Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 05/03/2021
Internship Batch: LISP01
Version:1.0
Data intake by Beril Chepkorir
Data intake reviewer:
Data storage location: https://github.com/dangerous2

Tabular data details:

Total number of observations	359392
Total number of files	4
Total number of features	16
Base format of the file	.csv
Size of the data	56.6+ MB

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

Proposed Approach:

Data Validation (Identification)
> I merged various datasets using common columns into a master data and analyzed it using various python techniques and visualization

Assumptions
Users feature of city dataset is treated as number of cab users in the city.
we have assumed that this can be other cab users as well (including Yellow and Pink cab)
We have outliers in our dataset
We assume that the profit is calculated by Price Charged -Cost of trip