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)

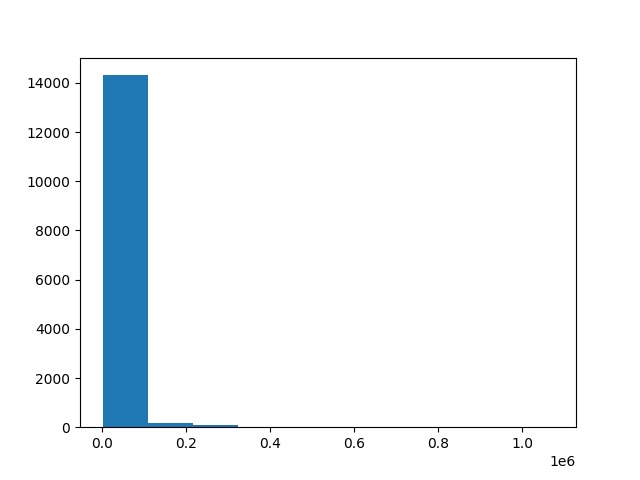
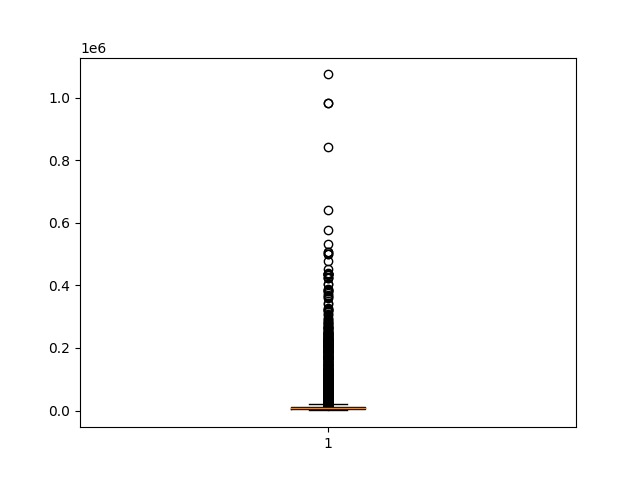
# Calculate the variance of the variable

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')

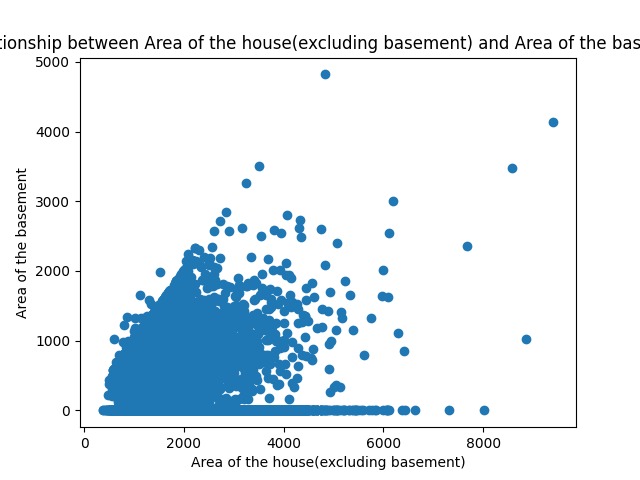
plt.colorbar()

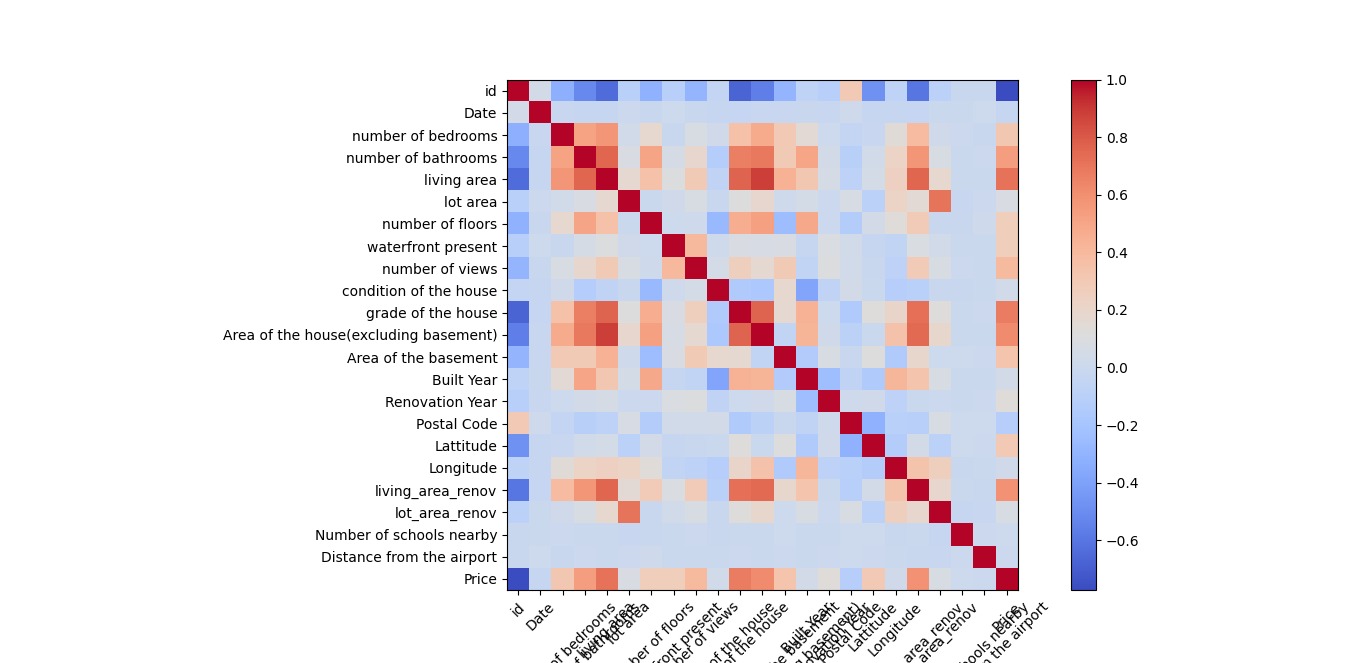
tick\_marks = np.arange(len(correlation\_matrix.columns))

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()

