Introduction

Yelp is a well-known website where users can post and read evaluations for a variety of businesses, including restaurants and cafes. The success or failure of a small business can sometimes be determined by its online ratings, making Yelp an indispensable platform. This study's objective is to analyse a dataset containing restaurant and café Yelp ratings in order to draw statistical conclusions and identify patterns.

This dataset, named "yelp_reviews.csv," is crammed with useful information. User_id and business_id provide anonymized identification for reviewers and businesses, respectively. Each evaluation also contains a date, allowing us to examine tendencies over time. By analysing this data set, we can learn about reviewers' attitudes, rating distributions, trends in review length, and the correlations between these and other factors. These findings will shed light on the dynamics and potential influence of restaurant and café Yelp reviews.

Analysis

Statistical Summary

From the below output of the descriptive analysis of the variables in the dataset, it can be drawn that the median star rating, number of positive words, and net sentiment score are all above average, indicating that the majority of Yelp reviews are positive. Standard deviations and the prevalence of lower-rated reviews, negative words, and lower net sentiment scores indicate, however, that ratings and feelings fluctuate. Despite the fact that most comments are concise, occasionally reviewers provide extensive commentary.

```
> stars_summary
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   3.000   4.000   3.743   5.000   5.000
> review_length_summary
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0    48.0    92.0   125.6   165.0  1047.0
> pos_words_summary
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   3.000   6.000   7.072   9.000  94.000
> neg_words_summary
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    0.00    2.00    2.55    4.00   65.00
> net_sentiment_summary
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-59.000   1.000   4.000   4.522   7.000  80.000
```

Trends you see in this data
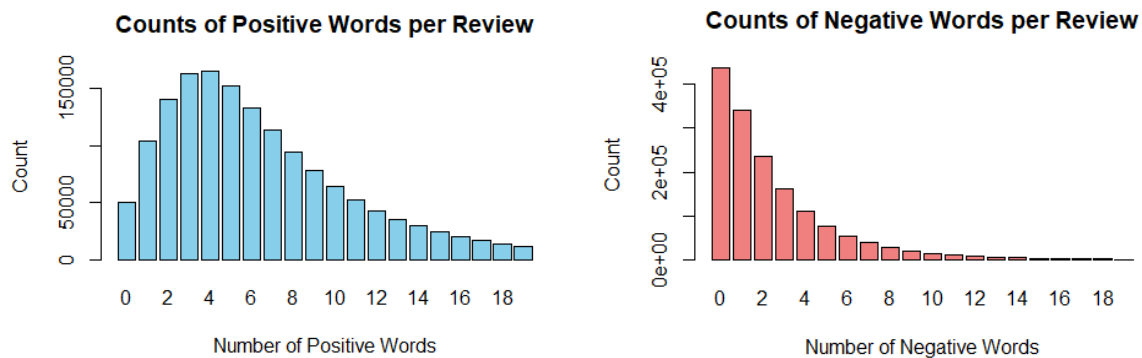
The above graphs indicate that the average number of positive words in reviews is lower than expected, indicating a right-skewed normal distribution. In addition, the other plot for negative words reveals a declining trend in the average number of negative words per review, suggesting that reviewers are becoming less critical on the whole. These tendencies may be indicative of an improvement in review quality over time and a favourable review climate in general.
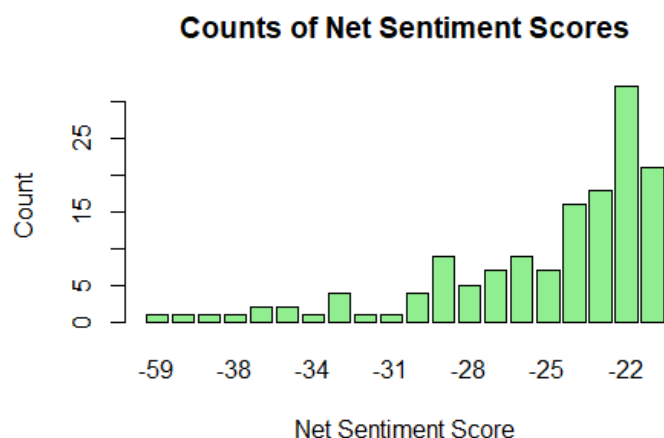
The graph above displays a left-skewed distribution, indicating that the overwhelming majority of Yelp reviewers hold favourable opinions of the businesses they have rated. This indicates that evaluators had, on average, a positive experience and rarely expressed discontent.

The average review length per star category has been calculated, and the results have been graphically represented for presentation and discussion. Due to the continuous nature of evaluation duration, the arithmetic mean should be used to calculate an aggregate average. And also, the arithmetic mean as the average provides a clear indication of the average review length across all rating tiers.

Average Review Length per Star Category
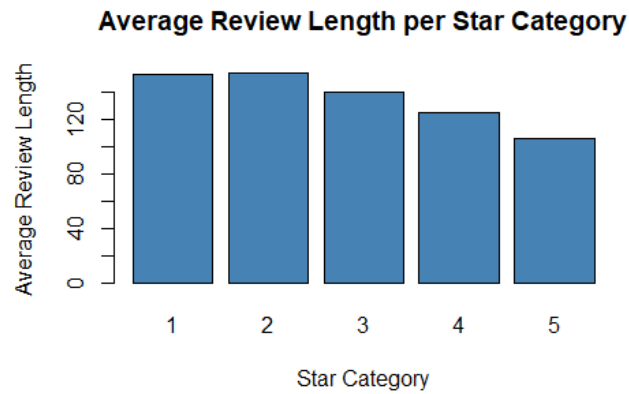
Based on the data shown above, it seems that reviewers tend to give higher-rated companies shorter ratings and lower-rated businesses lengthier evaluations.

Usefulness of Reviews by Star Rating

As indicated by the above small box with numerous whiskers and outliers in the box and whisker plot, there appears to be a wide variety of usefulness ratings and opinions of usefulness within each star rating category. The subjective nature of user opinions is emphasised, as is the need to consider multiple perspectives when evaluating the value of reviews.

Relationship between Usefulness and Review Leng
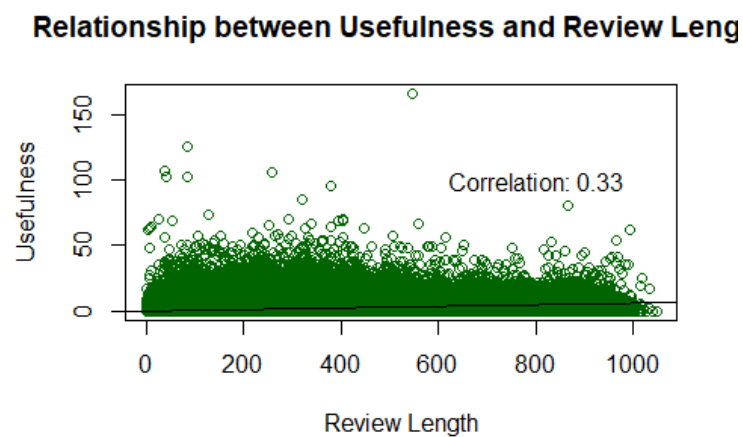
The correlation between review length and usefulness is moderately positive (0.33 correlation coefficient) with slightly increasing trend line which suggests that the longer the review is, the more useful it becomes. This means that, lengthier reviews have a higher likelihood of being helpful.

Number of Yelp Reviews per Day
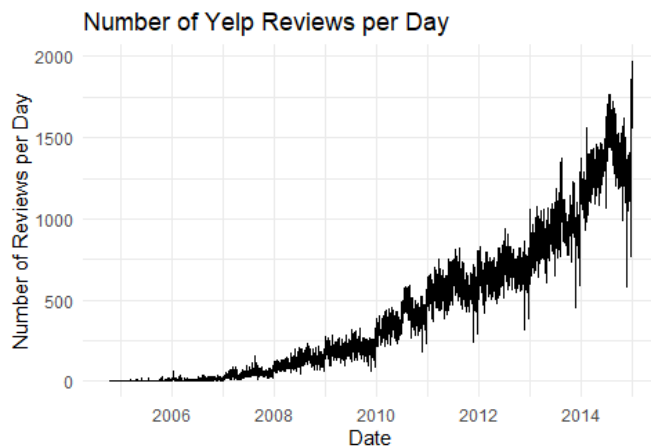
As per the above plot, the number of reviews per day is increasing over time. In general, an increase in the number of Yelp reviews each day is indicative of increasing Yelp's involvement and impact among local businesses.

Best Business & Customer

Best Business Criteria:

- Businesses whose average review score is four or more stars are considered to be among the best in the industry. Customers have had nothing but good experiences, which is great to hear.
- Businesses that have received the most five-star ratings are considered to have the most favourable evaluations. A high number of five-star ratings is indicative of satisfied customers and a solid reputation.
- Businesses with the fewest negative ratings are considered to be in the top category. This is indicative of happy customers and a successful transaction.
- The Businesses whose reviews have garnered the most likes or comments are shown. This indicates that users are engaged and enthusiastic.

By following the criteria for best business, "tAdd__IgXQEknDDicEbRgQ" is the best business.

business_id	average_rating	positive_reviews	negative_reviews	total_votes	total_reviews
<chr>	<dbl>	<int>	<int>	<int>	<int>
tAdd__IgXQEknDDicEbRgQ	5	68	1	180	69

Best Customer Criteria:

- Top Users based on the amount of "useful" votes their reviews earned. This means that people find their evaluations to be informative and useful.

- Users who have gotten the most upvotes or comments on their reviews are considered to have the most favourable feedback. This indicates that consumers find value in and agree with their reviews.
- Reviews from individuals who regularly provide 5 stars are considered reliable. This points to an accurate and favourable evaluation of companies.
- Those users who have posted the most reviews are considered the most active. This shows that you actively participate in and contribute to the Yelp community.

By following the criteria for best customer, "kGgAARL2UmvCcTRfiscjug" is the best customer.

```
user_id                    helpful_votes positive_feedback average_rating total_reviews
<chr>                              <int>             <int>          <dbl>         <int>
kGgAARL2UmvCcTRfiscjug              8785             19365           3.63          1427
```

Conclusion

As a result of analysing Yelp reviews, a number of conclusions were drawn:

- The majority of reviews were positive, with the vast majority awarding the product four or five stars. Fewer reviews awarded the businesses one or two stars, indicating that the majority of customers were satisfied.
- There was a wide range of review durations, ranging from very brief to relatively extensive. Customers tended to leave shorter reviews for establishments with higher star ratings, indicating greater satisfaction with their experiences.
- The distribution of the number of positive words per review was biased to the right, indicating that the majority of reviews contained few positive comments. There was, however, a declining trend in the number of negative terms per review, indicating that they were used less frequently overall.
- A left-skewed distribution of net sentiment ratings indicates that the overwhelming majority of reviews were positive. This would suggest that consumers had more positive experiences and were more inclined to post positive ratings than negative ones.
- There was a correlation between review star rating, review duration, and review utility, as measured by the number of votes each review received. This indicates that reviews regarded beneficial had a higher average rating and were longer in length.
- As the site has become more popular, the daily average number of reviews posted has increased, demonstrating its reliability and credibility.

These results provide organisations with vital information that they can use to prioritise customer satisfaction and increase their star ratings and number of positive reviews. Users may take the dependability and applicability of reviews into account. The increasing daily volume of reviews is evidence of the web's ability to make or destroy a business.