

Data Intake Report

Name: eCommerce Behavior File Ingestion and Schema Validation

Report date: February 6, 2023

Internship Batch: LISUM17

Version: 1.0

Data intake by: Elissa Kuon

Data intake reviewer: Data Glacier

Data storage location: https://github.com/ekuon/Data_Glacier_Internship/tree/main/Week%206

Tabular data details: eCommerce.csv

Total number of observations	42448764
Total number of files	1
Total number of features	9
Base format of the file	CSV
Size of the data	5.67 GB

Tabular data details: eCommerce.csv.gz

Total number of observations	N/A
Total number of files	89
Total number of features	N/A
Base format of the file	CSV.GZ
Size of the data	1.69 GB

Proposed Approach:

- No assumptions are made as we are focused on file ingestion and schema validation due to how big the data is