# Bangladesh Army University Of Science And Technology, BAUST

# DHAKA HOME RENT PRICE PREDICTION APP

| | |
|---|---|
| Department | CSE |
| Course Code | CSE 4252 |
| Course Title | Data Ware-housing and Data Mining Sessional |
| Course Credit | 3 |

# TABLE OF CONTENT


- Introduction.
- Project Summary.
- How It works?
- Architecture Diagram.
- Hardware & Software specification.
- Pictures of the project’s practical demo.
- Future Work.

# INTRODUCTION

**Problems faced during renting a house:**

- It’s a stressfull thing.
- Tenant are generally not aware of factors that influence the house rent.
- Hence real estate agents are trusted with the communication between Tenant & Landlord , This just create a middle man and increases the cost.

# Project Summary

- Our project is a machine learning app, based on certain specifications of your future home it will try to guess the most accurate price.

- Information such as Location, Bedroom Number, Bathroom Number, Land Size etc.

# How It Works

- Collecting Data: First step was to collect data. We collected data from https://www.bproperty.com .
- Then preprocess the collecting data.
- After preprocessing the data we trained the model using machine learning algorithm which in this case is linear regression.
- Based on the generated graphs we predict the cost of the house rent.

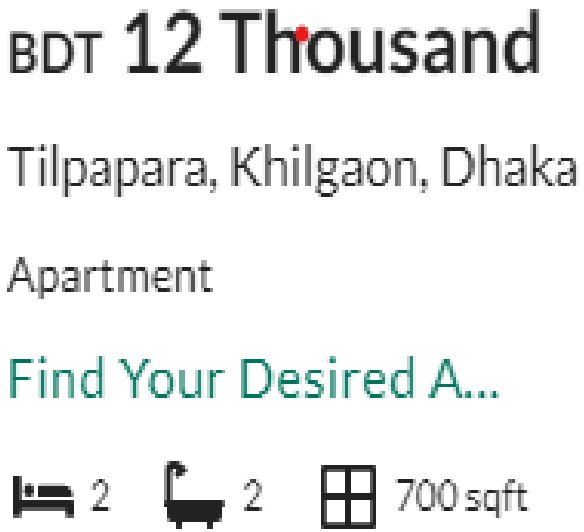
# Architecture Diagram

# Specification

| Hardware | Software | Libraries |
|---|---|---|
| Laptop | Framework –Flask<br>Design- Html, CSS | Numpy,Pandas,Seaborn,Matplotlib |
| Ram – Min 4 GB | Virtual Environment - Anaconda Navigator | BeautifulSoup,request,csv |
| HDD – Min 500 GB | Language – Python, Javascript | Scikit-learn |
| Internet – wifi, broadband | IDE - Spyder, Jupyter Notebook, VS Code | Pickle, Json |

# Collecting Data

- URL: https://www.bproperty.com/en/dhaka/properties-for-rent




```
<span class="f343d9ce" aria-label="Price">12 Thousand</span>
```

```
<div class="d6e81fd0">
  <div class="_6d2ea5a7"></div>
  <div class="cd6d5974 d8b3c34d">…</div> flex
  <div class="_4b74b8bb">
    <div class="_7afabd84" aria-label="Location">Tilpapara,
    Khilgaon, Dhaka</div>
    <div class="_9a4e3964" aria-label="Type">Apartment</div>
    <h2 class="_7f17f34f" aria-label="Title">Find Your Desired
    Apartment At This Ready 700 Sq Ft Flat For Rent At Khilgaon
    </h2>
```

**Save Those data in a datahouse.csv file in write mode.**

# Data Preprocessing

```
df = pd.read_csv('datahouse29.csv')
```

```
df.shape
```

```
(28683, 5)
```

```
df2 = df.drop_duplicates()
```

```
df2.shape
```

```
(15547, 5)
```

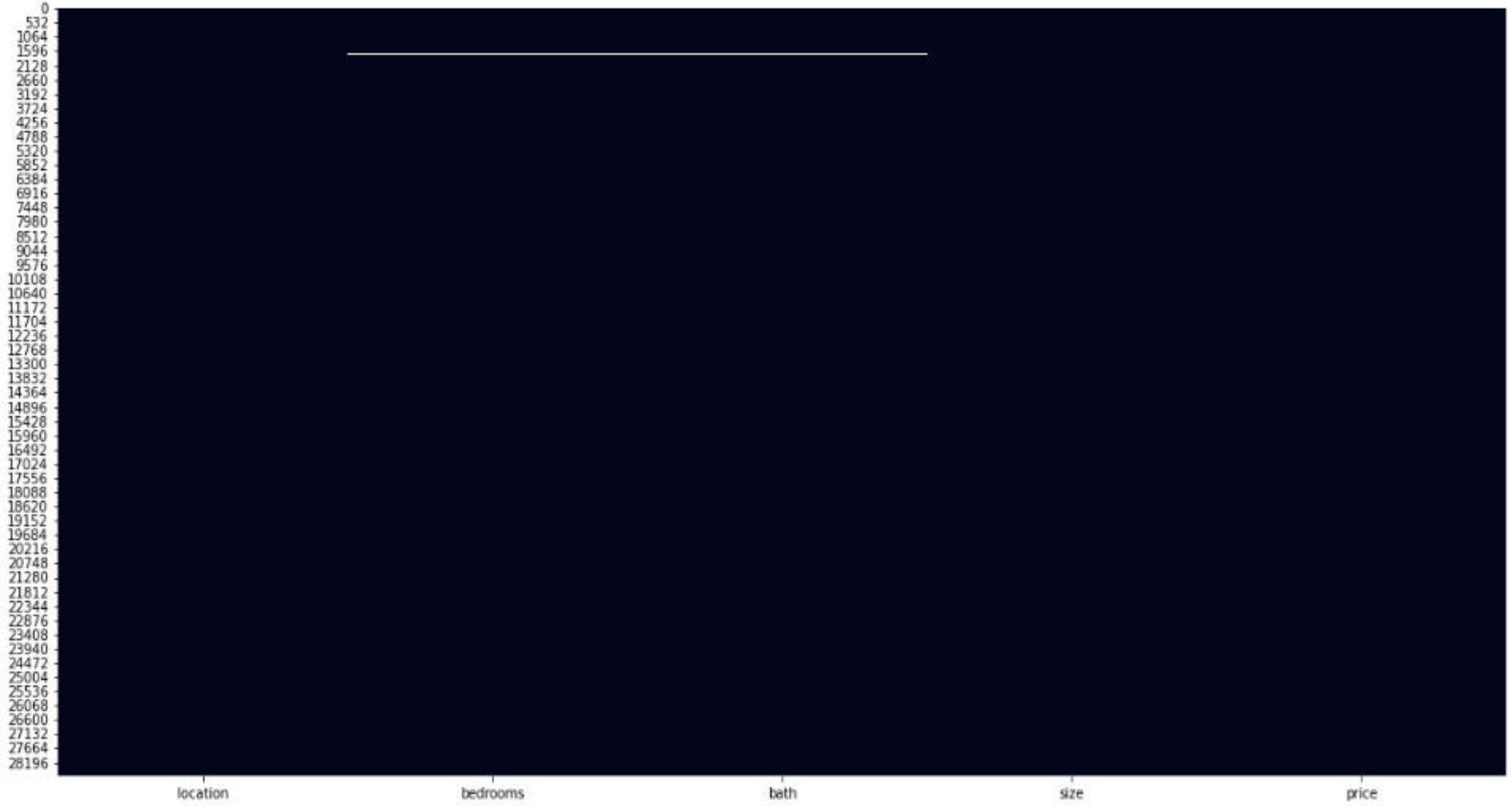
```
df.describe()
```

| | bedrooms | bath | size | price |
|---|---|---|---|---|
| count | 28682.000000 | 28682.000000 | 28683.000000 | 28683.000000 |
| mean | 2.520187 | 2.372638 | 1067.924868 | 19.898163 |
| std | 0.781235 | 0.781200 | 830.246843 | 23.865889 |
| min | 1.000000 | 1.000000 | 200.000000 | 1.100000 |
| 25% | 2.000000 | 2.000000 | 700.000000 | 13.000000 |
| 50% | 2.000000 | 2.000000 | 900.000000 | 16.000000 |
| 75% | 3.000000 | 3.000000 | 1250.000000 | 21.000000 |
| max | 63.000000 | 10.000000 | 100000.000000 | 2100.000000 |

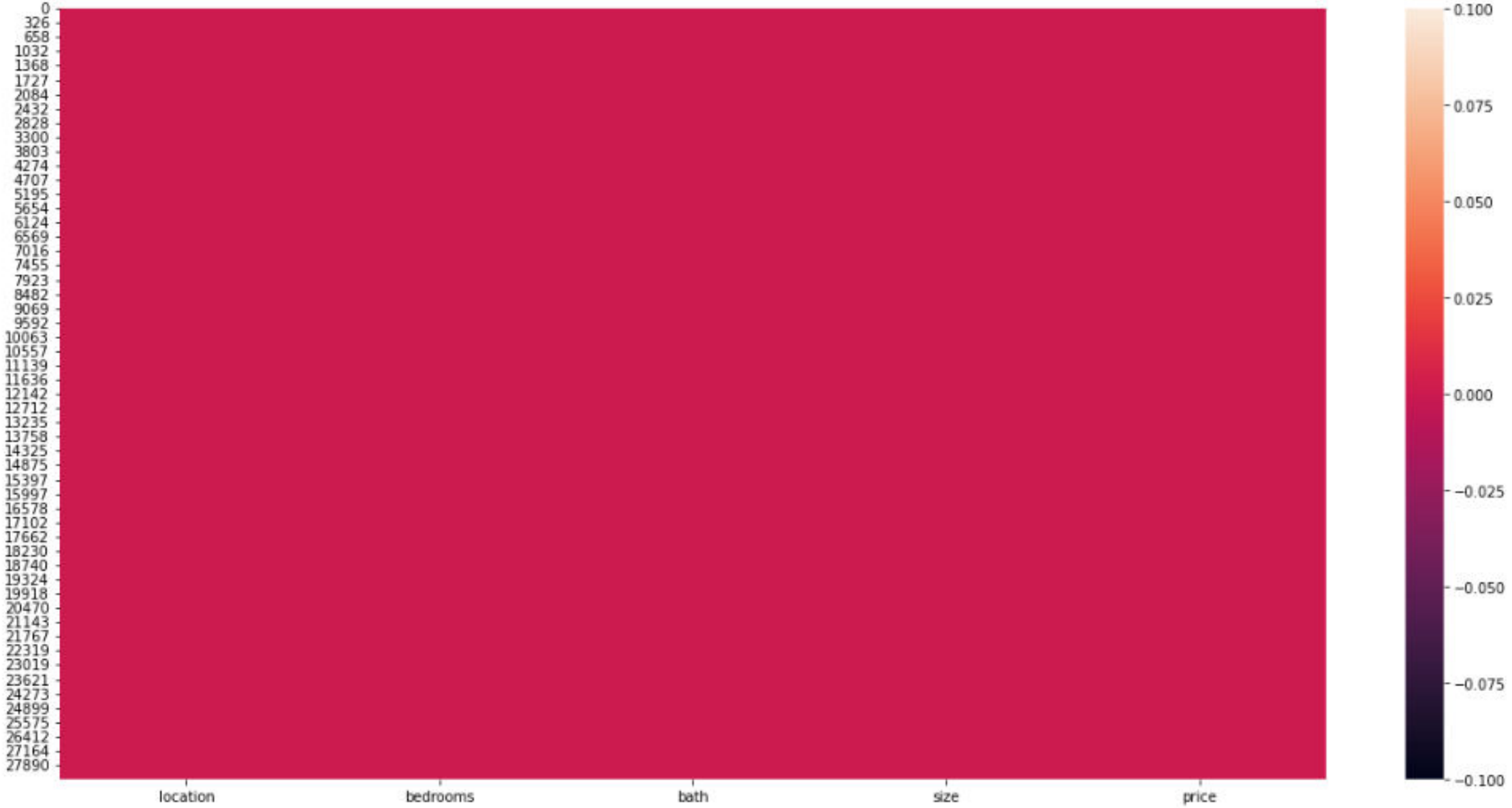
# Null value In Dataset

```
sns.heatmap(df.isnull(), cbar=False)
```

```
<AxesSubplot:>
```



# Heatmap


0
326
658
1032
1368
1727
2084
2432
2828
3300
3803
4274
4707
5195
5654
6124
6569
7016
7455
7923
8482
9069
9592
10063
10557
11139
11636
12142
12712
13235
13758
14325
14875
15397
15997
16578
17102
17662
18230
18740
19324
19918
20470
21143
21767
22319
23019
23621
24273
24899
25575
26412
27164
27890
location
bedrooms
bath
size
price
0.100
0.075
0.050
0.025
0.000
−0.025
−0.050
−0.075
−0.100

# Data Cleaning & Describe

```
df5 = df4[df4['bedrooms']<6]
df6 = df5[df5['bath']<6]
df7 = df6[df6['size']<3000]
df8 = df7[df7['size']>300]
df9 = df8[df8['price']<500]
df9 = df9[df9['price']>5]
df9.shape
```

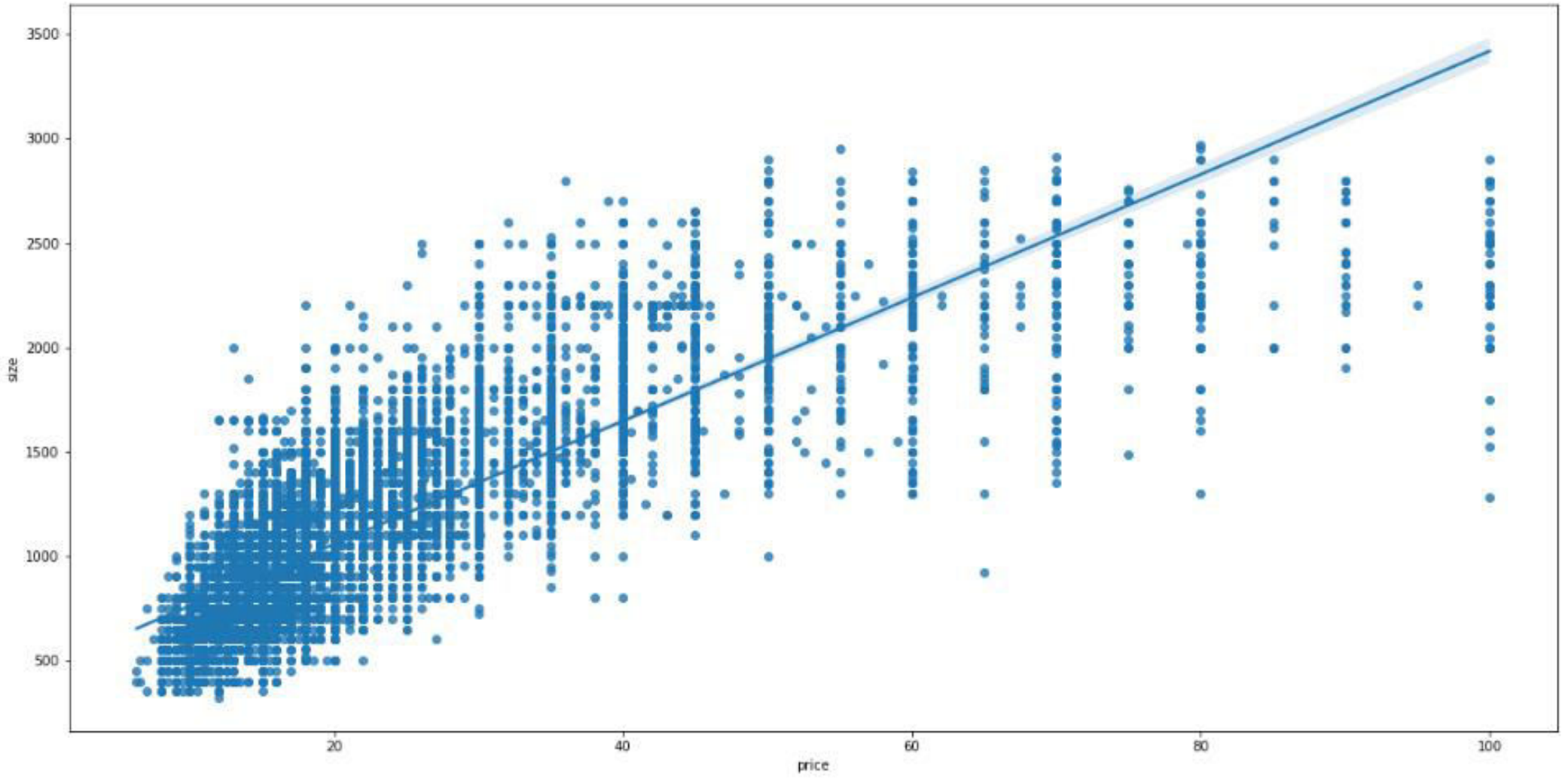
```
(15218, 5)
```

```
df9.describe()
```

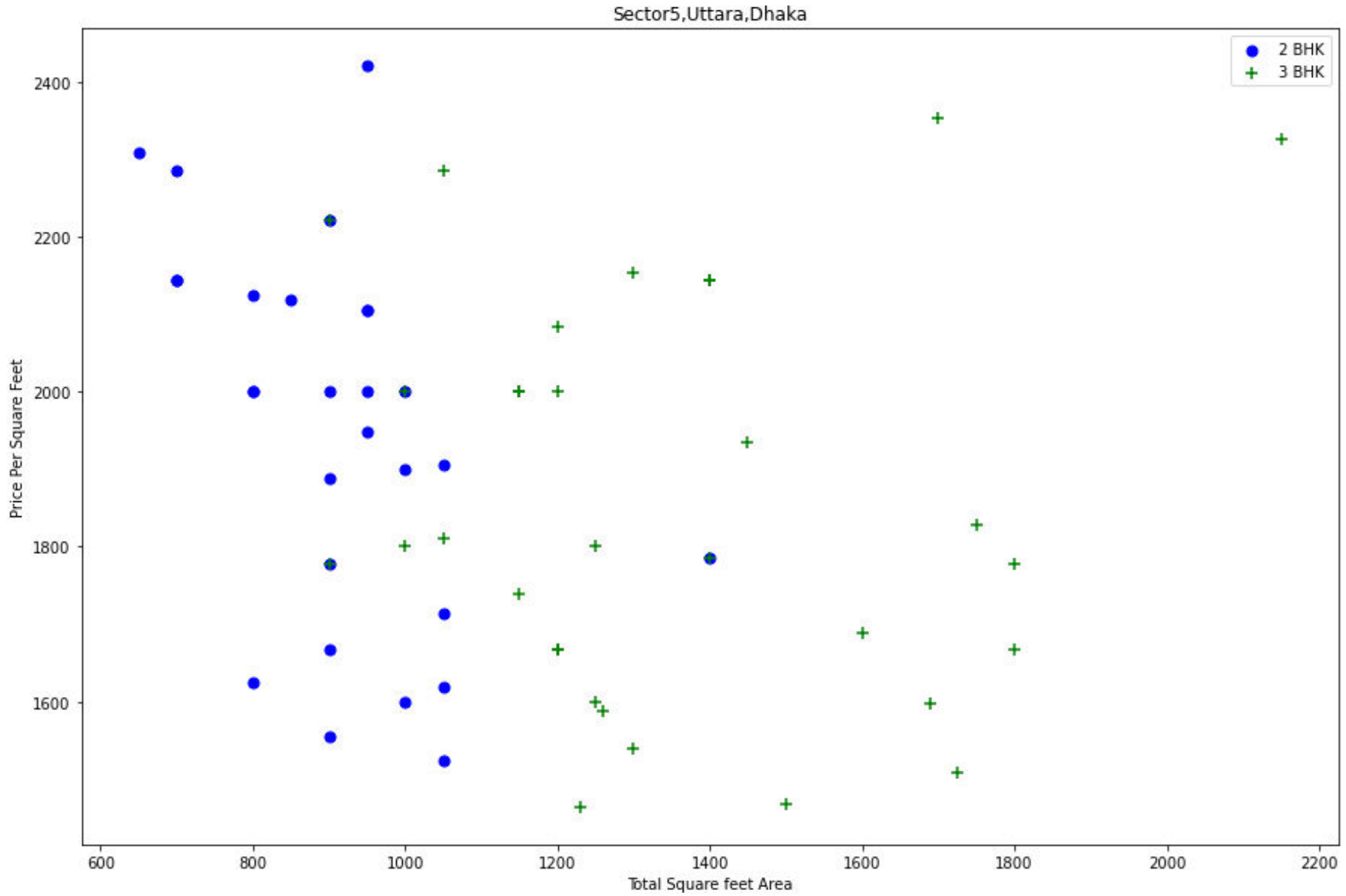
| | bedrooms | bath | size | price |
|---|---|---|---|---|
| count | 15218.000000 | 15218.000000 | 15218.000000 | 15218.000000 |
| mean | 2.554081 | 2.377579 | 1077.105861 | 20.711227 |
| std | 0.614067 | 0.809211 | 443.197003 | 12.442792 |
| min | 1.000000 | 1.000000 | 320.000000 | 6.300000 |
| 25% | 2.000000 | 2.000000 | 720.000000 | 13.500000 |
| 50% | 3.000000 | 2.000000 | 1000.000000 | 17.000000 |
| 75% | 3.000000 | 3.000000 | 1300.000000 | 24.000000 |
| max | 5.000000 | 5.000000 | 2970.000000 | 200.000000 |

# Dataset Regplot of size & price

```
sns.regplot(x="price", y="size", data=df9);
```

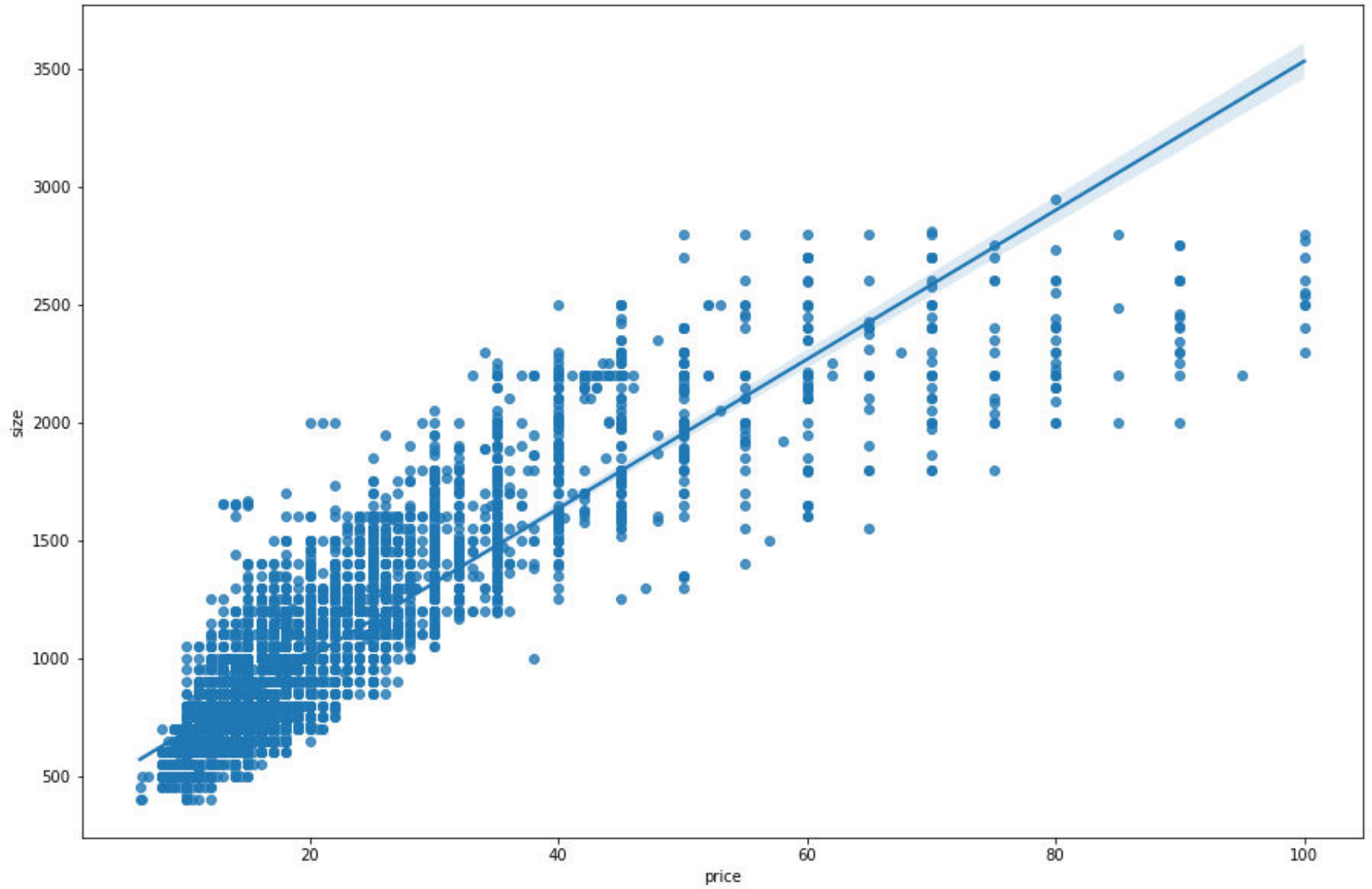


# Outliers Detection & Remove

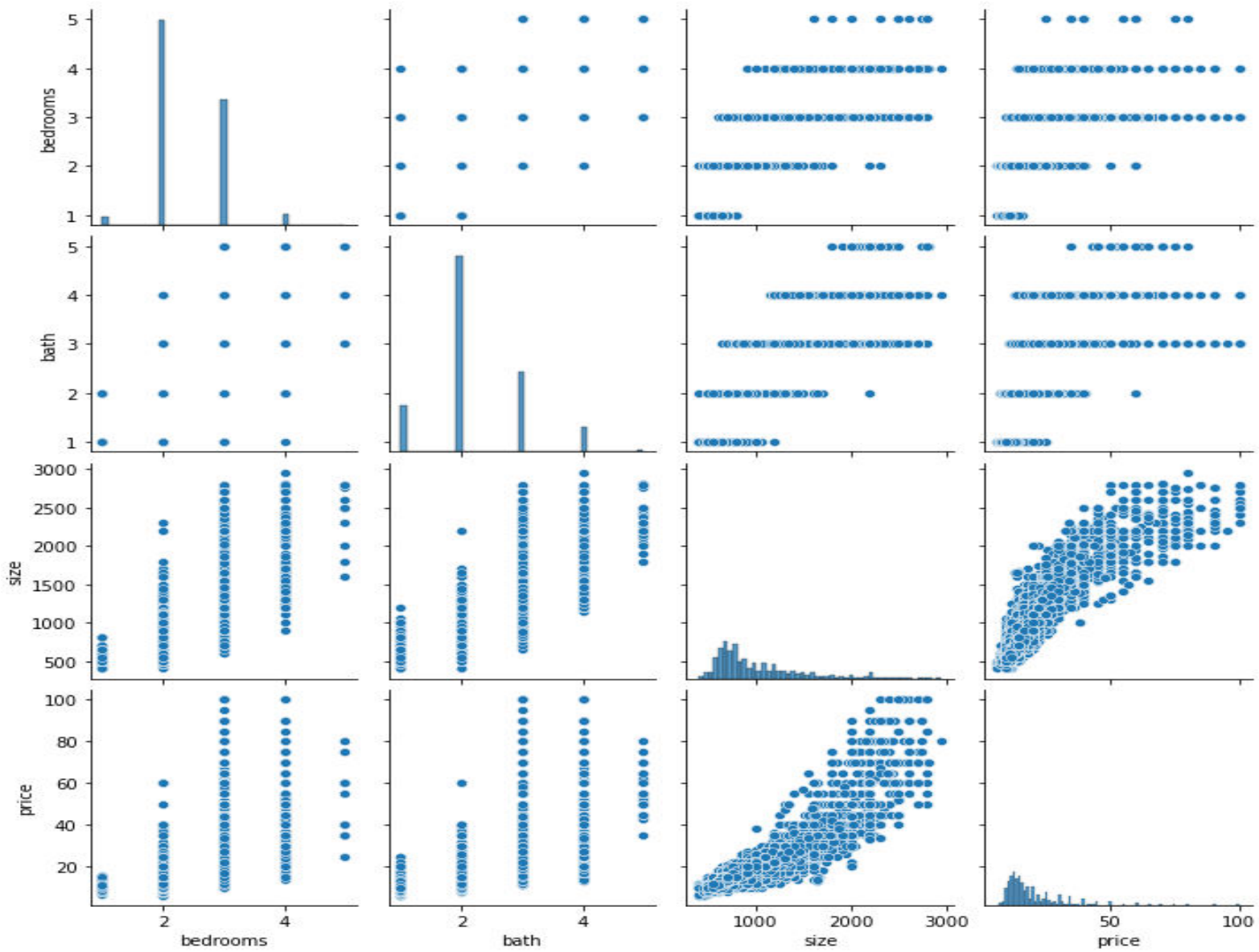


# Preprocessed Dataset Regplot of size &

```
sns.regplot(x="price", y="size", data=df12);
```



# Preprocessed Dataset Pairplot



# Machine Learning Models

```
from sklearn.linear_model import LinearRegression
lr_clf = LinearRegression()
lr_clf.fit(X_train,y_train)
lr_clf.score(X_test,y_test)
```

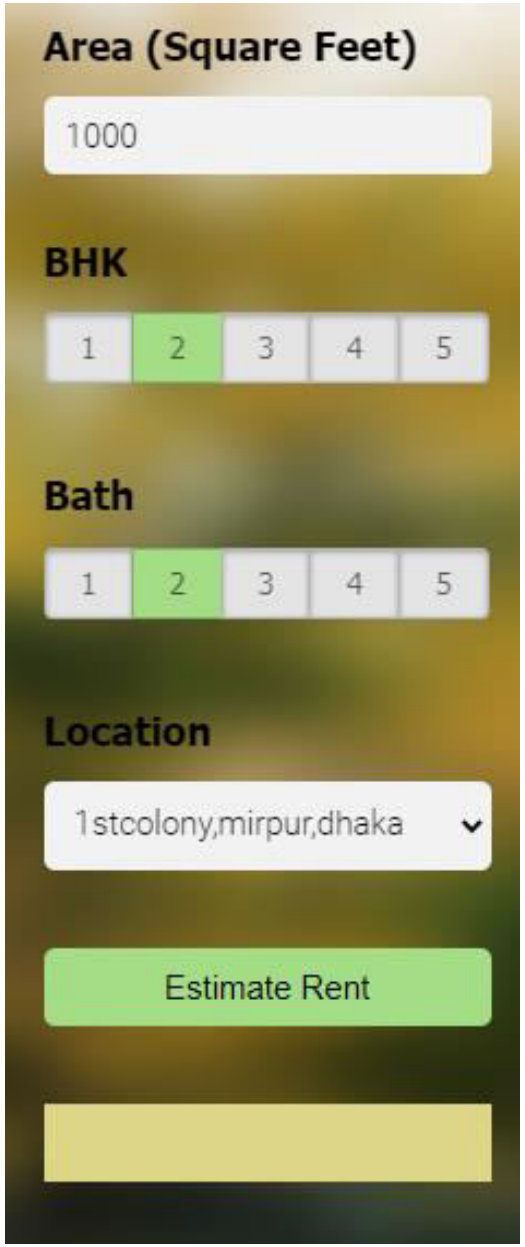
```
0.959423262825028
```

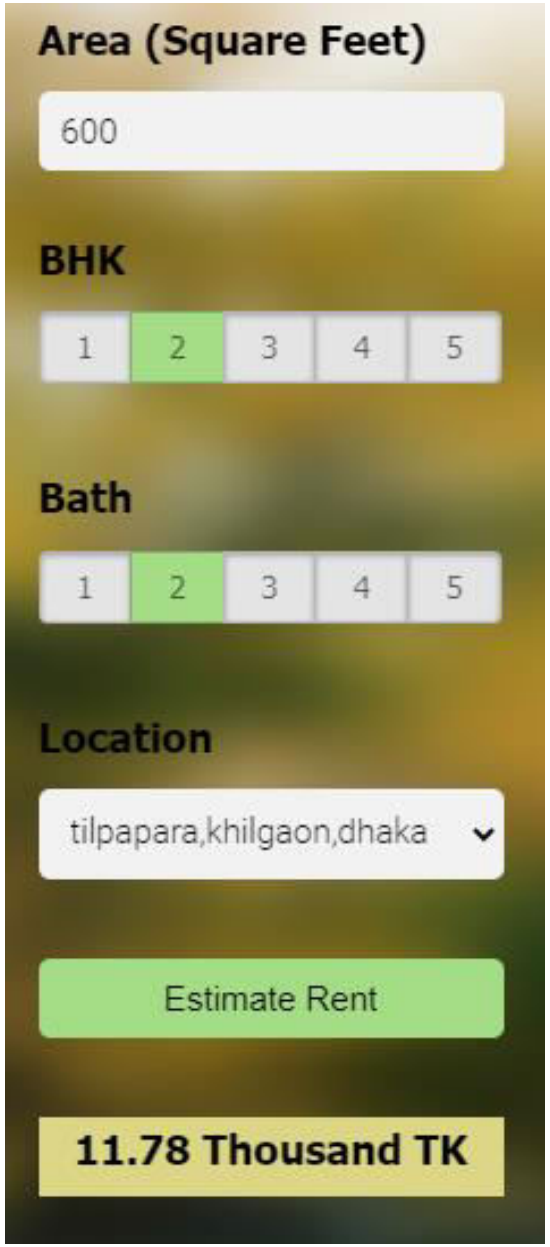
| | model | best_score | best_params |
|---|---|---|---|
| 0 | linear_regression | 0.952714 | {'normalize': True} |
| 1 | lasso | 0.779855 | {'alpha': 1, 'selection': 'random'} |
| 2 | decision_tree | 0.912485 | {'criterion': 'mse', 'splitter': 'best'} |

```
cross_val_score(LinearRegression(), X, y, cv=cv)
```

```
array([0.95942326, 0.95291503, 0.95925098, 0.95747018, 0.95820748])
```

# House Rent Prediction Page



# Future Work

The future works of this project include the following:

- We can add more feature in our dataset.
- We can use other advanced regression techniques, like Random Forest and Bayesian Ridge Algorithm.
- Since the data is highly correlated, we should try Elastic Net regression Technique.

Thank You!