Market Expansion using Data Science

Pratham Sharma

July 21, 2021

Introduction

Business Use-case

UP! Sports is a (hypothetical) retail company selling sporting goods, equipment, sports apparel, and much more, for various sporting activities, all under one roof.

Business is booming and the company has decided to expand its operations. For the same, the company is looking to setup stores in Canada, starting with the city of Toronto, Ontario. However, the company wants to make this decision using a strategic and methodical approach.

Acknowledging the power of data and analytics, UP! Sports has decided to use a data science based approach to support their decision-making process for business expansion.

Task Description

The task is to identify and group neighborhoods in the city of Toronto, Ontario that the company must target and setup their operations and stores in. Since the company specializes in sports-related consumer goods, the ideal neighborhood or groups of neighborhoods for the company would be neighborhoods with many sport-related facilities, such as gyms, yoga classes, soccer fields, and so on.

Thus, the task is to identify such neighborhoods and present an analysis, outlining which neighborhood(s) should UP! Sports setup their retail stores in, as well as the rationale behind the same.

Data

Data Description

The data that will be used for this business use-case is location data for neighborhoods in Toronto, Ontario, Canada. This data will be collected through Wikipedia, along with the Foursquare API.

The Wikipedia page "List of postal codes of Canada: M" gives details and list of all postal codes in Toronto, Ontario. Therefore, data for all neighborhoods in Toronto, including their names, respective boroughs and postal codes will be collected from Wikipedia by web scraping.

Further, data about all different venues in each neighborhood will be collected using the Foursquare Places API, using latitude and longitude values, which will be collected using geospatial data.

The final data will include the following:

1. Neighborhood name
2. Borough name
3. Postal Code
4. Latitude and longitude values
5. Venues in each neighborhood

This data will be analyzed to provide a solution to the discussed business problem.

Data Collection

The process of data collection begins with scraping the Wikipedia page “List of postal codes of Canada: M”, which includes information about different neighborhoods, their respective boroughs and postal code in Toronto, Ontario.

This page contains information about 103 neighborhoods of Toronto, which is scraped and collected using BeuatifulSoup. A sample of the initial data that will be collected is shown below.

Postal Code	**Borough**	**Neighborhood**
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M9A	Etobicoke	Islington Avenue
M1B	Scarborough	Malvern, Rouge

Further, data about coordinates of each neighborhood is obtained using the file "geospatial_coordinates.csv", which contains latitude and longitude values for different postal codes. The data collected from Wikipedia is joined with the geospatial data available to get coordinates (latitude and longitude) for each neighborhood. A sample of the data collected so far is shown below.

Postal Code	Borough	Neighborhood	Latitude	Longitude
M3A	North York	Parkwoods	43.75325	-79.32965
M4A	North York	Victoria Village	43.72588	-79.31557
M9A	Etobicoke	Islington Avenue	43.66785	-79.53224
M1B	Scarborough	Malvern, Rouge	43.80668	-79.19435

Next, data about venues in and around Toronto is collected using the Foursquare Places API. The Foursquare API returns data about 2000 different venues across 263 categories of venues.

The following table gives a description of the final data collected through Wikipedia and Foursquare API. Names of columns and their description is given.

Column Name	Description
Neighborhood	Name of the neighborhood in which venue is located
Neighborhood Latitude	Latitude of the neighborhood
Neighborhood Longitude	Longitude of the neighborhood
Venue	Name of venue
Venue Latitude	Latitude of venue
Venue Longitude	Longitude of venue
Venue Category	Category to which venue belongs

This is the final data that will be used for the analysis.

Methodology

To solve the problem at hand, neighborhoods and boroughs that have a high density of sports related venues, such as gyms, stadiums, etc., have to be identified. To do this, the required data has already been collected, through web scraping and through the Foursquare API.

The first step of the analysis involves filtering the collected data to include only those venues that are related to sporting activities.

After filtering out such venues, some basic exploratory analysis is done to get more insight on the available data. This involves plotting some charts and maps to visualize the data. Through these visualizations, we will be able to identify the boroughs that have a high density of sports related venues.

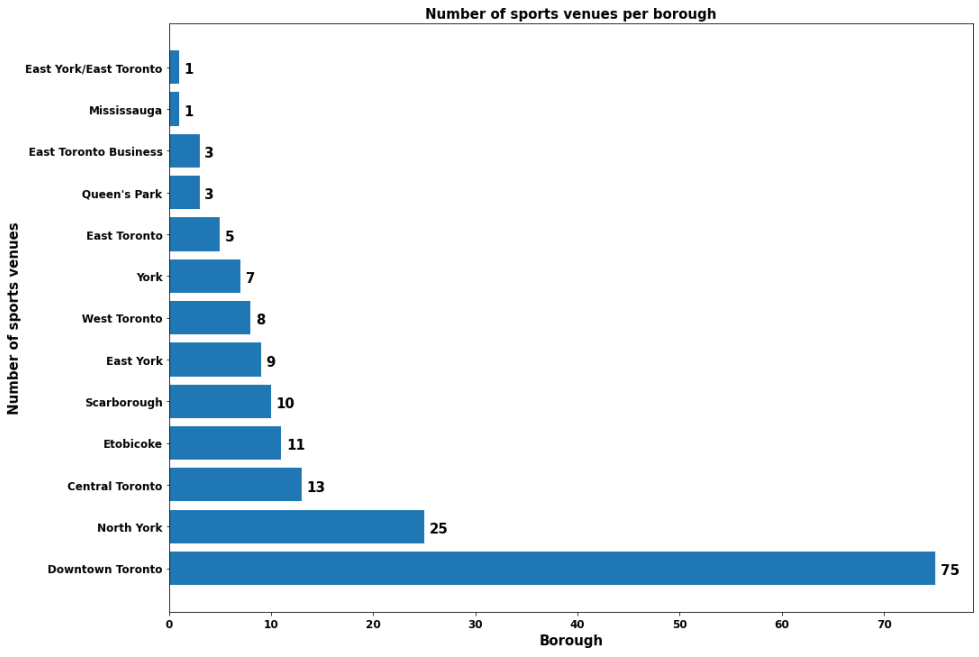
Finally, the selected venues are clustered to identify cluster centers, which will become the target locations for the company. For this, k-means clustering algorithm is used.

Analysis

Exploratory Data Analysis

After filtering the gathered data to retain only sport related venues, some basic exploratory data analysis is done.

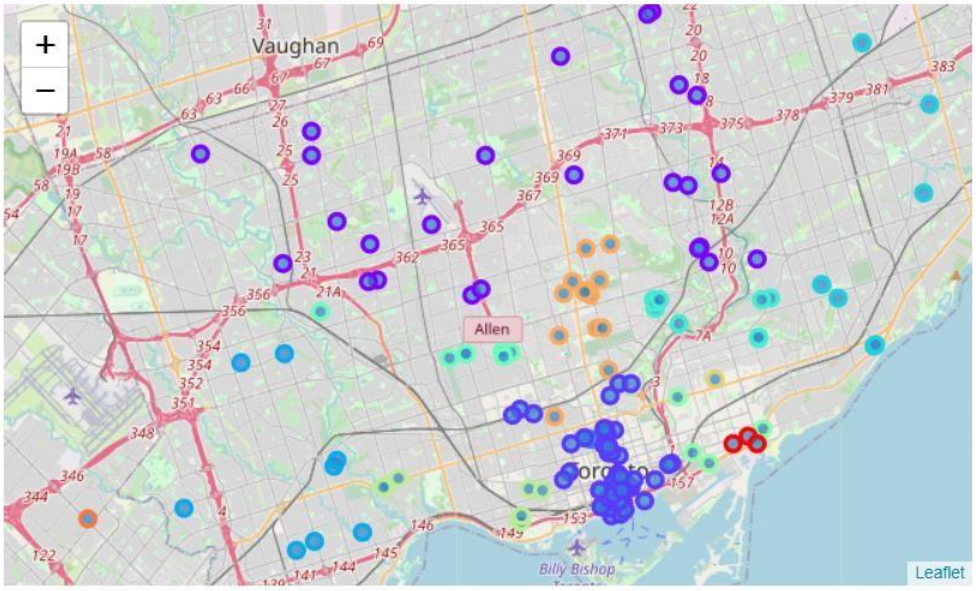
There are a total of 171 sports related venues in and around Toronto, located in different boroughs. The following bar chart shows the number of sports related venues in each borough.



By far, Downtown Toronto has the greatest number of sports related venues (78). North York ranks 2nd in terms of the number of sports venues in the borough. However, these venues are scattered around North York i.e., the density of sports related venues in North York is quite low.

Targeting neighborhoods in Downtown Toronto can be a great starting point for the company to set up their first store in Toronto, Ontario.

Further, let's take a look at a map of Toronto showing the different sports venues, based on the data collected so far.



As expected, a lot of the sports venues are located close to each other in the borough of Downtown Toronto (as shown on the map in dark blue color).

A considerably high density of sports related venues is also observed in the surrounding boroughs of Central Toronto, East Toronto, East Toronto Business, West Toronto and East York.

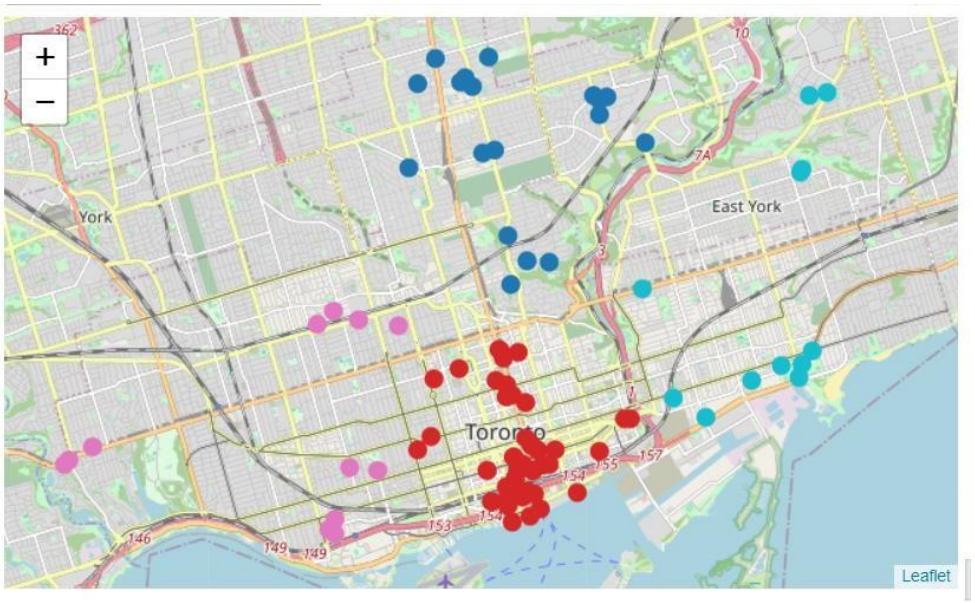
Clustering Neighborhoods

For further analysis, let's narrow down our focus to these selected boroughs having a high density of sports related venues and find some suitable neighborhoods in these boroughs.

To select suitable neighborhoods in these boroughs, clustering is done based on the latitude and longitude values of each neighborhood. The cluster centroids will then correspond to the required suitable neighborhoods.

The venues located in these boroughs are clustered into 4 groups, based on the venue coordinates, using k-means clustering algorithm.

The following map shows the groups of venues clustered using k-means clustering algorithm, shown in different colors.

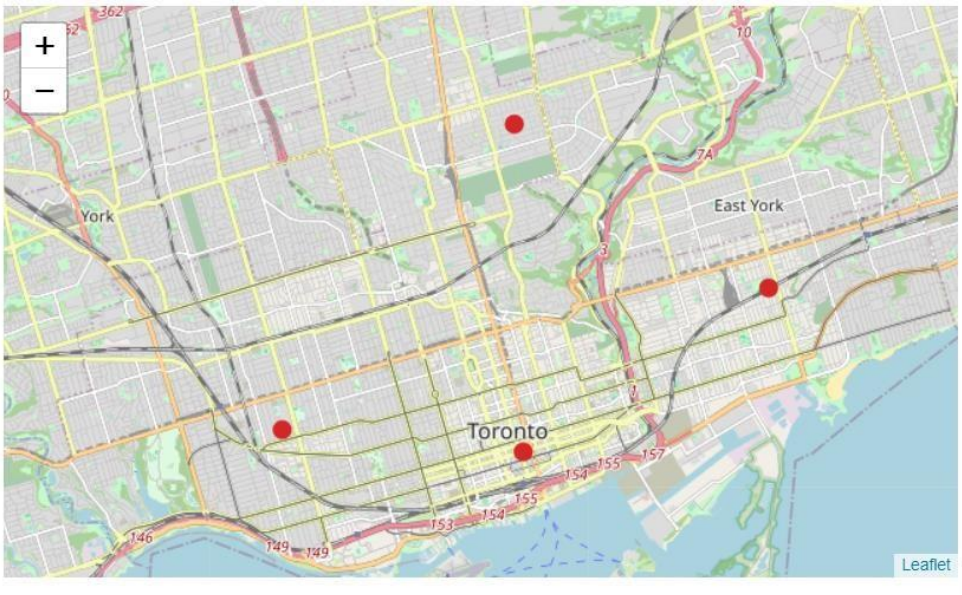


After grouping all the sports related venues in and around Toronto, Ontario, into 4 groups, the centers of these groups are determined. This information is stored in a DataFrame, shown below.

Latitude	Longitude	Address
43.65301	-79.42022	382, Shaw Street, University—Rosedale, Old Toronto, Toronto, Golden Horseshoe, Ontario, M6G 1C5, Canada

43.68502	-79.31830	12, Glebemount Avenue, Danforth Village, Beaches—East York, Old Toronto, Toronto, Golden Horseshoe, Ontario, M4C 1K9, Canada
43.70570	-79.38158	357, Manor Road East, Davisville, Don Valley West, Old Toronto, Toronto, Golden Horseshoe, Ontario, M4S 1S3, Canada
43.64974	-79.37879	Tim Hortons, Grand Opera Lane, King East, Toronto Centre, Old Toronto, Toronto, Golden Horseshoe, Ontario, M5H, Canada

Hence, we have narrowed down to 4 locations that can be targeted by the company to setup their first retail store in Toronto, Ontario, Canada. The target locations are shown with red markers on the map shown below.



Results and Discussion

Our analysis shows that there are 171 sports related venues in and around Toronto, Ontario, located in 103 neighborhoods.

A high density of such venues is observed in Downtown Toronto as well as the surrounding boroughs of East Toronto, East Toronto Business, Central Toronto, West Toronto, and East York.

Upon grouping these venues into 4 different clusters, we have identified 4 locations that the company can target in order to setup their first retail store in Toronto, Ontario, Canada.

These locations are:

1. 382, Shaw Street, University—Rosedale, Old Toronto, Toronto, Golden Horseshoe, Ontario, M6G 1C5, Canada
2. 12, Glebemount Avenue, Danforth Village, Beaches—East York, Old Toronto, Toronto, Golden Horseshoe, Ontario, M4C 1K9, Canada
3. 357, Manor Road East, Davisville, Don Valley West, Old Toronto, Toronto, Golden Horseshoe, Ontario, M4S 1S3, Canada
4. Tim Hortons, Grand Opera Lane, King East, Toronto Centre, Old Toronto, Toronto, Golden Horseshoe, Ontario, M5H, Canada

Therefore, the results of our analysis provide valuable information to the company which can be used to solve the problem at hand.

Conclusion

Hence, we have successfully identified 4 prime locations that can be targeted by UP! Sports as locations to setup their first retail store in Toronto, Ontario, Canada.