



Yelp Text Analysis:

“How to improve restaurant reputation even just one star using Yelp user reviews”

Reza Taeb
Springboard Bootcamp
Data Science Capstone Project
Spring-Summer 2018



Problem Statement

- Restaurants owners should be aware of their business current state and they need to be updated frequently in order to survive in the market especially in the big cities with more powerful competitors.
- Business reputation is one of the main area that owners should always take care of that and improve it.
- In the online world, restaurants reputations can be reflected in the customers' reviews on online platforms such as Yelp.



Audience

- current restaurants' owners
- Future restaurants' owners
- Other businesses' owners



Dataset

- The dataset is provided by Yelp for its round 10th challenge. The whole dataset includes several datasets (business, checkins, photos, review, tip, user).
- The datasets had been downloaded from Yelp server on January 9, 2018 and had been kept locally during this project.



Data Wrangling

- Filtering “Restaurants” from original dataset
- Filtering “Non_English” reviews
- Adding “Country” column
- Filtering 7 most popular types of Food (Adding “food_type” column)
- Adding “word_count” and “word_list” column
- Punctuation removal
- Stop words removal
- Lemmatization



Findings

- In general, restaurants “rating” has left-skewed distribution with the 3,5-4 peak.
- In general, restaurants “review counts” across “rating” has left-skewed distribution with 4-4.5 peak.
- People in UK give better ratings to restaurants than people in USA and Canada, Although American and Canadian people, respectively write more reviews for restaurants on Yelp.
- People tend to write more reviews in restaurants that serve Japanese, American , and Thai foods.
- By passing time (from 2006 to 2017), users get more attentive when it comes to give ratings to restaurants and restaurants get more diverse ratings (standard deviation increased 32 percent in 10 years) while the average of ratings for all restaurants stayed around the same number (around 3,65).



Findings

- Type of food can absolutely impact the restaurants' ratings (Japanese, Indian and Thai foods usually get higher ratings).
- Type of food can impact the users' tendency to write more reviews (Japanese food restaurants receives the highest number of reviews compared to other types of food).
- Reviews with lower ratings were found more funny and useful by others compared to the higher rating reviews.
- In general, lower rating reviews have longer length (more words in review's text).
- Some common N-Grams (N=5) statements in whole reviews are : “Can’t wait to go back”. “If you are looking for”, “The quality of the food” and “The best I’ve ever had”,
- By looking to the Ngrams in each rating group, we found out that negative direct recommendation (example: would never recommend) is very common in 1 star even compared to 2 star ratings.



Findings

- “Contrast” is very common in 3 star reviews. Users usually bring both positive and negative comments in their reviews along with words such as “But” , “On the other hand” and etc.
- The obvious difference between 4 star and 5 star reviews is observing more “excitement” and “exaggerated” statements in 5 star reviews compared to 4 star reviews.
- Applying TF-IDF technique, we observed that words and expressions that represent feelings even in their styles (wowowow, awwwwwwful, amaaaazing, looooove) have high impact on reviews’ message.
- Applying Multinomial Naive Bayes algorithm, just using the texts of reviews the “ratings” can be predicted in so many cases (72 percent precision and recall for 1 star difference and 86 percent precision and recall ratios for 2 star difference -between 3 and 5 stars- classification).



Recommendations

- First impression is very important for shaping users' opinions about a restaurants, therefore it is suggested to pay special attention to the first time customers.
- Excitement shows high positive correlation with perfect reviews, restaurants owners have to come up with some ideas to boost customers excitement in their visits.
- Still the quality of food is the first factor that users care about. Do not sacrifice it for anything else.
- Try to induce some suggestions in the customers' minds. "Suggestions" (by other users and staff) can boost restaurant's reputation. This concept may have connection with "signature product/service".



Suggestions For Future Research

- Improving the mentioned classification model to get better precision and recall ratios
- Customized classification model for each country
- Customized classification model for each type of food