**SVKM’s NMIMS**

**Mukesh Patel School of Technology Management & Engineering**

Program: B.Tech\MBA.Tech\MBA.Tech AI

**Course: Machine Learning**

**Experiment No.03**

PART A

(PART A : TO BE REFFERED BY STUDENTS)

**A.1 Aim: To understand and implement data visualization techniques**

**A.2 Prerequisite:**

Python Programming, Pandas library, Numpy Library, MatplotLib, Seaborn Library

**A.3 Outcome:**

**After successful completion of this experiment students will be able to:**

* 1. Read different types of data files (csv, excel, text file etc.)
  2. Understand usage of different types of Python libraries for plotting data
  3. Plotting of data using different types of plots

**A.4 Theory:**

Data visualization is a form of visual communication. It involves the creation and study of the visual representation of data. It translates the data to a more natural form for the human mind to comprehend and pick out patterns or points of interest.

**Matplotlib and seaborn** are among the common libraries for visualizing data in Python.

Matplotlib is a python library used extensively for the visualization of data. While Seaborn is a python library based on matplotlib. Seaborn provides a high-level interface for drawing attractive and informative statistical graphics.

Most common types of plots used in data visualization:

* Scatter plot
* Pair plot
* Box plot
* Violin plot
* Distribution plot
* Joint plot
* Bar chart
* Line plot

**Scatter Plot:**

It is one of the most commonly used plots for simple data visualization. It gives us a representation of where each point in the entire dataset are present with respect to any 2 or 3 features (or columns)

**Pair plot:**

Pair plot will help us create us a (n x n) figure where the diagonal plots will be histogram plot of the feature corresponding to that row and rest of the plots are the combination of feature from each row in y axis and feature from each column in x axis.

**Box plot:**

A box plot (or box-and-whisker plot) shows the distribution of quantitative data in a way that facilitates comparisons between variables or across levels of a categorical variable. The box shows the quartiles of the dataset while the whiskers extend to show the rest of the distribution.

**Violin Plot**

The violin plots can be inferred as a combination of Box plot at the middle and distribution plots (Kernel Density Estimation) on both side of the data. This can give us the details of distribution like whether the distribution is multimodal, Skewness etc.

**Joint Plot:**

Join plots can do both univariate as well as bivariate analysis. The main plot will give us a bivariate analysis, whereas on the top and right side we will get univariate plots of both the variables that were considered. It makes our job easy by getting both scatter plots for bivariate and Distribution plot for univariate, both in a single plot.

**Tasks:**

1. Read “seeds.csv” file into data frame.
2. Explore the dataset by using head and describe.
3. Find the number of samples per type. Plot a histplot for the count.
4. Plot a scatter plot for Kernel Width vs Length. Write your inference.
5. Plot a Jointplot to understand relation between Perimeter and Compactness. Write your inference.
6. Plot a Scatter plot to compare Perimeter and Compactness. Different type should have different colours. (hint use legend).
7. Plot a Box plot to understand correlation between compactness and type.
8. Plot a Box & Strip plots to understand correlation between compactness and type. State your inference.
9. Plot a Box & Strip plots to understand correlation between perimeter and type.
10. Plot a Violin and strip subplots to understand correlation between compactness and type. State your inference
11. Plot a Kernel Density Estimation plots to understand correlation between compactness and type. State your inference
12. Plot a pair plot to understand all characteristics with type being the main parameter. State your inference
13. Plot a pair plot to understand all characteristics with type being the main parameter. (the main parameter, with KDE instead of histogram in diagonal subplots)
14. An Andrews curve to display separability of data according to Type.
15. Plot a bar plot for below X and Y values.

X= [2,8,10] Y= [11,16,9]

X2= [2,3,6] Y2= [4,16,9]

PART B

(PART B : TO BE COMPLETED BY STUDENTS)

***(Students must submit the soft copy as per following segments within two hours of the practical.)***

|  |  |
| --- | --- |
| Roll No. I066 | Name: Srihari Thyagarajan |
| Class: B Tech Artificial Intelligence | Batch: B2 |
| Date of Experiment: 12/01/2023 | Date of Submission: 12/01/2023 |
| Grade : |  |

**B.1 Tasks**

**B.4 Conclusion:**

*(Students must write the conclusion in their own words.)*