



Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



Department of Computer Science and Engineering (Data Science)

MACHINE LEARNING MINIPROJECT

NAME:DIVYESH KHUNT

SAPID :60009210116

60009210116

Divyesh khunt

HOUSE PRICE PRDICTION

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

```
df=pd.read_csv("/content/Train.csv")
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 29451 entries, 0 to 29450
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   POSTED_BY             29451 non-null  object
1   UNDER_CONSTRUCTION   29451 non-null  int64
2   RERA                  29451 non-null  int64
3   BHK_NO.               29451 non-null  int64
4   BHK_OR_RK             29451 non-null  object
5   SQUARE_FT             29451 non-null  float64
6   READY_TO_MOVE         29451 non-null  int64
7   RESALE                29451 non-null  int64
8   ADDRESS               29451 non-null  object
9   LONGITUDE             29451 non-null  float64
10  LATITUDE              29451 non-null  float64
11  TARGET(PRICE_IN_LACS) 29451 non-null  float64
dtypes: float64(4), int64(5), object(3)
memory usage: 2.7+ MB
```

```
[20] df.columns
```

```
Index(['POSTED_BY', 'UNDER_CONSTRUCTION', 'RERA', 'BHK_NO.', 'BHK_OR_RK',
       'SQUARE_FT', 'READY_TO_MOVE', 'RESALE', 'ADDRESS', 'LONGITUDE',
       'LATITUDE', 'TARGET(PRICE_IN_LACS)'],
      dtype='object')
```



Department of Computer Science and Engineering (Data Science)

df.head(7)

	POSTED_BY	UNDER_CONSTRUCTION	RERA	BHK_NO.	BHK_OR_RK	SQUARE_FT	READY_TO_MOVE	RESALE	ADDRESS	LONGITUDE	LATITUDE	TARGET(PRICE_IN_LACS)
0	Owner	0	0	2	BHK	1300.236407	1	1	Ksf Layout,Bangalore	12.969910	77.597960	55.0
1	Dealer	0	0	2	BHK	1275.000000	1	1	Vishweshwara Nagar,Mysore	12.274538	76.644605	51.0
2	Owner	0	0	2	BHK	933.159722	1	1	Jigani,Bangalore	12.778033	77.632191	43.0
3	Owner	0	1	2	BHK	929.921143	1	1	Sector-1 Vaishali,Ghaziabad	28.642300	77.344500	62.5
4	Dealer	1	0	2	BHK	999.009247	0	1	New Town,Kolkata	22.592200	88.484911	60.5
5	Owner	0	0	3	BHK	1250.000000	1	1	South Chittoor,Kochi	10.033280	76.282571	42.0
6	Dealer	0	0	3	BHK	1495.053957	1	1	Sodala,Jaipur	26.916347	75.795600	66.5

```
# Splitting the "ADDRESS" column into "CITY" and "STATE" columns
df[["CITY", "STATE"]] = df["ADDRESS"].str.split(",", n=1, expand=True)
```

```
# Dropping "ADDRESS" column
df = df.drop(columns=["ADDRESS"])
```

df.head()

	POSTED_BY	UNDER_CONSTRUCTION	RERA	BHK_NO.	BHK_OR_RK	SQUARE_FT	READY_TO_MOVE	RESALE	LONGITUDE	LATITUDE	TARGET(PRICE_IN_LACS)	CITY	STATE
0	Owner	0	0	2	BHK	1300.236407	1	1	12.969910	77.597960	55.0	Ksf Layout	Bangalore
1	Dealer	0	0	2	BHK	1275.000000	1	1	12.274538	76.644605	51.0	Vishweshwara Nagar	Mysore
2	Owner	0	0	2	BHK	933.159722	1	1	12.778033	77.632191	43.0	Jigani	Bangalore
3	Owner	0	1	2	BHK	929.921143	1	1	28.642300	77.344500	62.5	Sector-1 Vaishali	Ghaziabad
4	Dealer	1	0	2	BHK	999.009247	0	1	22.592200	88.484911	60.5	New Town	Kolkata



df.dtypes



```
POSTED_BY          object
UNDER_CONSTRUCTION  int64
RERA                int64
BHK_NO.             int64
BHK_OR_RK           object
SQUARE_FT           float64
READY_TO_MOVE       int64
RESALE              int64
LONGITUDE           float64
LATITUDE            float64
TARGET(PRICE_IN_LACS) float64
CITY                object
STATE               object
dtype: object
```



Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



Department of Computer Science and Engineering (Data Science)

```
df.isnull().sum()
```

```

POSTED_BY      0
UNDER_CONSTRUCTION  0
RERA            0
BHK_NO.        0
BHK_OR_RK      0
SQUARE_FT      0
READY_TO_MOVE  0
RESALE         0
LONGITUDE      0
LATITUDE       0
TARGET(PRICE_IN_LACS)  0
CITY           0
STATE         0
dtype: int64

```

```
df.describe()
```

	UNDER_CONSTRUCTION	RERA	BHK_NO.	SQUARE_FT	READY_TO_MOVE	RESALE	LONGITUDE	LATITUDE	TARGET(PRICE_IN_LACS)
count	29451.000000	29451.000000	29451.000000	2.945100e+04	29451.000000	29451.000000	29451.000000	29451.000000	29451.000000
mean	0.179756	0.317918	2.392279	1.980217e+04	0.820244	0.929578	21.300255	76.837695	142.898746
std	0.383991	0.465675	0.879091	1.901335e+06	0.383991	0.255861	6.205306	10.557747	656.880713
min	0.000000	0.000000	1.000000	3.000000e+00	0.000000	0.000000	-37.713008	-121.761248	0.250000
25%	0.000000	0.000000	2.000000	9.000211e+02	1.000000	1.000000	18.452663	73.798100	38.000000
50%	0.000000	0.000000	2.000000	1.175057e+03	1.000000	1.000000	20.750000	77.324137	62.000000
75%	0.000000	1.000000	3.000000	1.550688e+03	1.000000	1.000000	26.900926	77.828740	100.000000
max	1.000000	1.000000	20.000000	2.545455e+08	1.000000	1.000000	59.912884	152.962676	30000.000000

```

#top 5 expensive houses
df.nlargest(n=5, columns='TARGET(PRICE_IN_LACS)')

```

	POSTED_BY	UNDER_CONSTRUCTION	RERA	BHK_NO.	BHK_OR_RK	SQUARE_FT	READY_TO_MOVE	RESALE	LONGITUDE	LATITUDE	TARGET(PRICE_IN_LACS)	CITY	STATE
11238	Dealer	0	0	3	BHK	1.875000e+08	1	1	13.018900	77.596300	30000.0	R.T. Nagar	Bangalore
10745	Owner	0	0	3	BHK	2.545455e+08	1	1	13.021000	77.938000	28000.0	Malur	Bangalore
15724	Owner	0	0	2	BHK	8.064516e+07	1	1	12.795926	77.331535	25000.0	Lakkasandra	Bangalore
5967	Dealer	1	1	2	BHK	5.422570e+04	0	0	19.180000	72.963330	9990.0	Thane West	Lalitpur
10637	Dealer	1	1	3	BHK	8.322835e+04	0	0	18.627000	73.782900	9910.0	Chinchwad	Pune



Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

(Autonomous College Affiliated to the University of Mumbai)

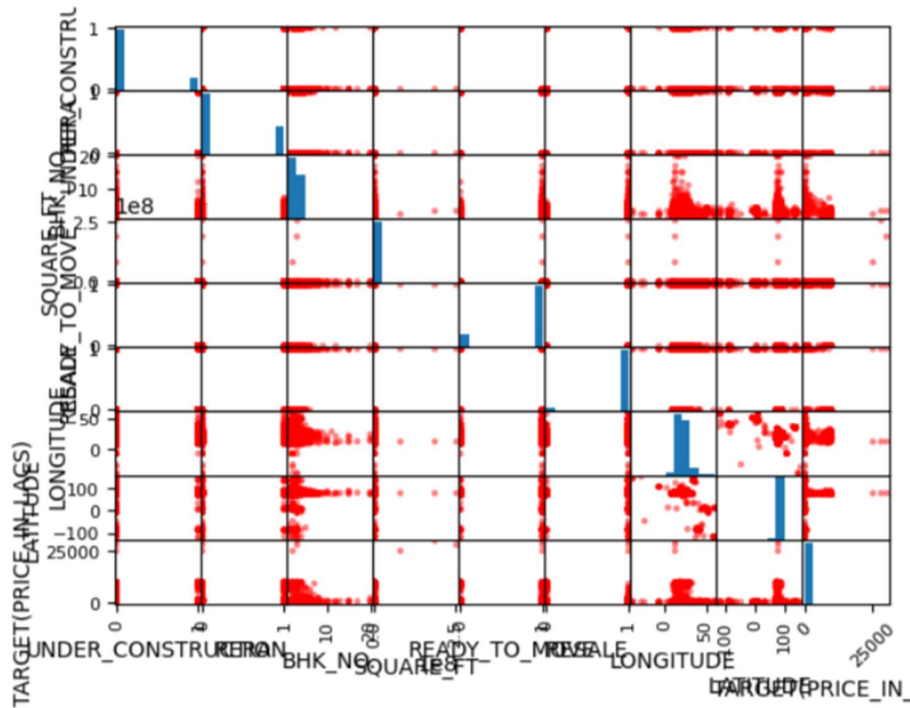
NAAC Accredited with "A" Grade (CGPA : 3.18)



Department of Computer Science and Engineering (Data Science)

```
[58] pd.plotting.scatter_matrix(df,color="red")
```

```
plt.figure(figsize=(50,24))  
plt.xticks(rotation=45)  
plt.yticks(rotation=45)  
plt.show()
```





Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



Department of Computer Science and Engineering (Data Science)

▶ # Compute the correlation matrix

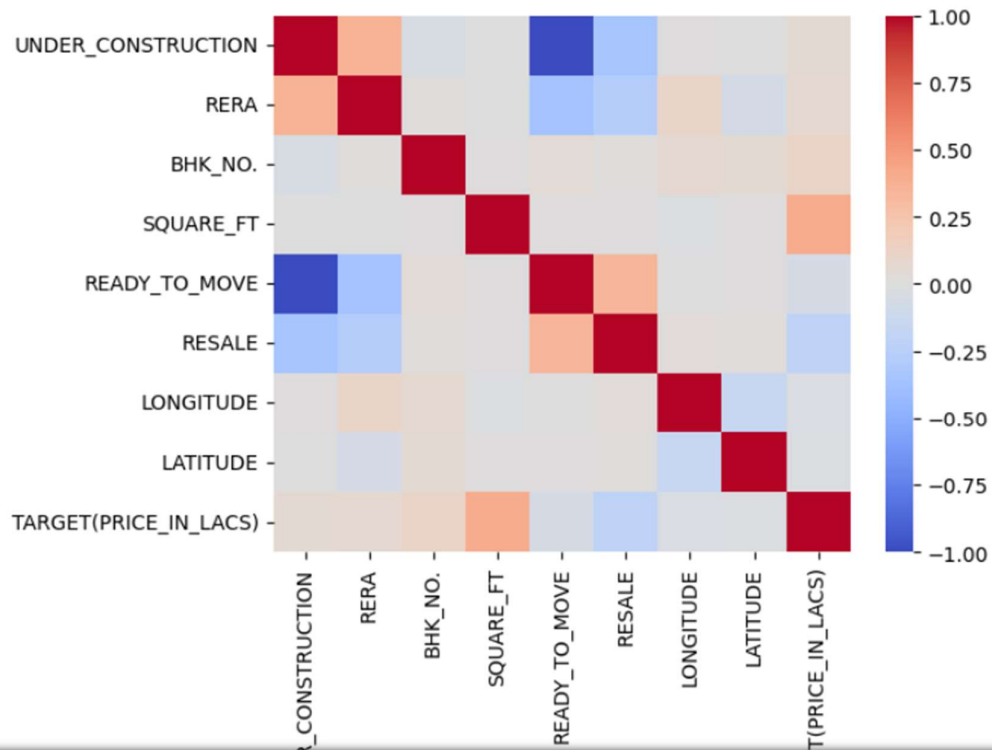
```
corr_matrix = df.corr()
```

```
sns.heatmap(corr_matrix, cmap='coolwarm')
```

```
sns.heatmap(corr_matrix, cmap='coolwarm')
```

```
plt.show()
```

↳ <ipython-input-84-6f04e91fc256>:2: FutureWarning: The default value of numeric_only in DataFrame.corr_matrix = df.corr()





Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

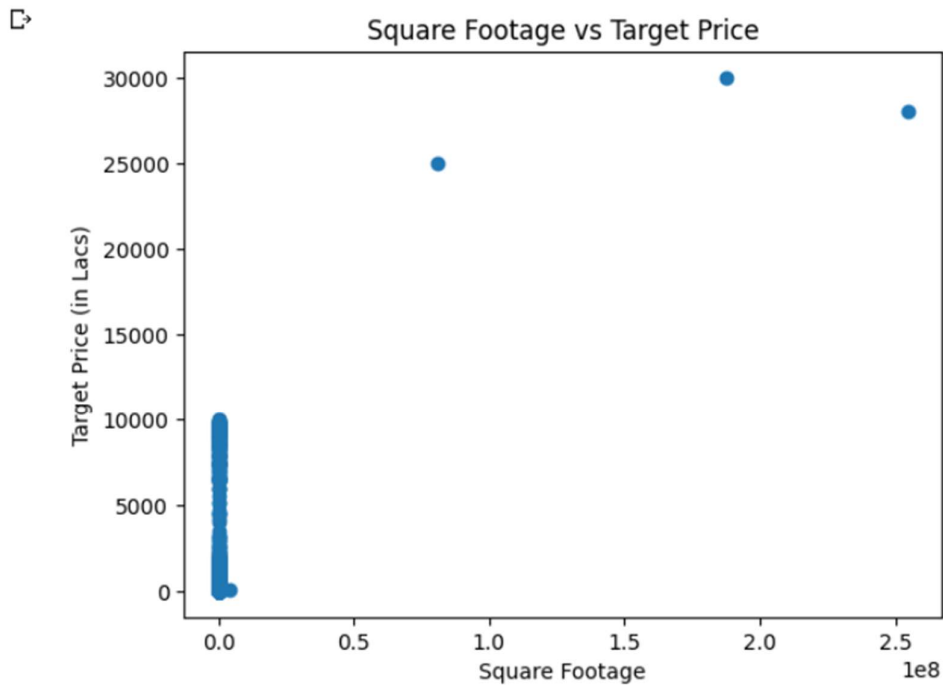
(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



Department of Computer Science and Engineering (Data Science)

```
plt.scatter(df["SQUARE_FT"], df["TARGET(PRICE_IN_LACS)"])  
plt.xlabel("Square Footage")  
plt.ylabel("Target Price (in Lacs)")  
plt.title("Square Footage vs Target Price")  
plt.show()
```





Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

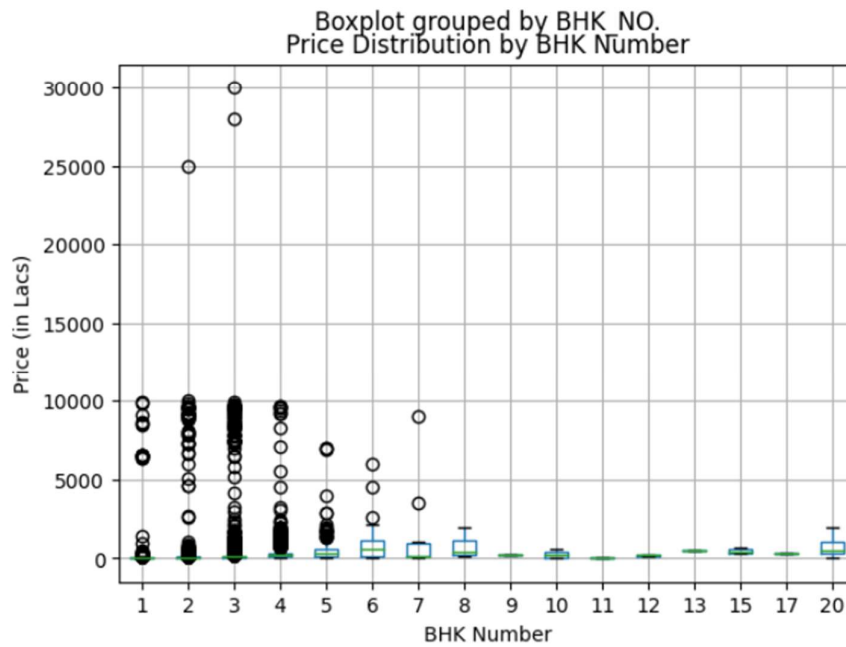
(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



Department of Computer Science and Engineering (Data Science)

```
# Create a box plot of price by BHK number
df.boxplot(column='TARGET(PRICE_IN_LACS)', by='BHK_NO.')
plt.title('Price Distribution by BHK Number')
plt.xlabel('BHK Number')
plt.ylabel('Price (in Lacs)')
plt.show()
```





Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

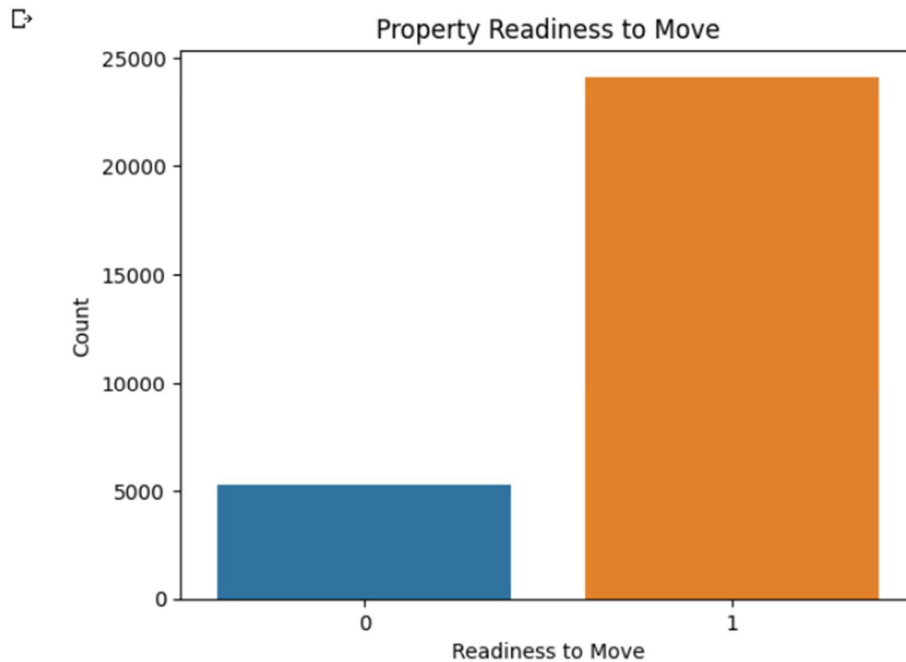
(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



Department of Computer Science and Engineering (Data Science)

```
sns.countplot(x='READY_TO_MOVE', data=df)
plt.title('Property Readiness to Move')
plt.xlabel('Readiness to Move')
plt.ylabel('Count')
plt.show()
```





Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

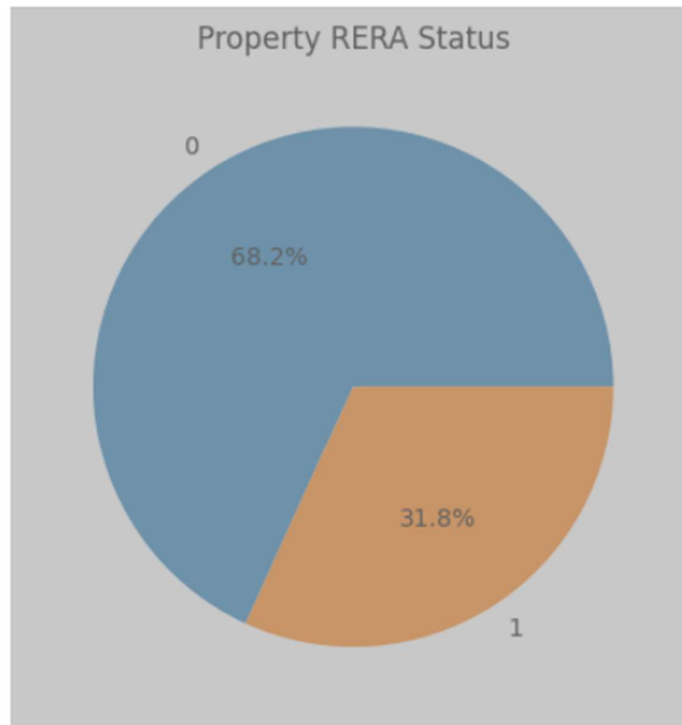
(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



Department of Computer Science and Engineering (Data Science)

```
counts = df['RERA'].value_counts()  
plt.pie(counts, labels=counts.index, autopct='%1.1f%%')  
plt.title('Property RERA Status')  
plt.show()
```





Department of Computer Science and Engineering (Data Science)

```
grouped = df.groupby(['BHK_NO.', 'POSTED_BY'])  
# Count the number of properties in each group  
counts = grouped.size()  
# Create a pie chart of property ownership by BHK number  
for bhk in counts.index.levels[0]:  
    bhk_counts = counts.loc[bhk]  
    plt.figure()  
    plt.pie(bhk_counts, labels=bhk_counts.index, autopct='%1.1f%%')  
    plt.title(f'Property Ownership for {bhk} BHK')  
  
# Show the plots  
plt.show()
```

