

Digital Empowerment Pakistan

“Project”

|  |  |
| --- | --- |
| Intern Name | Muhammad Attaullah |
| DEP Number | DEP935 |
| Batch | July 2024 |

Topic:

“Predicting House Price”

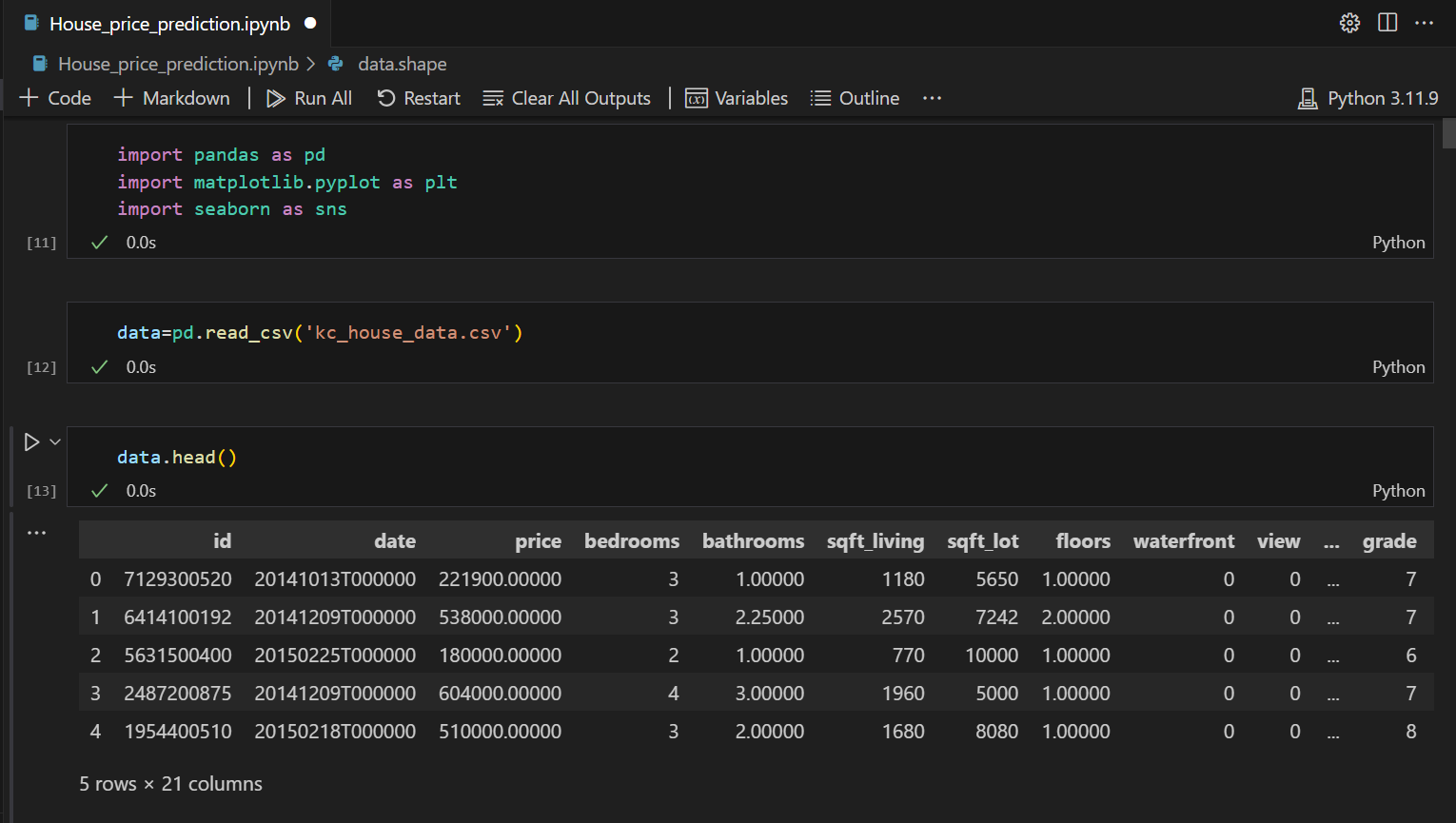
Introduction:

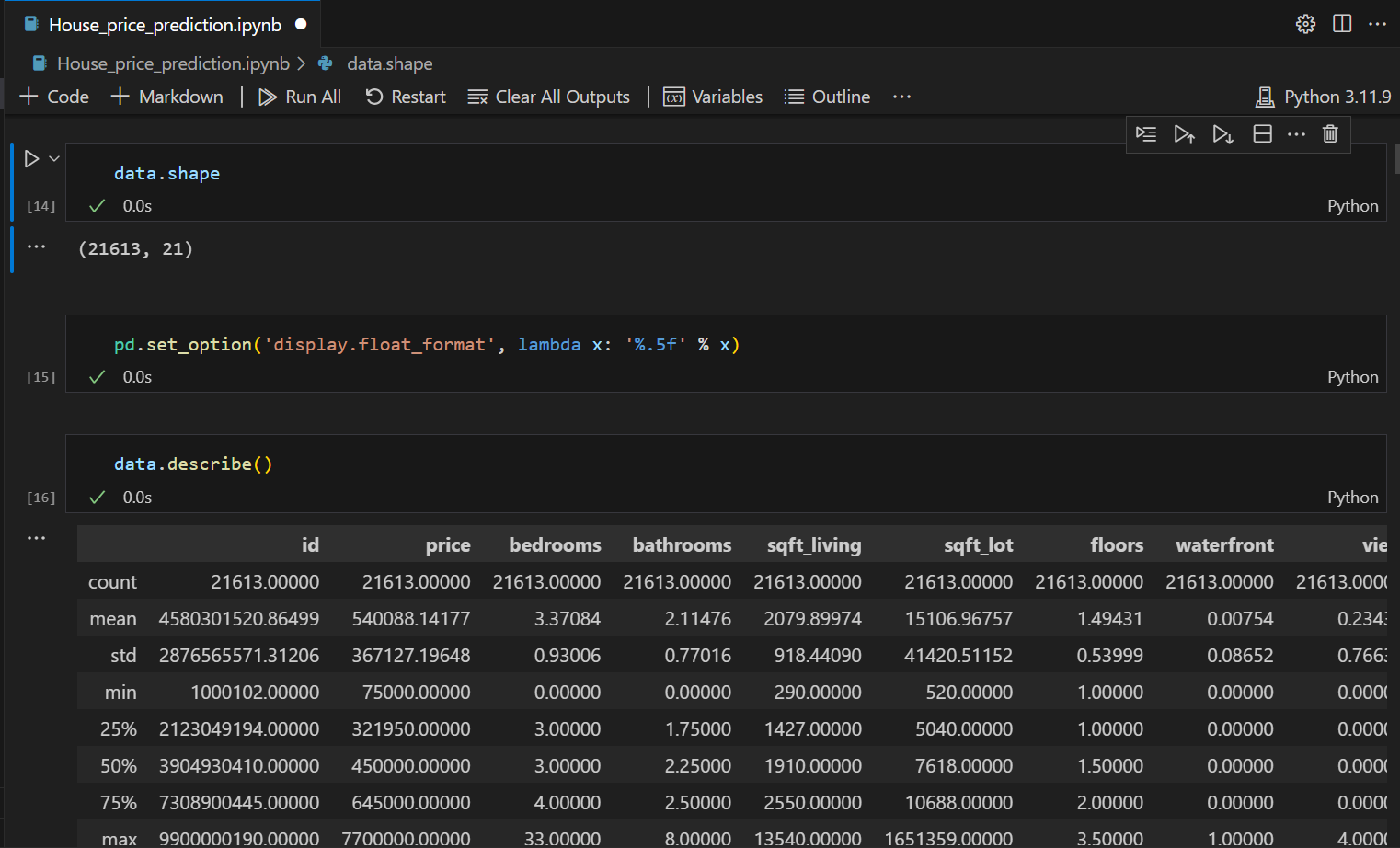
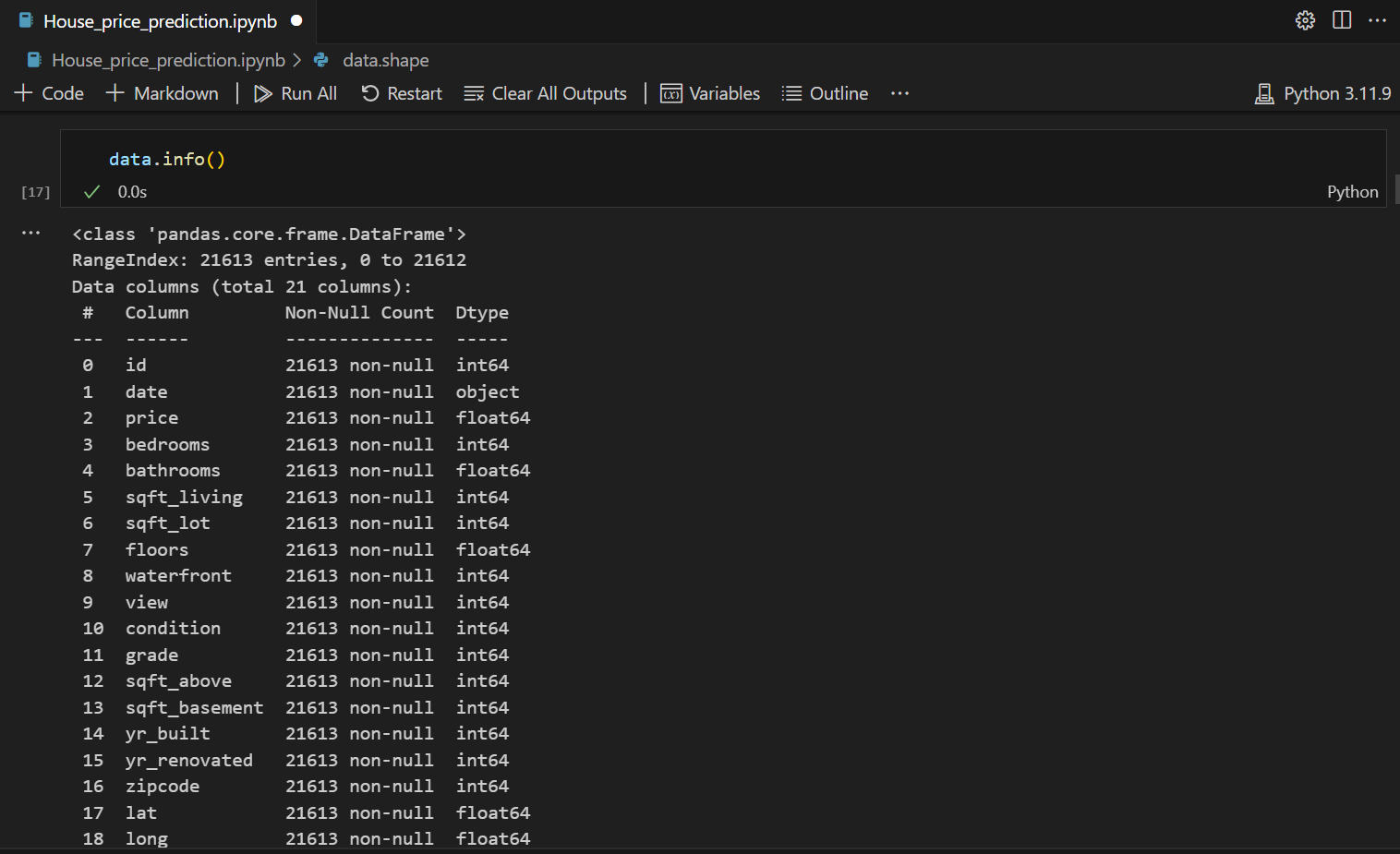
Predicting house prices using machine learning involves building a model that can estimate the value of a property based on various features. In this project we predict the price of house base on secondary data. There are many steps that are used to make data clean and then we train model base on the clean data.

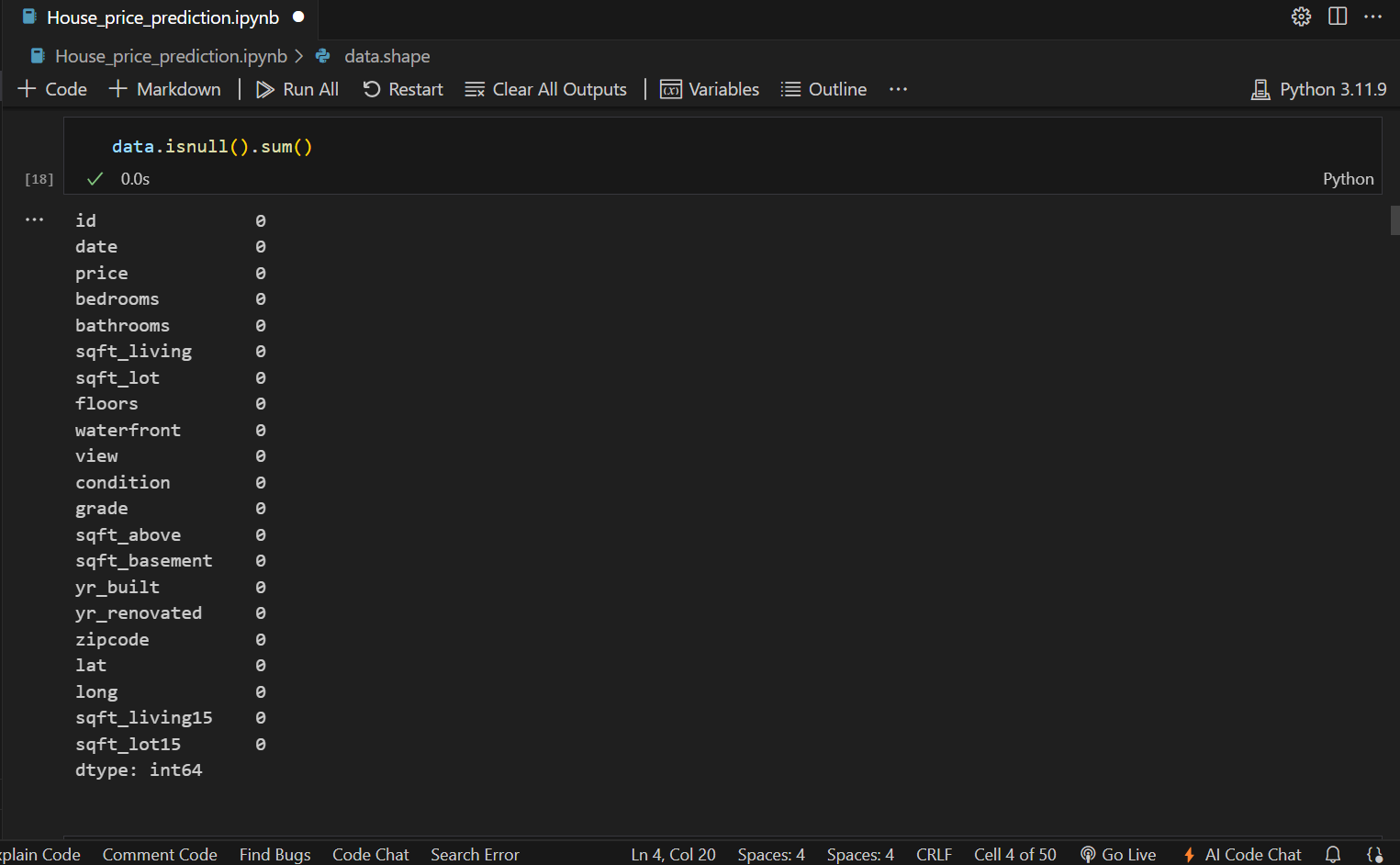
1. EDA and Preprocessing:

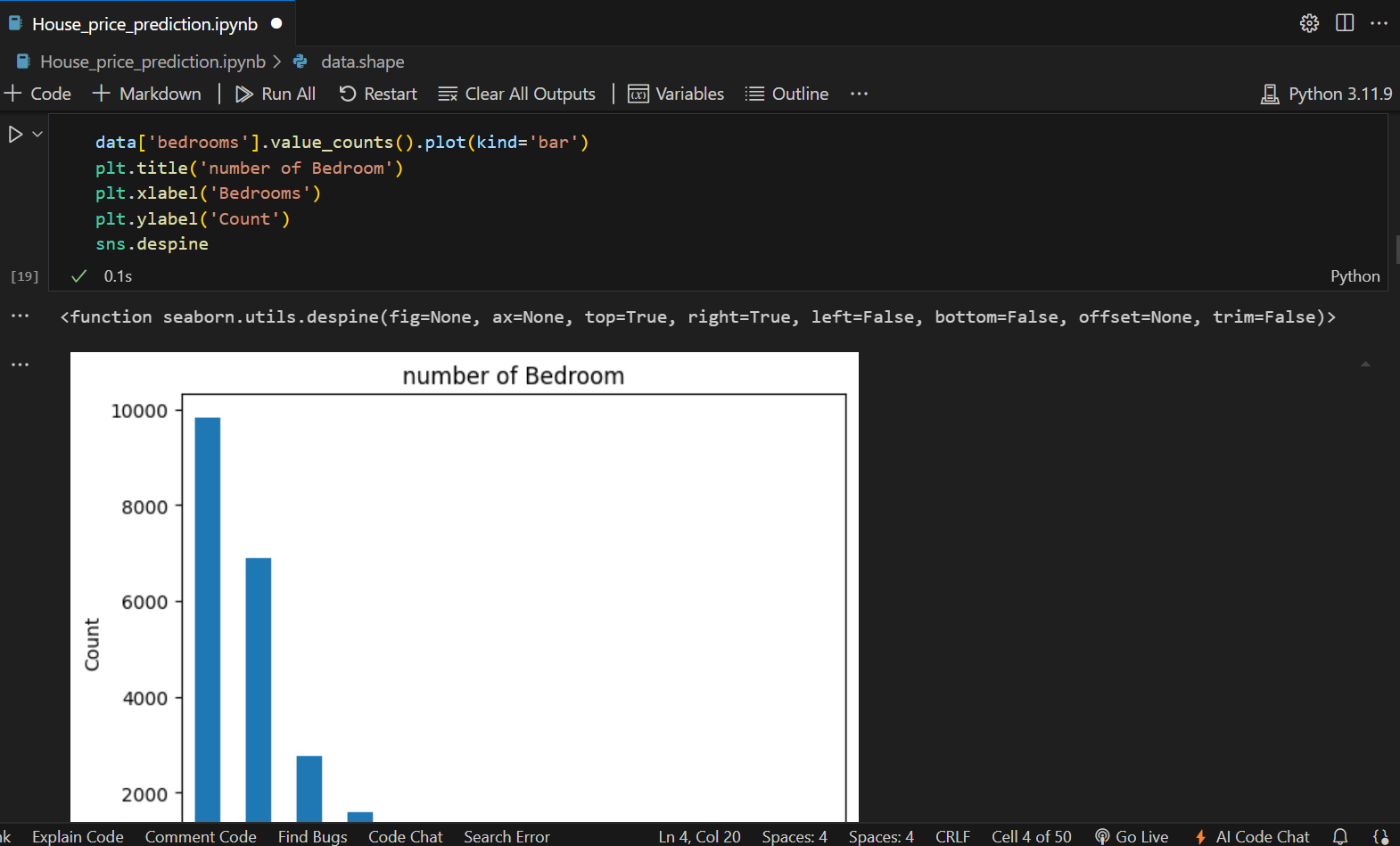
Exploratory Data Analysis (EDA) is a crucial step in the data analysis process where analysts use statistical graphics and other data visualization methods to examine datasets and summarize their main characteristics, often with the help of visual tools.

Preprocessing is a critical step in the data analysis and machine learning pipeline. It involves preparing and transforming raw data into a clean and usable format. Effective preprocessing improves the performance of machine learning models by ensuring that the data is consistent, relevant, and free of errors or biases.

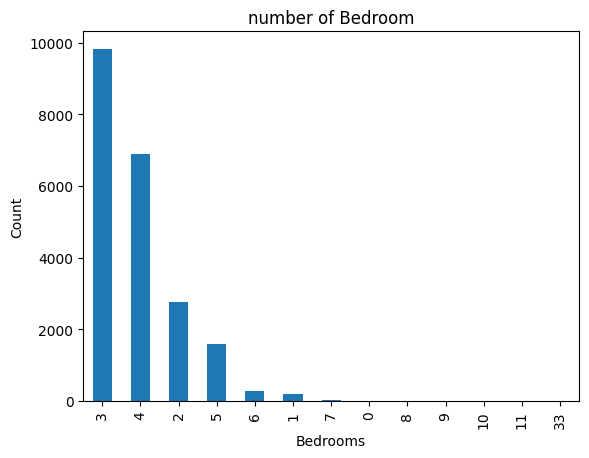


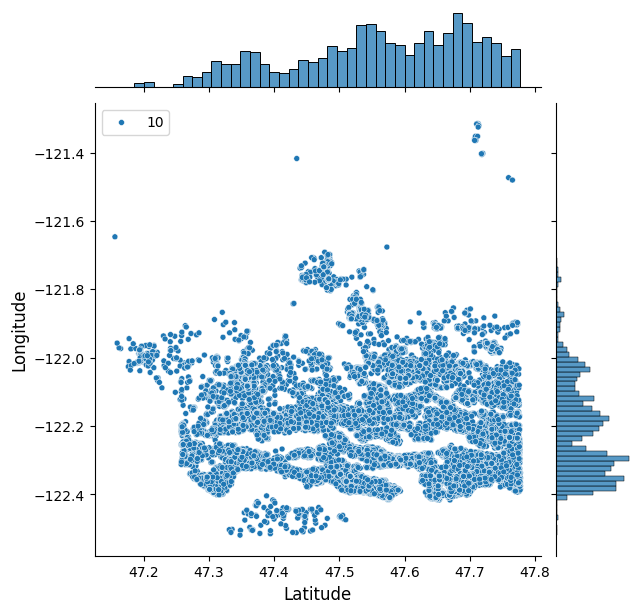


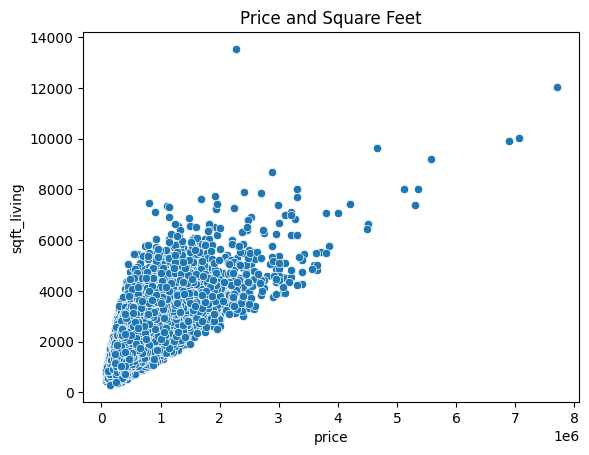


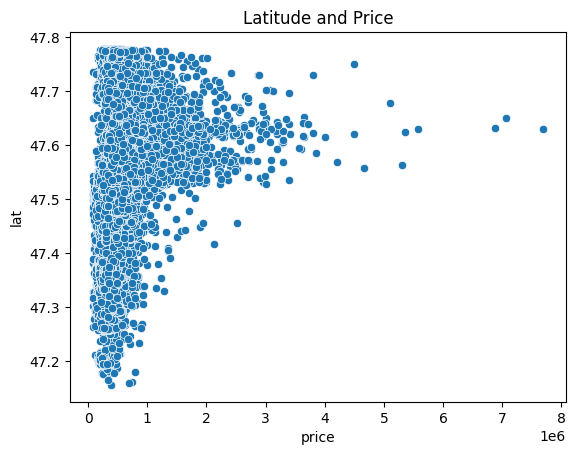


**Graphs:**

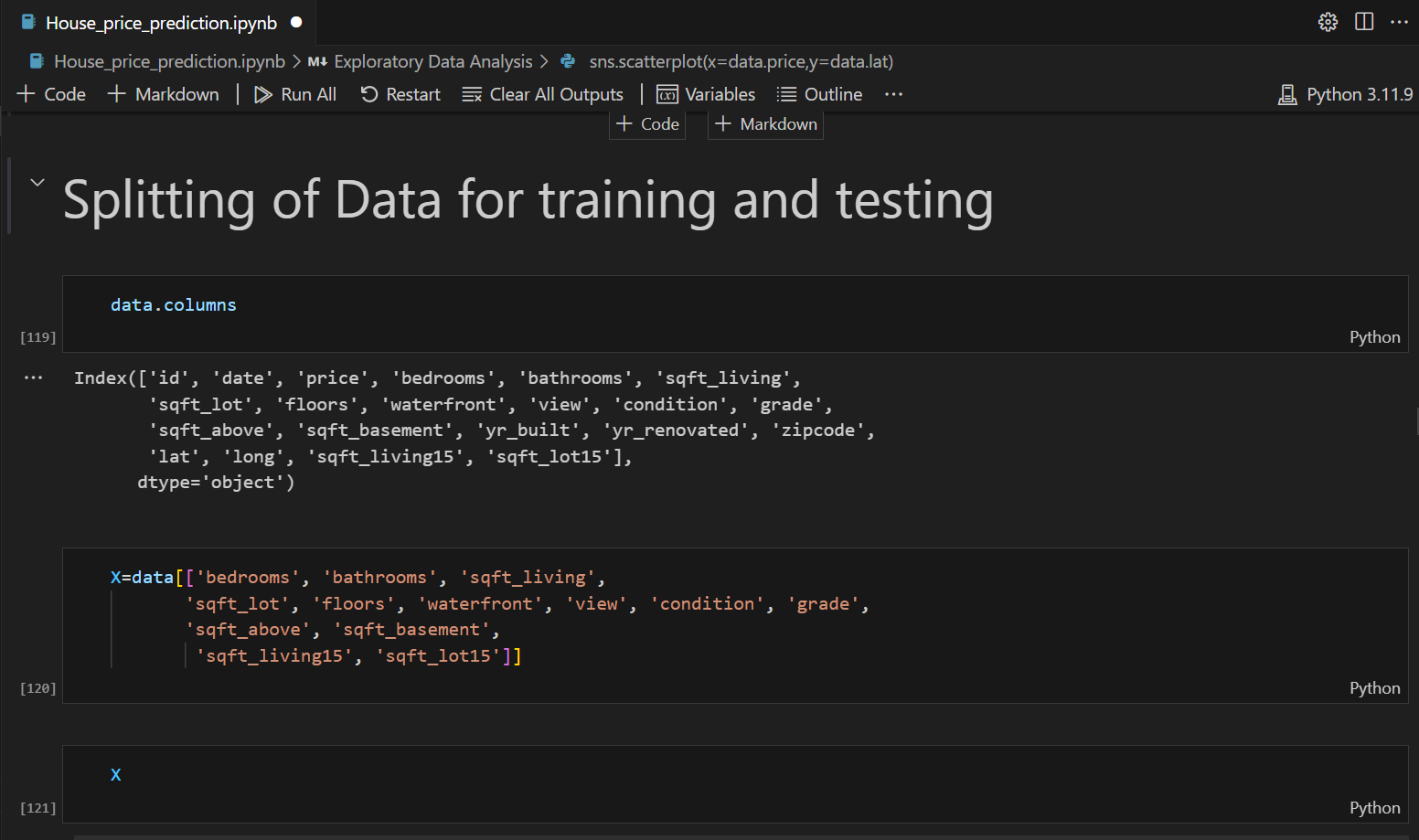
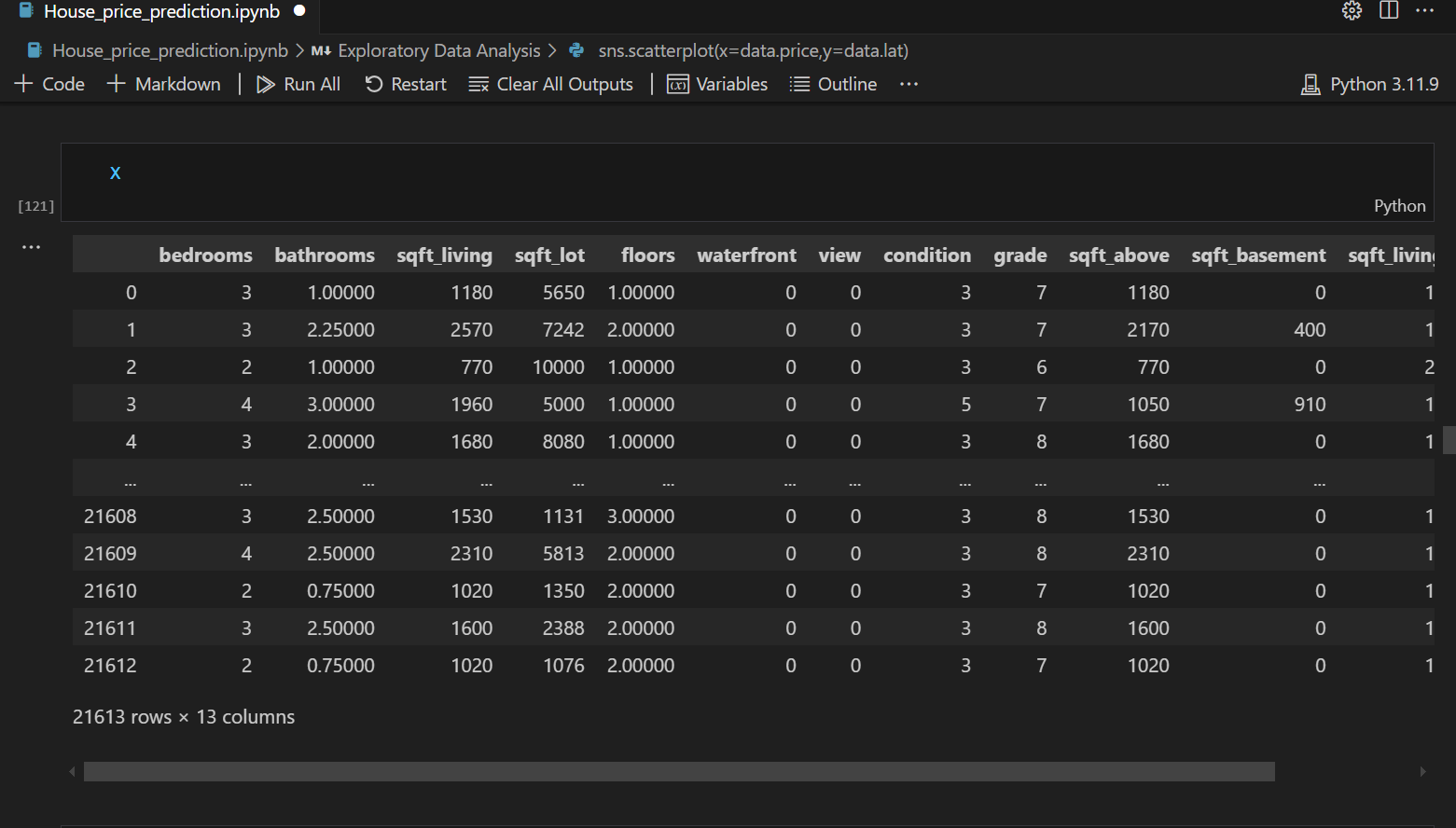
****

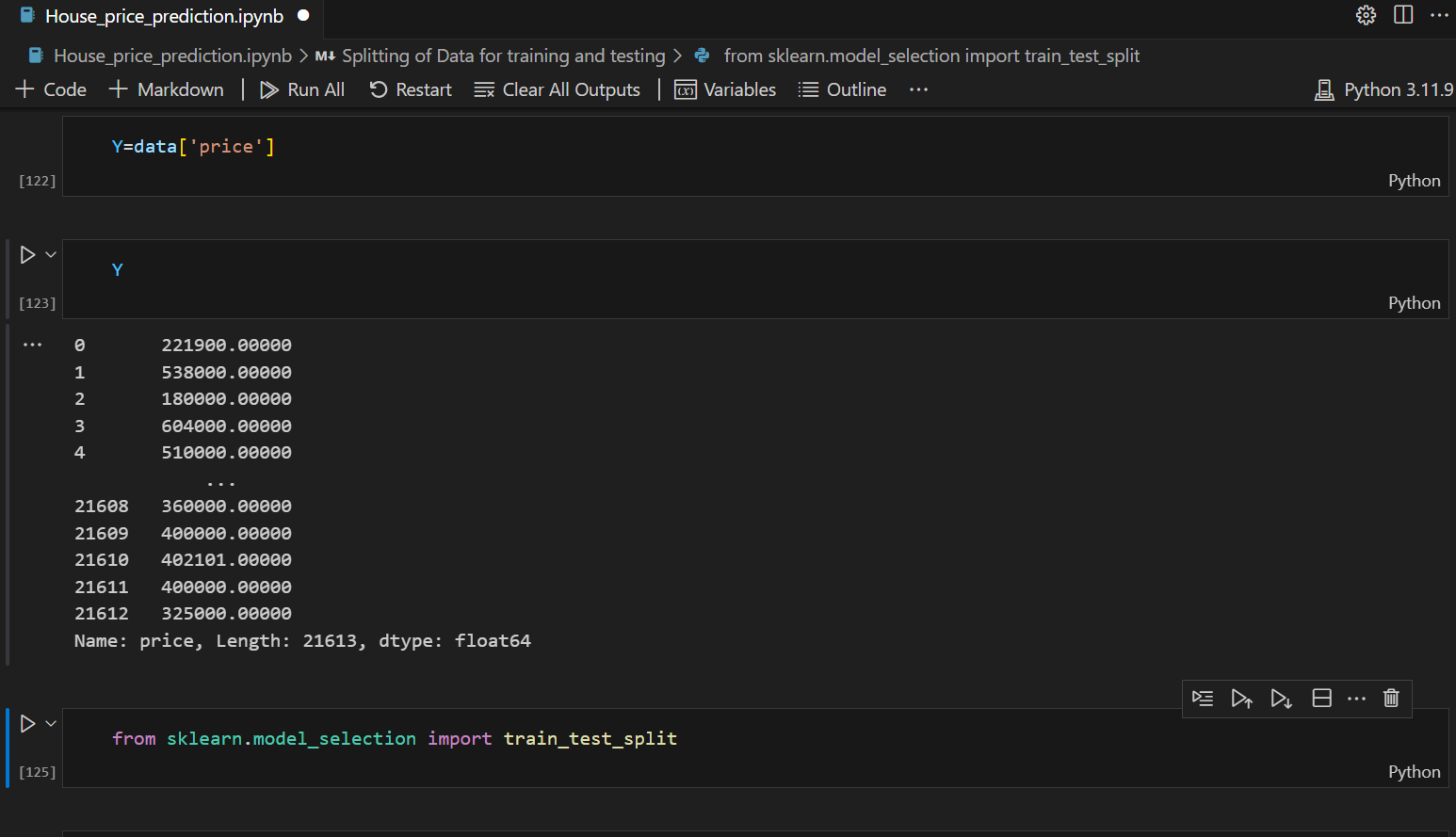


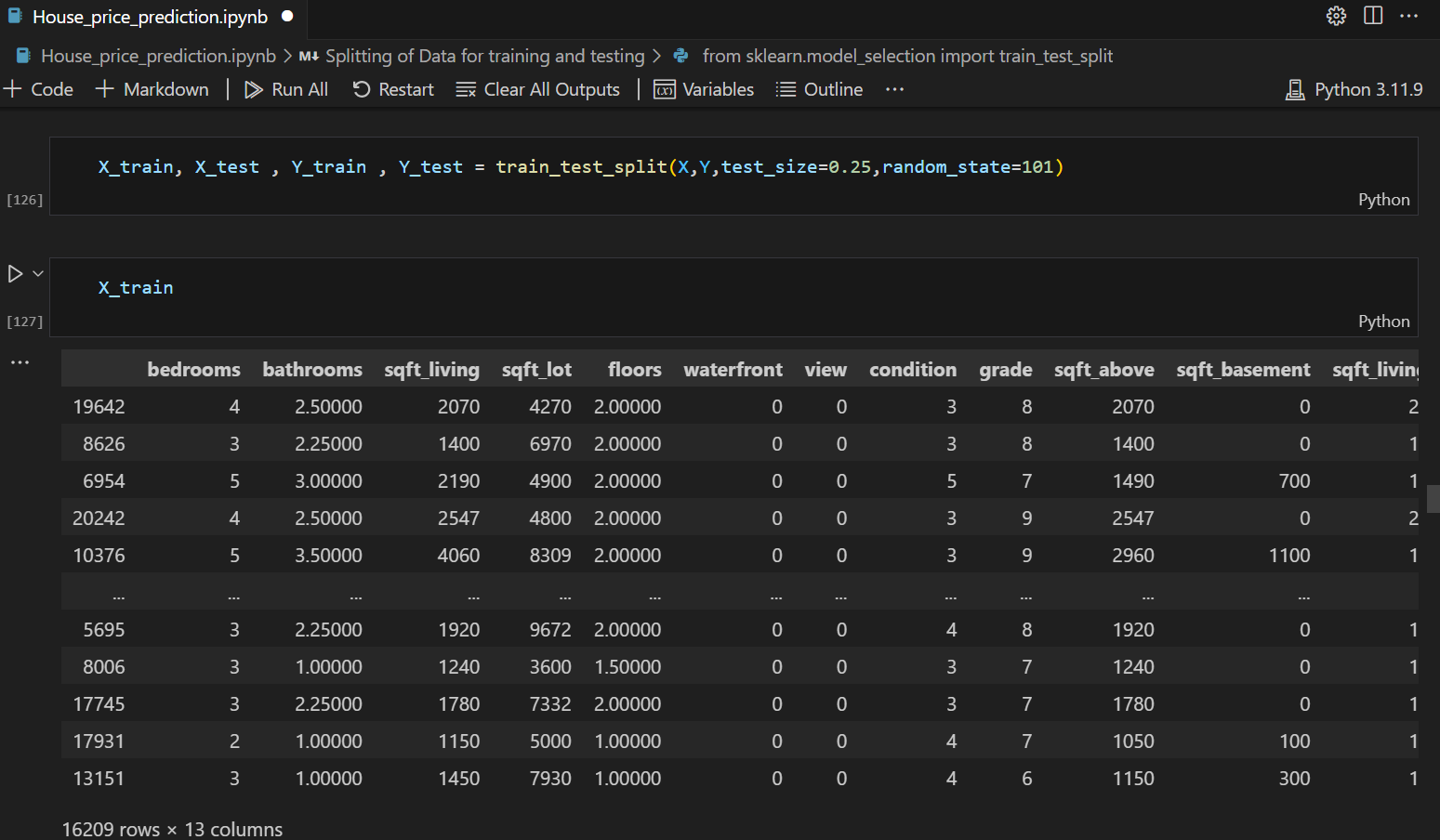


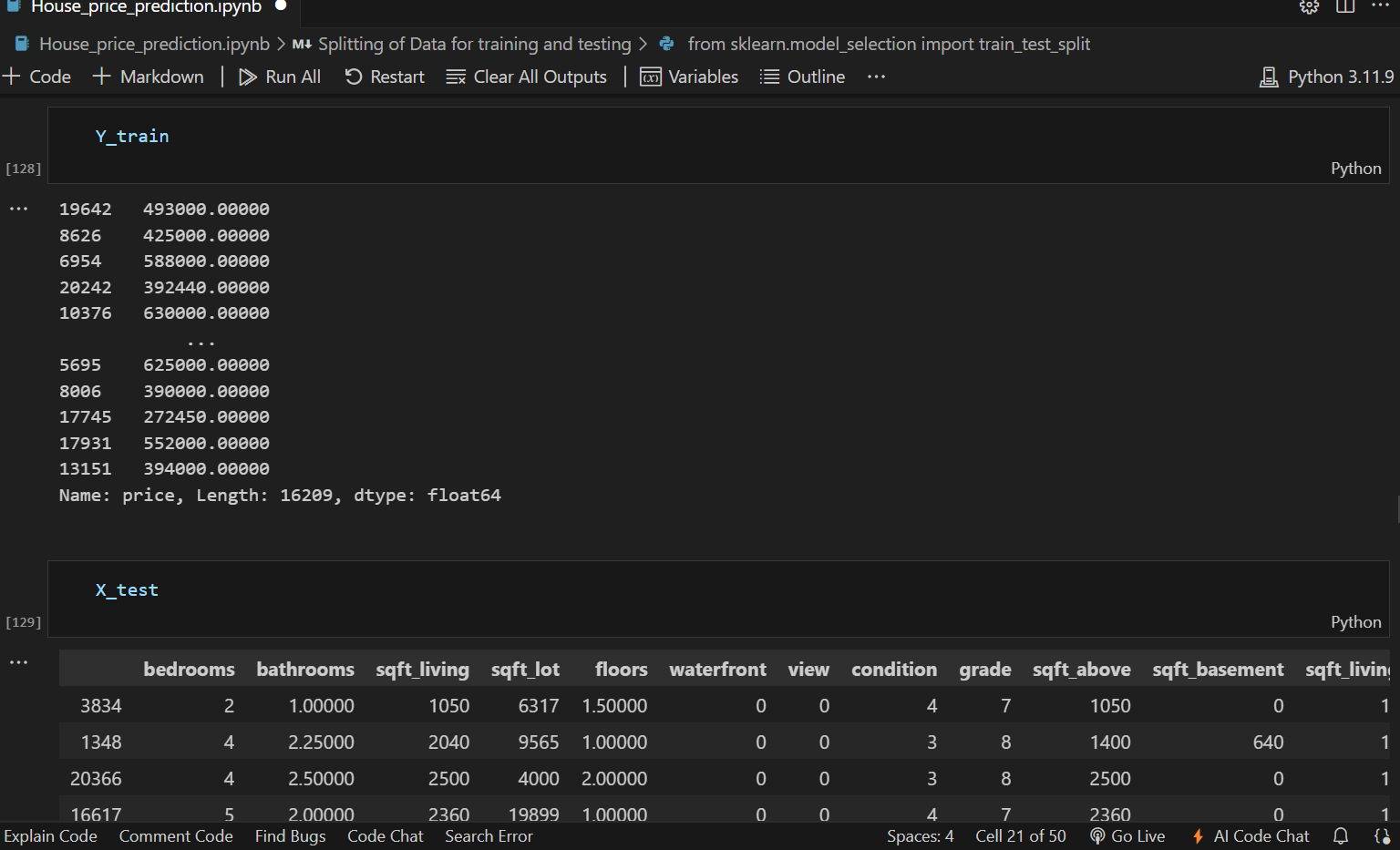


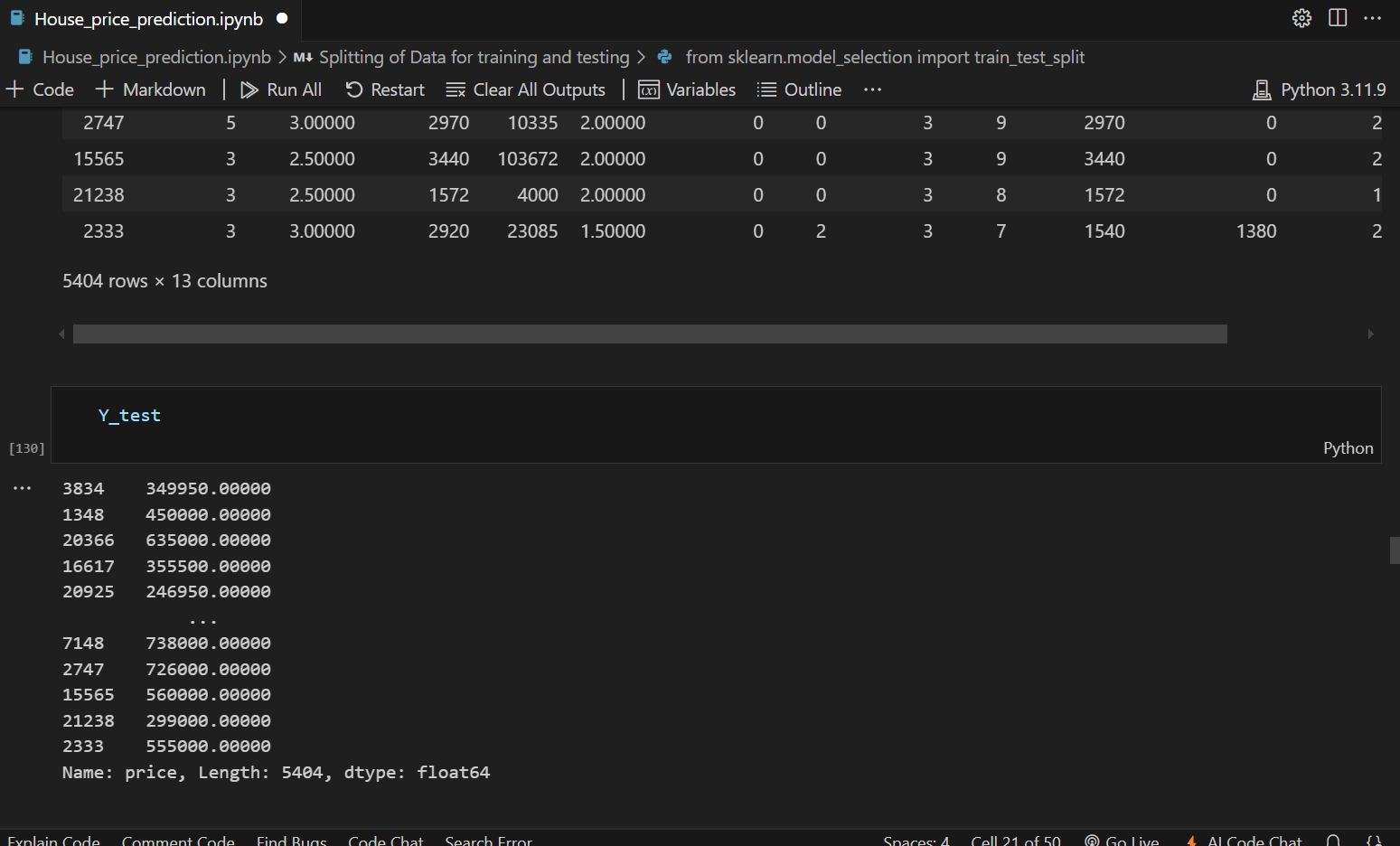
**Splitting of Data:**

In this we divide our data into two variables called x and y. The x variables basically that data, we give our model to learn from it and y data is that data, gives us the prediction.we also create some new variables that is easily understandable for machine is x\_train,y\_train,x\_test,y\_test. After this we apply standardition matrics on our data. Let’s see in the code:



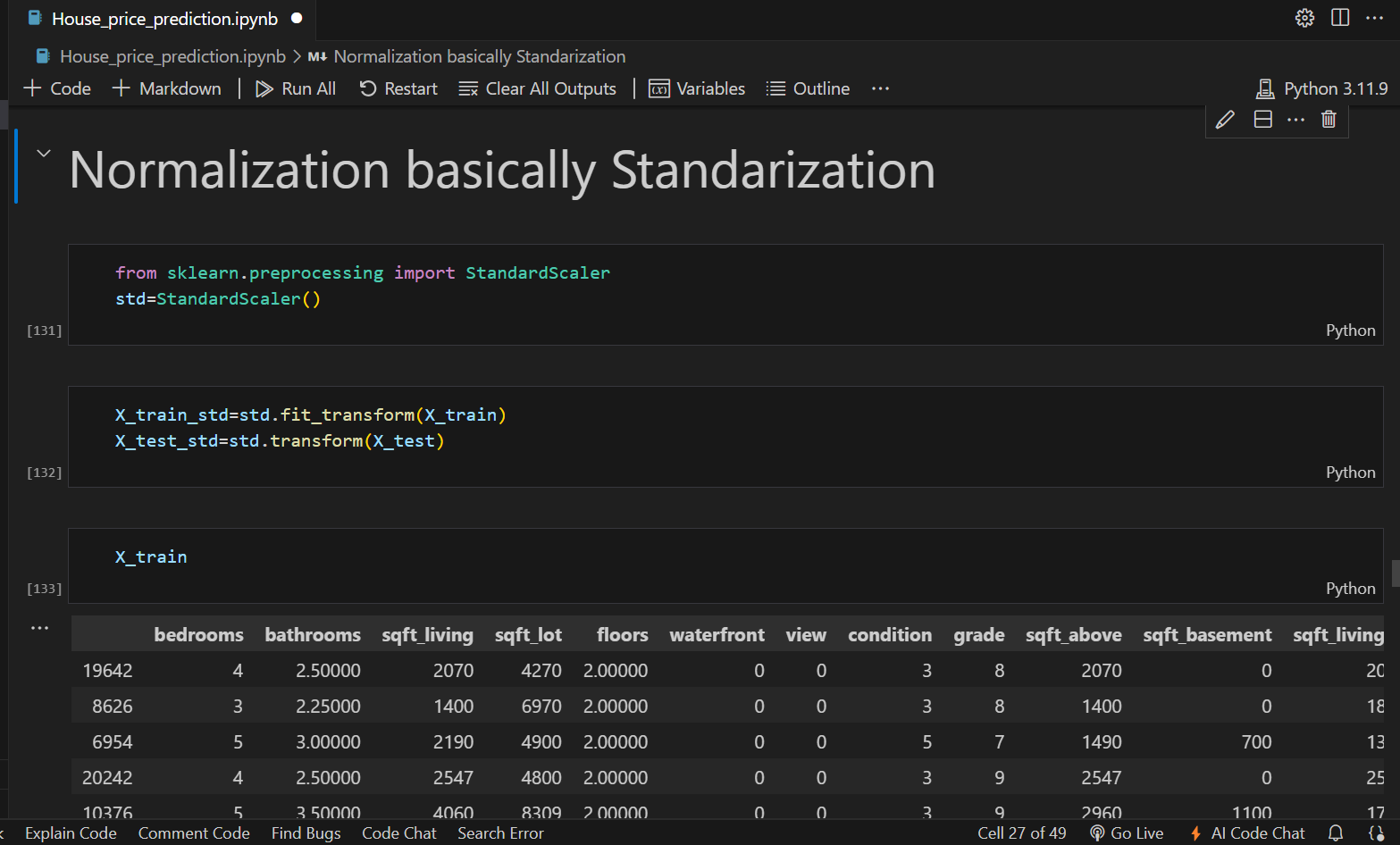


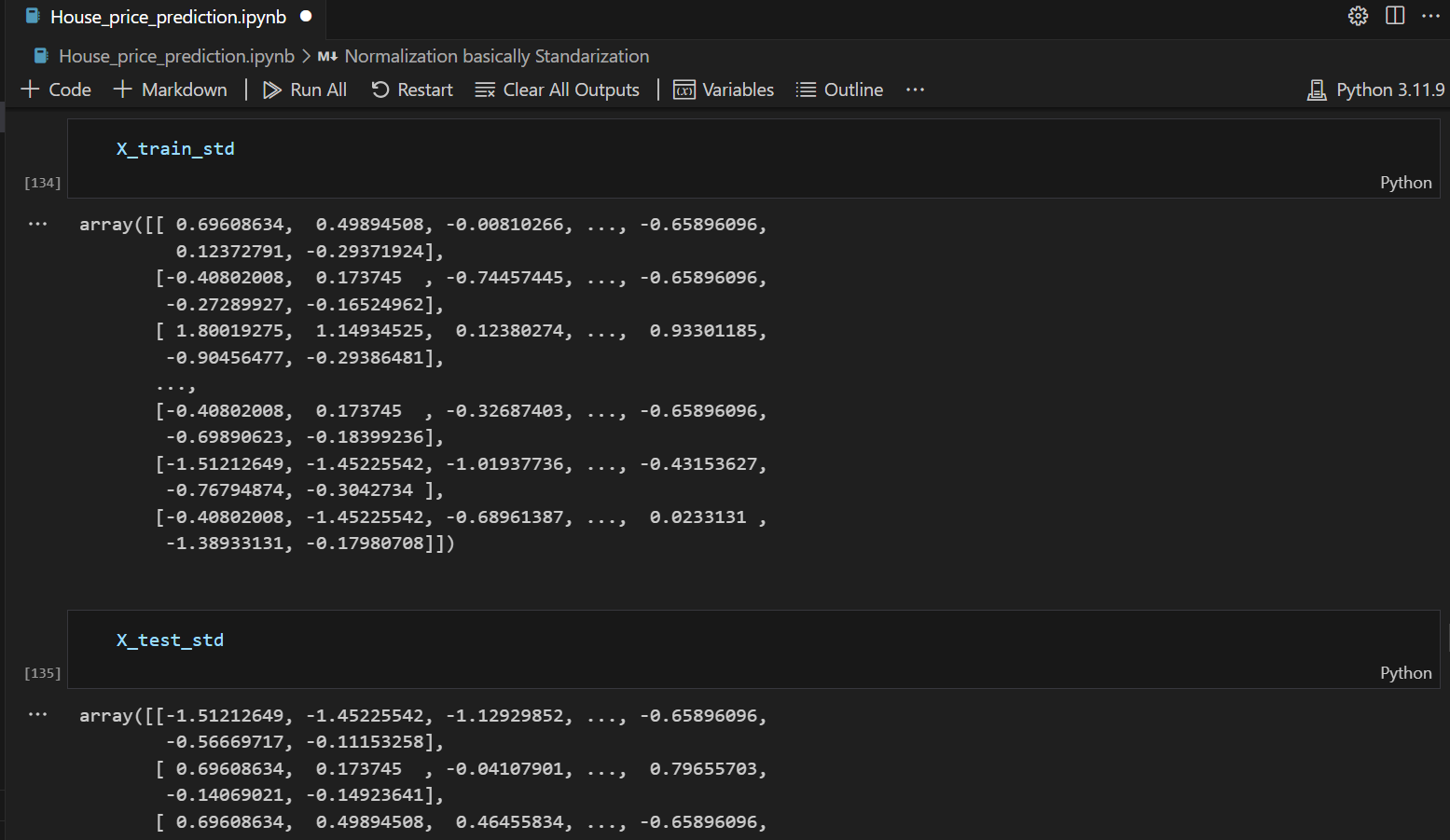


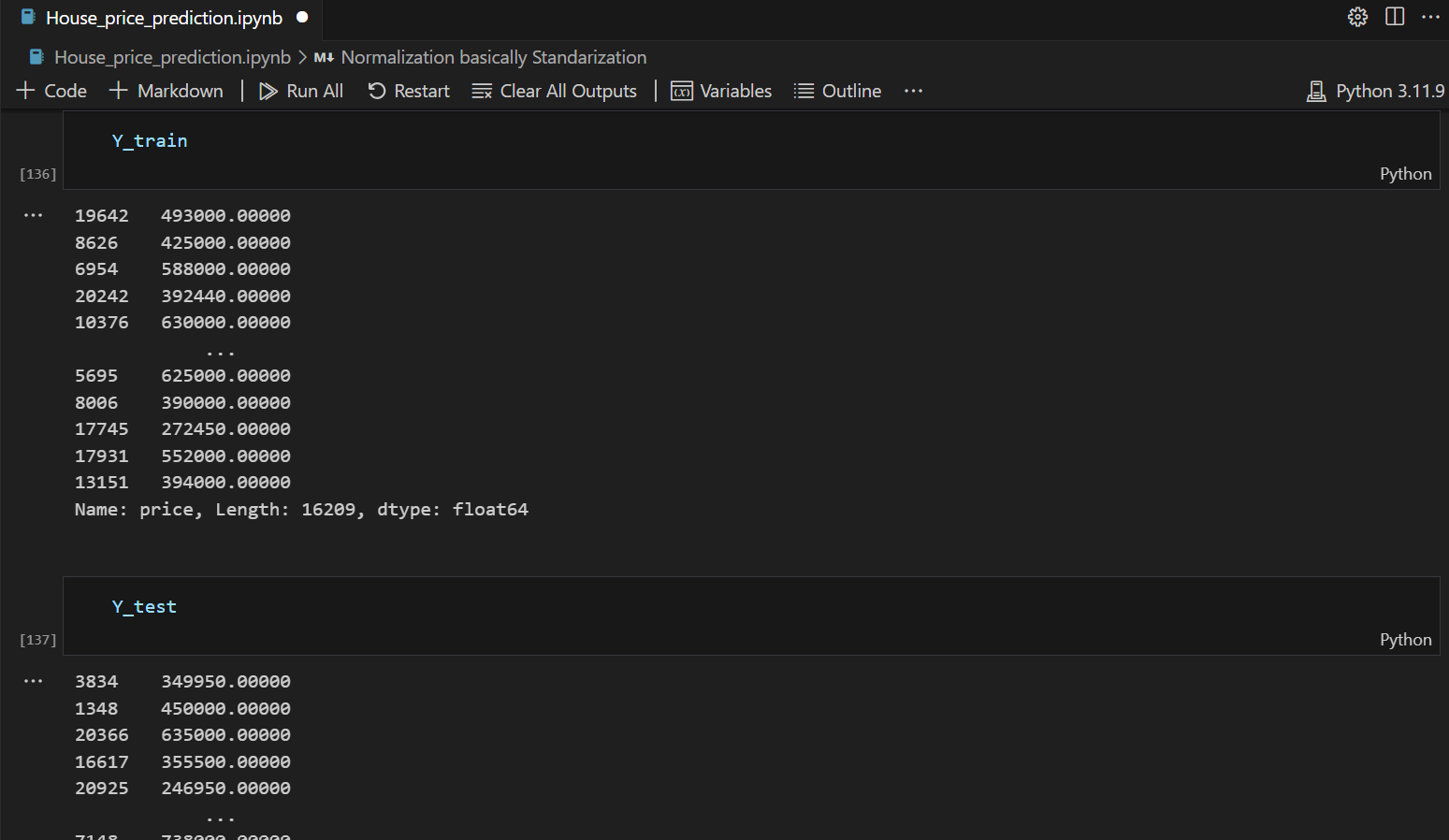


**Normalization:**

In this we normalize the data by using standar normal distribution approach.

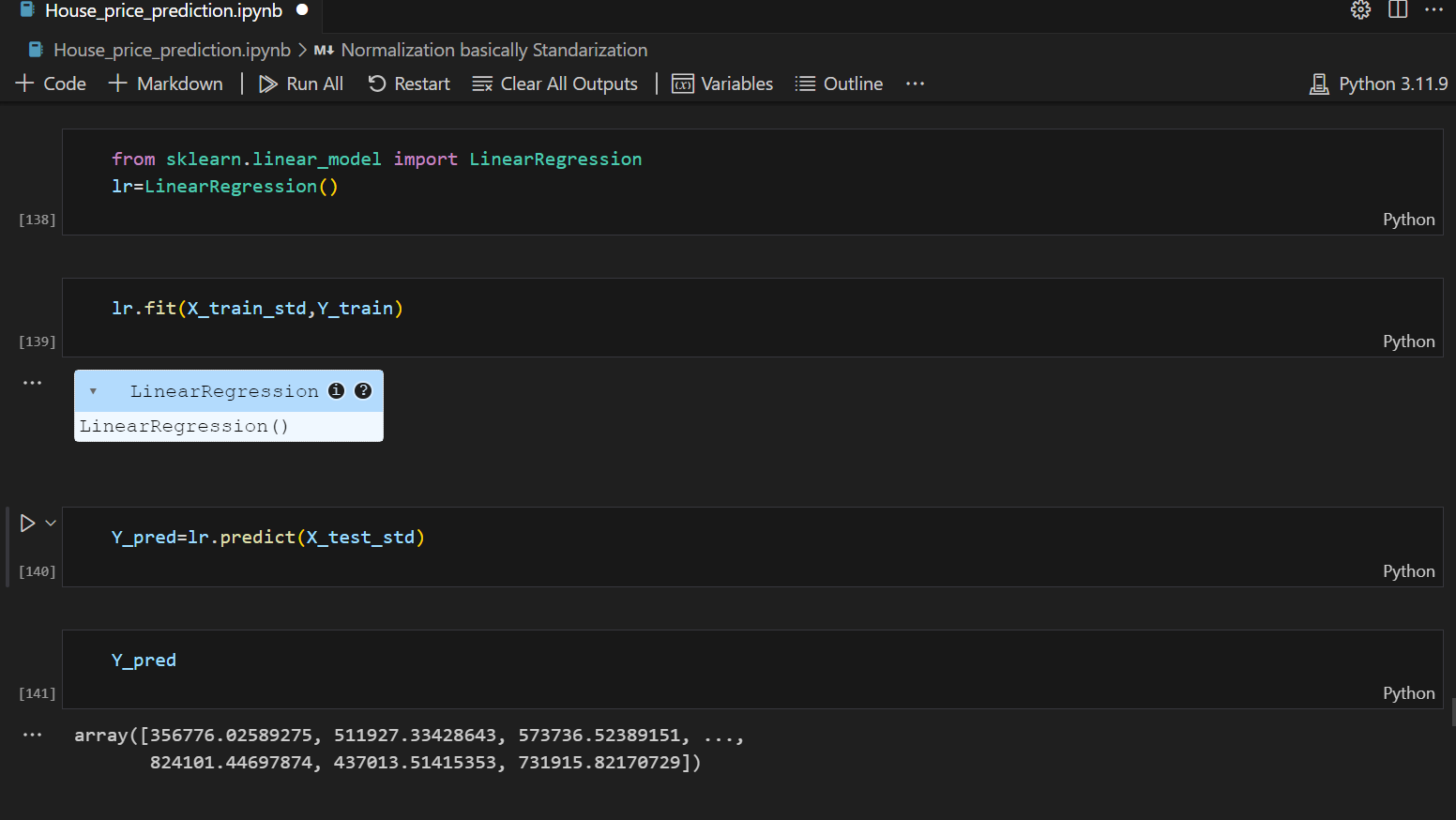
Normalization is the process in which we make that clean and tidy to easily or correctly understand.

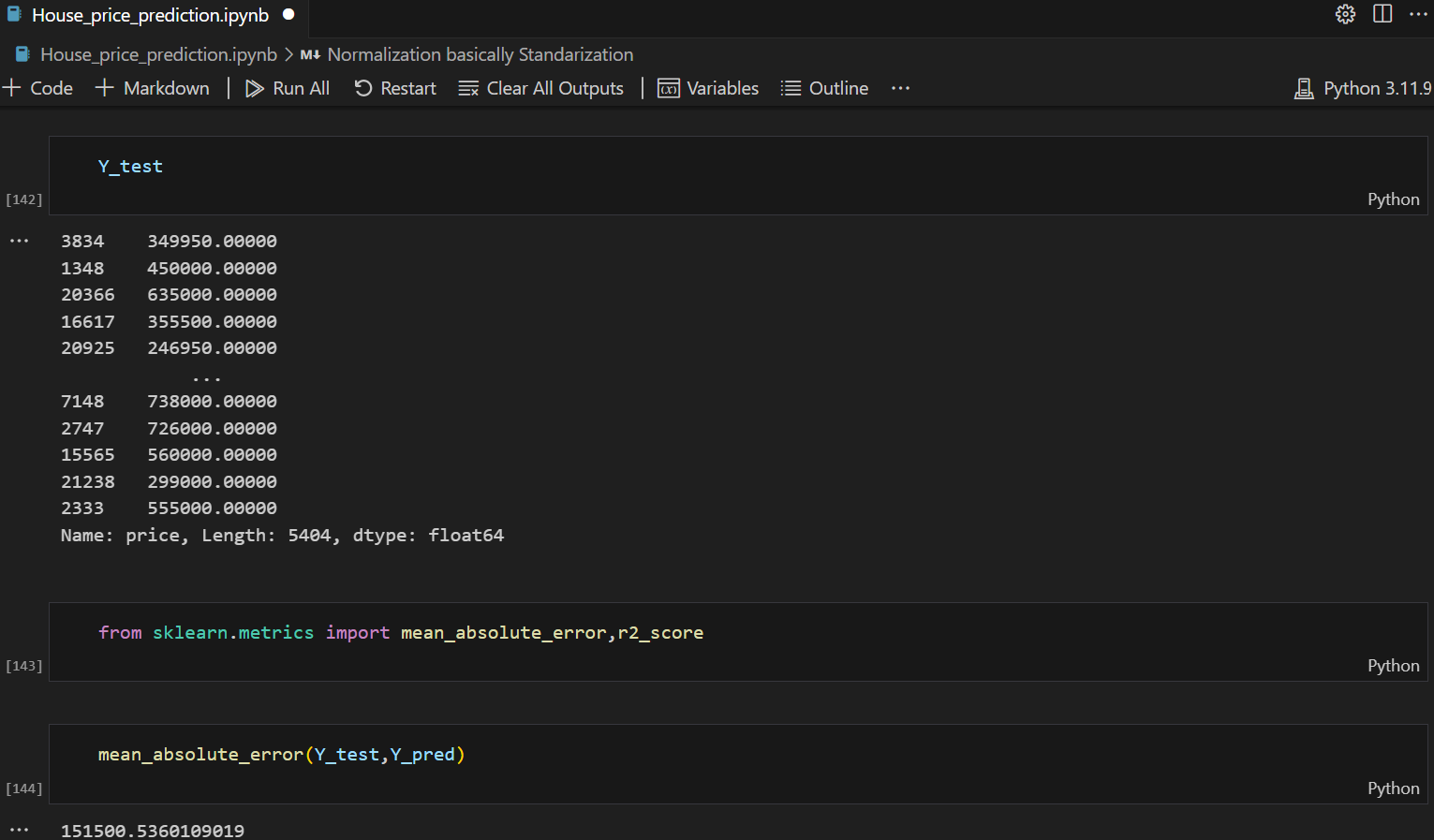


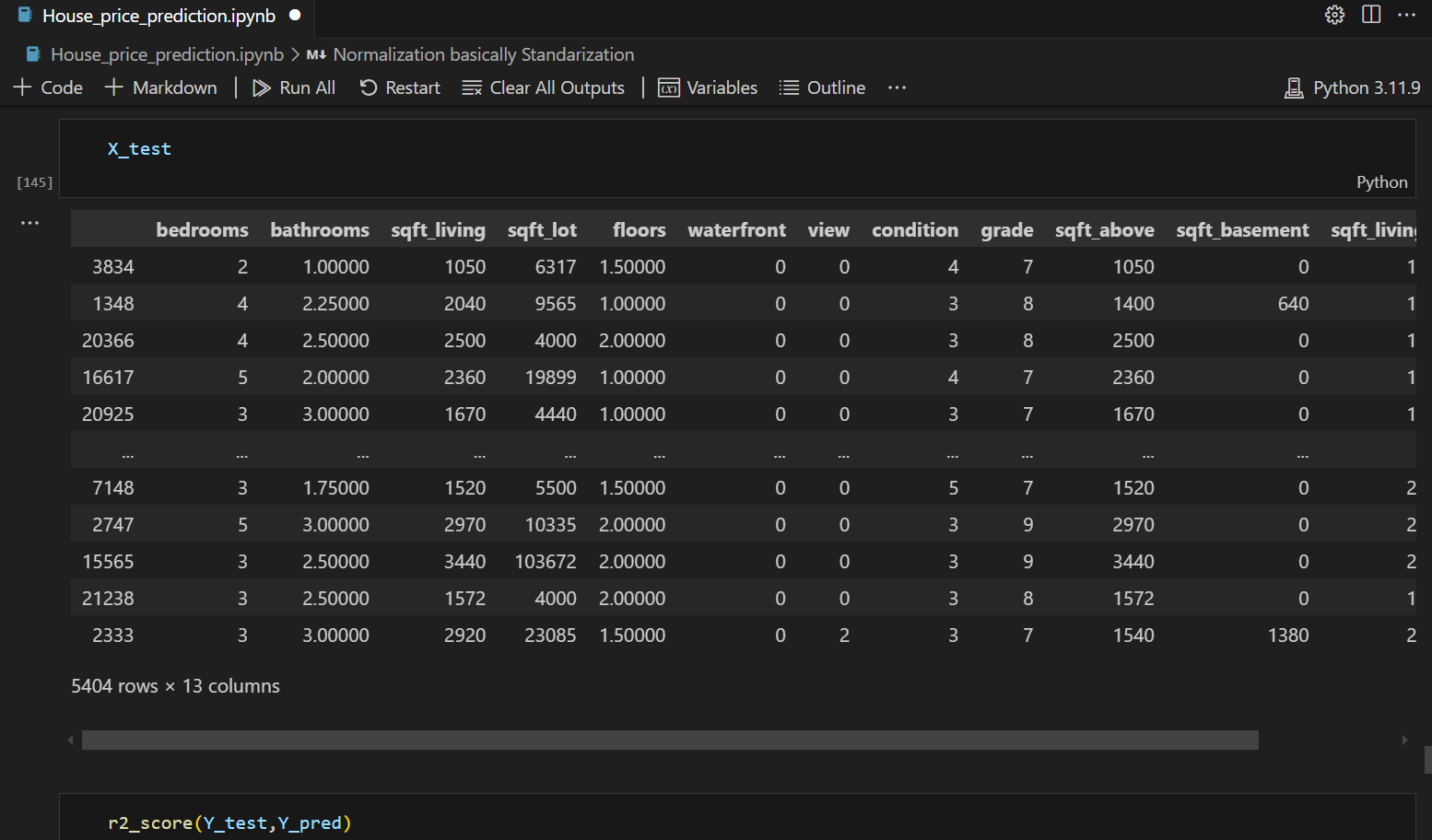


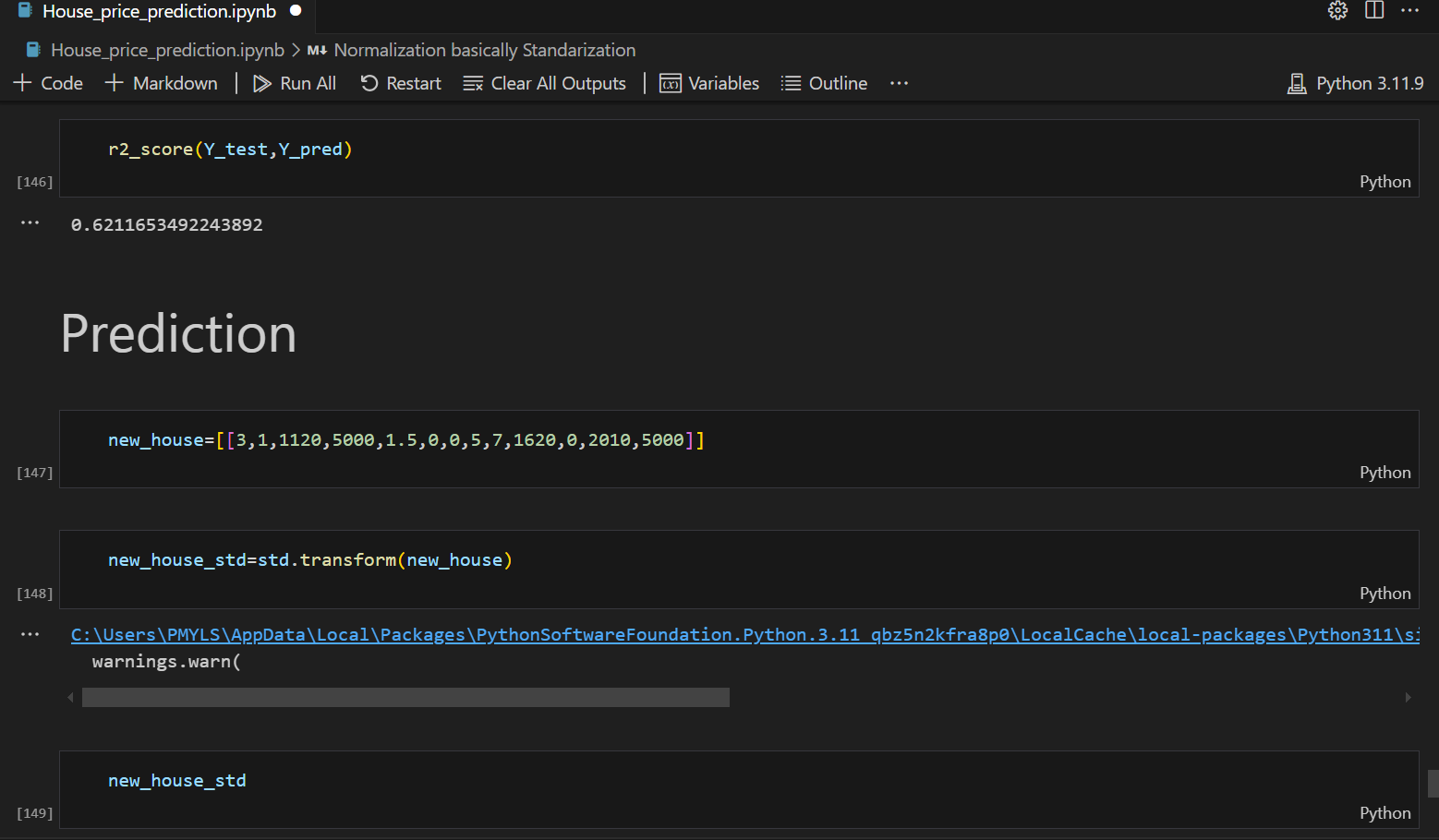
**Model:**

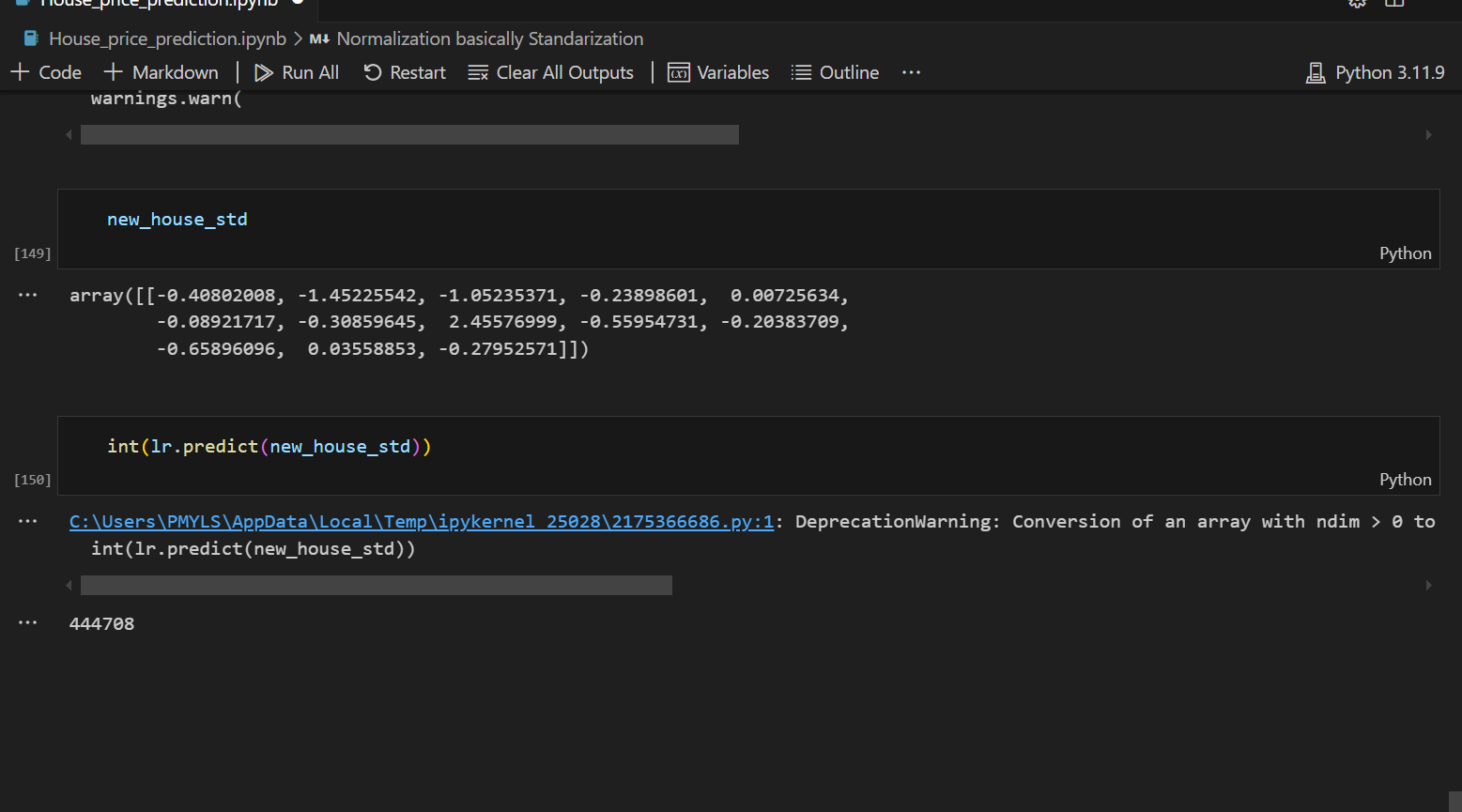
Model basically use to predict our outcomes based on our label data.In this we are going to create our model base on our data. Let’s start:











**Conclusion:**

In this project we create a model based on data that we get from Kaggle link is: <https://www.kaggle.com/datasets/shivachandel/kc-house-data> I get the secondary that of house prices based on different variables like bedrooms, bathrooms etc, I perform some analysis on it normalize data after I train the model base on the train data that I have in the code.ss