

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.

1. Create an objective quantitative algorithm or set of algorithms to develop a descriptive roller coaster rating/ranking system based only on roller coaster numerical and descriptive specification data (e.g., speed, duration of ride, steel or wood, drop).
2. Use your algorithm(s) to develop your “Top 10 Roller Coasters in the World” list. Compare and discuss the rating/ranking results and descriptions from your team’s algorithm(s) with at least two other rating/ranking systems found online.
3. Describe the concept and design for a user-friendly *app* that uses your algorithm(s) to help a potential roller coaster rider find a roller coaster that she or he would want to ride. NOTE: You DO NOT need to program and/or write code for the app. You are developing the concept and design for the app only.

Non-technical requirements: Create a short (5-8 minutes) and well-rehearsed slide show presentation describing your new algorithm, results, and app. Your audience is the editorial board for *Rollercoaster Magazine*.

Technical requirements:

- Your code should be linted along with markdown cells that describe what each Jupyter Notebook cell is doing and describing your algorithm.
- Use datetime capabilities to convert duration into seconds.
- 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. Consider using linear regression or other techniques to fill in missing data.)