In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import linear_model
```

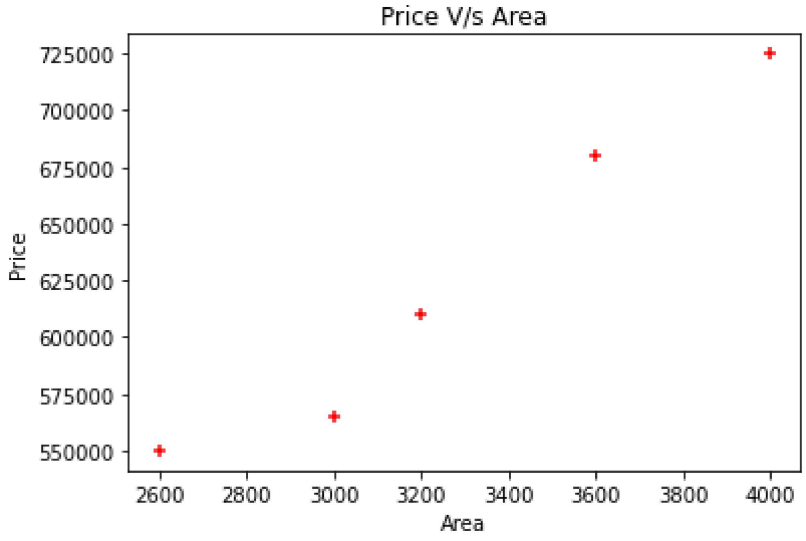
In [2]:
```
df = pd.read_csv('housing_data.csv')        # DataFrame creation
```

In [3]:
```
df
```

Out[3]:

	Area	Price
0	2600	550000
1	3000	565000
2	3200	610000
3	3600	680000
4	4000	725000

In [4]:
```
plt.scatter(df['Area'], df['Price'], color = 'red', marker = '+')
plt.title('Price V/s Area')
plt.xlabel('Area')
plt.ylabel('Price')
plt.show()
```



In [5]:
```
lin_reg = linear_model.LinearRegression()      #Linear Regression object
```

In [6]:
```
lin_reg.fit(df[['Area']], df.Price)            # First argument should be a 2D array
```

Out[6]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

In [7]:
```
lin_reg.predict([[3300]])                    # Price prediction for Area valu
e of 3300 sq units
```

Out[7]:
```
array([628715.75342466])
```

In [8]:
```
lin_reg.coef_                # Slope value 'm'
```

Out[8]:
```
array([135.78767123])
```

In [9]:
```
lin_reg.intercept_           # Intercept value 'b'
```

Out[9]:
```
180616.43835616432
```

In [10]:
```
# y  = mx + c
y = (135.78767123*3300) + 180616.43835616432
```

In [11]:
```
y                                # Same as predicted by our model
```

Out[11]:
```
628715.7534151643
```

In [12]:
```
df_2 = pd.read_csv('to_be_predicted.csv')
```

In [13]:
```
df_2
```

Out[13]:

	Area	Predicted Prices
0	1000	316404.109589
1	1500	384297.945205
2	2000	452191.780822
3	2500	520085.616438
4	2700	547243.150685
5	3000	587979.452055
6	3500	655873.287671
7	4000	723767.123288
8	4300	764503.424658
9	5100	873133.561644

In [14]:
```
predicted_prices = lin_reg.predict(df_2[['Area']])
```

In [15]:
```
predicted_prices
```

Out[15]:
```
array([316404.10958904, 384297.94520548, 452191.78082192, 520085.61643836,
       547243.15068493, 587979.45205479, 655873.28767123, 723767.12328767,
       764503.42465753, 873133.56164384])
```

In [16]:
```
# Assigning new column of predicted priced to df_2
df_2['Predicted Prices'] = predicted_prices
```

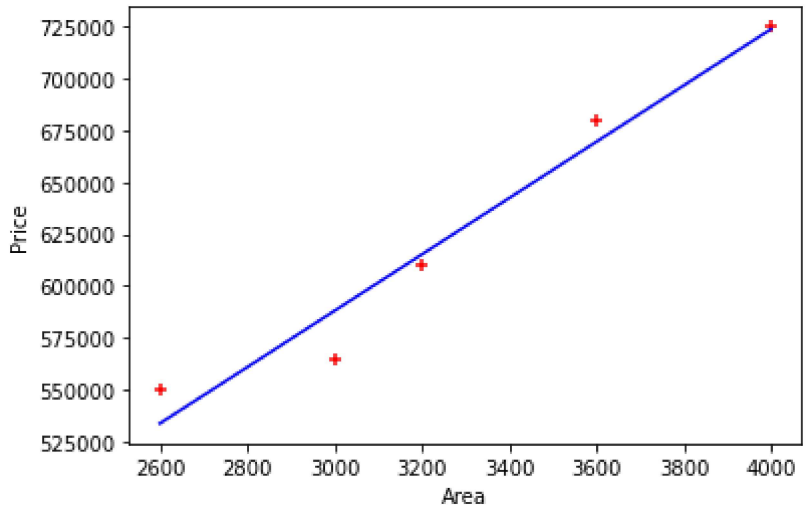
In [17]:
```
df_2
```

Out[17]:

	Area	Predicted Prices
0	1000	316404.109589
1	1500	384297.945205
2	2000	452191.780822
3	2500	520085.616438
4	2700	547243.150685
5	3000	587979.452055
6	3500	655873.287671
7	4000	723767.123288
8	4300	764503.424658
9	5100	873133.561644

In [18]:
```
df_2.to_csv('to_be_predicted.csv', index = False)
```

In [19]:
```
plt.scatter(df['Area'], df['Price'], color = 'red', marker = '+')
plt.plot(df['Area'],lin_reg.predict(df[['Area']]), color = 'blue')
plt.xlabel('Area')
plt.ylabel('Price')
plt.show()
```



In [ ]:
```

```