Capstone Project - The Battle of Neighborhoods

1. **Introduction**

**Problem background:**

Banhmi, a World Famous Vietnamese Cuisine Phenomenon!

“That’s a symphony in a sandwich” is how Anthony Bourdain described the most famous Vietnamese sandwich called the Bánh Mì.

In 2011, along with Pho, Banh My (bread) was added to Oxford Dictionary as a representative of Vietnamese cuisine.

For people who didnt know about Banhmi, it is is the Vietnamese word for bread.

In Vietnamese cuisine, it also refers to a type of baguette which is often split lengthwise and filled with various savory ingredients as a sandwich and served as a meal. A typical Vietnamese sandwich is a fusion of meats and vegetables from native Vietnamese cuisine such as chả lụa (pork sausage), coriander leaf (cilantro), cucumber, pickled carrots, and pickled daikon combined with condiments from French cuisine such as pâté, along with chili and mayonnaise.

**Problem description:**

Banhmi Phuong is one of the most famous brand in Vietnam. Now, the brand has a plan to expand its business in Toronto. The first job is finding the most suitable place for its.

1. Data

Based on definition of our problem, factors that will influence our decission are:

* number of existing restaurants in the neighborhood (any type of restaurant)
* number of and distance to Italian restaurants in the neighborhood, if any
* distance of neighborhood from city center

I decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

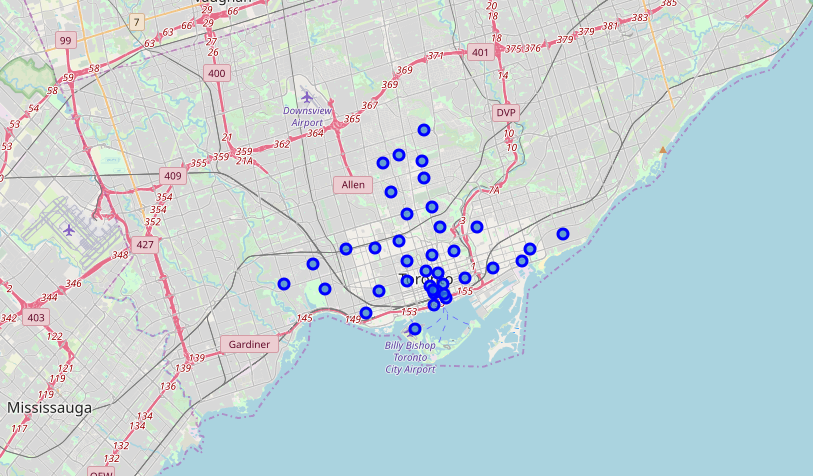
Following data sources will be needed to extract/generate the required information:

* centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **geopy**
* number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**
* coordinate of Toronto center will be obtained using  **geopy**

1. Methodology

In first step we have collected the required data: location and type (category) of every restaurant in Toronto center.

Second step in our analysis will be calculation and exploration, show the revenues in the map.



In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

1. Results

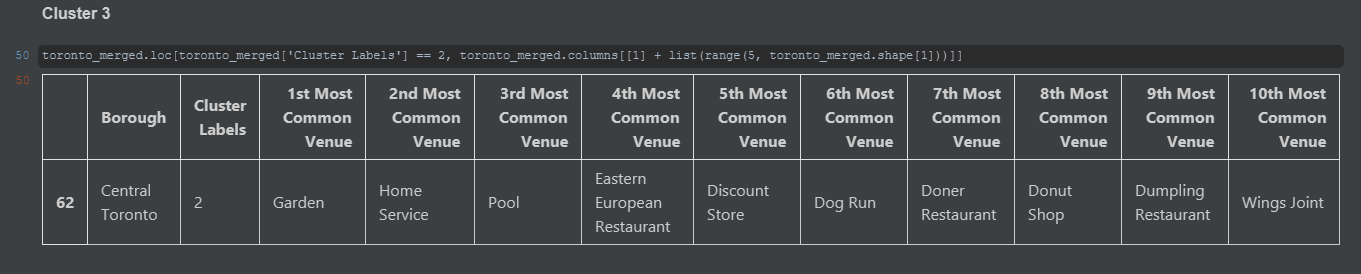
Cluster 1:



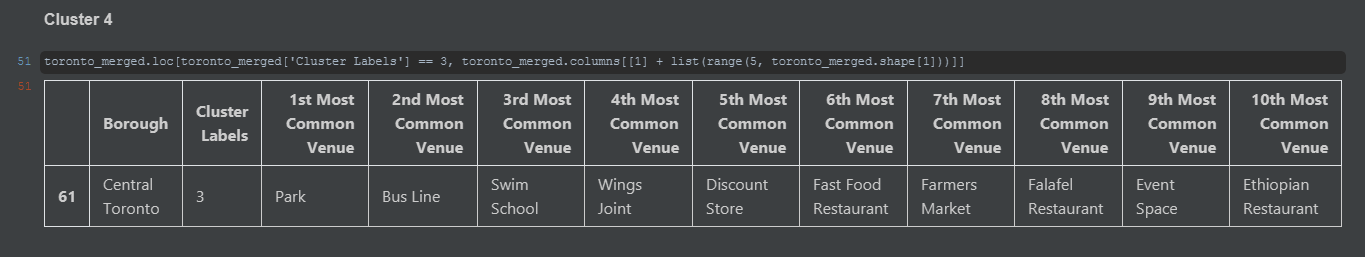
Cluster 2:



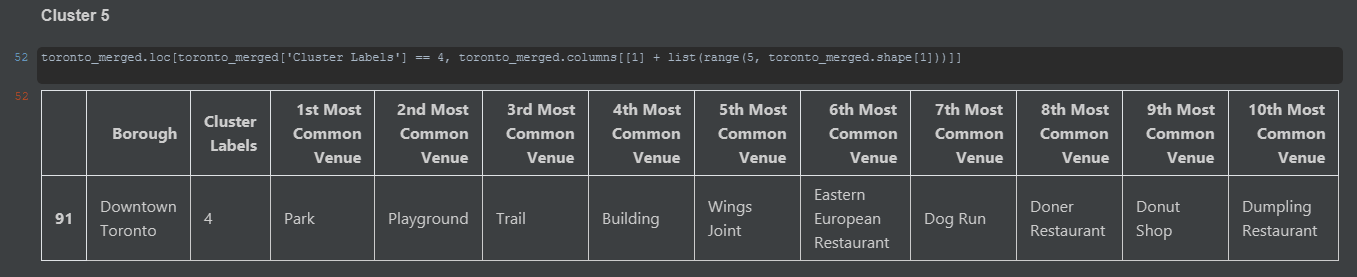
Cluter 3:



Cluster 4:



Cluster 5:



Map cluster:

