Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 14-May-2023
Internship Batch: LISUM21
Version:
Data intake by: Aditi Dadariya
Data intake reviewer: Data Glacier
Data storage location: https://github.com/aditidadariya/G2MInsightForCabInvestmentFirm

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	Cab_Data.csv
Size of the data	21.2 MB

Tabular data details: City

Total number of observations	20
Total number of files	1
Total number of features	3
Base format of the file	City.csv
Size of the data	759 bytes

Tabular data details: Customer_ID

Total number of observations	49171
Total number of files	1
Total number of features	4
Base format of the file	Customer_ID.csv
Size of the data	1.1 MB

Tabular data details: Transaction_ID

Total number of observations	440098
Total number of files	1
Total number of features	3
Base format of the file	Transaction_ID.csv
Size of the data	9 MB

Proposed Approach:
- There were no missing values in any of the datasets.
- Downloaded US Holidays dataset from Kaggle. Filtered the dates from 2016 to 2018, to understand the days off during these years. Details of Holiday dataset is below.
- **Tabular data details: US Holiday Dates (2004-2021)**

Total number of observations	57
Total number of files	1
Total number of features	6
Base format of the file	US Holiday Dates (2004-2021).csv
Size of the data	16 KB