Aligarh College of Engineering and Technology

(APPROVED BY A.I.C.T.E & AFFILIATED TO AKTU, LUCKNOW)

Session: 2020 - 2024

A Project Report on

House Price Prediction

Exploring the House Prices in Bengaluru

Team Members :

Anurag Gupta	**(2001090100004)**
Arpit Kumar Nauhwara	**(2001090100006)**
Hemant Sharma	**(2001090100021)**

Submitted To

Dr. Anand Sharma
HOD of CSE

Mr. Rohit Yadav
Project Incharge/Guide

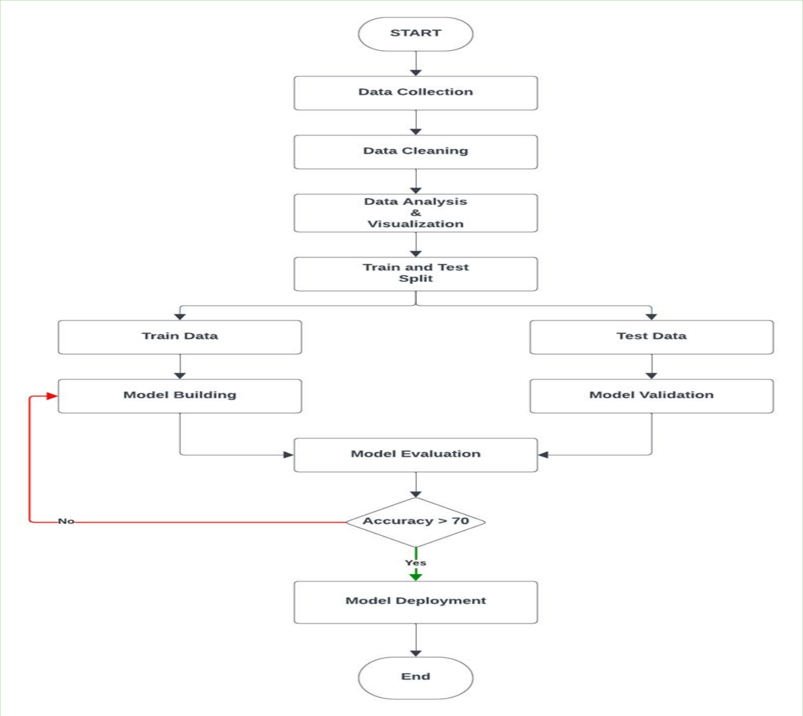
PROBLEM STATEMENT,

- It is very difficult to search house at places like Bengaluru. Even if you find a house it is very difficult to get a perfect price for the same.
- To overcome such problem Machine Learning technique can be used.
- Using Data Analysis and Machine Learning Technique it will be easy to know the price of house based on the area available, number of bedrooms, facilities available.

OBJECTIVE,

The objective of the House Price Prediction Project is to develop an accurate machine learning model forecasting residential property prices in Bangalore. It aims to provide precise predictions, offer insights into market dynamics, ensure transparency, generalize well to unseen data, and empower users with a user-friendly interface. By achieving these goals, the project seeks to contribute to a transparent and informed real estate market, enabling stakeholders to make confident decisions in property transactions.

WORKFLOW,

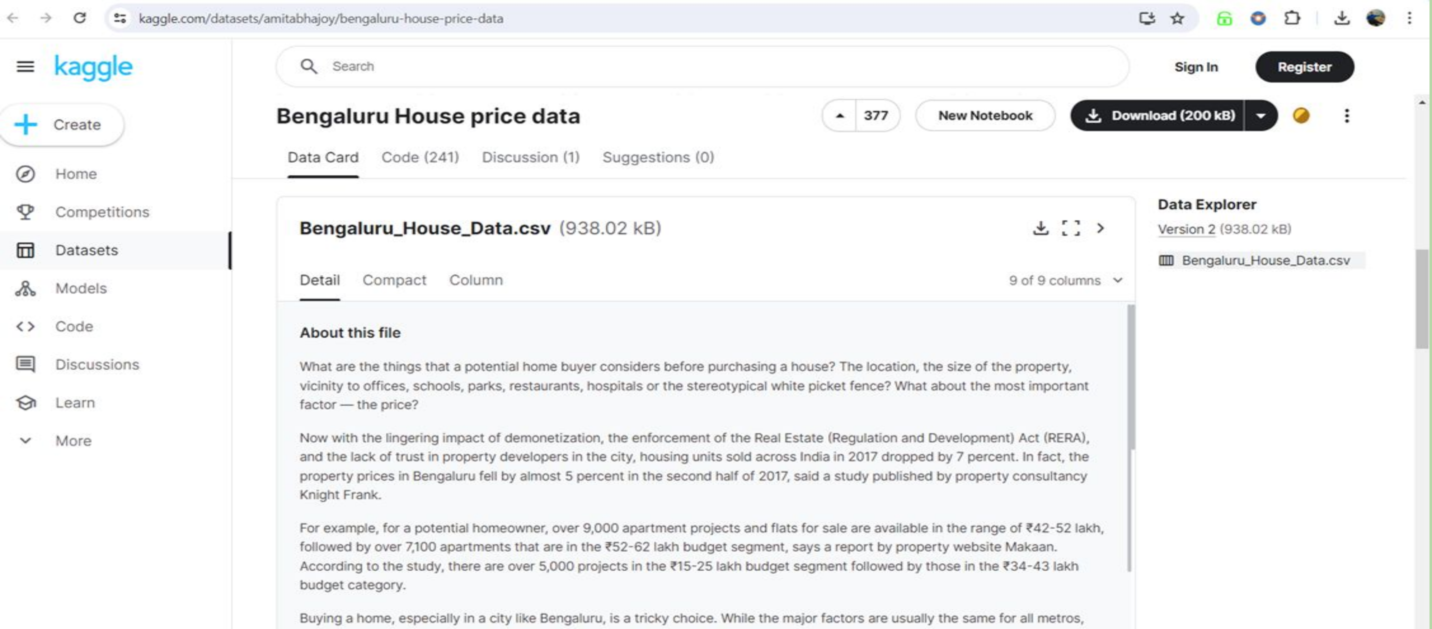

START
Data Collection
Data Cleaning
Data Analysis
&
Visualization
Train and Test
Split
Train Data
Test Data
Model Building
Model Validation
Model Evaluation
Accuracy > 70
No
Yes
Model Deployment
End

TOOLS AND TECHNOLOGIES,

1. Python.
2. Numpy and Pandas for data cleaning.
3. Matplotlib and Seaborn for data visualization.
4. Sklearn for model building.
5. Google Collab, Jupyter notebook, and visual studio code as IDE.
6. Python, flask for http server.
7. HTML/CSS for UI.

DATA COLLECTION,

.



Bengaluru House price data

Bengaluru_House_Data.csv (938.02 kB)

About this file

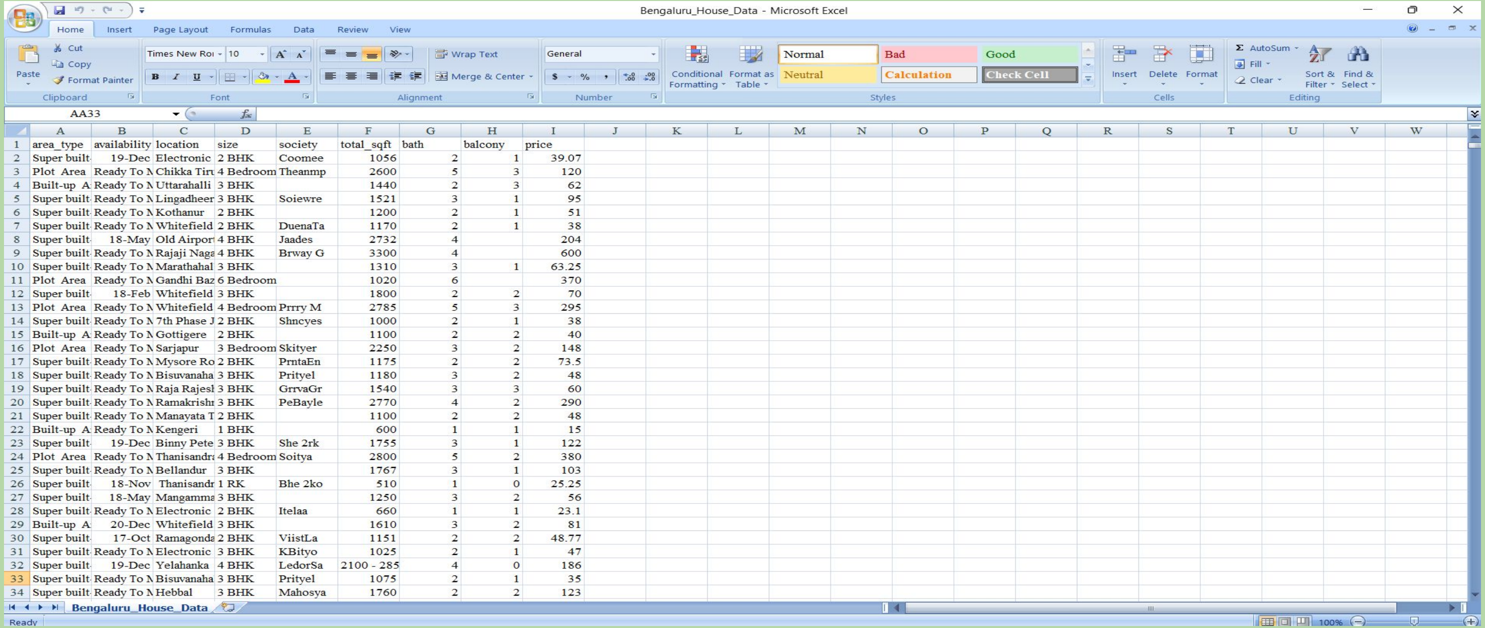
What are the things that a potential home buyer considers before purchasing a house? The location, the size of the property, vicinity to offices, schools, parks, restaurants, hospitals or the stereotypical white picket fence? What about the most important factor — the price?

Now with the lingering impact of demonetization, the enforcement of the Real Estate (Regulation and Development) Act (RERA), and the lack of trust in property developers in the city, housing units sold across India in 2017 dropped by 7 percent. In fact, the property prices in Bengaluru fell by almost 5 percent in the second half of 2017, said a study published by property consultancy Knight Frank.

For example, for a potential homeowner, over 9,000 apartment projects and flats for sale are available in the range of ₹42-52 lakh, followed by over 7,100 apartments that are in the ₹52-62 lakh budget segment, says a report by property website Makaan. According to the study, there are over 5,000 projects in the ₹15-25 lakh budget segment followed by those in the ₹34-43 lakh budget category.

Buying a home, especially in a city like Bengaluru, is a tricky choice. While the major factors are usually the same for all metros,

DATASET OVERVIEW,

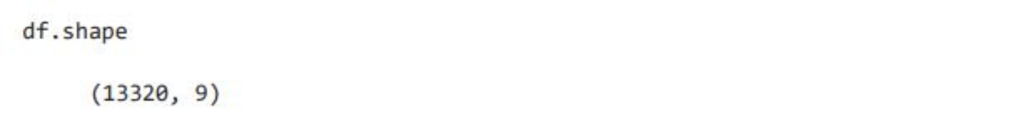


	A	B	C	D	E	F	G	H	I
1	area_type	availability	location	size	society	total_sqft	bath	balcony	price
2	Super built-	19-Dec	Electronic	2 BHK	Coomee	1056	2	1	39.07
3	Plot Area	Ready To M	Chikka Tiru	4 Bedroom	Theanmp	2600	5	3	120
4	Built-up A	Ready To M	Uttarahalli	3 BHK		1440	2	3	62
5	Super built-	Ready To M	Lingadheer	3 BHK	Soiewre	1521	3	1	95
6	Super built-	Ready To M	Kothanur	2 BHK		1200	2	1	51
7	Super built-	Ready To M	Whitefield	2 BHK	DuenaTa	1170	2	1	38
8	Super built-	18-May	Old Airport	4 BHK	Jaades	2732	4		204
9	Super built-	Ready To M	Rajaji Naga	4 BHK	Brway G	3300	4		600
10	Super built-	Ready To M	Marathahal	3 BHK		1310	3	1	63.25
11	Plot Area	Ready To M	Gandhi Baz	6 Bedroom		1020	6		370
12	Super built-	18-Feb	Whitefield	3 BHK		1800	2	2	70
13	Plot Area	Ready To M	Whitefield	4 Bedroom	Prrry M	2785	5	3	295
14	Super built-	Ready To M	7th Phase J	2 BHK	Shncyes	1000	2	1	38
15	Built-up A	Ready To M	Gottigere	2 BHK		1100	2	2	40
16	Plot Area	Ready To M	Sarjapur	3 Bedroom	Skityer	2250	3	2	148
17	Super built-	Ready To M	Mysore Ro	2 BHK	PrntaEn	1175	2	2	73.5
18	Super built-	Ready To M	Bisuvanaha	3 BHK	Prityel	1180	3	2	48
19	Super built-	Ready To M	Raja Rajesh	3 BHK	GrrvaGr	1540	3	3	60
20	Super built-	Ready To M	Ramakrishn	3 BHK	PeBayle	2770	4	2	290
21	Super built-	Ready To M	Manayata T	2 BHK		1100	2	2	48
22	Built-up A	Ready To M	Kengeri	1 BHK		600	1	1	15
23	Super built-	19-Dec	Binny Pete	3 BHK	She 2rk	1755	3	1	122
24	Plot Area	Ready To M	Thanisandra	4 Bedroom	Soitya	2800	5	2	380
25	Super built-	Ready To M	Bellandur	3 BHK		1767	3	1	103
26	Super built-	18-Nov	Thanisandr	1 RK	Bhe 2ko	510	1	0	25.25
27	Super built-	18-May	Mangamma	3 BHK		1250	3	2	56
28	Super built-	Ready To M	Electronic	2 BHK	Itelaa	660	1	1	23.1
29	Built-up A	20-Dec	Whitefield	3 BHK		1610	3	2	81
30	Super built-	17-Oct	Ramagonda	2 BHK	ViistLa	1151	2	2	48.77
31	Super built-	Ready To M	Electronic	3 BHK	KBityo	1025	2	1	47
32	Super built-	19-Dec	Yelahanka	4 BHK	LedorSa	2100 - 285	4	0	186
33	Super built-	Ready To M	Bisuvanaha	3 BHK	Prityel	1075	2	1	35
34	Super built-	Ready To M	Hebbal	3 BHK	Mahosya	1760	2	2	123

ABOUT DATASET,

1. **Area type :** Description of the area.
2. **Availability :** when it is ready.
3. **Location :** Where it is located in Bengaluru.
4. **Size :** BHK or Bedrooms.
5. **Society :** To which society it belongs.
6. **Total sqft :** Size of the property in sq.ft.
7. **Bath :** No. of Bathrooms.
8. **Balcony :** No. of the Balcony.
9. **Price :** Value of the property in lakhs (Indian Rupee - ₹).

Now Check the shape of Dataset,

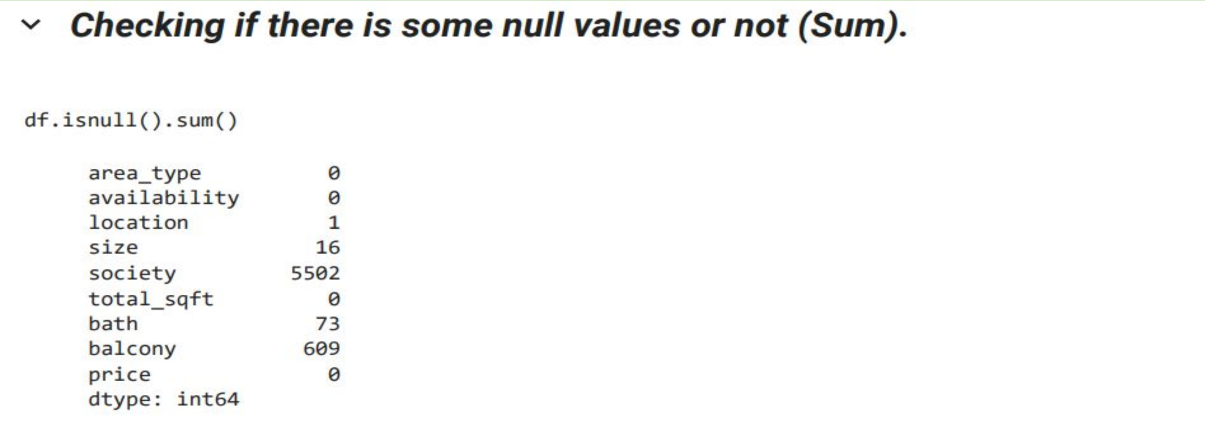


```
df.shape

     (13320, 9)
```

DATA CLEANING AND PREPROCESSING,

Data cleaning and preprocessing involves preparing raw data for analysis by removing errors, and irrelevant information.

It includes tasks like handling missing values.

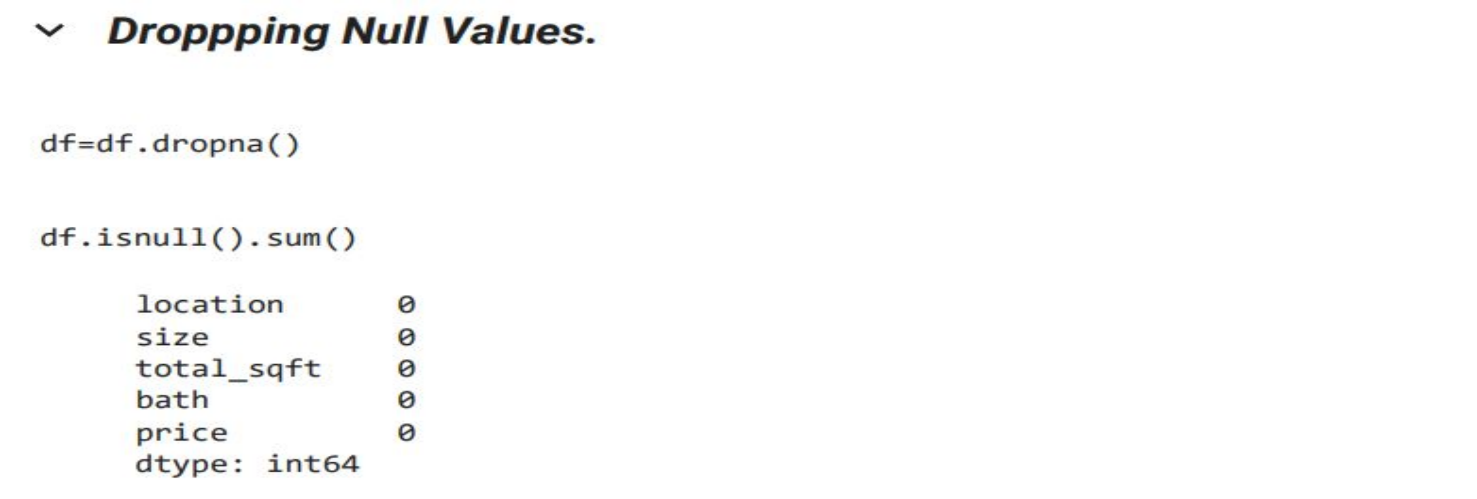


Checking if there is some null values or not (Sum).

```
df.isnull().sum()

     area_type          0
     availability       0
     location           1
     size              16
     society         5502
     total_sqft         0
     bath              73
     balcony          609
     price              0
     dtype: int64
```

DATA CLEANING AND PREPROCESSING,

After Checking Null Value, we drop Null value by using command “df.dropna()”.



Droppping Null Values.

```
df=df.dropna()
```

```
df.isnull().sum()
```

```
location      0
size          0
total_sqft    0
bath          0
price         0
dtype: int64
```

DATA VISUALIZATION,

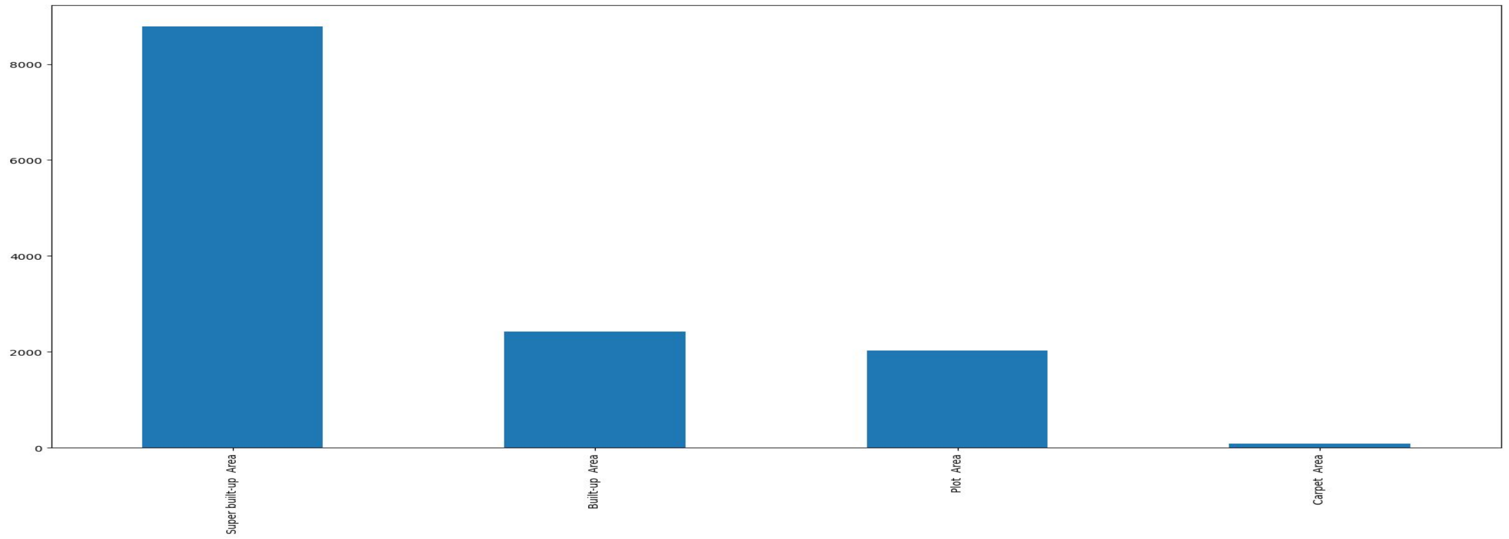
Data visualization is about presenting data in graphical or visual formats like,

1. Charts,
2. Graphs,
3. Maps.

It helps make complex data easier to understand, allowing us to spot patterns, trends, and insights more quickly and effectively.

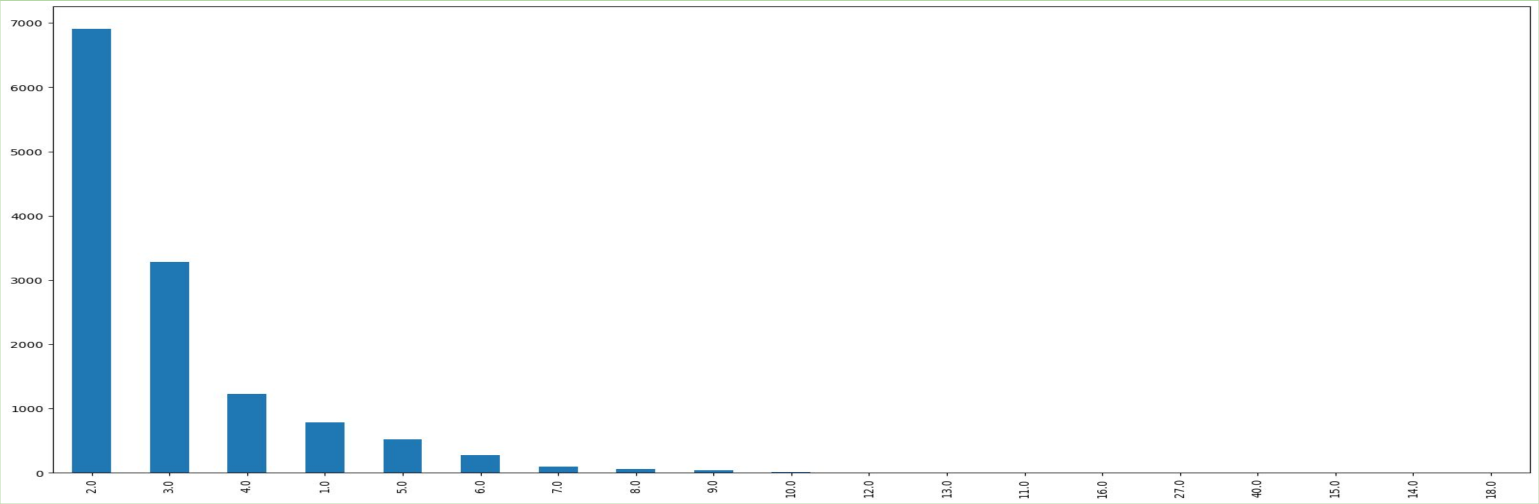
DATA VISUALIZATION,

1. BAR CHART : In the given Bar Chart, it shows that which area type have how much Property.



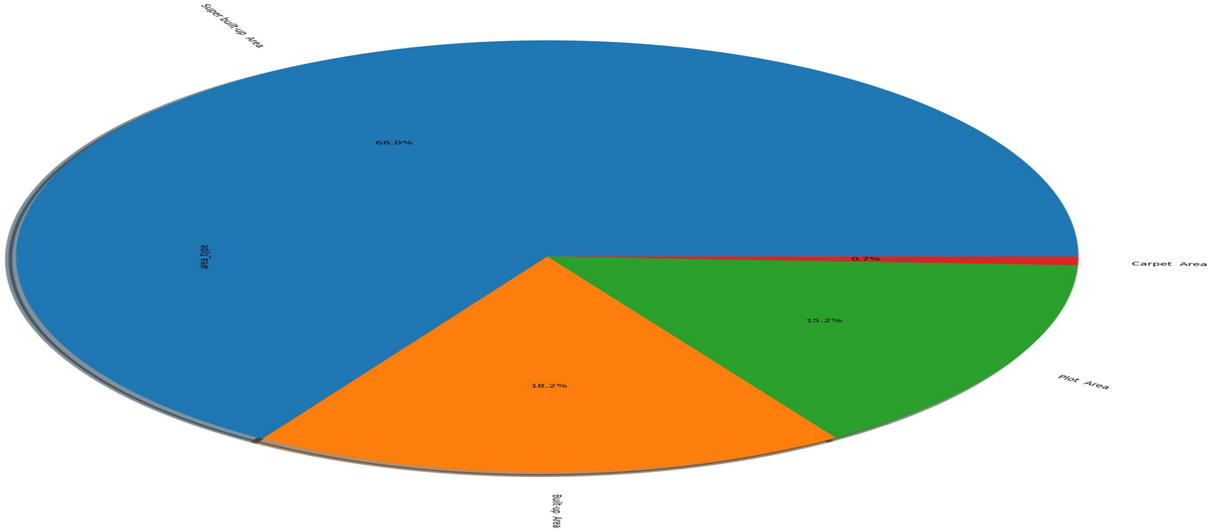
DATA VISUALIZATION,

2. BAR CHART : In the given Bar Chart, it shows that how much property have how much bathrooms.



DATA VISUALIZATION,

3. <u>PIE CHART</u> : In the given Pie Chart, it shows that which area type have how much Property in Percentage (%).

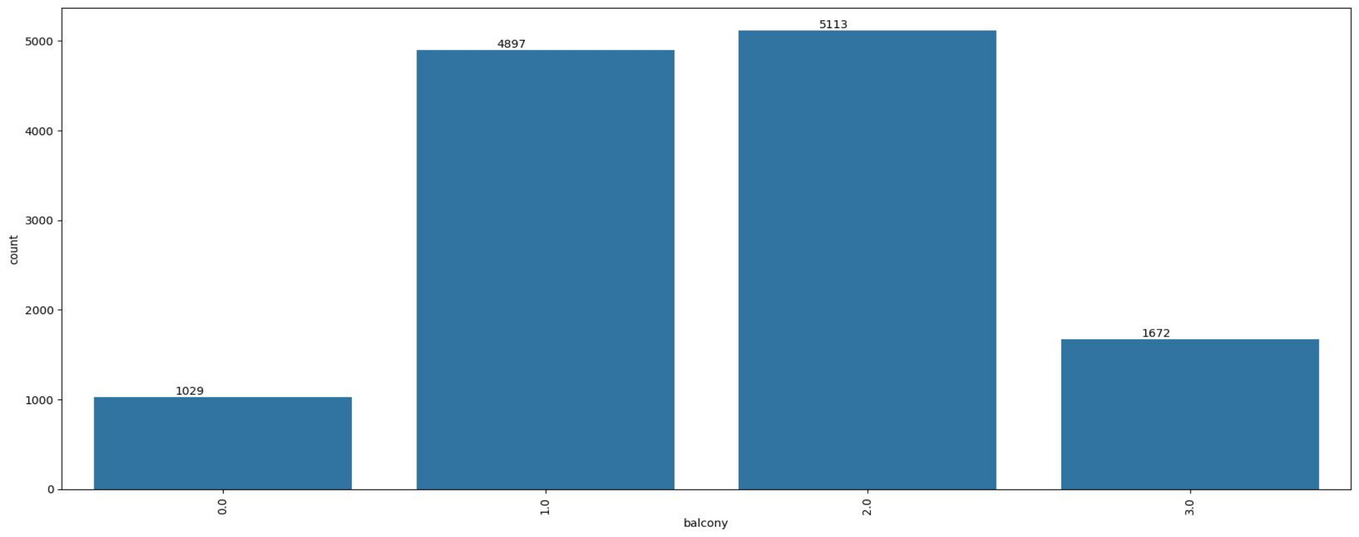


<u>Area Type in %.</u>

- *Super built-up area. (66.6%)*
- *Built-up Area. (18.2%)*
- *Plot Area. (15.2%)*
- *Carpet Area. (0.7%)*

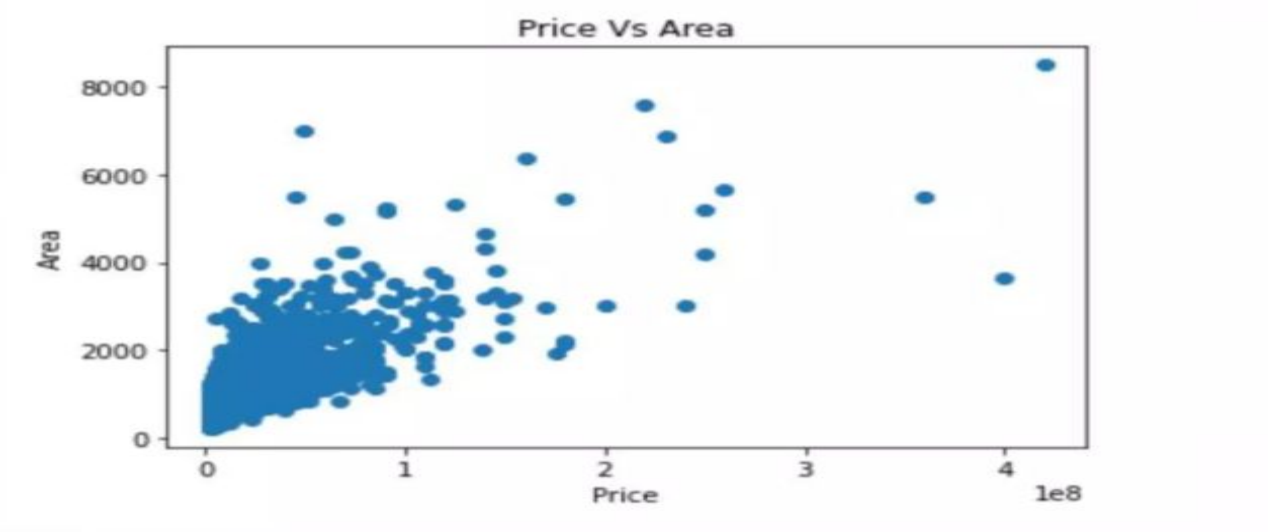
DATA VISUALIZATION,

4. BAR CHART : In the given Bar Chart, it shows that how many Property have how much balcony.



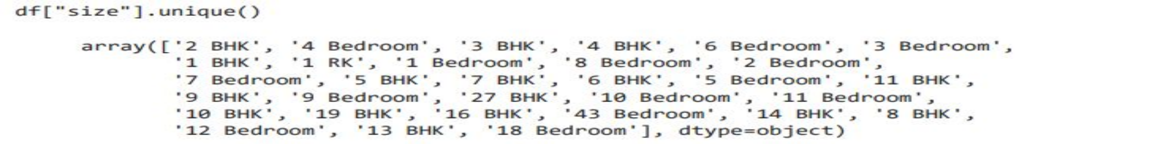
DATA VISUALIZATION,

5. <u>SCATTER PLOT</u> : In the given Scatter plot it shows that as the 'Price' of house increases with increase in 'Area' of house.



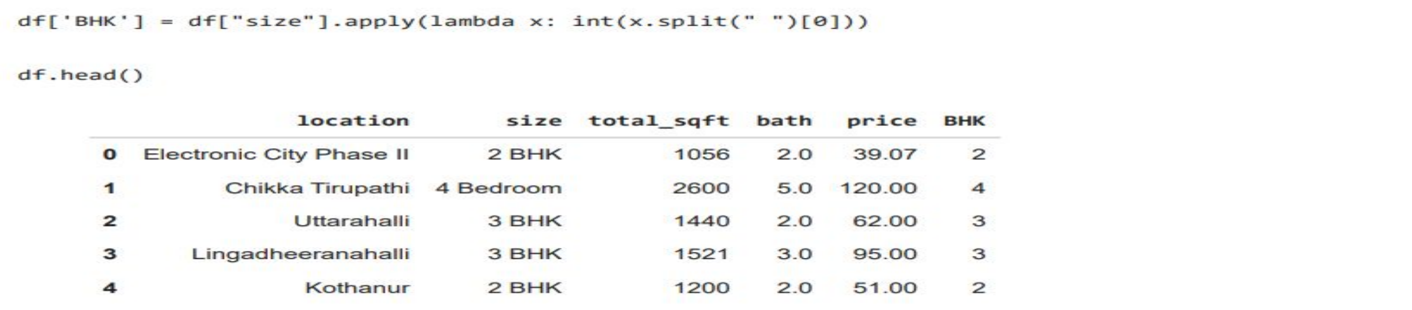
FEATURE ENGINEERING,

Firstly We get a Unique Function from Size,



```
df["size"].unique()

     array(['2 BHK', '4 Bedroom', '3 BHK', '4 BHK', '6 Bedroom', '3 Bedroom',
            '1 BHK', '1 RK', '1 Bedroom', '8 Bedroom', '2 Bedroom',
            '7 Bedroom', '5 BHK', '7 BHK', '6 BHK', '5 Bedroom', '11 BHK',
            '9 BHK', '9 Bedroom', '27 BHK', '10 Bedroom', '11 Bedroom',
            '10 BHK', '19 BHK', '16 BHK', '43 Bedroom', '14 BHK', '8 BHK',
            '12 Bedroom', '13 BHK', '18 Bedroom'], dtype=object)
```

From the above we can clearly see that Bedroom is represented with 2 different methods. One is BHK and the other one is Bedroom. So we are making a new column called BHK and we are discarding all the units (like BHK, Bedroom).

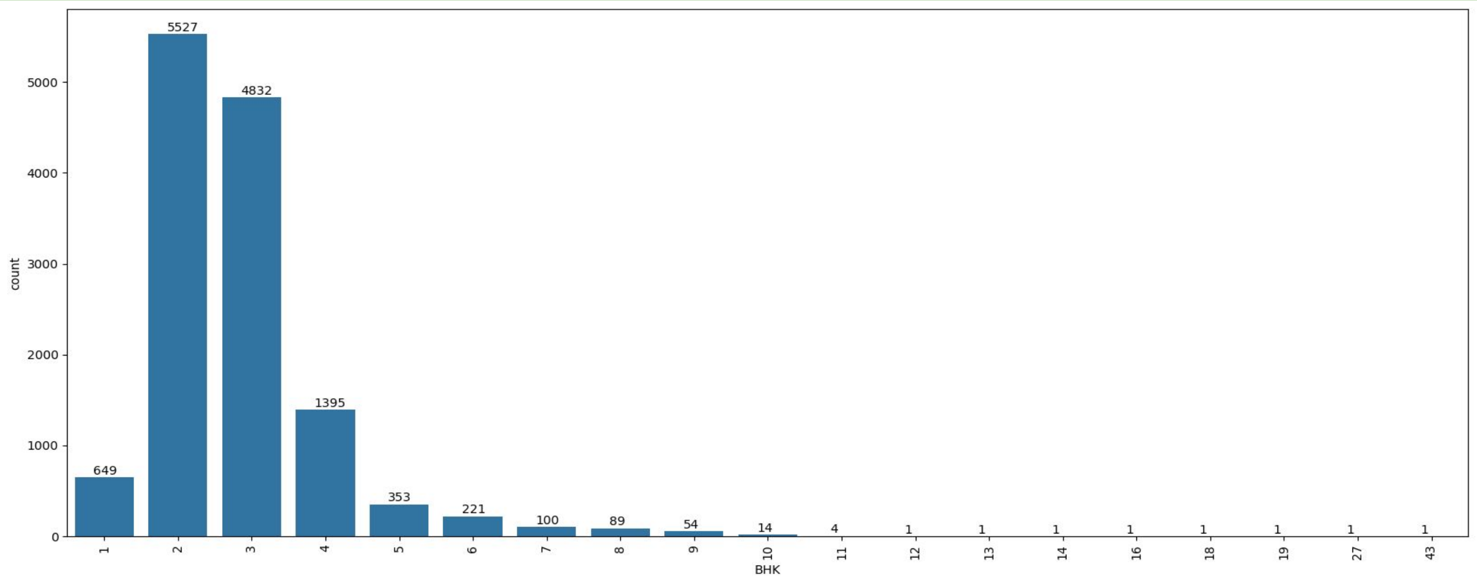


```
df['BHK'] = df["size"].apply(lambda x: int(x.split(" ")[0]))

df.head()
```

	location	size	total_sqft	bath	price	BHK
0	Electronic City Phase II	2 BHK	1056	2.0	39.07	2
1	Chikka Tirupathi	4 Bedroom	2600	5.0	120.00	4
2	Uttarahalli	3 BHK	1440	2.0	62.00	3
3	Lingadheeranahalli	3 BHK	1521	3.0	95.00	3
4	Kothanur	2 BHK	1200	2.0	51.00	2

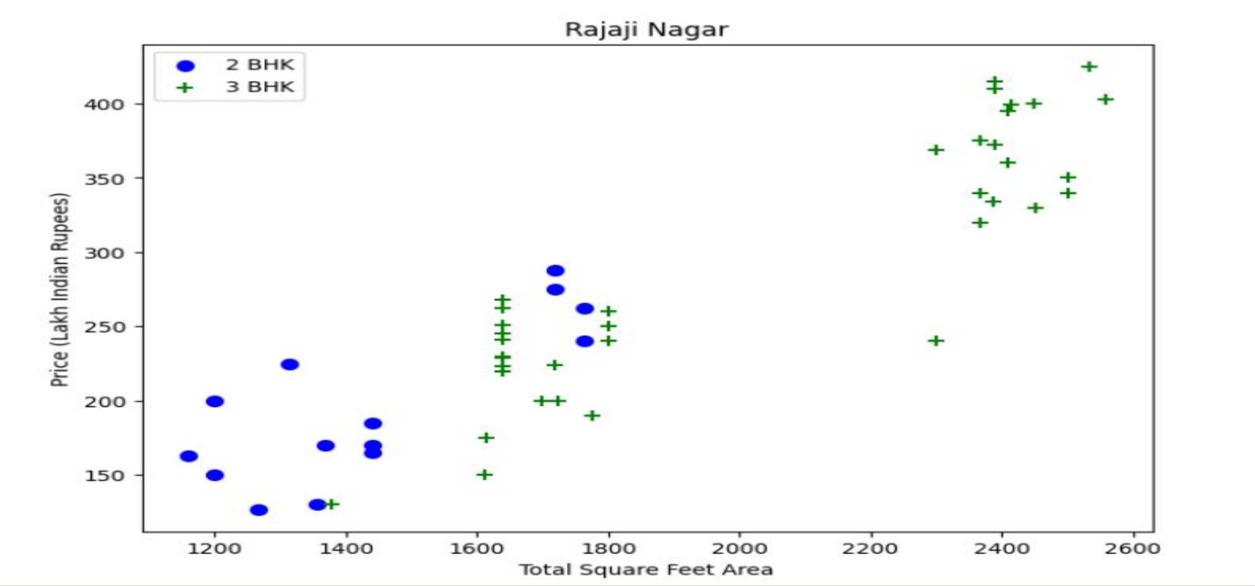
FEATURE ENGINEERING,

Secondly we created a Bar Chart, which shows how much property have how much BHK.



FEATURE ENGINEERING,

Now, we Plotting the Scatter Chart for 2 BHK and 3 BHK properties.



Aligarh College of Engineering and Technology

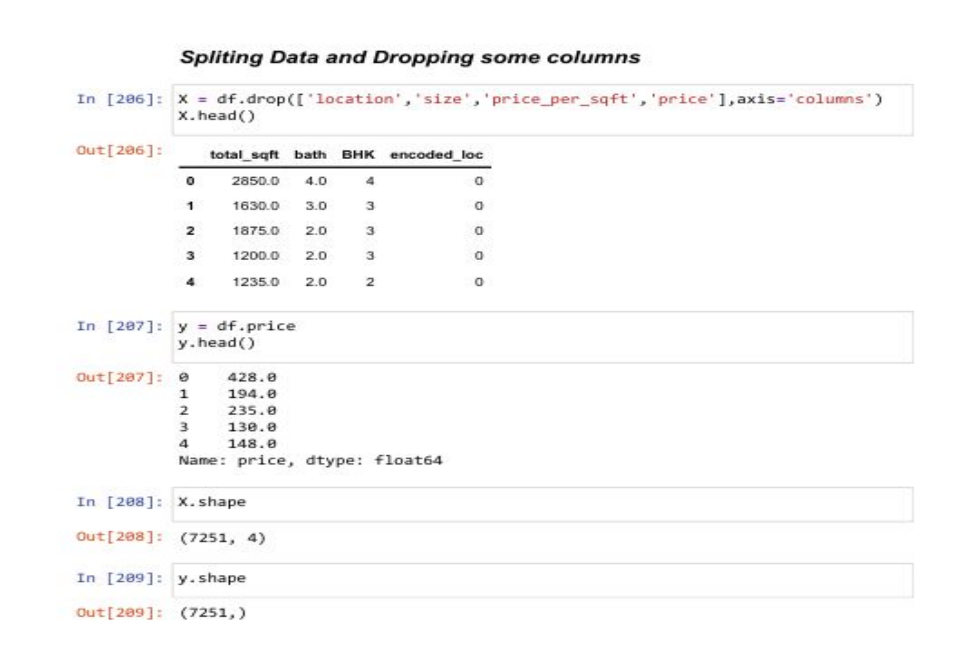
(APPROVED BY A.I.C.T.E & AFFILIATED TO AKTU, LUCKNOW)

THANK YOU

Hemant Sharma
(2001090100021)

MACHINE LEARNING MODEL DEVELOPMENT,

SPLITTING DATA,



Spliting Data and Dropping some columns

```
In [206]: X = df.drop(['location','size','price_per_sqft','price'],axis='columns')
          X.head()
```

Out[206]:

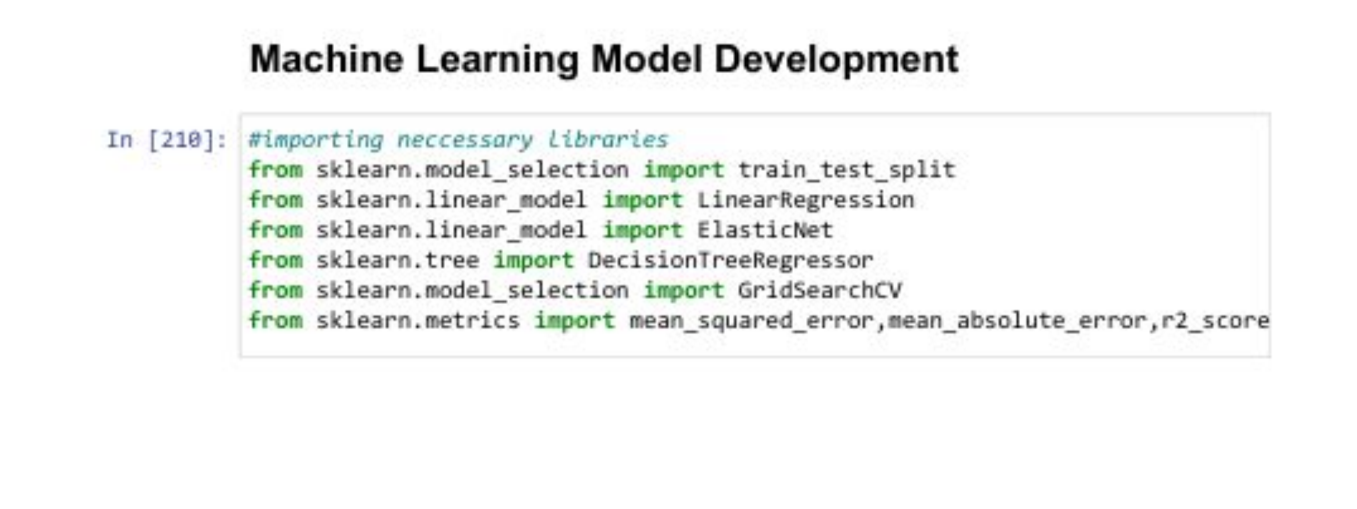
	total_sqft	bath	BHK	encoded_loc
0	2850.0	4.0	4	0
1	1630.0	3.0	3	0
2	1875.0	2.0	3	0
3	1200.0	2.0	3	0
4	1235.0	2.0	2	0

```
In [207]: y = df.price
          y.head()

Out[207]: 0    428.0
          1    194.0
          2    235.0
          3    130.0
          4    148.0
          Name: price, dtype: float64

In [208]: X.shape

Out[208]: (7251, 4)

In [209]: y.shape

Out[209]: (7251,)
```

Splitting the data for training and testing of model.

- Independent data as **X**
- Dependent data s **Y**

IMPORTING LIBRARIES,

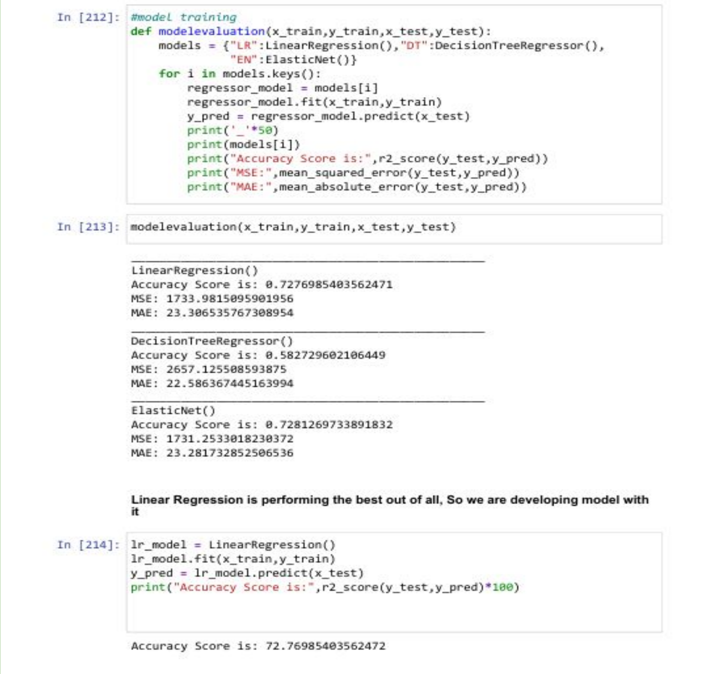


Machine Learning Model Development

In [210]:

```
#importing neccessary libraries
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import ElasticNet
from sklearn.tree import DecisionTreeRegressor
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import mean_squared_error,mean_absolute_error,r2_score
```

MODEL DEVELOPMENT,



```
In [212]: #model training
          def modelevaluation(x_train,y_train,x_test,y_test):
              models = {"LR":LinearRegression(),"DT":DecisionTreeRegressor(),
                        "EN":ElasticNet()}
              for i in models.keys():
                  regressor_model = models[i]
                  regressor_model.fit(x_train,y_train)
                  y_pred = regressor_model.predict(x_test)
                  print('_'*50)
                  print(models[i])
                  print("Accuracy Score is:",r2_score(y_test,y_pred))
                  print("MSE:",mean_squared_error(y_test,y_pred))
                  print("MAE:",mean_absolute_error(y_test,y_pred))
```

```
In [213]: modelevaluation(x_train,y_train,x_test,y_test)
```

```
__________________________________________________
LinearRegression()
Accuracy Score is: 0.7276985403562471
MSE: 1733.9815095901956
MAE: 23.306535767308954
__________________________________________________
DecisionTreeRegressor()
Accuracy Score is: 0.582729602106449
MSE: 2657.125508593875
MAE: 22.586367445163994
__________________________________________________
ElasticNet()
Accuracy Score is: 0.7281269733891832
MSE: 1731.2533018230372
MAE: 23.281732852506536
```

Linear Regression is performing the best out of all, So we are developing model with it

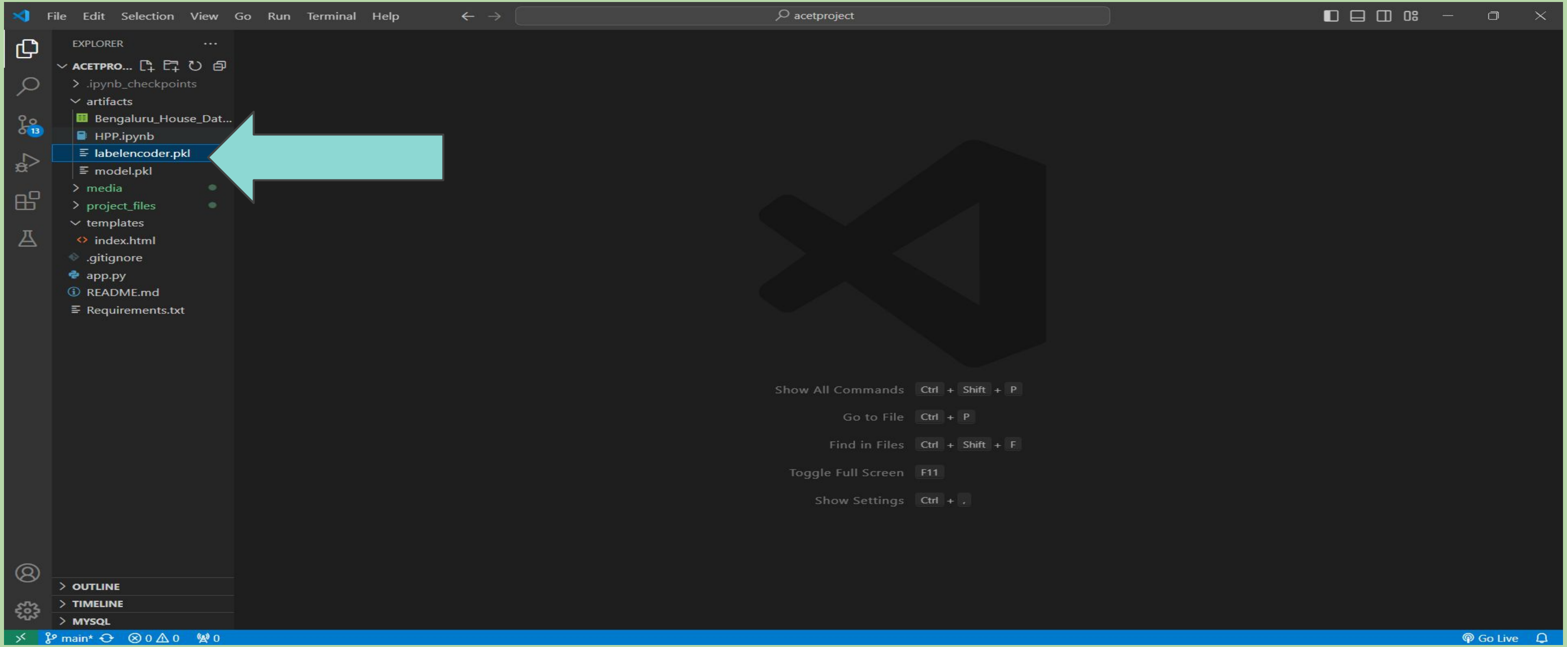
```
In [214]: lr_model = LinearRegression()
          lr_model.fit(x_train,y_train)
          y_pred = lr_model.predict(x_test)
          print("Accuracy Score is:",r2_score(y_test,y_pred)*100)
```

```
Accuracy Score is: 72.76985403562472
```

Building the ML Model by considering various algorithms and out of which we are selecting the best model.

- **Linear Regression** performing best of all model with **Accuracy Score** more than **70%**

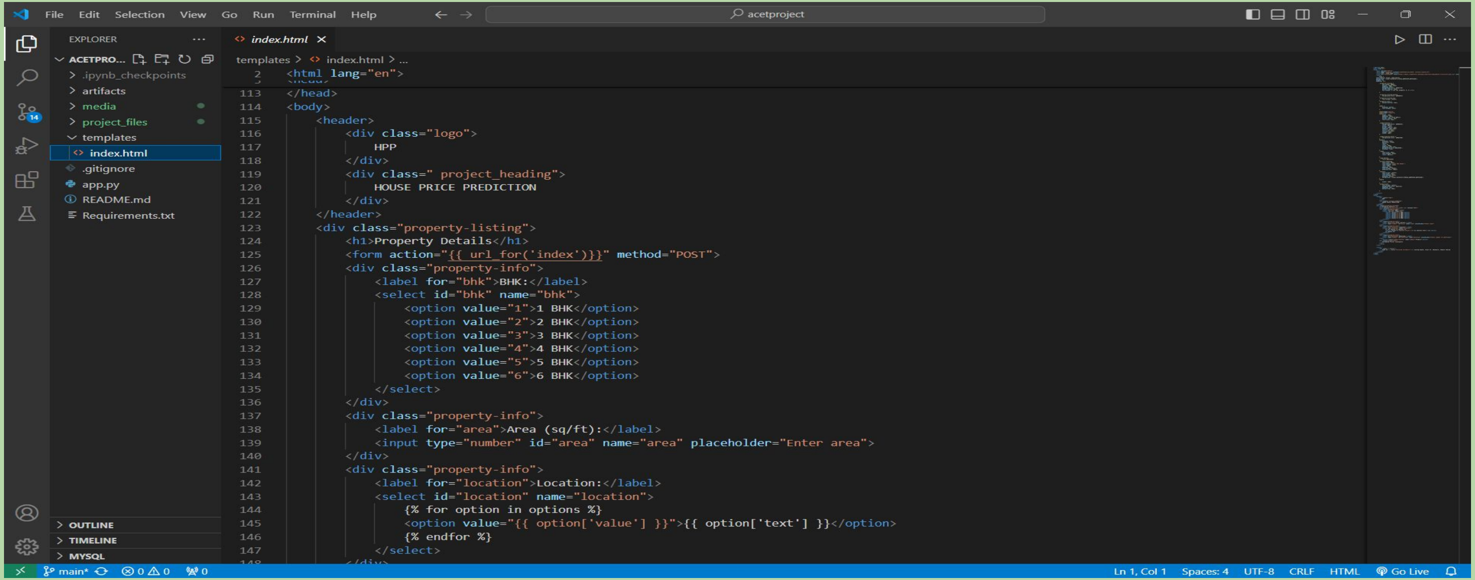
EXPORTING MODEL,



MODEL DEPLOYMENT,

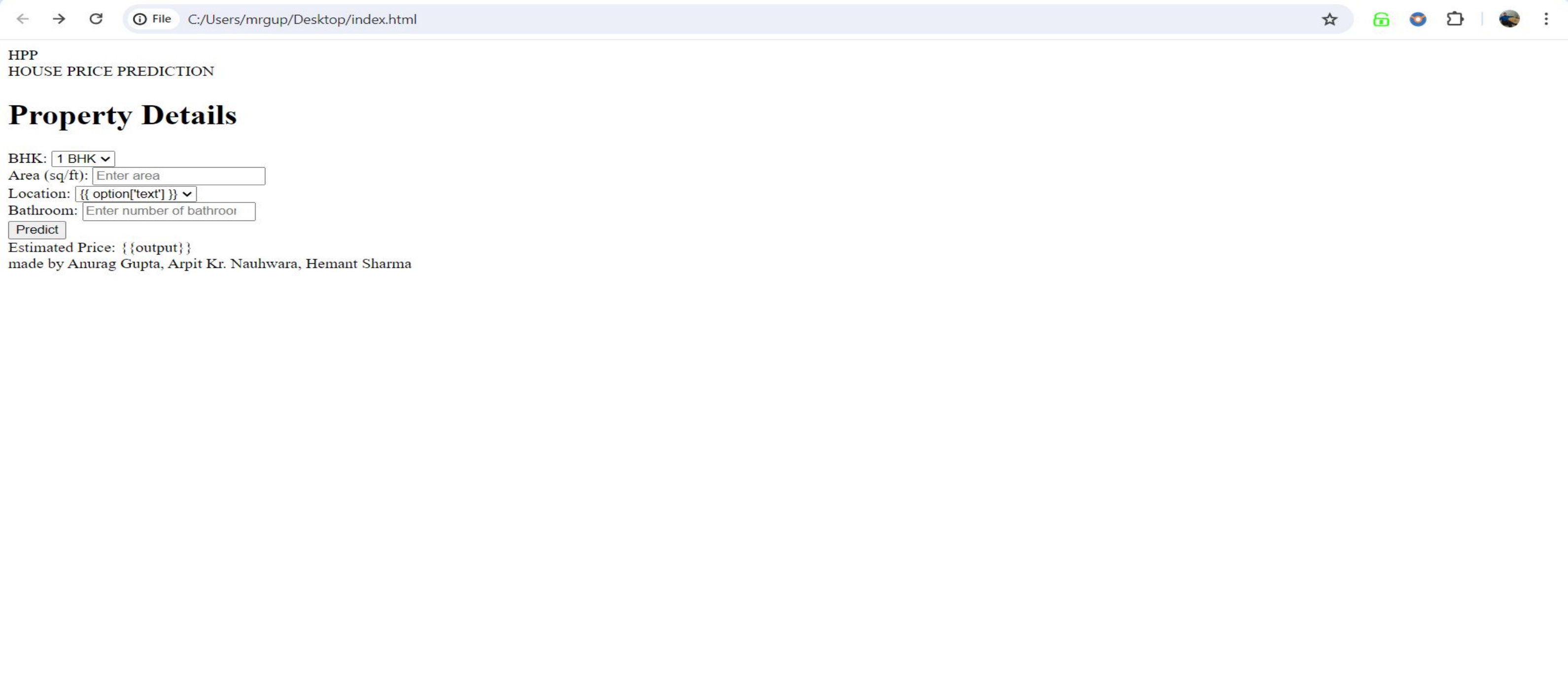
FRONTEND DEVELOPMENT,

1. HTML:



```html
<html lang="en">
</head>
<body>
    <header>
        <div class="logo">
            HPP
        </div>
        <div class=" project_heading">
            HOUSE PRICE PREDICTION
        </div>
    </header>
    <div class="property-listing">
        <h1>Property Details</h1>
        <form action="{{ url_for('index')}}" method="POST">
        <div class="property-info">
            <label for="bhk">BHK:</label>
            <select id="bhk" name="bhk">
                <option value="1">1 BHK</option>
                <option value="2">2 BHK</option>
                <option value="3">3 BHK</option>
                <option value="4">4 BHK</option>
                <option value="5">5 BHK</option>
                <option value="6">6 BHK</option>
            </select>
        </div>
        <div class="property-info">
            <label for="area">Area (sq/ft):</label>
            <input type="number" id="area" name="area" placeholder="Enter area">
        </div>
        <div class="property-info">
            <label for="location">Location:</label>
            <select id="location" name="location">
                {% for option in options %}
                <option value="{{ option['value'] }}">{{ option['text'] }}</option>
                {% endfor %}
            </select>
```

RAW WEBSITE,



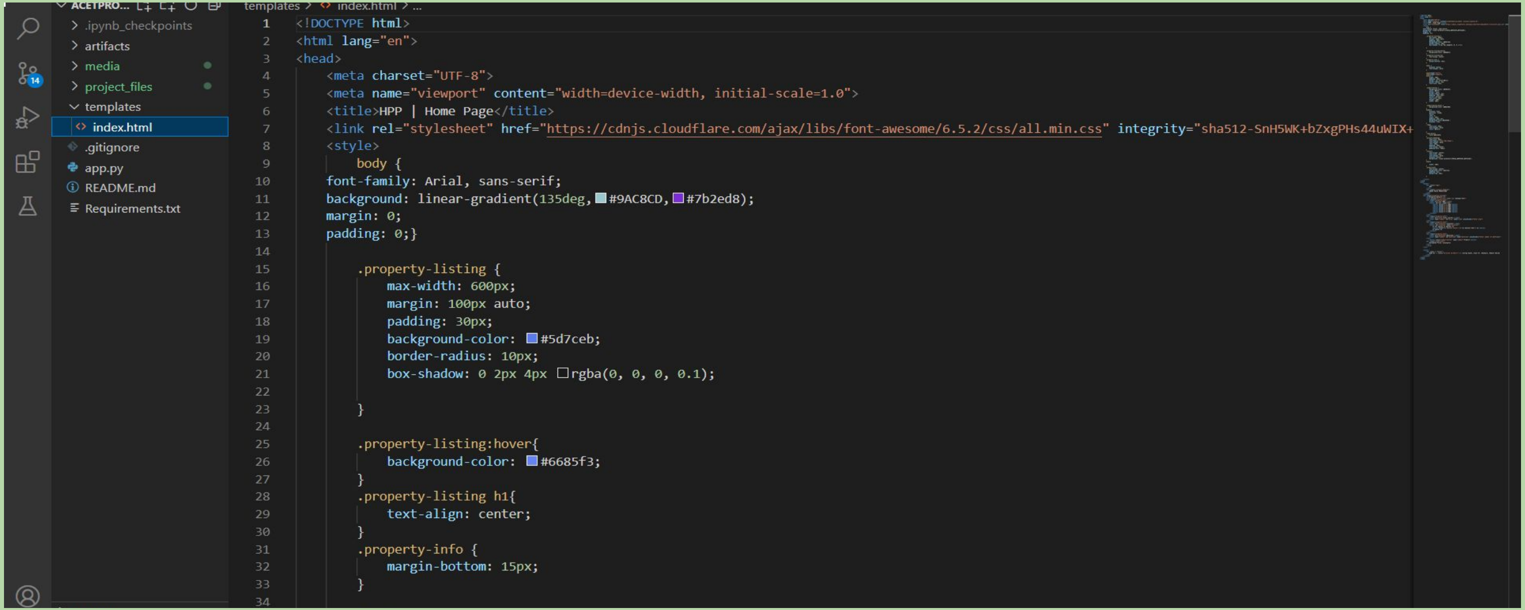
HPP
HOUSE PRICE PREDICTION

Property Details

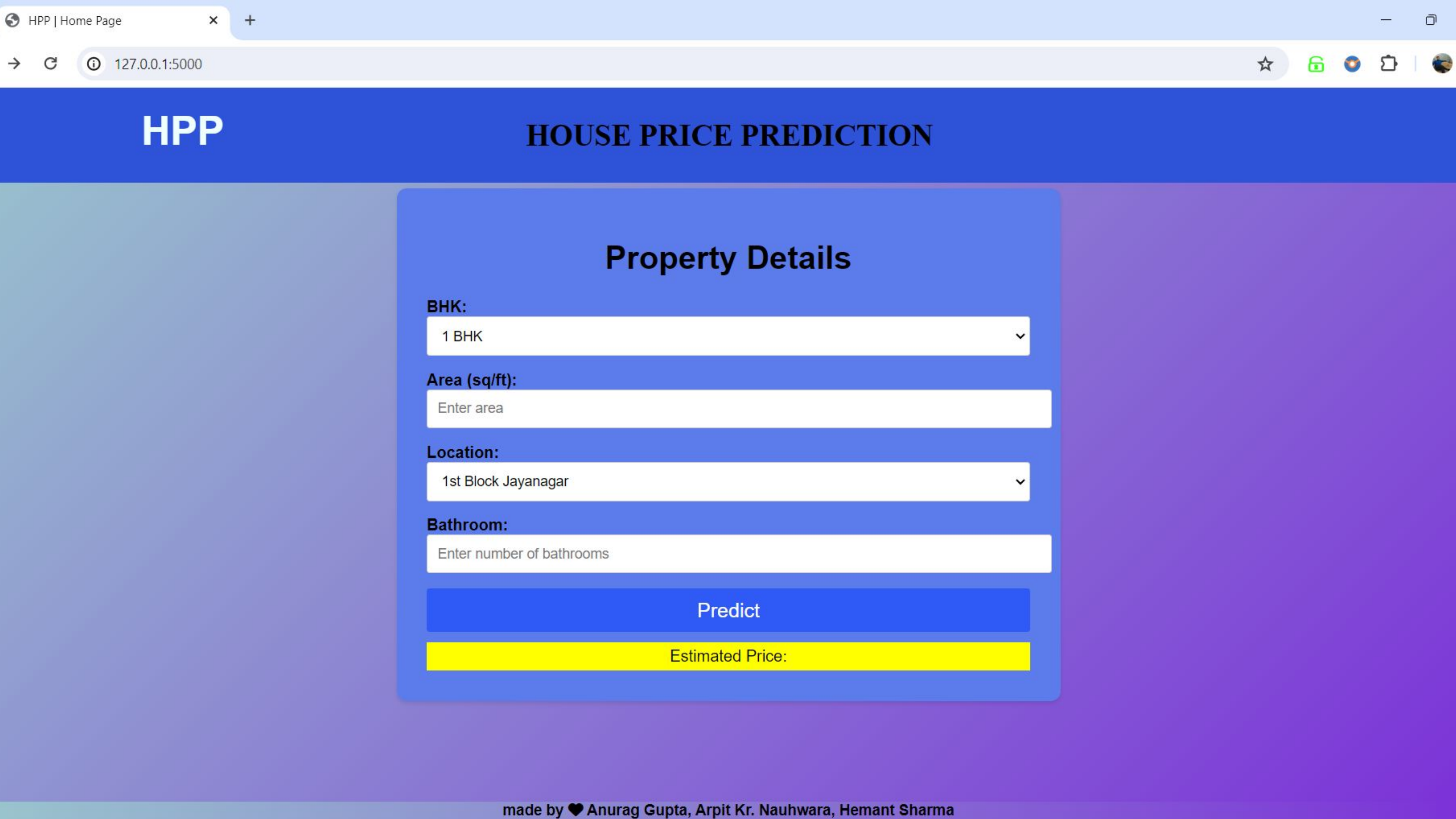
BHK: 1 BHK
Area (sq/ft): Enter area
Location: {{ option['text'] }}
Bathroom: Enter number of bathroor
Predict
Estimated Price: {{output}}
made by Anurag Gupta, Arpit Kr. Nauhwara, Hemant Sharma

FRONTEND DEVELOPMENT,

2. CSS



```
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>HPP | Home Page</title>
    <link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome/6.5.2/css/all.min.css" integrity="sha512-SnH5WK+bZxgPHs44uWIX+
    <style>
        body {
    font-family: Arial, sans-serif;
    background: linear-gradient(135deg, #9AC8CD, #7b2ed8);
    margin: 0;
    padding: 0;}

        .property-listing {
            max-width: 600px;
            margin: 100px auto;
            padding: 30px;
            background-color: #5d7ceb;
            border-radius: 10px;
            box-shadow: 0 2px 4px rgba(0, 0, 0, 0.1);

        }

        .property-listing:hover{
            background-color: #6685f3;
        }
        .property-listing h1{
            text-align: center;
        }
        .property-info {
            margin-bottom: 15px;
        }

```



HPP

HOUSE PRICE PREDICTION

Property Details

BHK: 1 BHK

Area (sq/ft): Enter area

Location: 1st Block Jayanagar

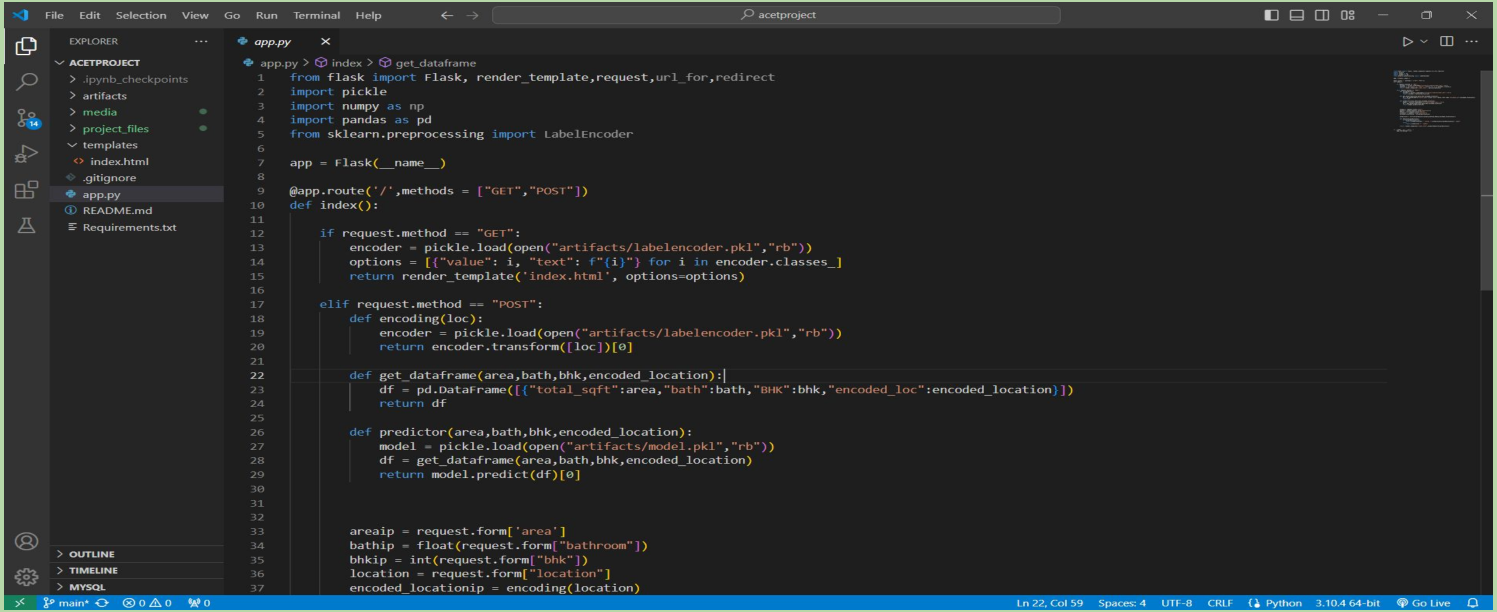
Bathroom: Enter number of bathrooms

Estimated Price:

made by ❤ Anurag Gupta, Arpit Kr. Nauhwara, Hemant Sharma

BACKEND DEVELOPMENT,

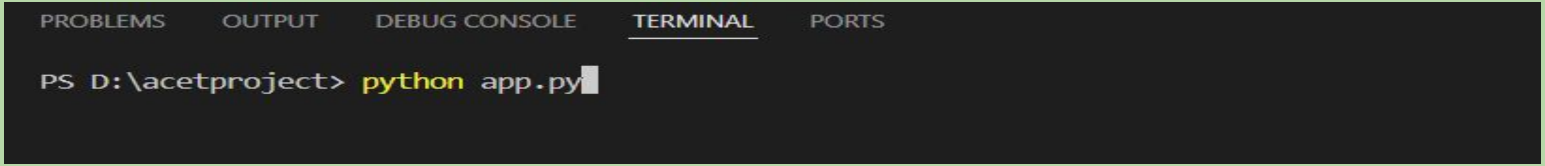
1. FLASK



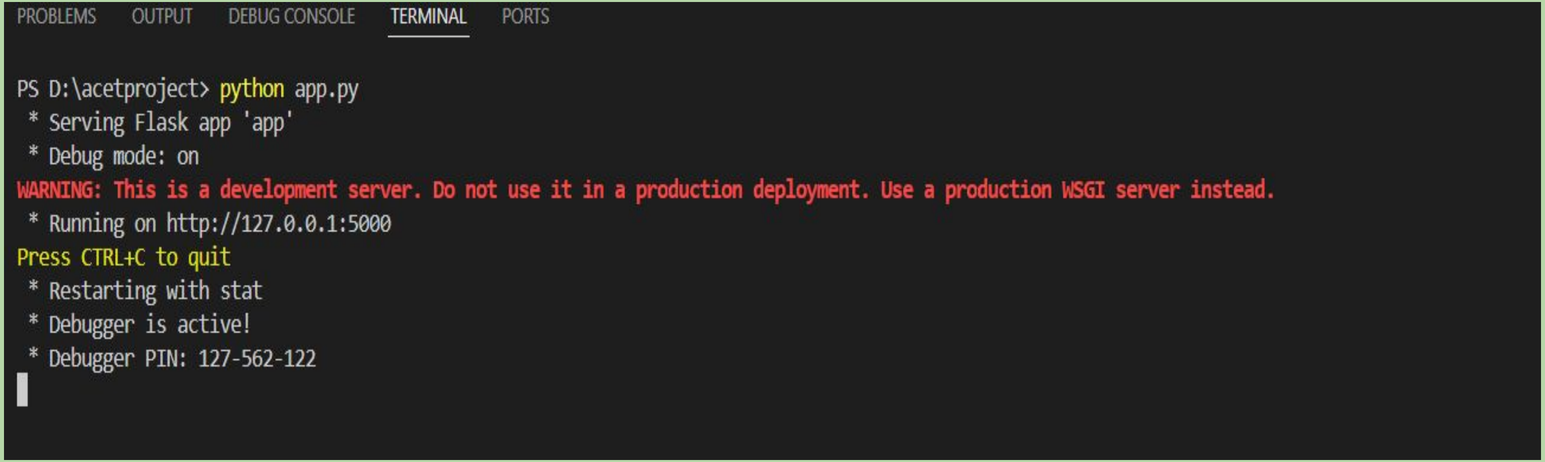
```python
from flask import Flask, render_template,request,url_for,redirect
import pickle
import numpy as np
import pandas as pd
from sklearn.preprocessing import LabelEncoder

app = Flask(__name__)

@app.route('/',methods = ["GET","POST"])
def index():

    if request.method == "GET":
        encoder = pickle.load(open("artifacts/labelencoder.pkl","rb"))
        options = [{"value": i, "text": f"{i}"} for i in encoder.classes_]
        return render_template('index.html', options=options)

    elif request.method == "POST":
        def encoding(loc):
            encoder = pickle.load(open("artifacts/labelencoder.pkl","rb"))
            return encoder.transform([loc])[0]

        def get_dataframe(area,bath,bhk,encoded_location):
            df = pd.DataFrame([{"total_sqft":area,"bath":bath,"BHK":bhk,"encoded_loc":encoded_location}])
            return df

        def predictor(area,bath,bhk,encoded_location):
            model = pickle.load(open("artifacts/model.pkl","rb"))
            df = get_dataframe(area,bath,bhk,encoded_location)
            return model.predict(df)[0]


        areaip = request.form['area']
        bathip = float(request.form["bathroom"])
        bhkip = int(request.form["bhk"])
        location = request.form["location"]
        encoded_locationip = encoding(location)
```

TUTORIAL,

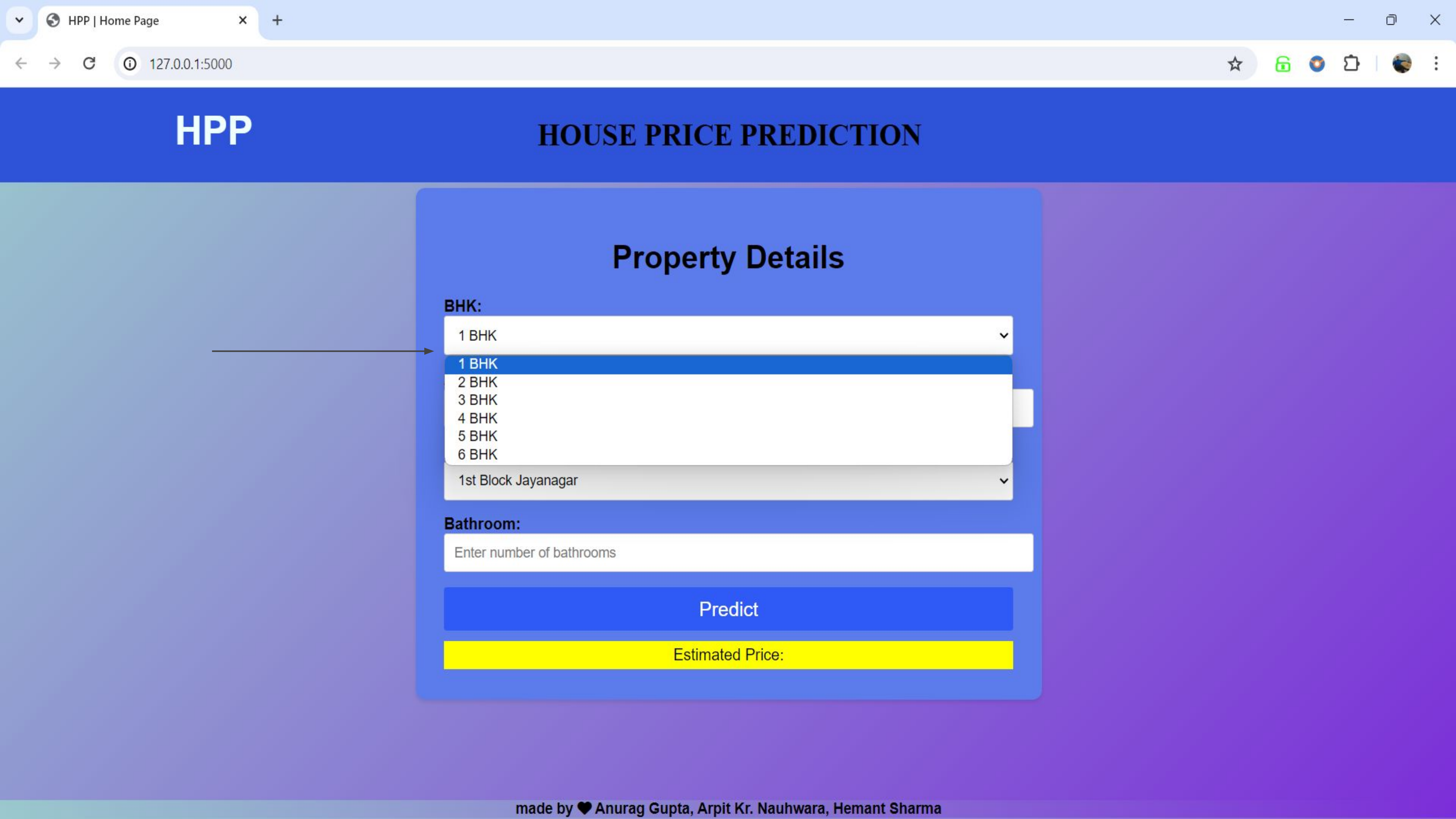
COMMAND TO RUN



```
PS D:\acetproject> python app.py
```



```
PS D:\acetproject> python app.py
 * Serving Flask app 'app'
 * Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
 * Running on http://127.0.0.1:5000
Press CTRL+C to quit
 * Restarting with stat
 * Debugger is active!
 * Debugger PIN: 127-562-122
```



HPP

HOUSE PRICE PREDICTION

Property Details

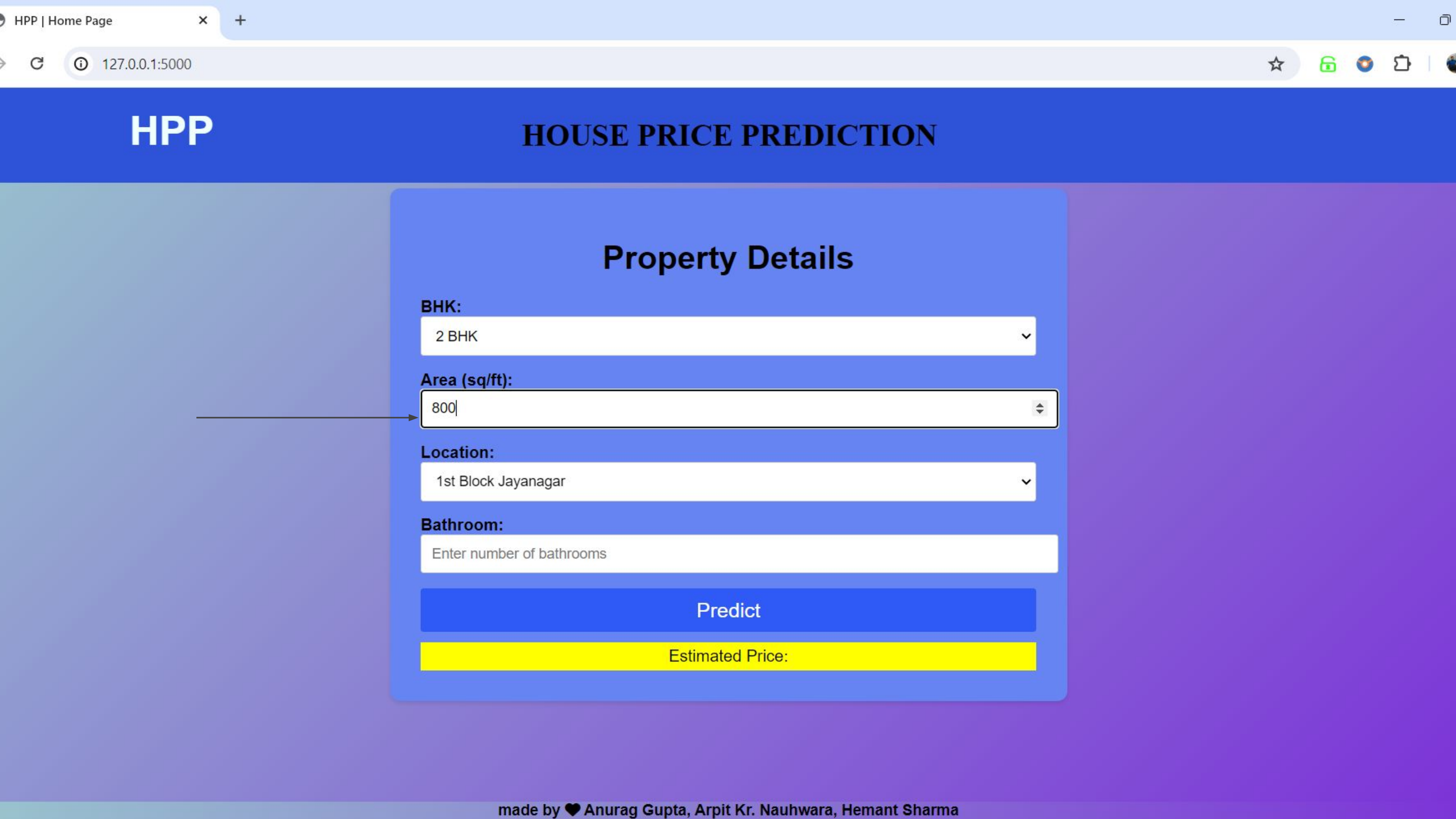
BHK: 1 BHK

1st Block Jayanagar

Bathroom: Enter number of bathrooms

Estimated Price:

made by ❤ Anurag Gupta, Arpit Kr. Nauhwara, Hemant Sharma



HPP

HOUSE PRICE PREDICTION

Property Details

BHK: 2 BHK

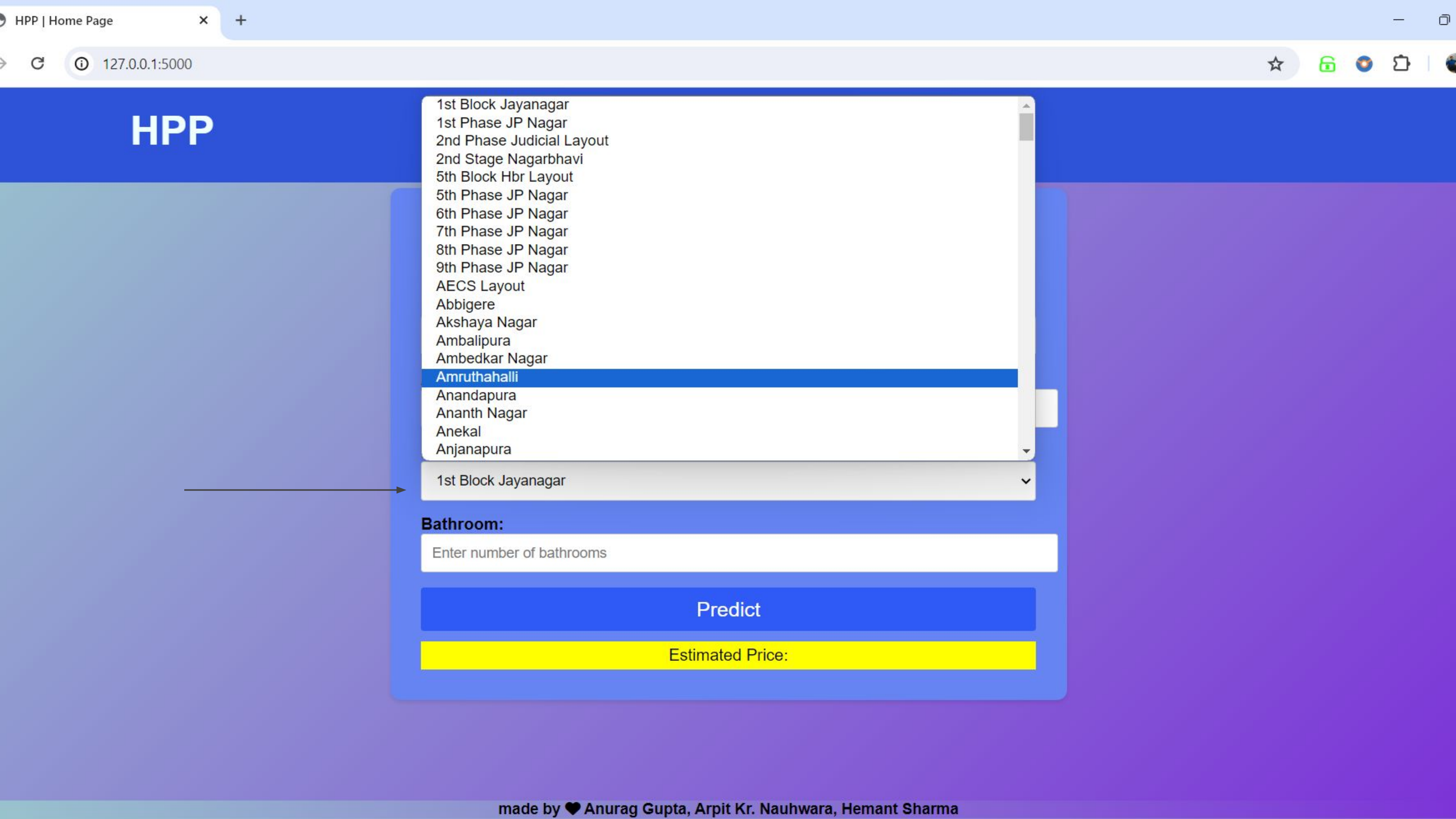
Area (sq/ft): 800

Location: 1st Block Jayanagar

Bathroom: Enter number of bathrooms

Estimated Price:

made by ❤ Anurag Gupta, Arpit Kr. Nauhwara, Hemant Sharma



HPP

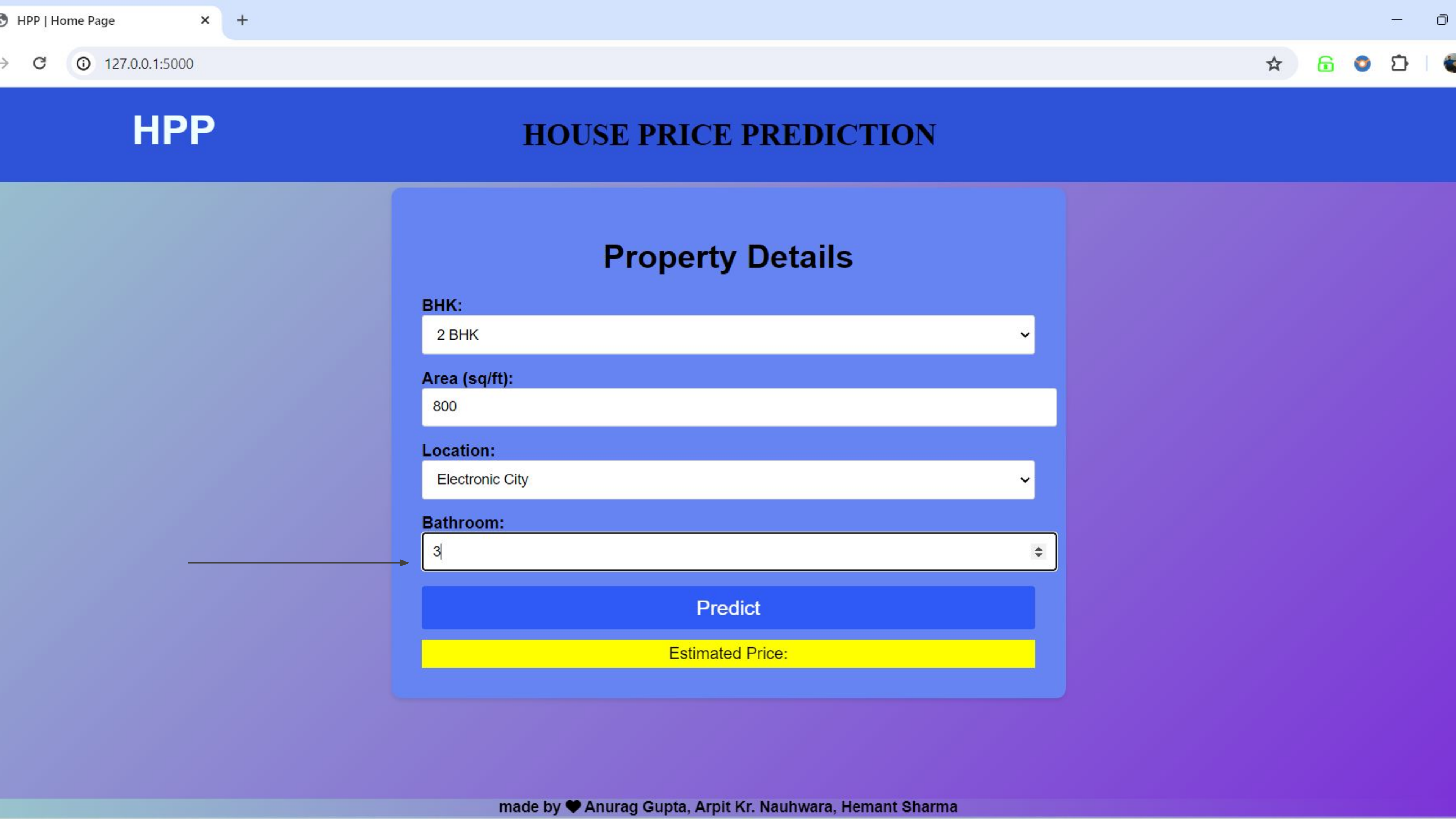
1st Block Jayanagar

Bathroom:

Enter number of bathrooms

Estimated Price:

made by ❤ Anurag Gupta, Arpit Kr. Nauhwara, Hemant Sharma



HOUSE PRICE PREDICTION

Property Details

BHK: 2 BHK

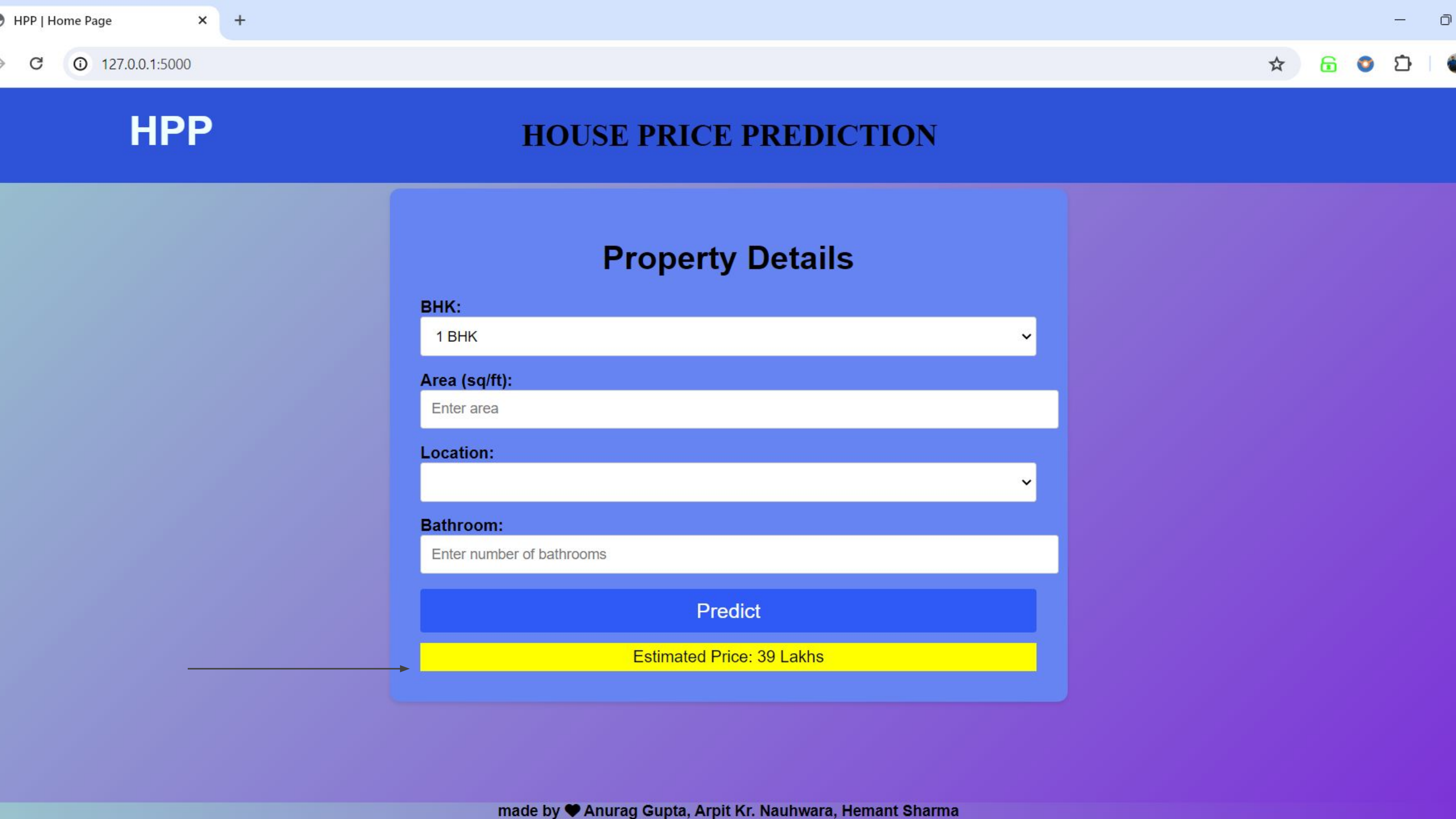
Area (sq/ft): 800

Location: Electronic City

Bathroom: 3

Estimated Price:

made by ❤ Anurag Gupta, Arpit Kr. Nauhwara, Hemant Sharma



HPP

HOUSE PRICE PREDICTION

Property Details

BHK:
1 BHK

Area (sq/ft):
Enter area

Location:

Bathroom:
Enter number of bathrooms

Estimated Price: 39 Lakhs

made by ❤ Anurag Gupta, Arpit Kr. Nauhwara, Hemant Sharma

(APPROVED BY A.I.C.T.E & AFFILIATED TO AKTU, LUCKNOW)

THANKYOU!

Anurag Gupta	(2001090100004)
Arpit Kumar Nauhwara	(2001090100006)
Hemant Sharma	(2001090100021)