

Data Intake Report

Name: <Data Analyst>

Report date: <08/12/2023>

Internship Batch:<LISUM23:30>

Version:<1.0>

Data intake by:<Dingyun hu>

Data intake reviewer:<intern who reviewed the report>

Data storage location: <<https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store?select=2019-Nov.csv>>

Tabular data details:

Total number of observations	<67501979>
Total number of files	<1>
Total number of features	<9>
Base format of the file	<.csv>
Size of the data	<9.01GB>

Note: Replicate same table with file name if you have more than one file.

Proposed Approach:

- Pandas and Dask are used to load the data
- Removed white spaces.