

```
In [1]: import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
import pandas as pd
```

```
In [2]: df= pd.read_csv('Downloads/USA_Housing.csv')
```

```
In [3]: df
```

Out[3]:

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Address
0	79545.458574	5.682861	7.009188	4.09	23086.800503	1.059034e+06	208 Michael Ferry 674\nLaurabury 37
1	79248.642455	6.002900	6.730821	3.09	40173.072174	1.505891e+06	188 Johnson V Suite 079\nKathleen,
2	61287.067179	5.865890	8.512727	5.13	36882.159400	1.058988e+06	9127 Eliza Stravenue\nDanielt WI 064
3	63345.240046	7.188236	5.586729	3.26	34310.242831	1.260617e+06	USS Barnett\nFP 4
4	59982.197226	5.040555	7.839388	4.23	26354.109472	6.309435e+05	USNS Raymond\r AE 0
...
4995	60567.944140	7.830362	6.137356	3.46	22837.361035	1.060194e+06	USNS Williams\r AP 30153-
4996	78491.275435	6.999135	6.576763	4.02	25616.115489	1.482618e+06	PSC 9258, 8489\nAPO AA 42
4997	63390.686886	7.250591	4.805081	2.13	33266.145490	1.030730e+06	4215 Tracy Ga Suite 076\nJoshua VA
4998	68001.331235	5.534388	7.130144	5.44	42625.620156	1.198657e+06	USS Wallace\nFP 7
4999	65510.581804	5.992305	6.792336	4.07	46501.283803	1.298950e+06	37778 George Ri Apt. 509\nEast H N

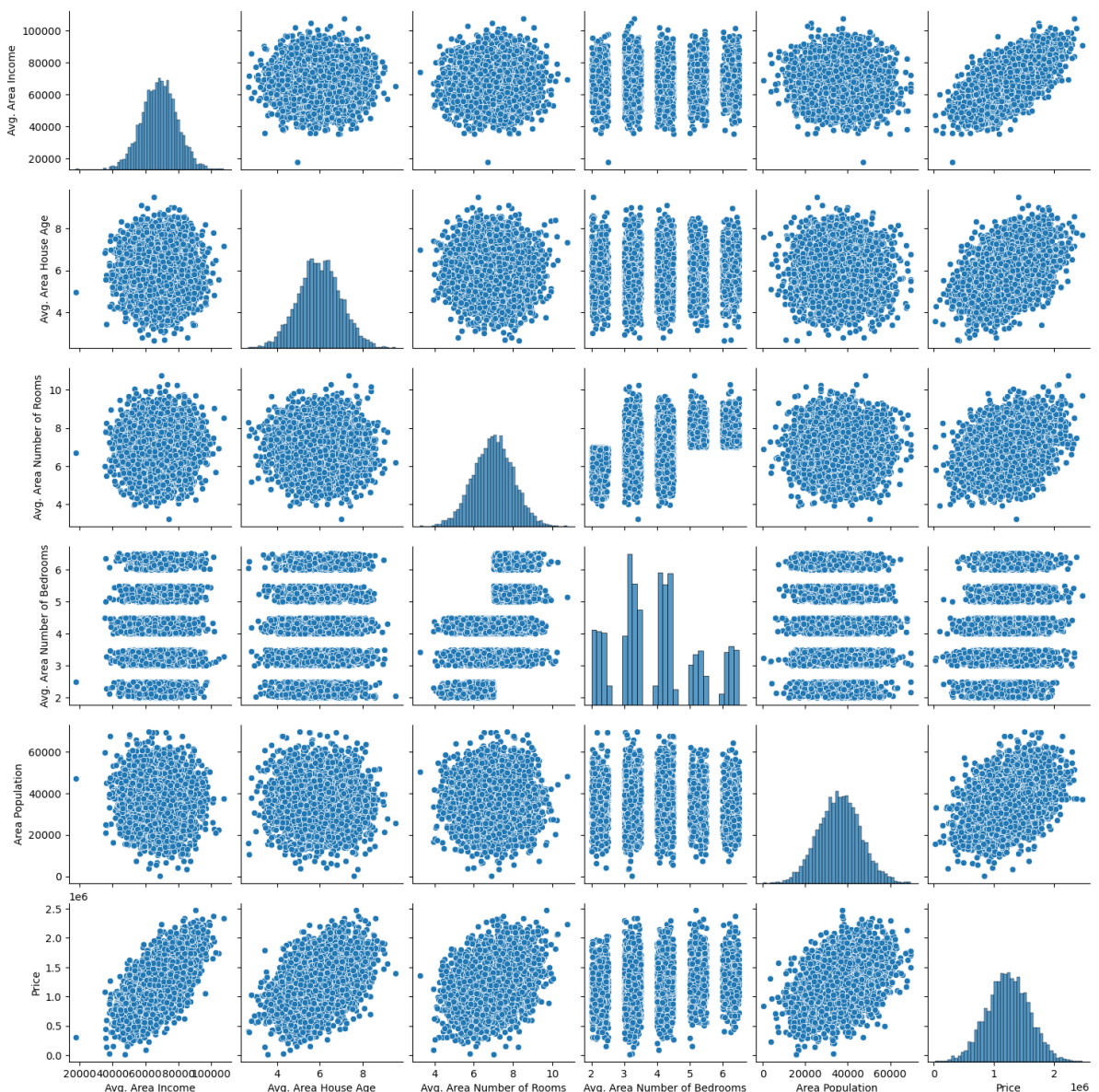
5000 rows × 7 columns

```
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5000 entries, 0 to 4999
Data columns (total 7 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   Avg. Area Income                      5000 non-null   float64
1   Avg. Area House Age                  5000 non-null   float64
2   Avg. Area Number of Rooms            5000 non-null   float64
3   Avg. Area Number of Bedrooms         5000 non-null   float64
4   Area Population                      5000 non-null   float64
5   Price                                5000 non-null   float64
6   Address                              5000 non-null   object
dtypes: float64(6), object(1)
memory usage: 273.6+ KB
```

```
In [51]: sns.pairplot(df)
```

```
Out[51]: <seaborn.axisgrid.PairGrid at 0x1a3e072ce80>
```



```
In [6]: df.corr()
```

```
C:\Users\Sony Vaio\AppData\Local\Temp\ipykernel_7532\1134722465.py:1: FutureWarning:
The default value of numeric_only in DataFrame.corr is deprecated. In a future
version, it will default to False. Select only valid columns or specify the value
of numeric_only to silence this warning.
df.corr()
```

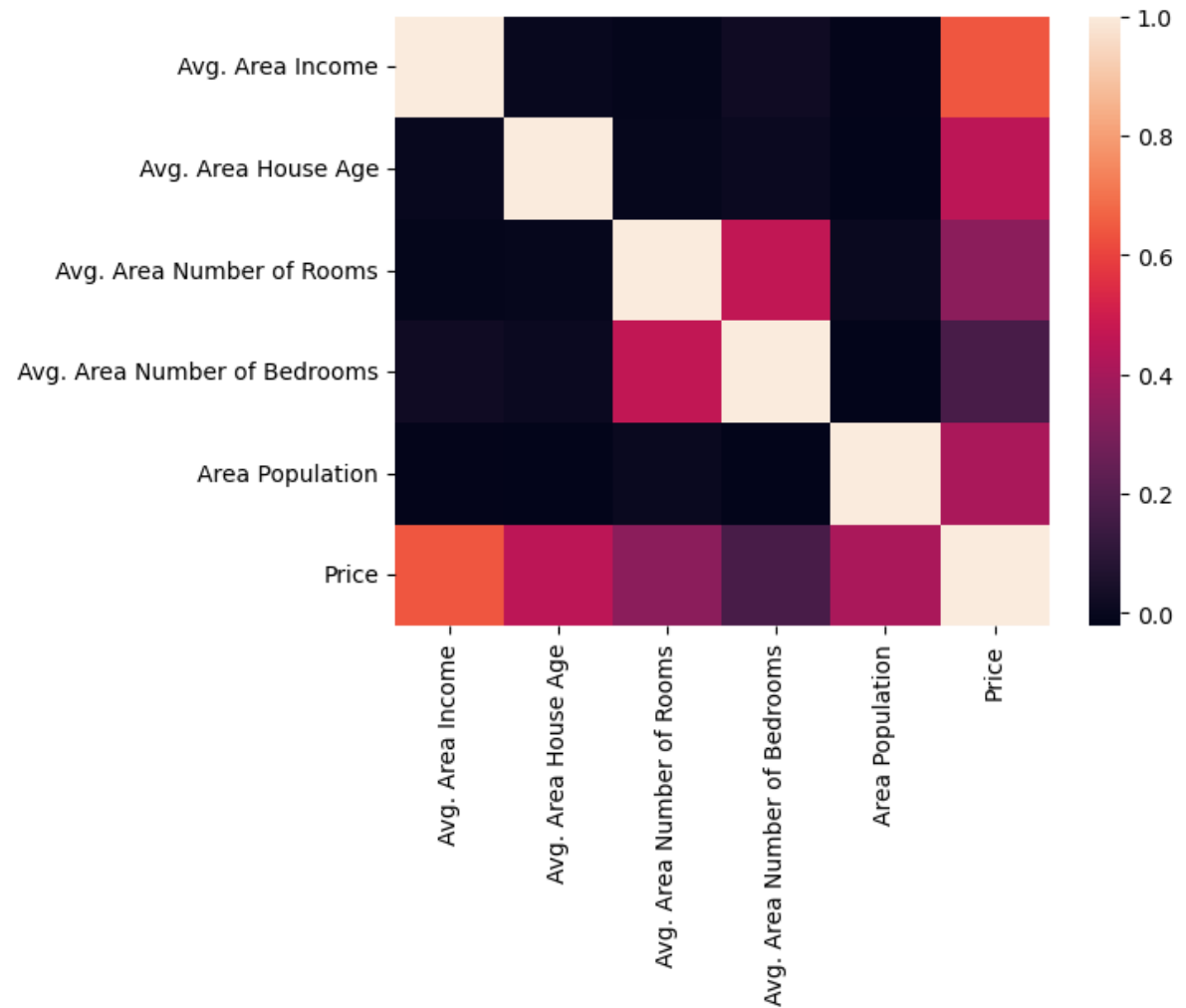
Out[6]:

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price
Avg. Area Income	1.000000	-0.002007	-0.011032	0.019788	-0.016234	0.639734
Avg. Area House Age	-0.002007	1.000000	-0.009428	0.006149	-0.018743	0.452543
Avg. Area Number of Rooms	-0.011032	-0.009428	1.000000	0.462695	0.002040	0.335664
Avg. Area Number of Bedrooms	0.019788	0.006149	0.462695	1.000000	-0.022168	0.171071
Area Population	-0.016234	-0.018743	0.002040	-0.022168	1.000000	0.408556
Price	0.639734	0.452543	0.335664	0.171071	0.408556	1.000000

In [7]: sns.heatmap(df.corr())

C:\Users\Sony Vaio\AppData\Local\Temp\ipykernel_7532\58359773.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
sns.heatmap(df.corr())

Out[7]: <Axes: >



```
In [8]: sns.distplot(df['Price'],kde='False')
```

C:\Users\Sony Vaio\AppData\Local\Temp\ipykernel_7532\2134802693.py:1: UserWarning:

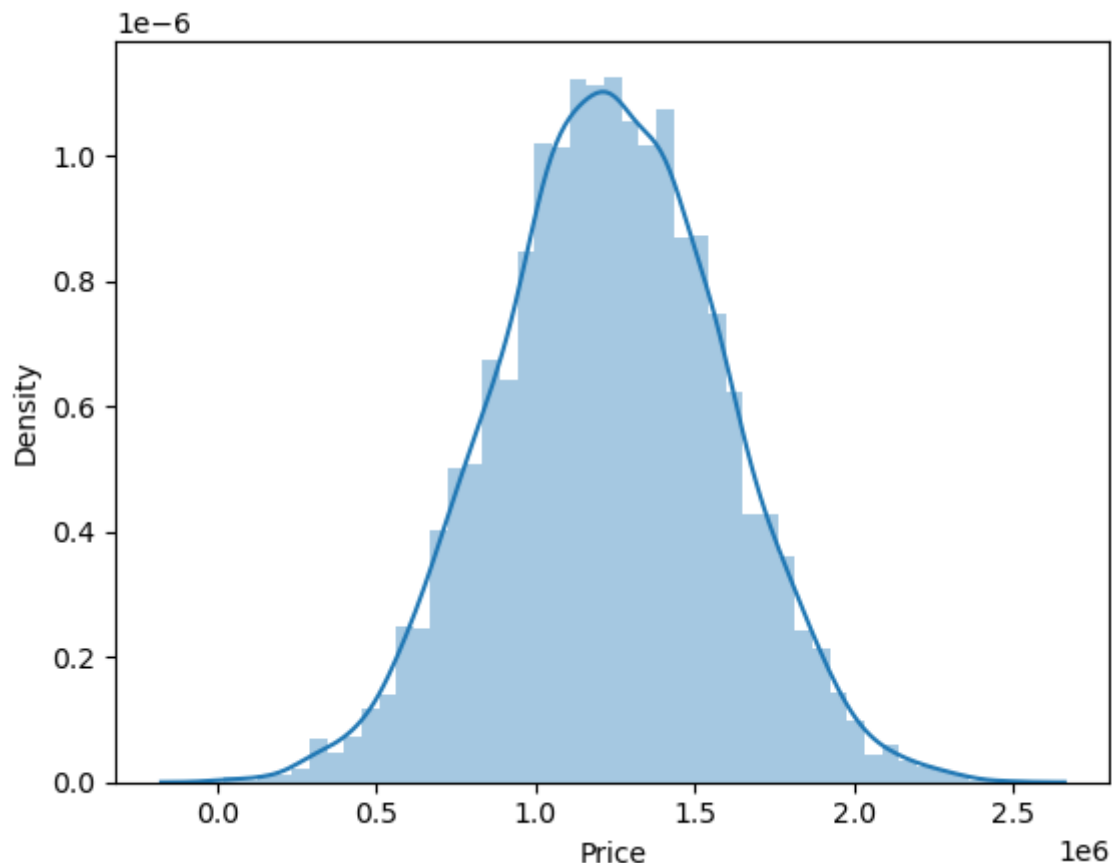
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(df['Price'],kde='False')  
<Axes: xlabel='Price', ylabel='Density'>
```

Out[8]:



```
In [9]: df.columns
```

```
Out[9]: Index(['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',  
              'Avg. Area Number of Bedrooms', 'Area Population', 'Price', 'Address'],  
            dtype='object')
```

```
In [10]: x=df[['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',  
              'Avg. Area Number of Bedrooms', 'Area Population']]  
         y=df['Price']
```

```
In [ ]:
```

```
In [54]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.4,random_state=100)
```

```
In [52]: x_train
```

Out[52]:

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population
327	61252.637672	5.189911	7.191573	4.43	40890.266029
2843	64163.541805	6.866750	6.933771	3.00	39354.789257
1960	58054.099478	5.408102	7.143542	6.13	35257.717593
4616	55317.203899	4.970859	7.145445	4.40	42317.895292
2399	62717.682889	6.022456	6.016159	2.01	38535.666604
...
4149	76717.097365	5.577625	7.197195	3.11	22276.373067
1890	54447.686052	6.148759	6.077189	2.40	50200.610108
350	81328.807304	6.483153	6.908830	3.01	31231.544403
79	64419.252638	6.954422	8.516160	6.16	39318.170755
3927	68742.987316	6.068788	7.316551	4.44	34008.881879

500 rows × 5 columns

In []:

In [57]: `X_train=np.array(x_train).reshape(15000,1)`

In [58]: `X_train.shape`

Out[58]: (15000, 1)

In [59]: `Y_train=np.array(y_train).reshape(3000,1)`

In [16]: `Y_train.shape`

Out[16]: (3000, 1)

In [60]: `lm=LinearRegression()`

In [61]: `lm.fit(x_train,y_train)`

Out[61]:

▼ LinearRegression

LinearRegression()

In [62]: `lm.intercept_`

Out[62]: -2628094.736394693

In [20]: `lm.coef_`

Out[20]: array([2.16101924e+01, 1.65387057e+05, 1.20222246e+05, -1.66224150e+01,
 1.52155668e+01])

In [63]: `coeff_df = pd.DataFrame(lm.coef_, x.columns, columns=['Coefficient'])`
`coeff_df`

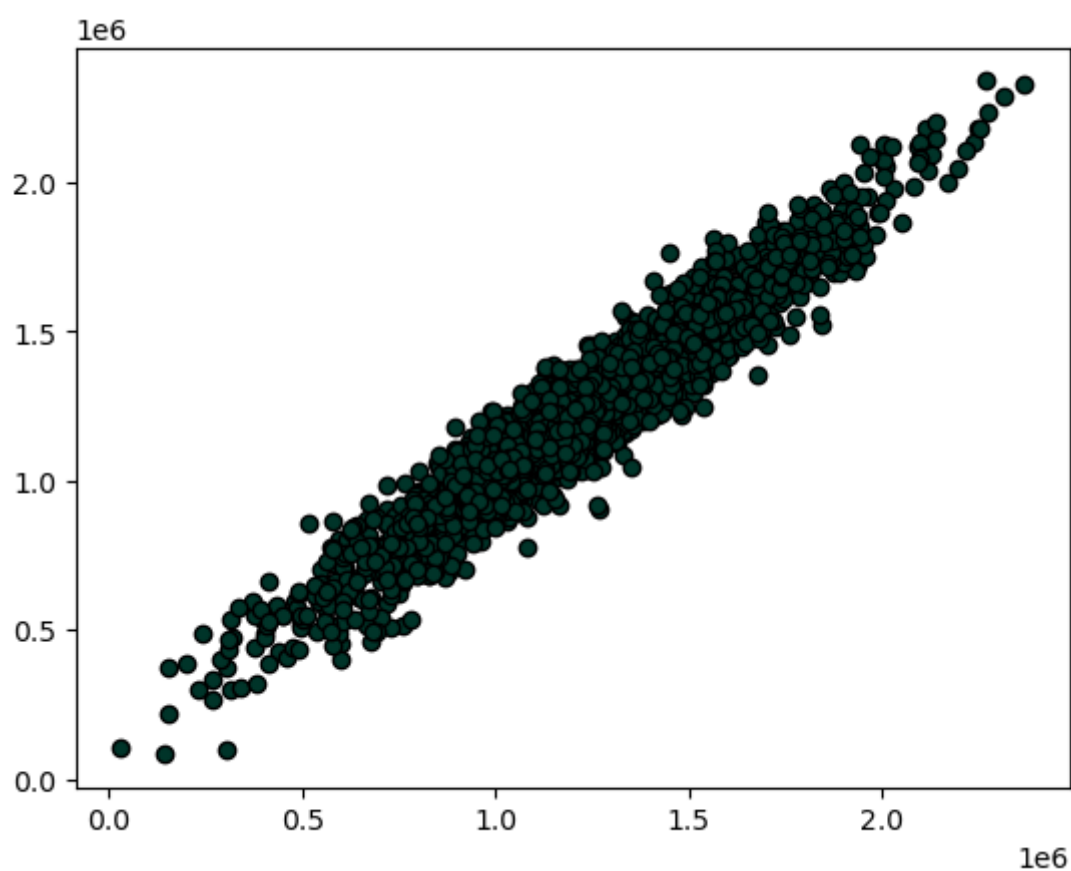
```
Out[63]:
```

	Coefficient
Avg. Area Income	21.610192
Avg. Area House Age	165387.057413
Avg. Area Number of Rooms	120222.245638
Avg. Area Number of Bedrooms	-16.622415
Area Population	15.215567

```
In [64]: X_test=np.array(x_test).reshape(10000,1)
```

```
In [78]: prediction=lm.predict(x_test)
plt.scatter(y_test, prediction, edgecolor='black',color='#003429')
```

```
Out[78]: <matplotlib.collections.PathCollection at 0x1a3ec066380>
```



```
In [ ]:
```

```
In [ ]:
```