

# Yelp Reviews and Food Types: A Comparative Analysis of Ratings, Sentiments, and Topics

Wenyu Liao<sup>1,†</sup>, Yiqing Shi<sup>1,†</sup>, Yujia Hu<sup>1,2</sup>, and Wei Quan<sup>1,2</sup>

<sup>1</sup> Department of Statistics and Data Science, BNU-HKBU United International College, Zhuhai 519087, China  
{q030026082,q030026125}@mail.uic.edu.cn

<sup>2</sup> Guangdong Provincial Key Laboratory IRADS, BNU-HKBU United International College, Zhuhai 519087, China  
{yujiahu,weiquan}@uic.edu.cn

†These authors contributed equally to this work and share first authorship

**Abstract.** This study examines the relationship between Yelp reviews and food types, investigating how ratings, sentiments, and topics vary across different types of food. Specifically, we analyze how ratings and sentiments of reviews vary across food types, cluster food types based on ratings and sentiments, infer review topics using machine learning models, and compare topic distributions among different food types. Our analyses reveal that some food types have similar ratings, sentiments, and topics distributions, while others have distinct patterns. We identify four clusters of food types based on ratings and sentiments and find that reviewers tend to focus on different topics when reviewing certain food types. These findings have important implications for understanding user behavior and cultural influence on digital media platforms and promoting cross-cultural understanding and appreciation.

**Keywords:** Yelp reviews · food types · cultural influence.

## 1 Introduction

Analyzing online reviews has become increasingly important over the past decade as more and more people rely on them to make decisions. Online reviews can provide valuable information about the quality, value, and overall experience of products and services from the perspective of other people.

Yelp is one of the largest online review platforms, with millions of reviews and ratings for businesses across various industries. According to [1], the large sample size of Yelp reviews reflects the dominant culture, which is continuously and historically rooted in favoring the white, Euro-centric experience. The average Yelp reviewer focuses on dirt floors, plastic stools, and other patrons who are non-white when reviewing non-European restaurants. However, the reviewers associate more positive characteristics when discussing European cuisines.

Stereotypes and biases can be deeply ingrained in cultural attitudes toward food, and these attitudes may manifest in the way that people review and rate

restaurants. Thus, this study investigates the relationship between Yelp reviews and food types, investigating how ratings, sentiments, and topics vary across different types of food in the Yelp dataset. Specifically, we analyze how ratings and sentiments of reviews vary across food types, cluster food types based on their ratings and sentiments, and compare topic distributions among different food types. We aim to answer the following research questions:

- 1. What is the mean rating distribution of food types? Do reviewers tend to rate different food types differently?
- 2. What is the proportion of sentiment distribution of food types? Do reviewers tend to express different sentiments towards different food types?
- 3. What are the clusters of food types based on reviewer ratings and sentiments?
- 4. What are the topic distributions of food types? Do reviewers tend to focus on different topics when they review different food types?

Our analyses can lead to various potential contributions, including understanding user behavior and cultural influence on digital media platforms. Our findings highlight the importance of recognizing and appreciating the diversity of food cultures and promoting cross-cultural understanding and appreciation.

## 2 Background

There has been a significant increase in research on online reviews, driven by the growing importance of online reviews as a tool for decision-making. Researchers have focused on various aspects of online reviews, such as analyzing online tourism reviews to explore racism [2], exploring characteristics of reviews and reviewers shape review frequency and continuity [3], among others.

Some of the previous work on Yelp data includes studying reviews, race, and gentrification [4], investigating the cultural impact on social commerce [5], identifying bias [6], evaluating reviews of hospitals and experiences of racism [7], examining whether a cuisine is part of mainstream [8], exploring authenticity [9, 10], investigating negative public sentiment during the COVID-19 pandemic [11], and comparing reviewing behaviors on Chinese and French restaurants [12].

While there is a growing body of research on online reviews and their impact on businesses and consumers, there is still a need for a more detailed analysis of Yelp reviews on different food types. The gap our study intends to bridge is the need for more specific insights into the relationship between Yelp reviews and food types. Our study aims to fill this gap by conducting a comparative analysis of ratings, sentiments, and topics of Yelp reviews in the US and providing specific insights into how reviews vary across food types. By doing so, our study can contribute to the growing body of research on online reviews and provide helpful information for businesses, policymakers, and researchers interested in understanding user behavior and cultural influence on digital media platforms.

### 3 Methods

#### 3.1 Data

In this study, we focus on restaurants in the US. We analyze the Yelp Open Dataset<sup>1</sup>, a subset of Yelp’s businesses, reviews, and user data for personal, educational, and academic purposes. We identify restaurants in ten popular food types in the US using Yelp restaurant categories<sup>2</sup>, see Table 1. Then, we remove duplicate restaurants (cross-reference in ten food types) and filter out only English reviews. Our final dataset includes 1,152,452 reviews on 7,141 restaurants.

**Table 1.** Food types, number of reviews, and categories

FoodType	Review#	Categories	FoodType	Review#	Categories
American	511,848	American American (New) American (Traditional)	French	18,811	French Mauritius Reunion
Japanese	125,260	Japanese Conveyor Belt Sushi Izakaya Japanese Curry Ramen Teppanyaki	Chinese	76,929	Chinese Cantonese Dim Sum Hainan Shanghainese Szechuan
Italian	141,160	Italian Calabrian Sardinian Sicilian Tuscan	Mexican	166,486	Mexican Tacos
Greek	23,050	Greek	Korean	24,691	Korean
Indian	23,190	Indian	Thai	41,027	Thai

#### 3.2 Measurements

Our dataset contains restaurant-level measurements, including food type, review count, star rating, and review-level measurement of star rating. We create topic and sentiment measurements at the review level for analysis purposes.

**Sentiment** We use Vader [13] to infer sentiments of reviews with three categories, including negative, neutral, and positive. Vader is a popular tool for sentiment analysis that uses a lexicon-based approach to infer the sentiment of a given text. It is particularly useful for analyzing short texts like tweets and product reviews, where traditional machine learning approaches may not be feasible due to the limited training data.

<sup>1</sup> <https://www.yelp.com/dataset>

<sup>2</sup> [https://blog.yelp.com/businesses/yelp\\_category\\_list/](https://blog.yelp.com/businesses/yelp_category_list/)

**Topic** We adopt the framework from previous studies [5, 14–16] to classify reviews into four major topics: food quality, service, ambiance, and price fairness. We train classification models with annotated data from [17], then apply these models to infer reviews topics. Each review can have one or multiple topics. Specifically, we apply Natural Language Toolkit (NLTK) [18] to process Yelp reviews. We employ a combination of grid search and cross-validation to train various machine learning models for the topic classification task, including Logistic Regression, Decision Tree, Support Vector Machine, Random Forest, and Naive Bayes. We use precision, recall, and F1-score as evaluation metrics. Then, we apply the best models in terms of the F1-score to infer topics for our reviews.

### 3.3 Analyses

**Kolmogorov-Smirnov test** The Kolmogorov-Smirnov (KS) test is a non-parametric statistical test for whether a frequency distribution is identical to the theoretical distribution or whether two samples come from the same distribution. It is a test that does not require assumptions about the population distribution. Given the large amount of data we have, the KS test is a highly efficient way to analyze the distribution of our sample, especially when dealing with large sample sizes. To further improve the reliability and accuracy of the KS test, we use the Bootstrap method to generate multiple re-sampled datasets and simulate the variability of the sample distribution. Using Bootstrap re-sampling, we can obtain a more reliable estimate of the test statistic and calculate confidence intervals that reflect the precision of our estimate.

**K-means Clustering Analysis** Clustering analysis helps us understand the similarities and differences in review ratings and sentiment scores among food types, enabling comparison of food types. We use the `fviz_nbclust` function to determine the optimal number of clusters. The function evaluates the clustering quality by calculating the within-cluster sum of squares (WSS) for different numbers of clusters. We choose the optimal number of clusters by observing the elbow or inflection point in the WSS curve. This point indicates a significant decrease in the improvement of the WSS as the number of clusters increases. Thus selecting this number of clusters yields better clustering results.

## 4 Results

### 4.1 RQ1. What is the mean rating distribution of food types? Do reviewers tend to rate different food types differently?

We use the Bootstrap KS test to compare whether the ratings of two food types come from the same distribution. The results suggest that all of the ratings of the food types are from different distributions, which implies that there are significant differences between the food types in terms of how they are rated.

This finding is important because it indicates that the food types may have different characteristics or experiences that affect their perceptions or ratings.

Fig. 1 shows that all ten food types have different mean ratings and that some are more likely to receive higher ratings than others. The food types arranged in descending order based on their mean ratings are Greek, Korean, French, Thai, Japanese, Indian, Italian, American, Mexican, and Chinese.

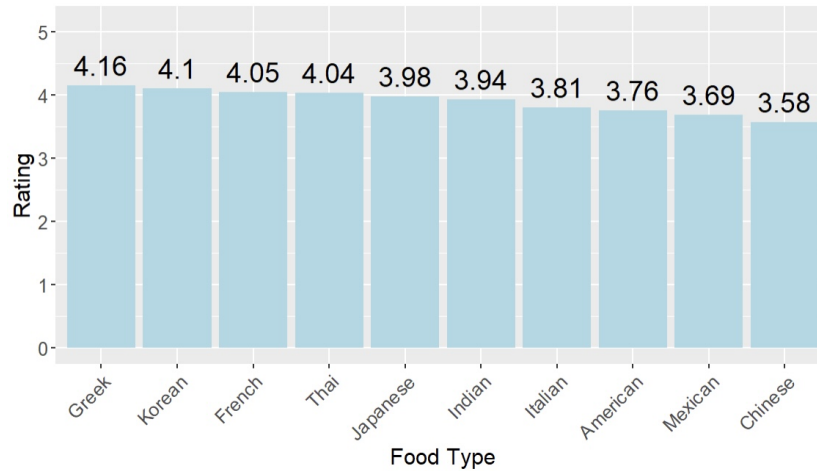


Fig. 1. Distribution of Mean Ratings of Food Types

4.2 RQ2. What is the proportion of sentiment distribution of food types? Do reviewers tend to express different sentiments towards different food types?

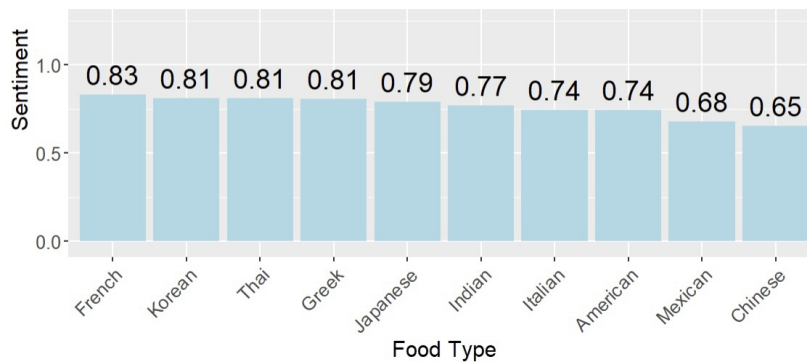


Fig. 2. Distribution of Proportions of Sentiments of Food Types

Similarly, we use the Bootstrap KS test to compare whether the sentiments of two food types come from the same distribution. The results show that American and Italian, Greek and Korean, Greek and Thai, and Korean and Thai have the same distribution. Other pairs of the sentiments of the food types are from different distributions, which implies that there are significant differences between the food types in terms of how reviewers express sentiments when they review.

Fig. 2 shows that some food types have different proportions of sentiments and that some are more likely to receive positive sentiments than others. The food types arranged in descending order based on their average sentiment scores are French, followed by a group of Korean, Thai, and Greek, then Japanese, Indian, another group of Italian and American, Mexican, and finally, Chinese.

#### 4.3 RQ3. What are the clusters of food types based on reviewer ratings and sentiments?

We use review star ratings and sentiments to cluster food types. We first standardize the data and conduct a normality test. We choose the optimal number of clusters to be four. Detailed results of food types and clusters are in Table 2.

**Table 2.** Food types and Clusters

	Cluster1	Cluster2	Cluster3	Cluster4
Food Types	Chinese, Mexican	French, Greek, Korean, Thai	Indian, Japanese	American, Italian

#### 4.4 RQ4. What are the topic distributions of food types? Do reviewers tend to focus on different topics when they review different food types?

We also use the Bootstrap KS test to compare whether the topics of two food types come from the same distribution. The results indicate that American, Italian, Japanese, and Korean have the same distribution; Indian and Thai have the same distribution; and Chinese and Mexican have the same distribution. The topics of other pairs of food types are from different distributions, which implies that there are significant differences between certain food types in terms of which topics reviewers focus on when they review.

Fig. 3. shows the topic distributions by food types in all reviews. Reviewers tend to focus more on ambiance and less on food quality across all food types. Above 40% of the reviews in Chinese, French, Greek, Indian, Italian, Japanese, Korean, and Mexican are about price fairness. Thai restaurants have less than 40% of reviews on price fairness, and American restaurants only have 8.5%. 24.2% of reviews on Thai restaurants mention services, followed by Greek, Indian, Chinese, Mexican, Italian, Japanese, Korean, French, and American.

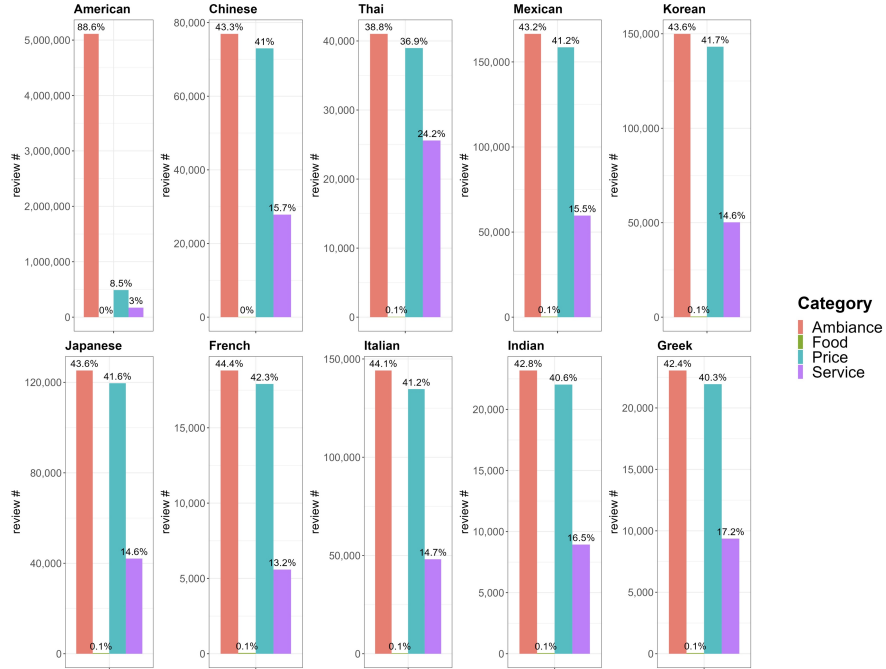


Fig. 3. Distributions of Topics by Food Types

## 5 Discussions

The findings of RQ1 and RQ2 suggest significant differences in the ratings and sentiments of different food types. Various factors, including cultural factors, historical context, globalization, and health and nutrition concerns, could influence these differences.

French cuisine has a long history of being associated with fine dining and upscale restaurants, which could lead to higher ratings due to perceptions of prestige and exclusivity. Similarly, Greek cuisine may be associated with healthy eating and the Mediterranean diet, which is popular in the US. Also, Greek and French cuisines may be more popular in urban areas of the US (where our Yelp reviews are from), with a higher concentration of upscale restaurants and foodies. This could lead to higher ratings due to a greater appreciation for these cuisines among restaurant reviewers. In the US, where there is a growing interest in healthy eating and wellness, these perceptions of health could contribute to higher ratings for Greek and French cuisines.

Korean cuisine is becoming increasingly popular in the US, with more Americans seeking Korean food and flavors. This growing popularity could lead to higher ratings for Korean restaurants as more people become familiar with and

appreciate Korean cuisine. Korean cuisine is known for its unique and bold flavors and use of ingredients that may be less common in other cuisines. This distinctiveness could lead to higher ratings due to the novelty and excitement of trying something new and different. Korean cuisine has become a popular topic on digital media platforms, with users sharing photos and videos of Korean dishes and restaurants. This exposure could lead to higher ratings for Korean restaurants as more people become interested in trying Korean food.

Stereotypes and biases could play a role in the lower ratings of Mexican and Chinese restaurants. For example, Mexican cuisine may be associated with fast food or cheap, greasy fare, while Chinese cuisine may be associated with low-quality ingredients. These stereotypes could lead to negative perceptions and lower ratings. Also, there may be more competition among Mexican and Chinese restaurants in the US, leading to lower ratings due to higher expectations and more discerning diners. Additionally, the demand for these cuisines may be higher, leading to more casual or low-quality restaurants opening to meet demand and lower ratings overall.

Looking at Table 2, Chinese and Mexican cuisines may share some similarities in flavor profiles and ingredients, such as spicy or savory seasonings and sauces. French, Greek, Korean, and Thai cuisines cluster together. These cuisines may share some similarities in terms of their emphasis on fresh ingredients, bold flavors, and complex preparation techniques. Additionally, these cuisines may be perceived as more exotic or adventurous compared to other cuisines, which could lead to higher ratings and more positive sentiment among diners. Indian and Japanese cuisines are grouped together. These cuisines may share some similarities in terms of their use of spices and seasonings, their focus on fresh ingredients, and their simple preparation methods. Additionally, both cuisines have gained popularity in recent years, with more Americans seeking authentic Indian and Japanese food. American and Italian cuisines are in the same cluster. They may share similarities regarding their focus on comfort food and familiar flavors. Additionally, both cuisines are widely available in the US, which could lead to some overlap in the types of restaurants and dishes most commonly reviewed. It is also possible that the sentimental attachment that many Americans have to classic American dishes and Italian cuisine could contribute to higher ratings and more positive sentiments.

Findings from RQ4 reveal that some food types have the same topic distributions while others do not. This could be due to various factors, such as cultural and regional preferences, differences in the types of restaurants and dishes most commonly reviewed, and variations in the dining experience associated with different cuisines. For example, the similarities in topic distribution among American, Japanese, Korean, and Italian cuisines could be due to their popularity and widespread availability in the US, leading to more consistent patterns in the types of reviews and feedback reviewers provide. On the other hand, the differences in topic distribution among other food types could be due to their unique flavor profiles, preparation methods, and cultural associations, contributing to variations in the types of experiences and feedback reviewers have.



Overall, these findings suggest significant differences in the topic distributions of different food types, and reviewers may focus on different topics when reviewing different types of cuisine. Further research would be needed to understand the reasons for these differences fully and identify the specific factors contributing to the topic distributions observed in this study.

## 6 Conclusions

In conclusion, our study has investigated the relationship between reviews and food types in the Yelp dataset. We have identified patterns and trends that shed light on reviewers' preferences and behavior through our analyses of rating and sentiment distributions, clusterings of food types, and topic distributions. Our findings suggest that different food types have distinct rating distributions and that reviewers tend to focus on different topics when reviewing specific types of food. Our study demonstrates the value of advanced analytical techniques for analyzing user-generated content.

While our study provides valuable insights, it is important to consider these limitations when interpreting the results. Our study relies on the Yelp dataset, which may only represent some US restaurants. Yelp users may be a self-selected group of reviewers who are more likely to write reviews, and the types of restaurants and food types most commonly reviewed may differ from the restaurant industry as a whole. Additionally, the Yelp dataset may be subject to biases and inaccuracies, such as fake reviews or reviews that do not reflect actual dining experiences. Our study focuses on a limited set of variables, such as rating, sentiment, and topic distributions. While these variables are informative, they may not provide a complete picture of reviewers' preferences and behavior. Additionally, our study does not provide contextual information about the restaurants. For example, it does not account for differences in restaurant quality, the dining experience, or cultural and regional preferences that may influence user ratings and sentiments. Further research that addresses these limitations could provide a more comprehensive understanding of user preferences and consumer behavior.

## Acknowledgment

This work is supported in part by the Guangdong Provincial Key Laboratory IRADS (2022B1212010006, R0400001-22) and UIC Research Grant with No. of UICR0700038-22 at BNU-HKBU United International College, Zhuhai, PR China.

## References

1. Sara, K.: Authenticity in online ethnic restaurant reviews: Revealing conflicted nationalism in multicultural consumption. In: Annual Meeting of the Agriculture, Food, and Human Values Society and the Association for the Study of Food and Society (2017)

2. Li, S., Li, G., Law, R., Paradies, Y.: Racism in tourism reviews. *Tourism Management* **80**, 104100 (2020)
3. Chen, H.N., Huang, C.Y.: An investigation into online reviewers' behavior. *European Journal of Marketing* **47**(10), 1758–1773 (2013)
4. Zukin, S., Lindeman, S., Hurson, L.: The omnivore's neighborhood? online restaurant reviews, race, and gentrification. *Journal of Consumer Culture* **17**(3), 459–479 (2017)
5. Nakayama, M., Wan, Y.: The cultural impact on social commerce: A sentiment analysis on yelp ethnic restaurant reviews. *Information & Management* **56**(2), 271–279 (2019)
6. Choi, S., et al.: An empirical study identifying bias in Yelp dataset. Ph.D. thesis, Massachusetts Institute of Technology (2021)
7. Tong, J., Andy, A.U., Merchant, R.M., Kelz, R.R.: Evaluation of online consumer reviews of hospitals and experiences of racism using qualitative methods. *JAMA Network Open* **4**(9), e2126118–e2126118 (2021)
8. Boch, A., Jiménez, T., Roesler, K.: Mainstream flavor: Ethnic cuisine and assimilation in the united states. *Social Currents* **8**(1), 64–85 (2021)
9. Le, T.H., Arcodia, C., Novais, M.A., Kralj, A., Phan, T.C.: Exploring the multidimensionality of authenticity in dining experiences using online reviews. *Tourism Management* **85**, 104292 (2021)
10. Yu, C., Margolin, D.: Food “taste” on yelp: displays of cultural omnivorousness and authenticity in dining choices and online reviews. *Food, Culture & Society* **26**(1), 3–23 (2023)
11. LaBelle-Hahn, A.: Pandemic-induced anti-asian sentiment in chicago: An exploratory analysis of restaurant yelp reviews (2022)
12. Quan, W., Li, Z., Kim, W.: Implicit discriminatory bias reflected in yelp reviews—a case study. In: 2022 IEEE International Conference on Big Data (Big Data). pp. 6799–6801. IEEE (2022)
13. Hutto, C., Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the international AAAI conference on web and social media. vol. 8, pp. 216–225 (2014)
14. Ha, J., Jang, S.S.: Perceived values, satisfaction, and behavioral intentions: The role of familiarity in korean restaurants. *International Journal of Hospitality Management* **29**(1), 2–13 (2010)
15. Jeong, E., Jang, S.S.: Restaurant experiences triggering positive electronic word-of-mouth (ewom) motivations. *International Journal of Hospitality Management* **30**(2), 356–366 (2011)
16. Ryu, K., Lee, H.R., Kim, W.G.: The influence of the quality of the physical environment, food, and service on restaurant image, customer perceived value, customer satisfaction, and behavioral intentions. *International journal of contemporary hospitality management* **24**(2), 200–223 (2012)
17. Sajnani, H., Saini, V., Kumar, K., Gabrielova, E., Choudary, P., Lopes, C.: Multi-label classification of reviews in yelp data. Technical report (2014)
18. Loper, E., Bird, S.: Nltk: The natural language toolkit. arXiv preprint cs/0205028 (2002)