

# Data Intake Report

Name: Data Ingestion Pipeline

Report date: 16-Jul-2021

Internship Batch: LISUM01

Data intake by: Yousef Elbayoumi

Data intake reviewer: Data Glacier team

Data storage location: [https://github.com/yousefhosam-b/DataGlacier\\_Internship/tree/main/Task6](https://github.com/yousefhosam-b/DataGlacier_Internship/tree/main/Task6)

## Tabular data details:

<b>Total number of observations</b>	22489348 rows
<b>Total number of files</b>	1 file
<b>Total number of features</b>	11 columns
<b>Base format of the file</b>	csv file
<b>Size of the data</b>	2.35 GB

## Proposed Approach:

- I found this dataset on Kaggle, it's under the name "UK Housing Prices Paid", the link of the Kaggle page is:  
[https://www.kaggle.com/hm-land-registry/uk-housing-prices-paid?select=price\\_paid\\_records.csv](https://www.kaggle.com/hm-land-registry/uk-housing-prices-paid?select=price_paid_records.csv)
- Reading the dataset was almost impossible on my pc in frameworks such as Spyder or PyCharm due to the huge size of data set, therefore I used Jupyter Notebook to read it and complete the other tasks.