

Data Intake Report

Name: <G2M insight for cab investment firm >
Report date: <06.06.2022>
Internship Batch:<LISUM10 : 30 >
Version:<1.0>
Data intake by:<Omar R.A. Ammar>
Data intake reviewer:<intern who reviewed the report>
Data storage location: <github>

Tabular data details:

Cab_Data.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	21.2 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	759 Bytes

Customer_ID.csv

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	11.1 MB

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	9 MB

Proposed Approach:

All the nullable data and the data with unreal values were removed, the rest of the data was not merged or modified in any other way. All the data that was repeated was removed as well (refer to code for the method of removing the data).