November 5, 2019

Ruohua Li

INFO 4360

University of Denver

word cloud visualization of Yelp reviews

**Data Sources**

In this project, 126 records are extracted from Yelp reviews of “Body Massage Wellness Spa", including comments and ratings. “Body Massage Wellness Spa” is a massage store located in Denver, who offers Acne Treatments, Back Treatments, Body Massages, and Body Scrubs etc. services. The main aim of this analysis is to explore the most frequent words used in reviews and help customers find out whether the store is good, what the great/bad service is and etc., and help the store owner improve their business. (see Appendix A)

**Summary of Results**

In order to illustrate more useful information through reviews, we make several plots using cleaning words, including:

* Bar Charts
* Word Cloud
* Bigram and Ngram

There are 2 bar charts that present the summary of Top 10 frequent words and the distribution of ratings. In the Top 10 Frequent Words chart, it shows words, “come” and “back”, take account for the first 2 biggest parts. In addition, the plot presents that customers use “great” and “best” multiple times and also mention “spa”, “time” and “experience”. the Distribution of Ratings chart demonstrates that “5 stars” occupies the biggest part, and then “4 starts” and “3 stars”. Thus, the massage store may provide great services for customers, and customers spend great time and have good experience there, so that they are willing to come back. Apart from top 10 words, we use the word cloud to depict more information.

In the word cloud, it shows that words, “therapist”, “relaxing”, “amazing”, and “nice”, appear several times. Also, this figure shows some words about services, like “facial”, “pain”, nail”, and etc. Thus, we can guess the massage store helps customers alleviate pressure and pain a lot, so they think it is a great place. To get deep understanding of the attitude in reviews, we can look at the bigram and the ngram.

In the table of the top 10 words frequency of bigram, it is very clear that people will come back, and they highly recommend this place. In addition, they mention “deep tissue” many times, and use “best ever” to show their positive feelings. In the table of trigram, “definitely come back” and “laser hair removal”, appears the most frequently, followed by “get last minute”, “last minute appointment”, and “deep tissue trigger”. We presume that the laser hair removal, spa, and deep tissue may be the store’s great services, and the following table prove it. The four-gram table shows people think the body wrap time is fair, they really enjoy the deep tissue and trigger points and they are able to get last minute appointment. However, the table demonstrates many negative comments as well, such as “unrelaxing day spa called”, and “want extremely unprofessional awkward”. Therefore, we can understand the reason that the store gets only 1 star. In addition, with the comparison of trigram and four-gram, whether the spa is good is vague, since the former shows the spa is highly recommended, but the latter illustrates the spa is unrelaxing. (See Appendix B)

Overall, the visualization of comments illustrates most customers have a great time there, just like the present of distribution of ratings. Additionally, the feedback of laser hair removal, deep tissue, and trigger points is great, but the feeling of spa is undetermined.

**Discussions**

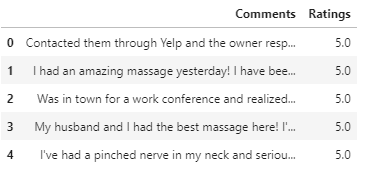
We can collect the general information of customers, like gender, to perform association rules. For example, we can explore whether women are more likely to enjoy deep tissue, and whether people who do spa are more willing to do body scrubs. This analysis may help the store provide some promotions and attract more customers.

**Appendices**

**Appendix A**

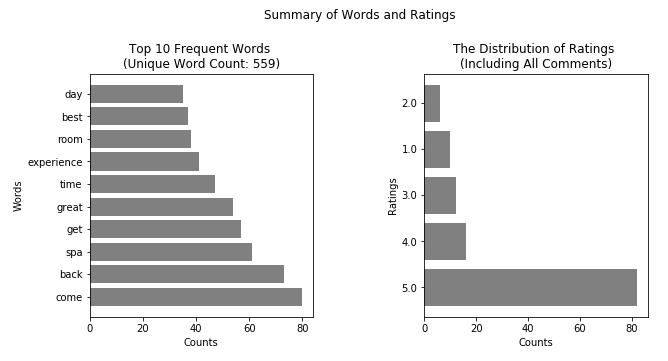
* The first 5 records extracted from Yelp

Website: https://www.yelp.com/biz/body-massage-wellness-spa-denver



**Appendix B**

* Bar charts for the summary of words and ratings



* Word Cloud



* Bar charts for the Top 10 frequent words in Bigram, Trigram, and Four-gram
  + Bigram



* + Trigram



* + Four-gram

