

POWER BI

POWER QUERY – ETL TOOL

Dataset Source - <https://www.kaggle.com/datasets/vasanthgx/dirty-metro-interstate-traffic-volume-dataset/data>

Aim - The goal of this project was to learn and practice Power Query by working with real-world data. The focus was on understanding and exploring the different features Power Query offers for data cleaning and transformation. After preparing the dataset, it was loaded into Power BI to create interactive and insightful dashboards. The project also included analyzing the data and performing forecasting, with all findings and results compiled into a detailed report at the end. This project served as a hands-on learning experience for both data preparation and visualization.

Explanation of Columns:

1. **traffic_volume:**
Represents the count of vehicles at a particular time. It's the primary metric for analyzing traffic patterns.
2. **holiday:**
Indicates whether the data point corresponds to a public holiday or not. It may contain specific holiday names or None.
3. **temp:**
Temperature recorded in Kelvin. Influences traffic patterns as weather conditions affect commute behaviour.
4. **rain_1h:**
Amount of rainfall in the last hour (in mm). Higher rainfall might correlate with reduced traffic volume or slower movement.
5. **snow_1h:**
Amount of snowfall in the last hour (in mm). Like rainfall, it affects road conditions and traffic flow.
6. **clouds_all:**
Percentage of cloud cover, ranging from 0 (clear sky) to 100 (completely overcast).
7. **weather_main:**
Broad classification of weather (e.g., Mist, Clouds, Drizzle). Useful for grouping and understanding traffic in different weather conditions.
8. **weather_description:**
Detailed description of the weather (e.g., overcast clouds, haze). Offers more granularity than weather_main.
9. **date_time:**
Date and time of the observation. Useful for time-based analysis like identifying peak hours, daily patterns, or seasonal trends.

Data Cleaning and Transformation (ETL Process)

The dataset underwent extensive cleaning and transformation to ensure reliability for visualization and insights. Key steps include:

1. Handling Missing Values:

- Replaced missing values in holiday with "No Holiday."
- Filled null values in temp using the daily average temperature.
 - Duplicated the dataset, Grouped Average Temperature with Date column on new dataset
 - Merged two datasets "Outer Join" with date column as foreign Key
 - Added conditional column to fill null values in temp with Average temperature of that day
 - Drop Original temperature and group by temperature column
 - Still data contains around 8 null values, Filtered data excluding null values.
- Replaced blanks and nulls in weather columns with meaningful descriptions based on existing data.
 - Used Custom Column and created a new column 'Weather' having weather main values. For null data appended values based on weather description column
 - Lastly deleted Weather main and weather description columns
- Set null or blank traffic volume to 0 and null Rain values to 0.

2. Data Enrichment:

- Added a Temp (°C) column by converting Kelvin to Celsius [temperature - 273.15]
- Split date_time into:
 - **Date:** To analyse traffic patterns by day.
 - **Hour:** To identify rush hours.
 - **Day of Week:** To observe weekday vs. weekend traffic.
- Created Conditional Column to indicate Traffic volume as Low, Moderate and High
- Defined time periods (Night, Morning, Afternoon, Evening).
 - Night (0 AM – 6 AM).
 - Morning (6 AM – 12 PM)
 - Afternoon (13 PM – 18 PM)
 - Evening (18 PM – 24 PM)

3. Optimization and Cleaning:

- Removed duplicates, unnecessary columns (date_time and Snow_1h), and records with excessive nulls.

Insights from Visualizations

Hourly Traffic Patterns

- **4 PM** recorded the highest traffic volume, contributing **41%** of the total.
- Traffic volumes peaked during **4 PM** and were lowest at **3 AM**.
- The range of traffic volumes across all hours was **1,224 to 1,411**.

Weekly Traffic Patterns

- Saturday and Sunday (**Weekends**) records low traffic volume in comparison to weekdays.
- Thursday accounted for **15.85%** of total Traffic volume

Weather and Traffic Volume

- **Cloudy weather** accounted for the highest traffic volume (31,891,757), representing **34.46%** of total traffic.
- **Squall** had the least traffic volume (420).
- A total of **11 weather types** were analyzed, with traffic volumes ranging significantly between weather conditions.

Holiday Traffic Trends

- **"No Holiday"** days dominated traffic, accounting for **99.90%** of total traffic volume (31,510 counts).
- Other holidays, such as **Martin Luther King Jr. Day** and **Memorial Day**, contributed marginally (4 counts each).

Yearly Traffic Trends (2012–2018)

- Traffic volumes showed a significant upward trend:
 - Increased by **219.88%** over the analyzed period.
 - **2015–2018** witnessed the steepest growth, with a **7090706** (86.29%) rise in traffic.
 - The sharpest incline occurred between **2015 and 2018**, where traffic jumped from **8,216,926 to 15,307,632**.

Traffic Volume Forecast for future Dates

- For the year **2019**, Power BI Forecast predicted **1,62.79,157.34** as total traffic volume
- Upper bound - **2,87,40,552.80**
- Lower bound - **38,17,761.88**

Conclusion

The analysis highlights distinct traffic patterns influenced by hour, weather, holidays, and year. Cloudy weather, non-holiday periods, and morning hours witness the highest traffic volumes. The upward traffic trend over the years indicates growth in vehicle movement or urbanization.

This structured report serves as a foundation for data-driven decisions regarding traffic management and forecasting.