MARCH 09, 2021

Dataset Analysis

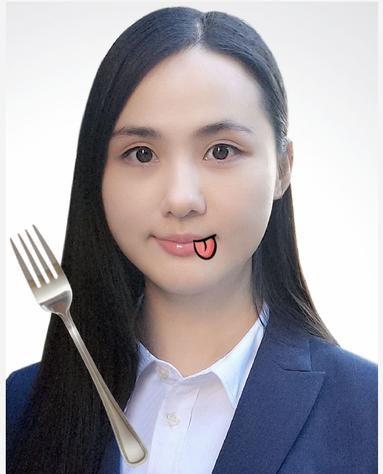
Team 7: Ying Zhang, Kristy Guo, Jiazheng Li, Barbara Liang

Meet the Team

Kristy

Barbara

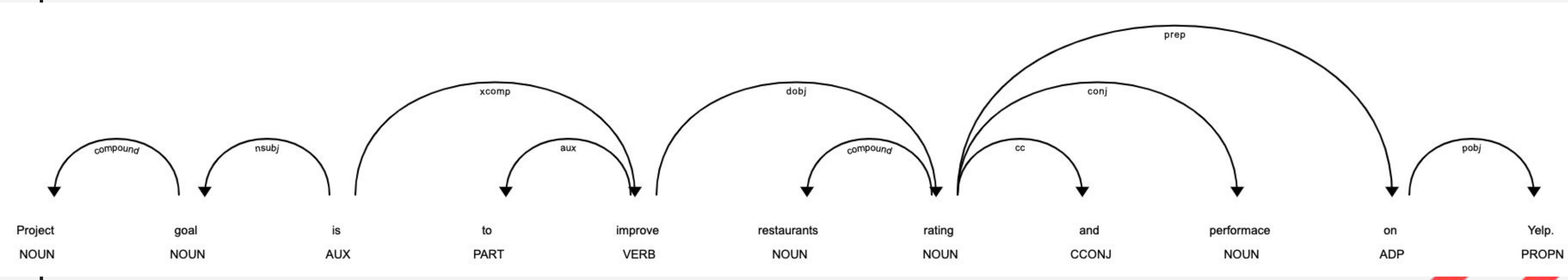
Jiazheng



Ying

Business Goal

- Predict star ratings
- Provide profound suggestions



Datasets

Source:
https://www.kaggle.com/yelp-dataset/yelp-dataset

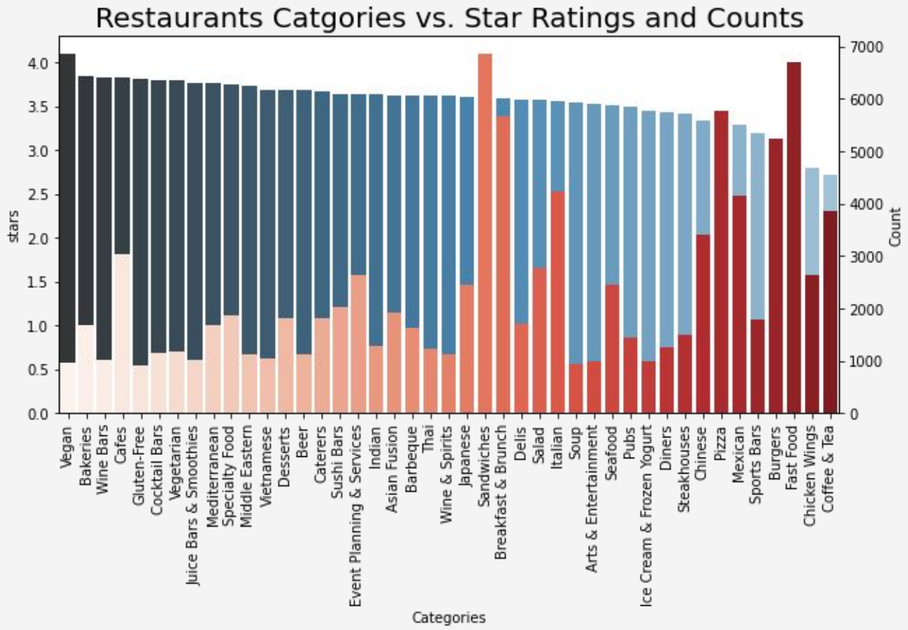
Data Cleaning:

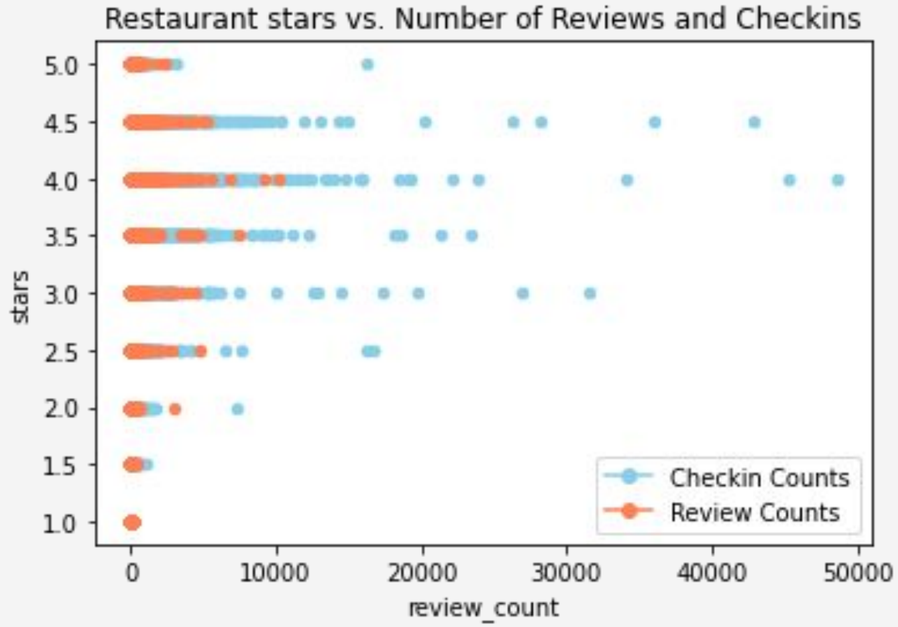
- Businesses containing “restaurant”
  - 49,947 restaurants information
- 56 features
  - text
  - attributes

Exploratory Analysis

Exploratory Data Analysis

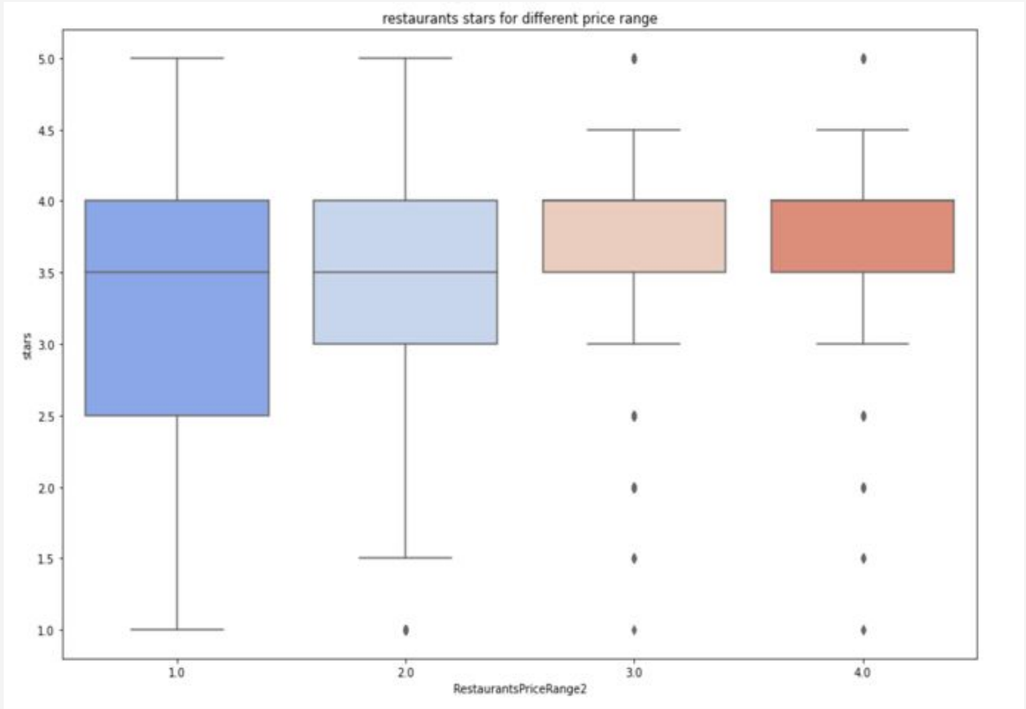
Restaurant stars ratings are compared with business attributes to observe what affects ratings





- Fast food and sandwiches categories have top two reviews counts
- Restaurants around 3.5 - 4.5 stars records the highest review count

Exploratory Data Analysis



- The average review stars has a positive linear relationship with the restaurants stars
- Restaurants with higher price range would like to have higher stars.

Predicting Stars

Supervised Machine Learning Models

```
f1 <- as.formula(stars ~ review_count + RestaurantsTakeOut +
BusinessAcceptsCreditCards + NoiseLevel + RestaurantsPriceRange2 + Alcohol + trendy
+ classy + lunch + dinner + brunch + RestaurantsAttire + GoodForKids +
RestaurantsReservations + RestaurantsGoodForGroups + HasTV + BikeParking +
RestaurantsDelivery + OutdoorSeating + WiFi + Caters)
```

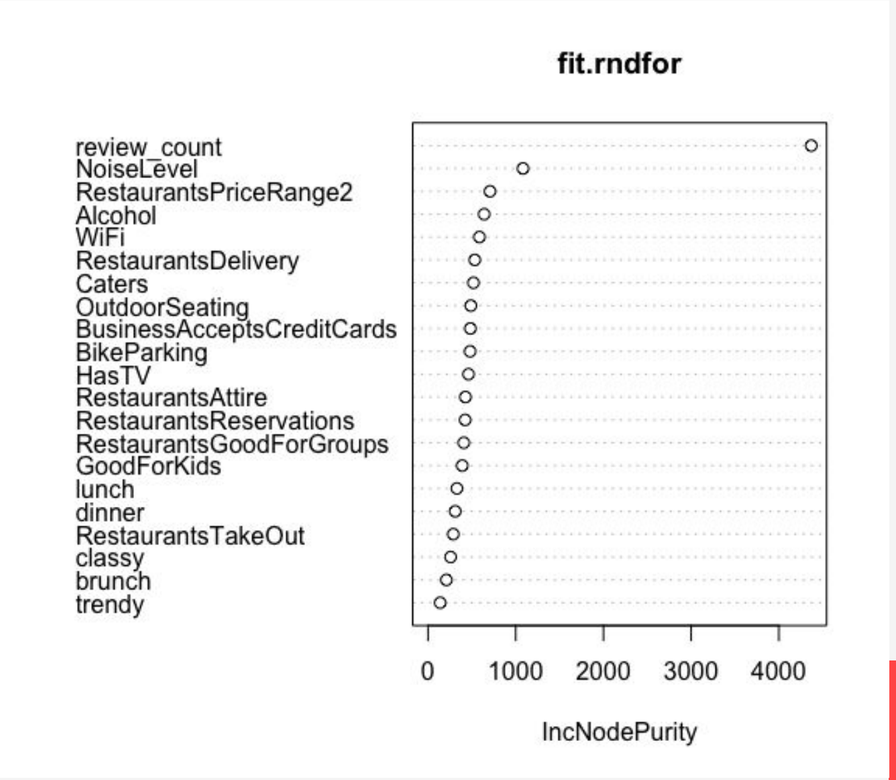
	MSE	R Squared Value
Simple Linear Model	0.8	0.12
Boosting Forest	0.46	0.51
Random Forest	0.21	0.638

Supervised Machine Learning Models - Random Forest

- The Random Forest model gave us the MSE of 0.21 and R squared value of 0.638 which is the best performance we had from predicted models.

- The variable of review_count played an important role.

- Other variables have smaller predictive power similarly.

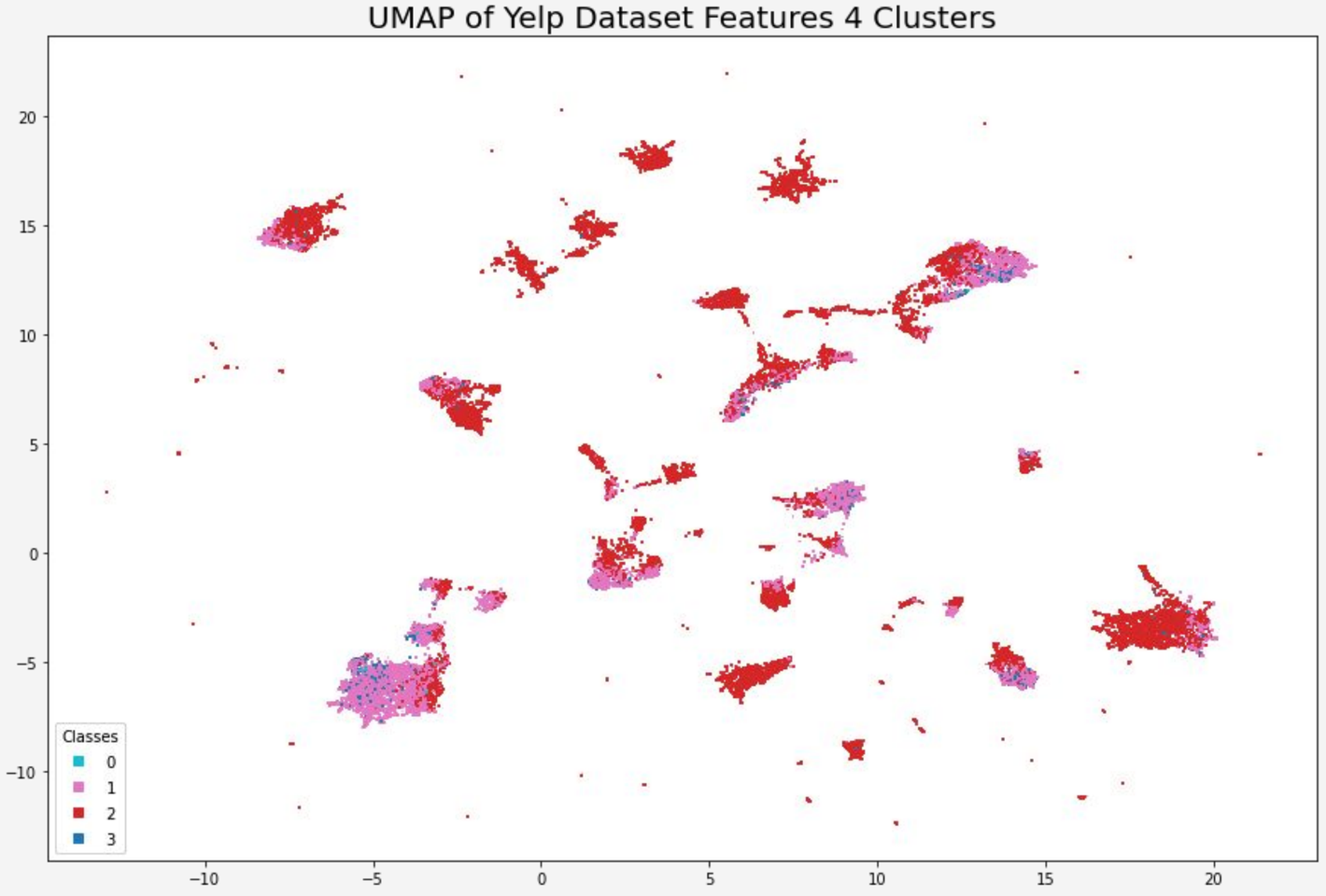


Clustering Restaurants

Eliminate reviewers subjectiveness and inconsistency

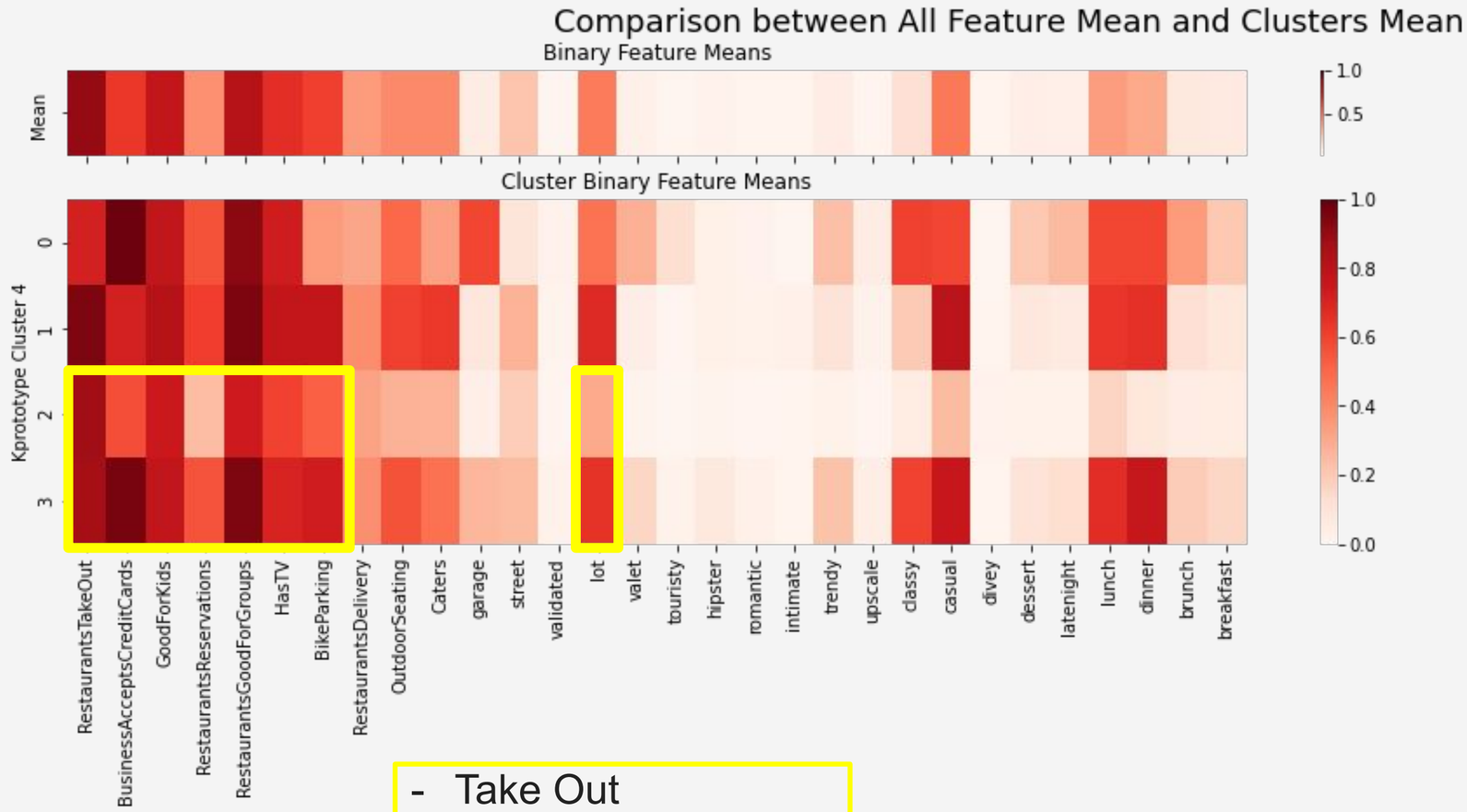
Clustering

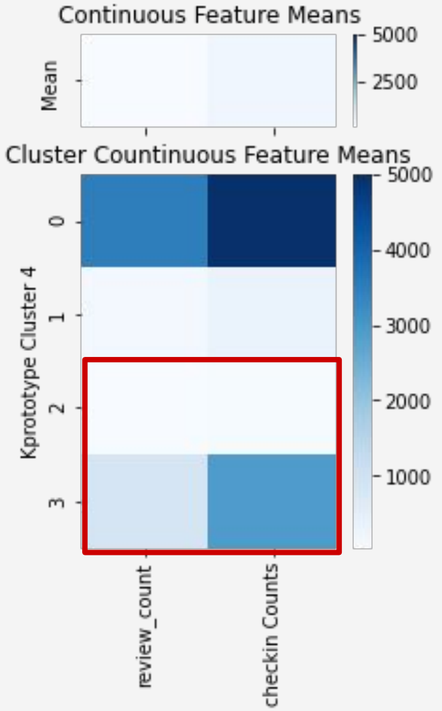
KPrototype with 4 clusters



Kprototype Cluster 4	stars	Count
0	3.851562	64
1	3.622035	14754
2	3.385310	26393
3	3.933197	1220

Clustering





- Take Out
- Parkings
- Accept Credit Cards
- Good for Groups

Review and Checkin Counts

Text Analysis

- word cloud,
- frequency plot,
- LDA topic model,
- sentiment analysis

General trends -

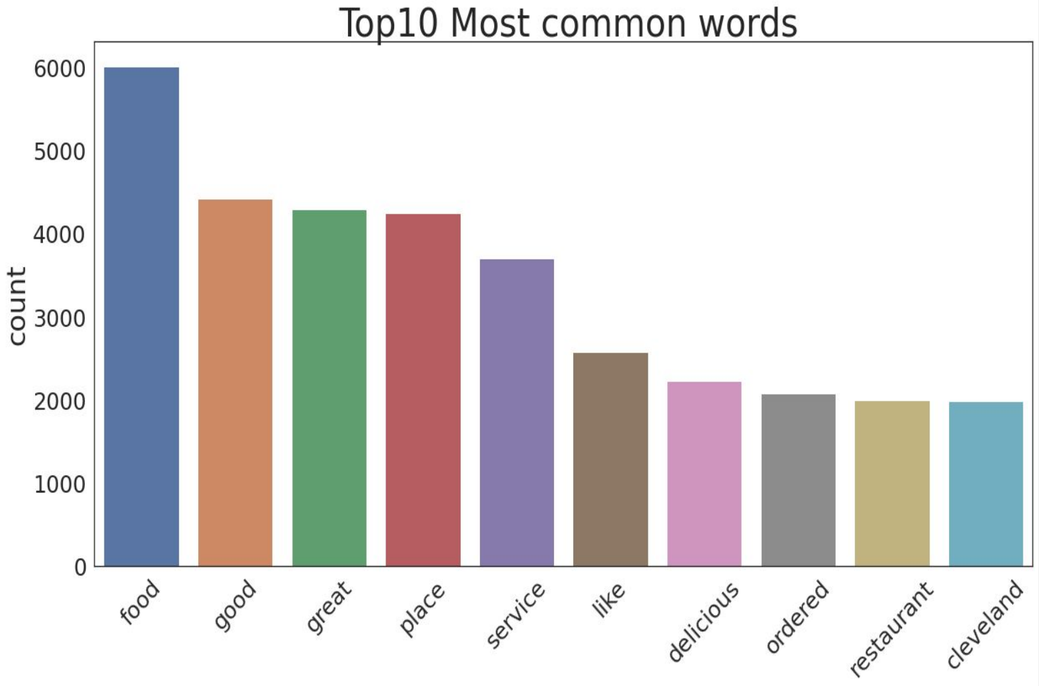
Customers care about food, service, and happy hour

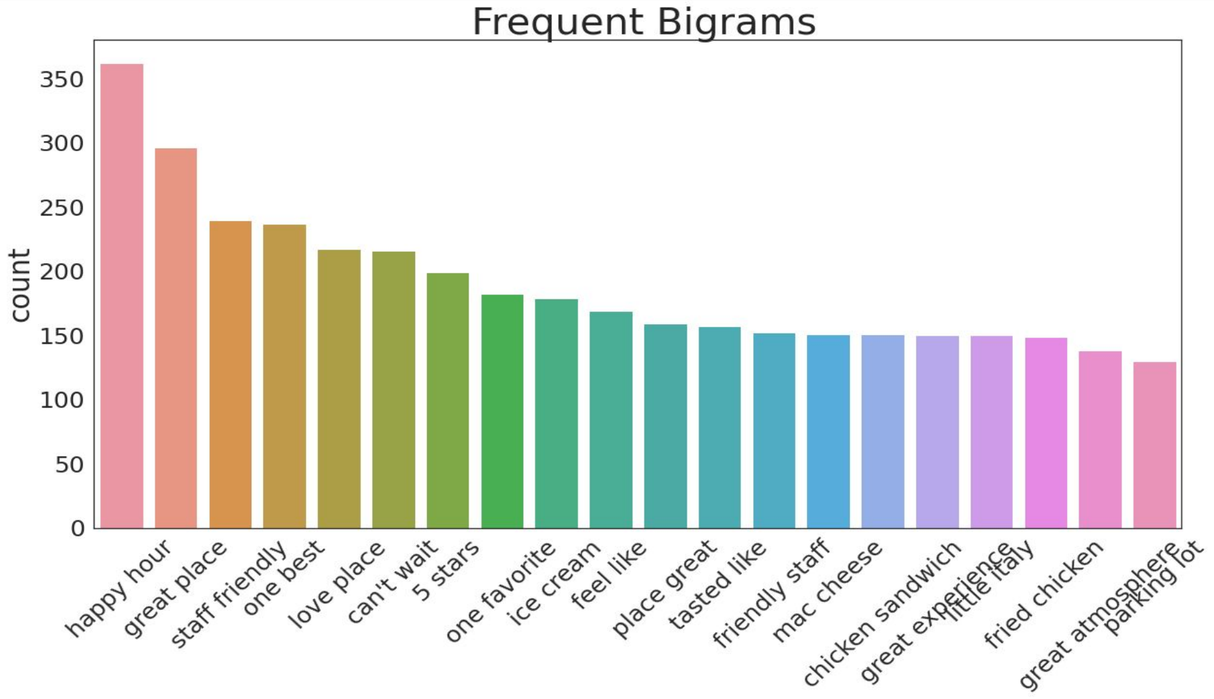
Food: 6004 times

Service: 3700 times

Delicious: 2230 times

Happy hour : 362 times





General Food Trend for Cleveland Restaurants



- The classic fast-food options appeared the most frequently
  - **Pizza, ice cream, chicken, fried chicken, chicken sandwich, and mac cheese**
- In terms of drinks
  - **beer** and bars.

Specific Food Trends

Healthy Foods/ Superfood

We notice that eating healthy meals is the trend now. The most popular restaurant - Townhall- offers a variety of food choices including gluten free food, veggie burger, grass-fed beef and so on.

Other fast food popular items

Other fast food popular items

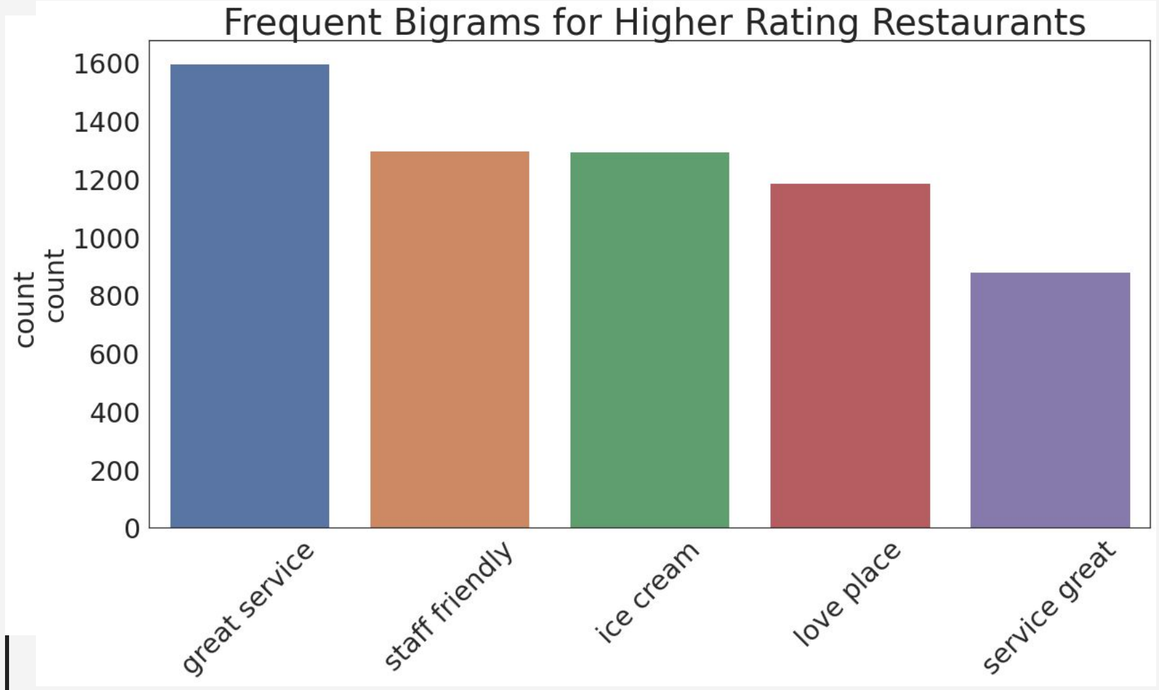
Sweet potato, fried rice, Taco

Asian Food Trend:

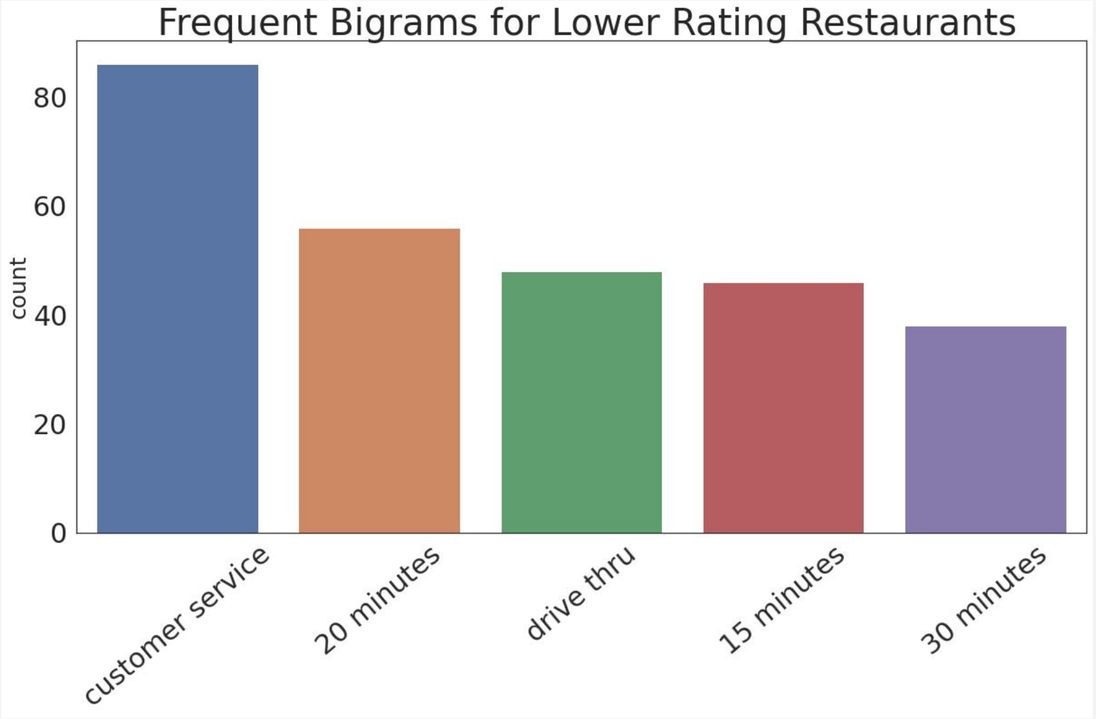
Soup, dumplings, ramen/noodles

High Rating V.S Low Rating

(>4.0)



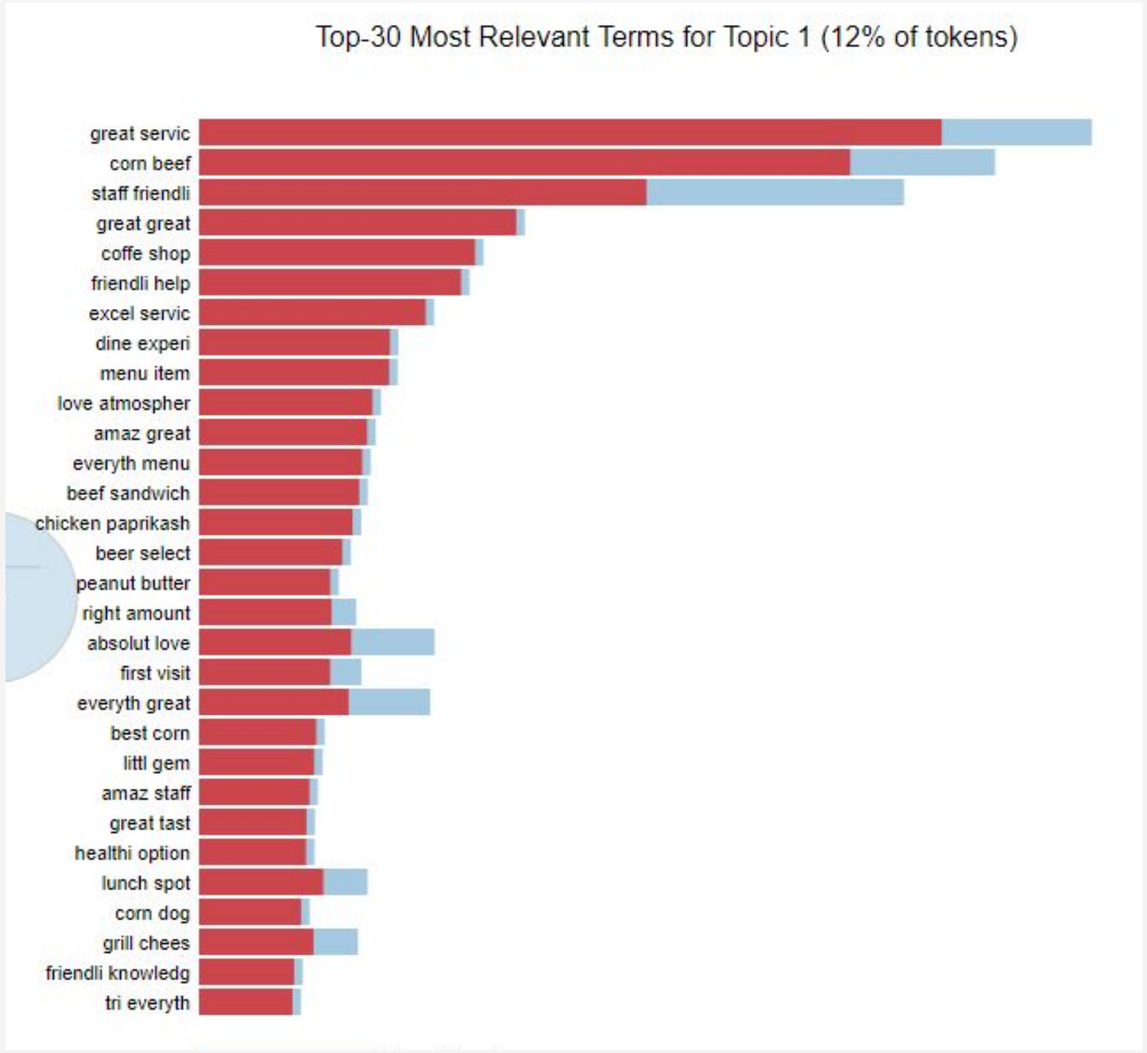
(< 3.5)

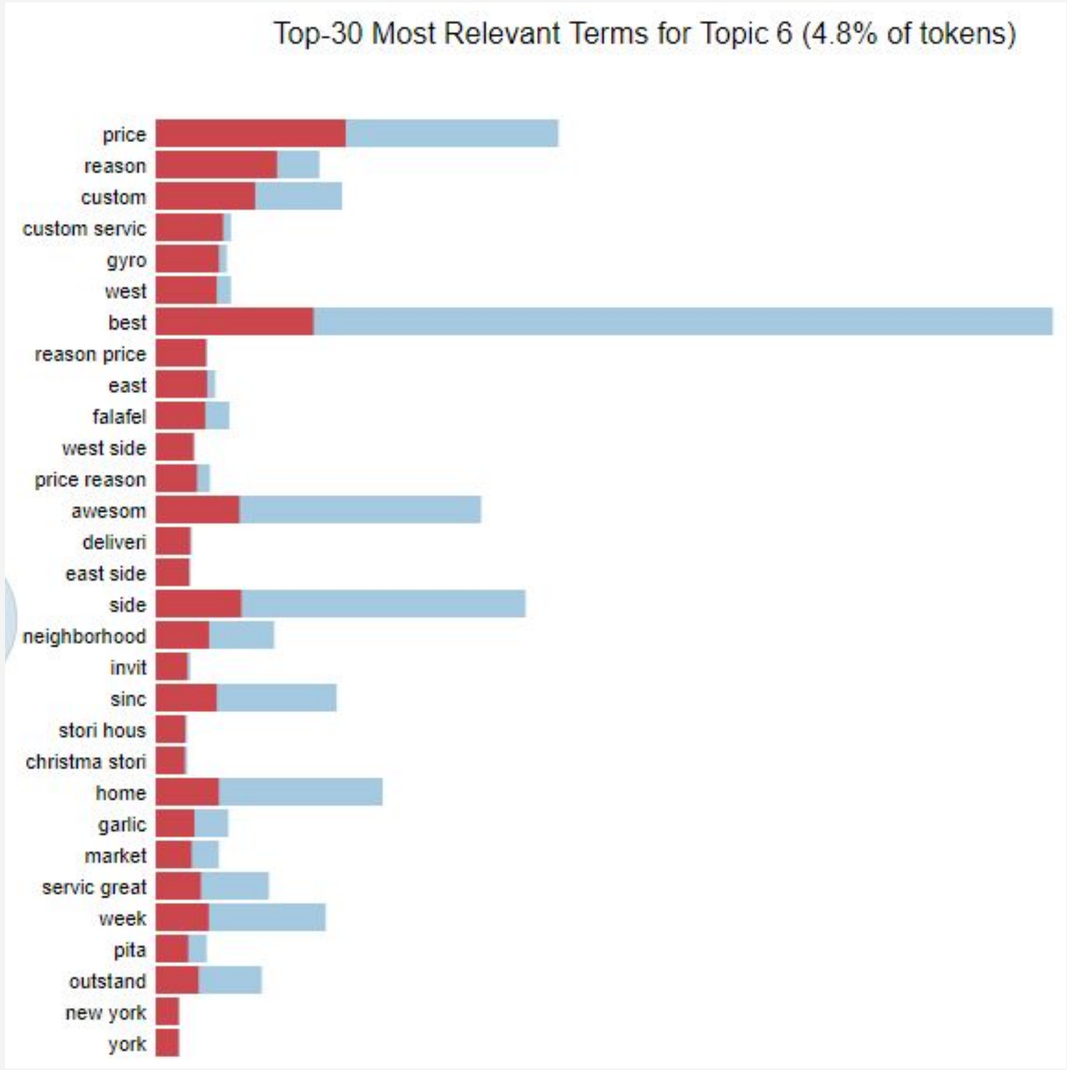


We can clearly see the quality of customer service is the main aspect. In addition, long waiting time seems to associate with the poor ratings.

High Rating Restaurants

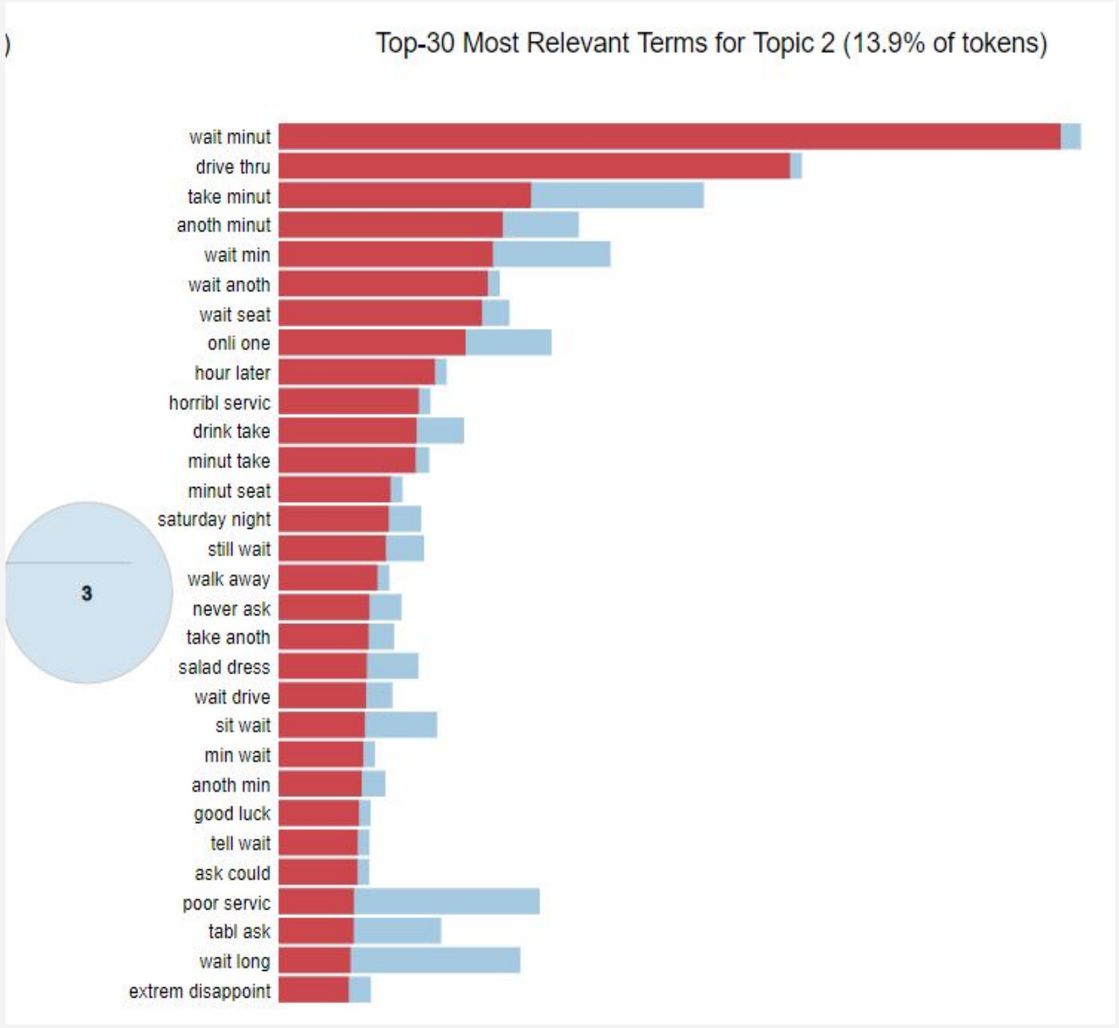
“High quality of service, reasonable price tend to attract a lot of positive comments and good ratings .”

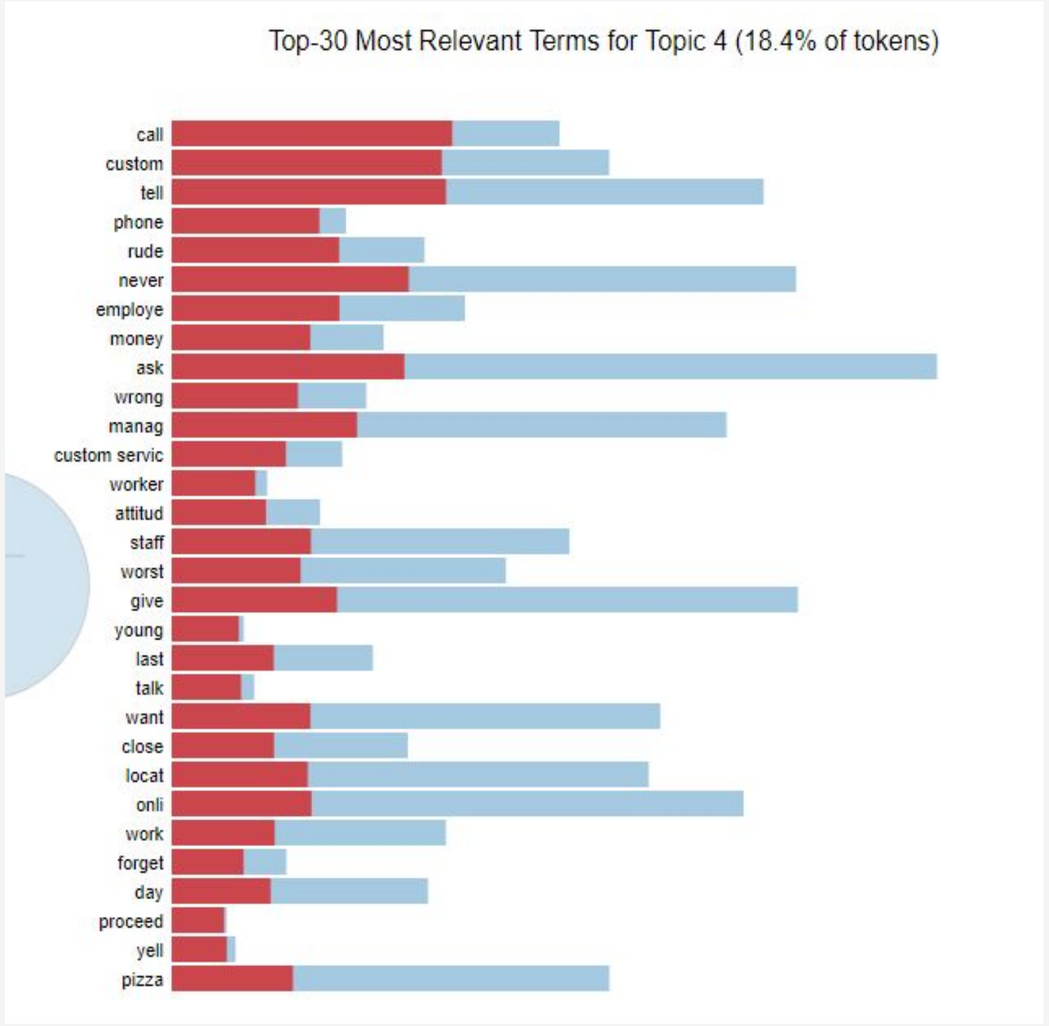




Low Rating Restaurants

“Long waiting time and poor attitude” might lead to more negative reviews and poor ratings.”





Gain more contexts

```
#look at service

text_lower.concordance("service",width=100)

text_lower.concordance("manager",width=150)
```

Service as keyword:

“Seriously horrible service! There is a lack of restaurant management,”

“Some of the absolute worst service I 've ever had. We waited 50 min”

“The service was extremely slow and our waitress kept ignoring… “

“it ruined my afternoon because he was so rude”

Manager as keyword:

“I was put on hold and told a manager was not available for several hours”

“The manager didn't seem to care about.”

“The manager was passive aggressive and made the experience terrible”

“The only positive is the manager was very nice.”

Recommendations

Reduce Wait Time:

1) Pre-schedule staff .

   Make sure there are enough employees during rush hours,

2) Accept reservations,

3) create fast-paced work flow and Improve table bussing efficiency

4) Prepare snacks for customers who are waiting

Improve service skills:

1)Listen to customers and solve problems immediately

2)Train employees in effective service techniques:

problem solving skills, definition of friendly services,

help the team maintain positive attitude by valuing employees

Analysis Insights and Conclusion

Good Ratings ?

Good Service!
Good Atmosphere!
Good Food !

Thank You.