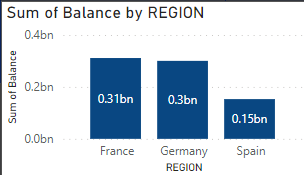
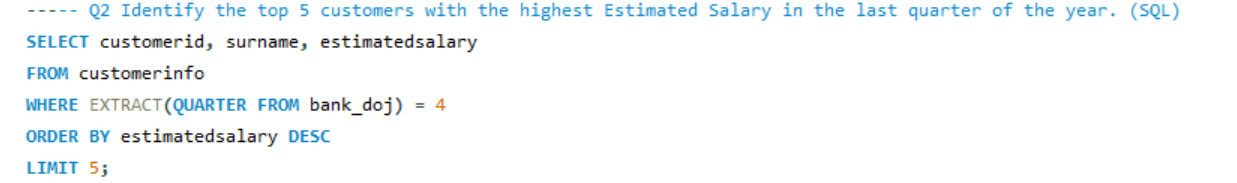
Objective Questions

1. Account Balance Distribution by Region:

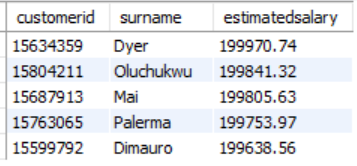
From the below given graph, it can be seen that France shows maximum sum of balance of about 0.31bn while Spain just show half of that of France that is 0.15bn.  


2. Top 5 Highest Estimated Salary Earners (Last Quarter)(SQL):

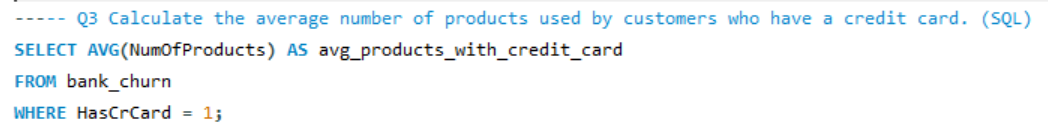
SQL query



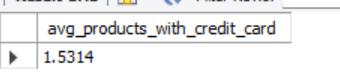
Result:



3. Average Number of Products for Customers with Credit Cards(SQL):

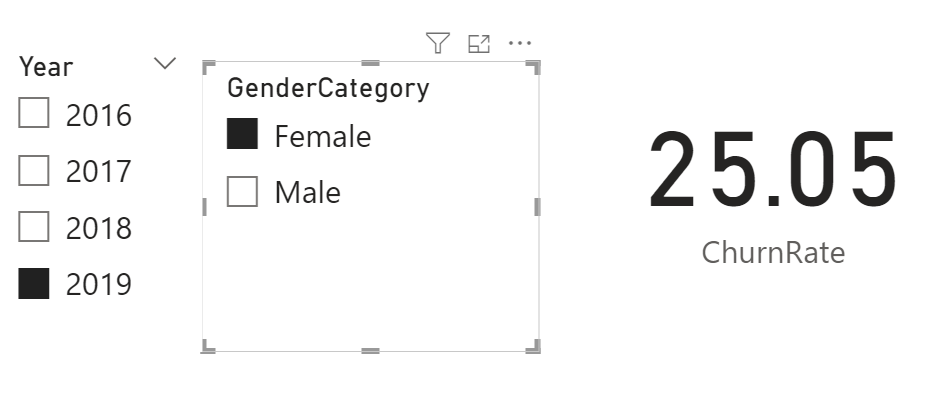
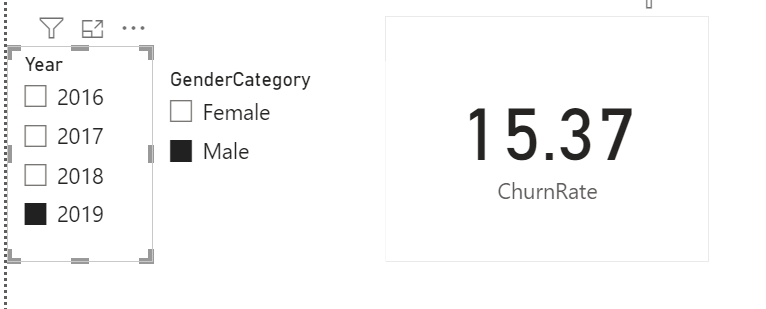
SQL query 

Result:



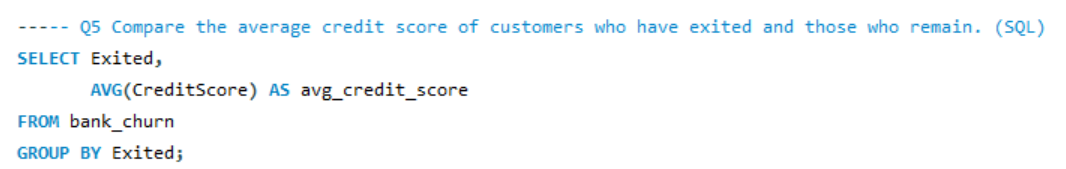
4. Churn Rate by Gender

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

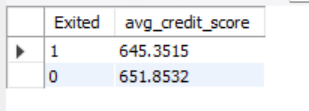
 

5. Average Credit Score of Exited vs. Retained Customers(SQL)

SQL query

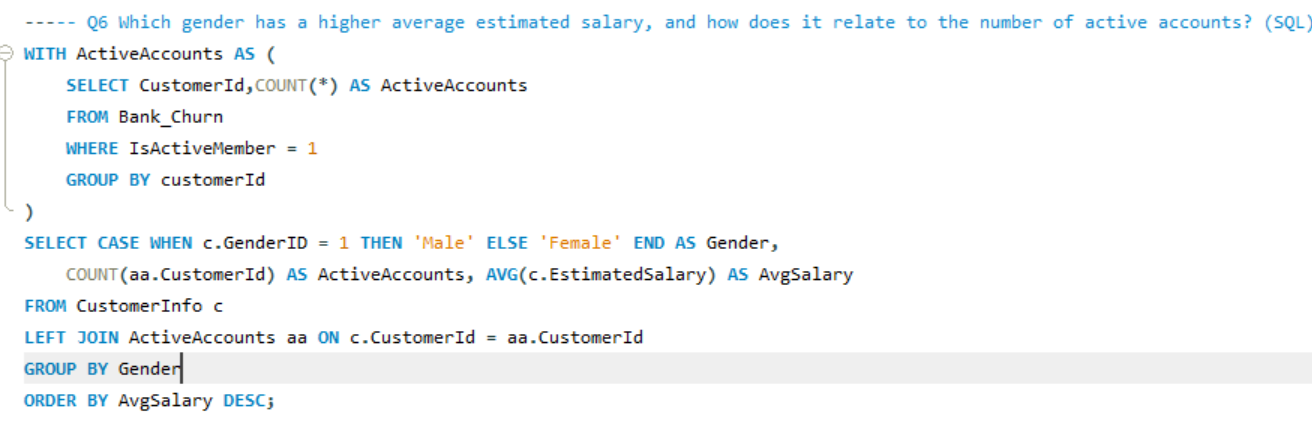


Result:

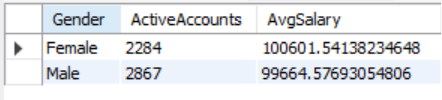


6. Gender with Higher Average Salary and its Relation to Active Accounts:

SQL query

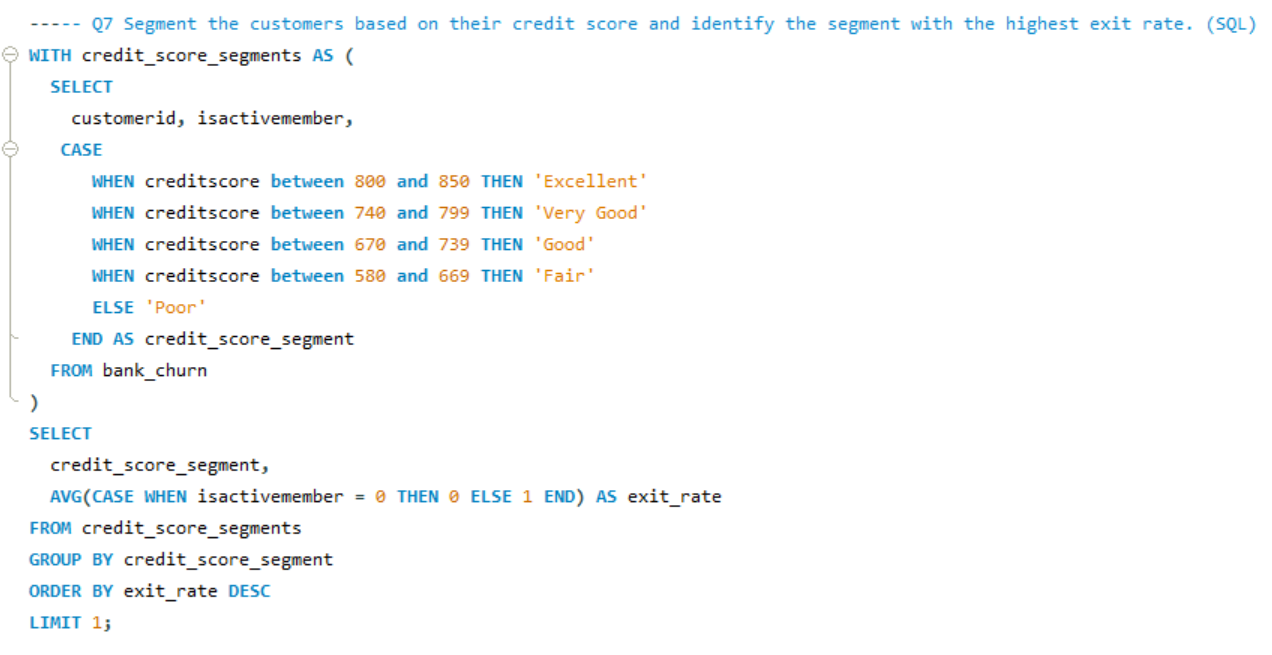


Result:

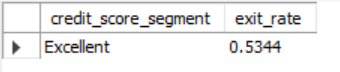


7. Customer Segment with Highest Exit Rate

This query utilizes a Common Table Expression (CTE) named credit\_score\_segments to categorize customers based on their credit score:

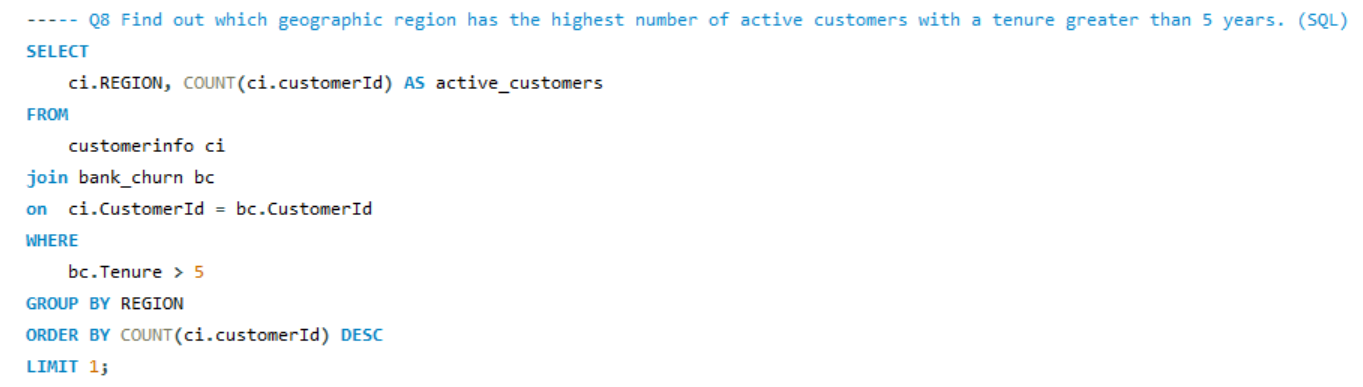


Result:



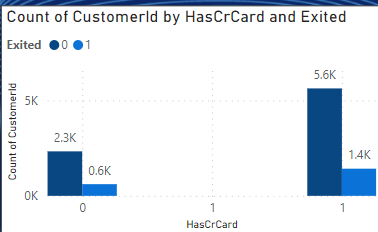
8. Geographic Region with Most Active Customers (Tenure > 5 Years):

SQL query

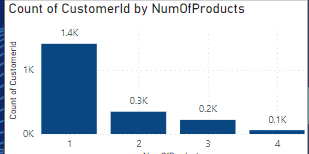


9. Impact of Credit Card on Customer Churn

The chart reveals a potentially higher churn rate for customers with credit cards (HasCrCard = 1) compared to those without credit cards.

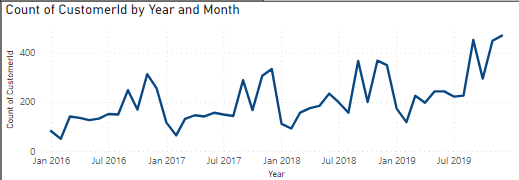


Most Common Number of Products for Exited Customers: From the below given graph it can be seen that product 1 (1.4k) was the mostly commonly used product by exited customers. Which is found out by applying filter to it. While product 4 (0.1k) was least used product by the exited customers.



11. 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 customer joining trend appears to be increasing over time. This indicates a positive growth in customer acquisition.There might be seasonal patterns present in the data. It seems that customer joins could potentially peak around the end of the year (December) but due to the limited data points, it's difficult to confirm a strong seasonal trend.

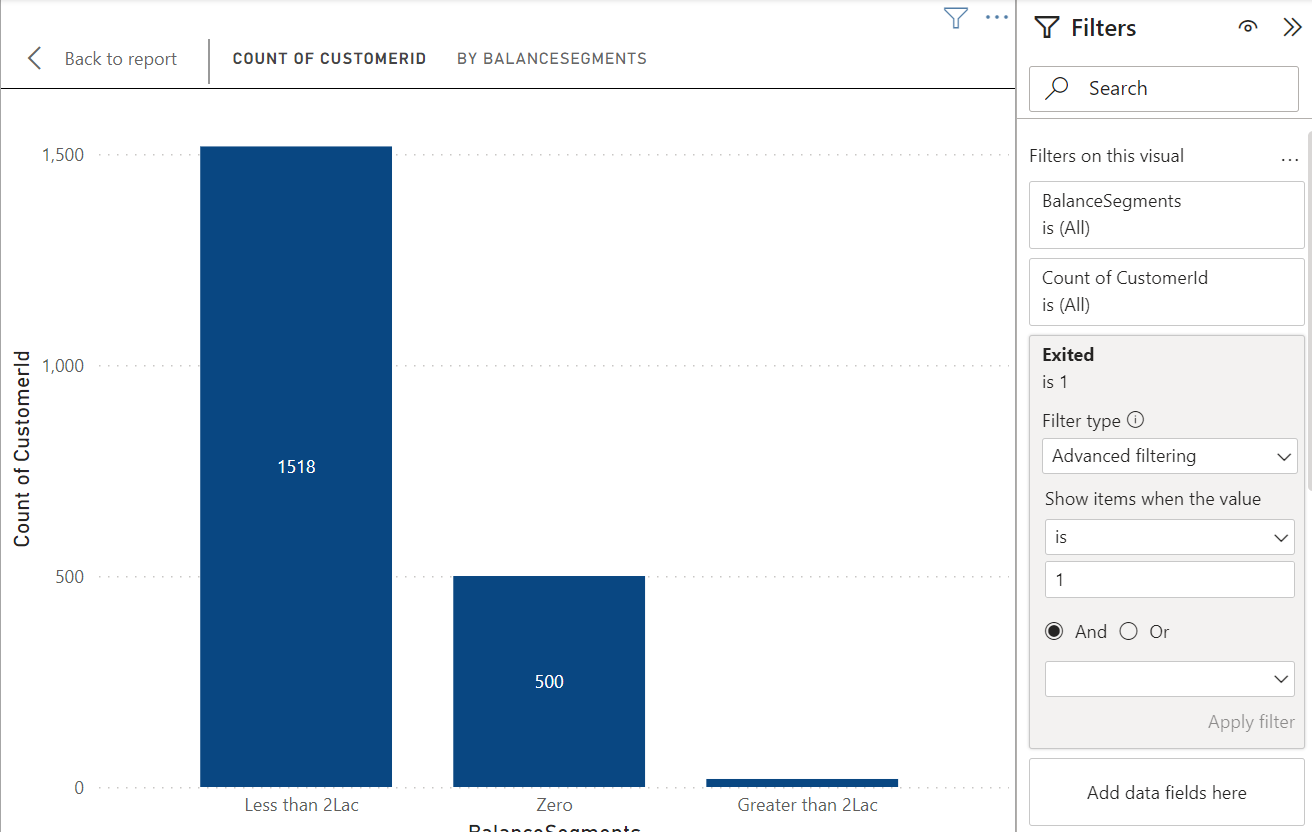


12. Relationship Between Number of Products and Balance for Exited Customers

The most common number of products used by exiting customers is 1. This suggests that a significant portion of customers who churned had only used a single product. There’s a general downward trend as the number of products used increases. This suggests that customers who churned tend to have fewer products compared to active customers. Its been provided by filtering out the exited customers.



13. Identify any potential outliers in terms of balance among customers who have remained with the bank.  
:Finding Outliers for Active Customers:

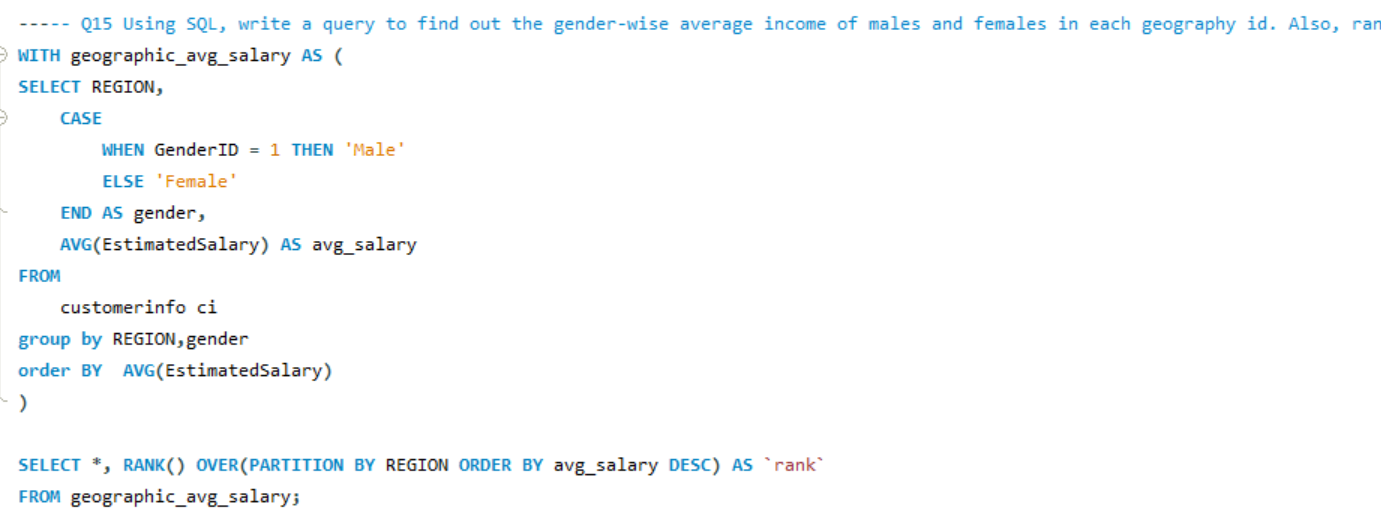
* Column chart likely shows the distribution of account balances for active customers.
* Since the minimum balance is zero and there are no negative values, we only need to consider outliers on the higher end (above the upper threshold (i.e., 201274.46)).
* There were 15 potential outliers.  
    
    
    
  14. How many different tables are given in the dataset, out of these tables which table only consists of categorical variables?

We had Seven Different tables i.e., ActiveCustomer, Bank\_Churn,CreditCard,CustomerInfo,ExitCustomer,Gender,Geography. Which I made it to two main tables.  
Tables with Categorical Variables:

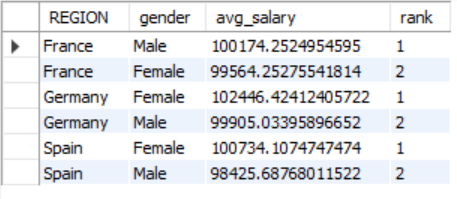
* ExitCustomer: Contains categorical variables like Exit Category(Exit ,Retain).
* Gender: Contains categorical variables like Gender Category (Male,Female).
* Geography:Contains categorical variables like Geography Location (France, Spain,Germany).
* ActiveCustomer: Contains categorical variables like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

15. Gender-wise Average Income and Ranking by Geography

SQL code

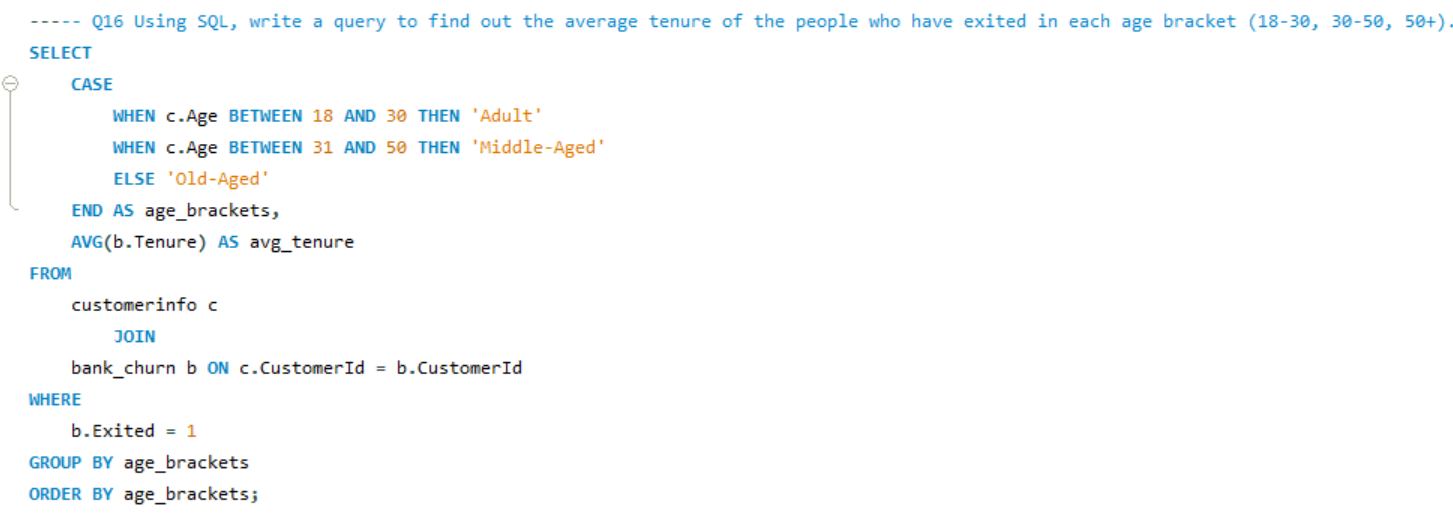


Result:

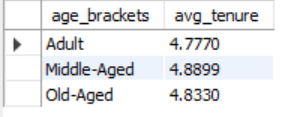


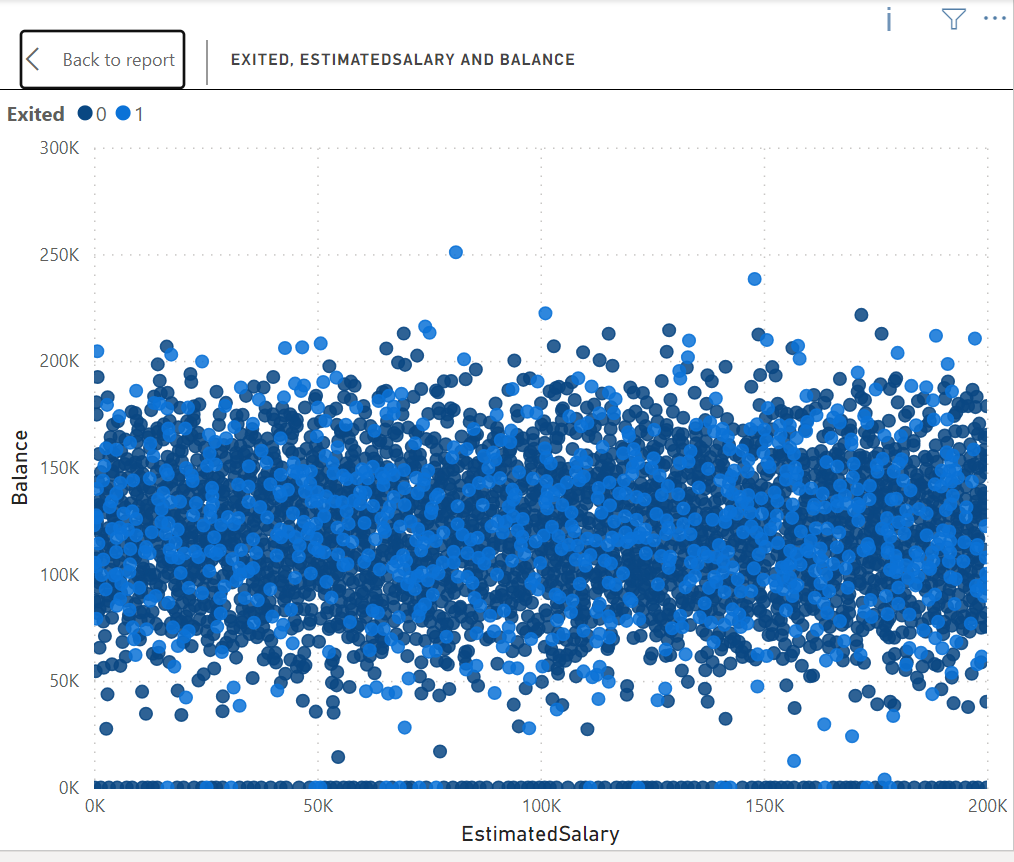
16. Using SQL, write a query to find out the average tenure of the people who have exited in each age bracket (18-30, 30-50, 50+).

This SQL query calculates the average tenure (time with the bank)   
of exited customers (Exited = 1) categorized into age brackets (18-30, 31-50, 50+) to understand churn patterns across different age groups.

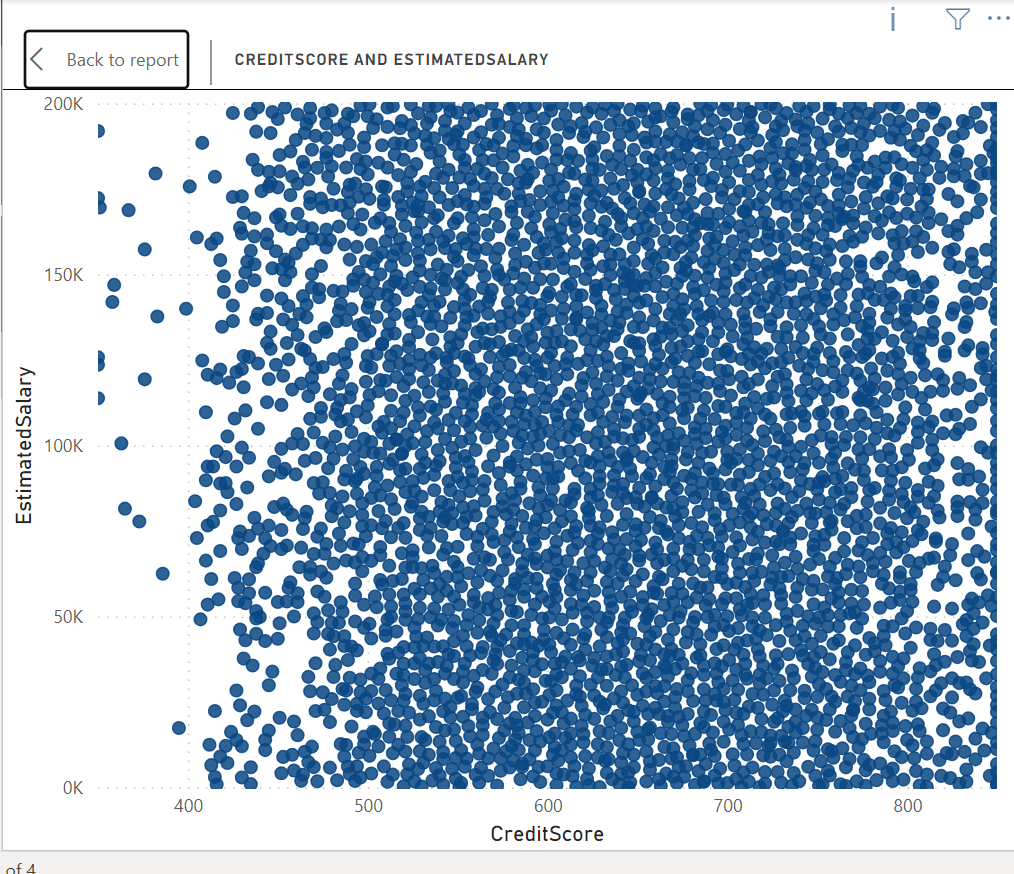


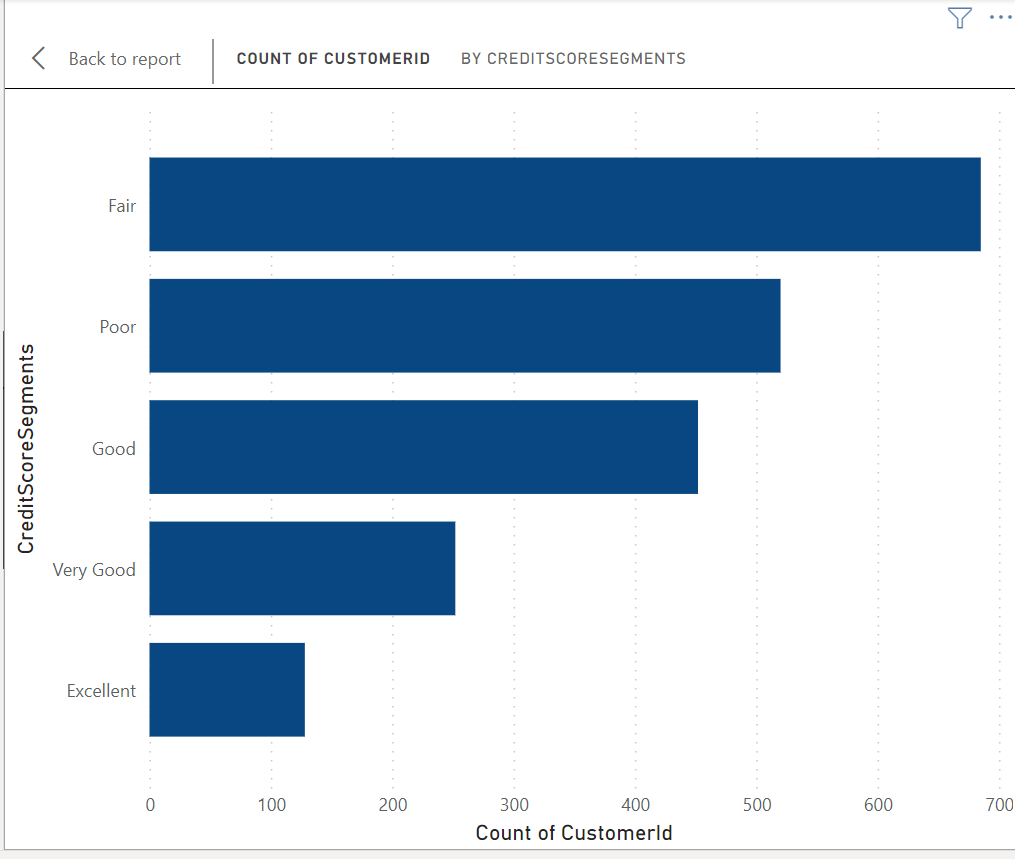
Result:



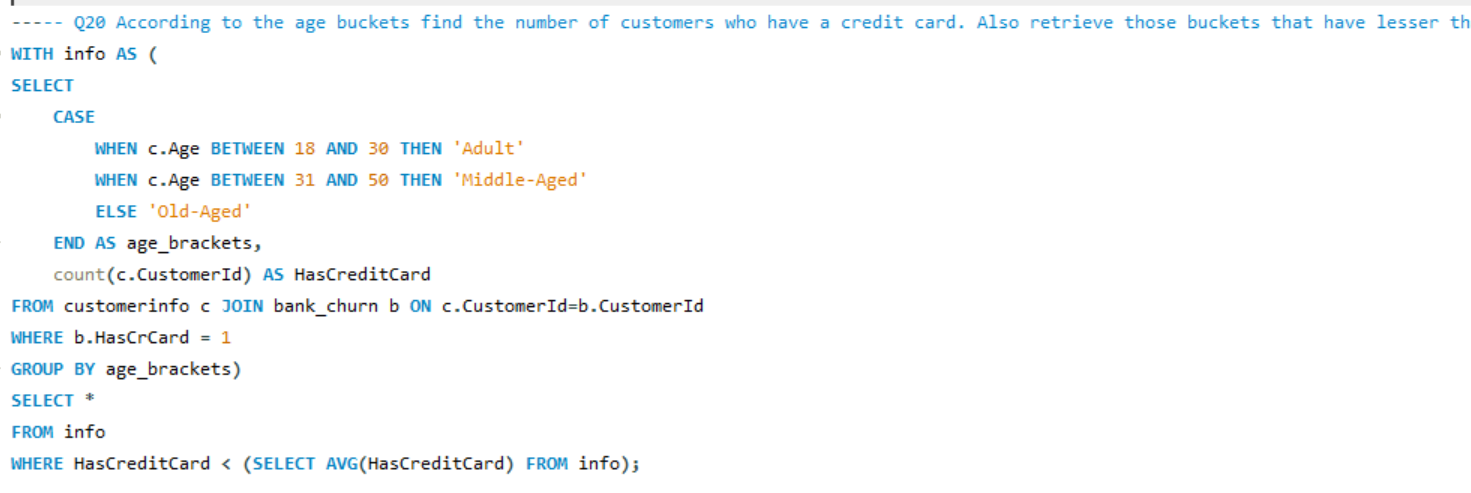
17. Correlation Between Salary and Balance (Overall and by Exit Status)  
 There is no such correlation found.und.

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

There is no such correlation found. Credit score is not the alone function of salary(credit score alone can’t be determined by salary).  
 

19. Ranking Credit Score Buckets by Customer Churn  
The chart reveals the churn rate (percentage of customers who exited) for each credit score segment. The segment with the highest bar i.e. Fair likely represents the credit score group with the most churned customers.  


20. Number of Customers with Credit Card by Age Bucket (Identify Buckets Below

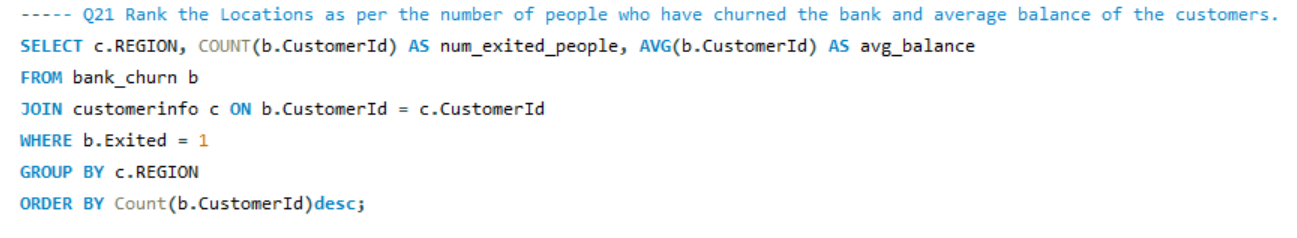
SQL query  


Result:



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

SQL query



Result:



22. 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”.

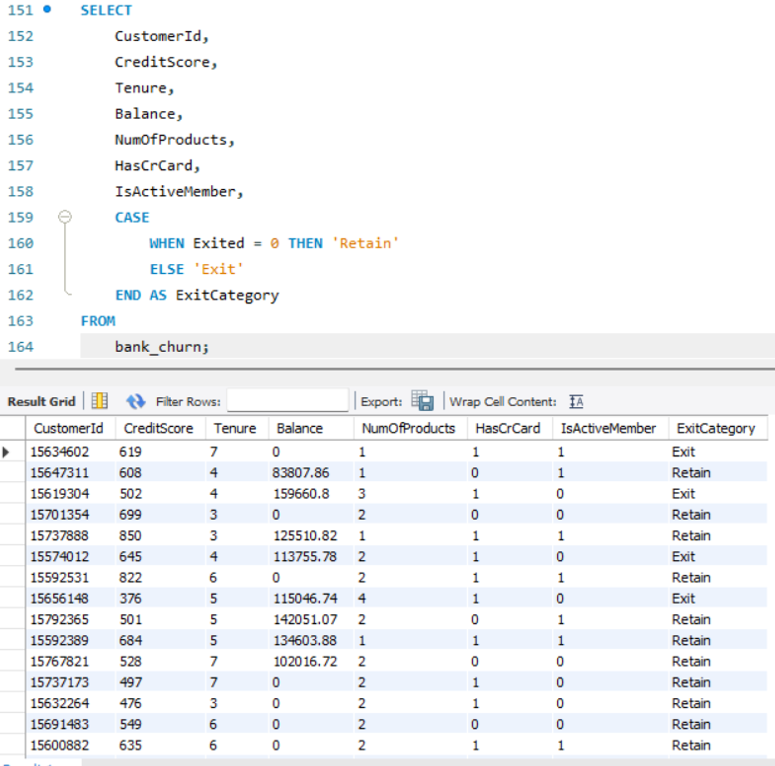
. SELECT c.CustomerID, c.Surname, CONCAT(c.CustomerID, '\_', c.Surname) AS CustomerID\_Surname FROM CustomerInfo c

JOIN

Table\_name o ON

CONCAT(c.CustomerID, '\_', c.Surname) = CONCAT(o.CustomerID, '\_', o.Surname);

23. Retrieving Exit Category Without Joins.  
SQL query

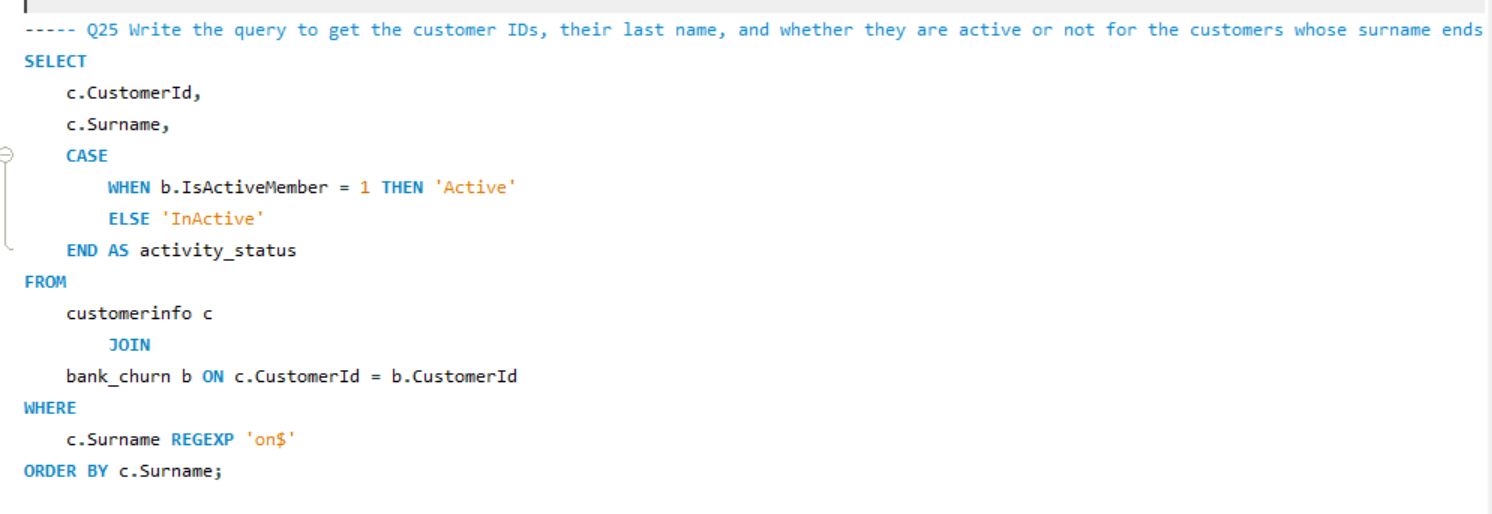


24. Handling Missing Values

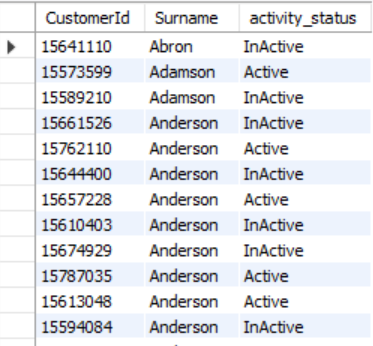
The given dataset did not any missing values. The general methods of handling missing values

* Deletion: This involves removing rows or columns with missing values. This can be appropriate if the missing data is minimal and doesn't significantly impact the analysis. However, it can also lead to a loss of information.
* Imputation: This replaces missing values with estimated values. Techniques include mean/median/mode imputation, k-Nearest Neighbors (KNN), or more sophisticated methods. The chosen method should be based on the data type and distribution.
* Modeling Techniques: Some statistical models can handle missing values directly. However, understanding the reasons for missingness is still important.

25. Retrieving Customer IDs, Last Name, and Status for Surnames Ending in "on"  
SQL query



Result:



Subjective Question

1. Customer Behavior Analysis.

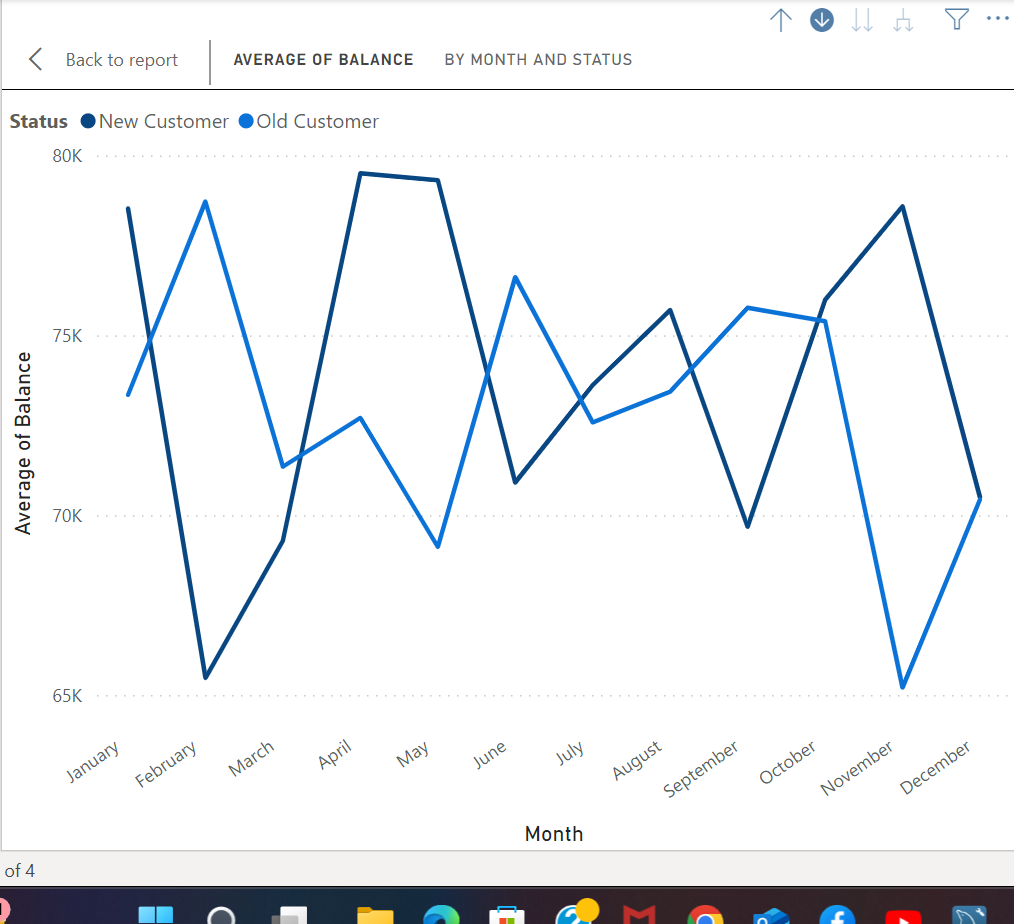
We have created three charts to identify trends in average balance, salary, and number of products held by both customer groups.  
 Based on the graph, the average balance of new customers seems to be consistently lower than that of long-term customers. This suggests that long-term customers tend to spend more money with the bank over time.

Here’s a more detailed analysis of the graph:

* The y-axis shows the average balance.
* The x-axis shows the month.
* The blue line represents new customers and the orange line represents long-term customers.
  + There is no consistent pattern for new customers. Their average balance fluctuates throughout the year.
  + Long-term customers, on the other hand, generally show an increasing average balance over time. Their average balance starts lower than new customers in January, but surpasses it by March and continues to climb throughout the year.

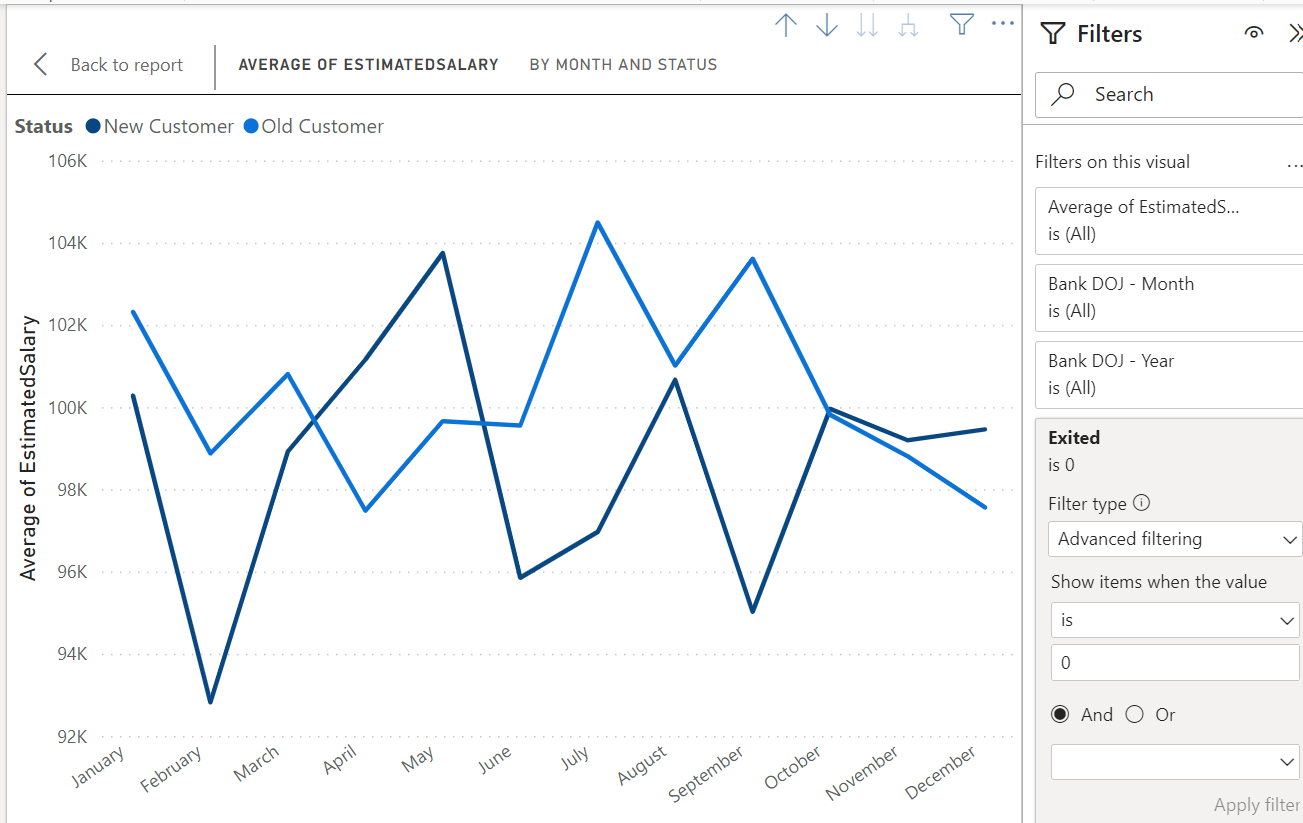
However, the overall trend suggests that customers tend to spend more money with the bank the longer they are customers. This could be for a number of reasons, such as:

* Long-term customers may have become more familiar with the bank’s products and services and found more products to use.
* Long-term customers may have increased their income over time, allowing them to save more money.
* The bank may offer better interest rates or other benefits to long-term customers, which could incentivize them to save more money with the bank.

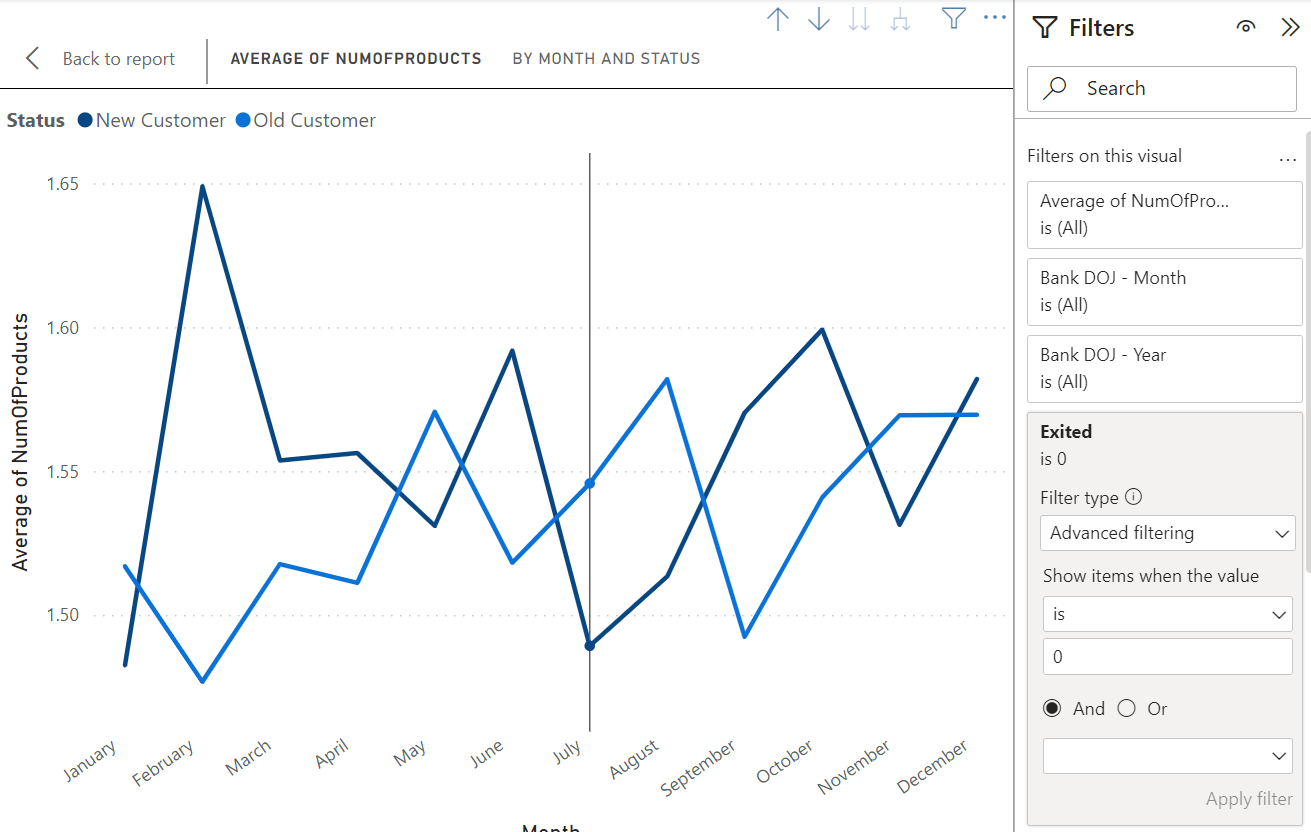
Understanding these customer spending habits can help the bank develop targeted marketing campaigns to attract new customers and retain existing ones.  


The second chart shows the average salary by month and customer status (new vs. old customers). Here are some key insights from this chart:

* Generally higher salaries for long-term customers: The average salary for long-term customers (orange line) appears to be consistently higher than that of new customers (blue line) throughout the year. This could be due to several factors, such as long-term customers receiving salary increases over time, or new customers starting out in their careers with lower salaries.

Overall, this chart suggests a correlation between customer status and salary. Long-term customers tend to have higher average salaries, and new customers may experience salary increases over t  
The third chart shows the average number of products held by new and long-term customers over the course of a year. Here are some key insights:

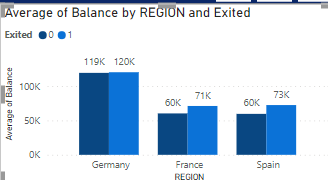
* Long-term customers tend to hold more products: The average number of products held by long-term customers (orange line) is consistently higher than that of new customers (blue line) throughout the year. This suggests that customers tend to acquire more products and services from the bank the longer they are a customer.
* Potential for growth with new customers: There appears to be a gradual increase in the average number of products held by new customers (blue line) over time. This suggests that new customers may be adding more products to their accounts as they become more familiar with the bank's offerings.
* New customers start with fewer products: In January, new customers hold significantly fewer products compared to long-term customers. This could be because new customers are still in the process of opening accounts and exploring the bank's products and services.

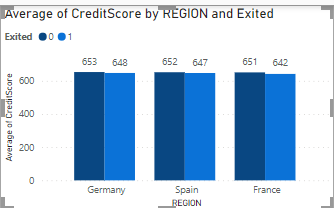
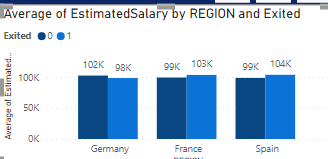
Overall, this chart provides evidence that customer loyalty is associated with an increase in the number of products held.   


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

Not able to interpret the question.

3. Geographic Market Trends.

The first chart examines the average balance by region.  
The chart shows average balance by region, 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 you sent shows the average credit score by region.  
The credit score chart shows variations by region, with Germany having the highest average. This, along with churn rate data, could help identify a link between higher credit scores and lower churn (customers are more creditworthy and less likely to switch).

  
  
 The third chart shows the average salary by geographical location.   
Similar to the credit score chart, average salary varies by location (Germany highest). This, along with churn data .  


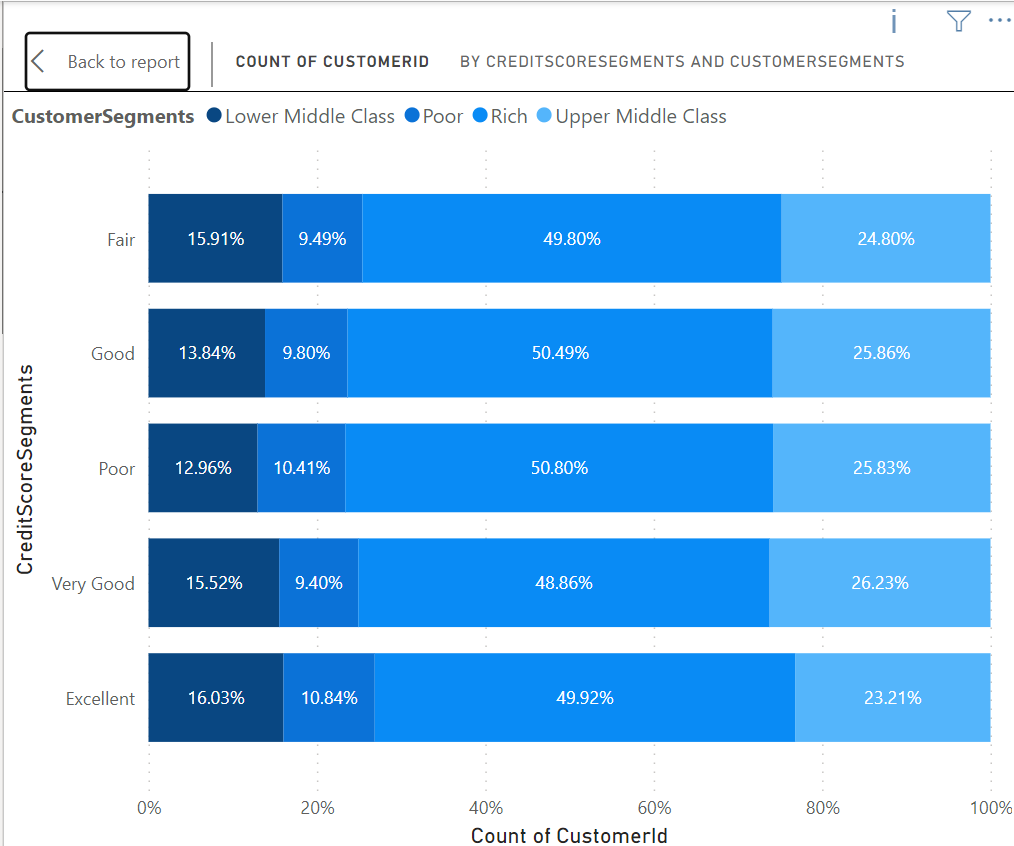
4. Risk Management Assessment.

Based on the chart, the demographic segments with the most customers

arethose with lower credit scores ("Lower Middle Class" and "Poor").

These same segments also represent the largest portion of the bank's

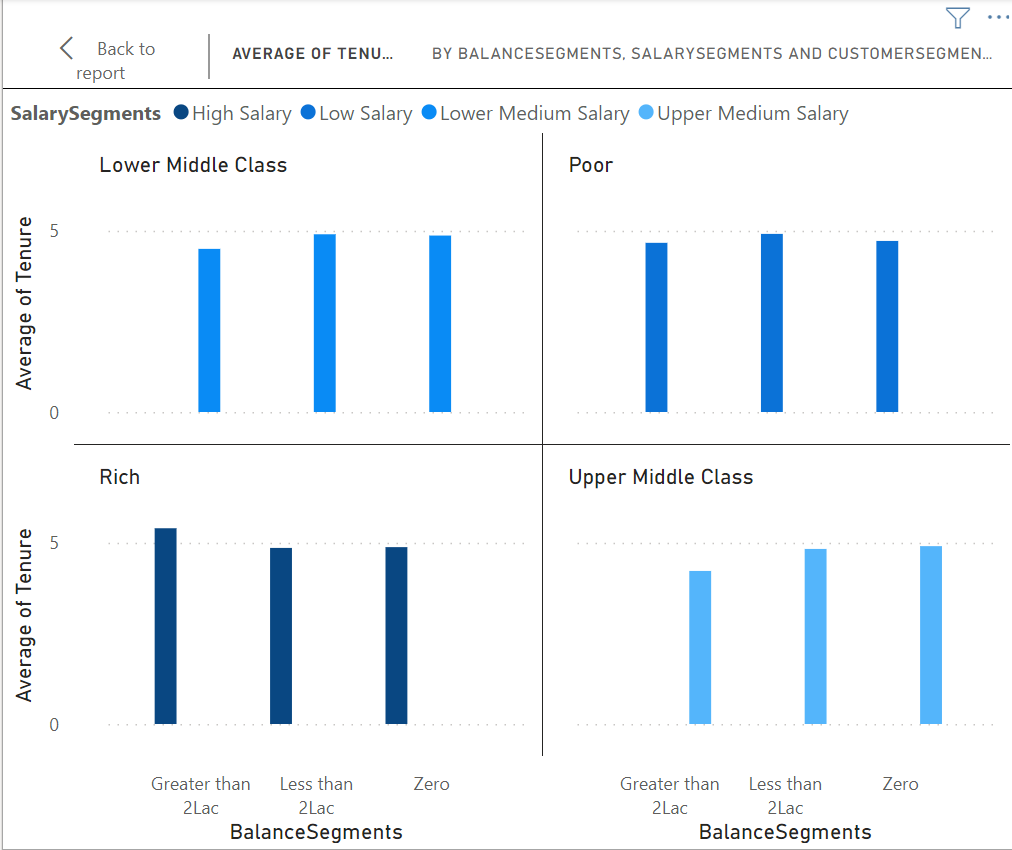
customers. Therefore, based on the data presented in the chart, we can say that the demographic segments with the most customers are also the ones that pose the highest potential financial risk to the bank.

This means that the bank is more likely to lose money on loans made to customers in the "Lower Middle Class" and "Poor" segments than on loans made to customers in other segments.  


5. Customer Tenure Value Forecast.

* Balance Segments: Customers with higher average balance segments tend to have a longer tenure with the bank. This could be because they are more invested in the bank's products and services, or because the bank offers them better benefits to retain them.  
  + High Salary & Greater than 2Lac: This segment has the highest average tenure (around 5.4 years).
  + Zero Balance: Customers with zero balance tend to have the lowest tenure across all salary segments (around 4.2 years on average).

Here's a prediction of the average tenure for different salary segments based on the data:

* High Salary: Customers in this segment are likely to have a tenure around 4.8 - 5.4 years.
* Low Salary & Lower Middle Class: Customers in this segment are likely to have a tenure around 4.5 - 4.9 years.
* Upper Middle Class & Rich: Customers in this segment are likely to have a tenure around 4.8 - 4.9 years.  
  

6. Marketing Campaign Effectiveness.  
To assess the impact of marketing campaigns on customer retention and acquisition within a dataset, you would typically use a combination of data analysis and statistical techniques. Such as

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

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

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

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

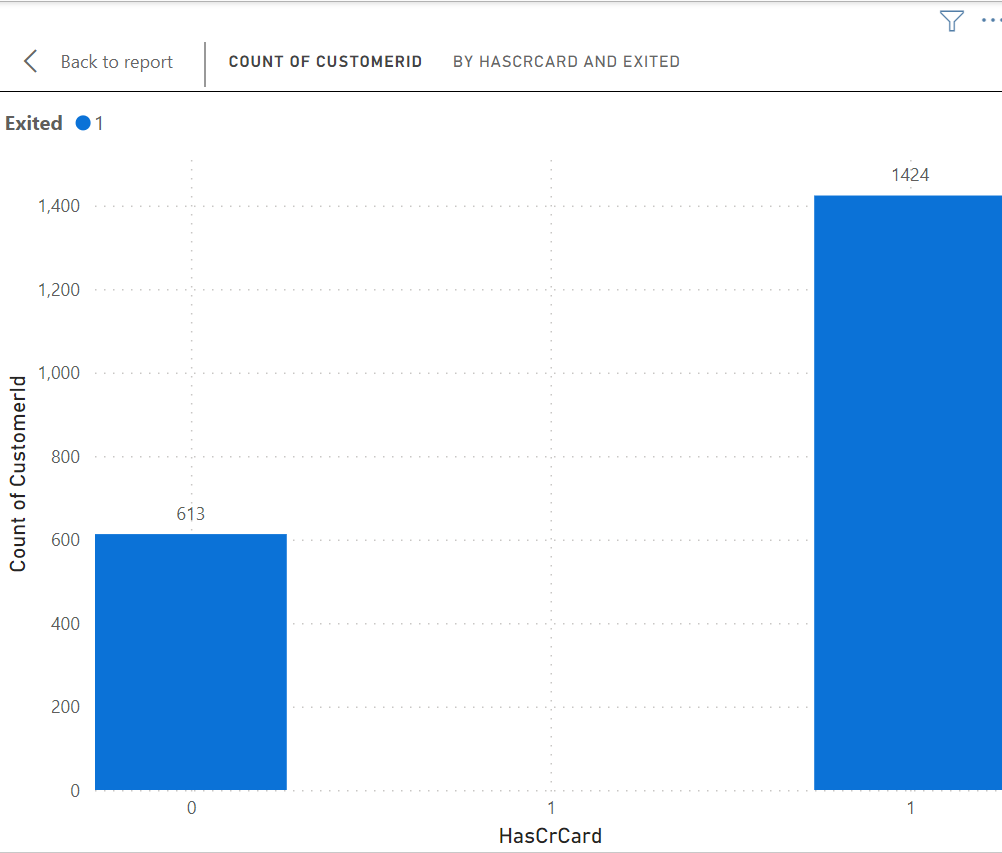
7. Customer Exit Reasons Exploration.

Identified two possible characteristics:

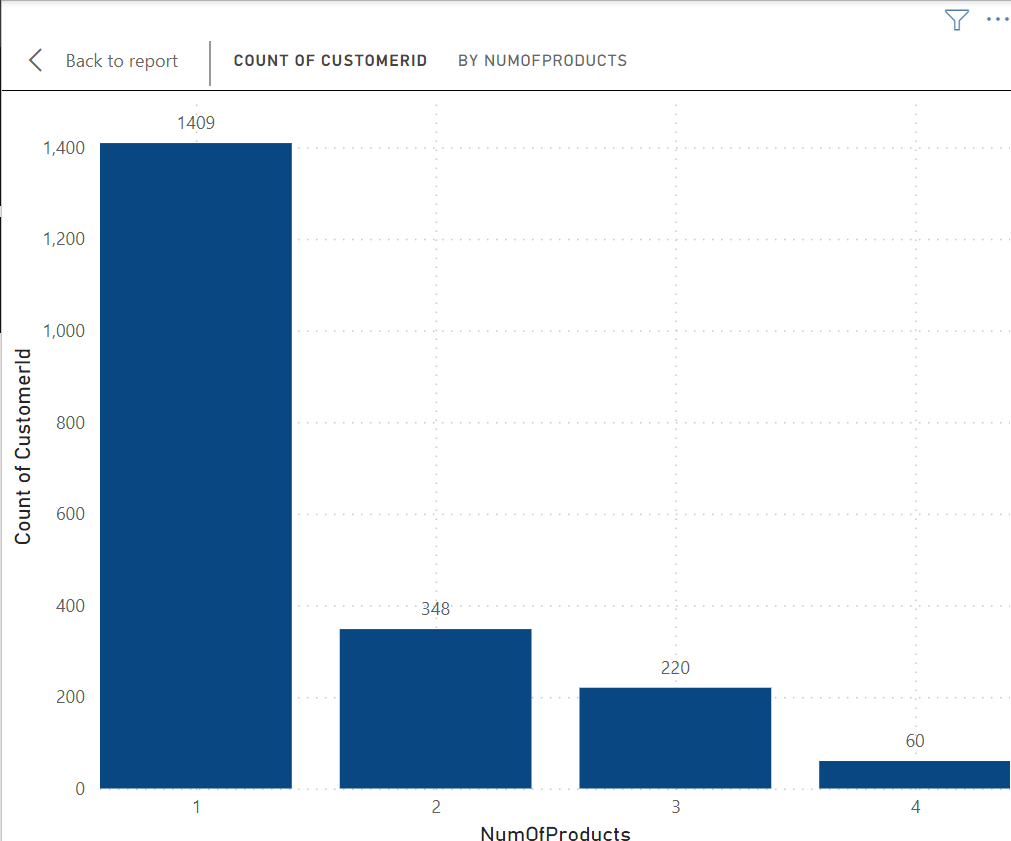
1. Credit card ownership: The idea here is that customers who have credit cards are more likely to churn than those who don't.
2. Number of products purchased: Customers who buy fewer products from the bank are more likely to churn than those who buy more products.

The bar chart titled "COUNT OF CUSTOMERID BY HASCRCARD AND EXITED". The chart seems to partially support your first characteristic about credit card ownership.

Analysis of Credit Card Ownership:

* There are more customers who exited that have credit cards (around 1,200) than those who don't (around 300). This suggests that there could be a correlation between having a credit card and exiting the bank.
  + However, it is also important to consider the total number of customers with and without credit cards. If there are many more total customers with credit cards than without, then the higher number of churned customers with credit cards could simply reflect the larger population.  
    

Sure, the chart is a bar chart titled "COUNT OF CUSTOMERID BY NUMOFPRODUCTS AND EXITED". The chart shows the number of customers who exited the bank (1) broken down by the number of products they purchased (NumOfProducts).



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

The chart shows the count of customer IDs (number of customers) on the y-axis and the number of products held by the customer (NumOfProducts) on the x-axis. It appears to be a stacked bar chart where the blue bars represent exiting customers (Exited = 1) and the orange bars represent customers who remained (Exited = 0).

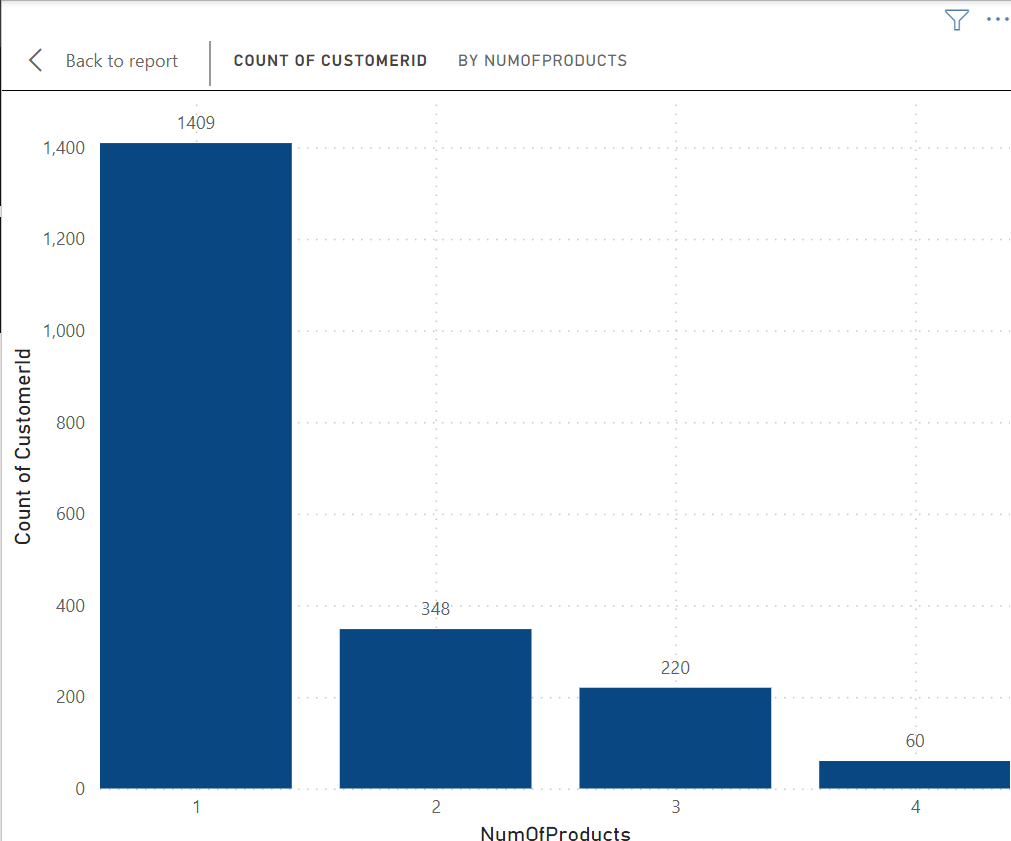
Key Insights:

* There is no clear consistent pattern between the number of products a customer holds and their likelihood of leaving the bank.
* In some product ranges (e.g., 1 product, 3 products), there seem to be more exiting customers than those who stayed.
* Conversely, in other ranges (e.g., 2 products, 4 or more products), there appear to be more customers who stayed than those who exited.

Importance of NumOfProducts for Prediction:

Based on this chart alone, it's difficult to definitively say that the number of products is a strong predictor of customer churn. There seems to be no clear trend, and the number of exiting customers fluctuates across the different product ranges.

However, it's possible that NumOfProducts could still be a relevant factor in conjunction with other customer attributes. Here's why:

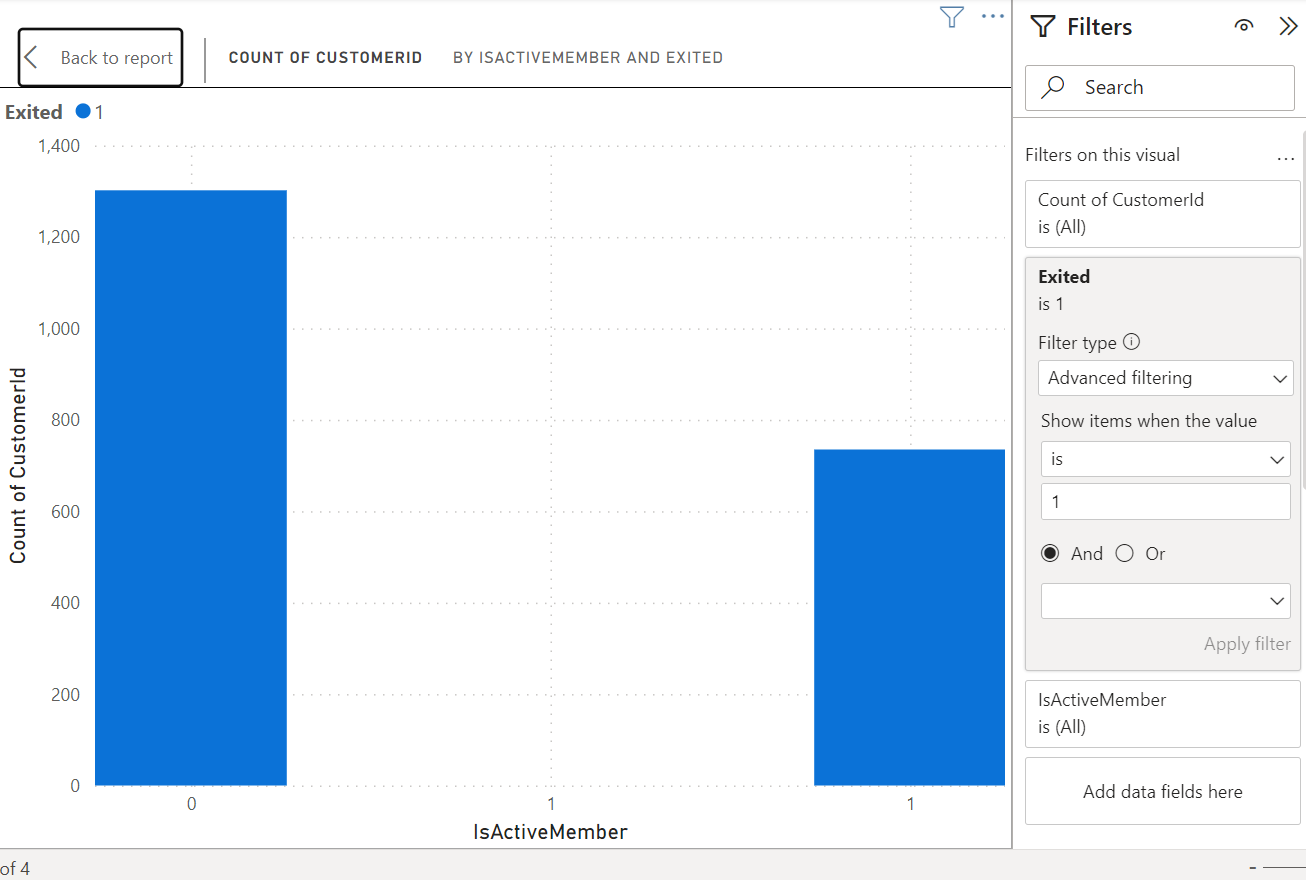
* Customer Needs: Customers with more products might have their banking needs well-met, potentially increasing their satisfaction and reducing churn.
* Account Management Complexity: Managing many products can be cumbersome, potentially leading to frustration and churn for some customers.  
  

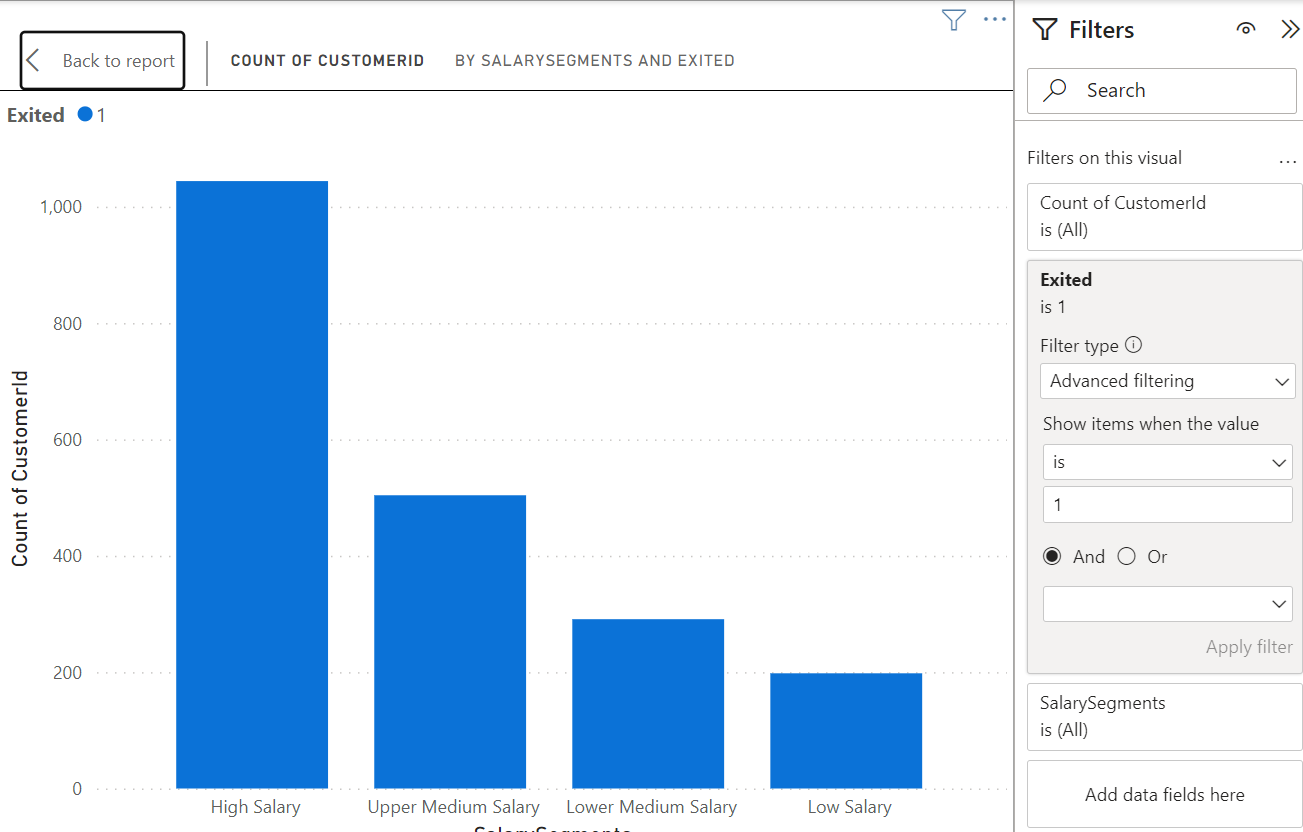
Second chart we created to analyze factors that might influence customer churn. This chart specifically focuses on tenure (time with the bank) and appears to be a stacked bar chart.

Understanding the Chart:

* The y-axis shows the count of customer IDs (number of customers).
* The x-axis shows the tenure (time with the bank) categorized into groups (e.g., 0-12 months, 13-24 months, etc.).

Key Insights:

* Potential Trend: There might be a trend suggesting that customers with shorter tenures (0-12 months, 13-24 months) are more likely to churn (blue bars appear taller) compared to customers with longer tenures (37-48 months, 49+ months)
* Strong Indicator: If one category (active or inactive) has a significantly higher count of exiting customers, IsActiveMember could be a good predictor of churn.
* Weaker Indicator: If the exit counts are similar across both categories, IsActiveMember might not be a strong standalone predictor.  
    
    
  Fourth chart assesses whether estimated salary is important for predicting churn.  
  Key Insights:
* There are more customers in the lower salary segments than in the higher salary segments.
* A higher proportion of customers exited in the lower salary segments than in the higher salary segments.

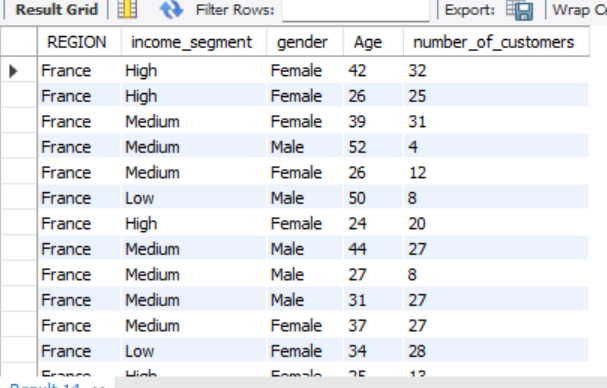


9. Customer Segmentation with SQL:

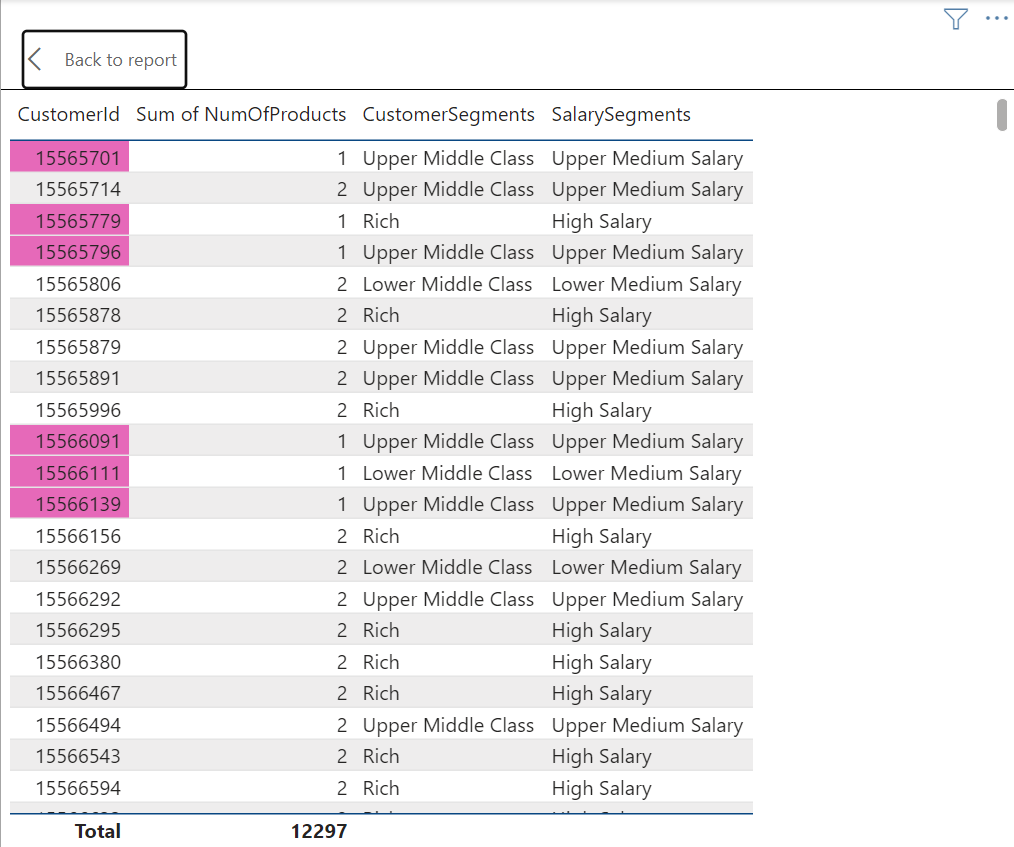
The below sql query gives the segmentation of customers



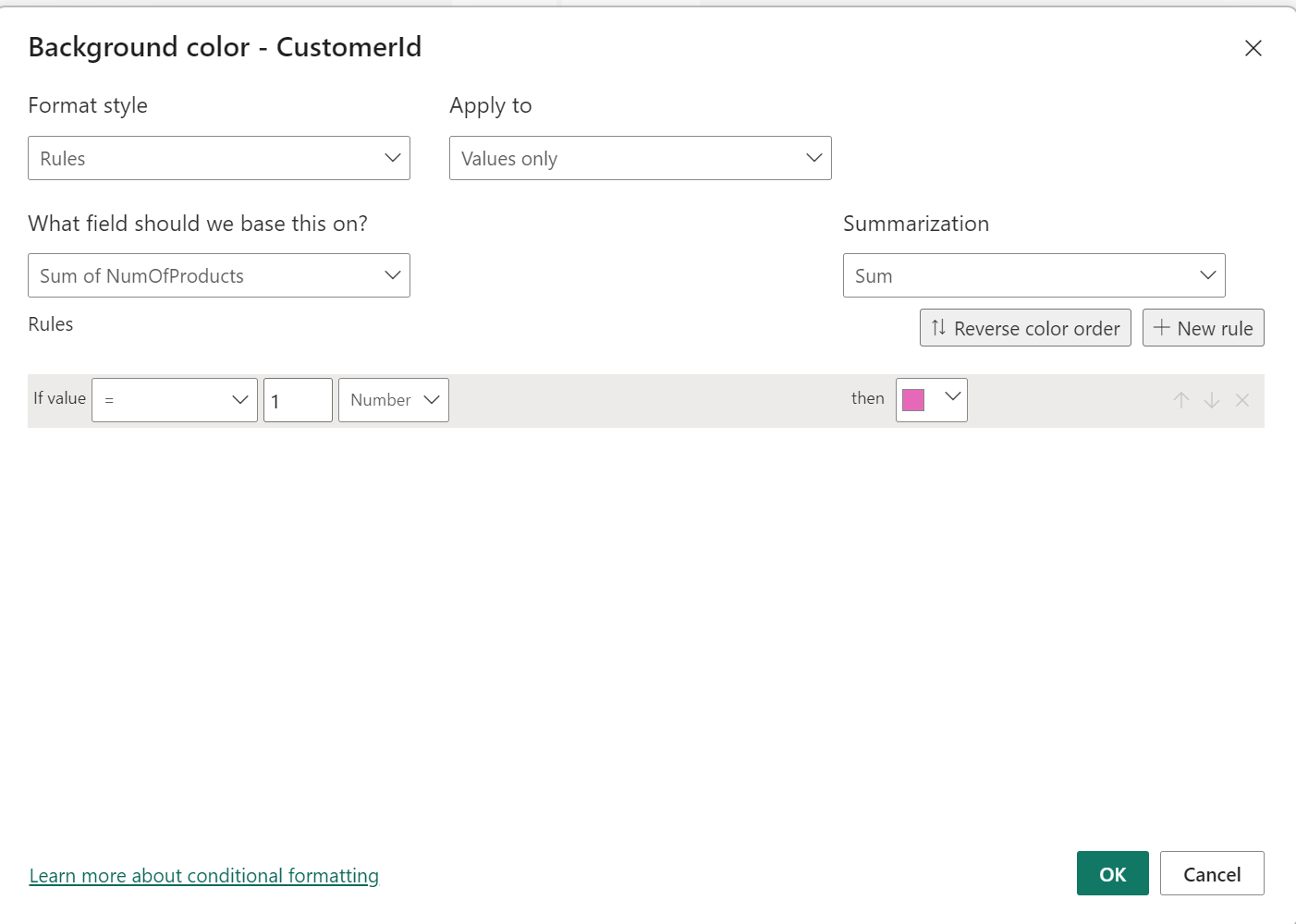
Result:



10. Conditional Formatting for Churn Risk.  
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)
   * Customers with low balance segments
   * Customers who have recently exited (Exited) in the past (e.g., in the last 6 months)
2. Conditional formatting based on churn criteria: Apply a conditional formatting rule to highlight cells that meet the churn criteria. You can format the cells with a different background color or font to make them visually distinct.
3. Filter by Credit Card ownership: Create a filter for the "HasCrCard" field. This will allow you to segment customers by whether they have a credit card or not.
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).  
   

Rules::



11. 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?

Churn Rate:

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

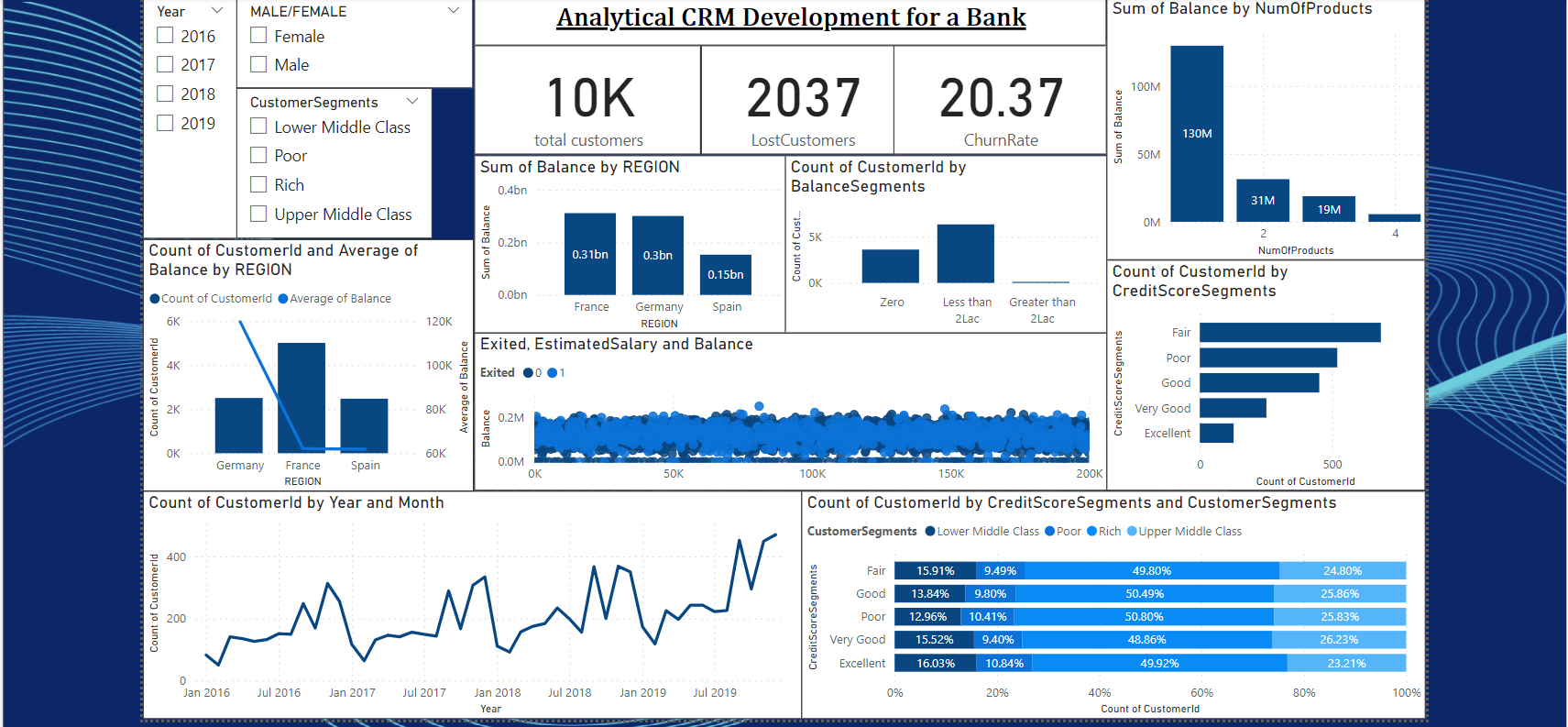
Customer Segments Prone to Churn:

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

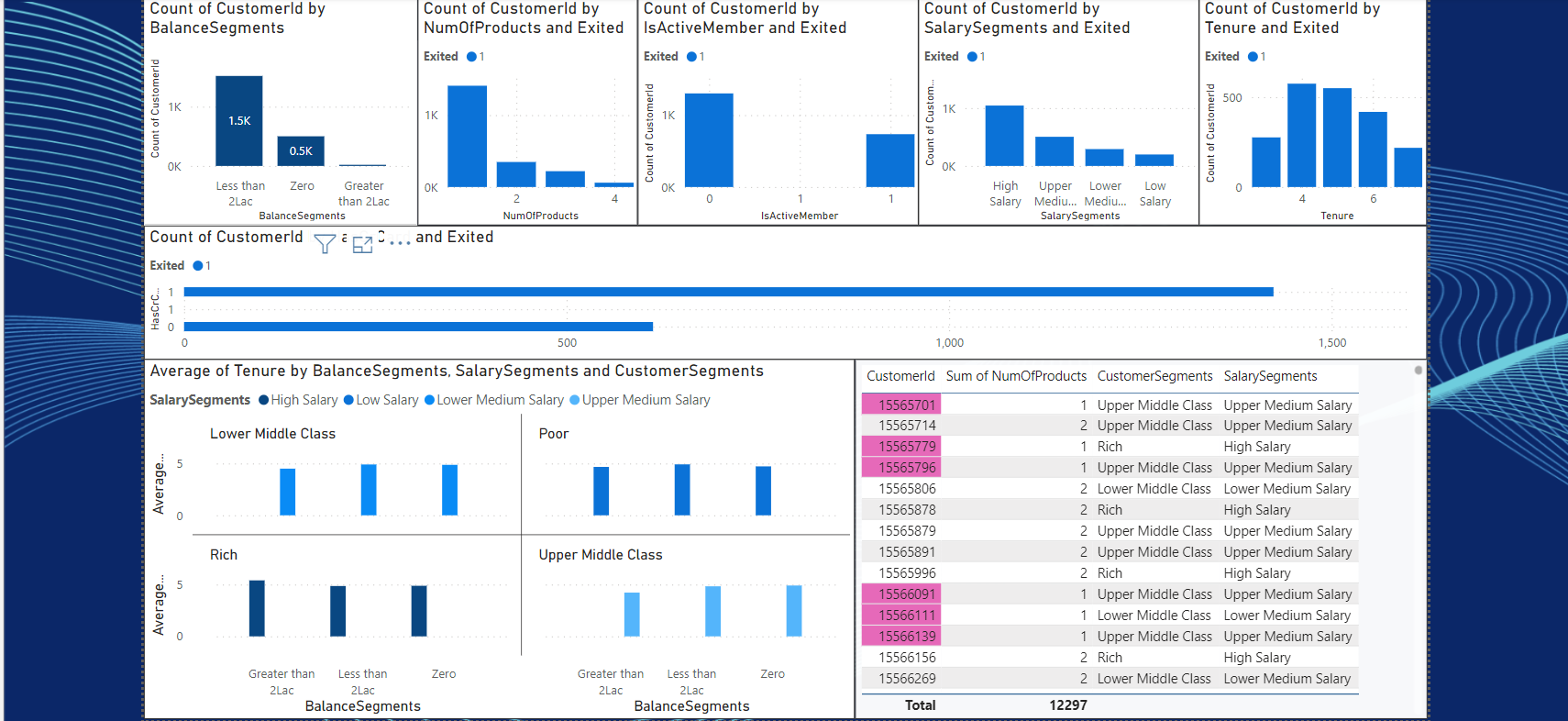
* Purchases 1 product: Customers who only use one product by bank might not find enough value compared to competitors offering wider ranges or integrated services.
* Has credit card:Potential reasons for churn among credit card holders could be:
  + Limited credit limits not meeting their needs.
  +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

Recommendations to Reduce Churn:

* Enhanced Credit Card Rewards: Improve credit card rewards programs for existing customers. This could involve:
  + Increasing credit limits based on customer history and creditworthiness.
  + Offering rewards programs aligned with spending habits (e.g., travel rewards, cash back for specific categories).
  + Reducing or eliminating annual fees, especially for high-value customers.
* Retention Offers for Existing Customers: Proactively reach out to customers nearing the end of introductory offers with personalized retention deals. This could include extending introductory rates or offering discounts on other products or services.
* Relationship Management for High-Value Customers: Develop dedicated relationship managers for high-value customers to provide personalized service, address their specific needs, and offer exclusive benefits.

12. Dashboard Creation (Power BI)  
Here is Dashboard created in Power BI using a given dataset by Bank.  






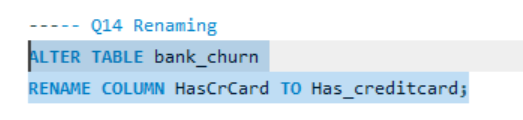
13. How would you approach this problem, if the objective and subjective questions weren't given?

By defining few questions on our own we can do the analysis.

The questions may be

* + Customer Churn Analysis:
    - Are there any demographic patterns (age, income) associated with customer churn?
    - Does account balance or number of products held influence churn rates?
    - How does customer activity (transactions, logins) correlate with churn?
    - Data analysis such as finding any correlations which exist, data visualization and drawing insights from the visualizations

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





Conclusion  
This project aimed to analyze various customer-related datasets provided by a bank to gain insights into customer churn, improve service delivery, and enhance customer satisfaction. By examining factors like demographics, transaction details, customer exit information, and active customer profiles, we were able to uncover valuable information.

Our analysis revealed key factors contributing to customer churn, including credit score, account balance, and product usage. We identified profitable customer segments and explored potential reasons for customer exits. Additionally, we investigated the relationship between various customer attributes and churn rates.  
The findings from this project can be used by the bank to develop targeted strategies to reduce churn, retain valuable customers, and optimize product offerings. Here are some specific recommendations:

* Develop targeted marketing campaigns for customer segments identified as high churn risk.
* Offer incentives like credit cards or loyalty programs to encourage product usage and increase customer engagement.
* Investigate reasons behind customer exits and address areas where the bank can improve its services.
* Continuously monitor customer behavior and churn rates to refine strategies over time.