# Sentiment Analysis – Restaurant reviews

Emily Spano

2024-04-04

library(readxl)  
library(tm)  
library(SnowballC)  
library(plyr)  
library(PerformanceAnalytics)  
library(ggplot2)

The goal of this assignment is to investigate the impact of customer reviews on restaurant rating and this will be done through sentiment analysis

# Read the dataset  
RestaurantReviews <- read.csv("C:/Users/emily/Documents/Reviews27.csv")  
  
# Check the structure of the dataframe   
str(RestaurantReviews)

## 'data.frame': 2000 obs. of 9 variables:  
## $ X : int 345364 177958 71125 482105 86119 129735 30121 160363 296591 457831 ...  
## $ restaurant\_name: chr "TAO\_Uptown" "Fabio\_Cucina\_Italiana" "New\_York\_Luncheonette" "Bustan\_NYC" ...  
## $ rating\_review : int 5 5 3 4 5 5 4 4 5 5 ...  
## $ review\_id : chr "review\_242199577" "review\_732278415" "review\_334713875" "review\_591028992" ...  
## $ title\_review : chr "Book a tabel in advance" "Excellent Meal" "Solid choice for breakfast -- if staying nearby" "so close to Excellent" ...  
## $ review\_full : chr "Here you get the best of club and eating. They have the laud music going all the time, and turn it into a night"| \_\_truncated\_\_ "We recently dined at Fabrio on a Tuesday evening around 8:30 pm. We had just arrived in NYC and the restaurant "| \_\_truncated\_\_ "I used to rave about this place. We stumbled upon it years ago and were smitten. An authentic, no-fuss diner wi"| \_\_truncated\_\_ "It was hot on fathers day. We sat out back in the Garden. Sun had heated it up and the buildings blocked any ai"| \_\_truncated\_\_ ...  
## $ date : chr "2014-11-30" "2019-12-12" "2015-12-23" "2018-06-26" ...  
## $ city : chr "New\_York\_City\_New\_York" "New\_York\_City\_New\_York" "New\_York\_City\_New\_York" "New\_York\_City\_New\_York" ...  
## $ author\_id : chr "UID\_135027" "UID\_118714" "UID\_54555" "UID\_14544" ...

# overview of the first 10 reviews  
head(RestaurantReviews$review\_full, 10)

## [1] "Here you get the best of club and eating. They have the laud music going all the time, and turn it into a night club in th night. The food was amazing. A lot of flavors and a helpful staff. Would go back."   
## [2] "We recently dined at Fabrio on a Tuesday evening around 8:30 pm. We had just arrived in NYC and the restaurant was recommended by our hotel. This turned out to be an excellent choice. We were greeted and promptly seated. The wine and bread were excellent. They had a few risotto dishes but my wife wanted to know if they could make a risotto like she had in Florence which had bolognese sauce in it. The Executive Chef Chicco Asante said to give them a few minutes. The staff produced an excellent dish which replicated the risotto from Florence. My wife loved the dish. I had the ORECHIETTE ALLA BARESE (ear shaped pasta, spicy italian sausage, broccili rabe and oasted garlic and olive oil) which was fantastic. The service was excellent. This was a delicious meal and we really appreciated the chef accommadating our request. Well worth a visit."   
## [3] "I used to rave about this place. We stumbled upon it years ago and were smitten. An authentic, no-fuss diner with stained menus, decent prices and quick service. We loved it. But as the years have gone by, the menus have gotten glossier, the prices have gotten higher, and it just seems like any other NYC diner. The food is fine, the service is still fast. But it's no longer worth the 8-10 block walk for us to dine here. There are too many other similar restaurants. However, if you are staying nearby, it's worth a visit -- the prices, though higher, are still a lot cheaper than a hotel breakfast!"   
## [4] "It was hot on fathers day. We sat out back in the Garden. Sun had heated it up and the buildings blocked any air circulation. but enough with conditions. Food was excellent, and they did cater to my grand daughter's allergies. Service was very good."   
## [5] "This cafe is a gem in the middle of the Village. Very small room and bar when you walk in and we were led into another small room for our dinner. Not sure if there are additional rooms or not. My husband, daughters and I went on a Saturday evening and there was live music and a man rolling complimentary cigars in the corner. Our server, Tony, was excellent and gave us great recommendations. For our appetizers, we shared empanadas and the crabmeat salad special. The filling on all the empanadas (chicken, beef and cheese) was plentiful and tasty. The outside was nice and flaky. The crabmeat salad was pretty big and also extremely tasty. Lots of chunks of real crab in it. For dinner, I had the ceviche mixto and was also packed with lots of seafood. Our other dishes were lechon and shredded beef. Huge portions! We also ordered black beans, yucca and sweet plantains. Lots of leftovers to take home! The mojitos were awesome as was the dark and stormy and mango martini. For dessert we shared a chocolate molten cake that was warm and had bittersweet chocolate. Our tres leches cake was good as well. This food IS definitely the real deal Cuban. Would definitely go back and would like to try for brunch too."  
## [6] "Just walking by and decided to take a look. They had a table and a piano player. Who could say no to that? Prices seemed slightly higher, but, as it turned out, portions were quite generous. My partner and I had lunch for the next day. Really nice wine pour. Again, the per glass price seemed high, but probably better value than buying a bottle. I had Rigatoni Buttera. Yumm!"   
## [7] "If you want New York pizza eat here. It's real pizza. Just really, really good! The other dishes we got were tasty as well! Best of all it didn't break the bank. We had a nice table right in front of the dessert case. There are other tables with better ambience further back in the restaurant. It's beautiful back there with hanging lanterns and greenery. This is a more romantic setting, so since we were eating with our kids we opted for this table."   
## [8] "We have been looking for places around MIdtown where we can walk over and enjoy a high quality Sushi menu without paying the Nobu rates. If you are not after the \"Decor\" , Natsumi is offering first class Sushi with great selection and all served to perfection without having to pay the \"Branded\" rates. All dishes are just made to perfection , service is very attentive and you have an extensive Sake Menu. Highly recommended - as it is neat the Theater district , reservation is a must"   
## [9] "Went here as part of a work function. We were in the private events room away from the main floor, thus I don't have general knowledge of the overall crowd, noise, regular service, etc. I do however have first hand knowledge of the food, which was outstanding! Everything was delicious... Duck Meatballs, Casarece, Wood-Fired Pork Chop, Rustic Potatoes, Spumani, etc."   
## [10] "Some of the best Mexican food I have ever had and awesome drinks! I've tasted almost everything on the menu and have never been disappointed. The music is loud and fun, the lighting is dark and intimate."

# overview of the first 10 review titles  
head(RestaurantReviews$title\_review, 10)

## [1] "Book a tabel in advance"   
## [2] "Excellent Meal"   
## [3] "Solid choice for breakfast -- if staying nearby"  
## [4] "so close to Excellent"   
## [5] "Great Neighborhood Cuban Cafe!"   
## [6] "Lucked into this one!"   
## [7] "Affordable Delicious Food in NYC!"   
## [8] "Great quality Sushi near Midtown"   
## [9] "Excellent food!"   
## [10] "Delicious and Fun!"

# create a new variables with the amount of characters in each review  
RestaurantReviews$nchar\_rev <- nchar(RestaurantReviews$review\_full)  
  
# describe and plot the amount of characters per review   
summary(RestaurantReviews$nchar\_rev)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 36.0 205.0 338.5 448.8 555.0 3205.0

#there is a lot of variation in terms of review length, the shortest review is 36 characters while the longest is 3205  
  
hist(RestaurantReviews$nchar\_rev)

A graph of a restaurant review

Description automatically generated

# distribution is very skewed to the left and has a long tail meaning that there are more short reviews than long ones  
  
  
# create a new variables with the amount of characters in each review title  
RestaurantReviews$nchar\_tit <- nchar(RestaurantReviews$title\_review)  
  
# describe and plot the amount of characters per review   
summary(RestaurantReviews$nchar\_tit)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.00 15.00 23.00 25.81 33.00 115.00

# also titles have a minimun of 2 characters and a maximum of 115  
  
hist(RestaurantReviews$nchar\_tit) # as the reviews above, there are more reviews with fewer characters

A graph of a restaurant review

Description automatically generated

#find out if the length of the reviews provide insights for the rating  
correlations\_rev <- cor(RestaurantReviews$nchar\_rev, RestaurantReviews$rating\_review)   
# negative correlation (-0.11) meaning that the longer the review the lower the rating but only slightly  
  
  
#find out if the length of the review title provide insights for the rating  
correlations\_tit <- cor(RestaurantReviews$nchar\_tit, RestaurantReviews$rating\_review)   
# negative correlation (-0.045). Even though the correlation is negative, it is almost zero meaning no correlation

# Step 1: Pre-processing Review Data for Sentiment Analysis

# Pre-processing is important to make the text consistent since it can often contain noise and irrelevant information which might interfere with the analysis.  
  
# Convert review text to lowercase for both review\_full and review\_title  
RestaurantReviews$review\_full <- tolower(RestaurantReviews$review\_full)  
RestaurantReviews$title\_review <- tolower(RestaurantReviews$title\_review)  
  
head(RestaurantReviews$review\_full, 10)

## [1] "here you get the best of club and eating. they have the laud music going all the time, and turn it into a night club in th night. the food was amazing. a lot of flavors and a helpful staff. would go back."   
## [2] "we recently dined at fabrio on a tuesday evening around 8:30 pm. we had just arrived in nyc and the restaurant was recommended by our hotel. this turned out to be an excellent choice. we were greeted and promptly seated. the wine and bread were excellent. they had a few risotto dishes but my wife wanted to know if they could make a risotto like she had in florence which had bolognese sauce in it. the executive chef chicco asante said to give them a few minutes. the staff produced an excellent dish which replicated the risotto from florence. my wife loved the dish. i had the orechiette alla barese (ear shaped pasta, spicy italian sausage, broccili rabe and oasted garlic and olive oil) which was fantastic. the service was excellent. this was a delicious meal and we really appreciated the chef accommadating our request. well worth a visit."   
## [3] "i used to rave about this place. we stumbled upon it years ago and were smitten. an authentic, no-fuss diner with stained menus, decent prices and quick service. we loved it. but as the years have gone by, the menus have gotten glossier, the prices have gotten higher, and it just seems like any other nyc diner. the food is fine, the service is still fast. but it's no longer worth the 8-10 block walk for us to dine here. there are too many other similar restaurants. however, if you are staying nearby, it's worth a visit -- the prices, though higher, are still a lot cheaper than a hotel breakfast!"   
## [4] "it was hot on fathers day. we sat out back in the garden. sun had heated it up and the buildings blocked any air circulation. but enough with conditions. food was excellent, and they did cater to my grand daughter's allergies. service was very good."   
## [5] "this cafe is a gem in the middle of the village. very small room and bar when you walk in and we were led into another small room for our dinner. not sure if there are additional rooms or not. my husband, daughters and i went on a saturday evening and there was live music and a man rolling complimentary cigars in the corner. our server, tony, was excellent and gave us great recommendations. for our appetizers, we shared empanadas and the crabmeat salad special. the filling on all the empanadas (chicken, beef and cheese) was plentiful and tasty. the outside was nice and flaky. the crabmeat salad was pretty big and also extremely tasty. lots of chunks of real crab in it. for dinner, i had the ceviche mixto and was also packed with lots of seafood. our other dishes were lechon and shredded beef. huge portions! we also ordered black beans, yucca and sweet plantains. lots of leftovers to take home! the mojitos were awesome as was the dark and stormy and mango martini. for dessert we shared a chocolate molten cake that was warm and had bittersweet chocolate. our tres leches cake was good as well. this food is definitely the real deal cuban. would definitely go back and would like to try for brunch too."  
## [6] "just walking by and decided to take a look. they had a table and a piano player. who could say no to that? prices seemed slightly higher, but, as it turned out, portions were quite generous. my partner and i had lunch for the next day. really nice wine pour. again, the per glass price seemed high, but probably better value than buying a bottle. i had rigatoni buttera. yumm!"   
## [7] "if you want new york pizza eat here. it's real pizza. just really, really good! the other dishes we got were tasty as well! best of all it didn't break the bank. we had a nice table right in front of the dessert case. there are other tables with better ambience further back in the restaurant. it's beautiful back there with hanging lanterns and greenery. this is a more romantic setting, so since we were eating with our kids we opted for this table."   
## [8] "we have been looking for places around midtown where we can walk over and enjoy a high quality sushi menu without paying the nobu rates. if you are not after the \"decor\" , natsumi is offering first class sushi with great selection and all served to perfection without having to pay the \"branded\" rates. all dishes are just made to perfection , service is very attentive and you have an extensive sake menu. highly recommended - as it is neat the theater district , reservation is a must"   
## [9] "went here as part of a work function. we were in the private events room away from the main floor, thus i don't have general knowledge of the overall crowd, noise, regular service, etc. i do however have first hand knowledge of the food, which was outstanding! everything was delicious... duck meatballs, casarece, wood-fired pork chop, rustic potatoes, spumani, etc."   
## [10] "some of the best mexican food i have ever had and awesome drinks! i've tasted almost everything on the menu and have never been disappointed. the music is loud and fun, the lighting is dark and intimate."

Upper cases have been removed because they only represent unnecessary elements of the text which do not provide insights to the sentiment analysis. The sentence “Here you get the best of club and eating.” becomes “here you get the best of club and eating.” Removing upper cases is needed as in certain types of text analysis such as frequency based, the word “Here” and “here” might be treated as two different words.

# remove punctuation from text   
RestaurantReviews$review\_full <- removePunctuation(RestaurantReviews$review\_full)  
RestaurantReviews$title\_review <- removePunctuation(RestaurantReviews$title\_review)  
  
head(RestaurantReviews$review\_full, 10)

## [1] "here you get the best of club and eating they have the laud music going all the time and turn it into a night club in th night the food was amazing a lot of flavors and a helpful staff would go back"   
## [2] "we recently dined at fabrio on a tuesday evening around 830 pm we had just arrived in nyc and the restaurant was recommended by our hotel this turned out to be an excellent choice we were greeted and promptly seated the wine and bread were excellent they had a few risotto dishes but my wife wanted to know if they could make a risotto like she had in florence which had bolognese sauce in it the executive chef chicco asante said to give them a few minutes the staff produced an excellent dish which replicated the risotto from florence my wife loved the dish i had the orechiette alla barese ear shaped pasta spicy italian sausage broccili rabe and oasted garlic and olive oil which was fantastic the service was excellent this was a delicious meal and we really appreciated the chef accommadating our request well worth a visit"   
## [3] "i used to rave about this place we stumbled upon it years ago and were smitten an authentic nofuss diner with stained menus decent prices and quick service we loved it but as the years have gone by the menus have gotten glossier the prices have gotten higher and it just seems like any other nyc diner the food is fine the service is still fast but its no longer worth the 810 block walk for us to dine here there are too many other similar restaurants however if you are staying nearby its worth a visit the prices though higher are still a lot cheaper than a hotel breakfast"   
## [4] "it was hot on fathers day we sat out back in the garden sun had heated it up and the buildings blocked any air circulation but enough with conditions food was excellent and they did cater to my grand daughters allergies service was very good"   
## [5] "this cafe is a gem in the middle of the village very small room and bar when you walk in and we were led into another small room for our dinner not sure if there are additional rooms or not my husband daughters and i went on a saturday evening and there was live music and a man rolling complimentary cigars in the corner our server tony was excellent and gave us great recommendations for our appetizers we shared empanadas and the crabmeat salad special the filling on all the empanadas chicken beef and cheese was plentiful and tasty the outside was nice and flaky the crabmeat salad was pretty big and also extremely tasty lots of chunks of real crab in it for dinner i had the ceviche mixto and was also packed with lots of seafood our other dishes were lechon and shredded beef huge portions we also ordered black beans yucca and sweet plantains lots of leftovers to take home the mojitos were awesome as was the dark and stormy and mango martini for dessert we shared a chocolate molten cake that was warm and had bittersweet chocolate our tres leches cake was good as well this food is definitely the real deal cuban would definitely go back and would like to try for brunch too"  
## [6] "just walking by and decided to take a look they had a table and a piano player who could say no to that prices seemed slightly higher but as it turned out portions were quite generous my partner and i had lunch for the next day really nice wine pour again the per glass price seemed high but probably better value than buying a bottle i had rigatoni buttera yumm"   
## [7] "if you want new york pizza eat here its real pizza just really really good the other dishes we got were tasty as well best of all it didnt break the bank we had a nice table right in front of the dessert case there are other tables with better ambience further back in the restaurant its beautiful back there with hanging lanterns and greenery this is a more romantic setting so since we were eating with our kids we opted for this table"   
## [8] "we have been looking for places around midtown where we can walk over and enjoy a high quality sushi menu without paying the nobu rates if you are not after the decor natsumi is offering first class sushi with great selection and all served to perfection without having to pay the branded rates all dishes are just made to perfection service is very attentive and you have an extensive sake menu highly recommended as it is neat the theater district reservation is a must"   
## [9] "went here as part of a work function we were in the private events room away from the main floor thus i dont have general knowledge of the overall crowd noise regular service etc i do however have first hand knowledge of the food which was outstanding everything was delicious duck meatballs casarece woodfired pork chop rustic potatoes spumani etc"   
## [10] "some of the best mexican food i have ever had and awesome drinks ive tasted almost everything on the menu and have never been disappointed the music is loud and fun the lighting is dark and intimate"

Although punctuation is very important for humans to read the text correctly, this is different for computers thus punctuation only creates noise in the corpus as they are non-textual elements which have no meaningful information. After having removed punctuation the sentence “here you get the best of club and eating. they have the laud music going all the time, and turn it into a night club in the night. the food was amazing. a lot of flavors and a helpful staff. would go back.” becomes “here you get the best of club and eating they have the laud music going all the time and turn it into a night club in the night the food was amazing a lot of flavors and a helpful staff would go back”

# remove number from text   
RestaurantReviews$review\_full <- removeNumbers(RestaurantReviews$review\_full)  
RestaurantReviews$title\_review <- removeNumbers(RestaurantReviews$title\_review)  
  
head(RestaurantReviews$review\_full, 13)

## [1] "here you get the best of club and eating they have the laud music going all the time and turn it into a night club in th night the food was amazing a lot of flavors and a helpful staff would go back"   
## [2] "we recently dined at fabrio on a tuesday evening around pm we had just arrived in nyc and the restaurant was recommended by our hotel this turned out to be an excellent choice we were greeted and promptly seated the wine and bread were excellent they had a few risotto dishes but my wife wanted to know if they could make a risotto like she had in florence which had bolognese sauce in it the executive chef chicco asante said to give them a few minutes the staff produced an excellent dish which replicated the risotto from florence my wife loved the dish i had the orechiette alla barese ear shaped pasta spicy italian sausage broccili rabe and oasted garlic and olive oil which was fantastic the service was excellent this was a delicious meal and we really appreciated the chef accommadating our request well worth a visit"   
## [3] "i used to rave about this place we stumbled upon it years ago and were smitten an authentic nofuss diner with stained menus decent prices and quick service we loved it but as the years have gone by the menus have gotten glossier the prices have gotten higher and it just seems like any other nyc diner the food is fine the service is still fast but its no longer worth the block walk for us to dine here there are too many other similar restaurants however if you are staying nearby its worth a visit the prices though higher are still a lot cheaper than a hotel breakfast"   
## [4] "it was hot on fathers day we sat out back in the garden sun had heated it up and the buildings blocked any air circulation but enough with conditions food was excellent and they did cater to my grand daughters allergies service was very good"   
## [5] "this cafe is a gem in the middle of the village very small room and bar when you walk in and we were led into another small room for our dinner not sure if there are additional rooms or not my husband daughters and i went on a saturday evening and there was live music and a man rolling complimentary cigars in the corner our server tony was excellent and gave us great recommendations for our appetizers we shared empanadas and the crabmeat salad special the filling on all the empanadas chicken beef and cheese was plentiful and tasty the outside was nice and flaky the crabmeat salad was pretty big and also extremely tasty lots of chunks of real crab in it for dinner i had the ceviche mixto and was also packed with lots of seafood our other dishes were lechon and shredded beef huge portions we also ordered black beans yucca and sweet plantains lots of leftovers to take home the mojitos were awesome as was the dark and stormy and mango martini for dessert we shared a chocolate molten cake that was warm and had bittersweet chocolate our tres leches cake was good as well this food is definitely the real deal cuban would definitely go back and would like to try for brunch too"  
## [6] "just walking by and decided to take a look they had a table and a piano player who could say no to that prices seemed slightly higher but as it turned out portions were quite generous my partner and i had lunch for the next day really nice wine pour again the per glass price seemed high but probably better value than buying a bottle i had rigatoni buttera yumm"   
## [7] "if you want new york pizza eat here its real pizza just really really good the other dishes we got were tasty as well best of all it didnt break the bank we had a nice table right in front of the dessert case there are other tables with better ambience further back in the restaurant its beautiful back there with hanging lanterns and greenery this is a more romantic setting so since we were eating with our kids we opted for this table"   
## [8] "we have been looking for places around midtown where we can walk over and enjoy a high quality sushi menu without paying the nobu rates if you are not after the decor natsumi is offering first class sushi with great selection and all served to perfection without having to pay the branded rates all dishes are just made to perfection service is very attentive and you have an extensive sake menu highly recommended as it is neat the theater district reservation is a must"   
## [9] "went here as part of a work function we were in the private events room away from the main floor thus i dont have general knowledge of the overall crowd noise regular service etc i do however have first hand knowledge of the food which was outstanding everything was delicious duck meatballs casarece woodfired pork chop rustic potatoes spumani etc"   
## [10] "some of the best mexican food i have ever had and awesome drinks ive tasted almost everything on the menu and have never been disappointed the music is loud and fun the lighting is dark and intimate"   
## [11] "amazing japanese food beautifully servedthe dessert was incrediblegood service as well attentive service would recommend highly"   
## [12] "i rarely have the whole family in nyc but we trekked in to remember our days living in ny with three children ages we tried the food court at grand central but it was packed without thinking that i was not on a business trip we hoofed it across the street to the pershing square cafe with no reservation and a family of five with three kids they were gracious and welcoming and found us a wonderful location to enjoy a luncheon the waiter could not have been friendlier kind and generous to our children helpful with suggestions and gave us all the time we needed with a slower eater in the clan the food was great and the atmosphere delightful for the holidays thank you"   
## [13] "we had a few bumps with service poor guy spilt water on my husband etc otherwise the food was delicious and would definitely go back there again maybe even after restaurant week"

Similar to punctuation, also numbers do not provide much information thus they are not relevant for text analysis. For instance the sentence “i rarely have the whole family in nyc but we trekked in to remember our days living in ny With three children ages **6-13** we tried the food court at grand central but it was packed” becomes “i rarely have the whole family in nyc but we trekked in to remember our days living in ny with three children ages we tried the food court at grand central but it was packed” In this case the information about children’s age was unnecessary.

# remove unnecessary white space  
RestaurantReviews$review\_full <- stripWhitespace(RestaurantReviews$review\_full)  
RestaurantReviews$title\_review <- stripWhitespace(RestaurantReviews$title\_review)  
  
# Double white spaces only creates noise thus these were removed.  
  
# remove potentially problematic symbols   
RestaurantReviews$review\_full <- gsub("[[:punct:]]", " ", RestaurantReviews$review\_full)  
RestaurantReviews$title\_review <- gsub("[[:punct:]]", " ", RestaurantReviews$title\_review)  
  
head(RestaurantReviews$review\_full, 10)

## [1] "here you get the best of club and eating they have the laud music going all the time and turn it into a night club in th night the food was amazing a lot of flavors and a helpful staff would go back"   
## [2] "we recently dined at fabrio on a tuesday evening around pm we had just arrived in nyc and the restaurant was recommended by our hotel this turned out to be an excellent choice we were greeted and promptly seated the wine and bread were excellent they had a few risotto dishes but my wife wanted to know if they could make a risotto like she had in florence which had bolognese sauce in it the executive chef chicco asante said to give them a few minutes the staff produced an excellent dish which replicated the risotto from florence my wife loved the dish i had the orechiette alla barese ear shaped pasta spicy italian sausage broccili rabe and oasted garlic and olive oil which was fantastic the service was excellent this was a delicious meal and we really appreciated the chef accommadating our request well worth a visit"   
## [3] "i used to rave about this place we stumbled upon it years ago and were smitten an authentic nofuss diner with stained menus decent prices and quick service we loved it but as the years have gone by the menus have gotten glossier the prices have gotten higher and it just seems like any other nyc diner the food is fine the service is still fast but its no longer worth the block walk for us to dine here there are too many other similar restaurants however if you are staying nearby its worth a visit the prices though higher are still a lot cheaper than a hotel breakfast"   
## [4] "it was hot on fathers day we sat out back in the garden sun had heated it up and the buildings blocked any air circulation but enough with conditions food was excellent and they did cater to my grand daughters allergies service was very good"   
## [5] "this cafe is a gem in the middle of the village very small room and bar when you walk in and we were led into another small room for our dinner not sure if there are additional rooms or not my husband daughters and i went on a saturday evening and there was live music and a man rolling complimentary cigars in the corner our server tony was excellent and gave us great recommendations for our appetizers we shared empanadas and the crabmeat salad special the filling on all the empanadas chicken beef and cheese was plentiful and tasty the outside was nice and flaky the crabmeat salad was pretty big and also extremely tasty lots of chunks of real crab in it for dinner i had the ceviche mixto and was also packed with lots of seafood our other dishes were lechon and shredded beef huge portions we also ordered black beans yucca and sweet plantains lots of leftovers to take home the mojitos were awesome as was the dark and stormy and mango martini for dessert we shared a chocolate molten cake that was warm and had bittersweet chocolate our tres leches cake was good as well this food is definitely the real deal cuban would definitely go back and would like to try for brunch too"  
## [6] "just walking by and decided to take a look they had a table and a piano player who could say no to that prices seemed slightly higher but as it turned out portions were quite generous my partner and i had lunch for the next day really nice wine pour again the per glass price seemed high but probably better value than buying a bottle i had rigatoni buttera yumm"   
## [7] "if you want new york pizza eat here its real pizza just really really good the other dishes we got were tasty as well best of all it didnt break the bank we had a nice table right in front of the dessert case there are other tables with better ambience further back in the restaurant its beautiful back there with hanging lanterns and greenery this is a more romantic setting so since we were eating with our kids we opted for this table"   
## [8] "we have been looking for places around midtown where we can walk over and enjoy a high quality sushi menu without paying the nobu rates if you are not after the decor natsumi is offering first class sushi with great selection and all served to perfection without having to pay the branded rates all dishes are just made to perfection service is very attentive and you have an extensive sake menu highly recommended as it is neat the theater district reservation is a must"   
## [9] "went here as part of a work function we were in the private events room away from the main floor thus i dont have general knowledge of the overall crowd noise regular service etc i do however have first hand knowledge of the food which was outstanding everything was delicious duck meatballs casarece woodfired pork chop rustic potatoes spumani etc"   
## [10] "some of the best mexican food i have ever had and awesome drinks ive tasted almost everything on the menu and have never been disappointed the music is loud and fun the lighting is dark and intimate"

Symbols as non-textual data are also removed to provide more consistency. after this change, the sentence “if you want new york pizza eat here its real pizza just really really good**!**” Becomes ”if you want new york pizza eat here its real pizza just really really good”

# remove stopwords  
stop\_words <- stopwords("english")  
RestaurantReviews$review\_full <- removeWords(RestaurantReviews$review\_full, stopwords("english"))  
RestaurantReviews$title\_review <- removeWords(RestaurantReviews$title\_review, stopwords("english"))  
  
head(RestaurantReviews$review\_full, 10)

## [1] " get best club eating laud music going time turn night club th night food amazing lot flavors helpful staff go back"   
## [2] " recently dined fabrio tuesday evening around pm just arrived nyc restaurant recommended hotel turned excellent choice greeted promptly seated wine bread excellent risotto dishes wife wanted know make risotto like florence bolognese sauce executive chef chicco asante said give minutes staff produced excellent dish replicated risotto florence wife loved dish orechiette alla barese ear shaped pasta spicy italian sausage broccili rabe oasted garlic olive oil fantastic service excellent delicious meal really appreciated chef accommadating request well worth visit"   
## [3] " used rave place stumbled upon years ago smitten authentic nofuss diner stained menus decent prices quick service loved years gone menus gotten glossier prices gotten higher just seems like nyc diner food fine service still fast longer worth block walk us dine many similar restaurants however staying nearby worth visit prices though higher still lot cheaper hotel breakfast"   
## [4] " hot fathers day sat back garden sun heated buildings blocked air circulation enough conditions food excellent cater grand daughters allergies service good"   
## [5] " cafe gem middle village small room bar walk led another small room dinner sure additional rooms husband daughters went saturday evening live music man rolling complimentary cigars corner server tony excellent gave us great recommendations appetizers shared empanadas crabmeat salad special filling empanadas chicken beef cheese plentiful tasty outside nice flaky crabmeat salad pretty big also extremely tasty lots chunks real crab dinner ceviche mixto also packed lots seafood dishes lechon shredded beef huge portions also ordered black beans yucca sweet plantains lots leftovers take home mojitos awesome dark stormy mango martini dessert shared chocolate molten cake warm bittersweet chocolate tres leches cake good well food definitely real deal cuban definitely go back like try brunch "  
## [6] "just walking decided take look table piano player say prices seemed slightly higher turned portions quite generous partner lunch next day really nice wine pour per glass price seemed high probably better value buying bottle rigatoni buttera yumm"   
## [7] " want new york pizza eat real pizza just really really good dishes got tasty well best didnt break bank nice table right front dessert case tables better ambience back restaurant beautiful back hanging lanterns greenery romantic setting since eating kids opted table"   
## [8] " looking places around midtown can walk enjoy high quality sushi menu without paying nobu rates decor natsumi offering first class sushi great selection served perfection without pay branded rates dishes just made perfection service attentive extensive sake menu highly recommended neat theater district reservation must"   
## [9] "went part work function private events room away main floor thus dont general knowledge overall crowd noise regular service etc however first hand knowledge food outstanding everything delicious duck meatballs casarece woodfired pork chop rustic potatoes spumani etc"   
## [10] " best mexican food ever awesome drinks ive tasted almost everything menu never disappointed music loud fun lighting dark intimate"

Removing stopwords can be done as these words do not provide much information to the corpus, after this passage, the reviews are usually a lot shorter. For example “some of the best mexican food i have ever had and awesome drinks ive tasted almost everything on the menu and have never been disappointed the music is loud and fun the lighting is dark and intimate” ” becomes ” best mexican food ever awesome drinks ive tasted almost everything menu never disappointed music loud fun lighting dark intimate”.

Here words such as ‘some of’, ‘have had’, articles and connecting words have been removed.

# Stemming   
RestaurantReviews$review\_full <- stemDocument(RestaurantReviews$review\_full, language = "english")  
RestaurantReviews$title\_review <- stemDocument(RestaurantReviews$title\_review, language = "english")  
  
head(RestaurantReviews$review\_full, 10)

## [1] "get best club eat laud music go time turn night club th night food amaz lot flavor help staff go back"   
## [2] "recent dine fabrio tuesday even around pm just arriv nyc restaur recommend hotel turn excel choic greet prompt seat wine bread excel risotto dish wife want know make risotto like florenc bolognes sauc execut chef chicco asant said give minut staff produc excel dish replic risotto florenc wife love dish orechiett alla bares ear shape pasta spici italian sausag broccili rabe oast garlic oliv oil fantast servic excel delici meal realli appreci chef accommad request well worth visit"   
## [3] "use rave place stumbl upon year ago smitten authent nofuss diner stain menus decent price quick servic love year gone menus gotten glossier price gotten higher just seem like nyc diner food fine servic still fast longer worth block walk us dine mani similar restaur howev stay nearbi worth visit price though higher still lot cheaper hotel breakfast"   
## [4] "hot father day sat back garden sun heat build block air circul enough condit food excel cater grand daughter allergi servic good"   
## [5] "cafe gem middl villag small room bar walk led anoth small room dinner sure addit room husband daughter went saturday even live music man roll complimentari cigar corner server toni excel gave us great recommend appet share empanada crabmeat salad special fill empanada chicken beef chees plenti tasti outsid nice flaki crabmeat salad pretti big also extrem tasti lot chunk real crab dinner cevich mixto also pack lot seafood dish lechon shred beef huge portion also order black bean yucca sweet plantain lot leftov take home mojito awesom dark stormi mango martini dessert share chocol molten cake warm bittersweet chocol tres lech cake good well food definit real deal cuban definit go back like tri brunch"  
## [6] "just walk decid take look tabl piano player say price seem slight higher turn portion quit generous partner lunch next day realli nice wine pour per glass price seem high probabl better valu buy bottl rigatoni buttera yumm"   
## [7] "want new york pizza eat real pizza just realli realli good dish got tasti well best didnt break bank nice tabl right front dessert case tabl better ambienc back restaur beauti back hang lantern greeneri romant set sinc eat kid opt tabl"   
## [8] "look place around midtown can walk enjoy high qualiti sushi menu without pay nobu rate decor natsumi offer first class sushi great select serv perfect without pay brand rate dish just made perfect servic attent extens sake menu high recommend neat theater district reserv must"   
## [9] "went part work function privat event room away main floor thus dont general knowledg overal crowd nois regular servic etc howev first hand knowledg food outstand everyth delici duck meatbal casarec woodfir pork chop rustic potato spumani etc"   
## [10] "best mexican food ever awesom drink ive tast almost everyth menu never disappoint music loud fun light dark intim"

Stemming is used to normalize words and bring them back to their original form, this step is important to reduce variation of different words by removing verb tenses or plural forms.

# Step 2: extract sentiment from the pre-processed data using sentiment r

library(sentimentr)  
  
# Perform sentiment analysis on the review\_full column using sentimentr  
sentiment\_scores\_review\_full <- sentiment(RestaurantReviews$review\_full)

# Perform sentiment analysis on the title\_review column  
sentiment\_scores\_title\_review <- sentiment(RestaurantReviews$title\_review)

# Add sentiment scores to the dataframe  
RestaurantReviews$sentiment\_score\_review\_full <- sentiment\_scores\_review\_full$sentiment  
RestaurantReviews$sentiment\_score\_title\_review <- sentiment\_scores\_title\_review$sentiment  
  
summary(sentiment\_scores\_review\_full)

## element\_id sentence\_id word\_count sentiment   
## Min. : 1.0 Min. :1 Min. : 5.00 Min. :-0.8884   
## 1st Qu.: 500.8 1st Qu.:1 1st Qu.: 20.00 1st Qu.: 0.2988   
## Median :1000.5 Median :1 Median : 32.00 Median : 0.5481   
## Mean :1000.5 Mean :1 Mean : 42.24 Mean : 0.5534   
## 3rd Qu.:1500.2 3rd Qu.:1 3rd Qu.: 52.00 3rd Qu.: 0.8002   
## Max. :2000.0 Max. :1 Max. :287.00 Max. : 1.8575

# Using Sentimentr the reviews' sentiment ranges from -8 to 1.8  
  
summary(sentiment\_scores\_title\_review)

## element\_id sentence\_id word\_count sentiment   
## Min. : 1.0 Min. :1 Min. : 1.000 Min. :-1.0000   
## 1st Qu.: 500.8 1st Qu.:1 1st Qu.: 2.000 1st Qu.: 0.0000   
## Median :1000.5 Median :1 Median : 3.000 Median : 0.3000   
## Mean :1000.5 Mean :1 Mean : 3.189 Mean : 0.2983   
## 3rd Qu.:1500.2 3rd Qu.:1 3rd Qu.: 4.000 3rd Qu.: 0.5303   
## Max. :2000.0 Max. :1 Max. :13.000 Max. : 1.5588   
## NA's :4

# Using Sentimentr the reviews' sentiment ranges from -1 to 1.5

# Perform sentiment analysis on the review\_full column using qdap  
Sys.setenv(JAVA\_HOME = "C:/Windows/WinSxS/amd64\_microsoft-windows-f..uetype-javanesetext\_31bf3856ad364e35\_10.0.22621.1\_none\_5efc0b9e3e2fdb16/javatext.ttf")  
library(qdap)

polarity\_score\_review\_full <- polarity(RestaurantReviews$review\_full)  
polaritymatrix <- (counts(polarity\_score\_review\_full))  
RestaurantReviews$polarity\_score\_review\_full <-as.numeric(polaritymatrix[,3])  
summary(RestaurantReviews$polarity\_score\_review\_full)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.8485 0.2500 0.5571 0.5781 0.8707 2.8284

# using qdap the polarity scores for review full change a bit, while the minimum remains the same, now the maximum polarity score is 2.8

# performing sentiment analysis on the title\_review column using qdap  
polarity\_score\_title\_review <- polarity(RestaurantReviews$title\_review)  
polaritymatrix\_tit <- (counts(polarity\_score\_title\_review))  
RestaurantReviews$polarity\_score\_title\_review <-as.numeric(polaritymatrix\_tit[,3])  
summary(RestaurantReviews$polarity\_score\_title\_review)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## -1.9677 0.0000 0.4472 0.3469 0.7071 3.6770 4

# using qdap the polarity scores for title review are -1.9 as minim value and 3.6 as maximum value   
  
# checking correlation between polarities   
cor(RestaurantReviews$polarity\_score\_review\_full, RestaurantReviews$sentiment\_score\_review\_full)

## [1] 0.712072 # highly correlated

# Step 3: Investigate the relationship between sentiment and restaurant rating

#correlation between review sentiment and restaurant rating  
cor(RestaurantReviews$rating\_review, RestaurantReviews$sentiment\_score\_review\_full)

## [1] 0.3275627

#correlation between title sentiment and restaurant rating

cor(RestaurantReviews$rating\_review, RestaurantReviews$sentiment\_score\_title\_review) #

## [1] 0.2778621

# There is a slight correlation between sentiment score and restaurant rating as well as sentiment title and rating. This simple statistical test can be used to have an overview on the relationship between these two scores.

# describe and plot polarity score   
hist(RestaurantReviews$sentiment\_score\_review\_full) # the sentiment score is normally distributed

A graph of a restaurant review

Description automatically generated

boxplot(sentiment\_score\_review\_full ~ rating\_review, data = RestaurantReviews, main = "polaty score per review", xlab = "rating", ylab = "polarity score ")

A graph with a number of lines and a number of points

Description automatically generated with medium confidence

The box plot is a great graph for this kind of analysis as it allows to see both positive and negative polarities, the mean for each star rating as well as outliers. As already anticipated by the correlation analysis, the higher the rating, the higher the polarity score, however from the boxplot it is interesting to see that while reviews rating 4 have very high positive outliers, for the 5 star reviews, outliers are not only extremely positive but there are also some very negative ones.

RestaurantReviews$date <- as.Date(RestaurantReviews$date)  
  
RestaurantReviews$year <- format(RestaurantReviews$date, "%Y")  
  
  
ggplot(RestaurantReviews, aes(x = year, y = sentiment\_score\_review\_full, fill = factor(rating\_review))) +  
 geom\_bar(stat = "summary", fun = "mean", position = "dodge") +  
 labs(x = "Year", y = "Mean Sentiment", fill = "Rating") +  
 ggtitle("Mean Sentiment by Rating Grouped by Year") +  
 facet\_wrap(~ rating\_review, ncol = 2) +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))

A graph of different colored bars

Description automatically generated

The histogram is another graph that can be used to visualize the mean sentiment per year for different star scores. Although this type of graph does not allow to visualize positive and negative polarities, it gives an overview of the polarity of the different star scores throughout the years. What stands out from this graph is that positive polarity for 4 and 5 star scores tend to slightly diminish over time whereas it remains stable in 3 star scores. It tends to be only negative for 1 star scores and it seems to be slightly increasing for 2 star scores. Overall, it can be seen the mean increase in polarity as the star rating increases.

# Step 4: regression model to investigate the relationship sentiment score and restaurant rating

library(fixest)  
library(lubridate)

# Convert the date column to Date format  
RestaurantReviews$date <- as.Date(RestaurantReviews$date)  
  
# Extract day of the week, month, and year of the review  
RestaurantReviews$day <- wday(RestaurantReviews$date)  
RestaurantReviews$month <- month(RestaurantReviews$date)  
RestaurantReviews$year <- year(RestaurantReviews$date)  
  
# Estimate linear regression model  
model <- feols(rating\_review ~ sentiment\_score\_review\_full | restaurant\_name + day + month + year, data = RestaurantReviews)   
  
# In this regression model rating review is the dependent variable while sentiment\_score is the independent variables. This model takes into account the name of the restaurant for duplicates and the date in which the review was written.  
  
  
# View summary of the model  
summary(model)

## OLS estimation, Dep. Var.: rating\_review  
## Observations: 2,000   
## Fixed-effects: restaurant\_name: 762, day: 7, month: 12, year: 16  
## Standard-errors: Clustered (restaurant\_name)   
## Estimate Std. Error t value Pr(>|t|)   
## sentiment\_score\_review\_full 0.907465 0.075679 11.991 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## RMSE: 0.703313 Adj. R2: 0.135164  
## Within R2: 0.114796

# Step 5: extending the model to investigate the effect of the day of the week

# Fit a linear regression model to understand the impact of the day of the week   
model\_day <- feols(rating\_review ~ sentiment\_score\_review\_full + day | restaurant\_name + month + year, data = RestaurantReviews)  
  
summary(model\_day)

## OLS estimation, Dep. Var.: rating\_review  
## Observations: 2,000   
## Fixed-effects: restaurant\_name: 762, month: 12, year: 16  
## Standard-errors: Clustered (restaurant\_name)   
## Estimate Std. Error t value Pr(>|t|)   
## sentiment\_score\_review\_full 0.908913 0.075811 11.989157 < 2.2e-16 \*\*\*  
## day 0.002295 0.013468 0.170394 0.86475   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## RMSE: 0.705451 Adj. R2: 0.133492  
## Within R2: 0.114896

# Fit a linear regression model for all variables   
model\_all\_variables <- lm(rating\_review ~ sentiment\_score\_review\_full \* wday(date) + month(date) + year(date) + nchar\_tit + nchar\_rev, data = RestaurantReviews)  
  
# View the summary of the model  
summary(model\_all\_variables)

##   
## Call:  
## lm(formula = rating\_review ~ sentiment\_score\_review\_full \* wday(date) +   
## month(date) + year(date) + nchar\_tit + nchar\_rev, data = RestaurantReviews)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.9248 -0.4962 0.1609 0.6843 1.7887   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -3.283e+01 1.688e+01 -1.945 0.0519  
## sentiment\_score\_review\_full 9.004e-01 1.242e-01 7.253 5.82e-13  
## wday(date) -7.882e-03 1.915e-02 -0.412 0.6807  
## month(date) -1.330e-02 5.848e-03 -2.274 0.0231  
## year(date) 1.830e-02 8.372e-03 2.186 0.0289  
## nchar\_tit -1.639e-03 1.450e-03 -1.130 0.2585  
## nchar\_rev -4.001e-04 5.622e-05 -7.117 1.53e-12  
## sentiment\_score\_review\_full:wday(date) 1.196e-02 2.914e-02 0.410 0.6816  
##   
## (Intercept) .   
## sentiment\_score\_review\_full \*\*\*  
## wday(date)   
## month(date) \*   
## year(date) \*   
## nchar\_tit   
## nchar\_rev \*\*\*  
## sentiment\_score\_review\_full:wday(date)   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.906 on 1992 degrees of freedom  
## Multiple R-squared: 0.1384, Adjusted R-squared: 0.1354   
## F-statistic: 45.72 on 7 and 1992 DF, p-value: < 2.2e-16

# Step 6: describing your findings

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t-value | p-value |
| Sentiment score | 0.907465 | 0.075679 | 11.991 | < 2.2e-16 \*\*\* |

The results from both models suggest that sentiment is a significant predictor of rating score. The first model, which takes into consideration confounding factors such as restaurant name for duplicates and the date in which the review was written, shows the significant impact of sentiment score on the reviews (0.90, p <0.05).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t-value | p-value |
| Sentiment score | 0.908913 | 0.075811 | 11.989157 | < 2.2e-16 \*\*\* |
| Day of the week | 0.002295 | 0.013468 | 0.170394 | 0.86475 |

In the second model, in which ratings remain the dependent variable but day of the week is added together with sentiment as independent variable, while sentiment keeps having a significant effect (0.90, p <0.05), the day of the week does not have a significant effect on the restaurant rating score (0.002, p > 0.05).

This can not only be seen with the p value which is only significant for the sentiment score whereas for the day of the week is not, but also from the coefficients. As a matter of fact, the coefficient for sentiment is .90 which means that on average, for each one-unit increase in sentiment score, the rating increases by 0.907 units. On the other hand, the coefficient of the day of the week is 0.002295 which suggest almost no effect at all meaning that the review ratings are not influenced by the day of the week when the review was written. Moreover the t-value is moderately high suggesting that the null hypothesis can be rejected, in other words sentiment has a strong effect on ratings, the opposite can be said for the day of the week which in turn has a very small t -value.

When taking into account all the variables, it can be seen that the number of characters in the review is highly significant in explaining the variation of ratings but also has a very small effect overall.

## This code has been partly generated with chatGPT