

# Sentiment Analysis of Disneyland Parks

**BANA 277: Customer & Social Analytics**

**Team A3:**

**Hridyangam Garg**

**Perla Ipanaque Pulido**

**Xinyi Li**

**Wenxi Su**

**Xiaohan Yang**

## TABLE OF CONTENTS

I.	Introduction.....	3
II.	Dataset.....	4
III.	Exploratory Data Analysis.....	5
IV.	Descriptive Analysis .....	7
V.	Text Analysis.....	9
VI.	Sentiment Analysis Findings .....	11
VII.	Recommendations.....	14
VIII.	Conclusion.....	15
	References.....	16

# I. Introduction

Disneyland parks are known worldwide as popular tourist destinations, attracting millions of visitors each year. Then, with such a large volume of visitors, there are differing opinions and sentiments towards the parks. In the following paper, we have done the evaluation of sentiment analysis about Disneyland parks.

Sentiment analysis is a technique that involves analyzing text data to determine the sentiment expressed in the text. In the context of Disneyland parks, sentiment analysis involves analyzing the feedback provided by park-goers to determine the overall sentiment towards the park. This analysis can provide valuable insights into the strengths and weaknesses of the park, allowing for targeted improvements to be made. Additionally, sentiment analysis can be used to identify patterns in visitor feedback, such as popular attractions or common complaints.

The *project's overall objective* with the *sentiment analysis* is to *recognize keywords* and *sentiment in customers* in order to answer the following questions:

- What can Disneyland focus on improving based on customer experience to increase brand engagement?
- How can Disneyland leverage the sentiment insights to develop more personalized experiences for its customers?
- What are the areas and reasons for customer dissatisfaction?

We will focus on this project on analyzing specifically the top three Disneyland parks in the world, Disneyland California, Paris, and Hong Kong. By using sentiment analysis, we can help Disney find areas that need improvement and identify reasons for customer dissatisfaction. Then, this information can be used to make the appropriate changes necessary to enhance visitor experience.

## II. Dataset

Data was sourced from Kaggle, including 42,000 reviews of 3 Disneyland branches - Paris, California, and Hong Kong, posted by visitors on Trip Advisor. By comparing the ratings and conducting a sentiment analysis based on the reviews of customers, we could learn about the customer attitude to different Disneyland branches in general and provide suggestions to help improve customer satisfaction.

### A) Dataset Content Description

Column	Description
Review_ID	Unique id given to each review
Rating	Ranging from 1 (least satisfied) to 5 (most satisfied)
Year_Month	When the reviewer visited the theme park
Reviewer_Location	Country or region of origin of the visitor
Review_Text	Comments made by visitor
Branch	Location of Disneyland Park

## Data Cleaning & Processing

To begin with, we dropped all the rows that contained missing values and then identified the number of unique values in the Review\_ID column. We found that the total number of unique Review IDs was 40,023, so in the following step we dropped duplicate values from the "Review ID" column, returning a clean dataset with distinct Review IDs.

Based on the rating scores, there were 21,908 individuals who gave a rating of 5, 10,086 individuals who gave a rating of 4, there were 4,782 individuals who gave a rating of 3, there were 1,929 individuals who gave a rating of 2, and there were 1,338 individuals who gave rating of 1.

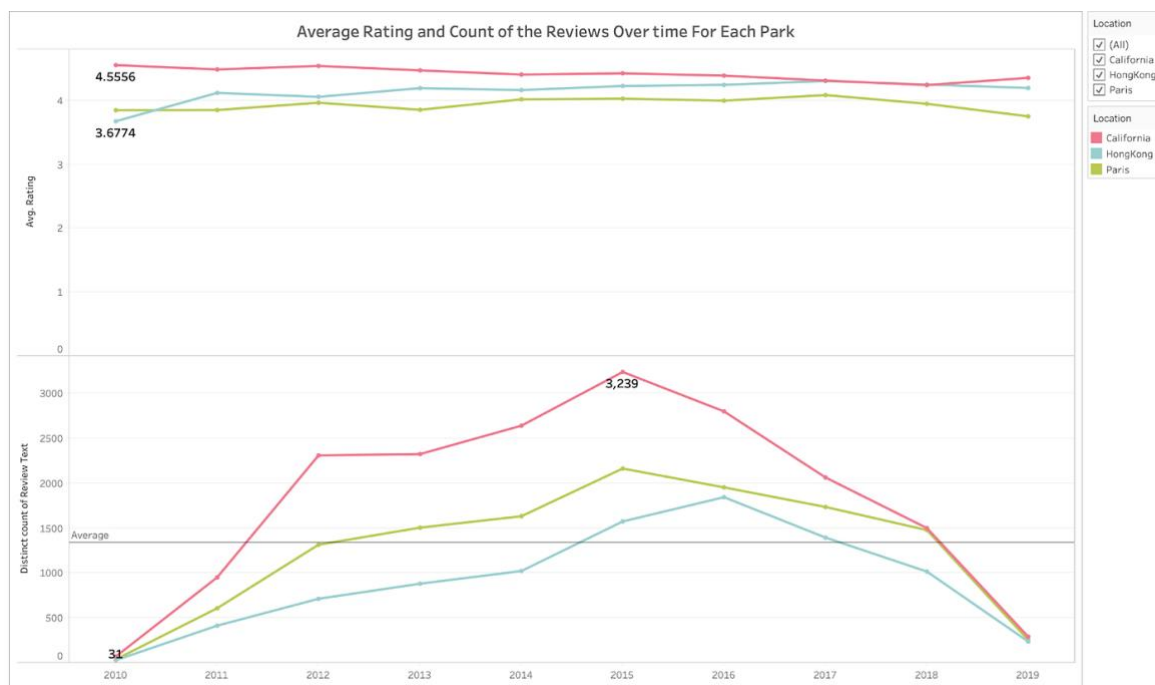
Later on, we performed pre-processing on the content of the "Review\_Text". We converted the content to lowercase, removed punctuation and numbers in texts, and then converted the reviews to tokens (i.e. broken into elements of a list). We then set stopwords to remove common words, which did not add any value to the analysis of the text, such as 'I' and 'you'. After that, we lemmatized the reviews and joined all the pre-processed review text to complete the processing of the "review" content.

In the end, we split the YEAR and Month into two different columns and dropped the Year Month column, making it easier to view and analyze the data. In this way, we could also conduct statistical analysis on ratings from year to year.

	Year	Count	Mean Rating	Median Rating	Rating Std Dev
0	2010	143	4.167832	5.0	1.216008
1	2011	1976	4.214575	5.0	1.087331
2	2012	4339	4.289007	5.0	0.990223
3	2013	4709	4.223614	5.0	1.060764
4	2014	5300	4.240943	5.0	1.035264
5	2015	6979	4.259206	5.0	1.030635
6	2016	6599	4.234581	5.0	1.031107
7	2017	5195	4.235804	5.0	1.055829
8	2018	3997	4.136853	5.0	1.135484
9	2019	786	4.111959	5.0	1.175193

### III. Exploratory Data Analysis

In order to understand the data, we did an exploratory data analysis, which involved analyzing all the ratings of all Disneyland parks around the world. By taking the average and count on the reviews over time for each park we found out that the *top three parks* with more and best reviews are *Disneyland in the following locations: California, Hong Kong and Paris*.

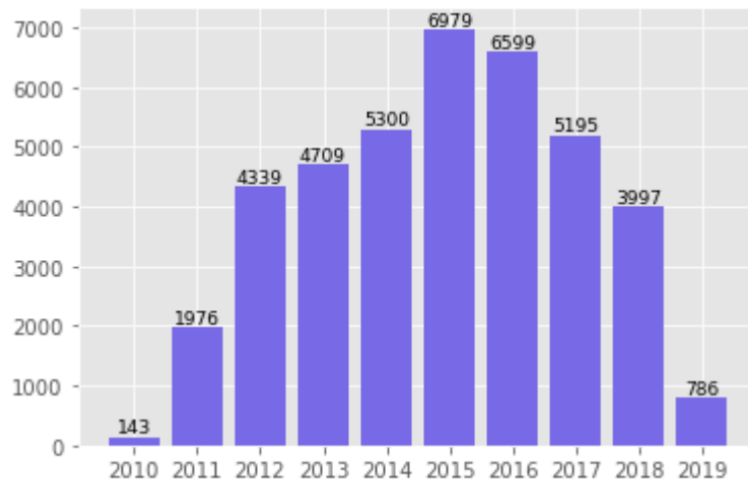


Graph 1: Average Rating and Count of the Reviews for each Disneyland Park



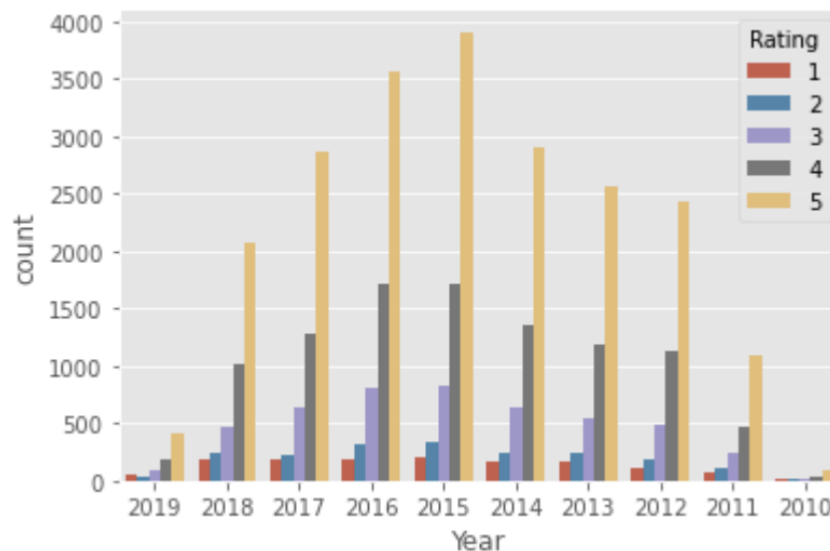
## IV. Descriptive analysis

Based on the year, we were able to perform a classification calculation on the data and observed that the count of rating followed a normal distribution. The ***highest count of ratings*** was observed in ***2015 and 2016***. Prior to 2015, the rating count increased each year, whereas, ***after 2015, the count decreased each year***.



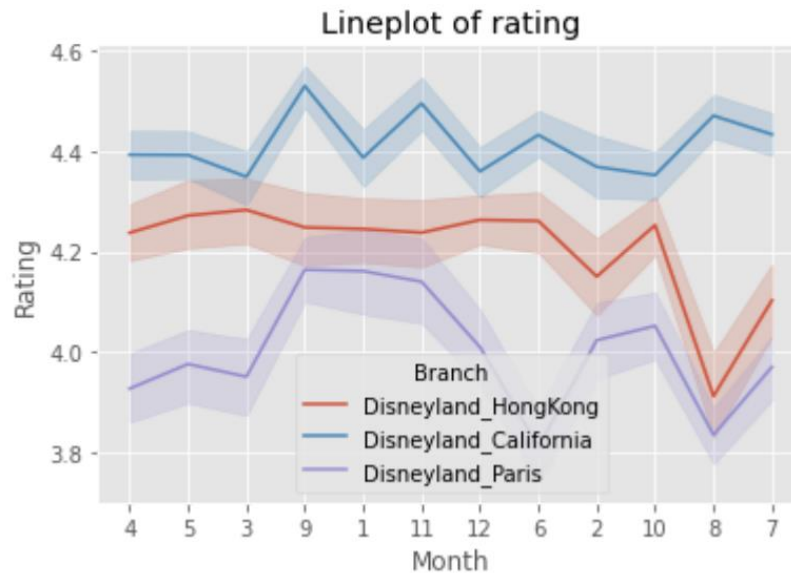
Graph 3: Total Count of Ratings from Year to Year

For all those years included in the data, we could find that ***most customers gave high ratings*** (more than 3). The bar chart shows that the rating count of 5 is particularly high.



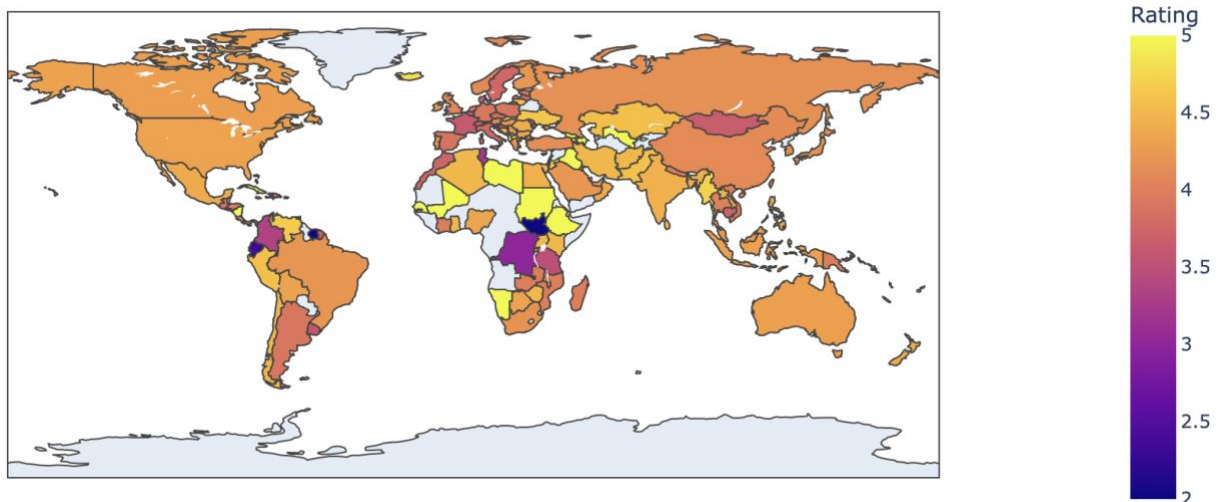
Graph 4: Total Count of Each Level of Rating by Years

We then proceeded to *rank the ratings for Disneyland branches* and found that *Disneyland California had a higher rating compared to Hong Kong and Paris*.



Graph 5: 'Ratings' vs 'Month' for the branches of Disneyland

Based on the location information, we observed that *reviewers located in North America, Asia, and Australia* were more likely to *give high ratings*, while *reviewers in Europe* gave relatively *lower ratings*. The graph indicates that reviewers located in *South Sudan, Suriname, and Ecuador* gave *extremely low ratings*, while those who were from *Mali, Libya, Ethiopia*, and some other regions rated *very high*. The main reason is that there are few rating data in those areas, thus generating extreme values, which could be regarded as outliers.



Graph 6: Distribution of Average Ratings for Each Country or Area



## V. Text Analysis

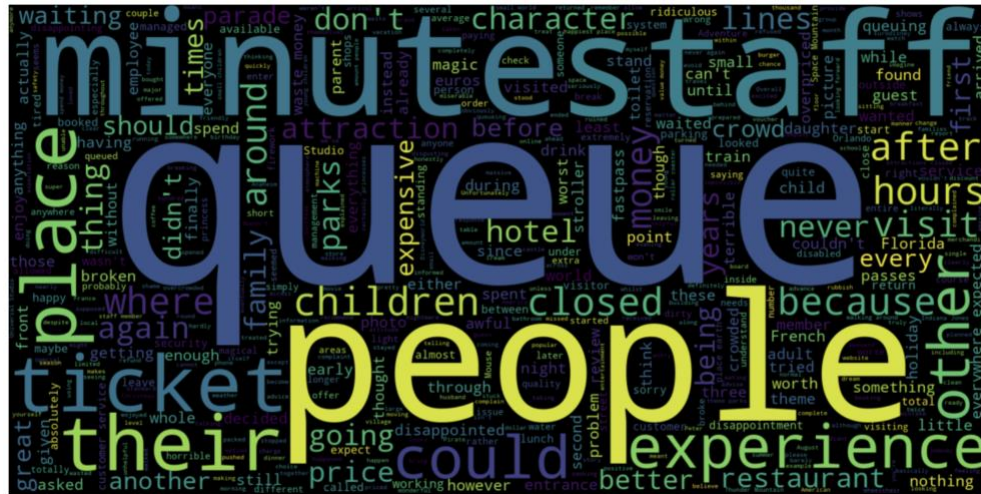
We generated *word clouds for review texts sorted by ratings* so that we found what reviewers with high ratings were satisfied with and what reviewers with low ratings complained about.

In this first text analysis graph, we identified the *key words* used by *reviewers with rating five*. In the graph below we can notice that most repetitive word used in the reviews is *great*, which can be interpreted as a *positive sentiment* about most of the parks. Furthermore, we have other most used words such as visit, first attraction, character, amazing, parks and people. We notice that these words are the most repetitive because these are things that people enjoy the most, such as



Graph 7: Keywords of Reviews with Ratings of 5

On another hand, we have that *the keywords with the lowest rating reviews* have other common words used. According to the text analysis, we can see that *the most negative words* used are *queue, people, minute staff, place, closed, ticket*. We can interpret that without a doubt the sentiment is negative because these people experience at the park hasn't been positive and these comments relate to the idea of having queue problems, conflict with other people in the park, finding some rides closed and having some issues finding tickets for some specific dates.



## Word Cloud: Disneyland California, Paris, and Hong Kong

We also made the analysis of sentiment by using word clouds for different Disneyland branches. In the following graphs we can see the most used words for each Disneyland Park and in the following pages we will explain the sentiment analysis and findings obtained from it.





Findings about *positive sentiment in reviews*:

- Disneyland California positive sentiment words related to this specific park are customer service, characters are described as very friendly, the park has a beautiful unique decoration during holidays and the favorite ride is highlighted to be space mountain.
- Disneyland Paris words that are related to a positive sentiment are stunning architecture and beautifully themed lands, from the enchanting Sleeping Beauty Castle to the immersive world of Ratatouille in the Walt Disney Studios Park. The rides are described as thrilling and fun. The food is defined as delicious.
- Disneyland Hong Kong has as positive sentiment words such as magical destination, unique and exciting. Other positive words that describe this park are clean and well-maintained, and the staff is friendly and helpful. Furthermore, about the rides the most used word is thrilling. Finally, the most positive sentiment points out that the park is a fantastic destination for families, couples, and anyone who loves Disney.

Findings about *negative sentiment in reviews*:

- Disneyland California: The customers show a negative sentiment related to the following issues. Long lines and crowds, family experience, food, accessibility, and strollers. We were able to find out that a unique problem at this Disneyland is focused on the topics related to stroller congestion, rental, parking, and theft. Most of the reviewers mentioned that cast members would sometimes move strollers if they were obstructing a path or cut their locks when necessary. This action can cause park guests to be confused, especially if they have no idea where their stroller was moved to.

- Disneyland Paris: The reviews show that the negative sentiment is associated to these topics. Long lines and crowds, customer service, bookings and accommodations, smoking, time of visit, and bad park experience. The main negative topic in this Disneyland is smoking. It looks like smoking is a common culture practice in France. Then, this Disneyland in Paris has created areas for smoking inside the park. Several reviewers complain about this issue and claim that it has “destroyed the magic.” Visitors express disgust about this and said that they feel it is not good for kids.
- Disneyland Hong Kong: Some of the negative topics identified are associated to queues, prices, cutting in line, and restaurants. The reviewers seem to complain a lot about people cutting in line and the food at the restaurants not being pleasant. It looks like the main issue about cutting lines is a great concern also noticed by Disneyland Hong Kong since they even had added in their visitor rules to be respectful between guests and be careful not to bump into, push or move ahead of others in the queue.

In summary, we found some strengths in three different locations for Disneyland, but we also discovered some specific areas that these Disneylands should improve. Some of the most common pain points for Disneyland in California, Paris and Hong Kong are the following: long wait times, crowded areas and dissatisfaction with some attractions and experiences related to capacity and guest interaction.

## VII. Recommendations

### Action to take based on sentiment analysis.

- Focus on improving the aspects of the park that are receiving negative reviews. Pay closer attention to the sentiments expressed in the reviews.
- Use the Random Forest Classifier to predict the sentiment polarity of future reviews, which could help identify potential issues and address them proactively. Keyword help to recognize strengths and weaknesses.
- Monitor the sentiment of reviews over time to track improvements in customer satisfaction.

### Recommendation applied to business.

- ***Leverage user-generated content*** to build engagement. Encouraging customers to share their experiences/photos on social media can help to build buzz and create a sense of community.
- ***Provide Entertainment in Line*** or interactive experiences, such as games or activities, to help pass the time and make the waiting experience more enjoyable.
- ***Technology*** to enhance the customer experience and increase engagement.  
For example, Disneyland could use augmented reality/virtual reality to create immersive experiences and keep people engaged.

## **VIII. Conclusion**

In conclusion, the sentiment analysis applied to Disneyland Parks reviews provides valuable insights about the public's perceptions and attitudes towards these theme parks. Moreover, through the analysis of the online reviews we were able to recognize the positive and negative sentiment and how these are related to specific words that can help Disney understand how their parks are being perceived by visitors. In consequence, recognize areas for improvement for the business. The sentiment analysis of Disneyland parks also helped to find patterns in visitor feedback, as well as insights into the strengths and weaknesses of the parks.

The results of sentiment analysis for Disneyland Park in California, Paris, and Hong Kong may vary based on the different park attractions and experience offered by each park, location and other factors. However, overall, it is clear that Disneyland is a beloved destination for many people around the world and generates the most positive sentiment across a wide range of online reviews.

Finally, by understanding the outcomes from the sentiment analysis, we were able to identify that Disneyland can gain a better understanding of their customers' needs and preferences, and tailor their marketing and customer experience strategies accordingly. Sentiment analysis can help Disneyland monitor and address any negative sentiment that arises, enabling them to maintain a positive reputation and continue to attract visitors from around the world.

## References

- Arush Chillar. Disneyland Reviews (2021) Kaggle. Retrieved from [www.kaggle.com/datasets/arushchillar/disneyland-reviews?resource=download](https://www.kaggle.com/datasets/arushchillar/disneyland-reviews?resource=download).
- Hedayati, S., & Mohammad, S. (2019). NRCLex. Retrieved from <https://pypi.org/project/NRCLex/>
- Hutto, C. J., & Gilbert, E. (2014). vaderSentiment. Retrieved from <https://pypi.org/project/vaderSentiment/>
- Loria, S. (2020). TextBlob. Retrieved from <https://pypi.org/project/textblob/>
- Jain, A. (2021, January 11). Assigning sentiment based on compound score: Vader or TextBlob. Analytics Vidhya. Retrieved from <https://www.analyticsvidhya.com/blog/2021/01/sentiment-analysis-vader-or-textblob/>
- Shah, N. (2019, March 6). Collocations: Identifying phrases that act like individual words in NLP. [Blog post]. Medium. Retrieved from <https://medium.com/@nicharuch/collocations-identifying-phrases-that-act-like-individual-words-in-nlp-f58a93a2f84a>
- Palash, D. (2020, June 11). Airlines on Twitter: Understanding customer complaints with NLP. [Blog post]. Analytics Vidhya. Retrieved from <https://medium.com/analytics-vidhya/airlines-on-twitter-understanding-customer-complaints-with-nlp-81278f2b68dc>