Analysis on yelp reviews

Huitong Kou, Zihang Wang, Peibin Rui

1.Goal

To provide

- Data-driven actionable plan to help business owners improve their ratings on Yelp.

- Key useful, analytical insights from the data (e.g. plots, tables, models, inferential methods, etc.) .

- A demonstration of your Shiny application.

2.Data Cleaning

Limited by the computing power, we chose to focus on restaurants whose category contained the word "steak".

The `yelp.ipynb` in `code` folder is used to clean the data and create the word frequency table and the large word embedding table with original views.

After filtering and combination, the `steak_cleaned.csv` has 33629 rows × 6085 columns.

3.Insights about Steak Restaurants

3.1 Different Types of Steak

Firstly, let's check the common types of steaks and how their ratings distribute. These steaks are also the ones that appear many times in the comments.
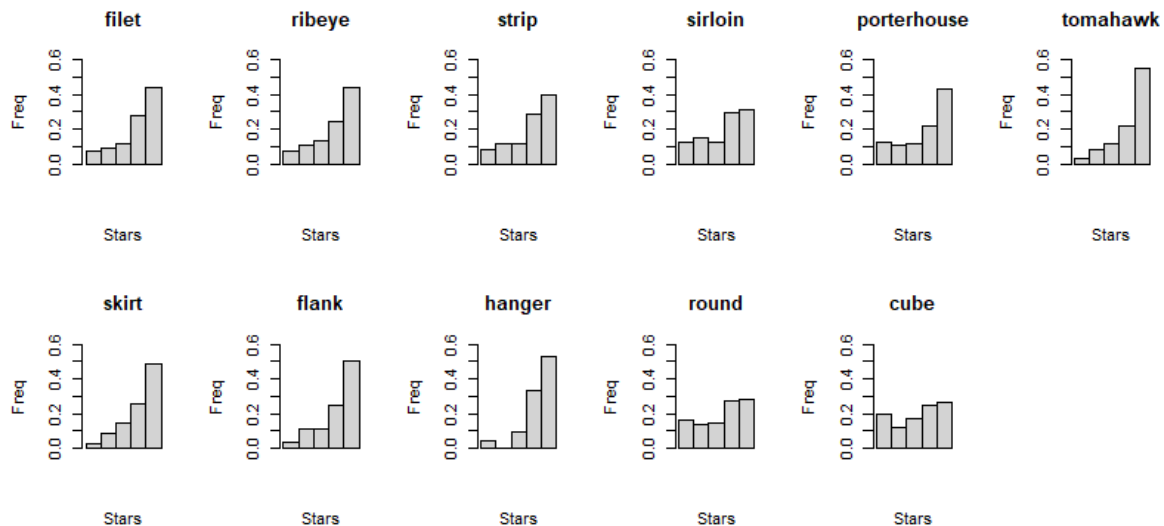
We found that tomahawk, skirt, hanger and flank steaks tends to obtain 5-star comments more easily. The portion of their 5-star comments is significantly over 45%.

However, customers seem to be more picky on sirloin, round and cube steak according to the data. The portion of their 5-star comments is roughly less than 30%.

3.2 Different Types of other food

todo

3.3 Different Non-food Items

todo list

We would like to investigate non-food items which will influence the review stars.

e.g.

food, drink, wifi, parking, price. . .

words distribution in reviews of different stars

focus on certain district like Madison

compare great and awful restaurant from words of food type, etc

statistical test?

. . .

3.4 Statistical Analysis

It's noticeable that the `business_city.json` contains many attributes of each restaurant. We decided to conduct t-test on some of the attributes to see whether different levels of some attributes can make a statistically significant difference on a restaurant's stars.

Firstly, let's check the overall star distributions of all steak restaurants.

Steak Restaurant Stars Histgram
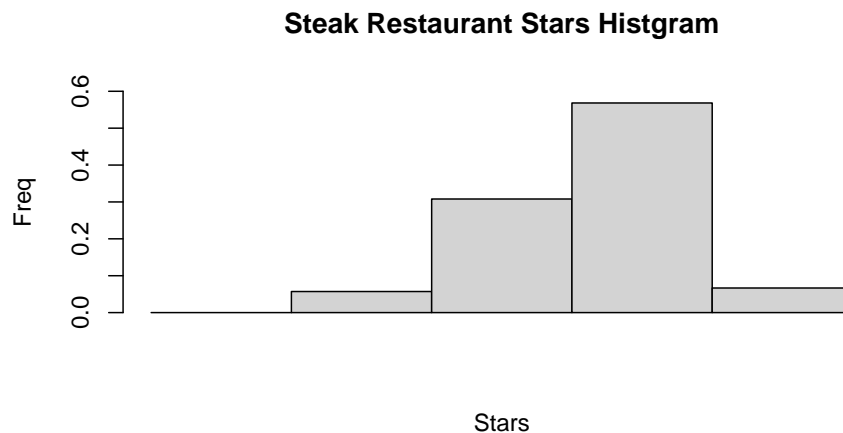
There are 7 attributes selected and the p-value of each t-test is as below.

attribute	p.value
RestaurantsReservations	0.040536672387
NoiseLevel	0.115368255456
RestaurantsAttire	0.000007760033
WiFi	0.799610824593
OutdoorSeating	0.044275411757
RestaurantsDelivery	0.012348344280
RestaurantsGoodForGroups	0.454369986827

With significance level of 0.05, the `RestaurantsReservations(True)` , `RestaurantsAttire(Dressy)`, `OutdoorSeating(True)` and `RestaurantsDelivery(False)` can statistically leads to significant higher stars. The `NoiseLevel(Quiet)`, `WiFi(Free)`, `RestaurantsGoodForGroups(True)` didn't statistically matter.

4.Actionable plan

4.1 Advice on opening a new steak restaurant

In the aspect of steaks:

- Focus on tomahawk, skirt, hanger and flank steaks at first and emphasize them on your menu.

- Filet, ribeye, strip and porterhouse steaks are not bad to consider.

- Make sirloin, round and cube steaks inconspicuous on your menu or avoid offering them.

In the aspect of other foods:

In the aspect of non-food items:

- Set up a proper reservation system is important.
- Offer well-designed attire to your waiters/waitresses.
- Look for possible locations with outdoor seating.
- Don't provide food delivery service.
- Avoid investing in WiFi or soundproof materials at the beginning.
- Allocating too much spaces for groups specially is unnecessary.

4.2 Advice on improving an existed steak restaurant

In the aspect of steaks:

- Try to improve tomahawk, skirt, hanger and flank steaks if one of them has brought your restaurant many low-star comments since the customers are not so picky about them.

- Advertise your sirloin, round and cube steaks if one of them has brought your restaurant high-star comments. It is very praiseworthy to have highly rated sirloin, round and cube steaks.

In the aspect of other foods:

In the aspect of non-food items:

- Improve or set up your reservation system.

- Pay attention to the attire of your waiters/waitresses.
- Check if it's possible to offer outdoor seating.
- Don't provide food delivery service.
- Allocating too much spaces for groups specially is unnecessary.

5.Shiny App

link