

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:

<div>Overview Alerts 10 Reproduction</div>			
Dataset statistics		Variable types	
Number of variables	17	Numeric	12
Number of observations	10000	Categorical	5
Missing cells	0		
Missing cells (%)	0.0%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	1.3 MiB		
Average record size in memory	136.0 B		

- 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)

Predict House Price in Paris

Area (in square meters)

Number of Rooms

has Yard? (1 for yes, 0 for no)

has Pool? (1 for yes, 0 for no)

floors

City Code

city Part Range (the higher the range, the more exclusive the neighbourhood is)

Number of Previous Owners

Year Built

renovated? (1 for yes, 0 for no)

has Storm Protection? (1 for yes, 0 for no)

B. The prediction is then displayed as shown below:

The screenshot shows a web browser window with the address bar displaying "127.0.0.1:5000/predict". The page has a dark background. A form with 11 input fields is centered on the page. The fields are labeled as follows:

- city Part Range (the higher the range, the more exclusive the neighbourhood is)
- Number of Previous Owners
- Year Built
- renovated? (1 for yes, 0 for no)
- has Storm Protection? (1 for yes, 0 for no)
- basement (in square meters)
- Attic (in square meters)
- Garage (in square feet)
- has Storage Room? (1 for yes, 0 for no)
- Number of guest Rooms

Below the form is a blue button labeled "Predict". Below the button, the prediction result is displayed: "House price should be \$ 40247.28".

2. Building and Saving the Model