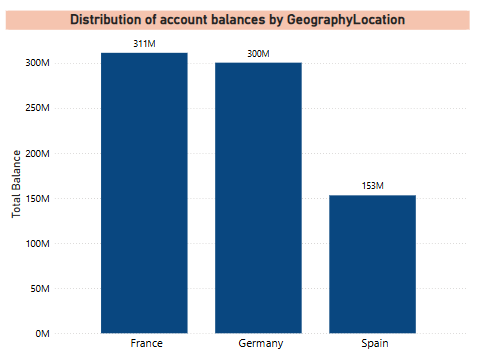
**Objective Answers**

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

**Ans.** The chart below shows the total account balances held by customers across different regions. This helps identify which geographic segments hold the largest share of deposits in the bank.

****

**Insights**

* France holds the highest total balance at ₹311M.
* Germany closely follows with ₹300M, while Spain lags behind at ₹153M.

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

**Ans.** To Identify the top 5 customers with the highest Estimated Salary in the last quarter of the year , I have used the following SQL query:

**SELECT**

**CustomerId,**

**EstimatedSalary,**

**Bank\_DOJ\_Date**

**FROM**

**customer\_info**

**WHERE**

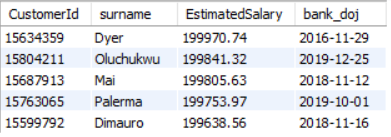
**MONTH(Bank\_DOJ\_Date) IN (10, 11, 12)**

**ORDER BY**

**EstimatedSalary DESC**

**LIMIT 5;**

**Result:**

****

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

**Ans.** To calculate the average number of products used by customers having a credit card , I have used the following SQL query:

**SELECT**

**AVG(NumOfProducts) AS Avg\_Products\_With\_CreditCard**

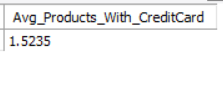
**FROM**

**Bank\_Churn**

**WHERE**

**HasCrCard = 1;**

**Result:**

****

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

**Ans.** To find the churn rate by gender for the most recent year , I have used the following SQL query:

**SELECT**

**g.gender,**

**COUNT(\*) AS total\_customers,**

**SUM(bc.Exited) AS churned\_customers,**

**ROUND(SUM(bc.Exited) / COUNT(\*) \* 100, 2) AS churn\_rate\_percent**

**FROM customer\_info ci**

**JOIN bank\_churn bc ON ci.CustomerId = bc.CustomerId**

**JOIN (**

**SELECT 1 AS GenderID, 'Male' AS gender**

**UNION**

**SELECT 2, 'Female'**

**) g ON ci.GenderID = g.GenderID**

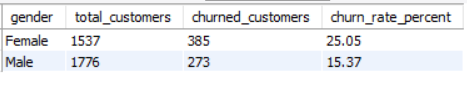
**WHERE YEAR(ci.Bank\_DOJ) = (**

**SELECT MAX(YEAR(Bank\_DOJ)) FROM customer\_info**

**)**

**GROUP BY g.gender;**

**Result:**

****

**Insights:**

* Female customers show a higher churn rate (25.05%) than males (15.37%), indicating a need for targeted retention efforts.
* Despite fewer in number, more females churned, highlighting a potential gap in engagement or satisfaction.

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

**Ans.** To compare the average credit score of customers who have existed and those who remain , I have used the below SQL query:

**SELECT**

**Exited,**

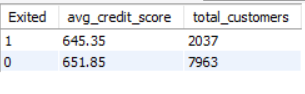
**ROUND(AVG(CreditScore), 2) AS avg\_credit\_score,**

**COUNT(\*) AS total\_customers**

**FROM bank\_churn**

**GROUP BY Exited;**

**Result:**

****

**Insights:**

* Exited customers have a lower average credit score (645.35) compared to those who stayed (651.85).
* Slightly lower credit scores may be linked to higher churn risk.

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

**Ans.**  To find which gender has a higher average estimated salary and how it is related to the number of active accounts, I have used the following SQL query:

**SELECT**

**g.gender,**

**ROUND(AVG(ci.EstimatedSalary), 2) AS avg\_estimated\_salary,**

**count(bc.Exited) AS active\_accounts**

**FROM customer\_info ci**

**JOIN bank\_churn bc ON ci.CustomerId = bc.CustomerId**

**JOIN (**

**SELECT 1 AS GenderID, 'Male' AS gender**

**UNION**

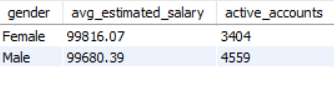
**SELECT 2, 'Female'**

**) g ON ci.GenderID = g.GenderID**

**where bc.Exited = 0**

**GROUP BY g.gender;**

**Result:**

****

**Insights:**

* Females have a higher average estimated salary even if they have a lower number of active accounts.
* Males have a slightly less average estimated salary as compared to females.

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

**Ans.** To segment the customers based on their credit score , I have defined the segment as given below:

**Low**: CreditScore < 600

**Medium**: 600 ≤ CreditScore < 700

**High**: CreditScore ≥ 700

**SQL query:**

**SELECT**

**CASE**

**WHEN CreditScore < 600 THEN 'Low'**

**WHEN CreditScore BETWEEN 600 AND 699 THEN 'Medium'**

**ELSE 'High'**

**END AS CreditSegment,**

**COUNT(\*) AS TotalCustomers,**

**SUM(Exited) AS ExitedCustomers,**

**ROUND(AVG(Exited) \* 100, 2) AS ExitRatePercentage**

**FROM bank\_churn**

**GROUP BY**

**CASE**

**WHEN CreditScore < 600 THEN 'Low'**

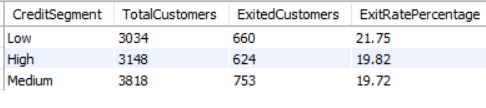
**WHEN CreditScore BETWEEN 600 AND 699 THEN 'Medium'**

**ELSE 'High'**

**END**

**ORDER BY ExitRatePercentage DESC;**

**Result:**



**Insights:**

* Customers with Low Credit Scores have the highest exit rate (21.75%), indicating a higher churn risk among financially weaker customers.
* Even though the Medium segment has the most customers, its exit rate (19.72%) is lower, suggesting that credit score positively correlates with retention up to a certain point.

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

**Ans.** To find out the region with the highest number of active customers with a tenure greater than 5 years , I have used the following SQL query:

**SELECT**

**g.GeographyLocation AS Region,**

**COUNT(\*) AS ActiveCustomersWithHighTenure**

**FROM bank\_churn bc**

**JOIN customer\_info ci ON bc.CustomerId = ci.CustomerId**

**JOIN geography g ON ci.GeographyID = g.GeographyID**

**WHERE bc.Exited = 0**

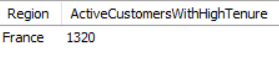
**AND bc.Tenure > 5**

**GROUP BY g.GeographyLocation**

**ORDER BY ActiveCustomersWithHighTenure DESC**

**LIMIT 1;**

**Result:**

****

**Insight:**

France has the highest number of active customers with a tenure greater than 5 years.

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

**Ans.** To find the impact of having a credit card on customer churn , I have used the following SQL query:

**SELECT**

**HasCrCard,**

**COUNT(\*) AS TotalCustomers,**

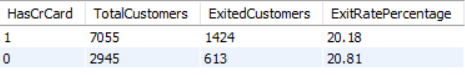
**SUM(Exited) AS ExitedCustomers,**

**ROUND(AVG(Exited) \* 100, 2) AS ExitRatePercentage**

**FROM bank\_churn**

**GROUP BY HasCrCard;**

**Result:**

****

**Insights:**

* Customers without a credit card have a slightly higher churn rate (20.81%) compared to those with a credit card (20.18%).
* Having a credit card shows a minor association with improved customer retention, though the difference is marginal.

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

**Ans.** To solve this question , I have used the following SQL query:

**SELECT**

**NumOfProducts,**

**COUNT(\*) AS ExitedCustomerCount**

**FROM bank\_churn**

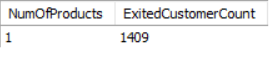
**WHERE Exited = 1**

**GROUP BY NumOfProducts**

**ORDER BY ExitedCustomerCount DESC**

**LIMIT 1;**

**Result:**

****

**Insight:**

**1** is the most common number of products used by customers who have exited.

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

**Ans.** To analyse the trend of customers joining over time , I have used the below SQL query:

**SELECT**

**YEAR(Bank\_DOJ) AS JoinYear,**

**MONTH(Bank\_DOJ) AS JoinMonth,**

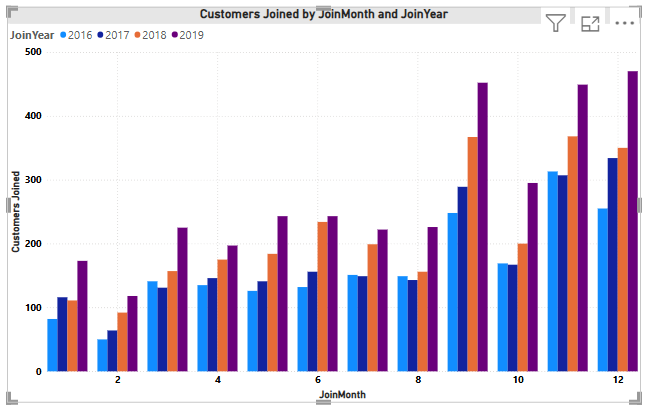
**COUNT(\*) AS CustomersJoined**

**FROM customer\_info**

**GROUP BY YEAR(Bank\_DOJ), MONTH(Bank\_DOJ)**

**ORDER BY JoinYear, JoinMonth;**

**Result:**

****

**Insights:**

* Most customers have joined in the year 2018 and 2019.
* Last quarter of the year is the peak quarter of joining.

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

**Ans.** Below is the SQL query used to analyze the relationship between the number of products and the account balance for those customers who have exited:

**SELECT**

**NumOfProducts,**

**COUNT(\*) AS ExitedCustomers,**

**ROUND(AVG(Balance), 2) AS AvgBalance**

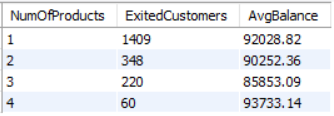
**FROM bank\_churn**

**WHERE Exited = 1**

**GROUP BY NumOfProducts**

**ORDER BY NumOfProducts;**

**Result:**

****

**Insights:**

* Most exited customers (1,409) had only 1 product, indicating limited product engagement may be linked to churn.
* Average balance generally decreases with more products, suggesting customers with multiple products may not always maintain higher balances.
* Customers with 4 products had a relatively high average balance (₹93,733) despite low volume, hinting at potential value loss from high-value customers.

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

**Ans.** To identify potential outliers in balance among customers who have not exited , I have used the following SQL query:

**-- Step 1: Get total count of active customers**

**SELECT COUNT(\*) AS TotalActive FROM bank\_churn WHERE Exited = 0;**

**-- Step 2: Get Q1(25th Percentile)**

**SELECT Balance AS Q1**

**FROM bank\_churn**

**WHERE Exited = 0**

**ORDER BY Balance**

**LIMIT 1 OFFSET 1989;**

**-- Step 3: Get Q3(75th Percentile)**

**SELECT Balance AS Q3**

**FROM bank\_churn**

**WHERE Exited = 0**

**ORDER BY Balance**

**LIMIT 1 OFFSET 5971;**

**-- Step 4: To find Outliers**

**SELECT \***

**FROM bank\_churn**

**WHERE Exited = 0**

**AND Balance > 315980.45;**

**Result:**

**TotalActive:**

****

**Q1 :**

****

**Q3 :**

****

**To Find Percentile Positions:**

* Q1 (25th percentile) ≈ 0.25 × 7963 = 1990.75 → Use OFFSET 1989
* Q3 (75th percentile) ≈ 0.75 × 7963 = 5972.25 → Use OFFSET 5971

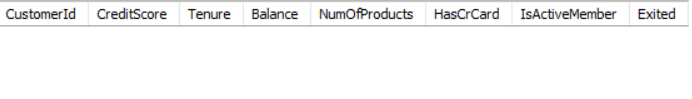
**To Calculate IQR and Thresholds:**

IQR = Q3 − Q1 = 126,392.18 − 0 = ₹126,392.18

Lower Bound = Q1 − 1.5 × IQR = 0 − 189,588.27 = −₹189,588.27  
 (Ignore this; balance can't be negative)

Upper Bound = Q3 + 1.5 × IQR = 126,392.18 + 189,588.27 = ₹315,980.45

**Final Outlier Result:**

****

**There are no significant high-balance outliers among active customers, based on the IQR method.**

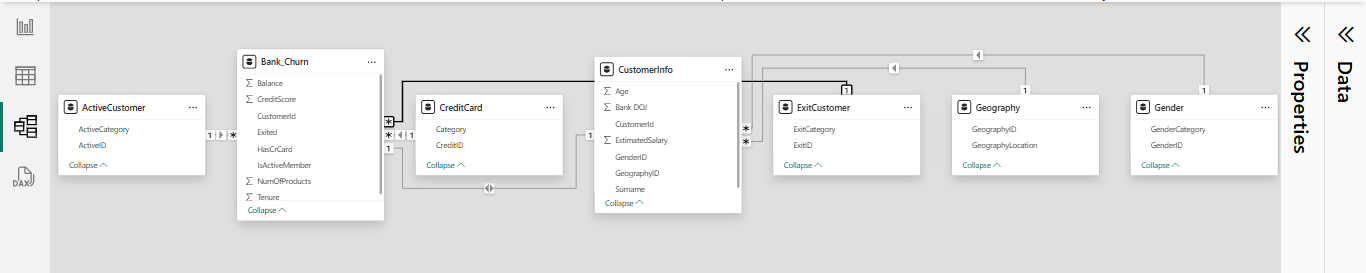
### **What This Tells Us:**

* Q1 = ₹0 means at least 25% of active customers have no balance at all.
* Q3 = ₹126,392.18 implies 75% of customers have balances below this.
* Since no active customer exceeds ₹3.15 lakhs, the distribution of balance is skewed toward the lower end with no extreme high values.

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

**Ans**. From the Power BI model view , there are 7 different tables in the dataset:

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

****

### **Table with Only Categorical Variables:**

**The table that consists only of categorical variables is:**

**➡️ Gender**

* Fields: GenderCategory, GenderID
* Both are categorical (no continuous/numeric measures)
* ExitCustomer, CreditCard, and Geography also mostly contain categorical fields, but they include ID columns that are used for relationships — some may still be treated as technical keys rather than "pure" categorical data.

**15. 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)**

**Ans.** To calculate the gender-wise average income for each GeographyID and then rank genders by that average within each geography, I have used the below SQL query:

**SELECT**

**GeographyID,**

**GenderID,**

**AVG(EstimatedSalary) AS Avg\_Income,**

**RANK() OVER (PARTITION BY GeographyID ORDER BY AVG(EstimatedSalary) DESC) AS Income\_Rank**

**FROM**

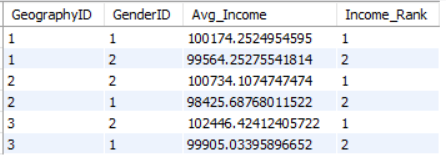
**customer\_info**

**GROUP BY**

**GeographyID,**

**GenderID;**

**Result:**

****

**Insights:**

* GeographyID 1: Males (GenderID 1) have a slightly higher average income than females.
* GeographyID 2: Males also lead in average income, but the gender gap is wider compared to GeographyID 1.
* GeographyID 3: Females (GenderID 2) earn more on average than males — the only geography with this reversal in ranking.
* Overall: Gender-based income differences vary by geography, with one region showing female income dominance.

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

**Ans.** To calculate the average tenure of exited customers grouped by age brackets (18–30, 31–50, 51+), I have written the following SQL query:

**SELECT**

**CASE**

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

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

**ELSE '51+'**

**END AS Age\_Bracket,**

**AVG(Tenure) AS Avg\_Tenure**

**FROM**

**bank\_churn bc**

**JOIN**

**customer\_info ci ON bc.CustomerId = ci.CustomerId**

**WHERE**

**Exited = 1**

**GROUP BY**

**CASE**

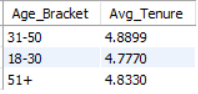
**WHEN Age BETWEEN 18 AND 30 THEN '18-30'**

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

**ELSE '51+'**

**END;**

**Result:**

****

**Insights:**

* **Age 31–50** has the **highest average tenure (4.89 years)** among exited customers, indicating longer retention before exit.
* **Age 18–30** group shows the **lowest average tenure (4.78 years)**, suggesting younger customers tend to leave earlier.
* **Age 51+** customers have a **moderate average tenure (4.83 years)**, slightly below the 31–50 group.

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

**Ans.** To calculate the correlation between salary and the account balance of the customer, I have used the following Excel formula:

**=CORREL(FILTER(Sheet4!E:E,H:H = 0),FILTER(D:D,H:H =0))**

Here filter is used to calculate for customers who have exited or not.

**Correlation** for active customers is **0.01719857062.**

**=CORREL(FILTER(Sheet4!E:E,H:H = 1),FILTER(D:D,H:H =1))**

**Correlation** for exited customers is **-0.01254045473.**

Based on calculated correlation values between EstimatedSalary and Balance:

* Exited customers: Correlation = −0.0125
* Non-exited customers: Correlation = +0.0172

### **Insights:**

* There is virtually no correlation between salary and balance for either exited or non-exited customers.
* The slightly negative correlation for exited customers (−0.0125) suggests that salary has no meaningful impact on their balance behavior.
* The slightly positive correlation for non-exited customers (0.0172) is also negligible and not statistically significant.
* This implies that EstimatedSalary is not a key driver of Balance, regardless of churn status.

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

**Ans.** To calculate the correlation between salary and the credit score of customers, I have used the following Excel formula:

**=CORREL(B:B,Sheet4!E:E)**

Result: **-0.001384292868**

**Insights:**

**Very Weak Negative Correlation**: The correlation value of -0.00138 indicates a very weak negative correlation between salary and credit score.

**Negligible Impact**: The correlation is so close to zero that it suggests there is almost no relationship between salary and credit score among the customers.

**No Strong Predictive Power**: Salary does not significantly predict the credit score of a customer.

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

**Ans.** To rank the credit score buckets by the number of churned customers, I have used the RANK() window function in SQL. This will assign a rank to each bucket based on the count of churned customers, where the bucket with the highest churn will be ranked first.

SQL query:

**SELECT**

**CreditScoreBucket,**

**ChurnedCustomers,**

**RANK() OVER (ORDER BY ChurnedCustomers DESC) AS Rnk**

**FROM (**

**SELECT**

**CASE**

**WHEN CreditScore BETWEEN 300 AND 499 THEN '300-499'**

**WHEN CreditScore BETWEEN 500 AND 599 THEN '500-599'**

**WHEN CreditScore BETWEEN 600 AND 699 THEN '600-699'**

**WHEN CreditScore BETWEEN 700 AND 799 THEN '700-799'**

**WHEN CreditScore BETWEEN 800 AND 900 THEN '800-900'**

**ELSE 'Other'**

**END AS CreditScoreBucket,**

**COUNT(\*) AS ChurnedCustomers**

**FROM bank\_churn**

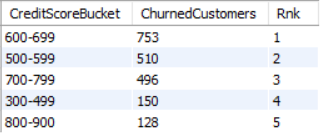
**WHERE Exited = 1**

**GROUP BY CreditScoreBucket**

**) AS BucketCounts**

**ORDER BY Rnk;**

**Result:**

****

**Insights:**

* **Highest Churn (600-699)**: Customers with credit scores between 600-699 have the highest churn (753), indicating higher exit risk.
* **Moderate Churn (500-599)**: The 500-599 range also shows significant churn (510), suggesting a moderate exit risk.
* **Lower Churn (700-799)**: Customers in the 700-799 range have a lower churn (496), indicating more stability.
* **Very Low Churn (300-499)**: The 300-499 range shows the least churn (150), possibly due to fewer customers or higher loyalty.
* **Lowest Churn (800-900)**: The 800-900 range has the least churn (128), suggesting that high credit scores are linked to stronger retention.

**20. 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.**

**Ans.** To analyze the number of customers who have a credit card in each age bucket, and retrieve only those buckets where the number of credit card holders is less than the average per bucket, I have used the following SQL query:

**WITH AgeBuckets AS (**

**SELECT**

**CASE**

**WHEN ci.Age BETWEEN 18 AND 25 THEN '18-25'**

**WHEN ci.Age BETWEEN 26 AND 35 THEN '26-35'**

**WHEN ci.Age BETWEEN 36 AND 45 THEN '36-45'**

**WHEN ci.Age BETWEEN 46 AND 55 THEN '46-55'**

**WHEN ci.Age BETWEEN 56 AND 65 THEN '56-65'**

**WHEN ci.Age > 65 THEN '65+'**

**ELSE 'Unknown'**

**END AS AgeBucket,**

**bc.HasCrCard**

**FROM customer\_info ci**

**JOIN bank\_churn bc ON ci.CustomerId = bc.CustomerId**

**),**

**CreditCardCounts AS (**

**SELECT**

**AgeBucket,**

**COUNT(\*) AS TotalCustomers,**

**SUM(CASE WHEN HasCrCard = 1 THEN 1 ELSE 0 END) AS CreditCardHolders**

**FROM AgeBuckets**

**GROUP BY AgeBucket**

**),**

**AverageCardHolders AS (**

**SELECT AVG(CreditCardHolders) AS AvgCardHoldersPerBucket**

**FROM CreditCardCounts**

**)**

**SELECT**

**c.AgeBucket,**

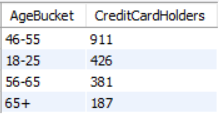
**c.CreditCardHolders**

**FROM CreditCardCounts c**

**JOIN AverageCardHolders a**

**ON c.CreditCardHolders < a.AvgCardHoldersPerBucket;**

**Result:**

****

**Insights:**

* 46–55 age group has the highest number of credit card holders below the average, indicating potential for increased engagement or targeting.
* 18–25 and 56–65 groups also fall below average, suggesting younger and older adults may be less likely to hold credit cards.
* 65+ group has the lowest number (187), possibly due to reduced financial activity or reliance on other payment methods.

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

**Ans.** To rank the locations by number of churned customers and average balance, I have used the following SQL query:

**SELECT**

**Location,**

**Num\_Churned\_Customers,**

**Avg\_Balance,**

**RANK() OVER (ORDER BY Num\_Churned\_Customers DESC, Avg\_Balance DESC) AS Location\_Rank**

**FROM (**

**SELECT**

**g.GeographyLocation AS Location,**

**COUNT(CASE WHEN bc.Exited = 1 THEN 1 END) AS Num\_Churned\_Customers,**

**AVG(bc.Balance) AS Avg\_Balance**

**FROM**

**customer\_info ci**

**JOIN**

**bank\_churn bc ON ci.CustomerId = bc.CustomerId**

**JOIN**

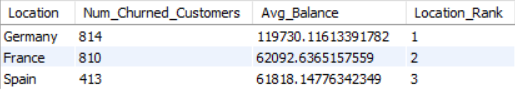
**geography g ON ci.GeographyID = g.GeographyID**

**GROUP BY**

**g.GeographyLocation**

**) AS RankedLocations;**

**Output:**

****

**Insights:**

* Germany ranks 1st with the highest number of churned customers (814) and the highest average balance (~119.7K) — indicating potential high-value customer loss.
* France ranks 2nd, also showing significant churn (810) but a much lower average balance (~62.1K) than Germany.
* Spain ranks 3rd, with the lowest churn (413) and slightly lower average balance (~61.8K) compared to France**.**

**22. 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”.**

**Ans.** As we have not a second table where the primary key is also a combination of CustomerID and Surname, so I have assumed that the second table is called CustomerDetails and has a column named CustomerID\_Surname.

SQl query to do this task:

**SELECT**

**ci.CustomerID,**

**ci.Surname,**

**cd.\***

**FROM**

**CustomerInfo ci**

**JOIN**

**CustomerDetails cd**

**ON CONCAT(ci.CustomerID, '\_', ci.Surname) = cd.CustomerID\_Surname;**

### **What This Does:**

* Dynamically constructs a CustomerID\_Surname key from the CustomerInfo table.
* Joins it with the same composite key in CustomerDetails.

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

**Ans.** Yes, we can add the ExitCategory from the ExitCustomers table to the Bank\_Churn table without using a JOIN, by using a subquery or CASE expression.

Since ExitID in ExitCustomers corresponds directly to the Exited column in Bank\_Churn, we can map it using a correlated subquery or a CASE statement.

SQL query:

**SELECT**

**CustomerId,**

**CreditScore,**

**Tenure,**

**Balance,**

**NumOfProducts,**

**HasCrCard,**

**IsActiveMember,**

**Exited,**

**CASE**

**WHEN Exited = 1 THEN 'Exit'**

**WHEN Exited = 0 THEN 'Retain'**

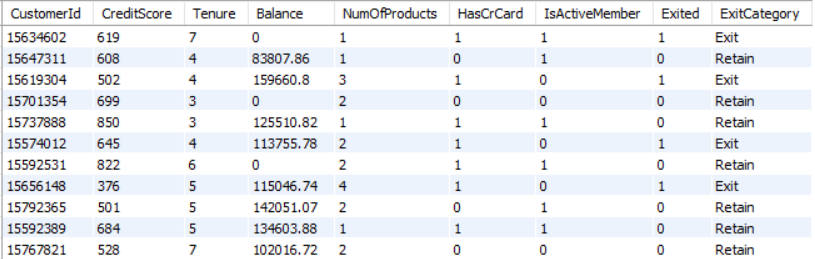
**ELSE 'Unknown'**

**END AS ExitCategory**

**FROM**

**Bank\_Churn;**

**Result:**

****

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

**Ans.** There were no missing values in any of the tables. However, in the Bank\_Churn table, the Balance column contained some entries with a value of 0.

I **did not replace or modify** the 0 values in the Balance column. I **left them as they are**.

### **Reason for not replacing the 0 values:**

* A 0 balance is **a valid and meaningful value**, not a missing one.
* It may indicate that the customer has **no funds in their account**, which is important for churn analysis.
* Replacing 0 with mean, median, or null would **distort the actual financial behavior** of such customers.
* In churn prediction, customers with a balance of 0 might be at **higher risk of leaving the bank**, so it's crucial to retain this signal.

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

**Ans.** To retrieve the Customer IDs, last names (surnames), and active status of customers whose surname ends with “on”, I have used the following SQL query:

**SELECT ci.CustomerId,**

**ci.Surname,**

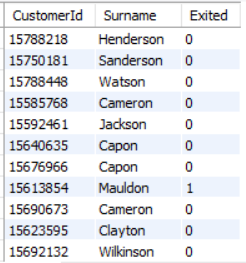
**bc.Exited**

**FROM customer\_info ci**

**JOIN bank\_churn bc ON ci.CustomerId = bc.CustomerId**

**WHERE ci.Surname LIKE '%on';**

**Result:**

****

These are some of the customers whose surname ends with “on” and their activity status.

**26. 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.**

**Ans.** Yes, there is a data discrepancy in the IsActiveMember and Exited columns. Here's the inconsistency:

### **Key Point:**

It is mentioned that Exited is accurate, which means we should trust this column.

### **Expected Relationship**:

* Typically, if a customer has exited (Exited = 1), they should not be an active member (IsActiveMember = 0).
* If Exited = 1 and IsActiveMember = 1, it implies the customer is marked as both active and exited, which is a contradiction.

There is a data inconsistency where IsActiveMember = 1 and Exited = 1, which contradicts the assumption that exited customers must no longer be active. So, I have considered the Exited column for my analysis, as it is confirmed to be accurate, it's smart to base our analysis on it. Ignoring the IsActiveMember column avoids misleading insights, especially when identifying churned customers.

**Subjective Answers**

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

**Ans.** To analyze customer behavior with a focus on spending habits of long-term vs. new customers, here’s a structured breakdown with observed patterns and their implications:

### **Approach:**

* Long-term customers: High tenure (e.g., ≥ 5 years)
* New customers: Low tenure (e.g., ≤ 2 years)
* Spending habits proxy: Use of Balance, NumOfProducts, and EstimatedSalary (actual spending data is unavailable)

SQL query:

**SELECT**

**CASE**

**WHEN bc.Tenure >= 5 THEN 'Long-term'**

**WHEN bc.Tenure <= 2 THEN 'New'**

**ELSE 'Mid-term'**

**END AS CustomerType,**

**COUNT(\*) AS TotalCustomers,**

**AVG(bc.Balance) AS AvgBalance,**

**AVG(bc.NumOfProducts) AS AvgNumOfProducts,**

**AVG(ci.EstimatedSalary) AS AvgEstimatedSalary,**

**AVG(bc.HasCrCard) AS CreditCardUsageRate,**

**AVG(bc.Exited) AS ChurnRate**

**FROM bank\_churn bc**

**JOIN customer\_info ci ON bc.CustomerId = ci.CustomerId**

**GROUP BY**

**CASE**

**WHEN bc.Tenure >= 5 THEN 'Long-term'**

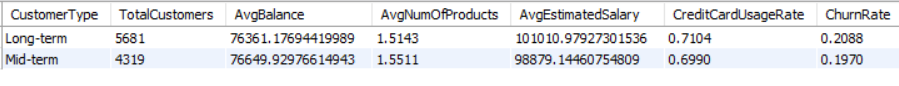
**WHEN bc.Tenure <= 2 THEN 'New'**

**ELSE 'Mid-term'**

**END**

**ORDER BY CustomerType;**

**Result:**

****

**Insights:**

* Mid-term customers maintain a slightly higher average balance (₹76,649.93) compared to long-term customers (₹76,361.18).
* This suggests that mid-term customers may be at a peak engagement stage in their lifecycle, exhibiting stronger transactional activity and trust in the bank.
* On average, mid-term customers use 1.55 products, while long-term customers use 1.51.
* Long-term customers report a higher average estimated salary (₹101,010.97) compared to mid-term customers (₹98,879.14).
* 71.04% of long-term customers hold credit cards, compared to 69.90% of mid-term customers.
* The churn rate is **higher among long-term customers** (20.88%) than mid-term customers (19.70%).

**Recommendations:**

* While mid-term customers show slightly higher engagement in balance and product usage, long-term customers hold greater income potential and deeper credit card adoption.
* The elevated churn rate among long-term customers signals a need for retention-focused strategies. These could include loyalty programs, regular satisfaction surveys, or proactive service enhancements to re-engage this high-value group.
* Overall, the findings suggest that customer loyalty does not automatically improve with tenure, and ongoing effort is required to maintain and grow the relationship.

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

**Ans**. To understand customer behavior more deeply and identify opportunities for cross-selling, I analyzed the product affinity—i.e., which bank products are most commonly used together.

**Approach:**

The key variables considered include:

* **NumOfProducts** (total number of products a customer uses)
* **HasCrCard** (1 = has credit card)
* **Balance > 0** (proxy for savings/investment usage)
* **Exited** (engagement indicator)

SQL query:

**SELECT**

**-- Product affinity combinations**

**SUM(CASE WHEN bc.HasCrCard = 1 AND bc.Balance > 0 THEN 1 ELSE 0 END) AS CrCard\_And\_Savings,**

**SUM(CASE WHEN bc.HasCrCard = 1 AND bc.NumOfProducts >= 2 THEN 1 ELSE 0 END) AS CrCard\_And\_MultiProduct,**

**SUM(CASE WHEN bc.Balance > 0 AND bc.NumOfProducts >= 2 THEN 1 ELSE 0 END) AS Savings\_And\_MultiProduct,**

**SUM(CASE WHEN bc.HasCrCard = 1 AND bc.Exited = 1 THEN 1 ELSE 0 END) AS CrCard\_And\_Active,**

**SUM(CASE WHEN bc.NumOfProducts = 1 AND ci.EstimatedSalary > 100000 THEN 1 ELSE 0 END) AS HighSalary\_SingleProduct,**

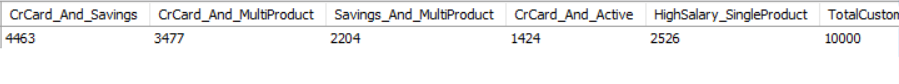
**COUNT(\*) AS TotalCustomers**

**FROM bank\_churn bc**

**JOIN customer\_info ci**

**ON bc.CustomerId = ci.CustomerId;**

**Result:**

****

**Insights and Recommendations:**

**CrCard\_And\_Savings (4,463 customers)**

* The most common combination, indicating strong affinity between credit card usage and savings/investment activity.
* Ideal base for promoting wealth management or investment products.

**CrCard\_And\_MultiProduct (3,477 customers)**

* A large segment uses a credit card along with multiple other products, suggesting high engagement and loyalty potential.
* These customers are well-suited for personalized premium offerings or bundled financial services.

**Savings\_And\_MultiProduct (2,204 customers)**

* Shows that many savings users also hold multiple products, pointing to cross-sell effectiveness in deposit-holding customers.

**CrCard\_And\_Active (1,424 customers)**

* A smaller but important segment of active credit card users, potentially high spenders or digitally engaged customers.
* They could be targeted for reward-based or usage-based retention strategies.

**HighSalary\_SingleProduct (2,526 customers)**

* A sizable group with high salary but only one product; represents a missed cross-sell opportunity.
* This segment should be prioritized for personalized upselling, especially for investment or credit services.

**Total Customers = 10,000**

* More than 40% of customers (4,463) already use both credit cards and savings—highlighting a natural product pairing.
* Cross-selling should focus on the HighSalary\_SingleProduct group and those not already in MultiProduct segments.

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

**Ans.** Here’s a detailed analysis of Geographic Market Trends based on the provided datasets, focusing on correlations between economic indicators, active accounts, and churn (Exited) across regions.

### **Approach:**

To analyze trends by geography, I merged customer\_info, bank\_churn, and geography datasets using CustomerId and GeographyID. The key columns analyzed:

* **EstimatedSalary** (economic indicator)
* **IsActiveMember** and **Exited** (activity and churn indicators)
* **Geography** (France, Spain, Germany)

**SQL query:**

**SELECT**

**g.GeographyLocation AS Region,**

**COUNT(c.CustomerId) AS Total\_Customers,**

**AVG(c.EstimatedSalary) AS Avg\_Estimated\_Salary,**

**SUM(CASE WHEN b.Exited = 0 THEN 1 ELSE 0 END) \* 100.0 / COUNT(\*) AS Active\_Percentage,**

**SUM(CASE WHEN b.Exited = 1 THEN 1 ELSE 0 END) \* 100.0 / COUNT(\*) AS Churn\_Percentage**

**FROM**

**customer\_info c**

**JOIN**

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

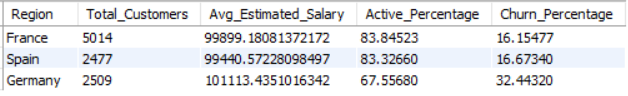
**JOIN**

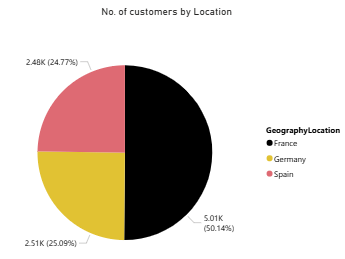
**geography g ON c.GeographyID = g.GeographyID**

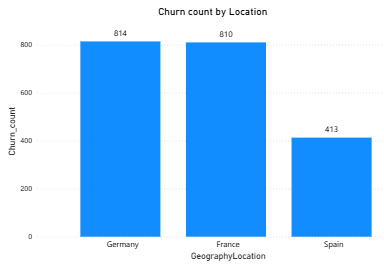
**GROUP BY**

**g.GeographyLocation;**

**Result:**

****

****

****

**Insights:**

* France and Spain show strong customer engagement with high active percentages (over 83%) and low churn (~16%).
* Germany, despite having the highest average salary, has the lowest active rate (67.6%) and highest churn (32.4%)—indicating a possible satisfaction or service issue.

**Recommendations:**

* Investigate churn causes in Germany—conduct surveys or analyze service gaps.
* Replicate successful strategies from France and Spain, such as better onboarding or loyalty programs.
* Target German customers with tailored offers to improve retention, especially high-income segments.

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

**Ans.** To assess financial risk based on customer profiles, we'll look for demographic segments with high churn (Exited = 1), low credit scores, and high balances.

**Approach:**

**Key Risk Indicators Used:**

* Churn Rate (Exited): A higher churn rate indicates a risk of customer loss.
* Average Credit Score: Lower scores suggest higher credit risk and lower repayment ability.
* Average Balance: High balances in high-churn segments indicate greater financial exposure for the bank.

**Segmentation Criteria:**

* Region (Geography): To understand regional variations in risk.
* Age Group: Bucketed into 10-year bands (e.g., 20s, 30s) to analyze generational behavior.

**SQL query:**

**SELECT**

**g.GeographyLocation AS Region,**

**FLOOR(c.Age / 10) \* 10 AS Age\_Group,**

**COUNT(c.CustomerId) AS Total\_Customers,**

**round(AVG(b.CreditScore),2) AS Avg\_CreditScore,**

**round(AVG(b.Balance),2) AS Avg\_Balance,**

**round(SUM(CASE WHEN b.Exited = 1 THEN 1 ELSE 0 END) \* 100.0 / COUNT(\*),2) AS Churn\_Percentage**

**FROM**

**customer\_info c**

**JOIN**

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

**JOIN**

**geography g ON c.GeographyID = g.GeographyID**

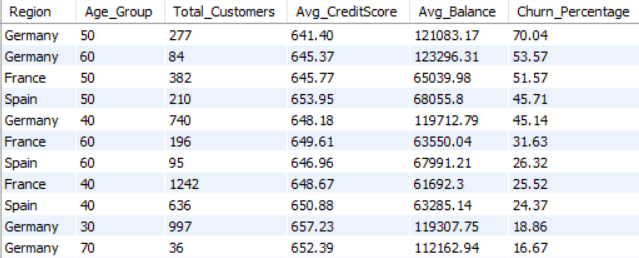
**GROUP BY**

**g.GeographyLocation, FLOOR(c.Age / 10) \* 10**

**ORDER BY**

**Churn\_Percentage DESC;**

**Result:**

****

### 

### 

### **Insights:**

1. **Germany - Age Group 50**
   * Highest churn rate (70%), low credit score (641.4), and high average balance (₹121K) → highest financial risk segment.

**2.**  **Germany - Age Group 60**

* Churn is **53.57%**, with **very high balance (₹123K)** → older customers with high funds are exiting at high rates.

**3. France - Age Group 50**

* Churn is 51.57% with moderate balance (₹65K) → indicates mid-life segment instability.

**4. Spain - Age Group 50**

* Churn is 45.71%, but credit score is slightly better (653.95) → needs moderate attention.

**5. Younger segments (30s and 40s)**

* Show lower churn rates (16%–26%), even when balances are high → relatively stable, lower risk segments.

**Recommendations:**

**1. Prioritize retention in Germany (50s and 60s):**

* Implement loyalty programs, conduct exit interviews, and offer personalized financial planning.

**2. Monitor high-balance customers closely:**

* Especially those with low credit scores and higher age groups—there’s significant financial exposure if they churn.

**3. Targeted credit improvement initiatives:**

* Offer credit education or restructuring tools to customers in risky age bands with lower credit scores.

**4. Protect mid-life customers in France and Spain:**

* Run engagement campaigns for 50s age group to improve loyalty and satisfaction.

**5. Strengthen onboarding and product bundling for younger groups (30s, 40s):**

* These are more stable and can be nurtured into high-value long-term customers.

**5. Customer Tenure Value Forecast: How would you use the available data to model and predict the lifetime (tenure) value in the bank of different customer segments?**

**Ans.** To forecast the customer tenure value, I have used the following approach:

**Approach:**

I combined the bank\_churn and customer\_info tables in Power BI using CustomerId to create a complete customer profile. Then, I calculated a new metric called **Tenure Value** using the formula:

**Tenure Value = Tenure × (Balance + EstimatedSalary / 12)**

This gives an estimate of each customer’s long-term value to the bank.

Next, I created customer segments based on credit score:

**CustomerSegment = SWITCH(**

**TRUE(),**

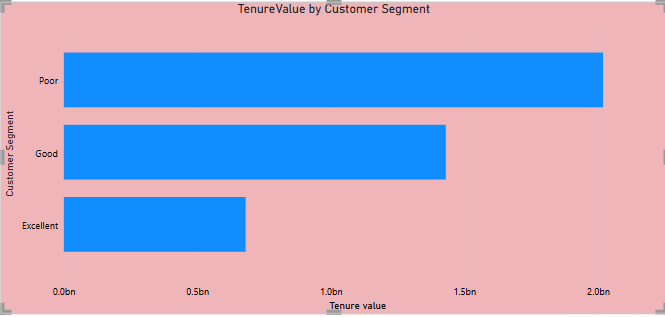
**[CreditScore] >= 750, "Excellent",**

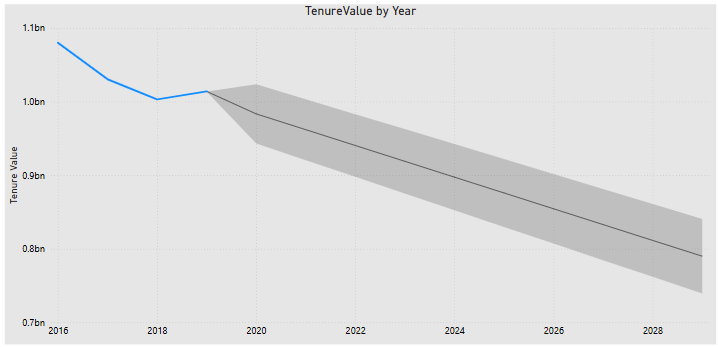
**[CreditScore] >= 650, "Good",**

**"Poor")**

and used Power BI visuals like bar charts and line charts (with forecasting) to analyze tenure value trends and segment performance. This helped identify high-value customer groups and forecast future value patterns.

**Result:**

****

****

### **Insights:**

1. **Poor Credit Segment Has Highest Tenure Value**  
    The "Poor" credit segment surprisingly contributes the most to overall tenure value. This suggests that even though these customers have lower credit scores, they are either staying longer or holding higher balances/income.
2. **Excellent Segment Contributes the Least** Customers in the "Excellent" credit segment show the lowest tenure value. This may indicate that financially strong customers either churn earlier or maintain lower balances.
3. **Declining Trend in Tenure Value Over Time** The line chart with forecasting shows a steady decline in tenure value starting around 2020, with predictions suggesting continued drop until 2028. This could indicate increasing customer churn or reduced balance/income contributions over time.

### **Recommendations:**

1. **Re-Evaluate Customer Value Assumptions** Do not assume high credit score customers are always the most valuable. Consider targeting the "Poor" segment with retention offers or financial advisory services to increase their lifetime value even further.
2. **Engage Excellent Segment More Effectively** Investigate why the "Excellent" segment underperforms in tenure value—perhaps due to lack of engagement or competitive offers from other banks. Tailored rewards, investment options, or premium services may help.
3. **Address Decline in Tenure Value** The downward trend suggests emerging customer dissatisfaction or stronger market competition. It may be time to strengthen onboarding, cross-selling, and loyalty strategies.
4. **Monitor Tenure Trends Yearly** Set up annual reviews of tenure value by segment to adapt your retention and marketing strategies based on changing customer behavior.

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

**Ans.** To assess the impact of marketing campaigns on customer retention and acquisition, the existing dataset provides a strong foundation for analyzing customer behavior but lacks critical campaign-specific information needed for a comprehensive evaluation.

The current dataset includes two main tables:

* bank\_churn: Contains indicators of customer activity and churn (e.g., Exited, IsActiveMember, HasCrCard, etc.).
* customer\_info: Contains demographic and financial data, along with Bank\_DOJ, which helps identify new customer acquisition timelines.

However, in order to evaluate the effectiveness of marketing campaigns, additional data is necessary—specifically, a marketing campaign table. This table should include details such as CustomerId, campaign type (email, SMS, phone), campaign dates, response indicators, and whether the campaign led to customer acquisition or retention. Without this data, it is not possible to directly attribute customer retention or acquisition to specific campaigns.

Once such data is available, Power BI would be an appropriate tool for this analysis. It can be used to:

* Track churn and acquisition trends before and after campaign periods.
* Segment customers based on campaign exposure and responses.
* Visualize key metrics such as retention rate, churn rate, and acquisition growth by demographic or geographic segments.

The analysis would begin by integrating the campaign data with existing customer and churn information. Then, KPIs like churn rate among campaign responders or acquisition rate post-campaign could be developed. These insights can be visualized in Power BI dashboards using trend lines, cohort analysis, and comparison charts to effectively communicate the impact of marketing efforts.

In summary, while the current dataset allows for basic churn and acquisition analysis, additional campaign-related data is essential to directly assess marketing campaign effectiveness. Once available, a combination of Power BI reporting and structured analysis would provide valuable insights into how marketing initiatives influence customer behavior.

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

**Ans.** To identify the common characteristics or trends among customers who have exited the bank, I began by analyzing the data from both the bank\_churn and customer\_info tables. The goal was not just to observe frequent patterns among churned customers, but to understand what differentiates them from retained customers.

**Approach:**

1. I focused on categorical and behavioral variables such as GenderID, GeographyID, NumOfProducts, Tenure, and HasCrCard**.**
2. To move beyond surface-level observations, I chose to calculate churn rates across these customer segments.
3. This involved comparing the number of customers who exited versus the total number of customers within each category.

I examined churn rate by:

* **Number of Products** to assess engagement levels.
* **Tenure** to understand whether loyalty decays over time.
* **Geography** to evaluate regional differences in retention.
* **Credit Card ownership** to see if financial product penetration affects churn.

**SQL query to for number of products:**

**SELECT**

**NumOfProducts,**

**COUNT(\*) AS Total\_Customers,**

**SUM(CASE WHEN Exited = 1 THEN 1 ELSE 0 END) AS Exited\_Customers,**

**ROUND(**

**SUM(CASE WHEN Exited = 1 THEN 1 ELSE 0 END) \* 100.0 / COUNT(\*),**

**2**

**) AS Churn\_Rate\_Percent**

**FROM**

**bank\_churn**

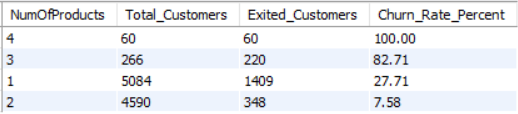
**GROUP BY**

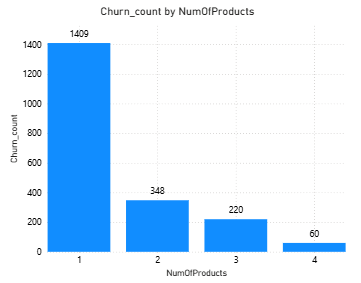
**NumOfProducts**

**ORDER BY**

**Churn\_Rate\_Percent DESC;**

**Result:**

****

****

**Insights:**

* Customers with 3 or more products have extremely high churn, especially those with 4 products (100%) and 3 products (82.71%), which is unusual and suggests dissatisfaction despite high engagement.
* Conversely, customers with 2 products are the most stable, having the lowest churn rate (7.58%).

**Recommendations:**

**Review Multi-Product Customer Strategy:**

* The 100% and 82.71% churn among 4- and 3-product holders is alarming. This could indicate product overload, complexity, or lack of value.
* Assess whether products are truly complementary or causingfrustration. Consider bundling or simplifying offerings.
* Implement proactive support or check-ins for customers holding multiple products.

**Encourage Adoption of Second Product:**

* Customers with 2 products are highly retained. Promote cross-selling strategies to encourage single-product customers to adopt one more relevant product.
* Offer incentives or educational content to highlight the benefits of adding a second product**.**

**SQL query for tenure:**

**SELECT**

**Tenure,**

**COUNT(\*) AS Total\_Customers,**

**SUM(CASE WHEN Exited = 1 THEN 1 ELSE 0 END) AS Exited\_Customers,**

**ROUND(**

**SUM(CASE WHEN Exited = 1 THEN 1 ELSE 0 END) \* 100.0 / COUNT(\*),**

**2**

**) AS Churn\_Rate\_Percent**

**FROM**

**bank\_churn**

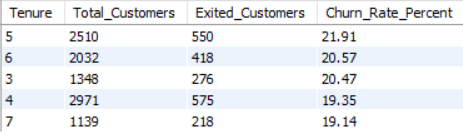
**GROUP BY**

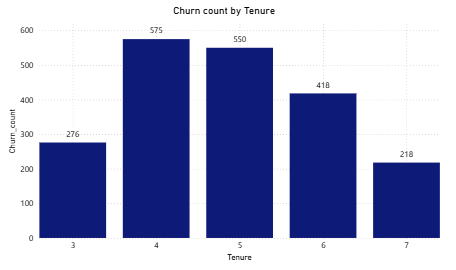
**Tenure**

**ORDER BY**

**Churn\_Rate\_Percent DESC;**

**Result:**

****

****

**Insights:**

* Customers with medium tenure (3–6 years) have the highest churn rates (~20–22%).
* These customers may feel undervalued over time or have unmet expectations after initial engagement.

**Recommendations:**

* Launch loyalty programs or personalized offers for customers in the 3–6 year tenure group to increase satisfaction and retention.
* Conduct feedback surveys specifically targeting this segment to understand their concerns.

**SQL query for Location:**

**SELECT**

**g.GeographyLocation,**

**COUNT(\*) AS Total\_Customers,**

**SUM(CASE WHEN bc.Exited = 1 THEN 1 ELSE 0 END) AS Exited\_Customers,**

**ROUND(**

**SUM(CASE WHEN bc.Exited = 1 THEN 1 ELSE 0 END) \* 100.0 / COUNT(\*),**

**2**

**) AS Churn\_Rate\_Percent**

**FROM**

**bank\_churn bc**

**join customer\_info c**

**on bc.customerID = c.customerID**

**join geography g**

**on c.GeographyID = g.GeographyID**

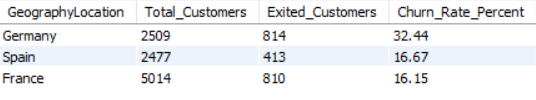
**GROUP BY**

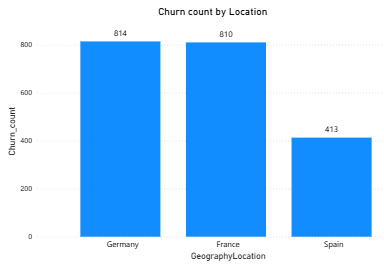
**g.GeographyLocation**

**ORDER BY**

**Churn\_Rate\_Percent DESC;**

**Result:**

****

****

**Insights:**

Germany has a significantly higher churn rate (32.44%) compared to Spain and France, indicating potential country-specific dissatisfaction or competition.

**Recommendations:**

* Perform a deep-dive analysis into the customer experience in Germany—review product offerings, customer service performance, and competitor actions.
* Tailor local marketing and retention strategies to better address customer needs in this market.

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

**Ans.** Yes, several of the variables — Tenure, NumOfProducts, IsActiveMember, and EstimatedSalary — can indeed be important for predicting whether a customer will exit the bank (churn).

**Approach:**

To evaluate the predictive importance of Tenure, NumOfProducts, IsActiveMember, and EstimatedSalary in customer churn, I adopted a segment-based analysis approach.

I began by calculating the churn rate within each variable category to uncover behavioral or demographic patterns. This involved analyzing how frequently customers exited the bank within different values or ranges of each feature.

**I examined:**

* Tenure, to understand if customer loyalty declines or strengthens over time, and to identify tenure bands with high exit rates.
* Number of Products, to assess whether owning more or fewer products impacts engagement or frustration levels leading to churn.
* IsActiveMember, to test the assumption that active members are more likely to stay, and whether inactivity is a churn signal.
* EstimatedSalary, to explore whether income levels play a role in customer satisfaction and retention.

**SQL query for number of products:**

**SELECT**

**NumOfProducts,**

**COUNT(\*) AS Total\_Customers,**

**SUM(CASE WHEN Exited = 1 THEN 1 ELSE 0 END) AS Exited\_Customers,**

**ROUND(**

**SUM(CASE WHEN Exited = 1 THEN 1 ELSE 0 END) \* 100.0 / COUNT(\*),**

**2**

**) AS Churn\_Rate\_Percent**

**FROM**

**bank\_churn**

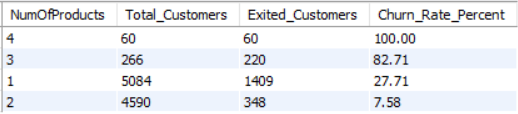
**GROUP BY**

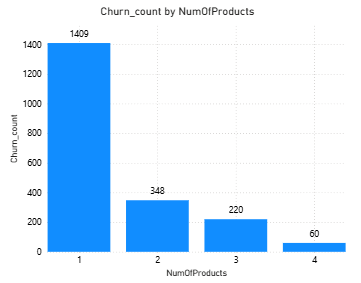
**NumOfProducts**

**ORDER BY**

**Churn\_Rate\_Percent DESC;**

**Result:**

****

****

**Insights:**

* Customers with 3 or more products have extremely high churn, especially those with 4 products (100%) and 3 products (82.71%), which is unusual and suggests dissatisfaction despite high engagement.
* Conversely, customers with 2 products are the most stable, having the lowest churn rate (7.58%).

**Recommendations:**

**Review Multi-Product Customer Strategy:**

* The 100% and 82.71% churn among 4- and 3-product holders is alarming. This could indicate product overload, complexity, or lack of value.
* Assess whether products are truly complementary or causingfrustration. Consider bundling or simplifying offerings.
* Implement proactive support or check-ins for customers holding multiple products.

**Encourage Adoption of Second Product:**

* Customers with 2 products are highly retained. Promote cross-selling strategies to encourage single-product customers to adopt one more relevant product.
* Offer incentives or educational content to highlight the benefits of adding a second product**.**

**SQL query for tenure:**

**SELECT**

**Tenure,**

**COUNT(\*) AS Total\_Customers,**

**SUM(CASE WHEN Exited = 1 THEN 1 ELSE 0 END) AS Exited\_Customers,**

**ROUND(**

**SUM(CASE WHEN Exited = 1 THEN 1 ELSE 0 END) \* 100.0 / COUNT(\*),**

**2**

**) AS Churn\_Rate\_Percent**

**FROM**

**bank\_churn**

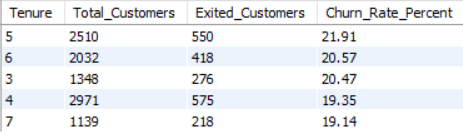
**GROUP BY**

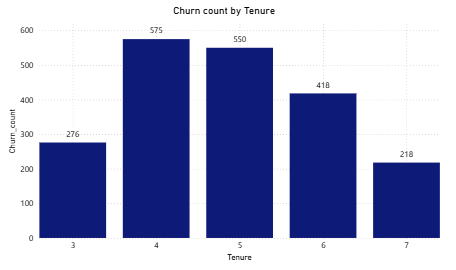
**Tenure**

**ORDER BY**

**Churn\_Rate\_Percent DESC;**

**Result:**

****

****

**Insights:**

* Customers with medium tenure (3–6 years) have the highest churn rates (~20–22%).
* These customers may feel undervalued over time or have unmet expectations after initial engagement.

**Recommendations:**

* Launch loyalty programs or personalized offers for customers in the 3–6 year tenure group to increase satisfaction and retention.
* Conduct feedback surveys specifically targeting this segment to understand their concerns.
* Evaluate whether long-term customers are being neglected or receiving fewer upgrades compared to new customers.

**SQL query for Estimated salary:**

**SELECT**

**CASE**

**WHEN ci.EstimatedSalary < 50000 THEN 'Under 50K'**

**WHEN ci.EstimatedSalary BETWEEN 50000 AND 100000 THEN '50K - 100K'**

**WHEN ci.EstimatedSalary BETWEEN 100001 AND 150000 THEN '100K - 150K'**

**ELSE '150K+'**

**END AS SalaryRange,**

**COUNT(bc.CustomerId) AS TotalCustomers,**

**SUM(bc.Exited) AS ExitedCustomers,**

**ROUND(SUM(bc.Exited) \* 100.0 / COUNT(bc.CustomerId), 2) AS ChurnRate**

**FROM**

**bank\_churn bc**

**JOIN**

**customer\_info ci ON bc.CustomerId = ci.CustomerId**

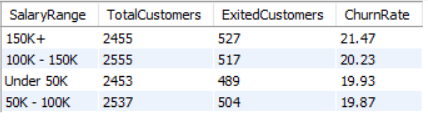
**GROUP BY**

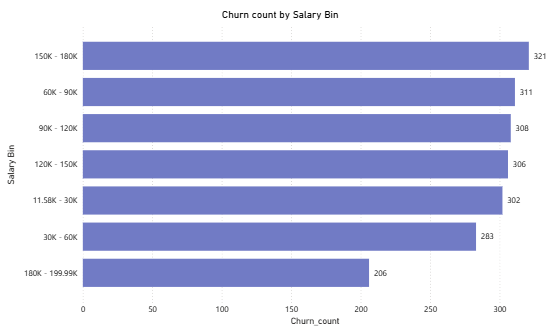
**SalaryRange**

**ORDER BY**

**ChurnRate DESC;**

**Result:**

****

****

**Insights:**

Customers with Estimated Salary above 150K have the highest churn rate (21.47%).

**Recommendations:**

* + Introduce or enhance premium banking services such as wealth management, exclusive credit cards, or dedicated relationship managers.
  + Conduct personalized outreach to high-income customers to better understand unmet needs.
  + Offer loyalty rewards or tier-based benefits for high-value clients to increase perceived value.

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

**Ans.** To effectively segment customers based on their demographics and account details, I followed a structured approach using SQL queries on two datasets: customer\_info and bank\_churn. My objective was to derive meaningful customer segments that could help the bank improve targeting, retention, and overall customer satisfaction**.**

**Approach:**

I created customer segments using SQL CASE statements and GROUP BY clauses based on:

* Age Groups: e.g., Under 30, 30–50, and Over 50.
* Gender and Churn Status: to analyze churn trends across genders.
* Geography and Activity: to understand regional engagement and loyalty.
* Credit Score Tiers: to classify customers into Low, Medium, and High creditworthiness.

These segments helped identify patterns like which age group or region had the highest churn or which customers were inactive but had high balances.

**SQL query for Segment by Age Group**

**SELECT**

**CASE**

**WHEN Age BETWEEN 18 AND 25 THEN '18-25'**

**WHEN Age BETWEEN 26 AND 35 THEN '26-35'**

**WHEN Age BETWEEN 36 AND 45 THEN '36-45'**

**WHEN Age BETWEEN 46 AND 55 THEN '46-55'**

**WHEN Age BETWEEN 56 AND 65 THEN '56-65'**

**ELSE '65+'**

**END AS Age\_Group,**

**COUNT(\*) AS Customer\_Count**

**FROM customer\_info**

**GROUP BY**

**CASE**

**WHEN Age BETWEEN 18 AND 25 THEN '18-25'**

**WHEN Age BETWEEN 26 AND 35 THEN '26-35'**

**WHEN Age BETWEEN 36 AND 45 THEN '36-45'**

**WHEN Age BETWEEN 46 AND 55 THEN '46-55'**

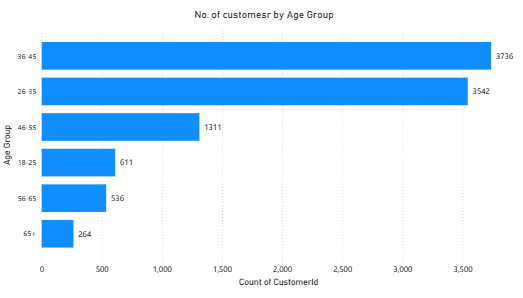
**WHEN Age BETWEEN 56 AND 65 THEN '56-65'**

**ELSE '65+'**

**END**

**ORDER BY Customer\_Count DESC;**

**Result:**

****

**Insights:**

* Majority (60%+) of customers are aged 26–45, indicating strong engagement from working professionals.
* 46–55 group is moderate, while 18–25, 56–65, and 65+ are underrepresented.

**Recommendations:**

* Focus marketing on 26–45 age group with tailored offers.
* Attract 18–25 with student-friendly products and digital incentives.
* Introduce senior-friendly services for 56+ customers to boost engagement.

**SQL query to Segment by Gender and Churn Status**

**SELECT**

**CASE GenderID**

**WHEN 1 THEN 'Male'**

**WHEN 2 THEN 'Female'**

**END AS Gender,**

**bc.Exited,**

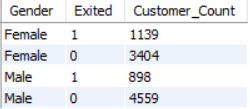
**COUNT(\*) AS Customer\_Count**

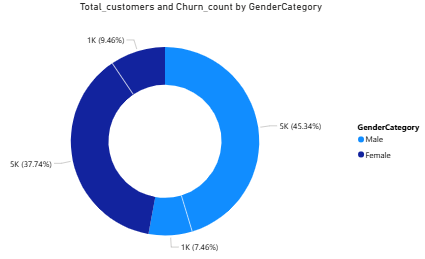
**FROM customer\_info ci**

**JOIN bank\_churn bc ON ci.CustomerId = bc.CustomerId**

**GROUP BY GenderID, bc.Exited;**

**Result:**

****

****

**Insights:**

* Female customers had a slightly higher churn rate compared to male customers.
* This may point to a gap in personalized services or financial product offerings tailored to women’s financial goals.

**Recommendations:**

* Enhance personalization in communication and offers for female customers.
* Introduce women-centric financial products, like savings plans or credit cards with lifestyle rewards.
* Conduct feedback surveys to understand specific needs and pain points.
* Offer financial literacy programs targeted at women to build trust and engagement.

**SQL query to Segment by Geography and Account Activity**

**SELECT g.GeographyLocation,**

**bc.Exited,**

**COUNT(\*) AS Customer\_Count**

**FROM customer\_info ci**

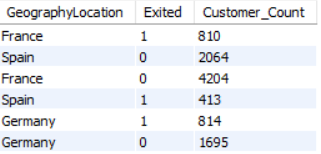
**JOIN bank\_churn bc ON ci.CustomerId = bc.CustomerId**

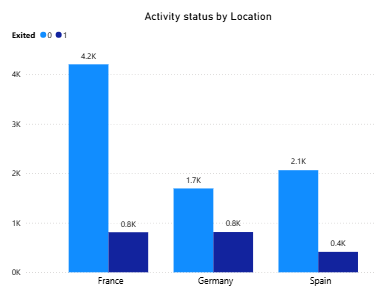
**join geography g**

**on g.GeographyID = ci.GeographyID**

**GROUP BY g.GeographyLocation, bc.Exited;**

**Result:**

****

****

**Insights:**

* Customers from Germany had the lowest engagement but also had higher churned customers.
* French customers formed the largest base and showed a balanced mix of active and inactive members.
* Spain has the lowest number of churned customers.

**Recommendations:**

**Germany:** Launch re-engagement campaigns and personalized retention offers to reduce churn.

**France:** Maintain engagement with consistent service quality and targeted promotions.

**Spain:** Analyze successful strategies and replicate them in other regions to minimize churn.

**SQL query to Segment by Credit Score Range**

**SELECT**

**CASE**

**WHEN CreditScore < 600 THEN 'Low'**

**WHEN CreditScore BETWEEN 600 AND 750 THEN 'Medium'**

**ELSE 'High'**

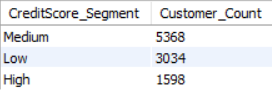
**END AS CreditScore\_Segment,**

**COUNT(\*) AS Customer\_Count**

**FROM bank\_churn**

**GROUP BY CreditScore\_Segment;**

**Result:**

****

**Insights:**

* Medium credit score (600–750) customers were the most stable, forming a key retention group.
* High credit score customers (>750), are fewer in number.

**Recommendations:**

* **Focus retention efforts** on medium credit score (600–750) customers with loyalty programs and tailored offers.
* **Attract high credit score customers** by promoting premium products, exclusive benefits, and personalized services.
* **Monitor and nurture** medium scorers to gradually upgrade them into the high-score segment.

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

**Ans.** To visually highlight customers at risk of churn and evaluate the impact of credit card rewards on retention, I have set up conditional formatting in Power BI using the following approach:

**Approach:**

**To Create a Churn Risk Indicator (DAX)**

Created a new column in Power BI to classify customers as "High Risk", "Medium Risk", or "Low Risk" based on factors like:

* Exited = 0
* CreditScore < 600
* Balance = 0
* NumOfProducts = 1
* HasCrCard = 0

**Dax Formula:**

**ChurnRisk =**

**SWITCH(**

**TRUE(),**

**Bank\_Churn[Exited] = 0 && bank\_churn[CreditScore] < 600 && bank\_churn[Balance] = 0, "High Risk",**

**Bank\_Churn[Exited] = 0 && bank\_churn[NumOfProducts] = 1, "Medium Risk",**

**"Low Risk" )**

**2. Applied Conditional Formatting to a Table**

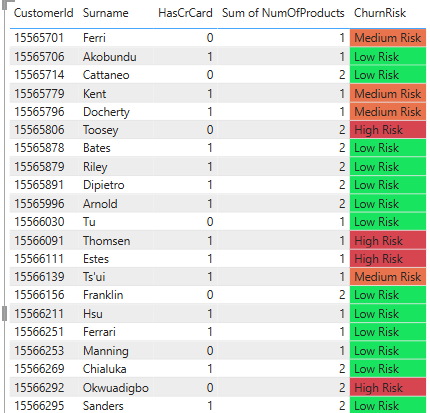
* **Added a table visual with fields like CustomerId, Surname, ChurnRisk, HasCrCard, Exited.**
* **Rules applied:**

Red for "High Risk"

Orange for "Medium Risk"

Green for "Low Risk"

**Result:**

****

**Insights:**

**High-Risk Customers Are Clearly Identified**

* Conditional formatting helped visually distinguish high-risk customers, primarily those who are inactive, have low credit scores, no balance, and only one product.
* These customers are likely disengaged and may require immediate retention actions.

**Recommendations:**

* Prioritize retention campaigns for high-risk customers with personalized offers and outreach.
* Encourage product bundling to increase engagement (e.g., offer discounts for using multiple products).
* Introduce balance incentives or small rewards to re-activate zero-balance accounts.
* Use targeted communication to educate and re-engage low credit score customers.

**11. What is the current churn rate per year and overall as well in the bank? Can you suggest some insights to the bank about which kind of customers are more likely to churn and what different strategies can be used to decrease the churn rate?**

**Ans.** To calculate current churn rate per year , I have used the following DAX formulas:

**Overall Churn Rate (%) =**

**DIVIDE(**

**COUNTROWS(FILTER(bank\_churn, bank\_churn[Exited] = 1)),**

**COUNTROWS(bank\_churn)**

**) \* 100**

AND

**Churn Rate by Year (%) =**

**VAR TotalCustomers = COUNTROWS(customer\_info)**

**VAR Churned =**

**CALCULATE(**

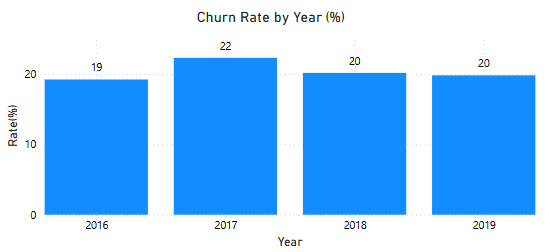
**COUNTROWS(bank\_churn),**

**bank\_churn[Exited] = 1**

**)**

**RETURN DIVIDE(Churned, TotalCustomers) \* 100**

**Result:**

****

**Insights:**

1. **2017 Had the Highest Churn (22%)**
   * Customers who joined in 2017 churned the most.
   * Indicates potential onboarding or engagement issues in that year.
2. **Churn Stabilized in Later Years**
   * In 2018 and 2019, churn stabilized at 20%, showing a slight improvement in customer retention efforts.
3. **2016 Cohort Showed Relatively Lower Churn**
   * Despite being the oldest cohort, 2016 customers had the lowest churn (19%)—suggesting that long-tenured customers may be more loyal if engaged well early on.
4. **No Consistent Downward Trend**
   * The churn rate did not decrease year-over-year, implying that strategies to reduce churn may not have been sustained or effective.

**Recommendations:**

**Improve Early Engagement for New Joiners**

* Since churn is often highest in the first 1–2 years, strengthen the onboarding process with personalized communications, tutorials, and product education.

**Target the 2017 Cohort for Recovery**

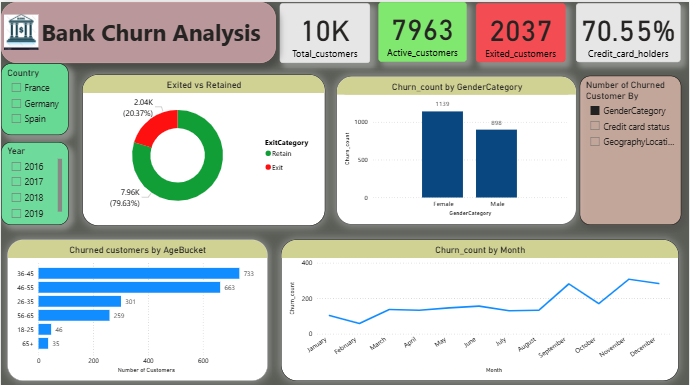
* Analyze the 2017 cohort in detail to identify why churn was highest—consider reaching out to similar remaining customers with re-engagement campaigns.

**Introduce Tenure-Based Loyalty Programs**

* Offer benefits or rewards to retain long-tenure customers, as they are more likely to stay loyal if they feel valued.

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

**Ans.** Below is the dashboard I have created in Power BI.



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

**Ans.** If no predefined questions were provided, I would take a structured, insight-driven approach to analyze customer churn and recommend strategies:

* **Define the Objective:** Focus on understanding churn patterns and identifying opportunities to improve customer retention and engagement.
* **Explore the Data:** Perform data cleaning and exploration to understand key features like age, credit score, balance, tenure, product usage, and activity status.
* **Establish Key Metrics:**
  + Overall and yearly churn rate
  + Product adoption rate (e.g., credit card, multiple products)
  + Inactivity levels and tenure-based segmentation
* **Create Visual Insights in Power BI:**
  + Churn trends by join year and geography
  + Credit card ownership vs. churn
  + Risk heatmaps for identifying at-risk segments
* **Identify Churn Drivers:** Analyze which customer characteristics (e.g., low credit score, inactivity, fewer products) are highly correlated with churn.
* **Develop Targeted Strategies:**
  + Personalized retention offers for high-risk customers
  + Bundling products and loyalty programs
  + Onboarding improvements for new customers
* **Optional:** Build a basic churn prediction model to assign risk scores and prioritize interventions.

This approach ensures a comprehensive understanding of churn while providing actionable, data-backed recommendations to reduce it.

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

**Ans.** To rename the column , I have used the following SQL query:

**ALTER TABLE Bank\_Churn**

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