Data Intake Report

Name: Week 2 Task: EDA Notebook
Report date: 13 June 2022
Internship Batch: LISUM 10
Version:1.0
Data intake by: Kashish Joshipura
Data intake reviewer:
Data storage location: https://github.com/kashish1928/DataSets

Tabular data details:

File Name	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

File Name	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	21.2 MB

File Name	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

File Name	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

File Name	US Holiday Dates (2004-2021).csv

Total number of observations	342
Total number of files	1
Total number of features	6
Base format of the file	csv
Size of the data	16 KB

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

Proposed Approach:

APPROACH

- **Data Exploration**
  Understanding the data and identify certain patterns in the data.

- **Investigate**
  Determine which cab is most profitable by creating and explaining the relationship.

- **Hypothesis**
  Recommending the most profitable company for XYZ to invest in.

ASSUMPTIONS

- The difference of Price Charged and Cost of Trip to be the Profit.

- The Date from USA Holiday data.csv to be Holiday in masterdata.csv.

- Neglect the outliers in Price Charged since we are calculating the profit.