Ramniranjan Jhunjhunwala College
MSC DSAI
EDA / VDA Project

Dataset: iris-m.xlsx. Refer iris.txt for dataset description.

Use Python and carry out EDA / VDA on Iris.csv and provide answers to following questions

1. What is structure of the dataset.
2. What are the data type of each columns?
3. What is the length of alpha numeric columns?
4. What are precision & scale of numeric columns?
5. What are the significant columns?
6. Identify significant columns of the dataset.
7. Find out for each significant column
   - Number of Null values
   - Number of zeros
8. For each significant column
   - Provide the obvious errors
9. For each numeric column
   - Replace null values & zeros with mean value of the column.
10. For each significant column
    - Provide the quartile summary along with the count, mean & sum
11. For each significant column
    - Provide the range, variance and standard deviation
12. For each significant column
    - Provide the count of outliers and their value
13. Are there any class variables? If yes,
    - provide frequency distribution table & chart for the same
14. For all numeric columns
    - Provide histogram
15. For all numeric variables
    - Provide box & whisker plot
16. For all numeric variables
    - Provide correlation table & graph
17. Based on the correlation table
    - Provide scatter plot relevant columns
18. Use relevant statistical test to check if population mean value of SepalLength is significantly different from 6.25. Use 0.06 as α
19. Use relevant statistical test to check if the population mean value of PetalWidth is no more than 1.5.
20. Prepare a summary report to interpret the above results.

Presentation

- The first section introduces you & your team
- Each of the above query is answered on a separate slide.
- The second last section should contain the interpretation of results.
- The last section should describe your experience of creating this project.

Project Submission

1. Project to be done as per groups assigned in your class.
2. Prepare the project using .py file.
3. The single .py files should be consolidated into a single zip file
   RJ-MSC-DS-GroupNo-GroupName.zip
   Eg RJC-MSC-DS-01-CodeMasters.zip
4. The .zip file needs to be submitted in "Project WOrk" assignment of Google Classroom
   Only one submission per group is required
5. The project needs to be submitted by _____ end-of-day.
6. The viva / presentation for the project will be held on _____ & _____.
7. Zoom meeting will be set up for each group.
8. Python related questions will be asked based on the not just the project but the full course.

Final Evaluation	**100**
Coding (how good is the code - same for all team members)	*20*
Documentation (inline comments & explanations- same for team members)	*20*
Accuracy & Interpretation (are the results correct same for team members)	*20*
Viva Per Individual (different for all team members)	*40*

Wishing You All The Best!!!