Shall We Go to Disneyland: Exploring Disneyland Reviews Through Visual Analytics

Patrick Zhang

Abstract— Whenever people plan a holiday, they would read the reviews of the venues they plan to visit to get an impression of whether they are worth visiting. By getting a thorough understanding of the reviews can shape people's decisions whether to go to a chosen venue or not. This report examines reviews of three Disneyland parks (California, Paris, and Hong Kong) from 2010 to 2019. Specifically, we aim to focus on understanding the reviews through sentimental analysis and visual analysis. We identify any relationships between the reviews and other features, such as the park, season and reviewer location through the use of Python visual tools and Tableau, a visual analytics software. In addition, we ulitize sentimental analysis to examine the atitudes of the reviews and park aspects that are common in positive and negative reviews.

1 Problem Statement

The Walt Disney Company is an American entertainment company founded by Walt and Roy O. Disney in 1923. For many years, the Disney amusement parks have been the place where children and those young at heart come and let their dreams come true. The original Disneyland, dubbed as "the Happiest Place on Earth", is in Anaheim, California and has been operation since 1955. At the time of writing, there are Disney parks in Anaheim, Orlando, Paris, Hong Kong, Shanghai, and Tokyo.

Throughout the report, we will analyse reviews from three Disneyland parks: California, Paris, and Hong Kong, and gain an understanding of them through visual analytics and sentimental analysis. We aim to answer the following research questions:

- Does the season affect the sentiment of the reviews at each Disneyland location?
- What are the most common aspects in positive and negative reviews?
- Which Disneyland Park has performed the best overall?

The dataset has sufficient information to allow us to utilize sentimental analysis and visual analysis. By using sentimental analysis, we can differentiate between a positive, neutral, and negative review by breaking down the reviews into different parts and analyse them individually. By analysing each part individually, we can find aspects that are common in positive and negative reviews. We can produce visuals to see trends of the reviews throughout the years. Furthermore, we can also see if there are any relationships between features like the park, date, and the reviewer location.

2 State of the Art

Fang and Zhan [1] utilized sentimental analysis on a set of Amazon reviews, dated between February and April 2014. They mainly focused on sentiment polarity categorization, which is one of the fundamental problems of sentiment analysis. Flaws that can have an impact on sentiment analysis included not being guaranteed a clear opinion and that some online data doesn't have any ground truth available. The main reason a clear opinion isn't always guaranteed is due to spammers posting spam that's meaningless or irrelevant. Without ground truth, it's not possible to distinguish between a positive, negative, or neutral opinion. Experiments for two categorizations (sentence-level and review-level) were performed. An algorithm for identifying negation phrases was developed, which takes each review and identifies phrases that include negations of either adjectives or verbs. Three classification models (random forest, Naïve Bayes, and SVM) were evaluated based on their performance in both sentence-level and review-level categorization. There were two limitations in the study. The first was that classifying reviews to the star ratings was proven difficult with review-level categorization, given that the F1 scores obtained from the experiments were lower than 0.5. The second was that the proposed sentiment analysis scheme doesn't work on reviews with implicit sentiments due to its reliance on sentiment tokens, which are words and phrases that convey sentiment.

Tun Thura Thet, Na and Khoo [5] evaluated an automatic sentiment analysis method on a set of movie reviews. Their study will focus on both sentiment orientation and a reviewer's sentiment strength on different movie aspects. The algorithm they developed breaks sentences into clauses and computes the clauses' sentiment scores based on the grammatical dependency structure and the words' sentiment scores. The sentences were broken into clauses for all the aspects to be analysed separately. An aspect's sentiment score is an average of all the clauses that discuss the same aspect. A sentence's sentiment score is the sum of all the aspects' calculated scores. During the experiment, there were sentences that were misclassified and the source for those errors came from either the algorithm, user error, prior score, or indirect expression.

Taking these two papers in consideration, they will provide this study a possible analytical process. To get the true sentiment of the reviews, each of them will be broken down into smaller parts. The rating alone isn't sufficient to find out how the reviewers truly feel about the park. Different aspects of the parks will be examined to find out the common aspects in positive and negative reviews.

3 Properties of the Data

The data came from Kaggle [2], and it consists of 42,656 reviews of the three Disneyland parks (California, Paris, or Hong Kong), which were extracted from Trip Advisor. All the reviews were from 2010 to 2019. The dataset contained six columns: *Review_ID*, *Rating*, *Year_Month*, *Reviewer_Location*, *Review_Text*, and *Disneyland_Branch*. New columns were derived from the existing columns to assist us with the analysis.

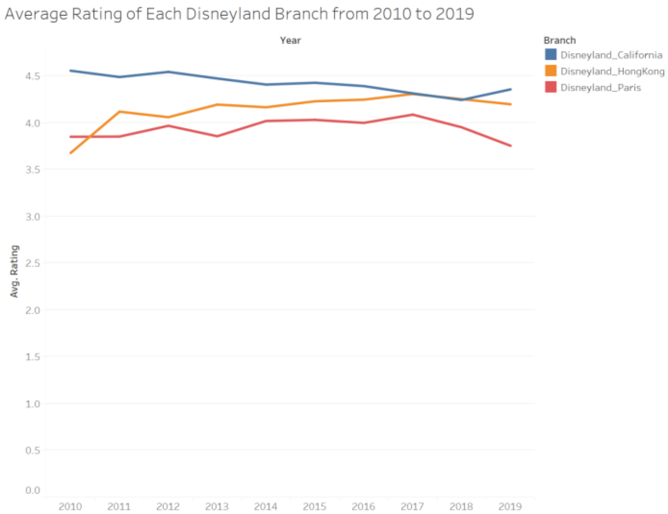

Fig. 1: Average Rating of Each Disneyland Park from 2010 to 2019

The *Review_ID* column is where each review is assigned a unique ID number. The *Rating* column is the reviewers' rating for the park they reviewed, ranging from 1 being unsatisfied to 5 for being satisfied. With the rating, we can derive the initial sentiment (positive, neutral, and negative) and compare it with the sentiment polarity score, which is the overall sentiment of the review. Figure 1 shows a line graph of each Disneyland park's average rating between 2010 and 2019. Based on the average ratings, Disneyland California performed the best overall. The *Year_Month* column is the date the reviewer visited the park. For the analysis, it will be split into *Year* and *Month* columns as we'll derive the seasons from the month and analyse trends throughout the years with visuals. The *Reviewer_Location* specifies the country the reviewer is from. Although it may not hold much relevance in the analysis, it will be great to visualize any patterns in the features such as the rating and the reviewer's location. The *Review_Text* column is the full review text that the reviewer left on the Trip Advisor site. Sentences from each review will be extracted and broken down into smaller parts for sentiment analysis. By breaking each review into smaller parts. Finally, the *Disneyland_Branch* column shows which Disneyland Park the reviewer reviewed.

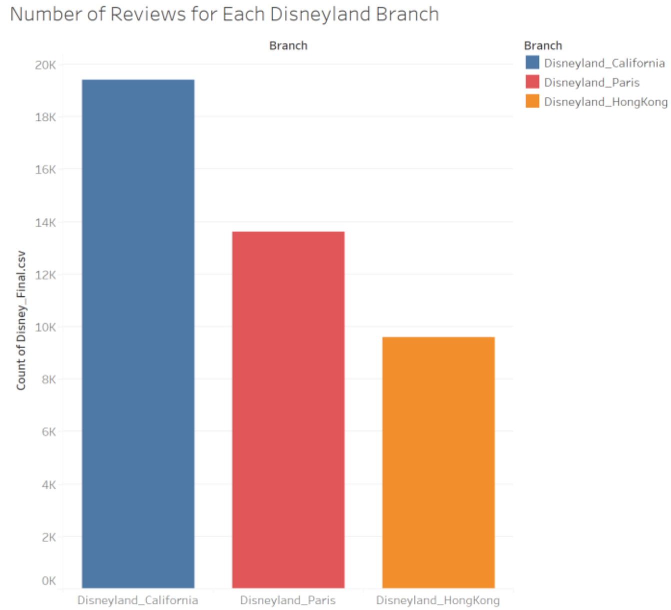

Fig. 2: Total Number of Reviews for Each Disneyland Park

The dataset had no missing values, which will make data pre-processing easier. However, it contained duplicate reviews and some of the rows had the string, "missing", in the *Year_Month* column. The duplicates were removed from the set, leaving 42,632 reviews remaining. Figure 1 shows the bar graph of number of reviews for each Disneyland branch after all duplicates where removed. The bar graph clearly showed that Disneyland California had the most reviews. For rows with a missing date, they weren't dropped as the main goal is to analyse the reviews and get insightful information of the overall sentiments for each of the reviews. All reviews will undergo through a process to remove punctuation, numerical values, stopwords, and converted into lowercase characters and each word in every review was lemmatised and tokenized.

4 Analysis

4.1 Approach

Figure 3 presents the workflow diagram of the approach to solving the proposed research questions. Below the diagram, each step will be discussed further in detail.

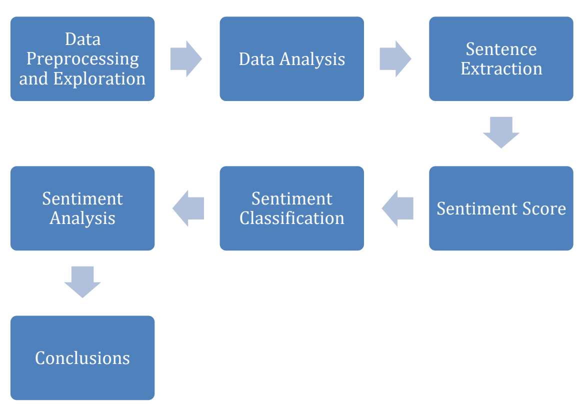

Fig. 3: Analysis Workflow

DATA PREPROCESSING AND EXPLORATION

The first step of the analysis is to load the dataset, do some pre-processing, and explore the dataset properties. The duplicates were removed and the rows with no date were kept as the main goal is to analyse the review texts. Initially, it was planned to remove the rows without a date. However, removing these rows not only is a loss of data, but without these reviews, the quality of the analysis can be affected. Given the existing columns, they can be used to create new columns that will be useful for the analysis.

DATA ANALYSIS

After obtaining a clean and processed dataset, we take it and conduct some data analysis on it. We will look for trends over the years and any relationships between the features. Tableau will be useful mainly to create visualisations. Python will be used for sentimental analysis to produce word clouds that contain the most common words and aspects in all the reviews.

SENTENCE EXTRACTION

All the reviews will be broken down into smaller parts. Punctuation, numerical values, and stopwords, are removed from the review texts. The stopwords list will be updated to filter out words that are irrelevant so that we can find which aspects of the park that are common in all the reviews. All uppercase letters are converted into lowercase characters and each word in every review was lemmatised and tokenized. Lemmatisation takes all infected forms of a word and treated them as a single entity, which is the word's dictionary form. Tokenization breaks sentences into small units called tokens, which can be words, subwords, or characters. Word clouds will be constructed to see which words are commons in all reviews, positive reviews, neutral reviews, and negative reviews.

SENTIMENT SCORE

As mentioned earlier, the rating alone isn't sufficient to determine the true sentiment of a review. The sentiment polarity score for each review will be obtained. Based on the content of the review, the polarity score obtained will range from -1 to 1, with 0 representing neutrality.

SENTIMENT CLASSIFICATION

Each review will be classified as positive, neutral, or negative based on the rating. That will be the initial sentiment. Once obtaining the polarity score, a new column called *Polarity Sentiment* will be created. A sentiment will be assigned to each review based on the polarity score.

SENTIMENT ANALYSIS

We will take all the reviews and find out which park aspects the reviewers talk about the most. Once we find out the common aspects, we investigate the positive, neutral, and negative reviews and see how these aspects impact the reviewers' sentiments.

CONCLUSIONS

Taking all the results collected from the study, we will reflect what has happened and explain whether the research questions have been answered or not. In addition, we will address areas of improvement and further work.

4.2 Process

Before the analysis was conducted, it's crucial to pre-process the data and afterwards, understand the properties of the dataset with exploratory data analysis. We will check the data types and confirm that there are no missing values and duplicates. The reason one must check the data types is to see if the data collected is in the correct format. It turned out that there were no missing values and all the data entered is in the correct format. However, some of the reviews had a missing date, as noted by the string, "missing", in the *Year_Month* column.

Once we obtained a clean and processed datsaset, we will analyse the dataset to find any trends that are useful for our analysis. As previous shown in Figure 1, Disneyland California had the largest number of reviews, meaning most people would prefer to go to there than Paris or Hong Kong.

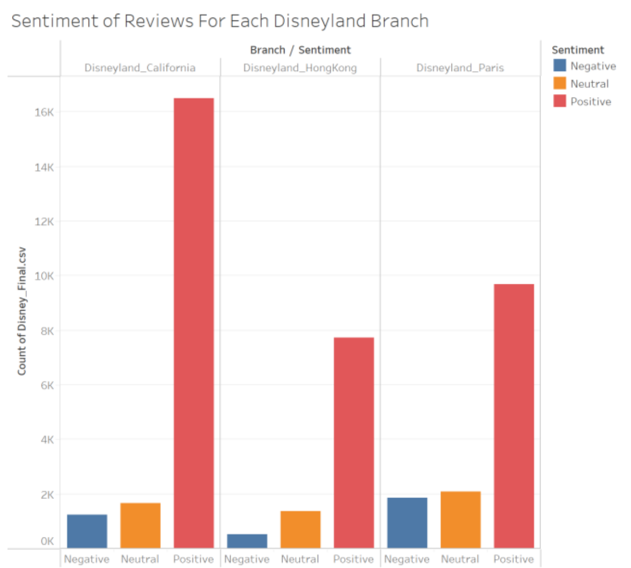

Fig. 4: Number of Review Per Season

We want to determine the initial sentiment based on the rating each reviewer gave for the parks. Each review was assigned a sentiment based on the rating. Ratings below 3 were considered negative, ratings above 3 were positive, and ratings that are 3 were neutral. Figure 3 showed that Disneyland California had the most positive reviews while Disneyland Paris had the most negative reviews.

We want to analyse trends over time, so we took the *Year_Month* column and split it into two new columns called *Year* and *Month*. First, the reviews with no date were separated from the original dataset and those with a date were broken down into *Year* and *Month* columns. Afterwards, Seasonality is taken into consideration, as one of the research questions relates season to sentiment. As a result, a column named *Season* was created based on the month. For example, if the review took place in January, it would be winter.

Reviews without a date were given the string, “N/A”. Review were grouped together by season. It turned out that most of the reviews took place during the summer, meaning a lot of people tend to visit Disneyland during the summer. That’s the time when the children are off school, and families tend to visit Disneyland. Figure 5 shows a line of the average rating of the parks over the years during each season. In general, Disneyland California performed the best while Disneyland Paris performed the worst.

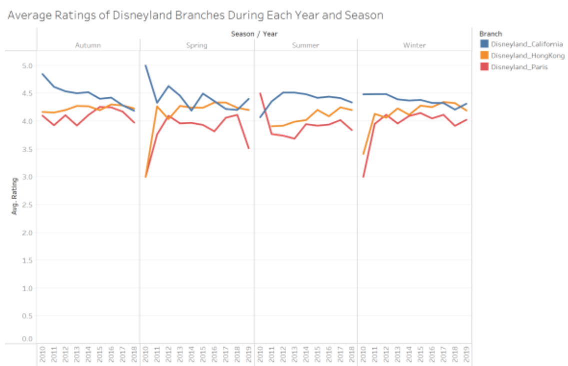

Fig. 5: Average Ratings Over the Seasons for Each Disneyland Park

After all the exploratory data analysis was complete, we move on to sentiment analysis. The reviews were broken down into smaller parts. Punctuation, numerical values, and stopwords, were removed from the review texts. All words were tokenized and lemmatised. Although we’ve divided the reviews by positive, neutral, and negative sentiments, we’ll focus mostly on positive and negative sentiments. We start by finding the most common words the all the reviews and then focus on the positive and negative reviews. To find the common words, WordClouds were constructed. The stopwords list was updated to filter out words that related to the park’s location, are too common or irrelevant to the analysis such as “park”, “Disney”, and “Disneyland”. The first WordCloud, as shown in Figure 6, shows all the words that are common in all the reviews. We can clearly see that a lot of reviews talk about aspects like the rides, the waiting times, and the line length.

Fig. 6: WordCloud Presenting Common Words in All Reviews

Given all the words common in all the reviews, it’s time to break them down based on positive and negative sentiments. Figure 7 shows all the WordClouds with words that are the most frequent in positive and negative reviews. Based on the WordClouds, we can tell the rides at Disneyland and the wait times are the most widely discussed aspects in both positive and negative reviews. Aside from the rides and wait times, the words “queue” and “line” also frequently appear in the negative reviews. Given that these words are common, it’s possible that they complained how long the lines were.

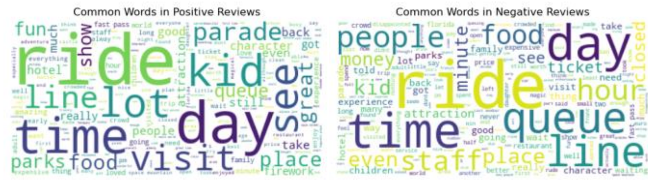

Fig. 7: WordCloud Presenting Common Words in Positive and Negative Reviews

Words aren’t enough to determine the true sentiment of a review. Now we look to calculate the polarity score of all the reviews. After getting the scores, we assigned a sentiment to each review in a new column called *Polarity Sentiment*. examined the distribution of them in histograms and boxplots, as shown in Figure 8. By looking at the graphs, it’s worth noting that some of the reviews labelled as positive have polarity scores below 0. Same can apply for the negative reviews, where there are reviews with polarity scores above 0.

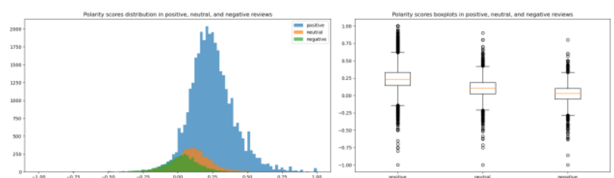

Fig. 8: Distribution and boxplots of polarity scores

One possible reason for this is that the reviews have both negative and positive words in them. Another reason is that some positive reviews had ratings higher than 3 but contained mostly negative words. Same can apply for negative reviews. For example, one reviewer from Australia gave a rating of 2 for their Disneyland Paris visit and the polarity score was 0.25, meaning it’s a positive sentiment. Even though the overall tone was negative, the words “great” and “good” may caused the polarity score to be above 0. Another reviewer from the United Kingdom gave a rating of 4 for their visit to Disneyland Hong Kong and the polarity score came out to be -0.087963. That reviewer had a good time at the park and said that there were no wait times. However, towards the end of the review, they mentioned that the number of restaurants were limited and that the bad part about the trip was that the castle was under renovations. These two aspects contributed to the negative polarity score.

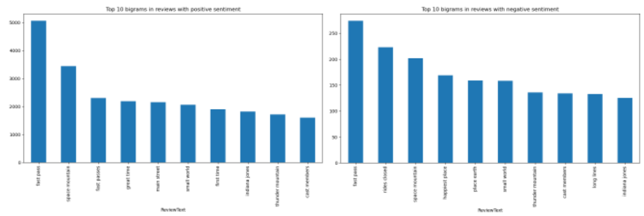

Fig. 9: Top 10 Bigrams in Positive and Negative Reviews

A single word is just the beginning to find out which parks aspects the reviewers talk about. Next, we explored bigrams, which are groups of two words from a sequence of tokens. Figure 9 shows the bar graphs of the top 10 bigrams in positive and negative reviews. “Fast pass” came out on top in both reviews. Looking at fast pass reviews, some say how useful the fast passes are while others say the fast pass doesn’t work up to their expectations. “Thunder mountain”, “indiana jones”, “space mountain”, and “small world” are bigrams that relate to names of the rides at the parks. Since they’re common in the reviews, it can infer that these are the most popular rides in the parks. There were some bigrams that allowed us to see the difference between negative and positive reviews. In the negative reviews, “long lines” and “rides closed” are common, meaning the reviewers complained a lot about long lines in the attractions and that the rides are closed due to either renovation or weather. In the positive reviews, “great time” meant that the reviewer had a great time at the park they reviewed, and “first time” meant that the reviewer came to the mentioned park for the first time and had a positive first impression.

Disneyland Paris has performed the worst among the three Disneyland parks since it had the most negative reviews and the lowest rating averages. We decided to analyse the negative reviews and see why it was the worst.

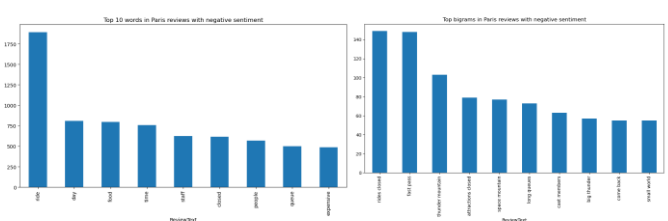

Fig. 10: Top 10 Words and Bigrams in Negative Paris Reviews

Figure 10 shows the top 10 words and bigrams that appeared in the negative Disneyland Paris reviews. Most of the Disneyland Paris reviewers complained about the rides, mostly about the rides being closed and the long lines. In addition, they complained about the cast members there being unfriendly and unenthusiastic. We decided to take a further step and filtered the reviews to see what the reviewers said about both the French language and people. It turned out that reviewers had problems with both. For the language, some reviewers said that the employees ignored them because they don’t know French. Others had wished that the employees learned English so that they can interact better with the foreign visitors. In terms of the people, reviewers said that they had problems with how the French behave like how they cut in the lines and that they’re tend to be rude to the park visitors.

4.3 Results

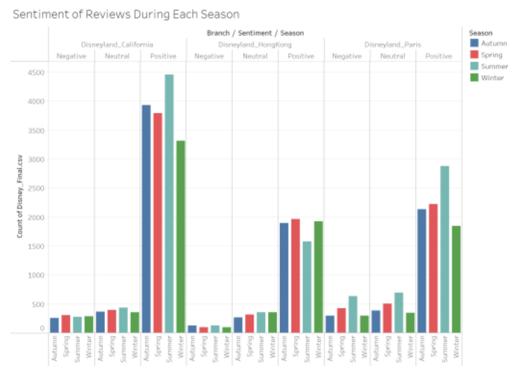

Fig. 11: Sentiment of Reviews During Each Seasons

The first research question asked whether season influences sentiment. Figure 11 shows a bar chart of the reviews at the parks during each season, given the sentiments. The California and Paris Parks received the most positive reviews during summer while the Hong Kong Park received the most positive reviews during autumn. The California Park received the most negative reviews during spring while the Paris and Hong Kong Parks received the most negative reviews during summer. Results will be further expanded in the next section.

The second research question asked about the common aspects in positive and negative reviews. The aspects that are common in positive and negative reviews are the rides and waiting times. Figure 12 shows the top 10 bigrams in the reviews and four of them are ride names.

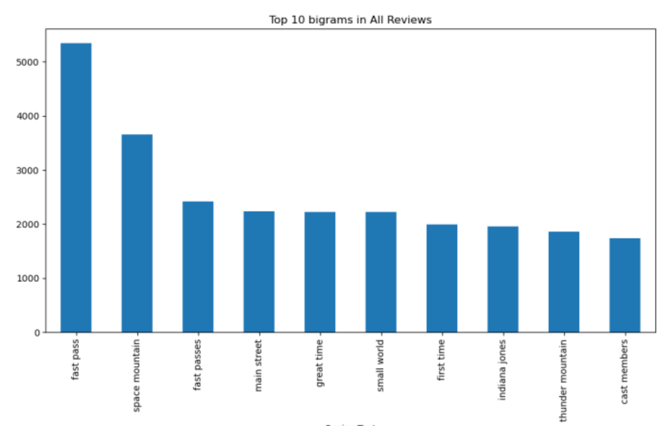

Fig. 12. Top 10 bigrams in the Reviews

The last research question asked about the best performing Disneyland Park. Disneyland California performed the best overall. The park had the highest rating average during the 9-year span. In addition, it had the largest number of positive reviews. While examining the bigrams for the positive California reviews, “happiest place” and “place earth” came out in the top 10. Disneyland California is best known as the “Happiest Place on Earth.”

5 Critical reflection

This study shows that sentiment analysis is crucial in industries like tourism and commerce. It's used to detect customers' complaints to improve their business or products. In terms of Disneyland reviews, they can help Disney understand the customers' pain points and work to improve the experiences at the parks.

The methods utilized worked well in this study, but the methods utilized aren't enough to determine the difference between positive and negative sentiments in a review. The approach was able to identify the common words in all the reviews and in positive and negative review. The polarity score helped to assist us in understanding the true sentiment of a review, although some of the sentiment scores don't alight with the rating. However, the approach used is limited to a certain degree. First, it doesn't help determine if a reviewer went into depth about a certain park aspect. For example, some of the reviews that talk about fast passes just mention the term without going further on about them. People reading the reviews should be able to get a feeling on whether to get fast passes or not. It also doesn't go in depth of the emotions the reviewers experienced at the parks. Readers should know how the reviewers felt at the attractions at the parks.

Expanding on the results for the first research question, we've dived into the negative reviews that mention weather. In the negative Hong Kong reviews, the reviewers went on days that were either too hot or on rainy days. During the summer, there are hot and humid temperatures and showers that affect the weather there. Tropical cyclones affect Hong Kong between July and September [3]. In the negative Paris reviews, rain and hot temperatures were also mentioned and a few reviewers said that Paris's weather was overcast. One reviewer even gave a tip to check the weather before going. For the California reviews, there were more reviews that mentioned weather during autumn and winter. Reviewers mostly mentioned strong winds and rainy weather there. Some of the Paris and Hong Kong reviewers said that the rides get shut down due to bad weather. One Hong Kong reviewer said that a cyclone warning caused the rides to close indefinitely.

Further work will include emotion detection, where we will examine the type of feeling and attitude of each reviewer, like what Singla, Randhawa and Jain have done with mobile phone reviews from Amazon. In addition to positive, negative, and neutral sentiments, we want to investigate emotions like fear, anger, and joy [4]. By extracting the emotions, we can see which park aspects causes the reviewers to feel a certain emotion. In addition, word segmentation will be utilized, where all words will be categorized into entities like date, time, nation, and language. By doing so, we can improve on determining what park aspects are common in positive and negative.

Table of word counts

Problem statement	243
State of the art	417
Properties of the data	412
Analysis: Approach	495
Analysis: Process	1374
Analysis: Results	200
Critical reflection	478

References

[1] Chillar, A. (2021). *Disneyland Reviews*. [online] Kaggle. Available at: https://www.kaggle.com/datasets/arushchillar/disneyland-reviews.

[2] Fang, X. and Zhan, J. (2015). Sentiment analysis using product review data. Journal of Big Data, 2(1). doi:10.1186/s40537-015-0015-2.

[3] Hong Kong Observatory (2020). *Climate of Hong Kong*. [online] www.hko.gov.hk. Available at: https://www.hko.gov.hk/en/cis/climahk.htm.

[4] Singla, Z., Randhawa, S. and Jain, S. (2017). STATISTICAL AND SENTIMENT ANALYSIS OF CONSUMER PRODUCT REVIEWS.

[5] Tun Thura Thet, Na, J.-C. and Khoo, C.S.G. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. Journal of Information Science, 36(6), pp.823–848. doi:10.1177/0165551510388123.