Study customer churn analytics for the banking industry

- Customer churn is the percentage of customers that stopped using your company's product or service over a time period.
- This will help a banking company understand what makes a customer leave and how to combat it.

by ***Soumyadeep Das***

Topics

- Explore DDL and DML Commands
- Summary statistics (Average Credit score & Balance)
- Identify the churned & non-churned
- Determining metrics to understand Customer Churn pattern
- Top metrics to understand customer churn pattern

Explore DDL command

```
-- Using DDL command Create for creating the table --

CREATE TABLE Churn_Modeling (
    RowNumber INT,
    CustomerId INT,
    Surname VARCHAR(255),
    CreditScore INT,
    Geography VARCHAR(255),
    Gender VARCHAR(255),
    Age INT,
    Tenure INT,
    Balance DECIMAL(10, 2),
    NumOfProducts INT,
    HasCrCard INT,
    IsActiveMember INT,
    EstimatedSalary DECIMAL(10, 2),
    Exited INT
);
```

```
--import the csv using copy--

COPY Churn_Modeling (RowNumber, CustomerId, Surname, CreditScore, Geography, Gender, Age, Tenure, Balance,
    NumOfProducts, HasCrCard, IsActiveMember, EstimatedSalary, Exited)
FROM 'C:\Users\soumy\Desktop\Internship Offer letters\Jupiter money\Churn Modeling.csv'
DELIMITER ','
CSV HEADER;
```

Exploring DML commands :

```
-- Using select to View all customers --

SELECT * FROM Churn_Modeling;
```

	rownumber integer	customerid integer	surname character varying (255)	creditscore integer	geography character varying (255)	gender character varying (255)	age integer	tenure integer	balance numeric (10,2)	numofproducts integer	hascrcard integer	isactivemember integer	estimatedsalary numeric (10,2)	exited integer
1	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88	1
2	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
3	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57	1
4	4	15701354	Boni	699	France	Female	39	1	0.00	2	0	0	93826.63	0
5	5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0
6	6	15574012	Chu	645	Spain	Male	44	8	113755.78	2	1	0	149756.71	1
7	7	15592531	Bartlett	822	France	Male	50	7	0.00	2	1	1	10062.80	0
8	8	15656148	Obinna	376	Germany	Female	29	4	115046.74	4	1	0	119346.88	1
9	9	15792365	He	501	France	Male	44	4	142051.07	2	0	1	74940.50	0
10	10	15592389	H?	684	France	Male	27	2	134603.88	1	1	1	71725.73	0
11	11	15767821	Bearce	528	France	Male	31	6	102016.72	2	0	0	80181.12	0
12	12	15737173	Andrews	497	Spain	Male	24	3	0.00	2	1	0	76390.01	0
13	13	15632264	Kay	476	France	Female	34	10	0.00	2	1	0	26260.98	0
14	14	15691483	Chin	549	France	Female	25	5	0.00	2	0	0	190857.79	0
15	15	15600882	Scott	635	Spain	Female	35	7	0.00	2	1	1	65951.65	0
16	16	15643966	Goforth	616	Germany	Male	45	3	143129.41	2	0	1	64327.26	0
17	17	15737452	Romeo	653	Germany	Male	58	1	132602.88	1	1	0	5097.67	1
18	18	15788218	Henderson	549	Spain	Female	24	9	0.00	2	1	1	14406.41	0
19	19	15661507	Muldrow	587	Spain	Male	45	6	0.00	1	0	0	158684.81	0
20	20	15568982	Hao	726	France	Female	24	6	0.00	2	1	1	54724.03	0
21	21	15577657	McDonald	732	France	Male	41	8	0.00	2	1	1	170886.17	0

Data Input

```sql
-- Using select Distribution of customers by age group --

SELECT
    CASE
        WHEN Age < 20 THEN 'Under 20'
        WHEN Age BETWEEN 20 AND 29 THEN '20-29'
        WHEN Age BETWEEN 30 AND 39 THEN '30-39'
        WHEN Age BETWEEN 40 AND 49 THEN '40-49'
        WHEN Age BETWEEN 50 AND 59 THEN '50-59'
        ELSE '60 and above'
    END AS AgeGroup,
    COUNT(*) AS NumberOfCustomers
FROM Churn_Modeling
GROUP BY AgeGroup
ORDER BY AgeGroup;
```

Data Output

	agegroup text	numberofcustomers bigint
1	20-29	1592
2	30-39	4346
3	40-49	2618
4	50-59	869
5	60 and above	526
6	Under 20	49

Customer Insights

```sql
-- Using select Top 10 customers with the highest balance --

SELECT * FROM Churn_Modeling
ORDER BY Balance DESC
LIMIT 10;
```

	rownumber integer	customerid integer	surname character varying (255)	creditscore integer	geography character varying (255)	gender character varying (255)	age integer	tenure integer	balance numeric (10,2)	numofproducts integer	hascrcard integer	isactivemember integer	estimatedsalary numeric (10,2)	exited integer
1	2093	15757408	Lo	655	Spain	Male	38	3	250898.09	3	0	1	81054.00	1
2	3281	15715622	To Rot	583	France	Female	57	3	238387.56	1	0	1	147964.99	1
3	8734	15714241	Haddon	749	Spain	Male	42	9	222267.63	1	0	0	101108.85	1
4	3589	15571958	McIntosh	489	Spain	Male	40	3	221532.80	1	1	0	171867.08	0
5	6718	15586674	Shaw	663	Spain	Female	58	5	216109.88	1	0	1	74176.71	1
6	1068	15599131	Dilke	650	Germany	Male	26	4	214346.96	2	1	0	128815.33	0
7	139	15594408	Chia	584	Spain	Female	48	2	213146.20	1	1	0	75161.25	1
8	1534	15769818	Moore	850	France	Female	37	3	212778.20	1	0	1	69372.88	0
9	3921	15620268	Thomson	634	Germany	Male	43	3	212696.32	1	1	0	115268.86	0
10	2710	15780212	Mao	592	France	Male	37	4	212692.97	1	0	0	176395.02	0

Data Input

```sql
-- age group distribution of exited vs. non-exited customers --

SELECT
    CASE
        WHEN Age < 20 THEN 'Under 20'
        WHEN Age BETWEEN 20 AND 29 THEN '20-29'
        WHEN Age BETWEEN 30 AND 39 THEN '30-39'
        WHEN Age BETWEEN 40 AND 49 THEN '40-49'
        WHEN Age BETWEEN 50 AND 59 THEN '50-59'
        ELSE '60 and above'
    END AS AgeGroup,
    Exited,
    COUNT(*) AS NumberOfCustomers
FROM Churn_Modeling
GROUP BY AgeGroup, Exited
ORDER BY AgeGroup, Exited;
```

Data Output

	agegroup text	exited integer	numberofcustomers bigint
1	20-29	0	1471
2	20-29	1	121
3	30-39	0	3873
4	30-39	1	473
5	40-49	0	1812
6	40-49	1	806
7	50-59	0	382
8	50-59	1	487
9	60 and above	0	379
10	60 and above	1	147
11	Under 20	0	46
12	Under 20	1	3

Churn Rate

Data Input

```
SELECT Exited, COUNT(*) AS NumberOfCustomers
FROM Churn_Modeling
GROUP BY Exited;
```

Data Output

	exited integer	numberofcustomers bigint
1	0	7963
2	1	2037

```sql
-- Query to calculate Retention Rate --

SELECT
    (SUM(CASE WHEN Exited = 0 THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS RetentionRate
FROM Churn_Modeling;
```

Data Output

	retentionrate numeric
1	79.6300000000000000

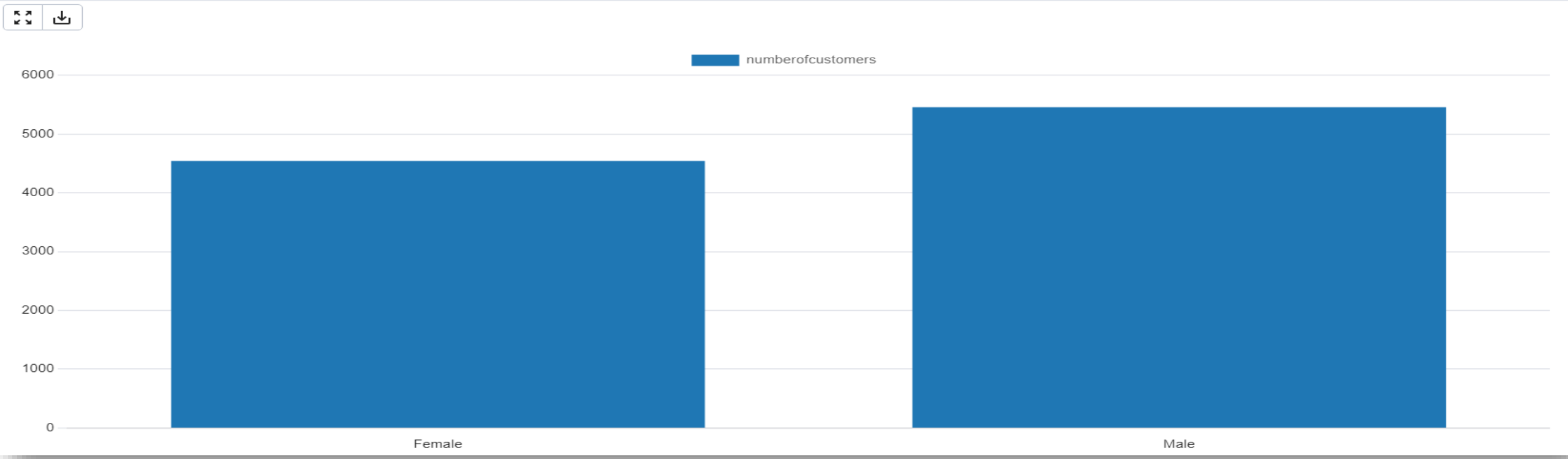
Data Input

```
-- Using select Count the number of customers by gender --

SELECT Gender, COUNT(*) AS NumberOfCustomers
FROM Churn_Modeling
GROUP BY Gender;
```

Data Output

	gender character varying (255)	numberofcustomers bigint
1	Female	4543
2	Male	5457



Summary Statistics

Data Input

```sql
-- the query to get summary statistics for numerical variables --

SELECT
    MIN(CreditScore) AS MinCreditScore,
    MAX(CreditScore) AS MaxCreditScore,
    AVG(CreditScore) AS AvgCreditScore,
    STDDEV(CreditScore) AS StdDevCreditScore,
    MIN(Age) AS MinAge,
    MAX(Age) AS MaxAge,
    AVG(Age) AS AvgAge,
    STDDEV(Age) AS StdDevAge,
    MIN(Tenure) AS MinTenure,
    MAX(Tenure) AS MaxTenure,
    AVG(Tenure) AS AvgTenure,
    STDDEV(Tenure) AS StdDevTenure,
    MIN(Balance) AS MinBalance,
    MAX(Balance) AS MaxBalance,
    AVG(Balance) AS AvgBalance,
    STDDEV(Balance) AS StdDevBalance,
    MIN(EstimatedSalary) AS MinEstimatedSalary,
    MAX(EstimatedSalary) AS MaxEstimatedSalary,
    AVG(EstimatedSalary) AS AvgEstimatedSalary,
    STDDEV(EstimatedSalary) AS StdDevEstimatedSalary
FROM
    Churn_Modeling;
```

Data Output

avgcreditscore numeric
650.528800000000000

avgbalance numeric
76485.889288000000

mincreditscore integer	maxcreditscore integer	avgcreditscore numeric	stddevcreditscore numeric	minage integer	maxage integer	avgage numeric	stddevage numeric	mintenure integer	maxtenure integer	avgtenure numeric
350	850	650.528800000000000	96.653298736130354	18	92	38.921800000000000	10.487806451704610	0	10	5.012800000000

stddevtenure numeric	minbalance numeric	maxbalance numeric	avgbalance numeric	stddevbalance numeric	minestimatedsalary numeric	maxestimatedsalary numeric	avgestimatedsalary numeric	stddevestimatedsalary numeric
2.892174377049683	0.00	250898.09	76485.889288000000	62397.40520239	11.58	199992.48	100090.239881000000	57510.49281770

How varied mean is in this case :

Ease of Calculation: The mean is straightforward to compute, requiring only the sum of the values divided by the number of values.
Understandability: It's an easily comprehensible measure, making it accessible for various stakeholders to grasp average customer behavior quickly.
Reflects Shifts: The mean is highly responsive to changes in data. Any variation in the average churn rate is promptly reflected, serving as a clear indicator of changes within the customer base.
Benchmarking: The mean facilitates easy comparison against historical data or benchmarks. By assessing the current mean churn rate against past data, banks can evaluate whether their situation is improving or deteriorating.
Strategic Insights: Understanding the average churn rate is crucial for strategic decision-making. A high mean churn rate might signal the need for interventions, such as enhancing customer service or providing better incentives to retain customers.
Identifying Trends: By calculating the mean churn rate for different customer segments, banks can identify which segments exhibit higher or lower churn rates. This information is vital for tailoring specific retention strategies for each segment.
Model Input: The mean churn rate serves as a valuable input for predictive models. By incorporating the average churn rate, banks can build more accurate models to forecast future churn rates, aiding in proactive customer retention efforts.

How varied median is in this case :

Customer Churn Data: Customer churn data can have extreme values, such as customers who churn almost immediately and those who stay for an exceptionally long time.
Mean vs. Median: The mean is sensitive to these outliers, which can distort the average. In contrast, the median is less affected by extreme values and provides a more stable measure of central tendency. This stability is crucial for understanding typical customer behavior without the influence of outliers.
Non-Normal Distributions: Churn data often does not follow a normal distribution. It may be skewed, for instance, if a large number of customers churn within a short period while a few remain loyal for a long time.
Central Value Representation: In skewed distributions, the median better represents the central value because it is the midpoint. The mean can be pulled in the direction of the skew, giving a distorted view of the central tendency. The median provides a clearer picture of the typical customer behavior in such cases.
Typical Customer Experience: The median represents the "typical" customer experience. For example, when analyzing the tenure of customers before they churn, the median tenure indicates the point at which half the customers have churned, and half have not.
Meaningful Comparisons: This makes the median a valuable metric for comparing customer segments or time periods, especially when the distributions of these segments or periods vary. It allows businesses to understand customer behavior more effectively.
Ease of Interpretation: The median is straightforward and easy to interpret. For instance, if the median tenure before churn is 12 months, it means that half the customers churn within a year.
Communication: This simplicity makes it easier to communicate insights to stakeholders who may not have a strong statistical background. It provides a clear, understandable measure that can guide decision-making.

Determine the metrics to understand Customer Churn

Statistical metrics are indispensable for predicting as well as understanding the customer churn. The patterns and variables that are mainly responsible for customer churn can be discovered by using the main statistics. Here is a comprehensive analysis using statistical metrics :

Data analysis forms a strong base for businesses on one hand,
on the other, it is a key tool for better customer support, and creating the right kind of product and service offers. Interpreting their insight of this data
companies can discover regular customer behavior, detect lack of service levels or discover the reasons behind a spate of customers who do not like offerings by the company.

Determining the top metrics are indeed crucial to evaluate reasons for customer outflow, allowing the businesses to take specific actions to lessen the risk of that situation and ensure customer retention.

Top metrics that will help understand the churn pattern :

I. **Credit Score**: Helps understand the financial reliability of customers. Lower credit scores may correlate with higher churn.

II. **Geography**: Identifies regional patterns and economic conditions affecting churn.

III. **Gender**: Analyzes if gender influences customer behavior and churn rates.

IV. **Age**: Assesses age-related trends in churn, as different age groups may have different needs.

V. **Tenure**: Longer tenure often indicates loyalty; shorter tenure might indicate a higher risk of churn.

VI. **Balance**: Indicates financial engagement. High or zero balances might show different churn behaviors.

VII. **Retention rate :** A crucial metric for understanding customer churn because it directly measures how well a company is keeping its customers over a given period

VIII. **Customer Lifetime Value (CLV)**: Plays crucial metric for understanding customer churn because it provides a comprehensive view of the financial impact of losing a customer.

IX. **Estimated Salary**: Higher salaries often indicate financial stability and lower churn risk.