

**FORECASTING HOUSE RENTS IN DHAKA CITY WITH
MACHINE LEARNING**

BY

**Dewan Abdullah Fahad
ID: 201-15-13586**

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

Supervised By
Md. Ferdouse Ahmed Foysal
Lecturer
Department of CSE
Daffodil International University

Co-Supervised By
Md. Shakhawath Hossain
Lecturer
Department of CSE
Daffodil International University

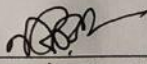


**DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
JANUARY 2024**

APPROVAL

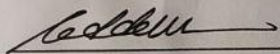
This Project titled “Forecasting House Rents in Dhaka City with Machine Learning”, submitted by Dewan Abdullah Fahad, ID No: 201-15-13586 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements 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 25 January, 2024.

BOARD OF EXAMINERS



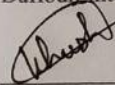
Narayan Ranjan Chakraborty
Associate Professor & Associate Head
Department of CSE
Daffodil International University

Chairman



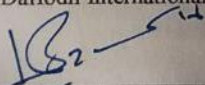
Saiful Islam
Assistant Professor
Department of CSE
Daffodil International University

Internal Examiner



Shayla Sharmin
Senior Lecturer
Department of CSE
Daffodil International University

Internal Examiner



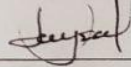
Dr. Md. Sazzadur Rahman
Professor
Institute of Information Technology
Jahangirnagar University

External Examiner

DECLARATION

I hereby declare that this project has been done by us under the supervision of **Md. Ferdouse Ahmed Foysal, Lecturer, Department of CSE** Daffodil International University. I 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:



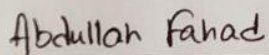
Md. Ferdouse Ahmed Foysal
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:



Md. Shakhawath Hossain
Lecturer
Department of CSE
Daffodil International University

Submitted by:



Dewan Abdullah Fahad
ID: 201-15-13586
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes me possible to complete the final year project/internship successfully.

I really grateful and wish my profound my indebtedness to **Md. Ferdouse Ahmed Foyzal, Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my 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 stage have made it possible to complete this project.

I would like to express my heartiest gratitude to **Dr. Sheak Rashed Haider Noori, Professor and Head**, Department of CSE, for his kind help to finish my project and to other faculty member and the staff of CSE department of Daffodil International University.

I would like to thank my entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

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

ABSTRACT

Bangladesh's capital city of Dhaka is become increasingly congested due to daily migration. They come to Dhaka for reasons like they have no money, looking for work, or sometimes because their families want them to shift to Dhaka city. So many people are coming to Dhaka that's why finding a place to live in Dhaka is getting really hard, especially for renting houses. Because it's hard to choose the right one. The house prices in Dhaka city are going up because there is a lot of demand for the house but there aren't enough houses for everyone. This study tries to find out how much a house should cost to rent in Dhaka City. In this study we see the different parts of Dhaka city's houses rent. Numerous elements are taken into account, such as the house's size, location, number of bedrooms, and number of bathrooms and predict the rent. In this work, we forecast housing rent using a variety of machine-learning regression algorithms and evaluate the accuracy of each model. The chosen algorithms that have been included consist of Lasso regression, Bayesian regression, Ridge regression, and Linear regression. The accuracy provided by each suggested model is nearly equal. After experimenting with a few different computer programs, they were all able to estimate the rent amounts rather accurately. The most accurate of these, known as Ridge Regression, was accurate 91.54% of the time. The rest followed closely, scoring correctly between 91.49% and 91.52% of the time. So, these machine-learning regression algorithms are helpful in figuring out how much it should cost to rent a house in Dhaka. It's a big deal because Dhaka has too many people looking for houses, and this could help them find places they can afford.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
CHAPTER	
CHAPTER 1: INTRODUCTION	01-05
1.1 Introduction	01
1.2 Motivation	04
1.3 Rationale of the Study	04
1.4 Research Questions	04
1.5 Expected Output	05
1.6 Project Management and Finance	05
1.7 Report Layout	05
CHAPTER 2: BACKGROUND	06-07
2.1 Preliminaries/Terminologies	06
2.2 Related Works	06
2.3 Comparative Analysis and Summary	07
2.4 Scope of the Problem	07
2.5 Challenges	07

CHAPTER 3: RESEARCH METHODOLOGY	08-24
3.1 Research Subject and Instrumentation	08
3.2 Data Collection Procedure/Dataset Utilized	08
3.3 Statistical Analysis	09
3.4 Proposed Methodology/Applied Mechanism	23
3.5 Implementation Requirements	24
CHAPTER 4: EXPERIMENTAL RESULT & DISCUSSION	25-29
4.1 Experimental Setup	25
4.2 Experimental Results & Analysis	25
4.3 Discussion	29
CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY	30-33
5.1 Impact on Society	30
5.2 Impact on Environment	31
5.3 Ethical Aspects	31
5.4 Sustainability Plan	32
CHAPTER 6: SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE WORKS	34-35
6.1 Summary of the Study	34
6.2 Conclusions	34
6.3 Implication for Further Study	34
REFERENCES	36-37

LIST OF FIGURES:

FIGURES	PAGE NO.
Figure. 1.1: Dhaka South City Corporation	03
Figure 1.2: Dhaka North City Corporation	03
Figure 3.3: Snap of Heatmap	10
Figure 3.4: Mirpur Dhaka	11
Figure 3.5: Avenue 1 Mirpur Dhaka	11
Figure 3.6: Avenue 2 Mirpur Dhaka	12
Figure 3.7: Avenue 3 Mirpur Dhaka	12
Figure 3.8: Block C Section 13 Mirpur Dhaka	13
Figure 3.9: East Shewrapara Mirpur Dhaka	13
Figure 3.10: Mirpur DOHS Mirpur Dhaka	14
Figure 3.11: Pallabi Mirpur Dhaka	14
Figure 3.12: Senpara Parbata Section 10 Mirpur Dhaka	15
Figure 3.13: West Kazipara Mirpur Dhaka	15
Figure 3.14: West Shewrapara Mirpur Dhaka	16
Figure 3.15: Azimpur Lalbagh Dhaka	17
Figure 3.16: Banani Dhaka	17
Figure 3.17: Baridhara Dhaka	18
Figure 3.18: Cantonment Dhaka	18
Figure 3.19: Dhanmondi Dhaka	19
Figure 3.20: Farmgate Dhaka	19
Figure 3.21: Gulshan 1 Dhaka	20

Figure 3.22: Mohammadpur Dhaka	20
Figure 3.23: Uttara Dhaka	21
Figure. 4.1: Linear Regression Accuracy	26
Figure 4.2: Bayesian Regression Accuracy	26
Figure 4.3: Lasso Regression Accuracy	27
Figure 4.4: Ridge Regression Accuracy	27

LIST OF TABLES

TABLES	PAGE NO.
Table 3.1: Features Attribute	09
Table 3.2: Dataset Portion	09
Table 3.3: With Unnamed Column	22
Table 3.4: Without Unnamed Column	22
Table 3.5: Outlier of Dataset	22
Table 3.6: Removing comma in Dataset	23
Table 4.1: Model Accuracy	25
Table 4.2: Prediction on Original Rent and Predicted Rent	28
Table 4.3: Location with the most number of houses	29

CHAPTER 1

INTRODUCTION

1.1 Introduction

The amount of available urban land is rapidly decreasing as a result of population growth and urban migration. Due to this, many middle-class and lower-class families in developing nations are unable to pay the high cost of housing. Bangladesh's capital city of Dhaka is renowned for being one of the megacities with the densest populations on Earth. Because Dhaka is home to over 23 million people and is the fourth largest metropolis in the world, renting a property there can be difficult [1].

Living in Dhaka has become increasingly challenging due to the relentless rise in house rents day by day. Simultaneously, the city's population is swelling. In 2011, According to the Population and Housing Census, the Dhaka division was home to more than 12 million people. However, the recent census in 2023 revealed a staggering increase, with the current population standing at 44 million in the Dhaka division. The population of the city has increased by roughly 4.2% a year [2][18]. In 2024 it is also increasing. This substantial surge in population further intensifies the struggle for accommodation in the city.

The migration from rural areas is the main cause of the population explosion in Dhaka City. Rental prices for homes grow sharply as more people move in and demand for housing rise. The migration of people from rural areas increases the demand for housing, which puts a tremendous amount of strain on the supply of housing and ultimately raises rental costs.

The landlords frequently set their own rental costs in Dhaka city, and occasionally these prices rise sharply year after year, making renting difficult for residents. It is suggested that a standard or guideline for figuring out house rents be established in order to reduce this problem and provide more consistency.

So the main goal of our study is to forecast Dhaka city housing rents using regression approaches. These sophisticated mathematical methods, which include Bayesian Regression, Lasso Regression, Ridge Regression, and Linear Regression, are used to identify patterns in data and generate predictions. By using these methods, researchers hope to develop models that can more accurately estimate or predict housing rents, which might lead to a more uniform and equitable pricing strategy. This might reduce the aggressive conduct of certain property owners who raise rents quickly.

Achieving accurate house price predictions necessitated the adoption of effective methodologies and techniques. The application of efficient techniques was crucial to the prediction of property values. We used a dataset that we obtained from bProperty.com for this study, which included rental pricing and various house attributes in Dhaka, Bangladesh. For the purpose of making accurate rental property predictions, high-quality data is crucial. And we also collect data locally. However, machine learning struggles to perform as intended when information is missing. These systems have difficulty making accurate rent pricing guesses when there is missing data. Thus, having complete information is essential to improving the ability to forecast house rents.

In Dhaka, housing costs have been rapidly increasing lately for unclear reasons. In this neighborhood, the majority of middle-class families rent their houses, and many are having trouble keeping up with the ongoing rent increases. In Dhaka, about 15.5 lakh people are homeowners while more than 90% of people live in rentals. Middle-class residents make up the bulk of Dhaka's residential land (65%), with only 20% of households owning their own residences and 80% renting [3]. Landlords are boosting rents as a result of the city's growing population and growing demand for rental properties.

When tenants don't satisfy their expectations, landlords have even been known to force them to vacate. Establishing a standard for house rents could help level the playing field for tenants in order to address this. To estimate rent, my suggestion is to use a neural network model that is based on thirteen distinct features. Although this approach has certain drawbacks, it may eventually deprive landlords of the authority to choose rent rates.

To streamline administrative procedures, the Dhaka City Corporation was divided into the Dhaka North City Corporation (DNCC) and the South City Corporation of Dhaka (DSCC). In other words, Old Dhaka, Elephant Road, Segunbagicha, Lalbagh, Shahbag, Wari, Arambagh, Motijheel, Malibagh, Paltan, Shantinagar, Khilgaon, Basabo, Mugdapara, Fakirapul, Arambagh, Motijheel, Shahjahanpur, Malibagh, Paltan, Jatrabari, Shantinagar, Green Road, and Green Road are some of the principal areas of the capital under Dhaka South City Corporation's jurisdiction. [4].

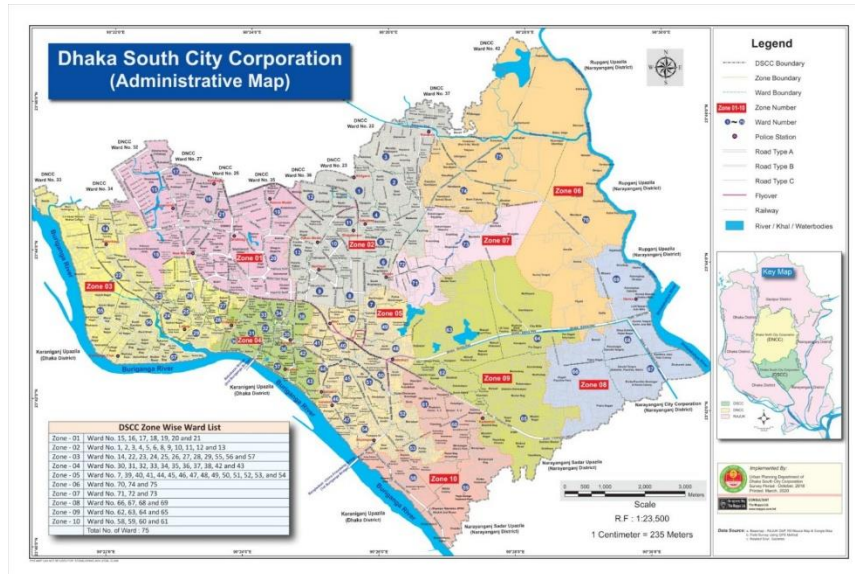


Figure 1.1: Dhaka South City Corporation

The independent Dhaka North City Corporation is in charge of overseeing Dhaka's 54 northern wards. Mirpur, West Uttara, East Uttara, Uttarkhan, Dakshin Khan, Biman Bandar, Khilkhet, Shahali, Vatara, Badda, Rampura, Tejgaon, Sher-E-Bangla nagar, Mohammadpur, Shilpanchal, Adabor, Darussalam, Turag, Pallabi, Rupnagar, Kafrul, Bhashantek, Cantonment, Banani, Hatirjheel, and Gulshan are among the other DNCC wards and areas included in this list. [5].

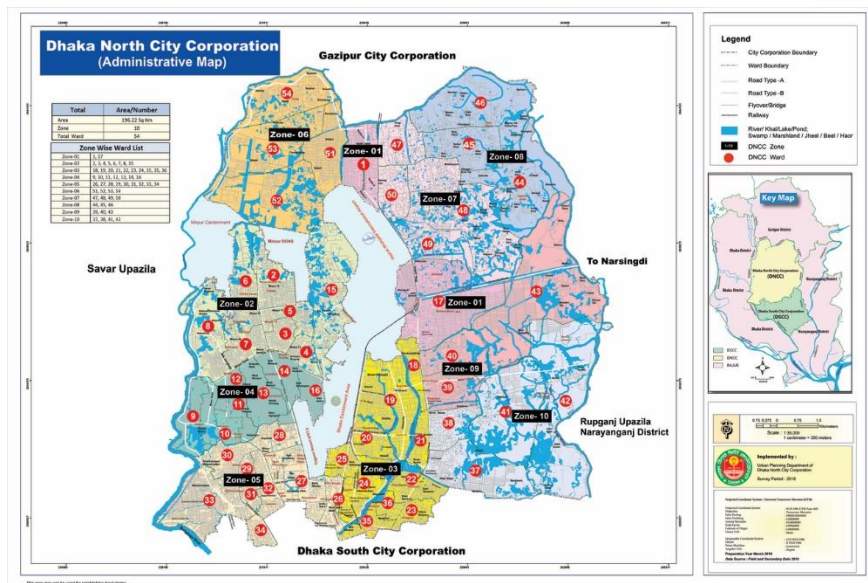


Figure 1.2: Dhaka North City Corporation

1.2 Motivation

- To create a model that can forecast the real rent for a home.
- Create this model so that individuals can look for renters who fit their ideal home.
- This model can raise awareness of rent for the home, preventing the owner from profit from excessive rental fees.
- The price of house rental varies depending on where in the city you live. People can locate specific locations they can afford by using this knowledge.

1.3 Rationale of the Study

In Dhaka city everyday lots of people come from various parts of Bangladesh. They come to Dhaka city for a better life. For a better income source. The main factor for coming to the city of Dhaka is poverty and employment scarcity, and the influence of family. And there are lots of job opportunities in Dhaka city rather than other cities in Bangladesh. So this factor attracts them to come to Dhaka city. This makes Dhaka city crowded and it increases the demand for renting houses. And this is the situation where landlords take advantage and propose their own desired rent and this is becoming problematic for middle class people. So our study model can forecast accurate house rents, giving people knowledge about fair renting pricing.

1.4 Research Questions

- Which variables have the most effects on home prices?
- Do machine learning models accurately estimate property prices based on provided features?
- Does the dataset show any noticeable patterns or trends that affect housing prices?
- Does the dataset contain any outliers, and if so, how do they impact the model's performance?
- What is the correlation between the features "Area," "Bed," and "Bath" and the goal variable "Price"?
- How can the created model be used in real estate valuation and investment decisions?
- Which features, in the model, are most important for predicting property prices?
- In terms of forecasting housing prices, how do various regression models compare?

1.5 Expected Output

The main objective of this study is to forecast house rents in Dhaka City. With this study, it finds a machine learning model that can give the best accuracy to forecast house rents. It can show the total number of houses in each area. And in this study, I find the best accuracy level with Ridge Regression. I was able to accomplish this goal and get an accuracy rate of 91.54%.

1.6 Project Management and Finance

I used a planned, systematic strategy to effectively handle a research task.

- This inquiry was conducted with a methodical approach to ensure effective growth.
- Planning and scheduling the tasks involved in data preprocessing, feature engineering, model training, evaluation, and deployment.
- Carried out data preprocessing, quality assurance, and collecting.
- Working on this research with my respected supervisors
- Task division, virtual meetings, and routine oversight all helped to speed the procedure.

1.7 Report Layout

There are five chapters in this report.

In Chapter 1, It presented the problem statement, motivation, and purpose of the housing rent situation in Dhaka city.

In Chapter 2, It highlighted a thorough background investigation along with an outline of the issue and its difficulties.

In Chapter 3, It provides detailed instructions on the research methodologies and data preprocessing, as well as how to develop each model and project a property's rent.

In Chapter 4, The discussion and explanation of the experimental results are given. The expected and actual rent for a residence are compared, and the accuracy of the model is also investigated.

In Chapter 5, It compares the benefits and effects of using machine learning for house rent forecasting, highlighting its significance in the field.

In Chapter 6, Discussion of the Conclusion.

CHAPTER 2

BACKGROUND

2.1 Preliminaries/Terminologies

Keywords: Machine Learning, Regression Model, Linear Regression, Ridge Regression, Bayesian Regression, Lasso Regression, Accuracy, Dataset, Feature Engineering

2.2 Related Works

There are a lot of factors responsible for high house rent in Dhaka city. The author of article [8] has investigated the possible influences on housing rent in Dhaka. The zonal variations of the building, the surrounding land, the distance from the road, the open space, the utility services, the construction, and the number of flats taken into consideration are some common variables they noted. They also stated that the home's owner raised the rent for their personal gain. Security and safety also claim that social standing and educational institutions are major factors in rising housing costs. People in Dhaka City are under a great deal of pressure every day. The apartment's rent was raised by the home's owner. In Bangladesh, middle-class individuals can begin their purchasing journey beyond the age of 40, as this is when they are either nearing retirement from their employment or have a stable position that allows them easy access to financing. The demand for urban housing is projected to be over 6 million units currently, and it is predicted that this figure would climb to 10.5 million units by 2030, making the shortage of affordable housing in Bangladesh a serious issue. This need is fueled by the fact that, by 2030, half of Bangladesh's population may reside in large cities, with an estimated 67 million people living in urban regions at present [9]. The writer of paper [10] collects 33224 housing rental data points for Chengdu in order to do visual analysis and forecasts based on an online housing platform. Initially, the distinctive factors such location, residential neighborhood, orientation, and method of collecting rent. The importance of the distinguishing characteristics of the rental data is taken into consideration while selecting transportation, building, etc are selected according to how important the distinctive qualities of the rental data are. The data visualization method shows that Chengdu renters are more interested in small-area and combined rental apartment types. It is becoming increasingly difficult for low-income groups worldwide to afford homes due to rising property values in major cities [11]. Rental housing can provide an affordable place to live for young people, newlywed households, and migrants. Renting has therefore become an essential component of owning a property. Worldwide, there are about 1.2 billion renters (Gilbert, 2016). More specifically, data indicates that in significant global cities including Shanghai, New York, Los Angeles, and Shenzhen, the rental rate is above 50% [11].

2.3 Comparative Analysis and Summary

The comparison analysis, which evaluates each machine learning algorithm's performance, forms the core of the thesis. Accuracy, computing efficiency, and interpretability are among the factors taken into account. The examination need to highlight the merits and demerits of every model, offering discernments into their suitability when predicting residential rental rates in Dhaka City.

2.4 Scope of the Problem

The goal of this study is to create a machine learning model that can forecast housing rents in Dhaka. In order to perform this prediction analysis, our research studies a variety of regression techniques, such as Bayesian, Ridge, Lasso, and Linear regression.

- Pandas eliminated the column that was unknown. Any column can be used with the `df.set_index()` function to set the dataframe's index.
- Category variables are strings that can be problematic for models. We create fictitious figures for these categories in order to fix this. These dummies enhance the performance of the models and provide more precise forecasts.
- Outlier values caused the scale of our graph to be incorrect. They were quite inconsistent, providing true readings up to 10% of the time and misleading readings at other times. To solve the issue, we eliminated these outliers.
- Prepare the data so that the next method can be used.
- A variety of regression techniques, including Lasso, Ridge, Linear, and Bayesian regression, have been employed.

2.5 Challenges

The main challenge in this study is that there is less related research on this topic. Thus, it was more difficult to gather information about house rents. And data collection is challenging for me because there are lots of houses in Dhaka City. And it's too hard to find every areas house rents data so easily.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

Research Subject:

This research subject is nothing but the forecasting of house rent prices in Dhaka City with machine learning. With this study, we are able to know that houses rent in different areas. We can learn about the various bed sizes that houses rent out.

Instrumentation:

We must first prepare the data in order to get better results. Next, we split it into two sections. A training section and a testing section. Next, we utilize the training data to construct the model and apply it to forecast results. We fixed unnamed columns, handled numerical variables, removed outliers, and removed commas from the obtained data to prepare it for analysis. Next, in order to create regression models, we divided the data into train and test sets. We examined the predictions made by the linear, ridge, lasso, and Bayesian regression models.

3.2 Data Collection Procedure/Dataset Utilized

Our dataset was collected from bproperty.com [12]. Bproperty is an all-inclusive virtual and physical platform that makes safe real estate transactions possible for Bangladeshi buyers, sellers, developers, tenants, and landlords. With the largest collection of property listings in the nation, it provides advice to both people and businesses, using its experience to help them make wise real estate decisions. And we also collected data from land lords. And on spot data collection.

Dataset Utilized

We are making sure the data is accurate and consistent when we analyze it. We are presenting the trustworthy and accurate data from our study. In addition to addressing typical problems such as outliers, missing data, altered data, data mining, and producing illuminating visual representations, we strive to reduce statistical errors.

3.3 Statistical Analysis

Dataset Features

There are a total of 5 attributes available in our dataset, including location, price, area, bed, and bath.

Features name	Features Description
Location	The house's location is specified in the location..
Area	The area gives us knowledge about the size of the full house.
Bed	Here we see the total number of bed in a house
Bath	Here we see the total number of bath in a house
Price	The Price class gives us knowledge about the price of the full house.

Table 3.1: Features Attribute

Dataset label and its properties

The class "Price" is crucial. Due to the fact that this class depends on independent attributes such as area, bed, bath, and location. Based on the independent qualities data, the pricing predicts various values.

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

Table 3.2: Dataset Portion

Dataset features correlation with the target column

It is essential to resolve any skewness and outliers in the price feature of the dataset prior to training. The main purpose of this stage is to reduce inaccuracy in models by removing pointless variables. It also makes it easier to improve accuracy. It also gives good results in feature engineering. To address skewed data, log transformation will be used, as shown in the figure, to address these problems.

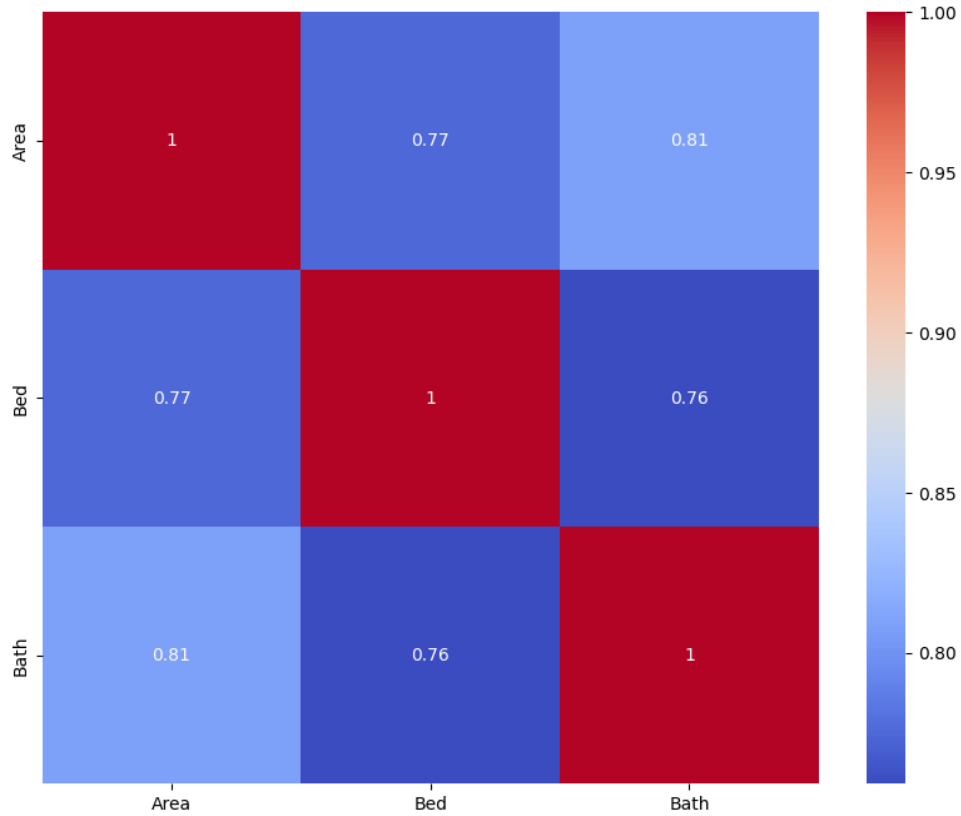


Figure 3.3: Snap of Heatmap

Dataset Insight

Below is a summary of every feature in the Scatterplot project along with how it is represented. We're displaying the actual house rent for different parts in Mirpur, specifically focusing on the amount of bedrooms.

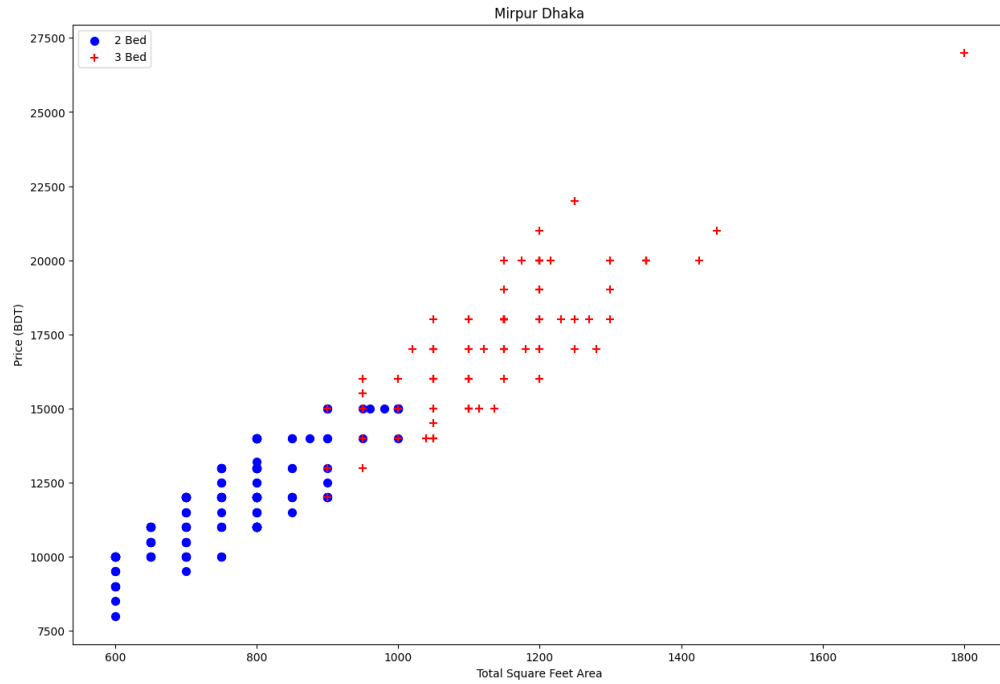


Figure 3.4: Mirpur Dhaka

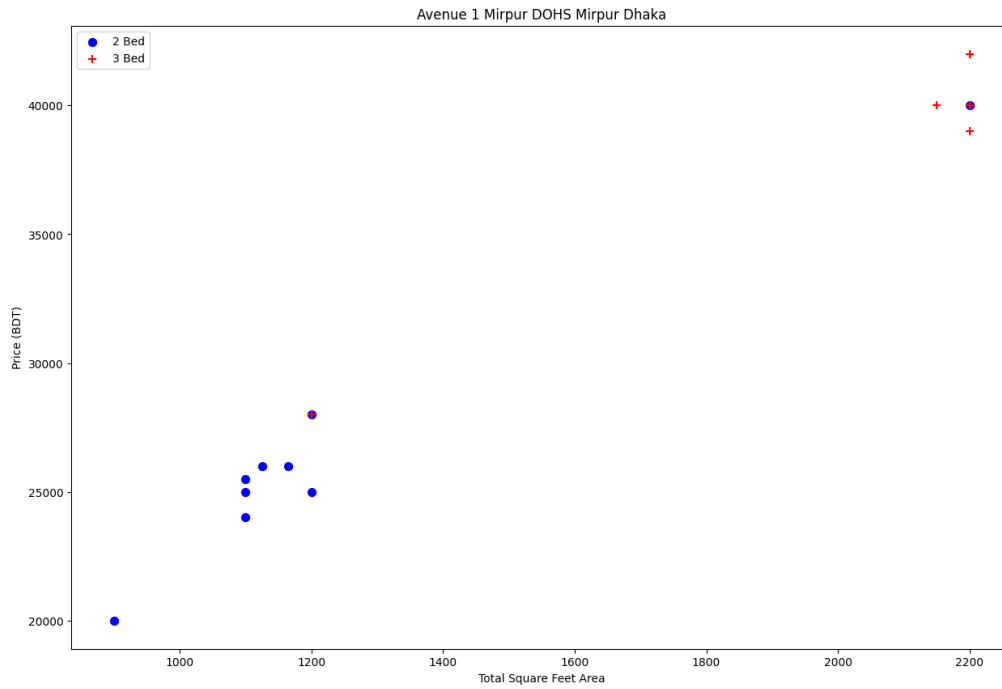


Figure 3.5: Avenue 1 Mirpur Dhaka

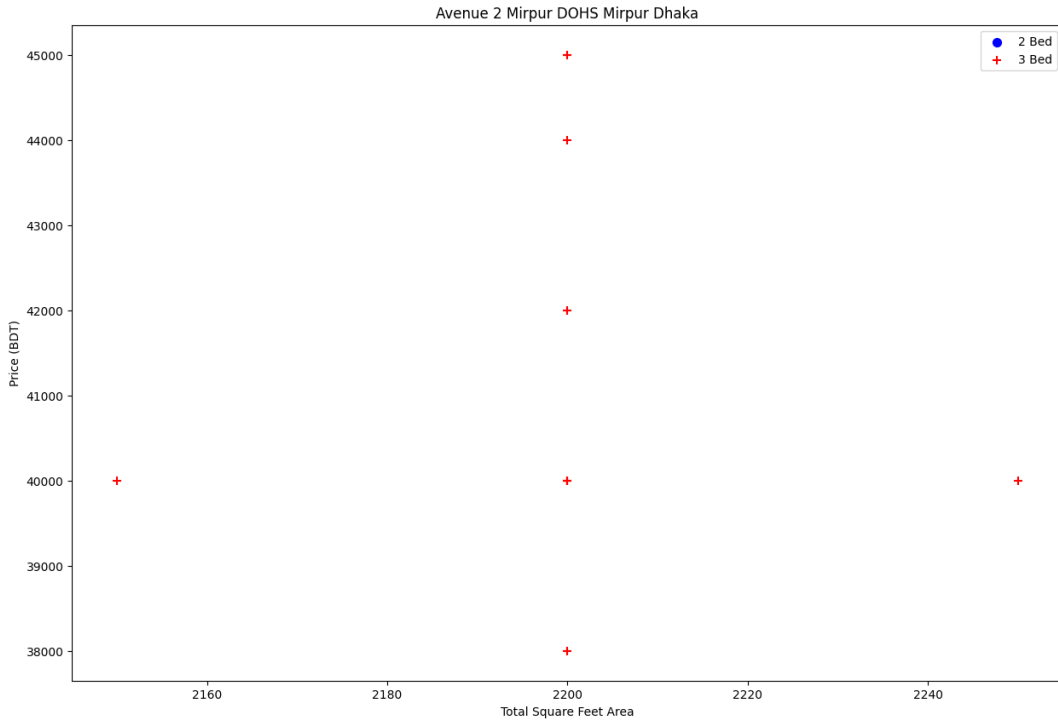


Figure 3.6: Avenue 2 Mirpur Dhaka

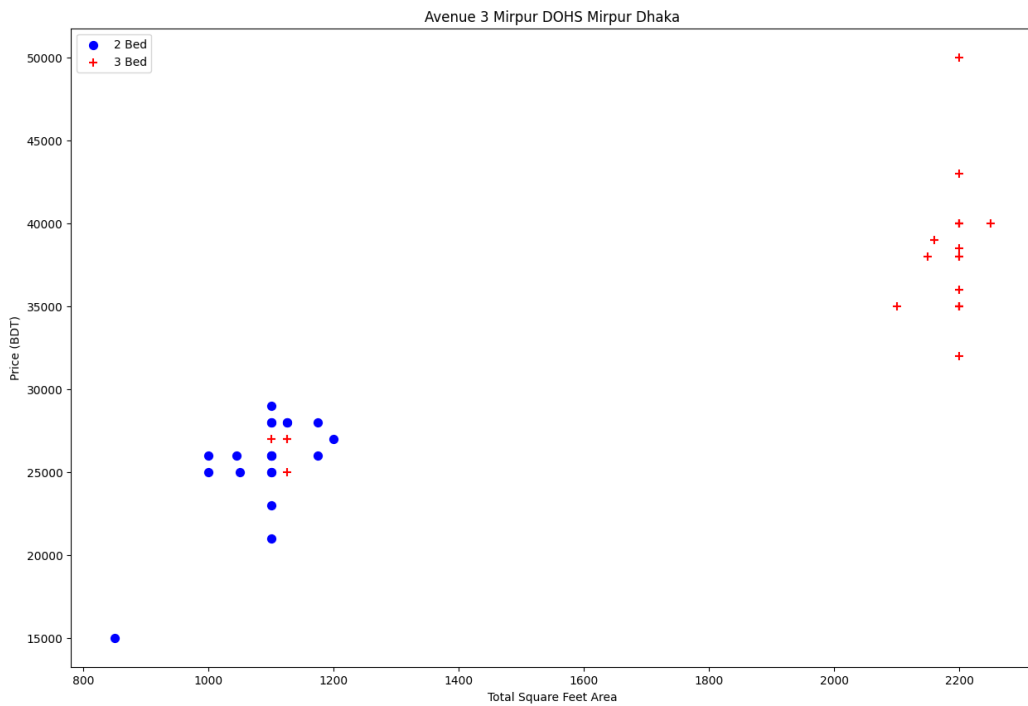


Figure 3.7: Avenue 3 Mirpur Dhaka

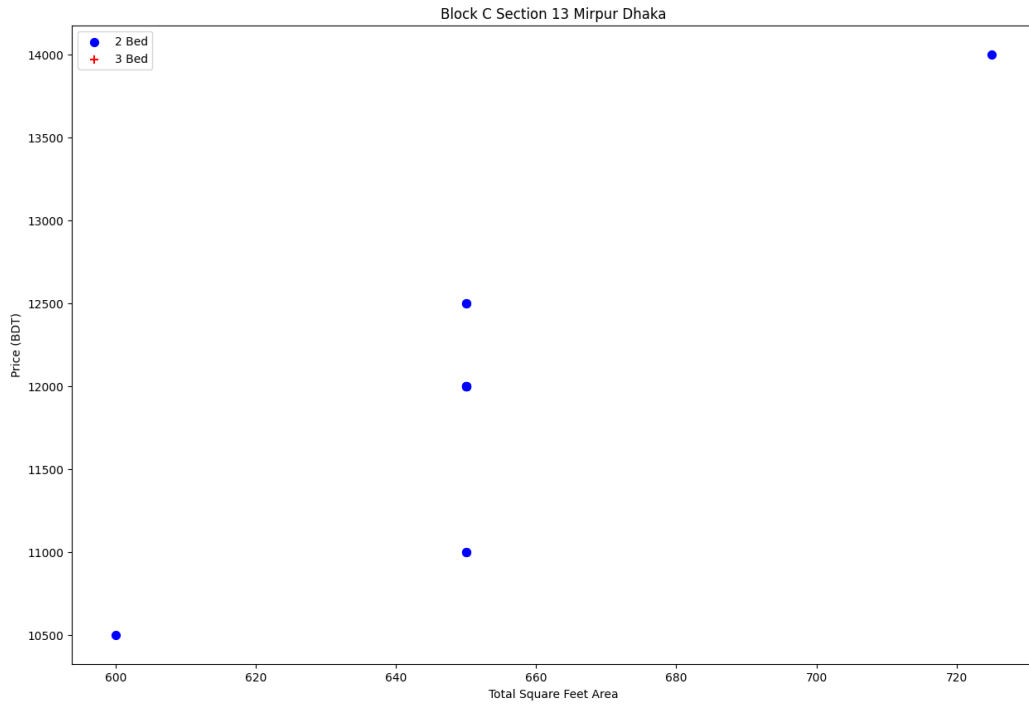


Figure 3.8: Block C Section 13 Mirpur Dhaka

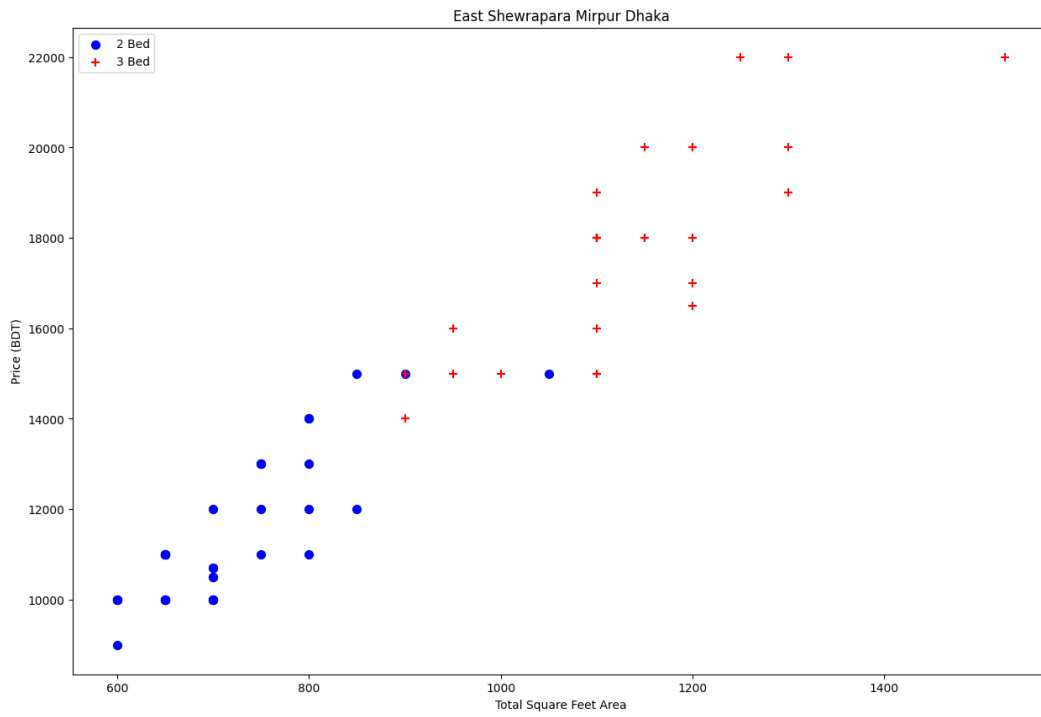


Figure 3.9: East Shewrapara Mirpur Dhaka

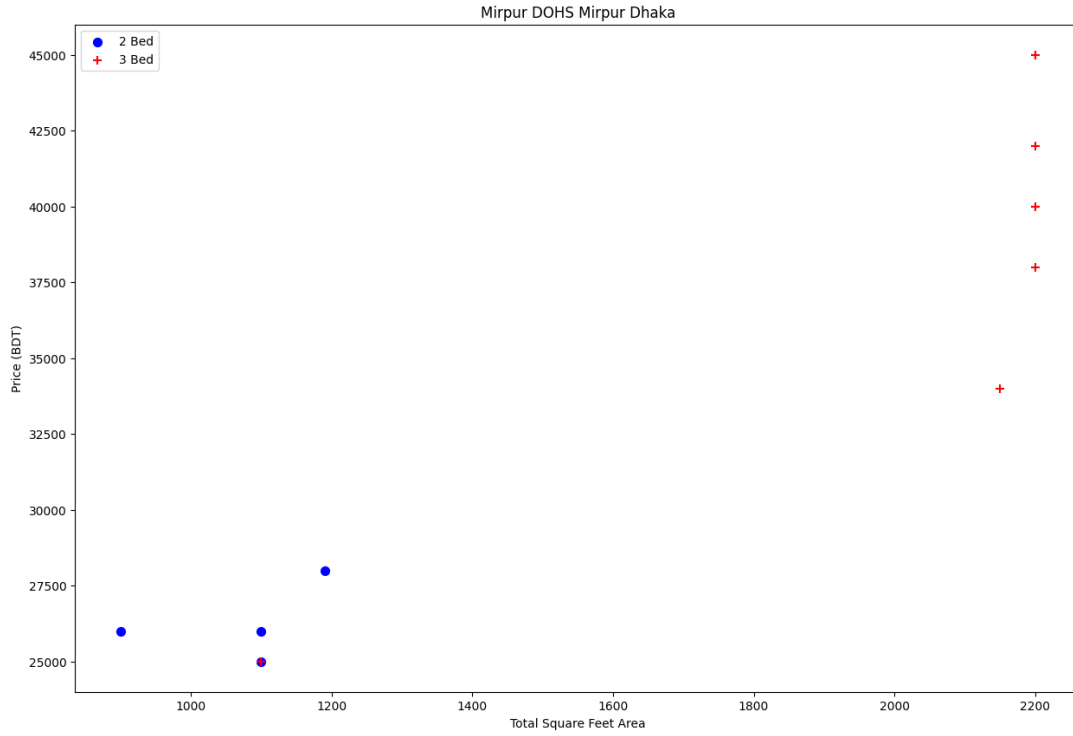


Figure 3.10: Mirpur DOHS Mirpur Dhaka

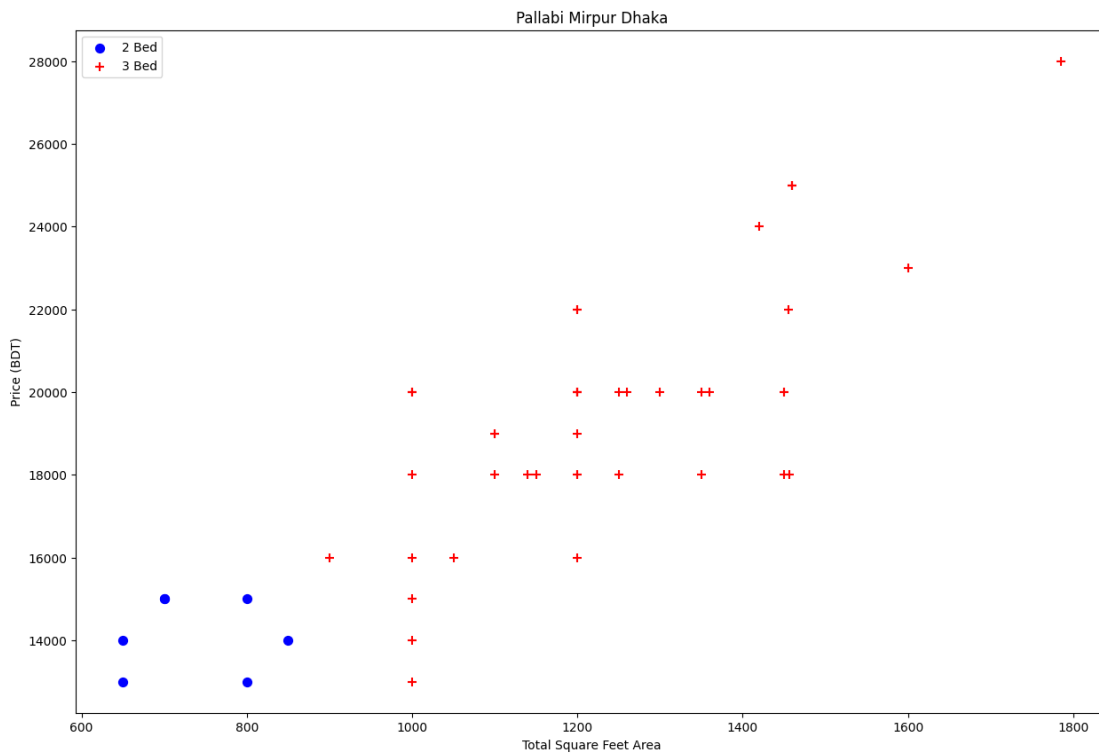


Figure 3.11: Pallabi Mirpur Dhaka

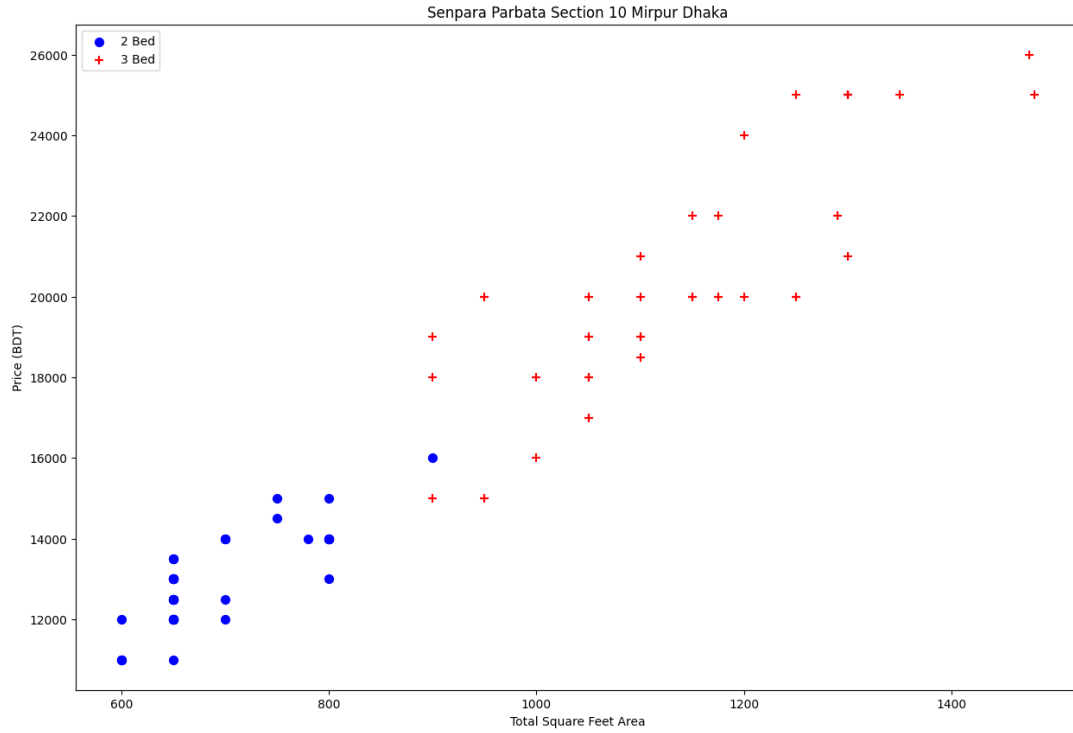


Figure 3.12: Senpara Parbata Section 10 Mirpur Dhaka

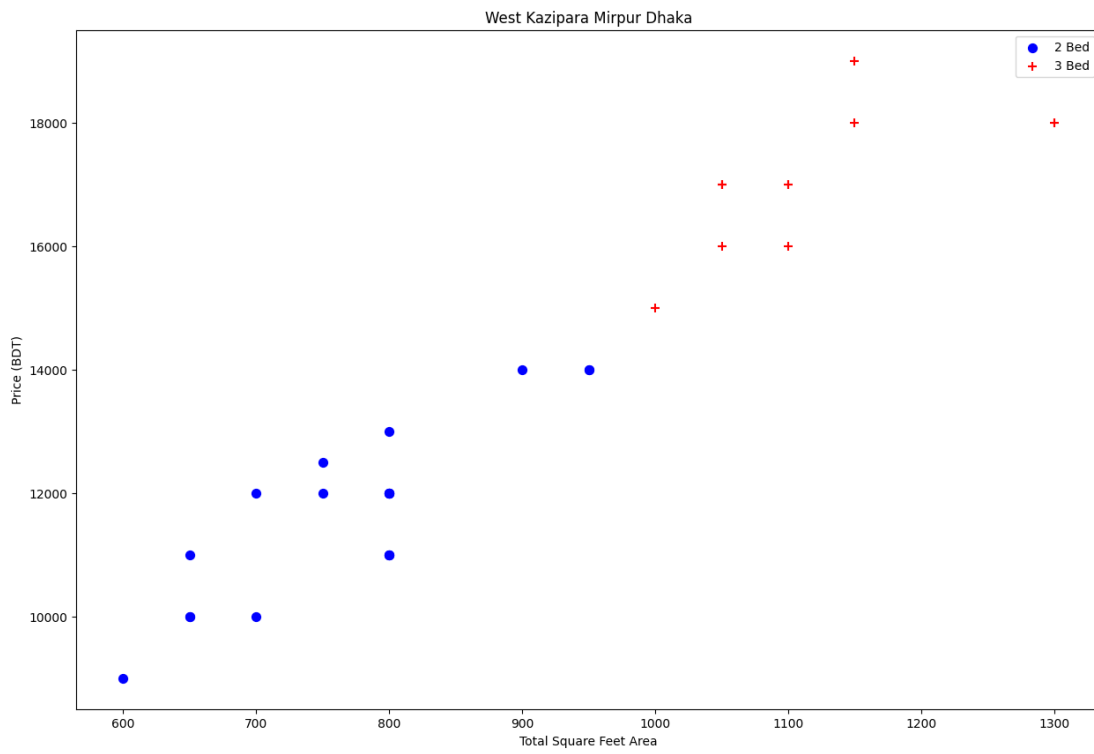


Figure 3.13: West Kazipara Mirpur Dhaka

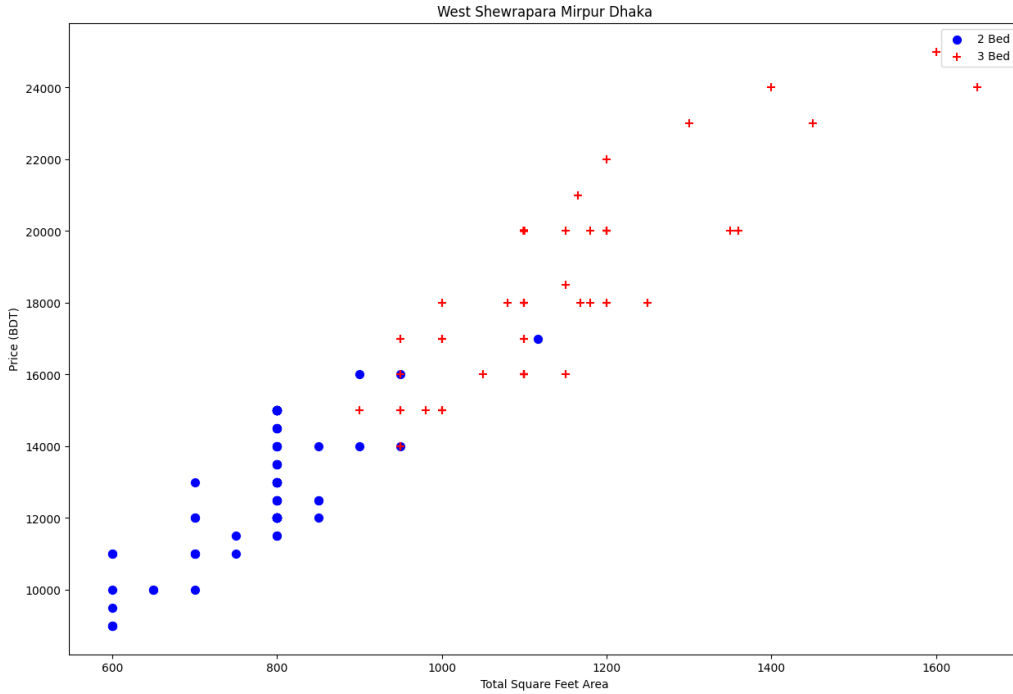


Figure 3.14: West Shewrapara Mirpur Dhaka

From the scatter chart of Mirpur Dhaka, we get that there are 2 bed houses and 3 bed houses available. The rent for two-bedroom houses ranges between 7500 and 15000 tk per month. We also see that there are 3 bed houses in 1000 sq ft. and this ends at 1800 sq ft. and the price range for 3-bed houses is 16500 to 27500 tk. Now we are going to take a look at another Mirpur area, which is East Shewrapara. We see that here two types of houses are available. 2 beds and 3 beds. Rent for 600sqft-900sqft 2bedroom houses ranges from 10000tk to 15000tk. and the 3-bedroom houses start at 900 sq ft and end at 1500 sq ft. And the rent starts at 14000 tk to 22000 tk. In East Shewrapara, a 1400-square-foot house rent is 22,000 tk. We found an interesting fact in our study: there are no 3-bed houses in Block C, Section 13, Mirpur Dhaka. And the house sq ft starts at 600 720, and the rent starts at 10500 tk and ends at 14000 tk for two-bedroom houses. The rent for a 1000sqft house in West Kazipara is 15000 tk, and the rent for an 800sqft house is 12000 tk. Now we see the Pallabi Mirpur Dhaka area. Here, the most number of houses is 3 beds. and here the 1200-square-foot house rent is 16000 tk, 17000 tk, 18000 tk, and 20000 tk. Now if we take a look at Mirpur DOHS, we see that in this area, 1100 sq ft of houses rent start at 26000 tk. And 2200 sq ft of houses rent start at 40000 tk and end at 45000 tk. Which contain three beds. So we can say that this is the most costly area in Mirpur, Dhaka.

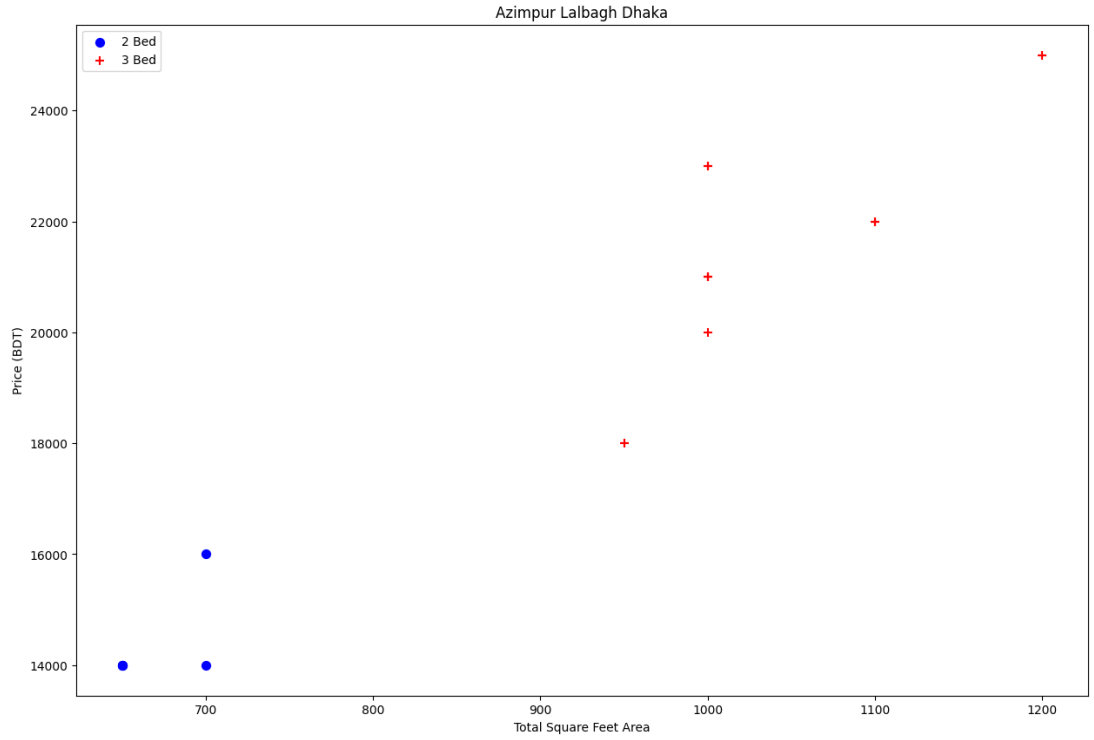


Figure 3.15: Azimpur Lalbagh Dhaka

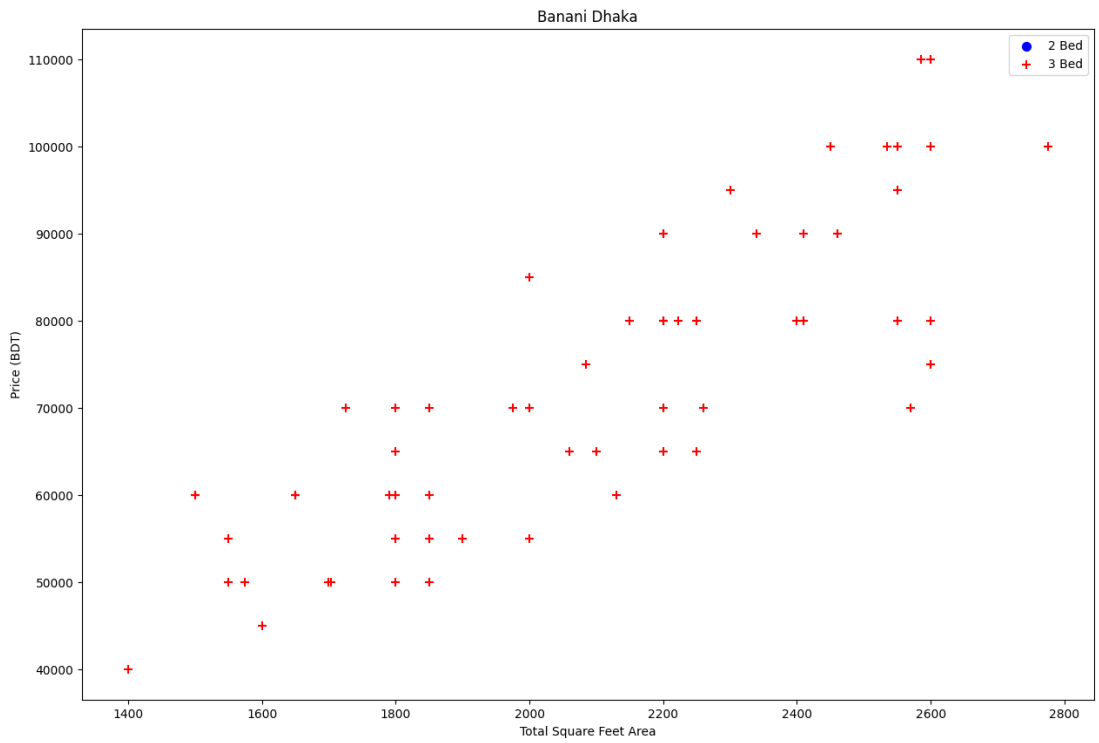


Figure 3.16: Banani Dhaka

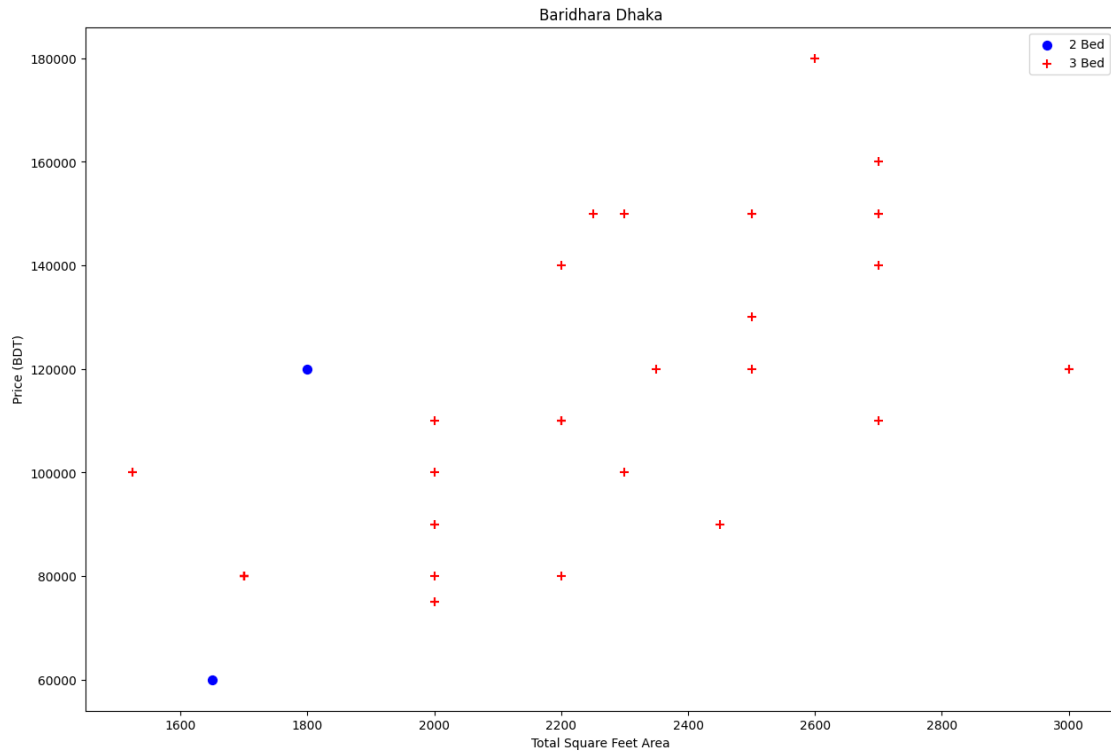


Figure 3.17: Baridhara Dhaka

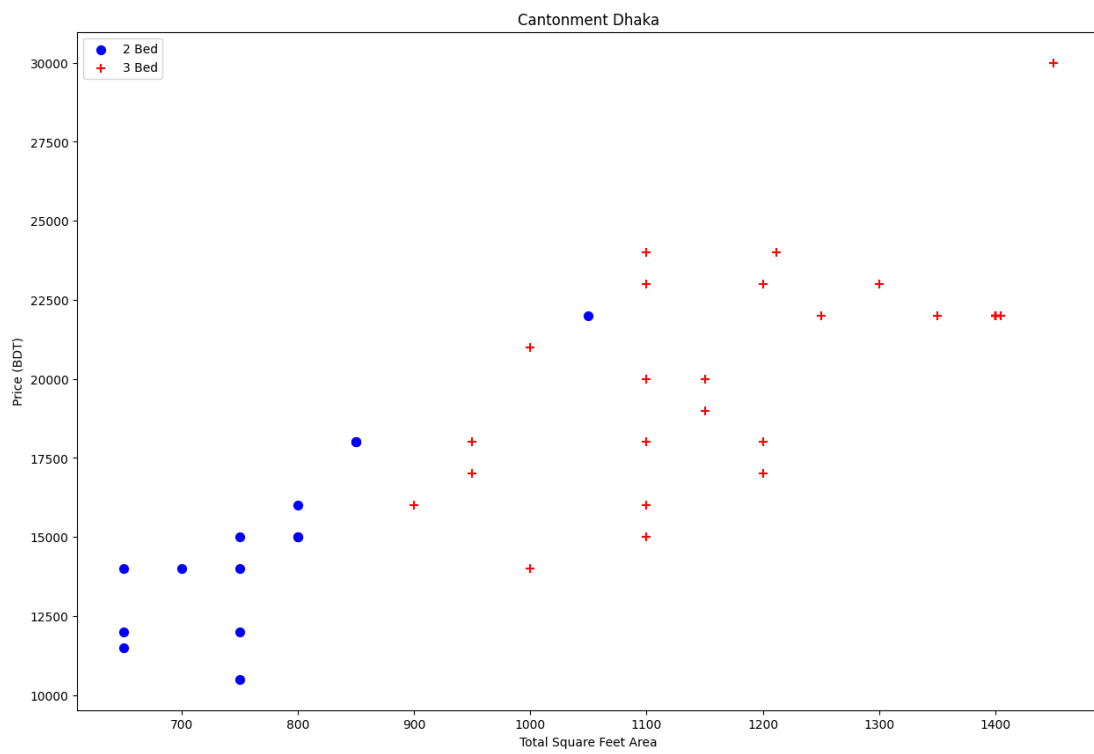


Figure 3.18: Cantonment Dhaka

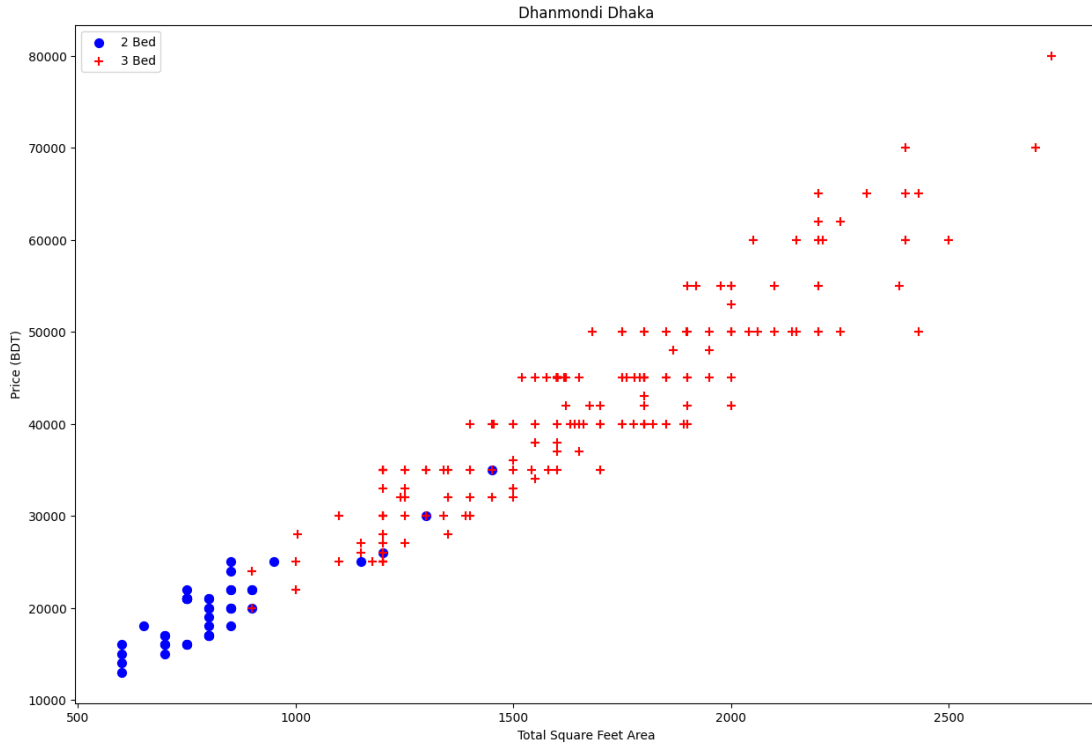


Figure 3.19: Dhanmondi Dhaka

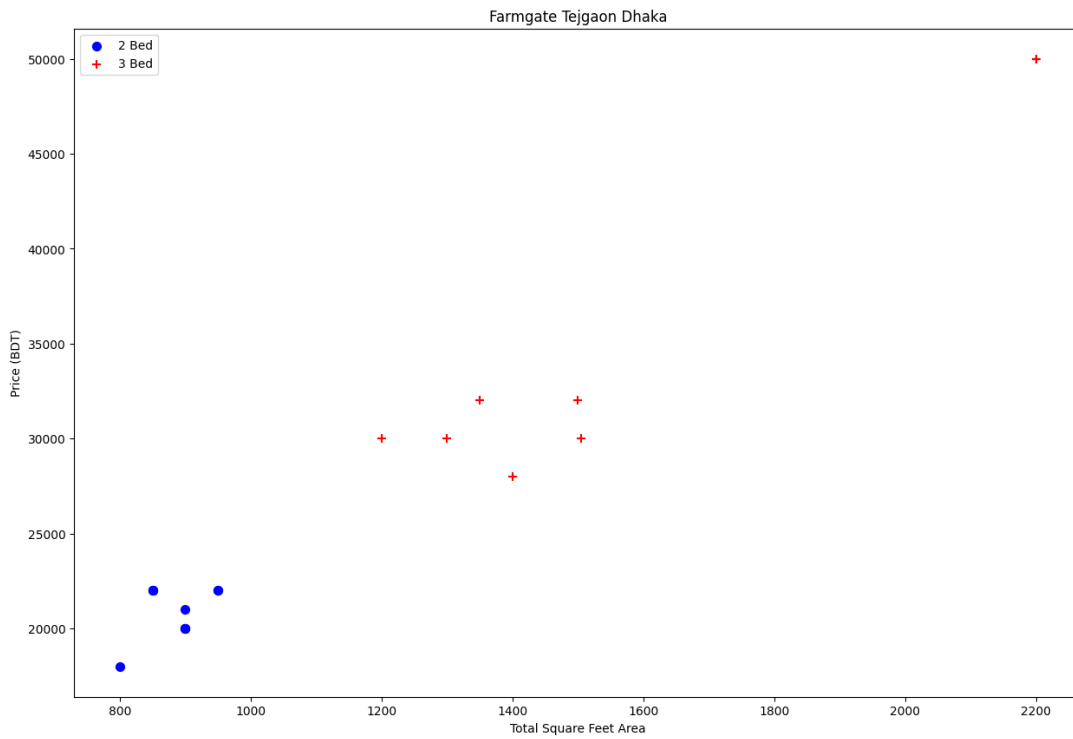


Figure 3.20: Farmgate Dhaka

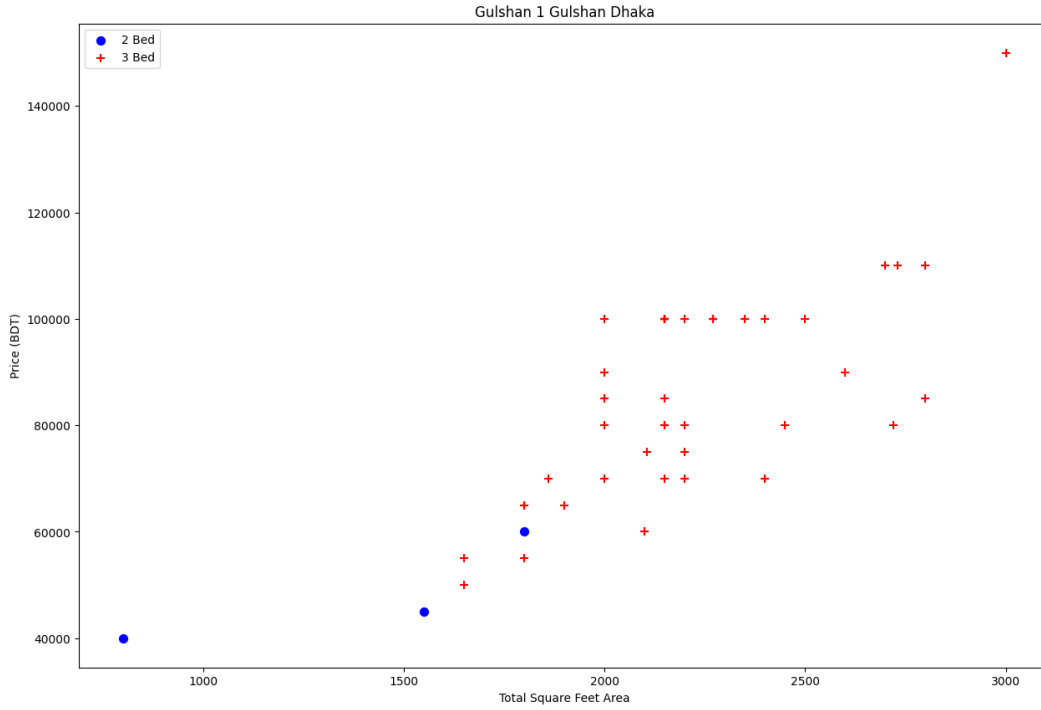


Figure 3.21: Gulshan 1 Dhaka

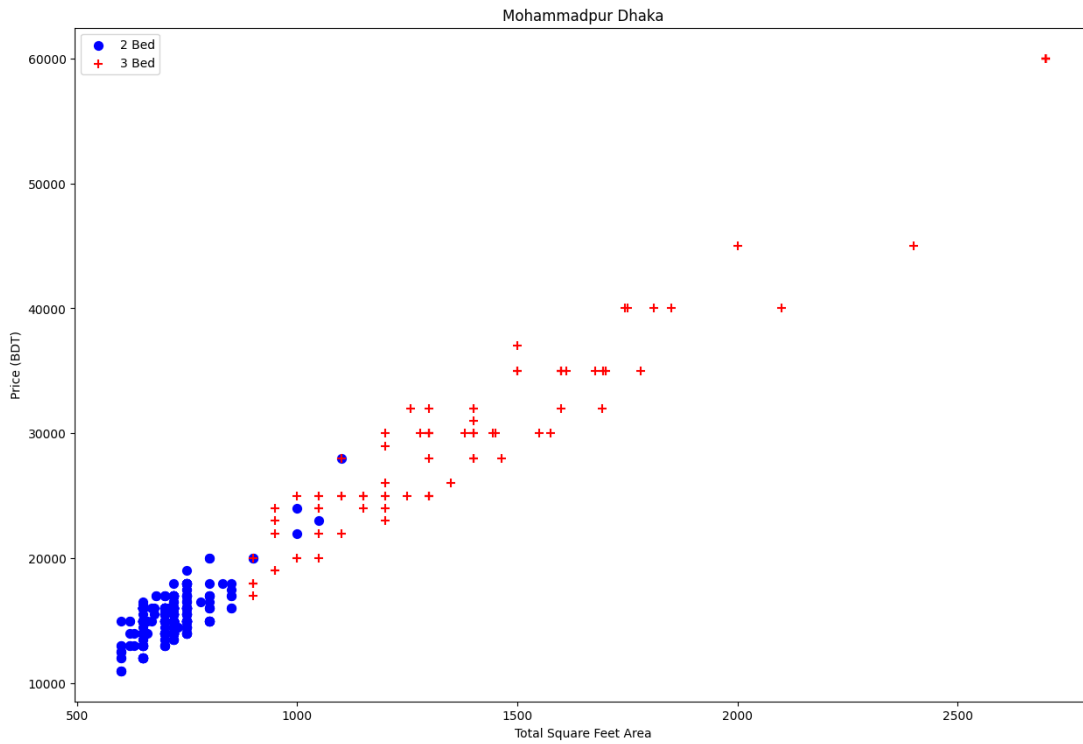


Figure 3.22: Mohammadpur Dhaka

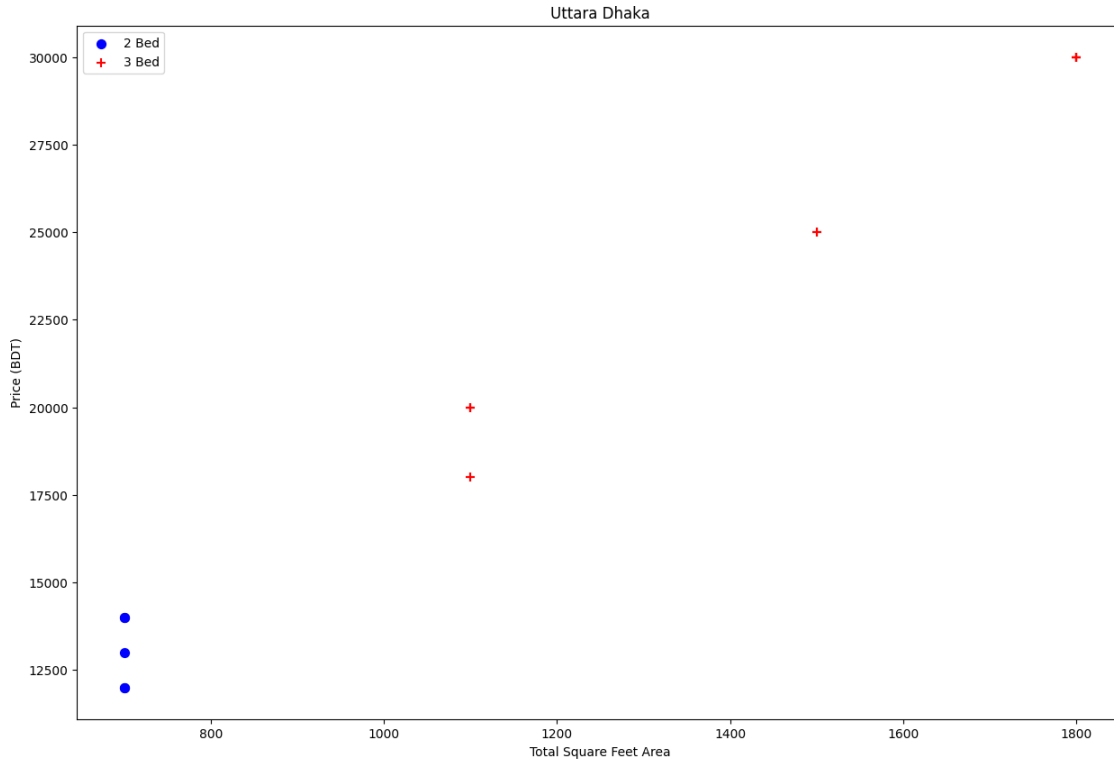


Figure 3.23: Uttara Dhaka

Feature Engineering

In this study we performed a number of feature engineering techniques. In order to accommodate both regularized linear regression models and multiple linear regression models (OLS), we transform the unnamed columns into dummy variables. In this study we ensured the limited use of unnamed columns for better performance. Introduced a new variable with enhanced usefulness by utilizing expert subject knowledge. Deleted extraneous punctuation and dealt with dataset abnormalities in order to enhance the data.

Data Preprocessing

This study's model is evaluated and trained using information from dataset. Among the data preparation steps are:

Removing the unnamed column

The unnamed column will be immediately dropped by pandas. To set the index for the data frame, call the `df.set_index()` function on any 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

Table 3.3: With 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

Table 3.4: Without Unnamed Column

Removing outlier

Our graph became inconsistent as a result of outlier values affecting its scale. The problem was caused by different numbers of outliers. Up to 10% of the data had misleading results, but the remaining measurements were accurate. We eliminated these outliers from our analysis to assure data accuracy.

		Location	Area	Bed	Bath	Price
0		Block H, Bashundhara R-A, Dhaka	1600	3	3	20 Thousand
1		Farmgate, Tejgaon, Dhaka	900	2	2	20 Thousand
2		Block B, Nobodoy Housing Society, Mohammadpur,...	1250	3	3	18 Thousand
3		Gulshan 1, Gulshan, Dhaka	2200	3	4	75 Thousand
4		Baridhara, Dhaka	2200	3	3	75 Thousand

Table 3.5: Outlier of Dataset

Remove commas from the area Column

Using the Find and Replace tool, we can locate and eliminate each and every comma from a cell. This method is limited to textual data.

	Location	Area	Bed	Bath	Price	price_per_sqft
0	Block H Bashundhara R-A Dhaka	1600	3	3	20000.0	12.500000
1	Farmgate Tejgaon Dhaka	900	2	2	20000.0	22.222222
2	Block B Nobodoy Housing Society Mohammadpur Dhaka	1250	3	3	18000.0	14.400000
3	Gulshan 1 Gulshan Dhaka	2200	3	4	75000.0	34.090909
4	Baridhara Dhaka	2200	3	3	75000.0	34.090909

Table 3.6: Removing comma in Dataset

Dataset Split

After we initially process the data, we split our dataset into training data and test data.

3.4 Proposed Methodology/Applied Mechanism

Model

The following regression techniques are used in this study: Bayesian, Ridge, Lasso, and linear regression.

Linear Regression

When there are two continuous things the Linear regression helps us understand the connection between them. One is used to predict, and the other responds. It shows a probable connection but doesn't guarantee an exact one. If one thing perfectly explains the other, then their relationship is considered more definite [13].

$$Y = \theta_1 * X + \theta_0$$

Ridge Regression

Ridge regression improves data overlapping. To address this problem, L2 regularization is used. Regular least-squares projections may be inaccurate when multicollinearity is present, displaying values that are distant from reality. Regression on the ridge helps to correct this [14].

$$RSS = \sum (y_i - \beta_0 - \sum \beta_j x_{ij})^2$$

Bayesian Regression

The mean of one parameter in Bayesian linear regression is represented by the weighted sum of the other variables. By determining the previous distribution of the regressors and other variables that describe the distribution of the regressor and, this kind of conditional modeling finally enables the regressor to be forecasted outside of the sample based on observations of the regression coefficients. Using Bayesian Regression can be quite beneficial when the dataset has insufficient or poorly distributed data [15] [16].

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

Lasso Regression

In statistics and machine learning, LASSO regression, also known as L1 regularization, is a method for predicting and identifying associations. It stands for Least Absolute Selection and Shrinkage Operator. The goal of LASSO regression is to understand model accuracy with simplicity. It achieves this by modifying the standard linear regression and applying a penalty that sets some values to zero, resulting in models that are more simple and accurate [17].

$$\sum (y_i - \sum_{j=1}^n x_{ij} \beta_j)^2 + \lambda \sum |\beta_j|$$

3.5 Implementation Requirements

There are certain conditions in order to put the model into practice.

- Ensure that an environment for Python is setup.
- Install the required libraries, including seaborn, scikit-learn, pandas, numpy, matplotlib, and any other necessary packages.
- Load the dataset into a pandas DataFrame.
- Handle outliers, fill in missing numbers, and prepare the data..
- Use one-hot encoding for variables that are categorical.
- Generate visuals (e.g. scatter plots, histograms) to gain early insights.
- Develop and execute code to train different regression models.

CHAPTER 4

EXPERIMENTAL RESULT AND DISCUSSION

4.1 Experimental Setup

We trained the dataset and generated independent predictions using the Linear, Ridge, LASSO, and Bayesian regression models.

4.2 Experimental Results & Analysis

Now let's examine our model's results, which are shown in the table below.

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

Table 4.1: Model Accuracy

After implementing all of our models we find the result. With the help of Linear Regression we get 91.49% accuracy. The accuracy provided by the Ridge regression model is 91.54%. From The Bayesian Regression and Lasso Regression we get 91.52% and 91.49% accuracy. After implementing all the models we see that the ridge regression model gives us the best outcome. So this gives a more accurate prediction than other regression models.

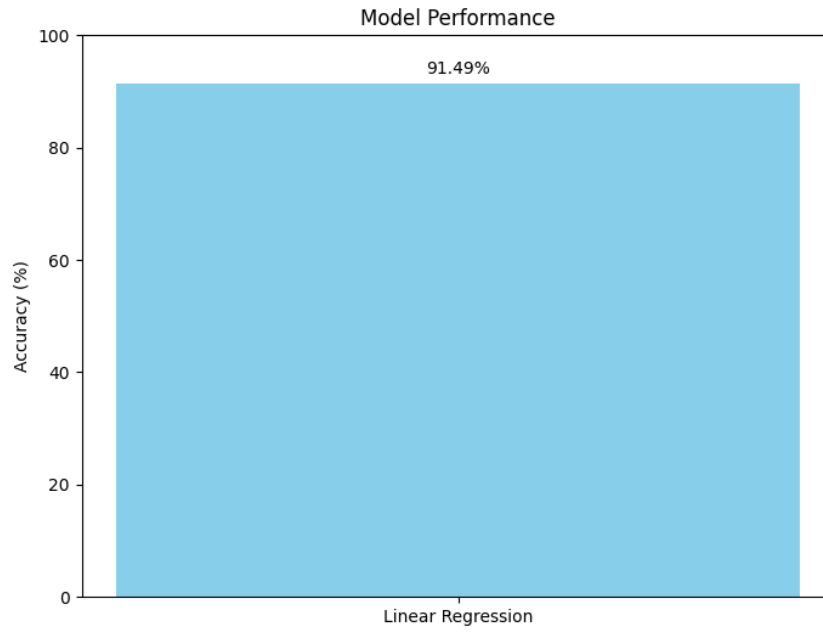


Figure 4.1: Linear Regression Accuracy

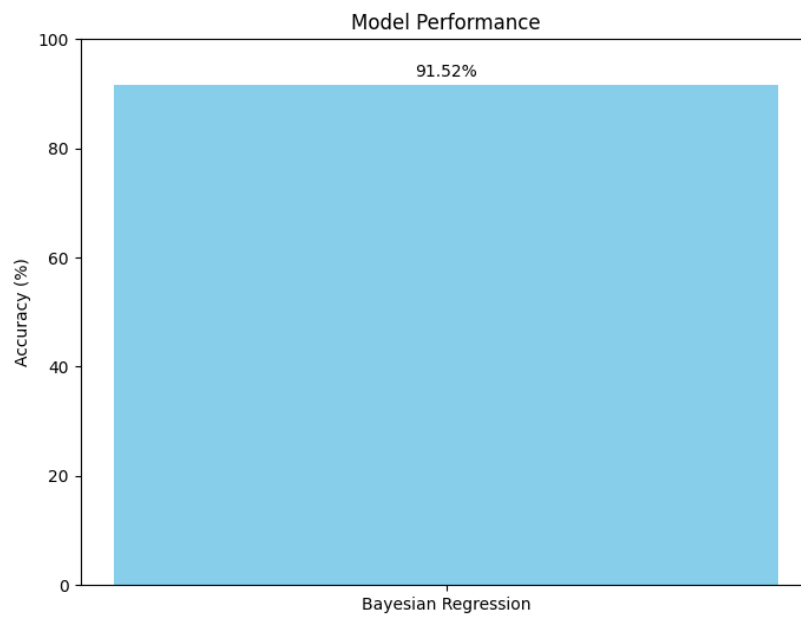


Figure 4.2: Bayesian Regression Accuracy

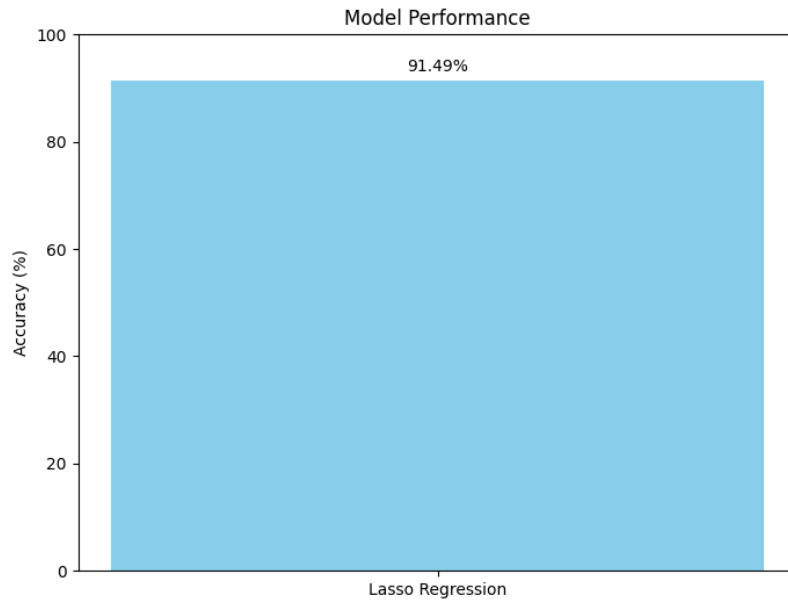


Figure 4.3: Lasso Regression Accuracy

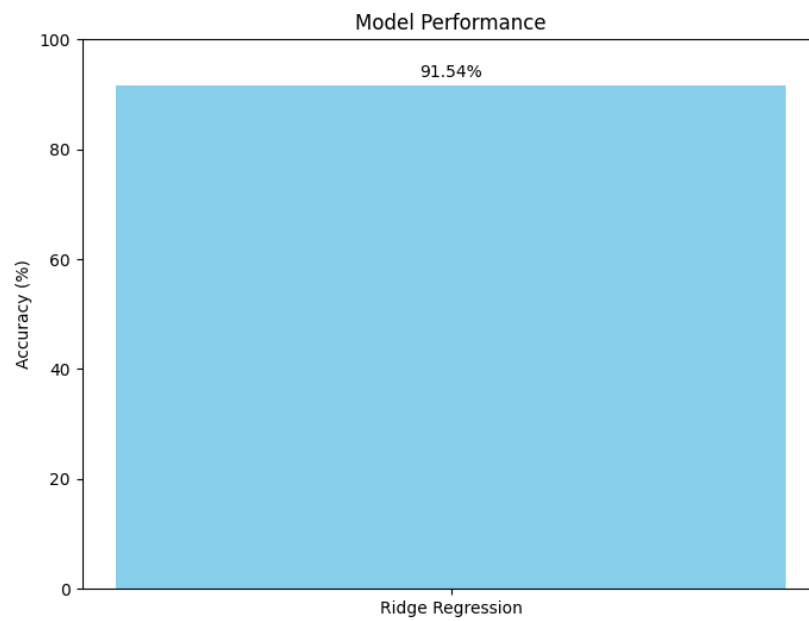


Figure 4.4: Ridge Regression Accuracy

Compare between prediction rent and original rent

And in this part the original house rent and the anticipated data are briefly compared.

Location	Area(sq ft)	Bed	Bath	Original Rent	Predicted Rent
East Shewrapara, Mirpur, Dhaka	1,300 sqft	3	3	22,000	23,165
West Shewrapara, Mirpur, Dhaka	1,150 sqft	3	3	16,000	18,587
Uttar Badda, Badda, Dhaka	880 sqft	2	2	13,000	18,512
Gulshan 1, Gulshan, Dhaka	2,200 sqft	3	4	75,000	96,266
Gulshan 21, Gulshan, Dhaka	1,600 sqft	3	3	50,000	91,276
Dhanmondi, Dhaka	2,660 sqft	3	4	85,000	71,819
Ibrahimpur, Dhaka	1,050 sqft	3	3	25,000	15,000

Table 4.2: Prediction on Original Rent and Predicted Rent

We have forecasted the data for the real dwelling rent in this table. We took into account the house's location, overall size, total number of bedrooms, and total number of bathrooms in order to estimate the rent. This house rent depends on the location of the bedroom bathroom. As per the study, we see that the residential areas have a higher rent than the non-residential areas. There are so many advantages in those areas.

Most numbers of Houses

Now with this study it finds total number of houses in each areas. Here is a list of the top ten locations with the most residences.

Location	Number of House
Dhanmondi	414
Mirpur	556
Shyamoli	285
Mohammadpur	757

Kallyanpur Mirpur	357
West Shewrapara Mirpur	292
PC Culture Housing Mohammadpur	274
Hazaribag Dhaka	260
Section 1 Mirpur Dhaka	241
Sector 13 Uttara	250

Table 4.3: Location with the most number of houses

Our dataset has 730 unique locations in total. The top 10 places with the most houses are displayed in Table.

4.3 Discussion

In this study, we used a regression model to forecast the house rent in Dhaka City. We use this to forecast something. And our study topic is to forecast house rent. That's why this machine learning model can fit perfectly in this scenario. The impact of various factors in predicting the target variable can be determined with the use of regression models. The research utilizes a variety of multiple regression models, such as Lasso, Bayesian, Ridge, and linear regression. This makes it possible to compare various algorithms and determine which one works best for the particular dataset. And in this study, ridge regression gives the highest accuracy of forecasting. The capacity of a model to produce precise predictions on unseen data is what determines its accuracy. Ridge regression may have produced the best accuracy in this particular code for a number of reasons. By penalizing large coefficients, the regularization term introduced by Ridge Regression helps prevent overfitting. When dealing with multicollinearity—highly correlated independent variables—this regularization can be quite helpful. Ridge regression can produce more reliable and broadly applicable predictions by penalizing big coefficients. And in our dataset, there are lots of these kinds of data. So Ridge regression gets perfect accuracy. Also it processes the dataset using feature engineering to produce an outstanding outcome. And this provides us with the most accurate result for Dhaka City's house rent forecast.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

The house price forecasting research may have a wide range of social effects. The possible effects that this initiative might have on society are listed below:

Landlords Making Well-Informed Decisions: Giving landlords a strong tool to use when deciding how much to charge for rentals is one of our project's main social effects. Landlords may create a more transparent and healthy rental market by setting competitive and fair pricing by accurately anticipating house rentals based on a variety of parameters.

Providing Knowledge to Residents: Residents are equipped with useful knowledge about the variables affecting house rents thanks to our machine learning algorithms. By having a better awareness of the factors that influence rental pricing, potential tenants are better equipped to make educated judgments about where to live. The landlord-tenant relationship becomes more egalitarian as a result of this empowerment.

Taking Care of Affordability of Housing: The knowledge gathered from this study may be useful in talks on how affordable housing is in Dhaka City. Policymakers and urban planners can address housing affordability issues and carry out focused interventions by having a thorough understanding of the major elements influencing rental pricing.

Transparency in the Real Estate Market: The total openness of Dhaka City's real estate market is improved by the transparency brought forth by this machine learning algorithms. Real estate agents, property developers, and prospective investors are among the many stakeholders who gain from this enhanced openness, which also promotes a more reliable and effective real estate ecosystem.

Minimizing Asymmetry in Information: This solution provides precise and data-driven insights to landlords and tenants, thereby mitigating information asymmetry in the housing market. A more equitable and competitive rental market where prices more closely reflect the true value of rental properties is one result of reducing information asymmetry.

Ethics-Related Considerations: Our dedication to responsible and ethical research processes is demonstrated by our acknowledgment and resolution of ethical considerations, such as potential social repercussions or bias in the data. This helps ensure the responsible deployment of machine learning technology in society.

5.2 Impact on Environment

The following are a few possible environmental factors to take into account for the project:

Urban Development That Is Sustainable: This project supports sustainable urban growth by offering precise rental estimates. An educated and open rental market facilitates data-driven decision-making about the development of housing infrastructure, which may result in more effective land use and less urban sprawl.

Optimizing Communication: Precise estimations of property values have the potential to impact residential preferences and, in turn, travel habits. People may be able to choose homes nearer to their places of employment if they have a better understanding of how much housing will cost in various locations, which would lower the overall environmental impact of commuting.

Encouragement of Eco-Friendly Communities: The knowledge gathered from this study can help landlords and real estate developers make decisions that are friendly to the environment. The establishment and upkeep of eco-friendly neighborhoods may be stimulated if rental markets place a premium on neighborhoods with sustainable practices or eco-friendly facilities.

Urban Planning and Green Spaces: Clear rental markets could help improve the choices made about landscaping in urban development. The rental market's appreciation of specific districts with lots of green space may incentivize city planners to give preservation and development of these kinds of areas top priority, thereby fostering environmental well-being and biodiversity.

Responsible Resource Distribution: This project's transparency helps ensure that resources are allocated responsibly in urban development. City planners can more effectively manage resources by analyzing housing patterns and demand, reducing the environmental impact of needless construction and infrastructure expansion.

5.3 Ethical Aspects

In every study, ethical considerations are essential, but they become even more important when forecasting property prices. The following are some crucial moral factors to take into account:

Algorithmic Fairness: Make sure that errors that might disadvantage specific demographics are avoided in the design and training of machine learning models used to predict housing values.

Data protection: To secure people's personal information, respect privacy standards and follow to data protection laws.

Anonymization: To avoid unwanted access or identification, make sure that each person's identity in the dataset is sufficiently anonymized.

Utilizing Predictive Models Responsibly: We acknowledge that our predictive algorithms have the ability to influence real-world housing decisions. Our models are intended to support decision-making, and users are urged to apply discretion and morality when acting.

Model Transparency: Attempt to analyze and clarify the models' forecast processes. Openness fosters confidence and makes it possible for interested parties to understand the factors influencing housing price forecasts.

5.4 Sustainability Plan

A sustainability plan created specifically for the house price forecast project is shown below:

Implement Continuous Model Monitoring: Monitor model performance and forecasts often to discover accuracy pass or decline. To maintain the model's efficacy and relevance, update it frequently with fresh data.

Ethical Standards and Tests: Establish ethical rules for the project, including concepts of justice, openness, and user privacy. Conduct audits to ensure consistency. Identify and resolve any potential biases or ethical issues that may develop over time by conducting routine ethical audits.

Data Management: To ensure the security, quality, and accuracy of the dataset used for training and prediction, establish strong data governance procedures. Create protocols for managing sensitive data while following by data protection laws.

Information Transfer and Documentation: To help collaborators share information, keep thorough documentation of the codebase, model architecture, and data preparation procedures. Conduct regular knowledge transfer sessions to ensure that all the collaborators are familiar with the project's details.

Scalability Considerations: Plan the project to be scalable in order to handle growth in user interactions and data volume in the future. Make sure that computing and infrastructure resources can handle the project's expansion on a regular basis.

Adaptability in Model Architecture: To ensure flexibility in model architecture, make it flexible and adaptive to changes in input features or machine learning techniques. Research and add new characteristics or techniques to improve the accuracy and interpretability of the model.

The house price forecast project's sustainability plan takes a multidirectional approach, addressing ethical, technical, and community-focused elements. The project's ethical considerations, community involvement, and regular reviews are crucial to ensuring its long-term success and appropriate use.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE WORKS

6.1 Summary of the Study

Administrative delays in securing planning licenses have resulted in rising rental costs and mismatches between supply and demand, which have negatively impacted the housing market. Despite their noble objectives, the national housing policy and building restrictions seem to be impeding innovation and cost-effective development [3].

The model predicts the rental price based on different factors, such as the number of bedrooms and bathrooms. The size of the house and the location. The number of bedrooms and bathrooms affects the price, and it also affects the price based on location. We see that in our study.

There is a strong demand for homes, but there are significant problems with access to mortgage financing and the availability of inexpensive solutions. In order to alleviate the lack of affordable housing, the government has launched programs like the Ashrayan Project; nonetheless, other measures may be required to solve these issues and improve finance accessibility [9].

6.2 Conclusion

This study introduces a model that can forecast The price range for a house rent in Dhaka with the help of advanced regression techniques. This model seeks to quickly assess property data and recommend appropriate pricing according to different dwelling attributes. We find the best accuracy from the Ridge regression model, which is 91.54%. Then came the lasso, Bayesian, and linear models, all of which displayed excellent and comparable accuracy. And with this model it can forecast the rent of house in Dhaka City. It can find the total number of unique houses in a areas. And with this it gives us a clear idea about house rental situation in our Dhaka City.

6.2 Implication for Further Study

With this study the Implication for Further Study is to build a web application so that all the people can easily get informaton about the house rent in Dhaka City. Here an user can easily find his desired house with his affordable rent. And in future examine more specific location-based elements that affect home values.Perform a time-series analysis to see how housing trends change over time. Examine the use of modern machine learning methods. Make an effort to improve the models' interpretability so that forecasts may be explained

more clearly, building user confidence and understanding. Work together with local communities to gather specific information about public services, facilities, and community projects. Examine how community-driven data might increase the precision of predictions.

REFERENCES:

- [1] World City Populations 2024, worldpopulationreview.com/world-cities. Accessed 6 Jan. 2024.
- [2] Census 2022: Dhaka Division Home to 44 Million People Now - Dhaka Tribune, www.dhakatribune.com/bangladesh/dhaka/291165/census-2022-dhaka-division-home-to-44-million. Accessed 2023.
- [3] Al Amin, Abdullah & Islam, Syed. (2023). Exploring Middle-Income Multi-Ownership Housing Affordability in Dhaka: A Case Study on Badda. 3. 12-21. 10.5281/zenodo.10016498.
- [4] Harun, and Md Rubel Sarker. "Dhaka South City Corporation Ward List (with Map)." Bproperty, 5 Dec. 2022, www.bproperty.com/blog/dhaka-south-city-corporation-ward-list/. Accessed 06 Dec. 2023.
- [5] Nafim, et al. "The Wards and Areas of DNCC Everyone Needs to Know." Bproperty, 12 Dec. 2021, www.bproperty.com/blog/wards-areas-dncc/. Accessed 06 Dec. 2023.
- [6] ঢাকা দক্ষিণ সিটি কর্পোরেশন, dsc.gov.bd/. Accessed 06 Dec. 2023.
- [7] ঢাকা উত্তর সিটি কর্পোরেশন, www.dncc.gov.bd/. Accessed 06 Dec. 2023.
- [8] Akter, Taslima, Mehedi Mahmudul Hasan, Akter Uz Zaman, Md. Rifat Hossain and Meher Nigar Neema. "Analysis of Potential Factors Bringing Disparity in House Rent of Dhaka City." Communication Systems and Applications (2013).
- [9] The Housing Market in Bangladesh in 2022 - Bproperty.Com, www.bproperty.com/blog/housing-market-in-bangladesh/. Accessed 06 Dec. 2023.
- [10] Yue Ming, Jie Zhang, Jiaming Qi, Tian Liao, Maolin Wang, and Lingli Zhang. 2020. Prediction and Analysis of Chengdu Housing Rent Based on XGBoost Algorithm. In Proceedings of the 3rd International Conference on Big Data Technologies (ICBDT '20). Association for Computing Machinery, New York, NY, USA, 1–5. <https://doi.org/10.1145/3422713.3422720>
- [11] Hu, Lirong, 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, 2019, pp. 657–673, doi:10.1016/j.landusepol.2018.12.030.
- [12] "Property & Real Estate for Sale and for Rent in Bangladesh." Bproperty.Com Bangladesh Property Portal, www.bproperty.com/. Accessed 06 Dec. 2023.
- [13] "Linear Regression in Machine Learning." GeeksforGeeks, 21 Dec. 2023, www.geeksforgeeks.org/ml-linear-regression/. Accessed 06 Dec. 2023.
- [14] Team, Great Learning. "What Is Ridge Regression?" Great Learning Blog: Free Resources What Matters to Shape Your Career!, 16 Nov. 2022, www.mygreatlearning.com/blog/what-is-ridge-regression/. Accessed 06 Dec. 2023.
- [15] "Introduction to Bayesian Linear Regression: Simplilearn." Simplilearn.Com, Simplilearn, 20 Feb. 2023, www.simplilearn.com/tutorials/data-science-tutorial/bayesian-linear-regression#:~:text=In%20contrast%20to%20conventional%20regression,variance%20and%20mean%20are%20normalized). Accessed 08 Dec. 2023.

[16] Koehrsen, Will. "Introduction to Bayesian Linear Regression." Medium, Towards Data Science, 20 Apr. 2018, towardsdatascience.com/introduction-to-bayesian-linear-regression-e66e60791ea7. Accessed 07 Dec. 2023.

[17] Team, Great Learning. "A Complete Understanding of Lasso Regression." Great Learning Blog: Free Resources What Matters to Shape Your Career!, 30 May 2023, www.mygreatlearning.com/blog/understanding-of-lasso-regression/#:~:text=LASSO%20regression%2C%20also%20known%20as,Absolute%20Shrinkage%20and%20Selection%20Operator. Accessed 07 Dec. 2023.

[18] "Dhaka, Bangladesh Metro Area Population 1950-2024." MacroTrends, www.macrotrends.net/cities/20119/dhaka/population. Accessed 01 Dec. 2023.

FORECASTING HOUSE RENTS IN DHAKA CITY WITH MACHINE LEARNING

ORIGINALITY REPORT

8%	7%	1%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
2	Submitted to Daffodil International University Student Paper	<1%
3	www.bproperty.com Internet Source	<1%
4	Submitted to Brunel University Student Paper	<1%
5	"The Security of Water, Food, Energy and Liveability of Cities", Springer Science and Business Media LLC, 2014 Publication	<1%
6	ocw.un-ihe.org Internet Source	<1%
7	Submitted to University of Technology, Sydney Student Paper	<1%
8	Submitted to Dundalk Institute of Technology Student Paper	<1%