In [3]:
```
import pandas as pd
df=pd.read_csv("D:\Home_Price.csv")
df
```

Out[3]:

	own	area	price
0	monroe township	2600	550000
1	monroe township	3000	565000
2	monroe township	3200	610000
3	monroe township	3600	680000
4	monroe township	4000	725000
5	west windsor	2600	585000
6	west windsor	2800	615000
7	west windsor	3300	650000
8	west windsor	3600	710000
9	robinsville	2600	575000
10	robinsville	2900	600000
11	robinsville	3100	620000
12	robinsville	3600	695000

In [4]:
```
df.rename(columns={'own':'Town'},inplace=True)
df
```

Out[4]:

	Town	area	price
0	monroe township	2600	550000
1	monroe township	3000	565000
2	monroe township	3200	610000
3	monroe township	3600	680000
4	monroe township	4000	725000
5	west windsor	2600	585000
6	west windsor	2800	615000
7	west windsor	3300	650000
8	west windsor	3600	710000
9	robinsville	2600	575000
10	robinsville	2900	600000
11	robinsville	3100	620000
12	robinsville	3600	695000

2.Dummy variables

In [6]:
```
dummies=pd.get_dummies(df.Town)
dummies
```

Out[6]:

	monroe township	robinsville	west windsor
0	1	0	0
1	1	0	0
2	1	0	0
3	1	0	0
4	1	0	0
5	0	0	1
6	0	0	1
7	0	0	1
8	0	0	1
9	0	1	0
10	0	1	0
11	0	1	0
12	0	1	0

3.Now we Merging dummy variables

In [7]:
```
marged=pd.concat([df,dummies],axis='columns')
marged
```

Out[7]:

	Town	area	price	monroe township	robinsville	west windsor
0	monroe township	2600	550000	1	0	0
1	monroe township	3000	565000	1	0	0
2	monroe township	3200	610000	1	0	0
3	monroe township	3600	680000	1	0	0
4	monroe township	4000	725000	1	0	0
5	west windsor	2600	585000	0	0	1
6	west windsor	2800	615000	0	0	1
7	west windsor	3300	650000	0	0	1
8	west windsor	3600	710000	0	0	1
9	robinsville	2600	575000	0	1	0
10	robinsville	2900	600000	0	1	0
11	robinsville	3100	620000	0	1	0
12	robinsville	3600	695000	0	1	0

In [8]:
```
# 4. Dropping ‘town’ column
```

In [9]:
```
new_dataFrame=marged.drop(['Town'],axis='columns')
```

In [10]:
```
new_dataFrame
```

Out[10]:

	area	price	monroe township	robinsville	west windsor
0	2600	550000	1	0	0
1	3000	565000	1	0	0
2	3200	610000	1	0	0
3	3600	680000	1	0	0
4	4000	725000	1	0	0
5	2600	585000	0	0	1
6	2800	615000	0	0	1
7	3300	650000	0	0	1
8	3600	710000	0	0	1
9	2600	575000	0	1	0
10	2900	600000	0	1	0
11	3100	620000	0	1	0
12	3600	695000	0	1	0

In [11]:
```
# 5. Dropping one dummy variable
```

In [12]:
```
# # Avoiding dummy variable trap I am dropping Monoroe column here.

# Note: You can drop anyone dummy variable column you want.
```

In [14]:
```
df1=new_dataFrame.drop(['monroe township'],axis='columns')
df1
```

Out[14]:

	area	price	robinsville	west windsor
0	2600	550000	0	0
1	3000	565000	0	0
2	3200	610000	0	0
3	3600	680000	0	0
4	4000	725000	0	0
5	2600	585000	0	1
6	2800	615000	0	1
7	3300	650000	0	1
8	3600	710000	0	1
9	2600	575000	1	0
10	2900	600000	1	0
11	3100	620000	1	0
12	3600	695000	1	0

In [15]:
```
# Independent variable take on x-axis
```

In [16]:
```
x=df1.drop(['price'],axis='columns')
x
```

Out[16]:

	area	robinsville	west windsor
0	2600	0	0
1	3000	0	0
2	3200	0	0
3	3600	0	0
4	4000	0	0
5	2600	0	1
6	2800	0	1
7	3300	0	1
8	3600	0	1
9	2600	1	0
10	2900	1	0
11	3100	1	0
12	3600	1	0

In [17]:
```
#  Dependent variable take on y axis
```

In [18]:
```
y=df1.price
y
```

Out[18]:
```
0     550000
1     565000
2     610000
3     680000
4     725000
5     585000
6     615000
7     650000
8     710000
9     575000
10    600000
11    620000
12    695000
Name: price, dtype: int64
```

In [19]:
```
#  Importing Linear Regression model
```

In [20]:
```
from sklearn.linear_model import LinearRegression
model=LinearRegression()
```

In [21]:
```
# fitting the model
```

In [22]:
```
model.fit(x,y)
```

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

In [23]:
```
#  Predicting the house price
```

In [24]:
```
model.predict(x)
```

Out[24]:
```
array([539709.7398409 , 590468.71640507, 615848.20468716, 666607.18125134,
       717366.15781552, 579723.71533004, 605103.20361213, 668551.92431735,
       706621.15674048, 565396.1513653 , 603465.38378843, 628844.87207052,
       692293.59277575])
```

In [25]:
```
#  Checking the accuracy of model
```

In [26]:
```
model.score(x,y)
```

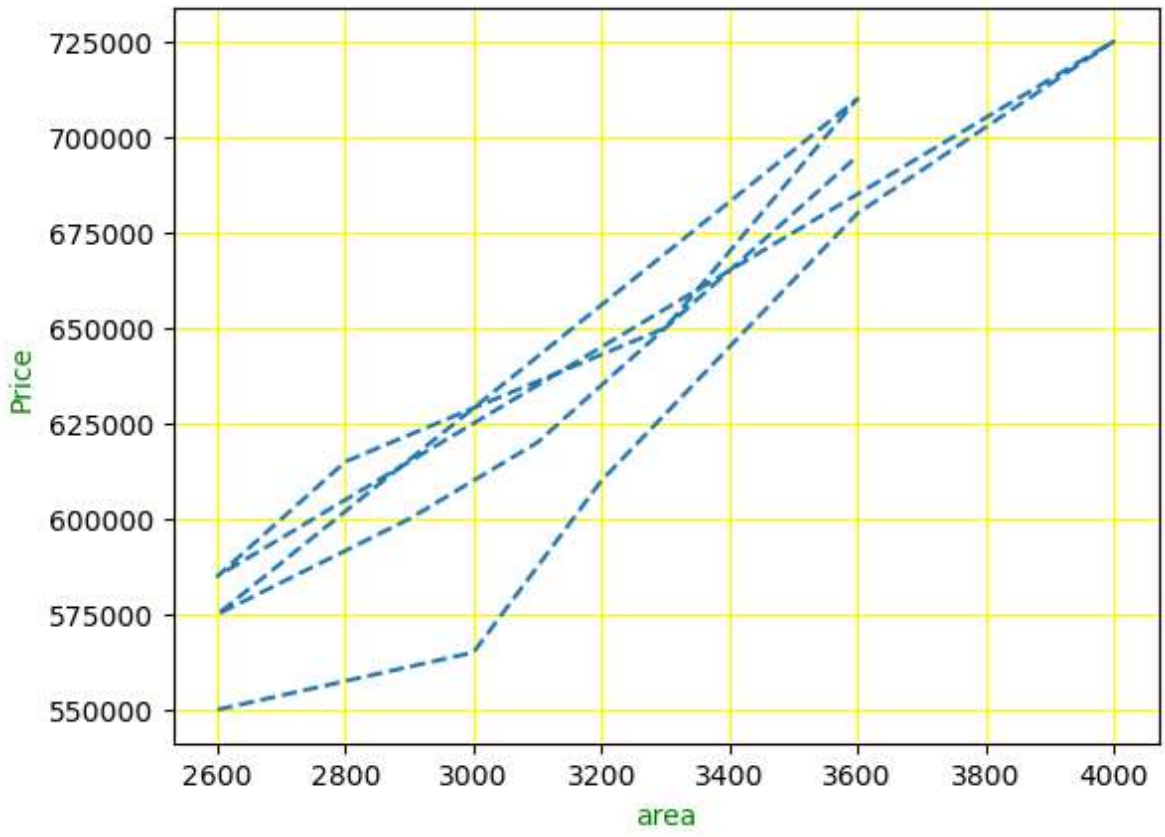
Out[26]:
```
0.9573929037221871
```

In [27]:
```
# Plot graph between area and price
```

In [53]:
```
area1=df1.drop(['area'],axis='columns')
x=df1.area
x
```

Out[53]:
```
0     2600
1     3000
2     3200
3     3600
4     4000
5     2600
6     2800
7     3300
8     3600
9     2600
10    2900
11    3100
12    3600
Name: area, dtype: int64
```

In [55]:
```
from matplotlib import pyplot as plt
plt.grid(color='yellow')
plt.ylabel('Price',color='Green')
plt.xlabel('area',color='green')
plt.plot(area,y,linestyle='--')
plt.show()
```



In [ ]: