

Data Intake Report

Name: Data Ingestion

Report date: 12 December 2023

Internship Batch: LISUM27

Version:1.0

Data intake by: Mufunwa Nemushungwa

Data intake reviewer:

Data storage location:

Tabular data details:

Total number of observations	22489348
Total number of files	
Total number of features	11
Base format of the file	.csv
Size of the data	2.41 GB

Proposed Approach:

- The deduplication validation approach is methodical and comprehensive. Within the dataset, the uniqueness of each Transaction unique identifier is verified, aiming to detect any inconsistencies or duplications.
- The dataset has unique identifiers and contains minimal missing points.