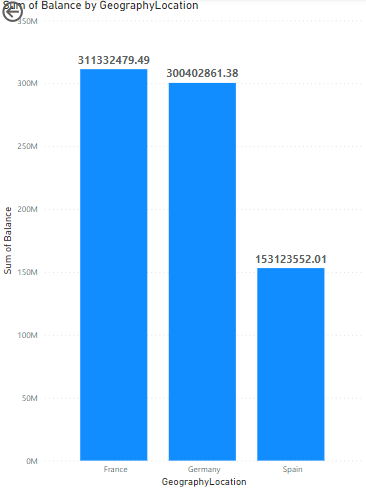
**Capstone Project:**

**Analytical CRM Development for a Bank**

Objective Questions:

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



According to given data we have following account balances in three different regions –

* France – 311332479.49
* Germany – 300402861.38
* Spain – 153123552.01

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

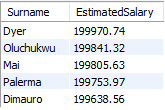
select Surname , EstimatedSalary from customerinfo

where quarter(BankDOJ) = 4

order by EstimatedSalary desc

limit 5

after applying the above formula we will get the following result –



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

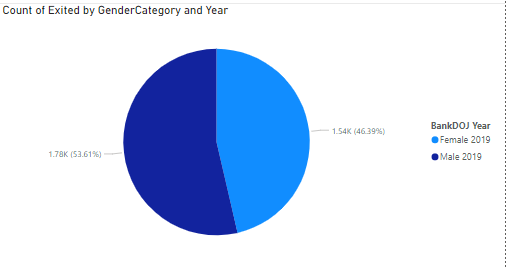
select avg(NumOfProducts) from bank\_churn

where HasCrCard = 1

after applying the above formula we will get the following result –



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



In the most recent year which is 2019 , 1776 males and 1537 females have left the bank

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

**select e.ExitCategory , avg(bc.CreditScore) as Avg\_cred\_sco**

**from bank\_churn bc**

**join exitcustomer e on e.ExitID = bc.Exited**

**group by e.ExitCategory**

after applying the above formula we will get the following result –



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

**select g.GenderCategory , avg(ci.EstimatedSalary) from customerinfo ci**

**join gender g on g.GenderID = ci.GenderID**

**join bank\_churn bc on bc.CustomerID = ci.CustomerID**

**where bc.Exited = 0**

**group by g.GenderCategory**

after applying the above formula we will get the following result –



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

**with custseg as (**

**select CustomerID , CreditScore ,**

**case**

**when CreditScore < 580 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'**

**else 'Excellent'**

**end as Credit\_Segment**

**from bank\_churn**

**where Exited = 1**

**)**

**select Credit\_Segment , count(\*) as totalexitcustomer from custseg**

**group by Credit\_Segment**

**order by totalexitcustomer desc**

**limit 1**

after applying the above formula we will get the following result –



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

**select g.GeographyLocation , count(bc.CustomerID) as activecustomers**

**from geography g**

**join customerinfo ci on ci.GeographyID = g.GeographyID**

**join bank\_churn bc on bc.CustomerID = ci.CustomerID**

**where bc.IsActiveMember = 1**

**and bc.Tenure>5**

**group by g.GeographyLocation**

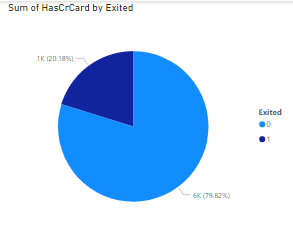
**order by activecustomers desc**

**limit 1**

after applying the above formula we will get the following result –



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



From the people having credit card, around 80% of them are retained by the bank.

Customer having credit card who left the bank – 1424

Customer having credit card who are with the bank – 5631

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

Around 65% of customers who left the bank had used only one product.

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.

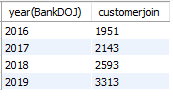
This SQL program will give the data of new customer yearly-

**select year(BankDOJ) , count(\*) as customerjoin from customerinfo**

**group by year(BankDOJ)**

**order by year(BankDOJ)**

after applying the above formula we will get the following result –



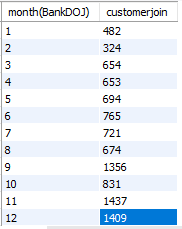
This SQL program will give the data of new customer monthly –

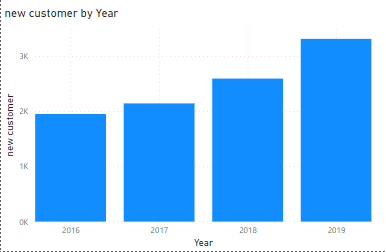
**select month(BankDOJ) , count(\*) as customerjoin from customerinfo**

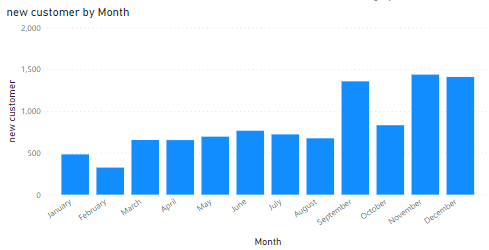
**group by month(BankDOJ)**

**order by month(BankDOJ)**

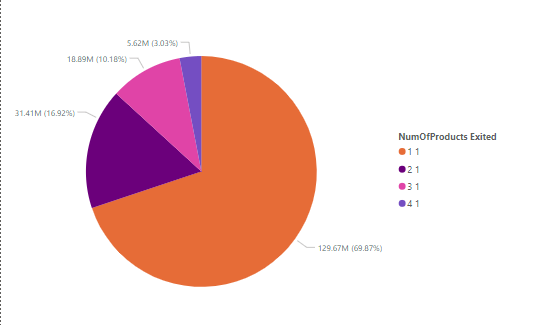
after applying the above formula we will get the following result –







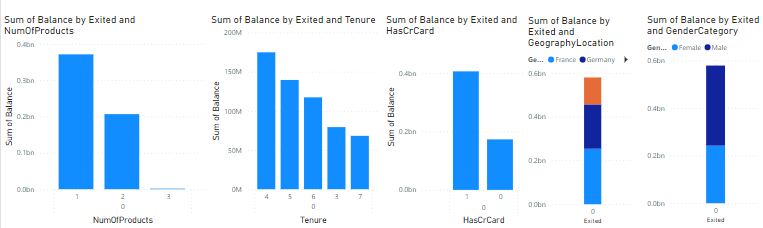
1. Analyze the relationship between the number of products and the account balance for customers who have exited.



The customers who left the bank have following balance according to the number of products used-

* 1. 129.67 million (approx. 70%)
  2. 31.41 million (approx. 17%)
  3. 18.89 million (approx. 10%)
  4. 5.62 million (approx. 3%)

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



If we will notice the above graphs we will notice that people who are remained with bank have following outliers –

* They use mostly 1 product
* Most of the customers are those who have completed 4 years with bank
* Most of the customers have credit card
* Many of the customers are from France
* More than half of the customers are male

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

total table - 7

categorical variable table – 5

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)

**select g.GeographyLocation , ge.GenderCategory , avg(ci.EstimatedSalary) as avg\_salary ,**

**rank() over (partition by GeographyLocation order by avg(ci.EstimatedSalary) desc) as ranking**

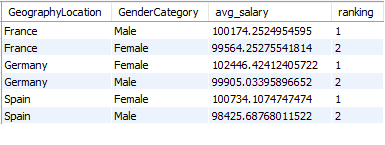
**from customerinfo ci**

**join geography g on g.GeographyID = ci.GeographyID**

**join gender ge on ge.GenderID = ci.GenderID**

**group by g.GeographyLocation , ge.GenderCategory**

after applying the above formula we will get the following result –



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+).

**select**

**case**

**when ci.Age >= 18 and ci.Age <= 30 then "18-30"**

**when ci.Age > 30 and ci.Age <= 50 then "30-50"**

**else "50+"**

**end as age\_bracket ,**

**avg(bc.Tenure) as avgtenure\_exitedcustomer**

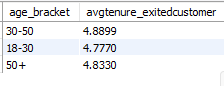
**from bank\_churn bc**

**join customerinfo ci on ci.CustomerID = bc.CustomerID**

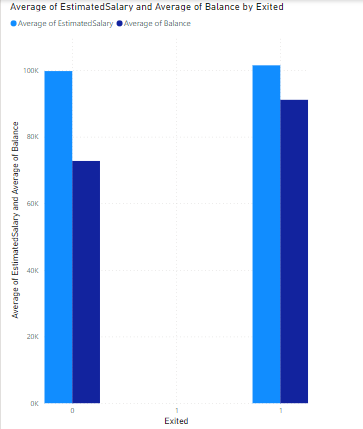
**where bc.Exited = 1**

**group by age\_bracket**

after applying the above formula we will get the following result –

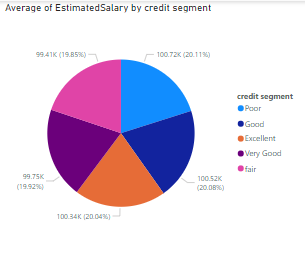


1. Is there any direct correlation between salary and the balance of the customers? And is it different for people who have exited or not?



If we observe the above graph we will find that people who have exited have more balance in their account compared to the people who remain with the bank

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



Observing the data, it don’t seems any relation between salary & credit score

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

**with custseg as (**

**select CustomerID , CreditScore ,**

**case**

**when CreditScore < 580 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'**

**else 'Excellent'**

**end as Credit\_Segment**

**from bank\_churn**

**where Exited = 1**

**)**

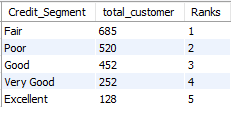
**select Credit\_Segment , count(CustomerID) as total\_customer , rank() over (order by count(CustomerID) desc) as Ranks**

**from custseg**

**group by Credit\_Segment**

**order by Ranks**

after applying the above formula we will get the following result –



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.

**select**

**case**

**when ci.Age >= 18 and ci.Age <= 30 then "18-30"**

**when ci.Age > 30 and ci.Age <= 50 then "30-50"**

**else "50"**

**end as age\_bracket ,**

**count(ci.CustomerID) as customer\_with\_cc**

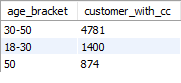
**from customerinfo ci**

**join bank\_churn bc on bc.CustomerID = ci.CustomerID**

**where HasCrCard = 1**

**group by age\_bracket**

after applying the above formula we will get the following result –



**with cte as (**

**select**

**case**

**when ci.Age >= 18 and ci.Age <= 30 then "18-30"**

**when ci.Age > 30 and ci.Age <= 50 then "30-50"**

**else "50"**

**end as age\_bracket ,**

**count(ci.CustomerID) as customer\_with\_cc**

**from customerinfo ci**

**join bank\_churn bc on bc.CustomerID = ci.CustomerID**

**where HasCrCard = 1**

**group by age\_bracket**

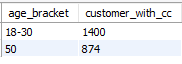
**)**

**select age\_bracket , customer\_with\_cc**

**from cte**

**where customer\_with\_cc < (select avg(customer\_with\_cc) from cte)**

after applying the above formula we will get the following result –



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

**select g.GeographyLocation , count(ci.CustomerID) as total\_customer ,**

**rank() over (order by count(ci.CustomerID) desc) as Ranks ,**

**avg(bc.Balance) as avg\_balance**

**from geography g**

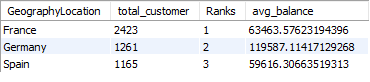
**join customerinfo ci on ci.GeographyID = g.GeographyID**

**join bank\_churn bc on bc.CustomerID = ci.CustomerID**

**where bc.IsActiveMember = 0**

**group by g.GeographyLocation**

after applying the above formula we will get the following result –



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”.

**ALTER TABLE customerinfo**

**ADD COLUMN CustomerID\_Surname VARCHAR(255);**

**UPDATE customerinfo**

**SET CustomerID\_Surname = CONCAT(CustomerId,'\_',Surname);**

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

**select \* ,**

**case**

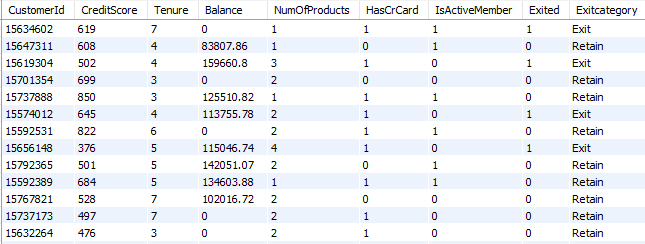
**when Exited = 1 then 'Exit'**

**else 'Retain'**

**end as Exitcategory**

**from bank\_churn**

after applying the above formula we will get the following result –

****

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

The missing values can be handled by using ISNULL function

Or by using the update and set function.

Suppose, we have a null value in the balance column of bank\_churn table; so to handle that we will use following fuctiion –

UPDATE bank\_churn

SET balance = (SELECT AVG(balance) FROM bank\_churn WHERE balance IS NOT NULL)

WHERE balance IS NULL

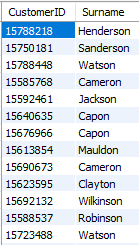
1. Write the query to get the customer IDs, their last name, and whether they are active or not for the customers whose surname ends with “on”.

**select CustomerID , Surname**

**from customerinfo**

**where Surname like "%on"**

after applying the above formula we will get the following result –

****

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.

According to given information, there are many data which shows customer who exited from the bank are active ,which can’t be correct.

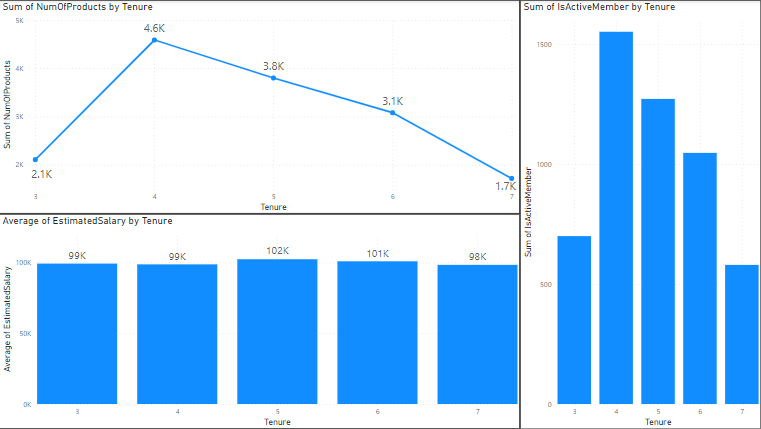
Those customer data could be seen by running below code -

**select \* from bank\_churn**

**where IsActiveMember = 1 and Exited = 1**

**Subjective Question:**

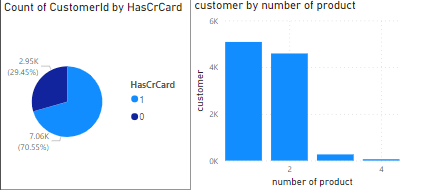
1. Customer Behavior Analysis: What patterns can be observed in the spending habits of long-term customers compared to new customers, and what might these patterns suggest about customer loyalty?

****

From the above data it can be observed that old customers spend less on buying products and are less active as compared to those customers who have completed around 4 years with the bank.

From the pattern it can said that as the customers spends the time in bank, they become less active and spend less on buying the product.

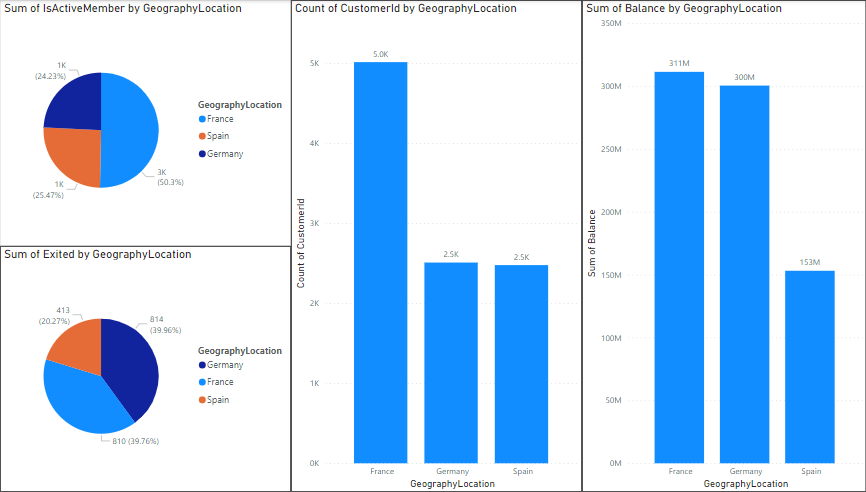
1. Product Affinity Study: Which bank products or services are most commonly used together, and how might this influence cross-selling strategies?



The credit card facility of the bank is most common service of the bank used by over 70% of the customers.

Most of the customers buy mostly one or two products from the bank.

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

****

According to the geographic trends, the number of customers from France(5014) is double than the customers from Germany(2509) and Spain(2477). But the balance of account is almost equal for France(311 Million) and Germany(300 Million) and half of that is of Spain(153 Million).

It can be seen that out of the total customers of different countries following of them are active –

1. France – 2591 out of 5014
2. Germany – 1248 out of 2509
3. Spain – 1312 out of 2477

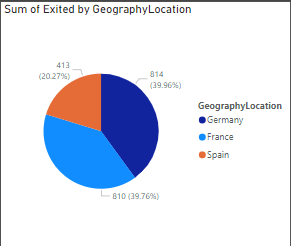
Out of the total active members France has 50% of the total active member then Spain 26% then Germany.24%

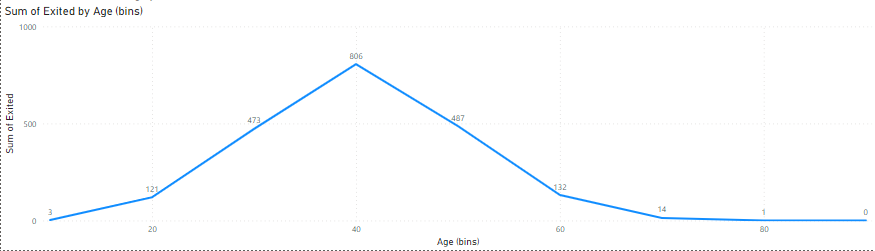
It can also be seen that out of the total customers of different countries following of them have exited the bank –

1. France – 810 out of 5014
2. Germany – 814 out of 2509
3. Spain – 413 out of 2477

Out of the total customers who left the bank France has 40% of the total exited customers then Germany 40% then Spain 20%.

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

****

****

The customers who are of the age between 40-50 are leaving the bank at very high rate which can turn into financial loss to the bank.

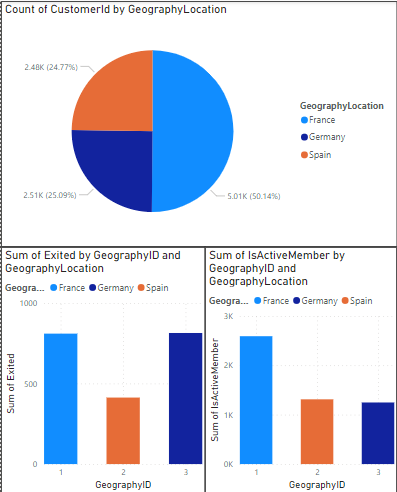
Also people from Germany and France are leaving the bank which could turn into financial loss.

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?

****

On observing the above data we can predict that most of the customers in every segment have an average tenure of 5 years.

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?

****

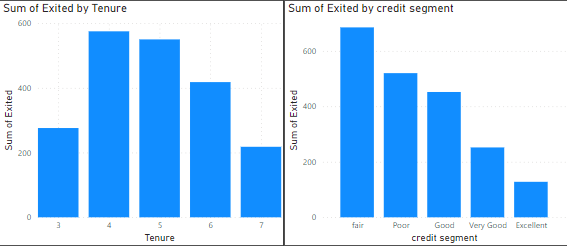
The bank should focus on area where the exit rate is high, from the above graph it can be seen that France & Germany has high exit and also bank should focus on the areas which have less active members like Germany & Spain.

Through the marketing campaign bank should promote its scheme and motivate the people of particular areas to join the bank.

To make the marketing campaign more effective, following points may help-

* If we could know the bank locations whether in remote area or in main city.
* Number of bank available in any area.
* What facilities people aspect from the bank

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

****

From the graphs, it can be seen that customers after completing 4-5 years start leaving the bank.

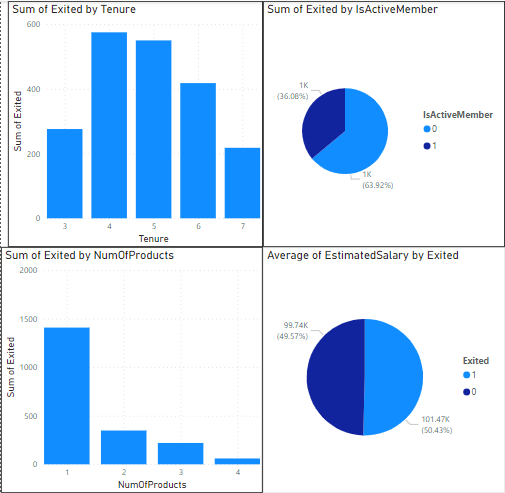
Also, customers with low credit score leave the bank.

Reasons for leaving-

May be people are coming to a location for job and after completing 4-5 years they get transferred.

May be people are not liking the services provided by the bank like high credit interests, low returns etc.

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

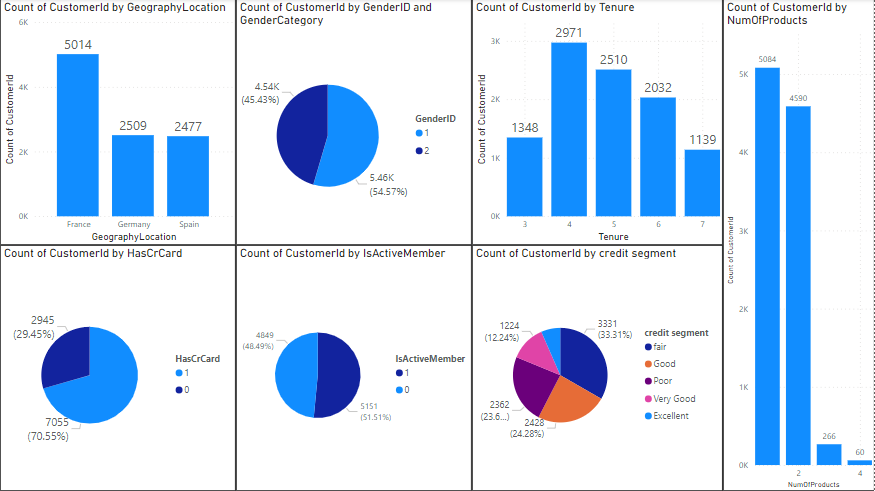
****

Yes, these are important factors to predict the future.

Some of the outputs from those factors are –

* Customers after completing 4-5 years with bank are leaving the bank
* Customers before leaving the bank are inactive
* Customers who bought only one product have high chance of leaving the bank
* Estimated salary don’t have much impact on prediction

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

****

After running the SQL queries we will get following details-

**select g.GeographyLocation , count(ci.CustomerId) as total\_customer**

**from customerinfo ci**

**join geography g on g.GeographyID = ci.GeographyID**

**group by g.GeographyLocation**

Total customers from different regions-

1. France – 5014
2. Germany – 2509
3. Spain – 2477

**select g.GenderCategory , count(ci.CustomerId) as total\_customer**

**from customerinfo ci**

**join gender g on g.GenderID = ci.GenderID**

**group by g.GenderCategory**

Customers by gender –

* Male – 5457
* Female – 4543

**select Tenure , count(Exited) from bank\_churn**

**group by Tenure**

**order by Tenure**

Churn rate by customer tenure –

* 3 years – 1348
* 4 years – 2971
* 5 years – 2510
* 6 years – 2032
* 7 years – 1139

**select NumOfProducts , count(CustomerId) from bank\_churn**

**group by NumOfProducts**

**order by NumOfProducts**

Customers buying the number of product from bank –

* One product – 5084
* Two product – 4590
* Three product - 266
* Four product – 60

**select count(\*) from bank\_churn**

**where HasCrCard = 1**

Customers who have credit card – 7055 customers

**select count(\*) from bank\_churn**

**where IsActiveMember = 1**

Customers who are active – 5151

**select**

**case**

**when CreditScore < 580 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'**

**else 'Excellent'**

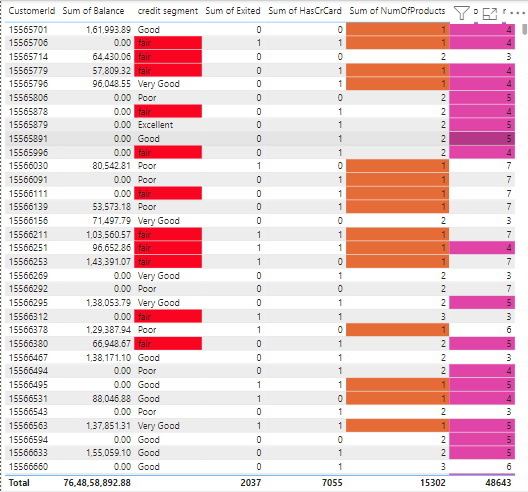
**end as Credit\_Segment , count(CustomerId)**

**from bank\_churn**

**group by Credit\_Segment**

Count of customers based on their credit score –

1. Poor (<580) - 2362
2. Fair (580-670) - 3331
3. Good (670-740) - 2428
4. Very good (740-800) - 1224
5. Excellent (>800) - 655
6. 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?

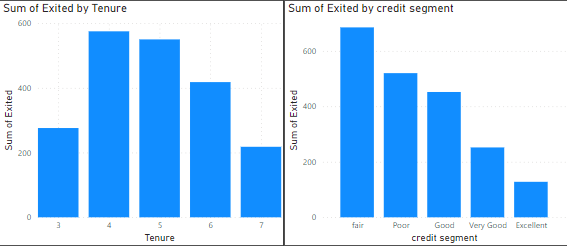
****

After applying the conditional formatting, the highlighted parts are at risk of churn.

If we start giving rewards on credit card it will benefit the 7055 customers which may help in decrease in churn rate.

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?

Current churn rate is 20.37%.

****

From the graphs, it can be seen that customers after completing 4-5 years start leaving the bank.

Also, customers with low credit score leave the bank.

Reasons for leaving-

May be people are coming to a location for job and after completing 4-5 years they get transferred.

May be people are not liking the services provided by the bank like high credit interests, low returns etc.

Remedies to decrease the churn rate –

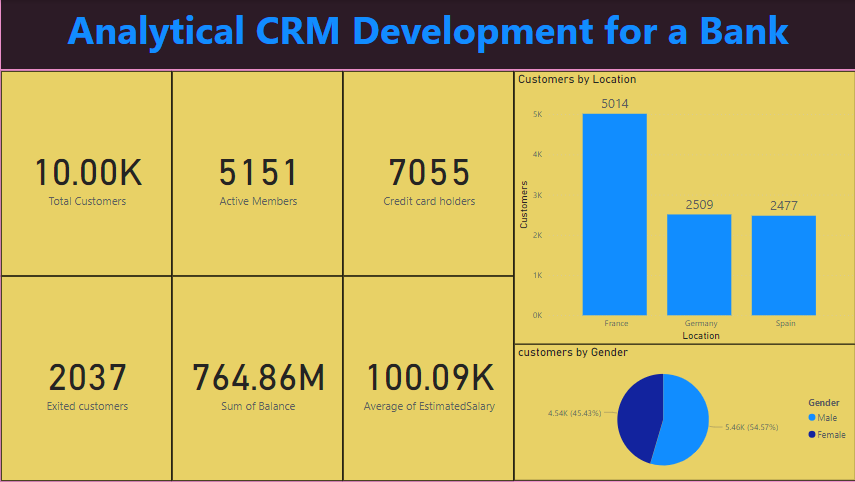
Personalized offers and discount

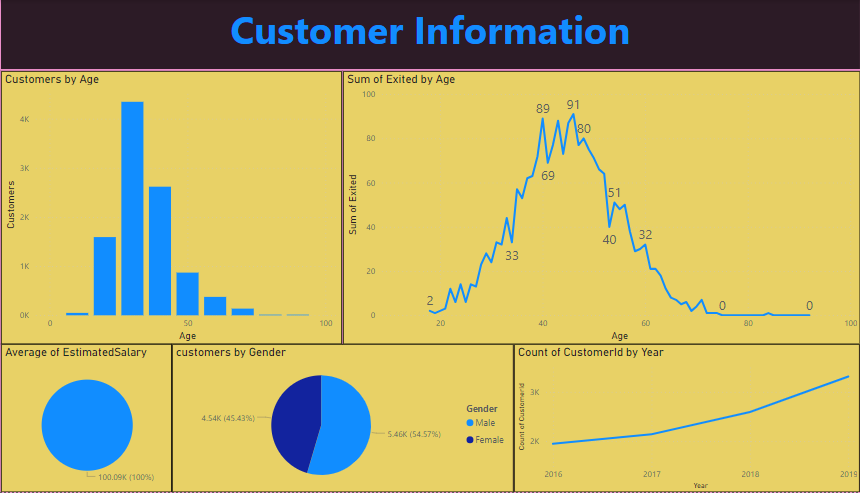
Enhanced customer service and support

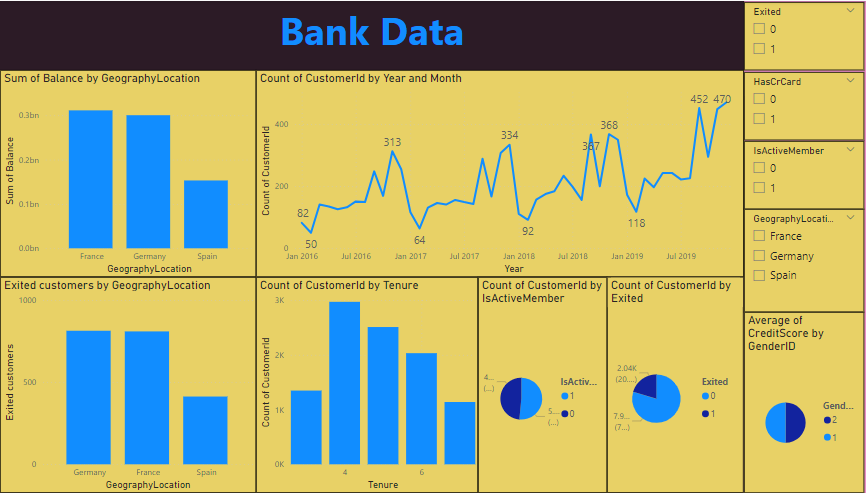
Loyalty programs and rewards

Proactive engagement for at-risk customers

1. 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?

On observing the data provided by the bank, we can find that the main concern of the bank is the churn of the customers. To reduce the churn rate we need to first understand the data, here are the following steps to approach the problem –

* **Data Understanding and Collection –** Collect relevant customer data: demographics, account information, credit score, balance, age, tenure, product usage, and churn status
* **Data Analysis –** summarising data , creating charts and tables
* **Model building –** finding the relations between the data and understanding the impact of each on other.
* **Interpretation and Insights –**

Identify significant features contributing to churn.

Segment customers based on churn

* **Actionable Recommendations –**

Develop targeted retention strategies:

Personalized offers and discount

Enhanced customer service and support

Loyalty programs and rewards

Proactive engagement for at-risk customers

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

ALTER TABLE bank\_churn

RENAME COLUMN HasCrCard TO Has\_Creditcard