Topic Modeling on Disneyland Reviews

Aastha Tandon

atandon@horizon.csueastbay.edu

Technical Report

June 19, 2023

California State University East Bay

College of Business and Economics

Abstract

The research focuses on conducting a comprehensive analysis of customer reviews for Disneyland branches in California, Paris, and Hong Kong. It utilizes topic modeling to analyze online customer reviews of Disneyland branches and identify the factors that lead to positive and negative reviews. The study explores how customers' spoken topics vary across different branches over time, providing valuable insights into customer experiences and preferences. Disneyland authorities can make use of this information to enhance visitor satisfaction, improve services, make data-driven decisions, and eventually improve their business outcomes. Ultimately, this research aims to help Disneyland maintain its position as a leading brand in the global theme park industry.

Contents

1 Introduction

1.1 Background

Disneyland is one of the most renowned theme parks in the world that attracts millions of visitors to its various branches annually. Like any service industry, customer feedback can play an important role for Disneyland authorities. Analyzing customers' reviews can help them to identify areas that need their focus and attention. The purpose of this research is to analyze customer reviews of Disneyland to identify the main topics that contribute to positive and negative reviews.

1.2 Research questions

- What are the key topics brought up in positive and negative customer reviews of Disneyland branches in California, Paris, and Hong Kong and the factors associated to it ?

- How do these topics vary across different months in each branch ?

- How do these topics vary across different seasons in each branch ?

- How can this information be used to enhance the customer experience and address any issues affecting customer satisfaction ?

1.3 Importance of Research

The results of this research can have significant implications for Disneyland's operations and customer service. Identifying the positive topics that attract customers can help Disneyland's authorities to further improve their services, enhance the customer experience, and increase customer footfall. On the other hand, identifying the negative topics can help Disneyland's authorities address the issues, improve their services, and ultimately enhance customer satisfaction.

Furthermore, the analysis of customer reviews based on monthly and seasonal changes can provide valuable insights into how Disneyland's authorities can cater to their customers' needs during specific times of the year and improve their services accordingly. Overall, these research findings can help Disneyland's authorities to better understand their customers and maintain their position as a leading brand.

2 Literature Review

Many studies have been conducted focusing on different service areas of Disneyland. One study utilized semantic network analysis of online customer reviews to explore global theme hotels, specifically those related to Disneyland [Zhang et al.(2022)Zhang, Mengying, and Kim]. Another paper presented an online review-based methodology called Importance-Performance Analysis (IPA), which assessed service attributes and derived improvement strategies[Bi et al.(2019)Bi, Liu, Fan, and Zhang].Additionally, sentiment analysis of Disneyland reviews was analyzed using multiple approaches[Pam et al.(2022)Pam, Ong, and Rahman]. to implement effective business strategies and improvements, it is crucial to pay attention to the specific areas that customers mention in their feedback. Topic modeling is a valuable methodology in this regard, as it helps identify topics and extract important points from customer feedback. A comprehensive review of topic modeling was conducted, highlighting its significance[Kherwa and Bansal(2019)]. Moreover,

a study utilized topic modeling to analyze online reviews of Disneyland, providing insights into visitor behavior and experiences.[Luo et al.(2020)Luo, Vu, Li, and Law]. What sets our study apart is its comparative analysis across multiple branches, exploration of temporal variations, and a specific focus on identifying key topics within customer reviews. By pinpointing these key topics, our research offers a more comprehensive understanding of the factors that drive satisfaction and dissatisfaction. The comparative and temporal analysis provides valuable insights for Disneyland authorities to enhance services and customer experiences, catering to specific regional needs and improving overall visitor satisfaction.

3 Data

The dataset used in this research is from Kaggle and is available at the following link: (https://www.kaggle.com/datasets/arushchillar/disneyland-reviews). The dataset contains a total of 42,656 reviews of three Disneyland branches - Paris, California, and Hong Kong, posted by visitors on Trip Advisor. Each review is identified by a unique *review_id* and includes information about the reviewer's rating, data of visit, reviewer location, review text, and Disneyland branch location. The data was first cleaned by removing records which has missing information in *date_of_visit* column, resulting in 40,043 reviews. After removing missing values, the distribution of reviews across the three Disneyland branches are as follows: 18,202 reviews from California, 12,694 reviews from Paris, and 9,147 reviews from Hong Kong.

In order to facilitate the analysis, we extracted the Year and Month from *date_of_visit* column and stored them in separate columns. Additionally, we grouped the months into seasons and stored them in a new column called *season* which can take any of the following values - Fall, Winter, Spring, and Summer. Specifically December, January and February represent winter; March, April, and May represent spring; June, July, and August represent summer; and September, October, and November represent fall. We also then performed a series of data preprocessing steps which includes tokenization of sentences to words, stemming, removal of stop words, and removal of commonly used words to remove noise or less meaningful words.

4 Analysis

4.1 Exploratory Data Analysis

First, we performed an exploratory analysis to understand the underlying patterns in the data. We look at average rating trend over the years for each branch in Figure 1. California has maintained a consistently high rating over the years, but there has been a slight decrease in its rating trend. Hong Kong has demonstrated significant improvement in its rating over the years, with a noticeable upward trend. Paris exhibited some improvement in its rating from 2014 to 2017, but it ended up with a much lower rating in 2019. Next, we also look at average rating trend aggregated for each month as shown in Figure 2. We notice September is the highest-rated month for California, June is the lowest-rated month for Paris, and August is the lowest-rated month for Hong Kong. However, there could be various factors contributing to these shifts in ratings which can be understood better by delving deep into the data to analyze the reasons.
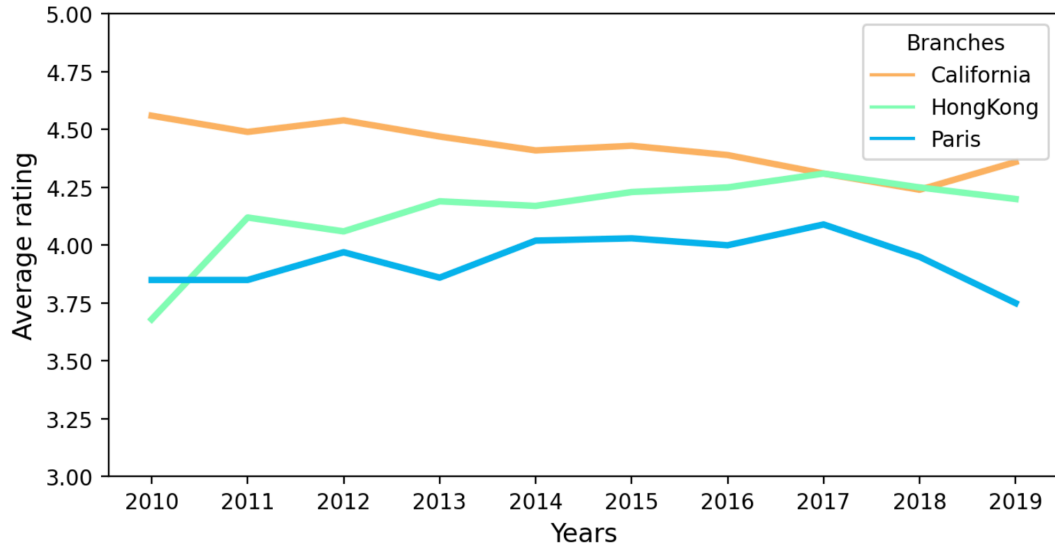
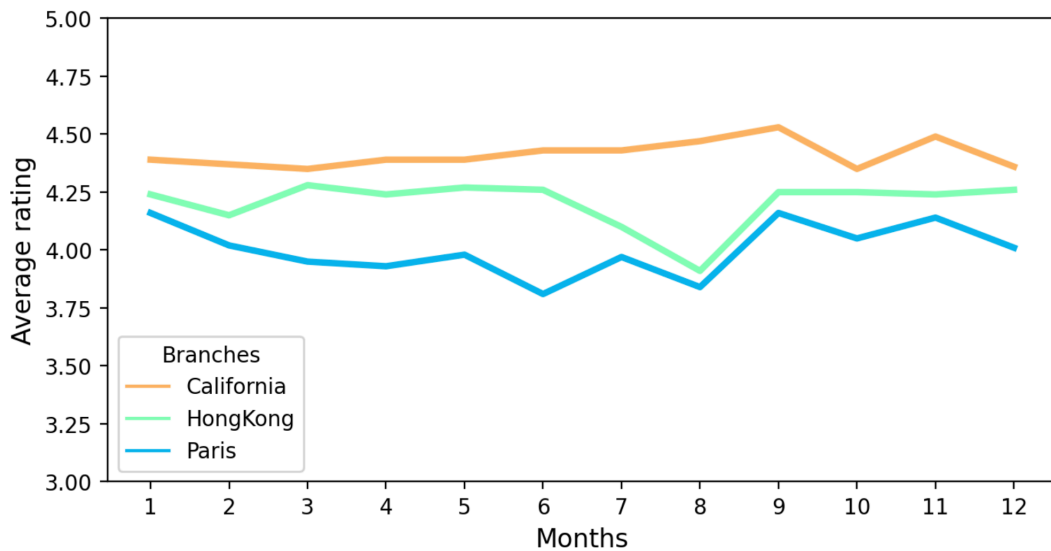Figure 1: Disneyland Average review rating over years

Figure 2: Disneyland Average review rating over months

Now we looked at total number of reviews aggregated over month for each branch. The intuition behind is we want to know which month has the highest footfall at each branch. Figure 3 shows that December is the busiest month for Disneyland Hong Kong, August for Disneyland Paris in Figure 4, and July for Disneyland California in Figure 5, as these months attract the highest number of visitors.

4.2 Topic Modeling for Positive and Negative Reviews

In order to perform topic modeling, we organized the data into three distinct groups based on the location for each branch separately. We classified the customer reviews into positive and negative buckets based on their rating scale of 1 to 5 for each branch. Ratings of 1 to 3 were classified in the negative bucket,
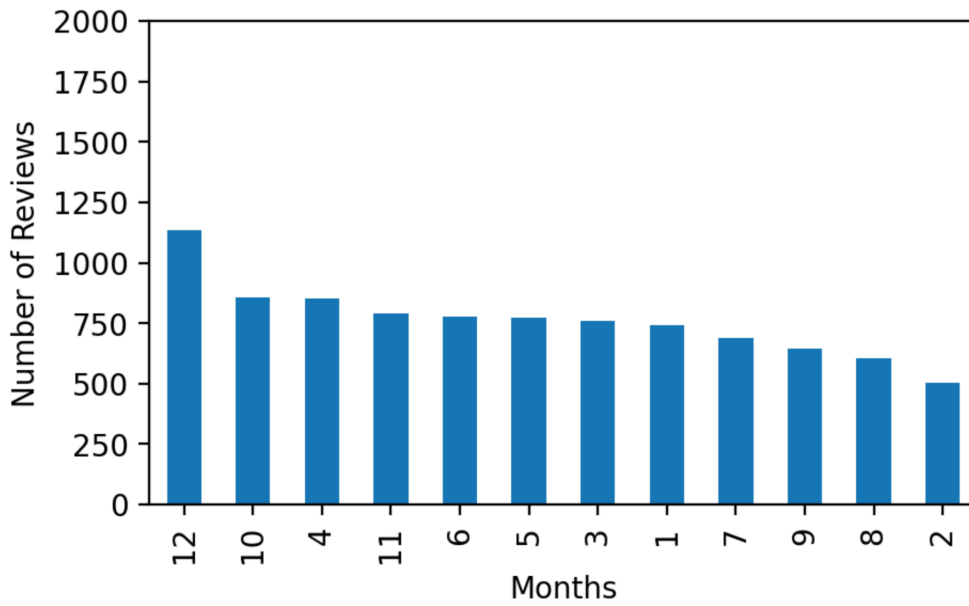
Figure 3: Peak time for visit in Disneyland Hong Kong. On x-axis we have month number.
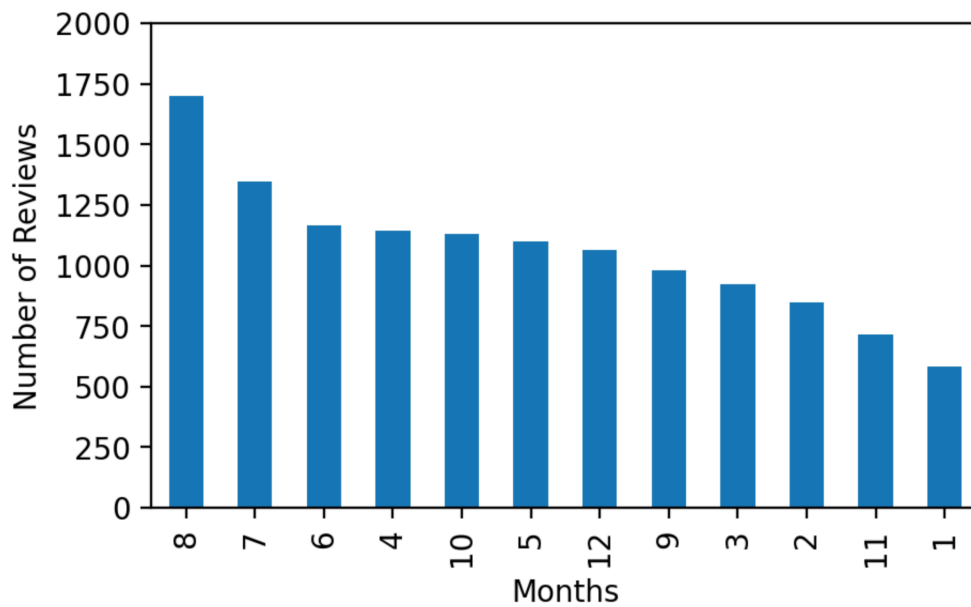

Figure 4: Peak time for visit in Disneyland Paris. On x-axis we have month number.

while 4 and 5 were considered in positive bucket. We then conducted LDA topic modeling on positive and negative buckets separately to identify the key positive and negative topics for each branch. We identified the number of topics by performing a grid search, where a higher coherence score gives a better model.
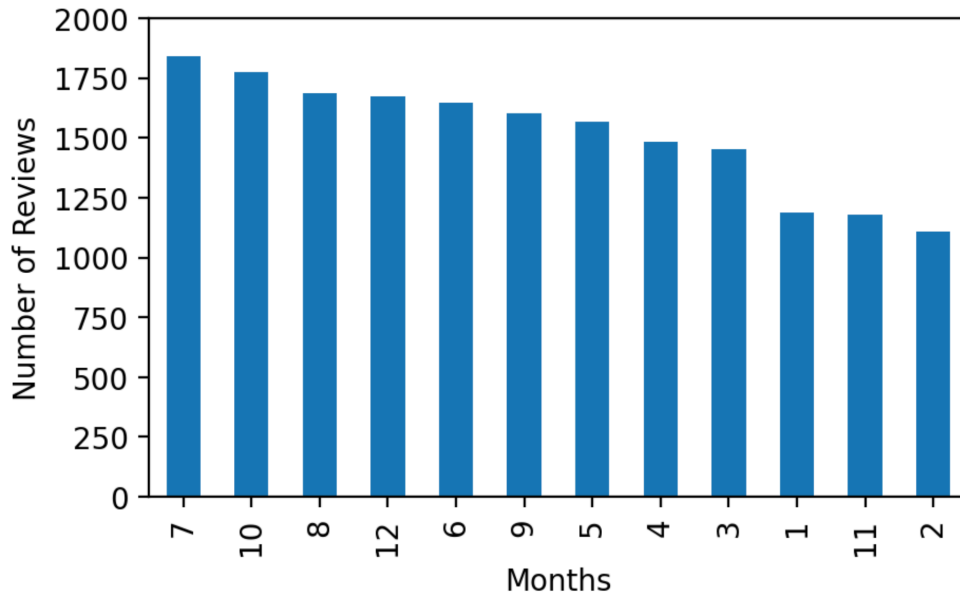
Figure 5: Peak time for visit in Disneyland California. On x-axis we have month number.

4.2.1 Disneyland California

We identified two positive topics and two negative topics for Disneyland California. Figure 6 describes the positive topics of Disneyland California as follows:

- Topic 0 - Theme Park Attractions and Shows Experience: This topic focuses on visitors' experiences with theme park attractions and shows. It may also include tips on avoiding long crowds and making the most of early hours.

- Topic 1 - Overall Theme Park Experience : This topic covers the overall experience of visiting the theme park, including food and dining, crowd levels, family-friendliness, and the general fun and love for the atmosphere of the park.

Figure 7 discusses the negative topics in Disneyland California mentioned as follows:

- Topic 0 - Disappointing Crowd Experience : The topic seems to be about the disappointment that the crowd experience brings. Words like "crowd", "close", and "long", "hour", "minute" suggests that the crowds are overwhelming causing long-hour delays, and the experience is not enjoyable.

- Topic 1 - Unsatisfactory Food and Price at Crowded Restaurant: This topic may be about unsatisfactory food and price along with the crowd spoiling the experience on trip.

4.2.2 Disneyland Paris

Figure 8 shows the positive topics of Disneyland Paris as follows :

- Topic 0 - Unforgettable Memories and Magical Experiences at Disneyland Paris: This topic could focus on positive reviews and experiences shared by visitors to Disneyland Paris. It explores the various aspects of the park that guests enjoyed, such as the rides, shows, hotel stays, and character interactions. The topic is mainly about visitors' great experience.

7

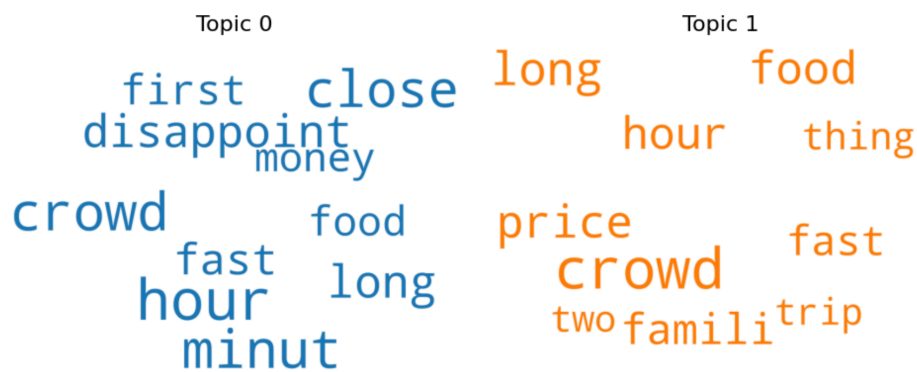Figure 6: Positive Topics in Disneyland California
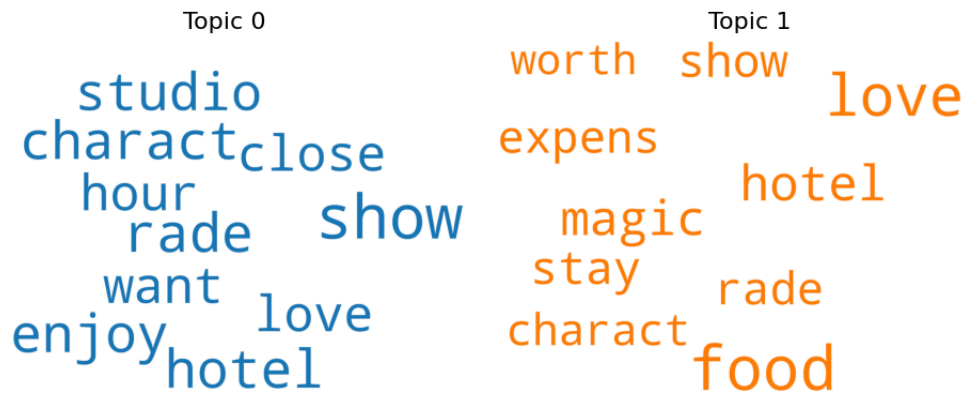

Figure 7: Negative Topics in Disneyland California

Figure 8: Positive Topics in Disneyland Paris

– Topic 1 - The Culinary Delights of Disneyland Paris: This topic seems to focus on the dining options available at Disneyland Paris, discussing the different types of food options available. It could also explore the cost of dining at Disneyland Paris and whether the experience is worth the expense. Additionally, it could cover any unique dining experiences available at Disneyland Paris, such as character dining or themed restaurants.
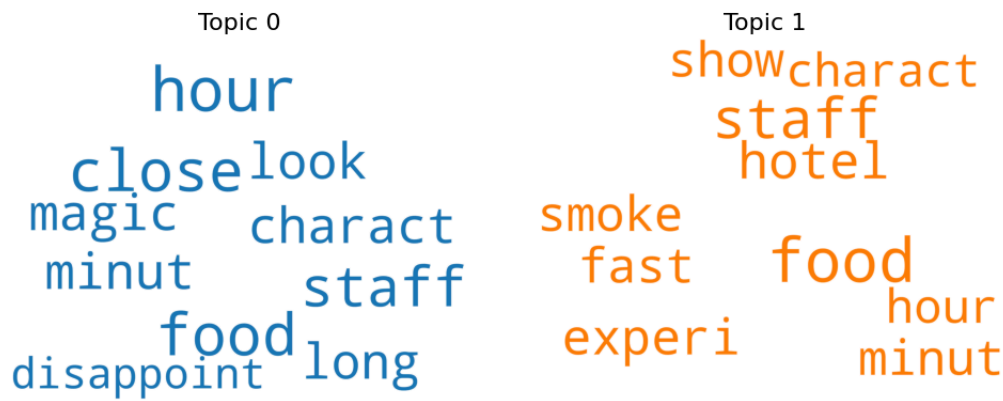
Figure 9: Negative Topics in Disneyland Paris

Figure 9 describes the negative topics of Disneyland Paris as follows :

– Topic 1 - Negative Reviews of Restaurant Food and Service: This topic is about negative reviews related to restaurant experiences. Customers might express their dissatisfaction with the food quality, staff being unhelpful or rude, long wait times, or feeling disappointed with their overall experience.

– Topic 2 - Negative Reviews of Hotel Accommodation: This topic is about negative reviews related to hotel accommodation experiences. Customers might express their dissatisfaction with the hotel's services and amenities such as a smoke-free environment. Additionally, guests may have negative feedback about their experiences with hotel shows and characters.
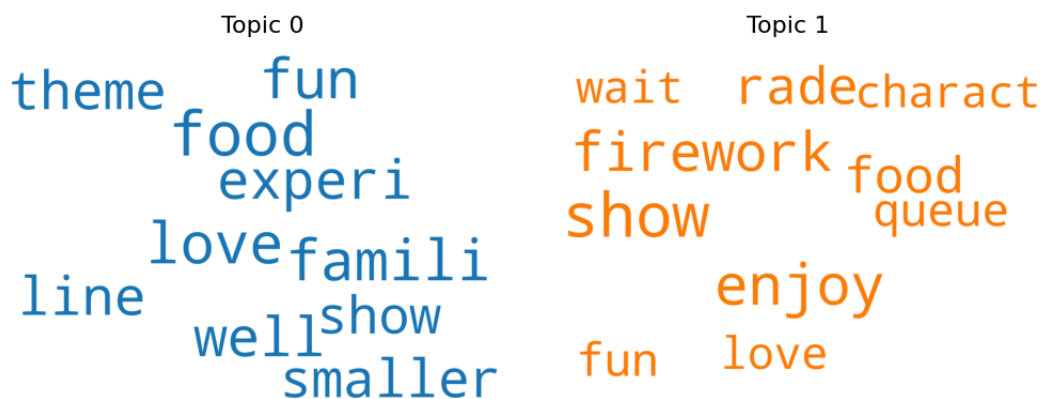
4.2.3 Disneyland Hong Kong

Figure 10: Positive Topics in Disneyland Hong Kong

Figure 10 shows the positive topics of Disneyland Hong Kong as follows :

– Topic 0 - Experience of Attending a Theme Park : This topic could explore the different aspects of visiting a theme park, such as the food, theme, shows, and overall experience. It could also delve
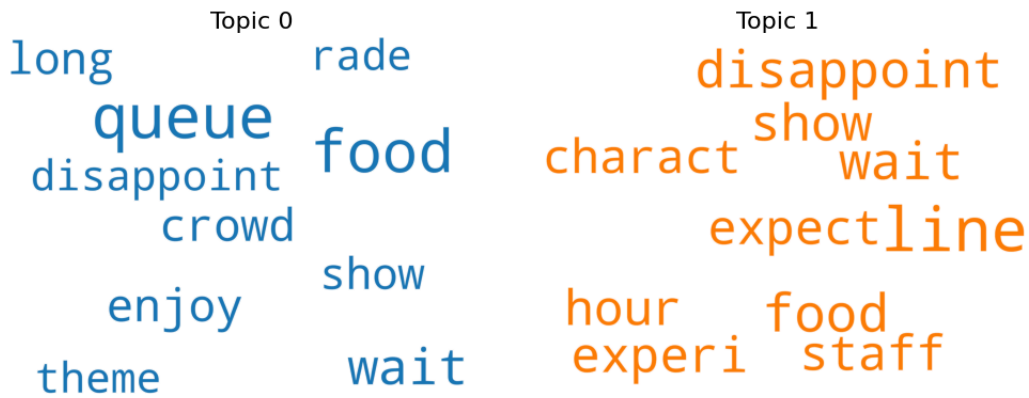
9

Figure 11: Negative Topics in Disneyland HongKong

into the excitement and fun of going to a theme park, and how it can create lasting memories with family and friends.

– Topic 1 - Enjoyment of Firework Shows: This topic could focus specifically on the experience of watching firework shows. The topic could also discuss the emotions and feelings associated with watching fireworks, such as joy, fun, and loving it all. Additionally, it could also be possible to discuss navigating the long queues to see the firework.

Figure 11 talks about the negative topics of Disneyland Hong Kong as follows :

– Topic 0 - Disappointing experience at a crowded theme park : The first negative topic explores the downsides of visiting popular theme parks, such as overcrowding, long wait times, and disappointment with attractions or food.

– Topic 1 - The frustrations and challenges of dining out at busy restaurants: This topic focuses on the challenges and frustrations of dining out at busy restaurants, such as dealing with long wait times, slow service, and food that fails to meet expectations.

4.3 Topic distribution across months

We also conducted topic modeling analysis across months for each branch. In this case, we did not classify the reviews as positive or negative. Instead, we learned *K* topics across all the reviews for a branch, and then computed the average topic distribution for each month. This enabled us to track the changes in topics across months and analyze the reasons for the changes.

4.3.1 Disneyland California : Topic distribution across months

We learned four topics across all the reviews for Disneyland California, and then computed the average topic distribution for each month to study the topic variation. Topics shown in Figure 12 are:

– Topic 0 - Crowd management : This topic can cover the management of the crowd at Disneyland California. The discussion can revolve around the strategies used by the management to control crowds, the peak hours, how the prices of the tickets affect the crowd, and the overall experience of the visitors.

Figure 12: Overall Topics in Disneyland California

– Topic 1 - Festive experiences : This topic can cover festive experiences, such as Halloween and Christmas celebrations. The discussion can be around different themes, the decorations, and the overall experience of celebrating these festivals at Disneyland California.

– Topic 2 - Fireworks and shows: The discussion can revolve around the different fireworks and shows, the timings, the character interaction, and the overall experience of watching these shows.

– Topic 3 - Fastpass enhancing Magical adventures : This topic can cover the different attractions, and experiences. The discussion can be around the most popular attractions, how to plan your visit using fastpass, and the overall experience of going on a magical adventure at Disneyland California.

We then computed the average topic distribution and can see from Figure 13 that Topic 0 is the most talked about and Topic 1 is the least talked about throughout the year.

4.3.2 Disneyland Paris : Topic distribution across months

Similarly like Disneyland California, we learned four topics across all the reviews for Disneyland Paris shown in Figure 14 as follows:

– Topic 0 - Character Experiences at Disneyland Paris: The topic discussed the impact of character interactions at rides, bookstores, shows, food restaurants, and hotels.
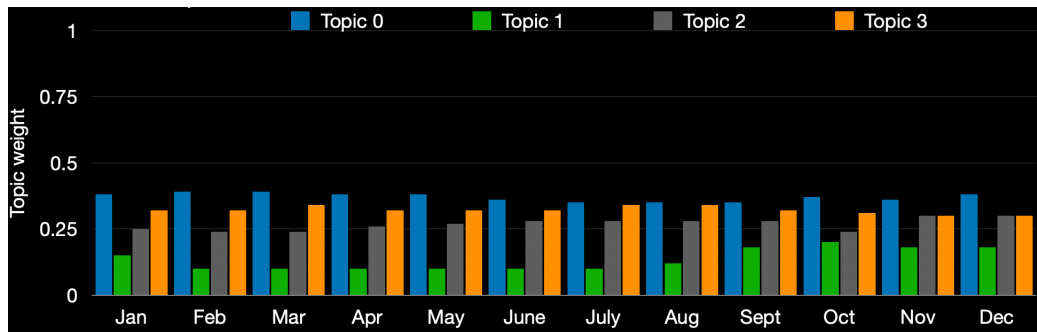
Figure 13: Mean topic distribution for each month in Disneyland California

– Topic 1 - Fast Pass for great overall experiences: This topic discussed different attractions at Disneyland Paris and how to maximize the experience by using the Fast Pass system.

– Topic 2 - Hotel Accommodations and Dining : This topic may be about the various hotels and restaurants available at Disneyland Paris, including options for different budgets and tastes.

– Topic 3 - Tips and Tricks for a Magical Visit : This topic seems to be about insider tips and advice on how to make the most of your visit to Disneyland Paris.

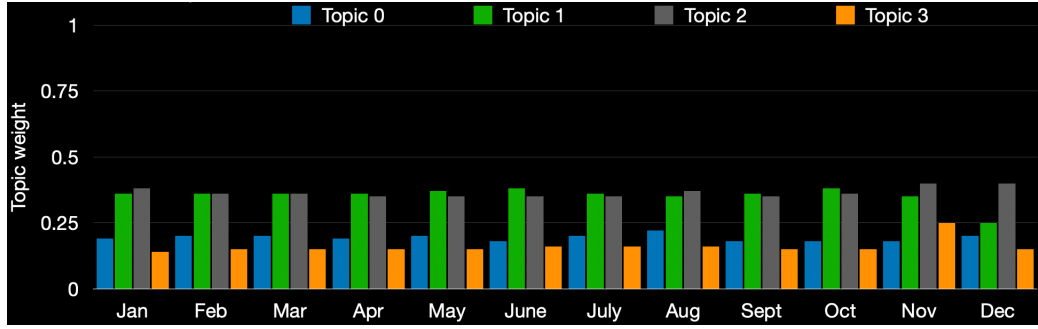Figure 14: Overall topic in Disneyland Paris

Figure 15: Mean topic distribution for each month in Disneyland Paris

After computing the average topic distribution we can observe from Figure 15, Topic 3 is the least discussed throughout. Topic 2 is the most discussed in January, August, November, and December. Topic 1 is most discussed in April, May, June July, September, and October.Topic 1 and 2 are equally discussed in Feb and March.

4.3.3 Disneyland Hong Kong : Topic distribution across months

As shown in Figure 16, we learned four topics across all the reviews for Disneyland Hong Kong as follows:

- Topic 0 - Theme park experiences : This topic seems to be on the general experience of visiting a theme park, including waiting in lines and watching fireworks, enjoying shows and attractions.

- Topic 1 - Food and dining: This topic can cover the food and dining options available at theme parks, including the quality of the food, the variety of options, the cost, and the overall dining experience.

- Topic 2 - Entertainment: This topic can explore the different forms of entertainment available at theme parks, such as shows, fireworks, and food as well as the characters and mascots that visitors can interact with.

- Topic 3 - Queue management : This topic can focus on the management of queues at theme parks, including how long visitors have to wait, the efficiency of the queue system, and ways to improve the overall queue experience.

From the average topic distribution from Figure 17, we can see that Topic 2 is the most discussed and Topic 3 is the least discussed across the year.

4.4 Topic distribution across seasons

The approach is similar to the topic modeling analysis across months. We already learned K topics across all the reviews for a branch while doing month-wise analysis. Here, we would compute the average topic distribution across seasons for those topics itself. This helped us to observe changes in topics across seasons and to analyze the reasons for the changes.
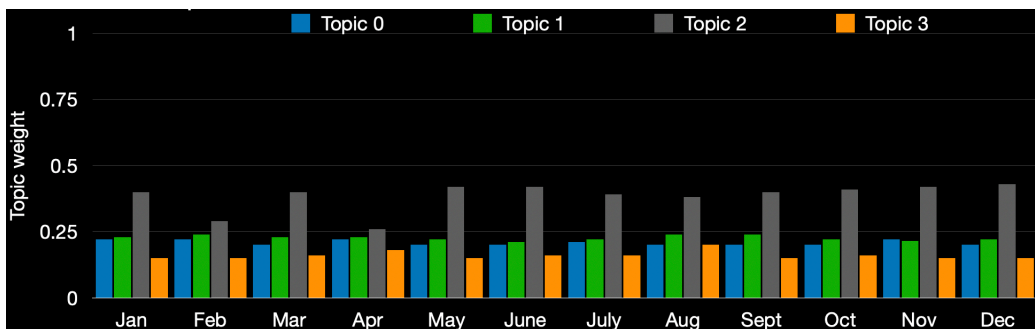
13

Figure 16: Overall topic in Disneyland Hong Kong

Figure 17: Mean topic distribution for each month in Disneyland Hong Kong

4.4.1 Disneyland California: Topic distribution across seasons

After computing the average topic distribution across seasons, we can see from Figure 18 that Topic 0 is the most talked and Topic 1 is the least talked across all the seasons. The importance of effective crowd management at Disneyland California cannot be underestimated as it affects every visitor to the park, with millions of guests each year. The festive experiences such as Halloween and Christmas celebrations, while enjoyed by many, may not be universally appealing and could potentially bring additional crowds, causing visitors to avoid the park during those times.
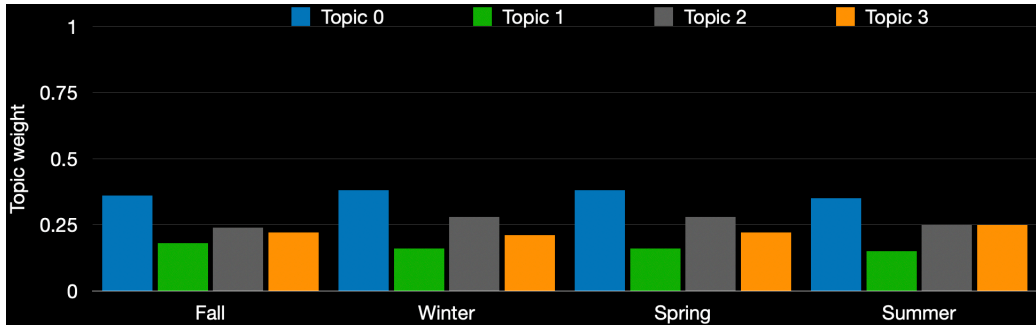
14

Figure 18: Mean topic distribution for each season in Disneyland California

4.4.2 Disneyland Paris: Topic distribution across seasons

From Figure 19, it can be seen that Topic 1 is slightly more discussed in Summer and Spring than Topic 2 and equally in Fall. Topic 2 is most discussed in Winter and Topic 3 is the least discussed. Topic 1 is slighlty more discussed in Summer, Spring than Topic 2 may be because crowd and long lines are major concern for themepark and fastpass can enhance the overall experience. Topic 2 is most discussed during winter as visitors are more focused on indoor experiences, such as dining and staying in hotels. Fall has an equal discussion on both topics due to moderate crowds and a balanced interest in both experiences.
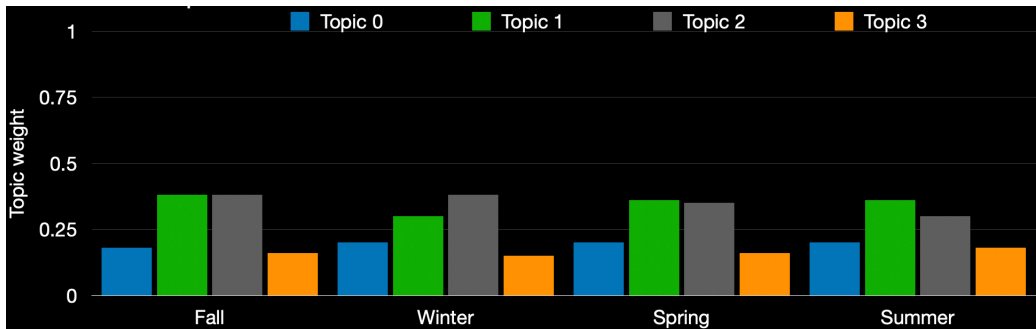

Figure 19: Mean topic distribution for each season in Disneyland Paris

4.4.3 Disneyland Hong Kong: Topic distribution across seasons

From Figure 20, it can ve seen that Topic 2 is most discussed and topic 3 is least discussed across the seasons.The reason why Topic 2 (entertainment) is most discussed and Topic 3 (queue management) is least discussed in Disneyland Hong Kong could be due to the importance, experience, perception, and marketing efforts. Visitors may be having such a great time enjoying the shows, parades, and attractions that they do not focus on discussing the queues. In this scenario, it is not necessarily that the queue management is quite good, but it may happen that the visitors are having such a positive experience with the entertainment options that it takes precedence in their discussions and memories of the park.
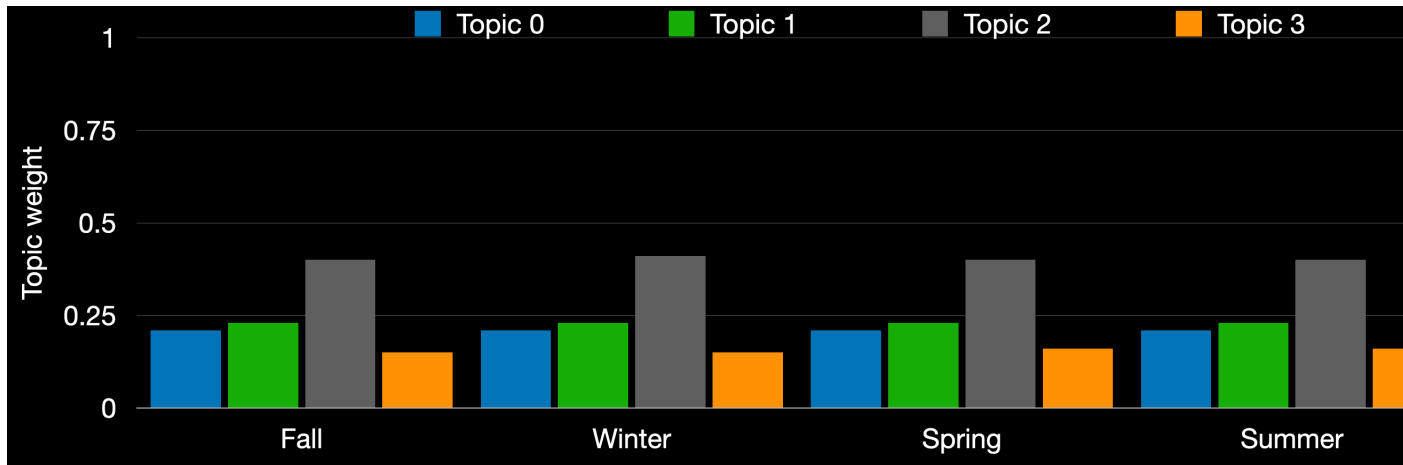
Figure 20: Mean topic distribution for each season in Disneyland Hong Kong

5   Summary

5.1   Impact of Results

5.1.1   Managerial Implications for Disneyland California

To improve visitors' experience and manage crowds during peak hours, Disneyland California should focus on implementing effective crowd management strategies, such as limiting visitors, improving ride queues, and increasing staff. It should also expand its festive experiences and offer unique events throughout the year, promoting them to attract more visitors. Enhancing the fireworks and parade shows, adding new attractions and experiences can help keep visitors engaged and returning. Disneyland could also consider implementing digital tools to help visitors plan their visit and make the most of their magical adventure. Improving the quality and variety of restaurant food by offering more choices, and discounts, and addressing visitor complaints can further enhance the visitors' experience.

5.1.2   Managerial Implications for Disneyland Paris

Disneyland Paris could improve food quality, staff training, reduce wait times, invest in hotel cleanliness and listen to guest feedback. They could introduce healthier food, a fast pass system, regular deep cleaning, and update amenities. Promotions and events could also be tailored to each season's interests, such as offering more Fast Pass options in summer and promoting unique hotel and restaurant features during winter. Guest feedback can be encouraged through surveys and social media monitoring to ensure visitors have the best experience possible.

5.1.3   Managerial Implications for Disneyland Hong Kong

To address the negative experiences associated with overcrowding, long wait times, and dining frustrations at Disneyland Hong Kong, several strategies could be implemented. These include managing visitor capacity, expanding attraction and dining options, improving communication, increasing staffing levels, using technology, and enhancing menu offerings. Furthermore, the park's popular entertainment options, such as shows, parades, fireworks, and character interactions, have generated significant buzz

16

among visitors. Disneyland Hong Kong could continue to focus on and expand these entertainment offerings to provide the best possible experience for its visitors.

5.2 Limitations

The research has several limitations that should be taken into consideration when interpreting the results. First, the analysis was limited to reviews posted on TripAdvisor, which may not represent the views of all visitors. Second, it only analyzed reviews in English, which may not be representative of the views of visitors who speak other languages. Finally, this research did not consider demographic variables, such as age and gender which may affect visitors' experiences and perception.

5.3 Future Research Direction

Future research could extend this study in several ways. First, the analysis could be expanded to include reviews from other platforms and in different languages. This would provide a more comprehensive view of visitors' experiences and perceptions. Second, future research could consider demographic variables to better understand the factors that influence visitors' experiences. Finally, future research could use more advanced natural language processing techniques to extract deeper insights from the reviews, such as sentiment analysis and emotion detection.

References

[Bi et al.(2019)Bi, Liu, Fan, and Zhang] Jian-Wu Bi, Yang Liu, Zhi-Ping Fan, and Jin Zhang. Wisdom of crowds: Conducting importance-performance analysis (ipa) through online reviews. *Tourism Management*, 70:460–478, 2019.

[Kherwa and Bansal(2019)] Pooja Kherwa and Poonam Bansal. Topic modeling: a comprehensive review. *EAI Endorsed transactions on scalable information systems*, 7(24), 2019.

[Luo et al.(2020)Luo, Vu, Li, and Law] Jian Ming Luo, Huy Quan Vu, Gang Li, and Rob Law. Topic modelling for theme park online reviews: Analysis of disneyland. *Journal of Travel & Tourism Marketing*, 37(2):272–285, 2020.

[Pam et al.(2022)Pam, Ong, and Rahman] Chun Onn Pam, Wen Eng Ong, and Amirah Rahman. Multiple approaches in sentiment analysis on disneyland reviews. *Open International Journal of Informatics*, 10(1):21–36, 2022.

[Zhang et al.(2022)Zhang, Mengying, and Kim] Xiaobing Zhang, T Mengying, and H Kim. Exploration of global theme hotels through semantic network analysis of online customer reviews: Focused on disneyland hotels. *J. Ind. Innov*, 38:215–241, 2022.