Data Intake Report

Name: Cab Investment Firm Data Analysis
Report date: 16/2/2022
Internship Batch: LISUM06
Version:<1.0>
Data intake by: Hassan Faheem
Data intake reviewer:
Data storage location: GitHub

Tabular data details: Cab Data

Total number of observations	359393
Total number of files	4
Total number of features	7
Base format of the file	.csv
Size of the data	20.1 MB

Tabular data details: City

Total number of observations	21
Total number of files	4
Total number of features	3
Base format of the file	.csv
Size of the data	759 Bytes

Tabular data details: Customer_ID

Total number of observations	49172
Total number of files	4
Total number of features	4
Base format of the file	.csv
Size of the data	1 MB

Tabular data details: Transaction_ID

Total number of observations	440099
Total number of files	4
Total number of features	3
Base format of the file	.csv
Size of the data	8.58 MB

Proposed Approach:

- Data Sources are merged on Customer ID, Transaction ID & City.
- There are outliers present in the feature Price Charged but they cannot be treated as outliers due to insufficient information for Trip Duration.
- The features of Price Charged & Cost of Trip are used for the calculation of Profit.
- The feature of Users in the dataset of City are assumed to be cab users of all cab companies including Yellow & Pink Cab Companies.