Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 09-05-2022
Internship Batch: LISUM09
Version: 0.1
Data intake by: Laâroussi Saâdeddine
Data intake reviewer: Laâroussi Saâdeddine
Data storage location: https://github.com/DataGlacier/DataSets

Tabular data details:

Cab_Data

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	20 663 Ko

City_Data

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	1 Ko

Customer_ID_Data

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	1027 Ko

Transaction_ID_Data

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	8788 Ko

HolidayUS_Data

Total number of observations	30
Total number of files	1
Total number of features	4
Base format of the file	csv
Size of the data	1 Ko

Proposed Approach:
- Cleaning data by checking null values and duplicate values
- Adding columns for Benefit, Benefit/KM, Age group, Income group, KM range, etc…
- Merging Cab_Data, Customer_ID_Data, Transaction_ID_Data. There is no need to merge City_Data or HolidayUS_Data to avoid redundancy of data
- Describing the data and finding correlation between numerical features to search for possible outliers
- Removing outliers
- Analyzing the data:
    - Finding the average profit per KM for each company in each year
    - Average profit per KM for each company in each month
    - Average profit per KM for each company in each day
    - Average profit per KM for each company in each day of the week
    - Profit per ride for each company
    - Profit Margin percentage year wise for each company
    - Total Profit per Gender each year for each company
    - Total Benefit per Payment Mode for each company
    - Benefit per City for each company
    - Benefit per age group for each company
    - Total Benefit per Income group for each company
    - Number of time customers returned (Customer retention)
    - Total Benefit per KM Travelled range
    - Total Benefit per Holiday
- Giving a recommendation in which company to invest