**CONTENTS**

|  |  |
| --- | --- |
| **Chapter** | **Page** |
| **Chapter 1**  **1.1 Introduction**  **1.2 Dataset**  **1.2.1 Area\_type** | **3**  **3**  **3** |
| **1.2.2 Availability**  **1.2.3 Location**  **1.2.4 Size** | **4**  **4**  **4** |
| 1**.2.5 Society**  **1.2.6 Total\_sqft**  **1.2.7 Bath**  **1.2.8 Balcony**  **1.2.9 Price** | **5**  **5**  **5**  **5**  **5** |
| **1.3 Problem Statement**  **1.4 Motivation**  **1.5 Summary**  **Chapter 2**  **2.1 Background**  **2.1.1 Introduction** | **6**  **6**  **6**  **6**  **6** |
| **2.2 Literal Review**  **2.3 Problem Analysis**  **2.4 Summary**  **Chapter 3**  **Proposed Model, Milestone, Schedule, DFD, Use Case Diagram, ER-Diagram**  **3.1 Proposed Model** | **7**  **7**  **7**  **7**  **7** |
| **3.2 Milestone**  **3.3 Schedule**  **3.4 DFD (Data Flow Diagram)** | **8**  **8**  **8** |
| **3.5 Use Case Diagram**  **3.6 ER-Diagram** | **9**  **9** |
| **Chapter 4**  **Implementation and Result**  **4.1 Implementation**  **4.1.1 Code** | **10**  **10**  **10** |
| **4.1.2 Output** | **16** |
| **4.2 System Setup** | **25** |
| **4.3 Results and Discussion**  **4.4 Existing System Analysis**  **4.5 Summary**  **Chapter 5**  **Standard, Ethics, Challenge**  **5.1 Standard** | **26**  **26**  **26**  **26**  **26** |
| **5.2 Ethics**  **5.3 Challenges**  **5.4 Summary**  **Chapter 6**  **Conclusion**  **6.1 Introduction** | **27**  **27**  **27**  **27**  **27** |
| **6.2 Future Work**  **Reference**  **Figures** | **28**  **28**  **29** |

**Chapter 1**

**1.1 Introduction**

Home Rent Prediction Systemis an AI-based program that can predict the house price by analyzing the location, size of the room, and the total amount of bed-room. Here, we used a dataset that contains various information about houses including house prices. Our program predicts rent by analyzing the used dataset. We used Python programming language which is one of the best high level programming language for Artificial Intelligence.

**1.2 Dataset**

Our project Home Rent Prediction System contains a dataset named Mirpur\_House\_Data which is in CSV (comma separated values) format. because, CSV is a simple file format used to store tabular data, such as a spreadsheet or database. CSV file stores tabular data (numbers and text) in plain text. Each line of the file is a data record.

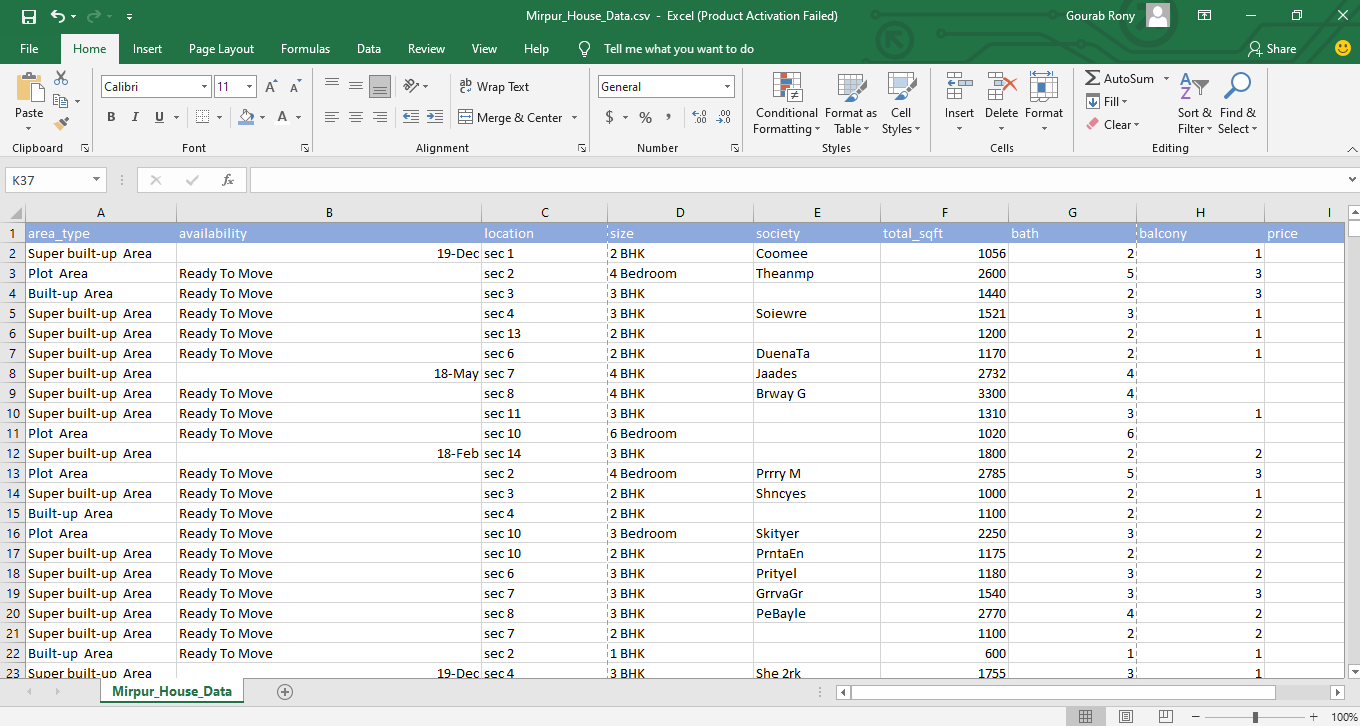


Figure 1: Mirpur house data

**1.2.1 Area\_type**

Area\_type is the first column in our dataset. It shows us different types of areas of the location. We have four types of areas in our dataset, Super built-up Area, Plot Area, Built-up Area, and Carpet Area.

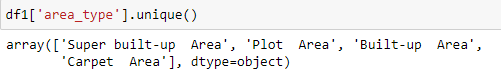


Figure 2: area\_type

**1.2.2 Availability**

Availability is the second column in our dataset. It shows us the current status of houses which are available for rent or not.

**1.2.3 Location**

Location is the third column in our dataset. It shows us the location of the houses. We have a total of fourteen unique locations in our dataset.

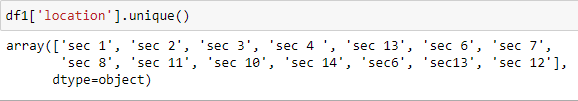


Figure 3: location

**1.2.4 Size**

Size is the fourth column in our dataset. It shows us the size of houses in BHK (bedroom, hall, and kitchen) format.

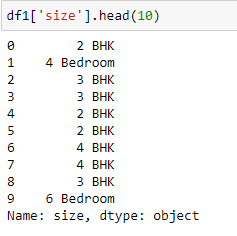


Figure 4: size

1**.2.5 Society**

Society is the fifth column in our dataset. It shows us different types of the society of houses.

**1.2.6 Total\_sqft**

Total\_sqft is the sixth column in our dataset. It shows us the total size of houses in the square feet format.

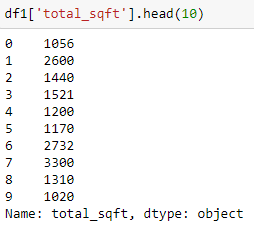


Figure 5: total\_sqft

**1.2.7 Bath**

Bath is the seventh column of our dataset. It shows us the total number of baths that houses contain.

**1.2.8 Balcony**

Balcony is the eighth column of our dataset. It shows us the total number of balconies that houses contain.

**1.2.9 Price**

Price is the last column in our dataset. It shows us the rent of houses.

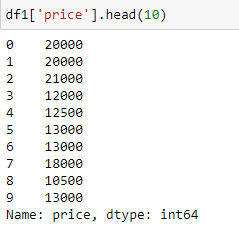


Figure 6: Price

**1.3 Problem Statement**

In Bangladesh's capital city Dhaka, searching for houses that are available for rent is a very hard-working job for the tenant. There are many websites that help them to make their job easy, but it is not worth it. Because tenant looks for houses by their rent and number of the room. The existing system is not helping the tenant with their need.

**1.4 Motivation**

A better idea is to use the system which can improve the current system and help the tenants. This system is providing each entity the facility to approach the tenant by predicts the rent so that it will become much easier to find houses which is available for rent.

**1.5 Summary**

To find a house which is available for rent is a very hard-working job. Our system can make the job a little easy. And tenants can save a lot of time to find their beautiful house.

**Chapter 2**

**2.1 Background**

**2.1.1 Introduction**

The majority of the people in Dhaka have come to Dhaka for a variety of economic and social reasons. They are staying and working in Dhaka permanently without a home of their own. The experiences and the house ownership process of the tenants generate knowledge and present a guideline for the policymakers and planners. According to research, about 68% of people who lived in Dhaka are tenants. So, it's very difficult for them to search for available houses on their own without any guidelines.

**2.2 Literal Review**

This work provides a literature review of the house finding process. We aim to help the tenant to save their time, make easy their job and improve the current system. With this system, tenants can find their house easier way and fast. In this system when tenants search by house square feet size, the number of rooms, and a number of bathrooms, it shows the suitable rent. So, our system can help the tenants.

**2.3 Problem Analysis**

In Dhaka, there are 60% to 65% of people are tenants. But the number of house owners is low who are willing to give rent their house. It is less than 20%. So, it's very hard to find a house for tenants. Tenants have to go door to door for searching the house or searching a house on different websites. On websites, tenants have only one option for finding the houses. it's the location option. So, it's a very stressful job to find a favorable house at an affordable price. Sometimes they cannot find any and waste their time.

**2.4 Summary**

Dhaka is the capital of Bangladesh with a population of 8,906,039. Every day thousands of people come here looking for hope. Our system is made for those hopeful tenants.

**Chapter 3**

**Proposed Model, Milestone, Schedule, DFD, Use Case Diagram, ER-Diagram**

**3.1 Proposed Model**

The scope of the project is that in a very short span provides the user with many facilities. It provides a very accurate prediction with an accuracy of more than 60%. So, users can trust it. It shows users the predicted rent when they give the location, total square feet of house, total bedroom, and total bath as input. The main purpose of this project is to reduce users' work to find their houses which are available for rent. This system predicts the rent by analyzing the given dataset used in the system. We modified the dataset so that the system can give the user accurate prediction.

**3.2 Milestone**

* To reduces the users' work as much as possible to search houses which are available for rent.
* To use a statistical dataset which can give us all information about the house rent.
* To make the system as simple as possible
* To use an algorithm that can give us our rent prediction result

**3.3 Schedule**

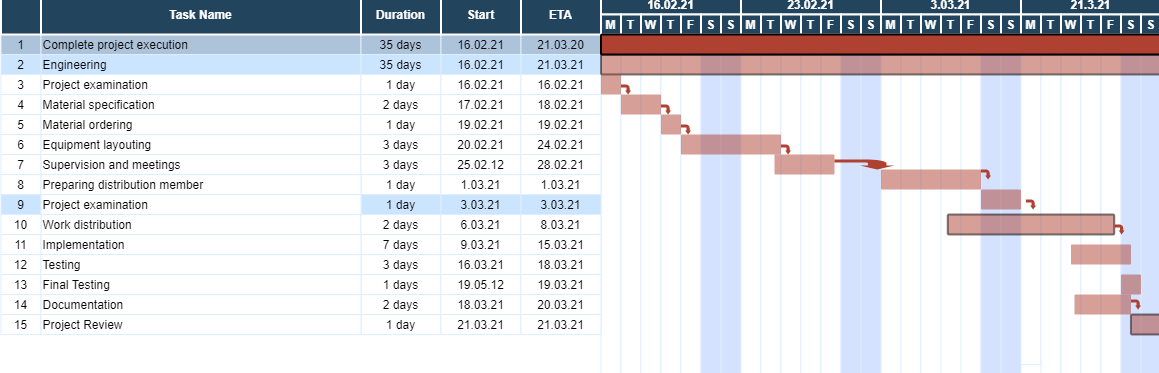
****

Figure 7: Schedule

**3.4 DFD (Data Flow Diagram)**

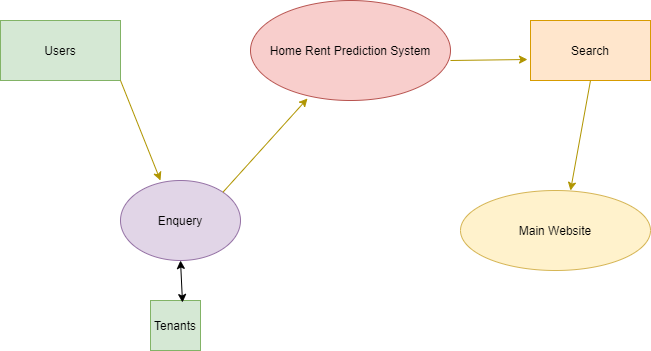
****

Figure 8: DFD

**3.5 Use Case Diagram**

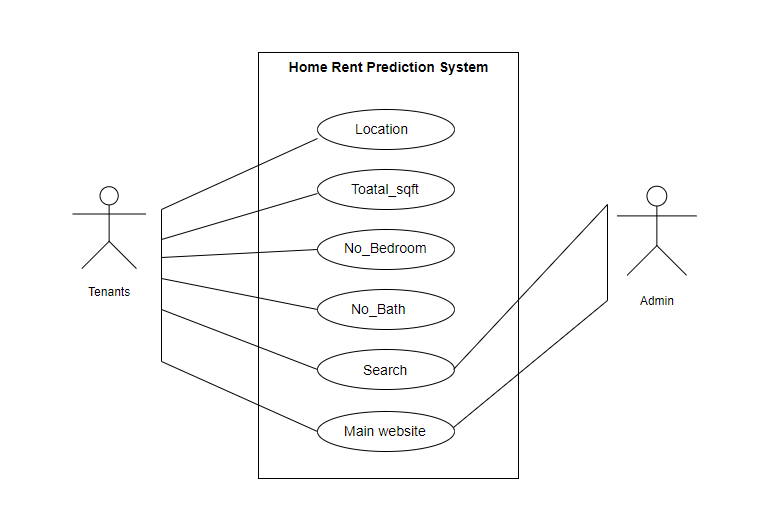


Figure 9: Use case diagram

**3.6 ER-Diagram**

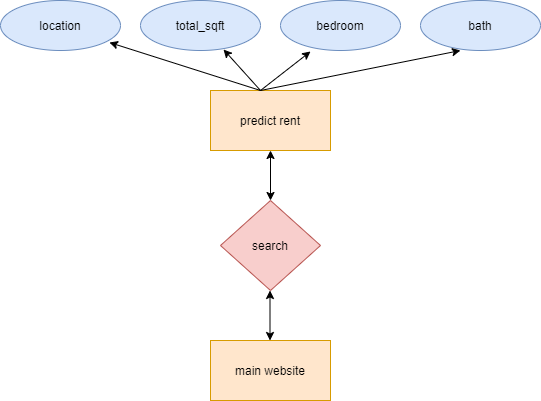
****

Figure 10: ER-diagram

**Chapter 4**

**Implementation and Result**

**4.1 Implementation**

Implementation is a basic expression of any quit project. To implement our project, we want a development model. We choosing the Linear Regression algorithm. Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. **Hypothesis function for Linear Regression: y = θ1 + θ2 \* x**

**Here,**

**x:** input  
**y:** labels  
**θ1:** intercept  
**θ2:** coefficient of x

**4.1.1 Code**

import pandas as pd

import numpy as np

from matplotlib import pyplot as plt

%matplotlib inline

import matplotlib

matplotlib.rcParams["figure.figsize"] = (20,10)

df1 = pd.read\_csv("Book2.csv")

df1.head()

df1.shape

df1.columns

df1['area\_type'].head(10)

df1['area\_type'].value\_counts()

df2 = df1.drop(['area\_type','society','balcony','availability'],axis='columns')

df2.shape

df2.isnull().sum()

df2.shape

df3 = df2.dropna()

df3.isnull().sum()

df3.shape

df3['bhk'] = df3['size'].apply(lambda x: int(x.split(' ')[0]))

df3.bhk.unique()

def is\_float(x):

try:

float(x)

except:

return False

return True

df3[~df3['total\_sqft'].apply(is\_float)].head(10)

def convert\_sqft\_to\_num(x):

tokens = x.split('-')

if len(tokens) == 2:

return (float(tokens[0])+float(tokens[1]))/2

try:

return float(x)

except:

return None

df4 = df3.copy()

df4.total\_sqft = df4.total\_sqft.apply(convert\_sqft\_to\_num)

df4 = df4[df4.total\_sqft.notnull()]

df4.head(2)

df4.loc[30]

df5 = df4.copy()

df5['price\_per\_sqft'] = df5['price']/df5['total\_sqft']

df5.head()

df5\_stats = df5['price\_per\_sqft'].describe()

df5\_stats

df5.to\_csv("bhp.csv",index=False)

df5.location = df5.location.apply(lambda x: x.strip())

location\_stats = df5['location'].value\_counts(ascending=False)

location\_stats

location\_stats.values.sum()

len(location\_stats[location\_stats>500])

len(location\_stats)

len(location\_stats[location\_stats<=500])

location\_stats\_less\_than\_500 = location\_stats[location\_stats<=500]

location\_stats\_less\_than\_500

len(df5.location.unique())

df5.location = df5.location.apply(lambda x: 'other' if x in location\_stats\_less\_than\_500 else x)

len(df5.location.unique())

df5.head(10)

df5[df5.total\_sqft/df5.bhk<300].head()

df5.shape

df6 = df5[~(df5.total\_sqft/df5.bhk<300)]

df6.shape

df6.price\_per\_sqft.describe()

def remove\_pps\_outliers(df):

df\_out = pd.DataFrame()

for key, subdf in df.groupby('location'):

m = np.mean(subdf.price\_per\_sqft)

st = np.std(subdf.price\_per\_sqft)

reduced\_df = subdf[(subdf.price\_per\_sqft>(m-st)) & (subdf.price\_per\_sqft<=(m+st))]

df\_out = pd.concat([df\_out,reduced\_df],ignore\_index=True)

return df\_out

df7 = remove\_pps\_outliers(df6)

df7.shape

def plot\_scatter\_chart(df,location):

bhk2 = df[(df.location==location) & (df.bhk==2)]

bhk3 = df[(df.location==location) & (df.bhk==3)]

matplotlib.rcParams['figure.figsize'] = (15,10)

plt.scatter(bhk2.total\_sqft,bhk2.price,color='blue',label='2 BHK', s=50)

plt.scatter(bhk3.total\_sqft,bhk3.price,marker='+', color='green',label='3 BHK', s=50)

plt.xlabel("Total Square Feet Area")

plt.ylabel("Price (Hazar Bangladesh Taka)")

plt.title(location)

plt.legend()

plot\_scatter\_chart(df7,"sec 2")

plot\_scatter\_chart(df7,"sec 10")

def remove\_bhk\_outliers(df):

exclude\_indices = np.array([])

for location, location\_df in df.groupby('location'):

bhk\_stats = {}

for bhk, bhk\_df in location\_df.groupby('bhk'):

bhk\_stats[bhk] = {

'mean': np.mean(bhk\_df.price\_per\_sqft),

'std': np.std(bhk\_df.price\_per\_sqft),

'count': bhk\_df.shape[0]

}

for bhk, bhk\_df in location\_df.groupby('bhk'):

stats = bhk\_stats.get(bhk-1)

if stats and stats['count']>5:

exclude\_indices = np.append(exclude\_indices, bhk\_df[bhk\_df.price\_per\_sqft<(stats['mean'])].index.values)

return df.drop(exclude\_indices,axis='index')

df8 = remove\_bhk\_outliers(df7)

# df8 = df7.copy()

df8.shape

plot\_scatter\_chart(df8,"sec 2")

plot\_scatter\_chart(df8,"sec 10")

import matplotlib

matplotlib.rcParams["figure.figsize"] = (20,10)

plt.hist(df8.price\_per\_sqft,rwidth=0.8)

plt.xlabel("Price Per Square Feet")

plt.ylabel("Count")

df8.bath.unique()

plt.hist(df8.bath,rwidth=0.8)

plt.xlabel("Number of bathrooms")

plt.ylabel("Count")

df8[df8.bath>10]

df8[df8.bath>df8.bhk+2]

df9 = df8[df8.bath<df8.bhk+2]

df9.shape

df9.head(2)

df10 = df9.drop(['size','price\_per\_sqft'],axis='columns')

df10.head(3)

dummies = pd.get\_dummies(df10.location)

dummies.head(3)

df11 = pd.concat([df10,dummies.drop('other',axis='columns')],axis='columns')

df11.head()

df12 = df11.drop('location',axis='columns')

df12.head(2)

df12.shape

X = df12.drop(['price'],axis='columns')

X.head(3)

X.shape

y = df12.price

y.head(3)

len(y)

from sklearn.model\_selection import train\_test\_split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X,y,test\_size=0.2,random\_state=10)

from sklearn.linear\_model import LinearRegression

lr\_clf = LinearRegression()

lr\_clf.fit(X\_train,y\_train)

lr\_clf.score(X\_test,y\_test)

from sklearn.model\_selection import ShuffleSplit

from sklearn.model\_selection import cross\_val\_score

cv = ShuffleSplit(n\_splits=5, test\_size=0.2, random\_state=0)

cross\_val\_score(LinearRegression(), X, y, cv=cv)

def predict\_price(location,sqft,bath,bhk):

loc\_index = np.where(X.columns==location)[0][0]

x = np.zeros(len(X.columns))

x[0] = sqft

x[1] = bath

x[2] = bhk

if loc\_index >= 0:

x[loc\_index] = 1

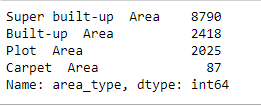
return lr\_clf.predict([x])[0]

predict\_price('sec 2',1000, 2, 2)

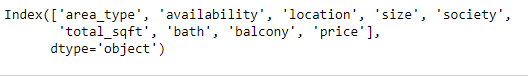
predict\_price('sec 10',1000, 3, 3)

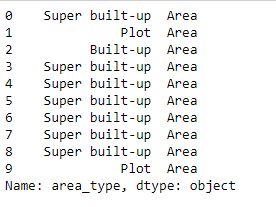
predict\_price('sec 2',850, 3, 2)

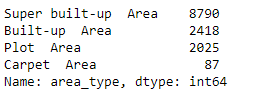
**4.1.2 Output**



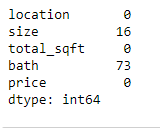


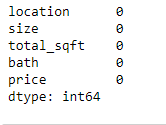








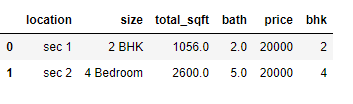


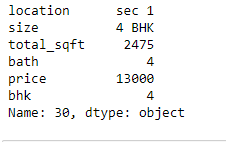




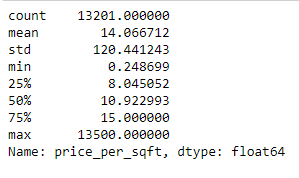


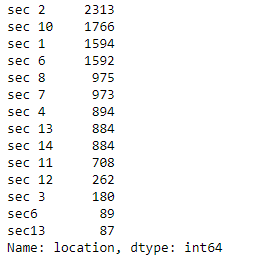










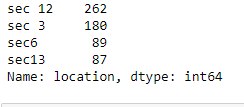














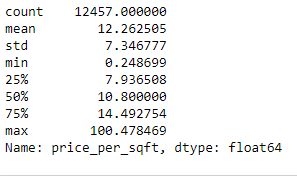




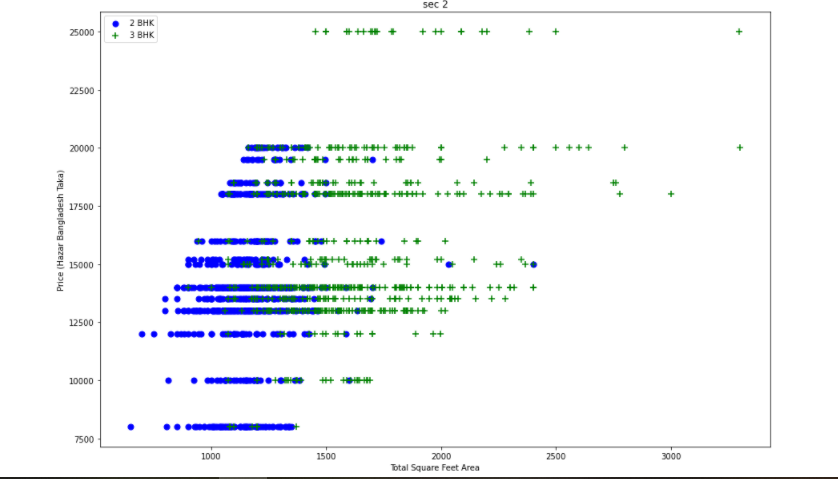


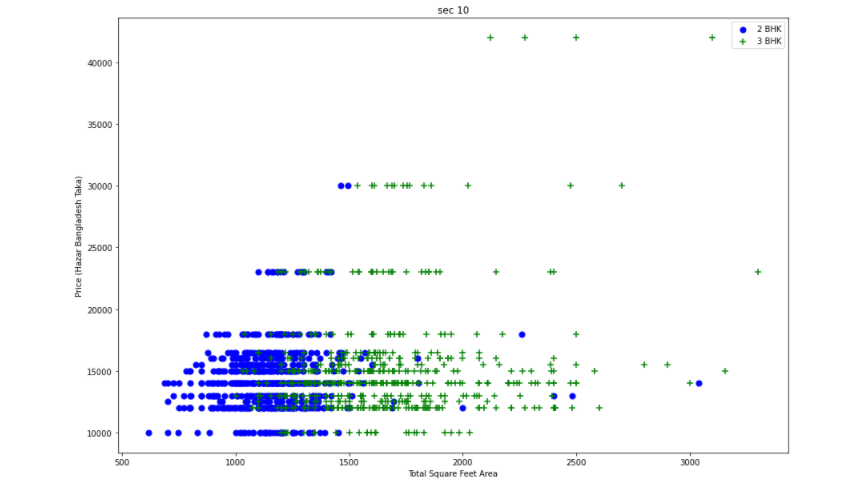




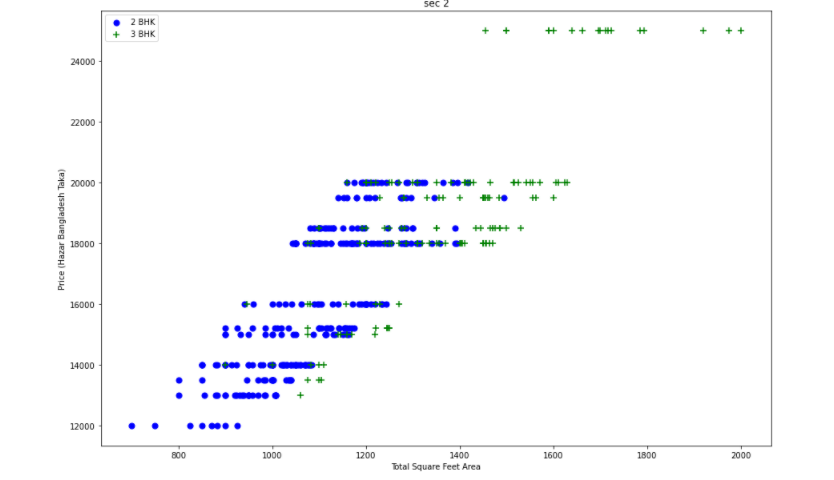


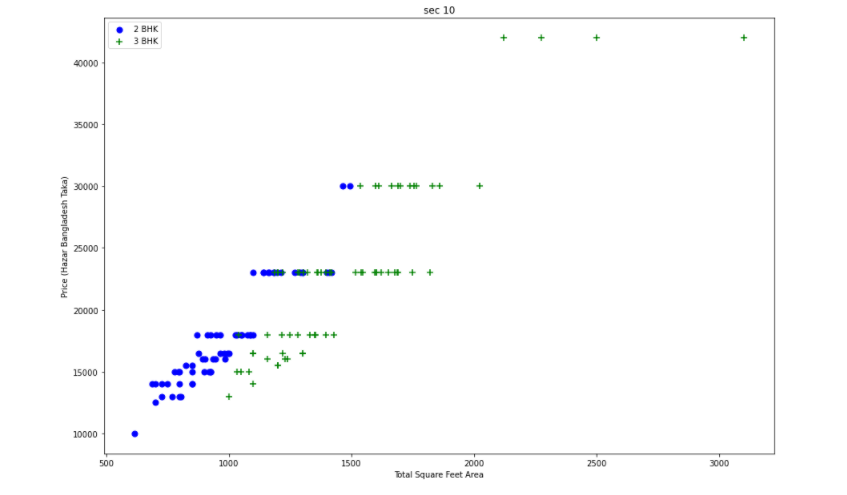


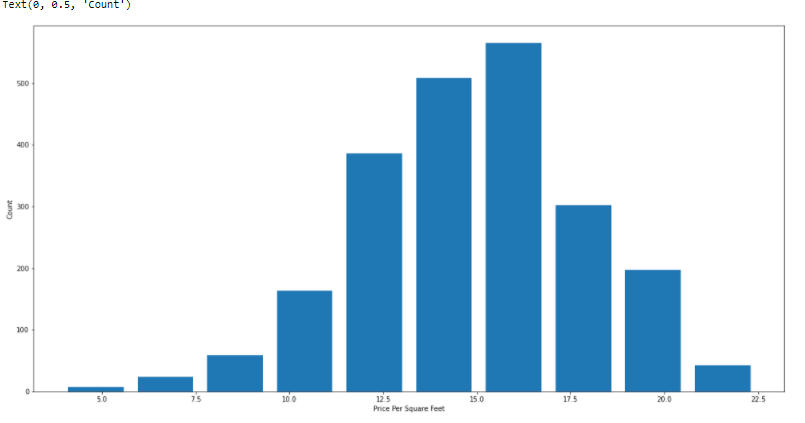




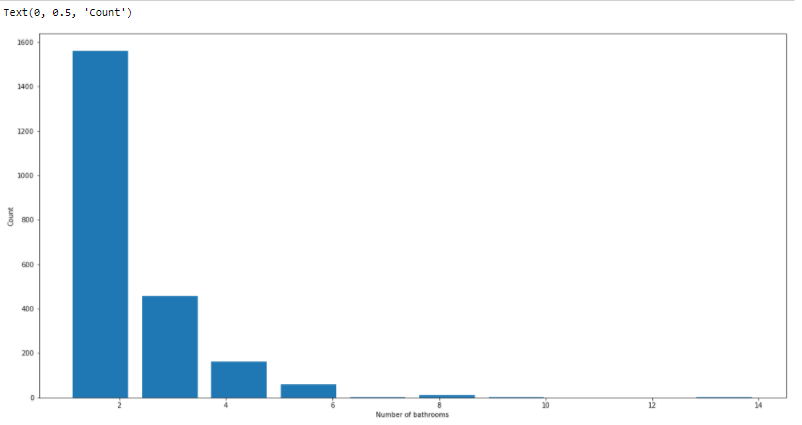


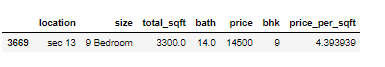


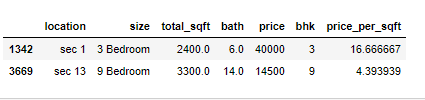






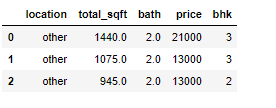


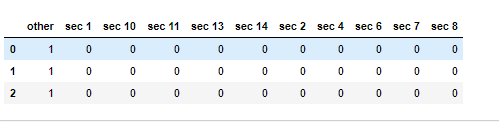


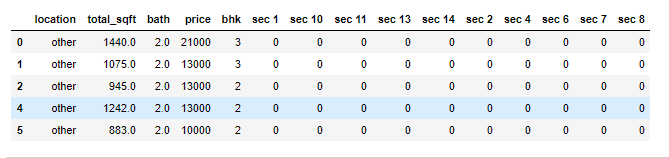


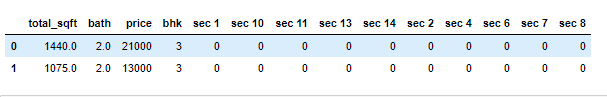




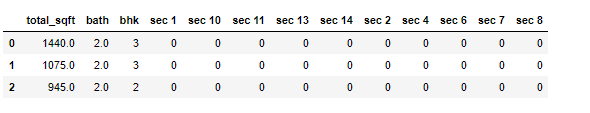














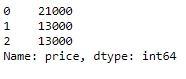














Figure 11: Ouput

**4.2 System Setup**

1. Package: Operating System (Any)
2. Software: Browser (Any), Python (new update version), Anaconda, Jupyter Notebook, Microsoft Excel
3. Hard Disk: 120 GB (minimum)
4. Ram: 2 GB (minimum)
5. Processor: Dual core or above

**4.3 Results and Discussion**

Home Rent Prediction Systemis an AI-based program that can predict the house price by analyzing the location, size of the room, and the total amount of bed-room. Here, we used a dataset that contains various information about houses including house prices. Our program predicts rent by analyzing the used dataset. We used Python programming language which is one of the best high level programming language for Artificial Intelligence.

The majority of the people in Dhaka have come to Dhaka for a variety of economic and social reasons. They are staying and working in Dhaka permanently without a home of their own. The experiences and the house ownership process of the tenants generate knowledge and present a guideline for the policymakers and planners. According to research, about 68% of people who lived in Dhaka are tenants. So, it's very difficult for them to search for available houses on their own without any guidelines.

To find a house which is available for rent is a very hard-working job. Our system can make the job a little easy. And tenants can save a lot of time to find their beautiful house.

**4.4 Existing System Analysis**

In the currently existing system, tenants search for houses on the website by searching the only location. So, it is a boring and frustrating job to do. They are wasting too much time with this process.

**4.5 Summary**

This system is providing high-quality prediction with maximum accuracy about House rent.

**Chapter 5**

**Standard, Ethics, Challenge**

**5.1 Standard**

Ensure an adequate, timely, and easily accessible dataset that gives us an accurate result. Ensure adequate system for the operation of a sustainable prediction system. Implement a top-quality management system within the service. Our vision is to possess a well-organized, coordinated, standardized, and quality insertion service that ensures adequate, safe, and timely shows the rent to the tenants. The tenants should understand the risks of view the predicted rent by the system.

**5.2 Ethics**

Home rent must be ethical. The how house owner and the tenant must follow the ethical rule. No fraud is acceptable in the home rent system. So always verify the information before renting a house. Tenants and the house owner can Develop straightforward contracts. House owner must be Reasonable with rent increases. Tenants should do proper maintenance of the house. Both tenants and house owners must respect privacy. A good tenant pays rent on time. But even the best tenant is going to occasionally make a mistake or have an issue where they miss a payment by a day or two. So, the house owner should show some leniency. The move-out process is often one of the biggest points of contention between landlords and tenants. Tenants want to get their security deposits back, while landlords typically look for any excuse to keep them. So, tenants shouldn't nitpick on move-out.

**5.3 Challenges**

The main challenge of building our project is to make a straightforward system for the tenants. To make our system we face many problems. But, due to hard teamwork and proper communication with our teammates, we come out through problems. To build a perfect dataset we are doing many things. Like, data load, handle NA values, Feature Engineering, Dimensionality Reduction, Outlier Removal Using Business Logic, Outlier Removal Using Standard Deviation and Mean, Outlier Removal Using Bathrooms Feature, One Hot Encoding for Location, build a Model, Test the model for few properties, etc. We use K Fold cross-validation to measure the accuracy of our Linear-Regression model.

Thus, we build our system and completed all the challenges.

**5.4 Summary**

To build a straightforward page with standards, ethics, and challenge it's not easy for us to end our project. But we tried our greatest to assist those tenants who seek houses each day. We believed that our system will help those tenants.

**Chapter 6**

**Conclusion**

**6.1 Introduction**

This system is easy to use for all. Home Rent Prediction Systemis an AI-based program that can predict the house price by analyzing the location, size of the room, and the total amount of bed-room. Here, we used a dataset that contains various information about houses including house prices. Our program predicts rent by analyzing the used dataset. We used Python programming language which is one of the best high level programming language for Artificial Intelligence.

The majority of the people in Dhaka have come to Dhaka for a variety of economic and social reasons. They are staying and working in Dhaka permanently without a home of their own. The experiences and the house ownership process of the tenants generate knowledge and present a guideline for the policymakers and planners. According to research, about 68% of people who lived in Dhaka are tenants. So, it's very difficult for them to search for available houses on their own without any guidelines.

The scope of the project is that in a very short span provides the user with many facilities. It provides a very accurate prediction with an accuracy of more than 60%. So, users can trust it. It shows users the predicted rent when they give the location, total square feet of house, total bedroom, and total bath as input.

To find a house which is available for rent is a very hard-working job. Our system can make the job a little easy. And tenants can save a lot of time to find their beautiful house.

**6.2 Future Work**

The Main future scope of our proposed model about the Home Rent Prediction system is to try it in the real world means we have to implement it with better results. We can improve this system with more accurate results. This system can be used in many websites by special features. It helps the websites to perform better and help the user to enjoy the sites.

**Reference**

[1]. Codebasics. “Machine Learning & Data Science Project” Jan-1, 2020. Available: <https://www.youtube.com/watch?v=rdfbcdP75KI&list=PLeo1K3hjS3uu7clOTtwsp94PcHbzqpAdg&index=1>  
[2]. Bproperty. Available: <https://www.bproperty.com/>

[3]. Rentalhomebd. Available: <https://rentalhomebd.com/>

[4]. Samsuddin Ahmed, Md. Mahbubur Rahman, Sabirah Islam, “House Rent Estimation in Dhaka City by Multi-Layer Perceptions Neural Network” Aug-2014 DOI: [10.14257/ijunesst.2014.7.4.26](https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.14257%2Fijunesst.2014.7.4.26?_sg%5B0%5D=A_rrow_eynN4BECsuK9kfiaTJhsBn1VxHZi7cM9Ge17bo1UFIEhbOSQbwyLiJLaq0x8cYqH9l_YiZZiQCmdr8Q65Kw.XyoK9x-1soeR-gYQx0m1_Hoq8SYtEVlRIUhtpDQOXcveDzOjFaBH5npprLPNZrPGe-t8tl32n4CyITnO1nVlJw) .

Available: <https://www.researchgate.net/publication/291197777_House_Rent_Estimation_in_Dhaka_City_by_Multi_Layer_Perceptions_Neural_Network>

[5]. Python Programmer. “What is Pandas? Why and How to Use Pandas in Python” May-24, 2018.

Available: <https://www.youtube.com/watch?v=dcqPhpY7tWk>

[6]. Python Programmer. “Watch this if you want to LEARN MATPLOTLIB for PYTHON!” Jun-28, 2018.

Available: <https://www.youtube.com/watch?v=6rKe2IEIu8c>

[7]. Corey Schafer. “Jupyter Notebook Tutorial: Introduction, Setup, and Walkthrough” Sep-22, 2016.

Available: <https://www.youtube.com/watch?v=HW29067qVWk>

[8]. Kaggle. “Bengaluru House Price Data” 2018.

Available: <https://www.kaggle.com/amitabhajoy/bengaluru-house-price-data>

**Figures**

|  |  |
| --- | --- |
| Figure 1 | Mirpur house data |
| Figure 2 | Area\_type |
| Figure 3 | location |
| Figure 4 | Size |
| Figure 5 | Total\_sqft |
| Figure 6 | Price |
| Figure 7 | Schedule |
| Figure 8 | DFD |
| Figure 9 | Use case diagram |
| Figure 10 | E-R diagram |
| Figure 11 | Output |