Pizzeria DiMari

Neighborhoods

Contents

Background & Objective

Background

Giuseppe DiMari is a renowned business man from little Italy, Manhattan. He has a few restaurants, mainly pizzerias, in that neighbourhood and he is thinking on expanding his operation into other neighbourhoods in other towns in America although he does not discard further growth in NY. Giuseppe has some friends in Chicago, Miami and Toronto. Thus he would like to select which neighbourhoods are the most similar to Little Italy based on the competition and nearby venues so that he can create a business in there. He is not just interested in one only neighbourhood, he wants a group of neighbourhoods he can chose from. If there are also neighbourhoods in NY that are similar to Little Italy, he would like to know those as well. Also, he is not particularly afraid of the competition, if anything, he prefers a neighbourhood with competition since he know that this will help his Pizzeria to stay at its best. After all, it will be his inexperienced brother, Julio DiMari, who will be managing the restaurant. And he needs some healthy competition to become a successful business man.

Objective

Identify which neighborhoods are the most similar to Little Italy(NY) in Miami, Chicago, Toronto and the broader neighborhoods from NY. From these, identify which ones have enough competition.

Target Audience & Data

Target Audience

The target audience is just Giuseppe and possibly Donny. Perhaps the brother Julio will also be there but this will be just a project of business owners seeking clear guidance on similar neighbourhoods to Little Italy.

Data

First, it will be downloaded from Wikipedia and previous modules the neighbourhoods from NY, Chicago, Toronto and Miami. For NY and Toronto we will use the data from other modules. For Miami and Chicago the data will be downloaded from Wikipedia

After the names of the neighbourhoods are obtained, we will obtain the latitude and longitude of each neighbourhood. We will also double check that no error has occurred (ie: neighbourhoods located in LA).

After the neighbourhoods and geospatial locations of these is obtained and these have been confirmed to belong to each city, we will download from Foursquare the 100 nearest venues to each geospatial address.

Method

1. We will obtain all the Neighbourhoods from Toronto, Miami, Chicago and NY. Neighbourhood data will be downloaded from Wikipedia or use existing data from other course modules. This data will be cleansed to ensure that these hoods are indeed in these cities. It is possible that some of the geolocation is incorrect, assigning latitudes and longitudes well outside the boundaries of the city. Thus, as part of the data preparation stage, the neighbourhoods will be plotted. Neighbourhoods located at +/-0.5 latitude or longitude from the center of the city will be eliminated from the data set.

2. We will obtain all the venues per neighbourhood from Foursquare. We will characterise each neighbourhood using the most common types of venue.

3. We will classify the neighbourhoods into clusters based on the types of common venues using K-mean clusters. Number of clusters will also be optimised so that there are no more than 25 similar neighbourhoods in that group. The list of neighbourhoods has to be sufficiently small for Giuseppe and his associates.

It has been chosen the K-Nearest neighbours since it is a simple and unsupervised classification algorithm. Also given the method uses distances to calculate "clusters" it is key the data is first normalised to ensure reliable clustering. Thus the "one hot encoding" method has been used to standardise the data. It is also very possible that the clusters consist in a significantly bigger group than other. In that case K will be increased so that the subset/cluster size is sufficiently small.

4. The cluster that contains Little Italy, NY, will be the target subgroup of neighbourhoods.

5. After clustering, we will find those neighbourhoods that do have a pizzeria or an Italian restaurant in the top 3 common venues. This is since the customer explicitly asked for some level of competition due to the ownership of the venue being potentially delegated to a family member.

Data Preparation- Neighborhoods

As anticipated in the methodology section, while obtaining the location for each neighborhood, it was observed that for Chicago and Miami some of the neighborhoods were incorrectly located. Images below using Folium
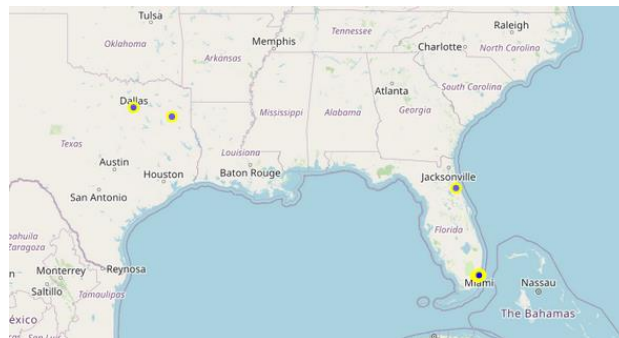

Figure 1 Chicago neighbourhoods

Figure 2 Miami neighbourhoods

A restriction was applied to the neighbourhoods so that the latitude and longitude cannot be outside +/-0.5 of the latitude and longitude of the city. Neighbourhoods that were incorrectly located were eliminated from the data set. With this, there are 406 neighbourhoods to compare.

Figure 3 After data preparation for latitude and longitude

Data Preparation- Venue Types

After the correct geolocation was obtained for each neighborhood, the 100 nearest venues to that location were downloaded. This resulted into 13081 venues. With this and to enable the one hot encoding methodology, the venues were classified by venue type. There were 468 different types of venues in the 406 neighborhoods.

Modeling – Data analysis

The K means clustering was performed. As defined in the methodology, it was expected that some groups would be significantly greater than others. With K=5, it was observed that the group in which Little Italy was contained was greater than 100 neighbourhoods. Thus an iterative process was completed. It was found that with K=200 groups would be sufficiently small to enable later comparison. At its core, the K-means algorithm simply tries to obtain the data points which are closer to each other based on a distance metric (i.e. Euclidian distance). Thus, increasing K simply permits classifying into smaller data sets.

After this analysis, some findings were observed

1. Most of the more similar neighbourhoods to Little Italy, are actually in NY

2. Within NY, the neighbourhoods more similar to Little Italy are actually in Manhattan.

3. There is no similar neighbourhood to Little Italy in Chicago

4. Between Miami and Toronto, there are more similar neighbourhoods in Miami

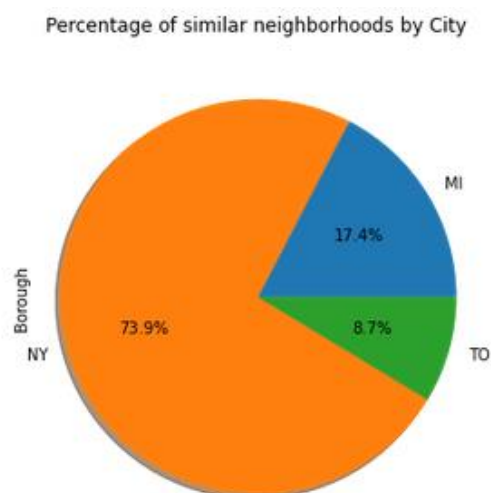To support these findings the following charts were plotted.

Figure 4: Percentage of similar neighbourhoods by City

Also, within NY City, several neighbourhoods could be found to similar venue characteristics as Little Italy. Although the vast majority are located in south Manhattan, there are some in Harlem, Brooklyn and the Bronx.
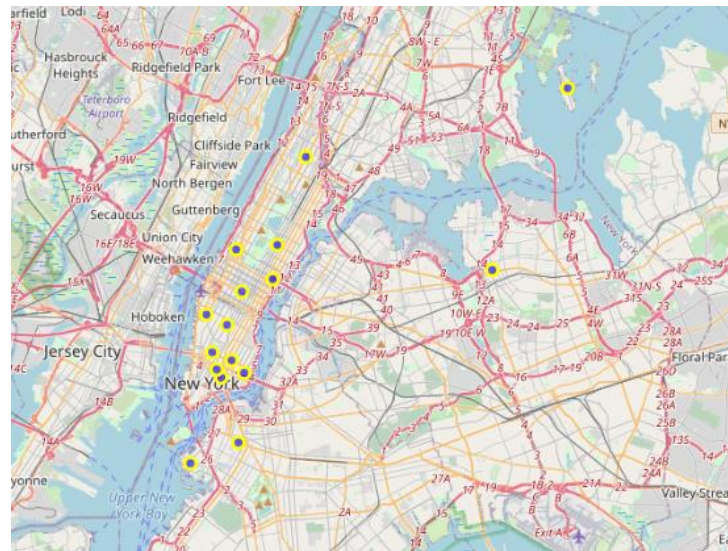

Figure 5 NY similar Neighbourhoods to Little Italy

Modelling – Competition

From the last cluster with K=200, it was filtered which neighbourhoods have in the top 3 common venues a Pizzeria or an Italian Restaurant. With this, the results still show that.

1. Most of the competition is actually in Manhattan, NY.

2. There are no neighbourhoods in Toronto with this criteria (neither Chicago)

3. In Miami there is also one neighbourhood that shares the characteristics of Little Italy and the competition desired.

Results

There are 23 neighbourhoods with similar venue profile as little Italy, Manhattan. Out of these, there are only four neighbourhoods with the desired level of competition. These neigbhoorhoods are:

	Borough	Neighborhood	Latitude	Longitude	City
209	Manhattan	Greenwich Village	40.7269	-73.9999	NY
340	Manhattan	Noho	40.7233	-73.9884	NY
368	Manhattan	Flatiron	40.7397	-73.9909	NY
399	Brickell	Brickell	25.7634	-80.1955	MI

Figure 6 Neighbourhoods

6

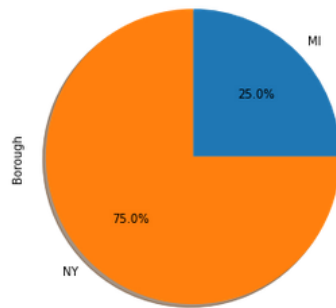Percentage of similar neighborhoods by City with competition

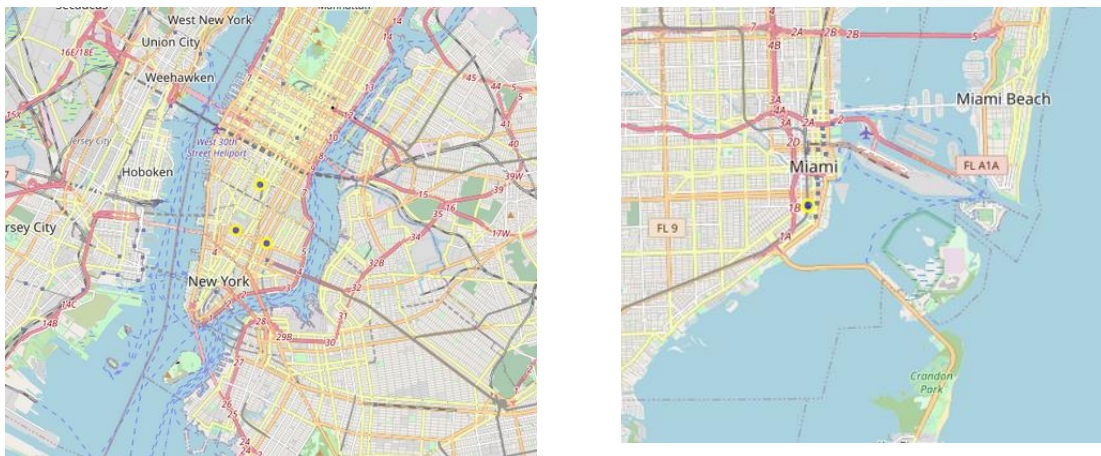**Figure 7 Neighbourhoods with competition**

Figure 8: Location of neighbourhoods NY Left; Miami, Right

Feedback

This material was presented to the DiMari family on Feb 2021. The reception was welcoming and grateful of the results and the perception was indeed aligned with the expectations from Julio and Giuseppe which were already thinking on starting a venture in Miami.