Pandas Session

Name:- Rahul Patil

In [3]:

```
import pandas as pd
import numpy as np
```

In [4]:

```
ab={
    "ID":[1,2,3,4,5],
    "Name":["Ravi","Jitu","Rahul","Deva","Sagar"],
    "City":["Pune","Dhule","jalgaon","Nashik","Mumbai"],
    "salary":[1000,2000,3000,4000,5000]
}
print(ab)
```

```
{'ID': [1, 2, 3, 4, 5], 'Name': ['Ravi', 'Jitu', 'Rahul', 'Deva', 'Sagar'],
'City': ['Pune', 'Dhule', 'jalgaon', 'Nashik', 'Mumbai'], 'salary': [1000, 2
000, 3000, 4000, 5000]}
```

In [5]:

```
data=pd.DataFrame(ab)
data
```

Out[5]:

	ID	Name	City	salary
0	1	Ravi	Pune	1000
1	2	Jitu	Dhule	2000
2	3	Rahul	jalgaon	3000
3	4	Deva	Nashik	4000
4	5	Sagar	Mumbai	5000

In [6]:

```
data.to_csv("dictionaries.csv",index=False)
```

In [7]:

```
data3=pd.read_csv("dictionaries.csv")
data3
```

Out[7]:

	ID	Name	City	salary
0	1	Ravi	Pune	1000
1	2	Jitu	Dhule	2000
2	3	Rahul	jalgaon	3000
3	4	Deva	Nashik	4000
4	5	Sagar	Mumbai	5000

In [9]:

```
dataset=pd.read_csv(r"C:\Users\Parimal\Downloads\Houseprice_by_City.csv")
```

In [10]:

```
dataset
```

Out[10]:

_FT	READY_TO_MOVE	RESALE	City	LONGITUDE	LATITUDE	TARGET(PRICE_IN_LACS)
3407	1	1	Bangalore	12.969910	77.597960	55.0
0000	1	1	Mysore	12.274538	76.644605	51.0
9722	1	1	Bangalore	12.778033	77.632191	43.0
1143	1	1	Ghaziabad	28.642300	77.344500	62.5
9247	0	1	Kolkata	22.592200	88.484911	60.5
...	...	...	...	...	...	...
0000	1	1	Agra	27.140626	78.043277	45.0
0769	1	1	Lake View Recidency	39.945409	-86.150721	16.0
1509	1	1	Jaipur	26.928785	75.828002	27.1
9009	1	1	Chennai	12.900150	80.227910	67.0
4194	1	1	Jaipur	26.832353	75.841749	27.8

In [11]:

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 29451 entries, 0 to 29450
Data columns (total 12 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   POSTED_BY              29451 non-null  object
 1   UNDER_CONSTRUCTION     29451 non-null  int64
 2   RERA                   29451 non-null  int64
 3   BHK_NO.                29451 non-null  int64
 4   BHK_OR_RK              29451 non-null  object
 5   SQUARE_FT              29451 non-null  float64
 6   READY_TO_MOVE          29451 non-null  int64
 7   RESALE                 29451 non-null  int64
 8   City                   29442 non-null  object
 9   LONGITUDE              29451 non-null  float64
 10  LATITUDE               29451 non-null  float64
 11  TARGET(PRICE_IN_LACS)  29451 non-null  float64
dtypes: float64(4), int64(5), object(3)
memory usage: 2.7+ MB
```

In [12]:

```
dataset.head()
```

Out[12]:

	POSTED_BY	UNDER_CONSTRUCTION	RERA	BHK_NO.	BHK_OR_RK	SQUARE_FT	READY
0	Owner	0	0	2	BHK	1300.236407	
1	Dealer	0	0	2	BHK	1275.000000	
2	Owner	0	0	2	BHK	933.159722	
3	Owner	0	1	2	BHK	929.921143	
4	Dealer	1	0	2	BHK	999.009247	

In [13]:

```
dataset.tail()
```

Out[13]:

	POSTED_BY	UNDER_CONSTRUCTION	RERA	BHK_NO.	BHK_OR_RK	SQUARE_FT	RE
29446	Owner	0	0	3	BHK	2500.000000	
29447	Owner	0	0	2	BHK	769.230769	
29448	Dealer	0	0	2	BHK	1022.641509	
29449	Owner	0	0	2	BHK	927.079009	
29450	Dealer	0	1	2	BHK	896.774194	

In [14]:

```
dataset.describe()
```

Out[14]:

	UNDER_CONSTRUCTION	RERA	BHK_NO.	SQUARE_FT	READY_TO_MOVE
count	29451.000000	29451.000000	29451.000000	2.945100e+04	29451.000000
mean	0.179756	0.317918	2.392279	1.980217e+04	0.820244
std	0.383991	0.465675	0.879091	1.901335e+06	0.383991
min	0.000000	0.000000	1.000000	3.000000e+00	0.000000
25%	0.000000	0.000000	2.000000	9.000211e+02	1.000000
50%	0.000000	0.000000	2.000000	1.175057e+03	1.000000
75%	0.000000	1.000000	3.000000	1.550688e+03	1.000000
max	1.000000	1.000000	20.000000	2.545455e+08	1.000000

In [15]:

```
data1=dataset.head(10)
```

In [16]:

```
data1.to_csv("home data.csv",index=False)
```

In [17]:

```
data2=pd.read_csv("home data.csv")
data2
```

Out[17]:

	POSTED_BY	UNDER_CONSTRUCTION	RERA	BHK_NO.	BHK_OR_RK	SQUARE_FT	READY
0	Owner	0	0	2	BHK	1300.236407	
1	Dealer	0	0	2	BHK	1275.000000	
2	Owner	0	0	2	BHK	933.159722	
3	Owner	0	1	2	BHK	929.921143	
4	Dealer	1	0	2	BHK	999.009247	
5	Owner	0	0	3	BHK	1250.000000	
6	Dealer	0	0	3	BHK	1495.053957	
7	Owner	0	1	3	BHK	1181.012946	
8	Dealer	0	1	2	BHK	1040.000000	
9	Owner	0	1	2	BHK	879.120879	

In [18]:

```
d=dataset["SQUARE_FT"].mean()
d
```

Out[18]:

```
19802.170190334633
```

In [19]:

```
ab=round(d,2)
ab
```

Out[19]:

```
19802.17
```

In [22]:

```
dataset["SQUARE_FT"].min()
```

Out[22]:

```
3.0
```

In [23]:

```
dataset["SQUARE_FT"].max()
```

Out[23]:

```
254545454.5
```

In [25]:

```
data=dataset[["SQUARE_FT","RERA","BHK_NO."]]
data
```

Out[25]:

	SQUARE_FT	RERA	BHK_NO.
0	1300.236407	0	2
1	1275.000000	0	2
2	933.159722	0	2
3	929.921143	1	2
4	999.009247	0	2
...	...	...	...
29446	2500.000000	0	3
29447	769.230769	0	2
29448	1022.641509	0	2
29449	927.079009	0	2
29450	896.774194	1	2

29451 rows × 3 columns

In [26]:

```
dataset.head(10)
```

Out[26]:

	POSTED_BY	UNDER_CONSTRUCTION	RERA	BHK_NO.	BHK_OR_RK	SQUARE_FT	READY
0	Owner	0	0	2	BHK	1300.236407	
1	Dealer	0	0	2	BHK	1275.000000	
2	Owner	0	0	2	BHK	933.159722	
3	Owner	0	1	2	BHK	929.921143	
4	Dealer	1	0	2	BHK	999.009247	
5	Owner	0	0	3	BHK	1250.000000	
6	Dealer	0	0	3	BHK	1495.053957	
7	Owner	0	1	3	BHK	1181.012946	
8	Dealer	0	1	2	BHK	1040.000000	
9	Owner	0	1	2	BHK	879.120879	

Working with Slicing

In [27]:

```
data=dataset.iloc[5:12,:5]
data
```

Out[27]:

	POSTED_BY	UNDER_CONSTRUCTION	RERA	BHK_NO.	BHK_OR_RK
5	Owner	0	0	3	BHK
6	Dealer	0	0	3	BHK
7	Owner	0	1	3	BHK
8	Dealer	0	1	2	BHK
9	Owner	0	1	2	BHK
10	Owner	0	0	3	BHK
11	Dealer	0	0	2	BHK

In [28]:

```
data1=dataset.iloc[6:12,[1,3,5]]
data1
```

Out[28]:

	UNDER_CONSTRUCTION	BHK_NO.	SQUARE_FT
6	0	3	1495.053957
7	0	3	1181.012946
8	0	2	1040.000000
9	0	2	879.120879
10	0	3	1350.308642
11	0	2	1333.010179

In [29]:

```
dataset["BHK_OR_RK"].unique()
```

Out[29]:

```
array(['BHK', 'RK'], dtype=object)
```

In [30]:

```
dataset["BHK_OR_RK"].nunique()
```

Out[30]:

```
2
```

In [32]:

```
dataset["BHK_OR_RK"].value_counts()
```

Out[32]:

```
BHK    29427
RK        24
Name: BHK_OR_RK, dtype: int64
```

Conditional Slicing

In [35]:

```
data=dataset.loc[dataset["City"]=="Bangalore"]
data
```

Out[35]:

E_FT	READY_TO_MOVE	RESALE	City	LONGITUDE	LATITUDE	TARGET(PRICE_IN_LACS)
36407	1	1	Bangalore	12.969910	77.597960	55.0
59722	1	1	Bangalore	12.778033	77.632191	43.0
00000	1	1	Bangalore	13.054202	77.674002	41.6
10179	1	1	Bangalore	13.040340	77.591347	110.0
77902	1	1	Bangalore	12.969910	77.597960	48.0
...	...	...	...	...	...	...
45182	1	1	Bangalore	13.031800	77.658300	42.0
40584	1	1	Bangalore	12.995893	77.761394	220.0
64497	1	1	Bangalore	13.019578	77.017574	97.0
28681	1	1	Bangalore	12.876691	77.599080	50.0
32676	0	1	Bangalore	12.995893	77.761394	62.0

In [37]:

```
data.shape
```

Out[37]:

```
(4319, 12)
```

In [38]:

```
data1=dataset.loc[(dataset["City"]=="Bangalore")|(dataset["City"]=="Ghaziabad")]
data1
```

Out[38]:

E_FT	READY_TO_MOVE	RESALE	City	LONGITUDE	LATITUDE	TARGET(PRICE_IN_LACS)
6407	1	1	Bangalore	12.969910	77.597960	55.0
9722	1	1	Bangalore	12.778033	77.632191	43.0
1143	1	1	Ghaziabad	28.642300	77.344500	62.5
0000	1	1	Bangalore	13.054202	77.674002	41.6
0179	1	1	Bangalore	13.040340	77.591347	110.0
...	...	...	...	...	...	...
5182	1	1	Bangalore	13.031800	77.658300	42.0
0584	1	1	Bangalore	12.995893	77.761394	220.0
4497	1	1	Bangalore	13.019578	77.017574	97.0
8681	1	1	Bangalore	12.876691	77.599080	50.0
2676	0	1	Bangalore	12.995893	77.761394	62.0

In [39]:

```
data1["City"].unique()
```

Out[39]:

```
array(['Bangalore', 'Ghaziabad'], dtype=object)
```

In [42]:

```
data2=dataset.loc[(dataset["City"]=="Bangalore")&(dataset["BHK_OR_RK"]=="BHK")]
data2
```

Out[42]:

	POSTED_BY	UNDER_CONSTRUCTION	RERA	BHK_NO.	BHK_OR_RK	SQUARE_FT	RE
0	Owner	0	0	2	BHK	1300.236407	
2	Owner	0	0	2	BHK	933.159722	
8	Dealer	0	1	2	BHK	1040.000000	
11	Dealer	0	0	2	BHK	1333.010179	
12	Owner	0	0	2	BHK	927.177902	
...	...	...	...	...	...	...	
29408	Dealer	0	0	2	BHK	1105.845182	
29415	Owner	0	0	4	BHK	3650.240584	
29421	Dealer	0	0	2	BHK	1409.064497	
29433	Owner	0	0	2	BHK	1195.028681	
29436	Owner	1	1	2	BHK	1175.132676	

4319 rows × 12 columns

In [46]:

```
data2=dataset.loc[((dataset["City"]=="Bangalore")|(dataset["City"]=="Ghaziabad"))&(dataset[
```

In [47]:

```
data2
```

Out[47]:

	POSTED_BY	UNDER_CONSTRUCTION	RERA	BHK_NO.	BHK_OR_RK	SQUARE_FT	RE
0	Owner	0	0	2	BHK	1300.236407	
2	Owner	0	0	2	BHK	933.159722	
3	Owner	0	1	2	BHK	929.921143	
8	Dealer	0	1	2	BHK	1040.000000	
11	Dealer	0	0	2	BHK	1333.010179	
...	...	...	...	...	...	...	
29408	Dealer	0	0	2	BHK	1105.845182	
29415	Owner	0	0	4	BHK	3650.240584	
29421	Dealer	0	0	2	BHK	1409.064497	
29433	Owner	0	0	2	BHK	1195.028681	
29436	Owner	1	1	2	BHK	1175.132676	

5404 rows × 12 columns

In [48]:

```
dataset.isnull().any()
```

Out[48]:

```
POSTED_BY               False
UNDER_CONSTRUCTION      False
RERA                    False
BHK_NO.                 False
BHK_OR_RK               False
SQUARE_FT               False
READY_TO_MOVE           False
RESALE                  False
City                     True
LONGITUDE               False
LATITUDE                False
TARGET(PRICE_IN_LACS)   False
dtype: bool
```

In [49]:

```
dataset.isnull().sum()
```

Out[49]:

```
POSTED_BY               0
UNDER_CONSTRUCTION      0
RERA                    0
BHK_NO.                 0
BHK_OR_RK               0
SQUARE_FT               0
READY_TO_MOVE           0
RESALE                  0
City                    9
LONGITUDE               0
LATITUDE                0
TARGET(PRICE_IN_LACS)   0
dtype: int64
```

In [ ]:

```

```