Data Intake Report

Name: Cab Industry G2M Insights
Report date: 13 April 2022
Internship Batch: LISUM08
Version:1.0
Data intake by: Garreth Lee
Data intake reviewer:
Data storage location: https://github.com/DataGlacier/DataSets

Tabular data details:

Cab_data.csv

Total number of observations	359,392
Total number of files	1
Total number of features	7
Base format of the file	.csv
Size of the data	21.2 MB

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

Customer_ID.csv

Total number of observations	49,171
Total number of files	1
Total number of features	4
Base format of the file	.csv
Size of the data	1.1 MB

Transactions.csv

Total number of observations	440,098
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	9 MB

Proposed Approach:

- Remove missing values and duplicated values (if any)
- Assumptions:
    - Some trips didn't register within the cab database, since there are more transactions than there are trips in the data. So, these additional transactions were omitted.
    - Assumes that there are other cab companies other than the Yellow Cab and the Pink Cab company
    - Assume that the users column from the cities table represents the total unique users from each city (including people who use other brands of cabs)