Name: Omkar Gaikwad
Batch: L1
Roll no. 31126

DSBDAL
Assignment 10

PICT, PUNE

Title: Data visualization III

Problem statement:

Download the iris flower dataset or any other dataset into a dataframe. scan dataset and give inference as:

1) List down features and their types available in the dataset.
2) create a histogram for each feature in the dataset to illustrate the feature distributions.
3) create a boxplot for each feature in dataset.
4) compare distributions and indentify outliers.

Objectives

- To understand various visualization techniques using seaborn python library.
- to apply appropriate plotting techniques to visualize all the features.
- Describe the observations made by using each plot graph.

learning outcomes:

students will be able to -

- perform basic data visualizations using appropriate graphs.
- Infer from the graphs about different features.

PICT, PUNE

Software and Hardware Requirements:

Linux OS (ubuntu). Intel core is 8th gen, (8 GB RAM) python 3.8, jupyter Notebook.

Theory:

Data visualization

It is the representation of data through use of common graphics, such as charts plots, info graphics and even animations. These visual displays of information communicate complex in a way that is easy way to understand.

seaborn library.

It is data visualization library built on top of matplotlib and closely integrated with pandas data structure. Visualization is central part of seaborn library which helps in exploration and understanding of the data.

Numeric data type.

It refers to a data type that can be stored and identified based on form of numbers and not any descriptive form.

Nominal data type

It is type of data that is used to label variables without providing any quantitative value.

PICT, PUNE

Observations:

- The nominal data types are - species, numeric data types are sepal len, sepal width, petal len, petal width.
- There are no null values in dataset.
- The attributes of Iris. versicolor and Iris-virginica are almost similar. They are, however, different in terms of sepal length and sepal width.
- Iris-setosa has a very small petal length and width as compared to the other two.
- The histogram depicts the variation of each feature wrt species.
- The box plot depicts distribution of each feature wrt species
- The box plot can also be used to find outliers

Conclusion

Thus we have successfully applied various visualisation techniques and inferred the variations of all the features wrt species in Iris dataset.