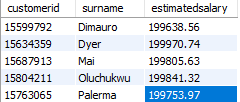
**OBJECTIVE QUESTIONS :**

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

The distribution of account balance across different regions is as follows: in France, the total account balance is 311 million; in Germany, it is 300 million; and in Spain, it is 153 million.

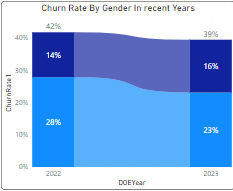
1. Identify the top 5 customers with the highest Estimated Salary in the last quarter of the year. (SQL)

These are the top 5 highest estimated salary in the last quarter of the year.

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

These are the average number of products used by the customer who has Credit card.

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

The churn rate by gender in the recent year is as follows: in 2022, the churn rate is 42%, with 14% for male gender and 28% for female gender; in 2023, the churn rate is 39%, with 16% for male gender and 23% for female gender. Comparing these recent years, 2022 has the higher churn rate.

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

 Comparing the average credit score for customers who have exited and those who have remained, exited customers have a higher average credit score than those who have remained.

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

 Here, you can see that genderID 2, which is female, has a higher average salary and fewer active accounts compared to genderId 1, which is male.

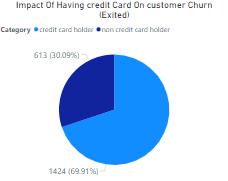
1. Segment the customers based on their credit score and identify the segment with the highest exit rate. (SQL)

After identifying the credit score segment, the segment with the highest exit rate is the LOW segment of credit scores.

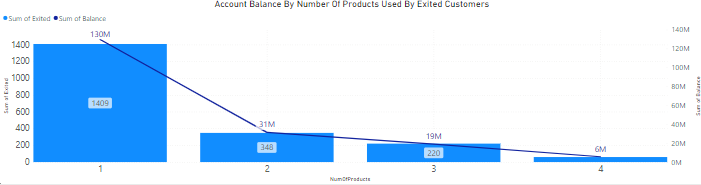
1. Find out which geographic region has the highest number of active customers with a tenure greater than 5 years. (SQL)

 The geographyID 1, which is France, has the highest count of active customers, with 797 having a tenure greater than 5 years.

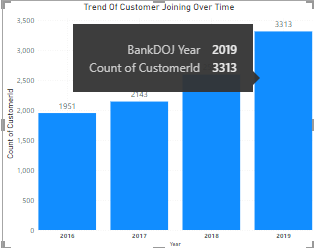
1. What is the impact of having a credit card on customer churn, based on the available data?

 Based on the available data, customers who have a credit card have a higher churn rate than those who don’t have a credit card. 69.91% of customers with a credit card have churned the bank. This might be due to mismanagement of credit cards by customers, leading to higher churn rates.

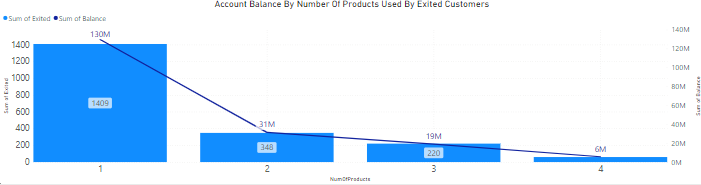
1. For customers who have exited, what is the most common number of products they had used?

  
  
The most common number of products used by customers who have exited is number 1, with 1409 customers using this product.

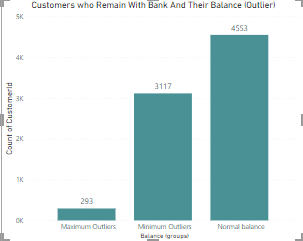
1. Examine the trend of customer joining over time and identify any seasonal patterns (yearly or monthly). Prepare the data through SQL and then visualize it.

 In this image, you can see that 2019 has the highest count of customers joining, indicating a trend of increasing customer numbers over time.

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

  
Product number 1 has the highest account balance of 130M compared to other products for exited customers. Product 1 also has the highest count of exited customers.

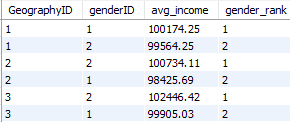
1. Identify any potential outliers in terms of balance among customers who have remained with the bank.

 The analysis reveals three distinct categories of customers based on their account balances. There are 293 customers classified as maximum outliers, indicating those with balances significantly higher than the majority. Additionally, 3,117 customers fall into the minimum outlier category, with zero balance. The majority, 4,553 customers, have balances within the normal range.

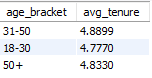
1. How many different tables are given in the dataset, out of these tables which table only consist of categorical variables?

In the dataset, there are 7 tables given, including ActiveCustomers, Bank\_Churn, CreditCard, Customerinfo, ExitCustomers, Gender, and Geography. The Bank\_Churn and Customerinfo tables consist of categorical variables such as age, gender, hascrcard, and GeographyID.

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

 In this table, you can see the gender-wise average income of male and female in each geographyID. After ranking the gender according to the average value, you can see that geographyID 1 has the first rank, which is a male gender, with an average salary of 100174.25.

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+).

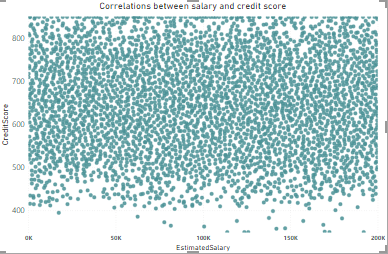
In the output you can see the age\_brackets and avg\_tenure for the people who have exited.

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

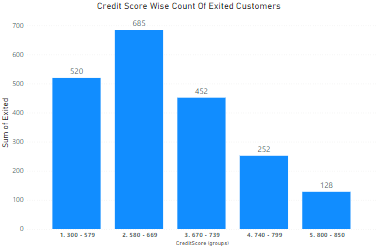
In the visualization, it is evident that there is no direct correlation between salary and account balance for customers, regardless of whether they have exited or not. Surprisingly, some customers with lower salaries have higher account balances, while others with higher salaries have lower balances. This suggests a complex relationship between salary and account balance, indicating that factors beyond salary alone influence the account balances of customers.

Top of Form

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

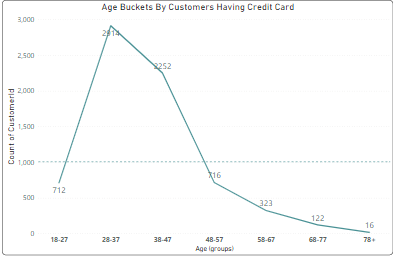


In the visualization, it is apparent that there is no correlation between the salary and credit score of customers. Surprisingly, some customers with high salaries have low credit scores, while others with low salaries have high credit scores. This lack of correlation suggests that factors other than salary significantly influence the credit scores of customers, highlighting the complexity of creditworthiness assessment.

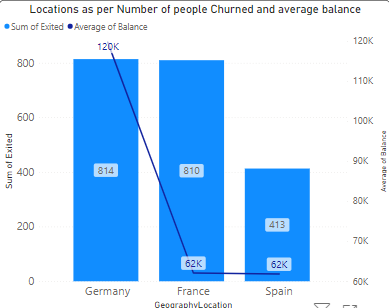
1. . Rank each bucket of credit score as per the number of customers who have churned the bank.

In the chart of credit score wise exited customers, the bucket of 580-669 has the highest number of customers, with a count of 685 compared to other credit score buckets. So, we can say that the 580-669 bucket of credit score has the highest churning count for the bank.

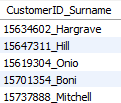
1. According to the age buckets find the number of customers who have a credit card. Also retrieve those buckets who have lesser than average number of credit cards per bucket.

In the analysis based on age buckets, it is observed that certain age groups have a lower number of customers with credit cards compared to the average number of credit cards per bucket. Specifically, age groups 18-27, 48-57, 58-67, 68-77, and 78+ exhibit this pattern, indicating a potential opportunity to target these segments for increasing credit card adoption. This insight highlights the importance of understanding age demographics in credit card marketing strategies.

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

Based on the data, Germany leads in the number of customers who have churned the bank, with a count of 814. Furthermore, Germany also boasts an average balance of 120,000, leading to its rank of 1 in terms of both churn count and average balance among the analyzed locations.

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”.

I have used this query for the specific question.

select concat(ci.customerID, '\_', ci.surname) as CustomerID\_Surname

from customerinfo ci

join bank\_churn bc ON ci.CustomerID = bc.customerID;

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

Yes, you can retrieve the "ExitCategory" from the "ExitCustomers" table to the "Bank\_Churn" table without using a JOIN by using a subquery.

SELECT

bc.\*,

(SELECT ec.ExitCategory FROM exitCustomer ec WHERE ec.ExitID = bc.Exited) AS ExitCategory

FROM

Bank\_Churn bc;

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

**Missing Values**

There were no missing values in the dataset, except for the BankDOJ column which was used to calculate the DOEyear based on the Tenure of customers.

**Handling Missing Values**

If there were missing values in the dataset, some common ways to handle them include:

**Removing Rows:** Delete rows with missing values, but this may lead to loss of valuable data.

**Mean/Median/Mode Imputation:** Fill missing values with the mean, median, or mode of the column. This is a simple method but may not be ideal for all cases.

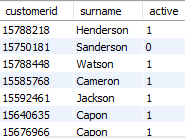
**Forward/Backward Fill:** Use the value from the previous or next row to fill missing values, which is useful for time-series data.

**Duplicates**

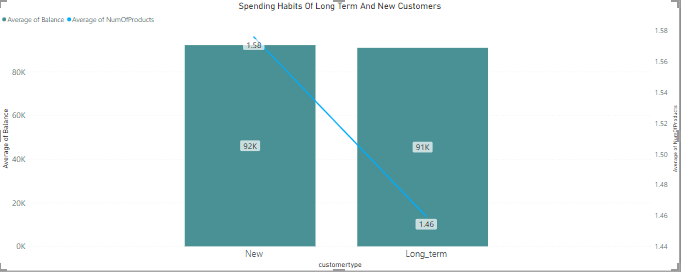
There were no duplicates present in the dataset, ensuring each record was unique based on the defined criteria.

By addressing these aspects, the dataset was cleaned and prepared for analysis, ensuring the integrity and quality of the data.

1. Write the query to get the customer ids, their last name and whether they are active or not for the customers whose surname ends with “on”.

This is the output for the customerid whos surname ends with on and whether they are active or not.

**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?  
   **Observation**:

New customers, on average, use more bank products (1.58 vs. 1.46) and maintain a slightly higher average balance (92k vs. 91k) compared to long-term customers. This indicates that new customers exhibit more active spending habits than long-term customers.

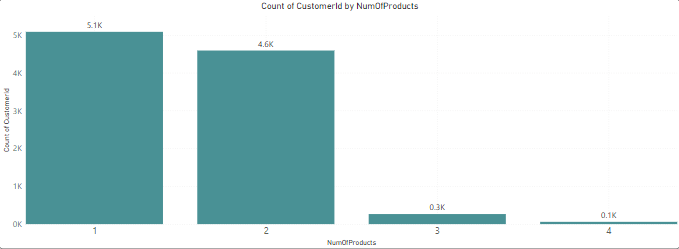
**Implication:**

The minimal difference in average balances suggests that while new customers are more engaged with products, their financial standing is not significantly higher than long-term customers. This analysis suggests that new customers may represent a segment with greater potential for growth and deeper engagement over time.

**Recommendation:**

To foster customer loyalty, the bank should focus on nurturing relationships with new customers, as they show a willingness to engage with a wider range of products. By providing tailored services and personalized experiences, the bank can encourage long-term loyalty and enhance customer satisfaction.

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



**Observation:**

5.1k customers are using 1 product ("Savings Account").

4.6k customers are using 2 products ("Savings Account" and "Credit Card").

0.3k customers are using 3 products ("Savings Account," "Credit Card," and "Personal Loan").

0.1k customers are using 4 products ("Savings Account," "Credit Card," "Personal Loan," and "Insurance").

**Implication:**

There is a substantial number of customers using 2 products, indicating a potential market for bundled offerings or targeted promotions to encourage additional product adoption.

However, the number of customers using 3 or 4 products is relatively low, suggesting a need for more personalized and targeted marketing strategies to increase adoption rates for these products.

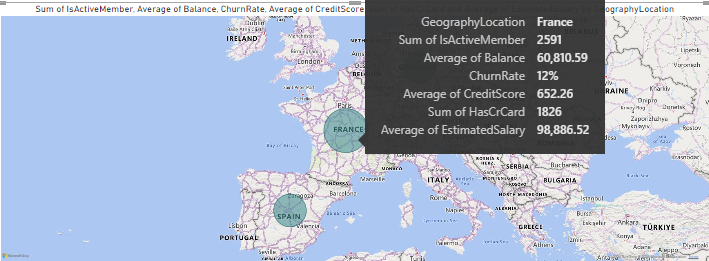
**Recommendation:**

To increase the number of customers using 3 products, the bank could offer bundled packages that include "Savings Account," "Credit Card," and "Personal Loan," along with incentives such as reduced fees or interest rates.

Similarly, for customers using 4 products, the bank could create tailored marketing campaigns that highlight the benefits of having all four products and offer incentives for customers to add the additional products to their portfolio.

By implementing these strategies, the bank can improve its cross-selling efforts, enhance customer engagement, and increase overall revenue.

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

**Observation**:

Across Germany, France, and Spain, there are notable differences in economic indicators and banking metrics. Germany has the highest average balance and churn rate, France leads in credit card usage, and Spain has the highest average estimated salary.

**Implication:**

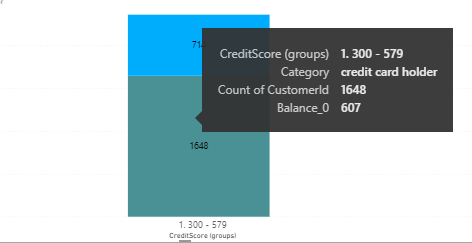
These differences suggest varying levels of customer engagement, banking behavior, and economic conditions in each region. Germany's high churn rate may indicate a need for improved customer retention strategies. France's high credit card usage could present opportunities for targeted promotions or product offerings. Spain's high average estimated salary may indicate a strong consumer base with potential for investment or wealth management products.

**Recommendation:**

For Germany, focusing on improving customer retention strategies, such as personalized offers or loyalty programs, could help reduce churn rates. In France, leveraging the popularity of credit cards by offering rewards or incentives could further increase customer engagement. In Spain, developing investment or wealth management products tailored to high-income earners could attract more customers and drive revenue growth.

Overall, understanding the correlation between economic indicators and banking metrics can help banks develop targeted strategies to meet the specific needs and preferences of customers in each region, ultimately improving customer satisfaction and driving business success.

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



**Observation:**

Created groups of credit scores and filtered the lowest credit scores customers who has 0 balnce in their account. Among customers with low credit scores (300-579), there are 1,648 credit card holders and 714 non-credit card holders.

Within this segment, there are 607 credit card holders and 267 non-credit card holders with a 0 balance in their account.

**Implication:**

Customers with low credit scores who also have credit cards may pose a higher financial risk to the bank, as they may have difficulty managing their finances or making timely payments.

The presence of a 0 balance in the account among these customers suggests a potential inability to repay debts or meet financial obligations.

**Recommendation:**

For customers with low credit scores who also have credit cards, the bank should closely monitor their account activity and payment behavior to identify any signs of financial distress early.

The bank could consider offering financial education programs or debt management assistance to help these customers improve their financial literacy and ability to manage their credit responsibly.

For customers with low credit scores who do not have credit cards, the bank could explore offering secured credit card options or alternative financial products that can help them build credit in a responsible manner.

By identifying and addressing the financial risks associated with these demographic segments, the bank can mitigate potential losses and better serve its customers' financial needs.

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

To model and predict the lifetime value of different customer segments, we can use the available data on customer demographics and behavior, such as geography locations, credit card usage, age, gender, and number of products used. Here's how we can approach it:

**Data Preparation:** Start by gathering historical data on customer interactions, purchases, and tenure. This data should include information on the customer's location, age, gender, credit card usage, and the number of products they use.

**Segmentation:** Divide the customer base into segments based on relevant factors such as geography, age group, gender, credit card usage, and number of products used. This segmentation will help us analyze and predict the lifetime value of different customer segments more effectively.

**Feature Engineering:** Create additional features that can help in predicting customer lifetime value, such as average transaction value, frequency of purchases, and customer tenure.

**Model Selection:** Choose appropriate machine learning models for prediction, such as regression models or machine learning algorithms like Random Forest or Gradient Boosting. These models can help us understand the relationship between customer demographics and behavior and their lifetime value.

**Model Training and Validation:** Split the data into training and validation sets. Train the model on the training data and validate its performance on the validation data to ensure its accuracy and reliability.

**Prediction:** Once the model is trained and validated, use it to predict the lifetime value of different customer segments. This prediction can help in understanding which segments are likely to be the most valuable over time.

**Insights and Recommendations:** Based on the predictions, identify the most valuable customer segments and develop targeted strategies to enhance their lifetime value. This could include personalized marketing campaigns, loyalty programs, or product offerings tailored to their needs.

By following these steps and leveraging the available data on customer demographics and behavior, we can effectively model and predict the lifetime value of different customer segments, enabling us to make informed decisions and drive business growth.

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

To assess the impact of marketing campaigns on customer retention and acquisition, you can follow these steps:

**Define Metrics**: Use metrics like retention rate (percentage of customers retained) and acquisition rate (percentage of new customers acquired).

**Data Preparation:** Merge churn and customer information tables based on customer ID.

**Analysis:**

Compare retention rates before and after each campaign.

Analyze acquisition rates during and after each campaign.

Segment data by demographics to identify responsive customer segments.

**Additional Information Needed:**

Details of each marketing campaign.

Customer interaction data (e.g., clicks, responses).

Competitor data and market trends.

Customer feedback on campaigns and services.

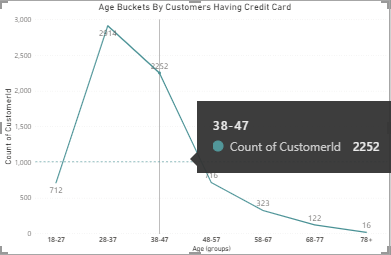
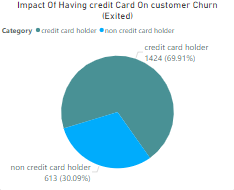
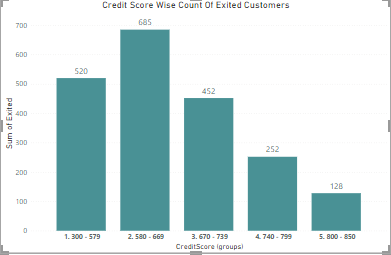
**Evaluation:**

Use statistical methods to determine the significance of changes in retention and acquisition rates.

Evaluate ROI by comparing campaign costs to revenue increase.

**Recommendations:** Provide data-driven recommendations for future campaigns, such as targeting specific segments or using different channels.

By following these steps and gathering the necessary information, you can assess the impact of marketing campaigns on customer retention and acquisition.

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

**Observation:**

Customers with credit cards have a higher churn rate (69.91%) compared to those without credit cards (30.09%).

The highest churn rate is observed in the credit score range of 580-669, with 685 customers exiting. This is followed by the credit score range of 300-579, with 520 customers exiting. The credit score range of 670-719 has 452 exited customers.

The age bracket of 28-37 has the highest number of customers with credit cards (2814), followed by the age bracket of 38-47 with 2252 customers. Other age brackets have fewer customers with credit cards, below the average line.

**Implication:**

Customers with credit cards may be more likely to churn, possibly due to dissatisfaction with the bank's services, rewards, or fees associated with the credit card.

Customers in the credit score range of 580-669 and 300-579 may be facing financial difficulties or may be dissatisfied with the bank's offerings, leading to higher churn rates.

The age bracket of 28-37, which has the highest number of customers with credit cards, may be more prone to churning. This could be due to life events such as job changes, relocation, or financial obligations.

**Recommendation:**

Offer personalized retention offers or rewards to customers with credit cards to improve their satisfaction and reduce churn.

Provide financial planning services or tailored products to customers in the credit score ranges of 580-669 and 300-579 to address their financial needs and improve retention.

Focus on understanding the specific needs and preferences of customers in the age bracket of 28-37 to enhance their loyalty and reduce churn.

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

We can assess the importance of 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' for predicting if a customer will leave the bank by considering their potential impact on customer behavior.

**Tenure:**

Insight: Tenure refers to the length of time a customer has been with the bank.

Potential Impact: Customers with longer tenure may have developed stronger relationships with the bank, leading to higher loyalty and lower churn rates. However, customers with very short tenure may not have fully integrated with the bank's services and could be more likely to leave.

Importance for Prediction: Tenure could be an important predictor, as customers with shorter tenure may be more likely to churn compared to those with longer tenure.

**NumOfProducts:**

Insight: NumOfProducts refers to the number of bank products/services a customer is using.

Potential Impact: Customers using more products may be more engaged with the bank and less likely to leave, as they have invested more in the relationship. However, customers with too many products may find it overwhelming or costly, increasing their likelihood to churn.

Importance for Prediction: NumOfProducts could be important, as it reflects the level of customer engagement and commitment to the bank.

**IsActiveMember:**

Insight: IsActiveMember indicates whether a customer is an active bank member.

Potential Impact: Active members are likely more engaged with the bank and satisfied with its services, leading to lower churn rates. Conversely, inactive members may be dissatisfied or disengaged, increasing their likelihood to leave.

Importance for Prediction: IsActiveMember could be a significant predictor, as inactive members may be more likely to churn compared to active members.

**EstimatedSalary:**

**Insight:** EstimatedSalary is the approximate annual salary of the customer.

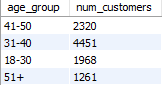
**Potential Impact:** Higher-income customers may have more financial stability and could be less likely to churn. However, they may also have higher expectations for service quality and personalized offerings.

**Importance for Prediction:** EstimatedSalary could be relevant, as customers with lower salaries may be more sensitive to fees or service issues, increasing their likelihood to churn.

In conclusion, 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' are potentially important for predicting if a customer will leave the bank. Analyzing these factors along with other relevant variables could provide valuable insights for predicting and reducing customer churn.

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

**Segment customers based on age groups:**

SELECT

CASE

WHEN age BETWEEN 18 AND 30 THEN '18-30'

WHEN age BETWEEN 31 AND 40 THEN '31-40'

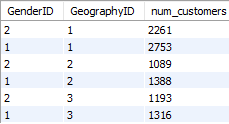
WHEN age BETWEEN 41 AND 50 THEN '41-50'

ELSE '51+'

END AS age\_group, COUNT(\*) AS num\_customers

FROM customerinfo

GROUP BY age\_group;

**Segment customers based on gender and geography:**

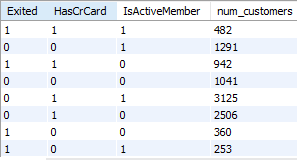
SELECT GenderID, GeographyID, COUNT(\*) AS num\_customers

FROM customerinfo

GROUP BY GenderID, GeographyID

ORDER BY GeographyID;

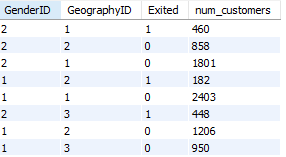
**Segment customers based on account details:**



SELECT Exited, HasCrCard, IsActiveMember, COUNT(\*) AS num\_customers

FROM bank\_churn

GROUP BY Exited, HasCrCard, IsActiveMember;

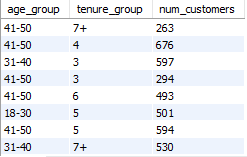
**Segment customers based on gender, geography, and churn status:**

SELECT GenderID, GeographyID, Exited, COUNT(\*) AS num\_customers

FROM customerinfo ci

JOIN bank\_churn bc ON ci.CustomerID = bc.CustomerID

GROUP BY GenderID, GeographyID, Exited;

**Segment customers based on age groups and tenure:**

SELECT

CASE

WHEN age BETWEEN 18 AND 30 THEN '18-30'

WHEN age BETWEEN 31 AND 40 THEN '31-40'

WHEN age BETWEEN 41 AND 50 THEN '41-50'

ELSE '51+'

END AS age\_group,

CASE

WHEN tenure = 3 THEN '3'

WHEN tenure = 4 THEN '4'

WHEN tenure = 5 THEN '5'

WHEN tenure = 6 THEN '6'

ELSE '7+'

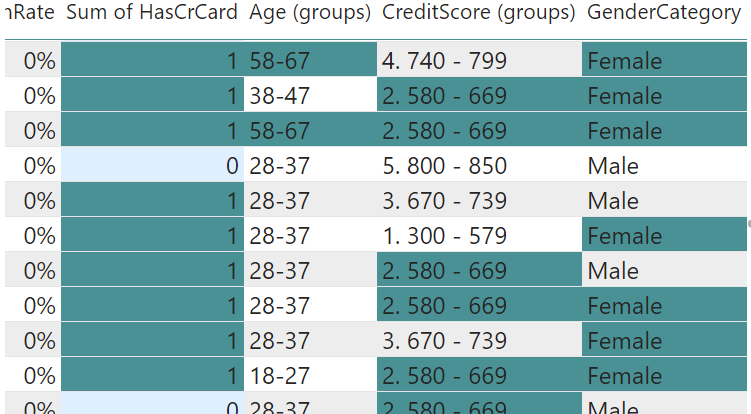
END AS tenure\_group, COUNT(\*) AS num\_customers

FROM customerinfo ci

JOIN bank\_churn bc ON ci.CustomerID = bc.CustomerID

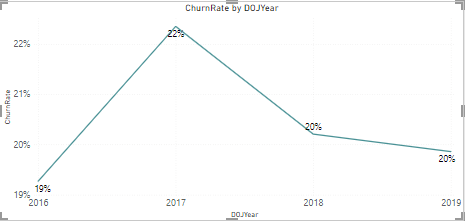
GROUP BY age\_group, tenure\_group;

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?



* **Churn Rate Definition:** Percentage of customers who have stopped using the service.
* **Retention Rate Definition:** Percentage of customers who continue to use the service.
* **Objective:** Evaluate impact of credit card rewards on retention and identify churn patterns.
* **Conditional Formatting: Highlights:**
* Customers with credit cards (higher churn).
* Age groups 45-57 and 58-67 (higher churn).
* Credit scores 580-669 (higher churn).
* Females (higher churn).
* **Results:**
* Customers with credit cards show higher churn.
* Age groups 45-57 and 58-67 have higher churn.
* Lower credit scores correlate with higher churn.
* Females exhibit higher churn.
* **Conclusion:**
* Enhance credit card rewards for better retention.
* Tailor retention strategies for specific age groups, credit score ranges, and genders.
* **Recommendations:**
* Improve credit card reward programs.
* Implement targeted retention strategies for different demographics.

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 are the different strategies that can be used to decrease the churn rate.



Churn Rates:

* In 2017, 19%
* In 2018, 22%
* In 2019, 20%

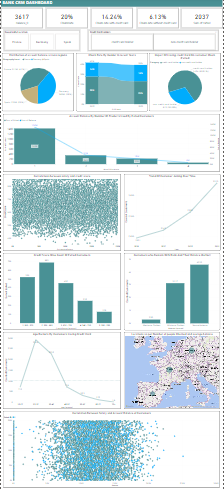
Key Insights:

* Customers aged 48-57 have a 58% churn rate.
* Customers aged 58-67 have a 40% churn rate.
* Customers using 1 product have a churn count of 1409.
* Customers using 2 products have a churn count of 348.
* Customers using 3 products have a churn count of 220.
* Customers using 4 products have a churn count of 60.

Strategies to Decrease Churn:

* Targeted marketing to 48-57 age group.
* Personalized offers for customers with 1 product.
* Improved customer service for 58-67 age group.
* Product bundling to incentivize multiple product usage.
* Customer loyalty programs for increased retention.
* By focusing on these strategies, the bank can potentially decrease its churn rate and improve customer retention.

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?

As a data analyst approaching customer churn analysis in a banking dataset without specific questions, I would:

* Familiarize with the dataset to understand its structure and variables.
* Conduct exploratory data analysis (EDA) to uncover patterns and relationships.
* Create new features or transform existing ones to enhance analysis.
* Analyze churn rates and factors using statistical tests or machine learning models.
* Develop interactive dashboards to visualize insights effectively.
* Summarize findings and provide actionable recommendations for reducing churn.
* Iterate on the analysis based on feedback and new insights to refine recommendations.

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

ALTER TABLE Bank\_Churn

RENAME COLUMN HasCrCard TO Has\_creditcard;