Battle of London Boroughs: New Fitness Centre

IBM Data Specialization Capstone Project

Rupangi Gupta

08/07/2020

London Fitness Industry

The 2019 State of the UK Fitness Industry Report[1] reveals that the UK health and fitness industry is healthier than it has ever been and is worth £5.1 billion.

- Number of fitness facilities in the UK up from 7,038 to 7,239 in 2019
- In last 12 months, the total UK membership grew by 4.7% and has broken the 10 million mark[2]

Along with the industry the customer appetite is also growing both vertically and horizontally. Customers are looking for fitness options clubbed with relaxation and enjoyment. Fitness is no longer restricted to committing few hours daily in gym- Fitness is now a Lifestyle Change.

Business Problem

The objective of this project is to segment London boroughs into clusters with clear view of current market on fitness centre categories. This will help fitness service provider to focus on specific boroughs based on fitness services they are offering.

Data Acquisition & Preparation

Data Acquisition:

- List of London boroughs & demographic data - acquired from free public data source[3].

- London Borough coordinates - acquired using geocoders library on Python

- Venue data for each borough - acquired using Foursquare API[4].

Data Preparation:

- Removed duplicate and irrelevant features. Filled missing values and gave relevant names to features.

Exploratory Data Analysis

Correlation heatmap

There is strong correlation between certain features like 'Number of households' & 'Total Population' OR 'BAME population' & 'Median Household income'.

However, there is no strong correlation between 'Parent Category' of the venue and any other features. Hence, no single feature is directly influencing the category of the fitness centre (⬅).
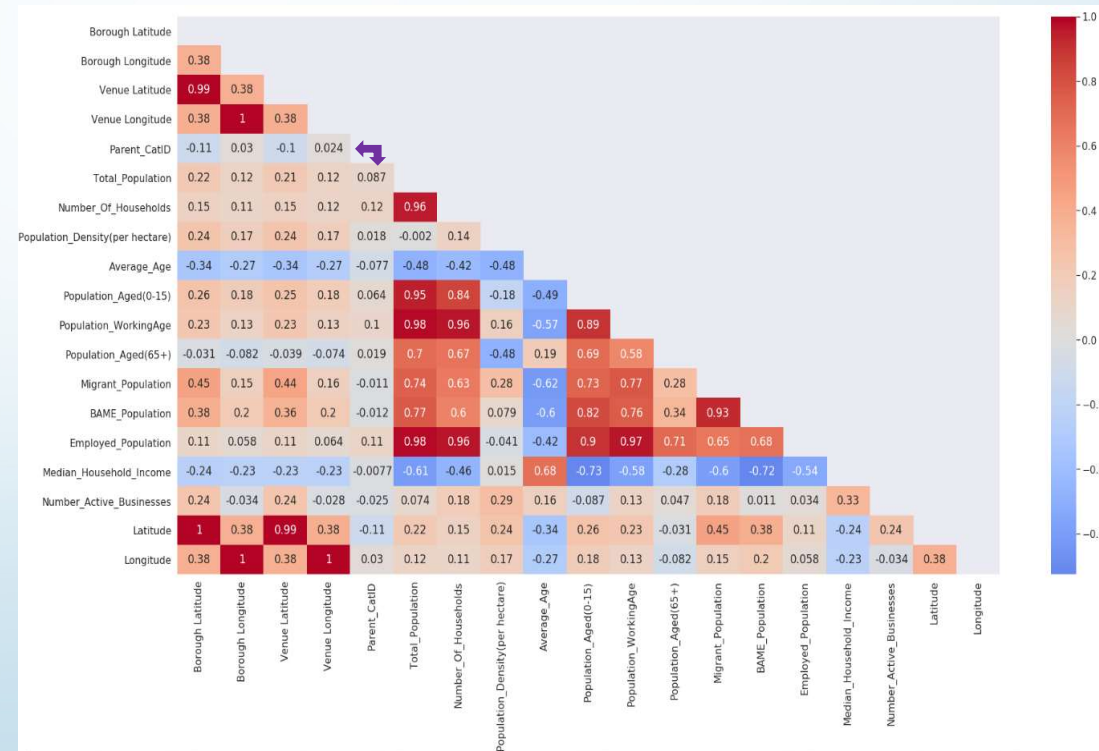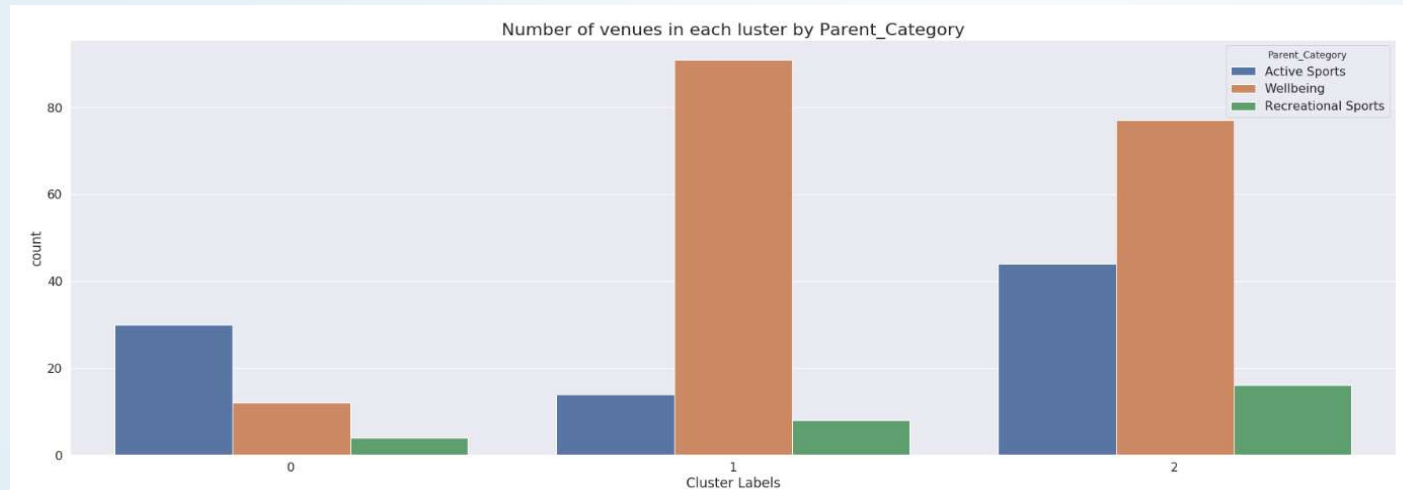
Exploratory Data Analysis

Analysed number of fitness centres by category for each borough. As we can see some boroughs like 'Wandsworth' have high concentration of venues as compared to 'Bexley'.
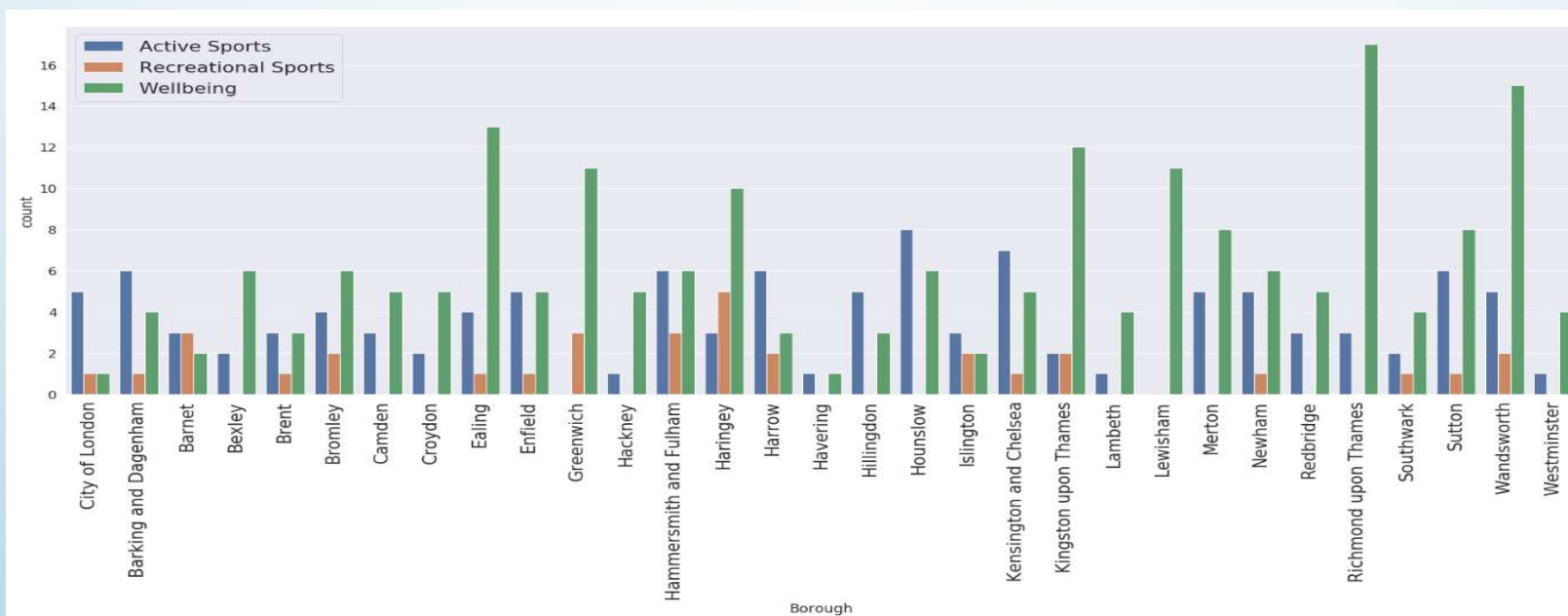
K-Means Clustering

Performed k-means clustering on venue data and clustered them into 3 buckets. When we examine the above graph we can cluster boroughs based on frequency of occurrence of each fitness centre category:

Cluster 0- 'High number of Active Sports Centres'

Cluster 1- 'High number of Wellbeing Centres'

Cluster 2- 'High number of All Fitness Centres'

Fitness Market Segmentation

Bubble map of London boroughs based on 3 clusters

- **Cluster 0**- 'High number of All Fitness Centres'

- **Cluster 1**- 'High number of Wellbeing Centres'

- **Cluster 2**- 'High number of Active Sports Centres'

The size of the bubble is based on 'Median Household Income'. Similar comparison can be done with other features.
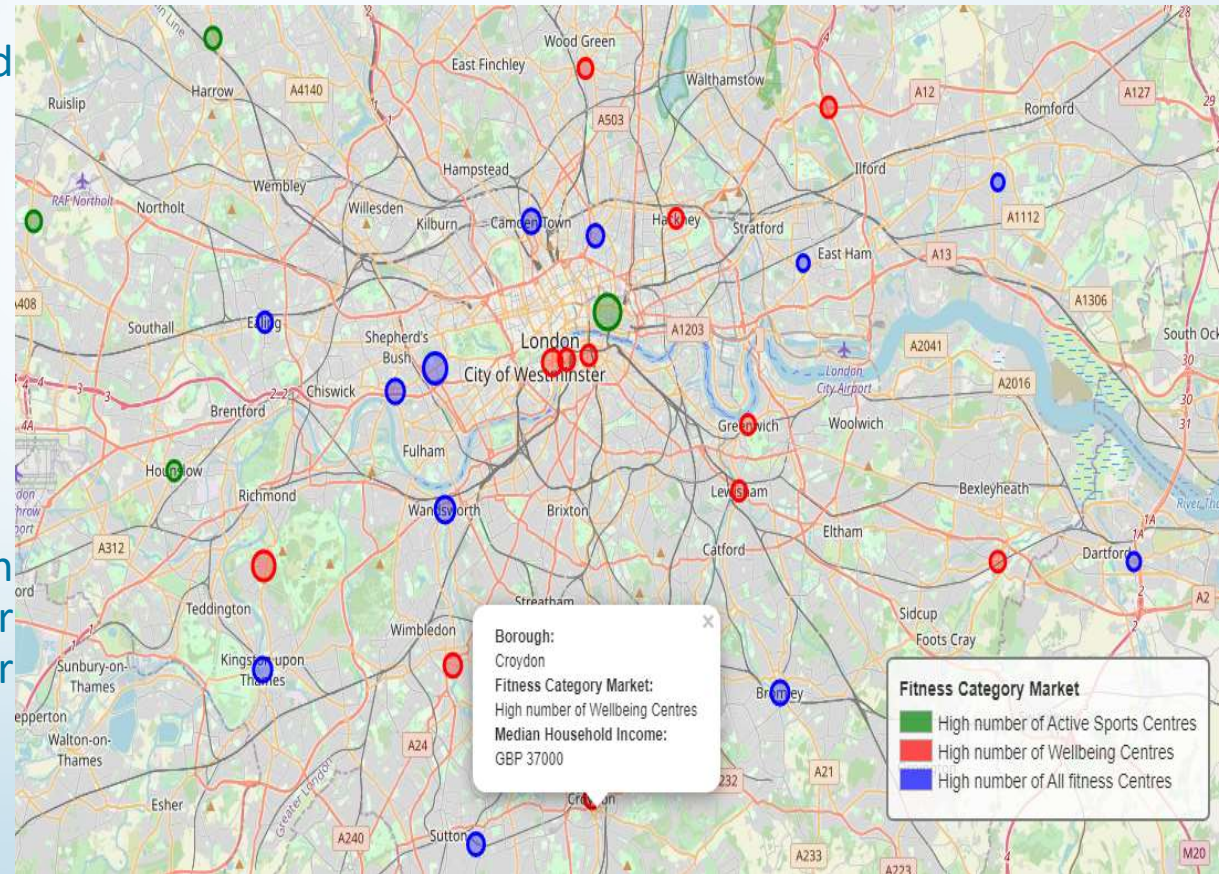
Recommendations

✓ We observed certain boroughs that are worth exploring further. For e.g. when we compare features like 'total population', 'population density' and 'median household income' of Cluster 1 & Cluster 0 boroughs. There are multiple boroughs in Cluster 1 like 'Croydon', 'Greenwich', 'Lewisham' that have similar range in all features but still very low number of Active Sports Centre indicating a good opportunity for a new Active Sports Centre

Cluster 1 - 'High number of Wellbeing Centres'

	Borough	Active Sports	Recreational Sports	Wellbeing	Fitness Category Market	Total Population	Population Density	Median Household income
3	Brent	3	0	3	High number of Wellbeing Centres	0.627053	0.259133	0.581421
6	City of London	6	1	2	High number of Wellbeing Centres	0.622433	0.715158	0.687677
7	Croydon	1	0	8	High number of Wellbeing Centres	0.992043	0.287110	0.581578
10	Greenwich	1	3	10	High number of Wellbeing Centres	0.718943	0.380254	0.555643
11	Hackney	1	0	7	High number of Wellbeing Centres	0.704055	0.925308	0.552342
21	Lambeth	1	0	6	High number of Wellbeing Centres	0.844199	0.788315	0.604998
22	Lewisham	1	0	9	High number of Wellbeing Centres	0.778747	0.554667	0.564288
25	Redbridge	1	0	6	High number of Wellbeing Centres	0.780801	0.346471	0.579378

Cluster 0 – 'High number of Active Sports Centres'

	Borough	Active Sports	Recreational Sports	Wellbeing	Fitness Category Market	Total Population	Population Density	Median Household income
0	Barking and Dagenham	2	0	4	High number of Active Sports Centres	0.022587	0.194728	1.000000
9	Enfield	4	2	2	High number of Active Sports Centres	0.854723	0.264725	0.520434
14	Harrow	4	2	2	High number of Active Sports Centres	0.647587	0.321277	0.611129
16	Hillingdon	5	0	2	High number of Active Sports Centres	0.772587	0.167171	0.582207
17	Hounslow	9	0	3	High number of Active Sports Centres	0.703799	0.314763	0.555329

Recommendations

✓ 'Cluster 2' clearly shows high density of all types of fitness centres hence we will not recommend them for a new fitness centre or suggest to perform extensive research before finalizing any of the boroughs from below list:

Cluster 2 - 'High number of All Fitness Centres'

	Borough	Active Sports	Recreational Sports	Wellbeing	Fitness Category Market	Total Population	Population Density	Median Household income
1	Barnet	3	3	2	High number of All Fitness Centres	0.536448	0.371945	0.462433
2	Bexley	2	0	6	High number of All Fitness Centres	1.000000	0.288597	0.637064
4	Bromley	4	2	6	High number of All Fitness Centres	0.852413	0.493617	0.505187
5	Camden	3	0	5	High number of All Fitness Centres	0.841632	0.140344	0.676831
6	City of London	5	1	1	High number of All Fitness Centres	0.622433	0.715158	0.687677
8	Ealing	4	1	13	High number of All Fitness Centres	0.902464	0.406764	0.566960
12	Hammersmith and Fulham	6	3	6	High number of All Fitness Centres	0.475616	0.726160	0.688777
15	Havering	1	0	1	High number of All Fitness Centres	0.652721	0.145448	0.576391
18	Islington	3	2	2	High number of All Fitness Centres	0.593429	1.000000	0.625432
19	Kensington and Chelsea	7	1	5	High number of All Fitness Centres	0.408111	0.842721	0.874253
20	Kingston upon Thames	2	2	12	High number of All Fitness Centres	0.450205	0.302487	0.690663
24	Newham	5	1	6	High number of All Fitness Centres	0.880133	0.608711	0.452373
28	Sutton	6	1	8	High number of All Fitness Centres	0.520021	0.296916	0.627790

Limitation & Future Research

- In this project we considered very limited factors and unfortunately couldn't establish very strong influence of any one factor on the fitness category of centres. Future research could device a methodology to further enhance this research with more factors and more data to determine preferred borough (or even postal code) to open a specific fitness centre.

- We also used free Sandbox Tier account of Foursquare API that came with limitation as to number of API calls and results we could fetch. Future research could make use of paid account to bypass this limitation and get better information like price bracket, social presence etc.

References

- [1] https://www.sportsthinktank.com/news/2019/05/the-2019-state-of-the-uk-fitness-industry-report
- [2] https://cilconsultants.com/base/assets/The-UK-health-and-fitness-industry-in-2019.pdf
- [3] https://data.london.gov.uk/dataset/london-borough-profiles
- [4] https://foursquare.com/developers/apps