Data Intake Report

Name: G2M Case Study
Report date: 07/05/2022
Internship Batch: LISUM11
Version: 1.0
Data intake by: Anthony Ghimpu
Data intake reviewer: (Individual Assignment)
Data storage location: Computer Disk

Tabular data details:

Total number of observations	359,392
Total number of files	1
Total number of features	7
Base format of the file	.csv
Size of the data	20.1 mb

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	4.00 KB

Total number of observations	49172
Total number of files	1
Total number of features	4
Base format of the file	.csv
Size of the data	1.00 mb

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	8.58 MB

Proposed Approach:

- Mention approach of dedup validation (identification)
    - Every observation will be checked for duplication in an iterative manner. If any duplicates arise, they will be dropped from the dataset.
- Mention your assumptions (if you assume any other thing for data quality analysis)
    - There are no outliers within the dataset. (Data is not skewed)
    - All missing data will result in dropping of observations.