DSA0410 – Fundamentals of Data Science

Day - 3

Name : Abinesh H

Reg.no.: 192424387

11. Scenario : You are a data scientist working for a company that sells products online. You have been tasked with creating a simple plot to show the sales of a product over time.

Question:

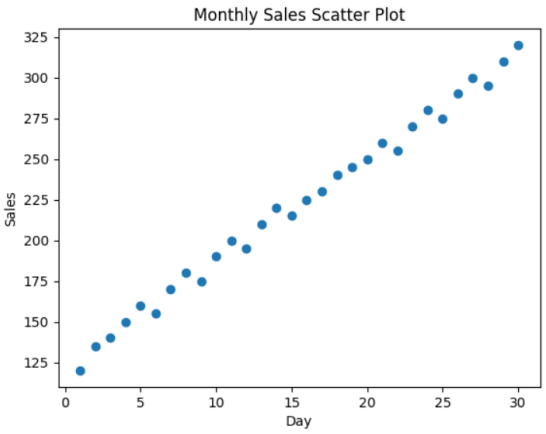
1. Write code to create a simple line plot in Python using Matplotlib to predict sales happened in a month?

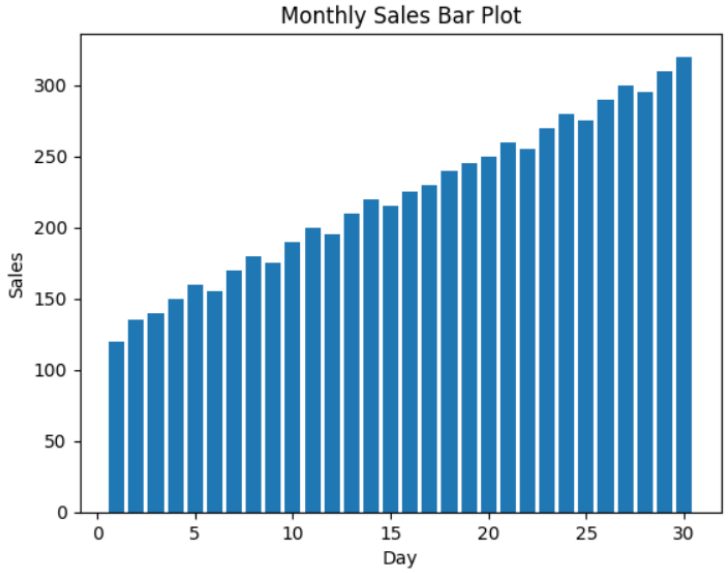
2. Write code to create a scatter plot in Python using Matplotlib to predict sales happened in a month?

3. Develop a Python program to create a bar plot of the monthly sales data.

```python
#exp11
import matplotlib.pyplot as plt

days = list(range(1, 31))
sales = [120, 135, 140, 150, 160, 155, 170, 180, 175, 190,
         200, 195, 210, 220, 215, 225, 230, 240, 245, 250,
         260, 255, 270, 280, 275, 290, 300, 295, 310, 320]
plt.plot(days, sales)
plt.xlabel("Day")
plt.ylabel("Sales")
plt.title("Monthly Sales Line Plot")
plt.show()
plt.scatter(days, sales)
plt.xlabel("Day")
plt.ylabel("Sales")
plt.title("Monthly Sales Scatter Plot")
plt.show()
plt.bar(days, sales)
plt.xlabel("Day")
plt.ylabel("Sales")
plt.title("Monthly Sales Bar Plot")
plt.show()
```





12. Scenario: You are working on a data analysis project that involves analyzing the monthly temperature and rainfall data for a city. You have a dataset containing the monthly temperature and rainfall values for each month of a year. Your task is to develop a Python program that generates line plots and scatter plots to visualize the temperature and rainfall data.

Question:

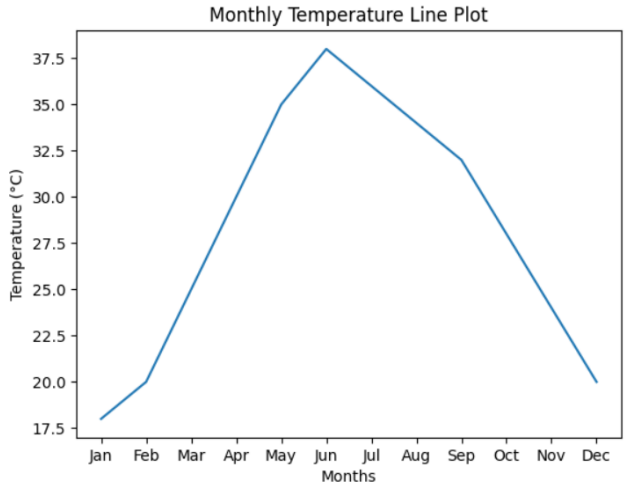
1. Develop a Python program to create a line plot of the monthly temperature data.

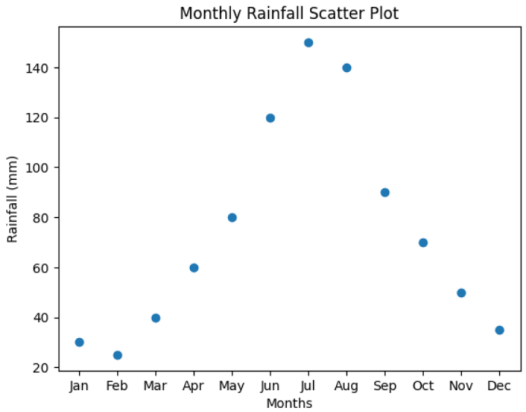
2: Develop a Python program to create a scatter plot of the monthly rainfall data.

```
#exp12
import matplotlib.pyplot as plt

months = ["Jan","Feb","Mar","Apr","May","Jun",
          "Jul","Aug","Sep","Oct","Nov","Dec"]

temperature = [18, 20, 25, 30, 35, 38, 36, 34, 32, 28, 24, 20]
rainfall = [30, 25, 40, 60, 80, 120, 150, 140, 90, 70, 50, 35]
plt.plot(months, temperature)
plt.xlabel("Months")
plt.ylabel("Temperature (°C)")
plt.title("Monthly Temperature Line Plot")
plt.show()
plt.scatter(months, rainfall)
plt.xlabel("Months")
plt.ylabel("Rainfall (mm)")
plt.title("Monthly Rainfall Scatter Plot")
plt.show()
```





13. Scenario: You are working on a text analysis project and need to determine the frequency distribution of words in a given text document. You have a text document named "sample_text.txt" containing a paragraph of text. Your task is to develop a Python program that reads the text document, processes the text, and generates a frequency distribution of the words.

Question: How would you develop a Python program to calculate the frequency distribution of words in a text document?

```python
#exp13
from collections import Counter

# Sample text dataset
text = """
Data science is an interdisciplinary field.
Data science uses scientific methods and algorithms.
Python is widely used in data science.
"""

# Convert text to lowercase
text = text.lower()

# Remove punctuation
for ch in ",.!?;:\"'()[]{}":
    text = text.replace(ch, "")

# Split text into words
words = text.split()

# Calculate word frequency
word_freq = Counter(words)

# Display frequency distribution
for word, count in word_freq.items():
    print(word, ":", count)
```

```
data : 3
science : 3
is : 2
an : 1
interdisciplinary : 1
field : 1
uses : 1
scientific : 1
methods : 1
and : 1
algorithms : 1
python : 1
widely : 1
used : 1
in : 1
```

14. Scenario: You are a data analyst working for a company that sells products online. You have been tasked with analyzing the sales data for the past month. The data is stored in a Pandas data frame.

Question: Develop a code in python to find the frequency distribution of the ages of the customers who have made a purchase in the past month.

```python
#exp14
import pandas as pd

# Sample sales dataset (customer ages)
data = {
    "Customer_ID": [101, 102, 103, 104, 105, 106, 107, 108, 109, 110],
    "Age": [22, 25, 22, 30, 35, 30, 25, 40, 35, 30],
    "Purchase_Amount": [500, 700, 450, 800, 900, 650, 720, 1000, 850, 780]
}

# Create DataFrame
sales_df = pd.DataFrame(data)

# Find frequency distribution of ages
age_frequency = sales_df["Age"].value_counts().sort_index()

# Display result
print(age_frequency)
```

```
Age
22    2
25    2
30    3
35    2
40    1
Name: count, dtype: int64
```

15. Scenario: You are a data analyst working for a social media platform. As part of your analysis, you have a dataset containing user interaction data, including the number of likes received by each post. Your task is to develop a Python program that calculates the frequency distribution of likes among the posts.

Question: Develop a Python program to calculate the frequency distribution of likes among the posts?

```python
#exp15
import pandas as pd

# Sample user interaction dataset
data = {
    "Post_ID": [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
    "Likes": [10, 25, 10, 50, 25, 30, 50, 10, 30, 25]
}

# Create DataFrame
df = pd.DataFrame(data)

# Calculate frequency distribution of likes
likes_frequency = df["Likes"].value_counts().sort_index()

# Display result
print(likes_frequency)
```

```
Likes
10    3
25    3
30    2
50    2
Name: count, dtype: int64
```