**Learners have to come up with a Report to support the answers to the following questions and suggestions**

**Objective Questions:**

1. What is the distribution of account balances across different regions?

: The bar chart titled “Sum of Balance by Geography Location” provides a comparative view of the sum of balances across three different countries: France, Germany, and Spain.

* **France** leads with the highest sum of balance, approximately at 0.31 billion.
* **Germany** follows closely with a sum of balance around 0.30 billion.
* **Spain**, however, has a significantly lower sum of balance, roughly at 0.15 billion.

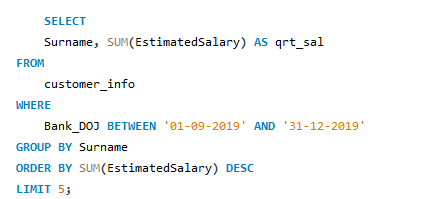
This data suggests that there are substantial differences in the sum of balances across these countries. France and Germany have nearly double the sum of balance compared to Spain. This could indicate varying economic conditions, financial behaviors, or banking policies in these countries. Further analysis would be required to understand the underlying reasons for these differences.

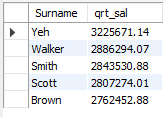
This chart is a valuable tool for visualizing and understanding the financial landscape across these three countries. It could serve as a basis for more in-depth economic or financial studies.

A graph with purple bars

Description automatically generated

1. Identify the top 5 customers with the highest Estimated Salary in the last quarter of the year. (SQL)



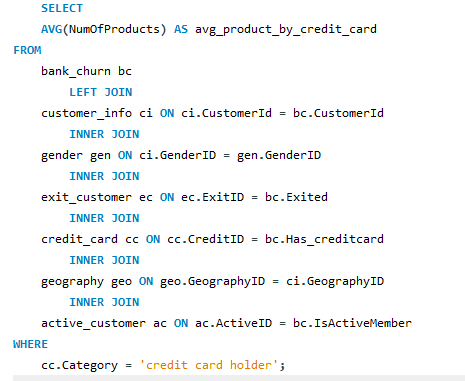


1. Calculate the average number of products used by customers who have a credit card. (SQL)

In our database, we have multiple tables that store information about our bank’s customers, their product usage, and whether they hold a credit card. We wanted to understand the average number of products used by customers who have a credit card.

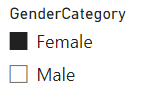
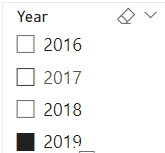
To achieve this, we wrote an SQL query that joins relevant tables and calculates the average number of products (NumOfProducts) used by customers who are credit card holders (cc.Category = 'credit card holder').

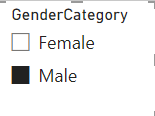
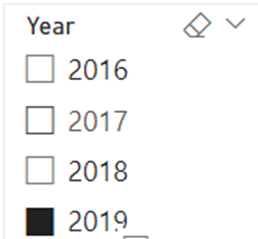
The SQL query returned an average of **1.5224**. This implies that, on average, customers who have a credit card use approximately 1 to 2 products, with a slight inclination towards 2. This insight can help us understand the product usage pattern among our credit card holding customers and can guide us in tailoring our product offerings and marketing strategies."





1. Determine the churn rate by gender for the most recent year in the dataset.

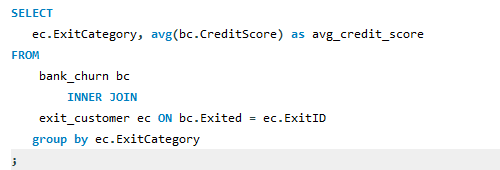




1. Compare the average credit score of customers who have exited and those who remain. (SQL)

The SQL query provided compares the average credit score of customers who have exited and those who remain. The query joins the bank\_churn and exit\_customer tables on the Exited and ExitID fields respectively. It then groups the results by the ExitCategory and calculates the average CreditScore for each group.

This indicates that the average credit score of customers who have exited is 645.3515, while the average credit score of customers who remain is 651.8532. Therefore, on average, customers who remain have a slightly higher credit score compared to those who have exited.





1. Which gender has a higher average estimated salary, and how does it relate to the number of active accounts? (SQL)

The query first creates a temporary table ActiveAccounts that counts the number of active accounts per customer. An account is considered active if the IsActiveMember field is 1.

Then, the main part of the query joins this ActiveAccounts table with the customer\_info table on the CustomerId field. It groups the results by gender and calculates two metrics for each group:

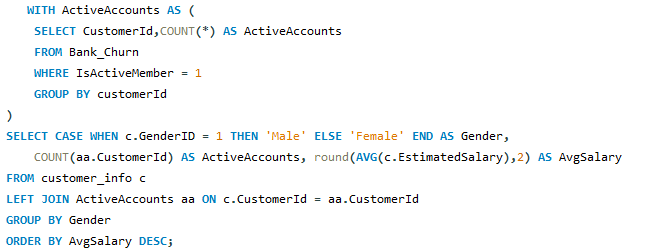
1. The number of active accounts (COUNT(aa.CustomerId))
2. The average estimated salary (AVG(c.EstimatedSalary))

The results are then ordered by the average estimated salary in descending order.

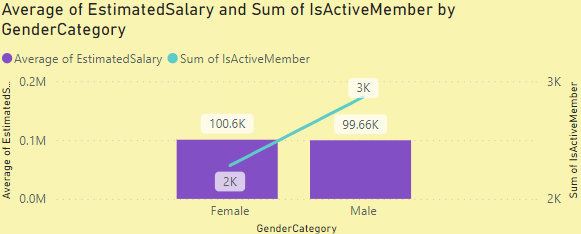
The output of your query shows that:

* Females, with **2284** active accounts, have an average estimated salary of **$100,601.54**.
* Males, with **2867** active accounts, have an average estimated salary of **$99,664.58**.

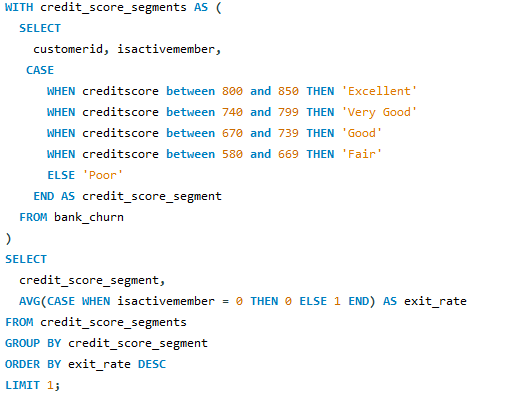
From these results, we can infer that while males have more active accounts, females have a slightly higher average estimated salary. This suggests that there may not be a direct correlation between the number of active accounts and the average estimated salary for each gender







1. Segment the customers based on their credit score and identify the segment with the highest exit rate. (SQL)



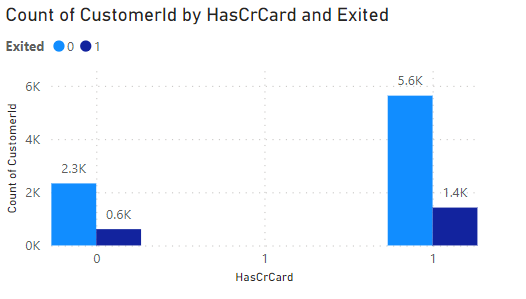


1. Find out which geographic region has the highest number of active customers with a tenure greater than 5 years. (SQL)





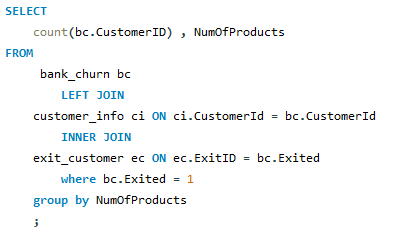
1. What is the impact of having a credit card on customer churn, based on the available data?

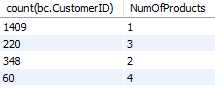


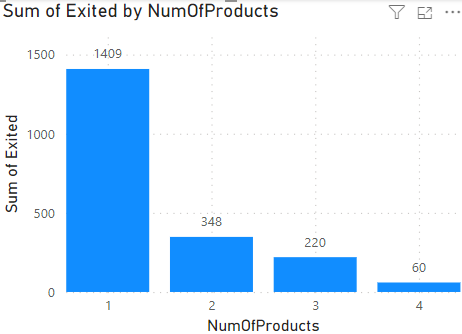
* Customers who do not have a credit card (HasCrCard = 0) show a lower count of churn (Exited = 1) compared to those who have stayed (Exited = 0).
* Similarly, customers who have a credit card (HasCrCard = 1) also show a lower count of churn compared to those who have stayed.
* However, in absolute numbers, the group of customers with a credit card has both a higher count of customers who have stayed and a higher count of customers who have churned.

“Based on the available data, it can be observed that while customer churn occurs regardless of whether a customer has a credit card or not, the group of customers with a credit card shows a larger absolute number of both retained and churned customers.

1. For customers who have exited, what is the most common number of products they have used?





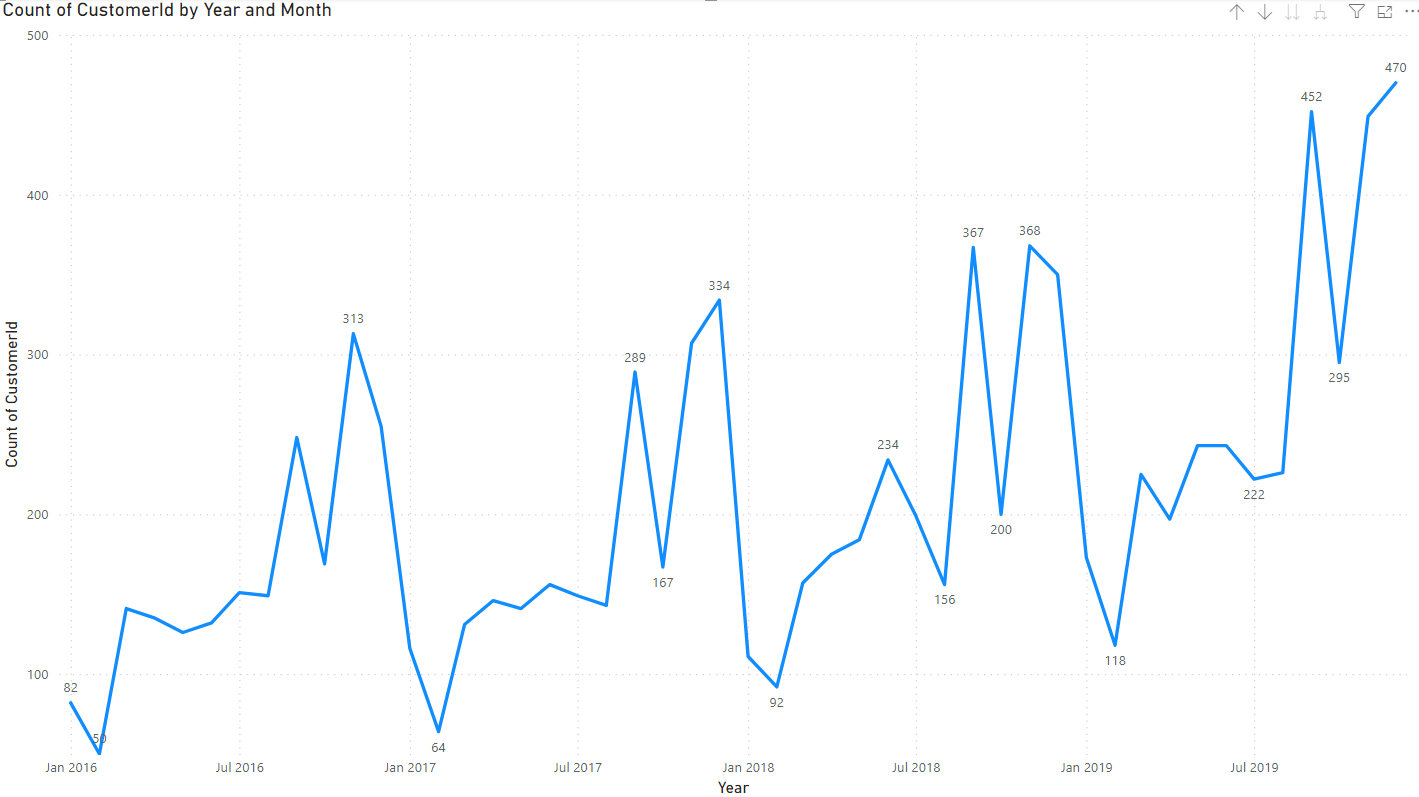


* **Title of the Graph**: The graph is titled “Sum of Exited by NumOfProducts”, which suggests it’s showing the total number of exits (possibly customers or products) for each number of products.
* **Number of Products**: The graph shows data for 1 to 4 products. There is no data for 0 products.
* **Sum of Exits**: The sum of exits decreases as the number of products increases. Specifically, the sum of exits is highest for 1 product (approximately 1409), and lowest for 4 products (around 60).
* **Trend**: There is a clear downward trend in the sum of exits as the number of products increases. This could imply that having more products leads to fewer exits, which might be a point of interest for further analysis or discussion.

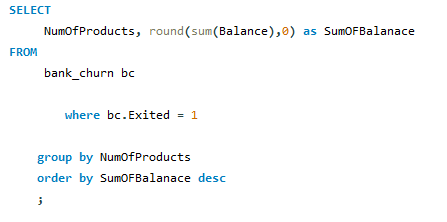
1. Examine the trend of customers joining over time and identify any seasonal patterns (yearly or monthly). Prepare the data through SQL and then visualize it.



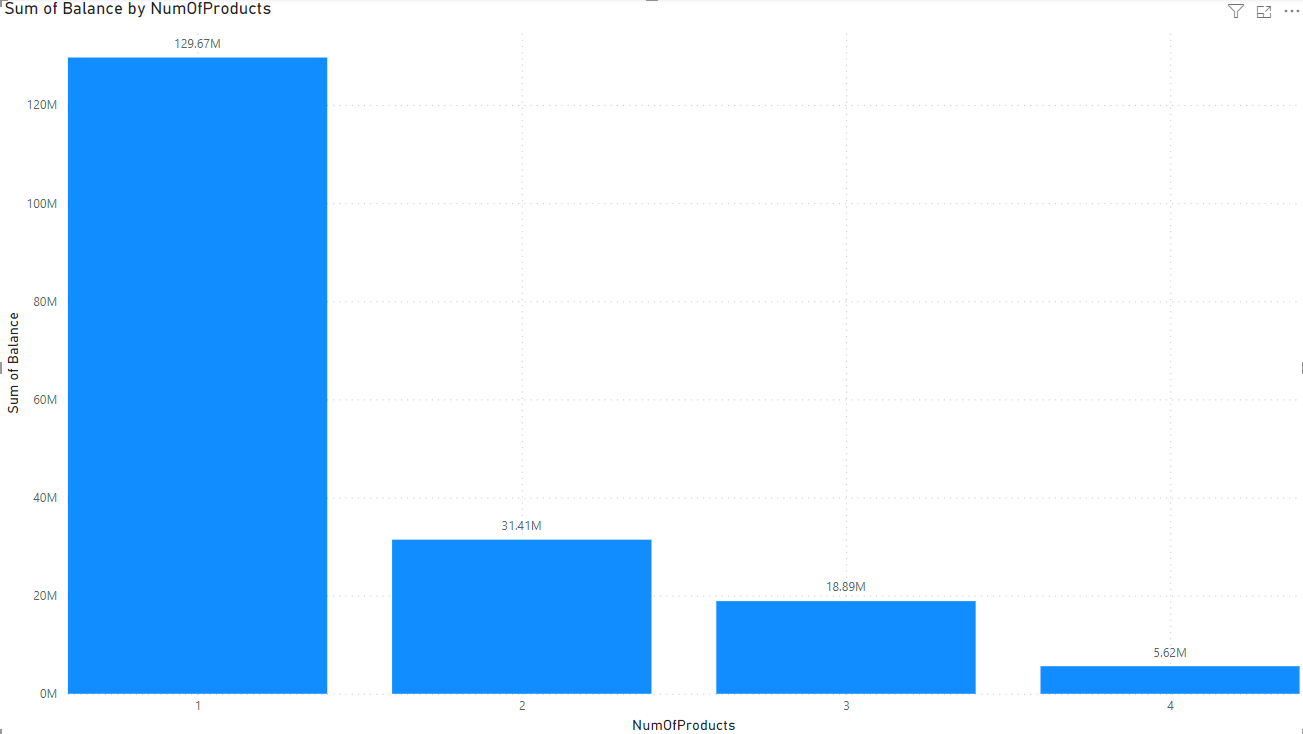




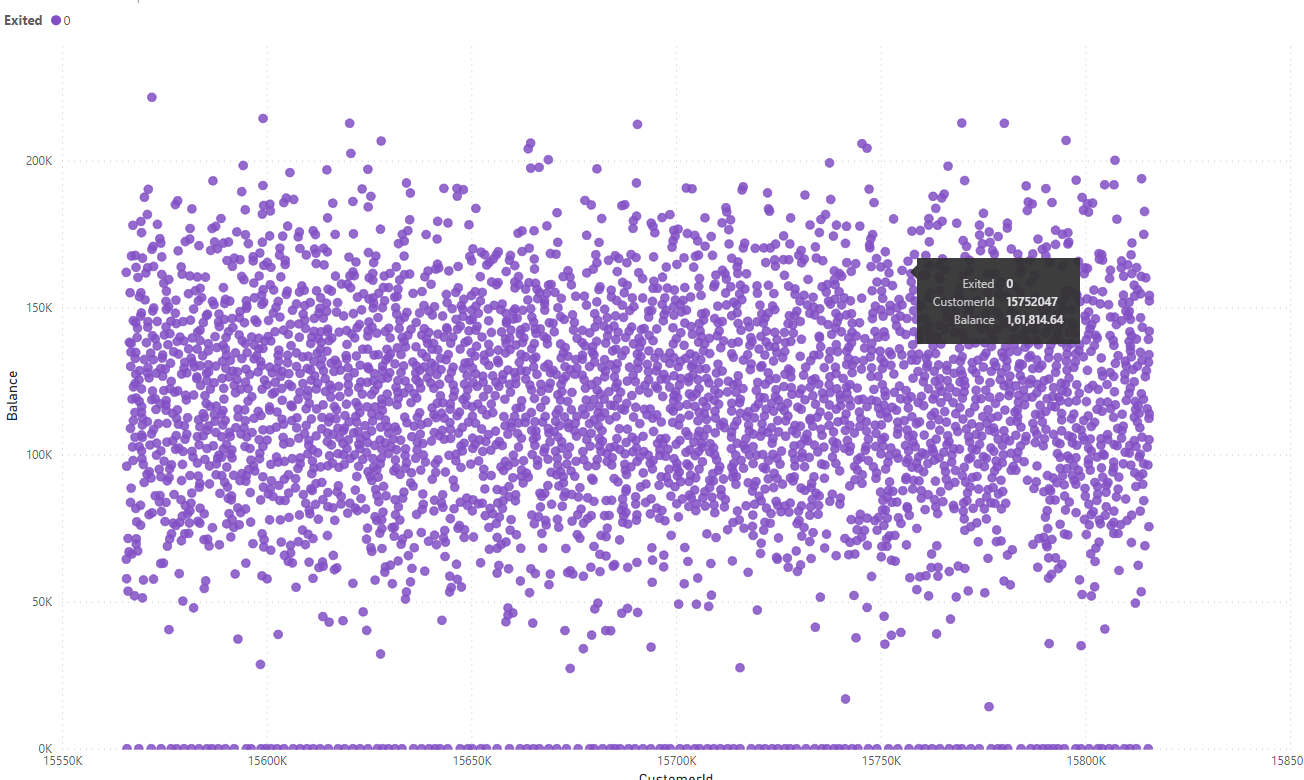
1. **Time Period**: The graph represents data from **January 2016 to July 2019**.
2. **Customer Count Fluctuations**: The count of CustomerId shows several fluctuations over this period, indicating varying levels of customer engagement or business growth.
3. **Notable Peaks**: There are notable peaks in **April 2016** with approximately 518 counts and in **July 2019** nearing the count of approximately 425.
4. **Significant Dip**: There is a significant dip near **July 2017** at about 267 counts.
5. **Overall Trend**: Despite the fluctuations, there seems to be an overall upward trend in the count of CustomerId over the years.
6. Analyze the relationship between the number of products and the account balance for customers who have exited.







1. **Number of Products**: The graph represents data for 1 to 4 products.
2. **Balance for One Product**: Customers with only one product have a substantially higher total balance, approximately 125M.
3. **Balance for Two Products**: The balance drops significantly for customers with two products, around 31.4M.
4. **Balance for Three and Four Products**: The balances for three and four products are much lower, approximately 3.8M and just above zero respectively.
5. **Overall Observation**: The graph indicates that as the number of products increases, the sum of the balance decreases.
6. Identify any potential outliers in terms of balance among customers who have remained with the bank.



A scatter plot is made by plotting Customer ID in X Axis, Balance in Y axis and Exited in Legend also Filter is done on Exited in ‘0’ that means those who are still retained.

After seeing the scatter plot it can be seen that there are total 15 outliers.

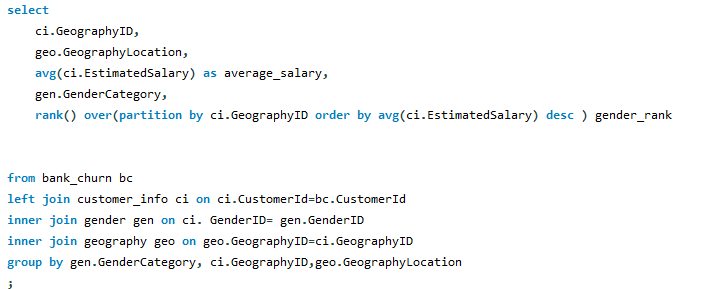
1. How many different tables are given in the dataset, out of these tables which table only consists of categorical variables?

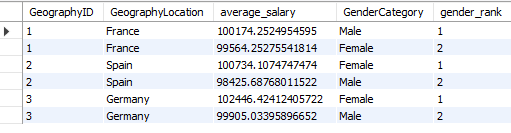
: We Have Seven Different tables i.e., ActiveCustomer,

Bank\_Churn,CreditCard,CustomerInfo,ExitCustomer,Gender,Geography.  Tables with Categorical Variables:

* CustomerInfo:Contains categorical variables like Surname.
* ExitCustomer: Contains categorical variables like Exit Category(Exit ,Retain).
* Gender: Contains categorical variables like Gender Category (Male,Female).
* Geography:Contains categorical variables like Geography Location (France, Spain,Germany).
* ActiveCustomer: Contains categorical variables like Active Category (Active Member , Inactive Member).
* CreditCard: Contains categorical variables like Category (Credit-card holder , Non-Credit card holder)

1. Using SQL, write a query to find out the gender-wise average income of males and females in each geography id. Also, rank the gender according to the average value. (SQL)





The SQL query provided is designed to calculate the average income of males and females in each geography id. It also ranks the genders according to the average income value. The query uses a combination of SELECT, FROM, LEFT JOIN, INNER JOIN, GROUP BY, and window function RANK() to achieve this.

The SELECT statement is used to select the required columns from the tables. It includes ci.GeographyID, geo.GeographyLocation, avg(ci.EstimatedSalary) as average\_salary, gen.GenderCategory, and rank() over ci.GeographyID ordered by avg(ci.EstimatedSalary) in descending order as gender\_rank.

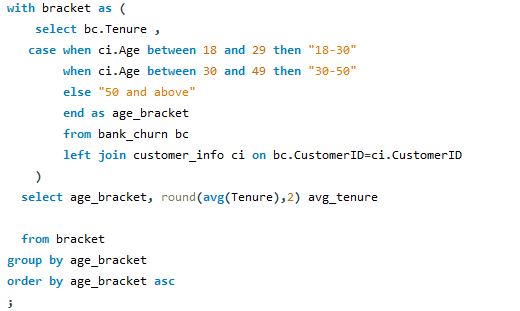
The FROM clause specifies the bank\_churn table as the main table. The LEFT JOIN and INNER JOIN clauses are used to combine rows from customer\_info, gender, and geography tables based on common columns.

The GROUP BY clause groups the result-set by gen.GenderCategory, ci.GeographyID, and geo.GeographyLocation.

The output of the query shows the GeographyID, GeographyLocation, average\_salary, GenderCategory, and gender\_rank. For example, in France (GeographyID 1), the average salary of males is higher than that of females, hence males are ranked 1. Similarly, in Spain (GeographyID 2) and Germany (GeographyID 3), females have a higher average salary, hence they are ranked 1.

This query is a powerful tool for analyzing gender-wise income distribution across different geographies. It can provide valuable insights for decision-making in areas such as marketing strategy, product development, and customer service.

1. Using SQL, write a query to find out the average tenure of the people who have exited in each age bracket (18-30, 30-50, 50+).





**Objective**: The SQL query was written to determine the average tenure of customers who have exited, categorized by different age brackets (18-30, 30-50, 50+).

**Method**: The query first creates a temporary table bracket that includes the Tenure and age\_bracket for each customer. The age\_bracket is determined based on the Age column from the customer\_info table.

**Results**: The query then calculates the average tenure (avg\_tenure) for each age\_bracket. The results are as follows:

For the age bracket 18-30, the average tenure of exited customers is 4.85.

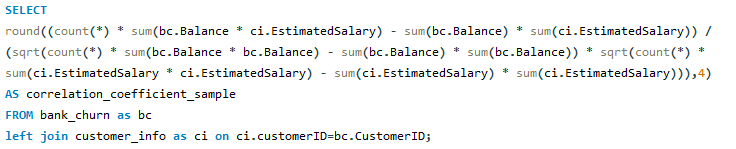
For the age bracket 30-50, the average tenure of exited customers is 4.87.

For the age bracket 50 and above, the average tenure of exited customers is 4.86.

**Conclusion**: The average tenure of exited customers is quite similar across all age brackets, indicating that age might not be a significant factor influencing the tenure of customers.

**Further** **Analysis**: Additional factors could be investigated to understand their impact on customer tenure, such as account balance, number of products, credit score, etc. This could provide a more comprehensive understanding of customer behaviour and help in developing effective customer retention strategies.

1. Is there any direct correlation between salary and the balance of the customers? And is it different for people who have exited or not?





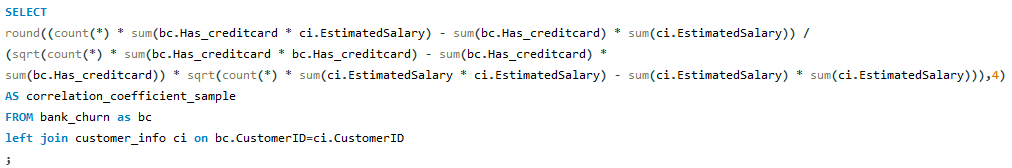
Objective: The aim was to investigate the correlation between the balance and estimated salary of customers. This was done to understand if there’s a direct relationship between these two variables.

Method: A SQL query was used to calculate the correlation coefficient between the Balance from the bank\_churn table and the EstimatedSalary from the customer\_info table. The correlation coefficient was calculated using the formula for Pearson’s correlation coefficient.

Result: The correlation coefficient obtained was **0.0128**.

Interpretation: The correlation coefficient ranges from -1 to 1. A value closer to 0 implies a weaker correlation. In this case, the correlation coefficient is very close to 0, which suggests that there is a very weak, if any, correlation between a customer’s balance and their estimated salary. **This means that a customer’s balance and their estimated salary do not directly influence each other**.

1. Is there any correlation between the salary and the Credit score of customers?

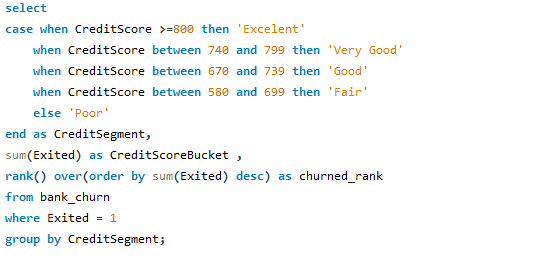


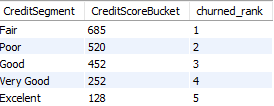


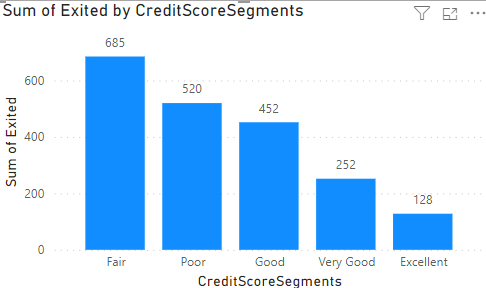
The output of query is -0.0099. This value is very close to zero, suggesting that there is no significant linear correlation between the presence of a credit card and the estimated salary of the customers in dataset.

In other words, having a credit card does not seem to be associated with higher or lower estimated salaries among customers. This could imply that the decision to have a credit card is not significantly influenced by the customer’s salary.

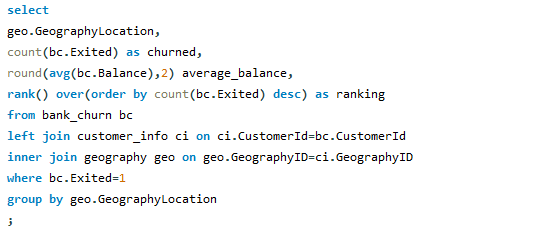
1. Rank each bucket of credit score as per the number of customers who have churned the bank.

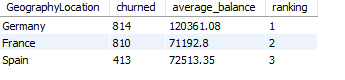






1. According to the age buckets find the number of customers who have a credit card. Also retrieve those buckets that have lesser than average number of credit cards per bucket.
2. Rank the Locations as per the number of people who have churned the bank and average balance of the customers.





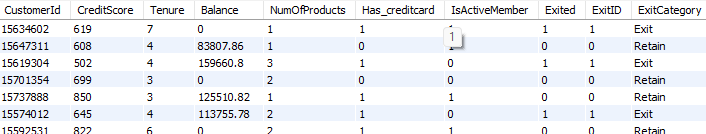
1. As we can see that the “CustomerInfo” table has the CustomerID and Surname, now if we have to join it with a table where the primary key is also a combination of CustomerID and Surname, come up with a column where the format is “CustomerID\_Surname”.





1. Without using “Join”, can we get the “ExitCategory” from ExitCustomers table to Bank\_Churn table? If yes do this using SQL.



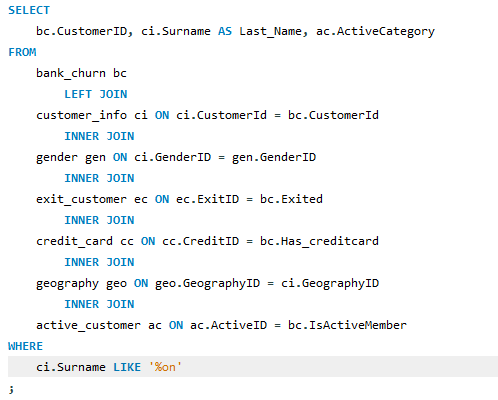


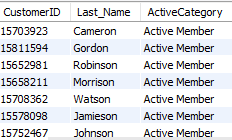
1. Were there any missing values in the data, using which tool did you replace them and what are the ways to handle them?

There are no missing data in our dataset that will result a good analysis because a missing data can leads to bias analysis, improper prediction.

If there could be null values it can be replace by mean , mode or median method.

1. Write the query to get the customer IDs, their last name, and whether they are active or not for the customers whose surname ends with “on”.



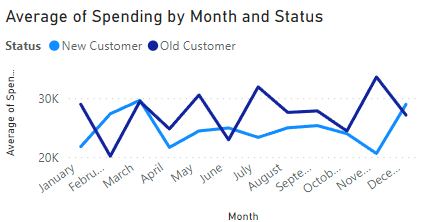


1. Can you observe any data disrupency in the Customer’s data? As a hint it’s present in the IsActiveMember and Exited columns. One more point to consider is that the data in the Exited Column is absolutely correct and accurate.

Ans : -There is no data discrepancy in the IsActiveMember column. There is only 1 and 0 in the column which is also available in Exited column.

**Subjective Question:**

1. Customer Behavior Analysis: What patterns can be observed in the spending habits of long-term customers compared to new customers, and what might these patterns suggest about customer loyalty?



1. Product Affinity Study: Which bank products or services are most commonly used together, and how might this influence cross-selling strategies?

Ans; Customers often use specific bank products together. Analyzing these pairings helps develop targeted cross-selling strategies to increase customer satisfaction and revenue.

Commonly Used Products (Examples):

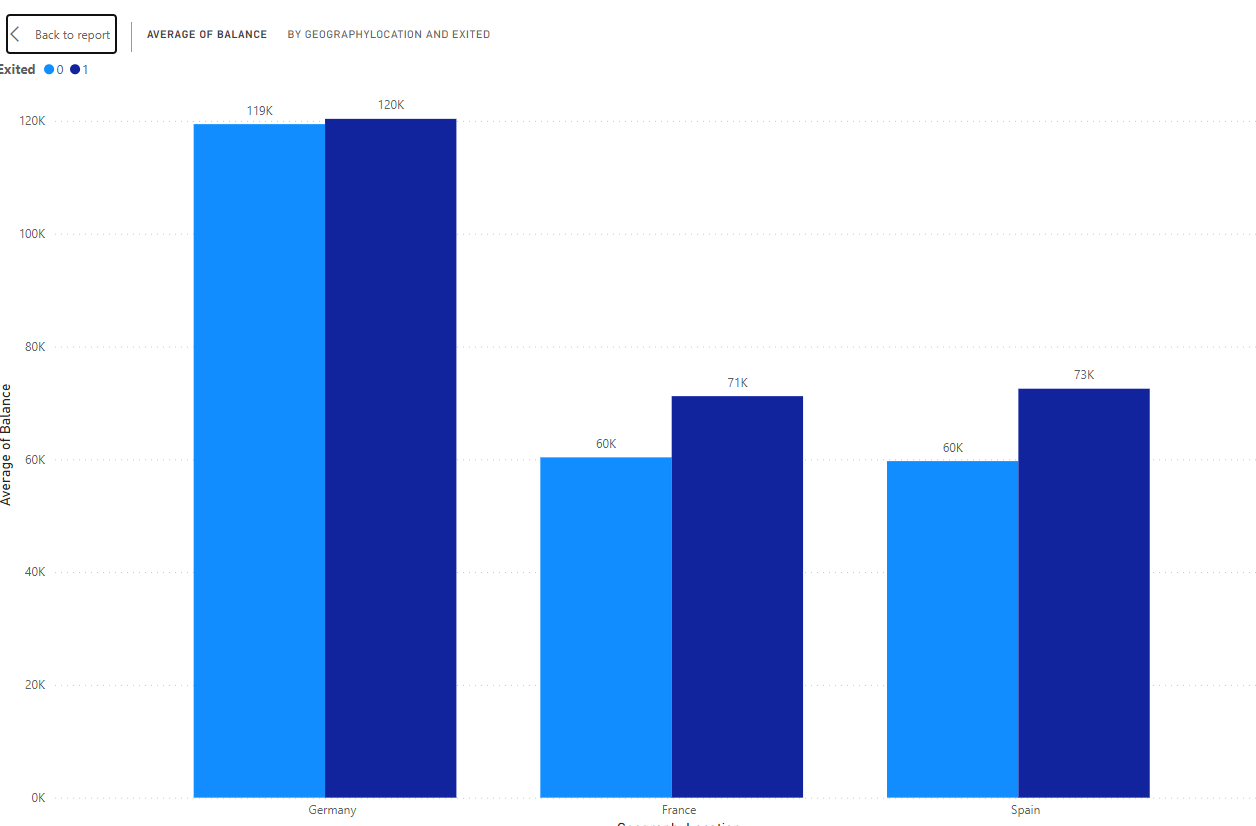
* Checking Accounts: Core for everyday transactions.
* Debit Cards: Linked to checking, providing convenient access to funds.
* Savings Accounts: Grow savings and often earn interest.
* Credit Cards: Offer a line of credit for purchases, requiring repayment with interest.
* Loans: Tailored financial solutions like mortgages or auto loans.

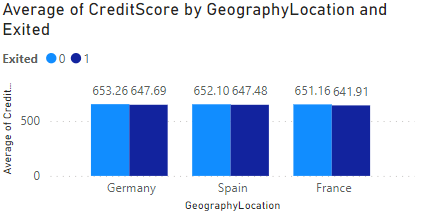
Cross-Selling Strategies:

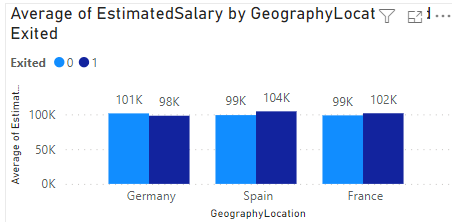
* Recommend complementary products: Checking account users might benefit from a debit card and online banking for easy management. Savings account holders could be recommended automatic transfers to boost saving habits or higher-interest options like CDs for larger balances.
* Personalize based on usage: Credit card users with travel habits could benefit from travel rewards cards. Loan seekers might be interested in bundled insurance options.
* Leverage digital platforms: Promote paperless statements and bill autopay through online/mobile banking. Offer investment options or financial tools accessible through these platforms.

By understanding product usage patterns and tailoring recommendations, banks can create a win-win situation: increased revenue and customers who feel their financial needs are being met.

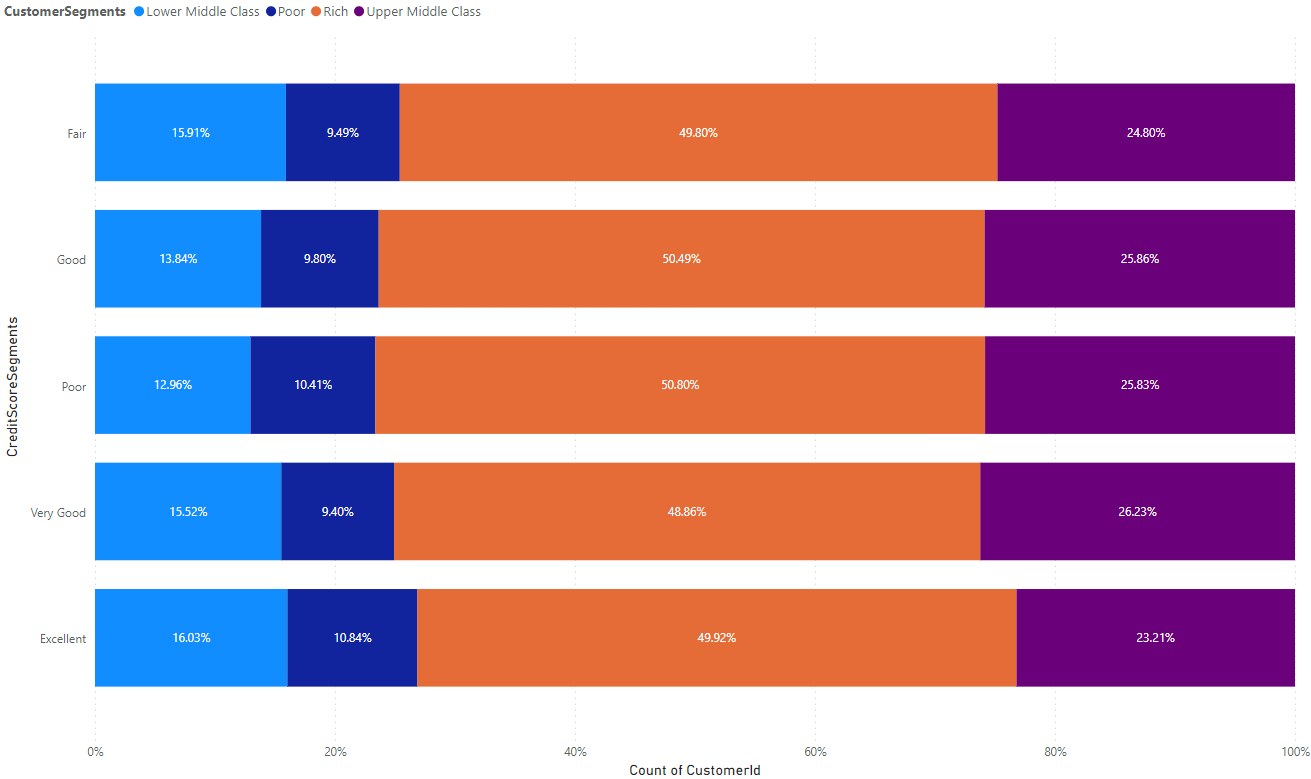
1. Geographic Market Trends: How do economic indicators in different geographic regions correlate with the number of active accounts and customer churn rates?







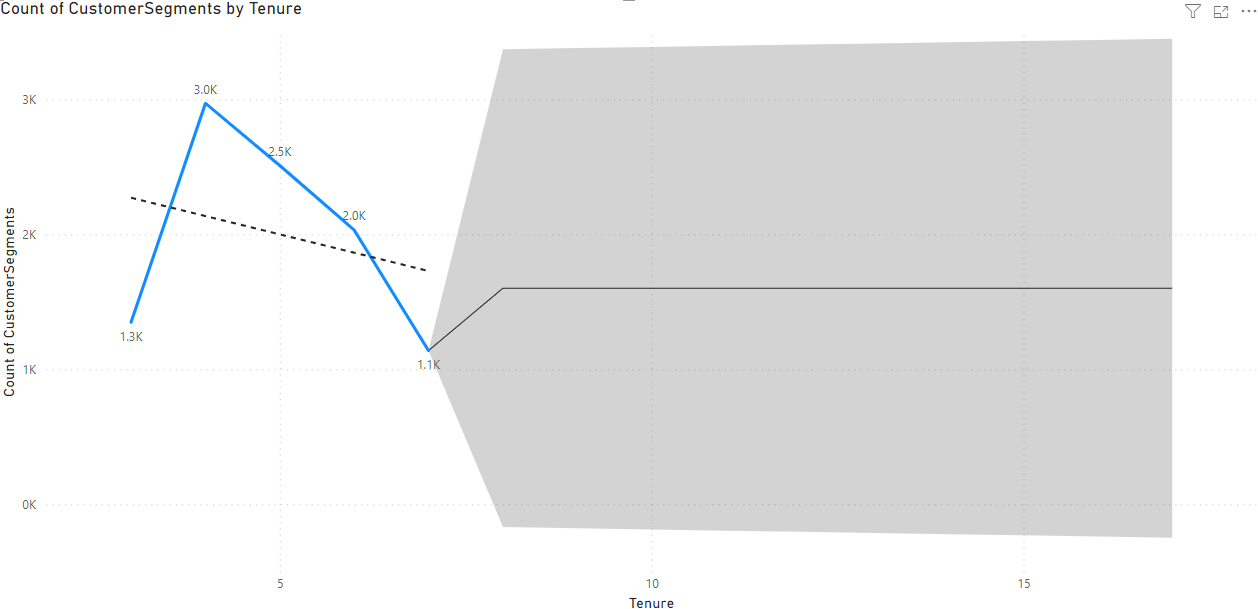
1. Risk Management Assessment: Based on customer profiles, which demographic segments appear to pose the highest financial risk to the bank, and why?



Based on the chart, the demographic segments with the most customers are those with lower credit scores ("Lower Middle Class" and "Poor"). These same segments also represent the largest portion of the bank's customers. Therefore, based on the data presented in the chart, we can say that the demographic segments with the most customers are also the ones that pose the highest potential financial risk to the bank.

There are other factors to consider when assessing customer risk besides creditworthiness, such as income and employment history. However, creditworthiness is a strong indicator of a customer's ability to repay loans. Customers with lower credit scores are statistically more likely to default on loans than customers with high credit scores. This means that the bank is more likely to lose money on loans made to customers in the "Lower Middle Class" and "Poor" segments than on loans made to customers in other segments.

1. Customer Tenure Value Forecast: How would you use the available data to model and predict the lifetime (tenure) value in the bank of different customer segments?



1. Marketing Campaign Effectiveness: How could you assess the impact of marketing campaigns on customer retention and acquisition within the dataset? What extra information would you need to solve this?

Ans: To assess the impact of marketing campaigns on customer retention and acquisition within a dataset, you would typically use a combination of data analysis and statistical techniques. Here's a general approach you could take:

1. Define Metrics: Define key metrics for customer retention and acquisition. For retention, you might use metrics like customer churn rate or retention rate. For acquisition, you might use metrics like new customer acquisition rate or customer acquisition cost (CAC).

2. Segment Data: Segment the data based on different marketing campaigns. This will allow you to analyze the impact of each campaign separately.

3. Calculate Metrics: Calculate the defined metrics for each segment and for each time period (e.g., monthly, quarterly, annually). This will help you understand how each campaign is affecting customer retention and acquisition over time.

4. Compare Results: Compare the metrics across different campaigns to identify which campaigns are most effective at retaining and acquiring customers.

5. Statistical Analysis: Use statistical tests (e.g., t-tests, ANOVA) to determine if the differences in metrics between campaigns are statistically significant.

6. Additional Information:To perform a comprehensive analysis, you may need additional information such as:

- Customer demographics: To understand if certain demographics respond better to certain campaigns.

- Campaign details: To understand the specifics of each campaign (e.g., duration, channels used, messaging).

- Competitor data: To understand the competitive landscape and how it might be impacting your results.

- External factors: Such as economic conditions, seasonality, or industry trends that might affect customer behavior.

By following this approach and gathering the necessary information, we can assess the impact of marketing campaigns on customer retention and acquisition within your dataset.

1. Customer Exit Reasons Exploration: Can you identify common characteristics or trends among customers who have exited that could explain their reasons for leaving?



By the reference of the chart we can easily see that Exit Customer in case of count, Average Balance ,Average Estimated salary and Num of the product buy are all together high as compared to the costumer who are retained. So it is very much important to analyse this churn reason and take actionable plan to tackle this bank internal challenges.

Some points are mentioned below for the possible reason of customer exits.

**Geographical Location:** Customers from Germany show a higher count of exited individuals compared to those from France and Spain.

**Credit Card Ownership:** The presence of a credit card does not show a significant difference in exit counts.

**Membership Activity**: Inactive members have a noticeably higher exit count than active members.

**Number of Products:** Customers with only one product have the highest exit count compared to those with more products.

These observations suggest that geographical location (specifically Germany), membership activity level (inactive members), and the number of products used (one product) are common characteristics among customers who have exited. This information could be useful in developing strategies to reduce customer churn

1. Are 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' important for predicting if a customer will leave the bank?

Ans:

**Tenure**: The number of years that the customer has been a client of the bank can be a significant factor. Normally, older clients are more loyal and less likely to leave a bank1.

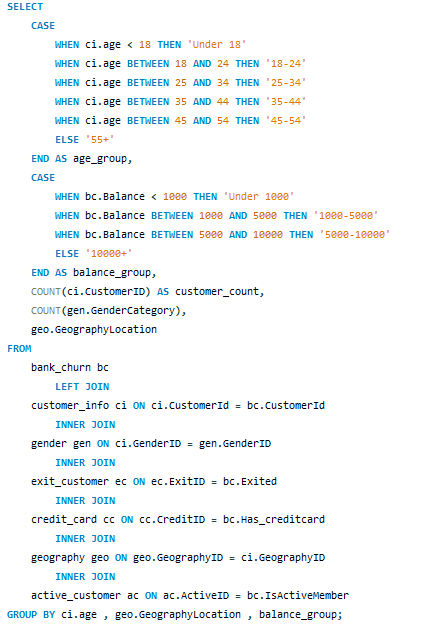
**NumOfProducts**: The number of products a customer has with the bank could potentially influence their decision to stay or leave. Customers with multiple products might have a higher level of engagement with the bank, which could reduce the likelihood of churn2.

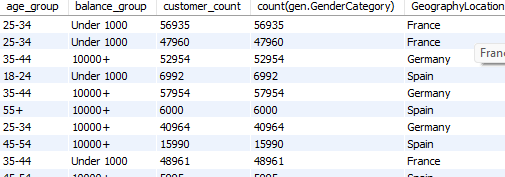
**IsActiveMember**: Whether a customer is active or not can be a strong predictor of churn. Active customers are likely to be more engaged with the bank’s services, and therefore, may be less likely to churn2.

**EstimatedSalary**: While one might assume that a customer’s estimated salary could influence their likelihood to churn, it has been found that estimated salary showed little to no correlation with churn3. This could be due to a variety of factors, such as the customer’s satisfaction with the bank’s services, which might not necessarily be tied to their income3.

Remember, these are general trends and may not apply to every customer. A comprehensive analysis should be conducted using appropriate statistical and machine learning techniques to determine the significance of these factors in predicting customer churn4. It’s also important to note that the importance of these factors can vary depending on the specific context and customer base of the bank

1. Utilize SQL queries to segment customers based on demographics and account details.



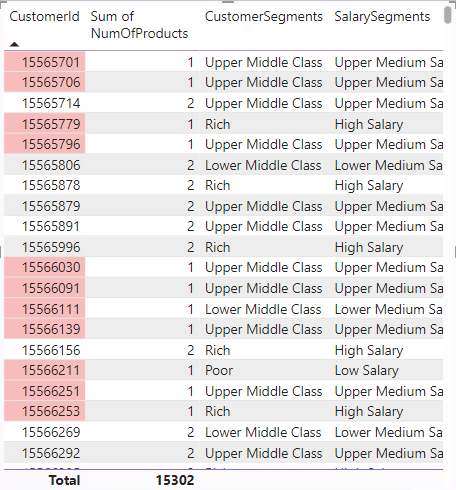


1. How can we create a conditional formatting setup to visually highlight customers at risk of churn and to evaluate the impact of credit card rewards on customer retention?

Ans:

1. Identify the churn criteria: Define the criteria to identify customers at risk of churn. This could be based on a combination of factors, such as:  
   1. Customers with a low number of products purchased (NumOfProducts)
   2. Customers with low balance segments
   3. Customers who have recently exited (Exited) in the past (e.g., in the last 6 months)
2. Conditional formatting based on churn criteria: Apply a conditional formatting rule to highlight cells that meet the churn criteria. You can format the cells with a different background color or font to make them visually distinct.
3. Filter by Credit Card ownership: Create a filter for the "HasCrCard" field. This will allow you to segment customers by whether they have a credit card or not.

Evaluate churn rate by Credit Card ownership: Analyze the churn rate (percentage of customers who exited) for customers with and without credit cards. You can calculate this by comparing the number of exited customers (where Exited = 1) to the total number of customers in each segment (HasCrCard = Yes or No).



1. What is the current churn rate per year and overall as well in the bank? Can you suggest some insights to the bank about which kind of customers are more likely to churn and what different strategies can be used to decrease the churn rate?

Ans: This analyzes customer churn rates and identifies segments most susceptible to churn. It also proposes strategies to decrease churn and improve customer retention.

Churn Rate:

* The overall churn rate for the bank is 20.37%.
* Year-on-year churn rates show some fluctuations:
  + 2016: 19.27%
  + 2017: 22.35% (highest)
  + 2018: 20.21%
  + 2019: 19.86% (lowest)

Customer Segments Prone to Churn:

Data analysis suggests a customer segment with a higher likelihood of churn:

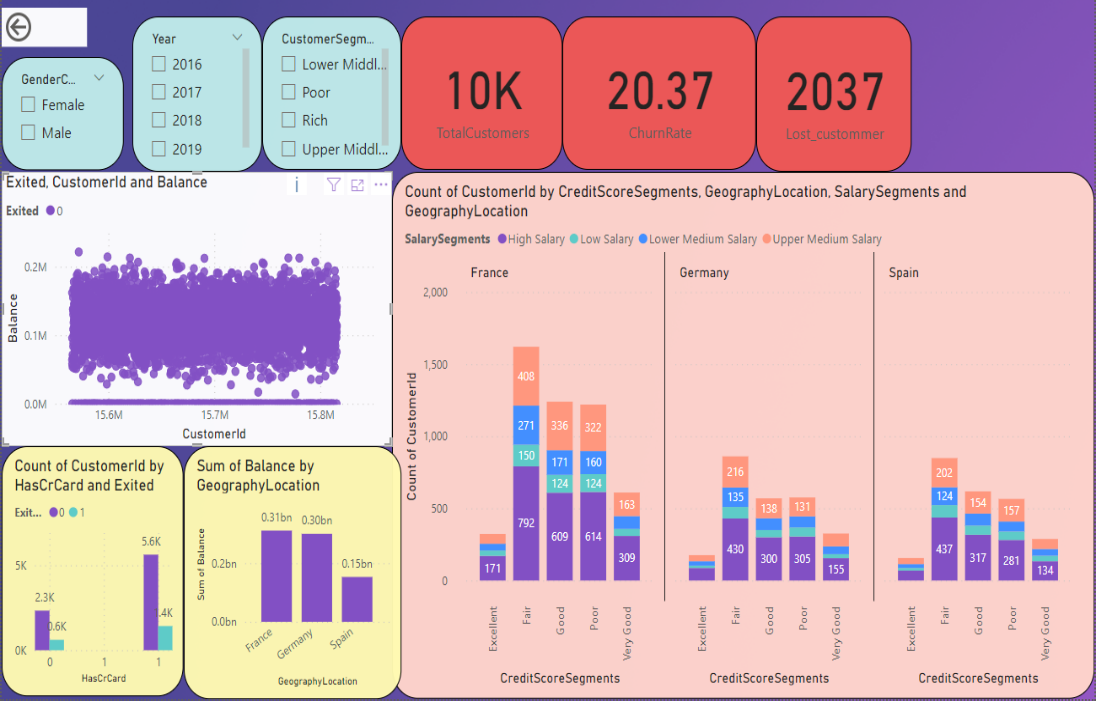
* Purchases 1 product: Customers who only use one product by bank might not find enough value compared to competitors offering wider ranges or integrated services.
* Has credit card:Potential reasons for churn among credit card holders could be:
  + Limited credit limits not meeting their needs.
  + Lack of rewards programs that incentivize them to keep the card.
  + High credit card fees.
* Tenure of 4-5 years: Customers with this tenure might be nearing the end of introductory offers or discounts, making them susceptible to competitor offers with better rates or features.
* High salary: High earners might have more options and be more likely to switch for a slightly better interest rate or benefit elsewhere.

Recommendations to Reduce Churn:

* Targeted Product Bundles: Create product bundles that cater to specific customer segments and needs. Offer these bundles to customers who only use one product, highlighting the additional benefits and potential cost savings.
* Enhanced Credit Card Rewards: Improve credit card rewards programs for existing customers. This could involve:
  + Increasing credit limits based on customer history and creditworthiness.
  + Offering rewards programs aligned with spending habits (e.g., travel rewards, cash back for specific categories).
  + Reducing or eliminating annual fees, especially for high-value customers.
* Retention Offers for Existing Customers: Proactively reach out to customers nearing the end of introductory offers with personalized retention deals. This could include extending introductory rates or offering discounts on other products or services.
* Customer Satisfaction Surveys: Regularly conduct customer satisfaction surveys to understand why customers churn. This can help identify areas for improvement and tailor retention strategies accordingly.

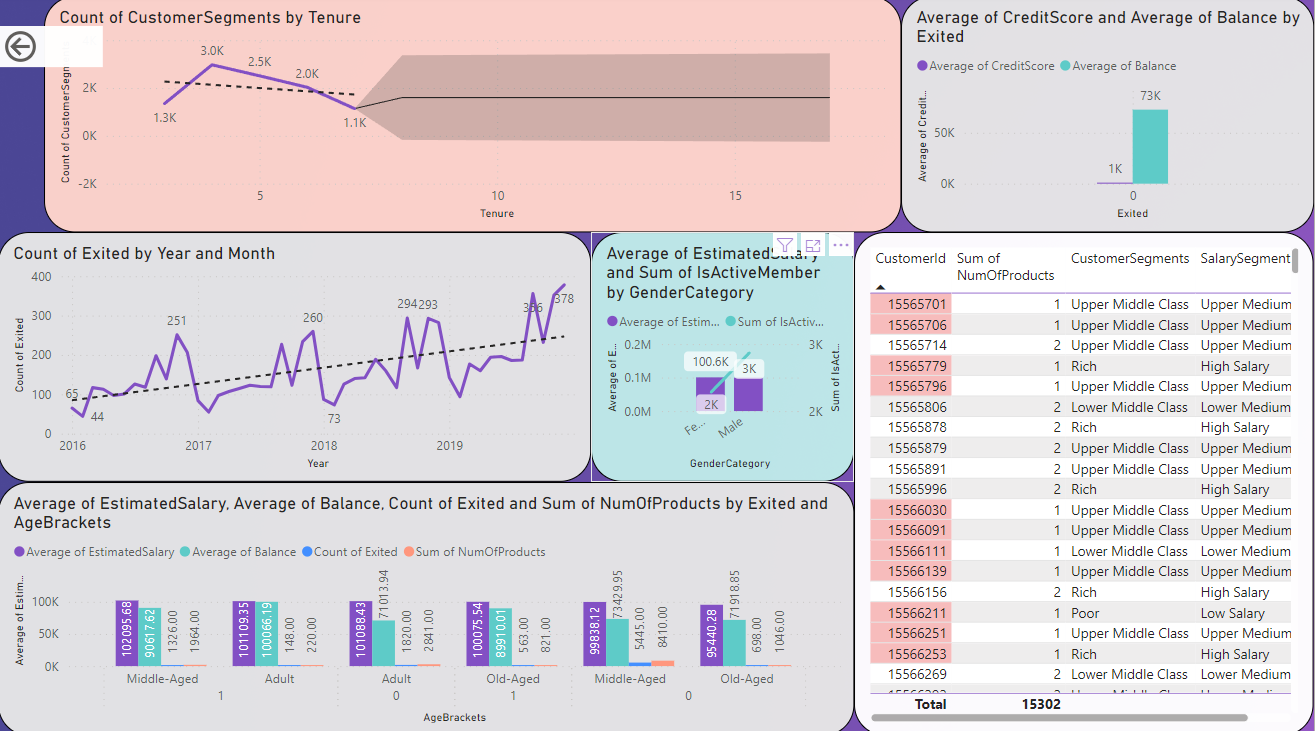
Relationship Management for High-Value Customers: Develop dedicated relationship managers for high-value customers to provide personalized service, address their specific needs, and offer exclusive benefits

1. Create a dashboard incorporating all the KPIs and visualization-related metrics. Use a slicer in order to assist in selection in the dashboard.









1. How would you approach this problem, if the objective and subjective questions weren't given?

Ans: By looking over given dataset, I can conclude that this dataset is of bank churn so according to this i will analyse.

1. In the “Bank\_Churn” table how can you modify the name of the “HasCrCard” column to “Has\_creditcard”?

Ans: This can be done by using SQL function named Alter which will allow to modify rename column will allow to change the column name. This is done by using a SQL query which is shown in image given bellow.

