

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 10/04/2021

Internship Batch LISUM04

Version: 1.0

Data intake by: Mohammed Wahba

Data intake reviewer:

Data storage location:

Tabular data details:

Total number of observations	359392
Total number of features	7
Base format of the file	.csv
Size of the data	21.2 MB

Tabular data details:

Total number of observations	440098
Total number of features	3
Base format of the file	.csv
Size of the data	8.58 MB

Tabular data details:

Total number of observations	49171
Total number of features	4
Base format of the file	.csv
Size of the data	1 MB

Tabular data details:

Total number of observations	20
Total number of features	3
Base format of the file	.csv
Size of the data	679 bytes

Note: Replicate same table with file name if you have more than one file.

Proposed Approach:

- Mention approach of dedup validation (identification)

The Transaction ID feature was full of duplicates as many customers have been used the Cab service multiple times.

- Mention your assumptions (if you assume any other thing for data quality analysis)

The Date of Travel feature was out of range, so to make the date qualify for the mention range between the year 2016 and 2018. The feature had to be cleaned