

Practical 10

AIM:

Download the Iris flower dataset or any other dataset into a DataFrame. (e.g., <https://archive.ics.uci.edu/ml/datasets/Iris>). Scan the dataset and give the inference as:

1. List down the features and their types (e.g., numeric, nominal) available in the dataset.
2. Create a histogram for each feature in the dataset to illustrate the feature distributions.
3. Create a box plot for each feature in the dataset.
4. Compare distributions and identify outliers.

REQUIREMENT:

AnacondaInstaller

Windows10OS

Linux

JupyterNotebook

THEORY:

What is DataVisualization?

Datavisualizationsimplyreferstotechniques thatareusedto communicate with insights from data through a visual representation. To simplify the thing, you can say that data visualization visually puts data. So we can easily understand it. One has to master different data visualization tools to become effective.

Data visualization is what will help you out. To learn more about data visualization and other visual representation in data science, check out our datascience certifications from recognized universities.

Categorical Data:

- Categorical data represents **categories or labels**
- Types of categorical data:
 - **Nominal:** No specific order (e.g., colors, names)
 - **Ordinal:** Has a specific order (e.g., ranks, grades)

Formats of Categorical Data:

- **Raw Data:** Individual observations
 - **Aggregated Data:** Counts of each category
 - **Cross-tabulated Data:** Data shown in table form (rows & columns)
-

Working with Categorical Variables:

- Represented as:
 - **Character vectors**
 - **Factors**
-

Advantages of Factors:

- Better control over **order of categories**
 - Levels remain preserved during **data operations**
-

Common Functions (R):

- **factor()**: Converts data into factor
- **levels()**: Shows categories (levels)
- **reorder()**: Changes order of levels

LibrariesUsed:

Seaborn:

Pandas:

Numpy

Sklearn:

CONCLUSION:

In this experiment we have studied about the data visualization and performed data visualization on the iris data set in different ways.