Data Intake Report

Name: G2M Cab Dataset Insights

Report date: 8/03/2021

Internship Batch: LISP01

Version:<1.0>

Data intake by: Ashish Rawat

Data intake reviewer:

Data storage location: https://github.com/DataGlacier/DataSets

**Tabular data details:**

|  |  |
| --- | --- |
| **Total number of observations** | 359392 |
| **Total number of files** | 1 |
| **Total number of features** | 7 |
| **Base format of the file** | CSV File |
| **Size of the data** | 22.088 MB |

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

|  |  |
| --- | --- |
| **Total number of observations** | 49171 |
| **Total number of files** | 1 |
| **Total number of features** | 4 |
| **Base format of the file** | CSV File |
| **Size of the data** | 1027 KB |

|  |  |
| --- | --- |
| **Total number of observations** | 440098 |
| **Total number of files** | 1 |
| **Total number of features** | 3 |
| **Base format of the file** | CSV File |
| **Size of the data** | 8788 KB |

**Proposed Approach:**

* Followed KDD (Knowledge Data Discovery) approach and analyzed each features of datasets thoroughly.
* Two assumptions:

1. Profit Margin = (Price Charged – Cost of Trip)/ Price Charged
2. Joins used to merge datasets were made using left and inner join primarily.