

Diane Quillao

Analyzing Disney Park Highlights

The Walt Disney Parks and Resorts Corporation is a leader in the theme park industry, both in annual revenue and attendance. As a fan of the Disney parks, I personally feel as if the Disney parks provide guests with an experience that is rarely matched by other theme parks. By performing topic modeling on the Disneyland (CA) and Disney California Adventure parks reviews, I was hoping to gain further insights on which aspects of the parks leave guests with a lasting impression. By looking at these topics, Disney can see which aspects of their parks they can improve on and also see what truly differentiates their parks from other theme parks.

To obtain the data, I used BeautifulSoup and Selenium to scrape reviews from TripAdvisor. There were around 30,000 reviews from 2010-2018. For preprocessing, I used TFIDF for LSA and NMF. I used CountVectorizer for LDA. I removed stop words, kept words that had 2 or more characters, and used an ngram range of 2-4. I found that using a range that included 1 ngram produced topics with very ambiguous words, such as “ticket,” “pass,” and “line.” I also knew that the names of most attractions and rides at the Disney parks have at least two words, which is why I kept the ngram range to 2-4.

I found that LSA and LDA did not produce insightful topics. When looking at the words list of each topic, the words did not necessarily relate to each other. The words of each topic were all related to Disneyland and Disney California Adventure, but as a topic, they didn't necessarily represent anything specific about these parks. LDA did a slightly better job than LSA. A topic would have words that were specific to one of the Disney parks. But other than that, there weren't that many insightful topics, despite changing the number of topics.

Using NMF produced much more meaningful topics. When using 100 topics, I created a dataframe that listed the top twenty words for each topic. The words for the topics were very specific and related to one another. There were topics for the Fast Pass system, long lines, Disney World, Disney employees, and the specific food at the Disney parks. When looking at the topics, I found that guests enjoy the unique snacks at Disneyland/Disney California Adventure, the option to enter the parks early if you are a resort guest, and how the parks make guests feel like a kid again. A few areas of improvement Disney can look into are the long lines, food prices, and crowd management.

To expand on this project, I would like to look into doing topic modeling for each park individually. For this project, I combined reviews for both Disneyland and Disney California Adventure. I would like to see if the topics would differ had I used LSA, NMF, and LDA on the reviews of the parks separately. If there were more reviews on TripAdvisor, I would also like to look into analyzing reviews by year. I found that users would talk about the newer rides but it would also be interesting to see which rides, old or new, are consistently talked about.