Run a Successful Chinese Restaurant in Philadelphia

Introduction/Motivation

In this report, we will conduct an in-depth study of Chinese restaurants in Philadelphia. Our primary motivation for this study is to provide a analysis of these restaurants using available data. Our goal is to provide a report for each restaurant: including the restaurant's current operations and some actionable recommendations.

We asked specific questions for our user: What is the current status of their businesses in Philadelphia,, including ratings, locations, attributes, and prices? How do their businesses in various neighborhoods of Philadelphia compare concerning price, service, and customer ratings? By understanding these issues, we hope to provide valuable advice to help every Chinese restaurant owner succeed.

Data Pre-Processing

The dataset utilized for this analysis comprises several sources, including two JSON files obtained from Yelp (business.json and reviews.json), a CSV file sourced from the U.S. Department of Transportation's Bureau of Transportation Statistics (Trip_by_Distance.csv), and two additional CSV files extracted from the U.S. Census Bureau (Asian.csv and Income.csv)[1].

We initially filtered the business.json file to extract information about restaurants specifically located in Philadelphia. From the filtered data, we further focus exclusively on Chinese restaurants based on their categorization. To gather customer feedback and reviews for these Chinese restaurants, we used the business IDs from the filtered business.json to locate and extract relevant reviews from the reviews.json dataset. As the Trip_by_Distance.csv dataset spanned from January 1, 2019, and the review records extended until January 19, 2022, we narrowed down our analysis to the period between January 1, 2019, and December 31, 2021, to maintain temporal consistency.

To analyze reviewer sentiment, we utilized the TextBlob package in Python to compute sentiment scores for reviews. Additionally, we employed the Census Bureau's data API to extract racial and income information based on zip code and time period, thereby offering insights into the demographics of the neighborhood.

We characterize a successful business as having a star rating greater than 4.0 or receiving reviews placing it in the top 25% among all Chinese restaurants from 2019 to 2021.

Exploratory Data Analysis (EDA)

The table below provides an overview of various types of Chinese restaurants in Philadelphia, including their frequency, average ratings, and the average number of reviews:

Category	Count	Star Mean	Review Mean
Asian Fusion	72	3.6	54.1
Dim Sum	48	3.6	57.2
Seafood	43	3.5	67.9
Szechuan	31	3.7	56.3
Noodles	30	3.8	66.4
Fast Food	16	3.2	71.1
Hot Pot	15	4.0	112.9

Table 1

"Asian Fusion" as the most prevalent category among Chinese restaurants in this city. However, it received the lowest average number of reviews. In contrast, it is quite intriguing to observe that, despite the fewer numbers of hot pot restaurants, they exhibit the highest average ratings along with the highest average number of reviews. This phenomenon can be attributed to the widespread fondness of Chinese diners for hot pot cuisine. On the other hand, "Fast Food" receive the lowest average star ratings. We hypothesize that this could be due to the impact of fast food preparation methods on the traditional flavors of Chinese cuisine. Fast food preparation may not align with the nuanced tastes often associated with Chinese dishes, which might explain the lower ratings in this category.

Here is a violin plot illustrating the relationship between restaurant attributes and star:

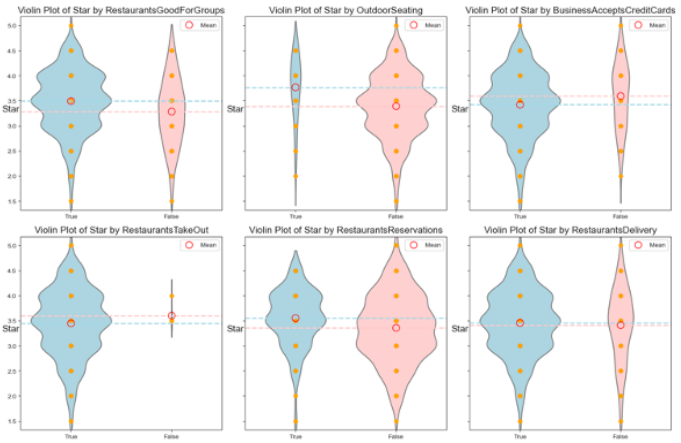

Fig 1

We found that restaurants with attributes such as 'Good For Groups', 'Outdoor Seating', 'Reservations', and 'Delivery' tend to have higher average scores than those lacking these features. The opposite is true for ‘Take Out’ and ‘Accepts Credit Cards’. The width of the violin chart represents the number of restaurants in this range under the specific score. This reveals that the majority of restaurants are group-friendly, typically do not offer outdoor seating, accept credit card payments, and offer both takeout and delivery services. This insight helps in understanding the prevalent trends in the restaurant industry and how specific attributes correlate with customer ratings.

The violin plot below is a visualization of the distribution of star ratings for different price ranges:

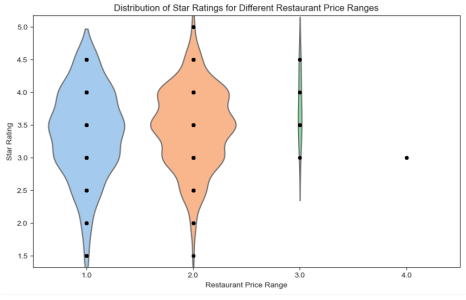

Fig 2

In the figure, the Price range from 1.0 to 4.0 represents the increasing approximate cost of a meal per person at a restaurant.1.0 = under $10, 2.0 = $11-$30, 3.0 = $31-$60, 4.0 = over $61. More than 90% of restaurants have a per capita cost of less than $30.

The chart presents the trend of the average number of monthly outings and the total number of restaurant reviews submitted in Philadelphia from 2019 to 2021. An 'outing' is quantified each time a person leaves their home and stays at a different location for more than ten minutes, with each stay over ten minutes being counted as one occurrence.

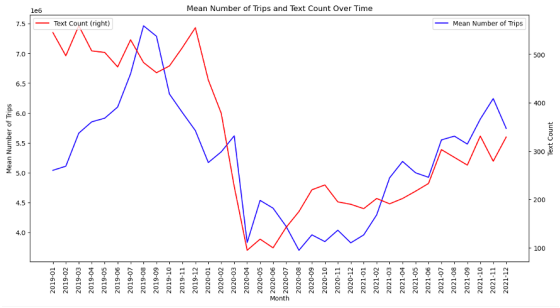

Fig 3

From the chart, there appears to be a correlation between the frequency of Philadelphians' outings and their interaction with restaurants, as seen through the volume of reviews. A significant downturn is evident in both metrics at the end of 2019, aligning with the emergence of the global pandemic, which had a profound impact on public life and the restaurant industry. Following the initial decline, there's a gradual resurgence in the mean number of outings, suggesting a cautious return to pre-pandemic social habits, possibly influenced by the development of vaccines and lifting of restrictions. However, the rebound in restaurant reviews is less pronounced, signaling a more tentative recovery in the dining sector.

In our initial analysis of word count, in figure 4, we observed that seafood and chicken emerged as the most popular food choices. Additionally, soup dumplings and noodles were the most frequently mentioned dishes, and over half of the reviewers indicated visiting Chinese restaurants with their friends.

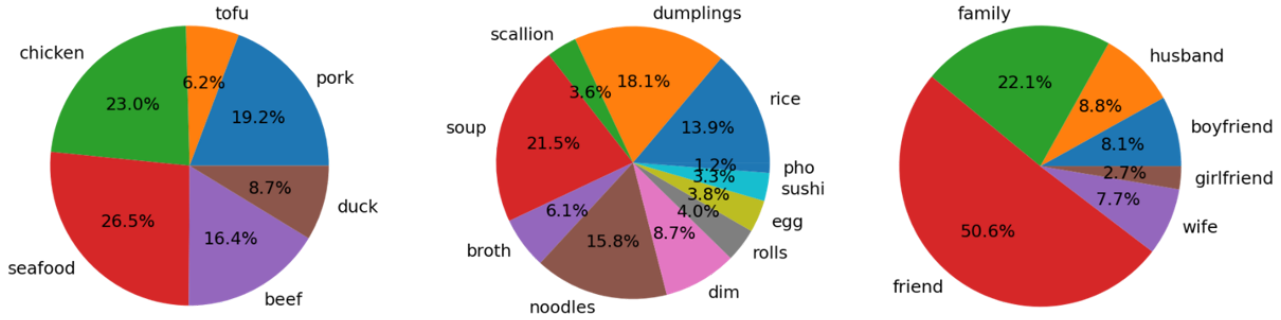

Fig 4

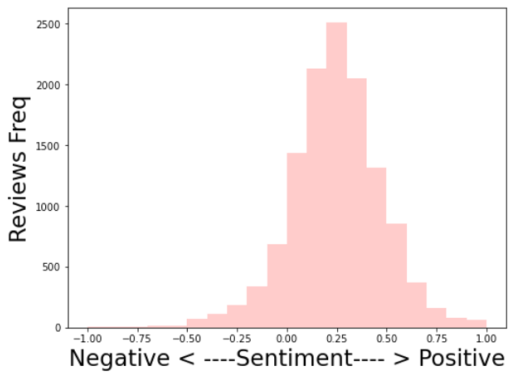

Fig 5

In sentiment analysis, the distribution shifts to the right, indicating that the majority of reviews express a positive sentiment.

The Comparisons and Suggestions

The distribution area of Chinese restaurants in Philadelphia is concentrated in different areas. Considering the influence of regional factors (e.g., race, population density, income, etc.) we therefore felt it was important to categorize the restaurants according to the different regions and analyze the neighbors and competitiveness of the restaurants according to the region in which they are located. We categorized areas in Philadelphia as Center, North, South, East, and West.

Subsequently, we computed the household income and Asian population proportion in different regions. Intuitively, a higher household income and a greater proportion of the Asian population in a neighborhood suggest a more favorable environment for a Chinese restaurant. This information offers an objective overview of the users' region. To findout how customer satisfaction about the restaurant. We determine satisfaction with price by identifying price-related words or phrases in reviews and calculating the mean and standard error of the mentioned-price reviews. The satisfaction with service is assessed using a similar way. By visualizing the distribution of sentiments in overall reviews, mentioned-price sentiments, and mentioned-service sentiments(see figure 6 as example), users can effortlessly identify the strengths and weaknesses. Based on the comparison results, we provide suggestions to our users.

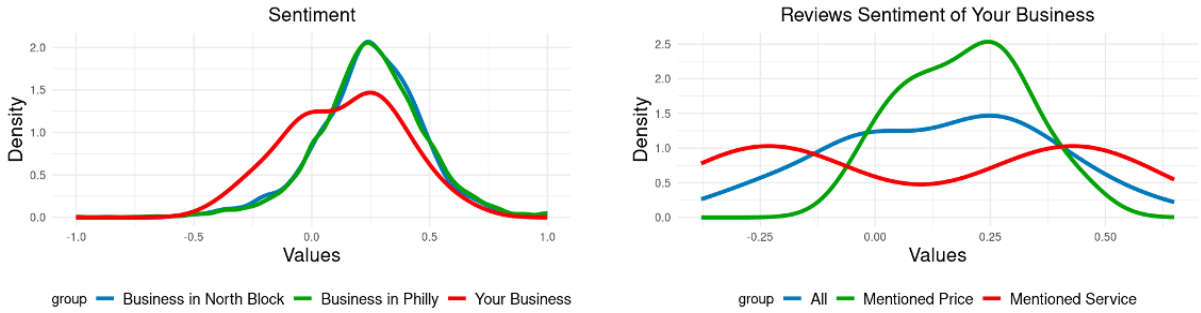

Fig 6

Conclusion:

In summary, our application delivers a detailed report on every Chinese restaurant in Philadelphia, encompassing ratings, locations, attributes, services, and pricing, along with tailored recommendations. It unveils key consumer preferences and prevailing market trends. There are some limitations of our study. While our data is comprehensive, it might not capture every facet of the restaurant landscape. Also, our methodology lacks the creation of predictive models, limiting our ability to forecast future trends and customers' behaviors. Despite these limitations, our analysis equips owners of Chinese restaurants in Philadelphia with insights, enabling them to better understand their current position and make informed improvements to align with customer needs and preferences.

Claim

ChatGPT is used in improving Shiny App performance, code debugging.

References

[1] U.S. Census Bureau. (n.d.). Data.Census.Gov. https://data.census.gov/

Contributions:

Contributions	Yeming Wei	Wanxin Tu	Yuman Wu
Presentation 1	Responsible for slides 6-8.	Responsible for slides 3-5.	Responsible for slides 2 and 9.
Presentation 2	Same contribution	Same contribution	Same contribution
Summary	Reviewed and edited summary.	Reviewed and edited summary.	
Code	Responsible for data cleaning code, and analysis section	Responsible for NLP models and reviewed data cleaning code.	
Shiny App	Co-responsible for Shiny app	Co-responsible for Shiny app	