In [1]:
```
#importing Importing Libraries
import pandas as pd
import numpy as np
import seaborn as sns
import mpl_toolkits

import matplotlib.pyplot as plt
%matplotlib inline
import warnings
warnings.filterwarnings("ignore")
from sklearn import linear_model
from sklearn import metrics
```

In [9]:
```
# Data
df = pd.read_csv("Dataset.csv")
```

In [10]:
```
df
```

Out[10]:

	Sqft	propertyprice
0	2500	800000
1	3500	1000000
2	3800	1500000
3	4500	2000000
4	5000	2500000

In [11]:
```
df.describe()
```

Out[11]:

	Sqft	propertyprice
count	5.00000	5.000000e+00
mean	3860.00000	1.560000e+06
std	960.72889	7.021396e+05
min	2500.00000	8.000000e+05
25%	3500.00000	1.000000e+06
50%	3800.00000	1.500000e+06
75%	4500.00000	2.000000e+06
max	5000.00000	2.500000e+06

In [12]:
```
df.isna().sum()
```

Out[12]:
```
Sqft            0
propertyprice   0
dtype: int64
```

In [15]:
```
#Visualizing using heatmap
sns.heatmap(df.corr(),cmap="YlGnBu",annot = True)
plt.show()
```

In [16]:
```
#Scatterplot
plt.figure(figsize=(12,6))
plt.scatter(data['Sqft'],data['propertyprice'],color = 'green')
plt.title('SQFT VS PRICE')
plt.xlabel('SQFT')
plt.ylabel('Property_Price')
plt.grid(True)
plt.show()
```

In [17]:
```
n_df = df.drop('propertyprice', axis = 'columns' )
```

In [18]:
```
1 n_df
```

Out[18]:

	Sqft
0	2500
1	3500
2	3800
3	4500
4	5000

In [19]:
```
1 #Sales = Target
2 target = df['propertyprice']
```

In [20]:
```
1 target
```

Out[20]:
```
0     800000
1    1000000
2    1500000
3    2000000
4    2500000
Name: propertyprice, dtype: int64
```

In [21]:
```
1 #Creating linear regression object
2 lr = linear_model.LinearRegression()
```

In [22]:
```
1 lr.fit(n_df,target)
```

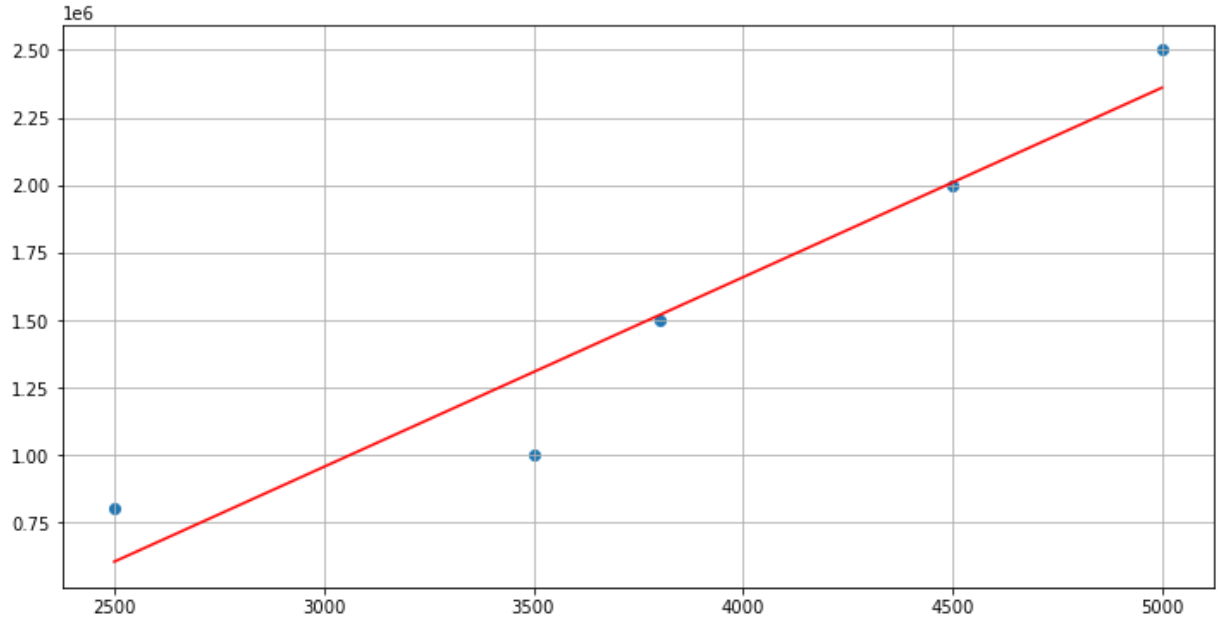
Out[22]:
```
LinearRegression()
```

In [23]:
```
1 #Predict price for given 3200 sqft value.
2 lr.predict([[3200]])
```

Out[23]:
```
array([1096641.38678223])
```

In [24]:
```
1 #Generate model prediction for given sqft area
2 y_predict = lr.predict(n_df)
```

In [33]:
```
# Visualize the predicte amount
plt.figure(figsize=(12,6))
plt.scatter(df['Sqft'],df['propertyprice'])
plt.plot(df['Sqft'],y_predict,'r')
plt.grid(True)
plt.show()
```



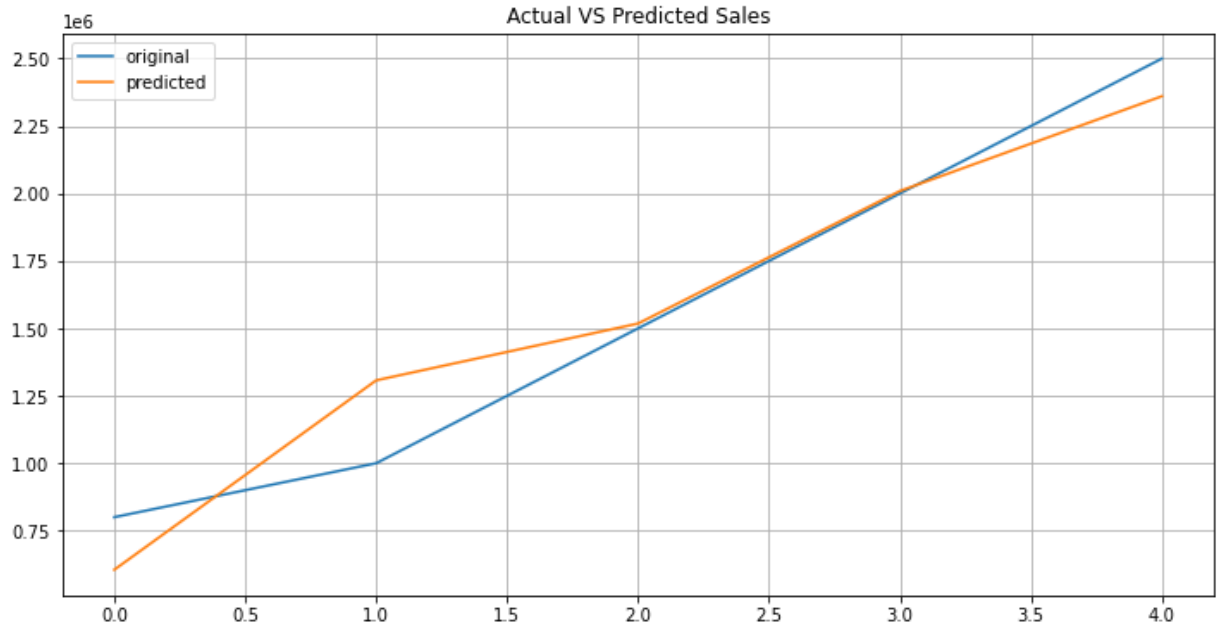
In [26]:
```
#Intercept value
print('Intercept:',lr.intercept_)
```

```
Intercept: -1149945.8288190686
```

In [28]:
```
#SLope Value
print('Slope:',lr.coef_)
```

```
Slope: [702.05850488]
```

In [34]:
```
plt.figure(figsize=(12,6))
x_ax = range(len(df['propertyprice']))
plt.plot(x_ax,df['propertyprice'],label ='original')
plt.plot(x_ax,y_predict,label='predicted')
plt.title("Actual VS Predicted Sales")
plt.legend()
plt.grid()
plt.show()
```



CONCLUSION:

1. We have clean data without any null values.

2. Property price increase gradually with increase in square foot value.

3. Based on trained Linear Model, if the house size is 3200 sqft then the house price will be approximate of 1096641.

In [ ]:
```
1
```