

For office use only
 T1 _____
 T2 _____
 T3 _____
 T4 _____

Team Control Number
 8744

For office use only
 F1 _____
 F2 _____
 F3 _____
 F4 _____

Problem Chosen
A

**2018
 HiMCM**

Summary Sheet

To many of the world’s roller coaster enthusiasts, a recurrent problem would be the decision of the next ride to experience. Indeed, as there are almost 4000 roller coasters across the globe, exactly which one should riders choose to go? To answer that question, we decided to create an algorithm that rates roller coasters objectively based on their statistical data, creating a comparable index and a ranking, effectively letting the roller coasters “speak for themselves.”

Before we begin modeling, we analyzed the dataset in the provided spreadsheet and found out that there are large numbers of missing, noisy, or inconsistent values in the data, and the variables in the dataset are not in the same scale. To solve this, we first normalized all factors using the Box-Cox Transformation and then standardized them into a unified range of [0,1]. Then, we use mean imputation and regression imputation to fill in the missing values and filter out the noise.

The first part of our model establishes the concept of a Thrill Index. We propose that the more thrilling the roller coaster is, the more people will like it, and thus should be ranked higher. We constructed this part of the model by aggregating the factors that may contribute the experience of thrill during a ride: the height, maximum speed, G-force, inversions, and type of roller coaster.

The second part of our model establishes the concept of a Discomfort Index. We propose that the more discomfort the riders feel during or after a ride, the fewer people will like the experience, and thus the roller coaster should be ranked lower. This part of the model is constructed around factors that may contribute to the discomfort of the rider: G-force and inversions. We used logistic functions to simulate people’s discomfort level caused by these two factors since they have non-linear relationships.

We think both the Thrill Index and the Discomfort Index are related to the overall experience of a rider, but none of them is comprehensive enough. Therefore, the third part of our model establishes the concept of a Comprehensive Index, which combines the effects of the Thrill Index and the Discomfort Index. The result of the Comprehensive Model is shown in the table below.

Rank	1	2	3	4	5	6	7	8	9	10
Name	Steel Dragon 2000	Fury 325	Millennium Force	Fujiyama	Leviathan	Formula Rossa	Desperado	Intimidator 305	Titan	Steel Vengeance

We also design an app named “Roller Ranker” for our model so that users can obtain our recommendations of highest-ranked roller coasters directly on their phone. In our app, users can view the default ranking, which is the one we ranked according to the Comprehensive Index, or they can customize settings concerning factors that may appear as more important to them, and our model will generate their personal recommendations using an adapted version of AHP (Analytic Hierarchy Process). Also, to improve the accuracy of the default rankings, we conduct a survey online about people’s preferences over different factors and add weight to the factors of the Comprehensive Model.

Press Release

Team 8744 releases a new algorithm, ranking and mobile app for thrill seekers worldwide.

(Foshan, Guangdong, China – November 13, 2018) Team 8744 in the HiMCM Competition today announced the availability of their new objectivity-promising mobile application, Roller Ranker, to users of the Android and iOS platforms through digital marketplaces such as Google Play and App Store. By now, the utility application has gained comprehensive data on many of the world's most popular roller coasters and would continue to expand in time, according to the spokesman of Team 8744, demonstrating a constant commitment to provide knowledge to rookies and enthusiasts alike. The initiative is part of Team 8744's mobile project, which focuses on reaching the millions of riders around the world whose primary method of Internet access is via a mobile device.

“Our team strives to remove bias in traditional methods of ranking determination, and for our customers around the world right now, having to refer to subjective experience and lack of factual science are the two major complications,” says Hunter Zhang, CEO of Roller Ranker. “We created a comprehensive index for those looking to compare real information and enhance their riding experience, minimizing the time wasted on decisions and enabling the pursuit of the most efficient solutions.”

Based on a large 300-entry dataset, Team 8744 determined the thrill, discomfort and overall experience riding on each of the individual roller coasters with statistical analysis. Each coaster was then rated against its competitors and ranked into a long list, presented in the application beside customization options for the user.

“We are delighted to bring our followers our latest achievement – the Ranker app that offers vital services with a significant outreach,” says Edison Chen, Chief Writer and Head of Technology. “With the app completed, we could conclude the first stage of our work and start the provisioning of Data-as-a-Service on demand to users regardless of their geographic location, bringing a vast organized knowledge source to underserved communities.”

The mobile application, in its design, would provide options for the user to either leave decisions to the core algorithms or provide their own preferences on roller coaster types, height, speed and so on.

For more information on Team 8744's Roller Rider and related algorithms, please visit their website, follow them on LinkedIn, YouTube and Facebook.

Content

1 - INTRODUCTION	4
1.1 BACKGROUND	4
1.2 PROBLEM RESTATEMENT	4
2 - ASSUMPTIONS AND VARIABLES	4
2.1 ASSUMPTIONS AND JUSTIFICATIONS.....	4
2.2 VARIABLES.....	5
3 - DATA ANALYSIS, AND MISSING VALUE IMPUTATION	6
3.1 DATA DESCRIPTION	6
3.2 CORRELATION ANALYSIS	7
3.3 MISSING VALUE IMPUTATION.....	8
4 - MODEL PART 1: DATA STANDARDIZATION	9
5 - MODEL PART 2: THRILL INDEX	10
5.1 INCORPORATING THE FEELING OF WEIGHTLESSNESS.....	10
5.2 INCORPORATING MAXIMUM SPEED	11
5.3 INCORPORATING ACCELERATION	12
5.4 INCORPORATING INVERSIONS	12
5.5 INCORPORATING THE TYPE OF ROLLER COASTERS.....	12
6 - MODEL PART 3: DISCOMFORT INDEX	14
6.1 INCORPORATING G FORCE	14
6.2 INCORPORATING INVERSION FREQUENCY	15
6.3 DISCOMFORT INDEX	16
7 - MODEL PART 4: COMPREHENSIVE INDEX	17
8 - APP: ROLLER RANKER	18
8.1 GOALS AND CONCEPTS OF THE APP	18
8.2 POTENTIAL USERS AND APP DESIGN	19
8.3 INCORPORATING CUSTOMIZED SETTINGS INTO THE COMPREHENSIVE MODEL	20
STEP 1: PREPARING THE DATA	21
STEP 2: CHECKING FOR INCONSISTENCIES	22
STEP 3: CALCULATING THE WEIGHT VECTOR	24
STEP 4: OUTPUT	25
9 - ADDING WEIGHTS TO THE COMPREHENSIVE MODEL	25
10 - RESULTS	27
10.1 COMPARISON WITH OTHER RANKINGS SYSTEMS	27
11 - EVALUATION	29
12 - REFERENCES	30
APPENDIX	31
APPENDIX A: THE PROTOTYPE OF OUR APPLICATION	31
APPENDIX B: CODE.....	32
APPENDIX C: OUR DATA SPREADSHEET	36
APPENDIX D: SURVEY RESULTS FOR PART 9.....	39

1 - Introduction

1.1 Background

Roller coasters are becoming one of the most symbolic landmarks in the eyes of visitors. The number of visitors to amusement parks and the number of roller coasters built every year arrives at unprecedented peaks, and many new coasters break records that those a decade ago would never even have a chance of achieving. Meanwhile, some enthusiasts miss the classic wooden feeling that differs from modern steel coasters, some challenge themselves to high drops and speeds, and seeking for the best personal experience is increasingly essential for visitors and potential riders. Current rankings about roller coasters are primarily based on subjective input with limited reference of objective input such as inversions and peak speed, which may lead to new visitors gaining biased conclusions.

Therefore, for the aforementioned new visitors' sake, we are developing a ranking system that looks at operational roller coasters worldwide in an objective stance in order to give a credible score and ranking based on calculations upon the roller coasters' data.

With the results, we evaluate the feasibility of our model, analyze its objectivity and statistical validity, and we demonstrate our model and the app in a news release for publishers worldwide, establishing a benchmark for future reviews and recommendations.

1.2 Problem Restatement

The inherent challenge in providing an objective ranking of roller coasters worldwide lies analyzing all data about all roller coasters. As current rankings rely heavily on subjective measurements, we have to differ from them and develop a mathematical model for an objective ranking system. This model would have to consider the most important factors that contribute to the experience of riding roller coasters, and these should all be present in the given dataset. The model then should be compared with current rankings. Finally, we must design a user-friendly app that helps people find roller coasters that they prefer to ride and compose a news release to highlight our model and app.

2 - Assumptions and Variables

2.1 Assumptions and Justifications

- **Assumption 1:** External factors that are not related to the roller coaster is ignored when different roller coasters are compared.
- **Justification:** Though the external factors (weather, infrastructure of the amusement park) may affect riders' overall experience, they do not reflect the experience of the ride itself.

- **Assumption 2:** The imputed data approximates the real data of the roller coasters.
- **Justification:** There are a lot of missing data in the provided dataset, and we use imputation to fill them in. We made this assumption because without the full data, valid comparison between roller coasters cannot be made.

- **Assumption 3:** People may feel discomfort after a ride.
- **Justification:** Although people come to ride roller coasters to experience the thrill, if a ride is too extreme, the person might feel sick. This is because roller roasters simulate an environment that people are not used to be in.

- **Assumption 4:** People start to feel discomfort when G-force approach 5 and inversion frequency approaches 0.133 inversions per second.
- **Justification:** These values are selected based on existing studies which can be found in the references.
- **Assumption 5:** The database of our application's server is well protected so that the integrity of our data is preserved.
- **Justification:** We assume only we can change the data in the server, so that no malicious actions are performed against our data.

2.2 Variables

We use not capitalized Latin or Greek letters for denoting factors of a roller coaster. It is important to note that all of these variables have been transformed into the same scale, which is explained in Part 4.

l = the length of a roller coaster.

δ = the drop of a roller coaster.

t = the duration of a ride of a roller coaster.

n_I = the number of inversions of a roller coaster.

ω = the inversion frequency of a roller coaster.

b_T = the numeric score deduced by the Borda point of the type of a roller coaster.

h_{max} = the maximum height of a roller coaster.

s_{max} = the maximum speed of a roller coaster.

g_{max} = the maximum magnitude of G Force of a roller coaster. Note that G Force is actually an acceleration.

We use bold, capitalized letter for denoting matrices.

A, **B**= examples of comparison matrix.

We use bold, not capitalized letters for denoting vectors.

0 = the zero vector.

g = the acceleration caused by the gravitational force on the earth surface.

w = the unadjusted weight vector of a user's comparison matrix.

w' = the adjusted weight vector of a user's comparison matrix.

w_c = the weight vector calculated from the survey.

We use capital letters for denoting indices.

TI = the Thrill Index.

DI = the Discomfort Index,

where DI_g = the discomfort index contributed by G-force,

and DI_ω = the discomfort index contributed by inversion frequency.

C = the Comprehensive Index.

CI = the Consistency Index.

RI = the Random Index.

We define the set of terms we incorporate in our model's formula as \mathcal{F} , and we assign each term an index as shown in the table below:

Factor	h_{max}	s_{max}	g_{max}	ω	b_T	$-DI_g$	$-DI_\omega$
Index of it in \mathcal{F}	1	2	3	4	5	6	7

3 - Data Analysis, and Missing Value Imputation

3.1 Data Description

The dataset that we will use throughout this text to develop our models is a subset of operational roller coasters whose height, speed, and/or drop are above the average of worldwide operating coasters. There are 300 entries in the dataset, each with thirteen (13) factors. As shown in table 3.1, there are 9 numerical factors, 3 categorical factors and one binary factor. Also listed are the unit or values each factor is calculated in, if applicable.

Numerical Factors	Year Opened Height (in feet) Speed (in mph) Length (in feet) Number of Inversions Drop (in feet) Duration (in minutes: seconds) G Force Vertical Angle (in degrees)
Categorical Factors	Construction Type Status
Binary Factor	Inversions (YES or NO)

Table 3.1: An overview of the variables in the dataset

We also calculated the mean, standard deviation, and other statistics of the numerical data to interpret the dataset better (missing values are not included in the calculation), the result of which is summarized in table 3.2.

	Year/Date Opened	Height (feet)	Speed (mph)	Length (feet)	Number of Inversions	Drop (feet)	Duration(sec)	G Force	Vertical Angle (degrees)
# of Observations	300.00	299.00	296.00	295.00	300.00	141.00	223.00	82.0	91.00
Mean	2000.66	135.52	59.68	3149.94	2.22	153.18	126.87	4.3	74.74
Standard Deviation	13.16	66.40	16.25	1454.48	2.60	73.85	47.35	0.6	17.94
Minimum	1924.00	28.96	28.00	215.00	0.00	27.00	28.00	2.8	45.00
25%	1996.00	98.00	49.70	2260.50	0.00	95.00	96.00	4.0	60.00
50%	2002.00	116.50	55.90	3024.80	1.00	144.00	120.00	4.3	77.00
75%	2009.00	169.00	70.00	4008.80	4.00	205.00	154.50	4.7	90.00
Maximum	2018.00	456.00	149.10	8133.20	14.00	418.00	325.00	5.2	121.00

Table 3.2: The statistics of the numerical data

3.2 Correlation Analysis

Correlation analysis of the dataset is a very important step to do before the modeling because in the process of creating an objective model, no single factor of a roller coaster should be emphasized more than any other. However, we can predict that several factors in the dataset are highly correlated. For example, *height* and *drop* conveys the same aspect of a roller coaster: the feeling of weightlessness; *Duration* and *Length* may also be correlated, as well as the *Number of Inversions* and *G force*.

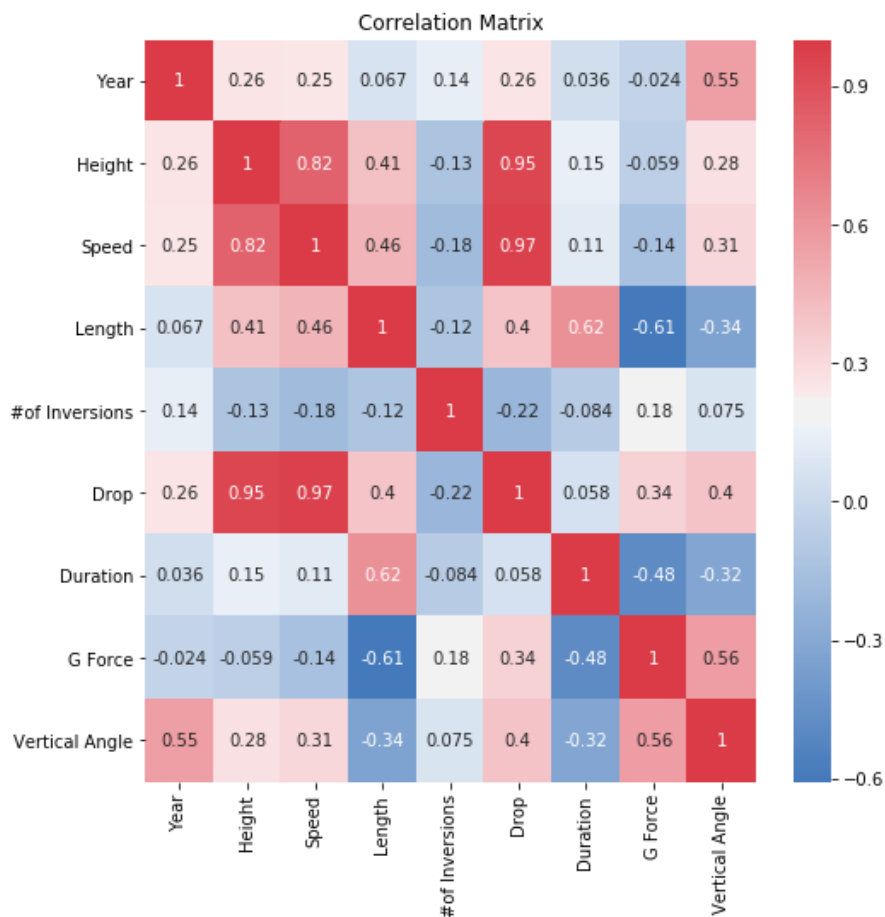


Figure 3.1 Correlation heatmap

To find out the exact correlation between factors, we plot a correlation heatmap using the Pearson correlation coefficient (r value) for the dataset, as shown in figure 3.1.

In the figure, it is shown that some of the factors in the dataset are highly correlated. For example, *Height* and *Drop* having a correlation coefficient of 0.95 and *Speed* and *Drop* having a correlation coefficient of 0.97 both indicates near perfect positive linear relationships. Also, *Length* and *G Force* exhibits a high possibility of an inversely proportional relationship. In the process of our analysis, prominent relationships of the graph are being considered in the modeling process, so all factors of a roller coaster are emphasized equally, preserving objectivity for our models.

3.3 Missing Value Imputation

There are a lot of missing data in our dataset, for example, 53% of the dataset is missing the *Drop* value, as shown in table 3.3.

	Year Opened	Height	Speed	Length	# of Inversions	Drop	Duration	G Force	Vertical Angle
% of Missing Values	0	0.3	1.3	1.67	0	53	25.67	72.67	69.67

Table 3.3: The table shows the percentage of missing values in the dataset

For factors such as *height*, *speed*, and *length*, which have a very low percentage of missing values, we used mean imputation to fill missing values.

However, for factors such as *drop*, *duration*, *G force*, and *vertical angles*, simple imputation with an arithmetic mean would not be appropriate because the percentage of missing values is too big that it will undermine the dataset.

Thus, we need another approach to recover these data. According to the correlation matrix shown in figure 3.1, the factors *height* and *drop* are highly correlated, with a correlation coefficient of 0.95; *Duration* and *length* also shows a strong correlation, with an r value of 0.62; *G force* is related to *length*, with an r value of -0.61; *Vertical angle* is correlated with *G force* as well, with the coefficient at 0.56. We plotted a linear regression plot between these factors to visualize this relationship. We determined that regression imputation, which approximates the missing data with other factors in the dataset, will be an effective approach to fill in the missing values.

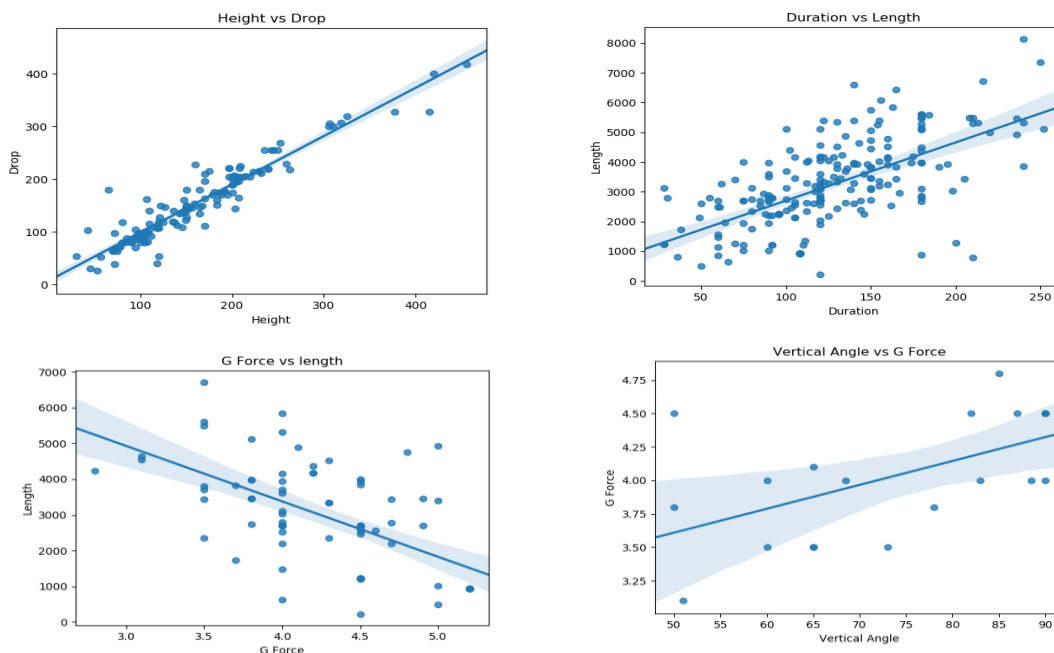


Figure 3.2: Linear regression plot for Height, Drop, Duration, Length, G force, and Vertical Angle

Since the factors with a high percentage of missing data are correlated with the factors with a small

percentage of missing data, we can impute the missing values using the least squares linear regression line, as shown in figure 3.2. With all the data filled, we can move on to our model.

4 - Model Part 1: Data Standardization

The dataset contains information of many factors; however, they are not in the same scale. In order to use the data in our model, it is crucial to give each value in the data a standardized score, so that they will be in the same range. To transform all factors into the same scale, it would be reasonable to standardize each factors according to its **percentile position** among other values of that factors.

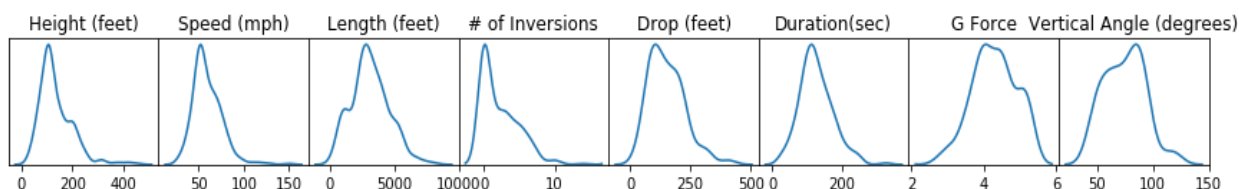


Figure 4.1: The distribution of eight numerical factors

However, the distributions of values of the numerical factors in the dataset are skewed, as shown in figure 4.1. To transform our skewed data into a normal distribution, Box-Cox transformation is applied. The effect of which is shown in figure 4.2.

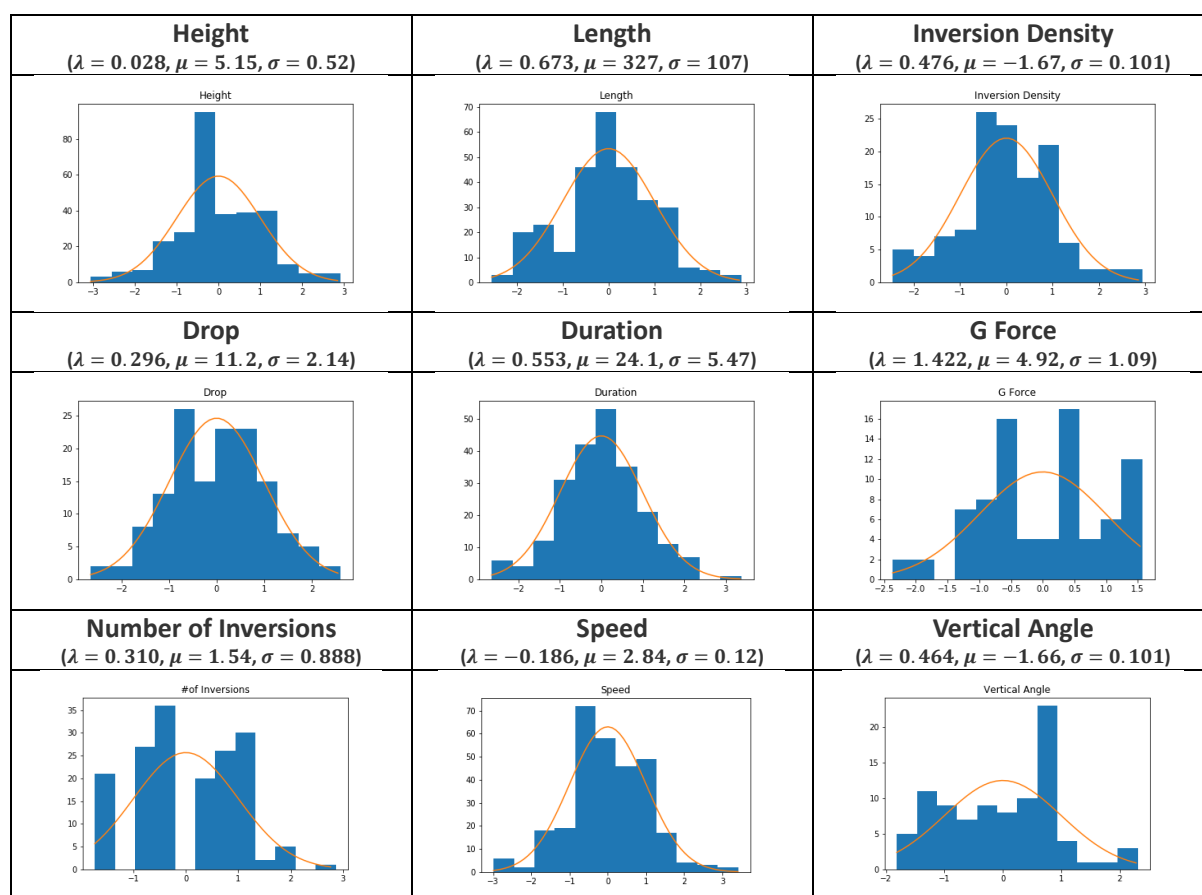


Figure 4.2: The distribution of eight numerical variables *before* Box-Cox transformation (blue) And the distribution of eight numerical variables *after* Box-Cox transformation (orange)

After converting the skewed data distribution into a normal one, we used the cumulative distribution

function of normal distributions to calculate the percentile position of all values in the dataset. With such transformation, all the values in our dataset will have the same range, from 0 to 1 inclusive. These transformed values can then be used in the following parts of our model.

5 - Model Part 2: Thrill Index

To create a descriptive roller coaster ranking system based only on numerical and descriptive specification data, we deliberated on methods of comparing different roller coasters, and concluded that the logical way is to develop a quantitative algorithm comparing how much thrill each of them brings to a rider. Theoretically, the more thrilling a roller coaster is, the more people will rate it higher, and thus its rank should be higher in our model. We can then stand on a statistical viewpoint and consider the aggregate effects of the rider population's opinions on an index evaluating the thrill of each roller coaster. Therefore, the goal of the first part of our model is to rank roller coasters according to the level of thrill they bring onto the average rider.

After researching into both objective sources such as Wikipedia and Encyclopaedia Britannica and subjective sources like Quora and roller coaster enthusiast forums, we concluded that there are several major factors overall that contributes to the experience or sensation of thrill:

- **The feeling of weightlessness,**
- **The maximum velocity,**
- **the maximum G-force experienced,**
- **Inversions, and**
- **The type of roller coaster.**

As these factors are statistically independent of each other, we would develop a formula that performs accumulation of each factor into a comparative index, namely *TI*, the Thrill Index, in the following text.

5.1 Incorporating the Feeling of Weightlessness

One of the most important reasons that people would want to ride a roller coaster is that the feeling of weightlessness when they are dropped from a high position is thrilling. Weightlessness or decreased weight is experienced when the acceleration downwards approaches 9.8m/s^2 or 32.17ft/s^2 , the gravitational acceleration near the surface of the earth, and the formula for weight in this scenario, $mg - ma$, approaches zero. The higher the roller coaster is designed, the longer a person on that roller coaster will be able to drop, experiencing larger downwards acceleration and thus weightlessness, and the more thrilling the experience.

In the provided dataset, there are two variables that should be directed related to the feeling of weightlessness: *height* and *drop*, both denoting the maximum value experienced in one trip on the roller coaster. However, when we plotted the two variables against each other in a scatterplot and drew the line of best fit, we discovered that these two variables are very highly correlated ($r=0.95$), as demonstrated in figure 5.1.

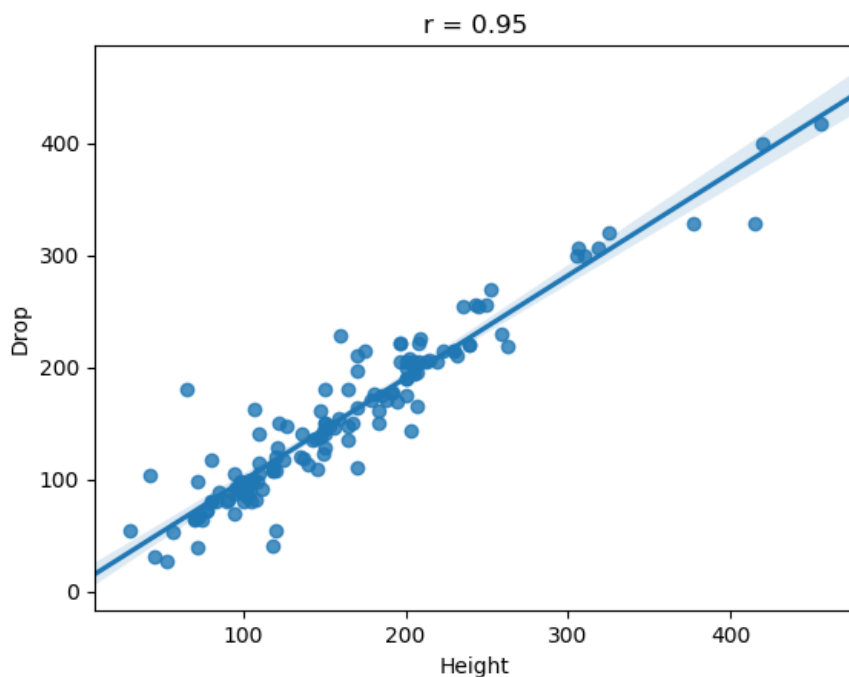


Figure 5.1 Scatterplot between height and drop

Since we can use any of the two variables to estimate the level of weightlessness that a rider is going to feel when falling, we decided to use the variable *height* and exclude the variable *drop*, as there is only a mere 1.67% of all entries in the dataset has the variable *height* missing, while an overwhelming 53% of all entries miss the variable *drop* in the dataset. And because of the high correlation between these variables, we can use a roller coaster's *height* to predict its *drop*.

Thus, the elementary formula for the thrill index is equal to the maximum height of the roller coaster.

$$TI = h_{max} \quad (5.1)$$

5.2 Incorporating Maximum Speed

After taking the feeling of weightlessness into account, we think it is reasonable that the maximum speed a roller coaster train is capable of reaching during one trip is also related to the feeling of thrill. For an analogy, people experiencing the thrill of putting their head outside of their cars' window when they are on a high-speed road such as a freeway is similar to the experience of riders on a roller coaster. The seemingly dangerous actions give people the sensation of thrill, and the feeling of air blowing right at people's face maximizes the sensation and produces excitement, which is where screaming aloud seems reasonable on a roller coaster. Therefore, the faster the roller coaster goes, the more thrilling the experience will be, and the thrill index should account for this factor as well.

So, the formula is adjusted into the following:

$$TI = h_{max} + s_{max} \quad (5.2)$$

5.3 Incorporating Acceleration

People come to an amusement park and ride a roller coaster to experience what they cannot in daily life, and vertical acceleration greater than 9.8m/s^2 or 32.17ft/s^2 , the natural gravitational acceleration of free fall near the surface of earth (denoted g), is sure to be one of these experiences. This greater acceleration is usually denoted in convention by the value of g-force, a scalar multiplier on the value of g that results in a downward vector of acceleration ($g \stackrel{\text{def}}{=} \frac{a}{g}$). Therefore, the higher the G force the roller coaster exerts on the rider, the higher the vertical or downwards acceleration is, and the more thrilling the experience becomes.

$$TI = h_{max} + s_{max} + g_{max} \quad (5.3)$$

5.4 Incorporating Inversions

After incorporating the physical traits and maximums of roller coasters, we propose that the design of the track of each roller coaster can also influence the overall experience of the rider, and should be viewed from a statistical standpoint. An effective way to increase the experience of thrill is increasing the number of inversions, which is when the rider and the train is positioned upside-down, usually at the top of a vertically circular track, from a ground perspective. This is another experience people do not encounter in daily life, and many would come to ride a roller coaster for it. Numerically and theoretically, the more inversions there are, the more thrilling a roller coaster ride becomes.

$$TI = h_{max} + s_{max} + g_{max} + n_I \quad (5.4)$$

However, when we are comparing two or more roller coasters having the same number of inversions, we propose that the coaster with a higher number of inversions per unit time should make the ride more intense, and thus more thrilling. Therefore, we divide the number of inversions of each roller coaster track by the one-time duration of its ride to get the inversion frequency ω of the roller coaster.

$$\omega \stackrel{\text{def}}{=} \frac{n_I}{t} \quad (5.5)$$

And so, the formula for the thrill index of each roller coaster should be readjusted to be the following, where the inversion frequency replaces the number of inversions:

$$TI = h_{max} + s_{max} + g_{max} + \omega \quad (5.6)$$

5.5 Incorporating the type of roller coasters

Finally, we should consider the many types of roller coasters that exist, some constructed specifically to attract enthusiast riders, and many of which are unique in experience. Thus, the type of the roller coaster contributes to the thrill a lot more than many would think. It is part of the design of the roller coaster, and as it affects each car in the roller coaster, goes on to affect the experience and thrill of the rider directly.

$$TI = h_{max} + s_{max} + g_{max} + \omega + b_T \quad (5.7)$$

But then, the type of roller coasters is a categorical variable, not numerical, so the calculation into the thrill index cannot be accomplished. So, we decided to convert the categorical variable into a numerical one, one that we could quantify based on values, while being as objective as possible. Arbitrary numbering is possible, but it would reduce the objectivity of our index, and so we decided to find the popularity of each type of roller coaster with samples statistically.

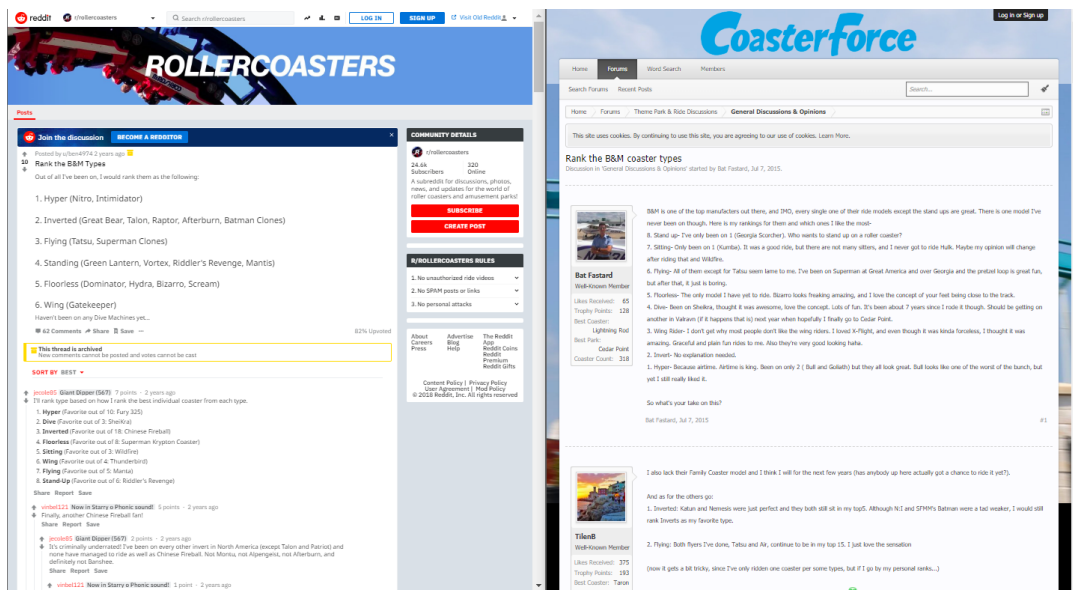


Figure 5.2: Data about the type of roller coaster are collected from websites

To determine the overall popularity of each type of roller coasters, we searched for existing data on the Internet that was voluntarily answered on forums (Reddit and CoasterForce) and collected a valid sample of 46 answering under the threads in a spreadsheet.

And so, we turned to voting theory, using the extended Borda Count method to determine the final

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	INDIVIDUAL RANKINGS													
2	Hyper	Inverted	Hyper	Hyper	Hyper	Hyper	Hyper	Wing	Inverted	Hyper	Inverted	Hyper	Hyper	Inverted
3	Inverted	Flying	Flying	Wing	Wing	Inverted	Inverted	Inverted	Hyper	Wing	Hyper	Inverted	Sit	Hyper
4	Wing	Sit	Wing	Suspended	Flying	Sit	Flying	Flying	Suspended	Flying	Flying	Flying	Inverted	Wing
5	Flying	Wing	Suspended	Sit	Inverted	Wing	Wing	Suspended	Sit	Inverted	Suspended	Wing	Suspended	Flying
6	Suspended	Hyper	Sit	Inverted	Suspended	Flying	Suspended	Sit	Wing	Suspended	Wing	Sit	Flying	Suspended
7	Sit	Suspended	Inverted	Stand	Sit	Suspended	Stand	Hyper	Flying	Sit	Sit	Suspended	Wing	Sit
8	Stand	Stand	Stand	Flying	Stand	Stand	Sit	Stand	Stand	Stand	Stand	Stand	Stand	Stand
9														

Figure 5.3: This figure shows the poll of different users on internet, rank from highest to lowest.

ranking of types of roller coasters. In the extended Borda Count method, each position on an answer is assigned points, 1 for the last place to 7 for the first place in this instance. The points are tallied for each type of roller coaster, and is sorted to determine the final ranking.

For Borda points to fit in the range of other data, we scale them into numbers from 0 to 1 by dividing the Borda points of each type by 264, the Borda points of the roller coaster that has a rank of number one, namely the type *Inverted*, as shown in table 5.1.

Rank	Types of Roller Coaster	Borda Points	Contribution to Thrill Index (0-1)
1	Inverted	264	1
2	Flying	186	0.704
3	Wing	178	0.674
4	Sit Down	156	0.591
5	Floorless	155	0.587
6	Stand	64	0.24

Table 5.1: The table shows the conversion of the type of a roller coaster to a numerical value using Borda Count Method

This ranking method naturally conforms to the Monotonicity Criterion, and in this instance, with a relatively large number of types, satisfies the Majority and Condorcet Criterion.

Note that in order to make the answers on the websites conform to our defined types of roller coasters, the type ‘Floorless’ is changed to ‘Suspended’, type ‘Dive’ is combined with ‘Sit Down’ and other small types are ignored as preferences are transitive.

6 - Model Part 3: Discomfort Index

It is a common sight to see people with a pale face right after a roller coaster ride – that is because roller coasters stimulate an experience that our body is not used to. After considering the thrill experienced during a roller coaster ride as one of our major ranking criteria, we propose the idea of accounting for the level of discomfort as well. A roller coaster may be extremely thrilling to ride, but if people experience discomfort on the trip, the overall experience of the ride can be seriously degraded.

After researching into the subject of discomfort on Wikipedia and Quora, we located two variables in the dataset that may contribute to a person’s feeling of discomfort – *G force* and *inversions*.

6.1 Incorporating G Force

Although a higher G force could emphasize the thrill of the overall experience of a roller coaster ride, it is also a major factor that causes discomfort in a typical roller coaster ride. Generally, a higher G force causes the discomfort to be greater on average for a rider.

According to previous studies, a human can tolerate an acceleration up to 4 G on average without feeling great discomfort. Above this threshold, symptoms like grey-out or even black-out are often reported by those untrained. Grey-out causes a person’s vision to lose hue, tunnel vision causes a person to lose peripheral vision, and black-out causes a complete loss of sight. Thus, G-force should be considered a factor that contributes to the probability of a rider experiencing discomfort.

We propose that the relationship between the probability of discomfort and G force is nonlinear, as most humans do not feel much discomfort from 1 G to 4 G acceleration, while it arises rapidly as G force approaches 5 and 6 G.

Therefore, we decided to use a logistic function, as shown in formula 6.1, to approximate the probability that the average rider will feel discomfort at a certain G force. The range of the predicted

probability should be between 0 and 1 inclusive because other variables in the dataset are scaled this way as explained in part 4.

$$DI_g = \frac{1}{1 + e^{45(0.88 - g_{max})}} \quad (6.1)$$

The parameter of the logistic regression is set with our assumption that the average rider can endure 5 G acceleration in average but cannot endure 6 G acceleration. As existing studies shows, 50% of the population cannot endure 5 G and 99.9% of the population cannot endure 6 G. Thus, we first transformed the critical values (5 G and 6 G) into the standardized scale, and then used them to determine the logistic function. The resulting points are (0.89, 0.5) for 5 G and (0.99, 0.99) for 6 G, and the shape of the logistic function is shown in figure 6.1.

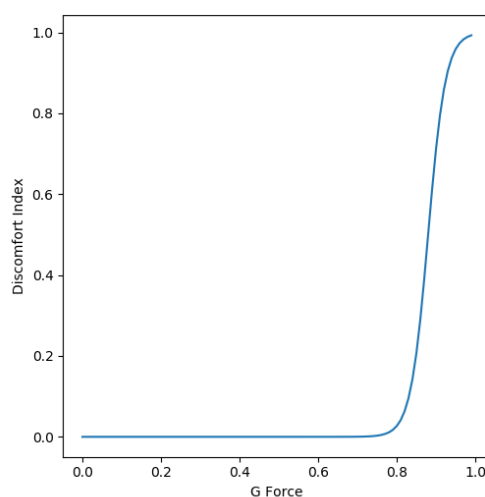


Figure 6.1: The graph of DI_g

The discomfort probability function starts with a long, continuous horizontal line when the G force, plotted in the horizontal axis, increases from 0 to 4, indicating very little discomfort, then arises sharply as it approaches 5. When G force exceeds 6, about all riders should feel sick and the probability of discomfort gets near 1 (the maximum).

6.2 Incorporating Inversion Frequency

Another factor that may cause discomfort during a roller coaster ride is the number of inversions. Many inversions in a single roller coaster ride causes the balance system inside the human brain to malfunction, leading to motion sickness and the feeling of discomfort. Generally, the more inversions there is, the more likely discomfort occurs for a rider.

When we are comparing two roller coaster rides with the same number of inversions, the one with a shorter duration will have a higher probability of generating more discomfort for the rider, since the average inversion frequency is higher on that ride. Remind that

$$\omega = \frac{n_I}{t}.$$

Frequent inversions can impair the balance system of human further, and may lead to complete loss of spatial vision and vomiting. The more frequent the inversion is, the dizzier people might feel. According to previous studies, about 90% of all people will have a strong feeling of dizziness when the inversion frequency is close to 0.133 inversions per second.

$$DI_{\omega} = \frac{1}{1 + e^{30(0.85-\omega)}} \quad (6.3)$$

Similar to our approach to G force, we use a logistic function to approximate the probability of discomfort. The range of the function is also from 0 to 1 inclusive. As the aforementioned study shows, the discomfort probability increases rapidly when inversion approaches 0.133 inversions per second [7], which after the Box-Cox transformation, described in part 4, is nearly equal to 0.922.

The resulting shape of the figure is as follows.

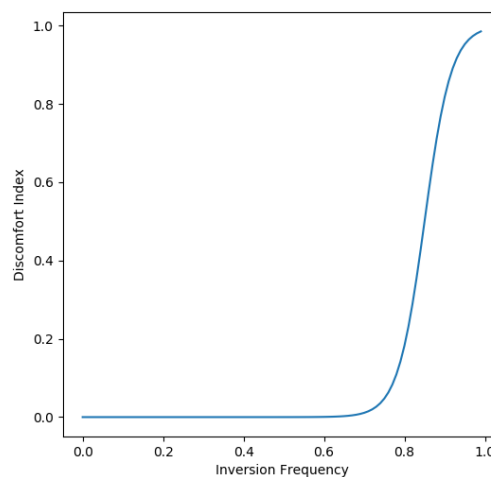


Figure 6.2 The graph of DI_{ω}

6.3 Discomfort Index

In conclusion, the discomfort index is equal to the sum of G-force probability of discomfort and the probability of inversion frequency discomfort, and its formula would be

$$DI = \frac{1}{1 + e^{45(0.88-G)}} + \frac{1}{1 + e^{30(0.85-\omega)}} \quad (6.4)$$

As shown below, we also plot this function in a 3D cube view for a better visualization. The discomfort is expressed in the plot both by the vertical axis and the colors of the rainbow, from purple representing the least discomfort and red representing the most.

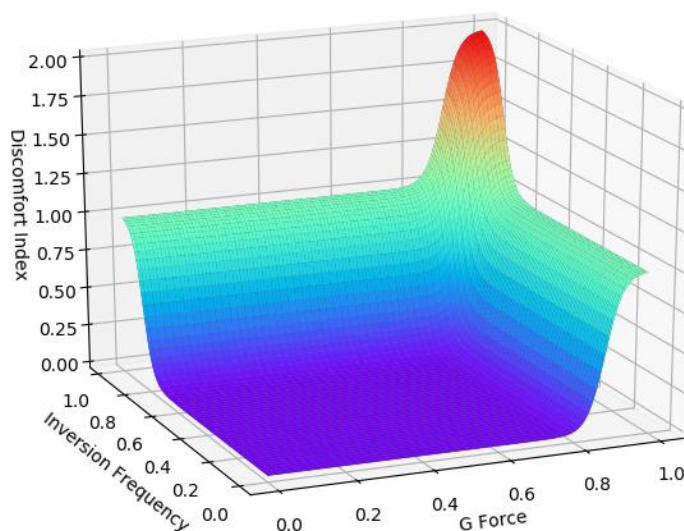


Figure 6.3: The 3D cube view plot of DI function

7 - Model Part 4: Comprehensive Index

As we have presented in the part 2 of this model, people come to ride a roller coaster in search of the “thrill”. The more thrilling the experience, the larger possibility that the rider will like it. We have then analyzed the many factors that contribute to this experience of thrill:

- The feeling of weightlessness,
- The maximum velocity,
- the maximum g-force experienced,
- Inversions, and
- The type of roller coaster.

And in model part 2, the index measuring thrill is the aggregate of these factors, as in the formula:

$$TI = h_{max} + s_{max} + g_{max} + \omega + b_T \quad (7.1)$$

As discussed in the third part of the model, the feeling of discomfort arises from the maximums of *G force* and *Inversion Frequency*, and this also affects the overall experience of a roller coaster ride. The formula to calculate discomfort takes both variables into account:

$$DI = \frac{1}{1 + e^{45(0.88-g_{max})}} + \frac{1}{1 + e^{30(0.85-\omega)}} \quad (7.2)$$

The higher the thrill index is, the larger possibility that riders will like it, but the higher the discomfort index is, the more likely that people will regret taking the ride. Both parts of the model aim to measure the experience of riding a particular roller coaster, but neither have taken all factors into account. Certainly, it is undesired that our ranking recommends roller coasters that simply have the highest drop, fastest speed, or the greatest number of inversions, while people feel horrible after riding it; equally undesired is that our ranking recommends people to ride roller coasters that does not create any discomfort but is not thrilling at all.

Therefore, an objective and comprehensive ranking system should take both sides into account with the equation

$$\mathbb{C} = TI - DI \quad (7.3)$$

which expands to

$$\begin{aligned} \mathbb{C} &= h_{max} + s_{max} + g_{max} + \omega + b_T - \frac{1}{1 + e^{45(0.88 - g_{max})}} - \frac{1}{1 + e^{30(0.85 - \omega)}} \\ &= \sum_{i=1}^7 \mathcal{F}_i \end{aligned} \quad (7.4)$$

where \mathbb{C} is the comprehensive index that measures the experience on any selected roller coaster.

By creating the equation as we did so, it is possible to penalize roller coasters that either gives excess emphasis to the experience of thrill or neglects the discomfort of the rider, or coasters that is incapable of thrill and emphasizes comfort overmuch.

The *Comprehensive* index give a more objective and comprehensive reference for the comparison of different roller coasters and offers insight into our rank of Top 10 Roller Coasters in the world. And as we aim to create an authoritative ranking that appeal both to roller enthusiasts and new riders, we can encourage roller coaster designers not to only go after the physical limits of a roller coaster (for example, speed, drop, and inversion frequency), but consider also the experience of discomfort the design might have on the average rider.

However, in this part, we omitted the relative importance of these factors affecting the final ranking (for example some might think the number of inversions that a roller coaster has is more important than its height, while others might think the opposite). To consider the relative importance, a weight should be added to each factors in the formula.

In order to take the relative importance between these factors into account of the Comprehensive Model, we added weights to each factor, which is explained in part 9.

8 - App: Roller ranker

After the construction for our model of ranking roller coasters is complete, for the convenience of both potential riders seeking for their next ride and enthusiasts seeking for thrill, we decided that it is necessary to design a mobile app named Roller Ranker so that they can receive the recommendations of our model directly in their phones.

8.1 Goals and Concepts of the App

1. User-Friendly

Our app should be designed to be as friendly to the user as possible, in ways that people will find it easy to use and understand.

2. Personalized

In our app, the users should be able to adjust settings according to their personal preferences

so that the app offers a customized recommendation for all individual users (for example, some might think the number of inversions that a roller coaster has is more important than its height, while others might think the opposite).

3. Self-improving

Our app should be able to learn the preference of users for different factors in our model automatically, so that our default recommendation will become more accurate in time.

8.2 Potential Users and App Design

Before designing the functions of the app, we shall first analyze the different groups of potential users of our app: those that are new to roller coasters, those who have ridden roller coasters before, and those who are roller coaster enthusiasts.

Those who are new to roller coasters might simply want our recommendation of roller coasters, which is based on our default ranking, as they do not know much about roller coasters.

Those who have some experience with riding roller coasters should know that some roller coasters can be extreme and causes discomfort (such as dizziness or grey-out), and others can be too boring and lacks excitement. Therefore, they may want to adjust the settings so that our model can recommend roller coasters that is perfect for them.

For Roller Coaster Enthusiasts who may have knowledge of every aspect of roller coasters, a very flexible customization of the settings may be desired so that they can find roller coasters that fits their specific preferences. For example, some enthusiasts may think the number of inversions that a roller coaster is more important than its height, and our app should let them express the preference.

Type of Users	What we offer
Beginner to roller coasters	Default Recommendation
Experienced riders	Customized setting for the weights of Thrill Index and Discomfort Index in the Comprehensive Model
Roller Coaster Enthusiasts	Customized setting for weights of all factors in the Comprehensive Model

Table 8.1: The table shows various offers to different type of users

To accommodate the need of those different types of users, our app should offer a default recommendation for beginners (ranked by the output of the Comprehensive index), an adjustable weight for the Thrill index and Discomfort index in our model for those who have some experience in riding roller coasters, and adjustable weights for all factors for Roller Coaster Enthusiasts.

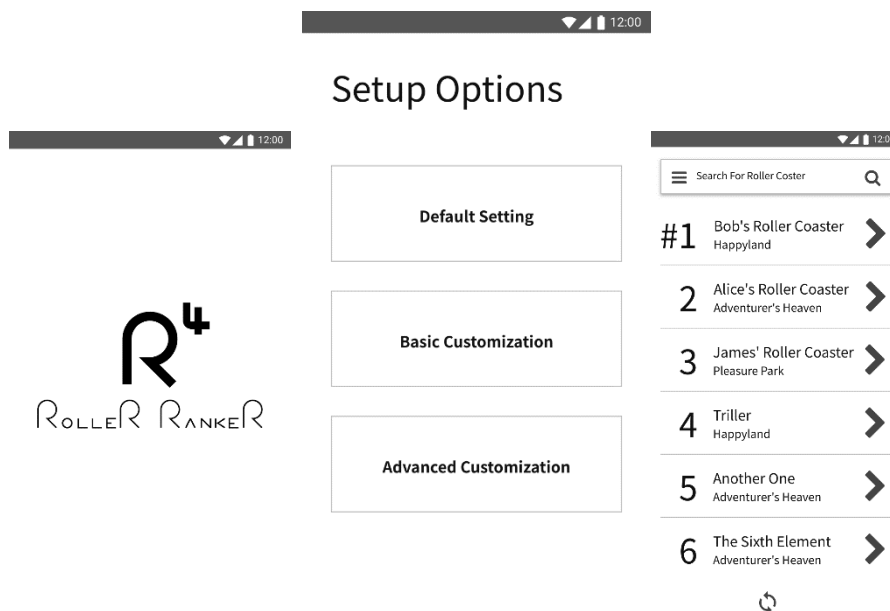


Figure 8.1 Prototype of “Roller Ranker”

Left: Front page of the app

Middle: 3 options for setup

Right: The personalized ranking

8.3 Incorporating customized settings into the Comprehensive Model

We have already developed a Comprehensive Model that takes many factors or variables into consideration. In that model, we assume that people will view each factor as having equal importance to them, so the default weight for each variable is set to be the same. However, as we have analyzed in part 8.2, different users might think of some factors as being more important to them than others. Therefore, we should incorporate the users’ setting into the Comprehensive Model so that our recommendation is flexible to their choice.

8.3.1 Basic Customization: Thrill Index vs Discomfort Index

For the basic customization algorithm, in a screen that asks for the user’s preference, a value x between 1 to 9 is retrieved from users’ input. The value x reflects the user’s absolute preference between Thrill and Comfort. We transform this value using $\alpha = \frac{x}{5}$ and $\beta = 1 - \frac{x}{5}$ and apply it in our Comprehensive Model formula for finding out a weighted index for our ranking. Thus, the formula based on this basic customization would be:

$$\mathbb{C} = \alpha \cdot DI - \beta \cdot TI \quad (8.1)$$

8.3.2 Advanced Customization: Detailed weights

For a more advanced customization process, in which a user can adjust weights for each individual factor in the Comprehensive Model, the user’s comparative preferences over each factor are collected in the Advanced Customization application interface.

This means that the method we use for basic customization, which handles an absolute preference,

cannot be applied here to comparative ones. Thus, we shall refer to the Analytic Hierarchy Process (AHP) method which uses relative importance of criteria and a pairwise comparison matrix to sort possible solutions to an operational problem. However, we only need a part of AHP, the Criteria layer, for our model. This is because the pairwise comparisons layer can be replaced by the similar scaling and normalization process in part 4.

We now use an example of the relative preferences of Katie, a roller coaster enthusiast and user of our app, on factors of roller coasters to illustrate how the modified method give the absolute weight of the factors. By these absolute weights, we can determine the preferences of the user exactly and recommend roller coasters by factors with larger weights.

For the sake of simplicity in our discussion, we first assign a number to each factor in the formula:

Number	1	2	3	4	5	6	7
Factor	h_{max}	s_{max}	g_{max}	ω	b_T	DI_g	DI_ω

Table 8.2: The table shows the number corresponding with specific factor

Using these numbers, we can denote the various aspects of the matrices in our following discussion with ease.

Step 1: Preparing the Data

First, users input their comparative preference for two factors at a time in our app. For example, Katie can slide the slider from 1 to 9 to indicate their comparative preference between the factors *height* and *speed*, as illustrated in figure 8.2.

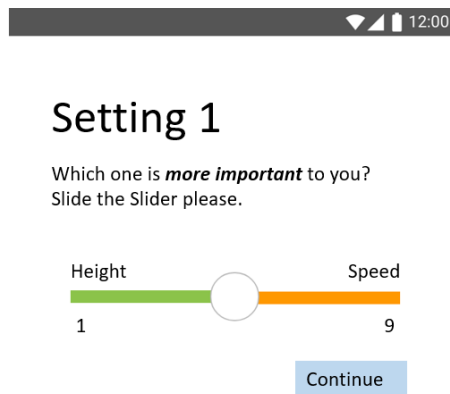


Figure 8.2: A sample setting page in the Advanced Customization setting

If Katie chose the values in the following table,

	h_{max}	s_{max}	g_{max}	ω	b_c	DI_g	DI_ω
h_{max}							
s_{max}	5.5						
g_{max}	6	5.25					
ω	6.5	5.5	5.17				
b_c	7	5.75	5.33	5.125			
DI_g	7.5	6	5.5	5.25	5		
DI_ω	8	6.25	5.67	5.375	5.9	5	

Table 8.3: The table shows the weight choosing by Katie

Then Katie's inputs can be transformed into values that can be used in the AHP by the formula

$$x' = \begin{cases} 2x - 9, & x \geq 5 \\ \frac{1}{11 - 2x}, & x < 5 \end{cases} \quad (8.2)$$

To calculate the relative weight for n different factors, a 7×7 comparison matrix A must be made to contain the user's relative preference data, produced following these rules:

- 1) $A_{ij} = \frac{1}{A_{ji}}$ (Rule of Reciprocity)
- 2) The value in matrix A_{ij} represents the comparative importance between factor i and factor j . If factor i is equally or more important when compared to factor j , the scale of A_{ij} is from 1-9, with 1 meaning equally important and 9 meaning factor i is preferred substantially more over factor j .

The transformed comparison matrix produced using Katie's preferences could be

$$A = \begin{bmatrix} 1 & 1/2 & 1/3 & 1/4 & 1/5 & 1/6 & 1/7 \\ 2 & 1 & 2/3 & 1/2 & 2/5 & 1/3 & 2/7 \\ 3 & 3/2 & 1 & 3/4 & 3/5 & 1/2 & 3/7 \\ 4 & 2 & 4/3 & 1 & 4/5 & 2/3 & 4/7 \\ 5 & 5/2 & 5/3 & 5/4 & 1 & 5/6 & 5/7 \\ 6 & 3 & 2 & 3/2 & 6/5 & 1 & 6/7 \\ 7 & 7/2 & 7/3 & 7/4 & 7/5 & 7/6 & 1 \end{bmatrix}$$

Or, if she chose some other preferences, the importance of which we will illustrate later, the matrix could be

$$B = \begin{bmatrix} 1 & 1/2 & 1/3 & 1/4 & 1/5 & 1/6 & 1/8 \\ 2 & 1 & 2/3 & 1/2 & 2/5 & 1/3 & 2/7 \\ 3 & 3/2 & 1 & 3/4 & 3/5 & 1/2 & 3/7 \\ 4 & 2 & 4/3 & 1 & 4/5 & 2/3 & 1/2 \\ 5 & 5/2 & 5/3 & 5/4 & 1 & 1 & 5/7 \\ 6 & 3 & 2 & 3/2 & 1 & 1 & 1 \\ 8 & 7/2 & 7/3 & 2 & 7/5 & 1 & 1 \end{bmatrix}$$

Step 2: Checking for Inconsistencies

Before applying a comparison matrix to calculate the relative importance of criteria, we need to check if it is rational, which is, not self-contradictory or inconsistent in opinions. For example, if A is preferred over B and B is preferred over C, it is irrational that C is preferred over A. Several steps are required to achieve this.

If a person's preference of the factors (being more or less important to them) is perfectly consistent, then the comparison matrix \mathbf{A} derived from his or her preferences should satisfy

$$\mathbf{A}_{ij} \cdot \mathbf{A}_{jk} = \mathbf{A}_{ik} \quad (8.3)$$

A matrix that satisfies this property is a consistent matrix. It can be proved that a $n \times n$ reciprocal matrix is consistent if and only if its greatest eigenvalue

$$\lambda_{max} = n \quad (8.4)$$

This illustrates a method to test if a person's opinion is perfectly consistent. For the Katie's example, the 7×7 comparison matrix \mathbf{A} is consistent because

$$\max\{\lambda: (\mathbf{A} - \lambda\mathbf{E})\mathbf{x} = \mathbf{0}\} = 7 \quad (8.5)$$

while the comparison matrix \mathbf{B} is not consistent because

$$\max\{\lambda: (\mathbf{B} - \lambda\mathbf{E})\mathbf{x} = \mathbf{0}\} \approx 7.011 > 7. \quad (8.6)$$

However, people do not always hold consistent opinions, and thus we need to use a consistency index CI to measure the extent to the consistency of a comparison matrix. For a $n \times n$ matrix with the greatest eigenvalue λ_{max} , its consistency index CI can be defined as

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (8.7)$$

Note that matrices having smaller CI is more consistent. The calculated CI is then compared with the random index RI . The random index RI is defined to be the average CI of random comparison matrices with same order. Typically, we can tolerate a matrix with a consistency index that satisfies

$$\frac{CI}{RI} < 0.1 \quad (8.8)$$

Returning to the example of Katie, as we know that the comparison matrix \mathbf{B} she produced is not perfectly consistent, to check whether the inconsistency of opinions represented by matrix \mathbf{B} is tolerable, we will first refer to table 8.2 for the value of RI when $n = 7$, which is $RI = 1.32$.

n	2	3	4	5	6	7	8	9	10
RI	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.51

Table 8.4: RI value for different n

Then, we test if

$$\frac{CI}{RI} = \frac{\frac{\lambda_{max} - n}{n - 1}}{RI} = \frac{\frac{7.011 - 7}{7 - 1}}{1.32} = 0.0014 < 0.1$$

is true. As $0.0014 < 0.1$ holds true, we can state that the inconsistency of matrix \mathbf{B} is in the tolerable range, and the data collected is considered valid.

Step 3: Calculating the Weight Vector

After checking the consistency of the comparison matrix, we continue with calculating the absolute weight of each factor. We first deduce the normalized pairwise comparison matrix $\bar{\mathbf{A}}$ from a $n \times n$ comparison matrix \mathbf{A} using the formula

$$\bar{A}_{ij} = \frac{A_{ij}}{\sum_k A_{kj}}, \quad (8.9)$$

then build the criteria weight vector \mathbf{w} using the formula

$$w_i = \frac{\sum_j A_{ij}}{n}. \quad (8.10)$$

Thus, the resulting normalized pairwise comparison matrices $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ for the comparison matrices \mathbf{A} and \mathbf{B} in Katie's example would be

$$\bar{\mathbf{A}} = \begin{bmatrix} \frac{1}{28} & \frac{1}{28} & \frac{1}{28} & \frac{1}{28} & \frac{1}{28} & \frac{1}{28} & \frac{1}{28} \\ \frac{1}{14} & \frac{1}{14} & \frac{1}{14} & \frac{1}{14} & \frac{1}{14} & \frac{1}{14} & \frac{1}{14} \\ \frac{3}{28} & \frac{3}{28} & \frac{3}{28} & \frac{3}{28} & \frac{3}{28} & \frac{3}{28} & \frac{3}{28} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{5}{28} & \frac{5}{28} & \frac{5}{28} & \frac{5}{28} & \frac{5}{28} & \frac{5}{28} & \frac{5}{28} \\ \frac{3}{14} & \frac{3}{14} & \frac{3}{14} & \frac{3}{14} & \frac{3}{14} & \frac{3}{14} & \frac{3}{14} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

and

$$\bar{B} = \begin{bmatrix} 1/29 & 1/28 & 1/28 & 1/29 & 1/27 & 1/28 & 7/227 \\ 2/29 & 1/14 & 1/14 & 2/29 & 2/27 & 1/14 & 16/227 \\ 3/29 & 3/28 & 3/28 & 3/29 & 1/9 & 3/28 & 24/227 \\ 4/29 & 1/7 & 1/7 & 4/29 & 4/27 & 1/7 & 28/227 \\ 5/29 & 5/28 & 5/28 & 5/29 & 5/27 & 3/14 & 40/227 \\ 6/29 & 3/14 & 3/14 & 6/29 & 5/27 & 3/14 & 56/227 \\ 8/29 & 1/4 & 1/4 & 8/29 & 7/27 & 3/14 & 56/227 \end{bmatrix}.$$

The criteria weight vectors w_A and w_B will be

$$w_A = \begin{bmatrix} 1/28 \\ 1/14 \\ 3/28 \\ 1/7 \\ 5/28 \\ 3/14 \\ 1/4 \end{bmatrix}$$

and

$$w_B = \begin{bmatrix} 1214239/34837326 \\ 1236163/17418618 \\ 1236163/11612412 \\ 1214239/8709309 \\ 1589639/8709309 \\ 1763141/8709309 \\ 2204656/8709309 \end{bmatrix}.$$

Step 4: Output

After Katie’s settings are converted to the criteria weight vectors w_A or w_B , we can apply these weights into the Comprehensive Model directly, as shown in the formula below.

$$\begin{aligned} \mathbb{C} &= w_1 h_{max} + w_2 s_{max} + w_3 g_{max} + w_4 \omega + w_5 b_T \\ &\quad - w_6 \frac{1}{1 + e^{45(0.88 - g_{max})}} - w_7 \frac{1}{1 + e^{30(0.85 - \omega)}} \\ &= \sum_{i=1}^7 w_i \mathcal{F}_i \end{aligned} \tag{8.11}$$

Using this formula, the app recalculates the Comprehensive index for each roller coasters, and rank them accordingly. With this customized ranking for each user, our recommendation would be flexible to individual preferences.

9 - Adding Weights to the Comprehensive Model

As we discussed in part 8.2, users that are new to roller coasters simply want to leave all the

customization as default and have our recommendation. The unweighted version of Comprehensive Model we built in part x gives a default ranking; however, it is under the assumption that people view each factor with equal importance, but it is certainly not the case in real life. After considering this problem, we think that we can improve the default recommendation by adding weights to each factor.

Recall that the unweighted version of the Comprehensive Model in part x:

$$\begin{aligned} \mathbb{C} &= h_{max} + s_{max} + g_{max} + \omega + b_T - \frac{1}{1 + e^{45(0.88-g_{max})}} - \frac{1}{1 + e^{30(0.85-\omega)}} \\ &= \sum_{i=1}^7 \mathcal{F}_i \end{aligned} \quad (9.1)$$

After we add weights to the Comprehensive Model, it can be represented by the formula

$$\begin{aligned} \mathbb{C} &= \mathbf{w}_1 h_{max} + \mathbf{w}_2 s_{max} + \mathbf{w}_3 g_{max} + \mathbf{w}_4 \omega + \mathbf{w}_5 b_T \\ &\quad - \mathbf{w}_6 \frac{1}{1 + e^{45(0.88-g_{max})}} - \mathbf{w}_7 \frac{1}{1 + e^{30(0.85-\omega)}} \\ &= \sum_{i=1}^7 \mathbf{w}_i \mathcal{F}_i \end{aligned} \quad (9.2)$$

where \mathbf{w}_c is the weight vector.

To find the values of \mathbf{w} , we used the same method in Section 8.3.2, the part where we customized a ranking for different users using the AHP method.

First of all, we conducted a survey online (the result of which is in Appendix D) to find people's comparative preference between each factors of our model (similar to **Section 8.3.2 Advanced Customization: Detailed weights** where we ask the users of our app to input their customized settings). We asked people questions like "is the height of the roller coaster or the speed of the roller coaster more important to you? please scale from 1 to 9, 5 means equally important". The study was done with 60 different users, each fully complete 21 questions about the preference of one factor over another scaling from 1 to 9.

Using the transformation described in Section 8.3.2, we calculate the 7×7 Comparison Matrix \mathbf{A}_c :

$$\mathbf{A}_c = \begin{bmatrix} 1 & 0.33 & 7.0 & 7.0 & 7.0 & 5.0 & 5.0 \\ 3 & 1 & 7.0 & 7.0 & 7.0 & 5.0 & 5.0 \\ 0.14 & 0.14 & 1 & 1 & 3 & 1 & 3 \\ 0.14 & 0.14 & 1 & 1 & 5 & 0.33 & 1 \\ 0.14 & 0.14 & 0.33 & 0.2 & 1 & 1 & 1 \\ 0.2 & 0.2 & 1 & 3 & 1 & 1 & 1 \\ 0.2 & 0.2 & 0.33 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (9.3)$$

After that, we follow the exact same steps in **Section 8.3.2 Advanced Customization: Detailed**

weights, and the resulting w_c is this:

$$w_c = \begin{bmatrix} 1.24 \\ 1.69 \\ 0.35 \\ 0.30 \\ 0.18 \\ 0.31 \\ 0.23 \end{bmatrix} \quad (9.4)$$

Thus, the final version of the Comprehensive Model is this:

$$\begin{aligned} \mathbb{C} &= 1.24 * h_{max} + 1.69 * s_{max} + 0.35 * g_{max} + 0,3 * \omega + 0.18 * b_T \\ &\quad - 0.31 * \frac{1}{1 + e^{45(0.88-g_{max})}} - 0.23 * \frac{1}{1 + e^{30(0.85-\omega)}} \\ &= \sum_{i=1}^7 w_c \mathcal{F}_i \end{aligned} \quad (9.5)$$

10 - Results

Using the weighted Comprehensive Model (part 9), we calculated the Comprehensive Index of each roller coasters; and each roller coasters are ranked accordingly, and the *Steel Dragon 2000* from *Nagashima Spa Land* park ranks number one, with comprehensive index equals to 1.72.

The following table shows our rank of “Top 10 Roller Coasters in the World”:

Name	Park	Comprehensive Index
Steel Dragon 2000	Nagashima Spa Land	1.72
Fury 325	Carowinds	1.71
Millennium Force	Cedar Point	1.70
Fujiyama	Fuji-Q Highland	1.66
Leviathan	Canada's Wonderland	1.65
Formula Rossa	Ferrari World Abu Dhabi	1.65
Desperado	Buffalo Bill's Resort & Casino	1.62
Intimidator 305	Kings Dominion	1.62
Titan	Six Flags Over Texas	1.62
Steel Vengeance	Cedar Point	1.61

Table 10.1: The table shows the ranking based on comprehensive index

10.1 Comparison with Other Rankings Systems

In order to evaluate our model, we compared our model with two other ranking systems we found online.

The first one is found on ranker.com, which is a large community (over 1200 voters on this topic) where everyone can upvote and downvote roller coasters that they like or dislike (with more than 9000 operations already), and the system will adjust its ranking according to user preference.

This model might represent the general preference of the population toward roller coasters, but it may be biased because this is a voluntary survey, which may result in undercoverage bias.

The second ranking system that we found is in a personal blog, written by a roller coaster enthusiast named Shannon George. This ranking reflects the preference of an expert who have ridden a great number of roller coasters. However, this ranking is very subjective.

The reason we choose these two rankings to compare with our model is because their ranking concept differs from our model’s greatly. Our model emphasizes on the balance between objectiveness and subjectiveness, while the ranking on ranker.com uses purely rider feedbacks and George’s ranking is based on the preference of one person. The comparison of the top 10 roller coasters that ranks in these three models are compared in Table 10.2.

It can be observed that there are many common roller coasters on our Top 10 ranking from Table 10.2.

Ranking	Our Model	Ranker.com	Enthusiast George
1	Steel Dragon 2000	Millennium Force	Bizarro
2	Fury 325	Steel Vengeance	Millennium Force
3	Millennium Force	Top Thrill Dragster	El Toro
4	Fujiyama	Maverick	Expedition GeForce
5	Leviathan	El Toro	The Voyage
6	Formula Rossa	Fury 325	Kingda Ka
7	Desperado	Intimidator 305	Intimidator 305
8	Intimidator 305	The Voyage	Goliath
9	Titan	Kingda Ka	Behemoth
10	Steel Vengeance	Apollo’s Chariot	Nemesis

Table 10.2: Comparison Table between our ranking and two other ranking found online (common items are marked with a common color)

11 - Evaluation

There are both strengths and weaknesses in our model:

Strengths

Our model is very **comprehensive**. We have taken many factors that may contribute to the ranking of roller coasters, including its maximum speed, maximum G force, height, inversion density, and its type. We have also considered that both the feeling of thrill and the feeling of discomfort contributes to overall experience of ride; therefore, the model also penalize extreme roller coasters.

Our model is very **objective**. This is because it is based on data and all data are standardized and normalized using Box-Cox transformation so that the unit of each variables will not affect the final ranking.

The results of our model **are easy to interpret**. The index is designed to be straightforward ‘bigger is better’ and one can effectively compare two roller coasters with the need to understand its mechanism.

The model quantifies the amount of thrilling each type of roller coasters brings to a rider with the Borda Count Method.

Our model is **flexible** for missing data. This is because we have analyzed correlation relationships between variables so that missing values can be filled by mean imputation or regression imputation.

The default recommendations of our app can improve itself by learning the customized settings of different users.

Limitations

Each variable in our calculation of the Comprehensive Model is not weighted, meaning that they are equally important when a person is judging a roller coaster. However, this is not the case in real life. This problem is solved in our app, which uses the customized setting of our users to offer recommendations that fits them the most.

The dataset that we used to rank our “Top 10 Roller Coasters in the World” only have the data of 300 roller coasters, which is far lesser than the total number (about 3950) of roller coasters worldwide. Therefore, our rank does not consider other roller coasters which are not included in the dataset.

There are a lot of missing data and they are replaced by means of imputation, which may be inaccurate and lead to biased results on some operational roller coasters

The Borda count method — which we used to quantify the amount of thrilling each type of roller coasters brings to a rider — do not satisfy all fairness criteria, as stated by Arrow’s impossibility theorem. It also does not satisfy the Independence-of-Irrelevant-Alternatives Criterion (IIA), which in this scenario means that the ranking is prone to change because of lowly-ranked roller coasters closing.

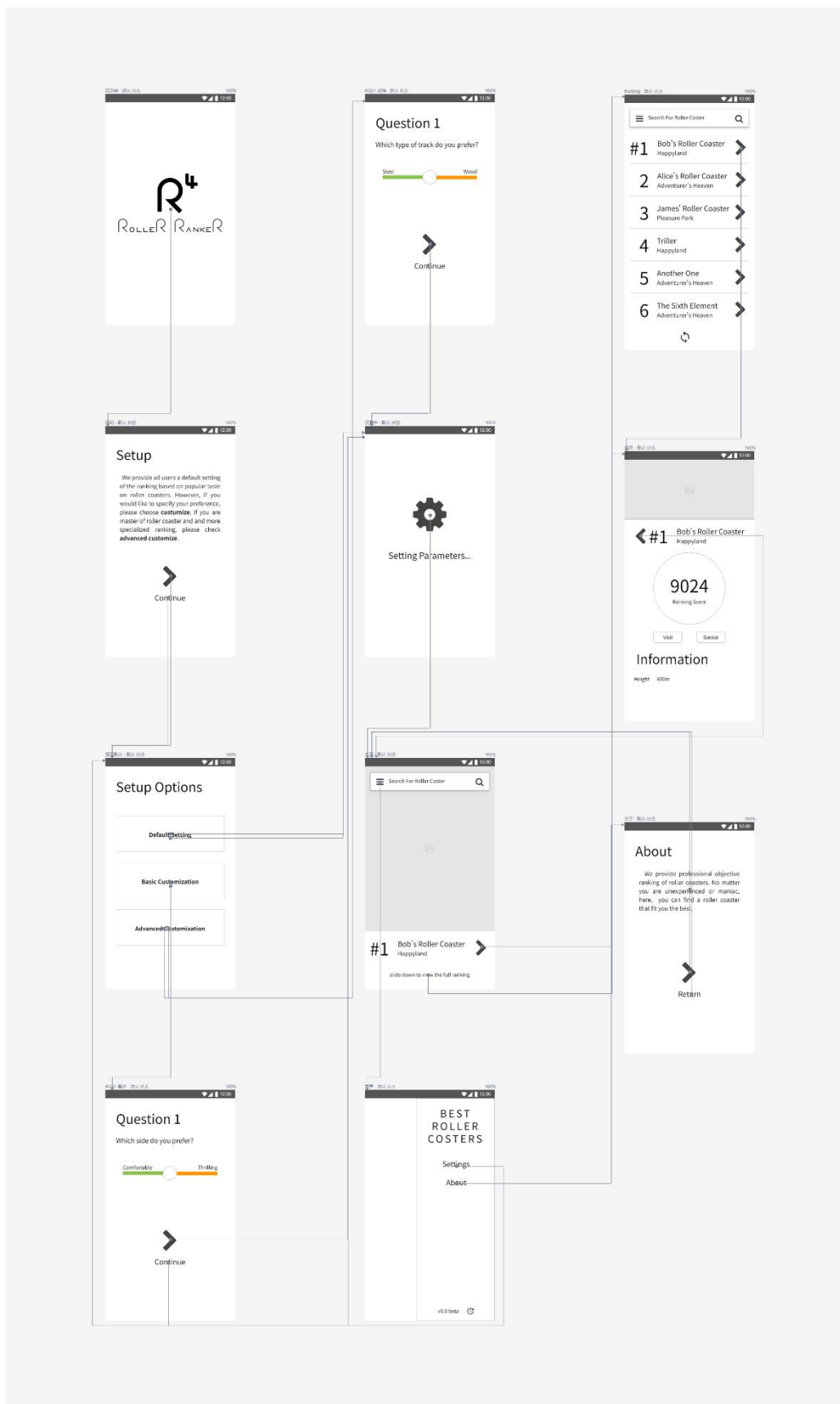
The Discomfort probability curve is derived from a study instead of real data and may be inaccurate of representation for the general public.

12 - References

1. “Random Roller Coaster.” Roller Coaster DataBase, www.rcdb.com.
2. “Ultimate Rollercoaster.” Coney Island | Roller Coaster History, www.ultimaterollercoaster.com.
3. “Coasterpedia The Roller Coaster Wiki.” Coasterpedia, coasterpedia.net.
4. “Rank the B&M Types • r/Rollercoasters.” Reddit, www.reddit.com/r/rollercoasters/comments/56sz3e/rank_the_bm_types/.
5. “Rank the B&M Coaster Types.” COASTERFORCE, coasterforce.com/forums/threads/rank-the-b-m-coaster-types.39119/.
6. Pandolf, Kent B., and R. E. Burr. Medical Aspects of Harsh Environments. Office of the Surgeon General, United States Army, 2001.
7. DeZio, Paul A, and James R Lackner. “Perceived Orientation, Motion, and Configuration of the Body during Viewing of an off-Vertical, Rotating Surface.” Perception & Psychophysics, Jan. 1986, pp. 39–46.
8. “Material Design.” Wikipedia, Wikimedia Foundation, 26 Oct. 2018, en.wikipedia.org/wiki/Material_Design.
9. Ergonomic Requirements for Office Work with Visual Display Terminals, ISO 9241-11, ISO, Geneva, 1998.

Appendix

Appendix A: The Prototype of Our Application



Appendix B: Code

Code #1 Data Analysis and Visualization

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
get_ipython().run_line_magic('matplotlib', '')

# Import dataset
data =
pd.read_excel("COMAP_RollerCoasterData_2018.xlsx")[0:300].drop(["Park", "City/Region", "City/State/Region", "Country/Region", "Geographic Region", "Status"], axis=1)

data

# Output basic statistics about the dataset
data.describe()

data = data.rename({"Year/Date Opened": "Year",
                   "Height (feet)": "Height",
                   "Speed (mph)": "Speed",
                   "Length (feet)": "Length",
                   "Number of Inversions": "#of Inversions",
                   "Drop (feet)": "Drop",
                   "Vertical Angle (degrees)": "Vertical Angle",
                   "Duration (min:sec)": "Duration"
                  },
                  axis="columns")

# Convert python Datetime Object to seconds
for i in range(data.shape[0]):
    if(type(data.Duration[i]) != float):
        data.Duration[i] = (data["Duration"][i].hour*60 + data["Duration"][i].minute)

# Type conversion
data["Inversion Density"] = data["#of Inversions"] / data["Duration"]
data["Inversion Density"] = pd.to_numeric(data["Inversion Density"]).astype(float)
data["Duration"] = pd.to_numeric(data["Duration"]).astype(float)
data["Duration"] = pd.to_numeric(data["Duration"]).astype(float)

# Correlation Matrix
data["Inversion Density"] = pd.to_numeric(data["Inversion Density"]).astype(float)
data["Duration"] = pd.to_numeric(data["Duration"]).astype(float)
Correlation = data.corr()
Correlation[abs(Correlation)>0.4]
plt.figure(figsize=(8,8))
plt.title("Correlation Matrix (|r| >= 0.4)")
sns.heatmap(Correlation[abs(Correlation)>0.4], xticklabels=Correlation.columns,
            yticklabels=Correlation.columns, annot=True,
            cmap = sns.diverging_palette(250, 10, as_cmap=True),)

# Regression Plots
plt.plot(data.Height, data.Drop, ".")
```



```
plt.title("Height vs Drop")
sns.regplot(x="Height", y="Drop", data=data,label="scatter");
sns.regplot(x="G Force", y="Length", data=data,scatter=False,label="no scatter");
sns.pairplot(data,kind="reg",diag_kind='kde')
```

Code #2 Comprehensive Model

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
get_ipython().run_line_magic('matplotlib', '')

# Import dataset
data = pd.read_excel("Standardized Data.xlsx")
data["Type Score"] = np.zeros(data.shape[0])
data["Thrill Index"] = np.zeros(data.shape[0])
data["Discomfort Index"] = np.zeros(data.shape[0])
data["Comprehensive Index"] = np.zeros(data.shape[0])
data = data.drop(["Park", "City/Region", "City/State/Region", "Country/Region", "Geographic
Region", "Status"],axis=1)

data

# Calculates type score
for i in range(0,data.shape[0]):
    if data["Type"][i] == "Inverted":
        data["Type Score"][i] = 1.0
    elif data["Type"][i] == "Flying":
        data["Type Score"][i] = 0.704
    elif data["Type"][i] == "Wing":
        data["Type Score"][i] = 0.674
    elif data["Type"][i] == "Sit Down":
        data["Type Score"][i] = 0.591
    elif data["Type"][i] == "Suspended":
        data["Type Score"][i] = 0.587
    elif data["Type"][i] == "Stand Up":
        data["Type Score"][i] = 0.24
    else:
        data["Type Score"][i] = 0.2

# Set Weight values
W_thrill = 1
W_discomfort = 1

Wheight = 0.28957165
Wspeed = 0.39279611
WG = 0.08114455
Wtype = 0.06912188
WInversion = 0.06912188
WDG = 0.0725201
WDI = 0.052985

# CalculateThrill Index
```

```
data["Thrill Index"] = data["Length"] + Wtype*data["Type Score"] + Wheight*data["Height"] +
Wspeed*data["Speed"] + WG*data["G Force"] + WInversion*data["Inversion Density"]
```

```
# Calculates discomfort index
```

```
x = np.array(data['G Force'])
```

```
y = np.array(data['Inversion Density'])
```

```
data["Discomfort Index"] = WDG*(1 / (1 + np.exp(45 * (0.88 - x)))) + WDI*(1 / (1 + np.exp(30 * (0.85 - y))))
```

```
# Calculates Overall
```

```
data["Comprehensive Index"] = W_thrill*data["Thrill Index"] - W_discomfort*data["Discomfort Index"]
```

```
# Sort by Comprehensive Index
```

```
sort = data.sort_values("Comprehensive Index", ascending=False)
```

```
sort
```

```
# Output
```

```
data.to_csv("Overall Ranking(Weighted).csv")
```

Code #2 AHP

```
import numpy as np
```

```
RI = 1.32
```

```
# Function to convert users' input to Comparison Matrix used by AHP
```

```
def t(x):
```

```
    if x>=5:
```

```
        return 2*x-9
```

```
    else:
```

```
        return 1/(11-2*x)
```

```
user_input = [[None, None, None, None, None, None, None],
```

```
              [6, None, None, None, None, None, None],
```

```
              [2, 2, None, None, None, None, None],
```

```
              [2, 2, 5, None, None, None, None],
```

```
              [2, 2, 4, 3, None, None, None],
```

```
              [3, 3, 5, 6, 5, None, None],
```

```
              [3, 3, 4, 5, 5, 5, None]]
```

```
# A = Comparison Matrix
```

```
A = np.ones((7,7))
```

```
for i in range(0,7):
```

```
    for j in range(0,7):
```

```
        if i==j:
```

```
            A[i][j]=1
```

```
            continue
```

```
        if user_input[i][j]==None:
```

```
            A[i][j]=1/t(user_input[j][i])
```

```
        else:
```

```
            A[i][j]=t(user_input[i][j])
```

```
l, _ = np.linalg.eig(A)
```

```
# Check for Inconsistency
```

```
if max(l) - 7 > 0.00000001:
```

```
    if (max(l)-7)/6/RI < 0.1:
```

```
        print('tolerable inconsistency')
    else:
        print('intolerable inconsistency')
        quit()

col_sum = np.sum(A, axis=0)

# Calculate Normalized matrix A
adjA = np.ones((7,7))
for i in range(0,7):
    for j in range(0,7):
        adjA[i][j] = A[i][j] / col_sum[j]

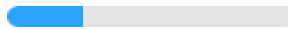
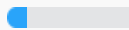







# Calculate Weight Vector
w = np.sum(adjA, axis=1)
w /= 7

print(w)
```


Appendix D: Survey Results for part 9

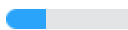

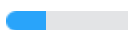

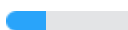


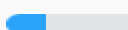

Q1: Height of Roller Coaster vs G Force of Roller Coaster

Mean score: 4.5

Options	Count	Proportion
1 Prefer Height the Most	17	 27.42%
2	6	 9.68%
3	3	 4.84%
4	4	 6.45%
5 same	12	 19.35%
6	3	 4.84%
7	1	 1.61%
8	4	 6.45%
9 Prefer G force the most	12	 19.35%
Total survey collected	62	

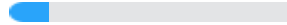








Q2: Height vs Speed

Mean score: 5.31

Options	Count	Proportion
1 Prefer height	10	 16.13%
2	3	 4.84%
3	10	 16.13%
4	2	 3.23%
5	10	 16.13%
6	1	 1.61%
7	4	 6.45%
8	7	 11.29%
9 Prefer Speed	15	 24.19%
Total survey collected	62	










Q3: Height vs Inversions

Mean score: 6.16

Options	Count	Proportion
1 Height	7	 11.29%
2	3	 4.84%
3	4	 6.45%
4	2	 3.23%
5	8	 12.9%
6	4	 6.45%
7	8	 12.9%
8	5	 8.06%
9 Inversions	21	 33.87%
Total survey collected	62	

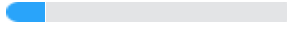








Q4: Height vs Type

Mean score: 4.73

Options	Count	Proportion
1 Height	11	 17.74%
2	4	 6.45%
3	10	 16.13%
4	4	 6.45%
5	10	 16.13%
6	6	 9.68%
7	5	 8.06%
8	1	 1.61%
9 Type	11	 17.74%
Total survey collected	62	










Q5: Height vs Discomfort arise from G force

Mean score: 5.71

Options	Count	Proportion
1 Height	8	 12.9%
2	1	 1.61%
3	4	 6.45%
4	3	 4.84%
5	17	 27.42%
6	3	 4.84%
7	5	 8.06%
8	7	 11.29%
9 Discomfort arise from G force	14	 22.58%
Total survey collected	62	

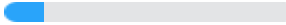
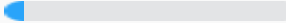
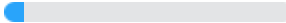
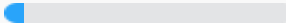
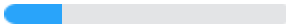
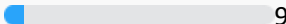
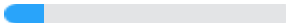
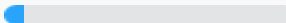
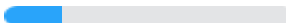
Q6: Height vs Discomfort arise from Inversions

Mean score: 6.32

Options	Count	Proportion
1 Height	5	 8.06%
2	2	 3.23%
3	5	 8.06%
4	2	 3.23%
5	10	 16.13%
6	2	 3.23%
7	8	 12.9%
8	10	 16.13%
9 Discomfort arise from Inversions	18	 29.03%
Total survey collected	62	

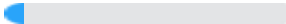
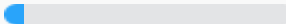
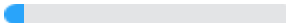
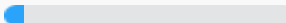
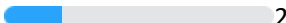
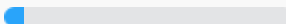
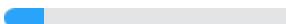
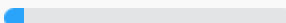
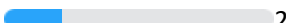
Q7: G force vs speed

Mean score: 5.37

Options	Count	Proportion
1 G Force	8	 12.9%
2	3	 4.84%
3	5	 8.06%
4	4	 6.45%
5	13	 20.97%
6	6	 9.68%
7	8	 12.9%
8	4	 6.45%
9 Speed	11	 17.74%
Total survey collected	62	

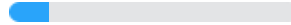

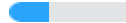






Q8: G force vs Inversions

Mean score: 5.94

Options	Count	Proportion
1 G force	3	 4.84%
2	4	 6.45%
3	4	 6.45%
4	4	 6.45%
5	14	 22.58%
6	5	 8.06%
7	9	 14.52%
8	5	 8.06%
9 Inversions	14	 22.58%
Total survey collected	62	









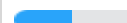
Q9: G force vs Type

Mean score: 5

Options	Count	Proportion
1 G force	10	 16.13%
2	2	 3.23%
3	8	 12.9%
4	6	 9.68%
5	12	 19.35%
6	7	 11.29%
7	2	 3.23%
8	3	 4.84%
9 Type	12	 19.35%
Total survey collected	62	

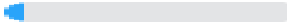
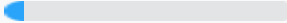
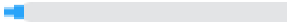
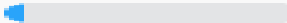
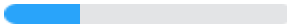
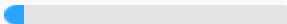
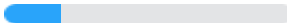
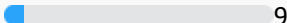
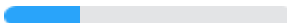
Q10: G force vs Discomfort G Force

Mean score: 5.71

Options	Count	Proportion
1 G force	6	 9.68%
2	1	 1.61%
3	4	 6.45%
4	2	 3.23%
5	18	 29.03%
6	5	 8.06%
7	13	 20.97%
8	2	 3.23%
9 Discomfort from G force	11	 17.74%
Total survey collected	62	

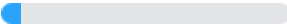
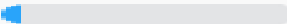
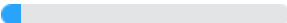
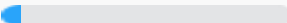
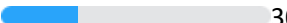
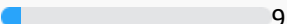
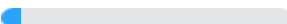
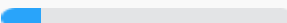
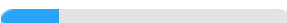
Q11: G force vs Discomfort from inversions

Mean score: 6.44

Options	Count	Proportion
1 G force	2	 3.23%
2	3	 4.84%
3	1	 1.61%
4	2	 3.23%
5	16	 25.81%
6	4	 6.45%
7	12	 19.35%
8	6	 9.68%
9 Discomfort from inversions	16	 25.81%
Total survey collected	62	

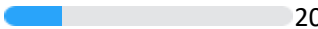
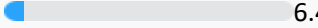
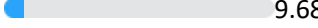
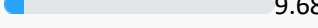
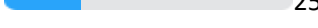
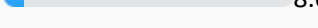
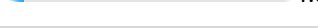
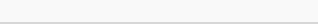

Q12: Speed vs Inversions

Mean score: 5.65

Options	Count	Proportion
1 Speed	5	 8.06%
2	2	 3.23%
3	5	 8.06%
4	3	 4.84%
5	19	 30.65%
6	6	 9.68%
7	4	 6.45%
8	7	 11.29%
9 Inversions	11	 17.74%
Total survey collected	62	


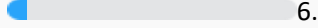
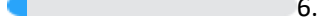
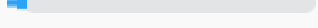
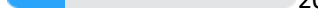
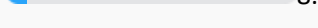



Q13: Speed vs Type

Mean score: 4.39

Options	Count	Proportion
1Speed	13	 20.97%
2	4	 6.45%
3	6	 9.68%
4	6	 9.68%
5	16	 25.81%
6	5	 8.06%
7	3	 4.84%
8	3	 4.84%
9 Type	6	 9.68%
Total survey collected	62	

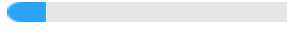
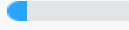

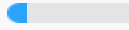
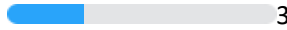
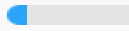
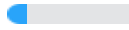
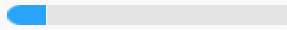
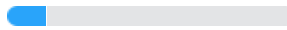
Q14: Speed vs Discomfort from G force

Mean score: 5.1

Options	Count	Proportion
1 Speed	12	 19.35%
2	4	 6.45%
3	4	 6.45%
4	1	 1.61%
5	13	 20.97%
6	5	 8.06%
7	6	 9.68%
8	10	 16.13%
9 Discomfort from G force	7	 11.29%
Total survey collected	62	

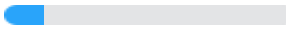

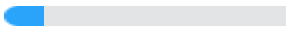

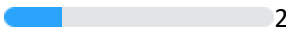
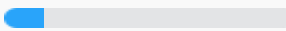


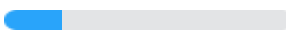
Q15: Speed vs Discomfort from inversions

Mean score: 5.37

Options	Count	Proportion
1 Speed	7	 11.29%
2	4	 6.45%
3	3	 4.84%
4	4	 6.45%
5	19	 30.65%
6	4	 6.45%
7	4	 6.45%
8	7	 11.29%
9 Discomfort from inversions	10	 16.13%
Total survey collected	62	

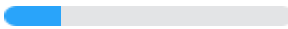



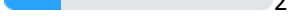

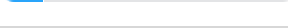


Q16: Type vs Inversions

Mean score: 5.23

Options	Count	Proportion
1 Inversions	8	 12.9%
2	3	 4.84%
3	7	 11.29%
4	3	 4.84%
5	14	 22.58%
6	8	 12.9%
7	5	 8.06%
8	2	 3.23%
9 Type	12	 19.35%
Total survey collected	62	


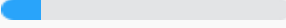
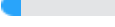

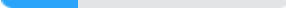
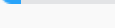



Q17: Type vs Discomfort from G force

Mean score: 4.84

Options	Count	Proportion
1 Type	12	 19.35%
2	5	 8.06%
3	4	 6.45%
4	2	 3.23%
5	14	 22.58%
6	5	 8.06%
7	10	 16.13%
8	2	 3.23%
9 Discomfort From G force	8	 12.9%
Total survey collected	62	

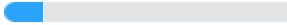








Q18: Type vs Inversions

Mean score: 4.63

Options	Count	Proportion
1 Type	11	 17.74%
2	8	 12.9%
3	4	 6.45%
4	1	 1.61%
5	17	 27.42%
6	5	 8.06%
7	5	 8.06%
8	5	 8.06%
9 Discomfort	6	 9.68%
Total survey collected	62	










Q19: Type vs Discomfort from Inversions

Mean score: 5.73

Options	Count	Proportion
1 Type	9	 14.52%
2	1	 1.61%
3	1	 1.61%
4	2	 3.23%
5	16	 25.81%
6	6	 9.68%
7	9	 14.52%
8	8	 12.9%
9 Discomfort from Inversions	10	 16.13%
Total survey collected	62	

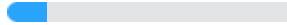
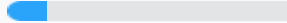
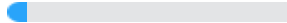
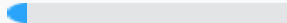
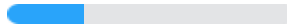
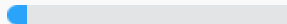
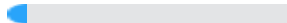
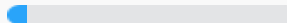
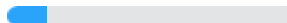
Q20: Type vs Discomfort from Inversions

Mean score: 5.82

Options	Count	Proportion
1 Type	5	 8.06%
2	3	 4.84%
3	5	 8.06%
4	1	 1.61%
5	15	 24.19%
6	6	 9.68%
7	8	 12.9%
8	7	 11.29%
9 Discomfort from Inversions	12	 19.35%
Total survey collected	62	

Q21: Discomfort from Inversions vs Discomfort from G force

Mean score: 4.79

Options	Count	Proportion
1 Discomfort from G force	10	 16.13%
2	7	 11.29%
3	4	 6.45%
4	3	 4.84%
5	17	 27.42%
6	5	 8.06%
7	3	 4.84%
8	4	 6.45%
9 Discomfort from Inversions	9	 14.52%
Total survey collected	62	