**Roller Coaster Project**

There are several Roller Coaster rating/ranking sites *online* that, while taking some *objective* measures into account, heavily rely on *subjective* input to determine the rating or ranking of a particular roller coaster (e.g., an “excitement” or “experience” score of an “expert” rider to measure “thrill”).

In addressing this problem, consider only roller coasters currently in operation. We have provided data for a subset of operating roller coasters whose height, speed, and/or drop are above the average of worldwide operating coasters. Therefore, we have not included family or kiddie coasters, nor have we included bobsled or mountain type coasters.

Questions to consider as you use pandas to perform analysis on this dataset:

1. How did you clean the data? What columns/rows did you keep? Which ones did you get rid of? Why? What missing data did you fill in? How? Why? You must deal with missing data thoughtfully (you may decide that some columns do not contain enough data to use and then drop them from the DataFrame entirely, but you can’t drop all of the missing data because then your dataset won’t be large enough to analyze.
2. What interesting observations did you find in the data? (For example: What is the oldest rollercoaster?)
3. What graphs did you find by plotting numerical variables? Any interesting scatterplots or histograms? Do any of the scatterplots warrant a line of best fit? If so, can you create one and interpret it?
4. Use datetime capabilities to create a duration column that contains the duration in total seconds and sort by duration. What are the longest and shortest durations?
5. Can you make any sense out of the CoasterBuzz rankings? If so, what factors do they seem to be weighting highly? If not, why might you be having trouble finding a pattern? <https://coasterbuzz.com/RollerCoasters/Top100>
6. Can you come up with your own ranking system? If you sort according to your own ranking, what are your top 10 roller coasters? How do they compare to the CoasterBuzz rankings?

Submit your findings in a single Jupyter Notebook. Your explanations should be written in Markdown and all of your code cells should run without errors. Please organize the notebook so that it is easy to read and follow.