EDA:

https://www.kaggle.com/swathiachath/kc-housesales-data

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

()sns.set

import warnings

from sklearn.model_selection import train_test_split

from sklearn.metrics import accuracy_score

('warnings.filterwarnings('ignore

from sklearn.linear_model import LogisticRegression

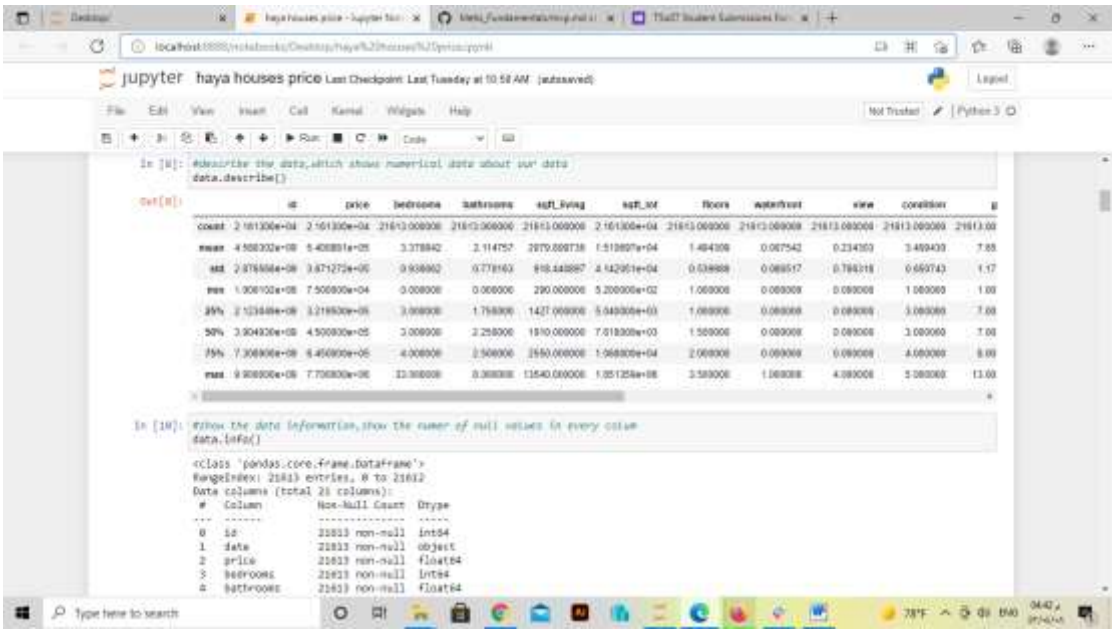
matplotlib inline%

("2- data=pd.read_csv("f:\\kc_house_data.csv

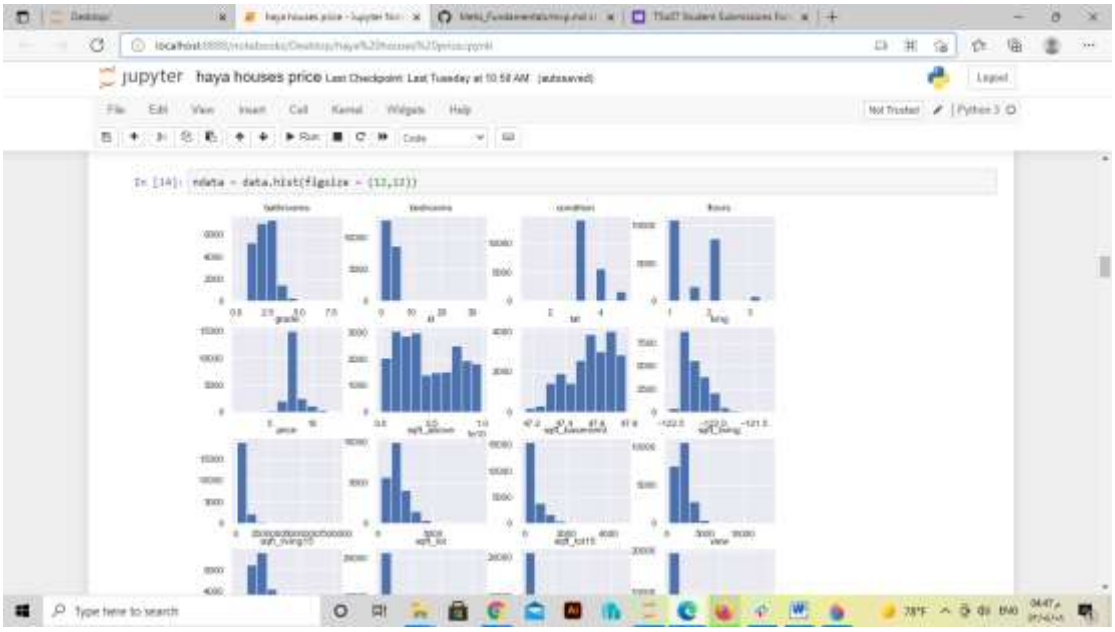
()data.head

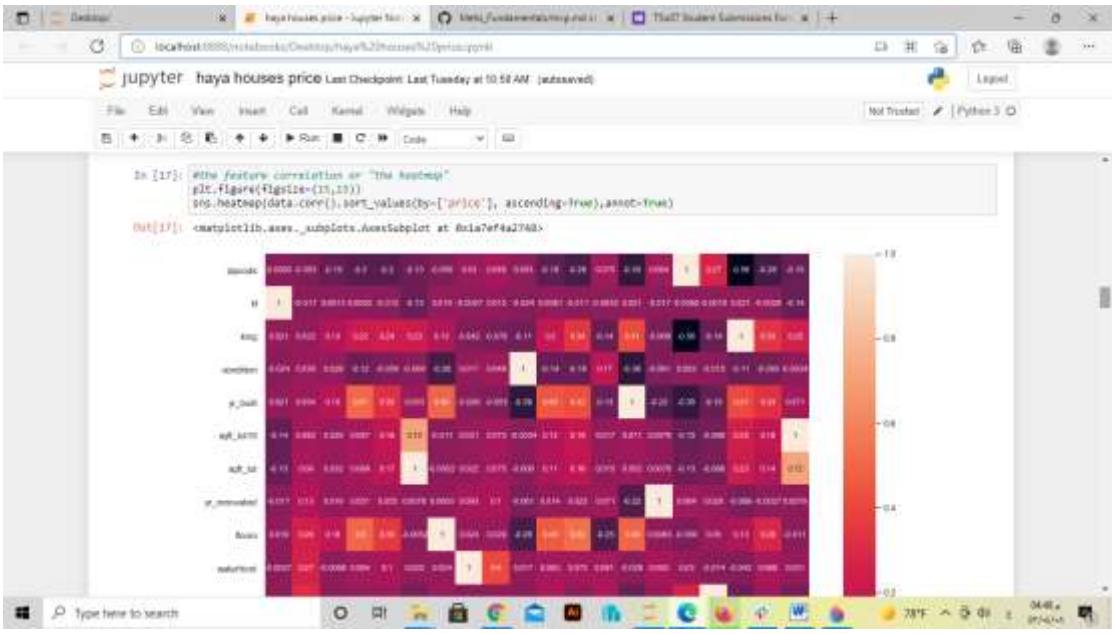
3- print('We have {} rows and {} columns'.format(data.shape[0],data.shape[

```
Output We have 21613 rows and 21 columns
```



EDA:


jupyter haya houses price
In [14]: data = data.hist(figsize = (12,12))


jupyter haya houses price
In [17]: plt.figure(figsize=(15,15))
sns.heatmap(data.corr().sort_values(by=['price'], ascending=True),annot=True)
Out[17]: <matplotlib.axes._subplots.AxesSubplot at 0x1a7ef4a2748>