THE ROLE THAT HOURS OF OPERATION PLAYS FOR RESTAURANTS AND THEIR AVERAGE STAR RATING ON YELP

INTRODUCTION

There are a few local food establishments in my neighborhood that keep very inconsistent or unstable hours of operation, which often impacts my choice of whether or not to visit these places for lunch or dinner. More often than not, these restaurants are small, owner-operated shops that are relatively new or just starting up. I often will have a good experience eating at one of these places but then when I return with family and friends, to our surprise, we will find it closed. This got me thinking about how inconsistent or inconvenient hours of operation impact certain restaurants and food services establishments. Can we draw specific conclusions about a restaurant based on their hours of operation?

With regards to the Yelp dataset that I will be working with for this analysis, my key research questions are targeted towards addressing this issue. *Are restaurants and other food services establishments that are open during the highest frequency checkin times receiving better reviews, on average? Does the text of the reviews or tips mention the convenience or consistency of the hours of operation? Are there patterns in the text of the reviews or tips, pertaining to this hours of operation issue, that can be used to predict whether a business is likely to receive a below or above average star rating?*

If there is a strong relationship between a restaurants average star rating and the amount of hours, or at least convenient hours, that these establishments are open we should be open to find patterns in the text of reviews and tips that will help us predict whether a particular restaurant is likely to have a below or above average star rating. This could be a significant finding for some smaller businesses as it would help them to see the impact that their hours of operation have on their overall success.

METHODS AND DATA

The following analysis is conducted on a collection of data provided by the Yelp Dataset Challenge 2014 Submission. This is a series of information about businesses and the reviews, tips and checkins that are associated to these businesses by Yelp users/reviewers. The original dataset consists of over 1.6 million reviews and over 500,000 tips from 366,000 users regarding over 61,000 businesses. For the sake of this analysis, I have chosen to focus on a 5% sample of the businesses in the original dataset. Furthermore, I have selected only restaurants and food services establishments for the focus of my analysis. My analysis sample consists of 1,502 restaurants. From these 1,502 restaurants, I have pulled the corresponding checkin data for 1,269 of these restaurants. I have also pulled nearly 60,000 reviews and a little under 19,000 tips regarding these restaurants.

The analysis will follow a step-by-step process. The first step will be to determine what are the busiest days and hours of operation for the restaurants in my sample. For this, I will explore the checkin data for each of the corresponding businesses in my restaurants subset. Checkins capture the date and time that customers log in when they are visiting a particular establishment. The checkin data will tell us what hours and days of the week are the most widely used for customers. Businesses will be rated on how available they are during these high frequency checkin times?

Once high volume checkin times are identified, a *checkin sum* will be calculated for each restaurant which represents the number of high frequency checkin time slots that their business is open. From

here, a simple regression model will be run to determine whether there is a significant relationship between checkin sum and the average star rating for a restaurant. This will indicate whether our initial research question is even worth asking. *Are restaurants and other food services establishments that are open during the highest frequency checkin times receiving better reviews, on average?* If the answer to this is 'no', that would indicate that there is not a meaningful relationship between hours and operation and a restaurants potential success or failure. However, if this answer is 'yes', then we have a reason to continuing exploring the data.

Once the relationship between checkin volume and star rating has been established, the next step will be to conduct an analysis of the text, specifically the text that users leave for both reviews and tips. For our purposes, I will focus specifically on just those restaurants that have a below average star rating and a below average checkin volume. Once I have identified this subset, I will use text mining techniques to determine whether there are any words or phrase patterns that can be found in both the reviews and tips for these below average restaurants. More specifically, I am hoping to find specific words or phrases that identify the lack of consistent hours of operation as a specific reason why the user rated that restaurant poorly.

Finally, once particular words or phrase patterns are identified from the text mining analysis, these phrases can then be reapplied back to the full dataset in an attempt to predict which restaurants are more likely than not to receive below average star ratings.

RESULTS

The first step of this analysis was to identify the high volume checkin times from the 1,269 tips that were associated with the 1,502 restaurants in my sample.


2
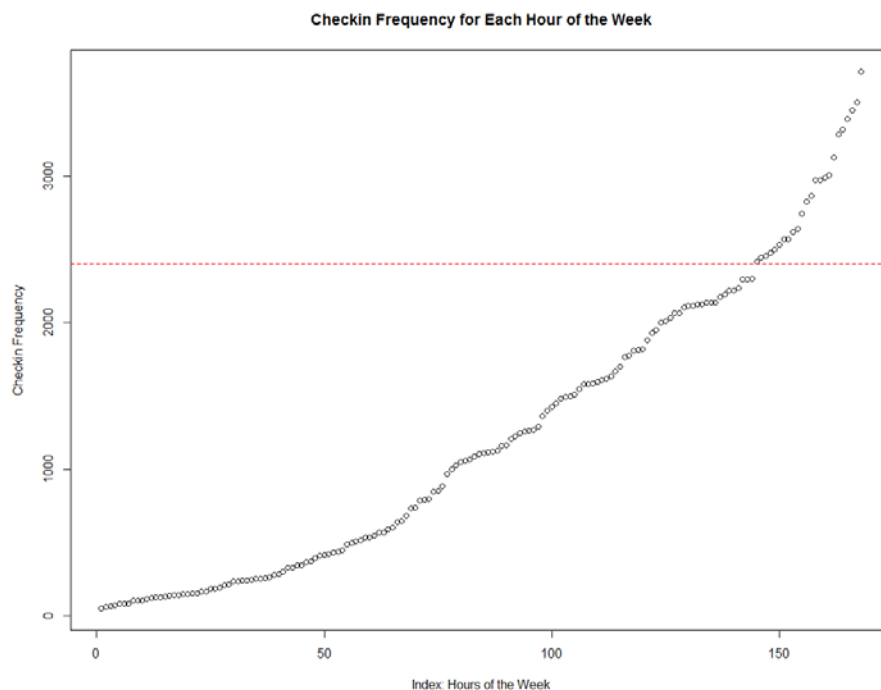
From the plot above we can see that there is a natural break around 2,400 checkins. All of the checkin times with a frequency of 2,400 or more were identified as the high frequency checkin times. There were 24 of these time slots, as follows (not in order of frequency):
- Monday (2): 6-7pm, 7-8pm
- Tuesday (2): 6-7pm, 7-8pm
- Wednesday (2): 6-7pm, 7-8pm
- Thursday (5): 12-1pm, 5-6pm, 6-7pm, 7-8pm, 8-9pm
- Friday (9): 11-12pm, 12-1pm, 1-2pm, 2-3pm, 4-5pm, 5-6pm, 6-7pm, 7-8pm, 8-9pm
- Saturday (4): 11-12pm, 12-1pm, 1-2pm, 6-7pm

For each restaurant, a total was calculated for the number of hour slots that each business was open during one of the 24 high frequency time slots, referred to as *checkin sum*. The checkin sum for each business ranges from 0 to 24 with an average checkin sum of 8.1 time slots. The assumption is that restaurants with a higher checkin sum (20+) have much more flexible and convenient hours of operation for consumers and vice versa for restaurants with a below average number of checkin sums.

Once these checkin sums were calculated for each restaurant, the next step was to run a simple regression model to determine whether there is any relationship between checkin sum and a restaurants average star rating. The regression plot and coefficient table below indicate that, even though this relationship looks somewhat scattered, there is a clear positive relationship between the two factors. Furthermore, the regression model results in a *p-value of less than 0.001 (***)*.
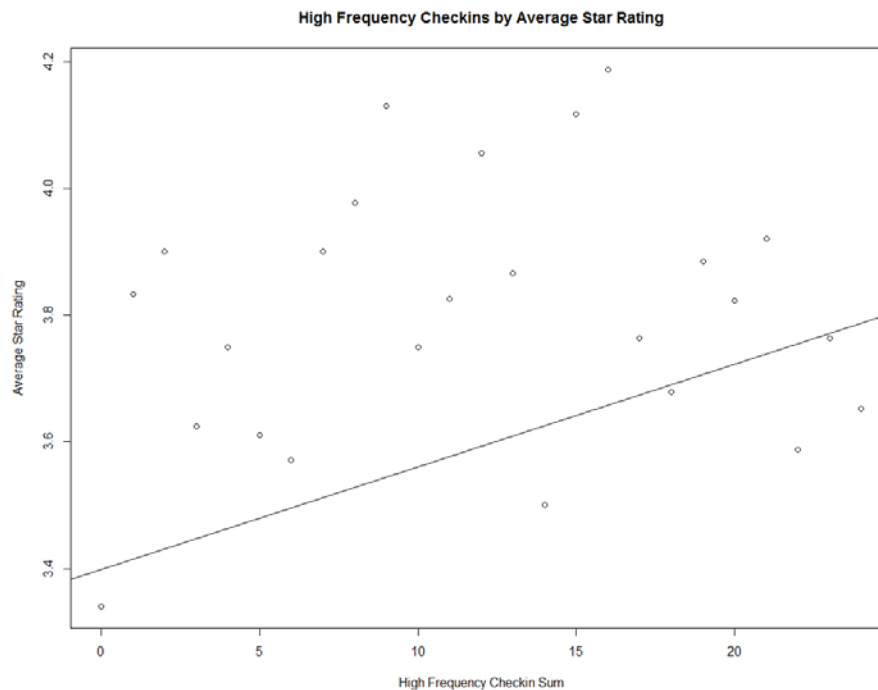

Regression Coefficients: *lm(stars~hrsSum,data=restaurants)*

	Estimate	Standard Error	t-value	p-value
(Intercept)	3.399064	0.024369	139.482	<2e-16 ***
Checkin Sum	0.016161	0.001919	8.423	<2e-16 ***

From these results we can conclude that there is a relationship between the total number of high frequency checkin hours that a restaurant is open and their average star rating. In order to conduct the next stage of the analysis, a subset of this restaurant data was pulled to focus on only those restaurants that had both a below average star rating and a below average checkin sum. This results in a subset of 576 restaurants.

For these remaining 576 restaurants, both the reviews and tips associated with these restaurants were pulled from the original dataset. The reviews were then converted into a corpus for the purpose of conducting a text mining analysis. Using the "NGramTokenizer" function from the *R Weka* package, the text from these reviews (N=6,916) were constructed into various n-gram models from 1 to 5 word sequences. The resulting models were analyzed for high frequency words and phrases that addressed issues regarding the hours of operation. Unfortunately, none of the high frequency words/phrases from the n-grams model referred to the hours of operation issue that I originally expected to find. The following is a word cloud for the unigram model and, as can be seen, none of the highest frequency words are associated with the hours of operation issue (i.e. "hours", "closed", "inconsistent", etc.).

Word Cloud for Reviews
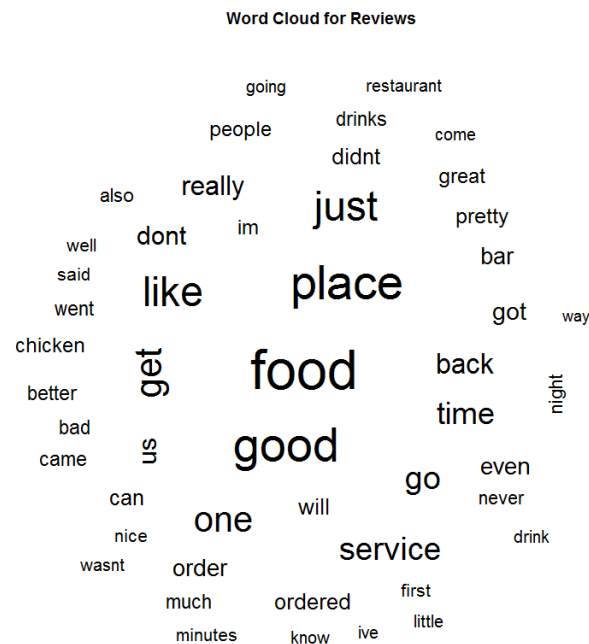
The same procedure was used to analyze the corresponding tips data (N=4,570). Once again the final n-gram models was analyzed to determine high frequency words or phrases that addressed the hours of operation issue, but just as was the case with the review text, the highest frequency words and phrases from the n-grams models (1 to 5-grams) did not address this particular issue. Words and phrases such as "hours", "closed", "inconsistent", "inconvenient", and so on, were not common themes in the final n-grams models.

Word Cloud for Tips

Furthermore, an attempt was made to make star rating predictions for each business based on the results of the n-grams models from both the reviews and tips but the results were inconclusive. The text from both reviews and tips mention many different things that cannot specifically be used to make predictions on that businesses average star rating, at least for my particular sample of the data.

DISCUSSION

Although we can conclude that there is a relationship between the total number of high frequency checkin hours that a restaurant is open and that restaurants average star rating, we do not have sufficient evidence from mining the text patterns in the reviews and tips to create a prediction algorithm based on the presence of specific words or phrases in the text. One potential reason for this is that there are other factors at play that determine whether a business receives a below average star rating and these other factors may be associated with the hours of operation issue or have precedent over it.

Another potential issue that I could be dealing with is that I only analyzed 5% of the original business sample. I chose a 5% sample due to processing speed and difficulty with running an analysis on such a large dataset. It is possible that with a full sample we would have much better luck at identifying the key words or phrases needed to create a prediction algorithm.

Future studies should consider beginning with the text mining step and conducting this analysis on the entire dataset. Once specific words or phrases are identified as common themes, then the corresponding factors in the dataset, where available, can be tested in a regression model to determine their relationship to the outcome variable of interest, which in this case was the average star rating.