**Introduction**

Globalization has become an increasingly relevant topic in the discussion of economic, political and social events. Barriers to internationalization continue to be breached at an incredible fast pace. From an strictly economic point of view, distances have been virtually reduced through technological advances. Permitting a greater fragmentation of supply chains were relative advantage defines the economic activity of the primary and secondary sectors of production in a given local economy.

But the tertiary sector, comprised of services, has been affected in a different manner. Big companies have realized the benefit of bringing people of different parts of the world, and with different views because of this, together and make increasingly bigger efforts to diversify. Individuals that were previously scared of moving to an unknown country with different values, are not as scared anymore and the values don’t seem so different. To make things simple, the cost-benefit analysis of moving from one location to another has shifted.

This paper, will focus on trying to help those people make a correct decision. This will be done by answering two questions. The first, what neighborhoods from two given cities are more similar or dissimilar. Assuming that the needs of a person were met in the neighborhood that they were previously in, what would be the best fit for him in a new city. The second, what would be the best neighborhood to start a specific type of business in a given city.

For this we will first use the k-means clustering method of classification to identify types of neighborhoods in the cities of Toronto and New York. We will then proceed to analyze the characteristics of this neighborhoods to determine how similar and dissimilar our clusters are. Followed by this we will analyze the specific types of venues in each of our clusters to determine which better fits different business ventures.

**Data**

* **New York**

We use a data set published by the New York City department of city planning from the year 2014. It was obtained from New York’s University (NYU) website and consists of 5 boroughs and 306 neighborhoods, each with its own latitude and longitude coordinates. We then use the foursquare API to obtain the venues at a 500 meter radius and the respective category of each one of the venues.

* **Toronto**

I repeat the same process for obtaining the venues. The difference is that the data regarding boroughs and neighborhoods, with their respective latitude and longitude values, was obtained form scraping the table of the following link: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

**A close up of a map

Description automatically generated**

**Methodology**

k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

We cluster the neighborhoods of the cities of Toronto and New York based on the density of venue categories in a 500meter radius of each neighborhood. **A close up of a map

Description automatically generated**

**Results**

We obtain 5 cluster. The first could be described as neighborhoods with close vicinity to parks, playgrounds and dog parks. It also has vicinity to distribution centers, bodegas and department stores. This would be the ideal neighborhood for a family lifestyle.

The second cluster would be ideal for social city living, comprised of varied venues meant for distinct social activities, such as coffee shops, pubs, restaurants and gyms. For a New Yorker living in the cluster locate din Brooklyn it would be an ideal moving spot.

The third cluster is comprised of neighborhoods that are near event centers, for people that are fans of big sporting events.

We ignore the fourth cluster due to lack of significant results.

The fifth cluster is comprised of locations farther away from the city center and near plenty of parks. The advantage is that it is also near many bus stops and means of transport. It would be ideal for those that enjoy the benefits of cluster number 2 but also want to enjoy the benefits of being farther away from the cities hotspots.

*Cluster 1***A screenshot of a cell phone

Description automatically generated**

*Cluster 2*

**A screenshot of a cell phone

Description automatically generated**

*Cluster 3*

**A screenshot of a cell phone

Description automatically generated**

*Cluster 4***A screenshot of a cell phone

Description automatically generated**

*Cluster 5***A screenshot of a cell phone

Description automatically generatedDiscussion and Conclusion**

We analyzed each of the cluster and concluded what best serves the needs of different profiles. Additionally, I would like to add that cluster number 2 would be best fit for business ventures regarding services of the third sector of the economy. In the future it would be interesting to obtain more data on the people that live in each of this cluster neighborhoods to make a more accurate profile of them and match them with the profile of people looking to move. Then we could compare both method and see which is more accurate at predicting the best neighborhood for someone, their previous neighborhoods characteristics, or the characteristics they share with a person in a given neighborhood.