Improving PA Restaurants through Yelp Reviews

Raymond Leong

Business Problem

Using Yelp, can we predict meaningful words associated in reviews to help improve **hospitality** or **food service** in the yelp's database?

Do we use an abbreviated review or a full review?

How can we translate these findings into data driven actions?

Stars 3 and above ratings for ALL buisnesses

Stakeholders

The stakeholders are the investors of the current 7076 restaurants in PA.

Why PA? It is good to start at examining restaurants that represents the majority of the reviews
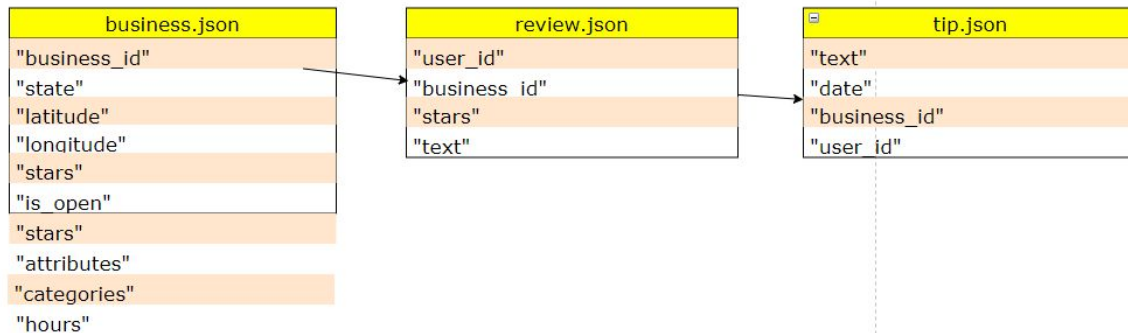

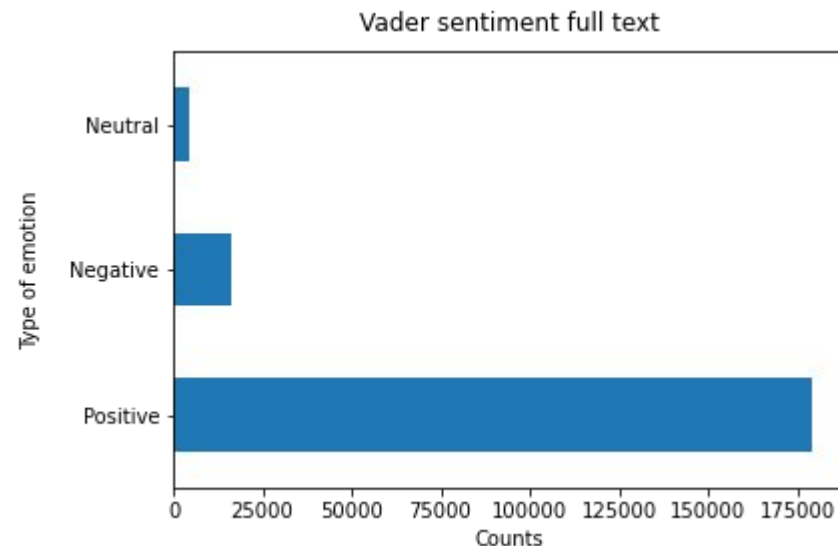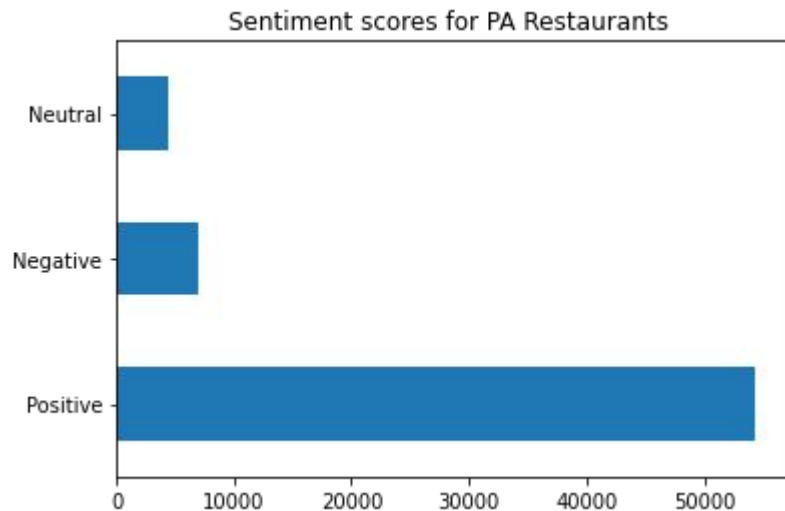Most Popular State with Restaurant Reviews.
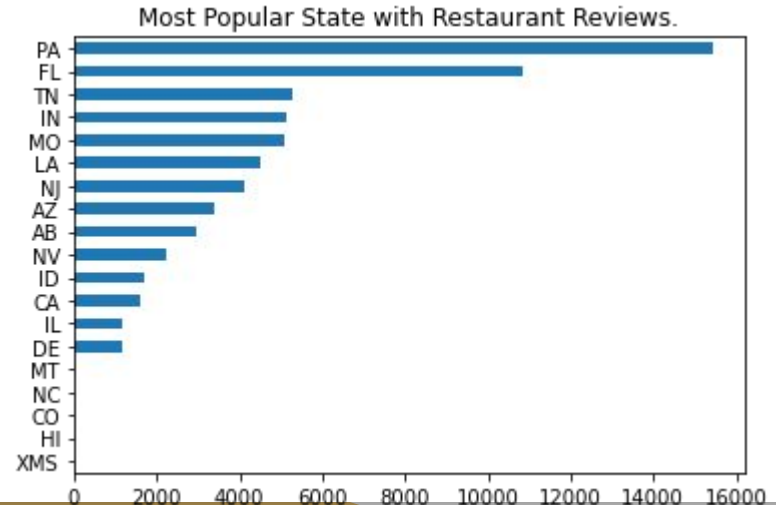
The Data

There are five files but only business.json, review.json, tips.json are relevant.

Due to machine limits, only 600,000 reviews were pulled from reviews.json

There are 150243 unique businesses, with 64,616 food establishments with 67,000 reviews for 7076 restaurants

business.json
"business_id"
"state"
"latitude"
"longitude"
"stars"
"is_open"
"stars"
"attributes"
"categories"
"hours"

review.json
"user_id"
"business_id"
"stars"
"text"

tip.json
"text"
"date"
"business_id"
"user_id"

Tips or full text?

Sentiment scores for PA Restaurants


Vader sentiment full text

We will be examining the full text instead of the abbreviated reviews because the distribution scores for the sentiment analysis are more similar.

Additionally our output variable: stars is only evident in the full text.

Time spent open during Weekends

Time for Friday, Saturday, Sunday: average range for 5 star restaurants was around 25.5 hours on the weekends with the upper quartile being in the 50 hour range. Linear Regression model failed for this FE.

Resturants and their Weekend hours
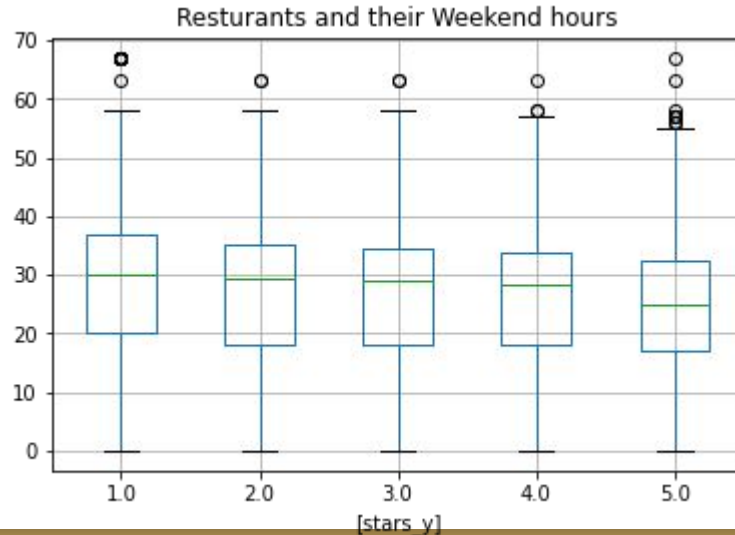
Model with CV

This model demonstrates popular phrases that can be made into business recommendations. Chicken, cheese, burger, beer are the most frequently appearing words.

'Night' is also common

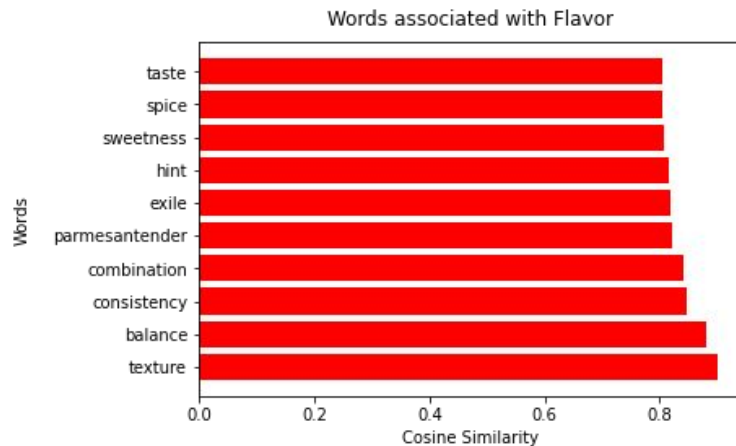Word analysis through Count Vectorization
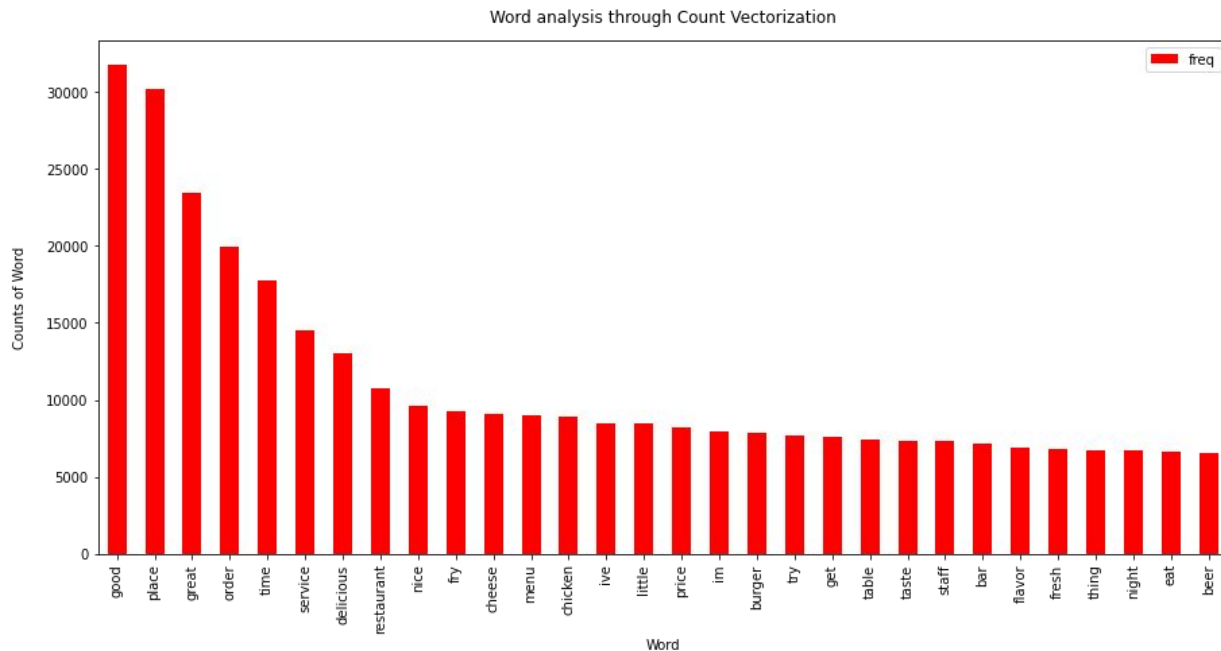
Model with NN 76% accuracy

The words; balance and texture appear most similar with flavor.

Some areas of flavor to explore might be spice and sweet flavors that have 'hints' of other 'combinations'

Additionally, consistency with flavor appears to high in similarity scores.


Words associated with Flavor

N grams for service

Waiter and Waitress show similar cosine similarity scores for certain behaviors.

For instance refilling is commonly associated with the job, so a business recommendation may be to advise your staff to walk around with water when busy.

Another task is probably taking orders from customers, don't overcomplicate things or arrive too early before they've made up their mind on their order.

n gram associated with waiter

n gram associated with Waitress
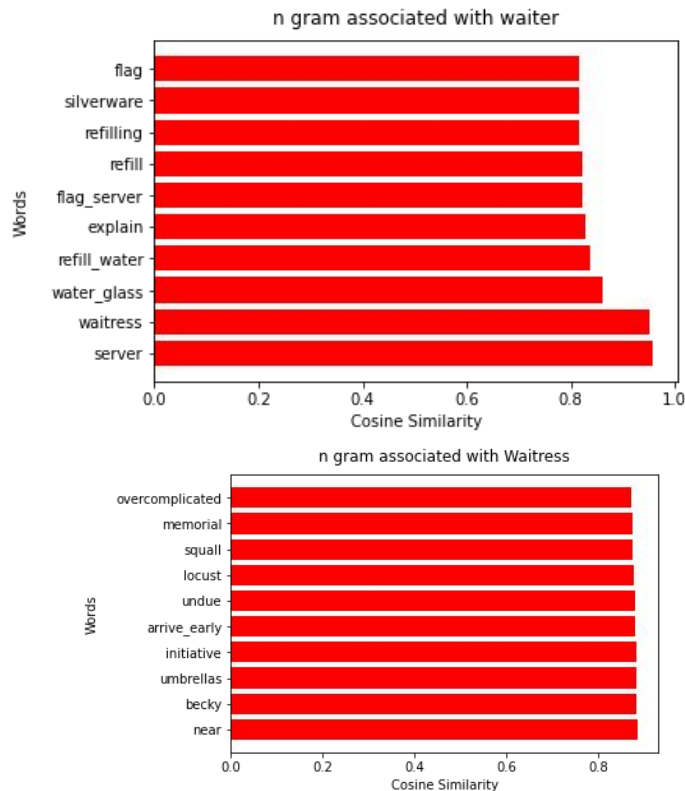
Final Statements

Due to the similar cosine scores in between words, we identified words most prevalent with the service or food quality domain.

We made some recommendations but we should use more of the dataset available because the scores are similar to each other.

We should incorporate the FE variables into another model to make more recommendations

Thank you!

Raymond Leong

https://github.com/RaymondLeong94

https://www.linkedin.com/in/raymond-leong-b75b11231/