# Data Intake Report

**Name:** G2M insight for Cab Investment firm
**Report date:** 13-June-2024
**Internship Batch:** LISUM34
**Version:**1.0
**Data intake by:** Aparna Virupkashi
**Data intake reviewer:** Data Glacier
**Data storage location: https://github.com/Aparna-6309663/DataGlacier-VC/tree/branch1/Week-2**
**Tabular data details:** Cab Data

| | |
|---|---|
| **Total number of observations** | 359392 |
| **Total number of files** | 1 |
| **Total number of features** | 7 |
| **Base format of the file** | csv |
| **Size of the data** | 19.2 MB |

**Tabular data details:** Transaction_ID

| | |
|---|---|
| **Total number of observations** | 440098 |
| **Total number of files** | 1 |
| **Total number of features** | 3 |
| **Base format of the file** | csv |
| **Size of the data** | 8.58 MB |

**Tabular data details:** Customer_Data

| | |
|---|---|
| **Total number of observations** | 49171 |
| **Total number of files** | 1 |
| **Total number of features** | 4 |
| **Base format of the file** | csv |
| **Size of the data** | 1 MB |

**Tabular data details:** City

| | |
|---|---|
| **Total number of observations** | 20 |
| **Total number of files** | 1 |
| **Total number of features** | 3 |
| **Base format of the file** | csv |
| **Size of the data** | 4 KB |

**Proposed Approach:**

1. Data was sourced from GitHub.

2. The dataset was cleaned to remove any inconsistencies and checked for appropriate data types.

3. The relationships between different tables were identified and Each table was linked using primary and secondary keys.

4. Tables were modeled based on their identified relationships to ensure accurate analysis.

5. The profit of rides was calculated by keeping other factors constant.
Only the features 'Price_Charged' and 'Cost_of_Trip' were used to determine profit.

6. The 'Users' feature from the city dataset was considered as the total number of cab users in each city, including both yellow and pink cab users.