Data Intake Report

Name: G2M Case Study
Report date: <19/06/2021>
Internship Batch: LISUM01
Version: <1.0>
Data intake by: Data Glacier
Data intake reviewer: Diaz Aurelli Salsabila
Data storage location: <location URL eg: github, cloud>

Tabular data details:

Customer_ID

Total number of observations	4
Total number of files	1
Total number of features	49,172
Base format of the file	.csv
Size of the data	1.363 MB

Transaction_ID

Total number of observations	3
Total number of files	1
Total number of features	440,099
Base format of the file	.csv
Size of the data	8.788 MB

City

Total number of observations	3
Total number of files	1
Total number of features	20
Base format of the file	.csv
Size of the data	1 KB

Cab_data

Total number of observations	7
Total number of files	1
Total number of features	369,393
Base format of the file	.csv
Size of the data	21.124 MB

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

Proposed Approach:

- Assumptions
    - Outliers are present in Price_Charged feature but due to unavailability of trip duration details ,we are not treating this as outlier.
    - Profit of rides are calculated keeping other factors constant and only Price_Charged and Cost_of_Trip features used to calculate profit.
    - Users feature of city dataset is treated as number of cab users in the city. we have assumed that this can be other cab users as well (including Yellow and Pink cab)
- Approach :
    - Separate the observations to different cells in Microsoft Excel.
    - Import the file into the Jupiter for data processing.
    - Process the EDA notebook in Jupyter.