KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING
*School of Data Science and Analytics*

# Roller Coasters - What is your thrill?

Vanessa Sunar – Graduating December 2022

Advisors: Professor Susan Mathews Hardy, Dr. Austin Brown, Dr. Gene Ray

## INTRODUCTION

Roller Coasters have become an entertainment today especially due to their evolution over the years. They are now bigger, faster and scarier than ever before. This adrenaline boosting activity has certainly led to theme parks becoming a big business today. However, like many businesses, themes parks were also hit hard by the Covid-19 pandemic. It had great impact on their revenue, which led to some parks even "considering selling off property to raise capital."

This study researches various roller coasters in the U.S. and the different variables that relate to them. Length of ride, initial drop, duration of ride, maximum height, top speed, and number of inversions were analyzed with logistic regression, multiple linear regression, loess regression and graphics. Graphical analyses included scatter plots, times series plots and violin plots. These relationships give insights to what the public wants and can help determine roller coaster success.

The dataset was obtained from https://rcdb.com/196.htm and limited to roller coasters in the United States. Additional observations, which consist of roller coasters from years 2018 -2022, were added to to include top 10 roller coasters and the ratings for each roller coaster in the compiled data set. The increased sample dataset consists of 188 observations.

**By identifying the relationships and variables influencing roller coaster popularity and demand, this study seeks to serve as a guide for helping theme parks understand what makes roller coasters attractive to visitors. In this way they can improve marketing and make informed choices for new roller coaster installations.**

## METHODS

- **Logistic Regression:** used to determine whether roller coaster type can be predicted from duration, length, drop, number of inversions, and top speed.
- **Stratified Loess Regression:** used to determine the relationship for height and top speed for wooden and steel roller coasters.
- **Four-variable Stratified Scatterplot by Color and Bubbles:** used to study the interrelationships between maximum height, top speed, age of the roller coaster, and whether the roller coaster is wood or steel.
- **Box plot:** used to investigate whether a roller coaster's top 10 status or wooden/steel category is related to its rating.
- **Multiple Linear Regression:** used to determine if the length, number of inversions, and top speed of a coaster predict the rating it receives.
- **Timeseries:** used to visually look at the change in top speed and maximum height of roller coasters in the U.S. over time.
- **Violin Plot:** used to investigate the relationship between top speed and design of the coaster.
- **Geospatial Map:** locates top 10 roller coasters across the U.S. and displays how many top 10 categories the roller coaster was awarded. Larger-size dots indicated the number of top 10 awards for the roller coaster. The color indicates whether the roller coaster is wood or steel.

## REFERENCES

https://www.coaster101.com/2016/02/25/coasters-101-roller-coasters-removed/

https://coasterbuzz.com/RollerCoasters/Top100

https://www.semrush.com/blog/roller-coaster-marketing/

https://www.attractionsmanagement.com/index.cfm?pagetype=features&codeID=29837

## LINKEDIN

## GITHUB CODE

**Table 1: Best Logistic Model**

$$y = .223 + .032\,Duration - .001\,Length + .059\,Drop + 1.441\,NumOfInversions - .116\,TopSpeed$$

**Concordance Index** = 0.949221

**Table 1: Best Logistic Model to Predict Wood/Steel**

| Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| **Parameter** | **Estimate** | **Standard Error** | **DF** | **t Value** | **Pr > \|t\|** |
| **Intercept** | 0.2230 | 2.0742 | 138 | 0.11 | 0.9145 |
| **Duration** | 0.03209 | 0.01064 | 138 | 3.02 | 0.0031 |
| **Length** | -0.00140 | 0.000395 | 138 | -3.56 | 0.0005 |
| **Drop** | 0.05893 | 0.01572 | 138 | 3.75 | 0.0003 |
| **Num_of_Inversions** | 1.4409 | 0.4372 | 138 | 3.30 | 0.0012 |
| **Top_Speed** | -0.1158 | 0.06222 | 138 | -1.86 | 0.0649 |

**Figure 2: ROC Curve**

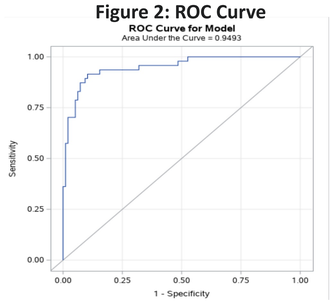



**Figure 3: Rating by Top 10 Status and Steel/Wood**

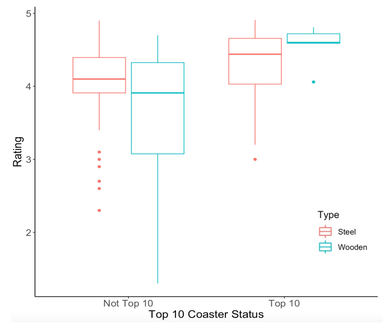



**Figure 4: Will you be traveling by a Top 10 roller coaster?**

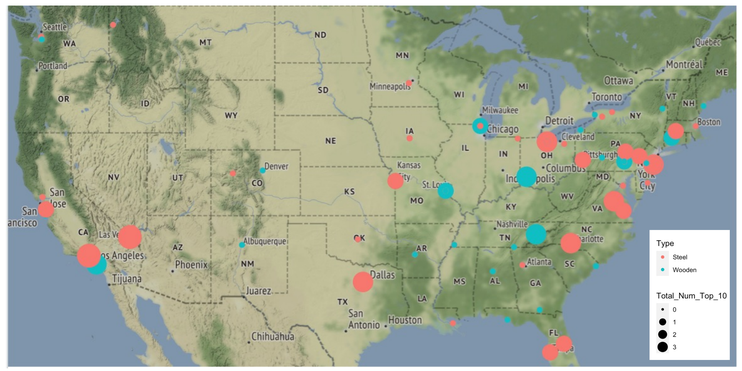



**Figure 5: Word Cloud of Recent Reviews**


**Figure 6: Scatterplot and Loess Curve of Height Predicting Top Speed for Wooden and Steel Roller Coasters**

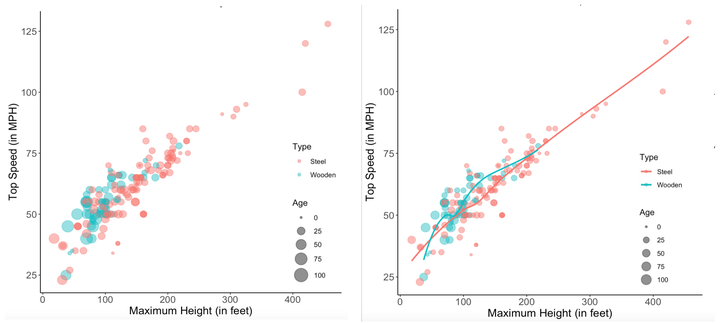



**Figure 7(a): Changes in Maximum Height Over Time**

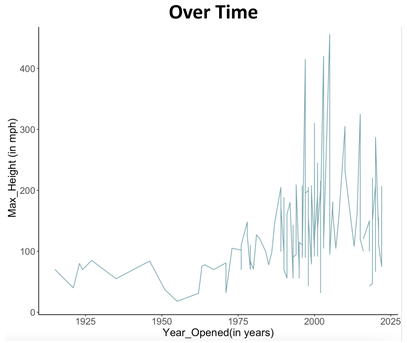



**Figure 8: How is speed distributed across different design categories?**

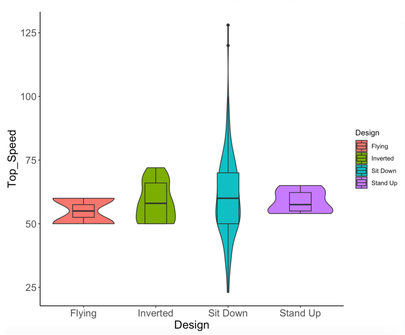



**Table 2: Regression Model to Predict Rating**

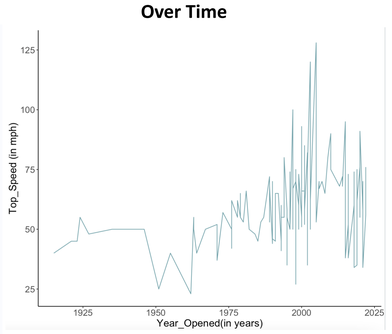
$$Rating = 2.991 + .0001\,Length + .058\,LNumOfInversions + .009\,TopSpeed$$

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 3 | 8.75702 | 2.91901 | 5.86 | 0.0010 |
| **Error** | 100 | 49.82418 | 0.49824 | | |
| **Corrected Total** | 103 | 58.58120 | | | |

| | | | |
|---|---|---|---|
| **Root MSE** | 0.70586 | **R-Square** | 0.1495 |
| **Dependent Mean** | 3.98740 | **Adj R-Sq** | 0.1240 |
| **Coeff Var** | 17.70231 | | |

| Parameter Estimates | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > \|t\|** | **Variance Inflation** |
| **Intercept** | Intercept | 1 | 2.99183 | 0.27417 | 10.91 | <.0001 | 0 |
| **Length** | Length | 1 | 0.00010012 | 0.00004865 | 2.06 | 0.0422 | 1.22305 |
| **Num_of_Inversions** | Num_of_Inversions | 1 | 0.05791 | 0.03135 | 1.85 | 0.0677 | 1.00068 |
| **Top_Speed** | Top_Speed | 1 | 0.00913 | 0.00454 | 2.01 | 0.0468 | 1.22232 |

**Figure 7(b): Changes in Top-Speed Over Time**




## RESULTS

**LOGISTIC REGRESSION:** Table 1 gives the model, Concordance Index and p-values for the coefficients. From the Concordance Index, we see that the model is able to distinguish between steel and wood 94.9% of the time.

**LOGISTIC BETA COEFFICIENTS** Table 1 displays the best predictors of whether the model is steel or wood. By exponentiating the coefficients, the relationships of the predictors with the outcome of steel/wood are discovered:

- **Duration:** The odds of a roller coaster being steel are 1.03 times the same odds of a roller that is one second shorter. **Longer duration rides are more likely to be steel.**
- **Length:** The odds of a roller coaster being steel are .93 times the same odds of another roller coaster being steel that is 50 feet shorter. **Longer length rides are less likely to be steel**.
- **Drop:** The odds of a roller coaster being steel are 1.06 times the same odds of a roller that has one foot less of a drop. **High drops are more likely to be steel.**
- **No of Inversions:** The odds of a roller coaster being steel are 4.2244 times the same odds of a roller coaster that has one less inversion. **A roller coaster with more inversions is more likely to be steel.**
- **Top_Speed:** The odds of a roller coaster being steel are .89 times the same odds of another roller coaster being steel that is 1 mile per hour slower. **Faster speed rides are less likely to be steel.**

**ROC CURVE:** Figure 2 displays the ROC curve. The Area Under the Curve (AUC) is the Concordance Index, 0.9493. A value close to one indicates high accuracy.

**RATING BY TOP 10 COASTER AND TYPE**: As displayed in Figure 3, the ratings of almost all wooden roller coasters in the top 10 exceed about 75% of the ratings of the top 10 steel roller coasters. Steel coaster ratings are less variable.

**GEOSPATIAL MAP:** Figure 4 displays top 10 roller coasters in the U.S. Bubble size indicates the number of top 10 lists to which the coaster belongs.

**WORD CLOUD:** Figure 5 displays words from ratings on the Captain Coaster website.

**MAXIMUM HEIGHT AND TOP SPEED BY DESIGN:** Figure 6 has two graphs, one with the loess curve and one without. The loess curve in the second plot shows that wooden roller coaster speeds cap out about 75 mph. For steel, however, there seems to be no limit to the top speed. Smaller dots for steel in the upper-right, indicate that newer steel roller coasters are higher and faster. In the scatterplot without the loess curve, it is easier to see that wooden roller coasters are in general older, shorter and slower.

**MAXIMUM HEIGHT OVER TIME:** Figure 7(a) shows the maximum height becomes extremely variable from about 1995 to 2005. After that time period, roller coasters seem to settle between the maximum height of 80 to 325 feet. In general, the maximum height of roller coasters has increased over time until recent years.. Starting in 2014, helix and other features seem to be dominating the roller coaster world more than height.

**TOP SPEED OVER TIME:** As shown in the Figure 7(b), the top speed follows the same pattern as maximum height. It becomes extremely variable from year 1995 to 2005. Again, starting in 2014, other features seem to be dominating the roller coaster world more than speed.

**DESIGN AND TOP SPEED**: Figure 8 shows that roller coasters with a Flying element have top speeds that are frequently 50 mph and 60 mph. Inverted designs are more likely to have top speeds in the lower 60th percentile of the group. Sit Down designs have the most variable speeds.

**PREDICTORS OF RATING:** As displayed in Table 2, the best model to predict rating of a roller coaster consists of the three variables - length, number of inversions and top speed.

- Length: As the length goes up 100 feet the rating increases by .01 on a five-point scale.
- No of Inversions: As the number of inversions increases by 1, the rating increases .06 on a five-point scale.
- Top Speed: As the top speed increases 10 mph, the rating increases by .1 on a five-point scale.

## CONCLUSION

- Longer duration, higher drops and more inversions are associated with steel coasters. Hence, it might be more profitable to have more steel coasters.
- Longer length and faster speeds are associated with wood coasters.
- Steel coasters are more likely to have higher ratings compared to wood.
- From research, wooden coasters are much more expensive to maintain.
- To maximize profit, steel coasters are a better investment.