ANALYSIS ON BANK CUSTOMER CHURN USING SQL

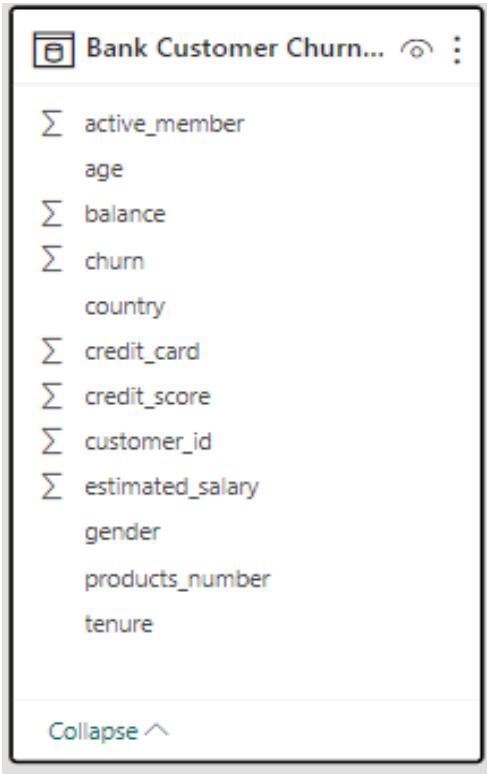
Introduction:

Customer churn prediction involves identifying which customers are likely to leave or unsubscribe from your service. Prediction plays a vital role in understanding customer behaviour. This is because acquiring new customers often costs more than retaining existing ones.

Customers have different behaviour and preferences, and reasons for cancelling their subscriptions. Therefore, it is important to actively communicate with each of them to keep them on your customer list. You need to know which marketing activities are most effective for individual customers and when they are most effective

About the data:

Bank churn customer prediction data is collected from "https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset". This data contains 10000 rows and 12 columns.



Objective

To determine which factor influence the customer churn.

Data Cleaning:

Before analysing the data, the important step is to cleaning the data. Data cleaning involves finding the missing value, detecting the outliers, removing the duplicate records etc.

```
CREATE DATABASE bank_churn;

use bank_churn;
SELECT * FROM bank_churn LIMIT 5;
SELECT COUNT(*) AS no_records FROM bank_churn;

describe bank_churn;
SELECT DISTINCT COUNT(customer_id) AS unique_id FROM bank_churn;
```

OUTPUT:

customer_id	credit_score	country	gender	age	tenure	balance	products_number	credit_card	active_member	estimated_salary	churn
15634602	619	France	Female	42	2	0	1	1	1	101348.88	1
15647311	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
15619304	502	France	Female	42	8	159660.8	3	1	0	113931.57	1
15701354	699	France	Female	39	1	0	2	0	0	93826.63	0
15737888	850	Spain	Female	43	2	125510.82	1	1	1	79084.1	0

no_records
10000

	Field	Type	Null	Key	Default	Extra
▶	customer_id	int	YES		NULL	
	credit_score	int	YES		NULL	
	country	text	YES		NULL	
	gender	text	YES		NULL	
	age	int	YES		NULL	
	tenure	int	YES		NULL	
	balance	double	YES		NULL	
	products_number	int	YES		NULL	
	credit_card	int	YES		NULL	
	active_member	int	YES		NULL	
	estimated_salary	double	YES		NULL	
	churn	varch...	YES		NULL	

Here, two columns contain categorical data, and the remaining columns are numerical. From this above result we observed that there are no duplicate records and there are no missing values. There total 10000 unique rows are present here.

```
select distinct country from bank_churn;
select distinct gender from bank_churn;
```

OUTPUT:

	country
▶	France
	Spain
	Germany

	gender
▶	Female
	Male

Here we can see that both two column had distinct class.

Outliers:

```
select "credit_score" as column_name ,avg(credit_score) as avarage,
max(credit_score) as maximum, min(credit_score) as minimum from bank_churn
union
select "age" as column_name,avg(age) as avarage,max(age) as maximum,
min(age) as minimum from bank_churn
union
select "tenure" as column_name,avg(tenure) as avarage,
max(tenure) as maximum, min(tenure) as minimum from bank_churn
union
select "products_numbers" as column_name,avg(products_number) as avarage,
max(products_number) as maximum, min(products_number) as minimum from bank_churn
union
select "credit_card" as column_name, avg(credit_card) as avarage,
max(credit_card) as maximum, min(credit_card) as minimum from bank_churn
union
select "active_member" as column_name, avg(active_member) as avarage,
max(active_member) as maximum, min(active_member) as minimum from bank_churn;
```

OUTPUT:

	column_name	avarage	maximum	minimum
▶	credit_score	650.5288	850	350
	age	38.9218	92	18
	tenure	5.0128	10	0
	products_numbers	1.5302	4	1
	credit_card	0.7055	1	0
	active_member	0.5151	1	0

Based on the results above, it appears that the average values in all fields are closely centered between the maximum and minimum values. This suggests a normal distribution of the data. Therefore, we can conclude that there are no outliers.

Data Preparation:

Here we are updating the columns, converting the 0 and 1 into 'No' and 'Yes'. We have changed the labels to make them more easily recognizable

```
ALTER TABLE bank_churn MODIFY churn VARCHAR(5) ;

UPDATE bank_churn SET churn =
CASE WHEN churn = 1 THEN 'yes'
ELSE 'no'
END;
```

```
alter table bank_churn  modify column credit_card varchar(5);
update bank_churn set credit_card =
case when credit_card=1 then "yes"
else "no"
end;
```

```
alter table bank_churn modify column active_member varchar(5);
update bank_churn set active_member=
case when active_member=1 then "yes"
else "no"
end;
```

Credit Score:

Based on the FICO score, credit scores falling within the range of 350 to 579 are categorized as 'poor,' scores within the range of 580 to 669 are categorized 'fair,' scores within the range of 670 to 739 are categorized 'good,' and scores within the range of 740 to 799 are categorized as 'very good.' Consequently, we have transformed all numerical values into these corresponding categorical labels

```sql
ALTER table bank_churn MODIFY credit_score varchar(10) ;

UPDATE bank_churn SET credit_score=
CASE WHEN credit_score BETWEEN 350 AND 579
THEN 'poor'
WHEN credit_score BETWEEN 580 AND 669
THEN 'fair'
WHEN credit_score BETWEEN 670 AND 739
THEN 'good'
WHEN credit_score BETWEEN 740 AND 799
THEN 'Very good'
ELSE
'excellent'
END;
```

Age:

We are updating the age values from numeric to categories, specifically 'teenager,' 'adults,' 'senior citizen,' and 'old senior citizen.

```sql
ALTER TABLE bank_churn MODIFY age VARCHAR(20);

UPDATE bank_churn SET age=
CASE
WHEN age BETWEEN 18 AND 35
THEN 'teenager'
WHEN age BETWEEN 36 AND 50
THEN 'adults'
WHEN age BETWEEN 51 AND 65
THEN 'senior citizen'
ELSE
'old senior citizen'
END;
```

Data Exploration:

1. Let us know the overall percentage of churn rate.

```
SELECT
    churn,no_customer,ROUND((no_customer/total)*100,2)
    "churn_rate in (%)"
FROM
    (SELECT churn,COUNT(churn) AS no_customer FROM
    bank_churn GROUP BY churn) AS total,
    (SELECT COUNT(churn) AS total FROM bank_churn) AS churned;
```

OUTPUT:

	churn	no_customer	churn_rate in (%)
▶	yes	2037	20.37
	no	7963	79.63

From the result above, we can see that 79.63% of customers have not left, while 20.37% of customers have chosen to leave. Let's determine the reasons why customers chose to leave.

2. Now, our work is to know the churn rate on each country,

```sql
SELECT
    total.country,total "number of customer",no_cust 'number_customer(churned)',
    ROUND((no_cust / total) * 100, 2) AS churn_rate
FROM
    (SELECT country, COUNT(customer_id) AS no_cust
    FROM bank_churn
    WHERE churn = 'yes'
    GROUP BY country) AS churned_customer
LEFT JOIN
    (SELECT country,COUNT(customer_id) AS total
    FROM bank_churn
    GROUP BY country) AS total
ON churned_customer.country=total.country;
```

OUTPUT:

country	number of customer	number_customer(churned)	churn_rate
France	5014	810	16.15
Spain	2477	413	16.67
Germany	2509	814	32.44

The churn rates vary among the countries, with Germany having the highest churn rate at 32.44%, followed by Spain at 16.67% and France at 16.15%.

Germany's has higher churn rate so let’s do further investigation to understand the factors contributing to this higher rate compared to the other countries.

3. Now let us look the overall churn rate based on the gender.

```sql
SELECT
    total.gender, number_churned_customer,
    (number_churned_customer / total) * 100 AS churn_rate
FROM
    (SELECT gender, COUNT(customer_id) AS number_churned_customer
    FROM bank_churn WHERE churn="yes" GROUP BY gender) AS gender_details
INNER join
    (SELECT gender, COUNT(customer_id) AS total
    FROM bank_churn GROUP BY gender) AS total
ON gender_details.gender=total.gender;
```

OUTPUT:

gender	number_churned_customer	churn_rate
Female	1139	25.0715
Male	898	16.4559

From this output, we can observe that females have the highest churn rate, i.e., 25.07%, while males have a churn rate of 16.45%. However, our focus is on investigating the churn rate in Germany. Therefore, let us determine the churn rate in Germany based on gender

```sql
WITH churned_customer AS
    (SELECT gender, COUNT(customer_id)  as churned_cust FROM
    bank_churn where country="Germany" and churn="yes"
    GROUP BY gender),
number_customer AS
    (SELECT gender, COUNT(customer_id) as no_cust FROM
    bank_churn WHERE country="Germany"
    GROUP BY gender)
SELECT
    cc.gender, churned_cust, (churned_cust/no_cust)*100 "churn rate(%)"
    FROM churned_customer as cc
LEFT JOIN
    number_customer as nc ON cc.gender=nc.gender;
```

OUTPUT:

	gender	churned_cust	churn rate(%)
▶	Female	448	37.5524
	Male	366	27.8116

The churn rate for females (37.5524%) is higher than that for males (27.8116%). This suggests that, females in Germany are experiencing a higher rate of churn compared to males. So, we can say that female customers are not satisfied with their facilities.

4. Now let us know, which age group of customers are not satisfied with bank facilities.

```sql
SELECT
    churned_customer.age,total_churn.total "number of customer",churned_customer,
    ROUND((churned_customer/total)*100 ,2) "churn_rate in (%)"
FROM
    (SELECT age, COUNT(customer_id) as total FROM bank_churn
    GROUP BY age) as total_churn
LEFT JOIN
    (SELECT age, COUNT(age) AS churned_customer From bank_churn
    WHERE churn="yes" GROUP BY age) as churned_customer
ON
    total_churn.age=churned_customer.age;
```

OUTPUT:

age	number of customer	churned_customer	churn_rate in (%)
adults	4586	1127	24.57
teenager	4153	347	8.36
senior citizen	997	528	52.96
old senior citizen	264	35	13.26

In the senior citizen age group (between the ages of 51 to 65), there is the highest churn rate at 52.96% compared to other age groups. So we conclude that the age between 51 and 65 customers are not satisfied with bank facilities.

5. Now let us determine the feature of credit card can influence the risk of leaving the bank.

```sql
WITH total AS
    (SELECT credit_card, count(customer_id) AS total FROM
    bank_churn GROUP BY credit_card),
    churned_customer AS
    (SELECT credit_card,count(customer_id) AS no_churned_customers FROM
    bank_churn WHERE churn="yes" GROUP BY credit_card)
SELECT
    tt.credit_card,tt.total"number of customer",
    cc.no_churned_customers " number of churned customer",
    (cc.no_churned_customers/tt.total)*100 "churn_rate(%)"FROM total AS tt
LEFT JOIN
    churned_customer AS cc ON tt.credit_card=cc.credit_card;
```

OUTPUT:

	credit_card	number of customer	number of churned customer	churn_rate(%)
▶	yes	7055	1424	20.1843
	no	2945	613	20.8149

In conclusion, it appears that both those who have a credit card and those who do not have, almost the same churn rate. Therefore, we can infer that the features of a credit card may not significantly influence the risk of customers leaving the bank.

6. Now let us determine the churn rate on active member.

```sql
WITH total AS
    (SELECT active_member, count(customer_id) AS total FROM
    bank_churn GROUP BY active_member),
    churned_customer AS
    (SELECT active_member,count(customer_id) AS no_churned_customers FROM
    bank_churn WHERE churn="yes" GROUP BY active_member)
SELECT
    tt.active_member,cc.no_churned_customers "numberof churned customer",
    (cc.no_churned_customers/tt.total)*100 "churn_rate(%)"FROM total AS tt
LEFT JOIN
    churned_customer AS cc ON tt.active_member=cc.active_member;
```

OUTPUT:

active_member	numberof churned customer	churn_rate(%)
yes	735	14.2691
no	1302	26.8509

Customers who are active members in the bank have lower churn rate 14.27% compared to non-active members.

So, being an active member in the bank seems to be linked with a lower risk of leaving. Encouraging and maintaining active membership could be important in reducing the number of customers who leave the bank.

7. Let us determine the percentage churn rate of non-active members in each country.

```sql
WITH country_total AS
    (SELECT country, COUNT(customer_id) AS total FROM
    bank_churn WHERE active_member="no"  GROUP BY country),
churned_cust AS
    (SELECT country, COUNT(customer_id) AS churned_cust FROM
    bank_churn WHERE active_member="no" and churn="yes" GROUP BY
    country)
SELECT
    tt.country,cc.churned_cust,(cc.churned_cust/tt.total)*100
    "churn rate (%)" from country_total as tt
LEFT JOIN
    churned_cust as cc ON tt.country=cc.country;
```

OUTPUT:

country	churned_cust	churn rate (%)
France	512	21.1308
Spain	272	23.3476
Germany	518	41.0785

Here, the churn rate of inactive customers, In Germany is high (41.0785%) compared to the other two country. So Germany was high risk in leaving the bank.

8. Let us determine the churn rate based on tenure

```sql
WITH TotalCustomers AS (
    SELECT tenure, count(tenure) AS no_customer
    FROM bank_churn GROUP BY tenure
),
ChurnedCustomers AS (
    SELECT tenure, COUNT(tenure) AS no_churned_customer
    FROM bank_churn WHERE churn = "yes"
    GROUP BY tenure
)
SELECT TC.tenure, TC.no_customer, CC.no_churned_customer,
 (CC.no_churned_customer / TC.no_customer) * 100 AS churn_rate
FROM TotalCustomers TC
LEFT JOIN ChurnedCustomers CC ON TC.tenure = CC.tenure order by tenure;
```

OUTPUT:

tenure	no_customer	no_churned_customer	churn_rate
0	413	95	23.0024
1	1035	232	22.4155
2	1048	201	19.1794
3	1009	213	21.1100
4	989	203	20.5258
5	1012	209	20.6522
6	967	196	20.2689
7	1028	177	17.2179
8	1025	197	19.2195
9	984	213	21.6463
10	490	101	20.6122

Here, we observed that the churn rate is almost the same for all tenures, ranging between 17.21% and 23.00%. However, newly acquired customers show a higher churn rate (23.00% and 22.41%) compared to other customer segments.

9. Now let us determine how credit scores influence the churn rate.

```sql
SELECT
    total_churn.credit_score,churned_customer,
    ROUND((churned_customer/total)*100 ,2) "churn_rate in (%)"
FROM
    (SELECT credit_score, COUNT(customer_id) as total FROM bank_churn
    GROUP BY credit_score) as total_churn
LEFT JOIN
    (SELECT credit_score, COUNT(customer_id) AS churned_customer From bank_churn
    WHERE churn="yes" GROUP BY credit_score) as churned_customer
ON total_churn.credit_score=churned_customer.credit_score;
```

OUTPUT:

credit_score	churned_customer	churn_rate in (%)
fair	685	20.56
poor	520	22.02
good	452	18.62
excellent	128	19.54
Very good	252	20.59

Here, we can also state that there is not much difference in the churn rate based on class of credit score. But who are having the poor credit score having high churn rate.

10. The remaining two factors, balance, and estimated salary, may not directly impact the churn rate. However, there is a possibility of an indirect effect on customer churn. For instance, a higher account balance or estimated salary might contribute to increased customer satisfaction, potentially lowering the likelihood of churn. Conversely, lower balances or salary estimates might be associated with dissatisfaction, potentially influencing churn rates.

Insight Section:

- **Overall Churn Rate:**

  The overall churn rate stands at 20.37%, with 79.63% of customers choosing to stay.

- **Country-wise Churn Rates:**

  Germany has the highest churn rate at 32.44%, followed by Spain at 16.67% and France at 16.15%.

- **Gender Impact on Churn:**

  Females, overall, exhibit a higher churn rate (25.07%) compared to males (16.45%). In Germany, female churn is even more i.e, 37.55% suggesting dissatisfaction with bank facilities.

- **Age Group Analysis:**

  The senior citizen age group (51 to 65) has the highest churn rate at 52.96%, indicating dissatisfaction with bank facilities.

- **Credit Card Influence:**

  The presence or absence of a credit card does not significantly impact churn rates. Both groups show almost the same churn rate, suggesting credit card features may not be a major factor.

- **Active Membership and Churn:**

  Active members have a lower churn rate (14.27%) compared to non-active members, emphasizing the importance of encouraging and maintaining active membership.

- **Non-Active Members by Country:**

  Germany stands out with a high churn rate (41.08%) among inactive customers, signifying a higher risk of leaving the company.

- **Churn Rate by Tenure:**

  Churn rates are relatively consistent across all tenures, ranging between 17.21% and 23.00%. Newly acquired customers, however, show a higher churn rate (23.00% and 22.41%).

- **Credit Scores and Churn:**

  There is not much difference in churn rates based on the class of credit score. However, customers with poor credit scores tend to exhibit higher churn rates.

- **Balance and Estimated Salary:**

  While balance and estimated salary may not directly impact churn rate, there is a possibility of an indirect effect. Higher balances or salaries could contribute to increased customer satisfaction, potentially lowering churn rates.

Recommendation to reduce the churn rate:

- **Address Female Dissatisfaction in Germany:**

  Given the higher churn rate among females, particularly in Germany, conduct customer satisfaction surveys or focus groups to identify specific pain points and areas for improvement.

- **Target Senior Citizen Engagement:**

  Since the senior citizen age group exhibits the highest churn rate, implement targeted initiatives to improve services and experiences for this demographic.

- **Promote Active Membership:**

  Given that active members have a significantly lower churn rate, implement campaigns to encourage and reward active participation. Provide exclusive benefits, discounts, or personalized services to retain and engage active members.

- **Focus on new Customers:**

  The higher churn rate among newly acquired customers. Focus on newly customers to ensure that new customers quickly understand and appreciate the full range of services and benefits offered by the bank.

- **Improve Inactive Customer Engagement:**

  Develop targeted campaigns to re-engage inactive customers, especially in high-churn areas like Germany. Offer incentives, personalized promotions, or exclusive services to encourage their continued relationship with the bank.

- **Continuous Feedback Loop:**

  Establish a continuous feedback loop through regular customer surveys and feedback mechanisms. Actively listen to customer concerns and use the insights gained to make continuous improvements to services and offerings.

POWER BI DASHBOARD: