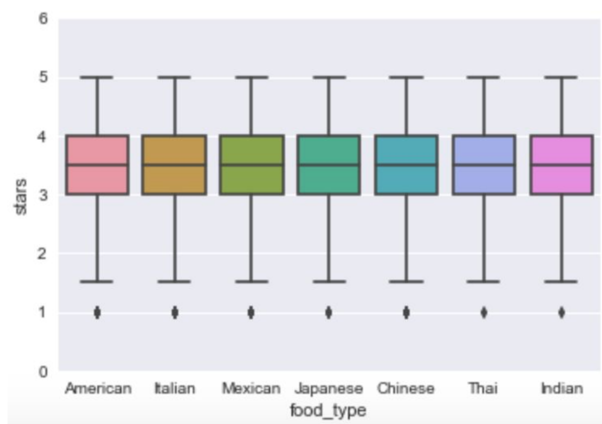
Data Science Capstone Project
Springboard Bootcamp, San Francisco
Reza Taeb
Spring 2018
Title : “Improving Restaurant Reputation Using Yelp User Reviews”
Part 3: Inferential Statistic

**How to improve restaurant reputation even just one star using Yelp user reviews !**
**“ The path from being normal to good, and good to perfect !"**

Based on previous part (storytelling), in this part I am trying to get more insight based on the statistics:

**1 - "What types of food do usually get higher ratings ? Is the difference across types of food in ratings statistically significant or not ? "**

| food_type | mean | std | count |
|---|---|---|---|
| American | 3.402928 | 0.713942 | 8401 |
| Chinese | 3.286468 | 0.689724 | 3318 |
| Indian | 3.579404 | 0.690748 | 1209 |
| Italian | 3.478221 | 0.773531 | 3788 |
| Japanese | 3.582657 | 0.662095 | 1972 |
| Mexican | 3.380827 | 0.768168 | 3604 |
| Thai | 3.582406 | 0.707131 | 989 |

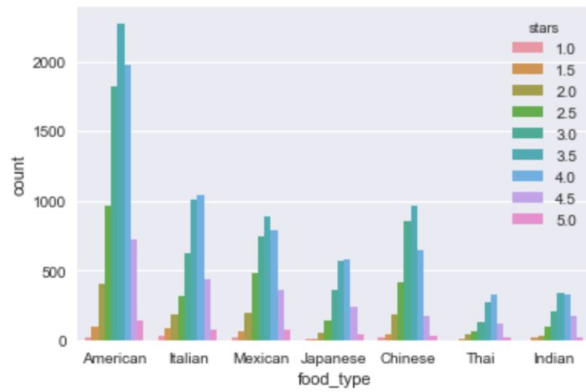



At the first glance, the above charts and table show that there is no significant difference across "type of food" in ratings. However it should be tested with some statistical tests.
Since the distributions are not normal (“stats.normaltest” in python), and ANOVA test could not be used and instead **Kruskal-Wallis**[1] was applied to check whether type of food affects the reviewers' ratings or not.
In this question, the null hypothesis is that there is no difference between the means of the ratings across different types of food.
Since the P-value of this test was very small (e-78). so we can reject the null hypothesis and we can say the **food type can impact the restaurant ratings.**

[1] *Kruskal-Wallis is used to compare the means of three or more samples when the sample distribution is not normal.*

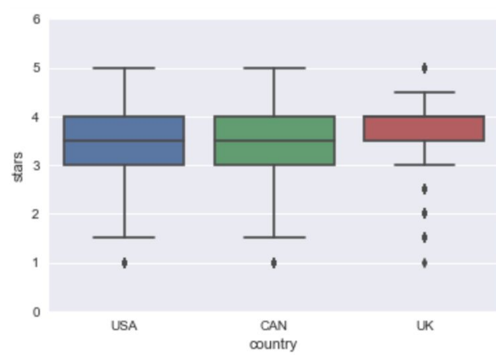
If the test has be done separately, we can notice that the "indian" , "thai" and "japanese" foods (highest earned ratings) are very similar, also "american" and "mexican" are similar in terms of the ratings that they received on average.

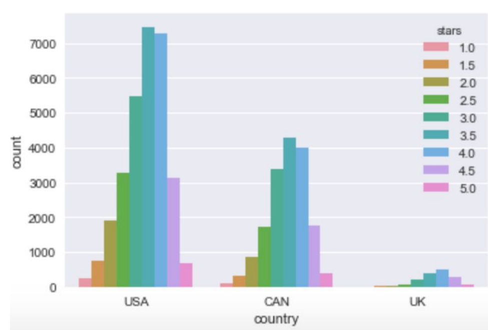
In conclusion, users tend to give the better ratings to 'indian', 'that' and 'japanese' restaurants and type of food matters in restaurant ratings.




**2 - "In which countries do users tend to give higher ratings to restaurants ? Does the "country" in which restaurant located impact the ratings that it is received and if so is this difference statistically significant or not ? "**

| country | mean | std | count |
|---|---|---|---|
| CAN | 3.426333 | 0.769820 | 16819 |
| UK | 3.765655 | 0.686447 | 1549 |
| USA | 3.395069 | 0.803634 | 30196 |



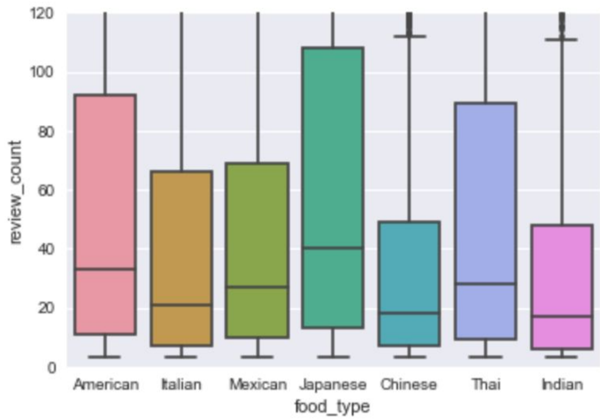




Like the previous part the ***Kruskal-Wallis*** test was used in this part.

The p-value is very small ( $p < 0.05$) so we can reject the null hypothesis (in all countries users tend to give similar ratings), and we could say ***the place of restaurant can impact the ratings*** that it received.

It seems that people in UK tend to give higher ratings to restaurants compared to the people in US and Canada.

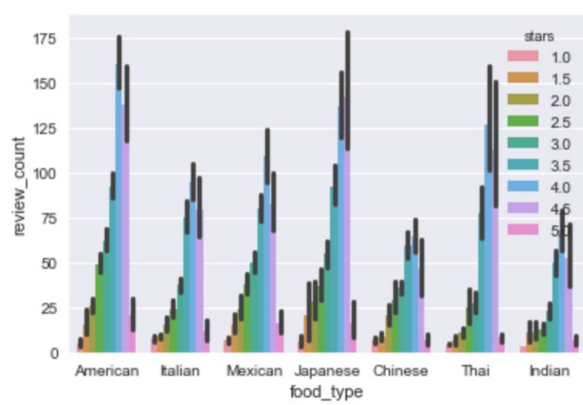
### 3 - "Which type of food has higher review_counts ? Is there a statistical difference between different types of food? "




| | mean | std | count | median |
|---|---|---|---|---|
| food_type | | | | |
| American | 95.094036 | 213.269908 | 8401 | 33 |
| Chinese | 46.087101 | 98.337412 | 3318 | 18 |
| Indian | 45.209264 | 84.383713 | 1209 | 17 |
| Italian | 64.285111 | 129.511216 | 3788 | 21 |
| Japanese | 97.183570 | 176.343833 | 1972 | 40 |
| Mexican | 69.134850 | 133.862843 | 3604 | 27 |
| Thai | 82.531850 | 184.999957 | 989 | 28 |

***Kruskal-Wallis*** is used again to check the rank of “review counts” across different types of food and check whether the type of food impacts the number of reviews or not.

The p-value of Kruskal-Wallis test is very small ( $p < 0.05$ ), so the null hypothesis can be rejected and we could say the ***types of food can impact the number of reviews in general***.



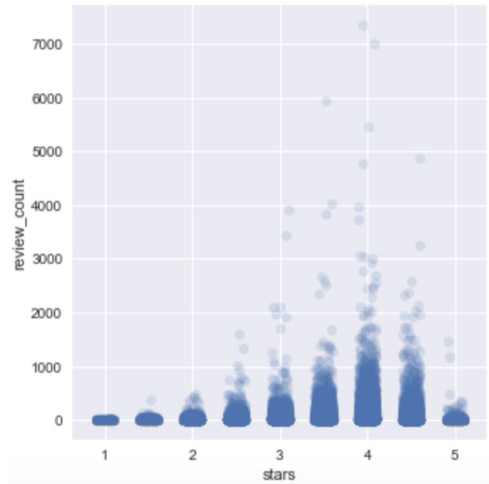

In general, Japanese and American restaurants usually get more reviews. On the other hand, Chinese and Indian restaurants get fewer reviews.

Data Science Capstone Project
Springboard Bootcamp, San Francisco
Reza Taeb
Spring 2018
Title : “Improving Restaurant Reputation Using Yelp User Reviews”
Part 3:  Inferential Statistic

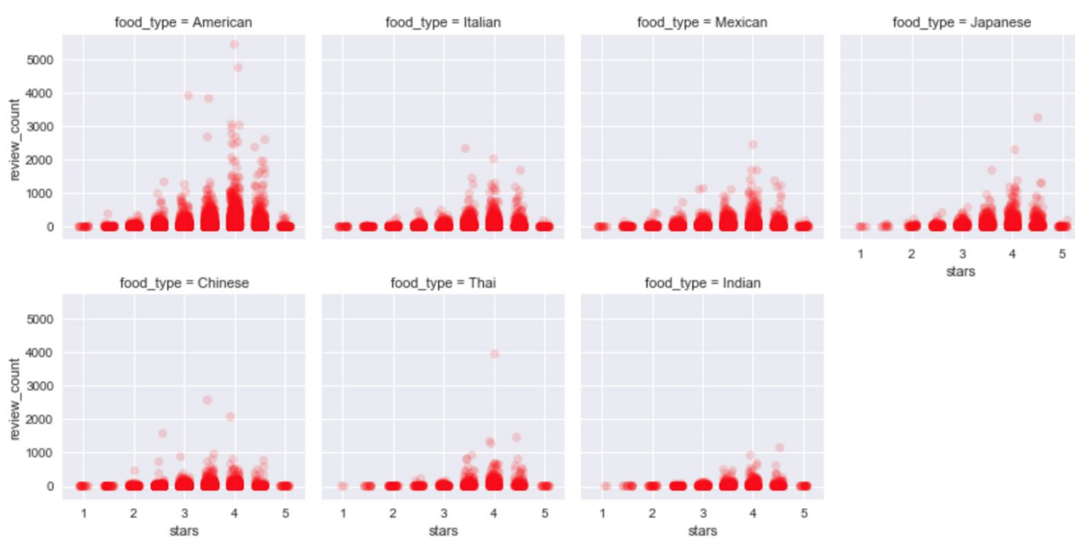
**4 - "Is there a relationship between number of reviews (review_count) and ratings? "**

The distribution of “review_count” across stars is seen in front. As this chart suggests, restaurants with ratings between 3.5 - 4.5 usually get more review compared to the very high and low ratings’ restaurants.

| | stars | mean | median |
|---|---|---|---|
| 0 | 1.0 | 5.620896 | 4 |
| 1 | 1.5 | 12.256198 | 7 |
| 2 | 2.0 | 17.632184 | 9 |
| 3 | 2.5 | 28.971750 | 12 |
| 4 | 3.0 | 41.974925 | 18 |
| 5 | 3.5 | 66.016231 | 27 |
| 6 | 4.0 | 95.941821 | 31 |
| 7 | 4.5 | 80.192841 | 18 |
| 8 | 5.0 | 16.641093 | 5 |




Below charts, show the distribution of “review_count” across stars in different types of food. The below charts also suggest that some types of food like American impact the distribution of “review_count” across “stars” more than other food types such as Chinese.




First, the pearson coefficient was calculated and the P-value is very small ( $p < 0.05$), so we can reject the null hypothesis and we can say ***there is a linear correlation*****. However the correlation is** ***very weak*** *(r = 0.13)*. let's see whether there is any non linear relationship or

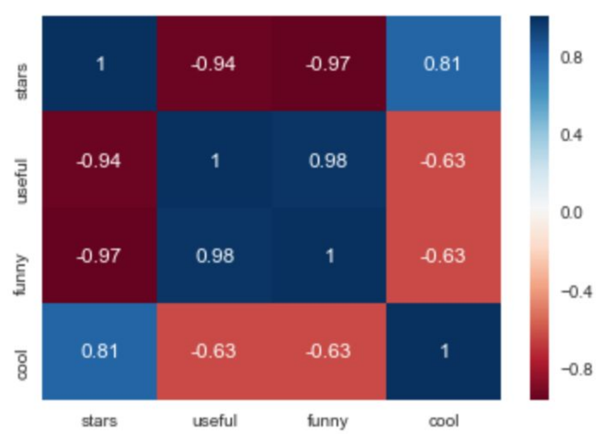
not. ( *"Spearman rank[2]"* correlation).
According to the result, the p-value is very small so we can reject the null hypothesis and we can say ***we can sort the different ratings (stars) based on their number of reviews, although this ranking is not very strong - we can assume it medium. (0.167).***
Based on the chart and the results, we can say 4 , 3.5 , and then 4.5 and 3 stars respectively have more number of reviews across ratings.

**5 - "Is there a correlation between review text features ( "cool" , "useful" , "funny" ) and "rating" ?**

The front chart shows the average of some features of review texts ( “useful” , “funny” , “cool” ) across “ratings”. The chart initially suggests that “useful” and “funny” features have negative direction with “ratings” and “cool” has relatively positive direction with “rating”.

| stars | useful | funny | cool |
|---|---|---|---|
| 1 | 1.364021 | 0.636963 | 0.257079 |
| 2 | 1.311154 | 0.588614 | 0.382864 |
| 3 | 1.148570 | 0.523031 | 0.583691 |
| 4 | 1.170537 | 0.481549 | 0.740233 |
| 5 | 0.933763 | 0.327923 | 0.559142 |

According to the above chart, heatmap, and p-values that calculated for each feature, ***" useful " and " funny "*** have a strong negative correlation with ratings and this correlation is *statistically significant* ( p-values are very small). In other words, **lower ratings lead to higher perceived useful and higher perceived funny by other users.**




The other feature, *"cool",* has a p-value more than 0.05 in two tests (pearson and spearman) , so it's not statistically significant and we can not tell anything about its relationship with stars.

[2] *Spearman rank* test whether of different groups have obvious different rankings or not. (The Spearman correlation between two variables is equal to the Pearson correlation between the rank values of those two variables; while Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships (whether linear or not) )