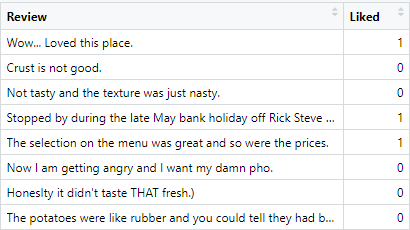
Introduction

Textual analysis better known as text analysis is a machine learning method for extracting useful information from unstructured text input. Text analysis tools are used by businesses to quickly ingest web data and documents and turn them into plain comprehensible insights upon which a business or a company can take actions in the future to boost revenue.

Text analysis can be used to extract specific information from hundreds of emails, such as keywords, names, or corporate information, or to categorize survey comments by sentiments or topic.

Data understanding

We will conduct text analysis upon restaurant review data. There are only two rows in this dataset. Customer Reviews and Liked are two rows in this data collection. Consumer reviews contain customer feedback in form of comments on restaurant's food or service, while the liked column indicates whether or not they enjoyed the meal. There are 1000 rows and two columns in the dataset. Customer reviews are in the review column, and likes are in the like column. If the review is positive, it has been labeled as 1; if it is negative, then it has been labelled as 0. Below is a sample of the dataset.



Raw dataset

Most restaurants seek customers for feedback, and based on the feedback, the restaurant can increase customer satisfaction, as customer feedback helps restaurants understand in which domain they need to improve in. As a result, reviews are critical to the restaurant's continued success. Hence, we will be performing text analysis on this dataset.

Upon this dataset we will perform two types of Text analysis:  
1. Sentiment Analysis: To understand the sentiment of the customers  
2. Bigram Analysis: To understand where to improve and what is working

Method 1: Simple Sentiment Analysis

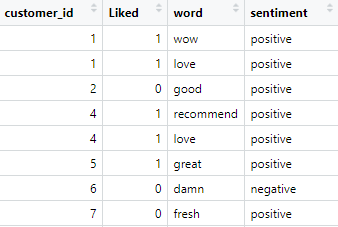
One of the most common uses of sentiment analysis in the business world is brand monitoring. Bad reviews can quickly accumulate on the internet, and the longer you wait to respond, the worse the situation can get. Negative brand mentions can be promptly alerted to you using sentiment analysis technologies.

Performing sentimental analysis on the restaurant dataset can help us track the image and reputation of the restaurant, allowing us to look at the success of the restaurant and what the customers think about the restaurant.

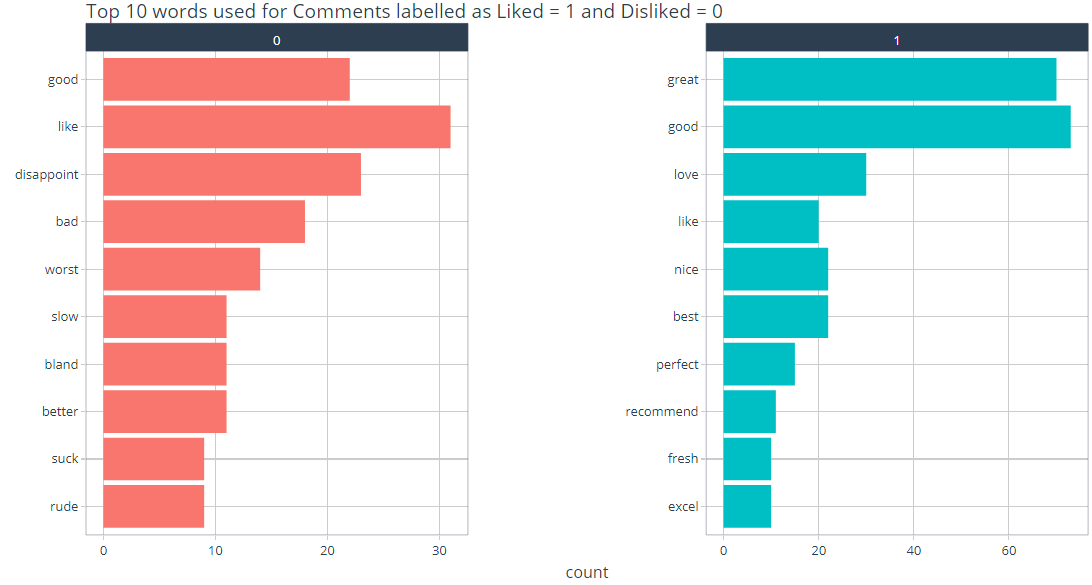
Approach

To perform the sentiment analysis on the restaurant review data, we first gave Customer-ID to each comment in the dataset. We do this because hereinafter I wish to break out each word from the comments into a separate row. Hence, the customer id label will help us understand which word belonged to which customer.

Thereon, we used ‘unnest\_tokens’ function to tokenize each word, we broke out each word and gave it its own row as shown in the figure below. Then words that carry positive and negative sentiments were labelled as such using the ‘get\_sentiments’ function from the ‘tidytext package’. Furthermore, we went on to remove stop words from the dataset. Stop words are words that are typically filtered out before a natural language is processed. These are the most common words in any language (articles, prepositions, pronouns, conjunctions, and so on), and they don't add any value to the text. In English stop words include "the," "a," "an," "so," and "what." Etc. If we don’t remove these words then the most common words in our analysis would be these words mentioned above. We used the ‘stopwords package’ in R to remove all the stop words. ‘Stopwords’ is a R Package that makes it simple to find stop words in over 50 languages.



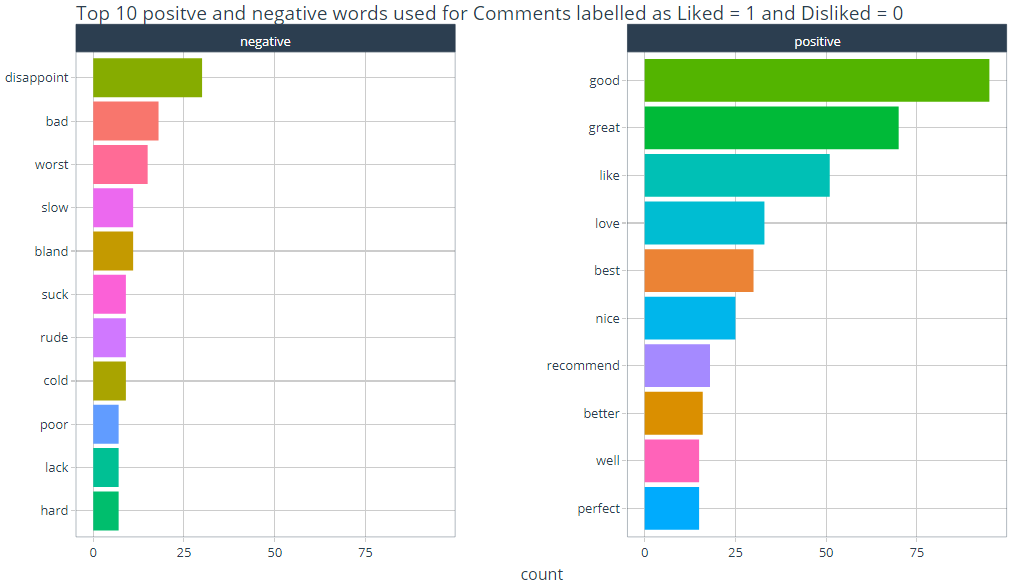
Sample of new Dataset after cleaning and wrangling



Top 10 words used in positive (1) and negative (0) comments

Assumptions: Upon plotting Top 10 words used in negative and positive comments we found that Positive words like: good, like, and best appear frequently in negative comments as well, indicating that they were likely used with other words to describe the service. As a result, rather we decide that we need to hunt for positive and negative words used in the comments rated as liked and disliked. Upon doing so we get the results mentioned below.

Results and recommendations



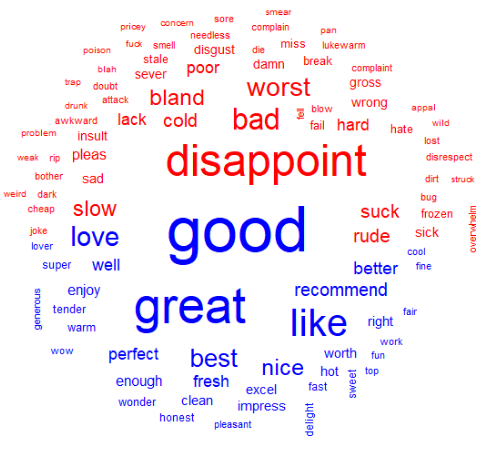
Top 10 Positive words used in positive (1) comments & Top 10 negative words used in Negative (0) comments

As clearly evident from both the insights above, positive words are more used in comments than negative words are used. The top positive word ‘good’ was used over 75 times, whereas the top negative word ‘disappoint’ was used only 25 times.

Positive words listed above only tells us what the customers positive sentiments are but it does not clearly indicate in what department is restaurant is doing well, meaning, the insights don’t tell us if the customer was happy with the food, service or ambience etc.

On the contrary the list of negative words do paint a rough picture for us. Negative words like ‘slow’ indicate that the customer might had experienced slow service. Moreover, negative words like ‘bland’ or ‘cold’ indicate that customer might have been served bland or cold food. In addition, negative word like ‘rude’ indicate that the customer might have encountered a rude attendant.

The Figure below displays the Negative and positive words in a word cloud format. A word cloud is a grouping of words that are displayed in various sizes. The larger and bolder the term, the more frequently it appears in a document and the more essential it is. In the figure below the Negative words are in red color and the positive words are in Blue.



A word cloud showing all the positive words (In blue) and negative words (In red)

Upon observing all the insights above: It is clear that there are more positive sentiments than negative sentiments. But the restaurant needs to not only serve fresh and hot food, they also need to step up their customer service quality. Excellent customer service can lead to customer satisfaction which will result in good word of mouth, which will subsequently increase the revenue as well. Hence, ensuring your customers contentment is critical to the restaurants continued success.

Method 2: Bigram Analysis

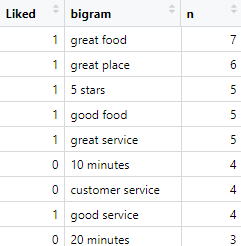
Though sentiment analysis only shows what the customers feel about your brand. It doesn’t give us a deep understanding about what the restaurant is doing right or what they are doing wrong and what needs to be improved.

We will now run a bigram analysis on the comment text to learn more about what the reviewers in this dataset are writing about. In text, bigrams are groups of words that are next to each other. The fact that bigrams often give additional information makes them more informative than simply calculating the frequency of individual words like what we had done in our first analysis. Example: We might expect positive reviews to regularly use the word "good," but this doesn't actually assist us to comprehend what is actually good. The bigram "good flavor" or "excellent presentation" reveals a great deal about what the reviewers like about the restaurant.

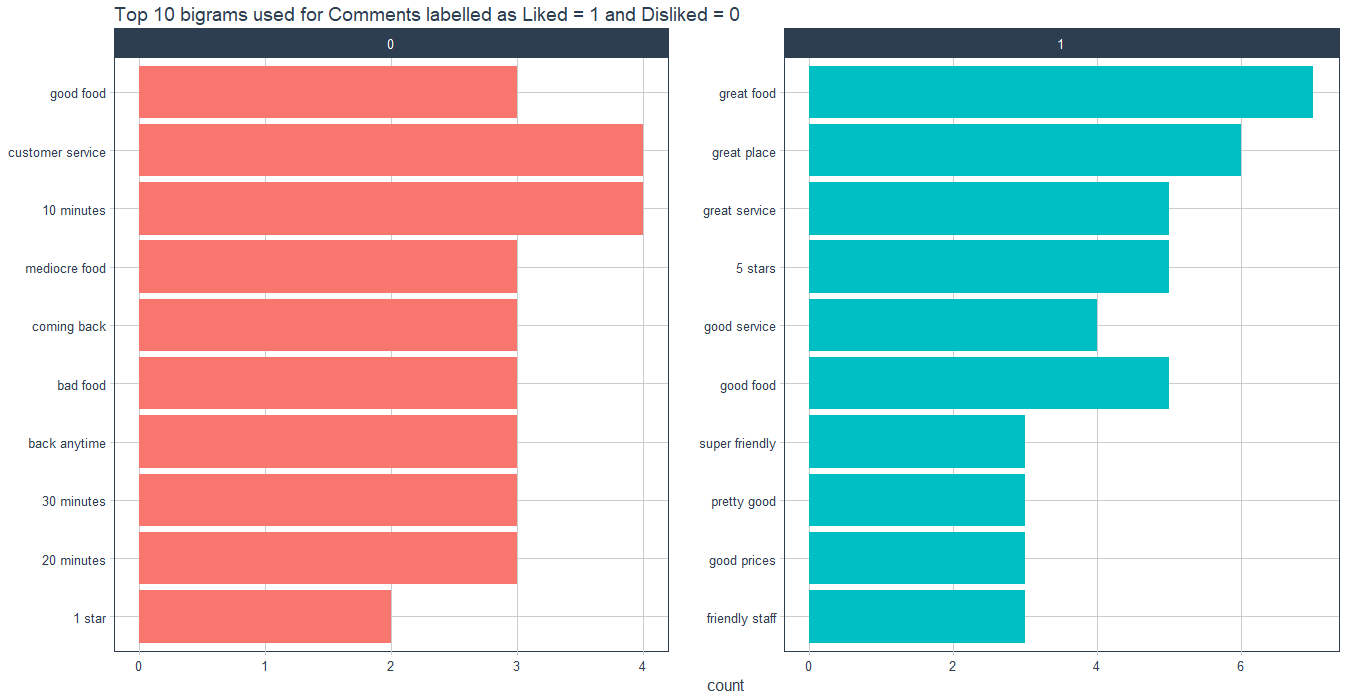
Approach

Now to perform bigram analysis on the dataset we use the ‘unnest\_tokens’ function again. We have been using the ‘unnest\_token’ function to tokenize each word in the previous analysis, which has been effective to extract the sentiment and sentiment frequency study we have done thus far. However, we may also utilize this function to perform bigram analysis as well. We now utilize this function to tokenize into n-grams, which are sequences of words. We achieve this by calling ‘unnest tokens()’' with the ‘token = "ngrams"’ option and setting ‘n’ to the number of words we want to capture in each n-gram. Here we set ‘n to 2’, we're looking at pairs of two words that are commonly referred to as "bigrams".

Another thing we do in order to better understand the underlying subject and the results of the evaluations is to eliminate words that don't assist us in communicating any information. Many words in the English language give a sentence structure but do not convey information. Articles like "the," "a," and "an" convey no context or information, for example. We're just as likely to encounter these words in favorable evaluations as we are in bad reviews, because they're so prevalent (and necessary for appropriate sentence building). We're going to eliminate them because they don't bring any value. These words are known as ‘stop words’ and finding and deleting them is one of the most crucial data preparation tasks for text analysis. After removing the stop words from each comment, we use ‘count()’ to examine the most common bigram & we get the dataset shown below.

  
Sample of bigram dataset after cleaning and wrangling

Results and recommendations



Top 10 Bigrams used in Positive and negative comments

As evident from the insights above, that the number of positive bigrams is more than negative bigrams. But the number of negative bigrams used are not that far behind and should not be taken lightly.

Positive Bigrams such as ‘great service’, ‘good service’, ‘super friendly’ and ‘friendly staff’ indicate towards an adequate customer service with which customer was very pleased.

Negative Bigrams such ‘coming back’ and ‘back anytime’ indicate that the customer was unhappy with his experience and he is unlikely to be coming back soon. Furthermore, negative bigrams such as ’10 minutes’, ‘ 20 minutes’ and ’30 Minutes’ indicate that the customer had to wait for that much time for his food to arrive and in addition, the negative bigram ‘customer service ’ suggest towards a poor customer service.



A word cloud showing all the positive and negative Bigrams used

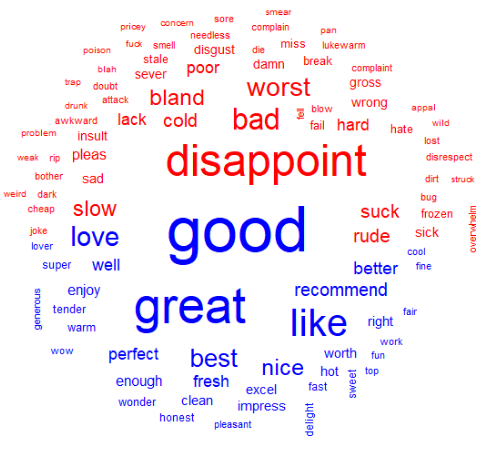
The negative bigrams and positive bigrams are displayed above in a word cloud fashion. A word cloud is a collection of words shown in different sizes. The more frequently the bigram appears in the document and the more important it is, hence the larger and bolder it is. Negative bigrams are shown in orange, while positive bigrams are in Turquoise.

Upon observing all the insights above: It is clear that there are more positive bigrams than negative bigrams. But the number of negative bigrams is not negligible, they are significantly high. From the above analysis we know that the customers liked some food items such ‘beef sandwich’’ and ‘vegas buffet’, whereas amongst the beverages like ‘beer’ seem to be very popular. However, customers did not like food items such as ‘lobster bisque’ and ‘crab legs’. Moreover, there are alarming number of negative bigrams such as ‘poor service’, ‘terrible service’, ’10 minutes’, ’20 minutes’ and many more, which indicate towards a concerning short coming in the restaurant’s customer service. Customers have often complained that they had to wait for a long time to be served their meals.

Customer service needs to be made number 1 priority. Customers despise waiting, therefore providing speedy service sets you apart from the competition. This gives your company a long-term, positive brand reputation in the market, which is crucial for dominating your industry. Customers always return to a company where they have a positive experience.

Comparing the results from both the analysis:

Showing the results from both the analysis



Simple Sentiment Analysis Bigram Analysis

Preferred analysis: One of the issues with the Sentiment analysis technique is that the context of a word might be just as important as its presence. Example even in a phrase like "I'm not happy and I don't like it!" the words "happy" and "like" will be considered as positive sentiments. We can evaluate how often sentiments-associated words are followed by "not" or other negating words by performing sentiment analysis on the bigram data. Hence bigram analysis gives us a clear picture, tells us exactly what is going wrong with the product and what is going right. Which helps the business to understand what department needs to improve.

From the first analysis we found that positive sentiments were used way more than negative sentiments, implying that most of the customers are very happy with the restaurant, only very few are unhappy with their experience. But bigram analysis painted a completely different picture, it showed us that the number of negative reviews is not insignificant and should not be overlooked.

The simple sentiment analysis simply reveals how customers feel about the restaurant. It doesn't provide us a clear picture of what the restaurant does well, what it does poorly, and what needs to be changed. The first analysis revealed that some reviewers were dissatisfied with the food, but the second analysis revealed particular foods were disliked, namely ‘lobster bisque’ and ‘crab legs’. Bigram analysis also showed us which food was liked by the customer.

According to the first analysis Customers are dissatisfied with the service, which has a broad definition. Customers may be dissatisfied with the quickness of service or the waiter's demeanor. Bigram's investigation revealed that the biggest concern was the slow customer service speed. Bigram analysis gives us a deep understanding of what the restaurant is doing right and what it isn’t.

Potential applications of text analysis methods in another context

In Online Business: Companies that sell their products online can assess the effectiveness of a product by utilizing sentiment analysis to see how the people react to their product. It will help to understand whether a feature of a product is as popular as it is supposed to be, or if customers have a different reaction to it entirely. Later on, marketing efforts can be tailored to target certain groups of people who have demonstrated an interest in that particular feature. It's very useful to know if a huge number of users have the same opinion about a particular product feature because If a lot of people hold a negative sentiment for a products feature then that feature can be discarded.

In Politics: We can use sentimental analysis to understand how the public reacts to news. The Obama administration used sentiment analysis to assess policy announcements in 2012. This type of analytics can also be used to investigate the number of negative or positive mentions of candidates in various news and media outlets. It can also help us to understand what the people want so that we can tailor policies that best cater to the people’s needs.

In Finance industry: Making solid investments in any business is a difficult affair. The stock market is notoriously volatile, with large swings in price occurring very frequently in a short period of time. Before investing, however, some of the factors that affect the stock market should be examined. Example: When picking between two vehicle businesses, for example, the sentiments expressed by the company about their most recent model can be weighed. This can show which company is performing better in the present market, allowing you to make an informed decision.

A brand can monitor their social media: Social media posts can contain some of the most candid thoughts on an enterprise. You can interact with customers in real time if you keep track of their comments. You'll be able to reply swiftly to negative or good feedback, and you'll have regular, reliable insights into your customers, which you can use to track your success from quarter to quarter.

Tracking the competition: For market and competition research, use sentiment analysis. Find out who among your competitors is getting positive press. Examine what exactly does the competitors customer like and incorporate some of the idea into your own enterprise.