Applied Data Science Capstone Project

Introduction

This is the final project in the IBM Data Science Professional Certificate. In this project I explore facilities in the Downtown Toronto area, focusing my attention on Italian restaurants. Here I take the role of an entrepreneur based in Canada who is aiming to open an Italian restaurant in the central business district of Toronto but, knowing that competition may be high in such a crowded place, I use data science methods and machine learning algorithms to find the best location to open my activity. So, this project is aimed to anyone who wants to open an Italian restaurant in the Downtown Toronto area.

Data

For this project I will need the list of postal codes and neighborhoods in Toronto, which will be scrapped from the Wikipedia page List of postal codes of Canada: M, along with their geographical coordinates that I will obtain from Geospatial data. Using the Foursquare API, I will obtain information about the venues in the area of interest in order to find Italian restaurants in Downtown Toronto.

Methodology

First of all, I use BeautifulSoup to scrap the list of postal codes and neighborhoods of Toronto from Wikipedia and then I clean the data by removing "Not assigned" entries and grouping neighborhoods with the same postal code. I then obtain the geospatial data and I merge this information with the postal code data to obtain the geographical coordinates of each neighborhood, limiting the analysis to Downtown Toronto.
I use the Foursquare API to pull the list of the top 100 venues in a 500 meter

radius from each neighborhood. With this data I obtain the name, category, and coordinates of each venue. Now I am able to find the Italian restaurants in Downtown Toronto.
Using the unsupervised machine learning algorithm K-means clustering, I divide the neighborhoods in Downtown Toronto into three clusters based on their frequency of occurence of Italian restaurants.

Results

With K-means clustering I now have three clusters of neighborhoods in downtown Toronto. Cluster 0, in red, indicates the area with the less amount of Italian restaurants. Cluster 1, in blue, is the area with a medium amount of Italian restaurants. Cluster 2, in green, indicates the neighborhoods with the highest amount of Italian restaurants.

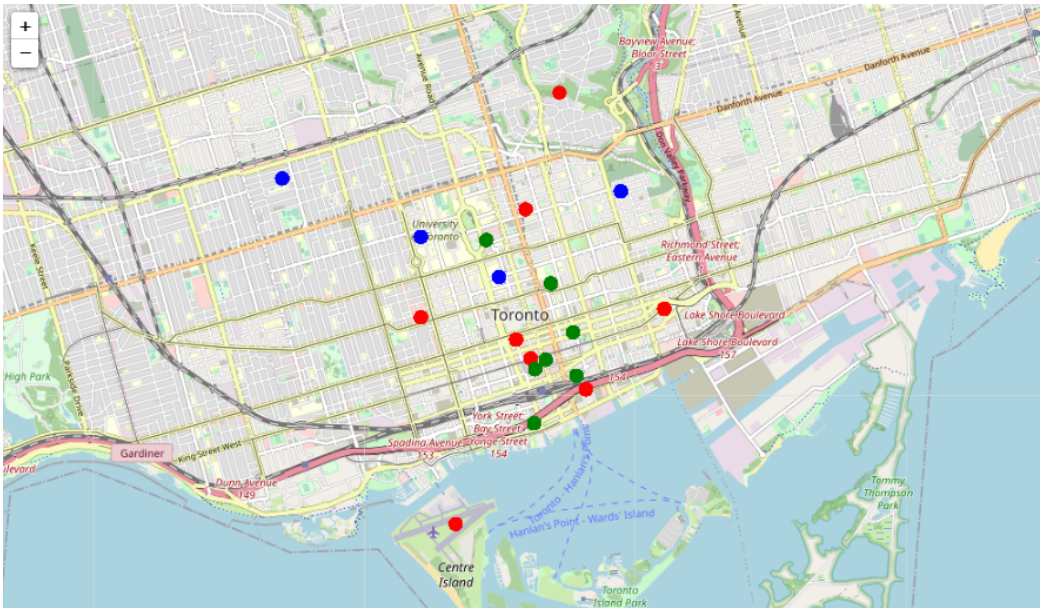

Figure 1: Clusters in Downtown Toronto

Discussion

With this analysis we can see that in Cluster 0 there are only two nearby Italian restaurants, located near First Canadian Place and Church and Wellesley.

Other neighborhoods in this area include Berczy Park, CN Tower, Kensington Market, Regent Park, Richmond and Rosedale.

Conclusion

After completing this project I would open my Italian restaurant in one of the locations of Cluster 0 listed above as these areas offer the less amount of competition and the best opportunity for success.