Statistical Measures

[5]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[ ]:
```
#The dataset house_price.csv which contains property prices in the city of banglore. Examine price per square feet
```

[7]:
```
#Load the dataset
df=pd.read_csv("C:\\Users\\hp\\Downloads\\house_price.csv")
df
```

[7]:

	location	size	total_sqft	bath	price	bhk	price_per_sqft
0	Electronic City Phase II	2 BHK	1056.0	2.0	39.07	2	3699
1	Chikka Tirupathi	4 Bedroom	2600.0	5.0	120.00	4	4615
2	Uttarahalli	3 BHK	1440.0	2.0	62.00	3	4305
3	Lingadheeranahalli	3 BHK	1521.0	3.0	95.00	3	6245
4	Kothanur	2 BHK	1200.0	2.0	51.00	2	4250
...	...	...	...	...	...	...	...
13195	Whitefield	5 Bedroom	3453.0	4.0	231.00	5	6689
13196	other	4 BHK	3600.0	5.0	400.00	4	11111
13197	Raja Rajeshwari Nagar	2 BHK	1141.0	2.0	60.00	2	5258
13198	Padmanabhanagar	4 BHK	4689.0	4.0	488.00	4	10407

[ ]:
```
#Q1.Perform Basic EDA
```

[6]:
```
df.head()
```

[6]:

	location	size	total_sqft	bath	price	bhk	price_per_sqft
0	Electronic City Phase II	2 BHK	1056.0	2.0	39.07	2	3699
1	Chikka Tirupathi	4 Bedroom	2600.0	5.0	120.00	4	4615
2	Uttarahalli	3 BHK	1440.0	2.0	62.00	3	4305
3	Lingadheeranahalli	3 BHK	1521.0	3.0	95.00	3	6245
4	Kothanur	2 BHK	1200.0	2.0	51.00	2	4250

[8]:
```
df.tail()
```

[8]:

	location	size	total_sqft	bath	price	bhk	price_per_sqft
13195	Whitefield	5 Bedroom	3453.0	4.0	231.0	5	6689
13196	other	4 BHK	3600.0	5.0	400.0	4	11111
13197	Raja Rajeshwari Nagar	2 BHK	1141.0	2.0	60.0	2	5258
13198	Padmanabhanagar	4 BHK	4689.0	4.0	488.0	4	10407
13199	Doddathoguru	1 BHK	550.0	1.0	17.0	1	3090

[16]:
```
df.describe()
```

jupyter Assignment-Statistical Measures Last Checkpoint: 20 hours ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[16]:
```
df.describe()
```

[16]:

	total_sqft	bath	price	bhk	price_per_sqft
count	13200.000000	13200.000000	13200.000000	13200.000000	1.320000e+04
mean	1555.302783	2.691136	112.276178	2.800833	7.920337e+03
std	1237.323445	1.338915	149.175995	1.292843	1.067272e+05
min	1.000000	1.000000	8.000000	1.000000	2.670000e+02
25%	1100.000000	2.000000	50.000000	2.000000	4.267000e+03
50%	1275.000000	2.000000	71.850000	3.000000	5.438000e+03
75%	1672.000000	3.000000	120.000000	3.000000	7.317000e+03
max	52272.000000	40.000000	3600.000000	43.000000	1.200000e+07

[10]:
```
df.shape
```

[10]:
```
(13200, 7)
```

[12]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13200 entries, 0 to 13199
Data columns (total 7 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   location        13200 non-null  object
 1   size            13200 non-null  object
 2   total_sqft      13200 non-null  float64
```

[28]:
```
import pandas as pd
import warnings
warnings.filterwarnings("ignore")
```

[10]:
```
df.shape
```

[10]: (13200, 7)

[12]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13200 entries, 0 to 13199
Data columns (total 7 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   location        13200 non-null  object 
 1   size            13200 non-null  object 
 2   total_sqft      13200 non-null  float64
 3   bath            13200 non-null  float64
 4   price           13200 non-null  float64
 5   bhk             13200 non-null  int64  
 6   price_per_sqft  13200 non-null  int64  
dtypes: float64(3), int64(2), object(2)
memory usage: 722.0+ KB
```

[18]:
```
df.columns
```

[18]:
```
Index(['location', 'size', 'total_sqft', 'bath', 'price', 'bhk',
       'price_per_sqft'],
      dtype='object')
```

[20]:
```
df.nunique()
```

[20]:
```
location          241
size               31
total_sqft       1972
bath               19
```

jupyter Assignment-Statistical Measures Last Checkpoint: 20 hours ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

```
 3   bath            13200 non-null  float64
 4   price           13200 non-null  float64
 5   bhk             13200 non-null  int64
 6   price_per_sqft  13200 non-null  int64
dtypes: float64(3), int64(2), object(2)
memory usage: 722.0+ KB
```

[18]:
```
df.columns
```

[18]:
```
Index(['location', 'size', 'total_sqft', 'bath', 'price', 'bhk',
       'price_per_sqft'],
      dtype='object')
```

[20]:
```
df.nunique()
```

[20]:
```
location          241
size               31
total_sqft       1972
bath               19
price            1952
bhk                19
price_per_sqft   4951
dtype: int64
```

[11]:
```
df.isnull().sum()
```

[11]:
```
location         0
size             0
total_sqft       0
bath             0
price            0
bhk              0
price_per_sqft   0
dtype: int64
```

jupyter Assignment-Statistical Measures Last Checkpoint: 20 hours ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[20]:
```
df.nunique()
```

[20]:
```
location          241
size               31
total_sqft       1972
bath               19
price            1952
bhk                19
price_per_sqft    4951
dtype: int64
```

[11]:
```
df.isnull().sum()
```

[11]:
```
location          0
size              0
total_sqft        0
bath              0
price             0
bhk               0
price_per_sqft    0
dtype: int64
```

[209]:
```
print("Duplicates:" , df.duplicated().sum())
```

```
Duplicates: 1049
```

[ ]:
```
#Q2.Detect the outliers using following methods and remove it using methods like
#trimming / capping/ imputation using mean or median
#a) Mean and Standard deviation
#b)Percentile method
#c) IQR(Inter quartile range method)
#d) Z Score method
```

Skilling | Home | Statistics | Project | Assignment-Statistic | Untitled3 | Home

localhost:8888/notebooks/Assignment-Statistical%20Measures.ipynb?

jupyter Assignment-Statistical Measures Last Checkpoint: 23 hours ago

File Edit View Run Kernel Settings Help

Trusted

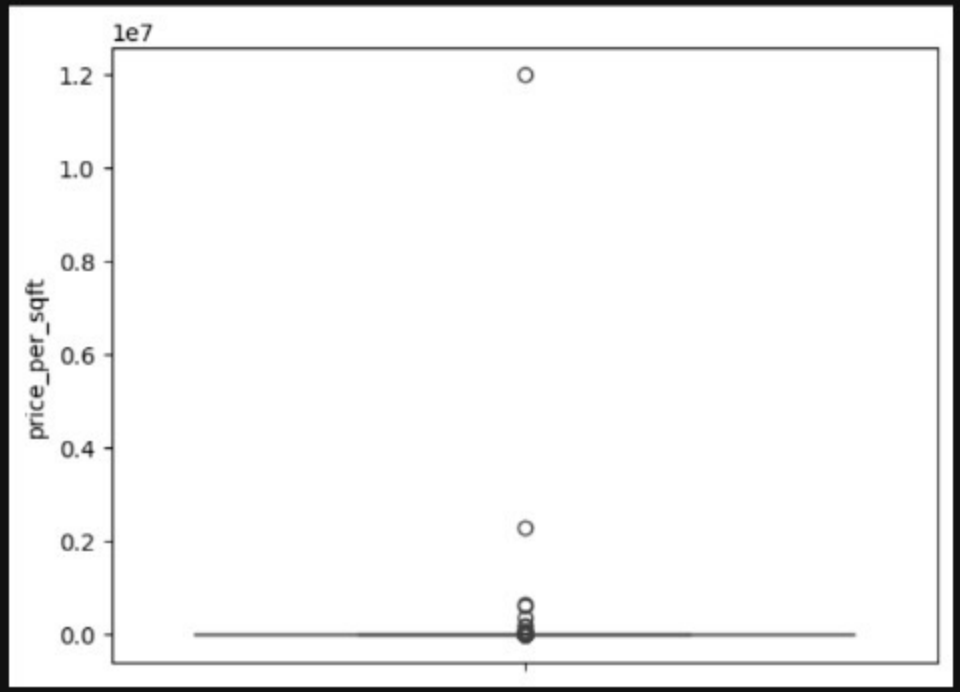
Code

JupyterLab Python 3 (ipykernel)

Duplicates: 1049

[16]:
```
sns.boxplot(df['price_per_sqft'])
```

[16]: <Axes: ylabel='price_per_sqft'>



[52]:
```
for i in df.columns:
    plt.figure(figsize=(12,5))

    # Histogram
    plt.subplot(1, 2, 1)
    sns.histplot(df[i], kde=True)
    plt.xlabel(i)
```

26°C Partly cloudy

Search

ENG IN

21:28 21-12-2024

Skilling | Home | Statistics | Project | Assignment-Statistic | Untitled3 | Home

localhost:8888/notebooks/Assignment-Statistical%20Measures.ipynb?

jupyter Assignment-Statistical Measures Last Checkpoint: 23 hours ago

File Edit View Run Kernel Settings Help

Trusted

Code

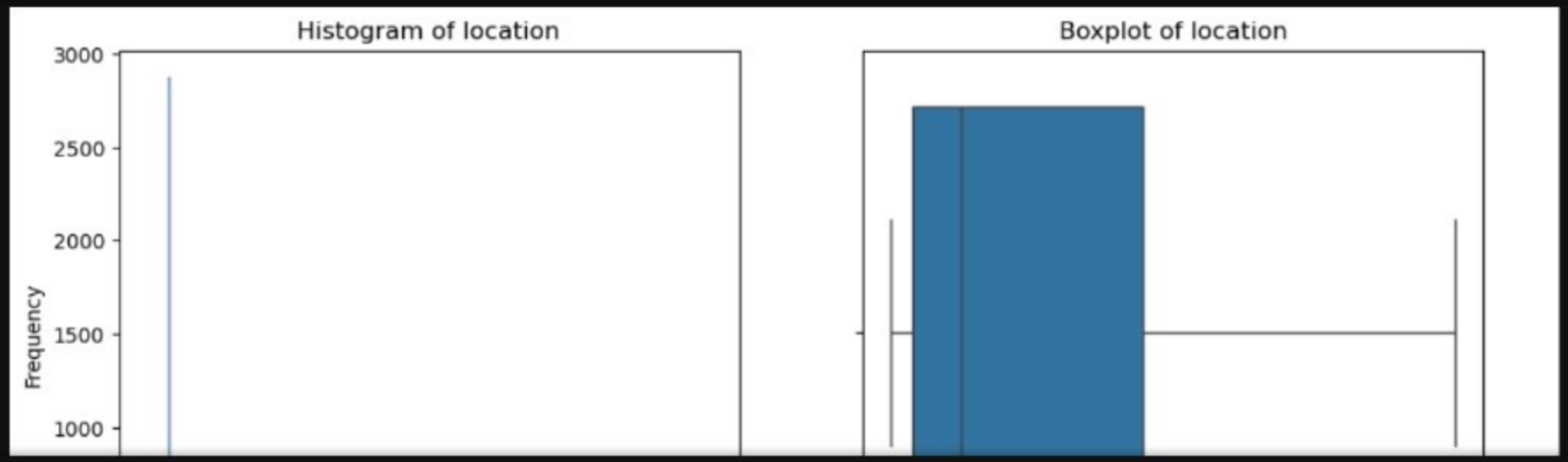
JupyterLab Python 3 (ipykernel)

[52]:
```
for i in df.columns:
    plt.figure(figsize=(12,5))

    # Histogram
    plt.subplot(1, 2, 1)
    sns.histplot(df[i], kde=True)
    plt.xlabel(i)
    plt.ylabel('Frequency')
    plt.title(f'Histogram of {i}')

    # Boxplot
    plt.subplot(1, 2, 2)
    sns.boxplot(x=df[i])
    plt.title(f'Boxplot of {i}')

    plt.show()
```



26°C Partly cloudy

Search

ENG IN

21:28 21-12-2024

Skilling | Home | Statistics | Project | Assignment-Statistic | Untitled3 | Home

localhost:8888/notebooks/Assignment-Statistical%20Measures.ipynb?

jupyter Assignment-Statistical Measures Last Checkpoint: 23 hours ago

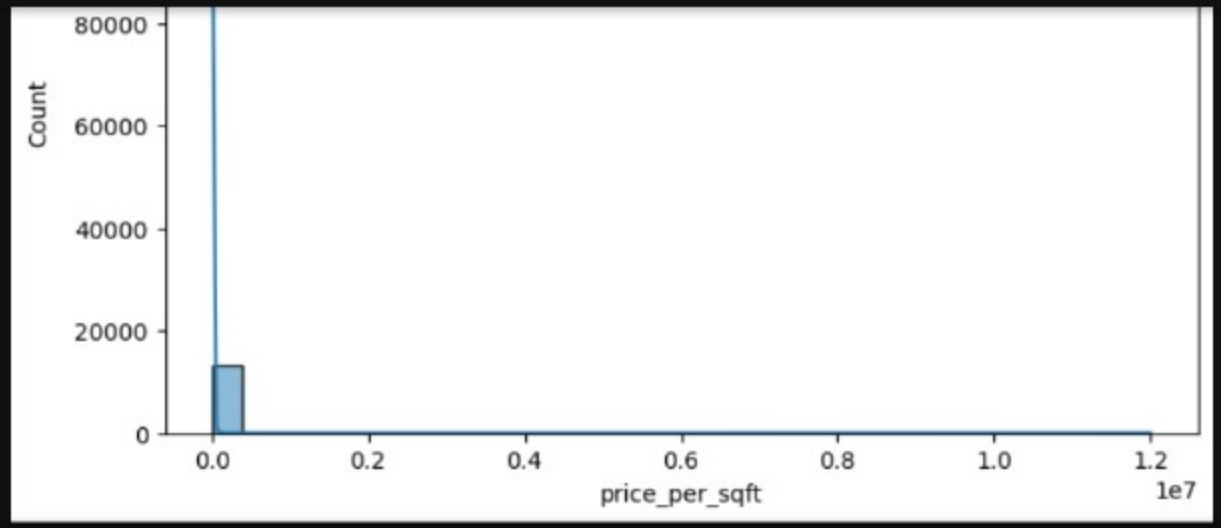
File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

[58]:
```
#Plot Histograms for Numerical Column
Numerical_columns = df.select_dtypes(include=[np.number]).columns
for col in Numerical_columns:
    plt.figure(figsize=(8,5))
    sns.histplot(df[col] , kde=True , bins=30)
    plt.title(f'Histogram of {col}')
    plt.show()
```



[70]:
```
df['price_per_sqft'].unique()
```

[70]:
```
array([ 3699,  4615,  4305, ...,  7423,  5020, 10407], dtype=int64)
```

[ ]:
```
#Q2.Detect the outliers using following methods and remove it using methods like
#trimming / capping/ imputation using mean or median
#a) Mean and Standard deviation
#b)Percentile method
#c) IQR(Inter quantile range method)
```

26°C Partly cloudy

Search

ENG IN

21:28 21-12-2024

Mean and Standard Deviation

[23]:
```
# finding the limits
Mean=df['price_per_sqft'].mean()
Std_dev=df['price_per_sqft'].std()
upper_limit =Mean + 3*Std_dev
lower_limit= Mean - 3*Std_dev
```

[27]:
```
#Detect Outliers
Outliers = df[(df['price_per_sqft'] < lower_limit) | (df['price_per_sqft'] > upper_limit)]
Outliers
```

[27]:

	location	size	total_sqft	bath	price	bhk	price_per_sqft
345	other	3 Bedroom	11.0	3.0	74.0	3	672727
1106	other	5 Bedroom	24.0	2.0	150.0	5	625000
4044	Sarjapur Road	4 Bedroom	1.0	4.0	120.0	4	12000000
4924	other	7 BHK	5.0	7.0	115.0	7	2300000
11447	Whitefield	4 Bedroom	60.0	4.0	218.0	4	363333

[41]:
```
#Remove Outliers using trimming
df_trimmed = df[(df['price_per_sqft'] >= lower_limit) & (df['price_per_sqft'] <= upper_limit)]
df_trimmed
```

[41]:

	location	size	total_sqft	bath	price	bhk	price_per_sqft
0	Electronic City Phase II	2 BHK	1056.0	2.0	39.07	2	3699
1	Chikka Tirupathi	4 Bedroom	2600.0	5.0	120.00	4	4615

4924	other	7 BHK	5.0	7.0	115.0	7	2300000
11447	Whitefield	4 Bedroom	60.0	4.0	218.0	4	363333

[41]:
```
#Remove Outliers using trimming
df_trimmed = df[(df['price_per_sqft'] >= lower_limit) & (df['price_per_sqft'] <= upper_limit)]
df_trimmed
```

[41]:

	location	size	total_sqft	bath	price	bhk	price_per_sqft
0	Electronic City Phase II	2 BHK	1056.0	2.0	39.07	2	3699
1	Chikka Tirupathi	4 Bedroom	2600.0	5.0	120.00	4	4615
2	Uttarahalli	3 BHK	1440.0	2.0	62.00	3	4305
3	Lingadheeranahalli	3 BHK	1521.0	3.0	95.00	3	6245
4	Kothanur	2 BHK	1200.0	2.0	51.00	2	4250
...	...	...	...	...	...	...	...
13195	Whitefield	5 Bedroom	3453.0	4.0	231.00	5	6689
13196	other	4 BHK	3600.0	5.0	400.00	4	11111
13197	Raja Rajeshwari Nagar	2 BHK	1141.0	2.0	60.00	2	5258
13198	Padmanabhanagar	4 BHK	4689.0	4.0	488.00	4	10407
13199	Doddathoguru	1 BHK	550.0	1.0	17.00	1	3090

13195 rows × 7 columns

Percentile Method

Percentile Method

[63]:
```
#Calculate Percentiles
lower_limit=df['price_per_sqft'].quantile(0.05)
Upper_limit=df['price_per_sqft'].quantile(0.95)
```

[65]:
```
#Detect Outliers
Outliers = df[(df['price_per_sqft'] < lower_limit) | (df['price_per_sqft'] > Upper_limit)]
Outliers
```

[65]:

	location	size	total_sqft	bath	price	bhk	price_per_sqft
7	Rajaji Nagar	4 BHK	3300.0	4.0	600.0	4	18181
9	other	6 Bedroom	1020.0	6.0	370.0	6	36274
20	Kengeri	1 BHK	600.0	1.0	15.0	1	2500
41	Sarjapur Road	3 BHK	1254.0	3.0	38.0	3	3030
45	HSR Layout	8 Bedroom	600.0	9.0	200.0	8	33333
...	...	...	...	...	...	...	...
13157	other	7 Bedroom	1400.0	7.0	218.0	7	15571
13185	Hulimavu	1 BHK	500.0	1.0	220.0	1	44000
13186	other	4 Bedroom	1200.0	5.0	325.0	4	27083
13191	Ramamurthy Nagar	7 Bedroom	1500.0	9.0	250.0	7	16666
13199	Doddathoguru	1 BHK	550.0	1.0	17.0	1	3090

13191	Ramamurthy Nagar	7 Bedroom	1500.0	9.0	250.0	7	16666
13199	Doddathoguru	1 BHK	550.0	1.0	17.0	1	3090

1320 rows × 7 columns

•[133]:
```
#Remove Ouliers
df_trimmed = df[(df['price_per_sqft'] >= lower_limit) & (df['price_per_sqft'] <= Upper_limit)]
df_trimmed
```

[133]:

	location	size	total_sqft	bath	price	bhk	price_per_sqft	zscore
20	Kengeri	1 BHK	600.0	1.0	15.00	1	2500	-0.050789
99	Chandapura	2 BHK	650.0	1.0	17.00	2	2615	-0.049711
130	Electronic City	2 BHK	880.0	1.0	16.50	2	1875	-0.056645
169	Attibele	1 BHK	450.0	1.0	11.00	1	2444	-0.051314
237	Chandapura	1 BHK	645.0	1.0	16.45	1	2550	-0.050320
...	...	...	...	...	...	...	...	...
12897	Kammasandra	3 BHK	1616.0	3.0	40.00	3	2475	-0.051023
12909	Bommasandra	2 BHK	950.0	2.0	25.00	2	2631	-0.049561
13019	Electronic City	2 BHK	750.0	2.0	19.50	2	2600	-0.049852
13028	Chandapura	3 BHK	1095.0	2.0	28.00	3	2557	-0.050255
13105	Chandapura	1 BHK	520.0	1.0	14.04	1	2700	-0.048915

230 rows × 8 columns

Inter Quartile Range Method

[74]:
```
#Calculate IQR
q1=df['price_per_sqft'].quantile(0.25)
q3=df['price_per_sqft'].quantile(0.75)
IQR = q3-q1
lower_limit=q1 - 1.5*IQR
Upper_limit=q3 - 1.5*IQR
```

[78]:
```
#Detect Outliers
Outliers=df[(df['price_per_sqft'] < lower_limit) | (df['price_per_sqft'] > Upper_limit)]
Outliers
```

[78]:

	location	size	total_sqft	bath	price	bhk	price_per_sqft
0	Electronic City Phase II	2 BHK	1056.0	2.0	39.07	2	3699
1	Chikka Tirupathi	4 Bedroom	2600.0	5.0	120.00	4	4615
2	Uttarahalli	3 BHK	1440.0	2.0	62.00	3	4305
3	Lingadheeranahalli	3 BHK	1521.0	3.0	95.00	3	6245
4	Kothanur	2 BHK	1200.0	2.0	51.00	2	4250
...	...	...	...	...	...	...	...
13195	Whitefield	5 Bedroom	3453.0	4.0	231.00	5	6689
13196	other	4 BHK	3600.0	5.0	400.00	4	11111
13197	Raja Rajeshwari Nagar	2 BHK	1141.0	2.0	60.00	2	5258
13198	Padmanabhanagar	4 BHK	4689.0	4.0	488.00	4	10407

13198	Padmanabhanagar	4 BHK	4689.0	4.0	488.00	4	10407
13199	Doddathoguru	1 BHK	550.0	1.0	17.00	1	3090

12970 rows × 7 columns

[101]:
```
#Removing Outliers
df_trimmed = df.loc[(df['price_per_sqft'] >= lower_limit) & (df['price_per_sqft'] <= Upper_limit)]
df_trimmed
```

[101]:

	location	size	total_sqft	bath	price	bhk	price_per_sqft	zscore
20	Kengeri	1 BHK	600.0	1.0	15.00	1	2500	-0.050789
99	Chandapura	2 BHK	650.0	1.0	17.00	2	2615	-0.049711
130	Electronic City	2 BHK	880.0	1.0	16.50	2	1875	-0.056645
169	Attibele	1 BHK	450.0	1.0	11.00	1	2444	-0.051314
237	Chandapura	1 BHK	645.0	1.0	16.45	1	2550	-0.050320
...	...	...	...	...	...	...	...	...
12897	Kammasandra	3 BHK	1616.0	3.0	40.00	3	2475	-0.051023
12909	Bommasandra	2 BHK	950.0	2.0	25.00	2	2631	-0.049561
13019	Electronic City	2 BHK	750.0	2.0	19.50	2	2600	-0.049852
13028	Chandapura	3 BHK	1095.0	2.0	28.00	3	2557	-0.050255
13105	Chandapura	1 BHK	520.0	1.0	14.04	1	2700	-0.048915

230 rows × 8 columns

Zscore Method

[97]:
```
#Find the Limits
Upper_limits = df['price_per_sqft'].mean() + 3*df['price_per_sqft'].std()
lower_limits = df['price_per_sqft'].mean() - 3*df['price_per_sqft'].std()
print('Upper_limit:',Upper_limit)
print('lower_limit:',lower_limit)
```

```
Upper_limit: 2742.0
lower_limit: -308.0
```

[113]:
```
#Detect Outliers
Outliers=df.loc[(df['price_per_sqft'] < lower_limit) | (df['price_per_sqft'] > Upper_limit)]
Outliers
```

[113]:

	location	size	total_sqft	bath	price	bhk	price_per_sqft	zscore
0	Electronic City Phase II	2 BHK	1056.0	2.0	39.07	2	3699	-0.039554
1	Chikka Tirupathi	4 Bedroom	2600.0	5.0	120.00	4	4615	-0.030971
2	Uttarahalli	3 BHK	1440.0	2.0	62.00	3	4305	-0.033876
3	Lingadheeranahalli	3 BHK	1521.0	3.0	95.00	3	6245	-0.015698
4	Kothanur	2 BHK	1200.0	2.0	51.00	2	4250	-0.034391
...	...	...	...	...	...	...	...	...
13195	Whitefield	5 Bedroom	3453.0	4.0	231.00	5	6689	-0.011538
13196	other	4 BHK	3600.0	5.0	400.00	4	11111	0.029897
13197	Raja Rajeshwari Nagar	2 BHK	1141.0	2.0	60.00	2	5258	-0.024946
13198	Padmanabhanagar	4 BHK	4689.0	4.0	488.00	4	10407	0.023300

13198	Padmanabhanagar	4 BHK	4689.0	4.0	488.00	4	10407	0.023300
13199	Doddathoguru	1 BHK	550.0	1.0	17.00	1	3090	-0.045260

12970 rows × 8 columns

[141]:
```
df_trimmed = df.loc[(df['price_per_sqft'] >= lower_limit) & (df['price_per_sqft'] <= Upper_limit)]
df_trimmed
```

[141]:

	location	size	total_sqft	bath	price	bhk	price_per_sqft	zscore
20	Kengeri	1 BHK	600.0	1.0	15.00	1	2500	-0.050789
99	Chandapura	2 BHK	650.0	1.0	17.00	2	2615	-0.049711
130	Electronic City	2 BHK	880.0	1.0	16.50	2	1875	-0.056645
169	Attibele	1 BHK	450.0	1.0	11.00	1	2444	-0.051314
237	Chandapura	1 BHK	645.0	1.0	16.45	1	2550	-0.050320
...	...	...	...	...	...	...	...	...
12897	Kammasandra	3 BHK	1616.0	3.0	40.00	3	2475	-0.051023
12909	Bommasandra	2 BHK	950.0	2.0	25.00	2	2631	-0.049561
13019	Electronic City	2 BHK	750.0	2.0	19.50	2	2600	-0.049852
13028	Chandapura	3 BHK	1095.0	2.0	28.00	3	2557	-0.050255
13105	Chandapura	1 BHK	520.0	1.0	14.04	1	2700	-0.048915

230 rows × 8 columns

Home | Project Python | Statistics | Assignment-Statistical Measures

localhost:8889/notebooks/Assignment-Statistical%20Measures.ipynb?

jupyter Assignment-Statistical Measures Last Checkpoint: 20 hours ago

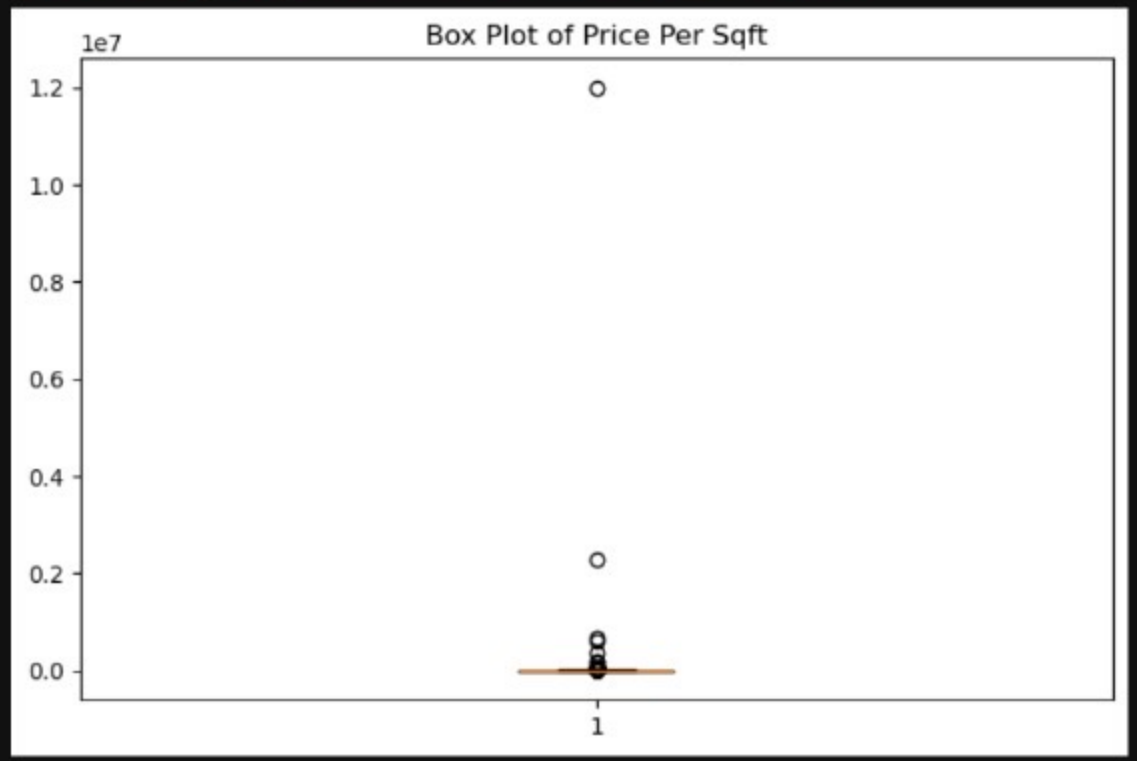
Trusted

File Edit View Run Kernel Settings Help

Code JupyterLab Python 3 (ipykernel)

[ ]:
```
#Q3.Create a box plot and use this to determine which method seems to work best to remove outliers for this data
```

•[153]:
```
#Create box Plot
plt.figure(figsize=(8,5))
plt.boxplot(df['price_per_sqft'])
plt.title('Box Plot of Price Per Sqft')
plt.show()
# For this data, The IQR method seems to work bestv to remove outliers
```



EUR/INR +0.48%

Search

ENG IN

19:00 21-12-2024

Home | Project Python | Statistics | Assignment-Statistical Measures

localhost:8889/notebooks/Assignment-Statistical%20Measures.ipynb?

jupyter Assignment-Statistical Measures Last Checkpoint: 20 hours ago

Trusted

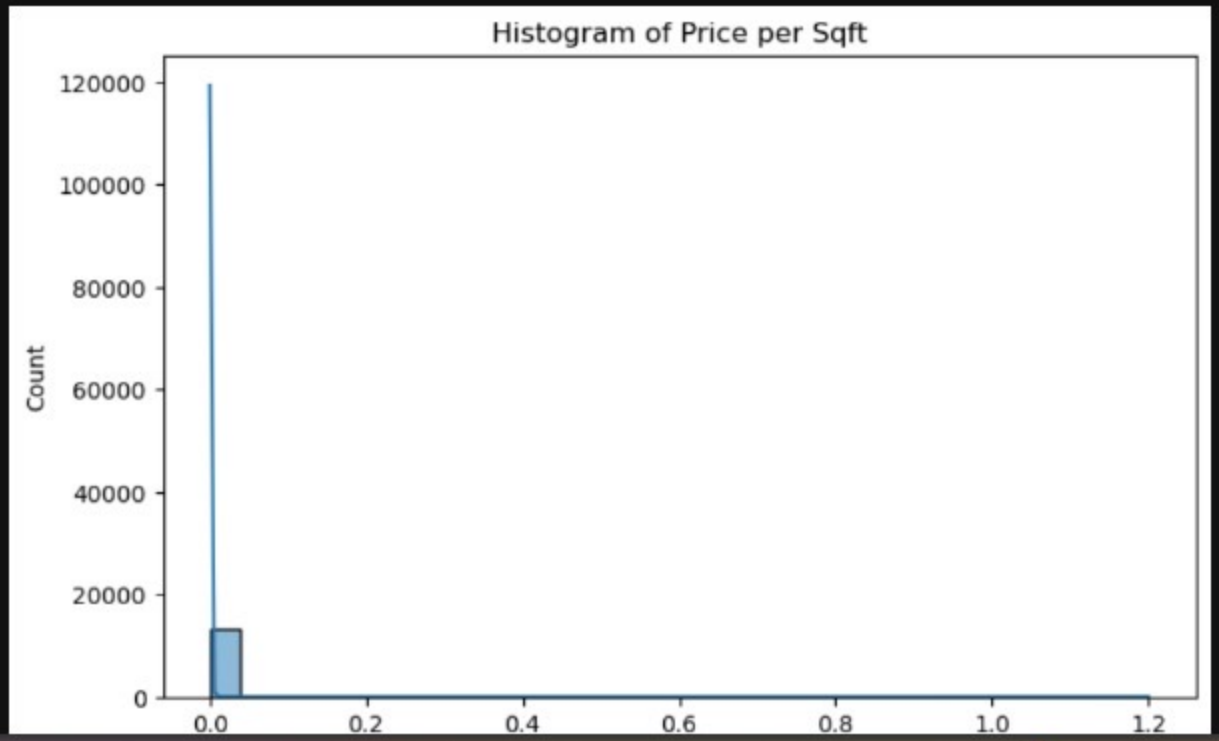
File Edit View Run Kernel Settings Help

Code JupyterLab Python 3 (ipykernel)

[ ]:
```
#Q4.Draw histplot to check the normality of the column(price per sqft column)
#and perform transformations if needed. Check the skewness and kurtosis before and after the transformation.
```

[155]:
```
from scipy.stats import skew,kurtosis
```

[169]:
```
#Draw Histogram
plt.figure(figsize=(8,5))
sns.histplot(df['price_per_sqft'],kde = True, bins=30)
plt.title('Histogram of Price per Sqft')
plt.show()
```



EUR/INR +0.48% Search ENG IN 19:00 21-12-2024

Home | Project Python | Statistics | Assignment-Statistical Measures

localhost:8889/notebooks/Assignment-Statistical%20Measures.ipynb?

jupyter Assignment-Statistical Measures Last Checkpoint: 20 hours ago

Trusted

File Edit View Run Kernel Settings Help

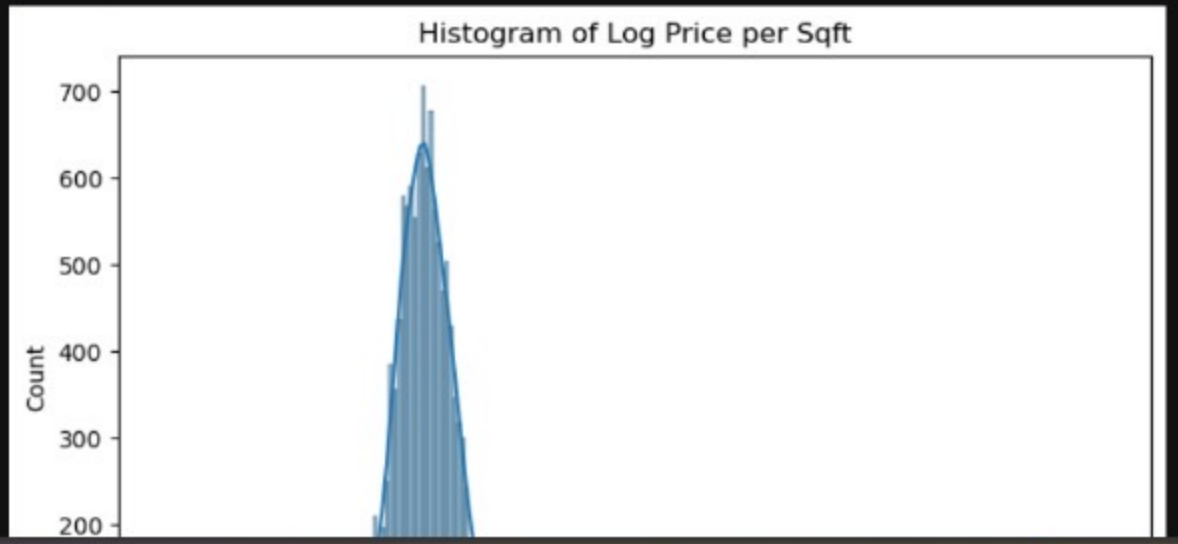
Code JupyterLab Python 3 (ipykernel)

[173]:
```
#Check Skewness and Kurtosis
Skewness_before = skew(df['price_per_sqft'])
kurtosis_before = kurtosis(df['price_per_sqft'])
print("Skewness before transformation:",Skewness_before)
print("Kurtosis before transformation:",kurtosis_before)
```

```
Skewness before transformation: 108.26875024325159
Kurtosis before transformation: 12090.633538860382
```

[185]:
```
#Perform log transformation
df['price_per_sqft_log'] = np.log(df['price_per_sqft'])
```

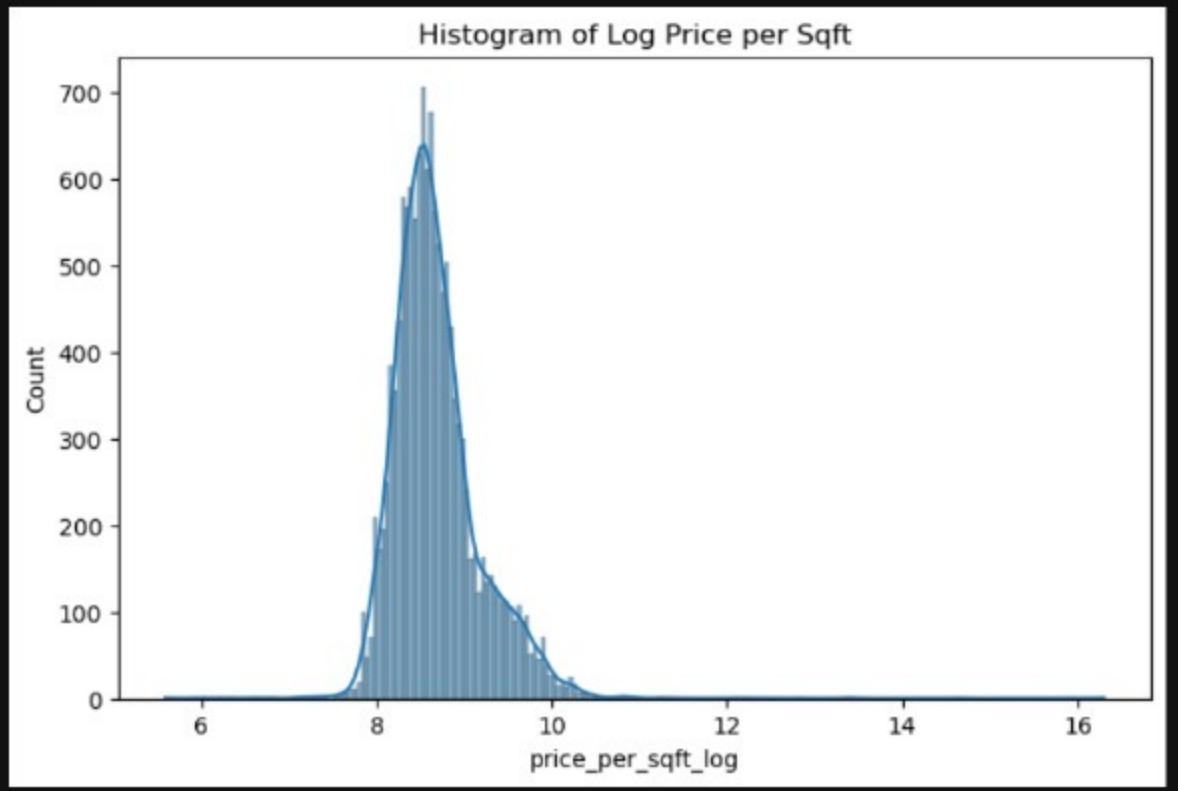
[189]:
```
#Draw Histogram after Transformation
plt.figure(figsize=(8,5))
sns.histplot(df['price_per_sqft_log'],kde=True)
plt.title('Histogram of Log Price per Sqft')
plt.show()
```



EUR/INR +0.48% Search ENG IN 19:00 21-12-2024

```
df['price_per_sqft_log'] = np.log(df['price_per_sqft'])
```

[189]:
```
#Draw HIstogram after Transformation
plt.figure(figsize=(8,5))
sns.histplot(df['price_per_sqft_log'],kde=True)
plt.title('Histogram of Log Price per Sqft')
plt.show()
```



[193]:
```
#Check skewnrss and kurtosis after transformation
```

Home | Project Python | Statistics | Assignment-Statistical Measures

localhost:8889/notebooks/Assignment-Statistical%20Measures.ipynb?

jupyter Assignment-Statistical Measures Last Checkpoint: 20 hours ago

File Edit View Run Kernel Settings Help

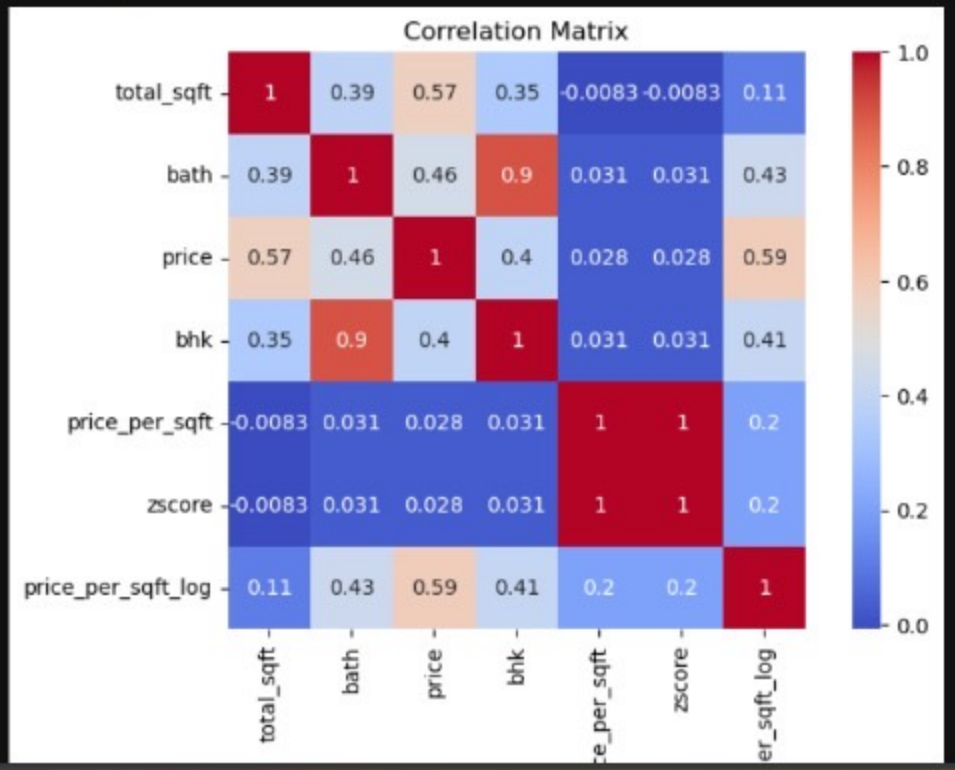
Trusted

Code

JupyterLab Python 3 (ipykernel)

[ ]:
```
#Q5.Check the correlation between all the numerical columns and plot heatmap.
```

[199]:
```
#Select numerical columns
Numerical_columns =df.select_dtypes(include=[np.number]).columns
#Calculate correlation
corr_matrix = df[Numerical_columns].corr()
#Plot Heatmap
plt.figure(figsize=(8,5))
sns.heatmap(corr_matrix , annot = True , cmap='coolwarm' ,square=True)
plt.title('Correlation Matrix')
plt.show()
```



EUR/INR +0.48%

Search

ENG IN

19:01 21-12-2024

Home | Project Python | Statistics | Assignment-Statistical Measures

localhost:8889/notebooks/Assignment-Statistical%20Measures.ipynb?

jupyter Assignment-Statistical Measures Last Checkpoint: 20 hours ago

File Edit View Run Kernel Settings Help

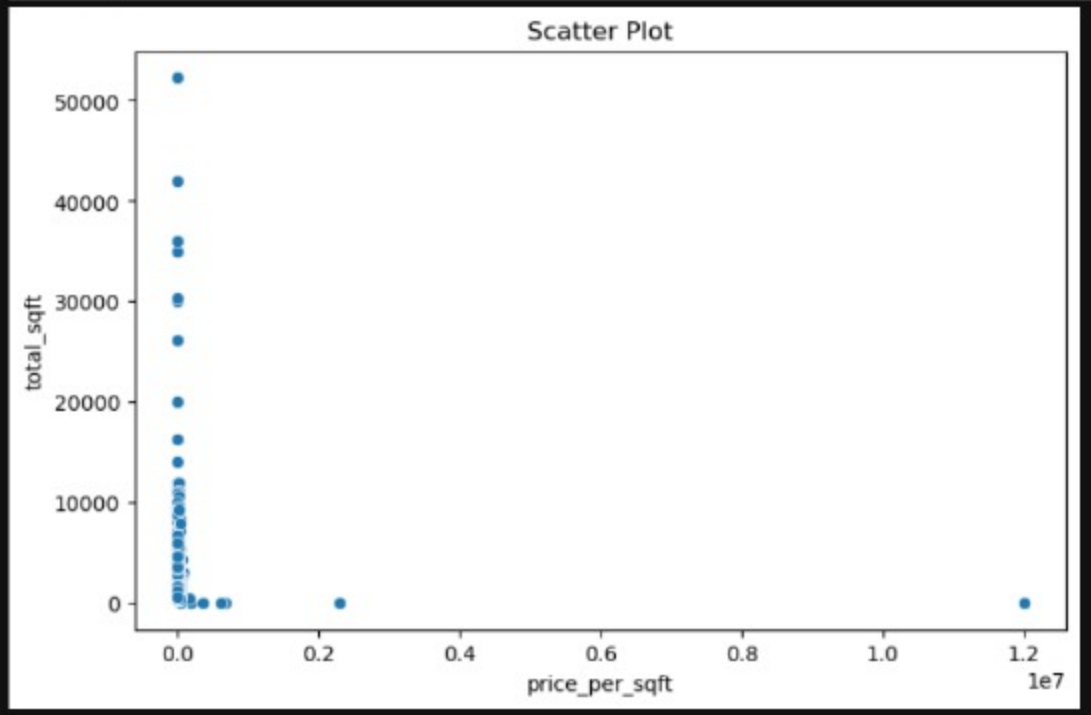
Trusted

Code

JupyterLab Python 3 (ipykernel)

[ ]:
```
#Q6.Draw Scatter plot between the variables to check the correlation between them.
```

[219]:
```
#Create scatter Plot
plt.figure(figsize=(8,5))
x_var='price_per_sqft'
y_var='total_sqft'
sns.scatterplot(x=x_var, y=y_var , data=df)
plt.title('Scatter Plot')
plt.xlabel(x_var)
plt.ylabel(y_var)
plt.show()
```



EUR/INR +0.48%

Search

ENG IN

19:01 21-12-2024