
**GRT INSTITUTE OF ENGINEERING AND TECHNOLOGY,** Tiruttani


# 1103-GRT INSTITUTE OF ENGINEERING AND TECHNOLOGY

## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

## PROJECT TITLE

***Future sales prediction***

Aswini.M

3rd yr, 5th sem

Reg no:110321104004

aswini30012004@gmail.com

# PHASE4: FUTURE SALES PREDICTION

Future sales prediction, also known as sales forecasting, is the process of estimating a company's future sales based on historical data, market trends, and various analytical techniques. Accurate sales predictions are crucial for businesses as they help in making informed decisions regarding inventory management, resource allocation, budget planning, and overall strategy.

**IMPLEMENTATION:**

- Load and preprocess your sales data.
- Split the data into features and the target variable.
- Split the data into training and testing sets for model evaluation.
- Create a linear regression model and train it on the training data.
- Use the model to make predictions on the test set and evaluate its performance.
- Visualize the results if needed.
- Finally, you can use the trained model to make predictions for future sales based on your future feature values.

In addition to these steps, it's essential to keep the business goals in mind when implementing the future sales prediction. Each business is unique, and the prediction strategy should align with the specific objectives.

# DATASET AND ITS IMPLEMENTATION

We've got the data set in the website called Kaggle(www.kaggle.com/data)

The data set which is respective to our project is sales.csv

(**https://www.kaggle.com/datasets/chakradharmattapalli/future-sales-prediction** )

The data set having the 4columns named TV,Radio,Newspaper,Sales and having 200 rows of datas.(numerical values).

# BEGIN THE PROJECT BY LOADING THE DATASET

THE SAMPLE DATA SET:

| TV | Radio | Newspaper | Sales |
|---|---|---|---|
| 203.1 | 37.8 | 69.2 | 22.1 |
| 44.5 | 39.3 | 45.1 | 10.4 |
| 17.2 | 45.9 | 69.3 | 12 |
| 151.5 | 41.3 | 58.5 | 16.5 |
| 180.8 | 10.8 | 58.4 | 17.9 |

## 1.IMPORT NECESSARY LIBRARIES

Here are the key libraries you may need and how to download them: make sure to activate your virtual environment first and then run the pip install commands within that environment.

- Numpy
- Pandas
- Matplotlib and seaborn

PROGRAM

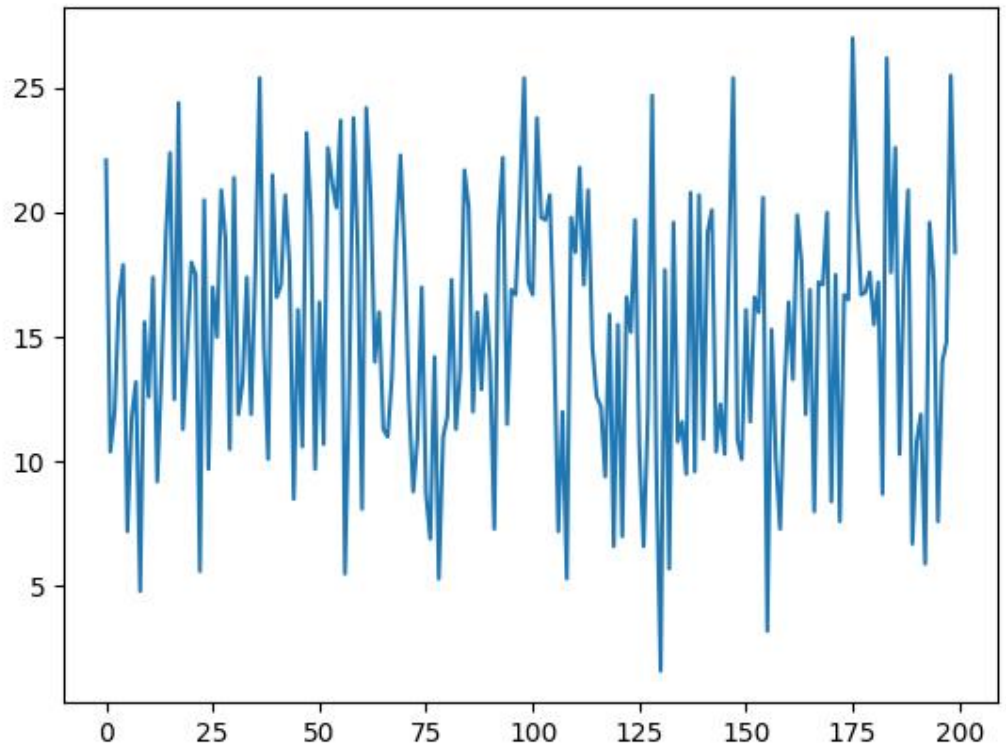
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

#load the dataset
data = pd.read_csv('C:/priya/Sales.csv')
print(data.head())
data['Sales'].plot()
plt.show()
data.plot(subplots=True, figsize=(4, 4))
plt.show()
print(data.describe())
print(data.isnull().sum())

# Create a scatter plot for tv vs sales
plt.figure(figsize=(8, 6))
plt.scatter(data['TV'], data['Sales'], c='b', marker='o', label='TV vs.
Sales')
plt.title('Scatter Plot of TV Advertising vs. Sales')
plt.xlabel('TV Advertising Budget')
plt.ylabel('Sales')
plt.legend()
plt.grid(True)
plt.show()

# Create a scatter plot for radio vs sales
plt.figure(figsize=(8, 6))
plt.scatter(data['Radio'], data['Sales'], c='r', marker='o', label='Radio
vs. Sales')
plt.title('Scatter Plot of Radio Advertising vs. Sales')
plt.xlabel('Radio Advertising Budget')
plt.ylabel('Sales')
plt.legend()
plt.grid(True)
plt.show()
```

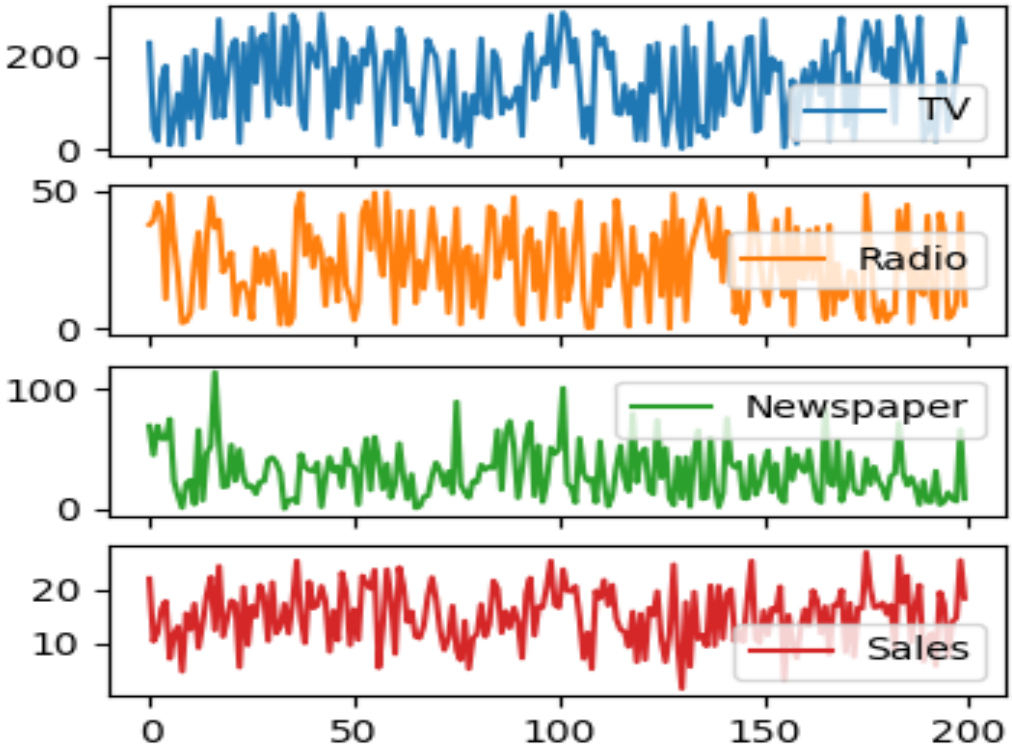
```
# Create a scatter plot for newspaper vs sales
plt.figure(figsize=(8, 6))
plt.scatter(data['Newspaper'], data['Sales'], c='g', marker='o',
label='Newspaper vs. Sales')
plt.title('Scatter Plot of Newspaper Advertising vs. Sales')
plt.xlabel('Newspaper Advertising Budget')
plt.ylabel('Sales')
plt.legend()
plt.grid(True)
plt.show()

#regression
# Sample data
TV = [230.1, 44.5, 17.2, 151.5, 180.8]
Radio = [17.8, 39.3, 45.9, 41.3, 10.8]
Newspaper = [69.2, 45.1, 69.3, 58.5, 58.4]
Sales = [22.1, 10.4, 12, 16.5, 17.9]

# Calculate the mean of each feature and the target
mean_TV = sum(TV) / len(TV)
mean_Radio = sum(Radio) / len(Radio)
mean_Newspaper = sum(Newspaper) / len(Newspaper)
mean_Sales = sum(Sales) / len(Sales)

# Calculate the coefficients
numerator = 0
denominator = 0
for i in range(len(TV)):
    numerator += (TV[i] - mean_TV) * (Sales[i] - mean_Sales)
    denominator += (TV[i] - mean_TV) ** 2

slope = numerator / denominator
intercept = mean_Sales - slope * mean_TV
# Now, you can make predictions
new_TV = 250  # Input a new value for TV
predicted_sales = intercept + slope * new_TV
```

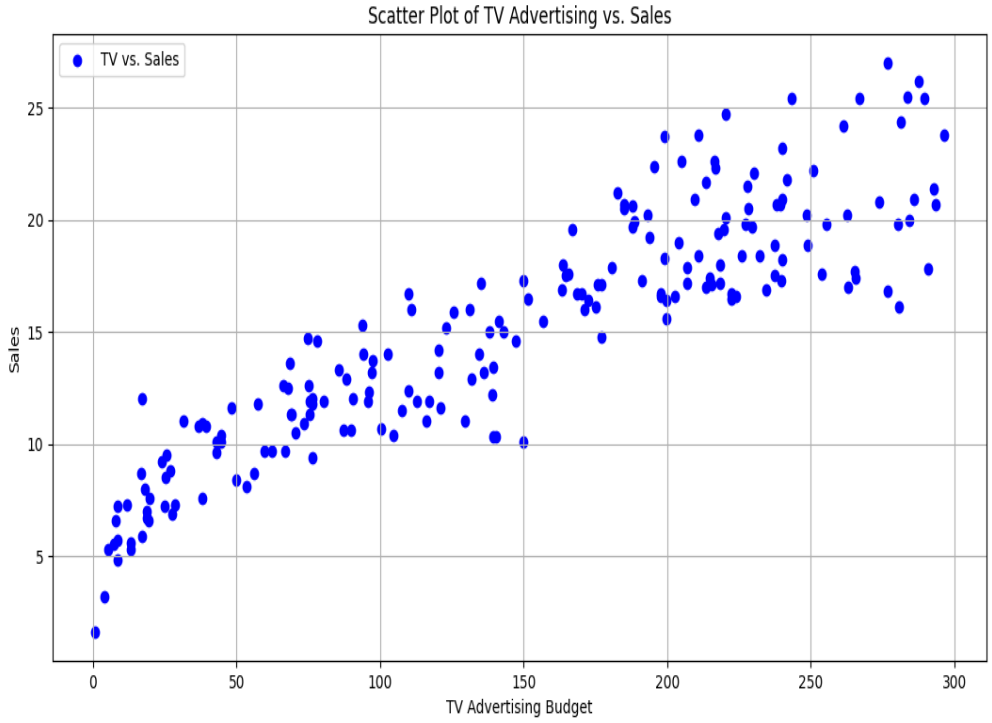
```
print(f"Slope (Coefficient): {slope}")
print(f"Intercept: {intercept}")
print(f"Predicted Sales for TV = {new_TV}: {predicted_sales}")

# Create an area plot
data[['TV', 'Radio', 'Newspaper', 'Sales']].plot.area(stacked=True)

# Add labels and a title
plt.xlabel('Data Points')
plt.ylabel('Values')
plt.title('Area Plot for TV, Radio, Newspaper, and Sales')
# Show the plot
plt.show()

# Extract the data for the columns you want to plot
tv_data = data['TV']
radio_data = data['Radio']
newspaper_data = data['Newspaper']
sales_data = data['Sales']

# Create a bar chart
plt.bar(['TV', 'Radio', 'Newspaper', 'Sales'], [tv_data.mean(),
radio_data.mean(), newspaper_data.mean(), sales_data.mean()])
plt.xlabel('Advertising Medium')
plt.ylabel('Mean Value')
plt.title('Mean Values for TV, Radio, Newspaper, and Sales')
plt.show()
```

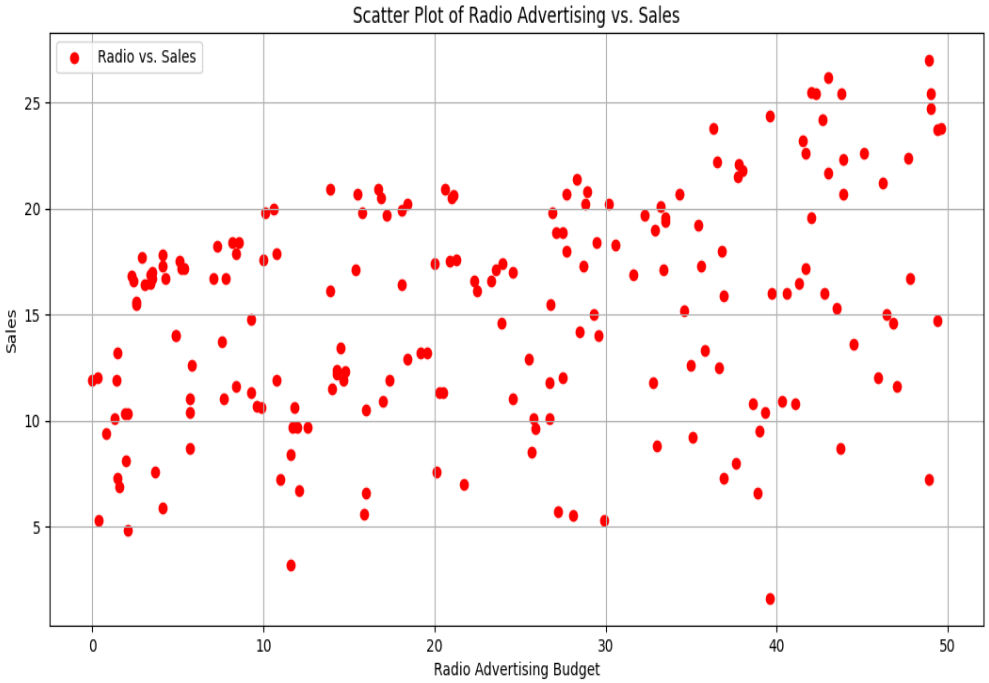
# OUTPUT

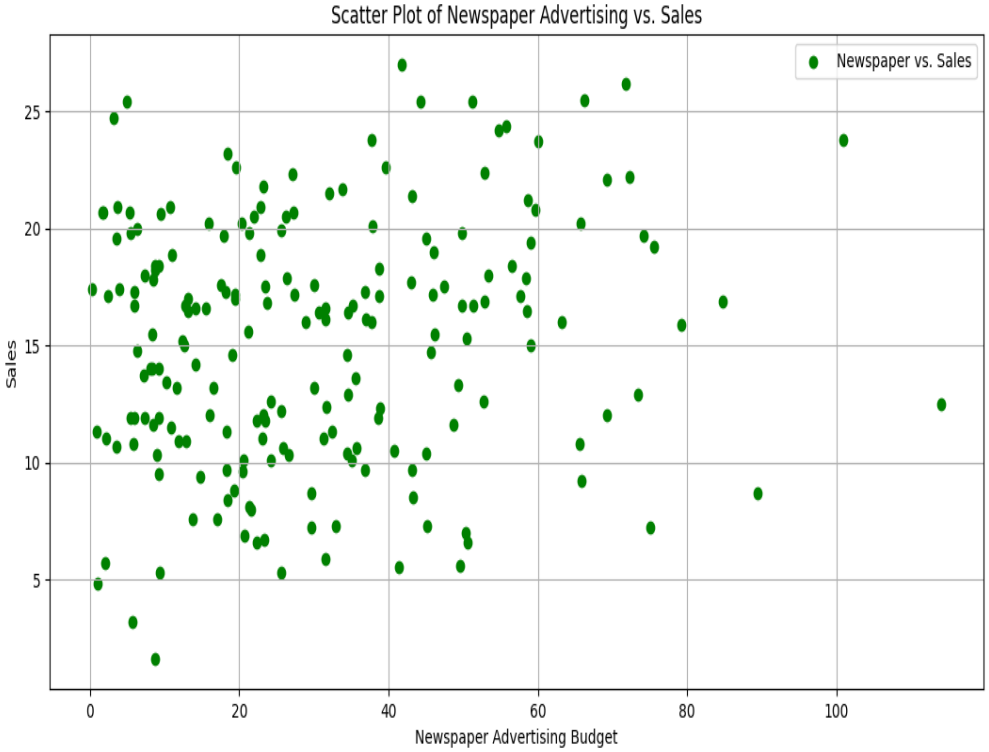
```
Run:  ads1
C:\Users\priya\PycharmProjects\pythonProjectlab\venv\Scripts\python.exe C:\Users\priya\PycharmProjects\pythonProjectlab\venv\ads1.py
      TV  Radio  Newspaper  Sales
0  230.1   37.8       69.2   22.1
1   44.5   39.3       45.1   10.4
2   17.2   45.9       69.3   12.0
3  151.5   41.3       58.5   16.5
4  180.8   10.8       58.4   17.9
                TV       Radio   Newspaper       Sales
count  200.000000  200.000000  200.000000  200.000000
mean   147.042500   23.264000   30.554000   15.130500
std     85.854236   14.846809   21.778621    5.283892
min      0.700000    0.000000    0.300000    1.600000
25%     74.375000    9.975000   12.750000   11.000000
50%    149.750000   22.900000   25.750000   16.000000
75%    218.825000   36.525000   45.100000   19.050000
max    296.400000   49.600000  114.000000   27.000000
TV           0
Radio        0
Newspaper    0
Sales        0
dtype: int64
Slope (Coefficient): 0.0498192084732535 7
Intercept: 9.561566398368491
Predicted Sales for TV = 250: 22.016368516681887
```

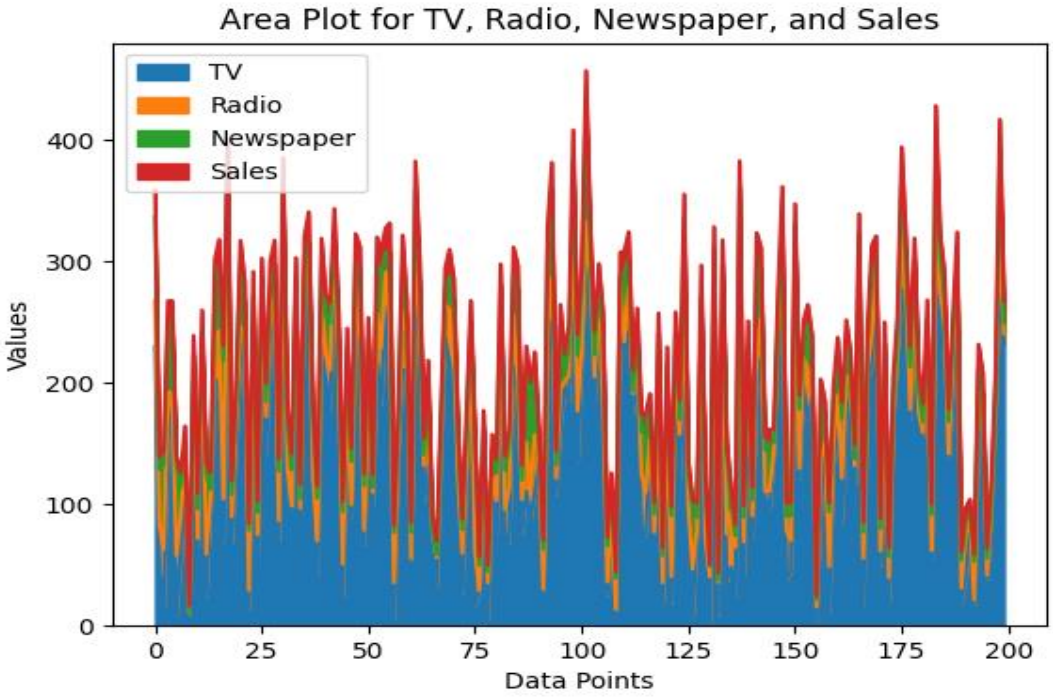


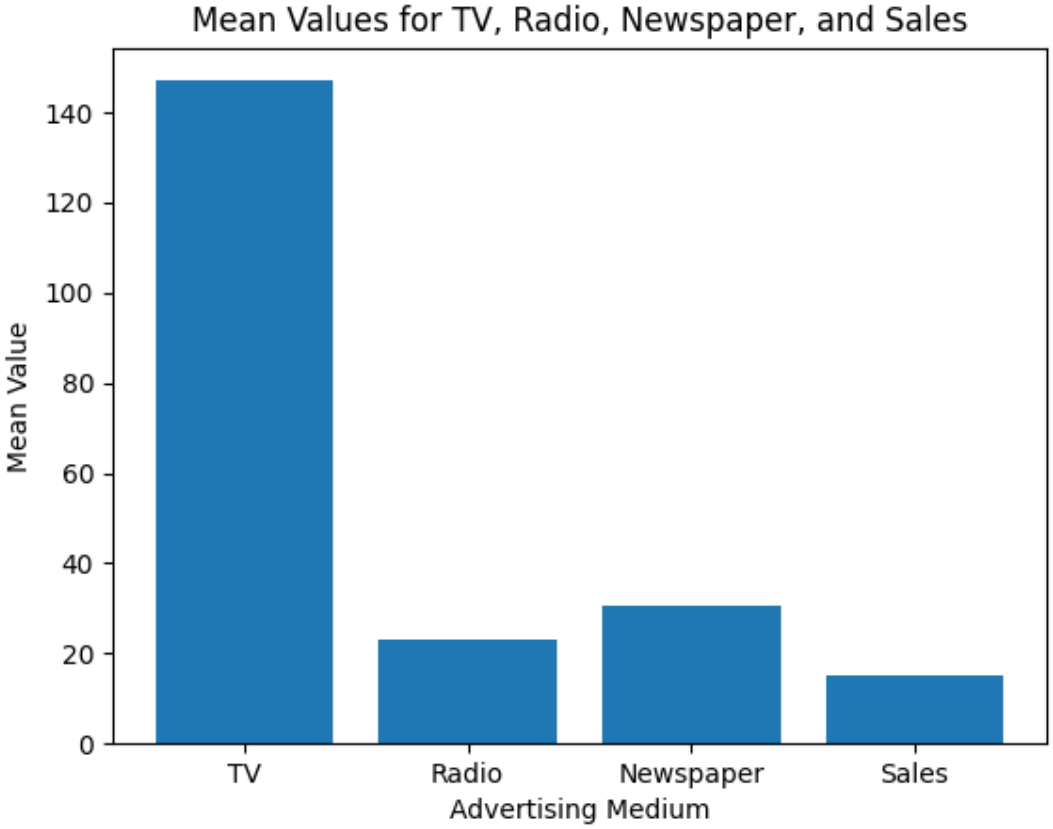

200
0
TV
50
0
Radio
100
0
Newspaper
20
10
Sales
0
50
100
150
200


Scatter Plot of TV Advertising vs. Sales
TV vs. Sales
Sales
TV Advertising Budget
25
20
15
10
5
0
50
100
150
200
250
300


Scatter Plot of Radio Advertising vs. Sales
Radio vs. Sales
Sales
Radio Advertising Budget
0
10
20
30
40
50
5
10
15
20
25


Scatter Plot of Newspaper Advertising vs. Sales
Newspaper vs. Sales
Sales
Newspaper Advertising Budget
0
20
40
60
80
100
5
10
15
20
25


Area Plot for TV, Radio, Newspaper, and Sales
TV
Radio
Newspaper
Sales
Values
0
100
200
300
400
0
25
50
75
100
125
150
175
200
Data Points


Mean Values for TV, Radio, Newspaper, and Sales
Mean Value
0
20
40
60
80
100
120
140
TV
Radio
Newspaper
Sales
Advertising Medium