Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 12.07.2022
Internship Batch: LISUM11
Version:1.0
Data intake by: Kristina Kaliagina
Data intake reviewer:
Data storage location: https://Kristina_Kaliagina@bitbucket.org/kristina_kaliagina/datasets.git

Tabular data details:

Total number of observations	848681
Total number of files	4
Total number of features	17
Base format of the file	.csv
Size of the data	30,45MB

Proposed Approach:
- Identification:
  - Time period of data is from 31/01/2016 to 31/12/2018.
  - The final data table consists of 359392 observations and 14 features, not counting the new ones that were created during the analysis.
  - Features and its types: Transaction ID (int), Date of Travel (datetime), Company (object), City (object), KM Travelled (float), Price Charged (float), Cost of Trip (float), Population (int), Users (int), Customer ID (int), Payment_Mode (object), Gender (object), Age (int), Income (USD/Month) (int).

- Assumptions:
  - The variable "Cost of Tript" includes all the company's expenses for the trip.
  - The "Customers ID" variable is part of the "Users" variable.
  - The new variable "Efficiency" is calculated using the following formula:
    $(Price\ Charged\ -\ Cost\ of\ Trip)/Price\ Charged$.
  - The client (investor) proceeds only from the provided data. Data that is not available does not matter for choosing a strategy.