| For office use only | Team Control Number | For office use only |
|---|---|---|
| T1 ________________ | 8554 | F1 ________________ |
| T2 ________________ | | F2 ________________ |
| T3 ________________ | Problem Chosen | F3 ________________ |
| T4 ________________ | A | F4 ________________ |

**2018**

**HiMCM**

**Summary Sheet**

(Your team's summary should be included as the first page of your electronic submission.)

Type a summary of your results on this page. Do not include the name of your school, advisor, or team members on this page.


## Summary

There are various kinds of roller coaster around the world, which appeal to many fans. In order to recommend special roller coasters to fans, some online ranking of roller coasters come out in different website. Those ranking are based on either specific objective indicator or subjective experience results. therefore, we should create a quantitative algorithm with multiple objective indicators and build a new ranking of roller coaster, compare our ranking with two other online ranking systems. In addition, as the APPs have embedded in our lives, so we also should design a friendly APP which can recommend different roller coaster to fans based on their preferences.

In the database of COMAP_RollerCoasterData_2018.xlsx, we find that a small amount of data doesn't fit the facts and some indicators have more missing data, so we should process the data. Firstly, we correct the wrong data, secondly, we divide the sample into different categories by cluster analysis, and then we made up for the missing data by the mean value of those categories.

For creating a quantitative algorithm, we should set several assumptions. The roller coaster is operating normally, the ticket price of the roller coaster and the weather does not affect our ranking, the height, speed, length, the number of inversions, duration, G force and vertical angle are directly proportion to the fans' experience. We rank the roller coaster with entropy weight method. According to the character of the indicator in the database, we construct five indexes, including ups and downs index, vertigo index, sustainability index, stimulation index and nostalgia index. After our calculation, Steel Dragon 2000, Smiler, Kinda Ka, Leviathan, Fury 325, Millennium Force, Top Thrill Dragster, Intimidator305, Fujiyama and Steel Vengeance rank top 10, Our ranking are partly identical to the other online ranking, and finally, we cut the last 100 roller coasters in our ranking, and calculate the ranking again, we find the result is very identical with our original ranking, which means our result is robustness.

According to our classification and ranking, we design a friendly APP to recommend the roller coaster. We classify five indexes into excitement, intensity and vertigo, and then we take the fans' preference into account. The fans can input his preference into our APP, he will find the roller coaster which rank the top.

We have three advantages, including the correction of error data and supplement of missing data, entropy weight method based on the preference, and friendly app design. Surely, a coin has two sides, our disadvantages include the supplement of missing data leading to the results bias, and the results depending on our preference.


Keyword: Cluster Analysis, Entropy Weight Method, Personalized Comprehensive Evaluation Model

# Rank it! Go to it! Ride it!

Have you been feeling the blood pumping in your heart, eager to find a roller coaster to let epinephrine wake you up? Have you had night that you couldn’t sleep, thinking about flying in the sky, hearing the screaming howling? Or have you been disappointed by those dull, low-quality roller coasters, wanting to say good-bye to the most exciting game in this world? Forget about those subjective assessed roller coaster ranking websites. Rank&Roll, a newly published app with the latest algorithm, will provide you with the most objective ranking in the world!

To meet the personal preferences of the users, Duncan’s team has designed an application to help each individual to get personal roller coaster ranking list. When first signed up, clients will complete the following processes in order to provided information to sort out the most suitable list.

They have already given out a top 10 list of roller coasters, as shown below:

| Name | Score |
|---|---|
| Steel Dragon 2000 | 5.05 |
| Smiler | 4.00 |
| Kingda Ka | 3.88 |
| Leviathan | 3.70 |
| Fury 325 | 3.68 |
| Millennium Force | 3.50 |
| Top Thrill Dragster | 3.26 |
| Intimidator 305 | 3.25 |
| Fujiyama | 3.05 |
| Steel Vengeance | 2.96 |

How well will this app perform? I expect it will be quite popular among roller coaster fans. Let’s wait and see.

# Get Ready for Scream

## Content

# 1. Introduction

Technology is booming. After going through a boom in electronic devices, people gradually get used to it. Thus, groups of people return to theme parks, seeking thrills in places such as roller coasters. Roller coaster lovers naturally want to find the most appropriate roller coaster. Because there is no objective ranking, they can only be based on the Internet-given rankings. However, the online ranking is usually given by individual evaluators according to their personal feelings or it's just formed by a bunch of subjective votes.

In order to solve this problem, our group decided to use a series of algorithms to obtain the ranking based on the main objective data of roller coaster. Then to justify the objectivity of the rank we have worked out, we will compare and discuss the rating results and descriptions from your team's algorithm with two other ranking systems found online. We will design and develop the concept of a user-friendly application in order to meet the personal preference of each individual. The application will base on our previous algorithm and apply screening system thus provide a personal roller coaster ranking. At last, our team will write a one-page non-technical News Release describing our new algorithm, results, and app.

When making the ranking of roller coasters, we have to fill in the information first. We decide to use cluster analysis to use the data of similar roller coasters to approximate the information. After the completion of information, we used entropy method with two layers of indexes to make the ranking list. To design the app, we also used the two-layer structure to rank the roller coasters. The options are designed in this way to enhance users' experience. We also used the collaborative filtering to give recommendation to old users, whose long-term preference has been saved by the app. To achieve the fixed amount of roller coasters to recommend, we used k-nearest-neighbor method to get the recommended roller coasters.

# 2. Assumptions and Justifications

**Assumption 1:** No accident will happen when the roller coasters are operating. That is to say, all the roller coasters are absolutely safe, so safety factors will not be considered in our ranking system.

**Justification:** The possibility that a roller coaster has an accident is approximately one in 250 million, which is so small. Plus, safety factor is not included in the database, so we have no way to take them into consideration.

**Assumption 2:** Our ranking system do not consider the price of the tickets for the roller coasters.

**Justification:** In our ranking system, only factors related riders' experience will be taken into consideration.

**Assumption 3:** Weather factors is not included in the ranking system.

**Justification:** As roller coasters are everywhere around the world, the climate at their locations may vary drastically, which might also affect the riding experience. However, it is extremely hard for us to collect so much data in such a short period of time. So, we decide not to consider the climate in the ranking system.

**Assumption 4:** A rise in height, speed, length, the number of inversions, duration, G force and vertical angle will all increase the riders' excitement without causing any unpleasantness.

**Justification:** In the Entropy Method we use, all the variables and final score is positively correlated, because we assume that people ride the roller coasters to look for excitement.

# 3 Index Needed

**Table 1 Index Needed**

| Index | Meaning of the indexes |
|---|---|
| $s_i$ | ups and downs index of the *ith* roller coaster， $i = 1, \cdots, 300.$ |
| $n_i$ | vertigo index of the *ith* roller coaster |
| $h_i$ | sustainability index of the *ith* roller coaster |
| $t_i$ | stimulation index of the *ith* roller coaster |
| $Y_i$ | nostalgia index of the *ith* roller coaster |
| $a_j$ | preference coefficient of the *jth* index， $j = 1, \cdots, 5.$ |

# 4. Data Analysis and Processing

## 4.1 Screening and Revision of Erroneous Data

By classifying and screening the data in COMAP_RollerCoasterData_2018, there are eight types, among which wood and steel are the same as construction. There are three roller coasters involved. After checking on the Internet, the error data can be modified as shown in Table 2. In the process of consulting the data, we find that there are differences in very few numerical data. Since it is impossible to verify all the data, we use COMAP_RollerCoasterData_2018 in the later calculation.

**Table 2 Correction of Erroneous Type Data**

| Name | Original Type | After modification |
|---|---|---|
| Terminator Salvation | Wood | Sit down |
| Cedar Creek Mine Ride | Steel | Sit down |
| Corkscrew（1976） | Steel | Sit down |

## 4.2 Estimation of Missing Data for a Few Speed and Length

Through the analysis of the data, we find that there are a lot of missing data are drop, duration, G force and vertical angle data, and only a few speed and length data are missing. For the missing data of speed and length, we use the "approximate complement" method to estimate.

Let's assume that the speed of roller coaster A is missing. Choose the speed data of roller coaster B, whose height and length data is closest to roller coaster A, with the same type and construction of roller coaster A as the estimated value of speed data of A. Set $h_0, l_0$ for value of the height and length of roller coaster A and $h_m, l_m$ for data of all roller coasters with the same types and construction of roller coaster A. The roller coaster which meets the following formula

$$\min_k \sqrt{\left(\frac{h_k - h_0}{h_0}\right)^2 + \left(\frac{l_k - l_0}{l_0}\right)^2} \tag{1}$$

is B, and its corresponding speed value is $v_B$. The estimated value of speed for A is

$v_0 = v_B$.

Similarly, missing length data can be estimated. If the length data $l_B$ of roller coaster meets the following formula,

$$\min_k \sqrt{\left(\frac{h_k - h_0}{h_0}\right)^2 + \left(\frac{v_k - v_0}{v_0}\right)^2} \tag{2}$$

the missing data can be estimated as

$l_0 = l_B$.

### 4.3 Estimation of Missing Data for Drop, Duration, G force and Vertical angle

Wooden roller coaster is often built along the hillside with rarely inversion. By analyzing the data of fifty groups of wooden roller coasters in COMAP_RollerCoasterData_2018, there are only 5 of them have inversions, and the number of inversions is not more than 3. Because people always think that wood is not as strong as steel in their subjective impression, the stimulation of insecurity of wooden roller coaster in riding is unmatched by steel. At the same time, because of the construction along the hillside, the feeling of ups and downs is much more intense than that of steel, when people ride a wooden roller coaster. However, the dizziness caused by inversion and rotation of wood is not comparable to that of steel. Therefore, the gap between these two kinds of construction roller coaster is still very obvious.

#### 4.3.1 Estimation of Missing Data for Roller Coaster Made of Wood

Although the construction process of roller coaster is developing over time, the physical principles are the same. Therefore, ignoring the construction time factor, we selected 50 roller coasters with construction as wood, and carry out cluster analysis of roller coasters with height, speed and length as basic features. Fifty roller coasters' clustering maps are obtained as shown in Figure 1. We believe that the size and structure of roller coasters in the same category are similar, and their characteristic attributes should also be similar

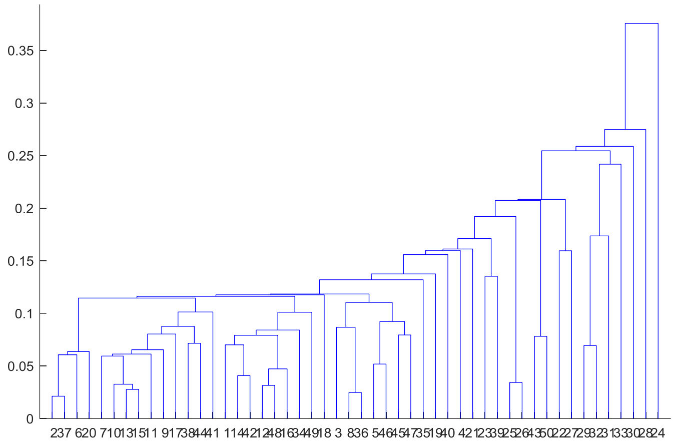


**Figure 1 Cluster Graph of 50 Wood Roller Coasters**

Of the 50 roller coasters, 18 are missing drop data, and the number of roller coasters missing data is less than the number of those not missing data. For roller coaster A with missing data, select in category with A the most similar roller coaster B with drop data, and use drop data of B to estimate that of A. If there are more than one such B, the average value of multiple drop data is used to estimate. The more the number of roller coasters not missing data is, the more detailed the clustering can be, and the closer the estimated value is to the real value.

For duration data, there are 14 roller coasters missing data. The situation is similar to that of drop data, and the data supplement method is the same as drop data.

There are 28 roller coasters missing data of vertical angle data, and the number of roller coasters with missing data is more than the number of those without missing data. It may appear that the roller coasters similar to the A in category are all missing vertical angle data. Therefore, we can consider to increase the number of roller coasters of the same kind as A by rough classification, so as to make roller coaster B without missing data as possible, which enables vertical angle data to be estimated. But this may increase the estimation error.

Only 7 of the 50 roller coasters have G force data, so it doesn't work to estimate with known data of the same category after cluster analysis. Because the 7 data may be a cluster, we should drop it. If the data is missing, it can be estimated according to the known data. Otherwise, other methods shall be used. By consulting relevant information on the internet, the G force of roller coaster will not exceed 2.5G without considering the car's own acceleration system. Because G force is an important indicator of the performance of roller coaster, it should be open data. If there is no relevant data or relevant data is difficult to find, it is reasonable that the roller coaster carriage does not have an acceleration system. In that case, its G force will not exceed 2.5G, and we might as well assume that it is 2G.

### 4.3.2 Estimation of Missing Data for Roller Coaster Made of Steel

The same method used for roller coasters made of wood can be used when estimating the missing data of roller coasters made of steel. We find from the data that 136 of 250 roller coasters made of steel have inversions and 114 not, so we should divide them into two parts according to whether they have inversions or not. For each part, the missing data can be estimated according to the method in Section 4.3.1.

When the roller coaster is built without inversion, we use height, speed, length as the three basic features for cluster analysis, while we use height, speed, length, number of

Inversion as the four basic features for cluster analysis, when it is built with inversion. The two-part cluster analysis schematic diagram is shown in Figure 2 and Figure 3, in which we only show the clustering results of the first 30 roller coasters in each part.

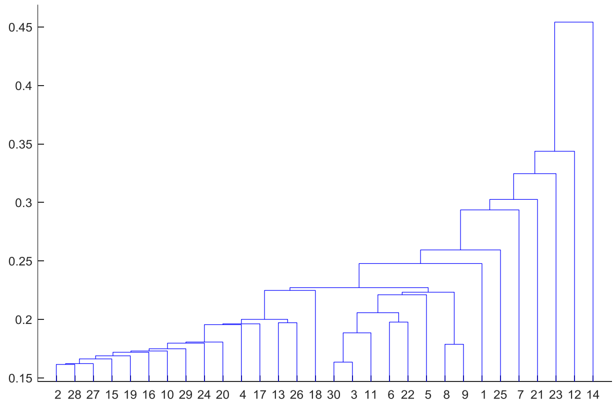


**Figure 2 Cluster Diagram with Inversion**

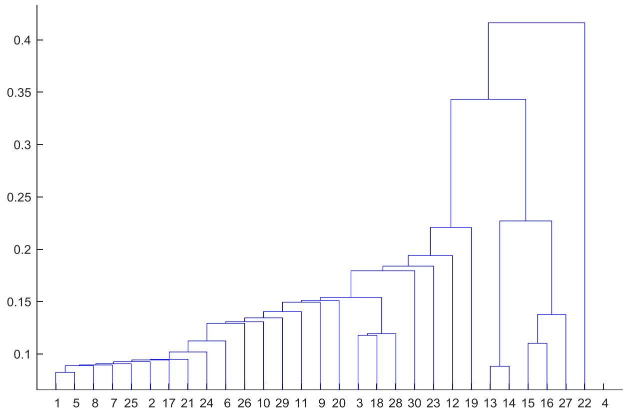


**Figure 3 Cluster Diagram without Inversion**

# 5. Evaluation Based on Entropy Weight Method

## 5.1 Analysis and Quantification of Characteristics

Through the method of section 4, we have revised and supplemented the data. For 300 roller coasters, each evaluation object contains 12 characteristics: type, construction, Year/Date, speed, length, inversions, number of inversions, drop, duration, G force and vertical angle. By consulting the relevant literature, passengers' ride experience varies when there are reversions or not. Roller coaster without inversions tend to use the alternating transformation of overweight and weightlessness caused by rapid changes of acceleration as the main stimulation. While those with inversions tend to use inversions and the vertigo caused by rotation as the main stimulation. The inversions characteristic can be 'yes' quantized to 1, and the 'no' quantification is 0.

Through the information online and the analysis of data, roller coasters made of wood rarely have inversions, while steel-made roller coaster does not. By comparing 50 wood-made roller coasters and 114 steel-made roller coasters which don't have inversions, The average data of height, speed and length are obtained in Table 3.

**Table 3 Comparison of Mean Values Between the Both**

| Construction | Height（feet） | Speed（mph） | Length（feet） |
|---|---|---|---|
| Wood | 110.22 | 56.045 | 3726.77 |
| Steel | 140.48 | 59.90 | 3040.80 |

Basically, these data can be found on the Internet, not estimated. As you can see from table 3, those steel-made roller coasters without inversions are higher and faster than wood-made ones, thus the average degree of stimulation is higher. The length data of wood-made roller coaster is higher than steel-made, which means that the average duration of stimulation from wood-made roller coaster is longer. The degree of Steel-made ones higher than the wood-made can be depicted by the following formula:

$$\left(\frac{140.48-110.22}{110.22}+\frac{59.9-56.045}{56.045}\right)\Big/2=0.172 \tag{3}$$

So when we quantify 'steel' to 1, the quantization value of 'wood' is 0.828.

According to the information on the internet, the sensation intensity of human body is different when choosing different types of roller coasters, even under the same level of stimulation. This is because the ability to control our body when experiencing different types of roller coasters varies a lot. According to the intensity of sensory strength from weak to strong, the arrangement and quantitative rules of the 6 types can be seen in Table 4.

**Table 4 Quantization of These 6 Types**

| Type | Sit down | Stand up | Wing | Flying | Suspended | Inverted |
|---|---|---|---|---|---|---|
| Quantization | 1 | 2 | 3 | 4 | 5 | 6 |

## 5.2 Evaluation Feature Extraction

### 5.2.1 Ups and Downs Index

According to the analysis of 5.1, we know that when roller coaster does not have inversions, the stimulation mode for passengers is mainly the alternating transformation of overweight and weightlessness caused by rapid change of acceleration. Therefore, we define the index of ups and downs, which can be characterized by characteristic speed, drop, G force and vertical angle. These four characteristics are positively correlated with this index.

Entropy represented, as we know, the overall degree of disorder of the whole system. The larger entropy is, the more information it would carry, which means a larger degree of disorder. Thus, a large entropy would mean a greater uncertainty of an index. Thus, a smaller entropy would mean a greater accuracy and contribution to the whole model. With this method, we can determine the factors that contribute most to the assessment equation, allowing us to work out the weight of each factor. At last, we can use the equation to give an overall assessment of the roller coasters.

When the system may be in several different states and the probability of each state is $p_i$（$i=1,2,\cdots,m$）, the entropy of the system can be defined：

$$e=-\frac{1}{\ln m}\sum_{i=1}^{m} p_i \ln p_i \tag{4}$$

Entropy weight method is an objective weighting method. In the specific use process, the entropy weight method calculates the entropy weight of each feature according to the degree of variation of each feature, and then amends the weight of each feature through the entropy weight, thus objectively obtaining the weight of each feature. We set $j=1,\cdots,4$ corresponding to characteristics: speed, drop, G force and vertical angle. $p_{ij}$ represents the proportion of the *ith* roller coaster about the *jth* characteristic:

$$p_{ij}=\frac{a_{ij}}{\sum_{i=1}^{300}a_{ij}},\ i=1,2,\cdots,300,\ j=1,\cdots,4 \tag{5}$$

Where, $a_{ij}$ represents the original data of the *ith* roller coaster's *jth* characteristic.

Calculate the entropy of the *jth* characteristic:

$$e_j=-\frac{1}{\ln 300}\sum_{i=1}^{300}p_{ij}\ln p_{ij},\ j=1,\cdots,4 \tag{6}$$

Calculate coefficient of variation of the *jth* characteristic:

$$g_j=1-e_j,\ j=1,\cdots,4 \tag{7}$$

Thus the weight of the *jth* characteristic:

$$w_j=\frac{g_j}{\sum_{j=1}^{4}g_j},\ j=1,\cdots,4 \tag{8}$$

So we can get the *ith* roller coaster's index of ups and downs:

$$s_i=\sum_{j=1}^{4}w_j p_{ij},\ i=1,2,\cdots,300 \tag{9}$$

After calculation, the weight coefficient of speed, drop, G force and vertical angle is 0.1864,0.5489,0.1967,0.0680 respectively, therefore

$$s_i=0.1864p_{i1}+0.5489p_{i2}+0.1967p_{13}+0.0680p_{14},\ i=1,2,\cdots,300 \tag{10}$$

From the weight coefficient, we can see that drop feature plays the most important role in the ups and downs index.

### 5.2.2 Vertigo Index

Those roller coasters with inversions tend to use inversions and the vertigo caused by rotation as the main stimulation. Thus we define the index of vertigo which can be described by characteristics inversions and number of inversion. But these two characteristics have strong correlation. As long as number of inversions is not zero, inversions must be 1. Otherwise, as long as number of inversions is 0, inversions must be 0, so only inversions index can be adopted. We need to normalize the number of inversions. The vertigo index $\tilde{n}_i$ standardized from the *ith* roller coaster' number of inversions is calculated as follows:

$$\tilde{n}_i=\frac{n_i-\min_i n_i}{\max_i n_i-\min_i n_i} \tag{11}$$

Where $n_i$ is the number of inversions of the *ith* roller coaster.

### 5.2.3 Sustainability Index

People who like roller coaster often like to pursue stimulation, and an important aspect of pursuing stimulation is the duration of stimulation, which is closely related to the scale of roller coaster. Therefore, the index can be characterized by height, length and duration, and the sustainability index is in positive phase with the three characteristics.

As the calculation of the ups and downs index, the value of the *ith* roller coaster's sustainability index $h_i$ can be obtained by using the Entropy Weight Method.

After calculation, the weight coefficient of height, length and duration is 0.0057,0.0032,0.9911 respectively, therefore

$$h_i = 0.0057 p_{i1} + 0.0032 p_{i2} + 0.9911 p_{13}, \ i = 1, 2, \cdots, 300 \tag{12}$$

From the weight coefficient, we can see that duration feature plays the most important role in the sustainability index.

### 5.2.4 Stimulation Index

Through the analysis of 5.1, we can see that when people ride roller coaster made of wood or steel, their subjective feelings are different. At the same time, when people choose different types of roller coasters, their subjective feelings of stimulus are different even when the other conditions are the same. Therefore, the stimulation index is depicted by characteristics construction and type.

According to the quantitative feature of type, the greater the value, the higher the corresponding stimulus sensitivity. In order to facilitate calculation, we need to normalize the type data. The type standardized data of roller coaster is calculated as follows:

$$\tilde{t}_i = \frac{t_i - \min_i t_i}{\max_i t_i - \min_i t_i}, \quad i = 1, \cdots, 300. \tag{13}$$

Where $t_i$ represents the type of the *ith* roller coaster. Set $c_i$ as the construction data of the *ith* roller coaster .Through the Entropy Weight Method, we can get the *ith* roller coaster's stimulation index $r_i$ .

### 5.2.5 Nostalgia Index

These 300 roller coasters are from 1924 to 2018. The time span is 94 years. The construction time of the *ith* roller coaster is indicated by $y_i$ .Set $y_i' = 2018 - y_i$ ,and the nostalgia index is :

$$Y_i = \frac{y_i' - \min_i y_i'}{\max_i y_i' - \min_i y_i'}, \quad i = 1, \cdots, 300. \tag{14}$$

This is a very personalized index.

## 5.3 Establishment and Solution of Evaluation Model

### 5.3.1 Comprehensive Evaluation Model Based on Passengers' Preference Degree

Section 5.2 proposes five evaluation indexes and the comprehensive evaluation here is based on these five aspects. We can design preference coefficient to match different preference levels and the comprehensive evaluation index of 300 roller coasters is:

$$f_i = a_1 \tilde{s}_i + a_2 \tilde{n}_i + a_3 h_i + a_4 r_i + a_5 Y_i, \ i = 1, \cdots, 300. \tag{15}$$

Where coefficients $a_1, \cdots, a_5$ represent preferences for different indexes. $a_1, \cdots, a_5 \geq 0$ and $\sum_{i=1}^{5} a_i = 1$. The larger the value of $a_i$, the more important its corresponding index is in the evaluation system. In order to eliminate the influence of dimension, it is necessary to normalize the data before calculating the comprehensive evaluation index. $\tilde{s}_i$ is normalized from $s_i$:

$$\tilde{s}_i = \frac{s_i - \min_i s_i}{\max_i s_i - \min_i s_i},\ i = 1, \cdots, 300. \tag{16}$$

When evaluating 300 roller coasters, the larger the value of $f_i$, the higher the corresponding roller coaster ranks. Based on the preference coefficient, we can make a personalized ranking scheme according to the different preferences of different passengers, which provides operational feasibility for us to design apps.

However, when we do not consider the personalized ranking scheme with preference degree and only hope to rank by characteristic data of roller coaster, we can still use the entropy weight method to determine the objective weighting coefficients of five indicators. So that the five indicators can be combined into a comprehensive evaluation index, ranking by calculating their comprehensive scores.

### 5.3.2 The Ranking Result and Comparison with Online Rankings

Using the entropy weight method to determine the weighting coefficient of the 5 indexes, we get a comprehensive evaluation index. After calculation, 300 roller coasters' comprehensive index scores are obtained, among which the top 10 roller coasters are listed in Table 5.

**Table 5 Overall Ranking of Roller Coasters**

| Name | Score |
|---|---|
| Steel Dragon 2000 | 5.04881085 |
| Smiler | 4.00226545 |
| Kingda Ka | 3.88475056 |
| Leviathan | 3.70368266 |
| Fury 325 | 3.68319054 |
| Millennium Force | 3.50044987 |
| Top Thrill Dragster | 3.26340141 |
| Intimidator 305 | 3.25427553 |
| Fujiyama | 3.05363808 |
| Steel Vengeance | 2.96345454 |

### 5.3.3 Comparison with Online Ranking

We found two online roller coaster rankings and made a comparison with our ranking, as shown in Table 6.

**Table 6 Result and Comparison with Online Rankings**

| Rank | Online1 | Online2 | Ours |
|---|---|---|---|
| 1 | Millennium Force | Kingda Ka | Steel Dragon 2000 |
| 2 | Steel Vengeance | Dodonpa | Smiler |
| 3 | Top Thrill Dragster | X2 | Kingda Ka |
| 4 | Maverick | Top Thrill Dragster | Leviathan |
| 5 | El Toro | Steel Dragon 2000 | Fury 325 |
| 6 | Fury 325 | Tower of Terror | Millennium Force |
| 7 | Intimidator 305 | Millennium Force | Top Thrill Dragster |
| 8 | The Voyage | Intimidator | Intimidator 305 |
| 9 | Kingda Ka | Thunder Dolphin | Fujiyama |
| 10 | Apollo's Chariot | Eejanaika | Steel Vengeance |

Note: two groups of online ranking data are derived from references.

After applying the algorithm presented in the previous text, we have made the following ranking.

To justify our ranking's reliability, we have searched several ranking systems online and have picked the two most authentic ranking to make comparison with our ranking.

The first ranking we will present is from the roller coaster system in "Ranker" community. Ranker as a ranking community, let people vote for everything, so the ranking on its web will be rather reliable and representative for the people who voted.

In the two rankings, there are 5 repetitions in the "top 10", which are Fury 325, Millennium Force, Kingda Ka, Intimidator 305 and Top Thrill Dragster. In our ranking, the order of those 5 from front to back is Kingda Ka, Fury 325, Millennium Force, Top Thrill Dragster and Intimidator 305. In another ranking, the sequence is Millennium Force, Top Thrill Dragster, Fury 325, Intimidator 305 and Kingda Ka.

The sequences of those two quite correspond except Kingda Ka.

Our second reference is from "THE TOP 10s". The same as "Ranker", "THE TOP 10s" is a ranking system which allows people to vote. As a result, this source is respectively reliable and Have reference value.

In the two rankings, there are 5 repetitions in the "top 10", which are Millennium Force, Kingda Ka, Intimidator 305, Steel Dragon 2000 and Top Thrill Dragster. In our ranking, the order of those 5 from front to back is Steel Dragon 2000, Kingda Ka, Millennium Force, Top Thrill Dragster and Intimidator 305. In another ranking, the sequence is Kingda Ka, top Thrill Dragster, Steel Dragon 2000, Millennium Force and intimidator 305.The sequences of those two quite correspond except Kingda Ka.

Moreover, as shown in our form, there are three repetitions which are Millennium Force, Top Thrill Dragster and Kingda Ka in the two ranking lists found online. This means that the repeated three roller coasters are confirmed to be eligible to be put in such high rankings. The three, furthermore, can all be found in our top lists, so a further confirmation that justify our objectivity is made.

Then comes the question: How can we explain the difference between the votes and our list? To answer this question, we have come up several possible explanations.

First, the ranking is based on the voters' personal experience. As we all know, different positions in the roller coaster can lead to different experiences. The more you sit back, the bigger the thrill. We have mentioned in our assumption that we would not consider the influence of positions of seats, so this might be a factor which lead the difference.

Moreover, the price of the tickets might also be an influential factor which may affect the votes. Sometimes a good roller coaster may not receive much votes for its high price. Likewise, we have mentioned that the price of the roller coaster will not make any influence in the assumption. This may also be a possible factor that causes the difference.

## 6. Sensitivity Analysis

In the above analysis, we have got the Roller coaster ranking by the comprehensive evaluation of multiple index with objective weight. For testing the robustness of our result, we use the new sample which drops the last 100 Roller coasters of our ranking, and then we calculate it again. the comparative result is shown in Table 7.

**Table 7 The Outcome of Two Different Samples**

| 300 Samples | | | 200 Samples | | |
|---|---|---|---|---|---|
| Name | Rank | Score | Name | Rank | Score |
| Steel Dragon 2000 | 1 | 5.05 | Steel Dragon 2000 | 1 | 5.03 |
| Smiler | 2 | 4.00 | Kingda Ka | 2 | 4.03 |
| Kingda ka | 3 | 3.88 | Smiler | 3 | 3.96 |
| Leviathan | 4 | 3.70 | Leviathan | 4 | 3.76 |
| Fury 325 | 5 | 3.68 | Fury 325 | 5 | 3.69 |
| Millennium Force | 6 | 3.50 | Millennium Force | 6 | 3.51 |
| Top Thrill Dragster | 7 | 3.26 | Top Thrill Dragster | 7 | 3.40 |
| Intimidator 305 | 8 | 3.05 | Intimidator 305 | 8 | 3.32 |
| Fujiyama | 9 | 2.96 | Fujiyama | 9 | 3.05 |
| Steel Vengeance | 10 | 2.95 | Coaster Through the Clouds | 10 | 2.99 |

Through the comparison, we find that the ranking of two samples is rather identical, so our method is robust and convincing.

# 7. Design for the App

### 7.1 Explanation of the Ranking System in App

In Section 5, our evaluation algorithm fully considers the individual ranking requirements of roller coaster for different passengers because of their different preferences. To this end, we design the app to achieve personalized recommendation based on the preferences of passengers. Users only need to fill in a few related preferences, then we can personalize the sorting according to the algorithm and give the recommended roller coaster.

Considering that some passengers may have some very special preferences, for example, some passengers are paranoid about taking roller coaster made of wood, while others only take steel. The comprehensive evaluation model we have established is not focused on the distinction, so it can be listed as an option alone. It can be embodied in the following aspects.

1. Do you prefer wood to steel? There are three options -- wood, steel and both.

If the passenger chooses wood, then we will choose in the 50 roller coasters made of wood. If the passenger chooses steel, then we will choose in the 250 roller coasters made of steel. If passengers choose “both”, then we will choose in all 300 roller coaster.

2. Which type of roller coaster do you prefer? There are two options -- sit down and others.

This is because 240 of the 300 roller coasters are sit down, up to 80%. If passengers choose sit down, we will further select the roller coasters which meets the first option and whose type is sit down. If the passenger chooses others, we will choose in the other 60 roller coaster.

3. Do you have nostalgia? Do you prefer the old roller coaster to the modern one? There are three options -- ancient, modern and both.

If the passenger chooses the old one, his nostalgia degree is higher, and the parameter $a_5$ can take a larger value. Because this index is not an important index when evaluating roller coaster, its value should not more than 0.5. If the passenger chooses modern, the value of $a_5$

is smaller. If passengers choose "both", it means that the index has little impact on passengers, thus $a_5 = 0$.

4. Do you have any request for the ride time of the roller coaster? The three option is longer, shorter, and no demanding.

If the passenger chooses a longer time, then the passenger pursues the persistence in the stimulation process, and the value of the parameters $a_3$ can be larger; if the passenger chooses a shorter time, it can be smaller; if the passenger does not require it, then $a_3 = 0$. Although this index is important, its size should not larger than the parameters of ups and downs index and vertigo index.

5. Do you prefer fast alternating A for weightlessness and overweight caused by ups and downs, to B for vertigo caused by inversions and rotation? The three options are A, B and both.

If the passenger chooses "both", the parameter is $a_1 = a_2$, if the passenger chooses A, then there is $a_2 = 0$, if the passenger chooses B, then there is $a_1 = 0$.

After sorting out the selected roller coasters screened through 1 and 2, the evaluation model is

$$f_j = a_1\tilde{s}_j + a_2\tilde{n}_j + a_3h_j + a_5Y_j$$

Where the non-negative parameters $a_1, a_2, a_3, a_5$ may be 0 either, and only need to satisfy the condition that their sum is 1. $j$ is the *jth* object of our roller coaster.

According to the calculation results, the individualized sorting of the selected roller coasters can be realized, and the best recommendation can be made to the passengers. For the simplicity of UI design interface and better customer self-evaluation, we classify ups and downs index, vertigo index, sustainability index, stimulation index and nostalgia index into tree feelings (excitement, intensity, vertigo), and the result of classification is in figure 4.

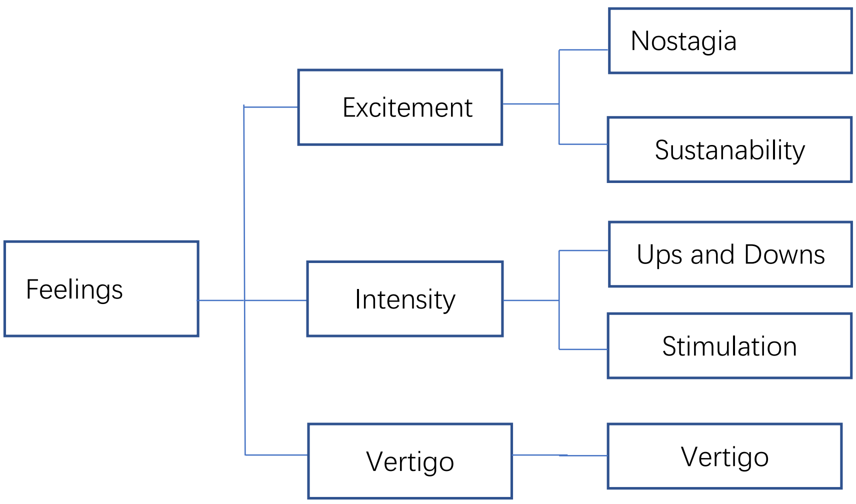


**Figure 4 The Classification of Five Feeling Indexes**

Our thinking figure of UI design is shown in Figure 5.

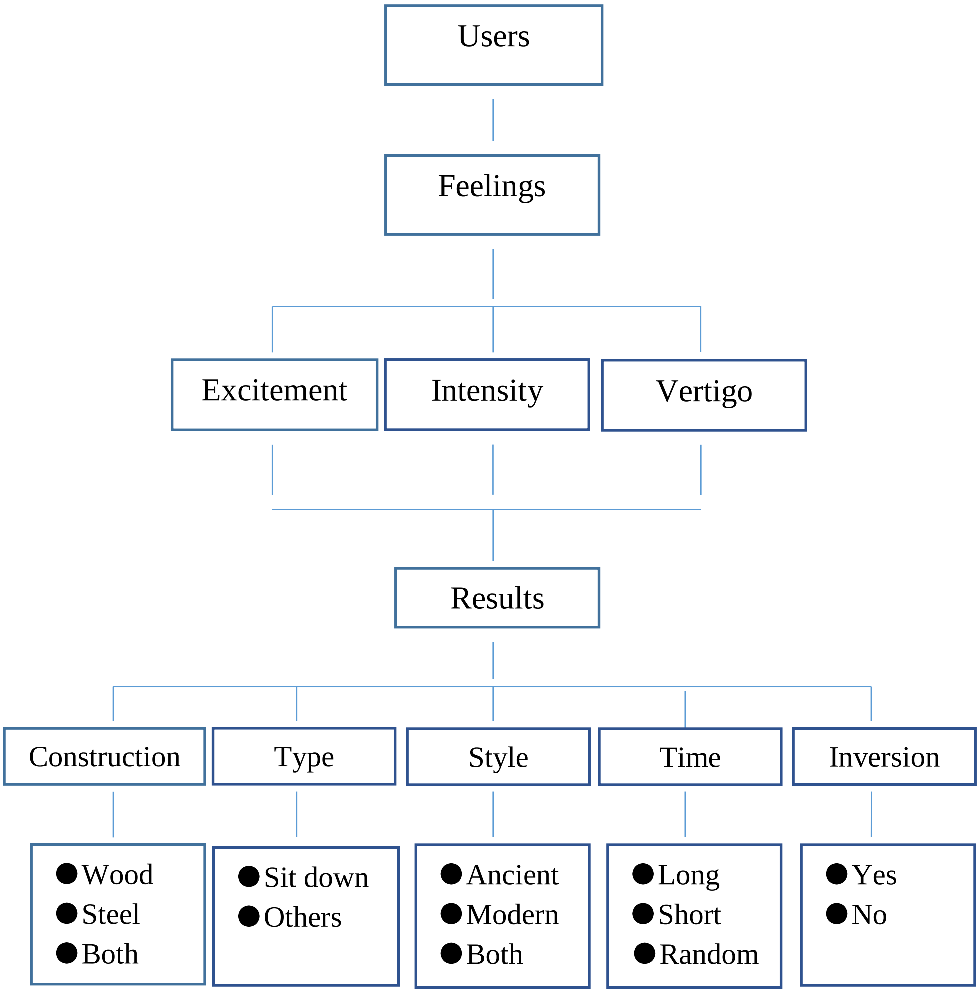


**Figure 5 Option Design**

### 7.2 Qualification for the Design

We define "User Friendly" as the clarity of the options, the numbers of options, and the speed a user can find the options. So, we follow the rules set by ISO (International Organization for Standardization) to determine the usability.

1.Learnability: How easy is it for users to accomplish basic tasks the first time they encounter the design?

2.Efficiency: Once users have learned the design, how quickly can they perform tasks?

3.Memorability: When users return to the design after a period of not using it, how easily can they re-establish proficiency?

4. Satisfaction: How pleasant is it to use the design?

We will follow the rules set above to qualify our design of app.

1.Learnability: Our app has only two major layers: a layer of feelings and a layer of construction and type. When the user sets his final preference, our app would provide an auto ranking, with almost no cost of learning.

2.Efficiency: Again, with only two major layers, the only thing to do to generate an auto ranking is two click of hand. Sometimes a user is not very sure what kind of feelings he wants to experience in the riding, so we also provide an overall choice, making them decide less and increasing their efficiency.

3.Memorability: Our structure of options is clear and involves many properties of roller coaster, so with a clear instruction, when a user see the app, he would instantly understand and re-establish proficiency.

4.Satisfaction: Though every user has his objective opinion of what is essential for a roller coaster to be satisfactory, our option design includes all the properties of roller coasters, which can satisfy almost every user with all the combinations possible.

### 7.3 UI Design of the App

We use a free UI designing web to create a simple UI design, as shown below in Figure 6.

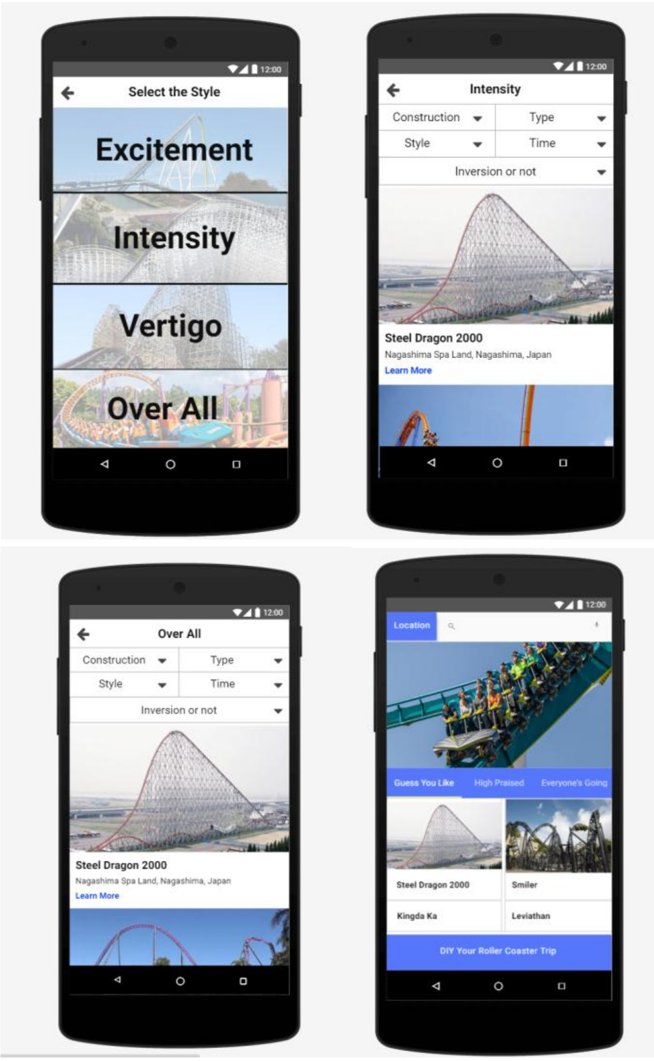


**Figure 6 Graphs of UI Design**

Note: all the graphs come from Internet search

Examples of rankings after choosing preference, see Figure7-Figure 9.

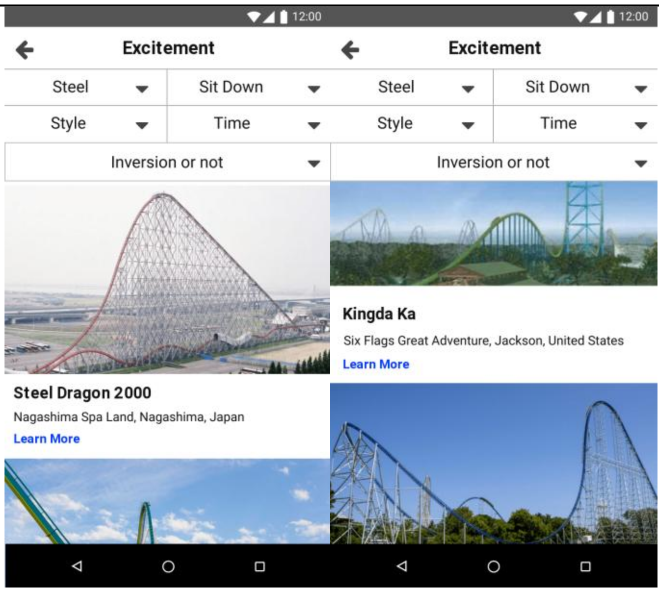


**Figure 7 Excitement + Steel + Sit Down**

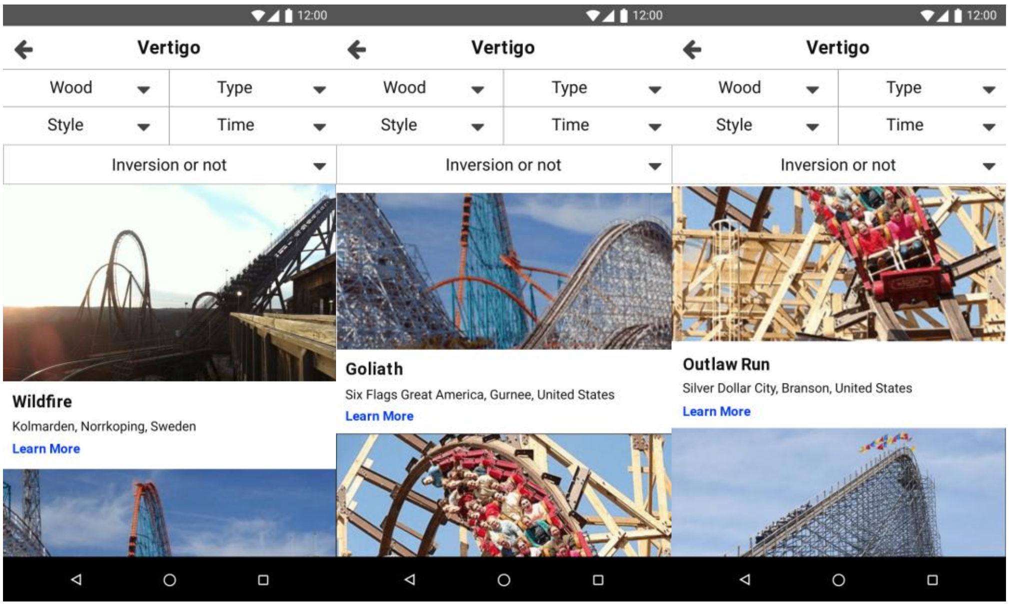


**Figure 8 Vertigo + Wood + All Types**

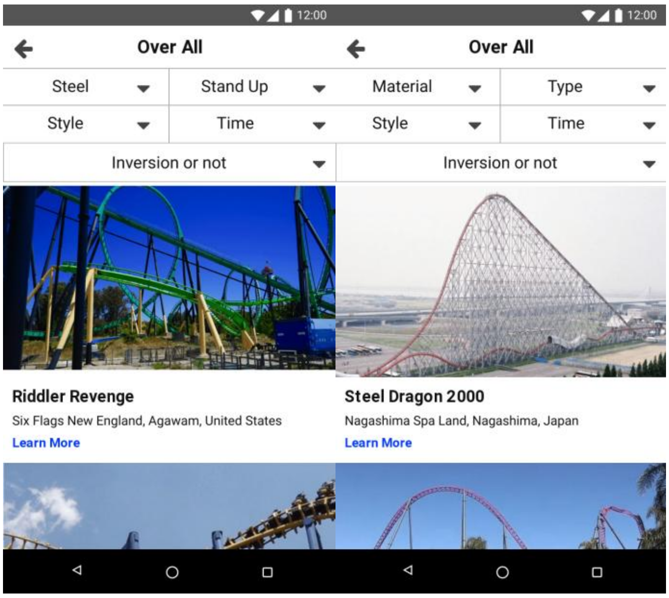


**Figure 9 Overall + Steel + Stand Up**

These three examples show how will the app generate the ranking based on personal preference.

# 8. Strength and Weakness of the model

## 8.1 Strength

When handling with the missing information, we didn't just delete them. Instead, we use cluster analysis to approximate the missing information. When handling with the missing information, we didn't just delete them. Instead, we use cluster analysis to approximate the missing information.

We use known data in the same category to estimate unknown data. Before clustering, we cluster according to different roller coaster characteristics of different constructions, which improves the accuracy of clustering, so the accuracy of missing data estimation is improved.

When establishing the evaluation model, we put forward five first-level indicators, among which there are several second-level indicators under the three indicators of ups and downs, sustainability and stimulation index. We use an entropy weight method to determine the weight coefficients of the second-level indicators. The weight coefficients determined by the entropy weight method are objective. These five indicators reflect the characteristics and performance of roller coaster from various aspects. Finally, a comprehensive evaluation index is established by preference coefficient. The preference coefficients in the model vary from person to person, which provides a feasible option for us to design apps to meet the individual needs of passengers.

## 8.2 Weakness

Although we have found a good way to supplement the missing data, there is a certain gap between the missing data and the real data, which will make the evaluation results inaccurate, especially in the G force data supplement. Because we build an algorithm through preference coefficients to facilitate the realization of passengers' personalized needs in apps, the ranking of 300 roller coasters is largely dependent on our preferences. Therefore, there will be great changes in results of sorting, when we focus on different angles, or the degree of preference for the same angle is different.

# 9. Conclusion

We built a function to assess the roller coaster with entropy method. When handling the incomplete information, we used cluster analysis and successfully filled the incomplete information. The clustering process is a fairly good cluster, with a cophenetic coefficient close to 1. The entropy method is an efficient way to determine whether an index is accurate and the contribution to the system. We also designed a app that perfectly fit the ISO standard of usability, with a visual UI on the phone. Of course, the model is far from perfect, with a positive coefficient for G force, since some people will have a comfortlessness under high g force condition.

# 10. Reference

[1]. https://www.ranker.com/crowdranked-list/best-roller-coasters?ref=lzyrltdlstszerg_rr
[2]. https://www.thetoptens.com/most-extreme-roller-coasters/

# 11.Appendix: The Ranking of All Roller Coaster

| Rank | Name | Score | Rank | Name | Score |
|---|---|---|---|---|---|
| 1 | Steel Dragon 2000 | 5.05 | 34 | Superman/ la Atracción de Acero | 2.27 |
| 2 | Smiler | 4.00 | 35 | Kraken | 2.26 |
| 3 | Kingda Ka | 3.88 | 36 | Bizarro | 2.25 |
| 4 | Leviathan | 3.70 | 37 | Intimidator | 2.13 |
| 5 | Fury 325 | 3.68 | 38 | Katun | 2.11 |
| 6 | Millennium Force | 3.50 | 39 | Goliath | 2.05 |
| 7 | Top Thrill Dragster | 3.26 | 40 | Diamondback | 1.92 |
| 8 | Intimidator 305 | 3.25 | 41 | Desperado | 1.88 |
| 9 | Fujiyama | 3.05 | 42 | Valravn | 1.83 |
| 10 | Steel Vengeance | 2.96 | 43 | Superman el Último Escape | 1.83 |
| 11 | Coaster Through the Clouds | 2.95 | 44 | 10 Inversion Roller Coaster | 1.81 |
| 12 | Viper | 2.95 | 45 | Colossus | 1.81 |
| 13 | Alpengeist | 2.91 | 46 | Superman: Escape from Krypton | 1.80 |
| 14 | Dragon Mountain | 2.89 | 47 | Dragon's Run | 1.79 |
| 15 | Medusa | 2.87 | 48 | Vortex | 1.78 |
| 16 | Banshee | 2.79 | 49 | Monster | 1.77 |
| 17 | Formula Rossa | 2.76 | 50 | Takabisha | 1.76 |
| 18 | Titan | 2.69 | 51 | Pyrenees | 1.73 |
| 19 | Scream! | 2.64 | 52 | Mamba | 1.68 |
| 20 | Dragon Khan | 2.61 | 53 | Big One | 1.68 |
| 21 | Soaring Dragon & Dancing Phoenix | 2.58 | 54 | Steel Force | 1.66 |
| 22 | GateKeeper | 2.55 | 55 | Eejanaika | 1.65 |
| 23 | Silver Star | 2.53 | 56 | Incredible Hulk | 1.64 |
| 24 | Riddler's Revenge | 2.52 | 57 | Nitro | 1.62 |
| 25 | Montu | 2.47 | 58 | Superman the Ride | 1.59 |
| 26 | Kumba | 2.44 | 59 | Wild Thing | 1.58 |
| 27 | Hyperion | 2.41 | 60 | Tower of Terror II | 1.53 |
| 28 | Shambhala | 2.38 | 61 | Beast | 1.52 |
| 29 | Superman Krypton Coaster | 2.34 | 62 | Phaethon | 1.52 |
| 30 | Helix | 2.33 | 63 | Schwur des Kärnan | 1.48 |
| 31 | Altair | 2.31 | 64 | Raptor | 1.43 |
| 32 | Crazy Coaster | 2.31 | 65 | Flight of the Phoenix | 1.42 |
| 33 | Velikolukskiy Myasokombinat-2 | 2.31 | 66 | Red Force | 1.40 |

(Continued Table)

| Rank | Name | Score | Rank | Name | Score |
|---|---|---|---|---|---|
| 67 | Flying Aces | 1.39 | 104 | Goliath | 0.60 |
| 68 | Afterburn | 1.37 | 105 | OCT Thrust SSC1000 | 0.59 |
| 69 | Big Apple Coaster | 1.36 | 106 | Batman - The Dark Knight | 0.57 |
| 70 | Gao | 1.35 | 107 | Jupiter | 0.53 |
| 71 | Silver Bullet | 1.35 | 108 | Goliath | 0.51 |
| 72 | Behemoth | 1.34 | 109 | T Express | 0.44 |
| 73 | Hydra the Revenge | 1.33 | 110 | Talon | 0.39 |
| 74 | Flash | 1.29 | 111 | Incredicoaster | 0.35 |
| 75 | Cannibal | 1.29 | 112 | Phantom's Revenge | 0.33 |
| 76 | Hyper Coaster | 1.27 | 113 | iSpeed | 0.31 |
| 77 | Wildfire | 1.26 | 114 | Iron Rattler | 0.28 |
| 78 | Dinoconda | 1.25 | 115 | Fly the Great Nor'Easter | 0.27 |
| 79 | Happy Angel | 1.23 | 116 | Lightning Rod | 0.27 |
| 80 | Raging Bull | 1.21 | 117 | Batman The Ride | 0.26 |
| 81 | Rougarou | 1.19 | 118 | Batman The Ride | 0.26 |
| 82 | X2 | 1.18 | 119 | Batman The Ride | 0.26 |
| 83 | Ride of Steel | 1.12 | 120 | Ultimate | 0.25 |
| 84 | Twisted Colossus | 1.09 | 121 | Batman The Ride | 0.23 |
| 85 | Superman - Ride Of Steel | 1.09 | 122 | Swarm | 0.19 |
| 86 | Voyage | 0.99 | 123 | Maverick | 0.18 |
| 87 | Griffon | 0.96 | 124 | El Toro | 0.14 |
| 88 | Goudurix | 0.92 | 125 | Goliath | 0.11 |
| 89 | Magnum XL-200 | 0.87 | 126 | Stunt Fall | 0.10 |
| 90 | Tatsu | 0.86 | 127 | Blue Hawk | 0.10 |
| 91 | Thunder Dolphin | 0.83 | 128 | Storm Runner | 0.09 |
| 92 | blue fire Megacoaster | 0.83 | 129 | Extreme Rusher | 0.07 |
| 93 | Velikolukskiy Myasokombinat | 0.82 | 130 | Hades 360 | 0.05 |
| 94 | Mako | 0.82 | 131 | Quimera | 0.03 |
| 95 | Do-Dodonpa | 0.81 | 132 | Batman The Ride | 0.01 |
| 96 | Apollo's Chariot | 0.80 | 133 | Wicked Cyclone | -0.01 |
| 97 | Firehawk | 0.75 | 134 | Goliath | -0.04 |
| 98 | Manta | 0.73 | 135 | Great White | -0.05 |
| 99 | Patriot | 0.69 | 136 | Skyrush | -0.05 |
| 100 | Batwing | 0.68 | 137 | Outlaw Run | -0.06 |
| 101 | Fahrenheit | 0.61 | 138 | Batman the Ride | -0.06 |
| 102 | Bullet Coaster | 0.60 | 139 | Desafio | -0.06 |
| 103 | Soaring with Dragon | 0.60 | 140 | Firewhip | -0.06 |

(Continued Table)

| Rank | Name | Score | Rank | Name | Score |
|---|---|---|---|---|---|
| 141 | Kong | -0.06 | 174 | Full Throttle | -0.59 |
| 142 | Limit | -0.06 | 175 | GhostRider | -0.59 |
| 143 | Mind Eraser | -0.06 | 176 | New Revolution | -0.61 |
| 144 | Mind Eraser | -0.06 | 177 | Rock 'n' Roller Coaster | -0.62 |
| 145 | MP-Xpress | -0.06 | 178 | Time Traveler | -0.63 |
| 146 | Raptor | -0.06 | 179 | Stealth | -0.65 |
| 147 | Riddler Revenge | -0.06 | 180 | Flight Deck | -0.65 |
| 148 | Flight of Fear | -0.06 | 181 | Storm Chaser | -0.66 |
| 149 | Batman: Arkham Asylum | -0.10 | 182 | Steel Eel | -0.67 |
| 150 | Flight Deck | -0.13 | 183 | Twister II | -0.67 |
| 151 | American Eagle | -0.15 | 184 | Saw - The Ride | -0.70 |
| 152 | Joker | -0.15 | 185 | Demon | -0.71 |
| 153 | Medusa Steel Coaster | -0.16 | 186 | Demon | -0.73 |
| 154 | Wodan Timbur Coaster | -0.16 | 187 | Gemini | -0.74 |
| 155 | Python in Bamboo Forest | -0.20 | 188 | Montezum | -0.75 |
| 156 | Shock Wave | -0.21 | 189 | Flight of Fear | -0.78 |
| 157 | Star Wars Hyperspace Mountain: Rebel Mission | -0.22 | 190 | Temple of the Night Hawk | -0.80 |
| 158 | New Texas Giant | -0.23 | 191 | Montana  Rusa | -0.83 |
| 159 | Journey to Atlantis | -0.30 | 192 | Colorado Adventure | -0.85 |
| 160 | Boss | -0.31 | 193 | Coaster Express | -0.86 |
| 161 | Mr. Freeze Reverse Blast | -0.34 | 194 | Big Thunder Mountain | -0.86 |
| 162 | Mr. Freeze Reverse Blast | -0.34 | 195 | Legend | -0.86 |
| 163 | Big Loop | -0.40 | 196 | Corkscrew | -0.89 |
| 164 | Snow Mountain Flying Dragon | -0.40 | 197 | Fluch von Novgorod | -0.91 |
| 165 | Expedition GeForce | -0.43 | 198 | Tonnerre de Zeus | -0.93 |
| 166 | Superman - Ultimate Flight | -0.43 | 199 | Invertigo | -0.93 |
| 167 | Nemisis Inferno | -0.43 | 200 | Steel Venom | -0.95 |
| 168 | RailBlazer | -0.50 | 201 | Corkscrew | -0.95 |
| 169 | Taron | -0.51 | 202 | Boomerang | -0.96 |
| 170 | Black Mamba | -0.51 | 203 | Boomerang | -0.96 |
| 171 | Boulder Dash | -0.52 | 204 | Boomerang | -0.96 |
| 172 | Poltergeist | -0.53 | 205 | Boomerang | -0.96 |
| 173 | Wicked Twister | -0.58 | 206 | Boomerang | -0.96 |

(Continued Table)

| Rank | Name | Score | Rank | Name | Score |
|---|---|---|---|---|---|
| 207 | Boomerang | -0.96 | 237 | Big Thunder Mountain Railroad | -1.43 |
| 208 | Boomerang | -0.96 | 238 | Sky Wheel | -1.45 |
| 209 | Boomerang | -0.96 | 239 | Mammut | -1.47 |
| 210 | Boomerang | -0.96 | 240 | Prowler | -1.48 |
| 211 | Boomerang Coast to Coaster | -0.96 | 241 | Mystic Timbers | -1.49 |
| 212 | Flashback | -0.96 | 242 | Wild One | -1.49 |
| 213 | Apocalypse | -0.97 | 243 | V2: Vertical Velocity | -1.53 |
| 214 | Balder | -0.98 | 244 | Jungle Trailblazer | -1.55 |
| 215 | Grizzly | -1.00 | 245 | Anaconda | -1.55 |
| 216 | Doble Loop | -1.01 | 246 | Alpina Blitz | -1.62 |
| 217 | Abismo | -1.05 | 247 | Bocaraca | -1.64 |
| 218 | Eurosat Can Can Coaster | -1.07 | 248 | Super Tornado | -1.64 |
| 219 | Timber Drop | -1.08 | 249 | Tornado | -1.64 |
| 220 | Green Lantern Coaster | -1.10 | 250 | Whirl Wind Looping Coaster | -1.64 |
| 221 | Adrenaline Peak | -1.11 | 251 | Terminator Salvation: The Coaster | -1.65 |
| 222 | Nemesis | -1.17 | 252 | Boardwalk Bullet | -1.66 |
| 223 | Timber Wolf | -1.19 | 253 | Sky Scream | -1.67 |
| 224 | Vertical Velocity | -1.19 | 254 | Furius Baco | -1.68 |
| 225 | Journey to Atlantis | -1.21 | 255 | Ranier Rush | -1.69 |
| 226 | Joker | -1.27 | 256 | Renegade | -1.71 |
| 227 | Comet | -1.27 | 257 | Montana Rusa | -1.71 |
| 228 | Star Mountain | -1.28 | 258 | Stampida | -1.74 |
| 229 | Atlantica SuperSplash | -1.28 | 259 | Giant Dipper | -1.75 |
| 230 | Xcelerator | -1.28 | 260 | Titan Cascabel | -1.77 |
| 231 | Crazy Bird | -1.29 | 261 | Timberhawk: Ride of Prey | -1.81 |
| 232 | Texas Tornado | -1.30 | 262 | Cedar Creek Mine Ride | -1.84 |
| 233 | Apocalypse the Ride | -1.35 | 263 | Viper | -1.85 |
| 234 | Racer | -1.39 | 264 | Mine Blower | -1.87 |
| 235 | Nessie Superrollercoaster | -1.39 | 265 | Montana Rusa | -1.88 |
| 236 | Katapul | -1.41 | 266 | Ravine Flyer II | -1.91 |

(Continued Table)

| Rank | Name | Score |
|---|---|---|
| 267 | El Toro | -1.91 |
| 268 | Half Pipe | -1.91 |
| 269 | Iron Dragon | -1.93 |
| 270 | Bandit | -1.93 |
| 271 | Spatiale Experience | -1.97 |
| 272 | Bat | -1.98 |
| 273 | Montezooma's Revenge | -2.02 |
| 274 | Cyclone | -2.03 |
| 275 | Raven | -2.03 |
| 276 | Oblivion | -2.04 |
| 277 | Phoenix | -2.07 |
| 278 | Ninja | -2.08 |
| 279 | Screamer | -2.08 |
| 280 | Coaster Thrill Ride | -2.12 |
| 281 | Wild Thing | -2.17 |
| 282 | SpeedSnake FREE | -2.21 |
| 283 | Whizzer | -2.27 |
| 284 | Giant Dipper | -2.30 |
| 285 | Piraten | -2.34 |
| 286 | Revenge of the Mummy the Ride | -2.34 |
| 287 | Kawazemi | -2.35 |
| 288 | Road Runner Express | -2.36 |
| 289 | Blue Streak | -2.49 |
| 290 | Pandemonium | -2.53 |
| 291 | Desert Race | -2.54 |
| 292 | Judge Roy Scream | -2.61 |
| 293 | Sidewinder | -2.68 |
| 294 | HeiBe Fahrt | -2.73 |
| 295 | Manta | -2.74 |
| 296 | Taunusblitz | -2.78 |
| 297 | Force One | -2.85 |
| 298 | Winjas | -2.95 |
| 299 | Pandemonium | -3.51 |
| 300 | Backlot Stunt Coaster | -3.59 |