

<b>Sr. No</b>	<b><u>Content</u></b>	<b>Page No.</b>
<b>1.</b>	Introduction	<b>2</b>
<b>2.</b>	Abstract	<b>2</b>
<b>3.</b>	Data-set	<b>3</b>
<b>4.</b>	Linear Regression	<b>4</b>
<b>5.</b>	Libraries used and modules	<b>5</b>
<b>6.</b>	User Interface	<b>6-8</b>
<b>7.</b>	References	<b>9</b>

# INTRODUCTION

Suppose that a new real-estate company is established. The company doesn't have any prior experience in this field and are not able to determine the appropriate price estimation for their properties. This is where our program comes into play. Using our program an appropriate fare estimation of a house can be done with about 92% accuracy. This program can work as an aid for the real estate dealers and help them in making better estimations for their properties.

## ABSTRACT

We have created a model using the concept of multiple linear regression. The data set used for training our model will be a data frame with 21 features of 21,613 houses in King Country, Washington. The data set was recorded between year 2014-2015. There will be two main features present in our program:

- I) Test Algorithm and analyse data: To analyse the data set and predict the accuracy using different algorithms. We will be using the graph plotting libraries to visualize the effect a particular feature is having at the house price individually. This type of analysis is very effective to understand the real life scenario. For ex, a house with 3 or more bedrooms sells at a higher price and same goes for other features as well which will help the builders to build an ideal house for targeted customers.
- II) Predictor: This is the most essential feature of the project and also our important contribution towards predicting house price using linear regression. Under this the client will be entering the details of house under the fields of our

interactive GUI. Once entered the predicted price (using regression) of the house will be displayed after pressing the done option.

## DATA-SET

The data set used by us is a .csv file of 21,613 houses of King Country, Washington and having 21 features each. The details of a particular house present are:

- House Id
- date of record
- house price
- total sqft\_lot
- total sqft\_living
- actual sqft\_lot
- actual sqft\_living
- bedrooms
- bathrooms
- floors,
- waterfronts
- views
- condition (the condition of the house on a scale of 1-5)
- grade (the rating of the quality of material used for construction of house on the scale of 1-15)
- sqft\_above
- sqft\_base
- yr\_built
- yr\_renovated
- zipcode
- latitude
- longitude

# LINEAR REGRESSION

In statistics, **linear regression** is a **linear** approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple **linear regression** else its **multiple regression**. It used in finance, investing, and other disciplines.

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_tX_t$$

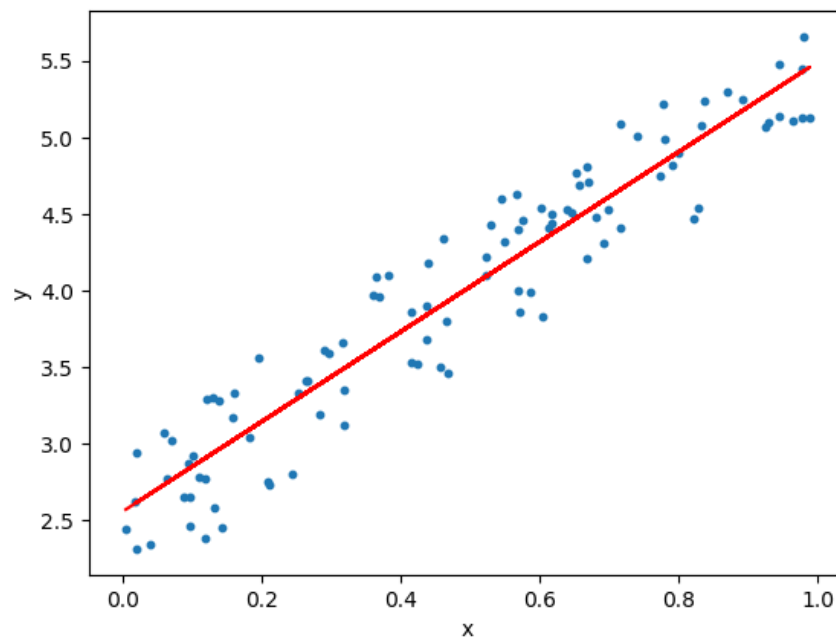
Dependent Variable (Y)

Independent Variables

(X1,X2,X3,...)

$$a = Y - (b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_tX_t)$$

$$b_1 = \frac{(\sum x_2^2)(\sum x_1y) - (\sum x_1x_2)(\sum x_2y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2}$$



## Libraries Used:

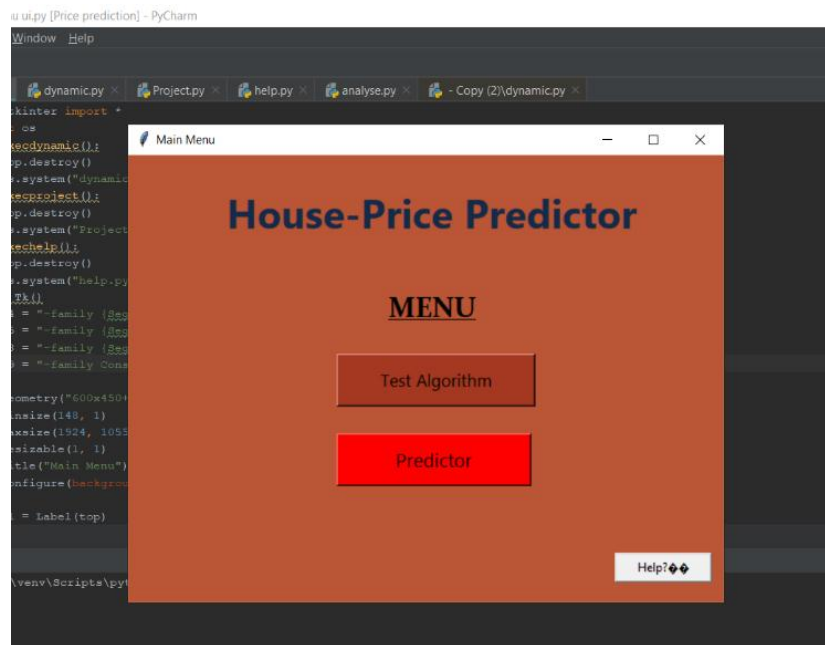
- Pandas : Importing the data sets
- Matplotlib: plotting the data sets in scatter plots
- Sci-kit learn: used for creating a training model, splitting the data set and predicting the price as well as accuracy of algorithm
- Tkinter: used for creating GUI
- OS: used for executing a different python script from current

## Modules:

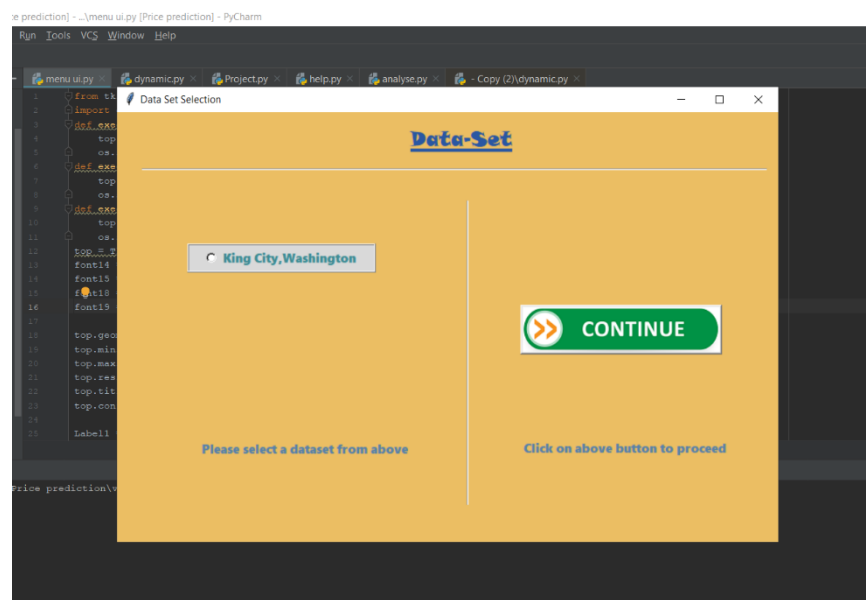
- I) Main-menu: this module contains buttons for main-menu.
- II) Data-set/algorithm testing: this module is used for analysing dataset with scatter plots and predicting accuracy of algorithm chose.
- III) Predictor: this module is a combination of fields for making the predicting feature dynamic and predict the price according to user inputs.

# User Interface

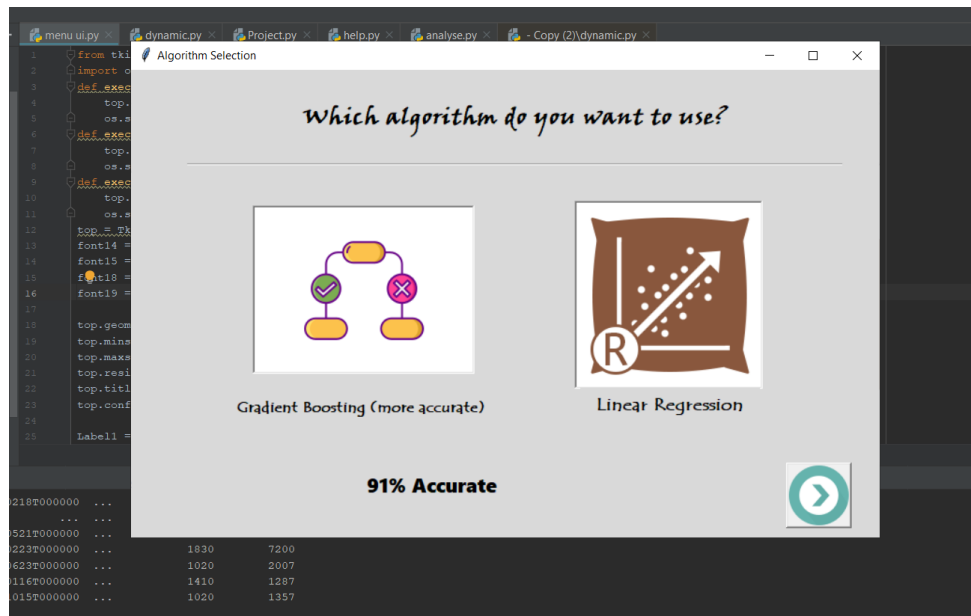
- 1) **Main-Menu:** Main-menu is a simple interface which provides 3 buttons to work on our dataset and algorithm, to use our predictor tool and to open the help section.



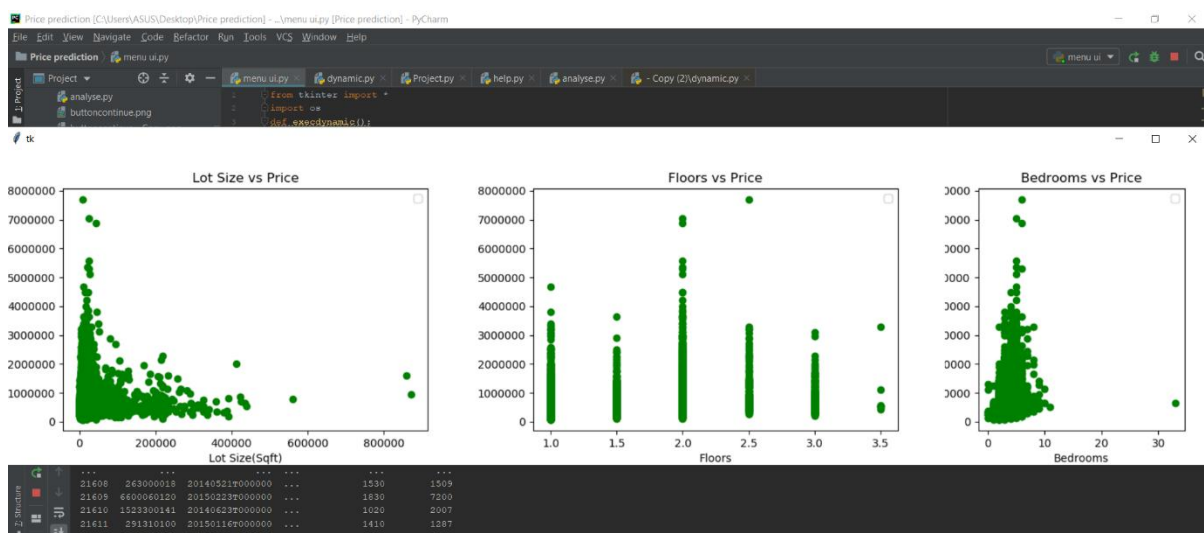
- 2) **Data-set selection:** In this particular interface, user will be provided an option to select the dataset. After selecting the user has to click on continue.



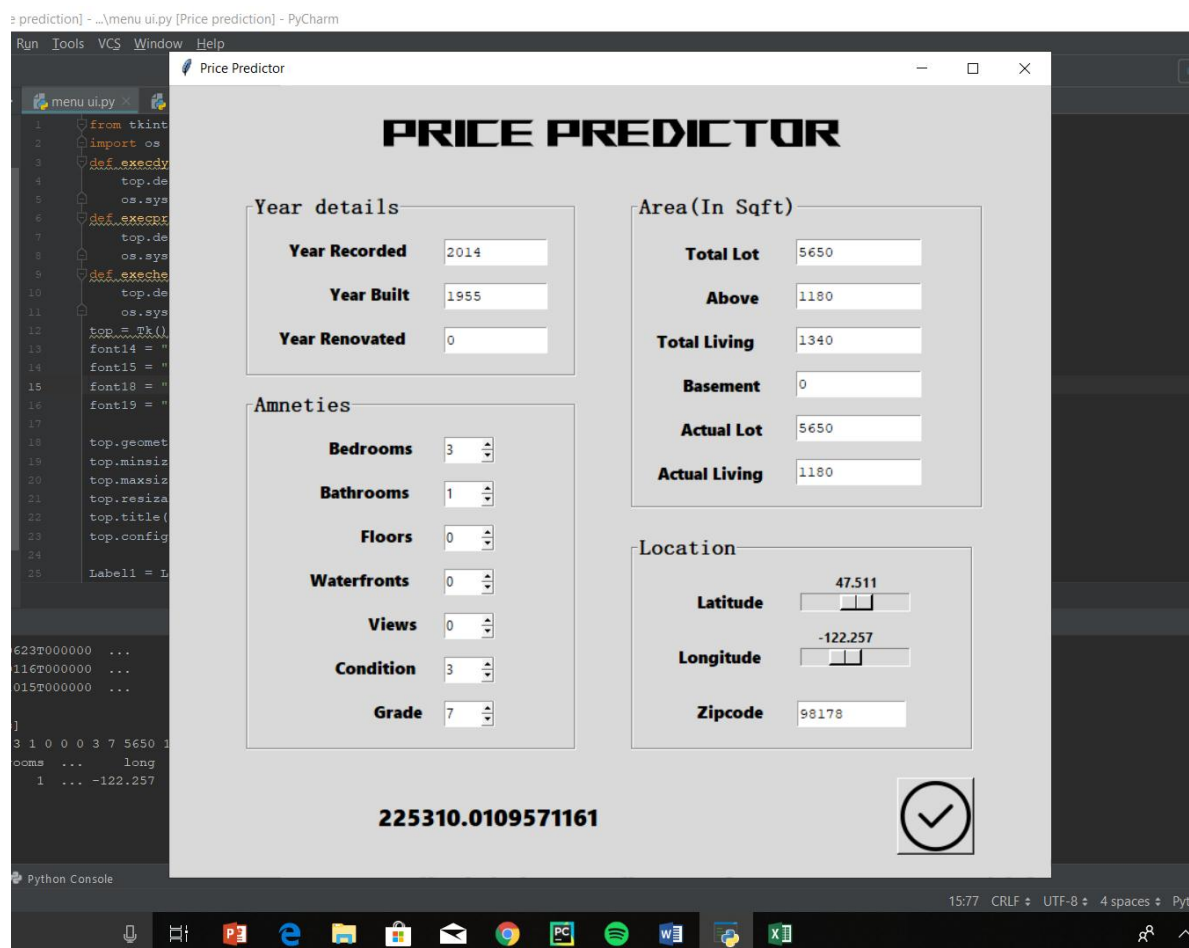
- 3) **Algorithm Testing:** In this interface user is given the option to select an algorithm and get its respective accuracy. The accuracy will be tested using the data set selected in the previous interface. Once done, the user has to click next to proceed to the analysing part.



- 4) **Analysing data:** In this module the dataset selected is analysed using scatter plot graphs. Currently the individual relationship of price with 3 independent variables are shown.



- 5) **Predictor:** This interface is the most essential functional requirement of our whole project and also our peak contribution towards predicting price using regression. The user will be provided with fields and scales to enter the data of any house the user wishes to predict the price of. Once all data is entered and on clicking done, the predicted price of the house will be displayed on the same interface.



- 6) **Help:** The last interface of our project is help section which is a simple help support page with information displayed on how to use our program.



## REFERENCES

- <https://towardsdatascience.com/simple-and-multiple-linear-regression-in-python-c928425168f9>
- <https://www.kdnuggets.com/2019/04/data-visualization-python-matplotlib-seaborn.html>
- <https://www.superdatascience.com/pages/machine-learning>
- [https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html?source=post\\_page-----d34fe8fad88f-  
-----](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html?source=post_page-----d34fe8fad88f-)
- <https://www.kaggle.com/shivachandel/kc-house-data>
- <https://towardsdatascience.com/boosting-the-accuracy-of-your-machine-learning-models-f878d6a2d185>
- <https://realpython.com/python-matplotlib-guide/>