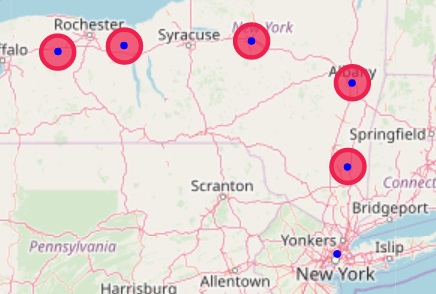
**Battle of Neighborhoods**

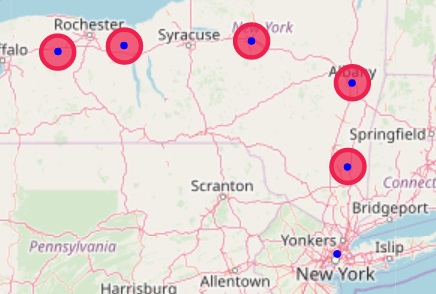
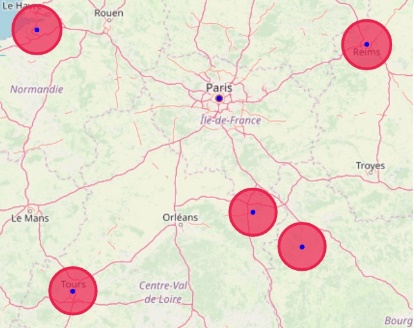
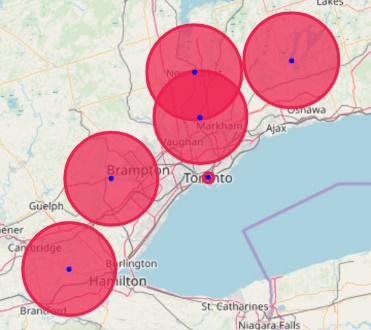
**Introduction**

Whether you are a civil engineer, a politician, or a CEO, it is important to understand the priorities of the citizens / consumers you are working for.

The goal of this project is to see if we can understand some regional similarities based on whether a region is a major urban center, or a rural / suburban area, and based on the country that the region is in.

**Data**

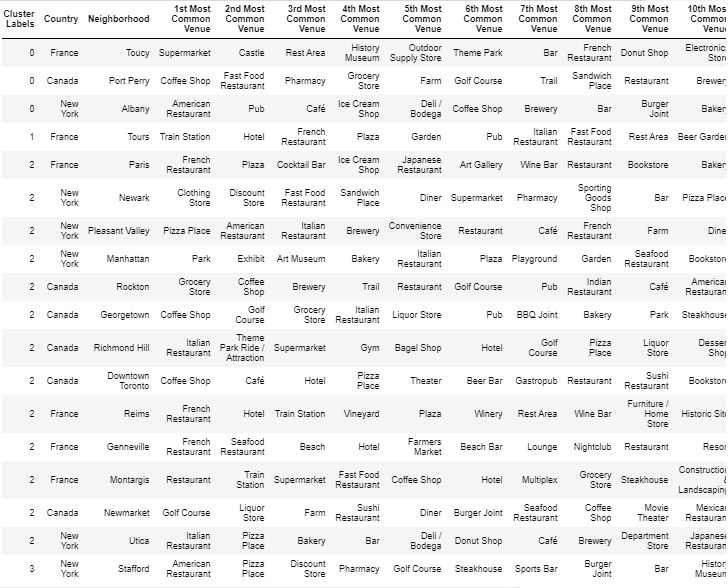
****The data that will be analyzed will be venue data using foursquare API. We will look at 3 different major urban centers (Toronto, Paris, and New York City), and several suburban/rural areas surrounding each city.

****

**Methodology**

By using K-means clustering and clustering each region based on its most common venues, the goal is to observe some correlation between regions in the same country, or between urban centers and rural areas.

**Results**

The table below shows the grouping of each region. The grouping does not appear to correlate with cities or rural areas, or within countries. 

**Discussion**

We are not able to infer much from the K-means model. I suspect this is most likely due to foursquare having too many different venue categories; for example, pub, bar, sports bar and beer bar are all considered different venues in spite of being very similar. The K-means algorithm does not know anything about how similar these venues are, it sees each one as completely unique. Because of this, many venues that might be similar in different regions are viewed as being completely unrelated by the K-means algorithm.

Though less likely, it is also possible that the distribution of venue types is completely unrelated to regional subculture. It could also be that the algorithm just needs more data to develop an accurate model.

**Conclusion**

We are not able to infer anything about regional subculture based on a standard k-means clustering of venue data returned by foursquare api. I believe this is due to foursquare api having too many different categories in its venue database. In order to develop a meaningful cluster model, the foursquare categories would first have to be grouped based on their own similarity, or just narrowed down to some more general categories. This further manipulation is however out of the scope of this project.