

Analysis on which Chain Serves the Best Chicken Sandwich
Based on Statistical and Computational Methods

Introduction/Motivation

The restaurants we are going to analyze are those who feature or serve chicken sandwiches. There are a number of restaurants who serve chicken sandwiches, however, the popularity of their chicken sandwiches varies a lot. Since we have data of customers' reviews, the best idea to analyze if a chicken sandwich is popular or not, as well as the reasoning as to why it is popular, is to analyze the restaurants' customer reviews and ratings.

Our first goal of analysis is to determine the chain of restaurants who serve the best chicken sandwich as well as find out why the chicken sandwiches from this chain are so highly sought after. Our second goal is to find out useful information from customers' reviews for each chain of restaurant who serves chicken sandwiches, to see what is the advantage and weakness of their recipe/process of making chicken sandwiches, along with some beneficial and detrimental aspects of their management and business style. These topics of interest lead us to the suggestions and recommendations for each business.

In order to find out the chain of restaurants who serves the best chicken sandwiches, we need quantitative ratings (stars) of each chain. Statistical methods can be used for analysis and inference.

To find out the advantages and weaknesses of each chain's dining experience, we need qualitative ratings (text) for each of the chains. The package NLTK in Python (TF-IDF with the help of sentiment analysis and concordance analysis) can facilitate the extraction of sentiment words as well as key words from text surrounding the topics of interest.

Background Information

The data we are using is downloaded from Canvas STAT628, Module 3. It is a subset of millions of reviews provided by Yelp, which is an Internet platform for users to write reviews of businesses. The data set consists of five .json files (a.k.a JavaScript Object Notation), and each file consists of high-dimension data. The data focuses on reviews of restaurants which are located in different locations around the United States.

The first task we completed was to come up with a list of chains who feature or serve chicken sandwiches. Our resulting list was as follows: Popeyes, KFC, Wendy's, Chick Fil A, Carl's Jr, McDonald's, and Burger King

We then opened business.json with Python to find out the corresponding business_id's for these chains. In review.json, we only kept the ratings (stars) and reviews (text) for these chains, as well as the business_id to keep track of the corresponding restaurant and chain.

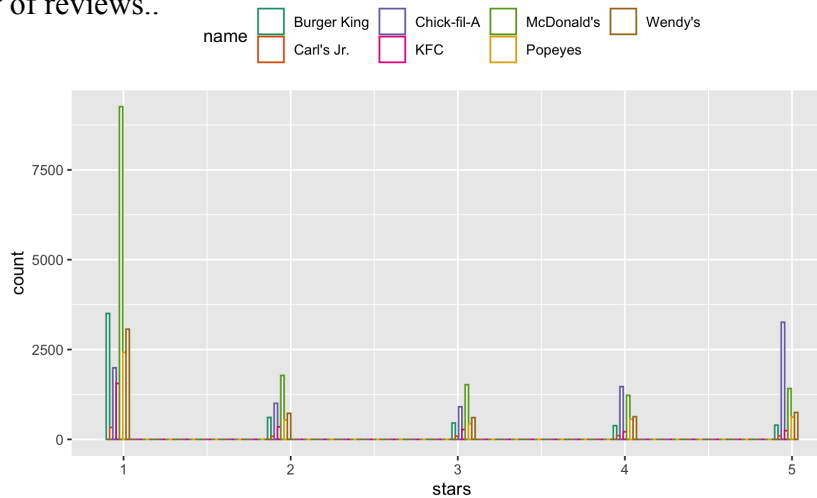
To search for the chain who serves the best chicken sandwiches, (1) we first test if the average rating of chicken sandwiches for each chain is the same as the average overall rating for each chain. In review.json, we keep the data whose review (text) contains "chicken sandwich" to get ratings of chicken sandwiches for each chain. (2) According to user_id, we then merge user.json and the data we filter from review.json to get users' ratings (in stars) for chicken sandwiches and average ratings (average stars). This can help us get the standardized ratings for chicken sandwiches for each chain.

To find out useful information from customers' reviews for each chain of restaurant who serves chicken sandwiches, so that we can see what are the advantages and weaknesses of their recipe of chicken sandwiches, and dining experience, we focus on review.json which is already filtered by business_id (it should be among the list of chains) and text which contains "chicken sandwich".

Exploratory Data Analysis

To explore which restaurant chain makes better chicken sandwich restaurants, we examine the overall distribution of ratings for all the 7 contenders. From the overall ratings distributions, we find that Chick-Fil-A leads in 5 star and 4 star ratings, providing us with

confidence to hypothesize that its chicken sandwich might also be well rated. On the other hand, McDonald's has the largest proportion of 1 star reviews, even though it leads in total number of reviews..



To further examine the actual performance of chicken sandwiches for each contender, we pull all the reviews containing chicken sandwiches as a sample for estimating chicken sandwich ratings. We then compare the estimated chicken sandwich reviews average rating with the overall average rating of the chains.

Restaurant Brand	AVG CS Stars	AVG Restaurant Stars	AVG # Reviews (Reviews about CS)
Chick-fil-A	3.64(+0.14)	3.50	126.60
Popeyes	2.48(+0.33)	2.15	66.22
KFC	2.29(+0.41)	1.88	22.24
Carl's Jr.	2.27(+0.06)	2.21	31.35
Wendy's	2.20(-0.09)	2.29	29.43
McDonald's	1.81(-0.24)	2.05	49.23
Burger King	1.62(-0.20)	1.82	26.48

As shown in the table above, we find Chick-fil-A leading in both average chicken sandwich rating and average restaurant overall rating. Meanwhile, Burger King receives lowest scores in both of these ratings. In terms of comparison between the average chicken sandwich rating and the average restaurant overall rating within the chains. The KFC leads with a 0.41 improvement for its chicken sandwich ratings compared to its overall average restaurant ratings, followed by Popeyes as a close second. In the next sections, we will perform testing, for each chain, on the difference between average restaurant overall rating and its average chicken sandwich rating.

Part 1: Key Findings About Chicken Sandwich chains

1. Hypothesis testing

We test whether chicken sandwich ratings (stars) and overall restaurant ratings (stars) are significantly different for each chain. We use the Welch Two Sample t-test. The p-value is $1.08e-09$. At the significant level $=0.05$, we reject the null hypothesis that chicken sandwich ratings (stars) and overall restaurant ratings (stars) are the same.

We also perform t tests for each chain. For Popeyes, the p-value is 0.0006636. For Burger King, the p-value is 0.02181. For Wendy's, the p-value is 0.7641. For McDonald's, the p-value is 0.2724. For Chick-Fil-A, the p-value is $1.29e-09$. For KFC, the p-value is 0.09424. For Carl's Jr., the p-value is 0.7377. We can conclude that for Popeyes, Burger King, and

Chick-Fil-A, chicken sandwich ratings (stars) are significantly different from overall restaurant ratings (stars), while we conclude the opposite for the rest of the restaurants. Thus, analyzing the difference between average overall restaurant rating and average chicken sandwich rating might only work for some restaurants. For the purpose of ranking the best chicken sandwiches, we should consider the restaurant's performance in both overall average rating and chicken sandwich rating. In this paper we believe that Chick-Fil-A should be crowned champion for best scores in both ratings! In the next sections, we provide analysis on the r

Part 2: Recommendations for Business Methods

Sentiment analysis is used to let computers simulate human reading, i.e. text reviews and respond with whether there was a positive or negative impression of the reviews. It divides reviews into three categories (Positive, Negative, and Neutral). We then extract key information from each set to find keywords in positive and negative reviews. We use the packages NLTK, pandas, matplotlib.pyplot, and re in Python to do sentiment analysis on customers' reviews on chicken sandwiches.

The steps are as follows. First, we clean the data frame in order to keep only chicken sandwich reviews as well as useful columns(name, user_id, text, stars). Second, we remove punctuation from all chicken sandwich reviews (texts), as well as stop words from the text strings(and, or, but, etc.). Then we assign each word in chicken sandwich reviews with tags according to their part of speech. For adjectives, the tag is J, for verbs, the tag is V, for nouns, the tag is N, and for adverbs, the tag is R. After tagging the words in chicken sandwich reviews, we can then use the built in lemmatize function on the remaining words so that they are in similar formats for analysis. For instance, we make words with different tenses the same and to remove inflectional endings. Last but not least, according to the lemmatized words in each chicken sandwich review, and based on the language base in NLTK, we get the subjectivity of each review and calculate the polarity of each review. Each review is assigned with a polarity score which ranges from -1 to +1. The closer the polarity score is to -1, the more negative the review is, vice versa. We set the criteria as follows. If the polarity score is within -1 and -0.025 we label the text as Negative. If the polarity score is within 0.025 and 1 we label the text as Positive. Otherwise the text is labelled as Neutral. We observe in total 1,223 reviews tagged as "Positive", 1,011 reviews tagged as "Negative", and 190 reviews tagged as "Neutral".

2. TF-IDF

We use TF-IDF for information extraction from the reviews. We perform TF-IDF on both positive and negative reviews datasets to get a better understanding on common successes and failures when making these sandwiches. We also perform independent TF-IDF analysis for each chain in order to provide unique suggestions for each chain. In this study, TF-IDF is realized using modules os, json, sys, pandas, matplotlib.pyplot, numpy and nltk in Python.

Our TF-IDF analysis makes use of the reviews separated into "Positive" and "Negative" categories by sentiment analysis from our Exploratory Data Analysis. For both the positive and negative reviews, we then proceed with the following steps. First, we clean the reviews by removing stop words and lemmatizing the text. This leaves us with clean and meaningful words for our next steps. We then vectorize the resulting reviews. Our next step is to call the TF-IDF transformer from the package scikit-learn to calculate the raw term frequency of our terms. We then apply our inverse document frequency script to the raw frequency. This is to

penalize frequencies that are too large and output a standardized list of frequencies with a smaller range. This output is the bag of words we get from one TF-IDF process.

3. Concordance analysis

Concordance Analysis is to find out what appears frequently around “chicken sandwich”. It also gives us insights on what people comment most often about the chicken sandwiches from each chain. Similarly to our sentiment analysis, we start with chicken sandwich reviews and separate them into positive reviews (4 and 5 star) and negative reviews (1 and 2 star) and then clean our data by removing stopwords and lemmatizing the remaining words. However, now we cut the strings using a concordance function to cut everything except for approximately 5 words on either side of “chicken sandwich”. We then calculate word frequency on the remaining strings. After this we removed words without much value (nouns or useless non-stopwords) until we had concordance frequency matrices that gave us a lot of insight about positives and negatives of chicken sandwiches in general.

Restaurant Recommendations

From our analyses above, we have the following suggestions for all chicken sandwich restaurants. With strong evidence to show the following, we believe that improvement in these 3 areas can improve customer satisfaction in any chain. First, one of the most important things would be to speed up the process of the food from the grill to the customer. ‘Slow’ and ‘wait’ are two of the most frequently used adjectives in 1 and 2 star reviews so it is critical to not have customers thinking that when they visit your establishment. This can be fulfilled by accurately forecasting orders and being properly staffed. This leads right into the second recommendation, that the food is always hot and fresh. Two more common words in negative reviews are cold and soggy and are very infrequent in higher reviews. ‘Just in Time’ sandwich preparation is ideal to satisfy this issue to minimize wait time while not sacrificing freshness. According to our analysis, reviews that were written positively often referred to free and new items. Therefore, we believe offering coupons, free treats, and interchanging new menu items greatly increases a customer's experience and will improve ratings.

According to our TF-IDF calculations supported by sentiment analysis results as well as concordance analysis results, we made some individualized suggestions for each chain on our Shiny Web Application. We use word frequency of each chain to determine what negative can be improved upon.

Conclusion

In this study, we conduct statistical testing on chicken sandwich ratings and restaurant overall ratings to help us determine which restaurant chain prevails in the “chicken sandwich war”. We find Chick-Fil-A the champion. In addition, we combined sentiment analysis, concordance analysis, with TF-IDF to produce insights, both generalized and personalized on how businesses should improve.

Contributions:

Luke VandenHeuvel:

1. Responsible for Sentiment and Concordance Analysis
2. Helped with Shiny.app
3. Contributed to Writing Final report

Xintong Shi:

1. Responsible for Shiny.app(worked with Luke), TF-IDF, and hypothesis testing.
2. Writing the report, parts: EDA. Editing: Sentiment Analysis, TF-IDF

Xinyue Zhu:

1. Writing the report: Introduction, Background Information, Methods.