Statistical Analysis of Yelp Restaurant Ratings

Prepared for

Dr. Nicolai Amann and Yan-Yu Chen

STA 141B

Contributions

Joshua Wei: Data Scraping & Cleaning

Michelle Lin: Data Processing & Transformations

Darren Lam: Visualizations & Version Control

Abstract
Yelp ratings are commonly used to guide restaurant choice, yet it is unclear whether a given star rating carries the same meaning across different cities, price levels, and cuisines. In this project, we investigate how these contextual factors shape the distribution of Yelp restaurant ratings across the United States. Using data collected from the Yelp Fusion API, we analyze approximately 100,000 restaurants sampled from 10 randomly selected cities in each U.S. state. We examine how ratings vary by price level, cuisine group, geographic location, and chain status through descriptive statistics and visualizations. Our findings show that Yelp ratings are highly compressed overall, but systematic differences emerge across price tiers, cuisine types, and states. These patterns suggest that Yelp star ratings should be interpreted relative to contextual factors rather than as universal measures of restaurant quality.

Reproducibility and Code Availability
All code used for data acquisition, cleaning, transformation, analysis, and visualization is publicly available at: *https://github.com/DarrenTheLamb01/Yelp-Rating-Analysis*

The repository contains all notebooks, processed datasets, and scripts required to reproduce the results presented in this report.

Research Questions
This project focuses on the following research questions:

1. How are Yelp ratings distributed overall across U.S. restaurants?
2. Do Yelp rating distributions differ across price levels?
3. How do ratings vary by cuisine type, and how does cuisine interact with price level?
4. Are there systematic geographic differences in average Yelp ratings across states?
5. Do chain restaurants exhibit different rating patterns compared to non-chain restaurants?

These questions are motivated by concerns that Yelp ratings reflect not only restaurant quality, but also consumer expectations, visibility biases, and contextual norms.

Introduction
Yelp reviews are fairly useful for the general public, but we believe that they are not a clean and accurate representation of true restaurant quality. First, they are written by a very specific group of customers, which are people who feel strongly enough to leave a review. These reviews are shaped by things like expectations, the internet, and even social influence so a star rating at a cheap and casual spot does not mean the same thing as the same rating at an omakase restaurant. On top of that, Yelp's ranking and filtering algorithms decide which restaurants are most visible, which adds another layer of bias. Because of all these factors, the three of us decided to focus our STA 141B final project on this issue. In particular, we wanted to investigate how city, price level, and cuisine influence Yelp rating distributions for restaurants. In the rest of this report, we first

describe our data sources and scraping process, then explain how we cleaned and transformed the raw data into an analysis ready dataframe. We then present visualizations and summaries that show how ratings are distributed across our three factors. Finally, we discuss what these patterns suggest about how to interpret Yelp stars and the limitations of using Yelp reviews as a measure of true restaurant quality.

Data Acquisition

For data acquisition, we used the Yelp Fusion API to collect restaurant data across a random set of cities in each of the 50 states We wrote a Python scraper using the requests library that authenticated with our API key, queried the correct endpoint for the restaurants category, and handled scraping which page with the limit and offset parameters to pull up to roughly 200 restaurants per city. To make our code as robust as possible, we added basic error handling as well as introduced small delays between calls to respect Yelp's rate limits. The raw responses arrived in nested JSON format, so we stored them locally as a single JSON file in a data/raw directory, keeping the original JSON format separate from our cleaned version, where we stored in a data/processed directory for efficient reuse without re-scraping.

Data Processing/Cleaning

The main areas we wanted to process and transform are:

1. Convert Columns: Change certain columns into different types for analyzation
2. Chain Restaurant: Identifying if restaurants are chains
3. Cuisines: Identifying and categorizing restaurants into major cuisine categories

The first area of the data we wanted to process was converting the price, categories, transactions, and categories_alias columns into data types that could be graphed and visualized. For the price column, we created a new column that converts a string into a float. For example, if a restaurant is rated $$$, the new column would contain a value of 3.0. The other three columns are all originally strings and must be converted to Python lists as analysis on these columns can only be done on lists, not strings.

The next question we wanted to address is whether or not these restaurants are chains. We did this by creating a new column called "is_chain" that outputs True or False values. True values mean that the restaurant is a chain and false values indicate that the restaurant is not a chain.

Lastly, the most important part of the data we wanted to process was categorizing the cuisines. As the categories column contains several elements, we created a column called "primary_cuisine" that takes the first category in each list as the primary cuisine. Next, we created a column called "primary_cuisine_clean" that standardizes the primary cuisine, removing capitalized letters, plural form, and more from the cuisines with the most inconsistencies. Then, we mapped all the cuisines into 12 major categories: American, Asian, European, Latin American, Middle Eastern, African, Dessert/Bakery, Cafe/Coffee, Bar/Pub, Seafood, Vegetarian/Vegan, and Fast Food. The cuisines that weren't successfully mapped fell into the Other category. To identify certain cuisines types that may have been missed, we ran code to identify the most common cuisines that ended up in Other and remapped them. This process was done several times to include as many cuisines as possible. Another issue we ran into was that some restaurants were categorized as Delis, Breakfast/Brunch, and more, but those aren't categorizable cuisines, so we needed to find a secondary cuisine. We ran code to find a cuisine that better fits the mapping for those restaurants in the Other category and reduced the number of Other restaurants from over 7000 to 3579. With the data processed and transformed, we were able to continue exploratory analysis and visualization.

Exploratory Analysis & Visualization

Overall Distribution of Yelp Ratings

Figure 1 shows the overall distribution of Yelp ratings across all restaurants. Ratings are heavily concentrated between 3.5 and 4.5 stars, with a median slightly above 4. This compression indicates that Yelp ratings do not span the full 1–5 scale evenly. As a result, relatively small numerical differences in ratings may correspond to meaningful perceived differences in restaurant quality.

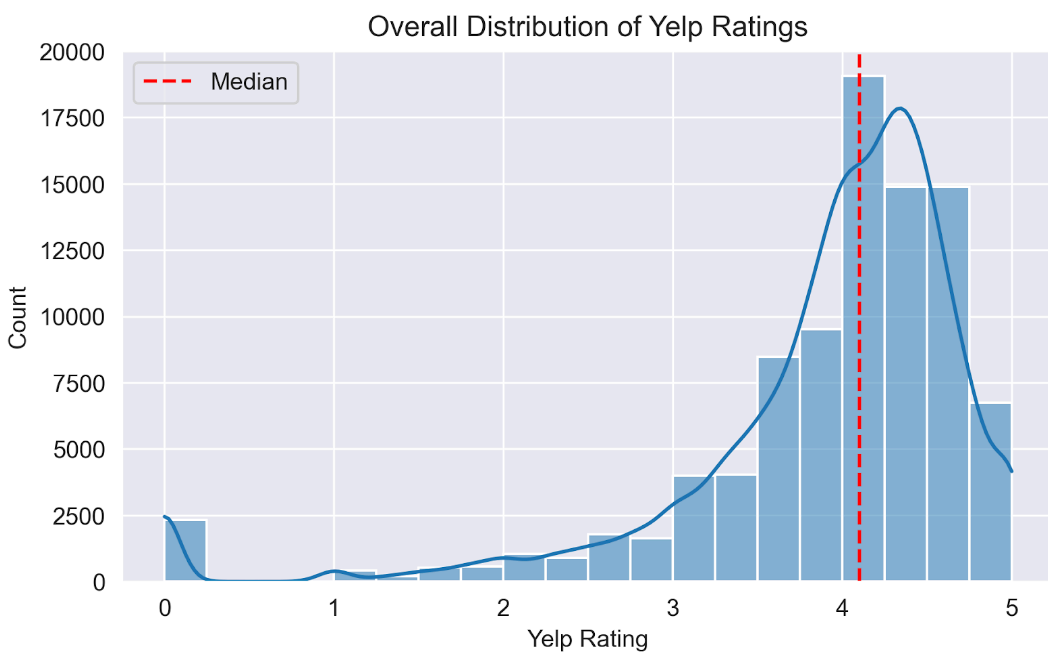

Figure 1. Overall distribution of Yelp ratings across all sampled restaurants.

This distribution motivates the need to examine ratings relative to contextual factors such as price and cuisine rather than interpreting them in isolation.

Yelp Ratings by Price Level

Figure 2 presents boxplots of Yelp ratings by price level. Median ratings increase modestly from $ to $$$, suggesting that higher-priced restaurants tend to receive slightly higher ratings. However, the most expensive category ($$$$) exhibits greater variability and a lower lower quartile, indicating that high-end restaurants may be more polarizing or subject to higher customer expectations.

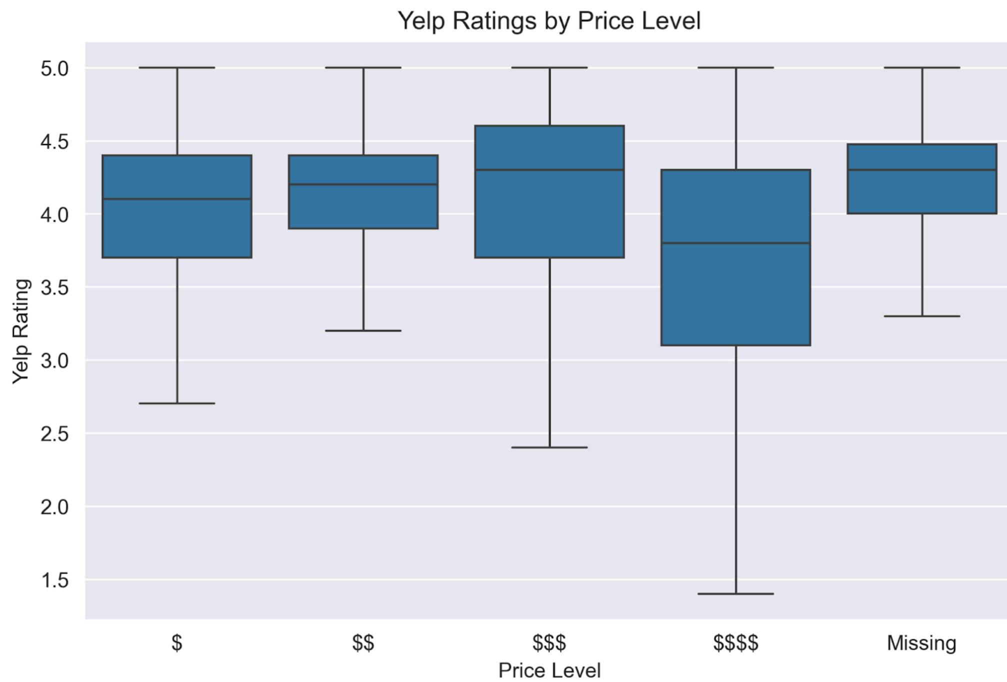

Figure 2. Yelp rating distributions by price level.

Restaurants with missing price information display a wide distribution, underscoring the importance of explicitly accounting for missingness rather than discarding these observations.

Yelp Rating by Cuisine Group

Figure 3 compares rating distributions across the 10 most common cuisine groups. While median ratings are broadly similar across cuisines, several notable differences emerge. Fast food restaurants show both lower medians and greater spread, while dessert and café-oriented establishments tend to receive higher and more consistent ratings.

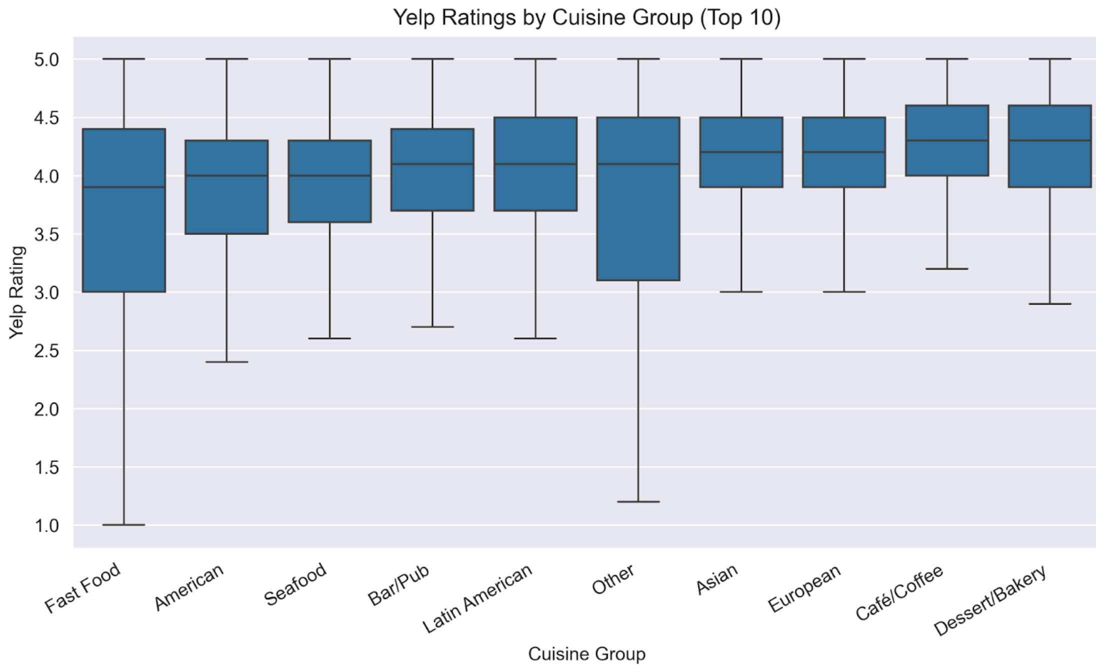

Figure 3. Yelp ratings by cuisine group for the ten most frequent cuisines.

These patterns suggest that customer expectations differ substantially by cuisine type, reinforcing the idea that Yelp ratings are context-dependent.

Interaction Between Cuisine and Price Level

Figure 4 displays a heatmap of mean Yelp ratings by cuisine group and price level. This visualization highlights interaction effects that are not visible when examining price or cuisine alone. For example, higher price levels are associated with higher ratings for some cuisines (e.g., Asian, European), but not for others (e.g., Fast Food).

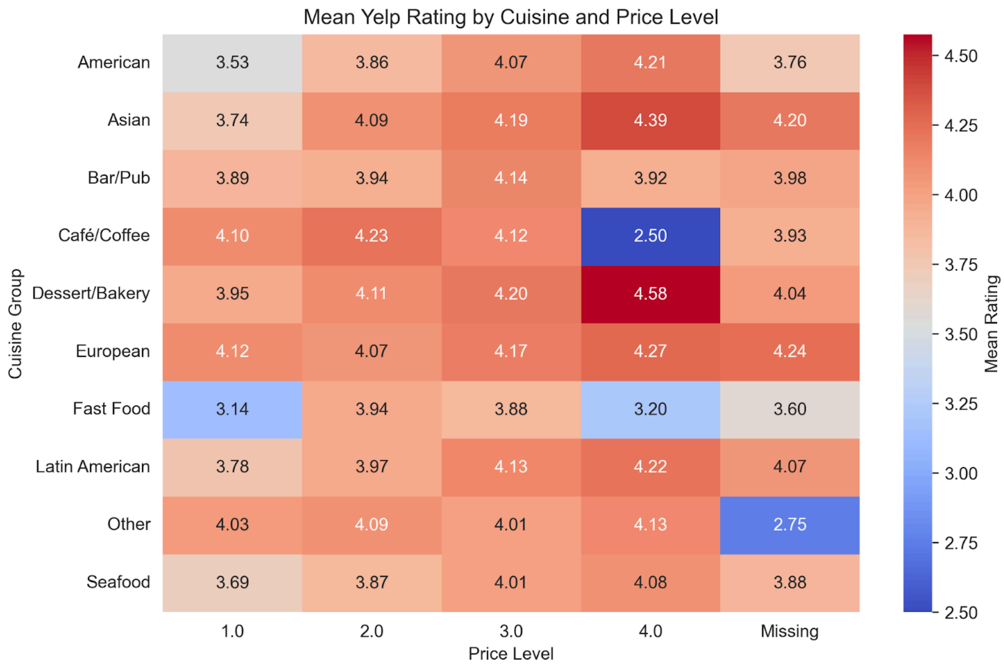

Figure 4. Mean Yelp rating by cuisine group and price level.

This interaction underscores that price does not have a uniform meaning across cuisines. A $$$$ rating in one cuisine category may reflect a very different dining experience and expectation set than in another.

Geographic Variation in Yelp Ratings

Figure 5 shows the mean Yelp rating by state. While state-level averages vary, the overall range is relatively narrow, reflecting the compressed nature of Yelp ratings. Nonetheless, consistent regional differences appear, suggesting that local reviewing norms, restaurant density, or cultural factors may influence how users assign ratings.

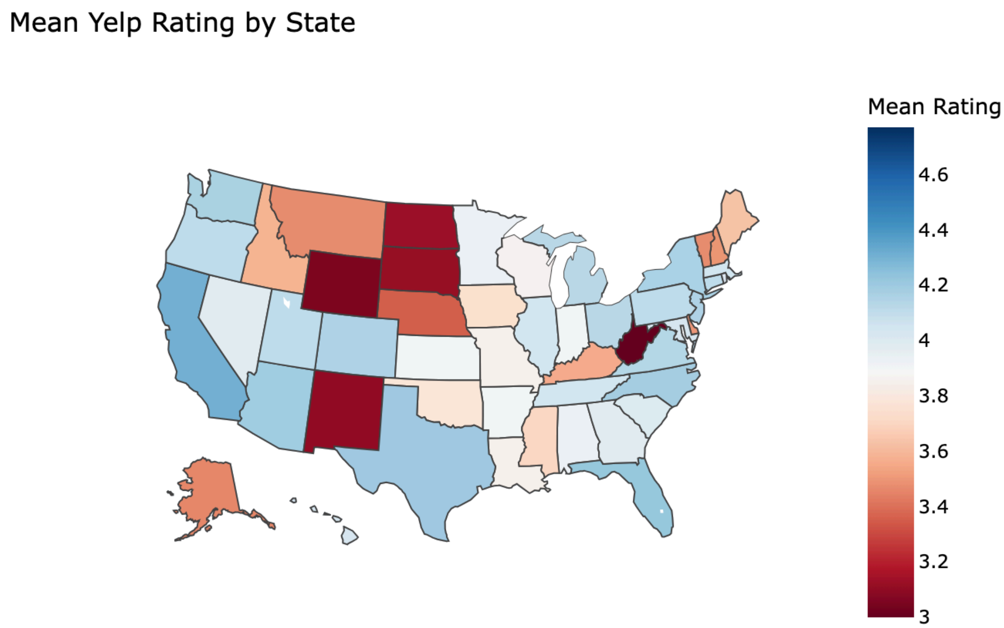

Figure 5. Mean Yelp rating by U.S. state.

These geographic patterns indicate that a given star rating may not be directly comparable across locations.

Chain vs. Non-chain Restaurants

Figure 6 compares rating distributions between chain and non-chain restaurants. Chain restaurants exhibit slightly higher medians and narrower interquartile ranges, suggesting more standardized customer experiences. In contrast, non-chain restaurants display greater variability, with both higher highs and lower lows.

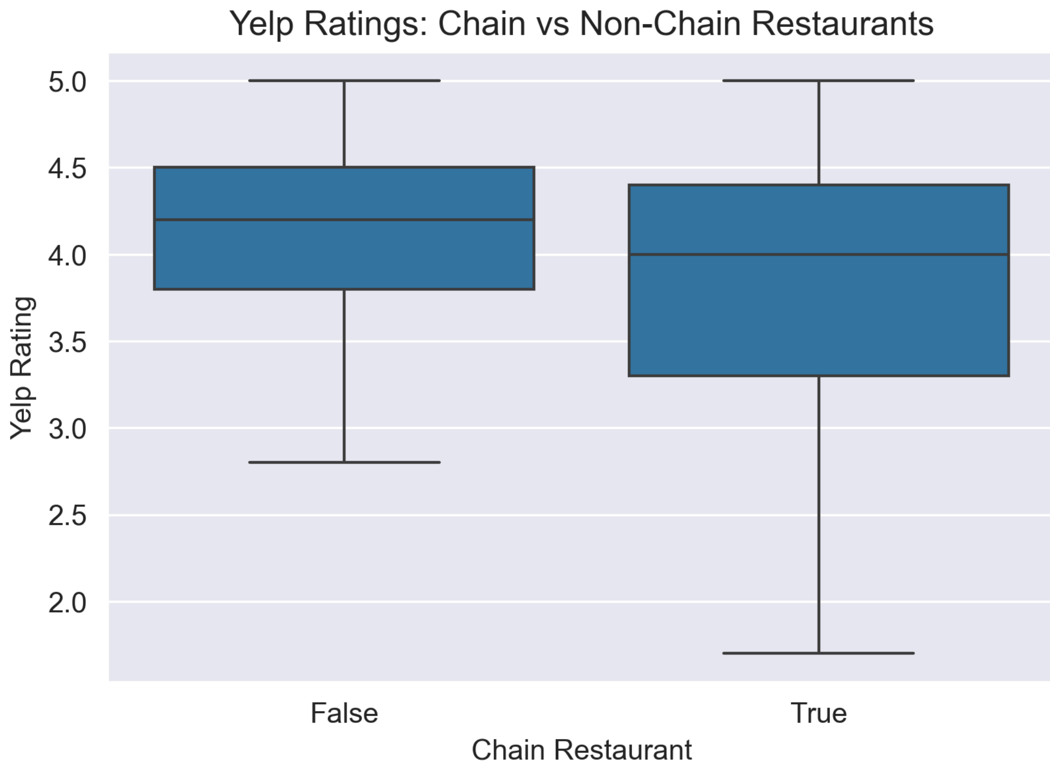

Figure 6. Yelp ratings for chain versus non-chain restaurants.

This finding aligns with the idea that independent restaurants face more heterogeneous expectations and experiences, which are reflected in their ratings.

Discussion

Across all analyses, a consistent theme emerges: Yelp ratings are not absolute measures of restaurant quality. Instead, they reflect a combination of quality, expectations, and contextual norms. Price level, cuisine type, geographic location, and chain status all shape how users assign ratings.

Importantly, many of the observed differences are modest in magnitude but meaningful given the compressed rating scale. This suggests that users and analysts should be cautious when comparing restaurants across different contexts using raw star ratings alone.

Limitations

This analysis is descriptive and does not establish causal relationships. Review counts and Yelp's internal ranking algorithms may introduce visibility and selection biases that are not fully captured here. Additionally, although restaurants were sampled broadly across the U.S., the use of fixed numbers of cities per state may not fully represent each state's restaurant landscape.

Future work could incorporate review text, temporal trends, or formal statistical modeling to better isolate drivers of rating variation.

Conclusion

Yelp ratings are widely used but context-dependent. Our analysis shows that ratings vary systematically across price levels, cuisine types, geographic regions, and chain status. These findings suggest that Yelp star ratings should be interpreted relative to similar restaurants rather than as universal indicators of quality. By highlighting these patterns, this project demonstrates the importance of contextualized data interpretation when working with large-scale consumer review data.

References

Yelp. (n.d.). Getting started with the Yelp Places API. Yelp Developers. https://docs.developer.yelp.com/docs/places-intro