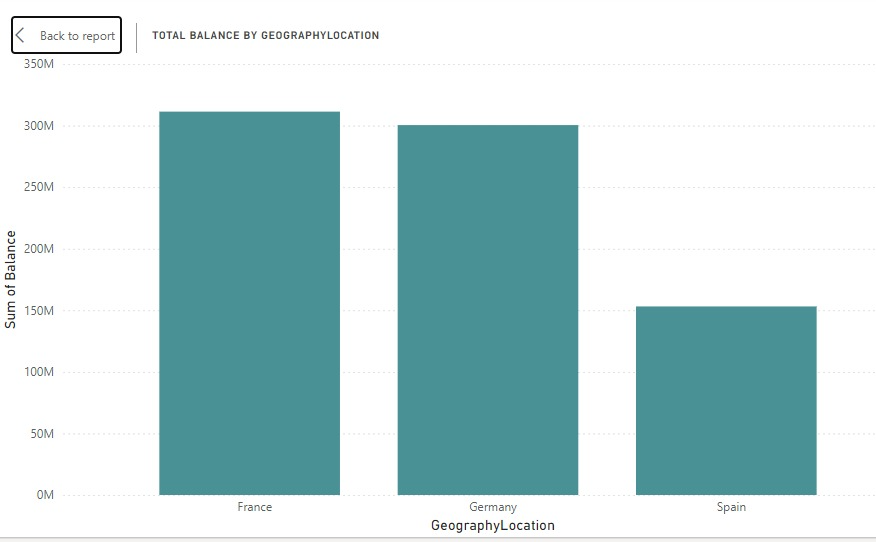
**Objective Questions**

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



The chart shows the distribution of account balance across different regions.

France has the highest account balance distribution followed by Germany.

Spain has the least account balance distribution.

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

This query can be used for identifying the customer id and salary of the top 5 customers with the highest estimated salary in the last quarter of the year.

**Query:**

SELECT c.CustomerId, c.Surname, c.EstimatedSalary,c.JoiningDate

FROM customerinfo c

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

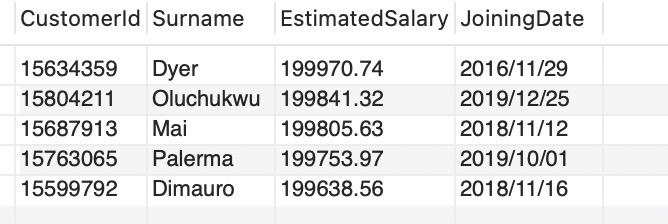
WHERE EXTRACT(QUARTER FROM c.JoiningDate) = 4

ORDER BY c.EstimatedSalary DESC

LIMIT 5;

**Query explanation:**

* Customerinfo is joined to bank\_churn on customer id to ensure relavant data from both tables.
* WHERE EXTRACT(QUARTER FROM c.JoiningDate) = 4 filters the data from the last quarter based on the joiningdate.
* Limit 5. It limits the results to the top 5 customers.



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

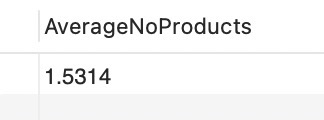
The average number of products used by the customers who have a credit card is 1.5314.

**Query:**

SELECT AVG(NumOfProducts) AS AverageNoProducts

FROM bank\_churn

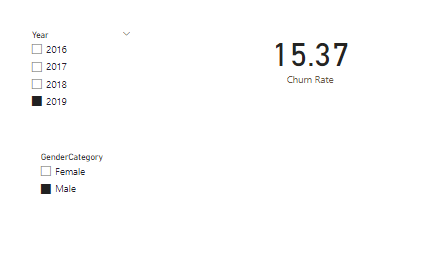
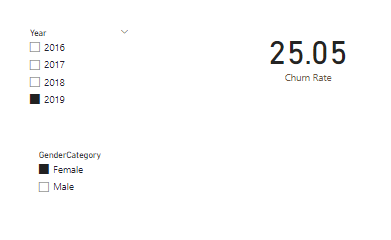
WHERE HasCrcard = 1;



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

The churn rate for the most recent year per gender can be developed by an interactive slicer. Here two slicers have been created one for the year and the another for Gender category. By using these slicers, we can filter the churn rate. A calculated measure,

(ChurnRate = DIVIDE([LostCustomers],[TotalCustomers])\*100) determines the churn rate.

****

**Conclusion:**

Churn rate for Males in 2019 is 15.37.

Churn rate for Females in 2019 is 25.05.

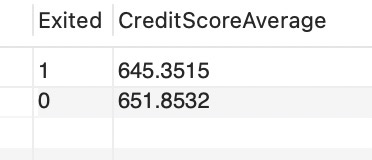
1. **Compare the average credit score of customers who have exited and those who remain. (SQL)**

**Query:**

SELECT Exited, AVG(CreditScore) AS CreditScoreAverage

FROM bank\_churn

GROUP BY Exited;



The average credit score for the people who have exited is 645.35.

The average credit score for the people who have remained is 651.85.

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

The query first finds the number of active accounts among the customers and then segregates it into Gender category. Now we find the total number of active accounts and average estimated salary for each Gender category.

**Query:**

WITH ActiveAccounts AS (

SELECT CustomerId, COUNT(\*) AS ActiveAccounts

FROM bank\_churn

WHERE IsActiveMember = 1

GROUP BY CustomerId

)

SELECT

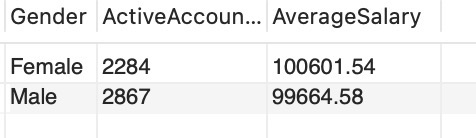
CASE WHEN ci.GenderID = 1 THEN 'Male' ELSE 'Female' END AS Gender,

COUNT(aa.CustomerId) AS ActiveAccounts, AVG(EstimatedSalary) AS AverageSalary

FROM customerinfo ci LEFT JOIN ActiveAccounts aa ON ci.CustomerId = aa.CustomerId

GROUP BY Gender

ORDER BY AverageSalary DESC;



**Conclusion:** Male have higher active accounts and the average salary of the females who have active accounts is slightly higher than the males.

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

To identify the customers with the highest exit rate based on segments we first divide the customer data into segment.

* The customers who have a credit score over 800 as “Excellent”
* The customers with credit score over 740 and less than 799 as “Very Good”.
* The customers with credit score over 670 and less than 739 as “Good”
* The customers with a credit score between 580 and 669 as “Fair” and the rest as “Poor”

Now we find the segment which has the highest exit rate.

**Query:**

WITH CreditScoreSegment AS (

SELECT CustomerId, IsActiveMember,

CASE

WHEN CreditScore BETWEEN 800 AND 850 THEN 'Excellent'

WHEN CreditScore BETWEEN 740 AND 799 THEN 'Very Good'

WHEN CreditScore BETWEEN 670 AND 739 THEN 'Good'

WHEN CreditScore BETWEEN 580 AND 669 THEN 'Fair'

ELSE 'Poor'

END AS CreditScoreSegment

FROM bank\_churn

)

SELECT CreditScoreSegment,

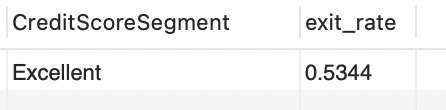
AVG(CASE WHEN IsActiveMember = 0 THEN 0 ELSE 1 END) AS exit\_rate

FROM CreditScoreSegment

GROUP BY CreditScoreSegment

ORDER BY exit\_rate DESC

LIMIT 1;



**Conclusion:** The “Excellent” segment has the highest exit rate among all the other segments.

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

The geographic region with the highest active customers with a tenure greater than 5 years is France with 1575 active customers.

**Query:**

SELECT g.GeographyLocation, COUNT(b.CustomerId) AS active\_customers

FROM geography g

INNER JOIN customerinfo c ON g.GeographyID = c.GeographyID

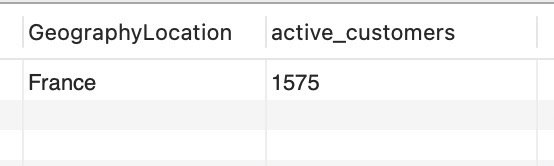
INNER JOIN bank\_churn b ON c.CustomerId = b.CustomerId

WHERE b.Tenure > 5

GROUP BY g.GeographyLocation

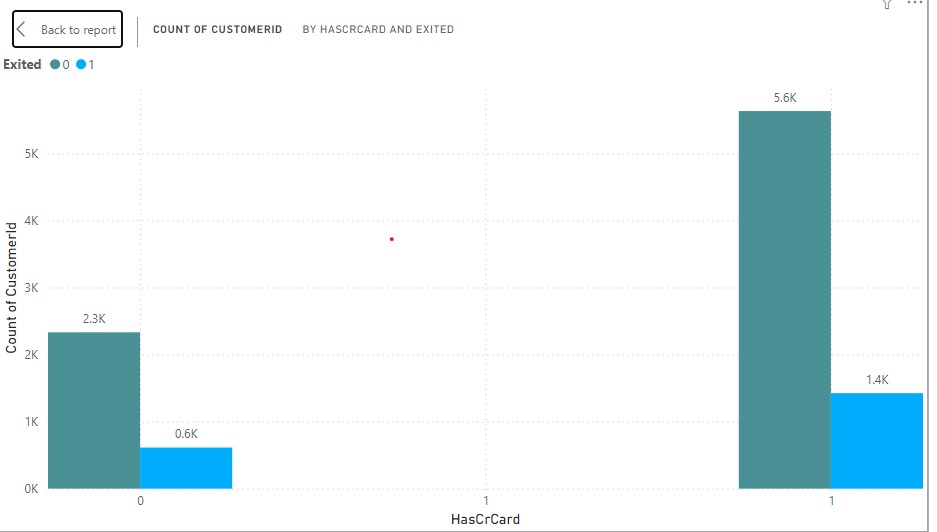
ORDER BY active\_customers DESC

LIMIT 1;



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

The provided graph compares customer churn (exit) rates between customers who do not have a credit card (HasCrCard = 0) and those who have a credit card (HasCrCard = 1).



* It appears that more customers with a credit card (HasCrCard = 1) have churned compared to those without a credit card.
* In comparison less people who don’t have a credit card have churned in comparison to the people who have a credit card.

**Conclusion:**

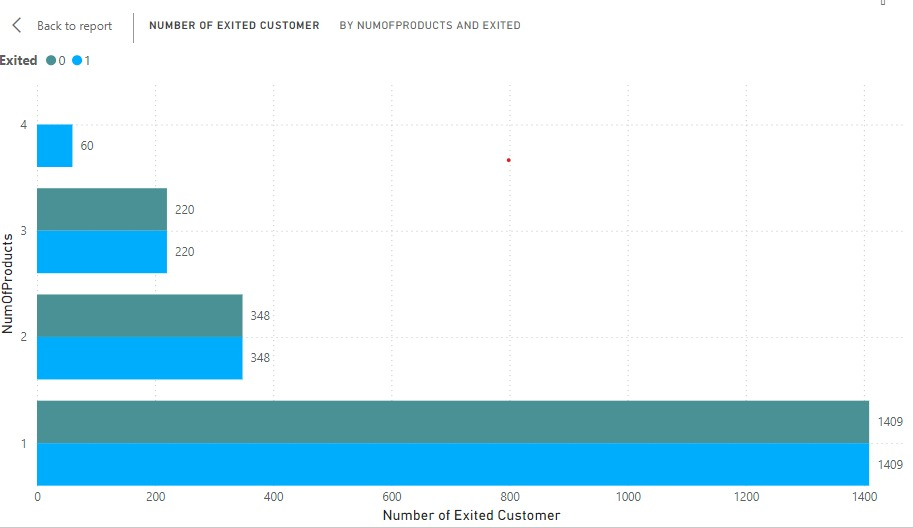
**Yes**, having a credit card seems to have a noticeable impact on customer churn. Customers with a credit card are more likely to churn.

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

The below chart provides the comparison between the number of products and the customers who have exited.

**Insights:**

* Product 1: The number of customers who have exited that use product 1 are 1409.
* Product 2: The number of customers who have exited that use product 2 are 348.
* Product 1: The number of customers who have exited that use product 3 are 220.
* Product 1: The number of customers who have exited that use product 4 are 60.

****

**Conclusion:** The customers who use 3 or 4 products are less likely to churn.

The customer who uses just 1 product are the customers who have churned the most. So, focusing on those customer base can help to decrease the churn rate.

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.**

The chart is a time series graph showing the count of customers joining the bank over time, likely year and month.

**Conclusion:**

* The customer joining trend appears to be increasing over time. This indicates a positive growth in customer acquisition.
* The number of customers joining the bank follow a seasonal pattern.
* Less number of customers join in the month of January.
* Highest number of customers join the month of July.

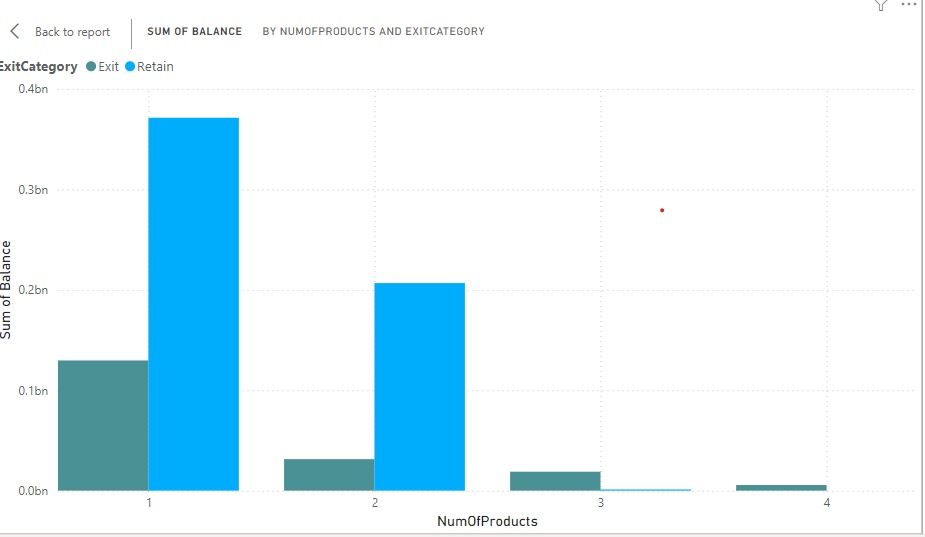
****

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

The chart describes the distribution of the number of products used by customers who have exited. The x-axis represents the number of products used, and the y-axis represents the count of customers who used that many products.

**Conclusion:**

* The most common number of products used by exiting customers is 1. This suggests that the customers who use least number of products are more likely to churn.
* As the number of products used increases. This suggests that customers who churned tend to have fewer products compared to active customers.
* The customers who have used 4 products have less balance.

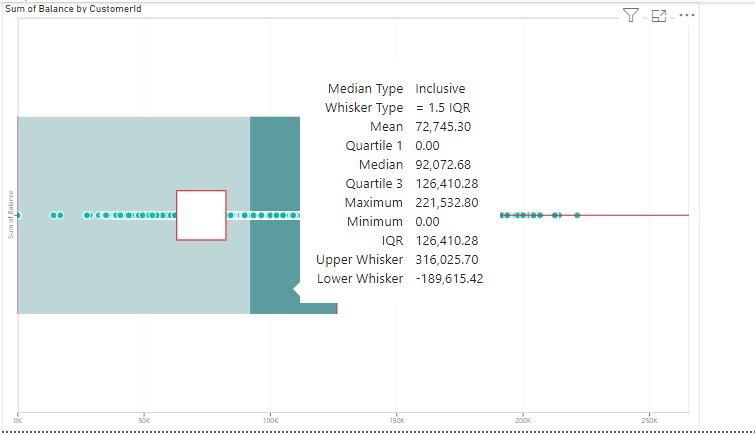
****

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

We can use Box plots or scatter plots are both useful tools for detecting outliers, but they offer different perspectives.

A **box plot** highlights outliers by visualizing the spread of data through quartiles. The whiskers represent data points within 1.5 times the interquartile range (IQR), and any points beyond them are considered potential outliers. This allows you to quickly identify customers with unusually high or low balances. The box itself shows the middle 50% of the data, with the line inside it indicating the median.

Outliers beyond the whiskers often represent customers who deviate significantly from the majority, either with very high or low balances, which could pose a risk to the bank.

****

**Detecting Outliers:**

The data points which are beyond the whiskers represent the outliers.

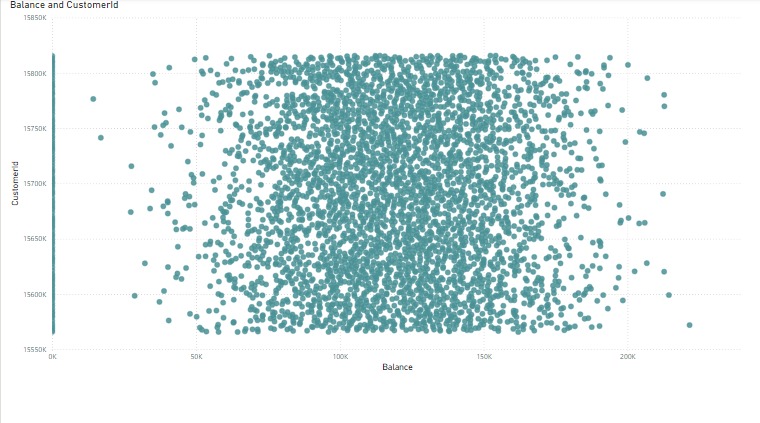
As per the dataset, The customers which have a balance greater than 316,025 or lower than -189,615 could act as outliers.

They could affect the data of the bank.

**Zero outliers** detected as all the data points are within the whiskers.

**Scatter plot** provides a more visual approach by displaying individual data points and their relationships. Outliers appear as points far removed from the main cluster of data, making it easy to spot customers with extreme values, such as balances that don’t align with expected trends.

Example, a customer with a low salary but a high balance might be flagged as an outlier in a scatter plot of balance vs. salary. Scatter plots can also reveal clusters and trends, helping to detect deviations from expected patterns. When combined with a trendline, outliers become even more apparent as points that stray far from the general trend. This helps in identifying customers whose behavior differs from the norm, possibly indicating financial risk.

****

**Conclusion:** No outliers have been found in the data.

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

The dataset has seven different tables they are, Active Customer, Bank Churn, Credit Card, Customer Info, Exit Customer, Gender, Geography.

**Tables with Categorical Variables:**

* Customer Info: Contains categorical variables like Surname.
* Exit Customer: Contains categorical variables like Exit Category.
* Gender: Contains categorical variables like Gender Category.
* Geography: Contains categorical variables like Geography Location.
* Active Customer: Contains categorical variables like Active Category.
* Credit Card: Contains categorical variables like Category.

**Conclusion:** There is no such table which contains only categorical variables.

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)**

This SQL query calculates the average income for males and females within each geographic location and assigns rank based on the average salary.

**Query:**

WITH AverageGeographySalary AS (

SELECT

g.GeographyLocation,

CASE

WHEN c.GenderID = 1 THEN 'Male'

ELSE 'Female'

END AS Gender,

AVG(c.EstimatedSalary) AS avg\_salary

FROM customerinfo c

INNER JOIN geography g ON c.GeographyID = g.GeographyID

GROUP BY g.GeographyLocation, c.GenderID

ORDER BY g.GeographyLocation

)

SELECT \*,

RANK() OVER (PARTITION BY GeographyLocation ORDER BY avg\_salary DESC) AS 'rank'

FROM AverageGeographySalary;



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+).**

To find the average tenure for each age bracket we first segregate the user base into three segments.

* People with age from 18-30 as adults.
* People with age from 31-50 as Middle Aged.
* People with age above 50 as Old Aged.

**Query:**

SELECT

CASE

WHEN age BETWEEN 18 AND 30 THEN 'Adult'

WHEN age BETWEEN 31 AND 50 THEN 'Middle-Aged'

ELSE 'Old-Aged'

END AS AgeBrackets,

AVG(b.tenure) AS avg\_tenure

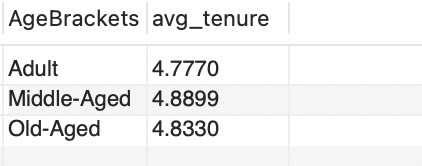
FROM customerinfo c

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

WHERE b.exited = 1

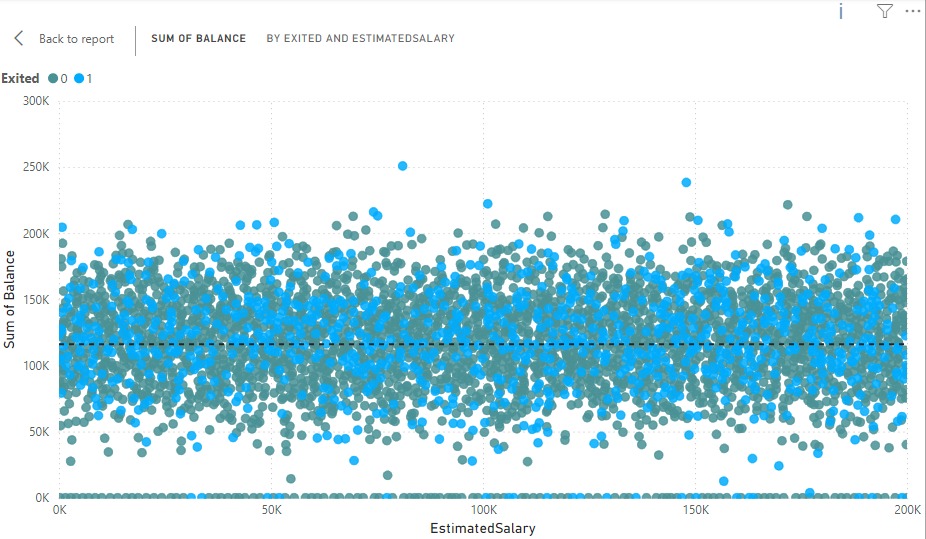
GROUP BY AgeBrackets

ORDER BY AgeBrackets;



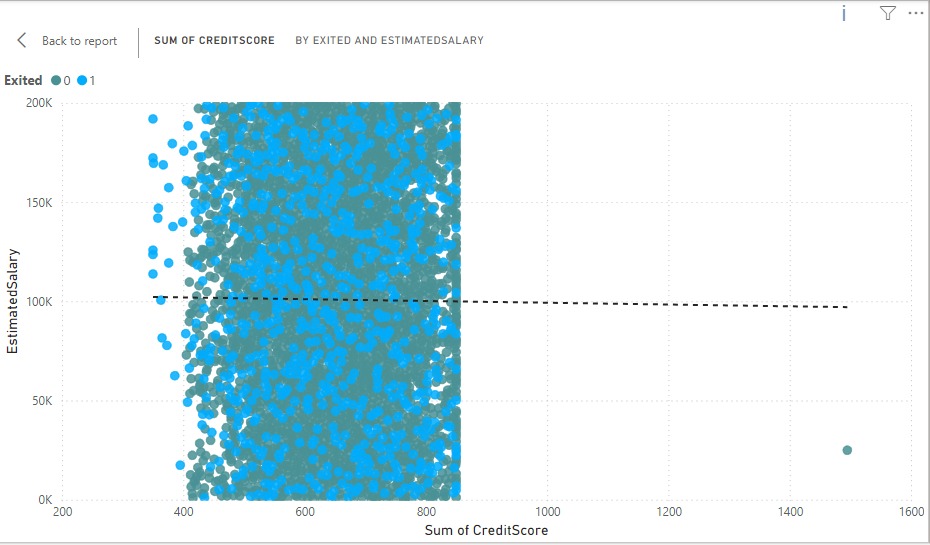
1. **Is there any direct correlation between salary and the balance of the customers? And is it different for people who have exited or not?**

In this analysis, we are examining the correlation between Balance and Estimated Salary. From the scatter plot, there does not appear to be a strong or clear relationship between these two variables. The data points are spread relatively uniformly across various salary and balance ranges, indicating that higher or lower salaries do not consistently correspond to higher or lower balances. Additionally, the two categories for Exited (represented by different shades on the plot) are evenly distributed throughout the chart. This suggests that customer churn (exiting behaviour) does not seem to be directly influenced by either salary or balance based solely on this visual representation.



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

There does not appear to be a strong or clear correlation between Credit Score and Estimated Salary. The data points are widely and evenly spread across all ranges of both variables, indicating no evident linear relationship. Additionally, the exit status (Exited 0 or 1) is uniformly distributed across different credit scores and salary ranges. This suggests that neither credit score nor estimated salary alone significantly predicts customer exit behaviour. Despite some variation, the graph does not visually present any clear pattern or correlation between these two variables and customer churn.

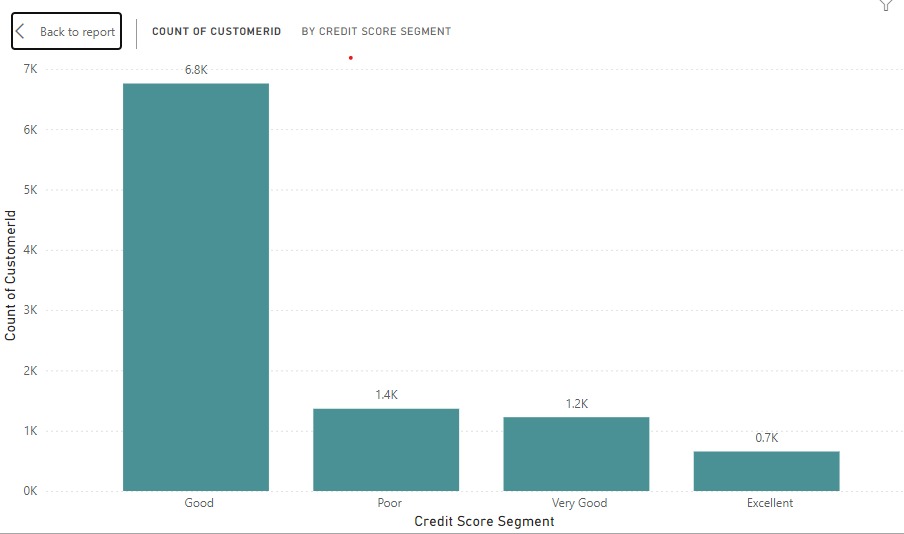


1. **Rank each bucket of credit score as per the number of customers who have churned the bank.**

The below chart provides the distribution of the customer base across different credit score segments.

Insights: On the basis of the graph the credit score segments are ranked by the count of customers as follows:

1. Good (6.8k customers)
2. Poor (1.4k customers)
3. Very Good (1.2k customers)
4. Excellent (0.7k customers)



**Conclusion:** the majority of customers fall into the "Good" credit score segment, followed by "Poor," "Very Good," and then "Excellent." This chart represents churned customers, the ranking indicates the distribution of churn across these segments.

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.**

The below query provides the numbers of customers who have a credit card based on the credit card. We first segregate the customers based on the age into age brackets and then find those buckets that have lesser than average number of credit cards per bucket.

**Query explanation:**

1. Common Table Expression (CTE): creditinfo:
   * This part defines a CTE named creditinfo.
   * It joins the customerinfo (c) and bank\_churn (b) tables on CustomerID.
   * A CASE statement categorizes Age into 'Adult' (18-30), 'Middle-Aged' (31-50), and 'Old-Aged' (above 50).
   * It filters for customers with credit cards (HasCrCard = 1).

**(Filtering is done before counting)**

* + It groups the data by age\_brackets and calculates:
  + HasCreditCard: The count of customers with credit cards within each age bracket.

1. Main Query:
   * This section selects all columns (\*) from the info CTE.
   * It then filters the results to include only age brackets where HasCreditCard is less than the average number of credit cards per bracket.
   * The average is calculated using a subquery that computes the AVG(HasCreditCard) from the info CTE.

**Query:**

WITH creditinfo AS (

SELECT

CASE

WHEN age BETWEEN 18 AND 30 THEN 'Adult'

WHEN age BETWEEN 31 AND 50 THEN 'Middle-Aged'

ELSE 'Old-Aged'

END AS AgeBrackets,

COUNT(c.CustomerId) AS HasCreditCard

FROM customerinfo c

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

WHERE b.Has\_creditcard = 1 **-- Ensures filtering is done before counting**

GROUP BY AgeBrackets

)

SELECT \*

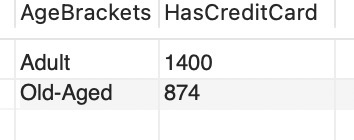
FROM creditinfo

WHERE HasCreditCard < (

SELECT AVG(HasCreditCard)

FROM creditinfo

);



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

**Query:**

SELECT

g.GeographyLocation,

COUNT(b.CustomerId) AS TotalExited,

AVG(b.Balance) AS avg\_bal

FROM bank\_churn b

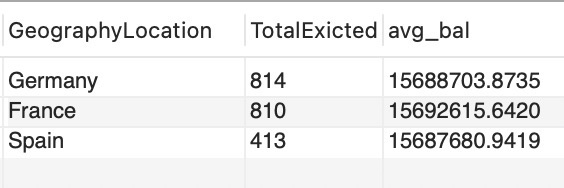
JOIN customerinfo c ON b.CustomerId = c.CustomerId

JOIN geography g ON c.GeographyID = g.GeographyID

WHERE b.Exited = 1

GROUP BY g.GeographyLocation

ORDER BY COUNT(b.CustomerId) DESC;



**Conclusion:**

* **Germany**: 814 customers exited, with an average balance of approximately 15,687,083.
* **France**: 810 customers exited, with an average balance of approximately 15,692,615.
* **Spain**: 413 customers exited, with an average balance of approximately 15,687,680.

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”.**

The column where the format is “CustomerID\_Surname” can be created in POWER BI using the formula:

CustomerID\_Surname = CustomerInfo[CustomerId]&” “&CustomerInfo[Surname]

Steps: In Table View -> Click on New Column -> Use the Formula.

By following the above steps you can create a new column in the prescribed fomat.



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

This code retrieves customer data from the Bank\_Churn table and adds a new column named "ExitCategory" to classify customers as 'Retain' (not churned) or 'Exit' (churned) based on the value in the Exited column (likely 0 for non-churned, non-zero for churned).

**Query:**

SELECT

CustomerId,

CreditScore,

Tenure,

Balance,

NumOfProducts,

HasCrCard,

IsActiveMember,

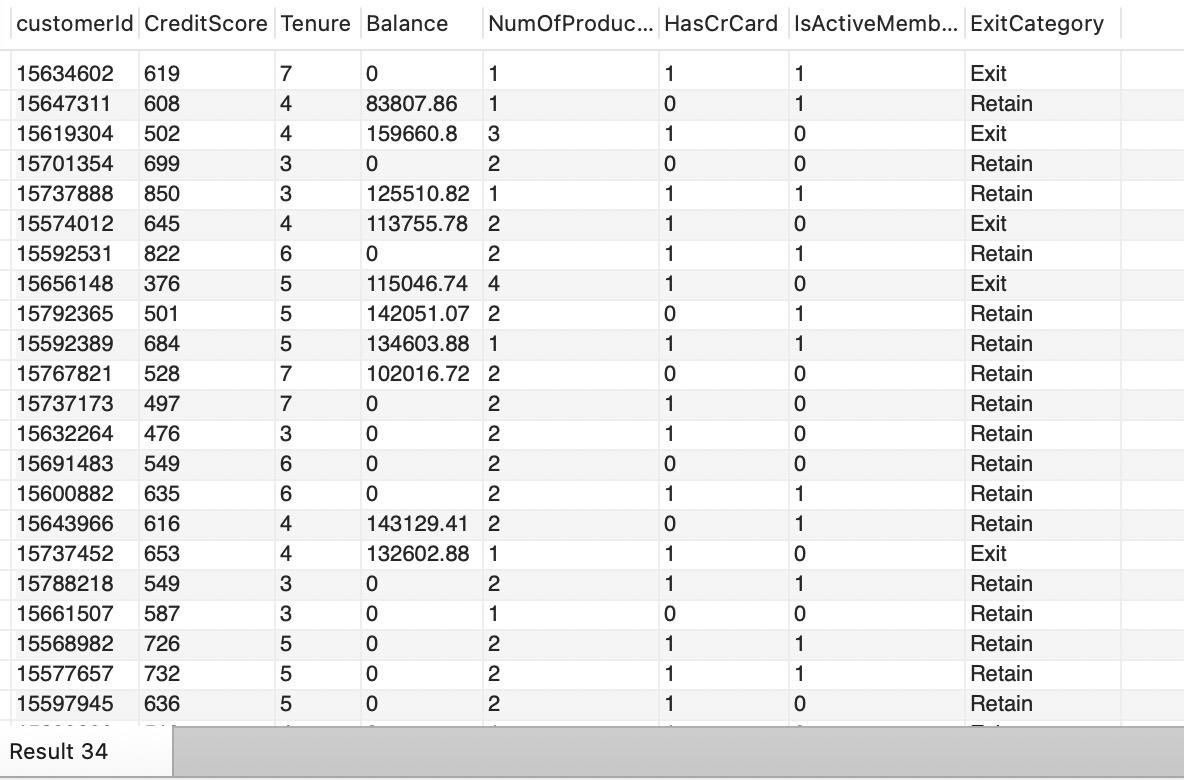
CASE

WHEN Exited = 0 THEN 'Retain'

ELSE 'Exit'

END AS ExitCategory

FROM bank\_churn;



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

**No,** there weren’t any missing values in the data. This eliminates the need for imputation techniques that might introduce assumptions or biases.

If there were missing values in the data, I would use these techniques to handle them:

* **Deletion**: This involves removing rows or columns with missing values.
* **Imputing missing values**: Fill missing values using statistical methods like mean, median, or values from other related columns.
* **Replace Missing Values with a Default Value**

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”.**

This SQL query identifies customers with surnames ending in "on"

**Query:**

SELECT

c.CustomerId,

c.Surname,

CASE

WHEN b.IsActiveMember = 1 THEN 'Active'

ELSE 'Inactive'

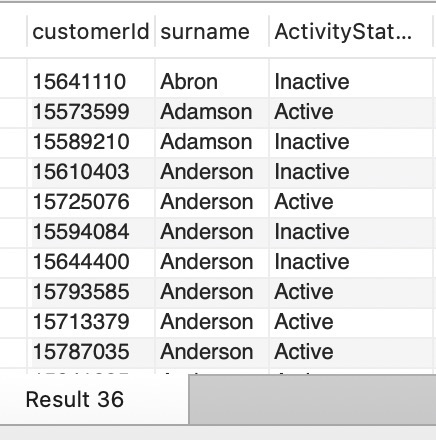
END AS ActivityStatus

FROM customerinfo c

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

WHERE c.Surname LIKE '%on'

ORDER BY c.Surname;



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.**

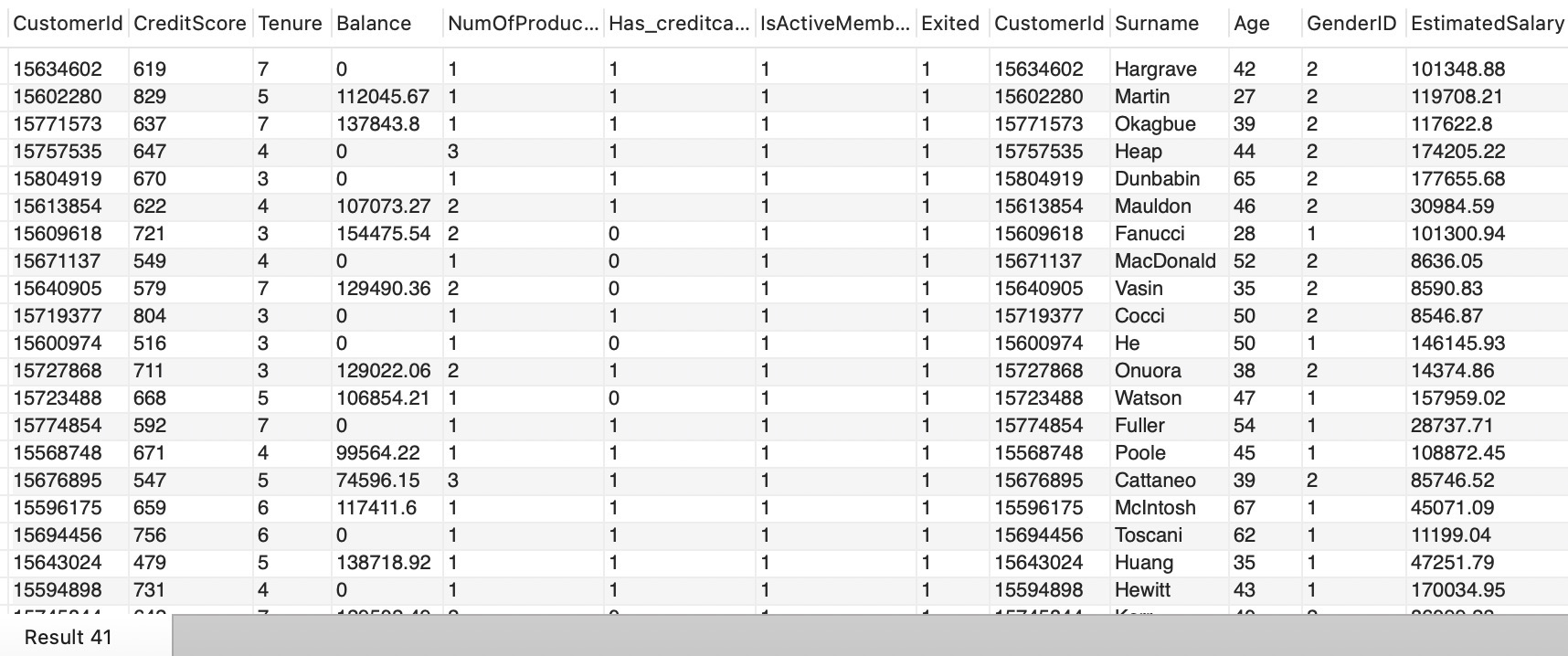
**Query:**

SELECT \*

FROM bank\_churn b join customerinfo c on b.CustomerId = c.CustomerId

WHERE b.Exited =1 and b.IsActiveMember =1;

Yes, there are data discrepancy in the Customer’s data. When the customer is Active which mean the customer is still in the bank and hence the data in IsActiveMember column is 1. When the customer exits the bank, the Exited column is 1. But here there is data which represents both the customer Exited and also an active member at the same time. This is due to data discrepancy.



**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?**

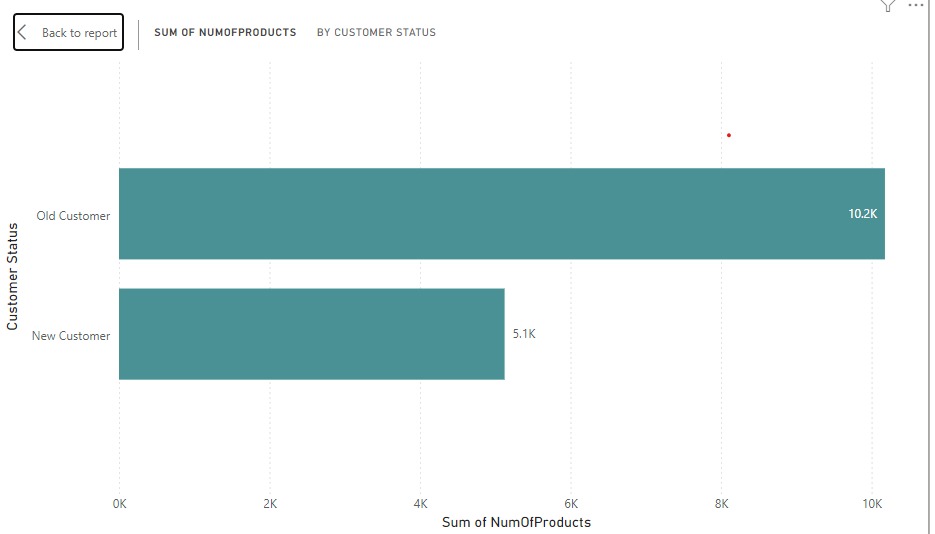
We first segregate the customers into long-term customers and short-term customers based on their date of joining in the bank. The customers who have joined in the year 2019 are new customers and the customers who have joined in all the previous years are old customers.

Now let’s compare few behavioral spending habits of the customers.

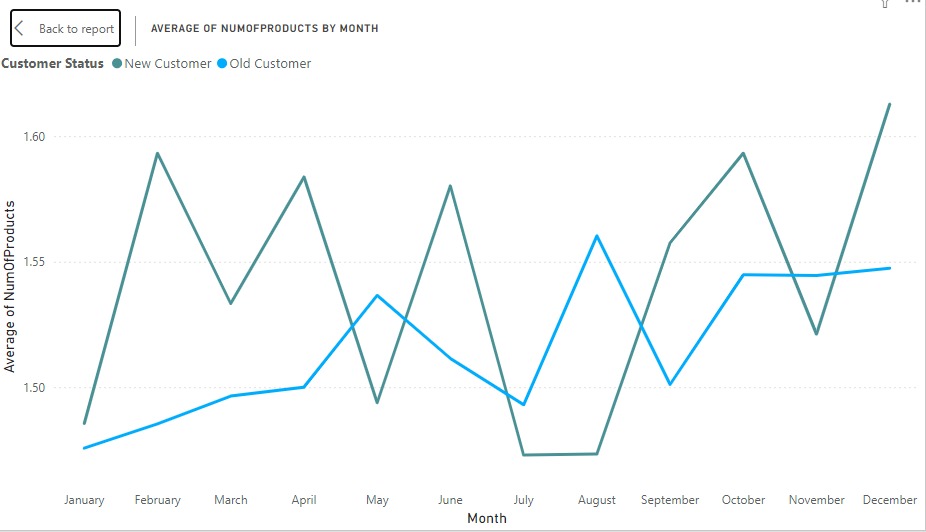
**Number of Products:**

Insights:

* Long-term customers consistently hold more products than new customers, indicating that they engage more deeply with the bank over time.
* There is a gradual increase in the average number of products held by new customers, implying that they adopt more of the bank’s offerings as they become familiar with its services.
* The Total Number of products used by old customers are 10.2k.
* The Total Number of products used by new customers are 5.1k.



* In January, new customers hold significantly fewer products compared to long-term customers, likely because they are still exploring the bank's services and opening accounts.

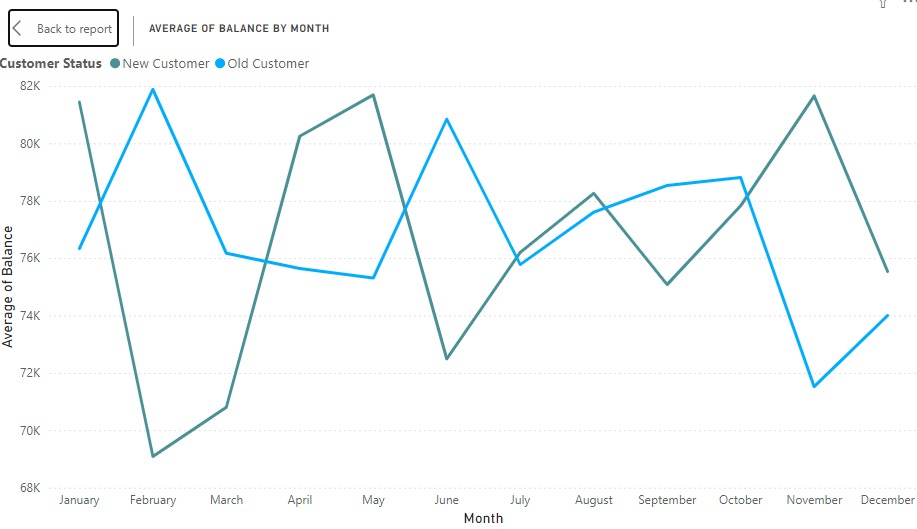


**Average Balance Analysis:**

The chart shows that new customers consistently have lower average balances compared to long-term customers. This indicates that long-term customers tend to spend or save more.

Insights:

* As their income grows over time, long-term customers might be able to save or invest more money.
* The bank may offer better interest rates or perks to long-term customers, incentivizing them to keep more money in their accounts.

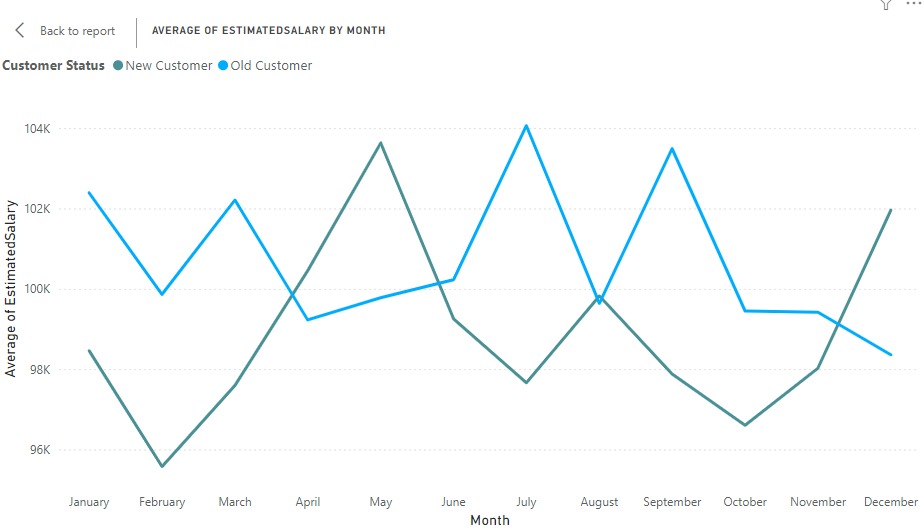


**Salary Analysis:**

The chart compares the average salary by month for new and long-term customers.

**Insights:**

* + The average salary for long-term customers is consistently higher than that of new customers.
  + New customers have lower average salaries initially, there is a slight upward trend throughout the year, suggesting salary increases or job changes leading to higher pay.



**Conclusion:**

* Old customers use twice as the number of products used by the new customers.
* The average salary for long-term customers is consistently higher than that of new customers.
* As their income grows over time, long-term customers might be able to save or invest more money.

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

Customers often tend to use specific bank products in combination, and understanding these pairings can help the bank develop effective cross-selling strategies. These strategies can boost customer satisfaction and increase revenue by offering relevant products that complement what customers are already using.

**Commonly Paired Bank Products:**

* Checking Accounts
* Debit Cards
* Savings Accounts
* Credit Cards
* Loans

**Cross-Selling Strategies:**

* **Offer complementary products:** Customers with checking accounts may benefit from being offered a debit card and online banking for easier money management. Those with savings accounts could be encouraged to set up automatic transfers to grow their savings or consider high-interest options like CDs for larger balances.
* **Personalize recommendations based on usage**: Customers who frequently use credit cards for travel could be offered travel rewards cards. Those seeking loans might be interested in bundled insurance products for added convenience and protection.
* **Utilize digital platforms:** Promote services like paperless statements and automatic bill payments through online or mobile banking. You can also highlight investment options or financial planning tools that customers can access via these digital platforms.

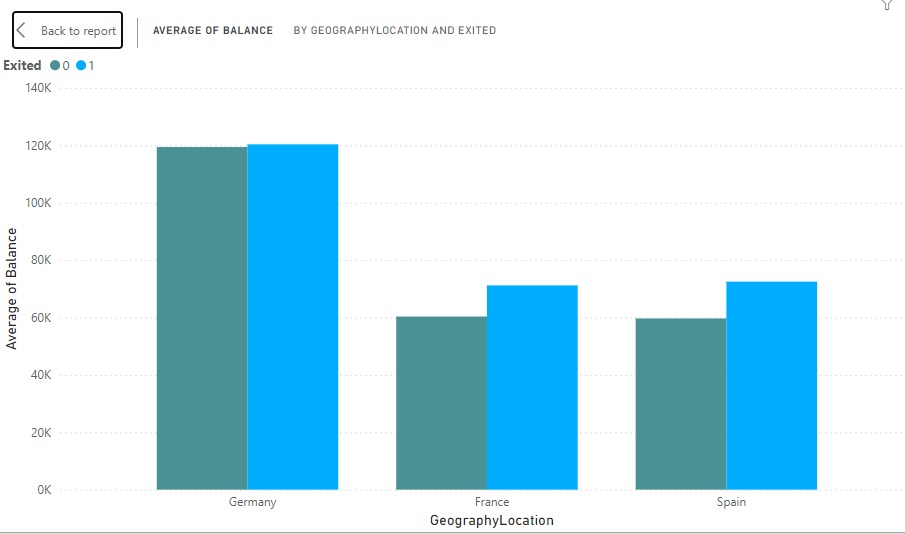
By analyzing product usage and tailoring recommendations accordingly, the bank can increase both revenue and customer satisfaction, as customers feel their financial needs are being met more effectively.

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

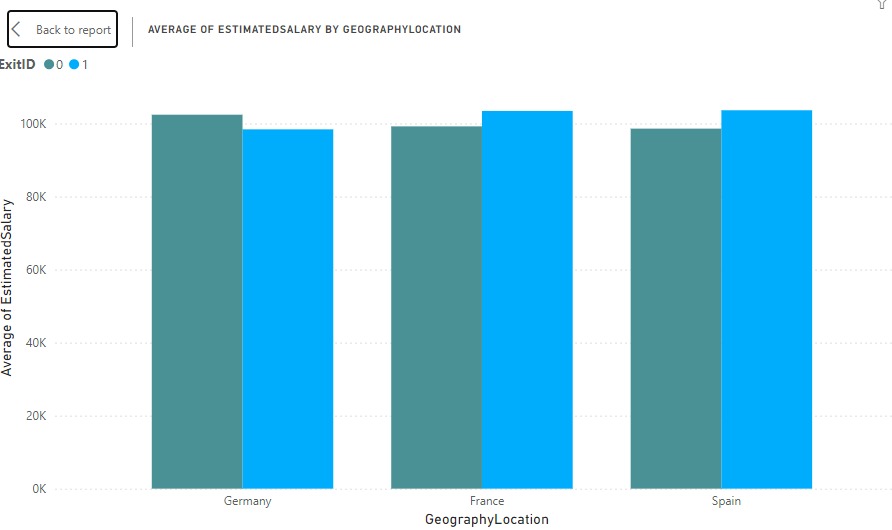
These charts provide insights on how economic indicators vary across geographic regions and their potential correlation with customer churn rates.

We'll analyze three key economic indicators: average balance, average salary, and credit score.

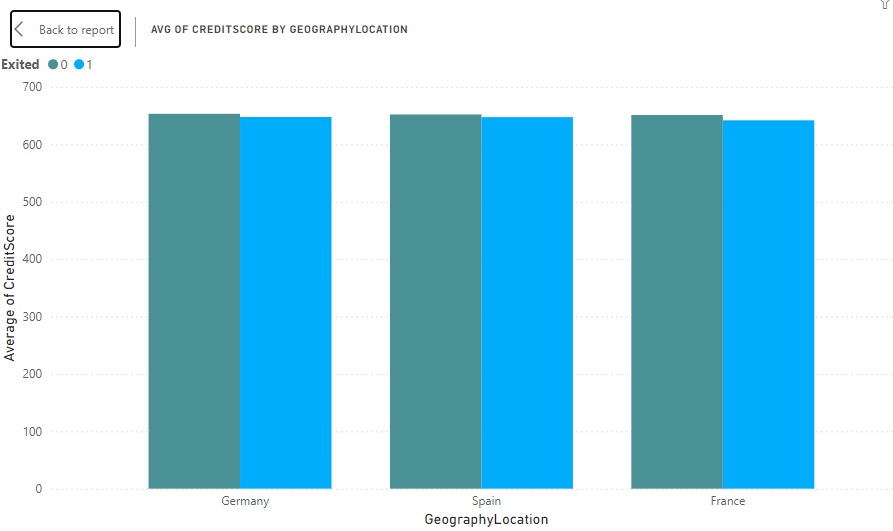
The first chart examines the average balance by geographic location.  
The chart shows average balance by location, with Germany having the highest. While it doesn't directly address churn, lower balance regions might have higher churn if customers there are more cost-sensitive.



The second chart shows the average credit score by geographical location.  
The credit score chart shows variations by location, with Germany having the highest average. This, along with churn rate data, could help identify a link between higher credit scores and lower churn.



The third chart shows the average salary by geographical location.   
Similar to the credit score chart, average salary varies by location.



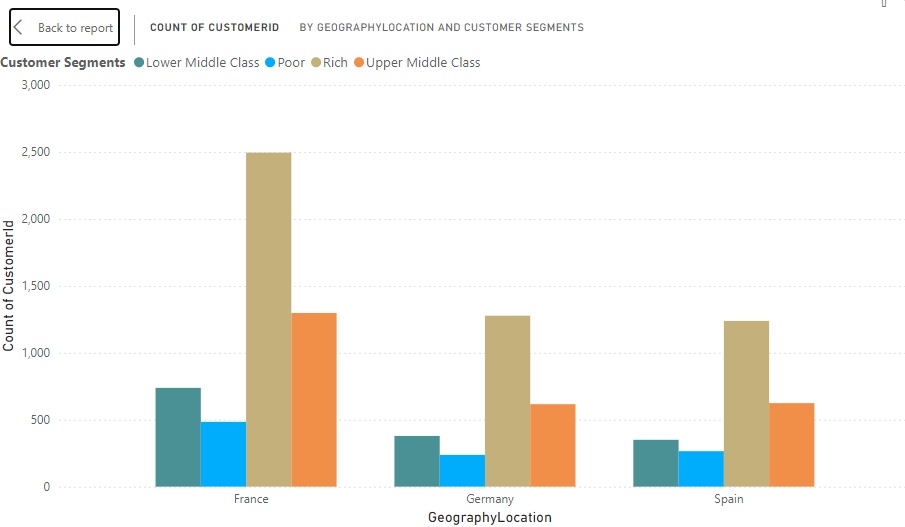
**Conclusion:**

The customers from Germany are most active account holders in the bank and also maintain the best credit score and bank balance.

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

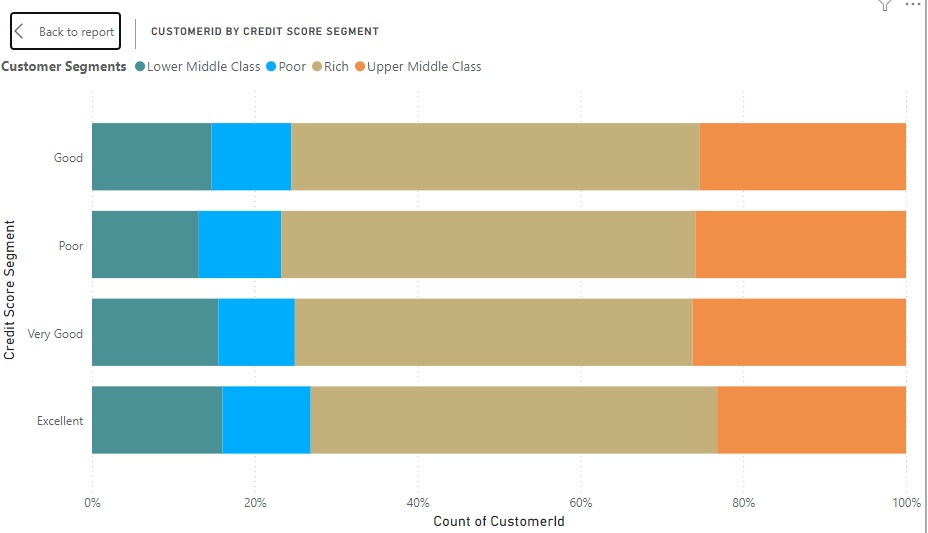
Firstly, the customer profiles can be divided into four different segments based on the salary they are Lower Middle Class, Poor, Rich and Upper Middle Class.

The charts provide the insights on the demographic segments with the most customers are those with lower credit scores.



**Insights:**

* + France has the greatest number of Poor and Lower Middle-Class people followed by Germany and Spain.



**Insights:**

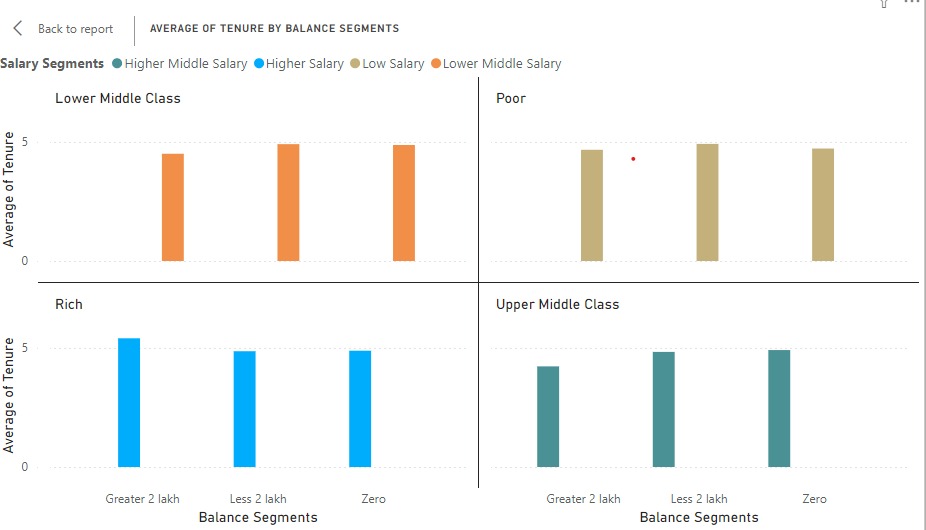
These demographic segments also represent the largest portion of the bank's customer base. Based on the data in the chart, it's clear that these groups (“lower middle class” and “poor”) not only have the most customers but also present the highest potential financial risk to the bank.

**Conclusion:**

* + While creditworthiness is a key factor in assessing customer risk, other elements such as income and employment history should also be considered.
  + That said, creditworthiness remains a strong predictor of a customer's ability to repay loans.
  + Customers with lower credit scores are statistically more likely to default on loans compared to those with higher credit scores.
  + This suggests that the bank faces a greater financial risk when lending to customers in the "Lower Middle Class" and "Poor" segments, as these groups are more likely to result in loan defaults than 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?**

The chart below presents the average tenure of customers by balance segments, salary segments, and overall customer segments.



**Insights:**

Balance Segments: Customers with higher average balances tend to have a longer tenure with the bank. This may be due to their higher level of investment in the bank’s products and services or the bank’s efforts to offer better benefits to retain them.

* + High Salary & Balance Greater than 2 Lakh: This segment has the longest average tenure, at approximately 5.4 years.
  + Zero Balance: Customers with zero balance have the shortest average tenure, around 4.2 years across all salary segments.

Salary Segments: There is a weak correlation between salary and tenure. On average, customers in higher salary segments have slightly longer tenures, but the difference is not significant across most segments.

* + High Salary: Customers in this segment are likely to have an average tenure of 4.8 to 5.4 years.
  + Low Salary & Lower Middle Class: Expected tenure for this segment is around 4.5 to 4.9 years.
  + Upper Middle Class & Rich: Average tenure in this segment is likely to be between 4.8 and 4.9 years.

**Conclusion:**

* + Balance segments play a more critical role in predicting customer tenure than salary segments.
  + Banks can use this insight to focus retention strategies on customers in higher balance segments, while also working to enhance customer satisfaction and product adoption across all segments to improve. overall tenure.

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?**

This approach can be used to give insights for the impact of marketing campaigns on customer retention and acquisition within a dataset.

1. Segment the Data: Segment the dataset based on the different marketing campaigns. This allows you to analyze each campaign individually and observe how it impacts customer retention and acquisition.

2. Define Key Metrics: Identify the key metrics for both retention and acquisition. For retention, customer churn rate, retention rate and customer acquisition rate.

3. Gather Additional Information: To ensure a thorough analysis, gather relevant data such as:

* + Customer demographics
  + Campaign specifics
  + Competitor data
  + External factors

4. Calculate Metrics: Calculate the key retention and acquisition metrics over relevant time periods (e.g., monthly, quarterly, or yearly). This helps to track how customer behavior changes over time in response to different campaigns.

5. Compare Results Across Campaigns: Compare the performance metrics across various campaigns to identify which are most effective at acquiring and retaining customers. This step helps highlight the campaigns driving success or those needing improvement.

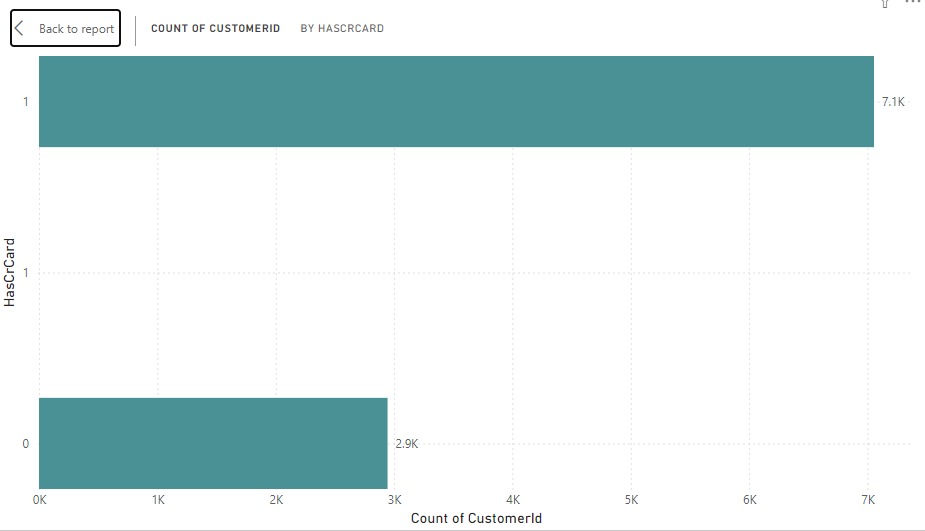
6. Conduct Statistical Analysis: Use statistical methods like t-tests or ANOVA to evaluate whether the differences in performance between campaigns are statistically significant.

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

To find the common characteristics or trends among customers who have exited that could explain their reasons for leaving we can use the below chart analysis.

Insights:

* Credit card ownership: The idea here is that customers who have credit cards are more likely to churn than those who don't.
* Number of products purchased: Customers who buy fewer products from the bank are more likely to churn than those who buy more products.

****

**Conclusion:**

**The main reasons for exit are:**

* Credit Card Ownership: The chart displays the number of customers who exited the bank (churned) categorized by whether they own a credit card (HasCrCard) or not. According to the chart, a significantly higher number of customers with credit cards exited (around 1,200) compared to those without credit cards (around 300).
* Less Products Purchased: The number of customers who exited the bank is highest for those who purchased no products (around 800), and the number of exits decreases as the number of products purchased increases. Customers who purchased four or more products have the lowest exit rate (approximately 60). This suggests that customers who engage more with the bank by purchasing multiple products are less likely to churn.

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

Analysis of Factors for Predicting Customer Churn:

The aim is to determine whether tenure, number of products held, active membership status, and estimated salary can predict whether a customer will leave the bank. The analysis is based on four charts that visualize these factors for both customers who exited and those who remained.

Insights on NumOfProducts:

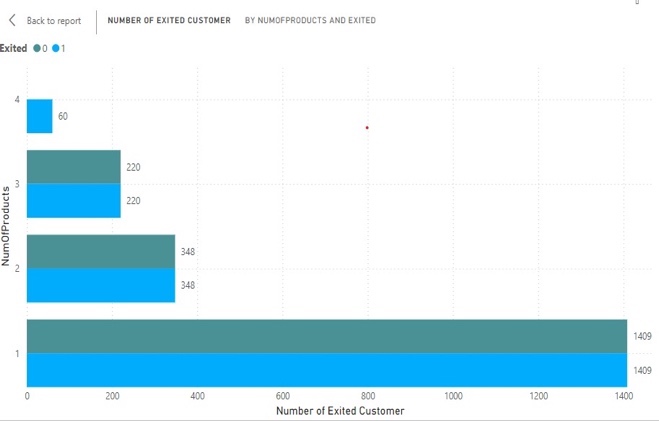
* The chart compares the number of products customers hold (NumOfProducts) with the likelihood of exiting or staying, represented by the count of customer IDs.

Observations:

* There is no consistent pattern between the number of products a customer holds and their likelihood of exiting.
* In some ranges, such as 1 product or 3 products, more customers exited than stayed.
* Conversely, for 2 products and 4 or more products, more customers stayed with the bank than those who exited.

Importance of NumOfProducts for Prediction:

* Based on this chart alone, it's hard to conclude that the number of products is a strong predictor of customer churn, as the trend fluctuates across product ranges.
* However, NumOfProducts could still play a role in conjunction with other factors:
  + Customer Needs: Customers who hold more products might have their needs more fully met, increasing satisfaction and reducing the likelihood of leaving.
  + Account Management: On the other hand, customers managing multiple products might feel overwhelmed, potentially leading to frustration and churn in some cases.

****

Comparison with Tenure:

While comparing the relationship with tenure of the customers and the exit rate of the customers we can find a key relationship that the customers who have a tenure of greater than or equal to four years are more likely to retain than others.

**Conclusion:**

While NumOfProducts by itself does not show a clear predictive trend for churn, it might be valuable when combined with other factors like tenure, IsActiveMember, and estimated salary. The interplay of these attributes could provide a more accurate prediction model for customer churn.

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

The customers are segmented mainly into three categories based on their estimated salary. They are Low, Medium and High segments. These segments are further divided in female and male customers based on their Geography.

**Query:**

SELECT

g.GeographyLocation,

CASE

WHEN EstimatedSalary < 50000 THEN 'Low'

WHEN EstimatedSalary < 100000 THEN 'Medium'

ELSE 'High'

END AS IncomeSegment,

CASE

WHEN c.GenderID = 1 THEN 'Male'

ELSE 'Female'

END AS Gender,

Age,

COUNT(c.CustomerId) AS NumberOfCustomers

FROM customerinfo c

JOIN geography g ON c.GeographyID = g.GeographyID

GROUP BY IncomeSegment, g.GeographyLocation, Gender, Age

ORDER BY g.GeographyLocation, Age;



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?**

The chart, which appears to be filter window, we can 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 by following these steps:

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:
   * Customers with a low number of products purchased (NumOfProducts)
   * Customer id
   * Customers Salary segments
   * Customer Segments
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.
4. 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?**

**Customer Churn Analysis and Retention Strategies:**

**Churn Rate Overview:**

* The bank’s overall churn rate is **20.37%**, with year-on-year fluctuations:
  + **2016**: 19.27%
  + **2017**: 22.35% (highest)
  + **2018**: 20.21%
  + **2019**: 19.86% (lowest)

**Customer Segments Prone to Churn:**

Analysis reveals several customer segments with higher likelihoods of churn:

* **Single Product Users**: Customers using only one product may not see enough value in the bank's offerings, particularly when competitors provide broader or more integrated services.
* **Credit Card Holders**: Potential factors for credit card holders churning include:
  + Insufficient credit limits.
  + Lack of appealing rewards programs.
  + High fees associated with the card.
* **Tenure of 4-5 Years**: These customers might be coming off promotional offers or discounts, making them susceptible to competitors offering more attractive rates or features.
* **High Salary Customers**: High earners may have more financial options and are more likely to switch banks for slightly better benefits or interest rates elsewhere.

**Recommendations to Reduce Churn:**

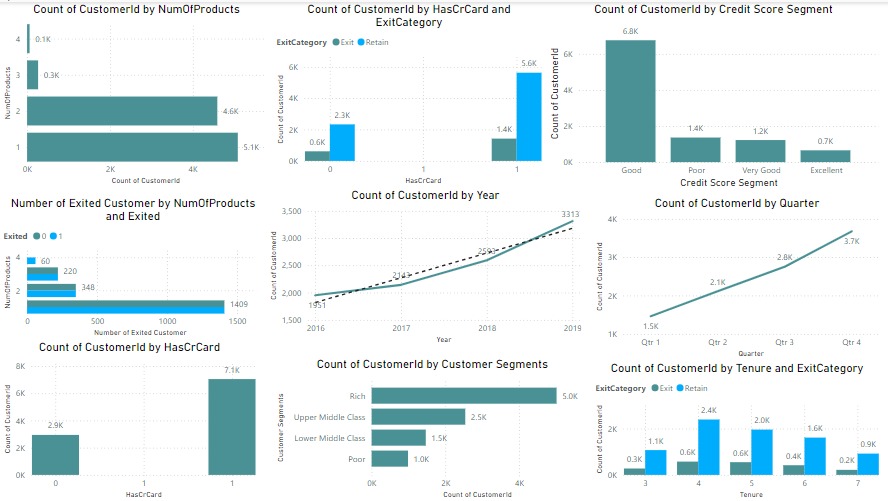
* **Targeted Product Bundles**:
  + Develop tailored product bundles to address the specific needs of customers who use only one product. Highlight the added value and potential cost savings of using multiple services.
* **Enhanced Credit Card Rewards**:
  + Improve the credit card offering by:
    - Increasing credit limits based on customer behavior and creditworthiness.
    - Aligning rewards programs with customer preferences (e.g., travel, cashback for certain categories).
    - Reducing or eliminating annual fees, especially for high-value customers.
* **Retention Offers for Existing Customers**:
  + For customers nearing the end of their introductory offers, proactively offer personalized retention deals, such as:
    - Extending the promotional rates.
    - Providing discounts on other products or services.
* **Customer Satisfaction Surveys**:
  + Regularly survey customers to gather insights into why they leave. Use this feedback to fine-tune retention strategies and address customer pain points.
* **Relationship Management for High-Value Customers**:
  + Assign dedicated relationship managers to high-value customers, offering them personalized services and exclusive benefits to strengthen loyalty and meet their individual needs.

By focusing on these strategies, the bank can reduce churn, retain high-value customers, and increase overall customer satisfaction.

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

**Dashboard:**

****

****

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

In the absence of explicit objective and subjective questions, a systematic approach to solving a problem can still be highly effective. Structured process to tackle issues by generating hypotheses, asking critical questions, and deriving actionable insights.

**1. Hypothesis Generation:**

**Start with Assumptions**: Begin by forming hypotheses about the problem or data you're analyzing. These can be informed by domain expertise, past experiences, industry norms, or initial observations.

**Example**: You might hypothesize that customers with higher account balances are less likely to churn, or that specific marketing channels drive higher conversions.

**2. Formulating Questions:**

**Craft Relevant Questions**: Develop a set of questions that will guide your analysis, focused on testing your hypotheses. These questions should drive your data exploration.

* **Sample Questions**:
  + **Customer Churn**:
    - Are certain demographic factors like age or income level tied to higher churn rates?
    - Does the number of products a customer holds reduce the likelihood of churn?
    - How do customer engagement metrics, such as transaction frequency or login activity, correlate with churn?
  + **Marketing Campaign Analysis**:
    - Which marketing channels (email, social media, paid ads) generate the best results in terms of customer acquisition or retention?
    - Is there a clear relationship between ad spend and campaign performance (conversions, ROI)?
    - How does different campaign messaging (promotional, educational) affect customer engagement rates?

**3. Data Exploration and Analysis:**

* **Examine the Data**: Use data exploration techniques like visualization, trend analysis, and statistical testing to find patterns and validate or reject your hypotheses.
* **Analyze Relationships**: For example, use correlation analysis to see if churn decreases with an increase in the number of products held, or run regression models to measure the impact of campaign channels on conversion rates.

**4. Insights and Recommendations:**

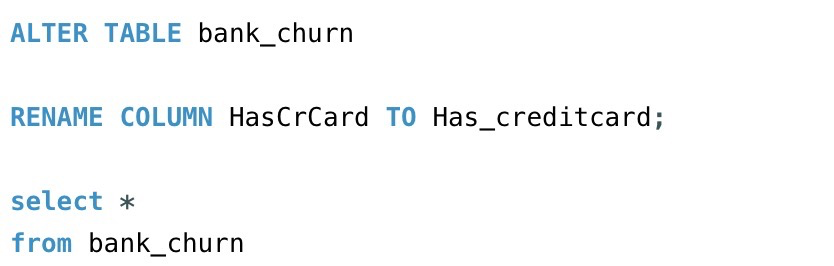
* **Draw Insights**: Summarize the results of your analysis into actionable insights. These insights should provide clarity on the key factors influencing the problem and offer guidance for decision-making.
* **Make Recommendations**: Based on your insights, propose strategies that could improve outcomes. For instance, if customer churn is linked to low product usage, you could recommend bundling more products to drive loyalty. If certain marketing channels outperform others, allocate more budget to those channels for better campaign results.

By following this approach, you systematically navigate through the data to derive valuable insights and make informed decisions, even without a predefined set of questions.

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

This SQL query changes the name of the “HasCrCard” column to “Has\_creditcard” in the Bank\_churn table. It is a more descriptive name which provides more clarity.

**Query:**



**Output:** The output shows that the column name “HasCrCard” had been change to “Has\_creditcard**”.**

