E-Commerce Bank Customer Churn Analysis

Aditya Anil Pawar

Department of Computer Science

University of Central Florida

April 30, 2024

1. Introduction: -

The Customer Churn, also known as customer attrition, customer turnover or customer defection is a degree of customer inactivity or disengagement that is observed by the bank over a certain amount of time.

Banks, insurance companies, streaming services companies and telecom service companies often use customer churn analysis and customer churn rates as one of their key business metrics because the cost of retaining existing customers is far less than acquiring a new one.

For instance, we have a client 'ABC Bank'. The bank has noticed an increased number of customers leaving the bank. To tackle this alarming situation, the bank has decided to collect the data of the past 6 months from the year 2016. 10,000 customers are selected randomly among three countries - France, Germany and Spain. The bank wanted to understand and get insights about customer churning so that the bank can upgrade or adapt new policies.

This analysis focuses on the behavior of bank customers who are more likely to leave the bank (i.e. close their bank account). I want to find out the most striking behaviors of customers through the Exploratory Data Analysis.

2. Data: -

The dataset provided via the Kaggle link is specifically curated for the analysis and prediction of customer churn within the e-commerce industry. It encompasses a comprehensive set of 5,691 records and 20 distinct attributes related to customer demographics, transactional history, and engagement metrics.

Key attributes include: -

- **CustomerId:** A unique identifier assigned to each customer. It is used to track individual customers without revealing personal information like their name.
- **Surname:** The family name or last name of the customer. It can be used to personalize communication, although it should be handled with care due to privacy considerations.
- **CreditScore:** A numerical expression based on a level analysis of a person's credit files, representing the creditworthiness of an individual. Higher scores typically indicate a better risk profile.
- **Country:** The country of residence of the customer. This can impact many aspects of a financial profile, including creditworthiness and product suitability.
- **Gender:** The gender of the customer. While not directly related to financial services, this demographic information can be useful for market segmentation and analysis.
- **Age:** The age of the customer. This is a crucial factor in financial modeling as it often correlates with income, financial needs, and risk tolerance.
- **Tenure:** In a financial context, this often refers to the number of years a customer has been with the bank or financial institution. It can be an indicator of customer loyalty and stability.
- **Balance:** The amount of money a customer has in their accounts with the financial institution. This can be a significant indicator of the customer's financial health and profitability for the bank.
- **NumOfProducts:** The number of products the customer has with the bank, such as checking accounts, savings accounts, loans, credit cards, etc. More products can indicate a deeper relationship with the bank.
- **HasCrCard:** A binary indicator (usually 1 for Yes, 0 for No) that specifies whether the customer has a credit card issued by the financial institution.
- **IsActiveMember:** Another binary indicator reflecting whether the customer is considered active, potentially based on recent transactions, logins to online banking, or other engagement metrics.
- **EstimatedSalary:** An estimate of the customer's annual income. This can be used to gauge their financial capability and to tailor financial products to their needs.
- **Exited:** Typically a binary indicator showing whether the customer has left or closed all accounts with the bank. This is the target variable for predicting churn in the financial services industry.

This dataset provides a rich blend of quantitative and categorical data, suitable for conducting detailed statistical analysis and building predictive models to understand and forecast customer churn. The broad range of features included allows for an exploration of complex relationships and patterns, facilitating a deep dive into the factors influencing customer loyalty and retention in e-commerce.

3. Analysis

3.1 Gender

3.1.1 What is the gender percentage of all customers who left the bank?

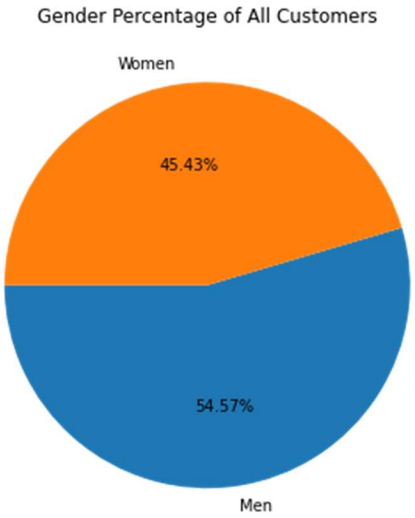

Figure 1: Gender percentage of all customers

The blue section of the pie chart represents 45.43% of the customers who left the bank, which could indicate the percentage of males if we assume the orange section represents females. Conversely, the orange section, which represents 54.57%, could indicate the percentage of females.

3.2 Estimated Salary with Age

3.2.1 What patterns can be observed in the estimated salary distribution across different age groups as shown in the bar chart, and how could these patterns potentially be explained by factors such as work experience, age-related workforce participation, or retirement trends?

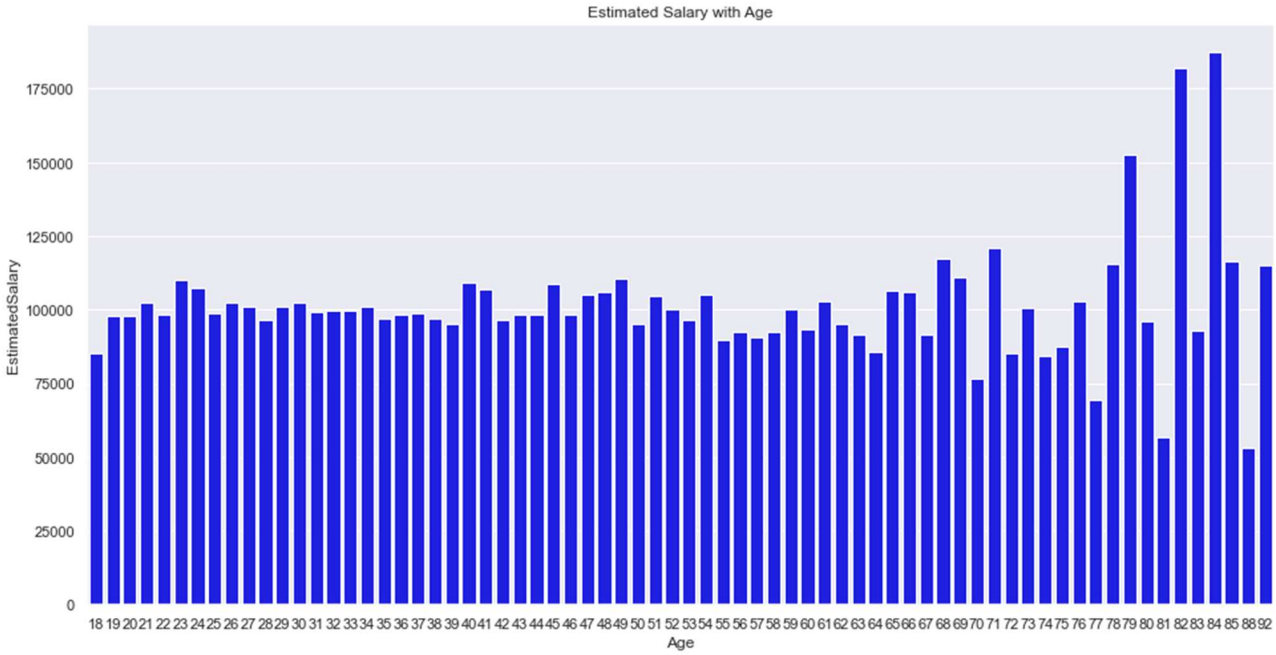

Figure 2: Estimated Salary with Age

From the bar plot, we see that after the age of 81, there is a rise in salary which goes above 175000. That may be because customers are generating income from blooming sources like businesses and investments.

3.3 Geography and Gender Distribution: -

3.3.1 How do the distributions of estimated salaries compare between males and females within France, Spain, and Germany, and what could be the underlying causes of any observed disparities in these distributions?

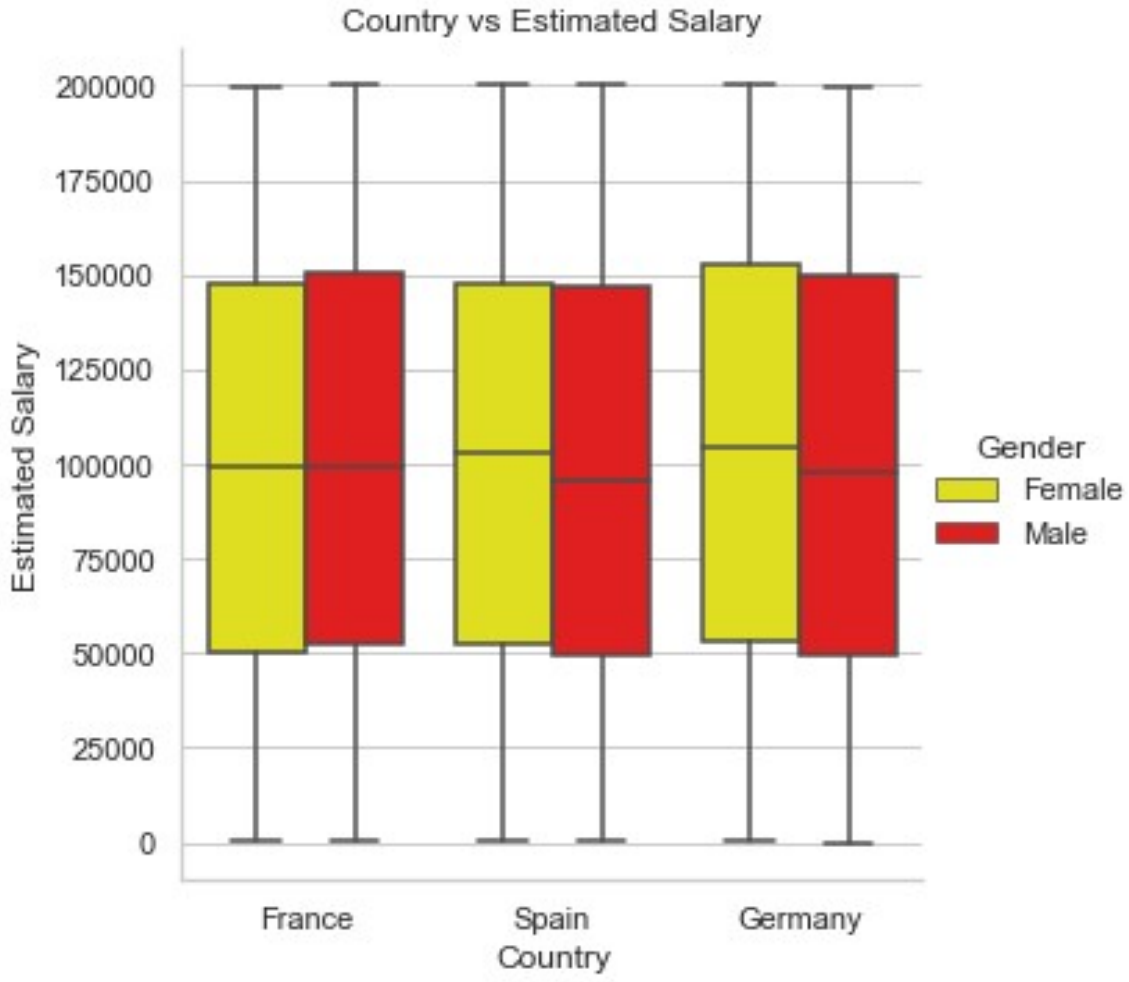

Figure 3: Country vs Estimated Salary

Based on the box plot, both men and women in France, Spain, and Germany seem to earn similar amounts of money, as shown by their median salaries. The ranges of salaries (from the lower to the upper end) also largely overlap between the genders. This means that there might not be a big difference in how much men and women earn in these countries. Any small differences in salary could be because of the types of jobs people have, their work experience, or whether they work part-time or full-time. Cultural factors and workplace equality policies in these countries might also affect the salary differences between men and women.

3.4 Age: -

3.4.1 What trends can be observed in the balance distribution across different age groups for France, Spain, and Germany?

Figure 4: Age Wise Balance Distribution

In this case, we made a scatter plot to look at how the customers' age relates to their bank balance. After creating the plot, we found that there isn't a clear relationship between a customer's age and their balance because the points are spread all over the graph without any clear pattern.

We noticed something interesting: both France and Germany have many accounts with almost no money in them. This is surprising because Germany has the highest number of customers who have stopped using the bank. This finding challenges the idea that customers with no money in their accounts are more likely to leave the bank.

3.5 Country Wise Balance Distribution

3.5.1 How does the distribution of customers' balances vary between France, Germany, and Spain, and what might these variations suggest about the financial behaviors or economic conditions in these three countries?

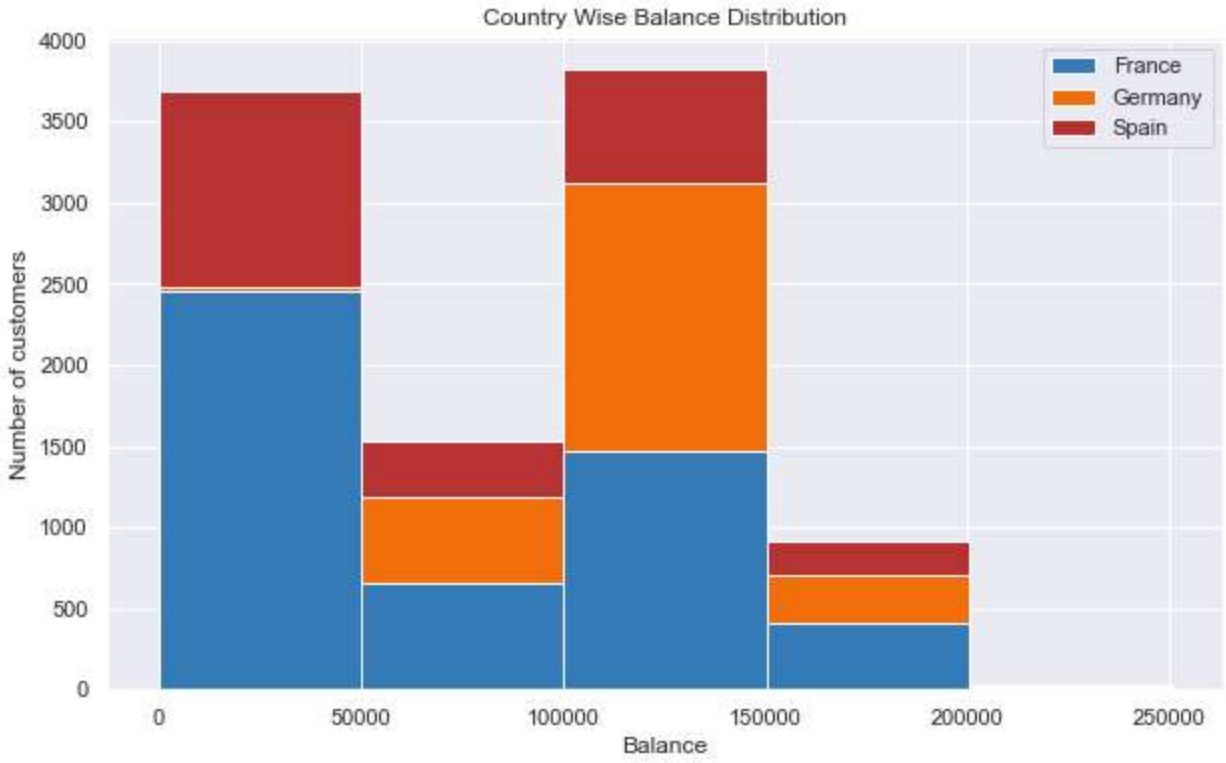

Figure 5: Country Wise Balance Distribution

The plot above shows the Balance of customers of different countries.

In the first bar you notice, it contains the customers who have a balance between (0 - 50000). France has the highest number of customers and then comes Spain and Germany has very less customers in this segment.

In the third bar, you notice Germany (100000 - 150000) has the maximum number of customers.

Based on the above plot, we can conclude that the accounts from Germany are very rich as compared to the other two countries. But if we see Plot 1, then we'll understand the maximum number of customers leaving the bank are from Germany.

3.6 Inferring who left the bank by taking age under consideration: -

3.6.1 What is the relationship between age and the likelihood of exiting as depicted in the histogram, and what might be the reasons for any significant differences in exit rates across different age groups?

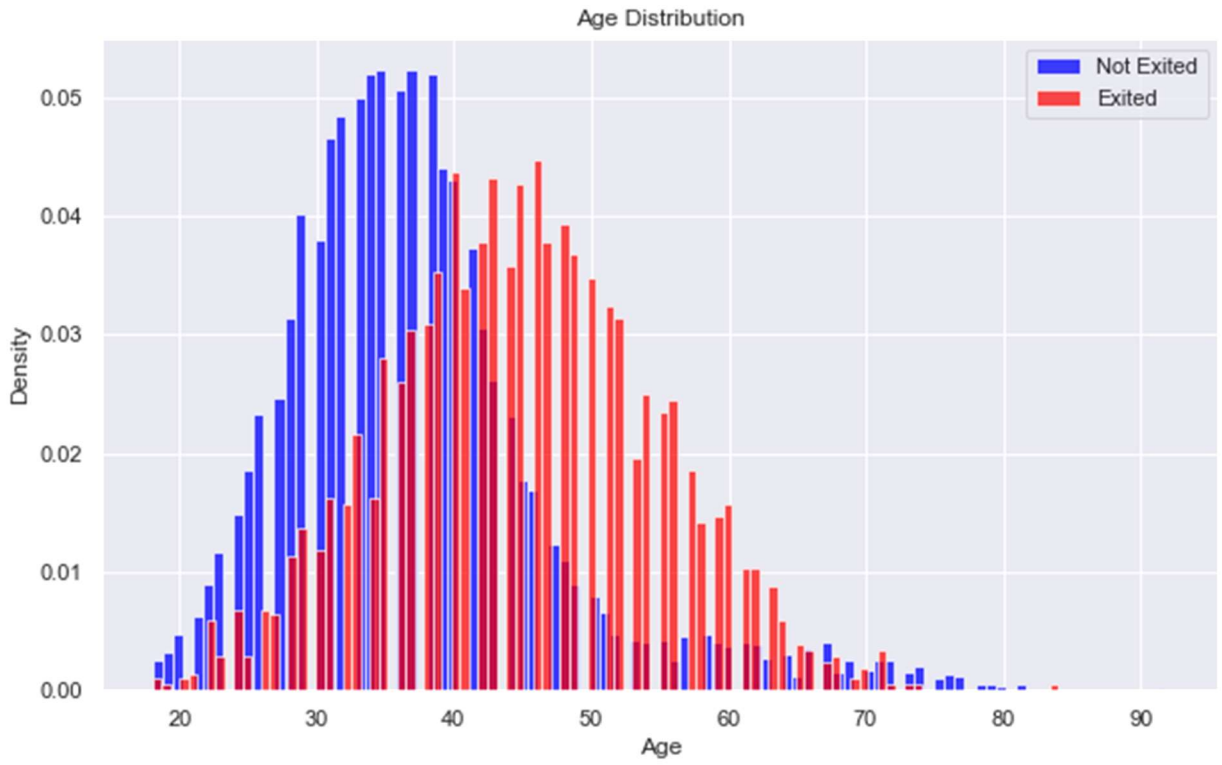

Figure 6: Age Distribution

The above histogram shows the age distribution according to the customer who stayed and left the bank. Here, we observe a very interesting trend:

Young adult customers (customers from 20-40) have not left the bank but customers after the age of 40 have mostly left the bank. The reason for this may be that other competitor banks offer better incentives (as the older customers have much experience with the bank and its services than the younger ones).

3.7 Inferring who left the bank by taking balance under consideration.

3.7.1 How does the balance distribution density differ between individuals who have exited and those who have not, and what financial behaviors or policy implications might be inferred from the observed distribution patterns?

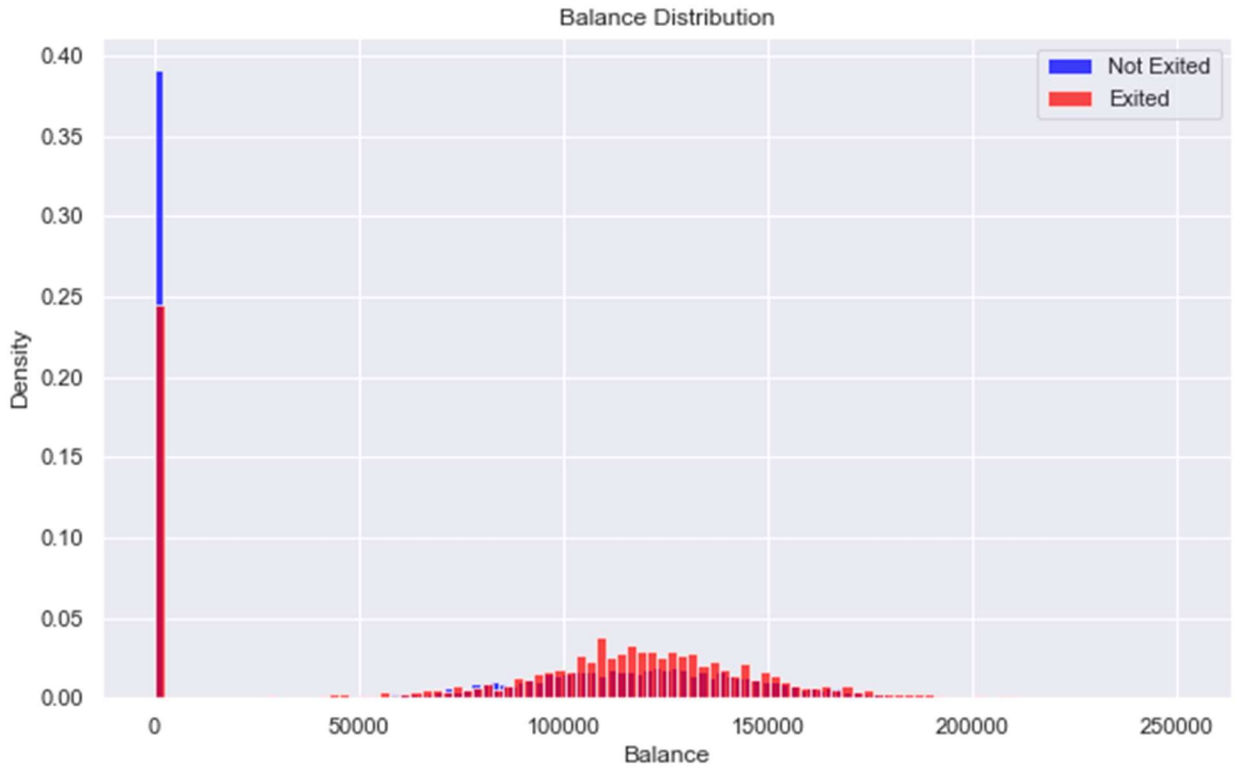

Figure 7: Balance Distribution

From the above visualization, we can see that the distribution of balance for the customers who left is normal. Also, for the customers with 0 bank balance there is a hike in the density- approximately 24% for the customers who left and 39% for the customers who stayed.

3.8 Inferring who left the bank by taking Credit Score of customer under consideration: -

3.8.1 How does the credit score density distribution vary between customers who have exited and those who have not, and what insights might this provide about the role of credit scores in customer retention strategies?

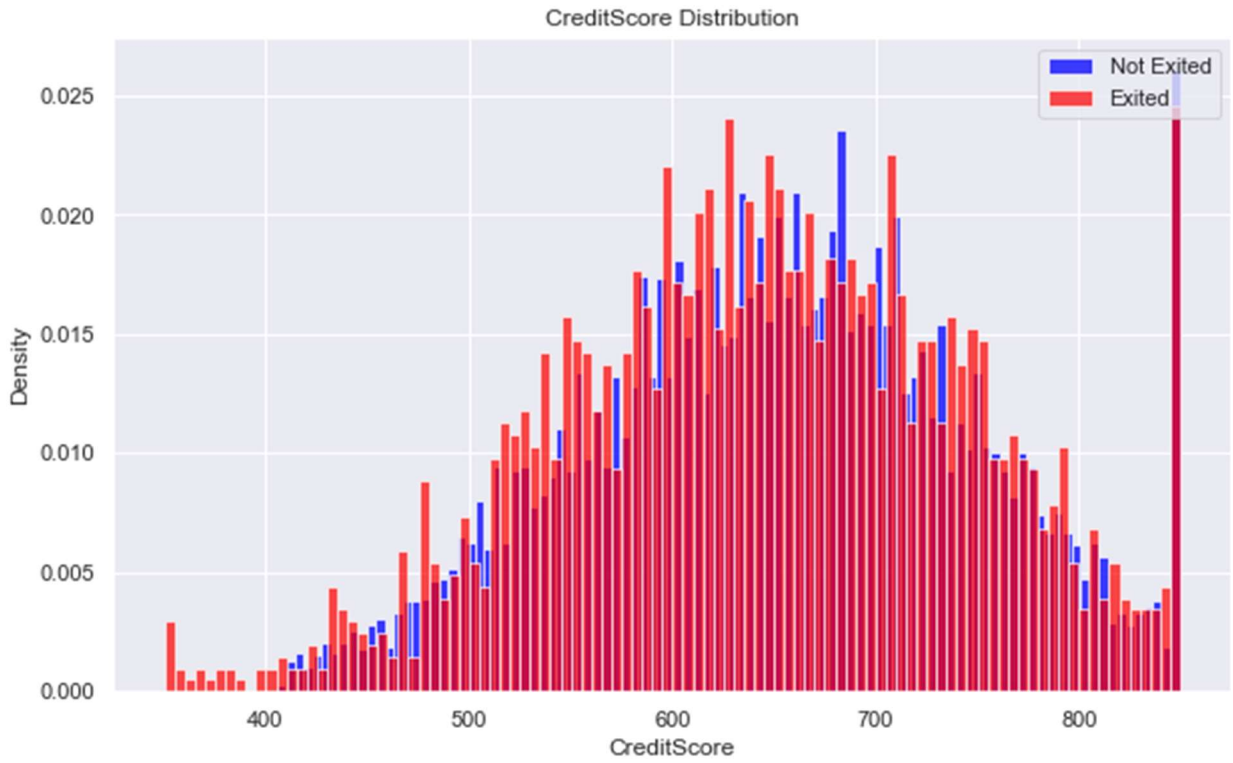

Figure 8: Credit Score Distribution

Here, we can observe that more customers with a high credit score are leaving the bank than the customers with a low credit score. One reason behind this may be that ABC Bank was not offering good services or premium services to their trustworthy candidates.

4. Asking and Answering Questions:-

Q.1) How does customer retention vary among France, Spain, and Germany as represented in the bar chart, and what factors might contribute to the differences in customer exit rates across these countries?

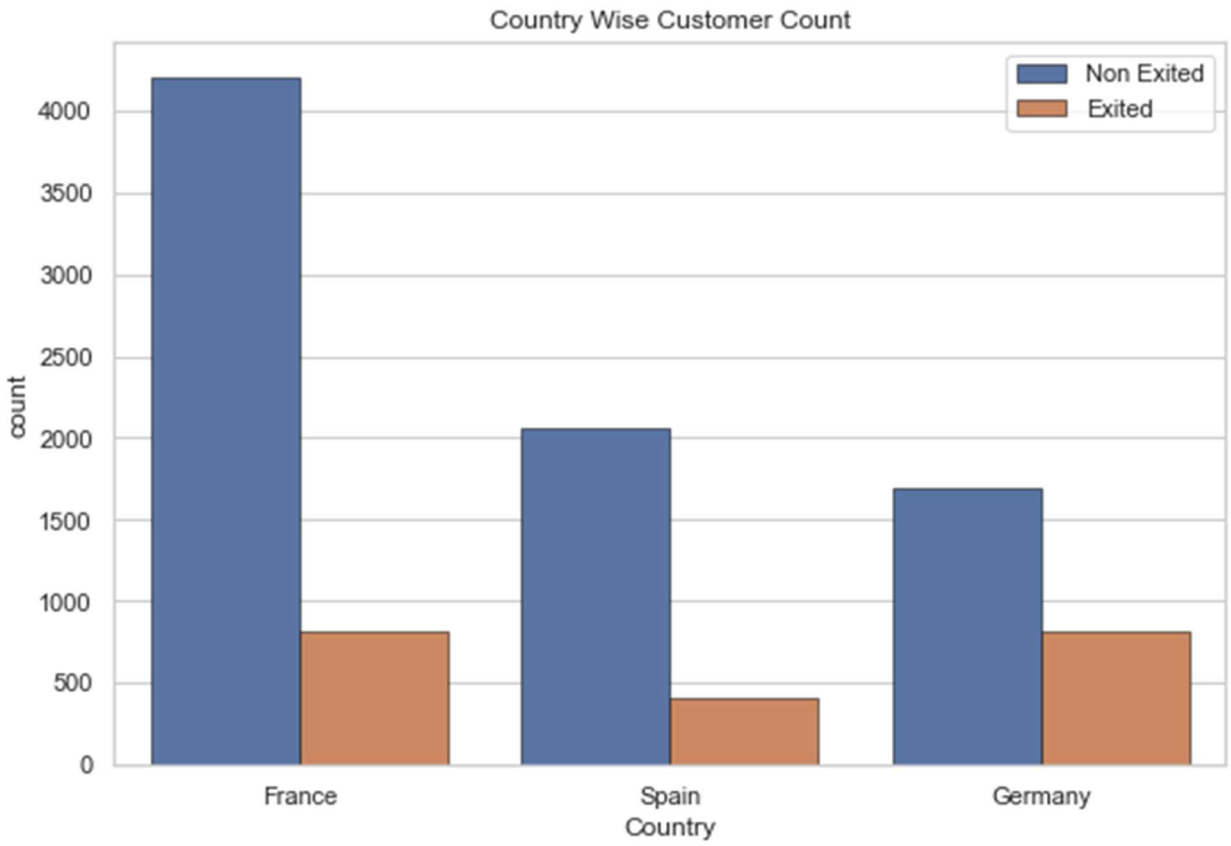

Figure 1: Country Wise Customer Count

The bar chart indicates that France has a significantly higher overall customer count, with more customers both exiting and staying compared to Spain and Germany. These observed differences suggest that country-specific factors, such as market size, customer satisfaction, economic conditions, competition, and perhaps service quality, may influence customer retention and exit rates differently across these countries.

Q.2) What factors could explain the variance in active versus not active bank memberships among France, Spain, and Germany as illustrated in the bar chart?

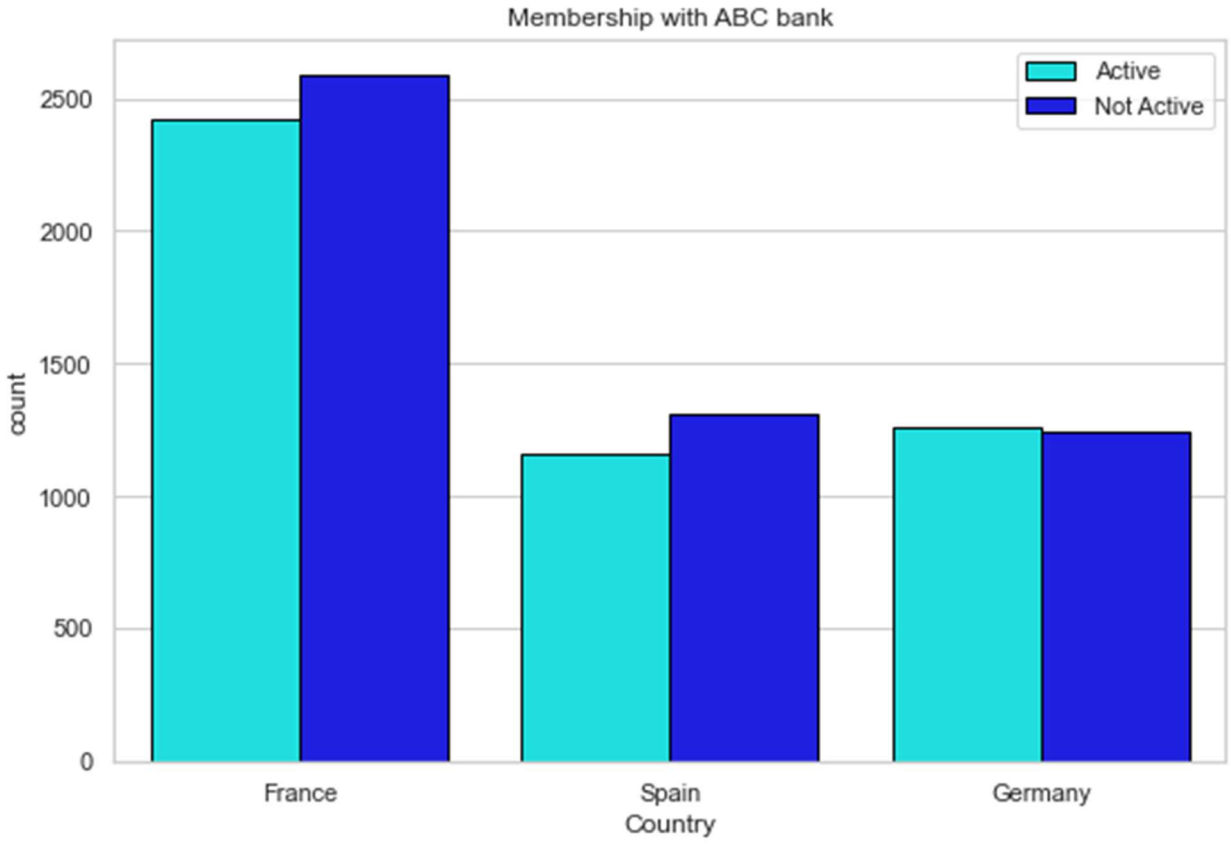

Figure 2: Membership with ABC bank

The bar chart shows that France has the highest number of both active and not active memberships with ABC bank, while Spain and Germany have a lower and nearly equal distribution of membership statuses. These variations might be influenced by country-specific marketing effectiveness, economic conditions, customer service quality, or the competitiveness of the banking sector within each country. From the count plot, we can see that there are approx. 2400 people who are active members from France, 1200 from Spain and 1300 from Germany. So, the total becomes 2400 + 1300 + 1400 = 5100 which is approximately correct as compared to the above estimated number.

Q.3) How many customers have a credit card?

Gender

Female 3192

Male 3863

From the above output, we can infer that there are a total of 7055 customers from the three countries(France, Spain and Germany) who have a credit card.

Q.3) A) What could be the reasons for the observed gender distribution in credit card ownership, as indicated by the pie chart showing 54.8% females and 45.2% males with credit cards?

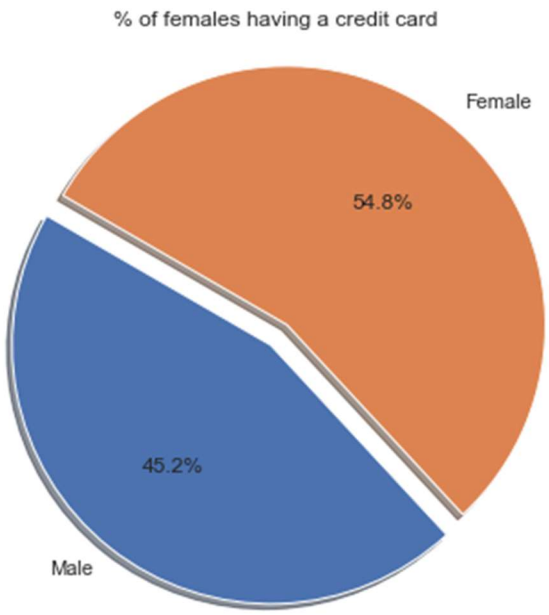

Figure 3: % of females having a credit card

The pie chart suggests that a slightly higher percentage of females than males have a credit card. This could indicate differences in financial behavior or access to credit facilities between genders, or it might reflect targeted marketing strategies of credit card companies that resonate more with female consumers.

Q.4) What factors may contribute to the significant difference in the number of candidates who chose to exit from the bank compared to those who remained, as shown in the bar chart?

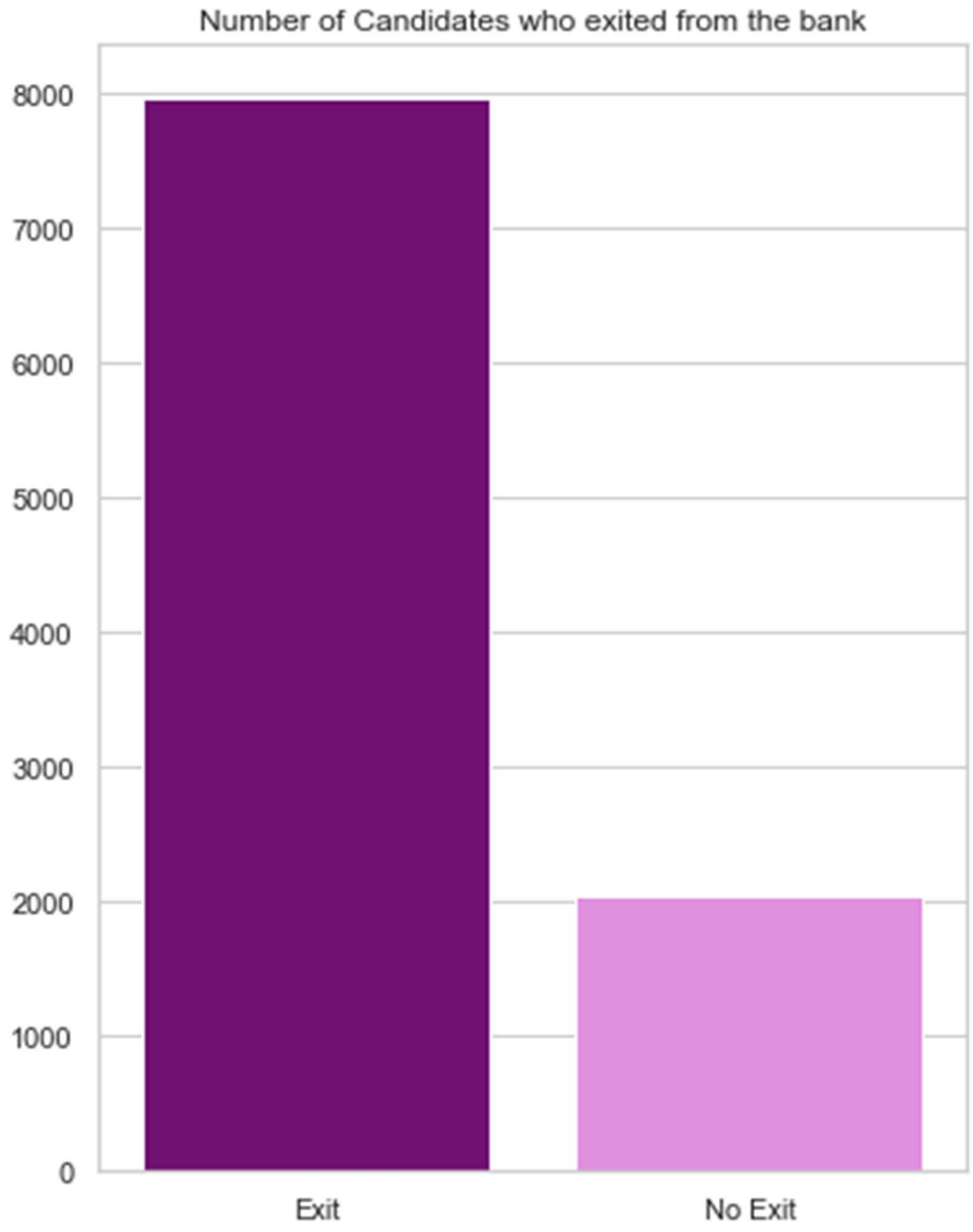

Figure 4: Number of candidates who exited from the bank

The bar chart indicates a substantially higher number of candidates who exited the bank compared to those who did not. This discrepancy might point to issues such as customer dissatisfaction, competitive financial products from other institutions, or changes in the bank's policies that may not align with the needs or preferences of a large segment of its clientele.

Q.5) What factors could explain the variation in the number of products across France, Germany, and Spain as depicted in the bar chart?

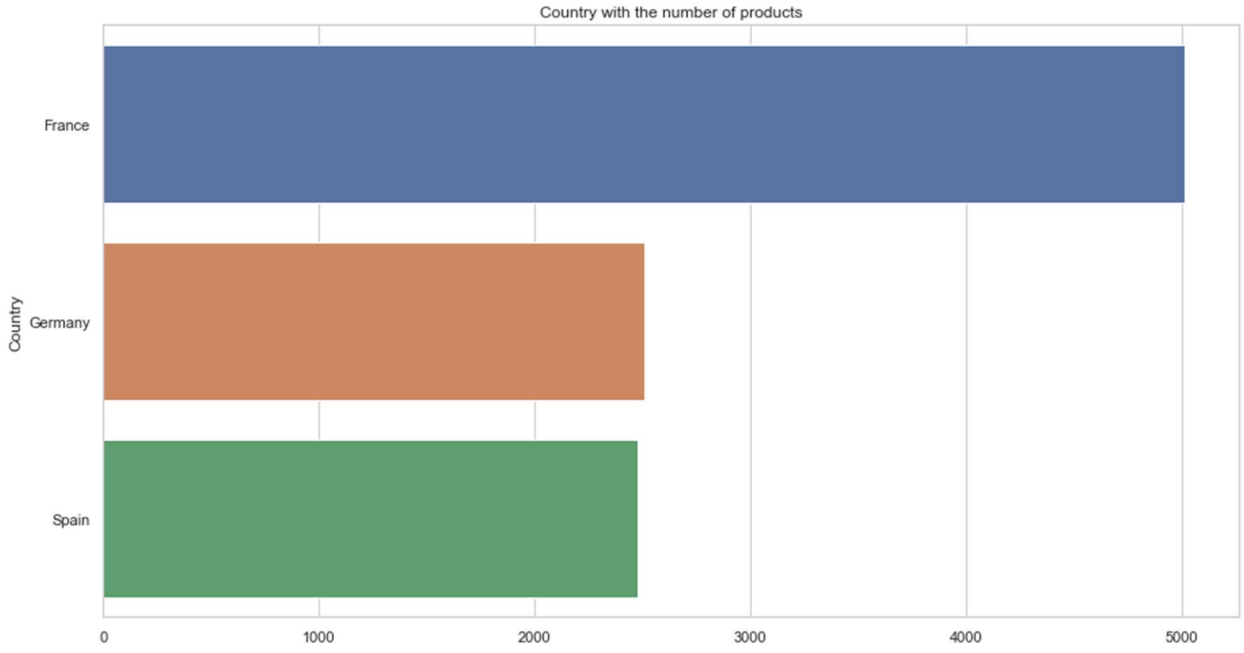

Figure 4: Country with the number of products

The bar chart reveals that France has a significantly higher number of products compared to Germany and Spain. This discrepancy might be due to France's larger market size, a more diverse product offering, higher consumer demand, or more aggressive marketing strategies employed by the bank in France.

Q.6) What can be inferred about the distribution of tenure periods from the data shown in the bar chart, particularly regarding the apparent increase in frequency at the 10-year mark?

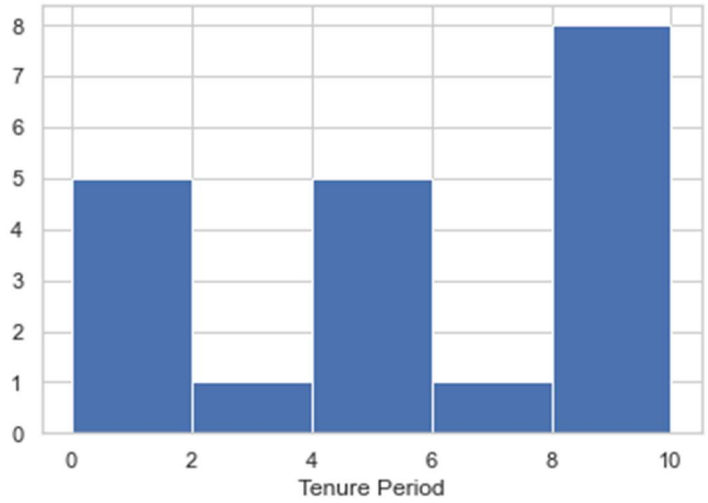

Figure 5: Tenure Period

The bar chart shows that the frequency for the 10-year tenure period is significantly higher than for the other periods displayed. This could suggest that there is a higher retention or occurrence rate for individuals or elements at this specific tenure milestone, which may indicate either a point of increased loyalty or a common duration after which a specific event occurs, such as contract renewal or eligibility for benefits.

- **Inferences and Conclusion: -**

We gathered many interesting facts about the customer churning of Bank ABC from this analysis. Below is the summary of few of them:-

- From the bar plot of Estimated Salary vs Age, we came across a very interesting thing. Generally, it is observed that as age increases, a person's salary increases. To check this assumption, we had plotted the estimated salary with age. If we see the plot, we can observe that our assumption is wrong. But we can see in the bar plot that, at the age of 81 and 88, customer is having a very low salary as compared to the other age groups. So, from this, we can infer that age does not play a major role in the salary of a customer.
- We can infer that Germany has multiple banks which offer better incentives to the customers as compared to the ABC Bank. That's why a few customers are moving out and opting for other banks. We don't have data of other banks so it's better that we speculate.
- From the visualization of Geography and Gender distribution, we can infer that the male customer estimated salary is higher than females in France and Spain, however in Germany female customer's estimated salary is higher than the male counterpart.
- From the histogram visualization of Age, we have inferred that the age of customers is inversely proportional to the customers exiting the bank (More younger customers (20-40) stayed in the bank and elder ones (above 40) left the bank).
- Maximum number of customers from 'ABC Bank' are churned (or have exited) from Germany.
- There are a smaller number of customers who have left the bank.
- France has a relatively high number of products as compared to Germany and Spain.

References:-

1. https://www.kaggle.com/datasets/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction/data
2. Ecommerce customer churn - rate prediction and prevention - https://www.klaviyo.com/uk/blog/customer-churn-rate-prediction
3. Customer Churn in Retail E-Commerce Business: Spatial and Machine Learning Approach - https://www.mdpi.com/0718-1876/17/1/9
4. Ecommerce Churn Rate: How To Calculate and Reduce Churn - https://www.shopify.com/blog/churn-rate-in-ecommerce