

EDS Assignment:4

Sarthak Vijay Jagadale(226)

Samruddhi Kinhikar(232)

Chirag Ghuge(222)

CODE:

```
import pandas as pd
```

```
df=pd.read_csv("/content/house.csv")
```

```
print(df)
```

```
Output: beds  baths  size size_units  lot_size lot_size_units  zip_code  price \
0      3    3.0  2850      sqft  4200.00      sqft    98119  1175000
1      4    5.0  3040      sqft  5002.00      sqft    98106  1057500
2      3    1.0  1290      sqft  6048.00      sqft    98125   799000
3      3    2.0  2360      sqft    0.28      acre    98188   565000
4      3    3.5  1942      sqft  1603.00      sqft    98107  1187000
..  ...  ...  ...  ...  ...  ...  ...
94     1    1.0   810      sqft  7500.00      sqft    98108   385000
95     3    2.0  1540      sqft  3755.00      sqft    98108   632500
96     3    2.5  1660      sqft  8450.00      sqft    98146   695000
97     3    2.0  1640      sqft  4320.00      sqft    98136   907500
98     2    1.0  1600      sqft  3750.00      sqft    98117   852500
```

Location

```
0    Kothrud
1    Balewadi
2      Pune
3    Alandi
4  Hinjewadi
..  ...
94  Hinjewadi
95  Lohegaon
96    Kothrud
97  Balewadi
98      Pune
```

[99 rows x 9 columns]

```
#maximum price of house
```

```
print(df['price'].max())
```

```
Output:2850000
```

```
#print price described
```

```
print(df['price'].describe())
```

```
Output:count    9.900000e+01
```

```
mean    9.263863e+05
```

```
std    5.119539e+05
```

```
min    1.970090e+05
```

```
25%    5.575000e+05
```

```
50%    7.990000e+05
```

```
75%    1.145000e+06
```

```
max    2.850000e+06
```

```
Name: price, dtype: float64
```

```
#print location with price more than 799000
```

```
print(df.loc[df['price']>799000])
```

```
Output: beds  baths  size size_units  lot_size lot_size_units  zip_code  price \
0      3    3.0  2850      sqft  4200.00      sqft    98119  1175000
1      4    5.0  3040      sqft  5002.00      sqft    98106  1057500
4      3    3.5  1942      sqft  1603.00      sqft    98107  1187000
7      4    6.0  3300      sqft  5810.00      sqft    98199  1795000
8      4    2.0  2060      sqft  4206.00      sqft    98144  1025000
10     3    2.5  1760      sqft  3630.00      sqft    98122  1135000
15     2    2.0  2241      sqft    1.09      acre    98112  2450000
16     3    1.5  2320      sqft  4000.00      sqft    98109  1275000
18     3    1.5  1750      sqft  2460.00      sqft    98102  925000
19     3    3.0  2340      sqft  6640.00      sqft    98126  974898
21     3    2.5  2320      sqft  3500.00      sqft    98103  1175000
28     4    2.5  2010      sqft  6000.00      sqft    98136  982000
32     4    4.0  3310      sqft  4313.00      sqft    98116  1215000
35     3    2.0  1910      sqft  2000.00      sqft    98109  1090000
37     3    2.0  2210      sqft  7200.00      sqft    98118  1425000
38     3    2.5  2168      sqft    NaN      NaN    98119  1700000
39     3    2.0  2760      sqft  4800.00      sqft    98105  1235000
40     2    2.5  2760      sqft    3.06      acre    98105  1440000
42     7    6.0  3780      sqft  5020.00      sqft    98119  1210000
43     4    3.0  2190      sqft  4000.00      sqft    98112  1685000
44     4    3.5  4027      sqft    0.30      acre    98199  2170000
45     4    2.5  3410      sqft  9602.00      sqft    98125  1700000
46     5    2.0  2470      sqft  3880.00      sqft    98107  905000
47     3    2.0  1850      sqft  3880.00      sqft    98117  890000
48     4    3.0  2430      sqft  5160.00      sqft    98106  995000
50     2    1.5  1290      sqft  5100.00      sqft    98117  869000
53     3    2.5  2220      sqft  8353.00      sqft    98177  1350000
56     3    3.0  2225      sqft    NaN      NaN    98101  2850000
59     3    2.5  1785      sqft  1500.00      sqft    98144  850000
60     3    1.5  1520      sqft  9405.00      sqft    98136  1200000
64     4    2.5  3200      sqft    0.28      acre    98118  1100000
65     2    2.0  1030      sqft  4507.00      sqft    98112  825000
66     3    2.5  1220      sqft  1095.00      sqft    98109  900000
69     2    2.0  1254      sqft    NaN      NaN    98121  1069000
70     4    3.0  3160      sqft  4830.00      sqft    98116  1425000
72     3    2.0  2798      sqft  3667.00      sqft    98109  1185000
73     4    2.0  2100      sqft  4799.00      sqft    98199  1000000
74     3    2.5  1620      sqft  5000.00      sqft    98117  885000
76     5    2.0  2180      sqft  4000.00      sqft    98144  927500
78     3    2.5  1830      sqft  1929.00      sqft    98116  1150000
81     2    1.0  2090      sqft  3290.00      sqft    98103  820000
83     5    5.0  2440      sqft    0.36      acre    98115  1505000
86     5    4.0  3510      sqft  5400.00      sqft    98199  1735000
87     5    3.0  2480      sqft  5300.00      sqft    98126  845000
89     9    6.0  6139      sqft  5750.00      sqft    98116  2800000
90     5    4.0  4770      sqft  9000.00      sqft    98115  1140000
93     5    3.5  4280      sqft  4000.00      sqft    98102  2300000
97     3    2.0  1640      sqft  4320.00      sqft    98136  907500
98     2    1.0  1600      sqft  3750.00      sqft    98117  852500
```

	Location
0	Kothrud
1	Balewadi
4	Hinjewadi
7	Balewadi
8	Pune
10	Hinjewadi
15	Alandi
16	Hinjewadi
18	Kothrud
19	Balewadi
21	Alandi
28	Hinjewadi
32	Pune
35	Lohegaon
37	Balewadi
38	Pune
39	Alandi
40	Hinjewadi
42	Kothrud
43	Balewadi
44	Pune
45	Alandi
46	Hinjewadi
47	Lohegaon
48	Kothrud
50	Pune
53	Lohegaon
56	Pune
59	Lohegaon
60	Kothrud
64	Hinjewadi
65	Lohegaon
66	Kothrud
69	Alandi
70	Hinjewadi
72	Kothrud
73	Balewadi
74	Pune
76	Hinjewadi
78	Kothrud
81	Alandi
83	Lohegaon
86	Pune
87	Alandi
89	Lohegaon
90	Kothrud
93	Alandi
97	Balewadi
98	Pune

```
#print location with price between 799000 and 1057500
```

```
print(df.loc[(df['price']>799000)&(df['price']<1057500)])
```

```
Output: beds  baths  size size_units  lot_size lot_size_units  zip_code  price \
```

8	4	2.0	2060	sqft	4206.0	sqft	98144	1025000
18	3	1.5	1750	sqft	2460.0	sqft	98102	925000
19	3	3.0	2340	sqft	6640.0	sqft	98126	974898
28	4	2.5	2010	sqft	6000.0	sqft	98136	982000
46	5	2.0	2470	sqft	3880.0	sqft	98107	905000
47	3	2.0	1850	sqft	3880.0	sqft	98117	890000
48	4	3.0	2430	sqft	5160.0	sqft	98106	995000
50	2	1.5	1290	sqft	5100.0	sqft	98117	869000
59	3	2.5	1785	sqft	1500.0	sqft	98144	850000
65	2	2.0	1030	sqft	4507.0	sqft	98112	825000
66	3	2.5	1220	sqft	1095.0	sqft	98109	900000
73	4	2.0	2100	sqft	4799.0	sqft	98199	1000000
74	3	2.5	1620	sqft	5000.0	sqft	98117	885000
76	5	2.0	2180	sqft	4000.0	sqft	98144	927500
81	2	1.0	2090	sqft	3290.0	sqft	98103	820000
87	5	3.0	2480	sqft	5300.0	sqft	98126	845000
97	3	2.0	1640	sqft	4320.0	sqft	98136	907500
98	2	1.0	1600	sqft	3750.0	sqft	98117	852500

Location

8	Pune
18	Kothrud
19	Balewadi
28	Hinjewadi
46	Hinjewadi
47	Lohegaon
48	Kothrud
50	Pune
59	Lohegaon
65	Lohegaon
66	Kothrud
73	Balewadi
74	Pune
76	Hinjewadi
81	Alandi
87	Alandi
97	Balewadi
98	Pune

```
#print median
print(df.median())
Output:beds      3.0
baths      2.0
size      1750.0
lot_size   4000.0
zip_code   98118.0
price     799000.0
dtype: float64
```

<ipython-input-7-541968dd2460>:13: FutureWarning: The default value of numeric_only in DataFrame.m
median is deprecated. In a future version, it will default to False. In addition, specifying 'numeric_only=None
' is deprecated. Select only valid columns or specify the value of numeric_only to silence this warning.
print(df.median())

```
#print mean of the data
print(df.mean())
Output:beds      2.959596
baths      2.202020
size      1895.808081
lot_size    3794.298824
zip_code    98127.171717
price      926386.262626
dtype: float64
```

```
# print grouped data
print(df.groupby('beds'))
Output:<pandas.core.groupby.generic.DataFrameGroupBy object at 0x7ff1c29a5390>
```

```
#print correlation between data
print(df.corr())
Output: beds   baths   size lot_size zip_code   price
beds    1.000000  0.748435  0.804521  0.230885  0.109841  0.534582
baths    0.748435  1.000000  0.766069  0.108129  0.070066  0.627121
size     0.804521  0.766069  1.000000  0.196935  0.085782  0.735033
lot_size 0.230885  0.108129  0.196935  1.000000 -0.102853  0.093168
zip_code 0.109841  0.070066  0.085782 -0.102853  1.000000 -0.009762
price    0.534582  0.627121  0.735033  0.093168 -0.009762  1.000000
```

```
#print covariance between data
print(df.cov())
Output: zip_code      price
beds      3.741703  3.517007e+05
baths      2.015976  3.484853e+05
size      2174.176149  3.597983e+08
lot_size  -7122.615623  1.115944e+08
zip_code    702.674294 -1.324775e+05
price   -132477.484333  2.620968e+11
```

```
#To check for missing values in a dataframe
df.isnull()
Output:beds baths size size_units lot_size lot_size_units zip_code price Location
0 False False False False False False False False False
1 False False False False False False False False False
2 False False False False False False False False False
3 False False False False False False False False False
4 False False False False False False False False False
... ..
94 False False False False False False False False False
95 False False False False False False False False False
96 False False False False False False False False False
97 False False False False False False False False False
98 False False False False False False False False False
99 rows × 9 columns
```

Colab link:https://colab.research.google.com/drive/1ah6RJH-IYiN2_wuQH8ETg8cZZKl4ksu?usp=sharing