Disney World Ride Wait Times

Gus Lipkin ~ glipkin6737@floridapoly.edu

April 18, 2022

Table of Contents

Introduction

Most people like going to Disney World for one reason or another. It could be the atmosphere, character meet and greets, theming, fireworks, exclusive dining options, rides, or any number of other reasons. Of the people that don't like going to Disney World, I'd hazard a guess that the large crowds and long ride wait times are at the top of most people's lists. While Disney does offer a way to cut long lines through its Lightning Lane program, it is an additional cost, limited use item, and is not available for every attraction (Disney 2022). There are some techniques to avoid the lines such as arriving to the park early, getting in line during a parade or fireworks, or visiting at different times of day or year. These techniques can show some success, but once too many people catch on, then the technique becomes useless. This is similar to how on a three lane highway, one lane may be moving faster until everyone wants to get into the "fast" lane, and then that lane becomes the slow lane. There are many websites that provide some level of insight into current and historical ride wait times such as touringplans.com, wdwpassport.com, and thrill-data.com. However, none attempt to forecast ride wait times to give their users a competitive advantage in choosing their next ride so as to minimize ride wait times. Although my method has not yet been tested as of writing, I will analyze historical data and publish decision trees that will empower Disney goers to choose their next ride based on current park data that is immediately available to them.

Literature Review

In Spring 2020, Lipkin et al Lipkin, Skoglund, and Pierstorff (2020) worked to predict the best time of year and day to visit Disney World based on historical seasonal trends. Based on an extensive exploration of the data and data visualization techniques, they recommended that September was the best time of year to visit, particularly a Wednesday afternoon. They suggested that this was mostly due to the fact that nearly every school in the country is in session during this time. They also considered that it may also have to do with the weather still being Florida's hot and rainy season, which is less than ideal for outdoor theme parks. In their next steps, they suggested using supervised machine learning to predict actual ride wait times based on posted ride wait times.

Nearly a year earlier, Mendoza, Wu, and Leung (2019) also studied data from Touring Plans (2021), and went so far as to develop a multiple regression model that explained approximately 80% of daily wait time variation with 15 minute average error. They took particular note of the seasonal lows in September and highs in December, but did not go so far as to make suggestions on when to visit Disney World based on their analysis.

Data

The bulk of the data used in this analysis was provided by Touring Plans (2021). It consists of fourteen main datasets, one each for the fourteen different attractions with data recorded. Each of these consists of a park date variable based on the day which the park opened, especially important for days where parks are open past midnight, a datetime variable that says when the data was recorded, and an actual and predicted wait time variable. Due to limitations in how Touring Plans recorded the data, there is either only an actual wait time or posted wait time, but not both for any given date time.

Touring Plans also provided an expansive metadata dataset with information from each day. While I won't discuss every variable here, I will highlight the some variables of note. Included are the ticket season, proximity to a holiday, historical maximum and minimum temperature as well as precipitation, opening and closing times, time of any fireworks, parades, or other shows and proportion of schools in session in different regions across the United States.

To supplement the weather data provided in the Touring Plans metadata dataset, I downloaded hourly weather data for 28.3772° N, 81.5707° W, the coordinates for Disney World from copernicus.eu. Using the Touring Plans data as a reference, I downloaded precipitation and temperature data for every hour that the parks are open.

All Touring Plans data was in `csv` files. However, after downloading, I noticed that the data did not go as far back as the data had for Lipkin, Skoglund, and Pierstorff (2020) and found that data. I combined the old data and the new data and dropped duplicate rows. The new metadata file also included twelve new columns. The volume of data needed from Copernicus exceeded their single download limit so I split the download request roughly in half by year. Each download was a `netcdf` file. In all cases, the data was combined in R.

Research Design

Aggregating the Data

There are two main issues with the wait time data from Touring Plans. The first is that data is not collected at consistent intervals and the second is that if the actual wait time is recorded, the posted wait time is not also recorded at the same time. These two issues then compound into smaller issues. The first of which is that traditional time series techniques become much more difficult to apply if data is not recorded at consistent intervals. The second is that wait times cannot be well compared across rides because there may not be a recording at the same time. While there is an indicator that a ride may be closed, there are also sometimes large gaps between recordings that are not well explained. To combat this, I took the mean recording every five minutes. I chose five minute intervals for two reasons. The first is that while line length can change significantly in five minutes, it is not likely, and the second is that five minutes is just enough time to really hoof it across any one of the four parks.

Choosing a Model

In all models, I would want the actual wait time to be lagged so that I am predicting what the ride wait time will be in five minutes. This way, model users will be able to get to the line before other people that see the line is getting shorter.

In *Final Project for Intro to Data Science*, I suggested that linear regression would be an appropriate next step. A linear regression would be more flexible with the sparse data. However, variable selection would be difficult such a large number of independent variables. In addition, linear regressions can be difficult to understand and implement while on the go at a theme park. Of course, nowadays everyone does have a calculator in their pocket, but I digress.

Logistic regression would follow the same pitfalls as linear regression, but would give the user even less information. Rather than predicting an actual wait time, I would only be able to predict if the next wait time would be less than the current wait time or not.

Analysis

Next Steps

Conclusion

References

2022. *Copernicus Climate Data Store*. Copernicus. https://doi.org/10.24381/cds.adbb2d47.

Disney. 2022. "Disney Genie Lightning Lane - Save Time Waiting in Line ..." 2022. https://disneyworld.disney.go.com/genie/lightning-lane/.

Lipkin, Gus, Hailey Skoglund, and Alex Pierstorff. 2020. "Final Project for Intro to Data Science." *Disney_ds*, April. https://github.com/guslipkin/disney_ds.

Mendoza, Dayanira, Wenbo Wu, and Mark T Leung. 2019. "Predicting the Expected Waiting Time of Popular Attractions in Walt Disney World." *Journal of Undergraduate Research & Scholarly Work*. https://rrpress.utsa.edu/bitstream/handle/20.500.12588/99/JURSW_Vol_6.Mendoza.pdf?sequence=1.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Touring Plans. 2021. *Disney World Ride Wait Time Datasets*. *Touringplans.com*. https://www.touringplans.com/walt-disney-world/crowd-calendar/#DataSets.

Appendix A (Code and Output)