# **Statistical Analysis for Capstone 1**

As we established in EDA over 80% of the reviews are 4 and 5 stars:

A screenshot of a cell phone

Description automatically generated

We also showed the nutritional values correlation matrix:

A screenshot of a cell phone

Description automatically generated

There is a strong correlation between calories and the total fats (specifically saturated fats), protein and carbohydrates contents. However, there is no strong correlation between the recipe rating and any of the mentioned recipe attributes.

We further investigated the data in EDA and have the following findings:

1. The best ingredient is garlic. It has the average recipe ranking of 4.453
2. The worst ingredient is baking powder. It has the average recipe ranking of 4.388
3. The best tag is 'beijing'. It has the average recipe ranking of 5.000
4. The worst tag is 'pressure-canning'. It has the average recipe ranking of 2.981
5. Successful recipes on average take longer to make and include 10 or more steps
6. In terms of nutritional value, successful recipes have more total fats, sugars and carbohydrates, but less sodium and saturated fats, and slightly less protein.

We performed two sample t-tests to confirm our EDA findings:  
  
1) The two independent samples were the recipes with a specific attribute and the recipes without it

2) We tested the null hypothesis 𝐻0 that the mean recipe rating for the two samples is identical  
  
2) The alternative hypothesis 𝐻𝑎 was that the rating means are different (e.g. the mean recipe rating is affected by a specific recipe attribute).  
  
3) If the t-test results were statistically significant (e.g. p-value > 𝑎, 𝑎 = 0.05), then we rejected the 𝐻0 and accepted the 𝐻𝑎.

As the result of the t-tests:

1. **Garlic** The test had p-value of 9.475783077096418e-18 which is very small, so we can reject the 𝐻0 and confirm that garlic in recipes contribute to a higher recipe rating
2. **Baking Powder** The test had p-value of 9.475783077096418e-18 which is very small, so we can reject the 𝐻0 and confirm that baking powder in recipes contribute to a lower recipe rating
3. ‘Beijing’ recipe tag The test had p-value of 0.13 > 𝑎a = 0.05, so we can reject the 𝐻𝑎. This means that 'beijing' tag positive effect on the recipe rating is statistically not significant.
4. **'pressure-canning' recipe tag** The test had p-value = 2.792736995754129e-10 which is very small, so we can reject the 𝐻𝑜. This means that 'pressure-canning' tag negative effect on the recipe rating is statistically significant
5. **Saturated Fats Content** was the only other attribute that affects the recipe rating in statistically significant way (the test p-value was 0.03). All the other recipe attributes do not significantly affect the recipe rating.