Assignment Questions

# Instructions:

There are 2 sets of questions listed below. Set-A is designed to test your Data Wrangling skills & Set-B is designed to test your visualization abilities.

For submission, you will need to share the following:

1. One Jupyter notebook containing the solution to the whole assignment
2. A pdf containing just the questions & solutions (Read: No Codes. Only the required outputs & explanations) – This should be obtained by exporting the jupyter notebook after hiding the code cells

Each question carries equal weightage (5 marks). Your solution will be graded on the following parameters:

1. Accuracy
2. Effectiveness of the explanation provided
3. Effort placed in beautifying the outputs (Statistics, Tables, visualizations) & making the result look professional

Some questions will have special instructions/requirements, highlighted in *italics.*

4 questions will have a HOTS (Higher order thinking skills) component. These are optional (for your knowledge) and will not be graded. However, if you complete ¾ of these, there will be a special reward for you.

# Set-A

Answer the following questions

1. Is the data clean? (*If not, clean it using best judgement before answering any of the following)*
2. What proportion of Pokémon have multiple types?
   1. What is the most common combination of types that Pokémon have?
3. How many mega Pokémon are there in the dataset (Identified by a prefix "Mega" in the name)?
4. List the top 5 Pokémon based on
   1. Highest stats (sum of all stats (HP--Speed))
   2. Number of battles fought
   3. Number of battles won
   4. Highest Win/Loss ratios (5 for each)
5. Are certain types naturally stronger than others, i.e.,
   1. Do certain Pokémon types have a higher percentage of wins?
   2. Do certain types have a higher value for certain stats?
6. For each Pokémon type, list the top 2 competitor types based on the number of battles that ended in victory or defeat (2 each for defeat and victory)
7. How are legendary Pokémon different from others?
   1. Do they have higher stats than other Pokémon? If so, which stats?
   2. Do they have a higher win percentage?
      1. How many Pokémon battles have legendary Pokémon lost?
8. Are Pokémon named in a totally random manner - Are certain alphabets more common in certain types of Pokémon? *(Ignore Pokémon having multiple types)*
   1. *(This is a HOTS question. Attempt if you want the challenge)* It is often said that Pokémon names are based on nature. For example, "Weedle" is a bug Pokémon (Not the presence of the word "Weed"). Across each type, find the Pokémon that have sub words present in the Oxford Dictionary (Word should have a size of at least 3 characters), along with the words present

# Set-B

Create visualizations to answer each of the following questions. Additional requirements in the viz are *italicized*.

1. How many Pokémon are present across different types (A Pokémon can have multiple types - type 1 & 2. Account for both of them)? *(Bar charts)*
2. Are dragon Pokémon generally stronger than others? (Compare across HP, Offense (Avg of Attack & Sp Atk), Defense (Avg of Defense & SP Def) and Speed) - *Create a BoxPlot with lines to highlight the mean, 20th & 80th percentile*
   1. *(Optional - HOTS) Make this process interactive using 2 drop-down boxes, for types. I should be able to select any 2 types and compare them across the mentioned stats*
3. Is the composition of Pokémon of different types changing significantly across generations? *(Area Curve)*
4. Which Pokémon types most commonly engage in battles with each other? *(Correlation plots)*
5. Compare the distribution of each of the stats (HP, Offense, Defense, Speed) between Generations 1,2,3. *(A grid of 4 charts (One for each stat). Each chart has 3 KDE(or Histogram) Plots (one for each generation))*
   1. *HOTS - Add mean, 75th percentile lines for each generation, color and label them accordingly and include them in the legend*