Assignment Instructions

Presentation

- Presentation is key. Ensure that your notebook is capable of explaining your insights and visualizations by itself. Do not rely on the report or the video to explain what is happening in your notebook. Section your questions and emphasize your results. **Do not** hide your final result in a sea of code or debugging cells.

  Examples:
  - If your question is on data cleaning, highlight the rows which need to be cleaned and show the results of your data cleaning before and after it has been applied on those rows.
  - If your question asks you to prove a statement using visualizations, ensure that you actually have a concluding statement after your graphs. **Do not** leave the conclusion unstated after visualizing the data in your notebook.

- It is recommended to have short bullet points explaining what you have done before each task, especially for non-visualization tasks. This will help us understand your approach to the problem and can help with partial marks even if you are unable to solve the entire question.

- Prioritise interpretability over design. While it is encouraged to have visually appealing graphs, make sure that you do not lose interpretability of the data in the pursuit of aesthetic visualizations. Refer to this link for some pointers on what not to do.

Insights

- The last section of your report will have to be dedicated to an out of the box pursuit. If you think you have a better way of cleaning the dataset or visualizing a question, or if you believe that you have noticed an interesting insight that can be gleaned from the data, add them at the end of your notebook and elaborate why you think you're right in your report or notebook and make sure you mention it in your recorded video. This carries weightage to your final scores.

Datasets and Questions

- You will be given randomized subsets of data from 3 sources, student performance, olympic medalist history and countrywise COVID information. Insights and correlations for each variation can be different and each variation is tracked during evaluation. Therefore, do not use any dataset except for the one assigned to you.

- Since the data has been shuffled, do not be surprised when the information in the dataset does not align with real world information.

- The hints provide important information that can help you with your approach to a question, read each question carefully and generate all the information it asks for.

- If you think a question is wrong or not solvable, please elaborate on why you think so in your notebook and your report, and if there is a noticeable issue with the question, drop a mail to **danish.ebadulla@gmail.com** or **hridhayks@gmail.com.**

Submission details

- Your assignment ID will be <Data source>_<DatasetID>. All submission files should be named <SRN>_<Assignment_ID> and submitted in a zip file with <SRN> as the name of the zip file/folder. Ensure that your SRN and assignment ID are also on the first page of your report and in a text cell in your notebook. **Assignment ID examples:** Student_31, Olympics_2, Covid_17