

Data Intake Report

Name: G2M Insight for Cab Investment Firm

Report date: 10/13/2024

Internship Batch: LISUM38

Version: 1.0

Data intake by: Serhat Uğur

Data intake reviewer: Data Glacier

Data storage location: <https://github.com/serhatugur/data-science-internship>

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	20,663 KB

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	1 KB

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	1,027 KB

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	8,788 KB

Proposed Approach: I identified unique columns and made sure that there were no duplicate columns. I corrected one data type, then I merged the whole dataset together and made an exploratory data analysis. Lastly, I did five different hypothesis tests.

