Data Intake Report

Name: <G2M insight for Cab Investment firm >
Report date: <07/15/2023>
Internship Batch:<LISUM23>
Version:<1.0>
Data intake by:<Wenjie Zheng>
Data intake reviewer:<None>
Data storage location: <https://github.com/wejie0824/hhh.git>

Tabular data details: 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	<16.5+ MB >

Tabular data details: City.csv

Total number of observations	<19>
Total number of files	<1>
Total number of features	<3>
Base format of the file	<csv>
Size of the data	<312.0+ bytes >

Tabular data details: 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	<1.3+ MB >

Tabular data details: Customer_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	<8.4+ MB >

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

Proposed Approach:
- Data Cleaning (merging 4 csv files together based on common attributes)
- Descriptive Statistics(Gender, Company, Price Charged, and etc)
- Bar Graphs or Box plots used to study the relationship between variables.