

Data Collection and Preprocessing Phase

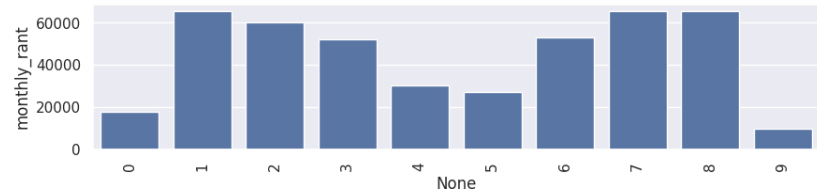
Date	28 June 2024
Team ID	740709
Project Title	House Rent Price Prediction Using Machine Learning
Maximum Marks	6 Marks

Data Exploration and Preprocessing Template

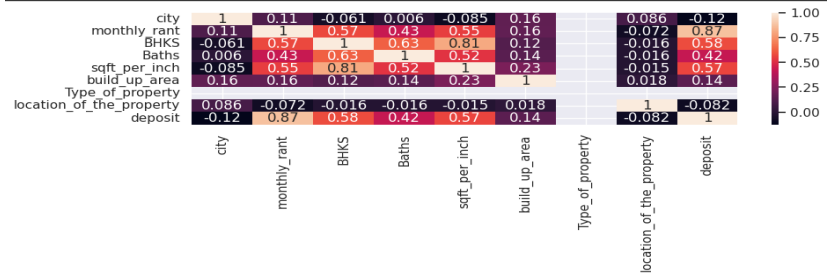
Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description																																																																																										
Data Overview	<div><div>▼ Descriptive Statistical</div><div><div><div></div><div>data.describe(include = "all")</div></div><table><thead><tr><th></th><th>city</th><th>monthly_rent</th><th>BMS</th><th>Baths</th><th>sqft_per_bch</th><th>build_up_area</th><th>Type_of_property</th><th>location_of_the_property</th><th>deposit</th></tr></thead><tbody><tr><td>count</td><td>82516.000000</td><td>82516.000000</td><td>82516.000000</td><td>82516.000000</td><td>82516.000000</td><td>82516.000000</td><td>82516.0</td><td>82516.000000</td><td>82516.000000</td></tr><tr><td>mean</td><td>1.682193</td><td>25780.219242</td><td>1.902936</td><td>6.483068</td><td>978.929627</td><td>1.733936</td><td>3.0</td><td>3684.225944</td><td>113721.581738</td></tr><tr><td>std</td><td>1.029670</td><td>17605.595475</td><td>0.755886</td><td>4.664011</td><td>451.079200</td><td>1.234548</td><td>0.0</td><td>2110.972993</td><td>86254.902980</td></tr><tr><td>min</td><td>0.000000</td><td>1000.000000</td><td>1.000000</td><td>0.000000</td><td>1.000000</td><td>0.000000</td><td>3.0</td><td>0.000000</td><td>0.000000</td></tr><tr><td>25%</td><td>1.000000</td><td>13000.000000</td><td>1.000000</td><td>0.000000</td><td>625.000000</td><td>1.000000</td><td>3.0</td><td>1990.000000</td><td>50000.000000</td></tr><tr><td>50%</td><td>2.000000</td><td>20000.000000</td><td>2.000000</td><td>8.000000</td><td>950.000000</td><td>2.000000</td><td>3.0</td><td>3482.000000</td><td>92500.000000</td></tr><tr><td>75%</td><td>2.000000</td><td>34000.000000</td><td>2.000000</td><td>8.000000</td><td>1200.000000</td><td>3.000000</td><td>3.0</td><td>5556.000000</td><td>150000.000000</td></tr><tr><td>max</td><td>3.000000</td><td>65500.000000</td><td>3.500000</td><td>20.000000</td><td>2382.500000</td><td>3.000000</td><td>3.0</td><td>7979.000000</td><td>300000.000000</td></tr></tbody></table></div></div>		city	monthly_rent	BMS	Baths	sqft_per_bch	build_up_area	Type_of_property	location_of_the_property	deposit	count	82516.000000	82516.000000	82516.000000	82516.000000	82516.000000	82516.000000	82516.0	82516.000000	82516.000000	mean	1.682193	25780.219242	1.902936	6.483068	978.929627	1.733936	3.0	3684.225944	113721.581738	std	1.029670	17605.595475	0.755886	4.664011	451.079200	1.234548	0.0	2110.972993	86254.902980	min	0.000000	1000.000000	1.000000	0.000000	1.000000	0.000000	3.0	0.000000	0.000000	25%	1.000000	13000.000000	1.000000	0.000000	625.000000	1.000000	3.0	1990.000000	50000.000000	50%	2.000000	20000.000000	2.000000	8.000000	950.000000	2.000000	3.0	3482.000000	92500.000000	75%	2.000000	34000.000000	2.000000	8.000000	1200.000000	3.000000	3.0	5556.000000	150000.000000	max	3.000000	65500.000000	3.500000	20.000000	2382.500000	3.000000	3.0	7979.000000	300000.000000
	city	monthly_rent	BMS	Baths	sqft_per_bch	build_up_area	Type_of_property	location_of_the_property	deposit																																																																																		
count	82516.000000	82516.000000	82516.000000	82516.000000	82516.000000	82516.000000	82516.0	82516.000000	82516.000000																																																																																		
mean	1.682193	25780.219242	1.902936	6.483068	978.929627	1.733936	3.0	3684.225944	113721.581738																																																																																		
std	1.029670	17605.595475	0.755886	4.664011	451.079200	1.234548	0.0	2110.972993	86254.902980																																																																																		
min	0.000000	1000.000000	1.000000	0.000000	1.000000	0.000000	3.0	0.000000	0.000000																																																																																		
25%	1.000000	13000.000000	1.000000	0.000000	625.000000	1.000000	3.0	1990.000000	50000.000000																																																																																		
50%	2.000000	20000.000000	2.000000	8.000000	950.000000	2.000000	3.0	3482.000000	92500.000000																																																																																		
75%	2.000000	34000.000000	2.000000	8.000000	1200.000000	3.000000	3.0	5556.000000	150000.000000																																																																																		
max	3.000000	65500.000000	3.500000	20.000000	2382.500000	3.000000	3.0	7979.000000	300000.000000																																																																																		
Univariate Analysis	<div><div><div><div>count</div><div>35000</div><div>30000</div><div>25000</div><div>20000</div><div>15000</div><div>10000</div><div>5000</div><div>0</div></div><div><div>0.0</div><div>1.0</div><div>2.0</div><div>3.0</div></div><div>build_up_area</div></div><div><div><div>count</div><div>30000</div><div>20000</div><div>10000</div><div>0</div></div><div><div>0.0</div><div>1.0</div><div>2.0</div><div>3.0</div></div><div>city</div></div><div><div><div>count</div><div>80000</div><div>70000</div><div>60000</div><div>50000</div><div>40000</div><div>30000</div><div>20000</div><div>10000</div><div>0</div></div><div><div>3.0</div></div><div>Type_of_property</div></div></div>																																																																																										

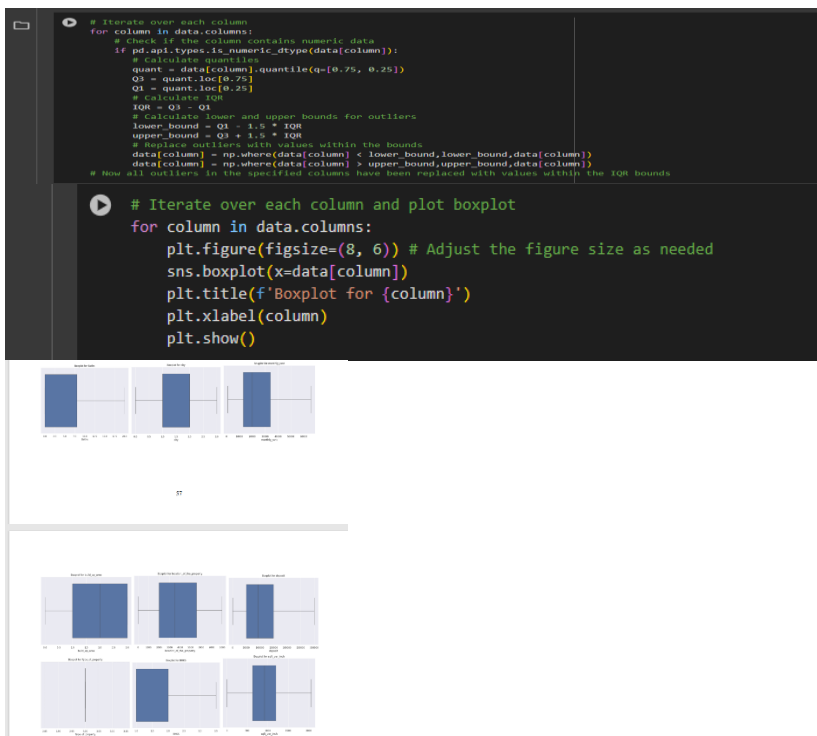
Bivariate Analysis



Multivariate Analysis



Outliers and Anomalies



Data Preprocessing Code Screenshots

Loading Data

Loading Dataset

```
#loading Dataset
data=pd.read_csv('/content/99acres_data (1).csv')
data
```

	city	monthly_rant	BHKS	Baths	sqft_per_inch	build_up_area	Type_of_property	location_of_the_property	deposit
0	mumbai	17500	1	2 Baths	470	Carpet Area	Residential	Kolshe Road	75000
1	mumbai	75000	3	3 Baths	1800	Super built-up Area	Residential	Sector 21 Nerul	400000
2	mumbai	60000	2	2 Baths	950	Super built-up Area	Residential	Wadala	200000
3	mumbai	52000	3	3 Baths	1300	Carpet Area	Residential	Hiranandani Estate	300000
4	mumbai	30000	1	1 Bath	550	Built-up Area	Residential	Kanjumarg (East)	150000
...
82511	bangalore	8400	1	1 Bath	1000	Carpet Area	Independent	Shivanagar	61000
82512	bangalore	9500	2	1 Bath	600	Built-up Area	Independent	2nd Stage Nagarbhavi	95000
82513	bangalore	7000	1	1 Bath	350	Built-up Area	Independent	Banagel Nagar	60000
82514	bangalore	5700	1	BHK	1000	Built-up Area	Independent	Channasandra	11400
82515	bangalore	5000	1	1R/1 Bath	200	Super built-up Area	Studio	Kamakshipalya	5000

82516 rows x 9 columns

Handling Missing Data

```
[ ] np.sum(data.isnull())
```

```
city          0
monthly_rant  0
BHKS          0
Baths         0
sqft_per_inch 0
build_up_area 0
Type_of_property 0
location_of_the_property 0
deposit       0
dtype: int64
```

Data Transformation

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()

[ ] data["city"]=le.fit_transform(data["city"])
    data["Baths"]=le.fit_transform(data["Baths"])
    data["build_up_area"]=le.fit_transform(data["build_up_area"])
    data["Type_of_property"]=le.fit_transform(data["Type_of_property"])
    data["location_of_the_property"]=le.fit_transform(data["location_of_the_property"])

[ ] data.head()
```

	city	monthly_rant	BHKS	Baths	sqft_per_inch	build_up_area	Type_of_property	location_of_the_property	deposit
0	2	17500	1	8	470	1	3	3024	75000
1	2	75000	3	11	1800	3	3	5558	400000
2	2	60000	2	8	950	3	3	7040	200000
3	2	52000	3	11	1300	1	3	2219	300000
4	2	30000	1	0	550	0	3	2756	150000

Save Processed Data

```
import pickle
pickle.dump(dt,open('mini.pkl','wb'))
```