

Analysis for Active Life Business

Xiangyu Wang

xwang2439

Zijin Wang

zwang2548

Jingshan Huang

jhuang456

Yicen Liu

liu943

Yuan Cao

cao234

December 11, 2020

1 Introduction

In Yelp official categories¹, “Active Life” businesses are basically related with fitness such as gym or yoga, all kinds of ball sports (basketball courts, badminton, golf) and outdoor activities like hiking or rock climbing. In Yelp dataset, there are more than 2.6 million reviews about this kind of businesses. We hope to get some information from these reviews or ratings and give business owners some suggestions about what their customers care about and focus on. We will use trend analysis, word frequency analysis with ordinal logistic regression and business attribute analysis with ANOVA to give our suggestions.

2 Trend Analysis

For “Active Life” businesses, we first evaluate the trend of rating for this category.

As Figure 1 shows, there are both increment and decrement, which make it hard to determine whether there is a obvious change within this category. In order to determine whether there is a statistically significant trend, we performed Mann-Kendall trend test and result is “no trend” or says, not significant ($p_value > 0.1$), which means neither increment and decrement in Figure 1 is a true change in “ActiveLife” businesses.

3 Word Frequency Count

In this part, we focus on those words that shows up most frequently and see how it can bring us with useful suggestions. Figure 2 shows them by different star levels.

First, we look at those bigrams that show up in the Top100 word with all star level. The most popular word is “customer service”, which shows up 13.45% in review with 1 star and 5% in 5 star. Its information is intuitive that business owner should pay more attention on customer service, like service attitude, responses time...We also notice that “30 minutes” shows up frequently, which tell us, again, improve service is important. Another word we should

¹https://blog.yelp.com/2018/01/yelp_category_list#section1

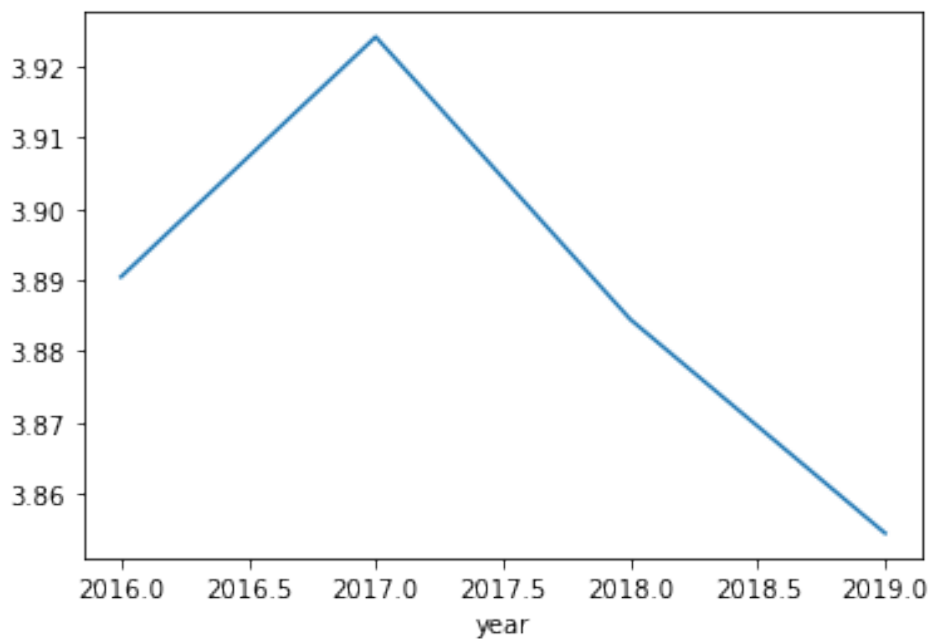


Figure 1: Average Star in Four Years

notice is “parking lot” that show up in all level of star with percentage 1% 2%. For more specifically, in bad review, some people mention that:

“I had the misfortune of parking in the YMCA’s large and mostly empty parking lot, which is not close to the YMCA building and looks a lot like a public parking area..”

”..outside the parking lot could use a little more lighting in certain area...”

“My only complaint is that outside the stadium, on the way to the parking lot near the parking garage, there was a really unpleasant sewer smell.”

And in those good review, people say:

”....this was a great! huge parking lot next to the stadium for \$12, which is average...”

“The trail and the parking lot were easy to find...”

“There is a safe parking lot for you....”

Basically, what important about parking lot is the position, capability, and its circumference. That is to say, build the parking lot more close to the activity place is better, what’s more, it should be large enough and easy to find. Business owner should also pay more attention on its cleanliness, lightness to import customer’s experience. Some word show high frequency in all star level but we found out they are meaningless like “front desk”, “las vegas”, “highly recommend”.

And then we pay attention to the word show up frequently in some specific star level. For example, “credit card” are the most popular word in 1 star review with show up rate 2.92%. People complain about some active place like gym stole their personal information by credit card, so business owner should pay more attention on the information security. We also notice that some business do not accept credit card which bring unconvinence to customers. Besides, word like “extremely rude”, “poor customer (service)”, “multiple times”, “3 months” show more

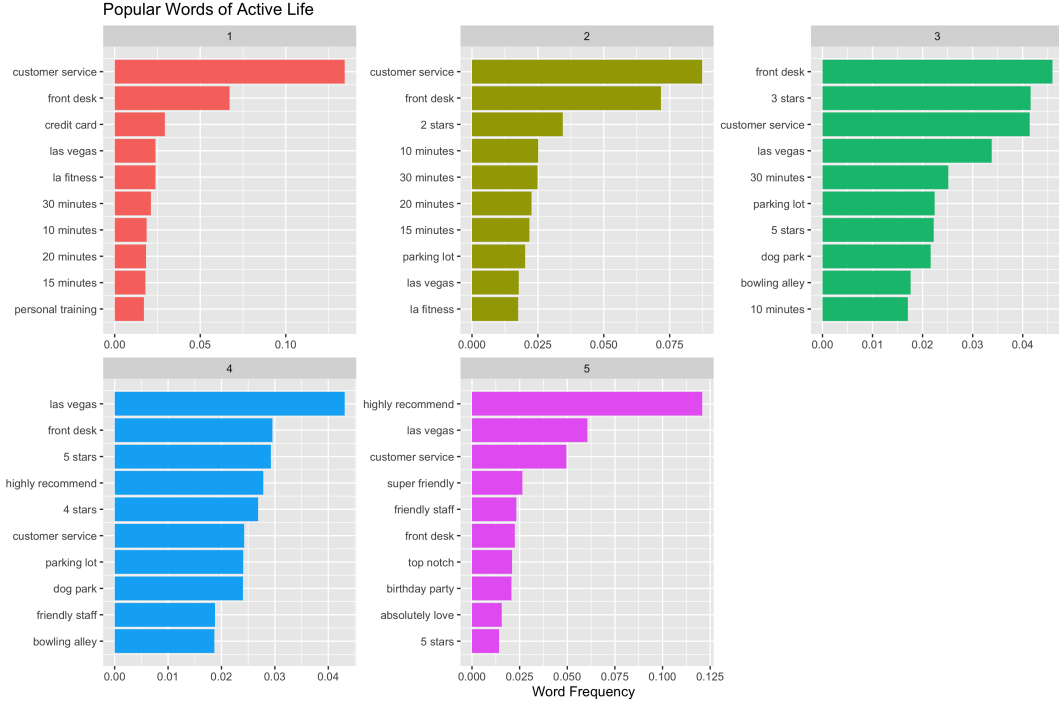


Figure 2: High Frequency Words

frequently in those bad review that tell us again improve customer service is very important.

In those word show up in good review, the most popular word is “super friendly”, “friendly staff” with show up rate 2.63% and 2.32%. For another aspect, we notice that people like “yoga studio”(1.11%) and “free weights”(1.40%) in active place. According to the raw text, we believe that providing free weights, machines and a well clean yoga studio can attract more customers and improve their experience.

4 Ordinal Logistic Regression

Besides using word frequency to do some exploring analysis, we also use ordinal logistic regression to assess the influence of these bigrams presence. We pick some high frequency words that shows up with different frequency in different star levels, such as the “customer service” mentioned above. To be specific, “customer service”, “credit card”, “minute”, “parking lot”, “staff”, “yoga studio” and “mini golf” are used to build our model.

In the result, some of them show quite importance to the change of review. The most influential bigrams are “credit card”, “customer service”, “minute” related such as “30 minutes” and “yoga studio”, with odds ratio to increase one level of rating as 0.139, 0.34, 0.442 and 1.683, respectively. For odds ratio less than 1 such as “credit card”, it means if a review does not contain “credit card”, then the odds ratio to increase one star level in review is $1/0.139 = 7.19$ times, which is a quite huge influence. For “yoga studio”, it means if a review contains this bigram, the the odds ratio to increase one level is 1.683 times.

Based on these model results, we also browse some related reviews, especially for low rating reviews. For example, the most important bigram “credit card”, many people mention that there are abnormal or some extra charge when they use credit card. So, we will suggest that business owners can strengthen the management for checkout staffs to decrease this kind of accidents. Similarly, suggestions based on other important words may include decreasing the waiting time for customers by simplifying service process or support some reservation service and considering adding yoga related projects (for gym business owners).

5 Business Attributes

In this analysis we first split the data into 23 subsets based on the business categories. Then we fitted an one-way anova model on each of the attribute in every subsets.

For the active life business, “bike”, “appointment”, “wheel”, “dog” and “parking” are significant and have positive effects on the rating, while “noise” has negative effects on the rating.

So, we will suggest the Active Life businesses reduce the noise around its location, offer bike parking place, offer wheel for disabled people, make it available and more suitable for people with pets(dog), improve parking ability and set up a appointment rule for better management.