Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 14/03/24
Internship Batch:LISUM31
Version:1.0
Data intake by: Fabio Pontecchiani
Data intake reviewer:
Data storage location:

Tabular data details:

Cab_Data.csv

Total number of observations	359392
Total number of files	4
Total number of features	7
Base format of the file	csv
Size of the data	21.2 MB

Customer_ID.csv

Total number of observations	49171
Total number of files	4
Total number of features	4
Base format of the file	csv
Size of the data	1.1 MB

Transaction_ID.csv

Total number of observations	440098
Total number of files	4
Total number of features	3
Base format of the file	csv
Size of the data	9 MB

City.csv

Total number of observations	20
Total number of files	4
Total number of features	3
Base format of the file	csv
Size of the data	759 bytes

Proposed Approach:

- Mention approach of dedup validation (identification)

  Checked shape and look of top lines of the data, missing values and data type, general data distribution with histograms.

- Mention your assumptions (if you assume any other thing for data quality analysis)

  Joined tables on unique values like Customer IDs, Transaction IDs and cities. Calculated margins and segment demographical parameters like age and group by salary bands to start the analysis.
  Compared the two cab companies and analysis of the profits through time 2016-2019. Study by people demographics and then performed a geographical analysis of the profits.