

2025 July 28

Text Analytics Project: Group 7

Transforming Yelp Reviews into Actionable Insights

By: Sudeeptha Sivarajan



AGENDA

- Overview
- Data Evaluation
- Key Insights: Ratings & Pricing
- Text Analysis: Word Clouds & Themes
- Sentiment & Topic Modeling
- Predictive Models
- Recommendations

Overview

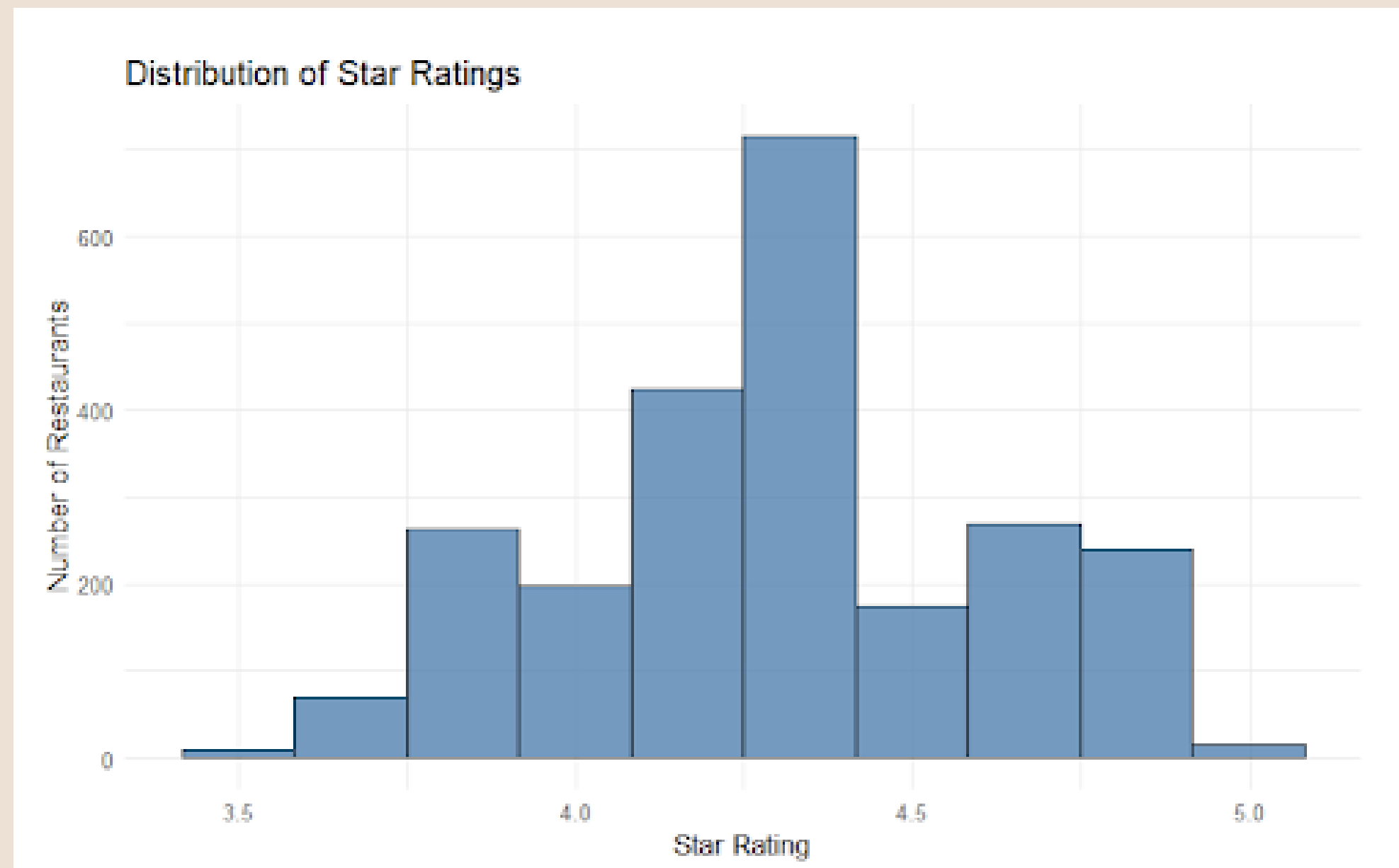
- The dataset has 2,381 reviews of top 240 restaurants in LA
- Key Variables: Star ratings, price categories, review texts.
- Methods: Text analytics, sentiment analysis, predictive modeling.
- Cleaned text: lemmatization, stopword removal.

Why It Matters:

- 90% of consumers read online reviews before dining
- Small improvements in ratings can boost revenue by 5-9%

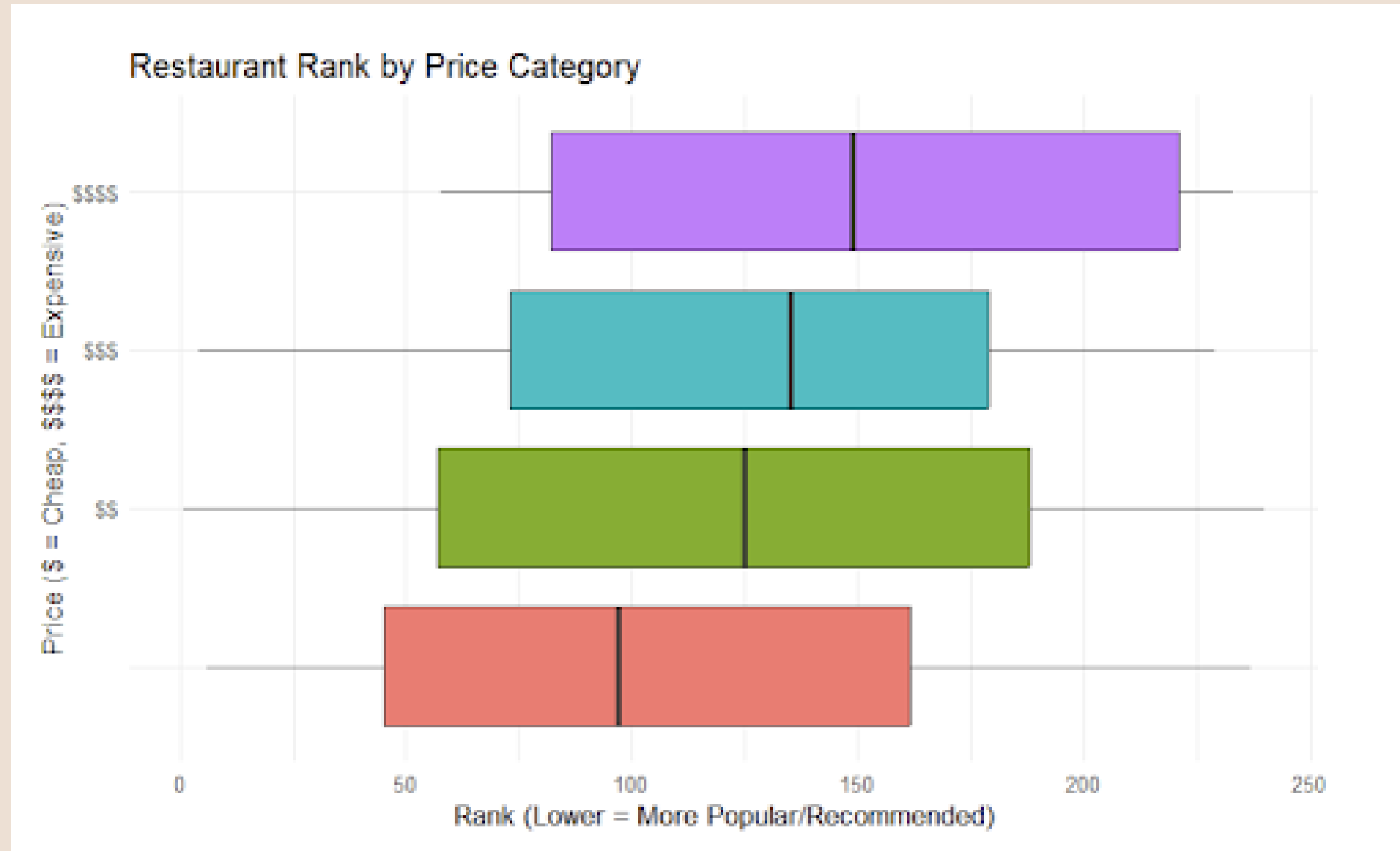
Star Rating Distribution

- Most ratings cluster between 4.0–4.5; extremes (≤ 3.5 or ≥ 4.8) are rare.
- Small range expected due to the nature of the dataset
- Implication: High baseline satisfaction; differentiation requires excellence.



Restaurant Rank by Price Category

- Ratings primarily around 4.0–4.5 which seems to be generally positive.
- Affordable restaurants are ranked better than expensive restaurants on average
- Popular keywords: food, service, and order.





3- Star



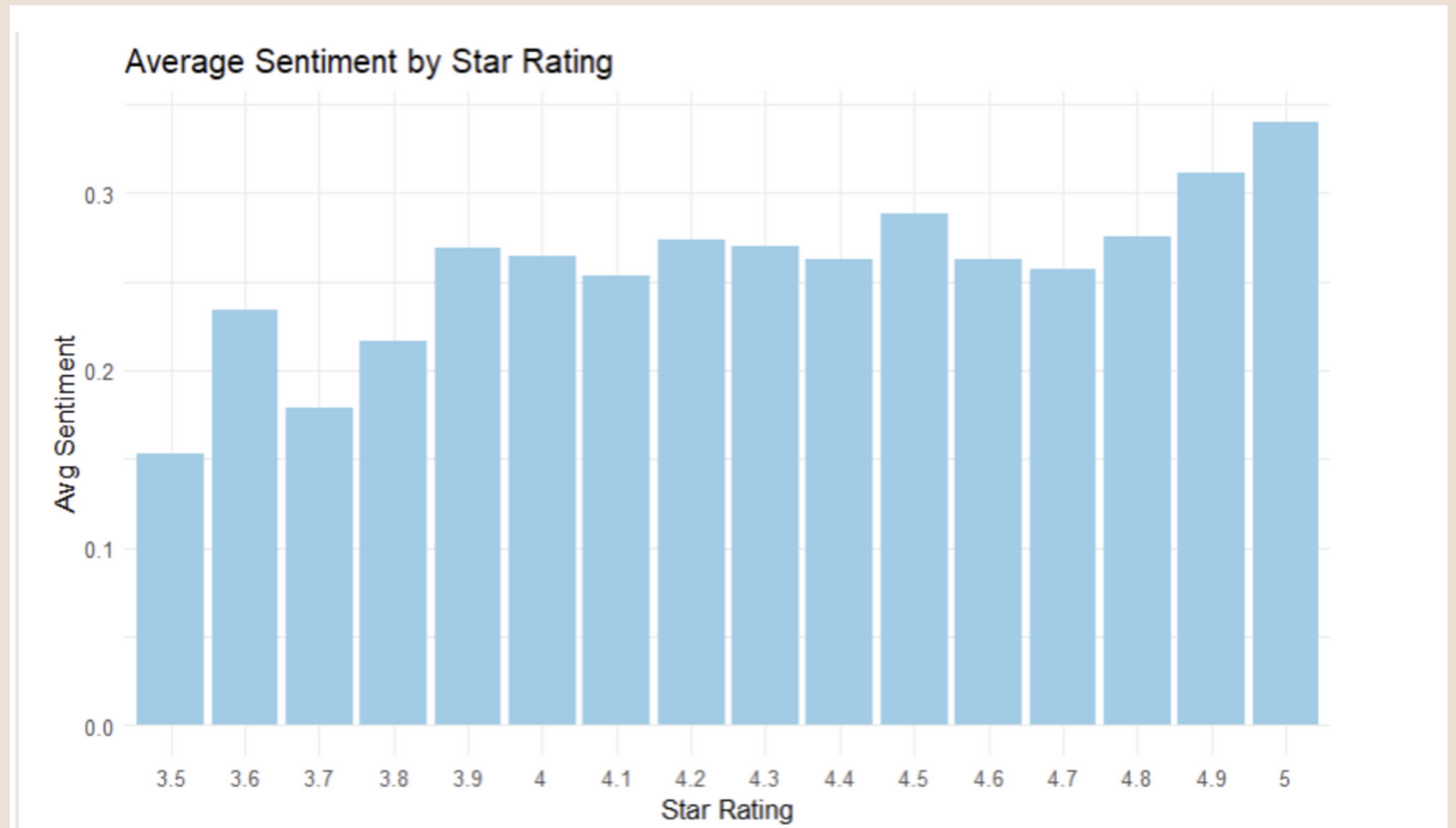
5- Star



4- Star

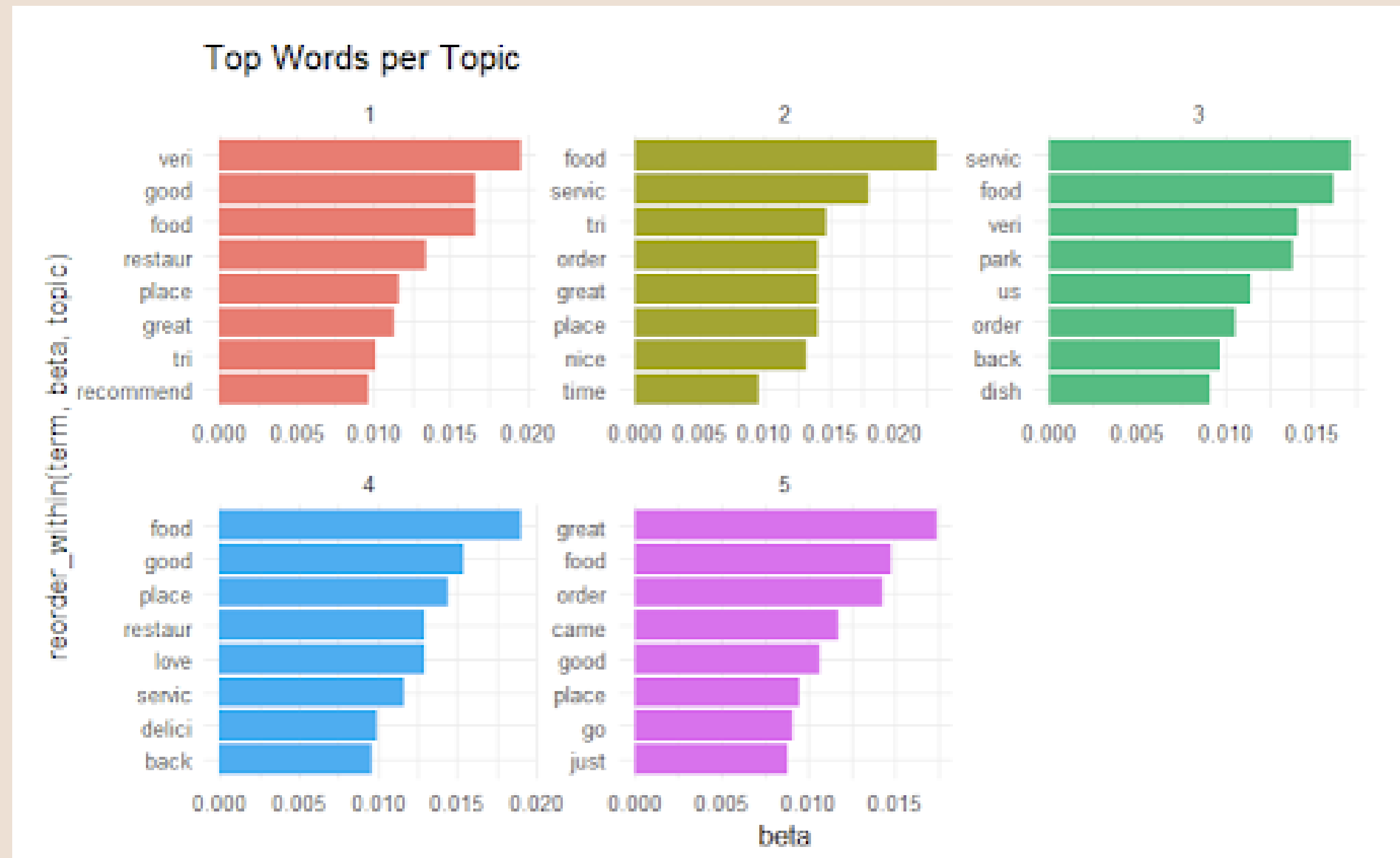
Sentiment analysis

- Higher ratings = more positive sentiment.
- 3-star = logistics issues (parking, wait).
- 4-star = personal service and authentic dishes
- 5-star = customers felt genuinely delighted and valued
- Topics: satisfaction, service, ambiance, emotion.



Topic Modeling

- 5 topics with the words with the highest betas
- Each topic appears to have a different focus or theme
- For example, topic 4 includes words like “delicious” and “love” that evoke strong emotions
- Topic 5 is action-oriented as it has words like “came”, “go”, and “order”



Modeling Results

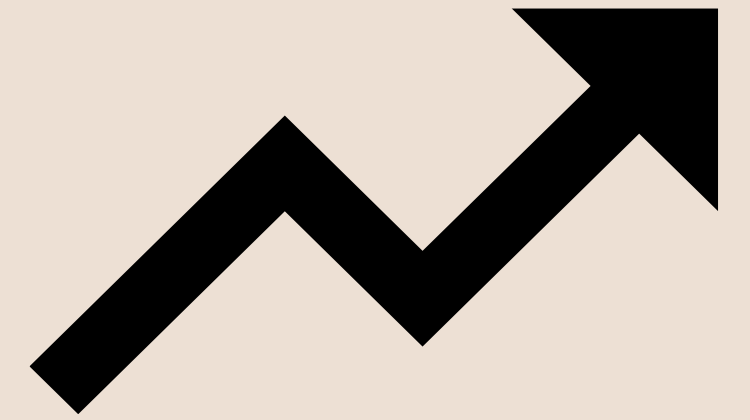
- Models: Linear Regression & KNN.
- Low R^2 indicating text alone isn't enough.
- KNN model showed high accuracy of 93%
- The regression model achieved an accuracy of 64%
- Heavily biased dataset with more number of cheaper restaurants

1. The **Regression model** performance:

```
## Confusion Matrix and Statistics
##
##              Reference
## Prediction  0    1
##           0 26   2
##           1 13   1
##
##
##              Accuracy : 0.6429
```

2. The **KNN model** performance:

```
## Confusion Matrix and Statistics
##
##              Reference
## Prediction  0    1
##           0 39   3
##           1  0   0
##
##
##              Accuracy : 0.9286
```



Business Recommendations

- Improve service, reduce operational friction.
- Offer distinct, flavorful menu items.
- Enhance ambiance and encourage photo sharing.
- Monitor sentiment trends over time.
- Use text analytics for reviews alongside other metrics

Thank you