

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 7/13/2024

Internship Batch:LISU35

Version:

Data intake by: Nuzat Siddiqua

Data intake reviewer:

Data storage location: https://github.com/nuzats19/Week2_Cab-InvestmentEDA

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 | 19.2+ MB |

Tabular data details: City Data

| | |
|------------------------------|--------------|
| 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 | 608.0+ bytes |

Tabular data details: Customer Data

| | |
|------------------------------|----------|
| 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 | 10.1+ MB |

Tabular data details: Transaction Data

| | |
|------------------------------|--------|
| 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 | 1.5+MB |

Proposed Approach:

- Checked for any missing values and outliers in critical columns such as 'Transaction ID', 'Customer ID', 'Date of Travel', 'Company', and 'City'. There is no null value or outliers. Using z score checked for any outliers.
- Checked for correlation between features.
- The date of travel column was in excel series form changed it real date.
- Profit is calculated using the price charged and cost of trips.
- Applied data formatting on date of travel column.
- Demands of cab over the summer and spring increases. Take a fall on winter and fall.