**Bank Customer Analytics Report**

Objective & Subjective Insights with SQL, Visualizations, and Strategic Reasoning



Project: Capstone Project

Prepared By: Kumar Prakash

**Table of Contents**

1. **Executive Summary**
2. **Objective Questions – Exact Answers**
3. **Objective Questions – Reasoned Answers**
4. **Subjective Questions – Strategic Analysis**
5. **Alternate Approach**
6. **Conclusion & Recommendations**
7. **References**

**Executive Summary**

* **This report analyzes customer data from a bank to uncover insights related to churn, customer behavior, and product usage.**
* **Using SQL queries, visualizations, and strategic reasoning, we answer both objective and subjective questions to support business decisions**

**Objective Questions with Answers**

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

SQL query:

SELECT

g.GeographyLocation AS Region,

ROUND(AVG(b.Balance), 2) AS `AvgBalance`,

ROUND(MIN(b.Balance), 2) AS MinBalance,

ROUND(MAX(b.Balance), 2) AS MaxBalance

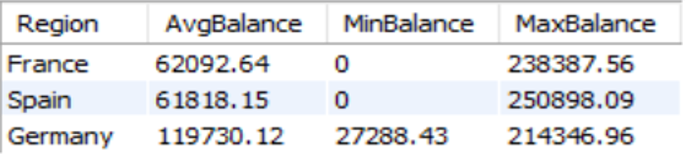
FROM Geography g

JOIN CustomerInfo c ON g.GeographyID = c.GeographyID

JOIN BankChurn b ON c.CustomerId = b.CustomerId

GROUP BY g.GeographyLocation;

Result table:



Insight:

* Span has lower Average Balance (i.e. 61818.15) and Germany has highest (i.e. 119730.12) among all three regions.
* Similarly, Spain has highest maximum balance (i.e. 250898.09) and Germany has highest minimum balance (i.e. 27288.43).

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

SQL query:

SELECT CustomerID, EstimatedSalary

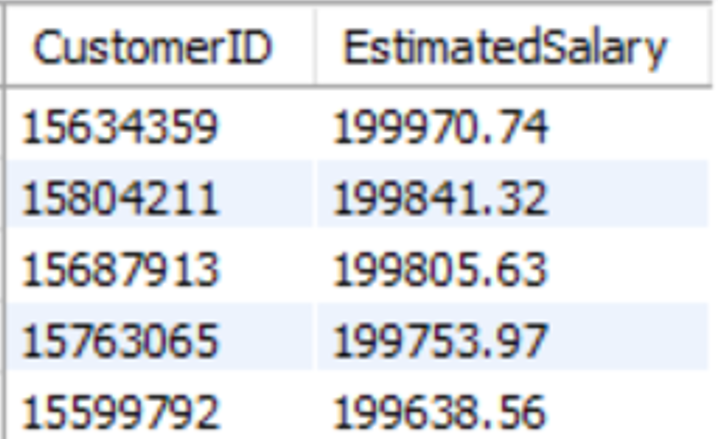
FROM CustomerInfo

WHERE MONTH(`Bank DOJ`) IN (10, 11, 12)

ORDER BY EstimatedSalary DESC

LIMIT 5;

Result table:



Insight:

* Among all customer, customer ID (i.e. 15634359) has highest Estimated salary (i.e. 199970.74).

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

SQL query:

SELECT ROUND(AVG(NumOfProducts)) AS AvgNumberOfProducts

FROM BankChurn

WHERE HasCrCard = 1;

Result table:



Insight:

* Average number of products by customer who have a credit card is: 2.

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

SQL query:

SELECT g.GenderCategory,

Round(((COUNT(CASE WHEN bc.Exited = 1 THEN 1 END) \* 100.0)/ COUNT(\*)), 2) AS `ChurnRate (in %)`

FROM BankChurn bc

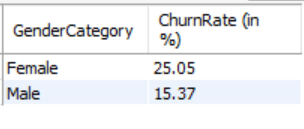
JOIN CustomerInfo ci ON bc.CustomerId = ci.CustomerId

JOIN Gender g ON ci.GenderID = g.GenderID

WHERE YEAR(`Bank DOJ`) = (SELECT MAX(YEAR(`Bank DOJ`)) FROM CustomerInfo)

GROUP BY g.GenderCategory;

Result table:



Insight:

* Churn Rate of Female is higher than Male in the most recent year.

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

SQL query:

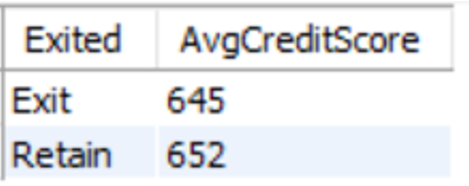
SELECT ec.ExitCategory as Exited, ROUND(AVG(bc.CreditScore)) AS AvgCreditScore

FROM ExitCustomer ec

JOIN BankChurn bc ON bc.Exited = ec.ExitID

GROUP BY ec.ExitCategory;

Result table:



Insight:

* Avgrage Credit Score of exited customers are low as compare to Retain customers.

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

SQL query:

SELECT g.GenderCategory,

ROUND(AVG(ci.EstimatedSalary), 2) AS AvgSalary,

COUNT(CASE WHEN bc.IsActiveMember = 1 THEN 1 END) AS ActiveAccounts

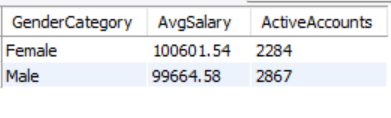
FROM CustomerInfo ci

JOIN BankChurn bc ON ci.CustomerId = bc.CustomerId

JOIN Gender g ON ci.GenderID = g.GenderID

GROUP BY g.GenderCategory;

Result table:



Insight:

* Male has highest number of active accounts as compare to Female but lower the Average Salary.

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

SQL query:

SELECT

CASE

WHEN CreditScore >= 800 THEN 'Excellent'

WHEN CreditScore >= 740 THEN 'Very Good'

WHEN CreditScore >= 670 THEN 'Good'

WHEN CreditScore >= 580 THEN 'Fair'

ELSE 'Poor'

END AS CreditSegment,

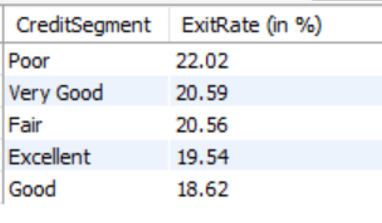
ROUND((COUNT(CASE WHEN Exited = 1 THEN 1 END) \* 100.0 / COUNT(\*)), 2) AS `ExitRate (in %)`

FROM BankChurn

GROUP BY CreditSegment

ORDER BY `ExitRate (in %)` desc;

Result table:



Insight:

* Poor (i.e. credit score < 580) has highest Exit Rate as 22.02% among all.

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

SQL query:

SELECT g.GeographyLocation, COUNT(\*) AS NumberOfActiveCustomers

FROM BankChurn bc

JOIN CustomerInfo ci ON bc.CustomerId = ci.CustomerId

JOIN Geography g ON ci.GeographyID = g.GeographyID

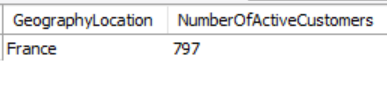
WHERE bc.Tenure > 5 AND bc.IsActiveMember = 1

GROUP BY g.GeographyLocation

ORDER BY NumberOfActiveCustomers DESC

LIMIT 1;

Result table:



Insight:

* France has highest Number Of Active Customers as 797 among others.

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

SQL query:

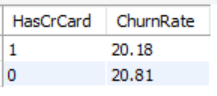
SELECT HasCrCard,

Round((COUNT(CASE WHEN Exited = 1 THEN 1 END) \* 100.0 / COUNT(\*)), 2) AS ChurnRate

FROM BankChurn

GROUP BY HasCrCard;

Result table:



Insight:

* There is no major impact of Has Credit Card on Churn Rate.

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

SQL query:

SELECT NumOfProducts, COUNT(\*) AS Frequency

FROM BankChurn

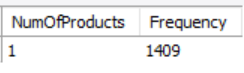
WHERE Exited = 1

GROUP BY NumOfProducts

ORDER BY Frequency DESC

LIMIT 1;

Result table:



Insight:

* Most common number of products among all are 1 with frequency: 1409.

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.

SQL query:

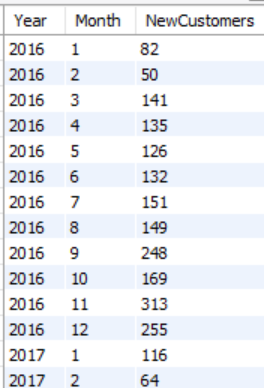
SELECT YEAR(`Bank DOJ`) AS Year, MONTH(`Bank DOJ`) AS Month, COUNT(\*) AS NewCustomers

FROM CustomerInfo

GROUP BY YEAR(`Bank DOJ`), MONTH(`Bank DOJ`)

ORDER BY Year, Month;

Result table:



Etc…

* I took the screenshot of few rows and showing in Result table section.

Insight:

**Yearly Growth Trend**

* There's a clear upward trajectory in customer acquisition from 2016 to 2019.
* Total new customers per year:
  + - **2016**: ~1,951
    - **2017**: ~2,143
    - **2018**: ~2,593
    - **2019**: ~3,313
* This suggests consistent growth, with 2019 showing a significant spike.

**Monthly Seasonality**

* Across all years, **September to December** consistently show peak customer acquisition.
* **November and December** are especially strong, possibly due to year-end financial planning, festive offers, or fiscal deadlines.
* **February** tends to be the lowest month, likely due to its shorter duration and post-holiday lull.

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

SQL query:

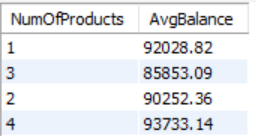
SELECT NumOfProducts, Round(AVG(Balance), 2) AS AvgBalance

FROM BankChurn

WHERE Exited = 1

GROUP BY NumOfProducts;

Result table:



Insight:

* Average Balance of Num Of Products as 4 has higher and Num Of Products as 3 has lower.

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

* To Identify outliers we have to calculate ZScore. For customer whose ZScore > 3 treated as a outlier.
* So, I calculated outlier by ABS(ZScore) > 3.
* Where ZScore = (Balance - AVG(Balance) / STDDEV(Balance)

SQL query:

SELECT CustomerId, Balance

FROM (

SELECT CustomerId, Balance,

(Balance - AVG(Balance) OVER()) / STDDEV(Balance) OVER() AS ZScore

FROM BankChurn

WHERE Exited = 0

) AS sub

WHERE ABS(ZScore) > 3;

Result table:

* We got empty results, it means there is no outliers among all.

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

SQL query:

SHOW COLUMNS FROM ActiveCustomer;

SHOW COLUMNS FROM BankChurn;

SHOW COLUMNS FROM CreditCard;

SHOW COLUMNS FROM Gender;

SHOW COLUMNS FROM ExitCustomer;

SHOW COLUMNS FROM Geography;

SHOW COLUMNS FROM CustomerInfo;

Insight:

* Total Tables = 7,
* There are no such table that contains only categorical variables.
* But, we have some column in the following table which contain only categorical value.
* For example:  
  Gender(GenderCategory), Geography(GeographyLocation), ExitCustomer(ExitCategory), ActiveCustomer(ActiveCategory), CreditCard(Category).

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)

SQL query:

SELECT ci.GeographyID, g.GenderCategory, Round(AVG(ci.EstimatedSalary), 2) AS AvgIncome,

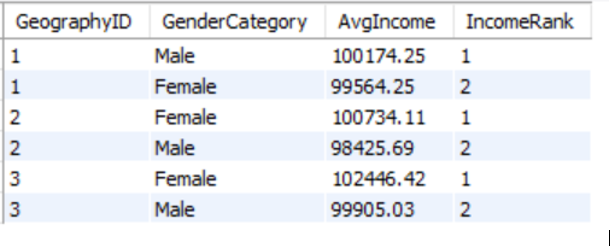
RANK() OVER (PARTITION BY ci.GeographyID ORDER BY AVG(ci.EstimatedSalary) DESC) AS IncomeRank

FROM CustomerInfo ci

JOIN Gender g ON ci.GenderID = g.GenderID

GROUP BY ci.GeographyID, g.GenderCategory;

Result table:



Insight:

* We got for GeographyID as 1 Male has highest AvgIncome but for others GeographyID (like 2, 3) Female has highest AvgIncome.

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+).

SQL query:

SELECT

CASE

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

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

ELSE '50+'

END AS AgeBracket,

Round(AVG(bc.Tenure), 2) AS AvgTenure

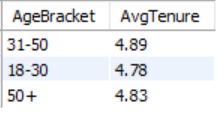
FROM CustomerInfo ci

JOIN BankChurn bc ON ci.CustomerId = bc.CustomerId

WHERE bc.Exited = 1

GROUP BY AgeBracket;

Result table:



Insight:

* Average Tenure for 18-30 is 4.78, for 31-50 is 4.89 and for 50+ is 4.83.

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?

SQL query:

SELECT bc.Exited, CORR(ci.EstimatedSalary, bc.Balance) AS Correlation

FROM CustomerInfo ci

JOIN BankChurn bc ON ci.CustomerId = bc.CustomerId

GROUP BY bc.Exited;

* Since for my laptop CORR function not supported. So, I used basic method to find correlation.

SELECT

bc.Exited,

Round((

SUM((ci.EstimatedSalary - avg\_salary) \* (bc.Balance - avg\_balance)) /

(SQRT(SUM(POWER(ci.EstimatedSalary - avg\_salary, 2))) \*

SQRT(SUM(POWER(bc.Balance - avg\_balance, 2))))

), 2) AS Correlation

FROM CustomerInfo ci

JOIN BankChurn bc ON ci.CustomerId = bc.CustomerId

JOIN (

SELECT Exited, AVG(EstimatedSalary) AS avg\_salary, AVG(Balance) AS avg\_balance

FROM CustomerInfo ci2

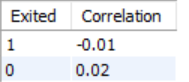
JOIN BankChurn bc2 ON ci2.CustomerId = bc2.CustomerId

GROUP BY Exited

) avgs ON bc.Exited = avgs.Exited

GROUP BY bc.Exited;

Result table:



Insight:

* Since Correlation for Exited = 1 is -0.01 and for exited=0 is 0.02. In both the case absolute value towards zero. It means both are not correlated.

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

SQL query:

SELECT

Round((

SUM((ci.EstimatedSalary - avg\_salary) \* (bc.CreditScore - avg\_credit)) /

(SQRT(SUM(POWER(ci.EstimatedSalary - avg\_salary, 2))) \*

SQRT(SUM(POWER(bc.CreditScore - avg\_credit, 2))))

), 2) AS SalaryCreditScoreCorrelation

FROM CustomerInfo ci

JOIN BankChurn bc ON ci.CustomerId = bc.CustomerId

JOIN (

SELECT

AVG(EstimatedSalary) AS avg\_salary,

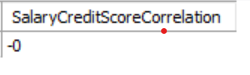
AVG(CreditScore) AS avg\_credit

FROM CustomerInfo ci2

JOIN BankChurn bc2 ON ci2.CustomerId = bc2.CustomerId

) avgs ON 1=1;

Result table:



Insight:

* I got the Salary Credit Score Correlation value as -0. means no correlation.

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

SQL query:

SELECT

CASE

WHEN CreditScore >= 800 THEN 'Excellent'

WHEN CreditScore >= 740 THEN 'Very Good'

WHEN CreditScore >= 670 THEN 'Good'

WHEN CreditScore >= 580 THEN 'Fair'

ELSE 'Poor'

END AS CreditBucket,

COUNT(\*) AS ChurnCount,

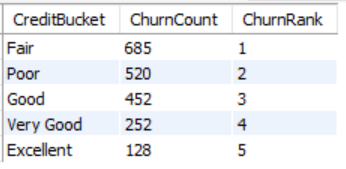
RANK() OVER (ORDER BY COUNT(\*) DESC) AS ChurnRank

FROM BankChurn

WHERE Exited = 1

GROUP BY CreditBucket;

Result table:



Insight:

* We got Fair as a rank 1 and Excellent as rank 5.

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

SQL query:

WITH AgeCredit AS (

SELECT

CASE

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

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

ELSE '50+'

END AS AgeBucket,

COUNT(CASE WHEN bc.HasCrCard = 1 THEN 1 END) AS CreditCardCount

FROM CustomerInfo ci

JOIN BankChurn bc ON ci.CustomerId = bc.CustomerId

GROUP BY AgeBucket

),

AvgCredit AS (

SELECT AVG(CreditCardCount) AS AvgCards FROM AgeCredit

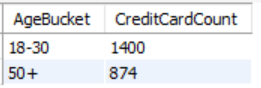
)

SELECT AgeBucket, CreditCardCount

FROM AgeCredit, AvgCredit

WHERE CreditCardCount < AvgCards;

Result table:



Insight:

* We got AgeBucket as 18-30 has 1400 CreditCardCount and 50+ has 874 counts.

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

SQL query:

SELECT g.GeographyLocation,

COUNT(CASE WHEN bc.Exited = 1 THEN 1 END) AS ChurnCount,

Round(AVG(bc.Balance), 2) AS AvgBalance,

RANK() OVER (ORDER BY COUNT(CASE WHEN bc.Exited = 1 THEN 1 END) DESC) AS ChurnRank

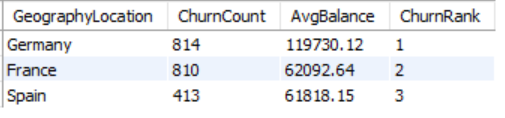
FROM BankChurn bc

JOIN CustomerInfo ci ON bc.CustomerId = ci.CustomerId

JOIN Geography g ON ci.GeographyID = g.GeographyID

GROUP BY g.GeographyLocation;

Result table:



Insight:

* From above query we got, Rank 1 for Germany and 3 for Span.

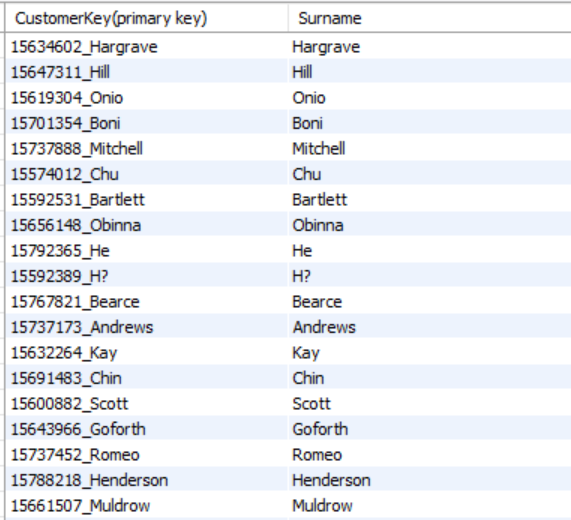
1. As we can see that the “CustomerInfo” table has the CustomerID and Surname, now if we have to join it with a table where the primary key is also a combination of CustomerID and Surname, come up with a column where the format is “CustomerID\_Surname”.

SQL query:

SELECT CONCAT(CustomerId, '\_', Surname) AS `CustomerKey(primary key)`, Surname

FROM CustomerInfo;

Result table:



Etc…

* I took the screenshot of few rows and showing in Result table section.

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

* Yes

SQL query:

SELECT CustomerId,

(SELECT ExitCategory FROM ExitCustomer ec WHERE ec.ExitID = bc.Exited) AS ExitCategory

FROM BankChurn bc;

Result table:



Etc…

* I took the screenshot of few rows and showing in Result table section.

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

* No, there is no missing values in the data.

Tool Used:

* Power BI or Python (Pandas),

Methods:

* Imputation (mean/median/mode),
* Domain-specific replacement,
* Deletion (if sparse),

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”.

SQL query:

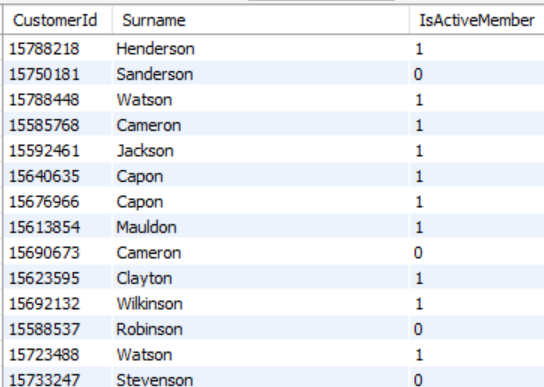
SELECT ci.CustomerId, ci.Surname, bc.IsActiveMember

FROM CustomerInfo ci

JOIN BankChurn bc ON ci.CustomerId = bc.CustomerId

WHERE ci.Surname LIKE '%on';

Result table:



Etc…

* I took the screenshot of few rows and showing in Result table section.

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.

* A customer cannot be both exited and active at the same time.
* To detect the Discrepancy, I used below query.

SQL query:

SELECT bc.CustomerId, bc.IsActiveMember, bc.Exited

FROM BankChurn bc

WHERE bc.IsActiveMember = 1 AND bc.Exited = 1;

* We got 735 records, means there are data quality issues.
* These rows represent customers who are incorrectly labelled as active despite having exited.
* Since the Exited column is confirmed to be accurate, the issue lies in the IsActiveMember flag.
* To fix this we can create a new column as CorrectedIsActiveMember or we can update the value in same column (IsActiveMember).
* Below, I am adding new column as CorrectedIsActiveMember. Along with CustomerId, IsActiveMember and Exited column. Also I update the flag (IsActiveMember) to 0 when Exited = 1 and IsActiveMember = 1.

SELECT

bc.CustomerId,

bc.IsActiveMember,

bc.Exited,

CASE

WHEN Exited = 1 THEN 0

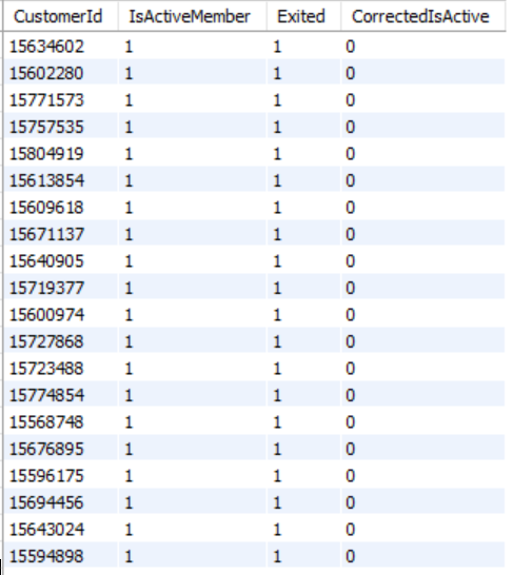
ELSE IsActiveMember

END AS CorrectedIsActive

FROM BankChurn bc

WHERE bc.IsActiveMember = 1 AND bc.Exited = 1;

Result table:



Etc…

* I took the screenshot of few rows and showing in Result table section.

**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?

SQL query:

SELECT

CASE WHEN Tenure > 5 THEN 'Long-Term' ELSE 'New' END AS CustomerType,

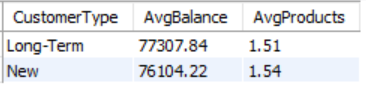
Round(AVG(Balance), 2) AS AvgBalance,

Round(AVG(NumOfProducts), 2) AS AvgProducts

FROM BankChurn

GROUP BY CustomerType;

Result table:



Insight:

* The data shows that long-term customers maintain a slightly higher average balance than new customers, suggesting stronger financial engagement over time.
* However, new customers use a marginally greater number of products, indicating initial experimentation with offerings.
* This suggests that loyal customers tend to focus on fewer, more substantial products, reflecting trust and satisfaction with the bank.
* Overall, loyalty appears linked more to quality of engagement rather than sheer quantity of product usage.

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

SQL query:

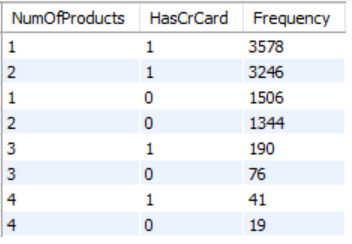
SELECT NumOfProducts, HasCrCard, COUNT(\*) AS Frequency

FROM BankChurn

GROUP BY NumOfProducts, HasCrCard

ORDER BY Frequency DESC;

Result table:



Insight:

* The results show that customers with 1 or 2 products and a credit card (HasCrCard = 1) are the most common groups, with 3,578 and 3,246 customers respectively.
* This indicates a strong affinity between holding a credit card and having multiple products, suggesting cross-selling opportunities around credit card holders.
* Lower frequencies for customers without credit cards imply banks can focus on promoting credit cards alongside other products to increase engagement.
* These patterns can guide targeted marketing and bundled offers to boost product adoption and deepen customer relationships.

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

SQL query:

SELECT g.GeographyLocation,

COUNT(CASE WHEN bc.Exited = 1 THEN 1 END) AS ChurnCount,

COUNT(CASE WHEN bc.IsActiveMember = 1 THEN 1 END) AS ActiveCount

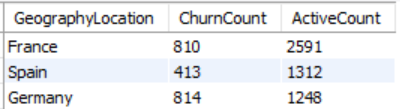
FROM BankChurn bc

JOIN CustomerInfo ci ON bc.CustomerId = ci.CustomerId

JOIN Geography g ON ci.GeographyID = g.GeographyID

GROUP BY g.GeographyLocation;

Result table:



Insight:

* The data shows that France has the highest number of active customers (2,591) with a churn count of 810, indicating a relatively stable but significant churn.
* Germany has a nearly equal churn count (814) but a much lower active customer base (1,248), suggesting a higher churn rate and weaker customer retention.
* Spain has the lowest churn (413) and active customers (1,312), reflecting moderate customer stability.
* These variations imply that geographic regions have different levels of customer loyalty, which banks should consider when tailoring retention and marketing strategies.

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

SQL query:

SELECT

CASE

WHEN Age < 30 THEN 'Young(<30)'

WHEN Age BETWEEN 30 AND 50 THEN 'Mid-Age(30-50)'

ELSE 'Senior(50+)'

END AS AgeGroup,

Round(AVG(CreditScore), 2) AS AvgCreditScore,

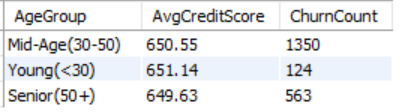
COUNT(CASE WHEN Exited = 1 THEN 1 END) AS ChurnCount

FROM CustomerInfo ci

JOIN BankChurn bc ON ci.CustomerId = bc.CustomerId

GROUP BY AgeGroup;

Result table:



Insight:

* The data shows that the Mid-Age group (30–50) has the highest churn count at 1,350, indicating a higher risk of customers leaving the bank.
* Despite the churn, the average credit scores are similar across all age groups, around 650, so creditworthiness is comparable.
* The Young group (<30) has the lowest churn count, suggesting lower risk in terms of customer attrition.
* Therefore, the Mid-Age segment poses the highest financial risk mainly due to higher churn, which can impact the bank’s stable revenue and growth.

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?

SQL query:

SELECT Age, GeographyID, HasCrCard, IsActiveMember,

AVG(Tenure) AS AvgTenure,

Round(AVG(Balance), 2) AS AvgBalance,

AVG(CreditScore) AS AvgCreditScore,

Round(AVG(EstimatedSalary), 2) AS AvgSalary,

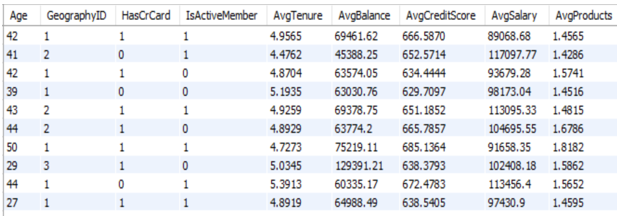
AVG(NumOfProducts) AS AvgProducts

FROM CustomerInfo ci

JOIN BankChurn bc ON ci.CustomerId = bc.CustomerId

GROUP BY Age, GeographyID, HasCrCard, IsActiveMember;

Result table:



Etc…

* I took the screenshot of few rows and showing in Result table section.

Insight:

* To model and predict lifetime (tenure) value for customer segments, the query groups customers by age, geography, credit card ownership, and activity status, then calculates average tenure within each segment.
* These results reveal which combinations (e.g., active credit card holders in specific regions and age brackets) tend to stay longer with the bank, identifying high-value customer segments.
* By identifying segments with higher average tenure, the bank can target retention strategies, prioritize valuable customer groups, and provide evidence-based guidance for resource allocation and personalized service offers to maximize long-term customer 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?
   * + Evaluate how marketing campaigns influence:
       - Customer retention (preventing churn)
       - Customer acquisition (bringing in new customers)
     + We need to define these key metrics:
       - Retention Rate, Acquisition Rate, Conversion Rate and Churn Rate Change
     + Required Data: In current dataset includes customer profiles and churn status, but does not include campaign data. To perform this analysis, you’d need:
       - CampaignID,
       - CampaignDate,
       - TargetSegment,
       - AcquisitionDate, and
       - RetentionStatus
     + Once Data is available. We use Visualization chart like Funnel, Line or Bar chart to show insights.
2. Customer Exit Reasons Exploration: Can you identify common characteristics or trends among customers who have exited that could explain their reasons for leaving?

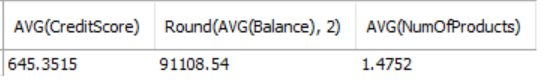
SQL query:

SELECT AVG(CreditScore), Round(AVG(Balance), 2), AVG(NumOfProducts)

FROM BankChurn

WHERE Exited = 1;

Result table:



Insight:

* The average credit score of customers who exited is relatively low at around 645, indicating potentially higher credit risk.
* Their average balance is moderately high (about ₹91,108), suggesting they had significant funds, which might make their churn impactful.
* The average number of products held (1.475) shows they were engaged but not deeply diversified in bank offerings.
* These trends suggest that customers leaving may be financially valuable but perhaps dissatisfied or seeking better options, emphasizing the need for proactive retention strategies focused on service quality and personalized offers.

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

* To find importance of these feature, we have to evaluate some matrics. Below query help to evaluate metrics...

SQL query:

SELECT

AVG(CASE WHEN Exited = 1 THEN Tenure ELSE NULL END) AS AvgTenure\_Exited,

AVG(CASE WHEN Exited = 0 THEN Tenure ELSE NULL END) AS AvgTenure\_Retained,

AVG(CASE WHEN Exited = 1 THEN NumOfProducts ELSE NULL END) AS AvgProducts\_Exited,

AVG(CASE WHEN Exited = 0 THEN NumOfProducts ELSE NULL END) AS AvgProducts\_Retained,

AVG(CASE WHEN Exited = 1 THEN EstimatedSalary ELSE NULL END) AS AvgSalary\_Exited,

AVG(CASE WHEN Exited = 0 THEN EstimatedSalary ELSE NULL END) AS AvgSalary\_Retained,

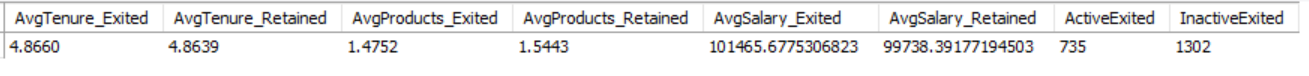
COUNT(CASE WHEN Exited = 1 AND IsActiveMember = 1 THEN 1 END) AS ActiveExited,

COUNT(CASE WHEN Exited = 1 AND IsActiveMember = 0 THEN 1 END) AS InactiveExited

FROM BankChurn bc

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

Result table:



Insight:

* IsActiveMember: is the most influential feature for predicting customer churn—customers who are inactive are nearly twice as likely to exit the bank.
* NumOfProducts: also shows a meaningful impact, with lower product usage correlating with higher churn, suggesting that deeper engagement reduces attrition.
* While EstimatedSalary and Tenure: show only marginal differences between exited and retained customers, they may still contribute when combined with other variables.
* Overall, focusing on reactivating inactive users and promoting multi-product adoption could significantly improve retention outcomes.

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

SQL query:

SELECT

g.GenderCategory,

CASE

WHEN Age < 30 THEN 'Young(<30)'

WHEN Age BETWEEN 30 AND 50 THEN 'Mid-Age(30-50)'

ELSE 'Senior(50+)'

END AS AgeGroup,

bc.NumOfProducts,

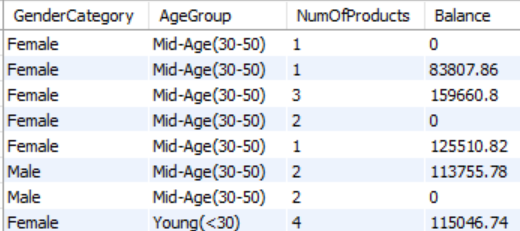
bc.Balance

FROM CustomerInfo ci

JOIN BankChurn bc ON ci.CustomerId = bc.CustomerId

JOIN Gender g ON ci.GenderID = g.GenderID;

Result table:



Etc…

* I took the screenshot of few rows and showing in Result table section.

Insight:

* This segmentation query groups customers by gender, age group, number of products, and account balance, providing a detailed profile of customer diversity.
* The output allows the bank to identify patterns, such as which demographic is associated with higher product usage or larger balances.
* For example:
  + both male and female customers in the 'Mid-Age(30-50)' group show variation in balance and product holdings, while younger customers may sometimes hold more products but with varied balances.

* Such insights enable targeted marketing, product recommendations, and risk assessment for specific customer segments.

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?

* Use the following indicators from your dataset:
  + Exited = 1 -> Confirmed churn
  + IsActiveMember = 0 -> Inactive
  + NumOfProducts < 2 -> Low engagement
  + CreditScore < 600 -> Financial risk
  + Balance = 0 -> Dormant account

* Then use conditional formating facility in Excel or Power BI and Charts to visually highlight customers at risk of churn.

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?

SQL query:

* churn rate per year:

SELECT YEAR(`Bank DOJ`) AS Year,

ROUND(COUNT(CASE WHEN bc.Exited = 1 THEN 1 END) \* 100.0 / COUNT(\*), 2) AS `ChurnRate(in %)`

FROM BankChurn bc

JOIN CustomerInfo ci ON bc.CustomerId = ci.CustomerId

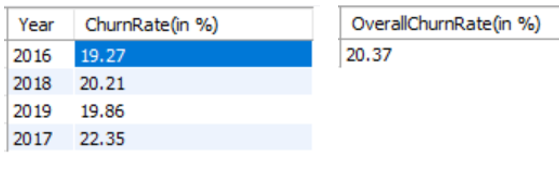
GROUP BY YEAR(`Bank DOJ`);

* Overall Churn:

SELECT ROUND(COUNT(CASE WHEN Exited = 1 THEN 1 END) \* 100.0 / COUNT(\*), 2) AS `OverallChurnRate(in %)`

FROM BankChurn;

Result table:



Insight:

* The overall churn rate at the bank is approximately 20.37%, which is higher than the typical benchmark of 5–10% for banks and may indicate underlying customer retention challenges.
* Customers most likely to churn often have lower product usage, lower activity levels (inactive members), and sometimes lower engagement or satisfaction based on previous analysis.
* Strategies to decrease churn include increasing customer engagement through personalized communications, regularly seeking feedback to address pain points, and offering more relevant or flexible products and services.
* The bank should use data segmentation to identify at-risk groups and prioritize proactive retention initiatives and customer satisfaction improvements.

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

* We will create a dashboard with the help of Power BI and focus on these things...
* KPIs: Churn Rate, Avgerage Balance, Product Usage, Tenure.
* Visuals: Bar charts, line graphs, heatmaps.
* Slicers: Geography, Age Group, Gender, Credit Score Bucket.
* Tool: Power BI or Tableau.

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

* We approach this problem like:

1. Explore data schema and relationships.

2. Identify key business goals (e.g., retention, risk).

3. Perform EDA (Exploratory Data Analysis).

4. Build hypotheses (e.g., “Does tenure affect churn?”).

5. Use SQL + visual tools to validate insights.

6. Present findings in a structured report.

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

SQL query:

ALTER TABLE BankChurn

RENAME COLUMN HasCrCard TO Has\_creditcard;

****