Where to open a pizzeria in Etobicoke?

by Adil Rakhmetov
Course: Applied Data Science Capstone Project
IBM Data Science Professional Certificate Program
February 2019

Introduction

This report is prepared on request from a small business entrepreneur to make data analysis to find a location which is the best fit to open a pizza delivery and walk-in service venue in the western part of Toronto - Etobicoke.

To have the most successful business for this kind of venues the location is the key factor. So we found out 2 main criteria for such places:

- Easy access for the potential customers,

- To be remotedly situated away from the similar competitors' venues or alternatives.

For this purpose we analyzed the data obtained from Foursquare API for the venues in this area arranged by the neighbourhoods in Etobicoke. We applied the clustering approach to find out the specific places regarding the venues located there. Then we analyzed the geographical distribution of existing pizzerias in Etobicoke to find the locations where this type of service will have the most demand.

Data Section

We used the data from public sources like Wikipedia, Google Maps and Foursquare API.

METHODOLOGY.

To solve this task we apply two approaches:

1. K means clustering to find out which areas of Etobicoke distinguish from each other depending on what venues and facilities they have.

2. Data visualization and chart analysis.

K-means clustering.

As we know, Etobicoke is the western borough of Toronto City. So we selected the postal codes which belong to Etobicoke from Wikipedia site (https://en.wikipedi-a.org/wiki/List_of_postal_codes_of_Canada:_M). Then we found the coordinates of each postal codes from Geocoder API and grouped around them the neighbourhoods. Next we obtained from Foursquare API the coordinates of all the venues in proximity of 1500 m to each postal code location (because the mean distance between the postal code locations in Etobicoke, estimated by Google maps scale, is around 3000 m). We applied K means clustering algorithm to the venues categories according to their popularity in each neighbourhood. The best result is obtained with clustering to 3 types. Analyzing the data of each cluster we found that the clusters 1 and 2 is more attractive for our purpose than the cluster 0, because they don't have the Pizzerias and Italian

restaurants among the top 10 most common venues on their territory. These are the
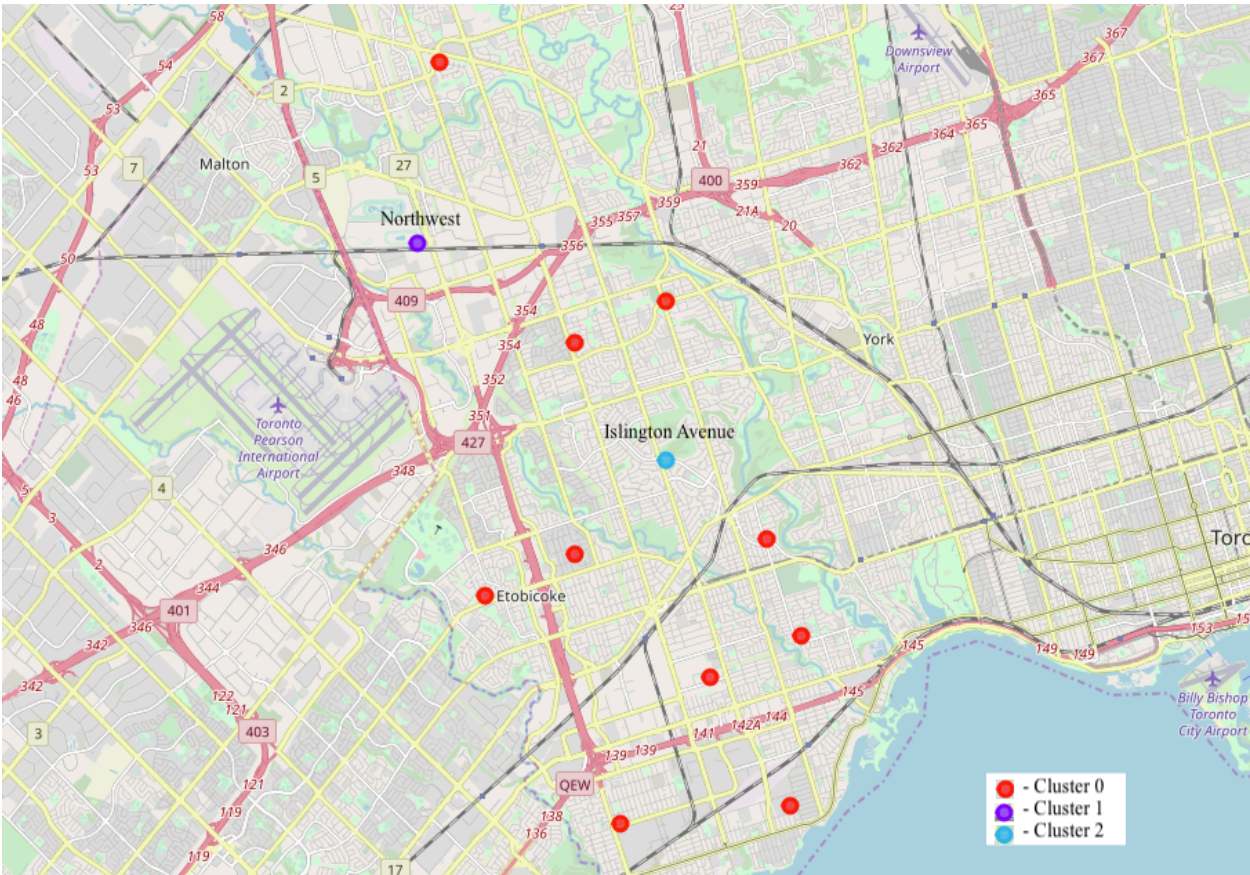
neighbourhoods "Islington Avenue" and "Northwest".

FIG. 1   K-MEANS CLUSTERED NEIGHBOURHOODS.

Cluster 1.

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Northwest	1	Coffee Shop	Sandwich Place	Rental Car Location	Paper / Office Supplies Store	Dog Run	General Travel	Mediterranean Restaurant	Hotel	Racecourse	Casino

Cluster 2.

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Islington Avenue	2	Café	Liquor Store	Playground	Grocery Store	Shopping Mall	Skating Rink	Camera Store	Pharmacy	Bus Stop	Bus Line

TABLE 1.   TOP 10 MOST COMMON VENUES IN CLUSTERS 1 AND 2.

Data visualization.

To be more specific to our task, we considered the venues which belong to the categories "Pizza Place" and "Italian Restaurant". We indicated them on the map of Etobicoke together with the clustered neighborhoods locations. It is obvious that neighbourhoods "Islington Avenue" and "Northwest" are the most remote areas from the pizza venues.
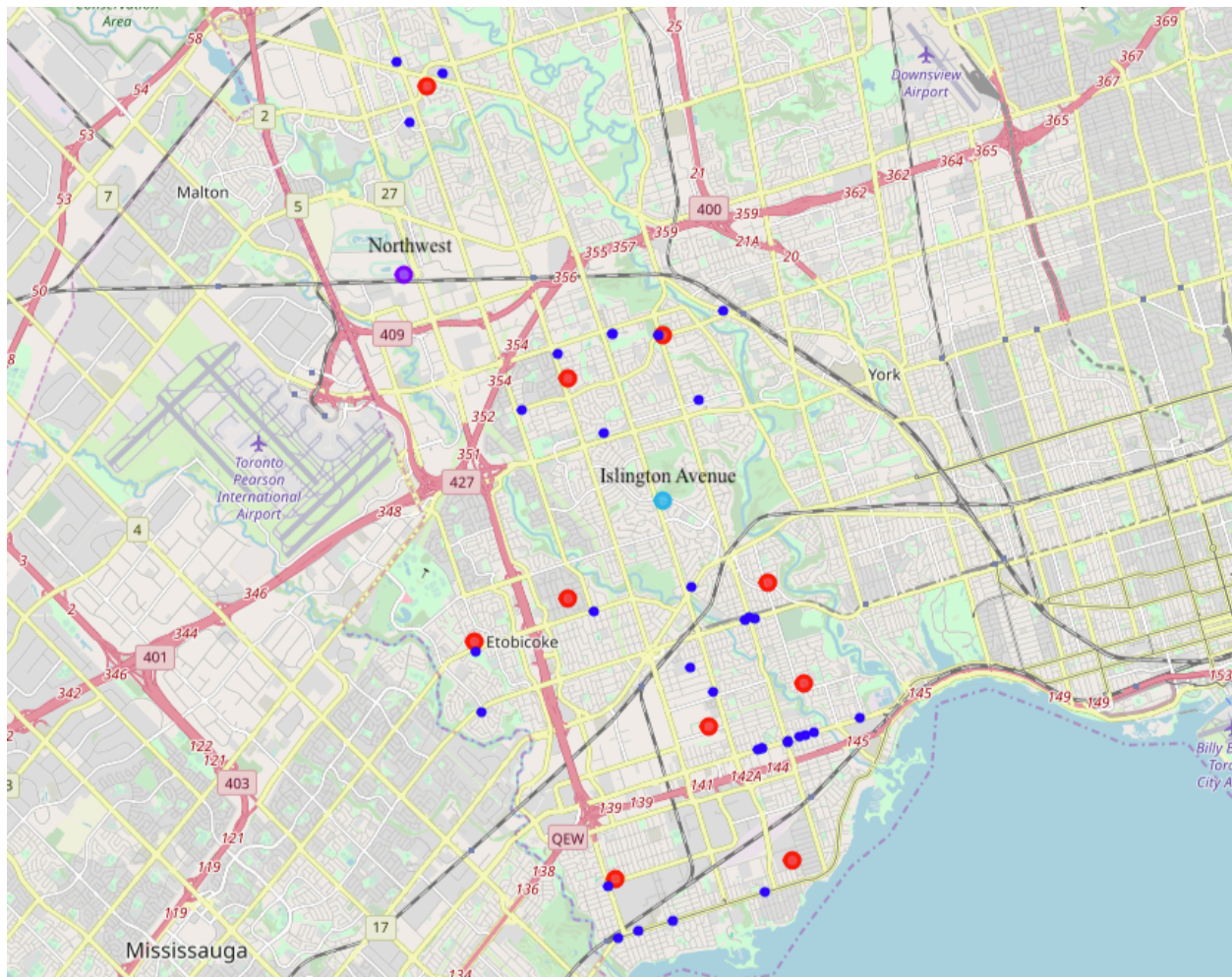

FIG 2. PIZZA VENUES AND NEIGHBOURHOODS

Also we counted how many venues of those categories are situated in each neighbourhood of Etobicoke and visualized it with the bar charts. Surprisingly, these two neighbourhoods have no any pizzerias or italian restaurants within the 1.5 km range.
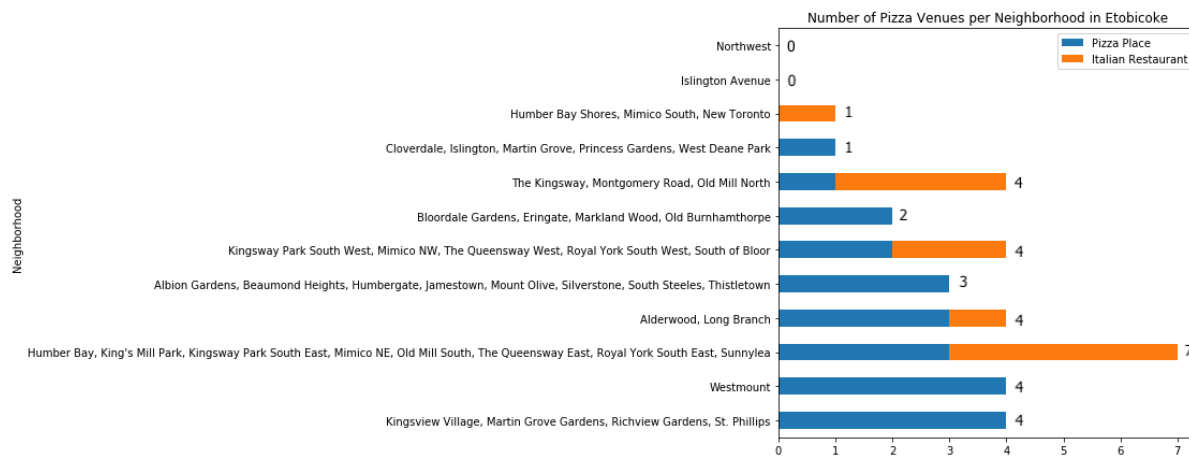

FIG 3.   NUMBER OF PIZZA VENUES PER NEIGHBOURHOOD IN ETOBICOKE.

Discussions and thoughts.

Analyzing the obtained data we found that despite we applied different approaches we received the same result concerning to our task. It proves that both approaches are realistic to our goals and should give the right aspects to our customer. Considering the environment of these two neighbourhoods we can see that "Northwest" is adjacent to Toronto International Airport, and "Islington Avenue" mostly contains the residential areas. Probably it would be more challenging to provide the delivery service in Northwest neighbourhood during the daytime due to heavy traffic conditions in that area. But in the evenings and at nights the delivery in the airport area with a lot of hotel and accommodation facilities the delivery service might be in big demand. For "Islington Avenue" neighbourhood the walk-in service is also would be popular due to many residential houses located within walking distance.

Conclusions.

1. Based on two independent approaches of data analyzing , K-means clustering and Data visualization - we found two perspective areas in Etobicoke to open a pizza venue - neighbourhoods "Northwest" and "Islington Avenue".

2. Depending on whether delivery or walk-in service is going to be primary, we recommend to consider "Northwest" neighbourhood to provide mostly delivery service, and "Islington Avenue" - for both delivery and walk-in service.