### SEC I – Executive Summary

This report leverages a deep statistical analysis of the Yelp dataset to provide a data-driven blueprint for launching a successful U.S. restaurant. The core finding is that customers overwhelmingly reward specialized, atmospheric dining experiences over generic, convenience-focused options.

*Our final recommendation is to open a classy and intimate Café or a high-end Dessert Bar in Portland, Oregon.*

This strategy is rooted in key data insights: Portland is a reliably high-rated market; specialized concepts like Cafes are statistically rewarded while Fast Food is heavily penalized; and a restaurant's ambience is the single most powerful driver of 5-star ratings. Additionally, focusing on and encouraging detailed, story-driven reviews will solidify the restaurant's top-tier reputation.

### SEC II – Introduction

Launching a new restaurant is a significant undertaking where early decisions can make the difference between a thriving business and a failed venture. To navigate the competitive U.S. market, this project moves beyond guesswork and instead uses a deep analysis of real-world customer data from the popular review platform, Yelp.

The objective is to create a data-driven blueprint for a successful new restaurant. By examining hundreds of thousands of reviews and restaurant profiles, this report answers the fundamental questions every new restaurant owner must face:

* **Where** is the most promising market to open a new restaurant?
* **What** type of restaurant concept and cuisine is most likely to succeed?
* **Which** facilities and services are essential for earning high ratings?
* **Why** do some restaurants get 5-star reviews while others struggle?
* **How** can we build a strong online reputation by encouraging positive, helpful reviews?

In addition, this analysis also aims to provide clear, actionable recommendations to the restaurant owner offering a strategic advantage in a challenging industry.

### SEC III – Analysis & Findings

**III. A. Analysis Methodology**

The recommendations in this report are derived from a rigorous statistical analysis of a large Yelp dataset, containing hundreds of thousands of restaurant profiles, user reviews, and ratings. To move beyond simple averages and uncover the true drivers of success, the following statistical techniques were employed:

* **Regression Modelling**: Used to identify and quantify the specific factors that have the most significant impact on a restaurant's star rating. This allowed us to understand the precise effect of attributes like ambience, price, and cuisine type while controlling for all other variables.
* **Hypothesis Testing (T-tests)**: Employed to confidently determine if offering a specific facility (e.g., reservations, valet parking) results in a statistically significant difference in average ratings compared to restaurants that do not offer it.
* **Confidence Intervals**: Used to identify the most reliably high-rated cities. This method considers both the average rating and the volume of reviews to find markets where success is not just a possibility, but a consistent pattern.

This multi-faceted approach ensures that our findings are not based on surface-level observations but on statistically significant relationships within the data.

**III. B. Data Analysis**

For the purpose of our analysis, we aim to answer five fundamental questions that every new restaurant owner must answer before entering the industry.

Following are the assumptions we have made for the purpose of this analysis:

* Confidence interval - 95%
* Margin of error - 5%
* z score = 1.96
* Level of Significance - 0.05

The questions, analysis and recommendations are as follows:

1. **Where** is the most promising market to open a new restaurant?

Rationale- To determine the best location for a new restaurant, we performed a statistical analysis to identify cities that are not just highly-rated, but reliably so. Instead of relying on a simple average, which can be skewed by a small number of reviews, we calculated the 95% confidence interval for the mean star rating in every city with a substantial number of restaurants.

By ranking cities based on the lower bound of this interval, we can identify markets where we are 95% confident that the true average rating is above a certain high threshold. This method prioritizes consistency and market health.

**Key Findings from the Data:**

The analysis identified five strong candidate cities:

|  |  |  |
| --- | --- | --- |
| **Top 5 Recommended Cities Based on 95% Confidence Interval of Mean Rating** | | |
| **City** | **State** | **CI\_lower\_bound** |
| Portland | OR | 3.806264 |
| Winter Park | FL | 3.675592 |
| Austin | TX | 3.654609 |
| Winthrop | MA | 3.636165 |
| Somerville | MA | 3.622077 |

|  |  |  |
| --- | --- | --- |
| **Final choice based on lower bound of 95% Confidence Interval of mean rating** | | |
| City | State | CI\_lower\_bound |
| Portland | OR | 3.806264 |

**Recommendation: Portland, Oregon**

The data clearly indicates that Portland, Oregon, is the top-recommended location.

We can be 95% confident that the true average restaurant rating in Portland is at least 3.81 stars, a significantly high benchmark. This suggests a mature and appreciative customer base with a strong culture of dining out and providing positive feedback, making it an ideal market to enter. Other strong candidates for consideration include Winter Park, Florida, and Austin, Texas.

1. **What** type of restaurant concept and cuisine is most likely to succeed?

Rationale- To answer this, we built a comprehensive regression model and analysed the data to isolate the impact of specific features—from cuisine type to ambiance and parking—on a restaurant's final star rating. This statistical technique highlights the factors that are most statistically significant and have the largest effect, providing a clear blueprint for success.

Our analysis reveals that the most successful restaurants prioritize a specific **atmosphere** and a **specialized offering** over a broad, generic approach.

**Key Findings from the Data:**

Our model identified the following attributes and cuisine types as having the most significant impact on star ratings. A positive "Effect on Stars" means the feature is associated with higher ratings, while a negative value is associated with lower ratings.

**Most Impactful Attributes:**

|  |  |
| --- | --- |
| **Attribute** | **Effect on Stars (Coefficient)** |
| ambience\_intimate | 0.231187 |
| businessparking\_street | 0.169919 |
| goodformeal\_latenight | -0.168973 |
| ambience\_classy | 0.151641 |
| ambience\_hipster | 0.123561 |

**Most Impactful Cuisine Categories:**

|  |  |
| --- | --- |
| **Cuisine Category** | **Effect on Stars (Coefficient)** |
| **Positive Impact** | |
| cat\_cafes | 0.189775 |
| cat\_caterers | 0.183327 |
| cat\_desserts | 0.176797 |
| **Negative Impact** | |
| cat\_fast\_food | -0.646272 |

**Interpretation & Recommendation**

The data points to a clear strategic direction.

* Atmosphere is Paramount: The most powerful positive drivers are related to ambiance. An intimate, classy, or hipster atmosphere provides a significant boost to ratings, indicating that customers reward a curated dining experience. Conversely, being known primarily as a late-night spot is a negative factor.
* Specialization Wins: The model shows a massive penalty for restaurants in the "Fast Food" category. In contrast, the concepts with the strongest positive association are specialized venues like "Cafes" and "Dessert" bars.

From the analysis, the ideal concept is a specialized establishment, such as a classy and intimate café or a high-end dessert bar. These establishments would leverage the powerful positive drivers of a great atmosphere and a focused, high-quality menu, while strategically avoiding the categories most heavily penalized by Yelp reviewers.

1. **Which** facilities and services are essential for earning high ratings?

Rationale- Here, we have used hypothesis testing (specifically, an independent samples t-test) to determine if offering a specific facility results in a statistically significant difference in star ratings. A low p-value (typically < 0.05) tells us the facility matters.

***Null Hypothesis (H0)****: The mean star rating of restaurants with the facility is the same as or less than the mean rating of those without it.*

***Alternative Hypothesis (Ha)****: The mean star rating of restaurants with the facility is greater than the mean rating of those without it.*

If the resulting p-value is less than 0.05, we reject the null hypothesis and conclude that the facility has a statistically significant positive impact.

**Key Findings from the Data:**

To identify which facilities, provide the best return on investment, a series of hypothesis tests (t-tests) were conducted. This analysis determined which amenities have a statistically significant positive association with higher star ratings, separating the "must-haves" from the "nice-to-haves."

The results provide a hierarchy of importance for the facilities a new restaurant should offer.

The analysis sorted facilities into two groups: those with a significant positive impact on ratings and those without.

|  |  |  |
| --- | --- | --- |
| **Facility** | **Impact on Star Ratings** | **Statistical Significance (P-value)** |
| **High-Impact Facilities** | | |
| Reservations | Positive | < 0.0001 |
| Outdoor Seating | Positive | < 0.0001 |
| WiFi | Positive | < 0.0001 |
| Caters | Positive | < 0.0001 |
| Bike Parking | Positive | < 0.0001 |
| **Lower-Impact Amenities** | | |
| Good For Kids | Not Significant | 0.0792 |
| Smoking | Not Significant | 1 |
| Wheelchair Accessible | Not Significant | 1 |
| Delivery | Not Significant | 1 |
| Take-Out | Not Significant | 1 |

**Interpretation & Recommendation**

The data indicates that customers of highly-rated restaurants value services that enhance the dining experience and align with modern lifestyle trends.

* Must-Have Facilities: Offering Reservations, Outdoor Seating, and WiFi are statistically proven drivers of higher ratings. These should be considered essential. Providing Catering services and Bike Parking are also strongly associated with positive reviews and represent key opportunities to attract a high-value customer base.
* Lower Priority Amenities: While services like Take-Out and Delivery are common, the data shows they do not correlate with higher ratings; they are viewed as convenience features, not markers of quality. Similarly, being "Good For Kids" or providing a smoking area are not statistically significant drivers of a better Yelp score. While Wheelchair Accessibility is an ethical and often legal necessity, it does not function as a differentiator for higher ratings in this dataset.

The strategic focus should be on implementing the five high-impact facilities to build a strong foundation for positive customer ratings.

1. **Why** do some restaurants get 5-star reviews while others struggle?

Rationale- To identify the key drivers of success, a comprehensive regression model was built to analyze hundreds of restaurant features simultaneously. This statistical technique isolates the most influential factors by measuring the impact of each one while holding all others constant. The results below highlight the features with the largest and most statistically significant effects. The features with the largest (positive or negative) coefficients that are also statistically significant (p-value < 0.05) are the most important.

**Key Findings from the Data:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Analysis for: Simple Facilities** | | | |
| **Feature** | **Coefficient** | **P-value** | **Interpretation** |
| price\_range | 0.0148 | 0.0231 | Statistically significant. Its presence is associated with an increase of 0.015 stars. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Analysis for: Ambience Features** | | | |
| **Feature** | **Coefficient** | **P-value** | **Interpretation** |
| ambience\_intimate | 0.2068 | 0 | Statistically significant. Its presence is associated with an increase of 0.207 stars. |
| ambience\_hipster | 0.1359 | 0 | Statistically significant. Its presence is associated with an increase of 0.136 stars. |
| ambience\_classy | 0.1312 | 0 | Statistically significant. Its presence is associated with an increase of 0.131 stars. |
| ambience\_touristy | -0.1058 | 0.0021 | Statistically significant. Its presence is associated with an decrease of 0.106 stars. |
| ambience\_upscale | 0.1055 | 0.0024 | Statistically significant. Its presence is associated with an increase of 0.106 stars. |
| ambience\_trendy | 0.0708 | 0 | Statistically significant. Its presence is associated with an increase of 0.071 stars. |
| ambience\_divey | 0.0674 | 0.0036 | Statistically significant. Its presence is associated with an increase of 0.067 stars. |
| ambience\_romantic | 0.0471 | 0.1344 | Not statistically significant. |
| ambience\_casual | 0.0238 | 0.003 | Statistically significant. Its presence is associated with an increase of 0.024 stars. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Analysis for: Business Parking Features** | | | |
| **Feature** | **Coefficient** | **P-value** | **Interpretation** |
| businessparking\_street | 0.1816 | 0 | Statistically significant. Its presence is associated with an increase of 0.182 stars. |
| businessparking\_garage | -0.0749 | 0 | Statistically significant. Its presence is associated with a decrease of 0.075 stars. |
| businessparking\_valet | -0.0679 | 0.0005 | Statistically significant. Its presence is associated with a decrease of 0.068 stars. |
| businessparking\_lot | 0.0366 | 0 | Statistically significant. Its presence is associated with an increase of 0.037 stars. |
| businessparking\_validated | 0.0351 | 0.1774 | Not statistically significant. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Analysis for: Good for Meal Features** | | | |
| **Feature** | **Coefficient** | **P-value** | **Interpretation** |
| goodformeal\_latenight | -0.1634 | 0 | Statistically significant. Its presence is associated with a decrease of 0.163 stars. |
| goodformeal\_lunch | 0.1018 | 0 | Statistically significant. Its presence is associated with an increase of 0.102 stars. |
| goodformeal\_dinner | 0.0776 | 0 | Statistically significant. Its presence is associated with an increase of 0.078 stars. |
| goodformeal\_breakfast | -0.0533 | 0.0011 | Statistically significant. Its presence is associated with a decrease of 0.053 stars. |
| goodformeal\_brunch | 0.0428 | 0.0105 | Statistically significant. Its presence is associated with an increase of 0.043 stars. |
| goodformeal\_dessert | -0.0427 | 0.0193 | Statistically significant. Its presence is associated with an decrease of 0.043 stars. |
| **Analysis for: Cuisine Categories** | | | |
| **Feature** | **Coefficient** | **P-value** | **Interpretation** |
| cat\_fast\_food | -0.696 | 0 | Statistically significant. Its presence is associated with a decrease of 0.696 stars. |
| cat\_chinese | -0.2884 | 0 | Statistically significant. Its presence is associated with a decrease of 0.288 stars. |
| cat\_chicken\_wings | -0.2833 | 0 | Statistically significant. Its presence is associated with a decrease of 0.283 stars. |
| cat\_burgers | -0.2422 | 0 | Statistically significant. Its presence is associated with a decrease of 0.242 stars. |
| cat\_pizza | -0.2412 | 0 | Statistically significant. Its presence is associated with a decrease of 0.241 stars. |
| cat\_cafes | 0.1897 | 0 | Statistically significant. Its presence is associated with an increase of 0.19 stars. |
| cat\_caterers | 0.1831 | 0 | Statistically significant. Its presence is associated with an increase of 0.183 stars. |
| cat\_desserts | 0.1771 | 0 | Statistically significant. Its presence is associated with an increase of 0.177 stars. |
| cat\_american\_traditional | -0.1664 | 0 | Statistically significant. Its presence is associated with a decrease of 0.166 stars. |
| cat\_cocktail\_bars | 0.1339 | 0 | Statistically significant. Its presence is associated with an increase of 0.134 stars. |
| cat\_coffee\_n\_tea | -0.117 | 0 | Statistically significant. Its presence is associated with a decrease of 0.117 stars. |
| cat\_delis | 0.1102 | 0 | Statistically significant. Its presence is associated with an increase of 0.11 stars. |
| cat\_asian\_fusion | 0.1017 | 0 | Statistically significant. Its presence is associated with an increase of 0.102 stars. |
| cat\_sushi\_bars | -0.0643 | 0.0009 | Statistically significant. Its presence is associated with a decrease of 0.064 stars. |
| cat\_breakfast\_n\_brunch | -0.0372 | 0.001 | Statistically significant. Its presence is associated with a decrease of 0.037 stars. |
| cat\_mexican | -0.0287 | 0.016 | Statistically significant. Its presence is associated with a decrease of 0.029 stars. |
| cat\_italian | -0.0255 | 0.0513 | Not statistically significant. |
| cat\_japanese | -0.0196 | 0.2556 | Not statistically significant. |
| cat\_bakeries | 0.0182 | 0.2758 | Not statistically significant. |
| cat\_american\_new | -0.0144 | 0.1749 | Not statistically significant. |
| cat\_seafood | 0.0139 | 0.3026 | Not statistically significant. |
| cat\_salad | 0.008 | 0.5679 | Not statistically significant. |

**Interpretation & Recommendations:**

The analysis reveals that Yelp ratings are most influenced by two overarching themes: the overall experience and atmosphere, and the restaurant's positioning.

* Experience Over Everything: The most powerful positive factors are related to the overall experience. A restaurant with an intimate, classy, or hipster ambience is rated significantly higher. This proves that customers are rating the entire vibe and experience, not just the food.
* The "Kiss of Death" Categories: The model's strongest signal is that certain categories are overwhelmingly associated with lower ratings. Being a Fast-Food restaurant is the single most damaging factor. Similarly, concepts centred around common, convenience-driven items like Pizza, Burgers, and Chicken Wings are heavily penalized in the ratings.
* Specialization is Rewarded: The concepts that are most positively rewarded are highly specialized. Cafes, Dessert shops, and Cocktail Bars all have a significant positive impact. This indicates that a focused, high-quality offering is a key strategy for achieving a high rating.

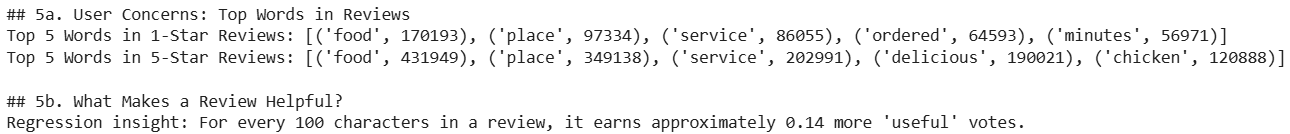
The most important factors for a high Yelp rating are: creating a special, atmospheric experience and avoiding association with generic, convenience-focused food categories.

1. **How** can we build a strong online reputation by encouraging positive, helpful reviews?

Rationale- For the purpose of answering this we have used a two-pronged approach.

* First (a), we analysed the most common words in 5-star vs. 1-star reviews to see what customers prioritize.
* Second(b), we used a regression model to confirm the link between review length and helpfulness ('useful' votes).

**Key Findings from the Data:**

****

1. A word frequency analysis reveals a clear distinction between what drives praise and what drives complaints:

* 1-Star Reviews (Complaints): When customers are unhappy, they focus on product, service failures and operational issues. After filtering out common words, the most frequent topics are related to place (97334), service (86055), ordered (64593) and minutes (56971). They also focus on food quality issues with words like food (170193). This shows that a negative experience is often rooted in poor product, service and execution.
* 5-Star Reviews (Praise): Happy customers focus on the experience and the food itself. The most common themes are overwhelmingly positive food (431949), place (349138), service (202991) and delicious (190021). They also tend to mention specific menu items they enjoyed like chicken (120888), indicating that a standout, memorable dish is a powerful driver of positive reviews.

1. What Makes a Review "Helpful"?

The regression analysis provides a clear, quantitative answer to what the Yelp community values in a review: detail and description.

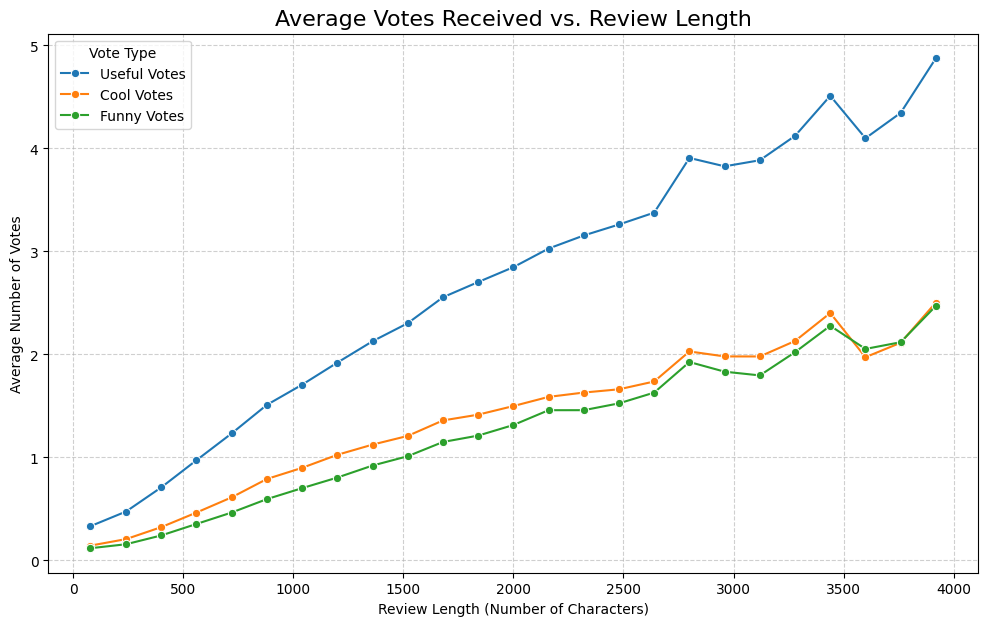
Key Insight: For every 100 characters a customer adds to their review, it is expected to earn approximately 0.14 more "useful" votes.

This shows that short, generic reviews like "It was good" have little impact. Longer, more descriptive reviews that tell a story about the experience—the food, the service, the atmosphere—are seen as more credible and helpful, which in turn boosts the restaurant's reputation.

### SEC IV – Strategic Recommendations

Similar to how we analysed the fundamental questions, we analysed a few strategic questions that the restaurant owner must make. They are as follows:

1. Are longer reviews associated with more useful/funny/cool votes?

A blue squares with white text

AI-generated content may be incorrect.

A screen shot of a graph

AI-generated content may be incorrect.

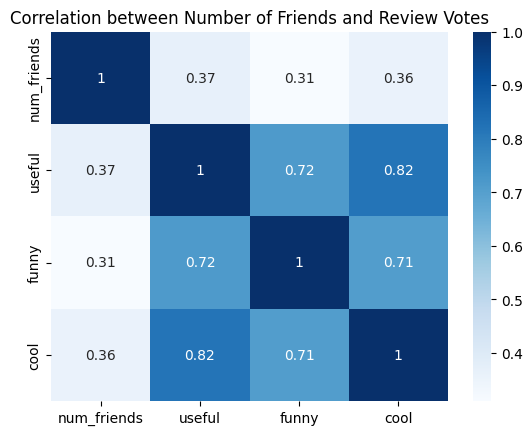
**Interpretation:** The data confirms a strong positive association between review length and votes. The line plot shows a clear upward trend for all vote types, and the correlation matrix reveals the strongest relationship is between length and "useful" votes (+0.27).

**Insight:** The Yelp community values and rewards detailed, descriptive feedback. While most reviews are short, helpfulness and credibility are directly linked to the amount of detail a reviewer provides.

**Strategic Recommendation:** To boost your Yelp page's credibility, actively encourage customers to write detailed reviews. Use specific prompts like, "What made your visit special?" to elicit longer, descriptive feedback that earns more "useful" votes.

1. Are reviews by users with more friends more likely to get votes?

A graph of a graph showing different colored lines

AI-generated content may be incorrect.

**Interpretation:** The data shows a strong, positive relationship between a user's friend count and the votes their reviews get. The line plot reveals this trend is not linear but accelerates dramatically, with reviews from users having 500+ friends receiving nearly 10 times more "useful" votes on average than users with no friends.

**Insight:** A user's friend count is a direct indicator of their influence on Yelp. A positive review from a well-connected user has significantly greater visibility and credibility within the community.

**Strategic Recommendation:** Prioritize engaging with "Yelp Influencers" (users with many friends). A prompt, professional response to their reviews can amplify positive exposure and yield a much higher return on all engagement efforts.

1. What are the busiest times?

A chart of different colors

AI-generated content may be incorrect.

**Interpretation:** The heatmap clearly shows two daily peaks in customer activity: a lunch rush from 12 PM to 2 PM and a much stronger dinner peak from 6 PM to 12 AM. The brightest colours reveal that the absolute busiest period of the week is the weekend, specifically Friday, Saturday, and Sunday evenings (continuing all the way into early Monday).

**Insight:** Customer dining and reviewing habits are highly predictable. The weekend dinner service is the most critical period for a restaurant's reputation, as this is when the highest volume of customers are forming their opinions.

**Strategic Recommendation:** Align staffing, inventory, and marketing efforts with these peaks. Ensure maximum staff coverage during weekend evenings to provide excellent service. Run promotions or special menu items during these high-traffic times to make the biggest impact and generate the most positive reviews.

1. How has the average review rating changed over time by state?

A graph of different colored lines

AI-generated content may be incorrect.

**Interpretation:** The plot shows a general upward trend in average star ratings from 2010 to 2021 across most states, suggesting either grade inflation or an overall improvement in restaurant quality. Drawing on the inference of the top three locations for the new restaurant, based on their respective states, Oregon (OR) has consistently been a high performer. Texas (TX) shows a steady and strong increase over the decade. Florida (FL) is comparatively a little more volatile, with a spike in 2021 before a drop.

**Insight:** While the national trend is positive, local market dynamics can be turbulent. A state with stable, high ratings like Oregon represents a mature and reliable market. A state with a sharp recent spike like Florida might be experiencing a temporary boom, making it a riskier, more unpredictable market.

**Strategic Recommendation:** For long-term stability, prioritize Oregon or Texas for the new restaurant. These states demonstrate consistent high performance and positive customer sentiment. Florida's recent volatility, despite its high peak, suggests a less predictable market that may be harder to navigate for a new business.

1. Are elite users more generous or harsher reviewers?

A close-up of a computer screen

AI-generated content may be incorrect.A chart of a distribution of stars

AI-generated content may be incorrect.

**Interpretation:** The hypothesis test (p < 0.0001) and box plot both prove that Elite users are not harsher. In fact, they are significantly more generous, with a higher median average rating (~4.0) compared to regular users (~3.8).

**Insight:** Elite users are a more consistent and predictable group of reviewers who are less likely to give extremely low ratings. They are enthusiasts who tend to reward positive experiences, rather than being overly critical.

**Strategic Recommendation:** Actively welcome and engage with Elite reviewers. A positive review from an Elite user is highly credible and comes from a user group that is statistically more generous and consistent than the average customer, making their feedback particularly valuable.

### SEC V – Conclusion

This analysis has systematically deconstructed the key factors that drive restaurant success on Yelp, moving beyond anecdotal evidence to create a data-driven way forward. The findings consistently point to a powerful central theme: today's customers reward specialized, atmospheric dining experiences over generic, convenience-focused options.

By leveraging statistical analysis, we have transformed the complex data into a clear, actionable strategy. The final recommendations are as follows:

1. **Location**: Launch in a stable, high-performing market like Portland, Oregon, where a strong culture of positive dining reviews provides a high floor for success.
2. **Concept**: The ideal business model is a classy and intimate Café or a specialized Dessert Bar. This concept focuses on a high-quality, niche offering and directly taps into the cuisine categories with the most significant positive impact on ratings, while avoiding the heavily penalized "Fast Food" category.
3. **Facilities & Atmosphere**: The primary investment should be in creating an exceptional ambience. An "intimate" and "classy" feel is the most powerful driver of high ratings. This should be complemented by essential, experience-enhancing facilities like offering reservations, providing WiFi, and having accessible street parking.
4. **Customer Engagement**: Success on Yelp is driven by detail. Encourage longer, story-driven reviews by prompting customers for specific feedback. Do not be intimidated by "Elite" reviewers; the data shows they are statistically more generous and consistent than the average user, making their positive feedback exceptionally valuable.

By adhering to this data-backed strategy—focusing on a specialized concept in a proven market, prioritizing a curated atmosphere, and encouraging detailed customer feedback—a new restaurant can significantly maximize its potential for achieving a 5-star reputation and long-term success.