

Data Intake Report

Name: Data Ingestion Pipeline

Report date: July 11, 2022

Internship Batch: LISUM 10

Version:<1.0>

Data intake by: Jeeyeon Shon

Data intake reviewer:

Data storage location: <https://www.kaggle.com/datasets/hm-land-registry/uk-housing-prices-paid>

Tabular data details:

price_paid_records.csv

Total number of observations	22489348
Total number of files	1
Total number of features	11
Base format of the file	.csv
Size of the data	2.41 GB

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

Proposed Approach:

- No duplicates or missing data
- The file is large in size: find other ways than pandas for faster reading