**Objective Questions**

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

**Ans ->**

select ci.GeographyID, g.GeographyLocation, round(sum(bc.Balance),2) as Balance

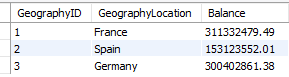
from customerinfo ci

join bank\_churn bc ON ci.CustomerId = bc.CustomerId

join geography g ON ci.GeographyID= g.GeographyID

group by 1,2

ORDER BY ci.GeographyID;



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

**Ans ->**

**Here we will find top5 customers with the highest salary**

select CustomerId, Surname, EstimatedSalary from customerinfo

where year(bankDOJ)= 2019 and 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)**

**Ans ->**

select avg(NumOfProducts) as avg\_product\_cc from bank\_churn where HasCrCard= 1;



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

**Ans ->**

select g.GenderCategory,

cast(count(case when exited= 1 then b.CustomerId end)\*100/ count(b.CustomerId) as decimal(10,2))

as churn\_rate

from bank\_churn b join customerinfo c on b.CustomerId= c.CustomerId

join gender g ON g.GenderID= c.GenderID

where year(bankDOJ)= 2019

group by 1;



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

**Ans ->**

select

Avg(Case when exited= 1 then creditscore end) as avg\_credit\_exited,

Avg(case when exited= 0 then CreditScore end) as avg\_credit\_remain

from bank\_churn;



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

**Ans ->**

select g.GenderCategory,round(avg(c.EstimatedSalary),2) as avg\_salary,

round(avg( case when a.ActiveID=1 then c.EstimatedSalary end),2) as avg\_salary\_active,

round(avg( case when a.ActiveID=0 then c.EstimatedSalary end),2) as avg\_salary\_inactive

from customerinfo c

inner join gender g

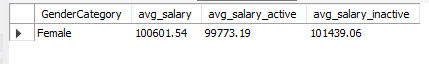
ON c.genderid= g.genderid

inner join bank\_churn b ON b.CustomerId= c.CustomerId

inner join activecustomer a ON b.IsActiveMember= a.ActiveID

group by g.GenderCategory

Order by avg\_salary desc limit 1 ;



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

**Ans ->**

select CreditScore, count(CustomerId) customer\_count from bank\_churn

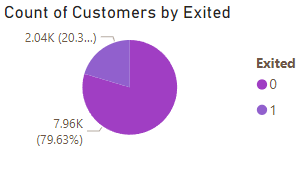
where exited= 1

group by CreditScore

order by customer\_count desc limit 1;



There are 7963 customers have retained and 2037 customers have exited. And this graph has showing the proper segmentation as the customers having credit score of 850 has the highest exit rate as 43 customers.



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

**Ans ->**

select g.GeographyLocation,count(case when a.ActiveCategory= 'Active Member' then 1 end )

as count\_active\_member from customerinfo c

inner join bank\_churn b ON c.CustomerId= b.CustomerId

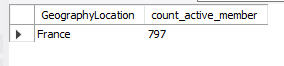
inner join geography g ON c.GeographyID= g.GeographyID

inner join activecustomer a ON b.IsActiveMember= a.ActiveID

where b.Tenure>5

group by g.GeographyLocation

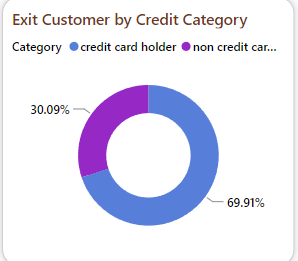
order by count\_active\_member desc limit 1;



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

**Ans ->**

From data exit customers are higher in customers having credit card (69.91%) rather than not having credit card (30.09%).

****

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

**Ans ->**

select NumOfProducts, count(NumOfProducts) as total\_count

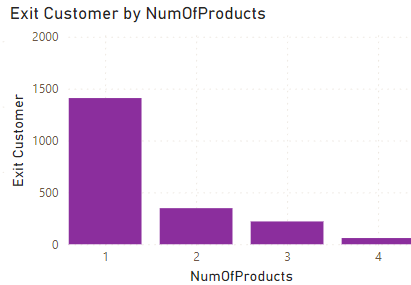
from bank\_churn

where exited= 1

group by NumOfProducts

order by total\_count desc limit 1;





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.**

**Ans ->**

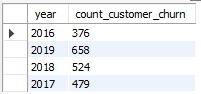
select year(bankDOJ) as year, count(c.CustomerId) as count\_customer\_churn

from bank\_churn b

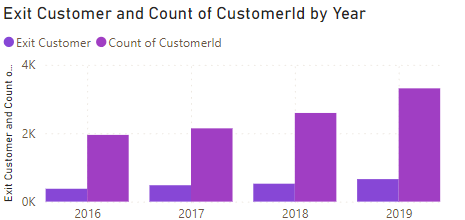
inner join customerinfo c ON b.CustomerId= c.CustomerId

where Exited= 1

group by year(bankdoj);

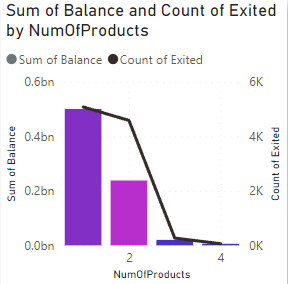


As per showing in the table in year and number of customers in four years.



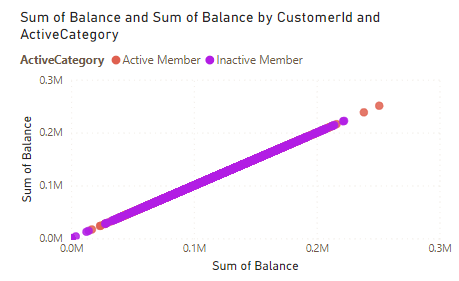
1. **Analyze the relationship between the number of products and the account balance for customers who have exited.**

**Ans ->** From the data the customers with less number of products have high exit rate.



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

**Ans ->** In the below scatter plot there are four outliers who have remained with bank.

****

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

**Ans ->**

**There are 7 different tables in dataset. And among them 2 tables (Bank Churn & Customer Info) consists of categorical variables.**

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)**

**Ans ->**

with temp as

(

select c.GeographyID,g.GenderCategory ,

round(AVG(c.EstimatedSalary),2) as avg\_salary

from customerinfo c

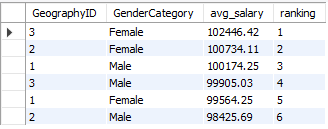
inner join gender g ON c.GenderID= g.GenderID

group by c.GeographyID,g.GenderCategory

)

select \*, rank() over(order by avg\_salary desc)

as ranking from temp ;



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+).**

**Ans ->**

with AgeBucket as

(

select c.CustomerId,c.surname,c.age,c.GenderID,

c.EstimatedSalary,c.GeographyID,c.bankDOJ,

b.CreditScore,b.tenure,b.balance,b.NumOfProducts,

b.HasCrCard,b.IsActiveMember,b.Exited,

case when c.age between 18 and 30 then '18-30'

when c.age between 31 and 50 then '30-50'

else '50+'

end as age\_bracket

from bank\_churn b

inner join customerinfo c ON b.CustomerId= c.CustomerId

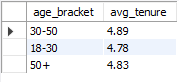
where exited=1

)

select age\_bracket, round(avg(tenure),2) avg\_tenure from AgeBucket

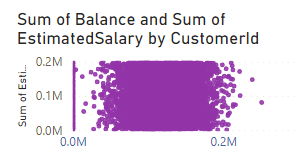
group by age\_bracket

;



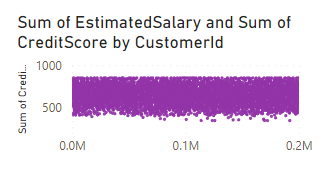
1. **Is there any direct correlation between salary and the balance of the customers? And is it different for people who have exited or not?**

**Ans -> From the following scatter plot we can know the correlation between salary and balance of the customers.**

****

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

**Ans ->** **From following scatter plot it is clear that there is no correlation between salary and credit score**

****

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

**Ans ->**

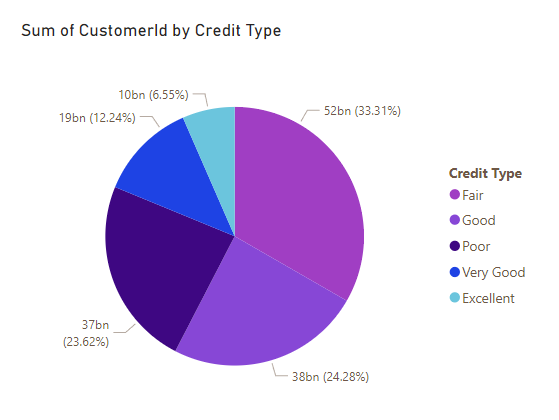
As per below graph 33.31% customers has fair credit score.

24.28% customers Good fair credit score.

23.62% customers has Poor credit score.

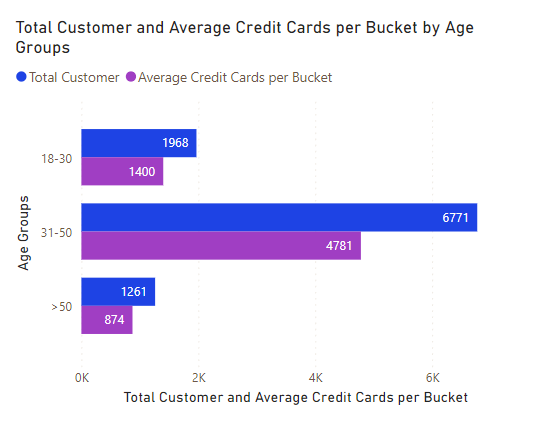
12.24% customers has Very Good credit score.

6.55% customers has Excellent credit score.

****

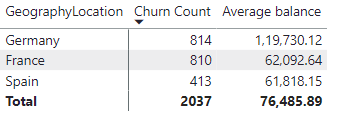
1. **According to the age buckets find the number of customers who have a credit card. Also retrieve those buckets that have lesser than average number of credit cards per bucket.**

**Ans ->** According to the age buckets the number of customers who have a credit card. And buckets those have lesser than number of credit cards per bucket. For age bucket **18-30** total customers are **1968** average number of credit cards per bucket is **1400**. For age bucket **31-50** total customers are **6771** average number of credit cards per bucket is **4781**. For age bucket 50 and above total customers are 1261 average number of credit cards per bucket is **874**.

****

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

**Ans ->** As per following table Germany is at 1st rank. 2nd one is France and last is 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”.**

**Ans ->** select

ci.CustomerId,

ci.Surname,

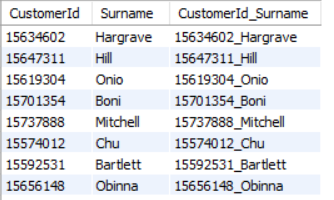
concat(ci.CustomerId,'\_',ci.Surname) as CustomerId\_Surname

from

customerinfo ci

join

bank\_churn ot on ci.CustomerId = ot.CustomerId;



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

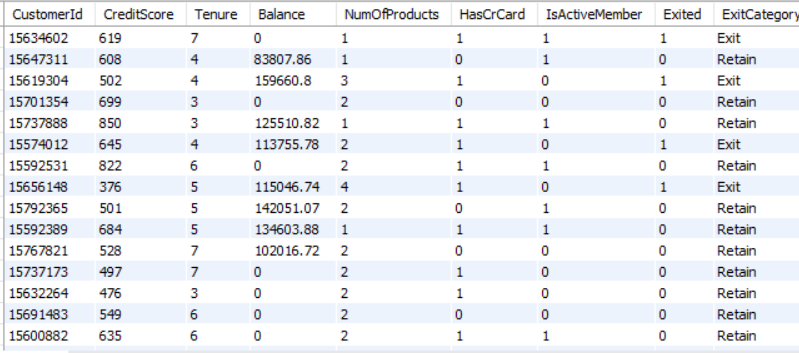
**Ans ->** SELECT

bc.\*,

(SELECT 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?**

**Ans ->**No there are no missing values but there are some null rows in the data. By using Power Query editor I handled them. And after that when I am getting the data to SQL then the data is getting corrupt so I made CSV file of that data (Separately) then I imported the data into SQL and then I made the connections.

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”.**

**Ans ->** SELECT

bc.CustomerId,

ci.Surname,

CASE WHEN bc.IsActiveMember = 1 THEN 'Active' ELSE 'Inactive' END AS ActiveStatus

FROM

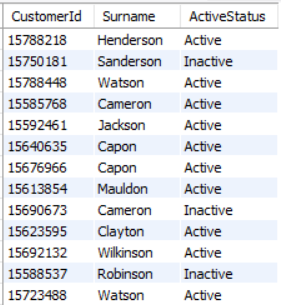
bank\_churn bc

JOIN

customerinfo ci ON bc.CustomerId = ci.CustomerId

WHERE

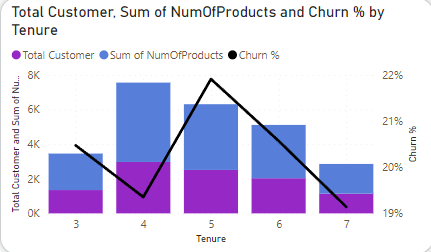
ci.Surname LIKE '%on';

****

**Data Analysis and Visualizations (Subjective Questions):**

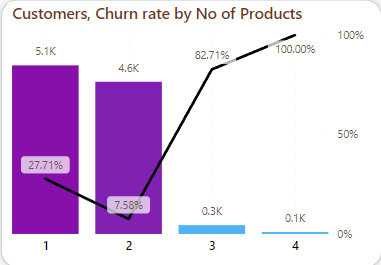
1. **Customer Behavior Analysis: What patterns can be observed in the spending habits of long-term customers compared to new customers, and what might these patterns suggest about customer loyalty?**

**Ans ->** From data it is Observed that Customers with Tenure 4 has spend more and customers with tenure 5 has higher churn rate and loyalty of customers is proportional to tenure, so with higher tenure customers are more loyal.

****

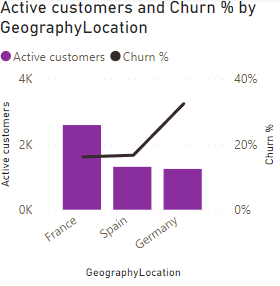
1. **Product Affinity Study: Which bank products or services are most commonly used together, and how might this influence cross-selling strategies?**

**Ans ->** From data it is been observed that Customers with less no of products has less churn rate compare to the customers who bought more no of products.

****

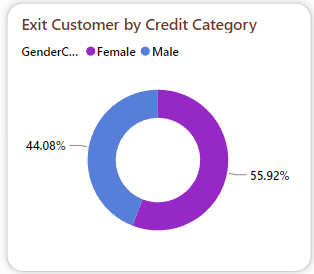
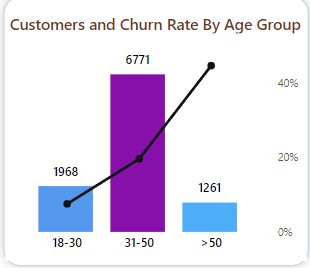
1. **Geographic Market Trends: How do economic indicators in different geographic regions correlate with the number of active accounts and customer churn rates?**

**Ans ->** From graph it is clear that France has more active customers and churn rate is max in Germany.

****

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

**Ans ->** From data it is clear that customer churn rate is proportional to the customer age. Female has higher churn rate compared to male.

****

1. **Customer Lifetime Value Forecast: How would you use the available data to model and predict the lifetime (tenure) value of different customer segments?**

**Ans ->** - Provide tailored offers and incentives based on customer preferences and behavior to encourage retention

- Implement loyalty programs that reward customers for their continued business and encourage them to stay with

the bank

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?**

**Ans ->**

Enhance the customer service experience service experiences by providing and personalized assistance, resolving issues efficiently, and addressing customer feedback

Marketing team should give a good offers and more security to customers age above 50, Also should offer more to customers who buys more no of products. They should also give more offers to credit card holders.

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

**Ans ->** Customers with age above 50 are not satisfied with bank terminology or schemes. Customers having credit score below 700 are also leaving. Germany and France are having high churn rates.

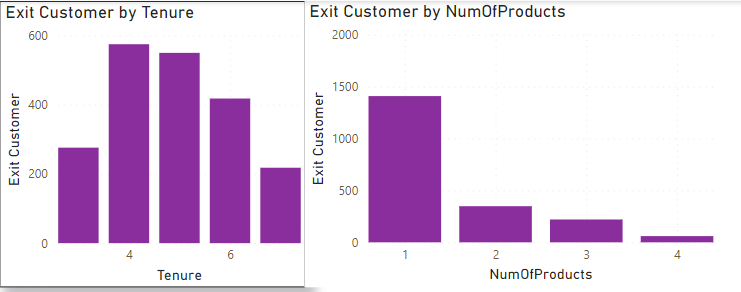
Low levels of customer satisfaction can lead to churn.

Dissatisfaction with customer Service, long wait times, and Unresolved issues are common reasons why customers churn banks

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

**Ans ->** From given dataset Customers with tenure 4 and 5 are more likely to exit. Customers with less no of products are

more likely to exit and we can not predict it from estimated salary.

****

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

**Ans ->** with AgeBucket as

(

select c.CustomerId,c.surname,c.age,c.GenderID,

c.EstimatedSalary,c.GeographyID,c.bankDOJ,

b.CreditScore,b.tenure,b.balance,b.NumOfProducts,

b.HasCrCard,b.IsActiveMember,b.Exited,

case when c.age between 18 and 30 then '18-30'

when c.age between 31 and 50 then '30-50'

else '50+'

end as age\_bracket

from bank\_churn b

inner join customerinfo c ON b.CustomerId= c.CustomerId

)

select age\_bracket, count(customerId) total\_customers from AgeBucket

group by age\_bracket

;

;

By Credit Score- with creditbucket as

(

select \*,

case when creditscore between 0 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 800 then 'Very Good'

else 'Excellent'

end as creditBucket

from bank\_churn

where exited = 1

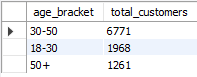
)

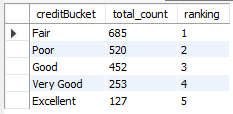
select creditbucket, count(customerid) as total\_count,

dense\_rank() over(order by count(customerid) desc) as ranking

from creditbucket

group by creditbucket





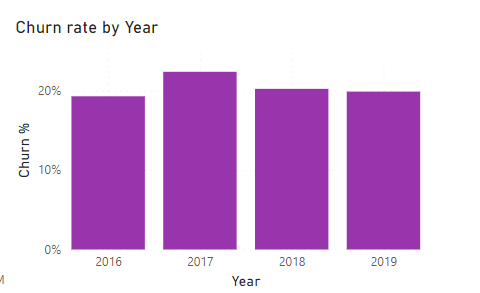
1. **How can we create a conditional formatting setup to visually highlight customers at risk of churn and to evaluate the impact of credit card rewards on customer retention?**

**Ans ->** We can create conditional formatting to visually highlight customers at risk on basis of age condition having age >50, and no of products <2 also customers with credit score < 700.

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 are the different strategies that can be used to decrease the churn rate.**

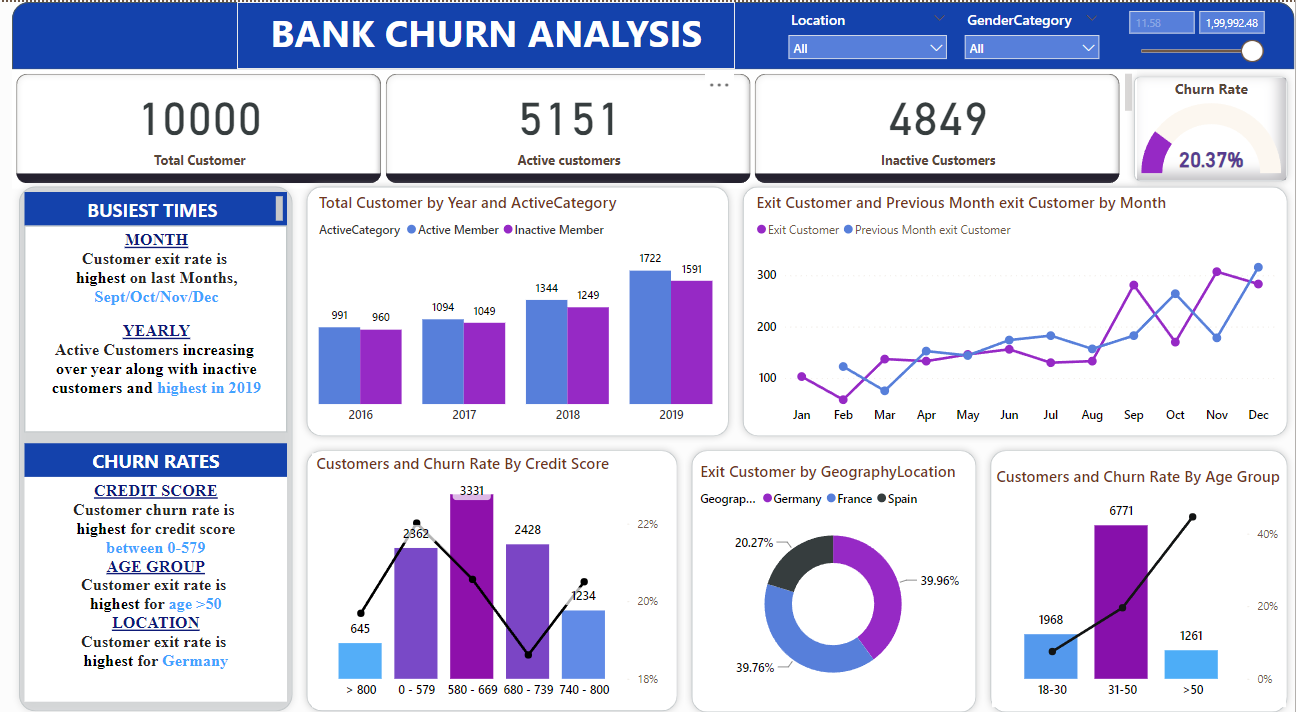
**Ans ->** Overall churn rate is 20.47 % . In year 2016- 19.27%, 2017- 22.35%, 2018- 20.21% and in 2019- 19.86%.

Customers above age 50 and customers having less no of product are more likely to churn. So if bank provides some good rewards, Offers and securities to customers from this category they might stay with bank for long period.

****

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

**Ans ->**

****

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

**Ans ->**As per my opinion if the objective and subjective questions weren’t given then the approaching the problem will be get quit harder than it looks. Because of those questions the SQL queries and the visualisations got easy to solve.

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

**Ans ->**

ALTER TABLE Bank\_Churn

RENAME COLUMN HasCrCard TO Has\_creditcard;