

Data Intake Report

Name: G2M Insight for Cab Investment Firm

Report date: 11/7/2024

Internship Batch: LISUM35

Version:1.0

Data intake by: Munirah Alfahaid

Data intake reviewer:

Data storage location:

Tabular data details:

Cab Data Dataset:

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	20663 KB

City Dataset:

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	1 KB

Customer ID Dataset:

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	1027 KB

Transaction ID Dataset:

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	8788 KB

Proposed Approach:

1. I individually analyzed and checked each dataset.
2. Then, I merged all four datasets to facilitate better analysis operations.
3. I added new columns using the existing columns.

Assumptions:

- The data analysis for both companies assumed the presence of external noise beyond the provided data.
- The analysis was conducted with the assumption that the dataset was limited to the period between 2016 and 2018.
- Only payments made by cash or card were considered.