

Simple linear regression - Exercise Solution

```

1 You are given a real estate dataset.
2
3 Real estate is one of those examples that every regression course goes
  through as it is extremely easy to understand and there is a (almost
  always) certain causal relationship to be found.
4
5 The data is located in the file: 'real_estate_price_size.csv'.
6
7 You are expected to create a simple linear regression (similar to the one
  in the lecture), using the new data.
8
9 Apart from that, please:
10 - Create a scatter plot (with or without a regression line)
11 - Calculate the R-squared
12 - Display the intercept and coefficient(s)
13 - Using the model make a prediction about an apartment with size 750
   sq.ft.
14
15 Note: In this exercise, the dependent variable is 'price', while the
   independent variable is 'size'.
16
17 Good luck!

```

```

1 ## Import the relevant libraries

```

In [15]:

```

1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5 sns.set()
6
7 from sklearn.linear_model import LinearRegression

```

```

1 ## Load the data

```

In [2]:

```

1 data = pd.read_csv('real_estate_price_size.csv')
2 data.head()

```

Out[2]:

	price	size
0	234314.144	643.09
1	228581.528	656.22
2	281626.336	487.29
3	401255.608	1504.75
4	458674.256	1275.46

In [3]: 1 data.describe()

Out[3]:

	price	size
count	100.000000	100.000000
mean	292289.470160	853.024200
std	77051.727525	297.941951
min	154282.128000	479.750000
25%	234280.148000	643.330000
50%	280590.716000	696.405000
75%	335723.696000	1029.322500
max	500681.128000	1842.510000

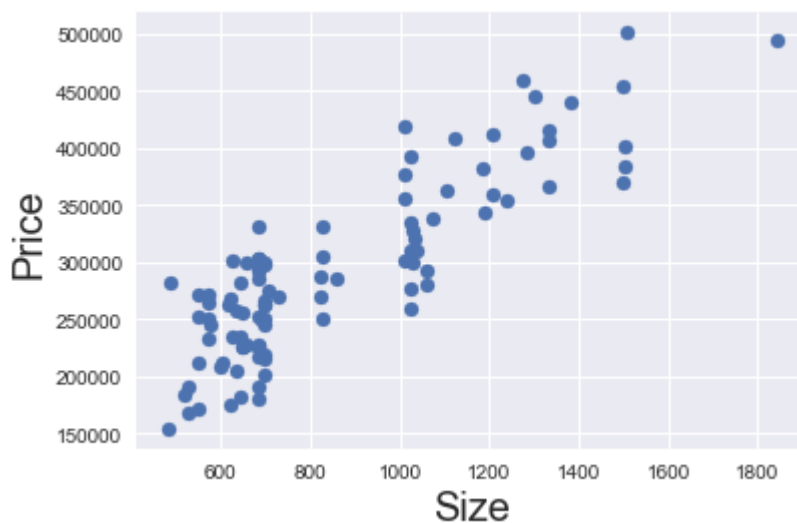
1 **## Create the regression**

1 **### Declare the dependent and the independent variables**

In [5]: 1 x = data['size']
2 y = data['price']

1 **### Explore the data**

In [6]: 1 plt.scatter(x,y)
2 plt.xlabel('Size',fontsize=20)
3 plt.ylabel('Price',fontsize=20)
4 plt.show()



1 **### Transform the inputs into a matrix (2D object)**

In [7]: 1 x_matrix = x.values.reshape(-1,1)

```
1 ### Regression itself
```

```
In [9]: 1 reg = LinearRegression()  
        2 reg.fit(x_matrix,y)
```

```
Out[9]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=False)
```

```
1 ### Calculate the R-squared
```

```
In [10]: 1 reg.score(x_matrix,y)
```

```
Out[10]: 0.7447391865847586
```

```
1 ### Find the intercept
```

```
In [12]: 1 reg.intercept_
```

```
Out[12]: 101912.60180122912
```

```
1 ### Find the coefficients
```

```
In [13]: 1 reg.coef_
```

```
Out[13]: array([223.17874259])
```

```
1 ### Making predictions
```

```
2  
3 You find an apartment online with a size of 750 sq.ft.  
4  
5 All else equal what should be its price according to the model?
```

```
In [14]: 1 reg.predict(750)
```

```
Out[14]: array([269296.65874718])
```