Predicting Wait Times:

Using Disney World Wait Time Data to Make Predictions

Jonathon Scroggins

DSC 630

Bellevue University

Professor Alsaleem

August 13, 2021

**Executive Summary**

This project will explore and use predictive analysis on the attraction wait times at The Walt Disney World Resort. The analysis will use data provided by touringplans.com on the wait times of various rides at The Walt Disney World Resort. The attraction wait times can help to determine how one plans their trip to The Walt Disney World Resort as wait times can help with the perception of how “busy” the parks are. These perceptions of the “busyness” can be assumed based on the assumption that the higher the average wait time the more people there would be at the parks making the lines at the attractions longer and therefore the average wait time would also increase. These predictions can then be used by travel planning companies to make suggestions on travel plans for tourists or the tourists themselves to make informed decisions on what they feel is the “best time” to go to the parks. The data used for this analysis is comprised of several attraction wait times and the calendar information for the parks, like ticket seasons, days of the weeks, and the datetime of the wait times. The data was imported into Python and cleaned. Each attraction was imported as a separate dataframe, cleaned and then all the wait times were merged into a single dataframe. The metadata file was also imported, cleaned, and merged with the wait time dataframe. After cleaning a preliminary analysis was performed on this data. This analysis looked at the averages of the data and used graphical analysis to look at things like the trend and distribution of the data. After cleaning and preliminary analysis, models were built around the data to predict the wait times. Decision Tree Regression models and Linear Regression models were chosen for the predictive analysis of the wait times. The models were then used to make predictions of what the wait times would be for the month of September 2021. These predictions showed the capability of the models to predict the wait times for future dates at the parks and allow one to have a more informed plan for a trip to the parks.

**Background**

After being on lockdown for so long Americans are ready to travel. According to a report by the U.S. Travel Association in May 2021, nearly ninety percent of Americans have travel plans within the next six months. COVID restrictions are being lifted and people are ready to get back to what they consider normal. One of the most popular destinations when people travel is still The Walt Disney World Resort. Since so many people are ready to get back to The Walt Disney World Resort and restrictions are being lifted, the crowds are beginning to come back, and people are starting to plan when they want to take their trips. Most people who visit the parks prefer to not have to wait for a ride, they want to get on the ride as fast as possible. To go along with that, most travelers prefer the crowds at the parks to be the lowest possible. That is where sites like touringplans.com or other travel agencies try to give their insight into planning trips to The Walt Disney World Resort with things like travel planning guides and crowd calendars. Being able to provide good planning tools and insights to the parks are what these places are selling. With that in mind, being able to build a good model to make predictions on when the ride wait times would be lower, provide more accurate insight to when to visit the parks to avoid crowds would make the travel guides more enticing. This project will look to investigate the wait times of certain popular rides at The Walt Disney World Resort to look for trends during the calendar year. The project will do this with the assumption that the longer the wait times for a ride the more people there are in the parks meaning that the crowd levels are higher during that time. This is also under the assumption that the “best time” to go to the parks is when the wait times are lower and therefore the crowd levels are lower as well. The project will investigate what factors would influence the wait times of the rides. The analysis will look at factors like the time of year, day of the week, or events going on at or around the parks that may influence the wait times and would therefore influence the crowd levels of the parks.

**Problem Statement**

When is the “best time” to visit The Walt Disney World Resort? Is there a way to predict the wait times of certain rides at The Walt Disney World Resort?

**Methods**

The wait time data set for ten rides were imported into Python and R. The rides imported were The Seven Dwarfs Mine Train (Magic Kingdom), Dinosaur (Animal Kingdom), Expedition Everest (Animal Kingdom), Avatar Flight of Passage (Animal Kingdom), The Na’vi River Journey (Animal Kingdom), The Pirates of the Caribbean (Magic Kingdom), Rock ‘n’ Rollercoaster (Hollywood Studios), Slinky Dog Coaster (Hollywood Studios), Spaceship Earth (Epcot), Toy Story Midway Mania (Hollywood Studios). Each of the rides was imported as a separate data set with the variables for date, datetime, the actual wait time and the posted wait time. Due to there not being a compatible number of datapoints for actual wait times across all rides, the decision was made to drop that variable in all the data sets. The data sets were then merged into one data set along with a meta data set that included all the calendar information for The Walt Disney World Resort. If the ride was open at the time, the data sets started on January 1, 2015. However, there were some of the rides that were not open until the year 2018. Therefore, the data was subset to only include data points from 2018 – 2019. The data also had notes in to signify when a ride was down or not open by placing the value -999 in the wait time slot. This was affecting the wait time average and therefore was changed to 0 for calculating purposes. Also, there are Null values in the data due to the nature of the time series data, but those were left in as they were not affecting the analysis at this time. After the data was cleaned and merged initial analysis was done on the data including but not limited to a graphical analysis to look at the distribution and averages of the attraction wait times. Initial graphical analysis was used to get an idea of how the data looked and be able to predict what models might be the best to build for these data sets. Figure 1 shows the average wait time for the rides for 2018, 2019 and the total average for both years.

Table

Description automatically generated

This chart shows that there is not a significant difference between the two years with the largest difference being for the Na’vi River Journey. This shows that there is similar behavior in the wait times for both years and for the overall average of the attraction. Next, Figures 2 and 3 show the distribution of the attraction wait times.

Timeline

Description automatically generated

Figure 2

Graphical user interface, timeline

Description automatically generated

Figure 3

Though most of them look skewed to the left, they all have a good distribution. Part of the reason for the skewness to the left could be from changing the -999 to 0. If a ride was not open for a good amount of time, there would be more 0’s in the dataframe. Over a 2-year period with refurbishments, weather and other factors that cause downtime the number of 0’s can be expected and would affect the wait times accordingly, that is why they should not be excluded from the predictive analysis.

Now we look at the trends in the data. First in figure 4 we see the trends for the average wait time per hour and the daily average wait time per hour is shown in Figure 5.

Chart, line chart

Description automatically generated

Chart

Description automatically generatedBoth figures show a significant trend in the data. There is a gradual increase in the early hours of the day for most attractions, that then decreases gradually through out the afternoon hours into evening as shown in Figure 4. There also is a trend for Friday – Monday to have higher average wait times and Tuesday – Thursday to have lower average wait times. The last feature to explore was the WDW ticket season. This is a feature assigned by the company to base the price of a ticket. If the demand for the day was deemed high then the ticket price would be considered peak, medium demand made the ticket regular and lower demand made the ticket value priced. This was a perceived value of the day based on how busy the day was though to be. As you can see in Figure 6, the average wait time also mirrored this analysis. The average wait times for a peak day, when the demand for the day was higher, were also higher.

Chart, bar chart

Description automatically generated

After doing the graphical analysis on the data, more analysis and models were built. The first analysis that was attempted was to do a time series analysis in R. With the trends in the data, and the fact that it was related to time, a time series was though to be a good model for the data. However, due to the nature of the data and the amount of the data, a time series analysis was proving to be more challenging than thought and a new method was sought. It was at this point that research was done on wait time analysis from other fields as well. A study was found titled *Predicting wait times in Pediatric OPHTHAMOLOGY outpatient Clinic using machine learning,* that showed how they used Random Forrest and Decision Trees to predict wait times. It was at this point that Decision Tree models, Linear Regression models and Logistic Regression models would be built.

**Results**

To build the models, the first step was to split the data into a training set and a testing set. The training set would be used to build and train the model, while the testing model would be used to see how well the model performs. The first Target Variable chosen was the WDW Ticket season. With this variable being a categorical variable, a Decision Tree Classifier model and a Logistic Regression model were trained with the data. The Decision Tree Classification model showed 99% accuracy, while the Logistic Regression model performed at 57% accuracy. However, at this point, it was decided to move on to the models for the attraction wait times. Due to the nature of the WDW Ticket Season variable being dependent on perceived crowd size and other unknown factors that are not in these datasets, it was determined that they would not be as beneficial of a model target.

The second set of models focused on the attraction wait times. An individual model was built for each of the ten attractions. In fact, three models were built for each individual attraction. The first two models used the same features: Day of the week, Day of the Year, Day of the month, Month of the year and Hour of the Day. With these features, a Decision Tree Regression model and a Linear Regression model was trained on the training set data. For the third model, a Decision Tree Regression was run and the WDW Ticket season variable, as dummy variables, was trained on the data. To assess these models, the Root Mean Squared Error (RMSE) was chosen. The RMSE is the distance the predicted value is from the actual value. In the case of this analysis, it would be the +/- the predicted value would be in minutes. So, if the predicted value is 15 minutes and the RMSE is 5 then the actual value would be assumed to be between 10 and 20. Figure 7 shows the results from the three models that were trained on the data.

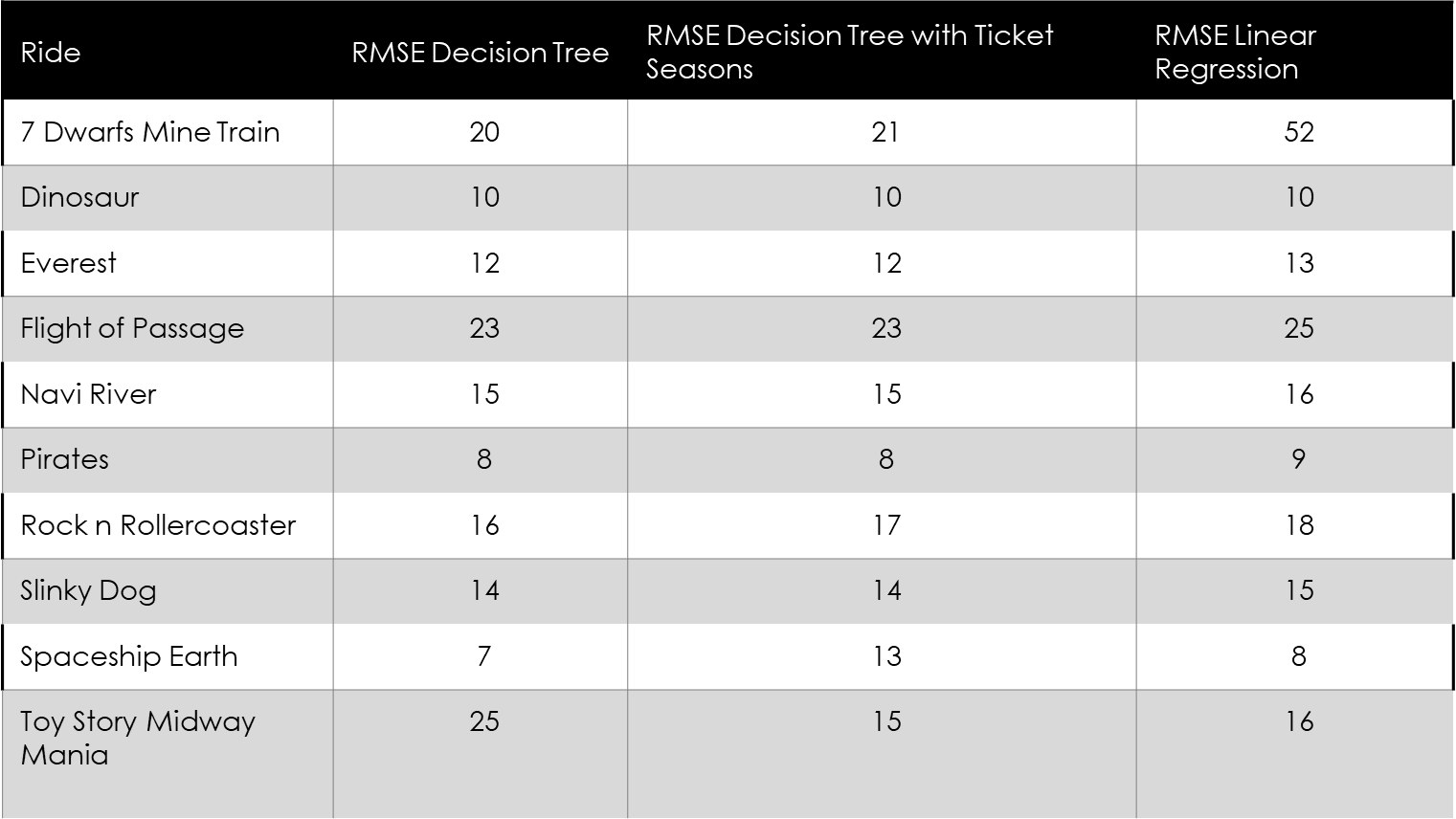


Figure 7

Overall, the models performed well, the one major exception would be the Linear Regression model for the Seven Dwarfs Mine Train. That model had the highest RMSE of 52, which when compared to all the other RMSE is much higher than desired and therefore it was decided to not use that model. When comparing the RMSEs from the Decision Tree Regression models, both are very similar without any significantly different values except for maybe Toy Story Midway Mania. However, due to the variable and unknown nature of the WDW Ticket Season, it was decided to go with the Decision Tree Regression model without the WDW Ticket Season feature.

Using this model, we can now predict the wait times for future dates at The Walt Disney World Resort. Therefore, a dataframe was created with the values for September 2021. That dataframe was then run through the model to get predictions for all the attraction’s wait times from the hours of 8am to 8pm since most parks are open during those hours. Figures 8 and 9 show the average wait time per day of the week for September 2021.

Table

Description automatically generated

A screenshot of a computer

Description automatically generated with low confidence

These averages and trends look like the ones from 2018 and 2019 with Friday-Monday being higher overall than Tuesday-Thursday. We can also look closer at a specific day as you can see in Figure 10 which shows the predicted wait times by hour of each attraction for September 1, 2021. Figure 11 shows the trend for the September 1, 2021, as well.

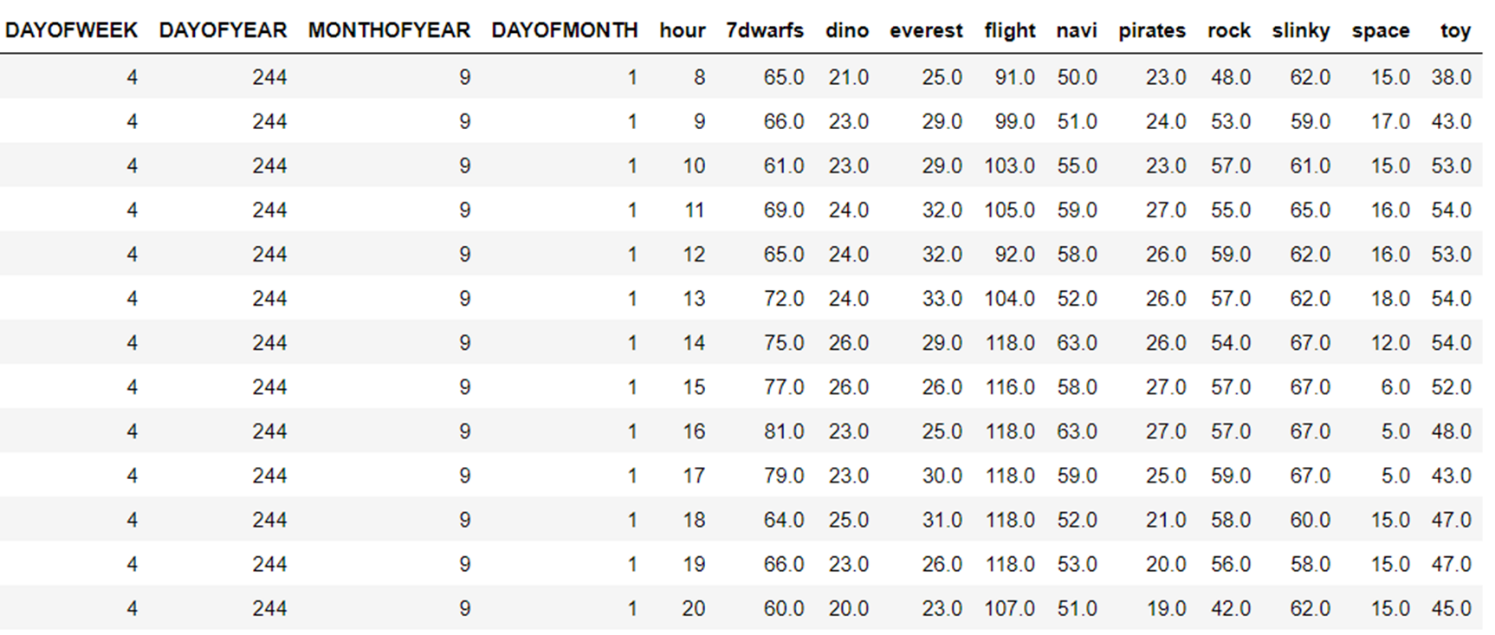


Figure 10

Chart, line chart

Description automatically generated

**Discussion/conclusion**

Based on the results of the models and the predictions for September 2021, I would say that it is possible to predict the wait times for attractions at The Walt Disney World Resort. Even though some may say the RMSEs of the models are a little high, I would argue that the margin that those RMSEs give is a good indicator of the time that can be expected to spend in the queue of a particular attraction. That margin can account for any unforeseen elements like weather and down times of an attraction. To improve upon these models and this analysis, one could gather information for more or all the Walt Disney World attractions. This would allow one to fully plan a trip at the parks. One could even build park specific models that could predict and help plan the day around the park. For example, there are a few of the rides from Animal Kingdom in this analysis. One could take those and use them to plan an Animal Kingdom specific day, and if other attractions were gathered as well, they can be used also. As mentioned before, there are other factors like weather, age of the ride, or location of the park that possibly affect the wait time and could be explored in future models.

**Acknowledgements**

I want to acknowledge and say thank you to the website touringplans.com for providing the datasets for the wait times of the attractions. I also want to say thank you to my fellow classmates that have given advice and peer reviewed all the milestones leading up to this final paper.

**Resources**

*"7\_dwarfs\_train.csv",* Disney World Ride Wait Time Datasets, TouringPlans.com, June 2018, https://www.touringplans.com/walt-disney-world/crowd-calendar/#DataSets, Accessed 9 June 2021

*"dinosaur.csv",* Disney World Ride Wait Time Datasets, TouringPlans.com, June 2018, https://www.touringplans.com/walt-disney-world/crowd-calendar/#DataSets, Accessed 9 June 2021

*"expedition\_everest.csv",* Disney World Ride Wait Time Datasets, TouringPlans.com, June 2018, https://www.touringplans.com/walt-disney-world/crowd-calendar/#DataSets, Accessed 9 June 2021

*"flight\_of\_passage.csv",* Disney World Ride Wait Time Datasets, TouringPlans.com, June 2018, https://www.touringplans.com/walt-disney-world/crowd-calendar/#DataSets, Accessed 9 June 2021

K, G. M. (2020, July 18). *Machine learning basics: Decision tree regression*. Medium. <https://towardsdatascience.com/machine-learning-basics-decision-tree-regression-1d73ea003fda>.

Lin, W.-C., Goldstein, I. H., Hribar, M. R., Sanders, D. S., & Chiang, M. F. (2020, March 4). *Predicting wait times in Pediatric OPHTHALMOLOGY outpatient Clinic using machine learning*. AMIA ... Annual Symposium proceedings. AMIA Symposium. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7153152/.

*"metadata.csv",* Disney World Ride Wait Time Datasets, TouringPlans.com, June 2018, https://www.touringplans.com/walt-disney-world/crowd-calendar/#DataSets, Accessed 9 June 2021

*Monthly Travel Recovery Data Report*. U.S. Travel Association. (2021, May 27). https://www.ustravel.org/research/monthly-travel-recovery-data-report.*"navi\_river.csv",* Disney World Ride Wait Time Datasets, TouringPlans.com, June 2018, https://www.touringplans.com/walt-disney-world/crowd-calendar/#DataSets, Accessed 9 June 2021

*"pirates\_of\_caribbean.csv",* Disney World Ride Wait Time Datasets, TouringPlans.com, June 2018, https://www.touringplans.com/walt-disney-world/crowd-calendar/#DataSets, Accessed 9 June 2021

*"rock\_n\_rollercoaster.csv",* Disney World Ride Wait Time Datasets, TouringPlans.com, June 2018, https://www.touringplans.com/walt-disney-world/crowd-calendar/#DataSets, Accessed 9 June 2021

*"slinky\_dog.csv",* Disney World Ride Wait Time Datasets, TouringPlans.com, June 2018, https://www.touringplans.com/walt-disney-world/crowd-calendar/#DataSets, Accessed 9 June 2021

*"spaceship\_earth.csv",* Disney World Ride Wait Time Datasets, TouringPlans.com, June 2018, https://www.touringplans.com/walt-disney-world/crowd-calendar/#DataSets, Accessed 9 June 2021

*"toy\_story\_mania.csv",* Disney World Ride Wait Time Datasets, TouringPlans.com, June 2018, https://www.touringplans.com/walt-disney-world/crowd-calendar/#DataSets, Accessed 9 June 2021