PICT, PUNE

Name: Aarkin Karnik
Roll no: 31140

DSBDA Lab
Assignment -3

- Title: Descriptive statistics - Measures of central tendancy & variability

- Problem Statement:
Perform the following operations on any open source dataset.

1) Provide summary statistics for a dataset with numeric variables grouped by one of the qualitative variable for example, if your categorised variable is age groups & quantitative variable is income, then provide summary statistics of income grouped by the age groups. Create a list that contains a numeric value for each response to the categorical variable.

2) Write a python program to display some basic statistics details like percentile, mean, standard deviation, etc. of the species of iris-setosa.

- Learning objective:
To understand the concept of statistical operation.

- S/W & H/W Requirements:
1) OS windows 10 (64 bit)
2) Programming Lang: Python 3.9
3) Jupyter Notebook
4) 8 GB RAM, Intel i5 - 8300H, 256 GB SSD

PICT, PUNE

Theory:

We have used two dataset for this arrangement

- Dataset 1: rbq.csv

The set consists of following columns

- Dataset 2: iris.csv

The dataset consists of following columns

Id, sepallengthcm, sepalliswidthcm, petallengthcm

Task 1:

Importing packages

- To load the pandas, numpy packages.

Importing datasets

- dataset = pd.read_csv("rbq.csv")

Print dataset

- dataset.head() (prints first 5 rows)

- info(): provides information about dataset.

- describe(): provides overview of numerical data.

- shape(): it gives the no. of rows & cols. in the dataset.

- isnull().shape(): it gives the null values in data frame.

- unique(): it gives unique values in the column.

PICT, PUNE

31140

- Measures of central tendancy

- Mean: Represents the arithmetic average of data.
- Median: It represents the middle value of data that separates the distribution into halves.
- Mode: It represents the most frequent value of a variable in the data set.

- Task 2:

Importing dataset

- import the iris dataset using pandas
  dataset: pd: read_csv ("iris.csv")

- print dataset
  dataset: head()

- pre processing
- shape
- info()
- describe
- isnull(). sum()

PICT, PUNE

31140

- Central tendancy:

- Mean:
Calculate mean of sepal Length, sepal width, petal Length, petal Width grouped by species that is 'iris-setar'.

- Median:
Calculate median of sepal Length, sepal width, petal Length, petal width grouped by species.

- Mode:
Calculate mod of sipal Length, sepal width, petal length, petal Width, grouped by specie

- Conclusion:
We successfully implemented assignment using statistical measures & understand the concept of descriptive statistics.