In [1]:
```
pip install seaborn
```

```
Requirement already satisfied: seaborn in c:\reddy\python37\lib\site-packages
(0.12.2)
Requirement already satisfied: numpy!=1.24.0,>=1.17 in c:\reddy\python37\lib
\site-packages (from seaborn) (1.21.0)
Requirement already satisfied: pandas>=0.25 in c:\reddy\python37\lib\site-pac
kages (from seaborn) (1.2.5)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in c:\reddy\python37\l
ib\site-packages (from seaborn) (3.4.2)
Requirement already satisfied: typing_extensions in c:\reddy\python37\lib\sit
e-packages (from seaborn) (4.0.1)
Requirement already satisfied: cycler>=0.10 in c:\reddy\python37\lib\site-pac
kages (from matplotlib!=3.6.1,>=3.1->seaborn) (0.10.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\reddy\python37\lib\sit
e-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.3.1)
Requirement already satisfied: pillow>=6.2.0 in c:\reddy\python37\lib\site-pa
ckages (from matplotlib!=3.6.1,>=3.1->seaborn) (8.2.0)
Requirement already satisfied: pyparsing>=2.2.1 in c:\reddy\python37\lib\site
-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (2.4.7)
Requirement already satisfied: python-dateutil>=2.7 in c:\reddy\python37\lib
\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2017.3 in c:\reddy\python37\lib\site-pac
kages (from pandas>=0.25->seaborn) (2021.1)
Requirement already satisfied: six in c:\reddy\python37\lib\site-packages (fr
om cycler>=0.10->matplotlib!=3.6.1,>=3.1->seaborn) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [2]:
```
!pip install scikit-learn
```

```
Requirement already satisfied: scikit-learn in c:\reddy\python37\lib\site-pac
kages (1.0.2)
Requirement already satisfied: numpy>=1.14.6 in c:\reddy\python37\lib\site-pa
ckages (from scikit-learn) (1.21.0)
Requirement already satisfied: scipy>=1.1.0 in c:\reddy\python37\lib\site-pac
kages (from scikit-learn) (1.7.3)
Requirement already satisfied: joblib>=0.11 in c:\reddy\python37\lib\site-pac
kages (from scikit-learn) (1.2.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\reddy\python37\lib
\site-packages (from scikit-learn) (3.1.0)
```

In [3]:
```
pip install matplotlib
```

```
Requirement already satisfied: matplotlib in c:\reddy\python37\lib\site-packa
ges (3.4.2)
Requirement already satisfied: cycler>=0.10 in c:\reddy\python37\lib\site-pac
kages (from matplotlib) (0.10.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\reddy\python37\lib\sit
e-packages (from matplotlib) (1.3.1)
Requirement already satisfied: numpy>=1.16 in c:\reddy\python37\lib\site-pack
ages (from matplotlib) (1.21.0)
Requirement already satisfied: pillow>=6.2.0 in c:\reddy\python37\lib\site-pa
ckages (from matplotlib) (8.2.0)
Requirement already satisfied: pyparsing>=2.2.1 in c:\reddy\python37\lib\site
-packages (from matplotlib) (2.4.7)
Requirement already satisfied: python-dateutil>=2.7 in c:\reddy\python37\lib
\site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: six in c:\reddy\python37\lib\site-packages (fr
om cycler>=0.10->matplotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [4]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn import preprocessing,svm
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

In [5]:
```
df=pd.read_csv(r"C:\Users\Mastan Reddy\Downloads\archive (3).zip")
df
```

0	1	3.888889
1	2	4.555556
2	3	5.222222
3	4	5.888889
4	5	6.555556
...	...	...
295	296	200.555556
296	297	201.222222
297	298	201.888889
298	299	1.888889
299	300	1.888889

300 rows × 2 columns

In [6]:
```
df=pd.read_csv(r"C:\Users\Mastan Reddy\Downloads\USA_Housing.csv")
df
```

Out[6]:

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Ad
0	79545.458574	5.682861	7.009188	4.09	23086.800503	1.059034e+06	208 Michael Ferr 674\nLaurabu 3
1	79248.642455	6.002900	6.730821	3.09	40173.072174	1.505891e+06	188 Johnson Suite 079\ Kathleen,
2	61287.067179	5.865890	8.512727	5.13	36882.159400	1.058988e+06	9127 Eliz Stravenue\nDanie WI 06
3	63345.240046	7.188236	5.586729	3.26	34310.242831	1.260617e+06	USS Barnett\nFF
4	59982.197226	5.040555	7.839388	4.23	26354.109472	6.309435e+05	USNS Raymond\ AE (
...	...	...	...	...	...	...	
4995	60567.944140	7.830362	6.137356	3.46	22837.361035	1.060194e+06	USNS Williams\ AP 30153
4996	78491.275435	6.999135	6.576763	4.02	25616.115489	1.482618e+06	PSC 925 8489\nAPO AA 4
4997	63390.686886	7.250591	4.805081	2.13	33266.145490	1.030730e+06	4215 Tracy G Suite 076\nJoshu VA
4998	68001.331235	5.534388	7.130144	5.44	42625.620156	1.198657e+06	USS Wallace\nFF
4999	65510.581804	5.992305	6.792336	4.07	46501.283803	1.298950e+06	37778 George R Apt. 509\nEast N

5000 rows × 7 columns

In [7]:
```
df.head()
```

Out[7]:

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Addres
0	79545.458574	5.682861	7.009188	4.09	23086.800503	1.059034e+06	208 Michael Ferry Ap 674\nLaurabury, N 3701
1	79248.642455	6.002900	6.730821	3.09	40173.072174	1.505891e+06	188 Johnson Viev Suite 079\nLal Kathleen, CA
2	61287.067179	5.865890	8.512727	5.13	36882.159400	1.058988e+06	9127 Elizabe Stravenue\nDanieltow WI 06482
3	63345.240046	7.188236	5.586729	3.26	34310.242831	1.260617e+06	USS Barnett\nFPO A 448
4	59982.197226	5.040555	7.839388	4.23	26354.109472	6.309435e+05	USNS Raymond\nFP AE 093

In [8]:
```
df.head(10)
```

Out[8]:

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Addres
0	79545.458574	5.682861	7.009188	4.09	23086.800503	1.059034e+06	208 Michael Ferry Ap 674\nLaurabury, N 3701
1	79248.642455	6.002900	6.730821	3.09	40173.072174	1.505891e+06	188 Johnson Viev Suite 079\nLal Kathleen, CA
2	61287.067179	5.865890	8.512727	5.13	36882.159400	1.058988e+06	9127 Elizabe Stravenue\nDanieltow WI 06482
3	63345.240046	7.188236	5.586729	3.26	34310.242831	1.260617e+06	USS Barnett\nFPO A 448
4	59982.197226	5.040555	7.839388	4.23	26354.109472	6.309435e+05	USNS Raymond\nFP AE 093
5	80175.754159	4.988408	6.104512	4.04	26748.428425	1.068138e+06	06039 Jennifer Islan Apt. 443\nTracypo KS
6	64698.463428	6.025336	8.147760	3.41	60828.249085	1.502056e+06	4759 Daniel Shoa Sui 442\nNguyenburgh, C
7	78394.339278	6.989780	6.620478	2.42	36516.358972	1.573937e+06	972 Joy Viaduct\nLake Williar TN 17778-64
8	59927.660813	5.362126	6.393121	2.30	29387.396003	7.988695e+05	USS Gilbert\nFPO A 209
9	81885.927184	4.423672	8.167688	6.10	40149.965749	1.545155e+06	Unit 9446 B 0958\nDPO AE 970

In [9]:
```
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 [10]:
```
df.describe()
```

Out[10]:

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5.000000e+03
mean	68583.108984	5.977222	6.987792	3.981330	36163.516039	1.232073e+06
std	10657.991214	0.991456	1.005833	1.234137	9925.650114	3.531176e+05
min	17796.631190	2.644304	3.236194	2.000000	172.610686	1.593866e+04
25%	61480.562388	5.322283	6.299250	3.140000	29403.928702	9.975771e+05
50%	68804.286404	5.970429	7.002902	4.050000	36199.406689	1.232669e+06
75%	75783.338666	6.650808	7.665871	4.490000	42861.290769	1.471210e+06
max	107701.748378	9.519088	10.759588	6.500000	69621.713378	2.469066e+06

In [11]:
```
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 [12]:
```
df.fillna(method='ffill',inplace=True)
```

In [13]:
```
x=np.array(df['Area']).reshape(-1,1)
y=np.array(df['pri']).reshape(-1,1)
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
C:\Reddy\Python37\lib\site-packages\pandas\core\indexes\base.py in get_loc(se
lf, key, method, tolerance)
   3080             try:
-> 3081                 return self._engine.get_loc(casted_key)
   3082             except KeyError as err:

pandas\_libs\index.pyx in pandas._libs.index.IndexEngine.get_loc()

pandas\_libs\index.pyx in pandas._libs.index.IndexEngine.get_loc()

pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHas
hTable.get_item()

pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHas
hTable.get_item()

KeyError: 'Area'

The above exception was the direct cause of the following exception:

KeyError                                  Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_4572\1691324301.py in <module>
----> 1 x=np.array(df['Area']).reshape(-1,1)
      2 y=np.array(df['pri']).reshape(-1,1)

C:\Reddy\Python37\lib\site-packages\pandas\core\frame.py in __getitem__(self,
key)
   3022             if self.columns.nlevels > 1:
   3023                 return self._getitem_multilevel(key)
-> 3024             indexer = self.columns.get_loc(key)
   3025             if is_integer(indexer):
   3026                 indexer = [indexer]

C:\Reddy\Python37\lib\site-packages\pandas\core\indexes\base.py in get_loc(se
lf, key, method, tolerance)
   3081                 return self._engine.get_loc(casted_key)
   3082             except KeyError as err:
-> 3083                 raise KeyError(key) from err
   3084
   3085         if tolerance is not None:

KeyError: 'Area'
```

In [ ]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25)
#splitting the data into training and testing data
regr=LinearRegression()
regr.fit(x_train,y_train)
print(regr.score(x_test,y_test))
```

In [ ]:

In [ ]:

In [ ]:

In [ ]: