

Introduction

For this report, I will be attempting to discover what that best city for opening a new pizza restaurant is in the vicinity of the north/northwestern suburbs of Detroit, Michigan. This area has a very large number of dining and bar options that vary both in quality and amount of competing dining options in that area. A restaurateur that is looking to open a new pizza restaurant will need to decide where to open her new place based on several qualities. This report will help the restaurateur to choose the best location for the new restaurant, giving it the greatest chance at thriving in this competitive environment. She is looking for an area that has other highly rated venues, including both restaurants and non-dining venues, but is not overly saturated with other pizza restaurants.

Data

The main source of data for this report came from the foursquare API. This provided location, name, id number, category, and rating for each venue that was looked at. The variables of highest interest were location and rating. Location was used to find areas that were not overcrowded with other pizza dining options. Rating was used to find which locations had the best rated venues.

Three websites were scraped to obtain preliminary data on these Detroit suburbs. The first was used to get the names of all the north/northwest suburbs of Detroit (umich):

<https://guides.lib.umich.edu/c.php?g=283069&p=1885993>. The second was used to obtain the county the suburb was located in (Wikipedia):

https://en.wikipedia.org/wiki/List_of_cities,_villages,_and_townships_in_Michigan. Latitudinal and longitudinal data was download from a .csv file from <https://www.aggdata.com/free/united-states-zip-codes>. The information from these websites was used to enter information into the foursquare API.

Methodology

Many python packages were needed for this project. These included BeautifulSoup, pandas, numpy, folium, matplotlib and geocoder. The umich website was scraped to obtain the names of the suburbs being looked at. Forty suburbs were taken from this website. The Wikipedia website was like wised scraped to obtain data on the suburbs county. These were both made into separate DataFrames and then combine into one. This DataFrame contained sixty-seven entries instead of forty, since one suburb may contain multiple zip codes in it. The csv file was read and combined with the first DataFrame (dropping unnecessary columns) to create one DataFrame contained all information needed to begin search the foursquare api. The columns of this DataFrame was called df_detroit_burbs and contained the columns suburb_name, zip, county, latitude, and longitude. One zip code was removed from the final DataFrame as it had no latitudinal or longitudinal data available. Removing this zip code did not influence the final recommendation as it was in an area that had many additional zip codes.

Folium was then used for preliminary data visualization of the suburb locations using latitudinal and longitudinal data. From the first visualization, it was apparent that two suburbs were out of the bounds of the area being studied. Romeo (zip code 48065) and Melvindale (zip code 48122) were subsequently removed from the DataFrame. The data was then visualized again for accuracy.

The foursquare API was then accessed using the information from the df_detroit_burbs DataFrame. Two functions were created to get the name and categories of venues located near each suburb. Due to the call restraints of the API, a limit of one hundred was put on the results for each location. The results were grouped by the number of venues return for each zip code. Using one hot

encoding the top ten most common venues were found for each suburb. A new DataFrame was created that removed any suburb that had 'pizza place' in the top ten common venues. A new function was used to return only the ID numbers of venues near each suburb, again with the limit of one hundred. These IDs were used to pull ratings for each venue. The foursquare API account used had a limit of five hundred premium calls a day. Due to this, the data was split in half with one half being ran on day one and the other half on day two. The ID and ratings data were saved to a pickle so that future calls would not need to be made. The ratings were grouped by zip code and the mean rating was found for each zip code. The mean ratings were combined with the DataFrame of the top ten most common venues and sorted by rating, with the highest rated suburb at the top of the list. This gave the highest rated single suburb with no pizza place.

Results

Table of mean venue ratings by zip code.

Zip Code	Suburb	Rating
48203	Highland Park	8.00
48220	Ferndale	7.63
48315	Utica	7.50
48084	Troy	7.42
48037	Southfield	7.33
48397	Warren	7.00
48090	Warren	7.00
48009	Birmingham	7.00
48099	Troy	7.00
48318	Utica	7.00
48311	Sterling Heights	7.00
48331	Farmington	7.00
48034	Southfield	7.00
48314	Sterling Heights	7.00
48071	Madison Heights	6.93
48051	New Baltimore	6.80
48212	Hamtramck	6.75
48007	Troy	6.70
48301	Bloomfield Hills	6.60
48303	Bloomfield Hills	6.42
48086	Southfield	6.42
48321	Auburn Hills	6.42
48068	Royal Oak	6.42
48332	Farmington	6.42
48333	Farmington	6.42
48343	Pontiac	6.42
48336	Farmington	6.33
48237	Oak Park	6.00
48326	Auburn Hills	6.00
48335	Farmington	6.00
48310	Sterling Heights	5.75

Discussion

At first glance, the best individual suburb to open a pizza place in the north/northwest metro Detroit area would be in Highland Park, specifically zip code 48023. This area does not have a pizza place in the top ten most common venues. It also has the highest average rating of all surrounding comparable areas. This area has a high number of what could be considered nicer venues, including art galleries, yoga studios, and other restaurants. Based on the top venues in the area, it would seem that Highland Park would attract a higher-class customer base. However, two suburbs come up three times in the results, Troy and Sterling Heights. The Troy zip codes without a pizza restaurant have an overall average venue rating of 7.04 and the Sterling Heights venues have an overall average rating of 6.58. These ratings are quite a bit lower than the Highland Park but there may be benefits to opening a pizza restaurant in one of these locations. They are both large locations without a large number of pizza places already dominating the area. Troy and Sterling Heights have abutting borders. Opening a pizza place in Troy could draw in pizza wanting customers from Sterling Heights and vice versa. For these reasons, it is recommended that the restaurateur open her new pizza restaurant in the Detroit suburb of Troy.

Conclusion

The Detroit suburbs are overflowing with places to eat. This is doubly true for pizza restaurants. A restaurateur looking to open a new pizza place in this area needs to be very selective. A location should be selected that does not have many other pizza dining options, has other highly liked venues, and will bring in a large customer base. From the data in this report, it is recommended that the restaurateur open her new pizza restaurant in Troy. The suburb is large but there are not many competing pizza places in the vicinity. The average rating of other venues is acceptable. The main benefit of opening a place here would be the draw of customers from the neighboring suburb of Sterling Heights, which is also a large suburb that is also lacking in other pizza dining options. A pizza restaurant located in this area would be able to thrive in the competitive metro Detroit area.