## COURSERA CAPSTONE: The Battle of the Neighborhoods

THE USE OF DATA SCIENCE TO DETERMINE OPTIMAL AREAS TO SET UP BUSINESS AROUND TORONTO

Report by: Jaison Amin

## Introduction: Business Problem

In this project, we will use Data Science to determine which location is the best for opening a new business(particularly food business such as restaurants, coffee shop, or bar) in Toronto.

In this Notebook, we will use Data Science to **determine which location is good for opening a new coffee shop**. There are two conditions that determines if the location is suitable for the business.

First, we will check **whether there are another popular coffee shop around the area**. We are looking for a location where **the top 3 places visited frequently are not coffee shop**. This would help us to analyze the competition around the area. Opening a new coffee shop.

Second, we need to ensure that **one of the top 3 location around the area is not a beach**. The area around the beach are hot, and the sales of hot coffee will be affected.

## Data

Based on our problem, there are several factors that influence our decision:

* top three location did not contain another coffee shop
* top three location did not contain beach

We 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 https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M
* number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**

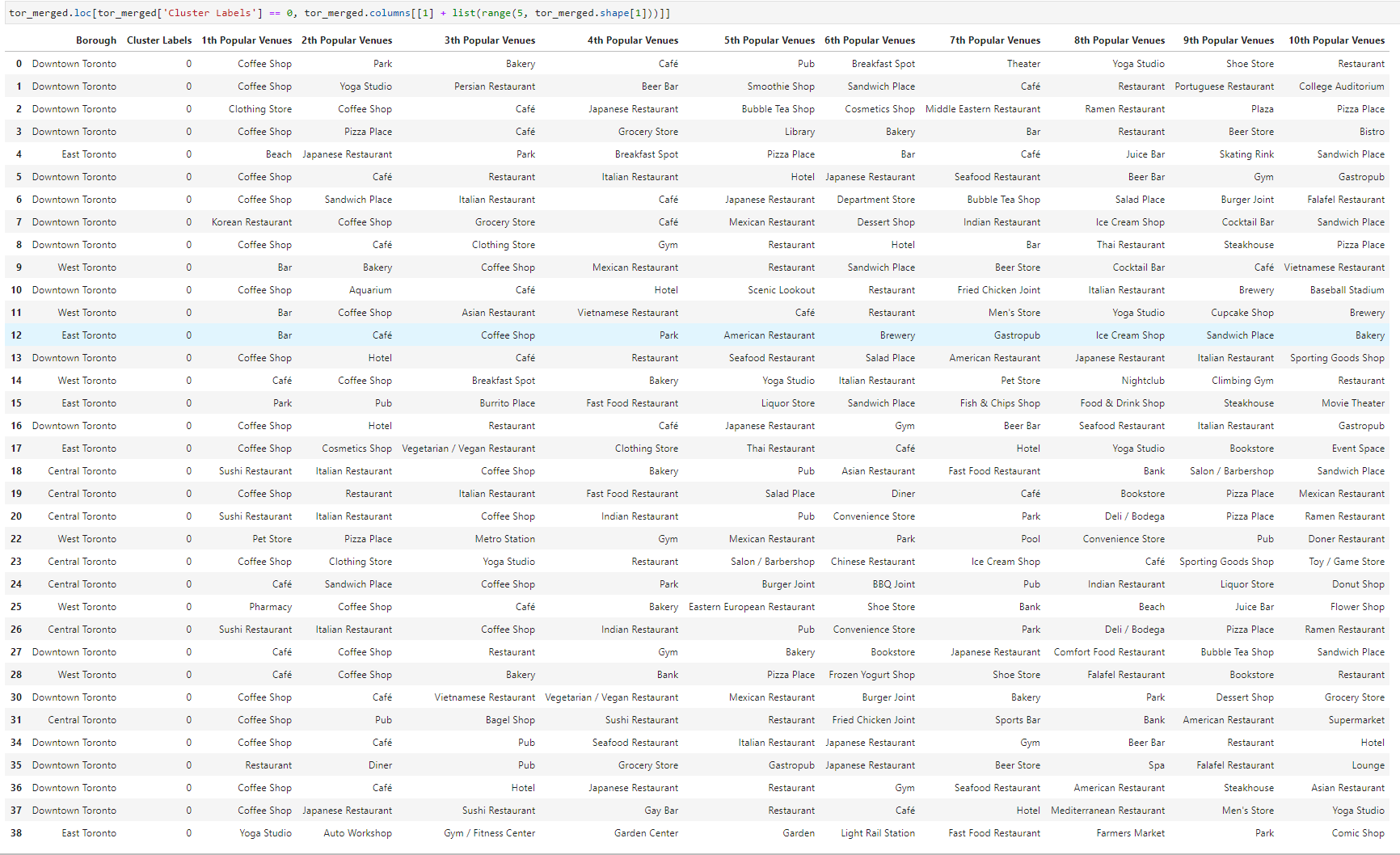
## Methodology

In this project we would use Foursquare API to determine the top 10places frequently visited around the area. We are looking for areas that has no coffee shops or areas with lesser known coffee shops. This would ensure that competition is not one-sided in that area.

The first thing we must do it to divide the location around Toronto into smaller areas, based on their borough. We only use data from Toronto borough (West Toronto, East Toronto, Central Toronto, and Downtown Toronto). After we separate the data, we analyze the top 10 places frequently visited by people around that area. We will then separate some regions into different clusters, and determine whether the area is suitable for opening a coffee shops by looking at the top 3 places frequently visited around that area. If the area has a coffee shop on their top 3, we will deduce the area as unfit for opening a coffee shop. Another rule that we implemented is that no beach on the top 3 frequently visited place. Selling hot beverages around a hotter area would affect sales. 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.

## Analysis

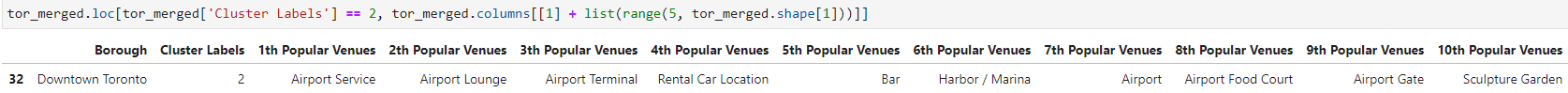
Cluster 0:



Cluster 1:



Cluster 2:



Cluster 3:



Cluster 4:



We conclude that most areas fall into the Cluster 0 category. This is because most areas that were fall into the cluster are dominated by restaurants and coffee shops. We can assume cluster 0 as the entertainment district, and many of them has another coffee shop that became the most frequent place visited around the area. Cluster 1, 2, 3, and 4 were the most suitable since all of them did not have both beach and coffee shop around the area.

## Results and Discussion

After dividing Toronto into different regions and analyze top 10 places frequently visited around Toronto, we conclude that areas on cluster 1, 2, 3, and 4 were the optimum places to set up a business there. Cluster 1 were mostly outdoor areas such as playground and campground. Cluster 2 is reserved for the Toronto Airport. Cluster 3 were dominated by exercise venues such as Tennis Court , Gym, and Park. We assume that Cluster 4 were in the suburban area where most places were designed as residential areas.

## Conclusion

The purpose of this project was to identify Toronto regions which has the better potential for setting up a new coffee shop. Clustering the regions into different cluster helps us to analyze the regions better. Areas in cluster 0 have another famous coffee shops around, hence, it is not advisable to open a new shop there, because competition would make it harder for businessowner to profit.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.