Mining Las Vegas Restaurant Reviews on Yelp

DANIEL BAE
THOMAS COCHRAN
PATRICK CONLEY

Project Description

Dataset:

- Yelp Open Dataset: https://www.yelp.com/dataset

Goal:

- Discover interesting patterns in Las Vegas restaurant review data that can improve business planning and service.

Questions

In this project, we sought to answer the following questions

1. What restaurant categories are frequently`` reviewed by Yelp users with low and high review counts?
2. Do restaurants with high or low average star reviews cluster around specific locations in the city?
3. Are there areas in the city where review sentiment is more negative or positive?
4. What are common text topics in Yelp low star and high star reviews?

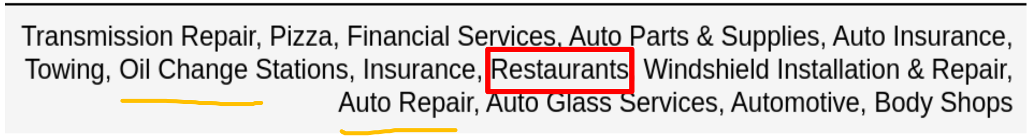
Data Preparation work

1. General data cleaning techniques
   - Duplication, normalization
   - Redundant attributes dropped
2. Text Data
   - Tokenize
   - Stop words
   - Lemmatization

Data Preparation work

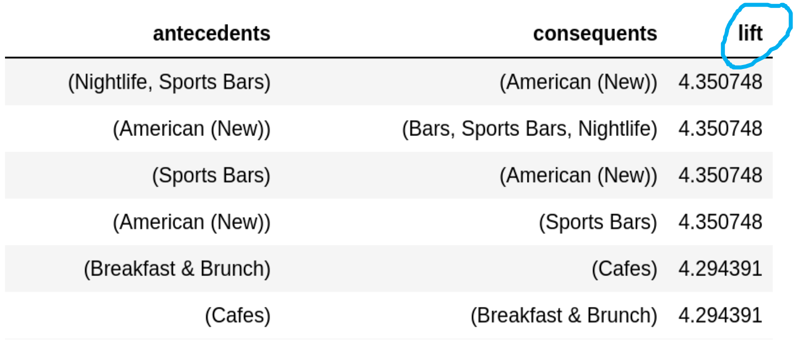
3. Filtering business categories

- Example of a business category list:



Transmission Repair, Pizza, Financial Services, Auto Parts & Supplies, Auto Insurance, Towing, Oil Change Stations, Insurance, Restaurants, Windshield Installation & Repair, Auto Repair, Auto Glass Services, Automotive, Body Shops

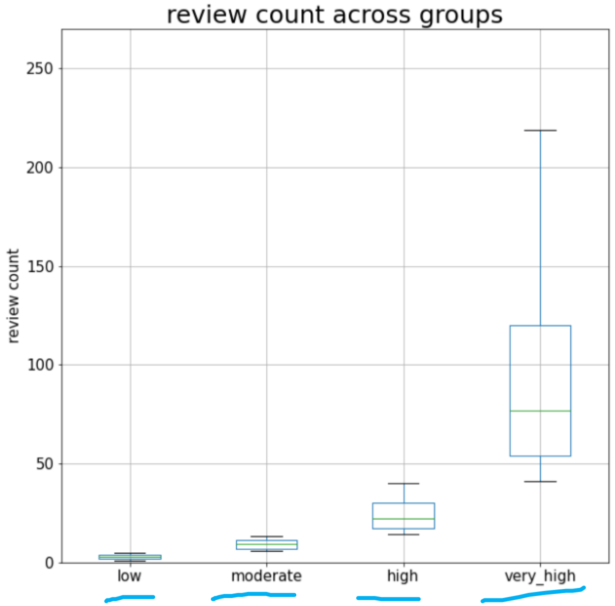
- Category list association rules:



antecedents	consequents	lift
(Nightlife, Sports Bars)	(American (New))	4.350748
(American (New))	(Bars, Sports Bars, Nightlife)	4.350748
(Sports Bars)	(American (New))	4.350748
(American (New))	(Sports Bars)	4.350748
(Breakfast & Brunch)	(Cafes)	4.294391
(Cafes)	(Breakfast & Brunch)	4.294391

4. Review count binning

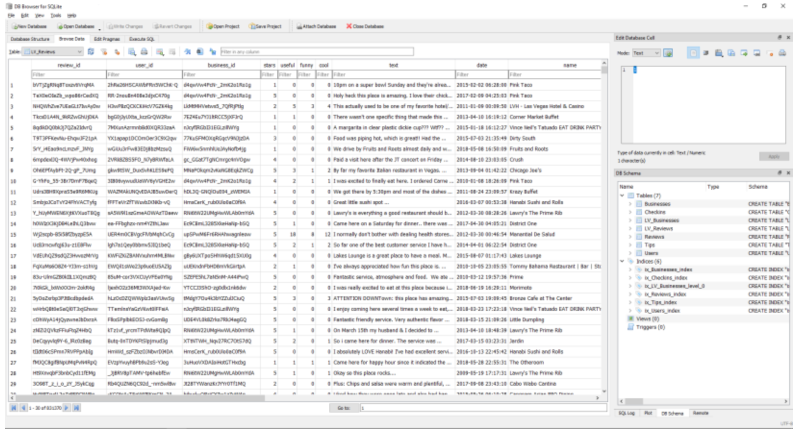
- Bin users by their number of reviews
- Log transform and bin using quartiles



Tools Used

Development Environment

- Python
- Jupyter Notebook
- DB Browser (sqlite)



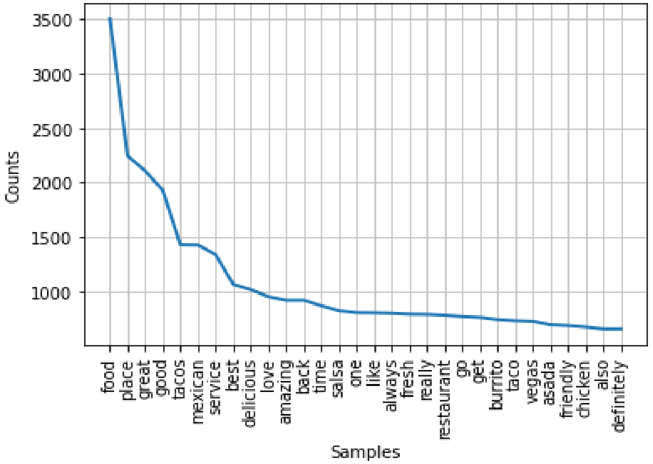
Version Control

- Git/GitHub

Tools for: Data analysis, Text Processing, GIS

- Pandas
- Numpy
- NLTK
- textblob
- Sklearn
- geopandas
- Seaborn/Matplotlib

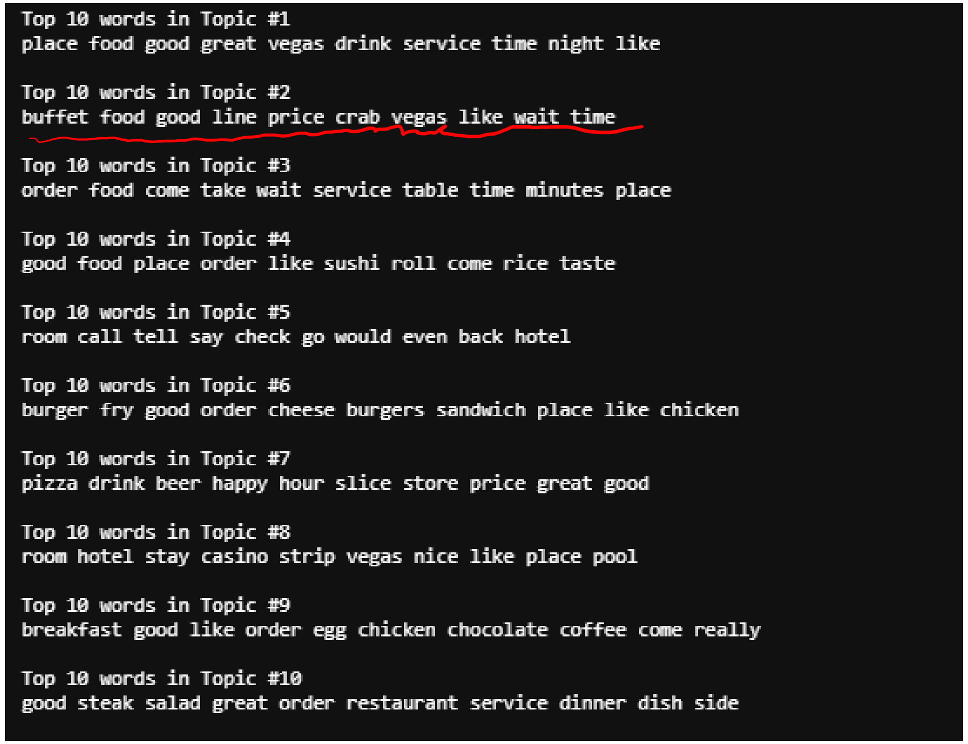
Frequency distribution for 30 most common tokens in 5-star Mexican restaurants



Clustering/Classification Methods and Results – Review Topics

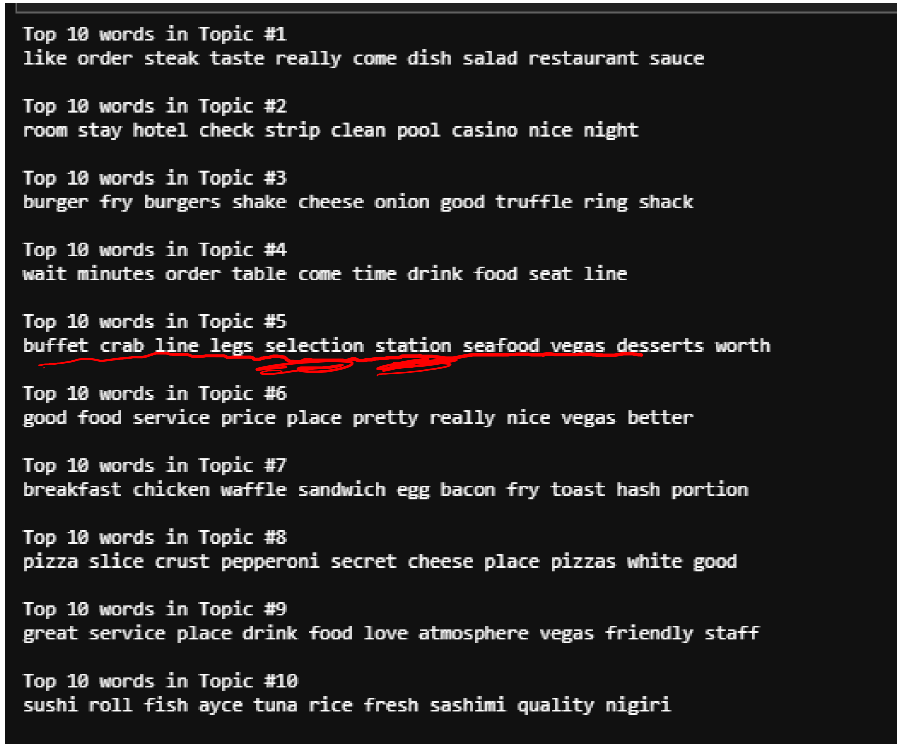
Topics: a collection of words that provide some context to a set of recurring themes

Modeling topics with LDA



```
Top 10 words in Topic #1
place food good great vegas drink service time night like

Top 10 words in Topic #2
buffet food good line price crab vegas like wait time

Top 10 words in Topic #3
order food come take wait service table time minutes place

Top 10 words in Topic #4
good food place order like sushi roll come rice taste

Top 10 words in Topic #5
room call tell say check go would even back hotel

Top 10 words in Topic #6
burger fry good order cheese burgers sandwich place like chicken

Top 10 words in Topic #7
pizza drink beer happy hour slice store price great good

Top 10 words in Topic #8
room hotel stay casino strip vegas nice like place pool

Top 10 words in Topic #9
breakfast good like order egg chicken chocolate coffee come really

Top 10 words in Topic #10
good steak salad great order restaurant service dinner dish side
```

Modeling topics with NMF



```
Top 10 words in Topic #1
like order steak taste really come dish salad restaurant sauce

Top 10 words in Topic #2
room stay hotel check strip clean pool casino nice night

Top 10 words in Topic #3
burger fry burgers shake cheese onion good truffle ring shack

Top 10 words in Topic #4
wait minutes order table come time drink food seat line

Top 10 words in Topic #5
buffet crab line legs selection station seafood vegas desserts worth

Top 10 words in Topic #6
good food service price place pretty really nice vegas better

Top 10 words in Topic #7
breakfast chicken waffle sandwich egg bacon fry toast hash portion

Top 10 words in Topic #8
pizza slice crust pepperoni secret cheese place pizzas white good

Top 10 words in Topic #9
great service place drink food love atmosphere vegas friendly staff

Top 10 words in Topic #10
sushi roll fish ayce tuna rice fresh sashimi quality nigiri
```

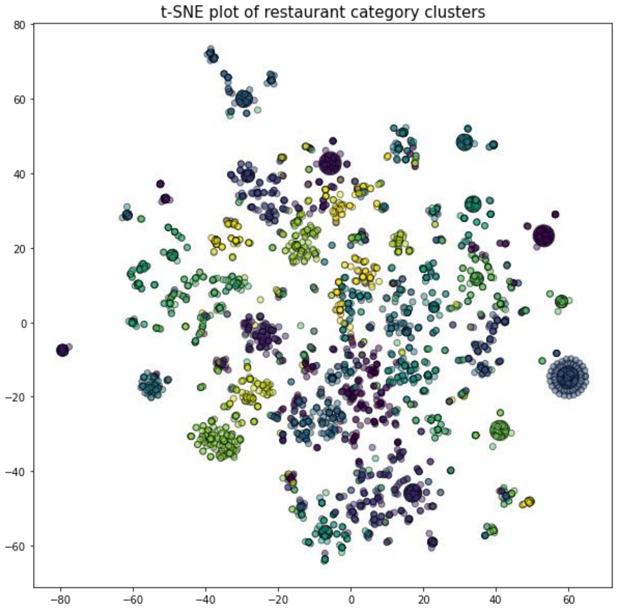
LDA Topics: Good star reviews (>=4 stars) by High review count users

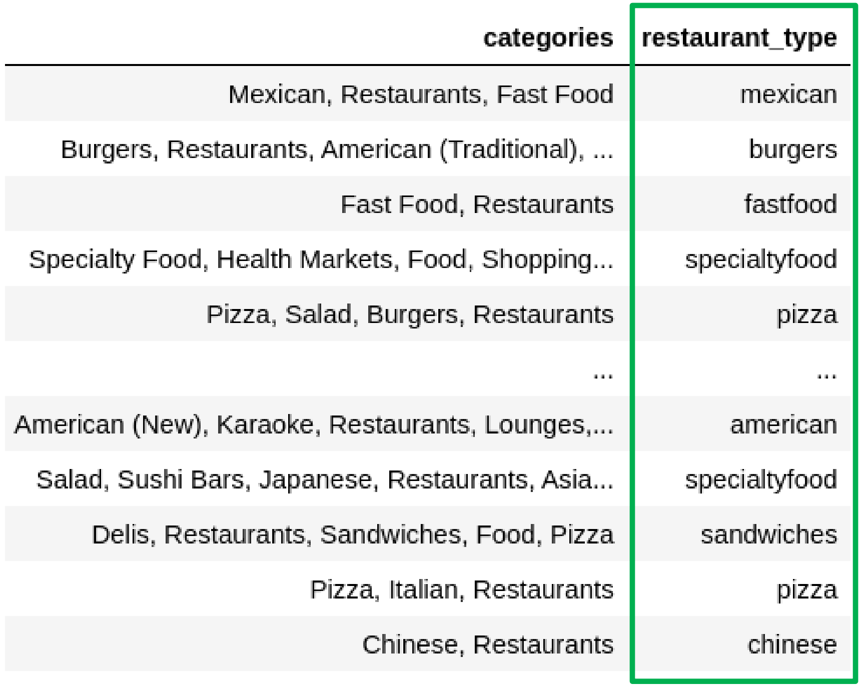
MNF Topics: Good star reviews (>=4 stars) by High review count users

Classification Methods and Results

- Classify restaurants using restaurant category lists
- Unsupervised classification
- Cluster category terms using their IDF weights

$$\boxed{tf\,idf}(t, d, D) = tf(t, d) \cdot idf(t, D)$$


t-SNE plot of restaurant category clusters



categories	restaurant_type
Mexican, Restaurants, Fast Food	mexican
Burgers, Restaurants, American (Traditional), ...	burgers
Fast Food, Restaurants	fastfood
Specialty Food, Health Markets, Food, Shopping...	specialtyfood
Pizza, Salad, Burgers, Restaurants	pizza
...	...
American (New), Karaoke, Restaurants, Lounges,...	american
Salad, Sushi Bars, Japanese, Restaurants, Asia...	specialtyfood
Delis, Restaurants, Sandwiches, Food, Pizza	sandwiches
Pizza, Italian, Restaurants	pizza
Chinese, Restaurants	chinese

Knowledge Gained

Restaurant types with high review counts:

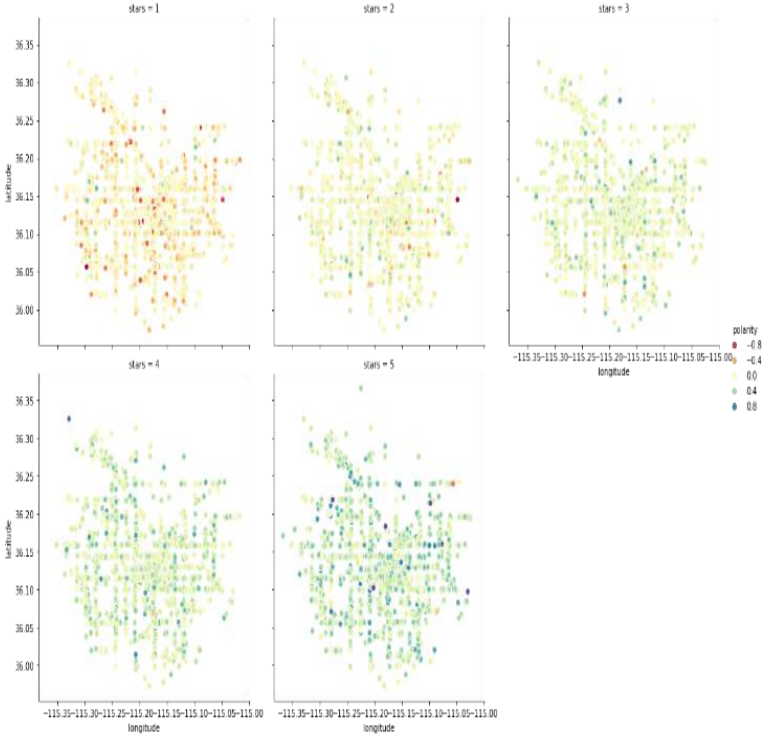
- Nightlife > American > Mexican > Fast food > Sandwiches > Coffee

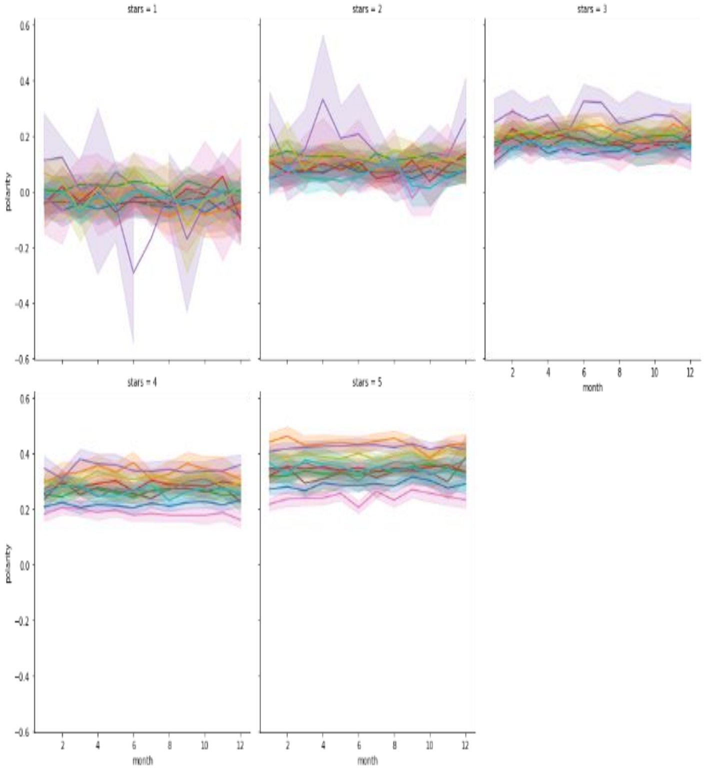
Restaurants with high star review near the city center:

- Sushi bars, bakeries, sandwiches

Restaurants with low star review near the city center:

- Mexican, fast food, coffee





Application of Knowledge

What Las Vegas restaurants generate the most hype?

Does restaurant location matter?

How can owners take advantage of Yelp user feedback?