Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 3/13/2023
Internship Batch: LISUM19
Version:<1.0>
Data intake by: Zhanqiu Guo
Data intake reviewer: Zhanqiu Guo
Data storage location: https://github.com/Zhanqiu-Guo/DataSets

Tabular data details: Cab_Data

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	19.2KB

Tabular data details: City

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	608B

Tabular data details: Customer_ID

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	1.5MB

Tabular data details: Transaction_ID

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	10.1KB

Proposed Approach:
- There are no NULL values in dataset.

• 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.