

MKTG 6640 Final Project - Group 1

Maddie Lee, Leah Ekblad, Caleb Call & Tami Salvador

Objective

Analysis of customer reviews from Los Angeles restaurants.

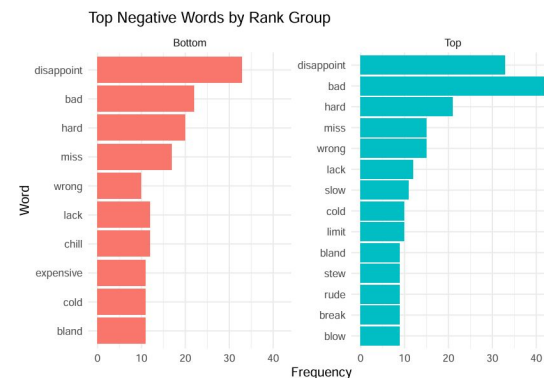
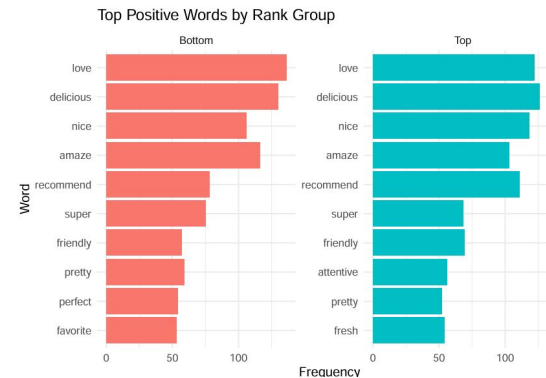
Comparison of Top 50 and Bottom 50 (based on Yelp rank) to identify:

- Common themes
- Sentiment patterns
- Operational differences



Restaurant success

The goal is to provide actionable insights to help restaurants improve their rankings and customer experience.



Process

Initial Cleaning

- Date Formatting
- Categorical Conversion
- Rank Grouping

Text Cleaning & Tokenization

- Preprocess text
- Tokenization
- Stop Word Removal
- Lemmatization

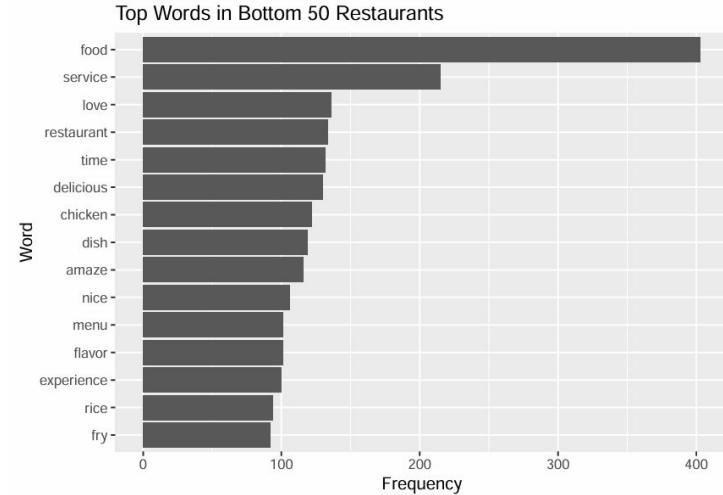
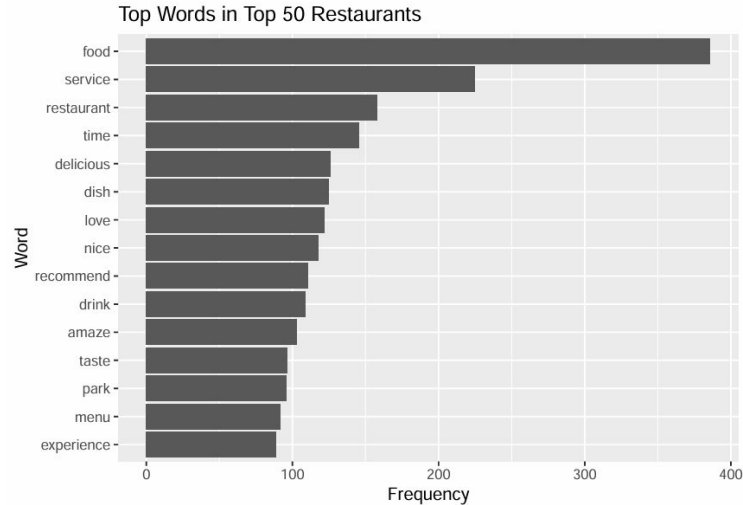
Document Structuring

```
summary(reviews)
```

```
##           Rank      CommentDate      Date
## Min.      : 1.0    Min.      :2011-12-07  Min.      :2023-09-17
## 1st Qu.: 61.0    1st Qu.:2023-08-18  1st Qu.:2023-09-17
## Median :119.0    Median :2023-09-03  Median :2023-09-17
## Mean   :120.2    Mean   :2023-08-08  Mean   :2023-09-17
## 3rd Qu.:180.0    3rd Qu.:2023-09-10  3rd Qu.:2023-09-17
## Max.   :240.0    Max.   :2023-09-17  Max.   :2023-09-17
##
##           RestaurantName  Comment      Address
## Sun Nong Dan             : 20    Length:2381    Length:2381
## Bottega Louie - WEHO     : 14    Class :character  Class :character
## Providence               : 14    Mode  :character  Mode  :character
## Morrison Atwater Village: 13
## Eat This Cafe            : 12
## Il Cielo                 : 12
## (Other)                  :2296
##      StarRating  NumberOfReviews      Style
## Min.      :3.500  Min.      : 2    Korean      : 65
## 1st Qu.:4.100  1st Qu.: 129    American (New), Cocktail Bars: 61
## Median :4.300  Median : 518    American (New)      : 60
## Mean   :4.305  Mean   :1142    Italian             : 41
## 3rd Qu.:4.500  3rd Qu.:1623    Korean, Soup        : 30
## Max.   :5.000  Max.   :10020    Thai                : 30
##                                     (Other)      :2094
##
##      Price
## Length:2381
## Class :character
## Mode  :character
##
```

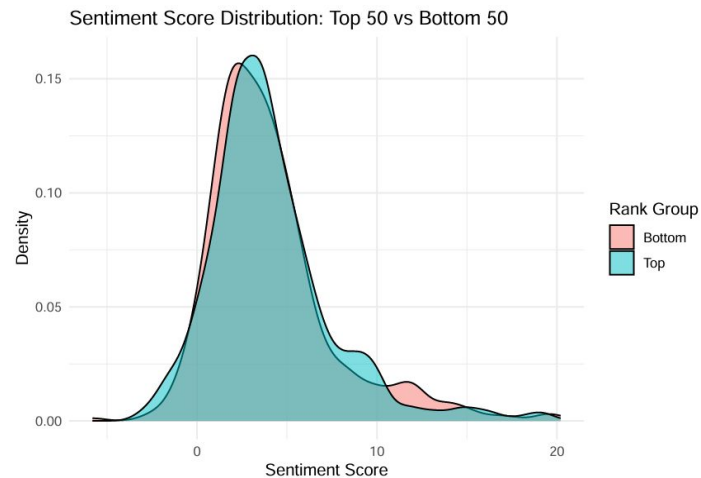
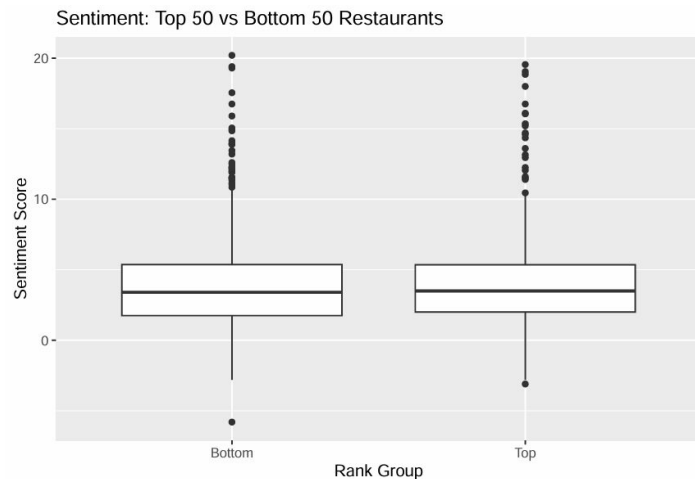
Process - Analysis Techniques

Frequency Analysis - Identified most frequent lemmatized words in Top vs. Bottom restaurants.



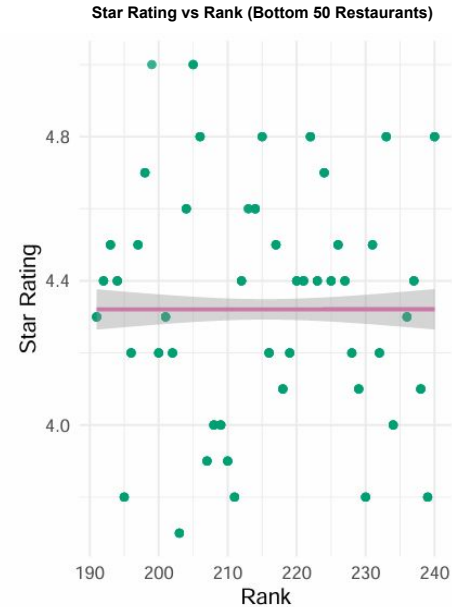
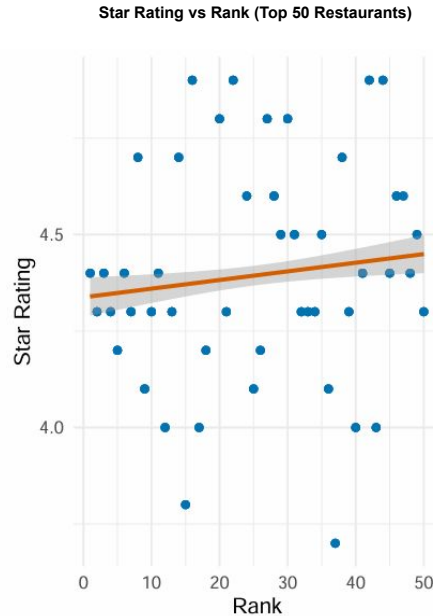
Process - Analysis Techniques

Sentiment Analysis - Explored and compared review patterns between the top and bottom ranked restaurants



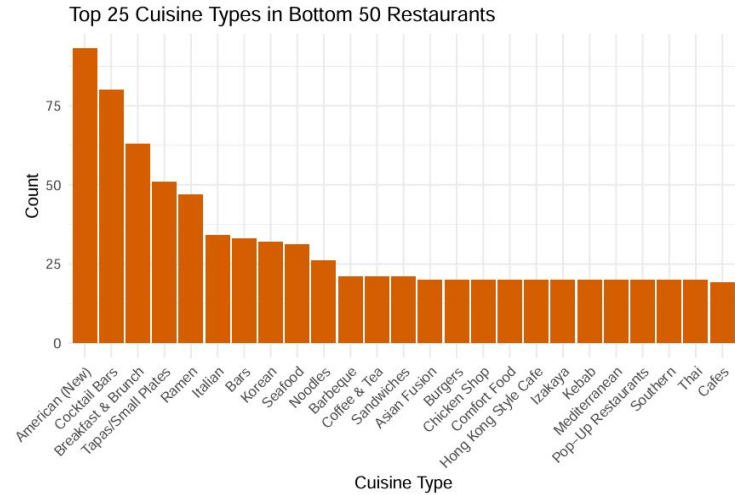
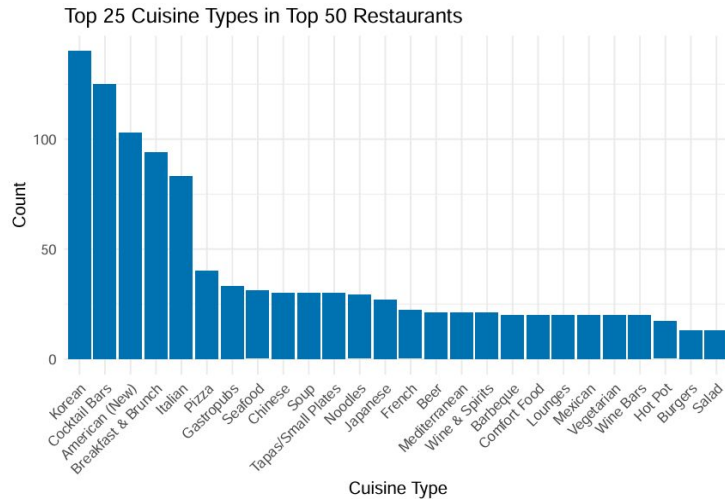
Process - Analysis Techniques

Star Rating Analysis - Per-restaurant average star rating with dot plots for Top and Bottom groups.



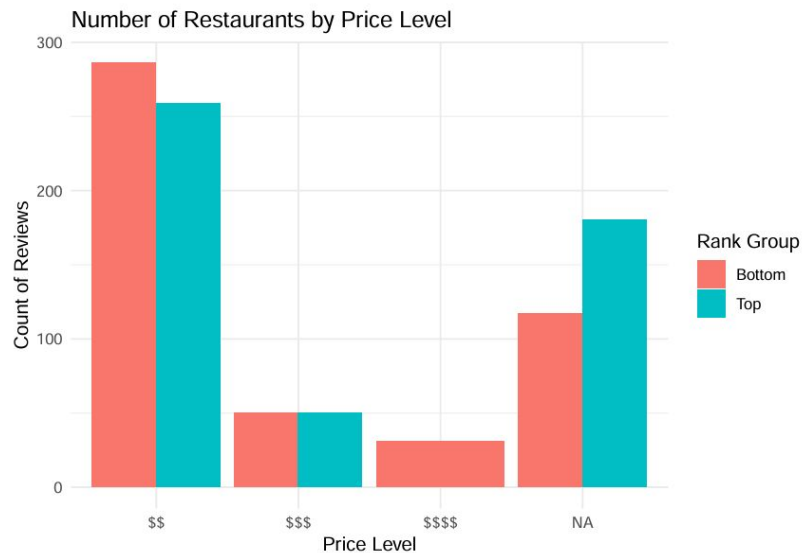
Process - Analysis Techniques

Cuisine Type Analysis - Counted and visualized frequency of cuisine types in both groups.



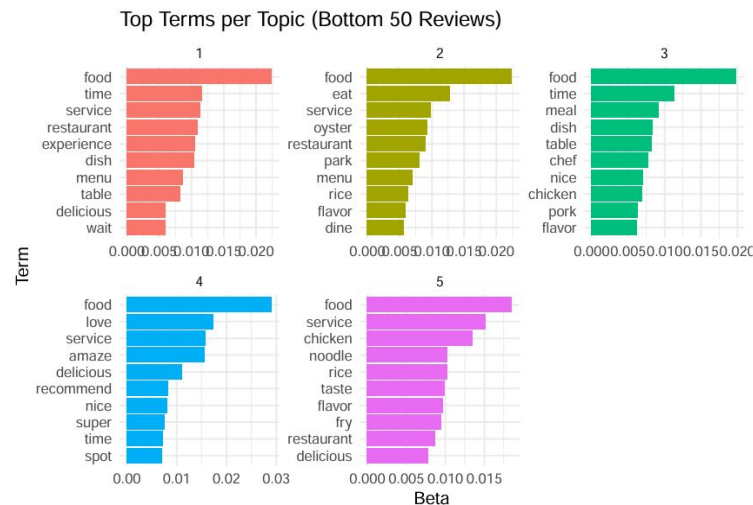
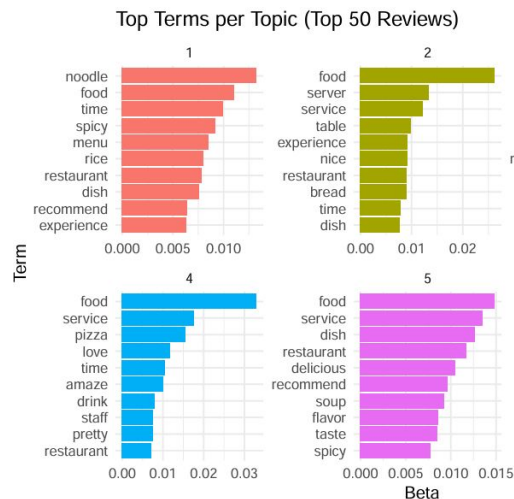
Process - Analysis Techniques

Price Level Analysis - Compared distribution of restaurants by price level between Top and Bottom groups using bar plots.



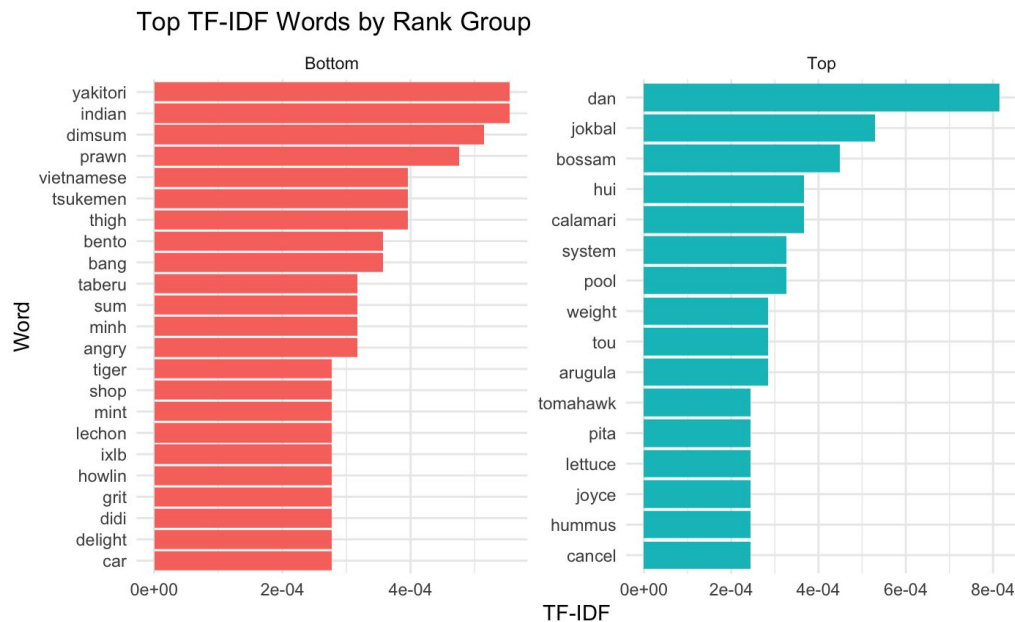
Process - Analysis Techniques

Topic Modeling (LDA) - Separated Models for Top 50 and Bottom 50 reviews and plotted top 10 terms for each topic using faceted bar charts.



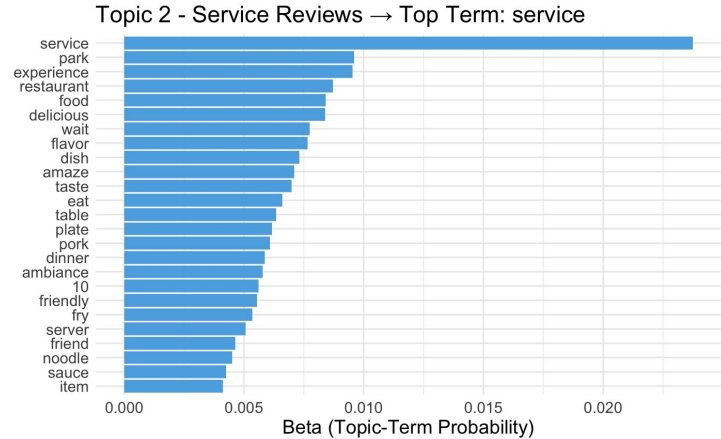
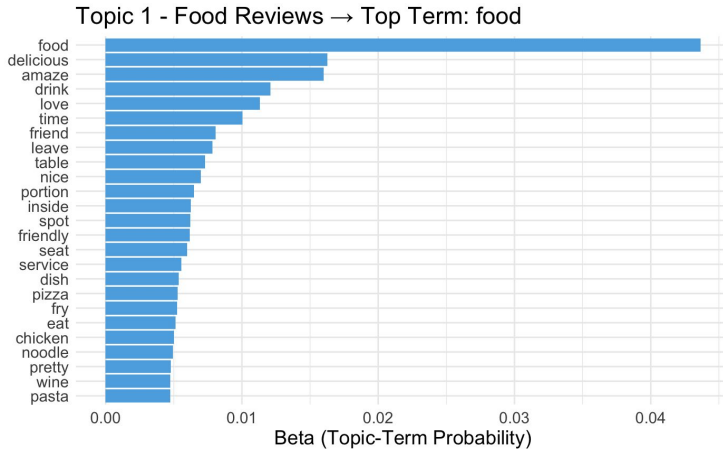
Process - Analysis Techniques

Distinctive Keywords by Rank Group (TF-IDF) - TF-IDF (Term Frequency–Inverse Document Frequency) highlights terms that are uniquely important in a group of documents. We calculated TF-IDF separately for reviews of the Top 50 and Bottom 50 LA restaurants.



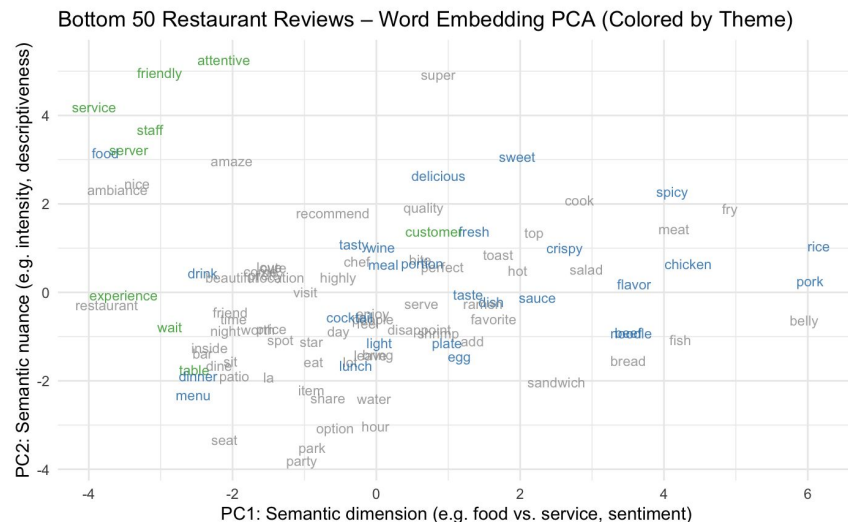
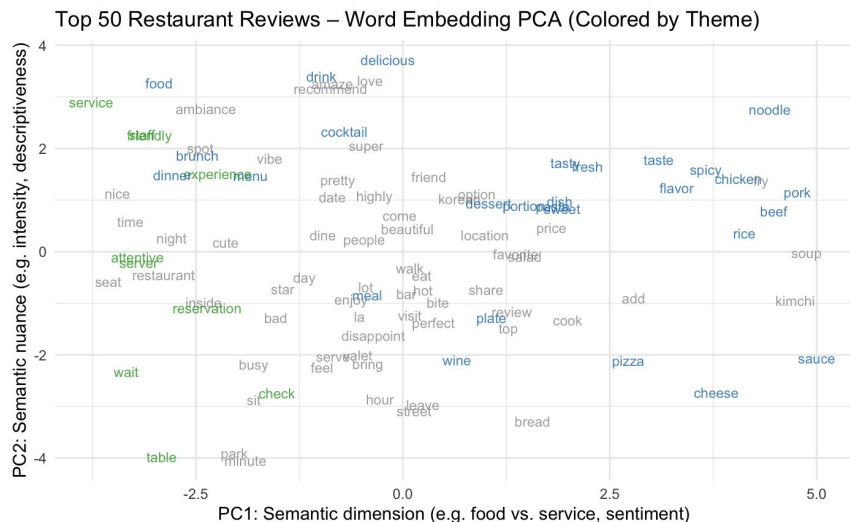
Process - Analysis Techniques

Keyword based filtering and LDA topic modeling - We used keyword-based filtering and LDA topic modeling to separately analyze reviews centered on *food* vs. *service*. Each model uncovered the dominant themes reviewers associate with those aspects of the restaurant experience.



Process - Analysis Techniques

Word Embedding - We trained separate GloVe models on Top 50 and Bottom 50 reviews, then visualized 2D PCA plots with words color-coded by theme: blue (food), green (service), and gray (other).



Results

Sentiment & Language:

- Top 50 reviews are emotionally rich and focused—especially around food.
- Bottom 50 reviews are less cohesive, often blending food and service issues.

TF-IDF Keywords:

- *Top 50*: Terms like *prawn*, *yakitori*, *dimsum* reflect upscale experiences.
- *Bottom 50*: Words like *lettuce*, *hummus*, *cancel* suggest casual or problematic themes.

Topic Modeling:

- *Top 50*: Themes center on food quality, ambiance, and recommendations.
- *Bottom 50*: Focused on wait times, inconsistencies, and vague service comments.

Word Embeddings:

- Top 50 reviews show tight semantic clusters by theme (food, service).
- Bottom 50 language is scattered, reflecting weaker thematic structure.

Cuisine & Pricing:

- Top 50 = more fusion/New American, higher price tiers (\$\$\$+).
- Bottom 50 = broader mix, lower price points.



Recommendations and Next Steps

- **Enhance Emotional Impact:** Focus messaging on food experience and ambiance to match language used in Top 50 reviews.
- **Fix Service Pain Points:** Address delays, inconsistent service, and unclear staff interactions.
- **Use Review Analytics:** Monitor sentiment trends and topic shifts for early warning signals.
- **Refine Comparison Groups (Next Step):** Future analysis should compare the Top 50 against LA restaurants with *low star ratings* (e.g., <3.5), not just the next 190 in the top 240. This would better highlight what truly separates top from struggling restaurants.



Thank you

Maddie Lee, Leah Ekblad, Caleb Call & Tami Salvador