EDS Theory Activity No. 1

Name:- Avinash Singh

Div:- CS2 Roll No:- CS2-68

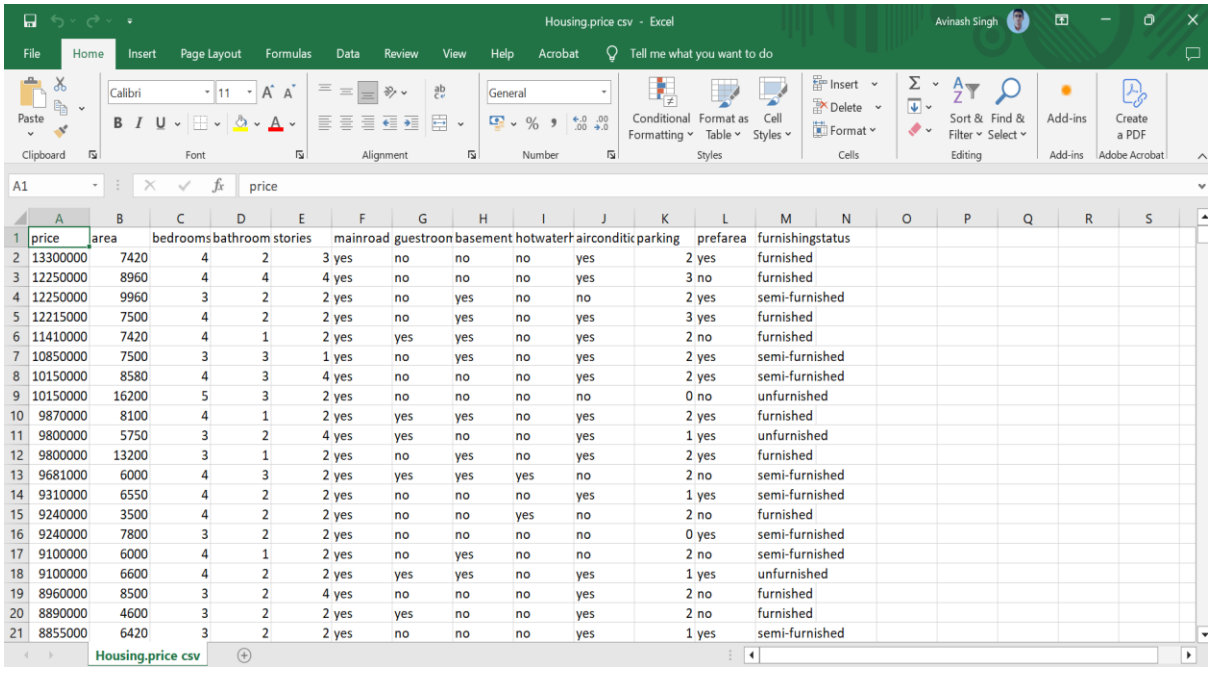
PRN:- 202401040311

Dataset Given :- House Price

Dataset download link:-

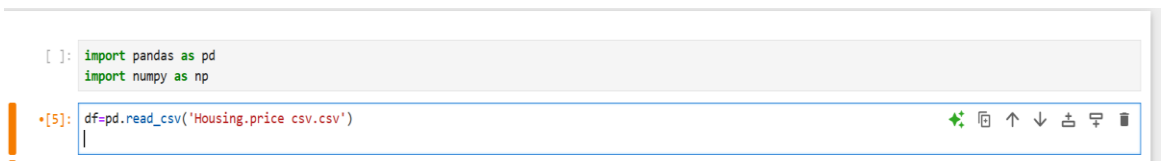
https://www.kaggle.com/datasets/yasserh/housing-prices-dataset

dataset Screen Shot:-



	price	area	bedrooms	bathroom	stories	mainroad	guestroom	basement	hotwaterh	airconditic	parking	prefarea	furnishingstatus
2	13300000	7420	4	2	3	yes	no	no	no	yes	2	yes	furnished
3	12250000	8960	4	4	4	yes	no	no	no	yes	3	no	furnished
4	12250000	9960	3	2	2	yes	no	yes	no	no	2	yes	semi-furnished
5	12215000	7500	4	2	2	yes	no	yes	no	yes	3	yes	furnished
6	11410000	7420	4	1	2	yes	yes	yes	no	yes	2	no	furnished
7	10850000	7500	3	3	1	yes	no	yes	no	yes	2	yes	semi-furnished
8	10150000	8580	4	3	4	yes	no	no	no	yes	2	yes	semi-furnished
9	10150000	16200	5	3	2	yes	no	no	no	no	0	no	unfurnished
10	9870000	8100	4	1	2	yes	yes	yes	no	yes	2	yes	furnished
11	9800000	5750	3	2	4	yes	yes	no	no	yes	1	yes	unfurnished
12	9800000	13200	3	1	2	yes	no	yes	no	yes	2	yes	furnished
13	9681000	6000	4	3	2	yes	yes	yes	yes	no	2	no	semi-furnished
14	9310000	6550	4	2	2	yes	no	no	no	yes	1	yes	semi-furnished
15	9240000	3500	4	2	2	yes	no	no	yes	no	2	no	furnished
16	9240000	7800	3	2	2	yes	no	no	no	no	0	yes	semi-furnished
17	9100000	6000	4	1	2	yes	no	yes	no	no	2	no	semi-furnished
18	9100000	6600	4	2	2	yes	yes	yes	no	yes	1	yes	unfurnished
19	8960000	8500	3	2	4	yes	no	no	no	yes	2	no	furnished
20	8890000	4600	3	2	2	yes	yes	no	no	yes	2	no	furnished
21	8855000	6420	3	2	2	yes	no	no	no	yes	1	yes	semi-furnished

Using jupyter notebook:-



Q1. What is the total number of houses in the dataset?

```
[11]: total_house=len(df)
      print(total_house)
```

```
545
```

```
[ ]:
```

Q2. How many houses have exactly 3 bedrooms?

```
[23]: Exactly_three=df[df['bedrooms']==3]

      print(len(Exactly_three))
```

```
300
```

Q3. What is the average price of all houses?

```
32]: Average_price=df['price'].mean()
     print(Average_price)
```

```
4766729.247706422
```

Q4. List all houses with 2 bathrooms and sort them by price.

```
[34]: two_bathrooms=df[df['bathrooms']==2]
      sorted_price=two_bathrooms.sort_values(by='price')
      print(sorted_price)
```

```
        price  area  bedrooms  bathrooms  stories mainroad guestroom basement  \
523   2380000  2787         4          2        2      yes        no       no
509   2590000  3600         2          2        2      yes        no      yes
446   3150000  3986         2          2        1       no       yes      yes
413   3430000  1950         3          2        2      yes        no      yes
390   3500000  2135         3          2        2       no        no       no
..        ...   ...       ...        ...      ...      ...       ...      ...
12    9310000  6550         4          2        2      yes        no       no
9     9800000  5750         3          2        4      yes       yes       no
3    12215000  7500         4          2        2      yes        no      yes
2    12250000  9960         3          2        2      yes        no      yes
0    13300000  7420         4          2        3      yes        no       no

    hotwaterheating airconditioning  parking prefarea furnishingstatus
523              no              no        0       no        furnished
509              no              no        1       no        furnished
446              no              no        1       no      unfurnished
413              no              no        0      yes      unfurnished
390              no              no        0       no      unfurnished
..              ...             ...      ...      ...              ...
12               no             yes        1      yes   semi-furnished
9                no             yes        1      yes      unfurnished
3                no             yes        3      yes        furnished
2                no              no        2      yes   semi-furnished
0                no             yes        2      yes        furnished
```

Q5. How many houses are furnished?

```
[40]: furnished_house=df[df['furnishingstatus']=='furnished']
      print(len(furnished_house))
```

```
140
```

Q6. What is the maximum area of houses with 4 bedrooms?

```
[42]: House_four_bedrooms=df[df['bedrooms']==4]['area'].max()
      print(House_four_bedrooms)
```

```
12090
```

Q7. Find the average price of houses located on a main road.

```
main_road=df[df['mainroad']=='yes']['price'].mean()
print(main_road)
```

```
4991777.329059829
```

Q8. Find the average price of houses that have hot water heating and are in a preferred area.

```
average_price=df[(df['hotwaterheating']=='yes') & (df['prefarea']=='yes')]['price'].mean()
print(average_price)
```

```
6104000.0
```

Q9. List the top 5 cheapest houses in the dataset.

```
cheapeast_house= df.sort_values(by='price').head(5)
print(cheapeast_house)
```

```
      price  area  bedrooms  bathrooms  stories mainroad guestroom basement  \
544  1750000  3850         3          1        2      yes        no       no
543  1750000  2910         3          1        1       no        no       no
542  1750000  3620         2          1        1      yes        no       no
541  1767150  2400         3          1        1       no        no       no
540  1820000  3000         2          1        1      yes        no      yes

    hotwaterheating airconditioning  parking prefarea furnishingstatus
544              no              no        0       no      unfurnished
543              no              no        0       no        furnished
542              no              no        0       no      unfurnished
541              no              no        0       no   semi-furnished
540              no              no        2       no      unfurnished
```

Q10. What is the total number of houses with a basement?

```
[21]: houses_with_basements=df[df['basement']=='yes']
      count_houses=len(houses_with_basements)
      print(count_houses)
```

```
191
```

Q11. Calculate the average area of houses with 2 stories.

```
[23]: Average_area=df[df['stories']==2]['area'].mean()
      print(Average_area)
```

```
4756.94537815126
```

Q12. How many houses have no parking spaces?

```
[33]: No_praking=df[df['parking']==0]
      print(len(No_praking))
```

```
299
```

Q13. Find the minimum price of houses in a preferred area.

```
[35]: minimum_price=df[df['prefarea']== 'yes']['price'].min()
      print(minimum_price)
```

```
2233000
```

Q14. List all houses with a guest room and sort by area.

```
[45]: List_houses=df[df['guestroom']=='yes'].sort_values(by='area')
      print(List_houses)
```

```
       price   area  bedrooms  bathrooms  stories mainroad guestroom basement  \
282  4270000   2175         3          1        2       no       yes      yes
346  3850000   2176         2          1        2      yes       yes       no
157  5495000   2817         4          2        2       no       yes      yes
236  4620000   2870         2          1        2      yes       yes      yes
184  5110000   3000         3          2        2      yes       yes      yes
..       ...    ...       ...        ...      ...      ...       ...      ...
172  5250000   8400         3          1        2      yes       yes      yes
57   7245000   9000         4          2        4      yes       yes       no
229  4690000   9667         4          2        2      yes       yes      yes
175  5250000   9800         4          2        2      yes       yes       no
191  5040000  10700         3          1        2      yes       yes      yes

    hotwaterheating airconditioning  parking prefarea furnishingstatus
282              no             yes        0       no      unfurnished
346              no              no        0      yes   semi-furnished
157              no              no        1       no       furnished
236              no              no        0      yes   semi-furnished
184              no              no        0       no       furnished
..              ...             ...      ...      ...             ...
172              no             yes        2      yes      unfurnished
57               no             yes        1      yes       furnished
229              no              no        1       no   semi-furnished
175              no              no        2       no   semi-furnished
191              no              no        0       no   semi-furnished

[97 rows x 13 columns]
```

Q15. What is the average price of unfurnished houses?

```
[55]: Average_price_unfurnished=round(df[df['furnishingstatus']=='unfurnished']['price'].mean(),2)
      print(Average_price_unfurnished)
```

```
4013831.46
```

Q16. How many houses have both 2 bedrooms and 1 bathroom?

```
houses=df[(df['bedrooms']==2) & (df['bathrooms']==1)]
print(len(houses))
print(houses)
```

```
128
       price    area  bedrooms  bathrooms  stories mainroad guestroom basement  \
61   7070000    8880         2          1        1      yes        no       no
66   6930000   13200         2          1        1      yes        no      yes
91   6419000    6750         2          1        1      yes       yes      yes
114  6020000    6800         2          1        1      yes       yes      yes
146  5600000   10500         2          1        1      yes        no       no
..       ...     ...       ...        ...      ...      ...       ...      ...
535  2100000    3360         2          1        1      yes        no       no
538  1890000    3649         2          1        1      yes        no       no
539  1855000    2990         2          1        1       no        no       no
540  1820000    3000         2          1        1      yes        no      yes
542  1750000    3620         2          1        1      yes        no       no

    hotwaterheating airconditioning  parking prefarea furnishingstatus
61               no             yes        1       no   semi-furnished
66              yes              no        1       no       furnished
91               no              no        2      yes       furnished
114              no              no        2       no       furnished
146              no              no        1       no   semi-furnished
..              ...             ...      ...      ...             ...
535              no              no        1       no      unfurnished
538              no              no        0       no      unfurnished
539              no              no        1       no      unfurnished
540              no              no        2       no      unfurnished
542              no              no        0       no      unfurnished
```

Q17. Group houses by the number of stories and calculate the average price for each group.

```
[67]: Group=round(df.groupby('stories')['price'].mean(),2)
      print(Group)
```

```
stories
1    4170658.59
2    4764073.53
3    5685435.90
4    7208449.76
Name: price, dtype: float64
```

Q18. What is the total number of houses with an area greater than 6000 sq.ft.?

```
[69]: Total_no_houses=df[df['area'] > 6000].value_counts()
      print(Total_no_houses)
```

```
price     area  bedrooms  bathrooms  stories  mainroad  guestroom  basement  hotwaterheating  airconditioning  parking  prefarea  furnishingstatus
2450000   7700  2         1          1        yes       no         no        no               no               0        no        unfurnished       1
6615000   10500 3         2          1        yes       no         yes       no               yes              1        yes       furnished         1
6300000   7200  3         2          1        yes       no         yes       no               yes              3        no        semi-furnished    1
          9000  3         1          1        yes       no         yes       no               no               1        yes       furnished         1
6419000   6750  2         1          1        yes       yes        yes       no               no               2        yes       furnished         1
                                                                                                                                                    ..
5215000   7200  3         1          2        yes       yes        yes       no               no               1        yes       furnished         1
5229000   7085  3         1          1        yes       yes        yes       no               no               2        yes       semi-furnished    1
5243000   6050  3         1          1        yes       no         yes       no               no               0        yes       semi-furnished    1
5250000   6540  4         2          2        no        no         no        no               yes              0        no        semi-furnished    1
13300000  7420  4         2          3        yes       no         no        no               yes              2        yes       furnished         1
Length: 157, dtype: int64
```

Q19. List all houses with a price less than 3 million and sort them by price.

```
[71]: list_of_houses=df[df['price'] < 3000000].sort_values(by='price')
      print(list_of_houses)
```

```
      price  area  bedrooms  bathrooms  stories mainroad guestroom basement  \
544  1750000  3850         3          1        2      yes        no       no
543  1750000  2910         3          1        1       no        no       no
542  1750000  3620         2          1        1      yes        no       no
541  1767150  2400         3          1        1       no        no       no
540  1820000  3000         2          1        1      yes        no      yes
..       ...   ...       ...        ...      ...      ...       ...      ...
478  2940000  3600         3          1        2       no        no       no
477  2940000  4960         2          1        1      yes        no       no
476  2940000  5850         3          1        2      yes        no      yes
475  2961000  3000         2          1        2      yes        no       no
474  2975000  4352         4          1        2       no        no       no

    hotwaterheating airconditioning  parking prefarea furnishingstatus
544              no              no        0       no      unfurnished
543              no              no        0       no        furnished
542              no              no        0       no      unfurnished
541              no              no        0       no   semi-furnished
540              no              no        2       no      unfurnished
..              ...             ...      ...      ...              ...
478              no              no        1       no      unfurnished
477              no              no        0       no      unfurnished
476              no              no        1       no      unfurnished
475              no              no        0       no   semi-furnished
474              no              no        1       no      unfurnished

[71 rows x 13 columns]
```

Q20. Group houses by furnishing status and find the minimum and maximum price for each group.

[79]:
```
group_houses=df.groupby('furnishingstatus')['price'].agg(['min','max'])
print(group_houses)
```

```
                     min       max
furnishingstatus
furnished        1750000  13300000
semi-furnished   1767150  12250000
unfurnished      1750000  10150000
```