

1. Introduction

1.1 Background

The COVID19 crisis has forced many restaurants to temporarily close their doors, limit their staff, and/or have reduced hours and seating. These factors have contributed to significant loss of revenue for restaurants, who rely on a steady stream of sales for day to day operations. Some restaurants have had to permanently close due to pandemic conditions. Other restaurants, however, have been able to stay a float and even thrive throughout the pandemic.

Restaurant X is a fine dining restaurant located in downtown Chicago, IL. Management of Restaurant X has noticed that they have been struggling since May 2020. The restaurant has seen decreased revenues, fewer customers, and even worse reviews. Management believes some of this may be due to COVID, but other factors may influence the popularity of Chicago restaurants.

1.2 Problem

As the pandemic slows and restaurants open up, stakeholders at Restaurant X want to make sure they are able to open successfully. Specifically, they want to be in the top 25 recommended Chicago, IL \$\$\$\$-\$\$\$\$ price range restaurants. Restaurant X has retained Claire-Firm to help them make it to this list.

1.3 Interest

Determining what factors contribute to a top 25 recommended restaurant in this post-pandemic world will be helpful for stakeholders because it may allow them to change aspects of their business to align with those factors that make a restaurant successful when coming out of these unprecedented times. The restaurant could then potentially reap benefits from making these changes. This could allow restaurants to avoid shutdown, reduce the number of unemployed service industry workers, and avoid depriving Chicago residents of the staple foods.

2. Data

2.1 Data Sources

In order to answer these questions, we need to understand:

1. Which \$\$\$\$-\$\$\$\$ Chicago, IL restaurants are in the top 25 recommended and which are not.
2. What are the characteristics of the two different sets of restaurants.

FourSquare's Place API data can help understand these.

We can use FourSquare's Search Venues API endpoint to find all \$\$\$-\$\$\$\$ price range restaurant venues in select neighborhoods around Chicago. We have to choose multiple neighborhoods so that we can get a dataset larger than 50 entries (which is what FourSquare limits you too with one query).

Next, we need to figure out which of these restaurants are in the top 25 Chicago, IL \$\$\$-\$\$\$\$ restaurants overall. To do this, we can consult FourSquare's Venue Recommendations API (<https://developer.foursquare.com/docs/api-reference/venues/explore/>), as well, which returns a list of recommended restaurants in order of rank.

Finally, we use FourSquare's Venue Details API endpoint to find characteristics of all of the venues \$\$\$-\$\$\$\$ price range restaurants including but not limited to distance from downtown and other Chicago attractions, hours, price tier, rating, whether the venue takes reservations, parking availability, whether a menu is available online, and restaurant type (i.e., Sushi, Chinese, American, Pizza, etc.), among other attributes.

2.2 Data Cleaning

General

High-priced restaurants (aka \$\$\$-\$\$\$\$ restaurants) within 250 meters of the centers of Chicago neighborhoods were scraped. The specific neighborhood centers include the Loop, West Loop, South Loop, Chinatown, Lake View, River North, Bucktown, Fulton Market, Lincoln Park, Uptown, Boystown, Logan Square, Wrigleyville, Streeterville, Noble Square, Ukrainian Village, Gold Coast, Old Town, Wicker Park, Humboldt Park, and the proper center of Chicago. These searches resulted in 200 unique restaurants.

Next, it was determined which of these 200 restaurants were in the top 25 restaurants among high-priced restaurants, where the top 25 was determined by a FourSquare generated popularity sort in July of 2021.

For those 200 restaurants, the details API was used to scrape attributes of the restaurants. These attributes included:

- The phone number, twitter handle, instagram handle, website url, number of photos, and Facebook id of the restaurant
- The hours of the restaurant
- Number of checkins, number of user currently at the location, number of likes, an average rating and whether the restaurant is verified
- Whether the following are available at or for the restaurant:
 - * Delivery

- * An online menu
- * Specials
- * Payment by credit card
- * Outdoor seating
- * Wi-Fi
- * Breakfast
- * Brunch
- * Lunch
- * Happy hour
- * Dinner
- * Dessert
- * Beer
- * Wine
- * Full Bar
- * Cocktails
- * Bar Service
- * Bar Snacks
- * A Tasting Menu
- * Reservations
- * An ATM
- * Wheelchair access
- * TVs
- * Private rooms
- * Live music
- * Jukeboxes
- * Music
- * Coat Check
- * Street parking
- * Valet parking
- * Parking in general
- * Public parking lots
- * Table service
- * Allowance of BYOB

Converting String Attribute Variables in to Dummy Variables

Some extracted attributes were in a string format of “Yes” or “No”. That is, if the restaurant has the attribute then “Yes” was indicated and, if not, then “No” was indicated. Dummy variables were generated for these attributes where 1 indicates the attribute is available at the restaurant and 0 indicates it was not. These attributes included credit cards, outdoor seating, wi-fi, reservations, ATM, restrooms, wheelchair accessible, TVs, private rooms, music, and coat check. Some restaurants did not have any value entered for these attributes. For these, the dummy value was entered as 0 because, even if these restaurants have these attributes they are not advertised

on FourSquare and so are therefore essentially a “No” for customers searching for features of this restaurants.

Other extracted attributes of restaurants were populated if the attribute was present at the restaurant and blank otherwise. The attributes were converted to dummy variables by coding populated fields as 1 and non-populated fields as 0. These attributes include phone, twitter, instagram, Facebook, url, breakfast, dessert, happy hour, lunch, brunch, dinner, beer, wine, full bar, cocktails, street parking, public lot, live music, jukebox, bar service, bar snacks, tasting menu, delivery, valet parking, BYO, parking table service, and public lot.

Feature Selection

Dummy variables that were available for all restaurants or not available for all restaurants were removed. For instance, all restaurants had phone numbers, so the phone number dummy was not included because it would not provide any useful information. Other dropped variables include checkin_counts, user_counts, visits_count, here_now, hours_days, phone_available, credit cards, and wheelchair accessible.

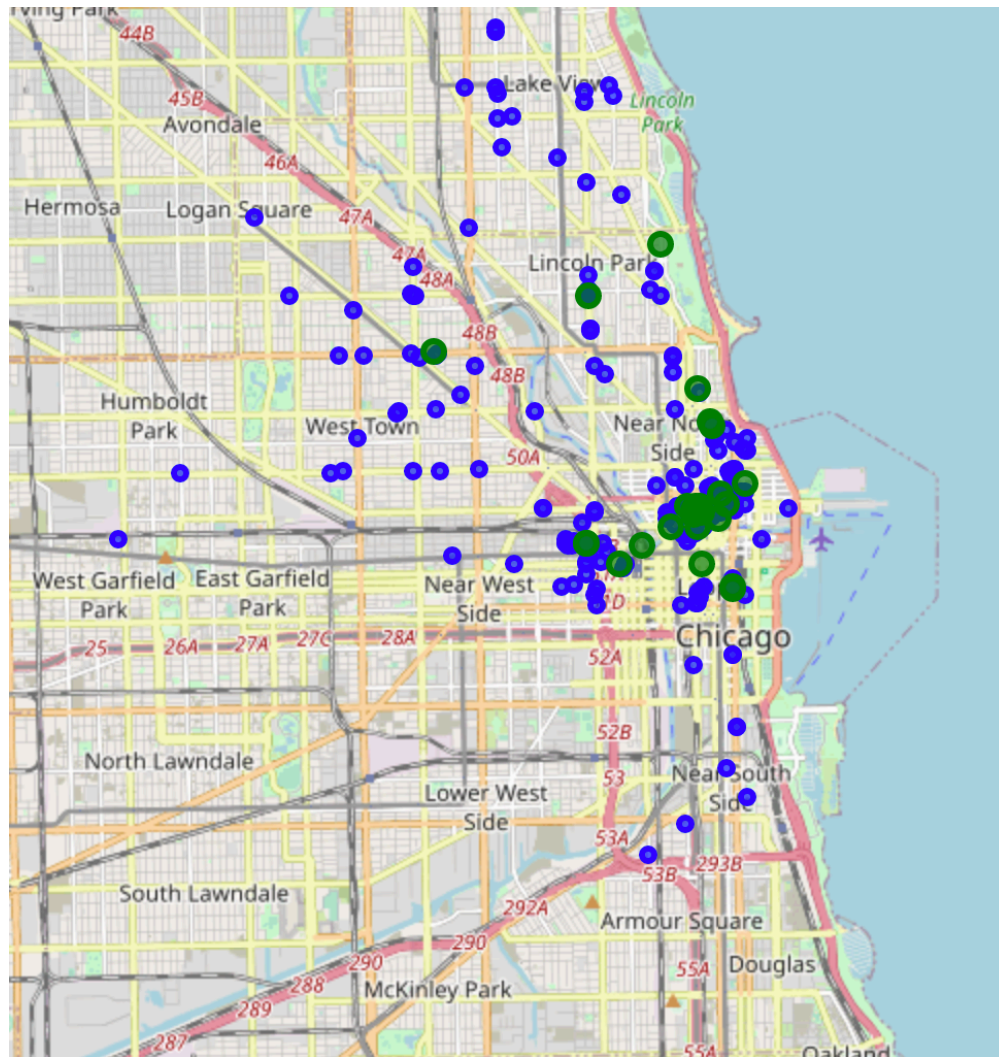
The other variables (as described in the general section, above, and cleaned in the manner outlined in the cleaning section, above) were all determined to be potentially predicative of whether a restaurant would be in the top 25 \$\$\$-\$\$\$\$ restaurants.

3. Methodology

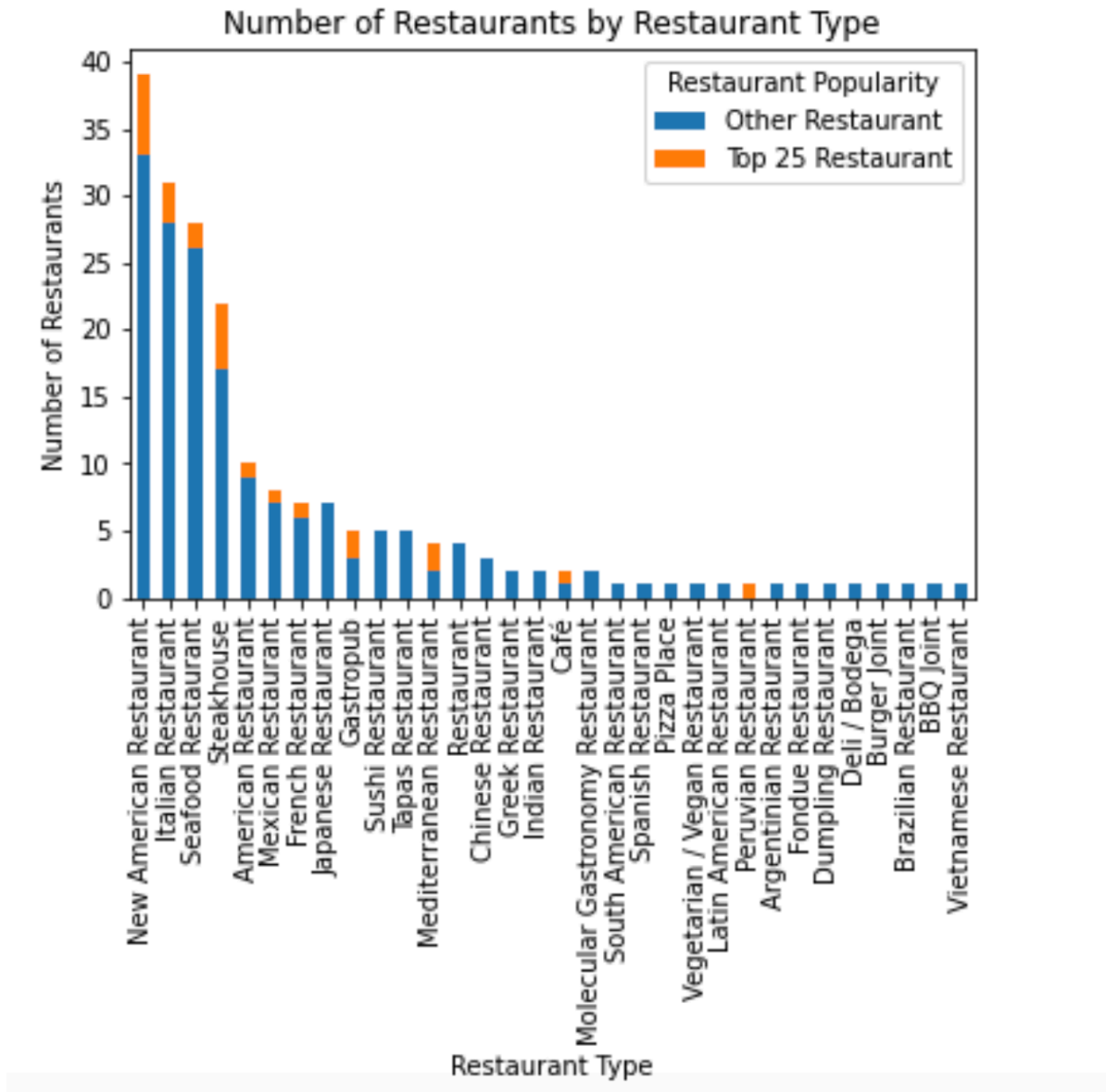
3.1 Exploratory Diagnostics

The following analyses were designed to determine what variables may help explain why some Chicago high-priced restaurants end up in the top 25 most popular in this price category and why others do not. Insights from these diagnostics are used to create a logistic regression model to determine what factors influence admittance to the top 25.

Primarily, one might guess that location is a big factor in what determines if a restaurant is popular or not. In this case, one would expect the top 25 restaurants to be located in Chicago’s downtown area, since those restaurants are central to tourist attractions such as Millennium Park, Michigan Avenue, the lake shore, and Chicago’s Loop area (to name a few). Outer areas such as Lincoln Park, Bucktown, and Ravenswood, while fun to Chicago natives, do not have the same level of “hype” or stream of customers (aka tourists) as Chicago’s downtown area. To test this hypothesis, the top 25 and other restaurants were plotted with Python’s folium library. As can be seen in the map, below, top 25 restaurants (shown in green) are, indeed, heavily concentrated in Chicago’s downtown area. However, other restaurants (shown in blue) are also located in Chicago’s downtown. Therefore, location is certainly not the only factor that may contribute to a restaurant being popular or not.

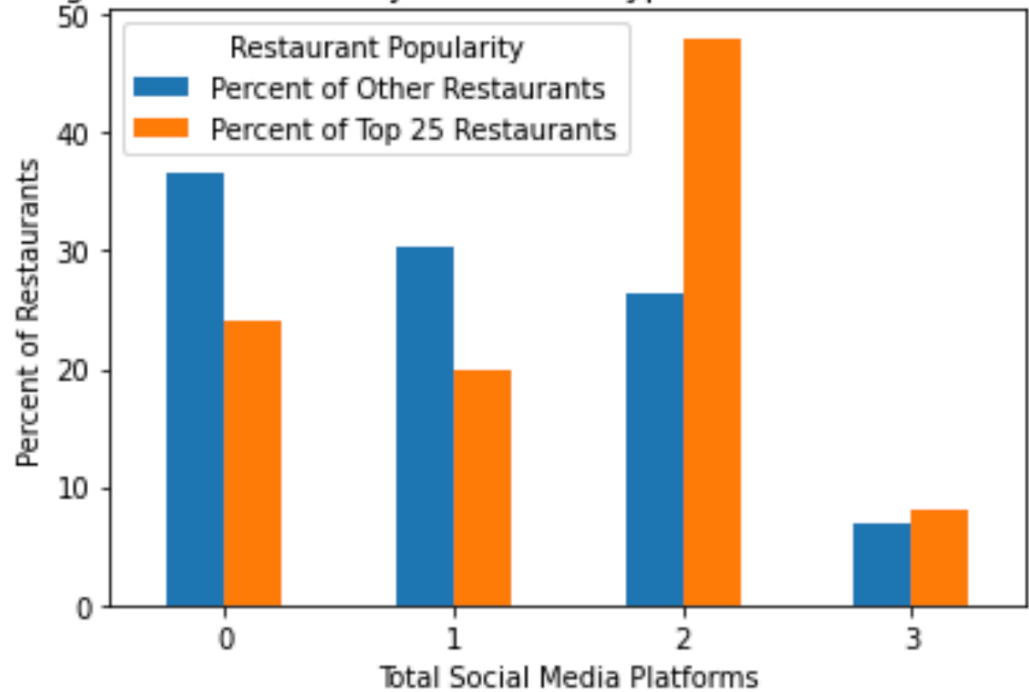


Second, the restaurant type might influence whether a restaurant is in the top 25 or not. New American, Italian, Seafood, and Steakhouse were among the top restaurant categories for both the top 25 and other high priced restaurants in Chicago.

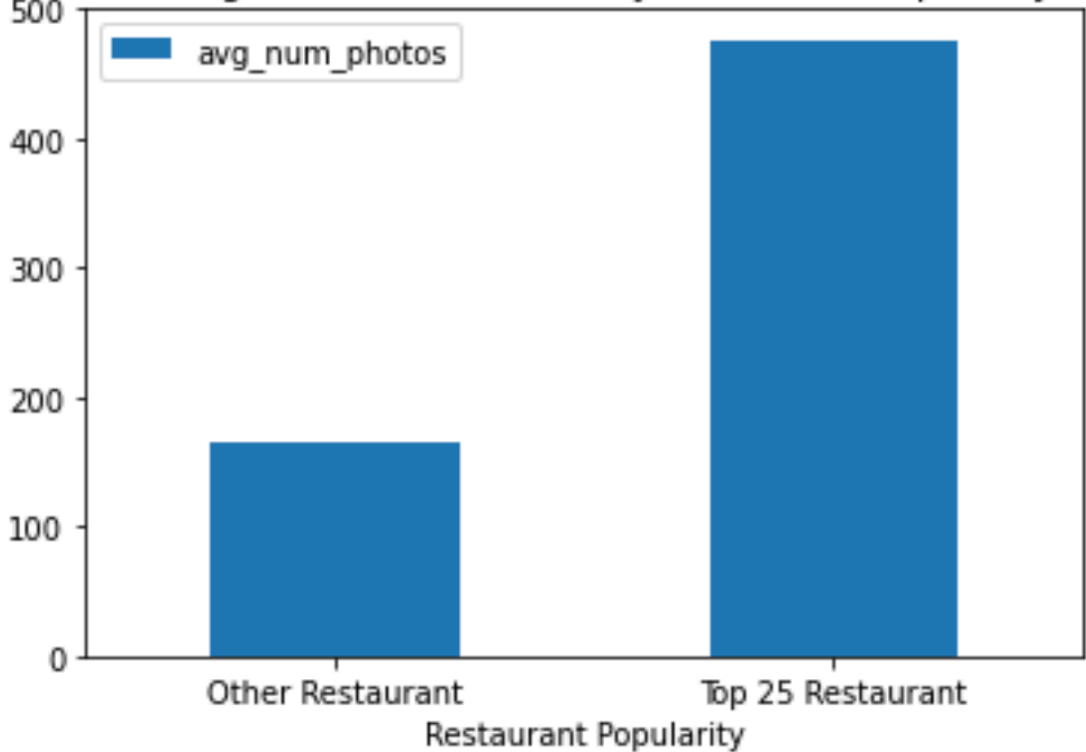


Third, one may assume that the number of social media platforms is directly proportional to the popularity of a restaurant. That is, with a larger social media presence comes a larger following and this leads to popularity. Below is a chart that shows the percentage of top 25 restaurants and percentage of other restaurants with a presence on 0 to 3 social media platforms, where the social media platforms include Facebook, Twitter, and Instagram. As can be observed, top 25 restaurants tend to have a larger social media presence. Additionally, the next chart shows the average number of photos per restaurant by top 25 v. Other restaurants. Top 25 restaurants have almost three times as many photos as other restaurants. This is not surprising as posting images

Percentage of Restaurants by Restaurant Type with 0-3 Social Media Platforms

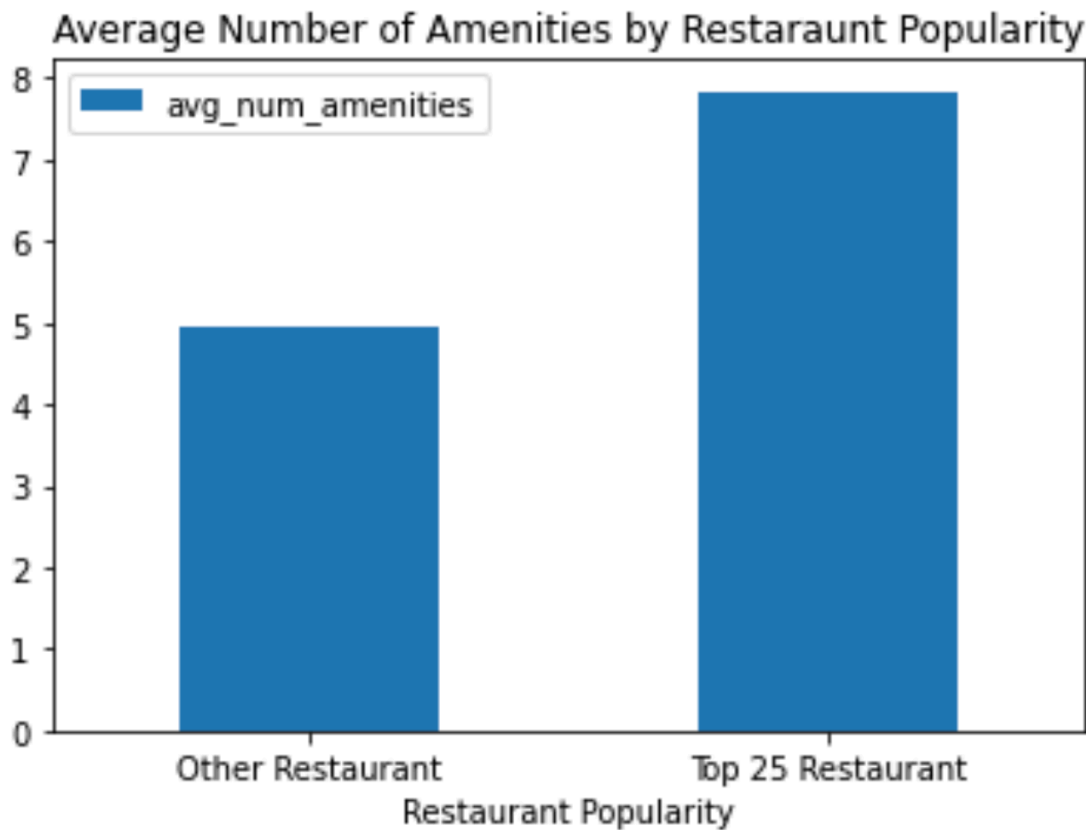


Average Number of Photos by Restaraunt Popularity



of food and atmosphere likely draws customers in. Additionally, as restaurants become more popular, the more patrons post pictures.

Fourth and finally, the total number of amenities may have something to do with how popular a restaurant is. Amenities like music, bar service, brunch, and parking may influence how desirable



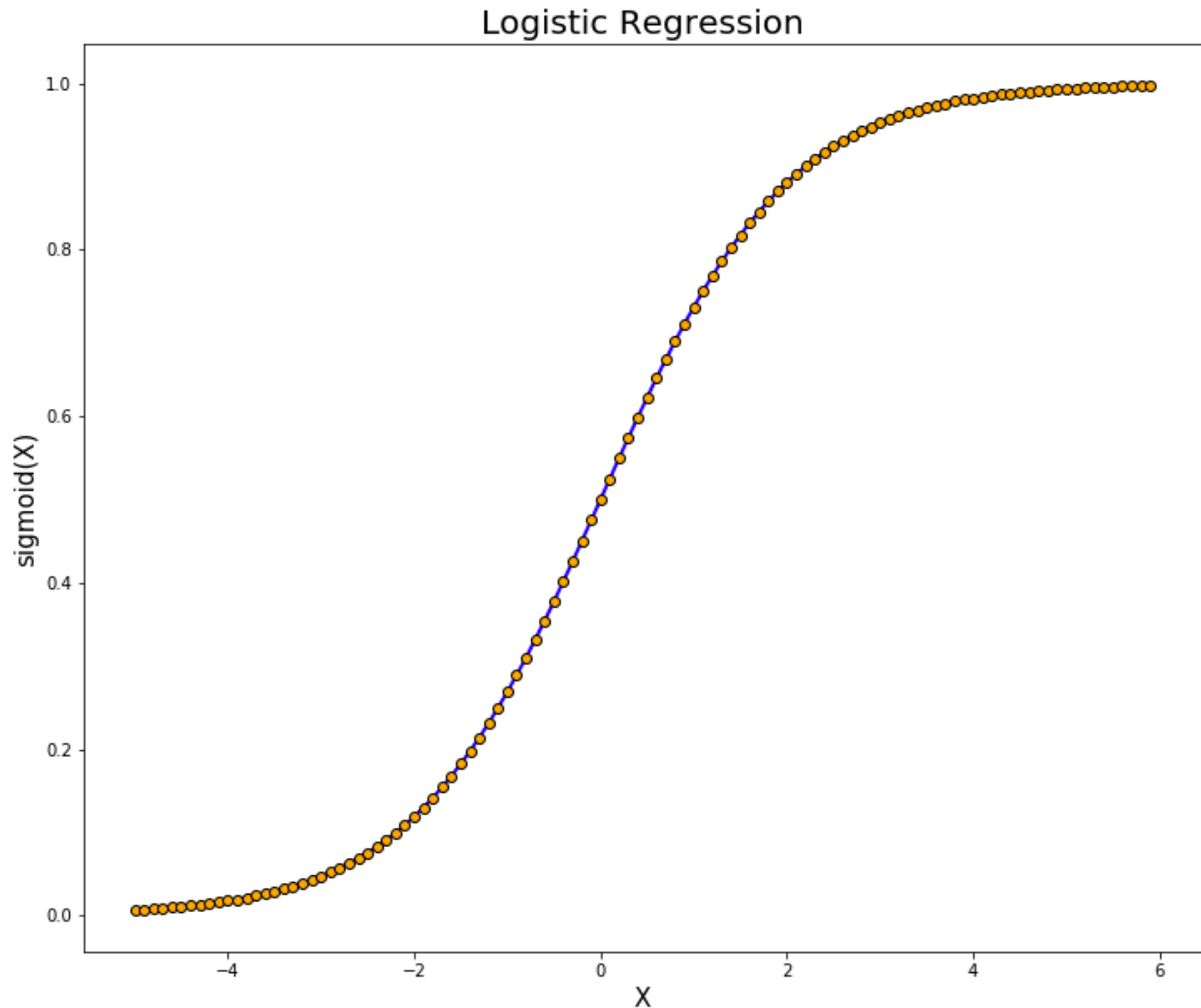
it is to go to a given restaurant. As can be seen in the diagram, below, top 25 restaurants tend to offer more amenities than other restaurants. The logistic regression that follows will determine if any particular amenity leads to a restaurant being more popular or not.

3.2 Analysis

Logistic regression is a machine learning, predictive model analysis that is conducted when the dependent variables is a binary variable. Like all regression analyses, the logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. I apply a logistic regression model analysis to determine what factors influence whether a company.

4. Results

The logistic regression model showed that the number of amenities and the location of the restaurant influenced whether a restaurant was in a top 25 restaurant or not.



5. Discussion / Conclusion

This suggests that Restaurant X may want to include more amenities such as music, parking, brunch, wifi, etc. in their offerings. Additionally, they may want to post more on social media to get their presence out.