APPLIED DATA SCIENCE CAPSTONE PROJECT

IBM PROFESSIONAL DATA SCIENCE CERTIFICATE

INTRODUCTION

- I explore facilities in the Downtown Toronto area, focusing my attention on Italian restaurants.
- I take the role of an entrepreneur based in Canada who is aiming to open an Italian restaurant in Downtown Toronto but, knowing that competition may be high in such a crowded place, I use data science methods to find the best location to open my activity.
- This project is aimed to anyone who wants to open an Italian restaurant in the Downtown Toronto area.

DATA

- I will need the list of postal codes and neighborhoods in Toronto, which will be scrapped from Wikipedia.
- I also need the geographical coordinates for each neighborhood, I will obtain those from the cocl.us website.
- 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.

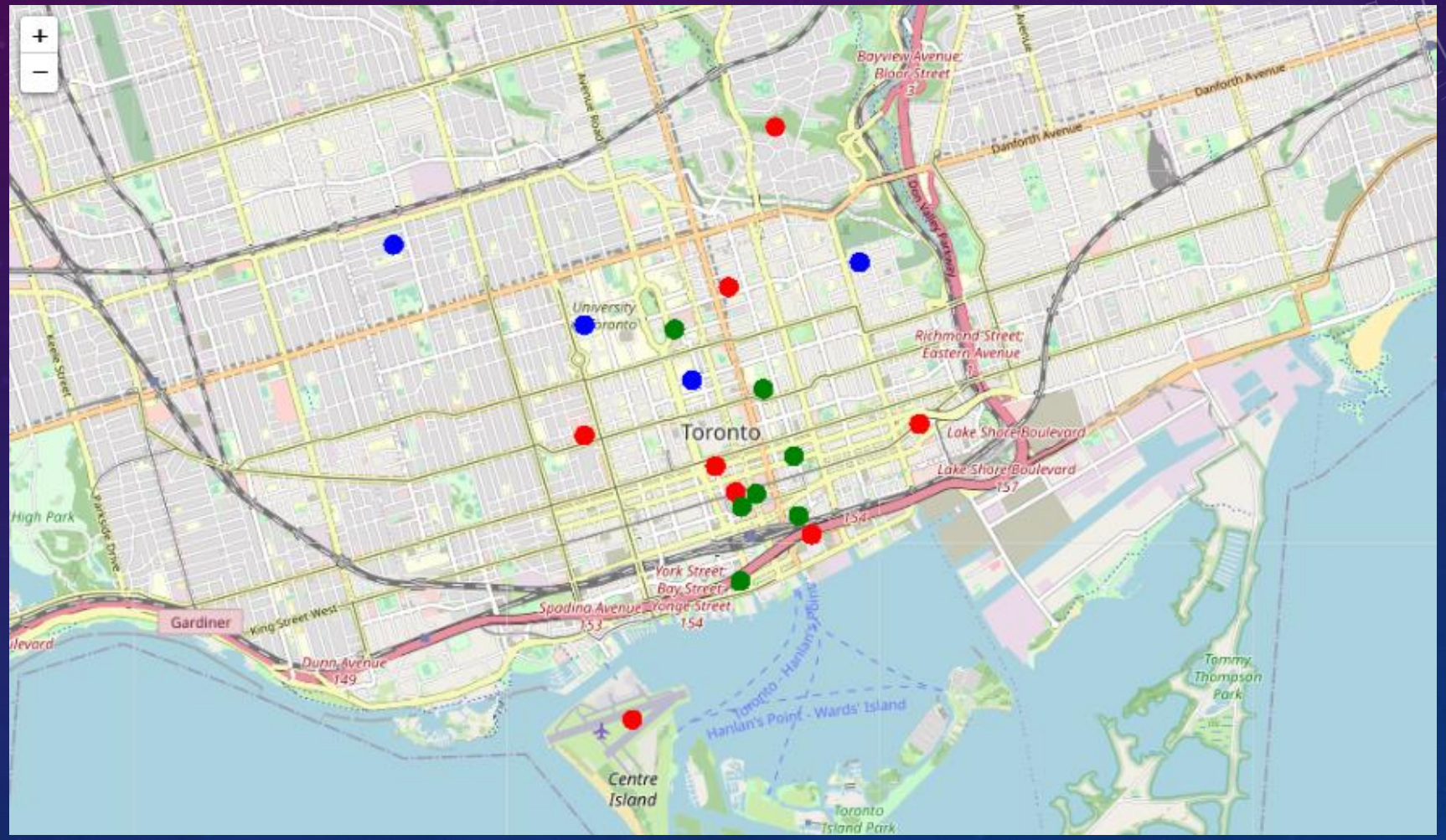
METHODOLOGY

- I use BeautifulSoup to scrap the list of postal codes and neighborhoods of Toronto from Wikipedia and then I clean the data.
- I 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. I obtain the name, category and coordinates of each venue. Now I am able to find the Italian restaurants in the area.
- Using 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, is 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, is the neighborhoods with the highest amount of Italian restaurants.

MAP OF THE CLUSTERS



DISCUSSION AND CONCLUSION

- In Cluster 0 there are only two nearby Italian restaurants, located near First Canadian Place and Church and Wellesley.
- Locations in Cluster 0 without Italian restaurants include: Berczy Park, CN Tower, Kensington Market, Regent Park, Richmond and Rosedale.
- I would open my Italian restaurant in one of these locations as these areas offer the less amount of competition and the best opportunity for success.