

# Data Intake Report

Name: <G2M Insight fir Cab Investment firm >

Report date: <14/03/2024>

Internship Batch:<LISUM31>

Version:<1.0>

Data intake by:<Bilikis Alayo>

Data intake reviewer:<intern who reviewed the report>

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

## Tabular data details:

File name: Cab\_Data.scv

<b>Total number of observations</b>	<359392>
<b>Total number of files</b>	<1>
<b>Total number of features</b>	<8>
<b>Base format of the file</b>	<.csv>
<b>Size of the data</b>	<2.9 MB>

File name: City.csv

<b>Total number of observations</b>	<20>
<b>Total number of files</b>	<1>
<b>Total number of features</b>	<3>
<b>Base format of the file</b>	<.csv>
<b>Size of the data</b>	<608 bytes>

File name: Customer\_ID.csv

<b>Total number of observations</b>	<49171>
<b>Total number of files</b>	<1>
<b>Total number of features</b>	<4>
<b>Base format of the file</b>	<.csv>
<b>Size of the data</b>	<1.5 MB>

File name: Transaction\_ID.csv

<b>Total number of observations</b>	<440098>
<b>Total number of files</b>	<1>
<b>Total number of features</b>	<3>
<b>Base format of the file</b>	<.csv>
<b>Size of the data</b>	<10 MB>

File name: UsHoliday.csv

<b>Total number of observations</b>	<342>
<b>Total number of files</b>	<1>
<b>Total number of features</b>	<6>
<b>Base format of the file</b>	<csv>
<b>Size of the data</b>	<16.2 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)

**Note: Convert this doc in pdf and provide the link of pdf file in your dashboard.  
Please do not forget to remove this section while converting the file into pdf.**