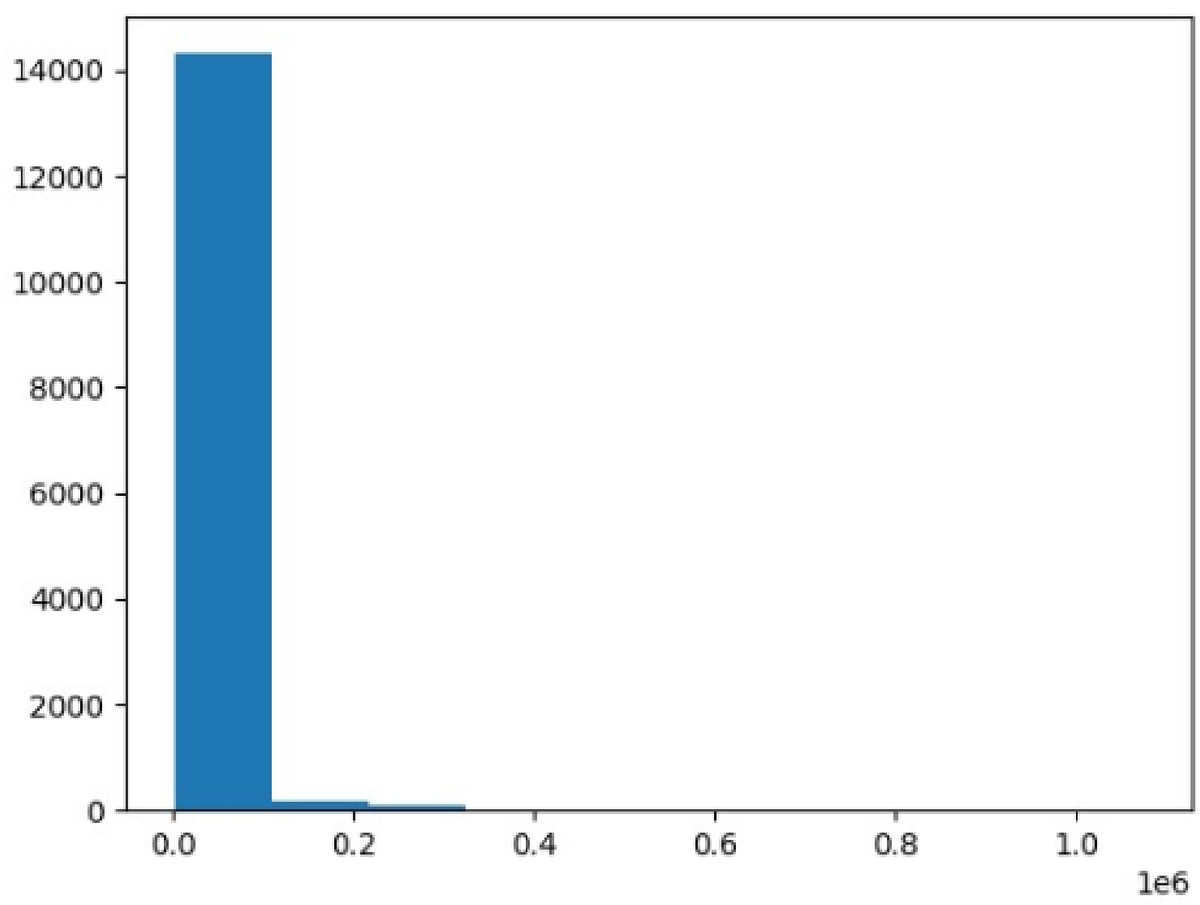
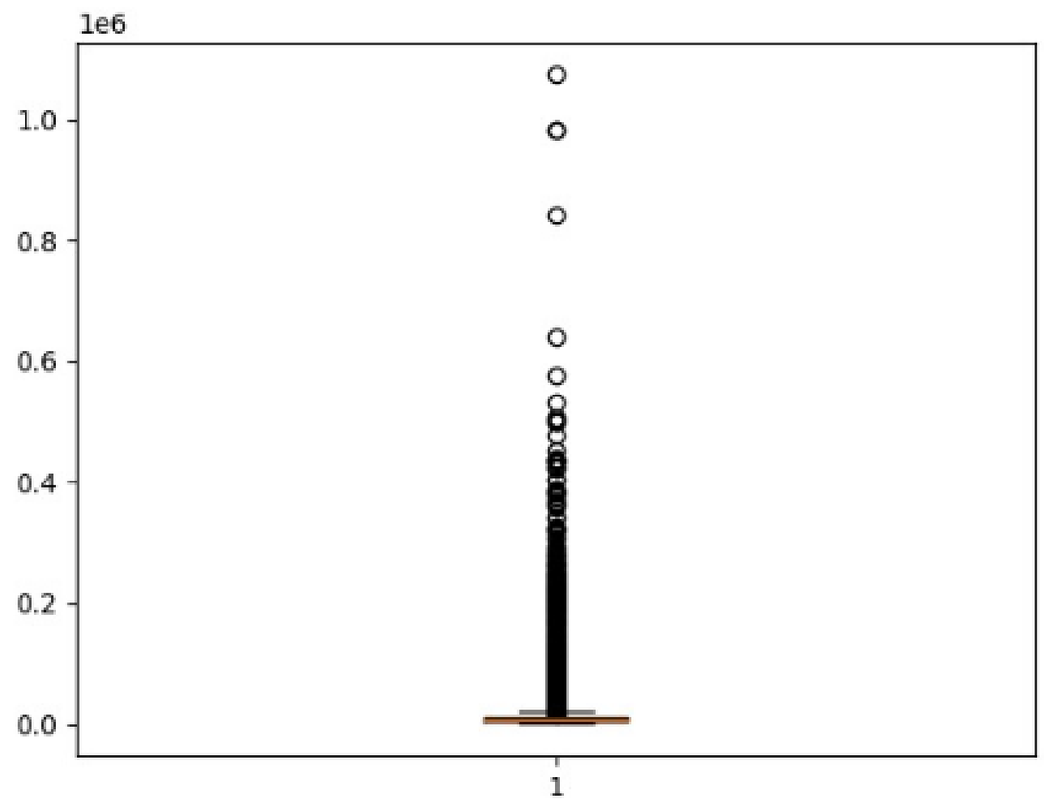
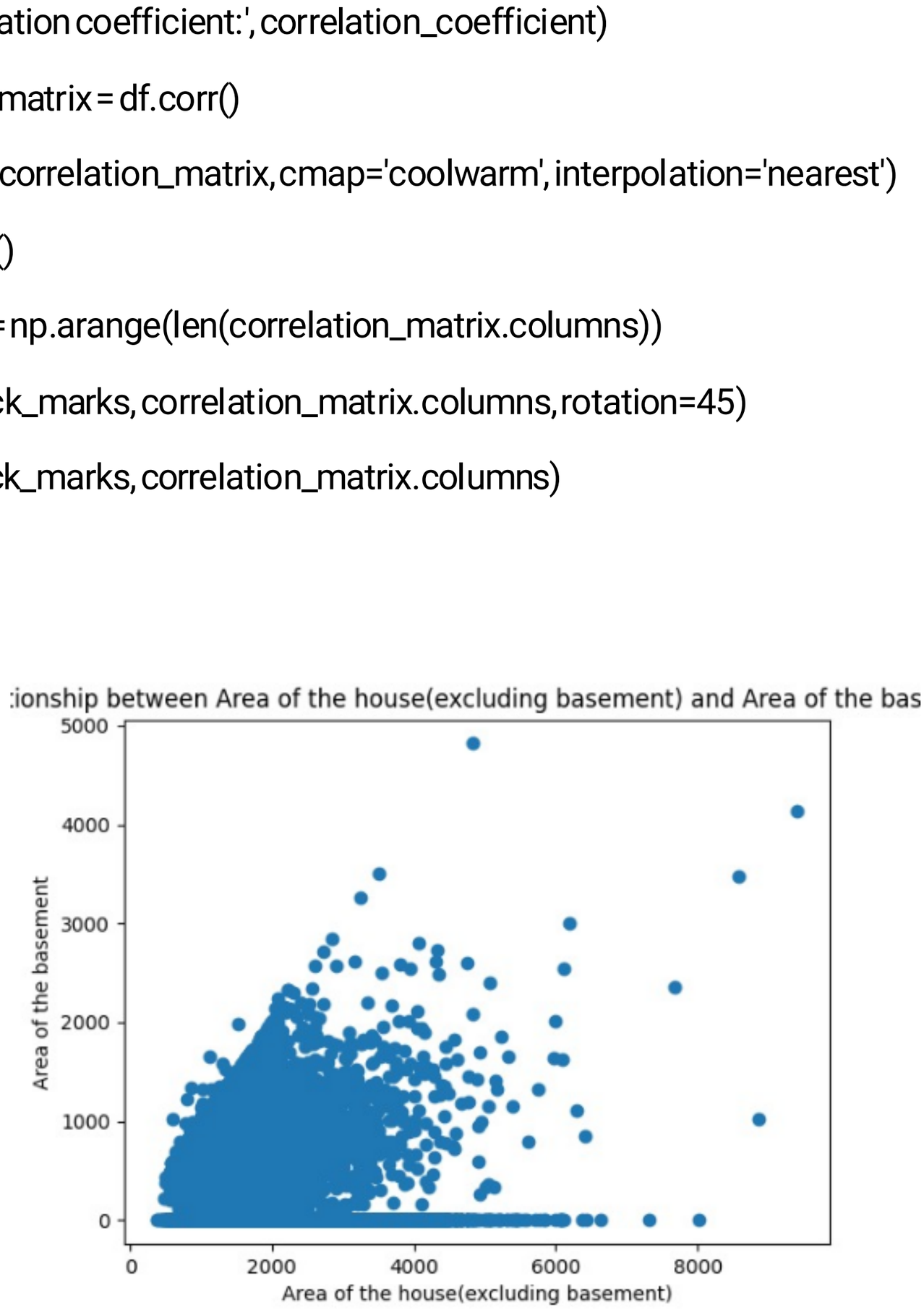
```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

data = pd.read_csv("E:\KVCET\TRAINING AND PLACEMENT CELL\Training Materials\Data Science\DS Experiments\dataset.csv")

# View the first 5 rows of the dataset
print(data.head())


# View the summary statistics of the dataset

print(data.describe())

# Plot a histogram of the variable

plt.hist(data['lot area'], bins=10)

plt.show()

# Calculate the mean of the variable

mean = np.mean(data['lot area'])

print("Mean:", mean)


# Calculate the median of the variable

median = np.median(data['lot area'])

print("Median:", median)


# Calculate the mode of the variable

mode = data['lot area'].mode()[0]

print("Mode:", mode)

# Calculate the standard deviation of the variable

std_dev = np.std(data['lot area'])

print("Standard deviation:", std_dev)


# Calculate the variance of the variable
```

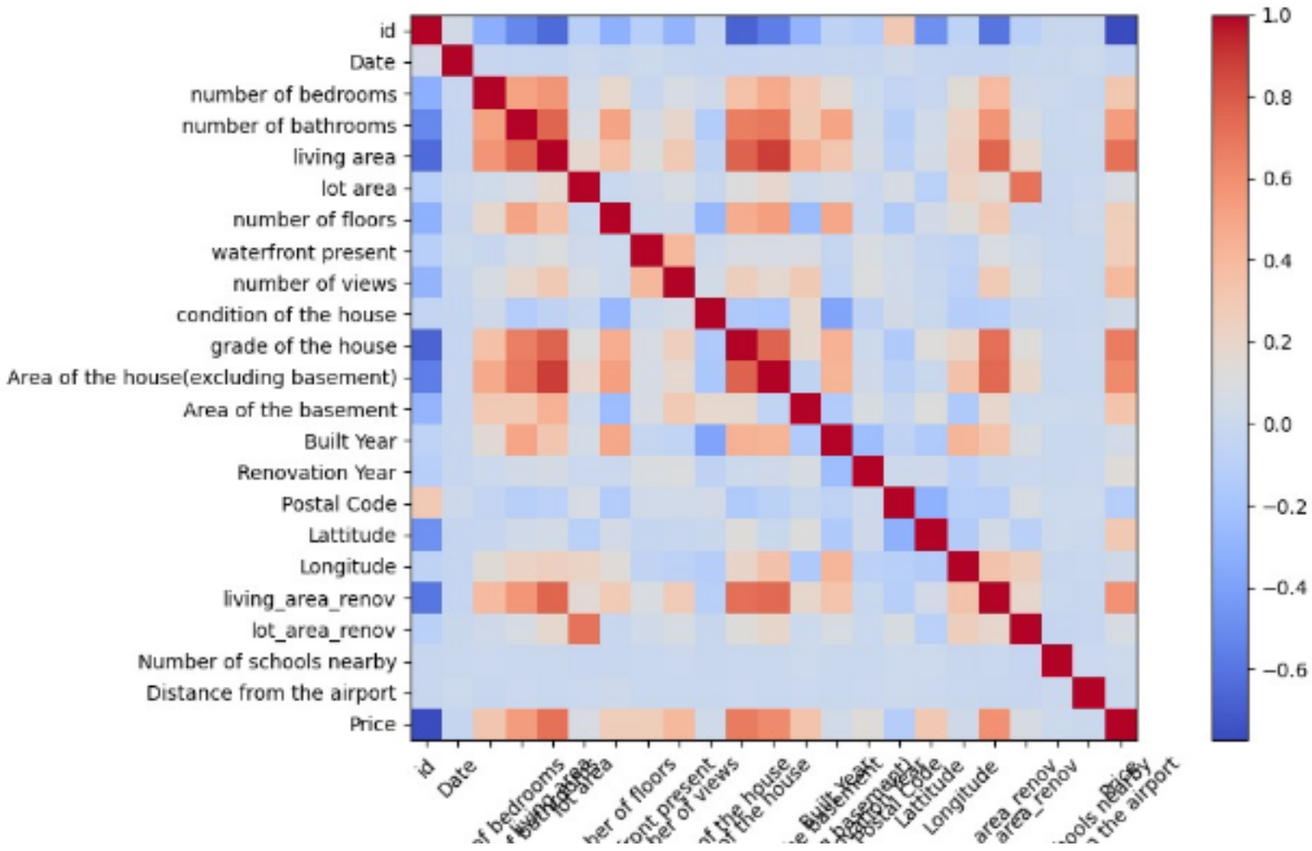
```
variance = np.var(data['lot area'])
print("Variance:",variance)
# Plot a boxplot of the variable
plt.boxplot(data['lot area'])
plt.show()
```





```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

df = pd.read_csv('E:\KVCET\TRAINING AND PLACEMENT CELL\Training Materials\Data Science\DS Experiments\dataset.csv')

plt.scatter(df['Area of the house(excluding basement)'], df['Area of the basement'])

plt.xlabel('Area of the house(excluding basement)')

plt.ylabel('Area of the basement')

plt.title('Relationship between Area of the house(excluding basement) and Area of the basement')

plt.show()

correlation_coefficient = np.corrcoef(df['Area of the house(excluding basement)'], df['Area of the basement'])[0,1]

print('Correlation coefficient:', correlation_coefficient)

correlation_matrix = df.corr()

plt.imshow(correlation_matrix, cmap='coolwarm', interpolation='nearest')

plt.colorbar()

tick_marks = np.arange(len(correlation_matrix.columns))

plt.xticks(tick_marks, correlation_matrix.columns, rotation=45)

plt.yticks(tick_marks, correlation_matrix.columns)

plt.show()
```






```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from sklearn import preprocessing

from sklearn.decomposition import PCA

df = pd.read_csv('E:\KVCET\TRAINING AND PLACEMENT CELL\Training Materials\Data Science\DS Experiments\dataset.csv')

X = df.drop(['grade of the house'],axis=1) #Remove the target variable from the dataset

X_scaled = preprocessing.scale(X) #Scale the features to have zero mean and unit variance

pca = PCA()
```

```
X_pca = pca.fit_transform(X_scaled)

plt.scatter(X_pca[:, 0], X_pca[:, 1])

plt.xlabel('Principal Component 1')

plt.ylabel('Principal Component 2')

plt.show()
```