Topic Modeling on Disneyland Reviews

May 2,2023

1

Aastha Tandon
MSBA, CSU East Bay

Table of Contents

1. Overview

2. Motivation

3. Dataset Description

4. Research Question

5. Data Preprocessing Steps

6. Exploratory Data Analysis (EDA)

7. Methodology

8. Conclusion and Future Work

Overview

- Disneyland is a popular theme park in many parts of the world with millions of visitors annually

- Customer feedback is important in the service industry, and Disneyland is no exception

- Our research focus on customer topics across branches and time, offering insights into preferences and experiences

Motivation

- By learning about the most spoken topics by the customers, Disneyland can enhance visitor satisfaction, refine services, and make data-driven decisions regarding their marketing strategies

- Disneyland can better tailor their offerings to meet their needs and expectations, ultimately improving both the customer experience and their business outcomes

Dataset Description

- Dataset - https://www.kaggle.com/datasets/arushchillar/disneyland-reviews

- Contains 42k reviews of three Disneyland branches - Paris, California, and Hong Kong, posted by visitors on Trip Advisor

- Includes reviewer's *Rating*, *Year_month* of visit, *reviewer* l*ocation*, *review text*, and *Disneyland branch location*

Research Questions

1. What are the key topics mentioned in positive and negative customer reviews across different branches and the factors associated to it ?

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

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

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

Data Cleaning and Preprocessing

- Cleaned dataset removing rows where missing entries in *Year_Month* column. Resulting in 40K reviews (18K - California, 12K- Paris, and 9K - Hong Kong)

- Extracted the "*Year*" and "*Month*" separately from the *Year_Month*

- Grouped the months into "Season" namely *Fall, Winter, Spring & Summer*

- Series of preprocessing like tokenization of sentences to words, stemming, removal of stop words and commonly used words based on multiple iterations

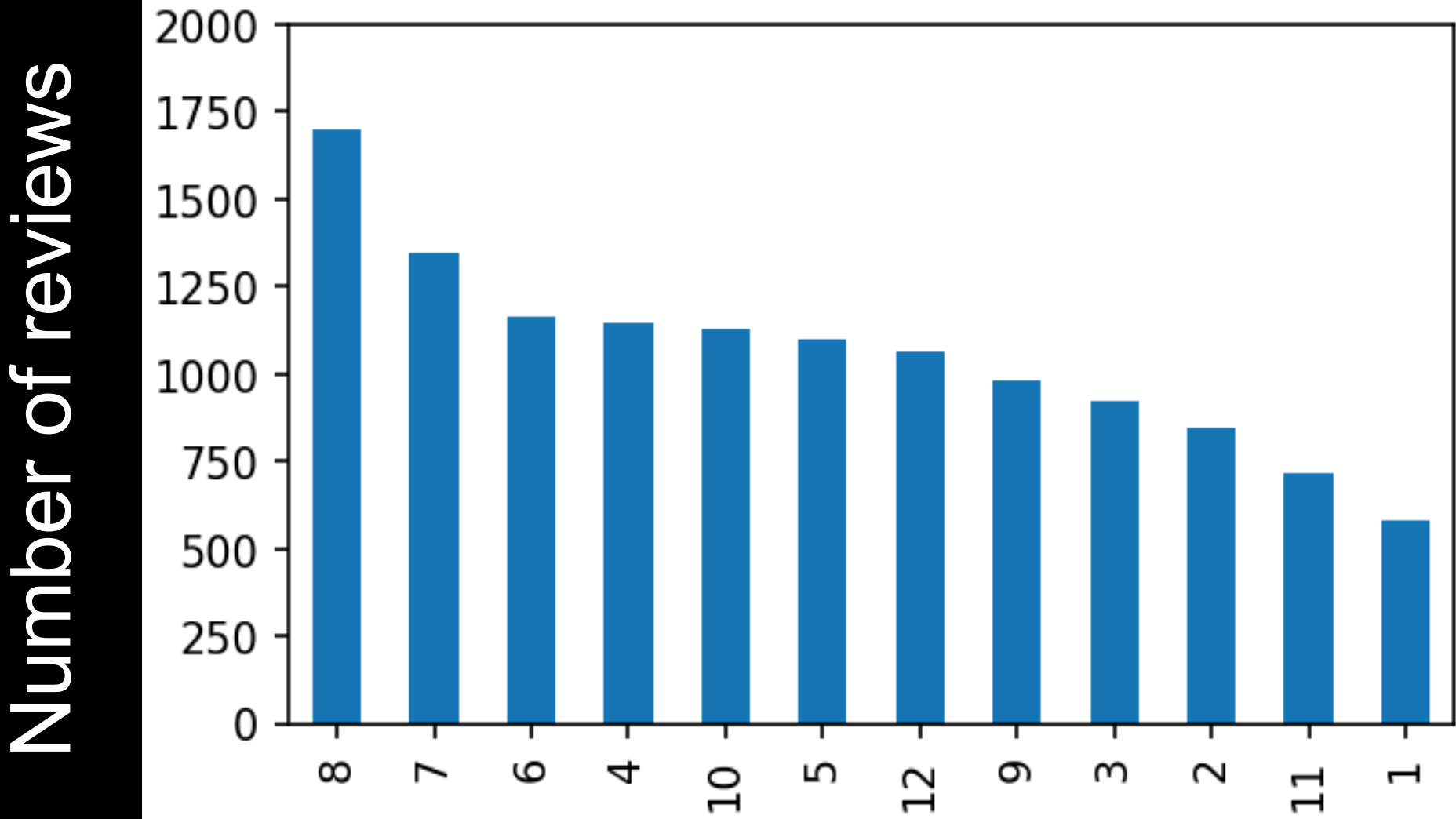# Exploratory Data Analysis

Change in average rating over years
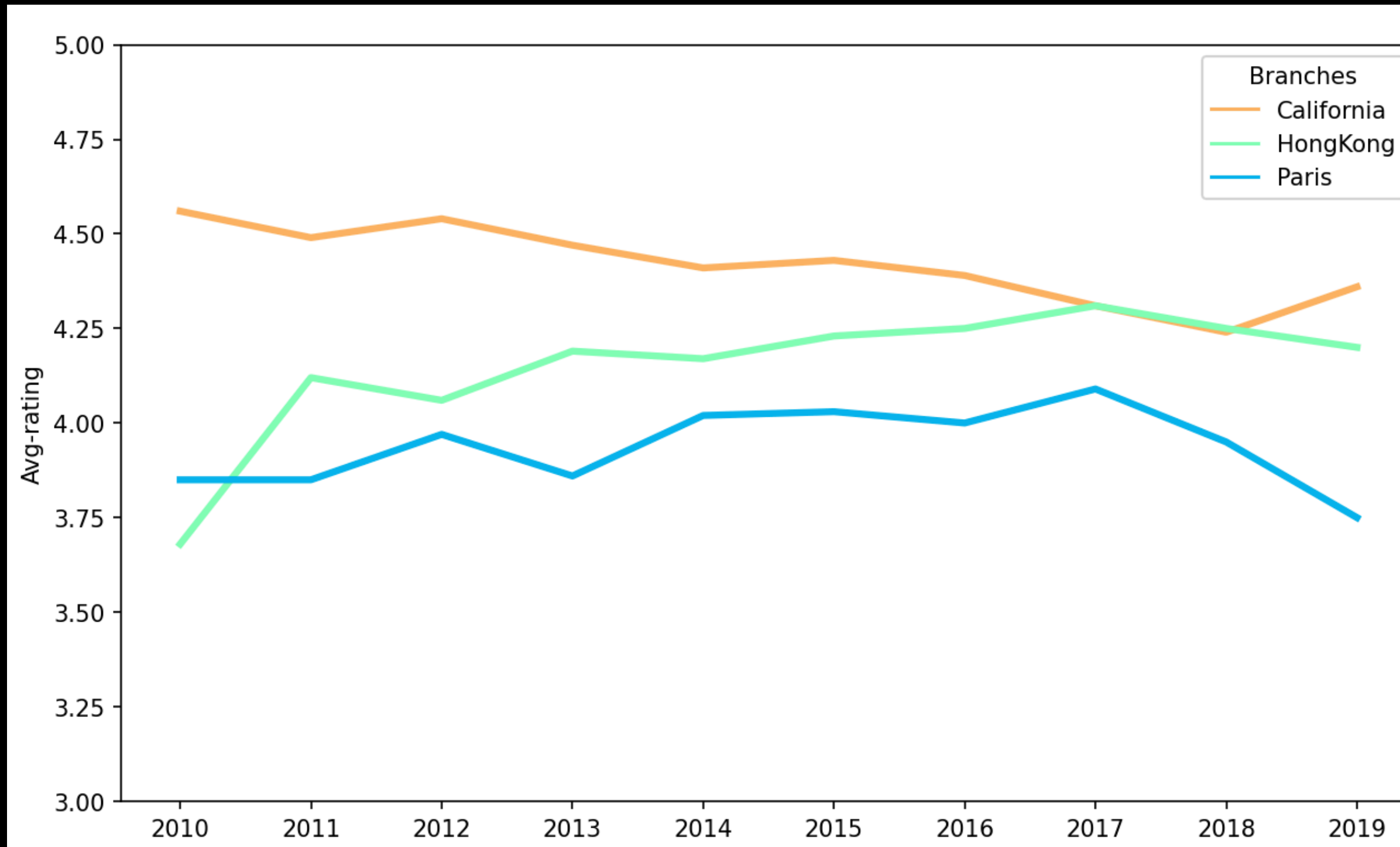
Change in average rating over months

Year
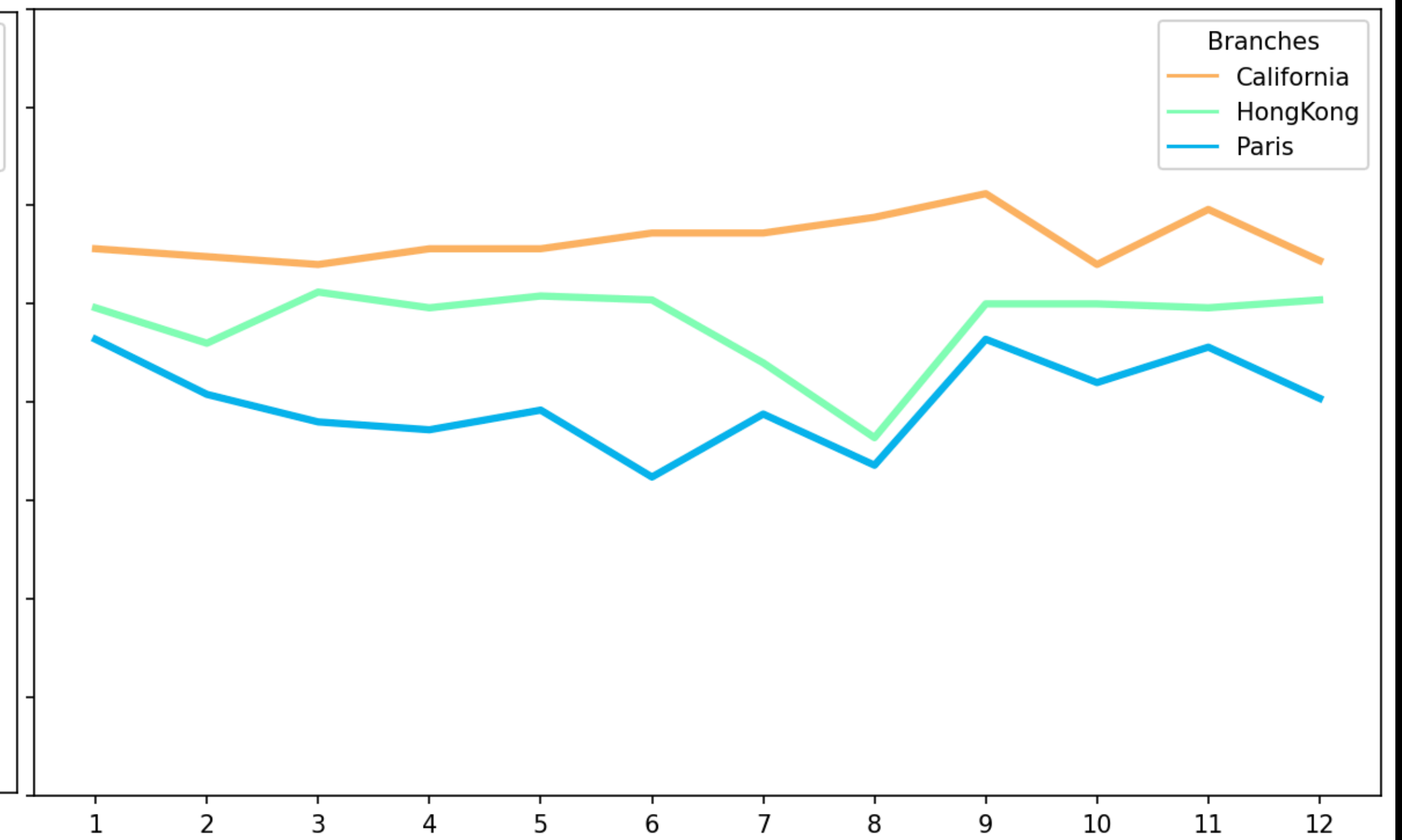
Month number
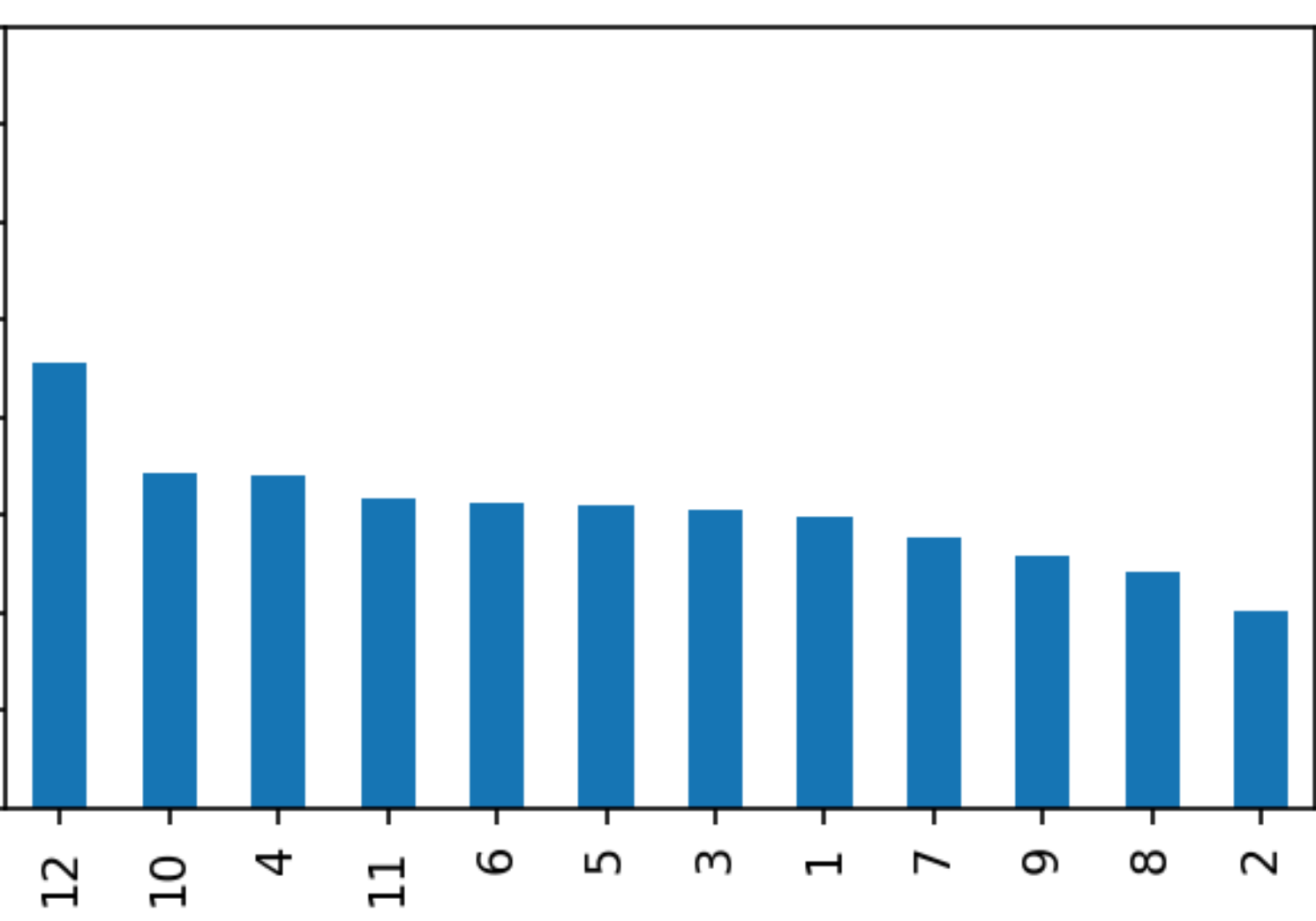
Exploratory Data Analysis
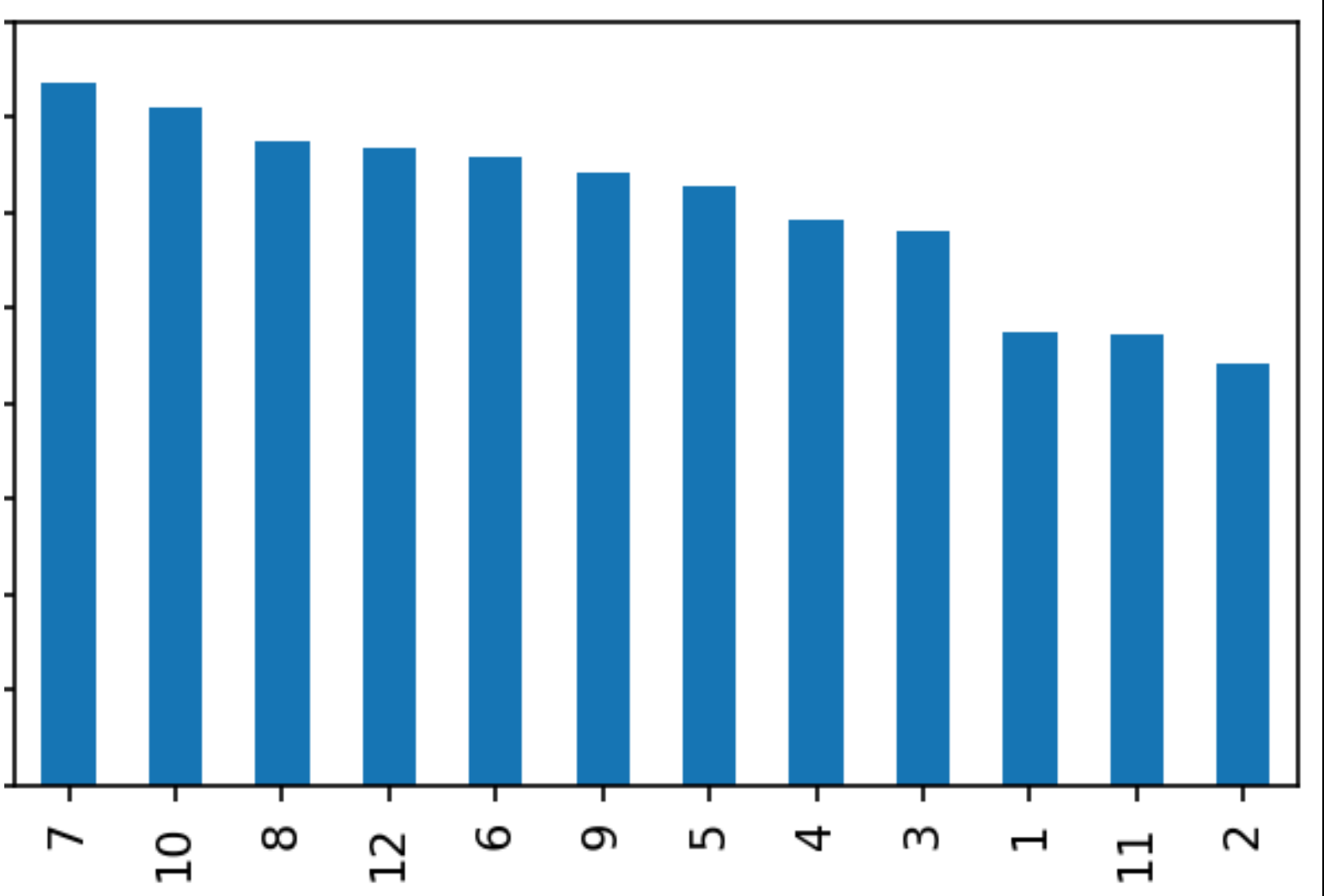
Analysis of peak time in Disneyland in different locations

Methodology

Topic modeling on positive and negative reviews

- All the analysis conducted for each branch separately

- Based on the rating scale of 1 to 5, we bucketed reviews in Negative (1 to 3) and Positive (4 & 5) bins

- Learning topics for positive and negative feedback separately by using LDA topic modeling

- Positive feedback reflects positive topics attracting customers

- Negative feedback leads to negative topics which might be harming the business

Topic modeling on Positive and Negative reviews
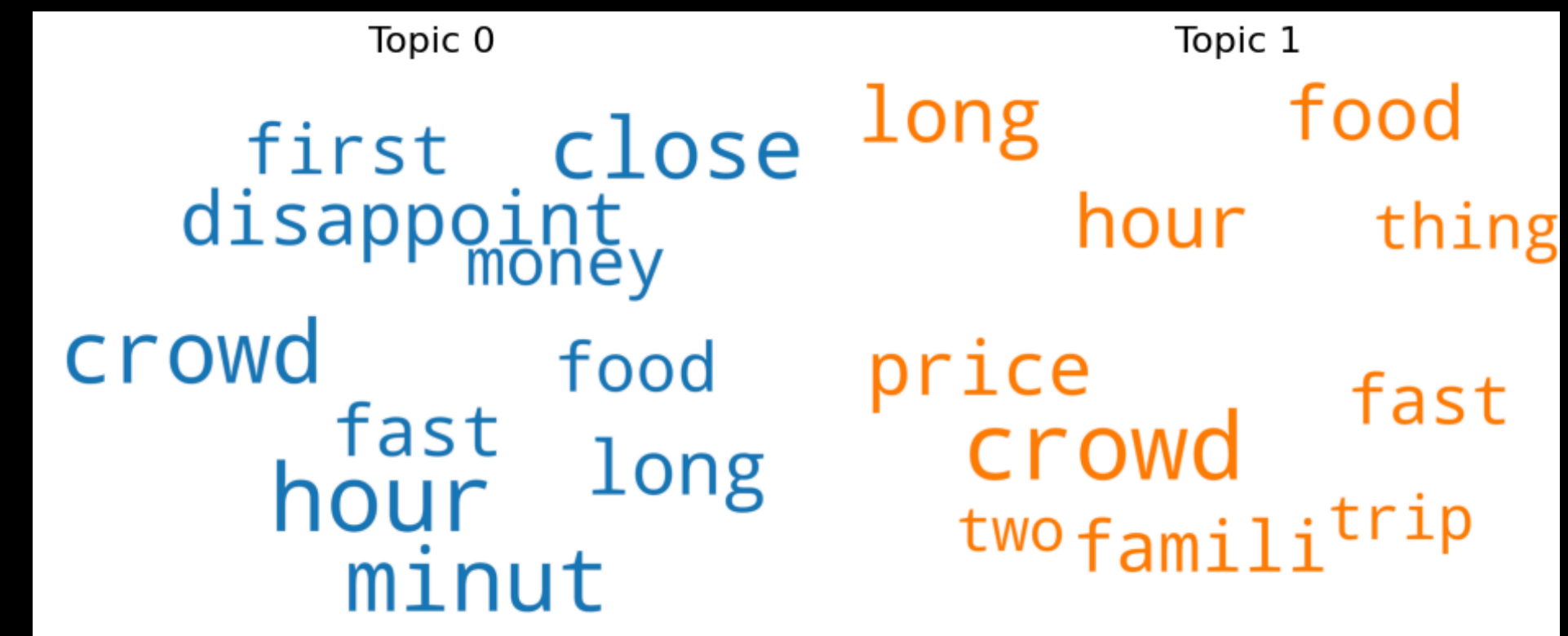
California
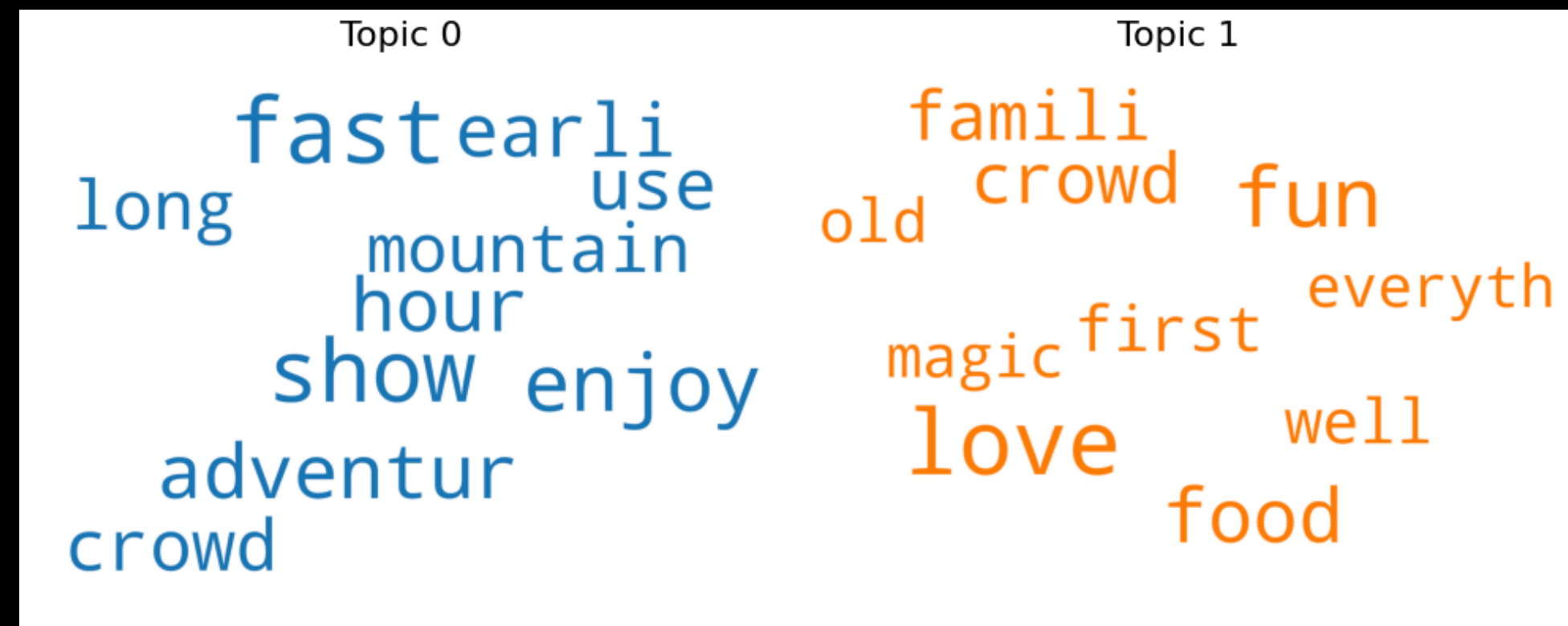
Positive review topics

Topic 0: Theme Park Attractions and Shows Experience
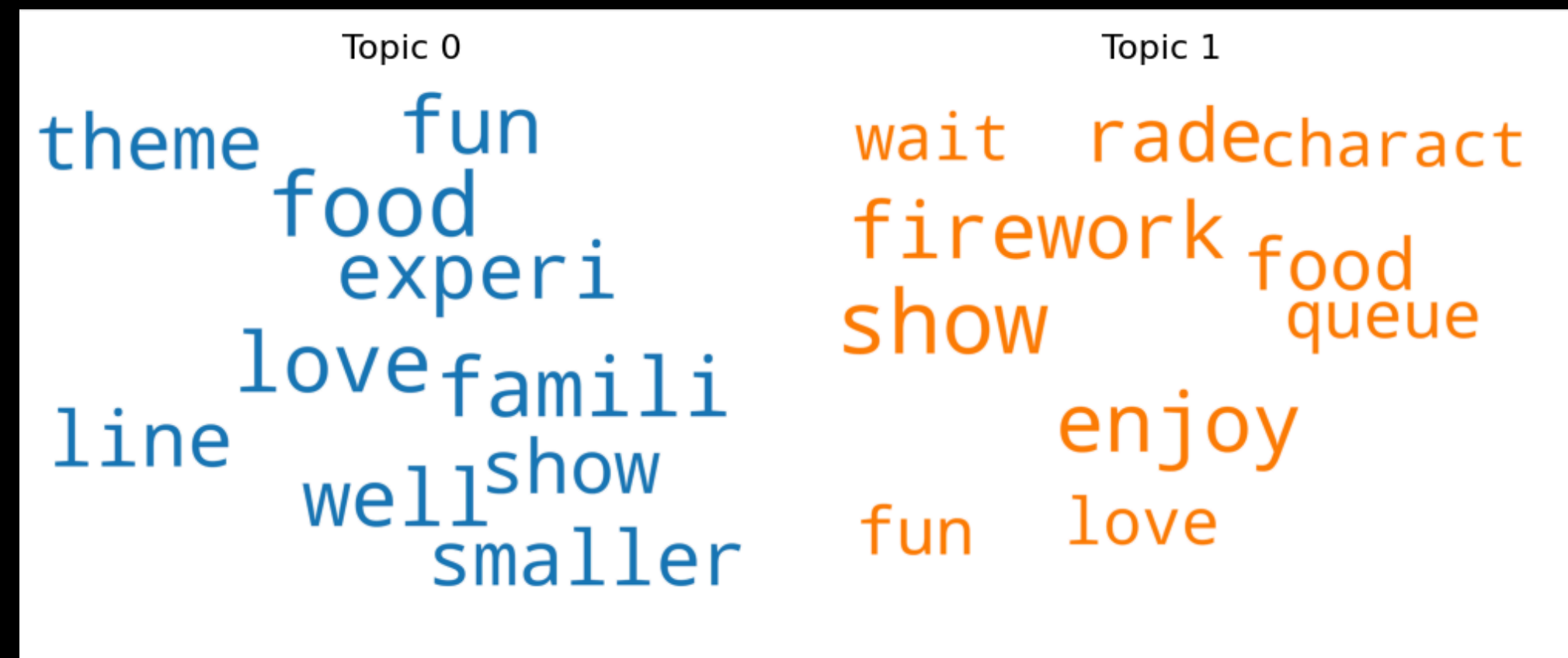
Topic 1: Overall Theme Park Experience

Negative review topics

Topic 0 : Disappointing Crowd Experience

Topic 1 : Unsatisfactory Food and Price at Crowded Restaurant
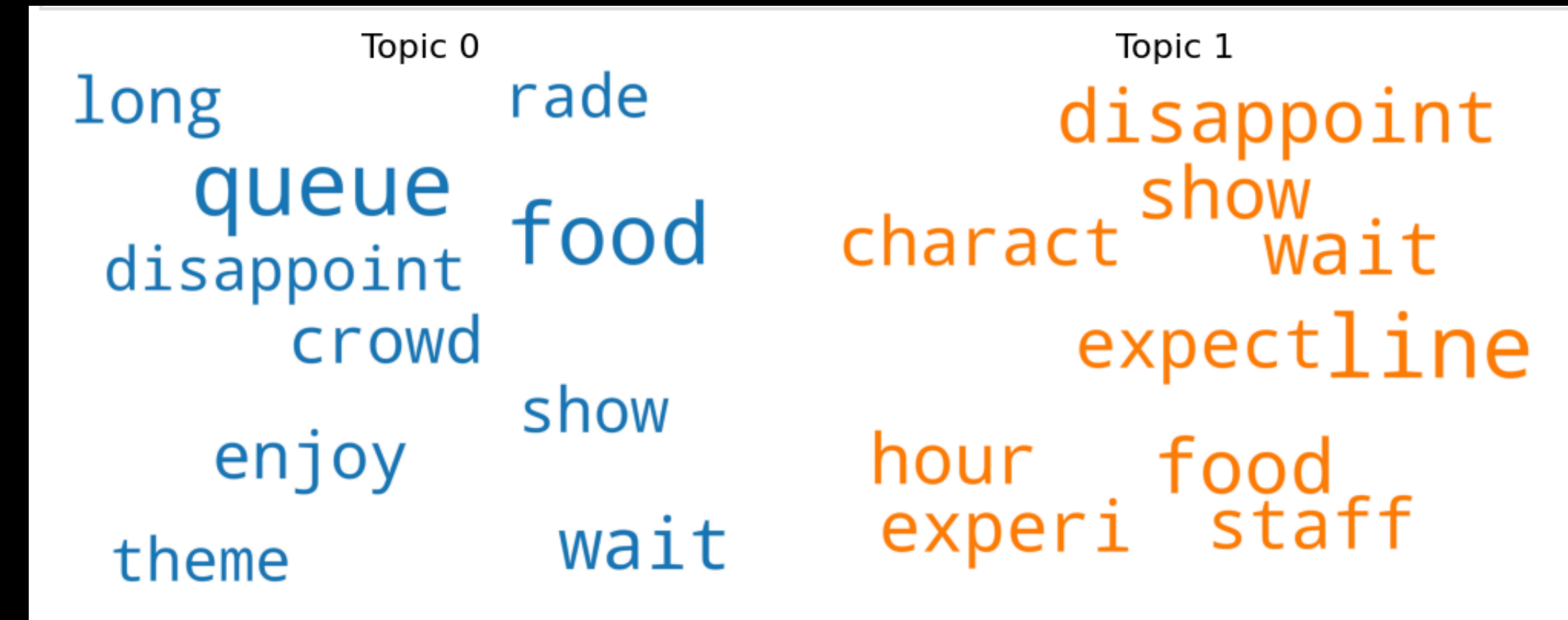
Topic modeling on Positive and Negative reviews

Hong Kong

Positive review topics

Negative review topics

Topic 0: Experience of Attending a Theme Park

Topic 1: Enjoyment of Firework Shows
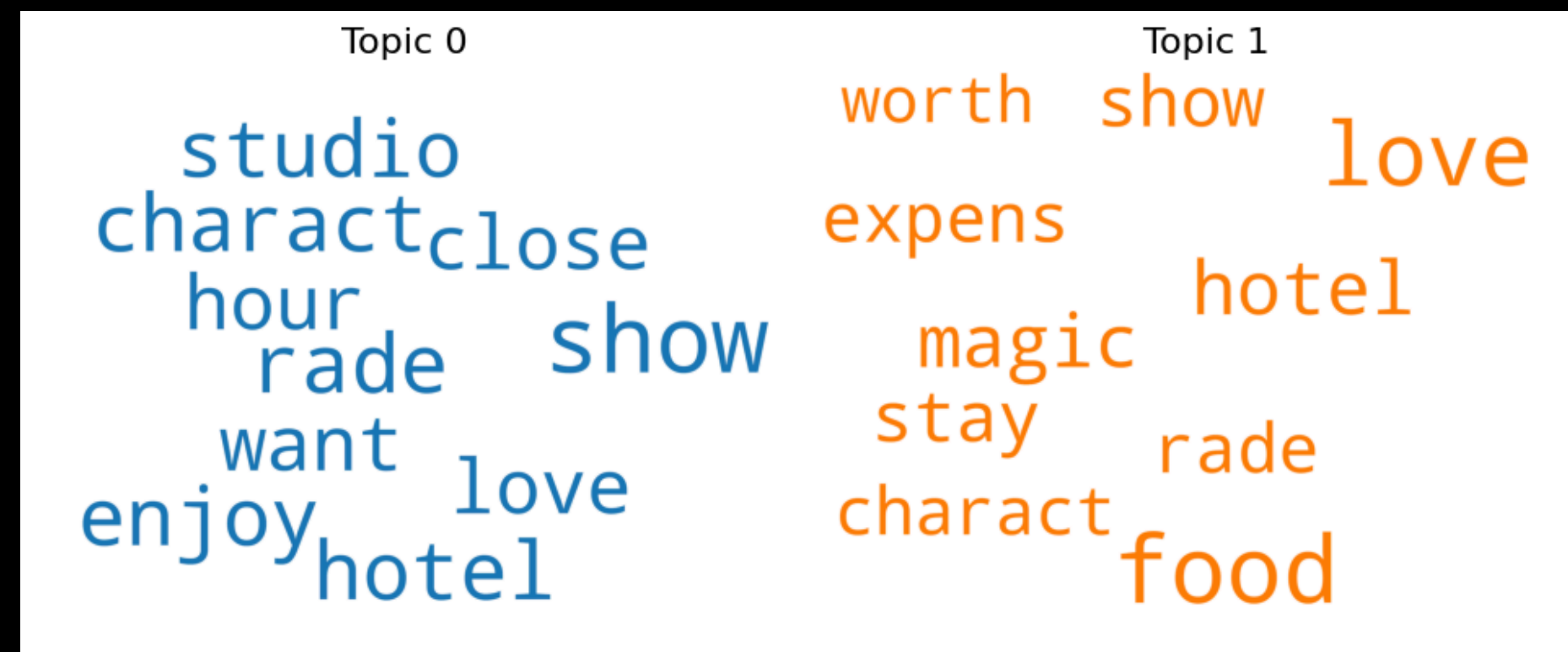
Topic 0: Disappointing experience at a crowded theme park

Topic 1: The frustrations of dining out at busy restaurants
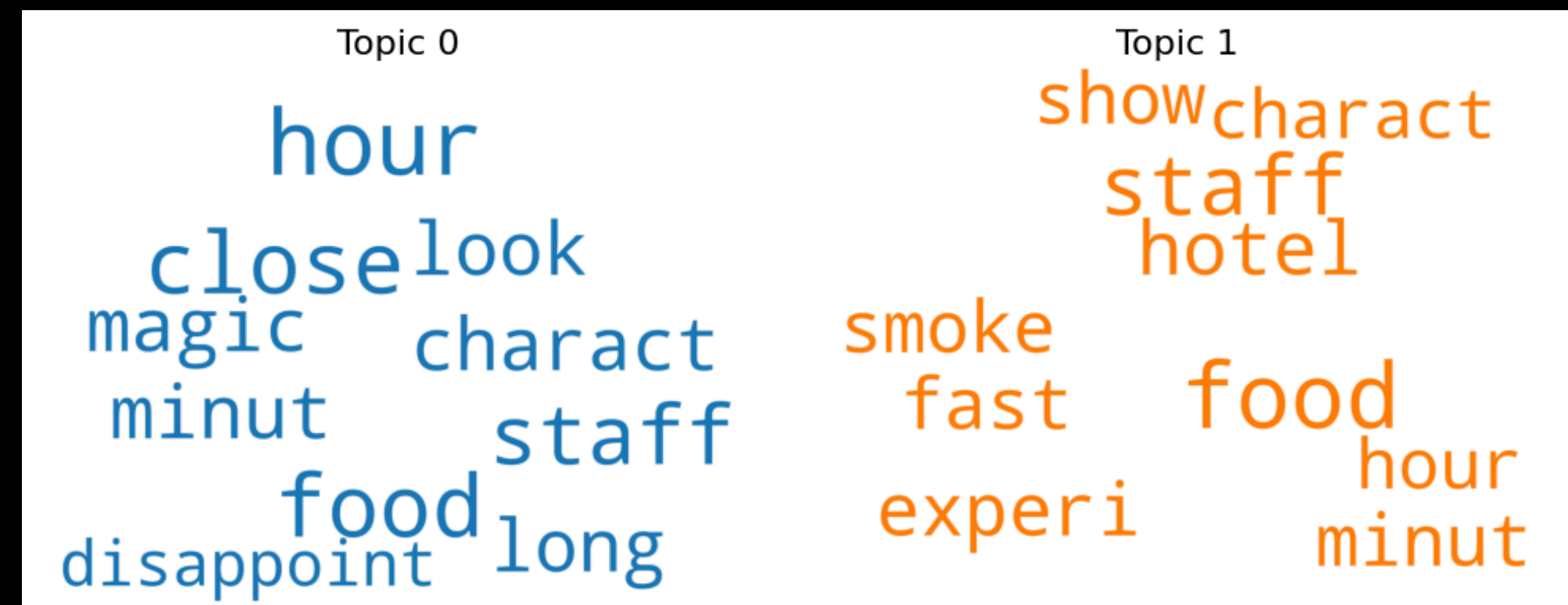
Topic modeling on Positive and Negative reviews

Paris

Positive review topics

Negative review topics

Topic 0: Unforgettable Memories and Magical Experiences

Topic 1: Culinary Delights of Disneyland Paris

Topic 0: Negative Reviews about Restaurant Food and Service

Topic 1: Negative Reviews about Hotel Accommodation

Methodology

Study topic distribution across months

- Analysis conducted on all the reviews of individual branch

- Learned K topics across all the reviews for a location

- Computed the average topic distribution for each month i.e from January to December

- This enabled us to track the changes in topics across months and analyze the reasons for the changes.

Topic modeling on all reviews

California

Identified 4 topics, encompassing both positive and negative aspects.
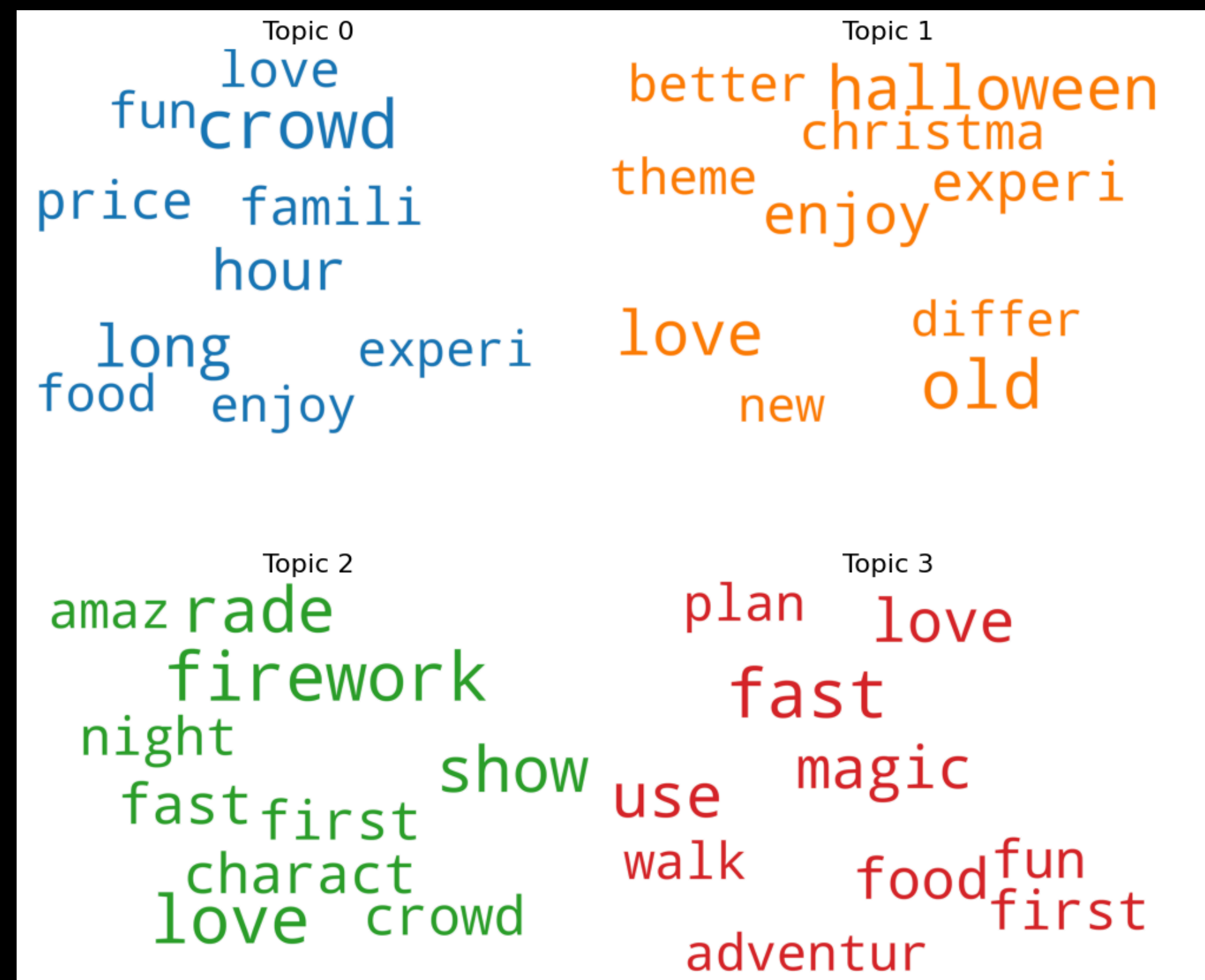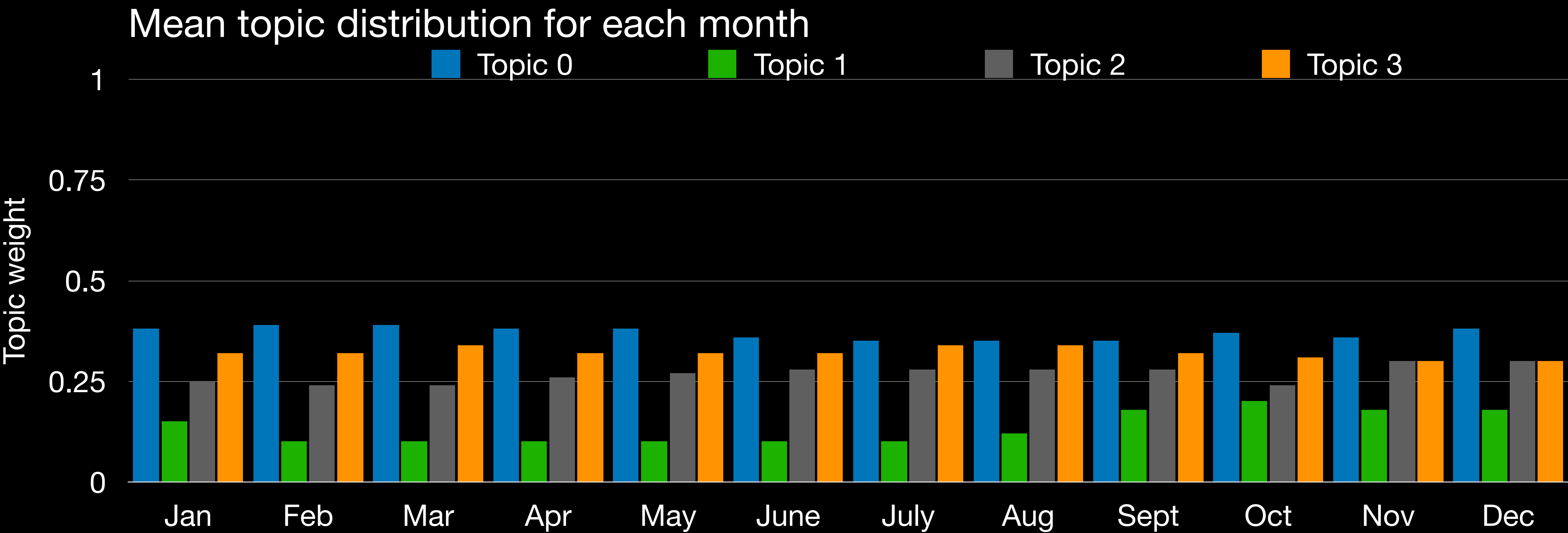

Topic 0 : Crowd management

Topic 1 : Festive experiences

Topic 2: Fireworks and parade shows

Topic 3 : Fastpass enhancing Magical adventures

Topic distribution across months

California

Mean topic distribution for each month

Topic 0 : Crowd management   Topic 1 : Festive Experiences

Topic 3 : Firework and Parade Show Topic 4 : Fast Pass enhancing Magic Experience

Topic 0 is most talked and Topic 1 is least talked in all the months of the year

Topic modeling on all reviews

Hong Kong

Identified 4 topics, encompassing both positive and negative aspects.
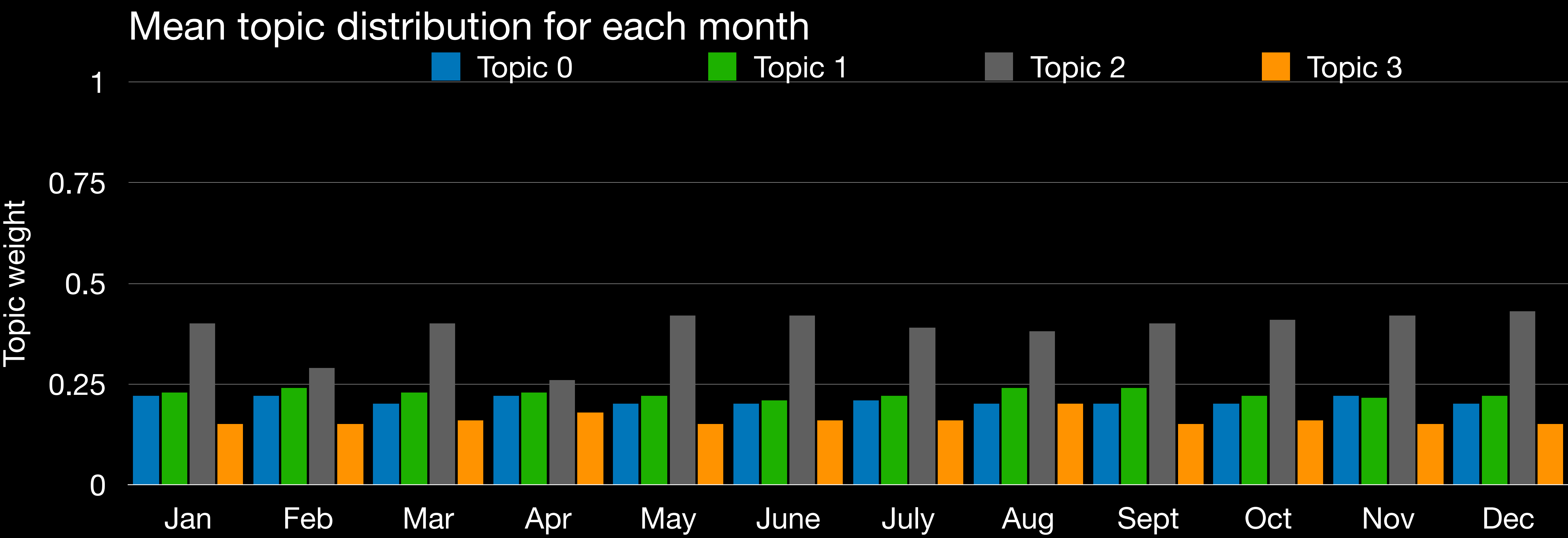

Topic 0: Theme park experiences

Topic 1: Food and dining

Topic 2: Entertainment

Topic 3: Queue management

Topic distribution across months

Hong Kong

Mean topic distribution for each month

Legend: ■ Topic 0 ■ Topic 1 ■ Topic 2 ■ Topic 3

Y-axis: Topic weight (0, 0.25, 0.5, 0.75, 1)

X-axis: Jan, Feb, Mar, Apr, May, June, July, Aug, Sept, Oct, Nov, Dec

Topic 0: Theme park experience   Topic 1: Food and dining
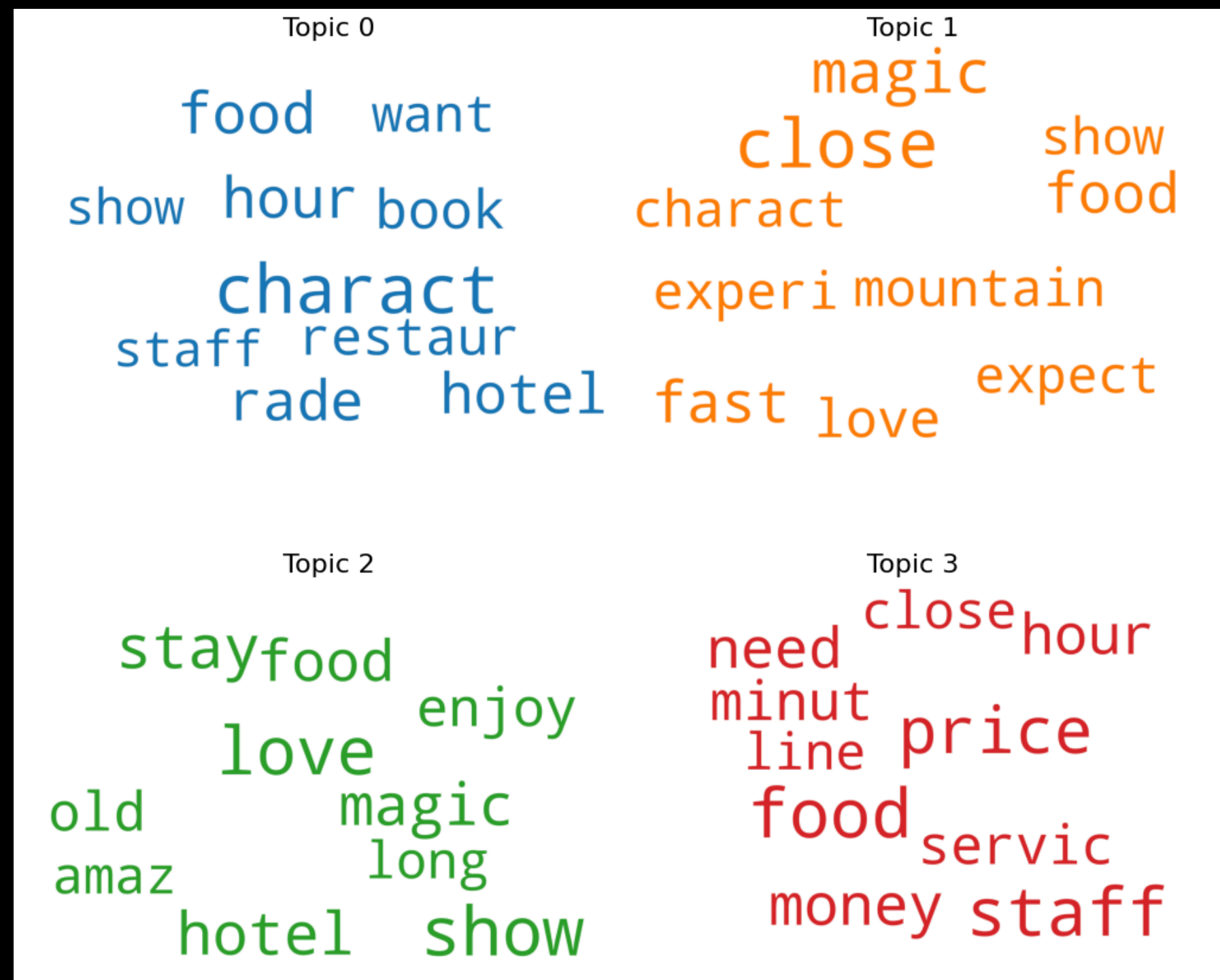
Topic 2: Entertainment   Topic 3: Queue management

We found Topic 2 most discussed and Topic 3 least discussed across the year

Topic modeling on all reviews

Paris

Identified 4 topics, encompassing both positive and negative aspects.

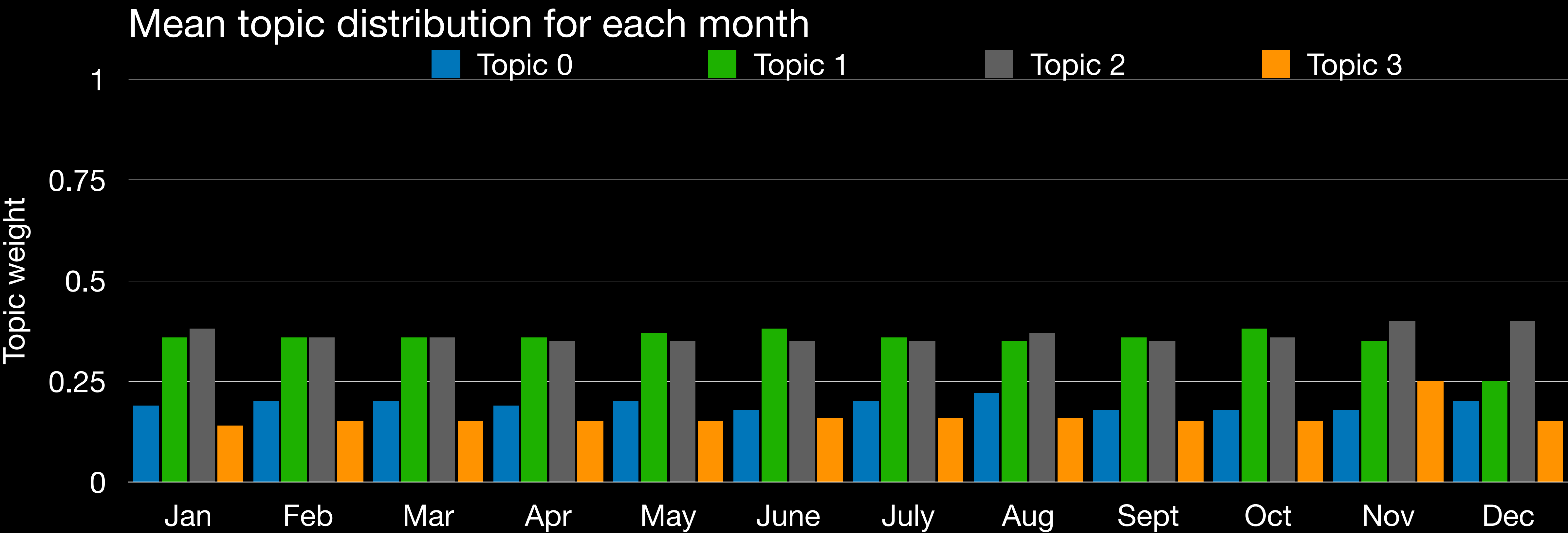Topic 0: Character Experiences at Disneyland Paris

Topic 1: Fast Pass for great overall experience

Topic 2: Hotel Accommodations and Dining

Topic 3:Tips and Tricks for a Magical Visit

Topic distribution across months

Paris

Mean topic distribution for each month

■ Topic 0   ■ Topic 1   ■ Topic 2   ■ Topic 3

Topic 0: Character Experience   Topic 1 : FastPass for great experience

Topic 2 : Hotel Accomodation & Dining Topic 3 : Tips & Tricks for Magical Visit

Topic 3 least discussed and competing Topic 1 and Topic 2 across the year

Methodology

Topic modeling for seasons on all reviews

- Analysis conducted on all the reviews of individual branch

- Learned k topics across all the reviews for a location

- Computed the average topic distribution for each season i.e. Fall, Summer ,Winter and Spring

- This enabled us to track the changes in topics across the seasons and analyze the reasons for the changes.

Recap: Topic modeling on all reviews

California

Identified 4 topics, encompassing both positive and negative aspects.
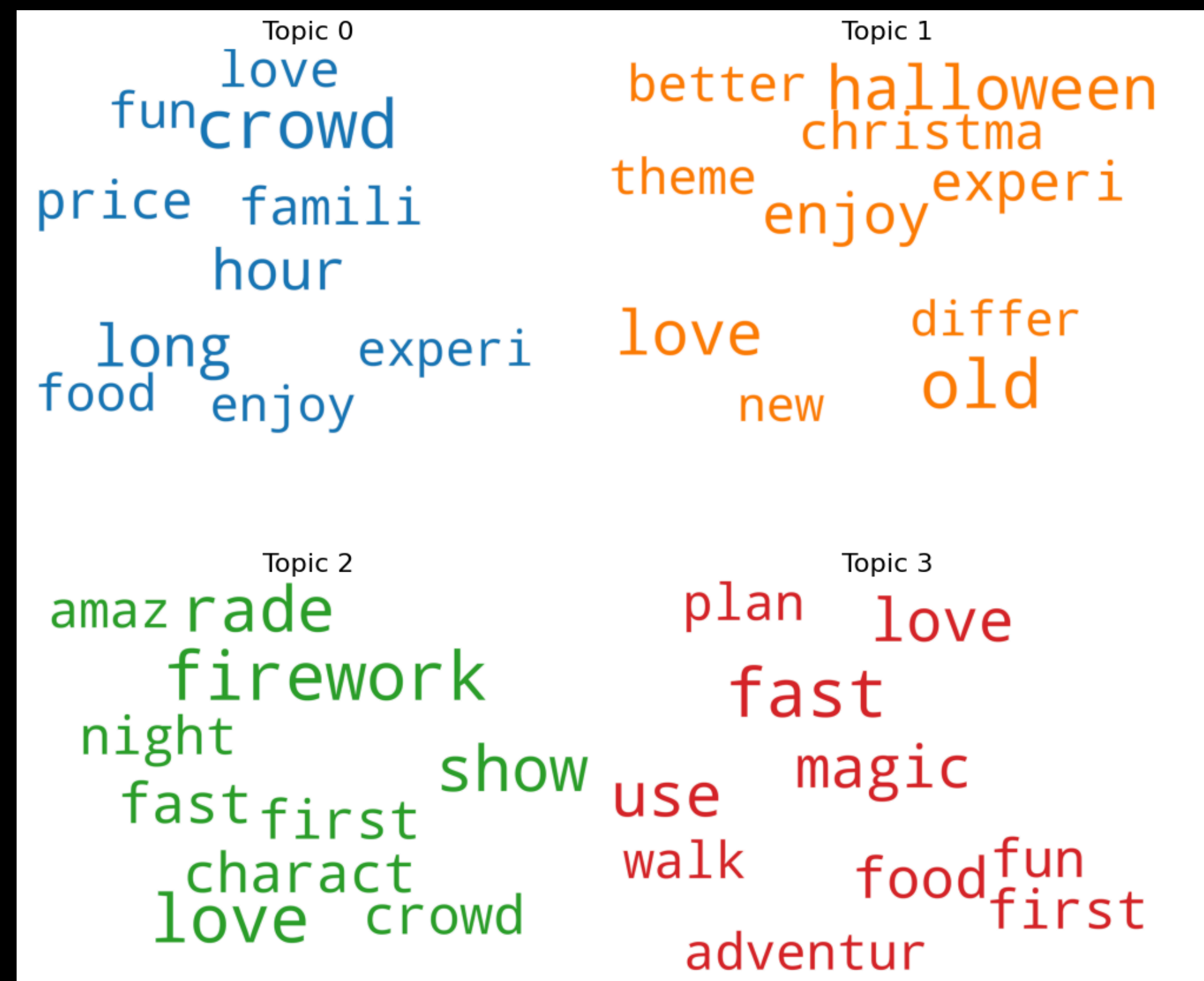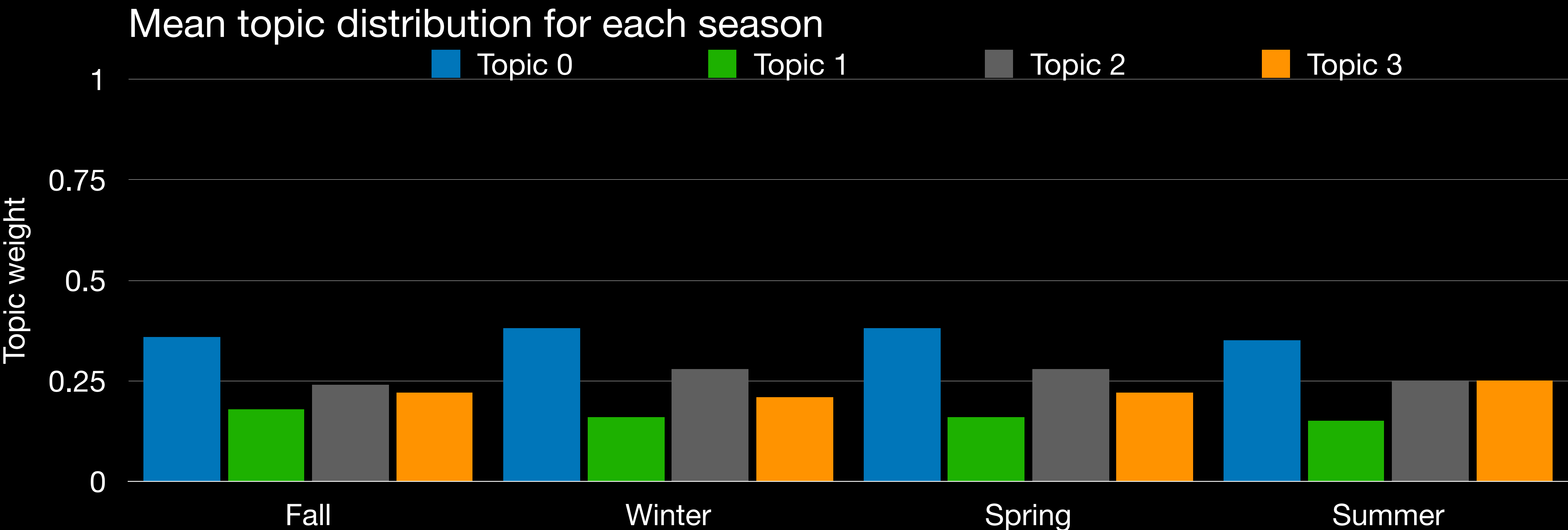

Topic 0 : Crowd management

Topic 1 : Festive experiences

Topic 2 : Fireworks and parade shows

Topic 3 : Fastpass enhancing Magical adventures

Topic distribution across seasons

California

Mean topic distribution for each season

Topic 0 : Crowd management   Topic 1 : Festive Experiences

Topic 3 : Firework and Parade Show   Topic 4 : Fast Pass enhancing Magic Experience

Topic 0 is most talked and Topic 1 is least talked across all the seasons

Recap: Topic modeling on all reviews

Hong Kong

Identified 4 topics, encompassing both positive and negative aspects.
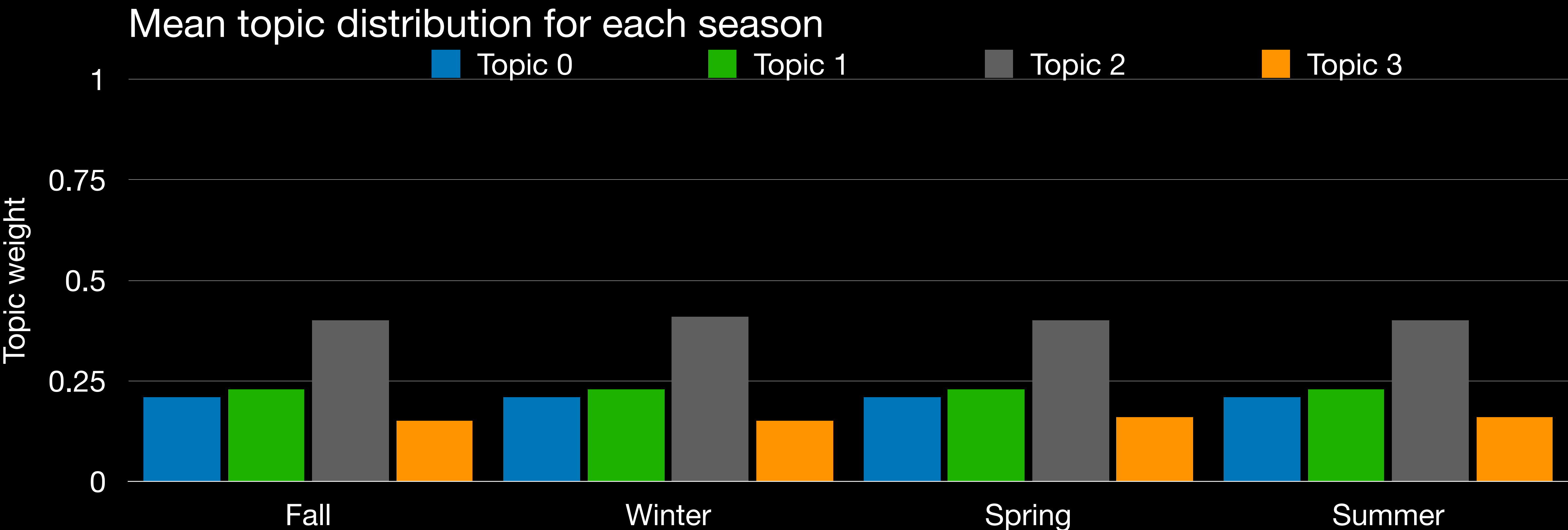

Topic 0: Theme park experiences

Topic 1: Food and dining

Topic 2: Entertainment

Topic 3: Queue management

Topic distribution across seasons

Hong Kong

Mean topic distribution for each season

Topic 0: Theme park experience   Topic 1: Food and dining
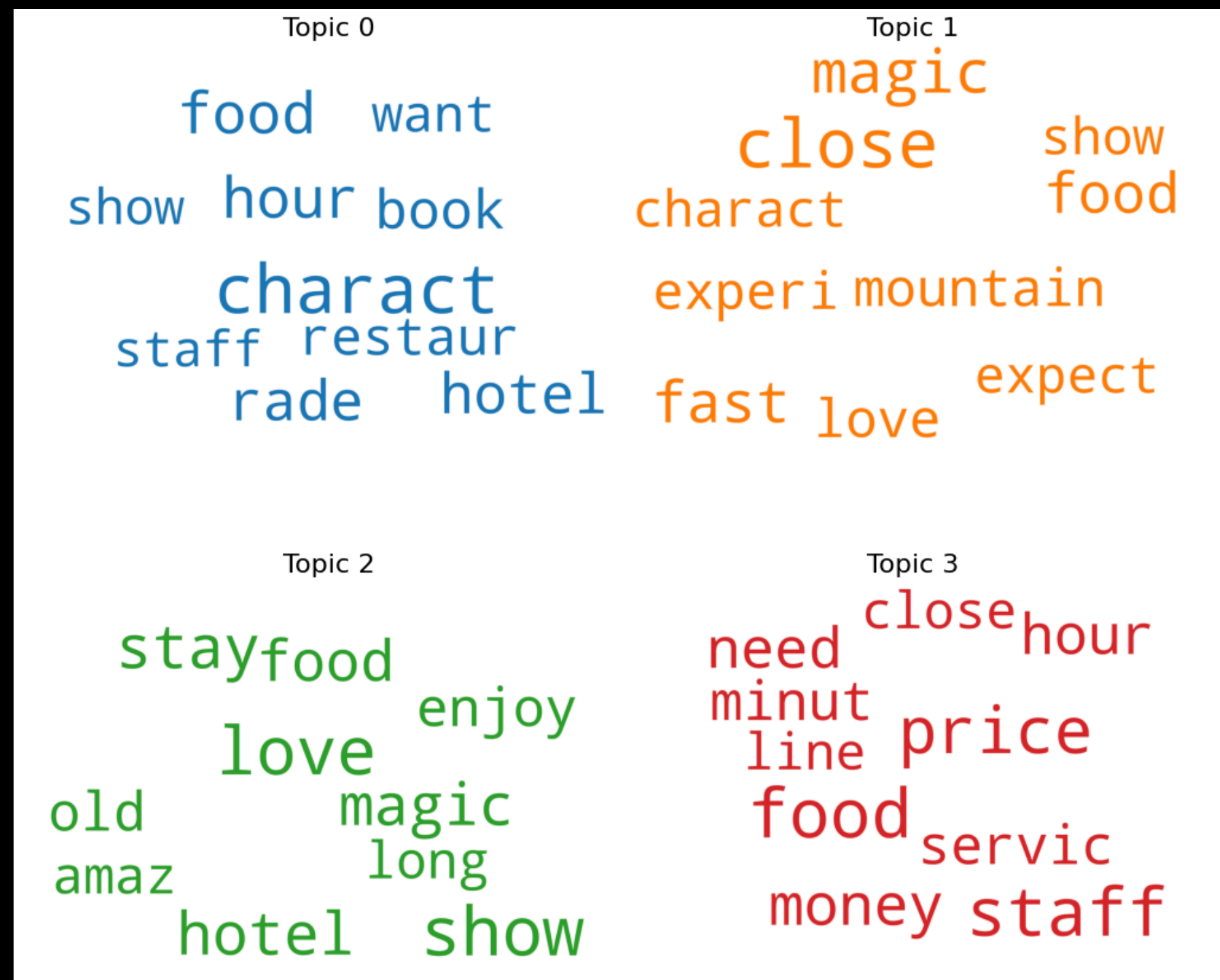
Topic 2: Entertainment   Topic 3: Queue management

Topic 2 is most discussed and Topic 3 is least discussed across the all the seasons

Recap: Topic modeling on all reviews

Paris

Identified 4 topics, encompassing both positive and negative aspects.

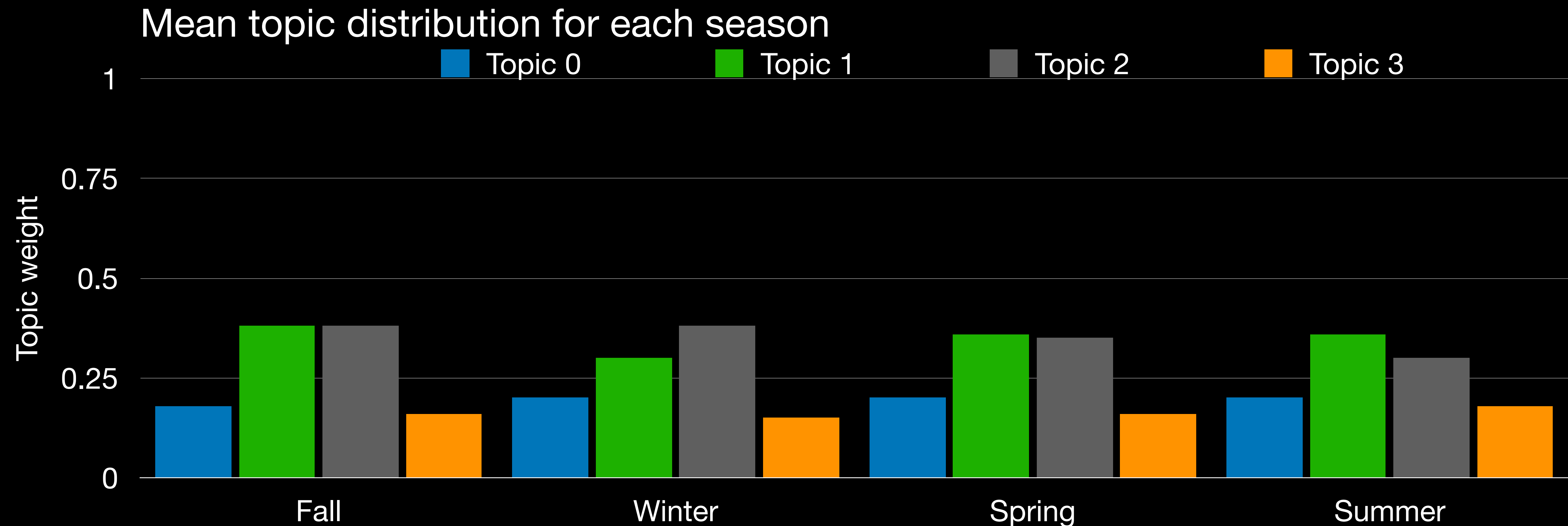Topic 0: Character Experiences at Disneyland Paris

Topic 1: Fast Pass for great overall experience

Topic 2: Hotel Accommodations and Dining

Topic 3:Tips and Tricks for a Magical Visit

Topic distribution across seasons

Paris

Mean topic distribution for each season

Topic 0: Character Experience   Topic 1 : FastPass for great experience

Topic 2 : Hotel Accomodation & Dining   Topic 3 : Tips & Tricks for Magical Visit

Topic 1 slightly more discussed in Summer & Spring than Topic 2 & equally in Fall. Topic 2 most discussed on Winter and Topic 3 is least discussed

Conclusion

Business insight for each Disneyland Branch

California	Hong Kong	Paris
• Implement effective crowd management strategies	• Manage visitor capacity	• Improve food quality
• Expand festive experiences	• Use technology and promote fast pass	• Staff training for better service
• Add new attractions and experiences	• Expand entertainment offerings	• Maintain hotel cleanliness and update amenities
• Implement digital tools to help visitors plan visit	• Increase menu offerings	• Fast pass to reduce wait times
• Improve the quality and variety of restaurant food	• Increase dining options	• Tailored promotions and seasonal events
• Timely Guest feedback	• Timely Guest feedback	• Timely Guest feedback

Limitation and Future Work

Limitations:

- Analysis limited to TripAdvisor reviews
- Not representative of visitors who write review in language other than English
- Did not consider demographic variables such as age and gender

Future Directions:

- Include reviews from other online platforms (like Yelp) and languages
- Consider demographic variables to better understand factors that influence experiences for each demographic group
- Use advanced NLP techniques for deeper insights such as - sentiment analysis, etc.