

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 08/11/22

Internship Batch: LISUM15

Version: 1.0

Data intake by: Christos Christoforou

Data intake reviewer:

Data storage location: <https://github.com/cchristoforou97/G2M-insights-for-Cab-Investment-firm>

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	22.418 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	1 KB

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	1027 KB

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	8.788 MB

Proposed Approach:

- Merge the 4 datasets.
- Check for duplicates.
- The data provided for the Population and Users in the City dataset does not make sense so we drop these columns. We cannot expect the Population and Users to stay the same in a three-year period. Also, the number of unique customers based on the Customer_ID column does not match the number of users based on the Users column.
- Check for null-values.
- EDA.
- Hypothesis Testing.
- Conclusions and Recommendations.