Disneyland Analysis

Jack Ogozaly

5/5/2021

Disneyland Analysis

This pdf gives a high level overview of an analysis of Disneyland reviews. This analysis primarily focuses on using natural language processing to extract insights from ~42,000 text reviews of various disneyland reviews.

The questions this analysis seeks to answer is:

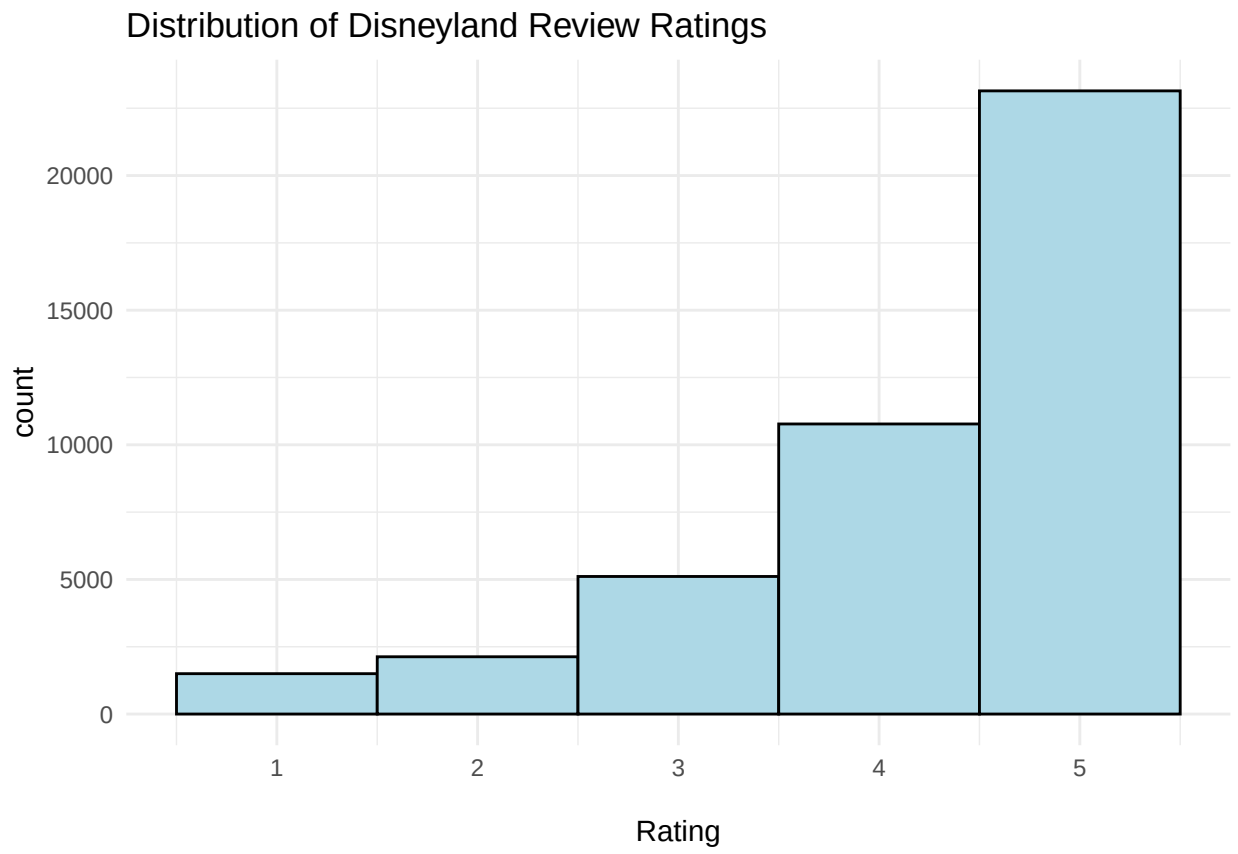
1) How are the different Disneyland branches performing?
2) What rides are reviewers talking the most about?
3) What are reviewers saying about Disneyland?
4) What are some positive things guests say about the parks?
5) What are some negative things guests say about the parks?

How are the different Disneyland branches performing?

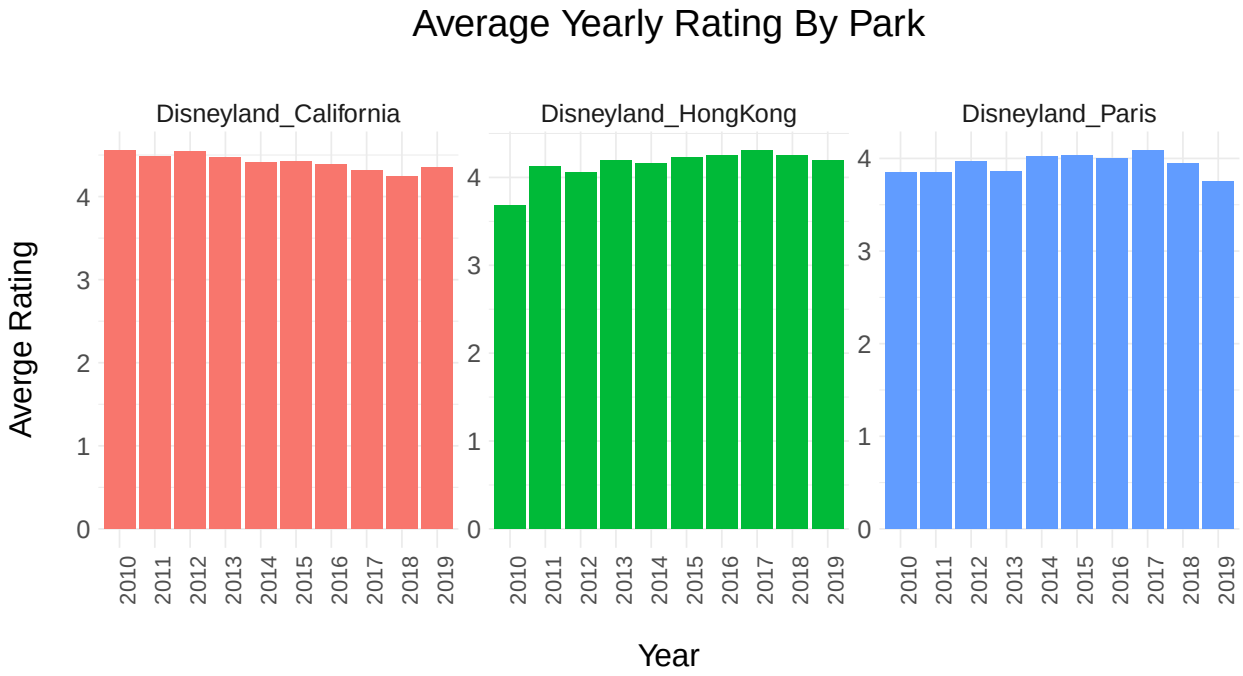
First, let's see what each branch's average rating is.

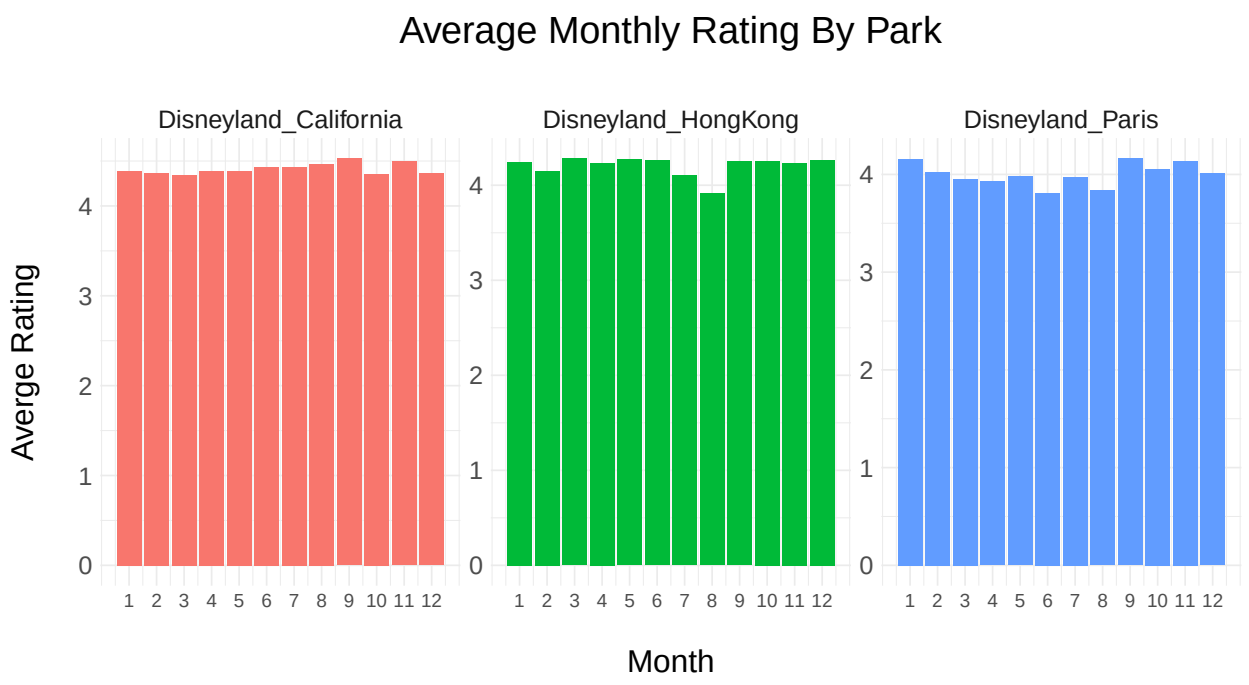
Branch	average_rating
Disneyland_California	4.405339
Disneyland_HongKong	4.204118
Disneyland_Paris	3.960012

Next, let's check the distrubution of ratings.



Overall, the reviews are fairly high and the average rating is ~4. Finally, let's see how the parks have performed for the past years and see if there's any seasonal fluctuation in ratings.





What rides are reviewers talking the most about?

For this question, we need to use fuzzy matching against a list of all the rides in the parks.

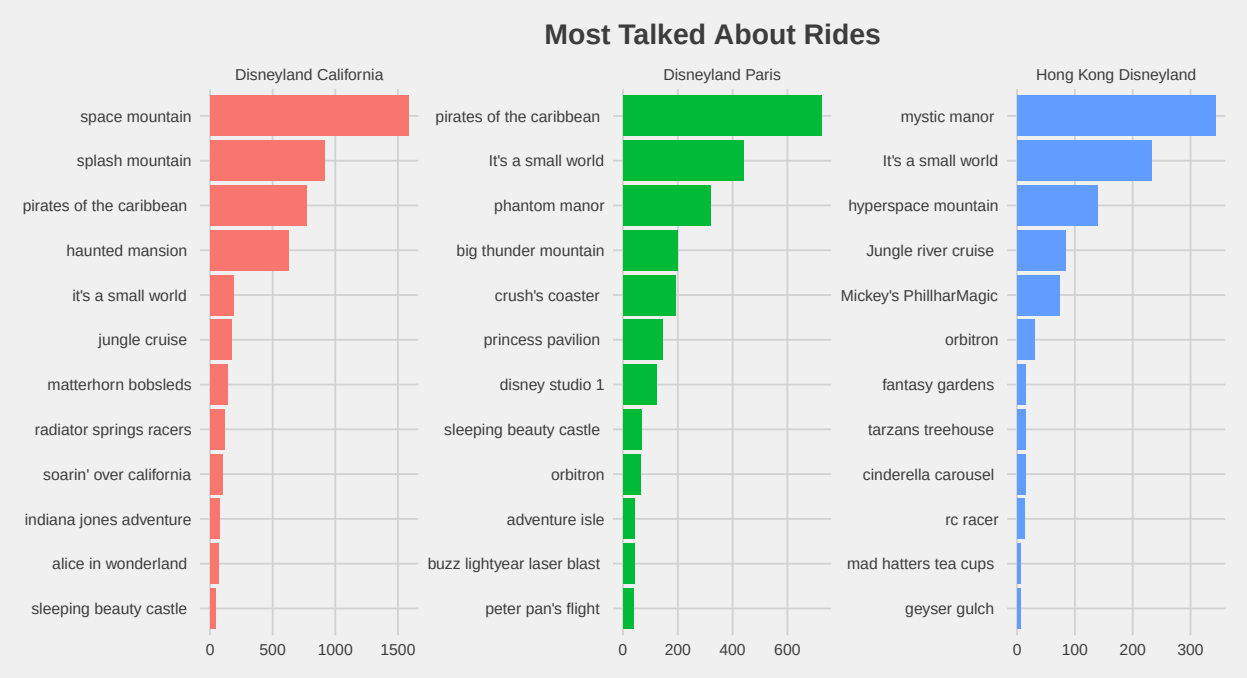
To do that, we'll just need to use a while loop to count how many approximate matches each ride has. We'll repeat this process for every park.

```
ride_data <- read.xlsx("Rides_Data.xlsx")

#Filter for Cali
cali_ride_data <- filter(ride_data, Park=="Disneyland California")
cali_data <- filter(disney_data, Branch=="Disneyland_California")
dataset <- cali_data$Review_Text

#Count cali ride popularity
rides <- cali_ride_data$Attraction
i=1
result <- vector("list", length(rides))
while (i <= length(rides)) {
  ride_to_search <- rides[i]
  result[[i]] <- length(test <- agrep(ride_to_search,dataset,value=T))
  i = i+1
}
#Collect results
df <- data.frame(matrix(unlist(result), nrow=length(result), byrow=TRUE))
cali_ride_data$occurence_count <- df$matrix.unlist.result...nrow...length.result...byrow...TRUE.
```

After we repeat this process for all the parks, we can combine our results into one dataframe and start making visualizations.

```
all_ride_data <- rbind(hong_kong_ride_data, paris_ride_data, cali_ride_data)
all_ride_data <-
  all_ride_data[order(all_ride_data$occurence_count, decreasing = TRUE),]
```

Now let's see what each park's most popular rides are.



This is interesting because it gives us another metric for customer engagement in the parks. Whereas Disney has easy access to ridership numbers in their parks, by using fuzzy matching and NLP we can see what rides customers talk about the most. Further, if we wanted to we could easily filter the most talked about rides for each rating.

Interestingly, Disneyland California has more instances of users discussing thrilling rides (such as space mountain), as compared to Paris and Hong Kong which have their most talked about rides being rides which are slower experiential rides (such as pirates).

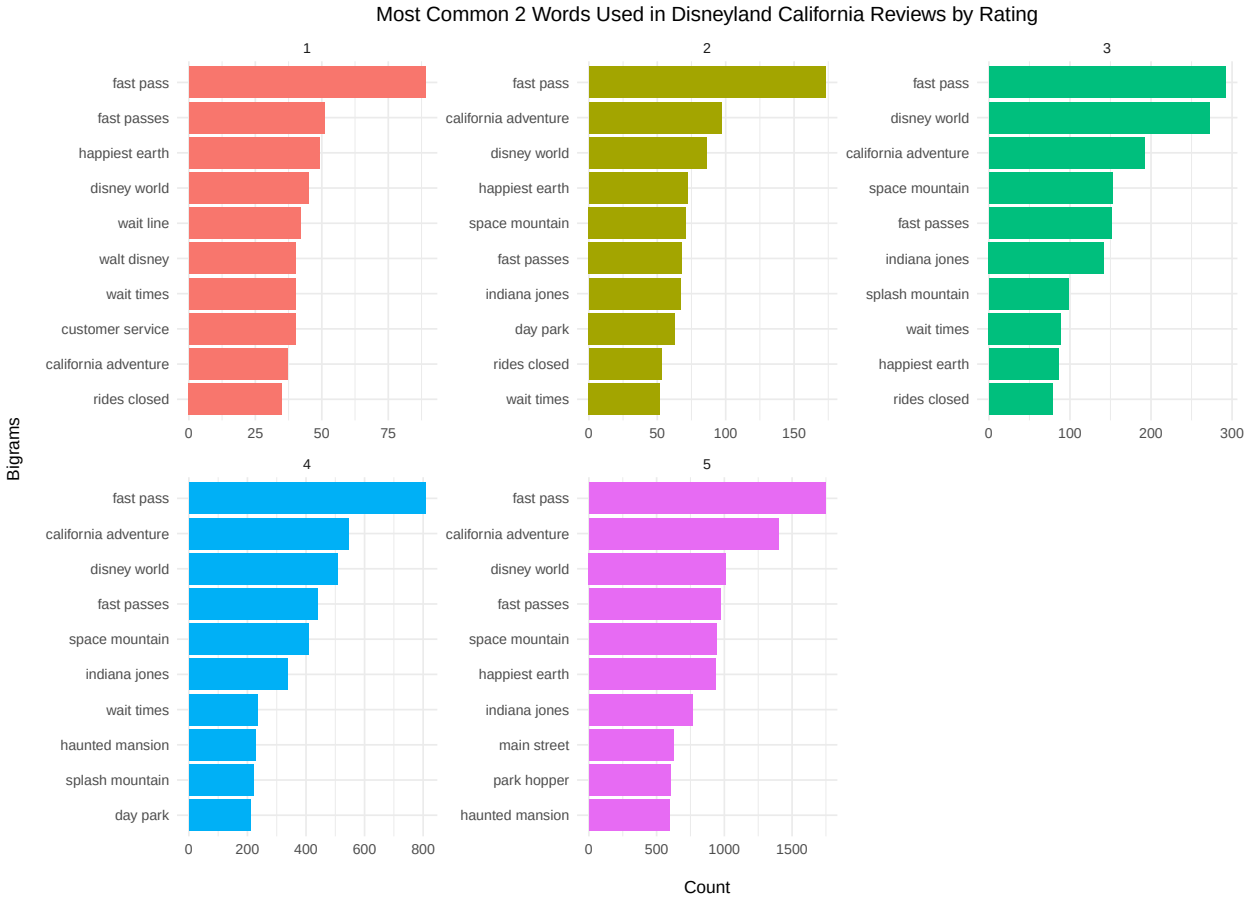
What are reviewers saying about Disneyland?

For this next section, let's see what people are generally saying about the Disneyland parks. To do this, we're going to find the most commmon 2 word combinations people are saying in these 42,000 reviews.

```
ngram_data <- disney_data
ngram_data$Rating <- as.factor(ngram_data$Rating)

#Clean our text field
ngram_data$clean_text <- ngram_data$Review_Text %>%
  removeNumbers() %>%
  tolower() %>%
  removePunctuation() %>%
  removeWords(stop_words$word)

#Make bigrams by branch and rating
```

```
bigram_by_branch_rating <- ngram_data %>%
  group_by(Branch, Rating) %>%
  unnest_tokens(word, clean_text,token = "ngrams", n = 2) %>%
  dplyr::count(word, sort = TRUE)%>%
  filter(row_number() <= 10)
```

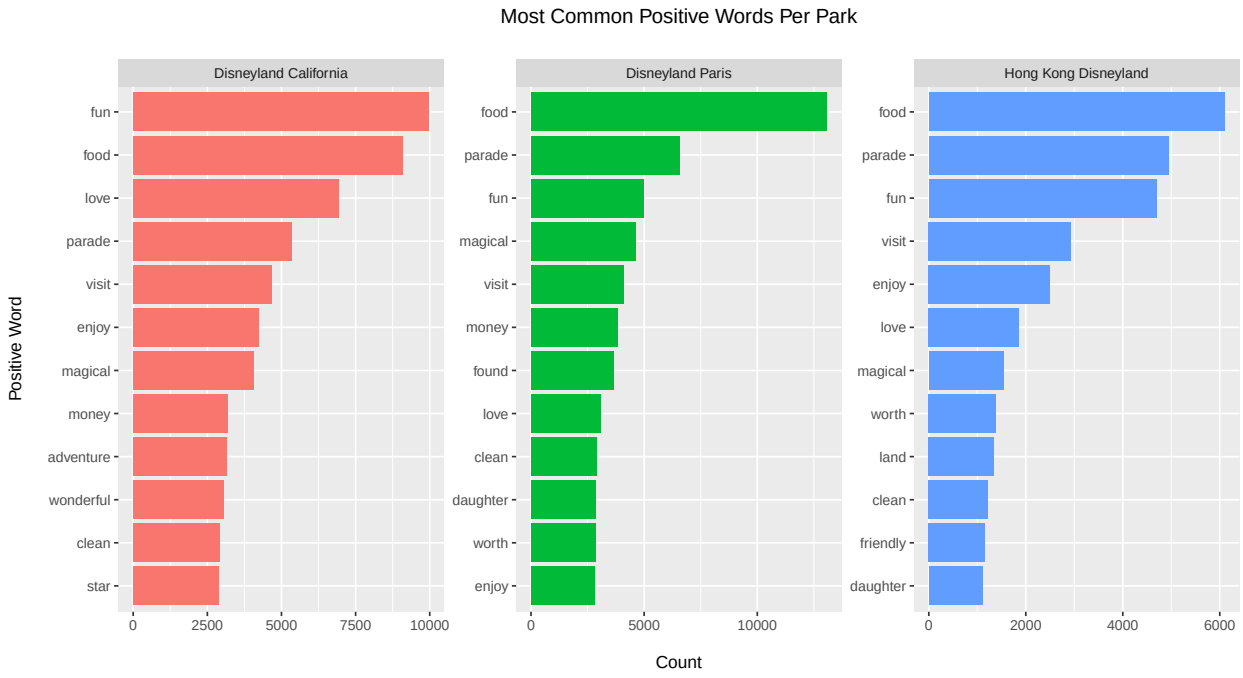
After running this code, we'll have a dataframe with the most common words said about each branch and for each rating. Let's see what people are most commonly saying at Disneyland California per rating.



As we can see from this graph, low rated reviews tend to mention things such as "wait times" more often than higher rated reviews.

What are some positive things guests say about the parks?

For this section, we just want to see the most common positive words said in all reviews. For this, we can use sentiment analysis and then view the positive words said about each Disneyland park.



What are some negative things guests say about the parks?

Similar to before, we want to see all the negative words used in reviews by park.