

IBM PHASE 2 PROJECT

PREDICTING HOUSE PRICE USING MACHINE LEARNING

Group members

- ▶ Dharani G
- ▶ Gopika shree G
- ▶ Ilakiya V
- ▶ Karthika M
- ▶ Keerthigadevi P
- ▶ Kowsalya R

Project outline

- ▶ Flowchart
- ▶ System design & Architecture
- ▶ Implementation
- ▶ Execution and output

Flow Chart

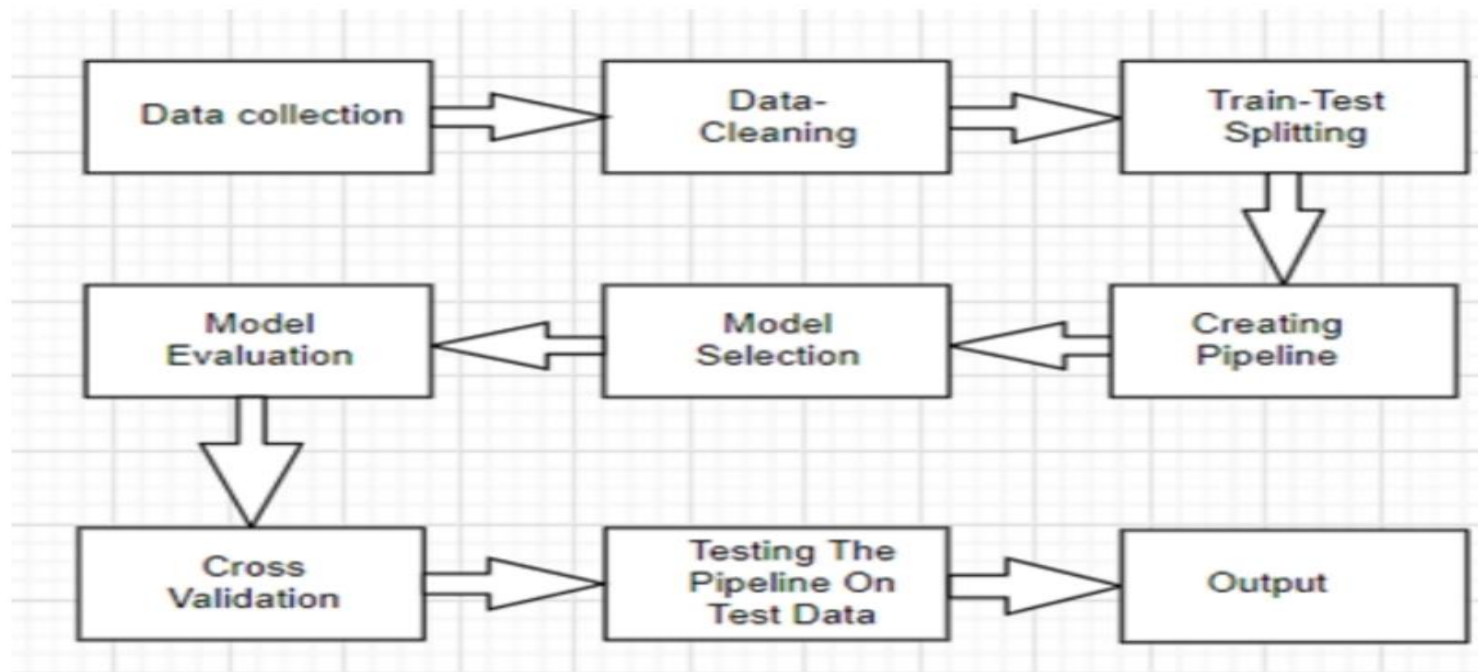


Fig 1. Research Flow Diagram

System design & Architecture

► Collection of data

We are collected data for Real estate from different online Real estate web side and repository.

In such data have features like 'ZN', 'INDUS', 'RAD', 'CHAS', 'LSTAT', 'CRIM', 'AGE', 'NOV' etc... ..

Dataset validity Is must otherwise there is no Point in analyzed the data .

System design & Architecture

► Data pre-processing

In this phase , our data is clean up.

There is might be missing values in our data set.

There are three ways to fill our missing values 1) get rid of our missing data Points, 2)Get rid of the whole attributes , 3)Set the values to some values(0, Mean or median)

System Design& Architecture

- ▶ Training and testing model

In this phase , data is broken down to two parts : training and testing.

80% of data used for training purpose and 20% data used for testing purpose

The trained model is used to save by 'joblib'.

Implementation

Reading the data to plot the graphs:

```
data = pd.read_csv("kc_house_data.csv")
```

Using the above data, code to show plot a relation between number of bedrooms and number of houses:

```
data['bedrooms'].value_counts().plot(kind = 'bar')  
plt.title('Number of bedrooms')  
plt.xlabel('bedrooms')  
plt.ylabel('Number of houses')  
plt.show()  
sns.despine()
```

Code to plot relation between Price and Living area:

```
plt.scatter(data.price, data.sqft_living)  
plt.title('price vs sqft living')  
plt.xlabel('price')  
plt.ylabel('Sqft area')  
plt.show()  
sns.despine()
```


Implementation

Plots relation between price and latitudes:

```
plt.scatter(data.price, data.lat)
plt.title('price vs latitude values')
plt.xlabel('price')
plt.ylabel('latitude values')
plt.show()
sns.despine()
```

Plots relation between price and area:

```
plt.scatter(data.price, (data.sqft_living + data.sqft_basement))
plt.title('price vs sqft area')
plt.xlabel('price')
plt.ylabel('area')
plt.show()
sns.despine()
```

Implementation

Plots relation between waterfront and price:

```
plt.scatter(data.waterfront, data.price)
plt.title('waterfront vs price')
plt.xlabel('waterfront')
plt.ylabel('price')
plt.show()
sns.despine()
```

Plots relation between condition and price:

```
plt.scatter(data.condition, data.price)
plt.title('condition vs price')
plt.xlabel('condition')
plt.ylabel('price')
plt.show()
sns.despine()
```

Execution and output

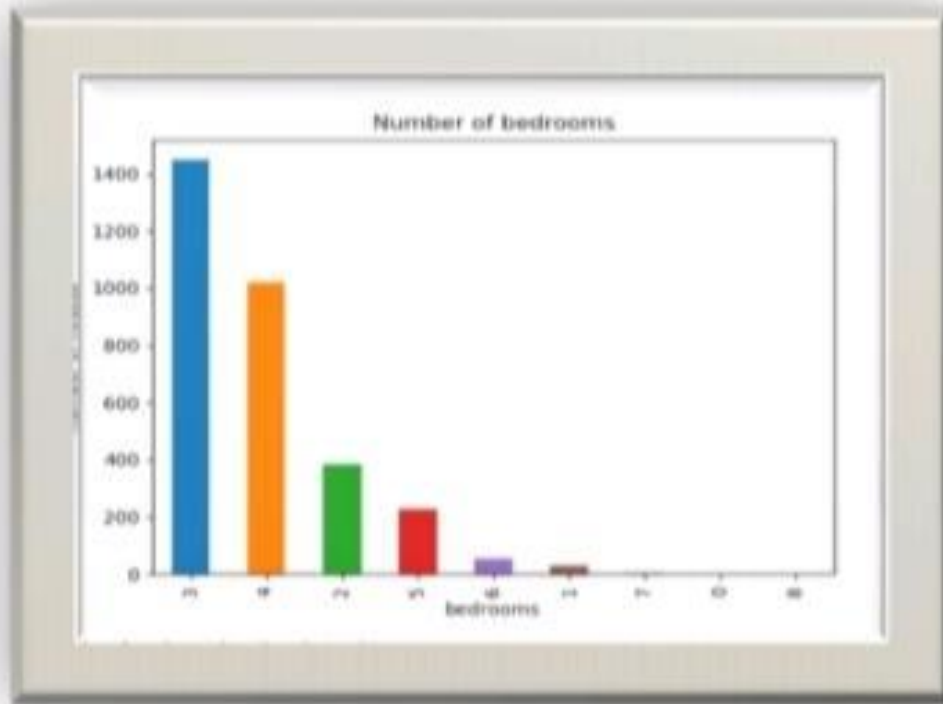
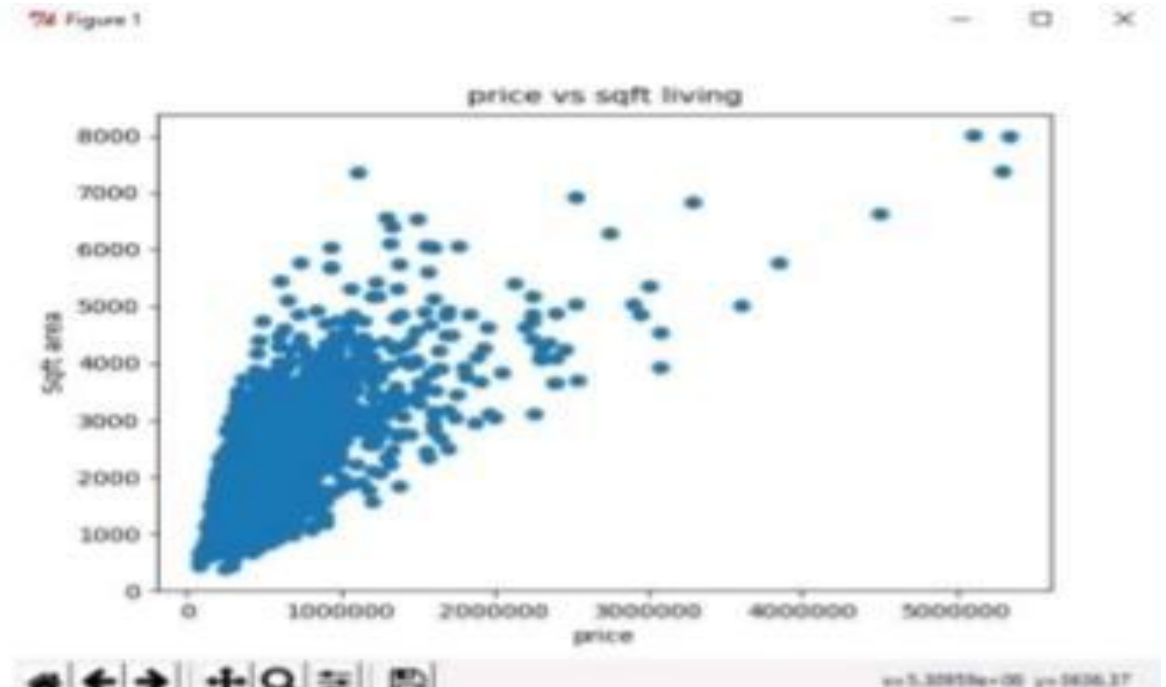


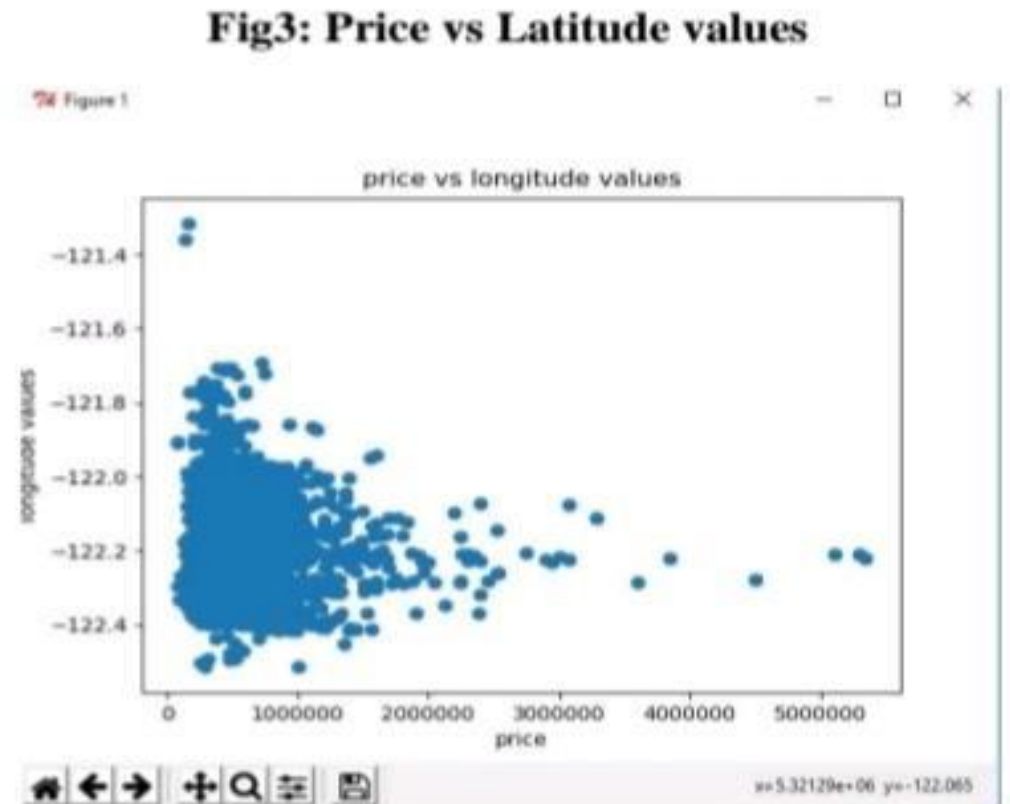
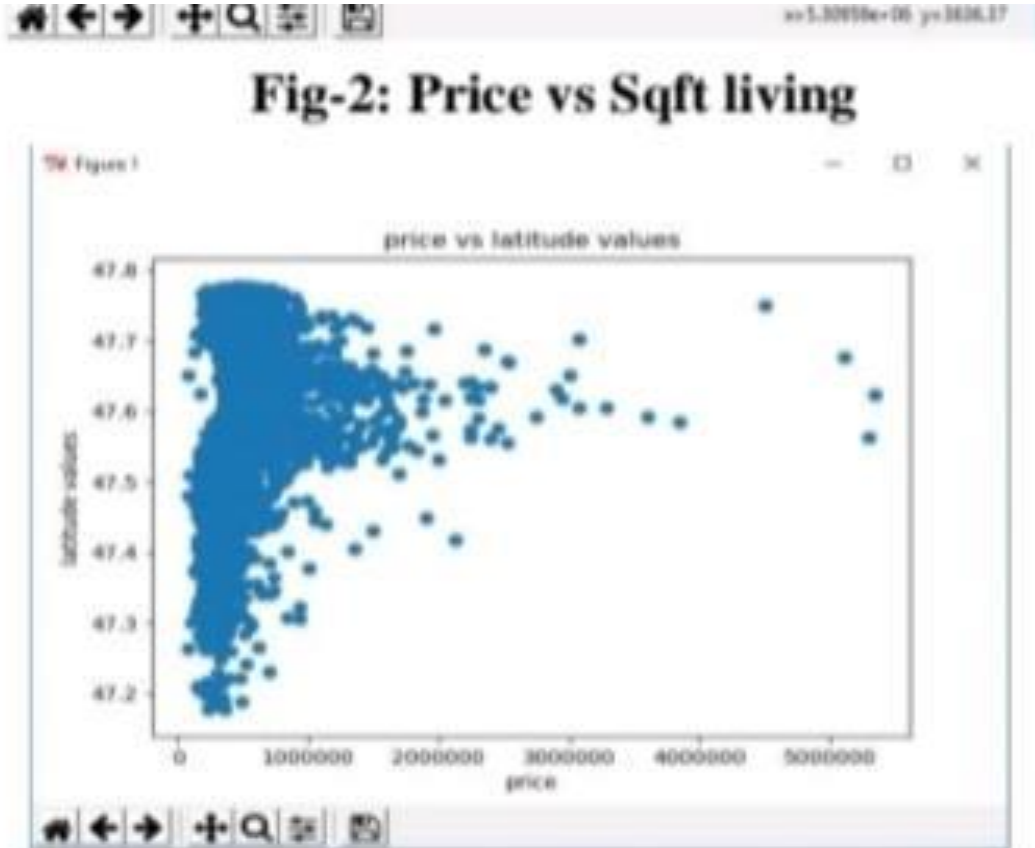
Fig – 1 : No. of houses vs no. of bedrooms

Figure 1

Figure 1



Execution and output



THANK YOU

