The Battle of Neighborhoods

Capstone project for Coursera IBM Data science professional certificate

Pierayeh Vahdani

- Introduction: Business Problem
- Data
- Methodology
- Results and Analysis
- Discussion
- Conclusion

Business Problem

In this project, I will compare the neighborhood of the city of Vancouver and the city of Toronto.

The purpose of the project is for a chain restaurant that would like to open a new branch in each city and tries to search to find the best location and a customized menu for each location.

The restaurant considers 2 factors to decide about the location and the menu as follow:

- The location should be in a popular neighborhood that can attract visitors.
- The menu for each location should includes foods that are less offered by other restaurants in the that neighborhood.

Data

Following data sources will be needed to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Wikipedia
- number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API
- Coordinate of Vancouver and its neighborhoods will be obtained using ‘the government of British Columbi's BC Address Geocoder website’**.**

Methodology

Based on definition of our problem, factors that will influence our decision are:

1- Location:

- Number of existing restaurants and other public venues such as parks, beaches, museums in the neighborhood.
- Distance of neighborhood from city center.

2- The menu:

- The dishes that are less offered by the restaurants in that neighborhood.

- Data from Wikipedia containing the borough, neighborhoods, and the associated postal codes is collected and converted to a pandas dataframe.

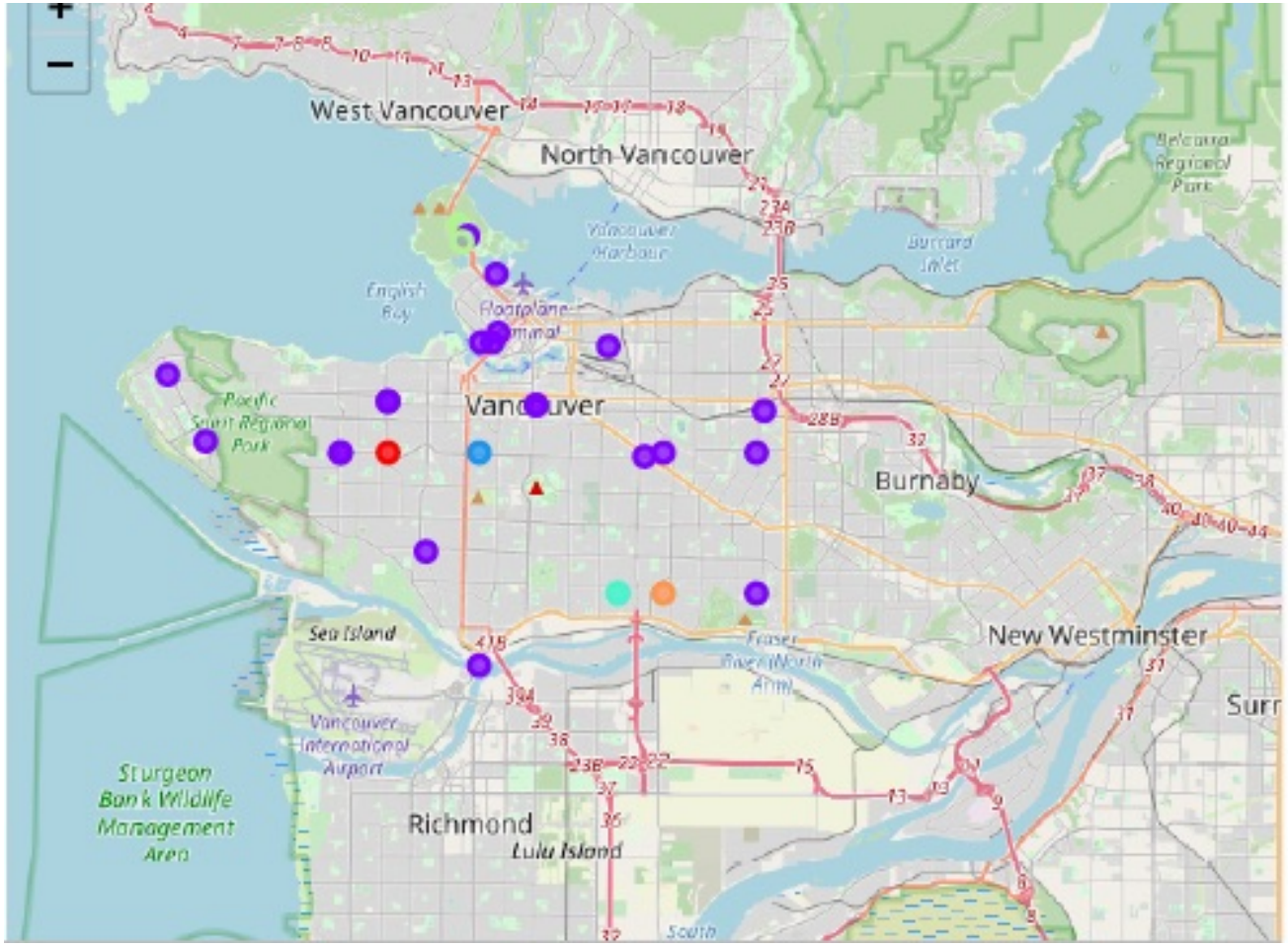
- Data including Vancouver borough, postal codes, and their corresponding geographical information is obtained from the government of BC address Geocoder website and converted to a pandas dataframe. I renamed some columns and re-indexd the dataframe.

- I used FourSquare to get 100 venues in Vancouver , then explored and clustered the neighborhoods in Vancouver and visualized the neighborhoods and how they clustered together.

The top 3 neighborhoods with the largest number of venues are:

- West Mount Pleasant, West Riley Park, Little Mountain
- North West Dunbar, Southlands, Chaldecutt, South University Endowment Lands
- SW Downtown

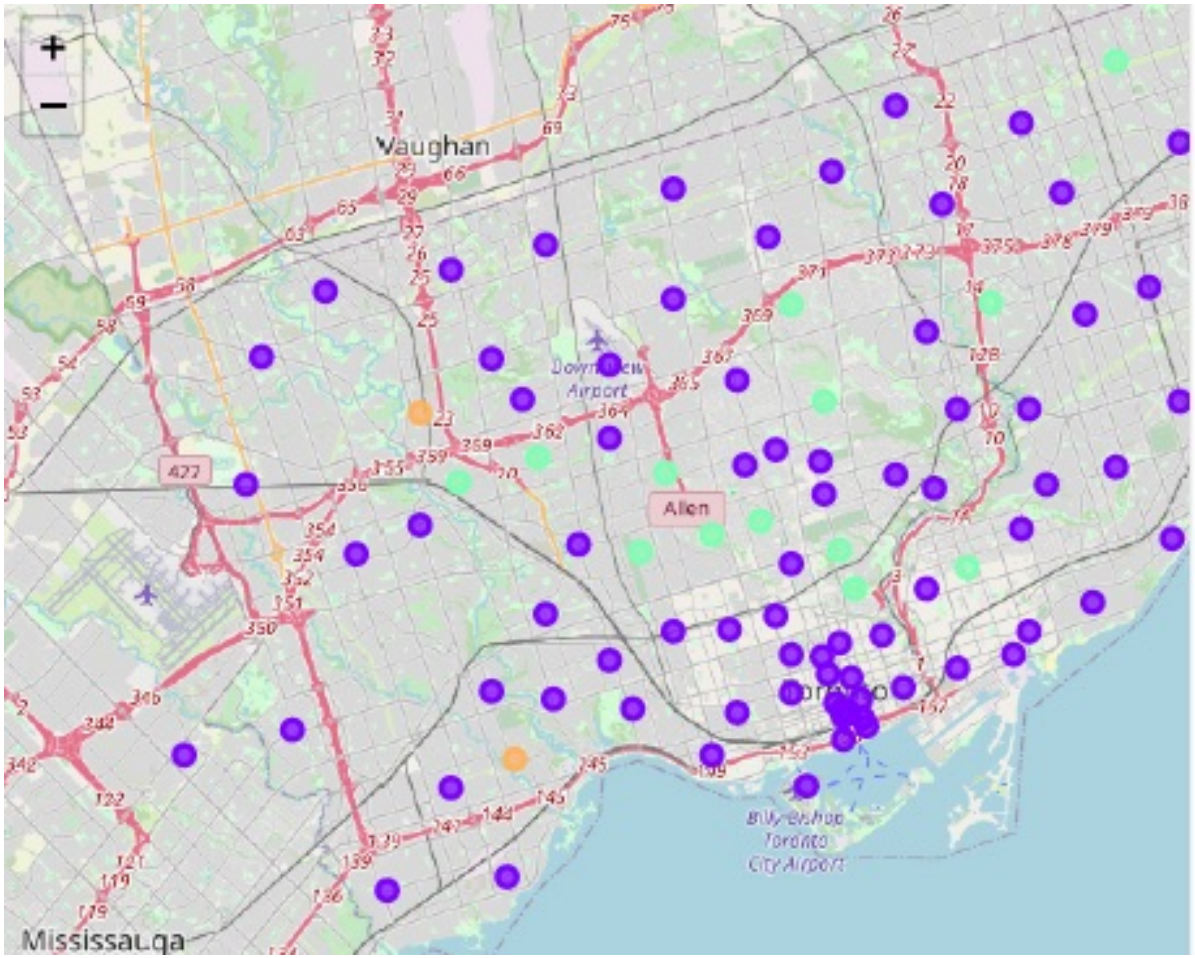
Map of Vancouver and different venues divided into 6 different cluster



I repeated the same process for the city of Toronto. The top 10 neighborhoods in Toronto with largest number of venues are:

- First Canadian Place
- Harbourfront East, Union Station, Toronto Islands
- Garden District, Ryerson
- Toronto Dominion Center, Design Exchange
- Commerce Court, Victoria Hotel
- Stn A P0 Boxes
- Richmond, Adelaide, King
- Church and Wellesley
- St. James Town
- Central Bay Street

Map of Toronto and different venues divided into 6 different cluster



Discussion

1. Vancouver neighborhoods

- All the top 3 neighborhoods in Vancouver are included in cluster number 2, which is the biggest cluster, shown in the figure with purple circles.
- Looking at top 10 most common venues in these neighborhoods, I concluded that Japanese, Chinese, Korean restaurants are the most common restaurants in these neighborhoods.
- The most common places in the second biggest cluster in Vancouver, cluster number 5, are Parks, lakes, and trials which attract many visitors specially during weekends.
- I would recommend to the chain restaurant to open a new branch in cluster number 2, best neighborhoods are those listed in the top 3 neighborhoods, or in the the neighborhoods within cluster number 5.
- I also recommend to customize the menu to include more Italian, Mexican, Mediterranean style of food.

- All top 10 neighborhoods in Toronto are included in cluster number 1, which is the biggest cluster, shown in the figure with purple circles.
- Looking at top 10 most common venues in these neighborhoods, I concluded that Japanese, Chinese, Korean restaurants are the most common restaurants in these neighborhoods.
- I would recommend to the chain restaurant to open a new branch in cluster number 2, best neighborhoods are those listed in the top 3 neighborhoods, or in the the neighborhoods within cluster number 5.
- I also recommend to customize the menu to include more Italian, Mexican, Mediterranean style of food.

Conclusion

- I demonstrated how to cluster and segment different neighborhoods in the central part of the city of Toronto and the city of Vancouver based on the venues in these neighborhoods.

- I used the results of this analysis to recommend to a chain restaurant the best neighborhoods in these two cities to open a new branch and a customized menu which includes the foods that are less commonly found in those neighborhoods.

- This project include Boroughs that are limited to the central part of the cities and not includes other interesting Boroughs, For example the data for North Vancouver and West Vancouver with potential attractive places was missing in this analysis.

- To improve the analysis, I would recommend to collect larger data set which includes other Boroughs to have a better understanding of different neighborhoods in both cities to open more branches in them.