Data Intake Report

Name:G2M insight for Cab Investment firm
Report date:7-03-2021
Internship Batch:LISP01
Version:<1.0>
Data intake by: Maureen Wanjiku Githaiga
Data intake reviewer: Maureen Githaiga
Data storage location: https://github.com/Reen-G/DataSets.git

Tabular data details:

Total number of observations	359392
File name	Cab_Data
Total number of features	7
Base format of the file	csv
Size of the data	19MB

Total number of observations	49171
File name	City
Total number of features	4
Base format of the file	csv
Size of the data	1.5MB

Total number of observations	20
File name	Customer_ID
Total number of features	3
Base format of the file	csv
Size of the data	608Bytes

Total number of observations	440098
File name	Transaction_ID
Total number of features	3
Base format of the file	CSV
Size of the data	10.1MB

Proposed Approach:
- Exploratory data analysis approach was taken,checking for unique and similar attributes in the data. Null values and duplicates also checked for data cleaning.
- Assumptions: The profit calculated using the price charged and cost of trip other factors constant.
- The outliers in the Price Charged weren't treated since enough data lacked to invalidate them.