Roller Coaster Ranking System

Team 8595

Summary

Roller Coasters have been ranked by many different organizations in the past, each giving a different list of which ones are the “best” roller coasters. The task was to make a ranking system, which primarily used objective data rather than the many subjective lists that already exist.

The first approach taken was to clean up the data provided. For method 1, data points that were not given were replaced with the average value. The next step was to quantify relative numerical indices for each given data type so that a ranking could be made by placing these values or a function including these values in a certain order. For the first method, it was decided that the means of the data types would be used as a “normal” value to which each data point could be compared. A scaled product was used to compute an aggregate index across data types and to get a final ranking. Another method utilized was filling in the missing data by making data models which relate different data types. If the data models fit the data set well, the model was used to fill in data where possible.

In the second method, the relative indices were based on medians as a “normal” value of the data types. To scale these indices, the differences were scaled down to the range [1, 2], and afterwards, scaled products were used to get a final ranking.

To compare the Top 10 lists that were found online to the Top 10 calculated from method 2, the median of differences in rankings was used to determine how similar the lists were.
When considering what to include in the app to receive the subjective data and combine it with the objective method 2, relative weights were added so that data types could be placed in order of importance. The user would have a chance to rank their favorite data types and from those rankings, weights would be derived. Location was used in the app to determine which roller coasters would be feasible options.

In conclusion, the final list created was a strong model for a roller coaster ranking system as it showed a strong relation with current subjective lists ranked by a sample of people.

News Release

HiMCM - News Release

November 16th, 2018

Team #8595 Roller Coaster Algorithm

Algorithm:

We created an algorithm which ranks the best roller coasters in the world, considering the age, the max height, the max speed, the length, and the drop angle of the roller coaster. We used a private weighting of the different aspects, with the max speed, height, and length of ride given high consideration, and the age and angle of the coaster given a lower consideration.

Results:

Top 10 Roller Coasters in the World (based on the algorithm)

1. Kingda Ka
2. Top Thrill Dragster
3. Fury 325
4. Steel Dragon 2000
5. Red Force
6. Leviathan
7. Millennium Force
8. Intimidator 305
9. Superman: Escape from Krypton
10. Hyperion

App:

Our mobile application generates a personalized list of the top ten roller coasters that match the preferences of the user. The way the app works is by first displaying five designated variables: roller coaster speed, height, length, date of roller coaster construction, and vertical angle. The user is now prompted to use the slider, of a scale between 1 – 10, for each named variable. This slider essentially allows the user to submit their personal level of importance given to each variable, with a 1 suggesting a low level of importance as well as a 10 suggesting a high level of importance. Once all input is accounted for, the app will replace the scale variables in our coded algorithm with values based off the survey data in the app. In addition, the app will allow the user to breakdown their newly formed Top 10 roller coaster ranking list based on feasibility. This will be completed by asking the user to enter a location or location radius. This will eliminate unfeasible options. After the algorithm has finished being computed, a Top 10 list will be displayed solely on the user's variable preferences.

Table of Contents

The Problem

Today, there are several online resources that have rankings and ratings of roller coasters in the United States and throughout the rest of the world. These ranking and rating systems take some objective measurements into account such as max speed, height, drop height, duration, year of opening, or length. However, most of the rankings are based off subjective input from reviewers and critics of roller coasters, who give each roller coaster a rank or level of excitement, experience, or thrill.

The provided data spreadsheet includes objective data on three hundred different roller coasters in the United States and the rest of the world. All of these roller coasters are in operation and several of their statistics are given in an organized table.

Goal:
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.

Assumptions

Assumptions Made

- Assumed that the following were important factors when evaluating roller coasters
- Year
- Speed
- Duration
- Inversion
- Construction
- Assumed that roller coaster ranking is solely determined by numerical roller coaster specifications themselves
- External subjective factors such as location preference will not be factored into the ranking system
- The extremes of age of roller coaster, that being oldest and newest, are the most sought after ranges. (There is an innate desire for the oldest and newest roller coasters, which increases their ranking)
- Height is highest height, drop is the length that the coaster travels during its drop
- Vertical angle is the angle below the horizon starting at the max height
- The given type of the roller coaster was not factored into the rankings

Assumptions Given

- In addressing this HiMCM 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.

Method 1

Process and Results

Four factors, including speed, height, length, and age of roller coaster, were considered in the first ranking system. Data for each of these factors was imported into Mathematica and a list was created with each element consisting of the name of the roller coaster and its respective speed, height, length, and age. The average of each ranking factor was calculated in order to get an accurate representation of all the data in that specific factor. Each ranking factor in each roller coaster was divided by the average of that ranking factor in order to get a "ranking" of each roller coaster's ranking factors with respect to the average value. For each roller coaster, all of its individual ranking factors were multiplied together to receive one, unscaled value that represents the rank of the ride with respect to all of the rides. At this point, the total compiled list consists of every single roller coaster matched with its respective unscaled ranking as calculated above. However, some of the unscaled ranks include null values because for some of the roller coasters, there is missing data. In order to deal with this problem, the roller coasters that had null values or missing values (9 roller coasters in total) were eliminated from the overall list, yielding 291 roller coasters in the list with full data and a calculated unscaled ranking. In order to scale each preliminary ranking into a whole number that is in the range from 1-10, the maximum ranking was identified, and all other rankings were divided by that value. Then, all the values were multiplied by ten in order to scale the final rankings into the range from 1-10.

Out[923]=

1	Steel Dragon 2000	10.
2	Fury 325	8.35037
3	Millennium Force	7.73104
4	Kingda Ka	7.41846
5	Formula Rossa	6.82052
6	Leviathan	6.31744
7	Top Thrill Dragster	5.74669
8	Intimidator 305	5.7208
9	Fujiyama	5.70156
10	Red Force	4.86176
11	Titan	4.50027
12	Shambhala	4.35884
13	Hyperion	4.34833
14	Coaster Through the Clouds	4.29279
15	Silver Star	4.08788
16	Nitro	4.03761
17	Desperado	3.98231
18	Diamondback	3.96958
19	Behemoth	3.84485
20	Intimidator	3.77988
21	Superman el Último Escape	3.72601
22	Goliath	3.65492
23	Steel Vengeance	3.57246
24	Superman the Ride	3.51663
25	Big One	3.51245

Strengths and Weaknesses

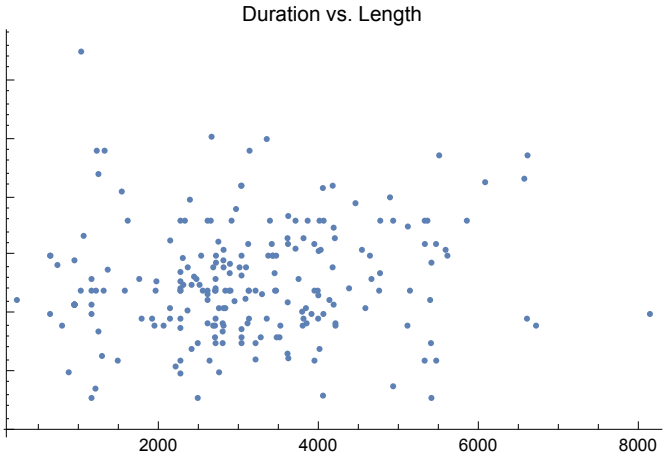
- Method one only uses four ranking factors and omits others such as g-force and duration. This is a weakness because the algorithm is not necessarily the best representation of all the data of all the roller coasters.
- Method one is very versatile because the data can easily be modified for other roller coasters and the algorithm would still return a desired ranking system.
- Method one was completely objective as it was solely based on the given data.
    - The scaling system used was not subjective because the rankings were manipulated such that they were in a range from 0-10 without weighting the rankings.

Method 2

Calculating Duration

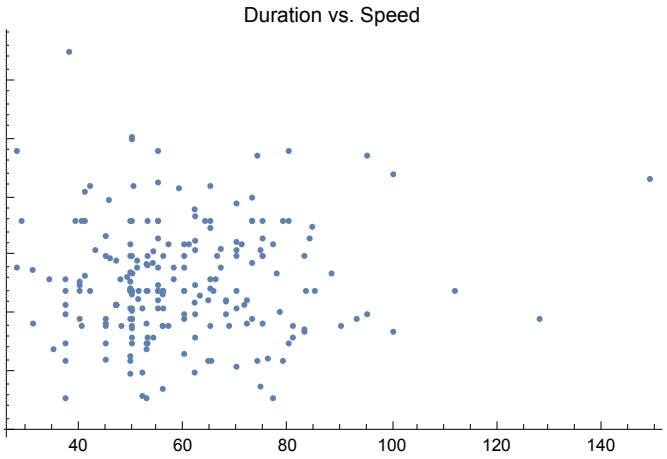
In Method 1, only 4 factors were used. Other factors were included in Model 2 to make the model stronger. In the given data set, not all of the values were given for duration. To fill in the missing data points, an attempt was made to find a correlation between duration and other factors such as speed, length, and height. The intent was to create a line of best fit through a scatter plot in order to solve for the missing data points in duration. The scatter plot of Duration vs. Length did not show any relationship.

Out[1043]=



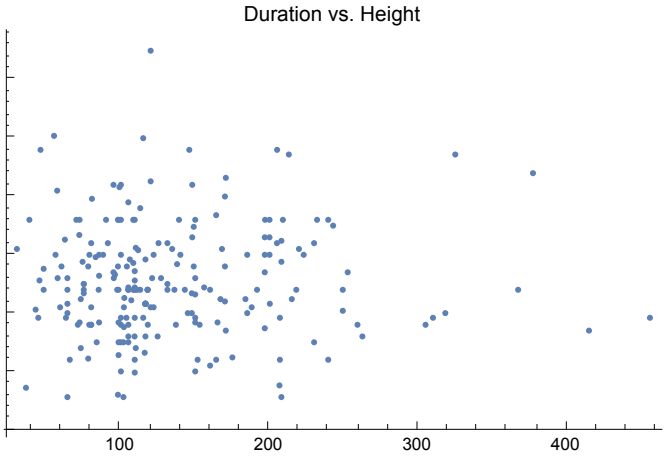
Since the Duration vs. Length did not show any relationship, a scatter plot of Duration vs. Speed was created. This plot did not show any relationship as well.

Out[1045]=



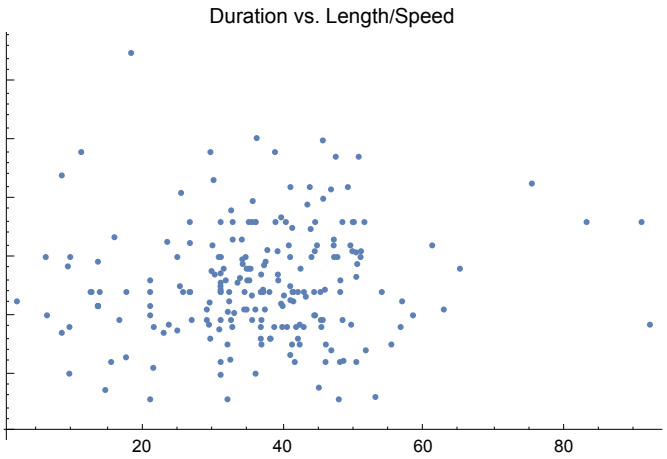
Since the Duration vs. Speed did not show any relationship too, a scatter plot of Duration vs. Height was created in the hopes of some relationship between them. This plot did not show any relationship as well.

Out[1047]=



No relationship was determined in the plot of Duration vs. Height. By dividing Length by Speed, a value would be found in seconds. So, there would be a better chance of a correlation between Duration vs. Length/Speed. A plot was created for Duration vs. Length/Speed.

Out[1049]=

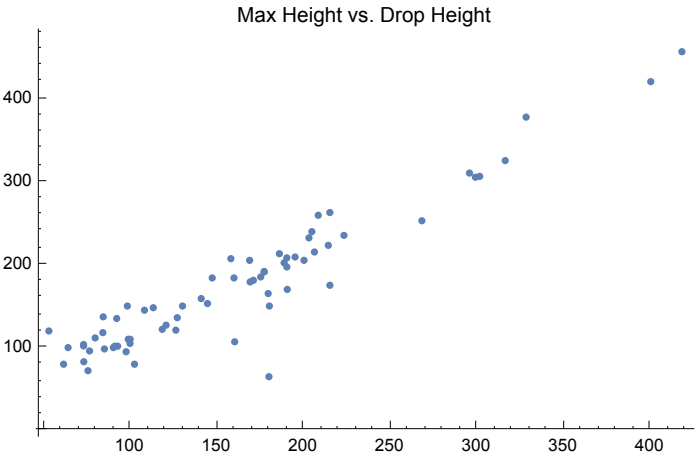


As there was no relationship between Duration and any of the factors that was tested, the use of Duration in the formula was determined to not be used. If the Duration was used, then the new data calculated to fill in the missing points would not be accurate, which would skew the final result.

Angle - Drop Length - Drop Height

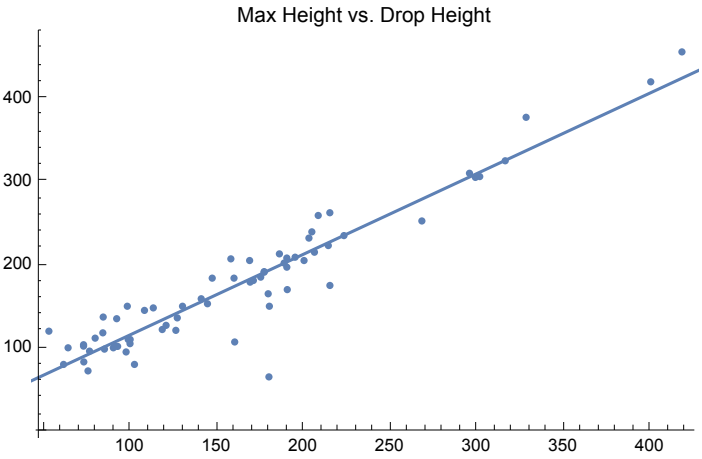
In the attempt to fill in missing data, an idea of correlating Vertical Angle, Drop Length, and Max Height was examined. By multiplying the length of the drop by the sine of the vertical angle, an approximate value of the height of the drop can be calculated. On typical roller coasters, the drop height is a little bit smaller than the max height, so there will be a linear correlation with the slope equal to about one. Therefore, a scatterplot was created between the max height and the drop height.

Out[1055]=



Using the data in the scatterplot, a line of best fit was found (Least-Squares Line). The equation of this line was y = 18.8448 + 0.97012x where y represents the max height and x is the drop height. Using this line of best fit, 98 additional values were added in place of the missing data (74 data points for the vertical angle and 24 data points for the max height).

Out[1059]=



Scale

This is a scale from 1-10 for the weights of the factors. V represents the weight of the speed, H for height, Y for years, L for lengths, and A for angle. The value after the decimal point represents the scale.

In[1060]:=
```
V = 1.9;
H = 1.8;
Y = 1.1;
L = 1.8;
A = 1.4;
```

Calculating Points

In this method, the medians of each aspect of the roller coasters was used to create an index of how good a certain value is compared to the provided values. A strength of this method is that it reduces any effect that outliers in the data may have because the median is a resistant value of the dataset. Each value in the dataset was divided by the median of the dataset in order to scale it down to values around 1. After this, the indices were scaled to the exact range [1, 2]. These indices were multiplied together to get the ranking score. Finally, for display, we scaled these scores to a maximum score of 10.

Out[1097]=

1	Kingda Ka	10.
2	Top Thrill Dragster	7.47994
3	Fury 325	7.0439
4	Steel Dragon 2000	6.80564
5	Red Force	6.43357
6	Millennium Force	5.26544
7	Leviathan	5.09352
8	Intimidator 305	4.47404
9	Superman: Escape from Krypton	3.67092
10	Hyperion	3.44313
11	Tower of Terror II	3.37022
12	Coaster Through the Clouds	3.08295
13	Fujiyama	2.98151
14	Shambhala	2.96231
15	Titan	2.70134
16	Desperado	2.13868
17	Behemoth	2.13595
18	Diamondback	2.08476
19	Silver Star	1.99473
20	Schwur des Kärnan	1.96668
21	Intimidator	1.96593
22	Nitro	1.95343
23	Steel Vengeance	1.94713
24	Superman the Ride	1.86895
25	Big One	1.81117

Strengths and Weaknesses

- Method two includes all the ranking factors used previously in method one and also vertical angle, making the ranking system more optimal and representative than the ranking system of method one.
- Method two is more optimal because a weighting system was used in order modify each ranking factor's weight such that factors that are more desired than others have a larger effect on the overall ranking.
  - This made the ranking system more similar to other ranking systems from outside sources.
  - However, this also made the ranking system somewhat subjective because it attempted to emulate human preference without using numerical data.
- The data set given had many missing data points forcing some of the roller coaster data to be removed from the overall algorithm.

Comparisons of the Rankings

Below is a table with rankings from the results in method 1 and method 2. Method 2 was used for the final ranking because it accounted for more factors.

Out[1104]=

No.	Method 1 Rankings	Method 1 Points	Method 2 Rankings	Method 2 Points
1	Steel Dragon 2000	10.	Kingda Ka	10.
2	Fury 325	8.35037	Top Thrill Dragster	7.47994
3	Millennium Force	7.73104	Fury 325	7.0439
4	Kingda Ka	7.41846	Steel Dragon 2000	6.80564
5	Formula Rossa	6.82052	Red Force	6.43357
6	Leviathan	6.31744	Millennium Force	5.26544
7	Top Thrill Dragster	5.74669	Leviathan	5.09352
8	Intimidator 305	5.7208	Intimidator 305	4.47404
9	Fujiyama	5.70156	Superman: Escape from Krypton	3.67092
10	Red Force	4.86176	Hyperion	3.44313
11	Titan	4.50027	Tower of Terror II	3.37022
12	Shambhala	4.35884	Coaster Through the Clouds	3.08295
13	Hyperion	4.34833	Fujiyama	2.98151
14	Coaster Through the Clouds	4.29279	Shambhala	2.96231
15	Silver Star	4.08788	Titan	2.70134
16	Nitro	4.03761	Desperado	2.13868
17	Desperado	3.98231	Behemoth	2.13595
18	Diamondback	3.96958	Diamondback	2.08476
19	Behemoth	3.84485	Silver Star	1.99473
20	Intimidator	3.77988	Schwur des Kärnan	1.96668
21	Superman el Último Escape	3.72601	Intimidator	1.96593
22	Goliath	3.65492	Nitro	1.95343
23	Steel Vengeance	3.57246	Steel Vengeance	1.94713
24	Superman the Ride	3.51663	Superman the Ride	1.86895
25	Big One	3.51245	Big One	1.81117

In order to test the results of the final list that we created, the list was compared to a the list of the Worlds Best Roller Coasters on Ranker.com. The top ten roller coasters on Ranker.com were compared to the ranking of the same roller coaster in the final list. The absolute value of the differences were taken in order to see how close the ranking in the final list were to a list created by peoples opinion. This showed how accurately the final list represented if a sample of people would like or dislike a roller coaster.

Out[1111]=

Name of Roller Coaster	Ranker.com Rankings	Our Rankings
Millennium Force	1	5
Steel Vengence	2	22
Top Thrill Dragster	3	2
Maverick	4	99
El Toro	5	51
Fury 325	6	3
Intimidator 305	7	8
Kingda Ka	8	1
Apollo's Chariot	9	44
Nitro	10	23

The median difference between the Ranker.com rankings and the Final rankings was 10. This shows that the strength of the model is strong as it closely represents a ranking by a sample of the population.

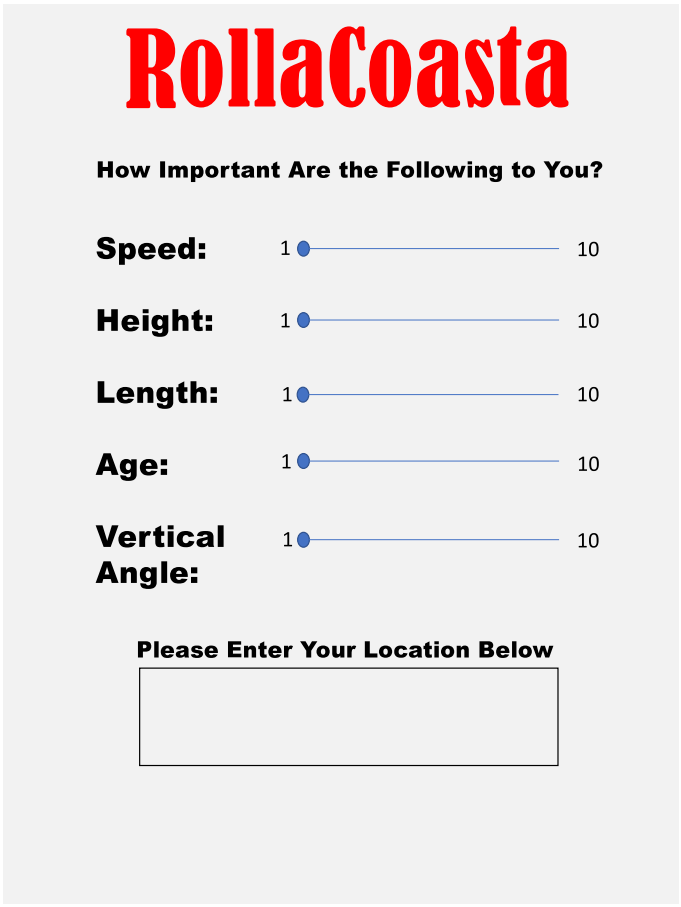
The final rankings were also compared to a ranking by Life'd. The median difference between the Life'd rankings and the Final rankings was 25.5. This was not as small as the difference between the Ranker.-com rankings and the Final rankings, but it is still close, showing that the model is strong.

Out[1119]=

Name of Roller Coaster	Life'd Rankings	Final Rankings
Superman the Ride	1	24
Millenium Force	2	7
El Toro	3	51
Expedition GeForce	4	62
The Voyage	5	43
Kingda Ka	6	1
Intimidator 305	7	8
Goliath	8	36
Behemoth	9	17
Nemesis	10	148

Application

The mobile application generates a personalized list of the top ten roller coasters that match the preferences of the user. The way the app works is by first displaying five designated variables: roller coaster speed, height, length, date of roller coaster construction, and vertical angle. The user is now prompted to use the slider, of a scale between 1 – 10, for each named variable. This slider essentially allows the user to submit their personal level of importance given to each variable, with a 1 suggesting a low level of importance as well as a 10 suggesting a high level of importance. Once all input is accounted for, the app will replace the scale variables in the coded algorithm with values based off the survey data in the app. In addition, the app will allow the user to breakdown their newly formed Top 10 roller coaster ranking list based on feasibility. This will be completed by asking the user to enter a location or location radius. This will eliminate unfeasible options. After the algorithm has finished being computed, a Top 10 list will be displayed solely on the user's variable preferences.



References

Annenberg Learner. (2016). Retrieved November 16, 2018, from https://www.learner.org/exhibits/park-physics/coaster.html

Top 100 Roller Coasters: The CoasterBuzz 100. (2018). Retrieved November 16, 2018, from https://coast-erbuzz.com/RollerCoasters/Top100

The Best Roller Coasters in the World. (2018). Retrieved November 16, 2018, from https://www.ranker.-com/crowdranked-list/best-roller-coasters

Top 25: Best Roller Coasters (in the World). (2012, July 27). Retrieved November 16, 2018, from http-s://www.lifed.com/top-25-best-roller-coasters-in-the-world