Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: <April 26, 2022>
Internship Batch: LISUM07
Version:<1.0>
Data intake by:Francisco Mendez
Data intake reviewer:
Data storage location: https://github.com/explorerjs32/cab_investment_firm_project

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	22.2 Mb

Tabular data details: City.csv

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	759 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	1.1 Mb

Tabular data details: Transaction_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	9 Mb

Proposed Approach:
- Combine all 4 data sets into one to find correlations between the data stored
- Used the combined data set to answer the following:
    - o Which company has more transactions per year and month?
    - o Which year and month seems to be the most profitable for each company?
    - o What is the average cost per distance traveled?
    - o Which company has more riders per city?
    - o Which city is more profitable to ride on?