Capstone Project: Exploratory Data Analysis With Pandas and Seaborn

Objective:

The objective of this capstone project is to perform an exploratory data analysis (EDA) on the "Ultra-Marathon Race" dataset available on Kaggle. You will replicate the analysis demonstrated in the provided YouTube tutorial video, ensuring your work aligns closely with the video to achieve a passing mark. The final deliverables include a Jupyter Lab notebook (or Google Colab notebook) and a PowerPoint presentation summarizing your findings.

Project Structure:

1. **Watch the YouTube Video:**
    - **Link:** YouTube Video Tutorial
    - Take detailed notes on the steps followed in the video.
    - Pay close attention to the data exploration techniques, visualizations, and any insights shared.
2. **Download the Dataset:**
    - **Kaggle Dataset:** The Big Dataset of Ultra Marathon Running
    - Load the dataset into your Jupyter Lab or Google Colab environment.
3. **Set Up the Environment:**
    - Install necessary libraries (pandas, numpy, matplotlib, seaborn, etc.).
    - Ensure your environment matches the setup used in the video for consistent results.
4. **Data Cleaning and Preprocessing:**
    - Follow the data cleaning steps as shown in the video.
    - Handle missing values, data type conversions, and any other preprocessing tasks mentioned.
5. **Exploratory Data Analysis (EDA):**
    - Perform the EDA as demonstrated in the video.
    - Include the same visualizations, such as histograms, scatter plots, box plots, etc.
    - Analyze key metrics and trends in the dataset, just as done in the tutorial.
6. **Insights and Observations:**

- Document the insights and observations made during the EDA.
- Ensure your conclusions align with those drawn in the video.

7. **Submit the Jupyter Lab/Google Colab Notebook:**
    - Ensure your notebook is well-organized and includes comments and markdown cells explaining each step.
    - Double-check that your work closely matches the video to ensure it meets the passing criteria.
8. **Create a PowerPoint Presentation:**
    - Summarize your findings in a concise PowerPoint presentation.
    - **Slides to Include:**
        1. **Title Slide:** Project title, your name, and date.
        2. **Introduction:** Briefly describe the dataset and the objective of the analysis.
        3. **Data Cleaning:** Key steps in data cleaning and preprocessing.
        4. **EDA Findings:** Visualizations and key insights discovered during the analysis.
        5. **Conclusion:** Summarize the main takeaways from the project.
    - Use visuals from your analysis to support your findings.
9. **Final Submission:**
    - Submit your Jupyter Lab or Google Colab notebook along with the PowerPoint presentation.
    - Ensure all files are clearly named and organized.

Evaluation Criteria:

- **Accuracy:** How well your work matches the analysis done in the video.
- **Clarity:** The organization and clarity of your notebook and presentation.
- **Insights:** The depth of insights and observations drawn from the data.
- **Presentation:** The effectiveness of your PowerPoint slides in communicating your findings.

Additional Tips:

- Double-check each step to ensure it aligns with the video.
- Use markdown cells in your notebook to explain your code and findings clearly.
- Keep your presentation slides clean, with minimal text and clear visuals.

This project will not only demonstrate your skills in exploratory data analysis but also your ability to follow and replicate complex data analysis workflows, which is a crucial skill for any data analyst.