TABLEAU PROJECT

Telecom Customer Churn Analytics Dashboard

Project title: Telecom Customer Churn Analytics

Objective:

The primary objective of this analysis is to identify patterns and factors that influence customer churn and to provide actionable insights that help the business reduce churn.

1) **Identify customers who are churning**

   - Determine which customers have left the service and analyze churn patterns across demographics, tenure, and customer segments.

2) **Understand key drivers of customer churn**

   - Analyze the impact of contract type, tenure, services subscribed, payment methods, and pricing on customer churn.

3) **Assess the revenue impact of churn**

   - Quantify revenue loss caused by churned customers and identify high-value customers at risk.

4) **Segment customers by churn risk**

   - Classify customers into low, medium, and high churn-risk groups based on behavioral and service-related factors.

5) **Provide data-driven insights to reduce churn**

   - Recommend actionable strategies to improve customer retention and minimize revenue loss.

1. Introduction: Telecom Customer Churn

Customer churn refers to the phenomenon where customers discontinue their services with a telecom provider over a given period. In the highly competitive telecom industry, churn is a critical business challenge because acquiring a new customer is significantly more expensive than retaining an existing one.

Telecom companies offer similar pricing and services, making customer loyalty fragile. Even minor dissatisfaction related to pricing, service quality, contract flexibility, or customer support can result in customer loss. Therefore, understanding churn behavior is essential for improving retention strategies and sustaining revenue growth.

2. Business Impact of Churn

Churn directly affects a company's **revenue, profitability, and market share**. High churn rates indicate customer dissatisfaction and lead to:

- Loss of recurring revenue
- Increased customer acquisition costs
- Reduced customer lifetime value
- Negative brand perception

From a business perspective, reducing churn by even a small percentage can significantly improve overall profitability. Hence, churn analysis is not only a data problem but a **strategic business priority**.

3. Key Factors Influencing Churn in Telecom

Based on industry trends and historical data, customer churn in telecom is commonly influenced by:

- Contract Type: Month-to-month customers tend to churn more frequently than long-term contract customers
- Tenure: New customers are more likely to churn compared to long-standing customers
- Pricing: Higher monthly charges often increase churn probability
- Service Quality: Lack of technical support or frequent service issues increases dissatisfaction
- Payment Method: Customers using manual payment methods show higher churn behavior

These factors help businesses prioritize targeted retention strategies rather than generic offers.

4. Role of Data Analytics in Churn Reduction

Data analytics enables telecom companies to move from reactive to proactive churn management. By analyzing historical customer data, businesses can:

- Detect early warning signs of churn
- Predict high-risk customers
- Design personalized retention offers
- Improve customer experience through targeted interventions

Using tools such as **Python for data analysis**, **SQL for querying**, and **Power BI/Tableau for visualization**, churn insights can be communicated effectively to both technical and non-technical stakeholders.

5. Expected Business Outcomes

An effective telecom churn analysis helps organizations:

- Reduce customer attrition
- Improve customer lifetime value
- Optimize marketing and retention costs
- Strengthen long-term customer relationships

Ultimately, churn analysis supports informed decision-making and sustainable business growth.

About the Dataset

Customer churn is a major concern in the telecom industry, where competition is high and switching costs for customers are low. This dataset is designed to help analyze customer behavior and identify patterns that lead to churn. By understanding these patterns, businesses can develop targeted customer retention strategies instead of relying on generic offers.

The dataset is sourced from **IBM Sample Data Sets** and is widely used for customer churn analysis and predictive modeling.

Dataset Description

Each row in the dataset represents an individual telecom customer, while each column captures a specific attribute related to the customer’s demographics, services, account details, and churn status.

The dataset provides a comprehensive view of customer behavior by combining service usage, billing information, and personal characteristics.

Key Information Included

1. Churn Information

- Indicates whether a customer has discontinued the service within the last month
- Target variable used to analyze and predict customer churn

2. Services Subscribed

- Phone service and multiple lines
- Internet service type
- Online security, online backup, and device protection
- Technical support
- Streaming TV and streaming movies

3. Customer Account Details

- Customer tenure (length of time with the company)
- Contract type
- Payment method
- Paperless billing option
- Monthly charges and total charges

4. Customer Demographics

- Gender
- Senior citizen indicator (age group)
- Partner and dependent status

ARCHITECTURE

1. Data Source

The dataset used for this end-to-end project is the **Telco Customer Churn dataset**, originally provided by **IBM Sample Data Sets** and widely used for customer retention and churn analysis use cases.

The dataset contains detailed information about Telecom customers, where **each row represents an individual customer** and **each column represents a customer attribute**.

Dataset Characteristics

- Data Format: CSV (Comma-Separated Values)

- Data Type: Structured data
- Number of Records: ~7,000 customers
- Granularity: Customer-level data

Key Data Categories

1. **Customer Demographics:** Gender, Senior Citizen status, Partner, Dependents.
2. **Service Details:** Phone service, Internet service, Online security, Tech support, Streaming services.
3. **Account & Billing Information:** Contract type, Payment method, Monthly charges, Total charges, Tenure.
4. **Target Variable:** Churn (Yes / No) – indicates whether the customer left the service.

This dataset is suitable for analyzing **customer behavior**, **churn drivers**, and **revenue impact**, making it ideal for a real-world Telecom analytics project.

2. Data Extraction

Data extraction was performed using **Python and SQL**, ensuring efficient data ingestion and seamless integration with downstream analytics and visualization tools.

Extraction Process

- The raw CSV file was imported into the Python environment using the **Pandas** library.
- Initial data inspection was conducted to understand schema, data types, and missing values.
- Relevant columns were selected based on business relevance for churn and revenue analysis.
- Cleaned and transformed datasets were prepared for loading into a **SQL database** for further analysis.

Tools & Technologies Used

- **Python:** Pandas, NumPy
- **SQL:** For structured querying and KPI generation
- **Jupyter Notebook:** For data extraction and preprocessing workflow

3.DATA CLEANING USING PYTHON

1. Importing Required Libraries

Python libraries were imported to handle data manipulation and numerical operations efficiently.

```
import pandas as pd
import numpy as np
```

Pandas display settings were adjusted to improve readability during analysis.

2. Loading the Dataset

The raw CSV file was loaded into a Pandas DataFrame for inspection.

```
df = pd.read_csv("/mnt/data/WA_Fn-UseC_-Telco-Customer-Churn (1).csv")
```

Initial checks included dataset shape, column structure, and data types to identify obvious data quality issues.

3. Standardizing Column Names

Column names were standardized by:

- Converting to lowercase
- Removing extra spaces
- Replacing spaces with underscores

This ensures consistency and avoids errors during querying and visualization.

```
df.columns = df.columns.str.strip().str.lower().str.replace(" ", "_")
```

4. Removing Duplicate Records

Duplicate customer records can distort analysis and KPIs. These were removed to maintain data integrity.

```
df.drop_duplicates(inplace=True)
```

5. Handling Incorrect Data Types

The total charges column contained numeric values stored as strings with blank entries. These were converted to numeric values.

```
df['totalcharges'] = pd.to_numeric(df['totalcharges'], errors='coerce')
```

6. Handling Missing Values

Missing values were identified, and a **median imputation strategy** was used for totalcharges to avoid skewing the data.

```
df['totalcharges'].fillna(df['totalcharges'].median(), inplace=True)
```

This approach is robust against outliers and commonly used in real-world business datasets.

7. Fixing Inconsistent Categorical Values

Binary categorical columns such as Partner, Dependents, and Churn were standardized to consistent values (Yes/No).

```
df[col] = df[col].str.strip().str.capitalize()
```

8. Cleaning Categorical Features

Extra spaces in categorical columns were removed to prevent category duplication during analysis.

```
df[col] = df[col].str.strip()
```

9. Outlier Treatment

- Negative tenure values were removed based on business logic
- Monthly charges outliers were treated using the **IQR method**
- This step ensures more reliable statistical analysis.

10. Feature Engineering

A new feature, avg_monthly_spend, was created to better understand customer spending behavior.

```
df['avg_monthly_spend'] = df['totalcharges'] / (df['tenure'] + 1)
```

11. Encoding the Target Variable

The churn column was encoded into a binary format for analysis and future modeling.

```
df['churn_flag'] = df['churn'].map({'Yes': 1, 'No': 0})
```

12. Saving the Cleaned Dataset

The cleaned dataset was saved for EDA and dashboard development.

```
df.to_csv("telco_churn_cleaned.csv", index=False)
```

EXPLORATORY DATA ANALYSIS (EDA)

1. Import Libraries & Data Loading

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

pd.set_option('display.max_columns', None)

# Load cleaned dataset
df = pd.read_csv("/mnt/data/telco_churn_cleaned.csv")

print(df.shape)
df.head()
```

2. Data Understanding

Basic structural checks were performed to understand distributions, data types, and summary statistics.

```
df.info()
df.describe()

# Separate numerical & categorical columns
num_cols = df.select_dtypes(include=['int64', 'float64']).columns
cat_cols = df.select_dtypes(include='object').columns

print("Numerical Columns:", num_cols)
print("Categorical Columns:", cat_cols)
```

3. Target Variable Analysis (Churn)

The churn distribution revealed a **class imbalance**, which reflects real-world business scenarios.

Key Insight:

- A smaller percentage of customers churn, but they represent **significant revenue risk**

```
df['churn'].value_counts()
df['churn'].value_counts(normalize=True) * 100
# CHURN DISTRIBUTION
sns.countplot(x='churn', data=df)
plt.title("Churn Distribution")
plt.show()
```

4. Univariate Analysis – Numerical Features

Tenure

- Most churned customers have low tenure
- Long-tenure customers show higher loyalty

```
plt.figure(figsize=(6,4))

sns.histplot(df['tenure'], bins=30, kde=True)

plt.title("Tenure Distribution")

plt.show()
```

Monthly Charges

- Customers with higher monthly charges tend to churn more

```
plt.figure(figsize=(6,4))

sns.histplot(df['monthlycharges'], bins=30, kde=True)

plt.title("Monthly Charges Distribution")

plt.show()
```

Total Charges

- Higher total charges correlate with longer customer relationships

```
plt.figure(figsize=(6,4))

sns.histplot(df['totalcharges'], bins=30, kde=True)

plt.title("Total Charges Distribution")

plt.show()
```

4. Univariate Analysis – Categorical Features

Categorical distributions helped identify dominant customer segments and rare categories, which are critical for segmentation analysis.

```
for col in cat_cols:
    plt.figure(figsize=(6,3))
    sns.countplot(y=col, data=df)
    plt.title(f"Distribution of {col}")
    plt.show()
```

5. Bivariate Analysis – Numerical vs Churn

Tenure vs Churn

- Customers with shorter tenure are significantly more likely to churn

```
plt.figure(figsize=(6,4))
sns.boxplot(x='churn', y='tenure', data=df)
plt.title("Tenure vs Churn")
plt.show()
```

Monthly Charges vs Churn

- High monthly charges increase churn probability

```
plt.figure(figsize=(6,4))
sns.boxplot(x='churn', y='monthlycharges', data=df)
plt.title("Monthly Charges vs Churn")
plt.show()
```

6. Bivariate Analysis – Categorical vs Churn

Churn rates were analyzed across contract types, payment methods, and services.

Key Findings:

- Month-to-month contracts have the highest churn
- Electronic check payment method shows higher churn risk

```
for col in cat_cols:

    if col != 'churn':

        churn_rate = pd.crosstab(df[col], df['churn'], normalize='index') * 100

        churn_rate.plot(kind='bar', stacked=True, figsize=(6,4))

        plt.title(f"Churn Rate by {col}")

        plt.ylabel("Percentage")

        plt.show()
```

7. Multivariate Analysis – Correlation

Correlation analysis confirmed strong relationships:

- Tenure and Total Charges (strong positive correlation)
- Monthly Charges and Churn (moderate correlation)

```
plt.figure(figsize=(10,6))

sns.heatmap(df[num_cols].corr(), annot=True, cmap='coolwarm')

plt.title("Correlation Heatmap")

plt.show()
```

These relationships align with telecom business expectations.

8. Key Business Metrics

Overall Churn Rate calculated using churn flag

 # Overall churn rate

churn_rate = df['churn_flag'].mean() * 100

print(f"Overall Churn Rate: {churn_rate:.2f}%")

 # Average metrics by churn

df.groupby('churn')[['tenure', 'monthlycharges', 'totalcharges']].mean()

- Comparison of tenure, monthly charges, and total charges by churn status

This quantifies the financial and behavioral differences between churned and retained customers.

Data Cleaning & Preparation

The dataset used for this analysis was cleaned and preprocessed using **Python in a Jupyter Notebook**.
The data cleaning process included handling missing values, correcting data types, removing inconsistencies, and validating key fields to ensure data accuracy and reliability for analysis.

The complete data cleaning workflow and code implementation are available at the link below:
https://drive.google.com/file/d/14CGhleVqPueN0lj7aKaPgneKiKYR9QgZ/view?usp=drive_link

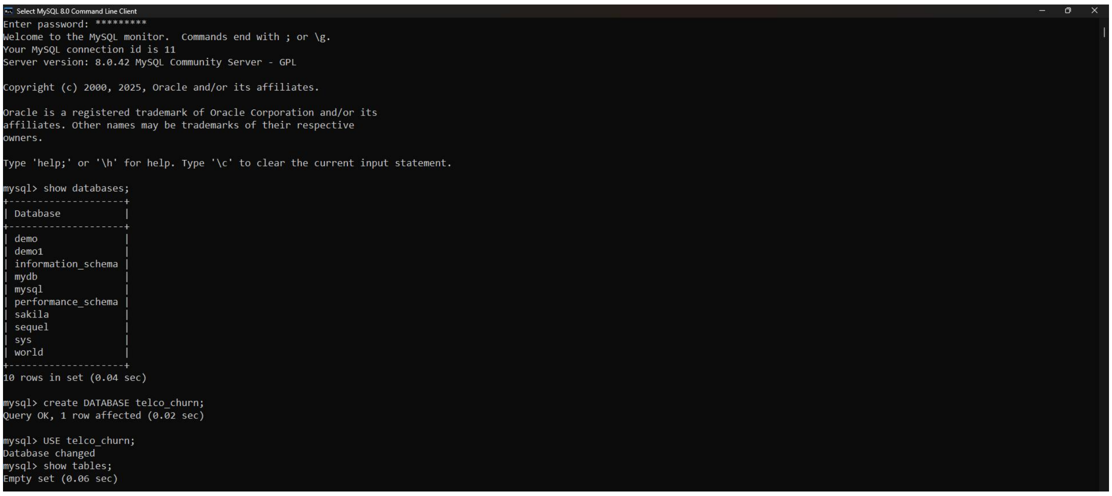
3. SQL

DATABASE & TABLE SETUP

A dedicated database was created to store the cleaned churn data produced from the Python pipeline.

CREATE DATABASE telco_churn;

USE telco_churn;



The table schema mirrors the cleaned dataset structure to ensure consistency across Python, SQL, and BI layers.

CREATE TABLE telco_churn_cleaned (

customerid VARCHAR(50),

gender VARCHAR(10),

seniorcitizen INT,

partner VARCHAR(5),

dependents VARCHAR(5),

tenure INT,

phoneservice VARCHAR(5),

multiplelines VARCHAR(50),

internetservice VARCHAR(50),

onlinesecurity VARCHAR(50),

onlinebackup VARCHAR(50),

deviceprotection VARCHAR(50),

techsupport VARCHAR(50),

streamingtv VARCHAR(50),

streamingmovies VARCHAR(50),

contract VARCHAR(50),

paperlessbilling VARCHAR(5),

paymentmethod VARCHAR(50),

monthlycharges DECIMAL(10,2),

totalcharges DECIMAL(10,2),

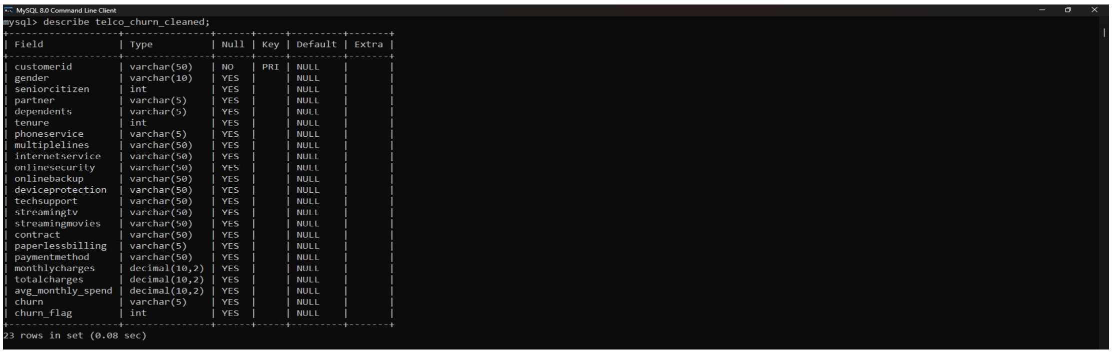
avg_monthly_spend DECIMAL(10,2),

churn VARCHAR(5),

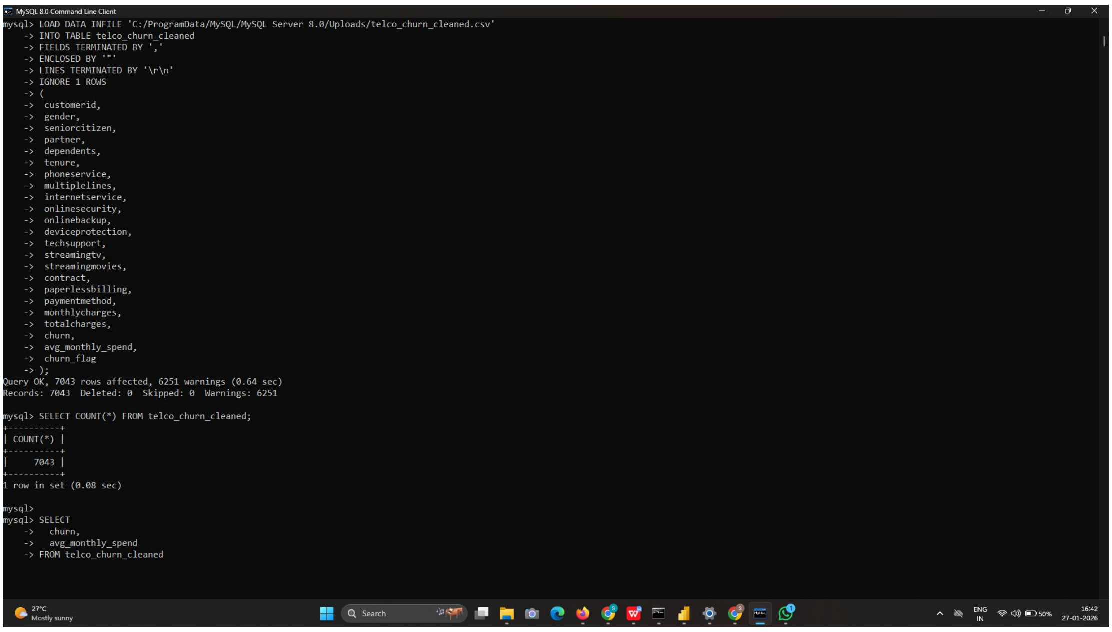
churn_flag INT

);

Data was loaded using CSV import tools such as LOAD DATA INFILE or MySQL Import Wizard.



```
MySQL 8.0 Command Line Client
mysql> describe telco_churn_cleaned;
+-------------------+---------------+------+-----+---------+-------+
| Field             | Type          | Null | Key | Default | Extra |
+-------------------+---------------+------+-----+---------+-------+
| customerid        | varchar(50)   | NO   | PRI | NULL    |       |
| gender            | varchar(10)   | YES  |     | NULL    |       |
| seniorcitizen     | int           | YES  |     | NULL    |       |
| partner           | varchar(5)    | YES  |     | NULL    |       |
| dependents        | varchar(5)    | YES  |     | NULL    |       |
| tenure            | int           | YES  |     | NULL    |       |
| phoneservice      | varchar(5)    | YES  |     | NULL    |       |
| multiplelines     | varchar(50)   | YES  |     | NULL    |       |
| internetservice   | varchar(50)   | YES  |     | NULL    |       |
| onlinesecurity    | varchar(50)   | YES  |     | NULL    |       |
| onlinebackup      | varchar(50)   | YES  |     | NULL    |       |
| deviceprotection  | varchar(50)   | YES  |     | NULL    |       |
| techsupport       | varchar(50)   | YES  |     | NULL    |       |
| streamingtv       | varchar(50)   | YES  |     | NULL    |       |
| streamingmovies   | varchar(50)   | YES  |     | NULL    |       |
| contract          | varchar(50)   | YES  |     | NULL    |       |
| paperlessbilling  | varchar(5)    | YES  |     | NULL    |       |
| paymentmethod     | varchar(50)   | YES  |     | NULL    |       |
| monthlycharges    | decimal(10,2) | YES  |     | NULL    |       |
| totalcharges      | decimal(10,2) | YES  |     | NULL    |       |
| avg_monthly_spend | decimal(10,2) | YES  |     | NULL    |       |
| churn             | varchar(5)    | YES  |     | NULL    |       |
| churn_flag        | int           | YES  |     | NULL    |       |
+-------------------+---------------+------+-----+---------+-------+
23 rows in set (0.08 sec)
```



```
MySQL 8.0 Command Line Client
mysql> LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/telco_churn_cleaned.csv'
    -> INTO TABLE telco_churn_cleaned
    -> FIELDS TERMINATED BY ','
    -> ENCLOSED BY '"'
    -> LINES TERMINATED BY '\r\n'
    -> IGNORE 1 ROWS
    -> (
    ->  customerid,
    ->  gender,
    ->  seniorcitizen,
    ->  partner,
    ->  dependents,
    ->  tenure,
    ->  phoneservice,
    ->  multiplelines,
    ->  internetservice,
    ->  onlinesecurity,
    ->  onlinebackup,
    ->  deviceprotection,
    ->  techsupport,
    ->  streamingtv,
    ->  streamingmovies,
    ->  contract,
    ->  paperlessbilling,
    ->  paymentmethod,
    ->  monthlycharges,
    ->  totalcharges,
    ->  churn,
    ->  avg_monthly_spend,
    ->  churn_flag
    -> );
Query OK, 7043 rows affected, 6251 warnings (0.64 sec)
Records: 7043  Deleted: 0  Skipped: 0  Warnings: 6251

mysql> SELECT COUNT(*) FROM telco_churn_cleaned;
+----------+
| COUNT(*) |
+----------+
|     7043 |
+----------+
1 row in set (0.08 sec)

mysql>
mysql> SELECT
    ->   churn,
    ->   avg_monthly_spend
    -> FROM telco_churn_cleaned
```


```
mysql> SELECT COUNT(*) FROM telco_churn_cleaned;
+----------+
| COUNT(*) |
+----------+
|     7043 |
+----------+
1 row in set (0.08 sec)

mysql>
mysql> SELECT
    ->   churn,
    ->   avg_monthly_spend
    -> FROM telco_churn_cleaned
    -> LIMIT 10;
+-------+-------------------+
| churn | avg_monthly_spend |
+-------+-------------------+
| No    |             59.33 |
| No    |             54.24 |
| Yes   |             56.17 |
| Yes   |             88.42 |
| Yes   |             66.85 |
| No    |             57.15 |
| No    |            109.78 |
| No    |             84.03 |
| No    |             42.54 |
| No    |             90.27 |
+-------+-------------------+
10 rows in set (0.01 sec)
```

Step 1: Overall Customer Overview

We already created a KPI view (vw_kpis), which gives a snapshot of total customers, churn rate, total revenue, and revenue lost due to churn.

Query:

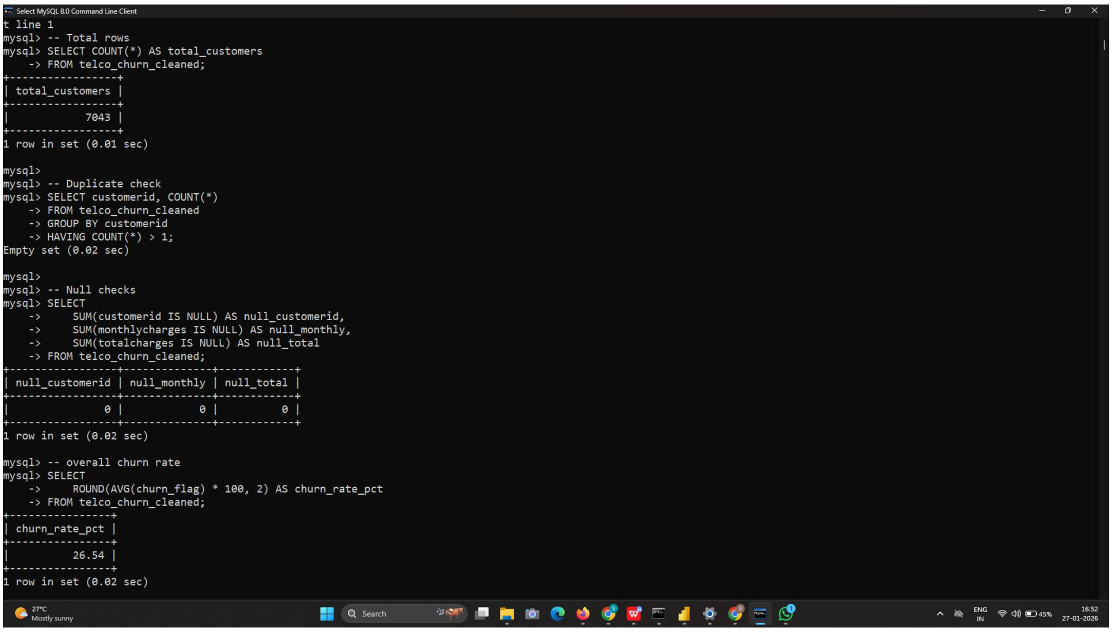
SELECT * FROM vw_kpis;

Sample Output:

total_customers	churn_rate_pct	total_revenue	churn_revenue_loss
7043	26.54	16,071,540.98	2,862,926.90

Observation 1:

- Overall **churn rate** is **26.54%**, which means roughly **1 in 4 customers leave**.
- **Revenue lost due to churn** is ~**2.86M**, significant compared to total revenue.



Step 2: Churn by Contract Type

Analyzing which contract types have the highest churn can guide retention strategies.

Query:

SELECT

contract,

COUNT(*) AS total_customers,

SUM(churn_flag) AS churned_customers,

ROUND(AVG(churn_flag) * 100, 2) AS churn_rate_pct

FROM telco_churn_cleaned

GROUP BY contract

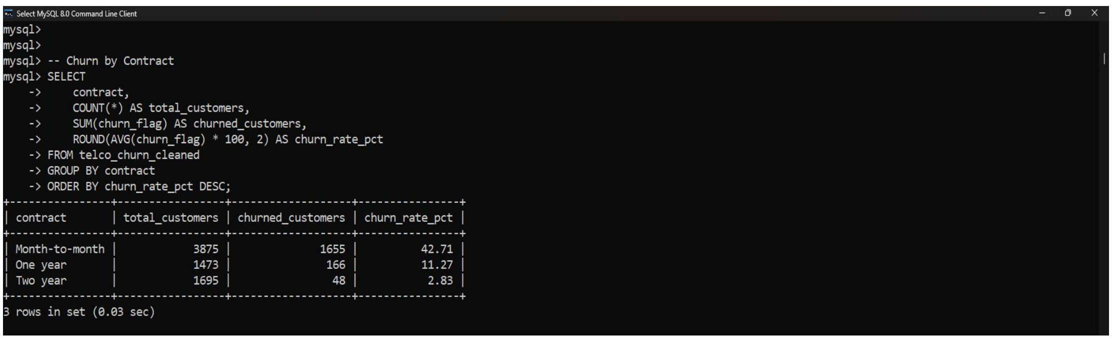
ORDER BY churn_rate_pct DESC;

Sample Output:

contract	**total_customers**	**churned_customers**	**churn_rate(%)**
Month-to-month	3875	1655	42.71
One year	1473	166	11.27
Two year	1695	48	2.83

Observation 2:

- **Month-to-month contracts** have the **highest churn rate** (**42.71%**).
- **Longer-term contracts (1 or 2 years)** have much **lower churn**, showing retention improves with contract length.



Step 3: Churn by Payment Method

Identifying risky payment methods helps target customers for proactive engagement.

Query:

```
SELECT
    paymentmethod,
    COUNT(*) AS customers,
    ROUND(AVG(churn_flag) * 100, 2) AS churn_rate_pct
FROM telco_churn_cleaned
GROUP BY paymentmethod
ORDER BY churn_rate_pct DESC;
```

Sample Output:

paymentmethod	customers	churn_rate_pct
Electronic check	2365	45.29
Mailed check	1612	19.11
Bank transfer (automatic)	1544	16.71
Credit card (automatic)	1522	15.24

Observation 3:

Customers using **Electronic check** have the **highest churn rate (45.29%)**.

Automated payments (credit card or bank transfer) correlate with **lower churn**, likely due to convenience.

```
mysql> -- Churn by Payment Method
mysql> SELECT
    ->     paymentmethod,
    ->     COUNT(*) AS customers,
    ->     ROUND(AVG(churn_flag) * 100, 2) AS churn_rate_pct
    -> FROM telco_churn_cleaned
    -> GROUP BY paymentmethod
    -> ORDER BY churn_rate_pct DESC;
+---------------------------+-----------+----------------+
| paymentmethod             | customers | churn_rate_pct |
+---------------------------+-----------+----------------+
| Electronic check          |      2365 |          45.29 |
| Mailed check              |      1612 |          19.11 |
| Bank transfer (automatic) |      1544 |          16.71 |
| Credit card (automatic)   |      1522 |          15.24 |
+---------------------------+-----------+----------------+
4 rows in set (0.01 sec)
```

Step 4: Tenure Analysis

Understanding churn across customer tenure can highlight **early churn risks**.

Query:

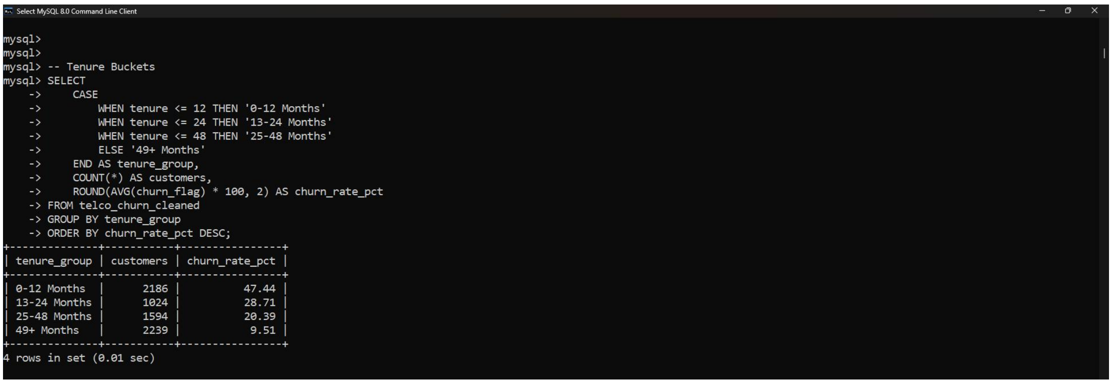
```
SELECT
   CASE
      WHEN tenure <= 12 THEN '0-12 Months'
      WHEN tenure <= 24 THEN '13-24 Months'
      WHEN tenure <= 48 THEN '25-48 Months'
      ELSE '49+ Months'
   END AS tenure_group,
   COUNT(*) AS customers,
   ROUND(AVG(churn_flag) * 100, 2) AS churn_rate_pct
FROM telco_churn_cleaned
GROUP BY tenure_group
ORDER BY churn_rate_pct DESC;
```

Sample Output:

tenure_group	customers	churn_rate_pct
0-12 Months	2186	47.44
13-24 Months	1024	28.71
25-48 Months	1594	20.39
49+ Months	2239	9.51

Observation 4:

- **Highest churn occurs in the first year (0-12 months)** at **47.44%**, confirming early-stage churn is critical.
- Retention efforts should **focus on new customers**.



```
Select MySQL 8.0 Command Line Client

mysql>
mysql>
mysql> -- Tenure Buckets
mysql> SELECT
    ->     CASE
    ->         WHEN tenure <= 12 THEN '0-12 Months'
    ->         WHEN tenure <= 24 THEN '13-24 Months'
    ->         WHEN tenure <= 48 THEN '25-48 Months'
    ->         ELSE '49+ Months'
    ->     END AS tenure_group,
    ->     COUNT(*) AS customers,
    ->     ROUND(AVG(churn_flag) * 100, 2) AS churn_rate_pct
    -> FROM telco_churn_cleaned
    -> GROUP BY tenure_group
    -> ORDER BY churn_rate_pct DESC;
+--------------+-----------+----------------+
| tenure_group | customers | churn_rate_pct |
+--------------+-----------+----------------+
| 0-12 Months  |      2186 |          47.44 |
| 13-24 Months |      1024 |          28.71 |
| 25-48 Months |      1594 |          20.39 |
| 49+ Months   |      2239 |           9.51 |
+--------------+-----------+----------------+
4 rows in set (0.01 sec)
```

Step 5: High-Value Customer Loss

Identifying **top-paying churned customers** helps prioritize retention strategies.

Query:

SELECT

customerid,

totalcharges,

monthlycharges,

tenure

FROM telco_churn_cleaned

WHERE churn = 'Yes'

ORDER BY totalcharges DESC

LIMIT 20;

Observation 5:

- Some churned customers have **total charges > 8,000**, representing **high revenue loss**.
- Targeted retention campaigns (discounts, loyalty programs) could protect revenue.

```
mysql> -- High Value Customer Loss
mysql> SELECT
    ->     customerid,
    ->     totalcharges,
    ->     monthlycharges,
    ->     tenure
    -> FROM telco_churn_cleaned
    -> WHERE churn = 'Yes'
    -> ORDER BY totalcharges DESC
    -> LIMIT 20;
+------------+--------------+----------------+--------+
| customerid | totalcharges | monthlycharges | tenure |
+------------+--------------+----------------+--------+
| 2889-FPWRM |      8684.80 |         117.80 |     72 |
| 0201-OAMXR |      8127.60 |         115.55 |     70 |
| 3886-CERTZ |      8109.80 |         109.25 |     72 |
| 1444-VVSGW |      7968.85 |         115.65 |     70 |
| 5271-YNWVR |      7856.00 |         113.15 |     68 |
| 8199-ZLLSA |      7804.15 |         118.35 |     67 |
| 9053-JZFKV |      7752.30 |         116.20 |     67 |
| 1555-DJEQW |      7723.90 |         114.20 |     70 |
| 3259-FDWOY |      7723.70 |         106.00 |     71 |
| 7317-GGVPB |      7690.90 |         108.60 |     71 |
| 0917-EZOLA |      7689.95 |         104.15 |     72 |
| 1984-FCOWB |      7674.55 |         109.50 |     70 |
| 3192-NQECA |      7611.85 |         110.00 |     68 |
| 5287-QWLKY |      7548.10 |         105.10 |     71 |
| 0748-RDGGM |      7534.65 |         109.50 |     70 |
| 2834-JRTUA |      7532.15 |         108.05 |     71 |
| 9835-ZIITK |      7491.75 |         110.85 |     66 |
| 2659-VXMWZ |      7482.10 |         111.30 |     67 |
| 5502-RLUYV |      7446.90 |         103.95 |     69 |
| 7632-MNYOY |      7432.05 |         110.90 |     66 |
+------------+--------------+----------------+--------+
20 rows in set (0.01 sec)
```

STEP:6 Tech Support vs Churn

To analyze whether **technical support availability** impacts customer churn.Tech support is a known **retention driver** in telecom businesses.Helps identify service gaps leading to churn.

Query

SELECT
techsupport,
COUNT(*) AS customers,
ROUND(AVG(churn_flag) * 100, 2) AS churn_rate_pctFROM telco_churn_cleanedGROUP BY techsupport;

Observation 6:

- Customers **without Tech Support** show a **significantly higher churn rate** compared to those with support.
- Customers with **active tech support** are more likely to stay, indicating that service quality plays a critical role in retention.

```
Select MySQL 8.0 Command Line Client
mysql>
mysql> -- Tech Support vs Churn
mysql> SELECT
    ->     techsupport,
    ->     COUNT(*) AS customers,
    ->     ROUND(AVG(churn_flag) * 100, 2) AS churn_rate_pct
    -> FROM telco_churn_cleaned
    -> GROUP BY techsupport;
+---------------------+-----------+----------------+
| techsupport         | customers | churn_rate_pct |
+---------------------+-----------+----------------+
| Yes                 |      2044 |          15.17 |
| No                  |      3473 |          41.64 |
| No internet service |      1526 |           7.40 |
+---------------------+-----------+----------------+
3 rows in set (0.01 sec)
```

STEP :7 Revenue Loss Due to Churn

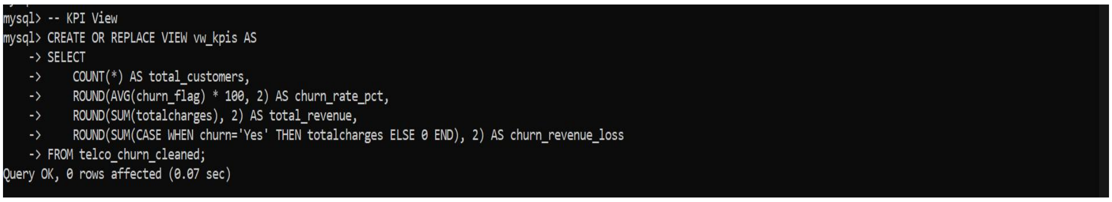
To quantify **financial impact** of churn.Converts churn from a **customer problem** into a **revenue problem**.Helps management prioritize churn reduction strategies.

Query

CREATE OR REPLACE VIEW vw_kpis ASSELECT
COUNT(*) AS total_customers,
ROUND(SUM(churn_flag) * 100 / COUNT(*), 2) AS churn_rate_pct,
ROUND(SUM(CASE WHEN churn_flag = 1 THEN totalcharges ELSE 0 END), 2)
AS churn_revenue_lossFROM telco_churn_cleaned;

Observation 7:

- Churned customers contribute to a **substantial revenue loss**.
- Even a small reduction in churn can result in **significant revenue recovery**.



```
mysql> -- KPI View
mysql> CREATE OR REPLACE VIEW vw_kpis AS
    -> SELECT
    ->     COUNT(*) AS total_customers,
    ->     ROUND(AVG(churn_flag) * 100, 2) AS churn_rate_pct,
    ->     ROUND(SUM(totalcharges), 2) AS total_revenue,
    ->     ROUND(SUM(CASE WHEN churn='Yes' THEN totalcharges ELSE 0 END), 2) AS churn_revenue_loss
    -> FROM telco_churn_cleaned;
Query OK, 0 rows affected (0.07 sec)
```

STEP :8 Contract Type vs Churn

To understand how **contract duration** influences churn behavior.Contract type directly reflects customer commitment level.

Query

CREATE OR REPLACE VIEW vw_contract_churn ASSELECT
contract,
COUNT(*) AS customers,
ROUND(AVG(churn_flag) * 100, 2) AS churn_rate_pctFROM telco_churn_cleanedGROUP BY contract;

Observation 8:

- **Month-to-month contracts** have the **highest churn rate**.
- **One-year and two-year contracts** show much lower churn.

```
mysql> -- Segmentation View
mysql> CREATE OR REPLACE VIEW vw_contract_churn AS
    -> SELECT
    ->     contract,
    ->     paymentmethod,
    ->     COUNT(*) AS customers,
    ->     ROUND(AVG(churn_flag) * 100, 2) AS churn_rate_pct
    -> FROM telco_churn_cleaned
    -> GROUP BY contract, paymentmethod;
Query OK, 0 rows affected (0.01 sec)
```

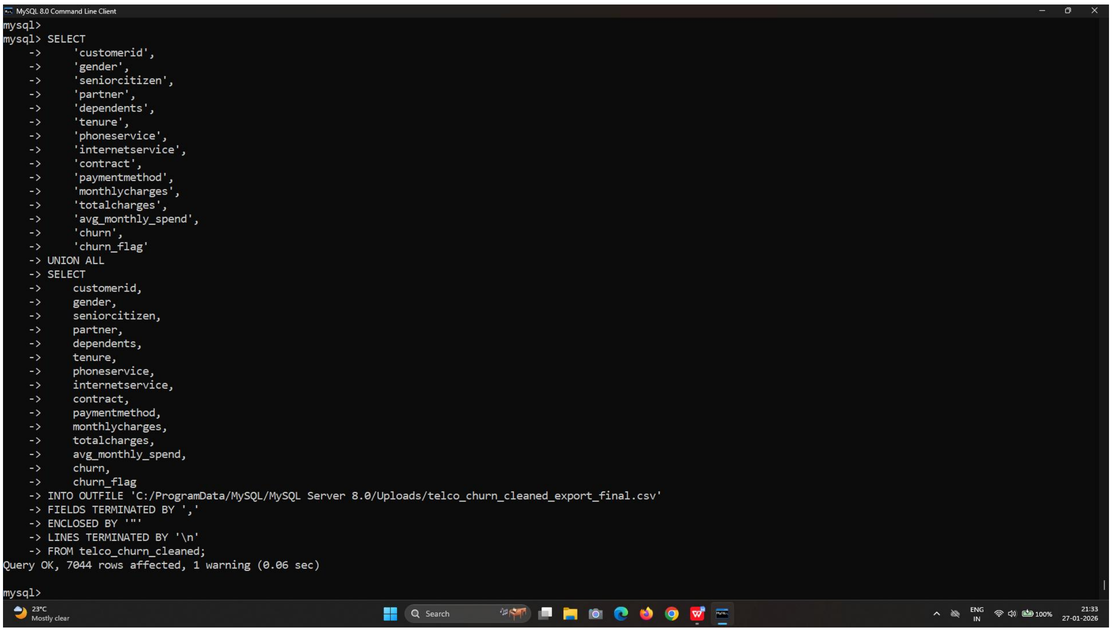
Dataset Overview

- Total Customers: **7,043**
- Total Revenue (all customers): **₹16,071,540.98**

- Overall Churn Rate: **26.54%**
- Revenue Lost Due to Churn: **₹2,862,926.90**

The dataset is clean: no duplicate customerids, no null values in key numerical columns (monthlycharges, totalcharges).


```
mysql> SELECT *
    -> INTO OUTFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/telco_churn_cleaned_export.csv'
    -> FIELDS TERMINATED BY ','
    -> ENCLOSED BY '"'
    -> LINES TERMINATED BY '\n'
    -> FROM telco_churn_cleaned;
Query OK, 7043 rows affected (0.12 sec)
```


```
mysql>
mysql> SELECT
    ->     'customerid',
    ->     'gender',
    ->     'seniorcitizen',
    ->     'partner',
    ->     'dependents',
    ->     'tenure',
    ->     'phoneservice',
    ->     'internetservice',
    ->     'contract',
    ->     'paymentmethod',
    ->     'monthlycharges',
    ->     'totalcharges',
    ->     'avg_monthly_spend',
    ->     'churn',
    ->     'churn_flag'
    -> UNION ALL
    -> SELECT
    ->     customerid,
    ->     gender,
    ->     seniorcitizen,
    ->     partner,
    ->     dependents,
    ->     tenure,
    ->     phoneservice,
    ->     internetservice,
    ->     contract,
    ->     paymentmethod,
    ->     monthlycharges,
    ->     totalcharges,
    ->     avg_monthly_spend,
    ->     churn,
    ->     churn_flag
    -> INTO OUTFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/telco_churn_cleaned_export_final.csv'
    -> FIELDS TERMINATED BY ','
    -> ENCLOSED BY '"'
    -> LINES TERMINATED BY '\n'
    -> FROM telco_churn_cleaned;
Query OK, 7044 rows affected, 1 warning (0.06 sec)

mysql>
```

Final Data Export for BI & Reporting

Query

SELECT
customerid,
gender,
seniorcitizen,
partner,
dependents,
tenure,
phoneservice,
internetservice,
contract,
paymentmethod,
monthlycharges,
totalcharges,
churn,

churn_flagFROM telco_churn_cleanedINTO OUTFILE'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/telco_churn_cleaned_export_final.csv'
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n';

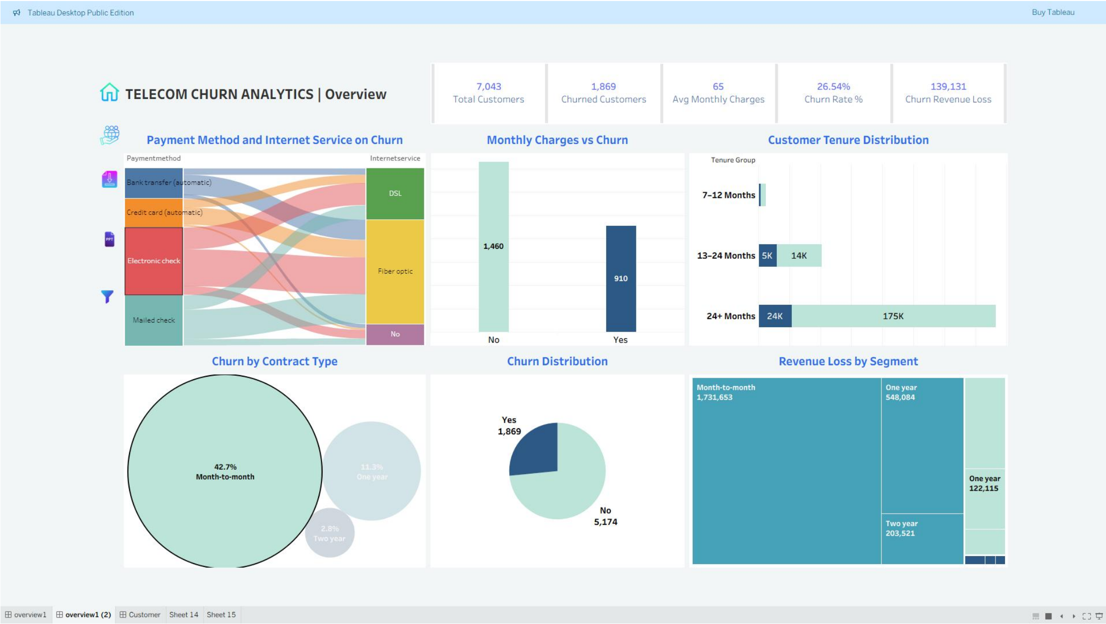
Why This Query Is Used

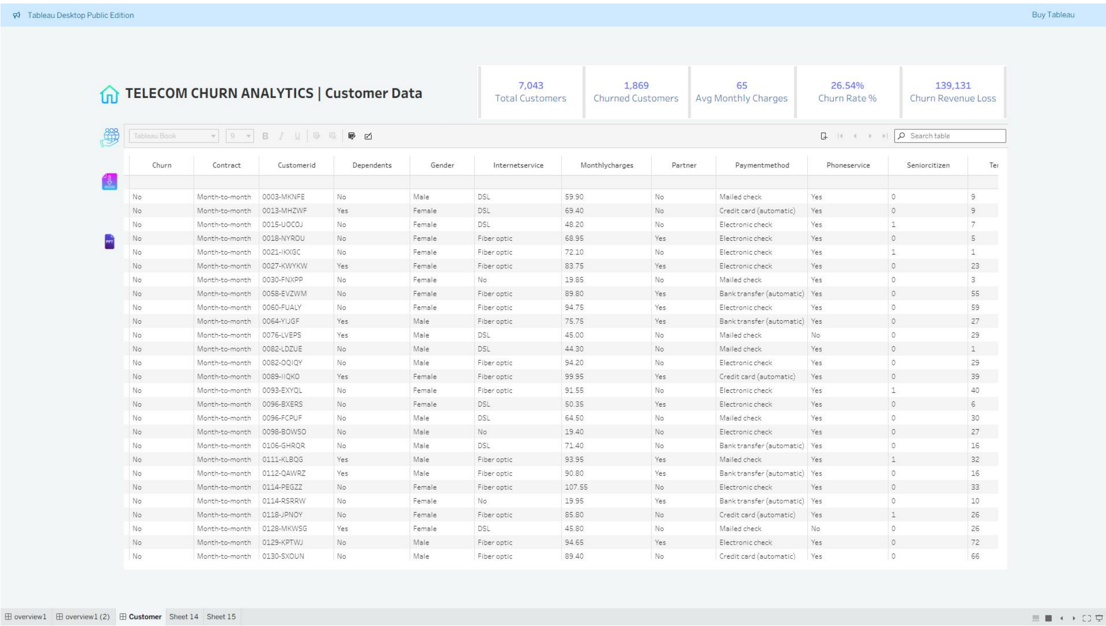
- To **export a fully cleaned and structured dataset** from MySQL.
- Acts as the **bridge between SQL and visualization tools** like Power BI or Tableau.
- Ensures only **business-relevant columns** are passed forward (no raw noise).
- Standard CSV formatting allows **easy ingestion into BI tools** without errors.
- Includes `churn_flag`, enabling direct KPI calculations and segmentation in dashboards.

Observation

- The dataset is now **analytics-ready** with:
  - Customer demographics
  - Service and contract details
  - Pricing information
  - Clear churn indicators (`churn`, `churn_flag`)
- No missing critical fields or inconsistent data types.
- Enables fast creation of:
  - Churn rate KPIs
  - Revenue loss metrics
  - Customer risk segmentation
  - Trend and comparison visuals

4. TABLEAU





1. Overall KPI Summary – Observation

- The dataset contains **7,043 total customers**, out of which **1,869 customers have churned**, resulting in a **churn rate of 26.54%**.
- The **average monthly charge is 65**, indicating a moderate pricing range across customers.

- Churn has led to a **total revenue loss of 139,131**, highlighting a significant financial impact.
- These KPIs indicate that **customer retention is a critical business concern** for the telecom company.

2. Payment Method & Internet Service vs Churn – Observation

- Customers using **Electronic Check** show the **highest churn contribution**, especially among **Fiber Optic users**.
- **Automatic payment methods** (Credit Card & Bank Transfer) are associated with **lower churn**, indicating better customer stability.
- **Fiber Optic customers** contribute more to churn compared to DSL, suggesting service-related dissatisfaction or pricing sensitivity.
- The interaction between **payment behavior and internet service type** plays a key role in predicting churn.

3. Monthly Charges vs Churn – Observation

- Customers with **higher monthly charges** are more likely to churn compared to those with lower charges.
- A noticeable difference exists between churned and retained customers, indicating **price sensitivity**.
- This suggests that **premium-priced customers require proactive retention strategies**.
- Pricing optimization or targeted discounts could help reduce churn among high-paying users.

4. Churn Distribution – Observation

- A majority of customers (**5,174**) have not churned, while **1,869 customers** have churned.
- Despite non-churned customers being higher in number, the **churn proportion is still significant**.
- The churned segment represents over **one-fourth of the customer base**.
- This emphasizes the importance of **early churn detection and prevention**.

5. Churn by Contract Type – Observation

- **Month-to-month contracts account for the highest churn (42.7%)**, making them the most unstable segment.
- **One-year contracts** show considerably lower churn, while **two-year contracts** have the least churn (2.8%).
- Longer contract durations are strongly associated with **higher customer retention**.
- Encouraging customers to move to **long-term contracts** can significantly reduce churn.

6. customer Tenure Distribution – Observation

- Customers with **shorter tenure (0–12 months)** show **higher churn concentration**.
- As customer tenure increases, the churn rate **gradually decreases**, indicating improved loyalty over time.
- Long-tenure customers form a stable revenue base with minimal churn risk.
- This highlights the need for **strong onboarding and early engagement strategies**.

7. Revenue Loss by Segment – Observation

- **Month-to-month contract customers contribute the highest revenue loss**, making them a high-risk segment.
- Customers on **one-year and two-year contracts** show comparatively lower revenue loss.
- Revenue loss is concentrated in **short-term contract and high-churn segments**.
- Targeting these segments with retention offers can help **recover lost revenue**.

Future Scope

- Customers on **month-to-month contracts** show higher churn, so future strategies can focus on **encouraging long-term contracts** to improve retention.
- **Electronic check users** have a higher tendency to churn; promoting **automatic payment methods** can help reduce churn.
- **Fiber optic customers** contribute significantly to churn, indicating the need for **service quality improvement and pricing optimization**.

- Customers with **lower tenure (0–12 months)** are more likely to churn, highlighting the importance of **strong onboarding and early engagement programs**.
- **High monthly charge customers** result in greater revenue loss when they churn, so **targeted retention offers** can be implemented.
- In the future, **predictive churn models** can be developed to proactively identify at-risk customers and take preventive actions.