Stat 628: Module 3 - Group 9 Executive Summary

1. Introduction

Yelp provides valuable information to consumers and businesses. It enables consumers to understand which businesses are offering perks they desire, and it allows businesses to understand what consumers enjoy about their businesses and what can be improved.

Yelp Elite are reviewers distinguished by Yelp as providing insightful reviews. To become a member of the Yelp elite reviewers must apply to join. Yelp describes these reviewers as "...a diverse community of passionate writers, photographers, and adventurers. Every experience you share on Yelp helps people discover the best of everything local."[1] We investigated the relationship and importance of Yelp elite reviewers on businesses, exploring the impact of having elite Yelp members reviewing restaurants, and correlating their review sentiment with star rating in the following month.

2. Data Description

Our data contain datatables with data on Yelp businesses, reviews, and users from 2005 to 2022. The user file contains information about if a user has ever been a yelp elite reviewer, and if so which years. The business data has information about each business. In particular it has attributes and categorical data about the type of business. The reviews datatable holds over 6.9 millions reviews with information about the details of the review.

2.1. Data Cleaning

The raw data contained a large amount of information across a number of businesses. To reduce the scope of our analysis we decided to only focus on restaurants. The restaurants we investigated were those that were categorized into one of 18 categories. The categories are Brazilian, British, Cajun, Chinese, Cuban, French, Indian, Italian, Japanese, Latin American, Mediterranean, Mexican, Middle Eastern, Modern European, Korean, Peruvian, Vietnamese, and American. We only included businesses with reviews in the provided reviews dataset because without any review data we were unable to complete our analysis. When completing the analysis of the average rating change we filtered out reviews that did not have at least one review 30 days before and a review 30 days after because we can't find an average rating of no reviews.

3. Exploratory Data Analysis

We began by doing an exploratory data analysis to understand if there could be a potential impact on reviews by yelp elite reviews. While completing this analysis we noticed that although there are substantially fewer yelp elite reviewers, their reviews appeared to reach a much wider audience. On average yelp elite reviewers had more fans, compliments, and reviews identified as useful by other yelp reviewers.

User Type	Total number of users	Average number of reviews per user	Average number of fans per user
Elites	90,000	224	23
Non-elites	1,900,000	13	0.4

In addition to having a wider reach, we also found that Yelp Elite reviewers were more likely to assign a 4 or 5 star rating to a restaurant and far less likely to assign a 1 star rating as shown in the distribution in Figure 1. The combination of widespread audience, useful reviews,

and tendency to rate higher could lead to other consumers being more likely to visit a restaurant after a Yelp elite member has written a review. Furthermore, for businesses with average rating below 4 stars and businesses with 0 reviews in the previous month, an elite review is likely to impose an uplifting effect.
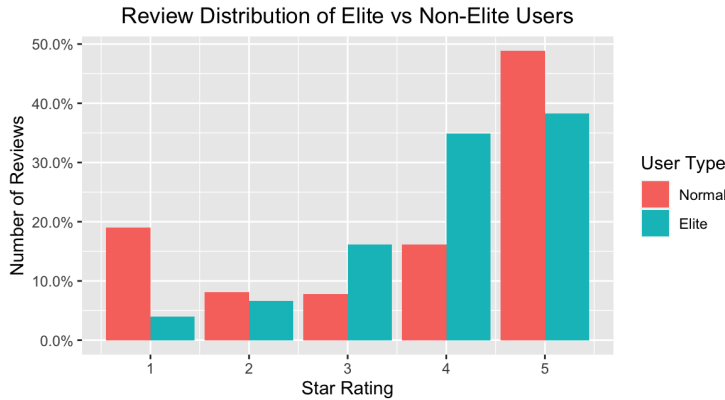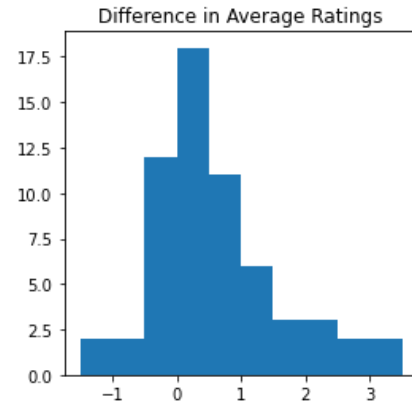

Figure 1: Rating Distribution

Figure 2: Average rating difference

4. Key Findings

To understand the potential differences on how businesses are impacted by elite yelp users writing reviews we conducted four different tests. We conducted a paired t-test, a difference-in-difference analysis, distribution comparison, and a sentiment analysis of the elite users' reviews.

4.1.1. Multiple Hypothesis Correction

In all of our statistical analyses we conducted hypothesis testing for multiple types of restaurants. Due to conducting numerous hypothesis tests a multiple hypothesis testing correction was required. We utilized the Holm-Bonferroni[2] method to control the family-error wise rate. The Holm method works by comparing increasingly less strict levels against the ordered p-values. We completed testing for 18 different types of restaurants and therefore had an adjusted alpha starting at 0.00278 and going to 0.05.

4.1.2. Paired t-test of Ratings and Review Counts

Our first analysis focused on treating the data as paired samples. In our hypothesis the event of an elite reviewer writing a review was the 'treatment' to the business. The number of reviews and the average rating of the reviews were found in the 30 days preceding and following the elite review. This was completed for each of 18 different types of restaurants. From this analysis only American restaurants had a significant increase in the number of reviews after a Yelp Elite member wrote a review with a p-value of 0.0018. We found no restaurants with a significant increase in their average rating.

While the paired t-test is able to detect a significant change in the data before and after a treatment, due to the Yelp data not coming from a randomized control trial we are unable to draw causality. To attempt to understand if there is a causal relationship we next completed a difference-in-difference analysis.

4.1.3. Difference-in-Differences

The difference-in-difference[3] analysis utilizes longitudinal data in an attempt to identify the differential effect of a treatment on the treatment group compared to the control group in an observational dataset. This is able to provide a deeper understanding than a paired t-test and give insight into how important elite users' reviews are to businesses compared to normal reviews.

To complete the difference-in-difference analysis we needed to calculate the mean difference for the number of reviews and rating before and after a Yelp elite member wrote a review, and the mean difference before and after a normal user wrote a review. We found this difference-in-differences for all reviews and calculated the differences.
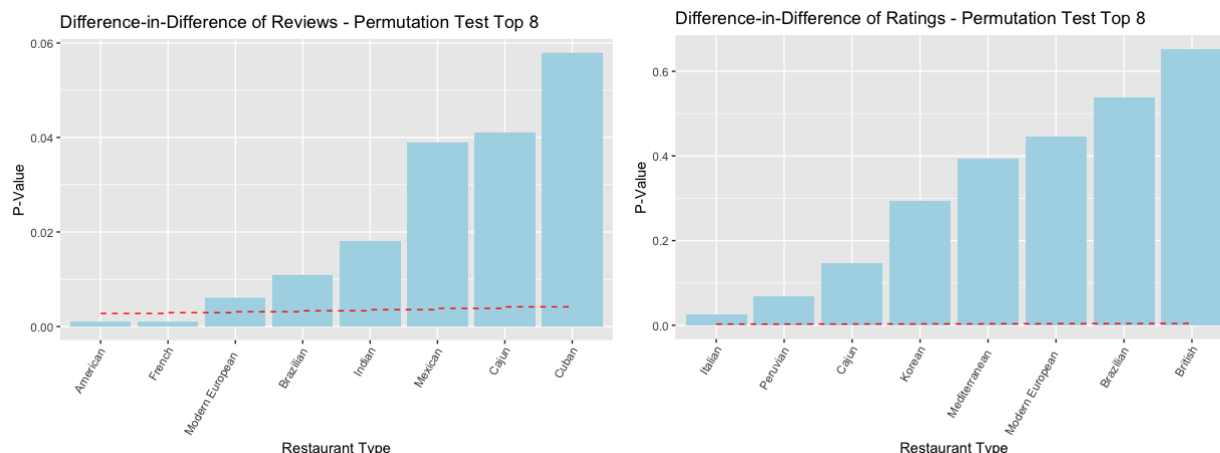

Figure 3: Difference-in-Difference Analysis

Our hypothesis for the difference-in-difference test is that the change in number of reviews for a business in the 30 days after Yelp elite user writes a review compared to the 30 days before is larger than the change in number of reviews in the 30 days after normal yelp user writes a review compared to the 30 days before. A second hypothesis with the same structure was completed for average rating as well.

We used permutation testing to create our null distribution. Using this null distribution we could then identify how extreme our observed data was for each of the types of restaurants. Our results indicate that for American and French restaurants elite Yelp members had a significant increase in the number of reviews while all other restaurants did not show significant results. Figure 3 is a bar chart showing the p-values obtained from our testing for number of reviews. When looking at the average rating we found that none of the results were statistically significant.

Although only a small subset of the types of restaurants were significant for the number of reviews, our original hypothesis did not involve any analysis of the rating assigned by the reviewer or the sentiment of their review. A Yelp elite member could have a stronger influence on other consumers, but if they are saying negative comments it could dissuade other individuals from visiting the restaurant.

4.1.4. Sentiment Correlation

Our next area of investigation was understanding if the sentiment of an elite Yelp reviewer correlates with the average rating in the month following their review. A strong correlation here could be indicative of Yelp elite members having a strong influence on how others perceive the restaurant. We use AFINN lexicon to calculate the sentiment scores. The sentiment scores are in [-1,1] and the larger the value, the positive the comments. We conducted the linear regression of average stars in the following one month of elite users' review on the sentiment scores. The coefficient of sentiment scores is 1.54 at 0.001 significant level. On average, sentiment scores increase by one unit and stars will increase by 1.54.

4.1.5. For Low-Rated and Unreviewed Businesses

We defined low rated businesses to have an average rating below 4 stars in the previous month. On average, the rating of those businesses increases by 0.52, and the number of reviews

of unreviewed businesses increases by 0.72, after a Yelp elite writes a review. The paired t-tests indicate that both increases are statistically significant.

In order to ascertain that the changes are results of elite reviews, for each elite review and the corresponding 2-month timeframe, we randomly picked another 1000 businesses, calculated the changes, compared the change of the business with the elite review versus the other 1000 reviews, and calculated the percentile. We repeated for all elite reviews and obtained a distribution of percentiles. If the null hypothesis holds true that elite reviews and non-elite reviews have the same effect on businesses during the same interval of time, then the distribution of percentiles follows normal distribution with mean at 50. For businesses with average rating below 4 stars, the ks test shows significant deviation from normal distribution, and 1-sample t-test shows that the mean is significantly larger than 50.
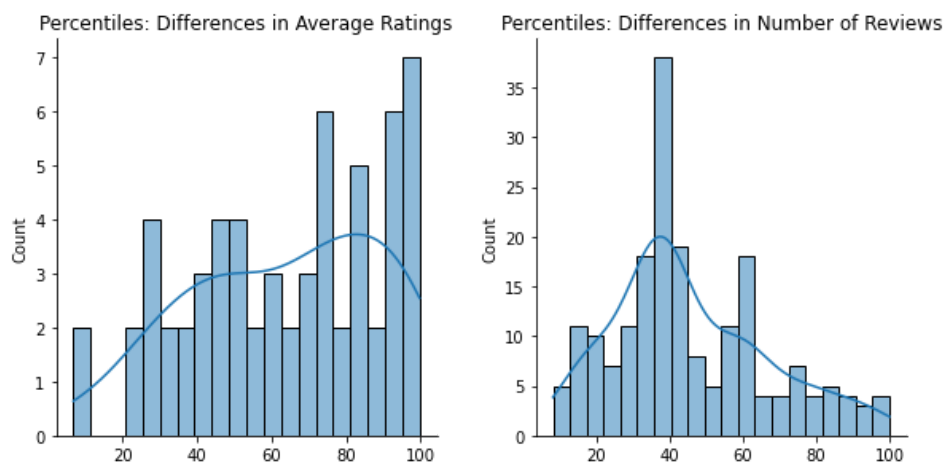

Figure 4: Distribution Comparison

5. Recommendations for Businesses

Given our results, our recommendation for restaurants is to not be concerned about having Yelp elite members come visit your restaurant unless you are an American or French restaurant, low-rated, or an unreviewed business. Their impact is no larger than the impact of non-elite Yelp reviewers. However, American and French restaurants could potentially see an increase in the number of reviews and by proxy the number of customers. Therefore they could look to incentive having elite Yelp members visit their restaurant.

For low-rated and unreviewed businesses, an elite review is significantly correlated with increases in either ratings or number of reviews. For the former type of businesses, this increase in ratings is more likely a result of the elite review, since changes due to non-elite reviews of the same time period are much weaker. However, for the latter type of businesses, since elite reviews have similar impacts as non-elite reviews, we cannot establish similar conclusions.

6. Conclusion

While at first glance it may appear that Yelp elite reviewers could have a significant impact on the future Yelp reviews and ratings of a business, we found in most situations their impact to be negligible. One major assumption about our analysis hinges on using the number of reviews as a proxy for the number of individuals that visit a business. This may not be a perfect metric, and it could even be that due to Yelp elite members often writing reviews that other users find useful and substantial, others are dissuaded from writing a review because they don't have any additional information to add. Extensions of this project could be done by finding a more precise metric of a restaurant's success, but without access to this data it would be challenging.

Works Cited

[1] https://www.yelp.com/elite

[2] Holm, S. (1979). "A simple sequentially rejective multiple test procedure". Scandinavian
      Journal of Statistics. 6 (2): 65–70

[3] Snow, John, 1855. On the Mode of Communication of Cholera. Edited by John Churchill.
      Second ed. London.

Member's Contribution to the Summary, the Presentation, and the Code

OH - Wrote code for some preprocessing, wrote code and report for paired t-test on restaurant types and difference-in-difference analysis. Made slides 2-6 of the presentation. Wrote some of introduction and data description.

CB - Wrote code for low-rated and unreviewed businesses in Jupyter Notebook. Wrote low-rated and unreviewed businesses section and part of recommendations section in summary. Did slides 7-8.

XW - Wrote code for sentiment analysis. Wrote sentiment correlation section.

ZT - Wrote code for shiny app. Wrote introduction and data description.