

The Battle of Neighborhoods



Introduction

In this project for the capstone course for Applied Data Science, we shall cluster and compare neighbourhoods in Toronto on the basis of popular venues. At the end of the exercise, we shall have multiple groups of neighborhoods based on the most popular venue. The goal is to better understand the differences between the neighborhoods in Toronto.

Business Problem

The goal of this project is to have a better understanding of the neighborhoods in Toronto. These informations could be used later by investors to choose in which neighborhood they should open new venues and develop their business.

Data

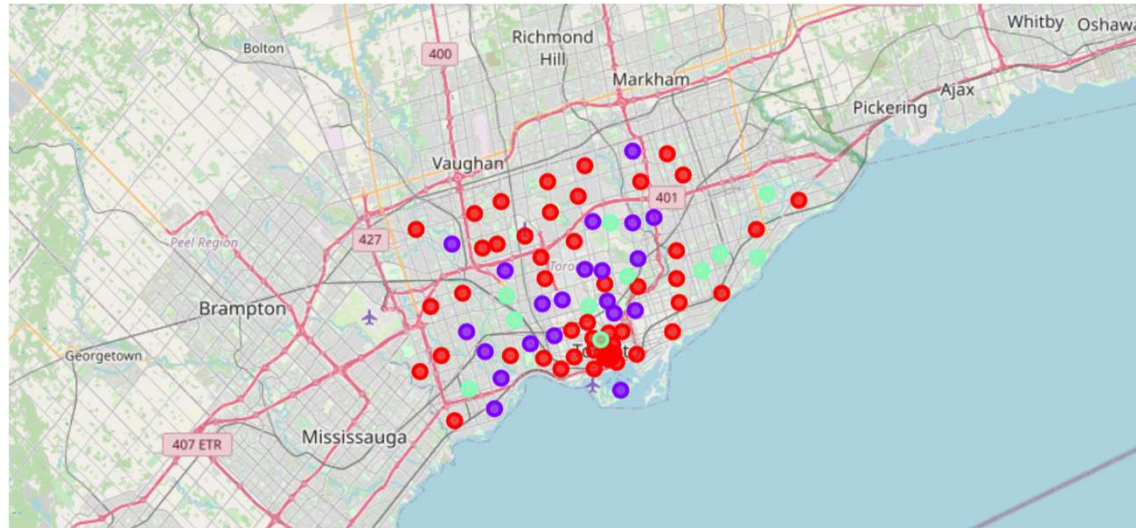
In this projet we will be using data collected from Foursquare; we will use the Foursquare API to explore neighborhoods in Toronto. You will use the **explore** function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters.

Methodology

In order to get a better vision of the venues distribution among the neighborhoods we're going to focus only of the top 10 venus in Toronto. Having limited the number of venus we shall be able to have a simplified view of neighborhoods différences. Once we have the data in the right format we're going to use kmeans clustering to divide the neighborhoods into 3 groups.

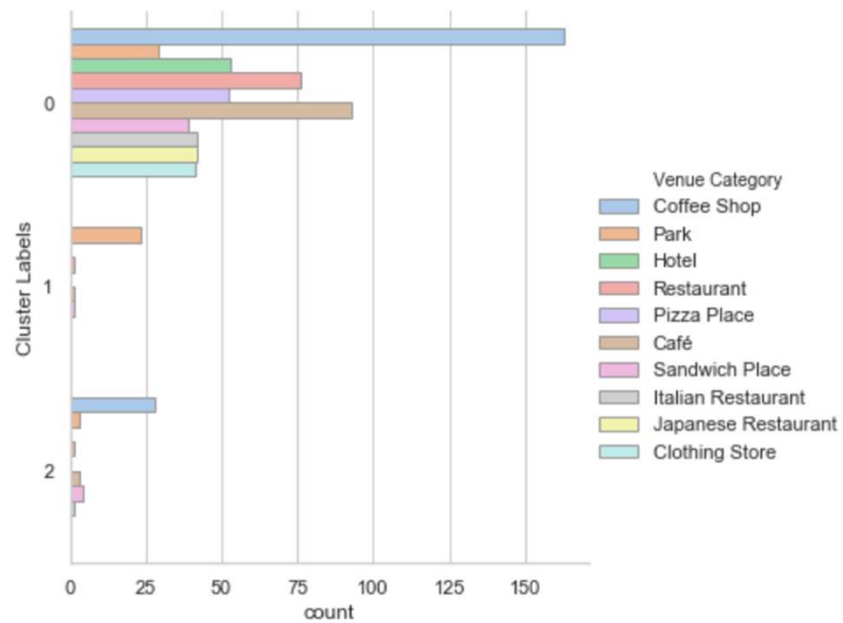
Results

We have now completed the clustering phase and have this map illustration of the 3 clusters:



Results

To better understand the difference between the cluster, we have studied the number of venues for each cluster by venue type:



Results

Cluster 0

It is the cluster indicated in red in the map above, it represent neighborhoods with a high concentration of venues, mainly food related venues such as coffee shop, café and restaurants.

Cluster 1

It is the cluster indicated in violet in the map above, it represent neighborhoods with a lower concentration of venues but a lot of parks nevertheless.

Cluster 2

It is the cluster indicated in light green in the map above, it represent neighborhoods with a lower concentration of venues but a lot of coffee shops nevertheless.

Discussion

There is an important factor which should be added to this analysis regarding the population of each neighborhood. there is certainly correlation between the demographic aspect and the venues available.

Conclusion

I believe this analysis present a quick overview of the differences between neighborhoods in term of venus which is very interrsting. Nevertheless this a lot of analysis that could be added to complete this work.