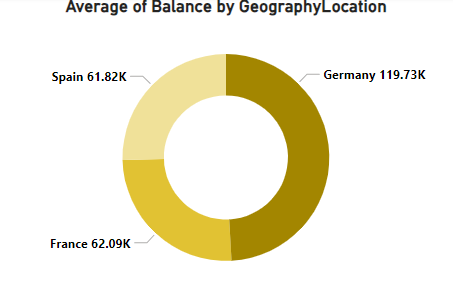
**Analytical CRM Development for a Bank**

**TASKS**

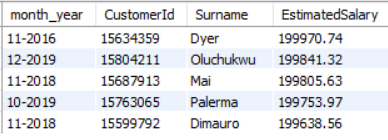
**OBJECTIVE QUESTIONS**

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



The above doughnut chart shows the distribution of average account balances across different regions. The average account balance for Spain, France and Germany are 61.82k, 62.09k and 119.73k respectively.

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



Above is the list of top 5 customers with the highest Estimated Salary in the last quarter of every year.

The query used to get the above result is –

SELECT date\_format(BankDOJ, '%m-%Y') AS month\_year, CustomerId, Surname, EstimatedSalary

FROM customerinfo

WHERE QUARTER(BankDOJ) = 4

ORDER BY EstimatedSalary DESC

LIMIT 5;

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



The above image shows the average products used by customers who have a credit card.

The query used to get the above result is –

SELECT AVG(NumOfProducts) AS AverageProductsUsed

FROM (

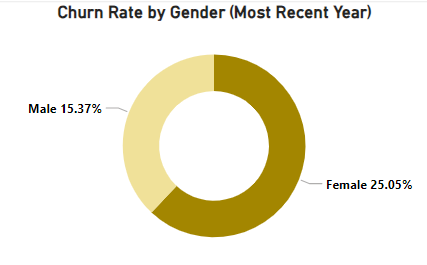
SELECT CustomerId, NumOfProducts

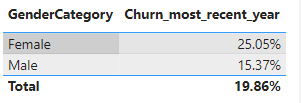
FROM bank\_churn

WHERE HasCrCard = 1

) dt ;

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





The above chart and table show the churn rate by gender for the most recent year.

The DAX formula used to calculate the churn rate is –

Churn\_most\_recent\_year =

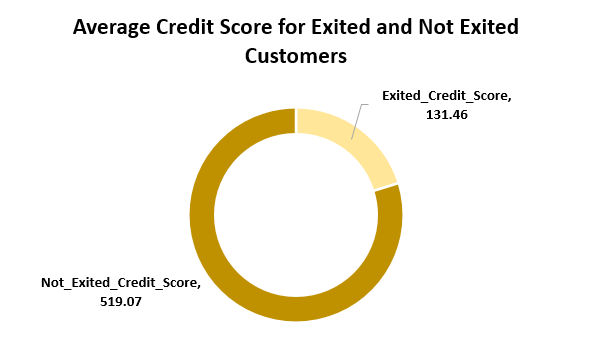
var most\_recent\_year = MAXX(CustomerInfo, YEAR(CustomerInfo[Bank DOJ]))

RETURN

DIVIDE(CALCULATE(COUNT(Bank\_Churn[Exited]), year(CustomerInfo[Bank DOJ]) = most\_recent\_year, Bank\_Churn[Exited] = 1),

                CALCULATE(COUNT(Bank\_Churn[Exited]), YEAR(CustomerInfo[Bank DOJ]) = most\_recent\_year))

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





The above visual shows that the average credit score for customers who have exited and those who remains (Not Exited) are 131.4581 and 519.0707 respectively. The data is derived from the SQL and Excel is used to create the visual.

The query used to get the above result is –

SELECT

AVG(CASE WHEN Exited = 1 THEN CreditScore ELSE 0 END) AS Exited\_Credit\_Score,

AVG(CASE WHEN Exited = 0 THEN CreditScore ELSE 0 END) AS Not\_Exited\_Credit\_Score

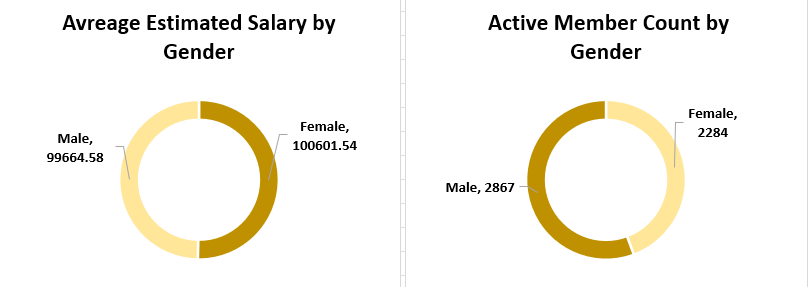
FROM (

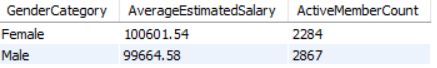
SELECT CustomerId, CreditScore, Exited

FROM bank\_churn

) dt;

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





The above doughnut chart shows the Average Estimated Salary and Active Member Count by Gender. The **average estimated salary for female (100601.54) is higher than male (99664.58), whereas the active female count (2284) is lower than male count (2867)**. The table below is formed using SQL query.

The query used to get the above result is –

SELECT GenderCategory, ROUND(AVG(EstimatedSalary), 2) AS AverageEstimatedSalary,

COUNT(CASE WHEN IsActiveMember = 1 THEN 1 END) AS ActiveMemberCount

FROM customerinfo c

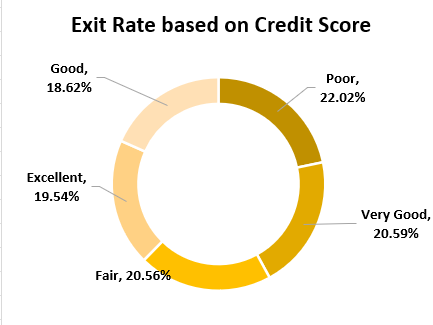
JOIN bank\_churn b ON c.CustomerId = b.CustomerId

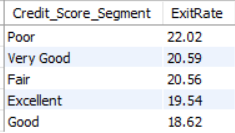
JOIN gender g ON c.GenderID = g.GenderID

GROUP BY 1

ORDER BY 2 DESC;

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

****



The customers are segmented based on Excellent, Very Good, Good, Fair and Poor Credit Score. The segment with highest exit rate is Poor segment and segment with lowest exit rate is very good.

Credit score:

Excellent: 800–850

Very Good: 740–799

Good: 670–739

Fair: 580–669

Poor: 300–579

The query was written in SQL and the doughnut chart was made on Excel.

The query used to get the result is –

SELECT Credit\_Score\_Segment,

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

FROM (

SELECT

CreditScore,

CASE WHEN CreditScore BETWEEN 300 AND 579 THEN 'Poor'

WHEN CreditScore BETWEEN 580 AND 669 THEN 'Fair'

WHEN CreditScore BETWEEN 670 AND 739 THEN 'Good'

WHEN CreditScore BETWEEN 740 AND 799 THEN 'Very Good'

WHEN CreditScore >= 800 THEN 'Excellent'

END AS Credit\_Score\_Segment,

Exited

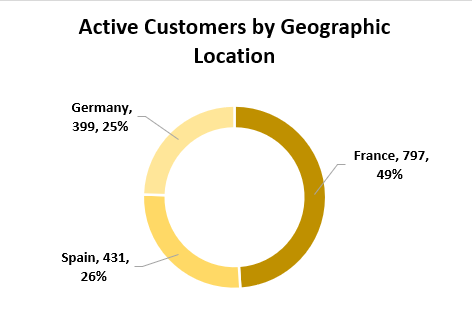
FROM bank\_churn

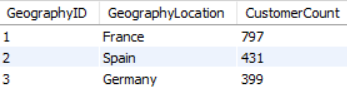
) dt

GROUP BY Credit\_Score\_Segment

ORDER BY ExitRate DESC;

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

****

****

The above doughnut chart shows the active customers in different geographic regions who have a tenure greater than 5 years. France (797) has the highest number of active customers. The query was written in SQL and the doughnut chart was made on Excel.

The query used to get the result is –

SELECT c.GeographyID, GeographyLocation, COUNT(c.CustomerId) as CustomerCount

FROM customerinfo c

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

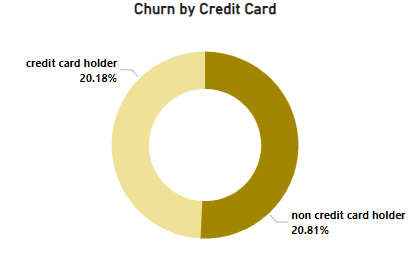
JOIN geography g ON c.GeographyID = g.GeographyID

WHERE b.IsActiveMember = 1 AND Tenure > 5

GROUP BY 1, 2

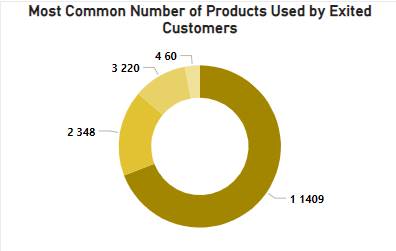
ORDER BY 3 DESC;

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

****

The impact of having a credit card on the customer churn is very low. So, it can be concluded that having a credit card does not influence the churn.

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

****

The above doughnut chart shows the Number of products and Count of each number. A filter is applied on the visual that filters only products that have been used by customers who exited.

Steps for getting the above doughnut chart –

1. Select doughnut chart option in the build visual pane.
2. Drag NumOfProducts to Legends field and NumOfProducts to Values field. Right click the NumOfProducts in Values field and select summarization as Count.
3. Drag the Exited in the Add data filter section on filter pane. Select Basic Filtering and tick mark the box with value 1.
4. **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.**

The above column chart shows the number of customers joined per year. We see that as the year is increasing the number of customers joined is also increasing.

The query used to get this result is –

SELECT YEAR(BankDOJ) as YearOfJoining, COUNT(CustomerId) as CustomersJoined

FROM customerinfo

GROUP BY 1

ORDER BY 1;

The above bar chart shows the number of customers joined each month over a period of 2016 to 2019. The most customers joined in the month of November, December and September and least customers joined in the month of April, January and February.

The query used to get the result is –

SELECT DATE\_FORMAT(BankDOJ, '%M') as MonthOfJoining, COUNT(CustomerId) as CustomersJoined

FROM customerinfo

GROUP BY 1;

The above line chart shows the trend of customer joined over the year 2016 to 2019. From this chart we can get the idea that in year which months got the highest and lowest number of customers.

The query used to get the above result is –

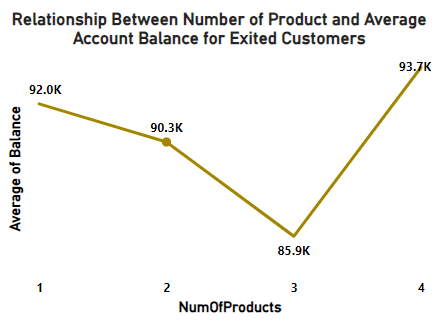
SELECT DATE\_FORMAT(BankDOJ, '%Y-%m') as MonthOfJoining, COUNT(CustomerId) as CustomersJoined

FROM customerinfo

GROUP BY 1

ORDER BY 2 DESC;

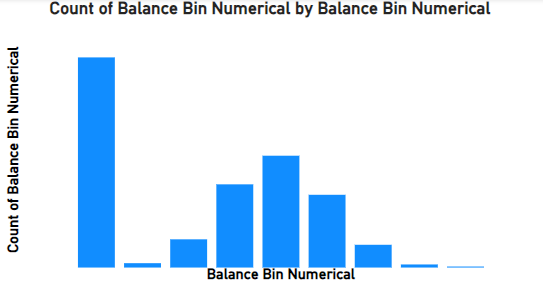
1. **Analyze the relationship between the number of products and the account balance for customers who have exited.**

****

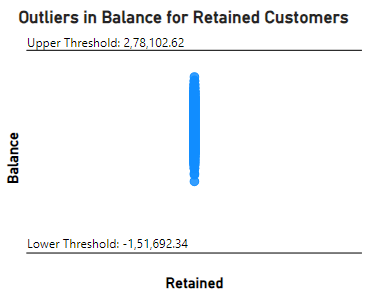
The above line graph shows the relationship between the number of products and the account balance for customers who have exited. We see that the as the number of products increases the account balance decreases till the number of products is 3, then for the customers who have ordered 4 number of products have the highest account balance. The chart was made using Power BI.

Steps for getting the above line chart –

1. Select the line chart from the build visual pane.
2. Drag NumOfProducts in X-axis field and drag Balance in the Y-axis field. Right the Balance in the y-axis field and select summarization as Average.
3. Drag the Exited in the Add data filter section on filter pane. Select Basic Filtering and tick mark the box with value 1.
4. **Identify any potential outliers in terms of balance among customers who have remained with the bank.**

****

The bar chart gives the distribution of Balance data who have retained. We see that this data is not a normal distribution. So, we used Inter-Quartile Range (IQR) method to find the outliers with a multiplier of 1.2 with IQR.

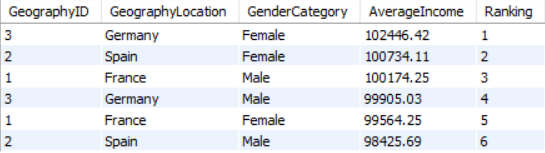


The above scatter plot is drawn between Retained Customers and Balance. There are no outliers present in the Balance data according to the IQR method.

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

There are 7 tables in the dataset. Out of these 7 tables, 5 tables only consist of categorical variables. These tables are – ActiveCustomer, CreditCard, ExitCustomer, Gender and Geography.

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)**

****

The above table shows the gender-wise average income of males and females in each geography id. The data is ranked based on the average income values with highest ranked 1. Germany and Spain has the highest average income for female.

The query used to get the result is –

SELECT c.GeographyID, GeographyLocation, GenderCategory, ROUND(AVG(EstimatedSalary), 2) as AverageIncome,

DENSE\_RANK() OVER(ORDER BY AVG(EstimatedSalary) DESC) as Ranking

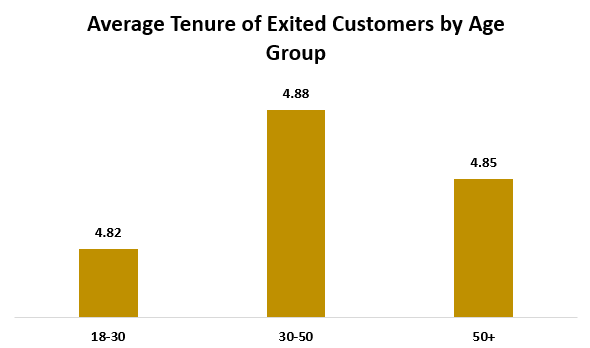
FROM customerinfo c

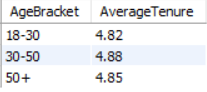
JOIN gender g ON c.GenderID = g.GenderID

JOIN geography geo ON c.GeographyID = geo.GeographyID

GROUP BY 1, 2, 3;

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+).**

****

****

The above table and chart show the average tenure of the customers who have exited in different age bracket. We see that customers in the age group 30-50 who have exited has the highest average tenure of 4.88 and 18-30 age group has the lowest average tenure of 4.82. The tables was formed in SQL and the chart is made in Excel.

The query used to get the result is –

SELECT

CASE WHEN Age BETWEEN 18 AND 29 THEN '18-30'

WHEN Age BETWEEN 30 AND 49 THEN '30-50'

ELSE '50+'

END AS AgeBracket,

ROUND(AVG(b.Tenure), 2) as AverageTenure

FROM customerinfo c

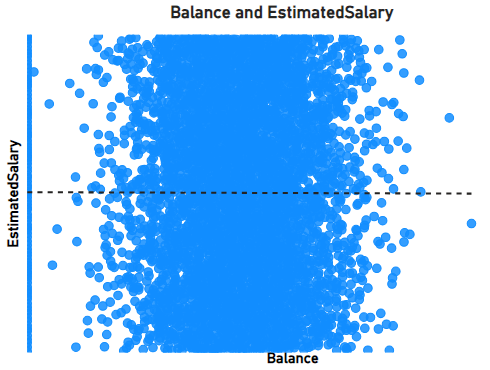
JOIN bank\_churn b ON c.CustomerId = b.CustomerId

WHERE b.Exited = 1

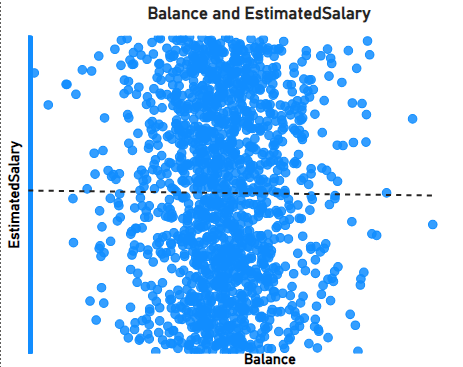
GROUP BY 1

ORDER BY 1;

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?**

****

The above scatter chart shows the correlation between salary and the balance of the customers. Trend line is a straight horizontal line, showing no correlation between balance and estimated salary.



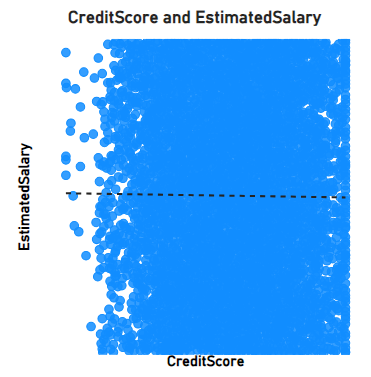
The above chart shows the correlation between balance and estimated salary for customers who have exited. In this case also, we are getting a straight horizontal trend line, showing no correlation. These charts were made using Power BI.

Steps for getting the above charts –

1. Select scatter plot from the Build Visual section.
2. Drag Balance to X Axis filed and Estimated Salary to Y Axis field.
3. In the Analytics Pane, turn ON the Trend Line option.

This will create the first chart.

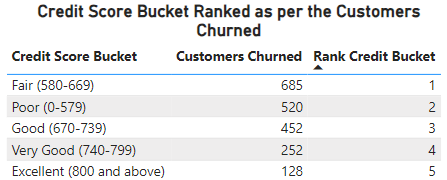
1. For getting the second scatter plot, drag the Exited column to the Add data fields in the field on this visual in the Filter pane.
2. Select Basic filtering from the drop down and select 1.
3. **Is there any correlation between the salary and the Credit score of customers?**

****

The above chart shows the correlation between estimated salary and credit score. A straight horizontal trend lines can be seen which implies that there is no correlation between estimated salary and credit score. The scatter chart is made using Power BI.

Steps for getting the above chart –

1. Select scatter plot from the Build Visual section.
2. Drag CreditScore to X Axis filed and EstimatedSalary to Y Axis field.
3. In the Analytics Pane, turn ON the Trend Line option.
4. **Rank each bucket of credit score as per the number of customers who have churned the bank.**

****

The above table shows the bucket of credit score as per the number of customers churned the bank. If the Customers Churn is high, the rank is low and vice-versa. We see the customers with Excellent (800 and above) credit score has lowest churn number and customers with fair (580-660) has the highest churn number. The above table was created on Power BI.

Steps for getting the above table –

1. Create a measure named “Credit Score Bracket” and a measure named “Rank Credit Bucket”.
2. Select the table option from the Build Visual section.
3. Drag Credit Score Bracket, Exited and Rank Credit Bucket into the Columns.
4. Drag the Exited in the Add data filter section on filter pane. Select Basic Filtering and tick mark the box with value 1.
5. Right click Exited in the Columns and then select summarization as Count.

Expression for creating measures –

1. **Credit Score Bucket =** SWITCH(TRUE(), Bank\_Churn[CreditScore] <= 579, "Poor (0-579)",
2. Bank\_Churn[CreditScore] <= 669, "Fair (580-669)",

Bank\_Churn[CreditScore] <= 739, "Good (670-739)",

Bank\_Churn[CreditScore] <= 799, "Very Good (740-799)",

Bank\_Churn[CreditScore] > 799, "Excellent (800 and above)")

1. **Rank Credit Bucket =**

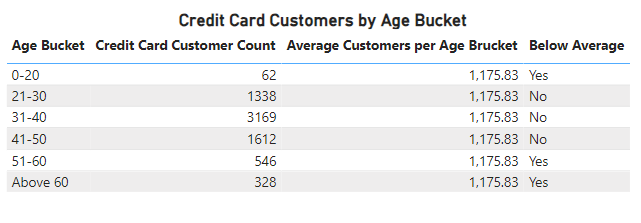
    RANKX(ALL(Bank\_Churn[Credit Score Bucket]),

CALCULATE(COUNT(Bank\_Churn[Exited]), Bank\_Churn[Exited] = 1),

    ,DESC,

Dense)

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 above table shows the number of customers who have a credit card in age buckets. It also shows the average customers per age bucket and has a Below Average column that shows if the customer count is below average or not. From the above table it can be concluded that customers in age bucket 0-20 have the lowest count and bucket 31-40 have the highest count. The above table was made in Power BI.

Steps for getting the above table –

1. Create a calculated column named “Age Bucket”.
2. Create measure “AverageCustomers having CreditCard per AgeBucket” and “Below Average”.
3. Select table from Build Visual section.
4. Drag Age Bucket, hasCrCard, AverageCustomers having CreditCard per AgeBucket and Below Average in the Columns. Name the columns accordingly.
5. Summarize the hasCrCard column to Count.
6. Drag the hasCrCard in the Add data filter section on filter pane. Select Basic Filtering and tick mark the box with value 1.

Expression for creating calculated column and measures –

1. **Age Bucket =** SWITCH(TRUE(), CustomerInfo[Age] <= 20, "0-20",

    CustomerInfo[Age] <= 30, "21-30",

    CustomerInfo[Age] <= 40, "31-40",

    CustomerInfo[Age] <= 50, "41-50",

    CustomerInfo[Age] <= 60, "51-60",

    "Above 60"

    )

1. **AverageCustomers having CreditCard per AgeBucket =**

    AVERAGEX(ALL(CustomerInfo[Age Bucket]),

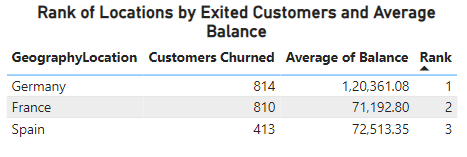
        CALCULATE(COUNT(Bank\_Churn[CustomerId]),

            Bank\_Churn[HasCrCard] = 1))

1. **Below Average =**

IF(CALCULATE(COUNT(Bank\_Churn[HasCrCard]), Bank\_Churn[HasCrCard] = 1) < [AverageCustomers having CreditCard per AgeBucket], "Yes", "No")

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

****

The above table shows the rank of locations as per the number of people who have churned the bank and average balance of the customers. The locations are ranked based on both Customers churned and Average of Balance with Customers Churned having higher weight of 70% and Average Balance having lower weight of 30%. The smaller number have been ranked higher and larger number have been ranked lower. We see that the Customers with location with most churn is Germany and location with least churn is Spain. The following table was made in Power BI.

Steps for getting the above table –

1. Create measures named “Rank based on numCustomers and AvBalance”.
2. Select table from Build Visual section.
3. Drag GeographyLocation, CustomerId, Balance and Rank based on numCustomers and AvBalance respectively in the Columns.
4. Summarize CustomerId as Count and Balance as Average. Change the names accordingly.
5. Drag the Exited in the Add data filter section on filter pane. Select Basic Filtering and tick mark the box with value 1.

Expression for creating measure –

**Rank based on numCustomers and AvBalance =**

RANKX(ALLSELECTED(Geography[GeographyLocation]),

RANKX(ALLSELECTED(Geography[GeographyLocation]), CALCULATE(COUNT(CustomerInfo[CustomerId]), Bank\_Churn[Exited]=1),,ASC,Skip) \* 0.7

+

RANKX(ALLSELECTED(Geography[GeographyLocation]),CALCULATE(AVERAGE(Bank\_Churn[Balance]), Bank\_Churn[Exited]=1),,ASC,Skip) \* 0.3,

,

DESC,Dense)

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”.**

Steps for creating a column named “CustomerID\_Surname” –

1. In the Data pane, click the CustomerInfo table.
2. In the Table tools section, click on New column.
3. Write the DAX expression as given below and press Enter. The column will be created in the CustomerInfo table.

Expression for creating column –

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

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 the “ExitCategory” from ExitCustomers table to Bank\_Churn table without using “Join” in SQL.

The query used to get the result is –

SELECT

bc.\*,

(SELECT ec.ExitCategory

FROM exitcustomer ec

WHERE ec.ExitID = bc.Exited) AS ExitCategory

FROM bank\_churn bc;

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

There were no missing values in the data.

Ways of handling missing values –

1. Removing Missing Data:
   * Remove rows when the missing values are in small fraction (<5%) and dropping them will not impact the analysis significantly.
   * Remove columns when missing values are in large proportion (>50%) and if the column is not essential to the analysis.
2. Imputing Values:
   * Mean Imputation –

Replace missing values with the mean of the available data in the column for numerical data with a symmetric distribution and no extreme outliers.

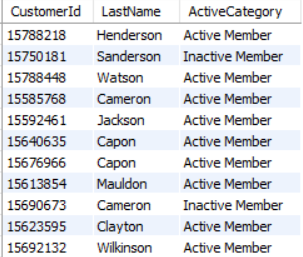
* + Median Imputation –

Replace missing values with the median values for numerical data with skewed distributions or outliers.

* + Mode –

Replace missing values with mode for categorical data.

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”.**

****

The above table is a snippet of the result we got from the analysis. There are 704 users with whose surname ends with “on”. Out of them 366 users are active members.

The query used to get the result is –

SELECT c.CustomerId, c.Surname AS LastName, ac.ActiveCategory

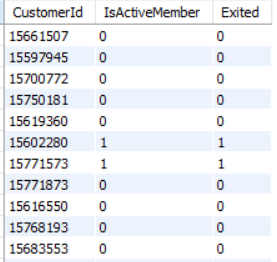
FROM customerinfo c

LEFT JOIN bank\_churn bc ON c.CustomerId = bc.CustomerId

LEFT JOIN activecustomer ac ON bc.IsActiveMember = ac.ActiveID

WHERE c.Surname like '%on';

1. **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.**

****

Out of the 10,000 rows in the provided dataset, 4,282 rows exhibit a data discrepancy where the columns for active membership status and customer exit status contain conflicting values. Specifically, a customer who has exited cannot simultaneously be marked as an active member, yet the data inconsistently assigns values of 1 to both columns, leading to ambiguity and potential inaccuracies in the dataset. The above table is the result of SQL query, that shows the data discrepancy.

The query used to get the result is –

SELECT CustomerId, IsActiveMember, Exited

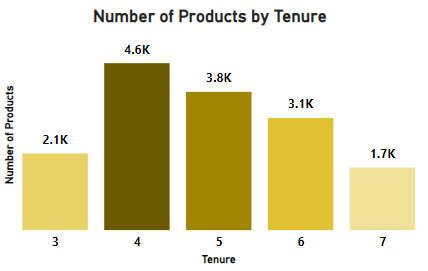
FROM bank\_churn

WHERE (IsActiveMember = 1 AND Exited = 1)

OR (IsActiveMember = 0 AND Exited = 0);

**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?**

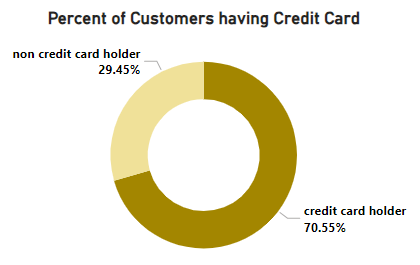


The above column chart shows the number of products bought by customers with tenures from 3 to 7 years. Spending is highest at 4 years, followed by 5 and 6 years, showing strong customer engagement during these years. It’s lower at 3 years and drops further at 7 years.

This suggests that mid-tenure customers are the most loyal, while newer and long-term customers may need more attention.

Steps for getting the above chart –

1. Select Stacked column chart from the Visualizations pane.
2. Drag Tenure to X Axis filed and NumOfProducts to Y Axis field.
3. Format your visual accordingly.
4. **Product Affinity Study: Which bank products or services are most commonly used together, and how might this influence cross-selling strategies?**



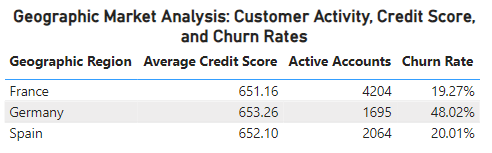
The above doughnut chart shows the percent of customers having credit card. We see that more than 70% of the customers have credit card. So, credit card presents opportunities for cross-selling strategies.

**Popular Combinations with Credit Card –** Since a large portion of customers already have a credit card, the bank could use this product to introduce related services, such as personal loans, insurance, or investment products. Customers who trust the bank with their credit card may be more likely to consider additional financial products.

**Cross-Selling Strategies –** The bank can target credit card holders with tailored offers for products that aligns well with their credit card usage like travel rewards, shopping on specific brands and products, or loans on specific products.

Steps for getting the above chart –

1. Select Doughnut chart from the Visualizations pane.
2. Drag Category from CrediCard Table to Legend filed and HasCrCard to Values field.
3. Format your visual accordingly.
4. **Geographic Market Trends: How do economic indicators in different geographic regions correlate with the number of active accounts and customer churn rates?**

****

The above table shows average credit score, active accounts and churn rate over different geographic regions.

We see that –

1. **France** has the highest number of active accounts and the lowest churn rate, suggesting strong customer retention and a stable customer base.
2. **Germany** has the lowest number of active accounts and the highest churn rate, indicating potential issues with customer retention or dissatisfaction.
3. **Spain** falls in between, with moderate active accounts and churn rate.

There is no clear link between the average credit score and churn rate in the regions, which suggests that other factors like customer service or competition might be affecting churn more than credit scores.

Steps for getting the above chart –

1. Select Table from the Visualizations pane.
2. Drag GeogrphyLoaction, CreditScore, Exited and Churn into the Columns field. Rename the columns as shown in the table.
3. In Filter, apply visual level basic filter as 0 to Exited.
4. Summarize the CreditScore as Average.
5. Format your visual accordingly.

Expression for creating measure –

Churn =

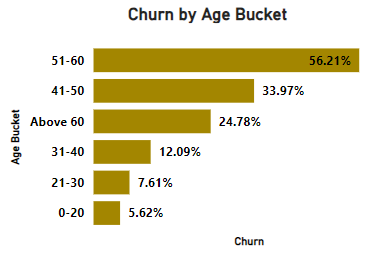
DIVIDE(

    CALCULATE(COUNT(Bank\_Churn[Exited]), Bank\_Churn[Exited] = 1),

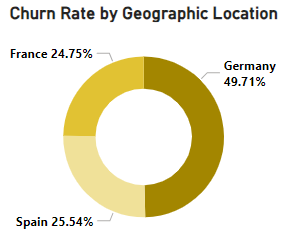
    CALCULATE(COUNT(Bank\_Churn[Exited]))

)

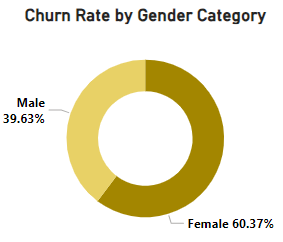
1. **Risk Management Assessment: Based on customer profiles, which demographic segments appear to pose the highest financial risk to the bank, and why?**



The above bar chart shows the churn rate of different age groups. We see that the churn rate is highest for the groups above 40 years of age.



The above doughnut chart shows the churn rate by geographic location. Germany has the highest churn rate.



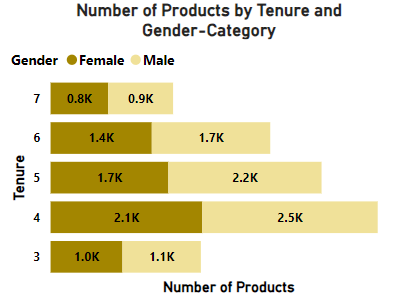
The above doughnut chart shows the churn rate by gender category. Female have higher churn rate than males.

Customer churn is directly related to the financial risk to bank. So, more the churn more is the financial risk. This is the reason that demographic segment with higher churn rate will pose a higher financial risk. From the above graphs we can conclude that –

1. Germany has the highest churn rate of 49.71%. So, has the highest financial risk.
2. Customers with age more than 40 pose higher financial risk.
3. Females with 60.37% churn rate, pose higher financial risk.

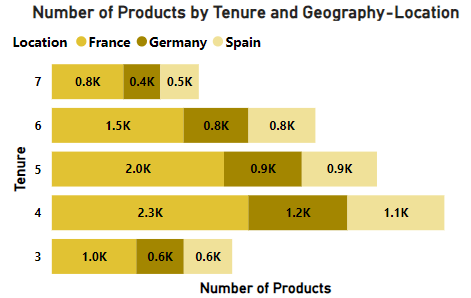
Steps for getting the above charts –

1. Churn by Age Bucket chart –
   1. Select bar chart from the Visualization pane.
   2. Darg Age Bucket into Y-axis and Churn into X-axis.
2. Churn by Geographic Location chart –
   1. Select Doughnut chart from the Visualization pane.
   2. Drag GrographyLocation into Legend section and Churn into Values.
3. Churn by Gender Category chart –
   1. Select Doughnut chart from the Visualization pane.
   2. Drag Gender Category into Legend section and Churn into Values.
4. **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 above stacked bar chart shows the number of products used by gender in different tenures. From the above chart we can conclude that, males have used more products than females.

The customers with highest use of products lies in tenure 4,5 and 6. So, male customers in tenure of 4 years are giving most value to the bank, while female customers in tenure of 7 years are giving least value to the bank.

****

The above stacked bar chart shows the number of products used by customers in different tenure and geographic location. From the above chart, we can conclude that customers in France have used the most and customers in Germany has used the least number of products.

The customers in France with tenure of 4 years use highest number of products contributing most value to the bank, while customers in Spain with tenure of 7 years use the least number of products contributing least value to the bank.

Steps for getting the above charts –

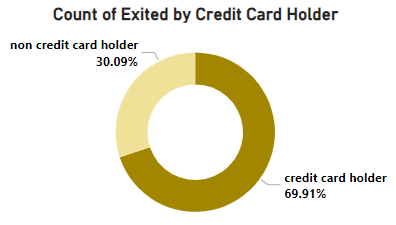
1. Number of Products by Tenure and Gender-Category chart –
   1. Select stacked bar chart from the Visualization pane.
   2. Darg Tenure into Y-axis, NumOfProducts into X-axis and GenderCategory into Legend.
2. Number of Products by Tenure and Geographic Location chart –
   1. Select stacked bar chart from the Visualization pane.
   2. Darg Tenure into Y-axis, NumOfProducts into X-axis and GeographyLocation into Legend.
3. **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 asses the impact of marketing campaigns on customer retention and acquisition –

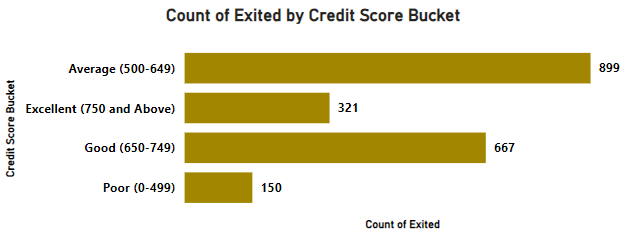
1. Retention – Compare churn rates (using Exited column), and customer tenure before and after the campaign.
2. Acquisition – Identify new customers using Bank DOJ and analyze demographics (Age, GenderCategory, GepgraphyLocation).
3. Customer Value – Balance, CreditScore and NumOfProducts for retained and acquired customers.

Extra information needed to solve this –

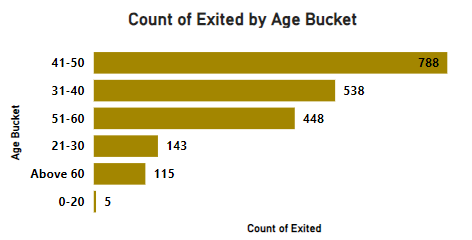
1. Campaign details like type, date.
2. Customer exposure to campaign.
3. Revenue generated.
4. **Customer Exit Reasons Exploration: Can you identify common characteristics or trends among customers who have exited that could explain their reasons for leaving?**

****

The above doughnut chart shows the count of exited customers by credit card holders. From the above chart we can conclude that the customers with credit card have exited more than the non-credit card holders.



The above bar chart shows the count of exited customers by credit score bucket. From the above chart we can conclude that the customers with average score (500-649) have highest exit numbers, followed by customers with good score (650-749).

****

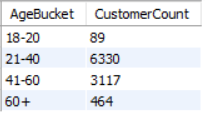
The above bar chart shows the count of exited customers by age bucket. From the above chart we can conclude that the customers begin to exit after the age of 31. As we can see that the highest number of customers have exited from 41-50 age group, followed by 31-40 and 51-60.

Steps for getting the above charts –

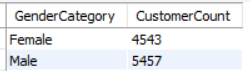
1. Count of Exited by Credit Card Holder chart –
   1. Select Doughnut chart from the Visualization pane.
   2. Darg Category from CreditCard table to Legend and Exited into Values. Summarize Exited as Count.
   3. Drag Exited into the Visual level filter in the Filters pane. Select Basic filtering and then select 1.
2. Count of Exited by Credit Score Bucket chart –
   1. Select bar chart from the Visualization pane.
   2. Darg Credit Score Bucket into Y-axis, Exited into X-axis and summarize Exited as Count.
   3. Drag Exited into the Visual level filter in the Filters pane. Select Basic filtering and then select 1.
3. Count of Exited by Age Bucket chart –
   1. Select bar chart from the Visualization pane.
   2. Darg Age Bucket into Y-axis, Exited into X-axis and summarize Exited as Count.
   3. Drag Exited into the Visual level filter in the Filters pane. Select Basic filtering and then select 1.
4. **Are 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' important for predicting if a customer will leave the bank?**

Yes, these factors are important for predicting customer churn:

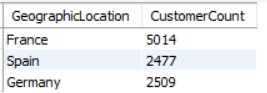
1. Tenure – Indicates customer loyalty and relationship duration. Shorter tenures may signal a higher likelihood of churn.
2. NumOfProducts – Reflects engagement level. Customers with fewer products may be less attached to the bank.
3. IsActiveMember – Active members are typically more engaged and less likely to leave.
4. EstimatedSalary – Could influence churn indirectly, as customers with higher salaries might seek better financial products elsewhere.
5. **Utilize SQL queries to segment customers based on demographics and account details.**

****

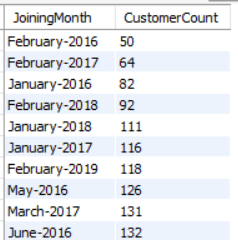
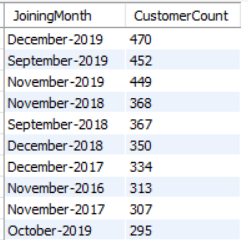
The above table shows the segmentation of customers based on age. It shows that majority of the customers age is in the range of 21-60.



The above table shows the gender-wise count of customers. It shows that the male numbers (5457) are higher than the female numbers (4543).

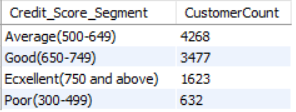


The above table shows the count of customers in different geographic locations. It shows that Fance, Spain and Germany have 5014, 2477 and 2509 customers respectively.

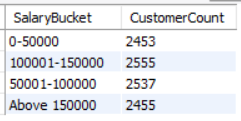
****

The above tables show the count of customers joined in month. The first table shows the 10 months in which highest number of customers joined while the second table shows the 10 months in which the lowest number of customers joined.

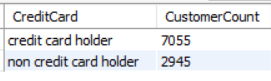
We see the customers mostly join in the months of September, October, November and December. The months in which least customers join are January and February.

****

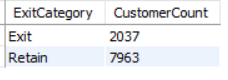
The above table shows the count of customers in different credit score segment. Most of the customers are in the average segment (4268), followed by good segment (3477).



The above table shows the count of customers for different salary buckets. The majority of customers lies in the salary bucket of 50001-100000 and 100001-150000.



The above table shows the count of customers based on having a credit card or not. It shows that 7055 customers have credit card whereas 2945 customers don’t have a credit card.



The above table shows the count of customers who have exited and retained. 7963 customers have retained while 2037 customers have exited.

The queries used to get the results are –

1. SELECT

CASE WHEN Age BETWEEN 0 AND 20 THEN '18-20'

WHEN Age BETWEEN 21 AND 40 THEN '21-40'

WHEN Age BETWEEN 41 AND 60 THEN '41-60'

ELSE '60+'

END AS AgeBucket,

COUNT(\*) AS CustomerCount

FROM customerinfo

GROUP BY 1

ORDER BY 1;

1. SELECT GenderCategory, COUNT(\*) AS CustomerCount

FROM customerinfo c

JOIN gender g ON c.GenderID = g.GenderID

GROUP BY 1;

1. SELECT GeographyLocation AS GeographicLocation, COUNT(\*) AS CustomerCount

FROM customerinfo c

JOIN geography g ON c.GeographyID = g.GeographyID

GROUP BY 1;

1. SELECT DATE\_FORMAT(BankDOJ, '%M-%Y') as JoiningMonth, COUNT(CustomerId) as CustomerCount

FROM customerinfo

GROUP BY 1

ORDER BY 2

-- ORDER BY 2 DESC

lIMIT 10;

1. SELECT

CASE WHEN CreditScore BETWEEN 300 AND 499 THEN 'Poor(300-499)'

WHEN CreditSCore BETWEEN 500 AND 649 THEN 'Average(500-649)'

WHEN CreditScore BETWEEN 650 AND 749 THEN 'Good(650-749)'

ELSE 'Ecxellent(750 and above)' END AS Credit\_Score\_Segment,

COUNT(CustomerId) AS CustomerCount

FROM bank\_churn

GROUP BY 1;

1. SELECT

CASE WHEN EstimatedSalary BETWEEN 0 AND 50000 THEN '0-50000'

WHEN EstimatedSalary BETWEEN 50001 AND 100000 THEN '50001-100000'

WHEN EstimatedSalary BETWEEN 100001 AND 150000 THEN '100001-150000'

ELSE 'Above 150000'

END AS SalaryBucket,

COUNT(\*) AS CustomerCount

FROM customerinfo

GROUP BY 1

ORDER BY 1;

1. SELECT Category as CreditCard, COUNT(CustomerId) as CustomerCount

FROM bank\_churn b

JOIN creditcard cc ON cc.CreditID = b.HasCrCard

GROUP BY 1;

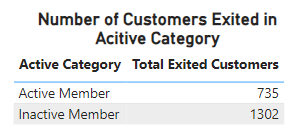
1. SELECT ExitCategory, COUNT(CustomerId) as CustomerCount

FROM bank\_churn b

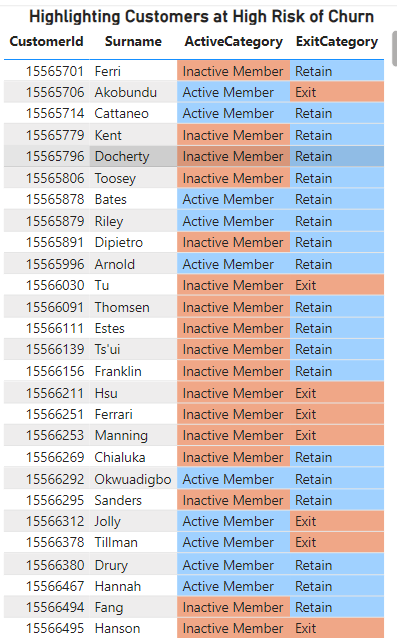
JOIN exitcustomer ec ON b.Exited = ec.ExitID

GROUP BY 1;

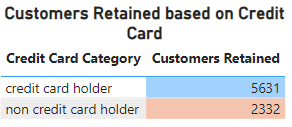
1. **How can we create a conditional formatting setup to visually highlight customers at risk of churn and to evaluate the impact of credit card rewards on customer retention?**



The above table shows the customer activity and the number of exited customers. We can conclude that the inactive members have exited more than the active members.



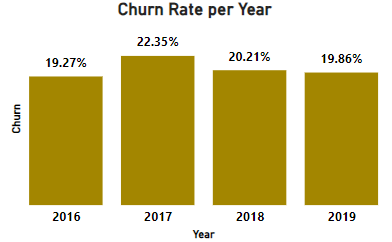
The above table is a snippet of the data. It highlights the customers at high risk of churn. As we have already seen from previous table that the inactive customers are mostly likely to churn. So, in this table inactive members and Customers who have exited are highlighted. The inactive members and the customers who have exited are highlighted in light red whereas the active members and the retained customers are highlighted in the light blue color.



The above table shows the customers retained based on credit card. The customers with credit card are retained more than the customers without credit card. The credit card category with higher value is highlighted in light blue whereas the category with lower number is highlighted in light red.

Steps for getting the above charts –

1. Number of Customer Exited in Active Category chart –
   1. Select table chart from the Visualization pane.
   2. Darg ActiveCategory and Exited in the Columns. Summarize Exited as Count.
   3. In the visual level filter, add Exited column. Select Basic filtering, check mark 1.
2. Highlighting Customers at High Risk of Churn chart –
   1. Select table chart from the Visualization pane.
   2. Darg CustomerId, Surname, ActiveCategory and ExitCategory in the Columns.
   3. Right click the ActiveCategory in the columns and select Background color from Conditional Formatting. Then on the Background color window, select Format Style as Rule, apply the rules according the table above. Click OK.
   4. Right click the ExitCategory in the columns and select Background color from Conditional Formatting. Then on the Background color window, select Format Style as Rule, apply the rules according the table above. Click OK.
3. Customers Retained based on Credit Card chart –
   1. Select table chart from the Visualization pane.
   2. Darg Category from CreditCard table and Exited in the Columns. Summarize Exited as Count.
   3. In the visual level filter, add Exited column. Select Basic filtering, check mark 0.
   4. Right click the Exited in the columns and select Background color from Conditional Formatting. Then on the Background color window, select Format Style as Gradient, apply the rules according the table above. Click OK.
4. **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?**

****

The above column chart shows the Churn Rate per year. It can be seen that the churn rate increased from 2016 to 2017 and decreased from 2017 to 2019.



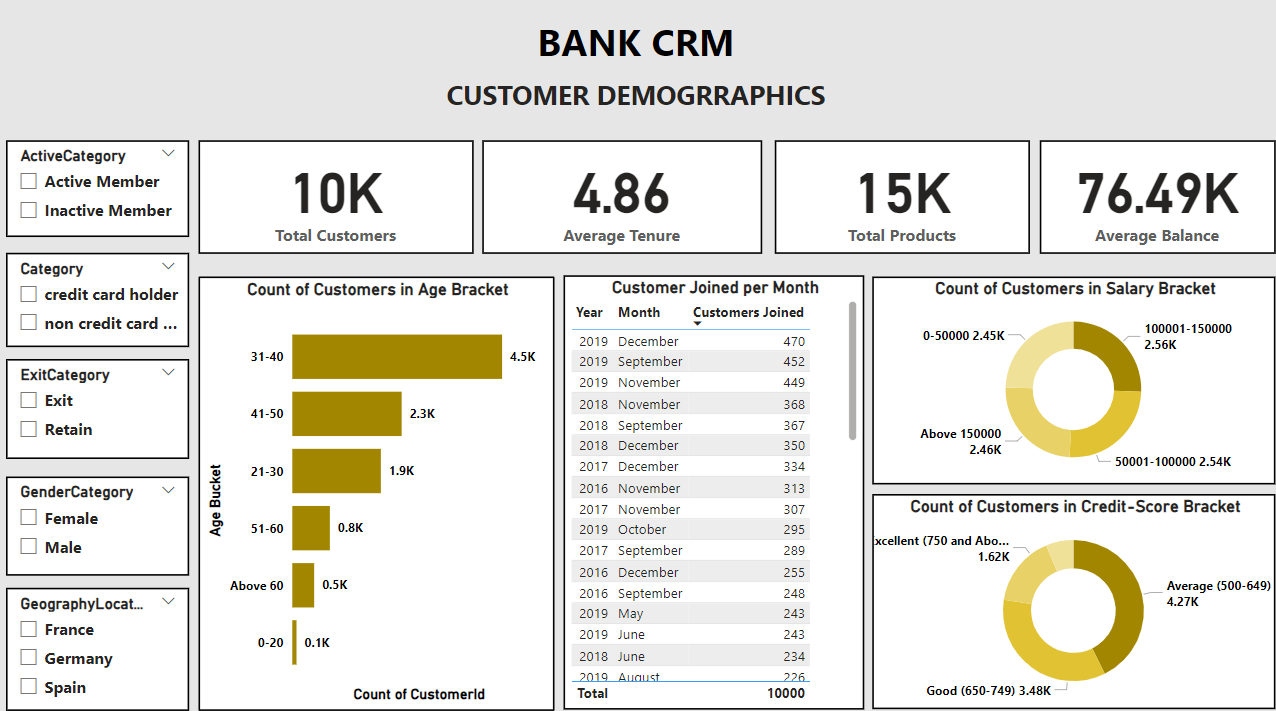
The above card shows that the overall churn rate is 20.37%.

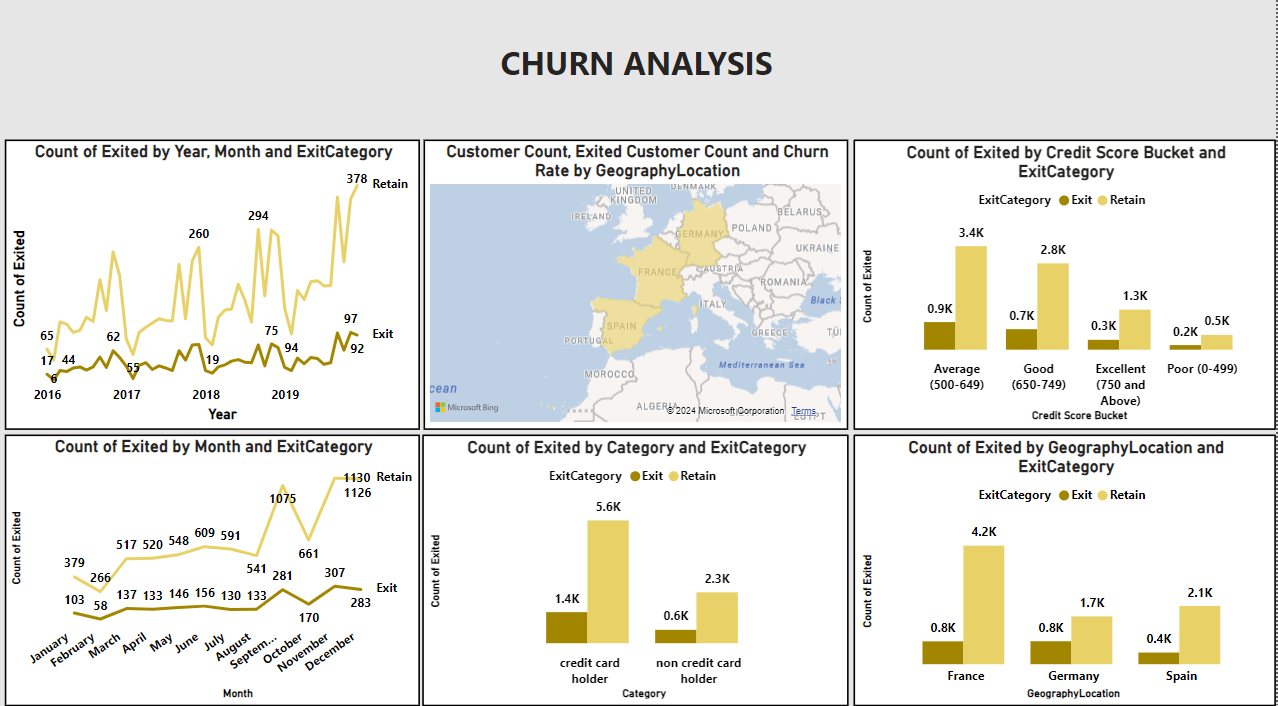
Insights on customers most likely to churn –

1. Customers in the above 40 years are most likely to exit [Refer Subjective question 4, Chart – Churn by Age Bucket].
2. Female customers (60.37%) have churned more than male customers (39.63%) in the past years [Refer Subjective question 4, Chart – Churn by Gender Category].
3. The number of customers having a credit card have churned (69.91%) more than the non-credit card customers (30.09%) [Refer Subjective question 7, Chart – Count of Exited by Credit Card Holder]. The customers with average and good credit score have exited more [Refer Subjective question 7, Chart – Count of Exited by Credit Scor Bucket].
4. The members who are Inactive have exited (1302) more than the active customers (735) [Refer Subjective question 10, Chart – Number of Customers in Active Category].

Steps for getting the above charts –

1. Churn Rate per Year chart –
   1. Select column chart from the Visualization pane.
   2. Darg Bank DOJ into X-axis and Churn into Y-axis.
   3. Do to formatting as per requirement.
2. Highlighting Customers at High Risk of Churn chart –
   1. Select card chart from the Visualization pane.
   2. Drag Churn into Fields. Change the name from Churn to Overall Churn Rate.
3. **Create a dashboard incorporating all the KPIs and visualization-related metrics. Use a slicer in order to assist in selection in the dashboard.**

****

****

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

Bank CRM (Customer Relationship Management) is the way banks use tools, technology, and strategies to manage and understand their interactions with customers. It helps banks build better relationships, keep customers happy, and encourage loyalty. The main goal is to improve customer satisfaction and grow the bank’s business.

If objective and subjective questions weren’t given, then the approach I would have used –

1. **Analyzing Customer Demographics –** Group customers based on their age, gender, location, and income to find the most valuable ones and offer services that match their needs.
2. **Analyze Churn –** Use tenure, credit score, and exit status to identify patterns in exiting customers and highlight at-risk groups.
3. **Enhance Retention –** Target high-risk segments with personalized rewards, loyalty programs, and personalized offers.
4. **Regional Insights –** Compare churn and behaviour trends across regions to customize strategies.
5. **Life Time Value –** Use data like how long customers stay, the products they use, and their financial details to predict how valuable they will be to the bank over time.

If we have the campaign data, then we can calculate the customer acquisition.

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

In SQL –

ALTER TABLE Bank\_Churn

RENAME COLUMN HasCrCard TO Has\_creditcard;

In Power BI –

Steps –

1. In Power BI Desktop, click on **Transform Data** from the ribbon to open the Power Query Editor.
2. In the left panel, select the **Bank\_Churn** table.
3. Right click on the column header **HasCrCard**.
4. Select rename from the drop-down menu or double-click the column header.
5. Type **Has\_creditcard** and press Enter.