

# Second Progress Report

of

## Project - I

**Subject Code: 4IT31**

**Academic Year 2021-22**

**Group Number : G - 13**

**Name of Students : Kalp Gohil 18IT405**  
**Man Desai 18IT404**  
**Param Shah 18IT413**  
**Jay Patel 17IT419**

**Topic : House Price Prediction**

**Guide Name : Prof. Prachi Shah**

**BACHELOR OF ENGINEERING**

*In*

**INFORMATION TECHNOLOGY**



**Birla Vishvakarma Mahavidyalaya**  
**Engineering College,**  
**Vallabh Vidhyanagar-388120**

## **Table of Content**

<b>Sr. No.</b>	<b>Topics</b>	<b>Page No.</b>
<b>1</b>	<b>Definition and Overview</b>	<b>3</b>
<b>2</b>	<b>Work Done on Modules</b>	<b>4</b>
<b>3</b>	<b>Remaining work and future direction</b>	
<b>4</b>	<b>Conclusion</b>	

## **1 - Definition and Overview :**

### **1.1 - Definition :**

Predict House Price based on some common parameters like number of rooms, total area, type of area, number of bathrooms, etc...

### **1.2 – Overview :**

As a part of second progress report we have completed data cleaning process. With that we had also completed data pre-processing part on raw data. WE had also completed some part of user interface like home page of website, signup page. Also user signing up process has been completed with backend that will store information about user.

## 2 - Work Done on Modules :

### ❖ Sign-Up / Login:

- **Description:** While designing website for any purpose (like for public use or for business part), it is very important to design user sign up page. This will help to analyze our website performance.
- In this system user will only be able to visit home page, about page, in case he/she is not login or is not valid user as per database.

### Code:

```
def register(request):  
    if request.method == "POST":  
        fname = request.POST['fname']  
        lname = request.POST['lname']  
        username = request.POST['username']  
        email = request.POST['email']
```

## 4IT31: Project - I

---

```
password1 = request.POST['password1']
password2 = request.POST['password2']

if password1 == password2:
    if User.objects.filter(username=username):
        messages.info(request, "Username already taken...!!")
        return redirect('register')
    elif User.objects.filter(email=email):
        messages.info(request, "Email already taken...!!")
        return redirect('register')
    else:
        user = User.objects.create_user(username=username, email=email,
password=password1,first_name=fname, last_name=lname)
        user.save()
        print("USER CREATED")
    else:
        messages.info(request, "Password does not match...!!")
        return redirect('register')
    return redirect('/')

else:
    return render(request, 'register.html')
```

## ❖ Data Cleaning & Pre-Processing:

- **Description:** In this module, first of all, data set has been imported, then data has been cleaned and pre processed in order to predict accurate price.

### Code:

- **Data Importing:**

```
import pandas as pd
house = pd.read_csv('Bengaluru_House_Data.csv')
house.head(5)
```

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00

- **Analyzing Data:**

```
house.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13320 entries, 0 to 13319
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  -
0   area_type      13320 non-null  object
1   availability  13320 non-null  object
2   location        13319 non-null  object
3   size            13304 non-null  object
4   society       7818 non-null  object
5   total_sqft      13320 non-null  object
6   bath            13247 non-null  float64
7   balcony       12711 non-null  float64
8   price           13320 non-null  float64
dtypes: float64(3), object(6)
memory usage: 936.7+ KB
```

## Things to be taken care of...!!

- area\_type, availability, society have different categorical data. So we have to drop this columns.

- location one None values.
- size has BHK, Bedroom attached with desire data. Remove this unnecessary string from data.
- size has many None values.
- size is as string not as int.
- total\_sqft has many None values.
- total\_sqft has Meter attached with desire data. Remove this unnecessary string from data.
- total\_sqft has many range values.
- total\_sqft is as string not as float.
- bath has many None values.
- balcony has many None values. Fill na with mean.
- balcony is stored as float. Change it to int to reduce memory used.
- Change name of column size to size\_BHK.
- Change name of column bath to bathroom.
- Change name of column price to price\_lakhs.

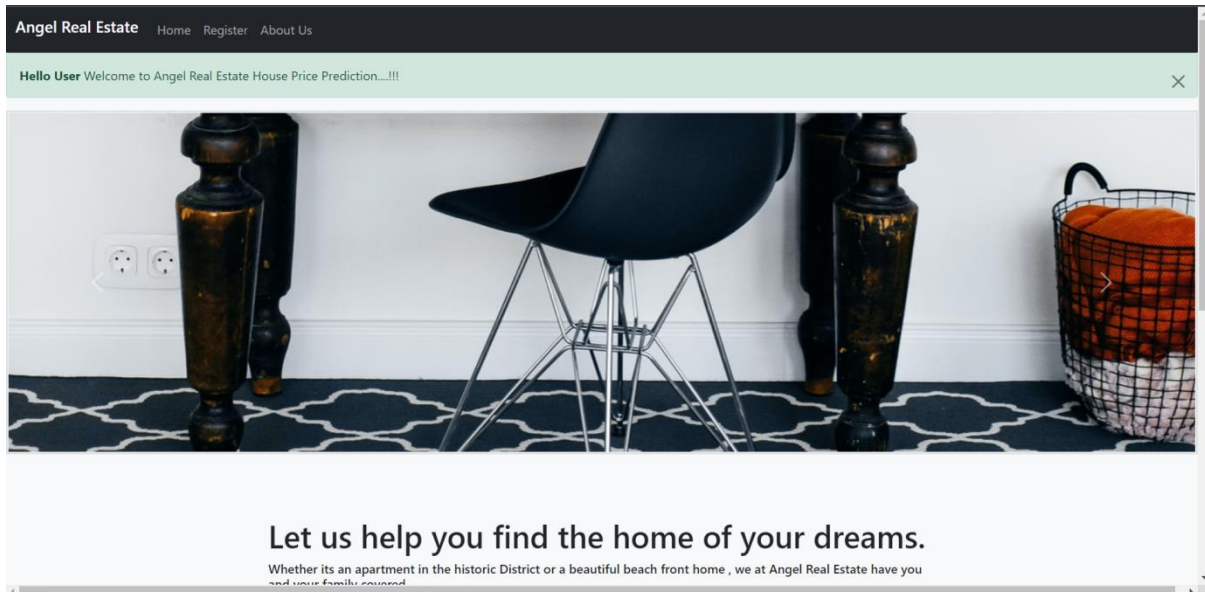
- **Dimensionality Reduction.**

Examine location which is a categorical data. We need to convert this categorical data in numerical type data using one hot encoding. But there are total 1295 unique values in location column. So after applying one hot encoding our data set will have very very large number of columns. So we need to dimensionally reduce location column.

```
location_state = house['location'].value_counts(ascending=False)
location_state_removal = location_state[location_state <= 10]
house['location'] = house['location'].apply(lambda x: "other" if x in
location_state_removal else x)
```

### ❖ UI Design:

- **Description:** In this module we had design one simple website based app having page like home-page, register-page, about-page,etc...





### 3 - Remaining work and Future Planning :

As Part of first mid we have completed around 40% - 45% work and we are remaining with following work module wise:

**1) Login and Signup:**

So, we have almost completed this module but if any changes or error occur in future during testing then we need to look into it.

**2) Model Training:**

- We are planning to train our model using Linear Regression but if time permits we will try to go for other options available to measure accuracy.

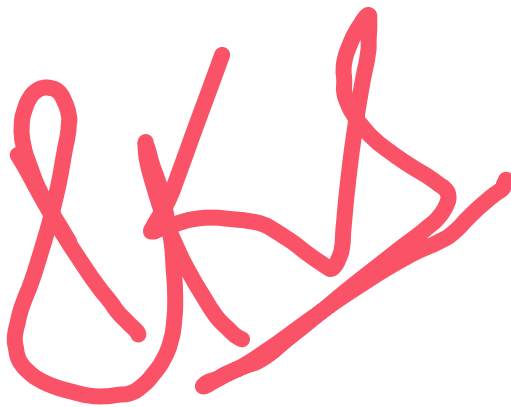
**3) UI Design:**

- In UI design we require complete login page and page that will interact with user to get parameters of a house whole price need to be predicted.
  - We have completed user login and creation of admin, But user logout page is still remaining.
  - And many more stuff is remaining in UI and Backend.
- So, next we are planning to complete almost 70% - 80% of project before the second mid presentation. In that time we will complete our model training part(and comparison with other method if time permits.) along with its UI integration and also some static pages like as mentioned above.

## 4 - Conclusion:

As part of first mid presentation first of we have completed our SRS as first progress report and this is the second progress report and also, we are done with around 40% - 45% work on actual system implementation. If some changes are required then we will take that into consideration and do changes in completed work.

\*\*\*\*\**The End*\*\*\*\*\*

A handwritten signature in red ink, consisting of stylized, overlapping loops and lines, positioned in the center of the page.