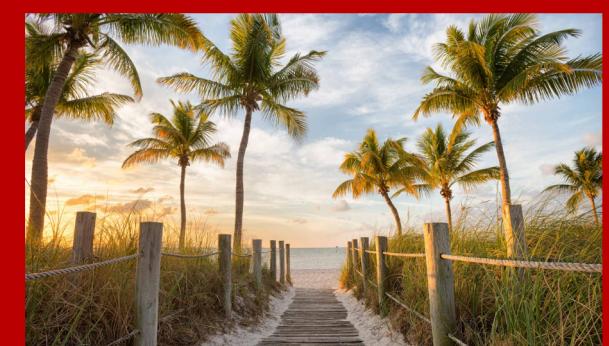


# FLORIDA RESTAURANT ANALYSIS

## Zijing Cheng, Nianqing Chen, Zixuan Li



### ABSTRACT

Dining plays a crucial role in life, especially dining experiences. In this digital world, choosing a restaurant became easier with the availability of apps like Yelp. This study focuses on exploring the influences on the customer reviews and the ratings by the restaurants' geographic locations. As well as the relationship between customer reviews and business status. The research aims to provide insights into how restaurant operations and customer feedback are related in the state of Florida, ultimately informing strategies for improving restaurant performance and customer satisfaction.

The restaurants are re-categorized by three ways: the **food region** such as Asian and American food, the **special providence** such as gluten free and seafood, and the **restaurant types** such as bar and steakhouse. Many categories are repetitive and not possible to avoid since everyone can give the restaurant category tags.

Classifications using random forest are performed based on the three main categories mentioned above. Additionally, the ratings changing by time is also visualized for further analyze.

### INTRODUCTION

Restaurants are an integral part of our daily lives, offering a space to socialize, celebrate, and explore diverse cuisines beyond home-cooked meals. Yelp, a popular review website, facilitates informed dining decisions by allowing customers to share their experiences and feedback on restaurants. With Yelp's platform, customers can evaluate restaurants' services and performance, leading to the discovery of new dining options and helping restaurants improve their services to meet customer needs.

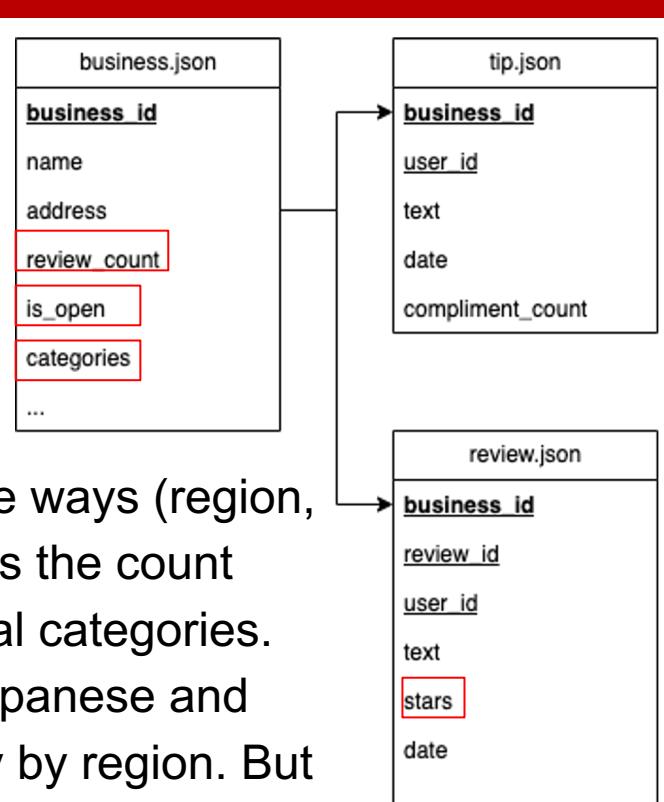
The data related to restaurants is significantly more than other categories such as theaters and other entertainments which support this study perfectly. Especially in the state of Pennsylvania and Florida. PA and FL are both popular choices for vacations. However, FL has more evenly distributed beaches and entertainments throughout the state. This provides diversity on geographic locations instead of the data in PA, which is mainly distributed in Philadelphia.

This study investigates the relationship between a restaurant's location, type of business, and customer reviews on Yelp. Specifically, it seeks to understand how location and cuisine type affect customer comments and rating scores on Yelp, as well as the impact of reviews on the business status of restaurants. By examining these relationships, we can gain a comprehensive understanding of the factors that influence restaurant performance and customer satisfaction.

### DATA

The data is from Yelp.com including information of merchants, reviews, star ratings, and tips from users. As shown in Figure 2, the scores are displayed by looking into the data summary.

By classifying all categories in three ways (region, special, service format), here shows the count types of restaurants in more general categories. Shown in Table 1. For example, Japanese and Chinese are both in Asian category by region. But they could be in the seafood category by special food providence.



Region Category	Count
American	2550
Asian	821
European	1060
Latin American	1022

Special Category	Count
Gluten-Free	42
Salad	422
Seafood	724
Vegan	124
Vegetarian	94

Format Category	Count
Bars	853
Desserts	282
Events	767
Food Delivery	124
Food Stores	289
Food Trucks	179
Juice Bars	170

Table 1: Summary by Three Categories

Attribute	N	Mean	SD
Stars	8177	3.627	0.7810
Review Count	8177	127.6	218.0105
Compliment Count	8177	0.7028	0.4570
Business Status	8177	0.7028	0.4570

Table 2: Summary of Variables in-use

### ANALYSIS I&II

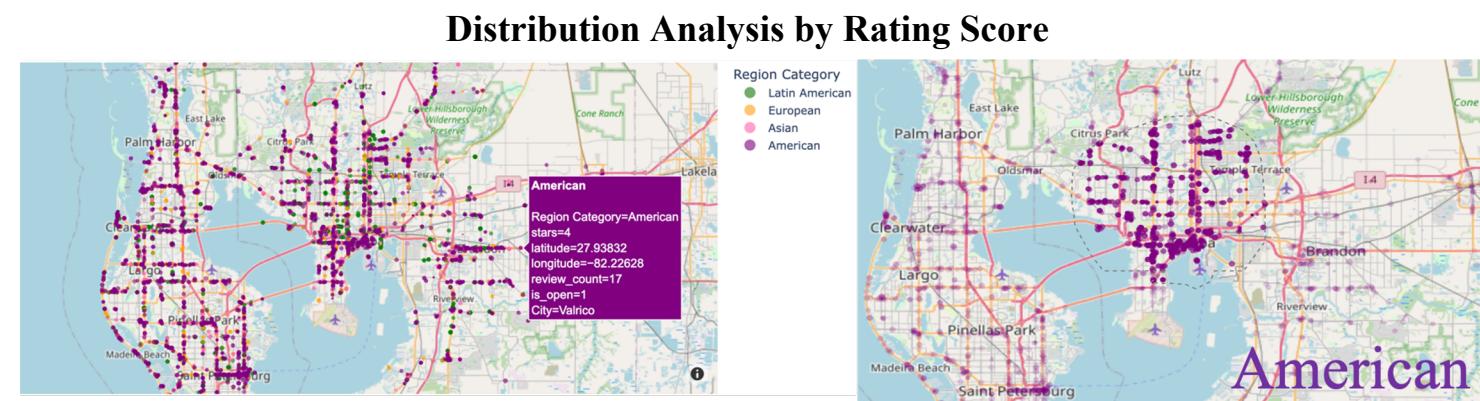


Figure 2: Correlation Distribution Map of Restaurant (By Region Category) and Rating Scores

Figure 2 reveals that **American restaurants** have a higher number and rating advantage compared to other types under the geographical criterion. They are consistently distributed throughout the region, with a concentration in the area surrounded by Citrus Park, Temple Terrace, and Tampa. This area is considered the best distribution area for American restaurants, while there is no significant distribution pattern for other types in the region.

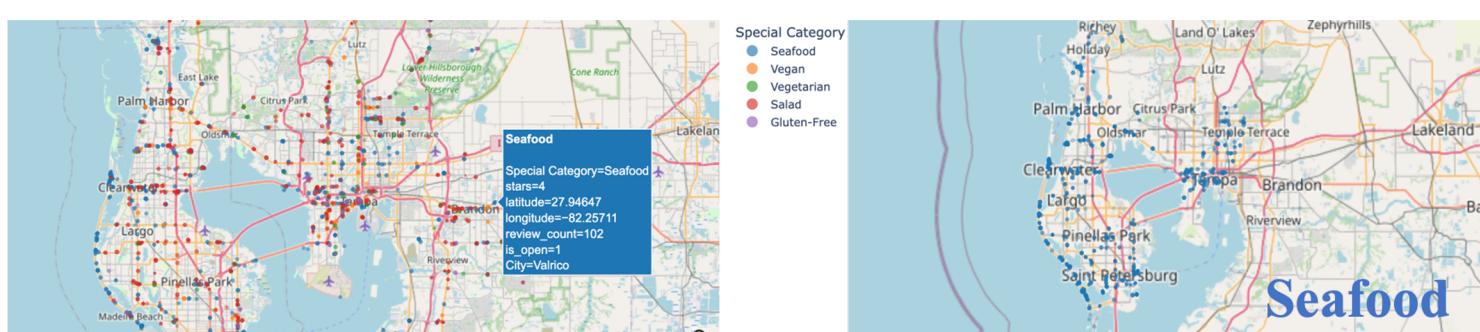


Figure 3: Correlation Distribution Map of Restaurant (By Special Category) and Rating Scores

Based on the food characteristics criterion, **Seafood restaurants** are the most consistently distributed type in the region. They are concentrated near Saint Petersburg and Old Tampa Bay and have a linear distribution pattern along the coast.

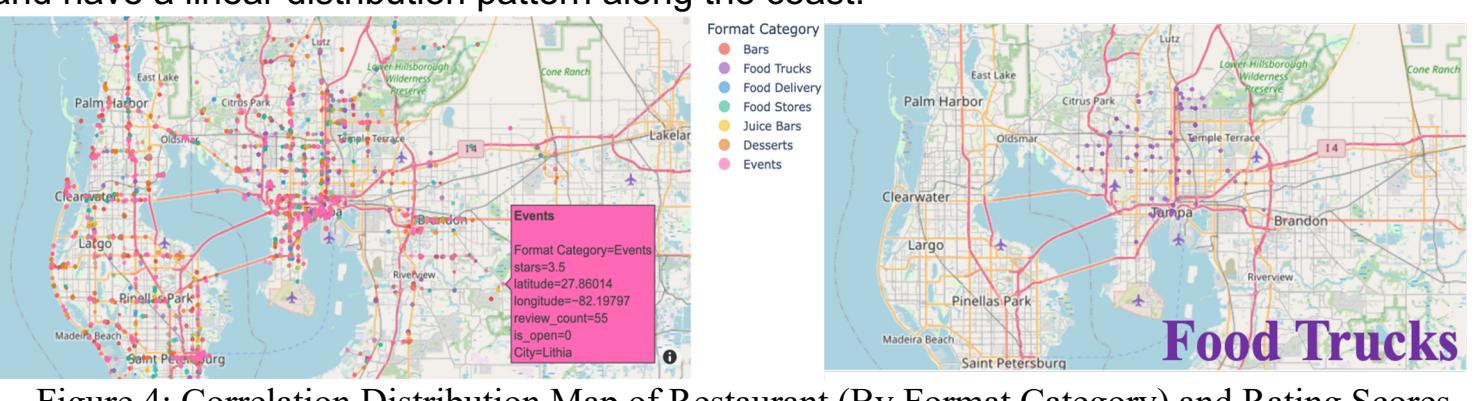


Figure 4: Correlation Distribution Map of Restaurant (By Format Category) and Rating Scores

Under the criterion of food type and dining style, **Food trucks** are the most consistently distributed type, even though they have a lower quantity compared to other restaurant types under the same criteria. They are concentrated in the Old Tampa Bay area, while other restaurant types have a more uniform distribution, indicating a more even level of star ratings.

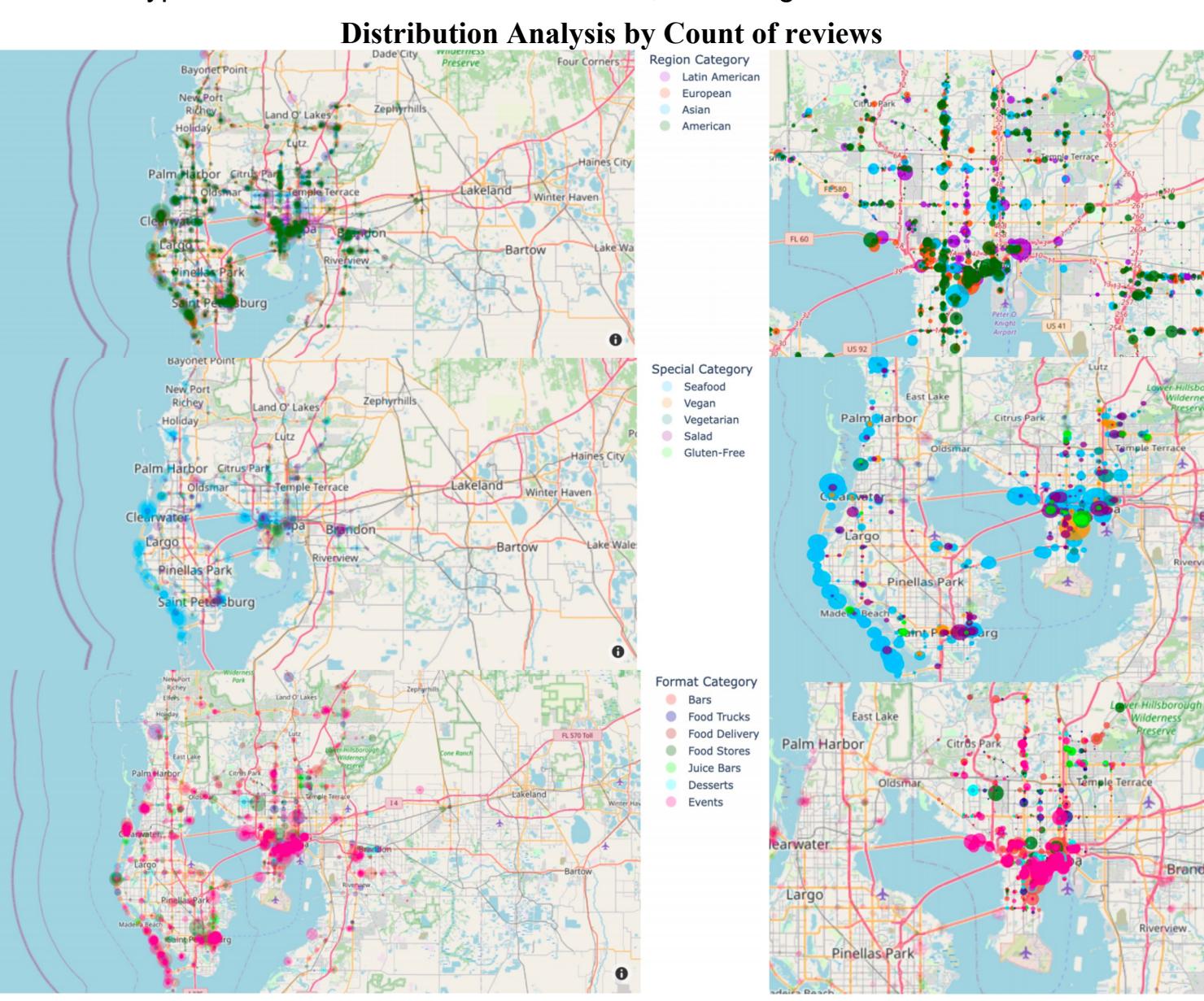


Figure 5: Correlation Distribution Map of Restaurant (By Three Classification Criteria) and Count of Reviews

The left section of Figure 5 displays the distribution of restaurant categories under three different criteria, with the position of markers determined by latitude and longitude, and the size of markers determined by **count of reviews**. The right section shows the most representative restaurant areas for each classification criteria.

In the upper left corner of the map, there is a large green area that represents American restaurants. This means that American restaurants are the most popular in this region and are found in many places. The bigger markers indicate that these restaurants are more representative of the region. There is also a special area in the upper right corner of the map. This area is located south of Citrus Park and Temple Terrace, and it is surrounded by Old Tampa Bay. In this area, restaurants of all categories are gathered, and people seem to be very interested in them. Many different types of restaurants have larger markers in this area, indicating that people are willing to leave reviews and share their opinions about them.

The middle left area of the map shows the distribution of restaurants based on their characteristics. The most prominent color and widespread distribution belong to seafood restaurants. They have a distinct pattern of distribution. In the middle right area, there are two distinct patterns of distribution: horizontal and vertical. The horizontal pattern covers Tampa International Airport, Tampa Heights, East Ybor, Harbor Island, and Palma Cera. The vertical pattern goes from Hudson in the north to Saint Pete Beach in the south, passing through New Port Richey, Holiday, Palm Harbor, Clearwater, Madeira Beach, and Treasure Island. The vertical pattern follows the coastline, while the horizontal pattern is concentrated in the Old Tampa Bay area. Salad restaurants have a unique distribution pattern, mainly concentrated in the Saint Petersburg area. Compared to seafood restaurants, they have a smaller and more concentrated distribution.

The lower left area of the map shows the distribution of restaurants based on their dining mode. The prominent pink color represents Events, which provides a venue for formal dining activities.

The map reveals a pattern of restaurant distribution. The areas where all types of restaurants are concentrated include Old Tampa Bay, Tampa Road, West Hillsborough Avenue, Bruce B Downs Boulevard, and the Oldsmar and Citrus Park areas within Temple Terrace. These areas offer a diverse range of restaurants and emphasize the dining habits and food delivery services that restaurants provide. Overall, the map highlights the popularity of functional restaurants, with a focus on formal dining experiences and a variety of dining options.

### ANALYSIS III

#### Points about Features Processing:

Feature	Summary
<b>Location:</b> Theater Beach Shopping Mall Museum	- Selected the locations of additional four representative consumption and sightseeing places, and use 2-mile as the dividing line to judge whether the restaurant is close to the four types of places.
<b>Category:</b> Region Format Special	- Performed Random Forest Model based on three category dataset. - Resigned restaurants' category into less groups to reduce tiny categories with three ways - Deleted all restaurants with multiple labels in each classification methods

### Results & Comparison

- Separate the three category datasets into 8:2 training and testing set
- Used random forest models for the category datasets to predict restaurants' business status of testing data set.

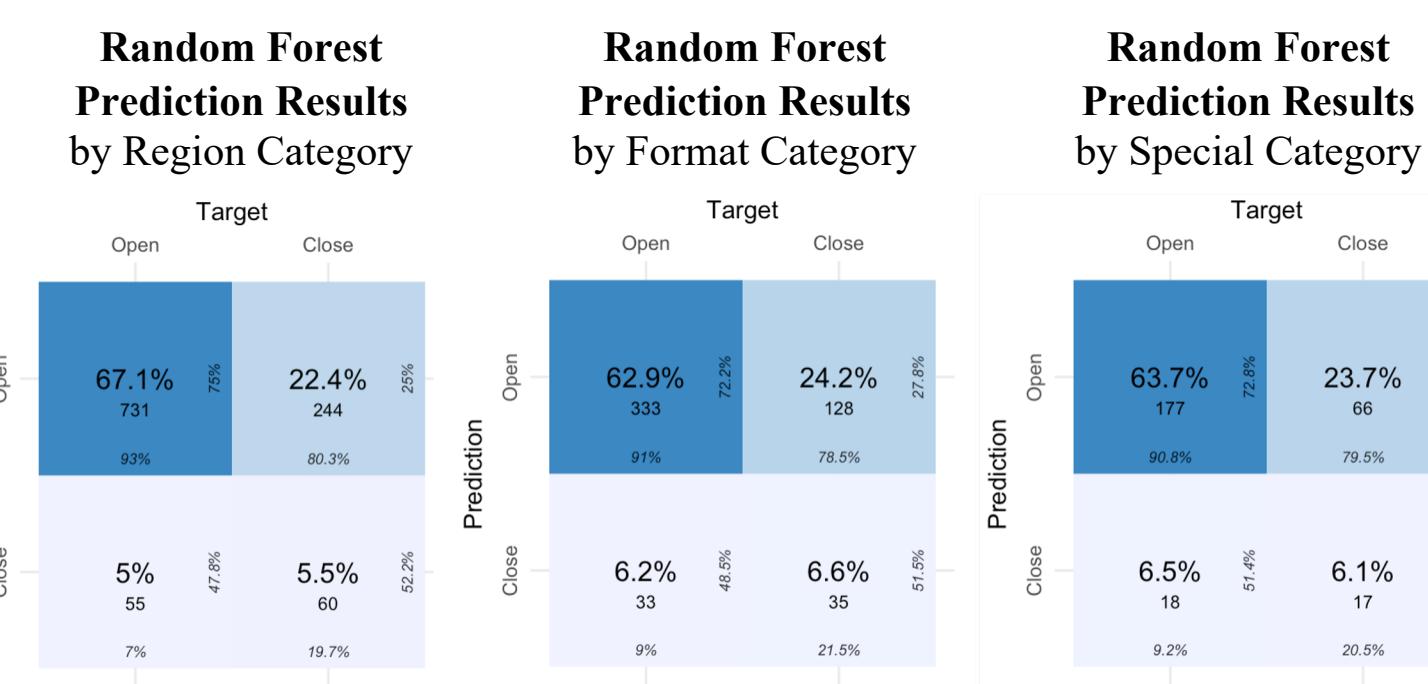


Figure 8: Confusion Matrix of Random Forest Model by Three Category Methods

### Random Forest Model Accuracy Results:

- Region Category:** 72.57%
- Format Category: 69.57%
- Special Category: 69.78%

According to the classification method, restaurants classification based on **food region** explained more on restaurants' business status, and we try to analyze people's preferences and evaluation trends for Florida restaurants in the past 15 years focusing on restaurant dish region classification.

Looking at the changes in the number of reviews and ratings for open and closed restaurants on Figure 7 & Figure 8, we came to the following conclusions:

- American and Latin American** restaurants are the **most frequently reviewed** types of restaurants in Florida while **Asian and European** restaurants are the **highest scored** types of restaurants in Florida.
- The ratings of open restaurants remain relatively stable, despite fluctuations in the number of reviews, in contrast to closed restaurants. The Ratings of open restaurants are **higher** than that of closed on average and the number of reviews of open restaurants are **much higher** than closed restaurants.

### CONCLUSION

In conclusion, our data analysis reveals that the Florida restaurant landscape is predominantly shaped by Latin American and American eateries, with Asian and European establishments serving as secondary influences. American and Latin American restaurants not only draw in the most customers but also elicit the most reviews, demonstrating their widespread appeal and popularity.

- While restaurant ratings remain relatively stable over time, a rise in positive reviews correlates with an increase in overall ratings. Asian restaurants boast the highest average ratings, though they have fewer total reviews compared to other categories. This may be due to a tendency among patrons to leave positive feedback more frequently than negative. As the total number of reviews grows, however, negative feedback becomes more common, leading to a decrease in overall ratings.
- In contrast, closed restaurants experience a decline in ratings as the number of reviews increases, particularly during 2016. This trend may stem from customers being more likely to leave negative reviews for shuttered establishments. The limited number of reviews for restaurants on the verge of closure in 2022 may introduce bias and result in less accurate ratings.
- By examining various classification criteria, we can conclude that American, seafood, and event-focused restaurants have the broadest reach and highest ratings. These establishments are favored by consumers, who actively provide feedback to inform others of their experiences. Consequently, restaurant owners and managers should take these trends into account when making strategic decisions, such as choosing a location, planning menus, and designing marketing campaigns, to better cater to their customers' tastes and preferences.