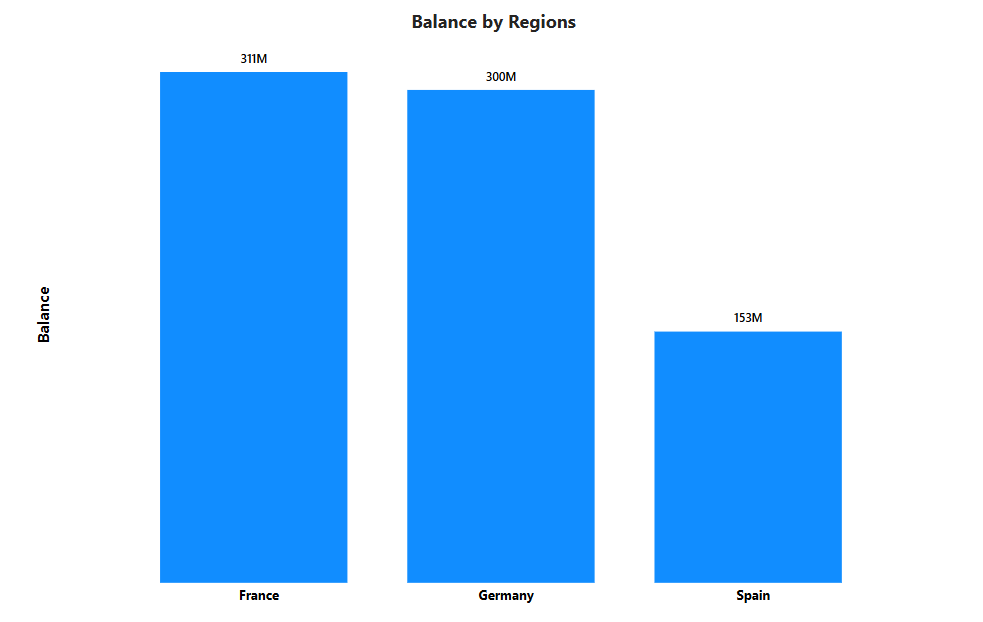
**Objective Questions:**

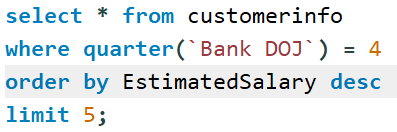
1. What is the distribution of account balances across different regions?



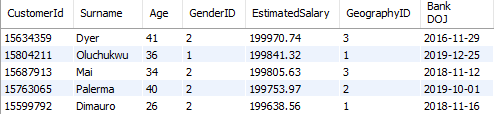
France has the highest account balance with the amount of 311M.

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

Query is

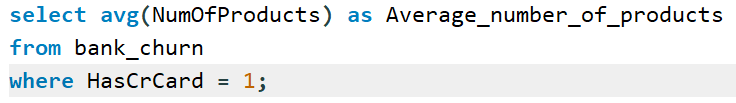


Answer is



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

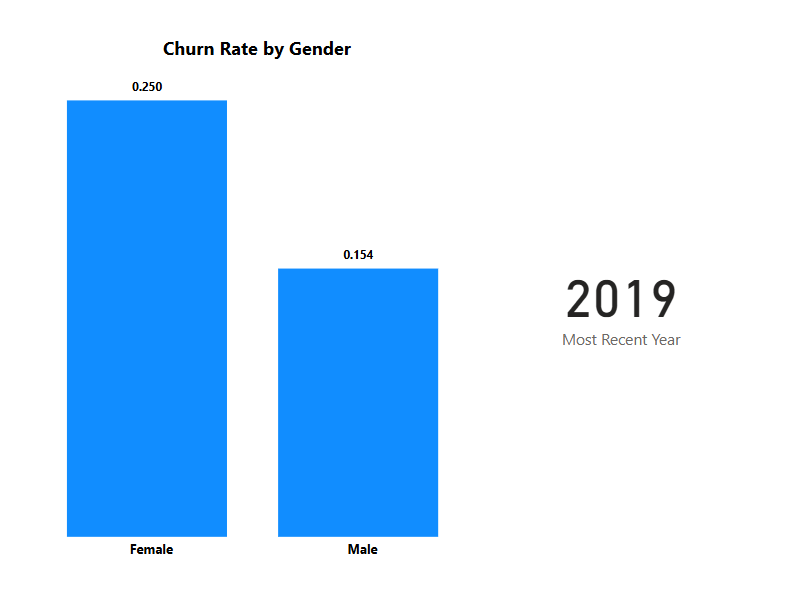
Query is



Answer is



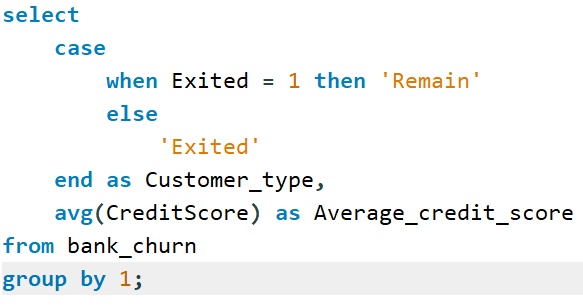
1. Determine the churn rate by gender for the most recent year in the dataset.



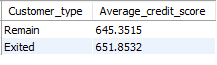
2019 is the most recent year in the dataset and in that year Females had 0.25 i.e. 25% churn rate, while males had 15% churn rate.

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

Query is

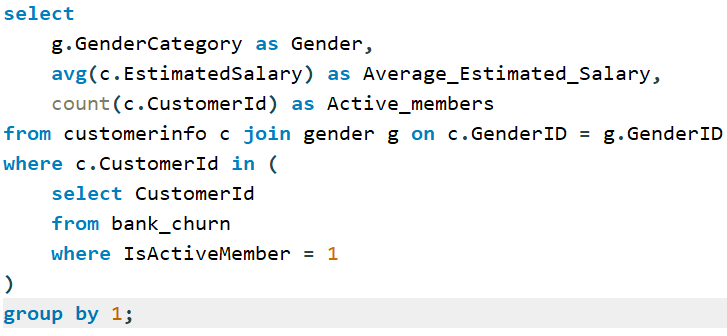


Answer is

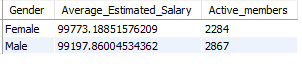


1. Which gender has a higher average estimated salary, and how does it relate to the number of active accounts? (SQL)

Query is

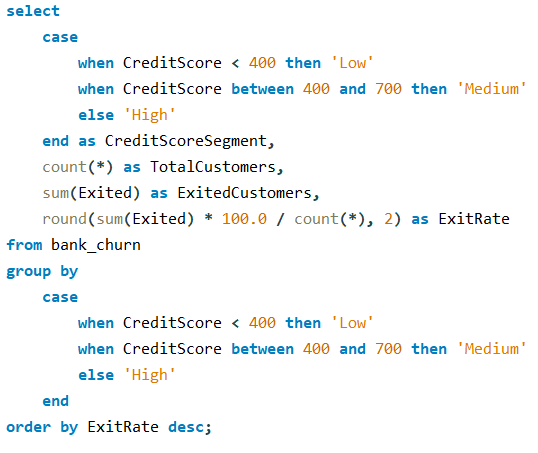


Answer is Females has highest average estimated salary who are active

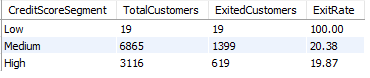


1. Segment the customers based on their credit score and identify the segment with the highest exit rate. (SQL)

Query is



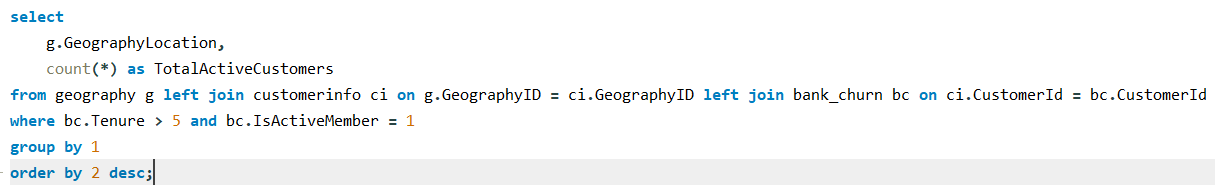
Answer is



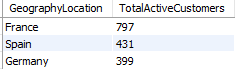
Low CreditScore customers have 100% exit rate.

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

Query is

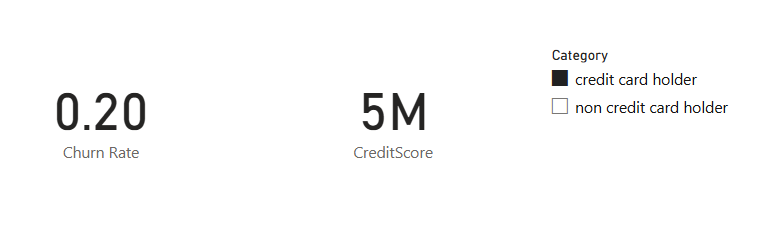
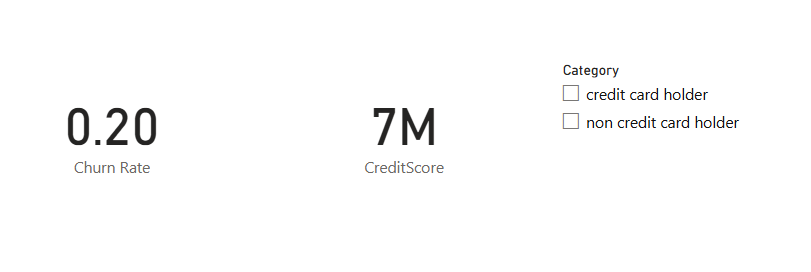


Answer is



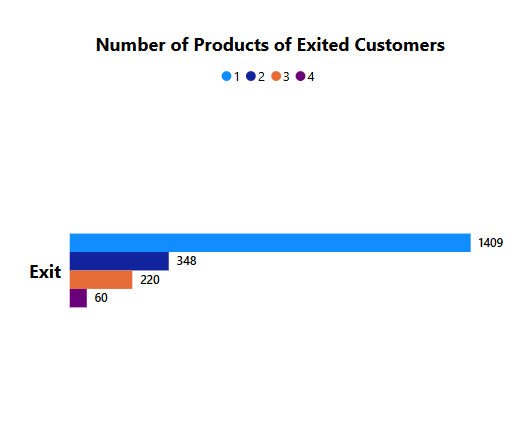
France has the highest number of Active members who have more than 5 years of tenure

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



There is no effect as such on whether having credit card or not but Credit score decreases for people having credit cards.

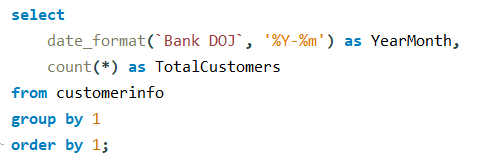
1. For customers who have exited, what is the most common number of products they have used?



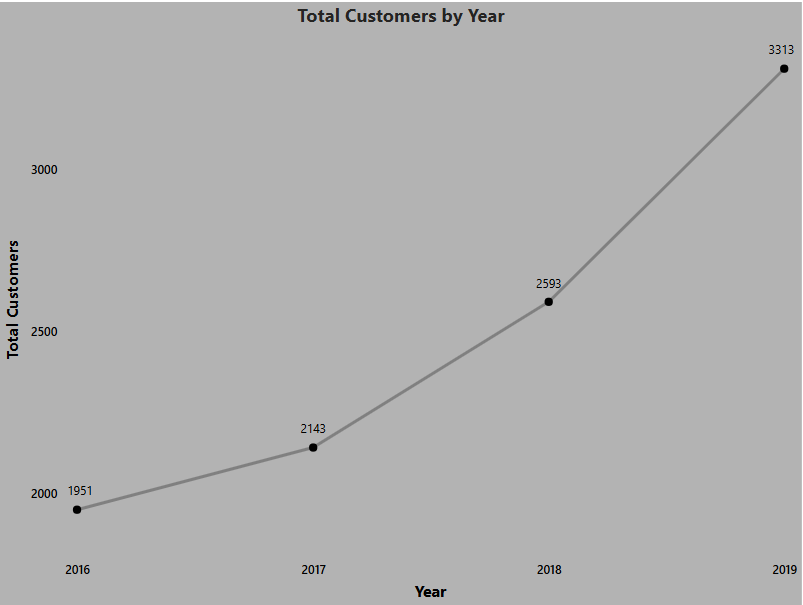
Product 1 is most purchased by the customers who exited while Product 4 is the least purchased.

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.

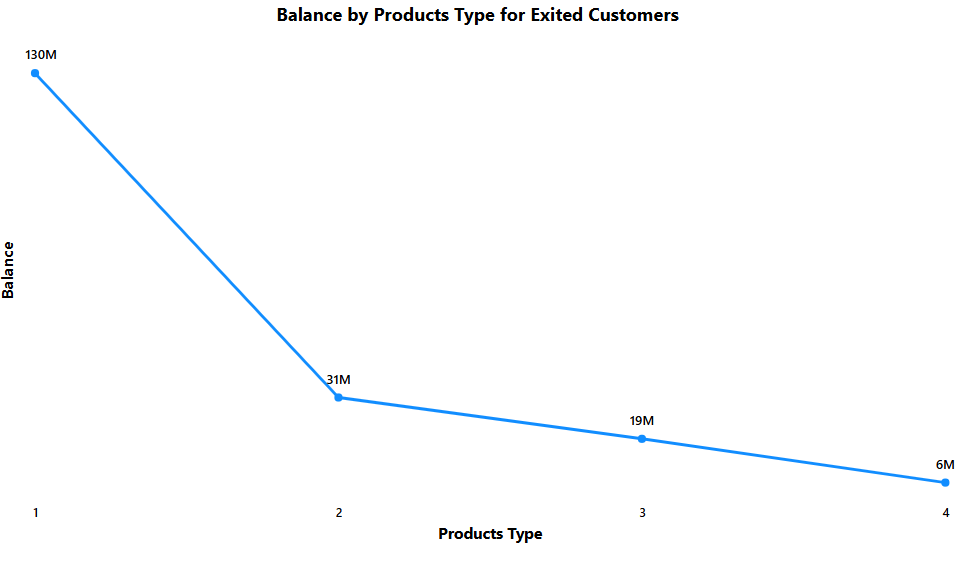
Query is



Graph is

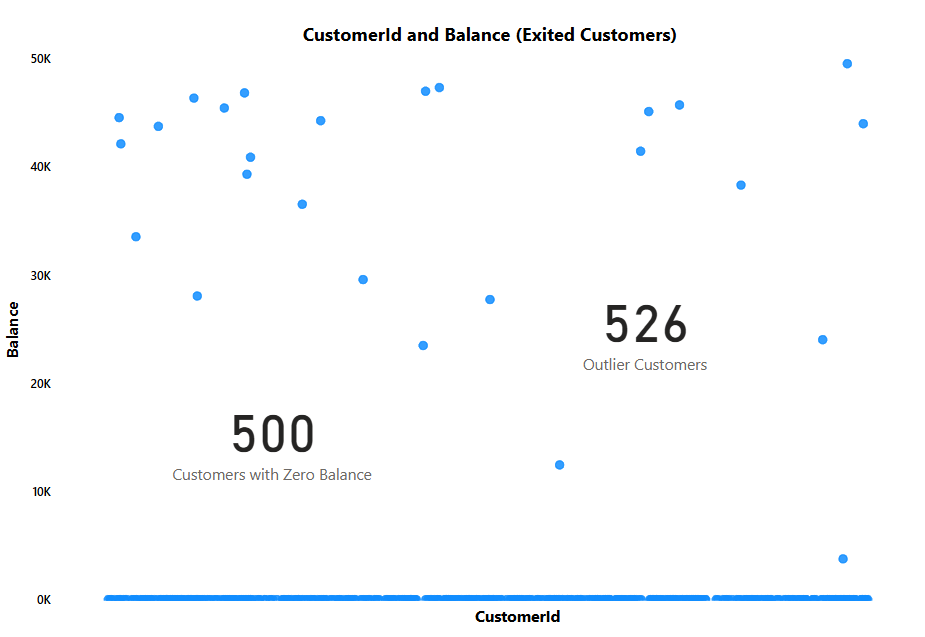


1. Analyze the relationship between the number of products and the account balance for customers who have exited.



As purchase of more products increased, account balances are going down and there is very long dip when customers buy 2 products rather than 1.

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



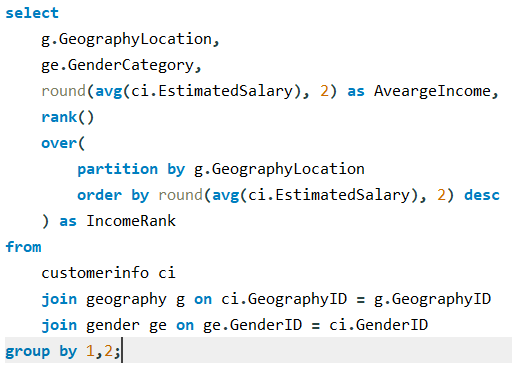
I considered outliers as customers having Balance less than 50K and total comes out to be 526 customers. In the scatter plot you can see in these also most customers around 500, have 0 balance.

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

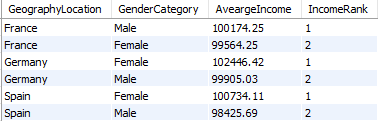
There are total 7 tables provided in the dataset. In those tables consisting only of categorical columns are activecustomer, creditcard, exitcustomer, gender and geography.

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)

Query is

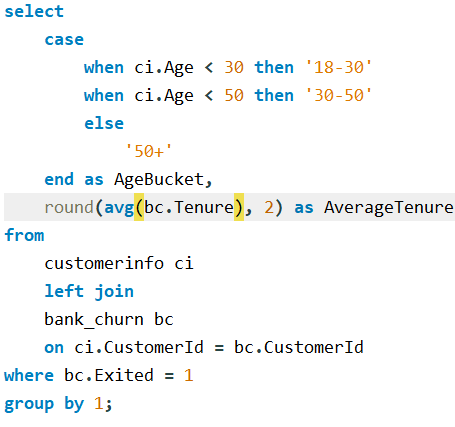


Answer is

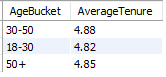


1. Using SQL, write a query to find out the average tenure of the people who have exited in each age bracket (18-30, 30-50, 50+).

Query is

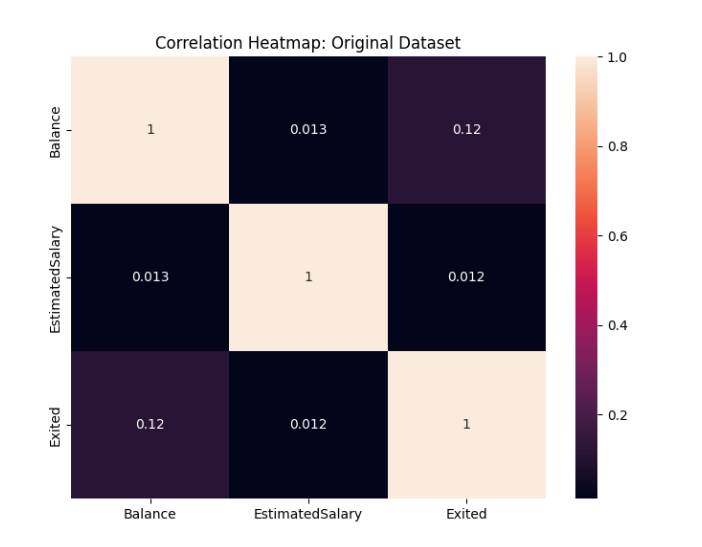
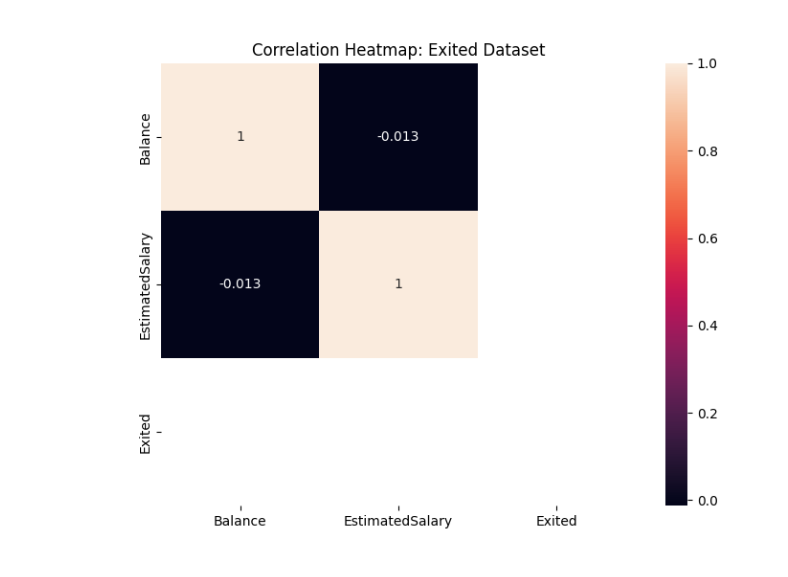


Answer is

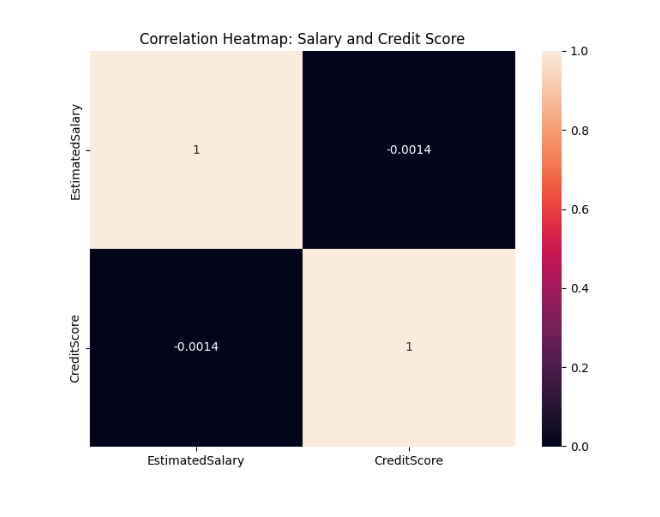


1. Is there any direct correlation between salary and the balance of the customers? And is it different for people who have exited or not?

These correlation heatmaps are generated using python scripts. There is positive but very small correlation between Salary and Balance which is 0.013, but for exited customers, the correlation is negative but small i.e. -0.013.

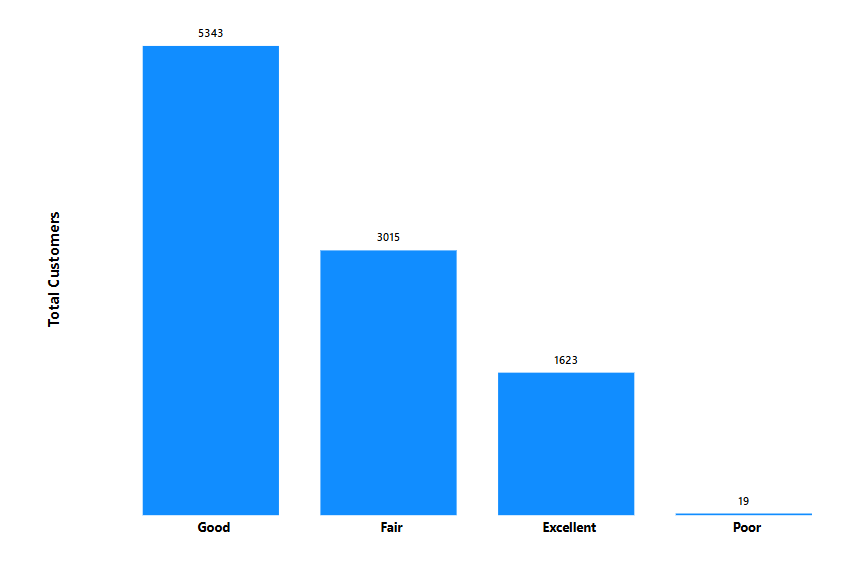


1. Is there any correlation between the salary and the Credit score of customers?



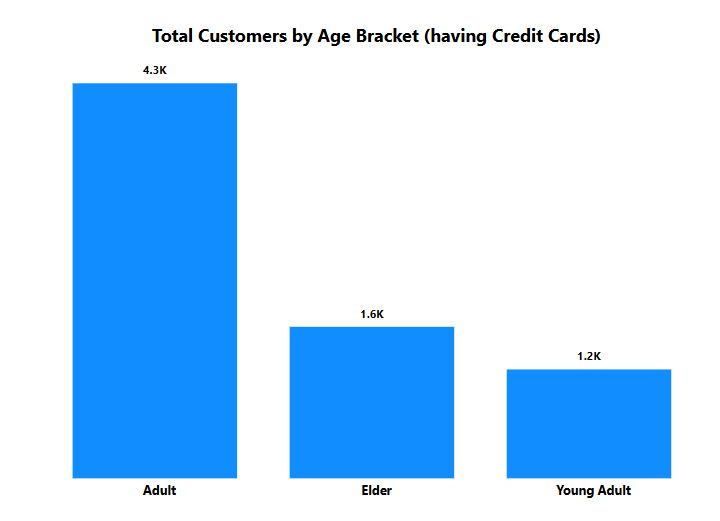
From the heatmap, we can see that there is negative correlation between Salary and Credit score but it is very close to zero.

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



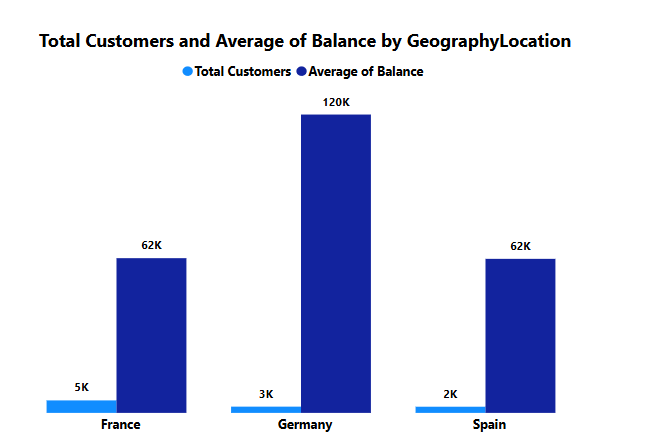
Ranking the Segments – Good, Fair, Excellent, Poor

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



Adult category (age from 30 to 45) has the highest number of customers who have credit cards.

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



In terms in Total customers churned, ranks of locations are France, Germany and Spain.

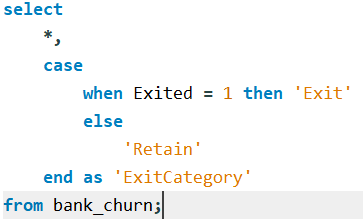
In terms of Average of Balance, ranking is Germany, France and Spain.

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”.

Yes we can create a calculated column in power bi, where we can concatenate both the columns to create a new column. This is the DAX formula

**CustomerID\_Surname = CONCATENATE(customerinfo[CustomerId], customerinfo[Surname])**

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



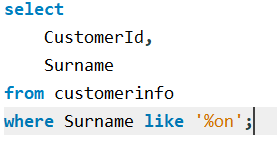
1. Were there any missing values in the data, using which tool did you replace them and what are the ways to handle them?

There are no missing values in the data.

There are several ways to deal with missing values

* Drop them – in this drop the rows or columns containing missing values based on the requirements.
* Fill them – Using Fill Up or Fill Down function or replacing missing values with mean/median/mode.
* Filter 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”.



1. Can you observe any data discrepancy 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.

Yes, there are discrepancies in the IsActiveMember column of bank\_churn table as if any customer exited than it should not be active.

**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?

**Observation**:  
Long-term customers (tenure > 5 years) typically hold higher balances (avg ~130,000) and use more products (avg ~2.2) than new customers (tenure ≤ 2 years), who have lower average balances (~70,000) and fewer products (avg ~1.5).  
**Implication**:  
This indicates stronger loyalty and engagement from long-term customers. They likely trust the bank more, suggesting higher cross-sell/up-sell potential.

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

**Observation**:  
The most common product combination is customers with both a *credit card and multiple accounts (NumOfProducts = 2 or more).*  
**Implication**:  
Bundling services like credit cards with account upgrades can enhance cross-sell efficiency.

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

**Observation**:  
Regions like Germany with higher economic engagement (higher avg. balance and salary) tend to have more active customers and lower churn rates.  
**Implication**:  
Focus marketing and retention strategies in regions showing economic growth for better ROI.

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

**High-Risk Segments**:

* Age group 18–30 with high salary but low tenure and multiple products (avg churn ~30%)
* Customers with NumOfProducts = 4 (churn rate ~45%)

**Implication**:  
Younger, high-salary customers with multiple products may churn due to unmet expectations. Targeted engagement is necessary.

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?

We can use the Linear Regression feature of Power BI or we can do it using Python also using Scikit learn.

In that we can use the features like Age, Credit Score, Balance, IsActiveMember, Salary, Products and our output variable will be Tenure

Insight: Helps forecast future revenue and improve LTV by targeting segments likely to stay longer.

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?

No campaign ID, source, or timestamp data available.

**Need**:  
Campaign tracking info (email click rates, acquisition dates) to analyze pre- and post-campaign effects on retention/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?

* Exited customers have higher product count (3–4) and lower activity rates (IsActiveMember = 0).
* Avg Credit Score is also lower by ~50 points.

**Conclusion**:  
Customers may exit due to complexity, poor engagement, or dissatisfaction with service quality.

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

* **IsActiveMember** and **NumOfProducts** have the highest correlation with churn.
* **EstimatedSalary** has minimal direct impact.

Yes, these features are predictive — especially **NumOfProducts** and **IsActiveMember**.

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

**Age Buckets with Credit Score Segmentation: -**

SELECT

CustomerID,

Age,

CreditScore,

CASE

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

WHEN Age BETWEEN 31 AND 50 THEN '31-50'

ELSE '50+'

END AS AgeBucket,

CASE

WHEN CreditScore < 580 THEN 'Poor'

WHEN CreditScore BETWEEN 580 AND 669 THEN 'Fair'

WHEN CreditScore BETWEEN 670 AND 739 THEN 'Good'

WHEN CreditScore BETWEEN 740 AND 799 THEN 'Very Good'

ELSE 'Excellent'

END AS CreditScoreSegment

FROM Bank\_Churn;

**Segment by Account Tenure and Product Usage: -**

SELECT

CustomerID,

Tenure,

NumOfProducts,

CASE

WHEN Tenure <= 2 THEN 'New'

WHEN Tenure BETWEEN 3 AND 5 THEN 'Mid-term'

ELSE 'Long-term'

END AS TenureSegment,

CASE

WHEN NumOfProducts = 1 THEN 'Single Product'

WHEN NumOfProducts = 2 THEN 'Double Product'

ELSE 'Multi Product'

END AS ProductUsageSegment

FROM Bank\_Churn;

**Combined Segmentation: Age, Gender, Geography, Product Usage: -**

SELECT

bc.CustomerID,

g.Gender,

ci.GeographyID,

bc.Age,

bc.NumOfProducts,

CASE

WHEN bc.Age < 30 THEN 'Young'

WHEN bc.Age BETWEEN 30 AND 50 THEN 'Middle-Aged'

ELSE 'Senior'

END AS AgeGroup,

CASE

WHEN bc.NumOfProducts = 1 THEN 'Low Engagement'

WHEN bc.NumOfProducts = 2 THEN 'Moderate Engagement'

ELSE 'High Engagement'

END AS EngagementLevel

FROM Bank\_Churn bc

JOIN Gender g ON bc.CustomerID = g.CustomerID

JOIN CustomerInfo ci ON bc.CustomerID = ci.CustomerID;

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?

First, we will create a calculated column

**RiskLevel = IF([Exited] = 1 && [IsActiveMember] = 0 && [NumOfProducts] > 2, "High Risk", "Low Risk")**

Then apply colour rules in a table/matrix for "High Risk". Also we can add card visuals to compare retention of credit card holders.

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?

* **Overall churn rate**: 20.37%
* **Annual churn**: Higher in Year 2 and Year 4.
* **At-risk profiles**: Young age (18–30), multiple products (3–4), inactive users.

Strategies: -

* Personalized engagement for high-product users
* Reactivation campaigns for inactive customers

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

KPIs created are: -

* Churn Rate
* Average Balance
* Total Customers
* Average Salary
* Average Credit Score

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

**Step 1 - Understand the Business Context**

It's crucial to understand what the bank cares about. In a CRM (Customer Relationship Management) context, some typical business goals might include:

* Reducing **customer churn**
* Improving **customer lifetime value (LTV)**
* Enhancing **cross-selling** and **up-selling**
* Increasing **customer satisfaction** and **engagement**

This understanding sets the foundation for relevant analysis.

**Step 2 - Data Exploration and Preparation**

**a. Schema Understanding:**

* Identify primary keys (like CustomerID, Surname)
* Understand column types (categorical vs numerical)
* Detect relationships for joining (e.g., CustomerID across tables)

**b. Data Cleaning:**

* Check for **null values**, **inconsistent formatting**, **duplicates**
* Convert categorical columns to consistent labels
* Handle outliers (e.g., very high balances or salaries)

**c. Data Merging:**

* Create a master table by merging all datasets on CustomerID using SQL or Power Query

SELECT \* FROM Bank\_Churn

LEFT JOIN CustomerInfo USING(CustomerID)

LEFT JOIN CreditCard USING(CustomerID)

LEFT JOIN Gender USING(GenderID)

LEFT JOIN Geography USING(GeographyID);

**Step 3 – Exploratory Data Analysis**

This is where the real story starts to unfold. I would use SQL for summaries and Power BI for visual trends:

**a. Customer Profile Analysis:**

* Age distribution, gender ratio, geography spread
* Average balance, credit score, salary per group

**b. Churn Analysis:**

* % of customers who have exited vs active
* Churn rate by age group, region, gender, credit score
* Identify any unexpected trends (e.g., customers with high salary still exiting)

**c. Behavioral Insights:**

* Product usage pattern (NumOfProducts vs Exited)
* IsActiveMember vs Exited
* Tenure vs Churn
* Credit Card usage vs Churn

**d. Time-Based Patterns:**

If there’s a timestamp (e.g., JoinDate), analyze:

* Monthly new customers
* Seasonal churn trends

**Step 4 - Visualization (Power BI Dashboard)**

Create an interactive dashboard with:

* KPIs: Total Customers, Active vs Exited, Churn %
* Filters: Region, Age Group, Gender
* Charts:
  + Bar chart for churn by product usage
  + Pie chart for credit card holders
  + Map for regional churn
  + Time series for customer acquisition

**Step 5 – Recommendations**

Based on analysis, provide insights:

* Customers with 3+ products and low tenure are likely to exit → focus on onboarding
* Regions with high churn need regional offers or branch-level engagement
* Encourage credit card usage (if it shows correlation with retention)

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

ALTER TABLE bank\_churn RENAME COLUMN HasCrCard TO Has\_creditcard;