

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: May 14, 2025

Internship Batch: LISUM45

Version: <1.0>

Data intake by: Bhavya Jha

Data intake reviewer:

Data storage location: [github/week2](#)

Tabular Data Details

Cab Dataset:

Total no. of observations	359392
Total no. of files	1
Total no. of features	7
Base format of the file	.csv
Size of the data	20.2 MB

City Dataset:

Total no. of observations	20
Total no. of files	1
Total no. of features	3
Base format of the file	.csv
Size of the data	759 bytes

Transaction ID Dataset:

Total no. of observations	440098
Total no. of files	1
Total no. of features	3
Base format of the file	.csv
Size of the data	8.58 MB

Customer ID Dataset:

Total no. of observations	49171
Total no. of files	1
Total no. of features	4
Base format of the file	.csv
Size of the data	1 MB

Proposed Approach:

1. Each dataset was checked for duplicates and null values, and all of them had none of these.
2. The Transaction_id dataset was merged with the Cab dataset
3. The Customer_id dataset was then merged into the above merged dataset.
4. All the merged datasets are used to plot graphs that help give insights into the required areas of investigation.

Assumptions:

1. It is assumed that the payments have been made by either card or cash.
2. The analysis has been performed for the given periods between the beginning of 2016 and the end of 2018.
3. Price Charged & Cost of Trip are the only factors considered when calculating the Margin.
4. The datasets were drawn from random selection.