

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 4/14/2023

Internship Batch: LIMSUM20

Version: 1.0

Data intake by: Darshan Patil

Data intake reviewer:

Data storage location:

Tabular data details:

CabData.csv

Total number of observations	267075
Total number of files	1
Total number of features	7
Base format of the file	.csv
Size of the data	1.1 MB

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	38 Bytes

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	53 KB

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	450 KB

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

Proposed Approach:

- Mention approach of dedup validation (identification):
 - While cleaning the data I used python commands in order to verify that there were not any duplicate values in the data sets
- Mention your assumptions (if you assume any other thing for data quality analysis)
 - The Cab_Data.csv dataset likely contains information such as the date and time of the transaction, pick-up and drop-off locations, trip distance, trip duration, fare, and surcharges, if any. The dataset also likely distinguishes between the two competing cab companies.
 - The Customer_ID.csv dataset probably contains demographic information about the customers, such as age, gender, and income level. This dataset will be essential for understanding the target audience of each cab company and identifying trends in user preferences.
 - The Transaction_ID.csv dataset connects transactions with customer information and payment methods. It is likely that the dataset includes information about the type of payment method used (e.g., cash, credit card, or mobile app) and transaction IDs that can be mapped to the corresponding trip details in the Cab_Data.csv dataset.
 - The City.csv dataset will help analyze the market size and penetration of each cab company in different cities across the US. This dataset may include the total population of each city, the number of cab users, and the market share of each cab company in that city.

**Note: Convert this doc in pdf and provide the link of pdf file in your dashboard.
Please do not forget to remove this section while converting the file into pdf.**