EXPERIMENT NO. 3

Title:

Implementing Line Plots, Area Plots, Histograms, Bar Charts, Pie Charts, Bubble Plots, Waffle Charts, and Word Clouds on Sample Data Points

Objective:

To create and interpret various types of plots including line plots, area plots, histograms, bar charts, pie charts, bubble plots, waffle charts, and word clouds using sample data points.

Brief Theory:

Introduction: Visualization is a powerful tool for understanding data. Different types of plots can reveal different aspects of the data. This experiment covers a wide range of plot types to provide a comprehensive understanding of data visualization techniques.

Plot Types and Their Uses

- **Line Plot:** Used to display data points over a continuous interval or time span.
- **Area Plot:** Similar to line plots but with the area below the line filled.
- **Histogram:** Used to represent the distribution of a continuous variable by dividing it into bins and counting the number of observations in each bin.
- **Bar Chart:** Used to compare the values of different categories.
- **Pie Chart:** Used to represent the proportions of different categories as slices of a pie.
- **Bubble Plot:** An extension of the scatter plot where each point is a bubble with its size representing a third variable.
- **Waffle Chart**: Used to represent parts of a whole in a grid-like fashion.
- **Word Cloud:** Used to visualize the frequency of words in a text dataset, where the size of each word indicates its frequency or importance.

Steps: -

Step 1: Import Necessary Libraries

First, we need to import the libraries required for data manipulation and visualization.

Code:-

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from wordcloud import WordCloud
import matplotlib.patches as patches
```

Step 2: Create Sample Data Points

Generate sample data points for each type of plot.

Code:-

```
# Sample data for plots
data = pd.DataFrame({
    'x': range(1, 11),
    'y': np.random.randint(1, 20, 10),
    'z': np.random.randint(1, 100, 10)
})
```

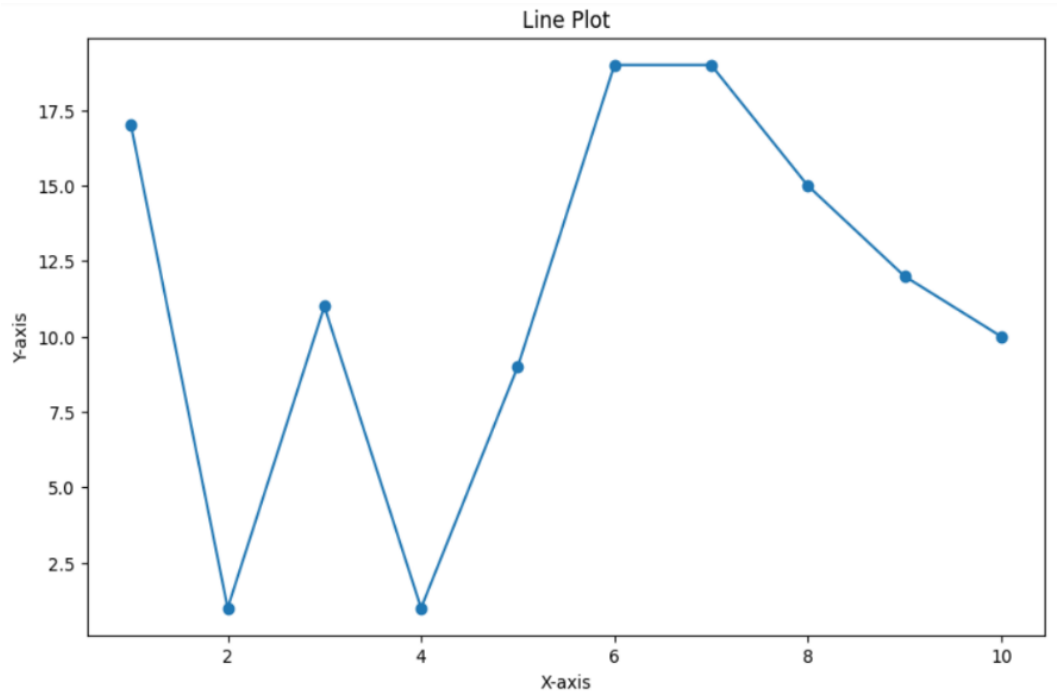
text = "data science machine learning data visualization word cloud python matplotlib seaborn numpy pandas"

Step 3: Line Plot

Create a line plot using the sample data points.

Code:-

```
plt.figure(figsize=(10, 6))
plt.plot(data['x'], data['y'], marker='o')
plt.title('Line Plot')
plt.xlabel('X-axis')
plt.ylabel('Y-axis')
plt.show()
```

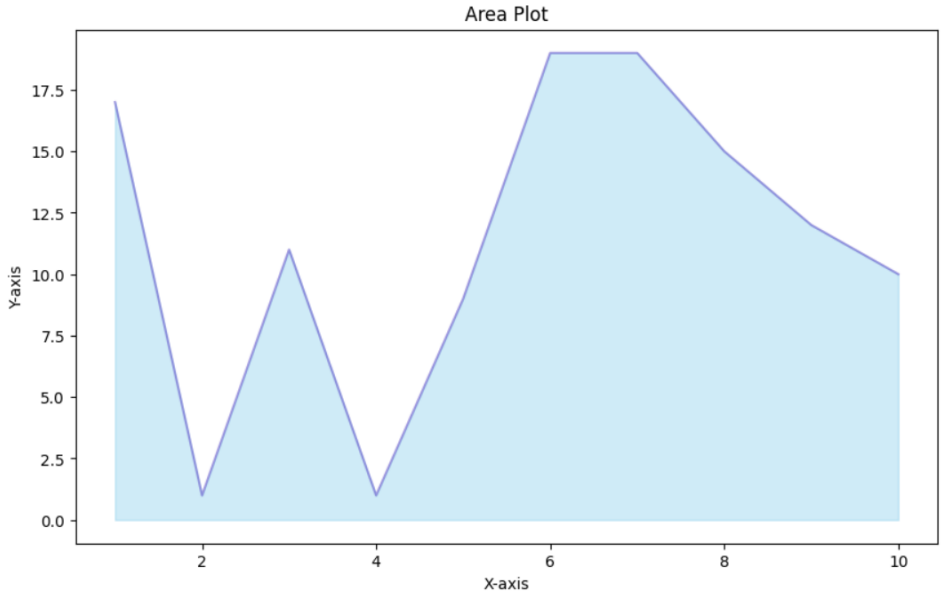


Step 4: Area Plot

Create an area plot using the sample data points.

Code:-

```
plt.figure(figsize=(10, 6))
plt.fill_between(data['x'], data['y'], color="skyblue", alpha=0.4)
plt.plot(data['x'], data['y'], color="Slateblue", alpha=0.6)
plt.title('Area Plot')
plt.xlabel('X-axis')
plt.ylabel('Y-axis')
```

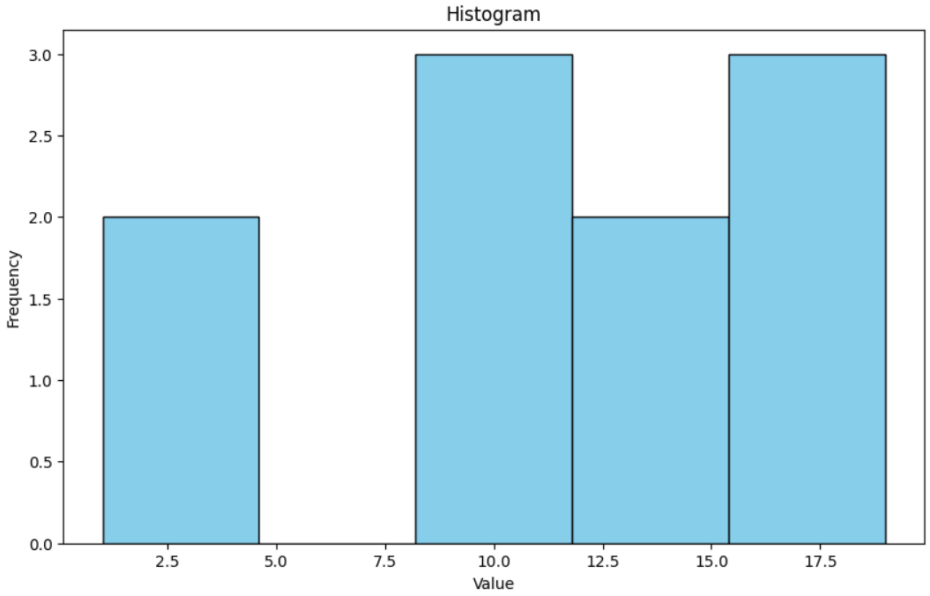


Step 5: Histogram

Create a histogram using the sample data points.

Code:-

```
plt.figure(figsize=(10, 6))
plt.hist(data['y'], bins=5, color='skyblue', edgecolor='black')
plt.title('Histogram')
plt.xlabel('Value')
plt.ylabel('Frequency')
plt.show()
```

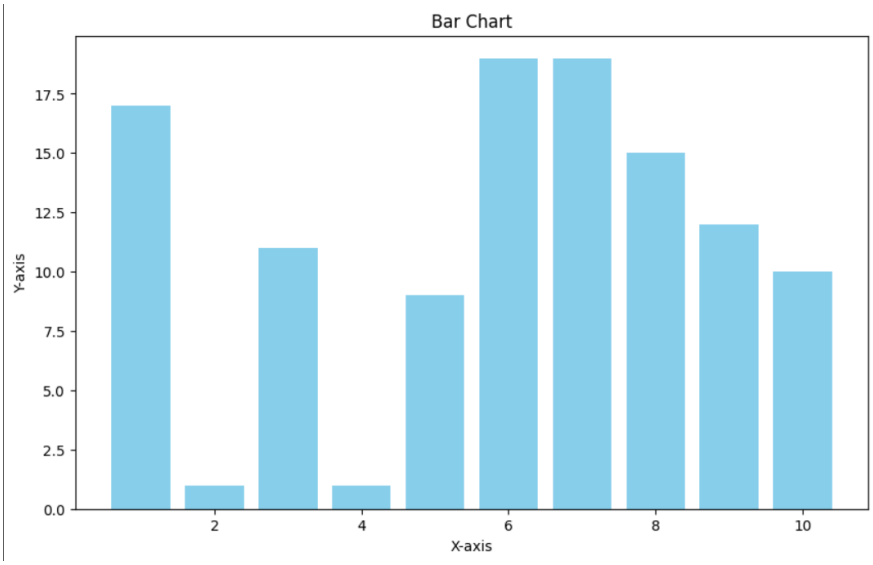


Step 6: Bar Chart

Create a bar chart using the sample data points.

Code:-

```
plt.figure(figsize=(10, 6))
plt.bar(data['x'], data['y'], color='skyblue')
plt.title('Bar Chart')
plt.xlabel('X-axis')
plt.ylabel('Y-axis')
plt.show()
```

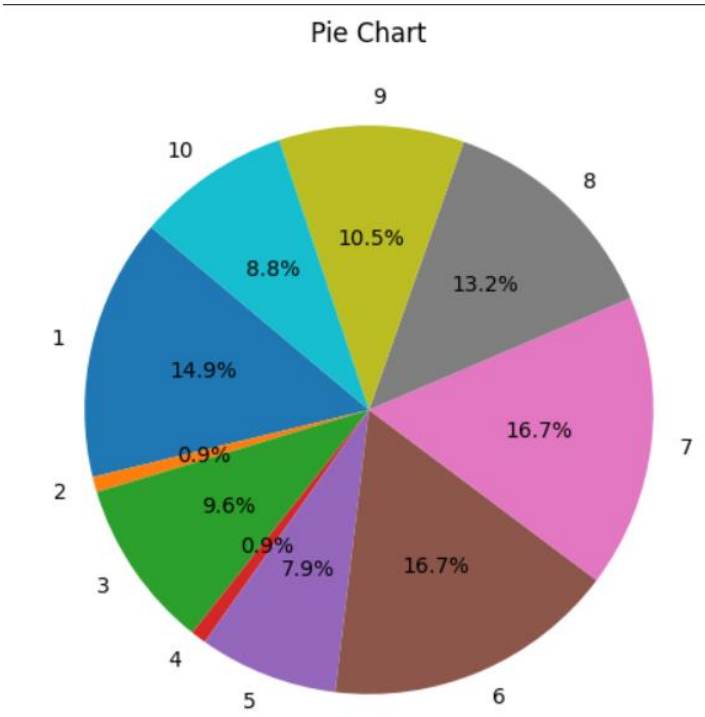


Step 7: Pie Chart

Create a pie chart using the sample data points.

Code:-

```
plt.figure(figsize=(10, 6))
plt.pie(data['y'], labels=data['x'], autopct='%1.1f%%', startangle=140)
plt.title('Pie Chart')
plt.show()
plt.show()
```

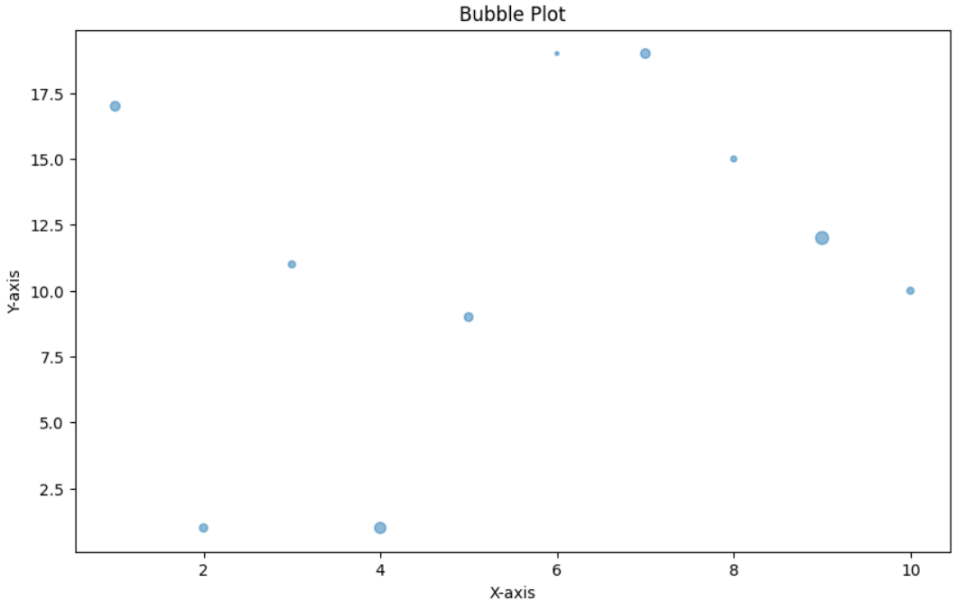


Step 8: Bubble Plot

Create a bubble plot using the sample data points.

Code:-

```
plt.figure(figsize=(10, 6))
plt.scatter(data['x'], data['y'], s=data['z'], alpha=0.5)
plt.title('Bubble Plot')
plt.xlabel('X-axis')
plt.ylabel('Y-axis')
plt.show()
```


Bubble Plot
Y-axis
X-axis
17.5
15.0
12.5
10.0
7.5
5.0
2.5
2
4
6
8
10

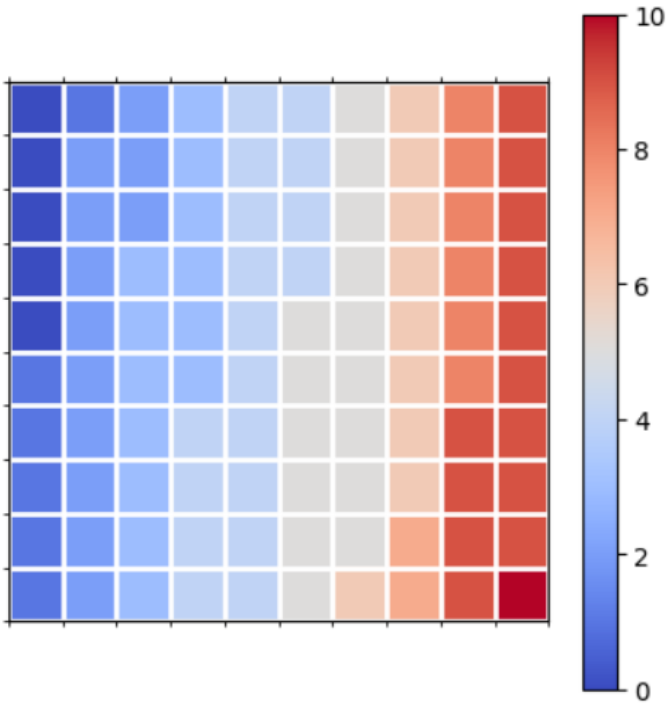
Step 9: Waffle Chart

Create a waffle chart using the sample data points.

Code:-

```
def make_waffle_chart(categories, values, height, width, colormap):
    total_values = sum(values)
    total_num_tiles = width * height
    category_proportions = [(float(value) / total_values) for value in values]
    tiles_per_category = [round(proportion * total_num_tiles) for proportion in
category_proportions]
    waffle_chart = np.zeros((height, width))
    category_index = 0
    tile_index = 0
    for col in range(width):
        for row in range(height):
            tile_index += 1
            if tile_index > sum(tiles_per_category[0:category_index + 1]):
                category_index += 1
                        waffle_chart[row, col] = category_index
    fig = plt.figure()
    colormap = plt.cm.get_cmap(colormap)
    plt.matshow(waffle_chart, cmap=colormap)
    plt.colorbar()
    ax = plt.gca()
    ax.set_xticks(np.arange(-.5, (width), 1), minor=True)
    ax.set_yticks(np.arange(-.5, (height), 1), minor=True)
    ax.grid(which='minor', color='w', linestyle='-', linewidth=2)
    plt.xticks([])
    plt.yticks([])
    values_cumsum = np.cumsum(values)
    total_values = sum(values)
    legend_handles = []
    for i, category in enumerate(categories):
        label_str = category + " (" + str(values[i]) + ")"
        color_val = colormap(float(values_cumsum[i]) / total_values)
        legend_handles.append(patches.Patch(color=color_val, label=label_str))
    plt.legend(handles=legend_handles, loc='best', bbox_to_anchor=(0.0, 0.0, 0.5, 0.5))
    plt.show()

categories = data['x']
values = data['y']
make_waffle_chart(categories, values, 10, 10, plt.cm.coolwarm)
```

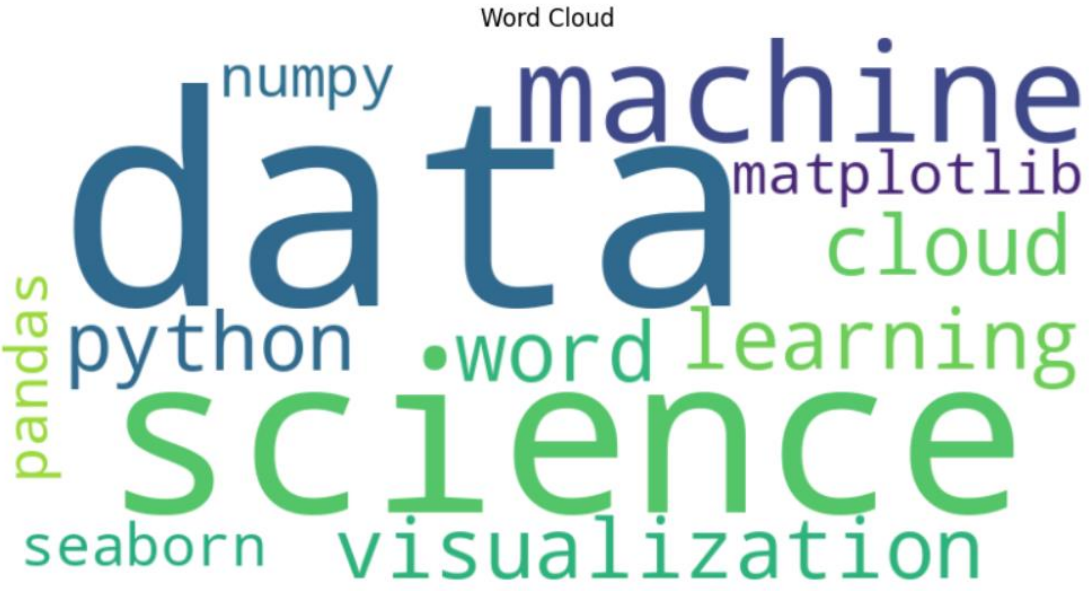


Step 10: Word Cloud

Create a word cloud using the sample text data.

Code:-

```
wordcloud = WordCloud(width=800, height=400, background_color='white').generate(text)
plt.figure(figsize=(10, 6))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title('Word Cloud')
plt.show()
```



Conclusion: -

In this experiment, we successfully implemented various types of plots including line plots, area plots, histograms, bar charts, pie charts, bubble plots, waffle charts, and word clouds using sample data points. Each type of plot provides unique insights and visual representations of the data, which are essential for data analysis and interpretation.

Practice Questions:

1. Using the data below, create a bubble plot where the x-axis represents the GDP (in trillion USD), the y-axis represents the population (in millions), and the size of the bubble represents the carbon emissions (in million tons): `Country A: GDP - 3, Population - 150, Emissions - 500; Country B: GDP - 5, Population - 200, Emissions - 800; Country C: GDP - 2, Population - 100, Emissions - 300; Country D: GDP - 4, Population - 250, Emissions - 600`. What insights can you gain from the bubble plot?
2. Given the following data for five products and their respective sales in units: `Product A: 120, Product B: 180, Product C: 140, Product D: 200, Product E: 160`, create a bar chart to represent the data. Which product had the highest and lowest sales?
3. A company's revenue is divided into four categories: `Product Sales: 40%, Services: 30%, Investments: 20%, Other: 10%`. Create a pie chart to represent the revenue distribution. How well does the pie chart represent the proportion of revenue from each category?
4. Generate a histogram using random normal data points (mean = 0, standard deviation = 1) and describe the shape of the distribution.

Expected Oral Questions

1. Why would you choose a line plot to represent the data? What are the key features of a line plot that make it suitable for this purpose?
2. How can you interpret the slope of the line in your plot? What does an upward or downward slope indicate in the context of your data?
3. How does an area plot differ from a line plot? What additional information does the area under the curve provide?
4. Can you explain how to interpret overlapping areas in an area plot?
5. What is the purpose of a histogram? How does it differ from a bar chart?
6. What are the main advantages of using a bar chart to represent categorical data?
7. How would you decide whether to use a vertical or horizontal bar chart for your data?
8. Why might a pie chart be chosen over other types of charts for certain datasets? What are the limitations of pie charts?
9. What are the three dimensions represented in a bubble plot, and how do they provide more

Information than a typical 2D scatter plot?

10. What makes a waffle chart an effective visualization for showing proportions? How does it compare to a pie chart?
11. What is the purpose of a word cloud, and how can it help in understanding textual data?
12. What are the limitations of word clouds? How might they lead to misinterpretations if not used carefully?

FAQs in Interviews

Q: What are the key factors to consider when choosing a type of chart or plot for data visualization?

A: The key factors include the nature of the data (quantitative or categorical), the message you want to convey, the complexity of the data, and the audience's familiarity with the chart types. For instance, line plots are great for trends over time, while pie charts are best for showing parts of a whole.

Q: How do you ensure that your visualizations are not misleading?

A: To avoid misleading visualizations, ensure that the scale is appropriate, data points are not omitted, colors and labels are clear, and the chart type chosen accurately reflects the data relationship. It's also important to provide context and avoid manipulating the visual to exaggerate findings.

Q: What challenges have you faced in data visualization, and how did you overcome them?

A: Common challenges include dealing with large datasets, choosing the correct visualization for complex data, and ensuring that the visualization is accessible and understandable to the audience. Overcoming these challenges often involves simplifying the data, using interactivity, or combining multiple visualizations for clarity.

Q: How can you highlight important data points in a line plot?

A: Important data points can be highlighted using markers, annotations, or different colors for specific sections of the line. You can also add trend lines or emphasize peaks and troughs with callouts.

Q: How can you interpret the shape of a histogram?

A: The shape of a histogram can indicate the distribution of the data (e.g., normal, skewed, bimodal). A symmetrical bell-shaped histogram suggests a normal distribution, while a skewed histogram indicates that the data has a longer tail on one side.