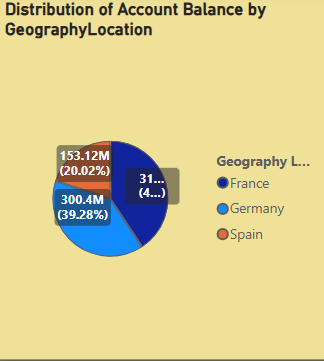
**Objective Questions:**

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

Ans: The distribution of account balances across different regions is below:

* **France** leads with the highest sum of balance, approximately at 311.33M (40.7%).
* **Germany** follows closely with a sum of balance around 300.4M (39.28%).
* **Spain**, however, has a significantly lower sum of balance, roughly at 153.12M (20.02%).



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

Ans: **SELECT**

**Surname, sum(EstimatedSalary )AS qrt\_sal**

**FROM**

**customer\_info**

**WHERE**

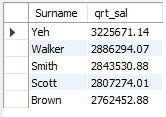
**Bank\_DOJ BETWEEN '01-09-2019' AND '31-12-2019'**

**GROUP BY Surname**

**ORDER BY sum(EstimatedSalary) DESC**

**LIMIT 5;**

* This query can be used for identifying the customer id and salary of the top 5 customers with the highest estimated salary in the last quarter of the year.



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

Ans:

* In the database, there are multiple tables that store information about the bank’s customers, their product usage, and whether they hold a credit card. The objective was to understand the average number of products used by customers who have a credit card.
* To achieve this, a SQL query is being written that joins relevant tables and calculates the average number of products (NumOfProducts) used by customers who are credit card holders (cc.Category = 'credit card holder').
* The SQL query returned an average of **1.5224**. This implies that, on average, customers who have a credit card use approximately 1 to 2 products, with a slight inclination towards 2. This insight can help understand the product usage pattern among our credit card holding customers and can guide us in tailoring our product offerings and marketing strategies.

**SQL Query:**

**SELECT**

**AVG(NumOfProducts) AS avg\_product\_by\_credit\_card**

**FROM**

**bank\_churn bc**

**LEFT JOIN**

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

**INNER JOIN**

**gender gen ON ci.GenderID = gen.GenderID**

**INNER JOIN**

**exit\_customer ec ON ec.ExitID = bc.Exited**

**INNER JOIN**

**credit\_card cc ON cc.CreditID = bc.Has\_creditcard**

**INNER JOIN**

**geography geo ON geo.GeographyID = ci.GeographyID**

**INNER JOIN**

**active\_customer ac ON ac.ActiveID = bc.IsActiveMember**

**WHERE**

**cc.Category = 'credit card holder';**



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

Ans: For the most recent year i.e., 2019 the churn rate for male category is 15.4 whereas for the female category the churn rate is 25.0.

* A DAX measure named Exit Rate was created using the formula:

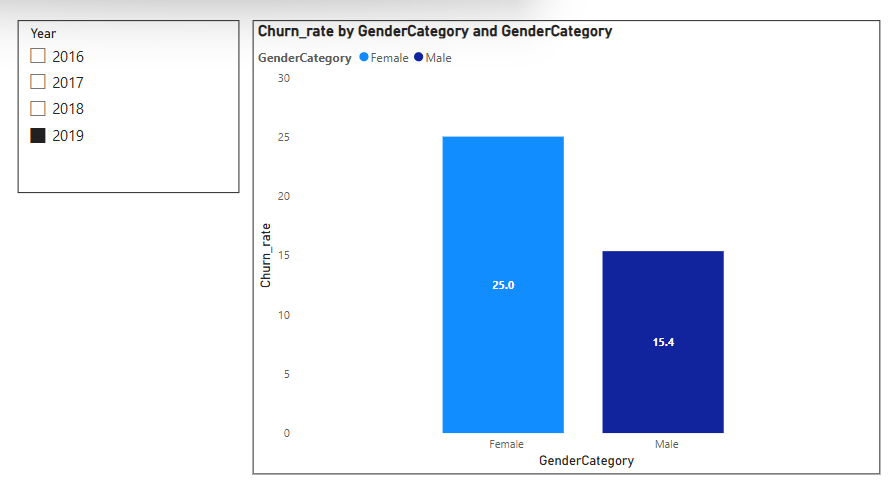
**Exit Rate =**

**DIVIDE(**

**CALCULATE(COUNT('Bank\_Churn'[CustomerId]), Bank\_Churn[Exited] = 1) \* 100,**

**COUNT(Bank\_Churn[CustomerId])**

**)**

* This will calculate the Exit rate in percentage. Then using the KPI card and applying the slicer of year with the most recent year and gender with male and female.

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

Ans:

* The SQL query provided compares the average credit score of customers who have exited and those who remain. The query joins the bank\_churn and exit\_customer tables on the Exited and ExitID fields respectively. It then groups the results by the ExitCategory and calculates the average CreditScore for each group.
* This indicates that the average credit score of customers who have exited is 645.3515, while the average credit score of customers who remain is 651.8532. Therefore, on average, customers who remain have a slightly higher credit score compared to those who have exited.

Code: **SELECT**

**ec.ExitCategory, avg(bc.CreditScore) as avg\_credit\_score**

**FROM**

**bank\_churn bc**

**INNER JOIN**

**exit\_customer ec ON bc.Exited = ec.ExitID**

**group by ec.ExitCategory**

**;**

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

Ans**: WITH ActiveAccounts AS (**

**SELECT CustomerId,COUNT(\*) AS ActiveAccounts**

**FROM Bank\_Churn**

**WHERE IsActiveMember = 1**

**GROUP BY customerId**

**)**

**SELECT CASE WHEN c.GenderID = 1 THEN 'Male' ELSE 'Female' END AS Gender,**

**COUNT(aa.CustomerId) AS ActiveAccounts, round(AVG(c.EstimatedSalary),2) AS AvgSalary**

**FROM customer\_info c**

**LEFT JOIN ActiveAccounts aa ON c.CustomerId = aa.CustomerId**

**GROUP BY Gender**

**ORDER BY AvgSalary DESC;**

* The query first creates a temporary table ActiveAccounts that counts the number of active accounts per customer. An account is considered active if the IsActiveMember field is 1.
* Then, the main part of the query joins this ActiveAccounts table with the customer\_info table on the CustomerId field. It groups the results by gender and calculates two metrics for each group:
* The number of active accounts (COUNT(aa.CustomerId))
* The average estimated salary (AVG(c.EstimatedSalary))

The results are then ordered by the average estimated salary in descending order.

The output of the query shows that:

* Females, with **2284** active accounts, have an average estimated salary of **$100,601.54**.
* Males, with **2867** active accounts, have an average estimated salary of **$99,664.58**.

From these results, we can infer that while males have more active accounts, females have a slightly higher average estimated salary. This suggests that there may not be a direct correlation between the number of active accounts and the average estimated salary for each gender



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

**Ans: WITH credit\_score\_segments AS (**

**SELECT**

**customerid, isactivemember,**

**CASE**

**WHEN creditscore between 800 and 850 THEN 'Excellent'**

**WHEN creditscore between 740 and 799 THEN 'Very Good'**

**WHEN creditscore between 670 and 739 THEN 'Good'**

**WHEN creditscore between 580 and 669 THEN 'Fair'**

**ELSE 'Poor'**

**END AS credit\_score\_segment**

**FROM bank\_churn**

**)**

**SELECT**

**credit\_score\_segment,**

**AVG(CASE WHEN isactivemember = 0 THEN 0 ELSE 1 END) AS exit\_rate**

**FROM credit\_score\_segments**

**GROUP BY credit\_score\_segment**

**ORDER BY exit\_rate DESC**

**LIMIT 1;**



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

**Ans: SELECT**

**geo.GeographyLocation,**

**COUNT(ac.ActiveCategory) AS active\_member\_count**

**FROM**

**bank\_churn bc**

**LEFT JOIN**

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

**INNER JOIN**

**gender gen ON ci.GenderID = gen.GenderID**

**INNER JOIN**

**exit\_customer ec ON ec.ExitID = bc.Exited**

**INNER JOIN**

**credit\_card cc ON cc.CreditID = bc.Has\_creditcard**

**INNER JOIN**

**geography geo ON geo.GeographyID = ci.GeographyID**

**INNER JOIN**

**active\_customer ac ON ac.ActiveID = bc.IsActiveMember**

**where**

**ac.ActiveCategory = 'Active Member'**

**AND bc.Tenure > 5**

**GROUP BY geo.GeographyLocation**

**ORDER BY ac.ActiveCategory DESC**

**limit 1**

**;**

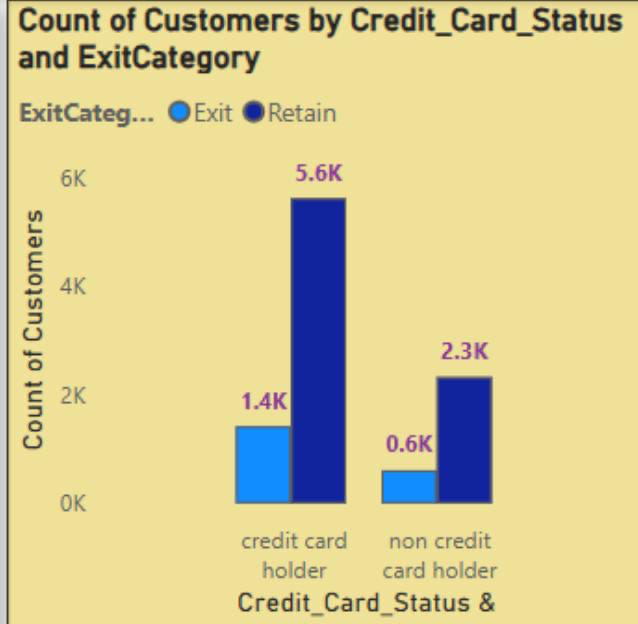


1. **What is the impact of having a credit card on customer churn, based on the available data?**

**Ans:**

* Customers who do not have a credit card (HasCrCard = 0) show a lower count of churn (Exited = 1) compared to those who have stayed (Exited = 0).
* Similarly, customers who have a credit card (HasCrCard = 1) also show a lower count of churn compared to those who have stayed.
* However, in absolute numbers, the group of customers with a credit card has both a higher count of customers who have stayed and a higher count of customers who have churned.

**Conclusion:** Based on the available data, it can be observed that while customer churn occurs regardless of whether a customer has a credit card or not, the group of customers with a credit card shows a larger absolute number of both retained and churned customers.

****

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

**Ans: SELECT**

**count(bc.CustomerID) , NumOfProducts**

**FROM**

**bank\_churn bc**

**LEFT JOIN**

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

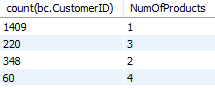
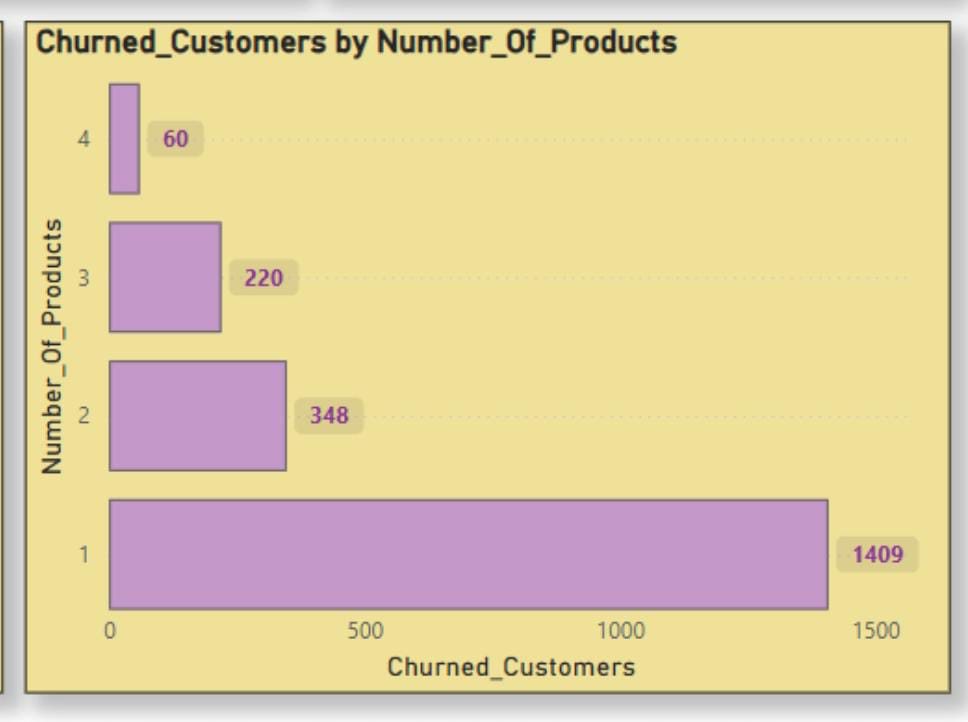
**INNER JOIN**

**exit\_customer ec ON ec.ExitID = bc.Exited**

**where bc.Exited = 1**

**group by NumOfProducts**

**;**

* The above graph provides the insights of customers who have exited by number of products.
* The graph shows data for 1 to 4 products. There is no data for 0 products.
* The number of customers who have exited is highest for 1 product (approximately 1409), and lowest for 4 products (around 60).

**Conclusion**:

* The customers who use 3 or 4 products are less likely to churn.
* The customers who use just 1 product are the customers who have churned the most. So, focusing on those customer bases can help to decrease the churn rate.

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

**Ans: SELECT**

**count(CustomerID),**

**extract(month from str\_to\_date(Bank\_DOJ,'%d-%m-%Y')) as "month\_Joining" ,**

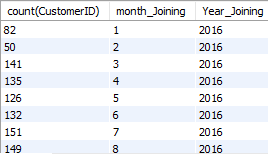
**extract(year from str\_to\_date(Bank\_DOJ,'%d-%m-%Y')) as "Year\_Joining"**

**FROM**

**customer\_info**

**group by month\_Joining, Year\_Joining**

**order by Year\_Joining,month\_Joining ;**





* **Time Period**: The graph represents data from2016 to 2019.
* **Customer Count Fluctuations**: The count of new customers shows several fluctuations over this period, indicating varying levels of customer engagement or business growth.
* **Notable Peaks**: Quarter 4 is the strongest quarter every year, with the highest number of new customers. 2019 saw the highest growth, ending with 1214 new customers in quarter 4.
* **Overall Trend**: Despite the fluctuations, there seems to be an overall upward quarterly trend starting lower in Q1 and peaking in Q4.
* **Conclusion:** The number of customers has increased per year in the bank, which is a good sign. Q4 consistently outperforms other quarters, suggesting effective year-end campaigns or seasonal banking behaviour.It is showing customers satisfaction with the bank services and products. From the data above, we could say that it is almost double in the 4 years of span.

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

**Ans: SELECT**

**NumOfProducts, round(sum(Balance),0) as SumOFBalanace**

**FROM**

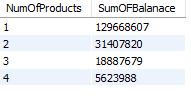
**bank\_churn bc**

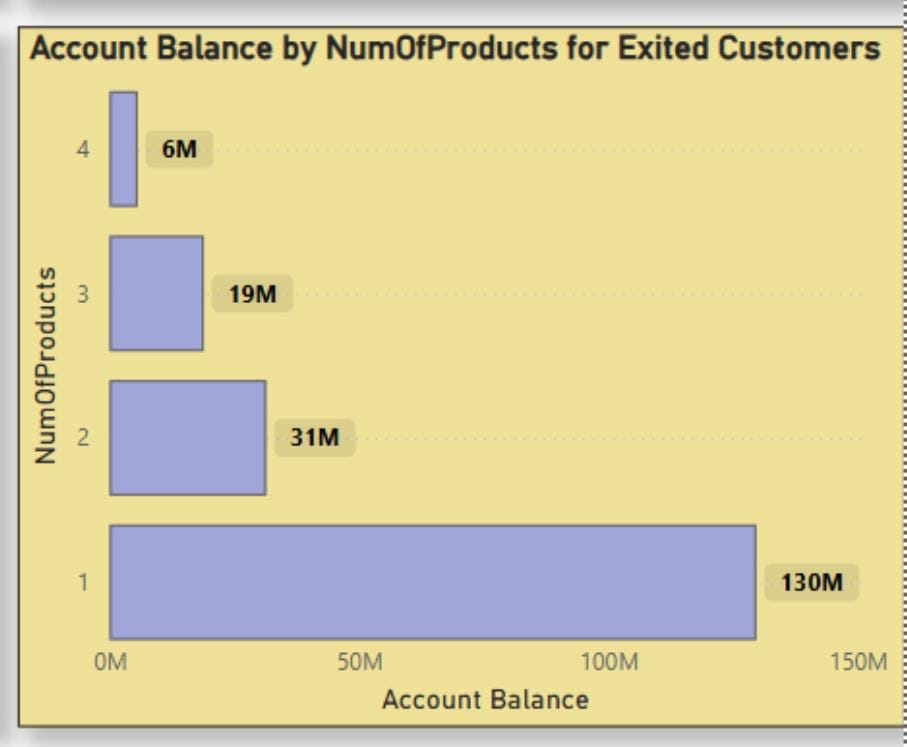
**where bc.Exited = 1**

**group by NumOfProducts**

**order by SumOFBalanace desc**

**;**



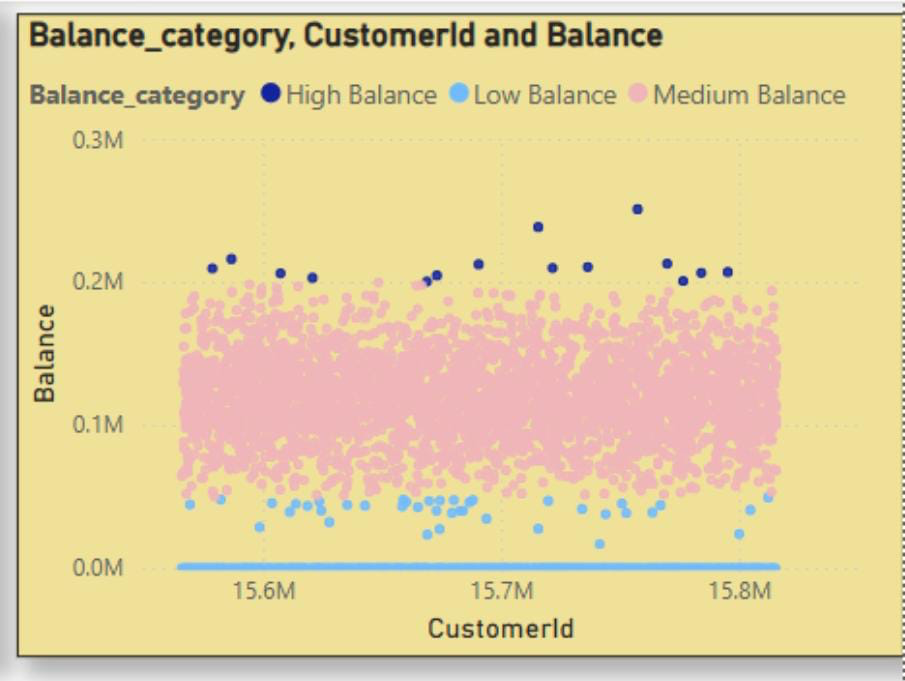


* The graph represents data for 1 to 4 products.
* The most common number of products used by exiting customers is 1. This suggests that the customers who use the least number of products are more likely to churn.
* As the number of products used increases. This suggests that customers who churned tend to have fewer products compared to active customers.
* The customers who have used 4 products have less balance.

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

**Ans:** A scatter plot of CustomerId versus Balance, categorized by Balance\_category, was used to identify potential outliers among customers who have been retained.

* **High Balance (dark blue dots):** These are the key outliers customers with balances exceeding 200,000. They are significantly fewer in number compared to those in the medium and low balance categories. Visibly distinct from the dense central cluster, these customers appear at the top of the chart, typically in the 220,000 to 300,000 range.
* **Medium Balance (pink dots):** This is the densest group, with balances ranging from 50,001 to 200,000, and it does not display any clear outliers, as it represents the dominant cluster.
* **Low Balance (light blue dots):** A few customers have very low or zero balances, which may be considered outliers at the lower end. However, given the large number of low-balance customers, only those with zero or near-zero balances truly stand out as outliers.



### **Conclusion:**

* High-balance customers (Balance > 200,000)are the most significant outliers.
* Zero-balance customersmay also be considered outliers at the lower end.

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

**Ans:** There are seven different tables i.e., ActiveCustomer, Bank\_Churn, CreditCard,CustomerInfo,ExitCustomer,Gender,Geography.

Tables with categorical variables:

* CustomerInfo:Contains categorical variables like Surname.
* ExitCustomer: Contains categorical variables like Exit Category(Exit ,Retain).
* Gender: Contains categorical variables like Gender Category (Male,Female).
* Geography:Contains categorical variables like Geography Location (France, Spain,Germany).
* ActiveCustomer: Contains categorical variables like Active Category (Active Member , Inactive Member).
* CreditCard: Contains categorical variables like Category (Credit-card holder , Non-Credit card holder).

**Conclusion:** There is no such table which contains only categorical variables.

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)**

**Ans: select**

**ci.GeographyID,**

**geo.GeographyLocation,**

**avg(ci.EstimatedSalary) as average\_salary,**

**gen.GenderCategory,**

**rank() over(partition by ci.GeographyID order by avg(ci.EstimatedSalary) desc ) gender\_rank**

**from bank\_churn bc**

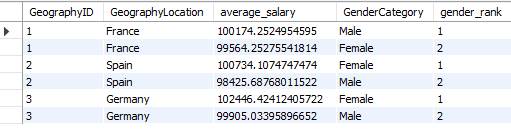
**left join customer\_info ci on ci.CustomerId=bc.CustomerId**

**inner join gender gen on ci. GenderID= gen.GenderID**

**inner join geography geo on geo.GeographyID=ci.GeographyID**

**group by gen.GenderCategory, ci.GeographyID,geo.GeographyLocation**

**;**



* The SQL query provided is designed to calculate the average income of males and females in each geography id. It also ranks the genders according to the average income value. The query uses a combination of SELECT, FROM, LEFT JOIN, INNER JOIN, GROUP BY, and window function RANK() to achieve this.
* The SELECT statement is used to select the required columns from the tables. It includes ci.GeographyID, geo.GeographyLocation, avg(ci.EstimatedSalary) as average\_salary, gen.GenderCategory, and rank(), over ci.GeographyID, ordered by avg(ci.EstimatedSalary) in descending order as gender\_rank.
* The FROM clause specifies the bank\_churn table as the main table. The LEFT JOIN and INNER JOIN clauses are used to combine rows from customer\_info, gender, and geography tables based on common columns.
* The GROUP BY clause groups the result set by gen.GenderCategory, ci.GeographyID, and geo.GeographyLocation.
* The output of the query shows the GeographyID, GeographyLocation, average\_salary,GenderCategory and gender\_rank. For example, in France (Geography ID 1) the average salary of males is higher than that of females, hence males are ranked 1. Similarly, in Spain ( Geography ID 2) and Germany (Geography ID 3) females have a higher average salary, hence they are ranked 1.

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

**Ans: with bracket as (**

**select bc.Tenure ,**

**case when ci.Age between 18 and 29 then "18-30"**

**when ci.Age between 30 and 49 then "30-50"**

**else "50 and above"**

**end as age\_bracket**

**from bank\_churn bc**

**left join customer\_info ci on bc.CustomerID=ci.CustomerID**

**)**

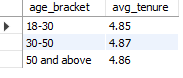
**select age\_bracket, round(avg(Tenure),2) avg\_tenure**

**from bracket**

**group by age\_bracket**

**order by age\_bracket asc**

**;**



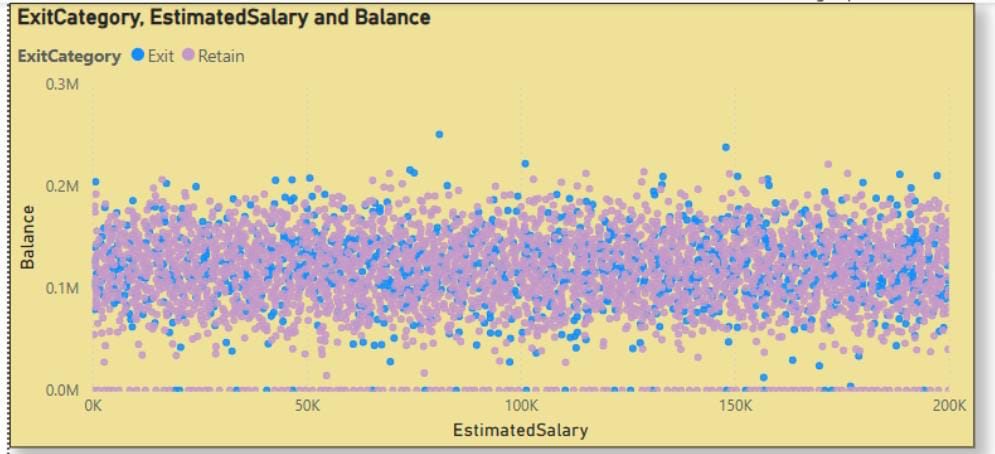
* The query first creates a temporary table bracket that includes the tenure and age\_bracket for each customer. The age\_bracket is determined based on the age column from the customer\_info table.
* The query then calculates the average tenure (avg\_tenure) for each age\_bracket. The results are as follows:
* For the age bracket 18-30, the average tenure of exited customers is 4.85.
* For the age bracket 30-50, the average tenure of exited customers is 4.87.
* For the age bracket 50 and above, the average tenure of exited customers is 4.86.
* The average tenure of exited customers is quite similar across all age brackets, indicating that age might not be a significant factor influencing the tenure of customers.
* Additional factors could be investigated to understand their impact on customer tenure, such as account balance, number of products, credit score, etc. This could provide a more comprehensive understanding of customer behaviour and help in developing effective customer retention strategies.

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

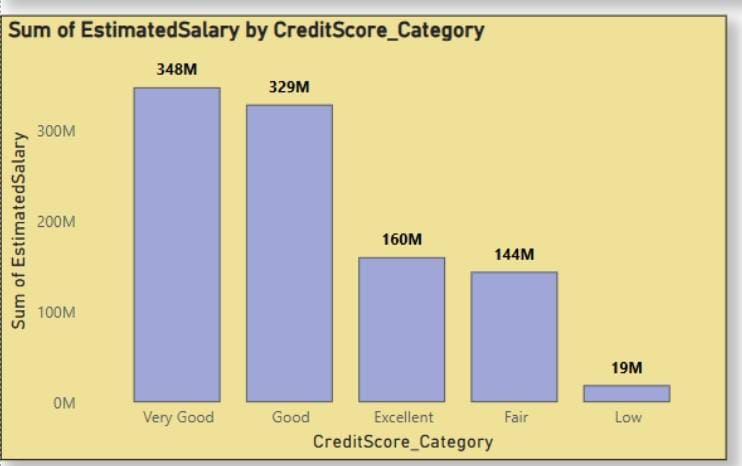
**Ans:** The given scatter plot examines the correlation between EstimatedSalary (X-Axis) vs. Balance (Y-Axis), with ExitCategory (Exited vs. Retained) shown in different colors.

**Key Observations:**

* **Lack of Correlation:** The data points are scattered randomly without a discernible upward or downward trend, indicating no significant relationship between Estimated Salary and Balance.
* **Exited vs. Retained Customers:** Blue (Exited) and Pink (Retained) data points are distributed across all salary and balance ranges, with no observable trend suggesting that customers with higher or lower salaries tend to have correspondingly higher or lower balances upon exit.
* **Low-Balance Customers:** Customers with zero or very low balances are present across various salary levels.
* **High-Balance Outliers:** A few customers have balances exceeding 200K, potentially indicating outliers



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

**Ans:** Below DAX function is created to categorize the Credit Score of customers for the available data:  
**CreditScoreCategory = SWITCH(TRUE(),  
Bank\_Churn[CreditScore]<= 450, "Low",  
Bank\_Churn[CreditScore] <= 550, "Fair",  
Bank\_Churn[CreditScore] <= 650, "Good",  
Bank\_Churn[CreditScore] <= 750, "Very Good",  
Bank\_Churn[CreditScore] > 750, "Excellent")  
**

* A positive correlation is observed between Estimated Salary and Credit Score.
* Customers in higher credit score categories (e.g., Very Good and Good) account for a significantly larger share of total estimated salaries, while those in lower categories (Fair and Low) contribute considerably less.
* This indicates that individuals with stronger credit profiles generally have higher income levels.

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

**Ans:** The chart ranks credit score categories by the number of churned customers (highest to lowest):

* Good: 689 customers
* Very Good: 667 customers
* Excellent: 313 customers
* Fair: 307 customers
* Low: 61 customers



**Conclusion**: Interestingly, the 'Good' and 'Very Good' categories despite having stronger credit scores exhibit the highest churn rates, highlighting a counterintuitive trend worth further exploration.

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

**Ans:**

**SQL Query:**

**WITH creditinfo AS (**

**SELECT**

**CASE**

**WHEN age BETWEEN 18 AND 30 THEN 'Adult'**

**WHEN age BETWEEN 31 AND 50 THEN 'Middle-Aged'**

**ELSE 'Old-Aged'**

**END AS AgeBrackets,**

**COUNT(c.CustomerId) AS HasCrCard**

**FROM customerinfo c**

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

**WHERE b.HasCrcard = 1 -- Ensures filtering is done before counting**

**GROUP BY AgeBrackets**

**)**

**SELECT \***

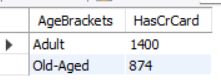
**FROM creditinfo**

**WHERE HasCrCard < (**

**SELECT AVG(HasCrCard)**

**FROM creditinfo**

**);**



The sql query above calculates the number of customers holding credit cards, segmented by age brackets. It identifies age groups with a below-average number of credit card holders compared to the overall distribution.

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

**Ans: select**

**geo.GeographyLocation,**

**count(bc.Exited) as churned,**

**round(avg(bc.Balance),2) average\_balance,**

**rank() over(order by count(bc.Exited) desc) as ranking**

**from bank\_churn bc**

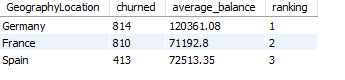
**left join customer\_info ci on ci.CustomerId=bc.CustomerId**

**inner join geography geo on geo.GeographyID=ci.GeographyID**

**where bc.Exited=1**

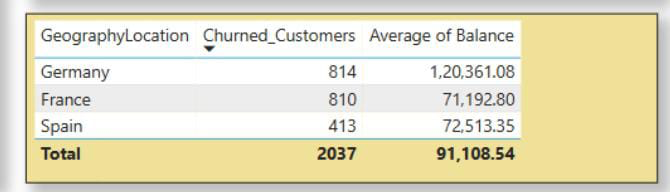
**group by geo.GeographyLocation**

**;**



* **Germany**: 814 customers exited, with an average balance of approximately 120361.08.
* **France**: 810 customers exited, with an average balance of approximately 71192.8.
* **Spain**: 413 customers exited, with an average balance of approximately 72513.35.

**Power BI:**

****

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”.**

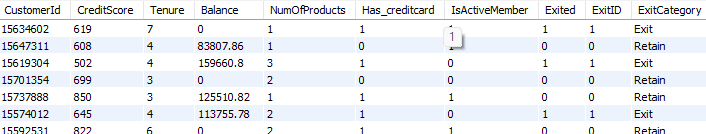
**Ans: select concat(CustomerID, ' ' ,Surname) as CustomerID\_Surname from customer\_info;**



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

**Ans:** The given code retrieves customer data from the Bank\_Churn table and adds a new column, 'ExitCategory', which classifies customers as 'Retain' (not churned) or 'Exit' (churned) based on the 'Exited' column, using a subquery.

**select \* from bank\_churn,exit\_customer where exit\_customer.ExitID=bank\_churn.Exited;**



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

**Ans:** No missing values were present in the dataset, eliminating the need for imputation techniques that could introduce assumptions or biases.

However, if missing values had been detected, the following strategies would have been considered:

* **Deletion:** Remove rows or columns containing missing values when appropriate.
* **Imputation:** Fill missing values using statistical methods such as mean, median, or inference from related columns.
* **Default Value Replacement:** Assign a predefined default value based on domain knowledge or business logic.

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

**Ans: SELECT**

**bc.CustomerID, ci.Surname AS Last\_Name, ac.ActiveCategory**

**FROM**

**bank\_churn bc**

**LEFT JOIN**

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

**INNER JOIN**

**gender gen ON ci.GenderID = gen.GenderID**

**INNER JOIN**

**exit\_customer ec ON ec.ExitID = bc.Exited**

**INNER JOIN**

**credit\_card cc ON cc.CreditID = bc.Has\_creditcard**

**INNER JOIN**

**geography geo ON geo.GeographyID = ci.GeographyID**

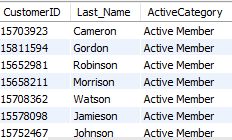
**INNER JOIN**

**active\_customer ac ON ac.ActiveID = bc.IsActiveMember**

**WHERE**

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

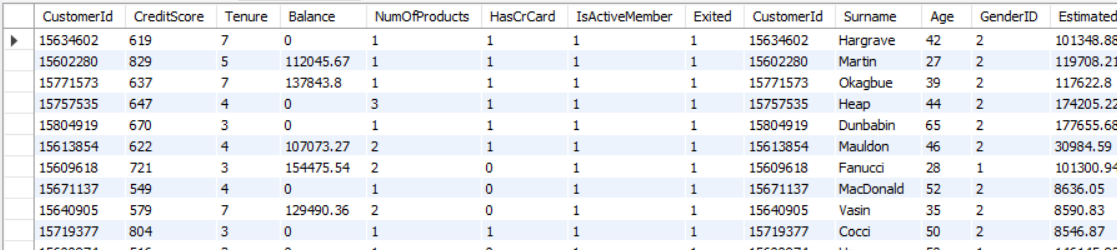
**;**



1. **Can you observe any data disrupency in the Customer’s data? As a hint it’s present in the IsActiveMember and Exited columns. One more point to consider is that the data in the Exited Column is absolutely correct and accurate.**

**Ans:** Yes, there is a data discrepancy in the customer records. Ideally, an active customer should have 'IsActiveMember' marked as 1, and an exited customer should have 'Exited' marked as 1.

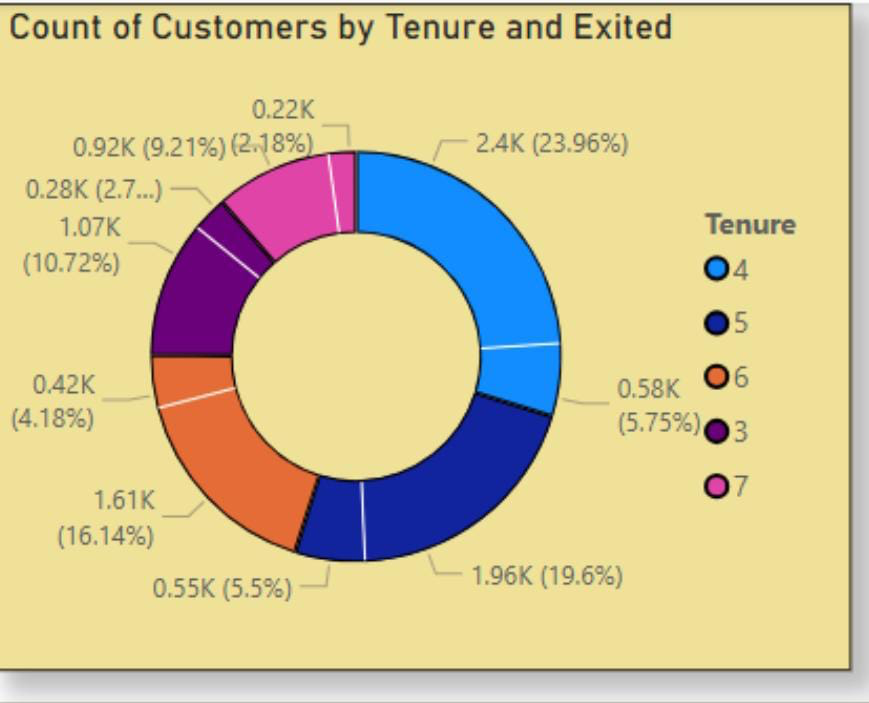
However, the dataset contains instances where both 'Exited' and 'IsActiveMember' are marked as 1 simultaneously indicating that a customer has exited the bank but is still flagged as active. This inconsistency highlights a data quality issue.



**Subjective Question:**

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

**Ans:**



**The exit rate by customer tenure is as follows:**

* **4 years:** 5.75%
* **5 years:** 5.5%
* **6 years:** 4.18%
* **3 years:** 2.76%
* **7 years:** 2.18%

These insights suggest that customers with longer tenures, particularly those with 7 years, are more loyal and less likely to exit the bank.

**Recommendations**:

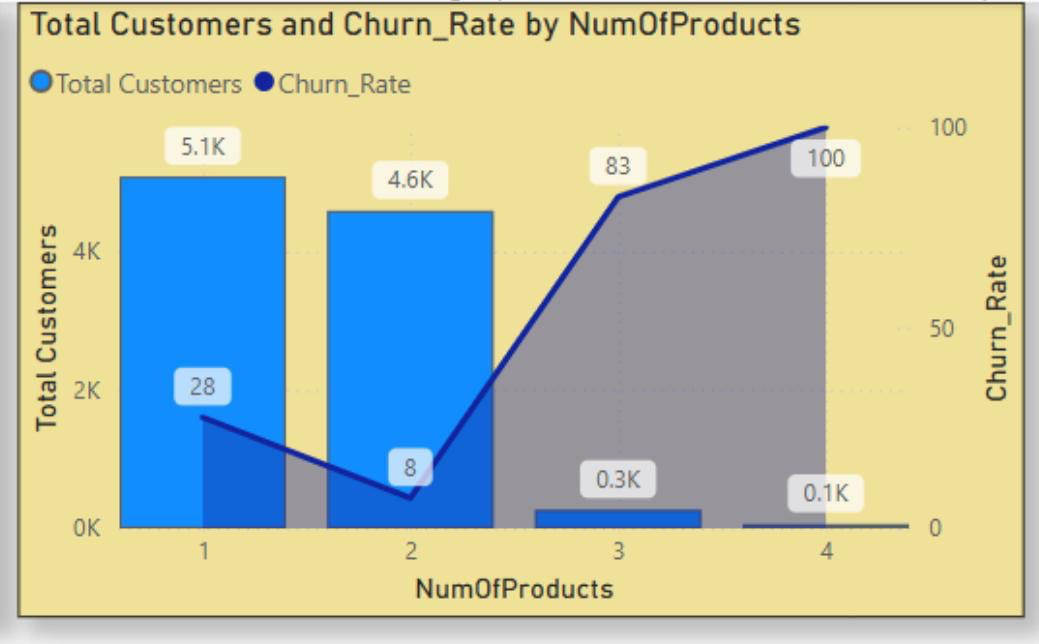
* **Prioritize the 4-Year Tenure Segment:**Closely monitor customers with 4 years of tenure through satisfaction surveys and personalized engagement, as they represent a significant portion of the base but exhibit elevated exit rates.
* **Introduce Tenure-Based Loyalty Programs:**Design targeted incentives for customers approaching 3 and 7 years of tenure to proactively reduce churn at these critical milestones.
* **Leverage Insights from Long-Tenure Customers:**Analyze spending patterns and behavioral trends among long-tenure customers to identify value-driving behaviors and apply those insights to nurture newer segments.
* **Strengthen Early Engagement (0–2 Years):**Implement onboarding and relationship-building initiatives during the first two years to improve the likelihood of customers progressing to more loyal mid-tenure stages.
* **Apply Predictive Churn Modeling:**Use predictive analytics to identify churn risk within key tenure brackets and deploy personalized retention offers before attrition occurs.

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

**Ans:**

**Approach:**

Analyze commonly paired products and service combinations to uncover usage patterns. Since higher product counts correlate with increased churn, refine cross-selling strategies to focus on value-driven bundles that promote engagement and long-term retention rather than simply increasing product count.



**Insights**:

* Customers with fewer products exhibit lower churn rates compared to those with multiple product holdings.
* The majority of customers hold 1 or 2 products (5.1K and 4.6K customers, respectively).
* Churn is lowest (8%) among customers with 2 products.
* Churn rises sharply for customers with 3 products (83%) and 4 products (100%), though these segments are very small.
* Deep cross-selling (3+ products) is rare, indicating limited customer appetite or product fit.
* Higher product count does not guarantee better retention; in fact, it correlates with higher churn.

**Recommendations:**

* Focus cross-sell efforts on single-product customers, encouraging the adoption of a second, well-aligned product.
* Avoid pushing more than two products, as churn risk increases significantly beyond that point.
* Leverage product affinity analysis to identify common and successful product pairings for smarter cross-sell strategies.
* Emphasize value and relevance of product bundles over simply increasing the number of products.
* Collect qualitative feedback from customers with high product counts to understand dissatisfaction drivers.
* Develop a churn prediction model that incorporates product count, engagement levels, and usage patterns to proactively identify at-risk customers.

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

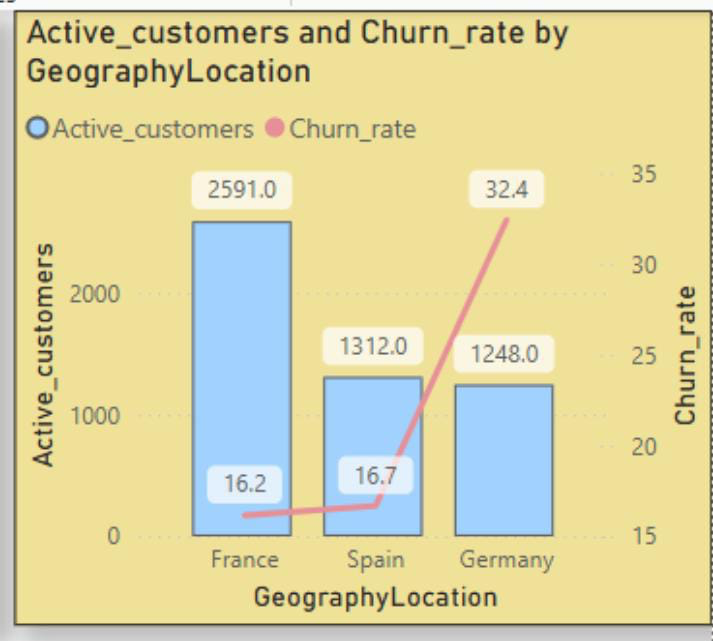
**Ans:** The visual clearly shows that France has the highest number of active members compared to the other two regions, while Germany exhibits the highest churn rate.

**Approach:**

* Analyze regional economic indicators in relation to active account volumes and churn rates.
* Notably, France shows the highest membership, while Germany exhibits elevated churn levels.
* Leverage these insights to develop region-specific retention and growth strategies aligned with local economic conditions and customer behavior.

**Regional Analysis:**

* France has the highest number of active customers (2,591) with a moderate churn rate of 16.2%, indicating strong customer retention and effective engagement strategies.
* Spain maintains 1,312 active customers and a slightly higher churn rate of 16.7%, requiring close monitoring.
* Germany shows the lowest active customer count (1,248) but a significantly high churn rate of 32.4%, suggesting possible dissatisfaction or low engagement despite a similar customer base to Spain.



### **Key Insights:**

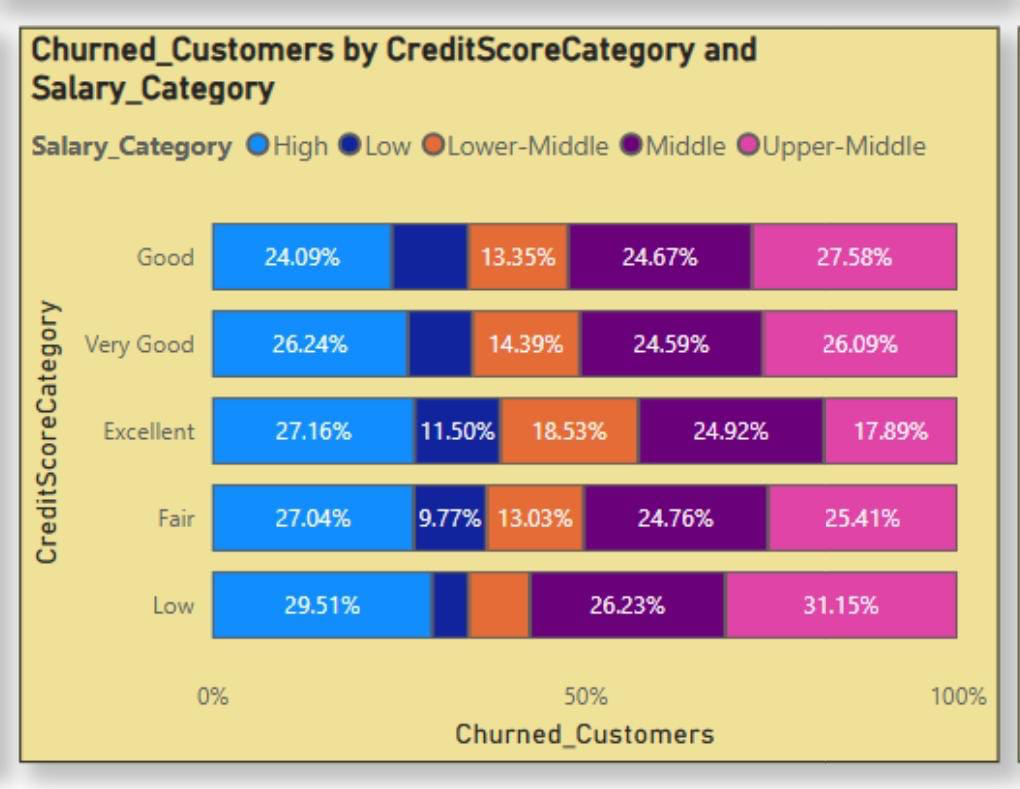
* **Germany’s elevated churn** points to potential service, satisfaction, or engagement gaps.
* **France’s performance** reflects strong retention strategies and potentially favorable economic or service conditions.
* **Spain’s churn** is increasing gradually and may soon require targeted intervention.

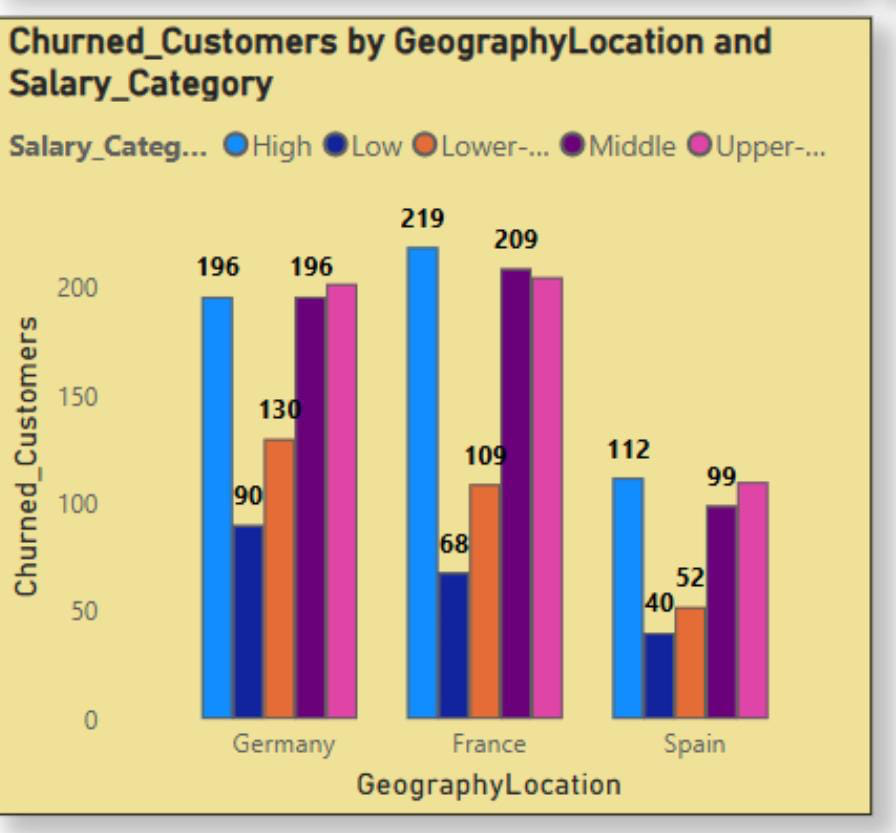
**Recommendations:**

* **Germany**:
* Conduct deep-dive analysis via customer feedback, surveys, and churn drivers.
* Enhance service quality, communication, and localized support.
* Introduce loyalty programs, retention offers, or personalized outreach.
* **France**:
* Identify and scale successful strategies (e.g., product bundling, reward systems, communication channels) to other regions.
* **Spain**:
* Monitor trends closely and prepare preemptive retention efforts if churn continues to rise.
* **Localized Strategies:**
* Tailor retention plans for regional customer behavior, economic indicators, and product preferences.
* Consider competition, cultural factors, and local market dynamics in campaign design.
* **Customer Engagement:**
* Launch targeted loyalty campaigns in high-risk regions (especially Germany) to improve retention and strengthen brand connection.

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

**Ans:**

****

****

**Insights** :

From the above graphs it can easily be interpreted as:

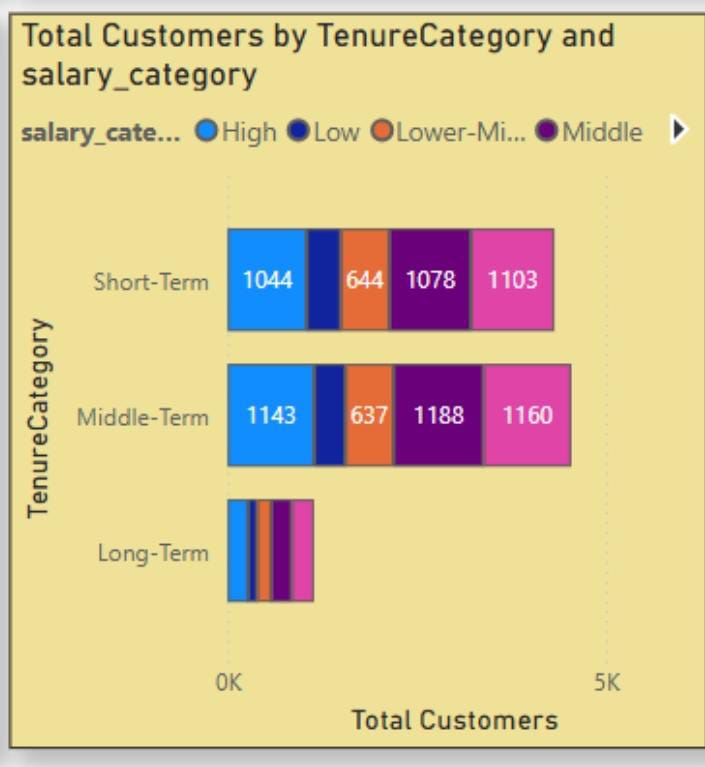
* Germany leads with the highest number of churned customers (814), particularly concentrated in the High, Middle, and Upper-Middle salary segments.
* France follows closely with 810 churned customers, while Spain reports 413.
* Low credit score customers are heavily represented in the Middle and Upper-Middle salary brackets.
* Notably, even customers with Good and Very Good credit scores exhibit churn within Middle and Upper-Middle salary ranges.
* High-income customers are not immune to churn, with losses observed across all credit categories.
* Middle-income customers span all risk levels, suggesting underlying retention risks that may not be obvious from salary or credit score alone.

**Conclusion:**

* While low-salary, low-credit score customers represent a clear risk segment, higher-salary customers with fair to good credit scores also warrant attention due to their substantial volume and observed churn.
* The bank should expand its monitoring to include high-income segments with declining or mid-tier credit profiles, as these customers pose a less obvious but significant risk for unexpected attrition.

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

**Ans:**



### **Insights: Customer Tenure Value Forecast**

### Short-Term and Middle-Term customers dominate the base retention efforts should be concentrated here.

### Long-Term customers, though fewer, offer high lifetime value; they require tailored nurturing strategies.

### Integrating Salary, Credit Score, and Tenure enhances the effectiveness of churn prediction models.

### Predictive models (e.g., Logistic Regression) can be used to forecast churn likelihood based on tenure segments.

### Upper-Middle salary segments show high potential for long-term value and they should be prioritized.

### Middle-Term customers present a strong opportunity for retention and upselling efforts.

### **Recommendations**

### Retain high-potential Short-Term customers with personalized engagement and early value delivery.

### Upsell and cross-sell to Middle-Term customers to increase tenure and overall profitability.

### Launch loyalty programs for Long-Term customers to maintain sustained revenue and strengthen relationships.

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

**Ans:** To assess the impact of marketing campaigns on customer retention and acquisition within a dataset, a combination of data analysis and statistical techniques can be used. These are some general approach:

1. **Define Metrics:** Define key metrics for customer retention and acquisition. For retention, we might use metrics like customer churn rate or retention rate. For acquisition, we might use metrics like new customer acquisition rate or customer acquisition cost (CAC).
2. **Segment Data:** Segment the data based on different marketing campaigns. This will allow us to analyze the impact of each campaign separately.
3. **Calculate Metrics:** Calculate the defined metrics for each segment and for each time period (e.g., monthly, quarterly, annually). This will help understand how each campaign is affecting customer retention and acquisition over time.
4. **Compare Results:** Compare the metrics across different campaigns to identify which campaigns are most effective at retaining and acquiring customers.
5. **Statistical Analysis:** Use statistical tests (e.g., t-tests, ANOVA) to determine if the differences in metrics between campaigns are statistically significant.
6. **Additional Information:**To perform a comprehensive analysis, some additional information might be needed such as:

* **Customer demographics:** To understand if certain demographics respond better to certain campaigns.
* **Campaign details:** To understand the specifics of each campaign (e.g., duration, channels used, messaging).
* **Competitor data:** To understand the competitive landscape and how it might be impacting your results.
* **External factors:** Such as economic conditions, seasonality, or industry trends that might affect customer behavior.

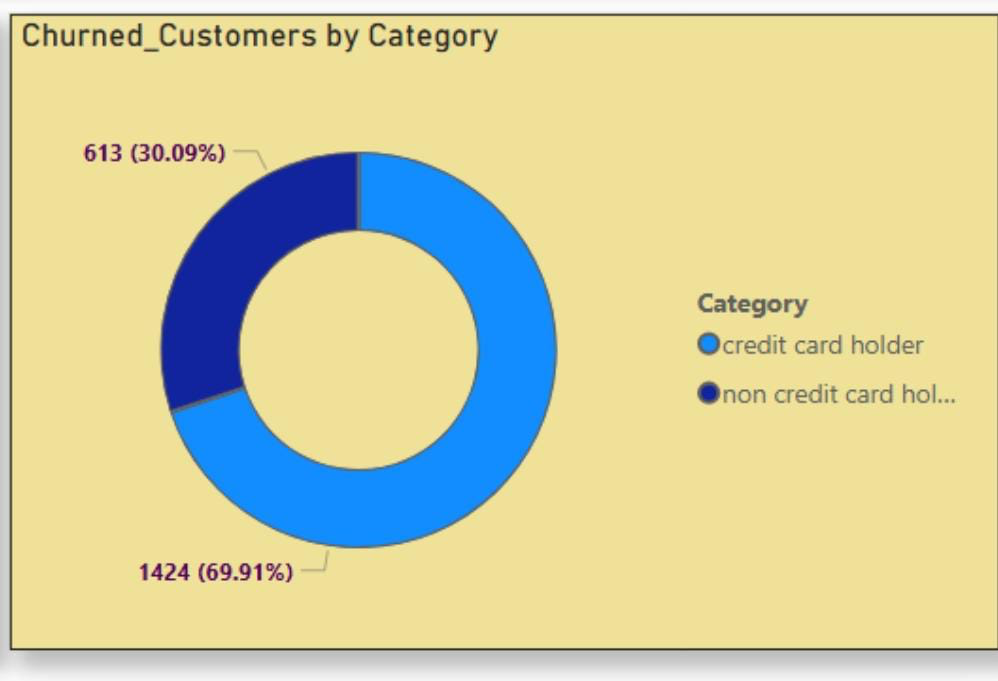
By following this approach and gathering the necessary information, we can assess the impact of marketing campaigns on customer retention and acquisition within the dataset.

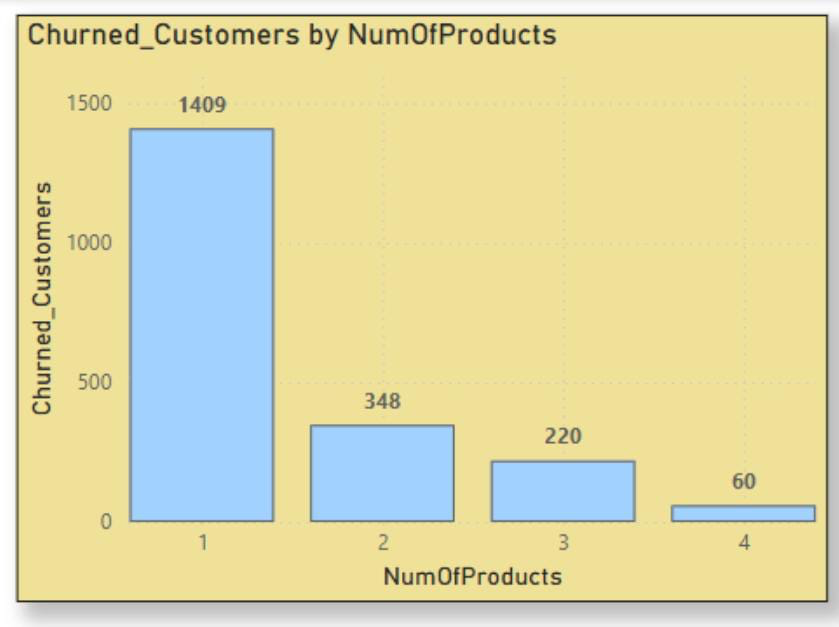
1. **Customer Exit Reasons Exploration: Can you identify common characteristics or trends among customers who have exited that could explain their reasons for leaving?**

**Ans:**  The following chart analysis helps identify common characteristics and trends among customers who have exited, providing insights into potential reasons for their departure.

**Insights:**

* **Credit Card Ownership:**Customers with credit cards show a higher likelihood of churn compared to those without.
* **Number of Products:**Customers with fewer product holdings are more prone to churn than those with multiple product engagements.





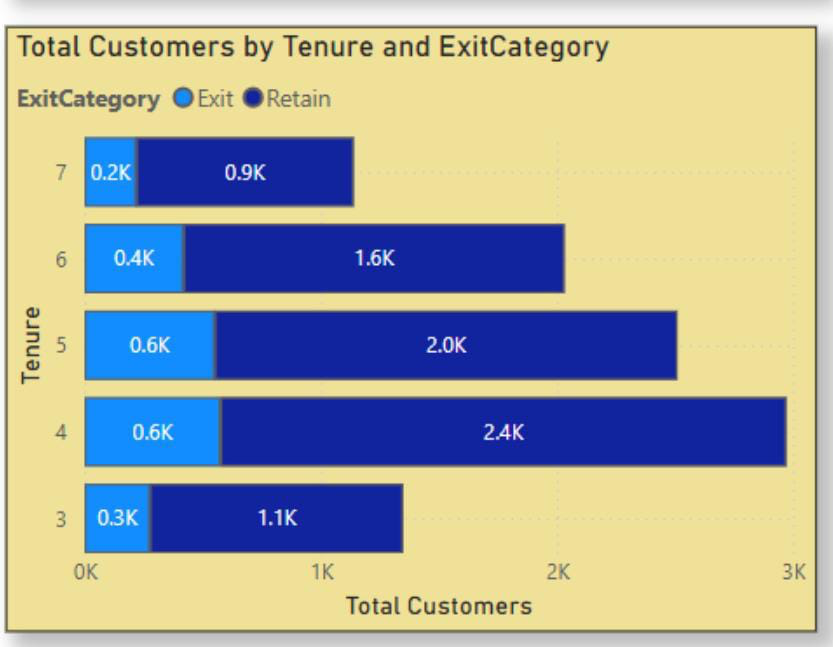
**Conclusion: Key Drivers of Customer Exit:**

* **Credit Card Ownership:**The chart indicates that a significantly higher number of customers with credit cards exited the bank (1,424) compared to those without credit cards (613). This suggests a potential correlation between credit card ownership and churn.
* **Low Product Engagement:**Customers who purchased only one product account for the highest number of exits (1,409). Exit counts decrease progressively as product count increases, with customers holding four or more products showing the lowest churn (approx. 60). This implies that greater product engagement correlates with higher retention.

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

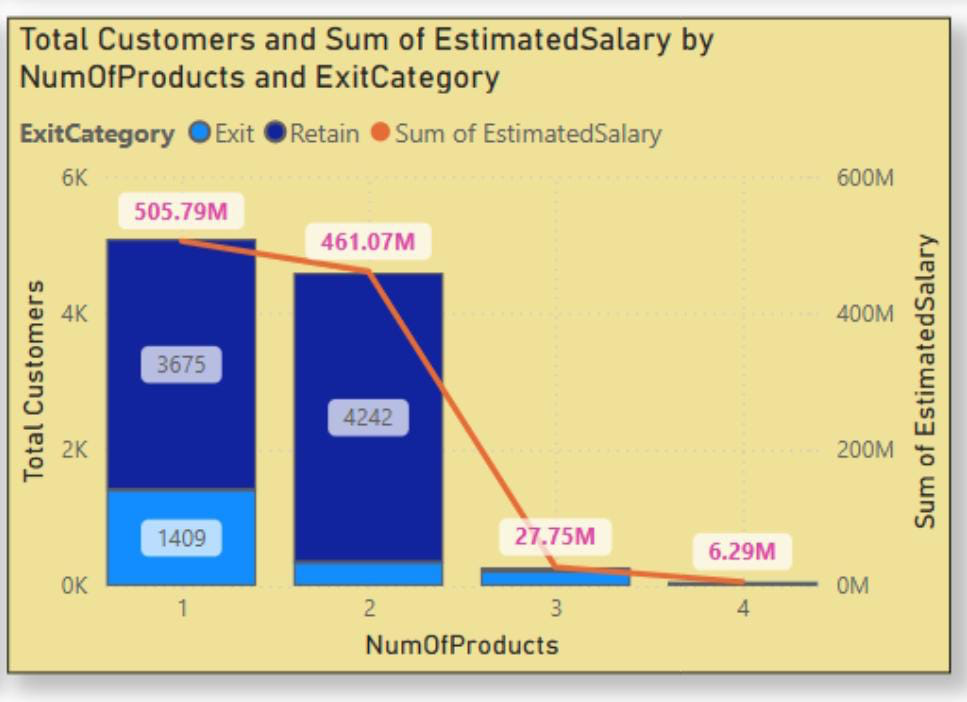
**Ans:** Yes, the given factors are important for predicting if a customer will leave the bank.

* **Tenure:** Customers with moderate tenure (4–6 years) show higher retention, though churn still occurs tenure contributes to retention but is not a standalone predictor.
* **Number of Products:** The majority of exits occur among customers with only 1 or 2 products. As product count increases, churn decreases, although the overall customer count also declines.
* **Estimated Salary:** No strong correlation observed between salary and churn. Exits occur across all salary ranges, making salary a weaker predictor.
* **Activity Status (IsActiveMember):** Active members tend to stay longer, making this a reliable feature for churn prediction and retention strategies.



**Conclusion**:

* Tenure, number of products, and activity status are key predictors of churn.
* Estimated salary, while useful for segmentation, is less impactful as a standalone predictor.



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

**Ans: SELECT**

**CASE**

**WHEN ci.age < 18 THEN 'Under 18'**

**WHEN ci.age BETWEEN 18 AND 24 THEN '18-24'**

**WHEN ci.age BETWEEN 25 AND 34 THEN '25-34'**

**WHEN ci.age BETWEEN 35 AND 44 THEN '35-44'**

**WHEN ci.age BETWEEN 45 AND 54 THEN '45-54'**

**ELSE '55+'**

**END AS age\_group,**

**CASE**

**WHEN bc.Balance < 1000 THEN 'Under 1000'**

**WHEN bc.Balance BETWEEN 1000 AND 5000 THEN '1000-5000'**

**WHEN bc.Balance BETWEEN 5000 AND 10000 THEN '5000-10000'**

**ELSE '10000+'**

**END AS balance\_group,**

**COUNT(ci.CustomerID) AS customer\_count,**

**COUNT(gen.GenderCategory),**

**geo.GeographyLocation**

**FROM**

**bank\_churn bc**

**LEFT JOIN**

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

**INNER JOIN**

**gender gen ON ci.GenderID = gen.GenderID**

**INNER JOIN**

**exit\_customer ec ON ec.ExitID = bc.Exited**

**INNER JOIN**

**credit\_card cc ON cc.CreditID = bc.Has\_creditcard**

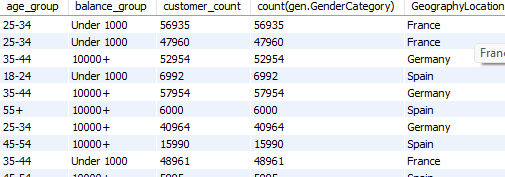
**INNER JOIN**

**geography geo ON geo.GeographyID = ci.GeographyID**

**INNER JOIN**

**active\_customer ac ON ac.ActiveID = bc.IsActiveMember**

**GROUP BY ci.age , geo.GeographyLocation , balance\_group;**

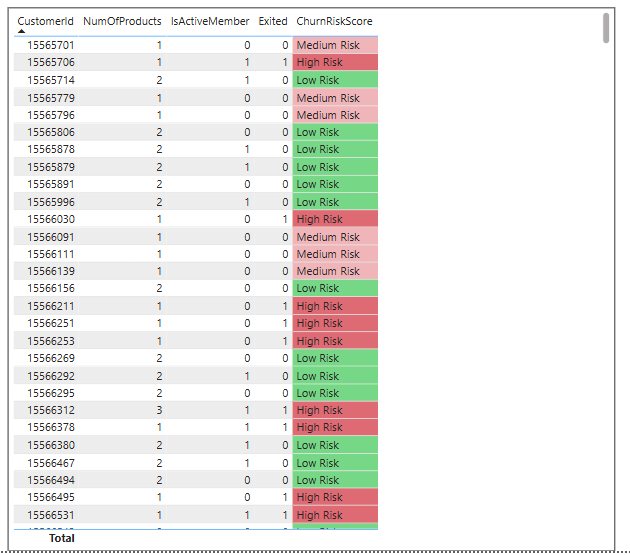


1. **How can we create a conditional formatting setup to visually highlight customers at risk of churn and to evaluate the impact of credit card rewards on customer retention?**

**Ans:** To enable conditional formatting for identifying customers at risk of churn, a new column was created to segment customers into three risk categories.

This classification is based on a combination of the number of products, churn status (Exited), and activity status (IsActiveMember).

**Power BI:**  
**ChurnRiskScore = IF(  
Bank\_Churn[Exited] = 1, "High Risk",  
IF(Bank\_Churn[NumOfProducts] = 1 && Bank\_Churn[IsActiveMember] = 0, "Medium Risk",   
"Low Risk")  
)**

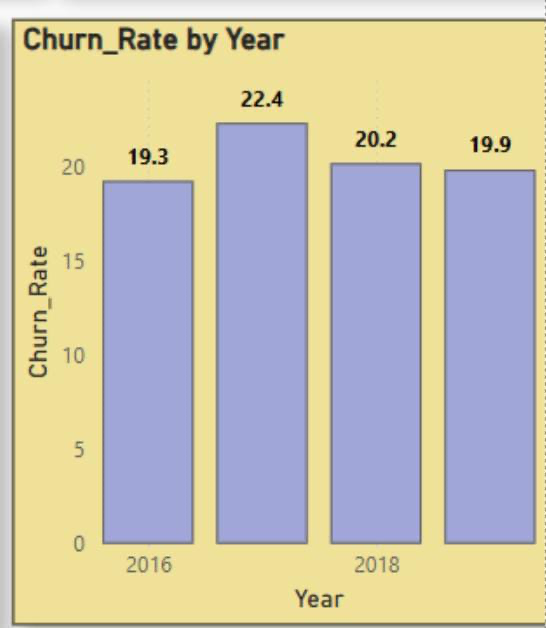
****

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

**Ans:**

1. **Churn Rate:**

* Theoverall churn rate for the bank is 20.37%.
* Year on year churn rates show some fluctuations:
* 2016: 19.3% (Lowest)
* 2017: 22.4% (Highest)
* 2018: 20.2%
* 2019: 19.9%
* This analyzes customer churn rates and identifies segments most susceptible to churn.



1. **Customer Segment prone to churn:** Data analysis suggests a customer segment with a higher likelihood of churn.
2. **Product Purchases:** Customers who only use one product by the bank might not find enough value compared to competitors offering wider ranges or integrated services.
3. **HasCreditCard:** Potential reasons for churn among credit card holders could be:

* Limited credit limits not meeting their needs.
* Lack of rewards programs that incentivize them to keep the card.
* High credit card fees.

1. **Tenure of 4-5 years:** Customers with this tenure might be nearing the end of introductory offers or discounts, making them susceptible to competitor offers with better rates or features.
2. **High Salary:** High earners might have more options and be more likely to switch for a slightly better interest rate or a benefit elsewhere.

**Recommendations to decrease churn rate:**

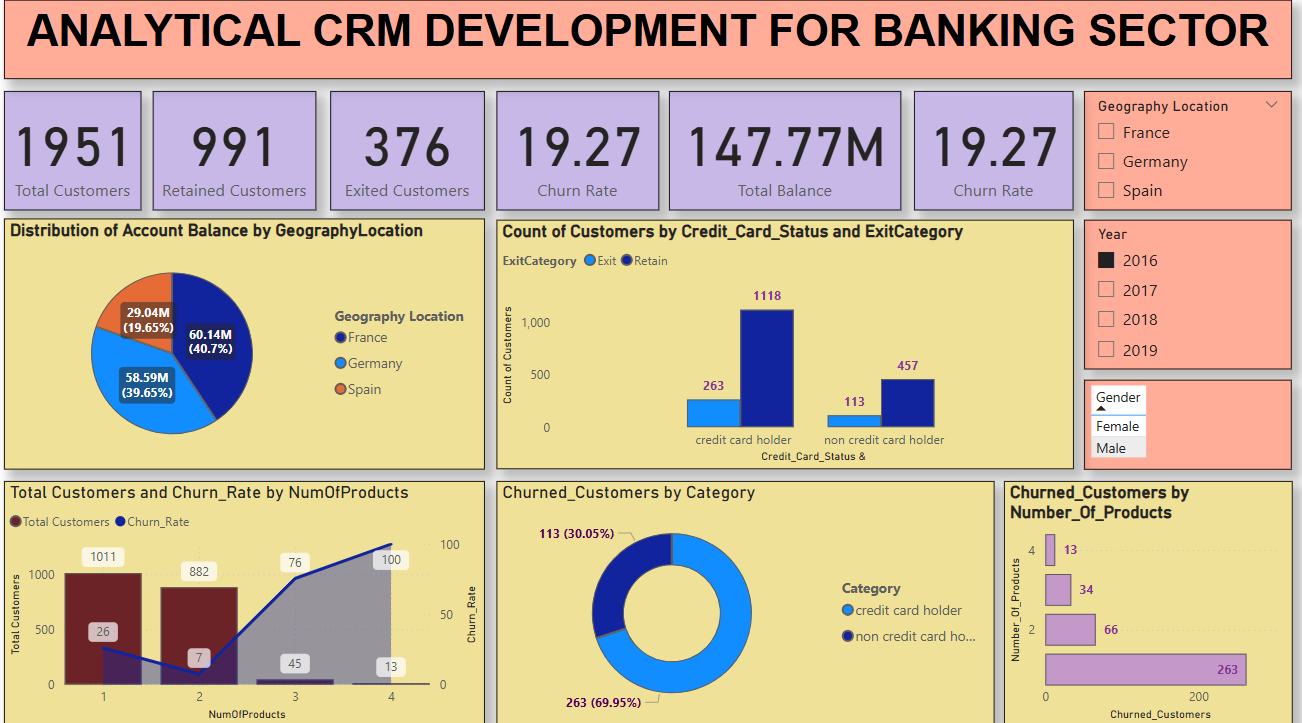
1. **Targeted Product Bundles:** Create product bundles that cater to specific customer segments and needs. Offer these bundles to customers who only use one product, highlighting the additional benefits and potential cost savings.
2. **Enhanced Credit Card Rewards**:Improve credit card rewards programs for exiting customers such as:

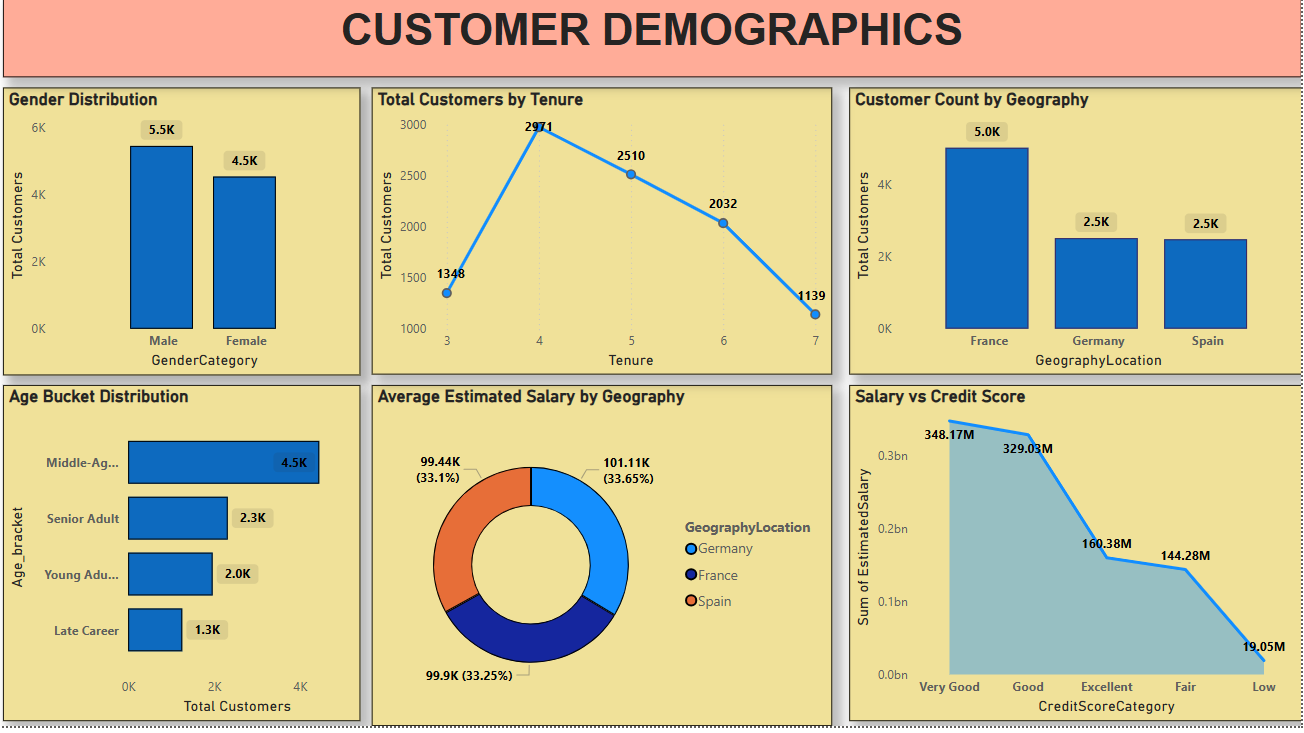
* Increasing credit limits based on customer history and credit worthiness.
* Offering reward programs aligned with spending habits ( e.g., travel rewards, cashback for specific categories).
* Reducing or eliminating annual fees, especially for high-value customers.

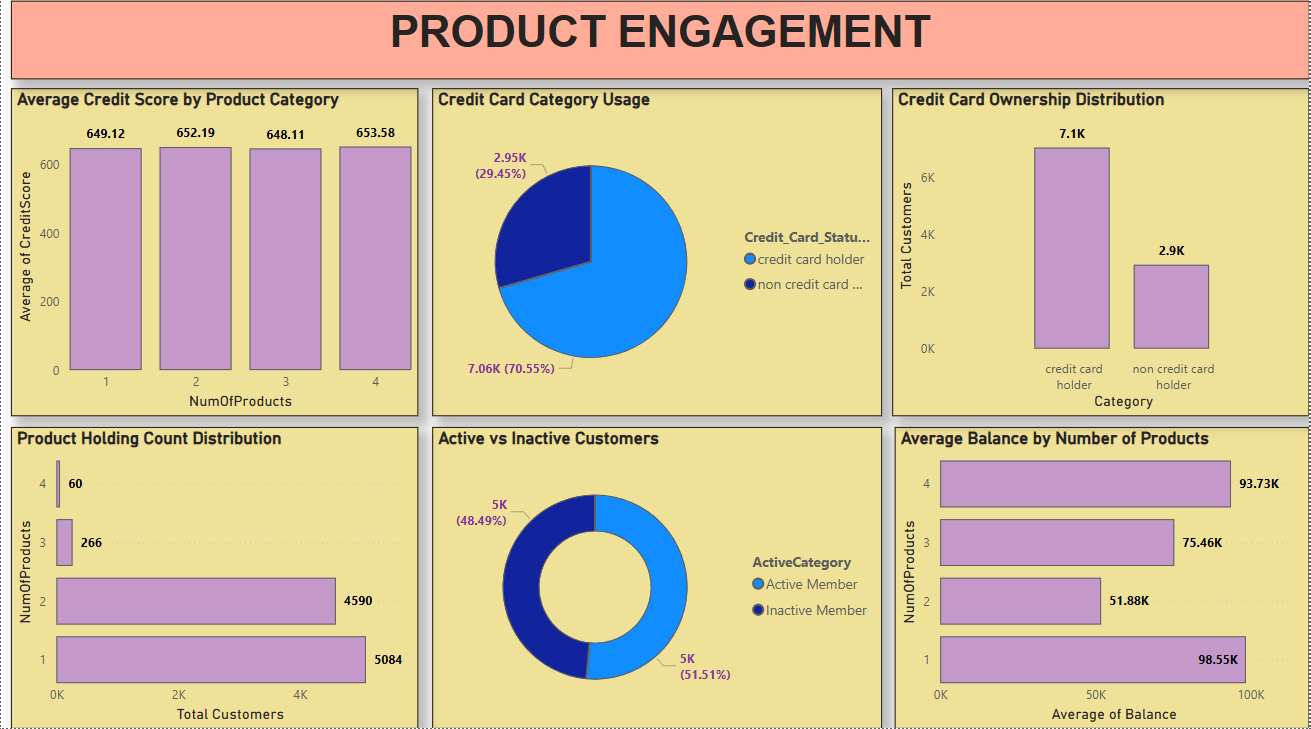
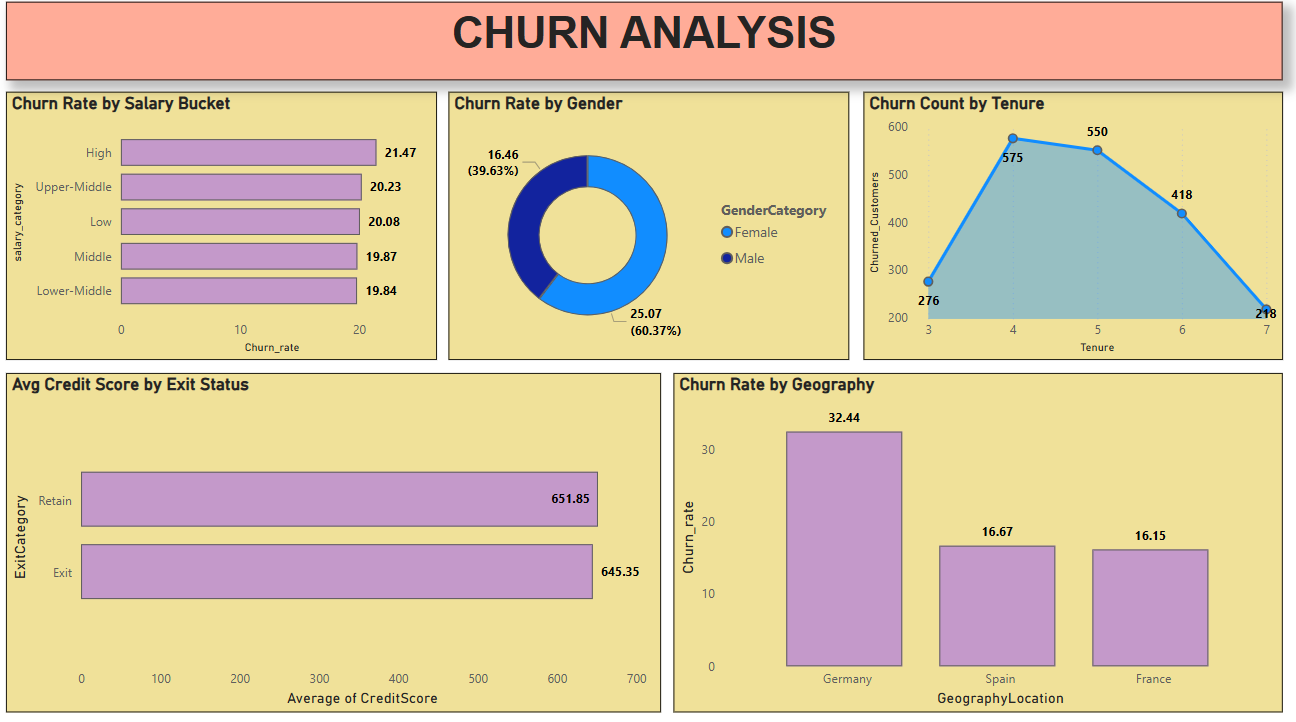
1. **Retention Offers for Existing Customers:** Proactively reach out to the customers nearing the end of introductory offers with personalized retention deals. This includes extending introductory rates or offering discounts on other products or services.
2. **Customer Satisfaction Surveys:** Regularly conduct customer satisfaction surveys to understand why customers churn. This can help identify areas for improvement and tailor retention strategies accordingly.
3. **Relationship Management for High-Value Customers:** Develop dedicated relationship managers for high-value customers to provide personalized service, address their specific needs, and offer exclusive benefits.

By focusing on these strategies, the bank can reduce churn, retain high-value customers, and increase overall customer satisfaction.

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

**Ans:** 





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

**Ans:**

1. **Contextual Understanding:** Begin with a thorough understanding of the analysis context and objectives. Engage with stakeholders or review relevant project documentation to uncover both explicit and implicit goals.
2. **Data Exploration:** Conduct exploratory data analysis (EDA) to familiarize yourself with the dataset. Assess its structure, distributions, and variable relationships, while identifying key patterns, anomalies, and trends.
3. **Insight Identification:**Based on initial exploration, identify potential insights or hypotheses. Consider both quantitative trends and qualitative factors that may influence outcomes within the domain.
4. **Hypothesis Generation:**Formulate hypotheses grounded in preliminary observations. These serve as a foundation for deeper analysis and validation.
5. **Iterative Analysis Process:**Employ a range of analytical techniques, statistical methods, machine learning, or qualitative approaches in an iterative manner. Tools such as Excel, Power BI, and SQL can be leveraged to explore and validate hypotheses effectively.
6. **Visualization and Communication:**Use clear, impactful data visualizations to present findings. Well-designed visuals help reveal insights and convey complex information to stakeholders in an intuitive format.
7. **Synthesis of Findings:**Consolidate findings into coherent narratives or thematic summaries. Highlight key patterns and overarching insights that bridge different aspects of the analysis.
8. **Feedback and Validation:** Validate results through peer review, expert consultation, or external benchmarks when possible. Solicit stakeholder feedback to ensure alignment with business needs and to enhance the relevance and clarity of the insights.
9. **In the “Bank\_Churn” table how can you modify the name of the “HasCrCard” column to “Has\_creditcard”?**

**Ans:** This can be done by using SQL function named Alter which will allow modification, rename column will allow to change the column name. This is done by using a SQL query which is shown in the code given below:

**Alter table bank\_churn**

**rename column HasCrCard to Has\_creditcard;**

**SELECT**

**\***

**FROM**

**Bank\_churn;**

The output shows that the column name “HasCrCard” had been changed to “Has\_creditcard**”.**

