Data Intake Report

Name: G2M insight for Cab Investment firm

Report Date : 12th November 2022

Internship Batch : LISUM15

Version:<1.0>

Data intake by: Abhimanyu Gangani Data intake reviewer :

Data storage location: https://bitbucket.org/agangani97/week-2-eda-and-hypothesis/

# 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** | 23 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** | City.csv |
| **Size of the data** | 759 bytes |

# Tabular data details: Customer\_ID

|  |  |
| --- | --- |
| **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** | 1.1 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** | 9 MB |

**Proposed Approach:**

* There are no duplicate and null records found in the dataset
* Outliers are present in ‘Price Charged’ feature. We are not treating this as outliers because of unavailability of more details. (We Assume these might be due to high end cars or harsh weather conditions.)
* There are no duplicated rows neither missing values in the datasets.
* The ‘Profit’ feature is calculated as follows: **Profit = Price Charged – Cost of Trip**