Richard Inniss
Applied Data Science Capstone

Animal Boarding

Exploring best location for Animal Boarding Service in Toronto using Data Science

Business Problem

The business problem we face is trying to ascertain for our stakeholders what would be the best location for an Animal Boarding Service in Toronto in order to maximize the number of clients.

I will attempt to leverage several of the tools discovered during the IBM Data Science Professional Certificate program, first by creating neighborhoods around Toronto by scraping some publically available web data on postal codes. Then, I will leverage services via API's from Foursquares to explore and identify pet stores, pet services, pet cafes, veterinarians and animal shelters contained in each of these neighborhoods.

Finally, using geopy, folium and K-mean clustering, I will locate all the venues on a map and find the ideal location to maximize the size of the clientele. This method will be based on the assumption that the above mentioned Pet Services are located in areas of high demand.

Data acquisition and cleaning

For this project, the data sources we decided to use are as follows:

1- Publically available list of postal code data for Toronto from the following Wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.

2- Publically available csv file found at http://cocl.us/Geospatial_data containing the geographical location data for each of the postal codes.

3- Data from Foursquares using an API call to explore specific venue types with the categoryId function. Specifically, the categoryId's I will be searching for are:

Category Name	Category ID
Pet Stores	4bf58dd8d48988d100951735
Pet Services	5032897c91d4c4b30a586d69
Pet Cafes	56aa371be4b08b9a8d573508
Veterinarians	4d954af4a243a5684765b473
Animal Shelters	4e52d2d203646f7c19daa8ae

Data acquisition and cleaning

- As The postal code data contains missing information for some of the boroughs as well as some of the neighborhoods, we needed to clean this by removing those postal codes having missing data.

- We also needed to clean some of the data acquired from Foursquares as some of the venues were duplicates based on the radius and proximity of certain neighborhoods.

- The original number of venues was 839, once the duplicates were eliminated the number of unique venues was 263, therefore there were quite a few duplicated in our original dataframe

Methodology

- First, we imported all required libraries.
- Second, we scraped the Toronto region postal code, borough and Neighborhood data from wiki page URL and created a dataframe
- Third, we merged the neighborhood dataframe with coordinates from the csv file

This is the resulting dataframe (first 5 rows):

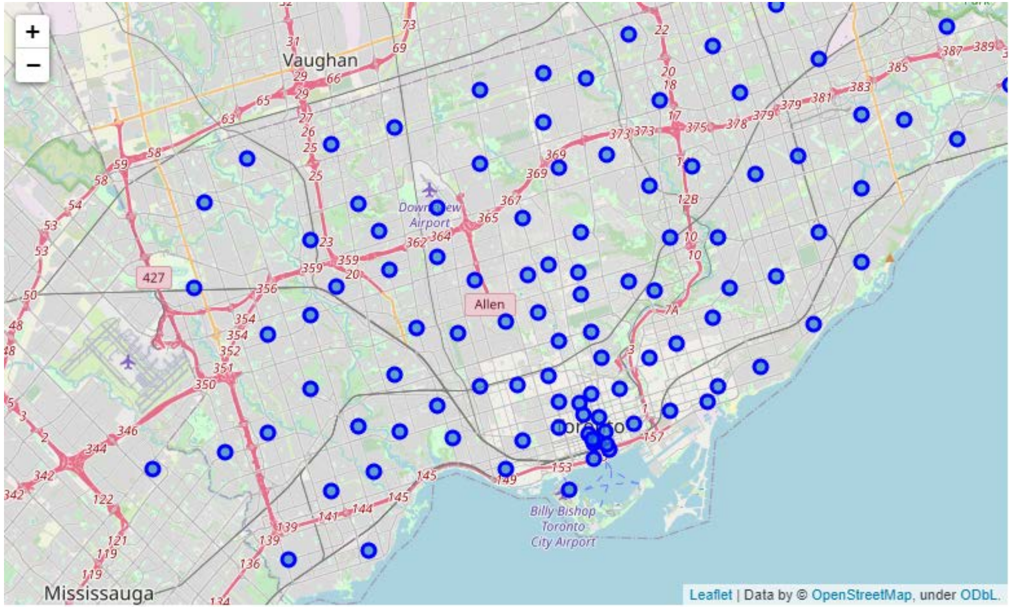
	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

Methodology

- Fourth, we explored the neighborhoods in Toronto and generated map.

We start by using geopy in order to obtain the lat/long coordinates for the city of Toronto. Then, using folium we generated a map of Toronto and add a marker for each one of the neighborhoods in our dataframe.

The resulting map can be seen bellow here:



Methodology

- Fifth, we extracted specific categoryId's from Foursquare for each of our neighborhoods.

  - ✓ As we will base on decisions going forward on the location of the venues, we add those coordinates and drop the neighborhood coordinates in our dataframe.
  - ✓ As some of the venues might be duplicates based on our radius of 2000 meters and the proximity of certain neighborhoods to each other, we remove all venue duplicates.
  - ✓ We then generated a map of Toronto using folium, this time adding markers for all the unique venues we had discovered (see map on right)

Methodology

- Sixed, we use math functions to find the optimal location of our Animal Boarding Service by locating the most central point "Centroid" based on all of the 263 venue latitudes and longitudes obtained from Foursquares.
- The following code (see code on right) was therefore used to calculate the central latitude and central longitude.
- The resulting coordinates after running each one of the venue locations in our code is:

central_Latitude: 43.70042977508153

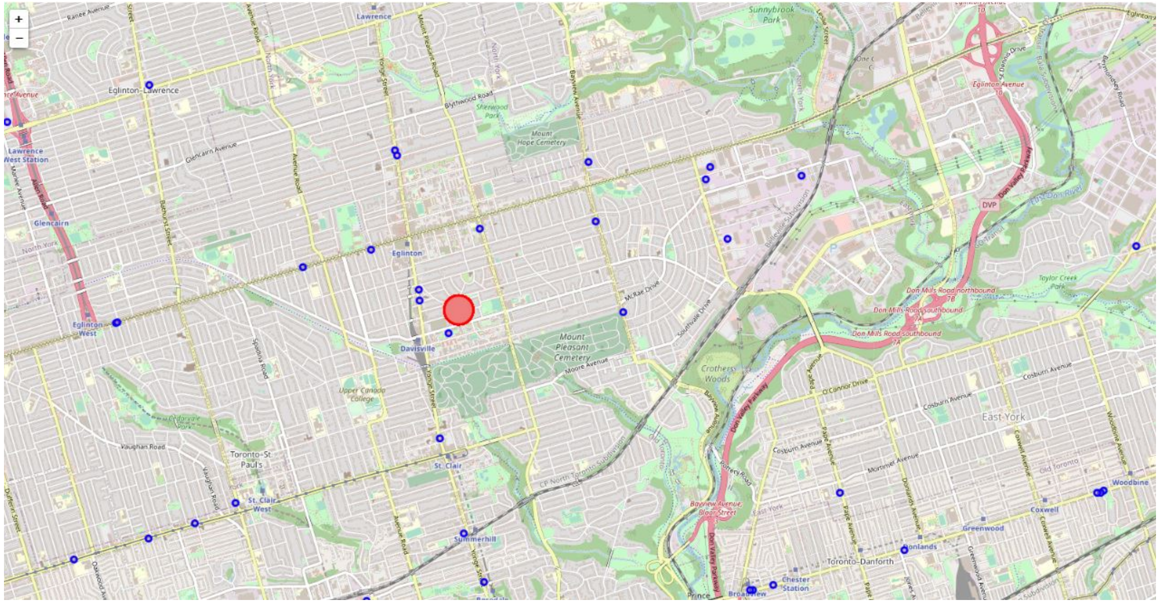
central_Longitude: -79.39244122265683

```
x = 0.0
y = 0.0
z = 0.0
for i, coord in toronto_venues.iterrows():
    Latitude = math.radians(coord.Vlatitude)
    Longitude = math.radians(coord.Vlongitude)
    x += math.cos(Latitude) * math.cos(Longitude)
    y += math.cos(Latitude) * math.sin(Longitude)
    z += math.sin(Latitude)
total = len(toronto_venues)
x = x / total
y = y / total
z = z / total
central_Longitude = math.atan2(y, x)
central_square_root = math.sqrt(x * x + y * y)
central_Latitude = math.atan2(z, central_square_root)
mean_location = {
    'Latitude': math.degrees(central_Latitude),
    'Longitude': math.degrees(central_Longitude)
    }
central_Latitude = math.degrees(central_Latitude)
central_Longitude = math.degrees(central_Longitude)
```

Results

- We can now generate the map of Toronto with markers for all the unique venues once again, but this time we add a large red marker to indicate the ideal location for our Animal Boarding Service and zoom in a little closer so we can see street names.

- Based on our previously calculated central cordinates, the ideal location for our Animal Boarding Service would be just south of Eglington near Yonge. This location has the potential of maximizing the size of the clientele as it is the most centrally located based on all of the related animal care services.



Conclusion

- The purpose of this project was to identify the best location of our Animal Boarding Service in order to in order to aid stakeholders identify the optimal location in Toronto.

- Using DataScience, we were able to identify the best location as being near Yonge and Eglington. We achieved this by leveraging data via an API with Foursquares to locate as many Pet relates services as possible in order to help us zero in on the optimal location for pet owner traffic.

- We did some additional exploration to see what other insights we can uncover using K-Means clustering. The insight we gathered for each of these clusters could be used for our stakeholders marketing efforts in fine tuning advertising of the Animal Boarding Service within each cluster.

- The final decision on optimal Animal Boarding Service location can be made by stakeholders based on either the optimal central location as previously mentioned, or by specific characteristics of neighborhoods and locations in each of the clusters, taking into consideration additional factors like attractiveness of each location (proximity to Veterinarians for example), Pet Services, etc.