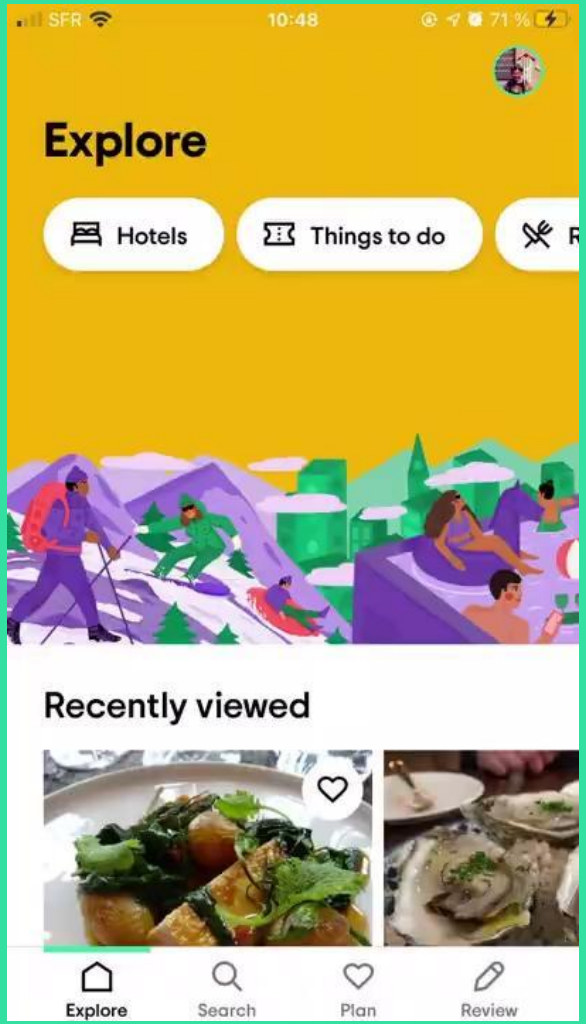
NLP Analysis on Restaurant Reviews

Tripadvisor®

NLP-filter

Anna Shin

IRONHACK
DAFT NOV21
Feb 11th, Paris



SFR 11:00 74 %

Septime

1 010 reviews 0.3 mi

#1,004 of 15,768 Restaurants in Paris

$$$$ • French • European • Vegetarian Friendly

Visit website Call Menu

Closed now

Opens 7:30 PM - 10:00 PM

About

Awards

Travelers' Choice 2021

MICHELIN

Dietary options

Vegetarian Friendly

Explore Search Plan Review

SFR 11:00 74 %

Septime

About Review Control

1,010 reviews

Excellent	580
Very good	232
Average	113
Poor	48
Terrible	37

Kevin B

5 contributions

Affordable and Comfortable for an Excellent Meal

Sep 2021 • Couples

Excellent meal and service for Americans in Paris. Went for lunch and did the tasting menu and wine pairing. Everything was gre...

Read more

Explore Search Plan Review

Quantitative scoring

- Michelin-stars
- Macarons, toques
- Customer rating

Qualitative feedback

- Food critics blogs
- Word-of-mouth
- Customer reviews

How do you score customer reviews?

Data Collection

download Octoparse

All reviews for one restaurant:

> 372 reviews

With 8 features:

> review title
> review text
> review author
> count of author's total reviews
> date of visit
> date of rating
> likes
> rating

Roadmap:

1. Clean, trim, process text

Lemmatization

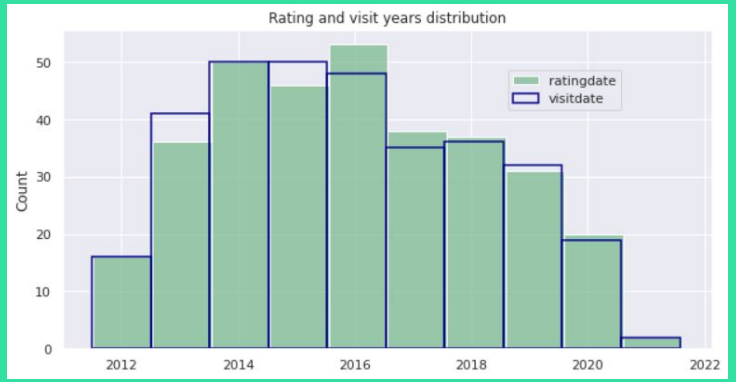
2. Compute polarity score on reviews

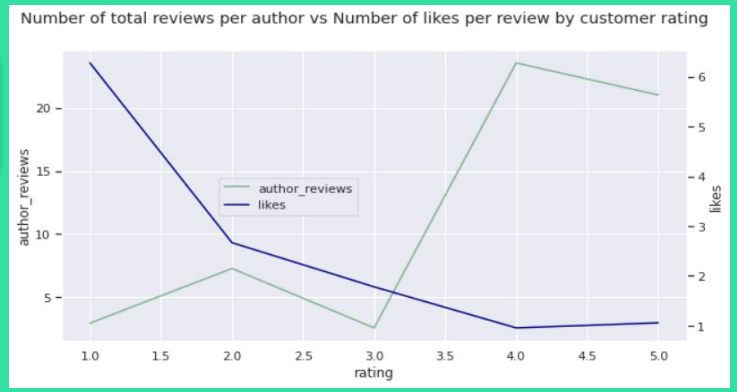
Sentiment Analysis

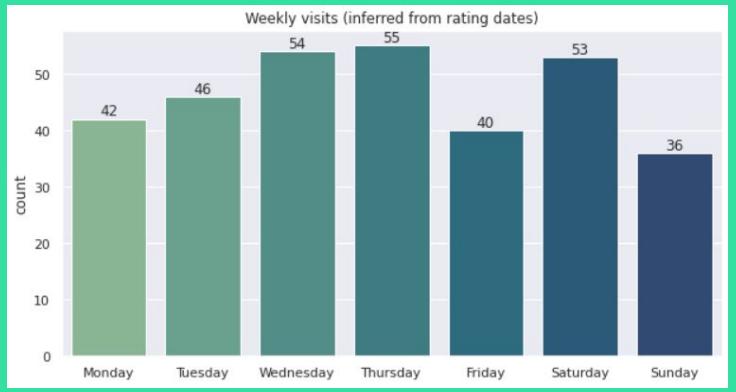
3. Display words that describe the place best

Word Cloud

EDA: explore the data







1. Lemmatization

import Spacy

Expand contractions

> “dunno” ➜ “do not know”

Lowercase text

> “SepTime” ➜ “septime”

Clean digits and special characters

> “hi-5!” ➜ “ “
> “ “ ➜ “ “

Remove stop words

> “I really loved this restaurant” ➜ “really loved”

Lemmatize

> “perfected tastings” ➜ “perfect tasting”

1. Lemmatization

Raw

'We had a wonderful experience. Great taste, great presentation. All the textures, flavors since the starter until the desert. Great chef',

Cleaned

'we had a wonderful experience great taste great presentation all the textures flavors since the starter until the desert great chef'

Lemmatized

'wonderful experience great taste great presentation texture flavor starter desert great chef',

```
import TextBlob
.sentiment.polarity
```

2. Sentiment Analysis

Rating = 5/5

"excellent food,totally fresh,simple ingredients mixed to haute cuisine,great,caring service,interesting wine and beverage-list,one of the very best restaurants in paris,yes,it's woth the sometimes troublesome reservation!"

Polarity = 0.85

Rating = 1/5

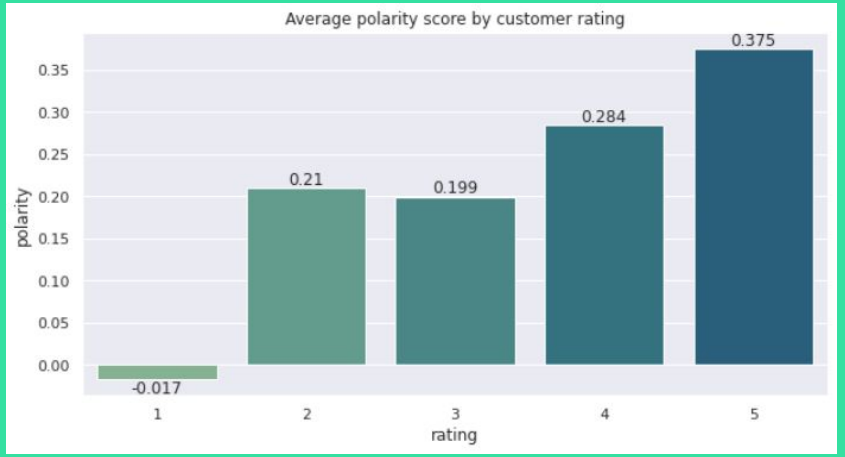
'Very little plates, pretentious dishes, pretentious cook, everybody left the restaurant being hungry (if not for the bread they keep serving to fill us up).'

Polarity score = -0.262

2. Sentiment Analysis

> Polarized extremes

> Ratings 2 & 3 share similar polarities



> Polarity consistent with rating

> Outlier identified and dropped

3. Word Cloud

```
import CountVectorizer
```

20 most occurring words and their frequencies

food	271	not	139
good	253	table	127
wine	209	great	121
menu	188	time	109
course	178	meal	109
reservation	160	staff	103
dish	157	place	98
paris	149	experience	92
service	149	like	84
lunch	147	taste	81

3. Word Cloud

Highest rating scores

Highest polarity scores

3. Word Cloud

Lowest rating scores

Lowest polarity scores

With this additional polarity-based wordcloud feature, Tripadvisor would be able to provide:

1. **Users** with a grouped, concise and customized overview of the restaurants's reviews

2. **Restaurants** with a visual tool to take corrective actions or to target their communication


Q&A
reservation
excellent
experience
good
food
book
place
dish
chef
delicious
service
serve
way
lunch
great
time
think
eat
enjoy
know
hard
nice
fresh
visit
worth
paris
staff
want
dinner
meal
find
menu
french
fish
taste
plate
love
ask
go
able
Anna Shin