Predicting Disneyland Review Sentiment Research Questions/Answers

David Berberena

Bellevue University

DSC 680 Applied Data Science

Amirfarrokh Iranitalab

November 17, 2024

Predicting Disneyland Review Sentiment Research Questions/Answers

Questions that I believe end users would be interested in knowing about the guests' sentiment concerning Disney parks are vast, but there are a few key ideas that come to mind that the general public may inquire as to how this study will use review data to benefit the park operations:

1. Does this study consider reviews for every Disney resort in the world today?

ANSWER: The data used within the research is comprised of TripAdvisor reviews from three of the six Disney theme park resorts, those being Disneyland California, Disneyland Paris, and Disneyland Hong Kong. For reviews concerning Walt Disney World in Orlando, FL, Tokyo Disneyland, and Shanghai Disneyland, more data would need to be collected.

2. What do the reviews cite as the most positive part about visiting a Disney park?

ANSWER: According to the study's initial exploratory data analysis, there are many encouraging words that are commonly found within reviews rated 5s and 4s, both of these ratings being considered highly positive and positive, respectively. Word clouds were generated within the study that showcased these word commonalities for each class, and the words can be seen as the descriptions of the most positive part about visiting a Disney park. The word clouds can be seen below:

Highly Positive Sentiment Word Cloud

Positive Sentiment Word Cloud

3. What do reviewers share about Disney parks that frustrate them the most, prompting
negative reviews?

ANSWER: In addition to the word clouds for the reviews rated 5 and 4, the research also

generated word clouds for negative reviews, or those rated 2 and 1. Here are the word clouds

for those reviews which show the elements of Disney parks that frustrate guests the most:

Negative Sentiment Word Cloud

Highly Negative Sentiment Word Cloud

4. How is sentiment determined from reviewer ratings 1 through 5?

ANSWER: The study converted these ratings to string classification groups with 5 being highly positive, 4 being positive, 3 being neutral, 2 being negative, and 1 being highly negative. These new monikers provide more insight as to the reviewer's sentiment when creating the review as opposed to a number value. The study assumes these sentiment class values as the true sentiment expressed by the reviewer, yet the research does dive into analyzing the reviews

with an automated text sentiment analyzer to see how the sentiment classes realized by the

analyzer differ from the actual sentiment class values.

5. Guests from which countries leave the most reviews on Disney park experiences?

ANSWER: Based on the data gathered for the study of reviews revolving around the three

Disneyland branches, a bar chart was created to visualize the top ten countries with the most

reviews in the data. This bar chart can be seen below:
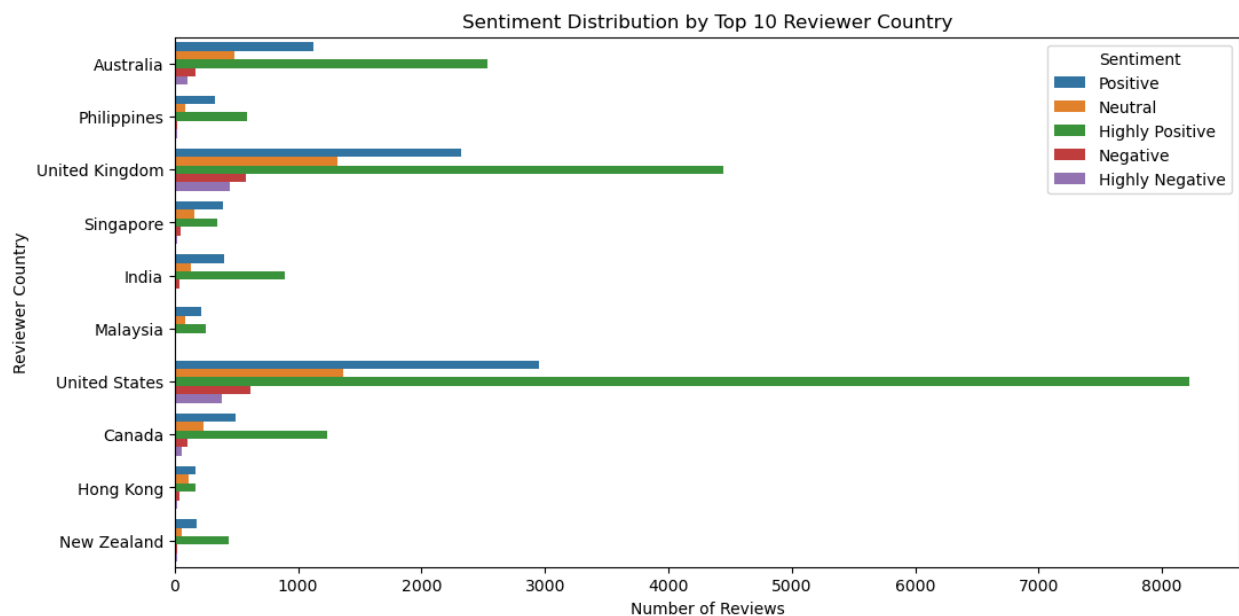
The top ten countries listed are the United States, the United Kingdom, Australia, Canada, India,

the Philippines, Singapore, New Zealand, Malaysia, and Hong Kong.

6. Which Disney park has the highest average monthly satisfaction rating?

ANSWER: The reviews featured in the dataset come from many years and months, so many line

charts were used to understand the underlying monthly rating trends. The line chart with the

most recent year comprising reviews from every month of that year will be shown below to see

the latest average monthly rating trends for the three Disneyland branches.

Average Monthly Ratings by Disneyland Branch in 2018
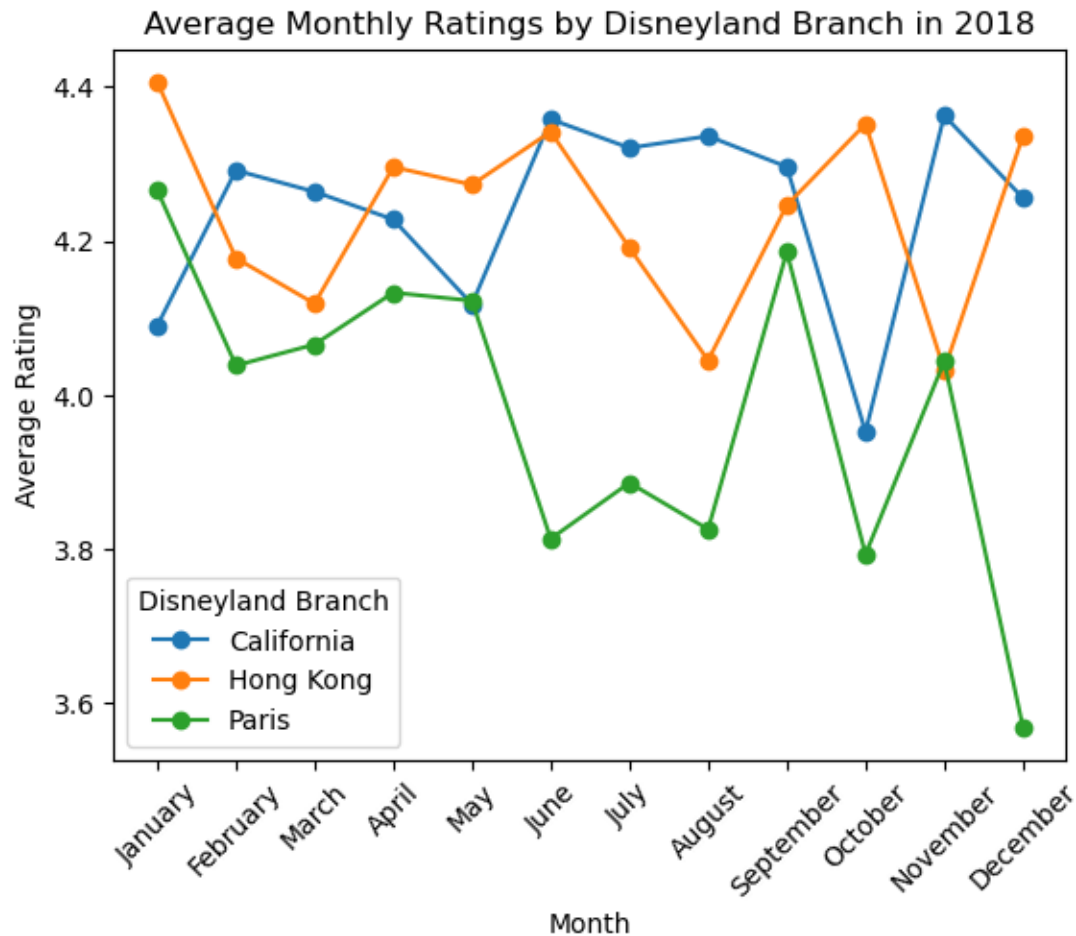
According to the line chart, Disneyland California looks to be the branch that has the highest average monthly satisfaction rating for the most months out of the year of 2018, the last year within the data that has a complete year's worth of review data. California holds the top spot in seven months out of the year, while Hong Kong has the highest satisfaction rating in the other five months, leaving Paris as the park with the lowest average monthly satisfaction rating.

7. What are the predictive models crafted in this study supposed to accomplish?

ANSWER: With sentiment analysis, it may seem ambiguous as to what the purpose of the predictive models is. Each model crafted within the confines of this educational research is intended to analyze the sentiment within each review, learn from the reviews, and predict the sentiment of future reviews based on the reviews it has already trained on. This

accomplishment is intended to help better understand the guest consensus concerning specific elements of Disney theme park operations globally, and given the sentiment of guests currently, begin working on perpetuating and expanding the positive sentiment feedback while also working to address the negative sentiment feedback for a better guest experience.

8. Why are techniques such as removing stopwords and porter stemming used?

ANSWER: These textual preprocessing techniques are used so that certain words, prefixes, and suffixes that are minimally important to the overall sentiment of the review are removed. The removal of such review elements helps to avoid predictive models mistakenly classifying more reviews as neutral than there actually are. If a review is saturated with words like 'the,' 'and,' 'or,' and 'it,' the overall sentiment of the review would fly closer to neutral due to the reduced number and impact of the other sentiment-indicating words. Preprocessing the review text brings the important words in the review to the forefront while leaving behind insignificant words that present as noise to the predictive models.

9. Why isn't the accuracy metric the only performance value used in this research?

ANSWER: While accuracy is a significant performance metric for classification problems such as the one outlined in the study, other metrics like the precision, recall, F-1 score, and confusion matrix statistics hold great value as they address certain questions hidden behind the definition of accuracy. Accuracy is typically considered the percentage of correct answers among total answers, yet how do we know what predictive models do with the incorrect answers? The other metrics come into play, as they address model behavior such as false positives, false negatives, and how the model classified each review, whether the model was correct in its predictions or not.

10. What can the Company do with this study's results should they be viable?

ANSWER: If the research provided viable results (which it did not unfortunately), the Walt Disney Company could better understand the roadblocks they are facing with their theme park guests which drive negative reviews, and work to directly address them. This analysis and subsequent data-driven decision to intervene in the guest experience for the sake of a better Disney park adventure would spur customer-business relations and perpetuate the brand reputation of the Walt Disney Company. The positive points mentioned in the reviews could be highlighted as well, with rewards being provided to those Cast Members who exhibit behaviors that continue the positive acts and/or elements showcased by guest feedback. Employee morale would be boosted in this way, driving the Company towards increased guest and employee retention.