

Finding the optimal location to open a fish market using data science

Background

For this project, I am creating a hypothetical scenario where I am tasked with finding the best location to open a new fish market in the Toronto area. With the global rise seafood consumption rates increasing, it is more prevalent to be selling seafood today. With this fact in mind, it looks like a promising business opportunity. The key is finding a suitable place to open the business. To find the best location, we would want to find a location that is near a seafood restaurant.

Problem

The objective is to find a good location to open a fish market in the Toronto area. We will use data science methods such as clustering to find the best location available.

Data

The data I will use consists of:

- List of neighborhoods in the Toronto area from Wikipedia scrapped using the Soup module in Python
- Latitude and Longitude of the neighborhoods provided by IBM
- Venue data from Foursquare API to help locate seafood restaurants

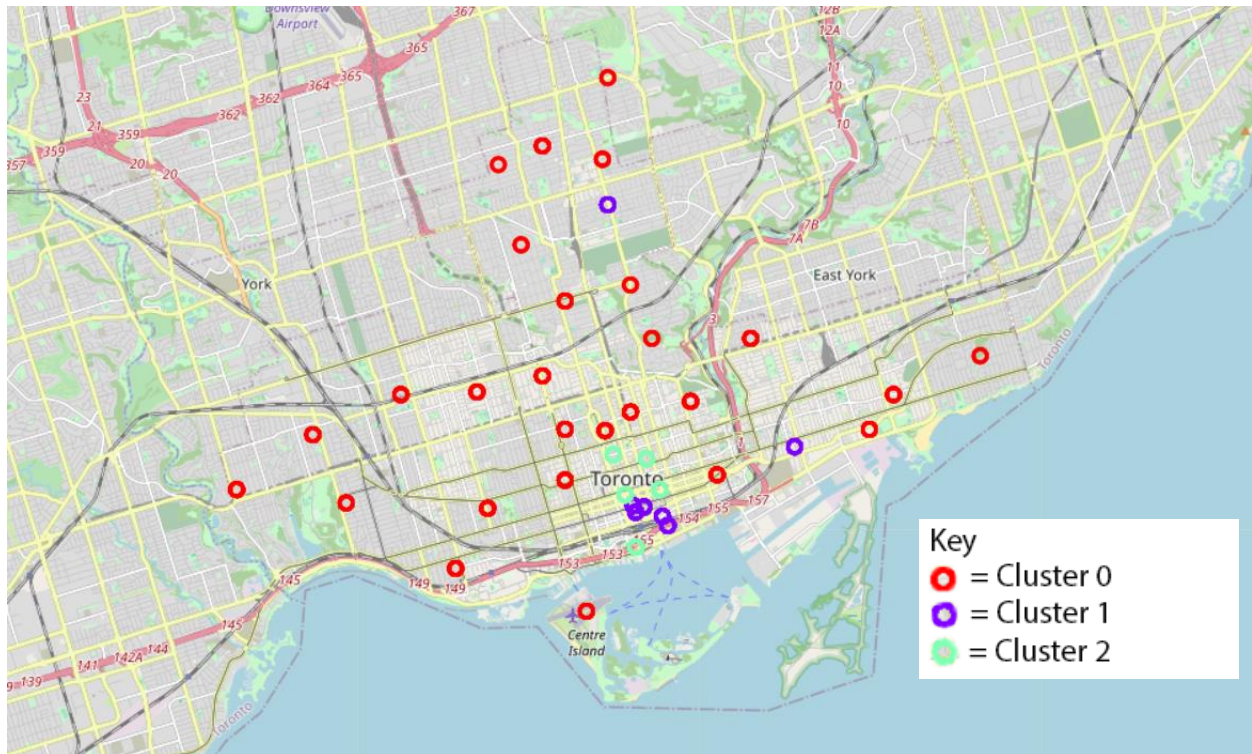
Methodology

To capture the landscape of Toronto and the many restaurants it has to offer, we must first locate data on the neighborhoods in Toronto. To do this we need a large database that has an accurate listing of neighborhoods. Wikipedia's page https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M has the information that we need. To extract the table into a usable data frame, we first had to use Beautiful Soup library that helps the user pull data out of HTML files. After cleaning the data, I then added the coordinates that the neighborhoods belonged to. IBM was kind enough to provide this in a file called "Geospatial_Coordinates.csv". I had to change some of the labels to match the original data frame that we just extracted, but after the change I merged this coordinate data frame to the original data frame with the neighborhood information using the postal code as the matching marker.

With a complete data frame that contained all the locational data, next was to get information on the venues that were in the Toronto area. To do this, I used Foursquare. Foursquare's application programming interface (API) allows the developer to find venues that you search for and are related to your search. This is useful for finding a specific type of restaurant. To utilize the API, I had to get a developer key that allowed me to make a certain amount of calls each day. After getting the key, I used

a function that allowed me to find a list of 100 venues in a 500-meter radius. The function also pulls longitude and latitude of these venues. By invoking the function, we get a very large data frame that contains the longitude and latitude of all the venues in Toronto in the Foursquare database. We then make a new data frame that is specified for looking for seafood restaurants. I then merge this data frame with the original data frame that contains the location data. To provide analysis of this data we can use clustering to group areas of Toronto, into more understandable information. I now use Folium to create a map to visualize the clusters. Along with this map, I also create data frames of the three individual clusters for deeper analysis.

Results



Map made with Folium showing the three clusters of data over various Toronto neighborhoods.

Cluster information:

- Cluster 0: Contains no seafood restaurants
- Cluster 1: Contains many seafood restaurants
- Cluster 2: Contains some seafood restaurants

Discussion

The map created gives a good visual representation of the seafood restaurants in the Toronto area. By looking closely at the map, we can see that most of the seafood restaurants are in the central downtown Toronto area. There are also some locations outside the downtown area that have many seafood restaurants. The point to the north in cluster 1 is Davisville Village, and the point to the east is Studio District. The rest of Toronto marked in cluster 0 with red points are places that have no seafood restaurants.

Conclusion

After analyzing the data frames and the map, I think that best place to construct a new fish market would have to be in the central downtown area of Toronto. I think that because this area contains the highest density of seafood restaurants. This will create the greatest demand allowing you to profit from the investment. This location also has another strong advantage that is lying near Lake Ontario which can act as way to obtain fish. I think you could make an argument for establishing a fish market in Davisville Village or Studio District if the rent was too high in downtown Toronto or the downtown Toronto is overly saturated with fish markets. If these are not costly factors that would outweigh profits, I would recommend establishing a fish market in downtown Toronto.

Limitations of Research

In my analysis I provide a conclusion of establishment of new business solely looking at the location of seafood restaurants. While this is a good start to understanding the problem, it does not fully analyze the multifaceted nature of establishing and new business. There are other factors that should be further analyzed if my conclusion is to be taken further. Such things are fixed costs such as rent or building costs, payroll, depreciation, and potential licenses. Variable costs could include delivery charges, sales commissions and marketing. Another way to analyze the market would be a survey of demand for different types of fish or lack of supply of fish in the Toronto area. By analyzing all these factors and my conclusion, you should have a better understanding viability of establishing a new fish market.