**Learners have to come up with a Report to support the answers to the following questions and suggestions**

**Objective Questions:**

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

Ans: SELECT geo.GeographyLocation,

ROUND(SUM(bc.Balance),2) AS balances

FROM bank\_churn bc

JOIN customerinfo ci USING (CustomerID)

JOIN geography geo USING (GeographyID)

GROUP BY geo.GeographyLocation ;

Output:

A screenshot of a computer

Description automatically generated

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

ANS:

SELECT \*

FROM customerinfo

WHERE month(BankDOJ) IN (10,11,12)

ORDER BY EstimatedSalary DESC

LIMIT 5;

Output:

A screenshot of a computer

Description automatically generated

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

Ans:

SELECT CustomerId, ROUND(AVG(NumOfProducts),0) As avg\_products

FROM bank\_churn

WHERE HasCrCard = 1

GROUP BY CustomerId;

Output:

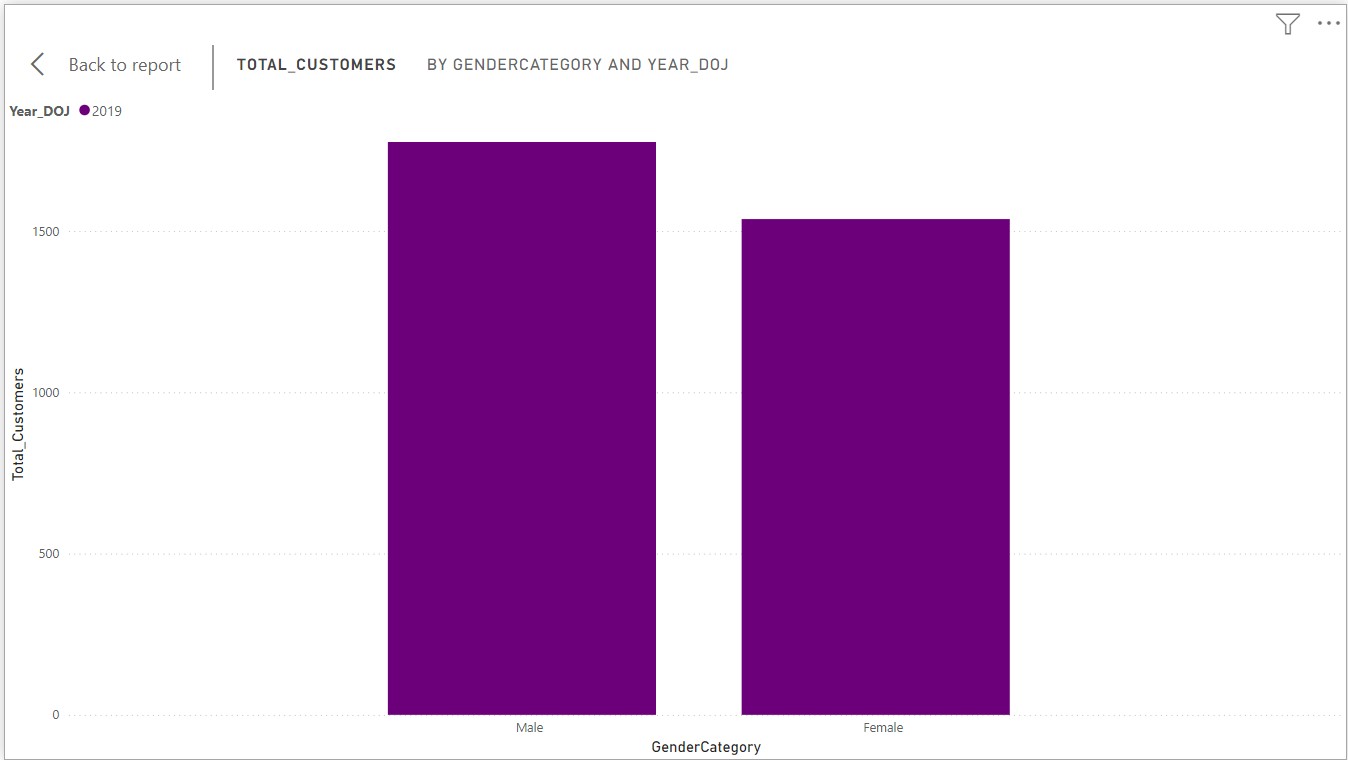
A screenshot of a computer

Description automatically generated

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

Ans:

* The bar graph shows the total number of customers by gender category and year DOJ for the year 2019.
* In the year 2019, there were more male customers than female customers. There were 1776 male customers and 1535 female customers.
* The total number of customers in the year 2019 is 3313.



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

**ANS:**

SELECT (CASE WHEN Exited = 1 THEN 'Exited' ELSE 'Remain'END ) AS exited\_remain,

AVG(CreditScore) AS avg\_creditscore

FROM bank\_churn

GROUP BY exited\_remain

Output:

A screenshot of a computer

Description automatically generated

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

ANS:

WITH active\_avg\_est\_salary AS

(SELECT g.GenderCategory AS gender, ROUND(AVG(c.EstimatedSalary),2) AS active\_avg\_est\_salary

FROM customerinfo c

JOIN bank\_churn b USING (CustomerID)

JOIN gender g USING (genderID)

WHERE IsActiveMember = 1

GROUP BY g.GenderCategory),

inactive\_avg\_est\_salary AS

(SELECT g.GenderCategory AS gender, ROUND(AVG(c.EstimatedSalary),2) AS inactive\_avg\_est\_salary

FROM customerinfo c

JOIN bank\_churn b USING (CustomerID)

JOIN gender g USING (genderID)

WHERE IsActiveMember = 0

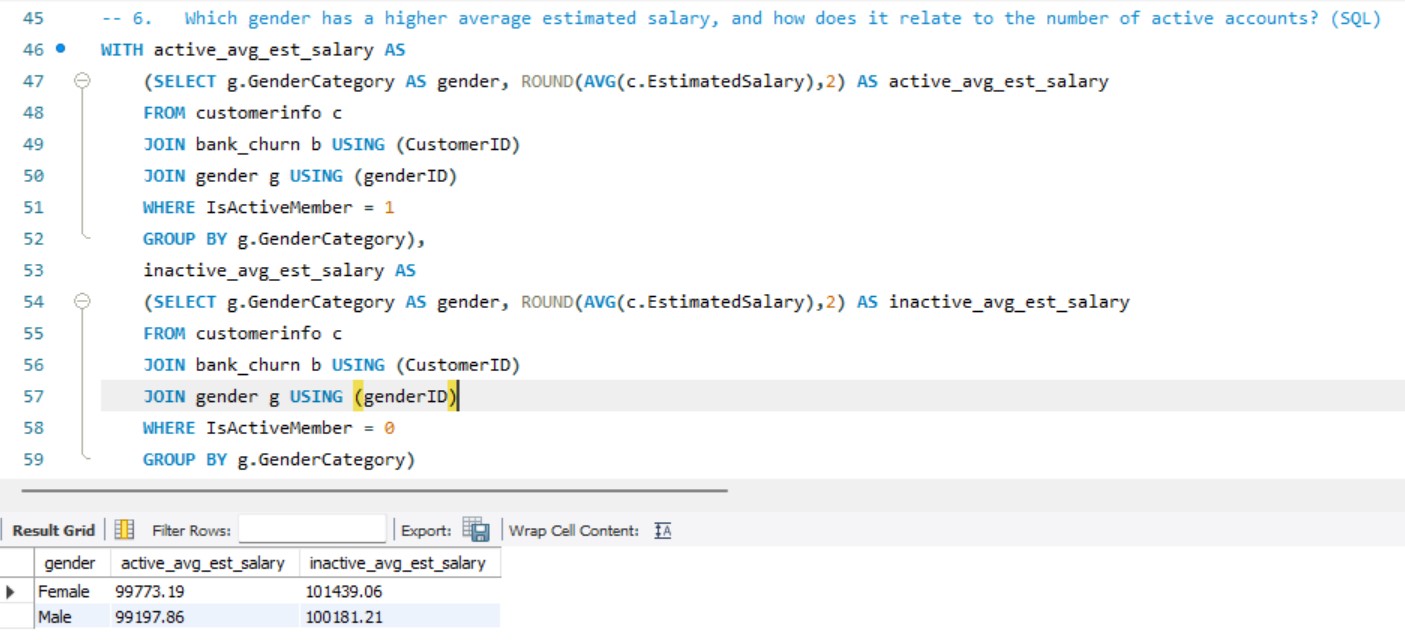
GROUP BY g.GenderCategory)

SELECT \*

FROM active\_avg\_est\_salary a

JOIN inactive\_avg\_est\_salary i USING(gender);

Output:



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

SELECT CASE WHEN CreditScore < 600 THEN 'Poor(Less Than 600)'

WHEN CreditScore >= 600 AND CreditScore < 700 THEN 'Fair(Between 600 And 700)'

WHEN CreditScore >= 700 AND CreditScore < 800 THEN 'Good(Between 700 And 800)'

ELSE 'Excellent(More than 800)'

END AS segments, Count(Exited) As cnt\_exited

FROM bank\_churn

WHERE Exited = 1

GROUP BY segments

ORDER By cnt\_exited DESC;

Output:

A screenshot of a computer

Description automatically generated

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

ANS: SELECT g.GeographyLocation, Count(c.CustomerId) AS active\_customers

FROM customerinfo c

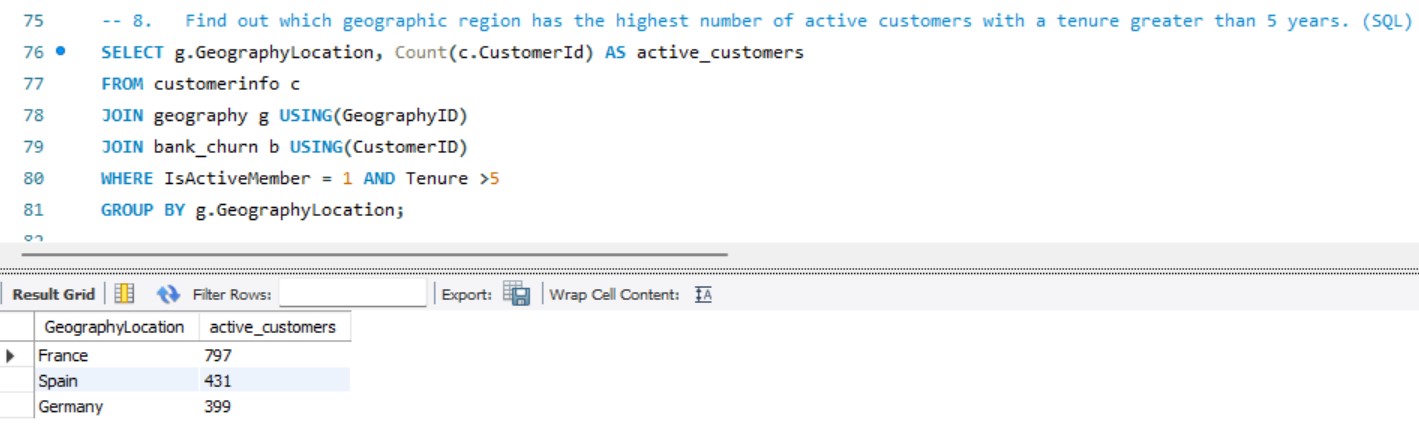
JOIN geography g USING(GeographyID)

JOIN bank\_churn b USING(CustomerID)

WHERE IsActiveMember = 1 AND Tenure >5

GROUP BY g.GeographyLocation

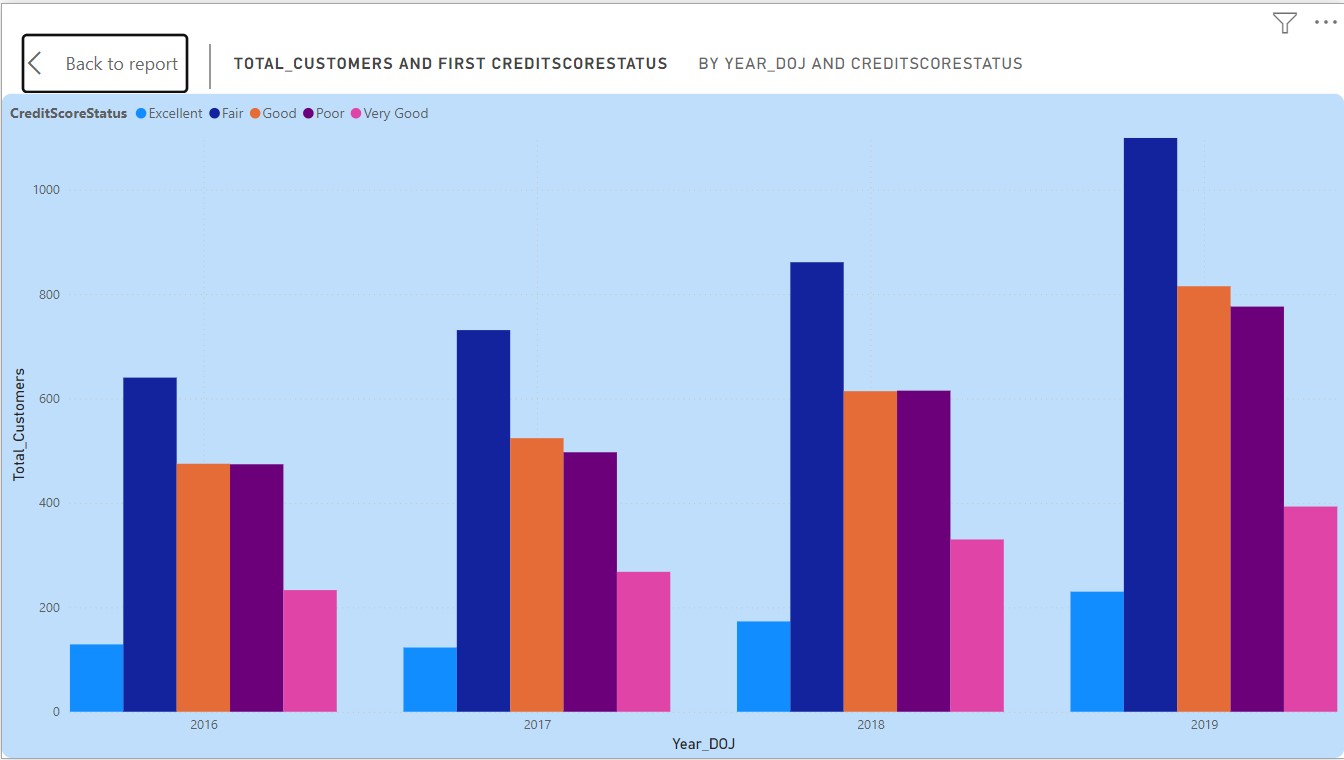
Output:



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

Ans:

* Exit Trend by Year and Credit Score: The number of exiting customers generally increases from 2016 to 2019, regardless of their credit score status.
* Fair Credit Scores Have Highest Exits: Customers with fair credit scores consistently show the highest exit rates across all years from 2016 to 2019.
* Excellent Scores See Lowest Exits: Those with excellent credit scores have the lowest exit rates, indicating higher retention among these customers.



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

Ans:

|  |  |
| --- | --- |
| NumOfProducts | ExitCustomers |
| 1 | 1409 |
| 2 | 348 |
| 3 | 220 |
| 4 | 60 |

* High Churn Rate for Single Product Users: Customers with only one product have the highest exit rate, with 1409 leaving.
* Decreasing Trend with More Products: Exit numbers significantly drop as customers use more products, with only 60 leaving when having four products.

SELECT NumOfProducts,count(CustomerId) AS no\_of\_customers

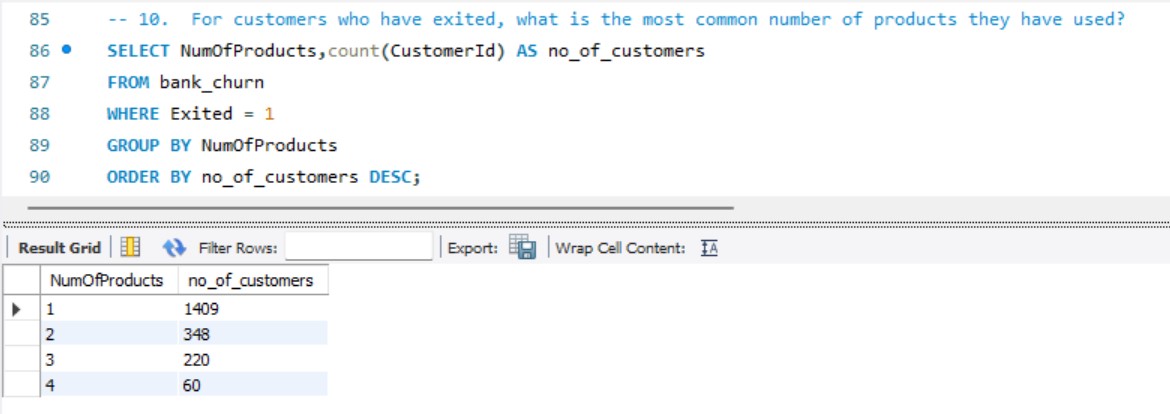
FROM bank\_churn

WHERE Exited = 1

GROUP BY NumOfProducts

ORDER BY no\_of\_customers DESC;

Output:



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.

Ans:

* Seasonal Peaks: The number of customers peaks annually in April and December, indicating possible seasonal factors influencing customer behavior.
* Overall Growth: There is a general upward trend in customer numbers from 2016 to 2019, showing consistent growth over the years.
* Fluctuating Monthly Patterns: Monthly customer numbers show significant fluctuations, with higher variability observed towards the end of the period.

SELECT YEAR(BankDOJ) AS join\_year,

MONTHNAME(BankDOJ) AS join\_month,

COUNT(CustomerID) AS Customers

FROM customerinfo

GROUP BY join\_year,join\_month

ORDER BY join\_year DESC,join\_month;

Output:

A screenshot of a computer

Description automatically generated

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

Ans: SELECT NumOfProducts,

Round(AVG(Balance),2) AS avg\_balance

FROM bank\_churn

WHERE Exited =1

GROUP BY NumOfProducts

Order BY NumOfProducts ASC;

Output:

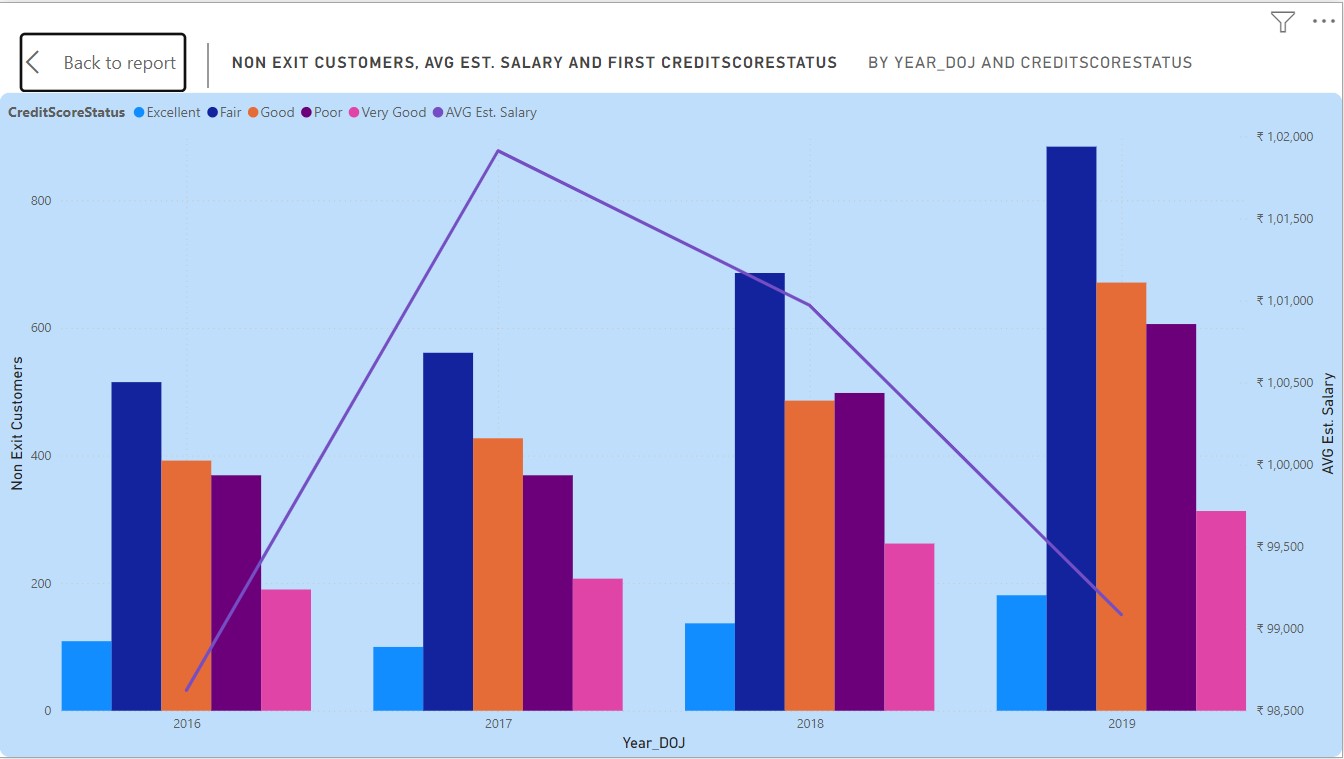
A screenshot of a computer

Description automatically generated

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

Ans:

* Non-Exit Customers by Credit Score: Customers with excellent credit scores consistently have the highest retention, while those with poor scores show lower retention.
* Average Salary Trend: Average estimated salary peaks in 2017 and declines thereafter, suggesting a possible correlation with customer retention over time.



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

ANS:

1.Activecustomers

2.Bank\_churn

3.CreditCard

4.CustomerInfo

5.ExitCustomer

6.Gender

7.Geography

Out of these tables Bank\_Churn and CustomerInfo are having the categorical Variables Like creditId, ActiveID, ExitID, GenderID and Geographical ID.

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)

WITH avg\_income AS

(SELECT gg.GeographyLocation,g.GenderCategory,ROUND(AVG(c.EstimatedSalary),2) AS average\_income

FROM customerinfo c

JOIN gender g USING (GenderID)

JOIN geography gg USING (GeographyID)

GROUP BY gg.GeographyLocation,g.GenderCategory

order by gg.GeographyLocation,g.GenderCategory)

SELECT \*,RANK() OVER(PARTITION BY GenderCategory ORDER BY average\_income DESC) AS rn

FROM avg\_income

Output:

A screenshot of a computer

Description automatically generated

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+).

SELECT CASE WHEN c.age BETWEEN 18 and 30 THEN '18-30'

WHEN c.age BETWEEN 30 AND 50 THEN '30-50'

ELSE '50+'

END AS age\_brackets,

AVG(bc.Tenure) AS avg\_tenure

FROM customerinfo c

JOIN bank\_churn bc USING (CustomerID)

WHERE bc.Exited = 1

GROUP BY age\_brackets

Output:

A screenshot of a computer

Description automatically generated

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?

Ans:

* The total no of retain customers are approximately 8K and exit customers are approximately 2k
* And the average estimated salary of exit customers is higher than retain customers.

A graph showing a bar and a rectangle

Description automatically generated

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

Ans: Customers having Fair credit score having more estimated salary 330 M of total salary Good and poor credit score customers having 244 M and 237 M total salary.

Who are having Top credit score those are very few.A graph of blue rectangular objects

Description automatically generated with medium confidence

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

Ans:

* The number of active customers appears to be increasing over the years. There were 991 active customers in 2016 and 1722 in 2019.
* This could indicate growth in customer acquisition for the company.

A graph with blue rectangles

Description automatically generated

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

ANS:

1.Activecustomers

2.Bank\_churn

3.CreditCard

4.CustomerInfo

5.ExitCustomer

6.Gender

7.Geography

Out of these tables Bank\_Churn having the categorical Variables.

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

WITH avg\_income\_location\_gender AS

(SELECT ci.GeographyID,

g.GenderCategory,

ROUND(AVG(ci.EstimatedSalary),2) As avg\_income

FROM customerinfo ci

JOIN gender g USING(GenderID)

GROUP BY ci.GeographyID,

g.GenderCategory

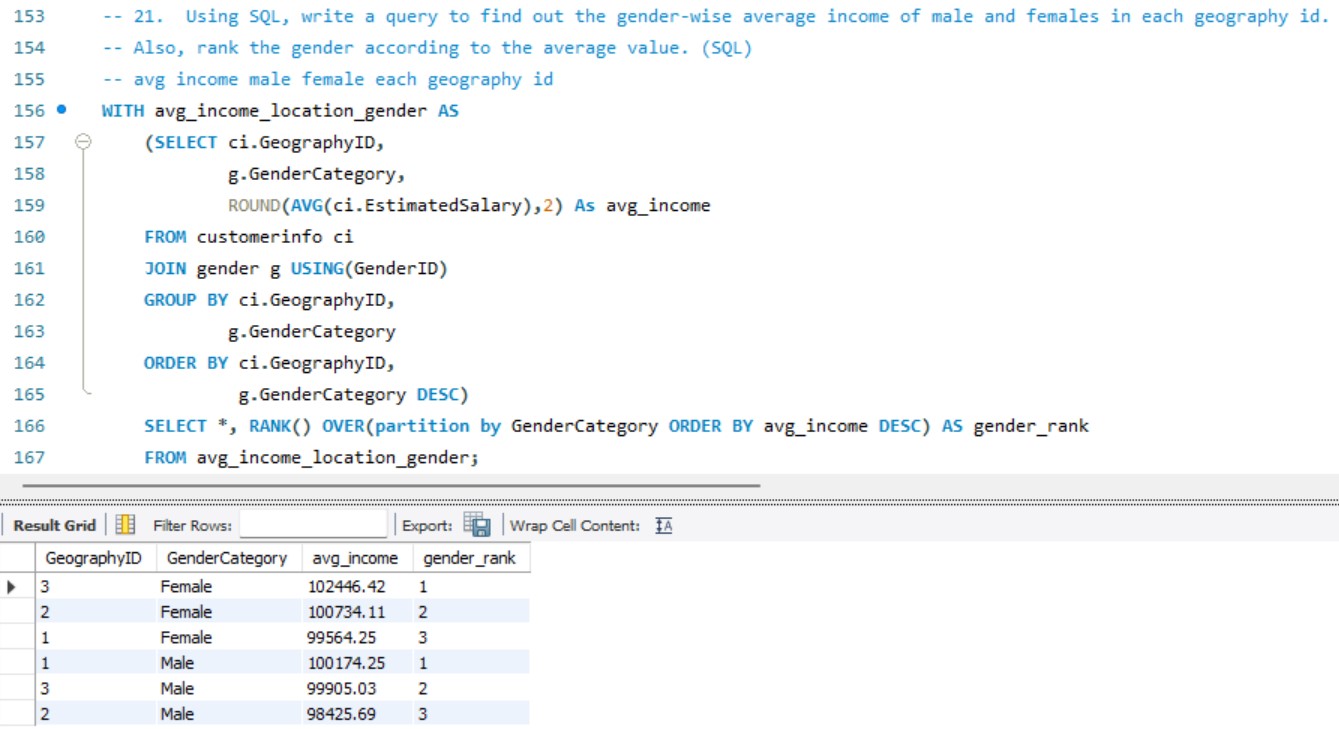
ORDER BY ci.GeographyID,

g.GenderCategory DESC)

SELECT \*, RANK() OVER(partition by GenderCategory ORDER BY avg\_income DESC) AS gender\_rank

FROM avg\_income\_location\_gender;

Output:



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+).

ANS:

With exit\_cust\_tenure As

(Select c.CustomerId,

c.Surname,

datediff(curdate(),c.BankDOJ) AS tenure\_years,

CASE WHEN c.Age> 50 THEN "50+"

WHEN c.Age BETWEEN 30 AND 50 THEN "30-50"

ELSE "18-30" END AS age\_brackets

FROM customerinfo c

JOIN bank\_churn ch On c.CustomerId = ch.CustomerId

-- JOIN exitcustomer e ON e.ExitID = ch.Exited

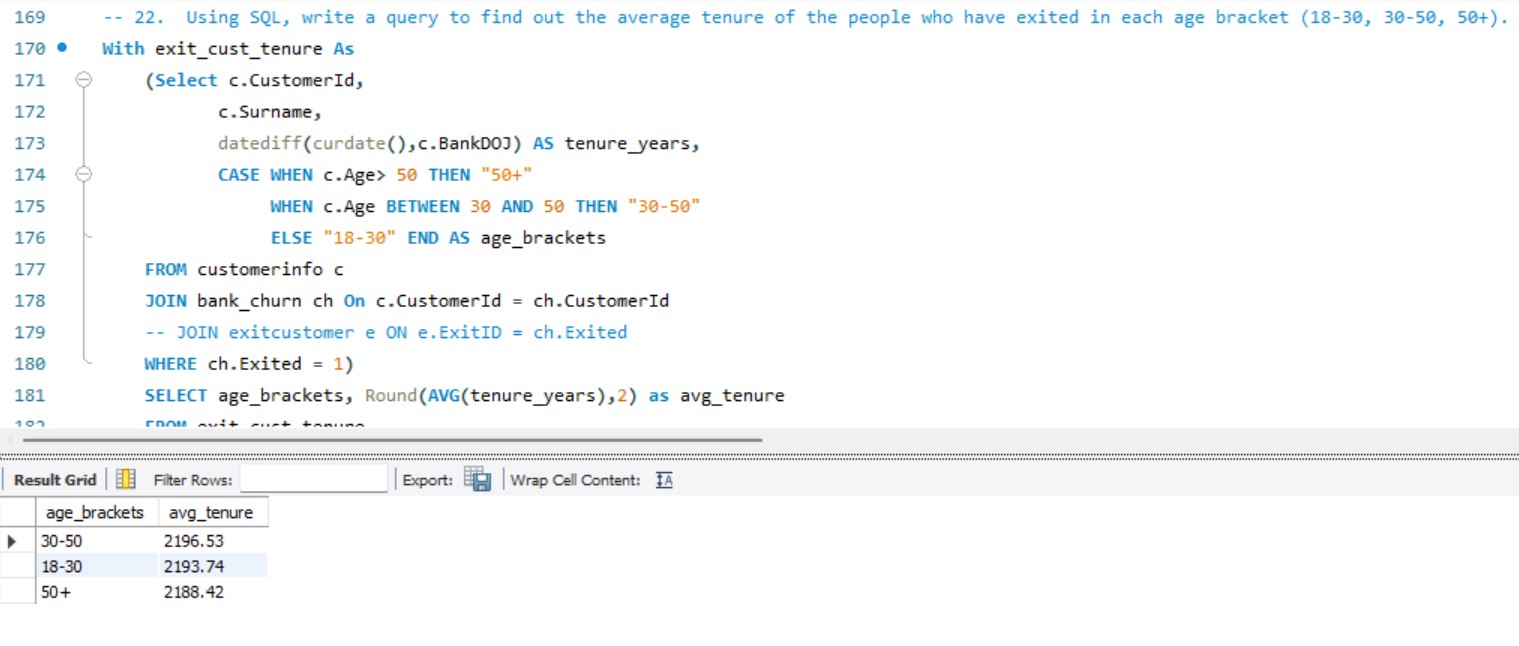
WHERE ch.Exited = 1)

SELECT age\_brackets, Round(AVG(tenure\_years),2) as avg\_tenure

FROM exit\_cust\_tenure

GROUP BY age\_brackets;

Output:



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

Ans:

* The total no of retain customers are approximately 8K and exit customers are approximately 2k
* And the average estimated salary of exit customers is higher than retain customers.

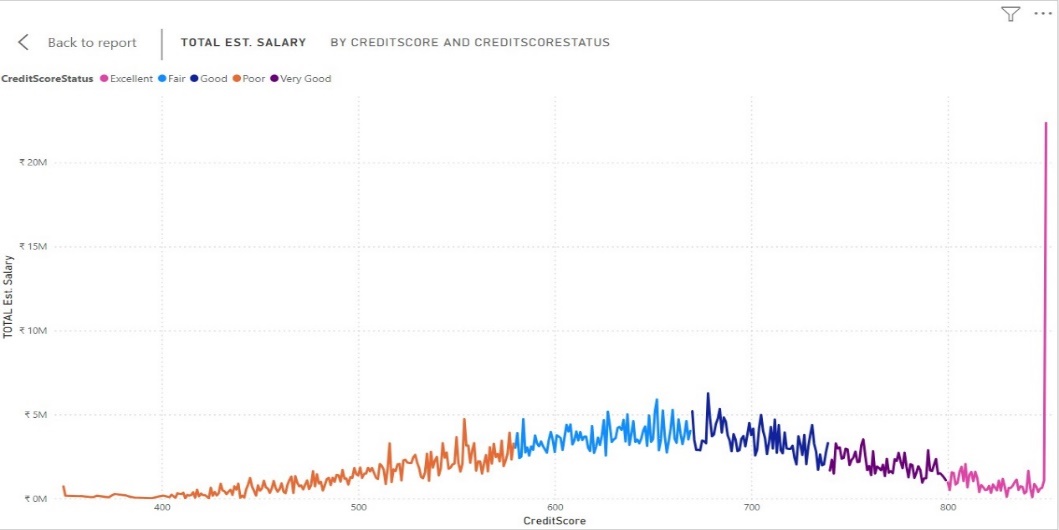
A graph showing a bar and a rectangle

Description automatically generated

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

Ans:

 People with credit scores between 550 and 750 tend to have high salaries, ranging from 2.5 million to 6 million rupees.

 The data suggests a positive correlation between credit score and salary, with the highest salary (22 million rupees) being associated with a very high credit score (nearly 850).

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”.

Ans:

SELECT DISTINCT b.CustomerID,c.Surname As last\_name, a.ActiveCategory

FROM customerinfo c

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

JOIN activecustomer a ON b.IsActiveMember = a.ActiveID

WHERE c.Surname LIKE "%on" AND a.ActiveCategory IS NOT NULL

Output:

A screenshot of a computer

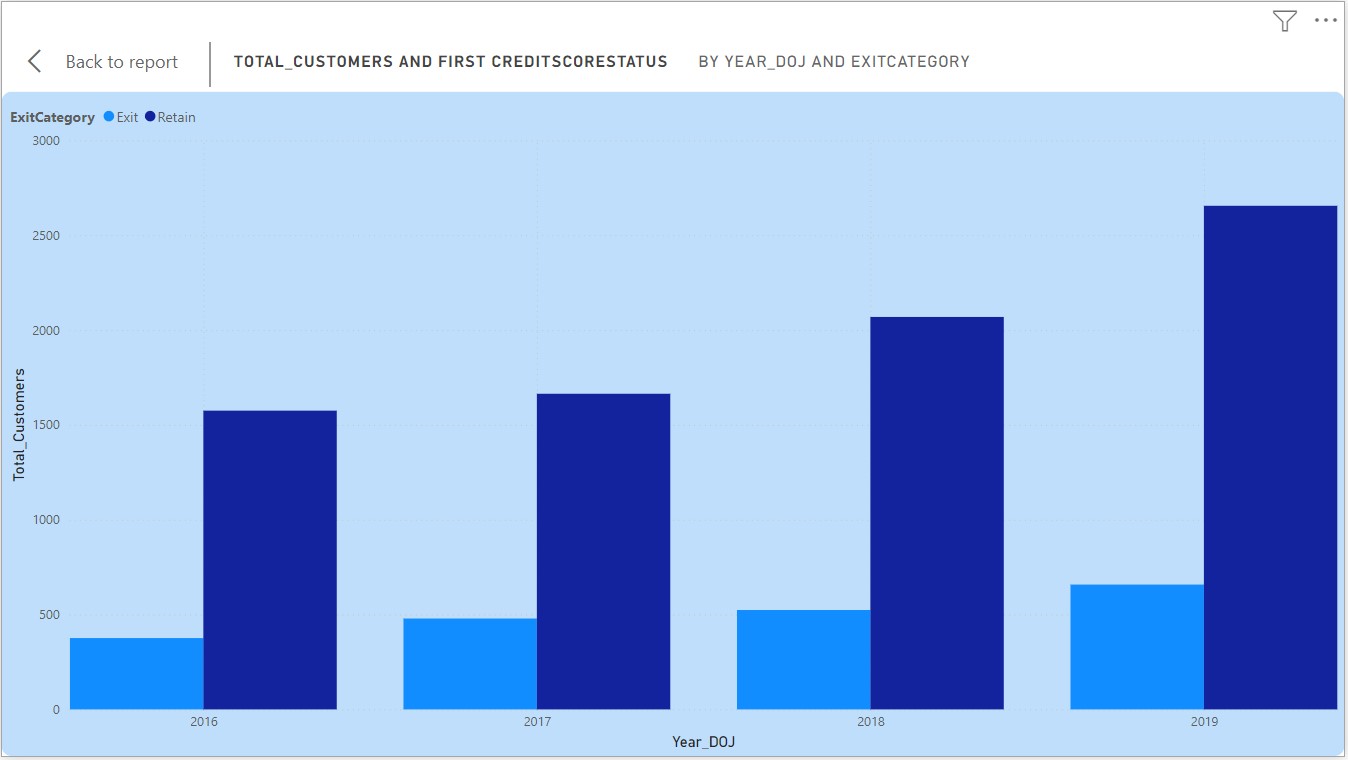
Description automatically generated

**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?

Ans:

* Over the years from 2016 to 2019, the total number of customers retained consistently increased, indicating growing customer loyalty.
* The number of customers exiting also rose annually, suggesting a possible parallel increase in the customer base and exit rate.



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

Ans:

Prepare the Database: Ensure that the customer data is loaded into a relational database.

Data Exploration: Use SQL queries to explore and understand the data.

Product Affinity Analysis: Use SQL to find common product combinations and generate insights for cross-selling strategies.

-- Distribution of Number of Products

SELECT NumOfProducts, COUNT(\*) AS Count

FROM customers

GROUP BY NumOfProducts

ORDER BY NumOfProducts;

-- Distribution of Credit Card Holding Status

SELECT HasCrCard, COUNT(\*) AS Count

FROM customers

GROUP BY HasCrCard;

-- Distribution of Active Members

SELECT IsActiveMember, COUNT(\*) AS Count

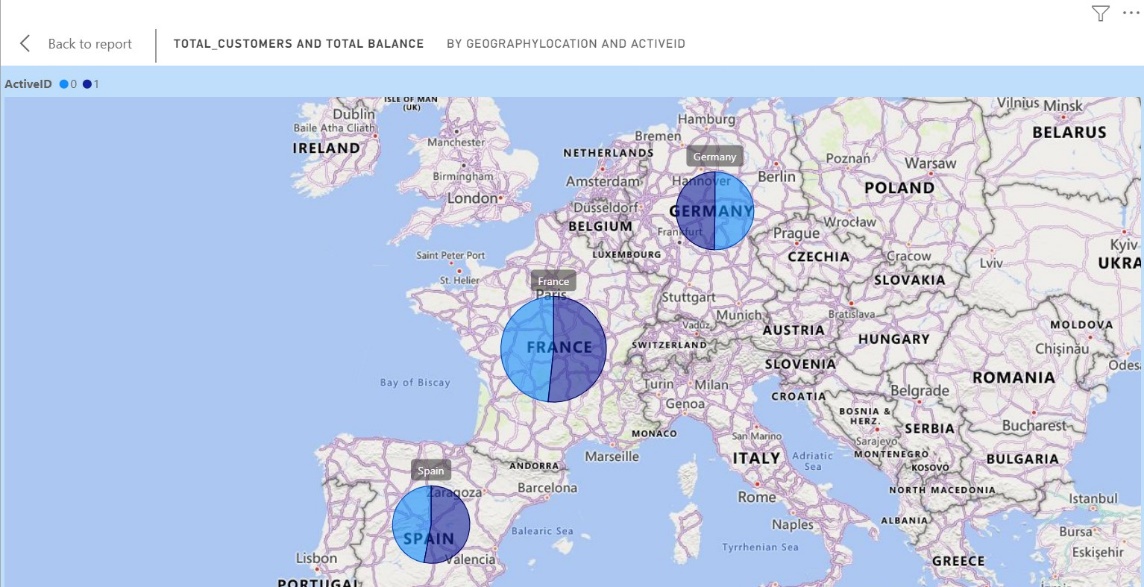
FROM customers

GROUP BY IsActiveMember;

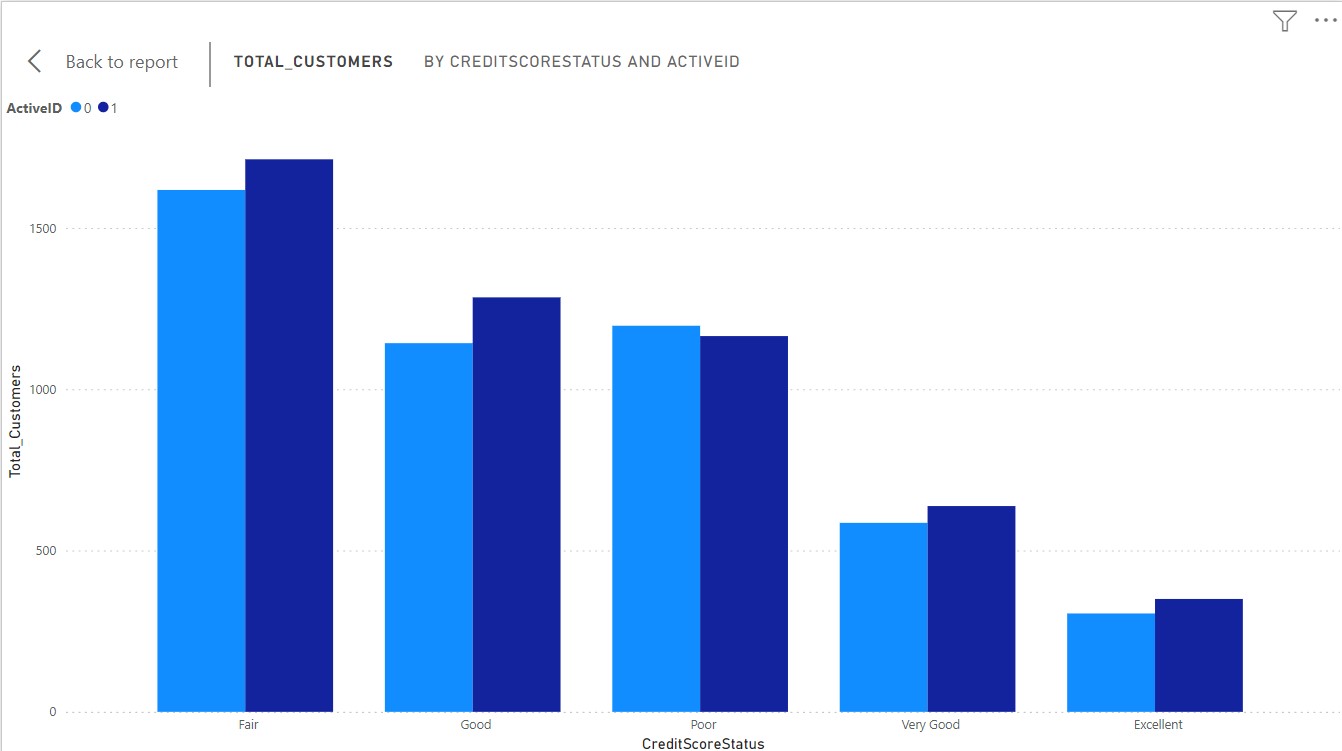
1. Geographic Market Trends: How do economic indicators in different geographic regions correlate with the number of active accounts and customer churn rates?

Ans:

* The table shows the total number of active members in France, Spain and Germany. France has the most active members (2591).
* Active members are categorized by their total balance and whether they are active or inactive.

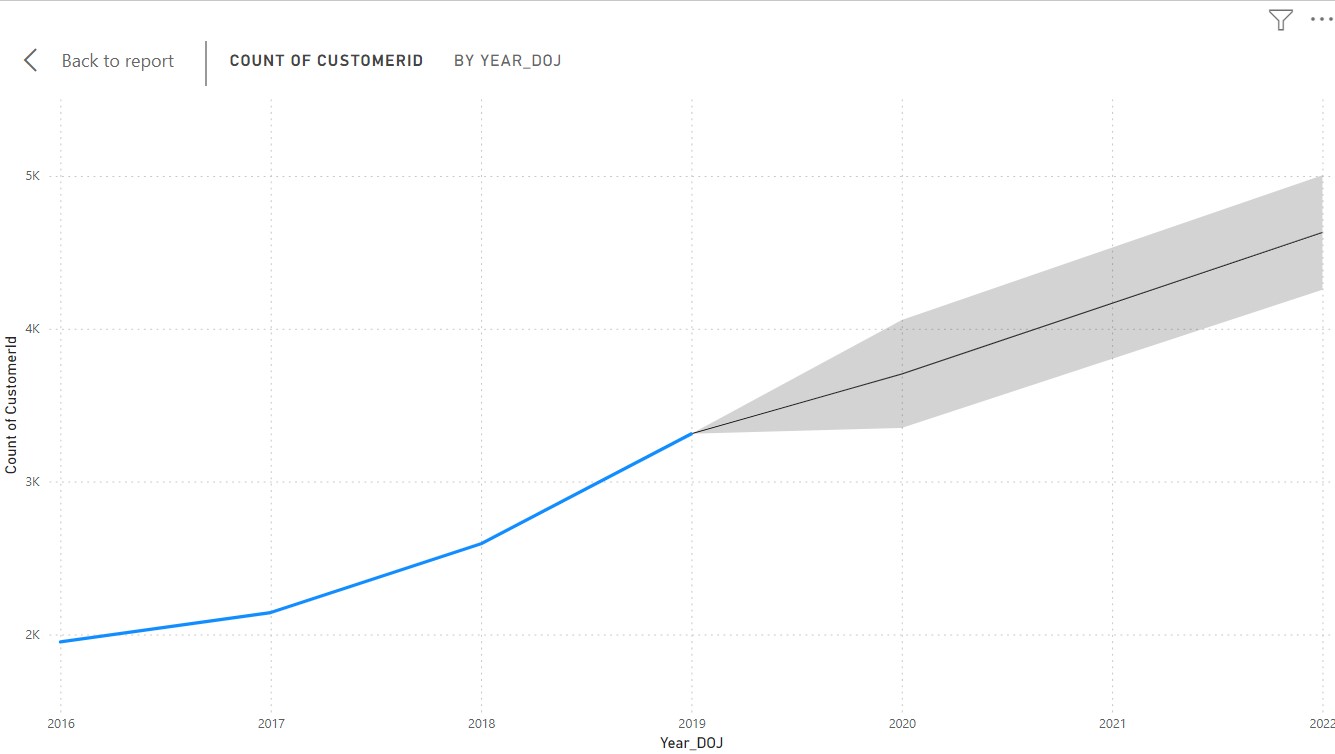


1. Risk Management Assessment: Based on customer profiles, which demographic segments appear to pose the highest financial risk to the bank, and why?



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?

Ans:

* The graph shows the number of customers who joined (DOJ) by year, according to a report. The x-axis is labelled "Year DOJ”, and the y-axis is labelled "Count of Customer id."
* This is the forecasting and prediction of customers for next few years.

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?

Ans:

**Compare Rates Pre vs. Post:** Track customer acquisition and retention rates before, during, and after a campaign to see if there's a lift.

**Segment by Campaign:** Analyze changes in customer behavior by segmenting them based on the marketing campaign they were exposed to.

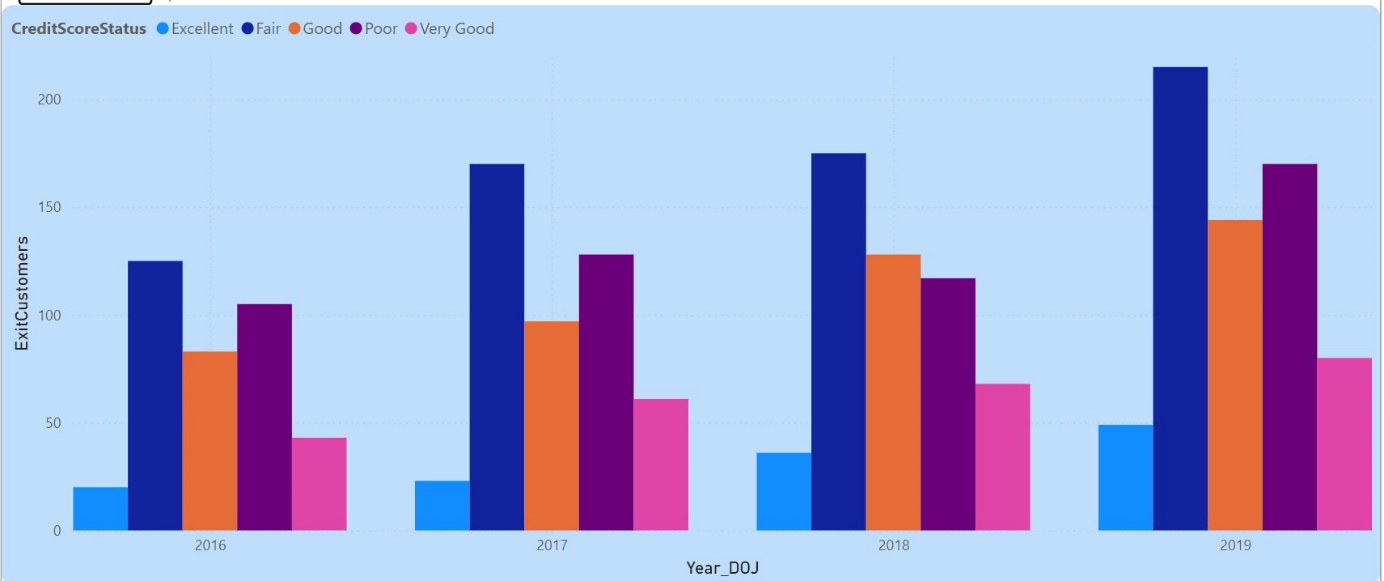
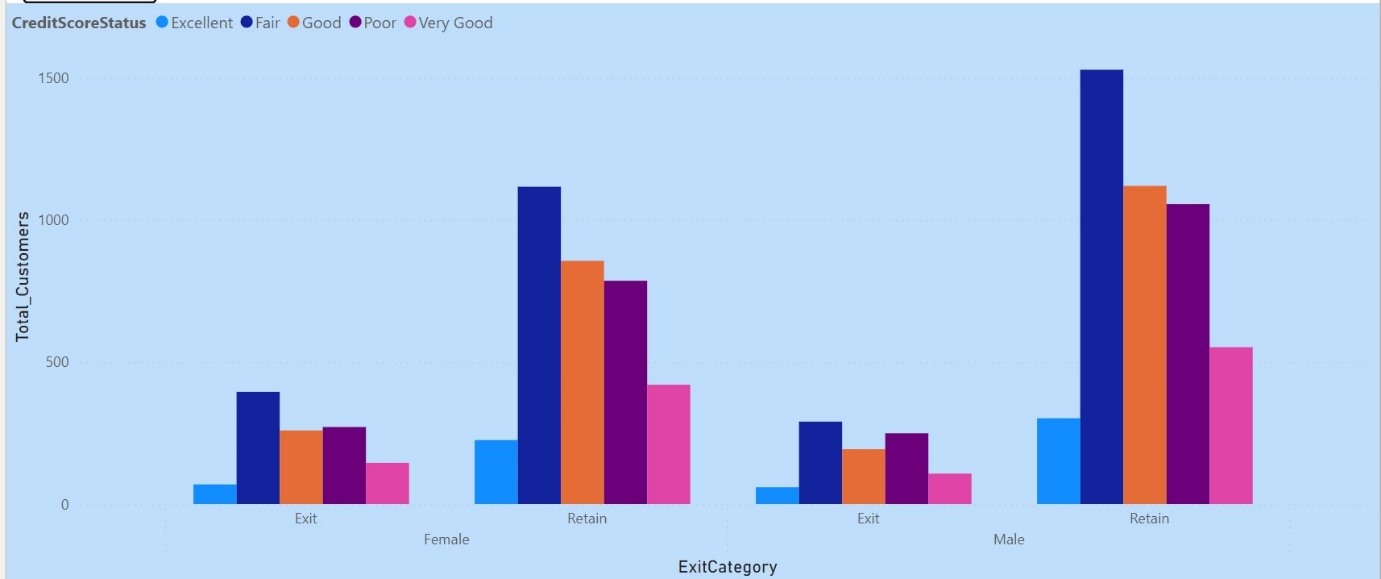
**Extra Information Needed:**

**Campaign Details:** Dates the campaign ran, target audience, and messaging used are crucial for understanding its reach and impact.

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

Ans:

* Exit category likely refers to whether a customer is leaving or staying with the company. Overall, there are more females than males across all credit score statuses and exit categories.
* Within the female category, there are more customers who are exiting with a fair credit score (395) than any other status. For males exiting, there are also more with a fair score (290) than any other status.



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

Ans:

**Longer Tenure Predicts Loyalty:** Customers who've been with the bank longer tend to be more satisfied and less likely to switch.

**Product Variety Reduces Churn:** Using multiple bank products (accounts, loans, investments) increases customer investment, reducing churn risk.

**Active Users Stay Engaged:** Customers who regularly use the bank's services are likely more content and less likely to leave.

**Estimated Salary Impacts Sensitivity:** Customers with higher estimated salaries might be more sensitive to fees and rates, potentially leading to churn if they find better deals elsewhere.

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

Ans:

To segment customers based on demographics and account details using SQL, we can create various segments by grouping customers on key attributes like age, geography, gender, credit score, tenure, balance, and product usage. Here are several SQL queries that will help you segment the customers based on these criteria:

SELECT CASE WHEN CreditScore < 600 THEN 'Poor(Less Than 600)'

WHEN CreditScore >= 600 AND CreditScore < 700 THEN 'Fair(Between 600 And 700)'

WHEN CreditScore >= 700 AND CreditScore < 800 THEN 'Good(Between 700 And 800)'

ELSE 'Excellent(More than 800)'

END AS segments, Count(Exited) As cnt\_exited

FROM bank\_churn

WHERE Exited = 1

GROUP BY segments

ORDER By cnt\_exited DESC;

WITH avg\_income AS

(SELECT gg.GeographyLocation,g.GenderCategory,ROUND(AVG(c.EstimatedSalary),2) AS average\_income

FROM customerinfo c

JOIN gender g USING (GenderID)

JOIN geography gg USING (GeographyID)

GROUP BY gg.GeographyLocation,g.GenderCategory

order by gg.GeographyLocation,g.GenderCategory)

SELECT \*,RANK() OVER(PARTITION BY GenderCategory ORDER BY average\_income DESC) AS rn

FROM avg\_income

SELECT CASE WHEN c.age BETWEEN 18 and 30 THEN '18-30'

WHEN c.age BETWEEN 30 AND 50 THEN '30-50'

ELSE '50+'

END AS age\_brackets,

AVG(bc.Tenure) AS avg\_tenure

FROM customerinfo c

JOIN bank\_churn bc USING (CustomerID)

WHERE bc.Exited = 1

GROUP BY age\_brackets

ORDER BY age\_brackets;

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?

Ans:

**Highlighting At-Risk Customers:**

1. **Identify Churn Risk Factors:** Analyze your data to determine factors that predict customer churn (e.g., low account balance, inactivity, missed payments).
2. **Define Conditional Formatting Rules:** Set up rules based on these factors. For example, highlight customers with a balance below $500 or those inactive for the past 6 months.
3. **Choose Visual Cues:** Use colors (red for high risk, yellow for medium) or symbols to visually identify at-risk customers.

**Evaluating Reward Program Impact:**

1. **Segment Customers:** Separate customers who use your credit card rewards program from those who don't.
2. **Compare Churn Rates:** Calculate customer churn rates (percentage who leave) for each segment over a set period.
3. **Conditional Formatting:** Apply formatting to highlight the churn rate differential. A significant difference between segments (lower churn rate for rewards users) suggests a positive program impact.

**Software Considerations:**

* The specific steps to implement these techniques will depend on your spreadsheet software (Excel, Google Sheets).
* Most offer conditional formatting options where you can define rules and visual cues.
* Pivot tables can help calculate churn rates by segmenting customer data.

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?

Ans:

* Customer Growth Trend: France consistently leads in customer numbers from 2016 to 2019, showing steady growth each year.
* Geographical Comparison: Germany and Spain have similar customer numbers each year, with Germany slightly ahead except in 2016.
* Annual Increase: Each geography shows a clear increase in customer numbers from 2016 to 2019, indicating overall market expansion.

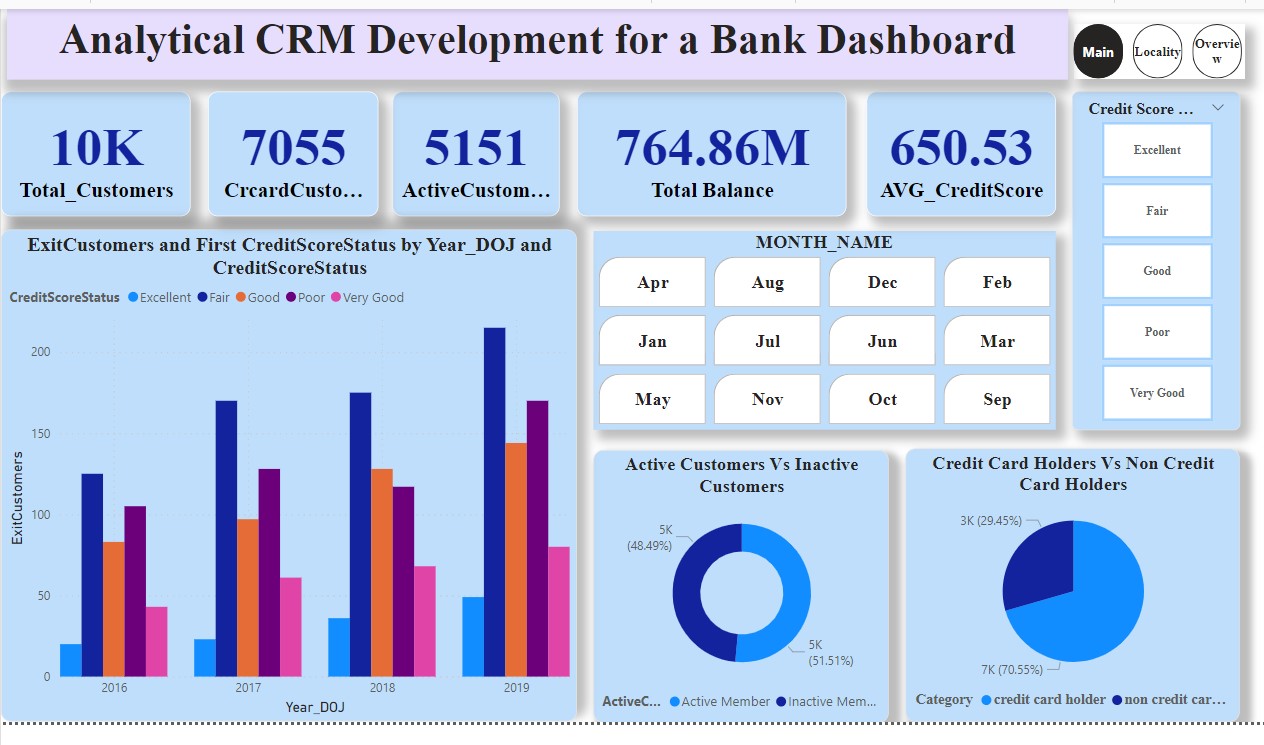
A graph showing different colored squares

Description automatically generated

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

Ans:

Here Creditscore and month\_name are the slicers used in 1st page. Year of Join and Geography Location slices are added in 2nd page.



A screenshot of a computer

Description automatically generated

A close-up of a graph

Description automatically generated

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

Ans:

**Data Analysis and Segmentation:** Analyse customer data to understand their behaviour, needs, and churn factors. Segment customers based on demographics, purchase history, and engagement levels to personalize marketing efforts.

**Targeted Interactions and Automation:** Design targeted marketing campaigns and outreach strategies for each customer segment.

**Performance Measurement and Improvement:** Track key metrics like customer acquisition cost, lifetime value, and churn rate. Analyze results to identify areas for improvement and iterate on CRM strategies for better outcomes.

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

Ans:

By using the SQL query rename the column like this.

ALTER TABLE Bank\_Churn

RENAME COLUMN HasCrCard TO Has\_creditcard;