Capstone Project - The Battle of Neighborhoods

I. 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.

II. 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:

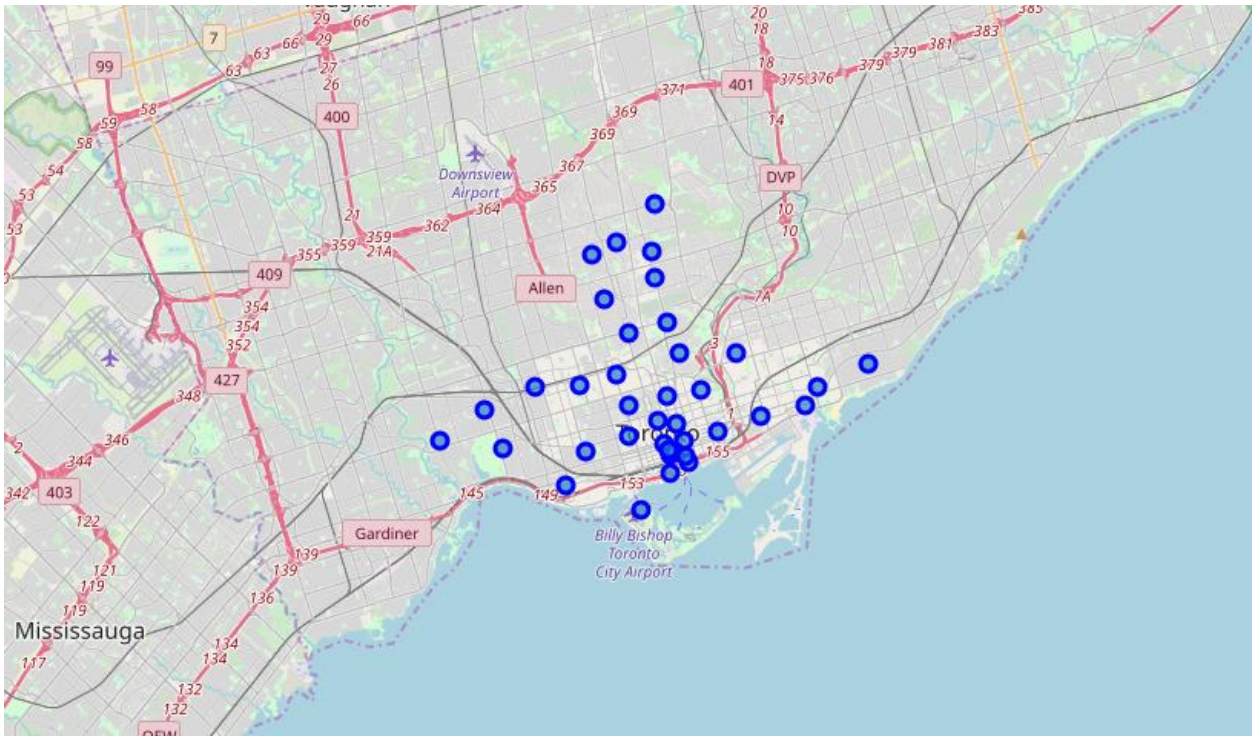
- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **geopy**
- number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**

- coordinate of Toronto center will be obtained using **geopy**

III. 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.

IV. Results

Cluster 1:

Cluster 1

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 0, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10t Co
2	Downtown Toronto	0	Coffee Shop	Bakery	Park	Café	Pub	Mexican Restaurant	Breakfast Spot	Theater	Gym / Fitness Center	Historic
9	Downtown Toronto	0	Coffee Shop	Clothing Store	Middle Eastern Restaurant	Cosmetics Shop	Café	Diner	Pizza Place	Sporting Goods Shop	Bookstore	Italian Restaur
15	Downtown Toronto	0	Coffee Shop	Café	Hotel	Restaurant	Italian Restaurant	Bakery	Cosmetics Shop	Cocktail Bar	Clothing Store	Breakfa
	East		Health				Fast Food	Farmers	Falafel	Event	Ethiopian	Electror

Cluster 2:

Cluster 2

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
83	Central Toronto	1	Park	Tennis Court	Wings Joint	Diner	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store

Cluster 3

Cluter 3:

Cluster 3

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
62	Central Toronto	2	Garden	Home Service	Pool	Eastern European Restaurant	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant	Wings Joint

Cluster 4:

Cluster 4

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 3, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
61	Central Toronto	3	Park	Bus Line	Swim School	Wings Joint	Discount Store	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant

Cluster 5:

Cluster 5

```
52 toronto_merged.loc[toronto_merged['Cluster Labels'] == 4, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

52

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
91	Downtown Toronto	4	Park	Playground	Trail	Building	Wings Joint	Eastern European Restaurant	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant

Map cluster: