Exploration of Yelp Reviews: Improving Reviews through Review Sentiment and Context

Avani Reddy, Laura Capalleja, Misha Desai, Matt Sheahan

Introduction

When trying to find the next place to get a bite to eat, you may look at Yelp to find the restaurants with the best ratings. But are all five-star reviews created equal? Factors such as where the restaurant is located, who is writing the review, what is written in the review, and when the review was written can all paint a more detailed picture than the number of stars assigned to a review. In this project, we are seeking to propose an optimal aggregation of reviews that normalizes the sentiment of reviews based on the reviewer and geographic location and prioritizes more recent reviews to get an accurate assessment of the restaurant's true rating. With a normalized review, customers can more accurately assess which restaurants are worth the visit.

Problem Definition

The problems with the current Yelp system can be summarized by human nature and a lack of transparency. We as humans tend to only review items our experiences that make a significant impression. This leads to an oversaturation of five-star reviews on Yelp, and if everything is rated five-stars how can you truly pick out the best places to eat? Human nature can also influence a potential discrepancy between the text of a review and the number of stars awarded in a review. Legitimate criticisms may be brought up in the text of a review but then a rating of four or five stars because the review does not want to seem overly negative. On the issue of transparency, Yelp keeps their exact algorithm secret, but reviews from multiple years ago can still be seen and still influence the overall review of a restaurant.

By extracting sentiment from the review text, determining the quality of the review, and weighing the review by recency, a new review score can be calculated that will allow consumers to better distinguish between good and great restaurants.

Who cares? A solution to this problem has a variety of potential benefactors. The average Yelp user will be better able to determine top of the line restaurants as reviews will no longer be skewed towards higher ratings and considering recency means that a restaurant that may have changed ownership in the last year will be better have a more accurate representation of the current experience. A business owner that uses Yelp will see benefit through having more transparency of the criteria that lead towards a higher review. The same can be said about the avid Yelp user who will better be able to improve their review quality in case they want to become more influential.

Impact and Measurement: To determine success, we will evaluate the difference between the normalized scores and actual non-normalized scores to see how much this varies. We are looking to see a more normal distribution for the new scores that is no longer skewed toward five stars.

Project Plan (equal contribution):

- 2 weeks: data preparation and cleaning (using Spark, Pandas, TextBlob, Python plotting libraries) [25% contribution from each team member]
- 2-3 weeks: algorithm testing (using PySpark) [70% Misha, 30% remaining members]
- 1-2 weeks: visualization (using Tableau) [80% Laura, 20% remaining members]
- 2 weeks: validation and analysis of algorithms [25% contribution from each team member]
- Poster and paper [50% Matt and 50% Avani]

Literature Survey

There has been extensive research showing that consumer reviews have had a strong impact on the popularity and demand for a given restaurant. Some studies have found that a 1-star increase in Yelp ratings led to a 5-9% increase in restaurant revenue [1]. Moreover, the growth of platforms like Yelp have also found that chain restaurants have declined in market share as Yelp usage has increased [1]. This begs the question: "What factors influence a user's rating of a given restaurant and how can consumers interpret all this data?"

The *AAA Diamond Rating Guidelines* for restaurants are based on three factors: food quality, service quality, and ambience. However, online restaurant reviews have been found to include additional factors that can skew a restaurant's rating, including price, reviewer bias, and review frequency [2,3]. Some research has even found that regional differences, such as GDP per capita and population density impact customer expectations of restaurants and reviews [4].

Today's rating methodology assumes that a restaurant's quality is constant over time. Since a new chef or new restaurant staff can have a dramatic impact on the restaurant quality, methods that have prioritized more recent Yelp reviews have yielded more accurate ratings [5].

Another avenue to improve the accuracy of a restaurant's rating is to find which groups of users may yield more accurate reviews. In one analysis, this was achieved by connecting fake reviewer groups using methods such as time of review and commonly reviewed products [6]. A similar approach may be possible to identify groups that may be more price sensitive or biased against a particular type of cuisine. In contrast, another analysis found that users with a 'Yelp Elite' classification tended to provide ratings with higher precision since they were more likely to incorporate feedback from previous reviews of the same restaurant [5]. We can also evaluate the user feedback on the reviews (i.e. "Useful" tags) to determine significance [7].

Other research has demonstrated that review sentiment has a strong impact on the restaurant rating [3,8]. Depending on the location and reviewer the weight of each portion of the review may be different and may cause a 3 star review in some places and a 4 star review in others. Their results suggested that recommended reviews are more likely to be generally positive and typically included complex sentences expressing substantial detail and varied sentiment between sentences [3].

Other methods used to determine sentiment in reviews required breaking down the sentences based on structure, punctuation, inflection, and other identifiers by feeding data into a pretrained sentiment analysis model, such as SentiStrength or Stanford Sentiment Analysis [9]. This paper used a black box model approach, and we could conversely tailor our approach rather than use the pretrained sentiment analysis models.

We can further explore regional differences in writing style by doing a sentiment analysis on other user input sources (I.e. other review sites, Twitter, etc.). The regional evaluation in these additional sources can then serve as an input to the sentiment analysis on the reviews [10].

If we want to investigate how different algorithms affect the normalization after sentiment analysis, we can try several approaches (e.g. SVM, Bayes, Genetic Algorithm, or a hybrid model approach) and use cross-validation to ensure we don't over-index on our modeled solution [11,12]. We can evaluate our approach with visualization techniques such as heatmaps, spider graphs, word clouds, etc. [13,14].

Intuition

How is it done today; what are the limits of current practice? When a user searches for a restaurant in Yelp, they can access a list of crowd-sourced reviews and view an average rating (based on a scale of 5 stars) that is based on all the reviews provided from previous users. While users can filter on most reviewed or best rated locations, there is no method for users to account for the sentiment of the reviews, reviewer patterns, geographic patterns. In addition, since Yelp uses an arithmetic mean of all user ratings, the current method does not account for the fact that a restaurant's quality can change over time.

What's new in your approach? Why will it be successful? Our approach will look at reviewer patterns and geographic patterns in order to normalize the restaurant review score, rather than calculating an average rating which is what is done today. This will be successful because we will provide a more localized score adjusted for regional patterns which will expose the true rating of the restaurant. Additionally, we will look at other variables that could result in us rating a reviewer's score higher, such as Yelp user data and when the review was written.

Approach: Yelp has a published collection of datasets including Reviews, User data, and Business attribution for a subset of their market. Using Pyspark, we have prepared the data for analysis by filtering to a relevant subset, creating meaningful groupings of the businesses and Yelpers, and extracting relevant information from the large tables.

Description of Approaches (Algorithms, User Interfaces)

Data Cleaning & Feature Engineering

Users: Yelp users have several pieces of data associated with them that we plan on using to help determine how trustworthy or useful the users' reviews are. This includes how long they've been an Elite user, the number of compliments they've received, and the types of tags their reviews have received (I.e. funny, useful, and/or cool). We normalized the Elite score to determine how many years the user has been Elite over the time period they have been a member of Yelp, to see if this helps us determine if long term Elite users have more trustworthy reviews. The normalized score is right skewed, telling us that users who have been Elite often were non-Elite users for several years too (Figure 1).

Businesses: Within Yelp, a business can choose to tag itself with multiple category key words. In the data, this is represented in a "Category" field which has the list of key words for each business. To find the most relevant categories, we extracted each individual key word and counted its usage. A total of 1,336 unique categories were found, including a variety of business types. We chose to filter our data to only the top 10 restaurant categories and assigned only one category per business.

Reviews: The Yelp Review dataset contains full review text data and includes the business that the review is written for, the user that wrote the review, and the rating provided by the user. By looking at the largest restaurant categories, we filter the Review dataset from 8.02 down to 3.7 million reviews. Next, we filter out states with a small number of reviews as there are not enough observations to yield statistically significant results. Based on this, Figure 2 confirms that our resulting data contains a distribution of reviews across each user rating. However, we can also see that the data is skewed towards more positive reviews. When we observe the distribution by state, we can see that, while the distribution may vary by region, ratings are still skewed towards 4/5-star user ratings. Due to user bias against providing strong negative ratings, we believe that the sentiment of the review may be a better indicator of the true review rating of a restaurant [15]. As a result, we then generate a polarity score for each review text using the TextBlob package. This package performs sentiment analysis by taking an average across the entire provided text and returns a polarity score between –1 and 1. Here, a score of –1 indicates a negative sentiment while a score of +1 indicates a positive sentiment. Based on the scaled results in Figure 3, we can see that this distribution has a more normalizing effect as the distribution is less skewed towards extreme positive values. Finally, we also create a new attribute to reflect the length of the review. This attribute is calculated by tokenizing the review text and removing stop-words.

Algorithm Scoring

Our methodology normalizes the data we have about users, reviews, and businesses to generate a final business rating. To generate a business rating, our algorithm calculates the **review quality** by combining 3 metrics:

- **Review engagement rank**: Each Yelp review can have a set of ‘useful’, ‘cool’, and ‘funny’ votes. In our model, we combine these votes to generate an engagement score. These votes indicate how other users perceived a review and help us measure the overall quality of the review. Once the engagement score is calculated, the score is normalized across all engagement scores for business reviews in the related state. Reviews with a higher engagement score are ranked as being ‘higher’ in quality than those with a lower engagement score.
- **Review length score**: Review length is another metric used to determine review quality. The review length is calculated by tokenizing the review text and removing stop word variables. The review length score is then calculated by normalizing the length of the review against other reviews from the same state.
- **User trustworthiness score**: To determine user “trustworthiness”, we normalized the number of ‘useful’ votes users received on their reviews. We did this by averaging the ‘useful’ votes they received by the number of reviews they had written. This prevented us from biasing a user that may have had one very successful review with many ‘useful’ votes.

The **review quality** results are then combined using the following calculation:

- **Recency of the review:** The algorithm gives higher weight to more recent reviews. The recency is determined by evaluating 6-month periods between the most recent date and date that the review was posted.
- **Polarity:** As described above, the polarity of the review is generated using tokenized and filtered text. The polarity is scaled to range between –10 (negative sentiment) and +10 (positive sentiment).

$$\text{Business Rating} = \frac{\sum_{i=1}^{N} \frac{polarity_i}{time_i} \times \left(f(length_i) + f(engagement_i) + f(user_i)\right)}{\sum_{i=1}^{N} 1 / time_i} \quad \text{where } f(x) \text{ calculates to percentile score for each metric } i$$

Visualization

Our visualization was built using Tableau and published to the Tableau Public Gallery. The dashboard includes two views aimed at two types of users; people looking to find the best rated restaurants and people who wish to understand how the algorithm works. The first view allows users to explore the highest rated restaurants within each category or state. As the user explores the restaurants, the most polarizing reviews (both positive and negative) will be displayed along with the most frequent words found in the reviews. The user can sort the list based on the highest adjusted scores or based on the biggest increase in score (from the original Yelp rating).

The second view is to allow the user to explore the methodology and get insight into the individual components of our algorithm. This view is advantageous over Yelp as Yelp currently lacks transparency in its scoring methodology. Understanding the components of the score would be particularly helpful for a business owner who wants to understand how to improve their overall rating.

Experiments, Evaluation, Details of Experiments/Observations

Experimentation

We will perform the following experiment to generate a more accurate business rating.

Experiment	Description
Find changes in sentiment across region	**Given** breakdown of regions in North America, state review data, and polarity **use** ANOVA **to** understand if sentiment differs significantly across different regions in North America
Find user's influential predictors of sentiment	**Given** Yelp users, their Elite normalized score, their average rating for businesses, the count of types useful/cool/funny tags **use** liner regression **to** identify which attributes affect sentiment, and therefore affect quality of the review
Scale reviews based on recency of review	**Given** date of the review, today's date **use** a self-defined algorithm **to** weigh recent reviews more heavily than old reviews
Generate a normalized review rating score	**Given** a review's engagement score, a review's length score, and a review's user trustworthiness score we will **explore** various algorithms **to** identify the optimal method **to** determine an accurate review rating score.
Generate an aggregated business rating	**Given** the polarity of the review, the recency of the review, and the quality of the review **explore** various methods **to** generate a final business rating.

Evaluation (Hypotheses)

In the next phase of our project, we will use our cleaned data and new features to validate or invalidate the following hypotheses.

ID	Hypothesis
1	There are significant regional differences in review sentiment.
2	A Yelp Elite user's ratings are more indicative of a business's quality than a standard, non-elite user.
3	The date (recency) of a review and the review length are influential factors in determining a restaurant's final business rating.

Details of the Experiments, Observations

We tested several hypotheses by performing experiments after cleaning our data, all listed above. After calculating polarity (the review sentiment), for the first hypothesis we investigated if there were regional differences in review sentiment that would impact the quality of the review. Conducting an ANOVA analysis comparing the polarity between states we received an incredibly small p-value, so we could reject the null hypothesis and concluded that at least two states had different average polarities. In order to better interpret the results, we mapped the states to different regions and conducted the same analysis with regions instead of state and concluded that the Western region of the US had a statistically higher polarity than all the other regions. Canada had the least polarizing reviews (neither negative nor positive) (Figure 4).

For the second hypothesis, we tested to see if Yelp Elite users had more impactful reviews than standard, non-Elite users by examining the relationship between normalized Elite status and polarity, along with several other review attributes. We discovered that the coefficient for the normalized Elite status had the second strongest effect on polarity, with a coefficient of –0.0743 (with the strongest effect on polarity coming from the rating the user provided) (Figure 7). Holding all else constant, this can be interpreted as the longer a user is Yelp Elite compared to how long they've been members of Yelp, the more likely they are to leave a negative polarity review. The model created was the strongest model we could create without introducing multicollinearity to the model. Additionally, the p-values showed that each of the predictors were statistically significant in predicting the response variable. The results of this model were also validated over a test subset of the data to ensure we didn't overfit the results to the original model.

For the third and final hypothesis, we aggregated the review data by individual business and fit a linear regression model where date of review was the independent variable and the original star rating was the dependent variable to each unique business id. Taking the coefficient of the independent variable as the trend of reviews, we were able to show that roughly 10% of all business had a trend greater than 0.001 (Figure 8) in either direction which shows that, holding all other things constant, over the course of a year a business's rating could change by around 0.365 stars. While the R-squared values on these regressions, were rather low due to the nature of the independent variable being in levels of either 1,2,3,4, or 5; the p-value for the dependent variable was consistently small enough to be considered significant even when the coefficient was less that 0.001. This shows that even if the trend is relatively small, taking recency into effect is a meaningful metric when compiling review scores

After proving the above hypotheses, we concluded that we could further investigate review quality and calculate new business scores by considering review quality, polarity, and review recency. We investigated how user trustworthiness, engagement with the review, and length of the review should be combined to calculate quality. Using this breakdown to measure score, we saw that the overall quality of reviews was skewed right, meaning that most reviews were low quality based on our analysis (Figure 9). Many of the reviews were short, and from users who were not providing useful reviews on average, and had overall low engagement with their reviews so this intuitively made sense. We then investigated how this review quality score, combined with the recency of the review and the polarity of the review, could re-calculate a new business rating. The results of this analysis yielded a more normal distribution of scores, most scores falling into the "3" star rating bucket (Figure 10), and less scores falling into "5" star rating. These results align with what we expected from our initial problem statement. Many businesses that have "5" star ratings may be made up of low-quality reviews and therefore should be closer to average (3-star rating).

Conclusions and Discussion

In conclusion, our experiments and analysis have resulted in a new business score that incorporates the quality of the review, recency of the review, and polarity (sentiment) of the review. Our visualizations created provide a breakdown for users looking to explore the new business ratings, along with a visualization to better understand the different components of our algorithm. While we believe we generated a new business score that reflects bias related to sentiment, recency of the review, and review quality, it would be wise to continue to validate this analysis by looking at incoming reviews for these businesses and seeing how they would be classified based on our approach.

Appendix

Figure 1: Histogram of Normalized Elite Scores (not including non-Elite users)

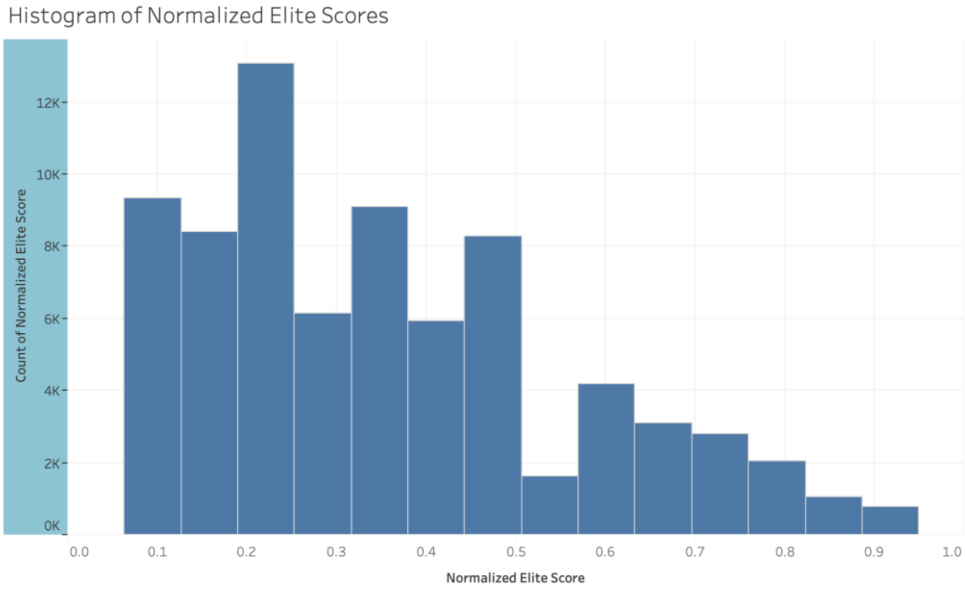


Figure 2: Total Reviews by User Rating (across all states)

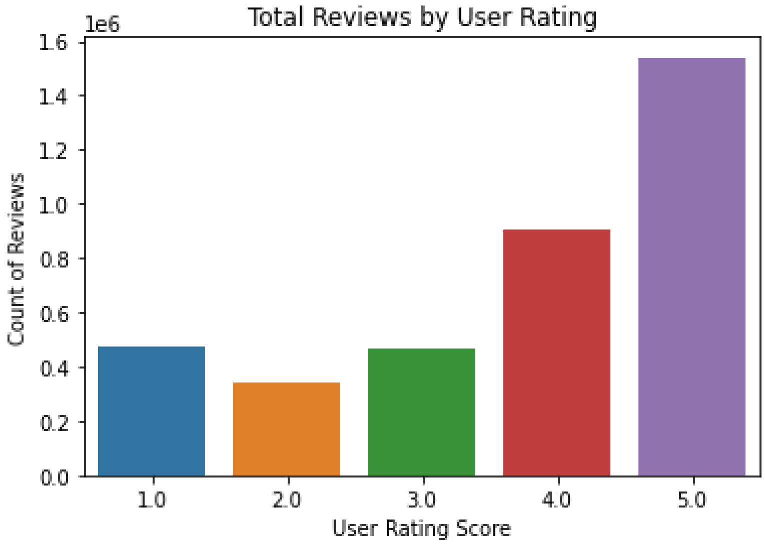


Figure 3: Histogram of Polarity Score

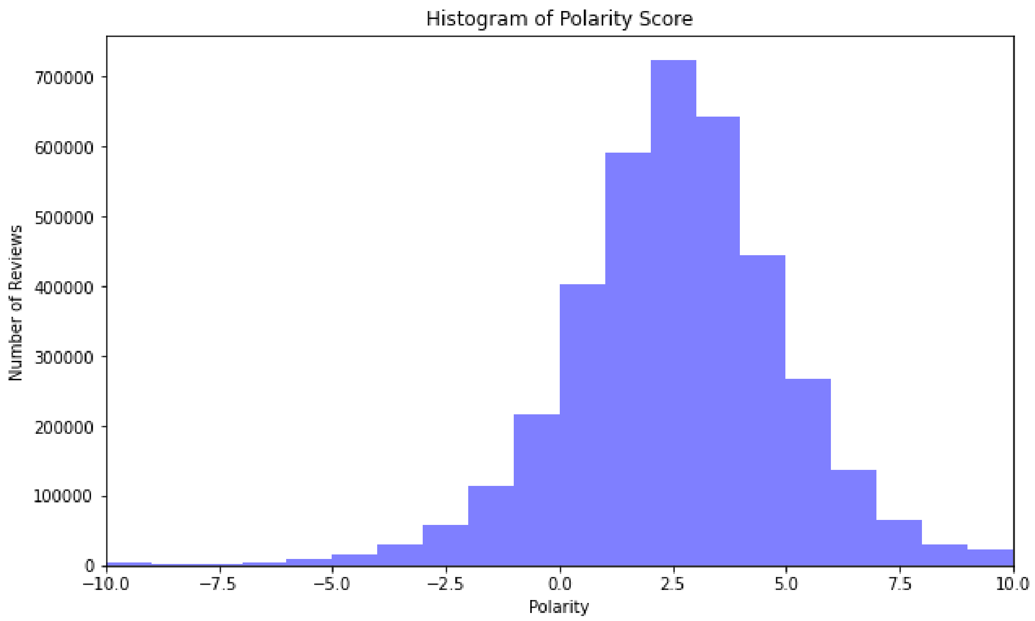
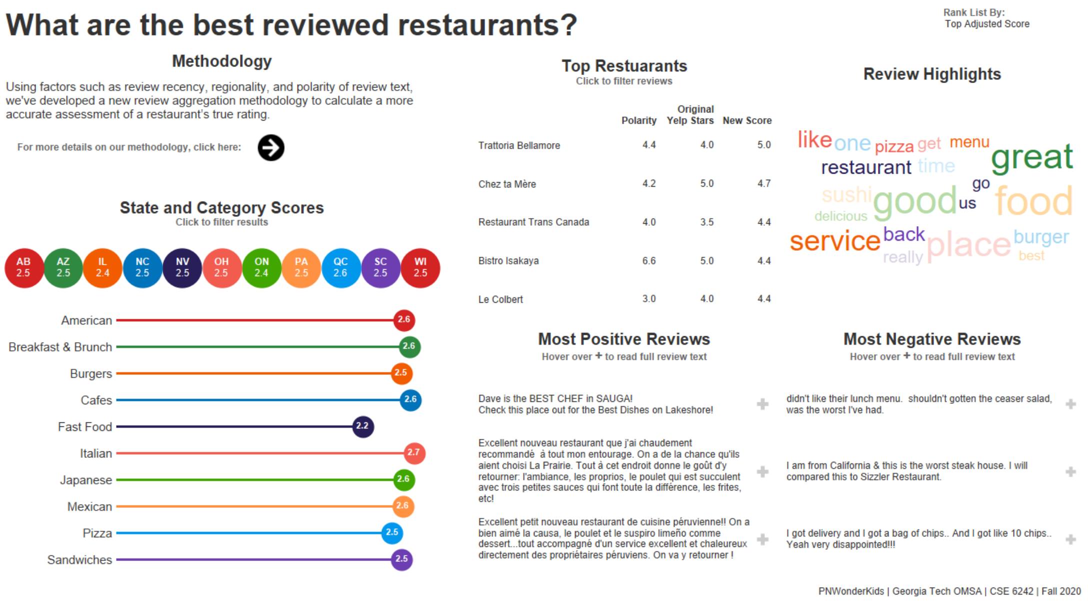


Figure 4: Final Visualization – Restaurant Explorer



What are the best reviewed restaurants?

Rank List By:
Top Adjusted Score

Methodology

Using factors such as review recency, regionality, and polarity of review text, we've developed a new review aggregation methodology to calculate a more accurate assessment of a restaurant's true rating.

For more details on our methodology, click here:

State and Category Scores
Click to filter results

Top Restuarants
Click to filter reviews

	Polarity	Original Yelp Stars	New Score
Trattoria Bellamore	4.4	4.0	5.0
Chez ta Mère	4.2	5.0	4.7
Restaurant Trans Canada	4.0	3.5	4.4
Bistro Isakaya	6.6	5.0	4.4
Le Colbert	3.0	4.0	4.4

Review Highlights

Most Positive Reviews
Hover over + to read full review text

Dave is the BEST CHEF in SAUGA!
Check this place out for the Best Dishes on Lakeshore! +

Excellent nouveau restaurant que j'ai chaudement recommandé à tout mon entourage. On a de la chance qu'ils aient choisi La Prairie. Tout à cet endroit donne le goût d'y retourner: l'ambiance, les proprios, le poulet qui est succulent avec trois petites sauces qui font toute la différence, les frites, etc! +

Excellent petit nouveau restaurant de cuisine péruvienne!! On a bien aimé la causa, le poulet et le suspiro limeño comme dessert...tout accompagné d'un service excellent et chaleureux directement des propriétaires péruviens. On va y retourner ! +

Most Negative Reviews
Hover over + to read full review text

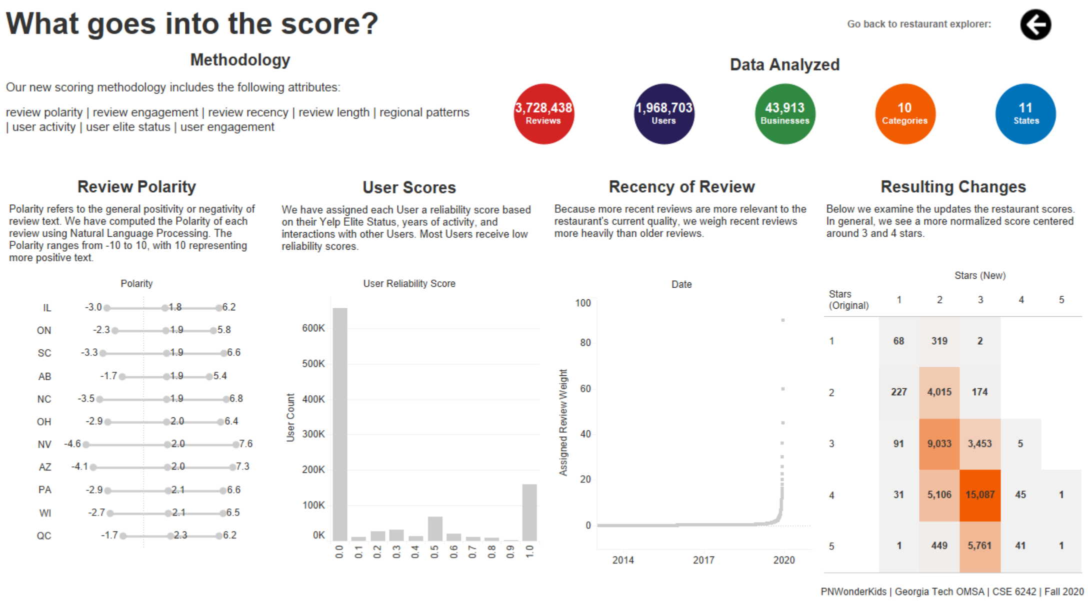
didn't like their lunch menu. shouldn't gotten the ceaser salad, was the worst I've had. +

I am from California & this is the worst steak house. I will compared this to Sizzler Restaurant. +

I got delivery and I got a bag of chips.. And I got like 10 chips.. Yeah very disappointed!!! +

PNWonderKids | Georgia Tech OMSA | CSE 6242 | Fall 2020

Figure 5: Final Visualization – Score Methodology



What goes into the score?

Go back to restaurant explorer:

Methodology

Our new scoring methodology includes the following attributes:

review polarity | review engagement | review recency | review length | regional patterns | user activity | user elite status | user engagement

Data Analyzed

3,728,438 Reviews

1,968,703 Users

43,913 Businesses

10 Categories

11 States

Review Polarity

Polarity refers to the general positivity or negativity of review text. We have computed the Polarity of each review using Natural Language Processing. The Polarity ranges from -10 to 10, with 10 representing more positive text.

User Scores

We have assigned each User a reliability score based on their Yelp Elite Status, years of activity, and interactions with other Users. Most Users receive low reliability scores.

Recency of Review

Because more recent reviews are more relevant to the restaurant's current quality, we weigh recent reviews more heavily than older reviews.

Resulting Changes

Below we examine the updates the restaurant scores. In general, we see a more normalized score centered around 3 and 4 stars.

	Stars (New)				
Stars (Original)	1	2	3	4	5
1	68	319	2		
2	227	4,015	174		
3	91	9,033	3,453	5	
4	31	5,106	15,087	45	1
5	1	449	5,761	41	1

PNWonderKids | Georgia Tech OMSA | CSE 6242 | Fall 2020

Figure 6: ANOVA of Polarity Across Regions

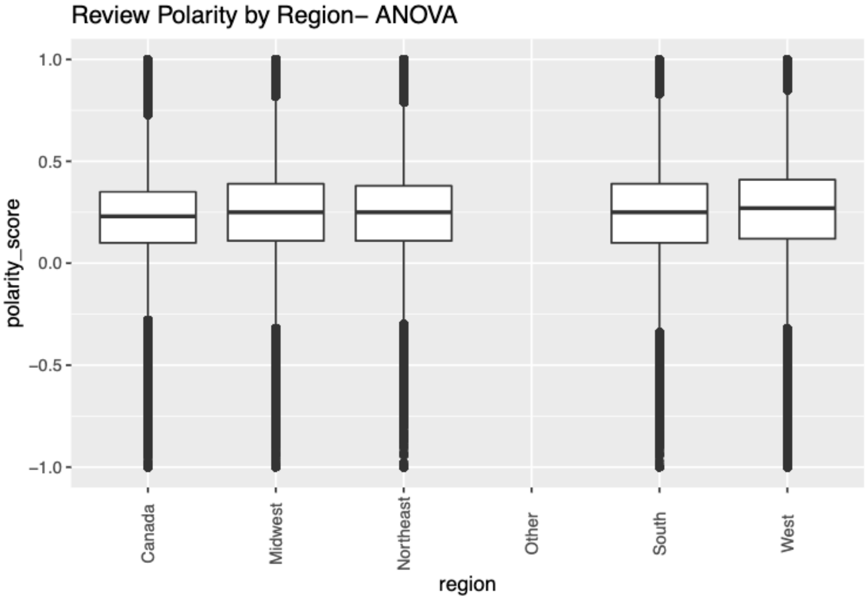
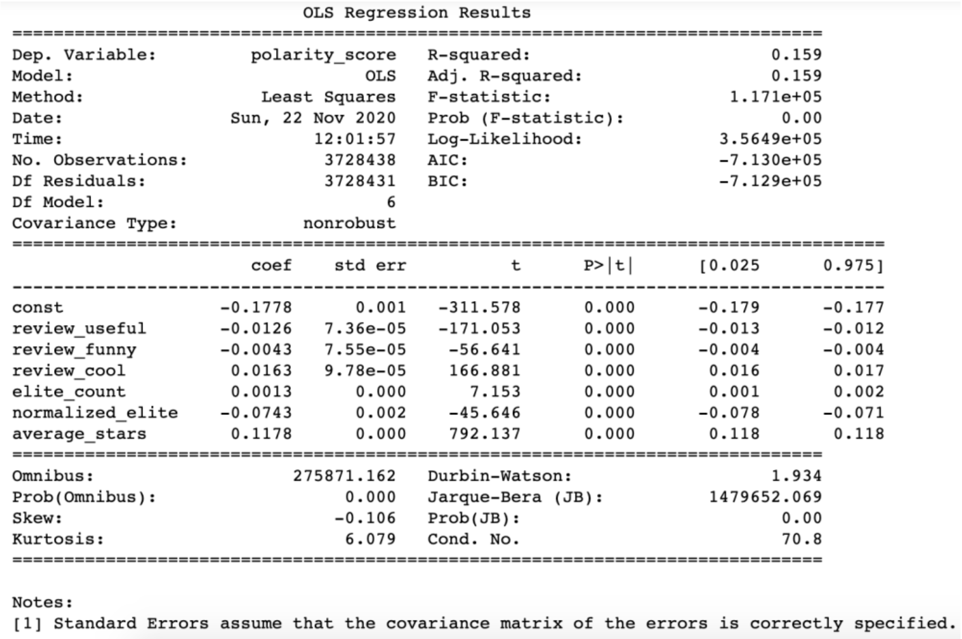


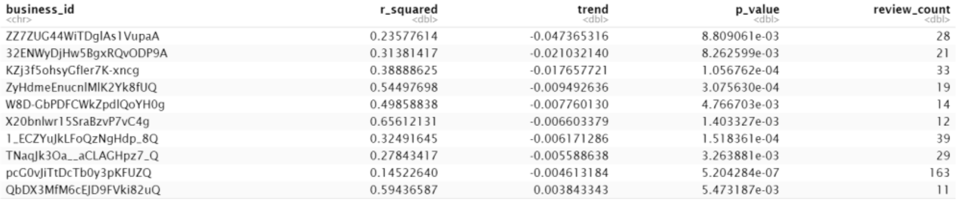
Figure 7: OLS Regression on User and Review Data vs Polarity



```
                            OLS Regression Results
==============================================================================
Dep. Variable:         polarity_score   R-squared:                       0.159
Model:                            OLS   Adj. R-squared:                  0.159
Method:                 Least Squares   F-statistic:                 1.171e+05
Date:                Sun, 22 Nov 2020   Prob (F-statistic):               0.00
Time:                        12:01:57   Log-Likelihood:             3.5649e+05
No. Observations:             3728438   AIC:                        -7.130e+05
Df Residuals:                 3728431   BIC:                        -7.129e+05
Df Model:                           6
Covariance Type:            nonrobust
======================================================================================
                        coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
const                -0.1778      0.001   -311.578      0.000      -0.179      -0.177
review_useful        -0.0126   7.36e-05   -171.053      0.000      -0.013      -0.012
review_funny         -0.0043   7.55e-05    -56.641      0.000      -0.004      -0.004
review_cool           0.0163   9.78e-05    166.881      0.000       0.016       0.017
elite_count           0.0013      0.000      7.153      0.000       0.001       0.002
normalized_elite     -0.0743      0.002    -45.646      0.000      -0.078      -0.071
average_stars         0.1178      0.000    792.137      0.000       0.118       0.118
==============================================================================
Omnibus:                   275871.162   Durbin-Watson:                   1.934
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1479652.069
Skew:                          -0.106   Prob(JB):                         0.00
Kurtosis:                       6.079   Cond. No.                         70.8
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 8: Linear Regression Output Examples with Significance



business_id	r_squared	trend	p_value	review_count
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
ZZ7ZUG44WiTDglAs1VupaA	0.23577614	-0.047365316	8.809061e-03	28
32ENWyDjHw5BgxRQvODP9A	0.31381417	-0.021032140	8.262599e-03	21
KZj3f5ohsyGfler7K-xncg	0.38888625	-0.017657721	1.056762e-04	33
ZyHdmeEnucnlMlK2Yk8fUQ	0.54497698	-0.009492636	3.075630e-04	19
W8D-GbPDFCWkZpdlQoYH0g	0.49858838	-0.007760130	4.766703e-03	14
X20bnlwr15SraBzvP7vC4g	0.65612131	-0.006603379	1.403327e-03	12
1_ECZYuJkLFoQzNgHdp_8Q	0.32491645	-0.006171286	1.518361e-04	39
TNaqJk3Oa__aCLAGHpz7_Q	0.27843417	-0.005588638	3.263881e-03	29
pcG0vJiTtDcTb0y3pKFUZQ	0.14522640	-0.004613184	5.204284e-07	163
QbDX3MfM6cEJD9FVki82uQ	0.59436587	0.003843343	5.473187e-03	11

Figure 9: Histogram of Review Quality

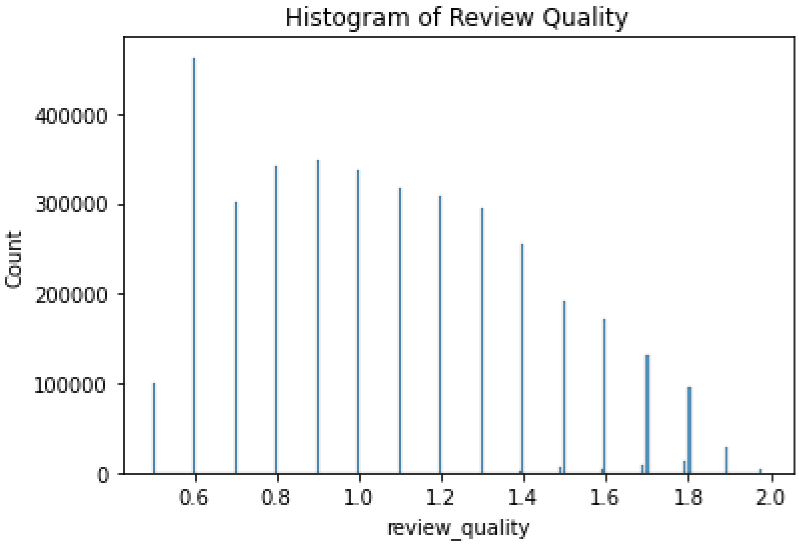
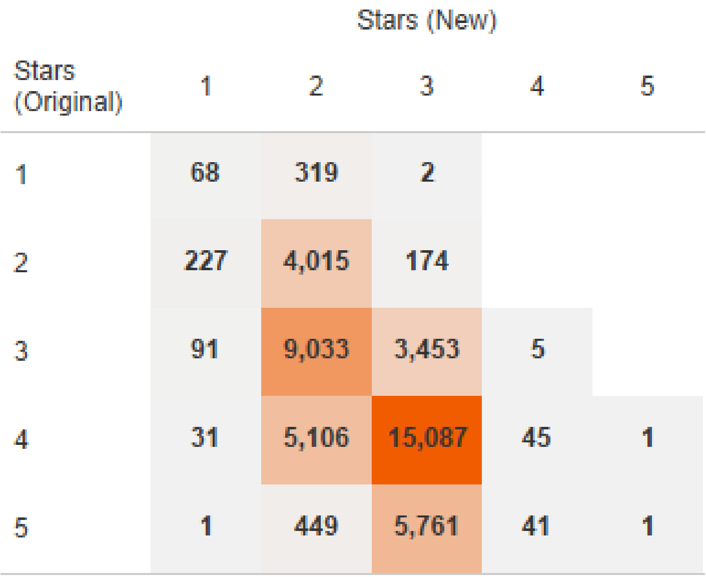


Figure 10: Before and After Distribution of Business Ratings



	Stars (New)				
Stars (Original)	1	2	3	4	5
1	68	319	2		
2	227	4,015	174		
3	91	9,033	3,453	5	
4	31	5,106	15,087	45	1
5	1	449	5,761	41	1

References

[1] Luca, M. “Reviews, Reputation, and Revenue: The Case of Yelp.com”, 2011. Retrieved from https://www.hbs.edu/faculty/Publication%20Files/12-016_a7e4a5a2-03f9-490d-b093-8f951238dba2.pdf

[2] Gan, Qiwei, and Yang Yu. "Restaurant Rating: Industrial Standard and Word-of-Mouth--A Text Mining and Multi-dimensional Sentiment Analysis." *2015 48th Hawaii International Conference on System Sciences*. *IEEE*, 2015.

[3] Yao, Yao; Angelov, Ivelin; Rasmus-Vorrath, Jack; Lee, Mooyoung; and Engels, Daniel W. "Yelp’s Review Filtering Algorithm," *SMU Data Science Review: Vol. 1 : No. 3 , Article 3*, 2018.

[4] Zhang, Ziqiong, Zili Zhang, and Rob Law. "Regional effects on customer satisfaction with restaurants." *International Journal of Contemporary Hospitality Management*, 2013.

[5] Luca, M., Lin, G., Dai, W., & Lee, J. “Optimal Aggregation of Consumer Ratings: An Application to Yelp.com”. *EconWeb*, 2014. Retrieved from http://econweb.umd.edu/~dai/dai_jin_lee_luca.pdf.

[6] Mukherjee, Arjun, Bing Liu, and Natalie Glance. "Spotting fake reviewer groups in consumer reviews." *Proceedings of the 21st international conference on World Wide Web*, 2012.

[7] Tucker, Tiana. “Online Word of Mouth: Characteristics of Yelp.com Reviews”. *The Elon Journal of Undergraduate Research in Communications: Vol.2: No.1*. 2011

[8] Nasukawa, Tetsuya, and Jeonghee Yi. "Sentiment analysis: Capturing favorability using natural language processing." *Proceedings of the 2nd international conference on Knowledge capture*, 2003.

[9] Vu, Huy Quan, Gang Li, Rob Law. “Exploring Tourist Dining Preferences Based On Restaurant Reviews.” *Journal of Travel Research*, 2017.

[10] B. Gupta et al. "Cross domain sentiment analysis using transfer learning." *2017 IEEE International Conference on Industrial and Information Systems (ICIIS)*, 2017.

[11] Govindarajan, M. “Sentiment Analysis of Restaurant Reviews Using Hybrid Classification Method.” *Proceedings of 2nd International IRF Conference*, 2014.

[12] Qiao, R. “Yelp Review Rating Prediction: Sentiment Analysis and the Neighborhood-Based Recommender.” *UCLA*, 2019.

[13] Yaakov Danone, Tsvi Kuflik, and Osnat Mokryn. “Visualizing Reviews Summaries as a Tool for Restaurants Recommendation”. 23rd International Conference on Intelligent User Interfaces (IUI '18). Association for Computing Machinery, 2018.

[14] Wang, J., Zhao, J., Guo, S., North, C., & Ramakrishnan, N. “ReCloud: Semantics-Based Word Cloud Visualization of User Reviews” *Graphics Interface Conference,* 2014. Retrieved from https://infovis.cs.vt.edu/sites/default/files/p151-wang.pdf

[15] Sutton, Dave. “4 Reasons Why (Most) Customer Satisfaction Surveys are Useless” *Top Right Marketing*, 2016. Retrieved from https://www.toprightpartners.com/insights/4-reasons-customer-satisfaction-surveys-useless/