SELECT \* FROM bank\_churn;

SELECT \* FROM customerinfo;

SELECT \* FROM activecustomer;

SELECT \* FROM creditcard;

SELECT \* FROM exitcustomer;

SELECT \* FROM gender;

SELECT \* FROM geography;

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

SELECT geo.GeographyLocation, ROUND(SUM(bc.Balance),2) AS balances

FROM bank\_churn bc

JOIN customerinfo ci USING (CustomerID)

JOIN geography geo USING (GeographyID)

GROUP BY geo.GeographyLocation ;

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

SELECT \*

FROM customerinfo

WHERE month(BankDOJ) IN (10,11,12)

ORDER BY EstimatedSalary DESC

LIMIT 5;

-- 3. Calculate the average number of products used by customers who have a credit card.

SELECT b.CustomerId, ROUND(AVG(NumOfProducts),0) As avg\_products

FROM bank\_churn b

LEFT JOIN customerinfo c ON b.CustomerID = c.CustomerID

WHERE HasCrCard = 1

GROUP BY b.CustomerId;

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

SELECT g.GenderCategory,MAX(YEAR(BankDOJ)) AS recent\_year, count(CustomerId) as no\_of\_customers

FROM customerinfo c

JOIN bank\_churn b USING (CustomerID)

JOIN gender g USING (genderID)

GROUP BY g.GenderCategory;

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

SELECT (CASE WHEN Exited = 1 THEN 'Exited' ELSE 'Remain'END ) AS exited\_remain,

AVG(CreditScore) AS avg\_creditscore

FROM bank\_churn

GROUP BY exited\_remain;

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

WITH active\_avg\_est\_salary AS

(SELECT g.GenderCategory AS gender, ROUND(AVG(c.EstimatedSalary),2) AS active\_avg\_est\_salary

FROM customerinfo c

JOIN bank\_churn b USING (CustomerID)

JOIN gender g USING (genderID)

WHERE IsActiveMember = 1

GROUP BY g.GenderCategory),

inactive\_avg\_est\_salary AS

(SELECT g.GenderCategory AS gender, ROUND(AVG(c.EstimatedSalary),2) AS inactive\_avg\_est\_salary

FROM customerinfo c

JOIN bank\_churn b USING (CustomerID)

JOIN gender g USING (genderID)

WHERE IsActiveMember = 0

GROUP BY g.GenderCategory)

SELECT \*

FROM active\_avg\_est\_salary a

JOIN inactive\_avg\_est\_salary i USING(gender);

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

SELECT CASE WHEN CreditScore < 600 THEN 'Poor(Less Than 600)'

WHEN CreditScore >= 600 AND CreditScore < 700 THEN 'Fair(Between 600 And 700)'

WHEN CreditScore >= 700 AND CreditScore < 800 THEN 'Good(Between 700 And 800)'

ELSE 'Excellent(More than 800)'

END AS segments, Count(Exited) As cnt\_exited

FROM bank\_churn

WHERE Exited = 1

GROUP BY segments

ORDER By cnt\_exited DESC;

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

SELECT g.GeographyLocation, Count(c.CustomerId) AS active\_customers

FROM customerinfo c

JOIN geography g USING(GeographyID)

JOIN bank\_churn b USING(CustomerID)

WHERE IsActiveMember = 1 AND Tenure >5

GROUP BY g.GeographyLocation;

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

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

SELECT NumOfProducts,count(CustomerId) AS no\_of\_customers

FROM bank\_churn

WHERE Exited = 1

GROUP BY NumOfProducts

ORDER BY no\_of\_customers DESC;

-- 11. 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.

SELECT YEAR(BankDOJ) AS join\_year,

MONTHNAME(BankDOJ) AS join\_month,

COUNT(CustomerID) AS Customers

FROM customerinfo

GROUP BY join\_year,join\_month

ORDER BY join\_year DESC,join\_month;

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

SELECT NumOfProducts,

Round(AVG(Balance),2) AS avg\_balance

FROM bank\_churn

WHERE Exited =1

GROUP BY NumOfProducts

Order BY NumOfProducts ASC;

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

-- 14. How many different tables are given in the dataset, out of these tables

-- which table only consists of categorical variables?

-- 15. 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)

-- Gender wise average income

-- geographical

-- rank gender avg values

WITH avg\_income AS

(SELECT gg.GeographyLocation,g.GenderCategory,ROUND(AVG(c.EstimatedSalary),2) AS average\_income

FROM customerinfo c

JOIN gender g USING (GenderID)

JOIN geography gg USING (GeographyID)

GROUP BY gg.GeographyLocation,g.GenderCategory

order by gg.GeographyLocation,g.GenderCategory)

SELECT \*,RANK() OVER(PARTITION BY GenderCategory ORDER BY average\_income DESC) AS rn

FROM avg\_income;

-- 16. 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+).

-- avg tenure

-- age bracket (18-30,30-50,50+)

SELECT CASE WHEN c.age BETWEEN 18 and 30 THEN '18-30'

WHEN c.age BETWEEN 30 AND 50 THEN '30-50'

ELSE '50+'

END AS age\_brackets,

AVG(bc.Tenure) AS avg\_tenure

FROM customerinfo c

JOIN bank\_churn bc USING (CustomerID)

WHERE bc.Exited = 1

GROUP BY age\_brackets

ORDER BY age\_brackets;

-- 17. Is there any direct correlation between salary and the balance of the customers?

-- And is it different for people who have exited or not?

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

-- 19. Identify any potential outliers in terms of spend among customers who have remained with the bank.

-- 20. How many different tables are given in the dataset,

-- out of these tables which table only consists of categorical variables?

-- 21. Using SQL, write a query to find out the gender-wise average income of male and females in each geography id.

-- Also, rank the gender according to the average value. (SQL)

-- avg income male female each geography id

WITH avg\_income\_location\_gender AS

(SELECT ci.GeographyID,

g.GenderCategory,

ROUND(AVG(ci.EstimatedSalary),2) As avg\_income

FROM customerinfo ci

JOIN gender g USING(GenderID)

GROUP BY ci.GeographyID,

g.GenderCategory

ORDER BY ci.GeographyID,

g.GenderCategory DESC)

SELECT \*, RANK() OVER(partition by GenderCategory ORDER BY avg\_income DESC) AS gender\_rank

FROM avg\_income\_location\_gender;

-- 22. 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+).

With exit\_cust\_tenure As

(Select c.CustomerId,

c.Surname,

datediff(curdate(),c.BankDOJ) AS tenure\_years,

CASE WHEN c.Age> 50 THEN "50+"

WHEN c.Age BETWEEN 30 AND 50 THEN "30-50"

ELSE "18-30" END AS age\_brackets

FROM customerinfo c

JOIN bank\_churn ch On c.CustomerId = ch.CustomerId

-- JOIN exitcustomer e ON e.ExitID = ch.Exited

WHERE ch.Exited = 1)

SELECT age\_brackets, Round(AVG(tenure\_years),2) as avg\_tenure

FROM exit\_cust\_tenure

GROUP BY age\_brackets;

-- 23. Is there any direct correlation between the salary and the balance of the customers?

-- And is it different for people who have exited or not?

-- 24. Is there any correlation between th e salary and Credit score of customers?

-- 25. 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 DISTINCT b.CustomerID,c.Surname As last\_name, a.ActiveCategory

FROM customerinfo c

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

JOIN activecustomer a ON b.IsActiveMember = a.ActiveID

WHERE c.Surname LIKE "%on"