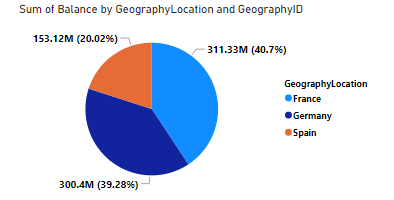
**Objective Questions**

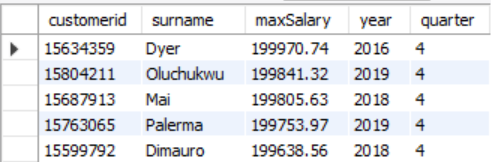
1. What is the distribution of account balance across different regions?



the distribution of account balance across different regions as follows

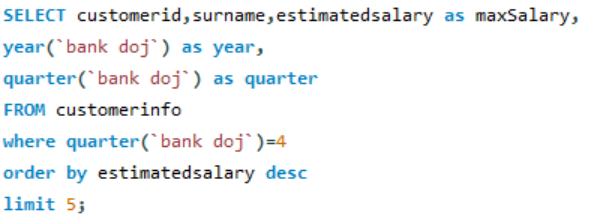
* france:311.33 million
* Germany:300.4 million
* Spain:153:12 million

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

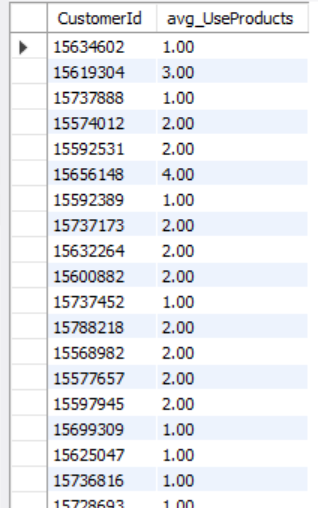


The above table give the top 5 customers with the highest estimated salary in the last quarter of the year

Sql query:

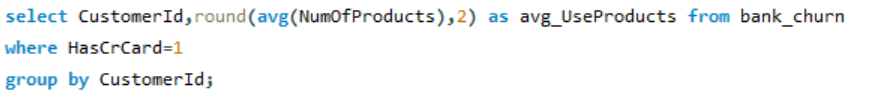


1. Calculate the average number of products used by customers who have a credit card. (SQL)



The above table gives the average number of products used by customers who have credit card

Query:



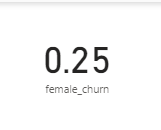
1. Determine the churn rate by gender for the most recent year in the dataset.

The recent year in the data set is 2019 and the churn rate for male is 0.15



Formula used: male\_churn = CALCULATE([male\_left]/[count\_of\_male],'bank customerinfo'[year]=2019)

Female churn rate is 0.25



Formula: female\_churn = CALCULATE([female\_left]/[count\_of\_female],'bank customerinfo'[year]=2019)

[ Note:

male\_left = CALCULATE(COUNT('bank customerinfo'[GenderID]),'bank gender'[GenderID]=1,'bank bank\_churn'[Exited]=1)

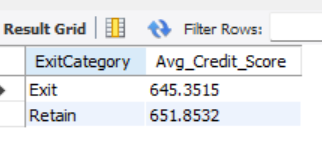
female\_left = CALCULATE(COUNT('bank customerinfo'[GenderID]),'bank gender'[GenderID]=2,'bank bank\_churn'[Exited]=1)

count\_of\_male = CALCULATE(COUNT('bank customerinfo'[GenderID]),'bank gender'[GenderID]=1)

count\_of\_female = CALCULATE(COUNT('bank customerinfo'[GenderID]),'bank gender'[GenderID]=2)

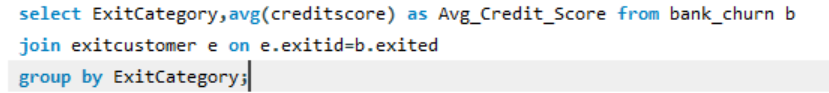
]

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



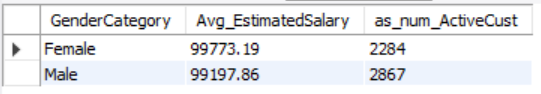
The average credit score of customers who have exited is less when compared to those who retain customers

Sql query:

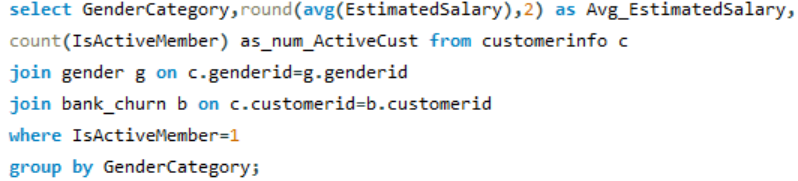


1. Which gender has a higher average estimated salary, and how does it relate to the number of active accounts? (SQL)

Female category have higher average estimated salary and this average estimated salary is not correlated with number of active accounts because female has higher average estimated salary but less number of active accounts when compared to the male category active accounts

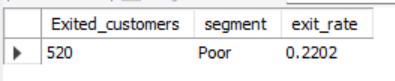


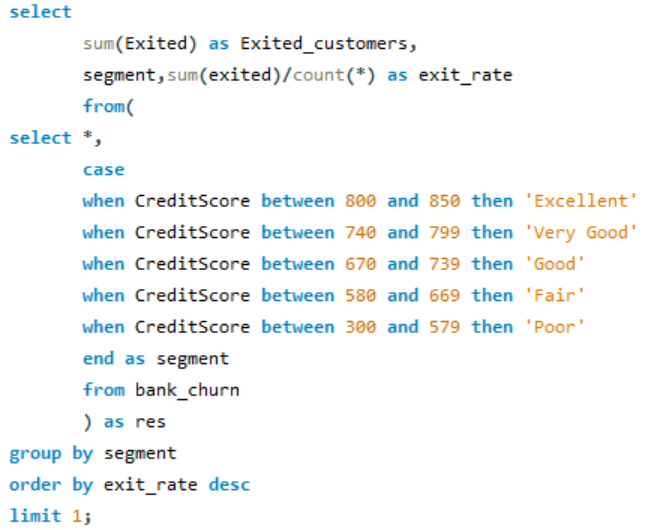
Sql query:



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

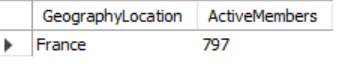
The poor segment has the highest exit rate of 0.2202

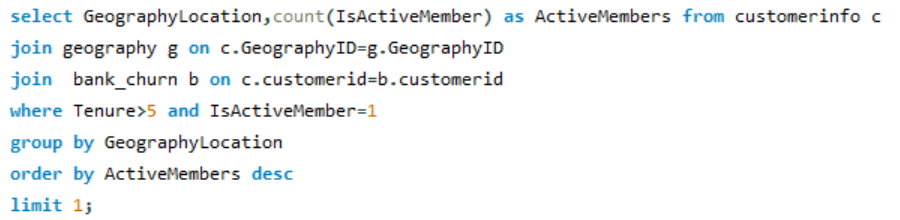


SQL query: 

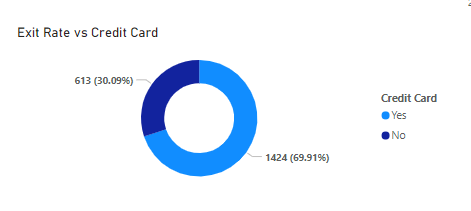
1. Find out which geographic region has the highest number of active customers with a tenure greater than 5 years. (SQL)

France has the highest number of active customers that is 797 active customers with a tenure greater than 5 years



Sql query: 

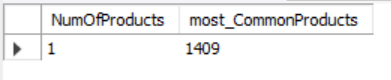
1. What is the impact of having a credit card on customer churn, based on the available data?



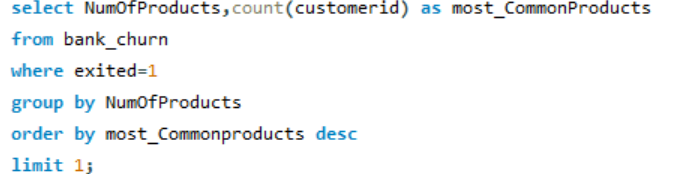
* The above chart indicates that customers who have a credit card are more likely to leave the bank compared to those who don't have one. This suggests that there might be something about our services related to credit cards that is causing dissatisfaction among these customers. Therefore, we should consider improvements specifically targeted towards our credit card services.

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

1 is the number of products which is the most common number of products used by the exited customer.

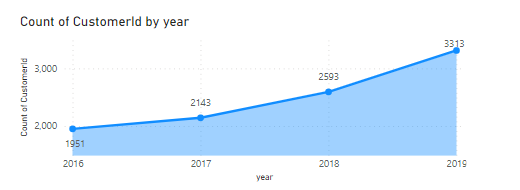


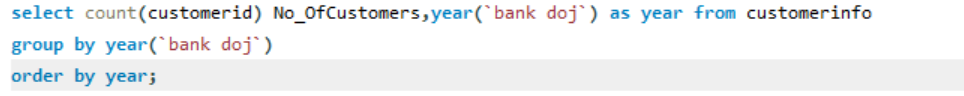
Sql query:



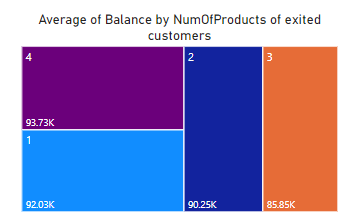
1. Examine the trend of customer joining over time and identify any seasonal patterns (yearly or monthly). Prepare the data through SQL and then visualize it.

The trend of customers increasing year on year, in 2019 the maximum number of customers can be observed

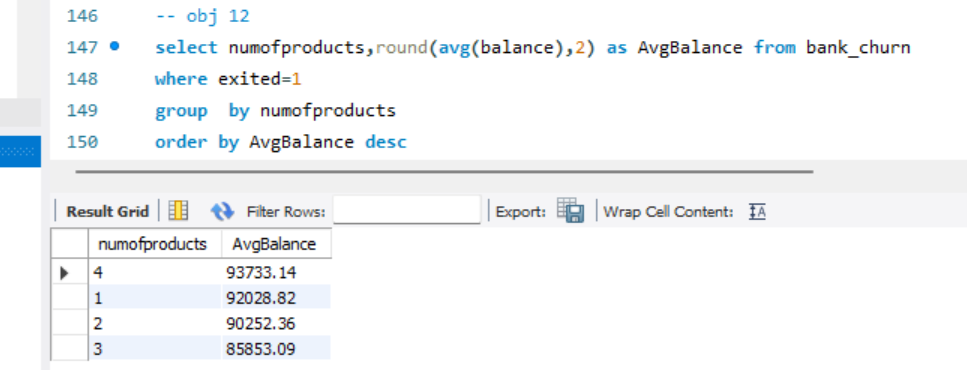


SQL query: 

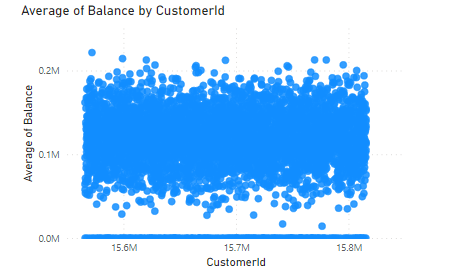
1. Analyze the relationship between the number of products and the account balance for customers who have exited.



Sql query:

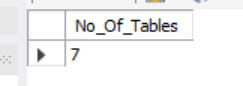


1. Identify any potential outliers in terms of balance among customers who have remained with the bank.



In the visual above, there are some datasets with notably high balances and others with very low balances, almost close to zero. These extreme cases represent the outliers in terms of balance among customers who have remained with the bank.

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



There are 7 tables in the given dataset

SQL query

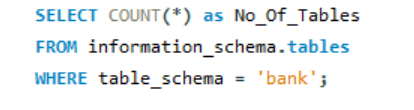
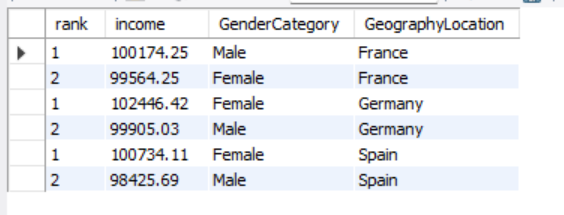


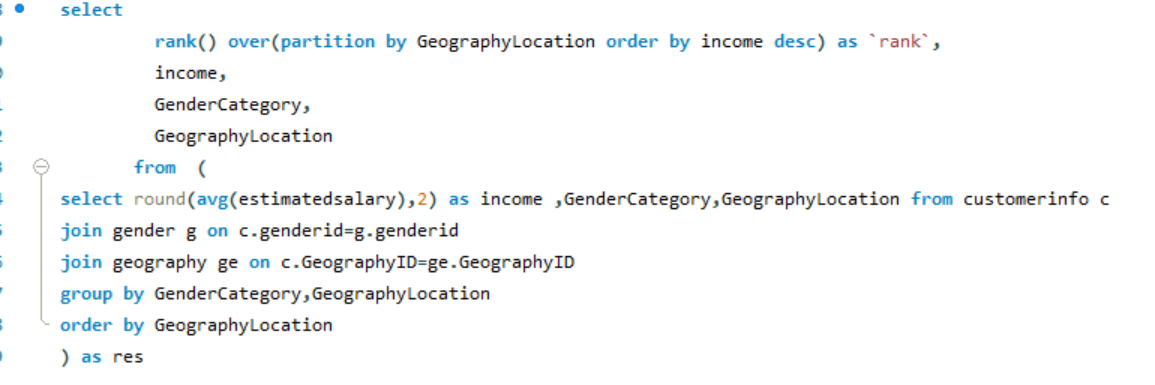
table only consists of categorical variables are :activecustomer ,creditcard, exitcustomer, gender, and geography. These are called categorical data because they consist of categories or groups into which data can be divided

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

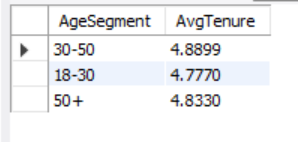


From the table above, we can discern the average income for each category within each geographic location. Additionally, the gender categories are ranked based on income within each category in the respective geographic locations.

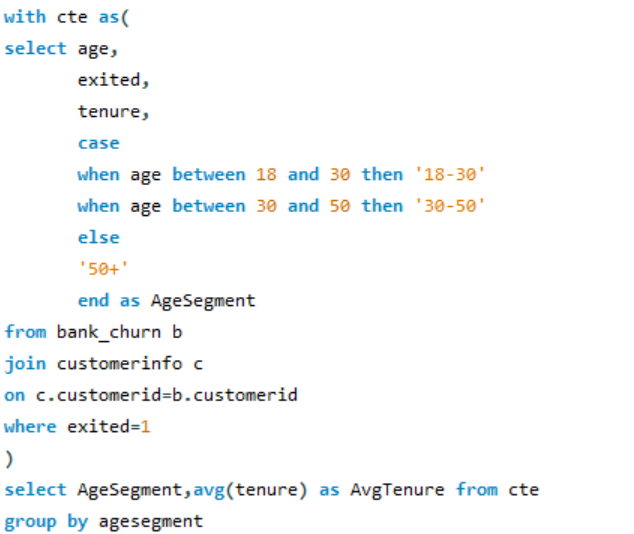
SQL query:



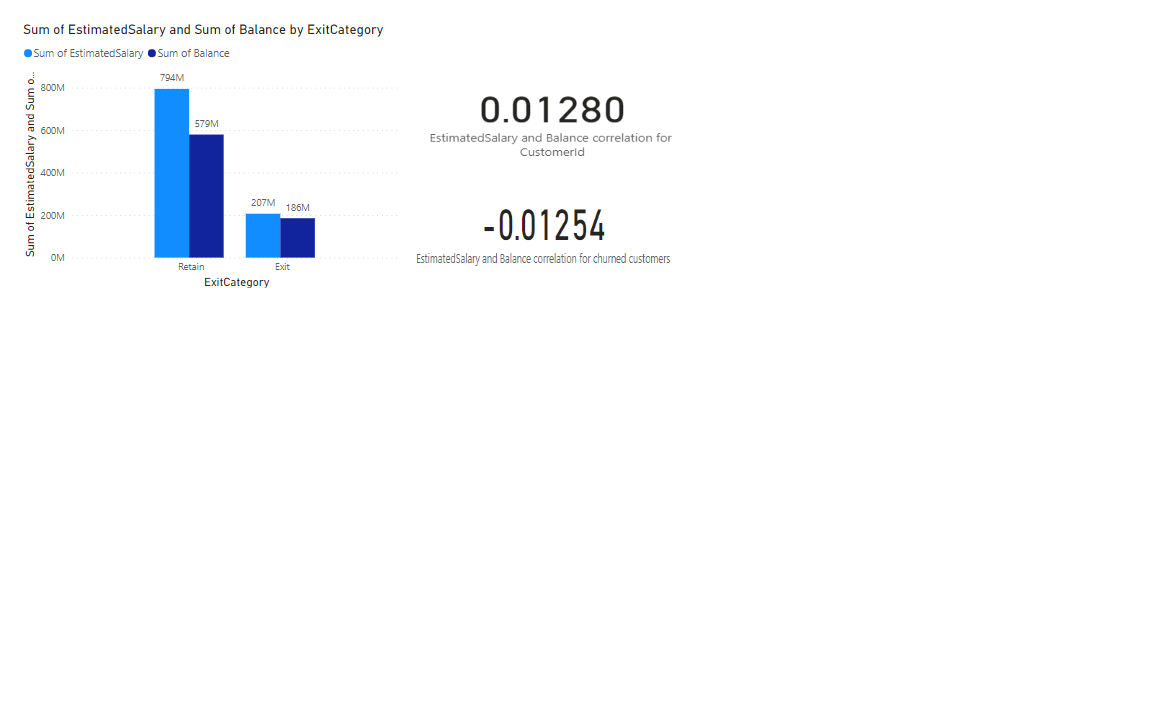
1. Using SQL, write a query to find out the average tenure of the people who have exited in each age bracket (18-30, 30-50, 50+).



SQL query:



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



In the analysis conducted, it was observed that there exists a weak correlation between the salary and balance of customers, with the correlation coefficient approximating zero. This suggests that there is little to no direct relationship between salary and balance.

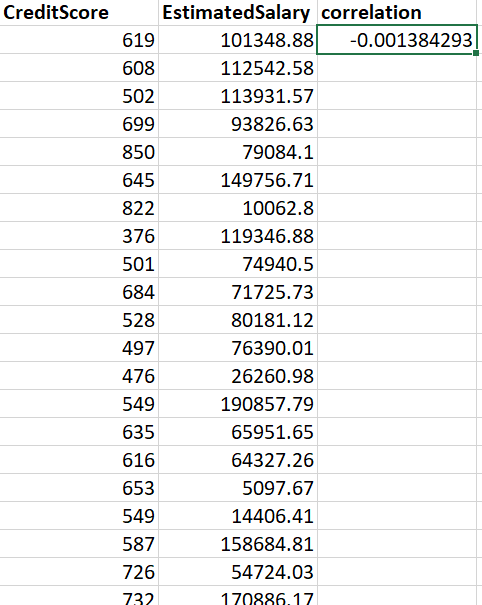
Furthermore, when specifically examining exited customers, a negative correlation was identified between salary and balance. This implies an inverse relationship, wherein higher salaries are associated with lower balances, and vice versa.

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

There is a negative correlation between salary and credit score

**Correlation=-0.001384293**

Excel formula used= =CORREL(K2:K10001,L2:L10001)

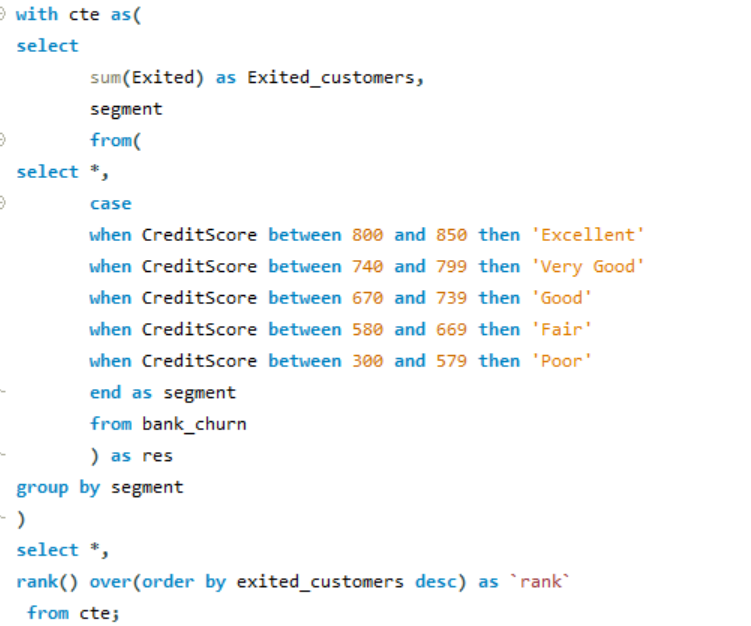


1. Rank each bucket of credit score as per the number of customers who have churned the bank.



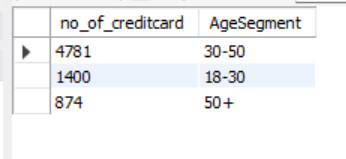
The above table ranks each bucket of credit score as per the number of customers who have churned the bank

SQL query

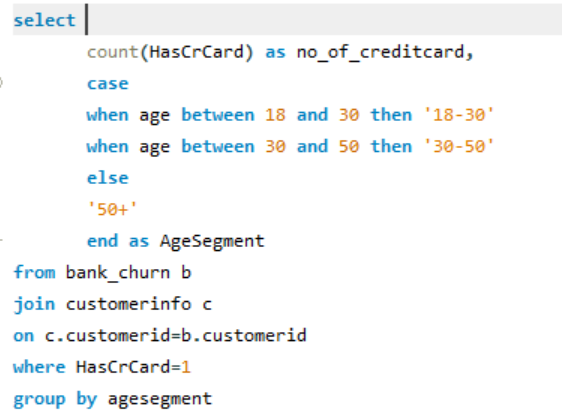


1. According to the age buckets find the number of customers who have a credit card. Also retrieve those buckets who have lesser than average number of credit cards per bucket.

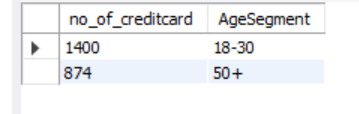
* According to the age buckets/segment number of customers having credit card are



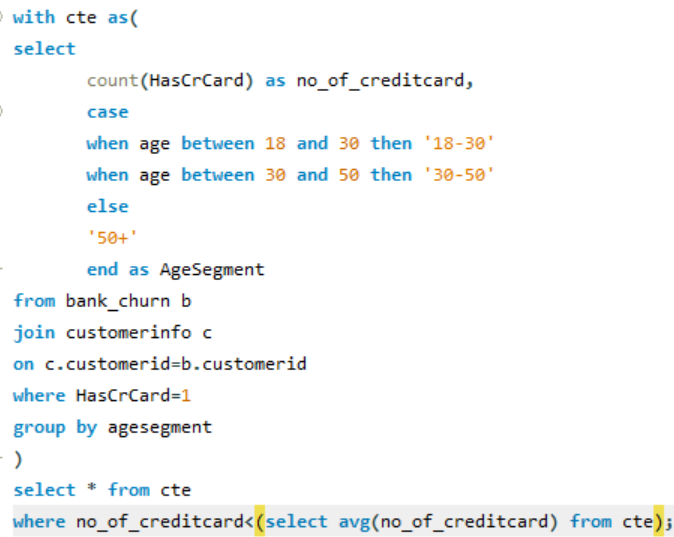
Sql query:



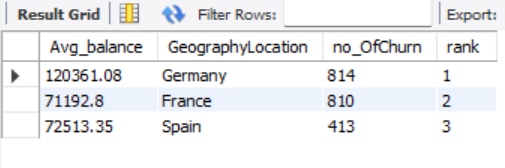
* Age buckets which have lesser than average number of credit cards per bucket are



SQL query

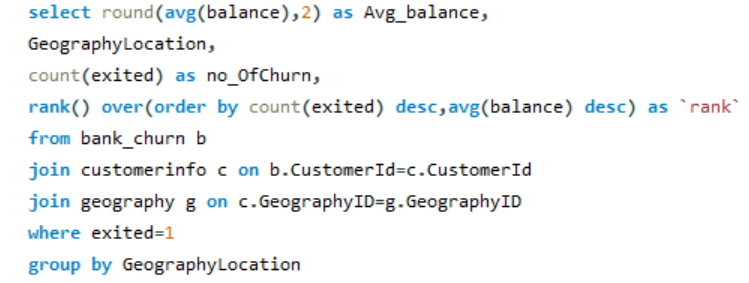


1. Rank the Locations as per the number of people who have churned the bank and average balance of the customers.



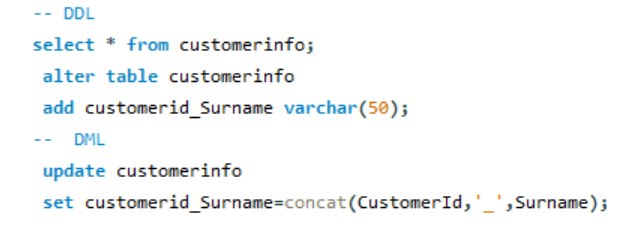
From the above table we will get the highest and the lowest rank of the locations based on number of people churned and average balance of customers

SQL query:

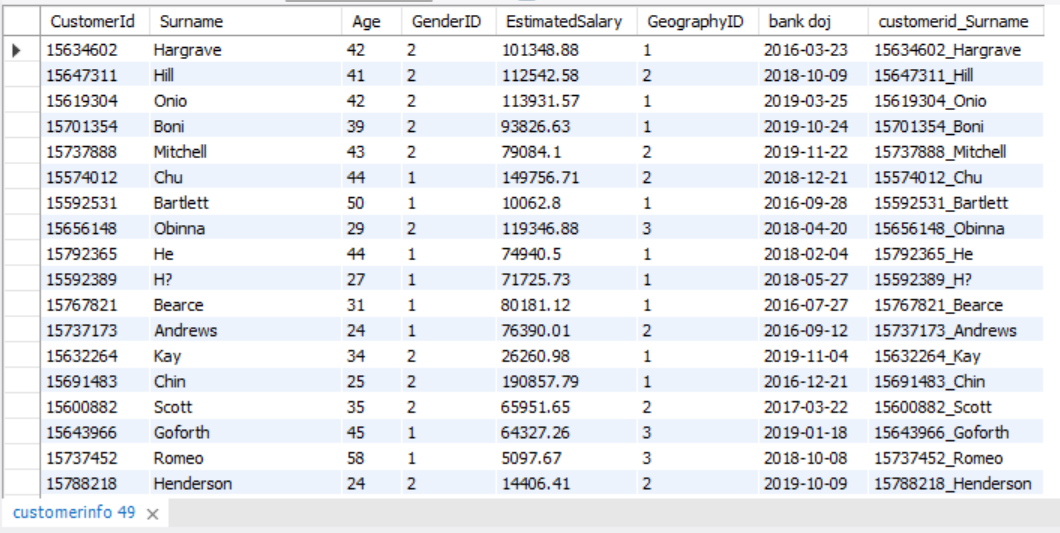


1. As we can see that the “CustomerInfo” table has the CustomerID and Surname, now if we have to join it with a table where the primary key is also a combination of CustomerID and Surname, come up with a column where the format is “CustomerID\_Surname”.

To accomplish this task, I began by altering the table structure to include a new column named CustomerID\_Surname using Data Definition Language (DDL). Then, I utilized Data Manipulation Language (DML) to update the records accordingly. This process ensured that the new column was populated with the desired format "CustomerID\_Surname", merging the CustomerID and Surname values.Sql query:

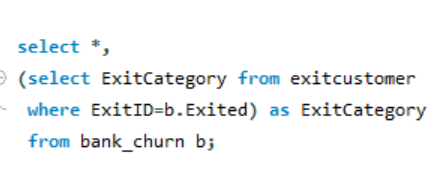


Output:



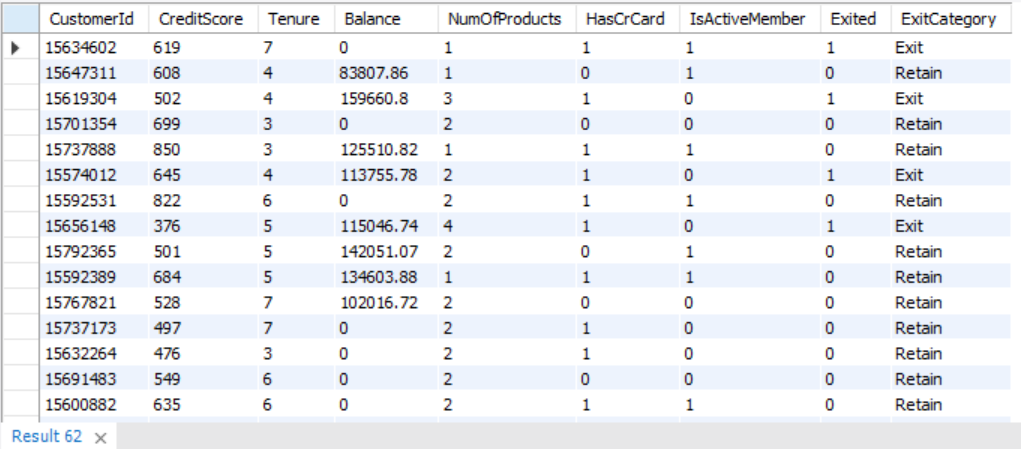
1. Without using “Join”, can we get the “ExitCategory” from ExitCustomers table to Bank\_Churn table? If yes do this using SQL.

Yes, without using join we can get “ExitCategory” from ExitCustomers table to the Bank\_Churn table using the **subquery**

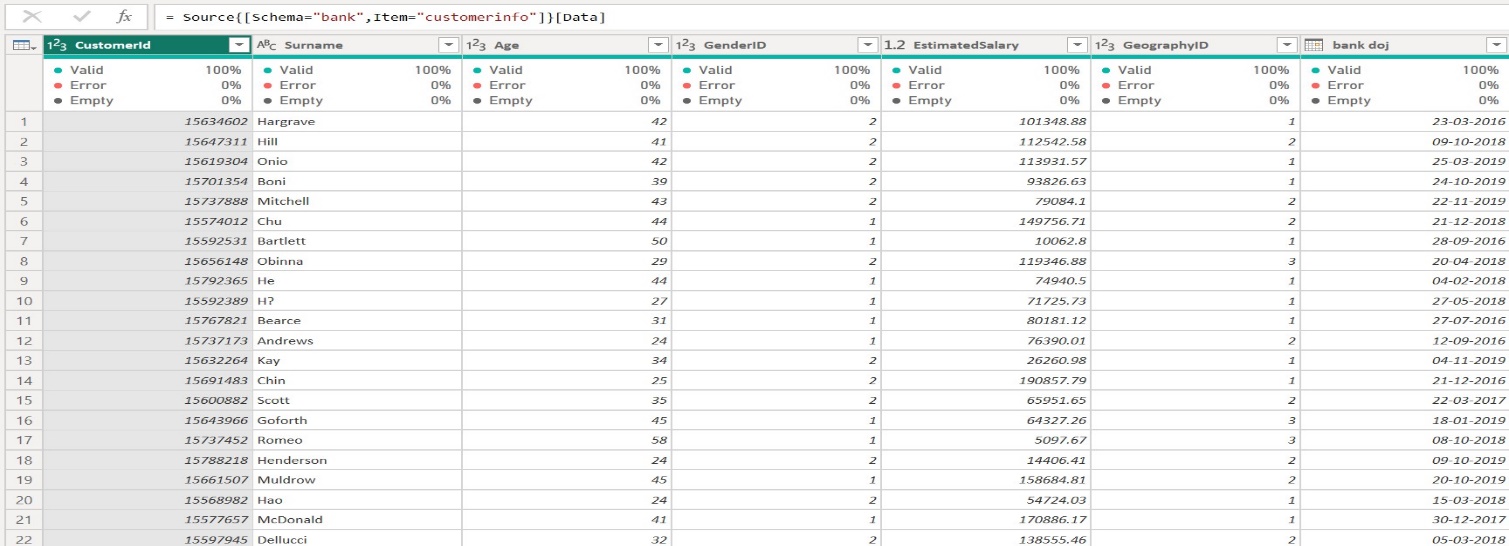


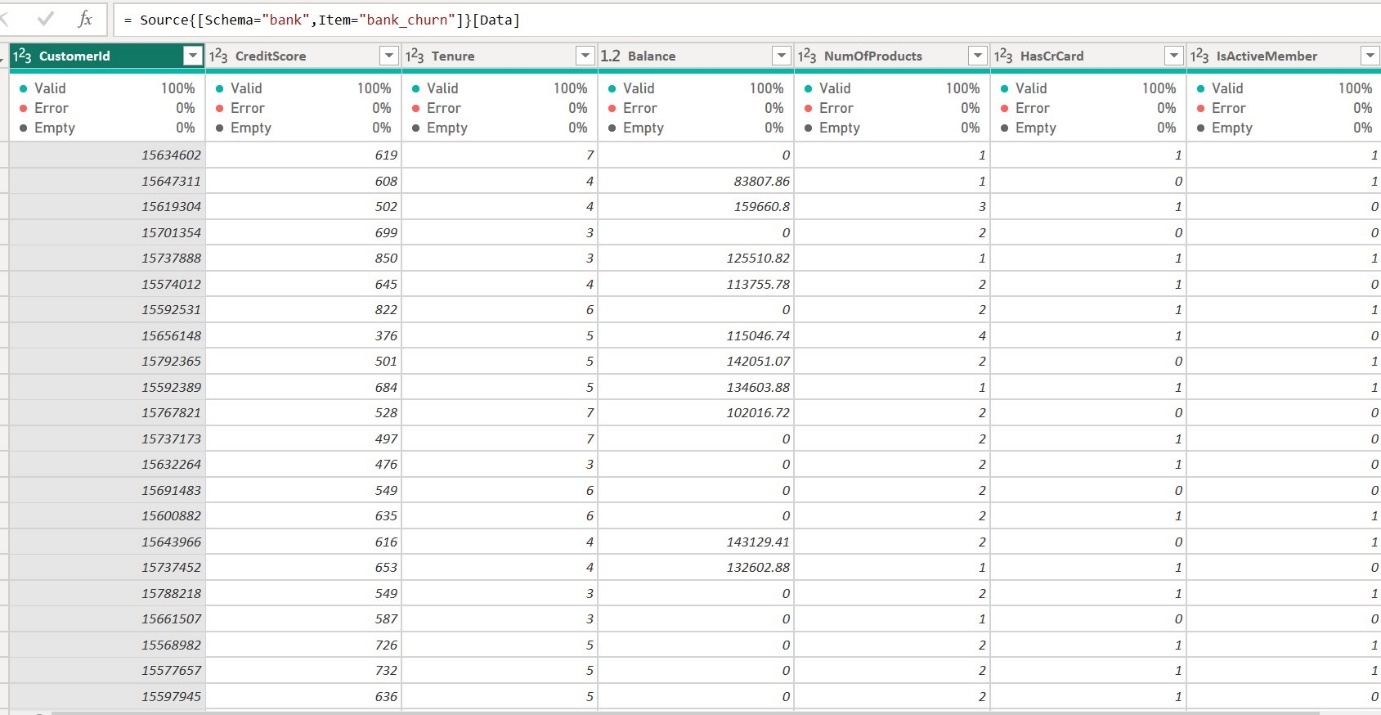
above query retrieves all columns from the `bank\_churn` table and adds a new column named `ExitCategory`. Within the query, a subquery is employed to fetch the `ExitCategory` from the `exitcustomer` table based on the condition where `ExitID` matches the `Exited` field in the main query's result set. This effectively brings in the `ExitCategory` information from the `exitcustomer` table into the `bank\_churn` table without using a join operation.

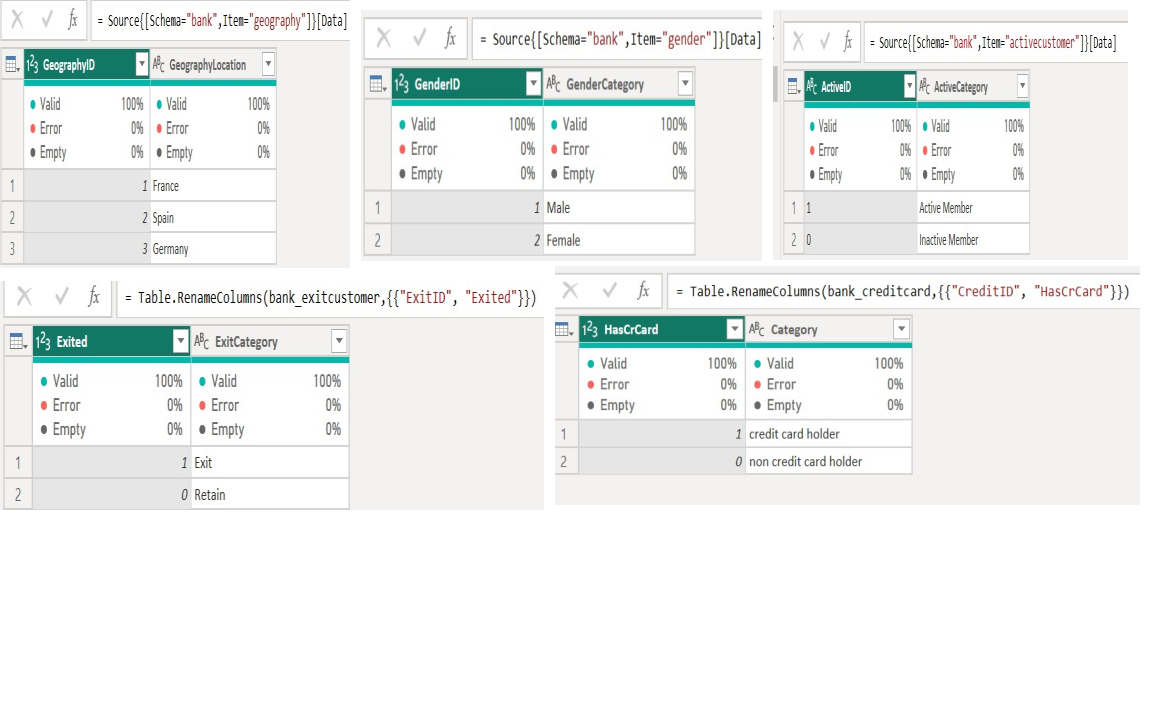
Output:



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







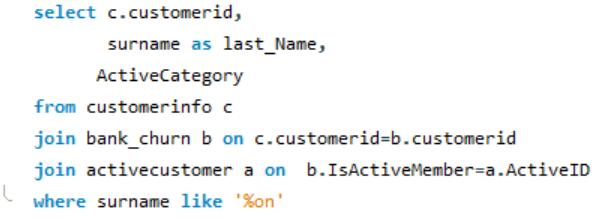
There are no missing values and null values in all the tables, column quality of the all tables is 100% and doesn’t have missing values or null values

However, in the event of missing values in any of the datasets we can handle them (using Power BI)

* Replacing the missing/null values with default values
* Remove the rows with missing/null values

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”.

Sql query:



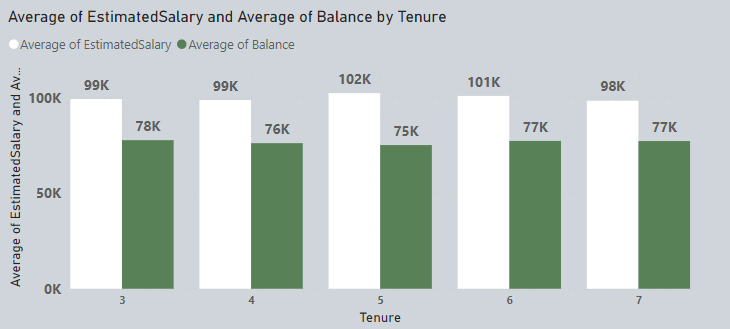
Output:



**Subjective Questions**

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?

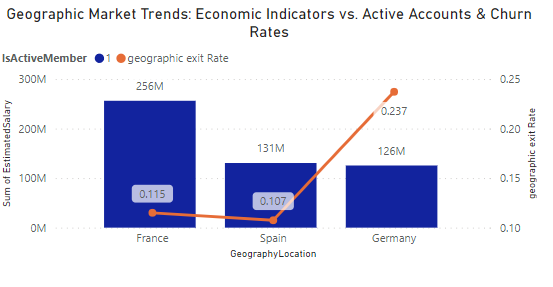
Looking at the average estimated salary helps us understand how much money customers make, giving us an idea of their income diversity. The average balance shows how much money is left in their accounts after spending. These numbers help us see how customers tend to spend their money. Interestingly, we don't see clear differences in spending habits between new customers and long-time ones. Both groups seem to have similar patterns in how much money they have left in their accounts.



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

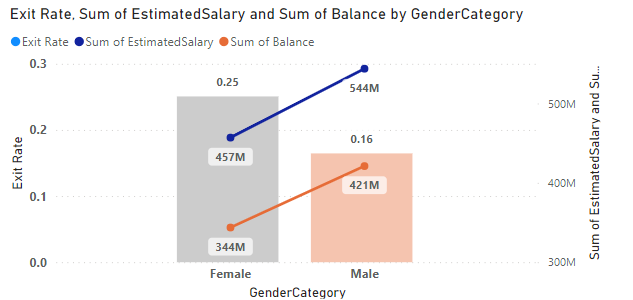
The dataset only includes the number of products bought by each customer, without details about specific products or combinations. Therefore, it's impossible to analyze product relationships or suggest cross-selling strategies using this data alone.

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

****

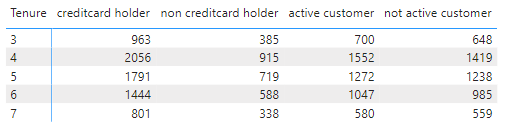
the chart shows that when the estimated salary is high in a location, there are more active accounts and fewer customers leaving (churn rate is low). Conversely, when the estimated salary is low, there are fewer active accounts and more customers leaving (churn rate is high). For example, France has a high estimated salary, which means it also has more active customers and fewer customers leaving (churn rate is low). On the other hand, Germany has a lower estimated salary, resulting in fewer active accounts and more customers leaving (churn rate is high).

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 data provided, we can see that more female categories are leaving the bank when compared to the male category. Also, female categories tend to have lower estimated salaries and maintain lower account balances compared to male category. Using these indicators, it seems that women pose a higher financial risk to the bank compared to men.

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



Based on the visual representation and the provided data, it can be inferred that customers who have a credit card play a significant role in generating interest income for the bank. This group of customers is likely to make a substantial contribution to the bank's revenue stream over time.

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?

To assess the impact of marketing campaigns on customer retention and acquisition within the dataset, I would adopt the following approach:

Segment customers: Divide customers into groups based on demographics, behavior, or other relevant factors to analyze campaign effectiveness for different segments.

Track campaign performance: Monitor campaign metrics like reach, engagement, and conversion rates over time to gauge effectiveness.

Establish a control group: Create a control group of customers not exposed to the campaign to compare against the exposed group, isolating the campaign's impact.

Measure changes: Compare the behavior, retention, and acquisition rates of the exposed group with the control group to assess the campaign's effectiveness in driving these metrics.

Analyze customer journey: Examine how customers interact with the campaign from awareness to conversion to understand its influence on retention and acquisition within each segment.

Collect additional data: Gather supplementary data such as customer feedback and detailed campaign performance metrics to gain deeper insights into campaign effectiveness. Historical data on previous marketing campaigns and their outcomes to identify trends and patterns.

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

It's difficult to fully grasp why customers leave using only limited data such as credit scores and membership status. While these factors offer some insight into customer behavior, they don't directly uncover the root causes of churn.

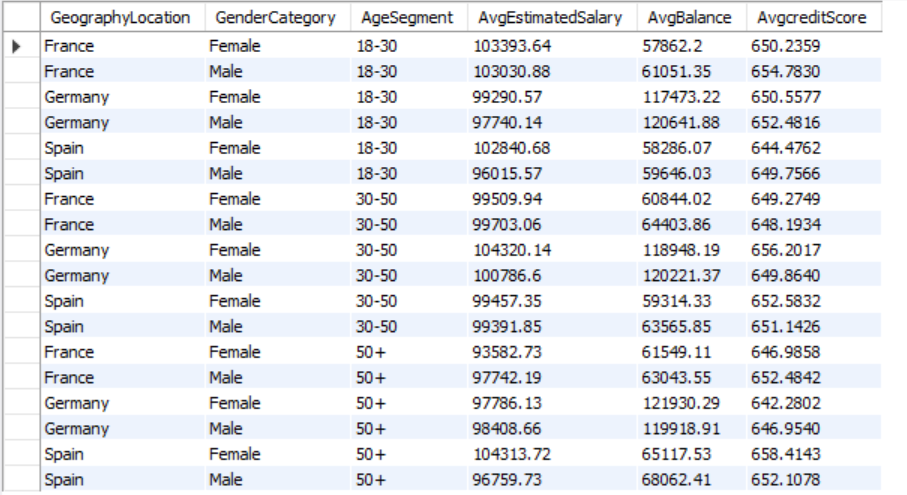
Customer attrition is influenced by a variety of factors, encompassing satisfaction with products/services, experiences with customer service, competitive dynamics, and external factors.

For a thorough understanding and effective management of customer churn, it's essential to collect direct feedback from customers through surveys and analyze their interactions in detail. Enhancing existing data with additional sources such as customer feedback can yield valuable insights into the reasons behind customer departure and strategies for reducing churn.

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

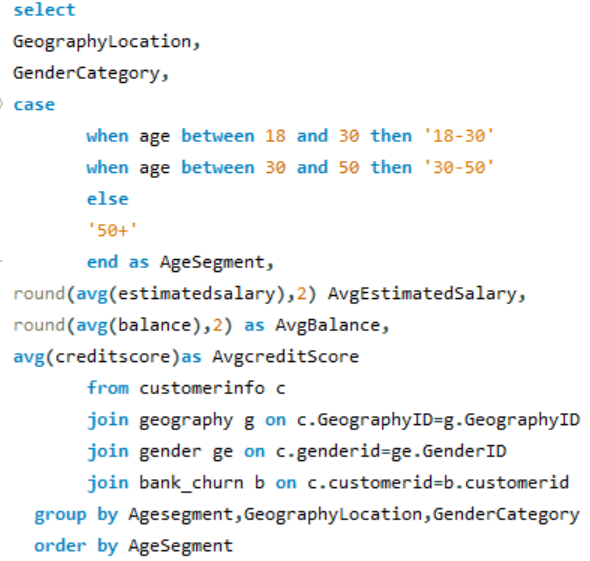
Predicting customer churn using only the provided attributes presents challenges. The dataset shows cases where even loyal and active customers have left the bank. Additionally, factors like the number of products and estimated salary don't consistently explain customer attrition. To accurately forecast churn, it's crucial to gather direct feedback from departing customers. This feedback offers insights into whether these attributes significantly contribute to predicting churn. Simply put, relying solely on numbers like tenure or product count isn't enough to understand why customers leave. We need to delve deeper into the reasons behind their decisions. That's where direct feedback becomes indispensable for refining our churn prediction models and enhancing customer retention strategies.

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



I've analyzed our customer data focusing on demographic information and account details. I've segmented our customers by age groups (18-30, 30-50, and 50+), and within each age group, I've looked at the average estimated salary, average account balance, and average credit score for each gender category and location. This breakdown provides us with valuable insights into the financial behavior and characteristics of our customer segments.

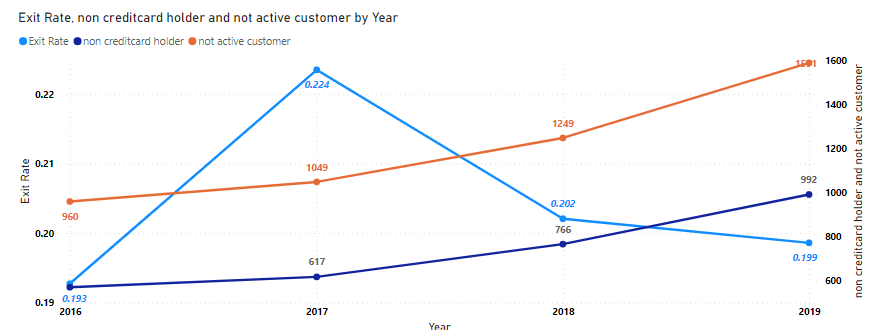
SQL query:



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?

* Our analysis of customer churn hits a roadblock when it comes to visualization. While conditional formatting helps pinpoint individual customers at risk, it's not effective for charts. These charts are crucial for understanding how factors like location and gender affect churn rates.
* Unfortunately, there's another gap in our analysis. Data on credit card rewards, a key element in assessing customer retention, is missing. This information is vital for getting a complete picture of why customers churn.

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 are the different strategies that can be used to decrease the churn rate.



Based on above analysis, I observed that the churn rate was notably high in 2017 but has shown a slight decrease since then. Additionally, we've noticed a consistent increase in the number of inactive customers and non-credit card holders each year. This trend could contribute to the rising churn rates among customers.

strategies that can be used to decrease the churn rate.

1. Improve Customer Engagement:

* Offer personalized communication and targeted marketing campaigns to re-engage inactive customers.
* Provide proactive customer support and assistance to address any issues or concerns promptly.

2. Promote Credit Card Usage:

* Encourage non-credit card holders to apply for credit cards by highlighting the benefits and rewards associated with card usage.
* Provide special promotions or discounts for credit card transactions to incentivize usage and increase customer retention.

3. Implement Customer Feedback Mechanisms:

* Solicit feedback from customers through surveys or feedback forms to identify pinpoints and areas for improvement.
* Use customer feedback to make data-driven decisions and prioritize initiatives that address customer needs and preferences.

4. Build Customer Loyalty Programs:

* Create loyalty programs that reward customers for their continued business and incentivize them to stay with the bank.
* Offer exclusive benefits or discounts for loyal customers to increase retention and encourage long-term relationships.

By implementing these strategies, the bank can work towards reducing churn rates, improving customer satisfaction, and fostering long-term relationships with its customers. It's essential to continuously monitor the effectiveness of these initiatives and adapt them as needed to ensure ongoing success in reducing churn and increasing customer retention.

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

I would start by conducting exploratory data analysis to understand the dataset's structure, variables, and distributions. Then, I would identify key metrics such as churn rate, customer demographics, active and non-active accounts, and financial indicators. Using SQL queries, Excel, or Power BI, I would analyze trends, correlations, and patterns in the data. For instance, I'd use SQL to calculate churn rates by segment, examine customer demographics, and identify correlations between variables. In Excel, I'd create pivot tables and charts to visualize trends and relationships. In Power BI, I'd build interactive dashboards to explore data visually and uncover insights. By leveraging these tools and techniques, I'd aim to identify actionable insights related to customer behavior, product affinity, geographic trends, risk factors, and marketing effectiveness.

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

In Power BI, we can easily rename a column by right-clicking on its name and selecting the option to rename it.

In SQL

ALTER TABLE bank\_churn

RENAME COLUMN HasCrCard TO Has\_creditcard;

Or

select hascrcard as Has\_creditcard from (

select \* from bank\_churn) as res;