Yelp Data Exploration and Analysis

ECE 143 FA20 Team 1
Bingxu(Vince) Chen, Ziwen Li, Zhimin Lin and Gaotong Wu
Presented by Gaotong Wu

Nov. 29, 2020
University of California, San Diego

The Yelp Dataset and Subset Selection

- Businesses and users information since Dec. 2004.

- 5.2 million user reviews and ratings from over 1.96 million users.

- 174,000 businesses that mostly spans 11 metropolitan areas (including some outside the US).

- In our project:
  - Consider only <u>food and beverage</u> providers (restaurants, bars, coffee shops, etc.)
  - Consider only businesses <u>in the US</u> that are <u>still open</u>.
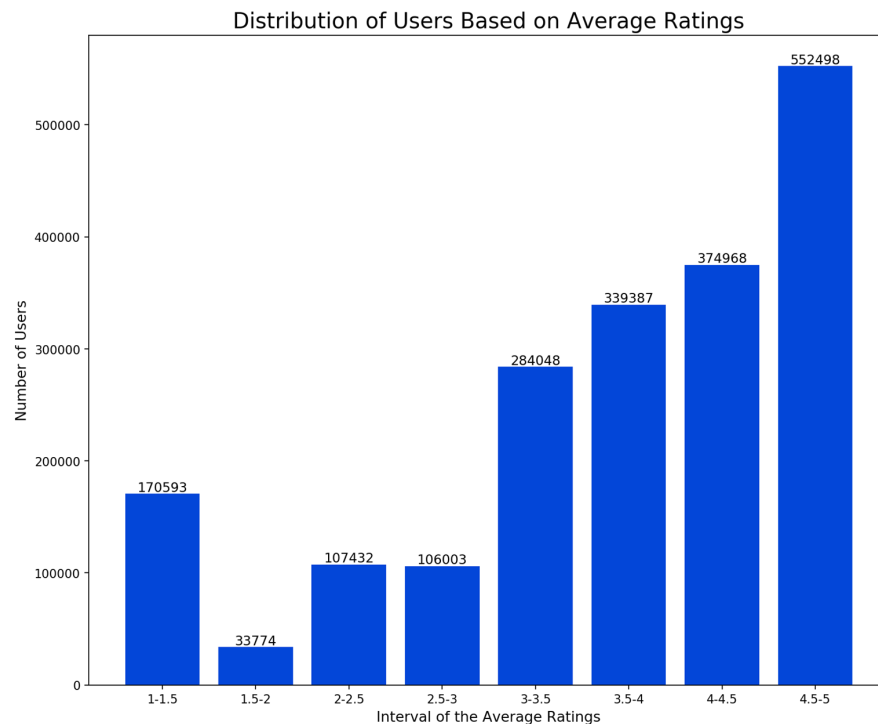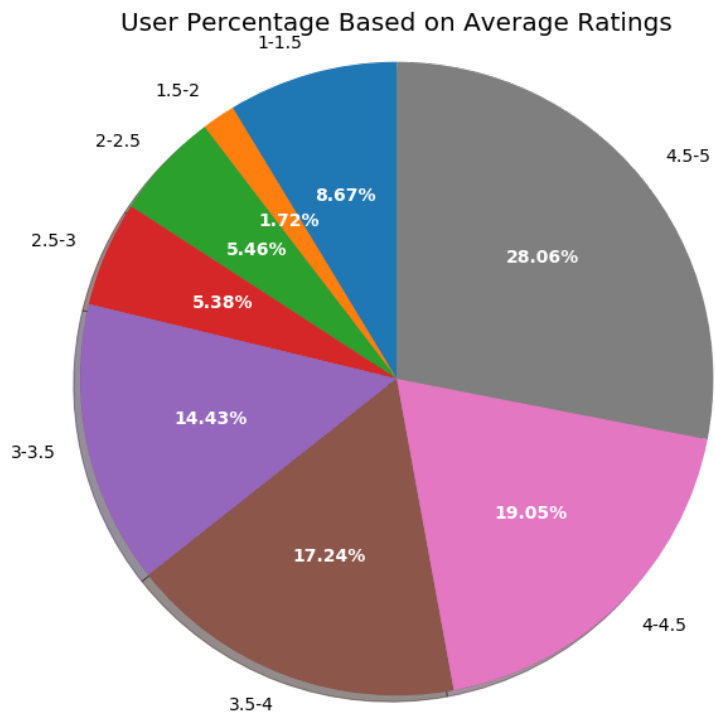 - Result: the number of businesses reduced to 22,730.

Methodology

- Statistics: grouping and counting based on different criteria, e.g. ratings, types of cuisines, etc.

- Word cloud: group review texts based on rating levels and then remove words (unigrams) with high frequencies. Then feed the texts into WordCloud library.

- Sentiment analysis: feed all review texts for a single restaurant into the analysis model one restaurant at a time.
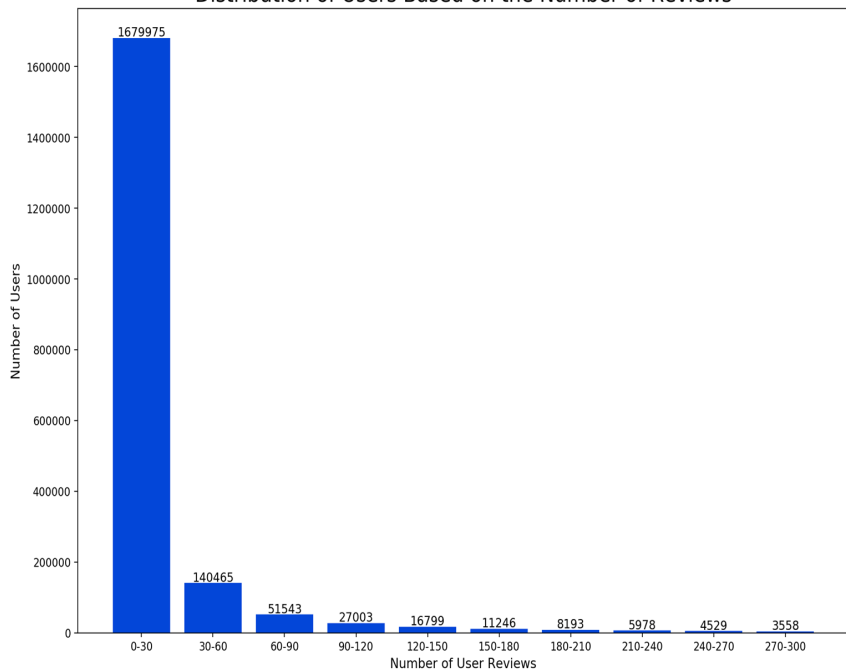
Users

# Most users tend to give high ratings

User Percentage Based on Average Ratings

- 1-1.5: 8.67%
- 1.5-2: 1.72%
- 2-2.5: 5.46%
- 2.5-3: 5.38%
- 3-3.5: 14.43%
- 3.5-4: 17.24%
- 4-4.5: 19.05%
- 4.5-5: 28.06%

Distribution of Users Based on Average Ratings

- 1-1.5: 170593
- 1.5-2: 33774
- 2-2.5: 107432
- 2.5-3: 106003
- 3-3.5: 284048
- 3.5-4: 339387
- 4-4.5: 374968
- 4.5-5: 552498

Most users write few reviews

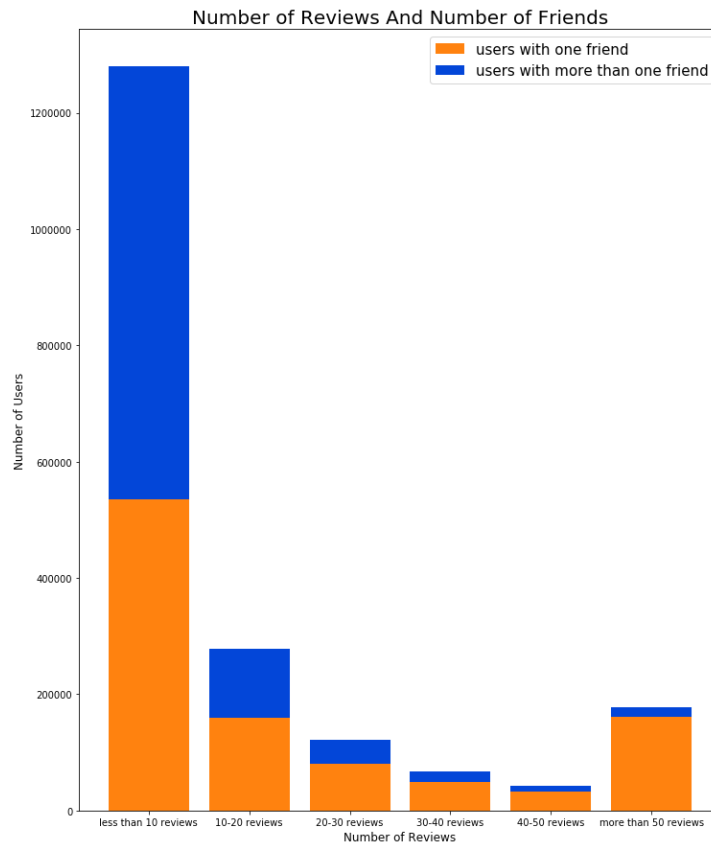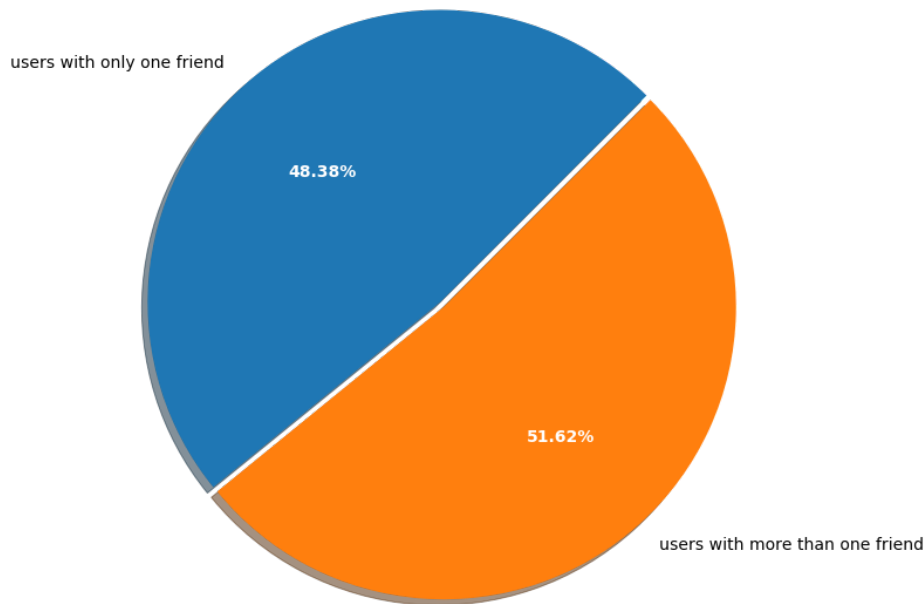Distribution of Users Based on the Number of Reviews

Distribution of Users with < 60 Reviews
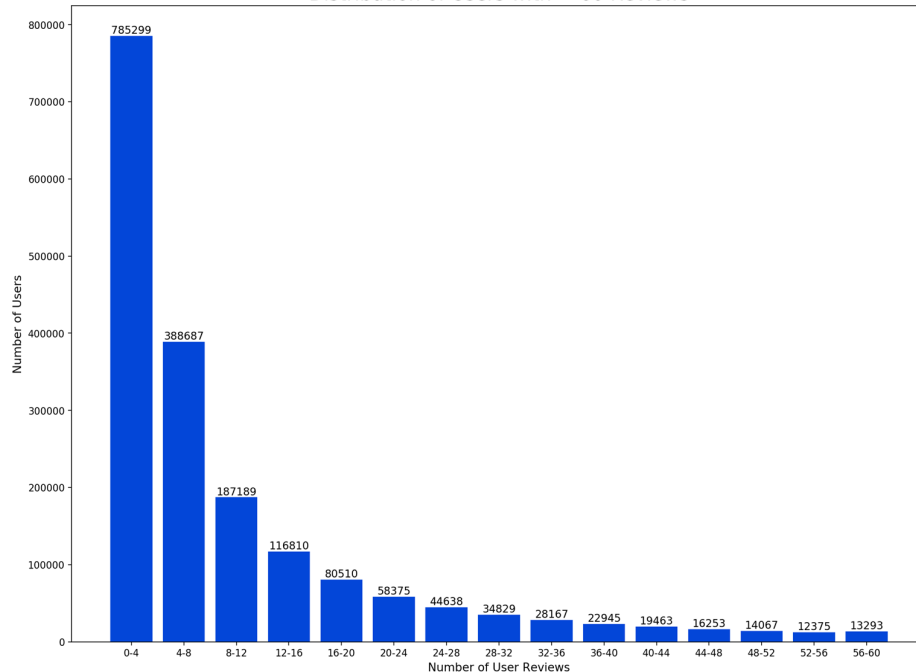
Almost half of the users have only 1 friend
(on Yelp)

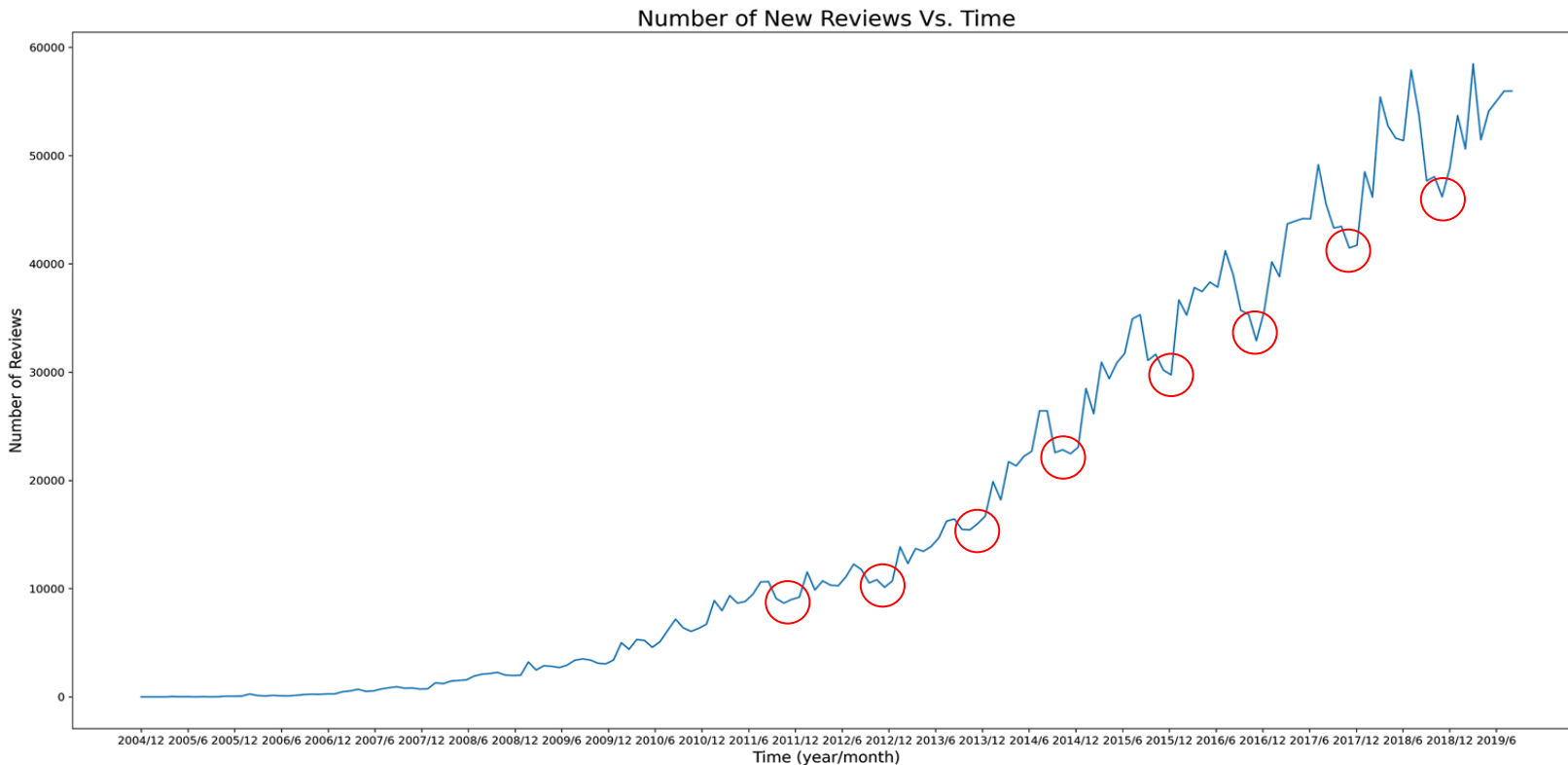Percentage of Users with One Friend and with Multiple Friends

users with only one friend

48.38%

51.62%

users with more than one friend

Number of Reviews And Number of Friends

- users with one friend
- users with more than one friend

Number of Users

1200000

1000000

800000

600000

400000

200000

0

less than 10 reviews   10-20 reviews   20-30 reviews   30-40 reviews   40-50 reviews   more than 50 reviews

Number of Reviews

Reviews and Ratings

Fewest new reviews in December. Always.

Number of New Reviews Vs. Time

Word Clouds:
What do customers say about...

| Businesses w/ 1 star rating | Businesses w/ 2 stars rating | Businesses w/ 3 stars rating |

Word Clouds:
What do customers say about...

Businesses w/ 4 stars rating

Businesses w/ 5 stars rating
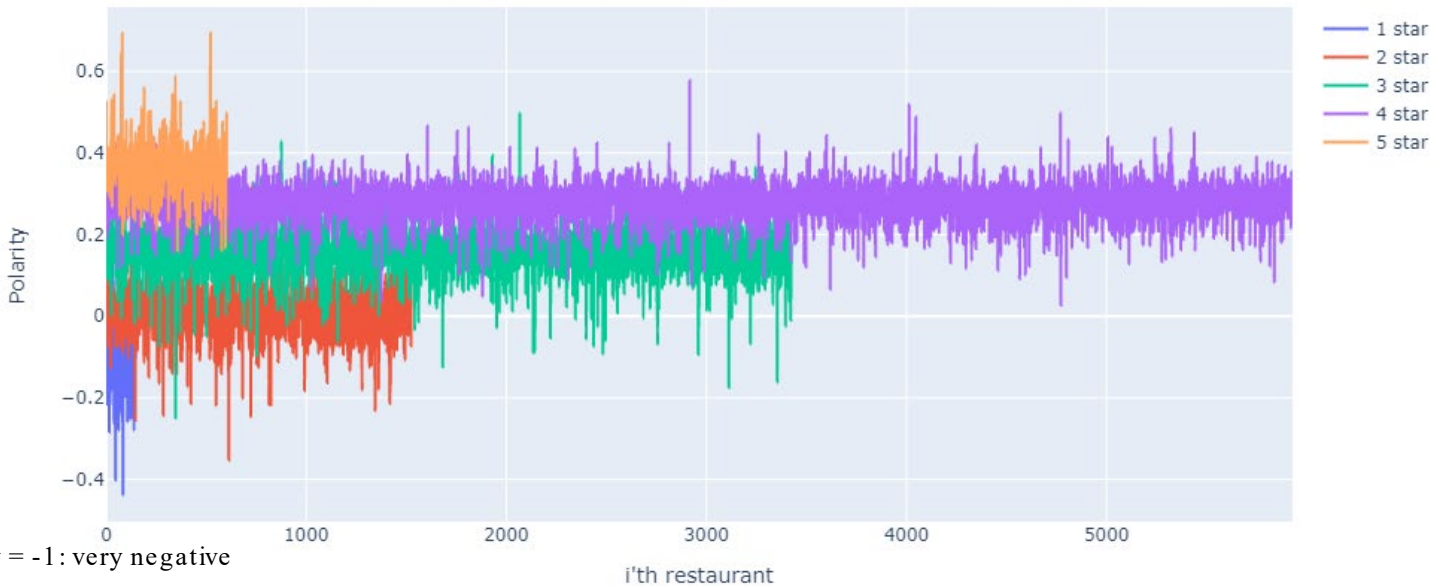
Sentiment Analysis: happy vs. unhappy

Polarity of Reviews with Different Ratings

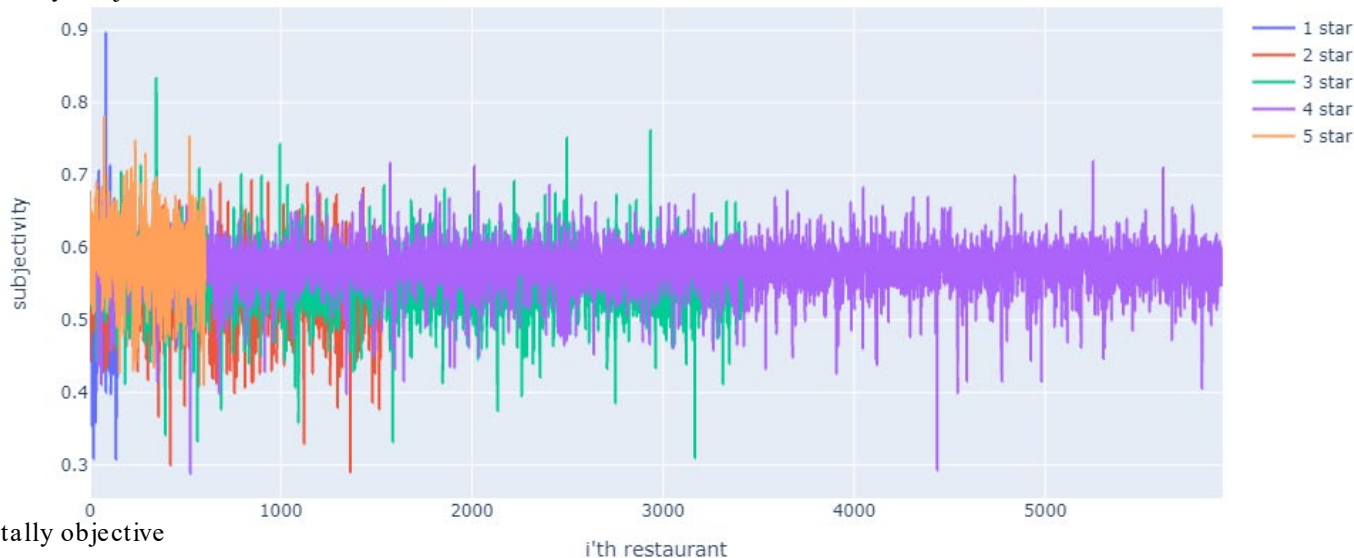Polarity = 1: very positive

Polarity = -1: very negative

Results (average): 1 star: -0.108; 2 stars: 0.011; 3 stars: 0.158; 4 stars: 0.280; 5 stars: 0.363

Sentiment Analysis: are reviews entirely subjective?

Subjectivity of Reviews with Different Ratings

Subjectivity = 1: totally subjective

Subjectivity = 0: totally objective

Results (average): 1 star: 0.56; 2 stars: 0.58; 3 stars: 0.55; 4 stars: 0.54; 5 stars: 0.59

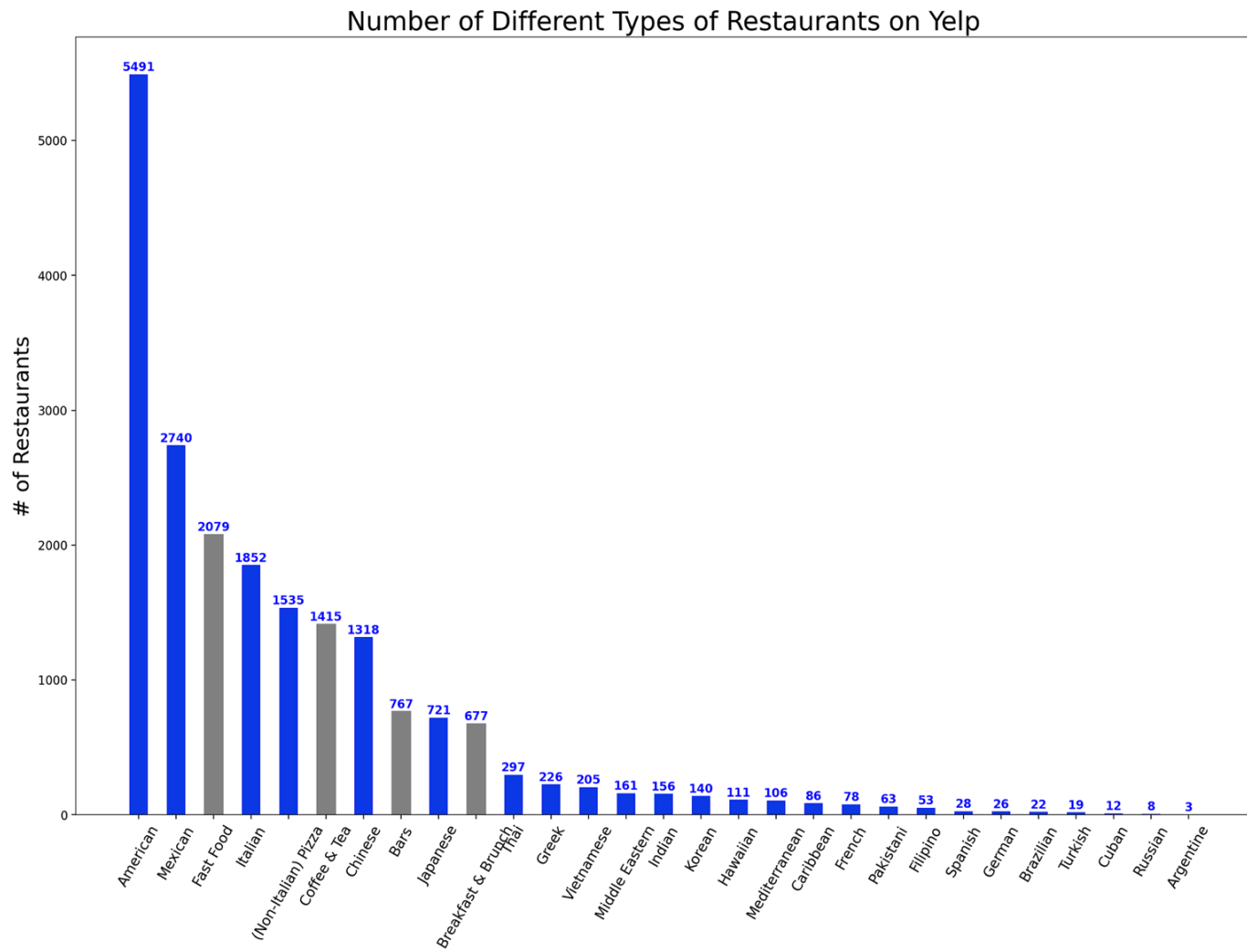# Some businesses receive higher ratings

- Those that are bad for kids generally rated higher.

- Those that are good for groups generally rated higher.

- People like quiet places.

Average Ratings for Restaurants with Certain Features

Cuisines

Overview

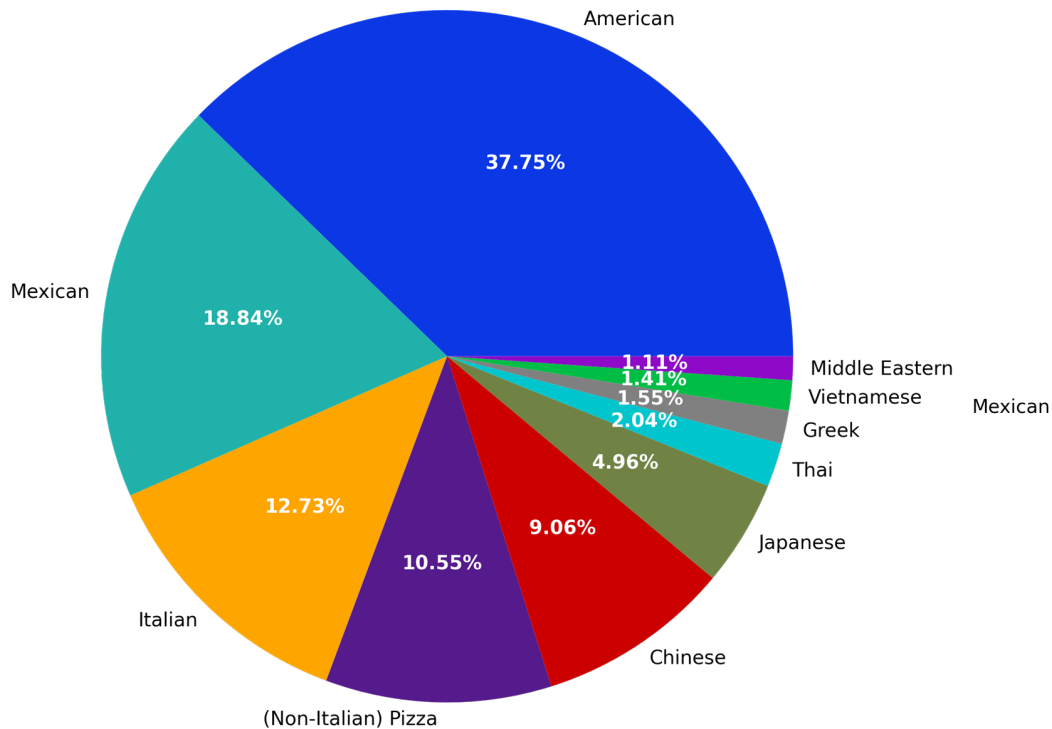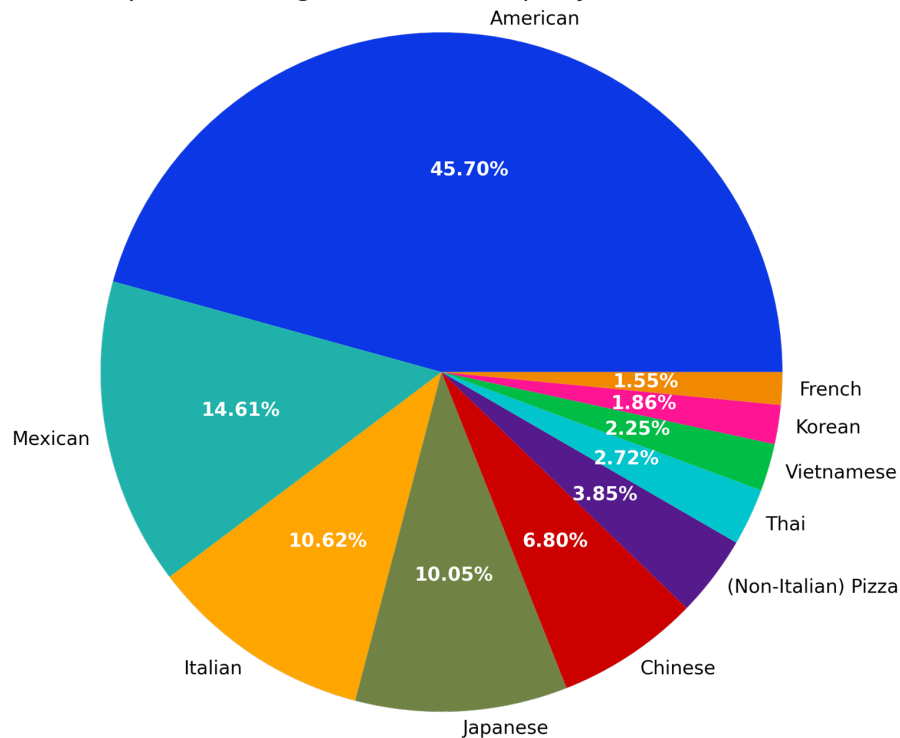Number of Different Types of Restaurants on Yelp
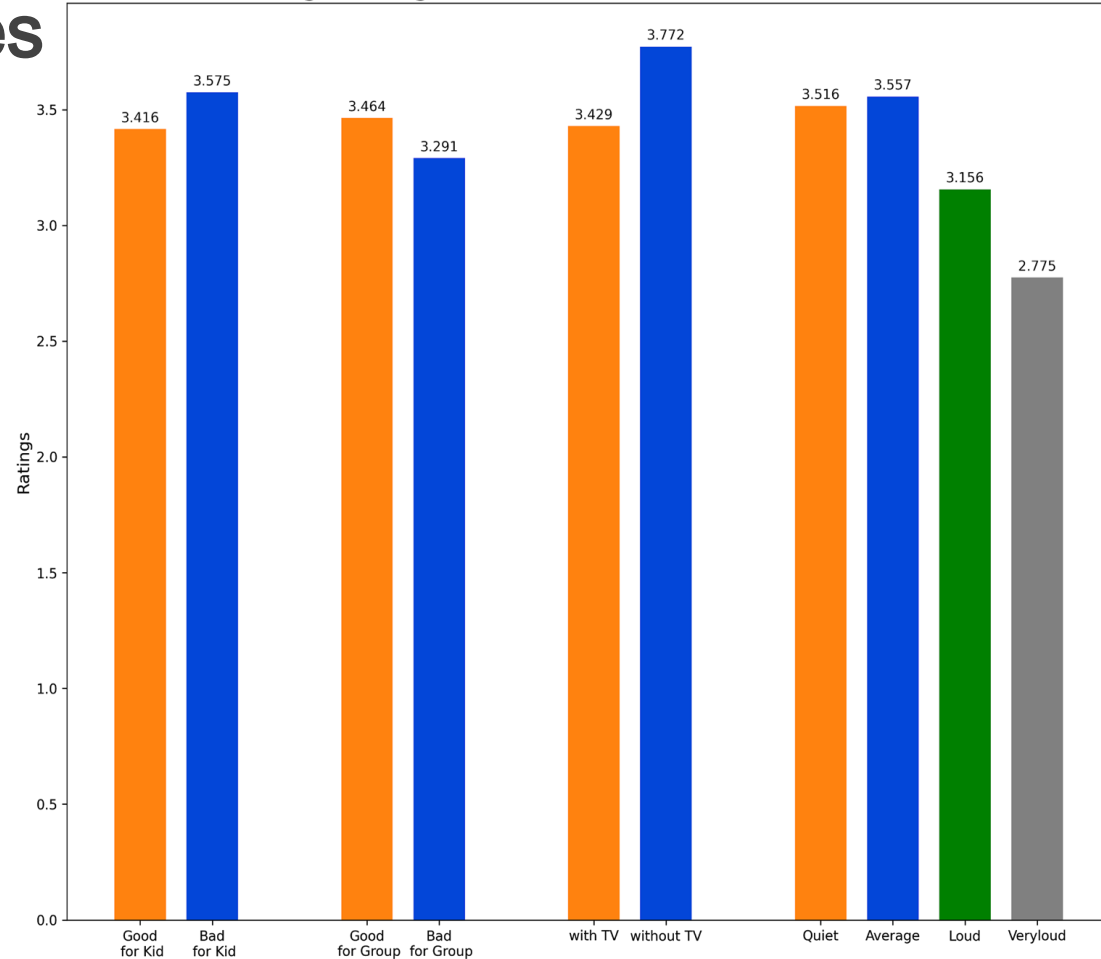
What is trending on Yelp?

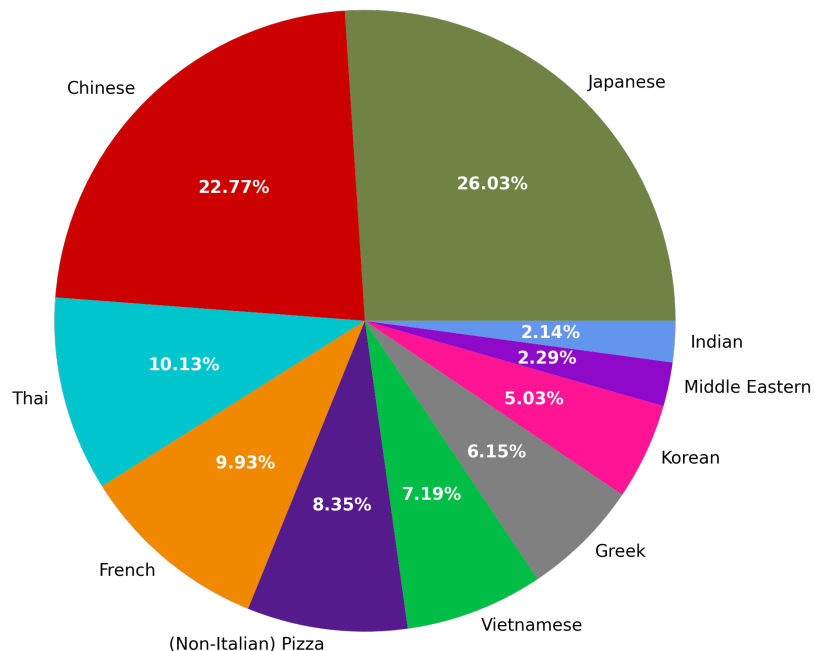Top 10 Trending Cuisines on Yelp (by # of restaurants)
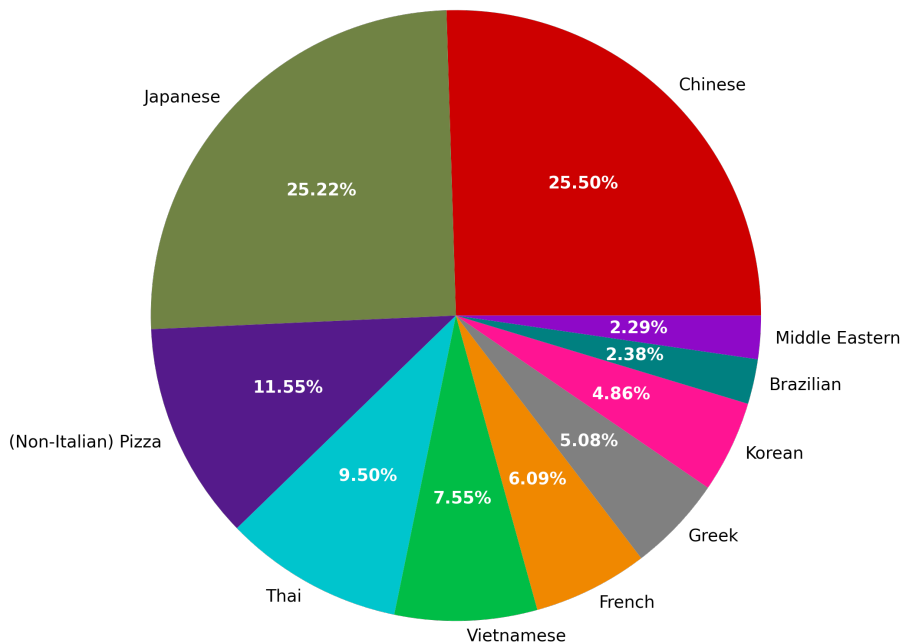
Top 10 Trending Cuisines on Yelp (by # of reviews)

(Active) Users' tastes change over time

Top 10 Trending Cuisines on Yelp in 2013 (by # of reviews)
(Excluding American, Mexican and Italian)

Top 10 Trending Cuisines on Yelp in 2019 (by # of reviews)
(Excluding American, Mexican and Italian)

Thank you!