

Yelp Ratings Mismatch Review Text Sentiment

DS 4002, Group 7: Dino
Papalabarakopoulos (leader), Pallavi
Mamillapalli, Cooper Bebeau

2/18/26

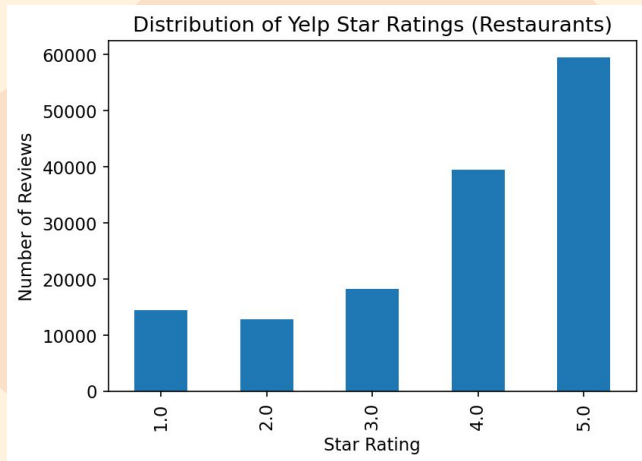


Motivation + Research Question

- Online reviews influence consumer decisions
- Star ratings summarize complex text
- Do ratings fully reflect written sentiment?
- Are mismatches systematic across cuisines?
- Hypothesis: Cuisine category is associated with the degree of sentiment-rating mismatch
- Research Question: Do some cuisine categories exhibit greater mismatches between textual sentiment and star ratings than others?
- Modeling Approach: Sentiment analysis + normalized ratings + statistical comparison across cuisines



Data & Acquisition



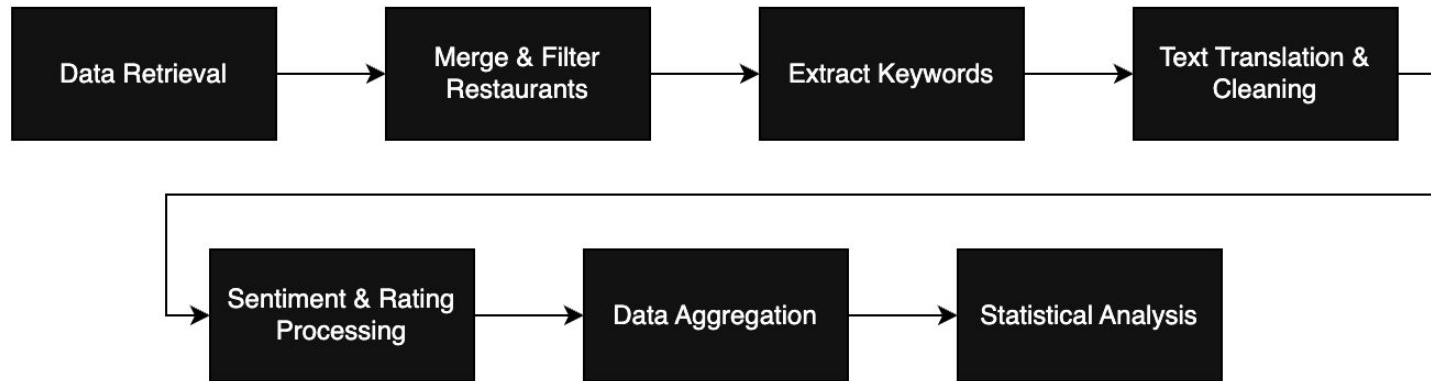
Data Source: Kaggle

- Kaggle Yelp Dataset
- 200,000 reviews (limited)
- Review dataset + Business dataset
- Focused on restaurants only

Final Dataset

- Review text
- Star rating
- Extracted cuisine tags
- Sentiment score (VADER)
- Normalized rating (0-1)
- Mismatch & direction metrics

Analysis Pipeline



Analysis Decision Making

Problem

- Many restaurants have multiple cuisine tags
 - Pick one? Pick most popular?
- Categories is stored as a String
- Reviews in different languages
- Handling large amounts of data

Impact

- Impacts ANOVA validity
- Affects statistical conclusions
- Strengthens robustness of findings

Our Approach

- Convert Categories to an array
- Identify and filter non-English reviews
- Limit review dataset to 200,000 entries
- Run statistics on exploded data set on cuisine and compare on multiple data set iterations with randomly assigned tags

Bias & Uncertainty

Bias Considerations

- Selection bias in Yelp reviews
- Translation inaccuracies (googletrans_py)
- Sentiment tool limitations
- Unequal cuisine sample sizes

Uncertainty Handling

- ANOVA at $\alpha = 0.05$
- Tukey HSD post-hoc tests
- Report top & bottom cuisines

Results and Conclusions

- As hypothesized, average sentiment-rating mismatches differed across the various cuisines (ANOVA, $p < 0.001$)
- Most cuisines carried slightly greater star rating than text sentiment on average
- No pattern sentiment-rating mismatch across cuisines from similar regions of the world (e.g. Thai, Chinese, Japanese)
- **Rating is an imperfect translation of review sentiment**

Top 5 Cuisines (by mean mismatch)

Cuisine	Mismatch	Direction
Chinese	0.295	-0.021
Italian	0.278	-0.001
Mexican	0.274	-0.006
Kosher	0.271	-0.070
Japanese	0.270	-0.006

Cuisine	Mismatch	Direction
Halal	0.198	-0.033
French	0.228	-0.039
Greek	0.229	-0.017
Korean	0.234	-0.008
Med.	0.238	0.000

Bottom 5 Cuisines (by mean mismatch)

Next Steps



Incorporate
geography



Explore cultural
explanations



Use other
sentiment
model

References

- GitHub Repository: <https://github.com/pallavim8/Project-1-Cavalier-Kaggles>
- [1] H. Li, H. Ji, J. M. Luo, and Z. Zhange, “Competition and restaurant online review manipulations: A dynamic panel data analysis,” *International Journal of Hospitality Management*, vol. 115, October 2023. [Online serial]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0278431923001792> [Accessed Feb. 1, 2026]
- [2] “Sentiment Analysis using VADER - Using Python”, [geeksforgeeks.org](https://www.geeksforgeeks.org/python/python-sentiment-analysis-using-vader/), para. 11, July 2025. [Online]. Available: <https://www.geeksforgeeks.org/python/python-sentiment-analysis-using-vader/>. [Accessed Feb. 1, 2026]
- [3] Yelp Inc., “Yelp Dataset,” [kaggle.com](https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset/data), n.d. [Online]. Available: <https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset/data>. [Accessed Feb. 8, 2026]
- [4] “Translate”, [googleapis.dev](https://googleapis.dev/nodejs/translate/latest/v2.Translate.html), n.d. [Online]. Available: <https://googleapis.dev/nodejs/translate/latest/v2.Translate.html>. [Accessed Feb. 8, 2026]

[illegible]