

Data Intake Report

Name: G2M Insight for Cab Investment Firm

Report date: 2/6/2022

Internship Batch: LISUM06

Version: 1.0

Data intake by: Clyde Wawire

Data intake reviewer: <intern who reviewed the report>

Data storage location: <https://github.com/clydewawire/G2M-Case-Study-Week2>

Tabular data details:

Cab.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	19.2MB

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	608Bytes

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	1.5MB

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	10.1MB

Proposed Approach:

- All 4 files are in .csv format.
- Data cleaning and Exploratory data analysis has been done in Jupyter Notebook.
- All four files are merged according to their references.
- The feature “Date of Travel” in Cab_data.csv was in Epoch format and has been converted into “Date time formant” using Excel
- It is assumed that Time period of data is from 31/01/2016 to 31/12/2018 and analysis has been done to recommend XYZ firm to invest on best cab company