Toronto, Canada (City of Neighborhoods): Neighborhood Clustering and Segmentation.

1 Introduction

1.1 Business Problem:

The battle of Neighborhoods is a data science project, aimed at grouping similar neighborhoods into the same clusters with the end result being useful information which a user can leverage on to take important business decisions. This project will be targeted on which neighborhoods look promising to set up a new African restaurant.

1.2 Background Description:

This projects focuses on Neighborhoods in Toronto, Canada. Toronto is the financial capital of Canada, so it makes sense that the commercial hub will be where most business will be aiming at due to the large purchasing power in the area. The diversity resulting from the movement of a lot of immigrants from several parts of the world for work and settlement is an added advantage because a variety of businesses can thrive that will meet the needs of different people.
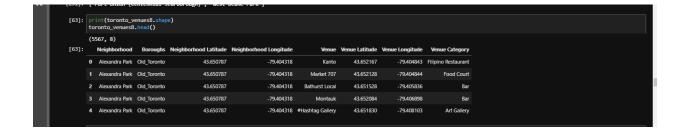
Toronto is the capital city of the Canadian province of Ontario. Toronto is an international center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. With a recorded population of 2,731,571 in 2016, it is the most populous city in Canada and the fourth most populous city in North America[1]. The city of Toronto is made up of 6 Boroughs and official 140 neighborhoods and more than 239 neighborhoods [2] including unofficial neighborhoods. Some of the neighborhoods are further subdivided into sub-boroughs such as the borough called "old" Toronto which is further sub dived into four boroughs. The other boroughs are East York, Etobicoke, North York, Scaborough and York.

Toronto being the most populous city in Canada has majority of the population in the area called "old" Toronto. When a potential African restaurant opener considers this, one might be tempted to say the best area will be "old" Toronto but a business is not solely dependent on a large population other factors such as existing competition, preference of the prospective customers are important dependent which this project will aim to tackle.

2 Data Description

The data used for this project comprised of the Neighborhoods name, position data (Latitude and Longitude) and the most popular venues in each neighborhood. The dataset was not readily available and so I had to:

- Use the Beautiful Soup library to scrap the names of the Neighborhood and their respective boroughs from the Wikipedia page. While the corresponding latitude and longitude data for each neighborhood was obtained using the Nominatim module from the the geopy library.

- The most popular venues for each neighborhood was obtained using the foursquare API

- The datasets were then eventually merged.

The data was grouped based on the basis of Neighborhood and the mean of the frequency of occurrence of certain venues were taken, this data was then fed into the Kmeans algorithm which then clustered similar Neighborhoods together.