# LONDON NEIGBORHOODS ANALYSIS


Abstract

Recommendation for suitable location to introduce new retail shops offering health nutrition products


Capstone Project

# Contents

# Introduction Business Objective and Problem

## Suitable Locations for New Retail store for Health Nutrition in City of London

**Description of Problem**

A client is planning to introduce their chain of stores which specializes in health based nutrition products, ranging from food supplements, vitamins, meal plans, juices, etc. They will be carrying out detail on ground market research activity to decide the exact target location of their stores, but for initial feasibility they need to narrow down the broader target neighborhood where they can start their kick start their detail market sizing/feasibility activity. For this purpose they have request to identify the neighborhoods in London which have higher concentration of population which will be interested in their product offering. Without utilizing existing public data the company could end up spending a large amount of money on ground market research encompassing the whole city.

**Criteria**

The client has already communicated that generally their stores are located in proximity to areas which have higher concentration of other facilities providing services to health/fitness enthusiasts. For this purpose we will be using London City Borough data (available on Wikipedia) and map it through FOURSQUARE to find out concentration of facilities which are of interest to health/fitness conscious people. And based on this analysis the suitable neighborhoods will be recommended to the client to initiate their detail on-ground market research.

# DATA Section

**List of London Boroughs**

**In this project, I will be using the following dataset to help solve problem - List of London Boroughs, and Foursquare API.**
Information on boroughs and their population & coordinates
Coordinates can be used to get neighborhood data from Foursquare. source: Wikipedia
url: https://en.wikipedia.org/wiki/List_of_London_boroughs

# Methodology

**Data Requirements**

The main districts in London are divided into 30 Boroughs (administrative districts)

The data regarding the districts in London needs to be researched and a suitable useable source identified. If it is found but is not in a useable form, data wrangling and cleaning will have to be performed.

The cleansed data will then be used alongside Foursquare data, which is readily available. Foursquare location data will be leveraged to explore or compare districts around London, identifying the areas where facilities providing services to our target market of health conscious individual are concentrated.

**Source Data**

Outline the initial data that is required: District data for London including names, location data. Research and find suitable sources for the district data for London. Access and explore the data to determine if it can be manipulated for our purposes.

**Data Cleaning**

Initial Data Wrangling and Cleaning: Clean the data and convert to a useable form as a dataframe.

Removed unnecessary strings from borough names

## Cleaning OF Data

```
# Strip unwanted texts
info['BoroughName'] = info['BoroughName'].map(lambda x: x.rstrip(']'))
info['BoroughName'] = info['BoroughName'].map(lambda x: x.rstrip('1234567890.'))
info['BoroughName'] = info['BoroughName'].str.replace('note','')
info['BoroughName'] = info['BoroughName'].map(lambda x: x.rstrip(' ['))
info.head()
```

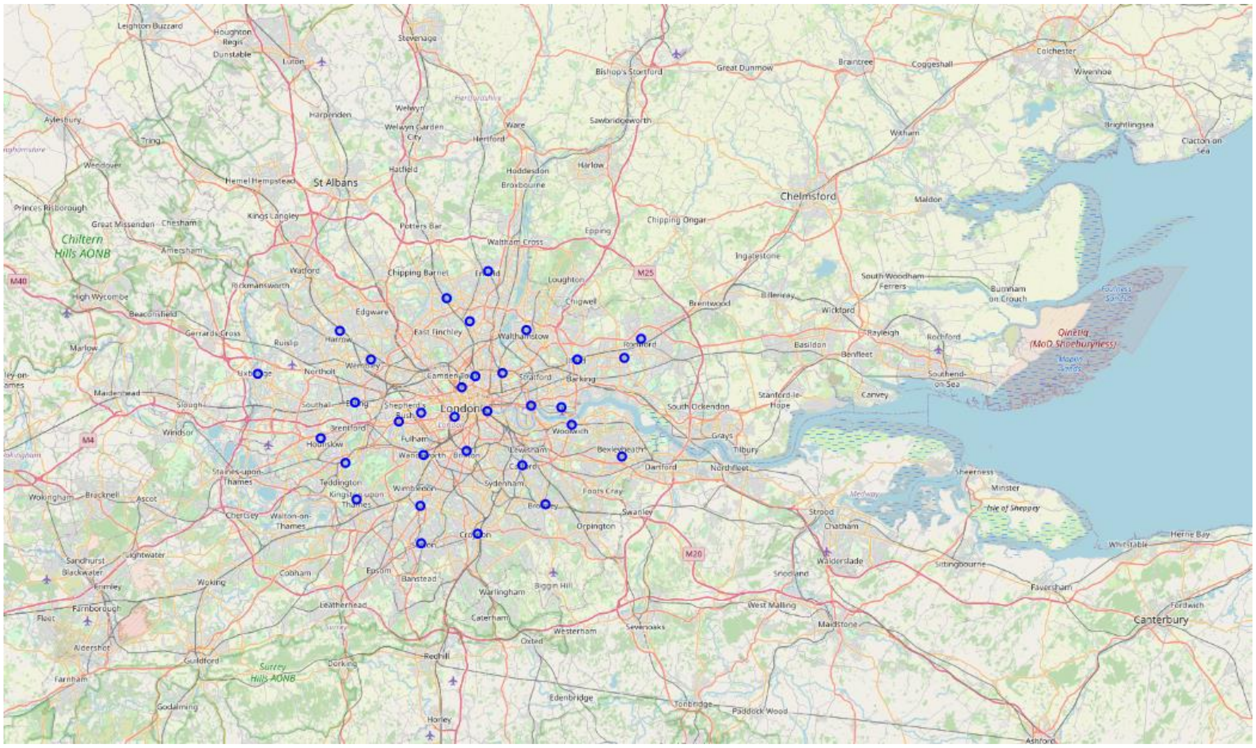
| | BoroughName | Population | Coordinates |
|---|---|---|---|
| 0 | Barking and Dagenham | 194,352 | 51°33′39″N 0°09′21″E / 51.5607°N 0.1557°E /... |
| 1 | Barnet | 369,088 | 51°37′31″N 0°09′06″W / 51.6252°N 0.1517°W /... |
| 2 | Bexley | 236,687 | 51°27′18″N 0°09′02″E / 51.4549°N 0.1505°E /... |
| 3 | Brent | 317,264 | 51°33′32″N 0°16′54″W / 51.5588°N 0.2817°W /... |
| 4 | Bromley | 317,899 | 51°24′14″N 0°01′11″E / 51.4039°N 0.0198°E /... |

Split coordinated to get Longitude and Latitude

```
# Clean coordinates
info[['Coordinates1','Coordinates2','Coordinates3']] = info['Coordinates'].str.split('/',expand=True)
info.head()
```

| | BoroughName | Population | Coordinates | Coordinates1 | Coordinates2 | Coordinates3 |
|---|---|---|---|---|---|---|
| 0 | Barking and Dagenham | 194,352 | 51°33′39″N 0°09′21″E / 51.5607°N 0.1557°E /... | 51°33′39″N 0°09′21″E | 51.5607°N 0.1557°E | 51.5607; 0.1557 (Barking and Dagenham) |
| 1 | Barnet | 369,088 | 51°37′31″N 0°09′06″W / 51.6252°N 0.1517°W /... | 51°37′31″N 0°09′06″W | 51.6252°N 0.1517°W | 51.6252; -0.1517 (Barnet) |
| 2 | Bexley | 236,687 | 51°27′18″N 0°09′02″E / 51.4549°N 0.1505°E /... | 51°27′18″N 0°09′02″E | 51.4549°N 0.1505°E | 51.4549; 0.1505 (Bexley) |
| 3 | Brent | 317,264 | 51°33′32″N 0°16′54″W / 51.5588°N 0.2817°W /... | 51°33′32″N 0°16′54″W | 51.5588°N 0.2817°W | 51.5588; -0.2817 (Brent) |
| 4 | Bromley | 317,899 | 51°24′14″N 0°01′11″E / 51.4039°N 0.0198°E /... | 51°24′14″N 0°01′11″E | 51.4039°N 0.0198°E | 51.4039; 0.0198 (Bromley) |

Plotted Boroughs Data on Maps for Visualization



# Data Analysis and Location Data:

**Explore London Neighborhood through FOURSQUARE**

### Found Top 100 venues that are in the neighborhood vicinity

## Clean Json Structure and convert it in dataframe

```
venues = results['response']['groups'][0]['items']

nearby_venues = json_normalize(venues) # flatten JSON

# filter columns
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
nearby_venues =nearby_venues.loc[:, filtered_columns]

# filter the category for each row
nearby_venues['venue.categories'] = nearby_venues.apply(get_category_type, axis=1)

# clean columns
nearby_venues.columns = [col.split(".")[-1] for col in nearby_venues.columns]

nearby_venues.head()
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/ipykernel_launcher.py:3: FutureWarning: pandas.i
  This is separate from the ipykernel package so we can avoid doing imports until
```

| | name | categories | lat | lng |
|---|---|---|---|---|
| 0 | Central Park | Park | 51.559560 | 0.161981 |
| 1 | Crowlands Heath Golf Course | Golf Course | 51.562457 | 0.155818 |
| 2 | Robert Clack Leisure Centre | Martial Arts Dojo | 51.560808 | 0.152704 |
| 3 | Morrisons | Supermarket | 51.559774 | 0.148752 |
| 4 | Beacontree Heath Leisure Centre | Gym / Fitness Center | 51.560997 | 0.148932 |

Analyze each neighborhood and find out the list of venues in each neighborhood

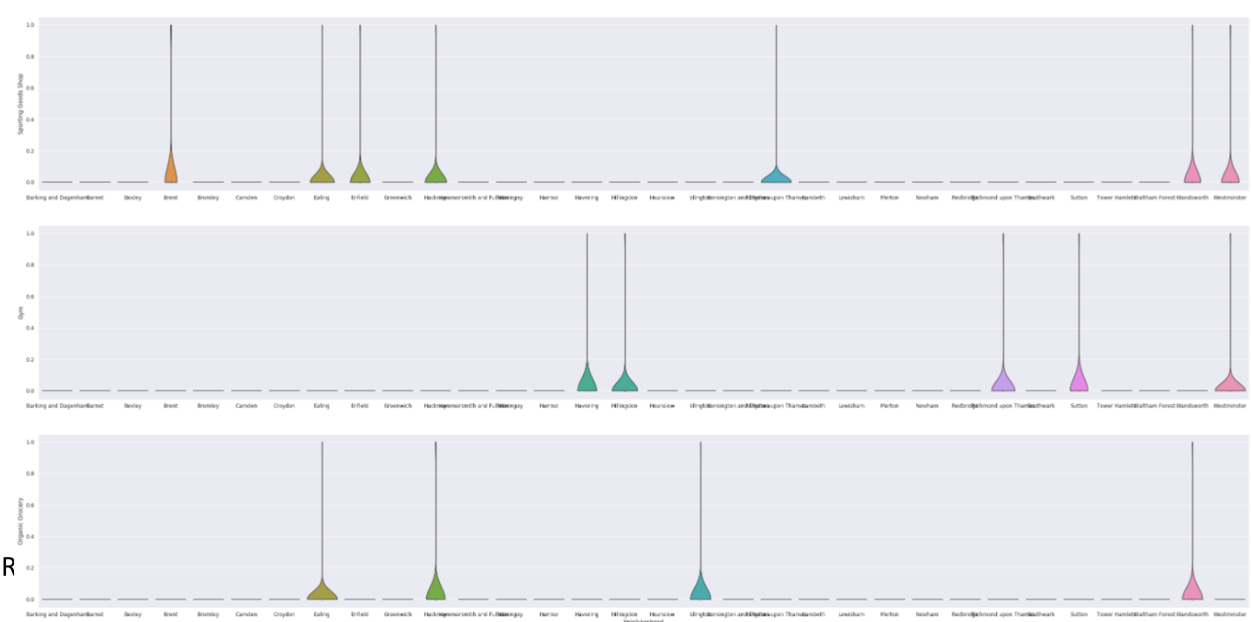
| | Neighborhood | African Restaurant | Airport | Airport Lounge | Airport Service | American Restaurant | Antique Shop | Argentinian Restaurant | Art Gallery | Art Museum | Arts & Crafts Store | Asian Restaurant | Australian Restaurant | BBQ Joint | Bakery | Bar | Bed & Breakfast | Beer Bar | Beer Garden | Beer Store | Bistro |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Barking and Dagenham | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | Barking and Dagenham | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | Barking and Dagenham | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | Barking and Dagenham | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | Barking and Dagenham | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Summarize data neighborhood wise and take the mean of frequency of each venue catagory

| | Neighborhood | African Restaurant | Airport | Airport Lounge | Airport Service | American Restaurant | Antique Shop | Argentinian Restaurant | Art Gallery | Art Museum | Arts & Crafts Store | Asian Restaurant | Australian Restaurant | BBQ Joint | Bakery | Bar |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Barking and Dagenham | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.000000 | 0.000000 |
| 1 | Barnet | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.000000 | 0.000000 |
| 2 | Bexley | 0.0 | 0.0 | 0.0 | 0.0 | 0.033333 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.033333 | 0.000000 |
| 3 | Brent | 0.0 | 0.0 | 0.0 | 0.0 | 0.026667 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.013333 | 0.0 | 0.0 | 0.000000 | 0.066667 |
| 4 | Bromley | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.025641 | 0.0 | 0.0 | 0.025641 | 0.025641 |

**Analysis Methodology**

We looked for the areas with higher concentration of venues already defined by client i.e 'Sporting Goods', 'GYM' and 'Yoga Studio'
Let's look at their frequency of occurrence for all the London neighborhoods, isolating the categorical venues These are the venue types that the client wants to have an abundant density of in the ideal store locations. I've used a violin plot from the Seaborn library - it is a great way to visualize frequency distribution datasets, they display a density estimation of the underlying distribution



R

Foursquare location data will be leveraged to explore or compare districts around London.
Data manipulation and analysis to derive subsets of the initial data.
Identifying the high traffic areas using data visualization and statistical analysis.

# Selcted neighborhoods:

so as we can see from the analysis there are 2 neighborhoods to open new stores - according to the criteria that they have the 2 of the 3 specified venues in a great frequency (Gyms and Organic Grocery). They are as follows:

Neighborhoods:
Wandsworth and Westminister

# Results

I guess it's not a surprise that these districts are all very centrally located in London. Locations fitting the criteria for popular venues would normally be in central locations in many cities of the world.

From this visualization it is clear that on a practical level, with no data to base decisions on, the circle of the 31 borough is very large, and researching and then visiting them all would be a daunting and time consuming task. We have narrowed the search area to suit the client's retail business.

# Discussion

This is the initial recommendations look for the initial stores locations, however after the on ground research, we could learn of additional insights and relevance for target market which can be used to rerun this analysis with additional parameter and help to locate suitable venues for additional stores.

# Conclusions

We have narrowed down the target locations for the customer to carry out the on ground research to further help in pinpointing the location of new healthy nutrition stores.

Without leveraging data to make focused decisions, the process could have been drawn out and resulted in new stores opening in sub-standard areas for this retailer. Data has helped to provide a better strategy and way forward, these data-driven decisions will lead to a better solution in the end.