Data Intake Report

Name: Cab Industry EDA
Report date: 30 August 2021
Internship Batch: LISUM03
Version:<1.0>
Data intake by: Prasad Mahajan
Data intake reviewer:
Data storage location: https://github.com/prasad847/Data-glacier

Tabular data details:

1. Cab_Data.csv

Total number of observations	359392
Total number of files	4
Total number of features	7
Base format of the file	.csv
Size of the data	22.4 MB

2. City.csv

Total number of observations	20
Total number of files	4
Total number of features	3
Base format of the file	.csv
Size of the data	1KB

3. Customer_ID

Total number of observations	49171
Total number of files	4
Total number of features	4
Base format of the file	.csv
Size of the data	1.02 MB

4. Transaction_ID

Total number of observations	440098
Total number of files	4
Total number of features	3
Base format of the file	.csv
Size of the data	8.78 MB

Proposed Approach:

- A XYZ private firm looking for investment in a Cab company. On behalf of that exploratory Data Analysis is done to help the company to make a right decision.
- Data understanding, Data Pre-processing and by observing trends via Data Visualization is done for a company to make a right decision about is investment.

Assumption Made:

- There is no Cab Strike during this period of data collected.
- Data for Total Cab Time Duration and Timing of the Cab ride is not known. Therefore outliers are not removed from the Cost of the trip
- There is no Cab data for San Fransico, so that column in City data is not used.
- Profit, Profit/km, Cost/Km were some of the columns generated using existing data transformation.
- Some of the transactions had Cost of trip more than the Price charged, which is very unlikely thing to happen. Therefore the inaccurate data is removed.