Evaluating Local Business Activity With Large Review Data Sets

Alan Chen[1*], Aaron Agarunov[2*], Julian Tsang[1*]

[*] Macaulay Honors College at [1] CUNY Baruch, [2] CUNY Hunter

[1] {alan.chen3, julian.tsang}@baruchmail.cuny.edu; [2] aaron.agarunov38@myhunter.cuny.edu

Abstract

In the global economy, businesses rise and fall based on brand reputation and competition, among other factors. Companies exist in large networks where their practices tend to influence their neighbors. These companies can be complementary or competitive as they can take on similar roles and locations, and receive different types of attention. In an effort to better understand these interactions, our paper studies how businesses survive in their local communities by observing large review data sets from Yelp. From an analysis of star ratings in reviews, we typically observe a negative linear trend between a business's Yelp review rating and its time of operation (decreasing positivity). In addition, we determine whether one business's ratings affect a nearby business's activity by examining rating performance and business review language. In searching for a correlation between business's ratings and business competition, we find that the earlier that users mention competitor businesses in their reviews, the higher the business's ratings tend to be. Our analysis begins to explore the patterns of a typical business's Yelp rating life cycle and gauge the impact a mention of a competitor in a user review could potentially have on the review rating itself.

1. Introduction

Since its inception in 2004, Yelp has grown to become an immensely popular review site, visited by customers and businesses alike. Yelp allows users to review a business with a text response accompanied by a rating between 1 and 5 stars. Each business's webpage features a list of attributes (e.g. pricing, hours of operation), reviews, and images. A business's star ratings and text reviews on Yelp are the two most significant metrics that potential customers use to determine if the business is worth visiting. In response to Yelp's popularity, many researchers have pursued analyses of the patterns and trends that might determine the factors that affect a business's ratings and text review content.

Our initial data is provided by the Yelp Academic Dataset (Challenge Round 7) [6]. Many of the highest rated businesses on Yelp join the service between 2005 and 2007. The amount of user ratings slowly increases at different rates, but the rating quality is typically inconsistent for individual businesses, and can begin at high or low extremes. At a certain stage of the Yelp review cycle, users tend to assign less 4 or 5 star review ratings and the average review embarks on a steady decline. To explain the presence of a tipping point from which reviews begin to falter, we first identify its occurrence among businesses in multiple cities. We focus on heavily populated American cities, namely Phoenix, Arizona, and then compare to smaller cities like Pittsburgh, Pennsylvania and Charlotte, North Carolina.

There are many problems with our approach to the Yelp corpus. In our paper, we choose to observe the factors we have quick access to. As a result, we cannot account for external stresses that might cause a change in activity for a city or business during an unforeseen change in a particular time period (for example, due to weather). In the broader spectrum, there are dozens of contributory elements that give rise to changes in Yelp popularity for a business. Yet, our paper aims to identify the causes that contribute most to the change, offering an explanation or methodology to changing trends on Yelp.

We idealize that a popular business experiences a gradual increase in their review count and activity on Yelp. We expect the increase in activity to reach a ceiling set by that business' characteristic attributes. Their business model may or may not create a proper environment for success; Yelp characterizes businesses that failed to shortcomings as "closed," though these businesses are not accounted for in our data. To address our concerns about business futures, we aim to observe the tipping point in the reviews over time at which the average ratings begin to decline. Competition is rampant in large cities and undeniably causes a consistent shift in consumers. It is not uncommon to find *regulars,* customers that remain loyal to a service distributor, to certain establishments. Our data might reveal how everyday consumers make choices between service providers by having multiple experiences with local businesses in the past. By observing how early a competitor is mentioned in the review text, we are inferring that the comparison of two businesses is contributing to a user's decision to make one choice over another.

In Section 2 we describe previous related works and how they improve on interpretation of the Yelp corpus. Section 3 presents the tools, applications, and statistical techniques used to examine the data, as well as how our proposed method fared in outputting useful data. Section 4 provides the output of our methods as rendered by Google Sheets and Microsoft Excel. Section 5 is used to explain the results and identify the trends in our data relating to review ratings over time, and correlation between review ratings and how early a competitor is mentioned in the review text.

2. Literature Review

In 2011, a New York Times article by David Segal brought attention to the idea that Yelp reviews can be easily fabricated. Many people have began services to sell fabricated reviews for businesses; this activity unfairly disrupts the Yelp review rating data. Although Yelp has taken steps to isolate fabricated reviews, Yelp has also removed genuine reviews as a result [1]. Yelp's internal maintenance of their reviews will likely disrupt the following literature and our work. Yelp does not share their source code or algorithms on the review filtration, which makes it difficult for us to account for all legitimate reviews. We are unable to address this issue and cannot make steps to resolve fake reviews, but it is something to consider when concluding our paper.

In a 2012 study, Michalis Potamias concludes that the first reviews of a business tend to be highly rated and begin to steadily drop in rating over time. Specifically the first review in Potamias' study averaged at 4.1 stars, the second review averaged at 4.0 stars, and the 20th review averaged at 3.7 stars [2]. Although Potamias was not searching for a tipping point for review ratings over time, his research provided insight into how review ratings generally aggregate and behave over time. Potamias' study does not focus on the tipping point and as such his data was extremely selective, choosing only the businesses that ran a Groupon deal at least once. Potamias used the Yelp API and found that his data collection ran into missing fields in roughly 23% of the reviews. Our research will show a more complete comparison of review ratings over time without singling out Groupon associated business and by utilizing data provided directly by Yelp.

In a 2015 study, Jao-ke Chin-Lee notes that “general” or “all-purpose” restaurants such as Chipotle or Flat Patties were found to have less review ratings over time, while niche businesses did not [3]. Although this observation was applied with a different framework, it is in line with Potamias’ observations, concluding that review ratings tend to drop over time. Chin-Lee employed the Yelp Academic Dataset. We did not focus on niche businesses, but certain businesses with a shorter operation time in our data tended to not have decreasing review ratings over time. We noticed that businesses with longer operation times tend to have decreasing review ratings over time. A hypothesis could be that restaurants eventually shift from their niche specialization to a “general” or “all-purpose” category once the consumer base expands by a certain factor. We note that Chin-Lee made his final conclusions for restaurants. However, much of our data observations include businesses in non-restaurant categories (e.g. nail salons).

In a 2014 study, Hajas, Gutierrez, and Krishnamoorthy concluded that restaurants near surrounding colleges experience a cyclical pattern in their published reviews. The dataset used by Hajas et al. encompassed 30 colleges, all of which seemed to align with a cyclical pattern. Although the paper does not resolve on a concrete conclusion to explain the cyclical pattern, it does support our findings of such patterns in businesses that we did not partition by college campuses [4].

3. Methods

Data Sets

Yelp provides a large corpus totaling to 2.2 million reviews and 77,000 businesses. Yelp assigns a value for the business_id variable to each review and business listing. We matched the 2.2 million reviews with the business dataset (after both were converted to a comma delimited format) and adjusted for missing values by aligning the business_id values. After merging, we can correct data outside our area of interest, such as entries in Arkansas (AR) meant for Arizona (AZ), or entries around Phoenix but within Scottsdale or Glendale, AZ. We measure the quantity of data provided for each city and perform basic visualizations such as the distribution of star ratings, price range, and how different cities compare in terms of attributes. We choose to first explore the businesses in Phoenix, Arizona to maximize the data we can study from small samples and neighborhoods. After sorting by the city variable, we use latitude and longitude coordinates to geofence the data using Zillow shapefiles and Google Maps neighborhood distinctions. Both services identify the boundaries to neighborhoods in major cities. We use our retrieved boundaries to extract separate data on Phoenix's urban villages, namely Central City, Encanto, and North Mountain. In addition to Phoenix, we worked similarly on Pittsburgh and Charlotte and their respective neighborhoods. Our paper aims to determine if regression analysis is consistent in each city between local regions and why that may be so. Businesses might operate at very competitive locations but offer unique services and attributes, yet experience the same average review rating dropoff that less successful businesses do. The variety in business composition poses a challenge to simple regression analysis and correlation plotting.

SAS, R and Excel are used to read original JSON files ported to the .csv (comma delimited) format. SAS and R are repeatedly used to create subsets of the data for faster processing. Each subset divides the data by a number of included or excluded variables.

Data Organization

To bring focus to the study, we first analyze the relationship between businesses over time. We sort the individual businesses by the number of reviews they have in descending order, allowing us to identify the businesses with the most reviews, which we speculate are the most impactful in their respective region. We quantify the time of operation for a business by the amount of quarters in a year.

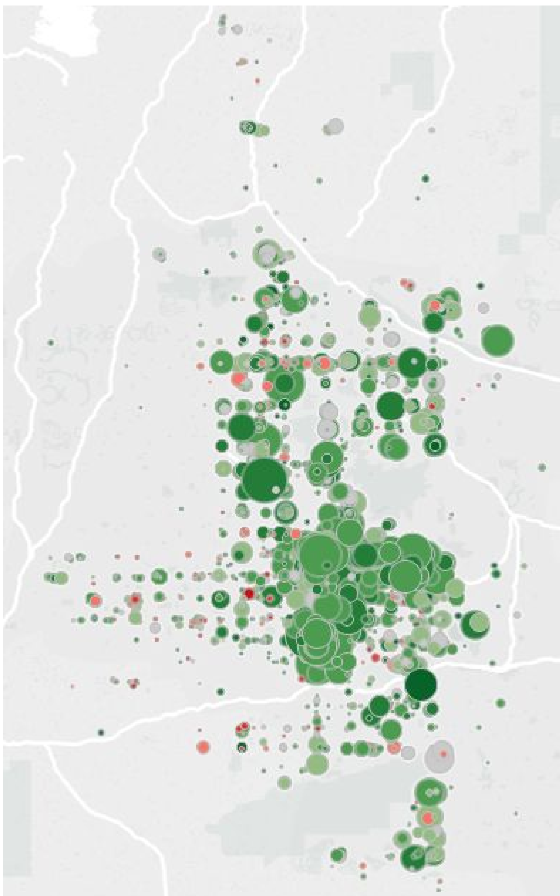
Using our initial dataset on Phoenix, we use Tableau software to generate a time series heat map (Figure 1.0.1 & 1.0.2) that show how review activity and ratings change over time. We notice how most activity was centered around specific neighborhoods. By observing our heat map we were able to select the the most active neighborhoods to include in the study, namely Encanto, Central City, and North Mountain.

We created multiple line graphs of review activity per month for each business and compared them. This allowed us to identify any trends and unique occurrences between the review activity of each business in relation to each other.

To understand review quality, we characterize a “positivity” metric. Positivity is based on how many 4 and 5 starred reviews a business has in a quarter relative to the total number of reviews in a given quarter. By graphing our results against quarters, the positivity metric allows us to observe if a business received more or less 4 and 5 star reviews over time. The trend line is useful for predicting the course of business' activity as seen by Yelp. If, for example, the trend is significantly negative, we have a metric to predict if a current 4-star establishment might end up decreasing by 0.5 stars within a set time period. The illustration of positive and negative trend lines is an introduction to quantifying how Yelp reviews and star ratings vary over time and compare in local areas.

We employ brief natural language processing (NLP) to observe how review ratings are affected by the mention of a competitor. The method consists of searching for language where a reviewer references a competitor in their review, and observe whether the reference is made with positive or negative implications. We restrict our search of businesses and reviews within the same neighborhood. We search for reviews that reference the most highly reviewed businesses in each neighborhood because we believe that they would have the most impact on their competitors and neighboring businesses. If a user mentions the name of a competitor in the review text, then they are likely performing a comparison, which would describe a rating impact on the business being reviewed. We applied SAS to search for names of the top five most popular business in a neighborhood. If the name of the business was mentioned, we would export the data to Excel. With Excel we created bar graphs to compare the frequency of each rating (1, 2, 3, 4, 5) and the exact location of a mentioned competitor in the review text using a character index. We utilize this process for our initial Phoenix dataset and repeat the process for the two other cities along with their associated neighborhoods. Identifying a mention of a business solely by its name is difficult due to several factors. A business name can be a common word such as "Matt", which could make it difficult to assert if the specific mention of "Matt" is referring to a person or the business we are searching for. Users also may spell the business name incorrectly, which would result in missing records. To verify our data, we would randomly check records in the data on Excel to see if the mention of the keyword we were looking for is in relation to the business or not.

4. Results



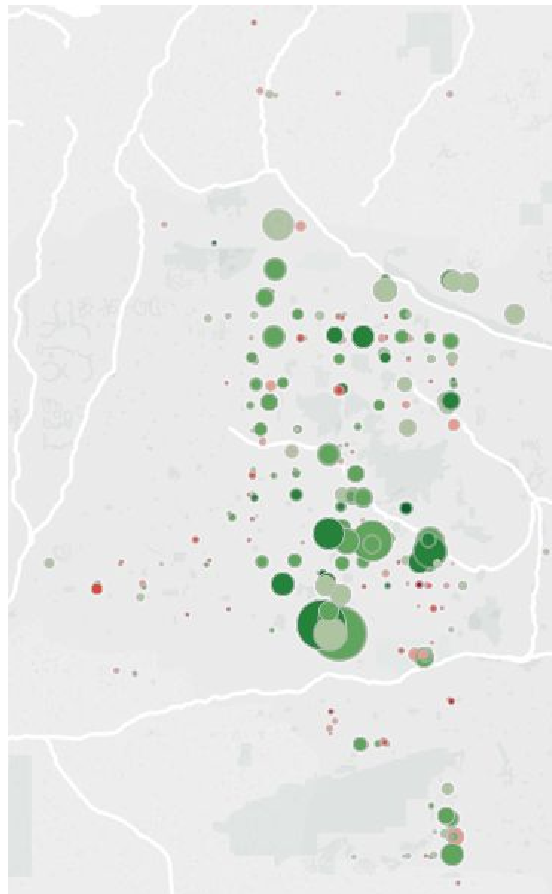


Figure 1.0.1 (left). Heat Map of Phoenix Business Activity (Size by review count).
The color of the circles indicate the review rating (1 → Red, 5 → Green) and the size of the circles indicate the number of review counts. Figure 1.0.1 allows us to determine which neighborhoods had the most review activity or largest cluster of bubbles. The neighborhoods we selected to compare are Central City, Encanto, and North Mountain.

Figure 1.0.2 (right). Heat Map of Phoenix Pizza Business Activity (Size by review count).
The top two businesses in Phoenix, AZ sorted by review count are *Pizza Bianco* and *Cibo,* both of which are pizzerias. Figure 1.0.2 identifies the same neighborhoods but is sorted by the pizza subcategory. These most reviewed pizzerias are very common in the three neighborhoods in question.

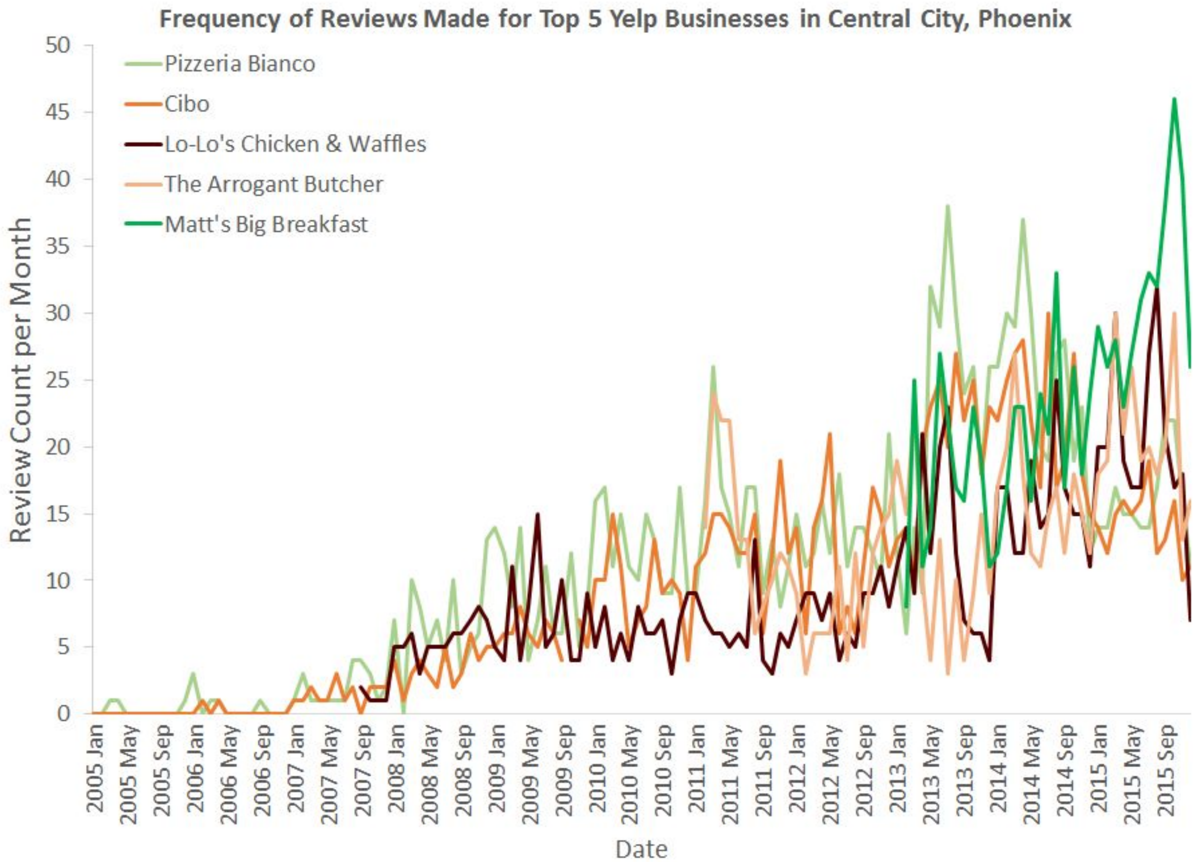


Figure 1.0.4. After identifying the top five Phoenix businesses in Central City, we graph the frequency of their user reviews per month. By simple peak observation, we note that businesses *do not* typically share peaks. One business tends to have the most reviews in a given month.

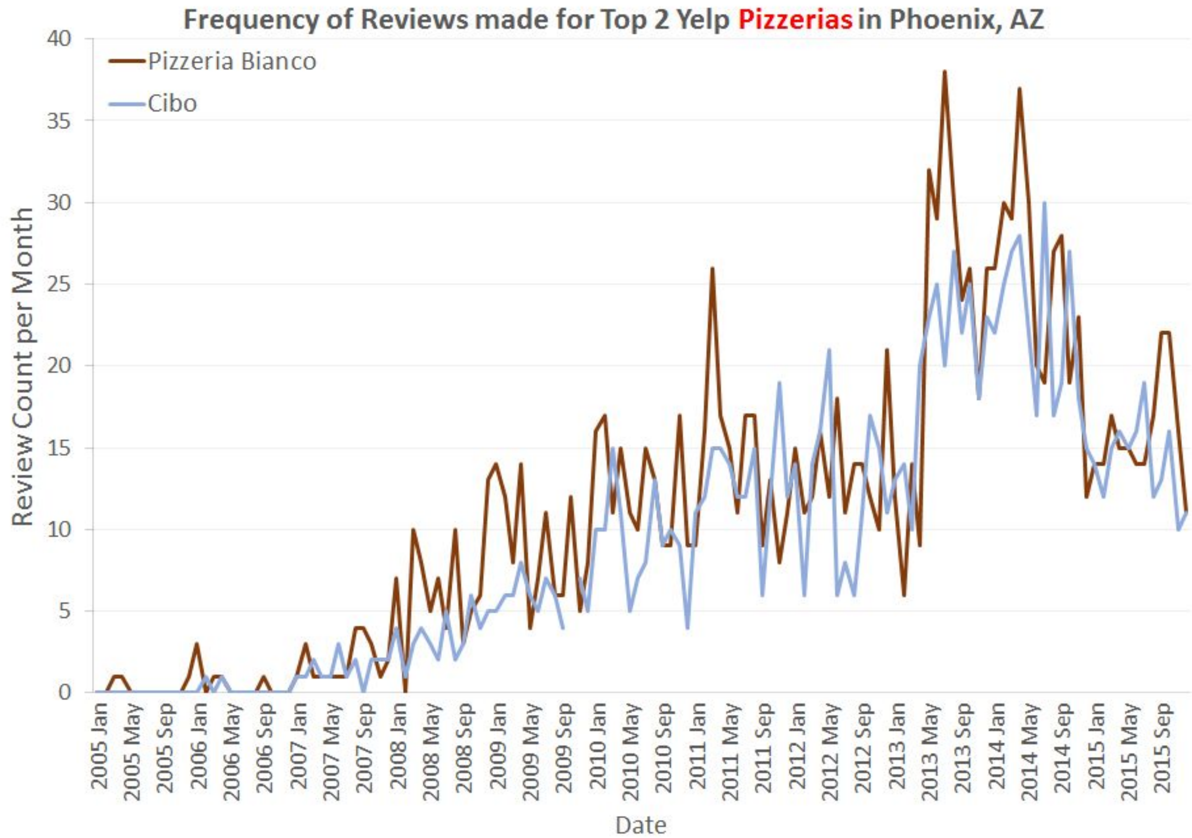
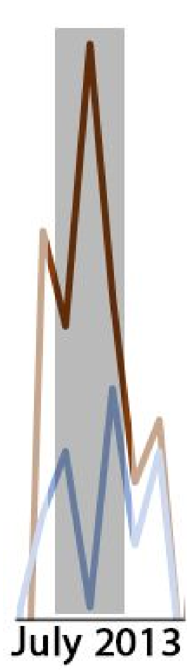


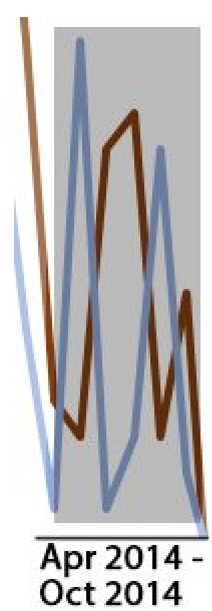
Figure 1.0.5. When we compare the review count trends for the top two pizzerias in Phoenix (coincidentally, these are also the top two Yelp businesses in Phoenix), we notice a frequent exchange in peaking activity.

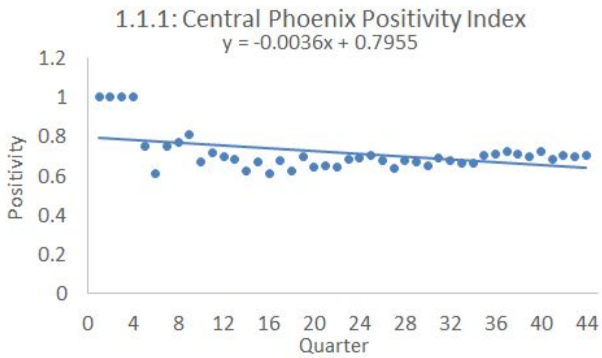
In certain months, when there is a peak for one pizzeria, the other pizzeria experiences a trough.

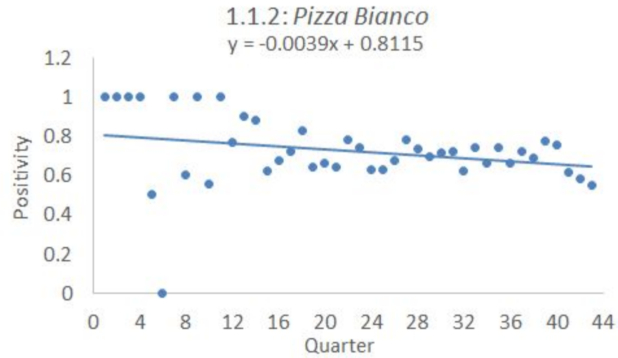
Example. In July 2013, Pizzeria Bianco experiences a peak while Cibo experiences a trough. We propose that direct competition between two pizzerias within 1.2 miles of each other is a viable explanation for the repetitive exchange of activity over 10 years of pizzeria activity in Phoenix [5]. If the consumer market stays consistent, the exchange describes how consumers shift between preferences over any given month.

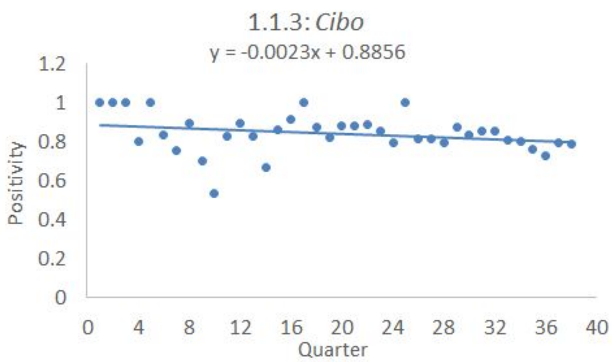
Figure 1.0.6 (right, one) & **Figure 1.0.7** (right, two) illustrate likely evidence of competition.

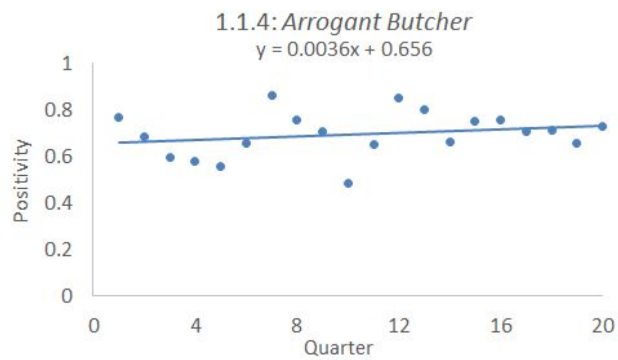


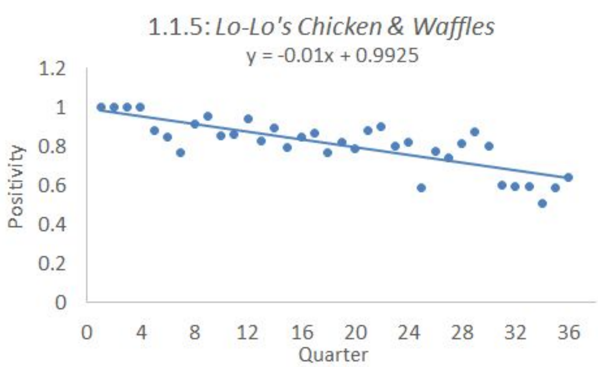


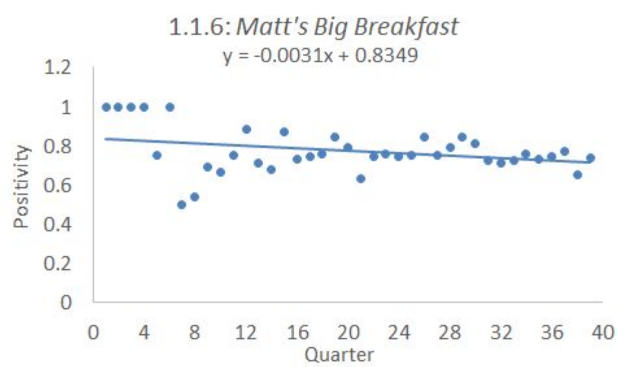












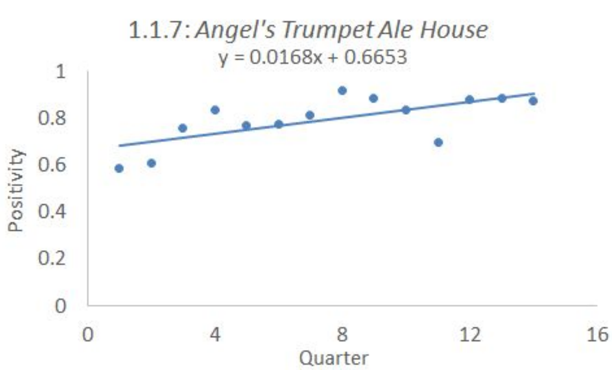


Figure 1. Phoenix, Arizona

Figure 1.1. Central City, Phoenix

Figure 1.1 (cont.). The review ratings in Central Phoenix are experiencing a decline: reviewers tend to provide less 4 or 5 star ratings over time. The more quarters a business has, the more likely it is to experience a decline in positivity (positivity is measured by the ratio of 4 and 5 star ratings).

Note. Though certain businesses do have positive indices (Figure 1.1.7), these businesses have existed for a much smaller total quarters than those with negative trends (Figures 1.1.2, 1.1.3, 1.1.5, and 1.1.6). We assume that if there was a tipping point from which ratings would drop, *Angel's Trumpet Ale House* might not have reached that tipping point.

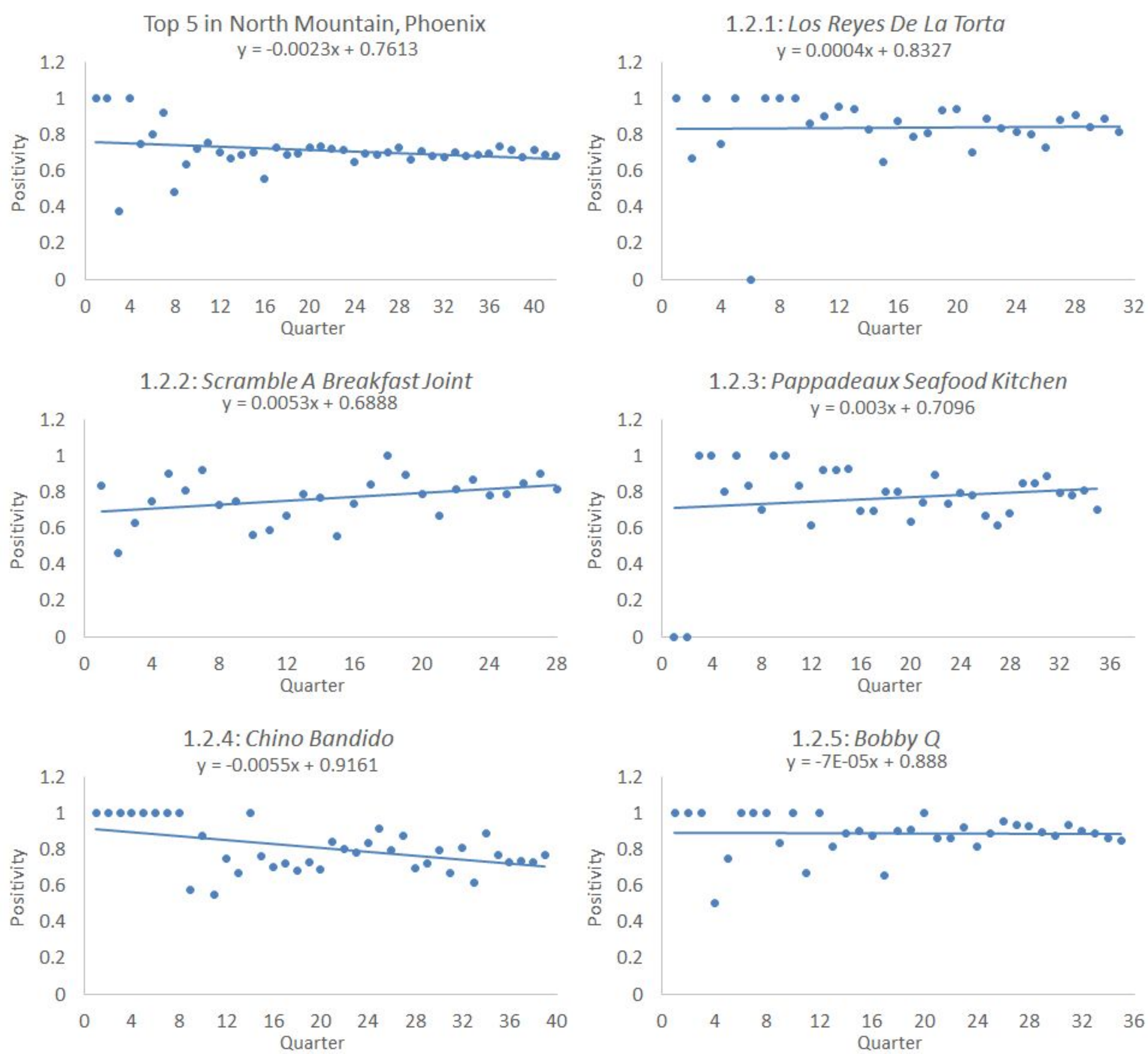
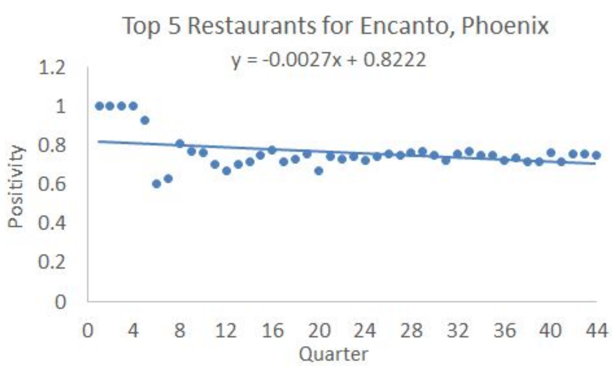
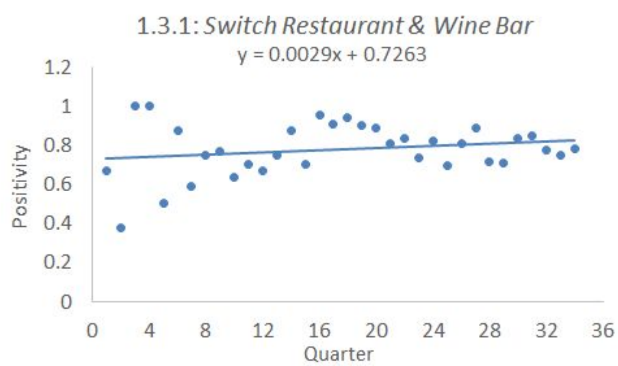


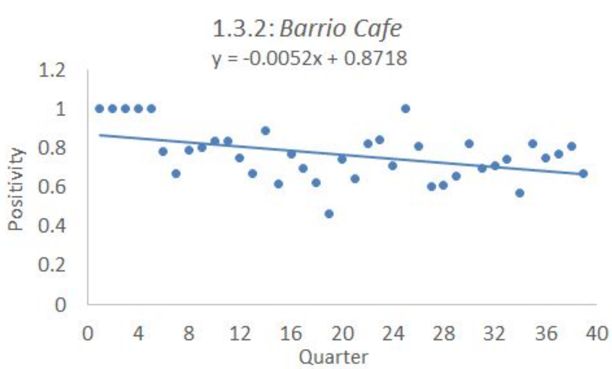
Figure 1.2. North Mountain, Phoenix

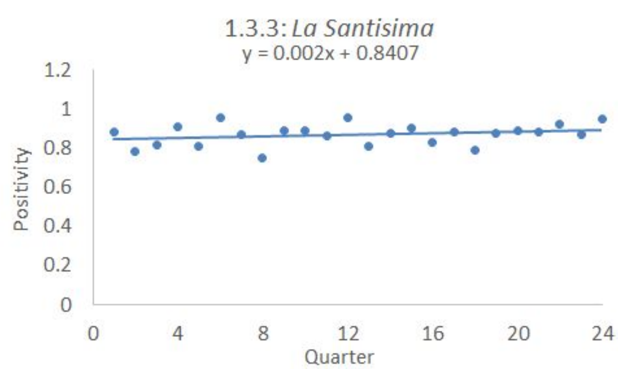
Figure 1.2 (cont.). The positivity in North Mountian seems to disagree with our findings in Central City. We notice that the most reviewed businesses in North Mountain do not experience a decline in positivity over each quarter.

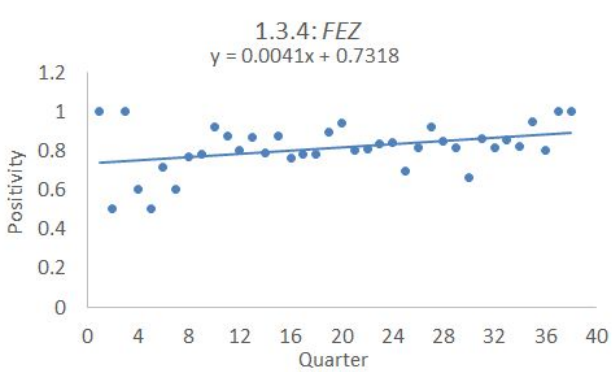
Whereas Figure 1.2.4 shows similar behavior to top businesses in Central City, Phoenix, Figures 1.2.1 and 1.2.5 have flat, horizontal trends. We account for the sudden change in trends by noting large variety in the total review counts used to produce the data. The top businesses in Phoenix have a review count range from 905 → 1613 whereas the top businesses in North Mountain have a range from 474 → 1224.











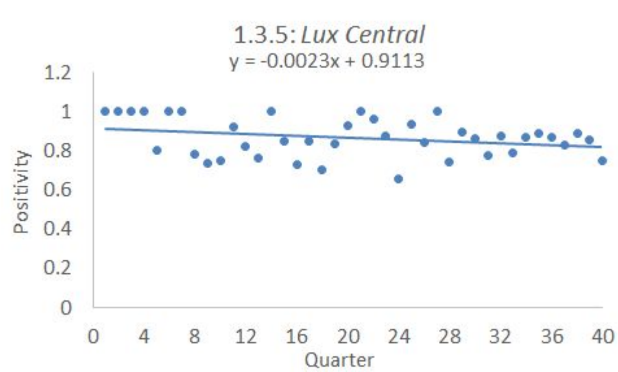


Figure 1.3. Encanto, Phoenix

Similar to the behavior in North Mountain businesses, results in Encanto seem to disagree with our findings in Central City. Encanto's business' review count ranges from 618 → 1302. It is important to note that the most reviewed establishment in North Mountain, *Lux Central* (Figure 1.3.5), has 1302 review ratings, a decreasing positivity, and been established for at least 40 quarters. These characteristics are highly similar to the businesses in Central Phoenix that display a negative trend, or decreasing positivity.

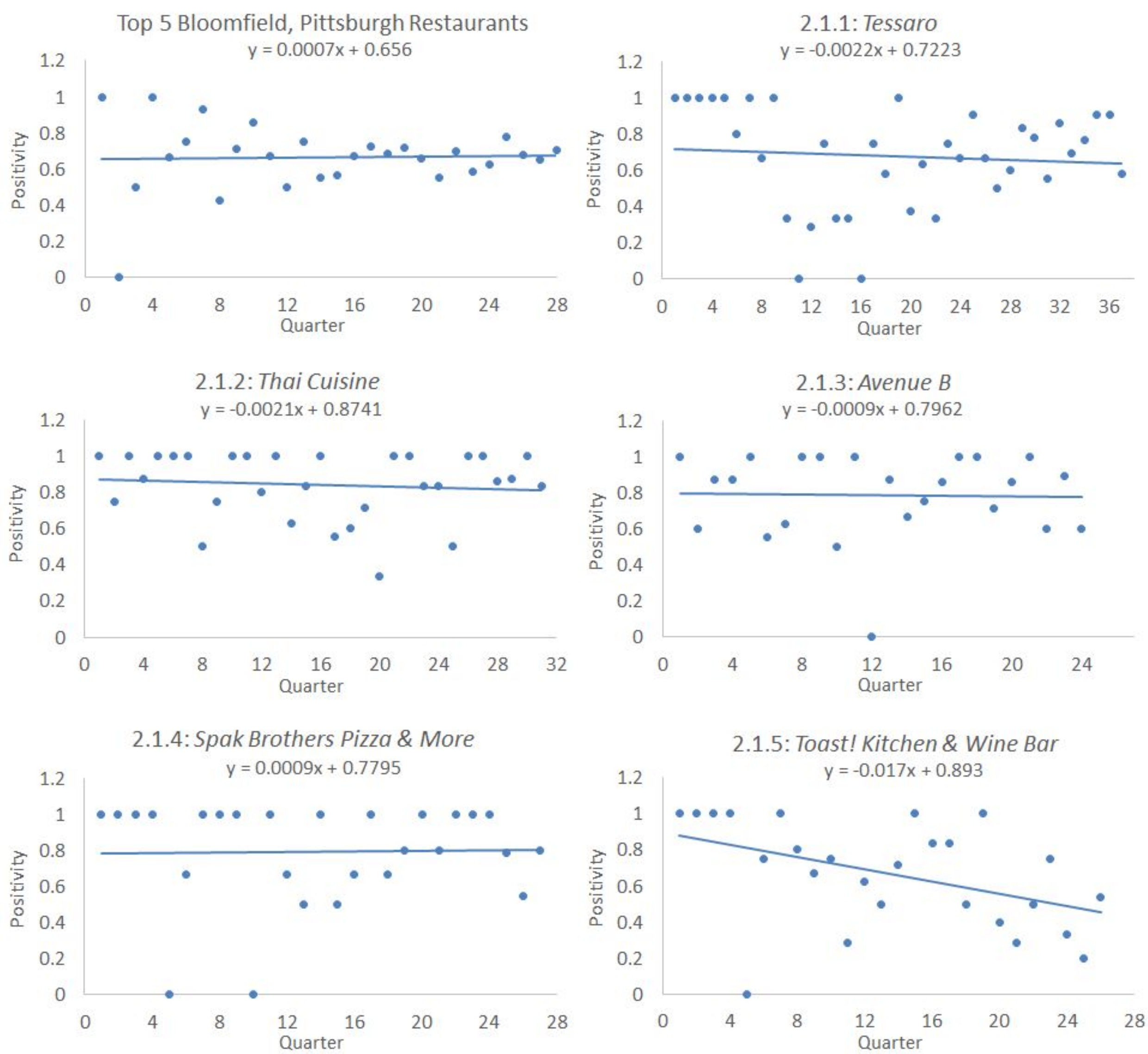


Figure 2. Pittsburgh, Pennsylvania

Figure 2.1. Bloomfield, Pittsburgh

In Bloomfield, Pittsburgh, only two of the businesses (Figures 2.1.1 & 2.1.2) support our proposed findings, but we believe this may be due to a lack of review numbers and that many of these businesses are still new and have not been reviewed by enough users. The review count of these businesses ranged from 145 → 268.

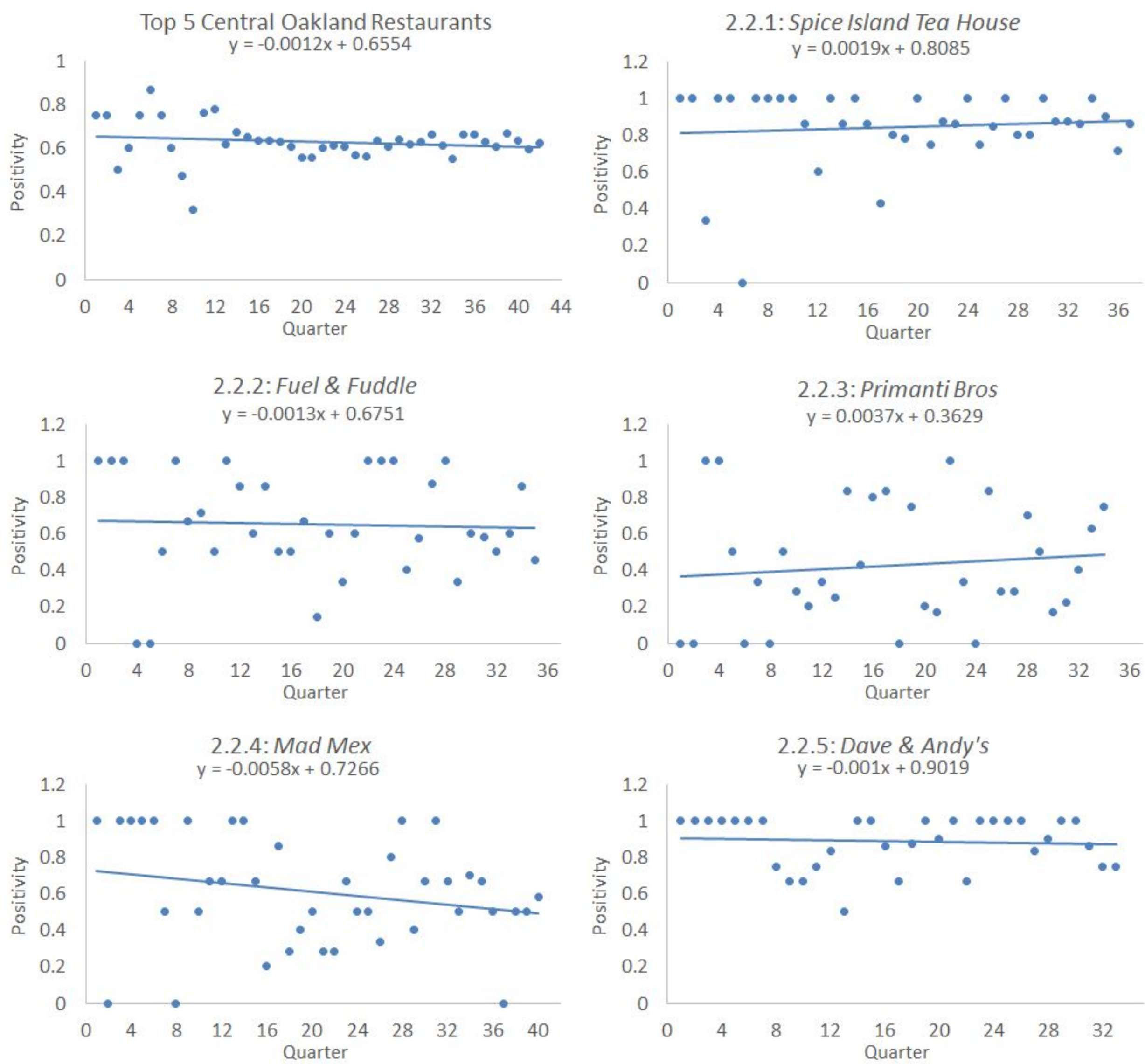


Figure 2.2. Central Oakland, Pittsburgh

In Central Oakland, Pittsburgh, only three businesses (2.2.2, 2.2.4, 2.2.5) supported our findings. There is a drastic difference between the review counts of businesses Central Oakland and the previous businesses we have examined. The review count of these businesses ranged from 180 → 230.

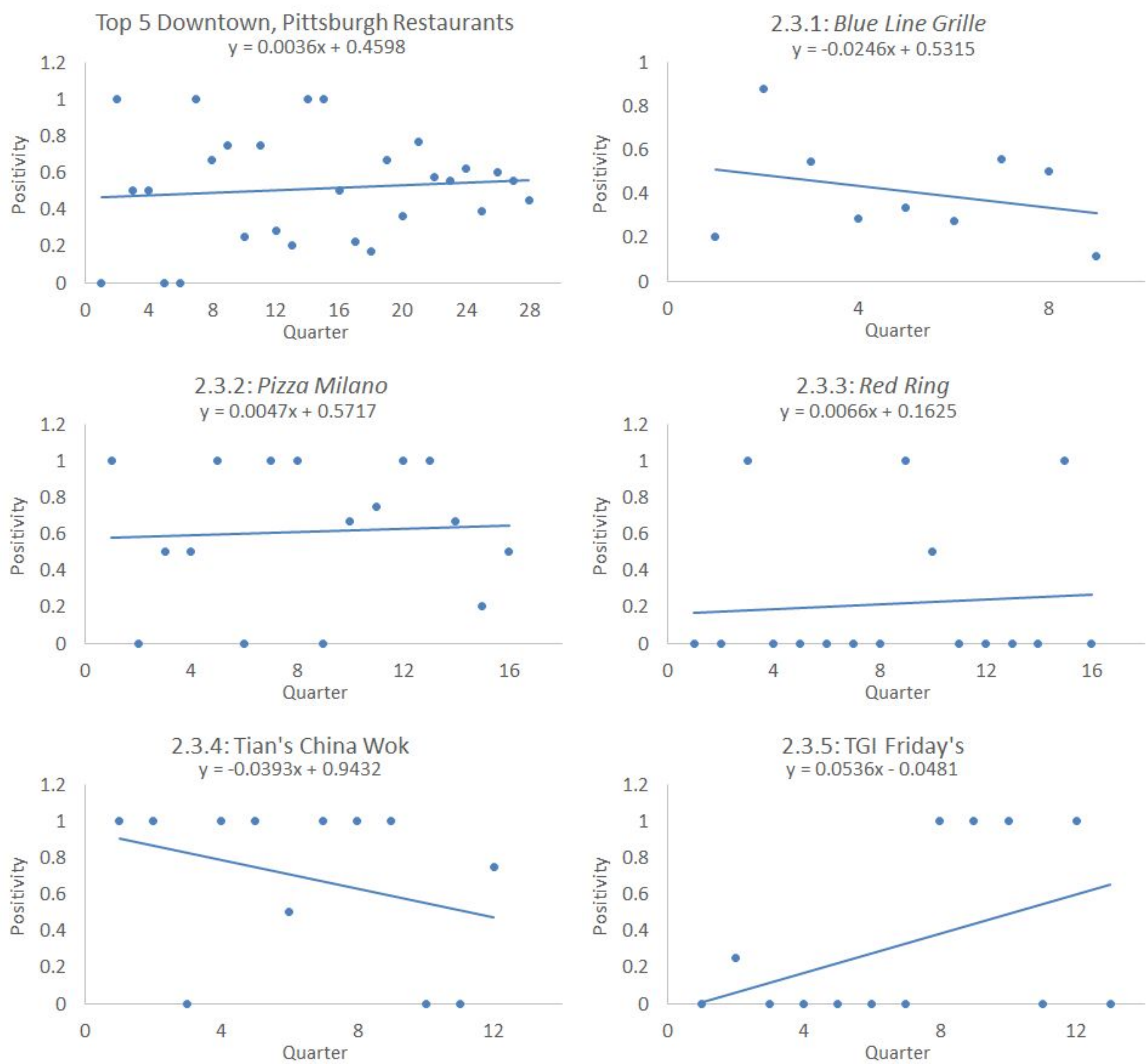


Figure 2.3. Downtown, Pittsburgh
In Downtown, Pittsburgh, the businesses we examine with the most review counts have at most 16 quarters. Businesses with less quarters listed on Yelp are more likely to see an increase in positivity over time. However, two businesses (Figures 2.3.1 and 2.3.4) have few quarters and decreasing positivity, which conflicts with our hypotheses.

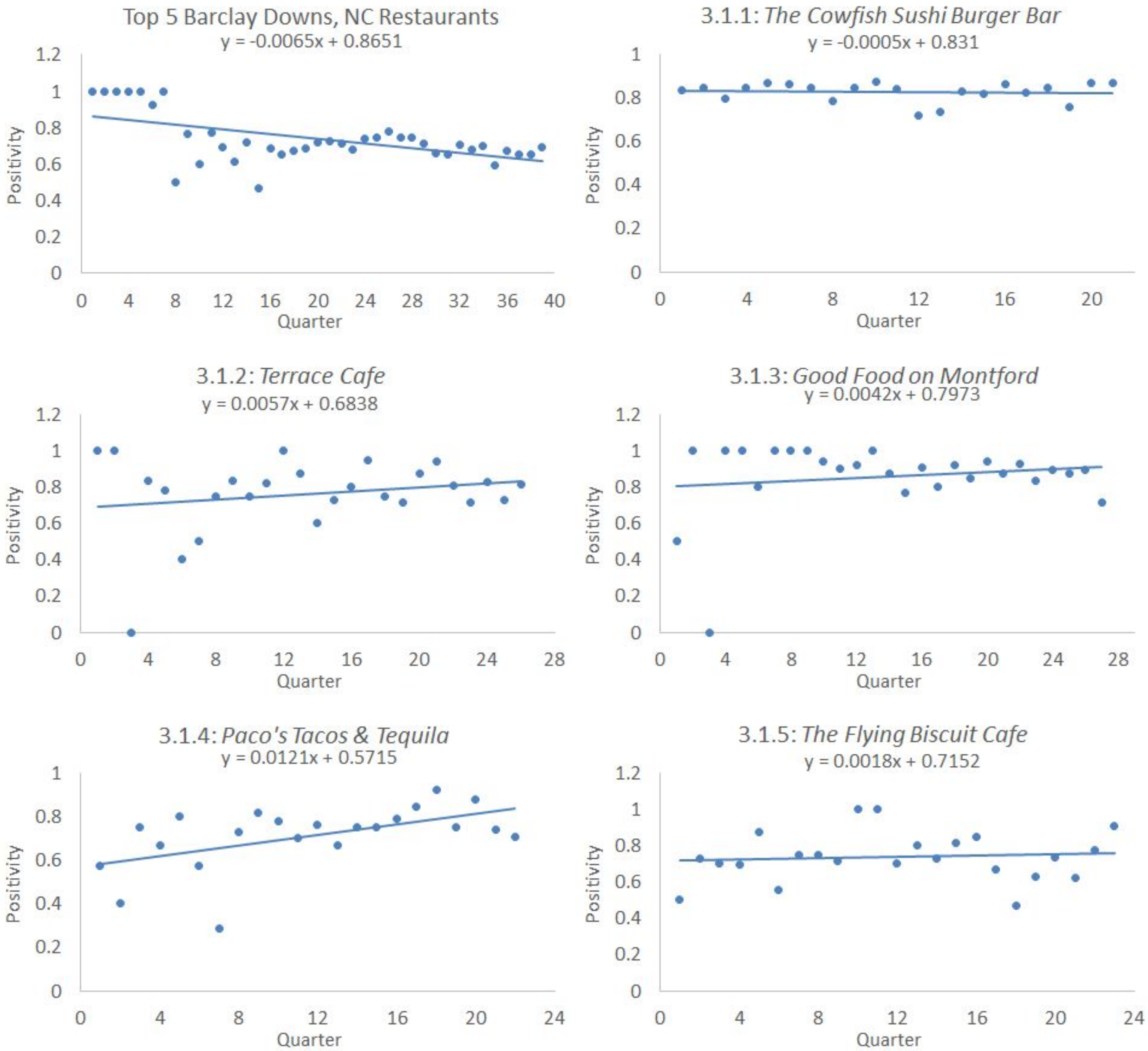


Figure 3. Charlotte, North Carolina

Figure 3.1. Barclay Downs, Charlotte

In Barclay Downs, Charlotte, none of the individual businesses directly support our analyses. The review count for these businesses ranges from 293 → 963. Among the businesses examined, the most quarters any business has is 28.

Note. The overall trend is negative, suggesting that when plotted over a larger period of time (i.e. >30 quarters), the overall trend overwhelms the majority five under study (Figures 3.1.2 - 3.1.5) and does suggest a decrease in positivity over time. This is further established by the Top 5 graph, which displays 40 quarters total and has a significant negative slope.

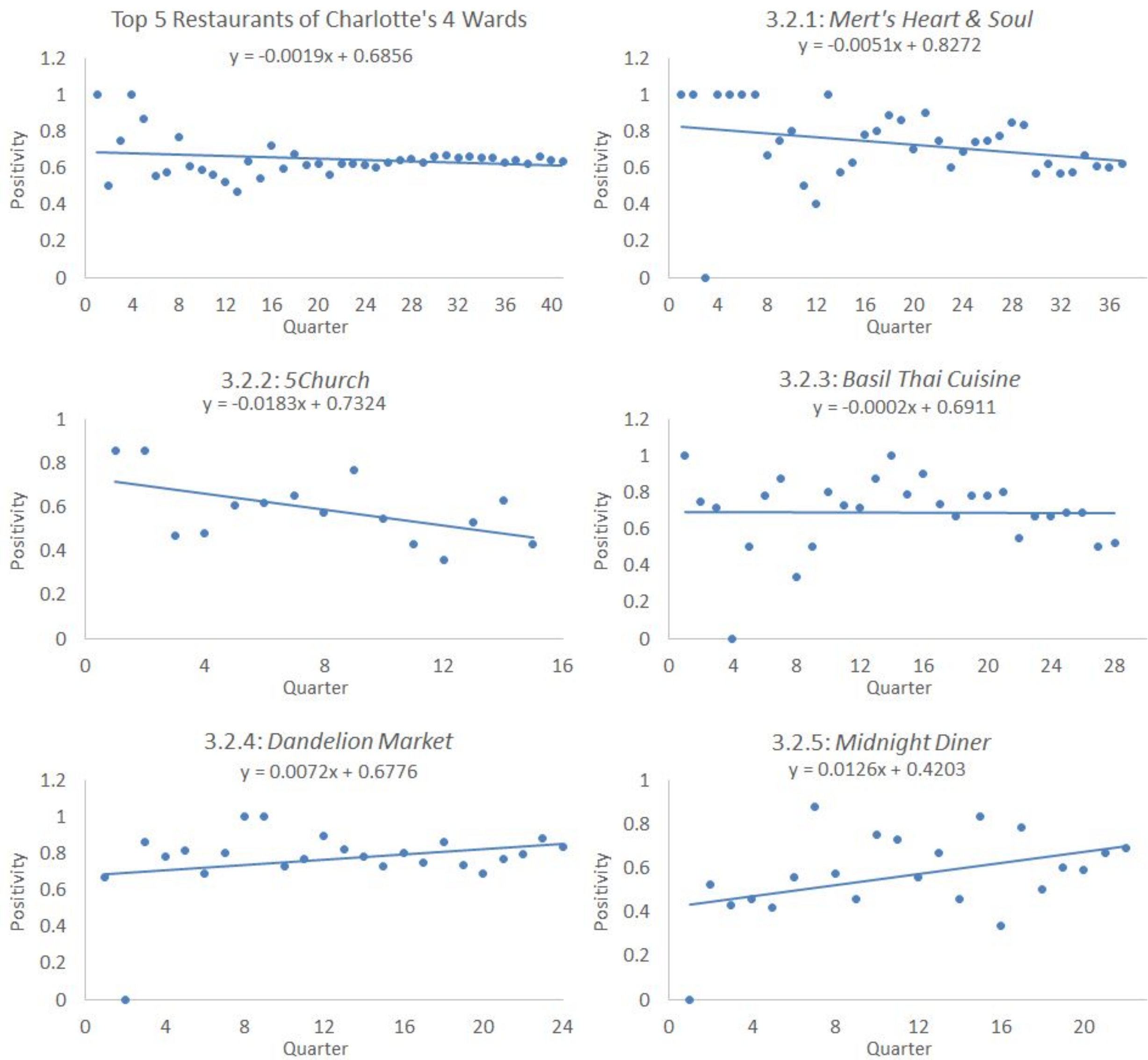


Figure 3.2. The Four Wards of Charlotte, NC

In Charlotte's Four Wards, one of the businesses (Figure 3.2.1) supports our proposed findings and has existed for more than 30 quarters. The review count for these businesses ranges from 307 → 582. The other businesses are relatively new and seem to be experiencing an increase in positivity. This is typical for businesses that have not been yet reached their tipping point.

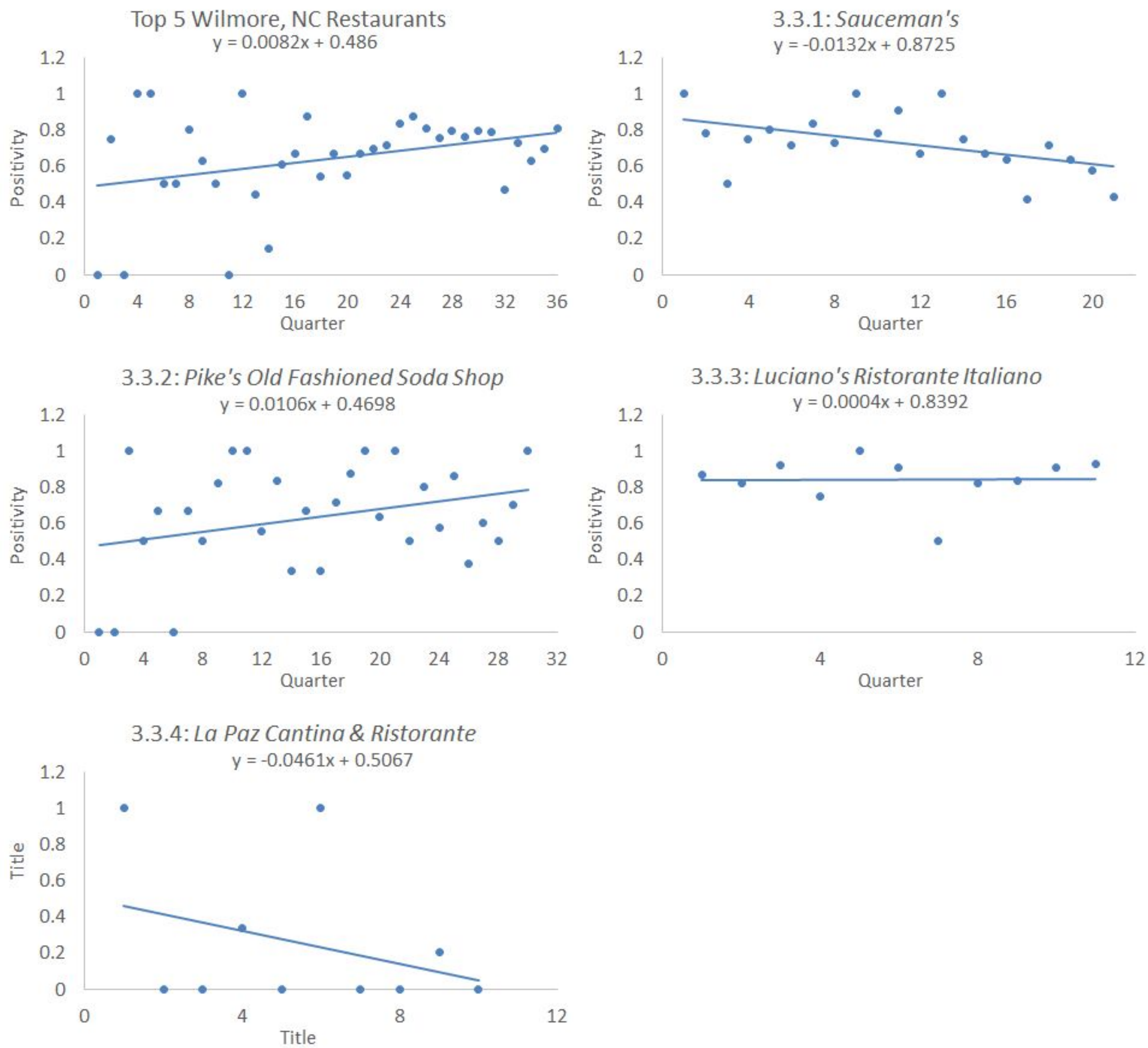


Figure 3.3. Wilmore, Charlotte

In Wilmore, Charlotte, none of the businesses seem to support our proposed findings. The review count for these businesses range from 21 → 181. It is difficult to quantify the data in this neighborhood because the review counts and quarters are not sufficiently great.

Figure 4. Central Phoenix, Arizona

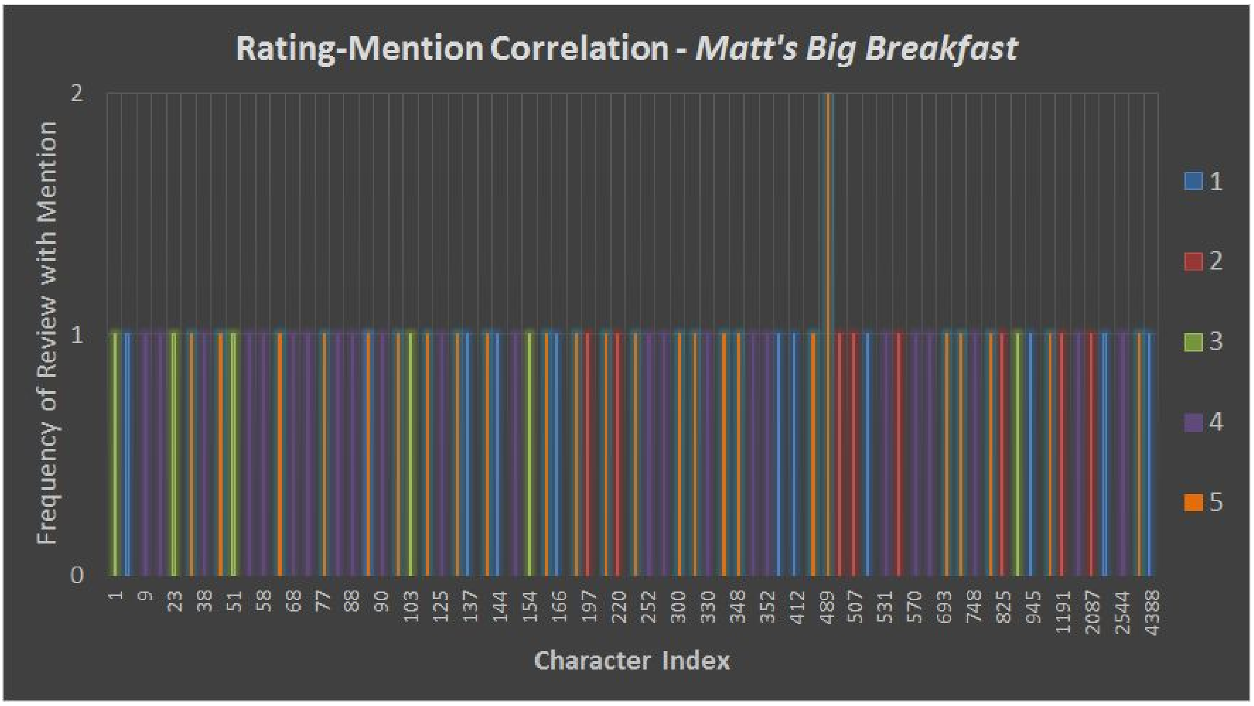


Figure 4.1. Matt's Big Breakfast Rating-Mention Correlation
Each of the ratings are color coded as shown. Each entry is a review that mentions the keyword(s) "Matt's Big Breakfast" in the review text. None of these entries are reviews about "Matt's Big Breakfast." The same holds true for the rating-mention graphs that follow.

Note. If a character index appears with a frequency of rating greater than 1, it signifies that the business (e.g. Matt's Big Breakfast) was mentioned in 2 reviews at the same exact character count location.

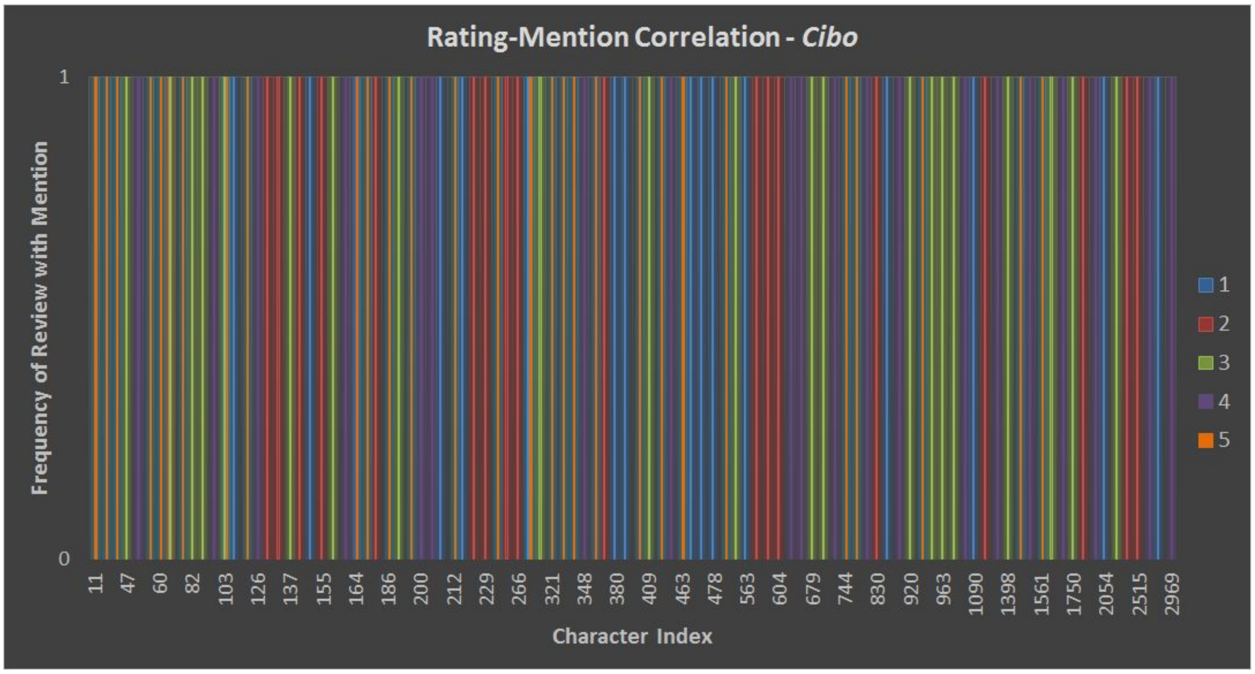


Figure 4.2. Cibo Rating-Mention Correlation
The five ratings are clustered before the 500 character count.

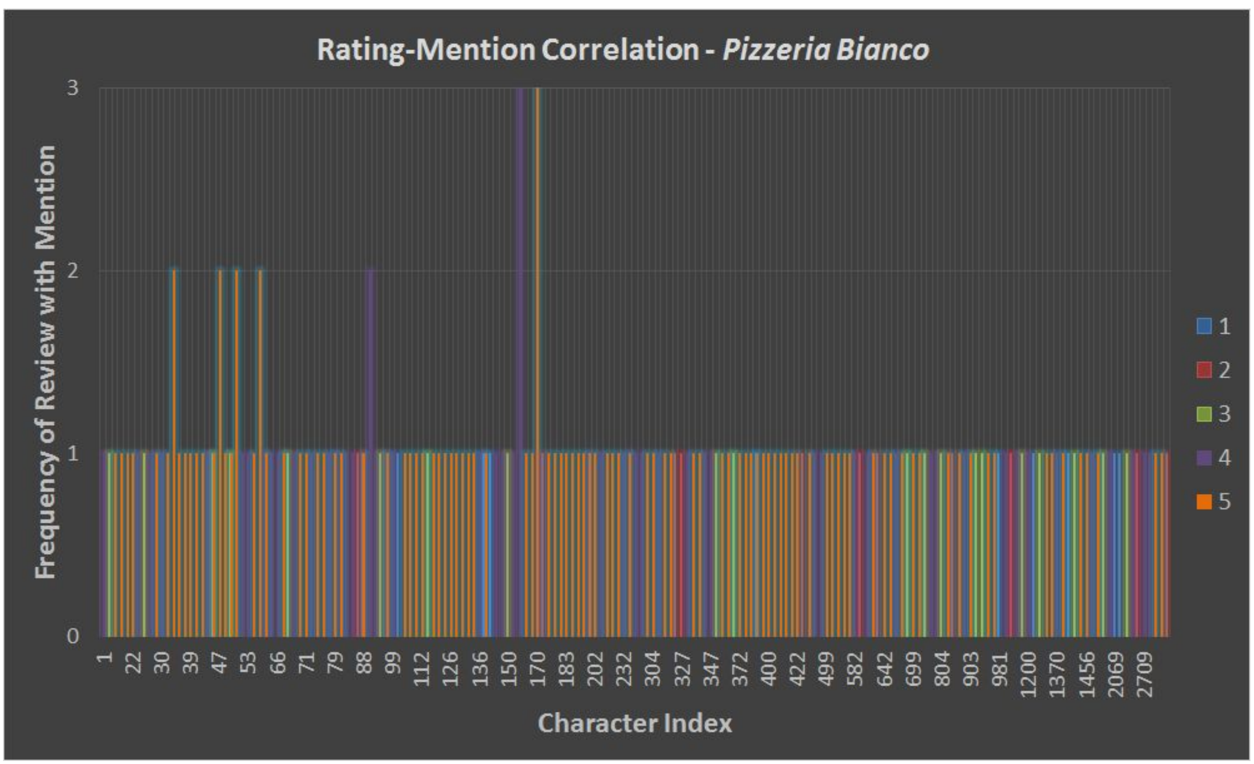


Figure 4.3. Pizzeria Bianco Rating-Mention Correlation
The clusters of five ratings are much more prevalent here, though it could be that Pizzeria Bianco has more five star ratings overall. There are several large clusters of five ratings prior to the 600 character count.

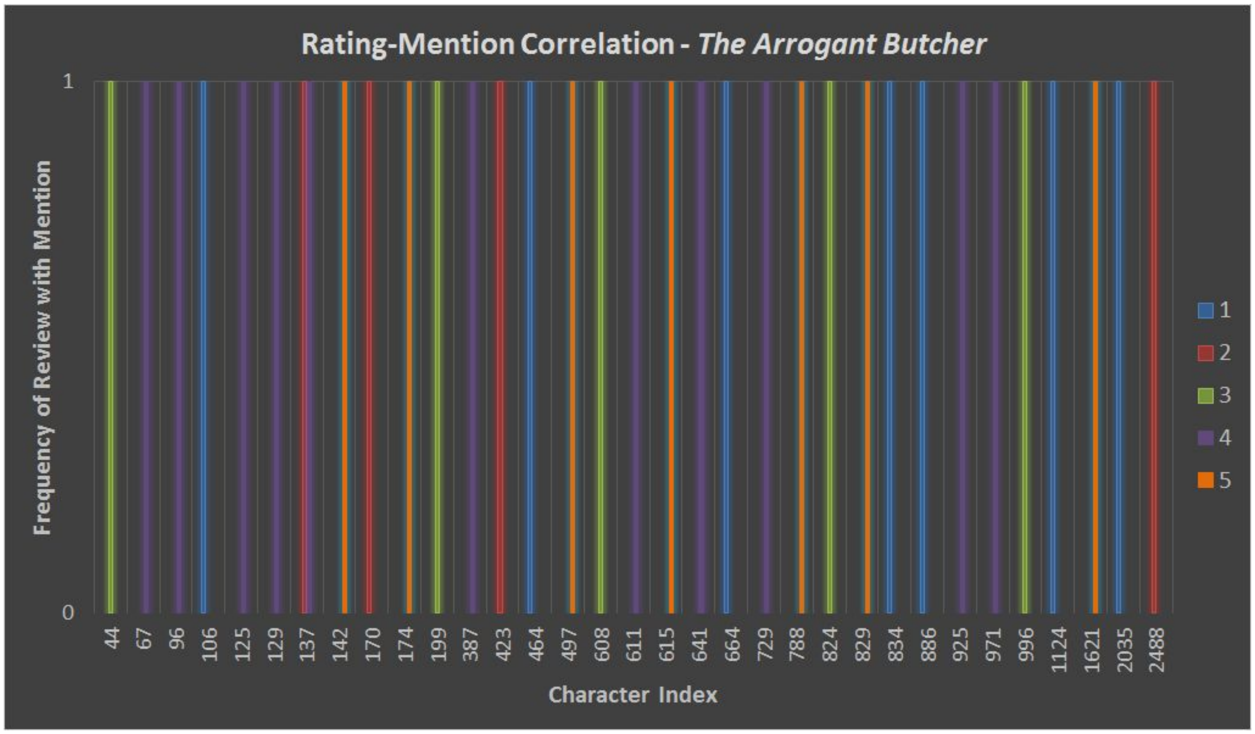


Figure 4.4. The Arrogant Butcher Rating-Mention Correlation

There are significantly less mentions of "The Arrogant Butcher" in the review data. Pizzerias may have more competitors in Phoenix than bars do and as a result will have more mentions in reviews overall.

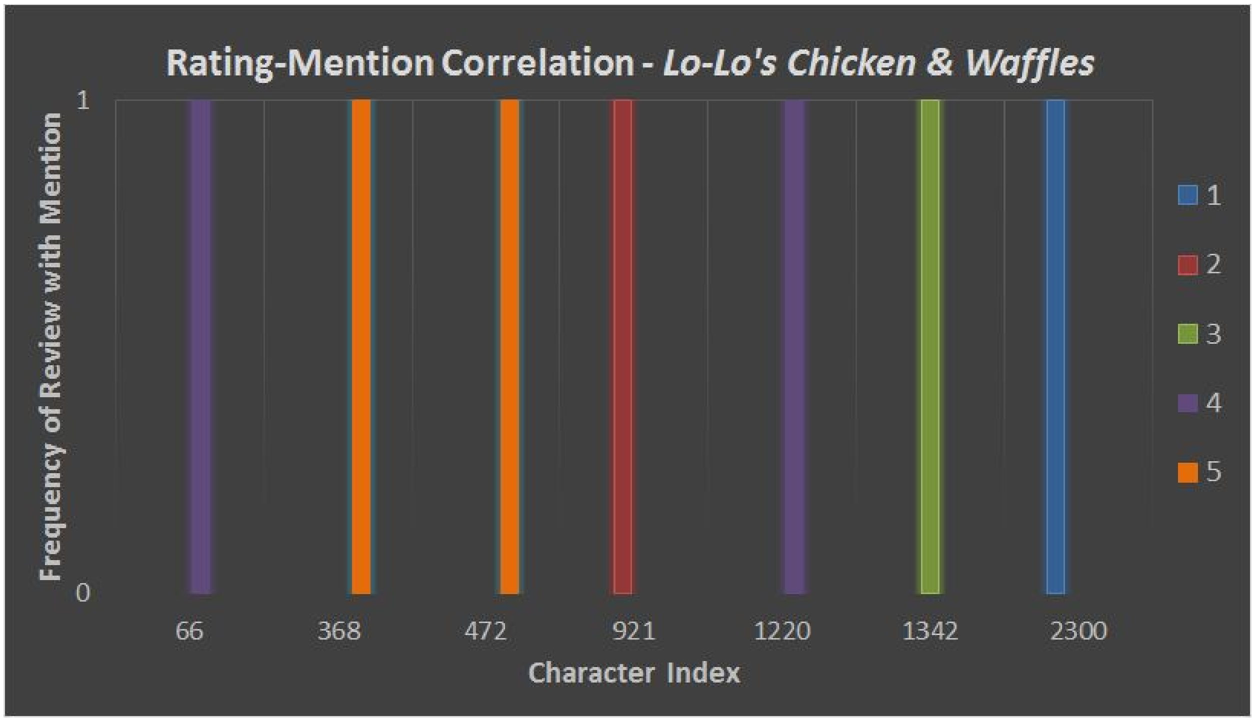


Figure 4.5. Lo-Lo's Chicken and Waffles Rating-Mention Correlation
There are significantly less mentions of "The Arrogant Butcher" in the review data. We cannot draw any conclusive information from this graph due to the lack of data available.

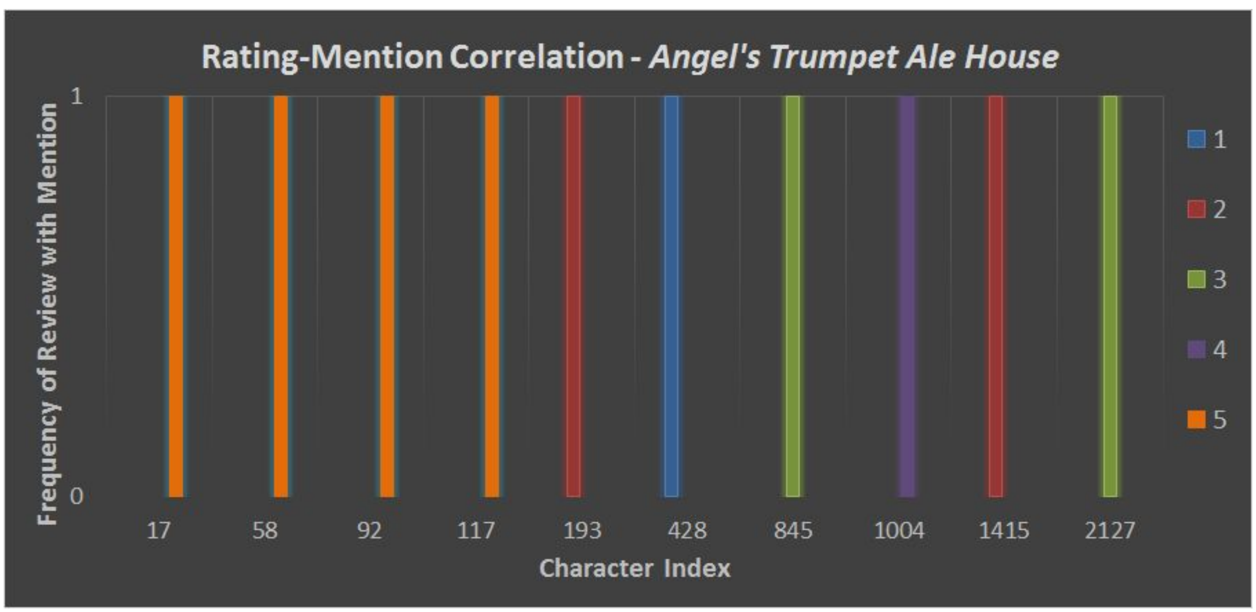


Figure 4.6. Angel's Trumpet Ale House Rating-Mention Correlation
There are significantly less mentions of "Angel's Trumpet Ale House" in the review data. When it is mentioned, we observe that the five rating are focused on early mentions.

5. Analysis

Warm Start Bias

Positivity index of a business is relatively higher early in a business's lifetime and goes down as time of operation increases. This observation is corroborated by the warm start bias thesis that was first introduced by Potamias [2]. His paper explains that a business in its initial stages gets exposure to a limited, niche audience that is more likely to give favorable reviews. Once a business reaches a certain level of general popularity, the audience becomes more diverse in taste and is more likely to give negative reviews. In the case of Central City, Phoenix, the positivity indices created show this phenomenon for businesses operating for over 30 business quarters. Businesses that are opened for less than 30 quarters tend to have a positive linear regression in positivity over time, meaning that the number of 4 and 5 star reviews relative to total reviews grows over time. This alternate scenario in which positivity increases over time can be explained by the fact that the businesses are still operating within their initial stages, still catering to audiences that are more receptive to them. There are some businesses that have operated for more than 30 quarters, but continue to display an increasing positivity. These businesses tend to have a much lower review count than those with decreasing positivity, which may point to the idea that the business is still serving a niche audience.

Seasonality

Although most positivity indices exhibit a negative linear regression, the points in the index charts show patterns of seasonality. Seasonality has an apparent effect on how many people give 4 and 5 star reviews. However, there is no clear pattern as to which quarters tend to show more positive reviews as quarters are not consistent yearly.

One possible explanation of seasonality affecting review quality is introduced by the paper written by Hajas, Gutierrez and Krishnamoorthy. By examining the cyclical nature of Yelp ratings, the researchers speculated that in order “to produce quality food, one has to continue to invest money on production, service and advertisement. However once the service is established and good ratings are obtained, continued investment on the components that contribute to the quality may drop” [4]. This drop in quality would result in a lower rating, which in turn would motivate management to reinvest money on production, service and advertisement to improve quality again. This management-investment cycle also gives clarity to the question of why quarters are not consistent with certain levels of positivity.

Rating-Mention Correlations

After using SAS to filter reviews based on whether they mention the most popular businesses in each neighborhood, we use Excel to see if there is a relationship between a review’s star rating and where their names are mentioned in the text. It is observed that if reviewers mention a competitor’s name early in their reviews, they are more likely to give a high rating. Conversely, the later reviewers mention a competitor, the more likely they are to give a low rating. After further examination of the language of the reviews that exhibit these behaviors, it is apparent that reviewers often mention a competitor early in a review to either offer a favorable comparison to a more-established business so as to better explain their experience or to mention how the business reviewed is better than the competitors. Alternatively, a review that mentions another business towards the end of the review is a result of a rant that points out the negative aspects of the reviewed business. The mention towards the end is usually a suggestion to avoid the reviewed business and to instead, try the alternative, mentioned business.

To better understand rating-mention correlation, it is also helpful to determine the tipping point at which a business mention in a review can determine whether that review would generally be favorable or not. From preliminary observation of the patterns that exist in the rating-mention charts for the six most popular restaurants in Central City, Phoenix, we speculate that a tipping point of 1000 characters exists. In other words, a review that mentions a competitor after the 1000th text character will generally be a negative review and a review that mentions a competitor before the 1000th text character will generally be a positive review. This observation is further verified when taking rating averages into consideration. For businesses in Central City, the average rating of a review that mentions a competitor before the 1000th character is 3.75 while the average rating of a review that mentions a competitor after 1000 characters is 3.02.

Restaurant	Ratings Before 1000 Characters	Ratings After 1000 Characters
Pizzeria Bianco	4.275	3.76
Cibo	3.325	3.0476
The Arrogant Butcher	3.33	2.667
Lo-Lo's Chicken and Waffles	4.00	2.667
Matt's Big Breakfast	3.707	3.00
Angels Trumpet Ale House	3.875	3.00
Average	3.75	3.02

6. Conclusion

Our analysis reveals that many businesses exist in tight clusters. From our heat map, it is clear that most review activity tends to be centered around the Central City/Encanto, Phoenix neighborhoods. This clustered area also includes businesses with the highest number of reviews, indicating a high level of preferential attachment. Yelp users will tend to gravitate toward more popular establishments as businesses with the highest review counts are constantly revisited and reviewed again.

Observing the patterns associated with the quantity of reviews should be complemented with observing patterns associated with their specific ratings and distribution. In this paper, we also identified and verified for a warm-start bias that appears in Yelp data. As time of operation increases for a business, positivity decreases, indicating a decreasing share of 4 and 5 star reviews over time. The explanation of this phenomenon may be that a business in its initial stages gets exposure to a limited, niche audience that are more receptive to the business. When the business reaches general popularity, it receives audiences that are more likely to give less favorable reviews since the preferences that the business caters to is more varied than those of its initial audience.

It is also evident that positivity is affected by seasonality as the number of 4 and 5 stars reviews given to a business fluctuates over a given year. This can be explained by a natural cyclical business behavior in which management of a business invests to improve its service when ratings are low and stops investing when ratings are at an acceptable level.

In addition, by measuring where a review would typically mention a competing business using a character index and measuring frequency, it is generally observable that reviews that mention a competitor early (specifically before the 1000th character in the review) tend to have a positive rating whereas reviews that mention a competitor later in the review tend to have a negative rating. This particular tipping point would have to be studied further since the tipping point is only valid if a certain number of reviews is reached. Furthermore, the lack of data available that show business mentions prohibits in-depth review.

The analysis and methodology presented in this paper illustrate the network behavior of local businesses as well as some of the underlying causes for such phenomena.

References

[1] Segal, David. A Rave, A Pan, or Just a Fake? *The New York Times*. 21 May 2011. Available at http://www.nytimes.com/2011/05/22/your-money/22haggler.html?_r=0

[2] Potamis, Michalis. The Warm Start Bias of Yelp Ratings. *Groupon*. 26 Feb. 2012. Available at http://arxiv.org/pdf/1202.5713v1.pdf

[3] Chin-Lee, Jao-ke. How to Win Ratings and Influence Reviewers: Preferential Attachment in Rating Systems. Bachelor's thesis, *Harvard College*. 2015. Available at http://nrs.harvard.edu/urn-3:HUL.InstRepos:14398548

[4] Hajas, Peter, Louis Gutierrez, and Mukkai S. Krishnamoorthy. Analysis of Yelp Reviews. *Rensselaer Polytechnic Institute.* 8 Jul. 2014. Available at http://arxiv.org/pdf/1407.1443v1.pdf

[5] Google Maps. Pizzeria Bianco to Cibo, Walking Distance. Available at https://goo.gl/t6755f

[6] Yelp Academic Dataset. Available at https://www.yelp.com/dataset_challenge