Applied Data Science Capstone
Gary Thomas

Peer-graded Assignment: Capstone Project - The Battle of Neighborhoods (Week 1): June 2019

Now that you have been equipped with the skills and the tools to use location data to explore a geographical location, over the course of two weeks, you will have the opportunity to be as creative as you want and come up with an idea to leverage the Foursquare location data to explore or compare neighborhoods or cities of your choice or to come up with a problem that you can use the Foursquare location data to solve.

A description of the problem and a discussion of the background. (**15 marks**)

1) Introduction-Problem Scope

Clearly define a problem or an idea of your choice, where you would need to leverage the Foursquare location data to solve or execute. Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem.

There is a groceries contractor in one of the boroughs of Toronto (Scarborough). This contractor provides places such as: Different types of Restaurants, Bakery, Breakfast Spot, Brewery and Café with fresh and high-quality groceries. The contractor wants to build a warehouse for the groceries it buys from villagers and farmers inside the borough, so that they will support more customers and bring better "Quality of Service" to the old customers.

For example, if the warehouse is close to those old and famous restaurants, then the vegetables and other groceries would be delivered to the restaurant in the right time and there would be no delay so the restaurant cooks can start their job from the morning and the Quality of Service will be high and this contractor will gain more reputation and income.

The contractor should build this warehouse where it is closest to its customers in order to minimize the cost of transportation in addition to the example above. which neighborhood (in that borough) would be a better choice for the contractor to build the warehouse in that neighborhood. Finding the right neighborhood is our mission and our recommender system will provide this contractor with a sorted list of neighborhoods in which the first element of the list will be the best suggested neighborhood. The idea of this study is to help people planning to open a new restaurant in Toronto to choose the right location by providing data about the income and population of each neighborhood as well as the competitors already present on the same regions.

2. Data Preparation and Challenges

Describe the data that you will be using to solve the problem or execute your idea. Remember that you will need to use the Foursquare location data to solve the problem or execute your idea. You can absolutely use other datasets in combination with the Foursquare location data. So make sure that you provide adequate explanation and discussion, with examples, of the data that you will be using, even if it is only Foursquare location data.

To provide the stakeholders the necessary information I'll be combining Toronto's 2016 Census that contains Population, Average income per Neighborhood with Toronto's Neighborhoods shapefile and Foursquare API to collect competitors on the same neighborhoods.
Toronto's Census data is publicly available at this website:

We will need geo-locational information about that specific borough and the neighborhoods in that borough. We specifically and technically mean the latitude and longitude numbers of that borough. We assume that it is "Scarborough" in Toronto. This is easily provided for us by the contractor, because the contractor has already made up his mind about the borough. The Postal Codes that fall into that borough (Scarborough) would also be sufficient for us. I fact we will first find neighborhoods inside Scarborough by their corresponding Postal Codes.

We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information, we will use "Foursquare" locational information. By locational information for each venue we mean basic and advanced information about that venue. For example, there is a venue in one of the neighborhoods.

As basic information, we can obtain its precise latitude and longitude and its distance from the center of the neighborhood. But we are looking for advanced information such as the category of that venue and whether this venue is a popular one in its category or maybe the average price of the services of this venue.

Moreover, to provide the stakeholders the necessary information I'll be combining Toronto's 2016 Census that contains Population, Average income per Neighborhood with Toronto's Neighborhoods shapefile and Foursquare API to collect competitors on the same neighborhoods.