Ratty Review Algorithms

Akshay Mehta

April 2023

1 Breakfast/Lunch/Dinner Scores

The data used to calculate the Breakfast/Lunch/Dinner scores are the average of all the reviews under their category. In other words, if Dinner had the foods, "Jambalaya", "Pizza", and "Pho", with the respective scores 3, 4, 5, then the dinner score would be the average of these three values such that:
Dinner Score $= \frac{3+4+5}{3} = 4$.

2 Individual Meal Scores

We decided not to make the individual meal scores on an averaging scale. The problem with review averages (what Amazon uses), is that it more highly places meals with few high reviews as compared to those with many reviews of the same proportion (i.e. 1 good review for 6 total = 10 good reviews for 60 total). For The Ratty Reviews, to score each food item, we have decided to use the Lower bound of Wilson score confidence interval for a Bernoulli parameter. The Bernoulli lower bound calculation, takes in creates a binomial distribution with our chosen 95% confidence interval and calculates the effective probability of success within the given population. The equation to calculate the Bernoulli lower bound is:

$$lower\_bound = \frac{p + \frac{z^2}{2n} - z\sqrt{\frac{p(1-p)}{n} + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}}.$$

p is the proportion of successes in the sample, n is the sample size, and z is the standard score corresponding to the desired level of confidence. The pitfall of this equation is it assigns items that are poor quality and that have no ratings both a value of 0. Our code avoids this as it utilizes an enumeration value to the rating scale, such that there is a value of none or rating (rating containing 1-5), so that it can parse that data without using a null operator. Since the formula, calculates a lower bound of confidence in a binomial distribution, its lowest value possible is 0 and highest 0.5, meaning if we take this value times 10, we get the 0-5 star rating scale. Thus, our equation for calculating ratings is

$$lower\_bound * 10 = rating$$