

Where to open a sports shop in London

Introduction:

A popular American sports shop, such as Fanatics wants to open stores in the U.K. They want to start by opening a store in London to test the waters. The sports shop should open this store in a borough where there is a high affinity for sports in general. As such, by exploring the most popular venues of each borough and finding the ones with the greatest number of gyms or sports facilities we can infer that the affinity for sports is high and therefore that borough will be a good place to start.

Data:

We will need a list of the boroughs in London with their corresponding geo-locational information, specifically latitude and longitude, the wikipedia page: https://en.wikipedia.org/wiki/List_of_London_boroughs was used as the source. The data was easy to scrape as the url contains a list of boroughs and the corresponding latitude and longitude values.

We will also need venue data for each borough. Foursquare can provide this information and a venue summary so they can be grouped into categories and examined.

Therefore with the borough list, their latitude and longitude coordinates and the venue list generated with Foursquare, we can find the most popular venues in each borough and choose the one with the greatest number of sport related venues.

Methodology:

The list of boroughs in London along with their latitude and longitude values was extracted from https://en.wikipedia.org/wiki/List_of_London_boroughs and processed into a data frame for analysis.

After gathering the borough data frame, we connected to the Foursquare API to gather venue information for each borough. The data was then processed and grouped into categories with mean information in each cell.

Our main goal was to clearly present venue category. One-hot coding was used to so this and display the count for each category. 5 clusters were generated using K-means clustering and the information was displayed as a table containing the cluster label and the top 10 most popular venues per borough.

Results:

The 5 clusters were examined and the top 10 venues were examined to find which had the greatest number of sport related facilities. This was done by displaying each cluster as a data frame containing the boroughs of that cluster.

Discussion:

From the clusters we can see that the results aren't very conclusive, mostly because by using borough data we are by default overlapping some boroughs. To improve this we could have retrieved a list of the individual areas in each borough and repeated the same progress. The usefulness of this investigation is to give a quick overview of the problem posed and by reducing the result to a single borough we can then just focus on that one and scout the best area to open up the shop.

Conclusion:

After examining the 5 clusters, we can recommend our stakeholders that Islington is the best borough to open up a sports shop as it is the only one with Gym as one of its top venues.