Data Intake Report

# Name: G2M insight for Cab Investment firm

Report date: 12/03/2022

Internship Batch: LISUM07

Version:1.0

Data intake by: Ezzuldin Zaky

Data intake reviewer: N/A

Data storage location: https://github.com/Ezz-Zaky/DS\_Internship\_Projects.git

**Tabular data details:**

Cab\_Data:

|  |  |
| --- | --- |
| **Total number of observations** | 359393 |
| **Total number of files** | 4 |
| **Total number of features** | 9 |
| **Base format of the file** | csv |
| **Size of the data** | 27.9MB |

City:

|  |  |
| --- | --- |
| **Total number of observations** | 20 |
| **Total number of files** | 4 |
| **Total number of features** | 3 |
| **Base format of the file** | csv |
| **Size of the data** | 0.553KB |

Customer\_ID:

|  |  |
| --- | --- |
| **Total number of observations** | 49171 |
| **Total number of files** | 4 |
| **Total number of features** | 4 |
| **Base format of the file** | csv |
| **Size of the data** | 1.027KB |

Transaction\_ID:

|  |  |
| --- | --- |
| **Total number of observations** | 440098 |
| **Total number of files** | 4 |
| **Total number of features** | 3 |
| **Base format of the file** | csv |
| **Size of the data** | 8.788MB |

**Proposed Approach:**

I will use Excel to remove duplicates and will assume nothing extra on top of the assumptions in the project description. EDA will be carried out on an ipynb file where various quantitative methods will be used, and hypotheses will be formed. The datasets will then be imported to a MySQL server to extract derived datasets based on the hypotheses. Visuals will be formed from these derived datasets and conclusions can be drawn about the two cab companies.