

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 11/14/2022

Internship Batch: LISUM15

Version:<1.0>

Data intake by: Xuyang Zhao

Data intake reviewer: Xuyang Zhao

Data storage location: <https://github.com/DataGlacier/DataSets.git>

Tabular data details:

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	21.5MB
File name	Cab_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	8.58MB
File name	Transaction_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	1MB
File name	Customer_ID

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	759bytes
File name	City

Proposed Approach:

- Use inner join to drop records with missing value when creating master data for EDA.
- The time period for analysis is from 31/01/2016 to 31/12/2018.