Yelp Reviews, A Good Indicator of Restaurants Closing?

Computational Content Analysis Final Project

Yuqing (Candice) Zhang

Abstract

In this research, I examined how do the mechanisms of ‘Food’, ‘Ambience’, and ‘Service’ might lead to restaurant closing. Contrary to the common belief, this study shows that these three factors are actually not significant factors when predicting restaurant closing. Instead, it may show that reviews overall, are not reliable.

Humans are innately pleased by good food and unique dining experiences. The National Restaurant Association showed that in 2005 (the latest year for which the statistics were available), the average household expenditure for restaurant food was $1,054 per person, or $2,634 per household (Sandlin, 2005). These numbers make owning a restaurant a good business opportunity for aspiring food-service entrepreneurs, even during economic slowdowns and when consumer prices are rising. On the other hand, the restaurant industry has always been among the most competitive and challenging ones to navigate through, and failures are nothing new. However, the struggles have left some wondering what is behind the phenomena of restaurant closing.

Nonetheless, there have been many articles talking about the determining factors of businesses' success, though not much has been done on researching why businesses fail. The purpose of this study is to examine factors of restaurants closing. This study will build on three previous studies on restaurant failure rates (e.g., Parsa et al. 2005) that have debunked the myth that nine out of ten restaurants fail in the first year. Building on the previous findings, instead of adopting traditional approaches, such as interpersonal interviews or surveys administrations, in my study, I will take on a different method to study why restaurants close by performing a detailed text analysis on Yelp reviews.

With more and more access and ease in using the Internet, an increasing number of people are writing reviews for their experiences (Hu, Liu, 2004). As a result, the number of reviews that a product receives grows rapidly over the recent years. Some popular products can get hundreds of reviews instantly at some large merchandise sites, providing cascades of complex information. However, whether the reviews are indeed useful for scientific analysis is controversial. Even though according to a recent report, almost seven out of ten people read online reviews before they buy something, but that's not likely to be all that helpful, according to a new research from the University of Colorado Boulder. In the study, researchers looked at user reviews for 1,272 items, including car seats, sunblock, smoke alarms and bike helmets. Their conclusion is that there's only a small link between the average user rating of an item and the actual quality of an item based on objective tests. In other words, just because a product has an average rating of five stars, it doesn't mean it's a great product. Similarly, nor does an average rating of one star mean a product is by all means bad.

"I do not think we should ignore user ratings in general," says Bart de Langhe, an assistant professor of marketing at the Leeds School of Business at the University of Colorado Boulder, and the study's author. "But our research suggests that we should rely much less on user ratings than we do now."

In addition, many reviews are long and heavy on anecdotes, with only a few sentences containing actual opinions on the product. With these features, therefore, if something goes wrong, the large quantity of review information makes it very difficult for product manufacturers to keep track of customer opinions of their products. On this front, there hence is a need to classify reviews into categories and analyze them accordingly. In this project, I study the problem of generating feature-based summaries of customer reviews of restaurants. Here, features broadly mean restaurants attributes and functions. Although the determinants of business failure are multidimensional, I focus on examining three main factors, food, ambience and service.

Reasons for Business Failure

Failure in the restaurant industry was studied by Parsa et al. (2005). They listed three key factors as contributing to restaurants closing: size and type of operation, competition and restaurant concept or segment. The third factor, restaurant concept, emerged from the finding that restaurant failure rates are lower in restaurants where there is a good decoration. Parsa and colleagues (2005) also interviewed several owners of failed restaurants and then highlighted personal reasons for restaurant failure. Factors included a high "demand of labor and time, poor food-quality controls or low perceived value, being undercapitalized or having poor financial management, and the quality of employees and service". Their results showed that ambience, as well as service are two important factors contributing to restaurant closing.

In addition, food is at the soul of a restaurant and plays a crucial role in restaurants success. Therefore, apart from the two aforementioned factors, I added a third factor, food as a critical element to examine as well.

In all, my hypothesis is thus formed as: Food, ambience and service are significant predictors in facilitating restaurants' continuation in business.

Furthermore, my secondary hypothesis is that: The content of reviews are good indicators of whether the restaurant would close or not.

Method

Data. Currently, Yelp is the most popular online consumer review website used for local business reviews and recommendations (Bird, 2015). It is an online consumer review website for shopping, restaurants, home and other services containing more than 83 million reviews (Yelp, 2016). The abundance of Yelp reviews provides sufficient starting point for doing text analysis. This year, Yelp presents the ninth round of its own dataset challenge. Yelp released a dataset that includes information about local businesses in 11 cities across four countries. All data were enclosed in five *MySQL* files. For my own convenience, I created a database in MySQL and import all the relevant tables, for example the "review" table which contains text reviews, business_id, user_id, etc. In my project, I used the sub dataset that only contains text reviews for restaurants to explore the interplay of my three matters of interest on restaurant closing. After I did some filtering work, my data contains 11,484 reviews of 500 closed restaurants and 21,285 reviews of 500 open restaurants.

My main method includes these main steps: (1) filtering the raw reviews from Yelp so that we only analyze reviews within our matter of interest; (2) tokenize, normalize and POS tagging the text; (3) generate graphs to get a closer look at the content reviews; (4) identifying features of the product that customers have expressed their opinions on, and finding out those sentences that talking about ambience, food or service; (4) if hypothesis 1 is proved, for all sentences under one feature, calculating a sentiment score; and (4) producing a summary using the discovered information. More details for each step are discussed below.

1. Filtering

Filtering was performed on the businesses so that I only include the restaurants in within my matter of interest. Using the *is_open* variable in the original dataset, I further split the restaurants into open (*is_open* = 1) and closed (*is_open* = 0) restaurants for further comparisons. After the filtering process I had two data files, one with 11,484 pieces of reviews from 500 closed restaurants and the other with 21,285 pieces of reviews from 500 open restaurants. Following analyses were done separately for each of the two data files.

2. Tokenizing, Normalizing and POS Tagging

Tokenization is a way to split text into tokens. These tokens could be paragraphs, sentences, or individual words. NLTK provides a number of tokenizers in the tokenize module.

Further normalization of the tokens aims to remove some 'stop words', stem the remaining words to remove suffixes, prefixes and (in some languages) infixes.

Part-of-Speech Tagging (POS). Product features are usually nouns or noun phrases in review sentences, so the part-of-speech tagging is crucial. My POS tagging comes from the Natural Language Tool Kit. I used the NLProcessor linguistic parser (NLTK, 2000) to parse each review to split text into sentences and to produce the part-of-speech tag for each word (whether the word is a noun, verb, adjective, etc).

3. Visualization of the reviews

Word Cloud: first of all, I perform a Word Cloud to gaze at and draw mystical, approximate inferences about important nouns and verbs in our corpus.

Word2Vec: I produce word vectors with deep learning via word2vec's "skip-gram and CBOW models". Inside the word2vec object, the words each have a vector. We can look at a few things that come from the word vectors. For example, to find similar vectors using cosine similarity, for this would help me with generate related words later.

I can also use dimension reduction to visualize the vectors. I start by selecting a subset I want to plot. I also use PCA to reduce the dimension.

Co-occurrence: then I visualize the text reviews by the sentence co-occurrence graph, as it suggests many more meaningful (more local) associations. But without filtering, it is too large. I first drop all the edges with weight below 5, then drop all the isolates. Moreover, I can find cliques, or completely connected sets of nodes.

4. Feature Sentence Extraction

4.1 *Word Bank*. In order to extract sentences about either food, ambience or service, I utilized the Corpora project (Corpora Project, 2016) from the GitHub, as well as the Thesaurus website to find words that pertain to these three features. I note here that Corpora is a collection of static corpora of words. For example, some of the words that related to food are given as follows: *food, bread, cooking, cuisine, drink, foodstuff, meal, pizza.*

4.2 *Part-of-Speech Tagging (POS).* Again, I used the NLProcessor linguistic parser (NLTK, 2000) to parse each review to split text into sentences and to produce the part-of-speech tag for each word (whether the word is a noun, verb, adjective, etc).

4.3 *Feature Sentence Extraction*. I then extracted all the nouns (‘*NN*’ or *‘NNJ’*) out of the sentences. If any of the nouns in the sentence contains words from any of the feature word banks, the sentence will be included in that category. For example, the sentence “The lettuce in that ice berg salad is so fresh!” will be added to the ‘Food’ category. After this task, I collected all sentences under each of the three categories for every restaurant.

4.4 *Conditional Frequency Distribution*: I start by finding frequency distributions for the feature sentences. The Conditional Frequency Distribution class reads in an iterable of tuples, the first element is the condition and the second the focal word.

Results

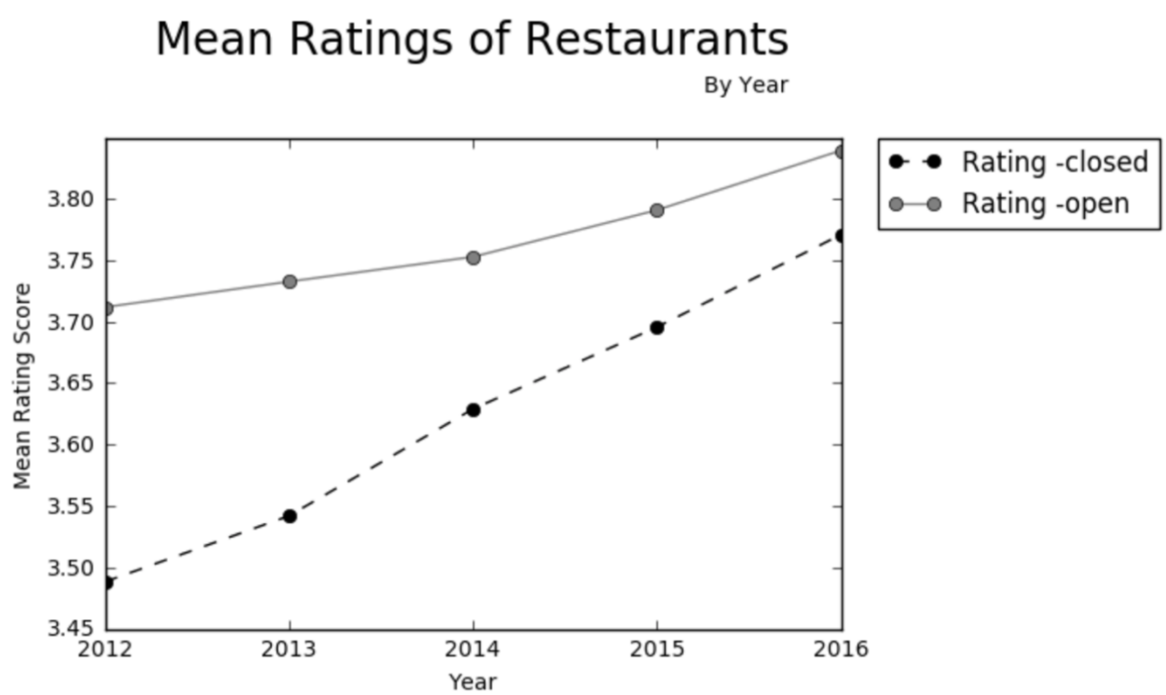


Figure 1 Mean ratings for open vs. close restaurants

Figure 1 shows the mean ratings for open and close restaurants accordingly. It is clear that the restaurants that are open received consistently higher rating than those that are close.

Figure 2 WordCloud of close (left) vs. open(right) reviews

The word cloud does give us much useful information, except that it seems like people like to talk about "place" more than food.

Figure 3 Top common super adjectives in reviews close(left) vs. open(right) restaurants

```
[('best', 1692),  [('best', 3262),
 ('least', 526),   ('least', 835),
 ('worst', 221),   ('worst', 442),
 ('honest', 189)   ('honest', 430),
 ('guest', 137)]   ('guest', 194)]
```

Initially I thought there would be many "bad" words in the reviews from close restaurants. Out of my expectation, however, the top super adjective word shown, is “best”. The frequency of the word “best” in the close restaurants, is only slightly lower than that from the open restaurants. This is quite interesting. What if we use other methods, will there be a difference?

Figure 4 Word2Vec Plot from reviews of close restaurants

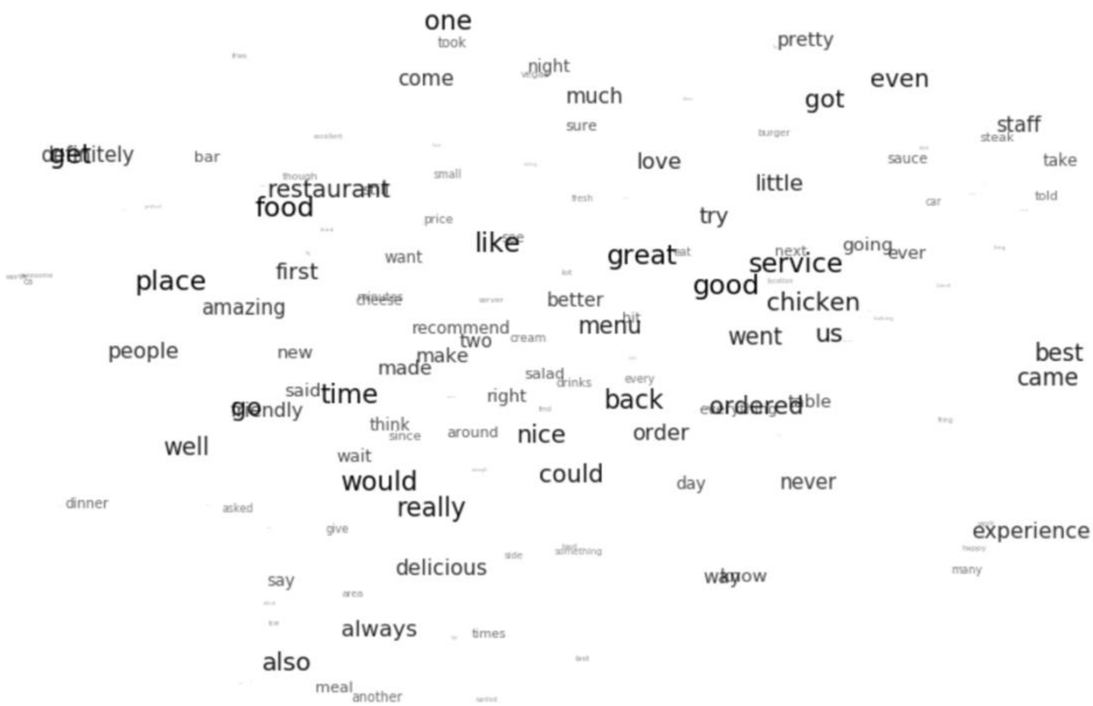
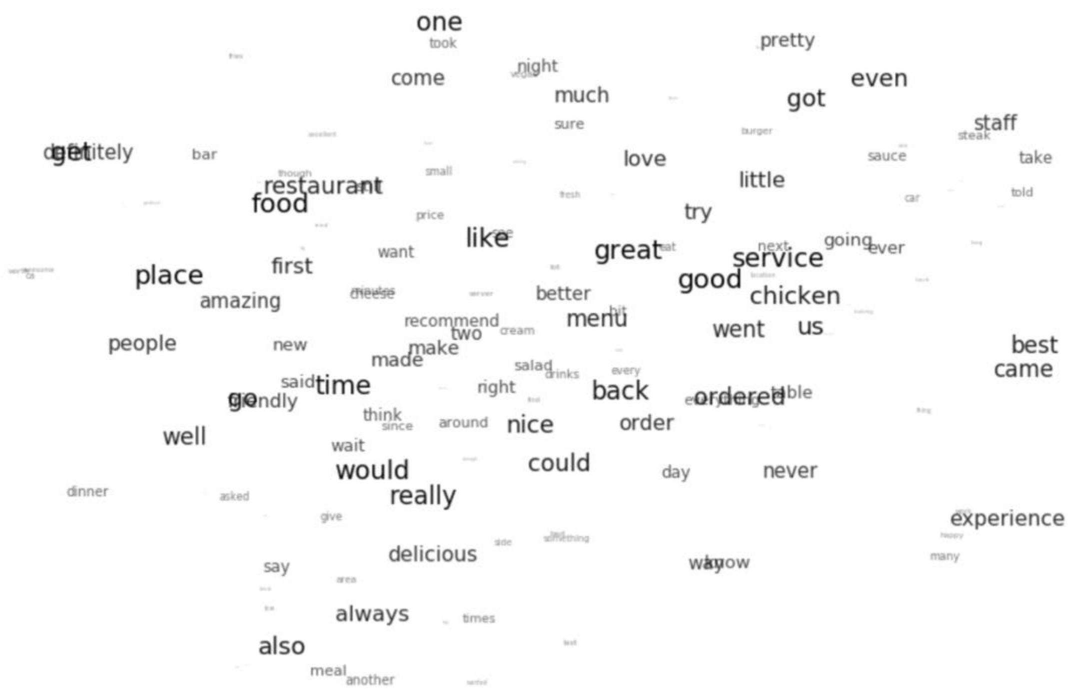


Figure 5 Word2Vec Plot from reviews of open restaurants



The Word2Vec plot, however, still does not show that “bad” words are prevailing in reviews from close restaurants.

Figure 6 Co-occurrence Plot of reviews in close restaurants

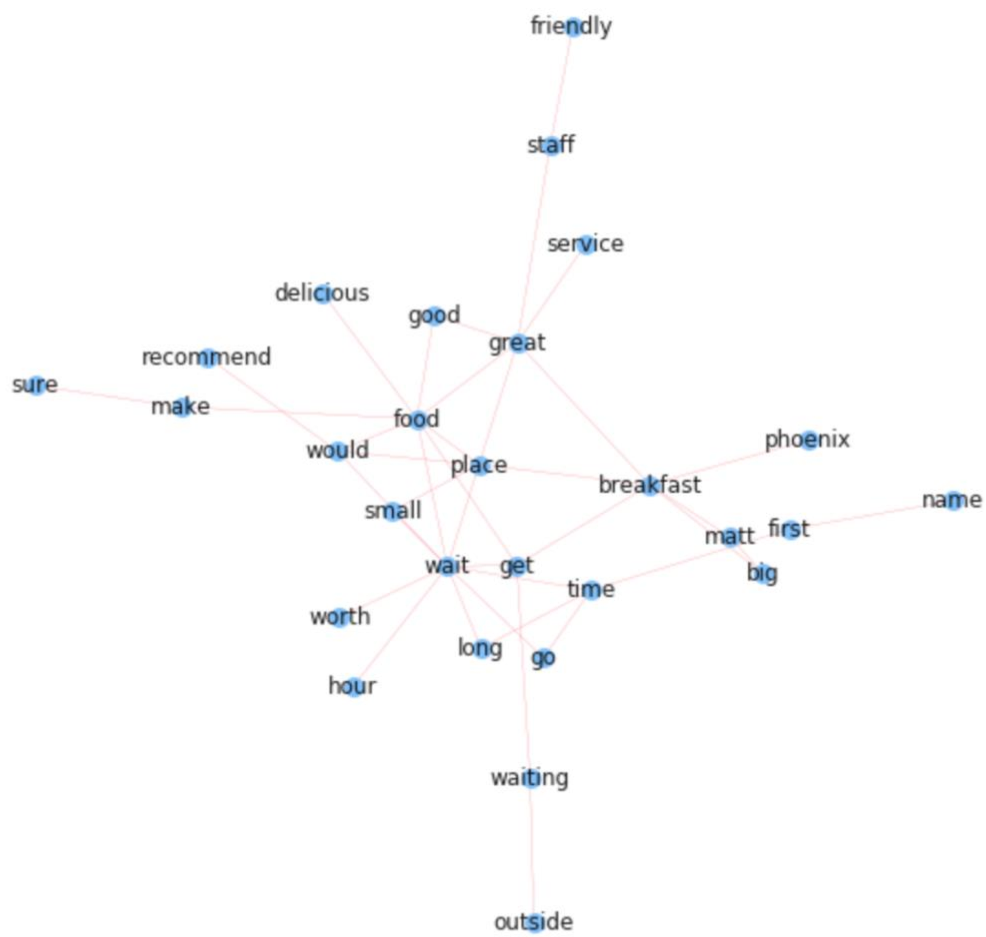
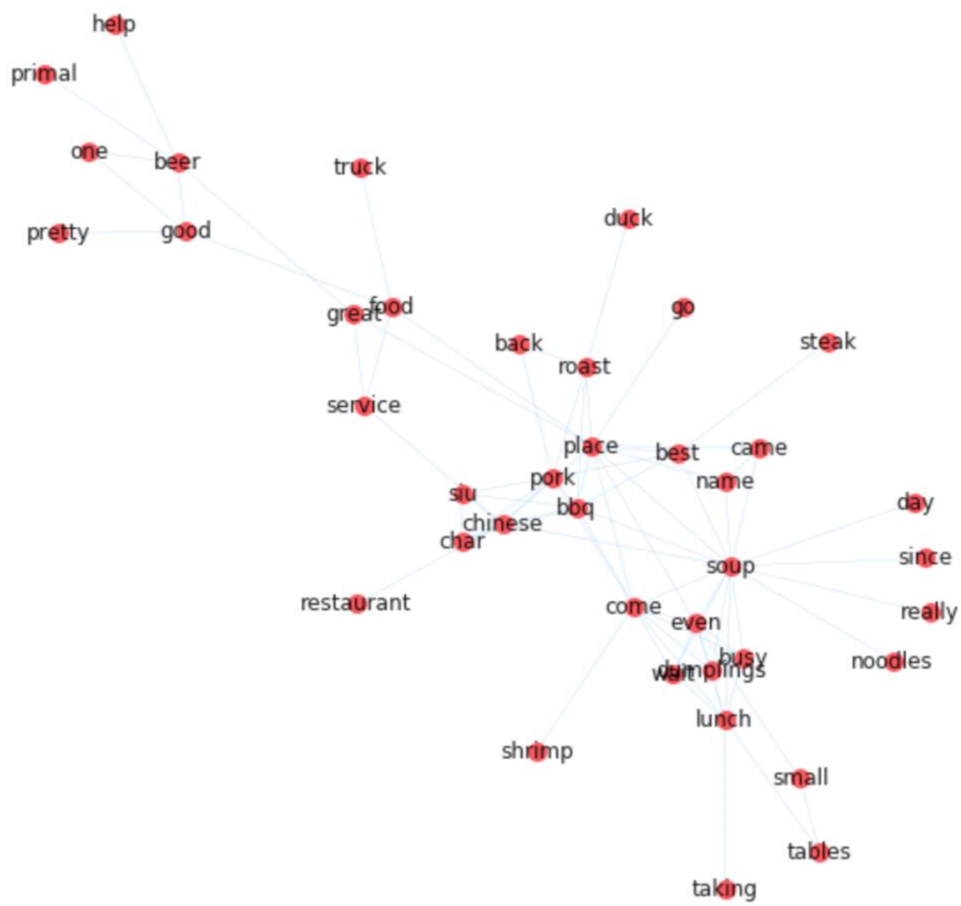
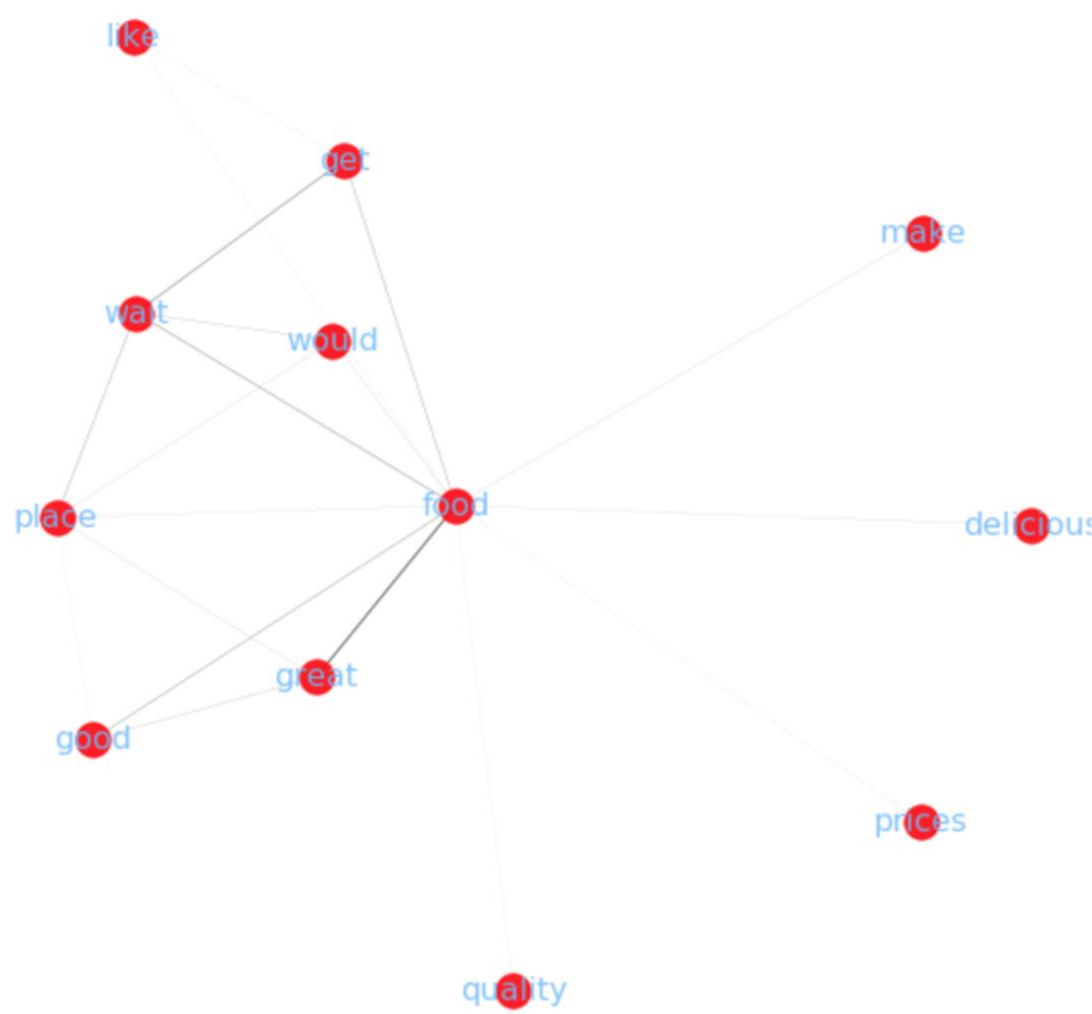


Figure 7 Co-occurrence Plot of reviews in open restaurants



Considering there might be a lot of random noise in the review, I decided to draw a Co-occurrence plot, as I think it suggests many more meaningful (more local) associations. But still, only including meaningful words still indicate that there isn't that much a difference in the sentiment around open vs. close restaurants' reviews.

Figure 8 Co-occurrence plot of "food" neighbors in reviews of close restaurants



After all the “failures” and surprises I got from the overall estimation of the reviews, I decided to dive into each category, and see if there will be any difference.

If we look more closely to the subgraph, those nodes that are within 1 or 2 network steps of the word “food” from close restaurants, are dominated by adjectives like, “delicious”, “good”.

Figure 9 Top adjectives around "food" in close restaurants

```
{'Outstanding', 'Good', 'fresh', 'favorite', 'exceptional', 'good', 'awesome', 'greasy', 'yummy', 'enough', 'amazin
g', 'Mediterranean', 'phenomenal', 'great', 'American', 'delicious', 'eastern', 'actual', 'basic', 'devoted', 'fast',
'Great'}
```

Some of the most common adjectives around the word "food" in close restaurants' reviews. It agrees with our co-occurrence graph, but still out of my expectation.

Then I thought maybe it is because I ignored so many sentences that relate to food, but do not contain the word "food", which biased my analysis. For example, "the salad is so bad." So I decided to generate my own feature based sentences, using method mentioned above in the method section. And below is an example of what it looks like.

26	The iced tea is terrific.	Food
27	I had to try the beef egg roll; quite differen...	Food
28	Great tabouleh, falafel and kibbeh.	Food
29	Very generous portions for the amount of money...	Food
...	...	...
42894	About seven minutes later the waiter said that...	Service
42895	I ordered the four cheese penne with chicken a...	Service
42896	But the overall customer service focus and the...	Service

Figure 10 Top adjectives around AMBIENCE, SERVICE, FOOD, related words in close restaurants

```
[('great', 919), ('good', 824), ('nice', 367), ('small', 300), ('new', 237)]

[('servic', 1365), ('great', 995), ('good', 847), ('nice', 338), ('slow', 181)]

[('good', 3187), ('great', 2025), ('fresh', 907), ('nice', 706), ('much', 578)]
```

Even I incorporated more sentences into each category, still, the sentiment around reviews from each these categories is overall positive, which I think is something need to be discussed.

Conclusion and Discussion

In the analog age, collecting data about behavior—who does what when—was expensive. Now, in the digital age, the behaviors of billions of people are constantly being recorded, stored, and analyzable (Salganik, Bit by Bit). The ever-rising flood of big data means that we have moved from a world where behavioral data was scarce to a world where behavioral data is plentiful. However, the big data itself is indeed clumsy, and the tools with which we apply to analyze the data is still imperfect. Almost always, there is an issue of the usefulness of the big data being examined.

In the past, although there has been some researchers studying the various factors leading to business failure, previous approaches were not able to collect information of large amount of restaurant customers.

In this study, I tried to tackle both issues. First, I had business administrative records data that Yelp created and collected as part of their routine activities. Critically, these business administrative records are not just about online behavior. By doing text analysis, however, I overruled my hypothesizes, the content of reviews are not good indicators of whether the restaurant would close or not in my case. It seems like there is no relationship between the reviews and whether the restaurant will close or not.

The reason why this happened is in need of a discussion. First of all, maybe it is because the relatively small size of my sample, for the sake of the analyzing speed, limit the predicting power of the reviews. In addition, adding more randomness to my sample selecting process might help. The mechanism of how Yelp arrange its data is unknown, therefore, instead of convenience sampling that I used, for further study, I will try to incorporate as many pieces of review from the dataset as I possibly can, and do random sampling.

Secondly, the mechanism that I used to determine the dependent variable, whether the restaurant is close or not might not be a good indicator whether a restaurant closes because it actually failed, or it is just because of, for example, the relocation of the restaurant owner. For further analysis, I need to do more field research before my analysis. Maybe just call the owner of some close restaurants and ask the reason they close.

Thirdly, maybe it is not because of me, it is due to the inherent flaws of the reviews. As previous study pointed that, the reason why the reviews are unreliable is because first, the average person has a hard time evaluating the objective quality of a product. To do that properly, someone would have to buy multiple alternatives of a product and put them through side-by-side testing. Instead, a consumer takes one product home and relies on other cues for quality. Second, some reviews are bogus. “About 5% of reviews are just fake,” de Langhe says. Some are posted by the company or employees of the company, or they’re posted by consumers who have received some incentive. Therefore, the phenomenon of reviews as not being a good indicator is not rare and specific in my case.

What’s more, the results of this study also suggest directions for further study on other factors, such as location and price range, as well as between-factor interaction effects. In addition, it calls for analyses on the subsets of the data to add richness to my results. For example, previous studies confirmed that Mexican style restaurants reported the highest failure rate, followed by sub shops, bakeries, coffee shops, and pizza restaurants. In contrast, cafeterias and seafood restaurants had the lowest cumulative failure rates (Parsa et al., 2005). With the richness of my dataset, I would also like to learn more about how different cuisines impact restaurant closure by a different approach from before, by performing review analysis.

References

Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Presented at the Eighth International Conference on Weblogs and Social Media (ICWSM-14).* Ann Arbor, MI.

Corpora Project. Available online at https://github.com/dariusk/corpora Sandlin, Eileen Figure (2005). Restaurant Success: The Main Ingredients. *Entrepreneur Magazine*. Available online at http://chefsblade.monster.com/benefits/articles/157-restaurant-success-the-main-ingredients

Parsa, H. G., J. Self, S. Sydnor-Busso, and H. J. Yoon. 2011. Why restaurants fail? Part II: The
impact of affiliation, loca- tion, and size on restaurant failures—Results from a survival analysis. *Journal of Foodservice Business Research* 14 (4): 360-79.

Parsa, H. G., J. I. Van der Rest, S. A. Smith, A. R. Parsa, and M. Bujisic. 2013. Relationship
between restaurant failures and demographic factors: Results from a secondary data analysis. EuroCHRIE Conference, Frieberg, Germany, October 16-19.

Hu, Minqing, Liu, Bing. 2004. Mining and Summarizing Customer Reviews. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* Pages 168-177. Seattle, WA, USA — August 22 - 25, 2004

Dave, K., Lawrence, S., and Pennock, D., 2003. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. WWW'03. Salganik, Matthew J. Bit by Bit: Social Research in the Digital Age, Princeton University Press, Open review edition C. Bird, (2015, May). Which local business reviews are better: Yelp or Google? [Online] Available: http://localvox.com/blog/local-business-reviews-yelp-or-google/

Yelp. (2016, June) An introduction to yelp metrics as of June 30, 2015. Yelp. [Online] Available: http://www.yelp.com/factsheet

NLProcessor–TextAnalysisToolkit.2000. http://www.infogistics.com/textanalysis.html