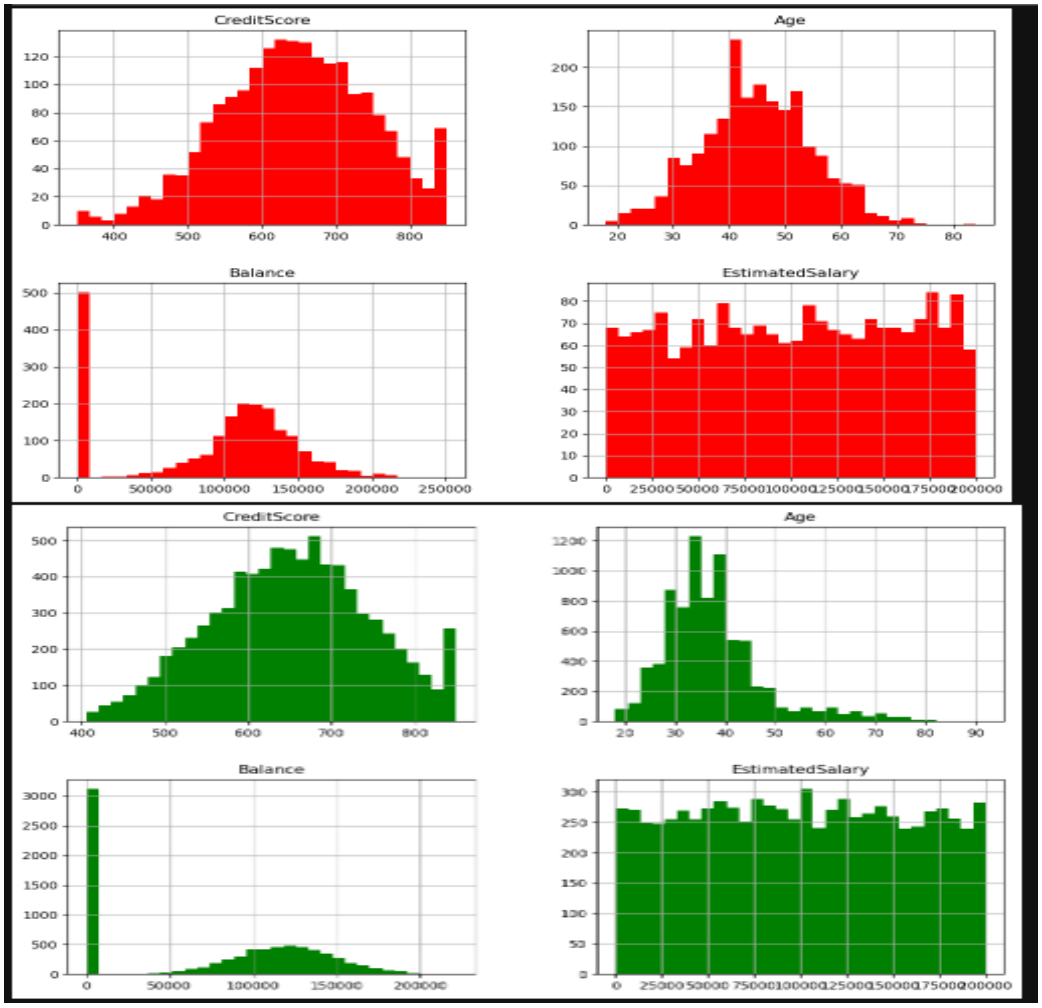

Figure 21

Pairplot:

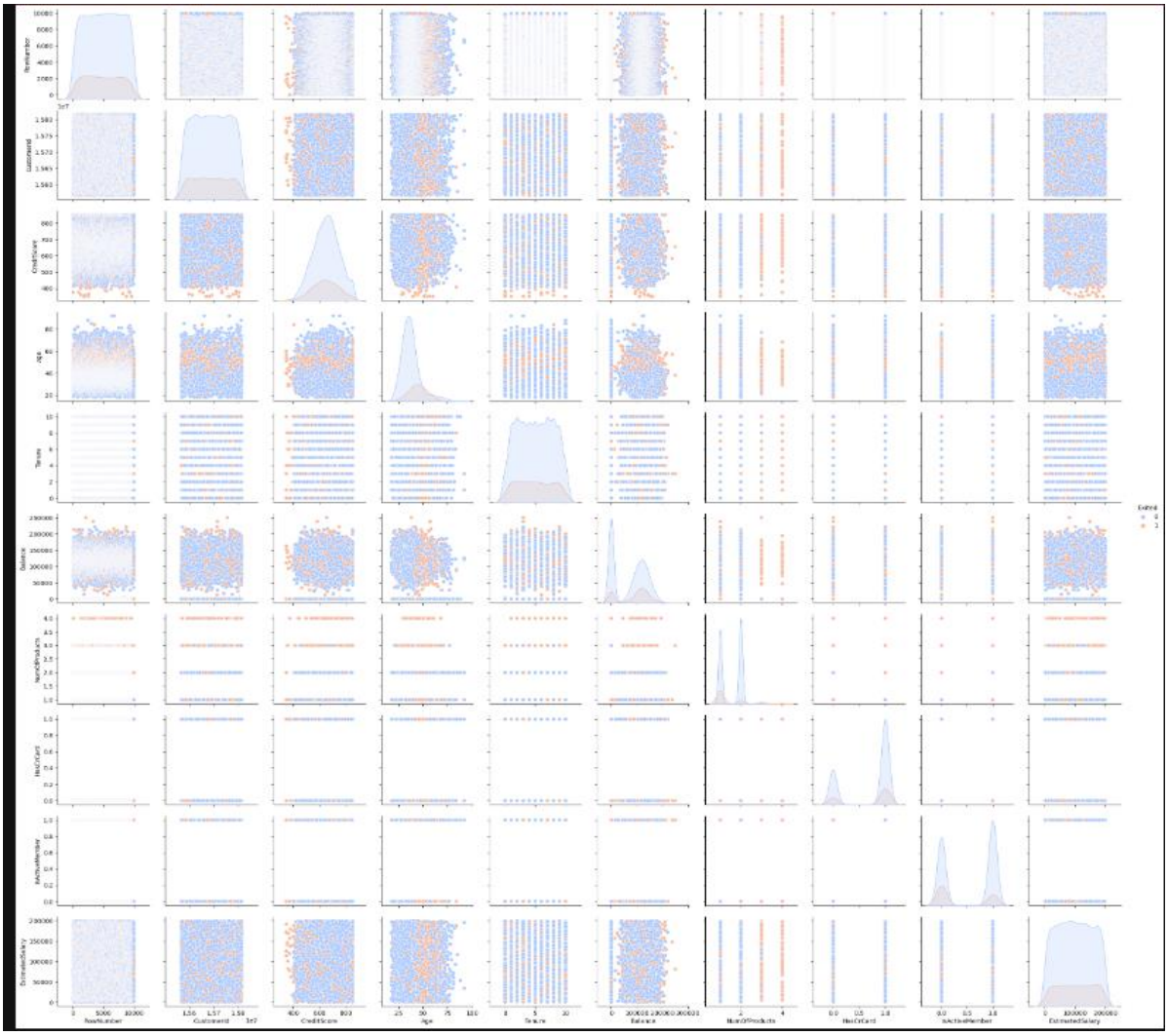

Figure 22

5. Conclusion

Summary of Findings

The analysis identified several key factors influencing customer churn:

- Balance: Lower account balances are linked to higher churn rates.
- Tenure: Shorter tenures correlate with increased churn, indicating newer customers are more likely to leave.
- NumOfProducts: Fewer products are associated with higher churn, suggesting that more products may improve retention.
- Age: Younger customers show a higher tendency to churn compared to older customers.
- IsActiveMember: Inactive members are significantly more likely to churn, emphasizing the importance of customer engagement.
- HasCrCard: Customers with a credit card are less likely to churn.

Correlation Insights

- Balance and Churn: Lower balances are strongly correlated with higher churn.
- Tenure and Churn: Shorter tenure is associated with a higher likelihood of churn.
- NumOfProducts and Churn: More products correlate with better retention.
- Age and Churn: Younger customers show a higher likelihood of churning compared to older customers.

Overall, factors such as financial stability, customer engagement, and product variety are crucial in influencing customer churn. Enhancing these areas could help reduce churn and improve retention.

6. Recommendations

Based on the findings, the following actions are recommended to reduce customer churn:

1. Boost Financial Stability: Provide incentives or financial advice to help customers maintain higher account balances.
2. Improve New Customer Experience: Strengthen onboarding with personalized offers and support for new customers.
3. Encourage Product Usage: Promote additional product adoption through cross-selling and bundled offers.
4. Engage Younger Customers: Develop targeted marketing and offers for younger demographics.
5. Reactivate Inactive Members: Launch campaigns to re-engage inactive customers with special promotions.
6. Promote Credit Card Usage: Increase awareness and benefits of credit cards to reduce churn.

Implementing these strategies can help enhance customer retention and satisfaction.
