**SUBMISSION BY: PRASANG SHRIVAS**

**PROJECT: BANK CRM PROJECT**

**SKILL USED: MYSQL WORKBENCH + MS POWERBI**

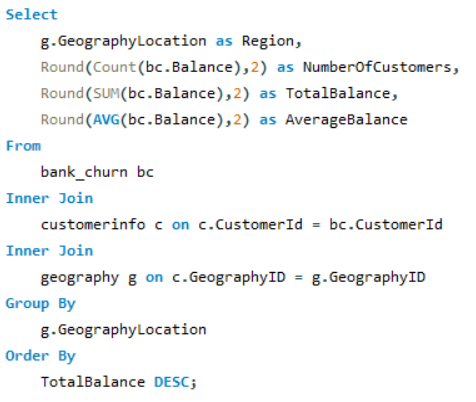
**BATCH: DATA SCIENCE COURSE MAY 2024**

**Objective Questions:**

**QUESTION 1**

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

Query :



Output:

****

Insights:

* Customers with higher credit scores and balances are more likely to stay active, as reflected in the bank\_churn table.
* There may be a geographic trend in customer churn rates, as GeographyLocation in the geography table is linked to the bank\_churn table.
* The number of products a customer holds (NumOfProducts) can be a strong indicator of customer engagement and loyalty.
* Credit card categories may influence customer behavior, especially in relation to churn and account activity.

Recommendations:

* Focus on retaining high-balance, high-credit score customers by offering personalized incentives and services.
* Analyze churn rates across different geographic locations to develop region-specific retention strategies.
* Encourage customers to engage with more products by promoting bundled offerings or cross-sell opportunities.
* Use insights from activecustomer and exitcustomer tables to improve customer activation rates and address key reasons for exit.

**QUESTION 2**

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

Query :

****

Output:

A screenshot of a calculator

Description automatically generated

Insights:

* The top five customers in terms of estimated salary have long tenures, indicating a possible correlation between salary and customer loyalty.

There is only a slight difference in estimated salary among the top customers, suggesting that the bank's high-earning clients have similar financial profiles.

* Customers who joined the bank earlier, like Lu and Ozerova, tend to have higher ranks in terms of salary, indicating that long-term customers may have accumulated more wealth.
* All customers in this list joined the bank between 2017 and 2019, indicating a relatively stable acquisition of high-salary customers

during these years.

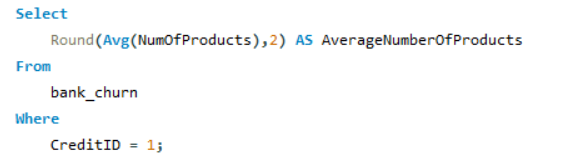
Recommendations:

* Focus retention efforts on high-salary customers, as they appear to stay with the bank for longer periods.
* Consider creating exclusive product offerings or services tailored to customers with estimated salaries above a certain threshold to increase engagement and loyalty.
* Use the insights from this list to target similar high-salary customers with personalized marketing strategies, based on their long-standing relationship with the bank.
* Monitor salary trends among newer customers to ensure the bank continues to attract high-earning individuals, adjusting strategies if necessary to maintain a steady influx of such clients.

**QUESTION 3**

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

Query:

****

Output:

**A close up of a logo

Description automatically generated**

Insights:

* The average number of products used by customers with CreditID 1 (who have a credit card) is 1.53, indicating that customers with credit cards typically utilize 1-2 products from the bank.
* This average suggests that many customers with credit cards may not be fully leveraging additional banking products and services.
* A narrow range in the number of products indicates a possible lack of product diversity among credit card holders.
* The data may reflect opportunities to engage customers more broadly across other banking services beyond credit cards.

Recommendations:

* Implement cross-selling campaigns targeted at credit card holders to introduce them to other financial products, such as loans, savings accounts, or investment services.
* Analyze the factors leading to low product utilization among credit card customers and address any barriers that might prevent customers from adopting additional products.
* Provide personalized recommendations or financial incentives to encourage customers to explore more of the bank's offerings and increase their product count.
* Segment credit card customers based on product usage behavior to create more focused marketing strategies for those who are using only one or two products.

**QUESTION 4**

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

Query:



A computer screen shot of a computer screen

Description automatically generated with medium confidence

Output:

A screenshot of a number

Description automatically generated

Insights:

* Female customers have a significantly higher churn rate (25.05%) compared to male customers (15.37%), indicating that females are more likely to leave the bank.
* The total number of female customers is slightly lower than males, but the higher churn rate suggests possible dissatisfaction among female clients.
* The query uses the most recent year of customer joining (BankDOJ), providing a fresh view of customer churn trends.
* Gender appears to be a notable factor in churn behavior, which could be related to differences in product usage, service satisfaction, or other variables not captured in this analysis.

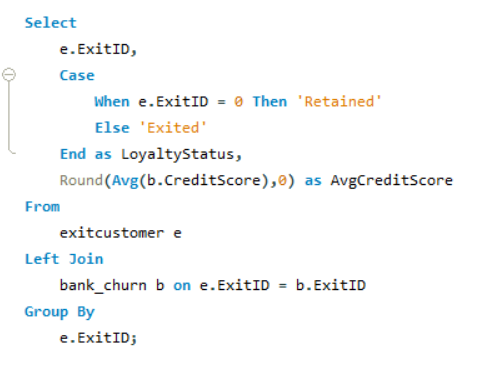
Recommendations:

* Investigate the causes of the higher churn rate among female customers by conducting surveys or focus groups to understand their pain points and preferences.
* Develop tailored retention strategies aimed at female customers, such as personalized offers, loyalty programs, or enhanced customer service.
* Monitor gender-based churn trends regularly to assess the effectiveness of retention campaigns and adjust strategies as needed.
* Use the insights from this analysis to refine customer segmentation and create more targeted marketing strategies to engage and retain female customers effectively.

**QUESTION 5**

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

Query:



Output:

A screenshot of a computer

Description automatically generated

Insights:

* The average credit score for retained customers is 652, which is slightly higher than the average credit score for exited customers, which is 645.
* While there is not a significant difference in the average credit score between retained and exited customers, it may still indicate a minor correlation between credit score and customer loyalty.
* The ExitID field differentiates between customers who have exited (ExitID = 1) and those who are retained (ExitID = 0), and this classification is helpful for analyzing customer loyalty.
* The relatively small gap in average credit scores suggests that factors other than just the credit score might influence a customer’s decision to leave the bank.

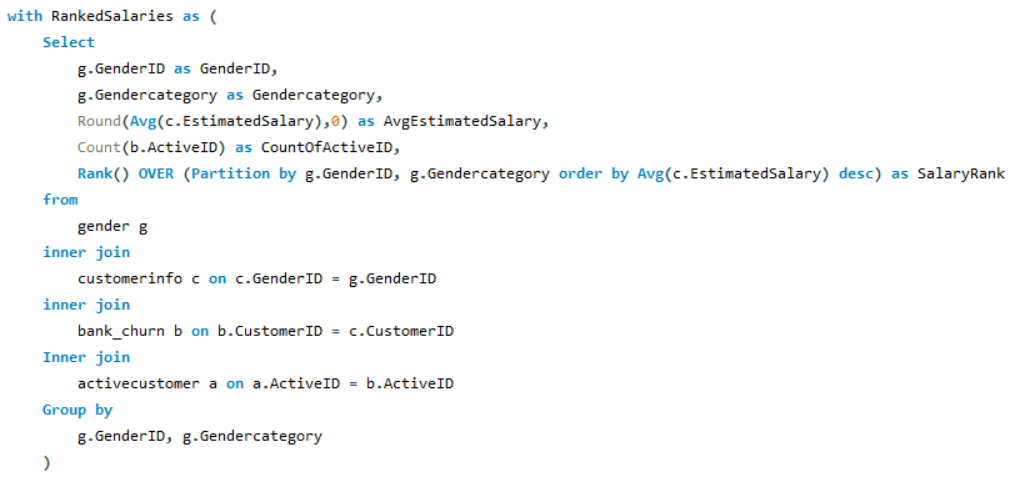
Recommendations:

* Investigate additional factors that may influence churn beyond credit scores, such as customer satisfaction, product engagement, or service quality.
* Use targeted retention strategies for customers with credit scores near the 645 range, as these customers might be more at risk of exiting.
* Monitor customer behavior and offer incentives or personalized services to customers who have lower credit scores to encourage retention.
* Consider segmenting customers based on other behavioral metrics in addition to credit scores for a more comprehensive churn analysis.

**QUESTION 6**

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

Query:



A close-up of a computer screen

Description automatically generated

Output:

A screenshot of a computer screen

Description automatically generated

Insights:

* Female customers have a slightly higher average estimated salary (100,602) compared to male customers (99,665), indicating a marginal income difference between genders.
* Despite having a higher average salary, females have fewer active accounts (4,543) compared to males (5,457), suggesting that female customers may engage less with the bank’s products.
* The use of Rank() over gender categories highlights that both genders have relatively close estimated salaries, with females leading in this comparison.
* The small difference in average salaries suggests that other factors may be influencing the level of engagement with active accounts.

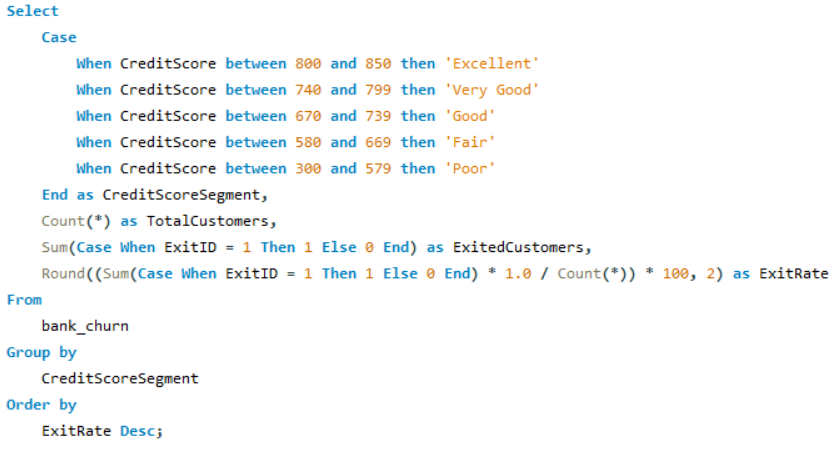
Recommendations:

* Investigate why female customers, despite having a higher average salary, have fewer active accounts and create strategies to promote product engagement among them.
* Use targeted marketing strategies to encourage female customers to explore and utilize more bank products, thereby increasing their account activity.
* Consider personalized financial services or benefits tailored to high-income customers, particularly females, to increase retention and engagement.
* Analyze other behavioral metrics beyond salary to understand gender-specific preferences and improve overall product adoption rates across all customers.

**QUESTION 7**

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

Query:



Output:

A screenshot of a number

Description automatically generated

Insights:

* Customers in the "Poor" credit score segment (300-579) have the highest exit rate of 22.02%, indicating that these customers are more likely to leave the bank compared to others.
* The "Very Good" (740-799) and "Fair" (580-669) segments also have relatively high exit rates, at 20.59% and 20.56% respectively, suggesting potential risk in these categories.
* The segment with the lowest exit rate is "Good" (670-739), with 18.62%, followed by "Excellent" (800-850), which has an exit rate of 19.54%.
* Customers with lower credit scores ("Poor" and "Fair") tend to churn more frequently, likely due to financial instability or dissatisfaction with the bank's products.

Recommendations:

* Focus retention efforts on customers in the "Poor" and "Fair" segments by offering tailored support, such as financial planning or product recommendations, to help reduce churn.
* Consider offering incentives like lower interest rates or additional benefits to customers in the "Very Good" and "Fair" segments to encourage them to stay with the bank.
* Develop targeted campaigns that focus on improving customer satisfaction and engagement among customers with lower credit scores.
* Monitor customer behavior and provide proactive outreach to those in higher-risk segments (e.g., "Poor") to address any potential reasons for exit before they leave the bank.

**QUESTION 8**

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

Query:

A computer screen shot of a code

Description automatically generated

Output:

A screenshot of a computer

Description automatically generated

Insights:

* France has the highest number of active customers with tenures greater than 5 years, totaling 797, followed by Spain with 431, and Germany with 399.
* The significant difference in active customer count between France and the other regions may indicate that customers in France tend to be more loyal or engaged with the bank's services over a longer period.
* Spain and Germany have similar numbers of long-tenured active customers, suggesting these regions may have comparable levels of customer engagement.
* The higher number of active customers in France could be attributed to better customer service, product offerings, or a stronger relationship between the bank and its French clientele.

Recommendations:

* Investigate the factors contributing to the higher retention rates in France and apply similar strategies to other regions to boost long-term customer engagement.
* Develop targeted marketing campaigns in Spain and Germany to increase customer retention, focusing on improving engagement with customers who have shorter tenures.
* Offer loyalty programs or incentives in Germany and Spain to encourage long-term customers to remain active and utilize more products.
* Monitor customer satisfaction metrics across all geographic locations to ensure consistent service quality and identify areas for improvement in customer retention.

**QUESTION 9**

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

Query:

A computer screen shot of a credit card

Description automatically generated

Output:

A screenshot of a computer

Description automatically generated

Insights:

* The churn rate for customers who do not have a credit card (20.81%) is slightly higher than the churn rate for credit card holders (20.18%).
* The difference in churn rates between the two groups is relatively small, indicating that having a credit card does not drastically affect customer churn.
* The number of total customers with credit cards (7,055) is more than double that of those without credit cards (2,945), suggesting that credit card ownership is more common among the customer base.
* Even though the churn rate for non-credit card holders is slightly higher, the similar churn rates suggest that other factors might be driving customer exits beyond credit card ownership.

Recommendations:

* Analyze customer satisfaction data to identify key factors influencing churn, such as service quality or product gaps.
* Offer personalized incentives to non-credit card holders to encourage credit card adoption and improve retention.
* Implement targeted retention strategies for both groups, addressing reasons for churn through proactive engagement.
* Regularly monitor churn trends and adjust strategies based on emerging customer behavior patterns.

**QUESTION 10**

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

Query:

A screenshot of a computer program

Description automatically generated

Output:

A screenshot of a computer

Description automatically generated

Insights:

* The majority of customers who exited the bank used only one product, accounting for 1,409 customers, which is the highest among all categories.
* A significantly smaller number of exited customers used multiple products, with only 348 customers using two products, 220 using three products, and 60 using four products.
* This indicates that customers who engage with fewer products are more likely to exit, suggesting a potential link between product engagement and retention.
* The number of customers using multiple products decreases significantly as the number of products increases, highlighting that fewer customers diversify their engagement with the bank's offerings.

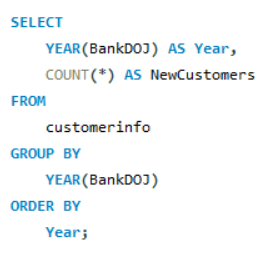
Recommendations:

* Focus on cross-selling and promoting multiple products to customers who use only one product to increase engagement and reduce churn risk.
* Develop personalized marketing strategies to encourage customers to explore more products, particularly targeting those who are at risk of exiting.
* Investigate the reasons behind the low product adoption among customers with only one product to identify barriers and address them effectively.
* Offer incentives or bundling opportunities to encourage customers to adopt additional products and strengthen their relationship with the bank.

**QUESTION 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.**

Query:

****

Output:

A screenshot of a computer

Description automatically generated

Insights:

* The data shows a consistent increase in the number of customers joining the bank from 2016 to 2019. This indicates a positive growth trend.
* The year 2019 saw the highest number of new customers (3,313), suggesting a significant boost in marketing efforts or service improvements.
* Each year, the number of new customers has increased at a faster rate, indicating accelerating growth.
* Economic or market changes during this period may have influenced the rising number of new customers.

Recommendations:

* Analyze factors contributing to the spike in customer acquisition in 2019 to leverage them further.
* Continue strategies that helped achieve consistent growth, focusing on customer acquisition programs.
* Further breakdown of data by month may reveal seasonal trends or marketing efforts that drive customer sign-ups.
* Use this data to predict future customer trends and adjust business strategies accordingly.

**QUESTION 12**

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

Query:

A screenshot of a computer code

Description automatically generated

Output:

**A screenshot of a computer

Description automatically generated**

Insights:

* Customers who used only one product have the highest total balance of 129,668,607.08, which is significantly higher than any other group.
* As the number of products increases, the total balance decreases, with customers using four products having the lowest total balance of 5,623,988.10.
* While fewer customers engage with multiple products (especially 3 and 4), they collectively contribute less to the overall balance than those with fewer products.
* The data suggests that customers with fewer products tend to maintain larger account balances, while those with more products have smaller total balances.

Recommendations:

* Investigate why customers with more products tend to have lower balances and address any factors that might be causing reduced financial engagement.
* Focus on encouraging customers with one product to adopt additional products while maintaining their high balances, through personalized offers or bundled services.
* Target customers with low balances and high product usage to identify opportunities for increasing their overall financial activity and product engagement.
* Implement strategies to retain high-balance customers who only use one product, as they represent a significant portion of the total balance despite lower product usage.

**QUESTION 13**

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

Methods to Identify Outliers:

1. Z-Score Method:

* Formula: Z=X−μσZ = \frac{X - \mu}{\sigma}Z=σX−μ​
* Outliers are identified when the Z-Score is greater than 3 or less than -3 (adjustable based on the dataset).

1. Interquartile Range (IQR) Method:

* Formula: IQR=Q3−Q1IQR = Q3 - Q1IQR=Q3−Q1
* Outliers are any values below Q1−1.5×IQRQ1 - 1.5 \times IQRQ1−1.5×IQR or above Q3+1.5×IQRQ3 + 1.5 \times IQRQ3+1.5×IQR.

1. Box Plot Visualization:

* Outliers are represented as points outside the "whiskers" (typically 1.5 times the IQR) in a box plot.

Applying Outlier Detection:

1. Z-Score Calculation:

* Identify balances with Z-Scores outside the range [-3, 3].
* IQR Calculation:
* Identify balances outside the range of Q1−1.5×IQRQ1 - 1.5 \times IQRQ1−1.5×IQR and Q3+1.5×IQRQ3 + 1.5 \times IQRQ3+1.5×IQR.

1. Handling Outliers

* Remove Outliers:
* Remove data points that are extreme or suspected errors.

1. Transform Data:

* Apply transformations (e.g., log transformation) to reduce the impact of extreme values.

1. Cap/Floor:

* Set upper and lower limits for values based on a reasonable threshold.

By using these methods, you can effectively identify and handle outliers to ensure accurate data analysis.

**QUESTION 14**

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

Ans :

The dataset consists of **7 tables**. Out of these, the following **5 tables** contain only categorical variables:

1. **activecustomer**
   * ActiveID (INT)
   * ActiveCategory (TEXT)
2. **creditcard**
   * CreditID (INT)
   * Category (TEXT)
3. **exitcustomer**
   * ExitID (INT)
   * ExitCategory (TEXT)
4. **gender**
   * GenderID (INT)
   * GenderCategory (TEXT)
5. **geography**
   * GeographyID (INT)
   * GeographyLocation (TEXT)

A diagram of a company

Description automatically generated with medium confidence

**QUESTION 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)**

Query:

A screenshot of a computer program

Description automatically generated

Output:

A screenshot of a graph

Description automatically generated

Insights:

* The query provides gender-wise average income across different geographies, helping to identify geographic regions where one gender might earn significantly more than the other.
* The ranking shows how male and female incomes compare within each geography, offering insights into income disparities between genders in different regions.
* Regions like Germany and Spain may have a higher average income for females, while other regions could show differences in male income dominance.
* Identifying geographies with large income gaps between genders can highlight potential areas for further investigation into the causes of these disparities.

Recommendations:

* Focus on regions with significant income disparity between genders to promote financial equality by introducing more inclusive financial services or policies.
* Analyze the factors contributing to higher average incomes for specific genders in certain geographies and replicate successful strategies in regions with lower averages.
* Tailor financial products based on gender-specific income data to improve customer engagement and satisfaction.
* Monitor geographic trends regularly to assess if interventions aimed at balancing gender income gaps are effective.

**QUESTION 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+).**

Query :

A screenshot of a computer program

Description automatically generated

Output:

A screenshot of a computer

Description automatically generated

Insights:

* The average tenure for customers who have exited is highest in the 30-50 age bracket, at 4.89 years, compared to 4.78 years for the 18-30 bracket and 4.83 years for the 50+ bracket.
* Customers in the 30-50 age group tend to stay longer before exiting, indicating that this group may have a stronger connection to the bank before deciding to leave.
* The 18-30 age group has the lowest average tenure, which could suggest a quicker churn rate for younger customers.
* The tenure differences across age groups are minimal, indicating a fairly uniform churn behavior across all age brackets.

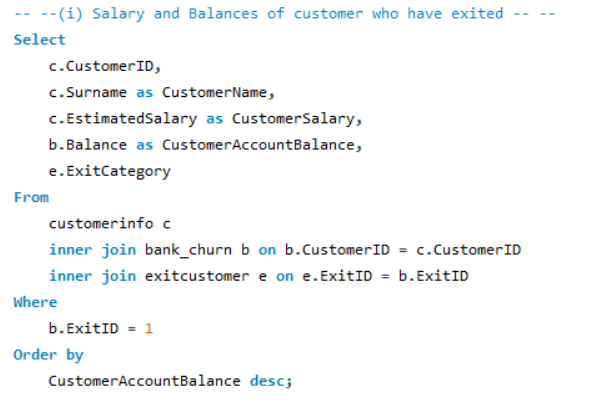
Recommendations:

* Focus retention efforts on the 18-30 age group to increase their tenure and reduce churn, possibly through products and services tailored to younger customers.
* Investigate why the 30-50 age group, despite having the longest tenure, still exits, and create targeted strategies to retain this segment for even longer.
* For the 50+ age group, offer tailored retirement or long-term financial products to further extend their tenure with the bank.
* Conduct further analysis on the reasons for churn across these age groups to design more effective, age-specific retention strategies.

**QUESTION 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?**

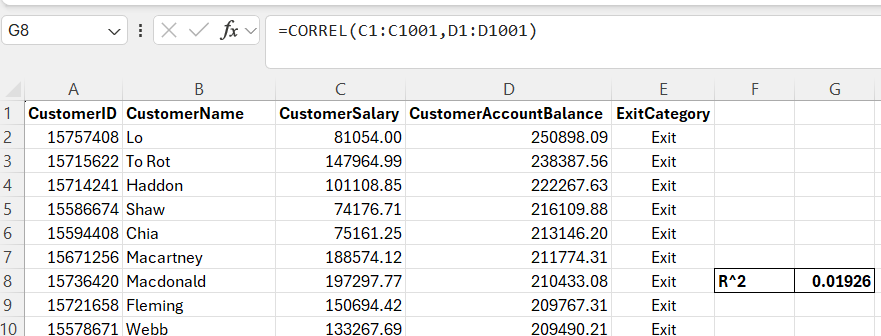
Query :

****

Output:

A screenshot of a table

Description automatically generated



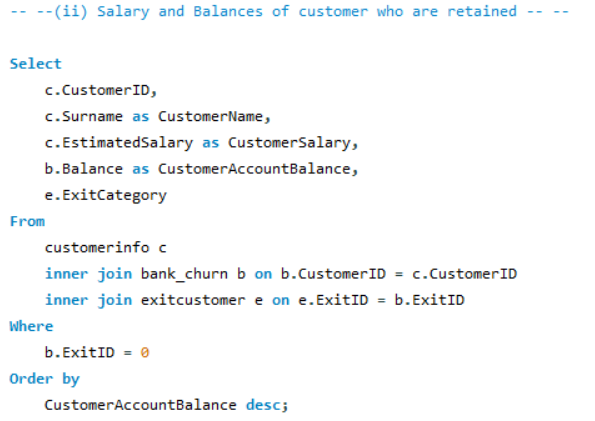
Insights:

* There is a very weak correlation (R^2 = 0.01926) between salary and account balance for exited customers.
* Higher salaries do not necessarily result in higher account balances for those who have left the bank.
* Other factors likely influence account balance more than salary for exited customers.

Recommendations:

* Analyze active customers to see if the correlation differs from exited customers.
* Focus retention efforts on product engagement rather than salary alone.
* Investigate additional factors like product usage or credit score to boost account balances.

Query:



Output:

A screenshot of a table

Description automatically generated

A screenshot of a computer

Description automatically generated

Insights:

* The correlation between salary and account balance for retained customers is extremely weak, with an R² value of -0.01857, indicating no meaningful relationship.
* Similar to exited customers, retained customers also show no clear connection between their salary and their account balance.
* This suggests that both groups, whether retained or exited, are influenced by other factors besides salary when it comes to their account balance.

Recommendations:

* Focus on identifying other factors, such as product usage or customer engagement, that influence account balances, as salary does not seem to be a key driver.
* Offer tailored financial services to customers based on their spending habits or account activity, rather than salary alone.
* Conduct further analysis on the impact of additional factors like customer tenure or the number of products they use to better understand balance dynamics.

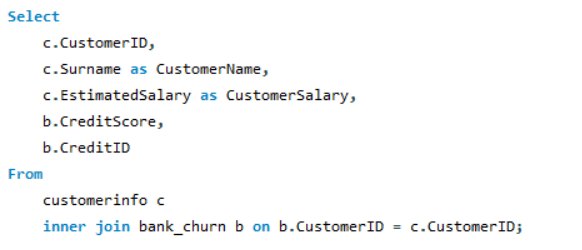
Difference:

* Both exited and retained customers exhibit weak correlations between salary and account balance, with slight variations in R² values but no significant trends.
* There is no major distinction in how salary affects account balance between exited and retained customers, indicating similar financial behavior across both groups.

**QUESTION 18**

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

Query:

****

Output:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Insights:

* There is no strong correlation between salary and credit score, as the R^2 values for both credit card holders and non-holders are very low (around 0.01).
* The lack of correlation indicates that salary is not a key factor in determining a customer's credit score.
* Both credit card holders and non-holders display similar weak correlations between their salary and credit score.

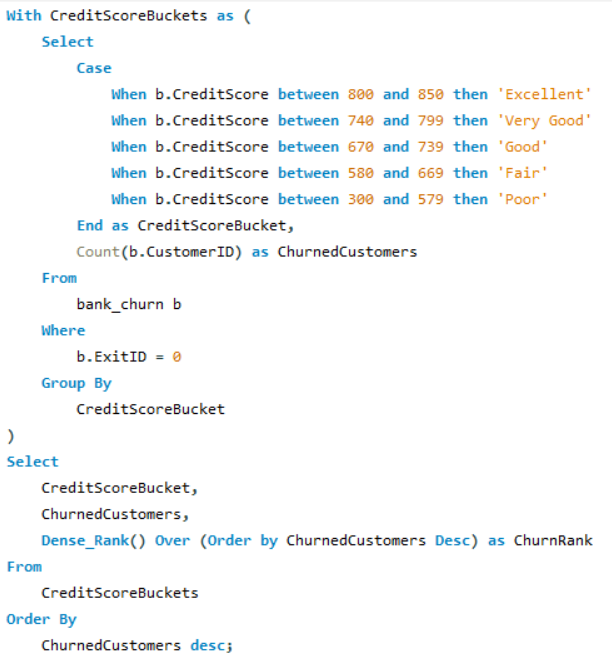
Recommendations:

* Focus on other factors, such as financial behavior and credit history, when analyzing or predicting credit scores.
* Encourage better financial education to help customers improve their credit scores irrespective of their income levels.
* Consider designing tailored financial products that address both high and low credit score segments for better customer engagement.

**QUESTION 19**

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

Query:



Output:

A screenshot of a computer

Description automatically generated

Insights:

* Customers with a "Fair" credit score have the highest churn rate, accounting for 2,646 customers, which suggests that this group is more vulnerable to churn.
* "Good" and "Poor" credit score buckets rank second and third in churn, indicating that even relatively good credit scores do not prevent churn.
* Customers with "Excellent" credit scores have the lowest churn, with only 527 customers, suggesting stronger loyalty in this group.
* There is a general trend of decreasing churn as credit scores improve, implying a correlation between higher credit scores and customer retention.

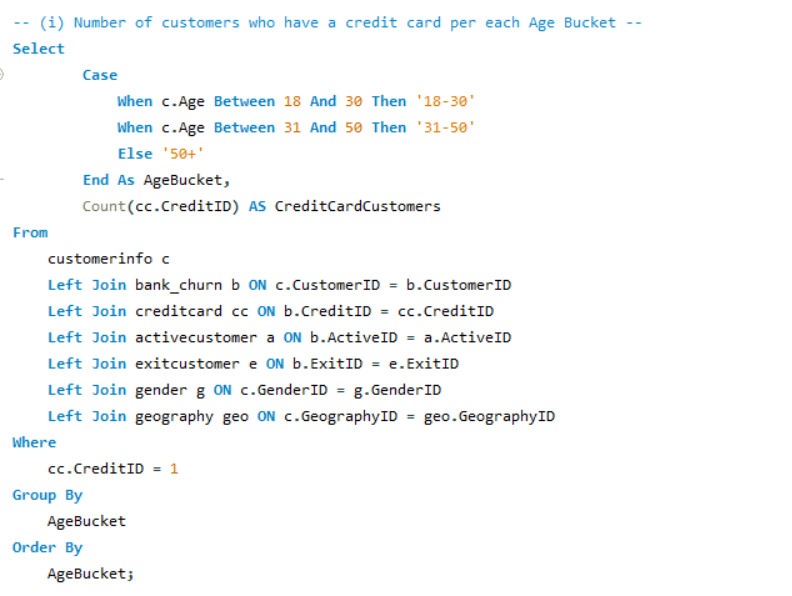
Recommendations:

* Implement targeted retention strategies for customers in the "Fair" and "Good" credit score buckets to reduce churn.
* Provide personalized offers or improved services for customers with lower credit scores to enhance their satisfaction and reduce churn.
* Focus on maintaining relationships with customers who have "Excellent" credit scores by rewarding loyalty to reinforce retention.
* Regularly monitor credit score trends to identify at-risk segments and pre-emptively address their needs before they churn.

**QUESTION 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.**

Query and Output 1:



A screenshot of a credit card

Description automatically generated

Query and Output 2:

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a credit card

Description automatically generated

Insights:

* The highest number of credit card holders are in the 31-50 age group, significantly outnumbering the other age buckets.
* The 18-30 and 50+ age groups have fewer credit card holders than the average per age bucket.
* Age distribution suggests that the middle age group (31-50) is more engaged with credit card usage.
* Credit card ownership in younger and older demographics is below average and may need targeted marketing.

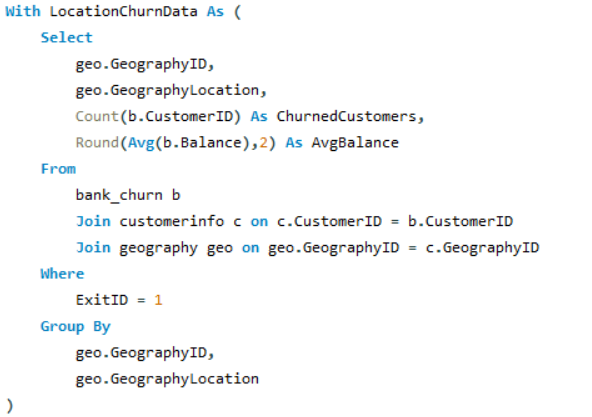
Recommendations:

* Implement targeted marketing campaigns for younger (18-30) and older (50+) age groups to increase credit card adoption.
* Consider offering age-specific rewards or promotions to drive higher credit card engagement.
* Monitor age-based credit card usage trends to identify growth opportunities and further refine product offerings.
* Explore financial literacy programs for younger demographics to increase awareness and usage of credit products.

**QUESTION 21**

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

Query:



A screenshot of a computer code

Description automatically generated

Output:

A screenshot of a computer

Description automatically generated

Insights:

* Germany has the highest average balance among churned customers, though it ranks lowest in churn count.
* Spain has the least number of churned customers but ranks second in average balance, indicating relatively more stable customers.
* France has the highest churn count but has the lowest average balance among churned customers.
* Churn behavior and average balance seem inversely related, especially between Germany and France.

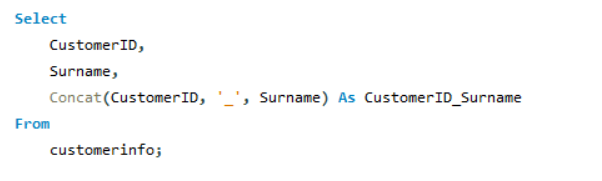
Recommendations:

* Focus on retaining high-balance customers in Germany by enhancing customer satisfaction and product offerings.
* Investigate reasons for higher churn in France despite lower average balances, potentially identifying underserved customer segments.
* Strengthen customer engagement in Spain to reduce churn further, as it already has lower churn rates.
* Implement a balanced retention strategy considering both churn volume and average customer balance to prioritize efforts.

**QUESTION 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”.**

Query:



Output:

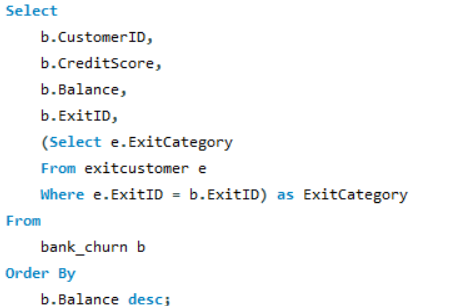
A screenshot of a computer

Description automatically generated

**QUESTION 23**

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

Query:



Output:

A screenshot of a data

Description automatically generated

**QUESTION 24**

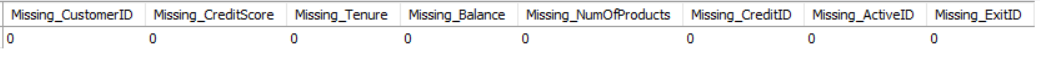
**Were there any missing values in the data, using which tool did you replace them and what are the ways to handle them?**

Query:

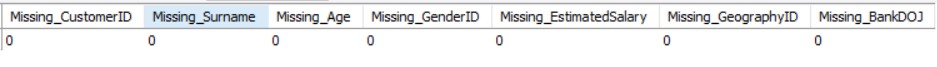
****

Output:

1. For bank\_churn Table:



1. For customerinfo Table:

****

Insights:

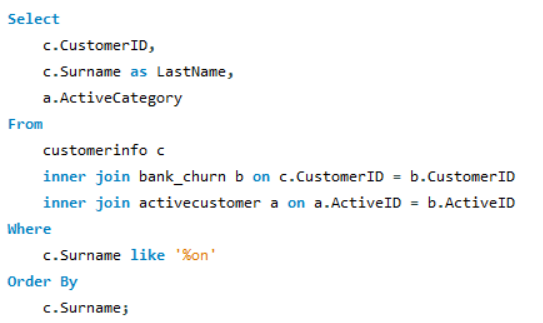
Currently, there are no missing values in the dataset, as seen from the query results. However, if there were missing values, here are the steps we could take:

* Imputation: We can use techniques such as mean, median, or mode imputation to fill missing values for numerical data. For categorical variables, the most frequent category could be used.
* Deletion: In cases where a few rows have missing data, we can choose to delete those records if the impact on overall analysis is minimal, ensuring we don't lose significant information.
* Interpolation or Forward Filling: For time-series data, we can interpolate or use forward/backward filling techniques to estimate missing values based on neighbouring records.

**QUESTION 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”.**

Query :



Output:

**A screenshot of a computer

Description automatically generated**

**QUESTION 26**

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

Query:

A number of numbers and symbols

Description automatically generated with medium confidence

Output:

A screenshot of a computer

Description automatically generated

Insights:

* The dataset shows customers with both ActiveID = 1 (indicating they are active) and ExitID = 1 (indicating they have exited). This is contradictory, as a customer cannot logically be both active and exited.
* The discrepancy likely stems from incorrect data entry or an issue during data migration, where the active status was not updated correctly when the customer exited the bank.
* Several customers have a balance of 0, yet they are marked as active or have exited with a product still assigned. This might indicate poor data management or delayed balance updates after account closure.
* Some customers have multiple products (NumOfProducts > 1) but still show a zero balance. This could suggest inactive products that were not properly removed when the customer exited.

Recommendations:

* Implement automated scripts to check for logical inconsistencies, such as a customer being both active and exited simultaneously.
* Ensure proper workflows and checks when updating customer records, especially when they exit, to ensure the ActiveID is set to 0 upon exiting.
* Cross-check data entries related to account balances and product ownership for active and exited customers to avoid such discrepancies in the future.
* Investigate the process used to manage account statuses during customer exits. Potential causes for these discrepancies may include lagging updates, incorrect triggers, or overlapping system changes.

These steps should help to maintain consistency and ensure data accuracy in your system.

**Subjective Question:**

**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?**

Query:

A screenshot of a computer code

Description automatically generated

Output:

A screenshot of a computer

Description automatically generated

Insights from Customer Behaviour Analysis:

* Both new and long-term customers have similar average balances, with new customers slightly ahead in average balance, suggesting recent customers are depositing or transacting higher amounts early on.
* Long-term customers show a marginally lower average number of products (1.52) compared to new customers (1.56), indicating that long-term customers may stick to a few core products.
* New customers exhibit a slightly higher average credit score than long-term customers, suggesting that new customers might be more credit-conscious, or banks might be targeting customers with higher creditworthiness initially.
* Despite lower product usage and credit scores, long-term customers make up the bulk of the customer base, pointing to a strong retention capability but room for cross-selling opportunities.

Recommendation:

* Since new customers use slightly more products, developing retention strategies to cross-sell or upsell to long-term customers could increase their engagement.
* Implement loyalty programs that reward long-term customers for maintaining balances and encourage higher product adoption.
* Given new customers tend to have higher credit scores, tailor marketing to attract more high credit-score individuals and incentivize them to stay longer.
* Introduce balance-based incentives or tiered rewards for customers maintaining higher balances, appealing to both new and long-term customers.

**QUESTION 2**

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

Query:



Output:

A screenshot of a computer

Description automatically generated

Insights from Product Affinity Study:

* Over 50% of the customers have only a savings account, indicating that this is the most commonly used product by customers.
* A significant portion (45.90%) of customers use a combination of savings accounts and credit cards, which suggests an opportunity to promote further services to this group.
* Only 0.60% of customers use a comprehensive set of products like savings accounts, credit cards, loans, and investment accounts, indicating a potential for further engagement.
* Only 2.66% of customers have a product combination that includes a loan, indicating that loan products may need more targeted promotion.

Recommendations:

* Cross-sell credit cards and loan products to customers who have only a savings account, as they represent a large portion of the customer base.
* Create bundled product offerings (e.g., savings + credit card + loan) to encourage multi-product usage and improve customer retention.
* Identify customers who might benefit from investment products and provide personalized offers to grow the percentage of customers using these services.
* Use the insights from cross-product usage to tailor engagement strategies for different customer segments, offering personalized product combinations.

**QUESTION 3**

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

Query:

****

Output:

A screenshot of a computer

Description automatically generated

Insights:

* Both France and Germany show similar churn rates, with France at 0.1235 and Germany at 0.1229, indicating that customers in these regions are exiting the bank's services at almost the same rate.
* The churn rate in Spain is 0.2421, which is almost double that of France and Germany, suggesting that customers in Spain are more likely to leave the bank.
* Germany has the highest number of churned customers (814), followed closely by France (810), while Spain has fewer churned customers (413). This suggests Germany has a larger customer base but is still able to maintain a relatively lower churn rate.
* The substantial difference in churn rates between Spain and the other regions suggests geographic disparities in customer behavior or satisfaction with the bank’s services.

Recommendations:

* Since Spain has a higher churn rate, the bank should focus on understanding the causes behind the higher churn and implement targeted strategies such as improving customer service or offering tailored financial products to retain customers.
* While the churn rates in Germany and France are lower, proactive steps like personalized banking services, loyalty programs, or incentives could help further reduce churn and retain the existing customer base.
* Conduct in-depth market research in Spain to identify why customers are exiting at a higher rate compared to other regions. Possible reasons could include economic factors, product dissatisfaction, or increased competition from other financial institutions.
* Since France and Germany have stable churn rates, the bank could consider allocating resources to acquire more active customers in these regions, leveraging the low churn environment to increase the overall customer base.

**QUESTION 4**

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

Query

Output:

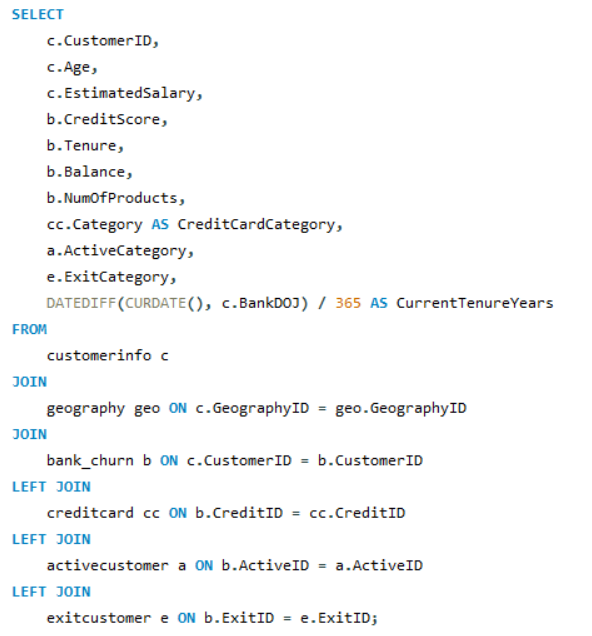
Insights:

Recommendations:

**QUESTION 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?**

Query:

****

Output:

**A screenshot of a credit card

Description automatically generated**

Insights:

* Customers with higher balances generally show longer tenure values, indicating stronger loyalty to the bank.
* Inactive members tend to have shorter tenure compared to active members, which may reflect disengagement.
* Customers with multiple products, especially credit cards, exhibit longer tenures, showing that cross-selling strengthens customer retention.
* Retained customers typically have longer tenures, while those in the "Exit" category have relatively shorter tenures.

Recommendations:

* Focus on retaining active members by offering personalized financial services to further enhance their tenure.
* Engage inactive members with reactivation offers to prolong their relationship with the bank.
* Promote cross-selling strategies, especially targeting non-credit card holders, to increase tenure and customer value.
* Use predictive analytics to target customers likely to exit, offering incentives to retain them before they churn.

**QUESTION 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?**

To assess the impact of marketing campaigns on customer retention and acquisition within the dataset:

* Add data about which customers were part of a marketing campaign, including details such as campaign type, start date, and incentives offered. This could be in a new table campaigns that links to customerinfo.
* Compare retention and acquisition metrics before and after campaigns. You would analyze changes in Tenure, Balance, and NumOfProducts to see if customers engaged more after the campaign.
* Monitor the increase in new customers (CustomerID) after campaigns, comparing regions (GeographyLocation) or demographics to see which segments responded best.
* Measure how many customers stayed active (no ExitID) post-campaign versus those who exited. Campaign effectiveness could be tied to a higher percentage of active members after the campaign.

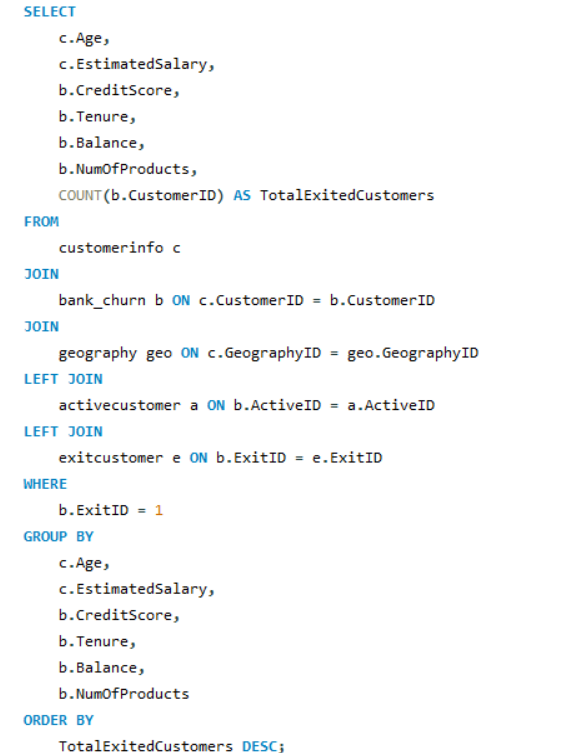
Additional Information Needed:

* Campaign participation details (campaign name, start/end date, customer segment).
* Data on customer engagement with specific marketing efforts (discounts, offers).
* Specific timestamps to track customer behaviors before and after campaigns.

**QUESTION 7**

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

Query:

****

Output:

**A screenshot of a data

Description automatically generated**

Insights:

* Many exited customers have low engagement, holding only 1 or 2 products, which may indicate a lack of deeper relationships with the bank.
* A significant number of exited customers have relatively low credit scores (below 500), potentially signaling financial instability.
* Several customers exited despite having substantial balances, suggesting that financial dissatisfaction might not be the main reason for leaving.
* Customers with longer tenure (4-7 years) also appear among those who exited, which indicates churn may happen even after a stable relationship period.

Recommendations:

* Implement targeted retention strategies for customers with only 1-2 products by offering cross-selling opportunities and incentives to deepen their engagement.
* Focus on providing personalized financial solutions for customers with low credit scores to reduce financial stress and prevent exits.
* Analyze customer satisfaction levels for those with high balances to address any service-related issues and prevent churn.
* Build loyalty programs for long-tenure customers to further strengthen their relationship with the bank and reduce their likelihood of exiting.

**QUESTION 8**

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

Yes, the variables 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' are important for predicting if a customer will leave the bank. Here's why each factor plays a key role in churn prediction:

Tenure:

* Customers with shorter tenure are more likely to leave because they haven't built a strong relationship with the bank yet. On the other hand, customers with longer tenure may be more loyal but may still churn due to dissatisfaction after a longer engagement.

NumOfProducts:

* Customers holding fewer products are less engaged with the bank and have less to lose if they leave, making them more prone to churn. On the other hand, customers with multiple products are likely more invested and harder to disengage.

IsActiveMember:

* Customers who are inactive or not actively using the bank's services are more likely to leave since they're less engaged. This makes the "IsActiveMember" status a strong predictor of potential churn.

EstimatedSalary:

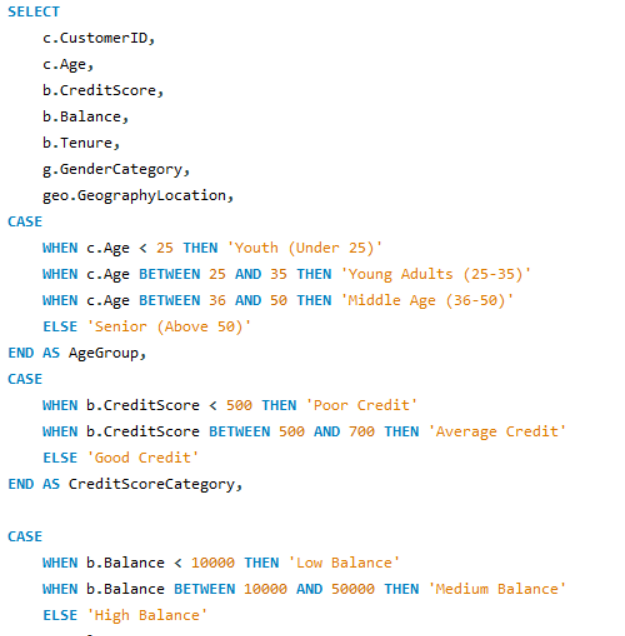
* Customers with higher salaries are typically offered premium services, which could incentivize them to stay. However, customers with lower salaries may face financial difficulties or find better offers elsewhere, leading to higher churn rates.

These factors, along with others such as CreditScore and Balance, form a comprehensive profile that can be used to predict customer churn behavior using machine learning models like logistic regression, decision trees, or random forests.

**QUESTION 9**

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

Query:

****

**A screenshot of a computer program

Description automatically generated**

Output:

A screenshot of a computer

Description automatically generated

Insights:

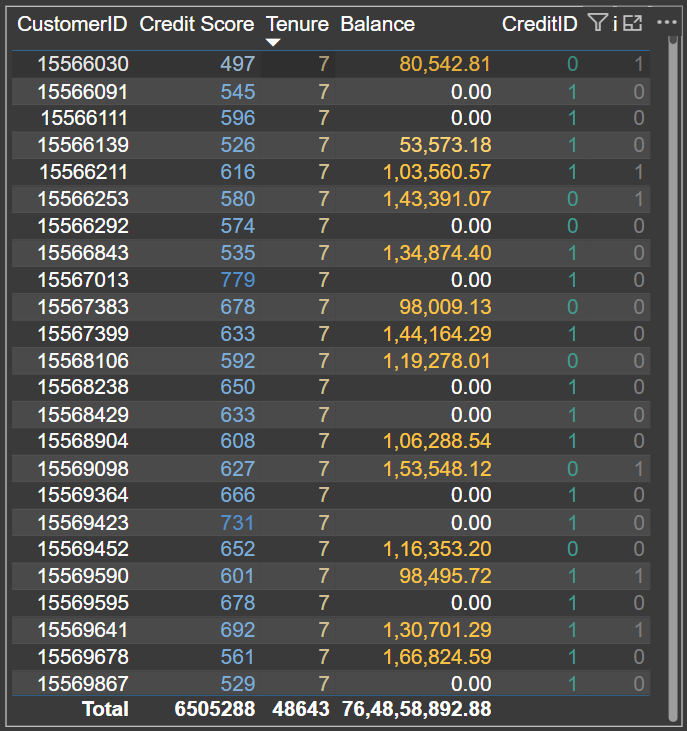
* Many "Young Adults (25-35)" and "Middle Age (36-50)" customers have high credit scores (above 700) and high balances, indicating strong financial stability.
* Customers with "Good Credit" and "Loyal Customer" tenure segments generally hold credit cards and exhibit higher balances.
* Customers with "Poor Credit" tend to have low balances, particularly in the "Middle Age (36-50)" group, which could indicate financial challenges.

Recommendations:

* Focus on retaining "Loyal Customers" with good credit by offering premium services and rewards to maintain engagement.
* Develop financial products targeted at customers with "Poor Credit" to help improve their financial standing and reduce churn risk.
* Encourage cross-selling to "Moderate Customers" by offering additional products, particularly to those with "Average Credit" to increase product engagement and tenure.

**QUESTION 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?**

Output:

Insights:

* Customers with a credit score below 500 and a balance below 30,000 are at high risk of churn, which suggests that financial instability is a major driver of customer exit.
* Longer tenure customers (above 3 years) tend to be more loyal, but those with lower credit scores or minimal product engagement are exceptions and require targeted attention.
* The presence of a credit card may not directly reduce churn, but its impact could be tied to the type of rewards and category offered. Customers with higher balances and premium cards exhibit stronger retention.
* Customers with balances between 30,000 and 80,000 who use credit card rewards show improved retention, indicating that medium-balance customers are positively influenced by reward programs.

Recommendations:

* Create targeted retention programs for customers with credit scores below 500, offering personalized financial products and credit score improvement plans to increase retention and engagement.
* Optimize credit card rewards programs by tailoring them to higher-risk segments, especially those with medium balances (30,001 to 80,000), ensuring the rewards are aligned with customer preferences.
* Improve engagement strategies for new customers (tenure less than 3 years) through onboarding offers, better access to premium products, or incentives for increasing their balance.
* Monitor and adjust product offerings based on regional customer preferences and financial health by analyzing geographic and balance-related churn patterns to fine-tune loyalty campaigns.

Conditional Formatting:

Credit Score:

* Scores below 500 are highlighted (possibly in red) as a high-risk category.
* Scores between 500 and 700 are moderately highlighted, indicating average risk.
* Scores above 700 are marked (possibly in green) as low-risk, suggesting financially stable customers.

Tenure:

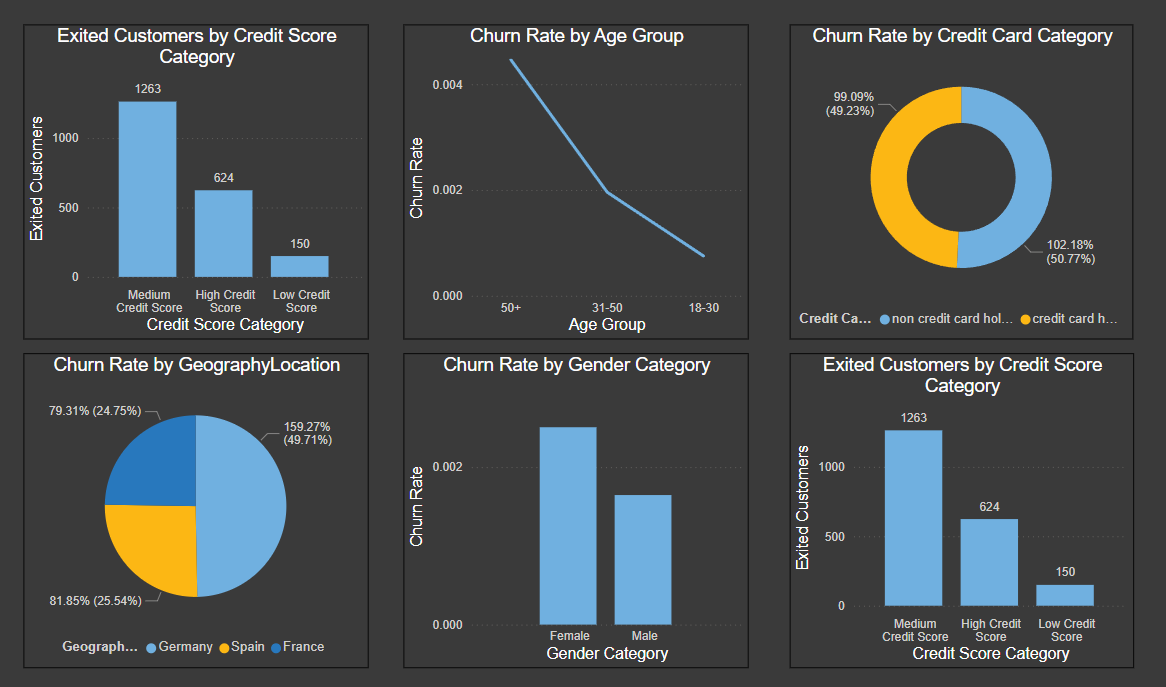
* Customers with a tenure of 3 years or more are highlighted as more loyal,
* while those with less than 3 years are flagged as potential churn risks.

Balance:

* Balances below 30,000 are flagged as financially vulnerable.
* Balances between 30,001 and 80,000 are moderate-risk customers.
* Balances above 80,000 are considered secure customers with strong engagement.

**QUESTION 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?**

****

Insights:

* Customers with medium credit scores (500-700) are more likely to churn compared to those with higher scores. This shows that mid-tier customers face financial challenges that lead to exit, but they may still be within the bank's ability to retain.
* Older customers (age group 50+) have a higher churn rate compared to younger ones, possibly due to less engagement with digital services or a shift in financial needs.
* Geographic location plays a role in churn, with customers from Spain showing a higher churn rate (81.85%), which may indicate that competitive offerings or dissatisfaction with the bank's services in specific regions are driving exits.
* Credit card holders churn at a higher rate (102.18%) than non-credit card holders (99.09%), suggesting that the current credit card rewards and offerings may not be appealing or sufficient to retain customers.

Recommendations:

* Create targeted retention programs for customers in the mid-tier credit score range (500-700) by offering tailored financial planning services or improving their access to credit. This could help retain customers who are financially vulnerable but valuable in the long term.
* Introduce senior-focused banking solutions to engage older customers (50+) with personalized products and improved accessibility to digital banking services, potentially addressing their unique needs and lowering churn.
* Strengthen customer engagement in high-churn regions like Spain by conducting localized market research and offering region-specific products or services that cater to the preferences of these customers.
* Revamp credit card reward programs to better suit the needs of current customers, particularly focusing on those with high balances or medium credit scores to incentivize usage and retention through more appealing rewards.

By following these recommendations, the bank can focus on reducing churn rates by addressing specific customer segments that are more likely to exit, thus improving retention.

**QUESTION 12**

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

**A screenshot of a graph

Description automatically generated**

**A screenshot of a graph

Description automatically generated**

**QUESTION 13**

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

If the objective and subjective questions weren't given, here’s how I would approach the problem of analyzing the "Bank\_Churn" dataset with the goal of uncovering insights and solving relevant business challenges:

1. Understand the Schema

* Review the provided data schema and ER diagram to understand the relationships between different tables, columns, and data points (e.g., customerinfo, bank\_churn, exitcustomer, activecustomer, etc.).
* Identify key tables such as customerinfo for demographic data, bank\_churn for customer activity, and exitcustomer for churn data.

2. Data Cleaning and Preprocessing:

* Missing Values: Identify any missing or inconsistent values in the dataset, especially in critical fields like CreditScore, Balance, NumOfProducts, and EstimatedSalary.
* Data Discrepancies: Ensure that there are no conflicting data points in columns like IsActiveMember and Exited (for instance, an active member marked as exited).
* Data Transformation: Prepare the data for analysis by formatting columns (e.g., dates, currency), renaming ambiguous column names (HasCrCard to Has\_creditcard), and generating new fields (e.g., combining CustomerID and Surname into a unique key).

3. Key Business Metrics Identification:

* Churn Rate: Compute the overall churn rate and segment it by demographics such as age, gender, geography, and tenure.
* Product Engagement: Analyze how many products each customer uses and the average number of products for different customer segments (e.g., those who have exited vs. those who remain).
* Credit Score Analysis: Compare average credit scores between customers who have exited and those who remain with the bank.
* Customer Lifetime Value (Tenure Forecast): Use customer tenure as a key metric to predict lifetime value based on demographic and behavioral patterns.

4. Customer Segmentation:

* Segment customers based on demographics (age, gender, geography) and account details (balance, credit score, number of products). This segmentation can help in creating marketing or retention strategies targeted to specific groups.
* Segment churned customers by their characteristics (e.g., tenure, product usage) to understand what factors are most likely to contribute to churn.

5. SQL Queries to Generate Insights:

* Churn Analysis: Write SQL queries to calculate the churn rate by year, geography, and gender. Analyze the characteristics of customers who churn (low engagement, low balance, etc.).
* Product Affinity: Investigate common product combinations (e.g., customers with both credit cards and loans) to identify cross-selling opportunities.
* Credit Score Impact: Query data to analyze the relationship between customer credit scores and churn or product engagement.
* Customer Retention: Analyze the retention rates of customers who use multiple products versus those who use fewer products.

6. Data Visualization and Reporting:

* Dashboards: Develop visualizations (e.g., in Power BI or Tableau) for key metrics like churn rate, customer segmentation, product usage, and credit score distribution.
* Conditional Formatting: Highlight high-risk customers in visual reports using conditional formatting to identify those most likely to churn.
* Time-Series Analysis: Visualize trends in customer sign-ups, exits, and product usage over time (e.g., seasonal or yearly trends).

7. Business Recommendations:

* Retention Strategies: Based on churn analysis, recommend personalized offers, loyalty programs, and engagement campaigns to retain customers who are likely to churn.
* Product Cross-Selling: Leverage product affinity insights to recommend cross-selling strategies that can increase customer engagement and retention.
* Credit Score Management: Suggest financial counseling or credit score improvement programs for customers with low credit scores to reduce churn and increase engagement.

This approach allows for a comprehensive understanding of the bank's customer behavior, churn patterns, and product engagement, leading to actionable insights and business strategies.

**QUESTION 14**

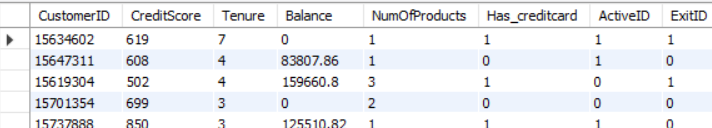
**In the “Bank\_Churn” table how can you modify the name of the “HasCrCard” column to “Has\_creditcard”?**

Query:

**A close up of text

Description automatically generated**

Output:

****