

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: June 14 2022

Internship Batch: LISUM10: 30

Version:1.0

Data intake by: Prince Kumar Lat

Data intake reviewer:

Data storage location: <https://github.com/princeklat03/Week2>

## Tabular data details: Cab\_Data.csv

Total number of observations	359392
Total number of files	
Total number of features	7
Base format of the file	.csv
Size of the data	22.2 Mb

## Tabular data details: City.csv

Total number of observations	20
Total number of files	
Total number of features	3
Base format of the file	.csv
Size of the data	759 bytes

## Tabular data details: Customer\_ID.csv

Total number of observations	49171
Total number of files	
Total number of features	4
Base format of the file	.csv
Size of the data	1.1 Mb

## Tabular data details: Transaction\_ID.csv

Total number of observations	440098
Total number of files	
Total number of features	3
Base format of the file	.csv
Size of the data	9 Mb

## Proposed Approach:

- Mention approach of dedup validation (identification)
- Mention your assumptions (if you assume any other thing for data quality analysis)
- **My response:** I have elaborated them in the ipynb files.