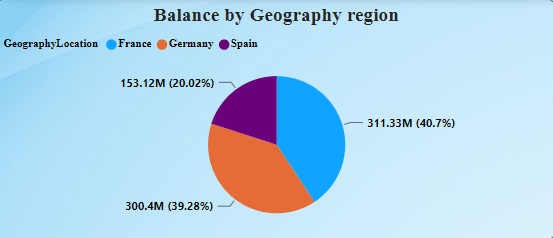
**Objective Questions:**

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

ANS- The numbers of customers account in France, Germany and Spain are so high, So, their account balances are also high. 

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

ANS- SELECT CustomerId, EstimatedSalary,Surname

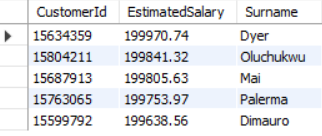
FROM customerinfo

WHERE extract(QUARTER from BankDOJ) = 4

ORDER BY EstimatedSalary DESC

limit 5;

OUTPUT:-



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

ANS- From the Query, the customers who have credit card mostly as their number of products. = avg(NO of product) = 1.5314

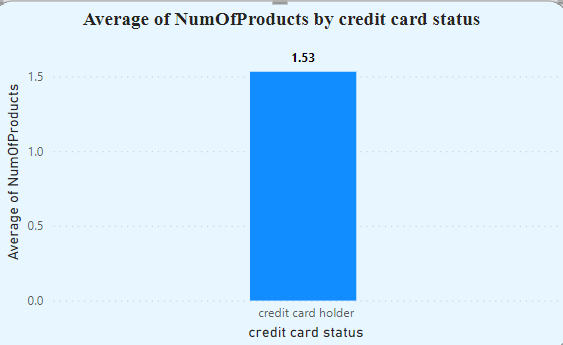
SELECT AVG(NumOfProducts) AS AvgProducts

FROM bank\_churn

WHERE HasCrCard = 1;

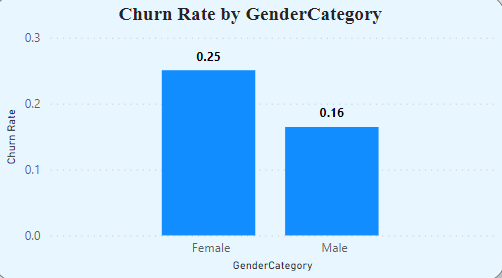
OUTPUT:-





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

ANS:-The churn rate for female is 0.25 while for males it stands at 0.16



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

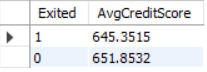
ANS:= The Customers who have left the bank possess an average credit score of 645.35, whereas those who have remained have an average credit score of 651.85.

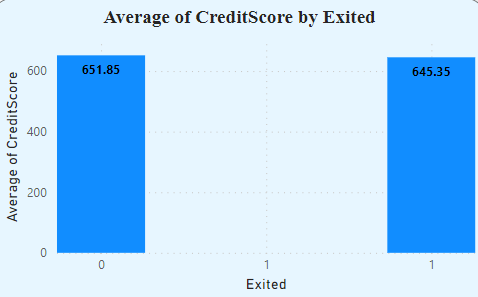
SELECT Exited, AVG(CreditScore) AS AvgCreditScore

FROM bank\_churn

GROUP BY Exited;

OUTPUT:=





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

ANS:- Female has highest average estimated salary of around 100601.54 even though they are having a smaller number of active accounts than the Male= 99664.57

SELECT

GenderID,

AVG(c.EstimatedSalary) AS avg\_Salary,

Count(distinct case when IsActiveMember=1 then c.customerid else null end ) AS number\_of\_activeaccount

FROM activecustomer a

JOIN

bank\_churn b ON b.IsActiveMember = a.ActiveID

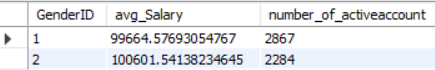
JOIN

customerinfo c ON b.CustomerId = c.CustomerId

GROUP BY

GenderID;

OUTPUT:-



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

ANS-

The customer’s credit score has been segmented as :

Excellent: 800–850,

Very Good: 740–799,

Good: 670–739,

Fair: 580–669,

Poor: 300–579.

The customer’s having Fair as their credit score worthiness has the highest exit rate is Poor (0.2202).

SELECT

CASE

WHEN CreditScore >= 800 THEN 'Excellent'

WHEN CreditScore >= 740 AND CreditScore < 800 THEN 'Very Good'

WHEN CreditScore >= 670 AND CreditScore < 740 THEN 'Good'

WHEN CreditScore >= 580 AND CreditScore < 670 THEN 'Fair'

WHEN CreditScore < 580 THEN 'Poor'

ELSE 'Unknown'

END AS CreditScoreSegment,

AVG(Exited) AS ExitRate

FROM bank\_churn

GROUP BY CreditScoreSegment

ORDER BY ExitRate DESC

LIMIT 1;

OUTPUT:-



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

ANS- Among geographic locations, France stands out with the highest count of active customers, totaling 1477, who have a tenure exceeding 5 years.

SELECT GeographyID,COUNT(c.CustomerId)AS CUST

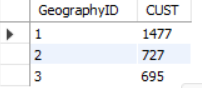
from customerinfo c

JOIN bank\_churn B ON c.CustomerId = B.CustomerId

WHERE IsActiveMember=1 AND Tenure>=5

GROUP BY GeographyID

ORDER BY CUST DESC;



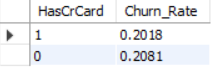
1. What is the impact of having a credit card on customer churn, based on the available data?

ANS:- The customers with credit card who remained in the bank is higher than the exited people. So, there is no impact of having a credit card on customer churn.

SELECT HasCrCard, AVG(Exited) AS Churn\_Rate

FROM bank\_churn

GROUP BY HasCrCard;



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

ANS- The most common number of products used by exited customers is 1.

SELECT NumOfProducts AS Most\_Common\_NumOfProducts,

COUNT(\*) AS Frequency

FROM bank\_churn

WHERE Exited = 1

GROUP BY NumOfProducts

ORDER BY Frequency DESC

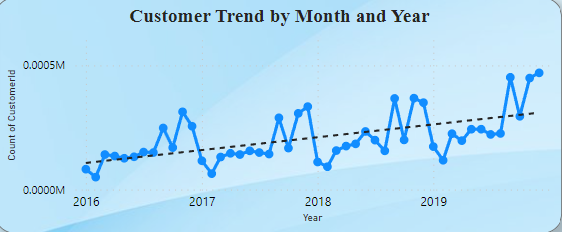
LIMIT 1;

OUTPUT:-



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.

ANS- Based on the provided dataset, we observed that the exit rate of customers is highest in the most recent year, 2019. Additionally, it is noteworthy that from 2016 to 2019, there's a trend of losing customers at a higher rate compared to preceding years, indicating an outlier period of customer attrition.



SELECT

YEAR(BankDOJ) AS Year,

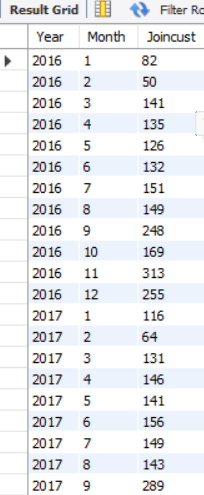
MONTH(BankDOJ) AS Month,

COUNT(\*) AS Joincust

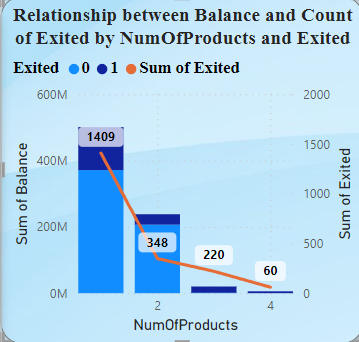
FROM customerinfo

GROUP BY YEAR(BankDOJ), MONTH(BankDOJ)

ORDER BY Year, Month;



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



ANS-

**For customers who had 1 product:**

**• Total balance: 129,668,607.**

**• Number of customers who exited: 1409**

**For customers who had 2 products:**

**• Total balance: 31,407,820.29**

**• Number of customers who exited: 348**

**For customers who had 3 products:**

**• Total balance: 18,887,679.16**

**• Number of customers who exited: 220**

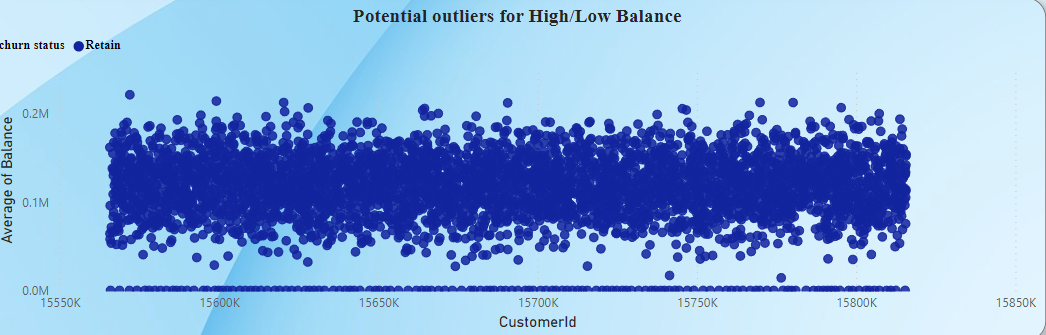
**For customers who had 4 products:**

**• Total balance: 5,623,988.10**

**• Number of customers who exited: 60**

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

ANS- The Most customers who do not have four products are typically retained by the bank.



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

ANS- As mention above, I utilized seven tables to generate a dashboard and established several measures, including the average customer balance, average credit score, average estimated salary, churn\_rate as well as the maximum estimated salary etc.

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)

ANS- By using Rank Function, I have ranked the gender-wise average income of males and females in each geography id

SELECT GeographyID,GenderID,

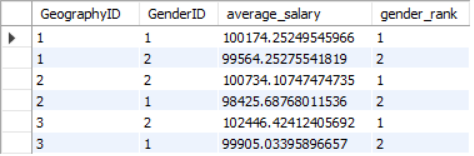
AVG(EstimatedSalary) AS average\_salary,

RANK() OVER (PARTITION BY GeographyID ORDER BY AVG(EstimatedSalary) DESC) AS gender\_rank

FROM customerinfo

GROUP BY GeographyID, GenderID;

OUTPUT:-



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+).

ANS- The average tenure of he customers with the age bracket of 30-50 is 4.8899which is the highest average tenure followed by the customers with the age bracket of 50+ is 4.8330, the customers with the age bracket of 18-30 is 4.7770.

SELECT

CASE

WHEN ci.Age BETWEEN 18 AND 30 THEN '18-30'

WHEN ci.Age BETWEEN 30 AND 50 THEN '30-50'

ELSE '50+'

END AS Age\_Bracket,

AVG(bc.Tenure) AS Avg\_Tenure

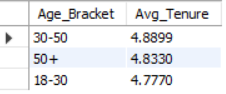
FROM customerinfo ci

INNER JOIN bank\_churn bc ON ci.CustomerId = bc.CustomerId

WHERE bc.Exited = 1

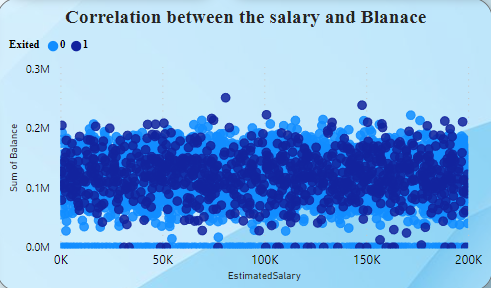
GROUP BY Age\_Bracket;

OUTPUT:-



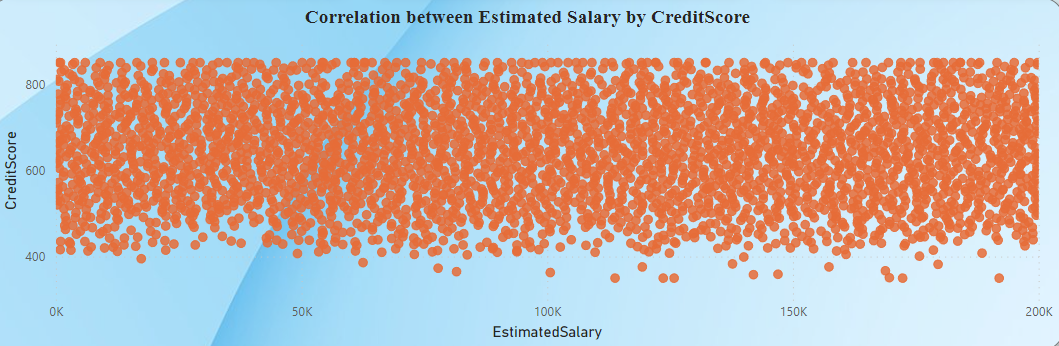
1. Is there any direct correlation between salary and the balance of the customers? And is it different for people who have exited or not?

ANS- From the below picture, we can get the conclusion as if there is increase in salary for both the retained customers and the exited customers, but there is no major changes in balance. So, there is no direct correlation between both salary and balance.



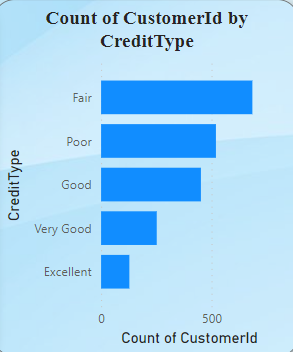
1. Is there any correlation between the salary and the Credit score of customers?

ANS- We use scatter plot chart to show mentioned Below, we can understand that there is correlation between salary and credit score



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

ANS-



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.

ANS- In our dataset, there are 1,261 customers aged 50 or above, out of which 874 have a credit card. Additionally, there are 1,968 customers aged between 18 and 30, with 1,400 of them having a credit card. These groups represent all the age brackets with fewer credit card holders compared to the average number of credit cards per bucket.

SELECT Age, COUNT(CustomerId) AS NumOfCust,

COUNT(HasCrCard) AS NumOfCrCard

FROM customerinfo

INNER JOIN bank\_churn USING(CustomerId) WHERE

HasCrCard = 1 GROUP BY Age;

WITH age\_num\_cr\_card AS

(SELECT Age,COUNT(HasCrCard) AS NumOfCrCard

FROM customerinfo

INNER JOIN bank\_churn USING(CustomerId)

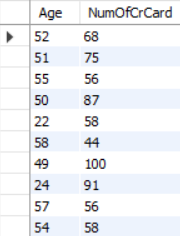
WHERE HasCrCard = 1 GROUP BY Age),

avg\_num\_cr\_card AS (SELECT AVG(NumOfCrCard) AS avg\_cr\_card FROM age\_num\_cr\_card)

SELECT Age,NumOfCrCard

FROM age\_num\_cr\_card,avg\_num\_cr\_card

WHERE NumOfCrCard < avg\_cr\_card;

OUTPUT: 

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

ANS-

* Germany got the first rank because of having 814 number of people and average balance of 120361.08 who exited the bank.
* France got the second rank because of having 810 number of people and average balance of 71192.8 who exited the bank.
* Spain got the third rank because of having 413 number of people and average balance of 72513.35 who exited the bank.

SELECT GeographyID,

AVG(Balance) AS AvgBalance,

COUNT(CustomerId) AS Totalcust,

RANK() OVER(ORDER BY COUNT(CustomerId) DESC) AS Ranking

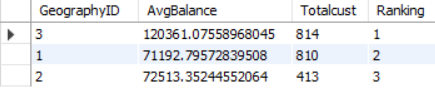
FROM bank\_churn INNER JOIN customerinfo USING(CustomerId) INNER JOIN geography USING(GeographyID)

WHERE Exited = 1

GROUP BY GeographyID

ORDER BY Totalcust DESC;

OUTPUT-



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”.

ANS- SELECT

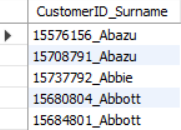
CONCAT(CI.CustomerID, '\_', CI.Surname) AS CustomerID\_Surname

FROM customerinfo CI

JOIN

customerinfo CI2 ON CI.CustomerID = CI2.CustomerID AND CI.Surname = CI2.Surname;

**OUTPUT-**

****

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

ANS- SELECT

bc.\*,

(SELECT ExitCategory

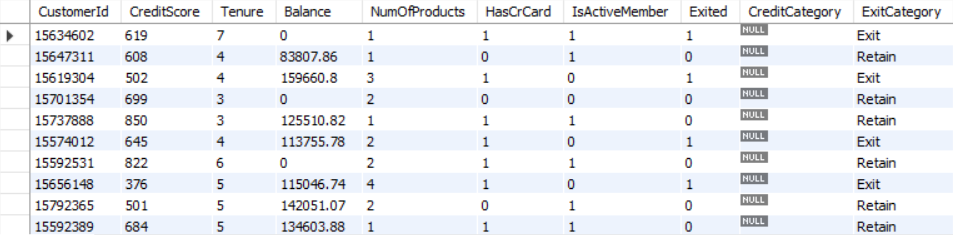
FROM exitcustomer ec

WHERE ec.ExitID = bc.Exited) AS ExitCategory

FROM

Bank\_Churn bc;

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?

ANS- In Power BI, I utilized the Power Query Editor tool to address missing data points within the dataset. Employing this tool, I identified and rectified instances where data was absent or contained empty cells. By leveraging the functionalities of the editor, I systematically replaced these missing values. This involved either assigning a specific predetermined value to the empty cells or dynamically filling them with the value from the preceding non-blank cell. This process ensured that our subsequent analysis was conducted on a dataset devoid of any gaps, thereby enhancing the reliability and accuracy of our insights.

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”.

ANS:-

SELECT

CustomerId,

Surname,

CASE

WHEN IsActiveMember = 1 THEN 'Active'

ELSE 'Inactive'

END AS Active\_Status

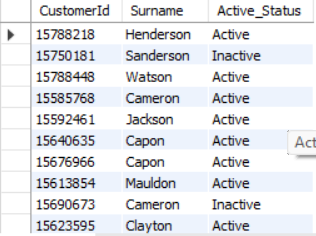
FROM

customerinfo JOIN bank\_churn using ( CustomerId)

WHERE

Surname LIKE '%on';

OUTPUT:-



**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? We can solve this with the tenure

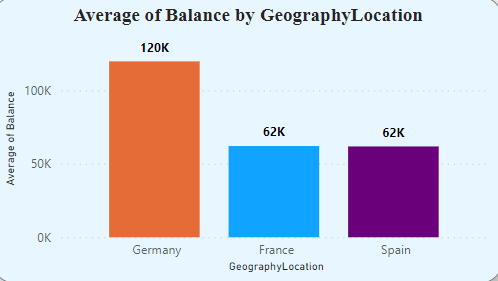
ANS- Observing the data, it's apparent that long-term retained customers have lower spending habits compared to new customers, which correlates with lower credit scores among the former group.

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

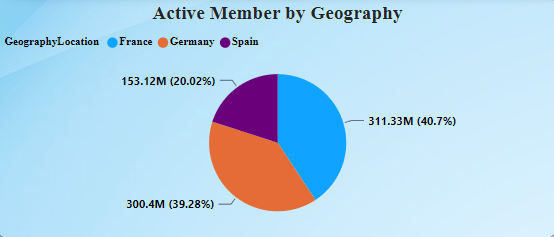
ANS- Without access to data on specific bank products within the datasets, it's challenging to determine which factors influence cross-selling strategies.

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

ANS- According to the stacked column chart visualization, France exhibits the lowest number of active accounts and the second-highest churn rates among the depicted countries. Conversely, Germany boasts the highest number of accounts but also registers the highest churn rates. Spain, on the other hand, demonstrates churn rates equivalent to those of France and holds the second smallest number of active accounts in comparison to the other countries represented.

ANS- 

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

ANS-

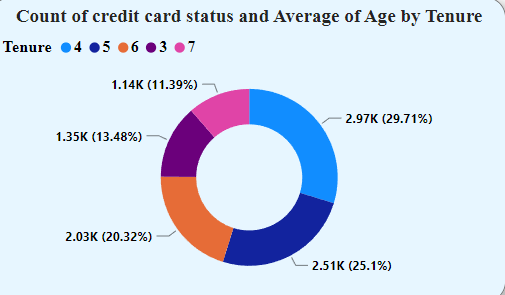
The reason why choose demographic segment is because age factor is common for both male and female which has high impact on bank’s financial risk. From the above visualizations, middle age bracket people have the highest rate of churn. At the age bracket between 30 and 50, all the people will try to put the investments, but at the time customers are exiting from the bank. It could cause high financial risk to the bank.

Spain = 20.02%

France = 40.7%

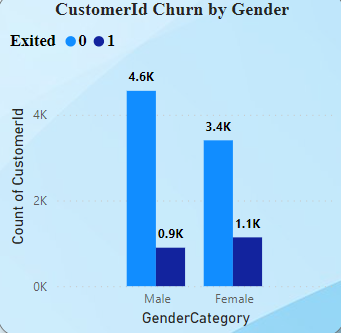
Spain = 39.28%

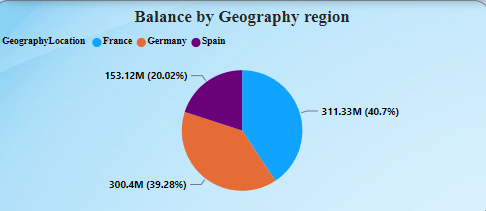
1. Customer Tenure Value Forecast: How would you use the available data to model and predict the lifetime (tenure) value in the bank of different customer segments?

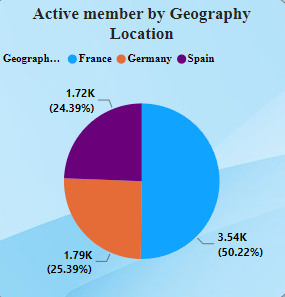
ANS- 

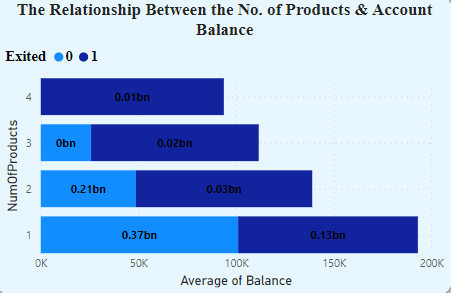
1. Marketing Campaign Effectiveness: How could you assess the impact of marketing campaigns on customer retention and acquisition within the dataset? What extra information would you need to solve this?

ANS- It appears that over the past year, the rate of acquiring new customers has surpassed the rate of retaining existing customers, regardless of gender. In order to mitigate this issue, one strategy could involve providing special offers to existing customers, potentially in the form of increased credit scores. By incentivizing loyalty among current customers, it's likely to encourage them to stay longer. This approach not only aims to improve retention rates among existing customers but also intends to foster a sense of value and longevity among new customers, thereby contributing to sustained customer retention.

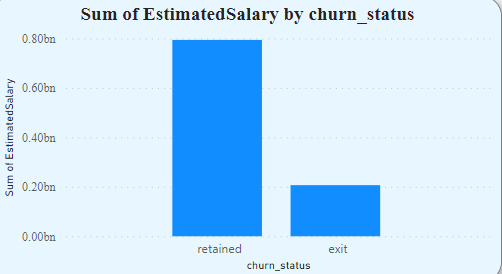




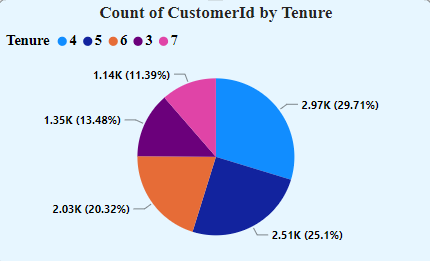


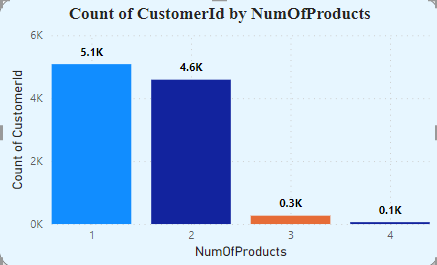


1. Customer Exit Reasons Exploration: Can you identify common characteristics or trends among customers who have exited that could explain their reasons for leaving?

ANS- After analyzing the provided visuals, it's evident that two key characteristics stand out among customers who have exited the bank: salary and credit score. A noticeable trend emerges where individuals with lower salaries tend to have lower credit scores. This observation suggests a potential correlation between financial stability, as indicated by salary, and creditworthiness, represented by credit score. The data implies that customers with lower salaries may struggle to maintain a satisfactory credit score, possibly leading them to consider exiting the bank. This insight underscores the importance of understanding the financial circumstances of customers and their impact on retention rates within the banking institution

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

ANS- Certainly, tenure, number of products, active membership status, and estimated salary play crucial roles in predicting customer churn. These factors collectively aid in assessing whether a customer can afford products within their estimated salary, manage payments within their tenure, and remain engaged as an active member of the bank.



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

**SELECT g.GeographyLocation AS Region,**

**COUNT(\*) AS NumCustomers,**

**MIN(b.Balance) AS MinBalance,**

**MAX(b.Balance) AS MaxBalance,**

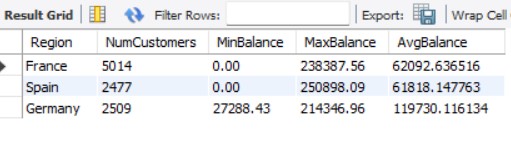
**AVG(b.Balance) AS AvgBalance**

**FROM bank\_churn b**

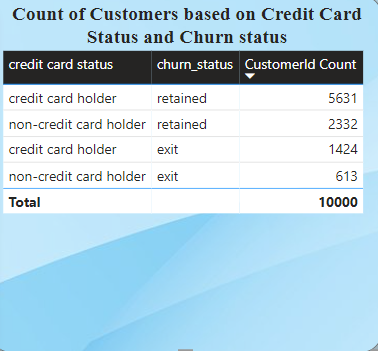
**JOIN customerinfo ci ON b.CustomerId = ci.CustomerId**

**JOIN geography g ON ci.GeographyID = g.GeographyID**

**GROUP BY g.GeographyLocation;**

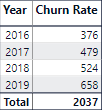
**OUTPUT:** 

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?

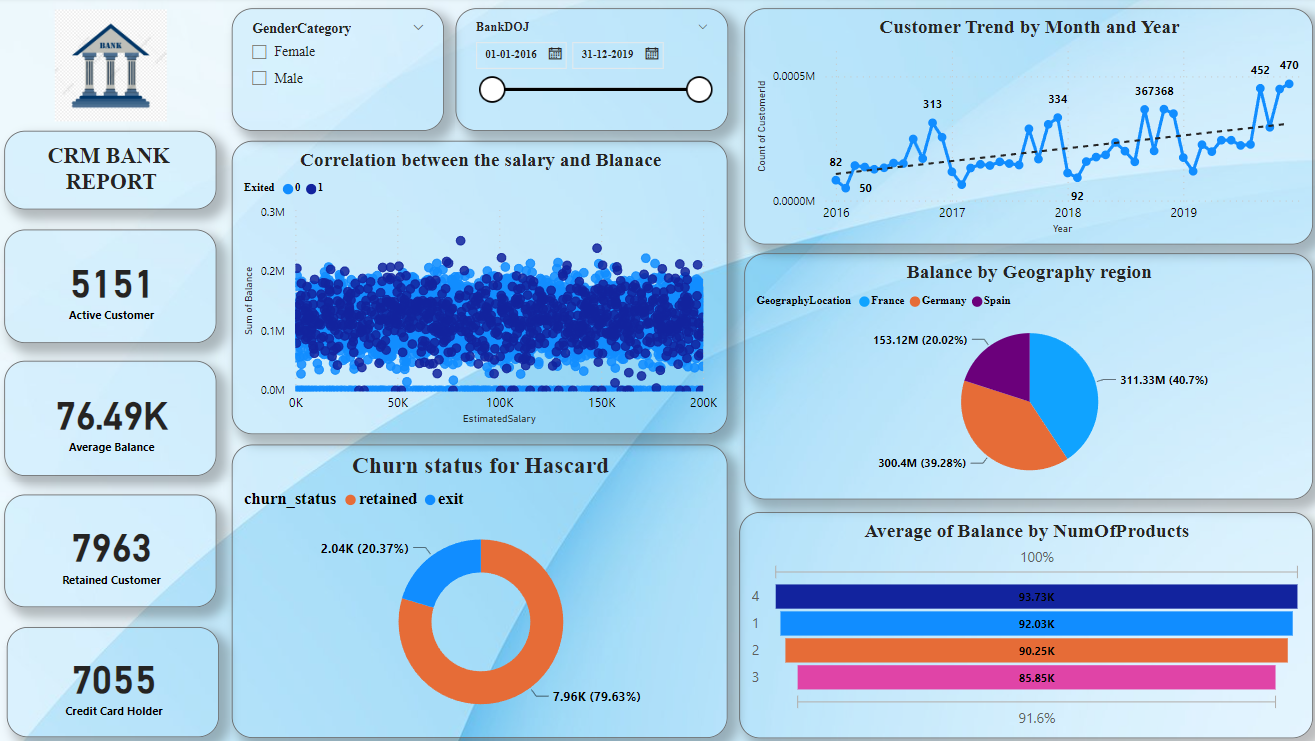
ANS- 

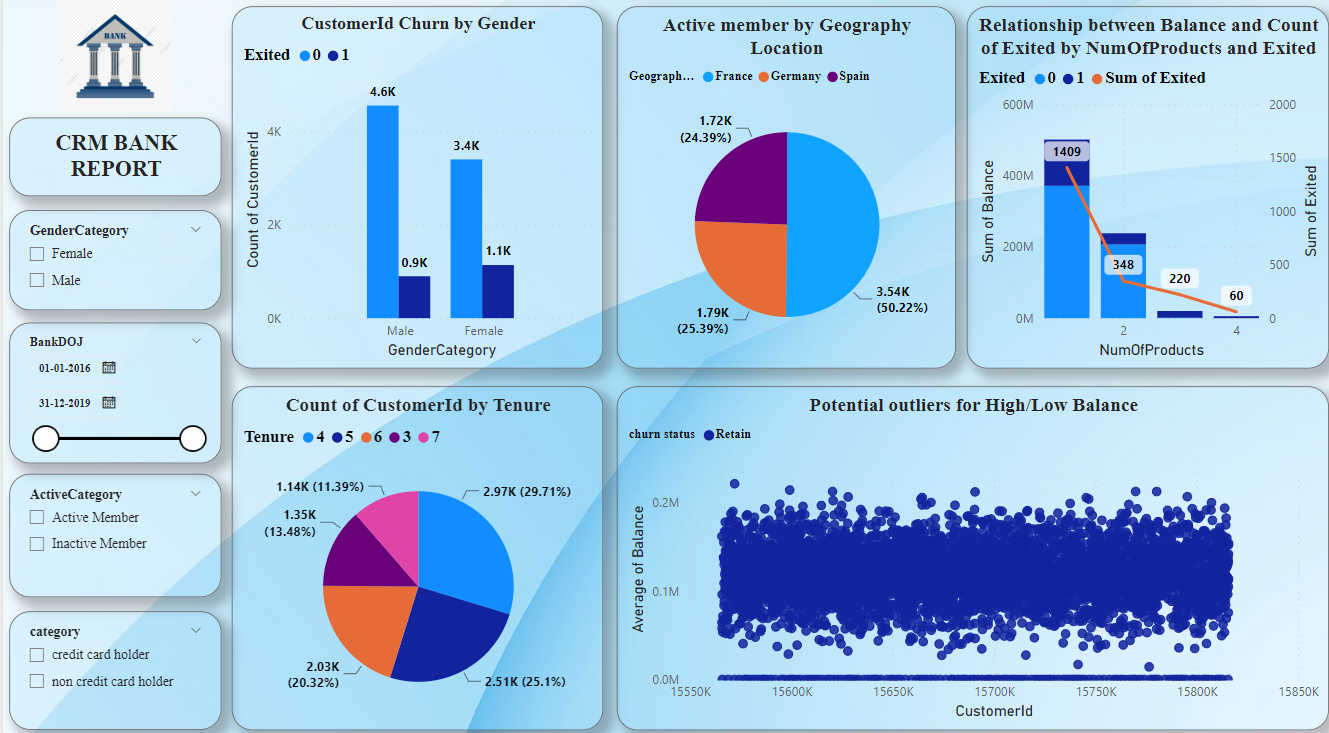
1. What is the current churn rate per year and overall as well in the bank? Can you suggest some insights to the bank about which kind of customers are more likely to churn and what different strategies can be used to decrease the churn rate?

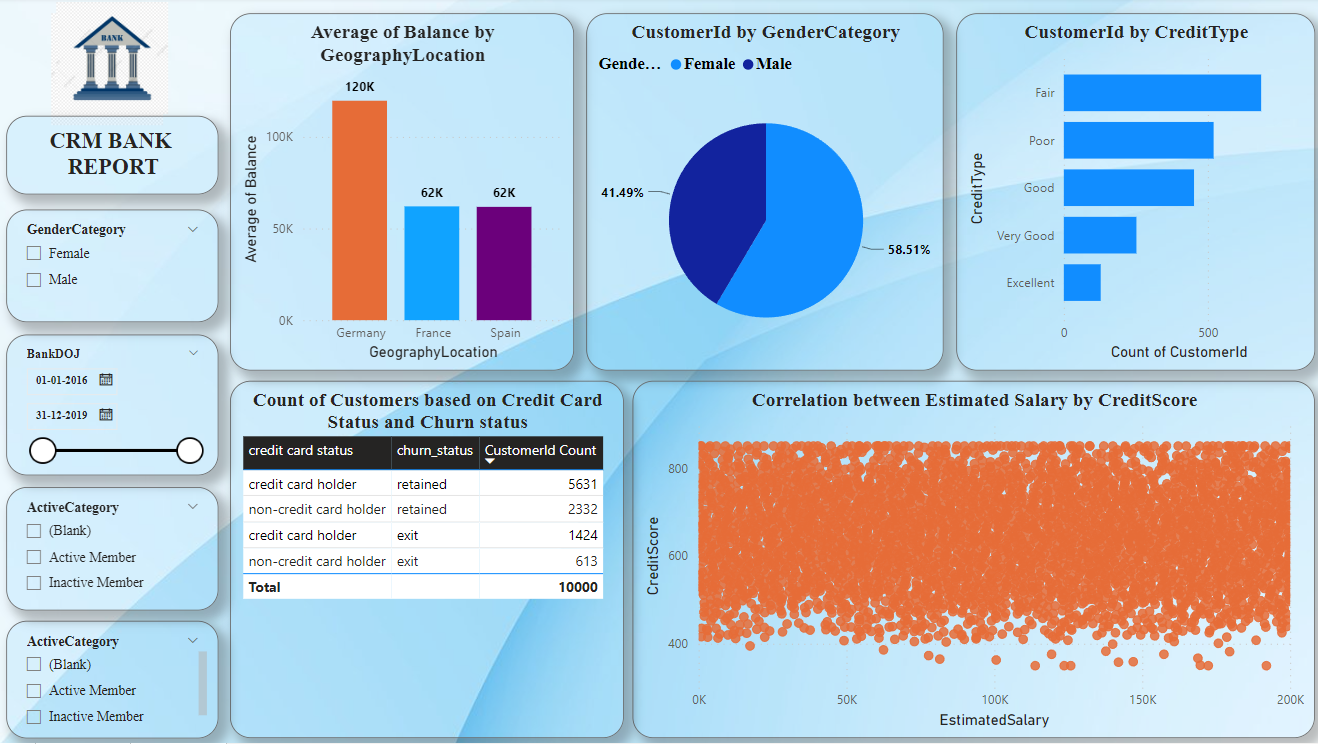
ANS- The current annual churn rate stands at 658 for the year 2019, contributing to a total churn rate of 2037. Notably, individuals aged between 30 and 50 exhibit the highest percentage of churn rate. To address this, offering an extended tenure period could provide these customers with additional time to manage their concerns effectively.



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

ANS- 





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

ANS –

**Understand the Database Schema:**

* Begin by examining the structure of the database. Identify tables, their columns, and the relationships between them.

**Data Profiling:**

* Conduct a data profiling exercise to understand the content, quality, and distribution of data within each table.
* Identify any data anomalies, such as missing values, duplicates, or outliers.

**Identify High-Impact Tables:**

* Determine which tables contain the most critical information for your project.
* Prioritize your analysis based on the significance of these tables to the overall project goals.

**Explore Data Relationships:**

* Use SQL queries to explore relationships between tables through joins and foreign key constraints.
* Understand how different entities in the database are related to each other.

**Aggregate and Summarize Data:**

* Aggregate data to generate summary statistics, such as counts, averages, or sums.
* This can help provide insights into overall trends and patterns in the data.

**Detect Trends and Patterns:**

* Analyze the data to identify any trends or patterns that may be present.
* Look for recurring themes or correlations between different attributes.
* Perform Descriptive Analysis:
* Use SQL functions to perform descriptive analysis, such as calculating percentiles, quartiles, or standard deviations.
* Understand the distribution of key variables within the dataset.

**Identify Data Quality Issues**:

* Use SQL queries to detect and address data quality issues, such as missing values, inconsistencies, or incorrect data types.
* Cleanse the data as necessary to ensure its accuracy and reliability.
* Generate Insights:
* Based on your analysis, generate insights and observations about the data.
* Consider how these insights align with the project objectives or potential business implications.

**Document Your Analysis:**

* Document your analysis process, including the SQL queries used, key findings, and any assumptions made.
* Provide context for your analysis and explain the significance of the insights uncovered.

**Iterate and Refine:**

* Review your analysis and iterate as necessary based on feedback or additional exploration.
* Refine your SQL queries or analysis techniques to uncover deeper insights or address any unanswered questions.

**Present Findings:**

* Present your findings to stakeholders, using visualizations or reports to communicate key insights effectively.
* Clearly articulate the implications of your analysis and any recommendations for further action.

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

ANS-

ALTER TABLE Bank\_Churn

RENAME COLUMN HasCrCard TO Has\_creditcard;