IBM Applied Data Science Capstone

BATTLE OF THE NEIGHBORHOODS

On directing tourists towards venues accessible by public transport.

Problem:

- Montreal is an international destination for tourists from around the world.
- Being new to the city, most tourists make the mistake of getting around by car, leading to traffic.
- Can we encourage the use of public transport by tourists by showing them which subway stations contain the best venues?

Data:

Two data sources will are used for this project. Montreal's Open Data Platform, and Foursquare's API.

Geolocation coordinates for Montreal's 68 subway stations.

Obtained form the City of Montreal's Open Data Platform.

Data on the venues nearby each of the subway stations.

Obtained by querying Foursquare's API

Methodology:

A lot of work had to be done on both dataframes in order to remove unnecessary information & format them.

- Transforming the shapefile obtained from the Open Data Format.
- Get venue data from Foursquare.
- Group venue data by subway station.
- Perform k-means clustering on the data.
- Visualize clusters on a map.

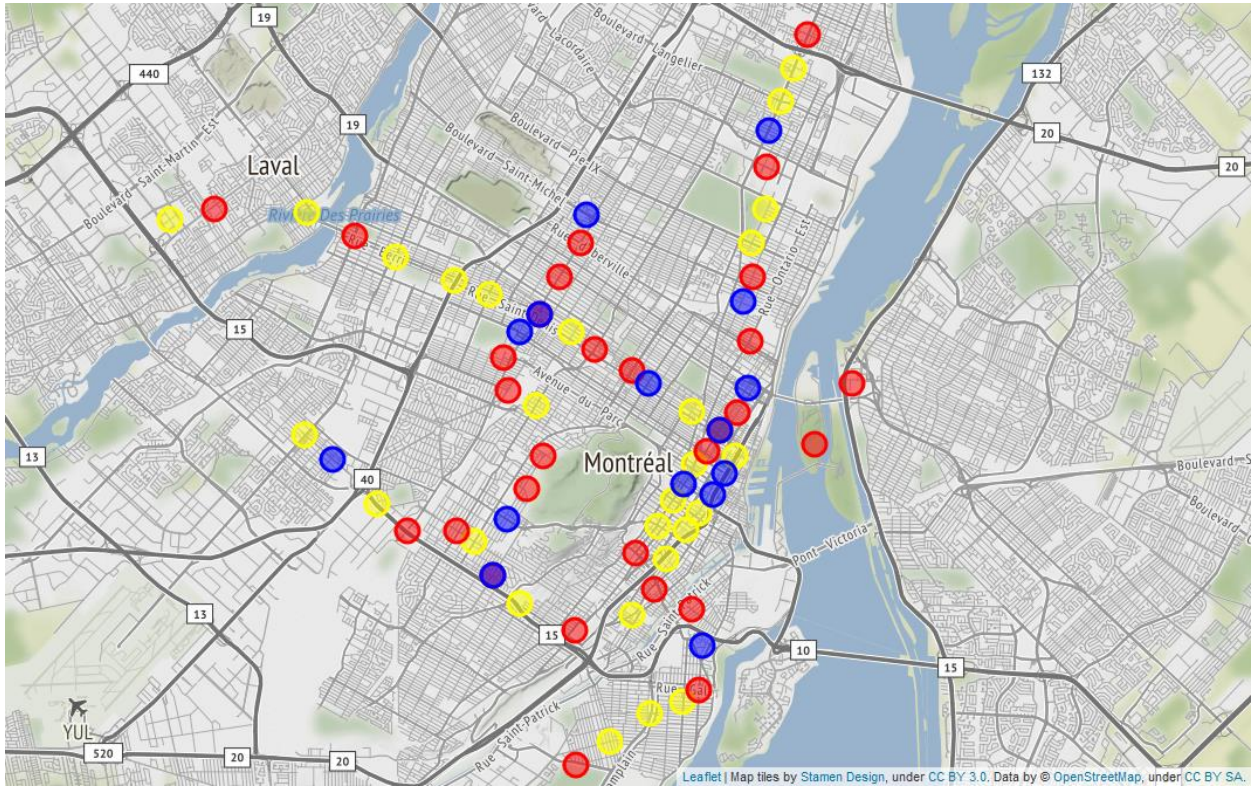
Results:

We have successfully separated the subway stations in 3 clusters:

Cluster 0: contains mostly stations in residential neighborhoods.

Cluster 1: contains mostly outdoors venues.

Cluster 2: contains restaurants and indoor venues.



Conclusion:

We have successfully separated the subway stations, and have hopefully provided tourists with a points of reference as to what they can access through the public transport system. By focusing on the particular clusters, they can make informed decisions as to what kind of venues they are interested in, and travel to those particular stations.