

Data Intake Report

Name: G2M Case Study

Report date: 20-Jun-2021

Internship Batch: LISUM01

Version:

Data intake by: Yousef Elbayoumi

Data intake reviewer:

Data storage location: https://github.com/yousefhosam-b/G2M_Cab_Task.git

Tabular data details:

Total number of observations	359392 rows
Total number of files	1 file
Total number of features	19 columns
Base format of the file	csv file
Size of the data	51.8 MB

Tabular data details:

Total number of observations	57 rows
Total number of files	1 file
Total number of features	6 columns
Base format of the file	csv file
Size of the data	3 KB

Proposed Approach:

- As I merged the given datasets and included one more about “Holidays”, there was a lot merging going on, therefore I had some duplicated values, I solved them with python, also I worked sometimes in Excel to change columns names or date format because it's faster and more accurate.
- I created new columns by doing some calculations or by just separating one column to several ones, that helped me to manage the analysis in faster way and to have more detailed results.
- I couldn't share my main dataset to GitHub as it's more than 25 MB, I can share somewhere else if it's needed.