Uber Trip Analysis - Data Model Documentation

---

1. Introduction

This document provides a comprehensive **data model documentation** for the **Uber Trip Analysis** project. It includes an **entity-relationship diagram (ERD)**, a **data dictionary**, and documentation of **calculated fields** used within the Power BI model.

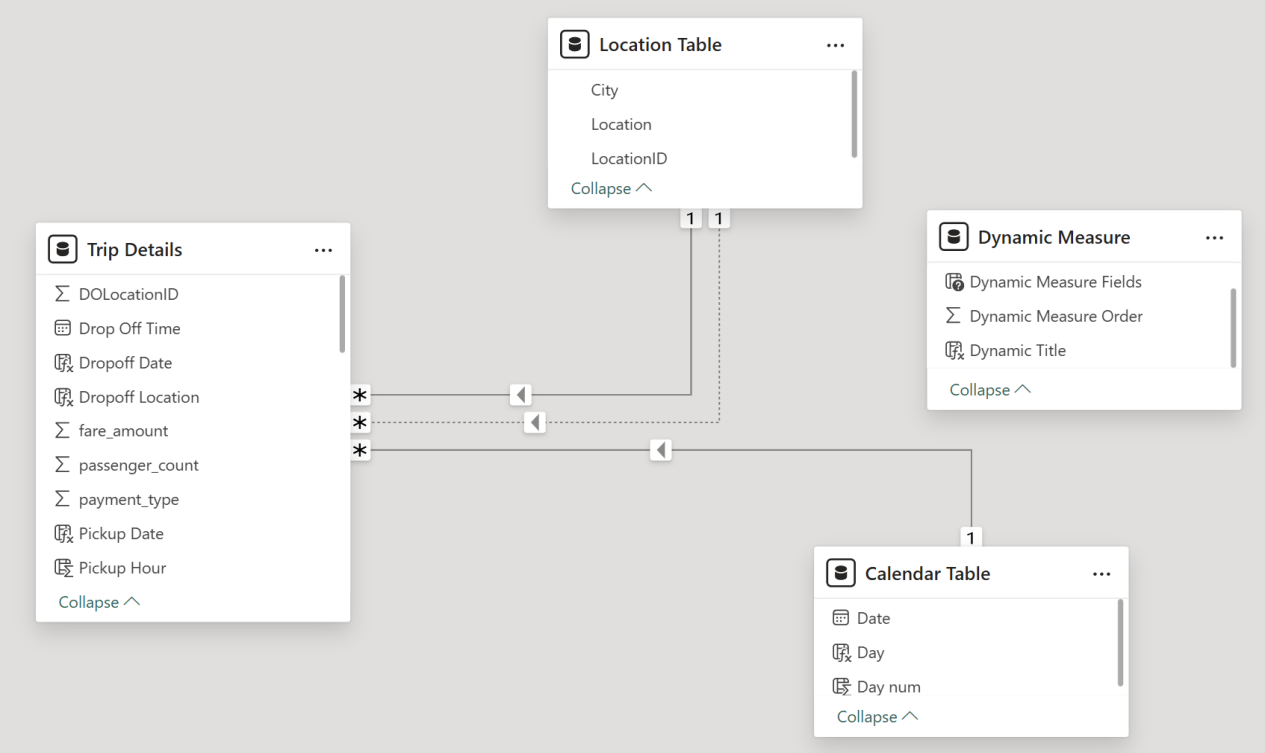
2. Entity-Relationship Diagram (ERD)

The data model follows a **star schema** structure with a central **fact table** (**Trip Details**) and multiple **dimension tables**. The relationships are as follows:

Key Relationships:

- **Trip Details (Fact Table)** → Connected to **Calendar Table** via Pickup Date.
- **Trip Details (Fact Table)** → Connected to **Location Table** via PULocationID (Active Relationship).
- **Trip Details (Fact Table)** → Connected to **Location Table** via DOLocationID (Inactive Relationship).

Entity-Relationship Diagram:



3. Data Dictionary

The following tables are used in the data model:

Table: Trip Details (Fact Table)

This table records detailed information about each Uber trip, including time, fare, distance, and payment type.

Column Name	Description
Trip ID	*Unique identifier for each trip.*
Pickup Time	*Timestamp of when the ride started.*
Drop Off Time	*Timestamp of when the ride ended.*
Passenger Count	*Number of passengers in the ride.*

Trip Distance	*Distance covered during the trip.*
PULocationID	*Pickup location ID (linked to Location Table).*
DOLocationID	*Drop-off location ID (linked to Location Table).*
Payment Type	*Payment method used (e.g., Credit Card, Cash).*
Fare Amount	*Base fare charged before additional fees.*
Surge Fee	*Extra charge applied during high demand periods.*
Vehicle	*Type of Uber service (UberX, UberXL, Uber Black, etc.).*

Table: Location Table (Dimension Table)

This table maps numeric Location IDs to actual area names.

Column Name	**Description**
LocationID	*Unique identifier for each location.*
Location	*Name of the area or neighborhood.*
City	*City where the location exists.*

Table: Calendar Table (Dimension Table)

This table provides a structured view of dates to facilitate time-based analysis.

Column Name	**Description**
Date	*Unique date value.*
Day	*Day of the week (e.g., Monday).*
Day Num	*Numeric representation of the day (1-7).*

4. Documentation of Calculated Fields

Several **calculated columns and measures** were implemented in Power BI to enhance data analysis.

Calculated Columns:

Trip Details Table:

- **Pickup Date** → *DATE(YEAR('Trip Details'[Pickup Time]), MONTH('Trip Details'[Pickup Time]), DAY('Trip Details'[Pickup Time]))*
- **Dropoff Date** → *DATE(YEAR('Trip Details'[Drop Off Time]), MONTH('Trip Details'[Drop Off Time]), DAY('Trip Details'[Drop Off Time]))*
- **Pickup Location** → *LOOKUPVALUE('Location Table'[Location], 'Location Table'[LocationID], 'Trip Details'[PULocationID])*
- **Dropoff Location** → *LOOKUPVALUE('Location Table'[Location], 'Location Table'[LocationID], 'Trip Details'[DOLocationID])*
- **Trip Type (Day/Night)** → Categorizes trips based on pickup time:

```
VAR HourOfDay = HOUR('Trip Details'[Pickup Time])
RETURN IF(HourOfDay >= 17 || HourOfDay < 6, "Night Trip", "Day Trip")
```

Dynamic Measures:

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.

KPI Measures:

- **Total Bookings:** *COUNT('Trip Details'[Trip ID])*
- **Total Booking Value:** *SUM('Trip Details'[Fare Amount]) + SUM('Trip Details'[Surge Fee])*
- **Avg Booking Value:** *DIVIDE(Total Booking Value, Total Bookings, BLANK())*
- **Total Trip Distance:** *SUM('Trip Details'[Trip Distance]) / 1000 & "K miles"*
- **Avg Trip Time:** *DATEDIFF('Trip Details'[Pickup Time], 'Trip Details'[Drop Off Time], MINUTE) & " min"*

Advanced Calculated Fields:

- **Most Frequent Pickup Point:**

TOPN(1, SUMMARIZE('Trip Details', 'Location Table'[Location], "PickupCount", COUNT('Trip Details'[Trip ID])), [PickupCount], DESC)

- **Most Frequent Dropoff Point:** Utilizes RANKX() and SUMMARIZE() functions to determine the top drop-off locations.
- **Farthest Trip:** Determines the longest trip using MAX('Trip Details'[Trip Distance]) and retrieves the associated pickup and drop-off locations.