Yelp Review and Rating Analysis

**Introduction/Background**

Nowadays, modern technology has affected people’s daily lives not only by providing information efficiently, but also influencing decision making. Yelp is an application people refer to when selecting businesses and restaurants to patronize. It serves as a platform to view restaurant reviews, ratings from other users out of five stars, and photos. Studies have shown that an extra half-star on a review nets restaurants 19% more reservations (Tepper). Yelp has not only changed how humans behave and interact with each other, but has also had a major impact on restaurants and their business as well. Thus, I wanted to investigate more about the reviews and ratings each restaurant received and how they related to each other in three major areas: polarities, bag-of-words/bigram and length of the review texts. Polarity classifies whether the expressed opinion in a sentence is positive, negative, or neutral by assigning a number ranging from -1 to 1 taking strength of the opinion into account. I wanted to see if the relationship between polarity and rating was consistent, and if not, what the trends actually were. Diving into the sentiments of the review texts allows us to see actual users’ thoughts. Conducting sentiment analysis involves bag-of-word and bigram techniques, which collect all the unique words or bigram (a pair of consecutive words) and represent each text with a sequence of 0s and 1s where 1 represents the existence of the feature. This allows us to extract the key factors that matter the customer experience the most.

**Methodology**

I scraped Yelp review data using the Yelp API (https://www.yelp.com/developers/documentation/v3). The files I used are “Business.json” which includes a “Business ID” and average “Stars” for each ID, and “Review.json” which includes the “Business ID”, “Stars”, “Text” from each user.

For the first analysis, I translated each review text into English and then calculated the polarity of each review from a user using the *TextBlob* Python package. Then, I ran a linear regression model using the *sklearn* package on the polarity and number of stars to calculate the correlation between the two.

For the second and third analysis, I wanted to find out what individual words/

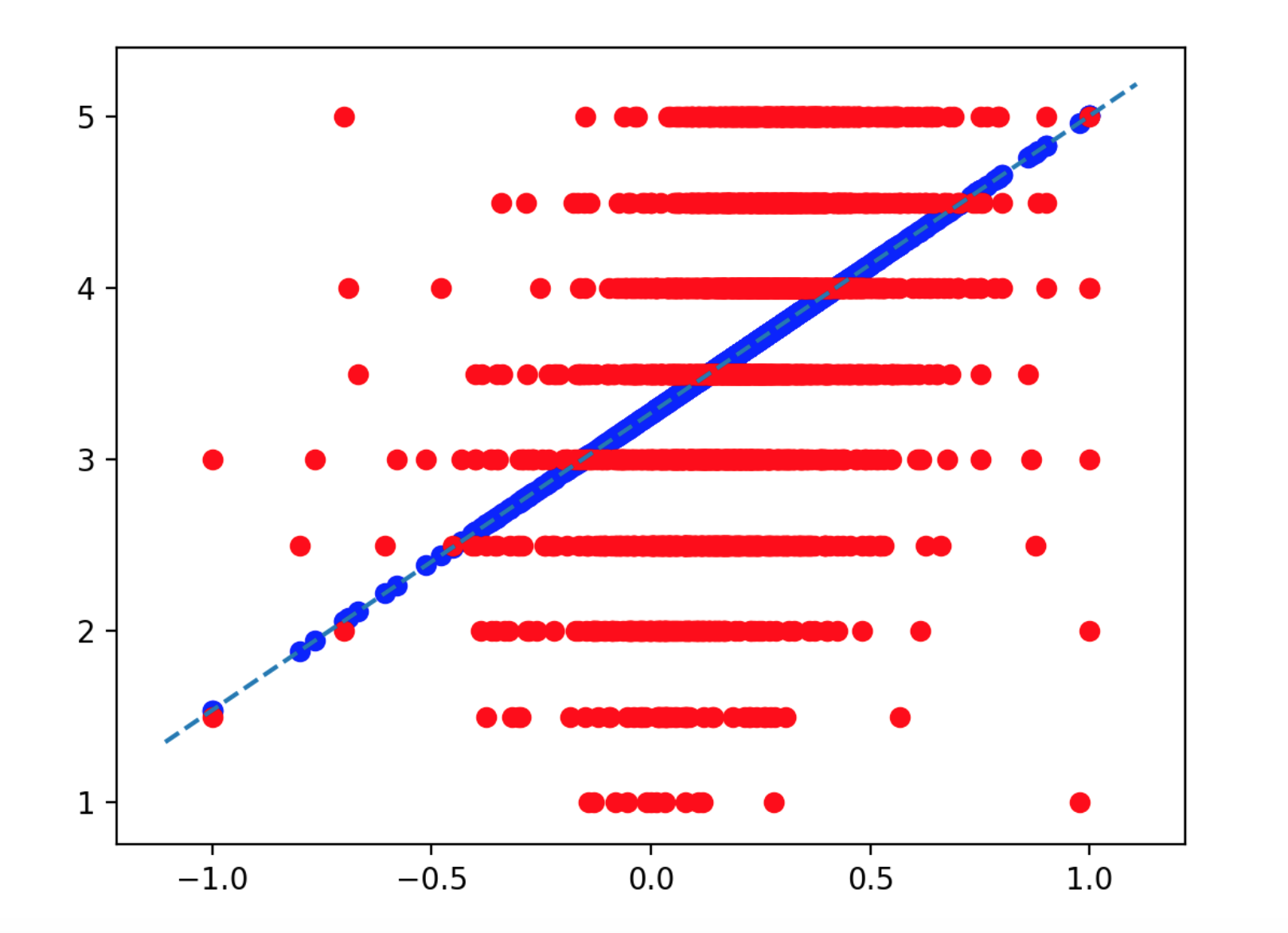
bigrams have more impact on the rating than other words/bigram. In order to find out the importance of the individual words, I first used the bag-of-words technique to collect the unique words in the universe of reviews with the *textmining* package and represented each review as a vector after filtering out the stop words (such as “and” , “it”, etc.). To find the importance of bigrams, I used BigramCollocationFinder from the NLTK package to first filter out stop words and then identified 50 bigrams with the highest PMI (Pointwise Mutual Information), or a measure of association between a feature and a class. Lastly, I used ExtratreesClassifier from the *sklearn* package to find the importance of a feature (a unique word or a bigram in each case) in terms of the impact it makes on the rating.

For the last analysis, I calculated the length of each review text and the ratings from each user review, and ran a linear regression model on the length and stars with the *sklearn* package.

**Results**

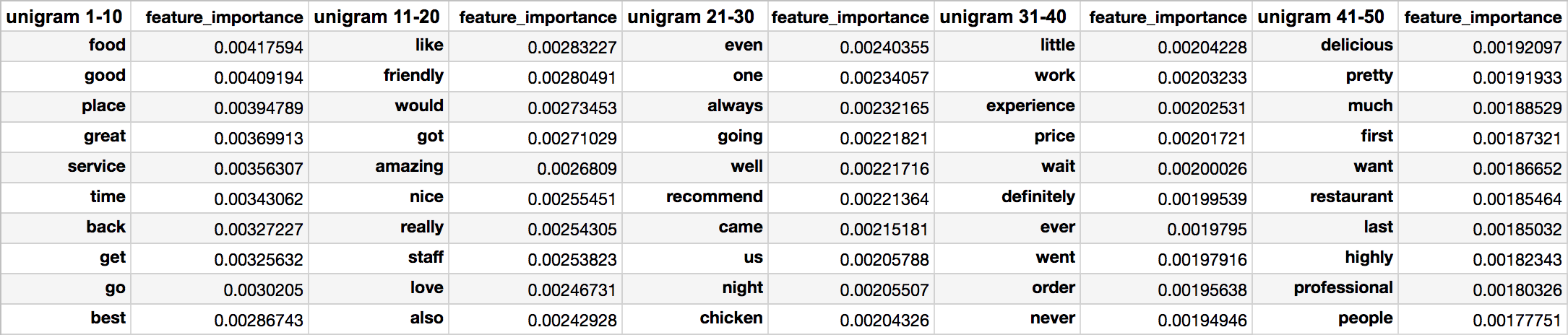
Results of the first analysis show that a higher polarity corresponds to a higher rating, and vice versa. The graph below shows polarity against stars with a linear regression line in blue.

Polarities

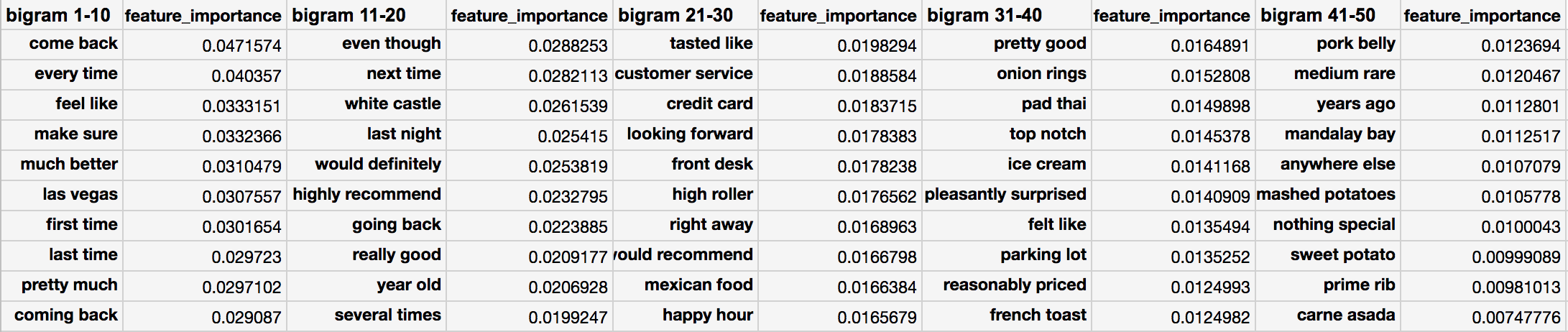


Stars

**Polarities vs Stars**

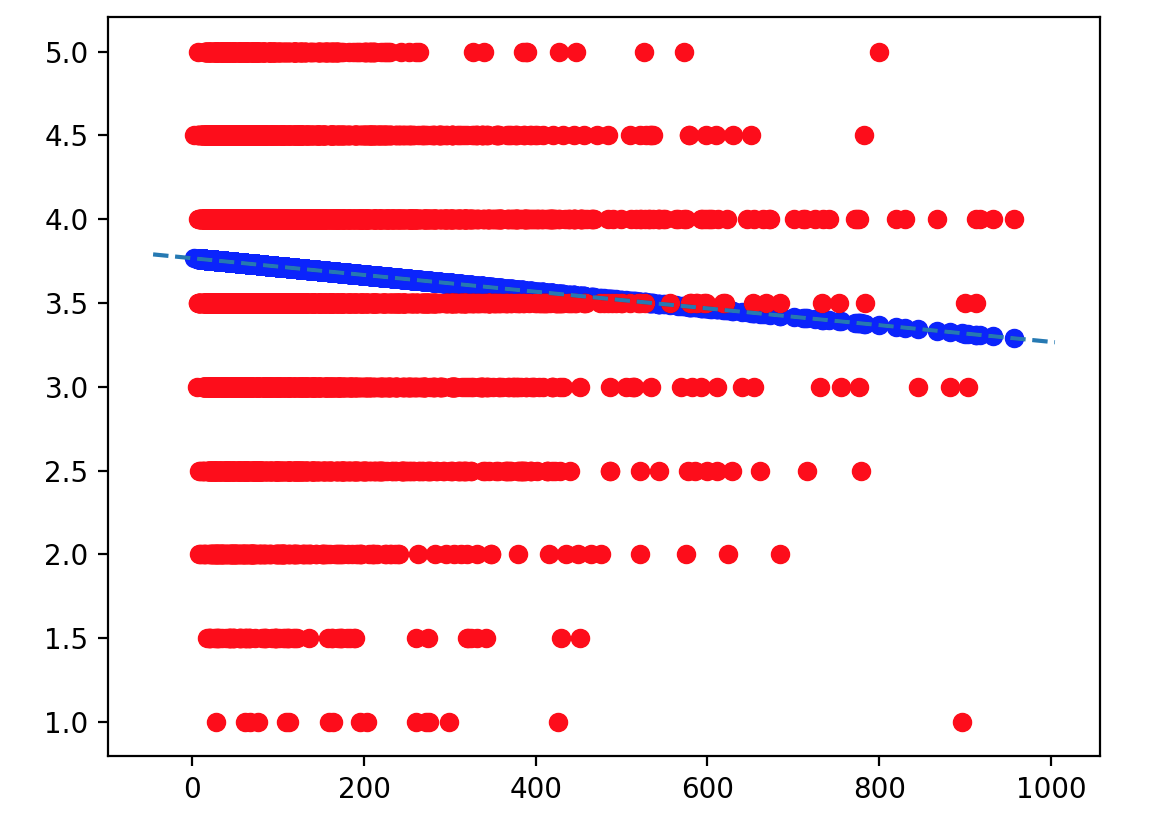
For the second and third analysis, we found the top 50 words/bigrams in terms of importance, shown below. 

Unigram - top 50 feature importance



Bigram - top 50 feature importance

Lastly, investigating the correlation between the length of the text review and the rating given by the user reveals a negative relationship between the two values.



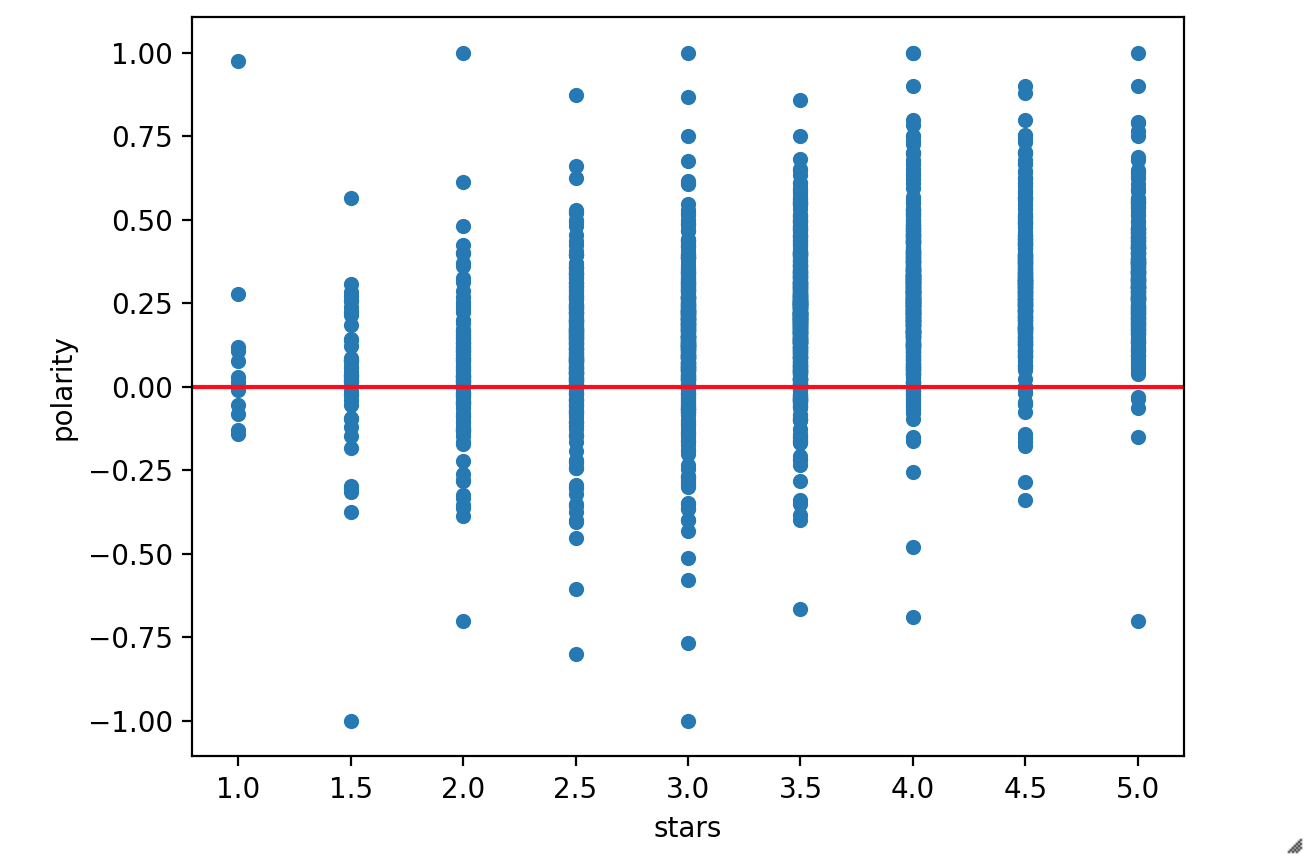
Stars

Length

**Length vs stars**

**Analysis**

In the first analysis of the relationship between stars and polarities, I noticed that although reviews of 4 stars and above tend to have mostly positive polarities, the negative polarities of reviews are not necessarily below 2 stars. Rather, they are spread out between 2 and 3.5 stars. This shows that people tend to give higher ratings even though they write negative comments. Typically, they rate in the range of 2 to 3.5 stars, and rarely 1 or 1.5.

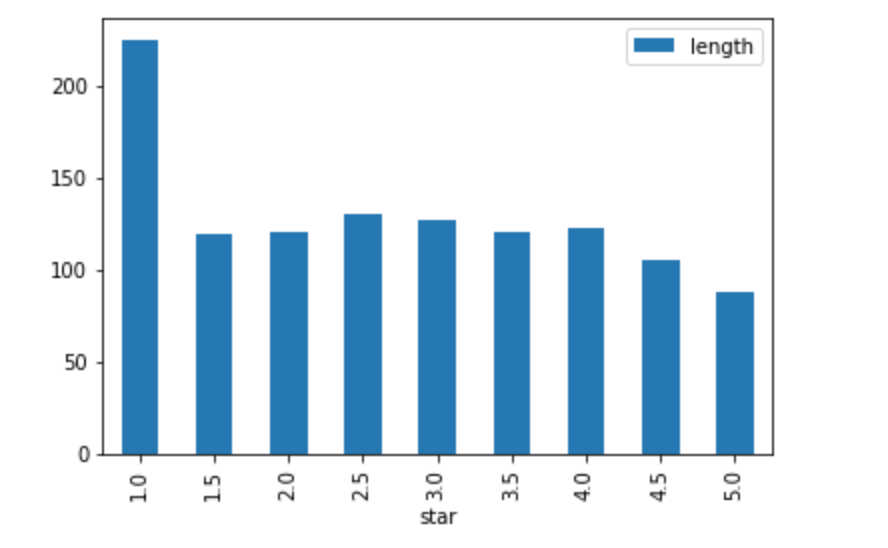


**Polarity against stars**

In the second analysis, words such as ‘food’, ‘service’ and ‘staff’ that appear on the list indicate that they are the key elements for a business in the service industry to succeed. Indeed, words like ‘friendly’, ‘professional’ in terms of service are also key words that led to higher ratings. Note that words such as ‘wait’ and ‘price’ are also on the list, showing that these are also important factors that customers are paying attention to.

In the third analysis, bigrams such as ‘customer service’, ‘reasonably priced’, ‘front desk’ and ‘parking lot’ on the list show what affects customers’ experiences the most. Specific dishes such as ‘pad thai’, ‘pork belly’, ‘prime rib’, ‘carne asada’ and ‘mashed potatoes’ also appeared on the list, implying that these dishes were very popular. Interestingly, key words such as ‘happy hour’, ‘las vegas’ and even ‘high roller’ are also on the lists, indicating that the current mood when people are in Vegas or during a promotion such as happy hour may have effects on the view of a restaurant, and are also crucial factors for scoring high ratings.

Lastly, the analysis on length and stars shows that text reviews for one-star ratings on average have the longest in length, while the length is the shortest for 4.5- and 5-star ratings. Below is the graph of review length against star ratings.



Length

Length v.s. stars

The graph implies that a good dining experience does not require many words to describe. People tend to put down short reviews, for example: “Definitely recommend!” or “Could not be happier than my decision.” The 1-star reviews tend to have longer paragraphs and mostly contain complaints. There was one 1-star review with 896 words in total (while the average review length is 118 words) describing how the Starbucks barista messed up the order, and the bad attitude the customer received when trying to confront the barista, then how they ended up requesting a refund after having the store manager involved; customers are more likely to use many words to rant about a bad experience than to rave about a good one.

**Conclusion**

This research revealed what factors in a review affected the ratings most. Since the rating has a significant impact on revenue, ensuring a decent one is non-negligible for restaurants to stay competitive. A restaurant can understand customers’ needs by doing similar analyses on reviews to find out what its particular strength and weaknesses are. For example, if the keyword ‘parking lot’ or ‘credit card’ appears on the bigram list, one can investigate if a lack of parking spaces or credit card machine causes inconvenience that further affects customer’s dining experience and leads to a low rating. In addition, if the signature dishes appear on the n-gram list, the owner can train the servers to recommend these dishes to customers while taking orders. Using the wealth of free information available, staying alert to social media, being flexible in business, and meeting customers’ needs are key elements for business growth and success strategies in the service industry.

**Citation**

Tepper, Rachel. “Yelp Study Shows Extra Half-Star Nets Restaurants 19% More Reservations.” *The Huffington Post*, TheHuffingtonPost.com, 6 Sept. 2012, www.huffingtonpost.com/2012/09/06/yelp-study-ratings-restaurant-reservations\_n\_1861720.html.