

Universal Studio Review Analysis

Bingxin Li, Ziyi Gao,
Shimin Liang, Yuxin Tang

Table of Contents

01 Goals

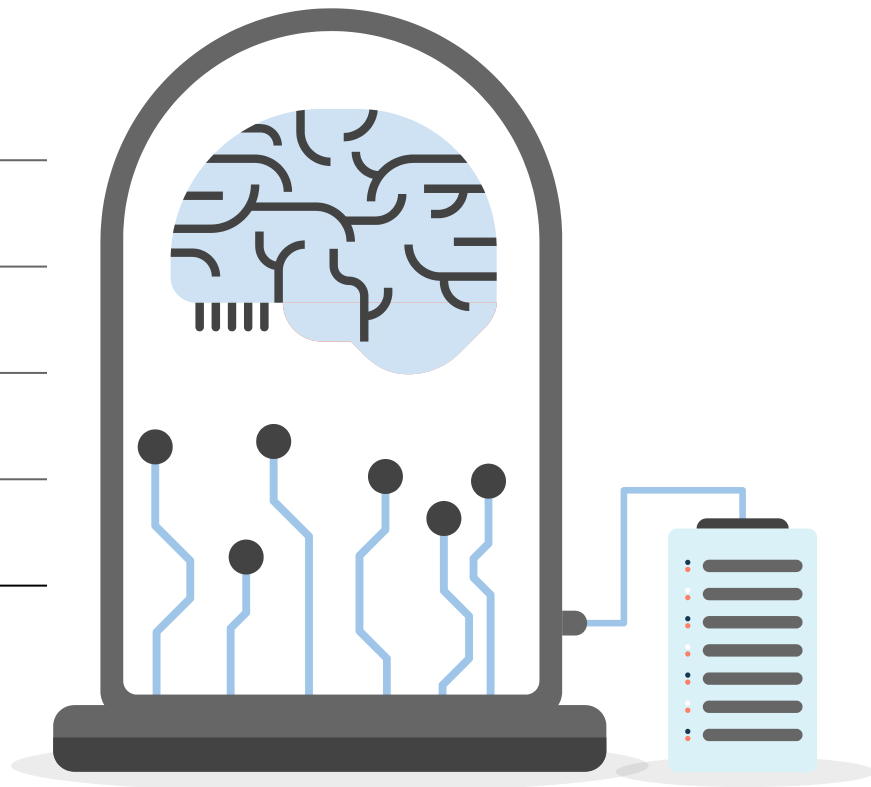
02 EDA & Text Preprocessing

03 N Gram Language Modeling

04 Topic Modeling

05 Feature Importance Modeling

06 Recommendation



Goals

01

Analysis

- Rating Trends
- General Performance
- Customer Segments



02

Implementation

- Solve problems that most poor reviews focused
- Keep advantages that most good reviews pointed out
- Balance between profit maximization and customer experience



Preparation

1. Exploratory Data Analysis
2. Text Preprocessing

Exploratory Data Analysis

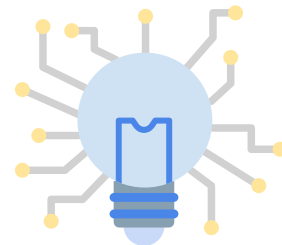


Universal Studio Databases

- The dataset includes 50,000+ reviews of 3 Universal Studios branches (Florida, Singapore, Japan).
- Each reviewer gave a rating (1-5) for their visit to Universal Studio and made a review with a title(title and review text have been combined in this project).
- Posted by visitors on the Tripadvisor website which is the world's largest travel guidance platform.
- Date from 2002 to 2021.

Exploratory Data Analysis

Customer Segments



Customer - Family

9.8%

Male	3443	son	1746
		dad	211
		husband	1486
Female	3262	daughter	1710
		mom	289
		wife	1263

Holiday

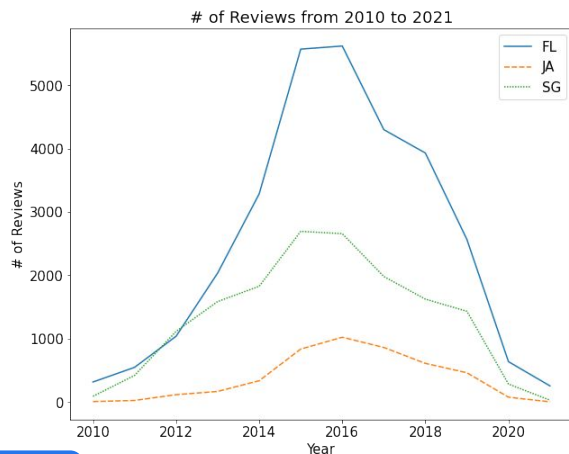
6.5%

Halloween	1457	Positive	1164 ★
		Negative	293
Christmas	944	Positive	778
		Negative	166
Birthday	382	Positive	329
		Negative	53
Anniversary	129	Positive	107
		Negative	22

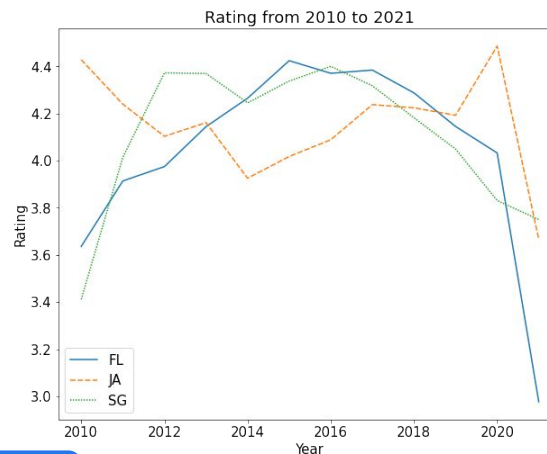
To identify broad customer segments, we counted the number of documents containing each of these words.

* For the customer table, the number does not sum up to the total because some words may appear together in 1 sentence.

Exploratory Data Analysis



- All three branches' review number began to decline in 2016, and only Florida still got hundred of reviews in 2021.
- Overall, the review number of Florida is far more than the other branches during these years.



- All three branches' rating level had a falling trend in recent years.
- Overall, the sharp decline in Florida is the most notable, and reached the lowest rating.

We decided to focus on the reviews in Florida

Text Preprocessing

General Cleaning

Combined title and review text;
Used textacy preprocessing package to remove Punctuation, Numbers, Currency_symbols, White Space

Lemmatization

Conducted Lemmatization to effectively group together inflected forms of words

01



04



02



03



Stop Words Modification

Used nltk to select the most commonly used stopwords;
Modified Stopwords by removing negative emotional words such as “no”, “don”, etc.

Stop Words Application

Applied the selected stopwords and removed them from the reviews

General Cleaning – Regex

Universal studio & Universal Park ...
Disney Park & Disney World...

Park

_UNIVERSAL_STUDIO_
DISNEY

Express Pass & Fast Ticket ...
Park Pass & Park Tickets...

Ticket

_EXPRESS_PASS_
_PARK_PASS_

Roller Coaster & Ride ...
Harry Potter & HarryPotter...
Hollywood Rip Ride & Rip Rocket...

Rides

Ride
_HARRY_POTTER_
_RIP_RIDE_ROCKIT_

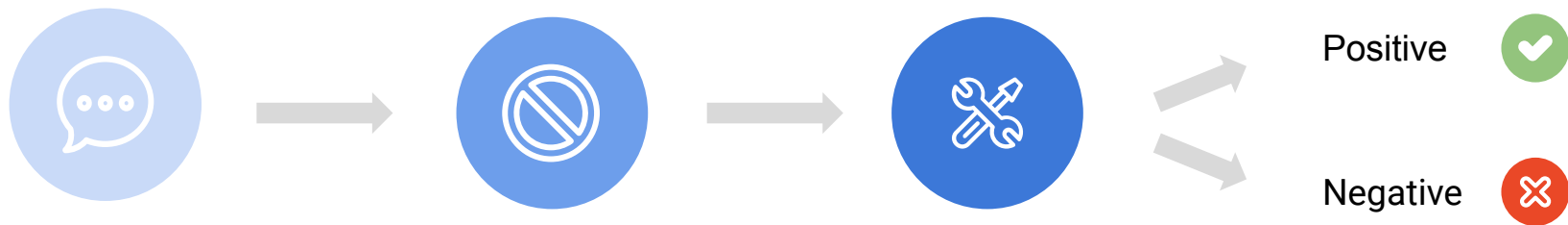




Modeling

1. N Gram Language Modeling
2. Topic Modeling
3. Feature Importance Modeling
(Random Forest)

N Gram Modeling



Token Pattern

- Size: 3 words per token
- Pattern: Words with 2+ alphabets

Thresholds

- Min_df: More than 0.1% of documents
- Max_features: Top 20 tokens to draw word clouds

Vectorizer

- Use TF-IDF Vectorizer

Sentiment Tagging

- Negative: Rating from 1 to 3
- Positive: Rating from 4 to 5

Word Cloud

Positive Sentiment Word Phrase

[illegible]

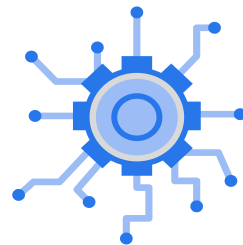
The positive sentiment trigrams show that customers are usually happy with single rider line, horror show and hard rock hotel / cafe.

Negative Sentiment Word Phrase

A word cloud on a light blue background, featuring various phrases of negative feedback. The most prominent words are "don't waste time", "never go back", "not go back", "not worth money", "waste time money", "park one day", "single rider line", "wait long line", "wait time minute", "wait line hour", "wait line minute", "not worth wait", "not worth price", "would not recommend", "go guest service", "win go back", "per person per", "make feel like", and "don't waste money". The words are arranged in a dense, overlapping manner, with some words appearing in a larger font size than others.

Conversely, guest service and long wait time are always things tourists complain about.

Topic Modeling for Positive Sentiment



Input Data

After several tests, we obtained the best results when n (size of ngram) is equal to 4 and topic number is equal to 3.



Output

Topic 1 - Single Rider Line

- use single rider line 96.8%
- “ Since most rides have single rider lines, we used those and were able to save ourselves a bunch of hours of waiting.”

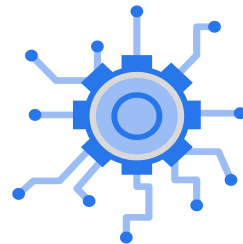
Topic 2 - Hard Rock Hotel

- stay hard rock hotel 93.5%
- stay one site hotel 1.7%
- “ We stayed at Hard Rock Hotel and due to this had express passes for all the ride and attractions, which is a massive bonus.”

Topic 3 - Rides

- men black alien attack 84.0%
- wizarding world diagonal alley 9.3%
- “ The last ride we rode before dinner was Men In Black Alien Attack. It pits you in a battle between invaders from outer space with plenty of humor, it is the best of its kind.”

Topic Modeling for Negative Sentiment



Input Data

After several tests, we obtained the best results when n (size of ngram) is equal to 3 and topic number is equal to 2.



Output

Topic 1 - Waste Money

- not worth money 25.0%
 - save money go 1.5%
 - don't waste money 1.1%
-
- "Way too expensive for what you get better if it were half the price."
 - "Waste of money, too expensive. I found LA and Singapore so much better."

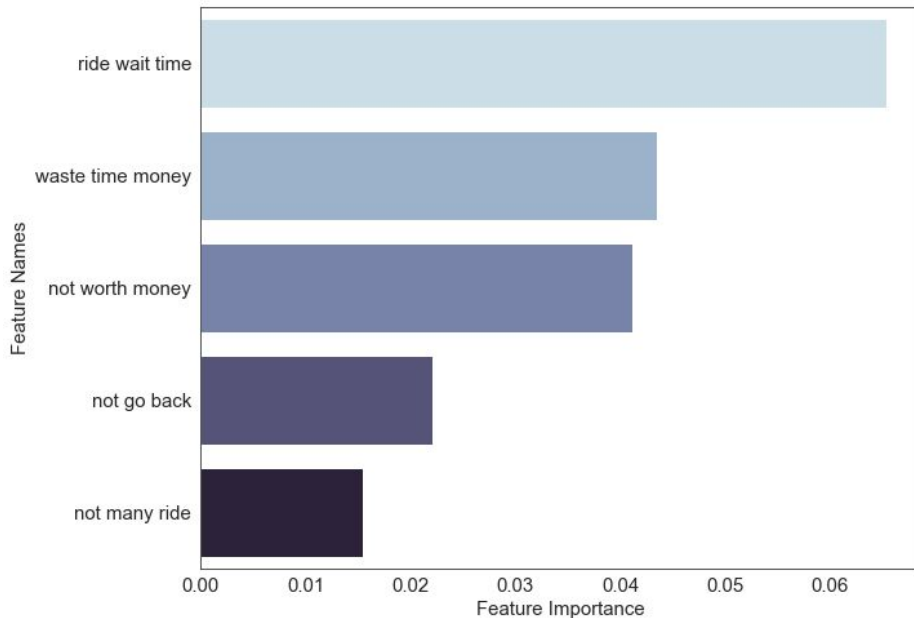
Topic 2 - Waste Time

- waste time money 13.1%
 - don't waste time 11.3%
 - waste time go 1.2%
-
- "Very long queues for not so funny attraction."
 - "Otherwise you're waiting 50+ minutes for one attraction. Considering November not a high season, waiting time then might be even worse."

Feature Importance for the Whole Dataset

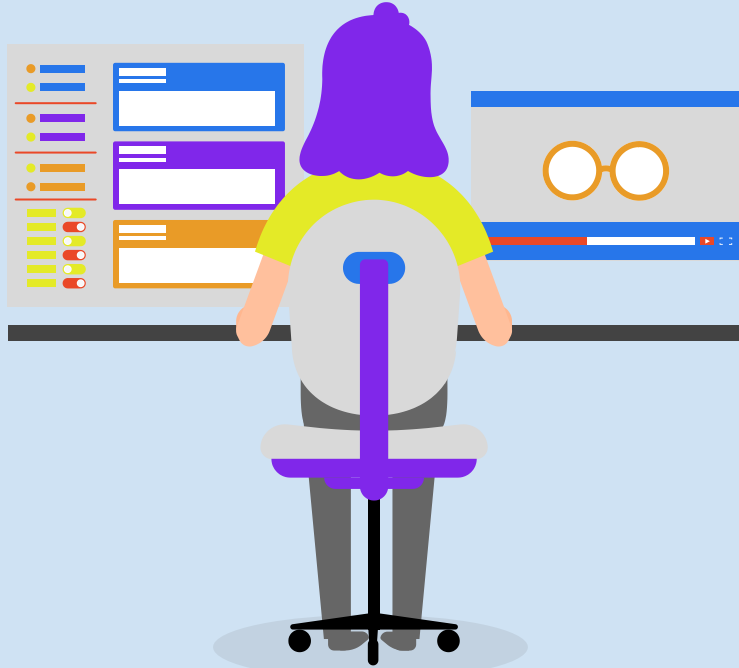
Random Forest

Top important features



As we can see from the left barplot the top factors that best separate positive/negative reviews are:

- Ride **wait time**
- Not worth **money**
- Not many **ride**



Business Implication

1. Recommendation regarding Positive Reviews
2. Recommendation regarding Negative Reviews

Recommendation

Positive

Single Rider

Further promote single riders' line, by encouraging people to volunteer to fill in any ride spots left empty by the larger groups, to reduce the queue time, and improve queue efficiency.

01

Hard Rock Hotel

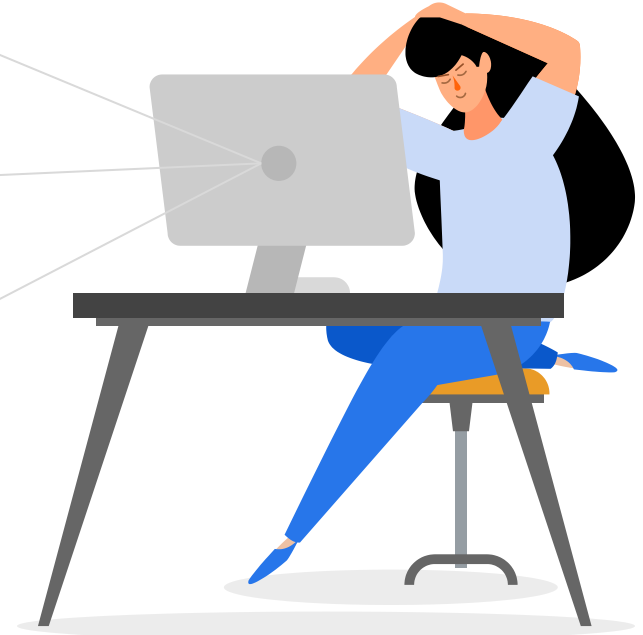
Corporate with more hotels of different levels, offer various vacation packages. For example, like the Hard Rock Hotel, offer Hotel & Free Universal Express vacation packages to meet the needs of different customers. Advertise more on different hotel choices and ticket packages.

02

Horror Nights

Introduce new popular haunted house IP, announce dates and put tickets on sale earlier. Extend the event time to allow more people to participate. Advertise more about what can be expected.

03



Recommendation

Negative

Waste Time

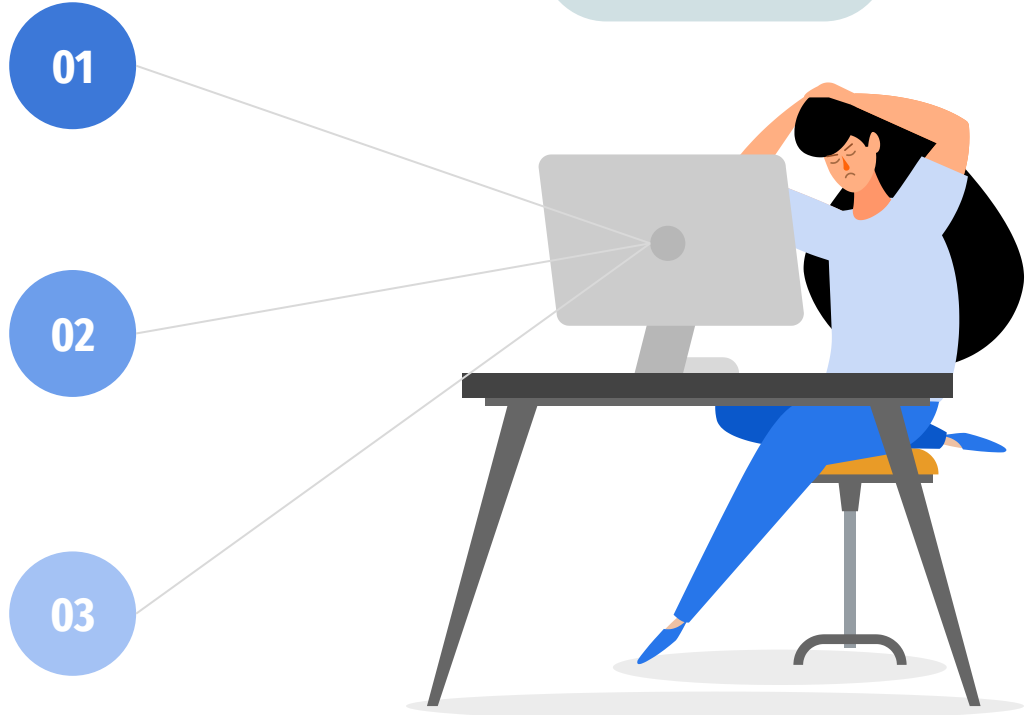
Add fun games and character shows in the theme park; Furnish the queue lane with more interactivity, like adding interacting displays about backgrounds of this ride; Further promote single riders' line; Create express pass for each ride.

Waste Money

Further research on elasticity between ticket price and customer volume, considering cost of other plans, adjust pricing strategies.

Too Few Rides

Conduct surveys to find visitors' preferences and base on results of popular rides and much-complained rides, develop new rides, shows, characters meets and greets of different themes to cater customers of all ages.



Conclusion

Customer Targeting

01

From the customer segment table, we can see reviews with families only account for 9.8% of total reviews. Since families are not the biggest customer segment for Universal Studio, they should target teenagers, and the young group more.

However, there is still possibility that large number of reviews did not mention their family members they were going with.

Business Development

02

Through the recommendations listed in previous pages, there are room for improvements such as adding more rides. However, these investments can take huge costs, and for company's overall profit, further research on cost-benefit analysis should be performed.

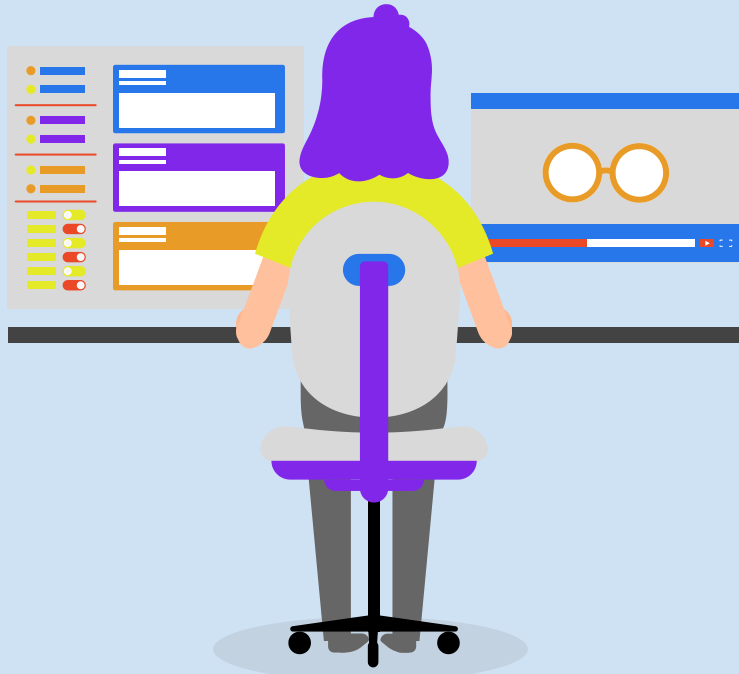
Further Improvements

03

This report is built based on reviews written on Tripadvisor. The result is not as comprehensive because it filtered visitors who do not use Tripadvisor.

To better understand the visitors, more researches and surveys can be done, and for business recommendations, more practical ideas can be proposed after further company and industry analysis.





Thank you