# housing-prices

March 31, 2024

[1]:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:

```
df = pd.read_csv("Housing.csv")
```

[3]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 545 entries, 0 to 544
Data columns (total 13 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   price             545 non-null    int64
 1   area              545 non-null    int64
 2   bedrooms          545 non-null    int64
 3   bathrooms         545 non-null    int64
 4   stories           545 non-null    int64
 5   mainroad          545 non-null    object
 6   guestroom         545 non-null    object
 7   basement          545 non-null    object
 8   hotwaterheating   545 non-null    object
 9   airconditioning   545 non-null    object
 10  parking           545 non-null    int64
 11  prefarea          545 non-null    object
 12  furnishingstatus  545 non-null    object
dtypes: int64(6), object(7)
memory usage: 55.5+ KB
```

[4]:

```
df.describe()
```

[4]:

```
              price         area    bedrooms   bathrooms     stories  \
count  5.450000e+02   545.000000  545.000000  545.000000  545.000000
mean   4.766729e+06  5150.541284    2.965138    1.286239    1.805505
std    1.870440e+06  2170.141023    0.738064    0.502470    0.867492
min    1.750000e+06  1650.000000    1.000000    1.000000    1.000000
25%    3.430000e+06  3600.000000    2.000000    1.000000    1.000000
50%    4.340000e+06  4600.000000    3.000000    1.000000    2.000000
```

```
75%    5.740000e+06   6360.000000    3.000000    2.000000    2.000000
max    1.330000e+07  16200.000000    6.000000    4.000000    4.000000

          parking
count  545.000000
mean     0.693578
std      0.861586
min      0.000000
25%      0.000000
50%      0.000000
75%      1.000000
max      3.000000
```

[6]:
```
df.shape
```

[6]:
```
(545, 13)
```

[10]:
```
df.isnull().sum()
```

[10]:
```
price               0
area                0
bedrooms            0
bathrooms           0
stories             0
mainroad            0
guestroom           0
basement            0
hotwaterheating     0
airconditioning     0
parking             0
prefarea            0
furnishingstatus    0
dtype: int64
```

[72]:
```
a = df.select_dtypes(include=["float64", "int64"])
a.corr()
```

[72]:
```
              price      area  bedrooms  bathrooms   stories   parking
price      1.000000  0.535997  0.366494   0.517545  0.420712  0.384394
area       0.535997  1.000000  0.151858   0.193820  0.083996  0.352980
bedrooms   0.366494  0.151858  1.000000   0.373930  0.408564  0.139270
bathrooms  0.517545  0.193820  0.373930   1.000000  0.326165  0.177496
stories    0.420712  0.083996  0.408564   0.326165  1.000000  0.045547
parking    0.384394  0.352980  0.139270   0.177496  0.045547  1.000000
```

[61]:
```
c = df.select_dtypes(include=["object"])
c
```

[61]:

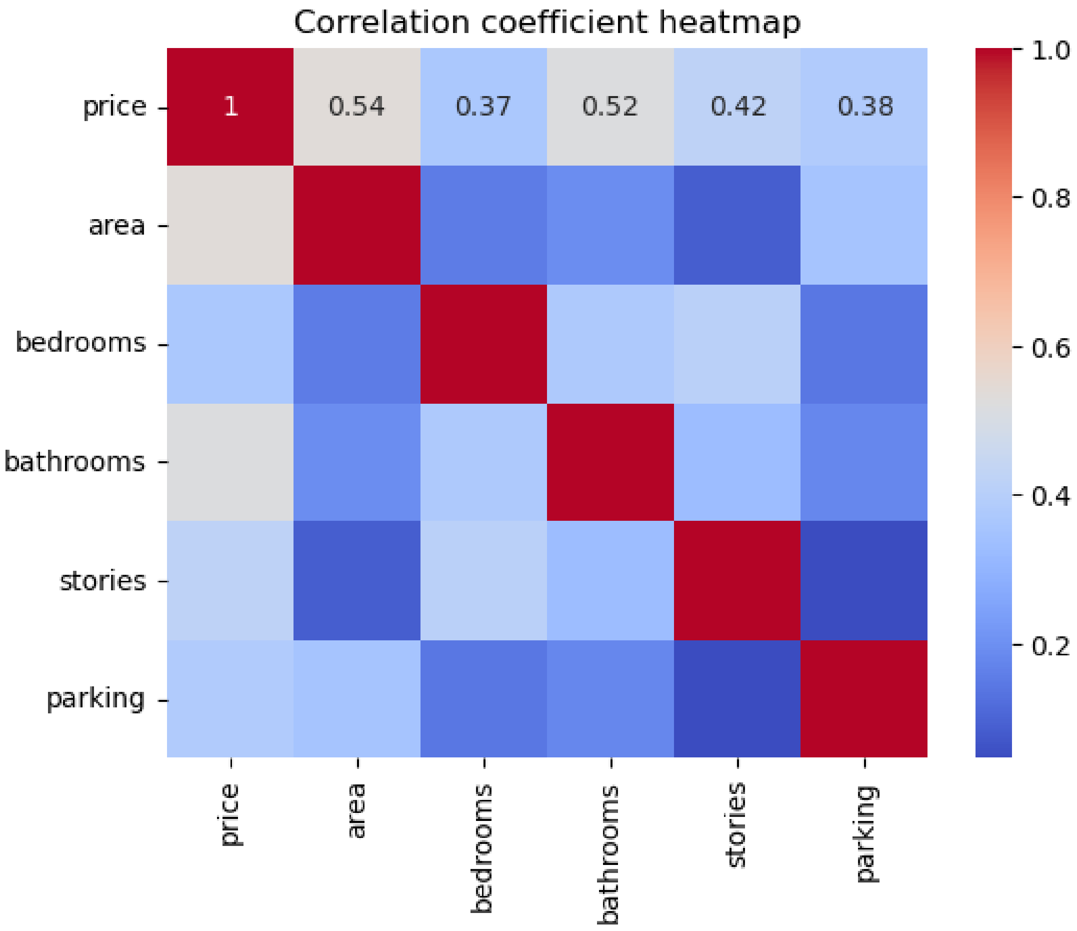
```
    mainroad guestroom basement hotwaterheating airconditioning prefarea  \
0        yes        no       no              no             yes      yes
1        yes        no       no              no             yes       no
2        yes        no      yes              no              no      yes
3        yes        no      yes              no             yes      yes
4        yes       yes      yes              no             yes       no
..       …       …      …             …             …      …
540      yes        no      yes              no              no       no
541       no        no       no              no              no       no
542      yes        no       no              no              no       no
543       no        no       no              no              no       no
544      yes        no       no              no              no       no

    furnishingstatus
0          furnished
1          furnished
2     semi-furnished
3          furnished
4          furnished
..               …
540      unfurnished
541   semi-furnished
542      unfurnished
543        furnished
544      unfurnished

[545 rows x 7 columns]
```

[46]:

```python
sns.heatmap(a, annot=True, cmap="coolwarm")
plt.title("Correlation coefficient heatmap")
plt.show()
```



### 0.0.1 Linear regression models

[79]:

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
from scipy import stats
lr = LinearRegression()
price = df[["price"]]
for i in range(5):
    b = df[a.columns[i + 1]]
    lr.fit(price, b)
    predict1 = lr.predict(price)
    r2 = r2_score(b, predict1)
    corr, p_value = stats.pearsonr(df["price"], b)
    print(f"Column: {a.columns[i + 1]}")
    print(f"Correlation Coefficient: {corr:.2f}")
    print(f"P-value: {p_value:.5f}")
    print(f"R-squared Score: {r2:.2f}")
    print("-------------------------")
```

```
Column: area
Correlation Coefficient: 0.54
P-value: 0.00000
R-squared Score: 0.29
-------------------------
Column: bedrooms
Correlation Coefficient: 0.37
P-value: 0.00000
R-squared Score: 0.13
-------------------------
Column: bathrooms
Correlation Coefficient: 0.52
P-value: 0.00000
R-squared Score: 0.27
-------------------------
Column: stories
Correlation Coefficient: 0.42
P-value: 0.00000
R-squared Score: 0.18
-------------------------
Column: parking
Correlation Coefficient: 0.38
P-value: 0.00000
R-squared Score: 0.15
-------------------------
```

### 0.0.2 Chi square test

[65]:

```
import pandas as pd
from scipy.stats import chi2_contingency

for column in c.columns:
    contingency_table = pd.crosstab(c[column], df['price'])
    chi2, p_value, _, _ = chi2_contingency(contingency_table)
    print(f"Chi-square test results for {column}:")
    print(f"Chi2 statistic: {chi2}")
    print(f"P-value: {p_value}")
    print("---------------------------------")
```

```
Chi-square test results for mainroad:
Chi2 statistic: 243.54213081891652
P-value: 0.1131151158786041
---------------------------------
Chi-square test results for guestroom:
Chi2 statistic: 302.3012447569234
```

```
P-value: 0.0001358840603675242
----------------------------------
Chi-square test results for basement:
Chi2 statistic: 264.63055645420036
P-value: 0.016889737890880243
----------------------------------
Chi-square test results for hotwaterheating:
Chi2 statistic: 235.1930903002226
P-value: 0.20202007185044277
----------------------------------
Chi-square test results for airconditioning:
Chi2 statistic: 290.700480170766
P-value: 0.0007184283964280383
----------------------------------
Chi-square test results for prefarea:
Chi2 statistic: 292.6905574192606
P-value: 0.0005457836771216665
----------------------------------
Chi-square test results for furnishingstatus:
Chi2 statistic: 509.0891908803235
P-value: 0.00886496129897742
----------------------------------
```