**Name:** Manish More

**Roll no:** 282010 **PRN:** 22310220

**Class :** SY CSE AI **Batch:** B1

**Assignment 3: Data Visualization**

**Problem Statement:**Visualize the data using R/Python by plotting the graphs for assignment no. 1 and 2. Consider suitable data set. Use Scatter plot, Bar plot, Box plot, Pie chart, Line Chart.

**Task Overview:**   
Visualize the data using Python by plotting various graphs for analyzing the Iris dataset. The visualization techniques include:

1. **Scatter Plot:** To analyze relationships between numerical variables.
2. **Bar Plot:** To compare feature averages for different categories.
3. **Box Plot:** To visualize data distribution and detect outliers.
4. **Pie Chart:** To represent proportions within categorical data.
5. **Line Chart:** To observe trends over sequences.

**Objective**

1. Gain hands-on experience with **Matplotlib** and **Seaborn** for data visualization.
2. Understand how different plots reveal insights into data.
3. Explore practical applications of visualization techniques.

**Tools and Resources**

* **Software:** Google Colab / Jupyter Notebook
* **Libraries:** Pandas, Matplotlib, Seaborn, scikit-learn

**Key Functions Used**

1. **Scatter Plot**: Analyze relationships between Sepal Length and Sepal Width by species.
2. **Bar Plot:** Compare mean feature values across different Iris species**.**
3. **Box Plot:** Identify data distribution and outliers in Iris features.
4. **Pie Chart:** Show proportions of species (created using custom logic).
5. **Line Chart:** Observe trends in feature measurements across sample indices.

**Methodology**

**1.Data Loading and Exploration**

1. Loaded the Iris dataset using load\_iris() from scikit-learn.
2. Converted the data into a Pandas DataFrame for easier manipulation.
3. Added species labels to the DataFrame.

**2.Visualization Process**

* Selected appropriate plots to represent relationships and trends.
* Generated plots using Matplotlib and Seaborn.
* Customized graphs with titles, labels, legends, and colors for readability.

**Advantages of Data Visualization**

1. **Simplifies Complex Data:** Makes patterns and relationships easier to understand.
2. **Detects Outliers:** Box plots reveal extreme data points.
3. **Enhances Data Interpretation:** Scatter plots and bar plots highlight key insights.
4. **Improves Decision-Making:** Visual tools make comparisons more intuitive.

**Challenges:**

* Choosing suitable plots for representing specific data characteristics.
* Properly handling plot scaling and customization for better readability.

**Conclusion**

This assignment provided practical experience in:

* Creating various plots for analyzing the Iris dataset.
* Understanding the importance of visualization for data analysis.
* Exploring visualization techniques applicable in data science projects.