

Data Intake Report

Name: G2M insight for Cab Investment Firm

Report date: 14th June 2023

Internship Batch: LISUM22

Version:<1.0>

Data intake by: Pavan Walvekar

Data intake reviewer: NA

Data storage location: <https://github.com/DataGlacier/DataSets> ,
<https://www.kaggle.com/datasets/donnetew/us-holiday-dates-2004-2021>

Tabular data details:

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

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

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

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

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

Proposed Approach:

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