

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 12/06/2022

Internship Batch:LISUM10

Version:1.0

Data intake by:Arda Baris Basaran

Data intake reviewer:<intern who reviewed the report>

Data storage location: <https://github.com/ardaxz99/Data-Glacier-Week2-and-Week-3/tree/main>

Tabular data details:

Cab.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	19.2 MB

City.csv

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	608 bytes

Customer ID.csv

Total number of observations	49170
Total number of files	1
Total number of features	4
Base format of the file	.csv
Size of the data	1.5MB

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	10.1 MB

Proposed Approach:

- Mention approach of dedup validation (identification)
All the above mentioned files to be merged and made into a master data, after which data features to be checked and studied
- Mention your assumptions (if you assume any other thing for data quality analysis)