```
import pandas as pd

import numpy as np

import scipy as sp

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split

from sklearn.linear_model import LinearRegression

from sklearn import metrics

data= pd.read_csv("housepriceprediction.csv")

data.head()

data.describe(include=[np.number])

data.isnull().sum()

names=['price','bedrooms','bathrooms','sqft_living','sqft_lot','floors','waterfront','view','conditi
on','grade','sqft_above','sqft_basement','zipcode','lat','long']

df=data[names]

correlations= df.corr()

fig=plt.figure()

ax=fig.add_subplot(111)

cax=ax.matshow(correlations,vmin=-1,vmax=1)

fig.colorbar(cax)

ticks=np.arange(0,15,1)

ax.set_xticks(ticks)

ax.set_yticks(ticks)

ax.set_xticklabels(names)

ax.set_yticklabels(names)
```

```
plt.show()
data.dtypes
sns.regplot(x='sqft_living',y='price',data=data)
sns.regplot(x='sqft_basement',y='price',data=data)
sns.regplot(x='sqft_above',y='price',data=data)
sns.stripplot(x='bedrooms', y='price',data=data)
sns.stripplot(x='bathrooms', y='price',data=data, size=5)
sns.stripplot(x='grade', y='price',data=data, size=5)
data=data[data['bedrooms'] < 10]
data=data[data['bathrooms']<8]
data.head()
c=['bedrooms','bathrooms','sqft_living','sqft_above','grade']
df=data[c]
df=pd.get_dummies(df,columns=['grade'], drop_first=True)
y=data['price']
x_train,x_test,y_train,y_test=train_test_split(df,y,train_size=0.8,random_state=42)
x_train.head()
reg=LinearRegression()
reg.fit(x_train,y_train)
print('Coefficients: \n', reg.coef_)
print(metrics.mean_squared_error(y_test, reg.predict(x_test)))
reg.score(x_test,y_test)
df=pd.get_dummies(data,columns=['waterfront','view','condition','grade','zipcode'],
drop_first=True)
y=data['price']
```

```
df= df.drop(['date','id','price'],axis=1)

x_train,x_test,y_train,y_test=train_test_split(df,y,train_size=0.8,random_state=42)

reg.fit(x_train,y_train)

print('Coefficients: \n', reg.coef_)

print(metrics.mean_squared_error(y_test, reg.predict(x_test)))

print(reg.score(x_test,y_test))
```