Module 4: Python Project

Dear Students,

Congratulations on completing the Python module!

As a culminating project, you'll be working with a dataset from ABC company, consisting of 458 rows and 9 columns. The company requires a comprehensive report detailing information about their employees across various teams. Your tasks include preprocessing the dataset, analyzing the data, and presenting your findings graphically. Here's a breakdown of what you need to do:

Preprocessing:
Correct the data in the "height" column by replacing it with random numbers between 150 and 180. Ensure data consistency and integrity before proceeding with analysis. (1 mark)

Analysis Tasks:
1. Determine the distribution of employees across each team and calculate the percentage split relative to the total number of employees. (2 marks)
2. Segregate employees based on their positions within the company. (2 marks)
3. Identify the predominant age group among employees. (2 marks)
4. Discover which team and position have the highest salary expenditure. (2 marks)
5. Investigate if there's any correlation between age and salary, and represent it visually. (2 marks)

Graphical Representation:
For each of the five analysis tasks, create appropriate visualizations to present your findings effectively. (5x2 = 10 marks)

Data Story:
Provide insights gained from the analysis, highlighting key trends, patterns, and correlations within the dataset. (3 marks)

Timely Submission:
Ensure timely submission of your project to earn an additional mark. (1 mark)
Total Marks : 25

By completing these tasks, you'll not only demonstrate your proficiency in Python programming but also your ability to analyze and communicate insights from real-world data.

Download the dataset from this link.

https://docs.google.com/spreadsheets/d/1VP9BE_eI2yl6uUHSm4mGiiwjRdoqCqnkcIjsv5Q2ex4/edit?usp=share_link
(P.T.O)

To submit the project:
1. Create a repository on your GitHub account.
2. Write a comprehensive overview of the project in the README file, summarizing the preprocessing steps, analysis tasks, graphical representations, insights gained, and any additional information.
3. Upload a well-organized Jupyter Notebook file containing the code for the project, ensuring proper formatting and documentation.
4. Include the dataset in the repository.
5. Submit the link to your repository in the designated section on Google Classroom.

Best of luck, and feel free to reach out if you have any questions or need assistance along the way!