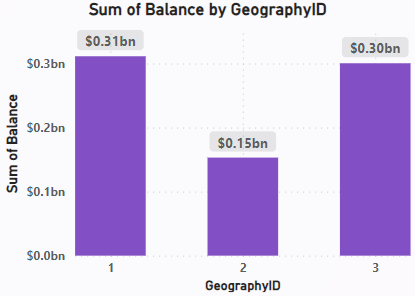
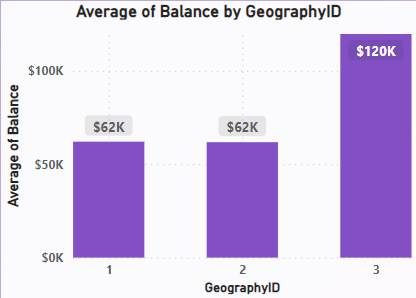
Bank CRM

Objective Questions

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

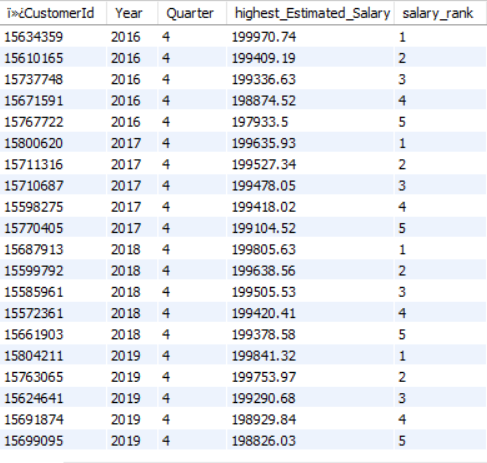
Answer. The distribution of account balances across different regions has been analyzed by considering both the total sum of balances and the average balance per customer in each region. Here are the findings:

These results indicate that while Regions 1 and 3 have similar total balances, Region 3 has a significantly higher average balance per customer compared to regions 1 and 2. This suggests that customers in region 3 tend to maintain higher individual account balances.

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

Answer. For each year, the top 5 customers with the highest Estimated Salary in the last quarter (Q4) are identified.



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

Answer. Based on the SQL query analysis, the average number of products used by customers who have a credit card is **1.5314**

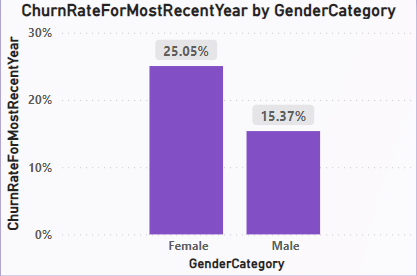


1. Determine the churn rate by gender for the most recent year in the dataset.

Answer. The churn rate by gender for the most recent year in the dataset has been calculated as follows:

GenderID 1 (Male): Churn Rate: **15.37%**

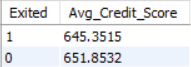
GenderID 2 (Female): Churn Rate: **25.05%**



This analysis indicates that the churn rate for males is **15.37%,** whereas for females it is **25.05%**. This suggests that female customers are more likely to churn compared to male customers in the most recent year.

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

Answer. Based on the SQL query analysis, the average credit score is as follows: Customers who have exited: **645** Customers who remain: **651**



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

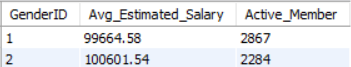
Answer. Based on the SQL query analysis, the results are as follows:

Gender ID 1 (Male):

Average Estimated Salary: **99,664.58** Number of Active Accounts: **2,867**

Gender ID 2 (Female):

Average Estimated Salary: **1,00,601.54** Number of Active Accounts: **2,284**



This indicates that while females tend to have a slightly higher average estimated salary, males are more likely to have active accounts.

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

Answer. The **"Poor"** credit score segment has the highest exit rate of **22.02%.**



This analysis indicates that customers with a "**Poor**" credit score are more likely to exit the bank compared to other segments.

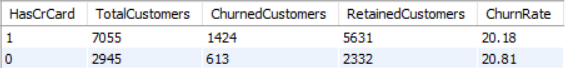
1. Find out which geographic region has the highest number of active customers with a tenure greater than 5 years. (SQL)

Answer. Based on the SQL query analysis, geography ID 1 has the highest number of active customers with a tenure greater than 5 years, totaling **797** active customers.



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

Answer. From the SQL query output, we can observe the following:

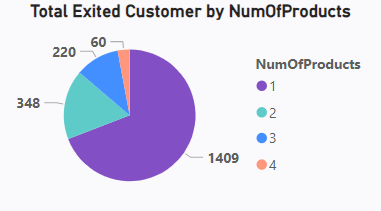


* Customers with a credit card (HasCrCard = 1) have a churn rate of 20.18%.
* Customers without a credit card (HasCrCard = 0) have a slightly higher churn rate of 20.81%.

**Conclusion:** There is a small difference in the churn rate between customers with and without a credit card. Customers without a credit card have a slightly higher churn rate compared to those with a credit card. This suggests that having a credit card might have a minor positive impact on customer retention. However, the difference is not substantial, indicating that other factors might play a more significant role in influencing customer churn.

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

Answer. The analysis revealed the following distribution of the number of products used by customers who have exited

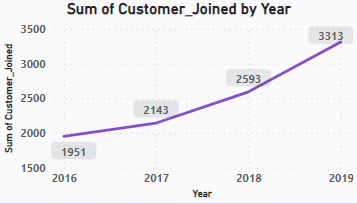


From this data, we can observe that the most common number of products used by customers who have exited is **1** product, with **1,409** customers falling into this category. This indicates that customers who have exited tend to use fewer products.

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.

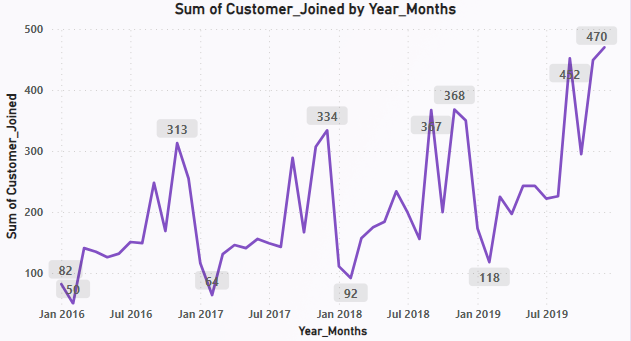
Answer.

* **Yearly Trend Analysis**



Consistent annual growth, with the largest increase in 2019.

* **Monthly Trend Analysis**

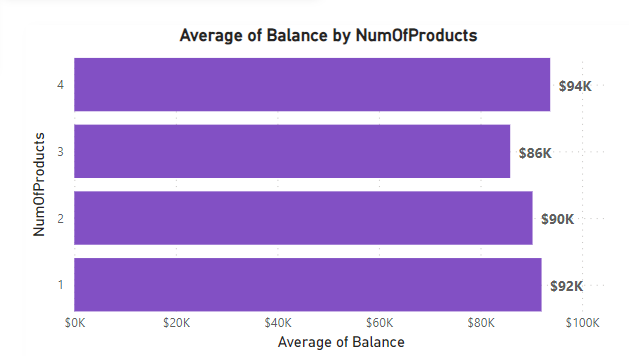


Significant monthly fluctuations, with peaks in January and July suggesting effective campaigns or seasonal factors.

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

Answer. I created a bar chart in Power BI to analyze the relationship between the number of products and the account balance for customers who have exited. To ensure the analysis was focused on exited customers, I applied a visual-level filter to include only rows where Exited = 1.

The bar chart displayed the average account balance for different numbers of products.

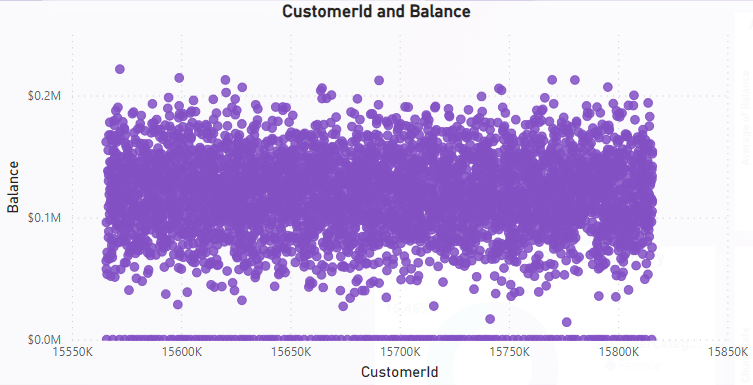


From this data, we can observe that customers who have exited and used 3 products tend to have the lowest average balance, while those with 4 products have the highest average balance.

However, the differences in average balances are relatively small, suggesting that the number of products has a modest impact on the account balance for exited customers.

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

Answer. I applied a visual-level filter to include only rows where Exited = 0



Upon analyzing the scatter chart, we observed that the majority of customer balances clustered within the $0 to $0.1M range. However, several data points were significantly higher, exceeding $0.2M, indicating potential outliers. These high-balance customers may represent high-value clients who require special attention. Additionally, there were a few points very close to $0, which might also be considered outliers.

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

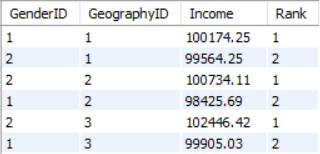
Answer. There are seven tables provided in the dataset. Among these, the following five tables consist only of categorical variables:

* **ActiveCustomer**
* **Gender**
* **Geography**
* **ExitCustomer**
* **CreditCard**

These tables provide categorical information about the customers, such as their active status, gender, geographic region, exit status, and credit card holding status.

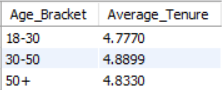
15.Using SQL, write a query to find out the gender-wise average income of males and females in each geography id. Also, rank the gender according to the average value. (SQL)

Answer. This ranking indicates that in Geography 1, males have a higher average income, while in Geography 2 and 3, females have a higher average income.



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+).

Answer. The average tenure of customers who have exited is as follows:

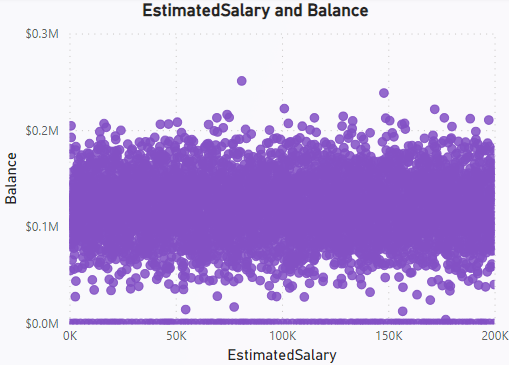


The results show that the average tenure for exited customers is relatively consistent across different age brackets, with the 30-50 age bracket having the highest average tenure at approximately 4.8899 years. This insight might indicate that the duration of tenure is not significantly different across these age groups for customers who have left the bank.

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

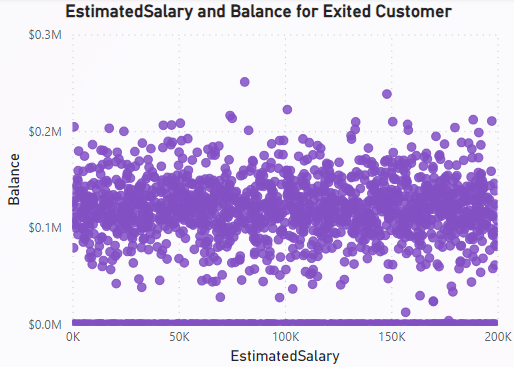
Answer.

**Overall Correlation**



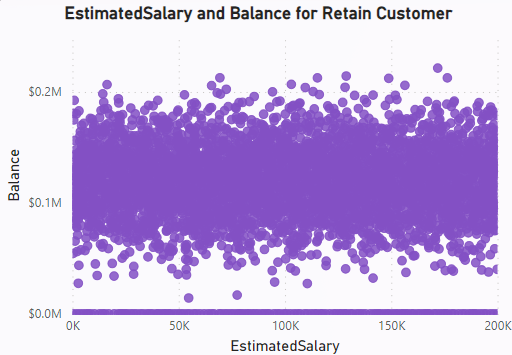
* A scatter plot was created to visualize the correlation between EstimatedSalary and Balance for all customers.
* The plot shows a widespread distribution of data points, indicating no strong correlation between salary and balance. Customers with varying salaries have similar balance distributions.
* This suggests that across the entire customer base, there is no direct relationship between how much a customer earns and how much they maintain as their account balance.

**Exited Customers**



* A scatter plot filtered for exited customers was analyzed to understand if salary impacts the likelihood of a customer exiting the bank.
* Similar to the overall correlation, the plot for exited customers shows no clear pattern or trend. Customers who have left the bank display a wide range of balances across all salary levels.
* This indicates that for customers who have exited the bank, their salary does not significantly affect their account balance. High or low salary levels do not predict whether a customer will have a higher or lower balance when they decide to leave.

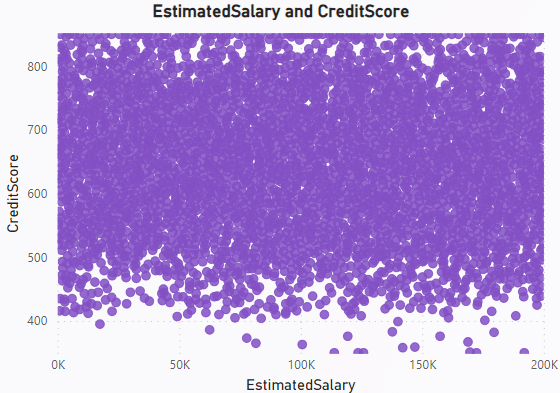
**Retained Customers**



* A scatter plot filtered for retained customers was analyzed to determine if salary impacts the retention of customers.
* The plot for retained customers, like the previous plots, shows no clear correlation. Customers who remain with the bank have diverse balances, regardless of their salaries.
* This suggests that for customers who have stayed with the bank, their salary does not significantly affect their account balance. Customers with higher or lower salaries can equally have varying account balances.

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

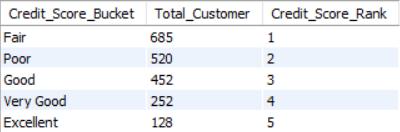
To analyze the correlation between the EstimatedSalary and CreditScore of customers, created a scatter plot.



* The scatter plot shows the distribution of customer data points based on their estimated salary and credit score.
* From the scatter plot, it is observed that there is no strong correlation between EstimatedSalary and CreditScore.
* Customers with varying salaries have similar credit scores, indicating that a customer's income level does not significantly impact their credit score.

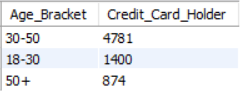
19.Rank each bucket of credit score as per the number of customers who have churned the bank.

Answer. From this, we can see that customers with a **"Fair"** credit score are the most likely to churn, followed by those with "Poor" and "Good" credit scores.

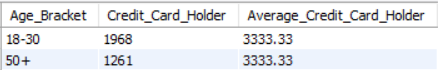


20.According to the age buckets find the number of customers who have a credit card. Also retrieve those buckets that have lesser than average number of credit cards per bucket.

Answer. Total number of customers with a credit card in each age bucket



Age buckets with fewer than the average number of credit card holders



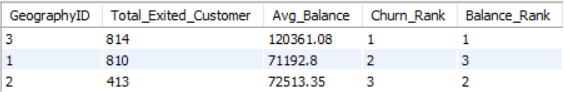
These results highlight that the age buckets 18-30 and 50+ have fewer credit card holders than the average across all age buckets

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

Answer. Based on the analysis, GeographyID 3 has both the highest number of churned customers and the highest average balance. This suggests that while customers in this region are more likely to churn, they also tend to have higher account balances.

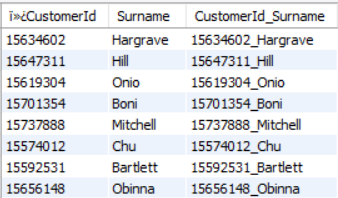
GeographyID 1 has a high churn rate but the lowest average balance.

GeographyID 2, with the lowest churn count and moderate average balance.



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”.

Answer. To create a column where the format is "CustomerID\_Surname", we can use the SQL CONCAT function.

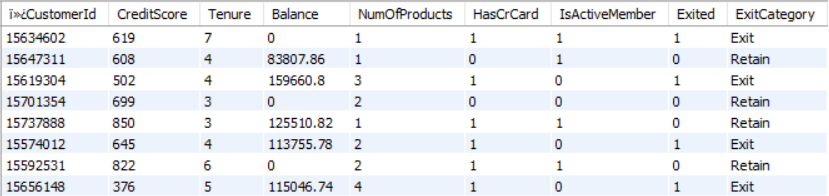


This will allow us to join the CustomerInfo table with another table where the primary key is a combination of CustomerID and Surname.

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

Answer. Yes, it is possible to get the **ExitCategory** from the **ExitCustomer** table to the **Bank\_Churn** table without using an explicit JOIN. We can achieve this by using a subquery.

The SQL query is in the SQL file.



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

Answer. There were no missing values in the given dataset. Power BI was used to verify and handle any potential missing values.

**Ways to Handle Missing Values:**

1. Re
2. move Missing Data: Delete rows or columns with missing values if they are minimal.
3. Impute Missing Data:

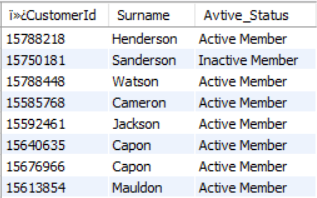
* Mean/Median Imputation: Replace with the mean or median of the column.
* Mode Imputation: Replace categorical data with the mode.

1. Power BI's Power Query Editor offers various tools for handling and imputing missing values effectively.

25.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”.

Answer. To retrieve the customer IDs, their last names, and their active status for customers whose surname ends with "on," I use the `LIKE` operator to filter the surnames.

The SQL query is in the SQL file.



26.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.

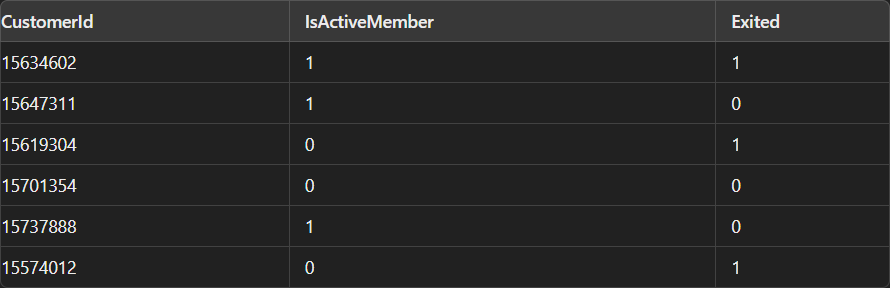
Answer. A potential discrepancy is observed in the `IsActiveMember` and `Exited` columns. Specifically, inactive customers (`IsActiveMember` = 0) are not consistently exiting the bank (`Exited` = 1), which raises questions about the accuracy of the `IsActiveMember` status.

**Data Details**

`IsActiveMember`: Indicates whether a customer is an active member (1 for active, 0 for inactive).

`Exited`: Indicates whether a customer has left the bank (1 for exited, 0 for retained). This column is accurate.

**Sample Data**



**Conclusion**

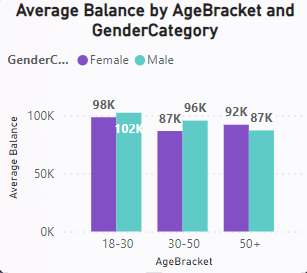
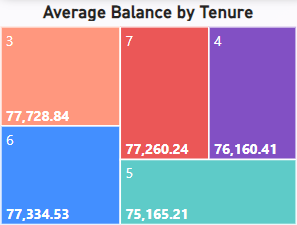
The discrepancy between inactive members and their exit status suggests a need to review the criteria for IsActiveMember to ensure it accurately reflects customer engagement.

Subjective Questions

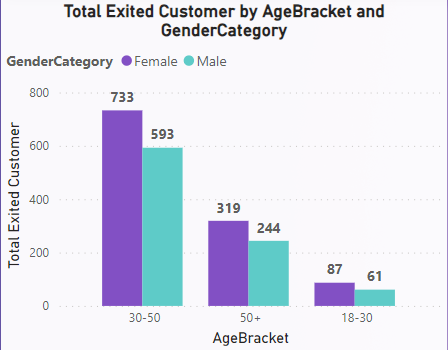
1.**Customer Behavior Analysis**: What patterns can be observed in the spending habits of long-term customers compared to new customers, and what might these patterns suggest about customer loyalty?

Answer.

* Patterns Observed:
* The spending habits of customers with different tenures and age brackets show variability in average balances.
* Younger customers, particularly females, exhibit higher average balances, suggesting a potential for higher spending capacity or savings among this demographic.
* Long-term customers with a tenure of 3 years have the highest average balance, indicating strong loyalty and engagement with the bank's services.



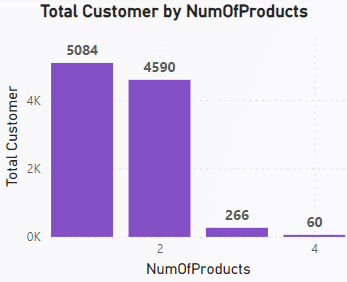
* Customer Loyalty
* Long-term customers (30-50) show higher churn rates, indicating lower satisfaction.
* Younger customers (18-30) demonstrate higher loyalty.
* Geographic differences necessitate region-specific retention strategies, especially in Germany.



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

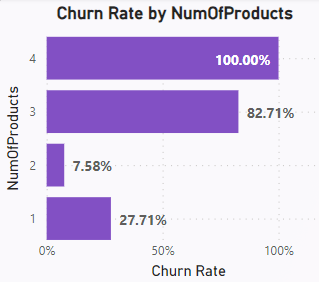
Answer. To identify which bank products are most commonly used together and how this influences cross-selling strategies, two charts were analyzed:

**Total Customers by Number of Products**



* Most customers use 1 or 2 products (5084 and 4590 respectively).
* Few customers use 3 or 4 products (266 and 60 respectively).

**Churn Rate by Number of Products**



* Single-product users have a higher churn rate (27.71%).
* The churn rate for 2-product users is significantly lower (7.58%).

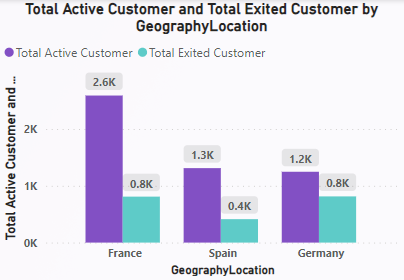
**Cross-Selling Strategies**

* Address High Churn in Single-Product Users:
* Focus on cross-selling additional products to single-product users to improve retention.
* Promote Multi-Product Use:
* Customers with multiple products show lower churn rates. Encourage single-product users to adopt more products.
* Targeted Marketing Campaigns:
* Create campaigns aimed at single-product users, showcasing the benefits of using multiple products.

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

Answer. Geographic Market Trends:

* Based on the visualization, I observe distinct trends in customer activity and churn across different geographic locations.

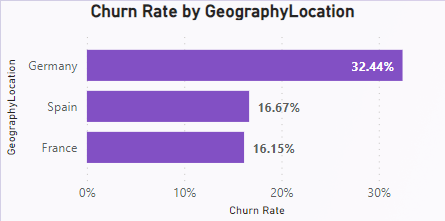


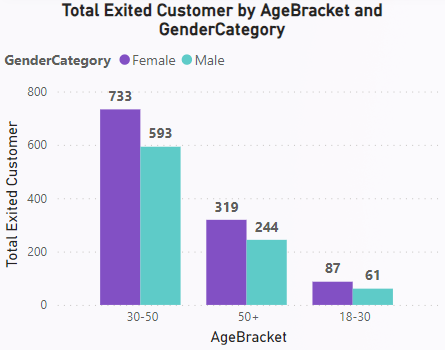
* France has the highest number of active customers and a significant but lower number of exited customers, indicating good retention.
* Spain shows moderate numbers with a lower churn rate, suggesting more stable customer behavior.
* Germany has a high churn rate, indicating a higher financial risk and potential issues with customer satisfaction.

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

Answer. Demographic Segments Posing Highest Financial Risk:

* Geographic risk: Germany with the highest churn rate (32.44%)
* Age and gender risk: Females aged 30-50 and 50+ show higher churn rates.
* Balance risk: High balances in the 18-30 age bracket, especially males ($102K)

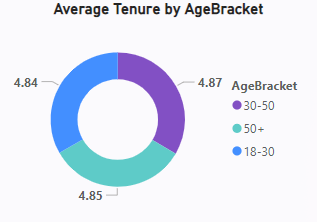




* Focus on improving satisfaction in Germany and among females aged 30-50 and 50+.
* Tailored engagement strategies to reduce churn and increase loyalty.

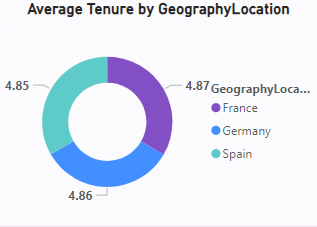
5.**Customer Tenure Value Forecast:** How would you use the available data to model and predict the lifetime (tenure) value in the bank of different customer segments?

Answer. **Average Tenure by Age Bracket:**



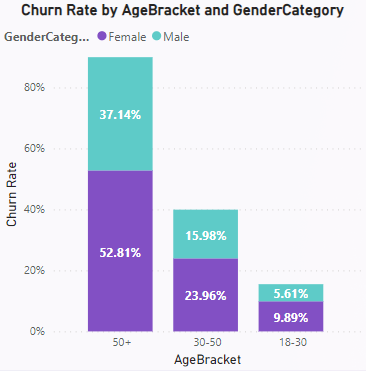
* The average tenure is highest in the 50+ age bracket (4.87 years).
* Recommendation: Enhance engagement and loyalty across all age groups.

**Average Tenure by Geography:**



* Average tenure is consistent across France (4.86 years), Germany (4.85 years), and Spain (4.87 years).
* Recommendation: Implement uniform retention strategies with regional customizations.

**Churn Rate by Age Bracket and Gender:**



* Customers aged 50+ have the highest churn rates, particularly females at 52.81%.
* Recommendation: Focus retention efforts on the 50+ age group, especially females.

**Conclusion**

Targeted retention efforts should be directed at older age groups, especially females aged 50+, who exhibit higher churn rates. Consistent retention strategies with regional adaptations will help predict and enhance customer lifetime value effectively.

6.**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?

Answer. Assessing the Impact on Customer Retention and Acquisition:

**Customer Retention**

* Compare churn rates before and after the marketing campaigns.
* Analyze if customers targeted by the campaigns have a lower churn rate.
* Look at engagement metrics, such as whether customers opened emails, clicked on links, or took advantage of special offers.

**Customer Acquisition**

* Measure the number of new customers acquired after the campaigns.
* Assess the quality of new customers by looking at their account balances, number of products used, and other relevant metrics.

**Additional Information Needed**

* Campaign Details: Information about what the campaigns were, when they ran, who they targeted, and what their goals were.
* Customer Response Data: Data on how customers interacted with the campaigns, such as click-through rates, purchases made, or services signed up for.
* Customer Demographics: Data on the characteristics of customers who responded to the campaigns, which can help tailor future campaigns.

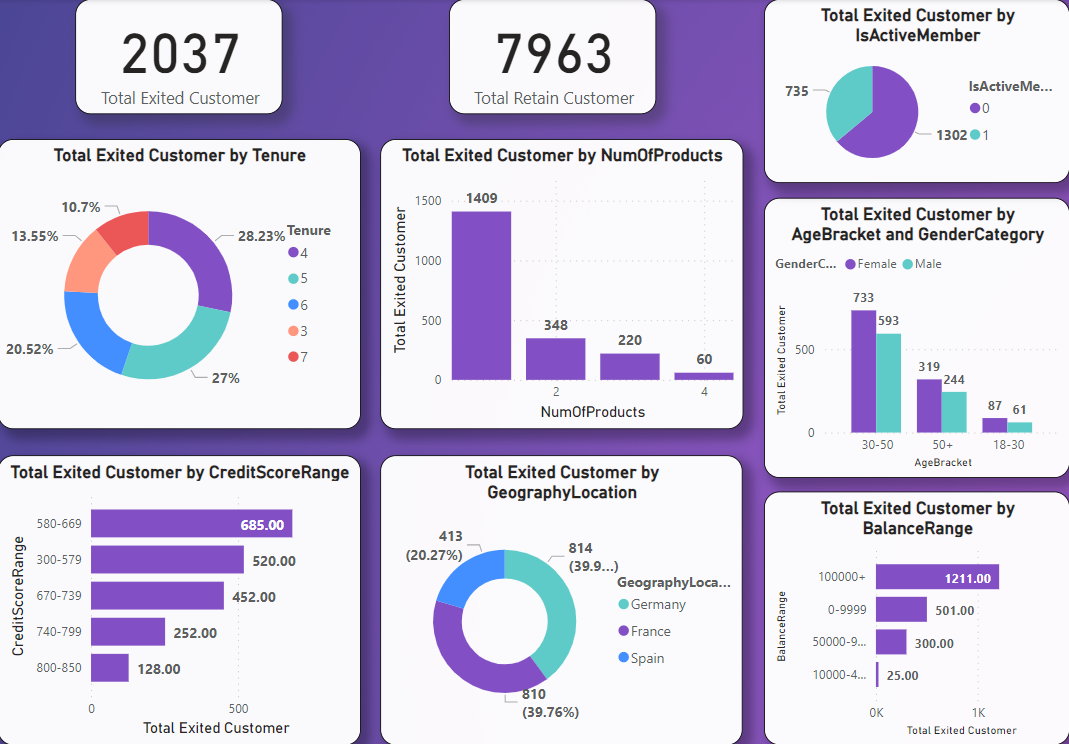
**Conclusion**

To effectively assess the impact of marketing campaigns on customer retention and acquisition, it is crucial to analyze customer engagement before and after the campaigns, gather detailed campaign information, and understand customer demographics and acquisition sources. This approach will provide insights into how marketing efforts are influencing customer behavior and help improve future campaign strategies.

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

Answer. **Analysis and Insights**

* A significant proportion of non-active members have exited, indicating that customer engagement and activity level are crucial for retention.
* Lower credit scores are correlated with higher exit rates. Customers with credit scores between 580-669 are most likely to exit, suggesting that financial instability might be a factor.
* Germany and France have the highest exit rates. This could be due to regional economic factors or competitive banking environments.



* The 30-50 age bracket, especially among females, shows the highest number of exits. This could be due to mid-life financial pressures or changing financial needs.
* Customers with only one product are more likely to leave. This suggests a need for better cross-selling strategies to increase customer engagement and retention.
* Customers with a tenure of 4 or 6 years are most likely to exit. This indicates critical points in the customer lifecycle where targeted retention efforts may be needed.
* Customers with higher balances (100000+) are more likely to exit. This might be due to higher expectations for service or better offers from competitors.

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

Answer.

1. Tenure: Tenure is a significant factor; shorter-tenure customers are more likely to churn.

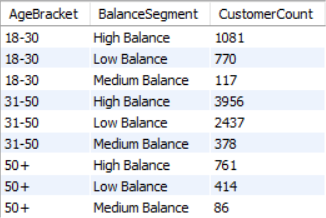
2. NumOfProducts: The number of products is a critical predictor of churn. Encouraging customers to use more products can enhance retention.

3. IsActiveMember: Customer activity status is a strong indicator of churn. Active members show greater loyalty.

4. EstimatedSalary: Estimated salary might not be as crucial for predicting churn compared to other factors like tenure, number of products, and activity status.

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

Answer.



* The majority of customers are in the '31-50' age group with high balances (3956 customers).
* The '18-30' age group has a significant number of high balance customers (1081), indicating younger customers with substantial balances.
* There are fewer customers in the '50+' age group, but a notable number still maintain high balances (761).
* Medium balance customers are fewer across all age groups, suggesting a polarization towards either high or low balances.

10.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?

Answer. To visually highlight customers at risk of churn and evaluate the impact of credit card rewards on customer retention, I used Power BI to implement conditional formatting and create insightful visualizations.

**1.Data Preparation**

Key columns: CustomerId, CreditScore, Tenure, Balance, NumOfProducts, HasCrCard, IsActiveMember, Exited, EstimatedSalary.

**2.Conditional Formatting**

Applied conditional formatting to highlight churn risk:

* Red for high churn risk (Exited = 1).
* Green for low churn risk (Exited = 0).

Visualized retention rates for customers with and without credit cards:

* Bar chart with conditional formatting to indicate higher retention rates in green and lower rates in red.

**3.Key Visualizations**

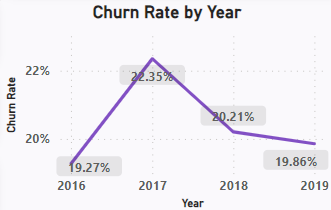
* Table: Highlighting churn risk and credit card ownership.
* Bar Chart: Comparing retention rates based on credit card ownership.
* Slicer: Filtering data by IsActiveMember to analyze churn and retention trends.

This setup helps in identifying at-risk customers and assessing the effectiveness of credit card rewards in retaining customers, guiding targeted marketing and retention strategies.

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?

Answer.

**1.Churn Rate Per Year**



**2.Overall Churn Rate:** The overall churn rate is 20.37%.

**3.Insights**

* Single Product Users: High churn rate of 27.71%.
* Lower Credit Scores: Higher churn in credit scores 580-669.
* Inactive Members: More likely to exit.
* Demographic Segments: Higher exit rates among 30-50 age group and females.
* Geography: Higher exit rates in France and Germany.
* Tenure: Highest exit rate within 1 year.

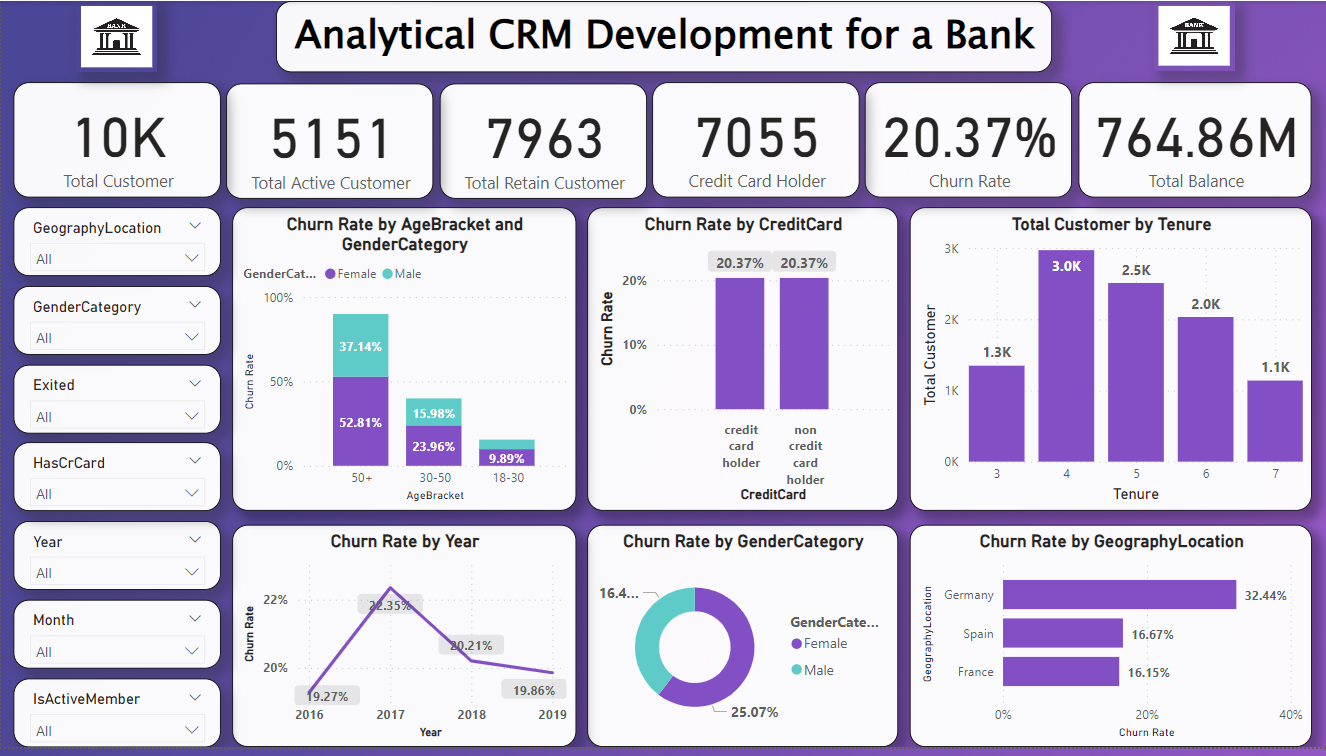
**4.Suggested Strategies**

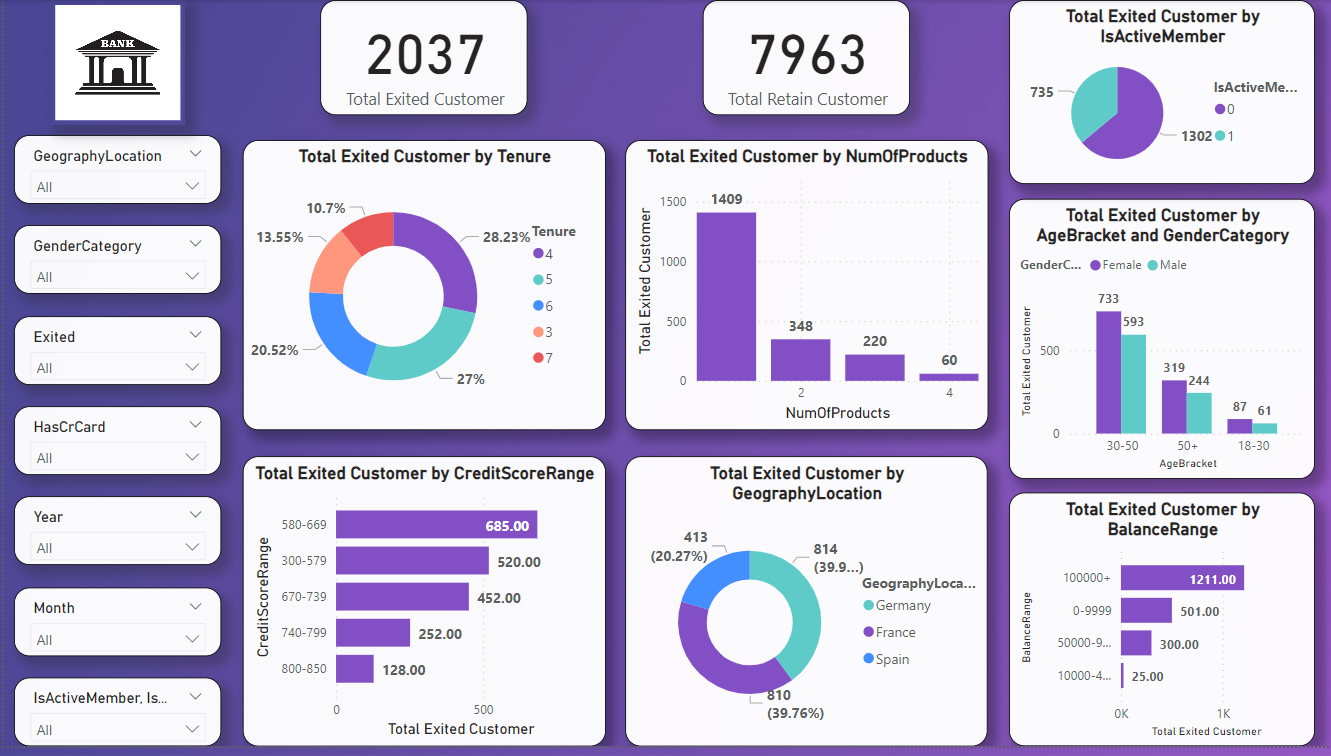
* Enhance Customer Engagement:
* Target single-product users with additional product offers.
* Reactivate inactive members with personalized engagement strategies.
* Credit Score Management:
* Provide financial advisory services for customers with lower credit scores.
* Offer tailored loan products or credit cards.
* Demographic-Specific Retention Programs:
* Design retention programs for the 30-50 age group and female customers.
* Conduct focus groups for specific needs and challenges.
* Geographical Focus:
* Implement region-specific retention strategies for France and Germany.
* Early Tenure Engagement:
* Develop onboarding programs for new customers.
* Provide proactive communication and support in the first year.

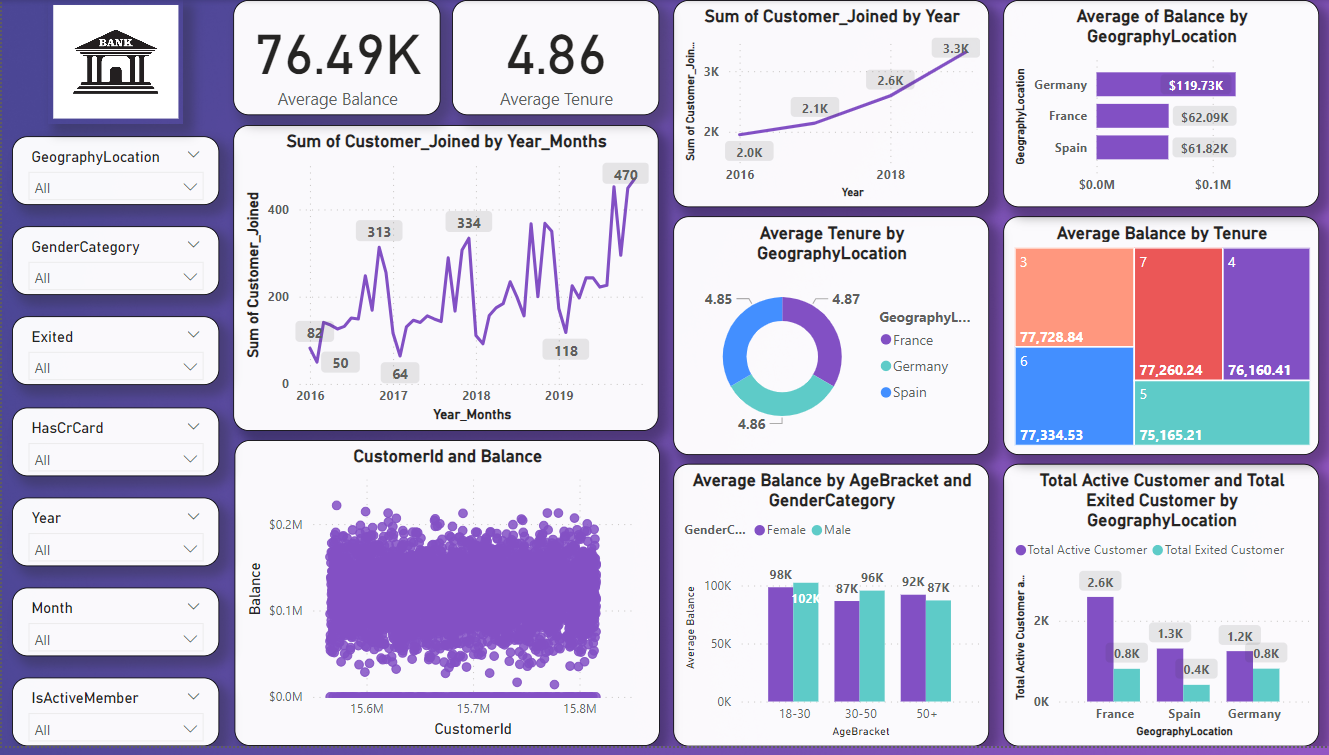
These strategies aim to enhance customer retention and reduce the overall churn rate.

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

Answer. The dashboard provides an in-depth analysis of various KPIs related to customer behavior, churn, demographics, and financial metrics. It includes interactive visualizations and slicers to assist in dynamic data selection and filtering.







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

Anwer. If the objective and subjective questions were not given, I would approach the problem of understanding the dataset and deriving actionable insights using a structured data analysis process. Here's how I would proceed:

**1. Understanding the Dataset:**

* Inspect the structure, types, and descriptions of the datasets provided.
* Identify key tables, columns, and their relationships.

**2. Data Cleaning and Preparation:**

* Identify and address missing or null values through imputation or removal.
* Ensure that data types are appropriate for analysis (e.g., numerical, categorical).
* Create derived columns (e.g., age brackets, balance ranges) for better analysis.

**3. Identifying Key Metrics:**

* Calculate overall churn rate and churn rate by different segments (e.g., age, geography).
* Segment customers based on demographics, product usage, and account details.

**4. Data Visualization:**

* Create visualizations to show the distribution of customers across different segments (e.g., age, gender, geography).
* Visualize churn rate by various factors such as credit score, account balance, and product usage.
* Show trends in customer retention and identify high-risk segments.

**5. Reporting and Dashboard Creation:**

* Develop interactive dashboards in tools like Power BI or Tableau to visualize key metrics and trends.
* Compile a comprehensive report detailing the analysis, findings, and recommendations.

By following this structured approach, I would be able to analyze the dataset comprehensively, uncover key insights, and make data-driven recommendations to address customer churn, improve retention, and enhance overall customer satisfaction.

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

Answer.

***ALTER TABLE Bank\_Churn***

***RENAME COLUMN HasCrCard TO Has\_creditcard;***

This query alters the Bank\_Churn table by renaming the column HasCrCard to Has\_creditcard.