ITALIAN RESTAURANT TARGET AREA IDENTIFICATION BATTLE OF THE NEIGHBORHOODS FINAL

TABLE OF CONTENTS

- Introduction

- Background

- Data

- Methodology

- Results

- Discussion

- Conclusion

INTRODUCTION

- Client is an Italian Restaurant entrepreneur who is searching for a neighborhood for a new Italian restaurant.

- Narrowed down to 2 areas Manhattan, and Downtown Toronto.

- I have been hired as a data scientist to analyze these areas.

- Client has strict criteria of neighborhoods with low competition for Italian cuisine, and a proportional number of overall restaurant venues, with a lack of Italian cuisine.

BACKGROUND

- NYC is the most densely populated city in USA, and Toronto is most densely populated city in Canada.

- Both cities do have Little Italy's section of high quantities of Italian cuisine.

- Both cities also have areas have many busines travelers and tourists that add to possible clientele.

- Either of these can be a prime opportunity, however my client's approach is key to identify low competition areas with many other restaurant venues. The approach has worked in many other cities.
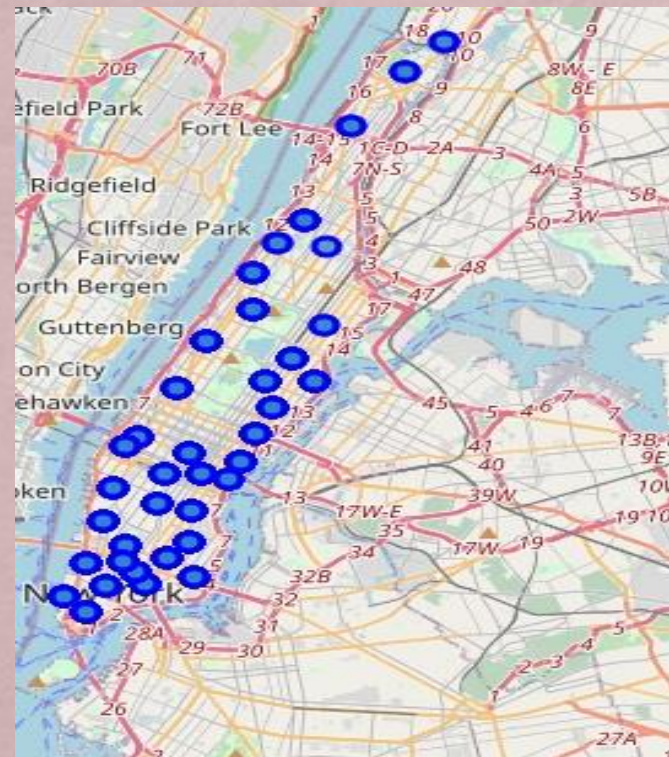
DATA

- Imported various libraries including Pandas, json, geopy, requests, sklearn, and folium to assist with this analysis.

- Analyzed Foursquare for venue categories that just included restaurants, and also venue categories that just included Italian restaurants.

- Pulled in the Wikipedia data for Toronto from a previous lab that needed scraped from the HTML website, and the NYC data that already had what we needed and no scraping was needed.

METHODOLOGY

- I cleansed the data pulling out any values of NA.

- Various issues I had to work through, one was the Wikipedia names continually changed from Post Code to Postal Code, and from Neighborhood to Neighbourhood so I had to rename columns.

- Key Issue was that when analyzing just the restaurant category in Foursquare not every neighborhood had a value returned, so I figured out I had to drop these rows when I did my clustering so I did not receive the error that the length did not match when merging in the last step for the KMeans.

METHODOLOGY CONTINUED

- Analyzed only Downtown Toronto and Manhattan.

- Rendered a visual map of these boroughs in Folium.

METHODOLOGY COTINUED

- Performed an API call to Foursquare for all restaurant venue categories, and then a second API call for just Italian restaurant venue categories.

- Created data frames for all of this data, and then used the one hot encoding to analyze the 10 most popular restaurant venues for each neighborhood.

- Clustered each of these neighborhoods to analyze similarities.

- Mapped the clusters so we can see what clusters meet my client's criteria.

METHODOLOGY CONTINUED

- Created data frames for just Italian restaurants category and visualized these for additional analytics on competition.

- The reasoning for using KMeans was to see how similar, dis-similar, homogeneous the groups of neighborhoods are.

- This information will be of key assistance in making a determination to see what areas are considered for placement of the new Italian restaurant.

RESULTS

- 1,365 restaurant venues in downtown Toronto, and 2,787 in Manhattan. These were maxed at 100 for Foursquare purposes.

- 3 Neighborhoods hit the max level of 100 Commerce Court, First Canadian Place, and Toronto Dominion Center. For Manhattan 12 neighborhoods maxed out at 100.

- There were 89 unique categories in Toronto, and 126 unique categories in Manhattan.

- When analyzing one hot for the most common restaurant venues Toronto neighborhoods had Italian restaurants in the top 5 in 13 of 37 neighborhoods, and 20 of 37 neighborhoods had Italian restaurants in the top 10 most common.

RESULTS CONTINUED

- When analyzing one hot for the most common restaurant venues Toronto neighborhoods had Italian restaurants in the top 5 in 22 of 40 neighborhoods, and 23 of 40 neighborhoods had Italian restaurants in the top 10 most common.

- However, NYC had 15 neighborhoods out of 40 where Italian restaurants had the number 1 ranked most common venue. Toronto had 0 neighborhoods where the ranking was 1.

- In the KMeans clustering for Toronto cluster 0 has 2 neighborhoods, cluster 1 has 32 neighborhoods, cluster 2 through 4 each had only 1 neighborhood. So, this analysis illustrates that 86% of the neighborhoods are together in 1 cluster for Toronto, so there is very high homogeneity.

RESULTS CONTINUED

- In the KMeans clustering for Manhattan cluster 0 has 10 neighborhoods, cluster 1 has 15 neighborhoods, cluster 2 has 13 neighborhoods, and cluster 3 and 4 each had only 1 neighborhood. So, this analysis illustrates that 95% of the neighborhoods are within 3 clusters for Manhattan, so there is a large amount of differentiation.

- For overall Italian restaurants Toronto has 202, and NYC had 893, none of the neighborhoods hit the max since we narrowed the category down to just Italian restaurants.

- The top 5 neighborhoods were in line with the neighborhoods with the highest volume of overall restaurant venues.

DISCUSSION

- Downtown Toronto and Manhattan had a similar number of neighborhoods, however the Manhattan area is 3 times larger in terms of square miles so that has to be taking into account as well.

- Manhattan had double the amount of restaurant venues, and more unique restaurants which is to be expected since the area is much larger.

- Toronto's neighborhood were much more homogeneous with 86% of those neighborhoods were in 1 KMeans cluster. Manhattan had more differentiation with the largest cluster only making up 38% of the neighborhoods.

DISCUSSION CONTINUED KMEANS CLUSTER MAP
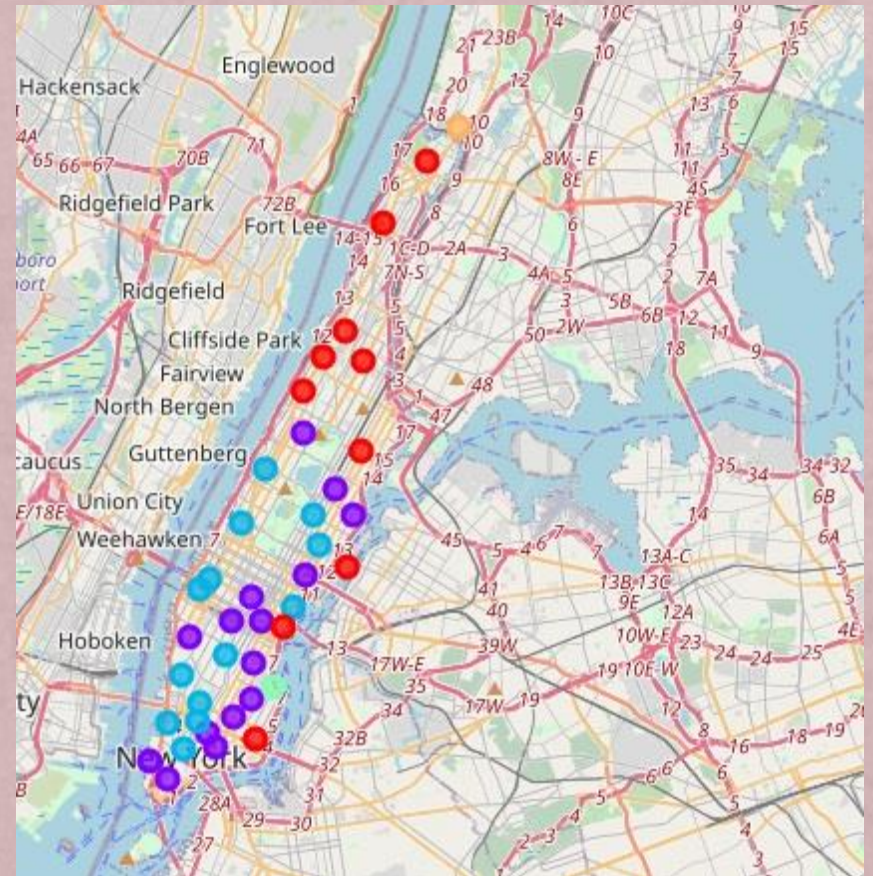
DISCUSSION CONTINUED

- In the one hot encoding illustrated that Toronto had 35% in the top 5 restaurant venues, and 54% in the top 10. NYC had 55% in the top 5 along with 58% in top 10. So, the top 10 were similar in terms of proportional percentages. However, NYC had a whopping 15 neighborhoods with an Italian restaurant as the number most common 1 venue. Toronto had 0 neighborhoods where an Italian restaurant made up the number 1.

- Analyzing total number of Italian restaurants both cities followed the similar trend of the venues with the highest overall restaurant venues also had the highest overall number of Italian restaurants for the most part.

DISCUSSION CONTINUED

- Manhattan had many more Italian restaurant venues, however the size is roughly 3 times larger in terms of square miles. I would still conclude that Italian restaurants are more popular in Manhattan since the discrepancy in volumes is even larger than the square mile difference.

- NYC also outnumbers Toronto 15 to 0 in neighborhoods where Italian restaurants are ranked number 1.

- Client needs an area with lower competition and a solid number of overall venues we can drill into the data and make some recommendations.

DISCUSSION CONTINUED

- First, my client has had success in large cities in low competition areas, however my client always takes into consideration the overall number of venues as well so they do not set up shop in an area that does not have many people going out to eat. So it is a fine line between low competition and a decent number of overall venues of what fits the bill for my client.

- Basically, we start by ruling out any top 5 ranked most common Italian restaurants venues in neighborhoods. For Toronto the clustering does not really show much differentiation, while NYC shows a lot of dis-similarity between neighborhoods. I would use the clusters that have Italian restaurants in them as being popular and choose the like neighborhoods with Italian restaurants that are lower than top 5 most common ranked or not on the list. Ruling out the top 5 removes almost 50% of the neighborhoods.

DISCUSSION CONTINUED

- When with the remaining venues I look for a median to higher volume total venue area with a lack of Italian restaurant competition. For NYC I exclude clusters 0 which has no Italian restaurants, cluster 2 which is highly competitive for Italian restaurants, and cluster 3 and 4 which have limited information. That leaves me with cluster 1 as my selection. I rule out the higher ranked Italian restaurants in that cluster that leaves me with 6 possibilities that meet my clients criteria Chinatown, Midtown, Murray Hill, Manhattan Valley, Battery Park, and Midtown South.

DISCUSSION CONTINUED

- Toronto has high homogeneity and a lack of Italian restaurants in top positions. For the Toronto clusters I exclude the 4 clusters that had 2 or less results, and choose the largest cluster. In the largest cluster I eliminate all of the neighborhoods that have an Italian restaurant rated in the top 10 most common venue per my client's requests. This leaves me with 14 Toronto neighborhoods Richmond, Queen's Park, The Beaches, Dufferin, Little Portugal, Lawrence Park, Davisville North, Forest Hill, The Annex, Kensington Market, Summerhill West, CN Tower, Rosedale, Church and Wellesley, and Studio District. This means in this growing diverse market there should be a large opportunity to open a new Italian restaurant.

DISCUSSION CONTINUED

- There will need be next steps in this final analysis of the location will be a deep dive into income per capita, rents, crime, safety, public transportation on these final neighborhoods in NYC and Toronto to make a final decision. The KMeans clustering, mapping, and Foursquare research in Italian and all restaurants categories is not enough to deliver a final answer to my client. We have narrowed the list from 77 neighborhoods down to 20 possibilities which may meet my client's criteria.

CONCLUSION

- New York and Toronto seem to be dis-similar in terms of restaurant venues. Toronto has high homogeneity within restaurant venues shown in clustering, while NYC does not. NYC has many neighborhoods with Italian restaurants as the number 1 most popular venue, Toronto does not. Following my client's instructions, and using KMeans clustering allows me to give my client a narrowed down for further research in income per capita, crime, transportation, tourists, foot traffic, ethnic make-up and place his restaurant in a popular growing area with a lack of good Italian restaurants.