# CUSTOMER CHURN IN BANKING ANALYSIS

NAME-ARYAMAN PILLAY

FINAL PROJECT

COURSE NO- MSBA 320

PROFESSOR NAME-SIAMAK ZADEH

## Table of Contents

# Introduction

In the Banking Sector, understanding Customer Behaviour, preferences ,etc is quite important for growth for any banking organisation. This dataset provides an overview of the factors affecting customer churn within the banking sector. Customers leaving the bank causes huge problems and challenges for businesses , leading to loss in revenue and causing harm to the reputation of the organisation. customer churn analysis has been focused on industries such as Information Technology, Online Video and Music Streaming Platforms, and e-commerce. However, with the increase usage  banking products and services and the rise of digital banking, customer churn has emerged as a pressing concern for banks globally. The study aims to identify the reasons for  customer churn within the banking sector, providing advanced analytics and statistical techniques to identify key drivers of churn and formulate actionable strategies for retention. The findings of the study are expected to provide valuable insights for banks and financial institutions, enabling them to improve  customer retention efforts, improve service offerings, and foster long-term relationships with their customer base.

# Overview of the Dataset

The Dataset consists of various variables like:

1**: CustomerID**: refers to the unique identification number of the customer.

2: **Surname:** refers to the last name of the. Customer.

3: **Creditscore:** indicates the likelihood of the customers repaying the money.

4: **Geography**: refers to various countries the customers are from

5: **Gender:** refers to whether the customer is a Male or a Female.

6: **Age:** refers to the Age of the Customer

7: **Tenure**: is a measure of the duration of the customer's relationship with the bank.

8**: Balance**: balance typically refers to the amount of money held in a customer's account or accounts with a financial institution. I

9: **Num of Products**: refers the number of financial products or services that a customer has or is associated with a bank or financial institution.

10**: HasCrCard**: This column or variables basically refers to whether the customer has a credit card or not.

11: **IsActiveMember:** This column basically means whether the customer is currrenly an active member of the bank or not.

12: **Estimated Salary**: This column refers to the salary amount of a customer

13**: Exited**: This basically refers to whether the customer has exited their relationship from the bank or not.

# Data Collection

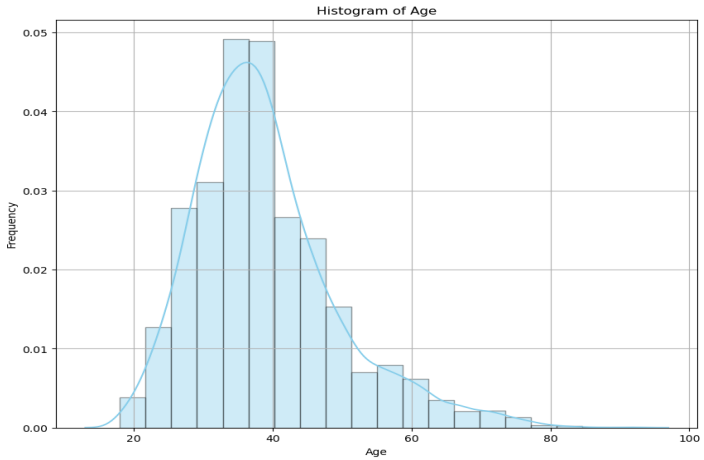
Data collection is an important step in any research or analysis process, providing the foundation upon which insights and conclusions are built. The dataset which has been selected for analysing customer churn in the banking industry was collected from Kaggle, a popular platform for sharing public datasets. Kaggle offers a diverse range of datasets ranging from various fields like finance, healthcare, OTT streaming, Marketing , HR etc. The platforms provide access to datasets users from any part of the world. Data collection from Kaggle offers several advantages:

1: **Accessibility**: Kaggle provides easy access to a wide range of datasets. Anyone can download the dataset from Kaggle and explore and analyse which dataset they would want access to.

2**: Quality Assurance**: Kaggle's Datasets undergo quality checks and validation. Kaggle's Datasets provides reliability ,integrity for analytical purposes.

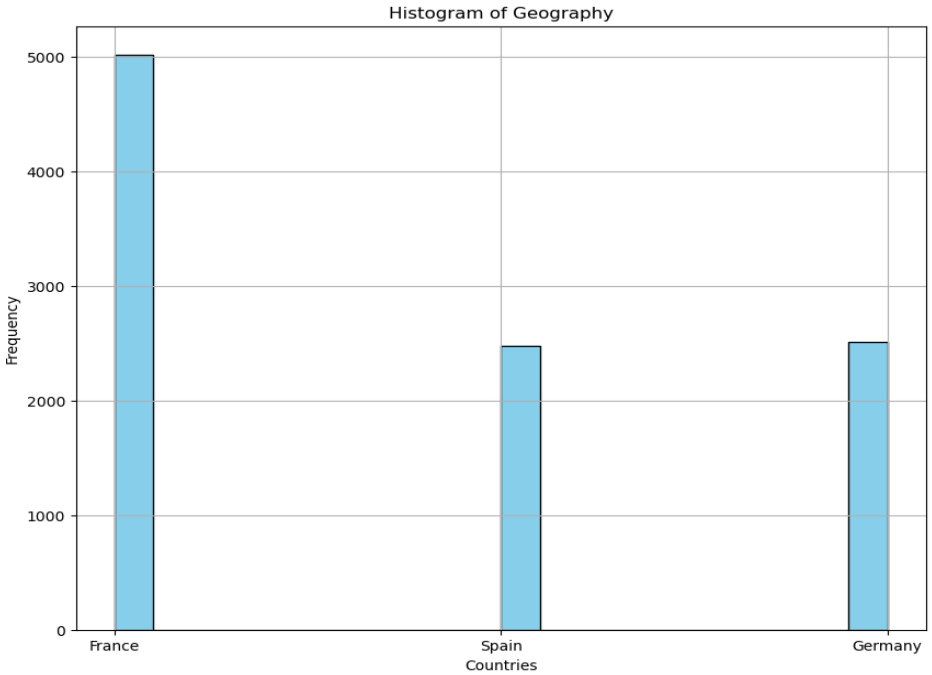
3: **Community Engagement**: Kaggle creates a vast community of data scientists , analysts and enthusiasts who actively interact with datasets throughout competitions, discussions etc.

# Data Analysis using Histograms, Count plots, Scatterplots and Heatmaps(Explolatory Data Analysis)



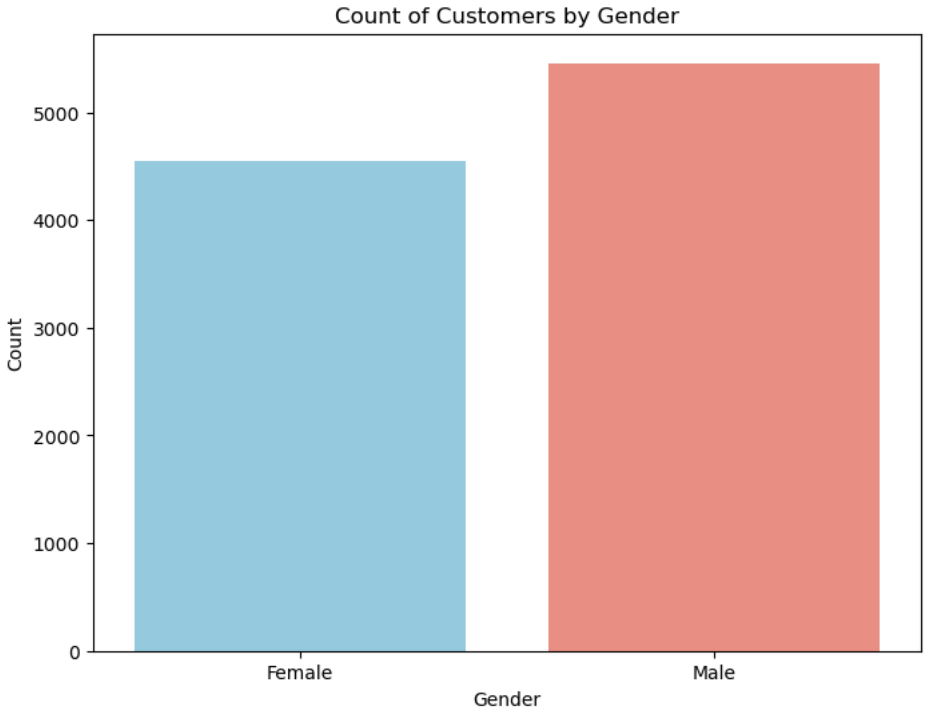

This is a Histogram of Frequency of Age column. According to the figure ,

- X- axis consists of age and the Y-axis consists of Frequency of Age.
- The histogram is skewed to the right which indicates that there are more data points on the younger side.
- The highest frequency ranges from 20-40. The frequency reduces after the age of 40. As the age increases, frequency decreases and the frequency peaks around the age between 20 and 40.




This figure depicts the frequency distribution of the European Countries. The Countries are France , Spain and Germany.

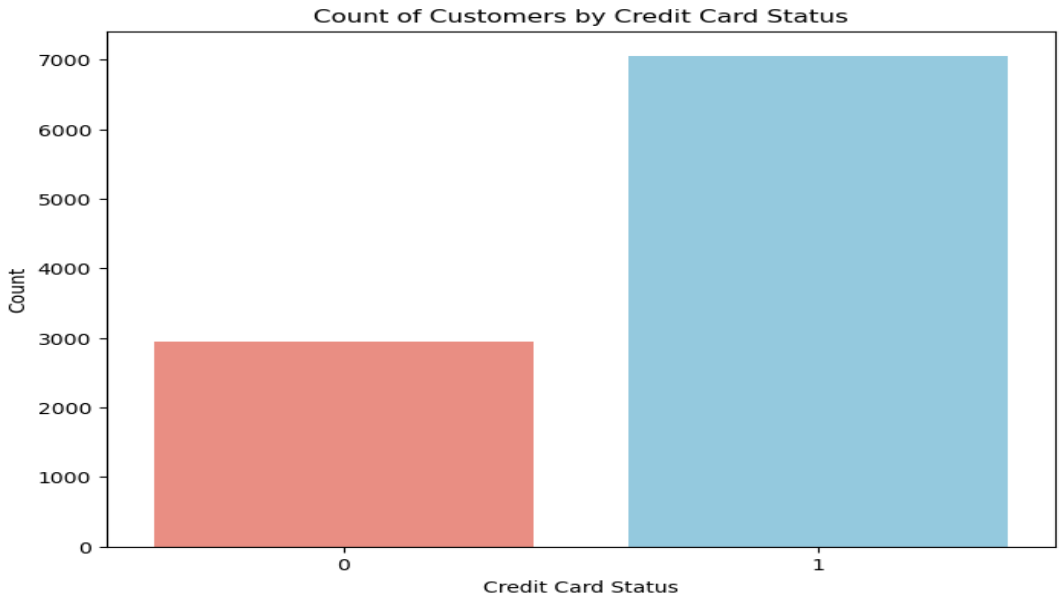
- The Frequency distribution for France is the highest . It touches 5000 on the Frequency axis.
- The Frequency distribution of Spain and Germany is the same. They both cross 2000 on the frequency axis.




This depicts a countplot of Gender .

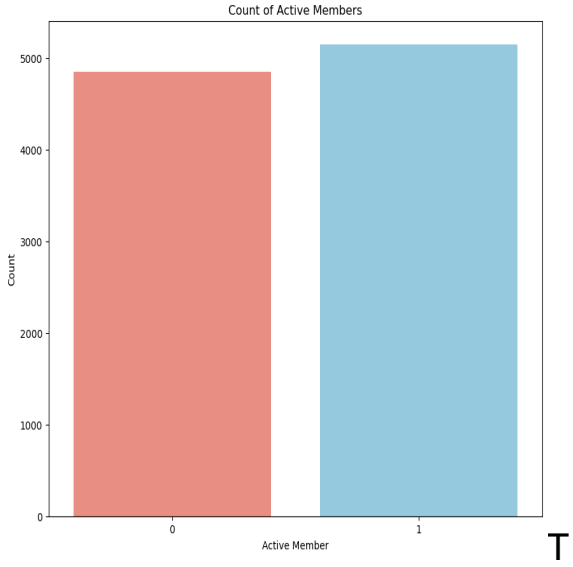
- Male counts crosses 5000 as can be  seen on the Count axis.
- Female count crosses 4000 as can be seen on the Count axis.

Male Customers are higher than female customers.




- The figure depicts a count plot of customers who have a credit and customers who do not have a credit card. There are two distinct categories labelled as “0” and “1.”  0 depicts customers who do not have a credit card. 1 depicts customers who have a credit card.

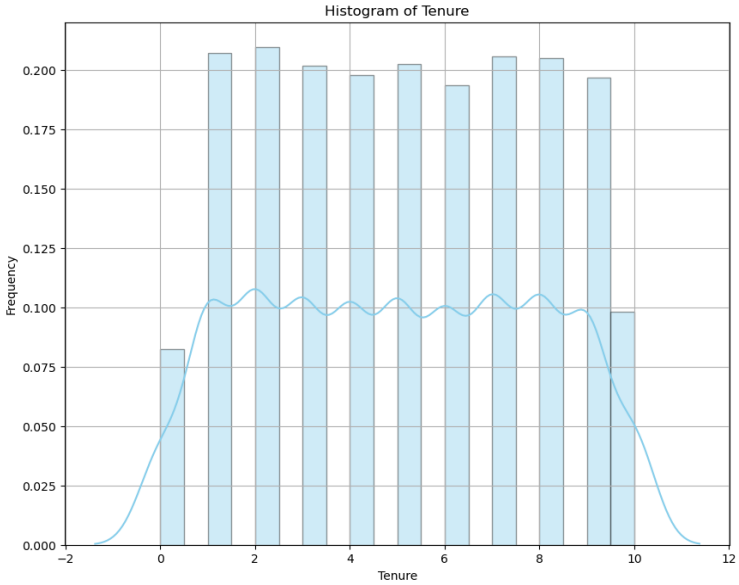
- The X- axis represents Credit Card status and the Y- axis represents the count of customers.
- The red bar corresponds to customers who do not have credit card. The blue bar corresponds to customers having a credit card.
- The count of credit card customers with status “1” which customers who have a credit card are more in number.
- The bar for the credit card status “1” crosses 6500 on  the count axis
- The bar for the credit card status “0” is somewhere close to 3000 on the count axis.




T

The figure depicts the counplot of active members.

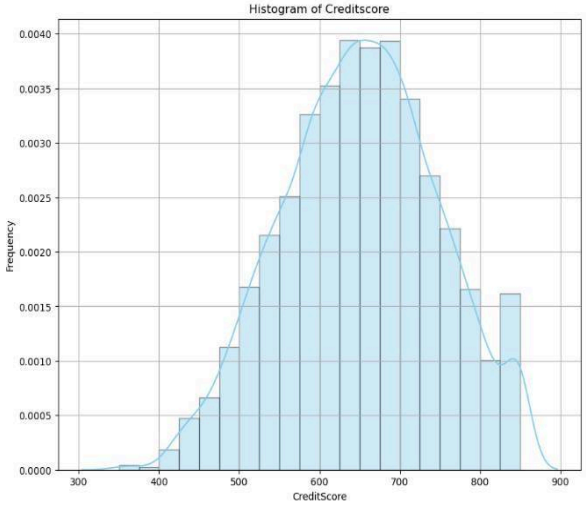
- There are two categories labelled as “0” and “1”. 0 indicates non-active member and 1 indicates that the member is active. The red bar is lablled “0” and the blue bar is labelled “1”
- The x-axis represents the categories of Active  Members.
- The y-axis  represents the count of Active Members.
- The count of Active members is more than the count of members who are not active.
- The count of active members touches 5000 on the count axis
- The count of members who are not active crosses more than 4000.




This figure depicts an histogram which represents distribution of Tenure, which means the number of years the customer was associated with the bank.

- The X- axis consists of Tenure which represents the number of years and the Y-axis consists of the Frequency of Tenure.
- As you can see in the figure the frequency is the highest for the Tenure value 2.
- The distribution peaks around a tenure value of 6.

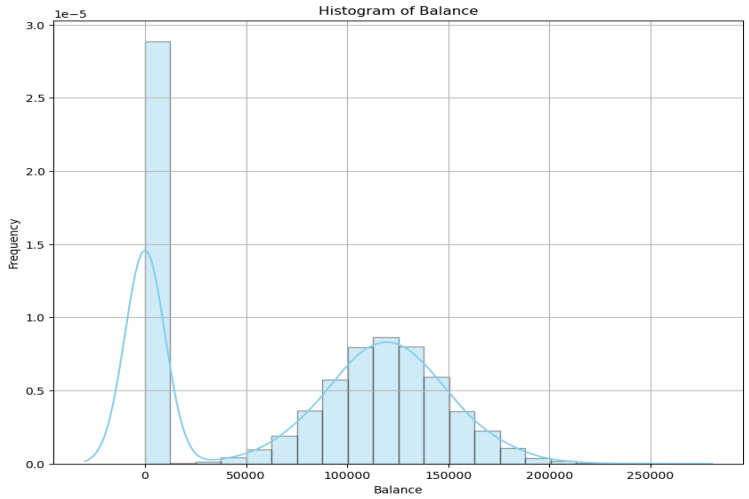
  Customers who have had a tenure with the bank for 2 years is the highest in number.




The figure depicts a histogram which represents Creditscore of various customers.

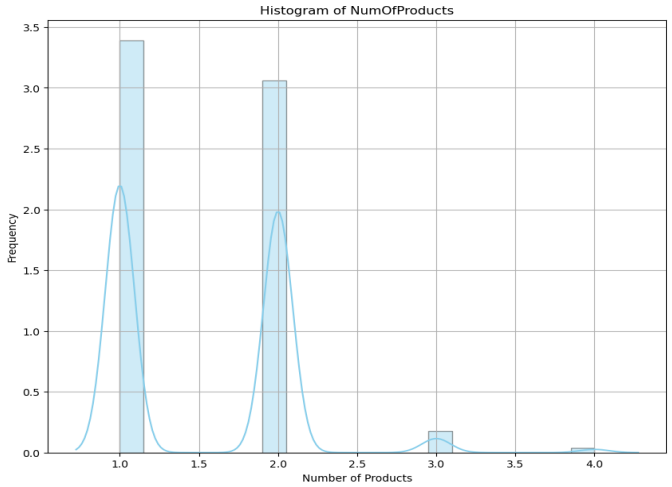
- The X-axis consists of Credit Score  values
- The Y- axis consists of the Frequency values of  Y-axis.
- Credit score ranging from 600 to 700 have the highest frequency.
- There is a significant drop in the frequencies beyond 700 which basically indicates that customers having a credit score beyond 700 are very few in number.

Customers having Credit Score between 600 to 700 are the most in number. Customers having a Creditscore beyond 700 are very less.11




The figure depicts a histogram which represents the balances in the accounts of customers.

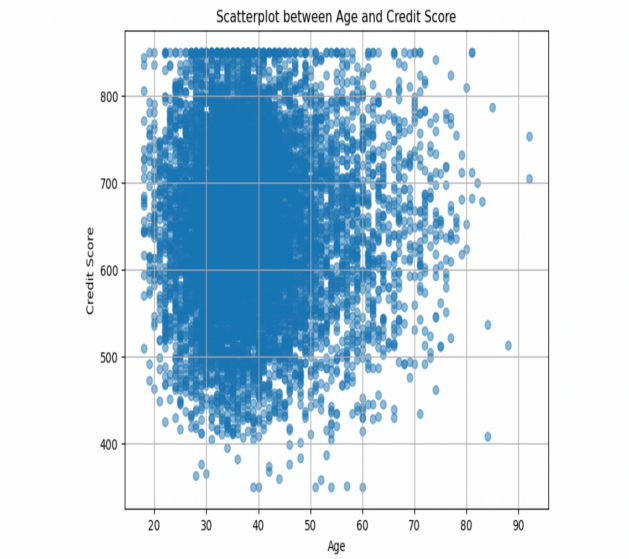
- The X- axis consists of Balance values and the Y-axis consists of the frequency of the Balance values.
- According to the figure , frequency of customers having a balance of 0 is the highest, indicating that these are the churned customers who have discontinued
- The values peak around 125000, indicating that these customers have maintained positive balances and are continuing their services with the bank, in other words are non-churned customers.




This figure depicts a histogram of the number of Products which refers to the number of financial products or services the customer has or is associated with the bank.

- The X- axis represents the number of products and the Y-axis represents the frequency of the number of the number of products.
- The frequency is highest for the customers who have only one product associated with the bank which is more than 3.0
- The frequency is slightly lower for customers who have two products associated with the bank which is slghtly more than 3.0 but is lower than customers having one product
- The frequency is lower for customers who have 3 or more products associated with the bank which is between 0 and 0.5 .

# Correlation Analysis

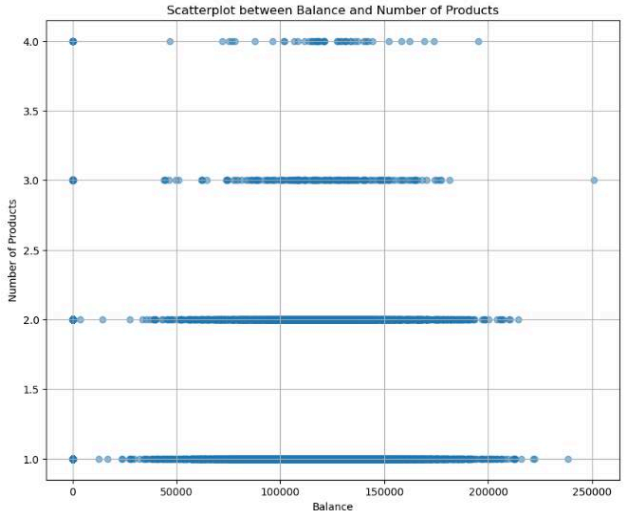



This figure depicts a scatterplot between variables Age and Credit Score

- X- axis repreents Age of customers and the Y-axis represents Credit score of Customers.
- There are a significant number of customers from the Age ranging from 30-60 , having a credit score ranging between 600-700.
- Customers who are at the age of 30 or less than 30 have very low credit scores
- There are Customers who are more than 70 have very low credit scores
- There are very few customers who have a credit score more than 800.

The scatterplot suggests that **age** and **credit score** are related, but the relationship is not linear, indicating that there is no consistent increase or decrease in the credit score because of increase in age.

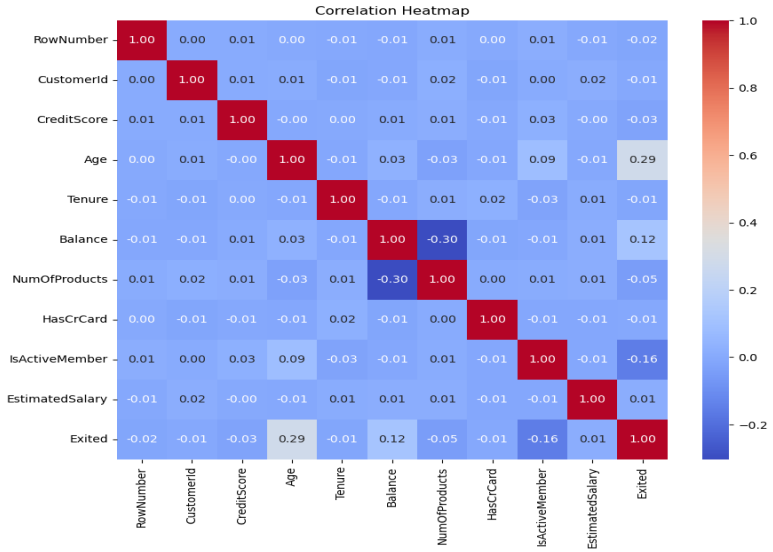
The correlation coefficient between Age and Credit Score is **-0.0039649055253900695** indicates a slight negative correlation, implying that as age increases, there might be a very slight tendency for the credit score to decrease, and vice versa, but this tendency is extremely weak.




This figure depicts a scatterplot between the Balance amounts of customers and the Number of Products the customers have associated with the bank.

- The X-axis consists of Balance amounts of Customers and the Y-Axis consists of the Number of Products.
- Most customers have 1-3 products irrespective of their balance.
- There is no strong correlation between Balance and Number of Products.

The correlation coefficient between Balance and Number of Products is **-0.3041797383605491,** indicates that there is a very weak correlation between Balance and Number of Products. There is a alight tendency of the Balance to decrease if the number  of products increases.




This figure depicts a heatmap which shows correlation values between various variables.

- Dark colors suggest strong correlation while lighter colors suggest weak or no correlation
- Positive correlations are indicated by Red or Orange color . Negative  and weakcorrelations are indicated by dark and light blue colors.
- There is a negative correlation  **-0.30**  between **Balance and Number of Products** which means that as the Balance increases, the number of products associated with bank decreases.

- There is a positive correlation **0.29** between **Age and exited** indicating that the as the Age increases, number of customers exiting or closing their accounts with the bank increases.
- There is a weak correlation **-0.00** between Age and Credit score indicating that there is no consistent increase or decrease in Credit score if the Age of the Customer increases.

## Analysis through Regression Models

Objectives of Regression Models:

1: **Understanding Relationships**: Regression models helps us in understanding the strength of the relationship between Independent and Dependent Variables. In this case through Regression Models , we will understand whether variable like CreditScore, Balance, Age, Num of Products have an effect on the outcome of the Dependent Variable “Exited” which is basically Customers leaving the bank.

2: **Evaluation**: Regression Models provides a framework for evaluating the the significant effect of the Independent Variables on the Dependent Variables.

3: **Comparison of Models**: Regression Analysis enables to determine which model is the best fit for the dataset by comparing various statistcis like R- squared, Mean- R-squared etc.

4**: Hypothesis testing**: Regerssion Models conducts Hypothesis testing about the relationship between various variables.

**model1= ols('Exited ~ NumOfProducts', data=df).fit()**

| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| NumOfProducts | 1.0 | 3.709236 | 3.709236 | 22.915223 | 0.000002 |
| Residual | 9998.0 | 1618.353864 | 0.161868 | NaN | NaN |

This is Model 1 which consists of a Dependent Variable “Exited” and an Independent Variable “NumofProducts”.
The results of this model indicates that the Number of Products has significant effect on the customer’s exit or leaving the bank as the PR(>F) is 0.00002.

**model2 = ols('Exited ~ NumOfProducts + Balance', data=df).fit()**

| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| NumOfProducts | 1.0 | 3.709236 | 3.709236 | 23.189890 | 1.489206e-06 |
| Balance | 1.0 | 19.328172 | 19.328172 | 120.838417 | 5.991070e-28 |
| Residual | 9997.0 | 1599.025692 | 0.159951 | NaN | NaN |

This is Model 2 which consists of a Dependent Variable “Exited” and 2 Independent varibles “NumofProducts” and “Balance”
The PR(>F) for Num of Products is 1.489206e-06 which is very low.
The PR(>F) for Balance is 5.991070e-28 which is also very low.

model3=ols('Exited ~ NumOfProducts+Balance +CreditScore ', data=df).fit()

Num of Products and Balance also has a significant effect on the customer’s exit from the bank or leaving the bank.
These variables cause a considerable effect on the variability of the Dependent Variable.
**model3=ols('Exited ~ NumOfProducts+Balance +CreditScore ', data=df).fit()**

| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| NumOfProducts | 1.0 | 3.709236 | 3.709236 | 23.205576 | 1.477138e-06 |
| Balance | 1.0 | 19.328172 | 19.328172 | 120.920154 | 5.752290e-28 |
| CreditScore | 1.0 | 1.240711 | 1.240711 | 7.762087 | 5.345480e-03 |
| Residual | 9996.0 | 1597.784981 | 0.159842 | NaN | NaN |

This is model 3 which consists of a Dependent Variable “Exited” and 3 Independent variables “NumofProducts” , “Balance”and “CreditScore”.
The PR(>F) for Num of Products is 1.477138e-06 which is very low
The PR(>F) for Balance is 5.752290e-28 which is extremely low
The PR(>F) for CreditScore is 5.345480e-03 which is also low. (weaker effect as compared to the above two)

All these 3 variables have a significant effect on the outcome of the Dependent Variable “Exited” , but will have varying degrees of importance

- All the 3 Models have a low p-values for the F-statistic, indicating that they have significant effect on the outcome of the Dependent Variable.
- Model 3 consists of 3 Independent Variables which is Number of Products, Balance and Credit Score. It has the lowest p-value which makes it the best fit Model.
- Model 3 also determines the most variability as compared to Model 1 and Model 2.
- However, in Model 3 , the CreditScore Variable has a slightly higher p-value for F statistic indicating that it will have a weaker effect on the Dependent Variable which is “Exited” as compared to the other 2 variables.

# Conclusion

The customer churn is influenced by various factors like Credit Scores, Balance of Customers, Number of Products associated with the bank etc, it is not affected by one factor. It is a combination of various factors which leads to customer churn. Though there was no linear relation, financial stability is important for having long -term relationships with the bank. Customers who have a decent account balance will have loyalty towards the bank. Offering a variety of products to customers is important as it will keep the customers interested, even if they are not interested in purchasing now.

What can businesses or Banks do with this information?

- Businesses or banks with this information can keep the customers with good Credit scores happy by offering them discounts priority customer service etc.
- Providing customers some financial resources to improve their credit score.
- Offering incentives and bonuses to customers to customers who maintain their account balances and who are into long-term investments.
- Providing consultancy or financial advice to customers to help them understand the benefits of investing in their accounts.
- Create personalized product recommendations  which would purely be based on customer spending habits, behavioural patterns etc.
- Market Research can be conducted to identify customer prefereneces offer products based on that.

By working on these strategies businesses and banks can attract and retain customers, improve customer loyalty and reduce customer churn rates. It is about creating strategies which can improve customer satisfaction and identifying customer needs and preferences.

# References


- Divu2001. “Customer Churn Rate” [Data set]. Kaggle.
- https://www.kaggle.com/datasets/divu2001/customer-churn-rate
- aryamanpillay. (2024). Final Project [Jupyter notebook]. Anaconda Server.
- http://localhost:8889/notebooks/Downloads/Final%20Project.ipynb

# Appendix: Regression Models, Codes, and Figures

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
from statsmodels.formula.api import ols
from statsmodels.stats.anova import anova_lm
from scipy.stats import norm
```

These are basically the libraries used for importing the data, creating histograms , scatterplots , regression analysis etc.

```
df=pd.read_csv("Churn_Modelling (1).csv")
```
**print(df) :** This basically reads the dataset and stores the Dataset in a dataframe df

**df.isnull().sum():**

```
RowNumber          0
CustomerId         0
Surname            0
CreditScore        0
Geography          0
Gender             0
Age                0
Tenure             0
Balance            0
NumOfProducts      0
HasCrCard          0
IsActiveMember     0
EstimatedSalary    0
Exited             0
dtype: int64
```

## df.head(20)

| | RowNumber | CustomerId | Surname | CreditScore | Geography | Gender | Age | Tenure | Balance | NumOfProducts | HasCrCard | IsActiveMember | EstimatedSalary | Exited |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 15634602 | Hargrave | 619 | France | Female | 42 | 2 | 0.00 | 1 | 1 | 1 | 101348.88 | 1 |
| **1** | 2 | 15647311 | Hill | 608 | Spain | Female | 41 | 1 | 83807.86 | 1 | 0 | 1 | 112542.58 | 0 |
| **2** | 3 | 15619304 | Onio | 502 | France | Female | 42 | 8 | 159660.80 | 3 | 1 | 0 | 113931.57 | 1 |
| **3** | 4 | 15701354 | Boni | 699 | France | Female | 39 | 1 | 0.00 | 2 | 0 | 0 | 93826.63 | 0 |
| **4** | 5 | 15737888 | Mitchell | 850 | Spain | Female | 43 | 2 | 125510.82 | 1 | 1 | 1 | 79084.10 | 0 |

In [6]:

**df.tail()**

| | RowNumber | CustomerId | Surname | CreditScore | Geography | Gender | Age | Tenure | Balance | NumOfProducts | HasCrCard | IsActiveMember | EstimatedSalary | Exited |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **9995** | 9996 | 15606229 | Obijiaku | 771 | France | Male | 39 | 5 | 0.00 | 2 | 1 | 0 | 96270.64 | 0 |
| **9996** | 9997 | 15569892 | Johnstone | 516 | France | Male | 35 | 10 | 57369.61 | 1 | 1 | 1 | 101699.77 | 0 |
| **9997** | 9998 | 15584532 | Liu | 709 | France | Female | 36 | 7 | 0.00 | 1 | 0 | 1 | 42085.58 | 1 |
| **9998** | 9999 | 15682355 | Sabbatini | 772 | Germany | Male | 42 | 3 | 75075.31 | 2 | 1 | 0 | 92888.52 | 1 |
| **9999** | 10000 | 15628319 | Walker | 792 | France | Female | 28 | 4 | 130142.79 | 1 | 1 | 0 | 38190.78 | 0 |

In [7]:

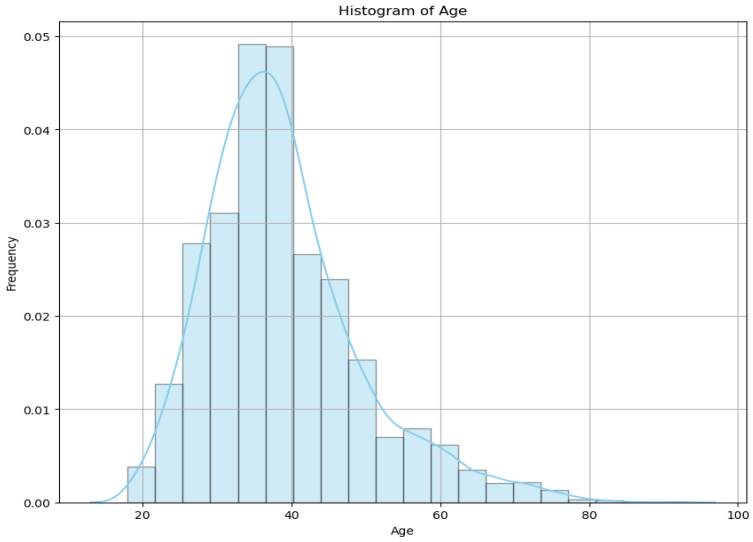
## df.describe()

| | RowNumber | CustomerId | CreditScore | Age | Tenure | Balance | NumOfProducts | HasCrCard | IsActiveMember | EstimatedSalary | Exited |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 10000.000000 | 1.000000e+04 | 10000.000000 | 10000.000000 | 10000.000000 | 10000.000000 | 10000.000000 | 10000.00000 | 10000.000000 | 10000.000000 | 10000.000000 |
| **mean** | 5000.50000 | 1.569094e+07 | 650.528800 | 38.921800 | 5.012800 | 76485.889288 | 1.530200 | 0.70550 | 0.515100 | 100090.239881 | 0.203700 |
| **std** | 2886.89568 | 7.193619e+04 | 96.653299 | 10.487806 | 2.892174 | 62397.405202 | 0.581654 | 0.45584 | 0.499797 | 57510.492818 | 0.402769 |
| **min** | 1.00000 | 1.556570e+07 | 350.000000 | 18.000000 | 0.000000 | 0.000000 | 1.000000 | 0.00000 | 0.000000 | 11.580000 | 0.000000 |
| **25%** | 2500.75000 | 1.562853e+07 | 584.000000 | 32.000000 | 3.000000 | 0.000000 | 1.000000 | 0.00000 | 0.000000 | 51002.110000 | 0.000000 |
| **50%** | 5000.50000 | 1.569074e+07 | 652.000000 | 37.000000 | 5.000000 | 97198.540000 | 1.000000 | 1.00000 | 1.000000 | 100193.915000 | 0.000000 |
| **75%** | 7500.25000 | 1.575323e+07 | 718.000000 | 44.000000 | 7.000000 | 127644.240000 | 2.000000 | 1.00000 | 1.000000 | 149388.247500 | 0.000000 |

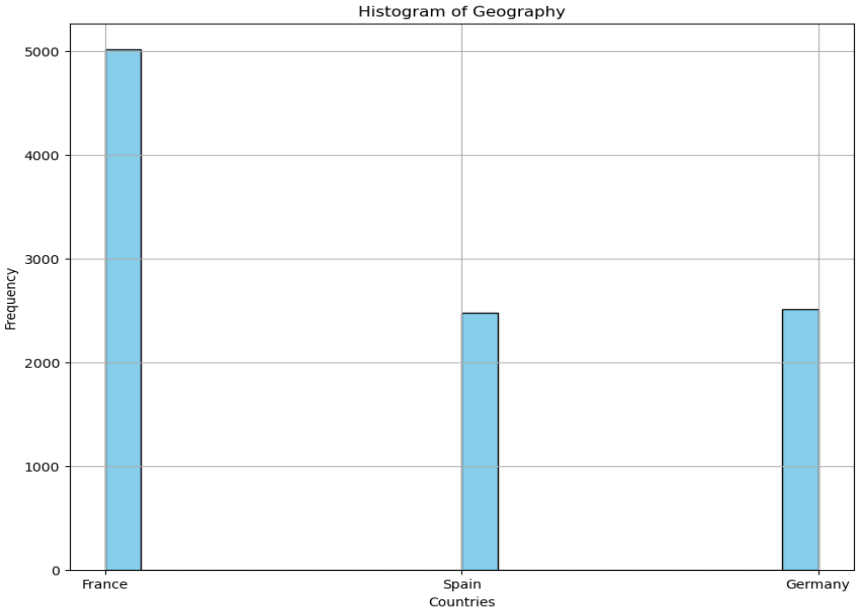
| | RowNumber | CustomerId | CreditScore | Age | Tenure | Balance | NumOfProducts | HasCrCard | IsActiveMember | EstimatedSalary | Exited |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **max** | 10000.000000 | 1.581569e+07 | 850.000000 | 92.000000 | 10.000000 | 250898.090000 | 4.000000 | 1.00000 | 1.000000 | 199992.480000 | 1.000000 |

In [8]:

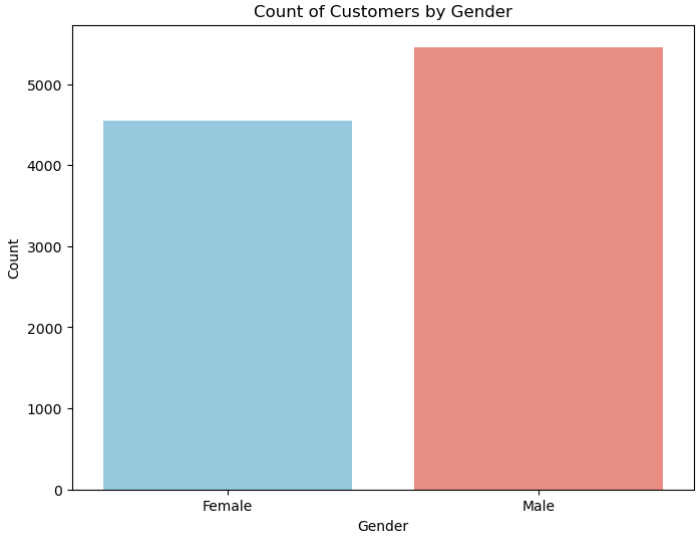
```
plt.figure(figsize =(10,8))
sns.distplot(df['Age'], bins=20, kde=True,hist=True, color='skyblue', hist_kws={'edgecolor': 'black'})
plt.title('Histogram of Age')  # Removed '=' sign, use parentheses for method calls
plt.xlabel('Age')  # Removed '=' sign, use parentheses for method calls
plt.ylabel('Frequency')
plt.grid(True)  # Removed '=' sign, use parentheses for method calls
plt.show()
```



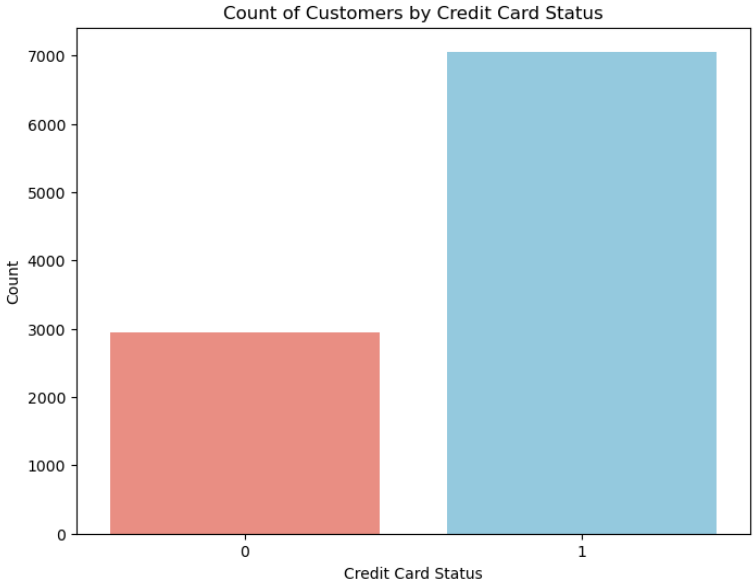
```
plt.figure(figsize=(10,8))
plt.hist(df['Geography'],bins=20,color='skyblue',edgecolor='black')
plt.title('Histogram of Geography')
plt.xlabel('Countries')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```


Histogram of Geography
Frequency
5000
4000
3000
2000
1000
0
France
Spain
Germany
Countries

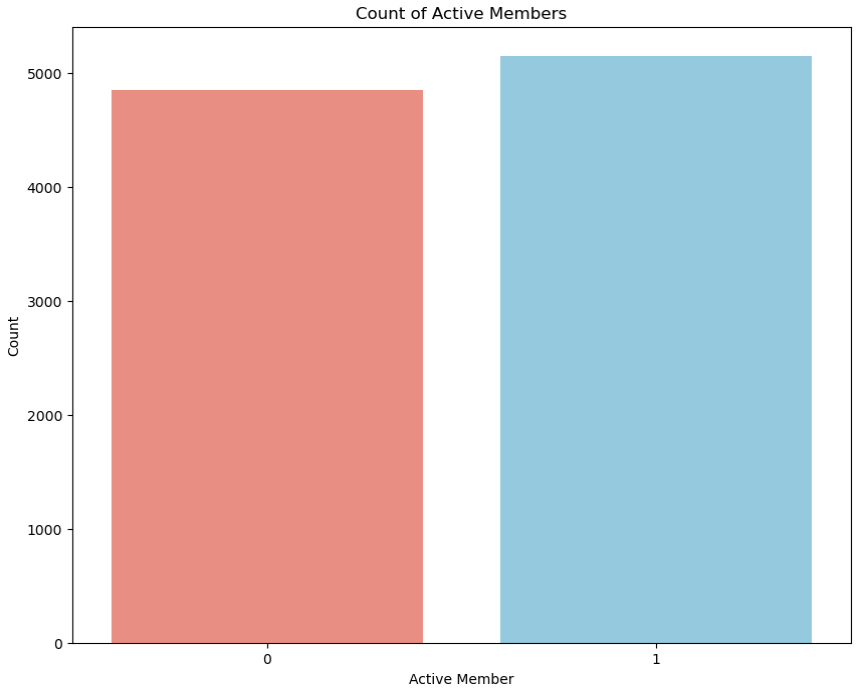
```
plt.figure(figsize=(8, 6))
sns.countplot(data=df, x='Gender', palette={'Female': 'skyblue', 'Male': 'salmon'})
plt.title('Count of Customers by Gender')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.show()
```



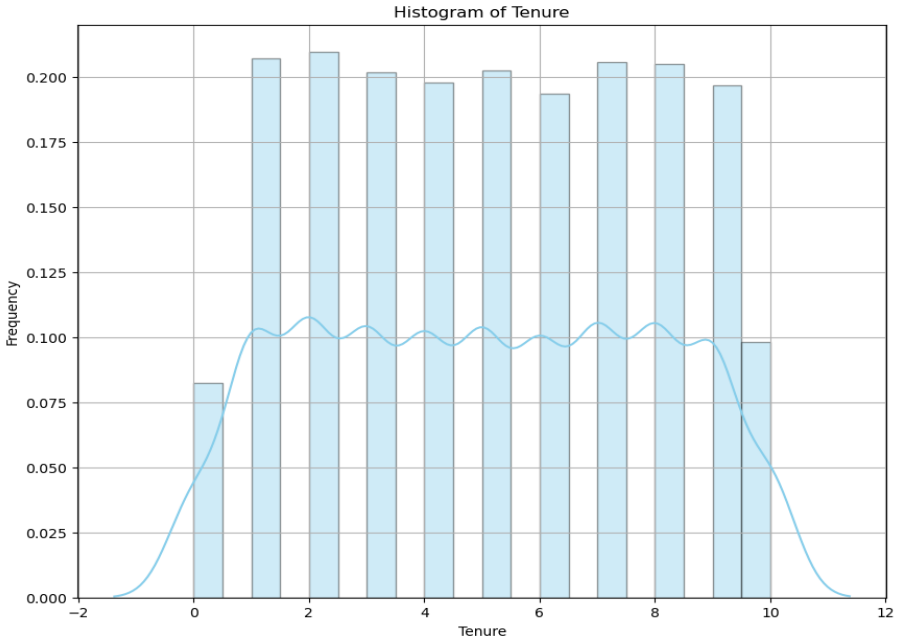
```
plt.figure(figsize=(8, 6))
sns.countplot(data=df, x='HasCrCard', palette={1: 'skyblue', 0: 'salmon'})
plt.title('Count of Customers by Credit Card Status')
plt.xlabel('Credit Card Status')
plt.ylabel('Count')
plt.show()
```



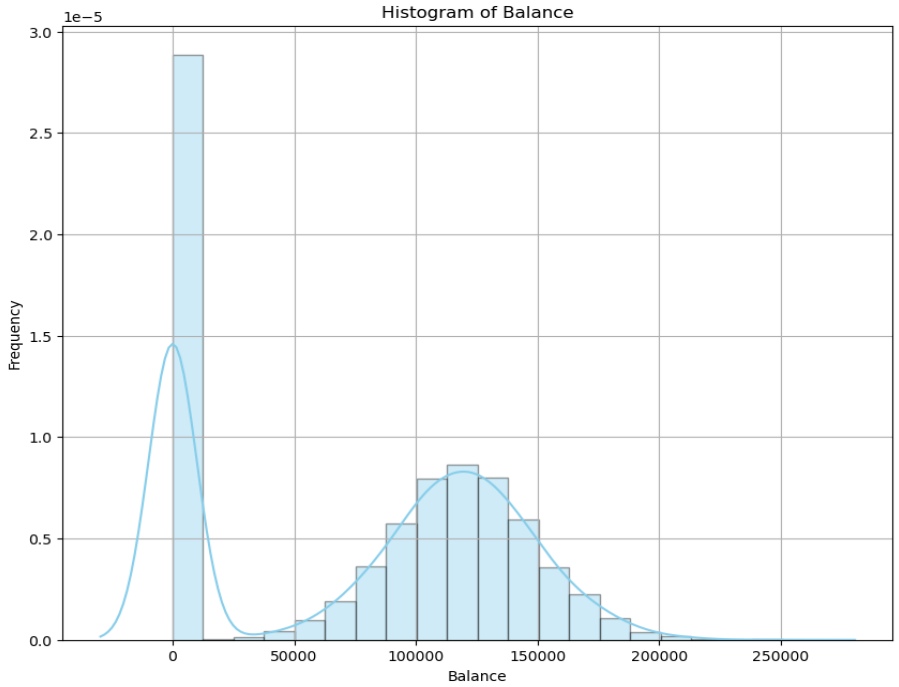
```
plt.figure(figsize=(10,8))
sns.countplot(data=df,x='IsActiveMember' ,palette={1: 'skyblue', 0: 'salmon'})
plt.title('Count of Active Members')
plt.xlabel('Active Member')
plt.ylabel('Count')
plt.show()
```



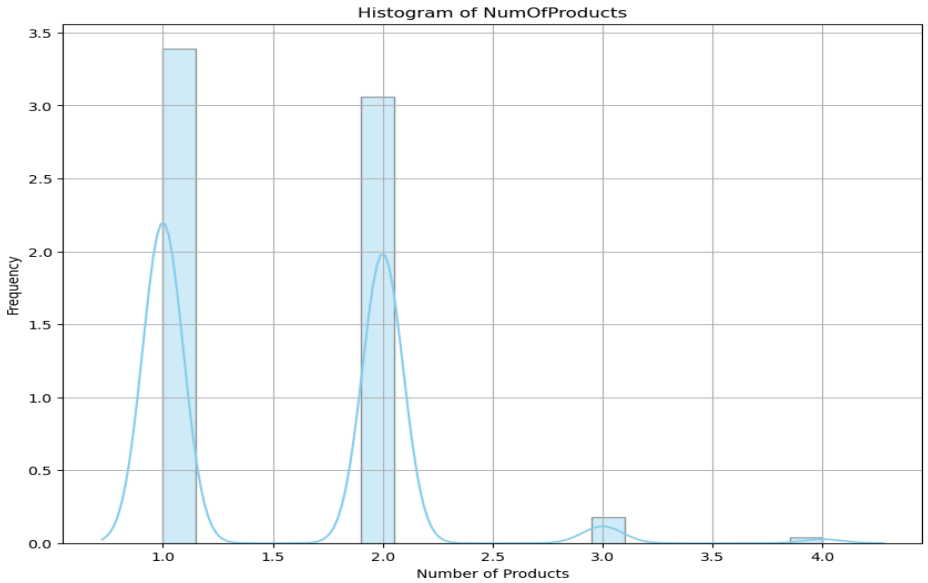
```
plt.figure(figsize=(10,8))
sns.distplot(df['Tenure'], bins=20, kde=True,hist=True, color='skyblue', hist_kws={'edgecolor': 'black'})
plt.title('Histogram of Tenure')
plt.xlabel('Tenure')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```



```
plt.figure(figsize=(10,8))
sns.distplot(df['Balance'], bins=20, kde=True,hist=True, color='skyblue', hist_kws={'edgecolor': 'black'})
plt.title('Histogram of Balance')
plt.xlabel('Balance')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```



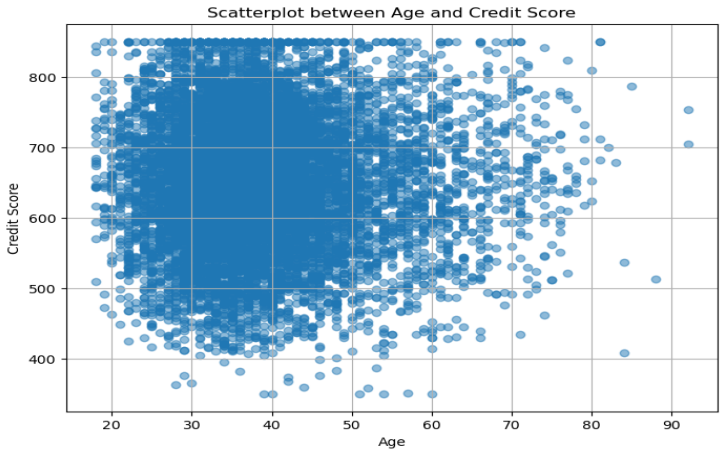
```
plt.figure(figsize=(10,8))
sns.distplot(df['NumOfProducts'], bins=20, kde=True,hist=True, color='skyblue', hist_kws={'edgecolor': 'black'})
plt.title('Histogram of NumOfProducts')
plt.xlabel('Number of Products')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```



```
correlation_coefficient = df['Age'].corr(df['CreditScore'])
print(correlation_coefficient)
```

```
-0.003964905525390069 5
```

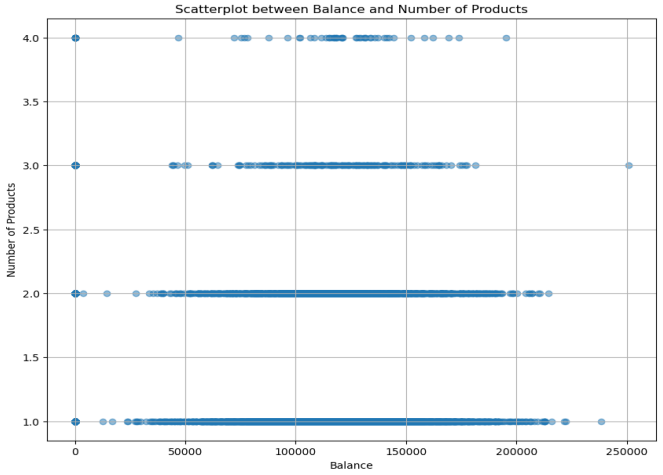
```
plt.figure(figsize=(8, 6))
plt.scatter(df['Age'], df['CreditScore'], alpha=0.5)
plt.title('Scatterplot between Age and Credit Score')
plt.xlabel('Age')
plt.ylabel('Credit Score')
plt.grid(True)
plt.show()
```



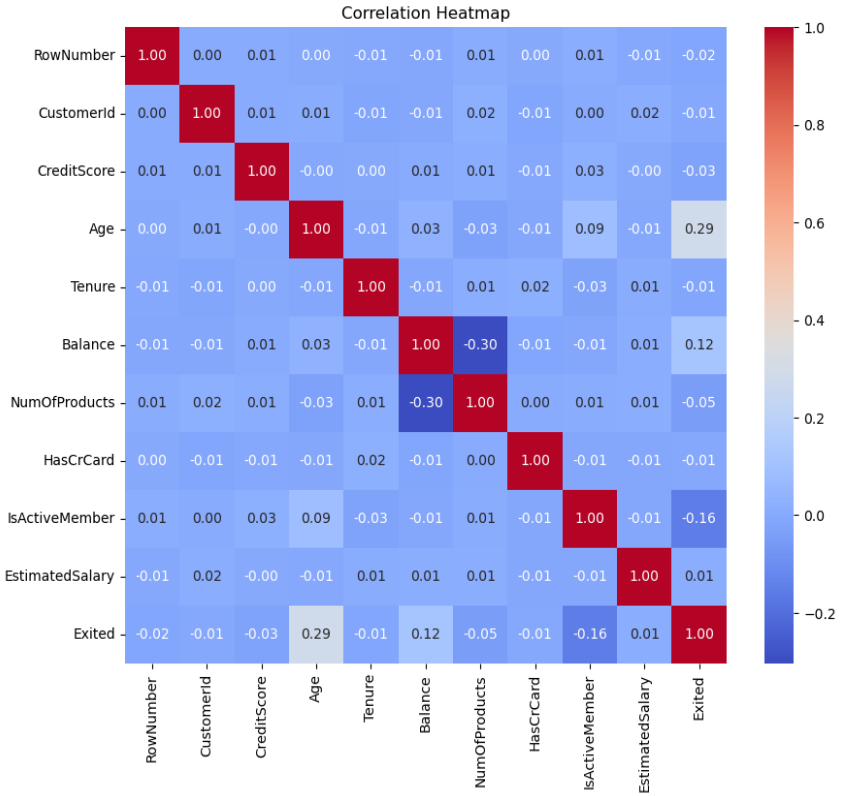
```
correlation_coefficient = df['Balance'].corr(df['NumOfProducts'])
print(correlation_coefficient)
```

```
-0.3041797383605491
```

```
plt.figure(figsize=(10,8))
plt.scatter(df['Balance'],df['NumOfProducts'],alpha=0.5)
plt.title('Scatterplot between Balance and Number of Products')
plt.xlabel('Balance')
plt.ylabel('Number of Products')
plt.grid(True)
plt.show()
```



```
# Create heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', fmt=".2f", annot_kws={"size": 10})
plt.title('Correlation Heatmap')
plt.show()
```



**model1 = ols('Exited ~ NumOfProducts', data=df).fit()**

**# Perform ANOVA**

**anova_table = anova_lm(model1)**

**# Print the ANOVA table**

**print(anova_table)**

```
                  df       sum_sq   mean_sq          F    PR(>F)
NumOfProducts     1.0     3.709236  3.709236  22.915223  0.000002
Residual       9998.0  1618.353864  0.161868        NaN       NaN
```

**model2 = ols('Exited ~ NumOfProducts + Balance', data=df).fit()**

**# Perform ANOVA**

**anova_table = anova_lm(model2)**

**# Print the ANOVA table**

**print(anova_table)**

```
                  df       sum_sq    mean_sq           F        PR(>F)
NumOfProducts     1.0     3.709236   3.709236   23.189890  1.489206e-06
Balance           1.0    19.328172  19.328172  120.838417  5.991070e-28
Residual       9997.0  1599.025692   0.159951         NaN           NaN
```

**model3=ols('Exited ~ NumOfProducts+Balance +CreditScore ', data=df).fit()**

**# Perform ANOVA**

**anova_table = anova_lm(model3)**

**# Print the ANOVA table**

**print(anova_table)**

| | df | sum_sq | mean_sq | F | PR(>F) | |
|---|---|---|---|---|---|---|
| NumOfProducts | 1.0 | 3.709236 | 3.709236 | 23.205576 | 1.477138e-06 | |
| Balance | 1.0 | 19.328172 | 19.328172 | 120.920154 | 5.752290e-28 | |
| CreditScore | 1.0 | 1.240711 | 1.240711 | 7.762087 | 5.345480e-03 | |
| Residual | 9996.0 | 1597.784981 | 0.159842 | NaN | NaN | |