

k-task-1-house-price-prediction-1

January 27, 2024

```
[2]: import numpy as np
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score
data=pd.read_csv("data.csv")
data
```

```
[2]:
```

	date	price	bedrooms	bathrooms	sqft_living \
0	2014-05-02 00:00:00	3.130000e+05	3.0	1.50	1340
1	2014-05-02 00:00:00	2.384000e+06	5.0	2.50	3650
2	2014-05-02 00:00:00	3.420000e+05	3.0	2.00	1930
3	2014-05-02 00:00:00	4.200000e+05	3.0	2.25	2000
4	2014-05-02 00:00:00	5.500000e+05	4.0	2.50	1940
...
4595	2014-07-09 00:00:00	3.081667e+05	3.0	1.75	1510
4596	2014-07-09 00:00:00	5.343333e+05	3.0	2.50	1460
4597	2014-07-09 00:00:00	4.169042e+05	3.0	2.50	3010
4598	2014-07-10 00:00:00	2.034000e+05	4.0	2.00	2090
4599	2014-07-10 00:00:00	2.206000e+05	3.0	2.50	1490

	sqft_lot	floors	waterfront	view	condition	sqft_above \
0	7912	1.5	0	0	3	1340
1	9050	2.0	0	4	5	3370
2	11947	1.0	0	0	4	1930
3	8030	1.0	0	0	4	1000
4	10500	1.0	0	0	4	1140
...
4595	6360	1.0	0	0	4	1510
4596	7573	2.0	0	0	3	1460
4597	7014	2.0	0	0	3	3010
4598	6630	1.0	0	0	3	1070
4599	8102	2.0	0	0	4	1490

	sqft_basement	yr_built	yr_renovated	street \
0	0	1955	2005	18810 Densmore Ave N
1	280	1921	0	709 W Blaine St

2	0	1966	0	26206-26214 143rd Ave SE
3	1000	1963	0	857 170th Pl NE
4	800	1976	1992	9105 170th Ave NE
...
4595	0	1954	1979	501 N 143rd St
4596	0	1983	2009	14855 SE 10th Pl
4597	0	2009	0	759 Ilwaco Pl NE
4598	1020	1974	0	5148 S Creston St
4599	0	1990	0	18717 SE 258th St

	city	state	zip	country
0	Shoreline	WA	98133	USA
1	Seattle	WA	98119	USA
2	Kent	WA	98042	USA
3	Bellevue	WA	98008	USA
4	Redmond	WA	98052	USA
...
4595	Seattle	WA	98133	USA
4596	Bellevue	WA	98007	USA
4597	Renton	WA	98059	USA
4598	Seattle	WA	98178	USA
4599	Covington	WA	98042	USA

[4600 rows x 18 columns]

```
[3]: data.sum()
```

```
[3]: date          2014-05-02 00:00:002014-05-02 00:00:002014-05-...
price              2539029746.976785
bedrooms           15644.0
bathrooms          9939.75
sqft_living        9840996
sqft_lot           68321574
floors             6955.5
waterfront         33
view               1107
condition          15878
sqft_above         8405421
sqft_basement      1435575
yr_built           9065617
yr_renovated       3719598
street             18810 Densmore Ave N709 W Blaine St26206-26214...
city               ShorelineSeattleKentBellevueRedmondSeattleRedm...
statezip           WA 98133WA 98119WA 98042WA 98008WA 98052WA 981...
country            USAUSAUSAUSAUSAUSAUSAUSAUSAUSAUSAUSAUSAUSA...
dtype: object
```

```
[4]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4600 entries, 0 to 4599
Data columns (total 18 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   date                  4600 non-null   object  
1   price                 4600 non-null   float64  
2   bedrooms              4600 non-null   float64  
3   bathrooms              4600 non-null   float64  
4   sqft_living           4600 non-null   int64  
5   sqft_lot              4600 non-null   int64  
6   floors                4600 non-null   float64  
7   waterfront            4600 non-null   int64  
8   view                  4600 non-null   int64  
9   condition             4600 non-null   int64  
10  sqft_above            4600 non-null   int64  
11  sqft_basement         4600 non-null   int64  
12  yr_built              4600 non-null   int64  
13  yr_renovated          4600 non-null   int64  
14  street                4600 non-null   object  
15  city                  4600 non-null   object  
16  statezip              4600 non-null   object  
17  country               4600 non-null   object  
dtypes: float64(4), int64(9), object(5)
memory usage: 647.0+ KB
```

```
[5]: data.describe()
```

```
[5]:
```

	price	bedrooms	bathrooms	sqft_living	sqft_lot \
count	4.600000e+03	4600.000000	4600.000000	4600.000000	4.600000e+03
mean	5.519630e+05	3.400870	2.160815	2139.346957	1.485252e+04
std	5.638347e+05	0.908848	0.783781	963.206916	3.588444e+04
min	0.000000e+00	0.000000	0.000000	370.000000	6.380000e+02
25%	3.228750e+05	3.000000	1.750000	1460.000000	5.000750e+03
50%	4.609435e+05	3.000000	2.250000	1980.000000	7.683000e+03
75%	6.549625e+05	4.000000	2.500000	2620.000000	1.100125e+04
max	2.659000e+07	9.000000	8.000000	13540.000000	1.074218e+06

	floors	waterfront	view	condition	sqft_above \
count	4600.000000	4600.000000	4600.000000	4600.000000	4600.000000
mean	1.512065	0.007174	0.240652	3.451739	1827.265435
std	0.538288	0.084404	0.778405	0.677230	862.168977
min	1.000000	0.000000	0.000000	1.000000	370.000000
25%	1.000000	0.000000	0.000000	3.000000	1190.000000
50%	1.500000	0.000000	0.000000	3.000000	1590.000000

75%	2.000000	0.000000	0.000000	4.000000	2300.000000
max	3.500000	1.000000	4.000000	5.000000	9410.000000

	sqft_basement	yr_built	yr_renovated
count	4600.000000	4600.000000	4600.000000
mean	312.081522	1970.786304	808.608261
std	464.137228	29.731848	979.414536
min	0.000000	1900.000000	0.000000
25%	0.000000	1951.000000	0.000000
50%	0.000000	1976.000000	0.000000
75%	610.000000	1997.000000	1999.000000
max	4820.000000	2014.000000	2014.000000

```
[6]: data.isnull()
```

```
[6]:
```

	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	\
0	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	
...	
4595	False	False	False	False	False	False	False	
4596	False	False	False	False	False	False	False	
4597	False	False	False	False	False	False	False	
4598	False	False	False	False	False	False	False	
4599	False	False	False	False	False	False	False	

	waterfront	view	condition	sqft_above	sqft_basement	yr_built	\
0	False	False	False	False	False	False	
1	False	False	False	False	False	False	
2	False	False	False	False	False	False	
3	False	False	False	False	False	False	
4	False	False	False	False	False	False	
...	
4595	False	False	False	False	False	False	
4596	False	False	False	False	False	False	
4597	False	False	False	False	False	False	
4598	False	False	False	False	False	False	
4599	False	False	False	False	False	False	

	yr_renovated	street	city	statezip	country
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False

4595	False	False	False	False	False
4596	False	False	False	False	False
4597	False	False	False	False	False
4598	False	False	False	False	False
4599	False	False	False	False	False

[4600 rows x 18 columns]

```
[9]: import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=(12, 12))
sns.heatmap(data.corr(), cmap = 'PiYG', fmt = '.2f', linewidths = 3, annot = True)
```

C:\Users\HP\AppData\Local\Temp\ipykernel_16056\990471310.py:4: 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(data.corr(), cmap = 'PiYG', fmt = '.2f', linewidths = 3, annot = True)
```

[9]: <Axes: >



[]: