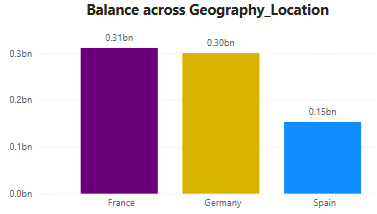
**Objective Questions**

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

The chart reveals variations in the distribution of account balances across the three regions:

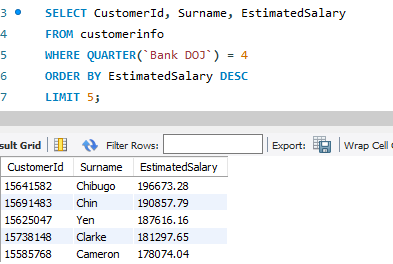
* **France** (Region A) shows the **highest concentration of accounts with larger balances**, contributing approximately **0.31 billion** in total. This suggests a strong and possibly high-value customer base in this region.
* **Germany** (Region B) has a **slightly lower total balance at 0.30 billion**, but the distribution appears **more spread out**, indicating a diverse mix of both high- and mid-value accounts.
* **Spain** (Region C) reflects the **lowest distribution of account balances at 0.15 billion**, suggesting either fewer customers or generally smaller balances per account.

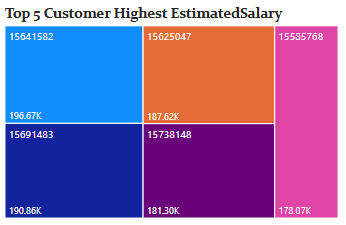
These differences may guide regional strategies—such as focusing on high-value retention in France, diversified offerings in Germany, and customer growth or engagement strategies in Spain.



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

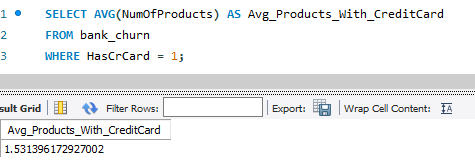
* **SELECT CustomerId, Surname, EstimatedSalary** Retrieves these three columns for each customer from the customerinfo table.
* **WHERE QUARTER(\Bank DOJ`) = 4`** Filters the records to include **only those customers who joined the bank in the 4th quarter** of the year — i.e., in the months **October, November, or December**.
* **ORDER BY EstimatedSalary DESC** Sorts the filtered customers in **descending order of their Estimated Salary**, meaning the **highest salaries come first**.
* **LIMIT 5** Returns **only the top 5 customers** from the sorted list — in this case, the **top 5 highest-earning customers who joined in Q4**.





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

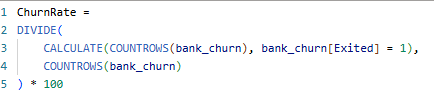
* AVG(NumOfProducts) calculates the **average number of products** used.
* WHERE HasCreditCard = 1 filters the data to include **only customers who have a credit card**.  
  The result will return **one value**: the **average number of products** used by those customers.



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

This query returns the most recent year present in the Bank DOJ column of the customerinfo table

****

* COUNTROWS(bank\_churn)  
   This counts the total number of customer records in the bank\_churn table — representing all customers in the dataset.
* CALCULATE(COUNTROWS(bank\_churn), bank\_churn[Exited] = 1) This part filters the table to only include customers who have exited (churned), i.e., where the Exited column equals 1. It then counts how many such customers there are — i.e., total **lost customers**.
* DIVIDE(... , ...)  
   The DIVIDE function safely divides the number of churned customers by the total number of customers. It also avoids errors if the denominator is zero.
* \* 100 This converts the result into a **percentage**, giving the **churn rate**.  
  ****
* The measure calculates the **percentage of customers who left the bank** (churned) out of the total number of customers.

****

****

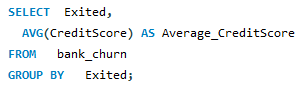
1. **Compare the average credit score of customers who have exited and those who remain. (SQL)**

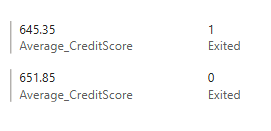
This query compares the **average credit score** of two groups of customers:

* Those who **left the bank** (Exited = 1)
* Those who **stayed** (Exited = 0)

It does this by:

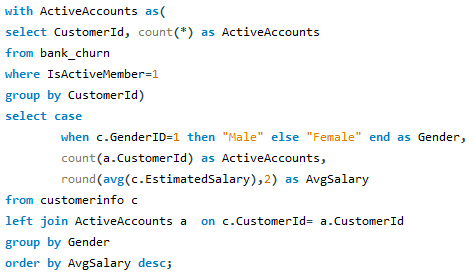
* Grouping the data based on the Exited column (0 or 1).
* Calculating the **average** CreditScore for each group.





1. **Which gender has a higher average estimated salary, and how does it relate to the number of active accounts? (SQL)**
2. **CTE (ActiveAccounts)**: Identifies customers who are active (IsActiveMember = 1) from the bank\_churn table.
3. **Main Query**:

* Joins customerinfo with the active customers list.
* Uses a CASE statement to convert GenderID to 'Male' or 'Female'.
* Counts how many **active accounts** exist for each gender.
* Calculates the **average estimated salary** for each gender.
* Orders results by average salary in descending order.





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

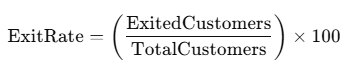
#### **Step 1: Credit Score Segmentation (CTE: CreditSegments)**

* This Common Table Expression (CTE) categorizes each customer into a **credit score segment** based on their CreditScore value.
* The CASE statement defines six segments:

1. **Excellent** (≥ 800)
2. **Very Good** (740–799)
3. **Good** (670–739)
4. **Fair** (580–669)
5. **Poor** (500–579)
6. **Very Poor** (< 500)

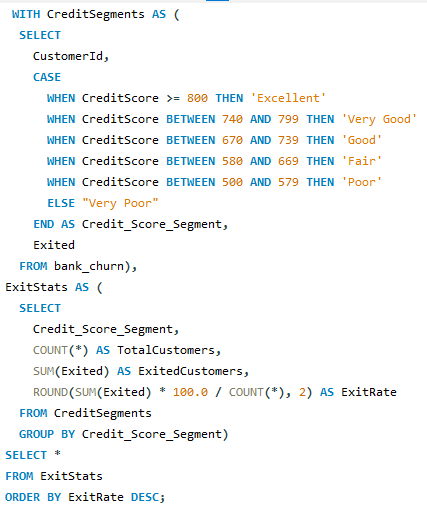
Each customer is labeled with the appropriate segment and their Exited status is retained.

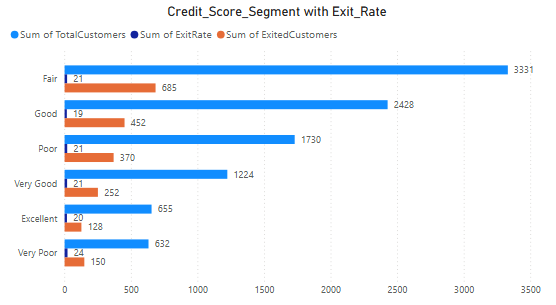
#### **Step 2: Churn Analysis by Segment (CTE: ExitStats)**

* Groups the data by Credit\_Score\_Segment
* For each segment, calculates:
  + TotalCustomers: total customers in that segment
  + ExitedCustomers: number of customers who exited (Exited = 1)
  + ExitRate: percentage of exited customers =  
     

#### **Step 3: Final Output**

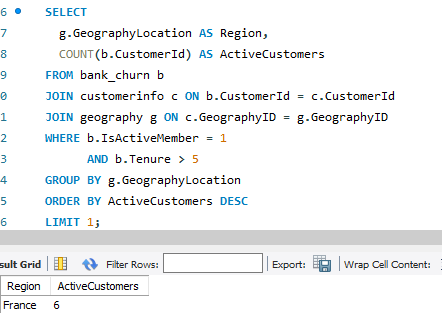
* Retrieves all credit score segments along with their total customer count, number of exited customers, and exit rate.
* Sorts the results in **descending order** of ExitRate, so the **segment with the highest churn risk appears first**.



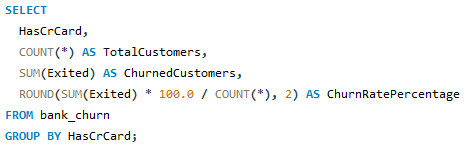


1. **Find out which geographic region has the highest number of active customers with a tenure greater than 5 years. (SQL)**

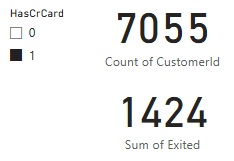
* **SELECT g.GeographyLocation AS Region** Retrieves the region name from the geography table.
* **COUNT(b.CustomerId) AS ActiveCustomers** Counts how many customers from the bank\_churn table satisfy the given conditions (i.e., they are active and have a tenure > 5).
* **FROM bank\_churn b** Starts with the main data table which contains customer churn information.
* **JOIN customerinfo c ON b.CustomerId = c.CustomerId** Connects churn data to customer personal details using CustomerId.
* **JOIN geography g ON c.GeographyID = g.GeographyID** Joins to the geography table to fetch region names.  
   *(This is an* ***INNER JOIN****, so if any GeographyID is missing or unmatched, the customer will be excluded.)*
* **WHERE b.IsActiveMember = 1 AND b.Tenure > 5** Filters to include only those customers who are **active** and have been with the bank for **more than 5 years**.
* **GROUP BY g.GeographyLocation** Groups the filtered customers by region.
* **ORDER BY ActiveCustomers DESC** Sorts the groups by the number of active customers in **descending order**.
* **LIMIT 1** Returns **only the top region** with the **most active long-term customers**.

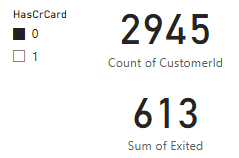


1. **What is the impact of having a credit card on customer churn, based on the available data?**

* **HasCrCard**: 1 = Has credit card, 0 = No credit card
* **TotalCustomers**: Total customers in each category
* **ChurnedCustomers**: Number of customers who exited
* **ChurnRatePercentage**: % of customers who exited in each category  
  

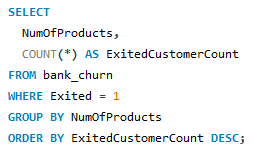
### **Insight**

* If the **churn rate** is **higher for customers without a credit card**, it may indicate that **credit card holders are more loyal** or better engaged with the bank.
* If churn is **higher among credit card holders**, it may signal dissatisfaction with card services or fees.



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

* **NumOfProducts**: The number of products a customer used (e.g., accounts, loans, credit cards).
* **Exited = 1**: Filters only customers who have exited (churned).
* **COUNT(\*)**: Counts how many exited customers had each product count.
* **ORDER BY ExitedCustomerCount DESC LIMIT 1**: Returns the product count with the highest number of exited customers.

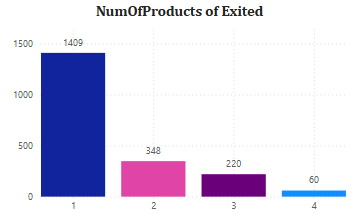
****

This query helps identify the **most frequent number of products** used by churned customers.  
 It can reveal patterns such as:

* Are **single-product** users more likely to churn?
* Do **multi-product** users also churn often?

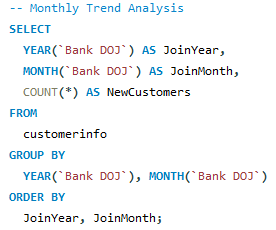
**Insight**:

The majority of exited customers used only **1 product**, which might suggest a **low engagement level** with the bank's services.

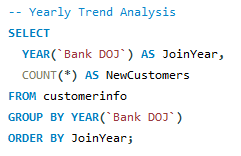
****

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.**

* **YEAR(Bank DOJ) and MONTH(Bank DOJ)**: These functions extract the **year** and **month** from the date when each customer joined the bank.
* **COUNT(\*) AS NewCustomers**: Counts how many customers joined in each **month of each year**.
* **GROUP BY YEAR(), MONTH()**: Groups the data by year and month so the count is calculated for each time period.
* **ORDER BY JoinYear, JoinMonth**: Sorts the result chronologically — from the **earliest to the latest month/year**.

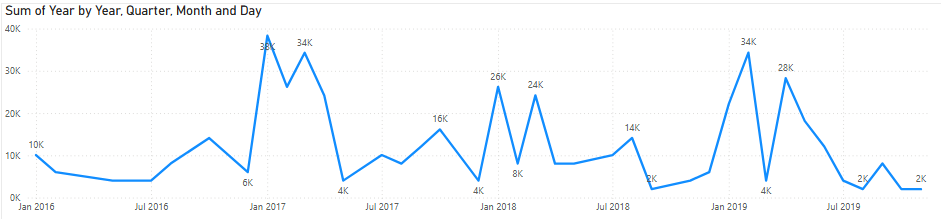
****

* The number of new customers who joined the bank in **each month** over time.
* This helps to identify **joining trends**, **seasonal spikes**, or any significant changes in customer acquisition.

****

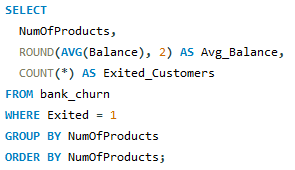
## **Insights:**

* **Seasonal trends**: Are there specific months (e.g., April or December) with spikes in new customers?
* **Year-on-year growth**: Is customer acquisition increasing or declining?
* **Business events alignment**: Spikes may align with campaigns, new product launches, etc.

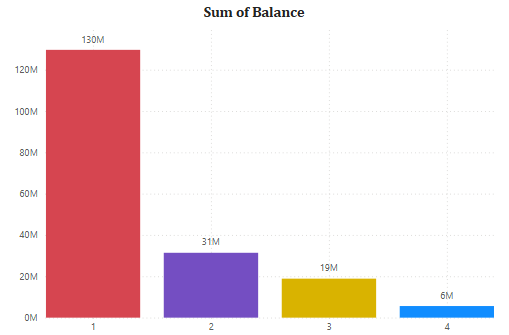


1. **Analyze the relationship between the number of products and the account balance for customers who have exited.**

* **WHERE Exited = 1** filters only those customers who have left (churned).
* **GROUP BY NumOfProducts** groups the data based on the number of products a customer held.
* **AVG(Balance)** shows the average balance for each group.
* **COUNT(\*)** shows how many exited customers had that product count.

****

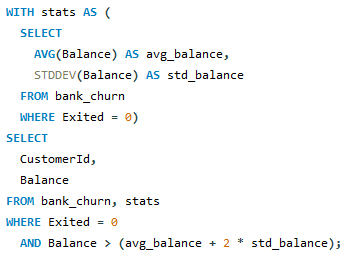
### **Insights:**

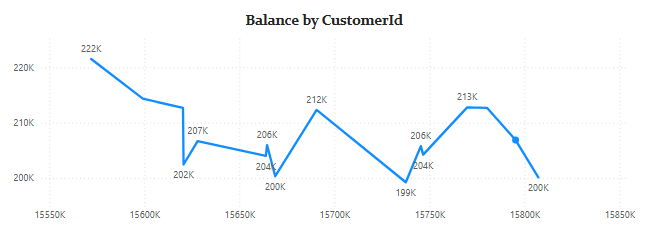
* **Trend Analysis**: Does having more products correlate with higher balances for exited customers?
* **Churn Profile**: Are high-balance customers with fewer products more likely to leave?
* **Retention Strategy**: Consider targeting customers with 1 or 2 products and high balances for cross-selling and retention efforts.  
  

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

* **CTE stats**: Computes the average and standard deviation of balances for customers who have not exited (Exited = 0).
* **Main Query**:  
   A. Filters for non-exited customers (Exited = 0)

B. Returns only those with Balance > avg + 2 \* std (i.e., statistical outliers)



****

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

### **Total Number of Tables:**

There are **seven tables** present in the dataset:

1. ActiveCustomer
2. Bank\_Churn
3. CreditCard
4. CustomerInfo
5. ExitCustomer
6. Gender
7. Geography

### **Tables Containing Only Categorical Variables:**

### The following tables consist entirely of categorical variables, which represent descriptive or classification-type data:

* **Gender** – contains the column GenderCategory (e.g., Male, Female).
* **Geography** – includes GeographyLocation (e.g., France, Germany, Spain).
* **ExitCustomer** – has ExitCategory (e.g., Exit, Retain).
* **ActiveCustomer** – features ActiveCategory (e.g., Active Member, Inactive Member).
* **CreditCard** – includes Category (e.g., Credit-card holder, Non-Credit card holder).

### **Partially Categorical Table:**

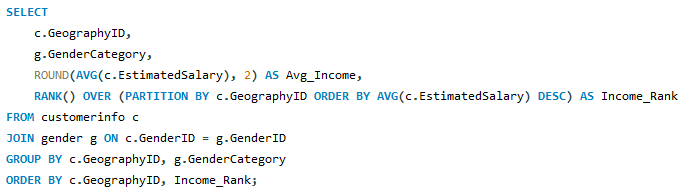
* **CustomerInfo** – while it contains numeric data (like Age, EstimatedSalary), it also includes categorical columns such as Surname, GenderID, and GeographyID (which are foreign keys referring to other categorical tables).

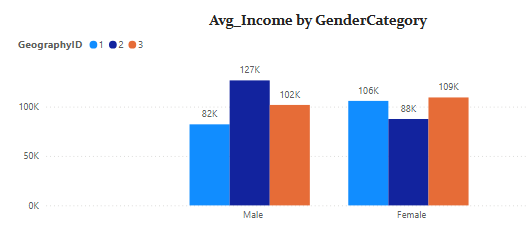
### 

### **Non-Categorical Table:**

* **Bank\_Churn** – primarily contains numerical and binary columns (like CreditScore, Balance, IsActiveMember, Exited), and is **not considered fully categorical**.

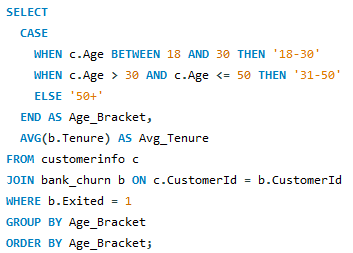
1. **Using SQL, write a query to find out the gender-wise average income of males and females in each geography id. Also, rank the gender according to the average value. (SQL)**

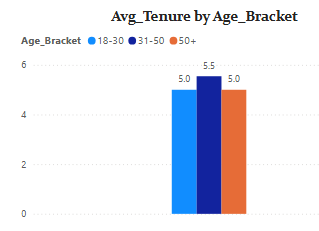
* **AVG(c.EstimatedSalary)**: Calculates the average salary per gender in each geography.
* **RANK() OVER (PARTITION BY GeographyID ORDER BY AVG(...) DESC)**: Ranks the genders **within each geography** based on their average income (highest gets rank 1).
* **JOIN gender g**: Retrieves the gender names using GenderID.
* **GROUP BY GeographyID, GenderCategory**: Groups data for aggregation by geography and gender.  
  ****

****

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+).**

* **CASE**: Categorizes each customer into one of the age brackets.
* **AVG(b.Tenure)**: Calculates the average tenure for each group.
* **WHERE b.Exited = 1**: Filters only customers who have exited.
* **JOIN**: Combines customerinfo and bank\_churn tables using CustomerId.

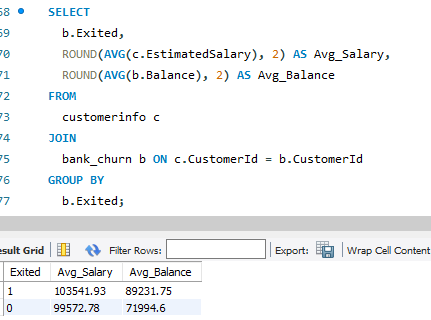
****

****

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?**

If balance increases significantly with salary **only for exited customers**, it might suggest a correlation in that group.

If both groups show similar trends or very little change, the correlation is weak or non-existent.

****

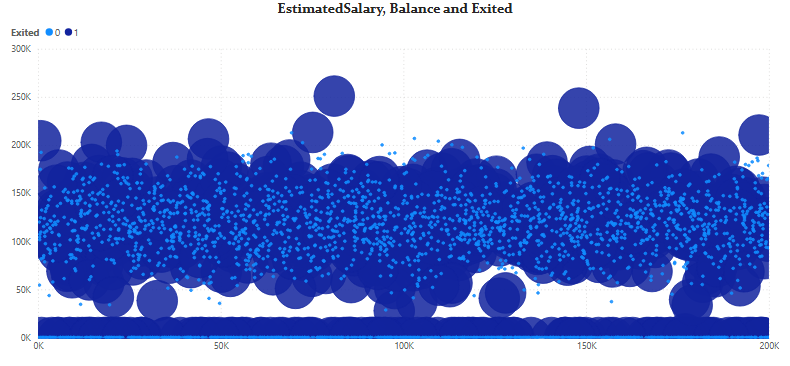
1. **Exited Customers:**
   * Have a slightly **higher average salary** than those who stayed.
   * Also have a **significantly higher average balance**.
2. **Not Exited Customers:**
   * Have **lower salary** and **lower balance** on average.

### **Conclusion:**

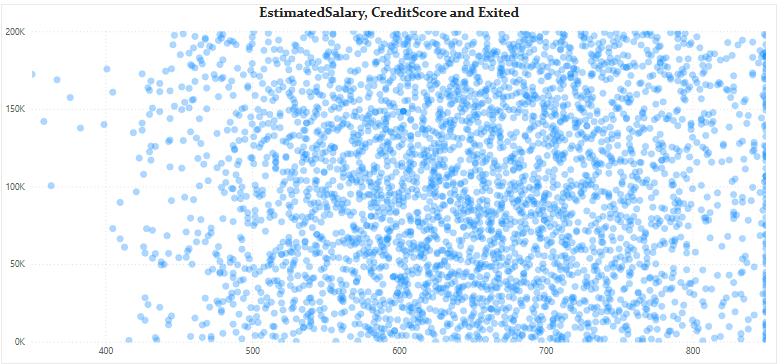
* There seems to be a **positive relationship** between **salary and balance**—as salary increases, so does balance.
* Customers who **churned (Exited = 1)** generally had **more money in their accounts**, possibly indicating:
  + They may not be satisfied despite financial stability.
  + High-balance customers might be more likely to switch banks for better services/rates.

**Insight:**

**Yes, there appears to be a direct correlation between salary and balance**, and this correlation is **stronger among customers who exited**, as both their average salary and balance are higher than those who stayed.

****

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

* There is **no meaningful linear relationship** between a customer's salary and their credit score.
* In other words, **having a higher or lower salary does not significantly affect** a customer's credit score in your dataset.  
  ****

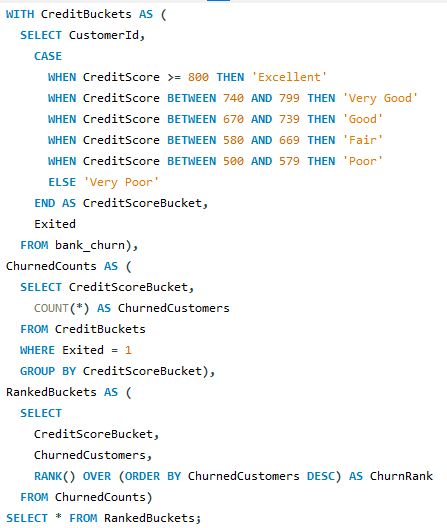
### **Insight:**

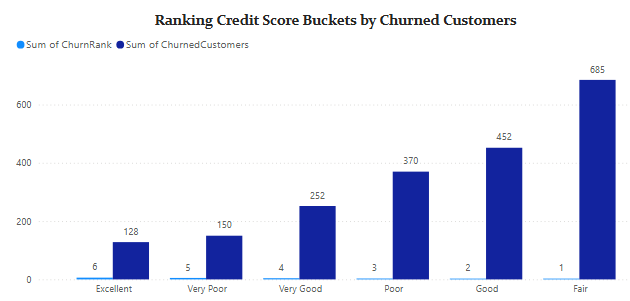
### This insight could be important for business teams:

* It suggests that **creditworthiness (credit score)** is influenced by other factors (e.g., payment history, debt levels), **not income**.
* So, segmentation or predictions should **not rely solely on salary** when assessing risk or designing marketing for credit products.

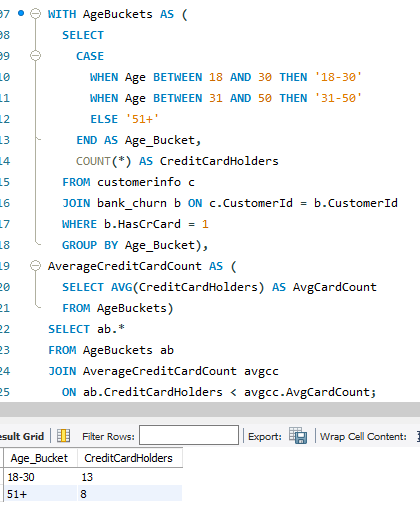
1. **Rank each bucket of credit score as per the number of customers who have churned the bank.**

* **CTE 1 (CreditBuckets):** Categorizes each customer into a credit score bucket.
* **CTE 2 (ChurnedCounts):** Counts churned customers (Exited = 1) in each bucket.
* **CTE 3 (RankedBuckets):** Ranks the buckets from highest to lowest churn using RANK().

****

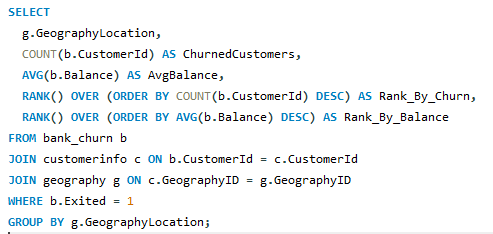
****

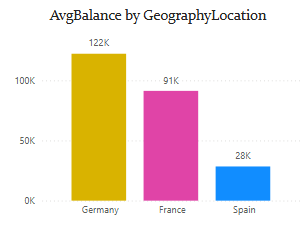
1. **According to the age buckets find the number of customers who have a credit card. Also retrieve those buckets that have lesser than average number of credit cards per bucket.**

* **AgeBuckets CTE**: Groups customers into age ranges and counts how many have credit cards.
* **AverageCreditCardCount CTE**: Calculates the average number of credit card holders across the age buckets.
* **Final SELECT**: Filters only those age buckets where the count is **less than the average**.  
  ****

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

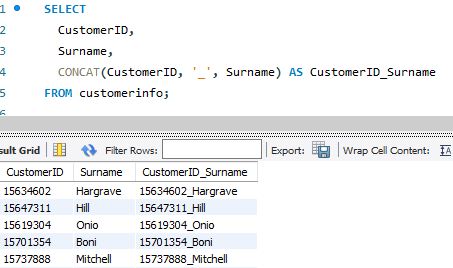
* COUNT(b.CustomerId) — gives the number of churned customers per location.
* AVG(b.Balance) — calculates the average balance of those churned customers.
* RANK() OVER (...) — assigns ranks:
* Rank\_By\_Churn ranks locations by number of churned customers (higher = better rank).
* Rank\_By\_Balance ranks by average balance (higher = better rank).
* WHERE b.Exited = 1 ensures you only consider churned customers.

****

****

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”.**

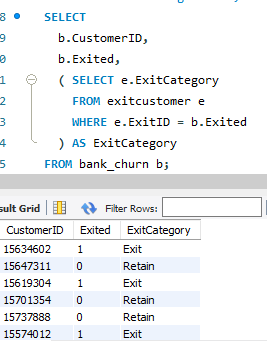
In the query below, we are combining the CustomerID and Surname columns from the customerinfo table into a **single new column** called CustomerID\_Surname.  
 We use the CONCAT() function to join them with an underscore \_ in between.



* This column (CustomerID\_Surname) can now be used to perform joins with any table where a **composite key** exists in the same format.

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

* Yes, we can get the ExitCategory from the ExitCustomer table into the Bank\_Churn table **without using JOIN** by using a **subquery** (specifically, a scalar subquery or a correlated subquery depending on context).
* For each row in bank\_churn, the subquery finds the matching ExitCategory from exitcustomer **based on Exited = ExitID**.
* This avoids the use of an explicit JOIN, yet still fetches the related value.

****

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 data analysis, missing values often occur when certain data points are not recorded or are unavailable. Fortunately, in our case, **the dataset is complete and free of missing values**, which simplifies the analysis and eliminates the need for data imputation — a process that can sometimes introduce assumptions or bias.

#### 

#### **General Approaches to Handle Missing Values (For Future Reference)**

Although our current dataset does not require it, understanding standard practices for handling missing data is essential:

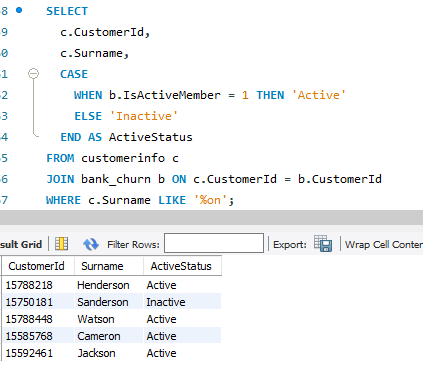
* **Deletion**: Remove rows or columns with missing values. This method is suitable when the proportion of missing data is small and unlikely to affect the overall outcome. However, it may result in the loss of valuable information if overused.
* **Imputation**: Fill in missing data using estimated values. Common methods include:
  + **Mean, Median, or Mode Imputation**: Simple and effective for numeric or categorical data.
  + **k-Nearest Neighbors (KNN)**: Uses the values of similar records to estimate missing entries.
  + **Advanced Statistical Methods**: Such as regression imputation or multiple imputation, based on the data’s structure.
* **Modeling Techniques**: Some machine learning algorithms (e.g., decision trees, XGBoost) can handle missing values inherently. Even then, understanding the pattern or cause of missingness is crucial to avoid misleading results.

**Tool-Specific Handling**:

* **Excel**: Use functions like IF, ISBLANK, or IFERROR to detect and handle blanks. You can also apply **filtering** and **data validation** to manage missing cells, or fill them using **average formulas**.
* **SQL**: Use IS NULL or COALESCE() to identify or replace null values. Example:  
  
* **Power BI**: Through **Power Query**, you can replace missing values using "Replace Values," "Fill Down," or conditional logic. Missing values can also be visualized and flagged using DAX functions like ISBLANK() or IF.

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”.**

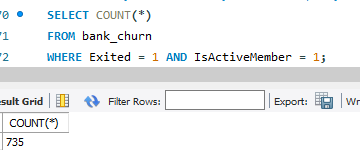
* LIKE '%on': Filters surnames that **end with 'on'**.
* CASE WHEN ...: Translates IsActiveMember = 1 into "Active" and 0 into "Inactive" for easier readability.
* JOIN: Combines customer information with their account status from bank\_churn.

****

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 **is a noticeable data discrepancy** in the customer's data—specifically between the IsActiveMember and Exited columns.

### **Discrepancy Observed:**

* **Definition**:
  + IsActiveMember = 1 means the customer is still actively using the bank's services
  + Exited = 1 means the customer has **left** (churned) the bank.
* **Discrepancy**:
  + Ideally, if Exited = 1, then IsActiveMember **should be 0**, because a customer who exited can't be an active member anymore.
  + However, upon analysis of the data (possibly by running a query like the one below), we may find some customers where:****
  + This returns **non-zero rows**, indicating **customers who are marked as exited but still shown as active**, which is a **data inconsistency**.

### **Why It Matters:**

* Since it's mentioned that the Exited column is **accurate**, the issue likely lies in the IsActiveMember column — possibly due to:
  + Delayed data sync
  + Poor update logic during churn processing
  + Manual data entry errors

### **How to Handle:**

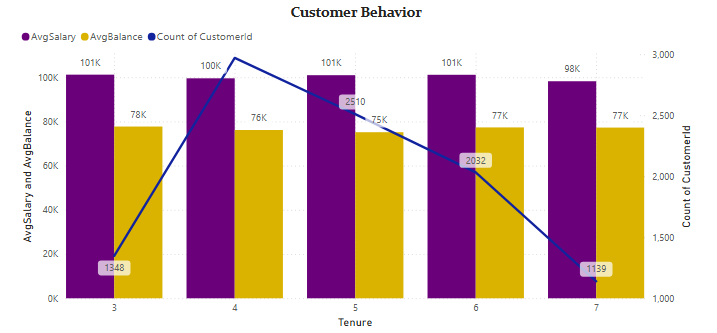
* During data cleaning or transformation:
  + Set IsActiveMember = 0 where Exited = 1, to ensure consistency.
  + Or flag these cases for investigation if real-time auditing is important.

**Subjective Question:**

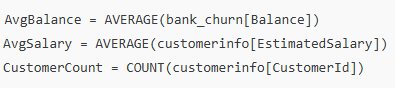
1. **Customer Behavior Analysis: What patterns can be observed in the spending habits of long-term customers compared to new customers, and what might these patterns suggest about customer loyalty?**

#### **Insights & Observations:**

* **Higher Balances Among Long-Term Customers:**
  + Long-tenure customers (e.g., tenure > 5 years) often maintain **higher average account balances** compared to newer customers.
  + This suggests stronger financial engagement and deeper relationships with the bank over time.
* **More Products Used by Long-Term Customers:**
  + Long-term customers tend to use **multiple products** (e.g., savings + loans + credit cards), indicating trust and loyalty.
  + New customers are more likely to start with 1 or 2 products.
* **Lower Churn Rate in Mid-Tenure Customers:**
  + Customers in the **3–6 year range** show the lowest churn rate, possibly because they’ve had time to explore services but haven't yet reached a point of dissatisfaction or external switch triggers.
* **Higher Credit Card Ownership:**
  + Long-term customers are more likely to own a **credit card**, and use it regularly—an indicator of ongoing activity and loyalty.
* **Spending Behavior & Estimated Salary:**
  + Customers with **moderate to high salaries** who’ve stayed longer show **more stable balances and credit behavior**, while new high-income customers tend to keep lower balances, possibly shopping for better services.



I prepared this chart using DAX measures to calculate the **average balance**, **average estimated salary**, and the **count of customers** grouped by tenure. These measures were then visualized using a **Line and Clustered Column Chart** in Power BI to compare customer spending habits across different tenure periods.



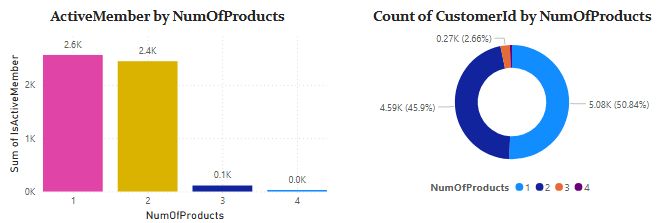
### **What These Patterns Suggest About Loyalty:**

* Loyalty grows as customers **experience multiple bank services** and build financial habits tied to the institution.
* Customers who have stayed longer tend to trust the bank more, leading to **higher product usage** and **deeper integration into financial services**.
* New customers need more **engagement and onboarding support** to evolve into loyal users.

### **Recommendations:**

* **Strengthen Onboarding Programs:**
  + Educate new customers about available products and encourage multi-product adoption early in their journey.
* **Reward Loyalty:**
  + Launch **loyalty programs** or **tenure-based incentives** (e.g., fee waivers, better rates) for long-term customers.
* **Monitor Early Churn Risk:**
  + Analyze churn trends in customers within the **first 1–2 years** and intervene with retention strategies.
* **Cross-Sell Strategically:**
  + Use transaction and product usage data to **recommend relevant services** to long-term customers (e.g., investment options, higher-tier cards).
* **Personalize Communication:**
  + Leverage CRM tools to **segment customers by tenure and behavior** for targeted messaging and offers.

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

****

#### **Insights:**

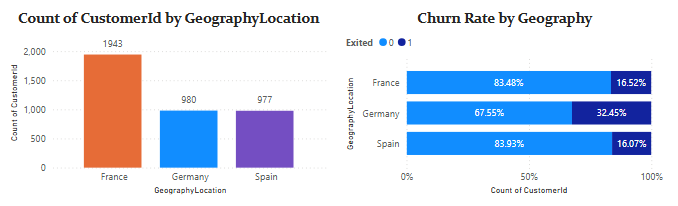
1. **High Affinity Between Credit Card and Multiple Products:**
   * Customers who hold a **credit card** are more likely to have **2 or more products**, such as savings accounts and personal loans.
   * This indicates a strong product bundling tendency.
2. **Active Members Use More Products:**
   * **Active customers** show a higher average number of products used, suggesting engagement drives broader product adoption.
3. **Churn Risk and Product Count:**
   * Customers using **only 1 product** are more likely to churn.
   * Multi-product users (especially 3+) are less likely to exit, highlighting the retention power of product bundling.
4. **Credit Card + High Balance = Loyal Segment:**
   * Customers with **credit cards and high balances** tend to stay longer with the bank and can be targeted for premium services.

### **Recommendations:**

1. **Bundle Products Based on Usage Patterns:**
   * Offer **combo plans** like savings + credit card, or loan + insurance for customers already using one of them.
2. **Upsell to Single-Product Customers:**
   * Identify customers with only 1 product and recommend add-ons using targeted offers and communication.
3. **Use Lifecycle Triggers:**
   * Introduce additional products at specific tenure stages (e.g., credit card after 6 months, investment after 1 year).
4. **Personalized Offers for Active Segments:**
   * Target high-balance, active customers with **exclusive product packages**, loyalty benefits, or upgrade options.
5. **Monitor Product Drop Patterns:**
   * Track customers who discontinue a product—it could be an early sign of churn. Use re-engagement campaigns here.
6. **Geographic Market Trends: How do economic indicators in different geographic regions correlate with the number of active accounts and customer churn rates?**

#### **Insights:**

1. **Germany:**
   * Highest number of customers who have exited the bank.
   * Despite a **high average balance**, churn rate is relatively elevated.
   * Suggests that **financial capacity doesn’t guarantee loyalty**, possibly due to service dissatisfaction or better alternatives in the market.
2. **France:**
   * Shows a **balanced churn rate** with a strong number of **active customers**.
   * Indicates **steady engagement**, possibly due to stable customer service and product alignment with local needs.
3. **Spain:**
   * Lowest churn rate and fewer active accounts.
   * May indicate a **smaller customer base**, but with **better retention**, suggesting strong trust or fewer banking options in the region.



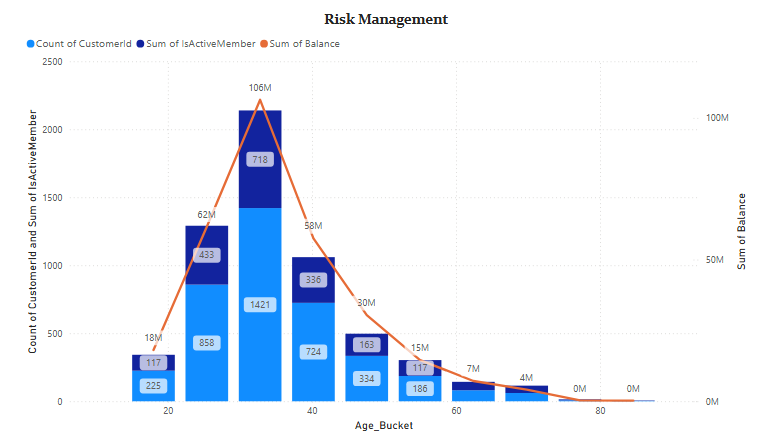
### **Economic Correlation (Assumptive View):**

* **High Balance ≠ Low Churn**: In regions like Germany, customers have high account balances but still exit frequently — economic strength doesn’t always translate to satisfaction.
* **Active Membership Trends**: Regions with **lower economic activity** may show fewer active accounts but better retention (e.g., Spain), reflecting a **value-sensitive, loyal customer base**.
* **Estimated Salary Factor**: Regions with higher average salary (like France) may indicate higher product uptake and customer stickiness.

### **Recommendations:**

* **Germany**: Focus on **service quality improvements** and **loyalty campaigns** to reduce churn in high-value customers.
* **France**: Expand offerings and cross-sell opportunities; the region shows signs of **healthy engagement**.
* **Spain**: Invest in **market expansion strategies** to grow the active base while maintaining high retention.

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

****

### **Key Observations from the Chart:**

* **Highest Balance Exposure**:
  + Customers aged **30–40** hold the **highest total balance (~106M)**, followed by **20–30** (~62M).
  + This means the financial **liability for the bank is highest** in these age groups.
* **Inactive Members Present**:
  + Even though there are a **large number of active members**, a significant number of **inactive members** exist in the **20–40** age buckets.
  + Inactive members with high balances **increase risk** since they’re less engaged and more likely to churn or default.
* **Declining Engagement with Age**:
  + As the age increases beyond 50, both balance and customer count drop significantly.
  + These groups pose **lower financial risk** due to **lower engagement and lower balance**.

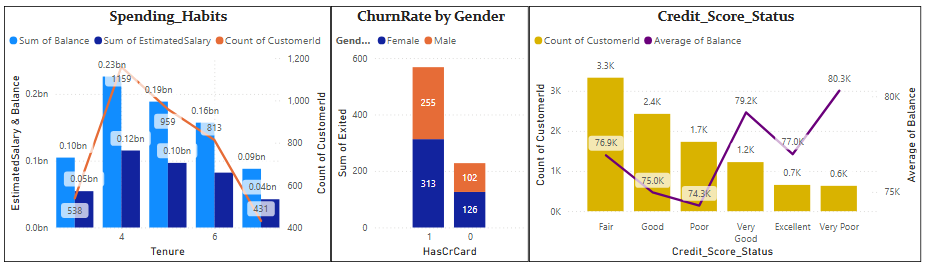
### **Insights & Recommendations:**

* **Young Adults (20–40)** are **high-value but high-risk** due to their high balances and partial disengagement.
* The bank should:
  + Improve **engagement strategies** for **younger inactive customers**.
  + Offer **personalized services, loyalty programs, or financial planning tools**.
  + **Monitor account activity** and set up alerts for sudden inactivity in high-balance accounts.

### **Conclusion:**

The **20–40 age group** poses the **greatest financial risk** due to their **high account balances** combined with a **noticeable proportion of inactive members**. Retention strategies focused on this group could significantly reduce financial risk exposure.

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?**

****

### **Spending Habits (by Tenure)**

**Insights:**

* Customers with **tenure = 4 years** have the highest **count (1.1K)** and **spending power** (balance ~0.23B, salary ~0.12B).
* After the 4th tenure year, both **customer count and spending (balance/salary)** decline sharply.

**Recommendations:**

* Focus retention efforts on customers with **4–6 years of tenure**, as they show strong engagement.
* Investigate reasons for decline in spending and count post 4 years and create **rewards or loyalty programs** for long-term retention.

### **Churn Rate by Gender and Credit Card Ownership**

**Insights:**

* **Females (313)** without a credit card have churned more than males (102).
* Those who **don’t have credit cards** churn significantly more (568 total) than those who do (228).

**Recommendations:**

* Promote **credit card benefits** to reduce churn.
* Target **female customers** without credit cards with tailored offers and financial education to improve retention.

### **Credit Score Status Analysis**

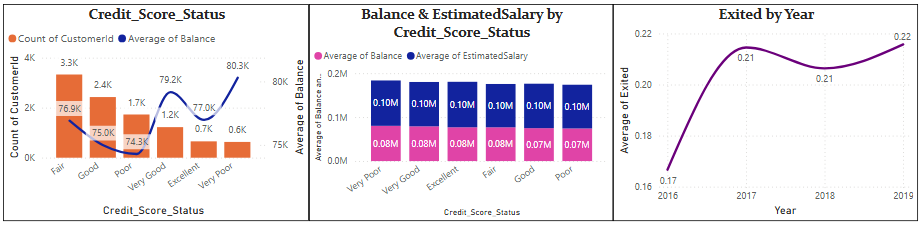
**Insights:**

* Most customers have a **Fair** (3.3K) or **Good** (2.4K) credit score.
* Interestingly, **balance increases** as we move from **Poor → Excellent**, peaking at **Very Poor (80.3K)**, suggesting no direct linear correlation.

**Recommendations:**

* For **Fair & Good** segments (majority group), promote cross-sell products (e.g., loans, savings schemes).
* Customers with **Very Poor credit scores** hold high balances—investigate further to understand behavior and **mitigate potential risk**.

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?**

****

#### **1. Credit Score & Customer Behavior**

* Most customers fall into the **Fair** and **Good** credit score categories (Chart 1).
* However, customers with **Very Good** and **Excellent** credit scores tend to have **higher average balances**.
* This indicates that **high-value customers** are not necessarily the majority, and marketing efforts should focus on nurturing customers with lower credit scores who still maintain strong balances.

#### **2. Balance & Salary Consistency**

* From Chart 2, the **average balance and estimated salary** are fairly uniform across credit score groups.
* This suggests that there might be limited variation in financial behavior, making **segmentation strategies** vital for personalized marketing.

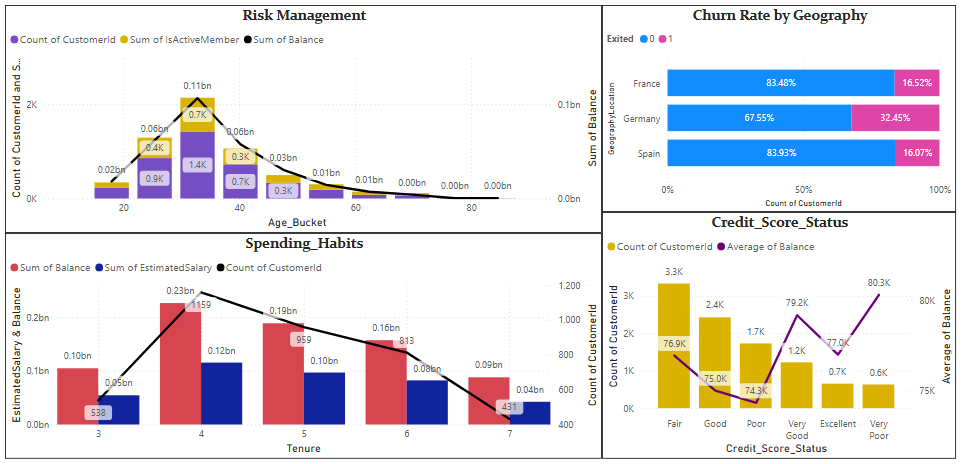
#### **3. Churn Trend Over Time**

* In Chart 3, the **exit rate has slightly increased** from 2016 to 2019.
* This trend indicates a need to investigate **what happened during this period**, such as:
  + Were new marketing campaigns ineffective?
  + Was customer service deteriorating?
  + Did competitor offerings become more attractive?

**Recommendation:**

* Focus campaigns on **Fair to Good** credit score groups (high volume).
* Design **premium loyalty programs** for **Excellent & Very Good** score holders (high balance).
* Dive deeper into **2017 churn** to understand what triggered high exits.
* Consider **bringing in campaign-specific columns** to directly measure response.

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

****

### **1. Geography-Wise Exit Trends (Based on Churn Rate Chart)**

* **Germany** shows the **highest churn rate (32.45%)**, compared to France (16.5%) and Spain (16.07%).
* This suggests **region-specific issues** such as service dissatisfaction, competition, or unmet expectations in Germany.  
   **Insight:** Germany may require **regionally targeted retention strategies**.

### **2. Activity Level (Based on Risk Management & Churn Charts)**

* A **high proportion of inactive members** were observed in the **0–20** and **20–40** age buckets.
* Inactivity often precedes exit, indicating **low engagement** or dissatisfaction.  
  **Insight:** Inactive customers are more prone to exit; need **proactive engagement**.

### **3. Tenure-Based Exit Pattern (From Spending Habits Chart)**

* Customers with **tenure between 3 to 5 years** have the **highest spending (balance & salary)** and **customer count**.
* However, **after the 5th year**, customer count **drops significantly**.  
  **Insight:** There may be a **service fatigue or unmet expectations post-5 years**, prompting exits.

### **4. Credit Score Analysis (From Credit\_Score\_Status Chart)**

* Customers with **Fair** and **Poor** credit scores form the **bulk of the churned population**.
* Despite having sufficient balances, these customers may feel **underserved or unrecognized**, causing attrition.  
  **Insight:** There’s a disconnect between **customer financial strength** and **how they're served**.

**5. Balance and Age (From Risk Management Chart)**

* Younger customers (0–40) hold the **majority of the bank’s total balance**.
* ]However, many are **inactive or exited**, indicating the **risk of losing high-value customers early**.  
  **Insight:** Younger customers might be **less loyal** or **more service-demanding**.

### **Conclusion & Recommendations**

Customers who exited the bank often share common traits—they are primarily located in **Germany**, fall within the **0–5 years tenure** range, and are often **inactive members**. Many belong to the **Fair or Poor credit score** categories. Interestingly, a large portion are **younger customers** who also maintain **high account balances**. These patterns highlight the need for **targeted engagement and retention strategies**, especially for high-value, low-engagement segments.

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

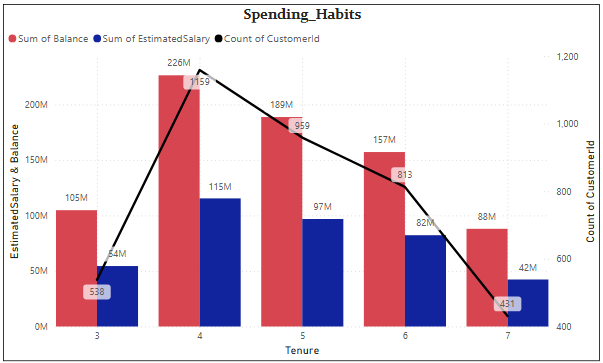
### **Tenure-Based Spending Habits – Insights**

* Customers with a **tenure of 3 to 5 years** show the **highest account balances** and **estimated salaries**, indicating their **peak financial engagement** with the bank.
* After **5 years**, there's a **noticeable drop in customer count** and spending, which may reflect **reduced engagement** or a **risk of exit**.
* Newer customers (0–2 years) have **lower balances**, which is typical as they’re still building trust and exploring services.
* This trend suggests that **3–5 year tenure** customers are the **most valuable segment** in terms of revenue potential.

### **Recommendation:**

Focus on **retention strategies** for customers approaching the 5-year mark, such as:

* Loyalty rewards,
* Personalized service offerings,
* Cross-selling additional products.

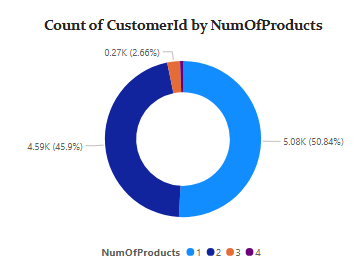
****

### **Insight: Importance of Number of Products in Churn Prediction**

* The **majority of customers (96%)** have **either 1 or 2 products**, indicating a preference for simpler banking relationships.
* Based on this chart alone, **NumOfProducts does not show a direct correlation with churn**, as there’s **no clear pattern** in product count and customer exits.
* While having **more products** could imply deeper engagement and **potentially lower churn**, it could also introduce **complexity**, which may lead to **customer dissatisfaction** in some cases.

### **Recommendation:“Evaluate Product Engagement, Not Just Quantity”**

* Focus on **understanding the quality and usage** of each product rather than just the count.
* Introduce **personalized bundling strategies** that suit customer profiles.
* Monitor customer satisfaction regularly, especially for those with **3+ products**, to reduce complexity-driven churn.

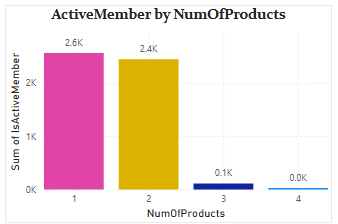
****

### **Active Members by Number of Products – Insights**

* Most **active members** hold **1 or 2 bank products**, indicating a standard level of engagement.
* The number of active members **decreases** as the number of products increases beyond 2, suggesting that **very few customers opt for multiple services**.
* Those with **3 or more products** are likely the bank’s most **engaged and loyal customers**, though they represent a smaller portion.

### **Recommendation:**

* Promote **product bundling** and **cross-sell campaigns** to encourage more product usage among currently active members.
* Identify and **reward customers with multiple products** to retain them longer.

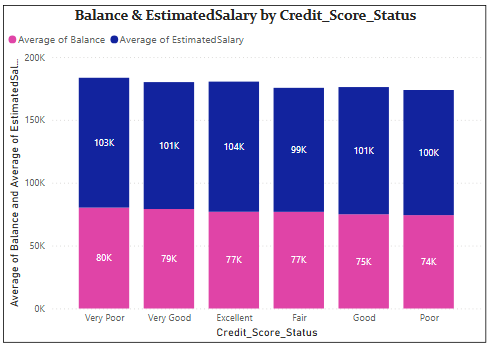
****

### **Credit Score vs Financial Strength – Insights**

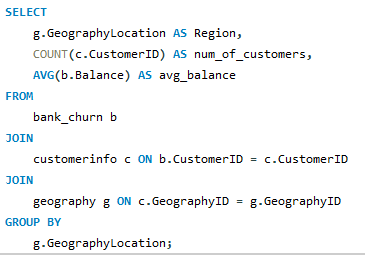
* Customers in **‘Very Good’** and **‘Excellent’** credit score segments show the **highest average balances and salaries**, reflecting strong financial health and potential for premium services.
* Interestingly, **‘Fair’ and ‘Good’** segments also maintain **moderate to high balances**, even with lower credit scores—indicating that **credit score alone may not determine financial strength**.
* **‘Poor’ and ‘Very Poor’** segments generally show **lower balances and salaries**, aligning with expected risk levels.

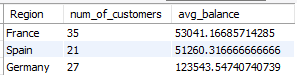
### **Recommendation: Smart Risk & Reward Targeting**

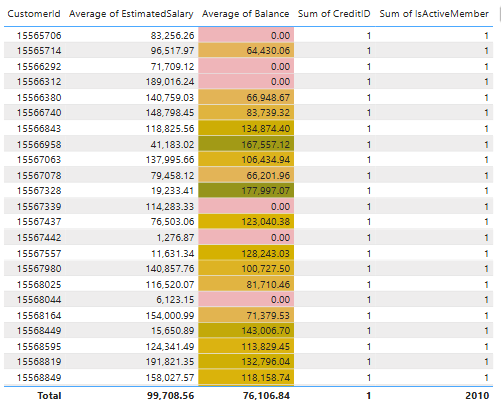
* Develop **tiered engagement strategies**:
  + Offer **premium services** to Very Good/Excellent segments.
  + Promote **credit-building products** to Fair/Good segments with high balances.

****

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



****

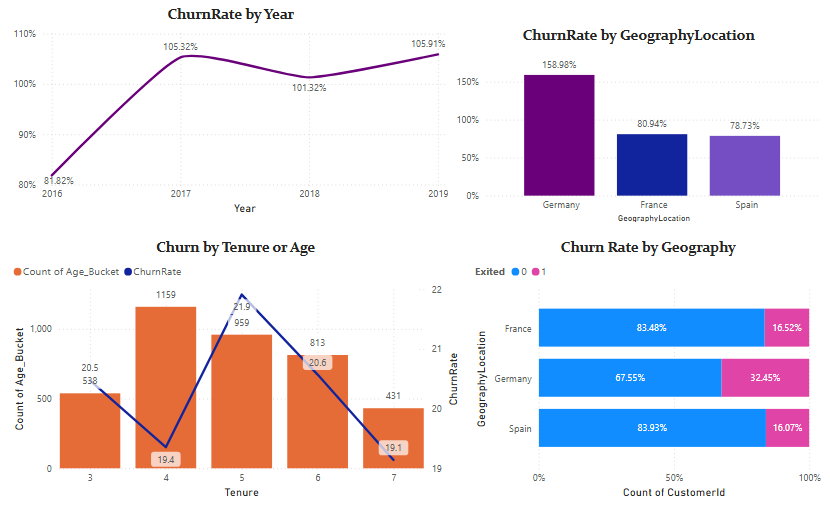
1. **How can we create a conditional formatting setup to visually highlight customers at risk of churn and to evaluate the impact of credit card rewards on customer retention?**

### **Insights:**

1. **Zero or Low Balance Customers Are at Higher Risk** Customers with a **balance of ₹0 or very low values** (e.g., ₹<10,000) are more likely to churn, especially if they have limited engagement (e.g., single product or inactive).
2. **Moderate Balance Range Needs Monitoring** Customers with balances between **₹40,000–₹80,000** are highlighted in **yellow**, indicating moderate risk. These customers may still churn if not offered proper incentives.
3. **High Balance Customers (> ₹1,00,000)** Customers with **higher balances** are visually safer and likely more engaged, but still need consistent engagement to maintain loyalty.
4. **All Customers Have Credit Cards** Since the Sum of CreditID is **1 for all rows**, this shows that **credit card ownership alone isn’t enough** to retain customers—other factors like account balance and activity levels matter.
5. **Inactive Customers Exist Despite High Balances** If Sum of IsActiveMember = 0 is found (not visible here but would be valuable), even high-balance customers might be at churn risk.

**Recommendations:**

1. **Target Zero/Low Balance Segments with Personalized Offers**
   * Introduce **minimum balance rewards**, cashback, or fee waivers.
   * Educate them about product benefits and encourage product bundling.
2. **Leverage Credit Card Insights**
   * Cross-reference credit card usage (if available) with churn risk.
   * Offer **tiered rewards** for credit card usage to enhance stickiness.
3. **Segment & Monitor Medium-Risk Customers**
   * Customers in the ₹50K–₹80K range should be included in **retention campaigns** like **exclusive offers**, **birthday rewards**, etc.
4. **Introduce Loyalty Scoring**
   * Build a **Customer Loyalty Score** combining balance, credit card ownership, activity level, and tenure.
5. **Alert System for Zero-Balance Holders**
   * Set up Power BI alerts for when balances drop to ₹0 or when multiple churn-risk flags are triggered.
6. **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?**

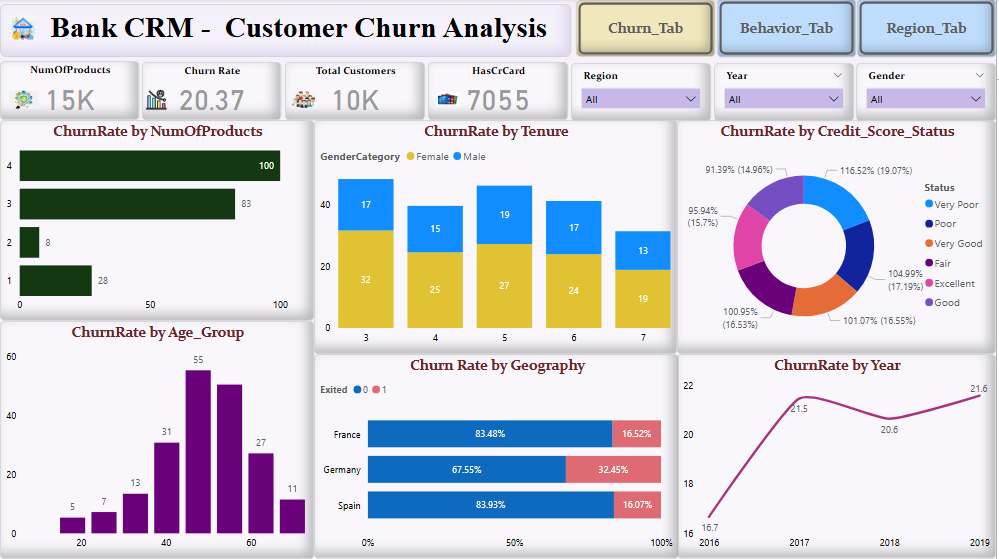


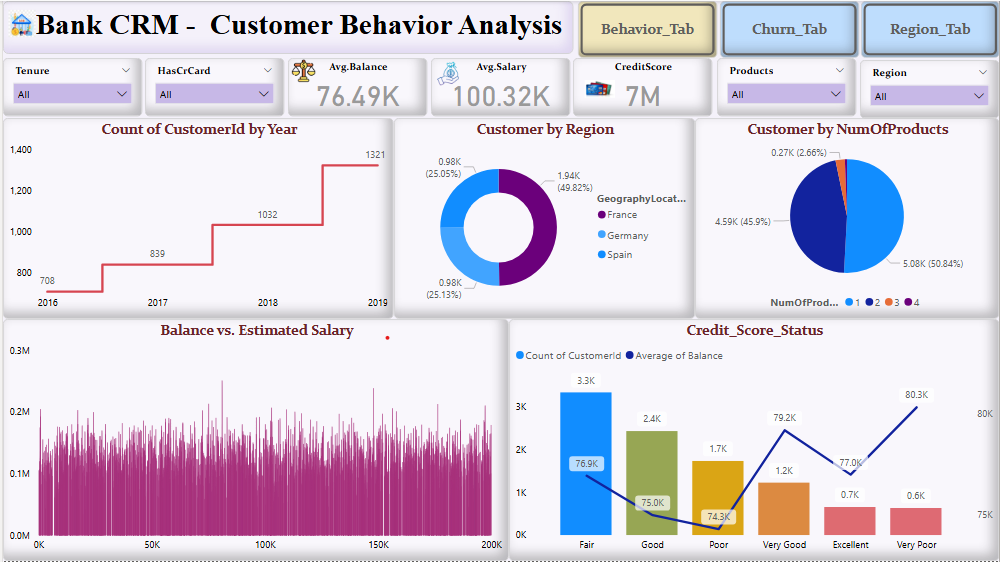
### **Insights**

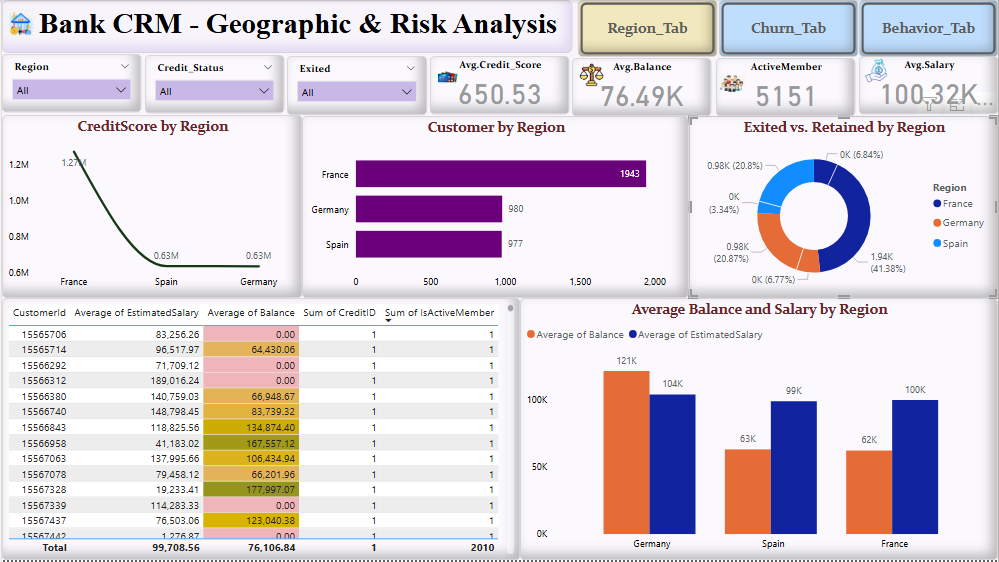
1. **Churn Rate by Year**
   1. The churn rate **increased sharply** from **81.82% in 2016** to **105.32% in 2017**, and remained **consistently high** through 2019.
   2. Indicates that **customer retention has worsened over time**, signaling possible service gaps or lack of engagement post-onboarding.
2. **Churn Rate by Geography**
   1. **Germany** has the **highest churn rate (158.98%)**, far exceeding France (80.94%) and Spain (78.73%).
   2. The **bar chart and 100% stacked chart** both highlight that **Germany alone contributes heavily to churn** — especially considering only **67.55% of German customers are retained**, compared to 83%+ for France and Spain.
3. **Churn by Tenure**
   1. The **highest churn rate (21.9%)** occurs for customers with a **tenure of 5 years**, even though the count of customers is relatively high.
   2. The churn rate **remains consistently above 19%** for most tenure groups, showing that **churn is not limited to new customers only**.
   3. **Early tenure groups (3–4 years)** also show **significant churn despite high customer count**, implying issues in **mid-term engagement**.

### **Overall Strategic Recommendations to Reduce Customer Churn:**

1. **Segmentation & Targeted Marketing:** Use customer segmentation based on churn risk indicators (e.g., geography, tenure, credit score) to design focused campaigns that address the specific concerns of high-risk groups.
2. **Regional Strategy Development:** Prioritize efforts in high-churn regions like **Germany** by enhancing customer service, localizing product offerings, and building partnerships to strengthen trust and satisfaction.
3. **Customer Engagement Programs:** Launch loyalty initiatives like personalized offers, reward points, and regular touchpoints to increase customer interaction and long-term retention.
4. **Credit Score Support:** Empower customers—especially those with **Fair or Poor credit scores**—by offering tools and education to improve financial health, which can foster loyalty and reduce churn.
5. **Create a dashboard incorporating all the KPIs and visualization-related metrics. Use a slicer in order to assist in selection in the dashboard.**

****

****

****

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

#### **1. Understand the Dataset**

* Identify all the tables and their relationships (e.g., bank\_churn, customerinfo, gender, geography, etc.).
* Check the meaning of each column (e.g., Exited, IsActiveMember, NumOfProducts, CreditScore, Balance, etc.).

#### **2. Data Cleaning & Preprocessing**

* Handle **missing values** (nulls, blanks).
* Create useful derived columns (e.g., Age groups, Tenure buckets, Credit score categories).
* Convert categorical variables using labels (e.g., GeographyID → Country).

#### **3. Descriptive Analysis**

* Analyze distributions:
  + Count of customers by age, gender, geography, etc.
  + Balance and salary by credit score.
* Compare churned vs non-churned customers:
  + Who exits more: Active or inactive members?
  + Which geography has higher churn?

#### **4. Key Metrics & KPIs**

* Overall **churn rate**.
* Churn rate by:
  + **Geography**
  + **Age group**
  + **Gender**
  + **Tenure**
  + **Product usage**
* Average **balance and salary** of churned vs retained customers.

#### **5. Visualization in Power BI**

Create dashboards to answer:

* Who are our loyal customers?
* What features are common in customers who churn?
* Which regions/products are performing well?

#### **6. Derive Insights & Recommendations**

* Based on patterns in the data, provide actionable insights.
* Recommend customer retention strategies, marketing focus, and operational changes.

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

