Clustering & Topic Modeling Yelp Reviews of Restaurants in IL

Katie Fan

Introduction

- Yelp collects a massive amount of reviews

- it can be time consuming to read all reviews

Yelp Restaurant Reviews
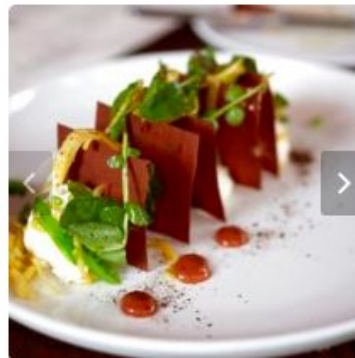
1. Girl & the Goat

★★★★ 8168 reviews

(312) 492-6262
809 W Randolph St
West Loop

$$$ · American (New), Bakeries, Coffee & Tea

"Having heard about Girl & the Goat for some time now, I took notice when a late night (10:45 p.m.) reservation was available recently. I'm so, so glad that I..." read more

3. The Purple Pig

★★★★ 6714 reviews

(312) 464-1744
500 N Michigan Ave
Near North Side

$$ · Tapas/Small Plates, Mediterranean, Breakfast & Brunch

"Context: Thursday night (around 6 PM) after work solo sitting at the bar. It was pretty busy, but I was able to snag a seat right away. Food & Drink: After..." read more
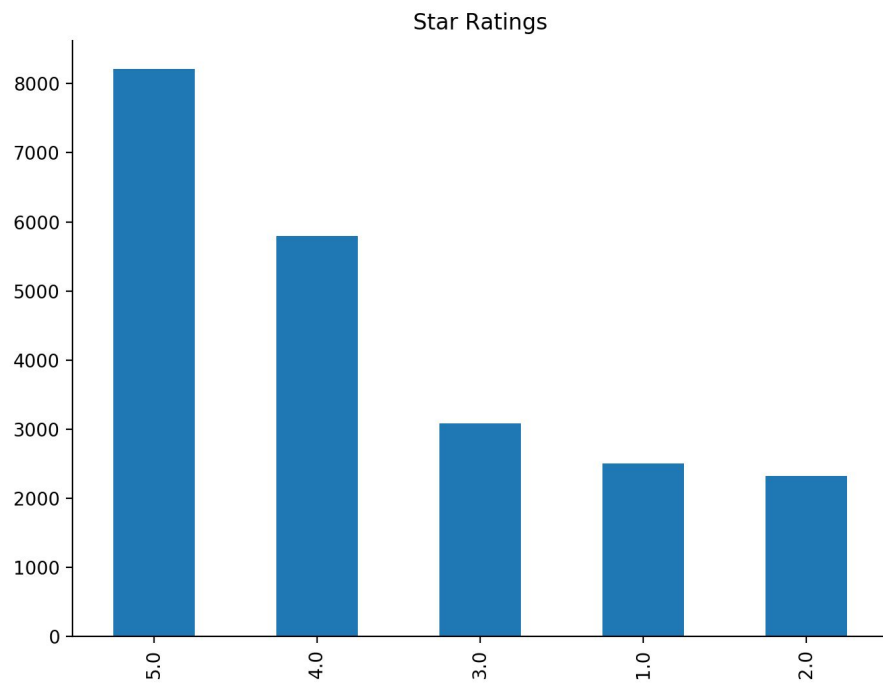
Too many reviews!

Objective

- to effectively develop insights from huge amounts of reviews for restaurants
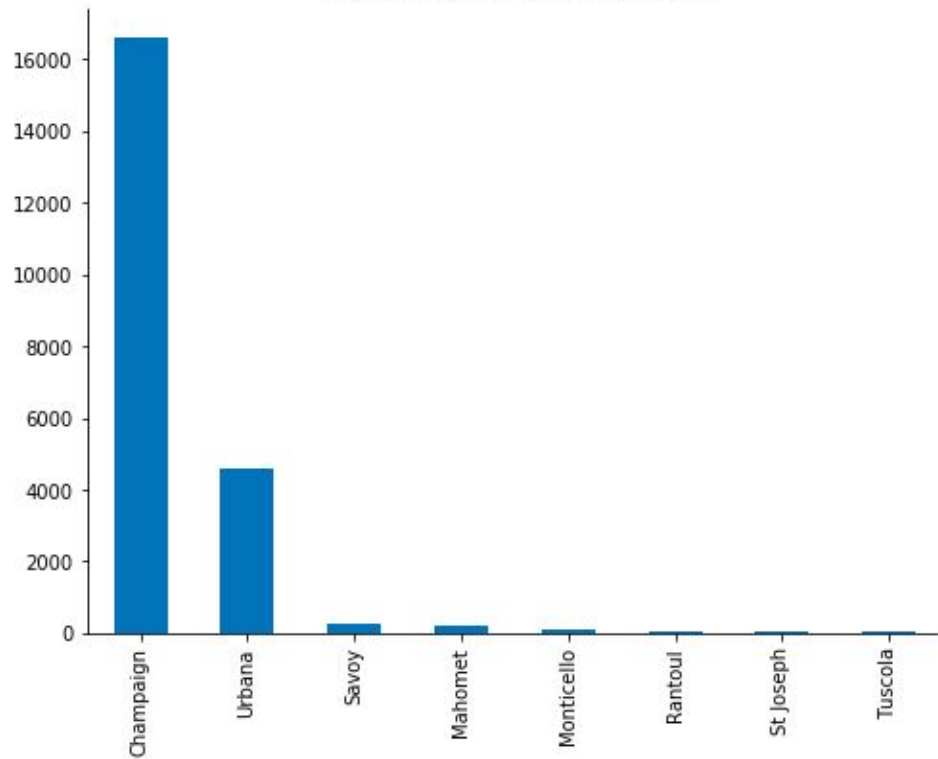
Data

- Data is from Yelp

- Total 5+ million reviews in the dataset

- Focused on restaurants in Illinois

- Chicago's restaurants are not in the dataset
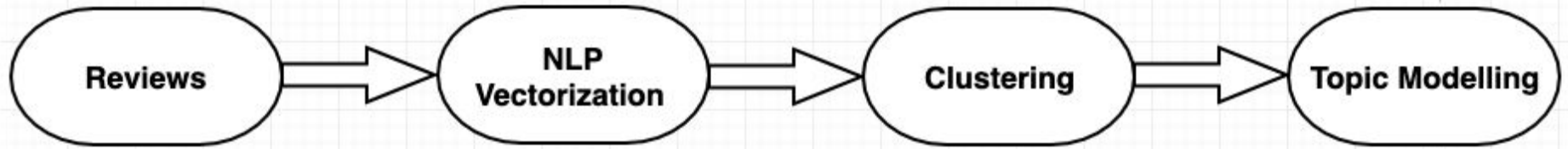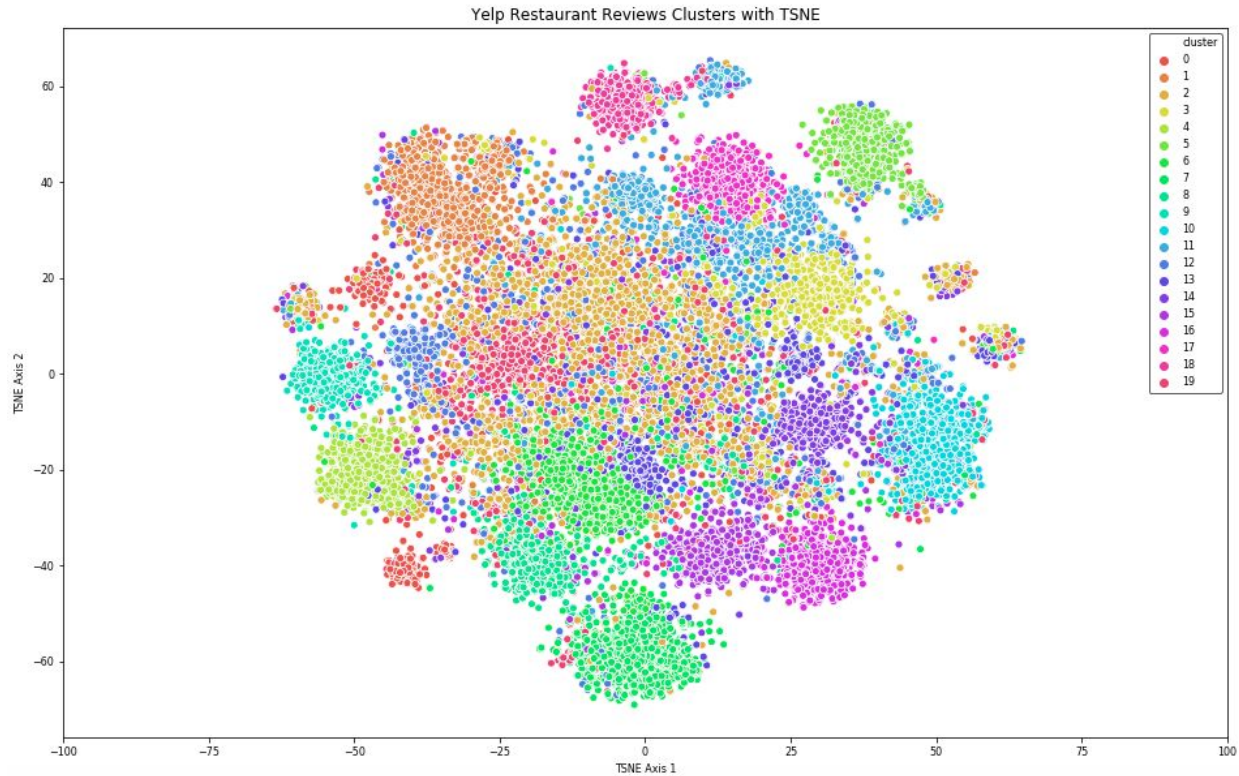
- 21,922 reviews were analyzed

Data

Data

Cities with Most Restaurants in IL

Approach

Results of K Means Clustering

Yelp Restaurant Reviews Clusters with TSNE

Cluster 1

Rainbow Garden

I have been going to this place for years. I swear they have gotten even better recently.

I got Chicken Eggplant last time I went and it was amazing. Food took no time to get to us. Can't wait to go back!

...

TGI Fridays

This place has terrible service! Poor customer service from management and poor table service. You should not have to wait 40 minutes for your food when everyone else is eating.

TGI Fridays

Slow slow slow!! There was four in our party and we were told the wait would be 40-50 minutes to be seated even though there were many empty tables. We told them we'd take any table, even in the bar area. We were seated at a bar table in about 25 minutes. It took over 10 minutes to get our drink order and another 10 minutes after that to take our food order.

Cluster 1

Total 1530 reviews in cluster 1

Results of Topic Modeling (NMF)

Topic ID	Top 5 Words	Topic
1	order, wait, time, minut, tabl	time
2	pizza, crust, deep, papa, chicago	pizza
3	great, food, place, servic, good	service
4	sushi, thai, chines, dish, roll	Asian food
5	good, fri, sauc, chicken, burger	fast food

Topic for One Review of Rainbow Garden

Restaurant: Rainbow Garden

Review:
Everything was pretty good but the beef in the Mongolian beef was very chewy and had a weird texture. I know this is always a 50/50 chance when trying a new Chinese restaurant. The rest of the food was very good! Their noodles are really good and the fried rice.

Topic Distribution:

topic: time	0.0385
topic: pizza	0.0383
topic: service	0.0384
topic: Asian food	0.8466
topic: fast food	0.0382

Conclusions

- Clustering and topic modeling are effective ways of extracting insights from large reviews data

- Could be the basis to build a useful data product for Yelp

Appendix

- Data: https://www.yelp.com/dataset

- NLP: https://en.wikipedia.org/wiki/Neuro-linguistic_programming

- Clustering: https://en.wikipedia.org/wiki/Cluster_analysis

- Topic Model: https://en.wikipedia.org/wiki/Topic_model

Thank you!