project1- price predicting

```
import pandas as pd

housing=pd.read_csv("data1.csv")

housing.head()

housing.info()

%matplotlib inline

import matplotlib.pyplot as plt

housing.hist(bins=50,figsize=(20,15))

from sklearn.model_selection import train_test_split

train_set,test_set=train_test_split(housing,test_size=0.2,random_state=42)

print(len(test_set))

print(len(train_set))

from sklearn.model_selection import StratifiedShuffleSplit

split=StratifiedShuffleSplit(n_splits=1,test_size=0.2,random_state=42)

for train_index,test_index in split.split(housing,housing['CHAS']):

    strat_train_set=housing.loc[train_index]

    strat_test_set=housing.loc[test_index]

strat_test_set['CHAS'].value_counts()

strat_train_set['CHAS'].value_counts()

housing=strat_train_set.copy()

corr_matrix=housing.corr()

corr_matrix['MEDV'].sort_values(ascending=False)

housing.plot(kind='scatter',x='RM',y='MEDV',alpha=0.8)

housing["TAXRM"]= housing['TAX']/housing['RM']

housing.head()

housing["TAXRM"]

corr_matrix=housing.corr()
```

```
corr_matrix['MEDV'].sort_values(ascending=False)
```

```
housing.plot(kind='scatter',x='TAXRM',y='MEDV',alpha=0.8)
```