Data Intake Report

Name: Insight for Cab Investment firm
Report date: March 06, 2021
Internship Batch: LISP01
Version: 1.0
Data intake by: Swapnil Vishwakarma
Data intake reviewer:
Data storage location: Github

Tabular data details:

Table name	Cab Data
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	19.2+ MB

Table name	City
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	608.0+ bytes

Table name	Customer
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	1.5+ MB

Table name	Transaction
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	10.1+ MB

Table name	US Holidays
Total number of observations	3288
Total number of files	1
Total number of features	5
Base format of the file	.csv
Size of the data	246.7+ KB

Assumptions made:

- Mean Price Charged is 423 USD whereas Median Price Charged is 386 USD which implies there are some outliers present.
- Mean Profit is 137 USD whereas Median Profit is around 82 USD which implies there are some outliers present.
- Profit is calculated using Price Charged and Cost of Trip only as there is no data for time of travel and time of the day.
- Profit per KM is calculated by using Profit and KM Travelled only as there is no data for time of travel and time of the day.
- Users feature of city dataset is treated as number of cab users in the city.
- Saturday and Sunday together is considered as one weekend.
- Demand of cabs over the weekend increases hence the Price Charged is also increased.