El Rango

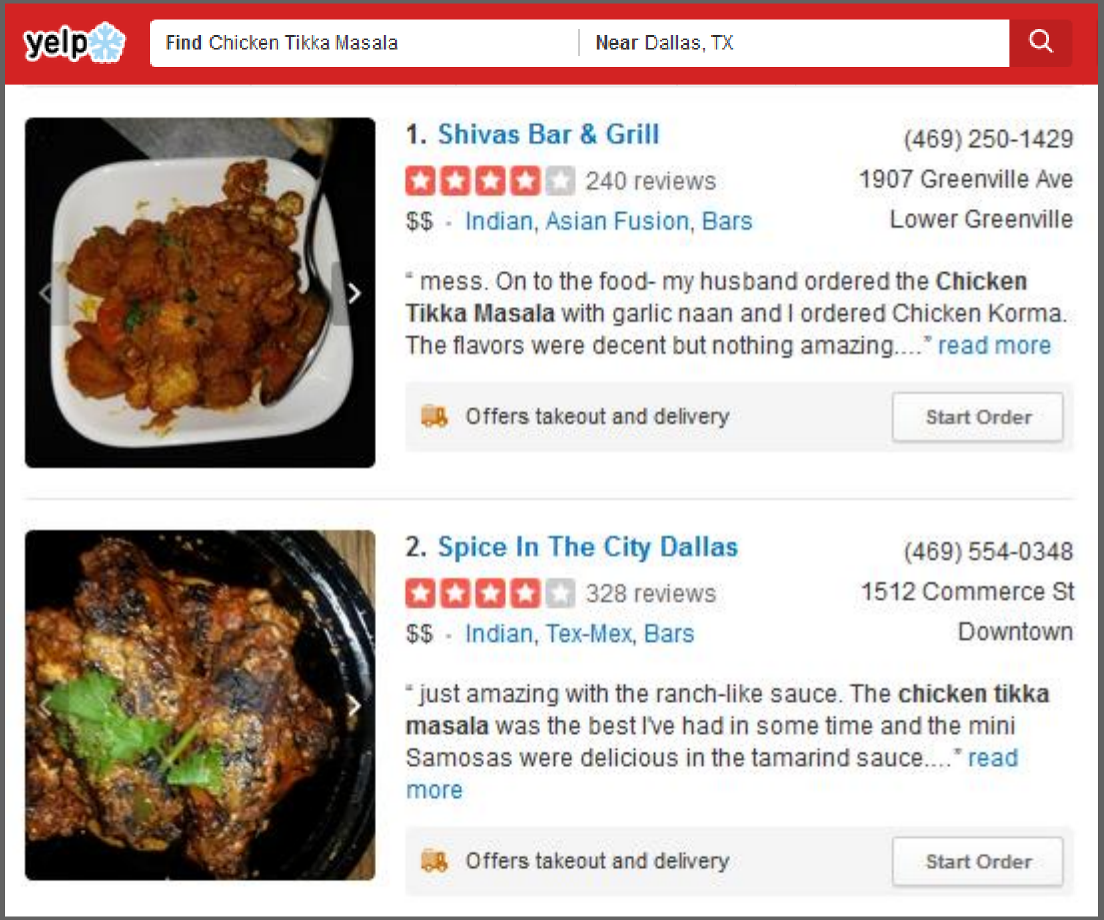
Ranking Restaurants by Dishes

Group 11 Section 2
Ibukun Phillips
Ananthapadmanabhan Sivasankaran
Yash Sharma
Muthuraja Palaniappan

Image credit: https://newsin.asia/sri-lanka-gets-tough-restaurants-maintain-hygiene-quality-service-food/

Why rank by dishes?

- Yelp, Zomato, Zagat, TripAdvisor
  - List by cuisines
  - Ranked based on ratings & distance
  - Search by dish only based on mentions
- Customer can’t find restaurants rated highly on a dish
- Restaurants can increase revenues with dish based ranking

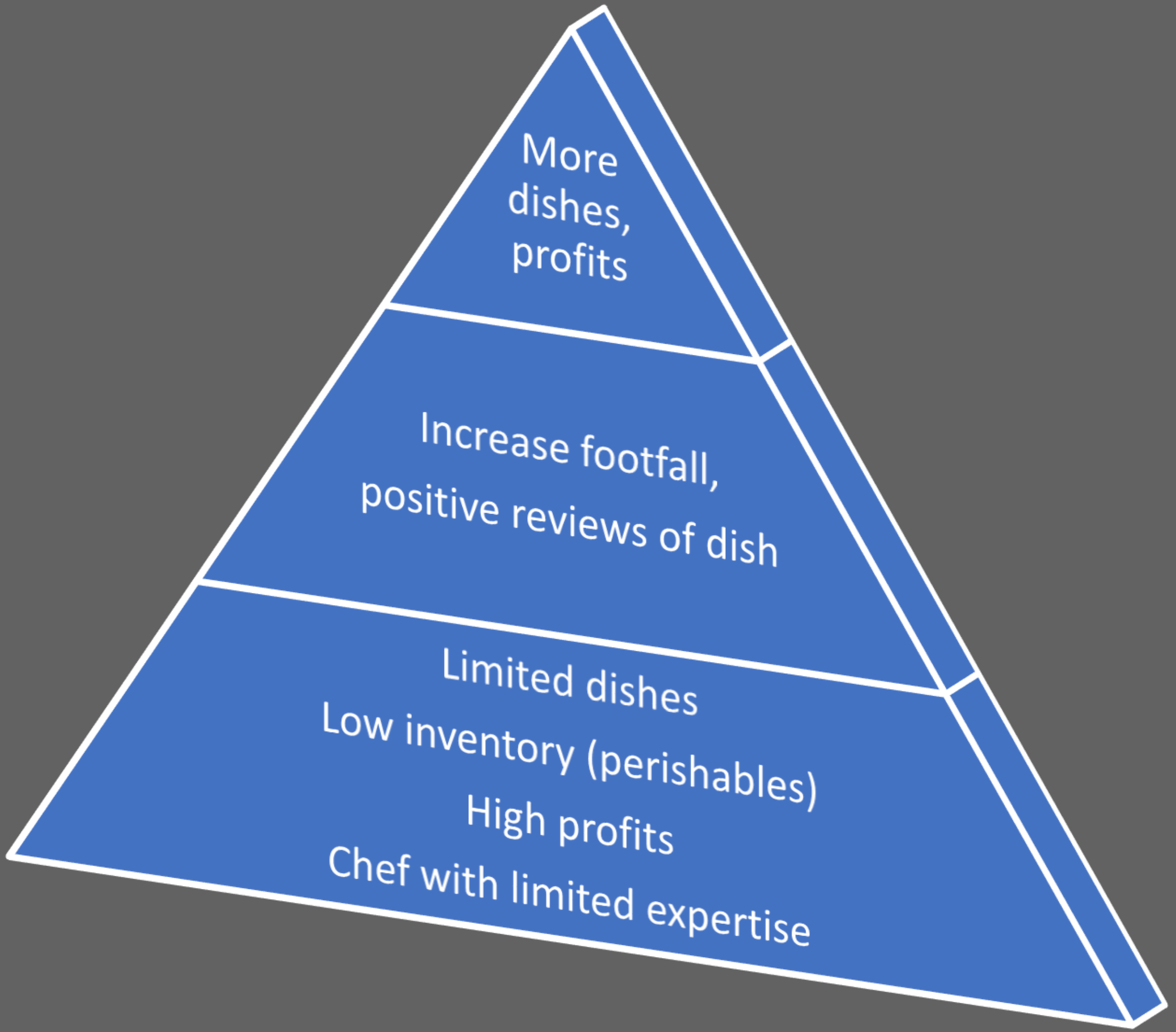


How El Rango helps the niche players differentiate

- Specialize in few dishes (Food costs account for 28-35% costs of restaurant*)
- Niche / new / small players will be able to compete against Goliaths
- Our analysis shows that restaurant revenues are significantly impacted due to dish sentiments

* source: https://smallbusiness.chron.com/common-food-labor-cost-percentages-14700.html

Enabling growth of restaurants



Business Model of El Rango

- One site to rank by dish reviews on Yelp, Zomato, Zagat, TripAdvisor, etc.
- Increase website traffic by
  - Enabling customers to comment on rankings on our Facebook page
  - Reducing dependence on Yelp, et al. by allowing customers to suggest good restaurants by dish on our site
  - Promoting through influencers (food bloggers, local tabloids, etc)
- Earn through
  - Advertisements of new / niche restaurants
  - Commissions on discount coupons sold on website

Observation

- Studies show ratings can reflect the sentiments of the corresponding reviews*
- TripAdvisor's restaurant rating reflects the aggregate of food, value and service ratings

Hypothesis : Higher Dish sentiment will lead to a higher overall rating (or revenues) for a restaurant

- We consider overall rating as a proxy for revenues

* Chen & Xu (2016)

Data Sources

- Web scraping and Text Mining were used in obtaining needed data from TripAdvisor
- Data was obtained for all San Francisco restaurants (hypothesis is tested for SF)
- Data obtained:

Overall rating	Review date
Review title	Dish Sentiment (derived from NLP)
Review content	Restaurant rating
Restaurant food rating	Restaurant service rating
Restaurant value rating	Price

Finding Popular dishes

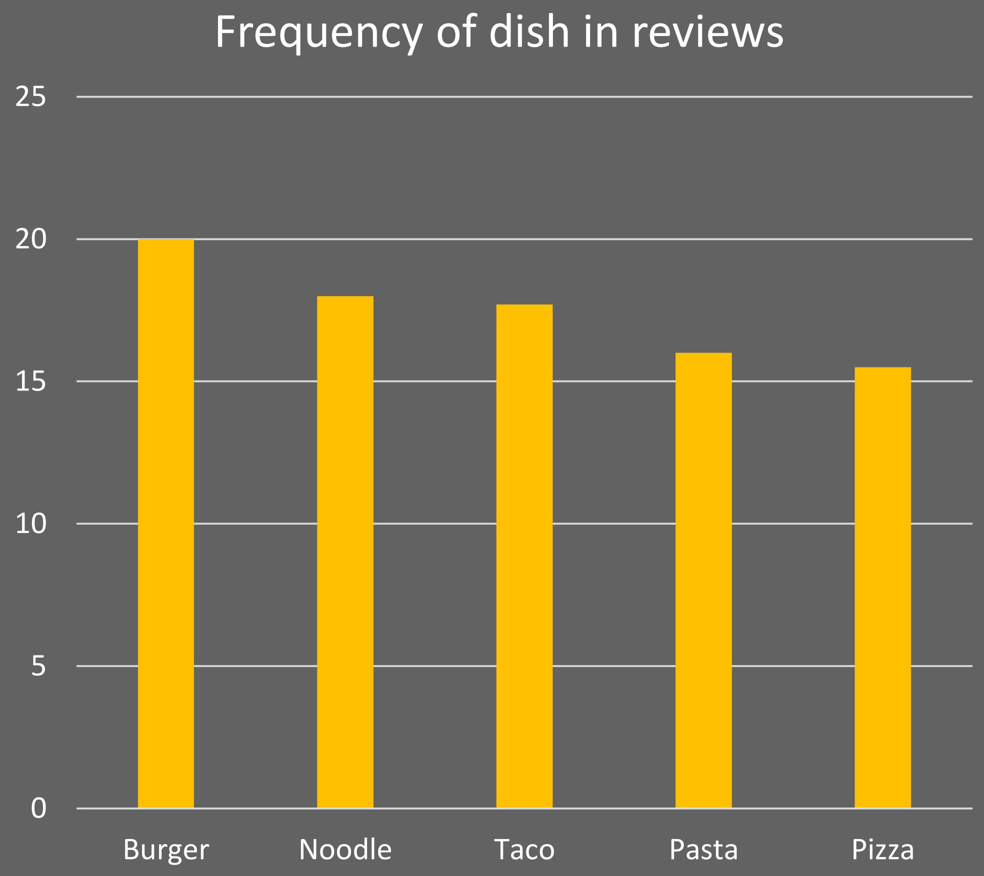
- We found the dishes that were mentioned the most in reviews by analyzing the frequency distribution table.

  a. Use the sysnet library to find food related keywords

  b. Draw the frequency distribution

  c. Find the top dishes

Food frequency distribution



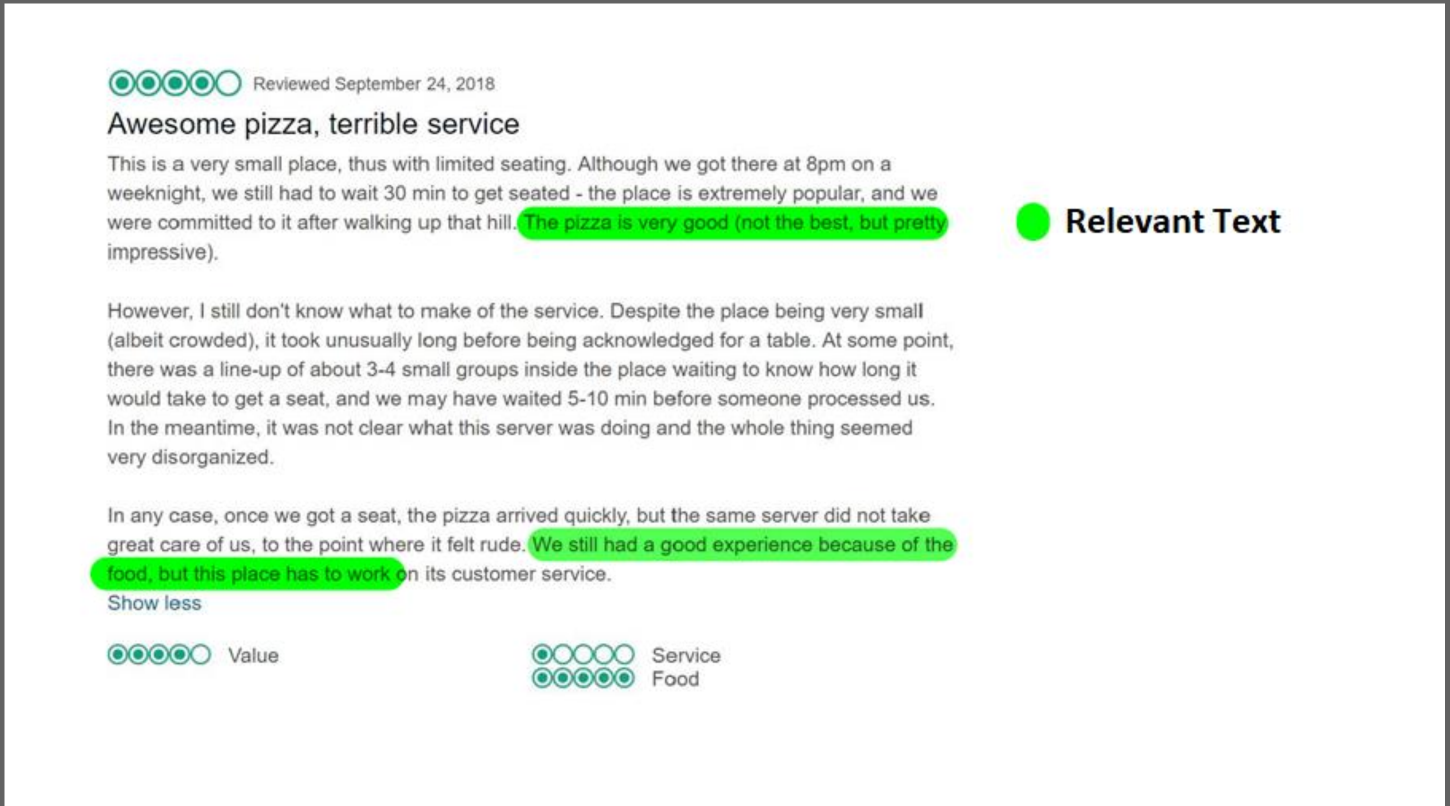
Most reviewed Dishes:

1. Burger
2. Noodle
3. Taco
4. Pasta
5. Pizza

Filtering relevant data

- The review contains text not relevant to dish sentiment such as
  - Ambience
  - Location
  - Delivery Service
  - Staff friendliness
  - Opening hours
  - ….

Sentiment analysis about dish


Reviewed September 24, 2018

Awesome pizza, terrible service

This is a very small place, thus with limited seating. Although we got there at 8pm on a weeknight, we still had to wait 30 min to get seated - the place is extremely popular, and we were committed to it after walking up that hill. The pizza is very good (not the best, but pretty impressive).

However, I still don't know what to make of the service. Despite the place being very small (albeit crowded), it took unusually long before being acknowledged for a table. At some point, there was a line-up of about 3-4 small groups inside the place waiting to know how long it would take to get a seat, and we may have waited 5-10 min before someone processed us. In the meantime, it was not clear what this server was doing and the whole thing seemed very disorganized.

In any case, once we got a seat, the pizza arrived quickly, but the same server did not take great care of us, to the point where it felt rude. We still had a good experience because of the food, but this place has to work on its customer service.

Show less

Value

Service

Food

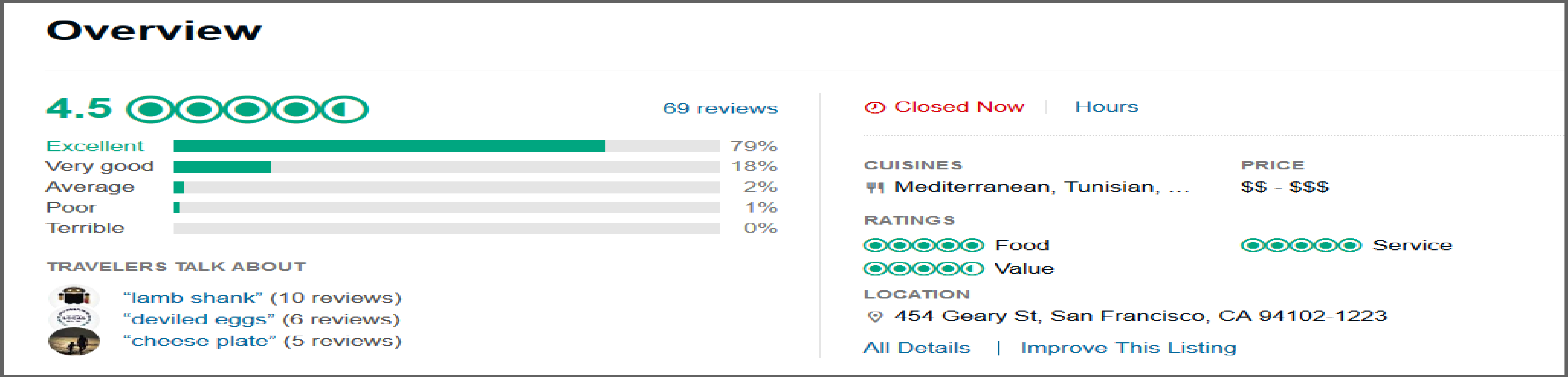
Relevant Text

This photo is taken from a review in TripAdvisor

We select only the sentences which have the required keyword mentioned at least once and then analyze its sentiments.

Factors currently considered by TA

Parameter	Coefficient	Significance
Intercept	0.98	High
Food rating	0.38	High
Service rating	0.24	High
Value rating	0.12	High
Price	0.03	Medium



This photo is taken from a restaurant review in TripAdvisor

Empirical Model

- The multiple linear regression model below was used to find the relationship between the independent variables and the dependent (overall rating)

$$\text{overall rating} = \beta_0 + \beta_1 * \text{price} + \beta_2 * \text{dish sentiment} + \beta_3 * \text{value rating} + \beta_4 * \text{service rating} + \beta_5 * \text{adjusted food rating}$$

Results of statistical tests

- Sentiment about dish significantly affects the overall rating while other factors are controlled (10% significance level)
- Price is less significant compared to other factors
- For few of the dishes, food rating is significant while dish sentiment is not
  - This implies the other foods in the restaurants have higher effect on overall rating

Regression results for Burger

Parameter	Coefficient	Significance
Intercept	0.86	High
Adjusted Food rating	0.37	High
Service rating	0.23	High
Value rating	0.17	High
Price	0.05	Medium
Burger	0.21	Medium

The dish (Burger) explains the change in overall rating

Note: Food rating is adjusted to remove the effect of dish sentiment

Regression results for Taco

Parameter	Coefficient	Significance
Intercept	1.2	High
Adjusted Food rating	0.33	High
Service rating	0.14	Medium
Value rating	0.20	High
Price	0.71	Low
Taco	0.29	Low

The dish (Taco) explains the change in overall rating

Note: Food rating is adjusted to remove the effect of dish sentiment

Suggested business actions

Focus on specialty dishes

Increase footfall
Reduce inventory costs

After popularity increases, extend the number of menu items
Increase footfall further

Advertise in El Rango
Popularize El Rango to customers to derive mutual benefit

Business actions & impact

- Hire chef who specializes on certain items
- Improve quality of limited food items
- Get good reviews on El Rango
- Increase revenues
- Improve ambience, service

Financial impact

- Tangibles
  - Revenues will increase (due to increase in restaurant rating in food review sites)
  - Costs will decrease owing to focus on fewer products (low inventory)

- Intangibles
  - Improved brand image
  - Increase in word-of-mouth popularity through El Rango
  - Cost-free marketing through El Rango rankings

Limitations

- Data
  - In future, we can collect larger data sources such as Yelp, Zagat, Zomato, Google+ as well. Collecting from one source might lead to convenience sampling
  - Limited features review rating data (food, service, ambience and values). Not all reviews provided full details of these
  - Location of the restaurant, parking facilities, etc. might be important features
- Approach
  - Sentiments derived from sentences rather than phrases
    - For ex, “pizza is good, but, service is bad…”

Gracias!

References

- Chen R., & Xu W. (2016). The determinants of online customer ratings. Electronic Commerce Research, 17(1), 31–50. https://link.springer.com/article/10.1007/s10660-016-9243-6
- Buckley, Steven. (June 29, 2018). Common Food & Labor Cost Percentages. https://smallbusiness.chron.com/common-food-labor-cost-percentages-14700.html
- Luca M. (2016). Reviews, Reputation and Revenue: The Case of Yelp. https://www.hbs.edu/faculty/Publication%20Files/12-016_a7e4a5a2-03f9-490d-b093-8f951238dba2.pdf

Appendix 1

```
Call:
lm(formula = Overall_rating ~ restaurant_food_rating + restaurant_service_rating +
    restaurant_value_rating + Price_fig + burger, data = deduped_data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.95657 -0.12803 -0.00233  0.14682  0.63576

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                0.85776    0.11088   7.736 5.71e-14 ***
restaurant_food_rating     0.37209    0.03556  10.463  < 2e-16 ***
restaurant_service_rating  0.22603    0.03291   6.868 1.95e-11 ***
restaurant_value_rating    0.16643    0.03273   5.085 5.21e-07 ***
Price_fig                  0.04533    0.01545   2.933  0.00351 **
burger                     0.20881    0.07379   2.830  0.00484 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2139 on 499 degrees of freedom
  (1087 observations deleted due to missingness)
Multiple R-squared:  0.6409,    Adjusted R-squared:  0.6373
F-statistic: 178.1 on 5 and 499 DF,  p-value: < 2.2e-16
```

Appendix 2

```
Call:
lm(formula = Overall_rating ~ restaurant_food_rating + restaurant_service_rating +
    restaurant_value_rating + Price_fig + taco, data = deduped_data)

Residuals:
    Min      1Q  Median      3Q     Max
-0.7963 -0.1123  0.0101  0.1404  0.5360

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                1.19950    0.17748   6.759 1.09e-10 ***
restaurant_food_rating     0.32717    0.05674   5.766 2.54e-08 ***
restaurant_service_rating  0.14181    0.05102   2.779  0.00589 **
restaurant_value_rating    0.20106    0.05074   3.962 9.85e-05 ***
Price_fig                  0.07171    0.02834   2.531  0.01204 *
taco                       0.29169    0.11562   2.523  0.01231 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2195 on 235 degrees of freedom
  (1351 observations deleted due to missingness)
Multiple R-squared:  0.5625,	Adjusted R-squared:  0.5532
F-statistic: 60.43 on 5 and 235 DF,  p-value: < 2.2e-16
```