Data Intake Report

Name: G2M Insight for Cab Investment Firm

Report date: April 14, 2023

Internship Batch: LISUM20

Version:1.0

Data intake by: Matthew Zhou

Data intake reviewer:

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

**Tabular data details:**

Combined Dataset

|  |  |
| --- | --- |
| **Total number of observations** | 359392 |
| **Total number of files** | 4 |
| **Total number of features** | 14 |
| **Base format of the file** | .csv |
| **Size of the data** | 31.3 MB |

Cab Data

|  |  |
| --- | --- |
| **Total number of observations** | 359392 |
| **Total number of features** | 7 |
| **Base format of the file** | .csv |
| **Size of the data** | 21.2 MB |

City

|  |  |
| --- | --- |
| **Total number of observations** | 19 |
| **Total number of features** | 3 |
| **Base format of the file** | .csv |
| **Size of the data** | 759 bytes |

Customer

|  |  |
| --- | --- |
| **Total number of observations** | 49171 |
| **Total number of features** | 4 |
| **Base format of the file** | .csv |
| **Size of the data** | 1.1 MB |

Transaction

|  |  |
| --- | --- |
| **Total number of observations** | 440098 |
| **Total number of features** | 3 |
| **Base format of the file** | .csv |
| **Size of the data** | 9MB |

**Proposed Approach:**

* Merge Cab Data, Customer, and Transaction datasets into one data frame.
* Create features for Year, Month, Day, Weekday from the Date of Travel feature.
* Create features for Profit and Profit/KM from the Price Charged, Cost of Travel, and KM travelled features.
* Analyze number of rides, users, profit, etc. over varying years, months, dates, demographics, and locations
* Assumptions – Profit is equal to Price Charged minus Cost of Travel.
  + Date of Travel is an excel serial number