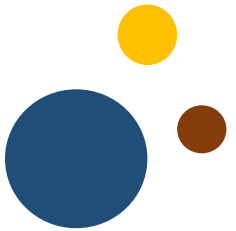
IBM PROFESSIONAL CERTIFICATE IN DATA SCIENCE CAPSTONE:

EMMIGREATION (Pty) Ltd:
Helping clients find similar neighbourhoods in foreign cities



INTRODUCTION

Emmigreation (Pty) Ltd

Emmigreation is a company that was founded to help people move seemlessly from one city to another. The company has noticed that when previous clients have moved to new neighbourhoods in new cities that happen to show some similarity to their previous home, they were far happier than those who moved to neighbourhoods that showed no similarities.

This has led to Emmigreation hiring an IBM Data Scientist to determine if there was a way that the company could propose specific neighbourhoods to their clients to help achieve this outcome using a more scientific approach. The following report will determine how this can be done as well as going through an existing client's options for moving into a new city.

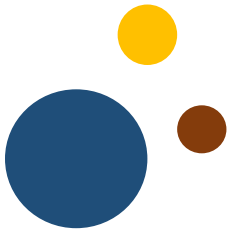
DATA AND METHODOLOGY

DATA

- The data that will be analysed will be the neighbourhoods of Toronto together with the neighbourhoods of New York. The data sets will be acquired the same way as before in the IBM Capstone course including the coordinates of each of the neighbourhoods.
- Using the Foursqaure API, we will be able link the closest/most-common amenities to these neighbourhoods to build up the data set required for the analysis.
- Once we have each neighbourhood from Toronto and New York City in one data set, with associated coordinates plus top 10 most common amenities within a certain radius, then we will be ready to run our analysis.

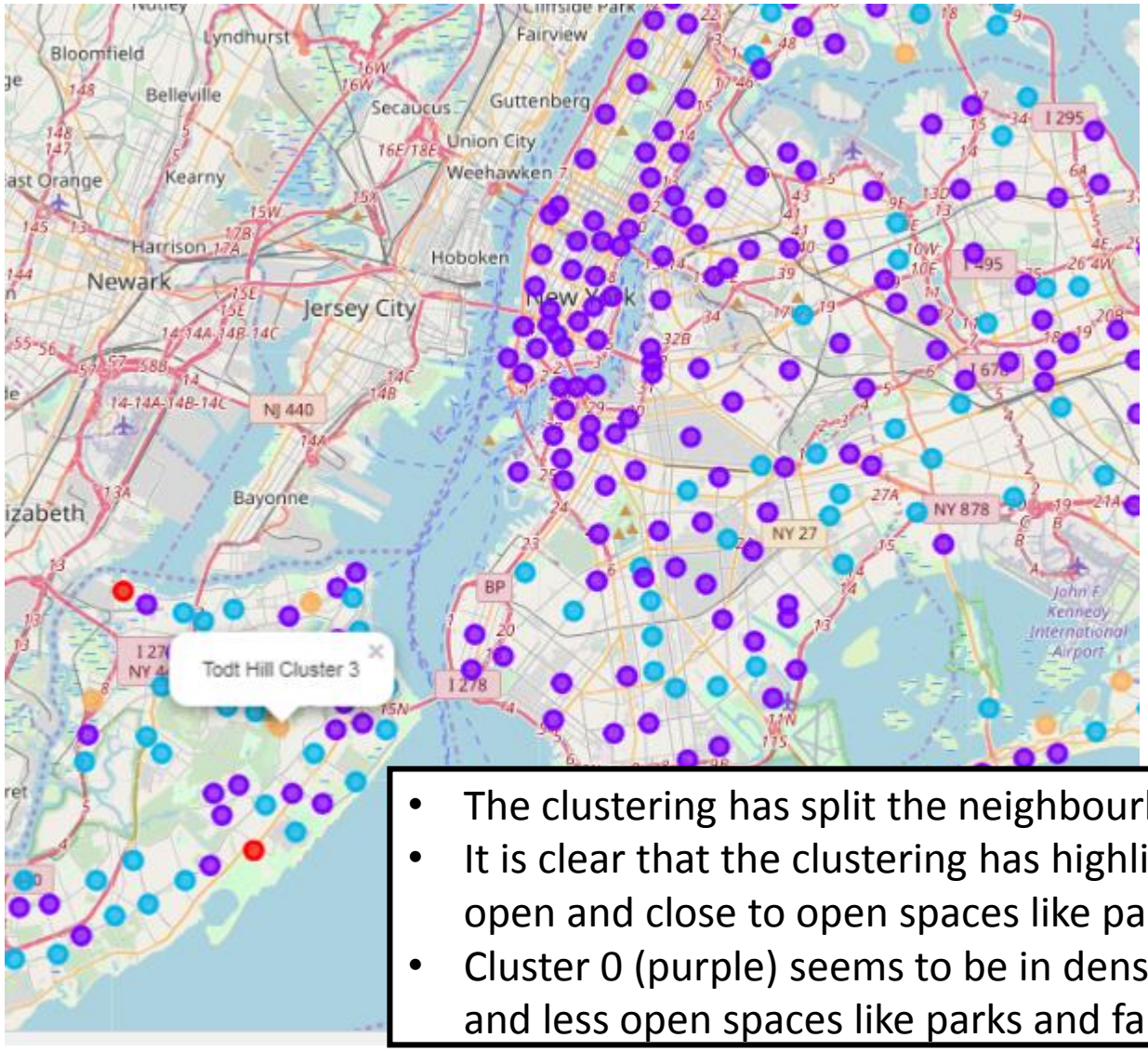
METHODOLOGY

- The methodology will use web scraping to pull data for Toronto and New York City neighbourhoods.
- The latitudes and longitudes of these neighbourhoods will then be linked and turned into a dataframe for analysis.
- The Foursquare API will then be utilised to link nearby amenities to the different neighbourhoods, which will then begin to build up profiles/characteristics for each neighbourhood.
- Using this dataset, we will then run a k-means clustering machine learning algorithm to determine which of the neighburhoods are similar. From this dataset, we can determine which Toronto neighbourhoods are similar to those of New York City.
- Emmigreation will be able to use this dataset to guide its clients on where they should consider moving to when moving in between these cities. This will achieve the outcomes set out in the introduction above.
- Lastly, we will look at one specific example of a client wanting to move from one neighbourhood to another across the aforementioned cities.

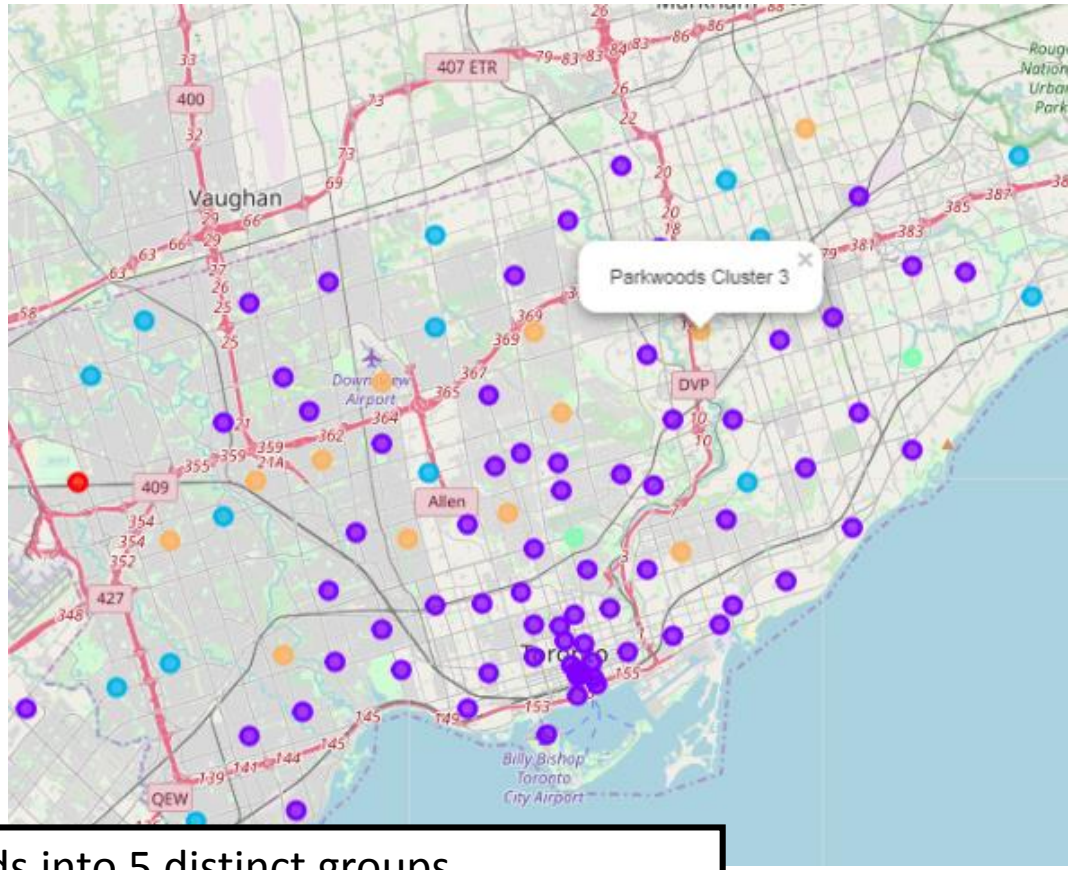


RESULTS

New York Neighbourhood Clusters



Toronto Neighbourhoods Clusters



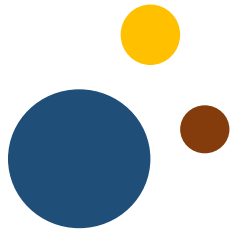
- The clustering has split the neighbourhoods into 5 distinct groups.
- It is clear that the clustering has highlighted neighbourhoods that are more open and close to open spaces like parks in cluster 3 (orange).
- Cluster 0 (purple) seems to be in denser areas with more high-rised buildings and less open spaces like parks and farms.

RESULTS: USE CASE

Jimbo Champster is a resident of Toronto and has recently take a job offer in the high-paced city of New York. He has signed with Emmigreation to assist him with his move and make it as pleasant a possible. Part of the new Emmigreation offering is to pinpoint neighbourhoods that will siilar to Jimbo's current neighbourhood to help make the move easier and more enjoyable.

Jimbo's current neighbourhood is that of Parkwoods nighbourhood in Toronto. If we look at the make up of the common amenities in the radius of the area, it is evident that this neighbourhood seems to be suited to someone who likes open spaces, nature and dining (particularly exotic dining such as Spanish and Ethiopian cuisine) See below for the profile of Parkwoods neighbourhood.

Borough	Neighborhood	City	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
North York	Parkwoods	Toronto	3	Food & Drink Shop	Park	Fast Food Restaurant	Women's Store	Farm	Electronics Store	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Event Service
Staten Island	Todt Hill	New York	3	Trail	Park	Women's Store	Farm	Electronics Store	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Event Service	Event Space



CONCLUSION

If Emmigreation were to propose a similar feel neighbourhood to that of Parkwoods for Jimbo, then Todt Hill was found to be in the same cluster. It is evident that Todt Hill has similar features to Parkwoods as it seems to have the same outdoors, open-space kind of profile with exotic dining options available for Jimbo to enjoy.

In this case, the k-means clustering algorithm has helped to pinpoint a great location for Jimbo to conisder moving to. This will definitely aid in making the move more pleasant and easy as well as helping Jimbo settle in his new spot much quicker than another neighbourhood which may be drastically different.

This same methodology can be appied to future clients to improve the client experience plus provide an added service offering for Emmigreation.