Exploring New York City Sports Venues

(Part 1)

Thomas Jacobs

13 June 2020

Content

1. Introduction

1.1. Background

Nowadays more people are health conscious and take part in various sport activities, not only to reach their ultimate physical potential, but also for reasons such as self-improvement, health improvement, etc.

The sports nutrition and sports equipment market is booming as people are becoming more concerned of what they eat in order to stay healthy and the quality of sports equipment used to reach their peak. Competition is fierce between some of the major players in sports nutrition in North American market – Nestlé S.A, Glanbia plc, Abbott Nutrition Inc, Yakult Honsha Co Ltd, The Coca-Cola Company, Reckitt Benckiser Group Plc, GNC Holdings, and Monster Beverage Corporation, among others. Major players in the sports equipment arena include Adidas, Nike, New Balance, etc.

Therefore, it is critical for a small, start-up company in this industry to ensure that it targets its products at the right target audience for optimal impact.

New York City has a total of 5 boroughs and 306 neighborhoods as indicated in https://en.wikipedia.org/wiki/Boroughs_of_New_York_City.

1.2. Problem

The marketing team of the start-up is highly skilled and experienced in their craft but are not local to New York and are not familiar with the different neighbourhoods. Their marketing budget is constraint and they want to ensure that every cent is optimally invested to get maximum return.

1.3. Value proposition

The aim of the project is to assist the marketing team in understanding their target area in order to develop the best marketing strategy for their portfolio of products. New York City will be clustered into the optimal number of different sports clusters to understand the most prevalent sports associated with each cluster.

2. Data

2.1. Initial datasets

We are interested in what sports activities people engage in and also where the different sports venues are located.

We will use the dataset available from https://geo.nyu.edu/catalog/nyu_2451_34572 which lists all 5 boroughs and the neighborhoods that exist in each borough as well as the corresponding geographical coordinates (latitude and longitude) of each neighbourhood.

We also require access to the Foursquare location data in order to get a dataset of all the different venue categories in the Manhattan area. Foursquare location data API calls will be executed to populate a dataset with all the different venue categories in the different neighbourhoods of Manhattan within a radius of 500m from each neighbourhood geographical coordinates. The dataset will also contain the geographical coordinates for these different venues.

2.2. Cleaning the datasets

We are however only interested in the target audience with includes all sports and health related venues. The full venue categories dataset will then be filtered to only extract the venue categories that are related to sports activities, e.g. gyms, health shops, stadiums, yoga studios, etc.

The sports dataset will contain the neighbourhood name, venue name, venue category and geographical coordinates.

2.3. Exploring the datasets

The sports dataset will be analysed to get a breakdown of the different sports activities in the different neighboorhoods. This will also assist in understanding which sports activities are prevalent.

The sports dataset will be analysed to get a graph of the breakdown of the number of different sports and health related venue categories in New York City's different boroughs. The top 5 venue categories in Manhattan will then be explored to see where they are located.

The next step would be to conduct statistical analysis on the sports dataset to cluster the dataset into the optimal number of clusters that will be crucial in the targeted marketing objective.