Yelp Reviews

Summarising Online Reviews via Sentiment Mining

1. Introduction

Yelp is an online platform that publishes crowd sourced reviews by users about business such as restaurants, hotels and shops. Users rate the business on a scale of 1 to 5 stars and provide written reviews about their experiences. The aim of this analysis is to gain insights into reviewer behaviour and review characteristics looking at the review ratings and volume but also from the review text by sentiment mining.

The data set used for this analysis contained over 1.5 million reviews, with each review containing 12 variables for analysis. Data ranges from 2004 to 2015.

2. Statistical Summary of Reviews

This section presents a comprehensive statistical summary of key variables in the Yelp dataset, including stars (rating), review length, and review sentiment by examination of averages, measures of variability, and distribution patterns.

2.1. Star Ratings

Table 1
Count of Reviews with Each Star Rating

Stars	1	2	3	4	5
Count	159811	140608	222719	466599	579527

Figure 1

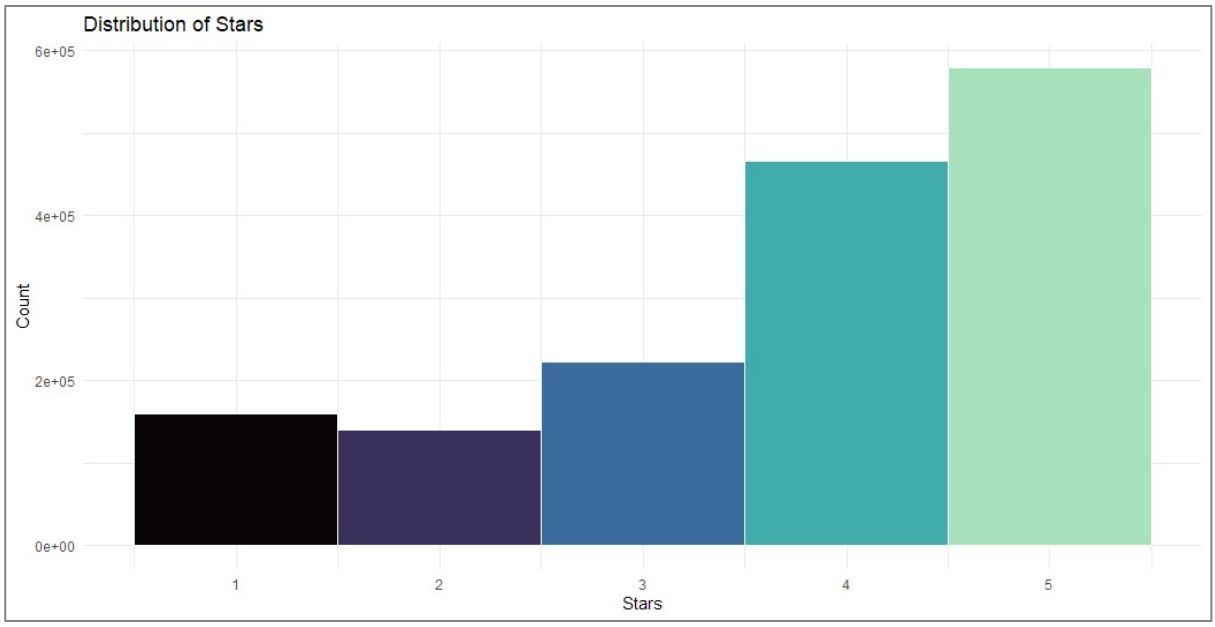


Table 2
Summary Statistics of Star Ratings

Variable	Mean	Median	Max	Min	Q1	Q3	SD
Stars	3.7	4	5	1	3	5	1.3

The median average star rating was 4 and mean average star ratings for reviews was 3.7/5 (with a standard deviation of 1.3), highlighting the data's positive skew.

The most common star rating received on Yelp was 5 stars (579,527) and the lowest was 2 stars (140,608).

The data showed a strong positive tendency in the reviews with 4- and 5-star reviews accounting for approximately two thirds of all reviews. This suggests users generally had favourable opinions of places they had visited.

2.2. Review Length

Figure 2

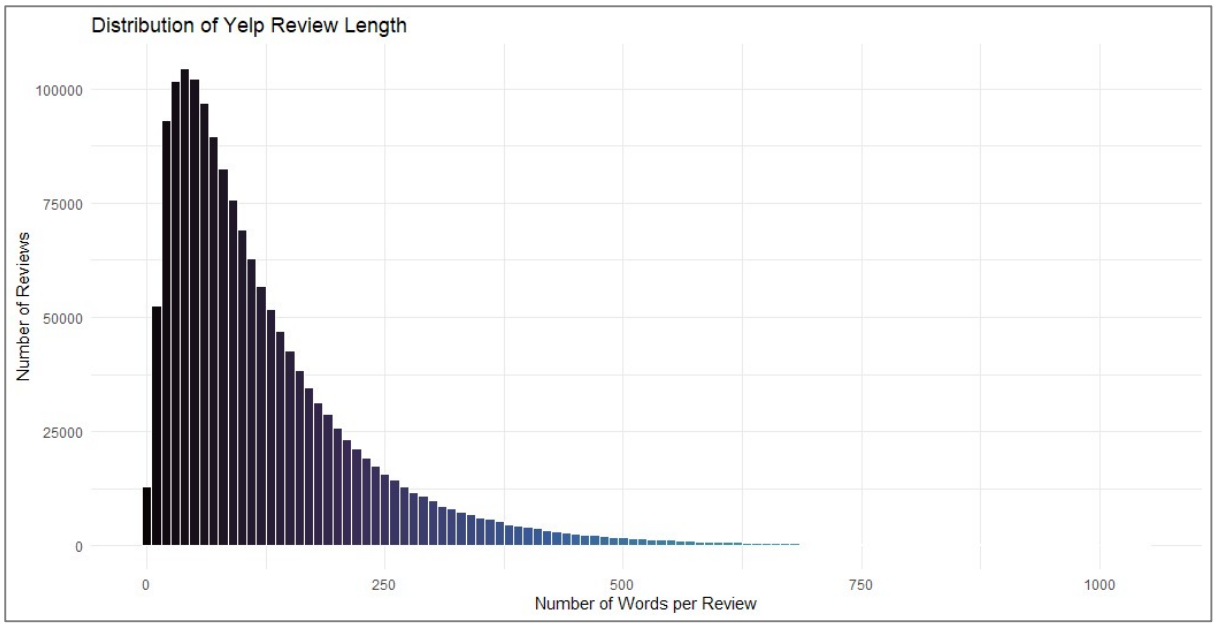


Table 3
Summary Statistics for Yelp Review Length

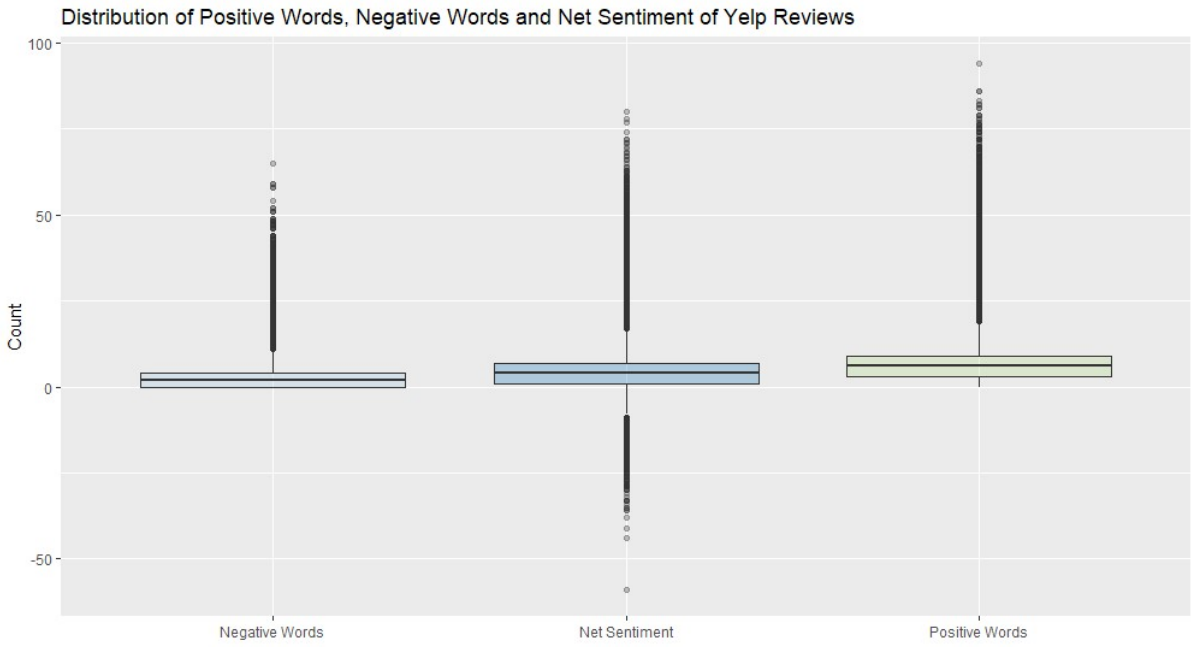
Variable	Mean	Median	Max	Min	Q1	Q3	SD
Review Length	126	92	1047	0	48	165	115.5

The median average word length of all reviews was 92 words with 50% of reviews having between 48 and 165 words, suggesting most reviews are relatively concise.

The longest review was 1047 words long. Reviews of zero length made up a relatively small proportion of the total reviews suggest that reviewers are enthusiastic about providing written feedback.

2.3. Sentiment

Figure 3



Variable	Mean	Median	Max	Min	Q1	Q3	SD
Negative Words	3	2	65	0	0	4	3.3
Net Sentiment	5	4	80	-59	1	7	5.2
Positive Words	7.1	6	94	0	3	9	5.9

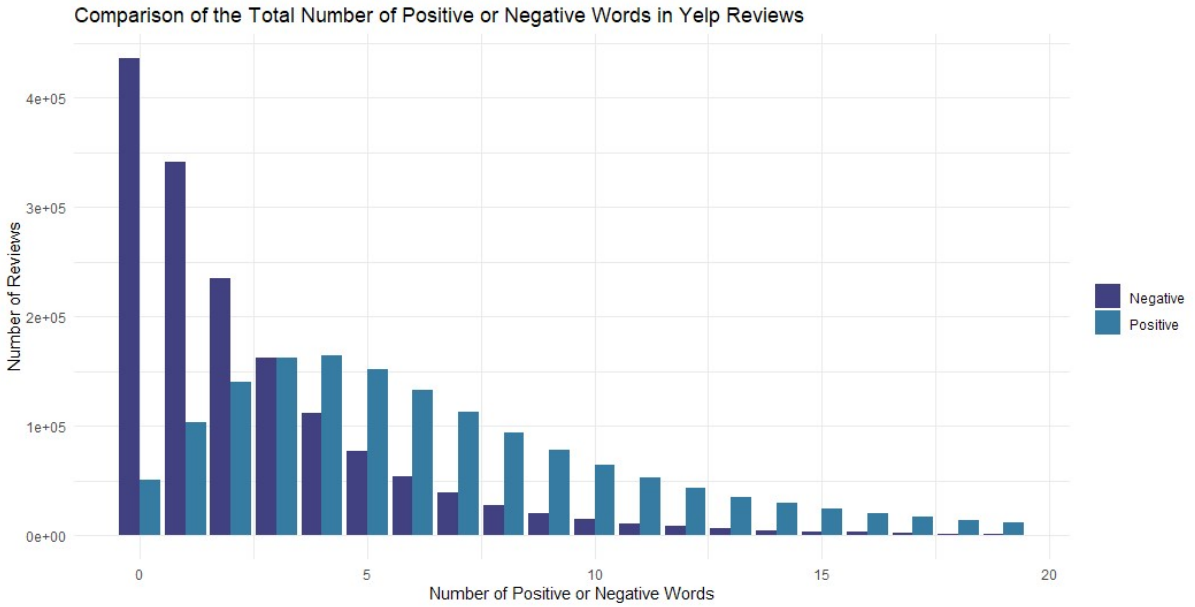
The sentiment of reviews was typically positive: the median average of positive words was higher than negative words (6 and 4 respectively) with mean average positive sentiment being 7.1.

The general positive sentiment observed here correlates with the skew for positive star ratings seen previously.

3. Sentiment Analysis

Net sentiment is the difference between the number of positive and negative words a review contains; a positive sentiment means more positive words were used and therefore implies a positive review.

Figure 4



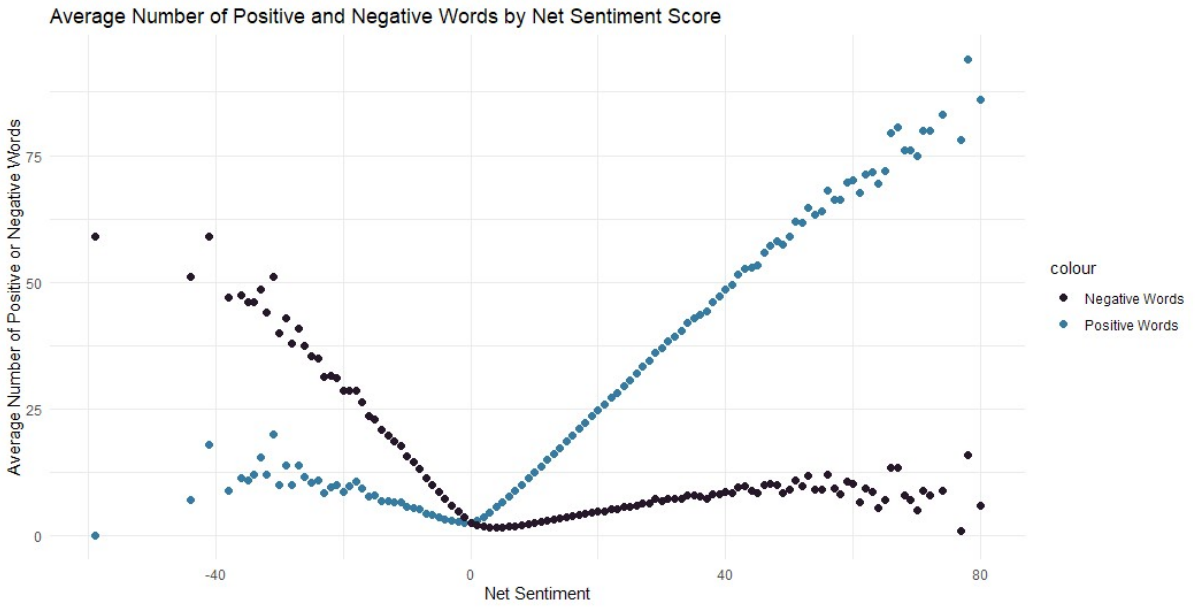
Approximately a quarter of all reviews (436,419) contained zero negative words with a notable downward trend as the number of words increases.

In contrast, the number of positive words shows a different distribution to negative with more of a bell-shaped distribution. 50,339 reviews contained no positive words increasing to the most common amount of 5 positive words (151,969 reviews). The number of positive words per review showed a more gradual decrease compared to negative words.

This data suggests that most reviews tend to have a low to moderate number of either negative or positive words. Reviews with large numbers of positive or negative words were uncommon.

Despite reviews tending to have low to moderate numbers of positive or negative words, the differences in distribution can be seen in the reviews' net sentiment:

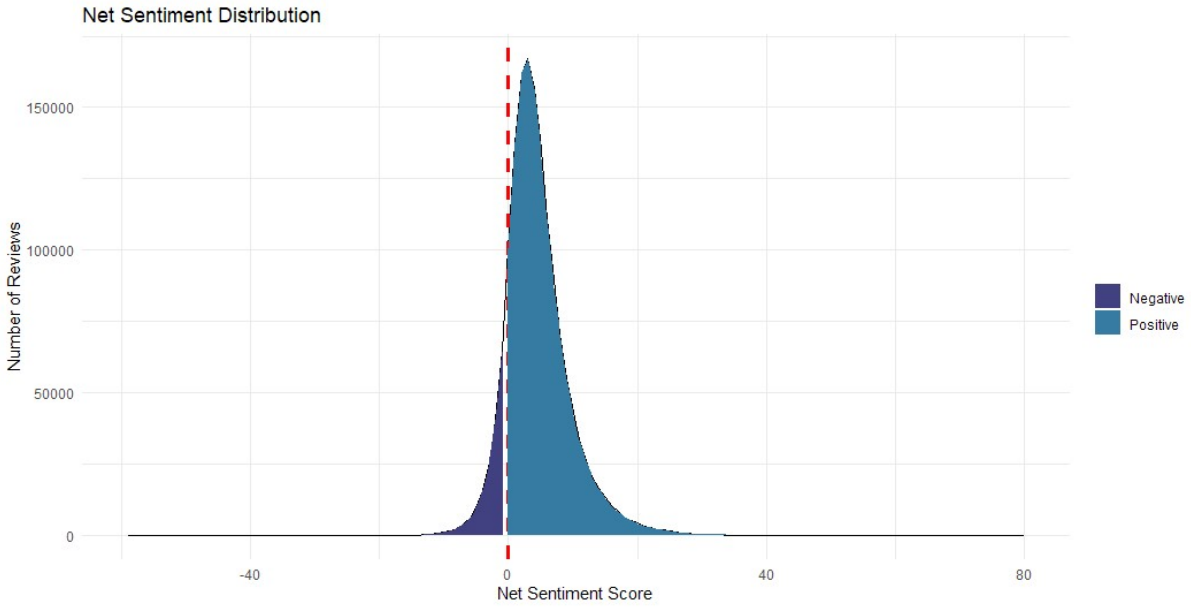
Figure 5



As net sentiment is derived from these two counts the 'positive' relationship each has on sentiment can be seen above. This chart also demonstrates the 'heterogeneous' nature of reviews in that, on average, they contain a mixture of positive and negative words. This suggests potential balance in reviews in that reviewers will often highlight the good as well as the bad (and vice versa).

Finally, the skew of net sentiment towards positive scores can be seen below. The average score (the peak) is greater than zero and the majority of the area under the curve is also above zero highlighting the strong positive skew.

Figure 6



4. Review Length Analysis

This analysis will investigate the relationship between average review length and star rating and provide insight into whether the customer experience affects the extent they are willing to share their experiences.

Figure 7

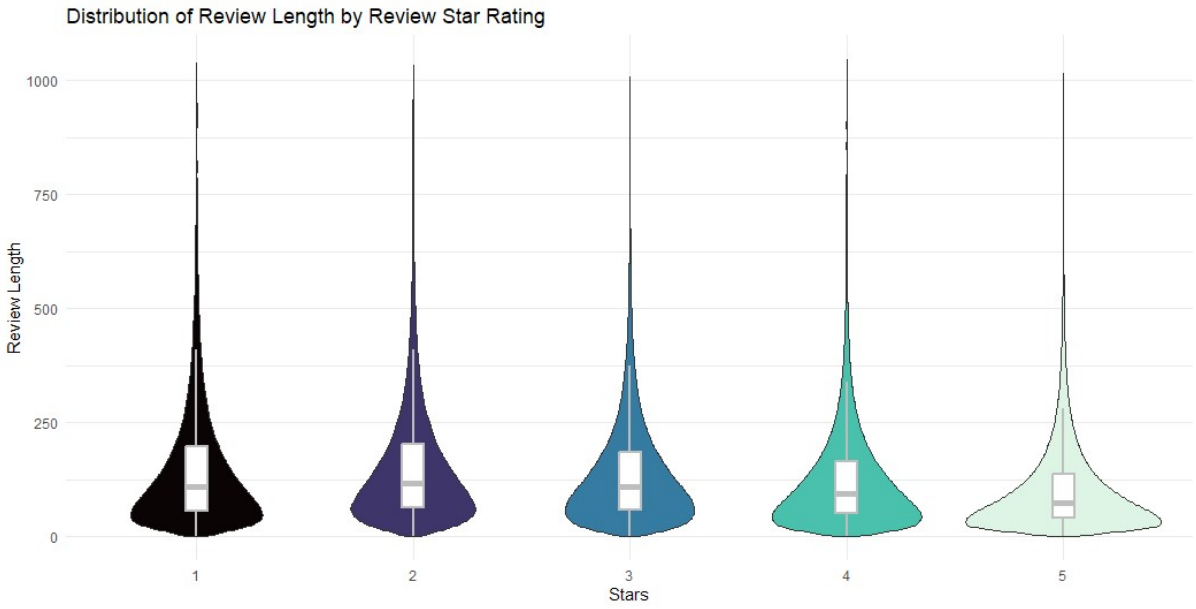


Table 4
Summary Statistics of Reviews at Each Star Rating

Stars	Median Length	Q1	Q3	Min	Max	Count
1	109	57	198	0	1039	159811
2	117	64	202	0	1033	140608
3	109	59	185	0	1009	222719
4	95	50	166	0	1047	466599
5	75	40	137	0	1017	579527

Analysis of review length by each star category showed that the data was not normally distributed with a strong positive skew (for longer reviews) in each category. This is explained by the presence of a small number of lengthy reviews in each category.

As the data is not normally distributed a mean average is not appropriate for assessing review lengths; the presence of large 'outlier' reviews will likely result in the mean average being overestimated. A median average will therefore be used.

One-star reviews had the highest median average length, 109 words while the shortest reviews were typically 5-star reviews with 75 words with a trend for decreasing length as star rating increases. Unhappier customers appear to have more to say.

Despite this trend the relationship between review length and star rating does not appear likely to be statistically significant. Taking into account the variability of reviews using the inter-quartile range (IQR), the median average length of 5-star reviews (75) is within the IQR (57-198) of 1-star reviews, suggesting that the difference may not be significant.

The word count of the longest reviews was notably similar (1009-1039 words) across all star categories, suggesting that reviewers' length of reviews was not related to the quality of their experience.

In conclusion, this data suggests review length may not be strongly influenced by the customers experience alone.

5. Useful Review Analysis

Yelp has a feature where users can 'vote' on whether they found another review 'useful'. This analysis will investigate whether voting a review as useful was influenced by review length (do users favour more concise or more in-depth reviews?) or by star rating (do users find warnings or endorsements more useful?).

It is worth noting that most reviews (>50%) on Yelp have 0 votes for usefulness:

Figure 8

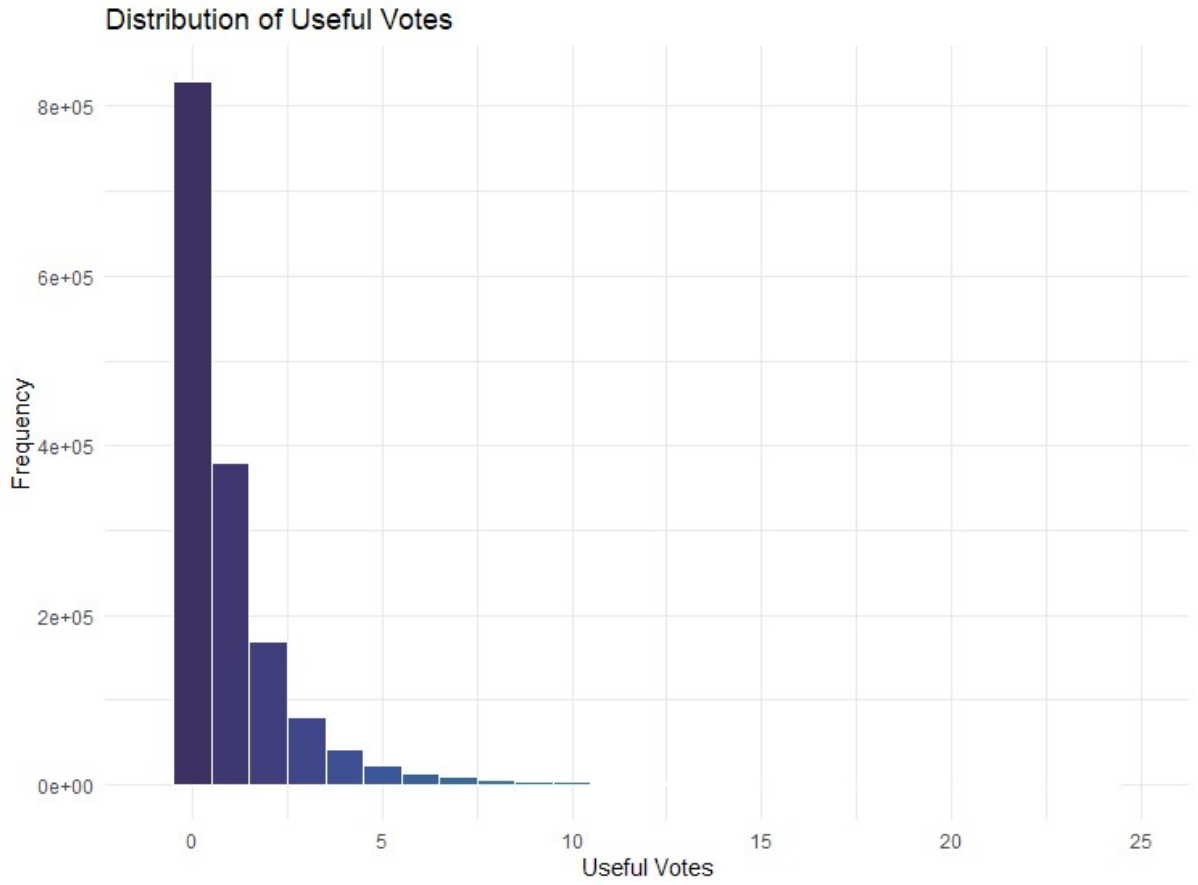


Table 5

Total Votes Cast	Min	Max	Reviews with zero votes	Reviews with >0 votes
1681655	0	166	829,315	739,949

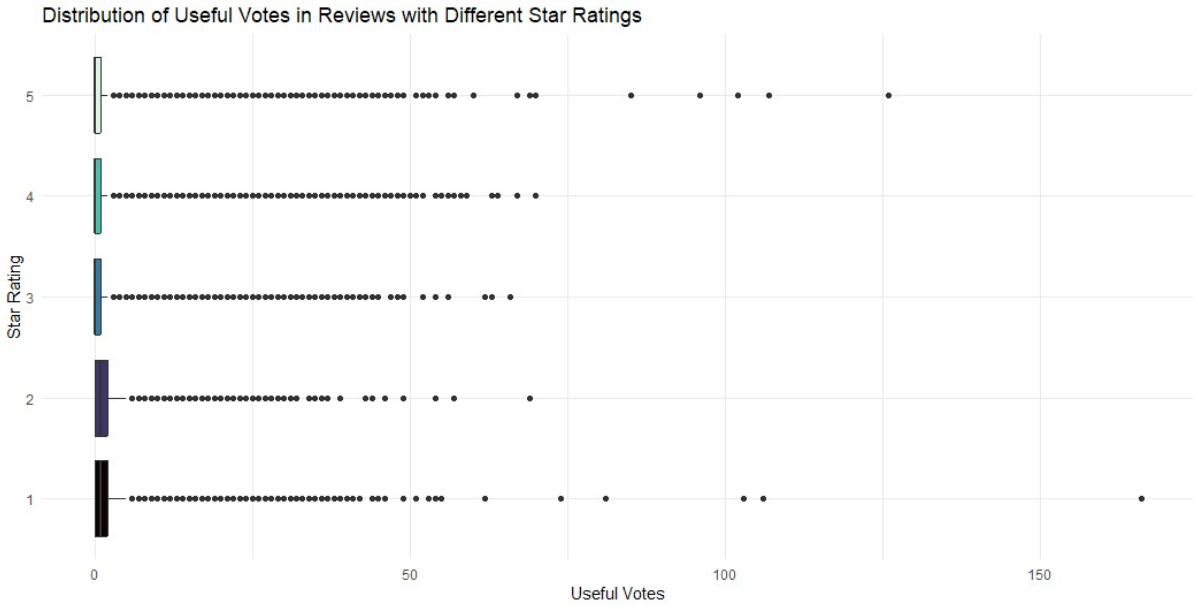
However, 739,99 reviews were voted on by Yelp users, for a total 1.68 million votes cast with the most 'popular' review receiving 166 votes.

5.1. Useful Votes and Star Rating Analysis

This analysis will investigate whether Yelp users find reviews with a lower or higher star rating more useful.

Table 6
Summary Statistics of Useful Votes at Each Star Category

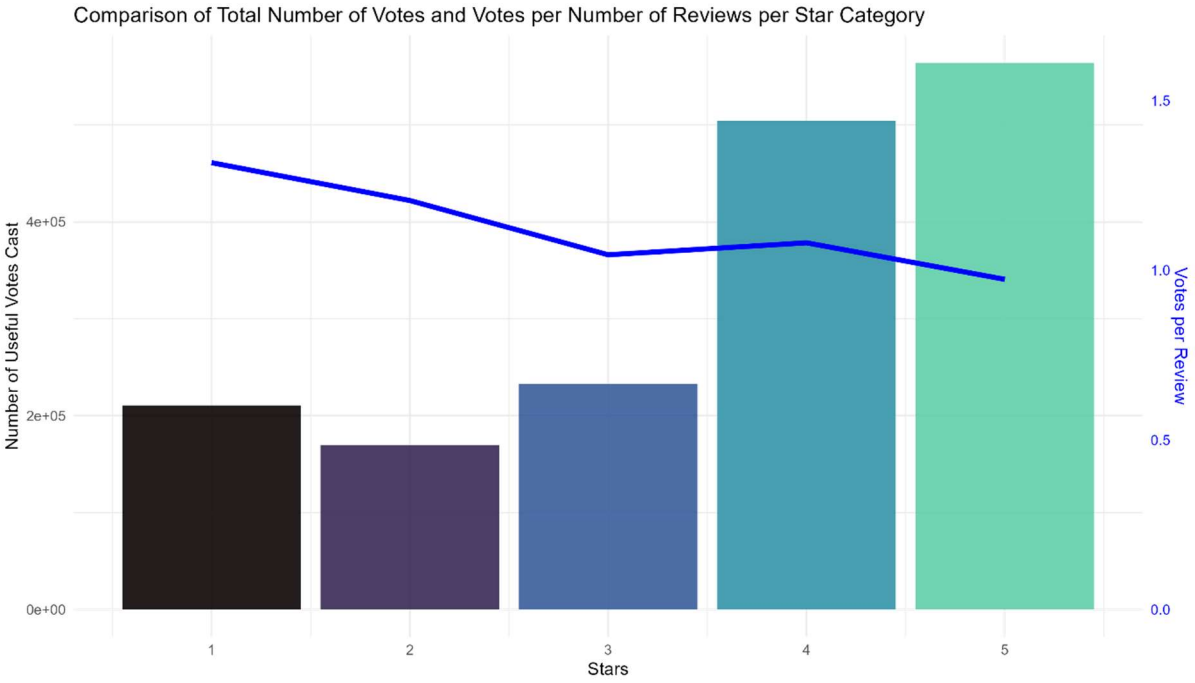
Stars	Median	Q1	Q3	Min	Max
1	1	0	2	0	166
2	1	0	2	0	69
3	0	0	1	0	66
4	0	0	1	0	70
5	0	0	1	0	126



Distribution analysis of the data grouped by the reviews' star rating shows a similar pattern with most reviews receiving low to no reviews. The outlier pattern ('extreme') review length shows similar patterns. This data highlights the longest review; 166 votes for useful on a 1-star review.

Analysis of median averages and IQR shows a trend for an increase in median average length of review as the number of stars decreases however it does not appear that this would be statistically significant.

Figure 9



Analysis of the volume shows that higher reviews (4 and 5 star) have more votes for usefulness. However, as noted previously, these reviews account for almost two thirds of this data set, making it an unreliable comparison. Factoring in the number of reviews and looking at "votes per review" there is again a trend for users finding lower rated reviews more useful.

In conclusion, there is a slight trend for users finding 1 star reviews the most useful. On average, these reviews appear more likely to be voted useful and have a slightly higher average number of votes. Further supporting the current findings, a 1-star review has also attracted the highest vote. This is a trend only and no statistically significant conclusions may be drawn.

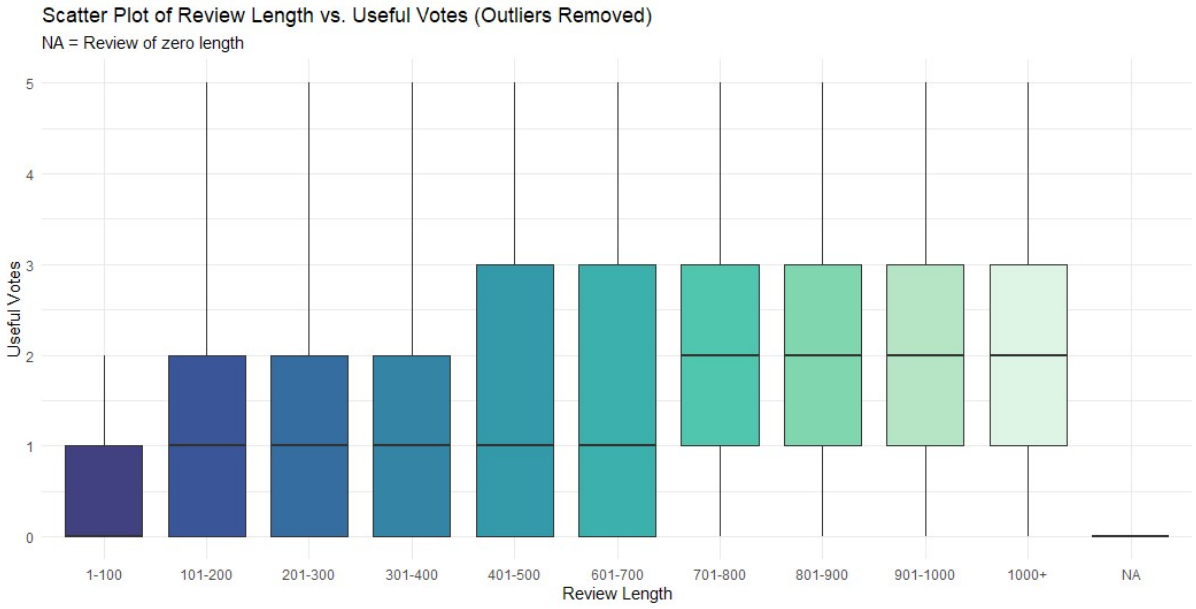
5.2. Useful Votes and Star Rating Analysis

The analysis below explores whether Yelp users find reviews with a lower or higher length more useful.

Table 7
Summary Statistics of Useful Votes by Review Length

Review length category	Median	Q1	Q3	Min	Max	Count (Reviews)	Sum (votes)
1-100	0	0	1	0	126	846768	500771
101-200	1	0	2	0	74	441304	550756
201-300	1	0	2	0	106	164518	309642
301-400	1	0	3	0	96	64294	155105
401-500	2	1	4	0	70	27312	77404
601-700	2	1	4	0	166	12228	39878
701-800	2	1	4	0	56	5876	20540
801-900	3	1	5	0	48	3096	11831
901-1000	2	1	5	0	81	2313	9792
1000+	3	1	5	0	62	1384	5867
NA	0	0	0	0	5	171	69

Figure 10

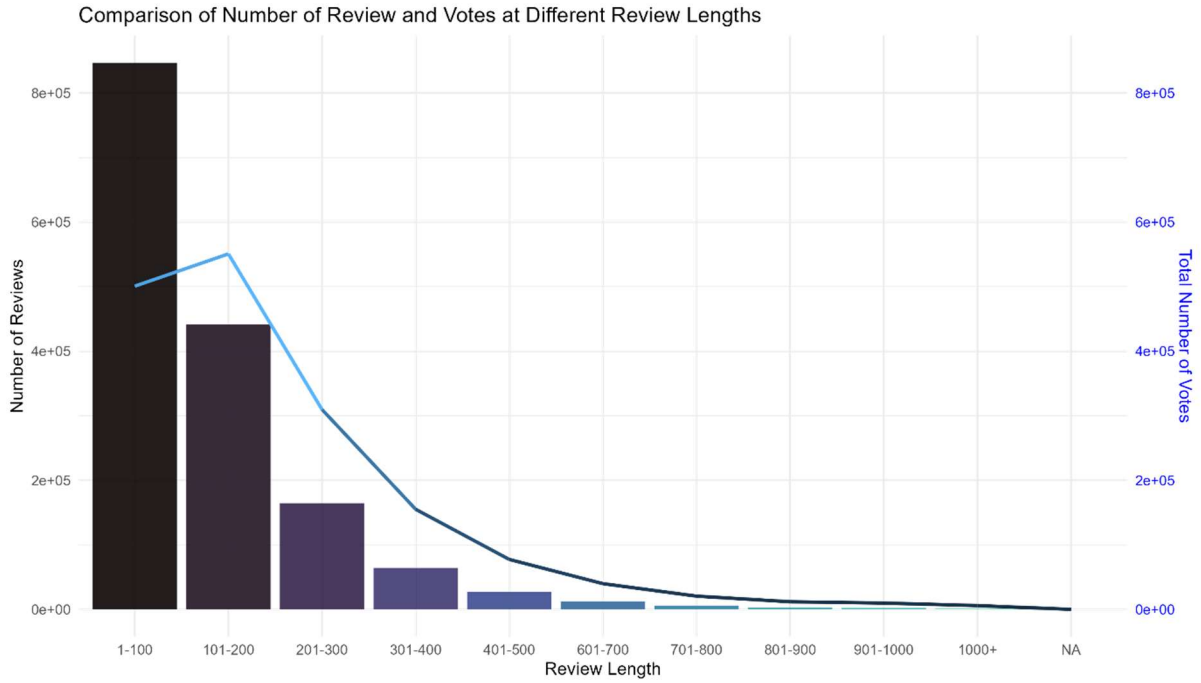


Analysis of central tendency and distribution showed a trend for more votes as review length increased with higher median averages and Q3's. The difference between short reviews (<100 words) and longer reviews (400 + words) may be statistically significant.

Vote counts were highest among shorter reviews (101-200 length attracted 550,756 votes) however it should be noted that short reviews made up a large proportion of reviews.

Analysis of volume data showed a correlation between the total number of votes received and the total number of reviews at a specific length; however, one notable departure from this trend is reviews 101-200 words in length, which showed an increase in votes despite a decreasing number of reviews (as indicated in *Figure 11*).

Figure 11



In conclusion, this analysis suggests a positive relationship between review length and the number of positive votes received.

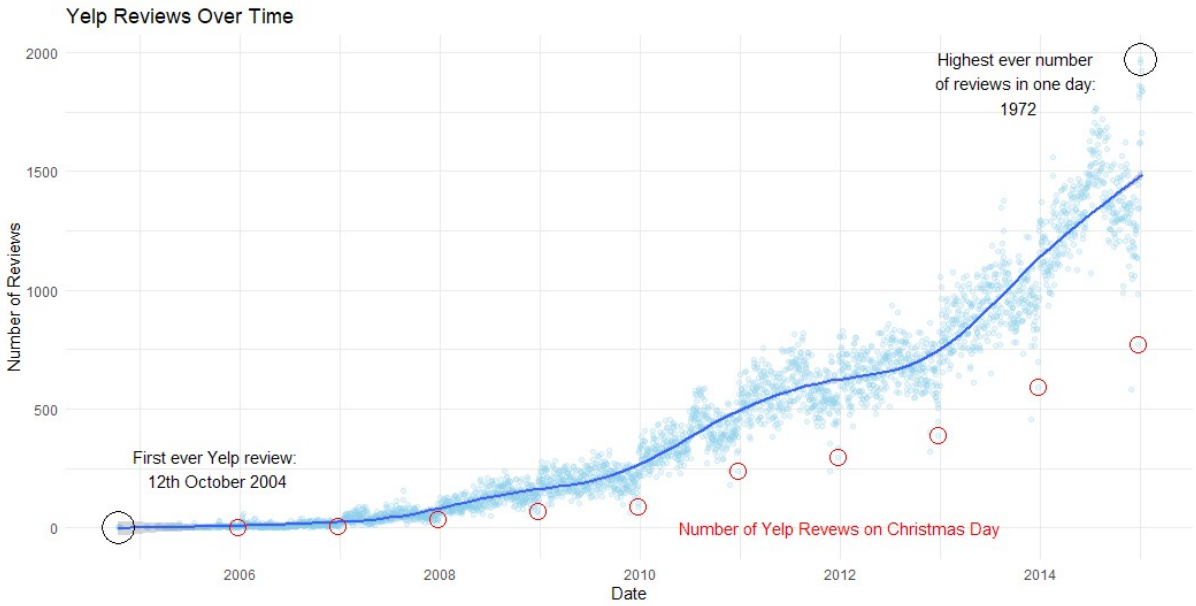
6. Reviews over Time Analysis

This analysis will investigate trends in the number of Yelp reviews between 2004 and 2015. Data has been grouped into years for ease and brevity of analysis.

Table 8
The Number of Yelp Reviews Over Time

Year	Number of Reviews	YOY Increase	YOY Increase (%)	YOY Growth (%)
2004	13	NA	NA	NA
2005	680	667	5130.8	NA
2006	4239	3559	523.4	433.6
2007	17724	13485	318.1	278.9
2008	45117	27393	154.6	103.1
2009	72948	27831	61.7	1.6
2010	137764	64816	88.9	132.9
2011	209429	71665	52.0	10.6
2012	244106	34677	16.6	-51.6
2013	336273	92167	37.8	165.8
2014	486306	150033	44.6	62.8
2015	14665	NA – Incomplete data		

Figure 12



Yelp has experienced varying rates of reviews per year since 2004. Using growth (the increase in the number of new reviews compared to the previous year) Yelp has consistently doubled (growth >100%) the number of reviews on its platform in 5 out of 9 years.

2009 and 2011 were notably flat for growth with the number of new reviews similar to the previous year. 2012 (34677 new reviews) was notably lower than the previous year (71665 new reviews) with -51.6% growth.

More recently, the platform appears to be growing strongly recently achieving its highest number of reviews in one day.

Interestingly, some Yelpers do post reviews on Christmas day, however they do appear to be in the minority as it appears to be one of the quietest days for Yelp activity each year. These Yelpers do appear to be happy though with an average review sentiment of 4.6.

In conclusion, Yelps growth has been inconsistent over the last decade with periods of exponential and relatively flat growth identified.

7. Best Business and Best Customer Analysis

This analysis will attempt to identify the best business and best user on Yelp.

7.1. Best Business on Yelp

Table 9
Summary Table for the 'Top 10' Yelp Businesses

Business ID	Number of Reviews	Average Star Rating	Average Net Sentiment	Weighted Star Rating	Weighted Sentiment
4bEjOyTaDG24SY 5TxsaUNQ	4137	4.1	6.3	17128	26216
2e2e7WgqU1Bnp xmQL5jbfw	3517	4.3	4.2	15145	14615
zt1TpTuJ6y9n551 sw9TaEg	3352	3.7	5.0	12569	16646
sIyHTizqAiGu12X MLX3N3g	2657	3.9	6.0	10300	16010
YNQgak-ZLtYJQxlDwN-qIg	2619	3.7	5.2	9748	13712

Figure 13

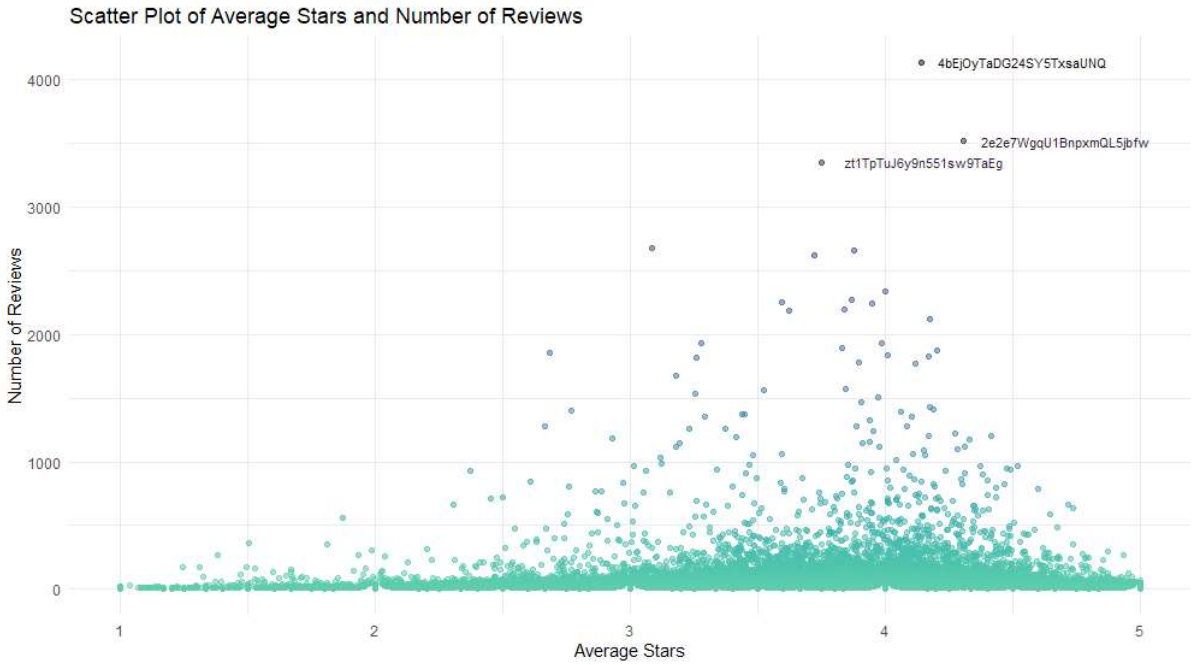
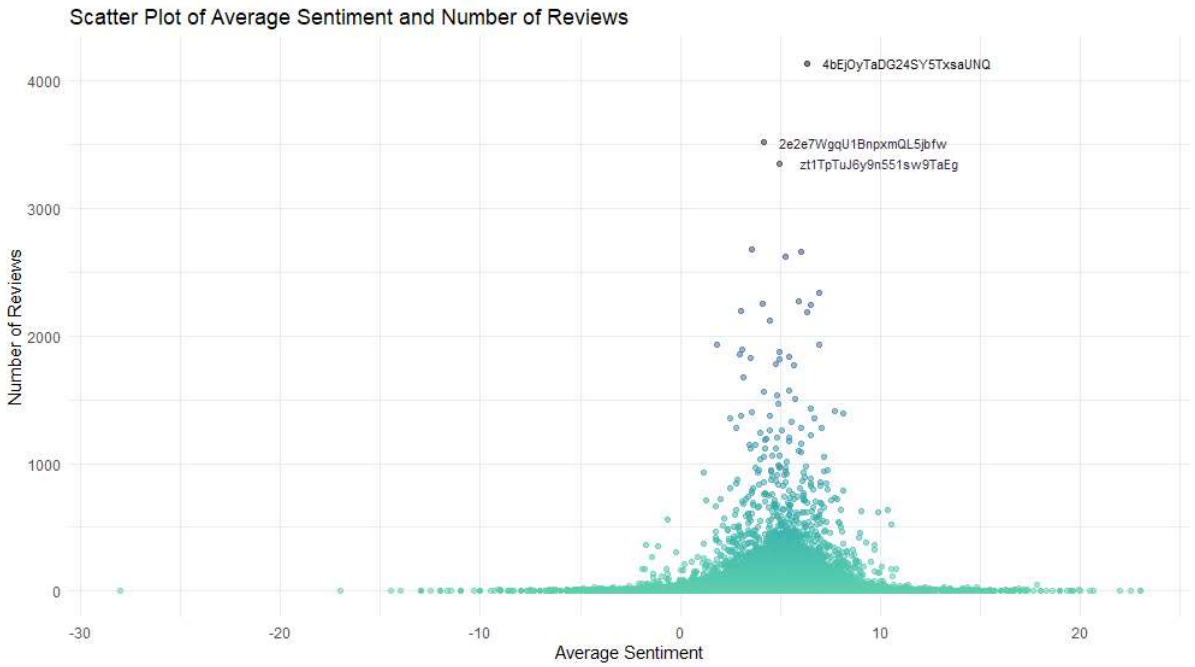


Figure 14



To identify the best business, average star review and average review sentiment are obvious choices. However, as Figures X and Y demonstrate, it is important to consider the number of reviews power those averages. For instance, a business with a 5.0 average review may not be the best business if the sum of its reviews was one.

Star ratings and net sentiment were therefore weighted by the total number of reviews as follows:

$$weighted\ star\ average = \frac{(stars * no.of\ reviews)}{no.of\ reviews}$$

This is a simple approach and does have its limitations; for example, it may favour more established businesses that have accrued more reviews than newer businesses. While this is an average and changes over time will be captured, it also does not factor in recent trends and may not be informative for a user looking for the best business 'today'.

Using this criteria, business "4bEjOyTaDG24SY5TxsaUNQ" can be said to be the best as it has the highest weighted average star rating and net sentiment rating. This business also has the most reviews out of any business assessed which may be driving this finding. This also demonstrates the consistency at which the business satisfies its customers.

7.2. Best User on Yelp

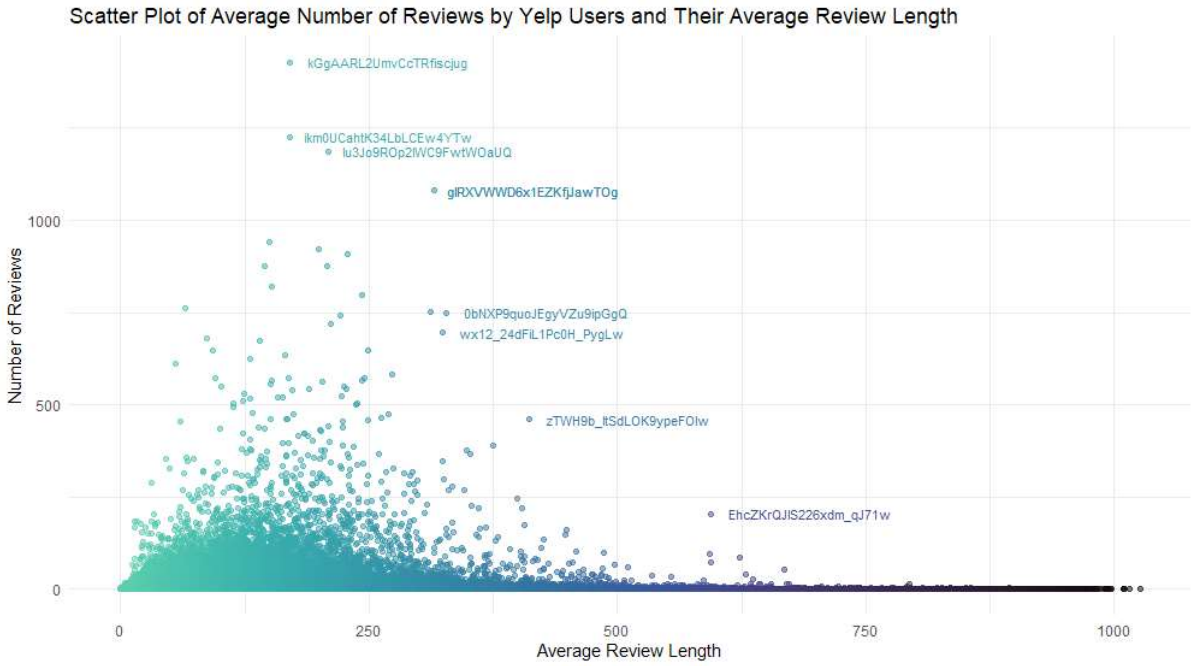
To identify the best user on Yelp, users will be assessed using 2 parameters: how prolific they are and how popular their reviews are.

7.2.1. Most prolific user on Yelp

Table 10
Most prolific users

User ID	No. of Reviews	Av. Review Length	Total Words Output
glRXVWWD6x1EZKfjJawTOg	1080	315	340510
lu3Jo9Rop2IWC9FwtWOaUQ	1186	209	248329
0bNXP9quoJEgyVZu9ipGgQ	750	328	245855
kGgAARL2UmvCcTRfiscjug	1427	170	242339
5lq4LkrviYgQ4LJNsBYHcA	753	311	234391
wx12_24dFiL1Pc0H_PygLw	695	324	225332
ikm0UcahtK34LbLCEw4Ytw	1225	170	208371
fczQCSmaWF78toLEmb0Zsw	908	229	207681
4ozupHULqGyO42s3zNUzOQ	796	243	193570
zTWH9b_ItSdLOK9ypeFOIw	461	411	189313

Figure 15

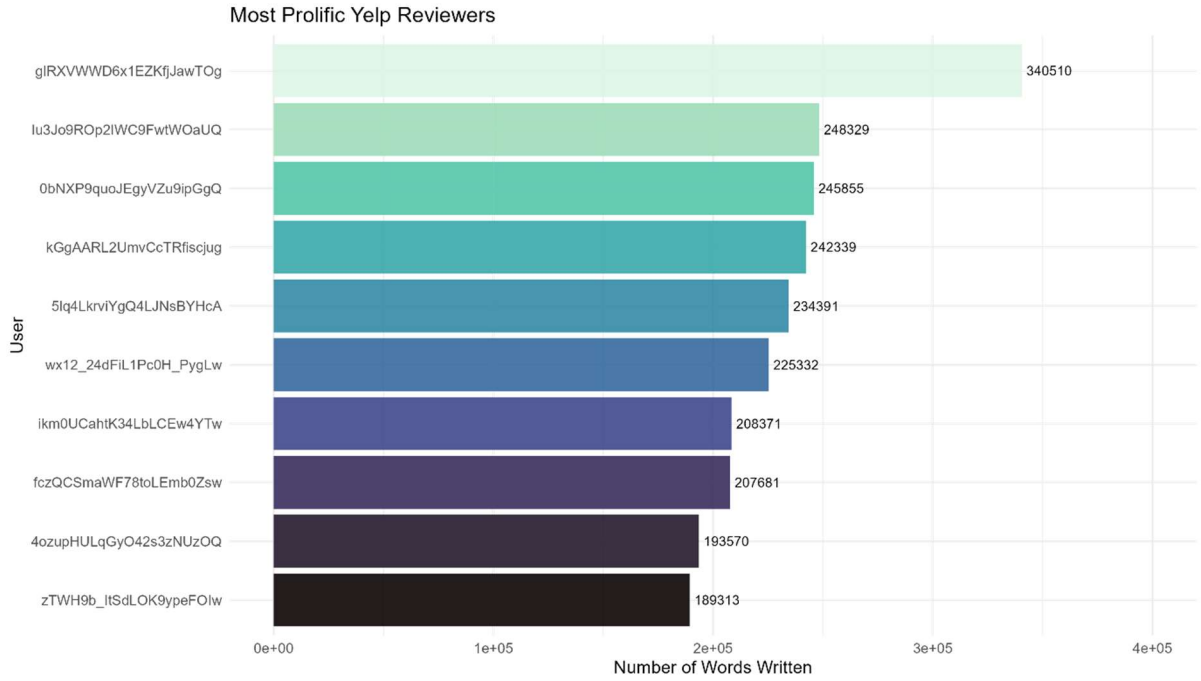


As a metric for analysis a user could be described as “prolific” if they write many and/or lengthy reviews.

Figure 15 shows the distribution of all users in terms of their mean average review length and the number of reviews that they have written. No clear winner can be identified from this plot as some users have many shorter reviews compared to others who have fewer but lengthier reviews.

A total words output was estimated from this data, which was the average review length multiplied by the number of the reviews.

Figure 16



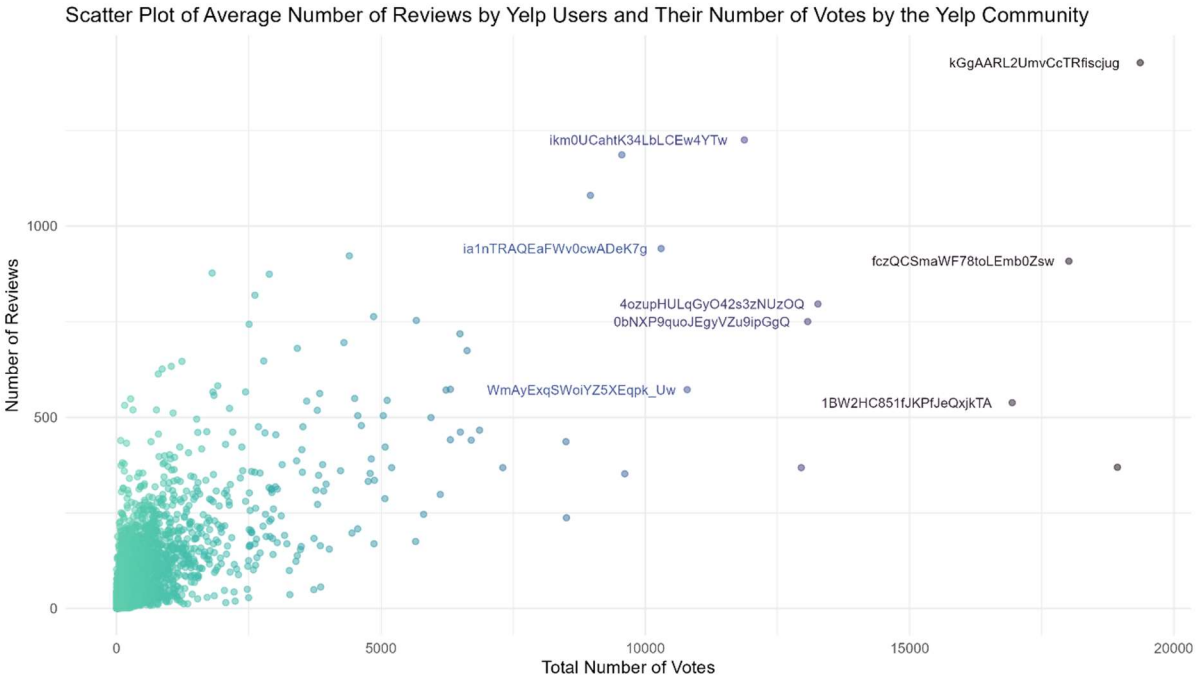
Using this metric, user glRXVWWD6x1EZKfjJawTOg with 1080 reviews and an average review length of 315 words can be described as Yelps most prolific user.

7.2.2. Most popular user on Yelp

Table 11
Most popular users

User ID	Number of Reviews	Total Votes	Votes per Review
C8Ztiwa7qWoPSMIivTeSfw	369	18933	51
2HmHgW3hRYvXYFmQyQtLuw	368	12953	35
1BW2HC851fJKPfJeQxjkTA	538	16943	31
NvDR3SPVPXrDB_dbKuGoWA	352	9614	27
MWt24-6bfv_OHLKhwMQ0Tw	298	6123	21
kSVYpNWA19wUplbdi0U0Uw	368	7306	20
fczQCSmaWF78toLEmb0Zsw	908	18015	20
P2kVk4clWyK4e4h14RhK-Q	436	8501	19
WmAyExqSWoiYZ5Xeqpk_Uw	572	10794	19
6kmu0mYbdpMIOZ6Y0eVsxg	287	5076	18

Figure 17



Reviewers may be described as popular by the Yelp community based on the number of votes their reviews received.

Assessment of the distribution shows some users with a large number of votes from a large number of reviews. It cannot be said for certain they are popular, perhaps just prolific, so the number of votes receiving was normalised against volume of reviews to assess how popular the users' reviews were on average.

This method has its limitations as it will favour long time users who have many reviews; however, users will have to consistently post popular reviews to have a good score.

Figure 18

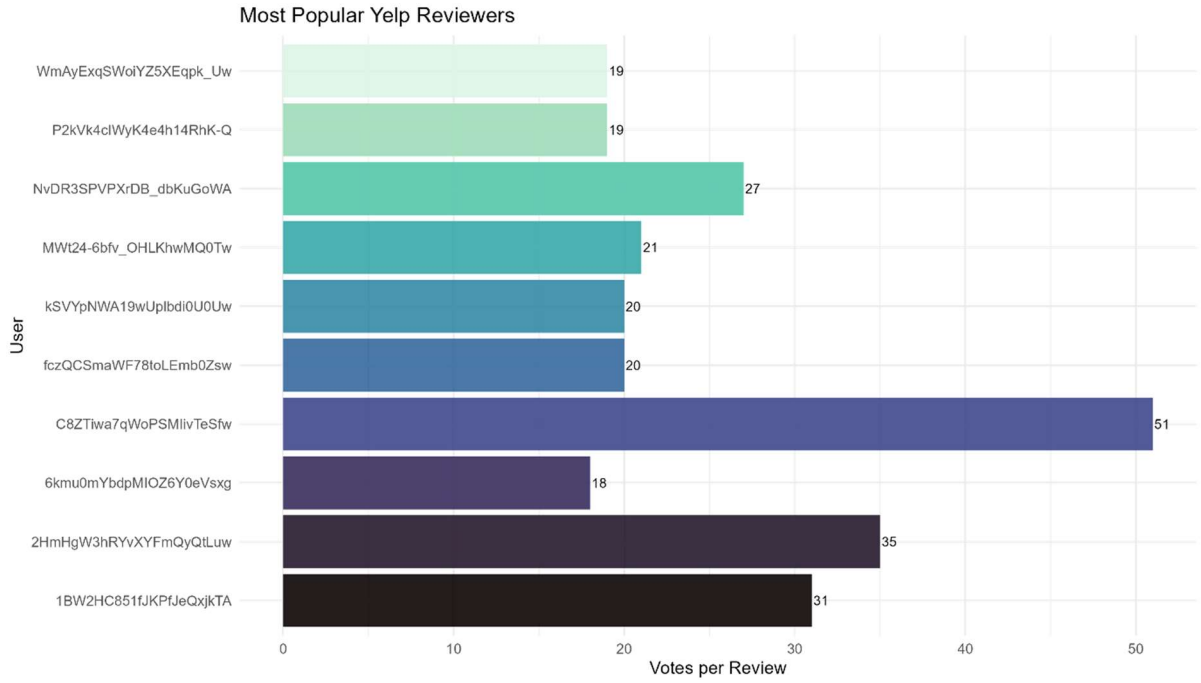
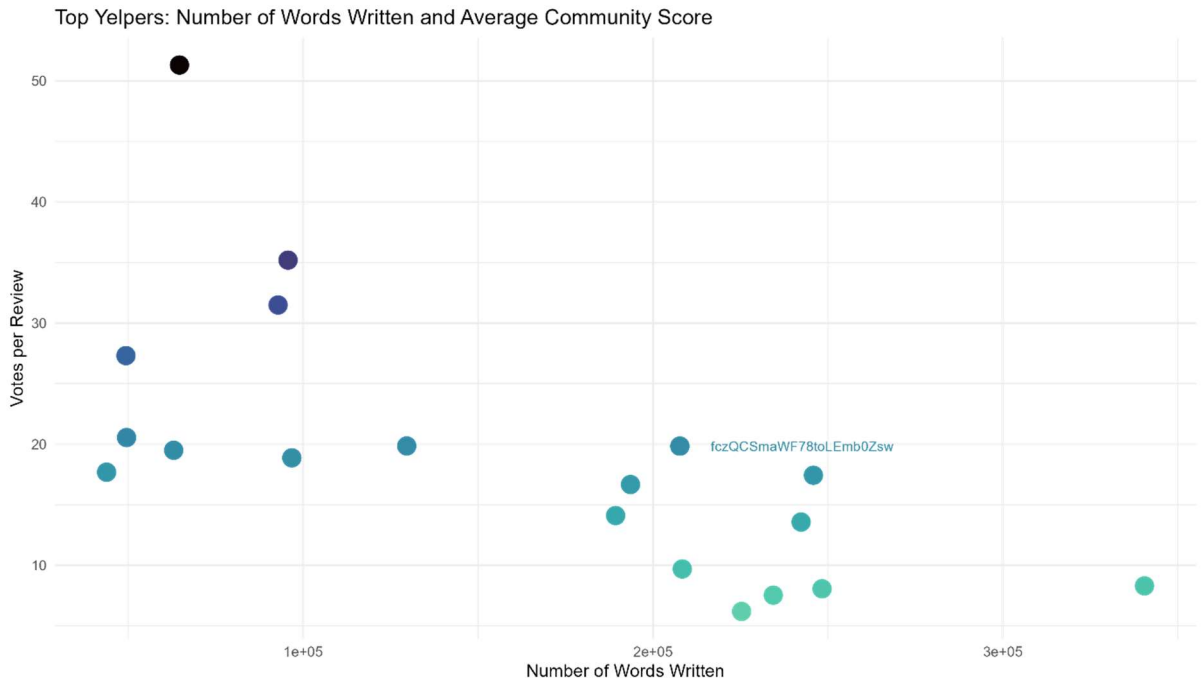


Figure 17 shows that user C8Ztiwa7qWoPSMIivTeSfw is the most popular reviewer with an average of 51 votes per review across their 369 reviews.

7.2.3. Best user on Yelp

Figure 19



Prolific and popular users are both of benefit to Yelp; prolific users provide the volume of reviews that allow Yelp to maintain its relevance to its customers. Popular users generate quality content for the site that ensure Yelp's reliability as a source of information.

Combining these two user qualities to determine the overall best user, one user is ranked in the top 10 for prolific and popular: fczQCSmaWF78toLEmb0Zsw. Given the quantity and quality of reviews they produce, fczQCSmaWF78toLEmb0Zsw can be determined as Yelp's best user.

8. Conclusions.

Reviews were generally positive in rating (mean = 3.7 stars) and sentiment as well as being short in nature (average = 92 words). Sentiment analysis showed that while reviews were biased towards having positive sentiment, reviews were balanced, containing both positive and negative words.

There was a trend for longer length reviews when star ratings were lower, but this relationship does not appear statistically significant, suggesting that review length may also be affected by other factors. There was a trend for longer reviews attracting more useful votes.

Yelp's growth has been inconsistent with years of rapid, exponential growth and years of limited growth.

The best business on Yelp was identified as the one having the highest average star rating and net sentiment rating, weighted against the total number of reviews and the best user on Yelp was selected based on the quality and quantity of reviews they produced.