

Data Intake Report

Name: G2M Insight for Cab Investment Firm

Report date: 13th November 2023

Internship Batch: LISUM27

Version:

Data intake by:

Data intake reviewer:

Data storage location: [<https://github.com/DataGlacier/DataSets>]

Tabular data details:

1. Cab_Data.csv

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	19.2 MB

2. 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	1 KB

3. 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	1 MB

4. 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	8788 KB

Proposed Approach:

- Data Type Conversion:
Data types of above four files were object, int64 and float64.

Checked and converted data types of features in the four files (Cab_Data, City_Data, Customer_ID, Transaction_ID).

Ensured appropriate data types for efficient analysis.

- **Date Format Correction and Feature Creation:**
Corrected the date format in the 'Date of Travel' feature in Cab_Data.
Created the 'Profit' column in Cab_Data based on the assumption that profit is calculated as the difference between 'Price Charged' and 'Cost of Trip'.
- **Function Usage in Pandas:**
Utilized functions in the pandas library for data manipulation and analysis.
- By following these steps and utilizing pandas functions, we ensured data consistency, corrected date formats, and created relevant features for a comprehensive analysis of the cab industry data.