Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 7th August 2022
Internship Batch: LISUM12
Version: 1.0
Data intake by: Dhvanil Prajapati
Data intake reviewer: NA
Data storage location:
1 to 4. https://github.com/DataGlacier/DataSets
5. US holiday dates: https://www.kaggle.com/datasets/donnetew/us-holiday-dates-2004-2021/download?datasetVersionNumber=1

Tabular data details:

1. Cab_Data.csv:

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 KB

2. City.csv:

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 KB

3. Customer_ID.csv:

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 KB

4. Transaction_ID.csv:

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 KB

5. US Holiday Dates

Total number of observations	342
Total number of files	1
Total number of features	6
Base format of the file	Csv
Size of the data	16 KB

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

Proposed Approach:
- Mention approach of dedup validation (identification)
- Mention your assumptions (if you assume any other thing for data quality analysis)