Importing libraries

```
In [1]:  import numpy as np
         import pandas as pd
```

```
In [2]:  df=pd.read_csv(r'C:\Users\user\Desktop\USA_Housing123.csv')
```

Find mean, median, mode and describe

```
In [3]:  print(df.mean())
```

```
Avg. Area Income                6.858311e+04
Avg. Area House Age             5.977222e+00
Avg. Area Number of Rooms       6.987792e+00
Avg. Area Number of Bedrooms    3.981330e+00
Area Population                 3.616352e+04
Price                           1.232073e+06
dtype: float64
```

```
In [4]:  print(df.median())
```

```
Avg. Area Income                6.880429e+04
Avg. Area House Age             5.970429e+00
Avg. Area Number of Rooms       7.002902e+00
Avg. Area Number of Bedrooms    4.050000e+00
Area Population                 3.619941e+04
Price                           1.232669e+06
dtype: float64
```

In [5]: `print(df.mode())`

```
        Avg. Area Income    Avg. Area House Age    Avg. Area Number of Rooms  \
0            17796.631190               2.644304                     3.236194
1            35454.714659               2.683043                     3.950225
2            35608.986237               2.797215                     3.950973
3            35797.323122               2.797619                     3.969632
4            35963.330809               2.922736                     4.027931
...                   ...                    ...                          ...
4995        101599.670580               8.973441                    10.024375
4996        101928.858060               8.991399                    10.144988
4997        102881.120902               9.008900                    10.219902
4998        104702.724257               9.125283                    10.280022
4999        107701.748378               9.519088                    10.759588

        Avg. Area Number of Bedrooms    Area Population          Price  \
0                               4.38        172.610686   1.593866e+04
1                                NaN       3285.450538   3.114052e+04
2                                NaN       3883.448164   8.859177e+04
3                                NaN       4114.489353   1.430274e+05
4                                NaN       5727.485885   1.515271e+05
...                              ...               ...            ...
4995                             NaN      68311.695822   2.318286e+06
4996                             NaN      69553.988327   2.330290e+06
4997                             NaN      69575.449464   2.332111e+06
4998                             NaN      69592.040236   2.370231e+06
4999                             NaN      69621.713378   2.469066e+06

                                                 Address
0          000 Adkins Crescent\nSouth Teresa, AS 49642-1348
1                  000 Todd Pines\nAshleyberg, KY 90207-1179
2                   001 Steve Plaza\nJessicastad, UT 25190
3          0010 Gregory Loaf\nSouth Ericfort, VA 34651-0718
4              00149 Raymond Knolls\nNew Jason, UT 75026
...                                                  ...
4995                  Unit 9774 Box 4511\nDPO AE 44963
4996                  Unit 9778 Box 2114\nDPO AP 59374
4997               Unit 9785 Box 0790\nDPO AP 60371-0797
4998                  Unit 9831 Box 7128\nDPO AA 54705
4999               Unit 9871 Box 9037\nDPO AP 37275-9289

[5000 rows x 7 columns]
```

In [6]: `print(df.describe())`

```
       Avg. Area Income  Avg. Area House Age  Avg. Area Number of Rooms  \
count       5000.000000          5000.000000                5000.000000
mean       68583.108984             5.977222                   6.987792
std        10657.991214             0.991456                   1.005833
min        17796.631190             2.644304                   3.236194
25%        61480.562388             5.322283                   6.299250
50%        68804.286404             5.970429                   7.002902
75%        75783.338666             6.650808                   7.665871
max       107701.748378             9.519088                  10.759588

       Avg. Area Number of Bedrooms  Area Population         Price
count                   5000.000000      5000.000000  5.000000e+03
mean                       3.981330     36163.516039  1.232073e+06
std                        1.234137      9925.650114  3.531176e+05
min                        2.000000       172.610686  1.593866e+04
25%                        3.140000     29403.928702  9.975771e+05
50%                        4.050000     36199.406689  1.232669e+06
75%                        4.490000     42861.290769  1.471210e+06
max                        6.500000     69621.713378  2.469066e+06
```

b) Find sum(), cumsum(), count, min and max values

In [7]: `print(df.sum())`

```
Avg. Area Income                                          342915544.91
9799
Avg. Area House Age                                           29886.11
0176
Avg. Area Number of Rooms                                    34938.95
9255
Avg. Area Number of Bedrooms                                     1990
6.65
Area Population                                          180817580.19
2873
Price                                                    6160363270.71
1784
Address                     208 Michael Ferry Apt. 674\nLaurabury, NE 370
1...
dtype: object
```

In [8]: `print(df.cumsum())`

```
        Avg. Area Income  Avg. Area House Age  Avg. Area Number of Rooms  \
0           7.954546e+04             5.682861                   7.009188
1           1.587941e+05            11.685761                  13.740009
2           2.200812e+05            17.551651                  22.252737
3           2.834264e+05            24.739887                  27.839465
4           3.434086e+05            29.780442                  35.678853
...                  ...                  ...                        ...
4995        3.426402e+08         29860.333757               34913.654931
4996        3.427186e+08         29867.332892               34920.231694
4997        3.427820e+08         29874.583483               34925.036775
4998        3.428500e+08         29880.117871               34932.166918
4999        3.429155e+08         29886.110176               34938.959255

        Avg. Area Number of Bedrooms  Area Population         Price  \
0                               4.09    2.308680e+04  1.059034e+06
1                               7.18    6.325987e+04  2.564924e+06
2                              12.31    1.001420e+05  3.623912e+06
3                              15.57    1.344523e+05  4.884529e+06
4                              19.80    1.608064e+05  5.515473e+06
...                              ...             ...           ...
4995                        19890.99    1.806696e+08  6.155352e+09
4996                        19895.01    1.806952e+08  6.156835e+09
4997                        19897.14    1.807285e+08  6.157866e+09
4998                        19902.58    1.807711e+08  6.159064e+09
4999                        19906.65    1.808176e+08  6.160363e+09

                                                 Address
0        208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
1        208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
2        208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
3        208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
4        208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
...                                                  ...
4995     208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
4996     208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
4997     208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
4998     208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
4999     208 Michael Ferry Apt. 674\nLaurabury, NE 3701...

[5000 rows x 7 columns]
```

In [9]: `print(df.count())`

```
Avg. Area Income                5000
Avg. Area House Age             5000
Avg. Area Number of Rooms       5000
Avg. Area Number of Bedrooms    5000
Area Population                 5000
Price                           5000
Address                         5000
dtype: int64
```

In [10]:
```python
print(df.min())
```

```
Avg. Area Income                                    17796.63
119
Avg. Area House Age                                    2.644
304
Avg. Area Number of Rooms                              3.236
194
Avg. Area Number of Bedrooms
2.0
Area Population                                      172.610
686
Price                                             15938.657
923
Address                         000 Adkins Crescent\nSouth Teresa, AS 49642-1
348
dtype: object
```

In [11]:
```python
print(df.max())
```

```
Avg. Area Income                        107701.748378
Avg. Area House Age                          9.519088
Avg. Area Number of Rooms                   10.759588
Avg. Area Number of Bedrooms                      6.5
Area Population                          69621.713378
Price                                  2469065.594175
Address                   Unit 9871 Box 9037\nDPO AP 37275-9289
dtype: object
```

c) Find covariance and correlation (spearman and pearsons)

In [12]:
```python
from numpy import cov
```

In [16]:
```python
cov(df['Avg. Area House Age'],df['Avg. Area Number of Rooms'])
```

Out[16]:
```
array([[ 0.98298536, -0.00940227],
       [-0.00940227,  1.01170049]])
```

In [17]:
```python
from scipy.stats import pearsonr
from scipy.stats import spearmanr
```

In [18]:
```python
spearmanr(df['Avg. Area House Age'],df['Avg. Area Number of Rooms'])
```

Out[18]:
```
SpearmanrResult(correlation=-0.0051978531999141275, pvalue=0.713280166966196
7)
```

In [ ]: