**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

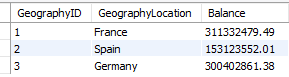
from customerinfo ci

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

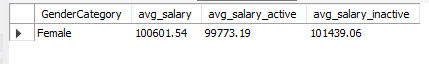
ON c.genderid= g.genderid

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;



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

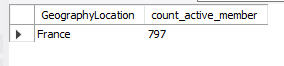
inner join geography g ON c.GeographyID= g.GeographyID

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

where b.Tenure>5

group by g.GeographyLocation

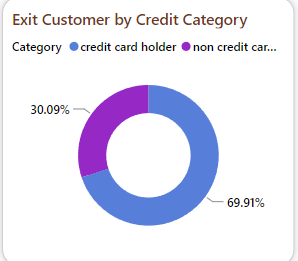
order by count\_active\_member desc limit 1;



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 ->**

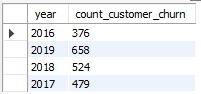
select year(bankDOJ) as year, count(c.CustomerId) as count\_customer\_churn

from bank\_churn b

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

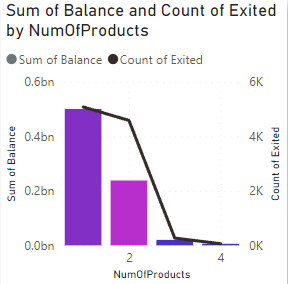
where Exited= 1

group by year(bankdoj);



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 ->**

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

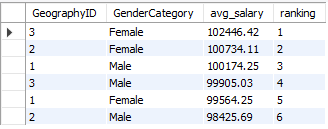
inner join gender g ON c.GenderID= g.GenderID

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

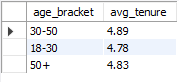
where exited=1

)

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

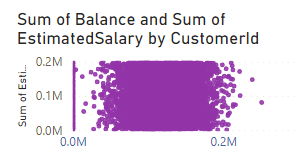
group by age\_bracket

;



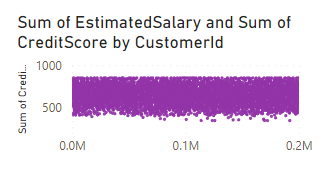
1. **Is there any direct correlation between salary and the balance of the customers? And is it different for people who have exited or not?**

**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 ->**

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

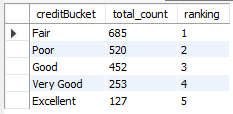
)

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

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

from creditbucket

group by creditbucket ;



1. **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 ->**

create view ageBucket1 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

);

with cte1 as

(select age\_bracket, count(customerid) total\_customer,

count(case when hascrcard=1 then customerid end) as count\_customer\_with\_credit

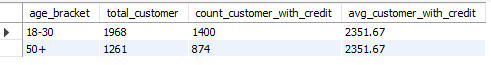
from agebucket1

group by 1)

select \*, round((select avg(count\_customer\_with\_credit) from cte1),2) as

avg\_customer\_with\_credit from cte1

having count\_customer\_with\_credit < (select avg(count\_customer\_with\_credit) from cte1);



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

**Ans ->**

with cte as(

select g.GeographyLocation, count(distinct b.CustomerId) as count\_churn

from bank\_churn b

join customerinfo c on b.CustomerId= c.CustomerId

join geography g ON c.GeographyID= g.GeographyID

where b.exited= 1

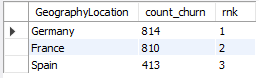
group by 1)

select \*, rank() over(order by count\_churn desc) as rnk from cte

;

**Average balance**

select avg(balance) from bank\_churn;





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 ->**

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

**Ans ->**

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 ->**

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 ->**