**Yelp business analysis Project**

**Introduction**:

Based on the dataset of Yelp consumer’s reviews, we wish to extract information, by conducting statistical analysis, to provide constructive suggestions for business owners for their business development in the industry of burgers.

**Data Cleaning**:

The dataset is large and unstructured. Our raw data is four JSON files, which contains the information of reviews (txt, stars, date, etc.); business (location, hours, etc.); users (name, review count, etc.); tips written by users (txt).

Since our research is focus on the industry of burgers, we picked the items which contain the keyword “Burgers” in reviews, business, users and tips; then write them into new .csv files, which are neat to work on.

**Solution Strategy**:

We extract information of the burger industry from three aspects--reviews, business and tips. (“users” does not contain useful information for our analysis for business development).

We conduct a narrative analysis on reviews and tips, the literature of which contain the useful information. Customers expressed what they care about when dinning out in sessions of reviews and tips, for example, “Great relaxed atmosphere”, “Good beer selection”.

We conduct an analysis on the business session by study the relationship of the rating of restaurants (restaurant stars) and the restaurants’ operation mode (for example, opening hours, locations, free Wifi service, etc.).

**Tips Analysis:**

Solution approach:

We aiming to extract the keywords in customers’ tips, by removing the punctation & unimportant words (the, are, to, etc.), and sorting the keywords based on frequency.

We will focus on both the frequency of the single word (ex. delicious) in the txt and the binary words (ex. French fries) in the txt.



Findings and Recommendations:

Single and binary words with most frequency are displayed in the charts below. These words represent what the customers mostly care about. We suggest the restaurant owner make improvement based on the words appeared in these two charts.

For example, the three most frequent binary words (great food, great service, happy hour) tell us the restaurant should put their main focus in food and service, and offer happy hour if possible.

Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated

**Review Analysis:**

Solution approach:

Similar to the tips analysis, we will sort the review text by frequency. One different point: the review data include the star rating customers give to the restaurant along with the reviews. So we will sorting the reviews to restaurants based on their rated stars: 5 star, 4 star and lower than 3 star.

Graphical user interface

Description automatically generated

Findings and Recommendations:

The plots below show the words frequency of restaurants with 5 star, 4 star and lower than 3 star.

5 star 4 star 3 star or lower

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generatedChart, line chart

Description automatically generated

The most frequent words appear in the reviews of 5 or 4 star restaurants indicate what restaurants should try to make improvement—For example, restaurant should try to provide “great service”, offer “cheese curds” and good “beer selection”, which are the words with most frequency.

On the other hand, words with most frequency in the review of restaurant with 3 star or lower indicate the characteristics of these group of restaurants. For example, the 3 words of frequency are “go back”, “drive through”, “10 minute”, which shows us customers probably will not go back to these restaurant again, so the restaurant should focus on attract one-time customer, while it would not be a good idea for them to put effort on returned customer.

**Business Analysis:**

**Opening hours:**

An analysis on the opening hours of the burger restaurants is shown below. We know that on average, restaurant open 6.9 days per week and open at 7:30 close at 23:10

A picture containing table

Description automatically generated

**Business categories:**

Burgers are sold in restaurants under different categories (ex. Fast food, bars). We try to sort what types these restaurants are, among higher rated restaurants (4 and above) and lower rated restaurants (3 & below).

As is shown in the plot, in 4 star or higher restaurants, burgers are mostly sold in bars, nightclubs and traditional American restaurants. So we recommend these type of restaurants should try to include burger in their menu.

In 3 star or lower restaurants, burgers are mostly sold in fast food, coffee & Tea and traditional American restaurants. So we recommend these type of restaurants not necessary to include burger in their menu.

To be notice: this is just one perspective to limited understand the restaurant rating and business categories. As we can see, traditional American restaurants appear in top 3 of both higher and lower rated restaurants. (which may because these type of restaurant just large in number)

Chart, line chart

Description automatically generated Chart

Description automatically generated

**Locations:**

By comparing the average rating of restaurants in different location, we find restaurants in OH and PA are average rated as 3.0 star; IL rated as 2.5 star; WI rated as 3.5 star. So we would recommend to open restaurants in WI for restaurants owners to develop their business.

Chart, box and whisker chart

Description automatically generated

**Attributes:**

Approach: Examine difference in average star ratings for places with/without the attribute and do two-sample t-test:

Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated Chart, box and whisker chart

Description automatically generated Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated

1. We notice the restaurants without delivery service are higher rated (1 star higher) than the restaurants with delivery. So we recommend the restaurant not to offer delivery service.
2. We notice the restaurants with lower noise are higher rated (0.5 star higher) than the restaurants with higher noise. We recommend not to take noise control at the restaurants.
3. We notice the restaurants with Wifi are lower rated (0.5 star lower) than the restaurants with Wifi. We recommend not to offer Wifi at the restaurant.
4. We notice the restaurants suitable for kids are lower rated (0.5 star lower) than the restaurants not suitable for kids. We recommend the restaurants to create more adult-only dining atmosphere.
5. We notice the restaurants with takeout service are higher rated (0.8 star higher) than the restaurants without takeout. We recommend the restaurants to offer takeout service.
6. We notice the restaurants with parking lot are higher rated than the restaurants without parking lot. We recommend the restaurants to offer parking lot and to offer relatively more parking spaces.
7. We notice the restaurants with TV are lower rated (0.5 star lower) than the restaurants without TV. We recommend not to install TVs in the restaurants.