



IBM-Coursera: Capstone Project Canadian Cities Comparison

Hugo Rocha De A.

Project Overview

- ▶ Moving to a different city can be quite a challenge as you require to take multiple decisions before you arrive to the final destination and there is limited information of the city's neighborhoods.
- ▶ This situation also applies to a scenario where a person wants to rent a flat/house by using Airbnb, but doesn't know if the selected selection is close to the most touristic venues or to a metro station/bus stop.
- ▶ This project presents a way to profile city's neighborhoods based on the venues found in them.
- ▶ The proposed scenario is that of a student trying to decide between moving to Montreal to study at McGill University or to Vancouver to study at UBC.

Data Obtention

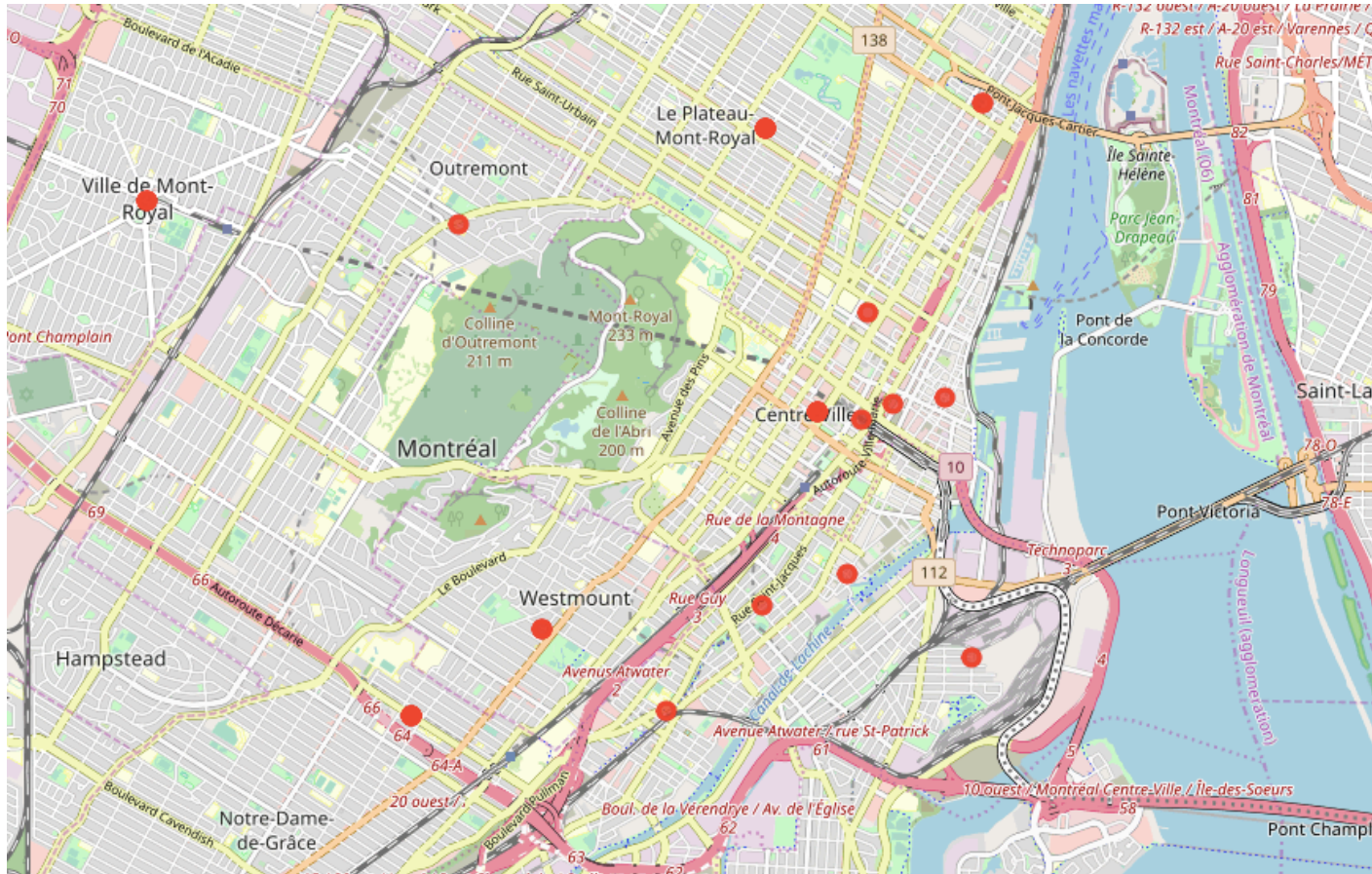
- ▶ I had to manually create each city's neighborhood data sets.
- ▶ Each data set contains information on:
 - ▶ Postal Code
 - ▶ Borough
 - ▶ Neighborhood
 - ▶ Neighborhood's Latitude
 - ▶ Neighborhood's Longitude
- ▶ Montreal's data set : 103 data points
- ▶ Vancouver's data set : 66 data points

Data Preparation

- ▶ In order to obtain the required venue data and carry out each city's neighborhoods clustering, these were the followed steps:
 - ▶ Reduce the number of studied neighborhoods, to those closer to the universities.
 - ▶ Use the Foursquare's API to obtain the venue data and transform it into data frames.
 - ▶ It was possible to identify 162 unique venue types in the selected Montreal's neighborhoods and 129 unique venue types in the selected Vancouver's neighborhoods.
 - ▶ Crop the obtained data into only the top 10 most visited venues.

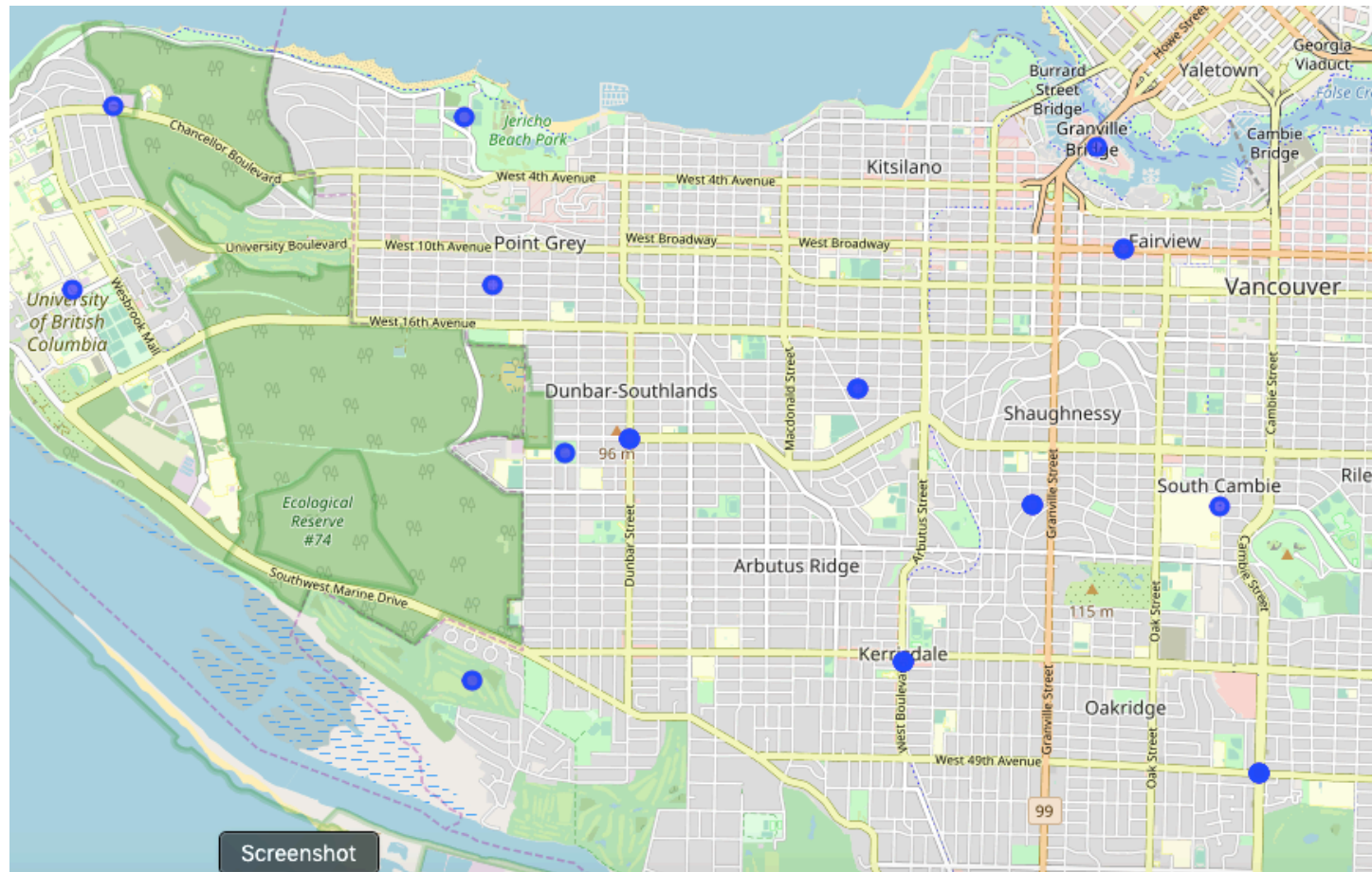
Data Preparation

- Geographical representation of Montreal's selected neighborhoods.



Data Preparation

- ▶ Geographical representation of Vancouver's selected neighborhoods.



Methodology

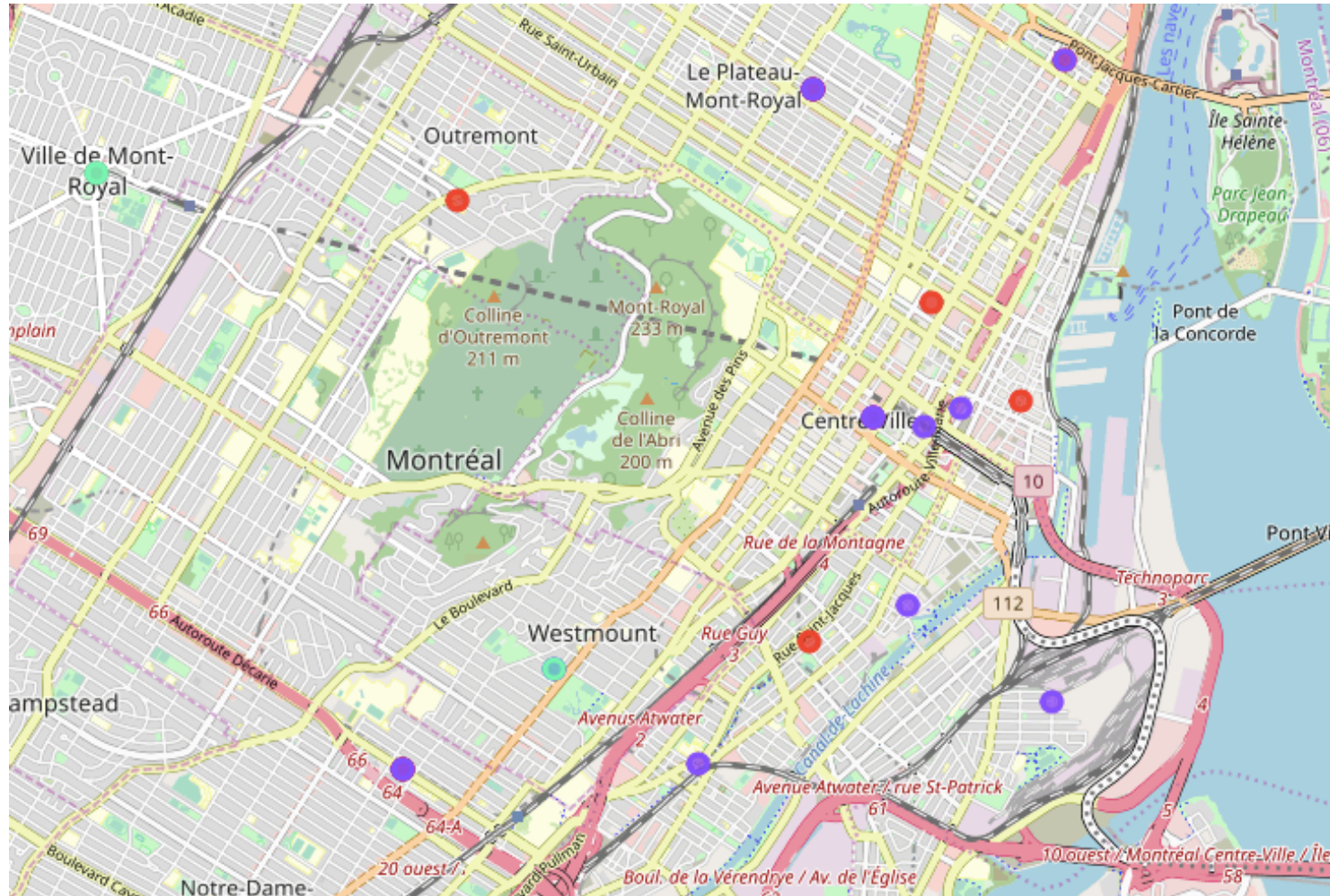
- ▶ In order to obtain the desired neighborhood segmentation, the following steps were followed:
 - ▶ Using the obtained top 10 most visited venues data frames, one hot encoding was performed to transform the categorical values into numerical values that can be feed to the K-Means clustering algorithm.
 - ▶ Feed the K-Means clustering algorithm with both data frames, the obtained categories are then joined with the data frames of each city's selected neighborhoods.
 - ▶ Finally, the data is geographically displayed in each city's map.

Results

- ▶ The found profiles of the selected Montreal's neighborhoods are:
 - ▶ Cluster 0 (red): The venues part of this cluster are mainly Asian restaurants, bakeries, coffee shops, hotels and bars.
 - ▶ Cluster 1 (purple): It is the largest clusters and the venues part of this cluster are a wide variety of restaurants, bakeries, coffee shops, tea shops, parks, banks, hotels and gyms.
 - ▶ Cluster 2 (green): The venues part of this cluster are mainly coffee shops, train stations, pharmacies, bakeries, parks, gyms, banks and some restaurants. Probably the neighborhoods part of this cluster, are living areas.

Results

- The found profiles of the selected Montreal's neighborhoods are:

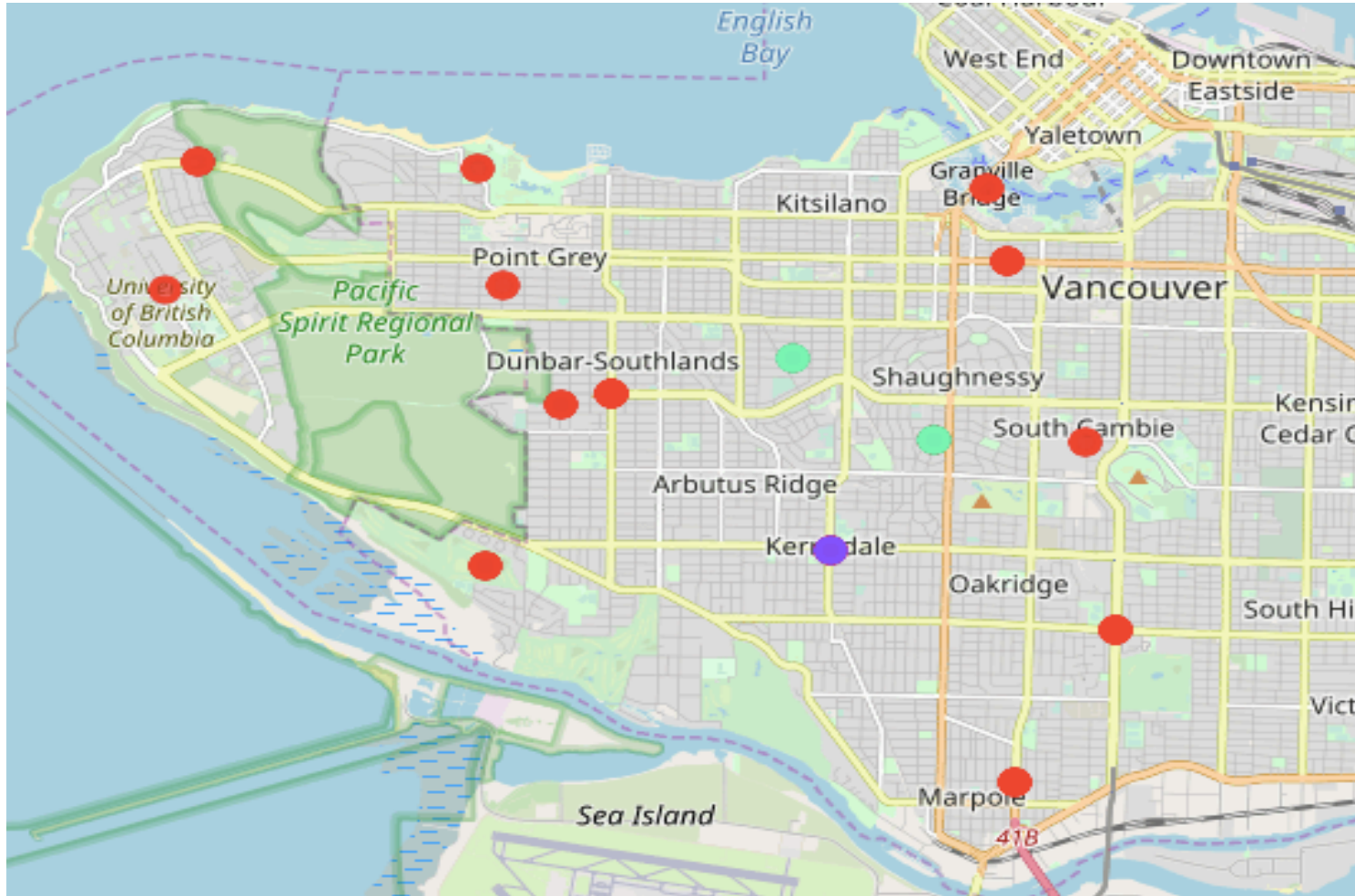


Results

- ▶ The found profiles of the selected Vancouver's neighborhoods are:
 - ▶ Cluster 0 (red): It is the largest cluster and the venues part of this cluster are restaurants, coffee shops, parks, tea shops, markets, grocery store, malls and sport centers.
 - ▶ Cluster 1 (purple): The venues part of this cluster are restaurants, coffee shops, parks, a golf course, gyms, tea shop, pharmacies and bars.
 - ▶ Cluster 2 (green): The venues part of this cluster are coffee shops, gardens, parks, restaurants, malls, bakeries, bus stops, furniture stores, art/crafts stores and event spaces. The neighborhoods in this cluster, are probably mainly living areas.

Results

- ▶ The found profiles of the selected Vancouver's neighborhoods are:



Results

- ▶ It was possible to observe that each city is very different.
- ▶ All of Vancouver's neighborhoods have parks, gardens or sport venues within them. In contrast, in Montreal, only the purple and green clusters have parks within them.
- ▶ Another interesting difference between these cities, is that only in the red Vancouver's cluster is possible to observe the presence of markets and grocery stores. In terms of restaurants, coffee shops and tea shops, both cities have a wide variety of this type of venues.

Conclusions

- ▶ It was possible to confirm that two cities can be compared based on the obtained profiles by K-Means clustering.
- ▶ An application based on this project can help a user to identify a flat/house within a neighborhood that matches the desired profile.
- ▶ Also, if it would be used to complement an application such as Airbnb, it would help clients to consider available rentals based on the surrounding venues and not only on price differentiation.