

Data Intake Report

Name: G2M Insight for Cab Investment Firm

Report date: 14/04/23

Internship Batch: LISUM20

Version: 1.0

Data intake by: Oana Cucu

Data intake reviewer: N/A

Data storage location: https://github.com/phonemes-and-speech-entrainment/data_science_test

Tabular data details:

Total number of observations	440098
Total number of files	7
Total number of features	17
Base format of the file	ipynb
Size of the data	44.3 MB

Note: Replicate same table with file name if you have more than one file.

Proposed Approach:

- Checked for NaN values (missing values), duplicates, unique values of columns, merged the various files into master data but did not use all of the columns
- Differences between the profitability of the two companies may be due to different types of investments (cost of trip versus price), length of trips, demographic information (age, income, gender, city), seasonality (week day, month, quarter), or it may vary with time due to unknown factors (economic year).