Telecom Churn Analysis Portfolio Project | Power BI + SQL + Python

Case Study

ANUSHA

8/30/24

DATA ANALYTICS

DATASET CONTENT:

Customer_ID *unique key

Gender
Age
Married
State

*Demographic and geographic info

Number_of_Referrals
Tenure_in_Months
Value_Deal
Phone_Service
Multiple_Lines
Internet_Service
Internet_Type
Online_Security
Online_Backup
Device_Protection_Plan
Premium_Support
Streaming_TV
Streaming_Movies
Streaming_Music
Unlimited_Data
Contract
Paperless_Billing

* Refers made by customers and various services customer had subscribed like internet ,streaming

Payment_Method
Monthly_Charge
Total_Charges
Total_Refunds
Total_Extra_Data_Charges
Total_Long_Distance_Charges
Total_Revenue

* Tells about payment and revenue related columns

Customer_Status ***
Churn_Category
Churn_Reason

* This talks about whether customer is churn or they are still with company

AGENDA

PROJECT GOAL

Create an entire ETL process in a database & a Power BI dashboard to utilize the Customer Data and achieve below goals:

1. Analyze Customer Data at below levels ,
    a. Demographic
    b. Geographic
    c. Payment & Account info
    d. Services
2. Study Churner profile & identify Areas for implementing Marketing campaigns.
3. Identify a Method to Predict Future Churners

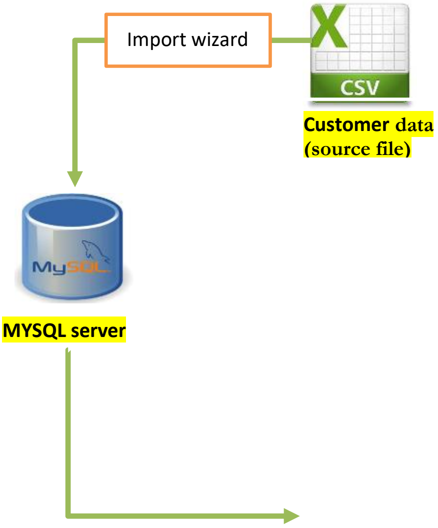
METRICS REQUIRED

1.Total Customers
2.Total Churn & Churn Rate
3.New Joiners

ETL FRAMEWORK

Our framework uses below components:

- CSV file – This is our source file
- SQL Server – Sql workbench – We will use its inbuild import wizard to transform & load the data
- MYSQL Workbench – This is where our final data will be loaded and host our data warehouse, tables & views for final usage.



NOTE: Database system is better at handling recurring data loads and maintaining data Integrity compared to an Excel file

Work Around SQL

- We would be working on SQL by following steps,
    1. Create Database
    2. Create Table
    3. Load Data in to the table.

Note: In real scenarios, first the raw data will be loaded to staging file which doesn't have alterations or modifications .so table name as "stg_churn".

- Using select statement check all data is loaded or not.
- Since we have lot of columns, we will be checking values of the **few of columns** and **understand the data distribution**.
- Will check for data exploration, like checking for null values, missing elements, or any duplications.

```
SELECT
   SUM(CASE WHEN Customer_ID=“ “   THEN 1 ELSE 0 END) AS
Customer_ID_Null_Count,
   SUM(CASE WHEN Gender=“ “  THEN 1 ELSE 0 END) AS Gender_Null_Count,
   SUM(CASE WHEN Age=“ “  THEN 1 ELSE 0 END) AS Age_Null_Count,
   SUM(CASE WHEN Married=“ “  THEN 1 ELSE 0 END) AS Married_Null_Count,
   SUM(CASE WHEN State =“ “  THEN 1 ELSE 0 END) AS State_Null_Count,
   SUM(CASE WHEN Number_of_Referrals =“ “  THEN 1 ELSE 0 END) AS
Number_of_Referrals_Null_Count,
   SUM(CASE WHEN Tenure_in_Months=“ “  THEN 1 ELSE 0 END) AS
Tenure_in_Months_Null_Count,
   SUM(CASE WHEN Value_Deal=“ “  THEN 1 ELSE 0 END) AS Value_Deal_Null_Count,
   SUM(CASE WHEN Phone_Service =“ “  THEN 1 ELSE 0 END) AS
Phone_Service_Null_Count,
   SUM(CASE WHEN Multiple_Lines =“ “  THEN 1 ELSE 0 END) AS
Multiple_Lines_Null_Count,
   SUM(CASE WHEN Internet_Service =“ “  THEN 1 ELSE 0 END) AS
Internet_Service_Null_Count,
   SUM(CASE WHEN Internet_Type =“ “  THEN 1 ELSE 0 END) AS
Internet_Type_Null_Count,
   SUM(CASE WHEN Online_Security =“ “  THEN 1 ELSE 0 END) AS
Online_Security_Null_Count,
   SUM(CASE WHEN Online_Backup=“ “  THEN 1 ELSE 0 END) AS
Online_Backup_Null_Count,
   SUM(CASE WHEN Device_Protection_Plan=“ “  THEN 1 ELSE 0 END) AS
Device_Protection_Plan_Null_Count,
   SUM(CASE WHEN Premium_Support=“ “  THEN 1 ELSE 0 END) AS
Premium_Support_Null_Count,
   SUM(CASE WHEN Streaming_TV=“ “  THEN 1 ELSE 0 END) AS
Streaming_TV_Null_Count,
   SUM(CASE WHEN Streaming_Movies=“ “  THEN 1 ELSE 0 END) AS
Streaming_Movies_Null_Count,
   SUM(CASE WHEN Streaming_Music =“ “  THEN 1 ELSE 0 END) AS
Streaming_Music_Null_Count,
   SUM(CASE WHEN Unlimited_Data=“ “  THEN 1 ELSE 0 END) AS
Unlimited_Data_Null_Count,
   SUM(CASE WHEN Contract =“ “  THEN 1 ELSE 0 END) AS Contract_Null_Count,
   SUM(CASE WHEN Paperless_Billing =“ “  THEN 1 ELSE 0 END) AS
Paperless_Billing_Null_Count,
   SUM(CASE WHEN Payment_Method=“ “  THEN 1 ELSE 0 END) AS
Payment_Method_Null_Count,
   SUM(CASE WHEN Monthly_Charge=“ “  THEN 1 ELSE 0 END) AS
Monthly_Charge_Null_Count,
   SUM(CASE WHEN Total_Charges=“ “  THEN 1 ELSE 0 END) AS
Total_Charges_Null_Count,
   SUM(CASE WHEN Total_Refunds =“ “  THEN 1 ELSE 0 END) AS
Total_Refunds_Null_Count,
   SUM(CASE WHEN Total_Extra_Data_Charges =“ “  THEN 1 ELSE 0 END) AS
Total_Extra_Data_Charges_Null_Count,
   SUM(CASE WHEN Total_Long_Distance_Charges =“ “  THEN 1 ELSE 0 END) AS
Total_Long_Distance_Charges_Null_Count,
   SUM(CASE WHEN Total_Revenue =“ “  THEN 1 ELSE 0 END) AS
Total_Revenue_Null_Count,
   SUM(CASE WHEN Customer_Status =“ “  THEN 1 ELSE 0 END) AS
Customer_Status_Null_Count,
   SUM(CASE WHEN Churn_Category =“ “  THEN 1 ELSE 0 END) AS
Churn_Category_Null_Count,
   SUM(CASE WHEN Churn_Reason=“ “  THEN 1 ELSE 0 END) AS
Churn_Reason_Null_Count
FROM stg_Churn;
```

Working on null/empty cell and replacing with required value,

```
CREATE TABLE prod_Churn AS
SELECT
   Customer_ID,
   Gender,
   Age,
   Married,
   State,
    Number_of_Referrals,
    Tenure_in_Months,
    IF(Value_Deal = ",'None', Value_Deal) AS Value_Deal,
    Phone_Service,
    IF(Multiple_Lines = ",'No', Multiple_Lines) AS Multiple_Lines,
    Internet_Service,
    IF(Internet_Type = ",'None', Internet_Type) AS Internet_Type,
    IF(Online_Security = ",'No', Online_Security) AS Online_Security,
    IF(Online_Backup = ",'No', Online_Backup) AS Online_Backup,
    IF(Device_Protection_Plan = ",'No', Device_Protection_Plan) AS Device_Protection_Plan,
    IF(Premium_Support = ",'No', Premium_Support) AS Premium_Support,
    IF(Streaming_TV = ",'No', Streaming_TV) AS Streaming_TV,
    IF(Streaming_Movies = ",'No', Streaming_Movies) AS Streaming_Movies,
    IF(Streaming_Music = ",'No', Streaming_Music) AS Streaming_Music,
    IF(Unlimited_Data = ",'No', Unlimited_Data) AS Unlimited_Data,
    Contract,
    Paperless_Billing,
    Payment_Method,
    Monthly_Charge,
    Total_Charges,
    Total_Refunds,
    Total_Extra_Data_Charges,
    Total_Long_Distance_Charges,
    Total_Revenue,
    Customer_Status,
    IF(Churn_Category = ",'Others', Churn_Category) AS Churn_Category,
    IF(Churn_Reason = ",'Others', Churn_Reason) AS Churn_Reason
FROM stg_Churn;
```

Data Exploration – Check Distinct Values

We will be checking on how the data were distributed among Demographic and geographic info.

```
SELECT gender, Count(gender) as Totalcount, Count(Gender)*100.0/(select Count(*) from
prod_churn) as percentage from prod_churn Group by Gender;
```

```
SELECT contract, count(contract) as Totalcontract,
 count(contract)*100.0/( SELECT Count(*) from prod_churn) as percentage
 from prod_churn
 Group by contract;
```

Now the important column is `Customer_Status ,` where it will give details of which customer **stayed** and which customer **churned** from organization .

```
SELECT Customer_Status, Count(Customer_Status) as TotalCount, Sum(Total_Revenue) as
TotalRev,
Sum(Total_Revenue) / (SELECT sum(Total_Revenue) from prod_Churn) * 100  as RevPercentage
from prod_Churn
Group by Customer_Status

SELECT State, Count(State) as TotalCount,
Count(State) * 1.0 / (SELECT Count(*) from prod_Churn)  as Percentage
from prod_Churn
Group by State
Order by Percentage desc
```

Creating View for PowerBI:

```
create view  vw_churnData as
        select * from prod_churn where Customer_status IN ("Churned",'Stayed');

Create view vw_JoinData as
        select * from prod_churn where customer_status = 'Joined';
```

DATA LOAD & TRANSFORMATION IN POWERBI

1. Open Power BI desktop - GetData → More →Mysql Database → server : 127.0.0.1:3306 , Database : db_churn
2. Once Connected , transfer data for creating measures and calculations for further analysis.

Column Level Transformation:

New Columns

1. Churn Status = if [Customer_Status] = "Churned" then 1 else 0
2. Modify Churn Status data type to numbers
3. Monthly Charge status=
   if [Monthly_Charge] > 20 then "<20"
   else if [Monthly_Charge] < 20 then "21-50"
   else if [Monthly_Charge]<100 then "51-100" else ">100"

New Table for referencing AGEGroup

1. Instead of creating in existing Prod_churn table I have created new table which will reference to Prod table
2. Keep only Age column and remove duplicates
3. Age Group = if [Age] < 20 then "< 20" else if [Age] < 36 then "20 - 35" else if [Age] < 51 then "36 - 50" else "> 50"
4. AgeGrpSorting = if [Age Group] = "< 20" then 1 else if [Age Group] = "20 - 35" then 2 else if [Age Group] = "36 - 50" then 3 else 4
5. Change data type of AgeGrpSorting

New Table for referencing TenureGroup

1. Keep only Tenure_in_Months and remove duplicates

2. Tenure Group = if [Tenure_in_Months] < 6 then "< 6 Months" else if [Tenure_in_Months] < 12 then "6-12 Months" else if [Tenure_in_Months] < 18 then "12-18 Months" else if [Tenure_in_Months] < 24 then "18-24 Months" else ">= 24 Months"
3. TenureGrpSorting = if [TenureGrp] = "< 6 Months" then 1 else if [TenureGrp] = "6-12 Months" then 2 else if [TenureGrp] = "12-18 Months" then 3 else if [TenureGrp] = "18-24 Months " then 4 else 5
4. Change data type of TenureGrpSorting

(Note: If I create duplicate table the duplicate table the duplicate table will also hits the sql if I click on refresh which impacts the performance.**)**

Creating Measures :

1. Total Customers = Count(prod_Churn[Customer_ID])

2. New Joiners = CALCULATE(COUNT(prod_Churn[Customer_ID]), prod_Churn[Customer_Status] = "Joined")

3. Total Churn = SUM(prod_Churn[Churn Status])

4. Churn Rate = [Total Churn] / [Total Customers]

Geographic Analysis:

Since we have heavy list of state details, I will be keeping **top 5 churn** details for both state churn rate and churn rate distribution.

Service view:

1. The last view I want to show all the services in the data which have an entry of "Yes" and "No" somehow create a visual which can summarize all.
2. If I create a chart with this it's going to be a mess so I prefer using a grid on this one before I can use a grid I need to transform this data.
3. Then making a Group based on the entries .

Churn Reason Page (Tooltip)

1. Churn Reason – Total Churn