Bengaluru Cab Traffic Analysis – Project Documentation

This project demonstrates an end-to-end data analytics pipeline using **Python**, **MySQL**, and **Power BI** to analyse synthetic cab trip data for Bengaluru city.

Objective:

To identify ride patterns, peak hour congestion, area-wise demand, and revenue distribution by analysing 2000+ cab trip records.

Tools & Technologies Used:

- **Python (pandas)** – Data cleaning and preprocessing
- **MySQL** – Data storage and SQL-based querying
- **Power BI** – Interactive dashboard and data visualization

Project Structure:

Bengaluru_Cab_Traffic_Analysis/

- data/
  - bengaluru_cab_data_raw.csv
  - bengaluru_cab_data_cleaned.csv
- python/
  - test.py
- sql/
  - create_table.sql
  - insights_queries.sql
- Power Bi/
  - Bengaluru_Traffic_Dashboard.pbix
  - Bengaluru_Traffic_Dashboard.pdf

Steps Performed:

1. **Data Cleaning (Python)**

- Removed missing and duplicate records
- Created “is_peak” column based on time of day
- Exported cleaned data to CSV

2. **Data Storage (MySQL)**

- Created a relational table
- Imported cleaned data using “LOAD DATA INFILE”
- Queried for traffic insights by area, time, and ride type

3. **Dashboard (Power BI)**

- Built visuals: KPIs, peak hour trends, ride type share, pickup area volume, heatmap
- Enabled slicers to filter by weekday, peak time, and ride type
- Key Insight: 34.8% of rides occur during peak hours

Sample SQL Insight:

```
SELECT pickup_area, COUNT (*) AS total_trips

FROM bengaluru_trips

GROUP BY pickup_area

ORDER BY total_trips DESC;
```

Key Insights:

- **Koramangala** and **Whitefield** had the highest trip demand
- **Ola** accounted for the highest revenue share
- Peak trip volume occurred between **8–10 AM** and **5–7 PM**
- **34.8%** of trips occurred during peak hours

Outcome:

A production-style project simulating real-world data flow from raw data to business-ready dashboards — suitable for resume, interview discussions, and BI portfolios.

Created By

Divyalakshmi PV
Aspiring Data Analyst | Python | SQL | Power BI