

Data Intake Report

Name : G2M insight for Cab Investment firm

Report date: 10.03.2023

Internship Batch:LISUM19

Version:<1.0>

Data intake by: Ecem AY

Data intake reviewer

Data storage location: <https://github.com/ecemay/week2>

Tabular data details: 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	

Tabular data details: Customer Data

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	

Tabular data details: Transaction Data

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	

Tabular data details: City Data

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	

Proposed Approach:

- Mention approach of dedup validation (identification)
 - **Exploratory Data Analysis**
- Mention your assumptions (if you assume any other thing for data quality analysis)
 - Yellow cab have higher amount of users and profit than pink cab but pink cab have more stability in profit changes over year.