Sentiment Analysis of
DISNEY PARKS
REVIEWS

Metis
Sandra Tran
October 2022

Categorize Reviews

Distribution of Ratings Across Parks

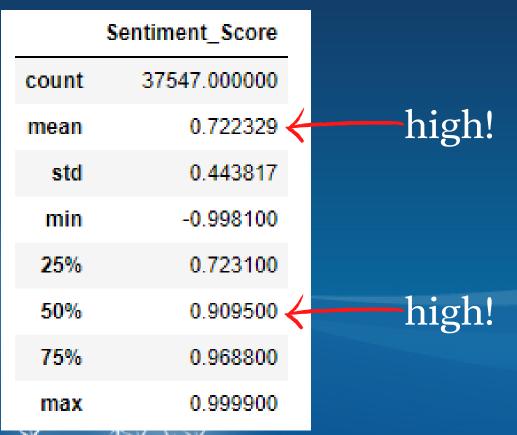- Add category 'Rating_Type'
  - **Remove neutral ratings '3'**
  - 'Positive' (4-, 5-star reviews)
  - 'Negative' (1-, 2-star reviews)
- ~90% - 'Positive'
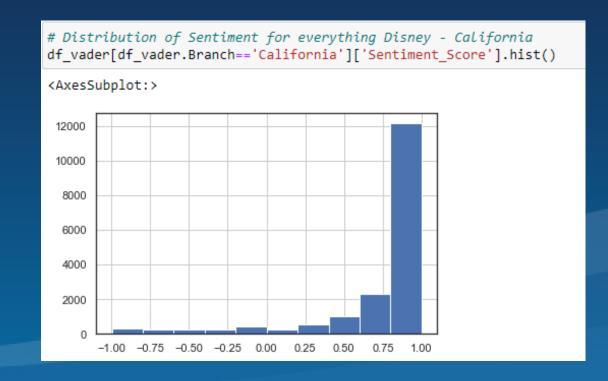- ~10% - 'Negative'

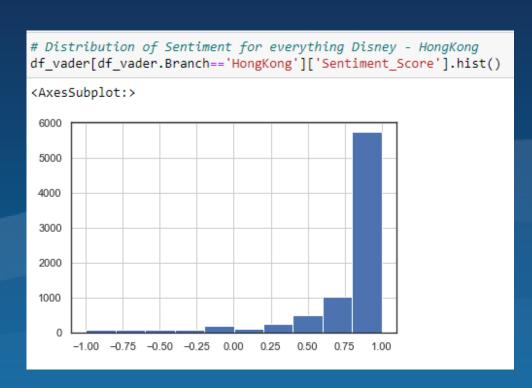| Rating	Reviews	Branch	Rating_Type
4	Great experience... Space Mountain is by far t...	California	Positive
5	Had a wonderful time visiting disneyland Paris...	Paris	Positive

Sentiment Analysis with VADER

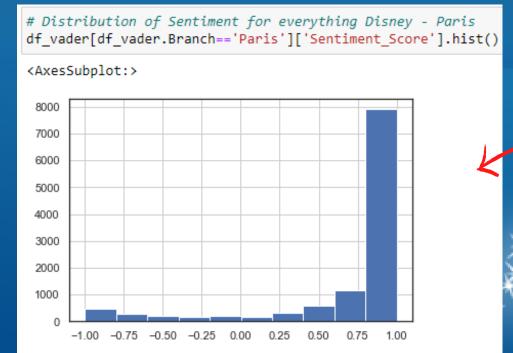Compound score normalized between -1 (most extreme negative) and +1 (most extreme positive)

(All Parks)

	Sentiment_Score
count	37547.000000
mean	0.722329
std	0.443817
min	-0.998100
25%	0.723100
50%	0.909500
75%	0.968800
max	0.999900

high! ← mean 0.722329

high! ← 50% 0.909500


```
# Distribution of Sentiment for everything Disney - California
df_vader[df_vader.Branch=='California']['Sentiment_Score'].hist()

<AxesSubplot:>
```



```
# Distribution of Sentiment for everything Disney - HongKong
df_vader[df_vader.Branch=='HongKong']['Sentiment_Score'].hist()

<AxesSubplot:>
```



```
# Distribution of Sentiment for everything Disney - Paris
df_vader[df_vader.Branch=='Paris']['Sentiment_Score'].hist()

<AxesSubplot:>
```

still positive!

More negative sentiment than neutral!

- Average sentiment (Positive): 0.79329
- Average sentiment (Negative): 0.05847

Interesting...

Reviewer_Location	Reviews	Branch	Rating_Type	Sentiment_Score
United States	What made me mad was we normally leave during the afternoon crowds and come back around 6pm but couldn't. There was a special event (Halloween party) thus we couldn't do that. Also, it was so crowded that we only spent a single day in the park and found other (free) stuff to do around LA instead, like going to the beach. As Disneyland closed early, there were no fastpasses available during the afternoon so we ended standing in huge lines. What a waste of time and we'll not be coming back anytime soon. Disneyland is packing in huge crowds so I'm sure they could care less.	California	Negative	0.9669

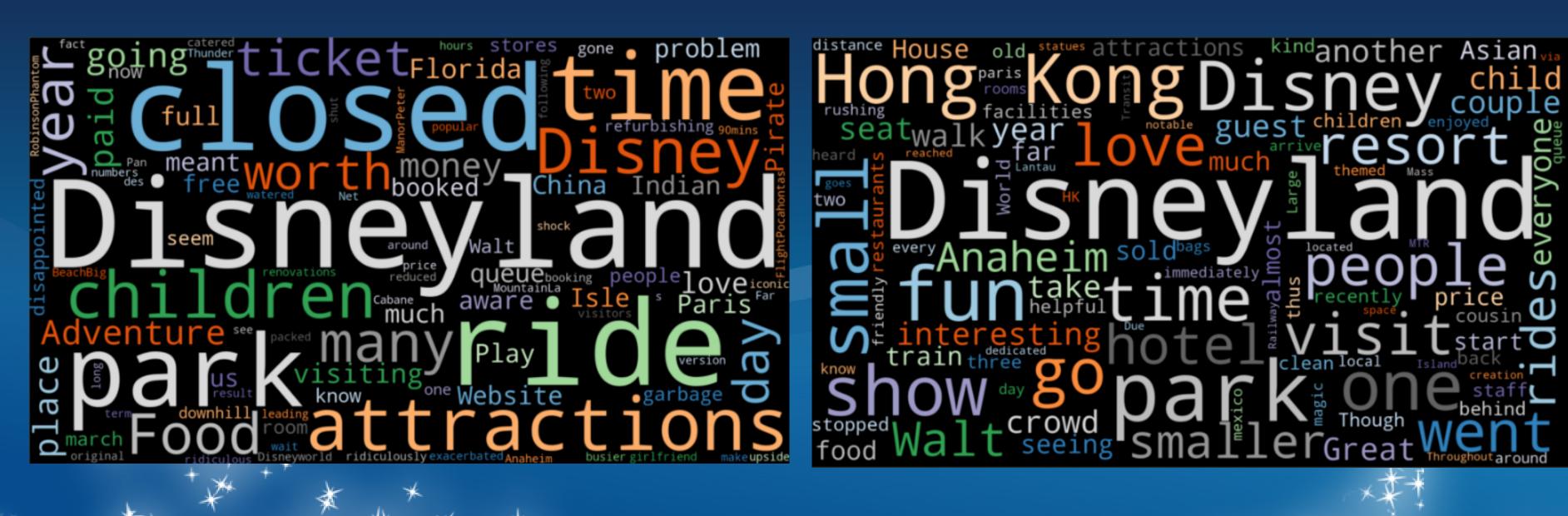| Reviewer_Location	Reviews	Branch	Rating_Type	Sentiment_Score
United States	so many rides for all ages and sizes! The Indiana Jones ride is my fave along with Pirates of the Caribbean and haunted house are great...space mountian is a must and when they reopen the matterhorn can't wait to go back agian been there several times!	California	Positive	-0.1984

- In a fraction of the reviews, the Sentiment_Score did <u>not</u> reflect the Rating_Type

Top 200 Words in Reviews

Top Words in Positive/Negative Reviews
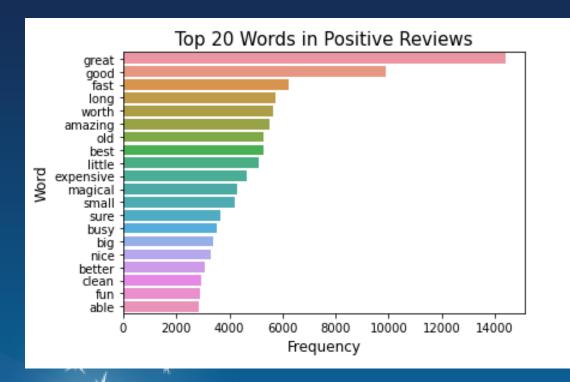
Can you tell?

Top Words in Positive/Negative Reviews

Negative
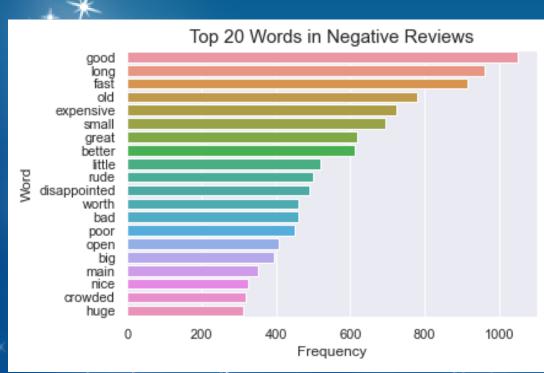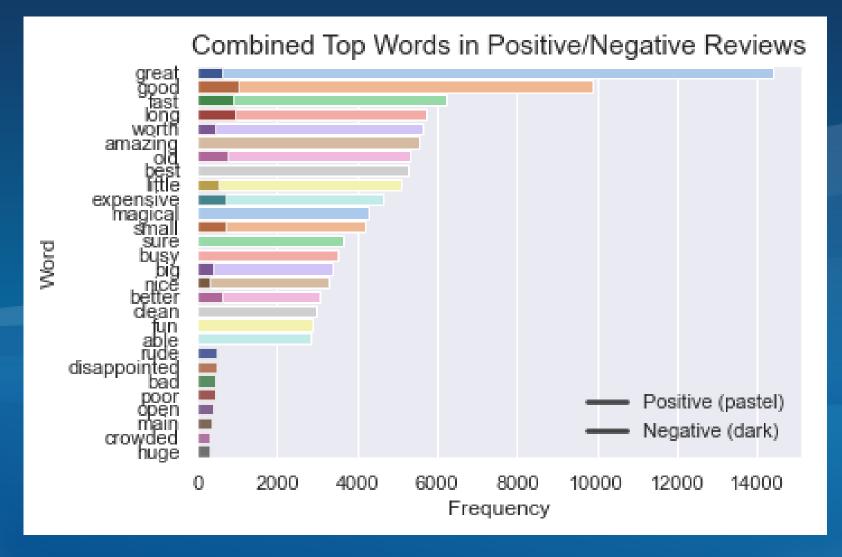
Positive

Top Words in Positive/Negative Reviews

- spAcy to preprocess, tokenize, extract adjectives

Exclusively:
Positive: **'amazing', 'best', 'magical', 'clean', 'fun'**
Negative: **'rude', 'disappointed', 'bad', 'poor', 'crowded'**

Top Words in Positive/Negative Reviews

- Dependency parsing with spaCy
- Finding adjective modifiers for specific nouns

"Food"

```
POSITIVE:
[('own', 312),
 ('fast', 309),
 ('good', 176),
 ('great', 129),
 ('expensive', 87),
 ('outside', 43),
 ('best', 40),
 ('available', 39),
 ('decent', 31),
 ('vegetarian', 31),
 ('indian', 30),
 ('chinese', 29),
 ('better', 28),
 ('nice', 28),
 ('much', 22),
 ('healthy', 21),
 ('asian', 19),
 ('overpriced', 19),
 ('western', 19),
 ('delicious', 18)]
```

```
NEGATIVE:
[('fast', 78),
 ('own', 39),
 ('expensive', 36),
 ('overpriced', 19),
 ('poor', 17),
 ('good', 16),
 ('terrible', 16),
 ('bad', 14),
 ('priced', 12),
 ('awful', 11),
 ('decent', 8),
 ('horrible', 8),
 ('mediocre', 8),
 ('same', 8),
 ('better', 7),
 ('great', 7),
 ('hot', 7),
 ('much', 6),
 ('only', 6),
 ('available', 5)]
```

Positive: **'great', 'vegetarian', 'chinese', 'indian', 'healthy', 'asian', 'western'**

- Reviewers enjoy the variety of food choices

Negative: **'expensive', 'overpriced', 'poor', 'terrible', 'bad', 'awful', 'mediocre'**

- Reviewers dislike the price and/or taste of the food

Topic Modeling

- CountVectorizer vs TfdifVectorizer
- LSA vs NMF vs CorEx
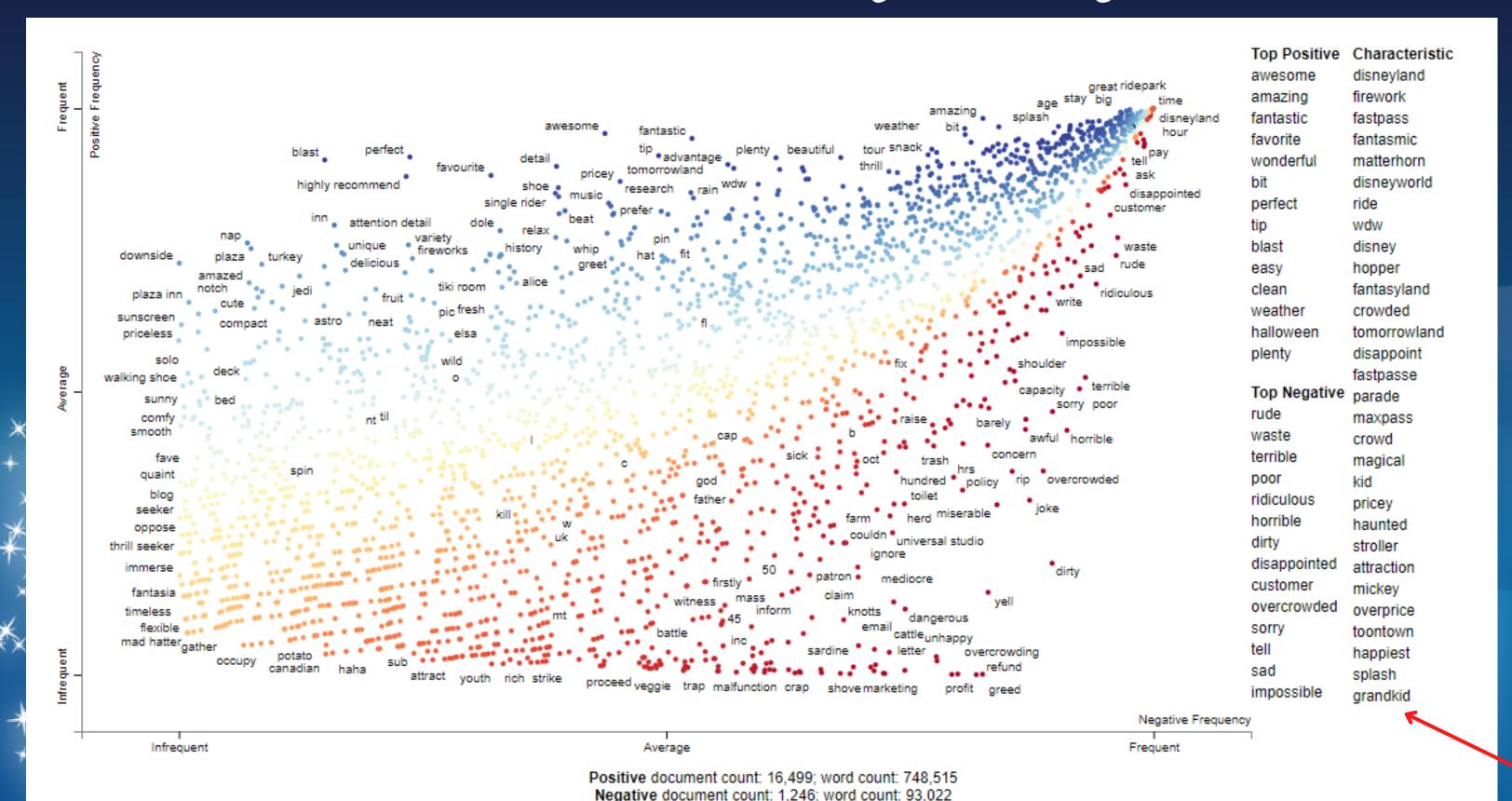- Chosen: CountVectorizer + CorEx with anchored words

```python
# Topics with anchored words
topics2 = topic_model2.get_topics()

for n,topic in enumerate(topics2):
    topic_words,_,_ = zip(*topic)
    print(f'{n}: {", ".join(topic_words)}')
```

```
0: disney, california, world, walt, downtown, fan, disneyland, fans, southern, original
1: disney, paris, florida, disneyland, orlando, dreams, rer, comparison, typical, states
2: disney, hong_kong, land, disneyland, smaller, compared, mtr, version, compare, scale
3: food, expensive, disney, drinks, drink, quality, prices, options, bring, overpriced
4: parade, fireworks, disney, night, watch, paint, miss, lion_king, floats, amazing
5: ride, park, day, just, time, people, did, rides, hotel, got
```

California: **"Original"**
Paris: **"Typical" - same as "States"**
Hong Kong: **"Smaller"**
Food: **"Expensive"**
Shows: **"Amazing"**

Scattertext - Disney California

Positive document count: 16,499; word count: 748,515
Negative document count: 1,246; word count: 93,022

Are Disney Parks *Magical* ?

YES!!

But..

1. food is expensive ⟶ bring your own snacks/food
2. parks are crowded ⟶ go on weekdays, non-holidays

Future Works

- Compare different Sentiment Analyzers (VADER, Flair, TextBlob, etc.)
  - Explore discrepancy between Rating and Sentiment Score
- Analyze more updated dataset
  - Include ALL parks (Tokyo, Shanghai)