**#California Housing Price Prediction**

**# Required libraries for this problem**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

**1) Loading the data**

dataset=pd.read\_excel('housing.xlsx')

#Printing first few rows of the data

df=dataset.head(n=20)

X=dataset.iloc[:,0:9].values

Y=dataset.iloc[:,9].values

dataset.isnull().sum()

**2)Handling the missing values**

from sklearn.preprocessing import Imputer

imputer=Imputer(missing\_values='NaN',strategy='mean',axis=0)

imputer=imputer.fit(X[:,[4]])

X[:,[4]]=imputer.fit\_transform(X[:,[4]])

df1=pd.DataFrame(X).values

df1.isnull().sum()

**3) Encoding the categorical data**

from sklearn.preprocessing import LabelEncoder,OneHotEncoder

labelencoder\_df1=LabelEncoder()

df1[:,8]=labelencoder\_df1.fit\_transform(df1[:,8])

df2=pd.DataFrame(df1)

df2.isnull().sum()

onehotencoder=OneHotEncoder(categorical\_features=[8])

df2=onehotencoder.fit\_transform(df2).toarray()

**4) Splitting the dataset**

from sklearn.model\_selection import train\_test\_split

X\_train,X\_test,Y\_train,Y\_test=train\_test\_split(X,Y,test\_size=0.2,random\_state=0)

**5) Standardizing the data**

from sklearn.preprocessing import StandardScaler

sc\_df2=StandardScaler()

X\_train=sc\_df2.fit\_transform(X\_train)

X\_test=sc\_df2.fit\_transform(X\_test)

**6) Performing the linear Regression**

from sklearn.linear\_model import LinearRegression

regressor=LinearRegression()

regressor.fit(X\_train,Y\_train)

regressor.coef\_

regressor.intercept\_

regressor.score(X\_train,Y\_train)

predicted=regressor.predict(X\_test)

expected=Y\_test

from sklearn import metrics

np.sqrt(metrics.mean\_squared\_error(expected,predicted))

plt.scatter(expected,predicted,color='blue')

plt.plot([0,600000],[-0,600000],'--k',color='black')

plt.title('California Housing Price Prediction')

plt.xlabel('Actual Price')

plt.ylabel('Expected Price')

plt.show()

**7)Bonus exercise :**

# California Housing Price Prediction # Bonus exercise

**#Required Libraries**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

**# Loading the data and splitting**

dataset=pd.read\_excel('housing1.xlsx')

X=dataset.iloc[:,[0]]

Y=dataset.iloc[:,[1]]

dataset.isnull().sum()

from sklearn.model\_selection import train\_test\_split

X\_train,X\_test,Y\_train,Y\_test=train\_test\_split(X,Y,test\_size=0.2,random\_state=0)

**#Performing linear Regression**

from sklearn.linear\_model import LinearRegression

regressor=LinearRegression()

regressor.fit(X\_train,Y\_train)

regressor.coef\_

regressor.intercept\_

y\_pred=regressor.predict(X\_test)

from sklearn import metrics

np.sqrt(metrics.mean\_squared\_error(Y\_test,y\_pred))

regressor.score(X\_train,Y\_train)

**#Plotting**

plt.scatter(X\_test,Y\_test,color='blue')

plt.plot(X\_test,y\_pred,color='black')

plt.xlabel('median house value')

plt.ylabel('median income')

plt.title('California Housing Price Prediction')

plt.show()