

Data Intake Report

Name: G2M Insights
Report date: 4/24/2023
Internship Batch: LISUM20
Version: 1.0
Data intake by: Henil Sha
Data intake reviewer:
Data storage location: <https://github.com/Henilshah9/G2M>

Tabular data details: Merged_data

Total number of observations	359392
Total number of files	1
Total number of features	15
Base format of the file	.csv
Size of the data	46.6+ MB

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	19.2+ MB

Tabular data details: City

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	608.0+ bytes

Tabular data details: Customer_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	1.5+ MB

Tabular data details: Transaction_ID

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:

- Merge dataframes based on their shared attributes.
- Dimension reduction may be applied for modeling as high collinearity is found between certain features.