



CLUSTERING BIXI STATIONS BASED ON NEARBY VENUES

Coursera Capstone Project - IBM Data Science Professional Certificate

Amen Memmi

Introduction

- Bixi: Bike sharing system in Montreal, QC, Canada
- News: Plans to extend Bixi bike sharing network
- Idea: why not study the use patterns by clustering the end stations according to their nearby venues.
- This could come very handy to Bixi when deciding the locations and sizes of the new docking stations.

Data used for this project

- Bixi Montreal network use data is available in open data section on their website:
- <https://www.bixi.com/en/open-data>
- The data used comprise all the bike trips during 2018 as well as the stations locations
- For the venues available nearby docking stations, we are using the Foursquare API which allows us to explore and acquire venues locations and categories for a given location

Methodology

- Focusing the analysis on one particular station: *St-André / Cherrier*, code of 6175
- Keep only the trips that started from our station
- Analyzing the most important and recurrent trips patterns: focusing on the most popular ones, i.e., selecting the trips that went to the top 100 destination stations from our station
- Using Foursquare API requests to know what is around the end stations: surroundings venues, in a 100m radius, as well as their locations and categories are provided for each end station
- For each end station, top 20 most commons venues were selected and used as features for clustering using **K-means**, expecting to capture use patterns in the trips' most common end stations.

Results: Clusters

For $k=5$, the stations are distributed in clusters as follows:

```
Cluster 0: 15 stations
Cluster 1: 18 stations
Cluster 2: 62 stations
Cluster 3: 2 stations
Cluster 4: 2 stations
```

The most common venues per cluster:

```
Most common venues in cluster 0 are: ['Yoga Studio' 'Hot Spring' 'Dog Run' 'Intersection']
Most common venues in cluster 1 are: ['Bakery' 'Supermarket' 'Gourmet Shop' 'Liquor Store' 'Grocery Store']
Most common venues in cluster 2 are: ['Restaurant' 'Bar' 'Coffee Shop' 'Food & Drink Shop' 'Sandwich Place']
Most common venues in cluster 3 are: ['Yoga Studio' 'Bus Station']
Most common venues in cluster 4 are: ['French Restaurant' 'Dumpling Restaurant']
```

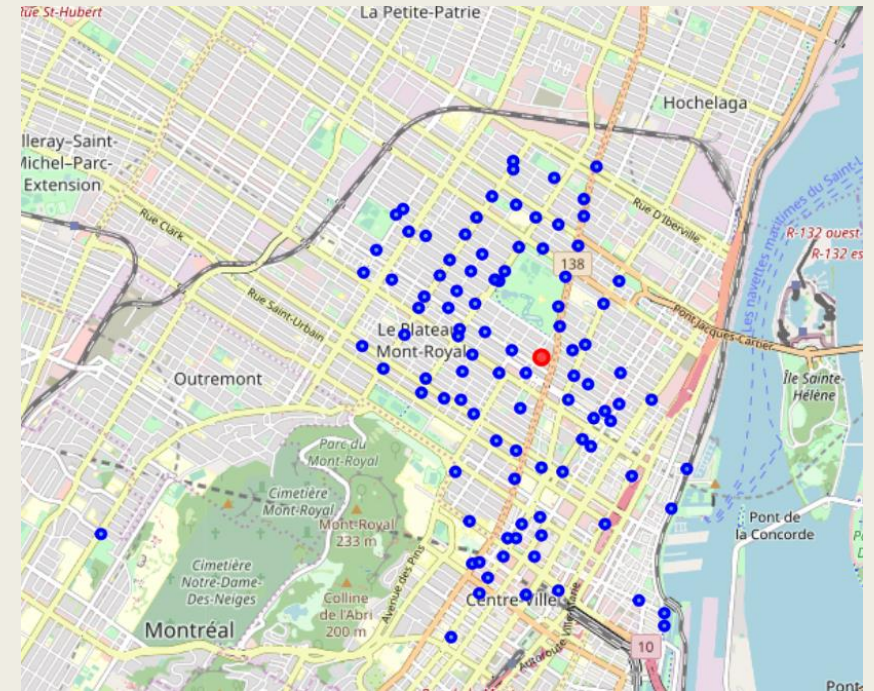


Figure 1: End Stations distribution

Results contd: Map

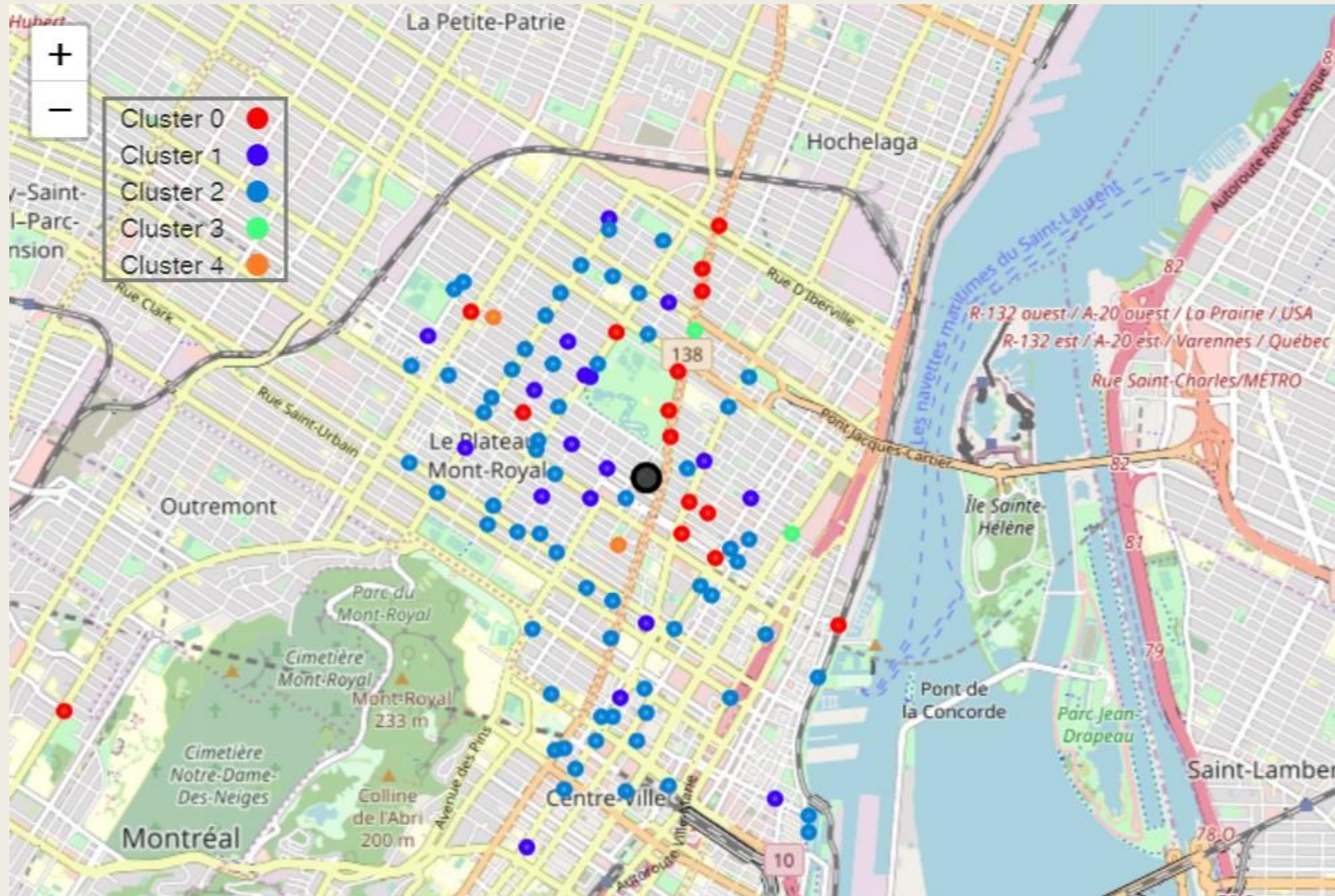


Figure 2: Stations' clusters distribution

Conclusion

- Interesting project
- make use of the Foursquare API venues for some interesting and potentially useful application
- promising results that could be generalized to all stations of the network
- while studying extension plans, such clusters help draw insights regarding the location and size of new dockings stations based on and optimized to match the venues and their categories in each neighborhood/ street.

A bright yellow rectangular sticky note is pinned to a light gray background with a single red pushpin at the top center. The note is slightly tilted and has a soft shadow beneath it. The words "Thank you!" are written in a dark gray, casual cursive script across the lower half of the note.

Thank you!