PICT, PUNE

Name: Aarkan Kornik
Roll: 31140

DSBDAL

Assignment 10

- Title: Data Visualization III

- Problem statement:
Download the iris flower dataset or any other dataset into a dataframe scan dataset & give inference as:

1) List down features & 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 & identify outliers.

- Objectives:

i) To understand various visualization techniques using seabern python library.

ii) To apply appropriate plotting techniques to visualize all the features.

iii) Describe the observations made by using each plot graph.

PICT, PUNE

3114 0

- Learning outcomes:
  Students will be able to:

i) Perform basic data visualizations using appropriate graphs.

ii) Infer from the graphs about different features.

- Software & Hardware Requirements
  Linux Os (Ubuntu). Intel core i3 10th 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, infographics & even animations. These visual displays of information communicate complex in a way that is easy way to understand.

Sea born library
It is data visualization library built on top of mat plot lib & closely integrated with pandas data structure. Visualization is central part of seaborn library which helps in exploration & understanding of data.

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

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

- Observations:

i) The nominal data types are species, numeric data types are sepal len, sepal width, petal len, petal width

ii) There are no null values in dataset.

iii) The attributes of Iris versicolor & Iris virginica are almost similar. They are however different in terms of sepal length & sepal width.

iv) Iris setosa has a very small petal length & width as compared to the other two.

v) The histogram depicts the variation of each feature w.r.t. species.

vi) The box plot depicts the distribution of each feature w.r.t. species.

PICT, PUNE

31140

vii) The box plot can also be used to find outliers.

- Conclusion:
Thus we have successfully applied various visualisation techniques & inferred the variations of all the features w.r.t the species in Iris dataset.