DATA CRUNCH

Level 1: Data Preprocessing (25 Points)

- **Outlier Detection (8 pts)**
    - Proper identification of outliers using statistical techniques.
    - Justification of the chosen method (IQR, Z-score, etc.).
- **Outlier Handling (8 pts)**
    - Correct approach for handling/removing outliers.
    - Explanation of how it improves data quality.
- **Data Normalization (9 pts)**
    - Appropriate scaling technique (Min-Max, Z-score, etc.).
    - Justification for normalization choice.

Level 2: Data Analysis & Visualization (45 Points)

- **K-Means Clustering (15 pts)**
    - Correct implementation with justification for cluster selection.
    - Effective use of the Elbow Method or Silhouette Score.
    - Meaningful interpretation of cluster groups.
- **Visualization Quality (15 pts)**
    - Effective and well-labeled graphs using Seaborn & Matplotlib.
    - Appropriate use of histograms, scatter plots, heatmaps, etc.
    - Aesthetics, clarity, and correctness.
- **Interpretability & Storytelling (15 pts)**
    - Insights are clearly communicated.
    - Logical flow between analysis steps.
    - Ability to explain complex findings in a simple way.

Level 3: Comprehensive Data Analysis (30 Points)

- **Depth of Insights (15 pts)**
    - Findings go beyond surface-level observations.
    - Contextual relevance to social media/healthcare.
    - Real-world implications discussed.
- **Report Structure & Presentation (10 pts)**

    - Well-organized report with introduction, methodology, results, and conclusion.
    - Effective use of visuals and narrative.
    - Clear, concise language and grammar.
- **Innovation & Advanced Techniques (5 pts)**
    - Use of PCA, hierarchical clustering, or other novel methods.
    - Unique insights that go beyond the basic analysis.

Additional Considerations

- **Bonus Points (Up to 5 pts):** Creativity in approach, use of additional models or techniques.
- **Penalty (Up to -5 pts):** Poor documentation, improper formatting, lack of explanations.

NOTE: There is not going to be any elimination at the end of any level. The entire competition is going to be evaluated.

Participants will have to share their entire Jupyter Notebook code and file on GOOGLE DRIVER and the link shared at GOOGLE SITES of their own.

Participants are also advised to share the screen-shots of their Snippets and the graphs, maps and other visualization techniques.`