DATS_6450_15

TIME SERIES MODELING & ANALYSIS

Instructor Name: Reza Jafri

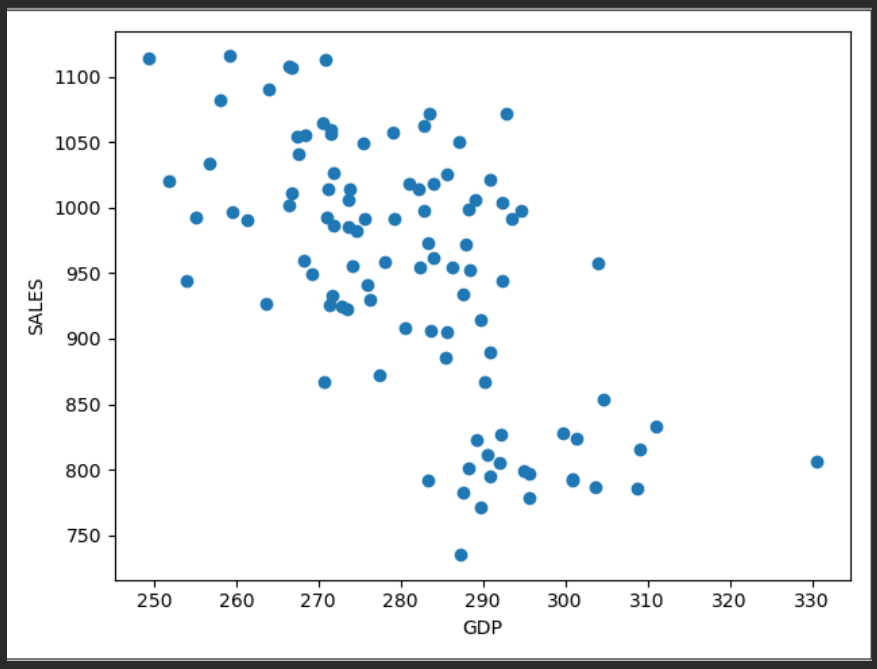
HW#: 2

Submitted by: Dinesh Kumar Padmanabhan

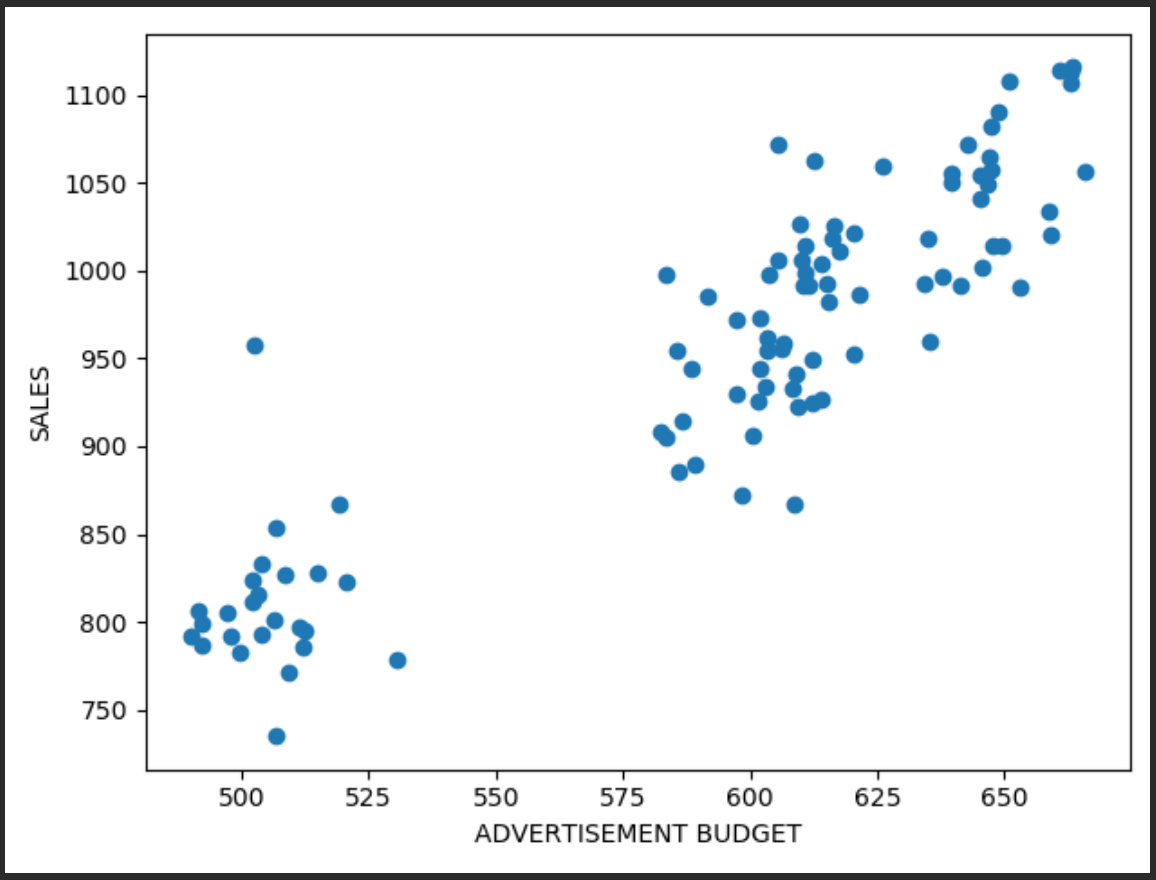
Date: 23-sept-2020

ANSWERS TO ASKED QUESTIONS

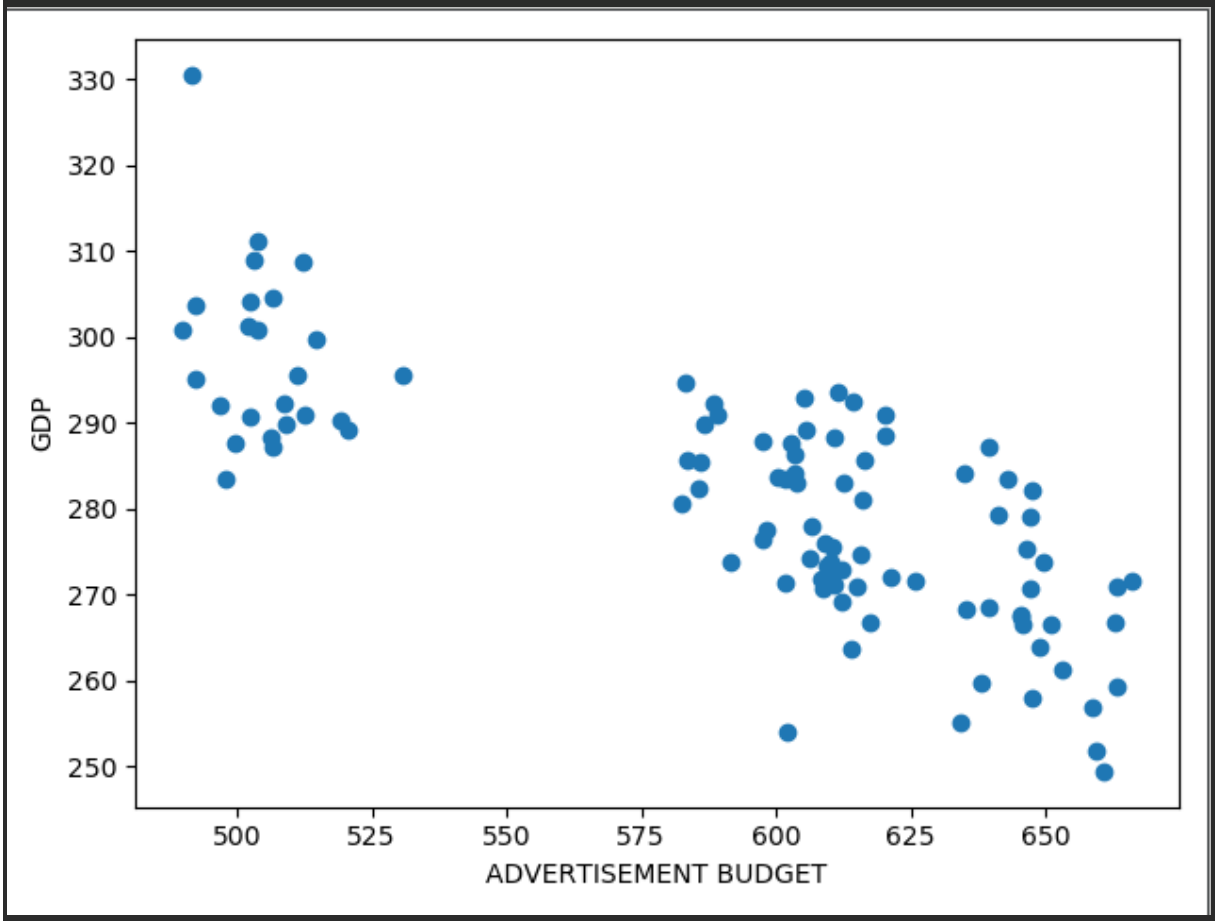
```
#Homework 2
# Using the Python program and “pandas”,  “matplotlib.pyplot” ,“numpy” and
“statsmodels” library
# perform the following tasks:

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import math


#%%=============================================================================
# 1: Load the time series data called tute1.
# The tute1 dataset is the same dataset used in LAB#1.
# Graph the scatter plot for Sales and GDP. (y-axis plot Sales and x-axis plot
GDP).
# Add the appropriate x-label and y-label.
# Do not add any title in this step. This needs to be updated in step 7.
# %%------------------------------------------------------------------------

df = pd.read_csv("tute1.csv")
plt.xlabel("GDP")
plt.ylabel("SALES")
plt.scatter(df["GDP"],df["Sales"])
plt.show()
```



```
#%%=============================================================================
# 2: Graph the scatter plot for Sales and AdBudget. (y-axis plot Sales and x-axis
plot AdBudget).
# Add the appropriate x-label and y-label.
# Do not add any title in this step. This needs to be updated in step 7.
# %%-----------------------------------------------------------------------------

plt.xlabel("ADVERTISEMENT BUDGET")
plt.ylabel("SALES")
plt.scatter(df["AdBudget"],df["Sales"])
plt.show()
```



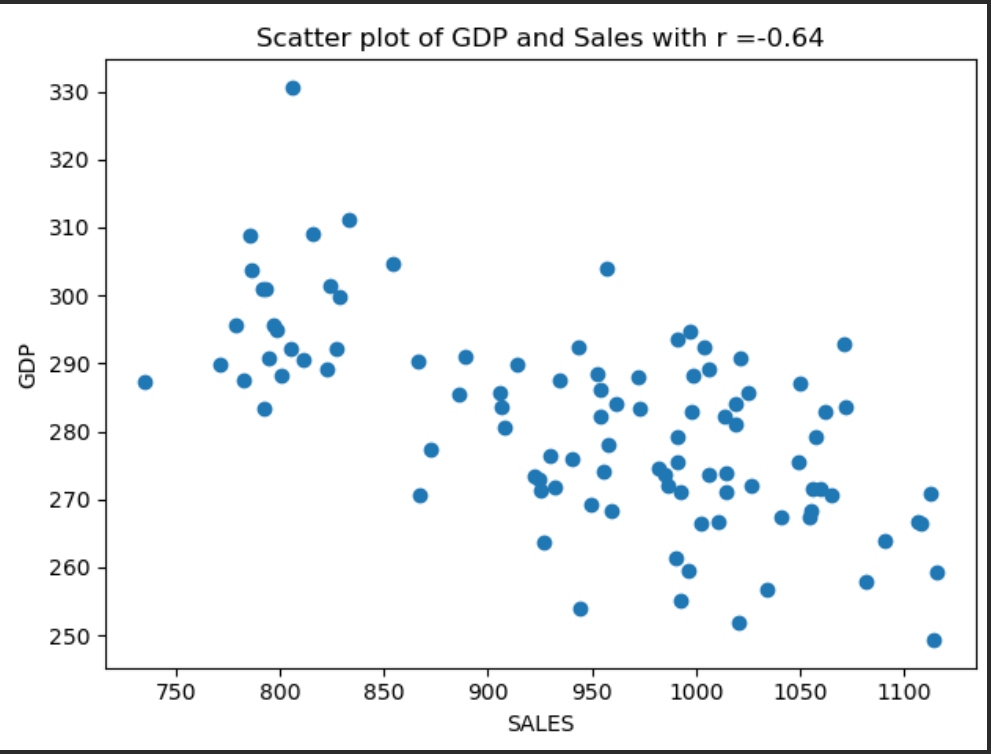
```
#%%=============================================================================
# 3: Graph the scatter plot for GDP and AdBudget. (y-axis plot GDP and x-axis plot
AdBudget).
# Add the appropriate x-label and y-label.
# Do not add any title in this step. This needs to be updated in step 7.
# %%-----------------------------------------------------------------------------

plt.xlabel("ADVERTISEMENT BUDGET")
plt.ylabel("GDP")
plt.scatter(df["AdBudget"],df["GDP"])
plt.show()
```



```
#%%=============================================================================
# 4: Call the function “correlation_coefficent_cal(x,y)” developed in the LAB#2
with y as the Sales data and
# the x as the GDP data.
# Save the correlation coefficient between these two variables as r_xy.
# Display the following message on the console:
# “The correlation coefficient between the Sales value and GDP is _________”.
# Does the r_xy value make sense with respect to the scatter plot graphed in
step 7.
# Explain why?
# %%-----------------------------------------------------------------------------

def correlation_Coefficient_cal(X, Y):
    n = len(X)
    sum_X = 0
    sum_Y = 0
    sum_XY = 0
    squareSum_X = 0
    squareSum_Y = 0
    i = 0
    while i < n:
        sum_X = sum_X + X[i]
        sum_Y = sum_Y + Y[i]
        sum_XY = sum_XY + X[i] * Y[i]

        squareSum_X = squareSum_X + X[i] * X[i]
        squareSum_Y = squareSum_Y + Y[i] * Y[i]

        i = i + 1

    r = (float)(n * sum_XY - sum_X * sum_Y) / \
        (float)(math.sqrt((n * squareSum_X -sum_X * sum_X) * (n * squareSum_Y -
sum_Y * sum_Y)))
    return r
# import LAB2
# LAB2.correlation_Coefficient_cal()
y = df["Sales"]
x = df["GDP"]
r_xy = (correlation_Coefficient_cal(x, y))
print('The correlation coefficient between the Sales value and GDP
is:{0:.2f}'.format(r_xy))
print('\n')

#OUTPUT
The correlation coefficient between the Sales value and GDP is:-0.64

'''
Yes it does make sense.
Negative values indicating a negative relationship between the sales value and
GDP.
'''
```

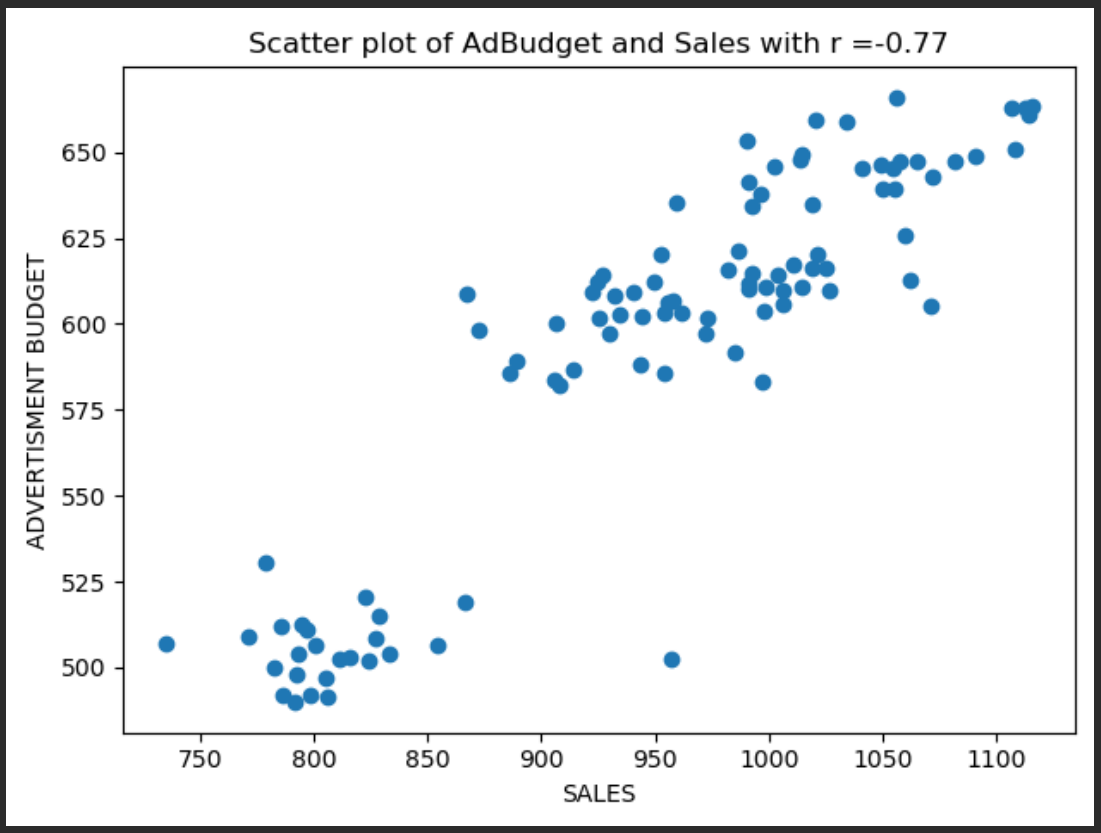
```
#%%=============================================================================
# 5: Call the function “correlation_coefficent_cal(x,z)” developed in LAB#2 with x
as the Sales data
# and the z as the AdBudget data.
# Save the correlation coefficient between these two variables as r_yz.
# Display the following message on the console:
# “The correlation coefficient between the Sales value and AdBudget is _________”.
# Does the r_yz value make sense with respect to the scatter plot graphed in
step 8.
# Explain why?
# %%------------------------------------------------------------------------

x = df["Sales"]
z = df["AdBudget"]
r_xz = (correlation_Coefficient_cal(x, z))
print('The correlation coefficient between the Sales value and AdBudget
is:{0:.2f}'.format(r_xz))
print('\n')

#OUTPUT
The correlation coefficient between the Sales value and AdBudget is:0.91

'''
No it does not make sense.
The correlation coefficient between the Sales value and AdBudget indicates
Negative values but the plot indicates the positive relationship.
'''

#%%=============================================================================
# 6: Call the function “correlation_coefficent_cal(y,z)” developed in LAB#2 with y
as the GDP data and the z as the AdBudget data.
# Save the correlation coefficient between these two variables as r_yz. Display
the following message on the console:
# “The correlation coefficient between the GDP value and AdBudget is _________”.
# Does the r_yz value make sense with respect to the scatter plot graphed in
step 8.
# Explain why?
# %%------------------------------------------------------------------------
y = df["GDP"]
z = df["AdBudget"]
r_yz = (correlation_Coefficient_cal(y, z))
print('The correlation coefficient between GDP and AdBudget
is:{0:.2f}'.format(r_yz))
print('\n')

#OUTPUT
The correlation coefficient between GDP and AdBudget is:-0.77

'''
No it does not make sense.
The correlation coefficient between the GDP value and AdBudget indicates positive
values but the plot indicates the negative relationship.
'''
```

```
#%%=========================================================================================
# 7:Include the r_xy, r_yz and r_xz in the title of the graphs developed in step 5
and 6.
# Write your code in a way that anytime r_xy, r_yz and r_xz value changes it
automatically updated on the figure title.
# Hint: you can use the following python command:
# plt.title("Scatter plot of GDP and Sales with r ={}".format(r_xy))
# %%-----------------------------------------------------------------------------------
print("Scatter plot of GDP and Sales")
plt.xlabel("SALES")
plt.ylabel("GDP")
plt.scatter(df["Sales"],df["GDP"])
plt.title("Scatter plot of GDP and Sales with r ={0:.2f}".format(r_xy))
plt.show()
```



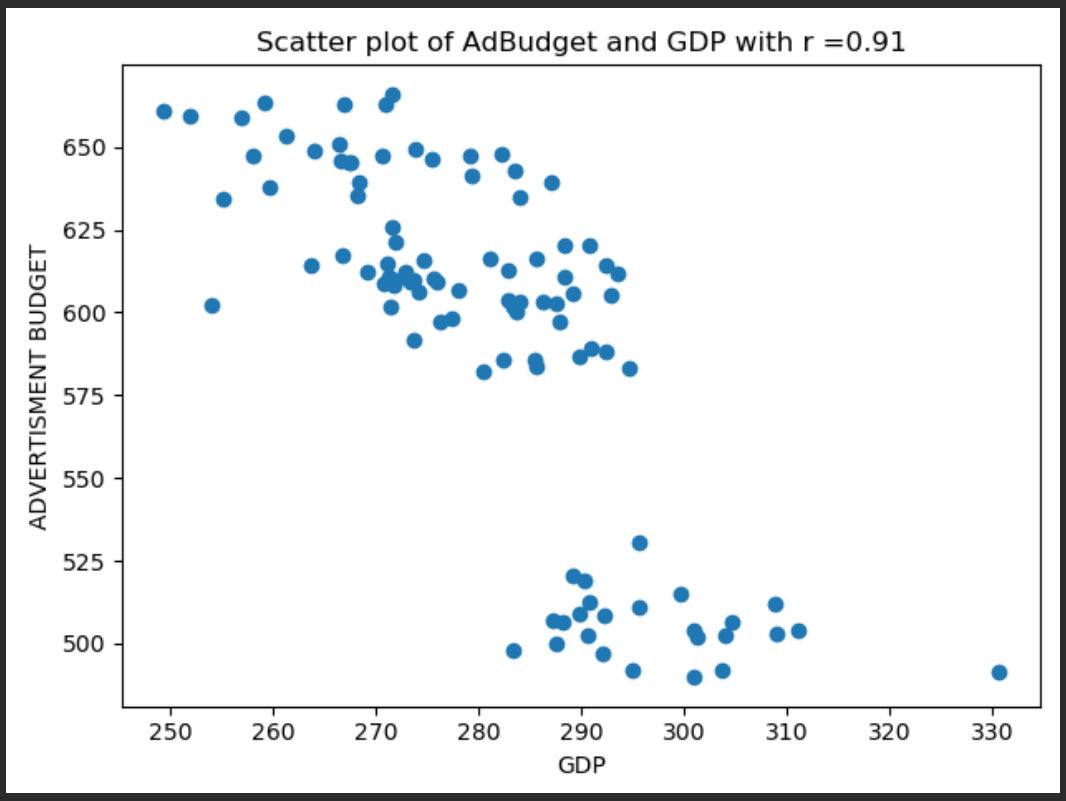

```
print('\n')
```

```
print("Scatter Plot of AdBudget and Sales")
plt.xlabel("SALES")
plt.ylabel("ADVERTISMENT BUDGET")
plt.scatter(df["Sales"],df["AdBudget"])
plt.title("Scatter plot of AdBudget and Sales with r ={0:.2f}".format(r_yz))
plt.show()
```




```
print('\n')
```

```
print("Scatter Plot of AdBudget and Sales")
plt.xlabel("GDP")
plt.ylabel("ADVERTISMENT BUDGET")
plt.scatter(df["GDP"],df["AdBudget"])
plt.title("Scatter plot of AdBudget and GDP with r ={0:.2f}".format(r_xz))
plt.show()
```




```
#%%=============================================================================
# 8:By looking at the correlation coefficients,
# write down your observation about the effect of AdBudget data and GDP data on
the Sales revenue?
# %%-----------------------------------------------------------------------------
'''
The correlation coefficient between the GDP  and AdBudget on sales is -0.64 and -
0.77 respectively which is
negatively correlated and the plot does make sense for the GDP and sales however
not for Adbudget and sales because
the r value is negative and the projection is positive.
'''
```

File - unknown

```
C:\ProgramData\Anaconda3\python.exe "C:\Program Files\JetBrains\PyCharm 2019.3.1\plugins\python\helpers\pydev\pydevconsole.py" --mode=client --port=59047

import sys; print('Python %s on %s' % (sys.version, sys.platform))
sys.path.extend(['C:\\Users\\nsree_000\\Desktop\\Python-Quiz', 'C:/Users/nsree_000/Desktop/Python-Quiz'])

Python 3.7.4 (default, Aug  9 2019, 18:34:13) [MSC v.1915 64 bit (AMD64)]
Type 'copyright', 'credits' or 'license' for more information
IPython 7.8.0 -- An enhanced Interactive Python. Type '?' for help.
PyDev console: using IPython 7.8.0

Python 3.7.4 (default, Aug  9 2019, 18:34:13) [MSC v.1915 64 bit (AMD64)] on win32
In[2]: runfile('C:/Users/nsree_000/Desktop/Python-Quiz/TIME SERIES/HW2.py', wdir='C:/Users/nsree_000/Desktop/Python-Quiz/TIME SERIES')
The correlation coefficient between the Sales value and GDP is:-0.64


The correlation coefficient between the Sales value and AdBudget is:0.91


The correlation coefficient between GDP and AdBudget is:-0.77


Scatter plot of GDP and Sales


Scatter Plot of AdBudget and Sales


Scatter Plot of AdBudget and Sales

```