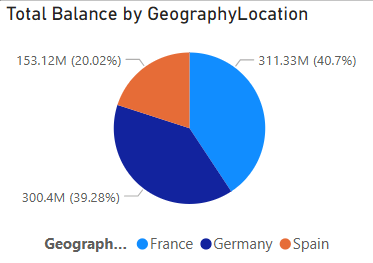
**Objective Questions:**

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

Below Column chart shows different regions and their total account balance.

We can see that France has the highest total balance followed by Germany.

And Spain has the least account balance distribution.



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

**Code:**

select CustomerId, Surname, EstimatedSalary, Bank\_DOJ

from customerinfo

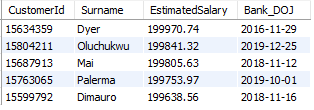
where quarter(Bank\_DOJ) = 4

order by EstimatedSalary desc

limit 5;

Above query returns details of top 5 customers with highest estimated salary who are joined in last quarter of a year.

**Output:**



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

**Code:**

select avg(NumofProducts) as Avg\_numof\_products

from bank\_churn

where HasCrCard = 1;

**Output:**



We can see that, the average number of products used by the customers who have a credit card is 1.5314.

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

Churn rate is exited customers by total customers. And it is done in power BI by using measures.

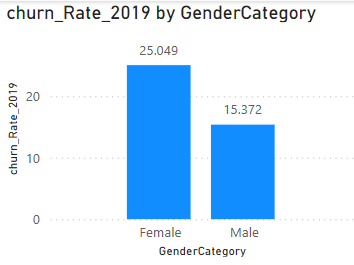
**Formulas used:**

lost\_cust = CALCULATE( DISTINCTCOUNT(CustomerInfo[CustomerId]), Bank\_Churn[Exited] = 1,YEAR( CustomerInfo[Bank DOJ])=2019 )

Tot\_cust = CALCULATE( DISTINCTCOUNT(CustomerInfo[CustomerId]), YEAR( CustomerInfo[Bank DOJ])=2019 )

churn\_Rate = ROUND( DIVIDE( [lost\_cust],[Tot\_cust] )\*100 ,3)

Above churn rate gives rate of exited customers in the last year 2019.



We can see that Churn rate for Males in 2019 is 15.37.

Churn rate for Females in 2019 is 25.05.

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

**Code:**

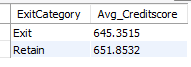
select ec.ExitCategory, avg(bc.CreditScore) as Avg\_Creditscore

from bank\_churn bc

join exitcustomer ec on bc.exited=ec.exitId

group by ec.ExitCategory;

**Output:**



We can say that, The average credit score for the exited people is 645.35.

The average credit score for the have remained people is 651.85.

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

**Code:**

select GenderCategory, round(avg(EstimatedSalary),2) as Avg\_estimated\_sal, count(bc.CustomerId) as Active\_Customers

from customerinfo ci

join gender g on ci.genderID=g.genderID

join bank\_churn bc on bc.customerID=ci.customerID

join activecustomer ac on bc.IsActiveMember=ac.ActiveID

where ActiveCategory='Active Member'

group by GenderCategory

**Output:**

A screenshot of a computer

Description automatically generated

We can see that,

Men have higher active accounts than women whereas the average salary of the women with active accounts is slightly higher than the men.

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

Lets create some segments for customers based on their credit score.

* The customers who have a credit score over 781 as “Excellent”
* The customers with credit score between 701 and 780 as “Very Good”.
* The customers with credit score between 611 and 700 as “Good”
* The customers with credit score between 510 and 610 as “Good” and the rest as “Poor”

After this segregation we can find number of exited customers on each credit score segment

**Code:**

with CreditScoreSegment as (

select CustomerId, Exited,

case when creditscore between 781 and 850 then 'Excellent'

when creditscore between 701 and 780 then 'Very Good'

when creditscore between 611 and 700 then 'Good'

when creditscore between 510 and 610 then 'Fair' else 'Poor'

end as CreditScoreSegment

from bank\_churn )

select CreditScoreSegment,

avg(case when Exited = 1 then 1 else 0 end) as Exit\_Rate

from creditscoresegment

group by creditscoresegment

order by exit\_rate desc

limit 1;

**Output:**



The “Poor” segment has the highest exit rate among all the other segments.

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

**Code:**

select g.GeographyLocation, count(b.CustomerId) as Active\_Customers

from geography g

join customerinfo c on g.geographyid = c.geographyid

join bank\_churn b on c.customerid = b.customerid

where b.tenure > 5 and b.IsActiveMember=1

group by g.geographylocation

order by active\_customers desc

limit 1;

**Output:**



We can see that the geographic region with highest active customers and with a tenure greater than 5 years is France with 797 active customers.

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

Below column chart we see the distribution of total customers among has or no credit cards and exit or retain factor.

A graph of a credit card

Description automatically generated

We can say that more customers exited who are having credit card than those who are not having credit card.

Looks like there is an impact of credit card on customer churn.

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

The below chart gives us the distribution of total exited customers based on the number of products they have used.

A graph with blue squares

Description automatically generated

From the data it is very clear that, the most number of products used by exited customers is 1 followed by 2 and 3.

And least used number of products is 4 among exited customers.

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.

**Code:**

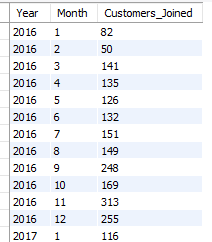
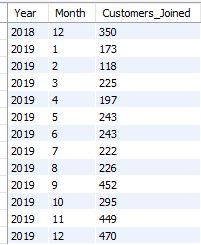
select year(Bank\_DOJ) as Year ,month(Bank\_DOJ) as Month, count(\*) as Customers\_Joined

from customerinfo

group by year(Bank\_DOJ),month(Bank\_DOJ)

order by year,month

**Output in MySQL:**

**Output in Power BI:**

Total customers joined every month in every year.

A graph of a line

Description automatically generated

The customer joining trend appears to be increasing and decreasing frequently but overall, it is increasing.

This indicates a growth in customer acquisition.

Every year, least number of customers joined are in January and Highest number of customers join the month of July

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

The chart tells about the distribution of the average balance of exited customers over Number of Products used.

A graph of blue rectangular bars

Description automatically generated with medium confidence

The average balance of exited customers is almost same. Where customers with 4 products used are highest in average balance and with 3 products are least.

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

**We can use scatter plots for detecting outliers. It** provides a more visual approach by displaying individual data points and their relationships. Outliers appear as points far removed from the main cluster of data, making it easy to spot customers with extreme values, such as balances that don’t align with expected trends.

We can find outliers by their appearance in the plot or by the formulas.

I have used below DAX, which creates a calculated coloumn to tell which point is “normal” and which is ” outlier”.

Outlier\_Flag =

VAR Q1\_Value = PERCENTILEX.INC(ALL(Bank\_Churn), Bank\_Churn[Balance], 0.25)

VAR Q3\_Value = PERCENTILEX.INC(ALL(Bank\_Churn), Bank\_Churn[Balance], 0.75)

VAR IQR\_Value = Q3\_Value - Q1\_Value

VAR LowerBound = Q1\_Value - (1.5 \* IQR\_Value)

VAR UpperBound = Q3\_Value + (1.5 \* IQR\_Value)

RETURN

    IF( Bank\_Churn[Balance] < LowerBound || Bank\_Churn[Balance] > UpperBound, "Outlier", "Normal" )

And gave “outlier\_flag” column to the scatter plot as a legend.

A blue dotted line graph

Description automatically generated with medium confidence

We can see from the plot, it is clear that there are no outliers in the data.

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

The dataset has seven different tables they are, Active Customer, Bank Churn, Credit Card, Customer Info, Exit Customer, Gender, Geography.

**Categorical columns** in a dataset contain **discrete values** that represent categories or groups rather than numerical measurements.

These columns are used for **classification** and **grouping** rather than mathematical operations.

**Tables and their Categorical Variables:**

* Customer Info: GeographyID, GenderID are categorical variables.
* Bank\_Churn: HasCrCard, isActiveMember and Exited are categorical variables.
* Exit Customer: ExitID and ExitCategory are categorical variables.
* Gender: GenderID and GenderCategory are categorical variables.
* Geography: GeographyID and GeographyCategory are categorical variables
* Active Customer: ActiveID and Active Category are categorical variables
* Credit Card: CreditID and CreditCategory are categorical variables.

Among the 7 tables, there are 5 tables which consists of only categorical variables. They are Exit Customer, Gender, Geography, Active Customer and Credit Card.

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.

This below query calculates the average income for males and females within each geographic location and assigns rank based on the average salary.

**Code:**

select geo.GeographyLocation, GenderCategory,

round(avg(c.estimatedsalary),2) as Avg\_salary,

rank() over (partition by GeographyLocation

order by avg(c.EstimatedSalary) desc) as 'Rank'

from customerinfo c

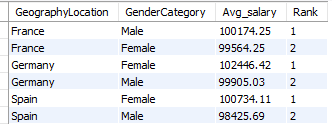
join geography geo on c.geographyid = geo.geographyid

join gender gn on gn.genderid=c.genderid

group by geo.geographylocation, GenderCategory

order by geo.geographylocation

**Output:**



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+).

To find the average tenure for each age bracket we first segregate the user base into three segments.

* People with age from 18-30 as **Adults.**
* People with age from 31-50 as **Middle Aged.**
* People with age above 50 as **Old Aged.**

**Code:** select case when age between 18 and 30 then 'Adults'

when age between 31 and 50 then 'Middle-aged'

else 'Old-aged' end as Age\_brackets,

avg(b.tenure) as Avg\_tenure

from customerinfo c

join bank\_churn b on c.customerid = b.customerid

where b.exited = 1

group by Age\_brackets

order by Age\_brackets

**Output:**

A screenshot of a computer

Description automatically generated

We can say almost every age bucket has almost same Average tenure.

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?

Below is the scatter plot for Salary and Balance.

A blue dots on a white background

Description automatically generated

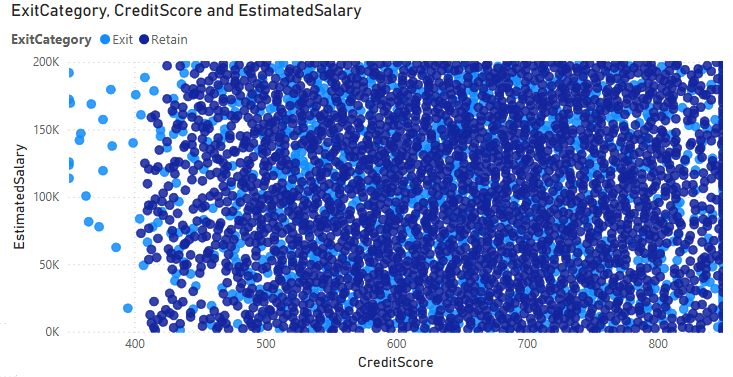
The above scatter plot does not show a strong or clear relationship between these variables.

The data points are evenly distributed across different salary and balance ranges, indicating no consistent pattern where higher or lower salaries correspond to higher or lower balances.

Additionally, the two **Exited** categories (represented by different shades) are dispersed throughout the chart, suggesting that customer churn is not directly influenced by salary or balance based on this visual representation.

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

Below is the scatter plot for Salary and CreditScore.



There is no strong or clear correlation between Credit Score and Estimated Salary. The data points are widely and evenly distributed across all ranges, indicating no evident linear relationship.

Additionally, the Exited status (0 or 1) appears uniformly spread across different credit scores and salary levels, suggesting that neither factor alone is a strong predictor of customer churn.

While some variation exists, the graph does not visually reveal any distinct pattern or correlation between these variables and exit behavior.

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

Below is the table I created to rank Credit Score Buckets by total churned Customers.

A screenshot of a computer

Description automatically generated

DAX formulas Used:

Credit Score Bucket = IF(Bank\_Churn[CreditScore]>=800,"Excellent",IF(Bank\_Churn[CreditScore]>=740, "Ve ry Good", IF(Bank\_Churn[CreditScore]>=670,"Good", IF(Bank\_Churn[CreditScore] >=580,"Fair","Poor"))))

Tot\_Exit\_cust = CALCULATE( DISTINCTCOUNT(CustomerInfo[CustomerId]), ExitCustomer[ExitCategory]="Exit")

Cr\_Score\_Ranked = RANKX( all( Bank\_Churn[Credit Score Bucket]), CALCULATE( DISTINCTCOUNT( CustomerInfo[CustomerId]), ExitCustomer[ExitCategory]="Exit" ) )

And remaining columns are just dragged and dropped in the matrix.

We can say that, the majority of exit customers comes into the "Good" credit score bucket, followed by "Poor," "Very Good," and then "Excellent."

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.

The below query first segregate the customers based on the age into age brackets and then find those buckets that have lesser than average number of credit cards per bucket.

**Code:**

with creditinfo as (

select case when age between 18 and 30 then 'Adult'

when age between 31 and 50 then 'Middle-aged'

else 'Old-aged' end as agebrackets,

count(c.customerid) as CrCard\_holders

from customerinfo c

join bank\_churn b on c.customerid = b.customerid

where b.hascrcard = 1

group by agebrackets

)

select \*

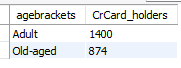
from creditinfo

where CrCard\_holders < (

select avg(CrCard\_holders)

from creditinfo );

**Output:**



Both Adult and Old Aged age Buckets have customers more than the average of all the age buckets.

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

Below is the table I created to Geography Location by total Customers.

A screenshot of a computer

Description automatically generated

DAX formula Used:

Tot\_Exit\_cust = CALCULATE( DISTINCTCOUNT(CustomerInfo[CustomerId]), ExitCustomer[ExitCategory]="Exit")

Geo\_rank = RANKX( ALL(Geography[GeographyLocation]), CALCULATE(DISTINCTCOUNT( CustomerInfo[CustomerId])) )

And remaining columns are just dragged and dropped in the matrix.

We can say that, the majority of exit customers are in "Germany", followed by "France" and then "Spain".

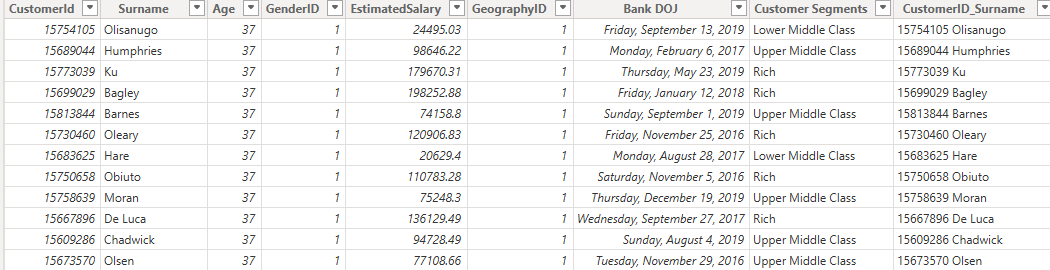
1. As we can see that the “CustomerInfo” table has the CustomerID and Surname, now if we have to join it with a table where the primary key is also a combination of CustomerID and Surname, come up with a column where the format is “CustomerID\_Surname”.

We can create “CustomerID\_Surname” column in Power BI by using create/add a column. Formula used:

CustomerID\_Surname = CustomerInfo[CustomerId]&" "&CustomerInfo[Surname]

Output can be seen in the Data view along with the table.

**Output:**



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

Yes, we can get Exit Category form the table ExitCustomers to bank\_churn table without using any joins between them.

By using Sub-Queries this can be done based upon the common column between the tables.

**Code:**

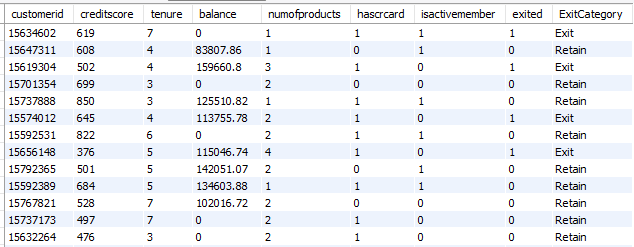
select customerid, creditscore, tenure, balance,

numofproducts, hascrcard, isactivemember, exited,

(select ExitCategory from exitcustomer ec where bc.exited = ec.exitID) as ExitCategory

from bank\_churn bc;

**Output:**



From the output we can see ExitCategory is returned form ExitCategory table and 1 in Exited column means Exit and 0 means Retain in ExitCategory.

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 weren’t any missing values in the data. This eliminates the need for imputation techniques that might introduce assumptions or biases.

If there were missing values in the data, I would use these techniques to handle them:

* **Deletion**: This involves removing rows or columns with missing values.
* **Imputing missing values**: Fill missing values using statstical methods like mean, median, or values from other related columns.
* **Replace Missing Values with a Default Value**

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”.

Below SQL query retrieves the customer's details like customerID,Last\_name and ActivityStatuts whose surname ends with “on”.

**Code:**

select c.CustomerId, c.Surname as Last\_name,

case when b.isactivemember = 1 then 'active'

else 'inactive' end as activitystatus

from customerinfo c

join bank\_churn b on c.customerid = b.customerid

where c.surname like '%on'

order by c.surname;

**Output:**

A screenshot of a computer

Description automatically generated

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.

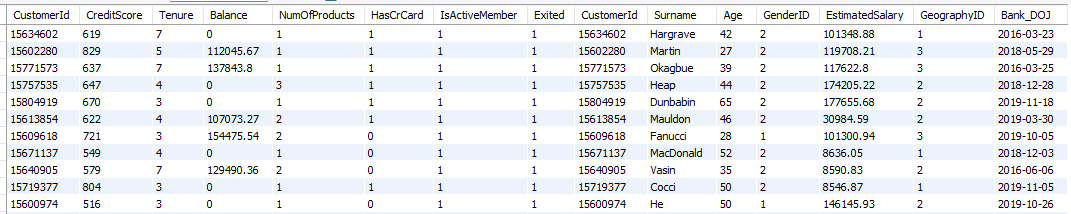
**Code:**

select \*

from bank\_churn b

join customerinfo c on b.customerid = c.customerid

where b.exited =1 and b.isactivemember =1;

**Output:** 

Yes, there is data discrepancy in the Customer’s data.

When the customer is Active which mean the customer is still in the bank and hence the data in IsActiveMember column is 1.

When the customer exits the bank, the Exited column is 1. But here there is data which tells us that the customer Exited and is an active member at the same time.

**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?

Let’s categorize customers into **long-term** and **short-term** based on their joining date. Customers who joined in **2019** are considered **new customers**, while those who joined in **previous years** are classified as **long-term customers**.

Below chart shows, total Old and New customers and their total products bought through the Bank’s credit card.

A graph of customer satisfaction

Description automatically generated with medium confidence

Old customers consistently bought more products than new customers, indicating that they use the bank’s credit card more.

The Total Number of products bought by 6.7k old customers is 10.2k.

The Total Number of products bought by 3.3k old customers is 5.1k.

A graph with blue lines and numbers

Description automatically generated

Above Chart shows Monthly Average products bought by Customers.

There is a steady increase in the average number of products purchased by new customers, suggesting that they adopt more bank services over time as they become familiar with them.

In January, new customers boughtt significantly fewer products than long-term customers, likely because they are still exploring the bank's offerings and setting up their accounts.

A pie chart with text

Description automatically generated

Above we can see tenure of the customers and products bought by customers aggregated by tenure period

Customers with 5, 6, 7 years of tenure are long term customers mean while customers with 3, 4 are short term customers.

Customers with a tenure of 4,5,6 are buying a greater number of products than the customers with a tenure of 3 or 7.

It is because of the less customers with a tenure of 3 or 7 years.

And presence of long-term customers shows belief that customers had in the bank.

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

Customers often use specific bank products together, and identifying these pairings can help the bank develop effective cross-selling strategies. These strategies enhance customer satisfaction and drive revenue by offering relevant products that complement existing ones.

**Commonly Paired Bank Products:**

* Checking Accounts
* Debit Cards
* Savings Accounts
* Credit Cards
* Loans

**Cross-Selling Strategies:**

* **Offer complementary products:** Customers with checking accounts may benefit from a debit card and online banking for easier money management. Similarly, those with savings accounts could be encouraged to set up automatic transfers or explore high-interest options like CDs for larger balances.
* **Personalize recommendations based on usage:** Customers who frequently use credit cards for travel could be offered travel rewards cards, while those applying for loans might find value in bundled insurance products for added convenience and protection.
* **Leverage digital platforms:** Promote features like paperless statements and automatic bill payments through online or mobile banking. Additionally, highlight investment options and financial planning tools that customers can access digitally.

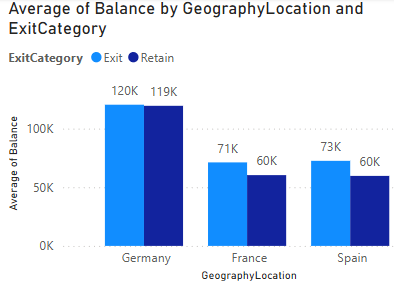
By analyzing product usage and tailoring recommendations, the bank can enhance customer experience while boosting revenue, ensuring that financial needs are met more effectively.

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

Let’s consider economic indicators like Balance, Estimated Salary and Credit Score to know about active and exited customers.

And consider Retain customers of Exit Category as Active customers.

And Below are the 3 different charts which provide insights on how economic indicators vary across geographic regions and their potential correlation with customer churn rates.



The above chart displays the average balance by location, with Germany having the highest balance. While it doesn’t directly indicate churn, regions with lower balances may experience higher churn rates if customers in those areas are more cost sensitive.

A graph of blue bars

Description automatically generated

The chart presents the average credit score by location, we can see all the locations have the nearly same average credit score of exit and active customers. When combined with churn rate data, this could help identify a potential link between **higher credit scores and lower churn rates**.

A graph of a number of blue bars

Description automatically generated

This chart shows the average salary by geographical location. Like the credit score chart, average salary is not varying much among the locations.

But when compared, the customers from Germany are relatively more active account holders in the bank and also maintain the best credit score and bank balance.

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

Customers can be divided into four different segments based on their Estimated salary. They are Lower Middle Class, Poor, Rich and Upper Middle Class.

Used DAX to create Calculated column call customer Segment.

Dax formula:

Customer Segments = IF(CustomerInfo[EstimatedSalary]<20000,"Poor",

                    IF(CustomerInfo[EstimatedSalary]<50000,"Lower Middle Class",

                    IF(CustomerInfo[EstimatedSalary]<100000,"Upper Middle Class","Rich")))

A graph of different colored squares

Description automatically generated

Above chart shows total customers in each demographic and they are distributed based on their financial status.

France has the greatest number of Poor and Lower Middle-Class people followed by Germany and Spain. But France has most Rich customers. Followed by Germany and Spain.

A graph of a number of different colored bars

Description automatically generated with medium confidence

Based on the data in the chart, it's clear that the groups “lower middle class” and “poor” are not only have the most customers but also present the highest potential financial risk to the bank.

**Conclusion:**

* While creditworthiness is an important factor in assessing customer risk, other aspects such as income and employment history should also be considered.
* However, creditworthiness remains a strong predictor of a customer’s ability to repay the loans.
* Customers with lower credit scores are statistically more likely to default on loans compared to those with higher scores.
* This indicates that the bank faces greater financial risk when lending to customers in the "Lower Middle Class" and "Poor" segments, as these groups have a higher likelihood of loan defaults.

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?

The chart below presents the average tenure of customers by balance segments and salary segments

Balance Segments: Customers with higher average balances tend to have longer tenures with the bank, possibly due to greater investment in the bank’s products or retention efforts.

* High Salary & Balance > 2 Lakh: This segment has the longest tenure, averaging 5.4 years.
* Zero Balance: Customers with zero balance have the shortest tenure, around 4.2 years across all salary levels.

Salary Segments: There is a weak correlation between salary and tenure. While higher salary customers tend to stay slightly longer, the difference is not significant.

* High Salary: Average tenure ranges between 4.8 to 5.4 years.
* Low Salary & Lower Middle Class: Expected tenure is around 4.5 to 4.9 years.
* Upper Middle Class & Rich: Likely to have an average tenure of 4.8 to 4.9 years.

A screenshot of a graph

Description automatically generated

Balance segments play a more critical role in predicting customer tenure than salary segments.

Banks can use this insight to focus retention strategies on customers in higher balance segments.While also working to enhance customer satisfaction and product adoption across all segments to improve overall tenure.

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?

A structured approach can provide valuable insights into how marketing campaigns influence customer retention and acquisition. By segmenting data, defining key metrics, and incorporating targeted strategies, businesses can optimize their marketing efforts for maximum effectiveness.

**Key Strategies for Measuring Campaign Impact:**

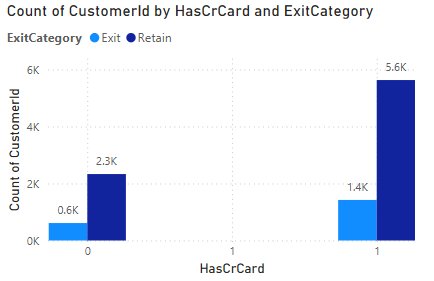
1. **Customer Segmentation:**
   * Group customers by **age, location, and product usage** to analyze campaign performance across different demographics.
   * Identify high-value segments, such as customers over **50**, who may benefit from **special offers and enhanced security features**.
2. **Trend Analysis:**
   * Track **active customer growth, exit rates, and product usage** over time to identify trends linked to marketing efforts.
   * Monitor how **customer behavior changes** in response to specific campaigns.
3. **Campaign Details & Timing:**
   * Maintain detailed records of **campaign content, delivery channels (online/offline), and launch dates**.
   * Link customer behavioral changes to specific marketing efforts for better strategy refinement.
4. **Customer Acquisition Channels:**
   * Track how customers were acquired (e.g., **referrals, online ads**).
   * Assess whether campaigns effectively **retain customers acquired through different channels**.
5. **Customer Lifetime Value (CLV):**
   * Calculate CLV to understand the **long-term financial impact** of campaigns.
   * Prioritize campaigns that **enhance retention and drive revenue growth**.

Enhancing Customer Experience for Better Retention:

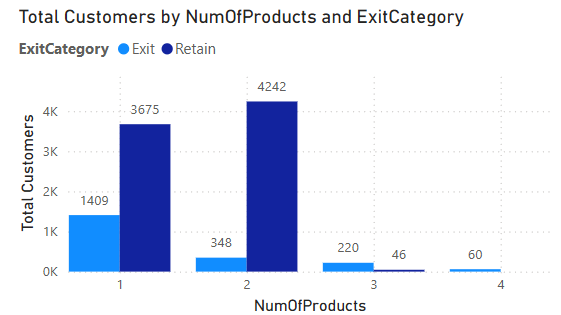
* Improved Customer Service: Provide personalized support, resolve issues quickly, and act on customer feedback to enhance satisfaction.
* Targeted Offers & Promotions: Offer incentives for customers purchasing multiple products and special promotions for credit card holders to encourage long-term engagement.

By integrating these strategies, businesses can make data-driven decisions to improve marketing effectiveness, enhance customer satisfaction, and increase retention and revenue.

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



Above chart can be used to know about the churned customers based on whether they are having credit or not.



Above chart can be used to know about the churned customers based on number of products bought by the customers.

After analyzing above 2 charts, we can say the main reasons for customer exit are as follows:

* **Credit Card Ownership:** The chart reveals that a significantly higher number of customers with credit cards exited the bank (around **1,424**) compared to those without credit cards (around **613**).
* **Fewer Products Purchased:** The highest number of exits occurred among customers who purchased **a single product** (around **1,409**). As the number of products purchased increases, the exit rate decreases. Customers who bought **four products** have the lowest exit rate, with only about **60** exits. This indicates that customers who engage more with the bank by purchasing multiple products are less likely to churn.

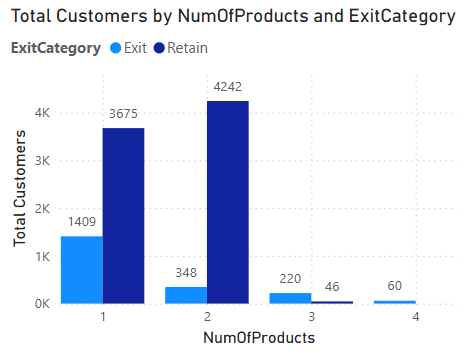
1. Are 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' important for predicting if a customer will leave the bank?

Let’s consider tenure, number of products and estimated salary to predict whether a customer will leave the bank by using exit category for churn and active customer data.

The analysis is based on below 3 charts that visualize these factors for both customers who exited and those who remained.

A screenshot of a computer screen

Description automatically generated

 A screenshot of a computer screen

Description automatically generated

**Tenure:**

The number of years that the customer has been a client of the bank can be a significant factor. Normally, older clients are more loyal and less likely to leave a bank1.

While comparing the relationship with tenure of the customers and the exit rate of the customers we can find a key relationship that the customers who have a tenure of greater than or equal to four years are more likely to retain than others.

**Number of Products:**

The number of products a customer has with the bank could potentially influence their decision to stay or leave. There is no consistent pattern between the number of products a customer holds and their likelihood of exiting.

Customers with multiple products might have a higher level of engagement with the bank, which could reduce the likelihood of churn.

**Estimated Salary:**

While one might assume that a customer’s estimated salary could influence their likelihood to churn, it has been found that estimated salary showed little to no correlation with churn.

This could be due to a variety of factors, such as the customer’s satisfaction with the bank’s services, which might not necessarily be tied to their income3.

Remember, these are general trends and may not apply to every customer. A comprehensive analysis should be conducted using appropriate statistical and machine learning techniques to determine the significance of these factors in predicting customer churn4. It’s also important to note that the importance of these factors can vary depending on the specific context and customer base of the bank

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

The customers are segmented mainly into three categories based on their estimated salary. They are Low, Medium and High segments. These segments are further divided in female and male customers based on their Geography.

select GeographyLocation,

case when estimatedsalary < 50000 then 'Low'

when estimatedsalary < 100000 then 'Medium'

else 'High'end as Income\_Segment, GenderCategory ,

count(c.customerid) as NumberofCustomers

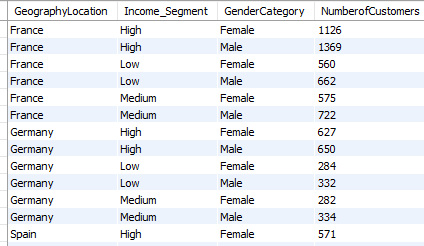
from customerinfo c

join geography g on c.geographyid = g.geographyid

join gender gn on c.genderid=gn.genderid

group by geographylocation, Income\_Segment, GenderCategory

order by geographylocation;



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?

We can do conditional formatting to visually highlight customers who are at risk of churn and to evaluate the impact of credit card rewards on customer retention in Power BI by the following below steps.

* Create a filter for the "HasCrCard" field. This will allow you to segment customers by whether they have a credit card or not.
* Define the criteria to identify customers at risk of churn. This could be based on a combination of factors, such as:

1. Customers with a low number of products purchased (NumOfProducts)
2. Customers in Low and Lower Middle Salary segment.
3. Customers in Zero and Less 2 Lakh Balance segment.

* Apply a conditional formatting rule to highlight cells that meet the churn criteria. You can format the cells with a different background color or font to make them visually distinct.
* For applying multiple conditions, we have to DAX founctions to highlight a cell.

DAX formula Used:

churn risk cust = IF( ( SELECTEDVALUE(Bank\_Churn[NumOfProducts]) = 1 && SELECTEDVALUE(CustomerInfo[Salary Segments]) in {"Low Salary", "Lower Middle Salary"} && SELECTEDVALUE(Bank\_Churn[Balance Segments]) in {"Zero","Less 2 lakh"}  ), 1 ,0)

OUTPUT:

A screenshot of a table with numbers and a number

Description automatically generated

Above table highlights the customers who are in edge of being a churn customer at the bank based on the considered criteria.

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?

**Churn Rate Overview:**

The bank’s overall churn rate is **20.37%**, with year-on-year fluctuations:

* + **2016**: 19.27%
  + **2017**: 22.35% (highest)
  + **2018**: 20.21%
  + **2019**: 19.86% (lowest)

A blue bar graph with numbers

Description automatically generated

Based upon the analysis we did in the before questions.

**Customer Segments Prone to Churn:**

* **Single Product Users**: Customers using only one product may not see enough value in the bank's offerings, particularly when competitors provide broader or more integrated services.
* **Credit Card Holders**: Potential factors for credit card holders churning include:
  + Insufficient credit limits.
  + Lack of appealing rewards programs.
  + High fees associated with the card.
* **Tenure of 4-5 Years**: These customers might be coming off promotional offers or discounts, making them susceptible to competitors offering more attractive rates or features.
* **High Salary Customers**: High earners may have more financial options and are more likely to switch banks for slightly better benefits or interest rates elsewhere.

**Recommendations to Reduce Churn:**

* **Targeted Product Bundles**:
  + Develop tailored product bundles to address the specific needs of customers who use only one product. Highlight the added value and potential cost savings of using multiple services.
* **Enhanced Credit Card Rewards**: Improve the credit card offering by:
  + Increasing credit limits based on customer behavior and creditworthiness.
  + Aligning rewards programs with customer preferences (e.g., travel, cashback for certain categories).
  + Reducing or eliminating annual fees, especially for high-value customers.
* **Retention Offers for Existing Customers**:
  + For customers nearing the end of their introductory offers, proactively offer personalized retention deals.
  + Like extending the promotional rates and Providing discounts on other products or services.
* **Customer Satisfaction Surveys**:
  + Regularly survey customers to gather insights into why they leave. Use this feedback to fine-tune retention strategies and address customer pain points.
* **Relationship Management for High-Value Customers**:
  + Assign dedicated relationship managers to high-value customers, offering them personalized services and exclusive benefits to strengthen loyalty and meet their individual needs.

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

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

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

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

**Structured Approach to Data Analysis**

1. **Understanding the Context:**
   * Gain a clear understanding of the **analysis objectives** and underlying business needs.
   * Consult **stakeholders** or review project documentation to uncover implicit expectations.
2. **Exploratory Data Analysis (EDA):**
   * Examine the dataset’s **structure, distributions, and relationships** to identify key patterns.
   * Detect any **outliers, anomalies, or trends** that may influence the analysis.
3. **Identifying Insights:**
   * Generate **initial insights** based on observed trends and relationships within the data.
   * Consider both **quantitative metrics** and **qualitative factors** relevant to the domain.
4. **Formulating Hypotheses:**
   * Develop **hypotheses** based on early findings to guide further investigation.
   * Use these hypotheses to refine analytical focus and explore potential causations.
5. **Iterative Analysis:**
   * Apply **various analytical techniques** such as **statistical analysis, machine learning, or data modeling**.
   * Utilize tools like **Excel, Power BI, and SQL** to experiment with different approaches and refine insights.
6. **Data Visualization & Communication:**
   * Leverage **charts, graphs, and dashboards** to represent insights effectively.
   * Use **visual storytelling** to highlight key patterns and make findings more accessible to stakeholders.
7. **Synthesizing Findings:**
   * Consolidate insights into a **cohesive narrative**, identifying overarching themes and trends.
8. **Validation & Stakeholder Feedback:**
   * Verify results through **peer reviews, expert assessments, or external benchmarks**.
   * Gather feedback from **stakeholders** to ensure the analysis aligns with business goals and provides actionable recommendations.

This structured approach enhances the effectiveness of data analysis, making insights more meaningful and impactful.

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

“Alter” is a DDL command in SQL. By running the below query we can change the Column name of a table.

**Code:**

alter table Bank\_Churn

rename column HasCrCard to Has\_creditcard ;



select \*

from bank\_churn;

Above code is to see the change of column name from “HasCrCard” to “Has\_creditcard**”.**

A screenshot of a data

Description automatically generated