

CYBERBULLS_DM_PROJECT-1.pptx

by Sai Harshitha Padala

Submission date: 07-Nov-2023 08:07AM (UTC-0500)

Submission ID: 2220543777

File name: CYBERBULLS_DM_PROJECT.pptx (2M)

Word count: 571

Character count: 3356

PREDICTING HOUSE RENTAL PRICES

Leveraging Data To Predict Rental Property



CYBER BULLS

YASHWANTH KRISHNA VARMA KOLUKULURI - U17838568

PURNA CHANDRA SHEKAR REDDY GOSALA-U63359957

SAI HARSHITHA PADALA- U91378601

DRITHI REDDY PEESARI-U03445314

NESHMA KASAMNENI - U22120891

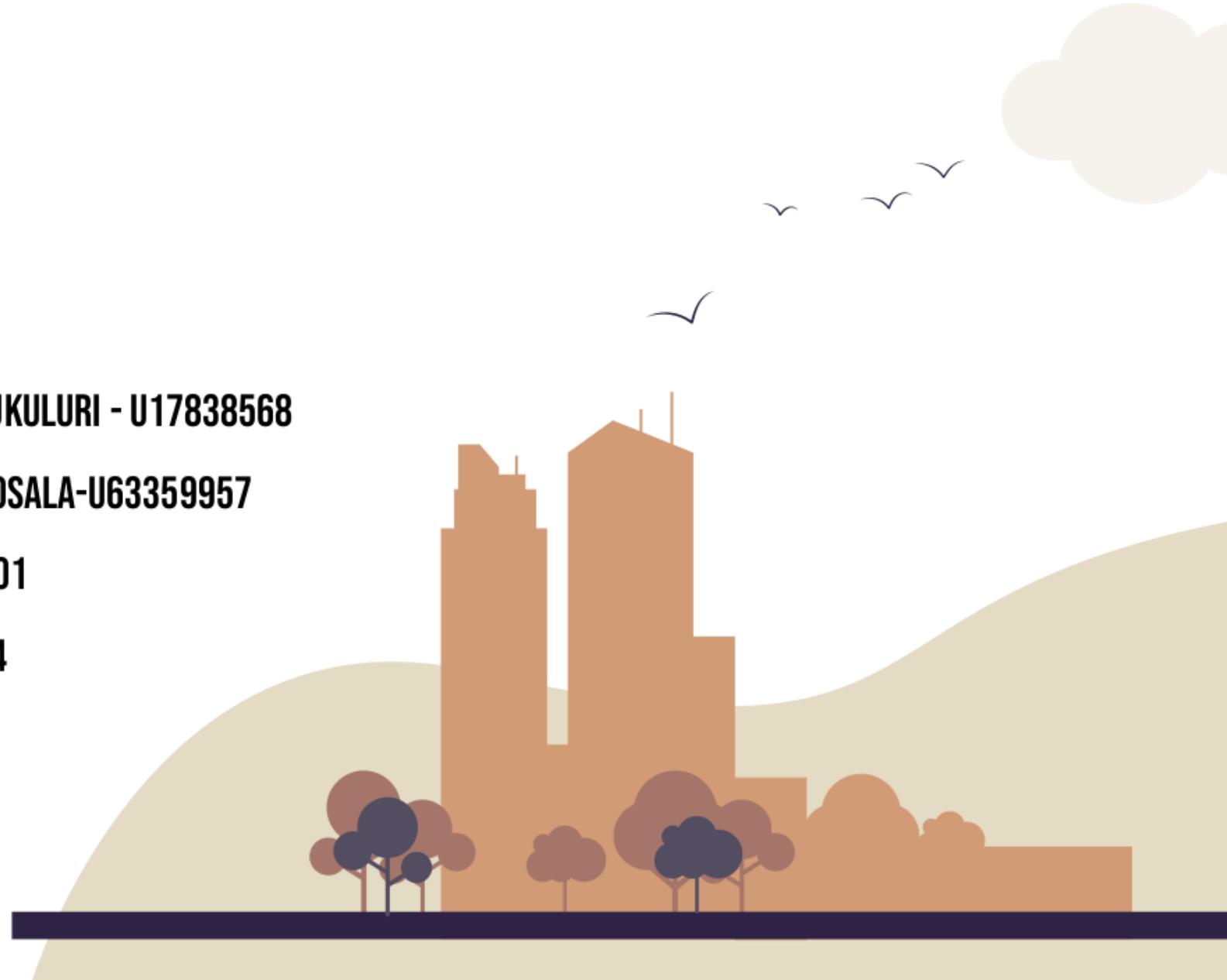


TABLE OF CONTENTS

INTRODUCTION AND BUSINESS PROBLEM

DATA OVERVIEW

OUR APPROACH

MODEL ANALYSIS

RESULTS

INTERPRETATIONS AND CONCLUSION

INTRODUCING THE BUSINESS PROBLEM

"A REAL ESTATE AGENCY IS LOOKING TO OPTIMIZE THE PROCESS OF ESTIMATING HOUSE RENTAL PRICES FOR BOTH PROPERTY OWNERS AND POTENTIAL RENTERS. THEY WANT TO ENSURE THAT RENTAL PRICES ARE ACCURATELY AND FAIRLY DETERMINED, TAKING INTO ACCOUNT THE VARIOUS FACTORS THAT INFLUENCE PRICING. THE AGENCY AIMS TO PROVIDE A DATA-DRIVEN SOLUTION THAT BENEFITS BOTH LANDLORDS AND TENANTS IN THE REAL ESTATE MARKET."

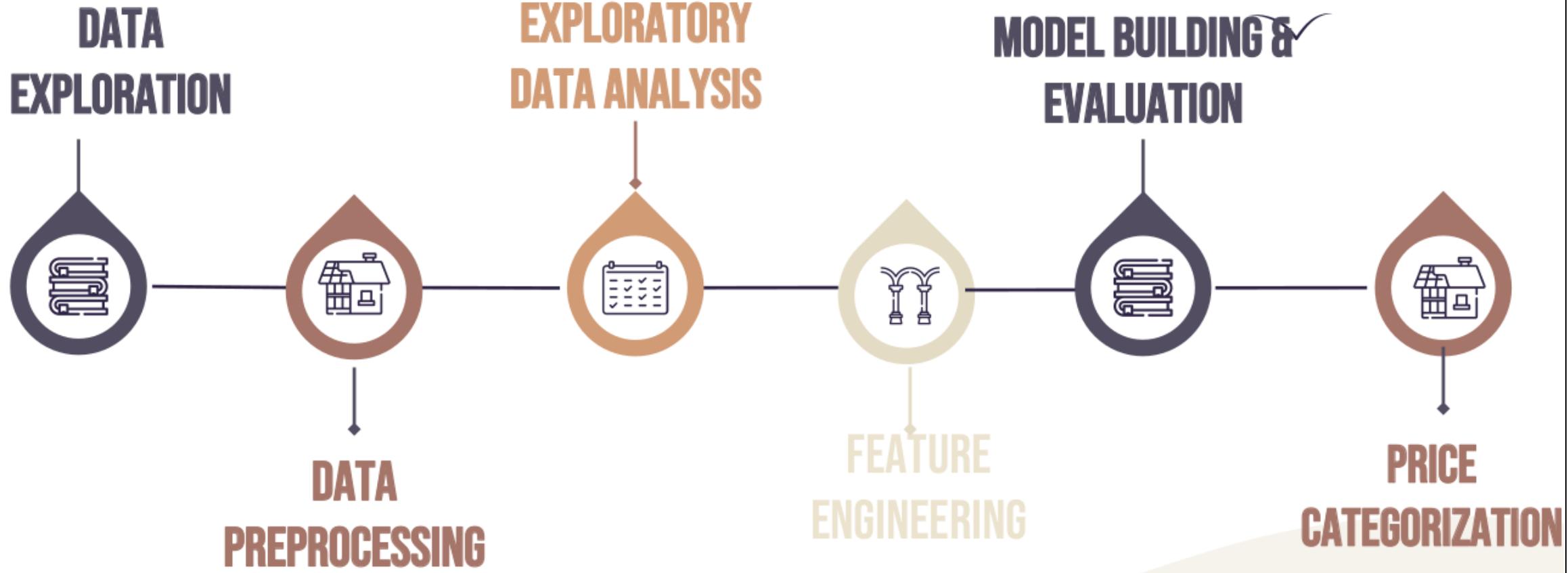
DATA OVERVIEW

22 COLUMNS
1,00,000 / 100K - ROWS



OUR APPROACH





1

DATA EXPLORATION



- DATA EXPLORATION IS THE FIRST STEP IN DATA ANALYSIS AND TYPICALLY INVOLVES SUMMARIZING THE MAIN CHARACTERISTICS OF A DATA SET, INCLUDING ITS SIZE.
- BEFORE ANALYZING A MIXED BATCH OF DATA STORED TOGETHER, IT'S IMPORTANT TO FIRST GET A HANDLE ON HOW MUCH DATA THERE IS, WHAT KINDS OF INFORMATION ARE INCLUDED IF ANY PIECES ARE MISSING, AND WHAT THE DATA MIGHT SHOW.
- WE DIVIDED THE DATA 8:2 FOR TRAINING AND TESTING PURPOSES RESPECTIVELY.

Dut[3]:

09	housing/rent/apartment	One BR 507 & 509 Esplanade	This unit is located at 507 & 509 Esplanade, R...	NaN	1.0	1.0	USD	No	Thumbnail	...	2195.0	Monthly	542.0	507 Esplanade
18	housing/rent/apartment	Three BR 146 Lochview Drive	This unit is located at 146 Lochview Drive, Ne...	NaN	1.5	3.0	USD	No	Thumbnail	...	1250.0	Monthly	1500.0	Lochview
86	housing/rent/apartment	Three BR 3101 Morningside Drive	This unit is located at 3101 Morningside Drive...	NaN	2.0	3.0	USD	No	Thumbnail	...	1395.0	Monthly	1650.0	Morningside
59	housing/rent/apartment	Two BR 209 Aegean Way	This unit is located at 209 Aegean Way, Vacavi...	NaN	1.0	2.0	USD	No	Thumbnail	...	1600.0	Monthly	820.0	209 Aegean Way
74	housing/rent/apartment	One BR 4805 Marquette NE	This unit is located at 4805 Marquette NE, Alb...	NaN	1.0	1.0	USD	No	Thumbnail	...	975.0	Monthly	624.0	4805 Marquette NE

columns

<																			>
---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	---

In [4]:

```
data.columns
```

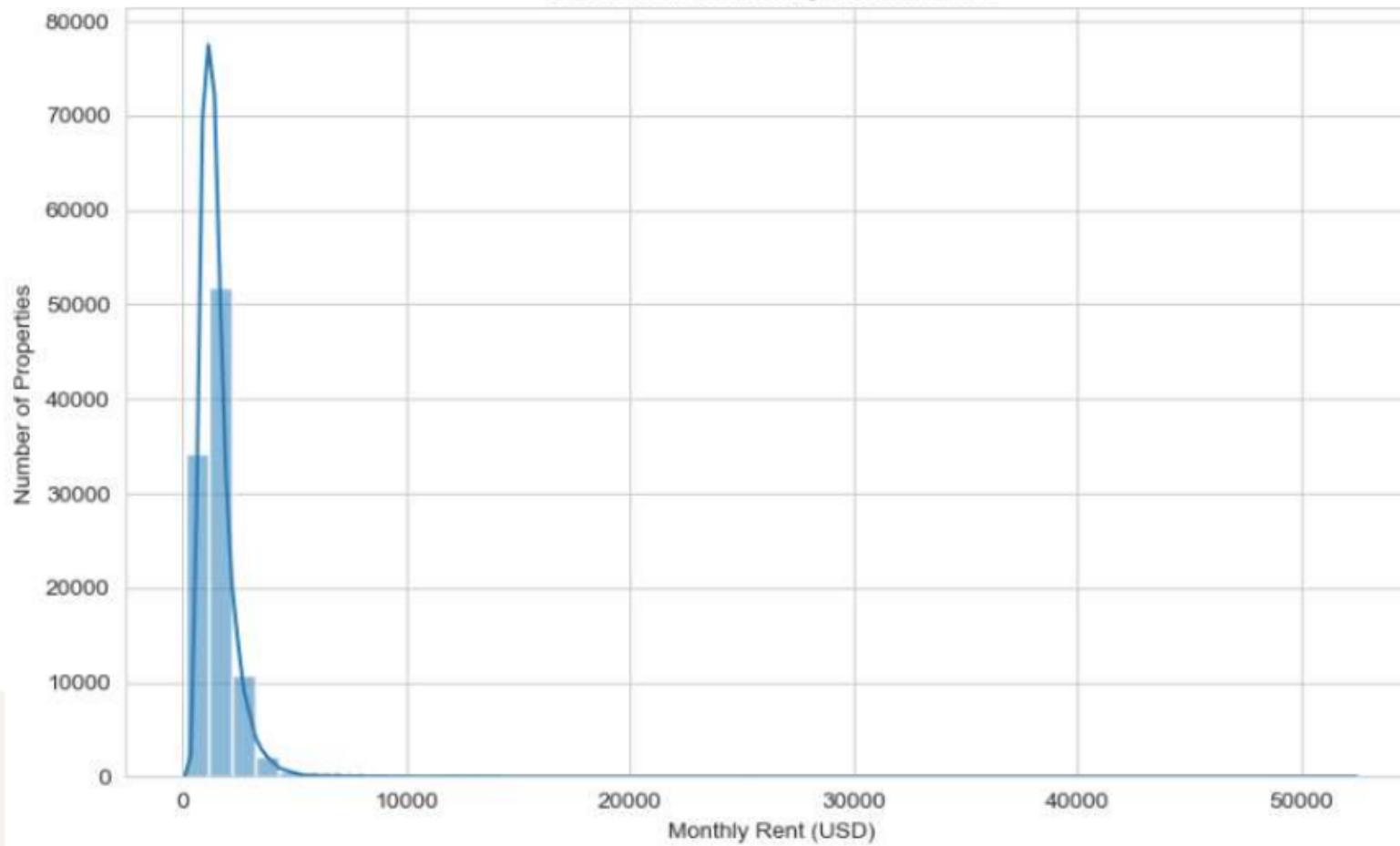
Dut[4]: Index(['id', 'category', 'title', 'body', 'amenities', 'bathrooms', 'bedrooms', 'currency', 'fee', 'has_photo', 'pets_allowed', 'price', 'price_display', 'price_type', 'square_feet', 'address', 'cityname', 'state', 'latitude', 'longitude', 'source', 'time'], dtype='object')

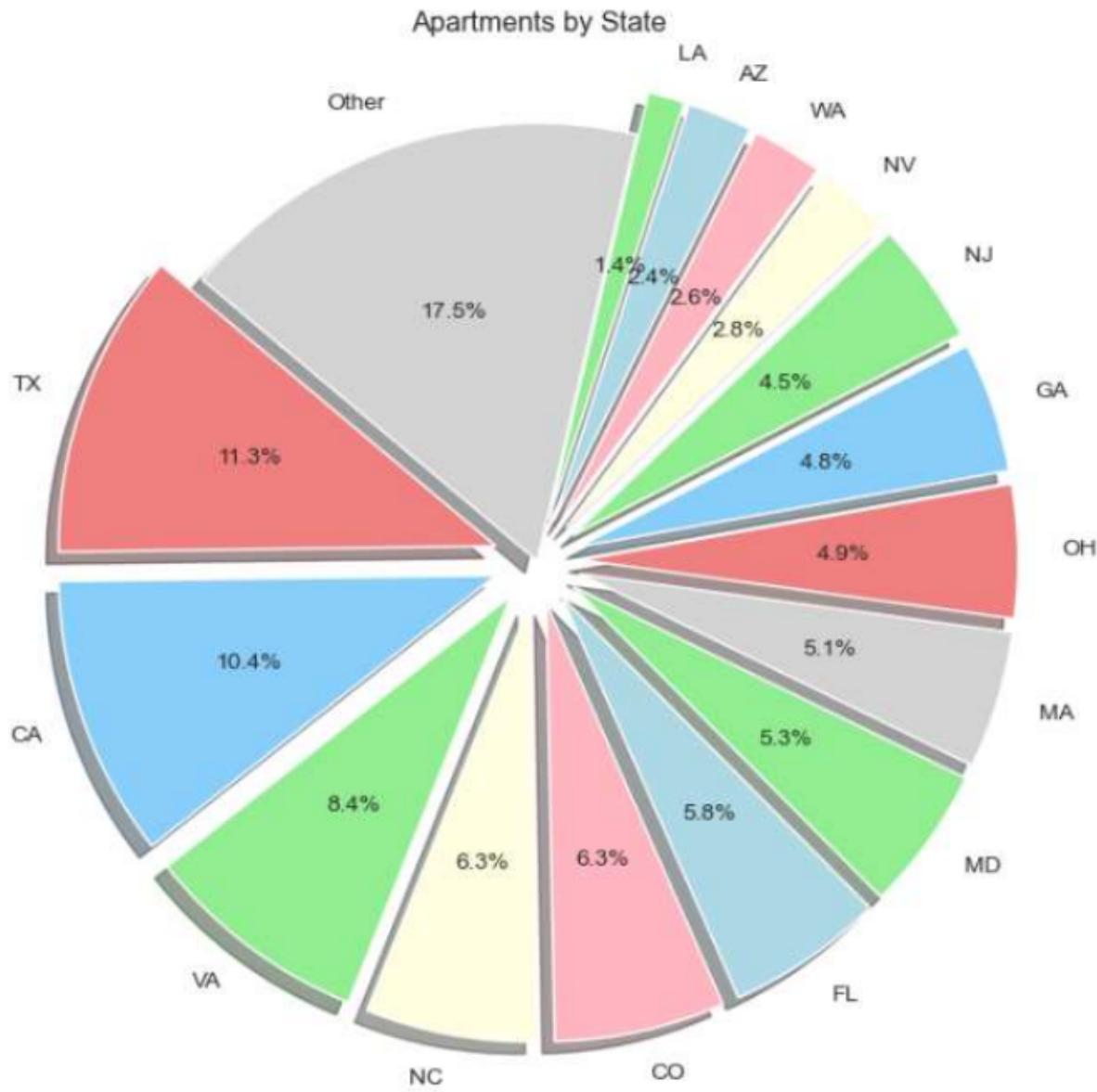
EXPLORATORY DATA ANALYSIS

- EXPLORATORY DATA ANALYSIS IS THE GRAPHICAL REPRESENTATION OF INFORMATION AND DATA. BY USING VISUAL ELEMENTS LIKE CHARTS, GRAPHS, AND MAPS, EDA TOOLS PROVIDE AN ACCESSIBLE WAY TO SEE AND UNDERSTAND TRENDS, OUTLIERS, AND PATTERNS IN DATA.
- IN THE WORLD OF BIG DATA, EDA TOOLS AND TECHNOLOGIES ARE ESSENTIAL TO ANALYZE MASSIVE AMOUNTS OF INFORMATION AND MAKE DATA-DRIVEN DECISIONS.

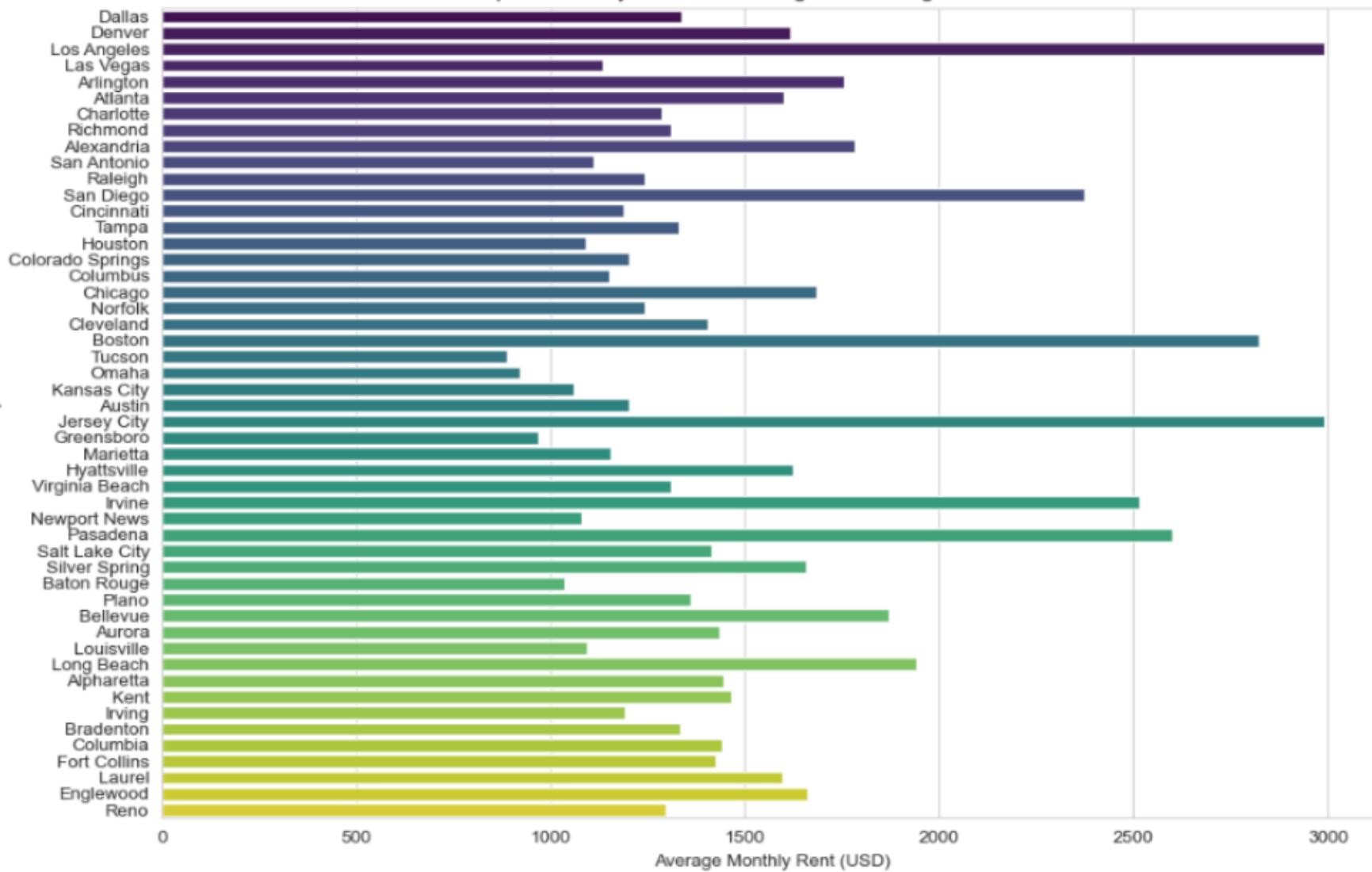
DATA VISUALIZATION GRAPHS

Distribution of Monthly Rental Prices

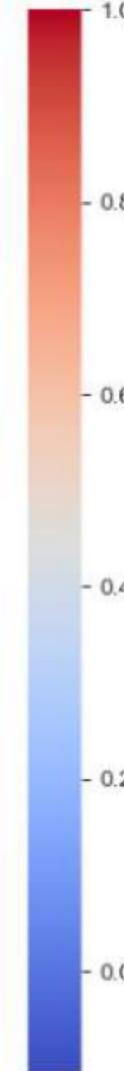
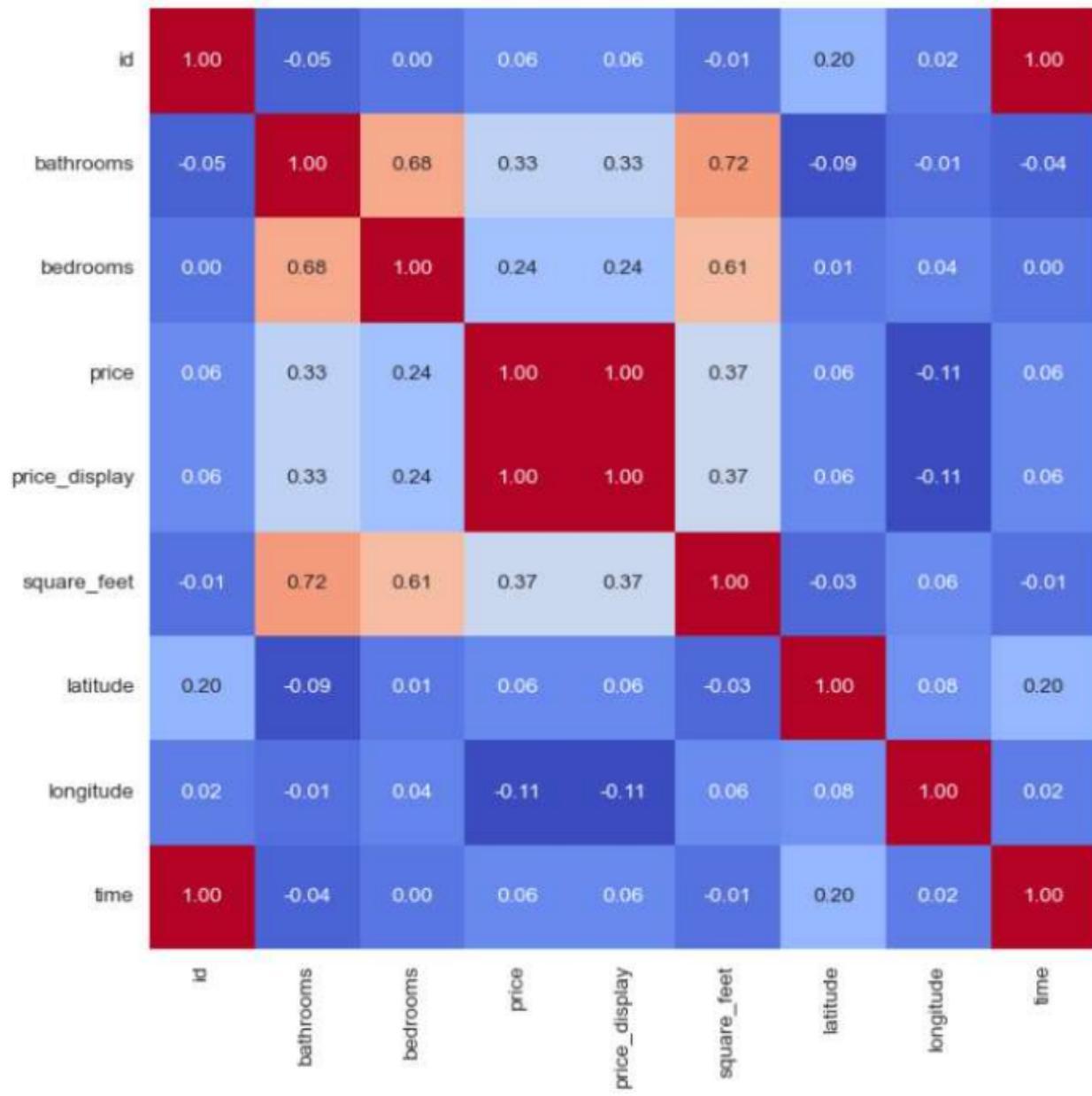




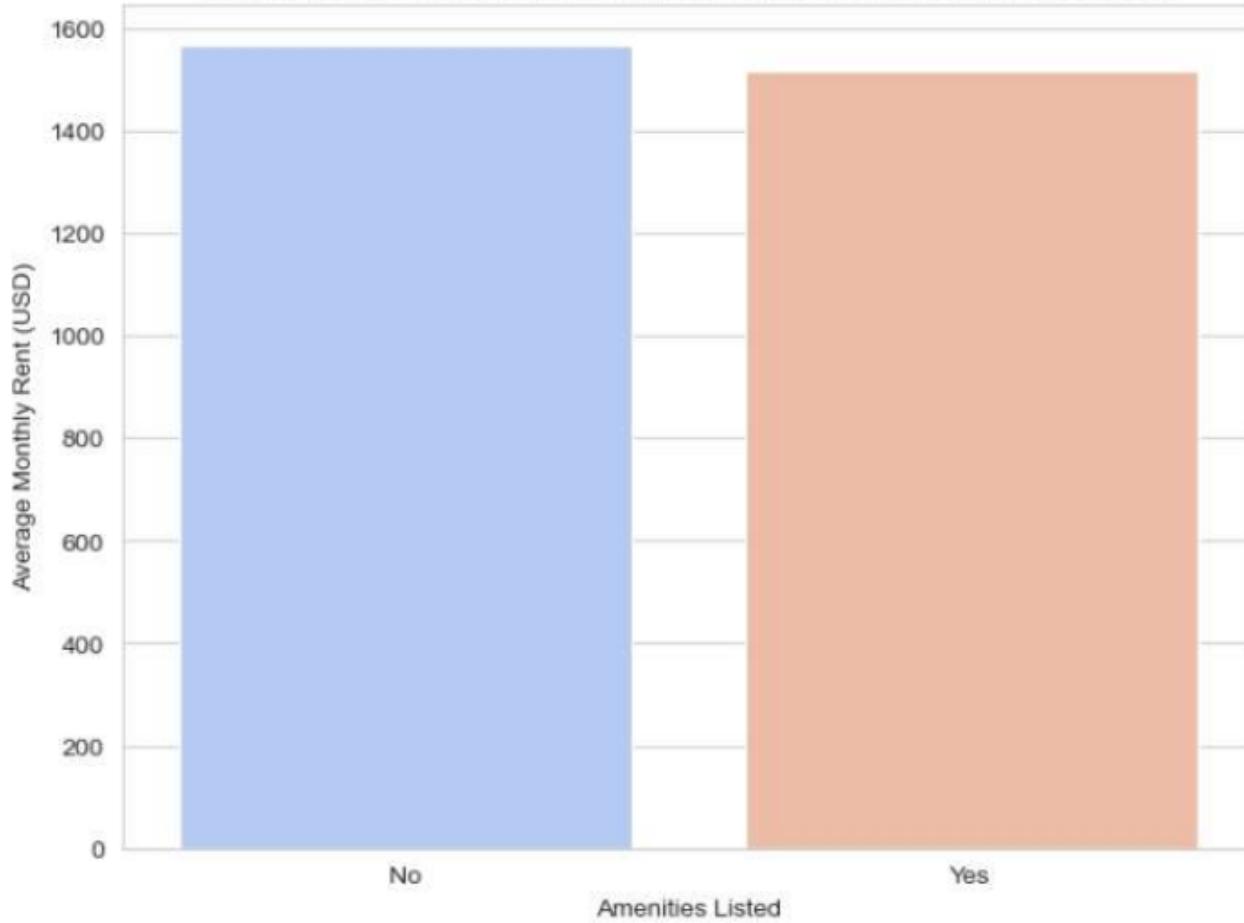
Top 50 Cities by Number of Listings and Average Rental Price

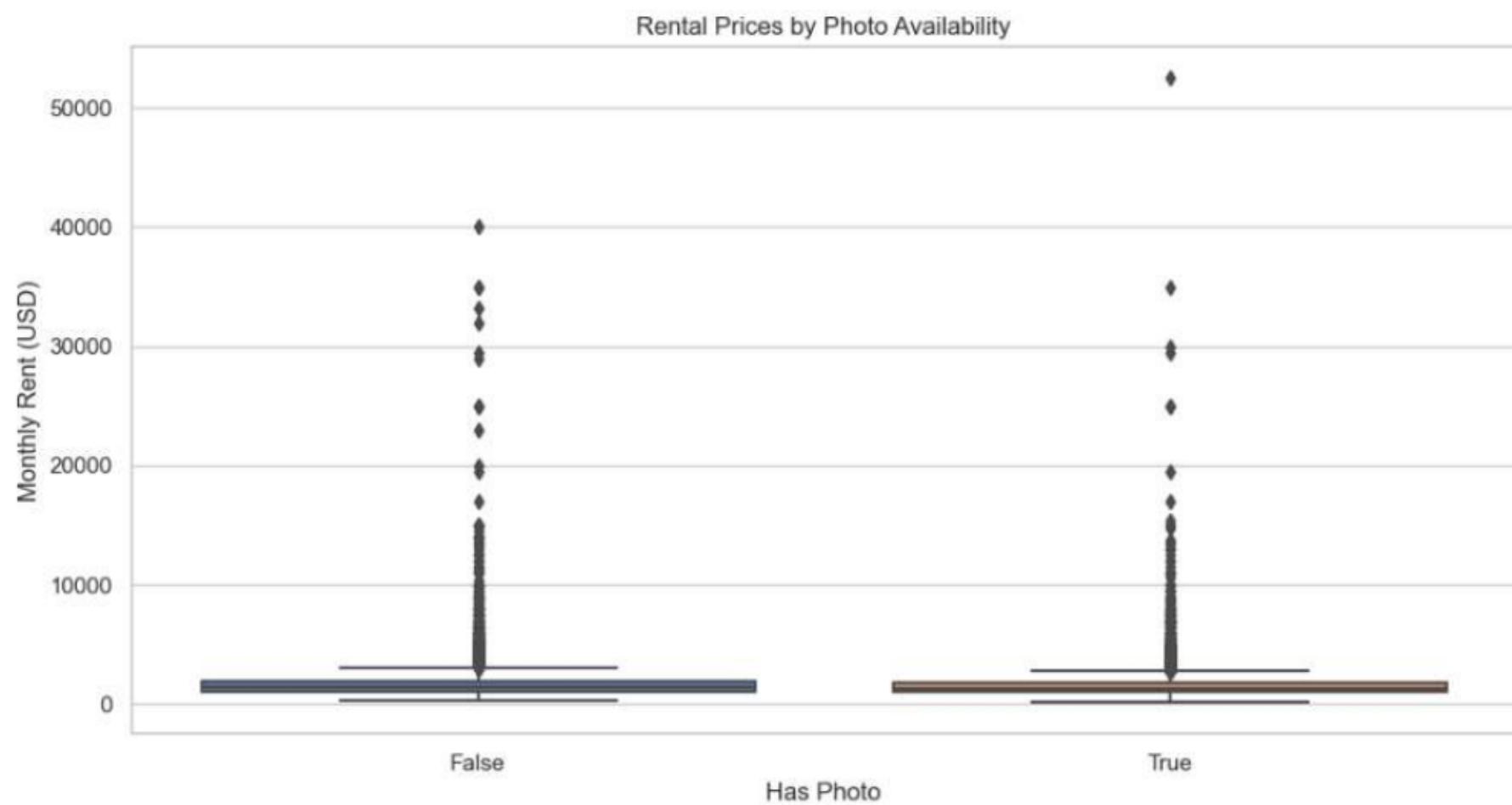


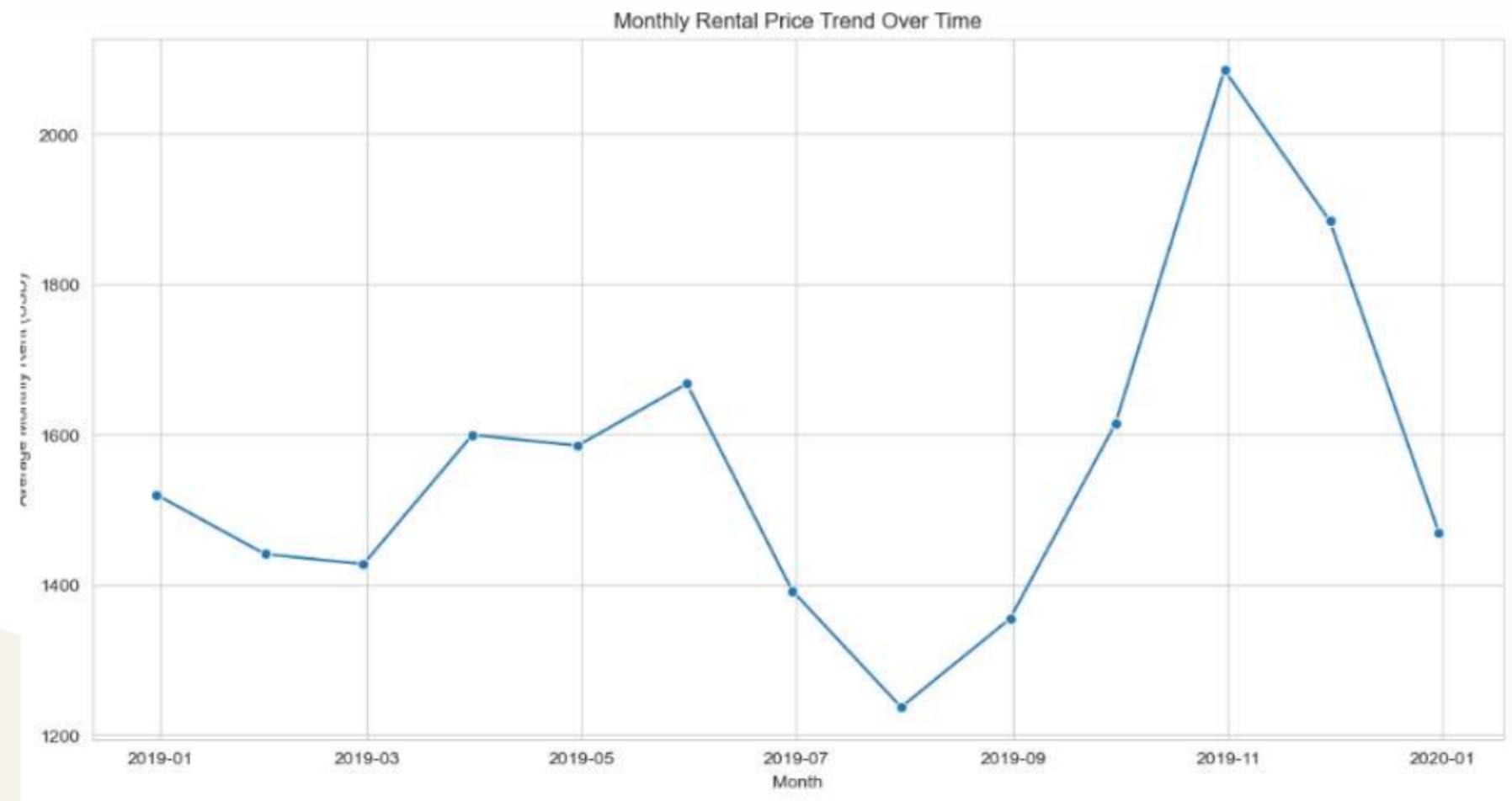
Correlation Matrix



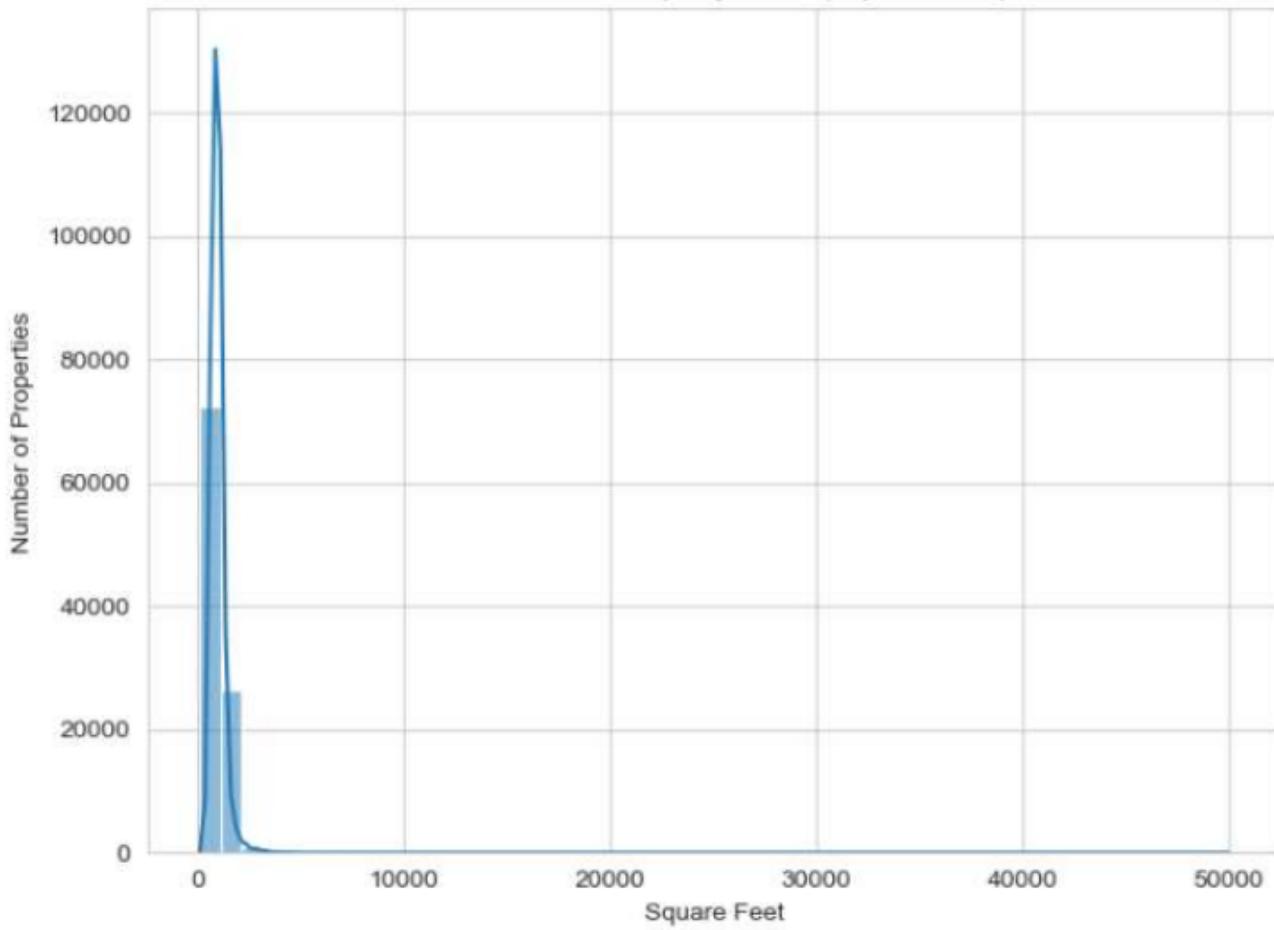
Comparison of Average Rental Prices With and Without Amenities Listed



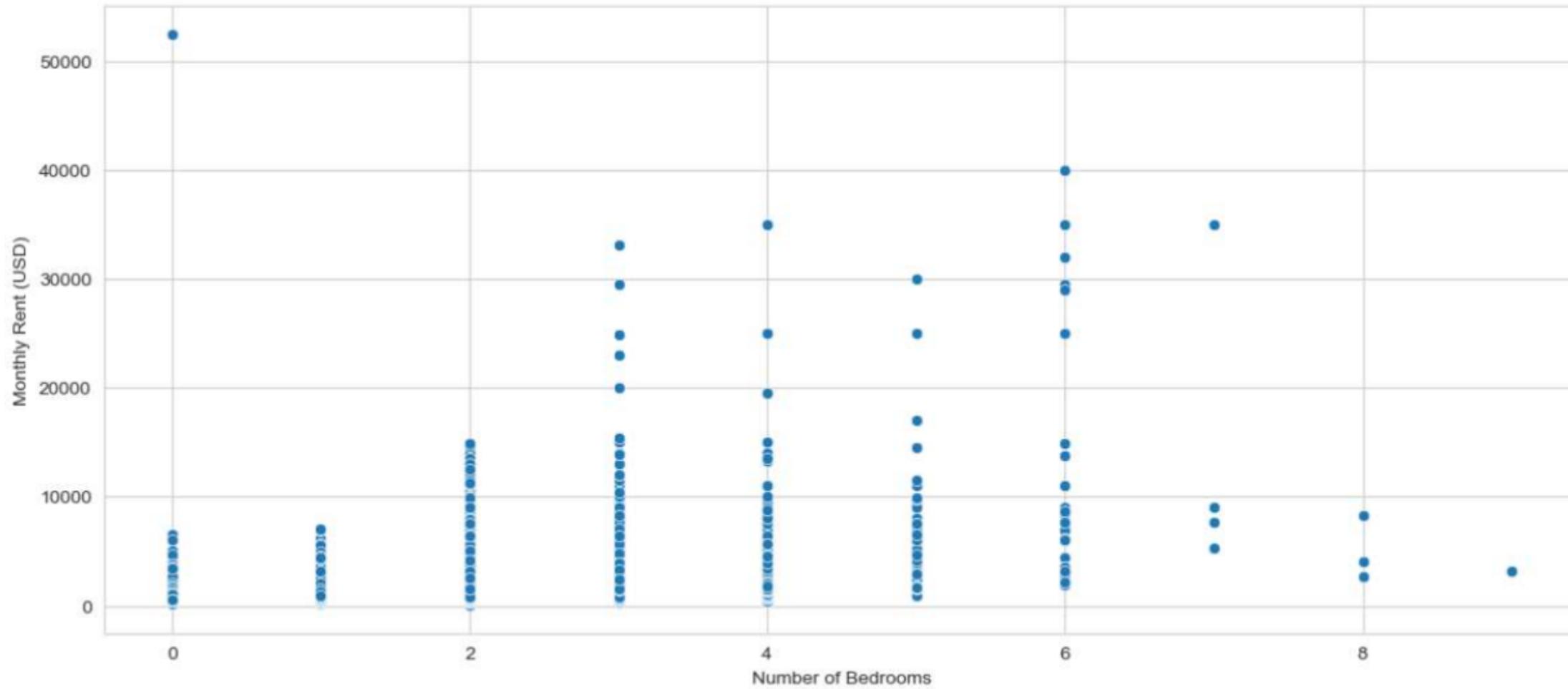




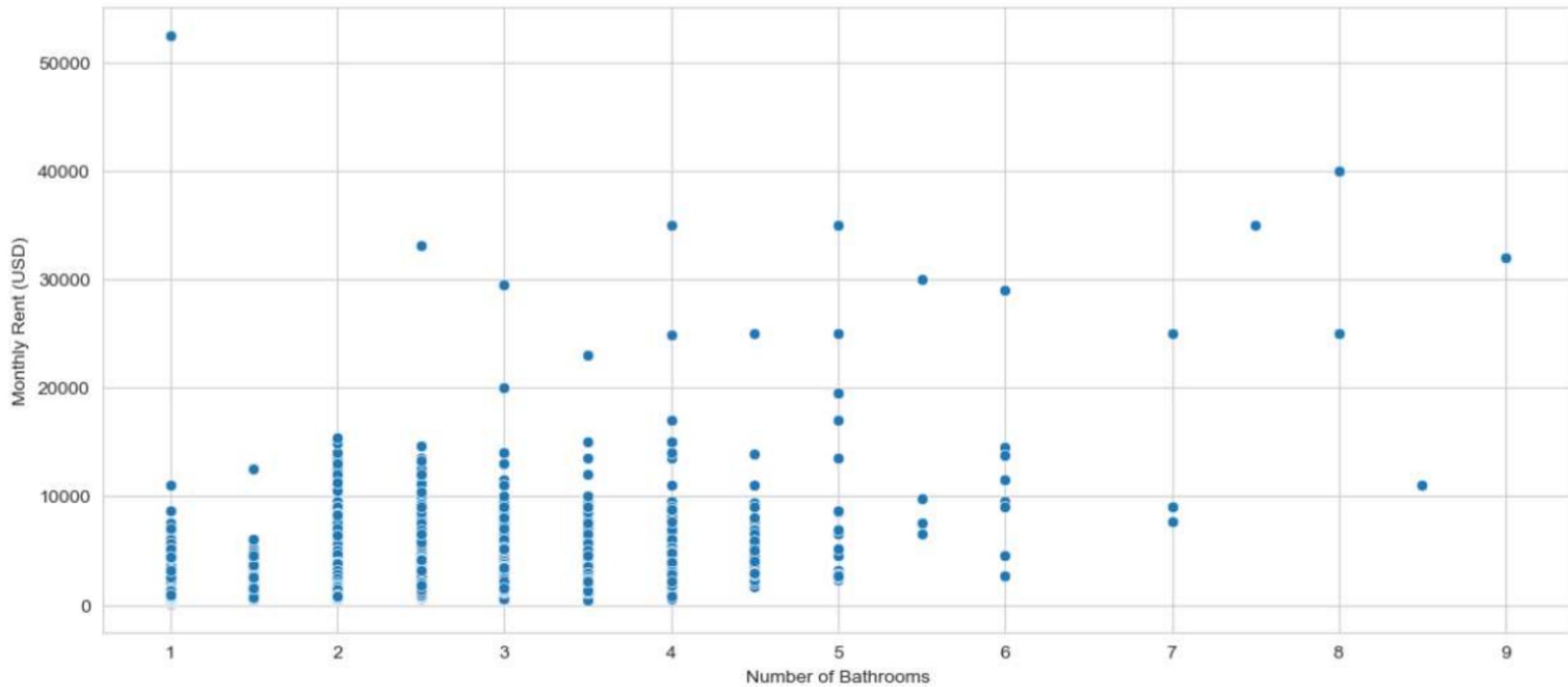
Distribution of Property Sizes (Square Feet)



Rental Price vs. Number of Bedrooms



Rental Price vs. Number of Bathrooms

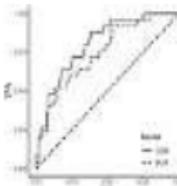




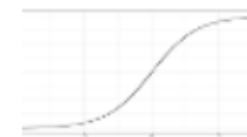
MODEL ANALYSIS

MODELS USED

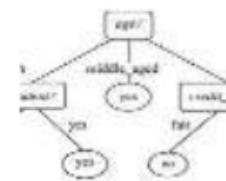
- GRADIENT BOOSTER REGRESSOR 3



- RANDOM FOREST REGRESSOR



- DECISION TREE REGRESSOR



- LINEAR REGRESSION

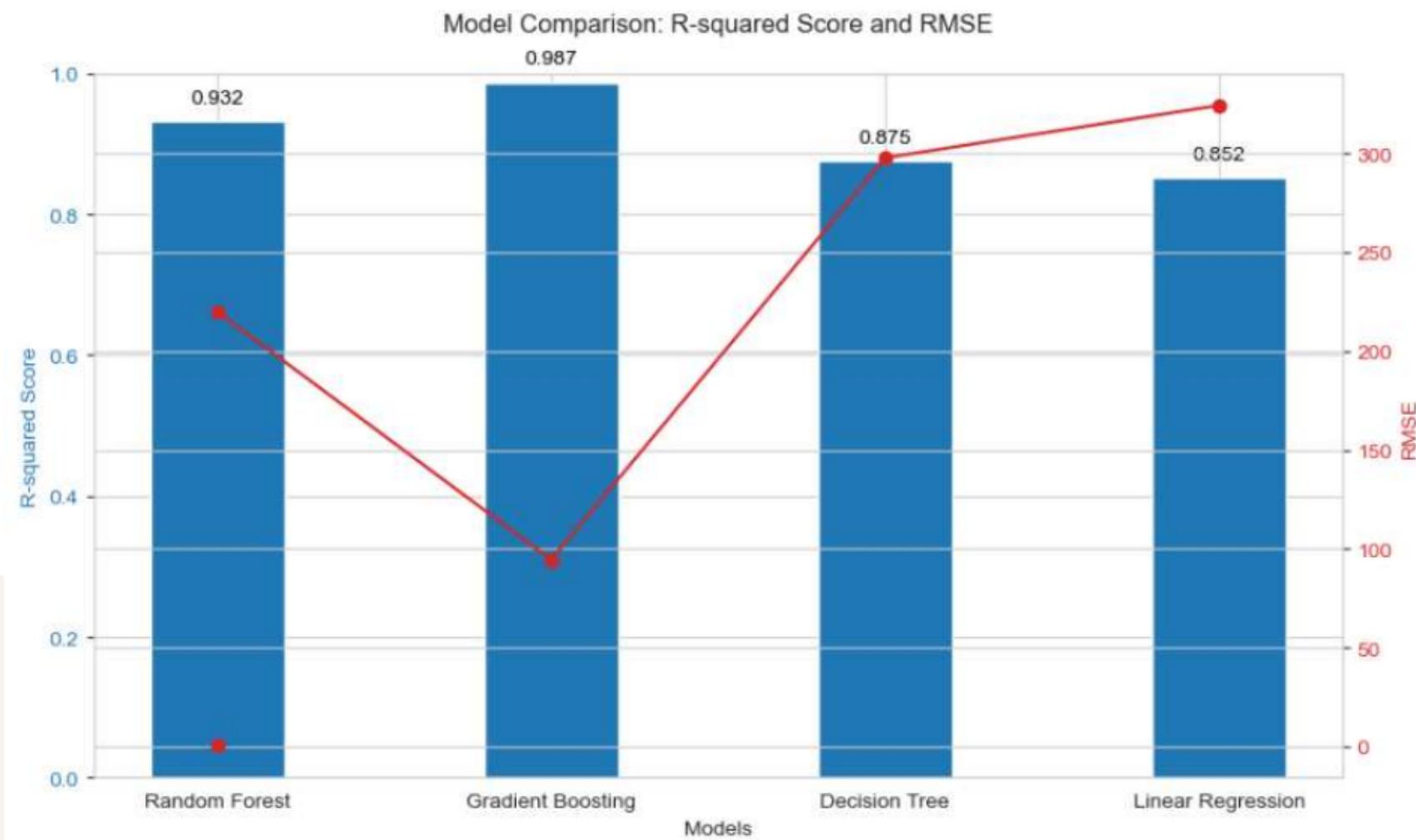
RESULTS



MODEL COMPARISON

	Model	MAE	MSE	RMSE	R2 Score
0	GradientBoostingRegressor	25.472811	8914.543433	94.416860	0.987467
1	DecisionTreeRegressor	158.432107	88637.112857	297.719856	0.875381
2	RandomForestRegressor	121.308996	48382.508090	219.960242	0.931977
3	LinearRegression	114.400269	105155.520510	324.276919	0.852157

MODEL COMPARISON:



INTERPRETATION AND CONCLUSION



BASED ON THE R² SCORE AND RMSE VALUES:

- THE GRADIENT BOOSTING REGRESSOR IS THE TOP-PERFORMING MODEL IN PREDICTING HOUSE RENT PRICES.
- IT ACHIEVES THE HIGHEST R-SQUARED SCORE AND THE LOWEST RMSE, INDICATING AN EXCELLENT FIT TO THE DATA AND THE MOST ACCURATE PREDICTIONS.
- THEREFORE, FOR THE SPECIFIC TASK OF HOUSE RENT PREDICTION, THE GRADIENT BOOSTING REGRESSOR IS THE PREFERRED CHOICE, OFFERING THE HIGHEST ACCURACY AND THE BEST FIT TO THE DATA ² AMONG THE EVALUATED MODELS.

- THE GRADIENT BOOSTING REGRESSOR STANDS OUT AS THE BEST MODEL BASED ON THE EVALUATION METRICS. IT HAS THE LOWEST MAE, MSE, AND RMSE, INDICATING THE SMALLEST PREDICTION ERRORS, AND THE HIGHEST R² SCORE, INDICATING IT EXPLAINS THE MOST VARIANCE IN THE TARGET VARIABLE.
- THE RANDOM FOREST REGRESSOR ALSO PERFORMS WELL, WITH COMPETITIVE MAE, MSE, RMSE, AND R² VALUES.
- THE DECISION TREE REGRESSOR IS LESS ACCURATE WITH HIGHER PREDICTION ERRORS COMPARED TO THE GRADIENT BOOSTING AND RANDOM FOREST MODELS.
- THE LINEAR REGRESSION MODEL, WHILE HAVING A GOOD R² SCORE, SHOWS HIGHER PREDICTION ERRORS IN TERMS OF MAE, MSE, AND RMSE.

RANDOM FOREST CLASSIFICATION

Based on the provided statistics for the "price" variable, we can create classification bins as follows:

- "**Low**" for prices between 200 and 1023 (25th percentile)
- "**Medium-Low**" for prices between 1024 and 1346 (50th percentile)
- "**Medium-High**" for prices between 1347 and 1758 (75th percentile)
- "**High**" for prices between 1759 and 35000 (maximum)



Confusion Matrix:

[0	1	0	0	0]
[0	3548	41	0	0]
[0	59	3448	88	0]
[0	1	70	3489	62]
[0	0	0	94	3418]]

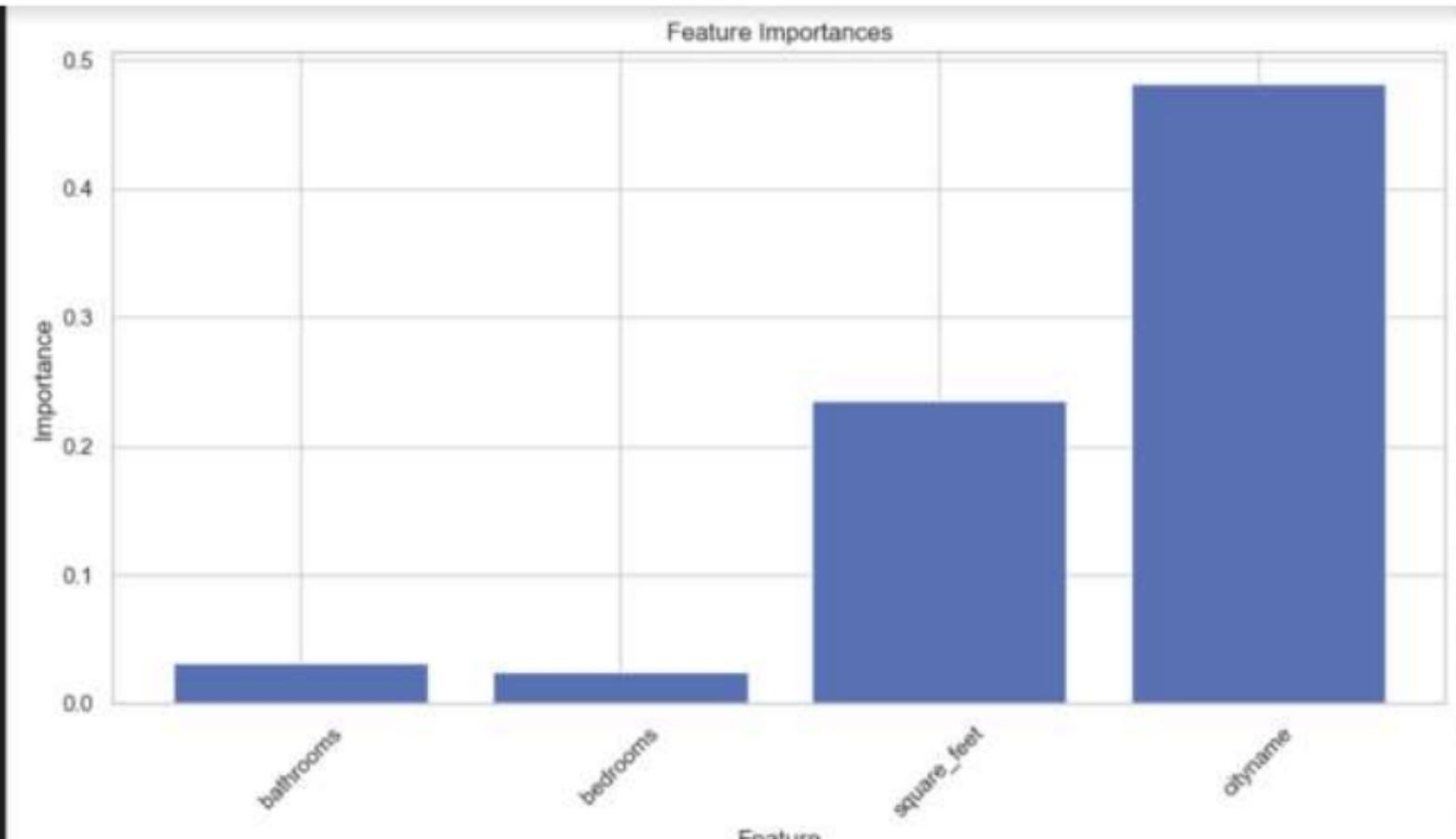
Accuracy: 0.9709476918779244

Classification Report:

	precision	recall	f1-score	support
-1	0.00	0.00	0.00	1
0	0.98	0.99	0.99	3589
1	0.97	0.96	0.96	3595
2	0.95	0.96	0.96	3622
3	0.98	0.97	0.98	3512
accuracy			0.97	14319
macro avg	0.78	0.78	0.78	14319
weighted avg	0.97	0.97	0.97	14319

The Random Forest classifier performs well in predicting price categories, achieving high accuracy and demonstrating good precision, recall, and F1-scores for the majority of categories.

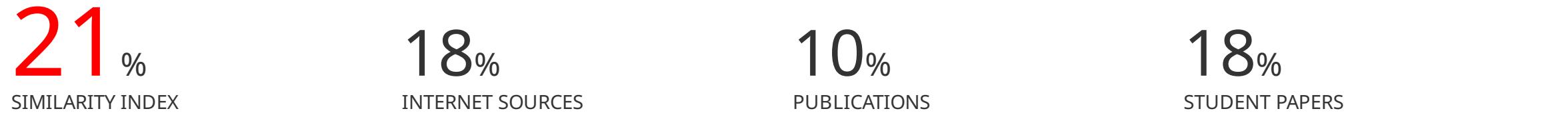
FEATURE IMPORTANCE



THANK YOU :)

~CYBER BULLS





PRIMARY SOURCES

- | | | |
|---|--|-----|
| 1 | www.irjmets.com
Internet Source | 16% |
| 2 | Submitted to University of Strathclyde
Student Paper | 2% |
| 3 | www.ijirset.com
Internet Source | 2% |
| 4 | Mohammad Alauthman, Ahmad al-Qerem, Someah Alangari, Ali Mohd Ali et al.
"Machine Learning for Accurate Software Development Cost Estimation in Economically and Technically Limited Environments", International Journal of Software Science and Computational Intelligence, 2023
Publication | 1% |

Exclude quotes

On

Exclude bibliography

On

Exclude matches

Off