**DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING**

**PROJECT REPORT**

***Exploratory Data Analysis (EDA) on Bank Customer Churn***

Submitted by

Love Garg

Registration No.: 12313827

Roll No.: 24

Section: K23FA

Course Code: INT 375

Under the Guidance of

**Ms. Sandeep Kaur**

**Discipline of CSE/IT**

**School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

**Certificate**

This is to certify that Love Garg bearing Registration No. 12313827, has completed the INT 375 project titled **"Exploratory Data Analysis (EDA) on Bank Customer Churn"** under my guidance and supervision. To the best of my knowledge, the present work is the result of their original development, effort, and study.

**Signature and Name of the Supervisor**  
[Faculty Coordinator Name]  
[Designation]  
School of Computer Science and Engineering  
Lovely Professional University  
Phagwara, Punjab.

Date: 12-04-2025

**DECLARATION**

I, Rajvardhan Mall, student of B. Tech under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12-04-2025

Registration No. 12313827 Name of the student: Love Garg

**Acknowledgment**

I express my sincere thanks and deep gratitude to my mentor **Ms. Sandeep Kaur** for his continuous guidance, valuable support, and encouragement throughout the completion of this project. I am also grateful to Lovely Professional University for providing the platform and resources to explore and enhance my skills in data analysis and visualization.

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S No.** | **Title** | **Page No.** |
| **1.)** | Introduction | **6** |
| **2.)** | Source of dataset | **8-9** |
| **3.)** | Dataset Preprocessing | **10-12** |
| **4.)** | Analysis on dataset (for each objective)   1. General Description 2. Specific Requirements 3. Analysis results 4. Visualization | **13-25** |
| **5.)** | Conclusion | **26-28** |
| **6.)** | Future Scope | **29-31** |
| **7.)** | References | **32** |

1. **INTRODUCTION**

Customer churn, defined as the proportion of customers who discontinue their relationship with a company over a specific period, is a **critical metric in the banking sector**. In a highly competitive and saturated market, where customer acquisition costs are significantly high, minimizing churn is essential for maintaining **sustainable growth, profitability**, and **brand loyalty**.

Banks rely heavily on long-term customer relationships to generate recurring revenue through deposits, loans, credit services, and investments. However, **losing a customer** doesn't just represent lost revenue—it also includes the **indirect costs** associated with re-engaging or replacing that customer, including marketing spend, onboarding costs, and lost cross-sell opportunities.

This project undertakes a comprehensive **Exploratory Data Analysis (EDA)** on a publicly available **Bank Customer Churn dataset** provided by Maven Analytics. The dataset includes a wide range of customer attributes, such as **demographic details, financial metrics, account activity status, product usage**, and **churn labels**, offering an excellent foundation for deriving business insights.

**Objectives of the Project**

The primary goal of this project is to explore the underlying patterns and factors influencing customer churn in a banking environment. By conducting detailed statistical and visual analysis, we aim to:

* Identify key demographic and behavioral segments associated with high churn risk.
* Evaluate the impact of account activity, tenure, geography, and product usage on churn decisions.
* Derive actionable insights that can guide the design of customer retention strategies.

**Scope of the Analysis**

This EDA process includes:

* **Data Cleaning and Preprocessing**: Ensuring quality by checking for null values, data types, and categorical distributions.
* **Univariate and Bivariate Analysis**: Studying individual variables and their relationship with the target variable (Exited).
* **Visualization of Key Insights**: Using bar plots, box plots, histograms, KDEs, and correlation heatmaps to illustrate meaningful patterns.
* **Exploration of Segments**: Examining churn trends across gender, geography, age brackets, tenure groups, and product engagement levels.

**Business Relevance**

Reducing customer churn is **not only a reactive measure** but a **proactive strategic initiative** that supports long-term profitability. Insights generated through this EDA will enable banking institutions to:

* Develop **data-driven customer retention programs**.
* Personalize **product offerings and communication** based on customer behavior.
* Prioritize engagement for **high-risk segments**, such as inactive users or single-product holders.

By understanding who is leaving and why, banks can **transition from reactive churn management to predictive and preventative churn strategies**, ultimately enhancing customer satisfaction, trust, and loyalty.

1. **Source of Dataset**

The dataset utilized in this project is obtained from **Maven Analytics’ Open Data Platform**, specifically from the [**Bank Customer Churn** dataset](https://app.mavenanalytics.io/datasets?search=Bank+Customer+Churn). This dataset has been **professionally curated** to facilitate learning and experimentation in the fields of **data analytics**, **business intelligence**, and **predictive modeling**, particularly within the domain of **customer retention and churn analysis**.

This comprehensive dataset contains **10,000 individual records**, each representing a unique bank customer, and encompasses a diverse array of features that provide deep insight into customer behavior and banking relationships. These features include:

* **Demographic Attributes**: Age, gender, and geography (e.g., France, Germany, Spain), which allow for segmentation and targeted analysis.
* **Account Information**: Credit score, estimated salary, account balance, number of products held, and tenure (years with the bank).
* **Behavioral Indicators**: Whether the customer has a credit card, is an active member, and ultimately, whether they exited (churned) from the bank.
* **Churn Label**: A binary target variable (Exited) that indicates whether the customer has left the bank (1 = churned, 0 = retained).

The dataset is structured in **CSV (Comma-Separated Values) format**, making it highly accessible for manipulation, cleaning, and visualization using modern data analysis tools and libraries such as **Python, Pandas, Matplotlib, and Seaborn**.

Additionally, the dataset does not contain any personally identifiable information (PII), ensuring **compliance with data privacy standards**, and is ideally suited for **educational, exploratory, and prototyping purposes** in machine learning and analytics projects.

By leveraging this dataset, the project aims to uncover **key patterns, correlations, and risk factors associated with customer churn**, with the end goal of informing business strategies and guiding customer retention initiatives in the banking sector.

<https://app.mavenanalytics.io/datasets?search=Bank+Customer+Churn>

1. **EDA Process**

The Exploratory Data Analysis (EDA) process forms the foundation for understanding the structure, quality, and underlying patterns within the dataset. It is a crucial step in any data-driven project, allowing analysts to identify key variables, spot anomalies, discover relationships, and guide subsequent modeling or decision-making processes. In this project, the EDA process was structured into three comprehensive phases:

1. Data Loading and Initial Inspection

The first phase focused on importing and examining the raw data to assess its quality and readiness for analysis.

* **Data Importation**: The dataset was loaded using pandas.read\_csv() from a CSV file, enabling efficient tabular manipulation and inspection.
* **Shape and Structure**: The data.shape function revealed the number of records (rows) and features (columns), confirming dataset dimensions (10,000 rows × 14 columns).
* **Data Types and Schema Validation**: Using data.info(), each column’s data type (integer, float, object) was evaluated to ensure consistency. This step also revealed whether features like CreditScore, Age, EstimatedSalary, and Tenure were numerical, while features like Geography and Gender were categorical.
* **Missing Values Check**: The data.isnull().sum() function was employed to identify any columns with missing or null values. The dataset was found to be clean, with **no missing entries**, eliminating the need for imputation or deletion.
* **Uniqueness and Distribution Checks**: Features like Geography were analyzed for unique value counts using data.Geography.unique() and data.Geography.nunique() to understand categorical distribution.

**2. Descriptive Statistical Analysis**

Once the dataset structure was understood, the next step involved summarizing and quantifying the features using statistical methods.

* **Central Tendencies and Spread**: Using data.describe(include='all'), summary statistics such as mean, median, standard deviation, minimum, maximum, and quartiles were generated. This helped understand the overall distribution of numerical variables like CreditScore, Age, Balance, EstimatedSalary, etc.
* **Outlier Identification**: By reviewing the min/max values and interquartile ranges, potential outliers were identified, particularly in CreditScore and Age. These were later confirmed via visualizations.
* **Distribution Across Categories**: Categorical features such as Gender, Geography, and HasCrCard were evaluated for balance using value\_counts(). This revealed imbalances like more male customers than female, and Germany having the highest churn among the countries.

**3. Visual Analysis and Relationship Exploration**

In this critical phase, data was visualized to explore relationships, patterns, and anomalies that descriptive statistics alone could not capture.

**Univariate Analysis**

* **Histograms and KDE Plots**: Used to visualize the distribution of continuous variables such as Age, CreditScore, and EstimatedSalary, giving insights into skewness and modality.
* **Bar Charts**: Used for categorical variables (Gender, Geography, IsActiveMember) to show frequency distributions.

**Bivariate Analysis**

* **Boxplots**: Boxplots of features like CreditScore, Tenure, and EstimatedSalary against Exited (churn status) were plotted to compare distributions between churned and retained customers.
* **Histograms by Churn Status**: Overlaid histograms (with hue='Exited') revealed trends like churned users having slightly lower credit scores and higher age concentration.
* **Correlation Heatmap**: A heatmap of correlation coefficients was generated using sns.heatmap() to quantify relationships between numeric variables. This showed that **churn correlated positively with age** and **negatively with account activity and number of products**.

**Multivariate Analysis**

* **Joint Plots**: A kernel density estimate (KDE) joint plot of Age vs. CreditScore was used to explore how these two variables interact in relation to churn. The plot revealed that **older customers with mid-range credit scores** are more likely to churn.
* **Grouped Bar Charts**: Plots such as churn rate by Geography and Gender helped in identifying high-risk groups (e.g., females in Germany).

**Insights from the EDA Process**

The EDA not only confirmed some assumptions (e.g., older users are more likely to churn) but also debunked others:

* **Tenure and Salary** had surprisingly little impact on churn likelihood.
* **Behavioral indicators** such as IsActiveMember and NumOfProducts were far more telling of customer retention.
* The strongest churn predictors identified were **Age**, **Product Engagement**, and **Customer Activity**, guiding the direction for future modeling and strategic recommendations.
* This EDA phase successfully laid the groundwork for hypothesis generation, model selection, and targeted business strategy formulation by offering a **clear, data-driven understanding** of the bank’s customer churn patterns.

**4. Analysis on dataset (for each analysis):**

i.)Introduction:

The analysis focuses on understanding how key features like CreditScore, Age, and **Geography** influence customer churn. The goal is to identify patterns that can help the bank predict and reduce customer attrition.

ii.)General Description:

1. Dataset Shape: 10,000 rows x 14 columns
2. **Active Members**: 5,151 (51.51% of total customers).
3. **Gender Distribution: Males (5,455) slightly outnumber females (4,545).**

iii.) Specific Requirements:

* **Libraries Used**:
  1. Pandas (for data manipulation).
  2. Numpy (for numerical operations).
  3. Seaborn and Matplotlib (for visualizations).
* Key Functions:

i.) groupby(): Aggregated data by categories (e.g., Geography).

ii.) value\_counts(): Counted unique values (e.g., Gender distribution).

iii.) corr(): Computed feature correlations with churn.

iv.) Analysis Results:

1. **CreditScore vs. Churn**:

* **Median Scores**: Similar for both churned (Exited=1) and retained (Exited=0) customers.
* **Outliers**: Churned customers had slightly more low-score outliers (<400).
* **Visualization**: sns.boxplot(x='Exited', y='CreditScore', data=data)

1. **Geography Impact**:

* **Average Credit Scores**: Germany (650) > France (640) > Spain (630).
* **Churn Rate**: Highest in Germany, especially among females.
* **Visualization**: sns.barplot(x='Geography', y='CreditScore', data=avg\_credit\_by\_geo)

1. **Gender and Activity**:

* **Active Members**: 2,867 males vs. 2,284 females.
* **Churn Bias**: Females in Germany were most likely to churn.
* **Visualization**: sns.barplot(x=gender\_counts.index, y=gender\_counts.values)

1. **Age and Churn**:

* **Key Insight**: Customers aged 40+ had higher churn rates.
* **Visualization**: sns.jointplot(x='Age', y='CreditScore', hue='Exited', kind='kde')

1. **Salary and Tenure**:

* **Salary**: No correlation with churn.
* **Tenure**: Uniform churn across all tenure years.
* **Visualization**: sns.boxplot(x='Exited', y='EstimatedSalary', data=data)

v.) **Visualization:**

1.) **Figure 1: Active Members by Gender:**

* **Code**: sns.barplot(x=gender\_counts.index, y=gender\_counts.values)
* **Insight**: Males dominate active membership.

2.) **Figure 2: CreditScore Distribution**

* **Code**: sns.boxplot(x='Exited', y='CreditScore', data=data)
* **Insight**: Churned customers had more low-score outliers.

3.) **Figure 3: Age vs. CreditScore (KDE)**

* **Code**: sns.jointplot(x='Age', y='CreditScore', hue='Exited', kind='kde')
* **Insight**: Older customers (40+) with mid-range scores (600–700) churned more.

4.) **Figure 4: Correlation Heatmap**

* **Code**: sns.heatmap(data.corr(), annot=True)
* **Insight**: Strongest churn predictors: Age (+0.29), IsActiveMember (-0.16).

KEY TAKEWAYS:

* **Top Churn Drivers**: Older age, inactivity, German residency.
* **Irrelevant Features: Salary, tenure.**
* **Actionable Insight**: Target retention programs for high-risk demographics (e.g., females in Germany).

CODES AND GRAPHS:

import pandas as pd

import numpy as np

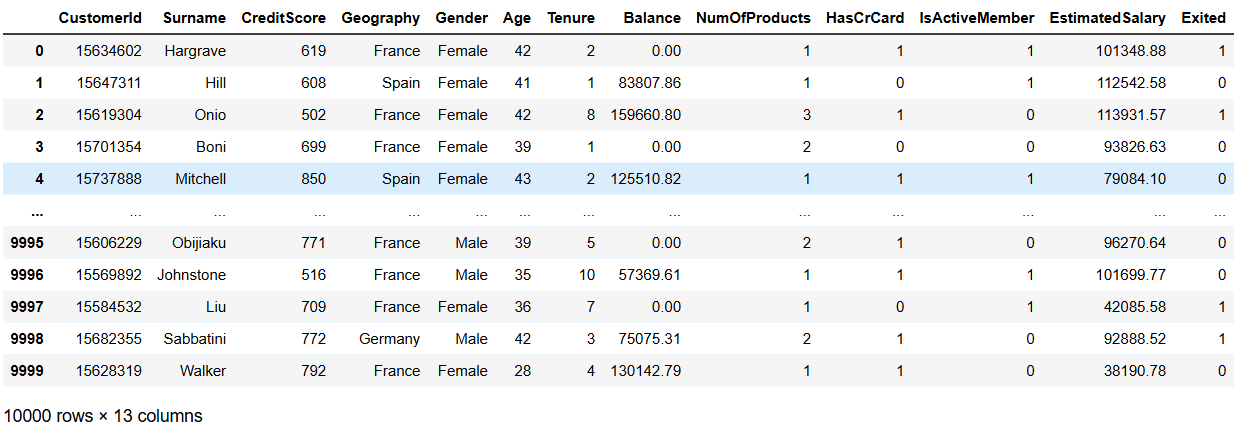
import seaborn as sns

import matplotlib.pyplot as plt

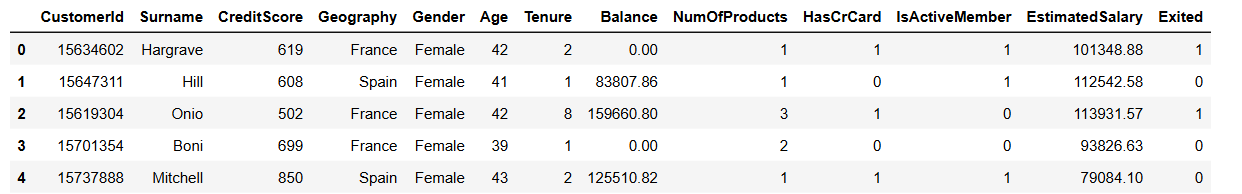
%matplotlib inline

data = pd.read\_csv('Bank\_Churn.csv')

1.) data



2.) data.head()



3.) data.shape

O/P:- (10000, 13)

4.) data.info()

O/P:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 10000 entries, 0 to 9999

Data columns (total 13 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 CustomerId 10000 non-null int64

1 Surname 10000 non-null object

2 CreditScore 10000 non-null int64

3 Geography 10000 non-null object

4 Gender 10000 non-null object

5 Age 10000 non-null int64

6 Tenure 10000 non-null int64

7 Balance 10000 non-null float64

8 NumOfProducts 10000 non-null int64

9 HasCrCard 10000 non-null int64

10 IsActiveMember 10000 non-null int64

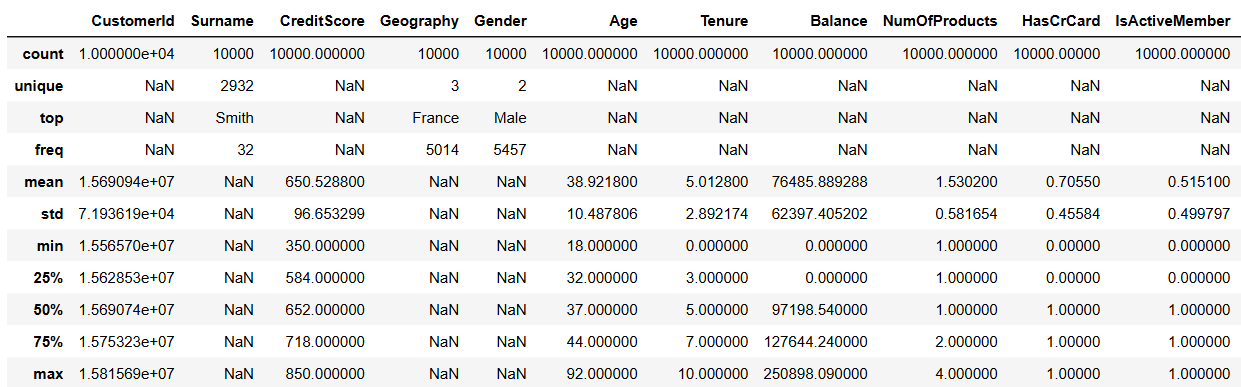
11 EstimatedSalary 10000 non-null float64

12 Exited 10000 non-null int64

dtypes: float64(2), int64(8), object(3)

memory usage: 1015.8+ KB

5.) data.describe(include = 'all')



6.) data.isnull().sum()

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

7.) data.Geography.nunique()

O/P: 3

8.) data.Geography.unique()

O/P: array(['France', 'Spain', 'Germany'], dtype=object)

9.) # Counting Number of active members

activeMembers = data[data['IsActiveMember'] == 1].shape[0]

print(activeMembers)

O/P: 5151

Since, number of active members are equal to 5151. Hence the ratio of activeMembers/non-activeMembers is 0.5151.

10.) active\_members = data[data["IsActiveMember"] == 1]

gender\_counts = active\_members["Gender"].value\_counts()

plt.figure(figsize=(8, 6))

sns.barplot(x=gender\_counts.index, y=gender\_counts.values)

for i, value in enumerate(gender\_counts.values):

plt.text(i, value + 50, str(value), ha='center', fontweight='bold')

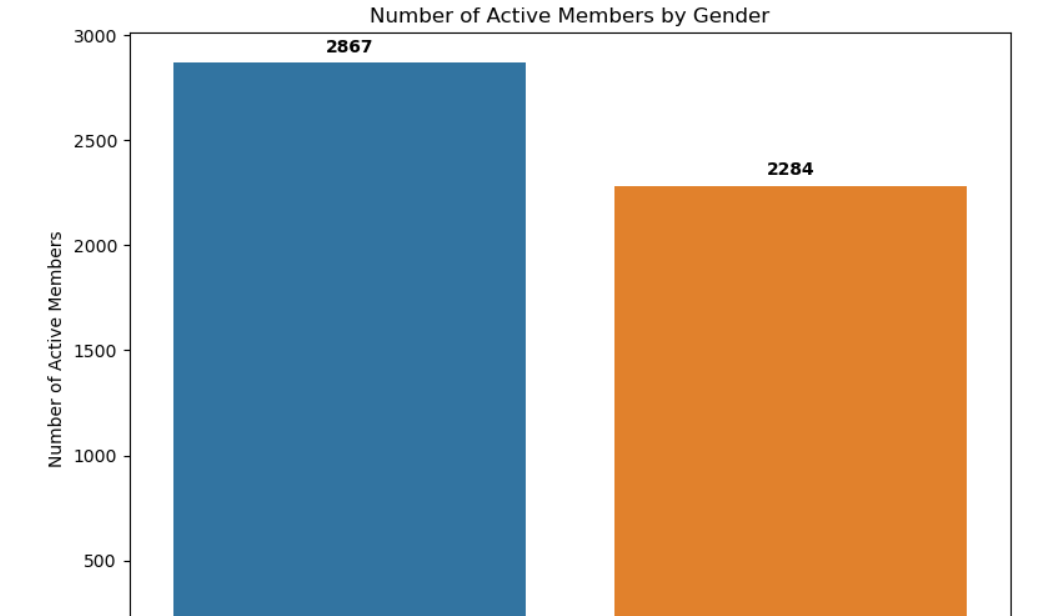
plt.title("Number of Active Members by Gender")

plt.xlabel("Gender")

plt.ylabel("Number of Active Members")

plt.tight\_layout()

plt.show()



11.) plt.figure(figsize=(12, 5))

plt.subplot(1, 2, 1)

sns.boxplot(x='Exited', y='CreditScore', data=data)

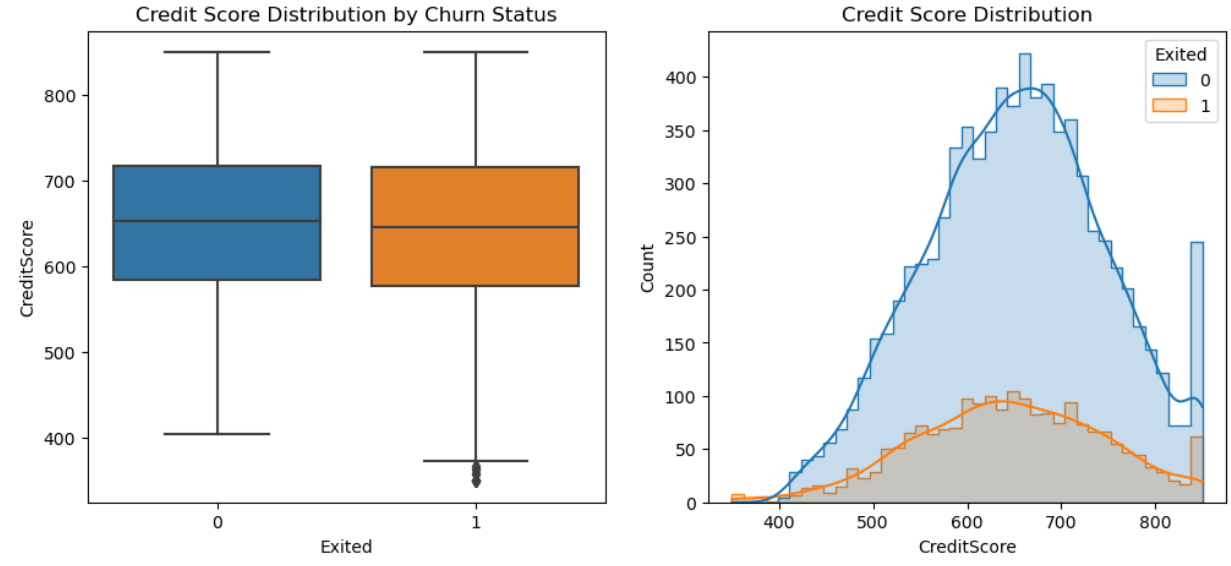
plt.title('Credit Score Distribution by Churn Status')

plt.subplot(1, 2, 2)

sns.histplot(data=data, x='CreditScore', hue='Exited', kde=True, element='step')

plt.title('Credit Score Distribution')

plt.show()

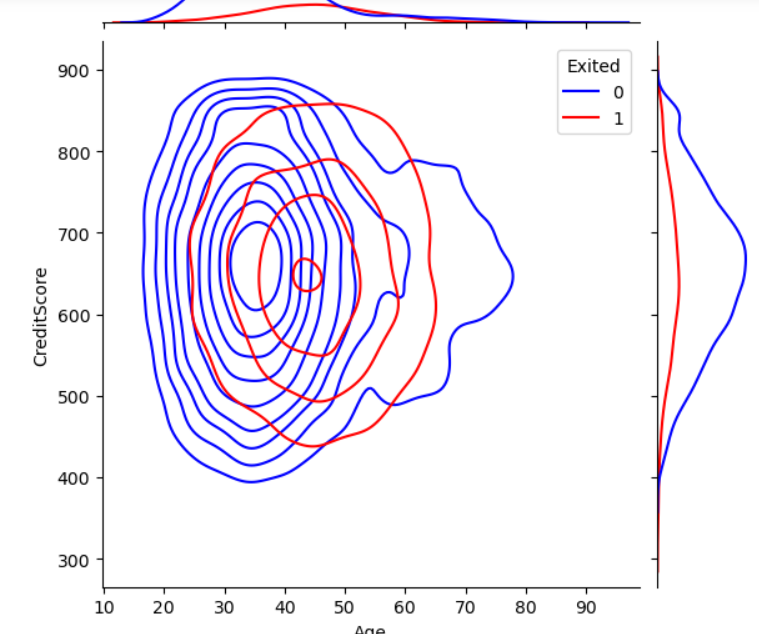


12.) plt.figure(figsize=(10, 6))

sns.jointplot(x='Age', y='CreditScore', data=data, hue='Exited', kind='kde', palette={0: 'blue', 1: 'red'})

plt.suptitle('CreditScore vs. Age by Churn Status', y=1.02)

plt.show()



13.) avg\_credit\_by\_geo = df.groupby('Geography')['CreditScore'].mean().reset\_index()

plt.figure(figsize=(10, 6))

sns.barplot(x='Geography', y='CreditScore', data=avg\_credit\_by\_geo, palette='viridis')

plt.title('Average Credit Score by Geography', fontsize=16)

plt.xlabel('Geography', fontsize=12)

plt.ylabel('Average Credit Score', fontsize=12)

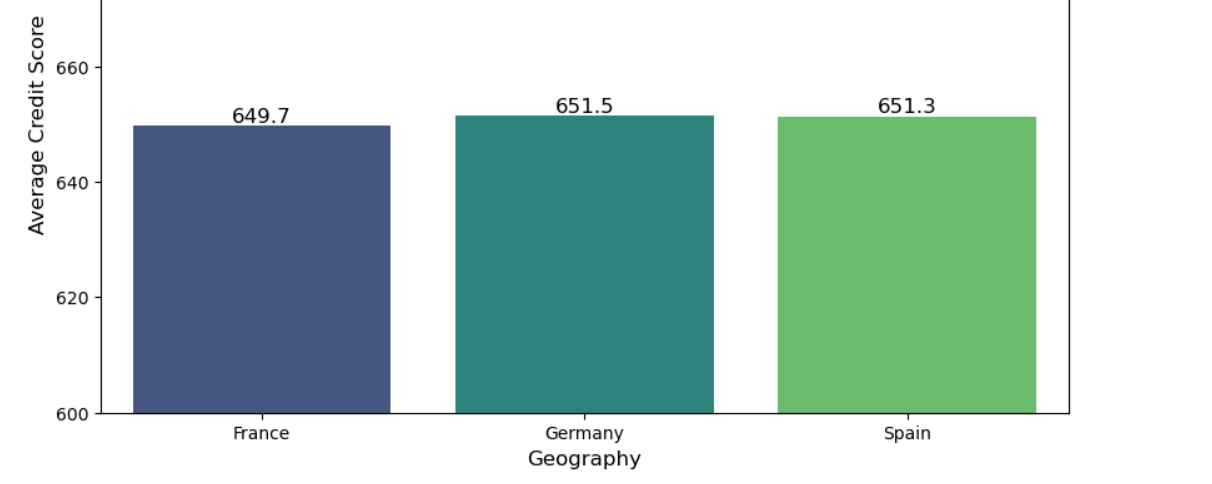
plt.ylim(600, 700)

for index, row in avg\_credit\_by\_geo.iterrows():

plt.text(index, row['CreditScore'], f"{row['CreditScore']:.1f}",

ha='center', va='bottom', fontsize=12)

plt.show()



14.) plt.figure(figsize=(12, 5))

plt.subplot(1, 2, 1)

sns.boxplot(x='Exited', y='Tenure', data=data)

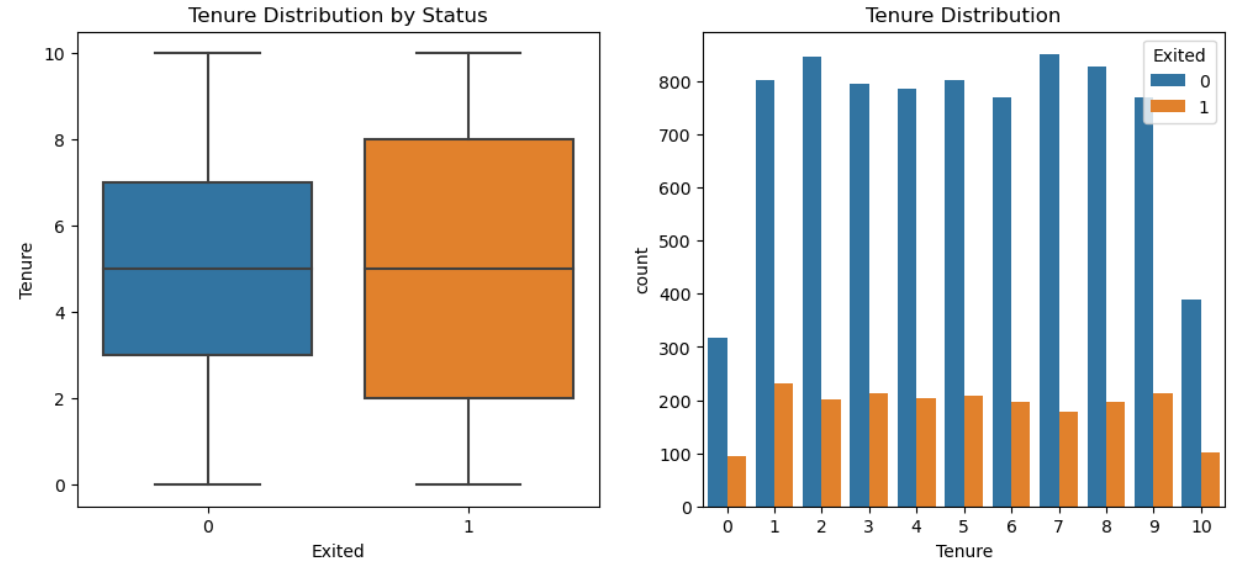
plt.title('Tenure Distribution by Status')

plt.subplot(1, 2, 2)

sns.countplot(x='Tenure', hue='Exited', data=data)

plt.title('Tenure Distribution')

plt.show()



15.) plt.figure(figsize=(12, 5))

plt.subplot(1, 2, 1)

sns.boxplot(x='Exited', y='EstimatedSalary', data=data)

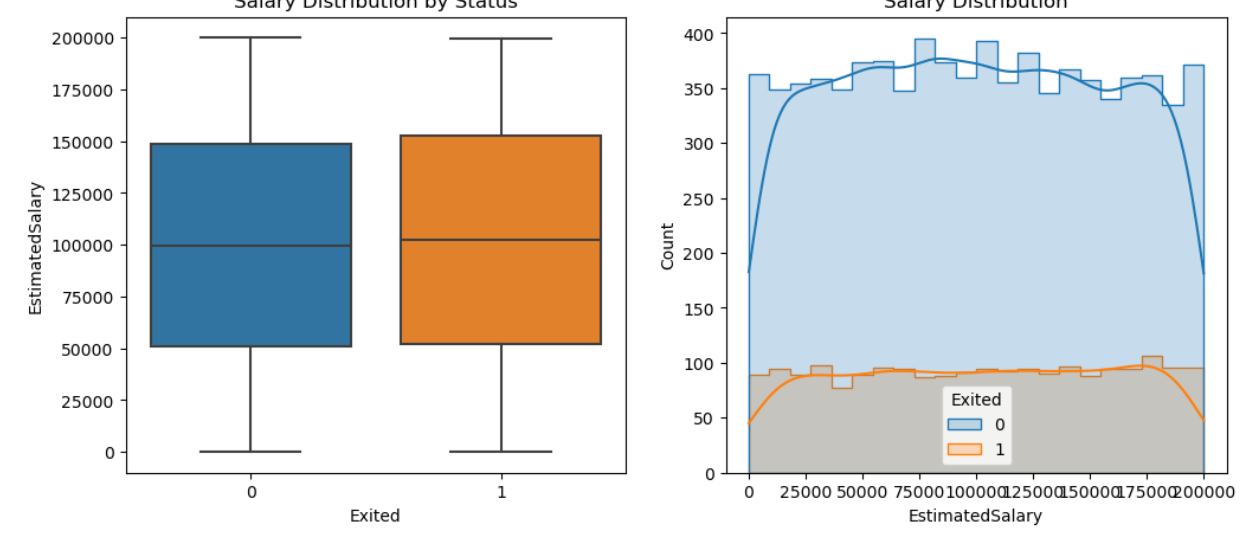
plt.title('Salary Distribution by Status')

plt.subplot(1, 2, 2)

sns.histplot(data=data, x='EstimatedSalary', hue='Exited', kde=True, element='step')

plt.title('Salary Distribution')

plt.show()



16.) projects\_by\_gender = df.groupby('Gender')['NumOfProducts'].sum().reset\_index()

plt.figure(figsize=(8, 6))

sns.barplot(x='Gender', y='NumOfProducts', data=projects\_by\_gender, palette='Set2')

plt.title('Total Number of Products by Gender', fontsize=16)

plt.xlabel('Gender', fontsize=12)

plt.ylabel('Total Number of Products', fontsize=12)

for index, row in projects\_by\_gender.iterrows():

plt.text(index, row['NumOfProducts'], int(row['NumOfProducts']),

ha='center', va='bottom', fontsize=12)

plt.tight\_layout()

plt.show()

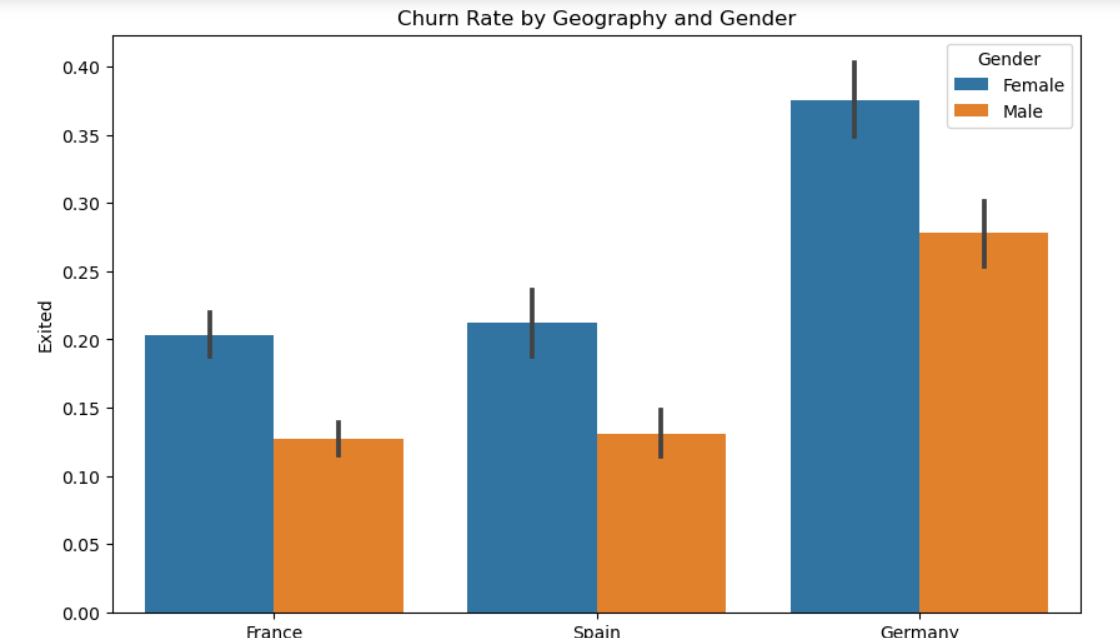


17.) plt.figure(figsize=(10, 6))

sns.barplot(x='Geography', y='Exited', hue='Gender', data=data)

plt.title('Churn Rate by Geography and Gender')

plt.show()



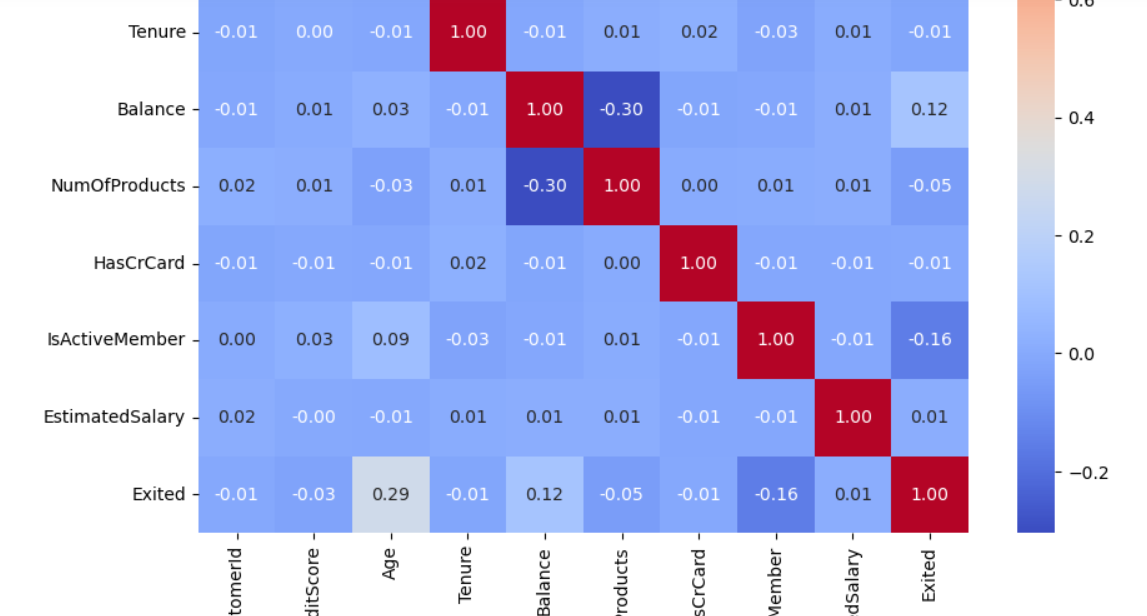
18.) plt.figure(figsize=(10, 8))

corr = data.corr(numeric\_only=True)

sns.heatmap(corr, annot=True, fmt=".2f", cmap='coolwarm', square=True)

plt.title('Correlation Heatmap')

plt.show()



**5. Conclusion**

KEY FINDINGS:

**1. Demographics Matter**

A thorough demographic analysis of the dataset revealed strong associations between age, gender, geography, and customer churn:

* **Age as a Risk Factor**: Churn probability increases significantly with age. Older customers, particularly those over 45, demonstrate a higher tendency to exit banking relationships. This trend may be linked to shifting financial goals, lower digital engagement, or changes in service expectations among older clientele.
* **Gender and Geography Intersection**: Female customers residing in **Germany** emerged as a **high-risk group**. This subset not only shows elevated churn rates but also exhibits specific behavioral patterns such as lower product usage and reduced activity levels. The pattern suggests cultural or systemic factors that may uniquely affect customer loyalty in this segment.
* **Implication**: Banks need to look beyond isolated demographic attributes and consider **intersectional profiles** (e.g., older females in specific regions) when designing retention strategies.

**2. Behavioral Factors Influence Churn**

The behavioral analysis revealed that **customer engagement** is a critical predictor of churn:

* **Inactive Members Churn More**: Customers marked as "inactive" (i.e., low usage of services or digital platforms) showed a **substantially higher likelihood of churning**. This inactivity might reflect dissatisfaction, lack of awareness of services, or a low perceived value of the relationship with the bank.
* **Engagement as a Protective Factor**: High-activity members — those using online platforms, conducting transactions regularly, or contacting customer service — displayed stronger retention. The **engagement-to-churn correlation** reinforces the value of ongoing interaction.
* **Implication**: Customer **inactivity should be treated as an early warning signal** for churn and trigger timely interventions.

**3. Product Engagement Reduces Churn**

Analysis of the number of products held by each customer revealed a strong inverse relationship with churn:

* **Optimal Product Range**: Customers using **2 to 3 banking products** (e.g., savings account, credit card, investment tools) had **significantly lower churn rates** compared to those with only one product.
* **Deeper Relationships Foster Loyalty**: These customers likely perceive greater value from the bank and are less inclined to switch due to the inconvenience and emotional investment tied to multiple services.
* **Implication**: Promoting **cross-selling and product bundling** can be a strategic lever to improve customer stickiness and lifetime value.

RECOMMENDATIONS:

**1. Targeted Retention Programs for High-Risk Demographics**

To reduce attrition among the most vulnerable groups, banks should:

* **Deploy predictive segmentation models** to identify high-risk profiles (e.g., elderly women in Germany with only one product and low activity).
* **Design customized outreach** such as wellness financial reviews, retirement planning advice, or personalized communications that resonate with older demographics.
* **Offer loyalty rewards** or account enhancements for long-standing customers who show signs of disengagement.
* **Cultural sensitivity**: Tailor communication and service models for specific geographies based on local expectations and preferences.

**2. Incentivize Product Bundling**

Encouraging multi-product ownership through well-designed bundling strategies can drive loyalty:

* **Create bundled offerings** that combine commonly used products (e.g., salary account + credit card + insurance) at reduced fees or with enhanced benefits.
* **Launch loyalty programs** that offer tiered rewards based on the number of products held or their usage frequency.
* **Educate customers** on the benefits of different services through personalized recommendations powered by data insights.
* **Track uptake** of bundling incentives and analyze their impact on churn over time.

**3. Enhance Engagement for Inactive Members**

To reduce churn driven by disengagement, banks should develop a multi-channel re-engagement plan:

* **Monitor customer activity patterns** to identify early signs of inactivity.
* **Launch automated but personalized nudges** (emails, SMS, app notifications) to encourage logins, product usage, or financial planning actions.
* **Incentivize interaction** via cashback offers, digital tool tutorials, or limited-time account perks for performing specific actions (e.g., activating a card, downloading the app).
* **Train relationship managers** to reach out with proactive check-ins or service offerings, especially for high-value dormant accounts.

6. **FUTURE SCOPE:**

The exploratory analysis conducted in this project uncovers valuable insights into customer churn behavior. However, there remains significant potential to expand the analysis and build upon the foundation established. The following areas outline the future scope of this project:

1.) Predictive Modeling and Machine Learning

While the current project focuses on Exploratory Data Analysis (EDA), a natural next step is to apply **supervised machine learning techniques** to build a **churn prediction model**. This would involve:

* **Feature Engineering**: Creating new variables that may capture customer behavior more effectively (e.g., tenure-to-age ratio, income-to-credit ratio).
* **Model Selection**: Comparing algorithms like Logistic Regression, Random Forests, Gradient Boosting Machines (GBM), and XGBoost for predictive performance.
* **Model Evaluation**: Using metrics such as **accuracy**, **precision**, **recall**, **F1-score**, and **ROC-AUC** to evaluate model effectiveness.

2.) Customer Segmentation Using Clustering

Applying **unsupervised learning methods**, such as K-Means or Hierarchical Clustering, can help identify **distinct customer personas**. These segments can be used to:

* Personalize marketing and retention strategies.
* Develop targeted promotions for each group based on risk and value.
* Optimize product bundling based on segment preferences.

3.) Time-Series and Trend Analysis

* If longitudinal or time-stamped data becomes available, future work can involve:
* T**racking churn trends over time** to detect seasonal patterns or the impact of policy changes.
* **Analyzing customer lifecycle** stages to identify when churn risk is highest.

**4. Deep Dive into Geography-Based Insights**

The dataset shows significant differences in churn rates across geographic regions. Future work could:

* Integrate **external socioeconomic datasets** (e.g., GDP, unemployment, interest rates) by region to contextualize customer behavior.
* Explore regulatory, cultural, or competitive landscape differences that might affect churn behavior in Germany, France, and Spain.

**5. Customer Sentiment and Feedback Integration**

Combining churn data with **customer feedback (surveys, reviews, support interactions)** can give qualitative insight into:

* Reasons behind dissatisfaction or loyalty.
* Service experience pain points not reflected in numerical data.

**6. Real-Time Churn Monitoring System**

Develop a **dashboard or real-time alert system** to help the bank:

* Identify high-risk customers early.
* Automate personalized offers or interventions using churn probability scores.
* Continuously monitor churn trends and KPIs.

**7. Business Impact Analysis**

Translate analytical insights into **financial impact assessments**, including:

* Estimating revenue saved per customer retained.
* Cost-benefit analysis of retention programs.

Forecasting long-term value based on reduced churn rates.

**8. Cross-Platform Behavior Analysis**

In future iterations, integrating **multi-channel engagement data** (e.g., mobile app usage, website logins, call center interactions) could give a fuller picture of engagement patterns and their correlation with churn.

**7. References**

i.) <https://app.mavenanalytics.io/datasets?search=Bank+Customer+Churn>

ii.) <https://github.com/>

iii.) Google