

Identify Restaurant Features Affecting Consumer Choice of Different Cuisines in the US

Final Report

Ashwathi Nair, Jingshu Zhang, Shanrong Zhou, Srushti Divyesh Sanghavi

Research Problem

America, the land of everyone's dreams, as we all know, is a place filled with passionate people, who work day in and day out to build their lives and make a mark. It's a highly competitive world we all live in today and working hard sure comes with a lot of expectations. One such would be to be able to have a light evening where one could relax and spend time with oneself, one's friends and families; maybe have a fun-filled weekend, sometimes void of responsibilities, to just have a wonderful time. So at times, we don't want to cook or we are not able to. Although we do want to be able to satiate our cravings for the kind of cuisine we want, maybe the cuisine we loved during our travels over the world.

The restaurant industry in America was a wonderfully expanding, \$12 million industry by 2019, facing changing dynamics after covid hit, although it started to revive back in 2021. Also, in due course of time, food deliveries saw a steady rise, and combined, these 2 sections of the food industry have been forecasted to continue to grow by leaps and bounds in the future.^{1 2}

As the US continues to have a cosmopolitan and diverse culture and set of people residing for education, work and travel, it becomes increasingly important for restaurants to be able to serve to them what they expect when they dine-in or take-out.

In today's digital era, online review has become a new form of social communication. Customers share their personal experiences and tell their perceptions towards restaurants on online platforms. The robust customer reviews not only help people to pick the preferred restaurants, but also deliver direct feedback for restaurants' strategic planning. Yelp is one of the largest platforms that allow customers in the US to post their reviews and give a rating to the restaurants. According to Statista, the cumulative number of reviews submitted to Yelp has

¹ Ariella, S. (2023, March 1). *45+ must-know us restaurant industry statistics [2023]: How many restaurants are in the US*. Zippia 45+ MustKnow US Restaurant Industry Statistics 2023 How Many Restaurants Are In The US Comments. Retrieved March 1, 2023, from <https://www.zippia.com/advice/restaurant-industry-statistics/#:~:text=The%20US%20restaurant%20industry%20made,down%20%24132.8%20billion%20from%202019>

² *Meal delivery - united states: Statista market forecast*. Statista. (n.d.). Retrieved March 1, 2023, from <https://www.statista.com/outlook/dmo/online-food-delivery/meal-delivery/united-states#revenue>

reached 244 millions in 2021, which is over double of the amount (121 millions) in 2016.³ Yelp data's growth potential shows the value for analysis on customer reviews.

Our research study wishes to focus on this area of concern for the restaurant industry. We want to study customer reviews to identify features affecting consumers' choice of different cuisines, in order to be able to recommend improvements to restaurants in America that specialize in international cuisines. We want to access the customer perceptions of restaurants that offering foreign cuisines in US, and identify the restaurant features that the customers care most about. Furthermore, we also want to deepen our understanding of the key aspects that drive customer perceptions to the most popular restaurants. This could help the restaurants to identify growth areas and attract more customers. Scaling up, America can start to feel more like home to immigrants within.

For doing so, we are analyzing the Yelp dataset with customer reviews for various foreign cuisines offered by restaurants in the US, using machine learning algorithms in text mining, including sentiment analysis, word cloud and classification trees.

Literature Review

Based on our research, people in America love Chinese, Thai, Indian, Korean, French and Japanese cuisines the most, and that is why we would like to focus our study mainly on these cuisines.

A study done on what customers expect from Chinese restaurants provides a lot of insights. It appears that there are more factors than taste, quality and price that matter to customers. There was a survey-based research and the results showed that variety, food safety, food temperature, presentation, service, aroma, cleanliness, decor, treatment and authenticity also matter immensely. These affect the rating that they provide to these restaurants.⁴

Thai restaurants in the US have also been rated as part of other studies for authenticity. One such study did so by sending out google forms in certain Metropolitan cities in the US. This revealed important insights like the healthiness of the food being served was extremely important. This study can be built on further if restaurant reviews were analyzed.⁵

A study was also conducted to understand the perceptions of customers eating Indian food and comparing it with the standards of the country. This was an online questionnaire with mostly

³ Published by S. Dixon, & 11, A. (2022, August 11). *Yelp: Cumulative number of reviews 2021*. Statista. Retrieved March 1, 2023, from <https://www.statista.com/statistics/278032/cumulative-number-of-reviews-submitted-to-yelp/>

⁴ Liu, Y., & Jang, S. C. (S. (2009). Perceptions of Chinese restaurants in the U.S.: What affects customer satisfaction and behavioral intentions? *International Journal of Hospitality Management*, 28(3), 338–348. <https://doi.org/10.1016/j.ijhm.2008.10.008>

⁵ AUAPINYAKUL, W., SIRIPONGDEE, S., & PIMDEE, P. (2022). *Antecedents of Customer Loyalty and Perceived Service Quality: A SEM Analysis of Thai Restaurant Brands*. *The Journal of Asian Finance, Economics and Business*, 9(7), 173–183. <https://doi.org/10.13106/JAFEB.2022.VOL9.NO7.0173>

ratings based on a few open ended questions. The study brought out interesting insights that that customers also expected to have better hygiene, trained staff and more value for money.⁶

Study and analysis of Korean restaurants in America was done for similar reasons and this highlighted more important criteria as well. It appears that authenticity, dining environment, service quality and music. This was also done by taking an online survey.⁷

As the above mentioned cuisines are very different from each other, we choose these to better understand consumer perceptions. Our study aims at focusing on whether these perceptions are highlighted in restaurant reviews, because customer reviews based on personal experiences would better portray the voice of customers rather than the surveys.

Research Questions and Analytical Techniques

This study aims to answer the following research questions:

1. What are the restaurant features that customers perceive as the most important for foreign cuisines in the US?
2. For the most popular restaurant serving foreign cuisines in the US, what are its key aspects that drive the overall customer perception?

We firstly understood the overall ratings and customer reviews of the restaurants via **exploratory data analysis**. We focused on exploring the two aspects of measuring the “best” restaurants, frequency of customer reviews and customer ratings.

We then grouped the restaurants based on the cuisine being served, conducted **sentiment analysis** and drew **word clouds** to assess the customer perceptions of these restaurants. From these analyses, we identified the features that customers care most about.

By comparing the sentiment score of various cuisine types of restaurants, we picked the most likable cuisine and the most popular restaurant serving this cuisine to build **classification trees**. The classification trees acted as a **sentiment classifier** to help us reveal the key aspects of this restaurant that drive overall customer perception.

Data Description and Preparation

Before choosing a restaurant for dining in or takeout, most individuals tend to refer to reviews. Yelp, being one of the biggest online platforms for finding restaurants based on reviews, is used for our study. The Yelp Dataset, which comprises 5 distinct files containing information on businesses, reviews for each of these businesses, users who wrote the reviews, as well as

⁶ Josiam, B. M., & Monteiro, P. A. (2004). Tandoori tastes: Perceptions of Indian restaurants in America. *International Journal of Contemporary Hospitality Management*, 16(1), 18–26.
<https://doi.org/10.1108/09596110410516525>

⁷ Jang, S. C. (S.), Ha, J., & Park, K. (2012). Effects of ethnic authenticity: Investigating Korean restaurant customers in the U.S. *International Journal of Hospitality Management*, 31(3), 990–1003.
<https://doi.org/10.1016/j.ijhm.2011.12.003>

check-in and tip data, are used for our analysis. For our study, we were mainly focusing on the review dataset and the business dataset.

The dataset description tables below show what the 5 datasets are describing. We illustrate the detailed explanation of data fields in business and review datasets in Appendix 1. The “business_id” links these two tables together.

<u>business dataset</u> : contains business data including location data, attributes, and categories.
<u>review dataset</u> : contains full review text data including the business_id that the review is written for.
<u>user dataset</u> : contains the user's friend mapping and all the metadata associated with the user.
<u>checkin dataset</u> : contains checkins' date and time on a business.
<u>tip dataset</u> : Tips are written by a user on a business, which tend to convey quick suggestions and are shorter than reviews.

Since the review dataset contains 7 million records, we used Python for the initial data processing due to R's limitations in speed. We began by analyzing the business dataset and filtered out restaurants serving **French, Japanese, Chinese, Korean, Indian, and Thai** cuisines. We then checked for missing values for further analysis and generated a new csv file with relevant cuisine data using R. Next, we read and examined the review dataset and merged it with the cuisine data in Python. We removed any non-essential columns and checked for any missing or null values within our final dataset.

The cleaned dataset contains these following data fields. We used this dataset for our data analysis.

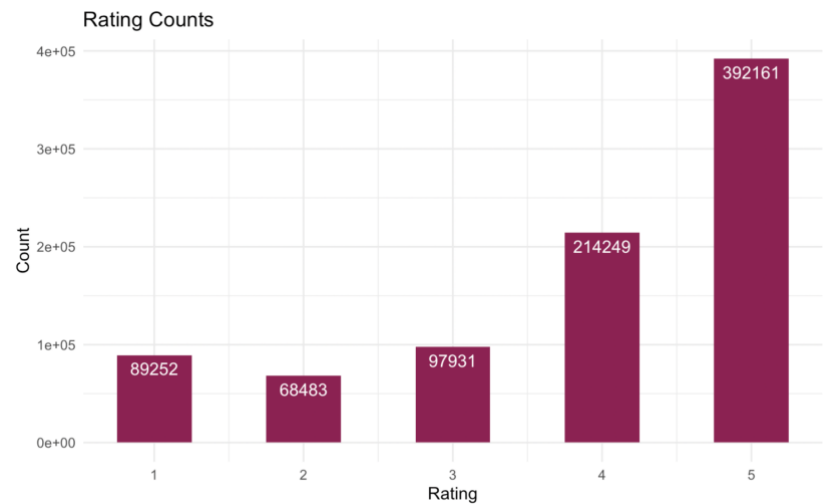
Field Name	Data Type	Description
business_id	string	unique string business id
name	string	the business's name
city	string	the city
state	string	state code, if applicable
restaurant_rating	float	business overall star rating, rounded to half-stars
categories	array	an array of strings of business categories
review_id	string	unique review id
user_id	string	unique user id
review_rating	integer	review's rating given by user
text	string	the review itself
date	string	review's date formatted YYYY-MM-DD

Data Analysis and Interpretation

We started with an exploratory data analysis to look at the rating and number of reviews across various foreign cuisines. Next, for research question 1, sentiment analysis and word clouds were leveraged to identify the restaurant features that customers care most about. Furthermore, for research question 2, we used classification trees to build a sentiment classifier to reveal the key aspects of the most popular restaurant that drive overall customer perception.

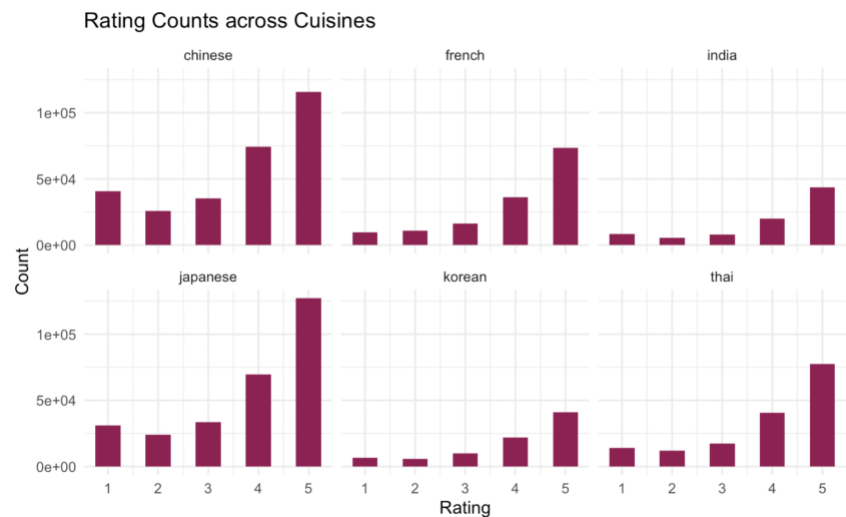
Exploratory Data Analysis

1. Overall Rating Distribution: Rating vs. Number of Reviews



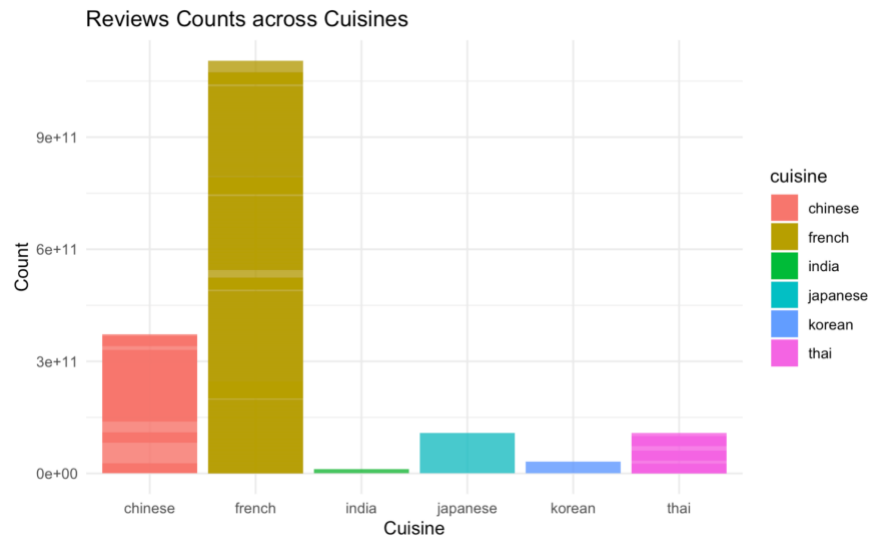
We firstly looked at the distribution of overall review ratings. We noticed that reviews with five stars are the most common. Rating of one is more prevalent than those with two. Customers are more likely to go to the trouble of submitting a review if they were really pleased or extremely disappointed.

2. Rating Distribution across Cuisines



We then looked at the rating distribution across cuisines. Except for French cuisine, all the other cuisines have a similar “U shape” distribution, which is similar to the overall rating distribution. However, French cuisine’s rating has a steady increase on the number of reviews. It indicates that the extremely negative reviews of French cuisine are relatively less than the other cuisines.

3. Number of Reviews across Cuisines



Among the number of reviews across the six foreign cuisines in US, French cuisine receives the most reviews from customers. It shows the highest popularity of French cuisine in the US.

Sentiment Analysis and Word Cloud

In the review text for each restaurant, there are a number of word lexicons that can be used to classify words as being positive or negative. We used afinn lexicon to categorize words as being positive and negative, and detailly distinct them based on how positive they are. Compared to the other sentiment lexicon like Bing, the afinn lexicon scores each word based on the extent to which it is positive or negative. We calculated the afinn scores for every cuisine. Meanwhile, we also compared average review ratings across various cuisines.

cuisines <chr>	average_afinn_scores <dbl>	average_ratings <dbl>
French	1.589387	4.038566
Thai	1.561937	3.963770
Indian	1.527623	3.990886
Korean	1.524306	3.993618
Japanese	1.440750	3.831592
Chinese	1.213658	3.679918

We can clearly see that French cuisine has the highest average afinn score and review rating, indicating that it is the restaurants serving French cuisine that receive the most positive review comments and ratings, compared to the other cuisines.

Next, we used Bing Lexicon to classify the sentiment of words in the review text, and created comparison word clouds for all the cuisines. This is an overall word cloud for all the restaurants serving the six foreign cuisines in the US.



The “service” is the most frequently mentioned term in customer reviews, indicating that customers perceive the restaurant dining service as the most important feature. We can also see that “delicious”, “time” and “menu” appeared second most times in customer reviews, showing that customers care about the taste, quality and waiting time of the restaurants as well. The menu of the restaurants offering various foreign cuisines is important, which might be related to the authenticity of the specific cuisine. Customers, especially people from foreign countries, can easily identify the authenticity from a menu and taste of the according foreign cuisine’s restaurant. The price also matters as shown by “worth” and “price”.

We then made comparison word clouds for all the cuisines separately. We attach all the six comparison word clouds in appendix 2 for clearer demonstration. This is an example for French cuisine's comparison word cloud of restaurant customer reviews. Among the positive words, "delicious", "nice", "amazing", "excellent", "fresh" and "friendly" are the most frequently used terms. It shows the importance of taste, quality and staff service to improve customer experience in French cuisine restaurants. Among the negative words, the most frequent words, "fried", "bad" and "disappointed" show that the cooking methods and services could most probably lead to negative experience.

A word cloud of customer feedback for 'The Fish Market'. The words are arranged in a circular pattern, with the most frequent words being 'delicious', 'amazing', 'love', 'nice', 'perfect', 'friendly', 'recommend', 'wonderful', 'excellent', 'great', 'fresh', 'lovely', 'cozy', 'pleasant', 'free', 'seasoned', 'horrible', 'worst', 'disappointing', 'awful', 'fried', 'bad', 'disappointed', 'split', 'poor', 'massive', 'disappointment', 'bland', 'cold', 'late', 'mon', 'dunk', 'decadent', 'cheap', 'expensive', 'long', 'desert', 'hype', 'weird', 'rude', 'miss', 'crowded', 'sac', 'crazy', 'bloody', 'complaint', 'difficult', 'dirty', 'medicore', 'lacking', 'overpriced', 'od', 'fabulous', 'helpful', 'tender', 'outstanding', 'awesome', 'knowledgeable', 'cute', 'incredible', 'clean', 'romantic', 'wow', 'attentive', 'super', 'pretty', 'impressed', 'enjoy', 'top', 'hot', 'happy', 'decent', 'perfection', 'warm', 'glad', 'favorite', 'wild', 'prices', 'squash', 'wrong', 'disappointing', 'awful', 'fried', 'bad', 'disappointed', 'split', 'poor', 'massive', 'disappointment', 'bland', 'cold', 'late', 'mon', 'dunk', 'decadent', 'cheap', 'expensive', 'long', 'desert', 'hype', 'weird', 'rude', 'miss', 'crowded', 'sac', 'crazy', 'bloody', 'complaint', 'difficult', 'dirty', 'medicore', 'lacking', 'overpriced', 'od'.

negative

To better summarize the most important restaurant features shown from customer reviews across cuisines, we extracted the five most frequent positive and negatives words in the following tables.

- Top positive 5 words of restaurant reviews across cuisines:

Chinese	delicious	nice	amazing	love	excellent
French	delicious	nice	love	fresh	friendly
Japanese	delicious	nice	love	amazing	friendly
Korean	delicious	love	nice	hot	fresh
Indian	fresh	delicious	nice	love	pretty
Thai	delicious	hot	nice	fresh	love

The taste and quality (“delicious” and “fresh”) of the food rank first among the positive reviews. Other than that, restaurant service and the environment are also priorities for customers, since “friendly” and “pretty” are top positive words. The other words, like “nice”, “amazing”, “excellent”, “love” and “hot”, are generally complimentary words that might not indicate specific restaurant features.

- Top negative 5 words of restaurants reviews across cuisines:

Chinese	fried	bad	disappointed	hard	cold
French	fried	bad	drunken	disappointed	hard
Japanese	bad	disappointed	fried	hard	bland
Korean	fried	bad	disappointed	hard	cold
Indian	fried	bad	disappointed	hard	cold
Thai	fried	bad	dim	sour	disappointed

From the top negative words, we can observe that “fried” is the cooking method that customers might not like or easily taste off. The meaning of “cold” and “hard” is a bit ambiguous. We are not sure whether “cold” is about the restaurant environment or food, and whether “hard” is about food or restaurant services. However, restaurants should avoid these certain perspectives if they’ve been complained by customers before. Interestingly, there are some common negative words related to specific cuisines that should raise restaurant’s attention. For example, Thai restaurants should be aware of sour when talking about the taste. Japanese restaurants should

avoid blandness, and French restaurants need to take care of drunken people that might disturb other customers.

Sentiment Classifier - Classification Trees

In this part, we looked deeper into the cuisine with the highest popularity and most positive customer ratings, French cuisine. We picked the most popular French restaurant which received the most customer reviews, Commander's Palace, for further research specifically. We would like to know what factors make the restaurant popular. To achieve this goal, we built classification trees as a sentiment classifier to help us reveal the key aspects of this restaurant that drive overall customer perception.

We firstly categorized reviews to positive and negative based on the review ratings. It is reasonable to assign a positive sentiment to 4- and 5-star reviews and a negative sentiment to 1- and 2-star reviews. A 3-star review would typically be considered neutral, but for the sake of simplicity, we focused on predicting positive or negative sentiment for now. This is because our objective is to train a model to recognize language that conveys positive or negative sentiment, and 3-star reviews are likely to contain both. In natural language, neutral sentiment usually implies the absence of words with emotional connotations. However, this is not the case with reviews, as customers who give a 3-star rating are likely to have enjoyed some aspects of the product or service while disliking others. Consequently, we believe that including a "neutral" category would adversely affect the accuracy of our model, and we should exclude it for the time being.

After creating the sparse document term matrix, we ruled out some adjectives that obviously influence the rating but cannot be interpreted as restaurant features (eg. nice, great, perfect, amazing, best, love), and only focused on those nouns that can directly indicate specific restaurant features (eg. waiter, time, flavor). A tree model is built to recognize the most important factors that affect ratings.

To improve the accuracy of the tree model, we tuned the hyperparameter using Grid Research and identified the cp of 0.002.

to attract more customers and increase the customer base, they should focus on improving the taste, restaurant service and environment. The positive customer reviews, word-of-mouth, can act as a social marketing promotion and make the restaurant more attractable. If the restaurants are going to keep the existing customers and increase the customer loyalty, the restaurants should follow closely on negative customer experience for further improvements. For example, as a Japanese restaurant, it could figure out ways to reduce blandness of food.

2. By looking at the single restaurant that received the most customer reviews, we revealed more specific aspects that drive customer perceptions to the most popular restaurant. The restaurant service, punctuality, food diversity, flavor, and serving approach matter a lot. They can consider whether to continue serving some specific items or not. They can also adjust the amount of food preparation and serving portion based on the insights of customer reviews. Taking the most popular French restaurant in our dataset as an example, the serving amount of martini can be reduced.

References

1. Ariella, S. (2023, March 1). *45+ must-know us restaurant industry statistics [2023]: How many restaurants are in the US*. Zippia 45+ MustKnow US Restaurant Industry Statistics 2023 How Many Restaurants Are In The US Comments. Retrieved March 1, 2023, from <https://www.zippia.com/advice/restaurant-industry-statistics/#:~:text=The%20US%20restaurant%20industry%20made,down%20%24132.8%20billion%20from%202019>
2. *Meal delivery - united states: Statista market forecast*. Statista. (n.d.). Retrieved March 1, 2023, from <https://www.statista.com/outlook/dmo/online-food-delivery/meal-delivery/united-states#revenue>
3. Published by S. Dixon, & 11, A. (2022, August 11). *Yelp: Cumulative number of reviews 2021*. Statista. Retrieved March 1, 2023, from <https://www.statista.com/statistics/278032/cumulative-number-of-reviews-submitted-to-yelp/>
4. Liu, Y., & Jang, S. C. (S. (2009). Perceptions of Chinese restaurants in the U.S.: What affects customer satisfaction and behavioral intentions? *International Journal of Hospitality Management*, 28(3), 338–348. <https://doi.org/10.1016/j.ijhm.2008.10.008>
5. AUAPINYAKUL, W., SIRIPONGDEE, S., & PIMDEE, P. (2022). *Antecedents of Customer Loyalty and Perceived Service Quality: A SEM Analysis of Thai Restaurant Brands*. *The Journal of Asian Finance, Economics and Business*, 9(7), 173–183. <https://doi.org/10.13106/JAFEB.2022.VOL9.NO7.0173>
6. Josiam, B. M., & Monteiro, P. A. (2004). Tandoori tastes: Perceptions of Indian restaurants in America. *International Journal of Contemporary Hospitality Management*, 16(1), 18–26. <https://doi.org/10.1108/09596110410516525>
7. Jang, S. C. (S., Ha, J., & Park, K. (2012). Effects of ethnic authenticity: Investigating Korean restaurant customers in the U.S. *International Journal of Hospitality Management*, 31(3), 990–1003. <https://doi.org/10.1016/j.ijhm.2011.12.003>

Appendixes

1. Raw Datasets Description

Variable Description Tables		
<u>business dataset</u> : contains business data including location data, attributes, and categories.		
Field Name	Data Type	Description
business_id	string	22 character unique string business id
name	string	the business's name
address	string	the full address of the business
city	string	the city
state	string	2 character state code, if applicable
postal code	string	the postal code
latitude	float	latitude
longitude	float	longitude
stars	float	star rating, rounded to half-stars
review_count	integer	number of reviews
is_open	integer	0 or 1 for closed or open, respectively
attributes	object	business attributes to values. note: some attribute values might be objects
categories	array	an array of strings of business categories
hours	object	an object of key day to value hours, hours are using a 24hr clock
<u>review dataset</u> : contains full review text data including the business_id that the review is written for.		
Field Name	Data Type	Description
review_id	string	22 character unique review id
user_id	string	22 character unique user id, maps to the user in user.json
business_id	string	22 character business id, maps to business in business.json
stars	integer	star rating
date	string	date formatted YYYY-MM-DD
text	string	the review itself
useful	integer	number of useful votes received

positive



negative

c. Japanese Cuisine

positive



negative

d. Korean Cuisine

positive



negative

e. Thai Cuisine

positive



negative

f. Chinese Cuisine

positive



negative