

Data Intake Report

Name: G2M Insight for Cab Investment Firm

Report date: 6/20/2023

Internship Batch: LISUM22 May 30 – Aug 30, 2023

Version: 1.0

Data intake by: Krishna Pratheek Seshabhattachar

Data intake reviewer: <>

Data storage location: <>

Tabular data details:

Cab Data:

Total number of observations	359392
Total number of features	7
Base format of the file	Csv
Size of the data	21.5 MB

City:

Total number of observations	20
Total number of features	3
Base format of the file	Csv
Size of the data	1 KB

Customer ID:

Total number of observations	49171
Total number of features	4
Base format of the file	Csv
Size of the data	1 MB

Transaction ID:

Total number of observations	440098
Total number of features	3
Base format of the file	Csv
Size of the data	8.58 MB

Proposed Approach:

- Performed Join operation on the 4 csv files to create a consolidated data set with appropriate relations.
- Data1 was created by joining Transaction ID and Customer ID on “Customer ID.”
- Data2 was created by joining Data1 and Cab Data on “Transaction ID.”
- Comb_data was created by joining Data2 and City on “City.”
- The outliers present in Price charged were ignored as there were no details available for duration of the trip.

- The profit column was calculated from arithmetic operation on price charged and cost of trip.
- The Users column shows the total number of cab users in a particular city using all companies including the pink and yellow cab companies present in the cab data file.