Santa Barbara Restaurant Analysis

This report covers a brief analysis of the restaurant found in the City of Santa Barbara. Prior to the analysis, let us get to know the city, and the rich culture the it has.

Santa Barbara is referred to “American Riviera”, with it’s Spanish colonial buildings style, and breathtaking mountainous and beach views than span along miles. It is a great place to wine and dine, visit historical museums, or just walk around during sunset. This report will focus more on the dining experience Santa Barbara provides by comparing the top 5 cuisines.

Data and Feature Engineering

The restaurants’ data was acquired from the pubic Yelp Dataset provided by Yelp in 2021. Data was split into 5 categories: business, user, reviews, check in, and tips. For the purpose of this analysis, we only focused on business and review datasets. Once the data was acquired, the majority of the wrangling was for the `categories` column. Each row in that column was presented as a string, text, of different categories the restaurant falls in. The string contained specific cuisines such as American, Japanese, etc and specific food item such as fries, sushi, etc. To simplify the strings, we found a dataset that broke down all the yelp categories into different columns. From there, we created a list of the main cuisines, and mapped it to the original `categories` column in order to keep the categories simple. One disadvantage, is that some restaurants might put more cuisines, that what they serve. For example, once the categories were obtained, we made word clouds, and we noticed Indian dishes in the Australian Cuisine cloud. We further searched up the restaurant, and it was an Indian restaurant. More accurate filtering can be done for further analysis. Further more, we were provided with the hours of each business is a dictionary form, so we pared the dictionary and created the columns `day`, `open_time`, and `close_time`. The business dataset contained the following columns.

Columns	Definition
name	Name of the business
address	Street address of the business
city	City where the business is located
state	State where the business is located
postal_code	ZIP or postal code
latitude	Geographic latitude coordinate
longitude	Geographic longitude coordinate
stars	Average star rating of the business
review_count	Total number of reviews
is_open	1 = open, 0 = closed
categories	List of categories assigned to the business
hours	Raw opening hours information
business_index	Unique index for each business in the dataset
day	Day of the week extracted from hours
open_time	Parsed opening time for that day
close_time	Parsed closing time for that day
hours_open	Total hours open for that day
categories_grouped	Simplified/grouped main category of the business

Once the categories were finalized in the categories_grouped column, we filtered reviews dataset to only contain the desired restaurants, and performed sentiment analysis on the reviews. The sentiment reviews, gave use 2 columns, one for the label `sentiment_prediction` (very positive, positive, very negative, negative, neutral) and the score of the sentiment accuracy `sentiment_score`. The table below shows some unique columns for the reviews dataset.

Columns	Definition
user_id	Unique ID of the user who wrote the review
stars	Star rating given by the user (1–5)
useful	Number of “useful” votes the review received
funny	Number of “funny” votes the review received
cool	Number of “cool” votes the review received
text	Full written content of the review
date	Date when the review was posted
sentiment_prediction	Predicted sentiment label for the review
sentiment_score	Numerical sentiment score returned by the model

Explanatory Data Analysis

The datasets are massive, and hold a lot of stories about the dining experience Santa Barbara has to offer. We limit the analysis to only the 5 most popular cuisines found in the city. The table below gives us an overview of these categories by looking at average star rating, number of reviews, and how many restaurants fall in each.

Category	Avg Stars	Total Reviews	Num Restaurants
American (New)	3.954	45577	151
Mexican	3.848	22443	102
American (Traditional)	3.619	14771	63
Italian	3.750	10789	40
Japanese	3.636	6915	33

A quick note, American (New) cuisine blends traditional cooking with global influences, modern techniques, and seasonal ingredients, a movement that began reshaping American dining in the 1980s. American (Traditional) is rooted in specific regional traditions shaped by local ingredients and cultural heritage. From Southern barbecue to New England clam chowder, these dishes reflect the flavors of place and history (Oriach 2025). From the table above, American (New) have the highest number of restaurants in the city, while Japanese has the lowest. The table motivated us to look into the population demographics of the City. DATA USA reported in 2023 that the city is predominantly White (Non-Hispanic) (53.4%),Other (Hispanic) (12.6%), Two Races Including Other (Hispanic) (12.1%), White (Hispanic) (9.86%), and Asian (Non-Hispanic) (3.66%). Thus, the low percentages of Asian population could have played part in not having as many Japanese, or other Asian cuisines, in the city. As for average star ratings, American (New) also had the highest average star rating of 3.954, while Japanese has the lowest 3.636. The average star rating of the different cuisines, begs the question of whether what people say (aka reviews) reflect the star ratings we see.

In Figure 1, we have a stacked bar plot that shows the distribution of positive, neutral, and negative reviews across our top categories. Overall, the majority of the reviews for all restaurants are primarily positive. Mexican cuisine has the highest percentage of positive reviews **76.4%**, while Italian has the lowest percentage being **69.3%.**

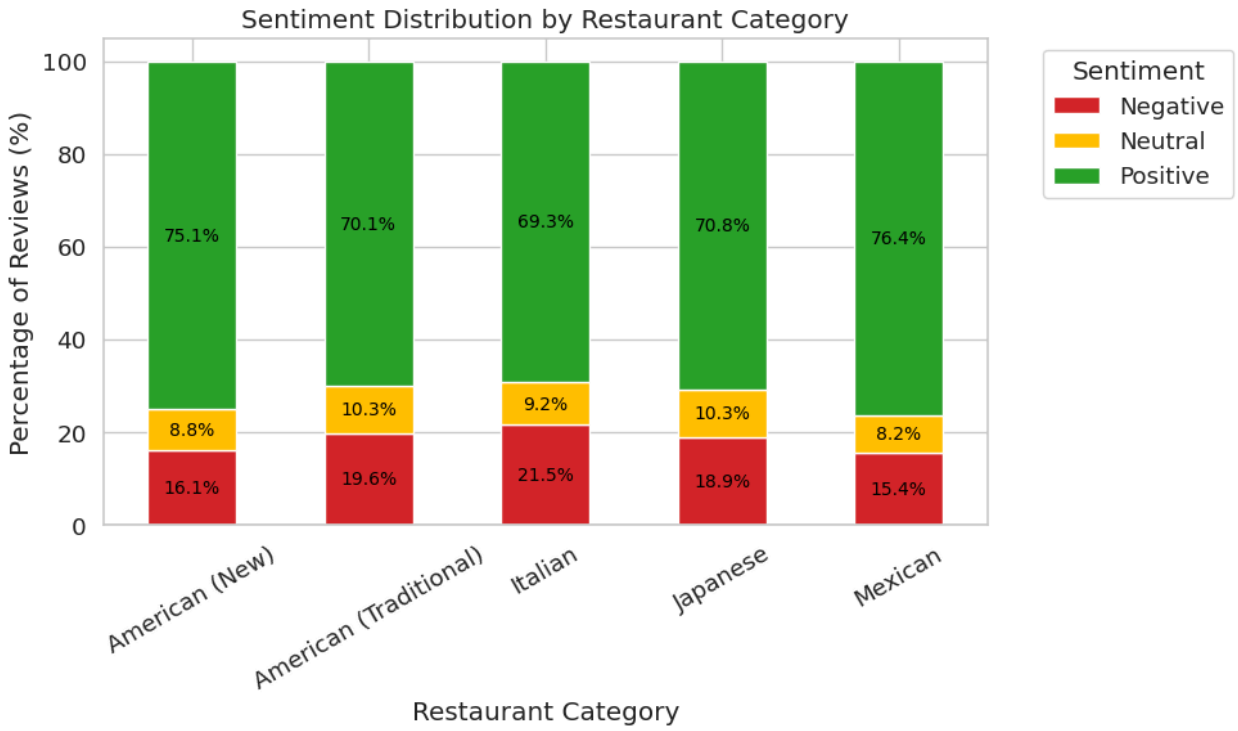

Figure 1

We wanted to further investigate Mexican and Italian cuisines by creating word clouds for their respective reviews. To highlight the positive aspects of Mexican cuisines, we filtered the reviews column to only consider those with 4 or more star rating. As for Italian, we filtered the reviews column to only include those with 2 or lesser star rating. Figure 2 below show the 2 word clouds side by side.

Figure 2

The green plot indicate the positive reviews. Some recurring words as shown in the plot are “delicious”, “friendly”, “great”, “nice”, and multiple food items such as “tacos”. This indicates that people value the friendly staff of these restaurants, the high quality food that is offered. As for the negative reviews, some prominent words are “time”, “ordered”,”server”, indicate that customers feel poorly towards the staff

and the ordering/waiting times. In both plots we can see that people don’t just want good food, but also good service.

Conclusion

This report was a brief analysis of the top 5 categories in Santa Barbara. We looked at some summary stats such as star ratings, and number of each category. We also explored sentiment analysis of the reviews and analyzed what made some categories better than others. This analysis could help those visiting the city to learn more about the different cuisines in the city.

Works Cited:

https://www.kaggle.com/datasets/hirakd/yelp-business-categories-list-complete-dataset

https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset

https://www.pods.com/blog/living-in-santa-barbara

https://www.cookunity.com/blog/new-american-cuisine

https://datausa.io/profile/geo/santa-barbara-ca/?compare=inglewood-ca