

Yelp Project Written Report

Aileen Wu (acw243), Kathryn Lee (kl2239), Pei-Jung Hsieh (ph489), Sydney Chien (cc2897)

NBA 6550 Business Data Analysis with SQL

Professor Nur Kaynar

Introduction

Online Review plays a key role in shaping consumer perceptions of whether to visit a restaurant and go back constantly. Our study examines the relationship between Yelp Restaurant Rating and review patterns to identify trends influencing customer engagement and satisfaction, specifically in Las Vegas (Appendix A). By analyzing key factors like review growth, rating trends, user activity levels, and sentiment analysis of review keywords, we want to understand whether higher-rated restaurants attract more reviews and the reasons contributing to the high performance. This paper also explores the correlation between active users' feedback and inactive users' feedback. Using SQL and visualizations, our findings offer insights for restaurant owners to optimize customer engagement and improve service quality.

Part I: Yearly Growth Trend Analysis: Review Count vs. Average Rating

This analysis highlights patterns in consumer engagement over time by examining the yearly growth trends in review count and average rating (Appendix B, Figure B-1). Between 2006 and 2007, there was a significant surge in review growth, peaking at over 250%, likely driven by the expansion of the Yelp platform and increased consumer adoption. (Appendix B, Figure B-2) However, after this spike, the growth rate of reviews steadily declined. While the absolute number of reviews continued to rise, the rate of increase eventually stabilized and approached 0%, suggesting a stagnation point in customer feedback growth.

Additionally, the average rating growth remained relatively stable throughout the period, with most customers consistently giving ratings around 3.5. This trend suggests that customer sentiment remained neutral over time, with little variation in overall satisfaction. A correlation analysis between the growth rate of reviews and ratings yielded a near-zero correlation of 0.003, indicating that an increase in the number of reviews does not necessarily lead to higher customer ratings. This finding underscores that while engagement levels fluctuate, consumer perception of quality remains largely unchanged.

Part II: User Activity Trends: Analyzing Overall Review Patterns of Active vs. Inactive Users

This section examines how active and inactive users contribute to restaurant ratings in Las Vegas. Since our analysis focuses on the restaurant industry in Las Vegas, we first filtered users who posted reviews and rated restaurants in this location. Users were then categorized based on their review activity: if the review count of a user is greater than the average review counts, it is divided into active users, otherwise inactive users. As the (Appendix C-1) shows, there are 1164 active users who posted reviews for restaurants in Las Vegas with an average rating of 3.80, while 2199 inactive users with an average rating of 3.72. To assess whether user activity status impacts ratings, we performed a correlation analysis between overall business ratings (1-5 stars) and all restaurant reviews in Las Vegas. The results, shown in (Appendix C-2),

indicate a very weak or no correlation (-0.02) between overall ratings and user status. Although this suggests no significant evidence that user activity level influences business ratings, we observed a key trend that as the number of Yelp users grows, the number of reviews with ratings ≥ 4 significantly increases based on the previous section (Appendix B, Figure B-1). This finding prompted us to further investigate restaurants with an average rating of 4 or higher, to see if user activity plays a larger role in highly rated businesses.

Part III: User Status and Rating Correlation: Do Active Users Influence Business Ratings? Focusing on average rating ≥ 4

Building on our previous findings, we applied the same analytical approach but focused specifically on restaurants with ratings ≥ 4 to assess whether user activity level plays a more significant role in high-rated restaurants. The analysis table (Appendix D-1) shows that there are 900 active users who reviewed Las Vegas restaurants rated ≥ 4 stars, with an average rating of 3.81, whereas 1,509 inactive users with an average rating of 3.91. Compared to the overall users in Las Vegas, more than 50% of both active and inactive users have reviewed restaurants rated ≥ 4 stars, showing greater user engagement in higher-rated establishments. Furthermore, in the overall dataset, active users rated slightly higher than inactive users (3.80 vs. 3.72, a +0.08 difference), but when focusing only on highly rated restaurants (≥ 4 stars) (Appendix D-2), this pattern reversed, with active users rating lower than inactive users (3.81 vs. 3.91, a -0.10 difference).

Continuing to evaluate whether more active users lead to higher business ratings, we got a correlation value -0.08, which presents a stronger correlation than the correlation (-0.02) between active user percentage and business ratings in the overall dataset (1-5 stars) (The visualization of the comparison could be viewed in Appendix D, Figure D-2.1). This suggests that active users are more critical when reviewing top-rated restaurants. Since active users are more critical reviewers, businesses should adopt proactive engagement strategies, such as personalized responses, loyalty rewards, and service improvements to mitigate negative feedback. Also, high-rated businesses can focus on incentivizing casual customers to submit reviews, which may help counterbalance stricter active user ratings as inactive users tend to leave slightly higher ratings. Restaurants should be aware that while an increasing number of active users may boost engagement, it may not always improve ratings. Managing expectations and balancing user feedback is key to long-term success.

Part IV: Sentiment Analysis: Positive & Negative Reviews

To further analyze user reviews and provide valuable insights for restaurants with 4- or 5-star ratings, the data is divided into positive and negative reviews. By selecting keywords associated with positive and negative tones, the analysis helps identify what restaurants should maintain and what areas need improvement. In addition to categorizing reviews as positive or

negative, the data is further broken down into categories such as service, cleanliness, food quality, wait time, price, and other factors (needs review) for a more detailed analysis.

IV-A: Positive Sentiment Analysis

The distribution of positive reviews shows that food quality (27.81%) and service quality (21.17%) are the most valued aspects of customers' dining experiences (Appendix E-1). Customers primarily emphasize taste, freshness, and preparation when they are evaluating food, while professionalism, friendliness, and helpfulness in service-related reviews. In contrast, other factors such as ambiance/cleanliness, wait time, and price were mentioned less frequently in the reviews (Appendix E-1). While these aspects may still affect the total review of the restaurant, they are considered less critical compared to food and service quality.

In positive service-related keywords, "friendly" (54.75%) was the most frequently mentioned word, showing the importance of having warm and welcoming interactions with customers (Appendix E-2). Other highly mentioned terms, such as being "helpful" (17.42%) and "professional" (15.92%) indicate the importance of being professional during their service. Similarly, when analyzing food-related keywords, "delicious" (44.94%), "fresh" (28.71%), and "tasty" (17.21%) were the keywords that were mentioned the most. This shows that flavor and ingredient quality are the strong contributors in a positive dining experience. Interestingly, "authentic" (4.91%) was mentioned less frequently, which shows that the authenticity of the dishes may not be as important as the overall freshness or taste of the food (Appendix E-3).

IV-B: Negative Sentiment Analysis

After splitting reviews to negative reviews into categories (Appendix E-4) using negative tone words, the report further calculates the percentages comparison for each category. Focusing on negative reviews within each category, Food Quality has the highest percentage (15.44%) compared to other categories (Appendix E-5, Figure E-5.1). Although we are analyzing reviews from restaurants with 4 and 5-star ratings, negative feedback still appears, though at a lower rate. Food Quality (15.44%) is the leading complaint, indicating that issues related to taste, freshness, or preparation persist despite overall positive experiences. To provide actionable business insights, we will extract and analyze key recurring terms within Food Quality complaints to identify which issues stand out (Appendix E-5). By comparing the frequency each keyword has presented in food quality, we can pinpoint specific problem areas that require attention.

After data analysis, "old" appears overwhelmingly more than any other term, suggesting freshness is a major concern (Appendix E-6, Figure E-6). There is 69.13% compared to any other keywords. "Cold & dry" are the next most common words appearing in the reviews. "Cold" has a percentage of 11.29% and "Dry" has 10.25%, adding up to 21.54%. "Cold" food complaints suggest possible delays in serving or inadequate heating methods. "Dry" could relate to

overcooked meats, stale baked goods, or improper food storage. Food's freshness and texture seem to be a bigger issue compared to taste or flavoring.

Summary and Business Recommendations

This analysis provides insights into consumer engagement trends, user activity patterns, and sentiment analysis for Yelp restaurant reviews in Las Vegas. While Yelp saw rapid review growth in its early stages, the growth rate has stabilized. The average rating has remained consistently neutral, with no significant correlation between review growth and ratings.

Another interesting finding is that the correlation between user activity level and business rating is weak. Active users tend to be more critical when giving reviews. This can suggest that businesses should pay more attention to feedback from highly engaged users. Inactive users tend to leave slightly higher ratings since they are normally more casual reviews. This means that as a restaurant gains more engagement, ratings may not necessarily increase. This can suggest restaurants should focus on maintaining food and service quality to improve business ratings.

Sentiment analysis on the customer reviews provides details of how restaurants can better their ratings by improving the quality and service. While positive reviews often highlight good service and ambiance, negative reviews are predominantly driven by food quality concerns, particularly issues related to freshness and texture. This suggests that the restaurant owners should invest more effort in food storage, preparation, and service time. On the service side, many customers value restaurant staff being friendly and professional, suggesting that response time and attentiveness are critical to customer satisfaction. Training staff to be more efficient while maintaining a friendly and welcoming attitude could enhance customer experience and lead to getting higher ratings.

Since frequent reviewers are more critical, Yelp can suggest businesses to implement personalized responses or loyalty programs to address their concerns. Businesses should also prioritize food quality improvements. A better understanding of review trends can help businesses proactively manage their Yelp visibility, enhance customer satisfaction, and ultimately increase rating over time.

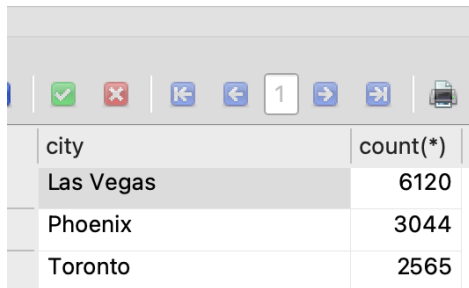
Appendix:

Appendix A

Pre condition:

- Business: restaurant
- City: Las Vegas

```
select distinct city, count(*)  
from business  
group by city  
order by count(*) desc;
```



city	count(*)
Las Vegas	6120
Phoenix	3044
Toronto	2565

- Stars: ≥ 4

Appendix B: Review Growth Trends

```
SELECT business_id, name, city, business.stars, strftime('%Y', date) AS year,  
COUNT(review_id) AS count
```

```
FROM review
```

```
JOIN business USING (business_id)
```

```
WHERE categories LIKE "%restaurant%"
```

```
AND (city LIKE "%Las Vegas%" OR city LIKE "%las vegas%")
```

```
AND business_id IN (
```

```
    SELECT business_id
```

```
    FROM review
```

```
    JOIN business USING (business_id)
```

```
    WHERE categories LIKE "%restaurant%"
```

```
    GROUP BY business_id
```

```
    HAVING COUNT(*) >= 2
```

```
)
```

```
GROUP BY name, business.stars, year, city
```

```
ORDER BY name, year, business.stars;
```

–Creating temporary table

```
CREATE TEMPORARY TABLE trend_table AS
```

```
SELECT business_id,
```

```

    name,
    review.stars,
    strftime('%Y', date) AS year
FROM review
JOIN business USING (business_id)
WHERE categories LIKE "%restaurant%"
AND (city LIKE "%Las Vegas%" OR city LIKE "%las vegas%")
AND business_id IN (
    SELECT business_id
    FROM review
    JOIN business USING (business_id)
    WHERE categories LIKE "%restaurant%"
    GROUP BY business_id
    HAVING COUNT(*) >= 2
)
ORDER BY business_id;

SELECT business_id, name, stars, year, COUNT(*) AS count
FROM trend_table
GROUP BY business_id, stars, year;

```

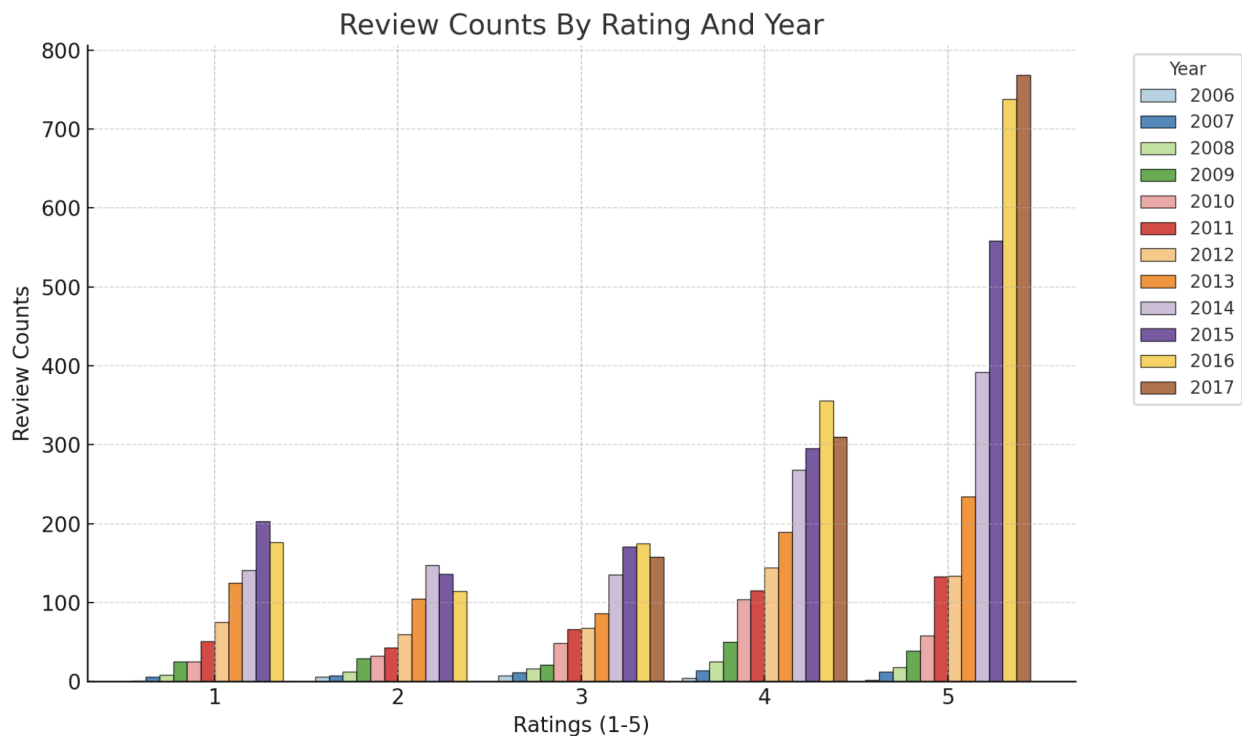


Figure B-1

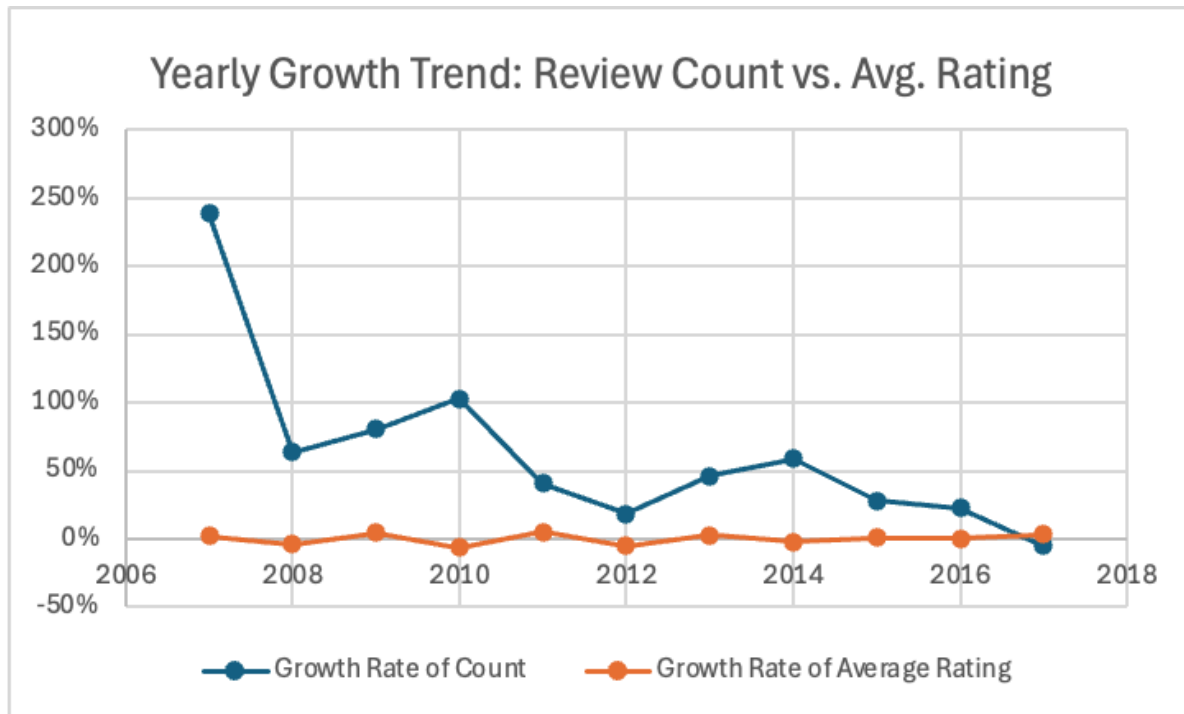


Figure B-2

Appendix C-1: User Activity Trends - Active vs. Inactive Users Data

```
-- divide user into active user and inactive user for all restaurant ratings (1-5)
with avg_review as (
  -- Compute the average review count for classification
  select round(avg(review_count), 2) as avg_reviews from user
),
filtered_users as (
  -- Select only users who reviewed restaurants in Las Vegas
  select distinct u.user_id, u.review_count, u.average_stars
  from review r
  join user u on r.user_id = u.user_id
  join business b on r.business_id = b.business_id
  where b.categories like '%Restaurant%' and (b.city like 'Las Vegas' or b.city like 'las vegas')
)
select
  case
    when review_count > (select avg_reviews from avg_review) then 'Active User'
    else 'Inactive User'
  end as "User Type",
  count(user_id) as user_count,
  round(avg(average_stars), 2) as avg_given_stars
```


from filtered_users
group by "User Type";

	User Type	user_count	avg_given_stars
1	Active User	1164	3.8
2	Inactive User	2199	3.72

Appendix C-2: Correlation Analysis (User Activity vs. Rating from 1-5 star)

```
with avg_values as (
  select avg(active_user_percentage) as avg_x, avg(business_rating) as avg_y
  from(
    select b.business_id, b.stars as business_rating,
    (100.0 * count(case when u.review_count > (select avg(review_count) from user) then
r.user_id end) / count(r.user_id))
    as active_user_percentage
    from review r
    join business b on r.business_id = b.business_id
    join user u on r.user_id = u.user_id
    where b.categories like '%Restaurant%' and (b.city like 'Las Vegas' or b.city like 'las vegas')
    group by b.business_id, b.stars)
  ),
data_points as (
  select b.business_id, b.stars as business_rating,
  (100.0 * count(case when u.review_count > (select avg(review_count) from user) then
r.user_id end) / count(r.user_id))
  as active_user_percentage,
  (100.0 * count(case when u.review_count > (select avg(review_count) from user) then
r.user_id end) / count(r.user_id))
  - (select avg_x from avg_values) as x_diff,
  b.stars - (select avg_y from avg_values) as y_diff
  from review r
  join business b on r.business_id = b.business_id
  join user u on r.user_id = u.user_id
  where b.categories like '%Restaurant%' and (b.city like 'Las Vegas' or b.city like 'las vegas')
  group by b.business_id, b.stars),
correlation_calc as (
  select sum(x_diff * y_diff) as numerator,
  sqrt(sum(x_diff * x_diff)) as denominator_x,
  sqrt(sum(y_diff * y_diff)) as denominator_y
```

```

from data_points)
select numerator / (denominator_x * denominator_y) as correlation_value
from correlation_calc;

```

	correlation_value
1	-0.02028050783984

Appendix D-1: Active vs. Inactive User Ratings in High-Rated Restaurants (Rating ≥ 4)

```

with avg_review as (
  -- Compute the average review count for classification
  select round(avg(review_count), 2) as avg_reviews from user
),
filtered_users as (
  -- Select only users who reviewed restaurants in Las Vegas with stars  $\geq 4$ 
  select distinct u.user_id, u.review_count, u.average_stars
  from review r
  join user u on r.user_id = u.user_id
  join business b on r.business_id = b.business_id
  where b.categories like '%Restaurant%' and (b.city like 'Las Vegas' or b.city like 'las vegas')
  and b.stars  $\geq 4$ 
)
select
  case
    when review_count > (select avg_reviews from avg_review) then 'Active User'
    else 'Inactive User'
  end as "User Type",
  count(user_id) as user_count,
  round(avg(average_stars), 2) as avg_given_stars
from filtered_users
group by "User Type";

```

	User Type	user_count	avg_given_stars
1	Active User	900	3.81
2	Inactive User	1509	3.91

Appendix D-2: Correlation Analysis (User Status vs Rating ≥ 4)

```

with avg_values as (
  select avg(active_user_percentage) as avg_x, avg(business_rating) as avg_y

```

```

from(
    select b.business_id, b.stars as business_rating,
    (100.0 * count(case when u.review_count > (select avg(review_count) from user) then
r.user_id end) / count(r.user_id))
    as active_user_percentage
    from review r
    join business b on r.business_id = b.business_id
    join user u on r.user_id = u.user_id
    where b.categories like '%Restaurant%' and (b.city like 'Las Vegas' or b.city like 'las vegas')
and b.stars >= 4
    group by b.business_id, b.stars)
),
data_points as (
    select b.business_id, b.stars as business_rating,
    (100.0 * count(case when u.review_count > (select avg(review_count) from user) then
r.user_id end) / count(r.user_id))
    as active_user_percentage,
    (100.0 * count(case when u.review_count > (select avg(review_count) from user) then
r.user_id end) / count(r.user_id))
    - (select avg_x from avg_values) as x_diff,
    b.stars - (select avg_y from avg_values) as y_diff
    from review r
    join business b on r.business_id = b.business_id
    join user u on r.user_id = u.user_id
    where b.categories like '%Restaurant%' and (b.city like 'Las Vegas' or b.city like 'las vegas')
and b.stars >= 4
    group by b.business_id, b.stars),
correlation_calc as (
    select sum(x_diff * y_diff) as numerator,
    sqrt(sum(x_diff * x_diff)) as denominator_x,
    sqrt(sum(y_diff * y_diff)) as denominator_y
    from data_points)
select numerator / (denominator_x * denominator_y) as correlation_value
from correlation_calc;

```

	correlation_value
1	-0.0808202835853

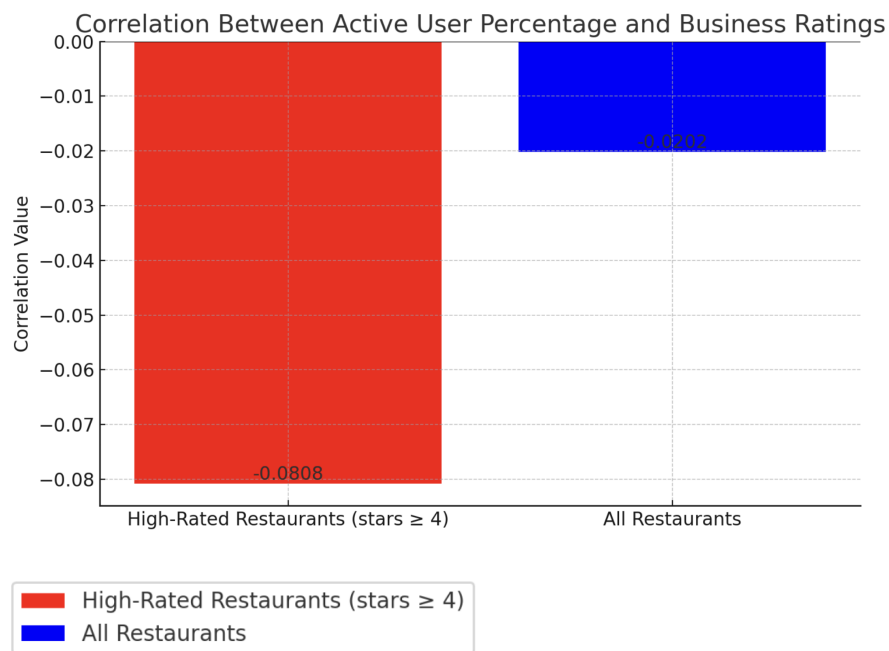


Figure D-2.1 Comparison of Active User Influence on Business Ratings Across All vs. High-Rated Restaurants – (Combined Visualization of Appendix B-2 & C-2 Data)

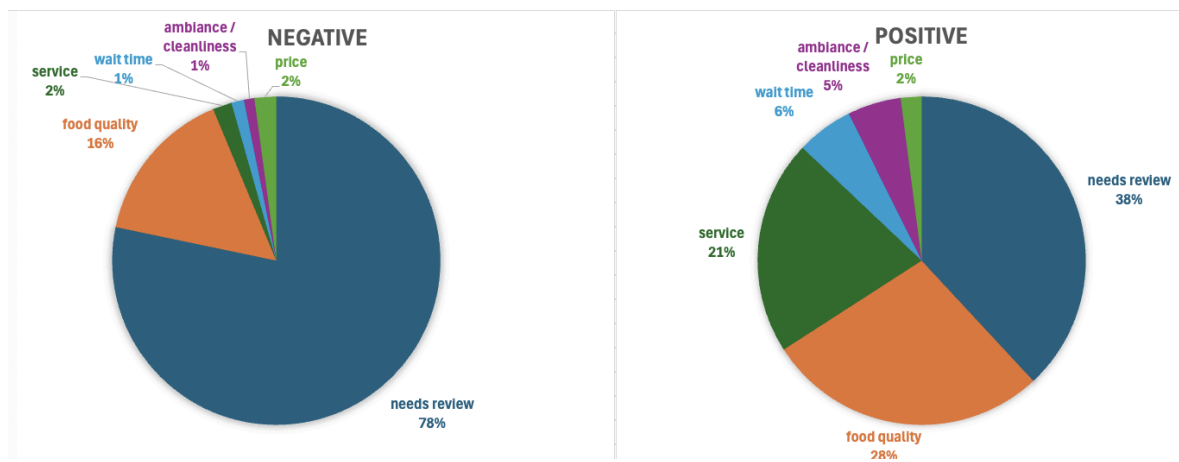


Figure E-1.1 Distribution of Positive and Negative Reviews Across Different Aspects of the Dining Experiences

Appendix E-1 (for Part IV-A)

```
select
  category,
  count(*) as review_count,
  round(100.0 * count(*) / sum(count(*)) over (), 2) as percentage
from (
  select r.text,
  case
```

-- Food Quality

```
  when r.text like '%delicious%'
    or r.text like '%tasty%'
    or r.text like '%fresh%'
    or r.text like '%amazing food%'
    or r.text like '%great taste%'
    or r.text like '%perfectly cooked%'
    or r.text like '%savory%'
    or r.text like '%juicy%'
    or r.text like '%crispy%'
    or r.text like '%tender%'
    or r.text like '%seasoned well%'
    or r.text like '%yummy%'
    or r.text like '%appetizing%'
    or r.text like '%incredible food%'
  then 'food quality'
```

-- Service

```
  when r.text like '%friendly%'
    or r.text like '%attentive%'
    or r.text like '%helpful%'
    or r.text like '%welcoming%'
    or r.text like '%great service%'
    or r.text like '%polite%'
    or r.text like '%wonderful staff%'
    or r.text like '%accommodating%'
    or r.text like '%amazing service%'
    or r.text like '%quick service%'
    or r.text like '%professional%'
    or r.text like '%kind staff%'
  then 'service'
```

-- *Ambiance & Cleanliness*

when r.text like '%beautiful decor%'
or r.text like '%clean%'
or r.text like '%cozy atmosphere%'
or r.text like '%great ambiance%'
or r.text like '%nice lighting%'
or r.text like '%relaxing%'
or r.text like '%comfortable seating%'
or r.text like '%spacious%'
or r.text like '%aesthetic%'
or r.text like '%good vibe%'
or r.text like '%elegant%'
or r.text like '%great view%'
or r.text like '%well-designed%'
or r.text like '%peaceful%'
then 'ambiance / cleanliness'

-- *Price Value*

when r.text like '%worth the price%'
or r.text like '%affordable%'
or r.text like '%good value%'
or r.text like '%fair price%'
or r.text like '%reasonable price%'
or r.text like '%good deal%'
or r.text like '%great price%'
or r.text like '%budget-friendly%'
or r.text like '%excellent value%'
or r.text like '%inexpensive%'
then 'price'

-- *Fast Service & Efficiency*

when r.text like '%fast%'
or r.text like '%quick%'
or r.text like '%efficient%'
or r.text like '%no wait%'
or r.text like '%seated immediately%'
or r.text like '%speedy%'
or r.text like '%on time%'
or r.text like '%got seated fast%'
then 'wait time'

```

-- Other
else 'needs review'
end as category
from review r
join business b on r.business_id = b.business_id
where r.stars >= 4
and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
) as categorized_reviews
group by category
order by review_count desc;

```

	category	review_count	percentage
1	needs review	3824	38.08
2	food quality	2792	27.81
3	service	2126	21.17
4	wait time	563	5.61
5	ambiance / cleanliness	529	5.27
6	price	207	2.06

Appendix E-2 (for Part IV-A)

– *Positive Service Percentage by Words*

```

select
  keyword,
  count(*) as word_count,
  round(100.0 * count(*) / (select sum(word_count)
    from (
      select count(*) as word_count from review r join business b on r.business_id
= b.business_id
      where r.stars >= 4
      and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
      and (
        r.text like '%friendly%'
        or r.text like '%attentive%'
        or r.text like '%helpful%'
        or r.text like '%welcoming%'
        or r.text like '%great service%'
        or r.text like '%polite%'
        or r.text like '%accommodating%'

```

```

        or r.text like '%amazing%'
        or r.text like '%quick service%'
        or r.text like '%professional%'
        or r.text like '%kind staff%'
    )
    ) total_mentions
), 2) as percentage
from (
    select 'friendly' as keyword from review r join business b on r.business_id = b.business_id
where r.text like '%friendly%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
union all
    select 'attentive' from review r join business b on r.business_id = b.business_id where r.text like
'%attentive%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
union all
    select 'helpful' from review r join business b on r.business_id = b.business_id where r.text like
'%helpful%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
union all
    select 'welcoming' from review r join business b on r.business_id = b.business_id where r.text
like '%welcoming%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
union all
    select 'great service' from review r join business b on r.business_id = b.business_id where r.text
like '%great service%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las
vegas%')
union all
    select 'polite' from review r join business b on r.business_id = b.business_id where r.text like
'%polite%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
union all
    select 'accommodating' from review r join business b on r.business_id = b.business_id where
r.text like '%accommodating%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like
'%las vegas%')
union all
    select 'amazing' from review r join business b on r.business_id = b.business_id where r.text like
'%amazing%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
union all
    select 'quick service' from review r join business b on r.business_id = b.business_id where
r.text like '%quick service%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las
vegas%')
union all

```



```

select 'professional' from review r join business b on r.business_id = b.business_id where r.text
like '%professional%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las
vegas%')
union all
select 'kind staff' from review r join business b on r.business_id = b.business_id where r.text
like '%kind staff%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
) as word_mentions
group by keyword
order by percentage desc;

```

	keyword	word_count	percentage
1	friendly	1637	54.75
2	helpful	521	17.42
3	professional	476	15.92
4	attentive	374	12.51
5	great service	298	9.97
6	accommodating	144	4.82
7	welcoming	118	3.95
8	polite	116	3.88
9	amazing service	34	1.14
10	quick service	19	0.64
11	kind staff	1	0.03

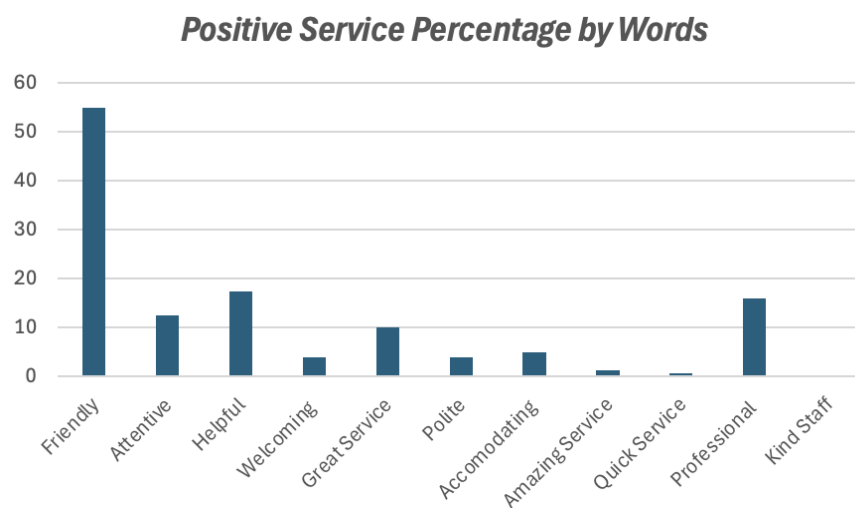


Figure E-2.1

Appendix E-3 (for Part IV-A)

– *Positive Food Percentage by Word*

```

select
  keyword,

```

```

count(*) as word_count,
round(100.0 * count(*) / (select sum(word_count)
    from (
        select count(*) as word_count from review r join business b on r.business_id
= b.business_id
        where r.stars >= 4
        and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
        and (
            r.text like '%delicious%'
            or r.text like '%tasty%'
            or r.text like '%fresh%'
            or r.text like '%yummy%'
            or r.text like '%appetizing%'
            or r.text like '%seasoned well%'
            or r.text like '%perfectly cooked%'
            or r.text like '%crispy%'
            or r.text like '%juicy%'
            or r.text like '%tender%'
            or r.text like '%authentic%'
            or r.text like '%amazing food%'
            or r.text like '%great taste%'
        )
    ) total_mentions
), 2) as percentage
from (
    select 'delicious' as keyword from review r join business b on r.business_id = b.business_id
where r.text like '%delicious%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like
'%las vegas%')
    union all
    select 'tasty' from review r join business b on r.business_id = b.business_id where r.text like
'%tasty%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
    union all
    select 'fresh' from review r join business b on r.business_id = b.business_id where r.text like
'%fresh%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
    union all
    select 'yummy' from review r join business b on r.business_id = b.business_id where r.text like
'%yummy%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
    union all
    select 'appetizing' from review r join business b on r.business_id = b.business_id where r.text
like '%appetizing%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las vegas%')

```

```

union all
select 'seasoned well' from review r join business b on r.business_id = b.business_id where
r.text like '%seasoned well%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las
vegas%')
union all
select 'perfectly cooked' from review r join business b on r.business_id = b.business_id where
r.text like '%perfectly cooked%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like
'%las vegas%')
union all
select 'crispy' from review r join business b on r.business_id = b.business_id where r.text like
'%crispy%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
union all
select 'juicy' from review r join business b on r.business_id = b.business_id where r.text like
'%juicy%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
union all
select 'tender' from review r join business b on r.business_id = b.business_id where r.text like
'%tender%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
union all
select 'authentic' from review r join business b on r.business_id = b.business_id where r.text
like '%authentic%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
union all
select 'amazing food' from review r join business b on r.business_id = b.business_id where
r.text like '%amazing food%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las
vegas%')
union all
select 'great taste' from review r join business b on r.business_id = b.business_id where r.text
like '%great taste%' and r.stars >= 4 and (b.city like '%Las Vegas%' or b.city like '%las vegas%')
) as word_mentions
group by keyword
order by percentage desc;

```

	keyword	word_count	percentage
1	delicious	1282	44.94
2	fresh	819	28.71
3	tasty	491	17.21
4	tender	380	13.32
5	yummy	295	10.34
6	crispy	190	6.66
7	authentic	140	4.91
8	juicy	133	4.66
9	amazing food	41	1.44
10	perfectly cooked	34	1.19
11	great taste	10	0.35
12	seasoned well	9	0.32
13	appetizing	8	0.28

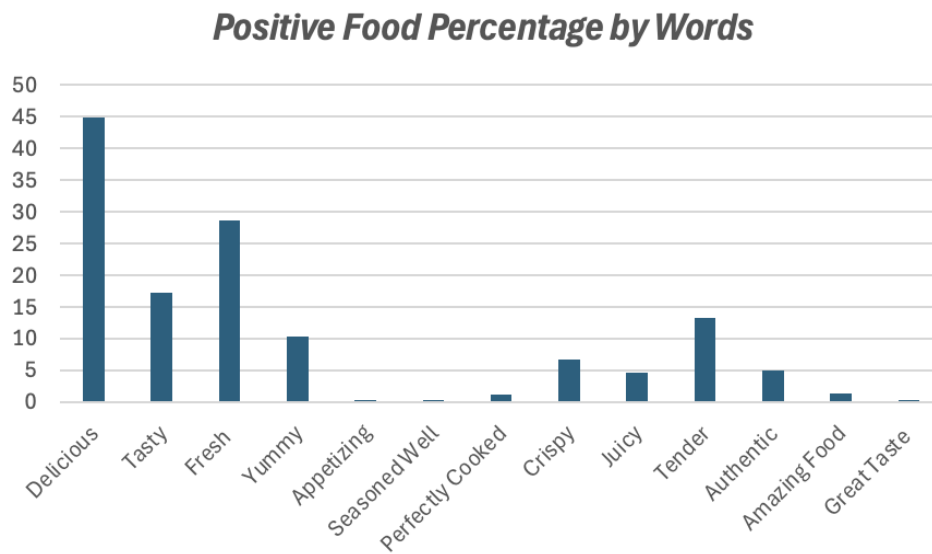


Figure E-3.1

Appendix E-4 (for Part IV-B)

Star Ratings with Negative Comments (with stars >=4)

SELECT category, COUNT(*) AS review_count

FROM(

 SELECT text,

 CASE

 -- Food Quality Issues (Further Expanded)

 WHEN text LIKE '%bland%' OR text LIKE '%cold%' OR text LIKE '%stale%'

 OR text LIKE '%tasteless%' OR text LIKE '%overcooked%' OR text LIKE

'%undercooked%'

 OR text LIKE '%spoiled%' OR text LIKE '%greasy%' OR text LIKE '%dry%'

 OR text LIKE '%burnt%' OR text LIKE '%soggy%' OR text LIKE '%flavorless%'

 OR text LIKE '%salty%' OR text LIKE '%rubbery%' OR text LIKE '%mushy%'

 OR text LIKE '%bad taste%' OR text LIKE '%not fresh%'

 OR text LIKE '%undercooked%' OR text LIKE '%weird taste%' OR text LIKE '%too

salty%'

 OR text LIKE '%too spicy%' OR text LIKE '%lacked seasoning%' OR text LIKE

'%tough%'

 OR text LIKE '%stringy%' OR text LIKE '%gamey%' OR text LIKE '%burned%'

 OR text LIKE '%strange texture%' OR text LIKE '%sour%' OR text LIKE '%rotten%'

 OR text LIKE '%old%' OR text LIKE '%disgusting%' OR text LIKE '%inedible%'

 OR text LIKE '%chemical taste%' OR text LIKE '%off-putting%' OR text LIKE

'%undercooked meat%'

 OR text LIKE '%raw chicken%' OR text LIKE '%frozen inside%' OR text LIKE

'%tasted spoiled%'

 OR text LIKE '%smelled bad%' OR text LIKE '%moldy%' OR text LIKE '%oily%' OR

text LIKE '%watery%'

 THEN 'Food Quality'

 -- Service Issues

 WHEN text LIKE '%slow%' OR text LIKE '%rude%' OR text LIKE '%unfriendly%'

 OR text LIKE '%ignored%' OR text LIKE '%unprofessional%' OR text LIKE

'%attitude%'

 OR text LIKE '%bad service%' OR text LIKE '%waiter never%' OR text LIKE '%forgot

order%'

 THEN 'Service'

 -- Ambiance & Cleanliness Issues

 WHEN text LIKE '%noisy%' OR text LIKE '%dirty%' OR text LIKE '%cramped%'

 OR text LIKE '%bad lighting%' OR text LIKE '%smelled%' OR text LIKE '%sticky%'

 OR text LIKE '%uncomfortable seating%' OR text LIKE '%too dark%' OR text LIKE

'%too bright%'

```

        THEN 'Ambiance / Cleanliness'
    -- Price Complaints
    WHEN text LIKE '%expensive%' OR text LIKE '%overpriced%' OR text LIKE '%not
worth%'
        OR text LIKE '%small portion%' OR text LIKE '%charged too much%' OR text LIKE
'%hidden fees%'
        THEN 'Price'
    -- Wait Time Issues
    WHEN text LIKE '%long wait%' OR text LIKE '%took forever%' OR text LIKE '%too
long%'
        OR text LIKE '%waiting for%' OR text LIKE '%delayed%' OR text LIKE '%had to
wait%'
        THEN 'Wait Time'
    -- If it doesn't match, mark for review instead of "Other"
    ELSE 'Needs Review'
END AS category
FROM review
WHERE stars >= 4
) AS categorized_reviews
GROUP BY category
ORDER BY review_count DESC;

```

Appendix E-5 (for for Part IV-B)

Create a Temporary Table

```

--Category Percentages--
CREATE TABLE negative_review_analysis AS
SELECT category, COUNT(*) AS review_count
FROM(
    SELECT text,
        CASE
            -- Food Quality Issues (Further Expanded)
            WHEN text LIKE '%bland%' OR text LIKE '%cold%' OR text LIKE '%stale%'
                OR text LIKE '%tasteless%' OR text LIKE '%overcooked%' OR text LIKE
'%undercooked%'
                OR text LIKE '%spoiled%' OR text LIKE '%greasy%' OR text LIKE '%dry%'
                OR text LIKE '%burnt%' OR text LIKE '%soggy%' OR text LIKE '%flavorless%'
                OR text LIKE '%salty%' OR text LIKE '%rubbery%' OR text LIKE '%mushy%'
                OR text LIKE '%bad taste%' OR text LIKE '%not fresh%'

```

OR text LIKE '%undercooked%' OR text LIKE '%weird taste%' OR text LIKE '%too salty%'

OR text LIKE '%too spicy%' OR text LIKE '%lacked seasoning%' OR text LIKE '%tough%'

OR text LIKE '%stringy%' OR text LIKE '%gamey%' OR text LIKE '%burned%'

OR text LIKE '%strange texture%' OR text LIKE '%sour%' OR text LIKE '%rotten%'

OR text LIKE '%old%' OR text LIKE '%disgusting%' OR text LIKE '%inedible%'

OR text LIKE '%chemical taste%' OR text LIKE '%off-putting%' OR text LIKE '%undercooked meat%'

OR text LIKE '%raw chicken%' OR text LIKE '%frozen inside%' OR text LIKE '%tasted spoiled%'

OR text LIKE '%smelled bad%' OR text LIKE '%moldy%' OR text LIKE '%oily%' OR text LIKE '%watery%'

THEN 'Food Quality'

-- Service Issues

WHEN text LIKE '%slow%' OR text LIKE '%rude%' OR text LIKE '%unfriendly%'

OR text LIKE '%ignored%' OR text LIKE '%unprofessional%' OR text LIKE '%attitude%'

OR text LIKE '%bad service%' OR text LIKE '%waiter never%' OR text LIKE '%forgot order%'

THEN 'Service'

-- Ambiance & Cleanliness Issues

WHEN text LIKE '%noisy%' OR text LIKE '%dirty%' OR text LIKE '%cramped%'

OR text LIKE '%bad lighting%' OR text LIKE '%smelled%' OR text LIKE '%sticky%'

OR text LIKE '%uncomfortable seating%' OR text LIKE '%too dark%' OR text LIKE '%too bright%'

THEN 'Ambiance / Cleanliness'

-- Price Complaints

WHEN text LIKE '%expensive%' OR text LIKE '%overpriced%' OR text LIKE '%not worth%'

OR text LIKE '%small portion%' OR text LIKE '%charged too much%' OR text LIKE '%hidden fees%'

THEN 'Price'

-- Wait Time Issues

WHEN text LIKE '%long wait%' OR text LIKE '%took forever%' OR text LIKE '%too long%'

OR text LIKE '%waiting for%' OR text LIKE '%delayed%' OR text LIKE '%had to wait%'

THEN 'Wait Time'

-- If it doesn't match, mark for review instead of "Other"

```

        ELSE 'Needs Review'
    END AS category
FROM review
WHERE stars >= 4
) AS categorized_reviews
GROUP BY category
ORDER BY review_count DESC;

```

Categorized Percentages with Negative Reviews

```

SELECT
    category,
    review_count,
    ROUND(100.0 * review_count / SUM(review_count) OVER (), 2) AS percentage
FROM negative_review_analysis
ORDER BY review_count DESC;

```

	category	review_count	percentage
1	Needs Review	25356	78.28
2	Food Quality	5000	15.44
3	Price	696	2.15
4	Service	605	1.87
5	Wait Time	406	1.25
6	Ambiance / Cleanliness	329	1.02

Figure E-5.1

Appendix E-6 (for for Part IV-B)

--Keywords Analysis in Percentages--

--Keywords Percentages--

```

WITH FoodQualityReviews AS (
    SELECT r.text
    FROM review r
    JOIN business b ON r.business_id = b.business_id
    WHERE r.stars >= 4
    AND b.city = 'Las Vegas'
    AND (

```



```

r.text LIKE '%bland%' OR r.text LIKE '%cold%' OR r.text LIKE '%stale%'
OR r.text LIKE '%tasteless%' OR r.text LIKE '%overcooked%' OR r.text LIKE
'%undercooked%'
OR r.text LIKE '%spoiled%' OR r.text LIKE '%greasy%' OR r.text LIKE '%dry%'
OR r.text LIKE '%burnt%' OR r.text LIKE '%soggy%' OR r.text LIKE '%flavorless%'
OR r.text LIKE '%salty%' OR r.text LIKE '%rubbery%' OR r.text LIKE '%mushy%'
OR r.text LIKE '%bad taste%' OR r.text LIKE '%not fresh%'
OR r.text LIKE '%weird taste%' OR r.text LIKE '%too salty%'
OR r.text LIKE '%too spicy%' OR r.text LIKE '%lacked seasoning%' OR r.text LIKE
'%tough%'
OR r.text LIKE '%stringy%' OR r.text LIKE '%gamey%' OR r.text LIKE '%burned%'
OR r.text LIKE '%strange texture%' OR r.text LIKE '%sour%' OR r.text LIKE '%rotten%'
OR r.text LIKE '%old%' OR r.text LIKE '%disgusting%' OR r.text LIKE '%inedible%'
OR r.text LIKE '%chemical taste%' OR r.text LIKE '%off-putting%' OR r.text LIKE
'%undercooked meat%'
OR r.text LIKE '%raw chicken%' OR r.text LIKE '%frozen inside%' OR r.text LIKE
'%tasted spoiled%'
OR r.text LIKE '%smelled bad%' OR r.text LIKE '%moldy%' OR r.text LIKE '%oily%' OR
r.text LIKE '%watery%'

```

```

)
)
SELECT keyword,
COUNT(*) AS keyword_count,
ROUND(100.0 * COUNT(*) / (SELECT COUNT(*) FROM FoodQualityReviews), 2) AS
percentage
FROM (
SELECT 'bland' AS keyword FROM FoodQualityReviews WHERE text LIKE '%bland%'
UNION ALL
SELECT 'cold' FROM FoodQualityReviews WHERE text LIKE '%cold%'
UNION ALL
SELECT 'stale' FROM FoodQualityReviews WHERE text LIKE '%stale%'
UNION ALL
SELECT 'tasteless' FROM FoodQualityReviews WHERE text LIKE '%tasteless%'
UNION ALL
SELECT 'overcooked' FROM FoodQualityReviews WHERE text LIKE '%overcooked%'
UNION ALL
SELECT 'undercooked' FROM FoodQualityReviews WHERE text LIKE '%undercooked%'
UNION ALL
SELECT 'spoiled' FROM FoodQualityReviews WHERE text LIKE '%spoiled%'
UNION ALL

```

```
SELECT 'greasy' FROM FoodQualityReviews WHERE text LIKE '%greasy%'
UNION ALL
SELECT 'dry' FROM FoodQualityReviews WHERE text LIKE '%dry%'
UNION ALL
SELECT 'burnt' FROM FoodQualityReviews WHERE text LIKE '%burnt%'
UNION ALL
SELECT 'soggy' FROM FoodQualityReviews WHERE text LIKE '%soggy%'
UNION ALL
SELECT 'flavorless' FROM FoodQualityReviews WHERE text LIKE '%flavorless%'
UNION ALL
SELECT 'salty' FROM FoodQualityReviews WHERE text LIKE '%salty%'
UNION ALL
SELECT 'rubbery' FROM FoodQualityReviews WHERE text LIKE '%rubbery%'
UNION ALL
SELECT 'mushy' FROM FoodQualityReviews WHERE text LIKE '%mushy%'
UNION ALL
SELECT 'bad taste' FROM FoodQualityReviews WHERE text LIKE '%bad taste%'
UNION ALL
SELECT 'not fresh' FROM FoodQualityReviews WHERE text LIKE '%not fresh%'
UNION ALL
SELECT 'weird taste' FROM FoodQualityReviews WHERE text LIKE '%weird taste%'
UNION ALL
SELECT 'too salty' FROM FoodQualityReviews WHERE text LIKE '%too salty%'
UNION ALL
SELECT 'too spicy' FROM FoodQualityReviews WHERE text LIKE '%too spicy%'
UNION ALL
SELECT 'lacked seasoning' FROM FoodQualityReviews WHERE text LIKE '%lacked
seasoning%'
UNION ALL
SELECT 'tough' FROM FoodQualityReviews WHERE text LIKE '%tough%'
UNION ALL
SELECT 'stringy' FROM FoodQualityReviews WHERE text LIKE '%stringy%'
UNION ALL
SELECT 'gamey' FROM FoodQualityReviews WHERE text LIKE '%gamey%'
UNION ALL
SELECT 'burned' FROM FoodQualityReviews WHERE text LIKE '%burned%'
UNION ALL
SELECT 'strange texture' FROM FoodQualityReviews WHERE text LIKE '%strange
texture%'
UNION ALL
```

```

SELECT 'sour' FROM FoodQualityReviews WHERE text LIKE '%sour%'
UNION ALL
SELECT 'rotten' FROM FoodQualityReviews WHERE text LIKE '%rotten%'
UNION ALL
SELECT 'old' FROM FoodQualityReviews WHERE text LIKE '%old%'
UNION ALL
SELECT 'disgusting' FROM FoodQualityReviews WHERE text LIKE '%disgusting%'
UNION ALL
SELECT 'inedible' FROM FoodQualityReviews WHERE text LIKE '%inedible%'
UNION ALL
SELECT 'chemical taste' FROM FoodQualityReviews WHERE text LIKE '%chemical taste%'
UNION ALL
SELECT 'off-putting' FROM FoodQualityReviews WHERE text LIKE '%off-putting%'
UNION ALL
SELECT 'undercooked meat' FROM FoodQualityReviews WHERE text LIKE '%undercooked
meat%'
UNION ALL
SELECT 'raw chicken' FROM FoodQualityReviews WHERE text LIKE '%raw chicken%'
UNION ALL
SELECT 'frozen inside' FROM FoodQualityReviews WHERE text LIKE '%frozen inside%'
UNION ALL
SELECT 'tasted spoiled' FROM FoodQualityReviews WHERE text LIKE '%tasted spoiled%'
UNION ALL
SELECT 'smelled bad' FROM FoodQualityReviews WHERE text LIKE '%smelled bad%'
UNION ALL
SELECT 'moldy' FROM FoodQualityReviews WHERE text LIKE '%moldy%'
UNION ALL
SELECT 'oily' FROM FoodQualityReviews WHERE text LIKE '%oily%'
UNION ALL
SELECT 'watery' FROM FoodQualityReviews WHERE text LIKE '%watery%'
) AS keyword_counts
GROUP BY keyword
ORDER BY keyword_count DESC;

```

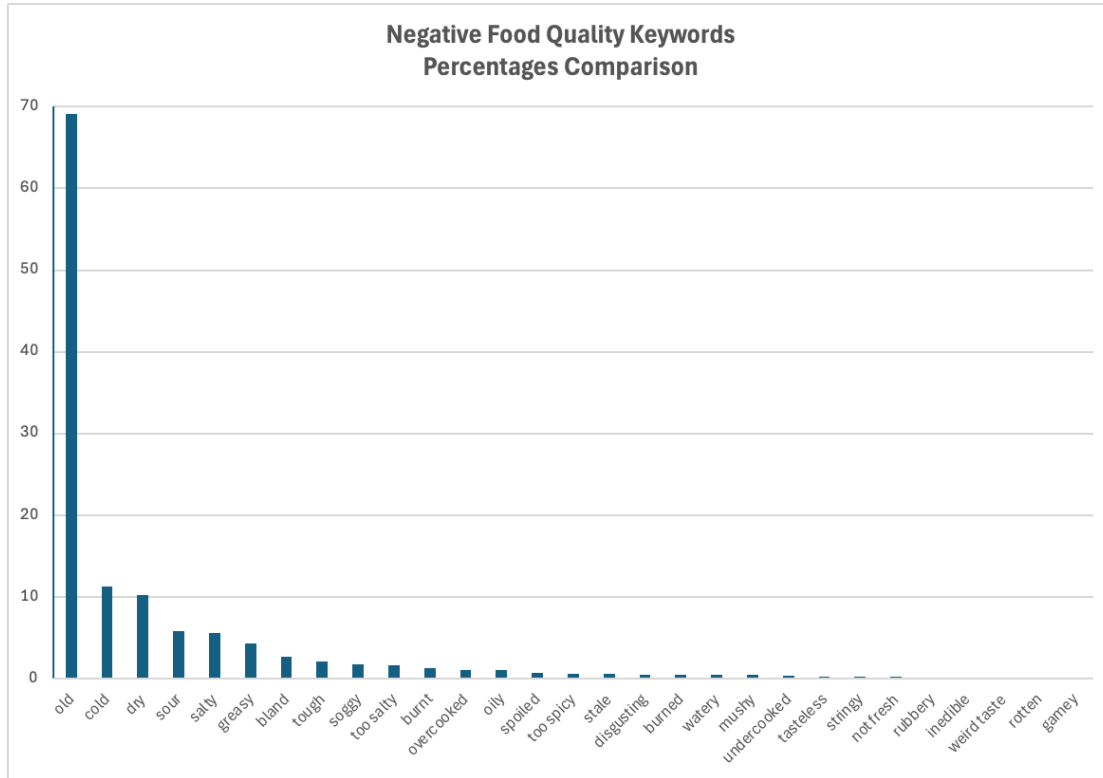


Figure E-6