Nicholas Mcallister

Coursera capstone project

Where to open up a bar in Toronto, Canada

Introduction

For this Ibm Capstone project I am imagining the scenario in which a group of business partners are interested in opening a bar in the Toronto area. The idea is that due to the increasing popularity of Toronto it has now become a major city for not only business but also as a destination spot for young people as both residents and travelers. This scenario presents a great opportunity for entrepreneurs to capitalize on the cities increasing popularity and open up a bar in Toronto's bustling nightlife scene. The idea is to open up a bar in a section of the city that is already dense with nightlife. This may sound counter intuitive as this would lead to more competition however when people go out they generally flock to parts of a city that are rich in nightlife so picking a spot in the thick of everything would increase exposure. With this in mind location location to open said bar is the most important factor to consider for these business partners. Therefore I am designing this project to help them find the most ideal location.

Business Problem

The goal of this project is to find the optimal location for business partners to open up a new bar in the Toronto area. By leveraging foursquare data and using the machine learning algorithm clustering this project aims to find the solution to this problem. Through the power of data science I plan to find the solution to the business question: Where is the most optimal location to open up a bar in Toronto, Canada?

Target Audience

Said business owners looking to open up a bar in Toronto, Canada

The Data

The following data will be used for this project

-List of all neighborhoods in Toronto

-Latitude and Longitude of these neighborhoods

-The Venue data of bars/nightlife in Toronto. This will allow us to find the high density areas for nightlife in Toronto

Extraction techniques

-Extracting the data for Toronto neighborhoods from wikipedia

-Loading a file containing the Latitude and longitude for the neighborhoods

-Making use of a Foursquare developer account levering the API to retrieve the venue data

related to these neighborhoods.

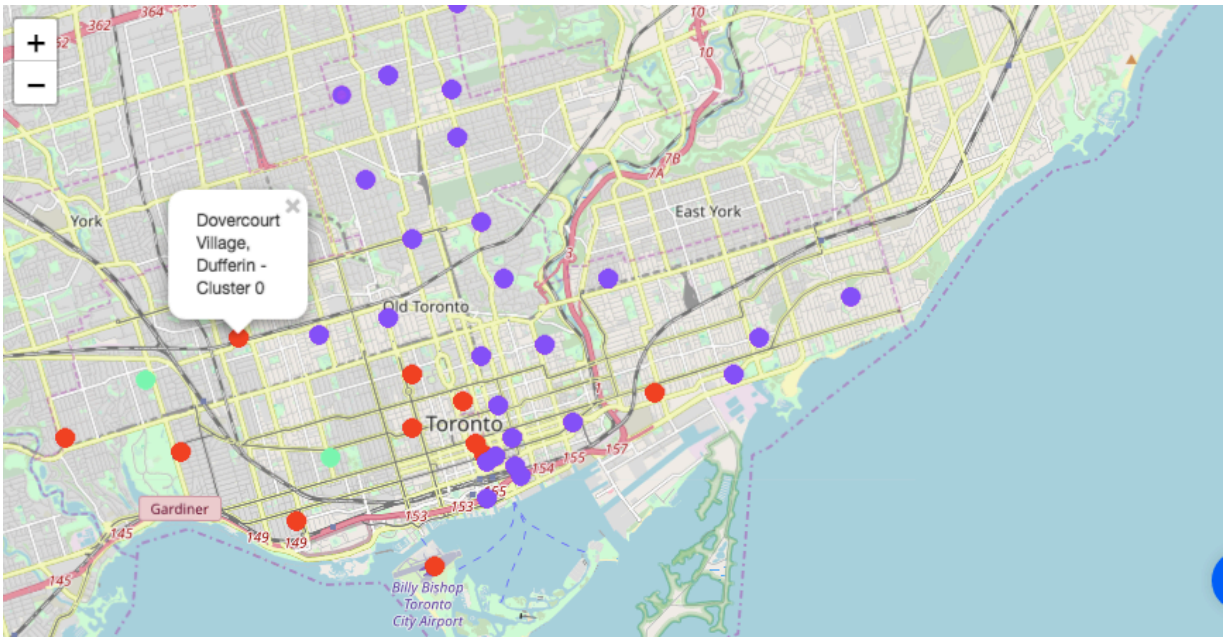
Methodology

To start things off creating a list of neighborhoods in Toronto, Canada was required. In order to do this I used the Beautiful soup package to extract the list of Canadian neighborhoods from the wikipedia page(“ `https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M”).`

`However the data frame created did not contain Longitude and Latitude data which is paramount for connecting with the`

Foursquare API. Therefore I had to create a new data frame by loading a csv file containing the geospatial data of these Toronto neighborhoods. From there I formatted both data frames in order to join them properly. This created a new data frame with all the information I needed to pull venue data via the Foursquare API. From there I created a map of Toronto using the folium package in order to verify that the location data was correct. At this point I connected to the foursquare Api using the necessary id information. Following that I pulled a list containing the top venues within a 500 meter radius, grouped the venues by neighborhood in to make the clustering process easier later on, and figured out the mean occurrence of each category in each neighborhood. My next step was filtering through the venues to get only the venues that were categorized as bars. To finish off my analysis I performed K means clustering on extracted data. I chose to set the clusters equal to 3 in order to create 3 separate centroids to allocate to the data to nearest centroids and create 3 separate clusters based off of the occurrence of bars in the various neighborhoods. Based on the concentration of the clusters I will confidently be able to suggest the ideal location to open the bar.

Results



K means clustering separated our data into 3 clusters based off of amount of bars in each area

-cluster 1- moderate amount of bars in the area

-cluster 2- no bars in the area

-cluster 3- many bars in the area

The results are shown above. Cluster 1 in red, cluster 2 in purple, and cluster 3 in green

Recommendations

As described all of the bars are contained in either cluster 1 or cluster 3. Knowing this we can immediately eliminate the idea of opening a bar in any areas contained in cluster 3. This eliminates east york and the northern part of Toronto completely as there would not be enough traffic going through these areas

during the later hours of the night when people generally enjoy the nightlife scene. Common logic would tell us to open up the bar in either cluster sectors located in Little Portugal or High Park. These two areas have the highest density of bars in all of Toronto and therefore probably get their share of nightlife enthusiasts coming in. However as you can see in the center of Downtown Toronto from the University of Toronto to the union station in the south we have 5 neighborhoods that belong to cluster 1. While these clusters may not have as many bars as those in cluster they are very close to each other in terms of distance. This could suggest that there are many people roaming from neighborhood to neighborhood and this makes sense because this area is in the heart of downtown. Because of this I would suggest that the bar be opened up somewhere in this area. Not only would you be near the university (college kids generally like to go out and have a good time) but you would also be near the union station therefore a lot of people would be coming and going through the area. Which would inevitably increase exposure to the bar.

Conclusion

There you have it. In this project we successfully identified the business problem, specified the data requirements, gathered/

extracted the data, and analyzed the data using k means in order to effectively provide the stake holder with a well informed decision on where to open up a bar in Toronto, Canada.

References

List of neighborhoods

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Foursquare API

https://developer.foursquare.com/