

Data Intake Report

Name: G2M Insight for Cab Investment Firm

Report date: 14-10-2023

Internship Batch: LISUM26

Version:1.0

Data intake by: Murat Kıran

Data intake reviewer: -

Data storage location: <https://github.com/murattkiran/G2M-insight-for-Cab-Investment-firm>

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	20.1 MB

Tabular data details: Customer_ID

Total number of observations	359392
Total number of files	1
Total number of features	4
Base format of the file	csv
Size of the data	1 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	8.58 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	759 bytes

Proposed Approach:

- Firstly, the “check_df” function was used to examine the general characteristics of the dataset.
- Subsequently, a 'check_duplicates' function was applied to investigate the presence of duplicate rows. The analysis results indicated that there were no repeated data entries in the datasets.
- Incorrect data types in the datasets were corrected. Right after that, I utilized the 'grab_col_names' function I created to capture categorical, numeric, and cardinal variables in the datasets. Then, the categorical and numeric variables were analyzed.
- It was determined that there are no missing values in the datasets. Furthermore, outlier detection in the datasets was performed using various functions.