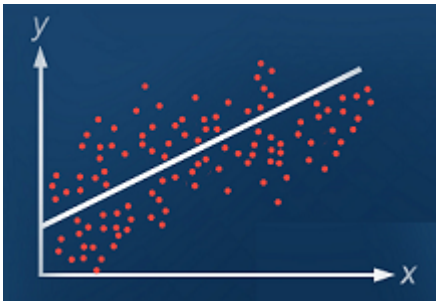
Linear Regression with one variable in Python



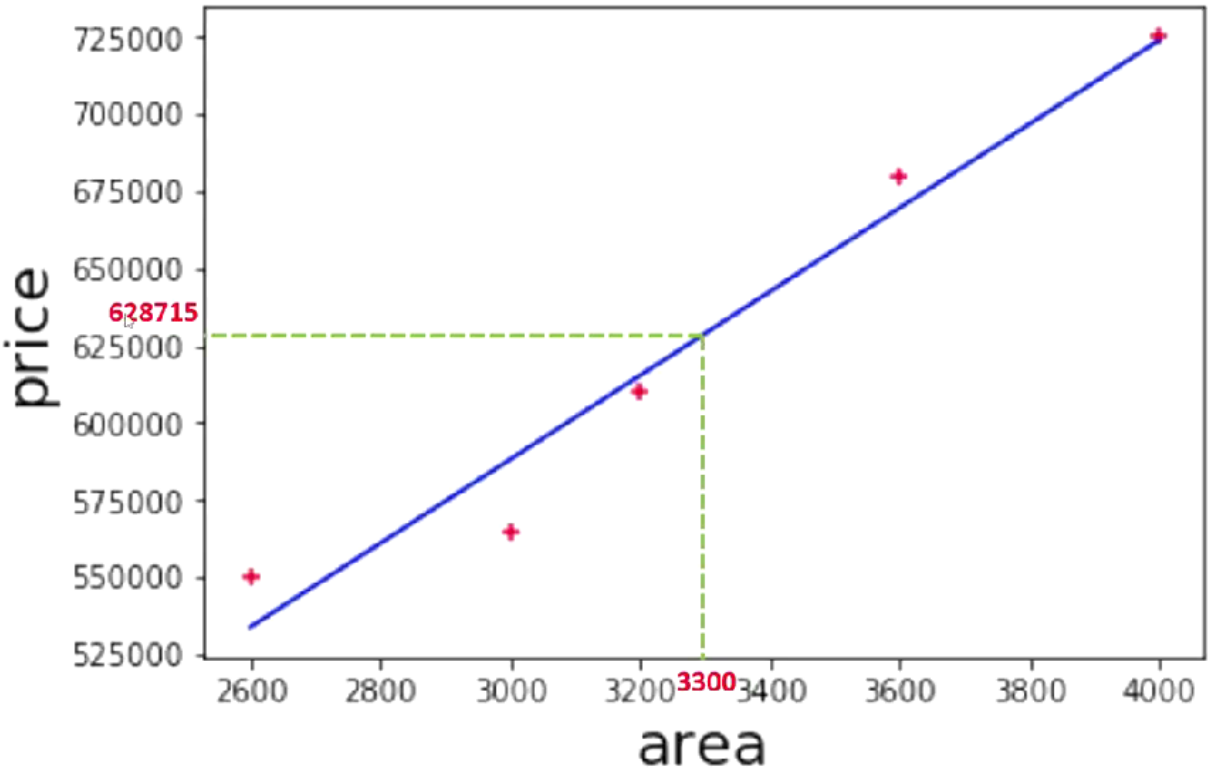
Outlines

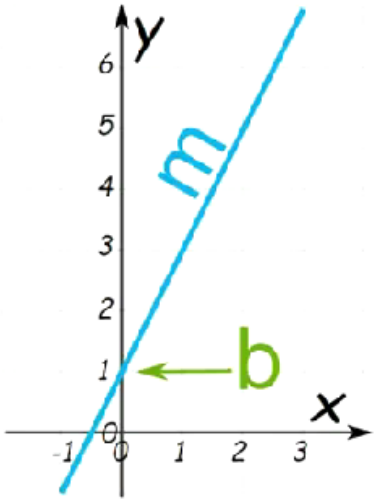
- Problem
- Write python code
- Import libraries
- Load data from csv file
- Plot scatter
- Linear Regression Model
- Prediction
- Calculate slope (m) & Intercept (b)
- Quiz
- How does linear equation look?
- Excercise 1
- Excercise 2

Problem

Based on the table shown below, what is the price of homes whose area are **3300** square feet & **5000** square feet?

area	price
2600	550000
3000	565000
3200	610000
3600	680000
4000	725000



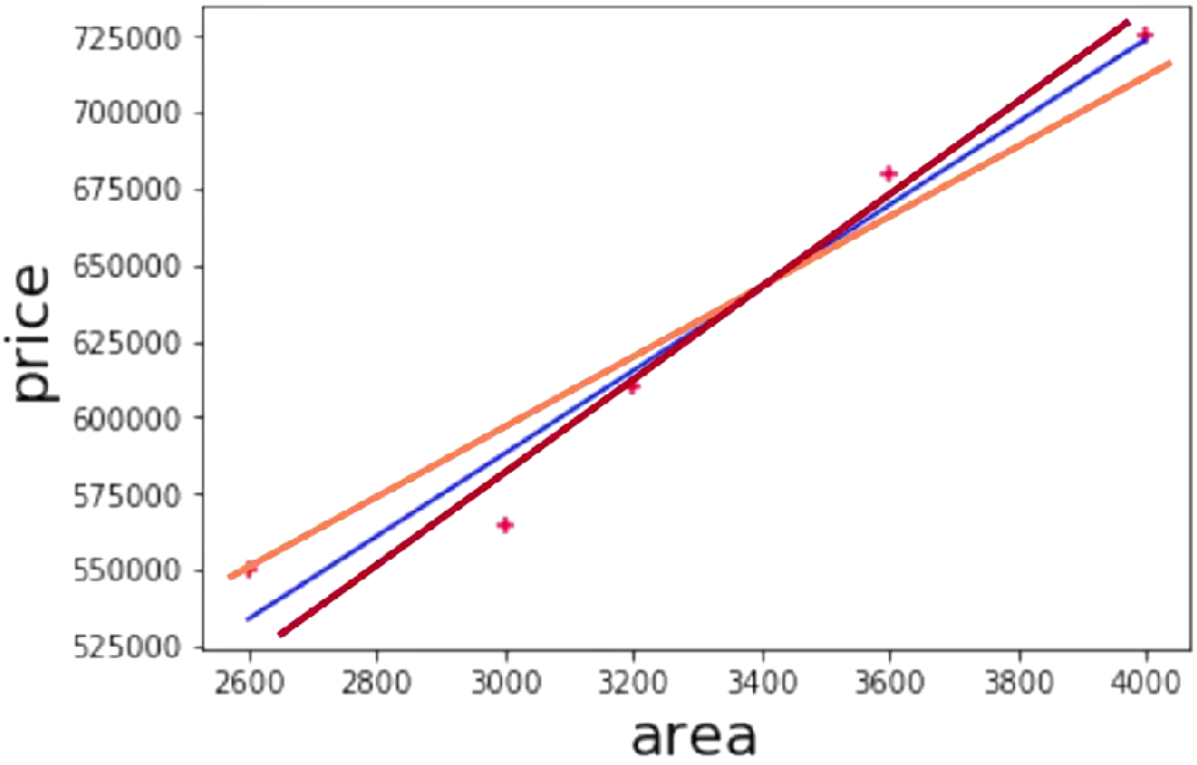


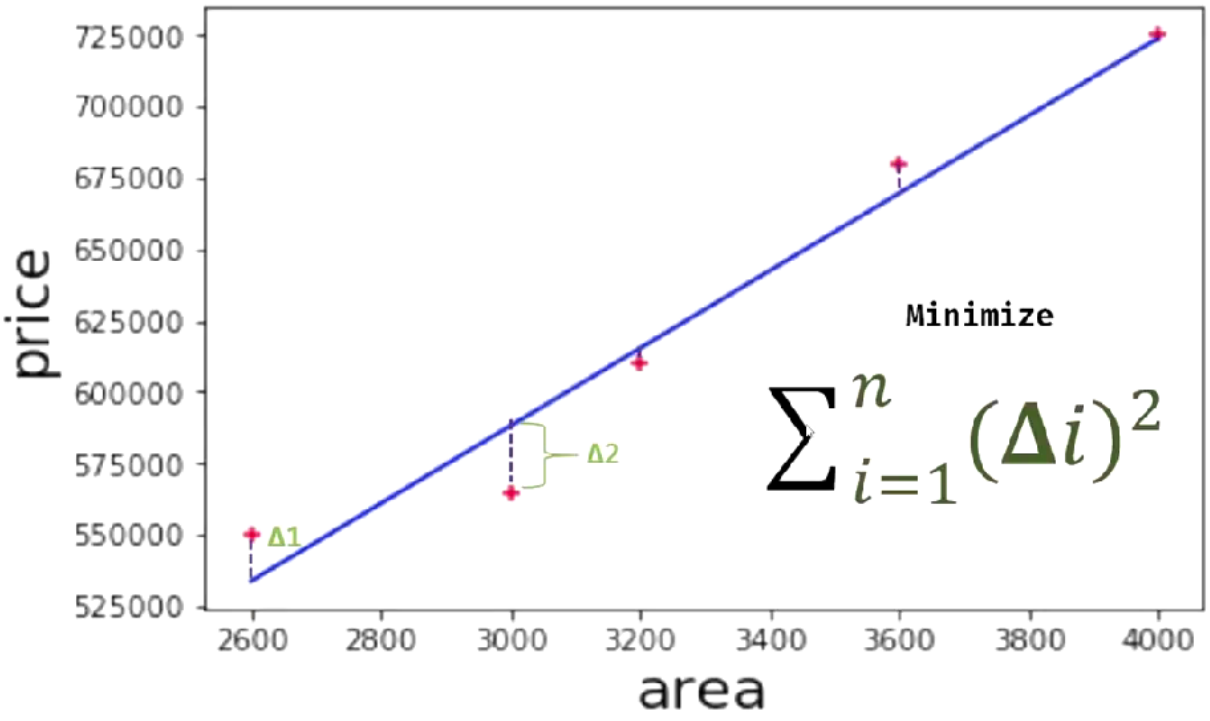
price = m * area + b

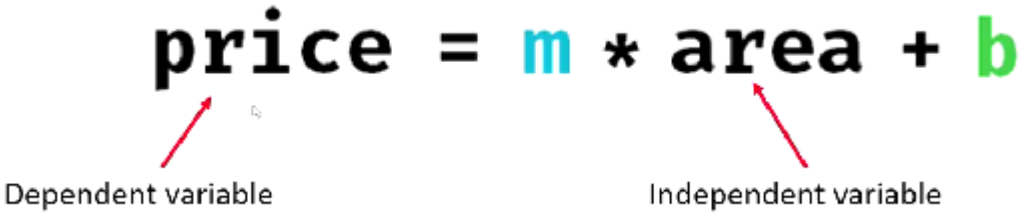
$$y = mx + b$$

Slope (or Gradient) Y Intercept

Reference: https://www.mathsisfun.com/algebra/linear-equations.html


price
area
725000
700000
675000
650000
625000
600000
575000
550000
525000
2600
2800
3000
3200
3400
3600
3800
4000


price
area
Δ1
Δ2
Minimize
$\sum_{i=1}^{n}(\Delta i)^2$



[Tip]: To execute the Python code in the code cell below, click on the cell to select it and press `Shift` + `Enter`.

Write Python Code

import libraries

In [45]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import linear_model
```

load data from csv file

In [6]:
```
# load data from csv file
# write the address of your folder where you save csv file
df = pd.read_csv('D:/Data_Science/My Github/Machine-Learning-with-Python/Linear_R
df
```

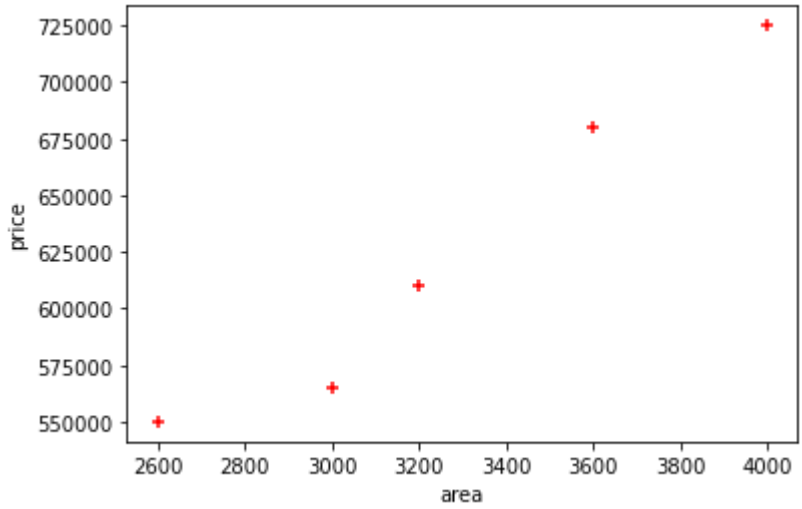
Out[6]:

	area	price
0	2600	550000
1	3000	565000
2	3200	610000
3	3600	680000
4	4000	725000

Plot scatter

In [12]:
```
%matplotlib inline
plt.xlabel('area')
plt.ylabel('price')
plt.scatter(df.area,df.price,color='red',marker='+')
```

Out[12]:
```
<matplotlib.collections.PathCollection at 0xb1f61c0>
```



Linear Regression Model

In [13]:
```
reg = linear_model.LinearRegression()
reg.fit(df[['area']],df.price)
```

Out[13]:
```
LinearRegression()
```

Prediction

In [16]:
```
reg.predict([[3300]])
```

Out[16]:
```
array([628715.75342466])
```

Calculate Slope (m) & Intercept (b)

In [19]:
```
print('m= %s , b= %s ' % (reg.coef_,reg.intercept_))
```

```
m= [135.78767123] , b= 180616.43835616432
```

In [23]:
```
m = reg.coef_
b = reg.intercept_
x=3300
y=m*x+b
y
```

Out[23]:
```
array([628715.75342466])
```

Quiz: Predict the price of 5000 feet square?

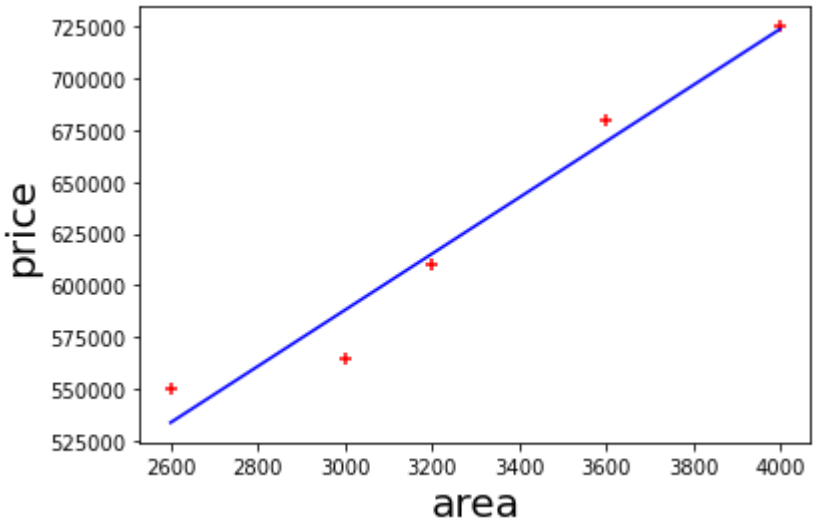
In [ ]:
```
# Write your code
```

Click here for the solution

How does Linear Equation Look?

In [33]:
```
%matplotlib inline
plt.xlabel('area', fontsize=20)
plt.ylabel('price', fontsize=20)
plt.scatter(df.area,df.price,color='red',marker='+')
plt.plot(df.area,reg.predict(df[['area']]),color='blue')
```

Out[33]:
```
[<matplotlib.lines.Line2D at 0xc82f0d0>]
```



Excercise 1:

A list of area is given in a csv file:

1. Load the csv file excercise.csv from your directory.
2. Predict the prices based on Linear Regression model.
3. Assign predictions to your data and print your new data.
4. write your new dataframe to a new csv file with a name prediction.csv.
5. Omit index from new data and write it in a new csv file named prediction1.csv

Click here for the solution part 1

Click here for the solution part 2

Click here for the solution part 3

Click here for the solution part 4

Click here for the solution part 5

Excercise 2:

A Canada's income is given in the folder from year 1970-2016. Your job is to predict the Canada's income in 2021?

Click here for the solution

Author

Ehsan Zia