Uber Trip Analysis Using Power BI - Technical Report

---

1. Introduction & Problem Statement

Problem Statement:

Ride-sharing platforms like Uber generate massive amounts of data daily. Understanding trip trends, revenue generation, and trip efficiency is crucial for optimizing operations. This report presents a **Power BI dashboard** analyzing Uber trip data to identify insights on ride demand, revenue, and customer behavior.

Key Business Questions Answered:

- What are the **total bookings** and **total revenue** over the given period?
- How does **trip demand** vary by **time, location, and payment type**?
- What are the **most popular pickup and drop-off locations**?
- What are the **average trip distance and duration**?
- How does **ride demand fluctuate daily and hourly**?

The **goal** of this report is to create a **data-driven visualization tool** using Power BI that allows stakeholders to make informed business decisions based on ride trends.

2. Data Sources & Collection Methodology

Data Sources:

- The dataset was sourced from **Kaggle**

  Kaggle - Uber Trip Data (June 2024)

- **Data Size:**
  - **Trip Details Table:** 103,728 rows, 11 columns
  - **Location Table:** 265 rows, 3 columns

Data Cleaning & Transformation Steps:

1. **Date & Time Parsing:** Extracted **Pickup Date**, **Dropoff Date**, and **Trip Duration** etc.
2. **Location Mapping:** Used LOOKUPVALUE() to associate **LocationID** with city names.
3. **Trip Categorization:** Identified **Day vs. Night trips** using the HOUR() function.
4. **Data Validation:** Ensured that missing values and duplicate records were handled.

3. Data Model Design & Implementation

Entity-Relationship (ER) Model:

The **star schema** includes the following relationships:

- **Trip Details (Fact Table)**
  - Connected to **Calendar Table** via Pickup Date (One-to-Many)
  - Connected to **Location Table** via PULocationID (Many-to-One, Active)
  - Connected to **Location Table** via DOLocationID (Many-to-One, Inactive)
- **Location Table (Dimension Table)** maps **LocationID** to city names.
- **Calendar Table (Dimension Table)** supports time-based analysis.

DAX Calculated Columns:

- **Pickup Date, Dropoff Date** → Extracted from timestamps.
- **Pickup & Dropoff Locations** → Used LOOKUPVALUE().
- **Trip Type (Day/Night)** → Categorized using VAR HourOfDay = HOUR(Pickup Time).

4. Visualization Approach & Tool Justification

Chosen Tool: Power BI

Justification:

- Efficient data modeling with DAX calculations.
- Interactive dashboards with slicers & filters.
- Scalability & support for large datasets.
- Advanced visualization options beyond basic charts.

Dashboard Overview:

The analysis is divided into **three interactive dashboards**:

1. **Overview Analysis** *(Key Metrics & Location Trends)*
2. **Time Analysis** *(Hourly & Daily Ride Trends)*
3. **Details Tab** *(Drill-Through Trip-Level Data)*

5. Documentation of Advanced Calculations (DAX Measures)

KPI Calculations:

- Total Bookings
- Total Booking Value
- Avg Booking Value
- Total Trip Distance
- Avg Trip Time

Advanced Measures:

- Most Frequent Pickup Point
- Most Frequent Dropoff Point
- Farthest Trip

Dynamic Measure Implementation:

A **Dynamic Measure Slicer** was created using the following DAX formula:

```
Dynamic Measure = {
   ("Total Bookings", NAMEOF('Trip Details'[Total Bookings]), 0),
   ("Total Booking Value", NAMEOF('Trip Details'[Total Booking Value]), 1),
   ("Total Trip Distance", NAMEOF('Trip Details'[Total Trip Distance Meaure]), 2)
}
```

This allows users to switch between key performance indicators dynamically.

6. Analysis of Findings & Insights

Key Findings:

1. **Total Bookings:** 103.7K trips were completed in June 2024.
2. **Revenue Insights:**
    - **Total Revenue:** $1.6M
    - **Avg Fare per Trip:** $15.0
3. **Trip Distance & Duration:**
    - **Total Distance:** 349K miles
    - **Avg Trip Distance:** 3 miles
    - **Avg Trip Time:** 16 minutes
4. **Peak Travel Times:**
    - Highest ride demand occurs between **6 AM - 9 AM and 5 PM - 8 PM**.
    - **Weekends see a 15% increase in bookings.**
5. **Most Popular Locations:**
    - **Most Frequent Pickup:** Penn Station/Madison Sq West
    - **Most Frequent Drop-off:** Upper East Side North

7. Challenges Encountered & Solutions Implemented

Challenges:

- **Data Inconsistencies:** Some records had missing or invalid trip durations.
- **Large Dataset Performance:** Rendering over 100K rows in Power BI affected performance.
- **Categorizing Night vs. Day Trips:** Time-based classification required additional DAX calculations.

Solutions Implemented:

- **Handled missing data** by filtering out incomplete records.
- **Optimized Power BI performance** by aggregating trip data before visualization.
- **Created dynamic measure switching** to enhance filtering without multiple visuals.

8. Future Enhancements & Recommendations

Planned Improvements:

- **Make dashboards mobile-responsive** for better accessibility.
- **Incorporate real-time data updates** to track ride demand dynamically.
- **Enhance user engagement** with drill-down maps for route visualizations.
- **Expand to multi-city analysis** for comparing ride demand across locations.

9. Conclusion

This Uber trip analysis successfully demonstrates how **Power BI** can be leveraged for data-driven decision-making. With **interactive dashboards, DAX-driven KPIs, and detailed insights**, stakeholders can optimize ride-sharing operations, predict demand fluctuations, and enhance customer experience.

Future enhancements will focus on **scalability, real-time analytics, and advanced geospatial mapping** to further improve data visualization and business strategy.