

House Rent Prediction in Dhaka City Using Machine Learning Approach

BY

Md Arman Hossen
ID: 191-15-12726

Md. Merazur Rahman
Id: 191-15-12723

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Dewan Mamun Raza
Senior Lecturer
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Md. Firoz Hasan
Lecturer
Department of Computer Science and Engineering
Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH

APPROVAL

This Project/internship titled “**House Rent Prediction in Dhaka City Using Machine Learning Approach**”, submitted by **Md Arman Hossen** and **Md. Merazur Rahman** ID No: 191-15-12726 and 191-15-12723 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 24-01-2023.

BOARD OF EXAMINERS

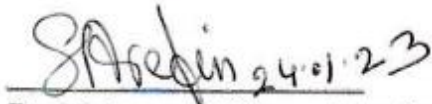


Dr. Touhid Bhuiyan

Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman

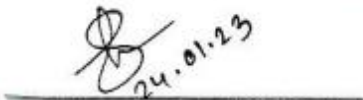


Dr. Mohammad Shamsul Arefin

Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

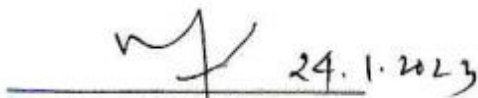


Md. Sabab Zulfiker

Senior Lecturer

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Ahmed Wasif Reza

Associate Professor

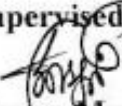
Department of Computer Science and Engineering
East West University

External Examiner

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Dewan Mamun Raza, Senior lecturer, Department of CSE, Daffodil International University.** We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

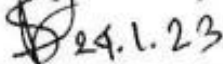


Dewan Mamun Raza

Senior lecturer

Department of Computer Science and Engineering
Daffodil International University

CO-Supervised by:

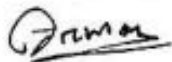


Md. Firoz Hasan

lecturer

Department of Computer Science and Engineering
Daffodil International University

Submitted by:



Md Arman Hossen

ID: 191-15-12726

Department of CSE

Daffodil International University



Md. Merazur Rahman

ID: 191-15-12723

Department of CSE

Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to Almighty Allah for His divine blessing which makes us possible to complete the final year project/internship successfully.

We are really grateful and wish our profound indebtedness to **Dewan Mamun Raza, Senior Lecturer**, Department of CSE, Daffodil International University, Dhaka, deep knowledge & keen interest of our supervisor in the field of Machine Learning to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Touhid Bhuiyan**, Head, Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of the CSE department of Daffodil International University.

Finally, must acknowledge with due respect the constant support and patients of my parents.

ABSTRACT

Dhaka is a densely populated megacity in Bangladesh and day by day population of Dhaka city is increasing. People are frequently migrating to Dhaka. The main driving forces behind migration to Dhaka City are poverty, job searching, and family pressure. As a result, Dhaka is currently overpopulated. So, with all available houses, it is tough to accommodate this overpopulated Dhaka city. For this reason, rental house is one of the most severe issues in this city. House rental prices are having an effect by various factors. So, it is very important to determine the house rent price. The main goal of our work is to analyze the different features of a house and predict the rental price of a house based on multiple factors using machine learning algorithms. So, this report's effort is to build a model that can predict house rent in Dhaka city. The area of Dhaka city namely Bashundhara, Gulshan, Mohammadpur, Uttara, Farmgate, Dhanmondi, Baridhara, Mirpur, Nikunja, Khilgaon, etc. Multiple factors including geographical location, house size, number of bedrooms, and number of bathrooms are considered. This work uses various machine-learning regression techniques to predict house rent and compare the accuracy of each algorithm. The following selected algorithms are - Linear Regression, Ridge Regression, Bayesian Regression, and Lasso Regression. All the proposed model gives almost similar accuracy. And our proposed model Ridge regression gives the highest accuracy with an accuracy of 91.54% on the other hand, Linear Regression, Bayesian Regression, and Lasso Regression give an accuracy of 91.49%, 91.52%, and 91.49%, respectively

TABLE OF CONTENTS

CONTENT	PAGE-NO
CHAPTER 01: INTRODUCTION	01-06
1.1 Introduction	01
1.2 Problem Statement	04
1.3 Motivation	05
1.4 Objective	05
1.5 Report Layout	05
CHAPTER 02: BACKGROUND STUDY	07-09
2.1 Related Work	07
2.2 Scope of the Problem	08
2.3 Challenges	09
CHAPTER 03: RESEARCH METHODOLOGY	10-33
3.1 Data collection	11
3.2 Data analysis	12
3.2.1 Dataset features	12
3.2.2 Dataset label and its properties	12
3.3 Features type, explain feature value in dataset	13
3.4 Dataset features correlation with the target column	14
3.5 Dataset insight (Graphs with explanation & details)	15
3.6 Feature Engineering	29
3.7 Data Preprocessing	29
3.8 Dataset split	31
3.9 Model	31
3.8.1 Linear Regression	31

3.8.2 Ridge Regression	31
3.8.3 Bayesian Regression	32
3.8.4 Lasso Regression	33
CHAPTER 04: EXPERIMENTAL RESULT AND DISCUSSION	34-37
4.1 Result	34
4.1.1 Accuracy of model	34
4.1.2 Compare between predicted rent and original rent	35
4.1.3 The Most number of houses	36
CHAPTER 05: CONCLUSION AND FUTURE WORK	38
5.1 Conclusion	38
5.2 Future Work	38
REFERENCES	39-41

LIST OF TABLES

TABLES	PAGE NO
Table 3.1 Features of Attribute	12
Table 3.2: A Portion of the dataset	13
Table 3.3: With Unnamed Column	29
Table 3.4: Without Unnamed Column	30
Table 3.5: Outlier of the dataset	30
Table 4.1. Predictors Score according to Model	34
Table 4.2. Original Rent vs Predicted Rent	35
Table 4.3: Locations with the most numbers of houses	36

LIST OF FIGURES

FIGURES	PAGE NO
Figure 1.1: Dhaka North City Corporation	3
Figure 1.2: Dhaka South City Corporation	4
Figure 2.1: A portion Dummy variable	8
Figure 3.1 Flow Diagram	10
Figure 3.2: A snap of Bproperty.com	11
Figure 3.3: Snap of Heatmap	14
Figure 3.4: Uttara house rent (depend on bedrooms)	18
Figure 3.5: House Rent Visualization (depends on bedrooms)	23
Figure 3.6: House Rent Visualization (depends on bathrooms)	27
Figure 3.7: Histogram for each attribute	28

CHAPTER 1

INTRODUCTION

1.1 Introduction

Increasing migration to cities and high population growth has resulted in a rapid decline of urban land, causing house prices to exceed the reach of middle- and low-income households in developing countries. The city of Dhaka is the nucleus of Bangladesh and is known as one of the world's most densely populated megacities. It's no secret that Dhaka is a megacity, and for this city, rent is a major problem. A population of over 22 million in Dhaka, the fourth largest city in the world, poses a big challenge [1]. Nowadays it is very tough to live in Dhaka because the house rent is increasing day by day continuously. On the other hand, the population also increasing as well. According to the Population and Housing Census 2011, over 12 million people lived in the Dhaka division. But the Population and Housing Census 2022 reported that nowadays 44 million people live in the Dhaka division [2]. So, the population growth rate at 1.74% in the division now stands. Between 1991 and 2001, population growth was 4.15 percent and 7 percent, respectively [3]. One of the main reasons for Population growth in Dhaka City has mainly been caused by immigration from rural areas. Consequently, house rent has been rising dramatically as a result of the increase in people's need for accommodation. And it is estimated that 65 percent or more of their monthly income goes toward paying rent [4]. The house owner fixes the rent for the house as per their wishes, even though it is seen that sometimes house rent increases yearly by the house owner three or four times. To stop this aggressive behavior of house owners, house rents should be fixed by a standard. So, we present here the most recent research on regression techniques that can be used to predict house rent in Dhaka city. These techniques include Linear Regression, Ridge Regression, Bayesian Regression, and Lasso Regression.

In order to get an accurate house price prediction, it was necessary to use some effective methods. In this paper, we have used a dataset from bProperty.com for the rental prices and various features of a house in Dhaka, Bangladesh. The quality of data plays a crucial role in the prediction of house rent, and missing information poses a challenge to machine learning algorithms.

In recent years, the cost of housing in Dhaka has been inexplicably rising. The majority of middle-class families in Dhaka live in leased homes. They are finding it difficult to deal with the unchecked increase in rent prices. Only roughly 20–25% of Dhaka's overall population own their home. The remaining 75–80% of people rent their homes. Due to the rising demand for rental homes in the city and the resulting increase in population, landlords are raising rents.

In rare instances, landlords would also put pressure on renters to vacate the property if the latter is unable to comply with the request. The government must establish a criterion for determining a house's rent. It lessens the unfairness toward the tenants. Here, I suggest a neural network design that can be utilized to calculate estimated rent. The house rent can be fixed upon the aforementioned thirteen features by enforcing this method of estimation, making it impossible for landlords to voluntarily determine the house rent. There are therefore certain drawbacks to this method.

The Local Government (City Corporation) Amendment Bill 2011, which was approved by the Bangladeshi Parliament, was used by the Awami League government to dissolve the Dhaka City Corporation on November 29, 2011. following its introduction to the Parliament on November 23. The city corporation will be divided into two corporations, North and South, with the southern wing controlling a larger portion of the city. Due to the fact that each corporation would be autonomous, Dhaka will have two mayors. According to the government, bifurcation would guarantee that the citizens of the city would receive better quality municipal services.[16]

A total of 54 wards make up the Dhaka North City Corporation, which includes the thanas of Mirpur, Mohammadpur, Sher-e-Bangla Nagar, Pallabi, Adabor, Kafrul, Dhaka Cantonment, Tejgaon, Gulshan, Rampura, Banani, Airport, Khilkhet, Vatara, Badda, Uttara, and others. Atiqul Islam serves as the current mayor of Dhaka North City Corporation.[16][17]

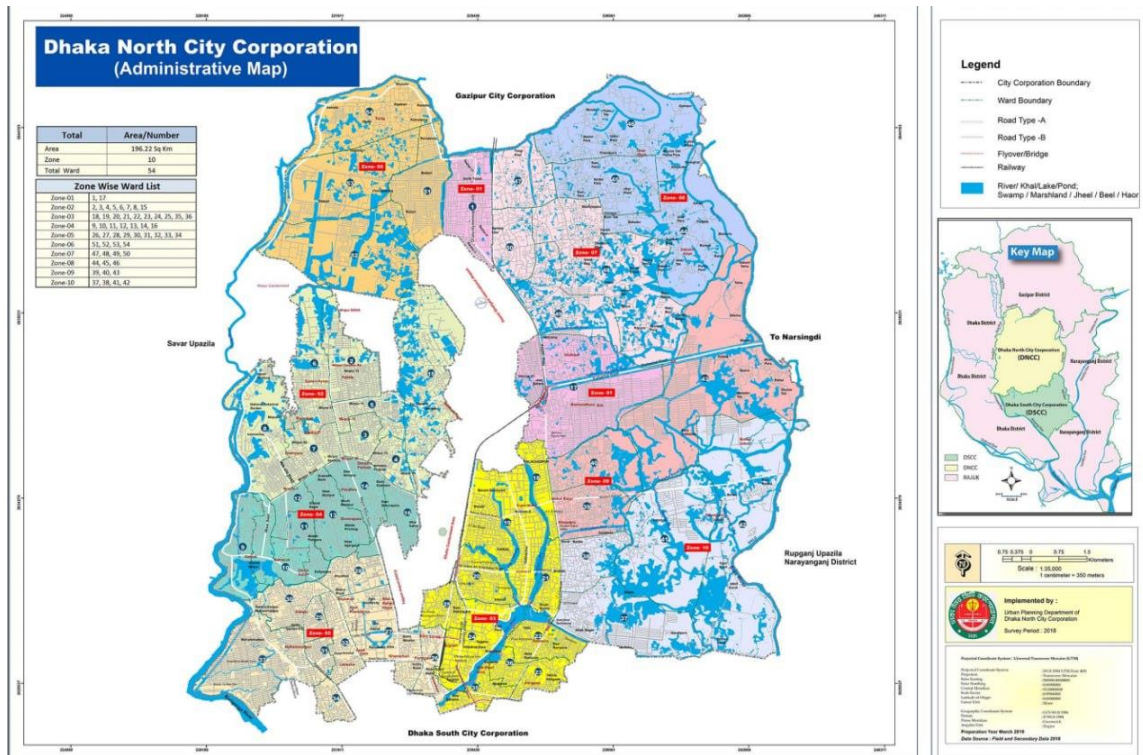


Figure 1.1: Dhaka North City Corporation

The 75 wards that makeup Dhaka South City Corporation include those in Paltan, Motijheel, Jatrabari, Kotwali, Sutrapur, Bangsal, Wari, Ramna, Gendaria, Chowkbazar, Lalbagh, Hazaribagh, Dhanmondi, Shahbagh, New Market, Khilgaon, Kamrangirchar, and other thanas. Sheikh Fazle Noor Taposh is the current mayor of Dhaka South City Corporation.[16][17]

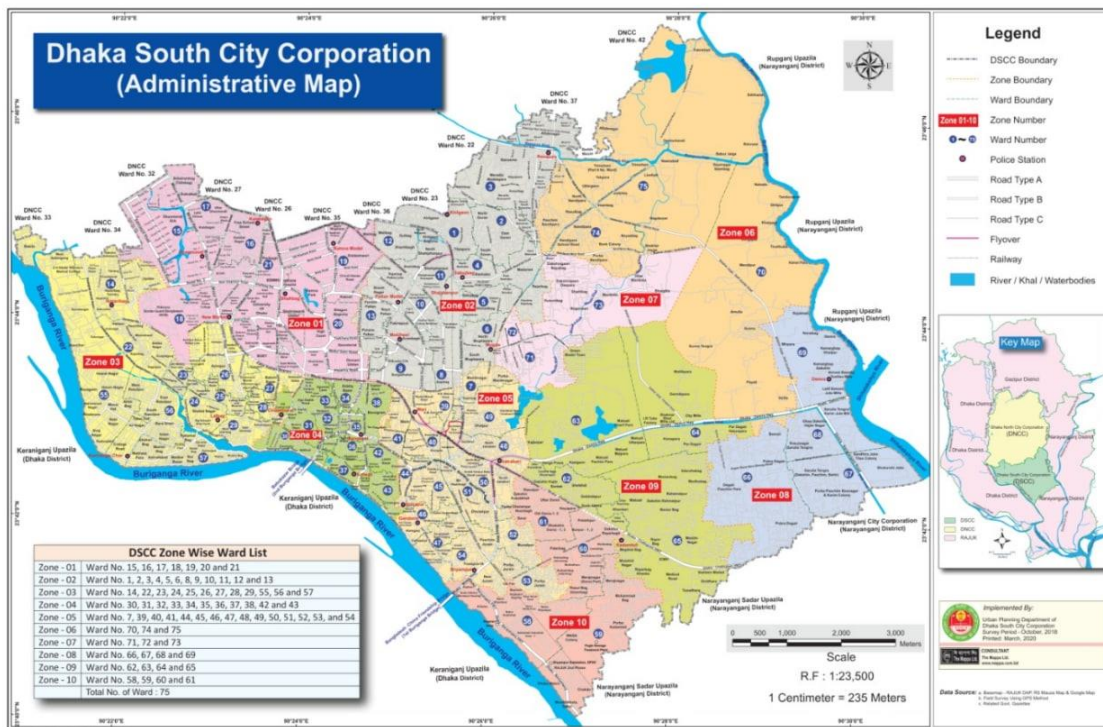


Figure 1.2: Dhaka South City Corporation

1.2 Problem Statement

Dhaka is the capital city of Bangladesh. It is a mega city. From other parts of the country, people are migrating to Dhaka city because Poverty, job searching, and family influence are the main push factors for migration to Dhaka, while better opportunities, prior migrants, and availability of jobs are the main pull factors behind migration. To lead a better life people to come to Dhaka. But Dhaka is now overpopulated. As a result, the house rent increasing day by day. The house owner takes full advantage. They increase their house rent unusually. As a result, middle-income people are suffering a lot. So, our proposed model can predict the actual house rent and people get acknowledged about the house rent.

1.3 Motivation

- House rent is a severe issue in Dhaka city.
- To develop a model that can predict the actual house rent.
- Promote the model so that people can search for their desired house for rent.
- Make people aware of house rent so that the house owner cannot take the advantage of overrent.
- Promote the different house rent based on different locations. As a result, people can find a specific house.

1.6 Objective

- The Main objective is to predict the house rent in Dhaka city.
- Remove the unnamed columns from the dataset.
- Remove the outlier from the dataset.
- Performing regression models like Linear regression, Ridge regression, Lasso regression, and Bayesian regression.
- Compare model accuracy with each other.
- Compare the predicted data with real data.
- Show house rent for a specific location with the highest number of houses.

1.3 Report Layout

This report is divided into five chapter

Chapter 1 gives an introduction to the housing rent situation in Dhaka city along with its problem statement, motivation, and objective.

Chapter 2 highlights a detailed background study with a scope of the problem and challenges.

Chapter 3 Describes the Data pre-processing and research methodology in a detailed description that how to build all the models and predict the house rent.

Chapter 4 Explains the experimental result and discussion where discusses the accuracy of the model, and compared the predicted house rent with actual house rent.

Chapter 5 Discussion about the conclusion and what can do in the future by this work.

CHAPTER 2

BACKGROUND STUDY

2.1 Related works

There are a lot of factors responsible for high house rent in Dhaka city. The author of [3] has study to find the potential factors that could be affected house rent in Dhaka city. There are a few common factors they mentioned which are zonal variations of building, surrounding land, distance from the road, open space, utility facilities, structure, and the number of flats considered. They also mentioned that the owner of the house increased the house rent for their own benefit. According to security and safety, educational institutions and social status also play an important role to increase house rent. The author [6] has mentioned another key factor that affects house rent in Dhaka city. That is the increasing population day by day in Dhaka city. According to the author [6] by 2030 the rate of urbanization will be 40 percent of the total population. Day by day There is tremendous pressure on people in Dhaka city. The owner of the house increased the rent for the apartment. In Bangladesh, medium-income people can start the buying process after 40 because at that time they are either close to retirement from their job or hold a good position in the job followed by comfortable access to the financing service. To meet the demand over the next 20 years, Bangladesh will need to build over 4 million new homes per year. Since it is difficult for the government to provide houses for those with low and middle incomes as promised in "Vision 2021," Tk. 1461.74 cores have been allocated by the government of Bangladesh for the land management and housing sectors for FY 2012–2013[17]. [21] Gathers 33224 pieces of data on Chengdu housing rental for visual analysis and prediction based on the online housing platform. First, the characteristic variables such as area, housing area, orientation, rent collection technique, transportation, structure, etc. are chosen based on the significance of the characteristic features of the rental data. The data visualization technique reveals that combined rental and small-area apartment types are more in demand among Chengdu tenants. Globally, rising property values in major cities have made it more difficult for low-income groups to afford to housing[22]. Migrants, young people, and newlywed households all require an affordable place to live, which rental housing can offer. As a result, renting has evolved into a necessary addition to home ownership.

Around 1.2 billion people reside in rental housing globally (Gilbert, 2016). More particularly, the statistics show that the rental rate has topped 50% in major international cities like New York, Los Angeles, Shenzhen, and Shanghai [22].

2.2 Scope of the Problem

The goal of the research is to develop a model and apply a machine-learning approach to predict house rent in Dhaka city. In this research, we have used a few regression approaches like Linear regression, Ridge regression, Lasso regression, and Bayesian regression.

- The unidentified column will be immediately deleted by Pandas. To set the index for the data frame, use the `df.set index()` method on any column.
- A string that threatens models is a category variable. Categorical variables need to have dummy variables established. These are numerical constants that will improve the performance of the models. Put categorical variables to use.

	Location	Area	Bed	Bath	Price	price_per_sqft	1st Colony Mirpur Dhaka	2nd Colony Mirpur Dhaka	3rd Colony Mirpur Dhaka	Aainusbag Dakshin Khan Dhaka	...	West Kalachandpur Dhaka	West Kazipara Mirpur Dhaka	Mollartek Dakshin Khan Dhaka	West Monipur Mirpur Dhaka	West Nakhalpara Tejgaon Dhaka
0	1st Colony Mirpur Dhaka	800	2	2	13500.0	16.875000	1	0	0	0	...	0	0	0	0	0
1	1st Colony Mirpur Dhaka	700	2	2	12000.0	17.142857	1	0	0	0	...	0	0	0	0	0
2	1st Colony Mirpur Dhaka	1100	3	3	20000.0	18.181818	1	0	0	0	...	0	0	0	0	0
3	1st Colony Mirpur Dhaka	1100	3	3	20000.0	18.181818	1	0	0	0	...	0	0	0	0	0

Figure 2.1: A portion Dummy variable

- The scale of our graph was unrealistic due to the outlier values. There was a problem since the number of outlier values varied. Sometimes we would receive completely accurate readings, but other times we would receive false readings that could account for as much as 10% of the data. We did away with outliers because of this.
- Pre-process the data for using the further algorithm.

- We have used Regression approaches such as Linear regression, Ridge regression, Lasso regression, and Bayesian regression.
- We have compared the original price of a house with the predicted price.

2.3 Challenges

Related works: The main challenge is there has little related research only about the house rent. So, collecting information about house rent was harder.

Data Source: Data collecting is a challenge for us. Because a few real estate companies provide generated data. The same dataset cannot be used every time. Because the rent of a house can be changed based on the situation after a few years

Outlier: It is difficult to create a reliable model through ablation research that can provide acceptable performance in noisy mammography, and it takes a lot of effort for different experimental setups.

CHAPTER 3

RESEARCH METHODOLOGY

To get a better outcome, data preprocessing is a must need. And then splitting the data into train and test sets. So that builds any model and then finds the predicted result. The model will be trained by training data and then predict the result.

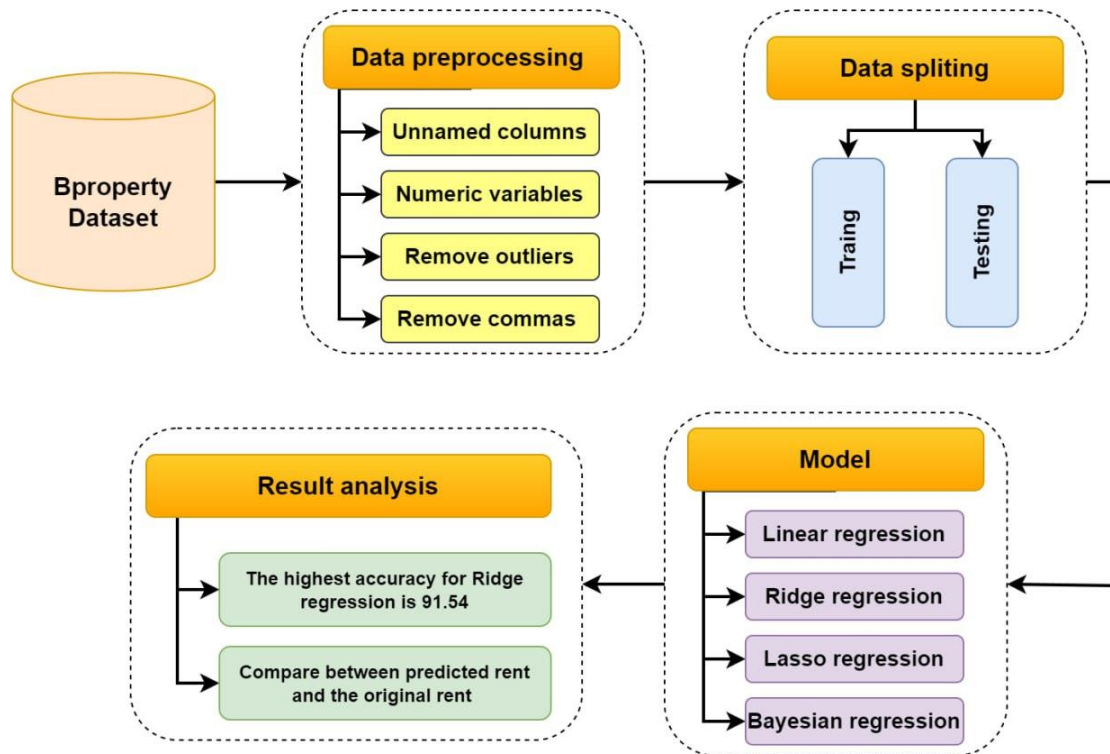


Figure 3.1 Flow Diagram

The above figure 1 is the flow diagram for the main work. The collected data is preprocessed including unnamed columns, numeric variables, remove outliers, and also remove commas. The whole data is split into a train set and a test set and builds a regression model. The regression

models are linear regression, ridge regression, lasso regression, and Bayesian regression. Each model predicts the result and then the result is analyzed. The ridge regression model gave the highest result with an accuracy of 91.54%.

3.1 Data collection

We collected the data set from bproperty.com [17]. With the goal of creating an online and offline marketplace where sellers, buyers, landlords, tenants, developers, and real estate investors can engage in concluding any real estate transaction securely and effectively, Bproperty is an online property portal that houses the largest collection of property listings in Bangladesh. Any person or business, regardless of their interest in real estate, can be guided by Bproperty, a reputable real estate specialist with an extensive understanding of Bangladesh's real estate market, to help them make the best choice to meet their real estate needs [18].

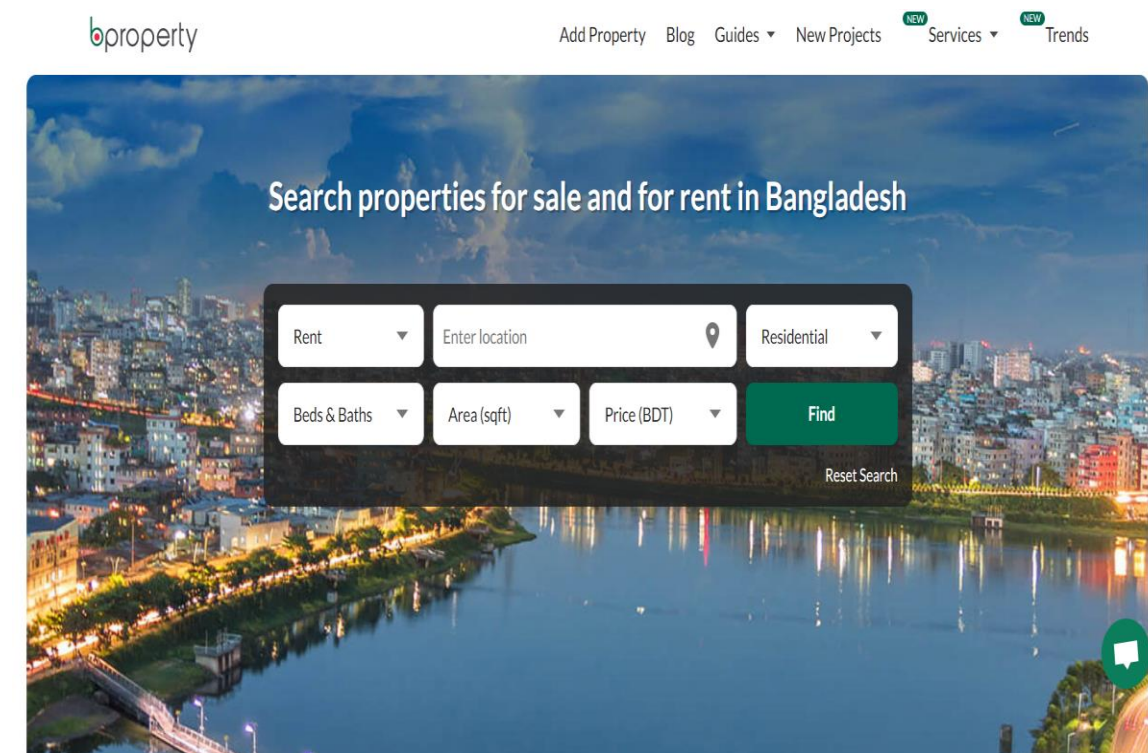


Figure 3.2: A snap of the Bproperty.com website

3.2 Data analysis

Data analysis is needed to make data more consistent. Presenting accurate and trustworthy data is the goal of data analysis in research. Avoid statistical errors as much as you can, and figure out how to handle common problems like outliers, missing data, changing data, data mining, or creating graphical representations.

3.2.1 Dataset features

In our dataset total of 5 attributes are available such as location, area, bed, bath, and price. And there are also 28000 instances or records available for the rental house in Dhaka city.

Table 3.1: Features Attribute

Features name	Features Description
Location	The location defines the position of that house. Is this house located in a residential area or non-residential?
Area	The area defines the total size of the house. It represents the size of the apartment in sq ft.
Bed	It represents the total number of bedrooms in that apartment.
Bath	It represents the total number of bathrooms in that apartment

3.2.2 Dataset label and its properties

The attribute 'Price' level is class. The attribute price depends on an independent attribute like location, area, bed, and bath. The price will predict different values. The location can be residential or non-residential as well. And bed and bath are the numerical value. So, there also could be a different value for the number of beds and a number of baths.

Table 3.2: A Portion of the dataset

Unnamed: 0		Location	Area	Bed	Bath	Price
0	0	Block H, Bashundhara R-A, Dhaka	1,600 sqft	3	3	20 Thousand
1	1	Farmgate, Tejgaon, Dhaka	900 sqft	2	2	20 Thousand
2	2	Block B, Nobodoy Housing Society, Mohammadpur,...	1,250 sqft	3	3	18 Thousand
3	3	Gulshan 1, Gulshan, Dhaka	2,200 sqft	3	4	75 Thousand
4	4	Baridhara, Dhaka	2,200 sqft	3	3	75 Thousand
5	5	Bashundhara R-A, Dhaka	3,000 sqft	4	5	50 Thousand
6	6	Baridhara, Dhaka	2,300 sqft	3	3	75 Thousand
7	7	PC Culture Housing, Mohammadpur, Dhaka	950 sqft	2	2	14 Thousand
8	8	Jigatola, Hazaribag, Dhaka	1,600 sqft	3	3	28 Thousand
9	9	West Kazipara, Mirpur, Dhaka	1,150 sqft	3	3	19 Thousand

3.3 Features type, explain feature value in the dataset

In the dataset, there are four independent attributes and one dependent attribute. Independent attributes are location, area, bed, and bath. On the other hand, the dependent attribute is price.

3.3.1 Location

The location represents the actual place for that house.

3.3.2 Area

It defines the actual size of the apartment in sq ft.

3.3.3 Bed

The bed actually defines, in that apart from how many bedrooms are available. Bedrooms are one of the important attributes of house rent.

3.3.4 Bath

Bathrooms are also other attributes that can change the rent of a house. It defines how many bathrooms are available in that apartment.

3.3.5 Price

The price attribute depends on all the attributes. The rent of a house can vary on location, area, bed, bath, etc. When the number of bedrooms and bathrooms increases then the rent will increase as well. On the other, if the location is a residential place, then the rent of the house will be higher.

3.4 Dataset features correlation with the target column

All the skewness and out-layers values for the feature "Price" need to be fixed before training the dataset. The models' error values will be reduced by removing unnecessary layers. Additionally, this will aid in collaborating with feature engineering to divide the price more evenly among all the features. Let's look at the skewed data and correct them using the log. Transformation in Figure.

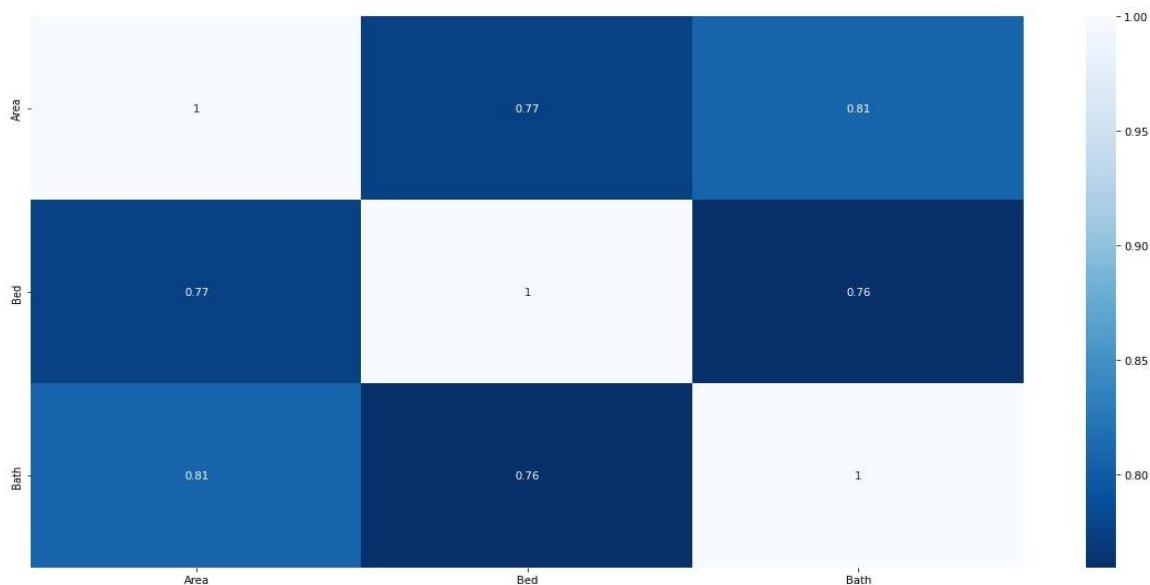


Figure 3.3: Snap of Heatmap

3.5 Dataset insight

Below you can see each feature and its outcome of the Scatterplot project. Based on the bedrooms we show the actual house rent for Uttara all sector.

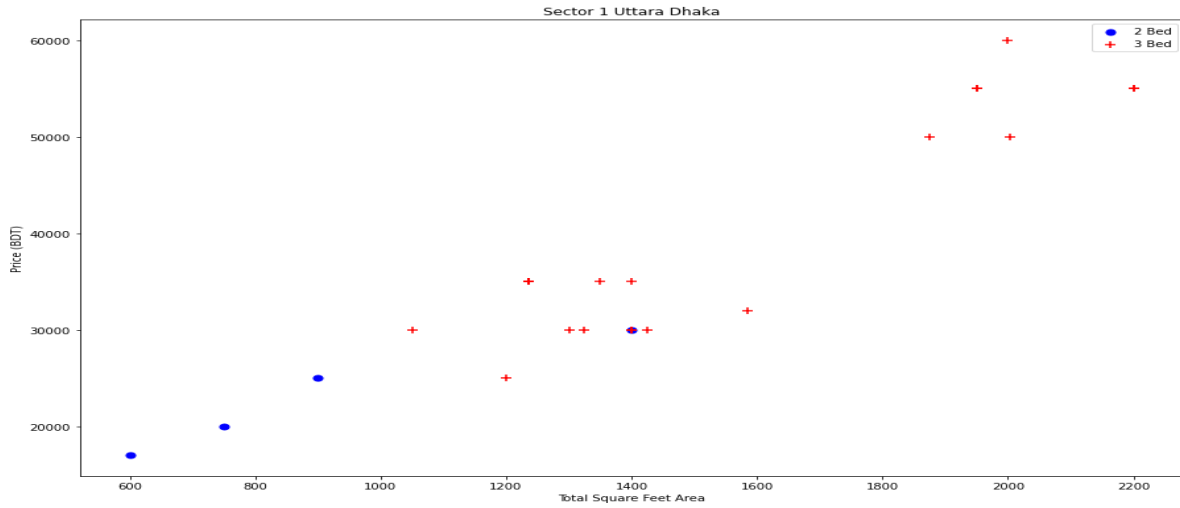


Figure 3.4.1: Uttara sector 1

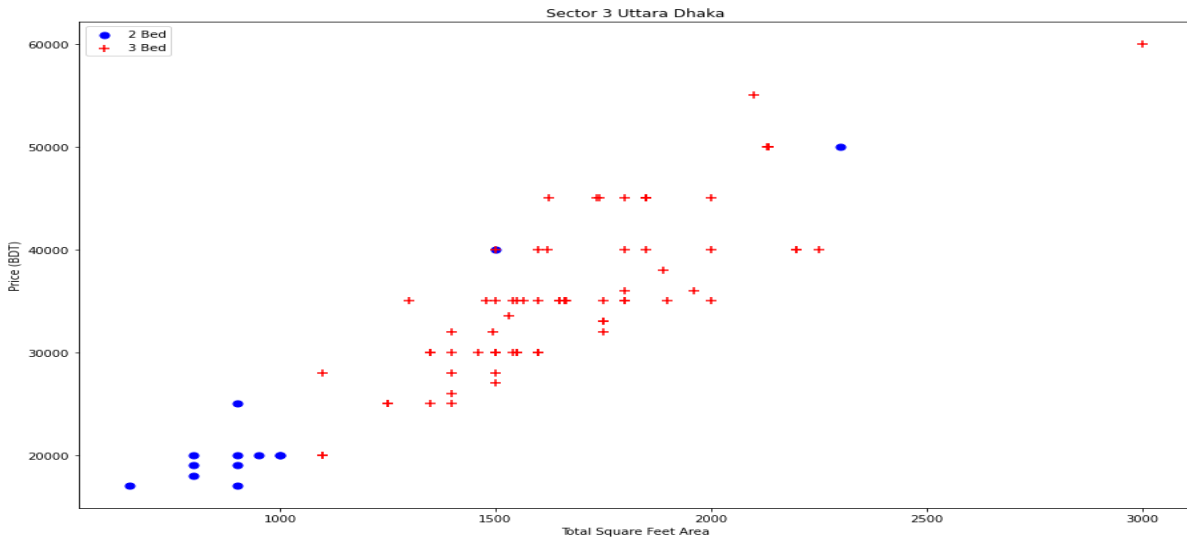


Figure 3.4.2: Uttara sector 3

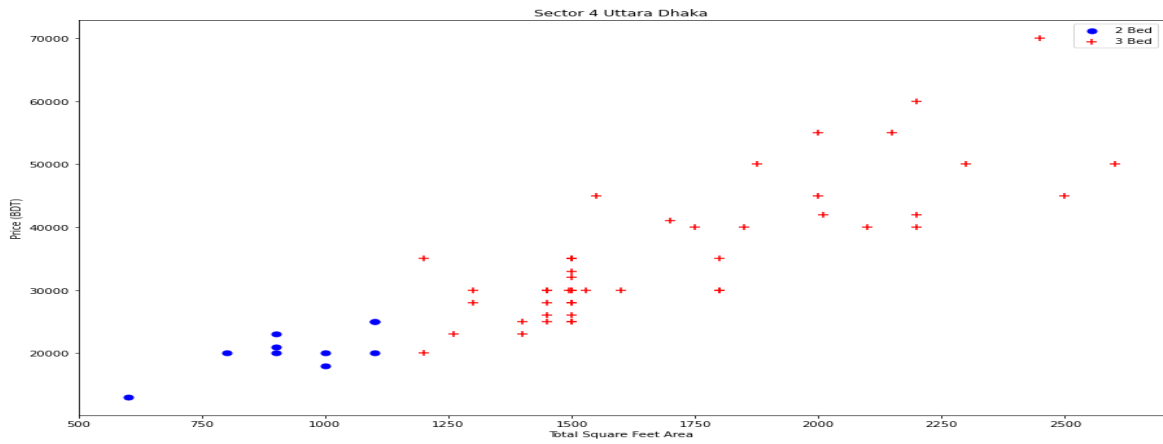


Figure 3.4.3: Uttara sector 4

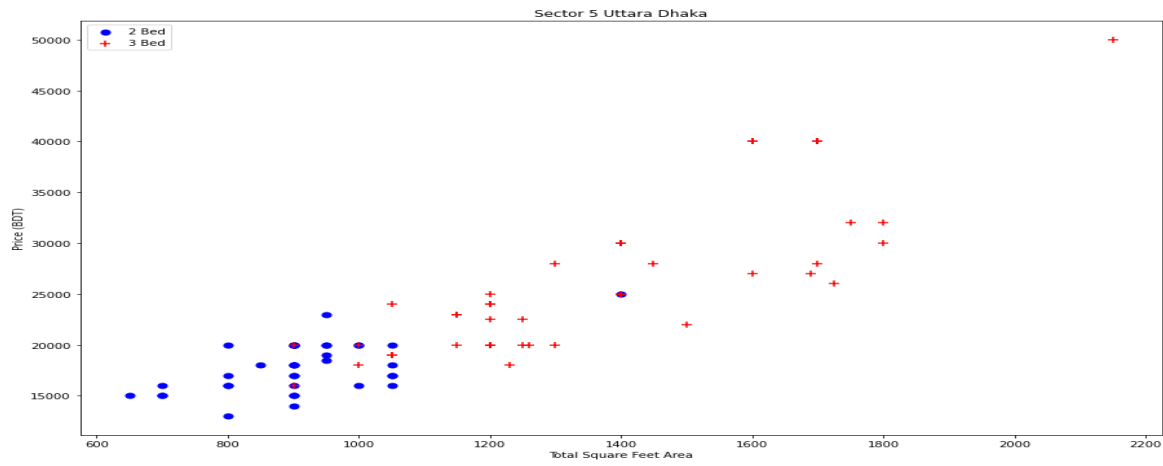


Figure 3.4.4: Uttara sector 5

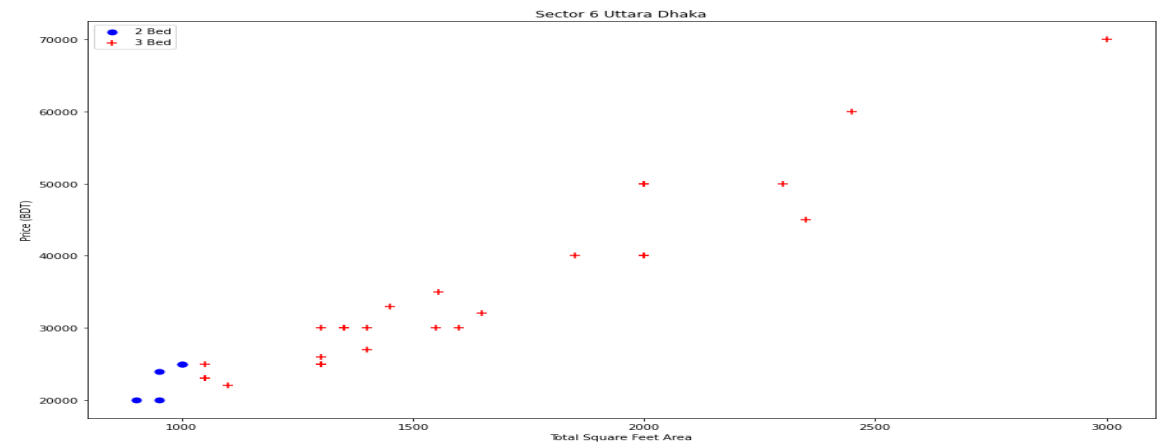


Figure 3.4.5: Uttara sector 6

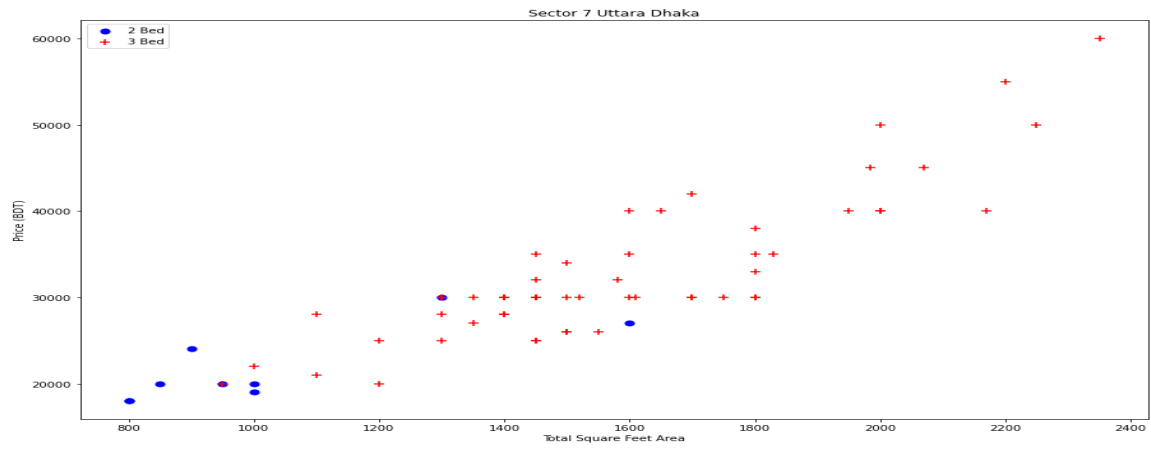


Figure 3.4.6: Uttara sector 7

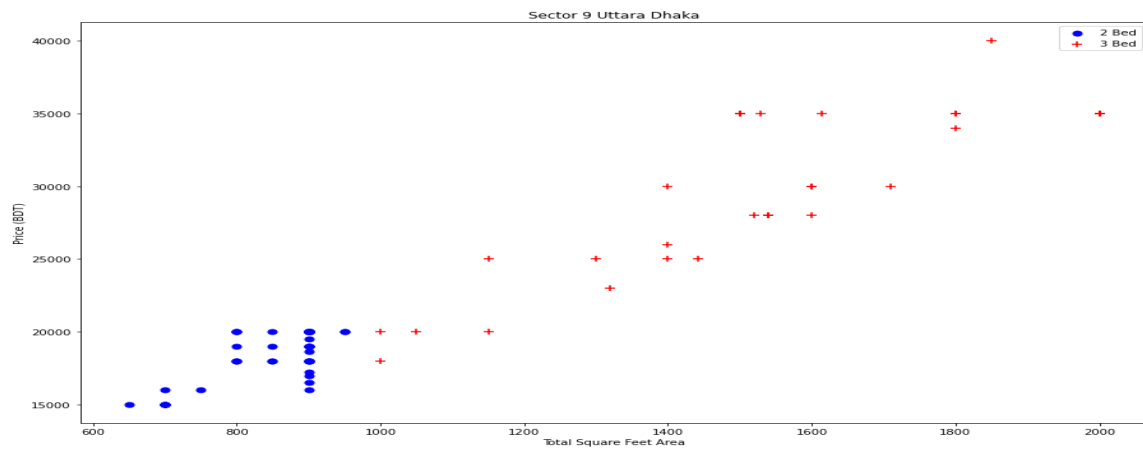


Figure 3.4.7: Uttara sector 9

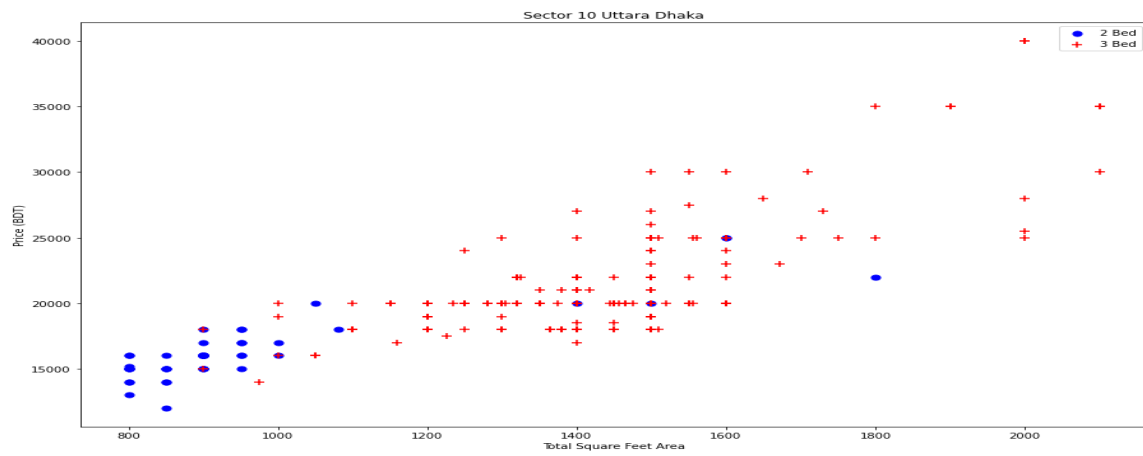


Figure 3.4.8: Uttara sector 10

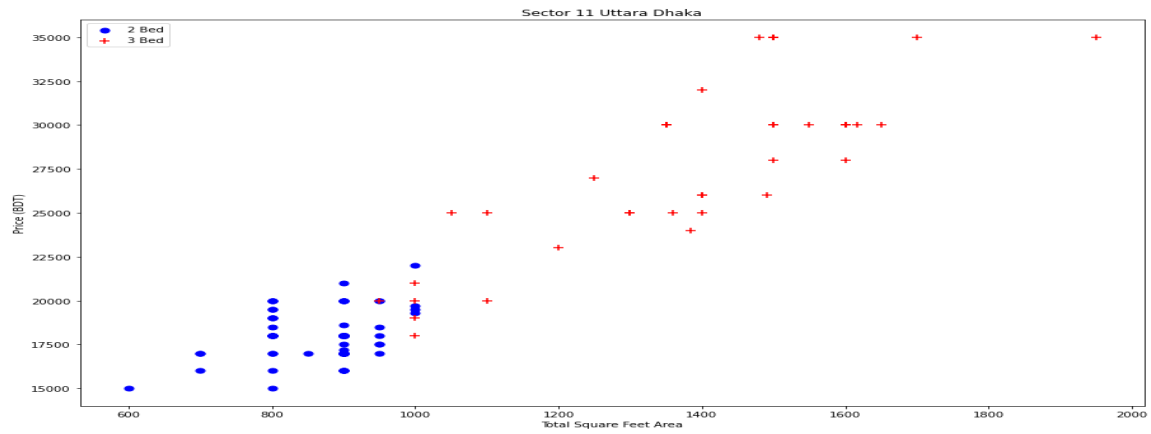


Figure 3.4.9: Uttara sector 11

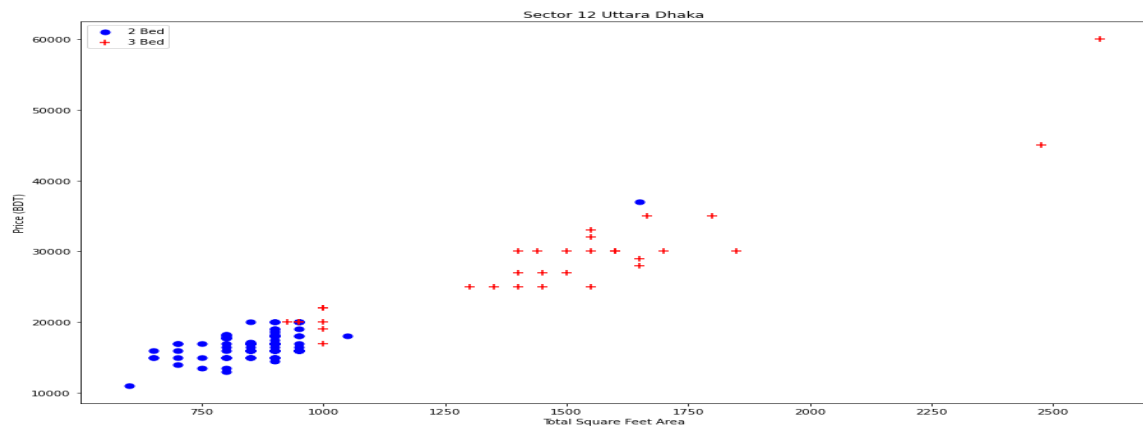


Figure 3.4.10: Uttara sector 12

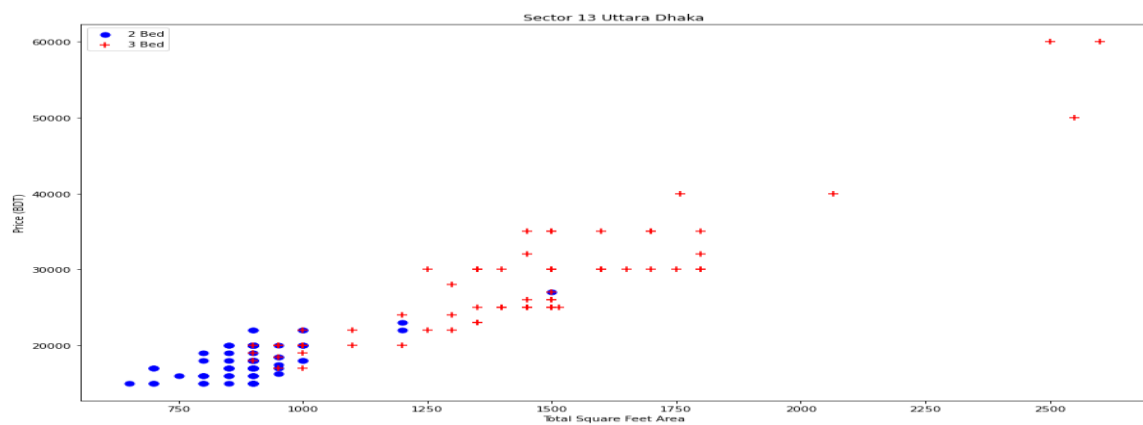


Figure 3.4.11: Uttara sector 13

Figure 3.4: Uttara house rent (depend on bedrooms)

The above scatter plot for Uttara shows the different prices for different sectors. In all sectors except sector 10 in Uttara, most of the apartments have 2 bedrooms but in sector 10 most of the apartments have 3 bedrooms. The above figure 6 shows us the rent based on the total size of the house in sq ft for 2 bedrooms and 3 bedrooms. In sector 1, the average rent for 2 bedrooms house is 24000 BDT to 27000 BDT, and for 3-bedroom houses 35000 BDT to 55000 BDT. In sector 3, we can rent a house within 20000 BDT for 2 bedrooms, and for 3 bedrooms the average rent is 35000 BDT to 45000 BDT. The average rent for a house is 30000BDT to 45000 BDT for 3 bedrooms in sector 4. Similar to sector 3, within 15000 BDT to 20000 BDT we can rent a house for 2 bedrooms in sector 5, and the average rent will be 20000 BDT to 30000 BDT for 3 bedrooms. The average rent of a house for 3 bedrooms is 35000 BDT in sector 6. In sector 7, for 2 bedrooms house, the average rent is 25000 BDT maximum. The average rent is 30000 BDT to 45000 BDT for 3 bedrooms house. 20000 BDT rent for 2 bedrooms in sector 9, and on average 25000 BDT to 30000 BDT rent for 3 bedrooms. In sector 12 and sector 13, The average rent is slightly similar. The average rent for 2 bedrooms house is 20000 BDT and the average rent for 3 bedrooms house is 25000 BDT to 35000 BDT. So, from all the scatter plots, we see that lower rent is in sector 12 and sector 13 and higher house rent is in sector 1.

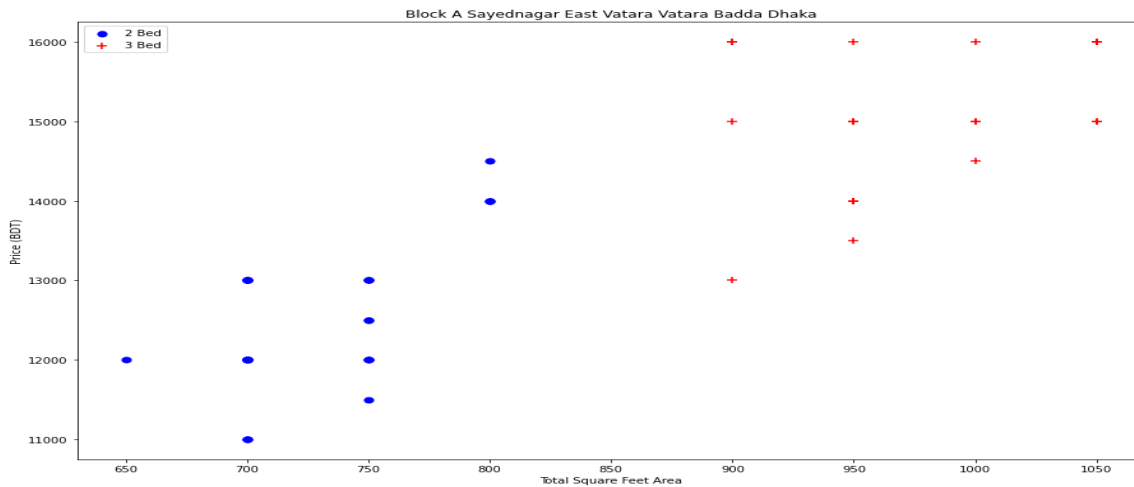


Figure 3.5.1: Badda

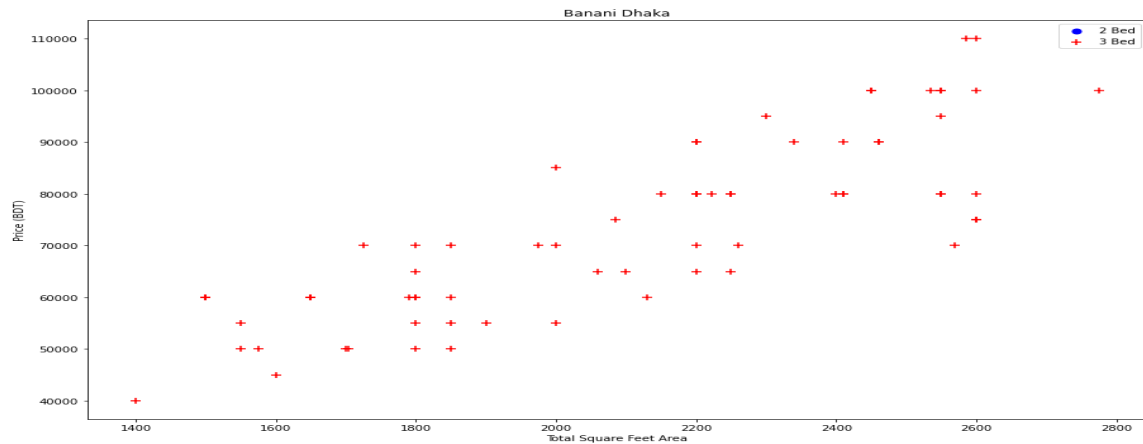


Figure 3.5.2: Banani

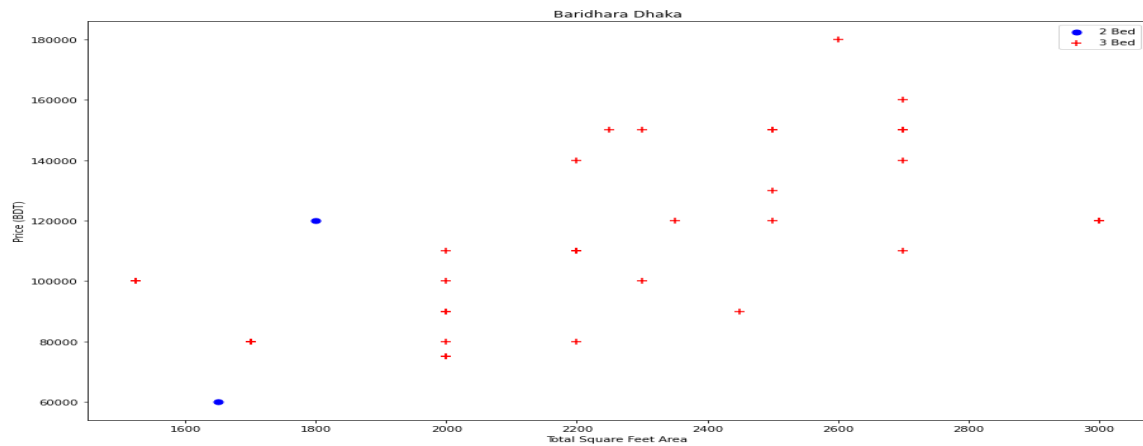


Figure 3.5.3: Baridhara

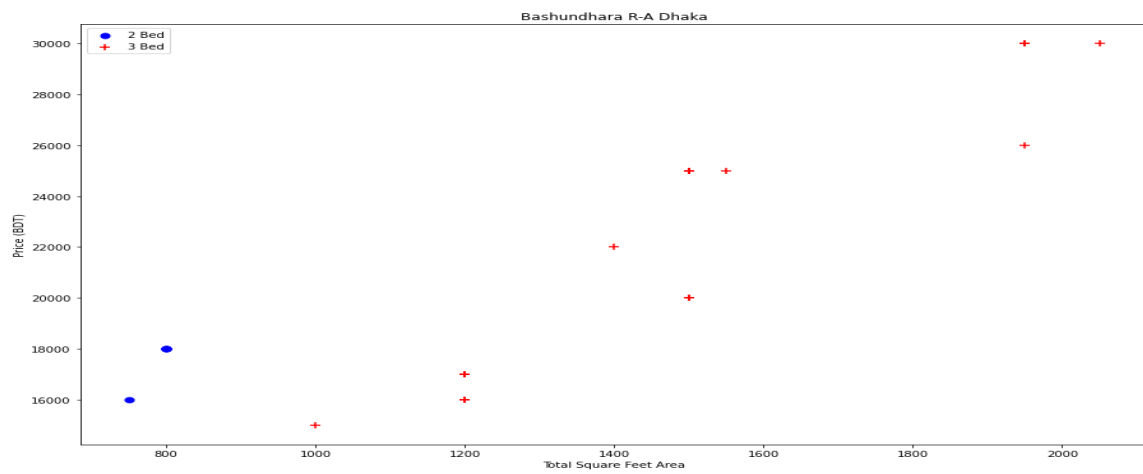


Figure 3.5.4: Bashundhara

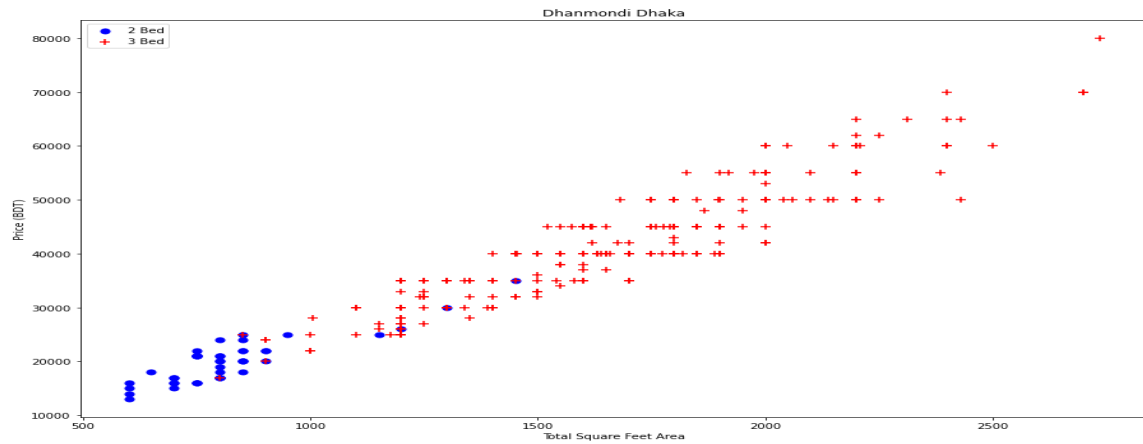


Figure 3.5.5: Dhanmondi

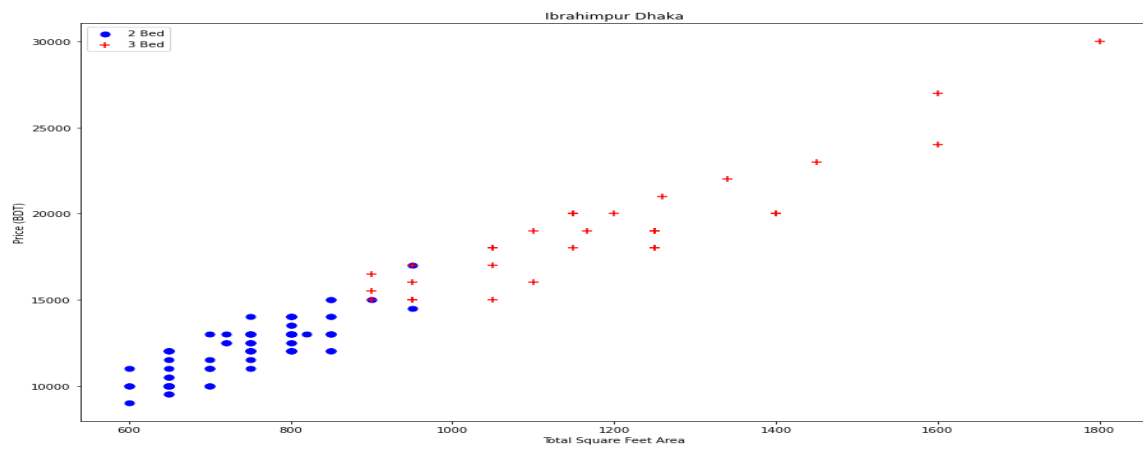


Figure 3.5.6: Ibrahimipur

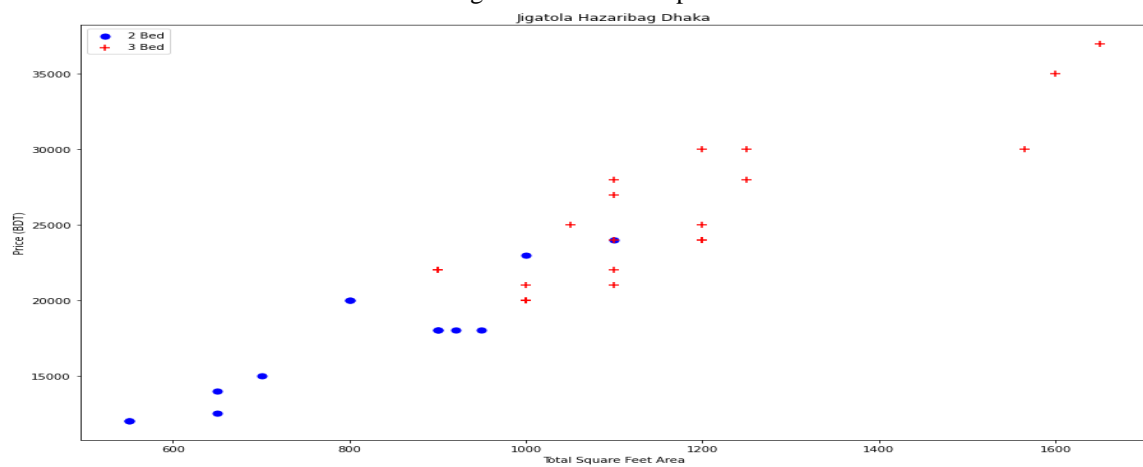


Figure 3.5.7: Jigatola Hazaribag

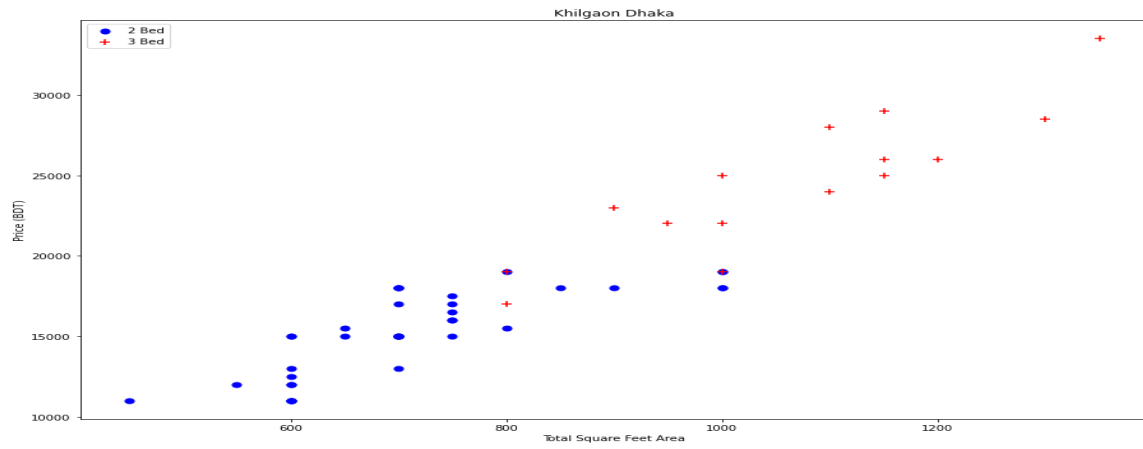


Figure 3.5.8: Khilgaon

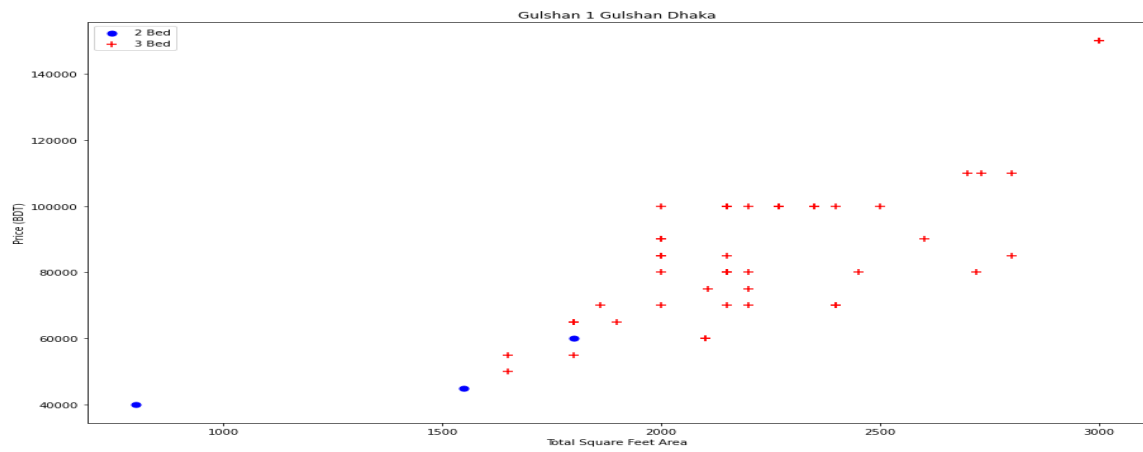


Figure 3.5.9: Gulshan

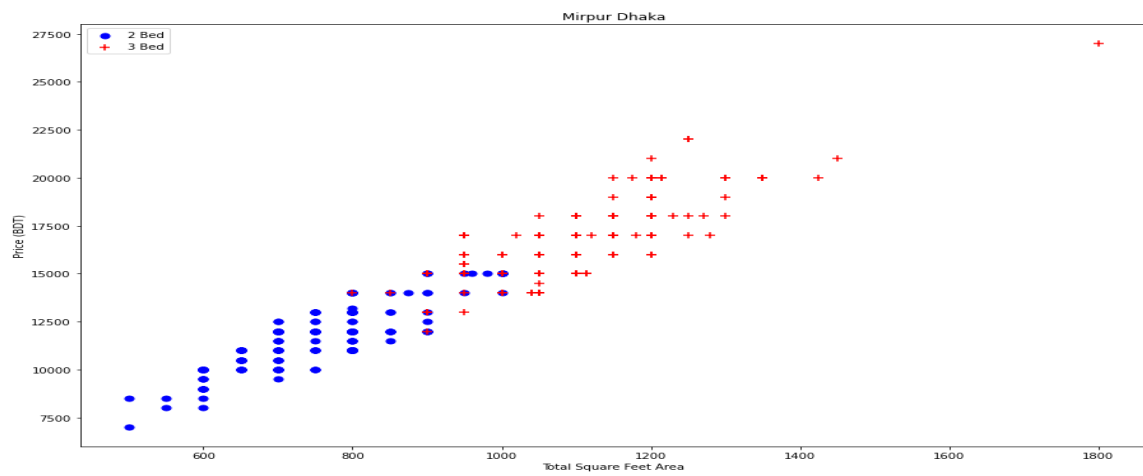


Figure 3.5.10: Mirpur

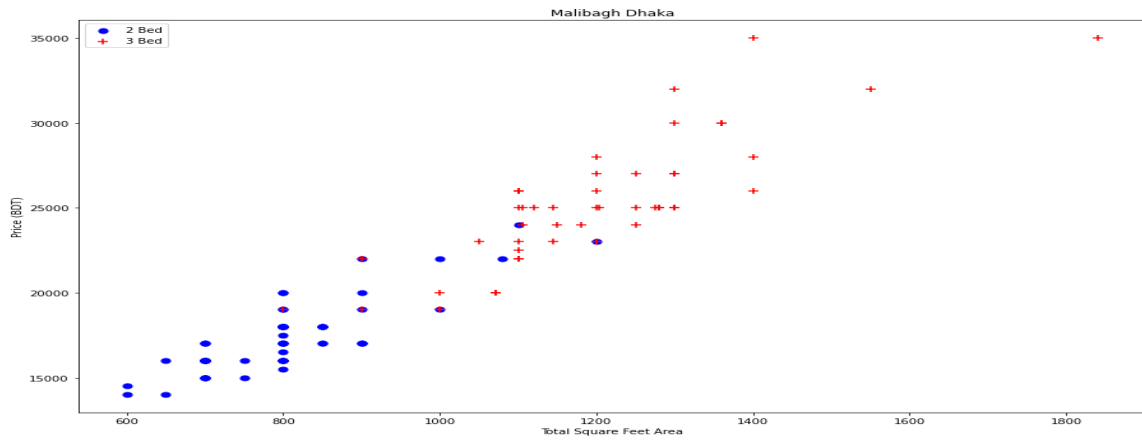


Figure 3.5.11: Mailbag

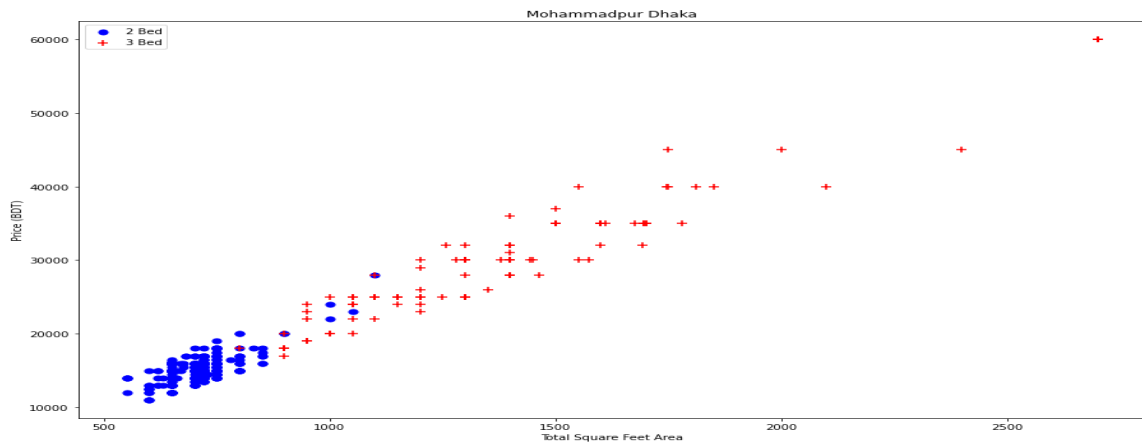


Figure 3.5.12: Mohammadpur

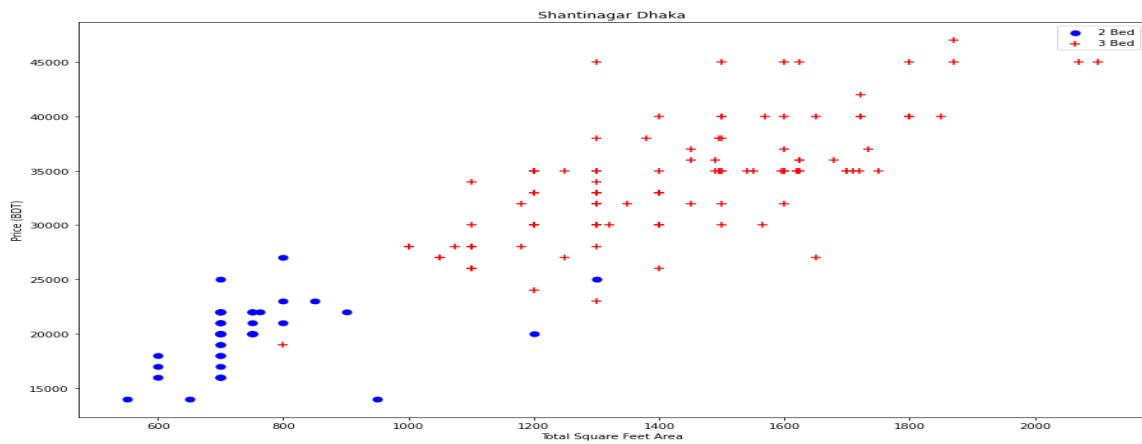


Figure 3.5.13: Shantinagar

Figure 3.5: House Rent Visualization (depends on bedrooms)

The above figure 3.5 shows us the rent of a house for different places in Dhaka city for 2 bedrooms and 2 bedrooms. We can see from the scatter plot most of the apartments have 3 bedrooms.

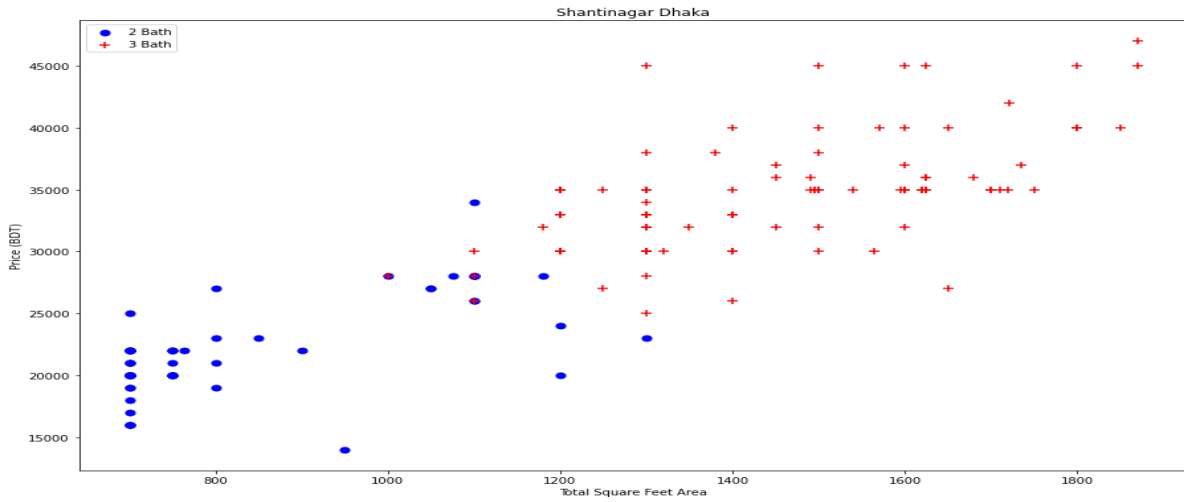


Figure 3.6.1: Shantinagar

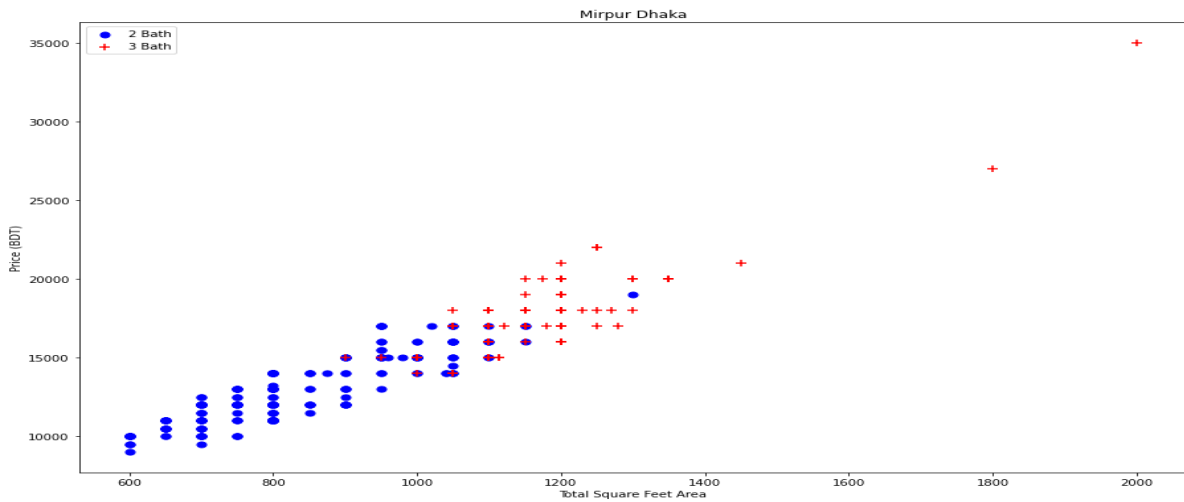


Figure 3.6.2: Mirpur

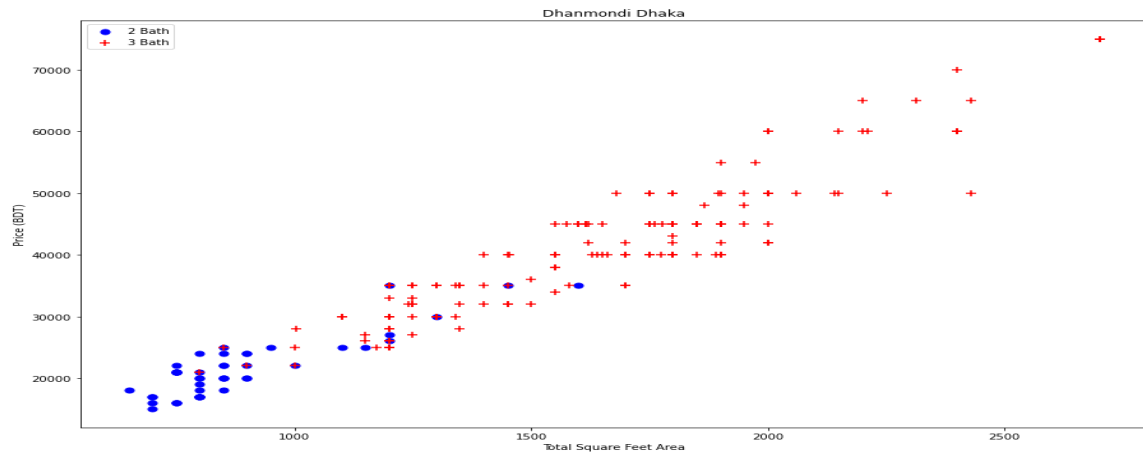


Figure 3.6.3: Dhanmondi

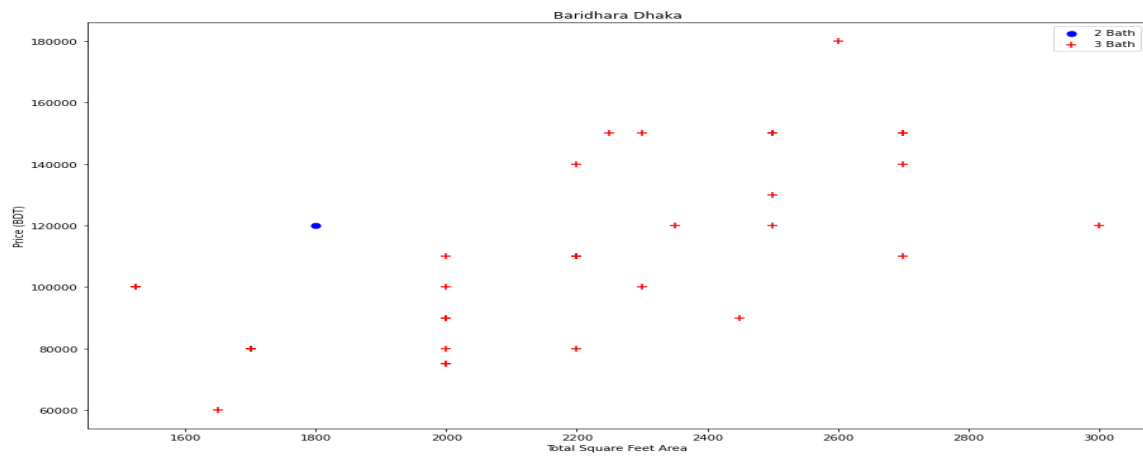


Figure 3.6.4: Baridhara

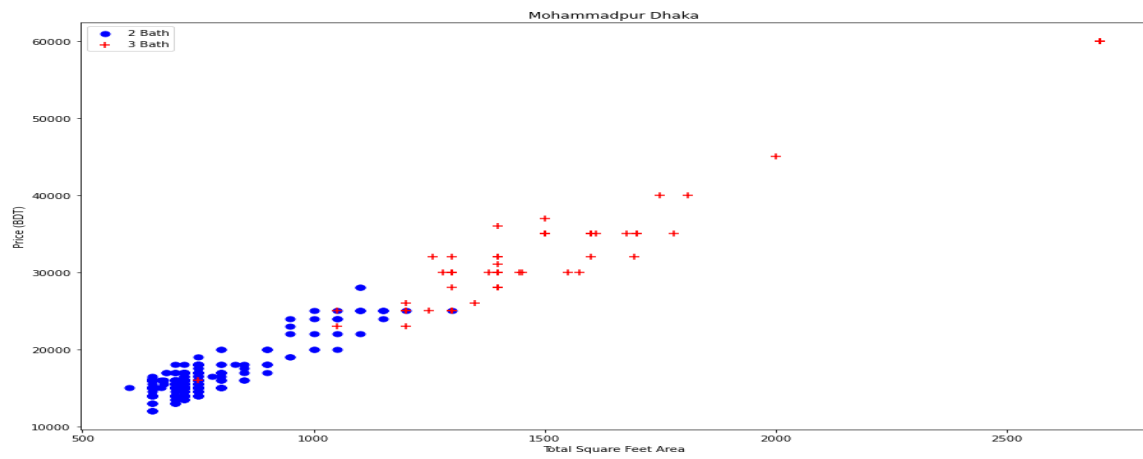


Figure 3.6.5: Mohammadpur

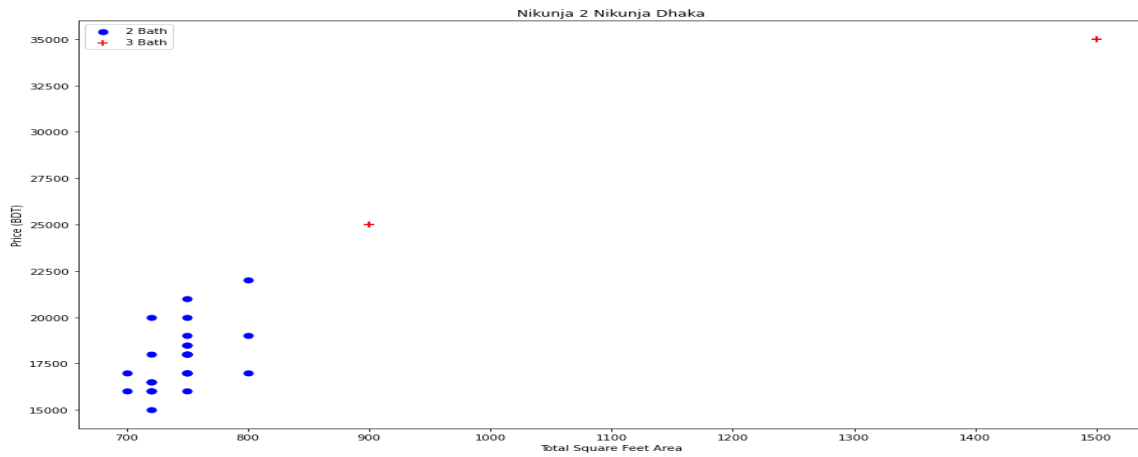


Figure 3.6.6: Nikunja-2

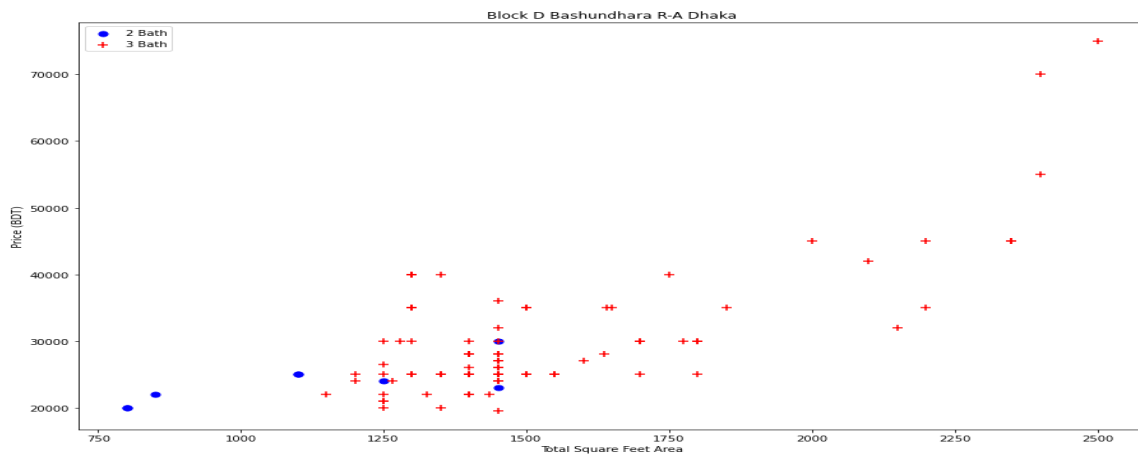


Figure 3.6.7: Bashundhara

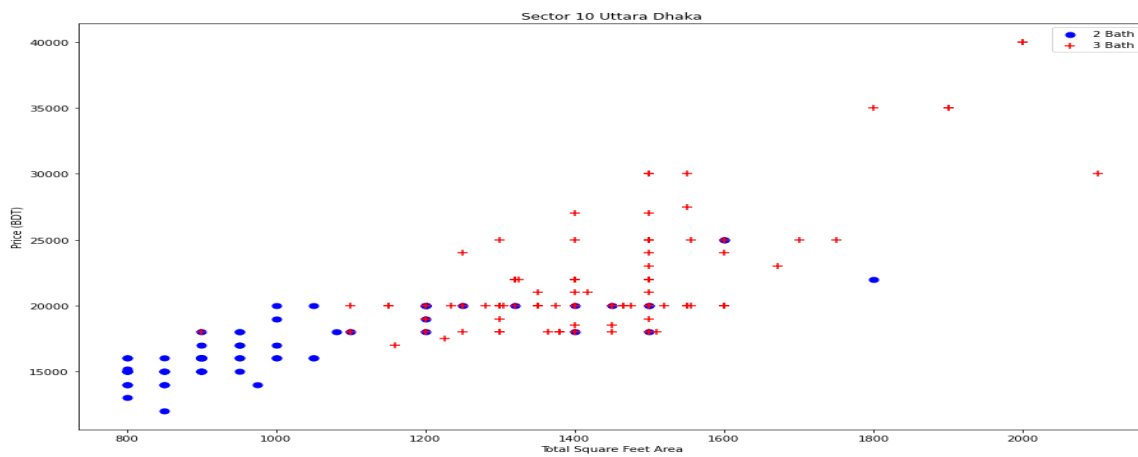


Figure 3.6.8: Uttara sector 10

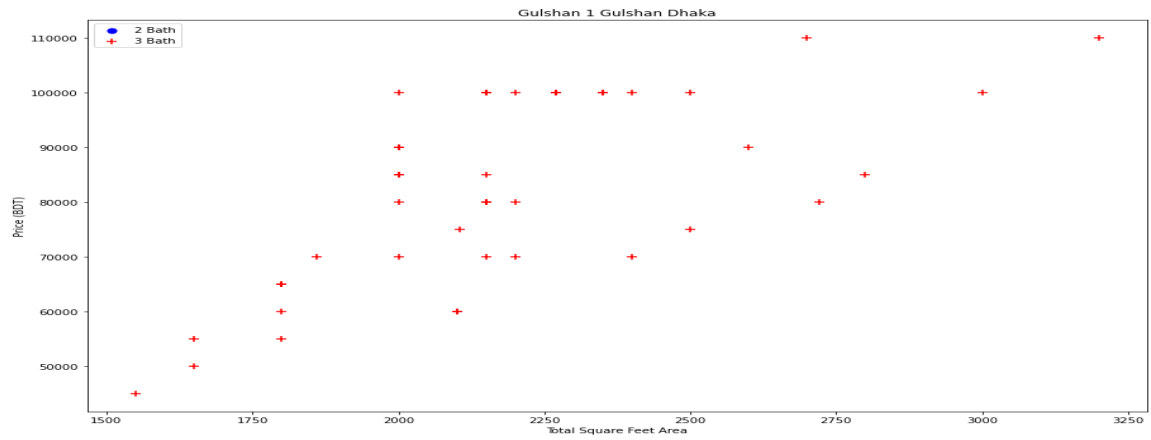


Figure 3.6.9: Gulshan 1

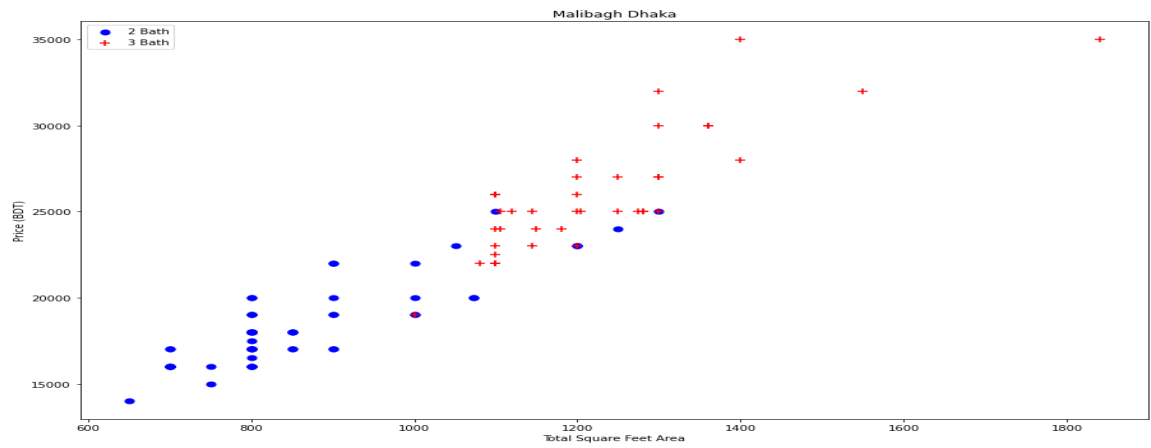


Figure 3.6.10: Malibagh

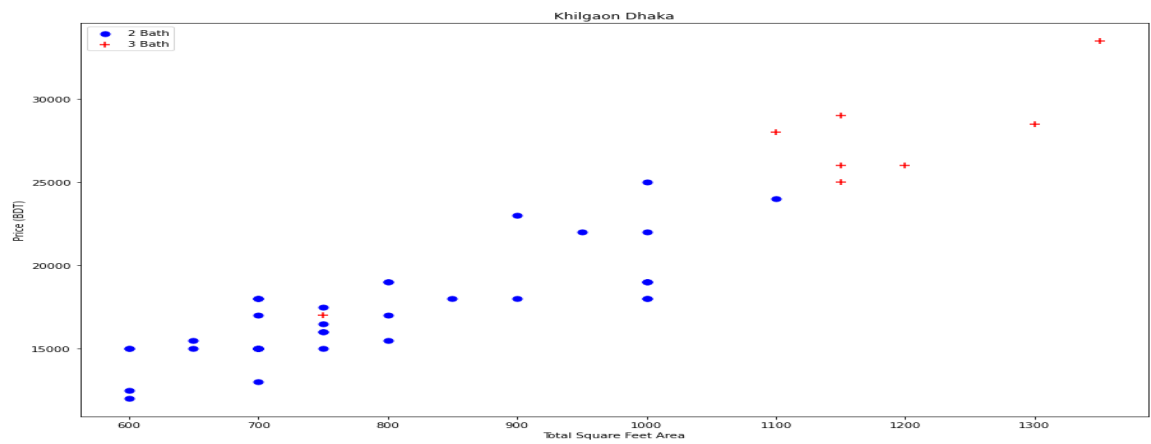


Figure 3.6.11: Khilgaon

Figure 3.6: House Rent Visualization (depends on bathrooms)

The above figure 3.6 shows us the visual representation of house rent based on the total number of bathrooms available in that apartment.

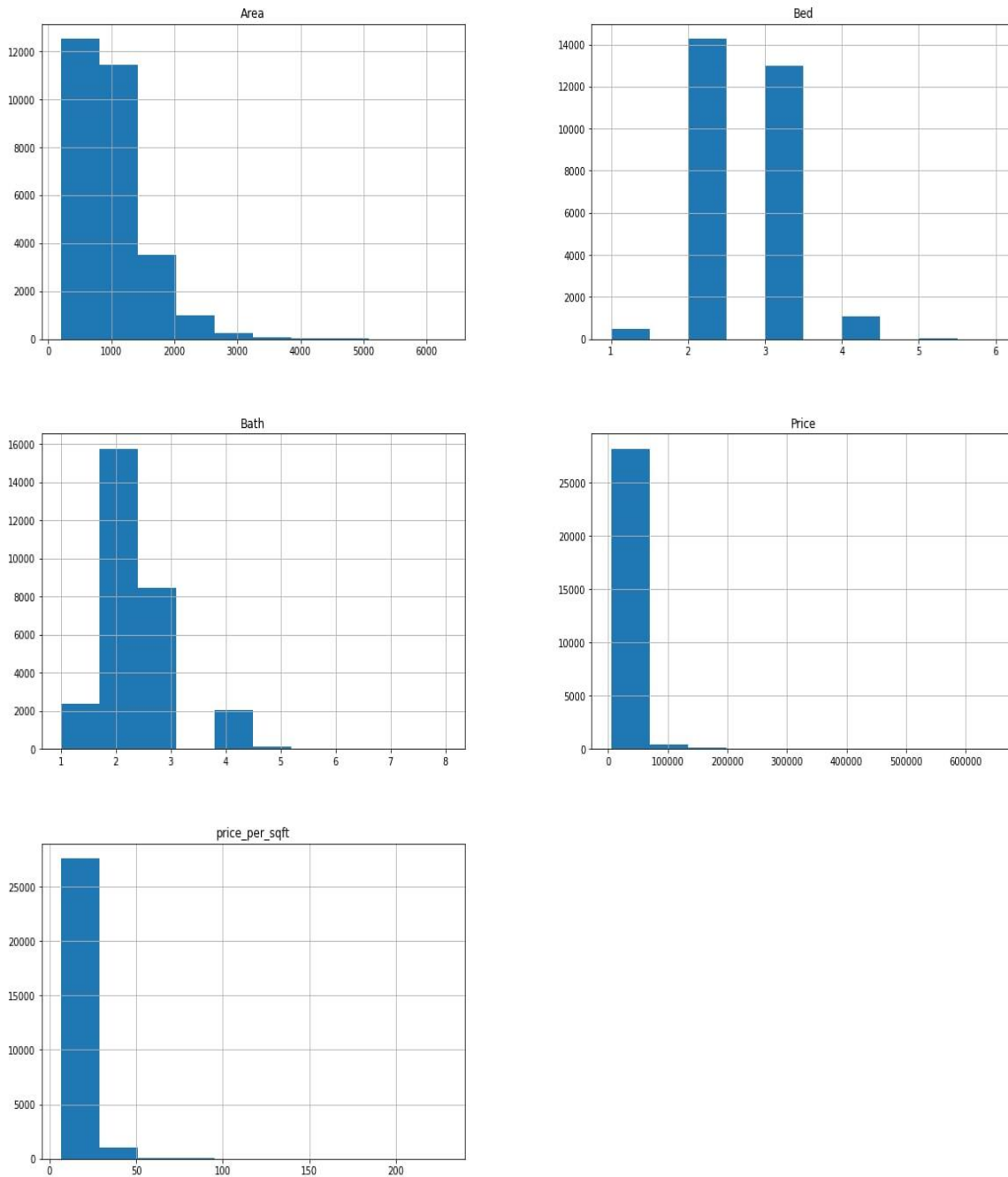


Figure 3.7: Histogram for each attribute

3.6 Feature Engineering

Performed a number of features engineering steps, including as each unnamed column must be turned into many dummy variables for regular multiple linear regression models (OLS) and regularized linear regression models, try to keep the number of unnamed columns as low as possible or suitable. Add a new variable with a promising functionality based on your subject expertise. Get rid of anomalies in the data set. Eliminate unnecessary punctuation from the dataset.

3.7 Data Preprocessing

The model in this study is evaluated and trained to utilize data from bProperty.com. The preprocessing of the data includes:

3.7.1 Removing the unnamed column

Pandas will automatically drop the unnamed column. The `df.set_index()` method can be called on any column to set the index for the data frame.

Table 3.3: With Unnamed Column

	Unnamed: 0	Location	Area	Bed	Bath	Price
0	0	Block H, Bashundhara R-A, Dhaka	1,600 sqft	3	3	20 Thousand
1	1	Farmgate, Tejgaon, Dhaka	900 sqft	2	2	20 Thousand
2	2	Block B, Nobodoy Housing Society, Mohammadpur,...	1,250 sqft	3	3	18 Thousand
3	3	Gulshan 1, Gulshan, Dhaka	2,200 sqft	3	4	75 Thousand
4	4	Baridhara, Dhaka	2,200 sqft	3	3	75 Thousand
5	5	Bashundhara R-A, Dhaka	3,000 sqft	4	5	50 Thousand
6	6	Baridhara, Dhaka	2,300 sqft	3	3	75 Thousand
7	7	PC Culture Housing, Mohammadpur, Dhaka	950 sqft	2	2	14 Thousand
8	8	Jigatola, Hazaribag, Dhaka	1,600 sqft	3	3	28 Thousand
9	9	West Kazipara, Mirpur, Dhaka	1,150 sqft	3	3	19 Thousand

Table 3.4: Without Unnamed Column

	Location	Area	Bed	Bath	Price
0	Block H, Bashundhara R-A, Dhaka	1,600 sqft	3	3	20 Thousand
1	Farmgate, Tejgaon, Dhaka	900 sqft	2	2	20 Thousand
2	Block B, Nobodoy Housing Society, Mohammadpur,...	1,250 sqft	3	3	18 Thousand
3	Gulshan 1, Gulshan, Dhaka	2,200 sqft	3	4	75 Thousand
4	Baridhara, Dhaka	2,200 sqft	3	3	75 Thousand

3.7.2 Numeric Variables

A categorical variable is a string that poses a threat to models. Dummy variables should be created for categorical variables. These are numerical constants that will help the models work better. Utilize categorical variables.

3.7.3 Removing Outliers

The outlier values made our graph's scale unrealistic. The challenge was that there was no fixed number of outlier values. We would sometimes get all valid deals but sometimes we would get erroneous readings that covered as much as 10% of the data. That's why we eliminated outliers.

Table 3.5: Outlier of dataset

	Location	Area	Bed	Bath	Price	price_per_sqft
39	other	800	3	2	15000.0	18.750000
86	Block C Section 13 Mirpur Dhaka	745	3	2	15000.0	20.134228
135	other	800	3	2	15000.0	18.750000
186	Nayatola Maghbazar Dhaka	550	2	2	13000.0	23.636364
191	Baitul Aman Housing Society Adabor Dhaka	550	2	1	8500.0	15.454545
193	Taltola Khilgaon Dhaka	550	2	1	11000.0	20.000000
195	South Monipur Mirpur Dhaka	850	3	2	13000.0	15.294118
205	other	800	3	2	15000.0	18.750000
252	Block C Kaderabad Housing Society Bosila Moham...	750	3	2	16000.0	21.333333
260	Faydabad Dakshin Khan Dhaka	800	3	2	8500.0	10.625000

3.7.4 Remove commas from the area Column

Using the Find and Replace feature, we can identify and eliminate all the commas from a cell. Only text data can be used using this method. This procedure will not get rid of a comma if there is one in your numbers because of formatting.

3.8 Dataset Split

After initial processing, we split our data set into training and test set. Where the Test set consists of 25 % data and the training set consists of 75 % data.

3.9 Model

The following regression models are used in this work. These regression models are Linear regression, Ridge regression, Lasso regression, and Bayesian regression.

3.9.1 Linear Regression

To establish the relationship between two continuous variables, linear regression is used. The first variable is a predictor or independent variable, and the second is a variable response or dependent variable. It seeks a statistically significant but non-deterministic link. If the other can precisely express one variable, it is said that the relationship between the two variables is deterministic.

$$Y = \theta_1 * X + \theta_0 \quad (1)$$

where y is the dependent variable, and x is the independent variable. Theta is the coefficient factor

3.9.2 Ridge Regression

Ridge regression is a method for analyzing multi-regression data that suffer from multicollinearity. Generally, square estimates are unbiased when multicollinearity occurs, but there are as variances are large, their actual value may differ greatly from what they appear to be. Ridge regression reduces the standard deviation by adding a degree of bias to regression estimates, errors are introduced. The net effect should be that estimates become more reliable. Edge relapse is adjusting

the least squares strategy to permit having one-sided assessors of the relapse coefficients in the relapse model. Edge relapse puts a specific type of imperative on boundaries.

$$RSS = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 \quad (2)$$

3.9.3 Bayesian Regression

In Bayesian regression, the mean of one boundary is described by a weighted number of different factors. This sort of restrictive demonstration expects to decide the earlier dispersion of the regressors as well as different factors depicting the designation of the regressed and in the long run, allows the out-of-test estimating of the regressed contingent on perceptions of the relapse coefficients.

The typical direct condition, where the appropriation of show style YY given by show style XX is Gaussian, is the most fundamental and famous variation of this model. The future can be resolved systematically for this model, and a particular arrangement of earlier probabilities for the boundaries is known as form priors. The rear ends mostly have all the more arbitrarily chosen priors.

Bayesian Relapse may be very useful when the dataset has too little or inadequately scattered information. Rather than customary relapse procedures, where the result is just gotten from a solitary number of each quality, a Bayesian Relapse model's result is gotten from a likelihood conveyance.

The outcome, "y," is created by an ordinary circulation (where the change and mean are standardized). The objective of the Bayesian Relapse Model is to recognize the 'back' dispersion again for model boundaries instead of the model boundaries themselves. The model boundaries will be supposed to follow a conveyance notwithstanding the result y.

$$y \sim N(\beta^T X, \sigma^2 I) \quad (3)$$

A normal (Gaussian) distribution with a mean and variance generates the output, y . The weight matrix is transposed by the prediction matrix, and the result is the mean for linear regression. The square of the standard deviation represents the variance.

Instead of attempting to isolate the one "best" value for each model parameter, Bayesian Linear Regression seeks to identify the posterior distribution of those values.

3.9.4 Lasso Regression

Lasso regression relapses and performs L1 regularization, which adds a punishment equivalent to the outright worth of the greatness of coefficients. This sort of regularization can bring about scanty models with not many coefficients; A few coefficients can become zero and kill the model. Bigger punishments bring about coefficient esteems more like zero, which is great for creating less difficult models. Then again, L2 regularization doesn't bring about the disposal of coefficients or scanty models. This makes the Lasso far more straightforward to decipher than the ridge. expected to follow dissemination notwithstanding the result y .

$$\sum_{i=1}^n (y_i - \sum_j x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (4)$$

Because the penalty function in the regression of Lasso utilizes absolute values rather than squares, it varies from the regression of the ridge. Because of the penalty (or analogous limitation of the total of the absolute values of the estimates), some parameter estimations end up being exactly zero. Estimates are further decreased to zero the more the penalty is applied.

CHAPTER 4

EXPERIMENTAL RESULT AND DISCUSSION

4.1 Result

We applied Linear, Ridge, LASSO (least absolute shrinkage and selection operator), and Bayesian regression models to train the dataset and produce predictions independently.

4.1.1 Accuracy of models

Analyze the outcomes of our model, provided in the table below.

Table 4.1. Predictors Score according to Model

Model Name	Accuracy (%)
Linear Regression	91.49
Ridge Regression	91.54
Bayesian Regression	91.52
Lasso Regression	91.49

As table 5 shows, every model predicts an almost identical outcome. The best outcome is given by Ridge Regression in this case, with a percentage of 91.54. With a percentage of 91.52, Bayesian Regression is then trailing Ridge Regression. However, with a percentage of 91.49, linear and lasso regression are practically identical to the two models before them.

4.1.2 Compare between prediction rent and original rent

Here is a short comparison between the original house rent and predicted data.

Table 4.2. Original Rent vs Predicted Rent

Area	Space(sq ft)	Bed	Bath	Original Rent	Predicted Rent
Matikata, Cantonment	1300	3	3	16300	17089
West Shewrapara, Mirpur	1200	3	3	20000	20149
Gulshan 2 Gulshan	2000	3	4	100000	101968
Sector 6, Uttara,	1850	3	3	44000	42600
Mohammadpur,	700	2	2	16000	15450
Dhanmondi,	2450	4	5	60000	59550
Banani, Dhaka	2250	3	3	65000	63700
Shyamoli Housing Shekhertek	1350	3	3	20000	21450
Uttar Badda, Badda,	1250	3	3	22000	23450

In table 6, we have predicted the data for the actual house rent. To predict the rent of an apartment we have considered the area of the house, total size(sqft) of the apartment, total bedrooms, and total bathrooms. Here are the few predicted results that are given which is predicted by the model. This rent of a house depends based on the location of the house, size, and total bedrooms and bathrooms in the apartment. From table 6, see a huge difference in house rent based on the area. In the residential area, the house rent is higher compared to the nonresidential area. In the residential area, there are extra facilities like security and safety and better educational institute.

The highest rent of a house in Gulshan and the lowest rent of a house in matikata cantonment Dhaka. The above table is like a shadow of a full dataset. The difference between the predicted price and the original price is highest for sectors 6, Uttara, and Dhaka, and the lowest difference is for West Shewrapara, Mirpur, and Dhaka.

4.1.3 Most numbers of Houses

Here, the top 20 locations with the most numbers of houses.

Table 4.3: Locations with the most numbers of houses

Location	Number of House
Mohammadpur	757
Mirpur	556
Block D Section 12 Mirpur	417
Dhanmondi	414
Block E Section 12 Mirpur	411
Sector 10 Uttara	357
Paikpara Ahmed Nagar Mirpur	352
Kallyanpur Mirpur	337
Section 12 Mirpur	311
Block B Section 12 Mirpur	307
Joar Sahara Dhaka 305 Block C Section 12 Mirpur	294
West Shewrapara Mirpur	292
Shyamoli	285
PC Culture Housing Mohammadpur	274

Hazaribag	260
South Baridhara Residential AreaD. I. T. Project	256
Block G Bashundhara R-A	255
Sector 13 Uttara	250
Uttar Badda	248
Baitul Aman Housing Society Adabor	241

We have a total of 730 unique locations in our dataset. Table 7 shows the top 20 locations with the most house. The highest house is located in Mohammadpur Dhaka and the lowest number of houses is 1 which is located on Tilpapara Road, Tilpapara Khilgaon Dhaka, Old Elephant Road Maghbazar Dhaka, Road No 16 Banani Dhaka, South Kutubkhali Kutubkhali Jatra Bari Dhaka, Majed Sardar Road Bangshal Dhaka.

CHAPTER 5

Conclusion and Future Work

5.1 Conclusion

In our research, we described a model for predicting the price of a rental house in Dhaka city that combines ensemble learning and advanced regression methods. We hope to acquire a result from this model that will enable us to quickly assess housing data in order to demonstrate appropriate prices for houses based on their provided sets of qualities. The model's accuracy is very comparable to that of other models. The Ridge Regression model gave us the best accuracy of 91.54%, with other models like linear, Bayesian, and lasso coming in a close second.

The individual prediction results are utilized as inputs for training the ensemble predictor, which culminates in the creation of the ensemble model. The benefit of this model is that it can accurately factor in several parameters to determine the rental price. The type of feature factoring from a separate dataset is another significant finding. When using dummy values and inter-variable interactions to forecast, categorical values have an impact on the models. The pioneer features for which the house price changes the greatest are the "total bed" and "total bath". The cost of renting a house also depends on the area; in Gulshan, for example, it will be more expensive than in Mirpur. Also, the residential area provides more security than the normal area. for this reason, the residential area is costlier than the normal area.

5.2 Future Work

We want to implement a web application based on this research. So that a consumer can easily get information about the house rent. Where a consumer can search for the desired house with the desire rent. They can filter by location, the total area of the house, and available bedrooms and bathrooms in the house as well.

REFERENCES

- [1] “World City Populations 2022,” Worldpopulationreview.com. [Online]. Available: <https://worldpopulationreview.com/world-cities>. [Accessed: 25-Dec-2022].
- [2] Tribune Desk, “Census 2022: Dhaka division home to 44 million people now,” Dhaka Tribune, 27-Jul-2022. [Online]. Available: <https://www.dhakatribune.com/dhaka/2022/07/27/census-2022-dhaka-home-to-44-million-people-now>. [Accessed: 25-Dec-2022].
- [3] T. Akter, M. M. Hasan, A. U. Zaman, M. R. Hossain, and M. N. Neema, “Analysis of potential factors bringing disparity in house rent of Dhaka city,” in Lecture Notes in Computer Science, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 408–421.
- [4] M. Samsuddin Ahmed, M. Rahman, and S. Islam, “House rent estimation in Dhaka city by multi-layer perceptions neural network,” International Journal of u-and e-Service, vol. 7, no. 4, pp. 287–300, 2014.
- [5] S. Shams, M. Mahruf, C. Shohel, and A. Ahsan, “Housing problems for middle and low income people in Bangladesh: Challenges of Dhaka Megacity,” Environment and Urbanization Asia, vol. 5, no. 1, pp. 175–184, 2014.
- [6] Wikipedia contributors, “Ridge regression,” Wikipedia, The Free Encyclopedia, 22-Dec-2022. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Ridge_regression&oldid=1128872620.
- [7] W. Koehrsen, “Introduction to Bayesian linear regression,” Towards Data Science, 14-Apr-2018. [Online]. Available: <https://towardsdatascience.com/introduction-to-bayesian-linear-regression-e66e60791ea7>. [Accessed: 25-Dec-2022].
- [8] “ML,” GeeksforGeeks, 13-Sep-2018. [Online]. Available: <https://www.geeksforgeeks.org/ml-linear-regression/>. [Accessed: 25-Dec-2022].

- [9] Simplilearn, “Introduction to Bayesian Linear Regression,” Simplilearn.com, 31-Oct-2022. [Online]. Available: <https://www.simplilearn.com/tutorials/data-science-tutorial/bayesian-linear-regression>. [Accessed: 25-Dec-2022].
- [10] Wikipedia contributors, “Lasso (statistics),” Wikipedia, The Free Encyclopedia, 09-Oct-2022. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=Lasso_\(statistics\)&oldid=1114996444](https://en.wikipedia.org/w/index.php?title=Lasso_(statistics)&oldid=1114996444).
- [11] A. A. Neloy, H. M. S. Haque, and M. M. Ul Islam, “Ensemble learning based rental apartment price prediction model by categorical features factoring,” in Proceedings of the 2019 11th International Conference on Machine Learning and Computing - ICMLC '19, 2019.
- [12] S. Ahmed, M. M. Rahman, and S. Islam, “House rent estimation in Dhaka city by multi layer perceptions neural network,” *Int. J. U- E- Serv. Sci. Technol.*, vol. 7, no. 4, pp. 287–300, 2014.
- [13] I. M. Badhan and A. Siddika, “Evaluating Rental Cost and Housing Ownership Affordability Criteria of Middle-Income Group in CBD-Tejgaon Residential Area Dhaka,” *Journal of Housing and Advancement in Interior Designing*, vol. 2, no. 1, 2019.
- [14] S. Law, “Defining Street-based Local Area and measuring its effect on house price using a hedonic price approach: The case study of Metropolitan London,” *Cities*, vol. 60, pp. 166–179, 2017.
- [15] Wikipedia contributors, “Dhaka City Corporation,” Wikipedia, The Free Encyclopedia, 01-Apr-2022. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Dhaka_City_Corporation&oldid=1080435298
- [16] S. H. Khan, “Dhaka South City Corporation ward list (with map) - bproperty,” A blog about homes, trends, tips & life | Bproperty, 09-Sep-2021. [Online]. Available: <https://www.bproperty.com/blog/dhaka-south-city-corporation-ward-list/>. [Accessed: 26-Dec-2022].

- [17] “Property & Real Estate for sale and for rent in Bangladesh,” Bproperty.com. [Online]. Available: <https://www.bproperty.com/>. [Accessed: 26-Dec-2022].
- [18] Tribune Desk, “Bproperty on the rise,” Dhaka Tribune, 22-Oct-2019. [Online]. Available: <https://archive.dhakatribune.com/business/real-estate/2019/10/22/bproperty-on-the-rise>. [Accessed: 26-Dec-2022].
- [19] M. Sharif Hossain, Innovations of Housing Financing Systems and the Implications in, vol. 1. Bangladesh.
- [20] K. I. Ahmed, Urban poor housing in Bangladesh and potential role of ACHR. Bangkok, Thailand: ACHR, 2007.
- [21] Y. Ming, J. Zhang, J. Qi, T. Liao, M. Wang, and L. Zhang, “Prediction and analysis of Chengdu housing rent based on XGBoost algorithm,” in Proceedings of the 2020 3rd International Conference on Big Data Technologies, 2020.
- [22] L. Hu et al., “Monitoring housing rental prices based on social media: An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies,” Land Use Policy, vol. 82, pp. 657–673, 2019.

House Rent Prediction in Dhaka City Using Machine Learning Approach

ORIGINALITY REPORT

20%

SIMILARITY INDEX

12%

INTERNET SOURCES

10%

PUBLICATIONS

9%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Intercollege

Student Paper

2%

2

dspace.daffodilvarsity.edu.bd:8080

Internet Source

2%

3

Asif Ahmed Nelay, H. M. Sadman Haque, Md. Mahmud Ul Islam. "Ensemble Learning Based Rental Apartment Price Prediction Model by Categorical Features Factoring", Proceedings of the 2019 11th International Conference on Machine Learning and Computing - ICMLC '19, 2019

Publication

1%

4

Submitted to University of Derby

Student Paper

1%

5

en.wikipedia.org

Internet Source

1%

6

Lecture Notes in Computer Science, 2013.

Publication

1%

www.dhakatribune.com