UNLOCKING THE MAGIC OR REVEALING THE MESS?

A Sentiment and Topic Analysis of Disneyland Customer Reviews

BAN200 – Text Mining and Sentiment Analysis

Group Project Proposal

Seneca Polytechnic

Mayra Geraldine Reinoso Varon,

Naveen Karan Krishna &

Thi Anh Tram Le

Table of Contents

Dataset Description and Source

We will use the publicly available dataset "Disneyland Reviews" from Kaggle (https://www.kaggle.com/datasets/arushchillar/disneyland-reviews). It includes over 40,000 guest reviews from Disneyland locations in California, Paris, and Hong Kong. Each review entry contains a unique ID, star rating (1–5), visit date, visitor location, review text, and park branch. This dataset is ideal for performing text mining and sentiment analysis to uncover patterns in customer satisfaction across different global parks. The data is under the CC0 1.0 Public Domain Dedication, where use, distribution, and adaptation are free, including for commercial purposes, without permission.

Business Problem

How can Disneyland use guest review data to improve customer satisfaction and enhance the park experience across global locations?

- Identify top drivers of positive vs. negative reviews and compare these themes across parks.
- Assess alignment between star ratings and sentiment in text
- Deliver data-driven recommendations to improve service, operations, and design

This project was chosen because Disneyland's brand relies heavily on guest experience. While star ratings provide a basic summary, open-ended reviews contain rich, underused insights that can help Disney pinpoint what guests truly value or dislike such as long wait times, staff friendliness, or food quality. By analyzing these reviews, we aim to discover both universal issues and location-specific trends that can be turned into actionable improvements.

Methodology

We analyzed the Disneyland Reviews dataset. We loaded the data, and we cleaned it by removing duplicate entries and reviews with missing visit dates. We also performed feature engineering by dividing the visit date into the year column and the month column to enable analysis by time.

Our exploratory data analysis (EDA) focused on rating distributions, review volume trends, review length, and reviewer locations. To extract sentiment, we applied standard text pre-processing (lowercase, tokenization, stop word removal, lemmatization), then the VADER sentiment analyser to classify reviews as positive, neutral, or negative. We also explored the alignment of star ratings and sentiment scores.

To identify top drivers of satisfaction and dissatisfaction, we employed TF-IDF to extract high-weight sentiment terms and Latent Dirichlet Allocation (LDA) topic modeling on bigrams to recognize recurring themes. The topics were manually labeled and tracked over time and by park location to compare guest priority across Disneyland branches.

Results and Evaluation

Our exploratory data analysis (EDA) revealed big trends in guest review patterns among the three Disneyland parks. On average, nearly 75% of reviews were 4- or 5-star ratings, which indicated overall guest satisfaction. Although 1-star reviews were the least frequent, they often presented strongly worded complaints and highlighted recurring issues that could significantly impact public perception. In terms of park-specific activity, Disneyland California dominated review volume, then Paris and Hong Kong. This trend repeated annually, most likely due to increased visitation, broader web-based engagement, or greater English-language accessibility of reviews for the California park.

Temporally, we experienced a review peak between the years 2015 and 2018, dropping sharply in 2020, most likely due to COVID-19-related closures. Seasonal peaks during holiday periods further verified the influence of high-season visitor seasons. Lastly, most reviews are short, with a median of 80 words. The mean is slightly higher (126 words) as some very long reviews pull it up. Most people write under 150 words, but some give much longer, detailed feedback, especially useful for more detailed analysis. These findings informed the rest of the analysis of guest sentiment and topics.

To achieve our initial objective, identifying top drivers of positive and negative reviews and comparing these themes across parks, we applied topic modeling (LDA) and TF-IDF to positive and negative review sentences. For positive reviews, four ongoing themes common across locations were: thrill rides and FastPass efficiency, magical atmosphere and cast interactions, efficient park design and traffic patterns, and family fun with character interaction. California and Paris visitors especially praised classic thrill attractions like Space Mountain and ease of movement in the parks with FastPass.

Conversely, Hong Kong visitors focused on family programming and high levels of interaction with characters and cast members, which aligns with local preferences. These findings highlight the importance of aligning park experiences with visitors' expectations: Western parks are characterized by immersive storytelling and operational ease, but emotional and character-based experiences play a greater part in Hong Kong.

On the other hand, negative reviews revealed certain operational challenges in parks. In Disneyland California, guest experience- and employee-related problems were the most common source of dissatisfaction, with a sharp increase in the latter part of 2015 and ride closures and technical glitches from 2014 to 2016. In Paris, service quality and general guest experience were the main drivers of negative sentiments. In Hong Kong, the dislikes were milder and constant throughout the period, with some occasional grievances being voiced here and there regarding ride closures and guest experience, which built to a modest crescendo in 2016.

These results indicate that while some of the dislikes, like long queues, are universal for all parks, others are intimately tied to local culture. This highlights the importance of implementing region-specific service improvements rather than applying uniform strategies across all locations

Achieving our second objective, assessing alignment between star ratings and textual sentiment, we compared the VADER sentiment scores of each review with its star rating. We saw a strong positive correlation with average sentiment scores of 0.81 for 5-star reviews, 0.73 for 4-star, 0.43 for 3-star, 0.15 for 2-star, and -0.09 for 1-star reviews. This trend confirms that higher star ratings are highly associated with more positive guest sentiment in review text. One interesting observation is to compare 4- and 5-star reviews: although 5-star reviews are nearly twice as common, their average sentiment (0.81) is only slightly above that of 4-star reviews (0.73).

This suggests that certain visitors are likely to record the visit as perfect despite some issues, whereas others with similar good experiences withhold one star, indicating differential expectations and rating expectations.

To further support our analysis findings, we generated a Wordcloud visualization from the cleaned review texts. This visual tool highlights the most frequently mentioned terms across all reviews, with larger words indicating higher frequency. Positive words like “magic”, “fun”, “ride”, “great”, and “amazing” suggest that many visitors describe their Disneyland experience in terms of excitement, emotional satisfaction and enjoyment. In addition, the presence of terms like “wait”, “line”, “hours”, “time”, and “crowded” show frequent mentions of queue times and park congestion. This supports our earlier insights of long lines and inefficient processes as major pain points.

Finally, our third objective was focused on providing data-driven suggestions on ways to improve service quality, operational designs, and guest satisfaction. To reduce frustration with ride closures, we recommend adding in-park real-time status updates on rides through the park app and enhanced on-park communication via signage or proactive cast member messaging. FastPass systems, though often appreciated, also generated frustration and concerns about equity. Improving access, clarifying usage instructions, and having equitable systems like tiered windows or internet lotteries during peak demand times might reduce unpleasant feelings.

Cast members consistently played a crucial role in the negative and positive experiences. Paris and Hong Kong could maximize service consistency and guest involvement by investing in supporting staff training and offering multilingual support. For long lines, enhancing line spacing with interactive elements, shade, and live wait-time data could deeply improve guest perception. In addition, tailoring park operations to local expectations is crucial: Hong Kong tourists value family and personality interactions, Paris tourists are primarily interested in thrill rides and easy access to rides, and California tourists crave immersion experiences and efficient park logistics. We also recommend ongoing examination of 3-star reviews, which are likely to have nuance or early-stage commentaries that can signal upcoming issues before they arise as serious problems.

In summary, Disneyland's global parks offer overall good experiences, but each location presents distinct themes that affect guest satisfaction. By addressing operational pain points through local solutions, enhancing emotional connection, service excellence, and communication, Disney can further improve its world-class guest experience.

Discussion of Challenges, Limitations, and Potential Improvements

Although this analysis yielded useful findings, several limitations need to be considered. Firstly, the sample set only contains reviews in the English language and could, therefore, underrepresent the opinions of non-English-speaking visitors, especially those in the Paris and Hong Kong parks, thus introducing a linguistic bias. Secondly, there is not always an obvious correlation between sentiment scores and star ratings, with some visitors giving high marks while being dissatisfied in the text. Third, qualitative judgment of LDA-derived topics is involved in topic interpretation, which may overlook low-frequency but significant issues. Finally, we were not exposed to internal operations data, e.g., ride downtimes, staff schedules, or maintenance logs, which would have enabled us to explore causal relationships between guest complaints and actual park conditions.

In future work, these limitations can be overcome in the following ways. Including multilingual reviews would increase representation and cultural relevance in the analysis. Cross-matching review data with internal event logs or business metrics can produce more insightful explanations of sentiment shifts. More advanced uses of natural language processing models, capable of detecting sarcasm, affective subtlety, or context, would enhance the accuracy of sentiment. Finally, breaking reviews by visit purpose (i.e., families, solo visitors, new visitors) could provide more personalized data and help support improvement in service and experience in a targeted manner.

Conclusion

This project demonstrates how sentiment analysis and text mining can provide deep insights from Disneyland guest feedback at global park locations. Overall rating and sentiment were positive, but we discovered some clear regional differences in guest expectations versus experience. California visitors emphasized immersive stories and operational efficiency, Paris visitors preferred thrill rides, and Hong Kong visitors enjoyed character interaction and family-friendly amenities. Major pain points, like extended lines, uneven service, and closure of rides, differed by park but also offered actionable areas for improvement.

While some of these limitations are still held, such as language representation and lack of internal operational data, our results lay down a firm foundation for data-driven improvement. With park experiences that are better attuned to local guest expectations, enhanced communication, and increased quality in services, Disneyland can continue to reinforce its global brand and deliver excellent, culturally appropriate experiences. Further research building on this work can include multilingual reviews, operational data, and advanced sentiment methods to further enhance decision-making.

Reference

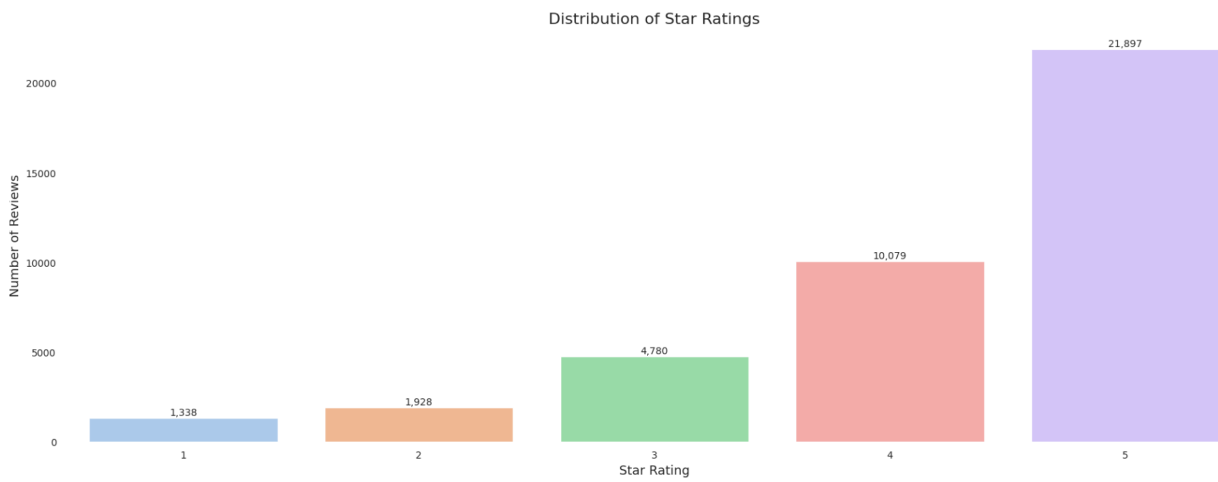
Chillar, A., (2020). *Disneyland Reviews* [Dataset]. Kaggle. https://www.kaggle.com/datasets/arushchillar/disneyland-reviews

Appendices

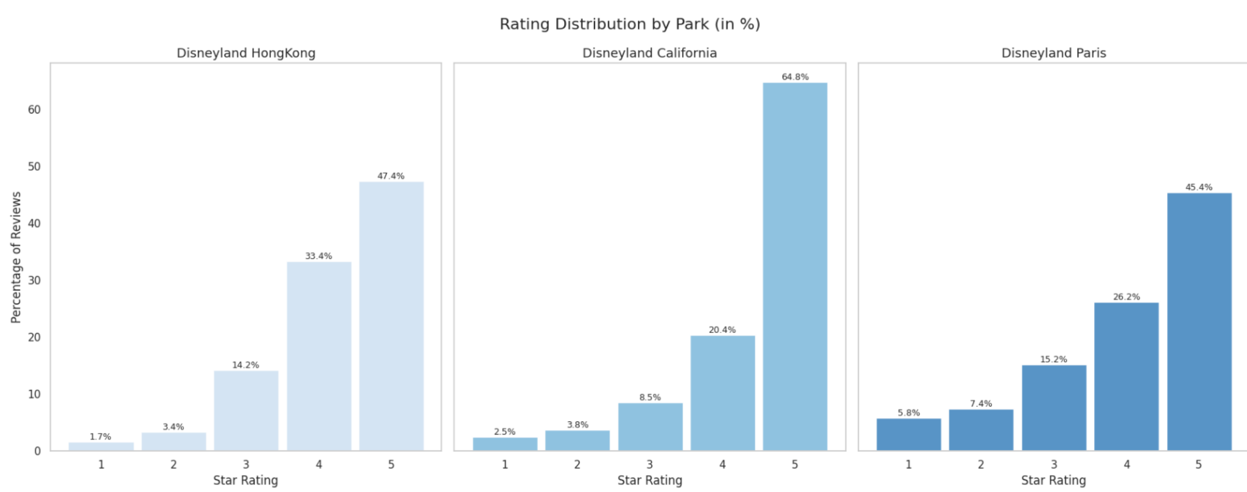
Appendix A: Project Repository

The full code, data processing steps, and visualizations used in this analysis can be accessed via the project's GitHub repository: https://github.com/NK-Mikey/Sentiment-Analysis

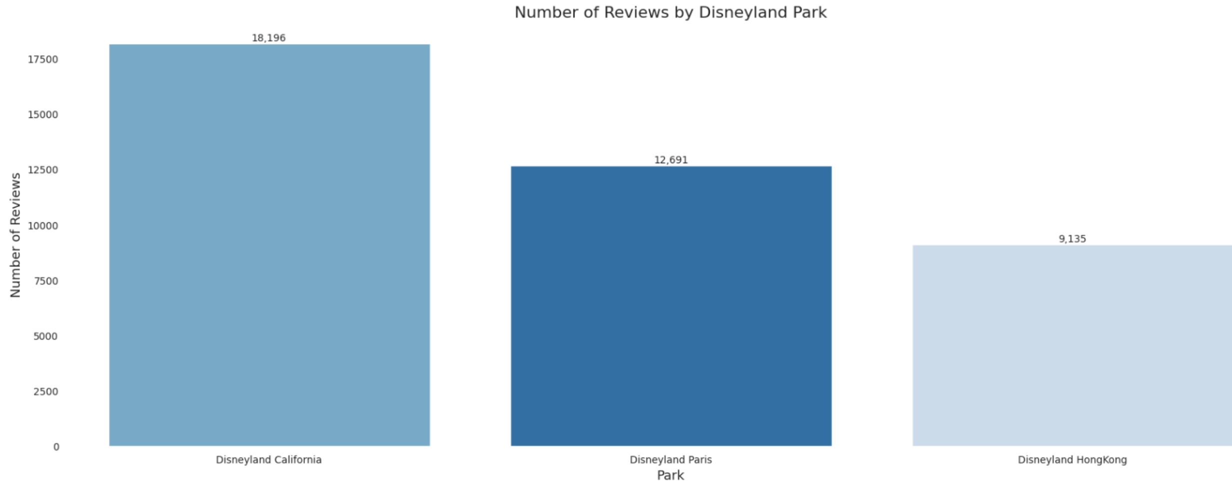
Appendix B



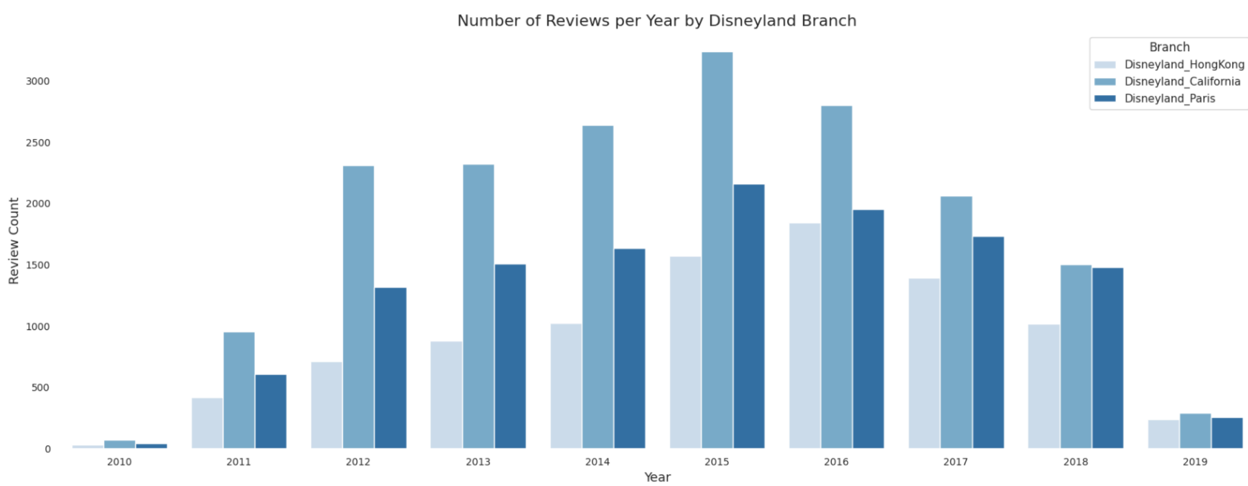
Appendix C



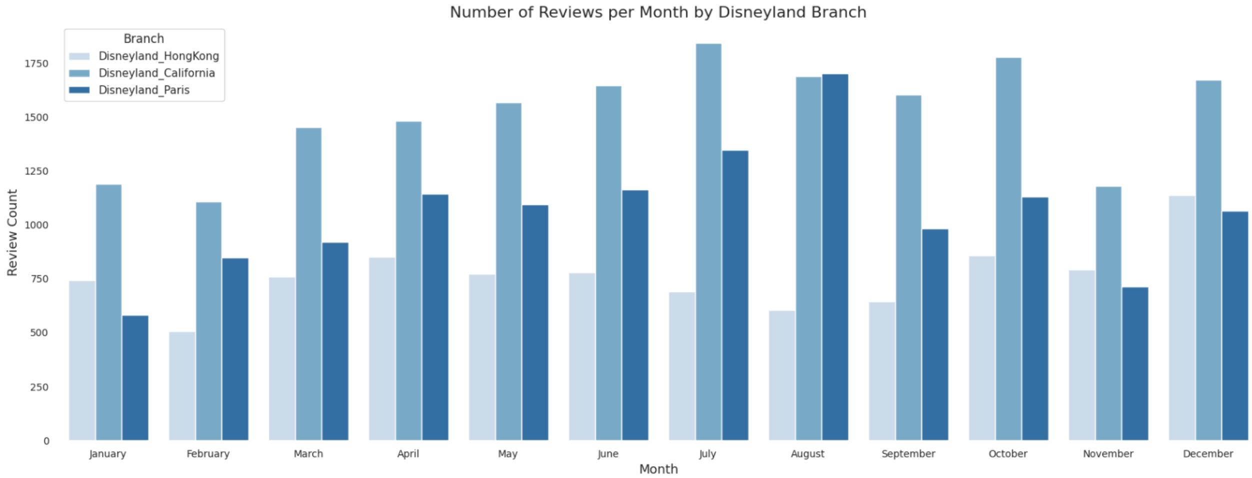
Appendix D



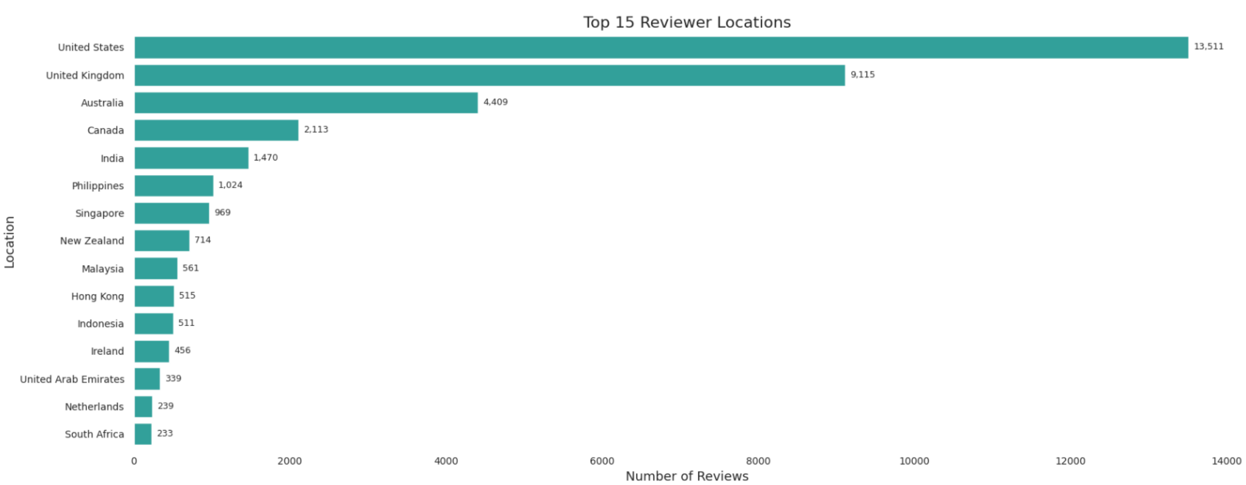
Appendix E



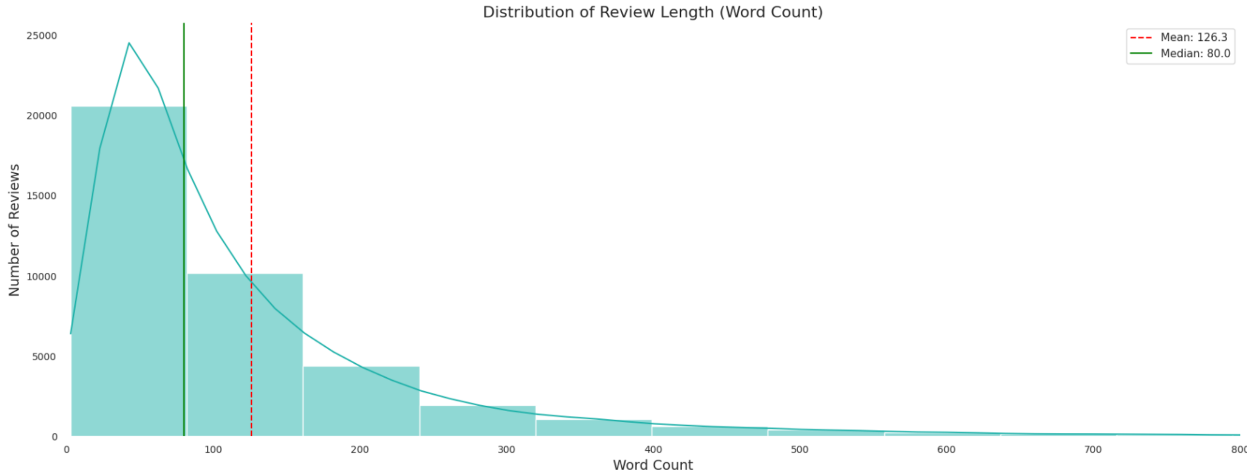
Appendix F



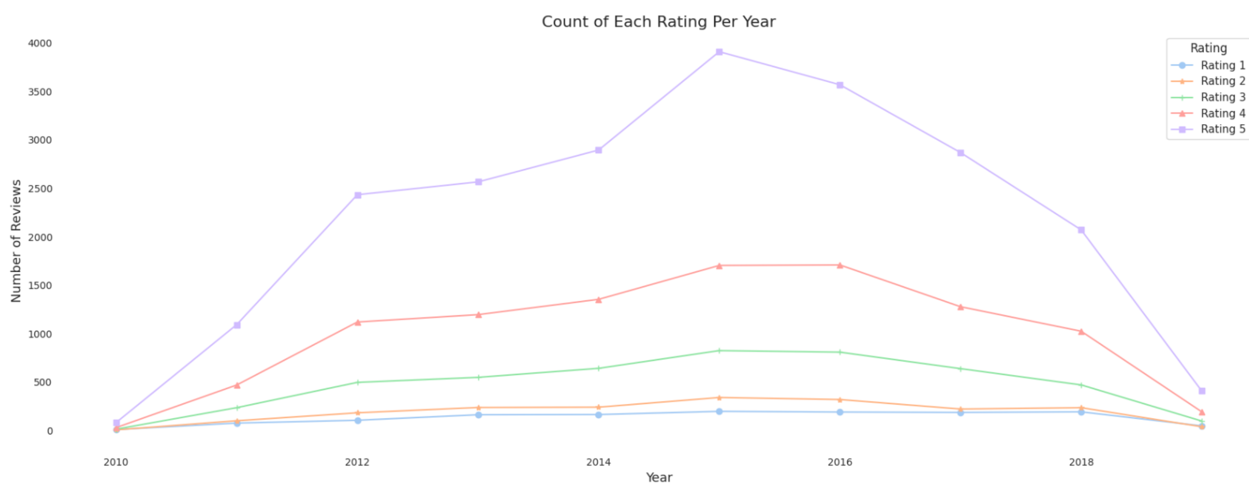
Appendix G



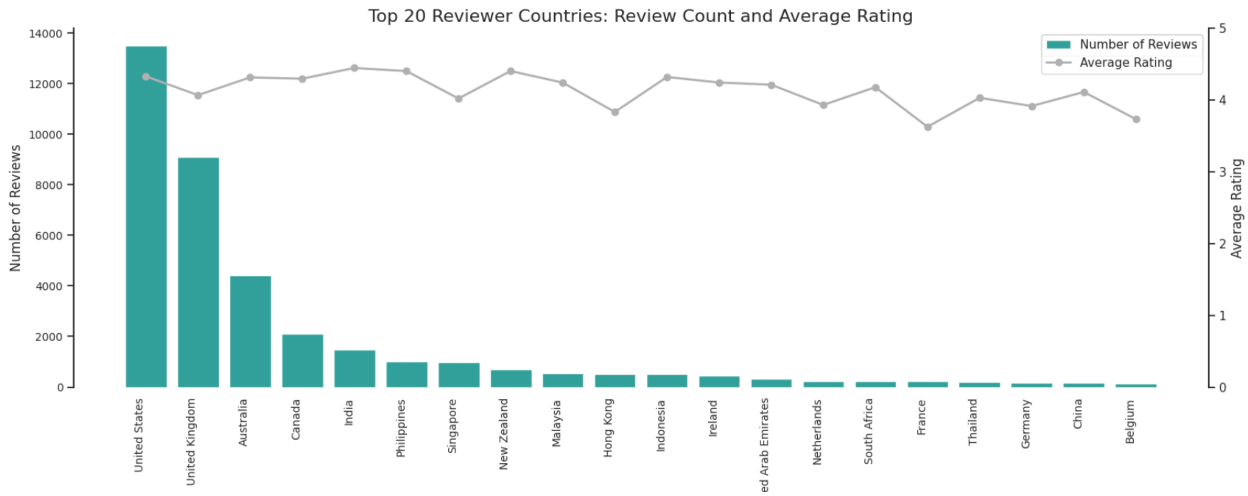
Appendix H



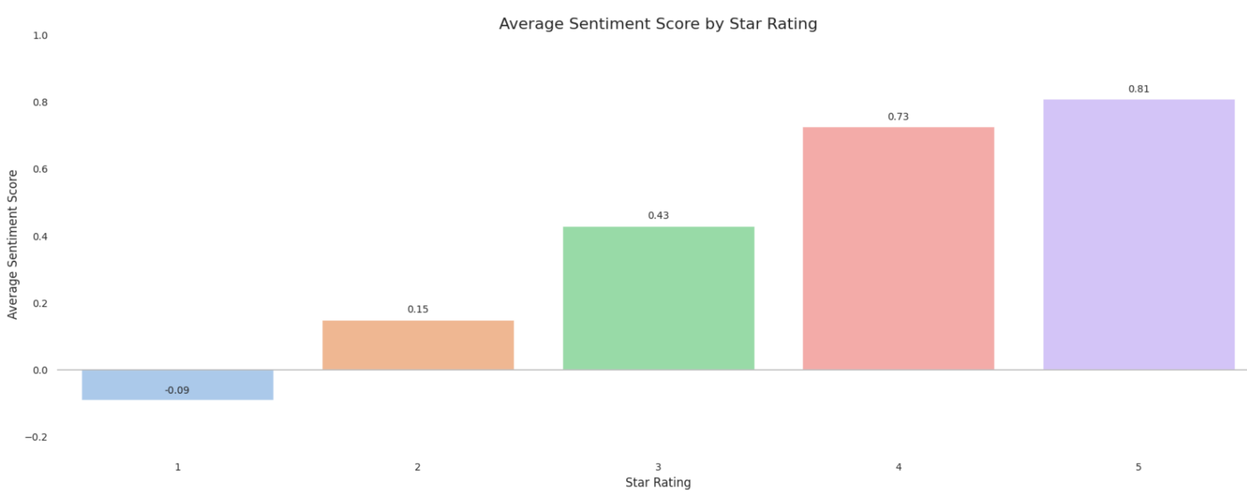
Appendix I



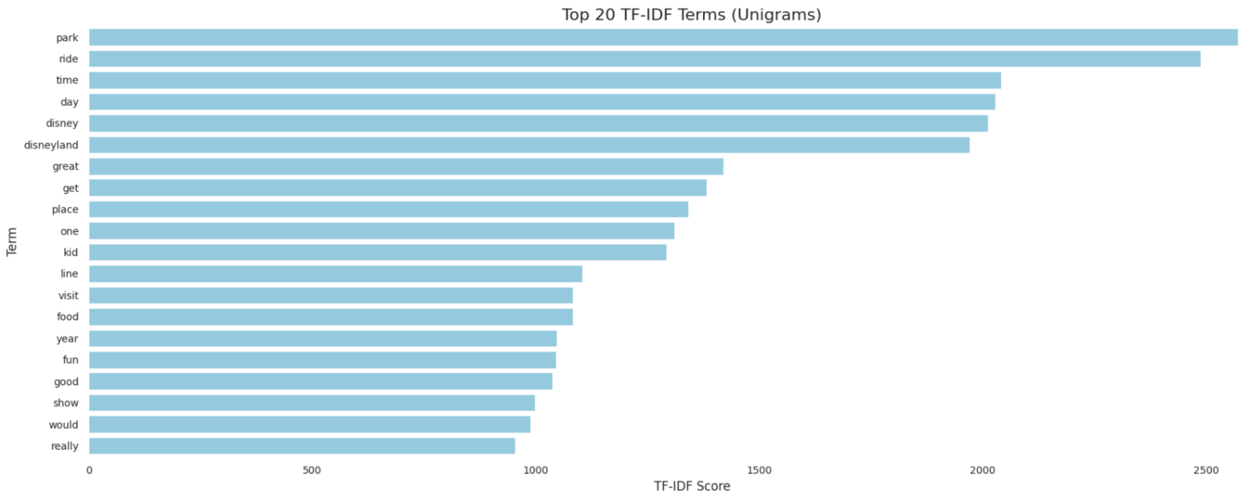
Appendix J



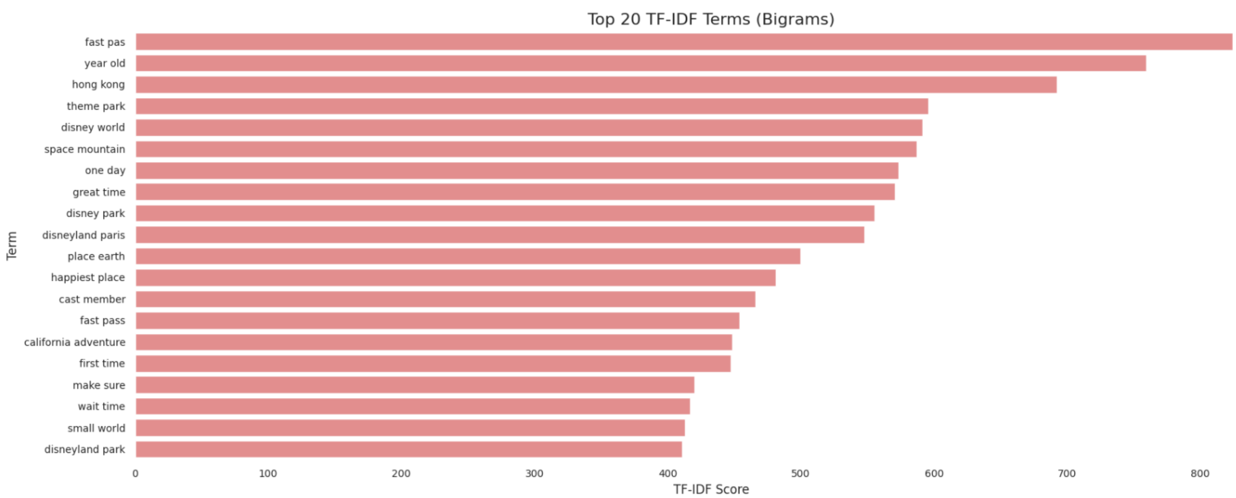
Appendix K



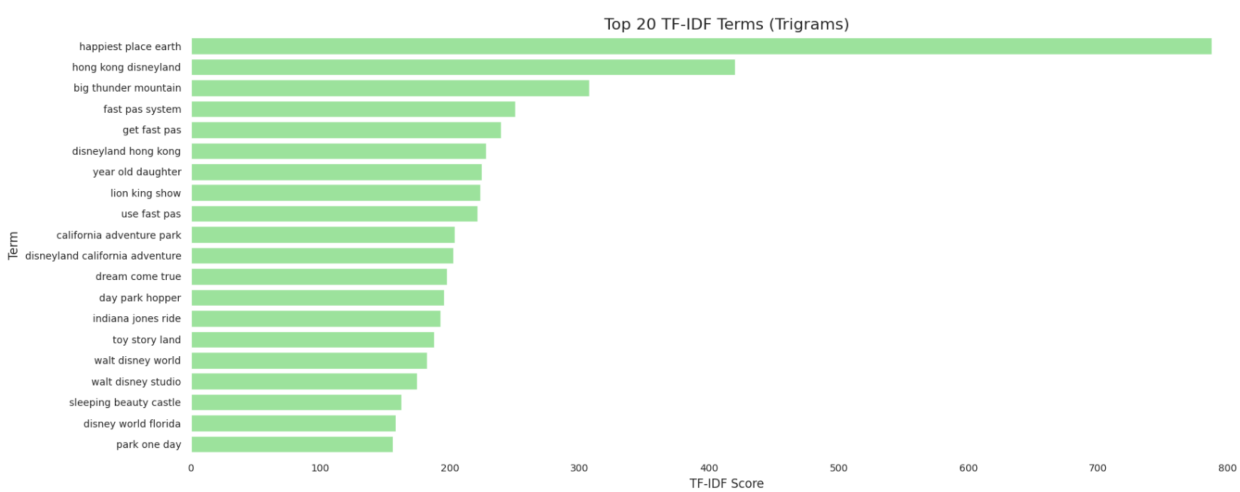
Appendix L



Top 20 TF-IDF Terms (Unigrams)
Term
park
ride
time
day
disney
disneyland
great
get
place
one
kid
line
visit
food
year
fun
good
show
would
really
0
500
1000
1500
2000
2500
TF-IDF Score

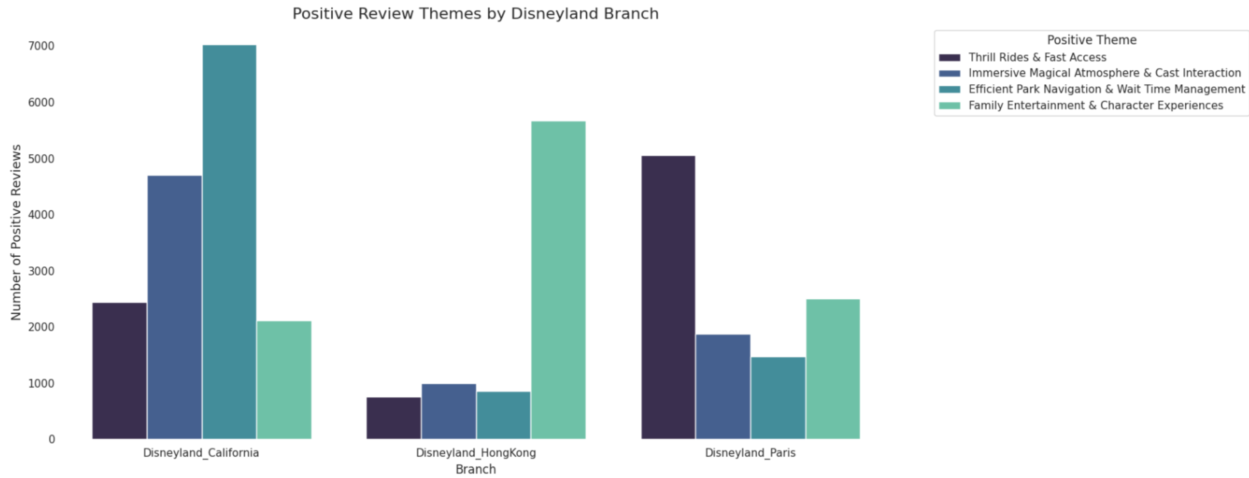


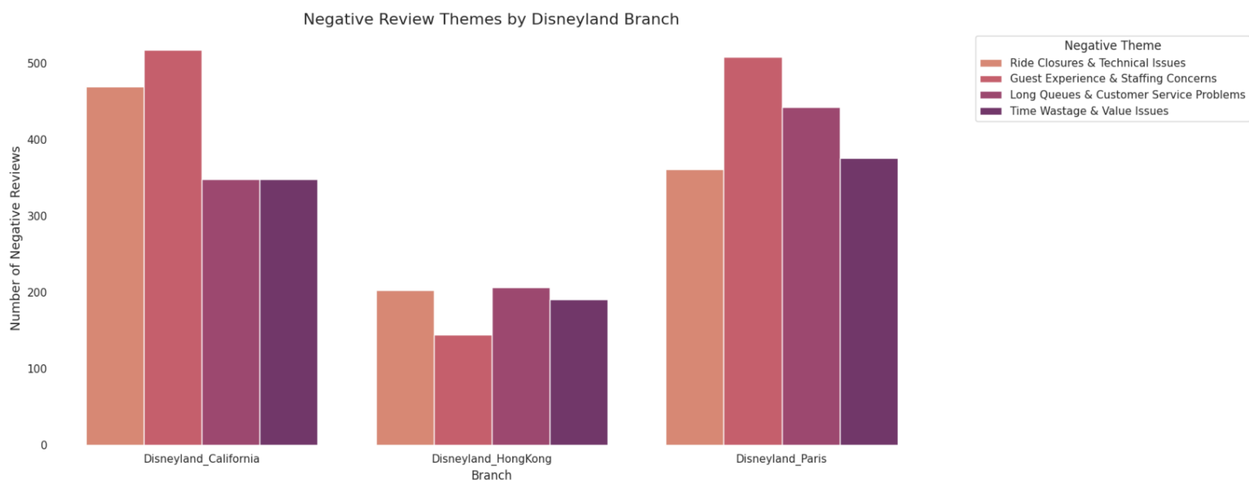
Top 20 TF-IDF Terms (Bigrams)
Term
fast pas
year old
hong kong
theme park
disney world
space mountain
one day
great time
disney park
disneyland paris
place earth
happiest place
cast member
fast pass
california adventure
first time
make sure
wait time
small world
disneyland park
0
100
200
300
400
500
600
700
800
TF-IDF Score



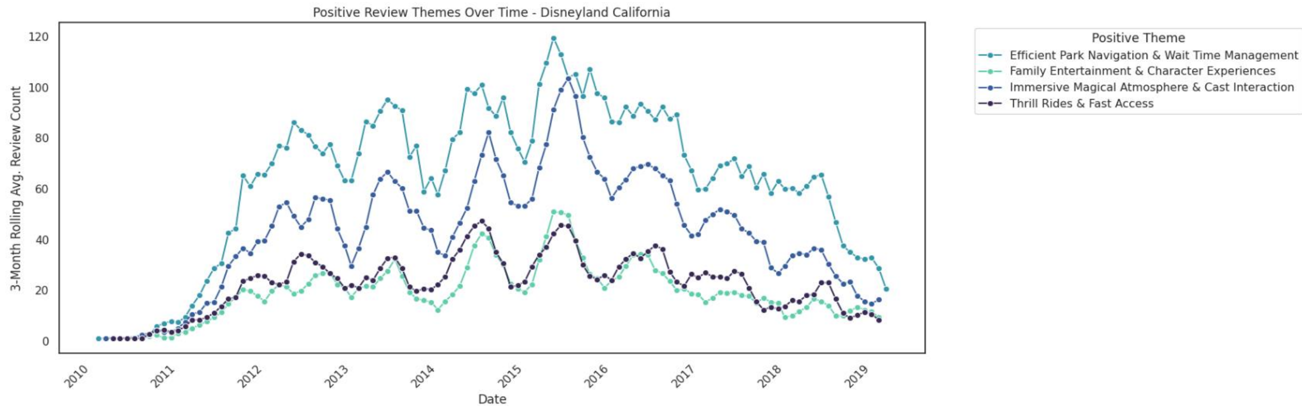
Top 20 TF-IDF Terms (Trigrams)
Term
happiest place earth
hong kong disneyland
big thunder mountain
fast pas system
get fast pas
disneyland hong kong
year old daughter
lion king show
use fast pas
california adventure park
disneyland california adventure
dream come true
day park hopper
indiana jones ride
toy story land
walt disney world
walt disney studio
sleeping beauty castle
disney world florida
park one day
0
100
200
300
400
500
600
700
800
TF-IDF Score

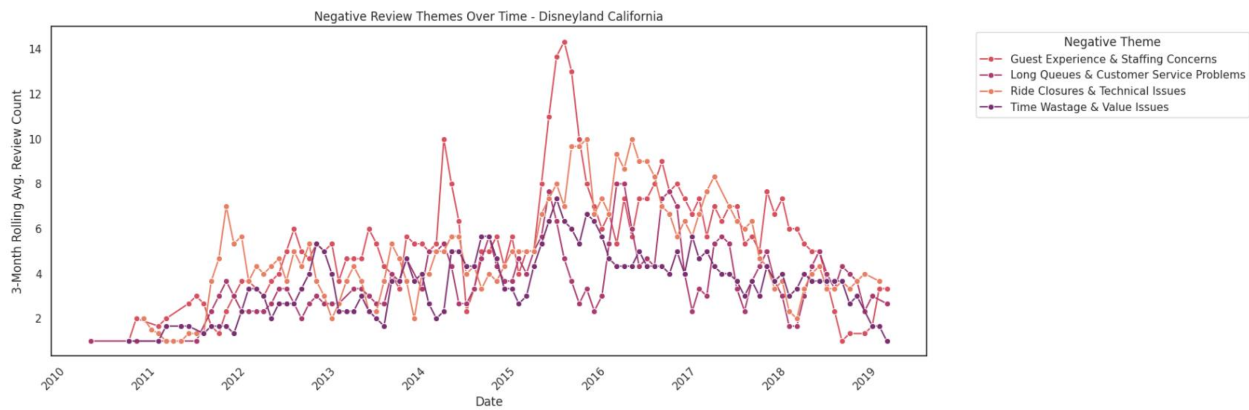
Appendix M

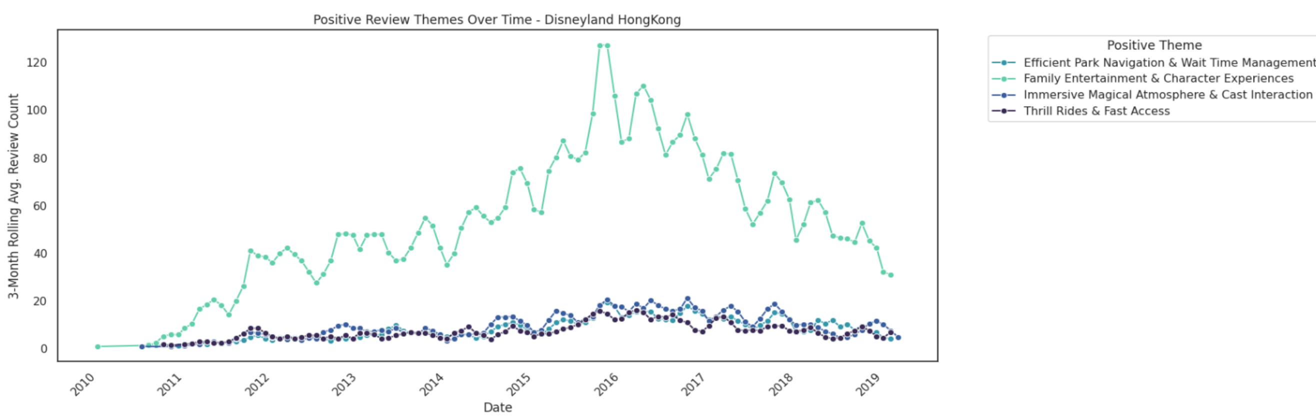

Positive Review Themes by Disneyland Branch
Number of Positive Reviews
0
1000
2000
3000
4000
5000
6000
7000
Disneyland_California
Disneyland_HongKong
Disneyland_Paris
Branch
Positive Theme
Thrill Rides & Fast Access
Immersive Magical Atmosphere & Cast Interaction
Efficient Park Navigation & Wait Time Management
Family Entertainment & Character Experiences

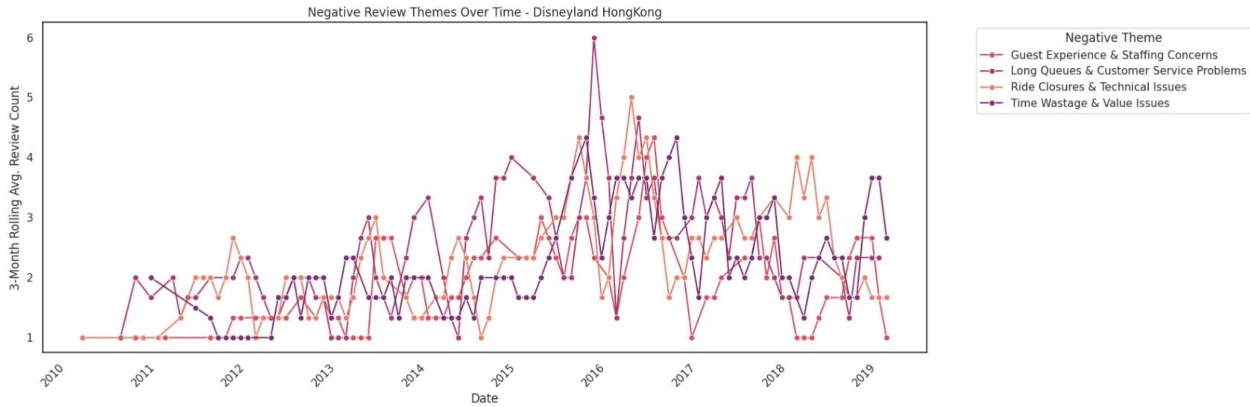

Negative Review Themes by Disneyland Branch
Number of Negative Reviews
0
100
200
300
400
500
Disneyland_California
Disneyland_HongKong
Disneyland_Paris
Branch
Negative Theme
Ride Closures & Technical Issues
Guest Experience & Staffing Concerns
Long Queues & Customer Service Problems
Time Wastage & Value Issues

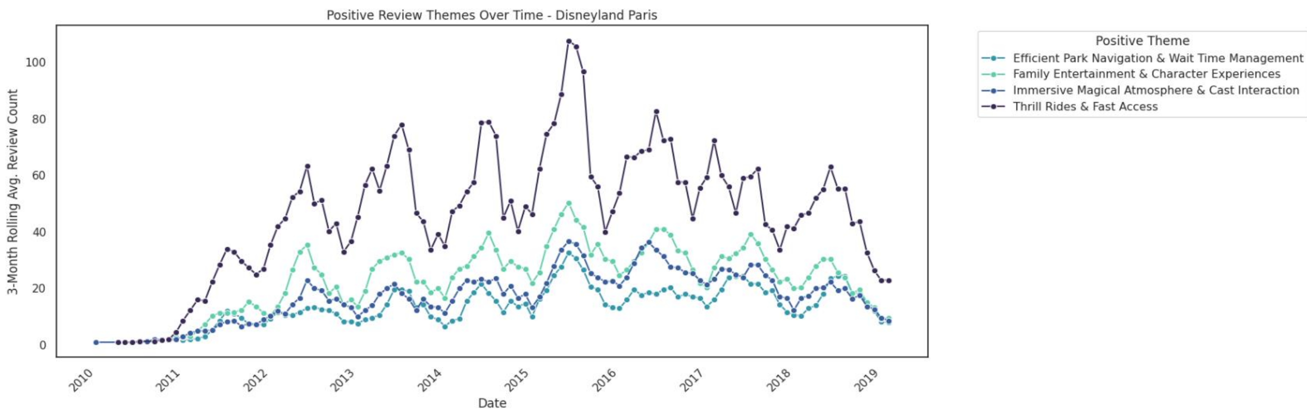
Appendix N

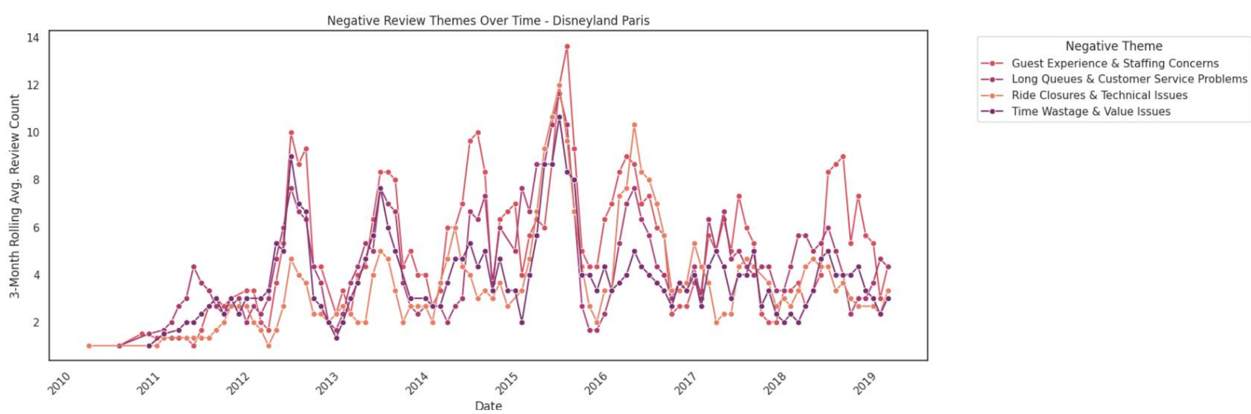

Positive Review Themes Over Time - Disneyland California
3-Month Rolling Avg. Review Count
Date
Positive Theme
Efficient Park Navigation & Wait Time Management
Family Entertainment & Character Experiences
Immersive Magical Atmosphere & Cast Interaction
Thrill Rides & Fast Access


Negative Review Themes Over Time - Disneyland California
3-Month Rolling Avg. Review Count
Date
Negative Theme
Guest Experience & Staffing Concerns
Long Queues & Customer Service Problems
Ride Closures & Technical Issues
Time Wastage & Value Issues


Positive Review Themes Over Time - Disneyland HongKong
3-Month Rolling Avg. Review Count
Date
Positive Theme
Efficient Park Navigation & Wait Time Management
Family Entertainment & Character Experiences
Immersive Magical Atmosphere & Cast Interaction
Thrill Rides & Fast Access


Negative Review Themes Over Time - Disneyland HongKong
3-Month Rolling Avg. Review Count
Date
Negative Theme
Guest Experience & Staffing Concerns
Long Queues & Customer Service Problems
Ride Closures & Technical Issues
Time Wastage & Value Issues


Positive Review Themes Over Time - Disneyland Paris
3-Month Rolling Avg. Review Count
Date
Positive Theme
Efficient Park Navigation & Wait Time Management
Family Entertainment & Character Experiences
Immersive Magical Atmosphere & Cast Interaction
Thrill Rides & Fast Access


Negative Review Themes Over Time - Disneyland Paris
3-Month Rolling Avg. Review Count
Date
Negative Theme
Guest Experience & Staffing Concerns
Long Queues & Customer Service Problems
Ride Closures & Technical Issues
Time Wastage & Value Issues

Appendix O


WordCloud of Disneyland Reviews
even
one
fast pa
year old
kid
attraction
went
space mountain
disneyland paris
hong kong
good
think
people
park
queue
child
got
everything
great
take
little
fast pass
going
although
nice
found
make sure
staff
way
disney world
theme park
still
one day
small world
need
know
amazing
say
experience
main street
really
crowd
great time
see
better
around
wait time
well
disney park
though
thing
much
make
cast member
fun
california adventure
happiest place
said
hotel
first
crowded
disneyland park
thought
never
try
love
disney character
day park
took
lot
left
find
best
restaurant
indiana jone
feel
look
busy
family
disappointed
especially
loved
made
told
first time
may
thunder mountain
many
seem
right
line
used
everyone
must
want
enjoy
definitely
etc
place earth
area