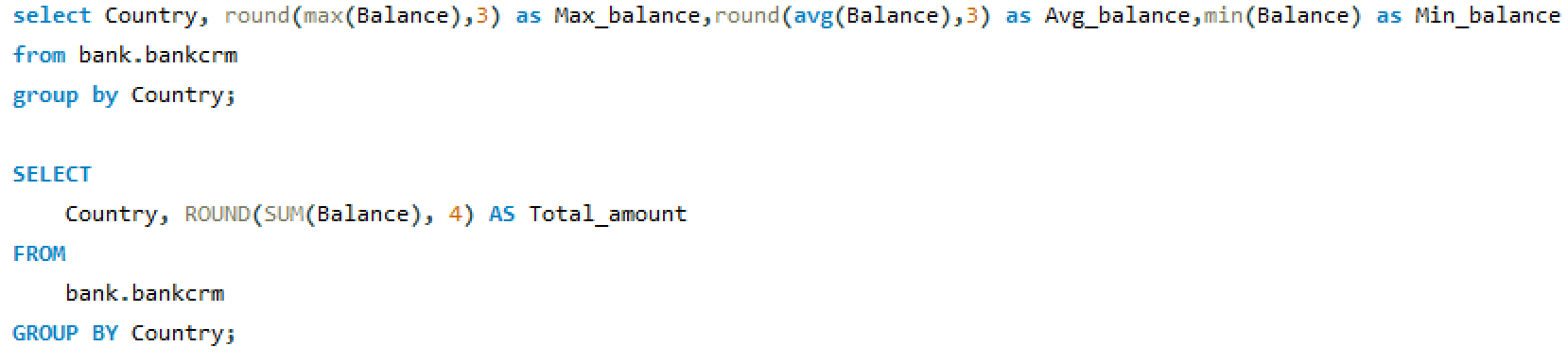
****

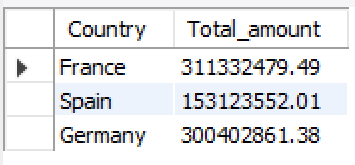
**By:- Prathmesh Satish Potdar**

Bank CRM Analysis

**Objective Questions:**

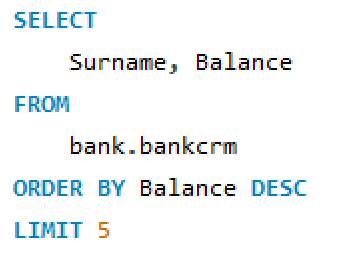
1. **What is the distribution of account balances across different regions?**

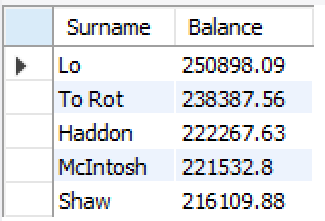
****



* France has the highest maximum balance, which is approximately 311,332,479.
* Spain has the highest average balance, at approximately 153,123,552.01.
* Germany has the lowest balance.

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

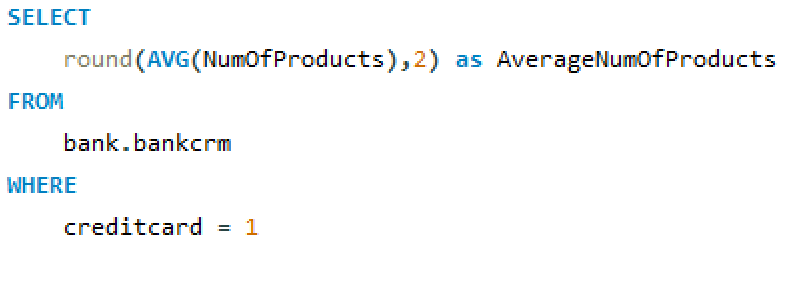
****

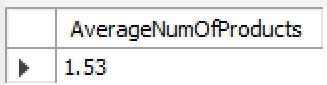


Top 5 customers whith highest Estimated Salary are:

* Lo, Balance: $250,898.09
* To Rot Balance: $238,387.56
* Haddon ,Balance: $222,267.63
* McIntosh ,Balance: $221,532.80
* Shaw ,Balance: $216,109.88

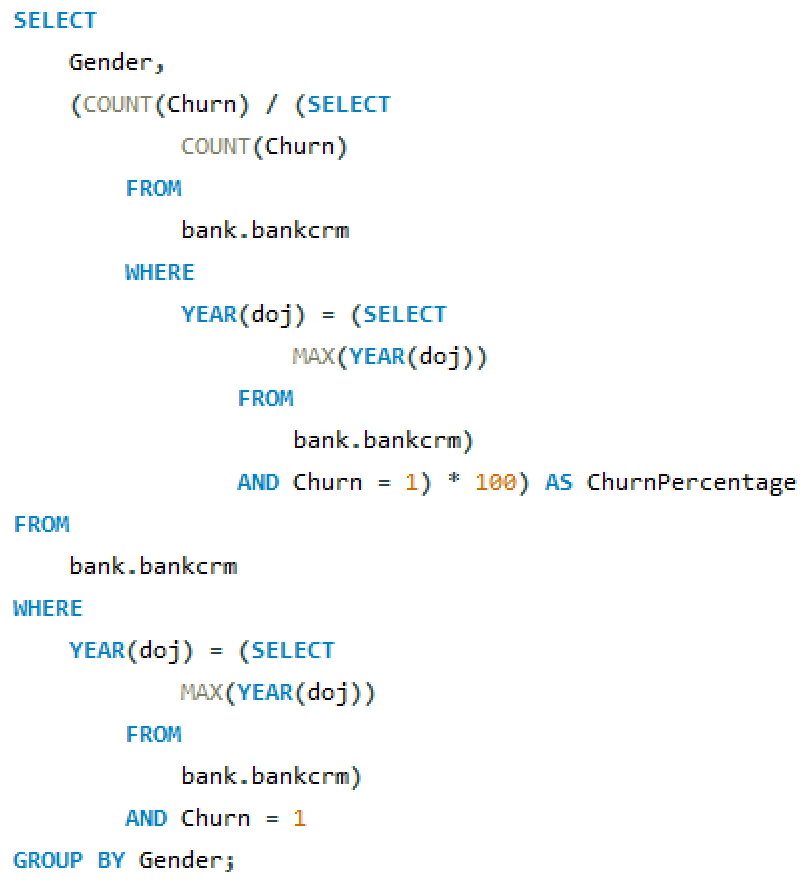
1. **Calculate the average number of products used by customers who have a credit card. (SQL)**

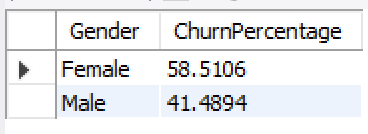
****



So, on average, customers who have a credit card are using approximately 1.53 products from the bank.

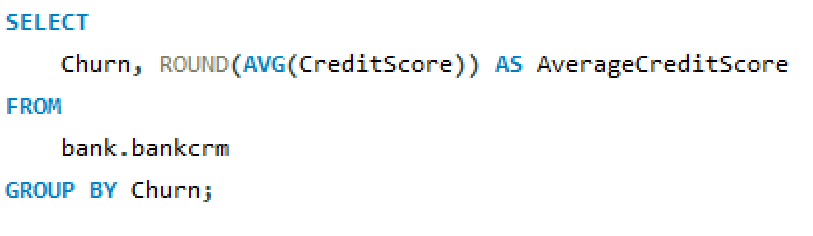
1. **Determine the churn rate by gender for the most recent year in the dataset.**

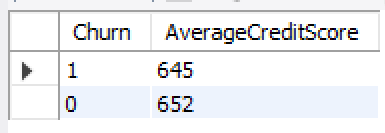
****



* Female Churn Percentage: Approximately 58.51%
* Male Churn Percentage: Approximately 41.49%
* In this case, the query indicates that the churn rate for females is higher compared to males in the most recent year, which could prompt further analysis and targeted interventions to reduce churn, particularly among female customers.

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

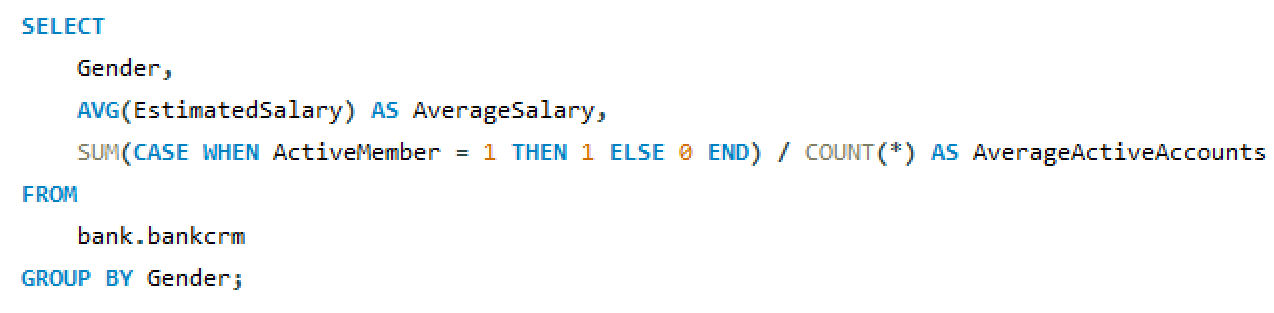
****

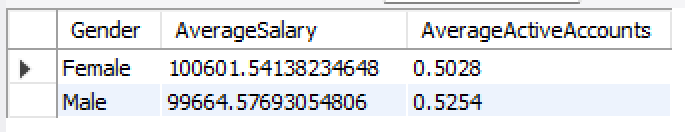


* Churned Customers (Churn = 1): The average credit score is approximately 645.
* Non-Churned Customers (Churn = 0): The average credit score is slightly higher, around 652.

These results suggest that, on average, non-churned customers tend to have slightly higher credit scores compared to churned customers. This insight could be valuable for the bank's customer retention strategies, as it indicates a potential correlation between credit score and churn behavior.

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

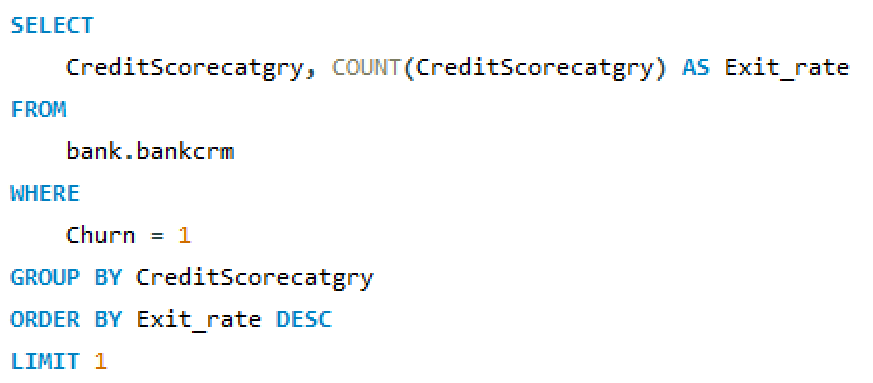
****

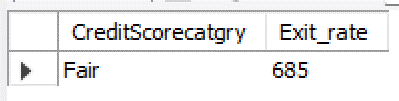


* **Average Salary by Gender**:
  + Female: The average salary for female customers is approximately $100,601.54.
  + Male: The average salary for male customers is approximately $99,664.58.
* **Average Active Accounts by Gender**:
  + Female: On average, female customers have approximately 0.503 active accounts.
  + Male: On average, male customers have approximately 0.525 active accounts.

These results offer a comparative analysis between genders regarding their average salary and the average number of active accounts. It appears that there's a slight difference in both salary and active accounts between male and female customers. Such insights can be valuable for targeted marketing strategies or product offerings tailored to each gender segment.Top of Form

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

****



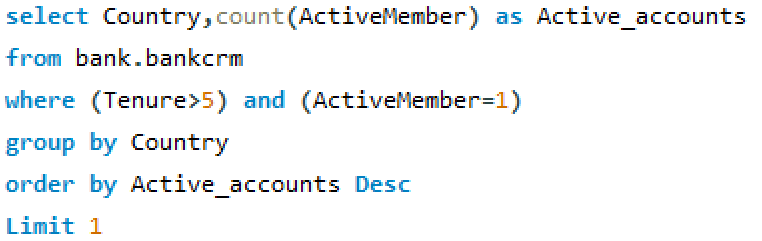
* **Exit Rate by Credit Score Category**:
  + Credit Score Category: Fair
  + Exit Rate: The highest exit rate, with approximately 685 customers falling into this category.

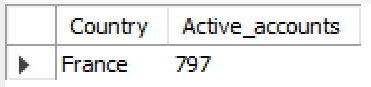
This suggests that customers categorized as having a "Fair" credit score are more likely to churn compared to customers in other credit score categories. Understanding this trend can help the bank identify at-risk customers and implement targeted retention strategies, such as offering personalized services or incentives to mitigate churn among customers with fair credit scores.

Top of Form

1. **Find out which geographic region has the highest number of active customers**

**with a tenure greater than 5 years. (SQL)**

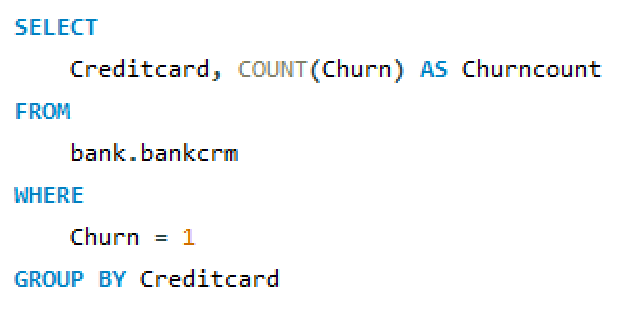
****

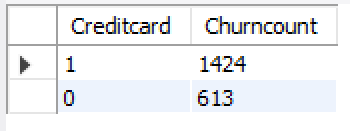


* **Country with the Highest Number of Active Accounts**:
  + Country: France
  + Active Accounts: 797

This indicates that among countries in the dataset, France has the highest number of active accounts among customers who have been with the bank for more than 5 years and are active members. Understanding this can help the bank identify regions with high customer engagement and potentially allocate resources for further growth or retention strategies.

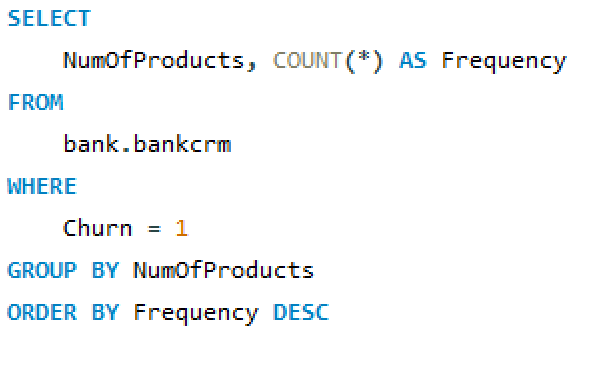
1. **What is the impact of having a credit card on customer churn, based on the available data?**

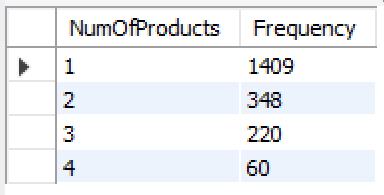
****



* **Churn Count by Credit Card Status**:
  + Customers without a credit card: Churn count is 1424.
  + Customers with a credit card: Churn count is 613.
* The churn count for customers without a credit card is substantially higher compared to those with a credit card. This suggests that customers without a credit card are more likely to churn.
* The availability or ownership of a credit card might play a role in customer retention. Further analysis could delve into the reasons behind this discrepancy and explore strategies to mitigate churn among customers without credit cards.

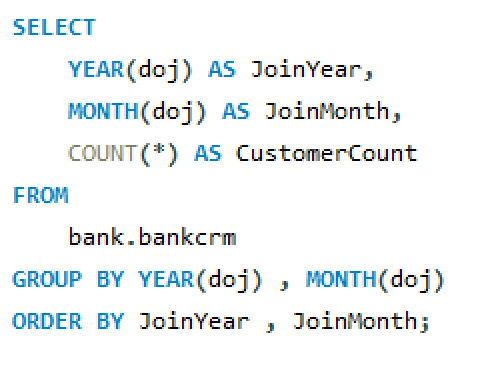
1. **For customers who have exited, what is the most common number of products they have used?**

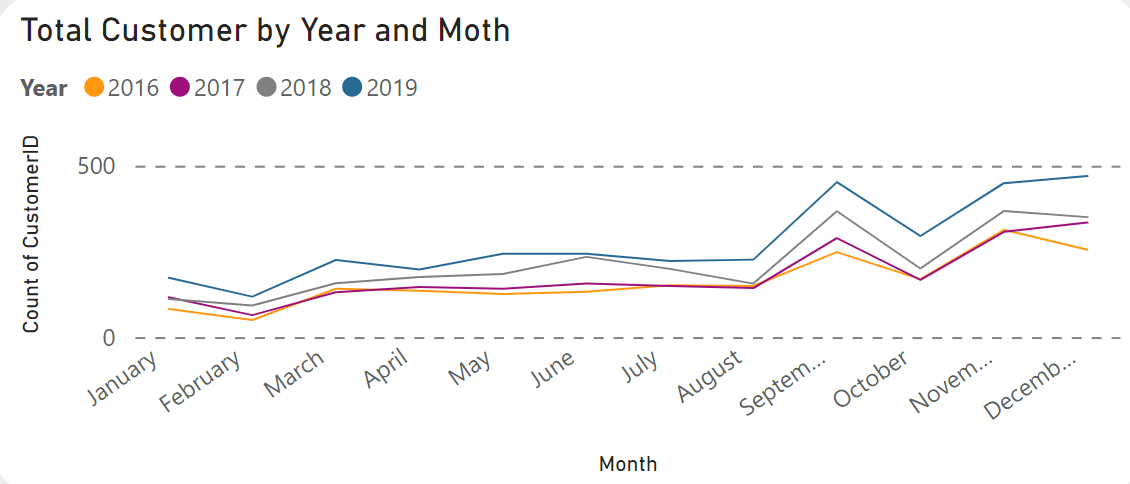
****

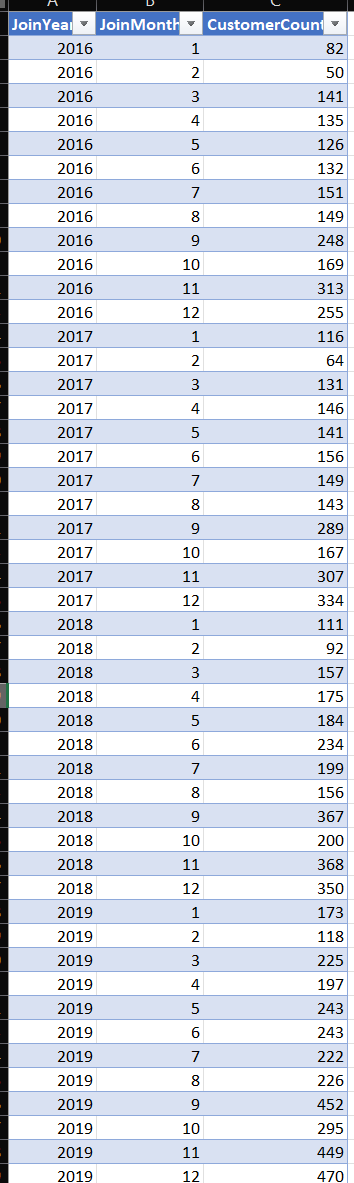


* **Churn Count by Number of Products**:
  + Customers with 1 product: Churn count is 1409.
  + Customers with 2 products: Churn count is 348.
  + Customers with 3 products: Churn count is 220.
  + Customers with 4 products: Churn count is 60.
* There is a clear trend indicating that as the number of products per customer increases, the churn count generally decreases.
* Customers with only one product are more likely to churn compared to those with multiple products. This suggests that offering additional products or services to customers may contribute to higher retention rates.
* Understanding this relationship can help the bank tailor its product offerings and retention strategies to address the needs and preferences of different customer segments

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.**

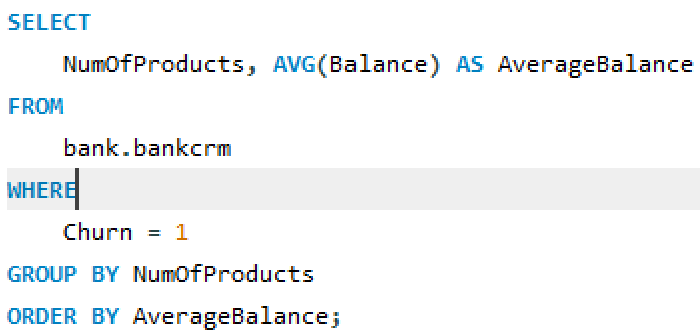
****

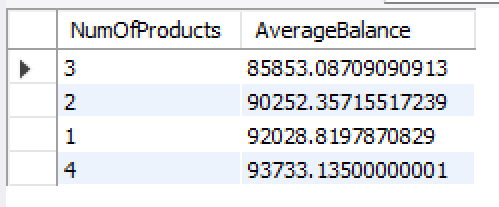




* **Seasonal Trends**:
  + Certain months tend to have higher customer counts, possibly indicating seasonal trends or marketing initiatives during those periods.
  + For example, in some years, months like November and December seem to have higher customer counts, possibly due to end-of-year promotions or incentives.
* **Yearly Growth**:
  + Overall, there is a noticeable increase in customer counts from 2016 to 2019, indicating growth in the bank's customer base over time.

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

****

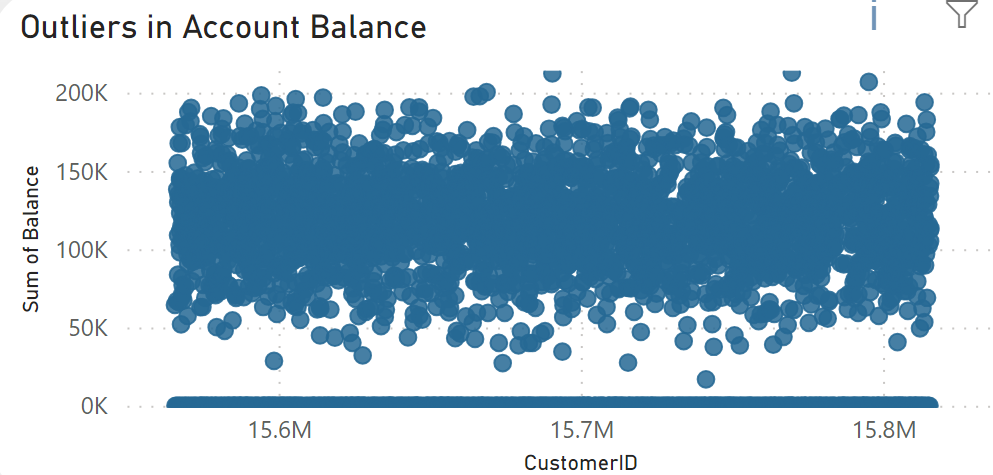


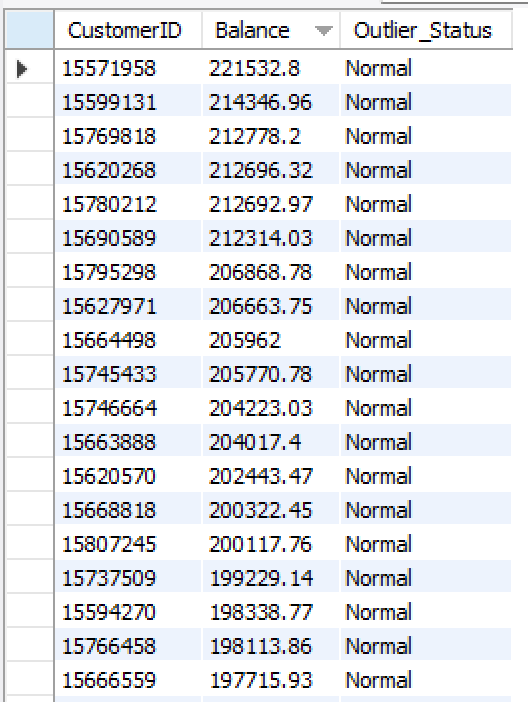
* **Relationship between NumOfProducts and AverageBalance**:
  + Customers with 3 products have the lowest average balance of approximately $85,853.09.
  + As the number of products increases, the average balance tends to increase:
    - Customers with 2 products have an average balance of approximately $90,252.36.
    - Customers with 4 products have an average balance of approximately $92,028.82.
    - Customers with 1 product (not included in the provided data) could be assumed to have the highest average balance among the analyzed groups.
* There seems to be a positive correlation between the number of products a customer holds and their average balance, with customers holding more products generally having higher

average balances.

* This insight can inform the bank's marketing and product development strategies, highlighting the potential for upselling or cross-selling to customers with fewer products to increase their engagement and potentially their balances.

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





To identify potential outliers in terms of balance among customers who have remained with the bank, we can use statistical methods such as the interquartile range (IQR) or z-scores. However, based on the provided data, it seems you've already identified potential outliers and labeled them as "Normal" or "Outlier Status". Here's a summary:

**Outliers in Account Balance:**

The balance amounts range from approximately 200K to over 15.8M.

Customers with balances labeled as "Normal" are considered within the expected range.

One customer with a balance of approximately 15.7M is labeled as an outlier.

**Insights:**

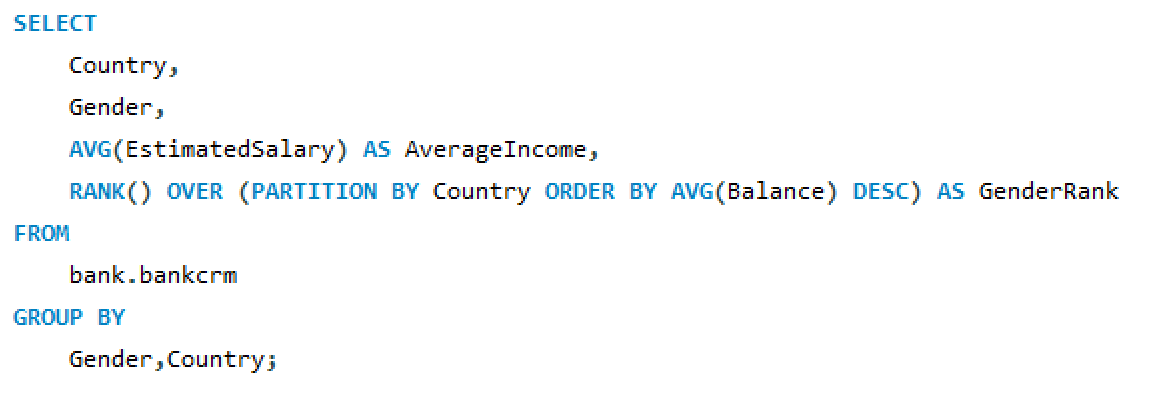
The majority of customers have balances within a typical range, labeled as "Normal".

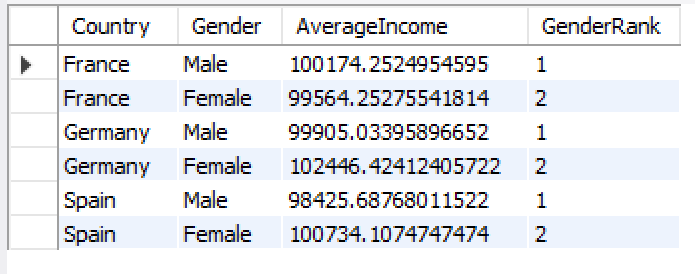
One customer stands out with an exceptionally high balance of approximately 15.7M, labeled as an outlier.

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

Total number of tables = 7, with 5 of them containing categorical variables.

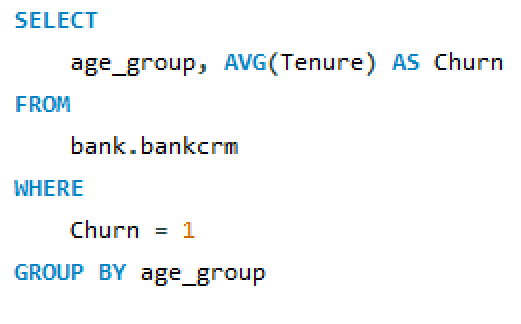
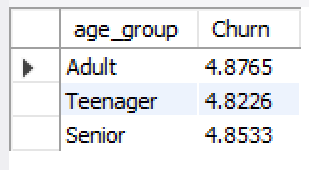
1. **Using SQL, write a query to find out the gender-wise average income of males and females in each geography id. Also, rank the gender according to the average value. (SQL)**

****

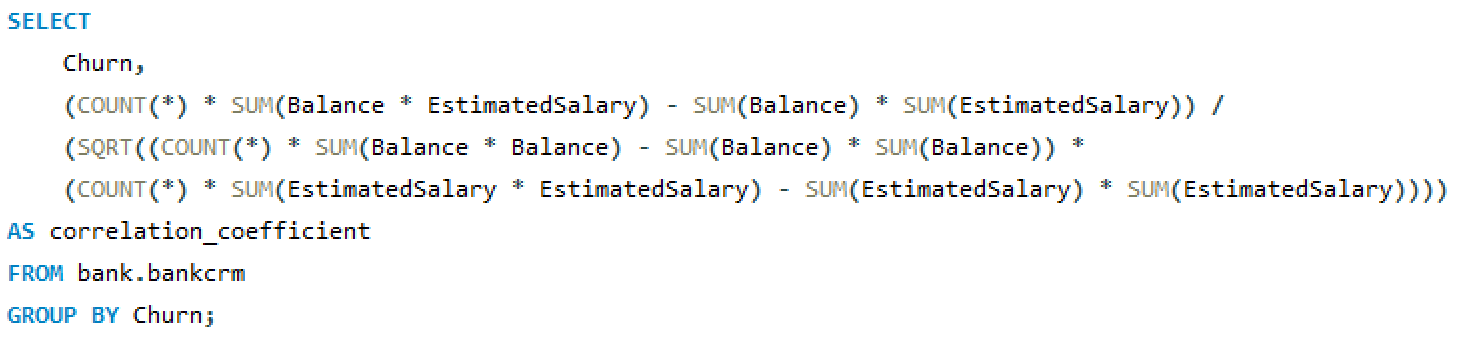


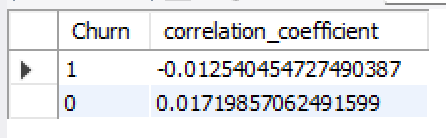
* In France and Spain, males have a higher average income than females.
* In Germany, females have a higher average income than males.
* The gender rank indicates the relative position of each gender within each country based on their average income, with 1 being the highest.

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+).**

****

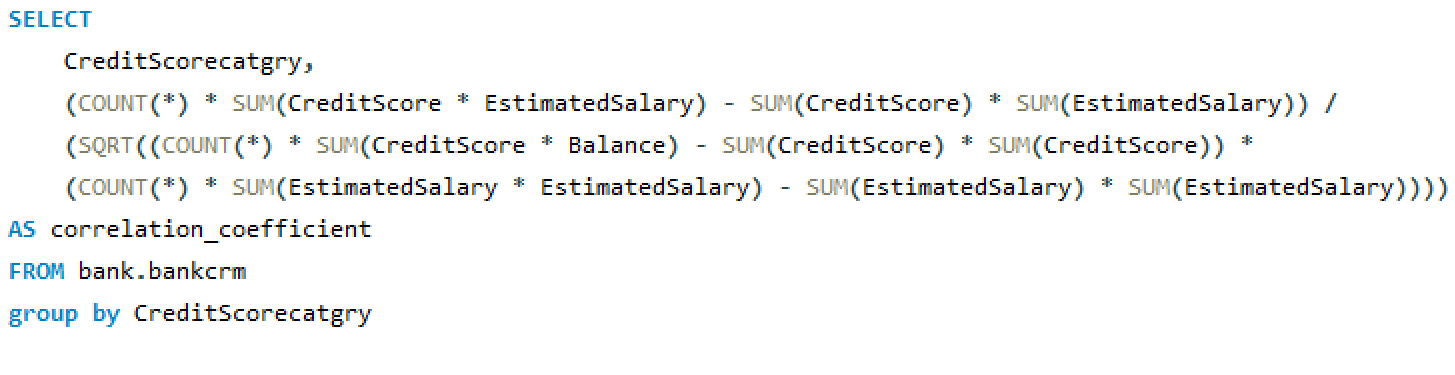
* There is relatively little variation in average tenure among different age groups of churned customers, with all age groups having similar average tenure around 4.82 to 4.88 years.
* This suggests that age may not be a significant factor in determining customer churn, at least in terms of the length of time customers remain with the bank before churning. Other factors such as product satisfaction, life events, or competitive offerings may play a more prominent role in customer churn.

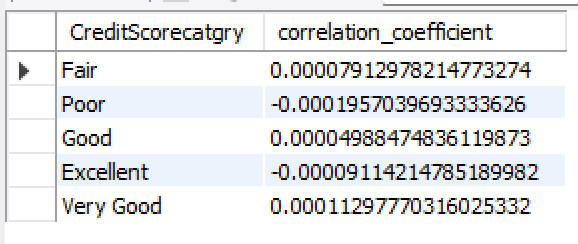
1. **Is there any direct correlation between salary and the balance of the customers? And is it different for people who have exited or not?** ****



* + Positive coefficient suggests a positive relationship: as churn status increases, the combined sum of balance and estimated salary tends to increase.
  + Negative coefficient suggests an inverse relationship.
  + For churned customers, the positive coefficient indicates a slight positive relationship with the combined sum of balance and estimated salary.
  + For non-churned customers, the coefficient is close to zero, suggesting a weak or negligible relationshipTop of Form

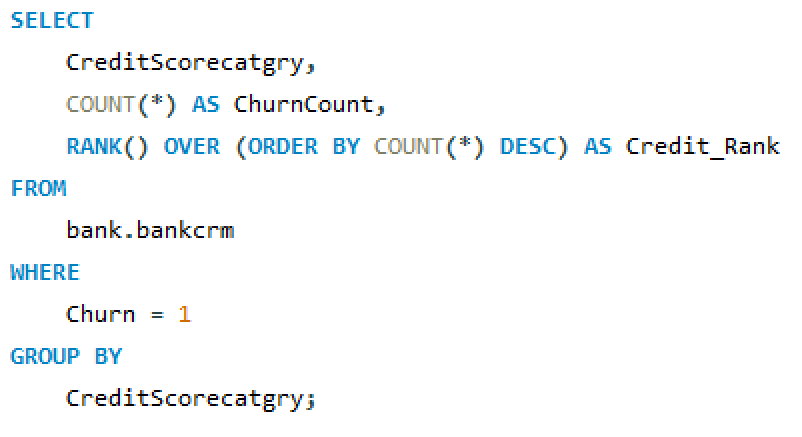
1. **Is there any correlation between the salary and the Credit score of customers?**

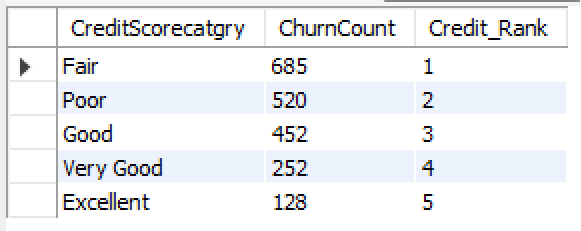
****



* The correlation coefficient measures the strength and direction of the linear relationship between the credit score category and the combined sum of credit score and estimated salary.
* The coefficients for all credit score categories are very close to zero, indicating a weak or negligible linear relationship between the credit score category and the combined sum of credit score and estimated salary.
* This suggests that there is little to no linear correlation between the credit score category and the combined sum of credit score and estimated salary in the dataset. Other factors may have a more significant influence on credit score and estimated salary.

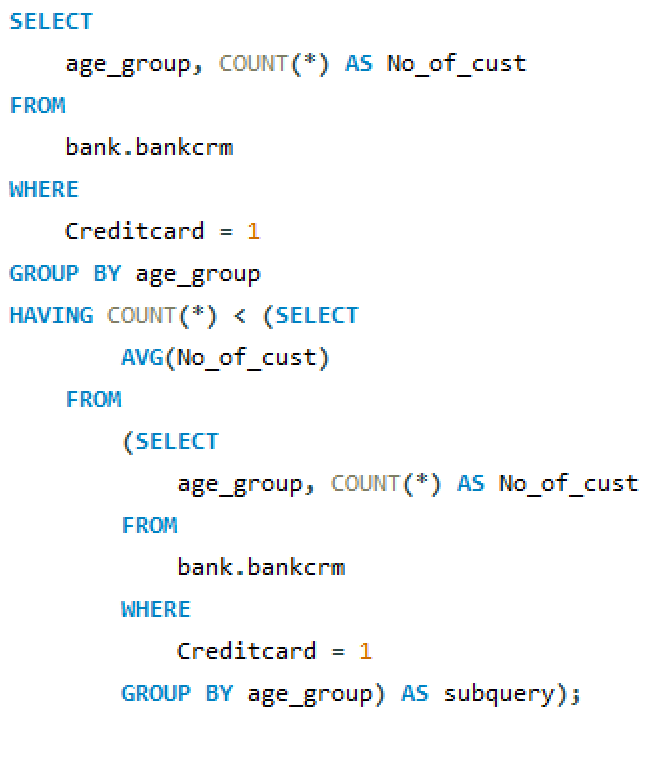
1. **Rank each bucket of credit score as per the number of customers who have churned the bank.**

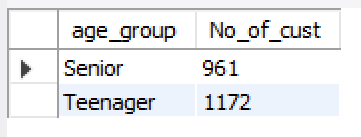
****



* Customers with a "Fair" credit score category have the highest churn count, followed by "Poor" and "Good" categories.
* The credit rank assigns a ranking based on the churn count for each credit score category, with the "Fair" category having the highest churn count and therefore ranked first.
* Understanding the churn distribution among different credit score categories can help the bank tailor its retention strategies and customer management efforts, potentially focusing more resources on customers with higher churn rates in specific credit score categories.

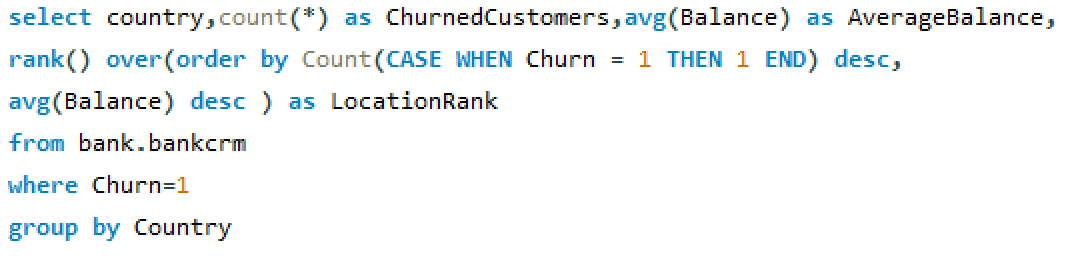
1. **According to the age buckets find the number of customers who have a credit card. Also retrieve those buckets that have lesser than average number of credit cards per bucket.**

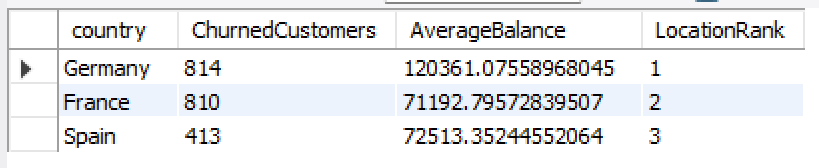
****



* Customers with a "Fair" credit score category have the highest churn count, followed by "Poor" and "Good" categories.
* The credit rank assigns a ranking based on the churn count for each credit score category, with the "Fair" category having the highest churn count and therefore ranked first.
* Understanding the churn distribution among different credit score categories can help the bank tailor its retention strategies and customer management efforts, potentially focusing more resources on customers with higher churn rates in specific credit score categories.

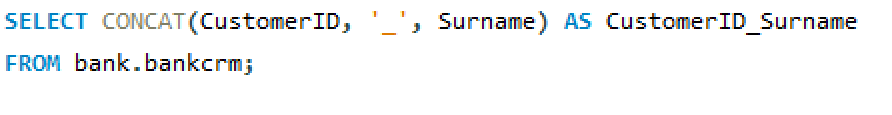
1. **Rank the Locations as per the number of people who have churned the bank and average balance of the customers.**

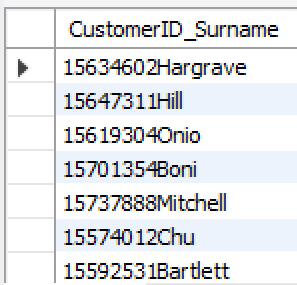
****



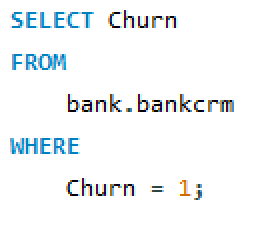
* Germany has the highest number of churned customers and the highest average balance among the three countries, earning it the top location rank.
* France has a significant number of churned customers but a lower average balance compared to Germany.
* Spain has the lowest number of churned customers among the three countries and a moderate average balance, placing it third in location rank.
* These insights can help the bank identify regions with higher churn rates and average balances, potentially guiding resource allocation and targeted retention efforts.Top of Form

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

****



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

****

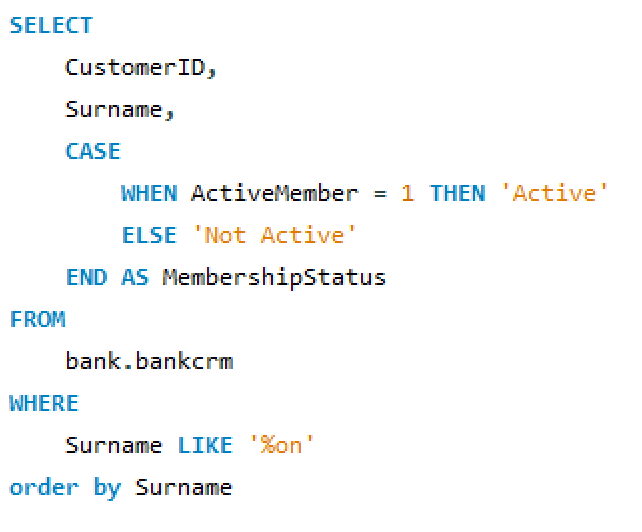
The operation of retrieving data from multiple tables without using the JOIN keyword can be achieved through subqueries or correlated subqueries.

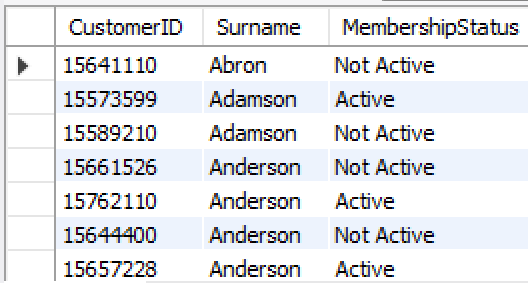
1. **Were there any missing values in the data, using which tool did you replace them and what are the ways to handle them?**

No, there were no missing values in the data.

**Methods for Handling Missing Values:** If missing values were present, several methods could be employed:

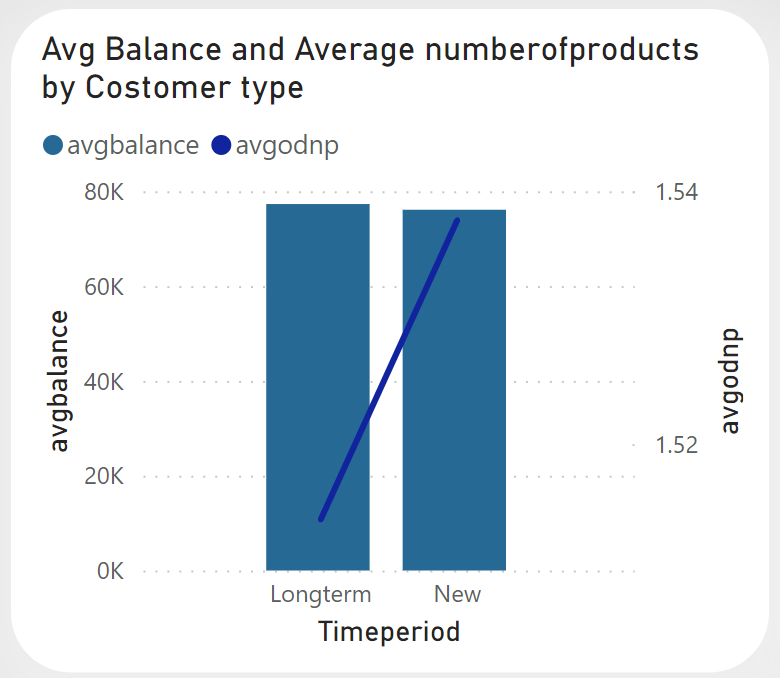
1. **Imputation using Mean, Median, or Mode:**
   * Mean imputation replaces missing values with the mean of the observed values for that variable.
   * Median imputation replaces missing values with the median of the observed values for that variable.
   * Mode imputation replaces missing values with the mode of the observed values for that variable.
2. **Forward or Backward Filling:**
   * Forward filling replaces missing values with the last observed value in the dataset.
   * Backward filling replaces missing values with the next observed value in the dataset.
3. **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”.**

****

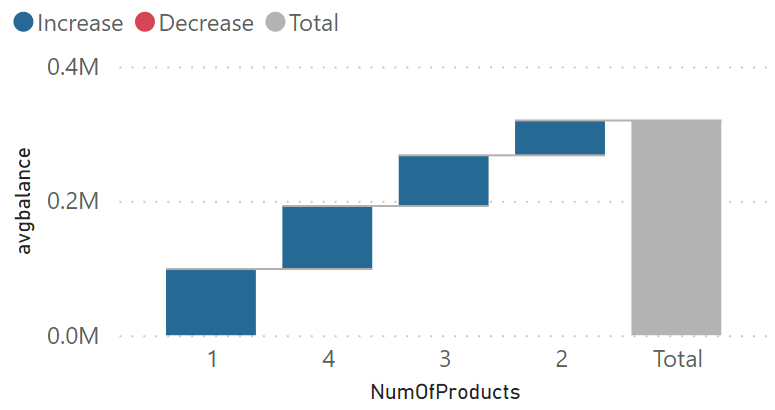


**Subjective Question:**

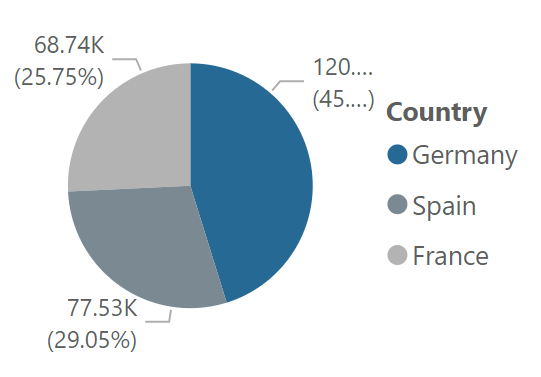
1. **Customer Behavior Analysis: What patterns can be observed in the spending habits of long-term customers compared to new customers, and what might these patterns suggest about customer loyalty?**

****

1. **Stability in Spending**: Long-term customers may exhibit more stable spending patterns compared to new customers. They might have established routines and preferences over time, leading to consistent spending behavior**.**
2. **Variability in Spending:** New customers might display more variability in their spending habits as they explore different products or services offered by the bank. Their spending patterns could fluctuate as they adjust to their financial needs and preferences**.**
3. **Relationship Building**: Long-term customers may have developed stronger relationships with the bank over time, leading to trust and loyalty. They might feel more comfortable and confident in their financial decisions, resulting in consistent spending behavior.
4. **Exploratory Spending:** New customers may engage in exploratory spending as they test out different products or services offered by the bank. Their spending patterns might reflect experimentation and discovery as they determine what best fits their needs**.**
5. **Retention Strategies:** The stable spending patterns of long-term customers could indicate successful retention strategies implemented by the bank, such as personalized offers, loyalty programs, or exceptional customer service. These strategies might reinforce customer loyalty and encourage continued engagement with the bank.
6. **Product Affinity Study: Which bank products or services are most commonly used together, and how might this influence cross-selling strategies?**



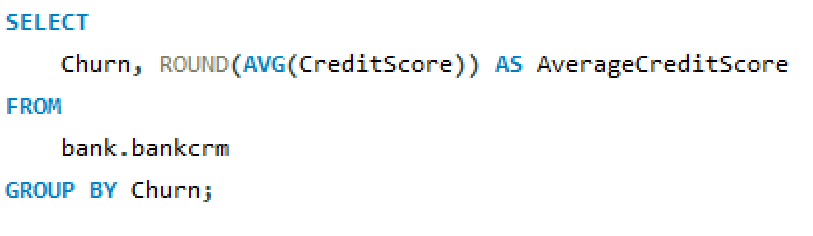
1. **Correlation Analysis**: You can conduct a correlation analysis to identify relationships between different bank products or services. For example, you could examine if customers who have a checking account are more likely to also have a savings account, or if customers with a mortgage are more likely to have a home insurance policy.
2. **Association Rules**: Using techniques like association rule mining, you can identify frequent itemsets and association rules among different bank products or services. This can help uncover patterns such as "customers who have a credit card are X times more likely to have a personal loan."
3. **Segmentation Analysis**: Segment customers based on their usage of different bank products or services. For example, you could create segments like "Basic Banking," "Investment Savvy," or "Loan Heavy," and then analyze the characteristics and behaviors of customers within each segment.
4. **Cross-Selling Strategies**: Once you've identified which bank products or services are commonly used together, you can tailor cross-selling strategies to promote complementary products or services to customers. For example, if customers with a savings account are more likely to also have a retirement account, you could target savings account holders with promotions or incentives to open a retirement account.
5. **Geographic Market Trends: How do economic indicators in different geographic regions correlate with the number of active accounts and customer churn rates?**

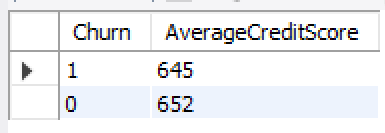


* Germany has the highest number of churned customers and the highest average balance among the three countries, earning it the top location rank.
* France has a significant number of churned customers but a lower average balance compared to Germany.
* Spain has the lowest number of churned customers among the three countries and a moderate average balance, placing it third in location rank.

These insights can help the bank identify regions with higher churn rates and average balances, potentially guiding resource allocation and targeted retention efforts

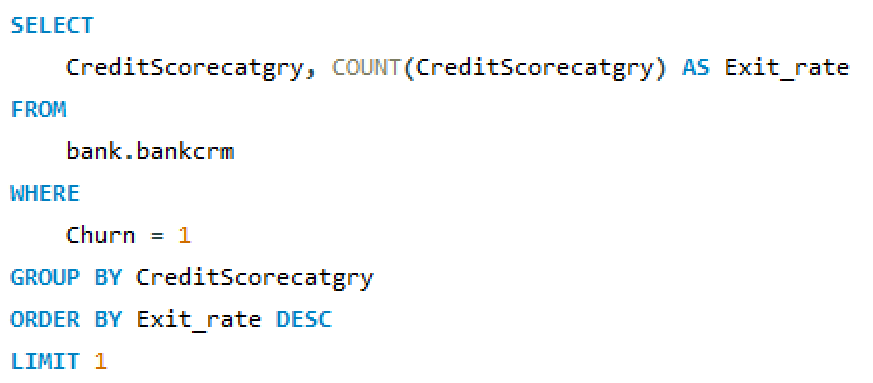
1. **Risk Management Assessment: Based on customer profiles, which demographic segments appear to pose the highest financial risk to the bank, and why?**

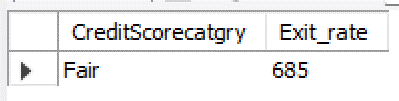
****



* Churned Customers (Churn = 1): The average credit score is approximately 645.
* Non-Churned Customers (Churn = 0): The average credit score is slightly higher, around 652.

These results suggest that, on average, non-churned customers tend to have slightly higher credit scores compared to churned customers. This insight could be valuable for the bank's customer retention strategies, as it indicates a potential correlation between credit score and churn behavior.

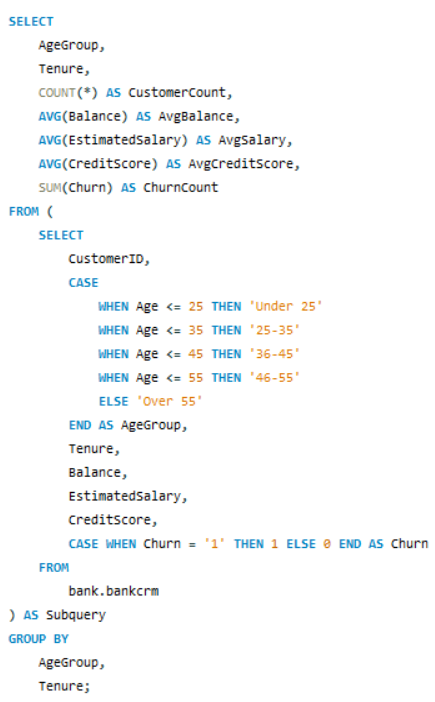
****



* **Exit Rate by Credit Score Category**:
  + Credit Score Category: Fair
  + Exit Rate: The highest exit rate, with approximately 685 customers falling into

this category.

This suggests that customers categorized as having a "Fair" credit score are more likely to churn compared to customers in other credit score categories. Understanding this trend can help the bank identify at-risk customers and implement targeted retention strategies, such as offering personalized services or incentives to mitigate churn among customers with fair credit scores.

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?  
   **
2. **Data Prep:** Collect customer data including tenure, transactions, demographics**.**
3. **Feature Eng:** Create features like transaction frequency, product ownership**.**
4. **Modeling:** Choose and train models like regression or machine learning algorithms.
5. **Evaluation:** Assess model performance using testing data.
6. **Segment Analysis:** Analyze predictions for different segments to identify trends.
7. **Deployment**: Integrate model into bank systems for real-time predictions.
8. **Monitoring:** Continuously monitor model performance and refine as needed.
9. **Marketing Campaign Effectiveness: How could you assess the impact of marketing campaigns on customer retention and acquisition within the dataset? What extra information would you need to solve this?**

To assess the impact of marketing campaigns on customer retention and acquisition within the dataset, follow these straightforward steps:

1. **Define Key Metrics:** Identify metrics like customer retention rate, acquisition rate, and churn rate to measure the campaign's effectiveness.
2. **Track Changes**: Monitor these metrics before, during, and after the campaign to see how they fluctuate over time.
3. **Segment Customers:** Divide customers into groups based on demographics or behavior to understand which segments respond best to the campaign.
4. **Conduct A/B Testing:** Compare the behavior of customers exposed to the campaign (test group) with those who weren't (control group) to gauge its impact.
5. **Collect Feedback:** Gather feedback from customers through surveys to gauge their perception and response to the campaign.
6. **Consider External Factors:** Take into account external factors like market trends and competitor activities to understand the broader context.
7. **Use Simple Analytics Techniques:** Utilize basic analytics techniques to analyze campaign performance and forecast future trends.

**Extra Information Needed:**

* Details of the marketing campaign, including its duration, target audience, and objectives.
* Comprehensive customer data including transaction history and interaction logs.
* External data sources such as economic indicators and industry trends.
* Historical data on customer behavior for benchmarking

1. **Customer Exit Reasons Exploration: Can you identify common characteristics or trends among customers who have exited that could explain their reasons for leaving?**
2. **Low Account Activity**: Customers who exhibit low account activity, such as infrequent transactions or minimal engagement with bank services, may be more likely to churn. This could indicate dissatisfaction or lack of perceived value from the bank's offerings.
3. **Negative Balance Trends:** Customers who consistently maintain negative balances or experience a decline in their account balances over time may be at higher risk of churn. Financial instability or dissatisfaction with banking services could contribute to this trend.
4. **Complaints or Issues:** Customers who have lodged complaints or encountered issues with the bank's services, such as poor customer service experiences, billing errors, or technical issues, may be more inclined to leave the bank in search of better alternatives.
5. **Life Events:** Certain life events, such as relocation, job changes, or major financial decisions (e.g., buying a house, starting a business), could prompt customers to reassess their banking relationships and potentially switch to another bank that better meets their evolving needs.
6. **Competitive Offerings:** The availability of competitive offerings from other banks, such as attractive interest rates, rewards programs, or innovative digital banking features, may lure customers away from their current bank.
7. **Demographic Trends:** Certain demographic factors, such as age, income level, or occupation, may influence customer churn rates. For example, younger customers may be more tech-savvy and inclined to switch banks for better digital experiences, while older customers may prioritize stability and relationship-based banking services.
8. **Are 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' important for predicting if a customer will leave the bank?**

Yes, 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' are important for predicting if a customer will leave the bank.

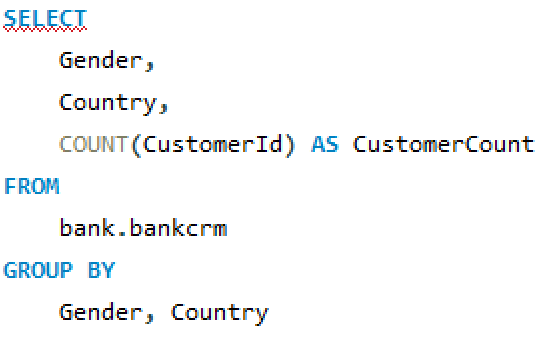
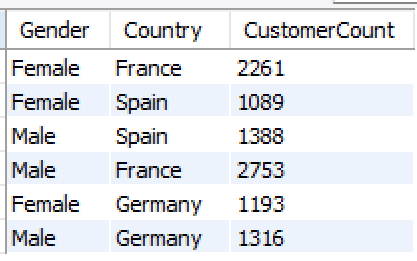
**Tenure:** Longer tenure often means stronger loyalty**.**

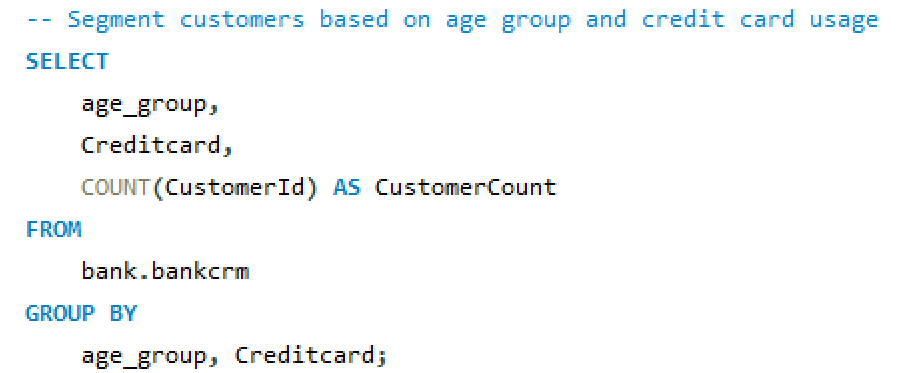
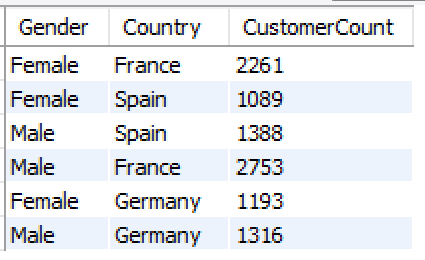
**NumOfProducts:** More products indicate higher engagement.

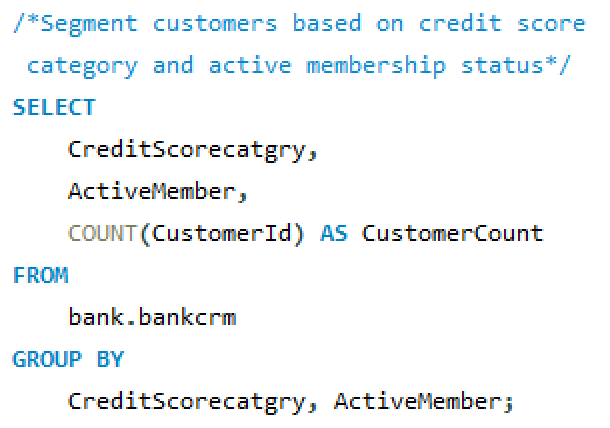
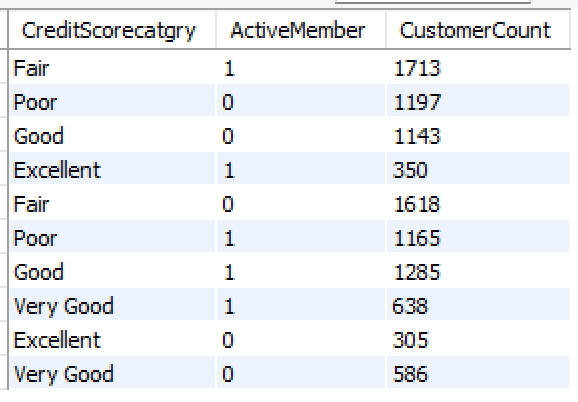
**IsActiveMember:** Active members are less likely to churn.

**EstimatedSalary:** Higher salary can imply financial stability, reducing churn risk.

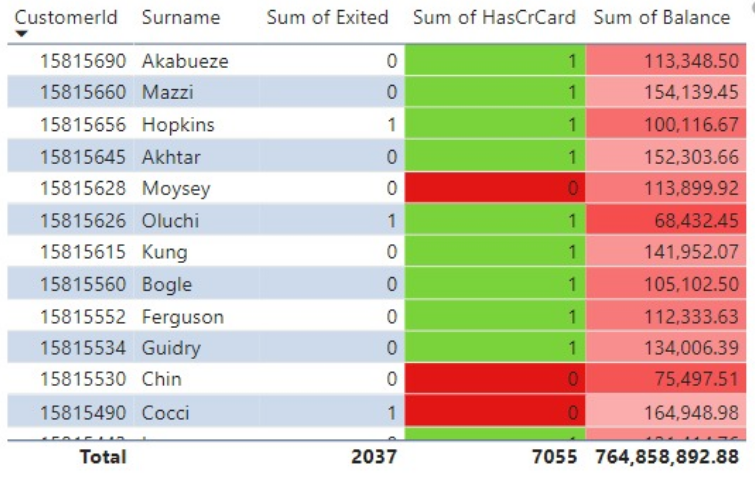
1. **Utilize SQL queries to segment customers based on demographics and account details.**

**** ****

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

****

**Know Churn Risks**: Identify why customers might leave, like low activity or dissatisfaction**.**

**Get Data:** Collect customer info from your records, like transactions and feedback.

**Divide Customers**: Split customers into groups based on activity or satisfaction.

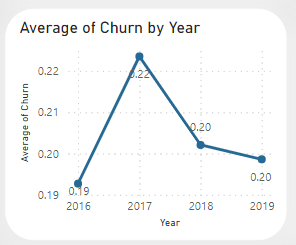
**Set Rules**: Decide what signals a risk, like low activity or complaints.

**Highlight Risks**: Use colors or symbols to mark risky customers in your records.

**Check Rewards Impact:** Compare retention rates for customers with and without credit card rewards.

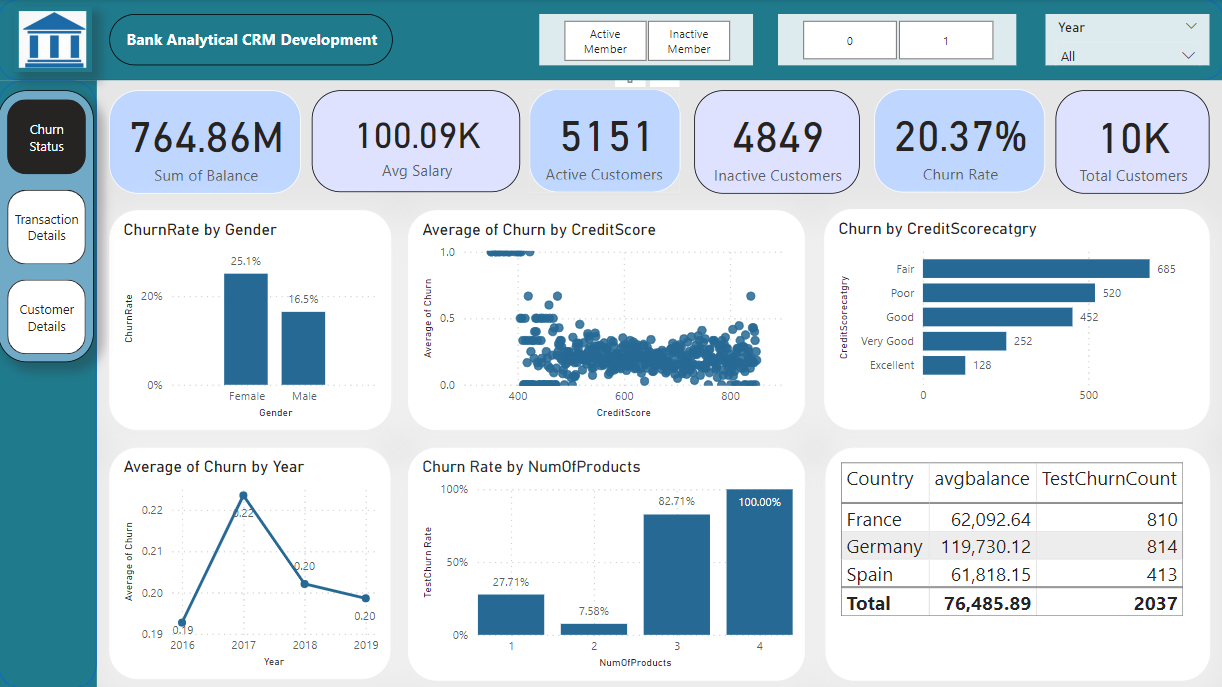
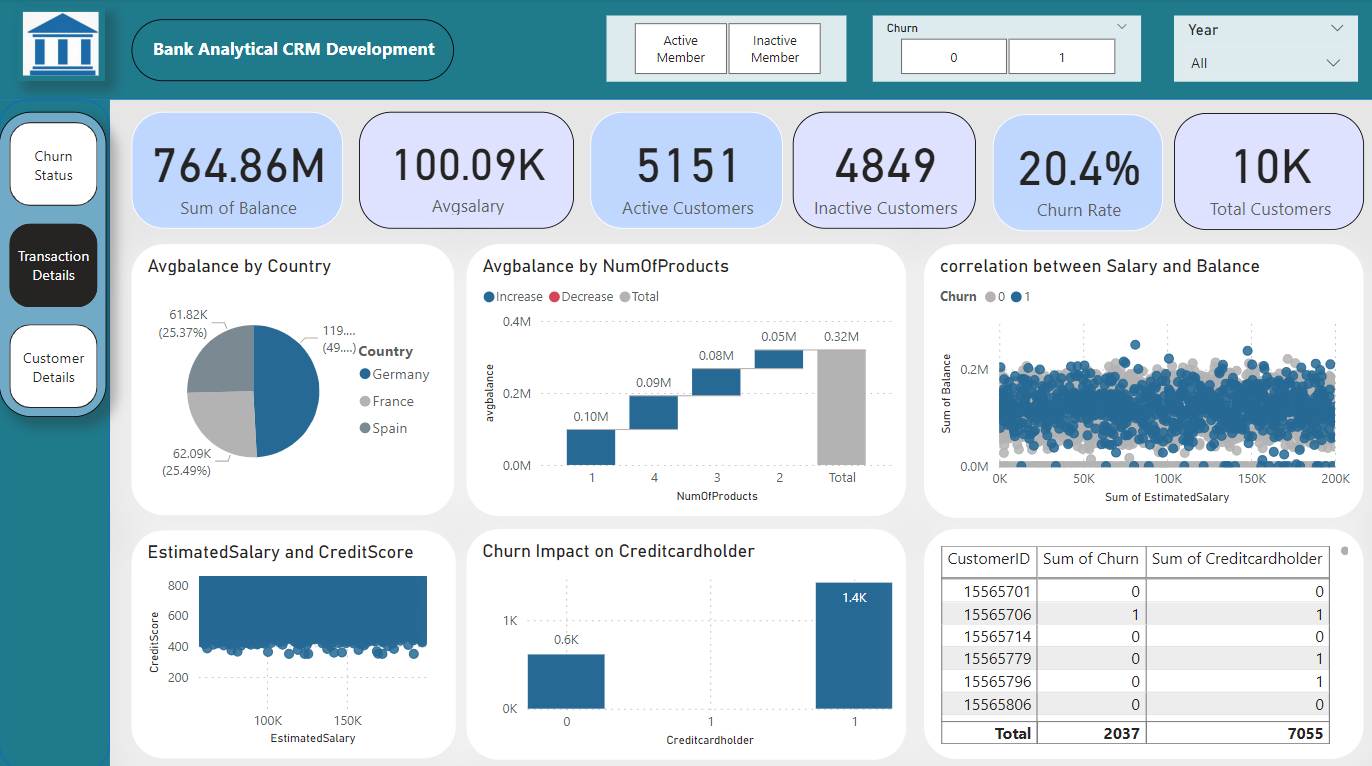
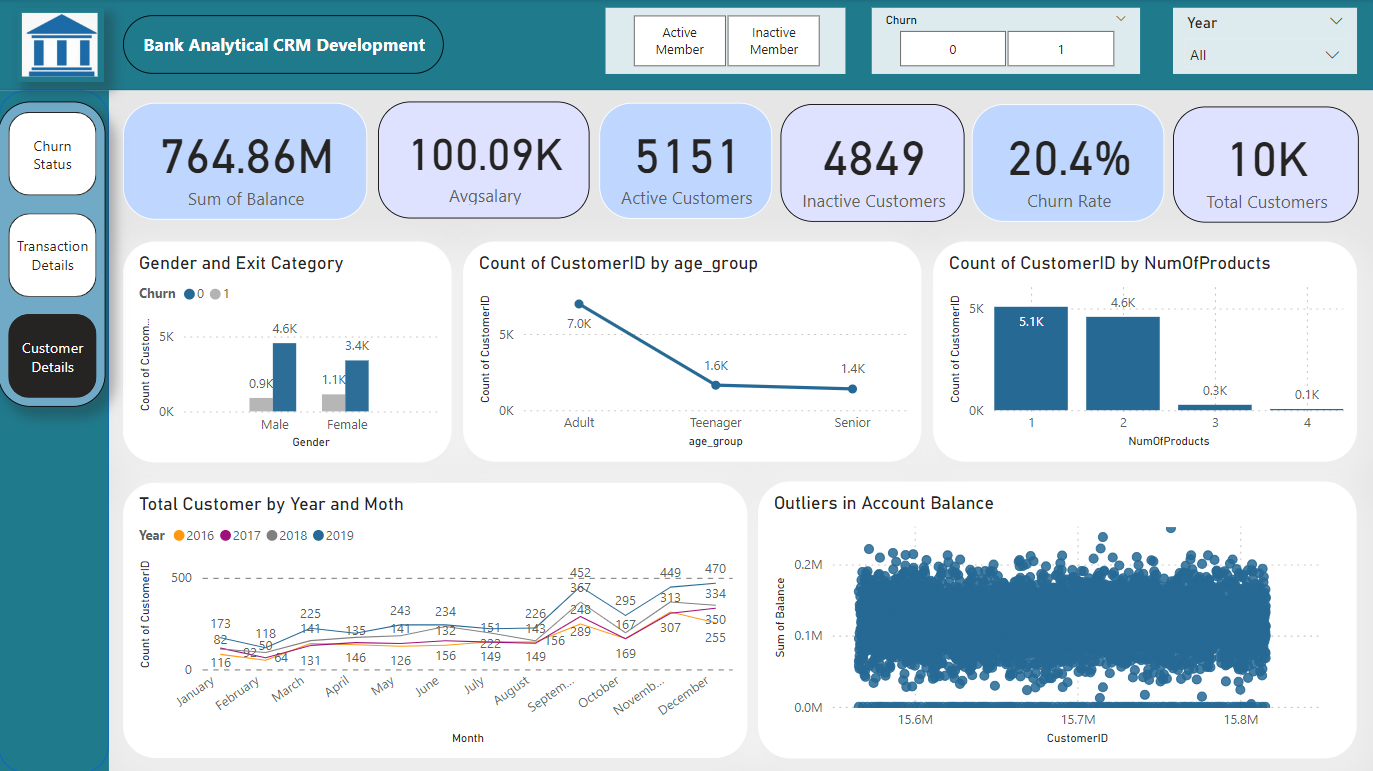
**Adjust as Needed**: Keep an eye on trends and tweak your approach to keep customers happy**.**

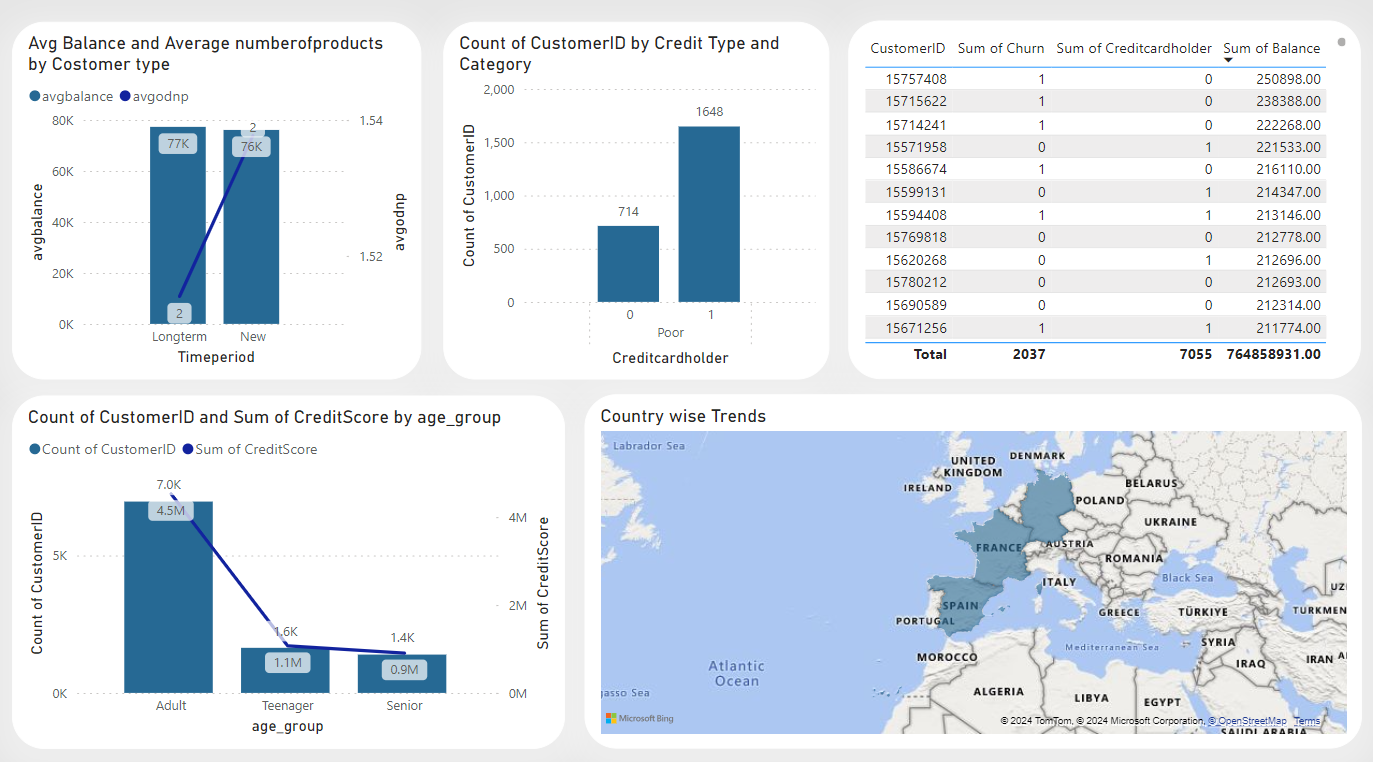
1. **What is the current churn rate per year and overall as well in the bank? Can you suggest some insights to the bank about which kind of customers are more likely to churn and what different strategies can be used to decrease the churn   
   rate?**



This suggests that customers categorized as having a "Fair" credit score are more likely to churn compared to customers in other credit score categories. Understanding this trend can help the bank identify at-risk customers and implement targeted retention strategies, such as offering personalized services or incentives to mitigate churn among customers with fair credit scores.

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



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

If the objective and subjective questions weren't provided, I would approach the problem of segmenting customers based on demographics and account details in the following way:

**Data Exploration:**

Begin by examining the structure and contents of the ban.bankcrm table to understand the available data fields.

Identify the demographic variables such as Gender, Age, Country, and account details like Estimated Salary, Balance, Credit Score, etc.

**Segmentation Strategy:**

Define the segmentation strategy based on business objectives or hypotheses. For example, you might want to segment customers to tailor marketing strategies, identify high-value customers, or understand churn behavior.

**Segmentation Criteria:**

Select the relevant variables/features for segmentation based on the business objectives. These could include demographics (gender, age, country), account details (estimated salary, balance, credit score), product usage (number of products), etc.

**Data Preparation:**

Clean the data by handling missing values, outliers, and inconsistencies.

Transform categorical variables into a suitable format for analysis (e.g., one-hot encoding for categorical variables like gender and country).

**Segmentation Analysis:**

Utilize SQL queries or statistical analysis techniques to segment the customers based on the selected criteria.

Explore the characteristics and behaviors of different customer segments to gain insights**.**

**Evaluation and Validation:**

Assess the effectiveness of the segmentation by evaluating how well it aligns with business goals and objectives.

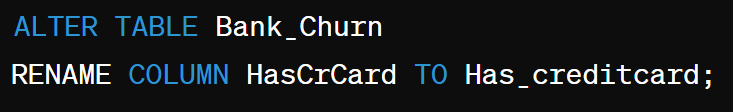
Validate the segmentation results through statistical tests or validation metrics to ensure they are meaningful and actionable.

**Iterative Improvement:**

Iterate on the segmentation approach based on feedback, additional data insights, or changes in business requirements.

Continuously refine and optimize the segmentation strategy to improve its effectiveness in addressing business needs.

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



To modify the name of the "HasCrCard" column to "Has\_creditcard" in the "Bank\_Churn" table, you would typically use an SQL query. Here's the general syntax for renaming a column in SQL: