Smart(er) Routing at Theme Parks

Oscar Li & Jerry Pan & Lei Qian

Duke University

Introduction

Disney World is a popular resort with 4 different theme parks: Magic Kingdom, Animal Kingdom, Epcot, and Hollywood Studios. On a busy day at a park in Disney World, getting on a ride can take more than an hour. At the end of the line for each ride there is an estimated wait time posted by Disney. We aim to predict the posted wait time for rides in Disney on a future day using historical posted wait time data. This could help park goers make better plans in going to Disney World.

Data and Processing

Our data set for wait times was provided by TouringPlans. The raw data consists of records of posted wait times scraped from Disney and recorded by TouringPlans app users. Each record associates a ride and a moment in time with the corresponding posted wait time. Directly comparing with wait times in the past is difficult in the raw data set as the wait times were not scraped at consistent intervals or at the same time each day. We processed the data such that each day is divided into time windows of a fixed duration and observations from the raw data that lay in the same time window were then averaged.

SDAY	SMONTH	SYEAR	SHOUR	SMIN	SPOSTMIN	SSINCEOPEN	userName
10	6	2014	10	1	20	61	marie23
10	6	2014	10	14	30	74	marie23
10	6	2014	10	22	30	82	marie23
10	6	2014	10	28	30	88	marie23
10	6	2014	10	36	30	96	marie23
10	6	2014	10	49	60	109	marie23
10	6	2014	10	57	60	117	marie23
10	6	2014	11	2	60	122	marie23
10	6	2014	11	8	60	128	marie23
10	6	2014	11	15	60	135	marie23
10	6	2014	11	21	50	141	marie23
10	6	2014	11	27	50	147	marie23
10	6	2014	11	36	40	156	marie23
10	6	2014	11	42	40	162	marie23
10	6	2014	11	51	40	171	marie23
10	6	2014	11	57	30	177	marie23

Table 1: Section of the Raw Data of Big Thunder Railroad

↓

date	window	AvgPost	NumPost
2014-06-10	10:00-10:10	20.00	1.00
2014-06-10	10:10-10:20	30.00	1.00
2014-06-10	10:20-10:30	30.00	2.00
2014-06-10	10:30-10:40	30.00	1.00
2014-06-10	10:40-10:50	60.00	1.00
2014-06-10	10:50-11:00	60.00	1.00
2014-06-10	11:00-11:10	60.00	2.00
2014-06-10	11:10-11:20	60.00	1.00
2014-06-10	11:20-11:30	50.00	2.00
2014-06-10	11:30-11:40	40.00	1.00
2014-06-10	11:40-11:50	40.00	1.00
2014-06-10	11:50-12:00	35.00	2.00

Table 2: Processed Data Using the Raw Data from Table 1

Data Visualization

After processing the raw data into a windowed format, we visualized the trend of how wait times of rides change throughout the day. We computed the rides' average posted wait times of every 20-minute window over the year 2014. Then, we plotted the average wait time graphs for 5 rides in Magic Kingdom and 5 rides in Animal Kingdom. We observe that the wait time curves for rides in the same park tend to follow a characteristic "shape" over the course of a day.

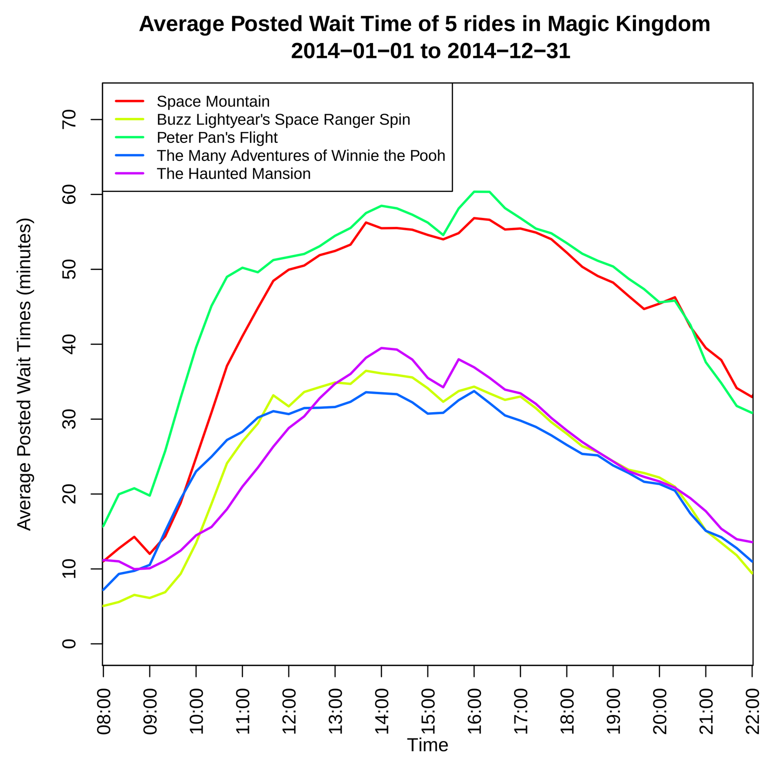

Figure 1: Magic Kingdom Park

In Figure 1 we can see that for Magic Kingdom, there is an increasing trend from opening time to around 2:00 p.m and a decreasing trend from 4:00 p.m. to closing. Between 2:00 and 4:00 p.m. there is a 'dip' where wait times tend to fall and then increase. This pattern is exclusive to Magic Kingdom whereas Animal Kingdom (Figure 2) can be more easily fitted with a downward parabola. The 'dip' in Magic Kingdom rides' graphs could be a result of the parade that occurs midday in Magic Kingdom. Since it is a major attraction, many people wait along the streets instead of queuing for rides during the parade.

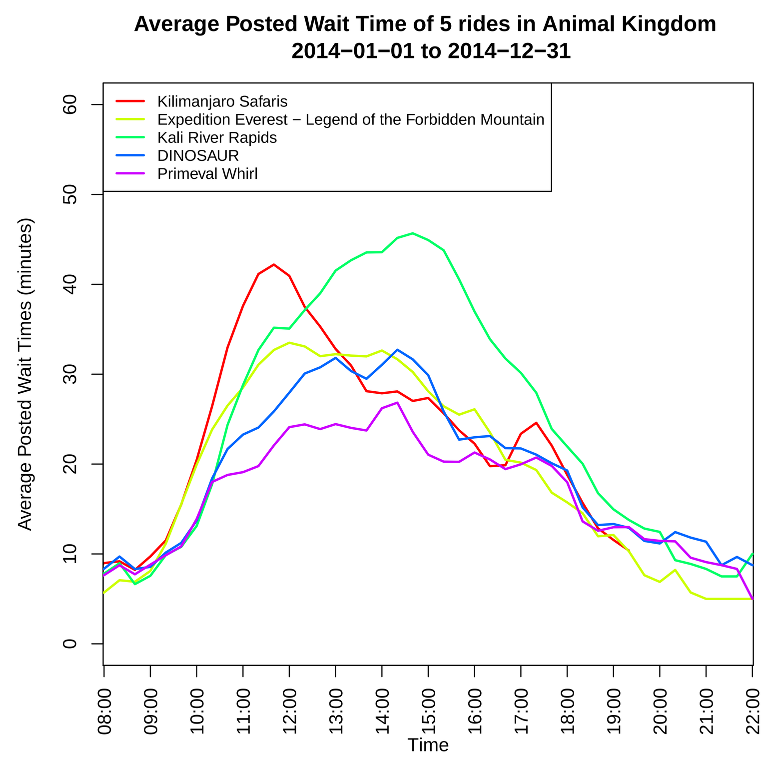

Figure 2: Animal Kingdom Park

The Many Adventures of Winnie the Pooh Ride Regression Model

$$\begin{aligned}\text{WinnieThePoohPostedWaitTime} = {} & \beta_0 + \beta_1 \cdot \text{WinnieThePoohHistWaitTime} + \beta_2 \cdot \text{WinnieThePoohLastWeekWaitTime} + \beta_3 \cdot \text{PeterPanHistWaitTime} + \beta_4 \cdot \text{BuzzLightyearHistWaitTime} + \beta_5 \cdot \text{SpaceMountainHistWaitTime} + \\ & \beta_6 \cdot \text{HauntedMansionHistWaitTime} + \beta_7 \cdot \text{LittleMermaidHistWaitTime} + \beta_8 \cdot \text{MonsterIncHistWaitTime} + \beta_9 \cdot \text{MadTeaPartyHistWaitTime} + \beta_{10} \cdot \text{EnchantedTalesW/BelleHistWaitTime} + \\ & \beta_{11} \cdot \text{Stitch'sGreatEscapeHistWaitTime} + \beta_{12} \cdot \text{DumboHistWaitTime} + \beta_{13} \cdot \text{JungleCruiseHistWaitTime} + \beta_{14} \cdot \mathbf{1}_{\text{weekend}} + \beta_{15} \cdot \mathbf{1}_{\text{spring}} + \beta_{16} \cdot \mathbf{1}_{\text{summer}} + \beta_{17} \cdot \mathbf{1}_{\text{winter}} + \beta_{18} \cdot \mathbf{1}_{\text{christmas}} + \\ & \beta_{19} \cdot \mathbf{1}_{\text{easter}} + \beta_{20} \cdot \mathbf{1}_{\text{halloween}} + \beta_{21} \cdot \mathbf{1}_{\text{july4}} + \beta_{22} \cdot \mathbf{1}_{\text{laborDay}} + \beta_{23} \cdot \mathbf{1}_{\text{memorial}} + \beta_{24} \cdot \mathbf{1}_{\text{mlk}} + \beta_{25} \cdot \mathbf{1}_{\text{newYears}} + \beta_{26} \cdot \mathbf{1}_{\text{thanksgiving}} + \beta_{27} \cdot \text{TimeSinceMidnight} + \beta_{28} \cdot \text{TimeSinceMidnight}^2 + \\ & \beta_{29} \cdot \text{TimeSinceMidnight} \cdot \mathbf{1}_{\text{afterDip}} + \beta_{30} \cdot \text{TimeSinceMidnight}^2 \cdot \mathbf{1}_{\text{afterDip}} + \beta_{31} \cdot \mathbf{1}_{\text{afterDip}}\end{aligned}$$

Predictive Regression Model

We built a predictive linear regression model with the focus on rides in Magic Kingdom. The model is built to predict the wait times a week in advance, and for every time window on that future date our model provides an estimate of the posted wait time. We manipulated the time variable to better fit the posted wait times, added historical information on the wait times of the ride we are predicting and other similar rides, and added variables depicting the type of day.

Transformation of Time

To predict the wait time of a future window, we want to incorporate the position of the window in the day as a predictor. As there is not a consistent increasing or decreasing trend between the wait time and time of day, we must transform time of day to better fit the model.

We first introduced the notion of a "chunk" into our model. The opening hours of a ride are split into "chunks" of time, where each chunk is the maximal period of time during which the wait time graph has a consistent increasing/decreasing trend. For each individual chunk, we fit the wait times in that chunk linearly.

Magic Kingdom has a "double" parabola shape, so we also used another model that splits the graph at the dip time and fitted separate quadratic curves for each time section. However, not every park has a definite double parabola shape, e.g. Animal Kingdom. Thus, we modified the double quadratic model into a single quadratic model, but with an optional second parabola defined only on the second chunk.

Historical and Categorical Data Processing

To predict the wait time for a ride in a given window, we added historical information of that ride and for other similar rides:

- The posted wait time of the ride in the same window exactly 7 days ago. This is the most recent same-time window for which we have data, due to the forward-looking prediction.
- The average posted wait time in the same window over the past 30 days for the ride we are predicting and other similar rides.

We also added categorical variables depicting properties of the day for each window, specifically:

- Months or the aggregation of these months into the four seasons.
- Differentiating between weekends and weekdays
- Major U.S. Holidays e.g July 4th, Christmas, New Year's, Thanksgiving, etc.

The historical information helps capture the wait time trends in the park and the categorical variables help represent the calendar effects that affect wait times in the park.

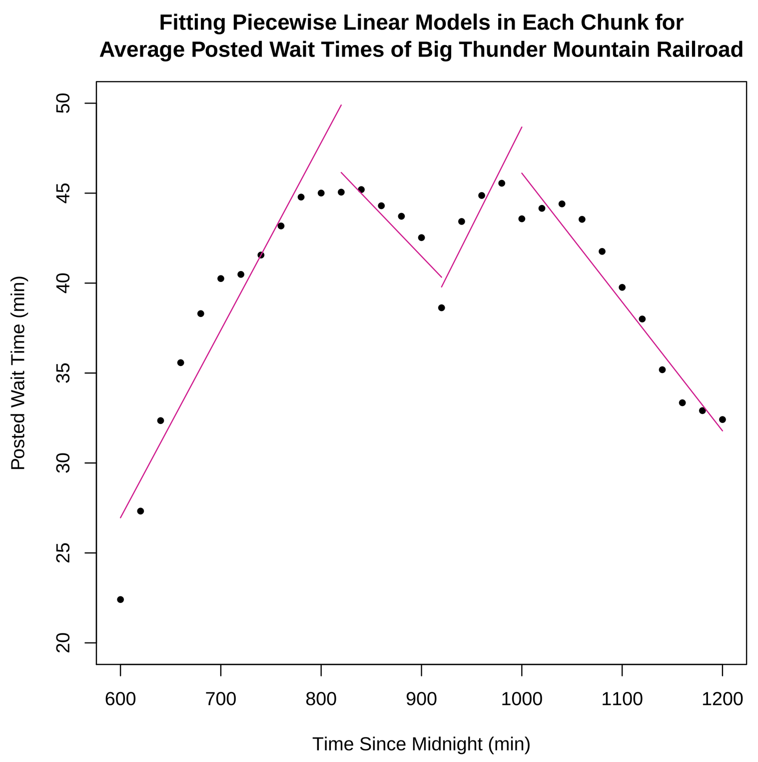

Figure 3: Linear Chunk

In Figure 3, we used local extrema of the curve to divide it into chunks which we then fit linearly. We also considered using a parabola for the hours before the midday 'dip' and another for the later hours. Our third method fits a parabola to the entire day and adds an 'optional' parabola from the 'dip' at midday to closing. This method is used in the Winnie the Pooh Ride Regression Model shown above. This method is as expressive as the double parabola method in capturing the 'dip' for Magic Kingdom but is still flexible enough to be used for other parks like Animal Kingdom where the midday 'dip' is absent.

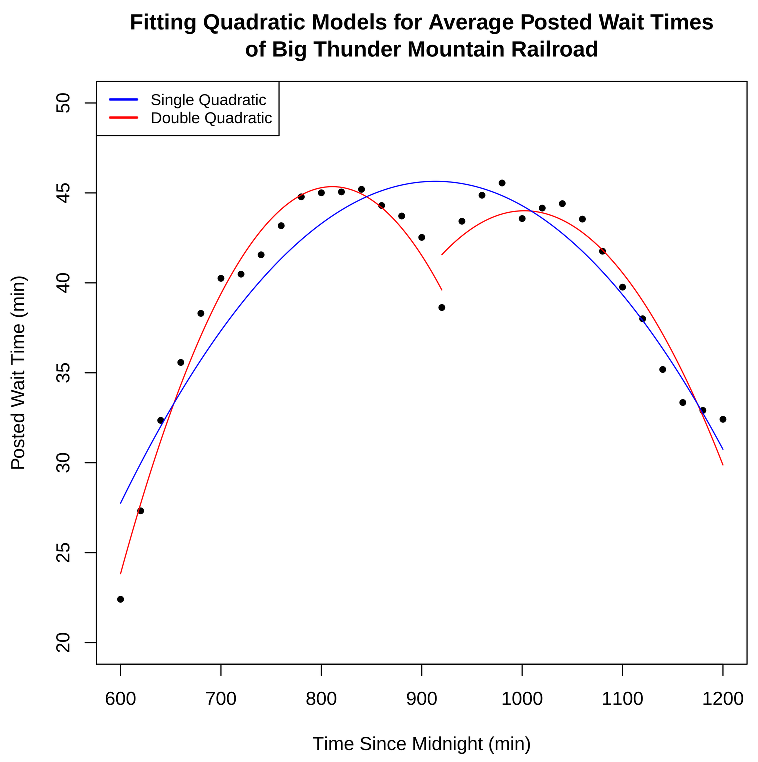

Figure 4: Optional Quadratic

Results

(i) Seasons summer, winter, and spring represent a $2.5 - 3.5$ minutes increase in posted wait time as compared to fall.

(ii) Christmas, July 4, New Year, and Thanksgiving can each increase posted wait time by ~ 10 minutes as compared to a normal day.

(iii) Easter and Labor Day actually represent a fall of ~ 5 minutes, which could be because these days are traditionally observed by other activities.

(iv) Of the historical posted wait times for all rides, the posted wait time for Winnie the Pooh exactly a week ago is the most significant.

(v) Other historical data typically also represent some increases and decreases in posted wait time.

(vi) Model Interpretation: The coefficient for Stitch's Great Escape is 0.33. Thus, a 10 minute increase in Stitch's Great Escape's Historical Wait Time will represent a 3.3 minute increase for Winnie the Pooh's Wait Times.

New Data Collection

Our data ranges from late 2009 to early 2016, but the most consistent and reliable information is concentrated in 2013-2015. Even then, the data is not evenly distributed throughout these three years and may be sparse in large chunks of time. To counter this issue, we have begun scraping posted wait times from Disney using an open source API in Node.js. We are collecting data for all rides in all four Disney World parks every 2 minutes from 8 a.m. to midnight every day. This new data set has much more detailed descriptions of rides' statuses such as "temporary closing" and "down for maintenance".

Future Work

- We want to run the models on the new data we are collecting to validate the models and make further adjustments.
- Future extensive analysis could include modeling the impact of temporary ride closings on wait times of other rides in the same park.
- More continuous data could enable us to do network level analysis in the park. This will ultimately allow us to better understand customers' strategic queuing behavior.

Acknowledgements

Many thanks to the following for their support and guidance:

Faculty Mentors: Alessandro Arlotto and Yehua Wei, Decision Sciences, Fuqua School of Business

Project Manager: Andrew Frazelle, Ph.D. Candidate, Decision Sciences

Data+ Team: Paul Bendich, Ariel Dawn, Katherine Peterson, Rhonda Stanfield, and Ashlee Valente

Data Provider: TouringPlans.com

This material is based upon work supported by the National Science Foundation under CAREER Award No. 1553274.

Contact Information

- Andrew Frazelle: andrew.frazelle@duke.edu
- Oscar Li: runliang.li@duke.edu
- Jerry Pan: jerry.pan@duke.edu
- Lei Qian: lei.qian@duke.edu