

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 7.03.2021

Internship Batch: LISP01

Version: 1.0

Data intake by: Popa Larisa

Data intake reviewer:

Data storage location:

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	1,27 KB

City:

Total number of observations	21
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	1 KB

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	20,663 KB

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 KB

Master_data:

Total number of observations	359392
Total number of files	1
Total number of features	17
Base format of the file	.csv
Size of the data	47,976 KB

Master_data file was obtained aggregating all 4 data sets. The features types were originally strings. In Excel were done some format changes: 'date_of_travel' to short date', 'population' and 'users' into number because in python there were met some errors during the process.

The others were changed in python.

The need for form change has come due to cases where ordering needed to be done, and the order was incorrect because parameters were not of numerical type.

No duplications were obtained.

Profit column was added to the data computed as 'price_charged' – 'cost_of_trip'.