

Data Intake Report 1

Name: G2M Case Study

Report date: 28/8/2021

Internship Batch: LISUM03

Version:1.0

Data intake by: Raghad Almeataz

Data intake reviewer:<intern who reviewed the report>

Data storage location: <https://github.com/Raghad-Almeataz/DS-G2M-insight-for-Cab-Investment-firm/tree/main/data>

Tabular data details:

Cab_Data.csv

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	22418 KB

City.csv

Total number of observations	19
Total number of files	1
Total number of features	3
Base format of the file	csv
Size of the data	1 KB

Customer_ID.csv

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.csv

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

Fulldataset.csv

Total number of observations	359392
Total number of files	1
Total number of features	18
Base format of the file	csv
Size of the data	45736 KB

Proposed Approach:

- All datasets are merged to produce Fulldataset.csv.
- Additional features added in Fulldataset.csv extracted from existing columns:
 - Profit Column: Subtracted costs from price charged.
 - Day, Month, and Year Columns: Extracted from Date of Travel Column.