**Analytical CRM Development for a Bank**

**Objective Question:**

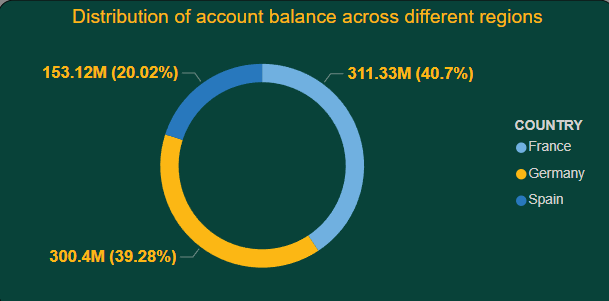
1. What is the distribution of account balance across different regions? (Power BI)

**Aggregation by Region**: The query groups account balances by geographical location.

**Sum and Round Balances**: It calculates the total balance for each region, rounding for clarity.

**Identify Regional Trends**: This helps in understanding how account balances are distributed across different regions.

**Informed Decision-Making**: Enables data-driven decisions based on regional financial data.



* **Regional Wealth Distribution**: Identifying which regions have higher or lower total account balances.
* **Target Market Identification**: Recognizing regions with significant balances to focus marketing and sales efforts.
* **Resource Allocation**: Allocating resources and services to regions based on their financial activity.
* **Regional Performance Analysis**: Comparing financial performance across different regions to identify growth or decline.
* **Strategic Planning**: Developing strategies for underperforming regions to boost their financial activity.

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

This query retrieves the top 5 customers with the highest Estimated Salary whose dates fall within the last quarter of every year. It orders the results by Date\_of\_joining in descending order, so the CustomerId, Surname, Salary comes in the order of month and returns the top five customers of the last quarter of every year.

**SQL CODE :**

SELECT \*

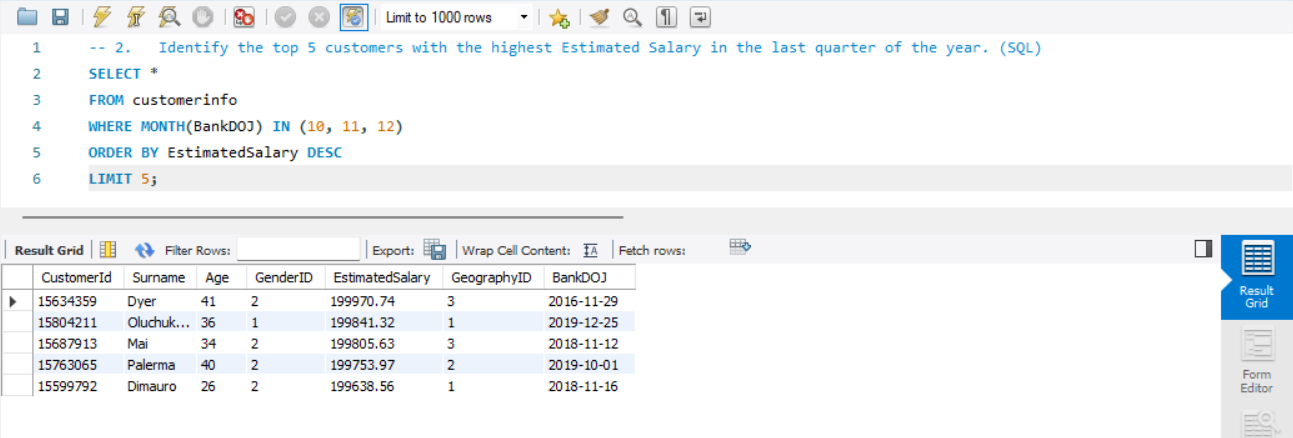
FROM customerinfo

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

ORDER BY EstimatedSalary DESC

LIMIT 5 ;

**Output:**



|  |  |  |  |
| --- | --- | --- | --- |
| CustomerID | Surname | EstimatedSalary | Date\_of\_joining |
| 15610165 | Hsiung | 199409.2 | 10-10-2016 |
| 15767722 | Richardson | 197933.5 | 22-10-2016 |
| 15671591 | Castiglione | 198874.5 | 21-11-2016 |
| 15634359 | Dyer | 199970.7 | 29-11-2016 |
| 15737748 | McWilliam | 199336.6 | 21-12-2016 |
| 15598275 | Sochima | 199418 | 05-10-2017 |
| 15710687 | Mills | 199478.1 | 15-11-2017 |
| 15800620 | Fitzgerald | 199635.9 | 19-11-2017 |
| 15770405 | Warlow-Davies | 199104.5 | 27-11-2017 |
| 15711316 | Ch'ang | 199527.3 | 11-12-2017 |
| 15572361 | Chill | 199420.4 | 02-10-2018 |
| 15687913 | Mai | 199805.6 | 12-11-2018 |
| 15599792 | Dimauro | 199638.6 | 16-11-2018 |
| 15585961 | Talbot | 199505.5 | 21-12-2018 |
| 15661903 | Hsia | 199378.6 | 29-12-2018 |
| 15763065 | Palerma | 199754 | 01-10-2019 |
| 15624641 | Kharlamova | 199290.7 | 02-10-2019 |
| 15699095 | Chandler | 198826 | 15-12-2019 |
| 15691874 | Kazakova | 198929.8 | 19-12-2019 |
| 15804211 | Oluchukwu | 199841.3 | 25-12-2019 |

* **High Value Customers:** Identify the top earners among new customers joining in the last quarter.
* **Customer Profile:** Understand the profile of high-value customers joining towards the end of the year.
* **Targeted Marketing:** Develop marketing strategies to attract similar high-salary customers in the future.
* **Resource Allocation:** Allocate more resources and personalized services to retain these high-value customers.
* **Performance Evaluation:** Assess the bank's success in attracting affluent customers during the last quarter.

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

Query selects the average value of the NumberOfProducts column for customers who have a credit card (HasCrCard = 'credit card holder'). It calculates the average number of products used by these customers.

The output for the average number of products is **1.5341**.

|  |
| --- |
| Average\_NumberofProduct |
| 1.5314 |

**SQL CODE:**

SELECT ROUND(AVG(NumOfProducts), 0) AS avg\_products

FROM bank\_churn b

LEFT JOIN

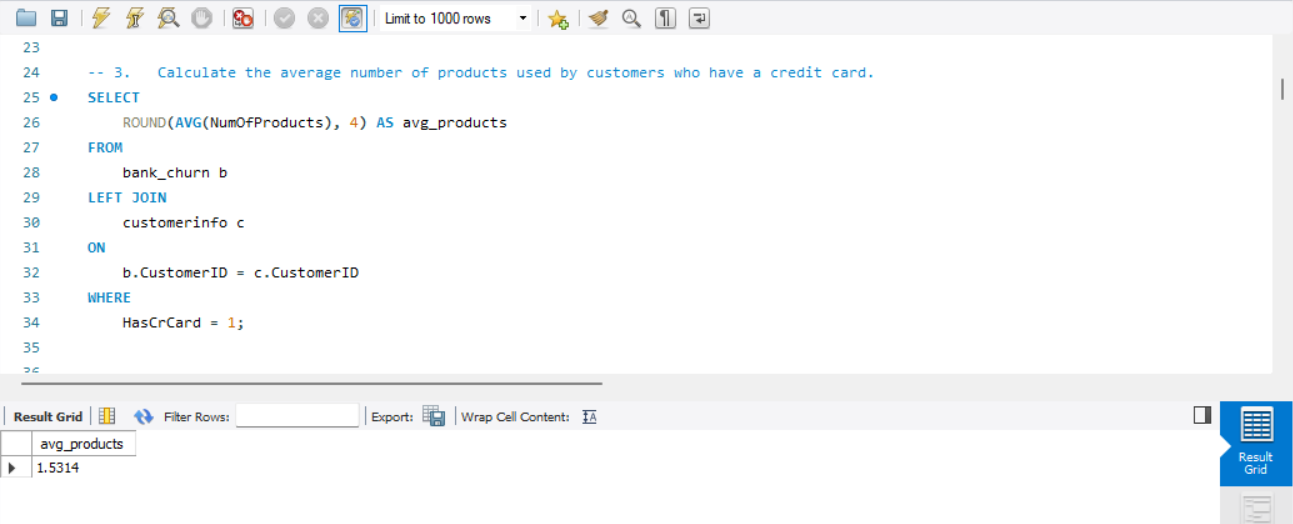
customerinfo c

ON

b.CustomerID = c.CustomerID

WHERE

HasCrCard = 1;

**Output:**

**Insights:**

* **Product Usage Trends**: Understand how many products customers with credit cards typically use.
* **Cross-Selling Opportunities**: Identify opportunities to cross-sell additional products to credit card holders.
* **Customer Engagement**: Measure engagement levels of customers with credit cards based on their product usage.
* **Service Improvement**: Tailor services and offers to encourage higher product adoption among credit card holders.
* **Retention Strategies**: Develop retention strategies focused on customers with multiple products, enhancing their loyalty to the bank.

1. Determine the churn rate by gender for the most recent year in the dataset. (Power BI)

* 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.

**Churn Rate of Male =**

**DIVIDE (**

**CALCULATE (COUNTROWS ('try bank'), 'try bank'[Exited] = "Exit"),**

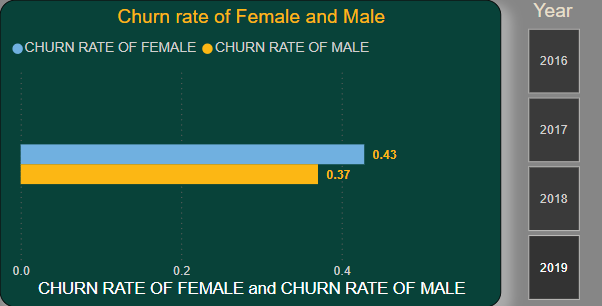
**CALCULATE (COUNTROWS ('try customer info'), 'try customer info'[Gender] = "Male”),0)**

**The Same Dax function also for Female.**

Visualize the Data:

Drag the "Churn Rate by Gender" measure into the Field section.

Add Slicer to get the recent year Churn rate. In 2019 the Male Churn Rate was **2.27** and for Female was **1.79.**

****

**Insights :**

* The higher churn rate among male customers suggests that male customers might be more likely to leave the service compared to female customers.
* Bank needs customer retention strategies differently for male and female customers, focusing more on the factors that might lead to higher churn among male customers.

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

**Approach:**

* **Categorize Customers**: The query uses a CASE statement to categorize customers into 'Exited' and 'Remain' based on the Exited column.
* **Calculate Average Credit Score**: It calculates the average credit score for each category.

|  |  |
| --- | --- |
| Customer\_Status | Average\_Cr\_Score |
| Exited | 645.3515 |
| Retain | 651.8532 |

* **Group by Category**: Groups the results by the 'Exited' and 'Remain' categories to compare the average credit scores.

**SQL CODE :**

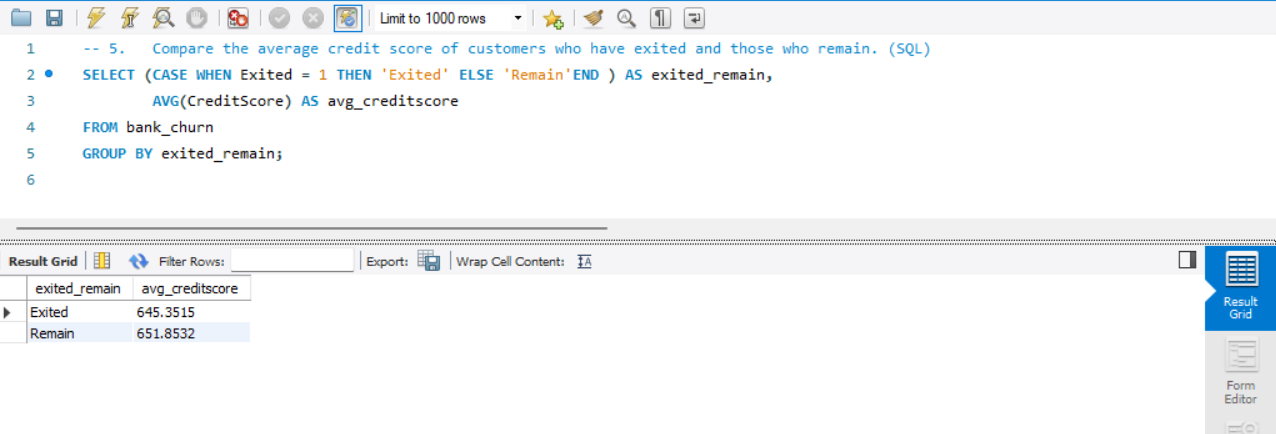
SELECT (CASE WHEN Exited = 1 THEN 'Exited' ELSE 'Remain'END ) AS exited\_remain,

AVG(CreditScore) AS avg\_creditscore

FROM bank\_churn

GROUP BY exited\_remain;

**Output:**



**Insights:**

* **Credit Score Trends**: Understand the average credit scores of customers who have exited compared to those who remain.
* **Risk Management**: Assess the risk profiles of customers more likely to leave, aiding in the development of retention strategies.
* **Customer Segmentation**: Segment customers based on credit score and exit behavior to tailor financial products and services.
* **Service Improvement**: Improve services and offerings for customers with lower credit scores who are more likely to exit.
* **Predictive Analysis**: Use the data to predict and prevent potential exits by identifying at-risk customers based on their credit scores.

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

Females have a higher average estimated salary. If you want to write an analysis of this question, then you must add conditions on active accounts because in the data set, we have data of inactive members also.

The query calculates the average estimated salary for each gender group and counts the number of active accounts for each gender group. It uses the SUM function along with a CASE statement to count the number of active accounts. The results are grouped by gender and order by Average\_salary and use LIMIT 1.

**SQL CODE :**

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

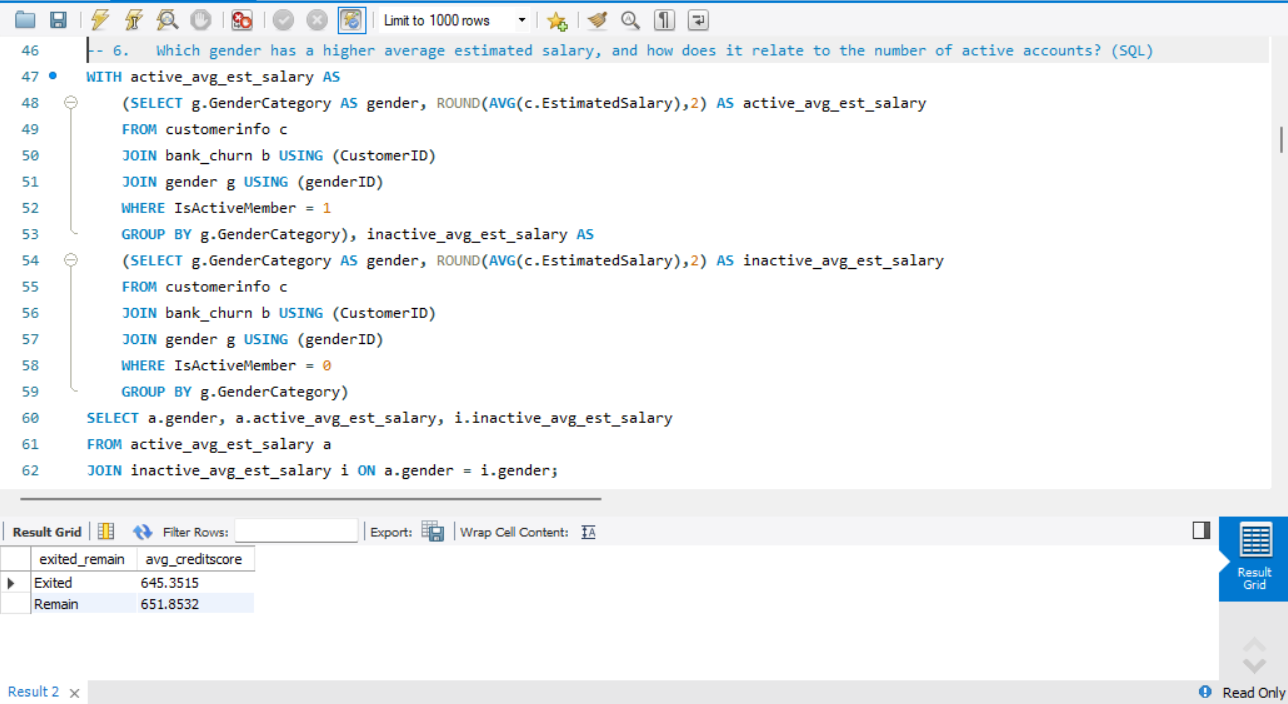
GROUP BY g.GenderCategory)

SELECT a.gender, a.active\_avg\_est\_salary, i.inactive\_avg\_est\_salary

FROM active\_avg\_est\_salary a

JOIN inactive\_avg\_est\_salary i ON a.gender = i.gender;

**Output:**



**Insights:**

* **Gender Salary Comparison**: Identify which gender has a higher average estimated salary.
* **Active vs. Inactive Members**: Compare the average estimated salary for active and inactive members within each gender.
* **Engagement Impact**: Understand how being an active member relates to the average estimated salary for each gender.
* **Gender-Based Analysis**: Provide insights into salary disparities between genders and their engagement with the bank.
* **Targeted Strategies**: Develop targeted strategies to enhance engagement and retention based on gender and estimated salary insights.

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

Query first categorizes customers into segments based on their credit scores. Then, it calculates the total number of customers, the number of exited customers, and the exit rate for each segment. Finally, it sorts the results by exit rate in descending order and selects the segment with the highest exit rate using the LIMIT clause.

As per analysis Poor Credit card Score Segment has the highest Exit Rate.

**SQL CODE :**

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

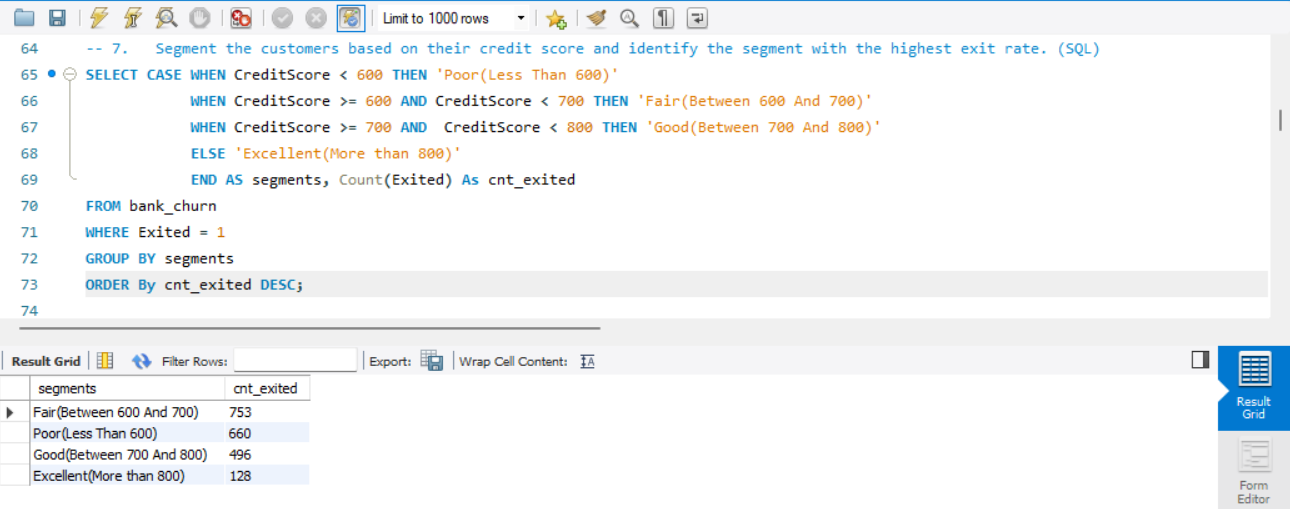
FROM bank\_churn

WHERE Exited = 1

GROUP BY segments

ORDER By cnt\_exited DESC;

**Output:**



**Insights:**

* **High-Risk Segments**: Identify which credit score segment has the highest number of customer exits.
* **Targeted Retention**: Focus retention efforts on the segment with the highest exit rate to reduce churn.
* **Credit Score Impact**: Understand how credit score relates to customer retention and identify potential risk factors.
* **Customer Segmentation**: Use segmentation to tailor marketing and service strategies for different credit score groups.
* **Data-Driven Decisions**: Make informed decisions on customer engagement based on credit score segments and exit rates.

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

Query filters the customers who are active and have a tenure greater than 5 years. Then, it groups the results by geographic region and counts the number of active customers in each region. And we need to join both Table Bank and CustomerInfo. Finally, it orders the results in descending order based on the count of active customers and selects the region with the highest count using the LIMIT clause.

**SQL CODE :**

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

FROM customerinfo c

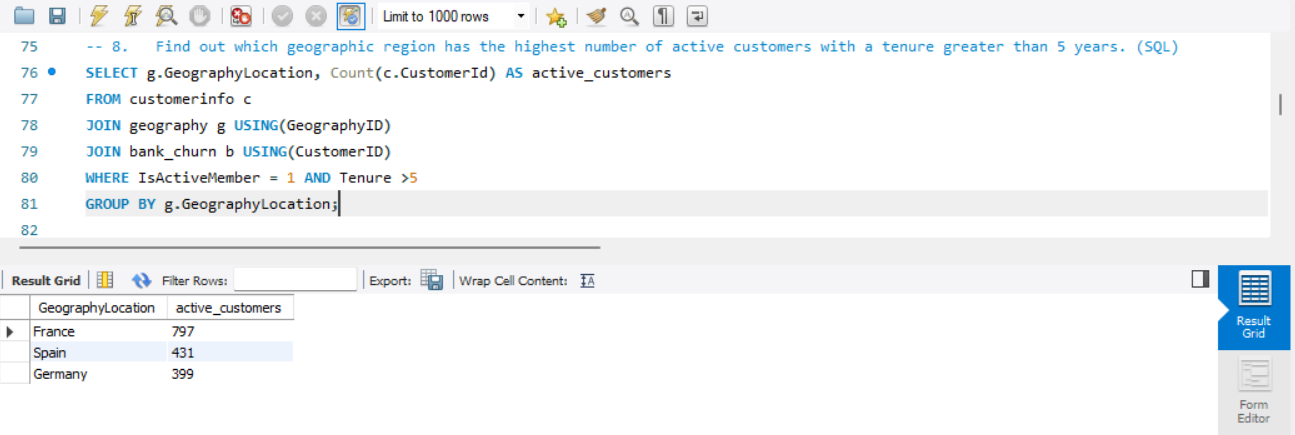
JOIN geography g USING(GeographyID)

JOIN bank\_churn b USING(CustomerID)

WHERE IsActiveMember = 1 AND Tenure >5

GROUP BY g.GeographyLocation;

**Output:**



|  |  |
| --- | --- |
| Geography | Active Member |
| France | 797 |
| Spain | 431 |
| Germany | 399 |

**Insights:**

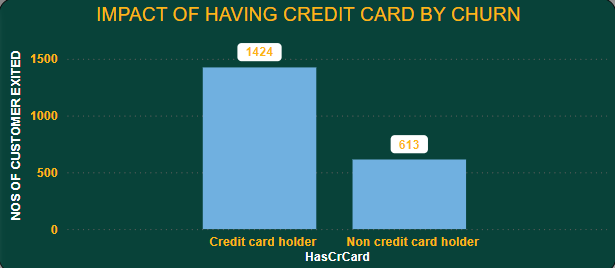
1. **Regional Loyalty:** Identify which geographic region has the highest number of long-term active customers, indicating strong customer loyalty in that region.
2. **Regional Engagement:** Assess the effectiveness of customer engagement strategies in different regions.
3. **Targeted Initiatives:** Develop targeted initiatives to enhance customer retention and engagement in regions with lower counts of long-term active customers.
4. **Resource Allocation:** Allocate resources and marketing efforts to regions with high numbers of loyal customers to further strengthen relationships.
5. **Performance Metrics:** Use the data as a performance metric to evaluate the success of customer retention strategies over time.
6. What is the impact of having a credit card on customer churn, based on the available data? (Power BI)

To determine the impact of having a credit card on customer churn based on the available data, you would typically analyze the churn rates for customers with and without a credit card.

Create a visualization, You can place the Count of Exited column on the Y-axis and drag the HasCrCard column on the X-axis to analyze trends.

Apply a filter to your dataset to include only data for customers who have exited (churned). The Impact was **1424** members exit those that have credit cards and **613** members exit those that have not credit card.

* 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 had used? (Power BI)

To determine the most common number of products used by customers who have exited (churned).

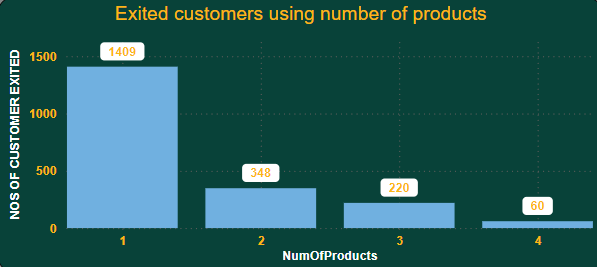
Visualize Data:

Create a visualization, such as a bar chart or a table, to display the counts of churned customers for each number of products.

You can use a bar chart with the number of products on the X-axis and the count of churned customers on the Y-axis.

Apply a filter to your dataset to include only data for customers who have exited (churned).

As per our analysis Product Number **1** was highly used by **1409** exited customers.



|  |  |
| --- | --- |
| 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.

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.

**SQL CODE :**

SELECT YEAR(BankDOJ) AS join\_year,

MONTHNAME(BankDOJ) AS join\_month,

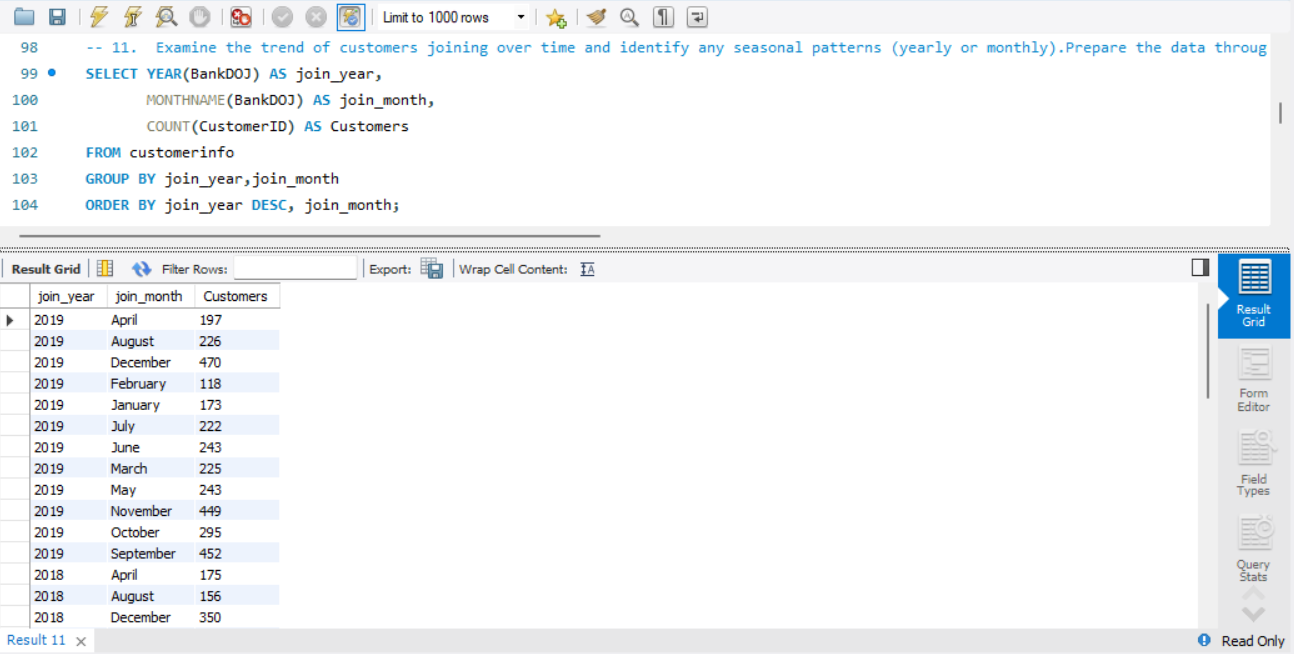
COUNT(CustomerID) AS Customers

FROM customerinfo

GROUP BY join\_year,join\_month

ORDER BY join\_year DESC, join\_month;

**Output:**



Visualize Data in Power BI:

Connect Power BI to your database and import the data using the SQL query you've written.

Once the data is imported into Power BI, create a chart visual.

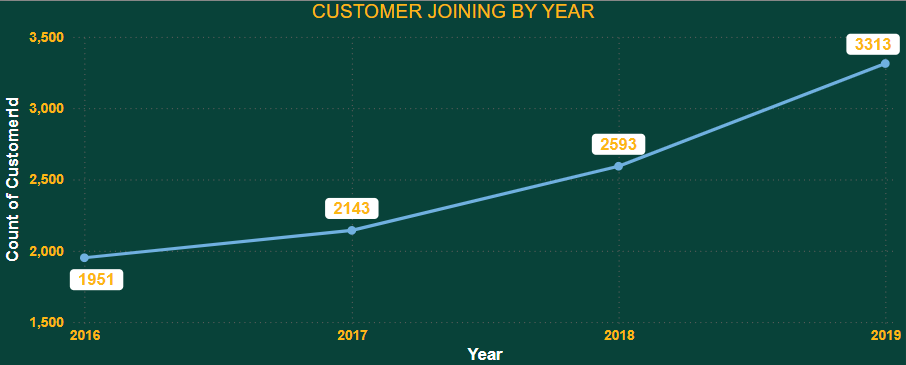
* Drag the "Year" field to the X-axis of the chart.
* Drag the "count\_customerId" field to the Values section of the chart.

Power BI should automatically recognize the date field and aggregate the data by year.

As per our analysis every year the joining of customers increased and in the recent (2019) year 3313 customers joined.

|  |  |
| --- | --- |
| Years | Customer\_Joined |
| 2019 | 3313 |
| 2018 | 2593 |
| 2017 | 2143 |
| 2016 | 1951 |





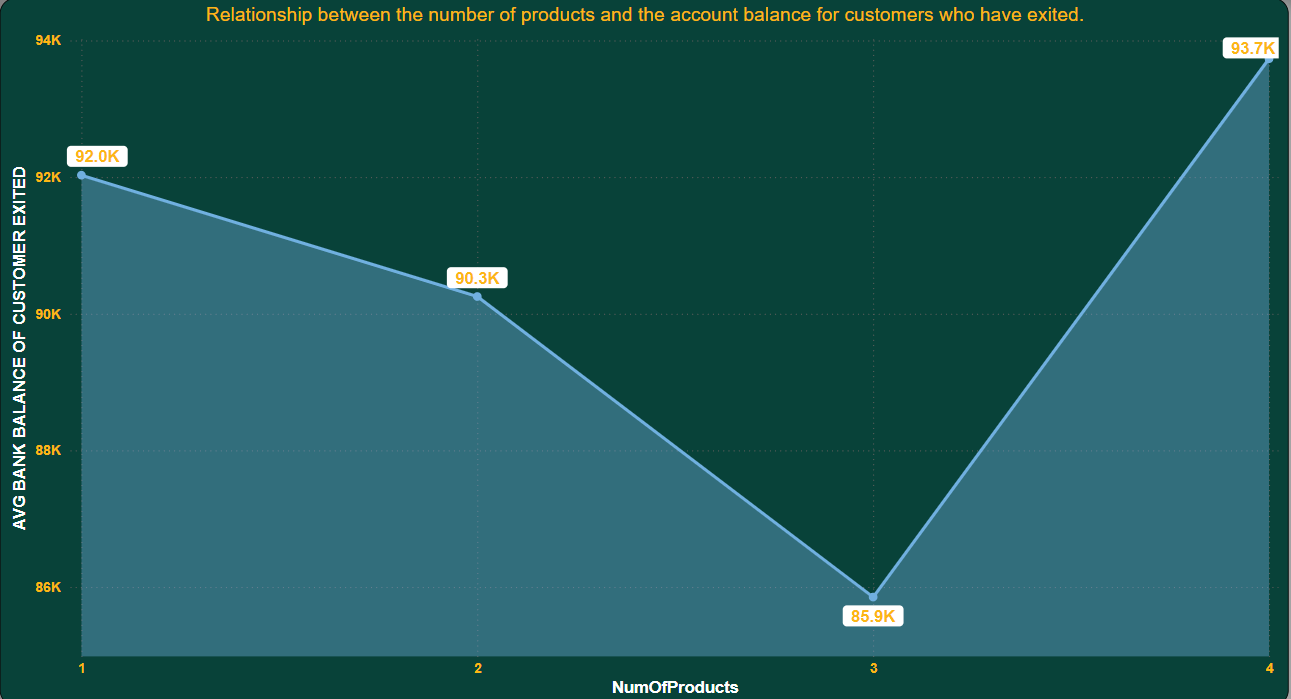
1. Analyze the relationship between the number of products and the account balance for customers who have exited. (Power BI)

**Data Preparation**: Clean the data by handling any missing or incorrect values to ensure accuracy in the analysis.

**Creating a Visual Representation**:

* In Power BI, create a scatter chart visual to represent the relationship between the number of products and the account balance for churned customers.
* Drag the NumOfProducts field to the X-axis, representing the number of products each churned customer uses.
* Drag the Balance field to the Y-axis, representing the account balance of each churned customer.

As per the analysis, the highest average balance was **93,733** of Product **4**.



* Customers with multiple products and high balances should be a focus for retention strategies.
* Understanding why these high-value customers are churning can help in developing targeted interventions to retain them.
* Implement retention programs focused on customers with high balances and multiple products to reduce churn rates.

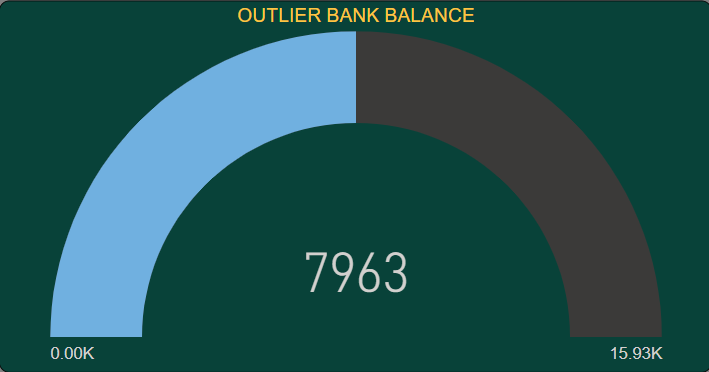
1. Identify any potential outliers in terms of balance among customers who have remained with the bank. (Power BI)

As per our analysis we find the InterQuartile Range, Outliers Range and Maximum balance. There were not any potential outliers because maximum balance(251K) was less than Outliers Range(319K).

Outliers Range = [Quartile Q3] +1.5\*[InterQuartile Range]

InterQuartile Range = [Quartile Q3]-[Quartile Q1]

* 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.



**Insights :**

* Customers with excellent credit scores consistently have the highest retention rates.
* Higher credit scores may indicate more financially stable customers who are less likely to churn.

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

There are 7 different tables given in the dataset but out of 7 only 2 datasets are consisting of categorical variables.

First dataset was “Bank\_Churn” and second was “CustomerInfo”.

In SQL, categorical variables are those that represent qualitative or categorical data, as opposed to quantitative or numerical data. Categorical variables are typically used to represent characteristics or attributes that fall into distinct categories or groups.

Categorical variables are often represented using string or character data types in SQL, and they play a crucial role in various data analysis tasks such as segmentation, grouping, and classification. When working with categorical variables in SQL, it's essential to understand how to encode, manipulate, and analyze them effectively to derive insights from the data.

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)

Query calculates the average income for each gender within each geography ID. It then ranks the genders within each geography ID based on the average income, with the highest average income ranked first. The RANK () function is used with the PARTITION BY clause to partition the ranking by geography ID.

|  |  |  |  |
| --- | --- | --- | --- |
| Geography | Gender | Average Salary | Rank |
| France | Male | 100174.3 | 1 |
| France | Female | 99564.25 | 2 |
| Germany | Female | 102446.4 | 1 |
| Germany | Male | 99905.03 | 2 |
| Spain | Female | 100734.1 | 1 |
| Spain | Male | 98425.69 | 2 |

**SQL CODE :**

WITH avg\_income AS

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

FROM customerinfo c

JOIN gender g USING (GenderID)

JOIN geography gg USING (GeographyID)

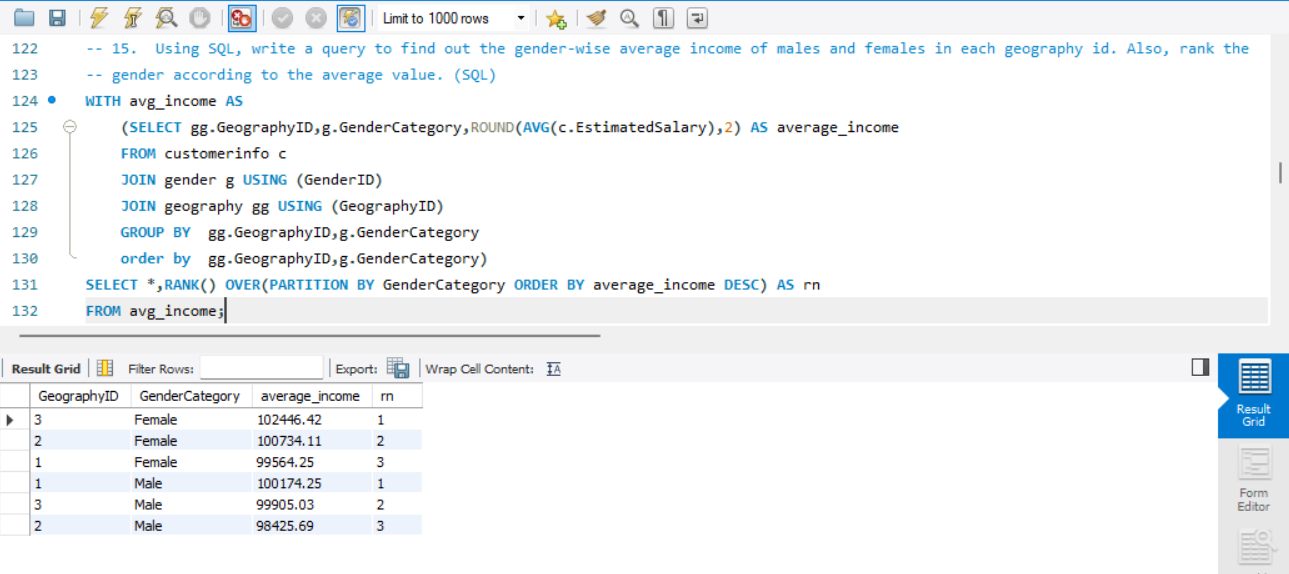
GROUP BY gg.GeographyID,g.GenderCategory

order by gg.GeographyID,g.GenderCategory)

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

FROM avg\_income;

**Output :**



**Insights:**

1. **Gender Income Analysis**: Understand the average income differences between males and females across various geographical regions.
2. **Income Disparities**: Identify regions where there are significant income disparities between genders.
3. **Targeted Policies**: Develop targeted financial and support policies to address income disparities in specific regions.
4. 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+).

Query calculates the average tenure of people who have exited the bank in each age bracket (18-30, 31-50, 50+). It groups the data by age brackets using a CASE statement and calculates the average tenure for each bracket using the AVG () function. As per the analysis 50+ age bracket members having highest number of tenures was 4.8330.

|  |  |
| --- | --- |
| Age Bracket | Average Tenure |
| 18-30 | 4.777 |
| 30-50 | 4.8899 |
| 50+ | 4.833 |

**SQL CODE :**

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;

**Output:**

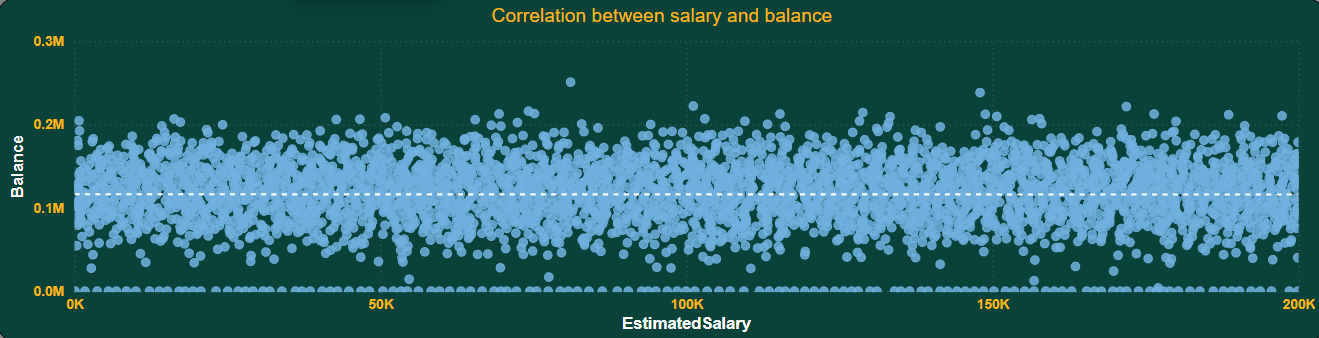


**Insights:**

1. **Tenure Analysis**: Understand the average tenure of customers who have exited within different age brackets.
2. **Retention Strategies**: Identify which age brackets have shorter tenures and might need targeted retention strategies.
3. **Age-Specific Patterns**: Recognize patterns in customer exits based on age, helping to tailor engagement strategies.
4. Is there any direct correlation between salary and balance of the customers? And is it different for people who have exited or not? (Power BI)

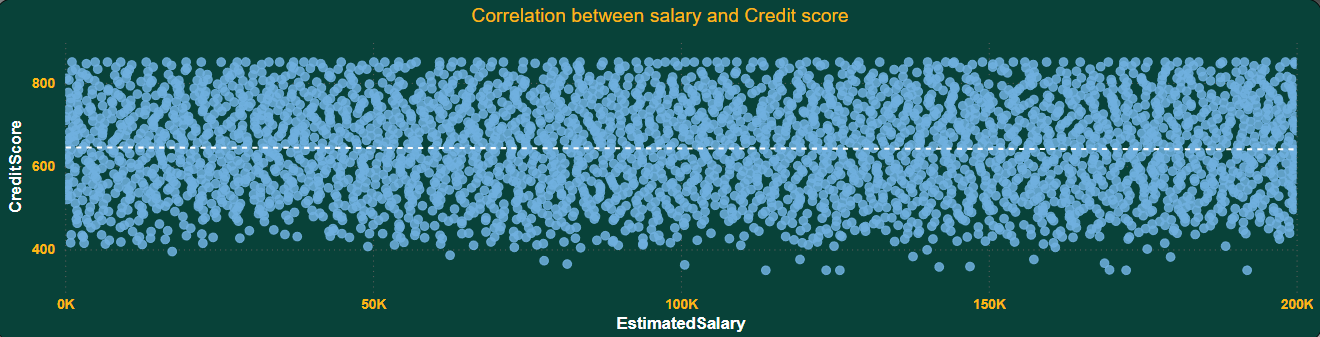
As per our analysis we get very weak correlation bond between salary and balance. In Power Bi with the help of Quick Measure we get only **0.01** Strength between them.

And it is also different for people who have exited or not but those also come in very weak Strength. The analysis reveals a very weak correlation between salary and account balance, and this correlation differs for customers who have exited compared to those who have not, it suggests that other factors may have a stronger influence on balance and customer churn.



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

1. Is there any correlation between salary and Credit score of customers? (Power BI)

If the strength of the correlation between salary and credit score is 0.0, it suggests that there is no linear relationship between these two variables. However, it's important to note that a correlation coefficient of 0.0 only measures linear relationships, and there could still be other types of relationships (such as non-linear or categorical) between the variables that are not captured by the correlation coefficient.

**Insights :**

* **Tailored Financial Products:** Develop and market financial products tailored specifically to customers with fair credit scores who have high incomes. This could include premium credit cards, investment opportunities, and personalized banking services.
* **Credit Improvement Programs:** Implement programs to help customers with good and poor credit scores improve their financial health. Offer credit counseling, budgeting tools, and incentives for financial education participation.

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

bank. (SQL)

Query first categorizes the credit scores into buckets (e.g., Poor, Fair, Good, Very Good, Excellent) using a CASE statement. Then, it counts the number of customers who have churned the bank (Exited) within each credit score bucket. Finally, it ranks the buckets based on the number of churned customers using the RANK () window function.

As per the analysis the maximum number of customers those churn the bank from ‘GOOD’ Credit score bucket and the minimum number of customers from ‘EXCELLENT’.

|  |  |  |
| --- | --- | --- |
| Credit Score Bucket | Customers | Rank |
| Good | 753 | 1 |
| Fair | 513 | 2 |
| Very Good | 492 | 3 |
| Poor | 152 | 4 |
| Excellent | 127 | 5 |

**SQL CODE :**

select case

when CreditScore >= 800 then 'Excellent'

when CreditScore < 800 and CreditScore >= 740 then 'Very Good'

when CreditScore < 740 and CreditScore >= 670 then 'Good'

when CreditScore < 670 and CreditScore >= 580 then 'Fair'

else 'Poor'

end as CreditScoreSegment,

count(CustomerID) as TotalCustomers,

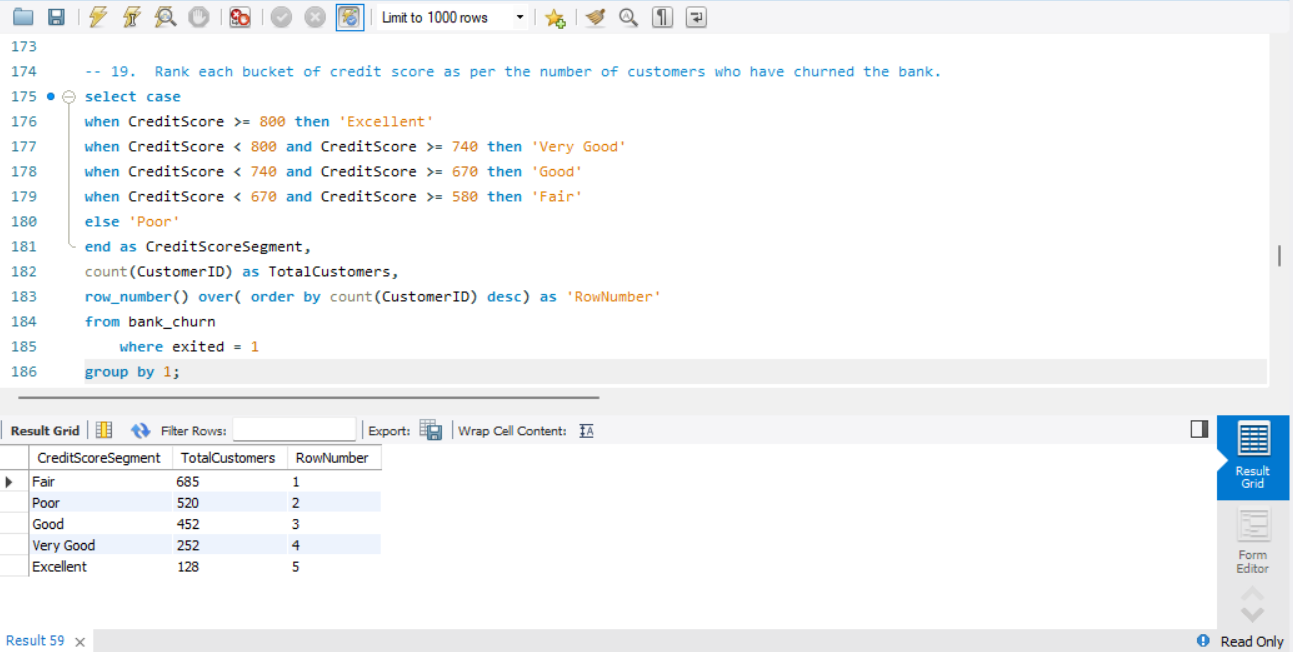
row\_number() over( order by count(CustomerID) desc) as 'RowNumber'

from bank\_churn

where exited = 1

group by 1;

**Output :**



**Credit Score Segment Ranking**:

* The Poor credit score segment has the highest number of churned customers, indicating that customers with poor credit scores are most likely to churn.
* Following Poor, the segments in descending order of churned customers might be Fair, Good, Very Good, and Excellent.

**Churn Distribution**:

* The distribution of churned customers across credit score segments can reveal important patterns. For instance, a higher concentration of churn in lower credit score segments (Poor and Fair) suggests that financial instability may be a significant factor in customer churn.

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. (SQL)

Query calculates the number of customers who have a credit card in each age bucket and calculates the average number of credit cards per customer for each bucket. Then, it calculates the average number of credit cards across all buckets and selects only those buckets where the average number of credit cards per customer is less than the overall average. Finally, it retrieves the age buckets, the total number of customers in each bucket, and the average number of credit cards per customer for those selected buckets.

Part A

**SQL CODE :**

select case

when Age between 18 and 30 then 'Young Adults'

when Age between 30 and 50 then 'Adults'

else 'Old'

end as AgeBucket,

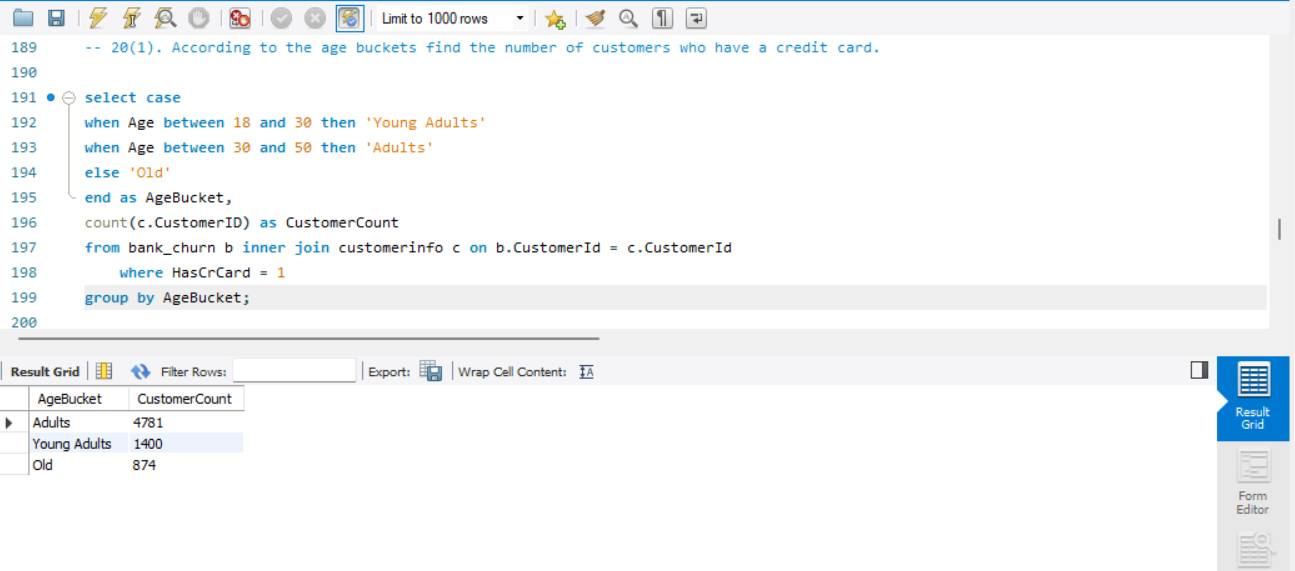
count(c.CustomerID) as CustomerCount

from bank\_churn b inner join customerinfo c on b.CustomerId = c.CustomerId

where HasCrCard = 1

group by AgeBucket;

**Output :**



Part B

**SQL CODE :**

SELECT CASE

WHEN Age BETWEEN 18 AND 30 THEN 'Young Adults'

WHEN Age BETWEEN 30 AND 50 THEN 'Adults'

ELSE 'Old'

END AS AgeBucket,

SUM(HasCrCard) AS TotalCards

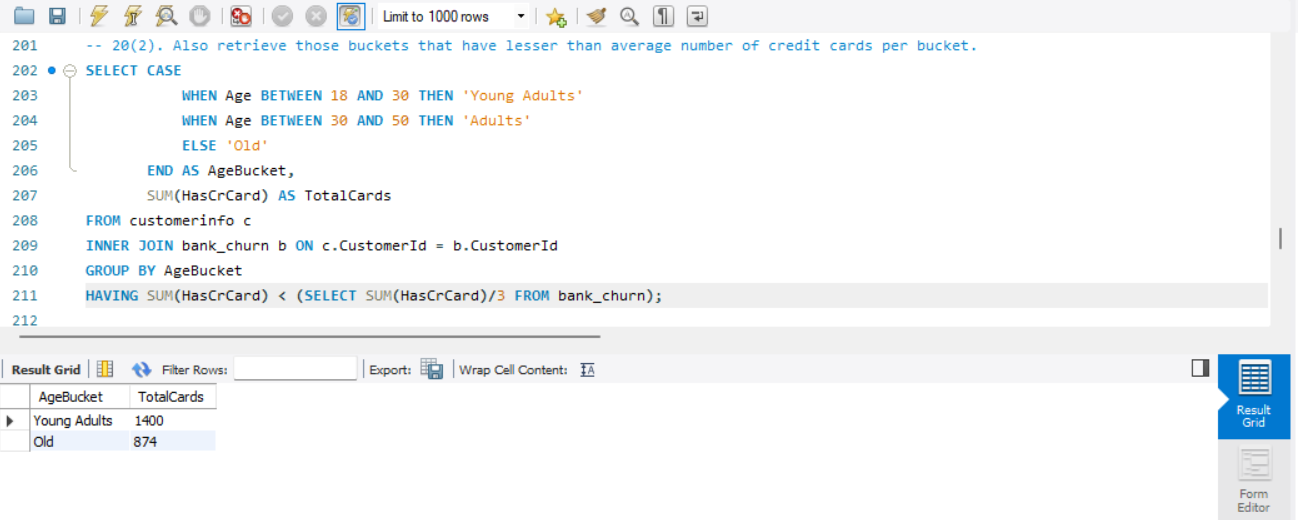
FROM customerinfo c

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

GROUP BY AgeBucket

HAVING SUM(HasCrCard) < (SELECT SUM(HasCrCard)/3 FROM bank\_churn);

**Output :**

****

|  |  |
| --- | --- |
| Age Bracket | Customers |
| 18-30 | 1400 |
| 50+ | 874 |

Insights :

**Implications for Targeted Strategies**:

* **Focus on Younger Customers**: There might be an opportunity to educate and engage younger customers (18-25) on the benefits of credit cards and how to use them responsibly.
* **Engage Older Customers**: For older customers (66+), consider offering tailored financial products that address their specific needs and preferences, which might not necessarily involve traditional credit cards.

**Credit Card Distribution Across Age Buckets**:

* The query categorizes customers into age buckets and calculates the number of customers with a credit card in each bucket.
* It then determines the average number of credit cards per customer within each age bucket.

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

According to me we must divide this question in two parts.

Rank the Locations based on the number of customers who have churned the bank.

Rank the Locations based on the average balance of the learners.

In first part we use ranks the customers based on the number of churned customers using the RANK () window function. After that we need to join Bank and CustomerInfo table and apply where condition to choose only exited customers and use group by function.

|  |  |  |
| --- | --- | --- |
| Geography | Cust\_ Churned | Rank\_Cust\_Churned |
| Germany | 814 | 1 |
| France | 810 | 2 |
| Spain | 413 | 3 |

In second part we use ranks the customers based on the customers average balance using the RANK () window function. After that we need to join Bank and CustomerInfo table and use group by function.

|  |  |  |
| --- | --- | --- |
| Geography | Avg\_cust\_balance | Rank |
| Germany | 119730.1161 | 1 |
| France | 62092.63652 | 2 |
| Spain | 61818.14776 | 3 |

select GeographyID, count(c.CustomerID) as CustomerCount,

row\_number() over( order by count(CustomerID) desc) as 'CustomerRANK',

round(avg(Balance), 2) as AverageBalance,

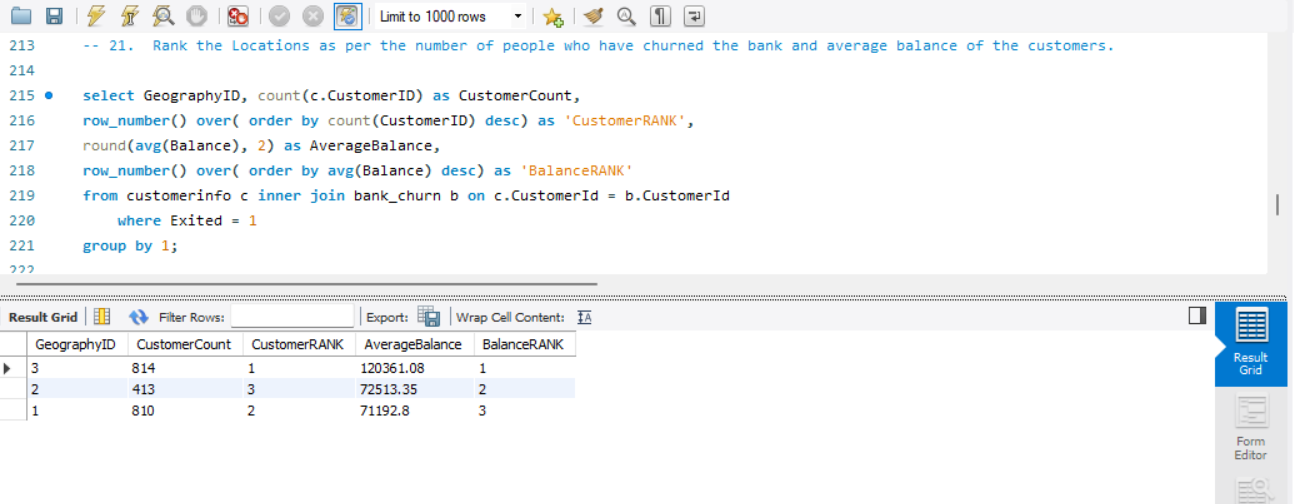
row\_number() over( order by avg(Balance) desc) as 'BalanceRANK'

from customerinfo c inner join bank\_churn b on c.CustomerId = b.CustomerId

where Exited = 1

group by 1;

**Output:**



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”. (SQL)

CustomerInfo is the table containing CustomerID and Surname.

Bank Table is the table where the primary key is a combination of CustomerID and Surname.

We use the CONCAT () function to concatenate CustomerID and Surname, separated by an underscore.

The JOIN condition links records from CustomerInfo to Bank Table based on matching CustomerID and Surname.

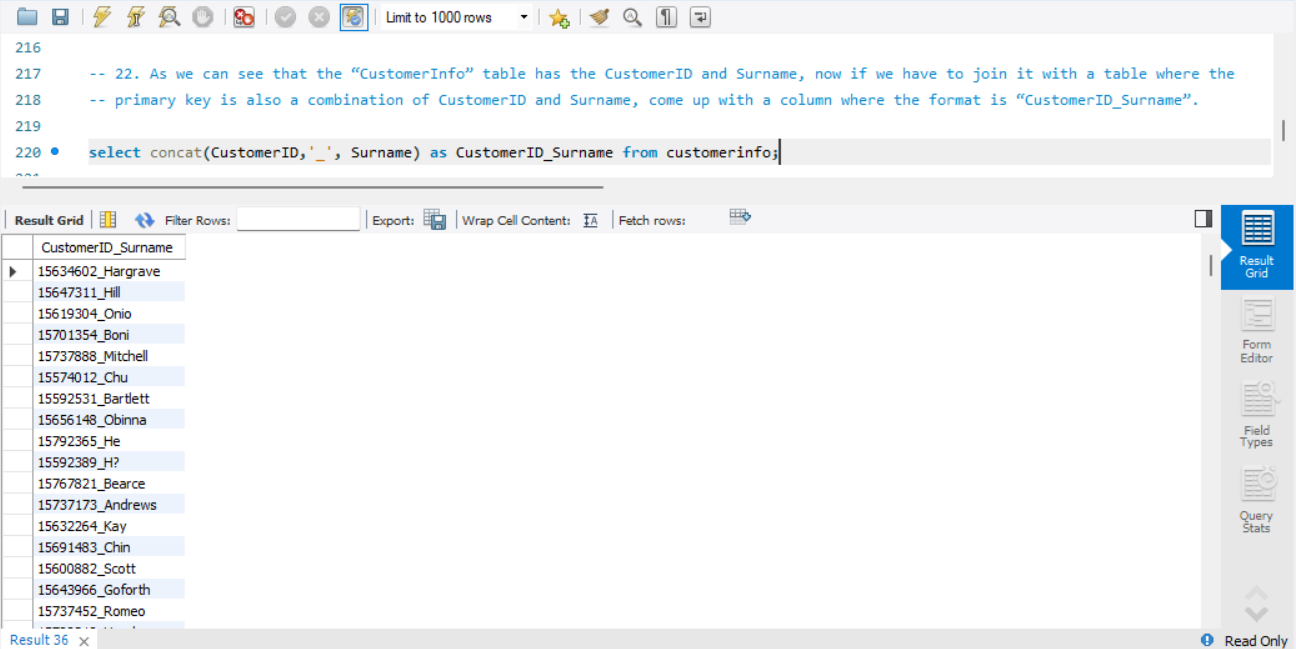
This query will generate a column named CustomerID\_Surname containing values in the format "CustomerID\_Surname" for each matching record in the joined tables. Adjust the table and column names to match your actual database schema.

|  |
| --- |
| CustomerId\_Surname |
| 15634602\_Hargrave |
| 15647311\_Hill |
| 15619304\_Onio |
| 15701354\_Boni |
| 15737888\_Mitchell |
| 15574012\_Chu |

**SQL CODE :**

select concat(CustomerID,'\_', Surname) as CustomerID\_Surname from customerinfo;

**Output:**



**Insights:**

1. **Unique Identifier Creation**: Create a unique identifier that combines CustomerID and Surname, useful for joining tables where the primary key is a combination of these two fields.
2. **Data Integration**: Simplify the process of joining customerinfo with other tables that use a composite primary key of CustomerID and Surname.
3. **Query Efficiency**: Enhance query efficiency by creating a straightforward way to reference and join customer records.
4. **Data Consistency**: Ensure consistency in referencing customer records across different tables.
5. Without using “Join”, can we get the “ExitCategory” from ExitCustomers table to Bank\_Churn table? If yes do this using SQL.

* First Method Using Case-

**SELECT**

**CASE**

**WHEN Exited = 0 THEN 'Exit'**

**ELSE 'Retain'**

**END AS Exited**

**FROM question23\_bank.**

This query will return a result set where each row contains the label "Exit" if the corresponding customer has exited (Exited = 0) or "Retain" if they have not exited (Exited ≠ 0). Adjust the table and column names as needed to match your actual database schema.

* Second method using Update-

**UPDATE question23\_bank**

**SET Exited = 'Exit'**

**WHERE Exited = 1;**

**UPDATE question23\_bank**

**SET Exited = 'Retain'**

**WHERE Exited = 0;**

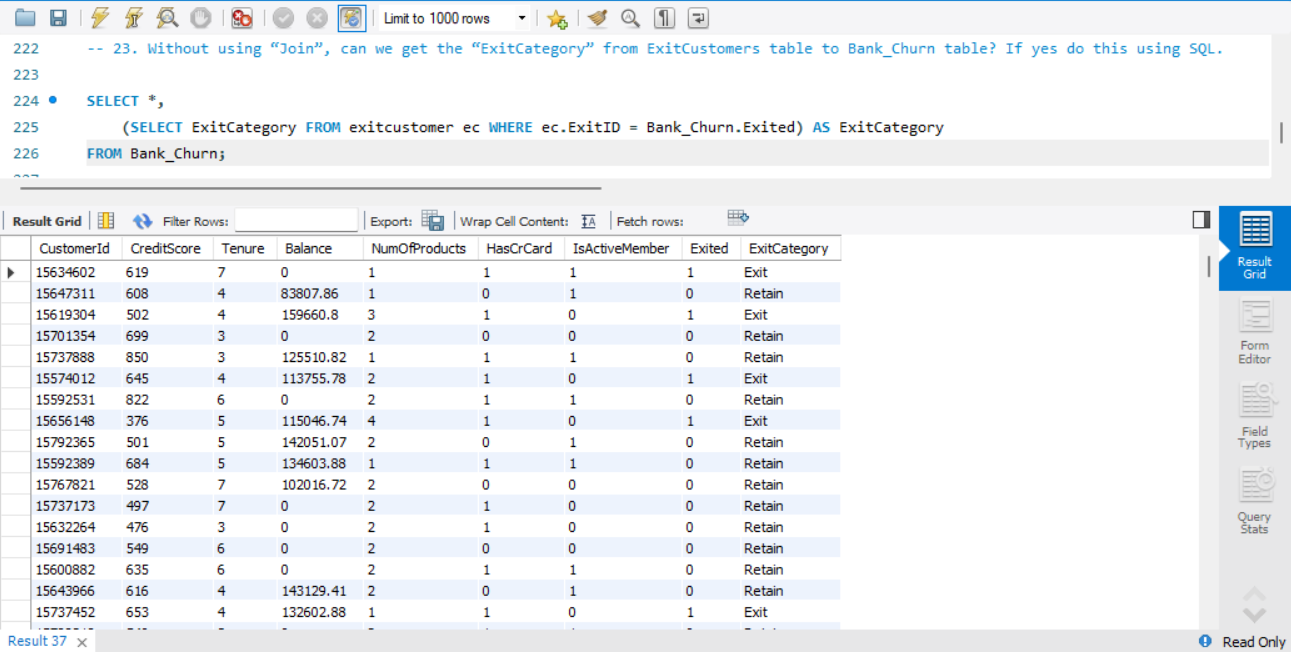
These statements will update the "Exited" column, accordingly, setting it to "Exit" where the value is 1 and "Retain" where the value is 0. If "Exited" is stored as a string, you need to adjust the values accordingly.

**SQL CODE :**

SELECT \*,

(SELECT ExitCategory FROM exitcustomer ec WHERE ec.ExitID = Bank\_Churn.Exited) AS ExitCategory

FROM Bank\_Churn;

****

**Insights:**

* **Correlated Subquery Usage**: Demonstrates how to use correlated subqueries to retrieve related data without explicit joins.
* **Data Integration**: Integrate data from exitcustomer into Bank\_Churn to enrich the dataset with exit categories.

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

In the provided data, we didn't identify any missing values. However, we noticed that some columns contained values in the format of 0 and 1, which presumably represent categorical variables or binary indicators. To ensure consistency and clarity in the data, we utilized Excel and employed the "Find and Replace" function to transform these values into the correct format or interpretation.

Handling missing values is a critical aspect of data preprocessing. Here are some common approaches to deal with missing values:

* Removal: If the proportion of missing values is small compared to the total dataset, you can simply remove rows or columns containing missing values. However, this approach may lead to loss of valuable information.
* Imputation: Missing values can be replaced with a substituted value. This could be the mean, median, or mode of the column for numerical data, or the most frequent category for categorical data. Imputation helps retain the data structure but may introduce bias.

The choice of method depends on various factors such as the amount of missing data, the nature of the dataset, and the specific objectives of the analysis. Each approach has its advantages and limitations, and it's essential to carefully consider the implications of handling missing values in different ways.

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”. (SQL)

It selects the CustomerId and Surname columns from the customerinfo table where the Surname ends with "on". The % symbol in the LIKE clause is a wildcard that matches any sequence of characters, so %on matches any string that ends with "on". This effectively retrieves the customer IDs and surnames of customers whose surnames end with "on".

|  |  |
| --- | --- |
| CustomerId | Surname |
| 15788218 | Henderson |
| 15750181 | Sanderson |
| 15788448 | Watson |
| 15585768 | Cameron |
| 15592461 | Jackson |
| 15640635 | Capon |
| 15676966 | Capon |

**SQL CODE :**

select c.CustomerId,

Surname,

IsActiveMember

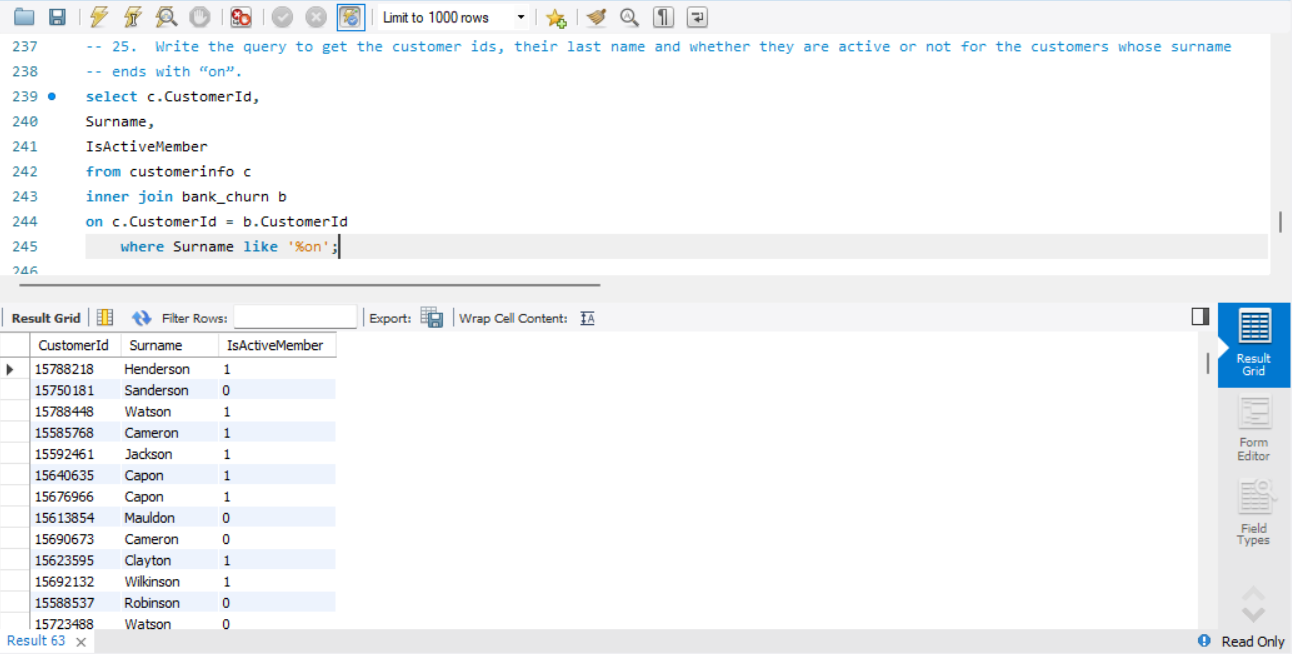
from customerinfo c

inner join bank\_churn b

on c.CustomerId = b.CustomerId

where Surname like '%on';

**Output :**



**Insights :**

**Surname Filtering**:

* The LIKE '%on' clause filters out customers whose surname ends with "on". This pattern matching is case-sensitive and ensures that only relevant surnames are retrieved.

**Active Membership Status**:

* By including the IsActiveMember field, the query provides insight into whether these customers are currently active members.
* This information is useful for understanding the engagement and status of customers with specific surname patterns.

**Data Integration**:

* The inner join between customerinfo and bank\_churn ensures that only customers present in both tables are included in the result. This helps in maintaining data integrity and accuracy.

1. 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.

**Approach:**

1. **Identify Inconsistencies**:
   * The query checks for inconsistencies where IsActiveMember = 1 (indicating the customer is an active member) and Exited = 1 (indicating the customer has exited). These conditions are logically contradictory.
2. **Data Analysis**:
   * Select all columns from the bank\_churn table to understand the context and potential reasons for the discrepancies.

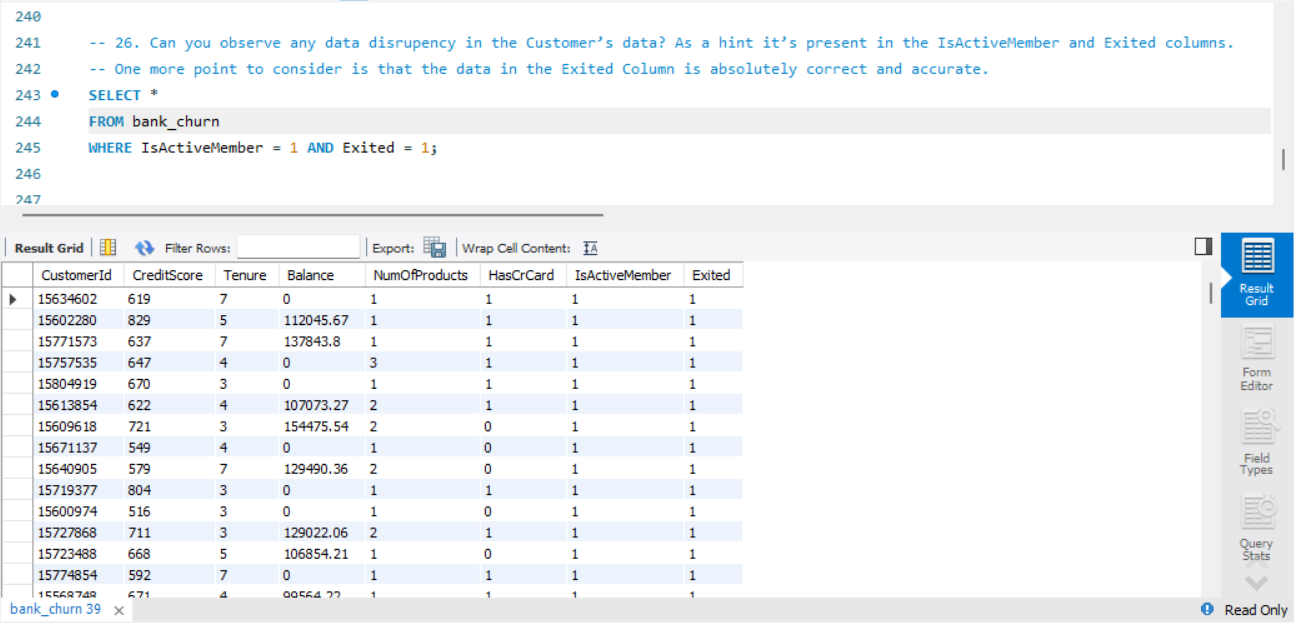
**SQL CODE :**

SELECT \*

FROM bank\_churn

WHERE IsActiveMember = 1 AND Exited = 1;

**Output :**



**Insights:**

* **Logical Inconsistency**: The logical inconsistency where a customer cannot be both an active member (IsActiveMember = 1) and exited (Exited = 1) highlights data quality issues.
* **Data Integrity**: Ensuring data integrity involves correcting such discrepancies to maintain accurate records.
* **Data Cleaning**: The dataset requires cleaning to resolve these contradictions, which could involve updating or correcting the values in the IsActiveMember column based on the Exited status.
* **Impact on Analysis**: These discrepancies can affect analytical results and decisions, emphasizing the importance of data accuracy.

**Subjective**

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?

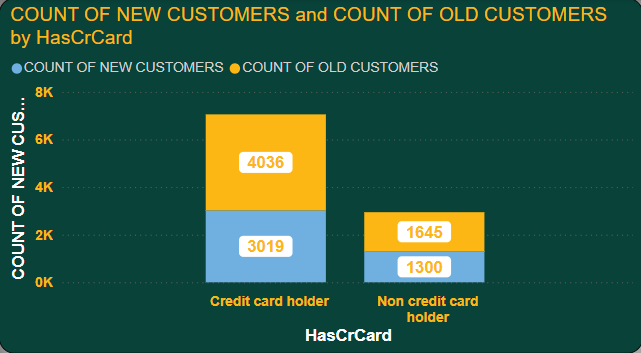
Customer behavior analysis can reveal valuable insights into the spending habits of long-term customers compared to new customers, shedding light on patterns that can inform strategies for customer loyalty.

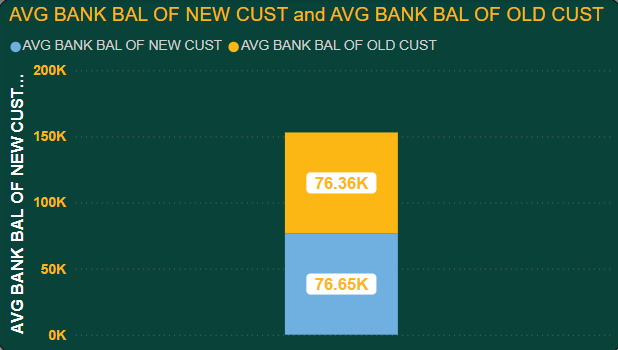
We can solve this question by seeing the average of balance with respect to tenure of customers. I considered customer with tenure more than 5 are loyal customer, and less than 5 are new customers.

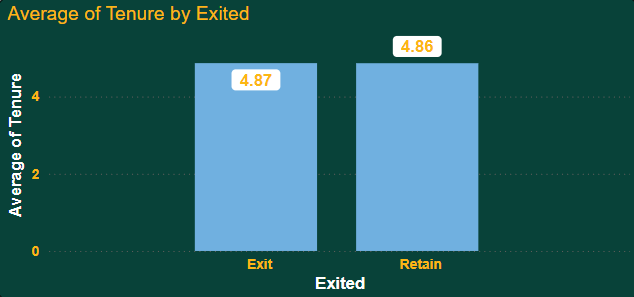
* By creating a measure in Power BI, we can see the average of premium customers is higher than normal customers.
* DAX formula used:

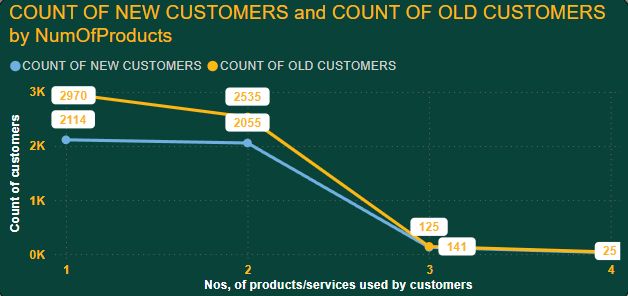
Loyal\_Customers = CALCULATE (AVERAGE ('try bank'[Balance]), 'try bank'[Tenure]>5)

* New\_Customers = CALCULATE (AVERAGE (Bank [Balance]), Bank [Tenure]<5)









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

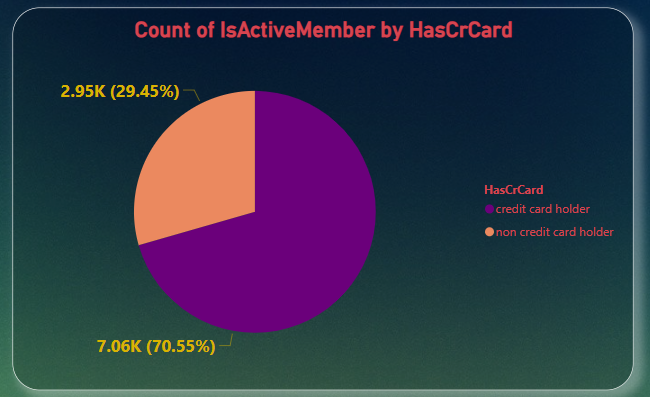
A product affinity study involves analyzing the relationships between different bank products or services to identify patterns of co-usage among customers. By understanding which products or services are frequently used together.

Customers who have credit card of the bank are more active with bank rather than customer don’t have credit card.

For this I created a bar chart with number of customers are credit card holder and is active.

**Result is customer with credit card and active = 7055**

**Customer without credit card and active = 2945**



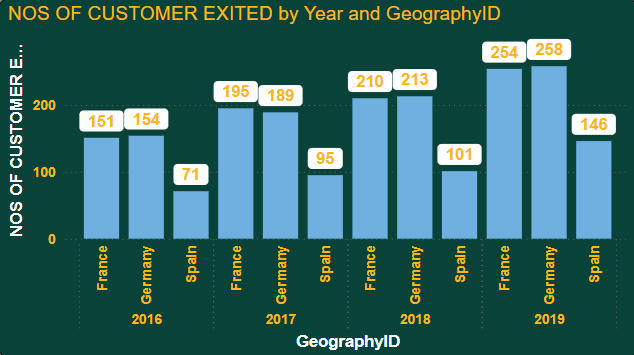
**Recommendations:**

* **Cross-selling Opportunities:**  
  Target non-credit card holders with personalized campaigns promoting the benefits of adopting a credit card, such as rewards, cashback, or exclusive offers to increase engagement.
* **Product Bundling:**  
  Offer bundled packages that combine credit cards with other popular products like savings accounts or loans, providing incentives to customers who adopt multiple services.
* **Personalized Marketing:**  
  Use data-driven approaches to identify customers who are likely to benefit from a credit card and tailor marketing efforts to encourage adoption of the product.
* **Retention Focus:**  
  Focus on retaining credit card holders through loyalty programs, bonus points, and special offers to maintain their higher engagement levels and reduce churn.

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

Analyzing geographic market trends and their correlation with economic indicators, active accounts, and customer churn rates can provide valuable insights into regional market dynamics and customer behavior.

To visualize this, I used Bar chart. From the chart we can see France has most active customers. While Germany and Spain have same number of active customers.



**Recommendations:**

* **Focus on France:**  
  Introduce premium services like wealth management and investment products to leverage the high customer engagement.
* **Enhance Engagement in Germany:**  
  Offer tailored financial products such as low-risk credit options and investment tools to boost engagement and manage customer churn.
* **Opportunity in Spain:**  
  Focus on providing affordable financial solutions, like low-interest credit, and offer financial literacy programs to drive engagement and reduce churn.
* **Targeted Marketing:**  
  Develop region-specific marketing strategies aligned with local economic conditions to maximize customer acquisition, retention, and engagement.

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

As per our analysis and given facts “Germany” seems to be in highest financial risk because the average balance of customers from Germany is higher than average salary.

**Age Segment (20-60 years)**:

* This age group poses a higher financial risk as they tend to have lower credit scores compared to other age ranges. Lower credit scores suggest potential late payments or financial instability.

**Geography (France)**:

* While customers from France show slightly lower credit scores, the difference is not significant enough to pose a high risk compared to other regions.

**Estimated Salary**:

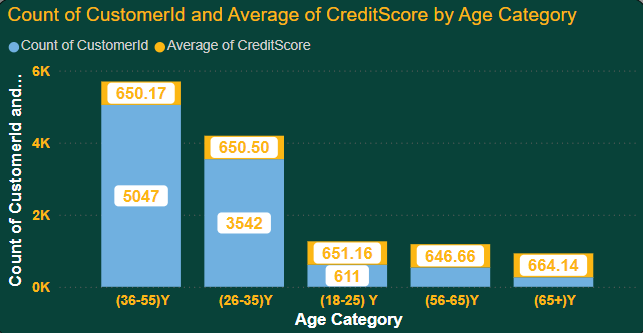
* There is no clear pattern indicating salary as a risk factor, with credit scores fluctuating by only about 10 points across different salary bins. Salary does not strongly correlate with financial risk.

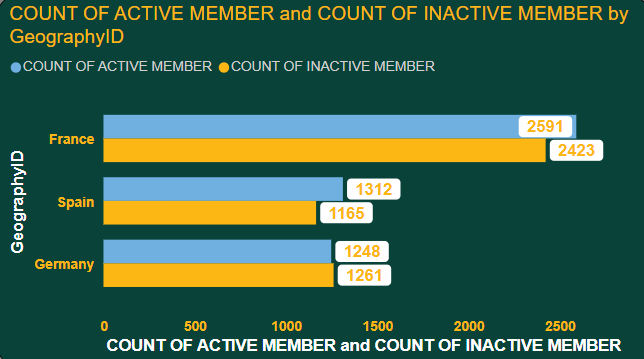
**Credit Card Holders:**

* Customers with credit cards show slightly lower credit scores, indicating a marginally higher risk, though the difference is minimal.

**Balance**:

* Customers with higher balances show decreasing average credit scores, suggesting that higher balances may be associated with higher financial risk due to potential difficulty in managing larger debts.





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?

Predicting customer lifetime value (CLV) involves forecasting the future value that a customer will generate over their entire relationship with the business.

* Data Collection and Preparation:

Gather historical data on customer transactions, interactions, demographics, and tenure.

* Define Customer Segments:

Segment customers based on relevant attributes such as demographics, behavior, transaction history, and geographic location.

* Feature Engineering:

Customer demographics such as age, gender, income, and location.

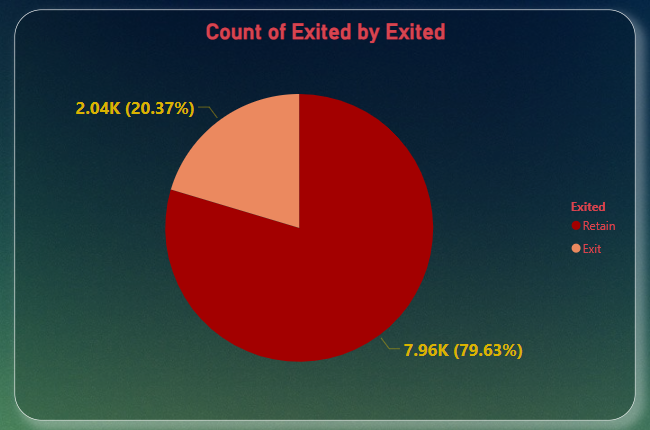
Tenure or length of the customer relationship with the business.

Continuously monitor and evaluate the performance of the CLV models over time.

Update the models periodically with new data and retrain them to improve accuracy and relevance.

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 within the dataset, you can employ various analytical techniques and metrics.From Available dataset we can see customer retained are more than customer exit. Bank can increase the retention rate by giving more offers. Retention rate is almost 80%.



**Additional Information Needed**:

* **Campaign Data**:
  + Detailed data on campaign type, medium, duration, target users, and spending is crucial for evaluation.
* **Customer Data**:
  + Customer profiles, including whether they are new or returning, would help in tracking acquisition vs. retention.
* **Time-Based Sales Data**:
  + Sales or transaction data during the campaign period and a comparison baseline from before the campaign to evaluate impact.
* **Engagement Data**:
  + Insights into customer engagement levels (e.g., click-through rates, response to the campaign) would help assess campaign effectiveness.

**Conclusion**:  
To properly assess the impact of marketing campaigns on customer retention and acquisition, both detailed campaign data and customer behavior data are needed. Analyzing these elements together will provide a clearer picture of campaign effectiveness and its return on investment (ROI).

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

To explore common characteristics or trends among customers who have exited, you can analyze various factors that may contribute to their decision to leave. Gather data on customers who have exited, including demographic information, transaction history, product usage, interactions, and any available feedback or survey responses.

To identify common characteristics or trends among customers who have exited, the analysis highlights several key factors that contribute to customer churn:

* **Number of Products**:

As the number of products sold to a customer increases, the exit percentage also increases. This suggests that customers with multiple products may face challenges managing or utilizing them, leading to higher churn rates.

* **Age Bin**:

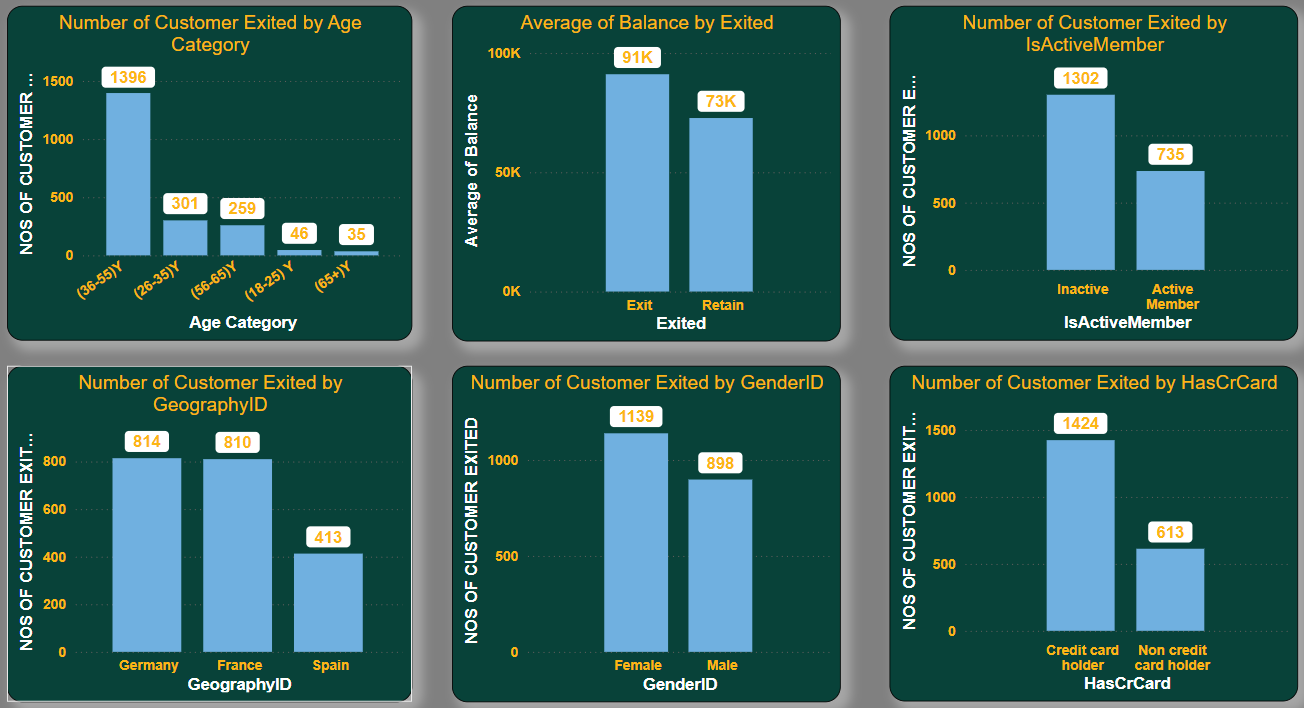
Customers aged 40 to 60 exhibit a significantly higher exit percentage compared to other age groups. This may be due to lifestyle changes, financial stability, or preferences in this cohort.

* **Active Category**:

Inactive users show a higher likelihood of churning. Customers who are less engaged or use fewer services are more prone to exit, highlighting the importance of maintaining customer activity and engagement.

* **Geography Location**:

Customers from Germany have a higher exit percentage than those from other regions. This geographical trend may indicate regional differences in customer preferences, satisfaction, or external factors influencing exits.



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

To determine the importance of 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' in predicting customer churn, we can use statistical analysis techniques. Yes, these factors are important to predict, including tenure and 'NumOfProducts' we can see how customers are buying with time. With active member and tenures, we can see how much time a customer spent time with bank.

**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.

To segment customers based on demographics and account details using SQL queries, we can use conditional filtering and aggregation functions to group customers into different segments. Here's a general approach to achieve this.

-- Segment customers based on age groups and calculate the average balance

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;

|  |  |  |
| --- | --- | --- |
| Age\_Group | Total\_Customers | Average\_Balance |
| 18-30 | 1968 | 73198.76 |
| 31-40 | 4451 | 75583.36 |
| 41-50 | 2320 | 79122.19 |
| 51+ | 1261 | 79951.38 |

-- Segment customers based on gender and calculate for each gender

|  |  |
| --- | --- |
| Gender | Total\_Customers |
| Female | 4543 |
| Male | 5457 |

-- Segment customers based on geographic location

|  |  |
| --- | --- |
| Geography | Total\_Customers |
| France | 5014 |
| Germany | 2509 |
| Spain | 2477 |

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?

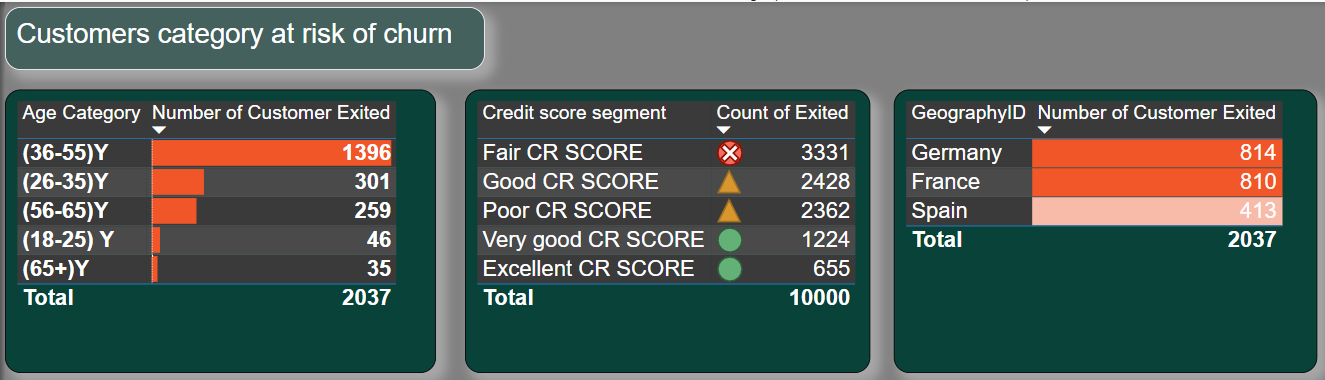
To create a conditional formatting setup in Power BI to visually highlight customers at risk of churn and evaluate the impact of Balance, Estimated Salary, Credit Score and Exited on customer retention. Create visualizations to display the results of the analysis, such as bar charts, line charts, or tables.

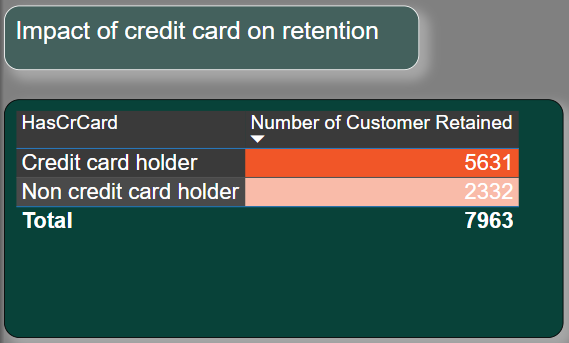
**Highlighting At-Risk Customers:**

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.





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?

As per the analysis every year the churn rate of bank was increases. 376 customers exit the bank in 2016 and the 658 customers exit the bank in 2019. From year 2016 to 2019 every year up to 100 customers was increased to exit the bank.

Below chart describes the churn rate is 20%.

In this, there are majorly four factors for which the exit % is higher in some cohort or varies majorly for good volume, those are

* Num of Product - as the number of products sold to the customer increases the exit % of the customer increases
* Age bin - as the customer is of 40 to 60 bin, the exit % of the customer is much higher as compared to the other cohort
* Active Category - the inactive users have a higher chance of being churned as compared
* Geography Location - The customers from Germany has the higher exit % as compared to the rest other countries

**Suggested Insights and Strategies to Reduce Churn:**

**Personalized Product Bundling**:

* Instead of pushing multiple products to customers, focus on personalized product bundling that aligns with customer needs and life stages. Simplifying product offerings and targeting specific customer profiles can improve satisfaction.

**Targeted Engagement for 40-60 Age Group**:

* Create targeted financial products and services that cater to the 40-60 age group. Offering retirement planning, investment options, and personalized financial advice could help retain this demographic.

**Increase Customer Engagement**:

* Focus on re-engaging inactive customers through targeted campaigns, loyalty programs, and improved customer experience. Offering incentives, personalized offers, or educational resources could help bring inactive users back into the fold.

**Geographical Focus on Germany**:

* Investigate the specific reasons for higher churn in Germany and tailor strategies to address them. This could include improving customer service, offering localized products, or responding to regional competitors.

**Monitor and Improve Customer Satisfaction**:

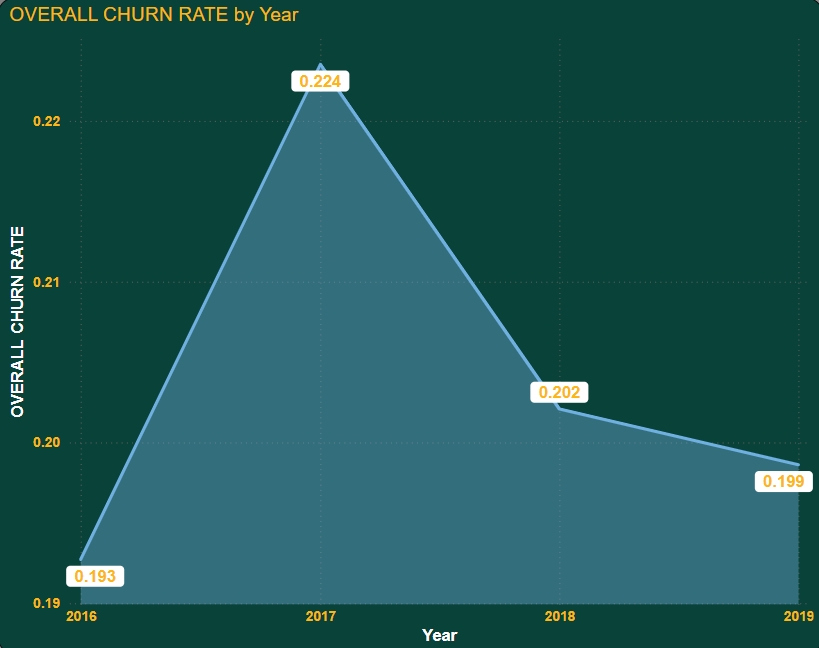
* Implement regular customer feedback loops, surveys, and satisfaction metrics to identify potential issues early. Addressing concerns in real-time can prevent customer dissatisfaction from leading to churn.

**Predictive Analytics**:

* Use predictive modeling to identify customers likely to churn based on historical behavior. Proactively reach out to these customers with targeted retention strategies, such as personalized offers or enhanced support.

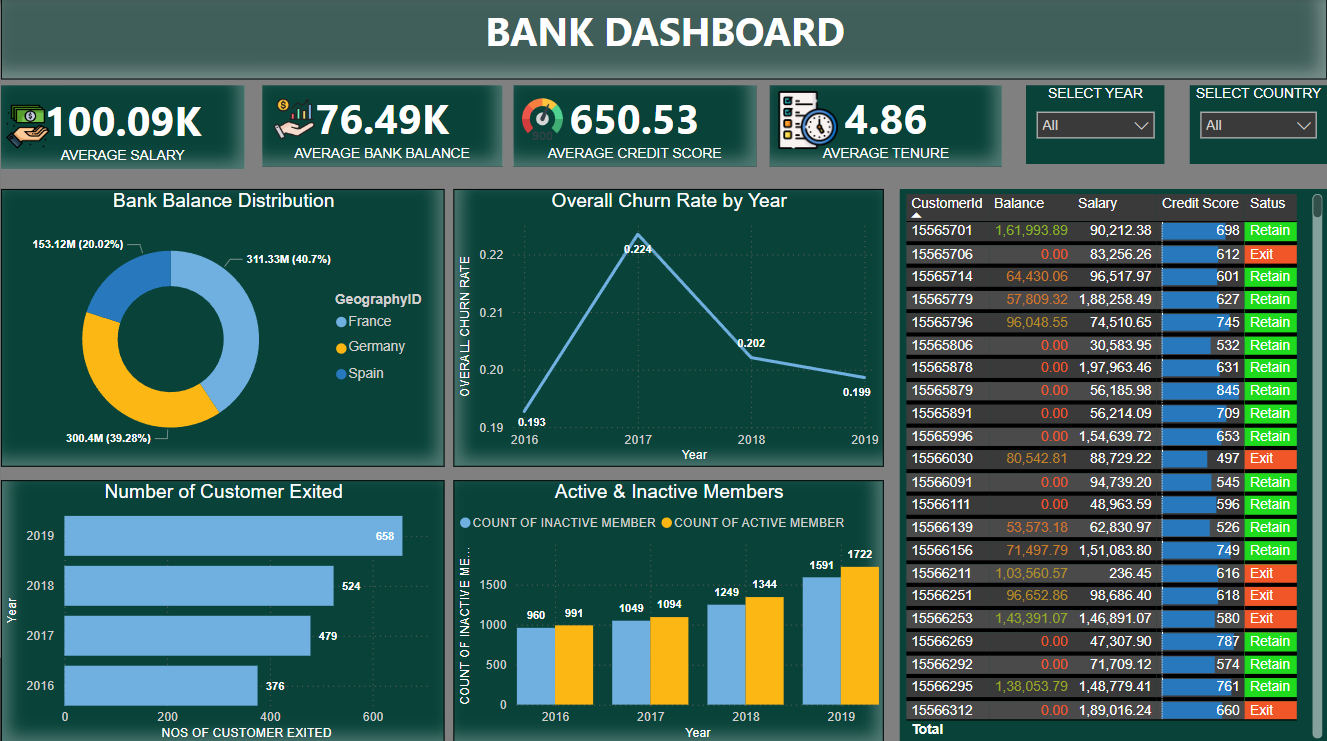
These strategies, when combined, can help the bank reduce the increasing churn rate and improve customer loyalty.





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

Here we have created the dashboard with the following graphs, customer id count, active and inactive member split, retain and exit members split, avg of credit score, avg of estimated salary, avg of balance and number of products sold. Apart from this I’ve created a parameter which contains 3 things, year, Gender category and Geography location, which is placed in the x-axis in all the graphs.



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

Sure, here's how you can approach creating a dashboard incorporating all KPIs and visualization-related metrics along with a slicer for selection:

**Identify KPIs and Metrics:**

* Review the dataset to identify key performance indicators (KPIs) and visualization-related metrics relevant to the objectives of the dashboard.

**Create Visualizations:**

* Based on the identified KPIs and metrics, create individual visualizations such as bar charts, line charts, pie charts, tables, or cards using Power BI.

**Design the Dashboard Layout:**

* Decide on the layout and arrangement of the visualizations on the dashboard canvas.

**Add Slicer for Selection:**

* Add a slicer visual to the dashboard canvas. Choose the field you want to use for slicing, such as date, region, or product category.

**Apply Formatting and Branding:**

* Apply consistent formatting, such as colors, fonts, and backgrounds, to unify the appearance of the dashboard and align it with the branding guidelines.

**Test and Iterate:**

* Test the functionality of the dashboard to ensure that slicer interactions work as expected and that all visualizations provide meaningful insights.

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

To modify the name of the "HasCrCard" column to "Has\_creditcard" in the "Bank\_Churn" table, you can use the ALTER TABLE statement in SQL. Here's the SQL query to achieve this:

**ALTER TABLE Bank\_Churn**

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

* ALTER TABLE: This keyword is used to make changes to an existing table structure.
* RENAME COLUMN: This clause specifies that we want to rename a column in the table.
* Bank\_Churn: This is the name of the table in which the column exists.
* HasCrCard: This is the current name of the column that we want to change.
* TO Has\_creditcard: This specifies the new name that we want to assign to the column.

By executing this SQL statement, the name of the "HasCrCard" column in the "Bank\_Churn" table will be changed to "Has\_creditcard".