

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 3/9/2021

Internship Batch: LISUM02

Version: 1.0

Data intake by: Aggelos Konioris

Data intake reviewer: Data Glacier

Data storage location: https://github.com/AGGELOS-KONIORIS/Week_2_deliverables

Tabular data details:

Name of the file	Cab_Data
Total number of observations	359.392
Total number of files	4
Total number of features	7
Base format of the file	csv
Size of the data	20.663KB

Tabular data details:

Name of the file	City
Total number of observations	20
Total number of files	4
Total number of features	3
Base format of the file	csv
Size of the data	1KB

Tabular data details:

Name of the file	Customer_ID
Total number of observations	49.171
Total number of files	4
Total number of features	4
Base format of the file	csv
Size of the data	1.027KB

Tabular data details:

Name of the file	Transaction_ID
Total number of observations	440.098
Total number of files	4
Total number of features	3
Base format of the file	csv

Size of the data	8.788KB
-------------------------	---------

We merge these four datasets into one dataset based on the identical columns in order to avoid duplicates and then we check for missing values.

Finally, these are the assumptions that have been made.

- 1) Total profit for each cab
- 2) Years profit analysis
- 3) Seasonality
- 4) Profit analysis based on age
- 5) Profit analysis based on city
- 6) KM travelled profit analysis
- 7) Profit analysis based on income