

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 09/10/2024

Internship Batch: LISUM34

Version: 1.0

Data intake by: Mahima Sadananda

Data intake reviewer: Data Glacier

Data storage location: https://github.com/MahimaSadananda/DataGlacier_Internship/tree/main

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 | 21.2 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 | 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 | 1.1MB |

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:

The 4 datasets were imported and analyzed to understand the columns, key identifier, data types, etc. They were merged to form a master data for further analysis. This was then checked for null and duplicate values. Several hypotheses were created to aid the analysis. Finally, the findings and recommendations were presented.

Assumptions:

1. There are only 2 cab companies in the US market.
2. The given data represents the majority and there are no outliers.
3. There is no other external factor that would affect the decision making for the investor apart from the parameters in the given dataset.