

Data Intake Report

Name: <G2M>

Report date: <Fri 6 January 2023>

Internship Batch:<LISUM17 >

Version:<1.0>

Data intake by:<Gangxin Li>

Data intake reviewer:<UNKNOWN>

Data storage location: <<https://github.com/DataGlacier/DataSets>>

Tabular data details:

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

Total number of observations	<440097>
Total number of files	<4>
Total number of features	<3>
Base format of the file	<.csv >
Size of the data	<9 MB>

Total number of observations	<18>
Total number of files	<4>
Total number of features	<3>
Base format of the file	<.csv >
Size of the data	<4 KB>

Total number of observations	<359391>
Total number of files	<4 >
Total number of features	<7>
Base format of the file	<.csv >
Size of the data	<21.2MB>

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.**