Data Intake Report

Name: Week 6: File ingestion and schema validation
Report date: 06/12/2024
Internship Batch: LISUM33
Version: 1.0
Data intake by: Jackson Taylor
Data intake reviewer:
Data storage location:

Tabular data details:

price_paid_records.csv

Total number of observations	22,489,348
Total number of files	
Total number of features	11
Base format of the file	.csv
Size of the data	2,405,685,902 B

Proposed Approach:

- All rows are used regardless of potential duplicate rows or values. No dedup validation is used. No assumptions are made about the data columns. All that matters for this assignment is that the size of the dataset is larger than 2 GB, the column names can be read, every record has fields (including empty fields) equal to the number of columns, and the number of data records can be found.