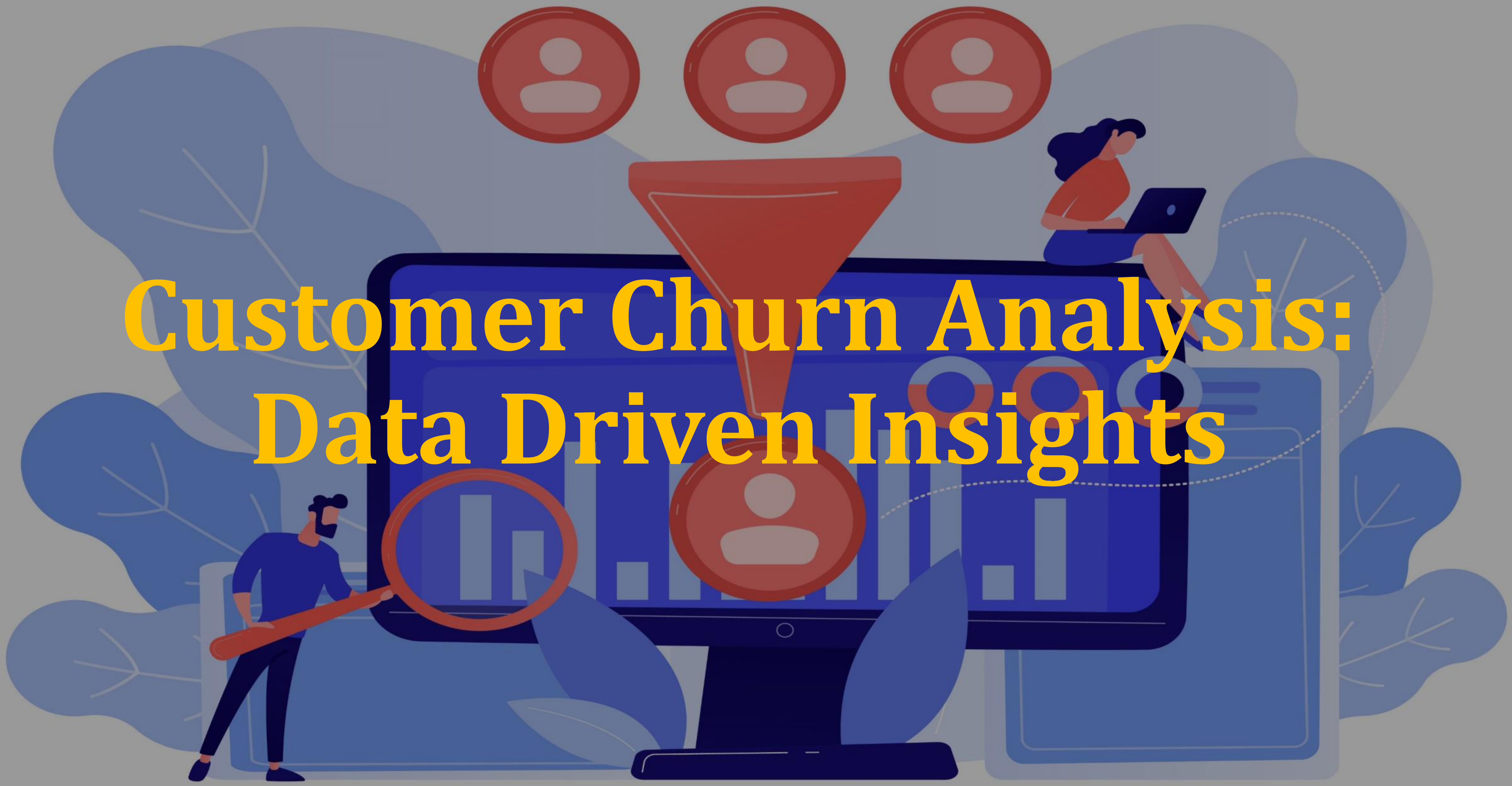
Customer Churn Analysis: Data Driven Insights



Project Overview

- In this project, I analyzed customer churn data from a banking institution to uncover key factors influencing attrition and derive actionable insights for retention strategies.
- Leveraging MS-SQL, I efficiently handled, queried, and processed the dataset, which consisted of 10,000 unique customer records.
- Each entry provided essential attributes such as Age, Country, Tenure, Gender, Balance, and Credit Score, enabling a comprehensive exploration of churn patterns.
- I executed data-driven operations, including data cleaning, transformation, and exploratory analysis, to ensure accuracy and consistency.
- By applying advanced SQL techniques, I extracted valuable insights that highlight behavioral trends, segment-specific risks, and strategic recommendations to mitigate churn.

Data Cleaning

1) Identify and Remove duplicate values

```
select
count(1) as Total_Rows,
count(distinct RowNumber) as RowNumber,
count(distinct CustomerId) as CustomerId
from Churn_Modelling;
```

```
with A as
(select
RowNumber,
CustomerId,
ROW_NUMBER() over (partition by CustomerId order by RowNumber asc) as ranking
from Churn_Modelling)
delete from Churn_Modelling
where RowNumber in (select RowNumber from A where ranking > 1);
```

Removing Duplicate values from CustomerId column

2) Dealing with Data Inconsistencies.

```sql
-- 3) Finding Data Inconsistencies
select distinct Geography from Churn_Modelling;
```

There were some inconsistencies
In the Geography column

	Geography
1	Germany
2	Spaine
3	Gremany
4	Spen
5	Germani
6	Germeny
7	France
8	Spain

```sql
update Churn_Modelling
set Geography =
case when Geography = 'Spaine' then 'Spain'
when Geography = 'Spen' then 'Spain'
when Geography = 'Gremany' then 'Germany'
when Geography = 'Germani' then 'Germany'
when Geography = 'Germeny' then 'Germany'
else Geography end;
```

Corrected inconsistencies to ensure the Data accuracy

Data Transformation

- Created few columns to categorise the numerical data.

1) Categorising customers into age groups

```sql
alter table Churn_Modelling
add Age_Group nvarchar(15);

update Churn_Modelling
set Age_Group = case
when Age >= 18 and Age < 26 then 'Young'
when Age >= 26 and Age < 41 then 'Adult'
when Age >= 41 and Age < 61 then 'Senior Adult'
else 'Senior Citizen' end;
```

2) Categorising customers based on Tenure

```sql
alter table Churn_Modelling
add Tenure_Group nvarchar(30);

update Churn_Modelling
set Tenure_Group = case
when Tenure between 0 and 1 then 'New Customers'
when Tenure between 2 and 4 then 'Established Customers'
when Tenure between 5 and 7 then 'Loyal Customers'
else 'Veteran Customers' end;
```

3) Categorising customers based on Salary

```sql
alter table Churn_Modelling
add Salary_Group nvarchar(30);

update Churn_Modelling
set Salary_Group = case
when EstimatedSalary < 20000 then 'Low'
when EstimatedSalary between 20000 and 100000 then 'Medium'
when EstimatedSalary > 100000 then 'High'
else 'N/A' end;
```

4) Categorising customers based on Credit Score

```sql
alter table Churn_Modelling
add CS_Group nvarchar(20);

update Churn_Modelling
set CS_Group = case
when CreditScore < 580 then 'Poor'
when CreditScore between 580 and 669 then 'Fair'
when CreditScore between 670 and 749 then 'Good'
else 'Excellent' end;
```

Through data preparation operations such as cleaning and transformations, we ensured the data was suitable for deeper analysis.
Now, let's explore the insights derived from our concise yet thorough examination.

Insights

- Total churned customers and their percentage

```sql
select count(1) as Churned_Customers,
count(1)*100.0/(select count(1)
from Churn_Modelling) as Churn_Percentage
from Churn_Modelling where Exited = 1;
```

Results | Messages

	Churned_Customers	Churn_Percentage
1	2037	20.370000000000

- Female customers exhibit a higher churn rate compared to their male counterparts. Additionally, the proportion of inactive accounts is notably greater among female customers than among males.

```sql
select
Gender,
count(1)*100.0/(select count(1) from Churn_Modelling) as Percent_Share,
sum(case when Exited = 1 then 1 else 0 end)*100.0/count(1)
as Churn_Rate,
sum(case when IsActiveMember = 0 then 1 else 0 end)*100.0/count(1)
as Inactive_Member_Percent
from Churn_Modelling group by Gender;
```

Results | Messages

	Gender	Percent_Share	Churn_Rate	Inactive_Member_Percent
1	Male	54.570000000000	16.455928165658	47.461975444383
2	Female	45.430000000000	25.071538630860	49.724851419766

- Despite the 26-40 and 41-60 age groups having a similar percentage of inactive customers, the churn rate among the **41-60 ('Senior Adult')** segment stands out at approximately 40%, indicating a significantly higher attrition rate. Meanwhile, younger customers (18-25) exhibit the lowest churn rate.

```sql
select
Age_Group,
count(1)*100.0/(select count(1) from Churn_Modelling)
as Percent_Share,
sum(case when Exited = 1 then 1 else 0 end)*100.0/count(1)
as Churn_Rate,
sum(case when IsActiveMember = 0 then 1 else 0 end)*100.0/count(1)
as Inactive_Member_Percent
from Churn_Modelling group by Age_Group;
```

Results | Messages

	Age_Group	Percent_Share	Churn_Rate	Inactive_Member_Percent
1	Adult	58.080000000000	11.019283746556	50.103305785123
2	Senior Adult	31.170000000000	39.653512993262	50.112287455887
3	Young	6.110000000000	7.528641571194	47.135842880523
4	Senior Citizen	4.640000000000	24.784482758620	19.181034482758

- Germany exhibits a significantly higher customer churn rate of 32%, nearly double that of other countries.

```sql
select
Geography,
sum(case when Exited = 1 then 1 else 0 end)*100.0/count(1)
as Churn_Rate,
sum(case when IsActiveMember = 0 then 1 else 0 end)*100.0/count(1)
as Inactive_Customers
from Churn_Modelling group by Geography;
```

Results | Messages

	Geography	Churn_Rate	Inactive_Customers
1	Germany	32.443204463929	50.259067357512
2	France	16.154766653370	48.324690865576
3	Spain	16.673395236172	47.032700847799

- This trend is particularly concerning among female customers, where churn reaches 37%, far exceeding the average of around 20% observed elsewhere.

Results | Messages

	Geography	Gender	Churn_Rate
1	Germany	Female	37.552388935456
2	France	Female	20.344980097302
3	Spain	Female	21.212121212121

- Since we already know that customer churn is highest among the 'Senior Adult' age group, the disproportionately high churn rate among female customers within this segment is particularly concerning and cannot be overlooked.

```sql
select
Age_Group, Gender,
sum(
case
when Exited = 1 then 1 else 0
end
)*100.0/count(1) as Churn_Rate
from Churn_Modelling
where Age_Group = 'Senior Adult'
group by Age_Group, Gender;
```

Results | Messages

	Age_Group	Gender	Churn_Rate
1	Senior Adult	Male	34.009840098400
2	Senior Adult	Female	45.808182427900

- Among male customers, those with poor credit scores (below 580) exhibit the highest churn rate. However, among female customers, the largest segment—those with fair credit scores (between 580 and 669)—experiences the highest churn.

Results | Messages

	Gender	CS_Group	Churn_Rate
1	Female	Excellent	24.668435013262
2	Female	Poor	25.638599810785
3	Female	Fair	26.124338624338
4	Female	Good	23.524590163934
5	Male	Good	15.300546448087
6	Male	Fair	15.942825728422
7	Male	Poor	19.080459770114
8	Male	Excellent	15.535097813578

- An emerging pattern reveals that customers who purchase three or four products are exhibiting a significantly higher churn rate.

```sql
select NumOfProducts as No_of_Products,
count(1) as Total_Customers,
sum(case when Exited = 1 then 1 else 0 end) as Churned_Customers,
sum(case when Exited = 1 then 1 else 0 end)*100.0/count(1)
as Churn_Rate
from Churn_Modelling
group by NumOfProducts
order by NumOfProducts;
```

Results | Messages

	No_of_Products	Total_Customers	Churned_Customers	Churn_Rate
1	1	5084	1409	27.714398111723
2	2	4590	348	7.581699346405
3	3	266	220	82.706766917293
4	4	60	60	100.000000000000

- We've already established that churn rates among inactive members are high.

```sql
select Total_Customers,
Active_Members,
Inactive_Members,
ChurnedActiveMembers*100.0/Active_Members as Churn_rate_active,
ChurnedInactiveMembers*100.0/Inactive_Members as Churn_rate_inactive
from
(select count(1) as Total_Customers,
sum(case when IsActiveMember = 1 then 1 else 0 end)
as Active_Members,
sum(case when IsActiveMember = 0 then 1 else 0 end)
as Inactive_Members,
sum(case when IsActiveMember = 1 and Exited =1 then
1 else 0 end) as ChurnedActiveMembers,
sum(case when IsActiveMember = 0 and Exited =1 then
1 else 0 end) as ChurnedInactiveMembers
from Churn_Modelling) A
```

Results | Messages

	Total_Customers	Active_Members	Inactive_Members	Churn_rate_active	Churn_rate_inactive
1	10000	5151	4849	14.269073966220	26.850897092183

- Inactive credit card holders exhibit a higher churn rate compared to inactive customers without a credit card. Additionally, active credit card holders experience lower churn compared to active members who do not have a credit card.

```sql
select
'CreditCard Holder' as Customer_type,
sum(case when HasCrCard = 1 then 1 else 0 end) as Total_Customers,
sum(case when HasCrCard = 1 and IsActiveMember = 0 then 1 else 0 end) as Inactive,
sum(case when HasCrCard = 1 and IsActiveMember = 1 then 1 else 0 end) as Active,
sum(case when HasCrCard = 1 and IsActiveMember = 0 and Exited = 1 then 1 else 0 end)*100.0/
sum(case when HasCrCard = 1 and IsActiveMember = 0 then 1 else 0 end) as Churn_Rate_among_inactive,
sum(case when HasCrCard = 1 and IsActiveMember = 1 and Exited = 1 then 1 else 0 end)*100.0/
sum(case when HasCrCard = 1 and IsActiveMember = 1 then 1 else 0 end) as Churn_Rate_among_active
from Churn_Modelling
union all
select 'Not CreditCard Holder' as Customer_type,
sum(case when HasCrCard = 0 then 1 else 0 end) as Total_Customers,
sum(case when HasCrCard = 0 and IsActiveMember = 0 then 1 else 0 end) as Inactive,
sum(case when HasCrCard = 0 and IsActiveMember = 1 then 1 else 0 end) as Active,
sum(case when HasCrCard = 0 and IsActiveMember = 0 and Exited = 1 then 1 else 0 end)*100.0/
sum(case when HasCrCard = 0 and IsActiveMember = 0 then 1 else 0 end) as Churn_Rate,
sum(case when HasCrCard = 0 and IsActiveMember = 1 and Exited = 1 then 1 else 0 end)*100.0/
sum(case when HasCrCard = 0 and IsActiveMember = 1 then 1 else 0 end) as Churn_Rate2
from Churn_Modelling
```

Results | Messages

	Customer_type	Total_Customers	Inactive	Active	Churn_Rate_among_inactive	Churn_Rate_among_Active
1	Card Holder	7055	3448	3607	27.320185614849	13.362905461602
2	Not Card Holder	2945	1401	1544	25.695931477516	16.386010362694

Summarizing Findings

- The overall churn rate stands at 20.37%.
- The highest churn is observed among female customers and those in the Senior Adult (41-60) age group.
- Germany has the highest churn rate at 32.44%, with German female customers experiencing an even higher rate of 37.55%.
- Inactive customers are a key driver of churn, suggesting that boosting customer engagement could help reduce attrition.
- Inactive customers with credit cards are more prone to churn compared to those without cards. However, among active customers, those with credit cards exhibit lower attrition rates.
- A majority of customers purchase only one or two products, with just 3.26% buying three or more.
- Customers purchasing more than two products experience churn rates exceeding 85%, potentially indicating dissatisfaction or a lack of product variety. To mitigate this, analyzing customer feedback can help identify the root cause of dissatisfaction.

- The Adult age group (26-40) comprises 58% of total customers, making it the largest segment. Churn within this group is minimal, yet inactivity remains high at 50%, presenting a strong opportunity for engagement strategies. Since this group holds the largest share across all salary segments—high, medium, and low—it has significant potential for revenue growth. Efforts should focus on reducing inactivity and strengthening customer interaction to maximize profitability.

Results | Messages

	Salary_Group	Age_Group	Percent_share
1	High	Adult	57.704590818363
2	Low	Adult	60.851926977687
3	Medium	Adult	57.867132867132

- The Senior Adult age group (41-60) ranks as the second-largest segment but exhibits the highest churn rate among all age groups. Despite this, customers in this group demonstrate a stronger tendency to purchase multiple products, with 6% buying more than one item, compared to 1.9% in Adults (26-40) and 1.6% in Young customers (18-25). Given this trend, optimizing strategies for product bundling and personalized engagement could enhance retention and drive higher-value transactions within this segment.

Results | Messages

	Age_Group	Percent_share
1	Adult	1.962809917355
2	Senior Adult	5.871029836381
3	Young	1.636661211129
4	Senior Citizen	4.094827586206

- The Young customer segment (18-25) holds strong potential for long-term profitability due to its low churn rate and high enthusiasm for investing in multiple services. Additionally, this group has the lowest inactivity rate, indicating strong engagement levels. By successfully attracting and retaining more young customers, businesses can drive sustained revenue growth while fostering a loyal customer base with a higher likelihood of exploring diverse offerings.

Closure

In this project, my primary focus was on enhancing my analytical abilities rather than simply leveraging SQL, a tool in which I am already proficient. Through the process, I discovered hidden trends, identified key patterns, and formulated data-driven recommendations to maximize profitability and expand the customer base.

To achieve this, I utilized various SQL concepts, including DQL, DDL, DML queries, GROUP BY, ORDER BY, and filtering commands like WHERE. Additionally, I incorporated advanced techniques such as CTEs, subqueries, and UNION ALL to extract meaningful insights.

These SQL functionalities enabled a deep dive into customer churn patterns, helping to identify high-risk customer clusters and uncover the key factors driving churn, ultimately facilitating informed strategic decisions.

Thank You !