Data Intake Report

Name: <Paris Housing Price Prediction>
Report date: <28/05/2021>
Internship Batch:< LISUM09>
Version:<1.0>
Data intake by:<Fahad Alothman>
Data intake reviewer:<NA>
Data storage location: <https://www.kaggle.com/datasets/mssmartypants/paris-housing-price-prediction>

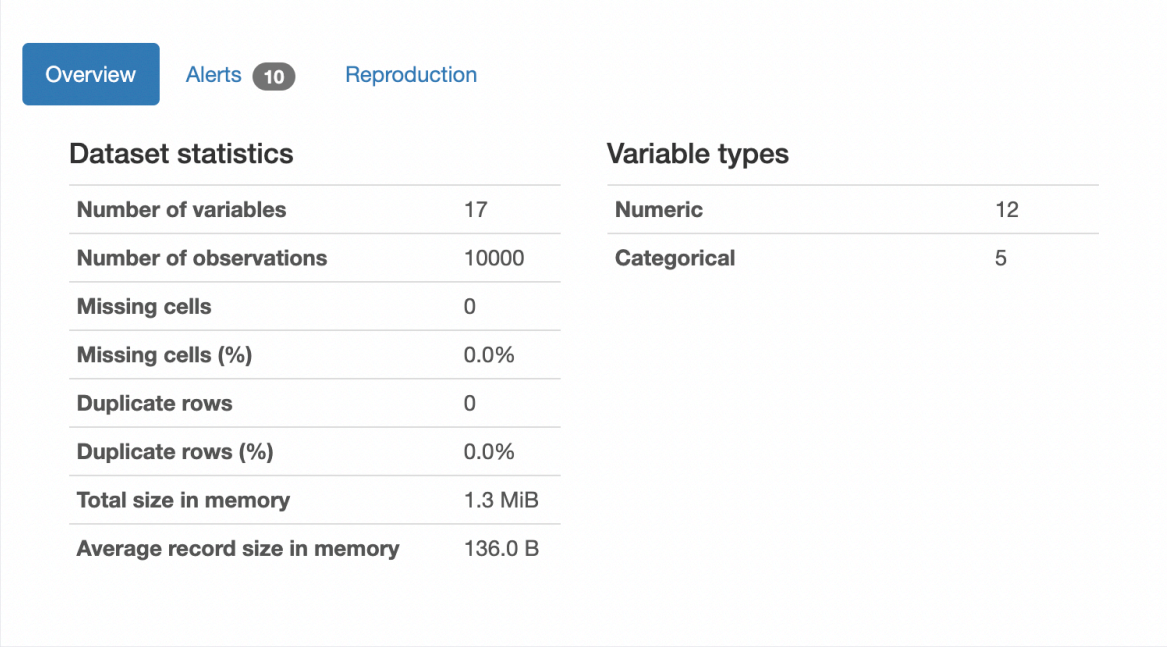
Tabular data details:

Total number of observations	10,000
Total number of files	1
Total number of features	17
Base format of the file	ParisHousing.csv
Size of the data	633 KB

Proposed Approach:

- Mention approach of dedup validation (identification):

There were no missing, inconsistent nor redundant rows in the dataset
As shown below using pandas_profiling:



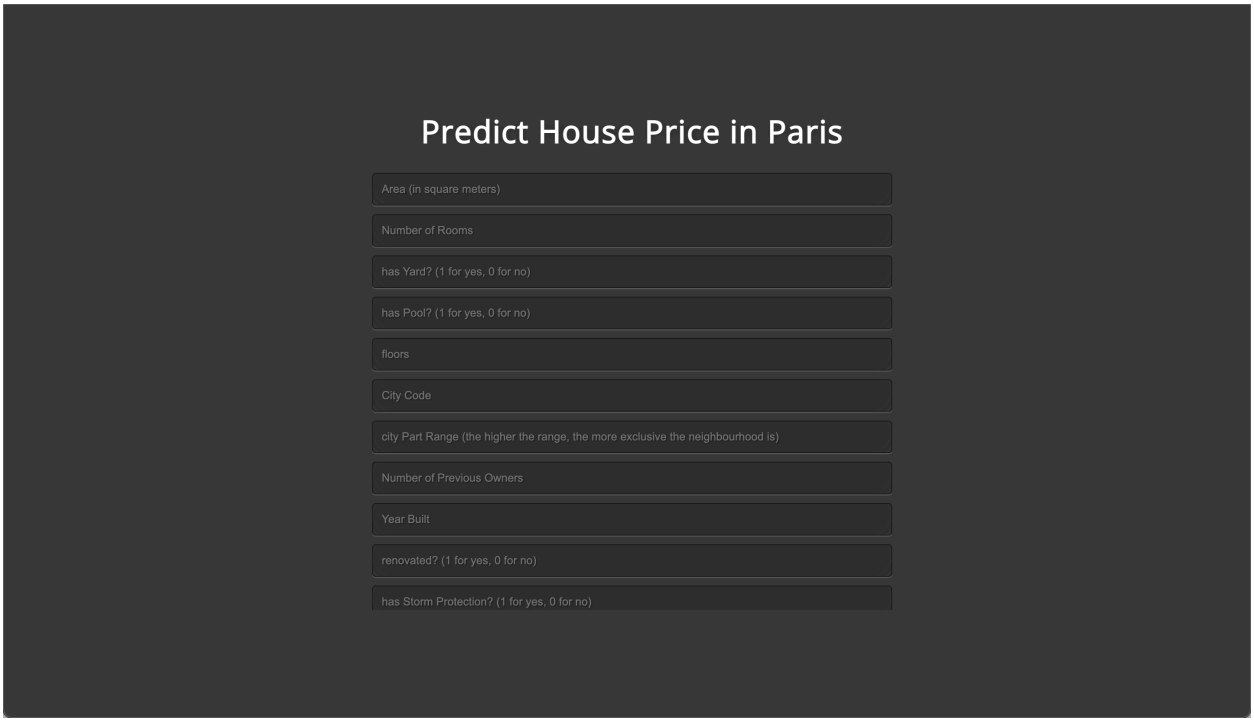
- Mention your assumptions (if you assume any other thing for data quality analysis)

1. Only 8 out of the 16 Features were required to build the linear regression Model.
2. The Root Mean Square Error chosen was 1891. That might seem high but if we take into account mean of the feature we want to predict (the price of the house) and the standard deviation, which are 4993448.0 & 2877424.0 respectively, we can safely assume that an RMSE of 1891 is appropriate and would be effective in building an accurate Model.

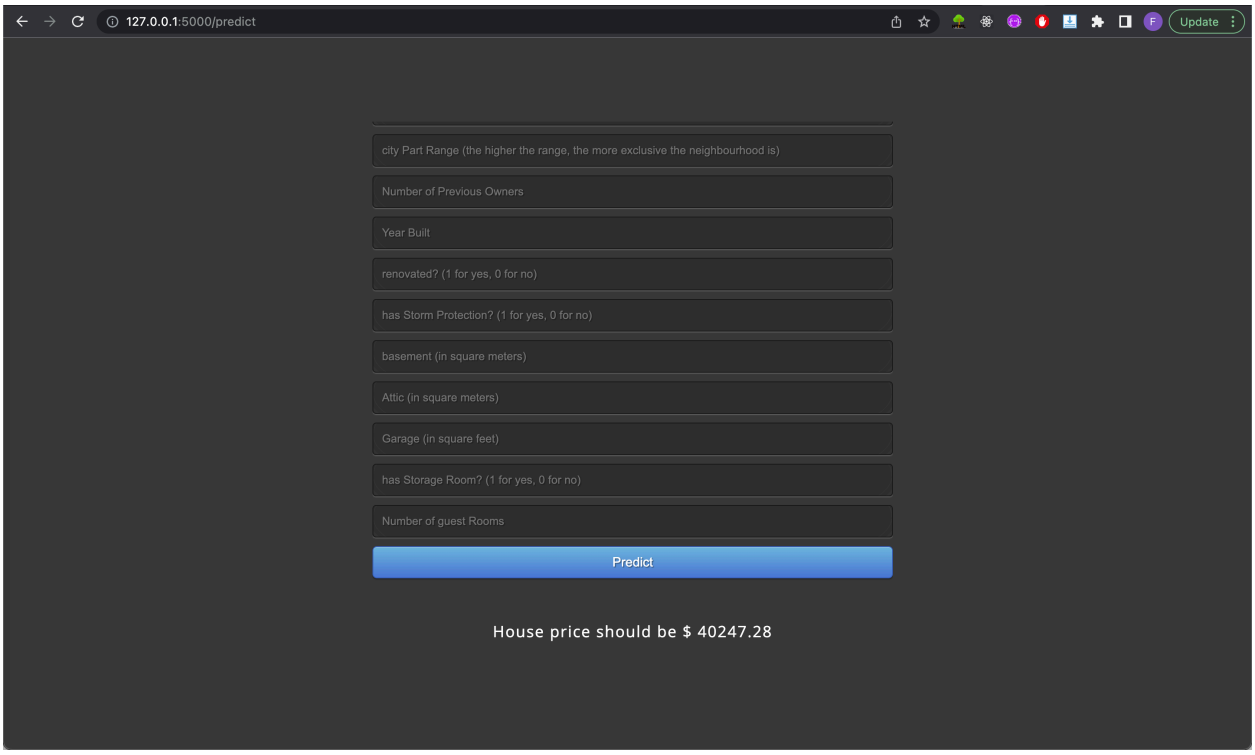
Deployment Steps:

1. Deploying the flask App

A. The user is expected to enter all 16 features of the house to predict the price (Note: I think we can reduce the inputs only to 8. But to more accurately predict the price, we decided to use all 16 features)



B. The prediction is then displayed as shown below:



2. Building and Saving the Model