Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 13 Feb 2024
Internship Batch: LISUM30
Version:1.0
Data intake by: Muhammad Nuril Huda
Data intake reviewer:
Data storage location:
https://github.com/MuhammadNurilHuda/G2M-Cab-Investment/blob/main/Dataset/Used_data.csv

https://github.com/MuhammadNurilHuda/G2M-Cab-Investment/blob/main/Dataset/City.csv

Tabular data details:

Total number of observations	359854
Total number of files	4
Total number of features	19
Base format of the file	.csv
Size of the data	53.9 MB

Total number of observations	21
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	759 Bytes

Proposed Approach:

- Used_data.csv is a combination of Cab_Data.csv, Customer_ID.csv, Transaction_ID.csv and US Holiday Dates (2004-2021).csv. Cab_Data.csv is transaction data. Customer_ID.csv data is left joined to Transaction_ID.csv, then the results are left joined to Cab_Data.csv. Then the results are combined with US Holiday Dates (2004-2021).csv data obtained open-source from Kaggle to get holiday data.
- City.csv contains the number of users and population in each city.
- Assumption for Data Quality Improvement: The Date of Travel column in Cab_Data.csv is Ms. Excel Serial Number, it's quite difficult to understand, the usage of datetime is more recommended