

project_google_reviews_kk

The Ole Store Restaurant – Google Reviews Analysis by Kyrylo Koltunov

My mini project centers on analyzing Google reviews for one of my favorite places to dine in Northfield, MN — The Ole Store Restaurant. This place is known for its Norwegian and European cuisine and holds a special place in the community. On Google Maps, its page has gained significant attention, with more than 800 reviews accumulated over the years, in addition to a wide range of photos uploaded both by the restaurant itself and by loyal customers.

The primary focus of my project is to examine the text of these reviews in order to better understand the psychology of customer responses and gain insight into how people actually feel about the place. By exploring the language of the reviews, I aim to identify patterns in customer satisfaction and critique. To accomplish this, I will consult a variety of sources that support both the technical and theoretical aspects of my investigation.

The process of collecting data initially proved to be a challenge. However, after exploring options on Reddit and GitHub, I was able to locate useful tools. In particular, I made use of a desktop application called Google Maps Scraper by Leads Extractor, which allowed me to extract the necessary data from The Ole Store Restaurant's Google Business page. The extracted dataset includes review ratings (stars), written comments, dates of submission, owner responses, and links to the original posts. This dataset provides the foundation for my analysis.

For reference, the Google Business and Google Maps page for the restaurant can be found here:

<https://www.google.com/maps/place/ole+store+restaurant/data=!4m2!3m1!1s0x87f653d70b2e9f99:x2434e177640b7bfc?sa=X&ved=1t:242&ictx=111>

(<https://www.google.com/maps/place/ole+store+restaurant/data=!4m2!3m1!1s0x87f653d70b2e9f99:x2434e177640b7bfc?sa=X&ved=1t:242&ictx=111>)

The software I used to collect the review data is available at Google Maps Scraper by Leads Extractor here:<https://www.leads-extractor.com/google-maps-scraper/> (<https://www.leads-extractor.com/google-maps-scraper/>)

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
## filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union
```

```
library(stringr)
library(ggplot2)
library(readxl)

data_kk <- read_excel("data_kk.xlsx")

onetofive <- data_kk %>%
  mutate(
    rating = as.numeric(rating),
    review_text = as.character(review_text),
    char_len = str_length(coalesce(review_text, ""))
  ) %>%

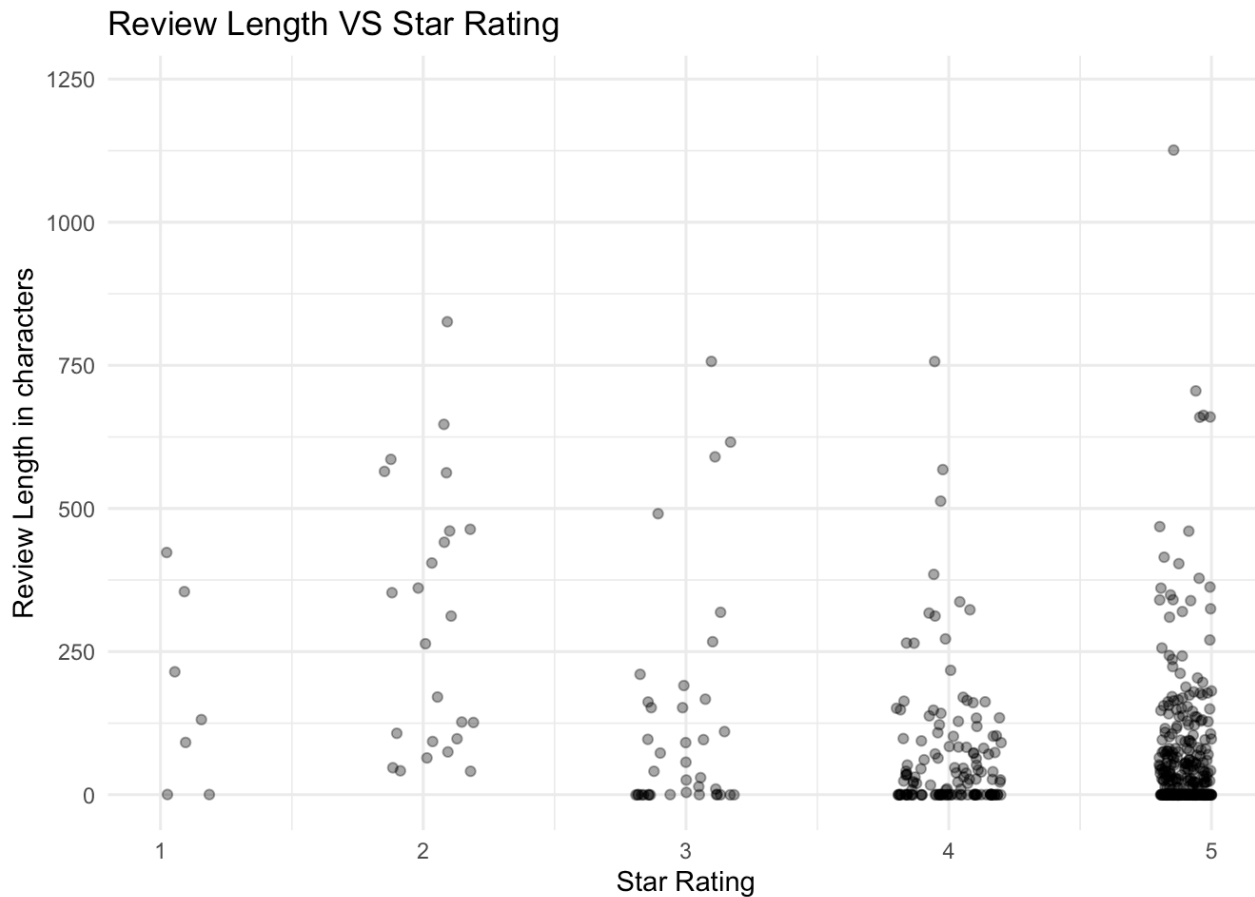
filter(rating >= 1 & rating <= 5)

table(onetofive$rating)
```

```
##
##   1   2   3   4   5
##  18  24  39 129 590
```

```
ggplot(onetofive, aes(x = rating, y = char_len)) +
  geom_point(alpha = 0.4, size = 1.5,
             position = position_jitter(width = 0.2)) +
  scale_x_continuous(breaks = 1:5, limits = c(1, 5)) +
  labs(
    title = "Review Length VS Star Rating",
    x = "Star Rating",
    y = "Review Length in characters"
  ) +
  theme_minimal()
```

```
## Warning: Removed 310 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



Alt-text: The scatterplot “Reviews Length vs. Star Rating” shows the relationship between the length of Google reviews and their star ratings. The x-axis represents the number of stars given (from 1 to 5), while the y-axis represents the review length in characters. Each dot corresponds to a single review across the 1–5 scale. The chart reveals that the higher the star rating, the longer the review tends to be. One-star reviews generally have the fewest characters, while five-star reviews are usually the longest. Overall, review length increases consistently across the categories, with five-star reviews making up the majority at the higher end.

The first scatterplot shows that the longest Google reviews were mostly written for 5-star ratings, while reviews with 4, 3, 2, or 1 stars tended to be shorter, and sometimes had no text at all. The smallest number of long reviews appeared in the 1-star category, where many comments were either extremely short or completely empty. A logical conclusion is that people who write long reviews on Google Maps usually focus on the positive aspects of a place, while 1-star reviews are often left without explanation. This suggests that, psychologically, people are more inclined to share details about what they truly enjoy, whereas negative reviews often provide little or no reasoning. Empty 1-star reviews may reflect small conflicts or just a quick bad impression, while people leaving 3-, 4-, or 5-star ratings tend to justify their opinions more, either in the hope of encouraging improvement or simply to share their insights. At the same time, it's important to remember that there are always exceptions. For example, an outlier in the 1-star category contained more than 600 characters, showing that when people find a place particularly upsetting, they can write very long and detailed negative reviews.

```
library(dplyr)
library(stringr)
library(ggplot2)
library(tidytext)
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

Library(RColorBrewer)

```

word_cloud_1 <- data_kk %>%
  mutate(rating = as.numeric(rating)) %>%
  filter(rating %in% c(1, 3, 5),
         !is.na(review_text))

tidy_reviews <- word_cloud_1 %>%
  unnest_tokens(word, review_text) %>%
  anti_join(stop_words, by = "word") %>%
  filter(str_detect(word, "^[a-z]+$"))

word_counts_1 <- tidy_reviews %>%
  count(rating, word, sort = TRUE)

make_wordcloud <- function(rating_value) {
  wc_data <- word_counts_1 %>% filter(rating == rating_value)
  wordcloud(words = wc_data$word,
            freq = wc_data$n,
            min.freq = 2,
            max.words = 100,
            colors = brewer.pal(8, "Dark2"),
            scale = c(4, 0.5),
            random.order = FALSE)
}

par(mfrow = c(1, 3))
make_wordcloud(5)
make_wordcloud(3)
make_wordcloud(1)

```

```

## Warning in wordcloud(words = wc_data$word, freq = wc_data$n, min.freq = 2, :
## months could not be fit on page. It will not be plotted.

```

```

## Warning in wordcloud(words = wc_data$word, freq = wc_data$n, min.freq = 2, :
## parmesan could not be fit on page. It will not be plotted.

```

```

## Warning in wordcloud(words = wc_data$word, freq = wc_data$n, min.freq = 2, :
## reviews could not be fit on page. It will not be plotted.

```

```

## Warning in wordcloud(words = wc_data$word, freq = wc_data$n, min.freq = 2, :
## server could not be fit on page. It will not be plotted.

```

```

## Warning in wordcloud(words = wc_data$word, freq = wc_data$n, min.freq = 2, :
## taste could not be fit on page. It will not be plotted.

```

```

## Warning in wordcloud(words = wc_data$word, freq = wc_data$n, min.freq = 2, :
## tasting could not be fit on page. It will not be plotted.

```

```
## Warning in wordcloud(words = wc_data$word, freq = wc_data$n, min.freq = 2, :
## times could not be fit on page. It will not be plotted.
```



```
## Warning in wordcloud(words = wc_data$word, freq = wc_data$n, min.freq = 2, :  
## water could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = wc_data$word, freq = wc_data$n, min.freq = 2, :  
## worse could not be fit on page. It will not be plotted.
```

The left-hand side word cloud presents the words most frequently used in 5-star reviews. One of the main ones is, of course, food, along with a list of adjectives such as amazing, friendly, delicious, fantastic, and so on. The overall spectrum of words is very positive, highlighting the restaurant's excellence in service, delicious dishes including desserts, and concluding that the staff is friendly. **The middle cloud** demonstrates all the 3-star reviews, concluding that some of the dishes were indeed overcooked, service was slow, and the use of adjectives is not as significant. The cloud mainly consists of nouns, pointing to the limitations of comments in this segment. **The third**, most negative cloud is focused on text analysis of 1-star reviews, concluding that the staff was not friendly, there was hair in the dishes, and breakfast was not good or was just average. There are also a lot of outliers like "girlfriend" or "left" because of extreme dissatisfaction in the 1-star reviews.

```
library(dplyr)
library(stringr)
library(tidytext)
library(ggplot2)

watermelon <- get_sentiments("bing")

reviews <- data_kk %>%
  mutate(text = ifelse(is.na(review_text) | review_text == "NA",
                        review_translated_text, review_text),
         text = str_to_lower(text),
         text = str_replace_all(text, "[[:punct:]]", ""),
         text = str_trim(text)) %>%
  unnest_tokens(word, text)

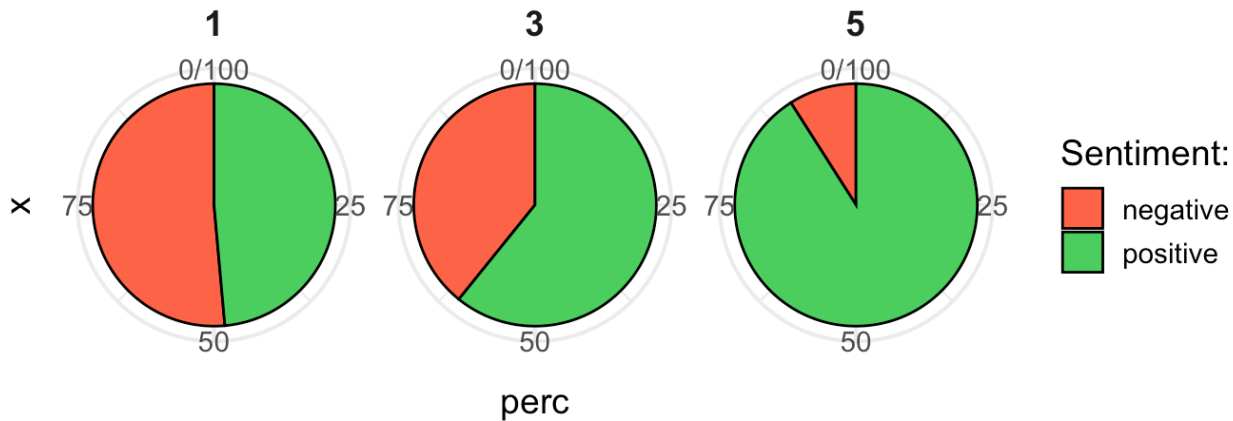
words1 <- reviews %>%
  inner_join(watermelon, by = "word")

sentiment_summary <- words1 %>%
  filter(rating %in% c(1, 3, 5)) %>%
  count(rating, sentiment) %>%
  group_by(rating) %>%
  mutate(perc = n / sum(n) * 100)

tone_colors <- c("positive" = "#50CF61", "negative" = "tomato")

ggplot(sentiment_summary, aes(x = "", y = perc, fill = sentiment)) +
  geom_col(width = 1, color = "black") +
  coord_polar(theta = "y") +
  facet_wrap(~rating) +
  scale_fill_manual(values = tone_colors) +
  labs(
    title = "Tone of Google Reviews",
    fill = "Sentiment:"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    strip.text = element_text(face = "bold", size = 14),
    plot.title = element_text(face = "bold", size = 16, hjust = 0.5)
  )
```

Tone of Google Reviews



The tone of Google reviews is indeed impressive. The tone in 5-star reviews is about 80% positive, while the 3-star reviews in the middle show around 60% positivity. The most exciting part is that even in negative 1-star comments, almost half of the reviews are written in a positive tone. This suggests that the place has a strong reputation among Google users who chose to leave reviews. While some people use Google Maps to either complain or express their fascination, there's also a group of people who don't fall into either category. With a sample size of 800, the results appear to be well balanced from a fairness perspective.

```
library(dplyr)
library(tidytext)
library(stringr)
library(igraph)
```

```
##
## Attaching package: 'igraph'
```

```
## The following objects are masked from 'package:dplyr':
##
##   as_data_frame, groups, union
```

```
## The following objects are masked from 'package:stats':
##
##   decompose, spectrum
```

```
## The following object is masked from 'package:base':  
##  
## union
```

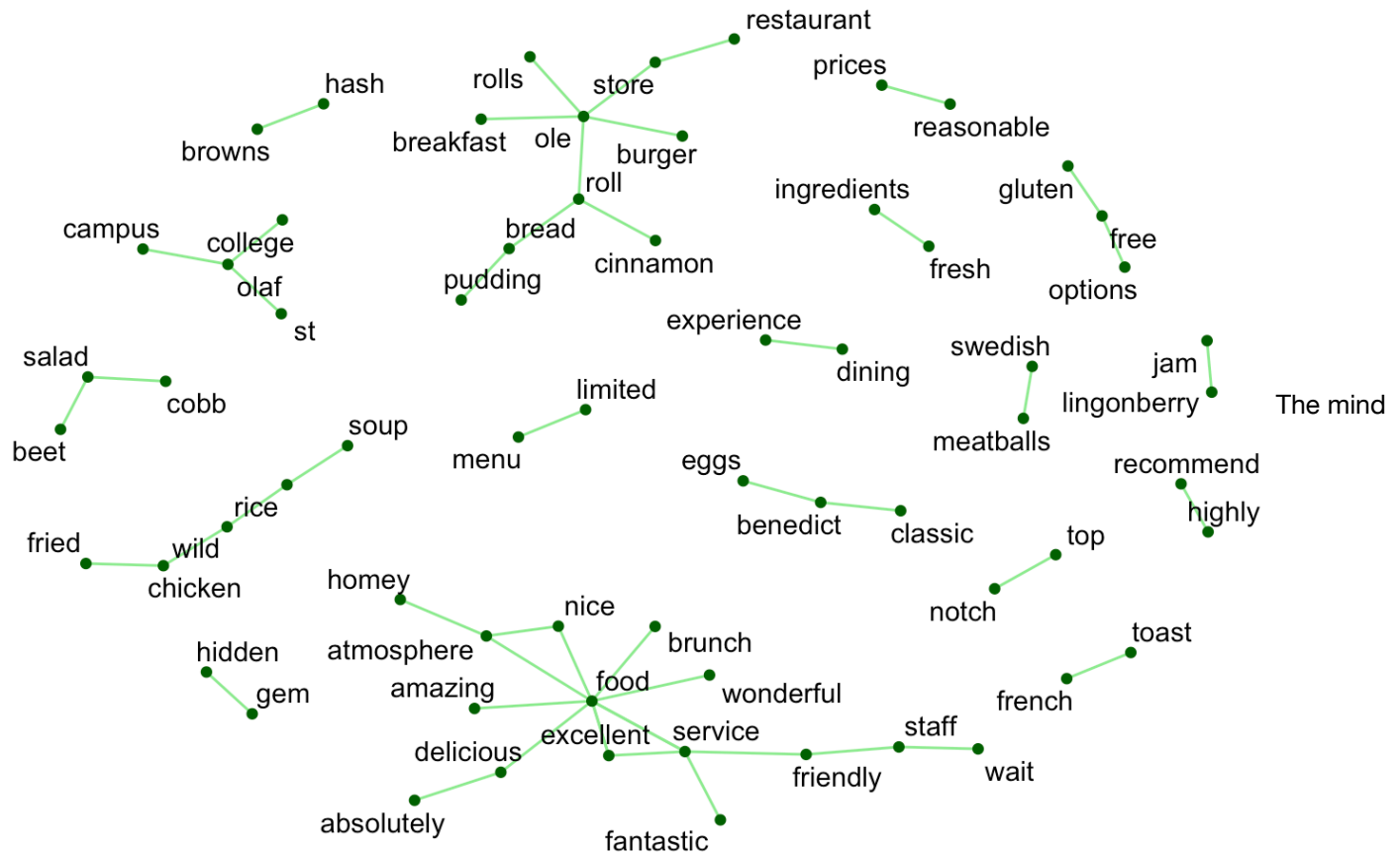
```
library(ggraph)  
library(tidyr)
```

```
##  
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:igraph':  
##  
## crossing
```

```
positive <- data_kk %>%  
  filter(rating %in% c(4, 5)) %>%  
  mutate(text = coalesce(review_text, review_translated_text),  
         text = str_to_lower(text),  
         text = str_replace_all(text, "[[:punct:]]", " "),  
         text = str_trim(text))  
  
bigrams <- positive %>%  
  unnest_tokens(bigram, text, token = "ngrams", n = 2) %>%  
  separate(bigram, c("word1", "word2"), sep = " ")  
  
bigrams_clean <- bigrams %>%  
  filter(!word1 %in% stop_words$word,  
         !word2 %in% stop_words$word,  
         !is.na(word1), !is.na(word2)) %>%  
  count(word1, word2, sort = TRUE)  
  
pos_graph <- graph_from_data_frame(bigrams_clean %>% filter(n > 2))  
  
ggraph(pos_graph, layout = "fr") +  
  geom_edge_link(color = "lightgreen") +  
  geom_node_point(color = "darkgreen") +  
  geom_node_text(aes(label = name), repel = TRUE) +  
  labs(title = "Mind Map of Positive Reviews") +  
  theme_void()
```


Mind Map of Positive Reviews



map extracted from Google reviews shows the possible correlations among the different subjects mentioned. One of the map lines focuses mainly on food, highlighting excellence, quality, great atmosphere, wonderful brunches, absolutely delicious and fantastic meals. The main strengths of the place are delicious, high-quality food, great service, fresh ingredients, a wide variety of breakfast options including the “ole burger,” reasonable prices, and a great overall selection of meals.

```
library(dplyr)
library(tidytext)
library(stringr)
library(widyr)
library(igraph)
library(gggraph)

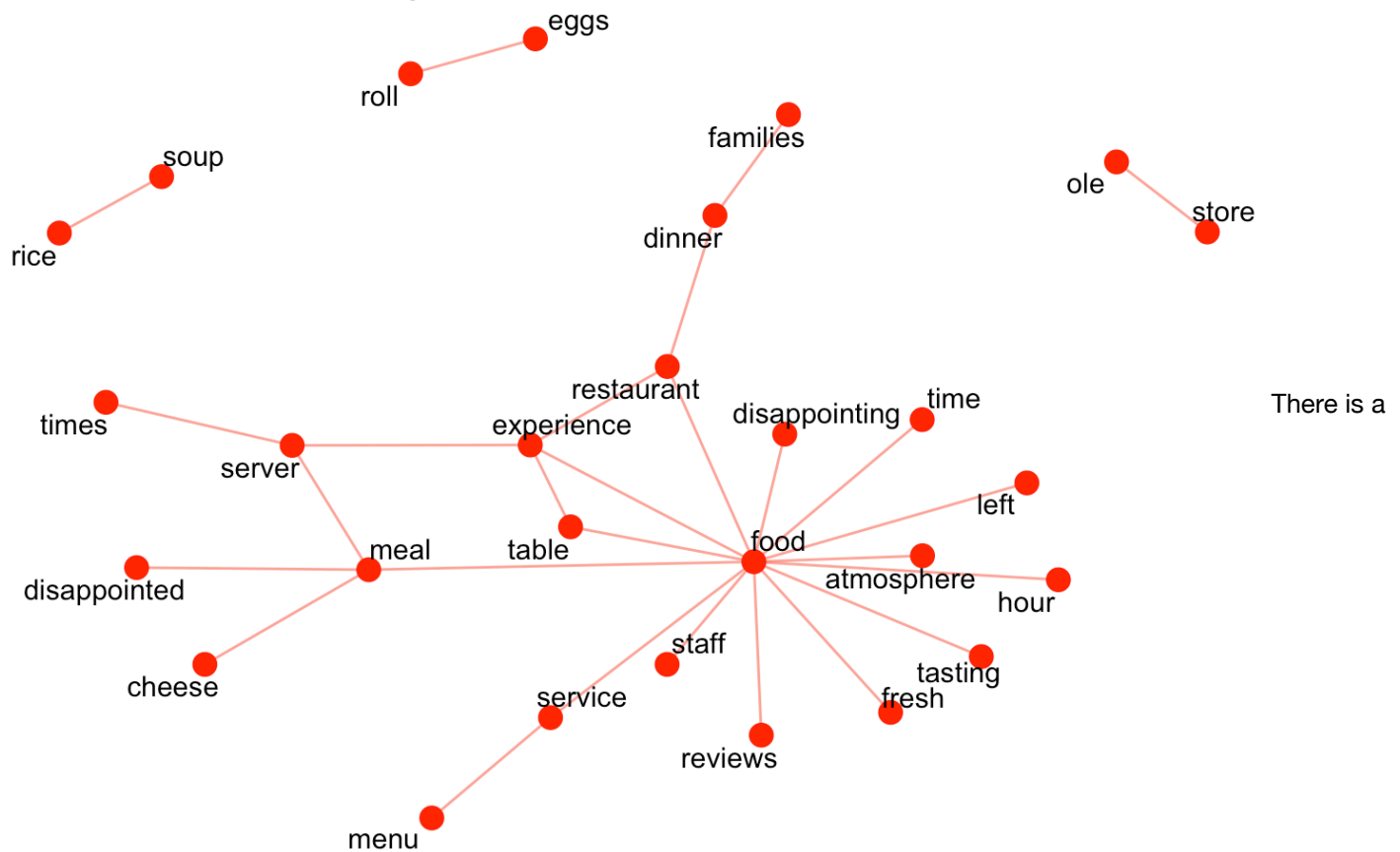
negative <- data_kk %>%
  filter(rating %in% c(1, 2)) %>%
  mutate(text = coalesce(review_text, review_translated_text),
         text = str_to_lower(text),
         text = str_replace_all(text, "[[:punct:]]", " "),
         text = str_trim(text),
         review_id = row_number()) %>%
  unnest_tokens(word, text) %>%
  filter(!word %in% stop_words$word, !is.na(word))

word_pairs <- negative %>%
  pairwise_count(word, review_id, sort = TRUE, upper = FALSE) %>%
  filter(n >= 3)

neg_graph <- graph_from_data_frame(word_pairs)

gggraph(neg_graph, layout = "fr") +
  geom_edge_link(color = "tomato", alpha = 0.6) +
  geom_node_point(color = "red", size = 4) +
  geom_node_text(aes(label = name), repel = TRUE, size = 4) +
  labs(title = "Connected Mind Map of Negative Reviews") +
  theme_void()
```

Connected Mind Map of Negative Reviews



food section present in the disadvantage word map as well, highlighting that the place does have some issues with cheese quality, disappointing food, rolled eggs, and rice soup. The table experience might also be concerning. Overall, the main weaknesses are related to the table experience, specific menu options, and some concerns about freshness.

In conclusion, the Ole Store Restaurant is truly a great place to enjoy a meal in Northfield, MN. The overall ratings are highly positive, emphasizing the freshness of dishes, friendly staff, and welcoming atmosphere. Like any other place, it also faces challenges with table assignments and certain menu items. The research also shows that people tend to write longer comments for positive reviews, while often leaving little or nothing for 1-star reviews, with some exceptions. My mini project not only covers the specific characteristics of the restaurant but also explores the psychological aspects of review writing.

My improvement areas could include using a larger number of visuals to highlight specific issues for deeper investigation. Additionally, I would like to explore how review scores depend on the time of year, month, or day. Unfortunately, due to Google review limitations, I was not able to export date information, but it would be useful to track positive and negative feedback across specific times and dates to identify clearer trends.