# Karanjot Singh

In [21]:

```
1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4 import numpy as np
```

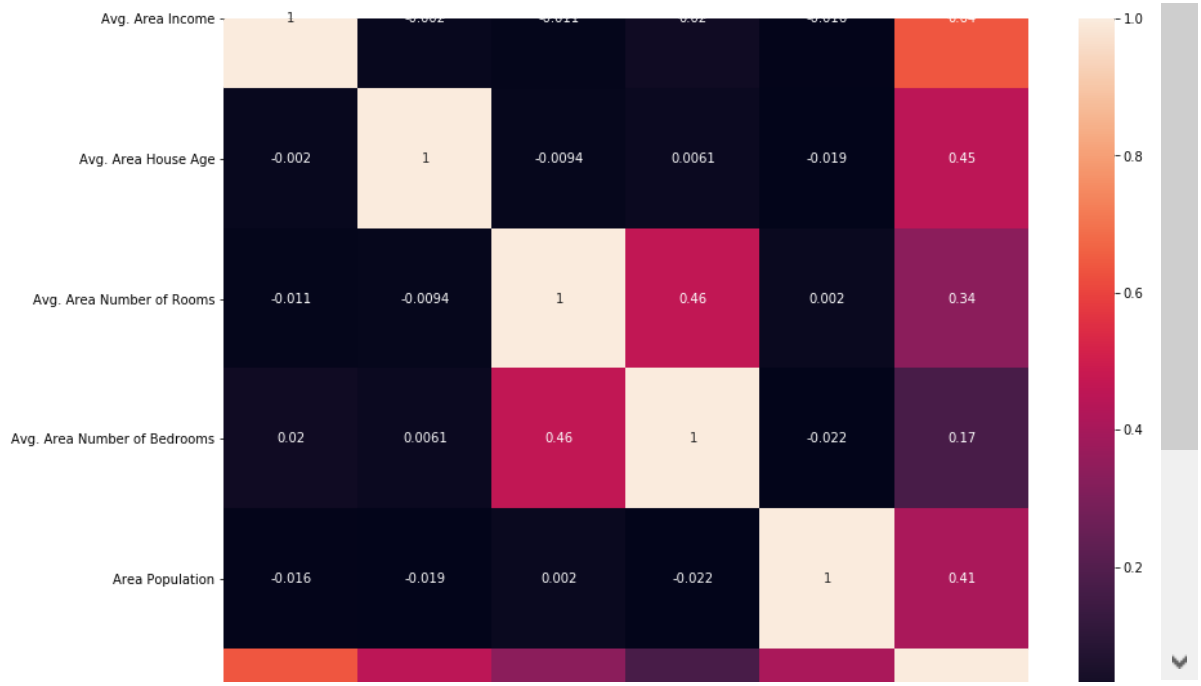
# Plot scatter graph between the price and best suited column.

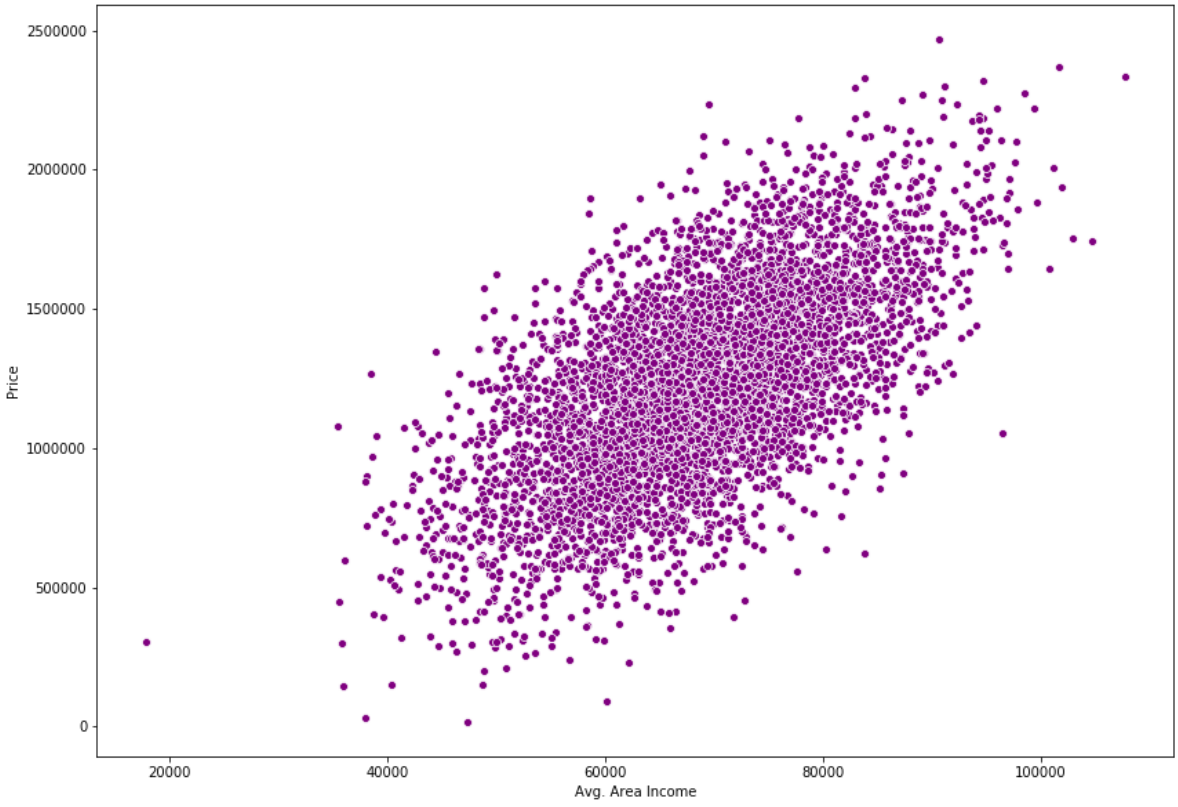
In [142]:

```
data = pd.read_csv(r'USA_Housing.csv')
#print(data)
plt.figure(figsize = (14,10))
corr  = data.corr()
print(corr)
sns.heatmap(corr,annot=True)
plt.figure(figsize = (14,10))
sns.scatterplot(data['Avg. Area Income'],data['Price'],color='purple')
plt.xlabel('Avg. Area Income')
plt.ylabel('Price')
plt.show()
```

```
                            Avg. Area Income  Avg. Area House Age  \
Avg. Area Income                    1.000000            -0.002007
Avg. Area House Age                -0.002007             1.000000
Avg. Area Number of Rooms          -0.011032            -0.009428
Avg. Area Number of Bedrooms        0.019788             0.006149
Area Population                    -0.016234            -0.018743
Price                               0.639734             0.452543

                            Avg. Area Number of Rooms  \
Avg. Area Income                            -0.011032
Avg. Area House Age                         -0.009428
Avg. Area Number of Rooms                    1.000000
Avg. Area Number of Bedrooms                 0.462695
Area Population                              0.002040
Price                                        0.335664

                            Avg. Area Number of Bedrooms  Area Population
\
Avg. Area Income                                0.019788        -0.016234
Avg. Area House Age                             0.006149        -0.018743
Avg. Area Number of Rooms                       0.462695         0.002040
Avg. Area Number of Bedrooms                    1.000000        -0.022168
Area Population                                -0.022168         1.000000
Price                                           0.171071         0.408556

                              Price
Avg. Area Income           0.639734
Avg. Area House Age        0.452543
Avg. Area Number of Rooms  0.335664
Avg. Area Number of Bedrooms  0.171071
Area Population            0.408556
Price                      1.000000
```

Plot scatter graph between the price and best suited column


Avg. Area Income
Avg. Area House Age
Avg. Area Number of Rooms
Avg. Area Number of Bedrooms
Area Population
1
-0.002
-0.0094
0.0061
-0.019
0.45
-0.011
-0.0094
0.46
0.002
0.34
0.02
0.0061
0.46
-0.022
0.17
-0.016
-0.019
0.002
-0.022
0.41
1.0
0.8
0.6
0.4
0.2


Price
2500000
2000000
1500000
1000000
500000
0
20000
40000
60000
80000
100000
Avg. Area Income

In [109]:

```
x = list(data['Avg. Area Income'])
y = list(data['Price'])
m = np.linspace(-100,100,100)
n = []
for i in m:
    k = 0
    for j in range(len(x)):
        k+=(((x[j]*i)-y[j])**2)/(len(x)*2)
    n.append(k)
print(n)
```

```
[33598054089934.332, 32459141032629.266, 31339888062646.043, 30240295179984.
93, 29160362384645.945, 28100089676628.69, 27059477055933.418, 2603852452256
0.242, 25037232076508.9, 24055599717779.49, 23093627446372.152, 221513152622
86.91, 21228663165523.387, 20325671156081.945, 19442339233962.41, 1857866739
9164.87, 17734655651689.3, 16910303991535.682, 16105612418704.064, 153205809
33194.375, 14555209535006.654, 13809498224140.895, 13083447000597.113, 12377
055864375.244, 11690324815475.414, 11023253853897.521, 10375842979641.652, 9
748092192707.625, 9140001493095.652, 8551570880805.63, 7982800355837.561, 74
33689918191.48, 6904239567867.334, 6394449304865.156, 5904319129184.947, 543
3849040826.708, 4983039039790.422, 4551889126076.114, 4140399299683.757, 374
8569560613.3853, 3376399908864.982, 3023890344438.5063, 2691040867334.025, 2
377851477551.506, 2084322175090.9458, 1810452959952.3486, 1556243832135.723
4, 1321694791641.058, 1106805838468.361, 911576972617.626, 736008194088.859
1, 580099502882.054, 443850898997.2133, 327262382434.33954, 230333953193.430
15, 153065611274.4849, 95457356677.50458, 57509189402.48965, 39221109449.439
644, 40593116818.3542, 61625211509.23377, 102317393522.07855, 162669662856.8
8794, 242682019513.662, 342354463492.40106, 461686994793.10486, 60067961341
5.775, 759332319360.4065, 937645112627.0096, 1135617993215.5671, 13532509611
26.0947, 1590544016358.591, 1847497158913.0522, 2124110388789.4692, 24203837
05987.868, 2736317110508.215, 3071910602350.5283, 3427164181514.831, 3802077
848001.061, 4196651601809.2837, 4610885442939.457, 5044779371391.581, 549833
3387165.6875, 5971547490261.757, 6464421680679.805, 6976955958419.82, 750915
0323481.768, 8061004775865.717, 8632519315571.588, 9223693942599.447, 983452
8656949.266, 10465023458621.078, 11115178347614.87, 11784993323930.52, 12474
468387568.25, 13183603538527.898, 13912398776809.521, 14660854102413.121, 15
428969515338.625, 16216745015586.135]
```

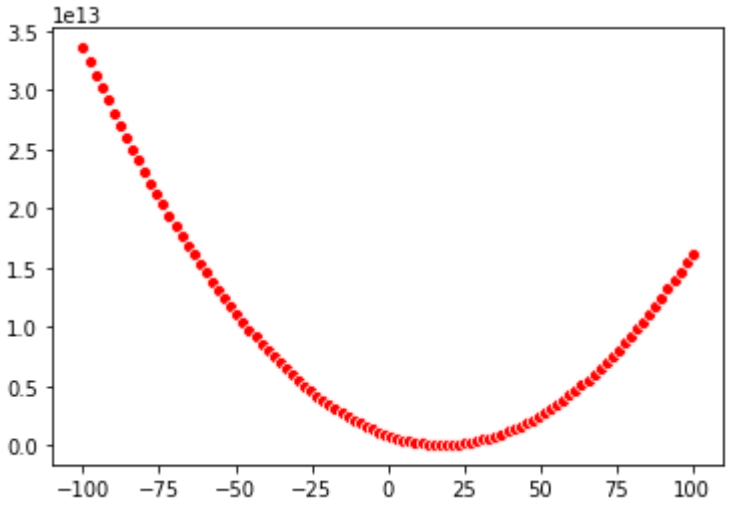
# Plot the graph between Cost function and parameters.

In [132]:

```
1 sns.scatterplot(m,n,color = 'Red')
```

Out[132]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x209e62d0b48>
```




# Find the value of parameter Q1 for which cost function is minimum.

In [139]:

```
1 min_value = m[n.index(min(n))]
2 print(min_value)
```

```
17.171717171717177
```

# Draw the Hypothesis.

In [138]:

```
l = list()
for i in range(len(x)):
    l.append(x[i]*min_value)
print(len(l))
plt.figure(figsize = (14,10))
sns.lineplot(x,l,color = 'Black')
sns.scatterplot(data['Avg. Area Income'],data['Price'],color = 'Purple')
```

```
5000
```

Out[138]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x209e6df52c8>
```