

# Opening a new business venue: from New York to London

## **Business Objective**

An entrepreneur owns an arty/concept coffee-shop in SoHo, Manhattan, New York and would like to open a new location in London.

The entrepreneur wants to select an area which has the same characteristics and vibes of SoHo, an area known for its commercialization and eclectic mix of venues ranging from restaurants and coffee-shops, to shopping boutiques and art galleries.

## **Datasets**

A list of all areas of London, containing area name and OS Grid Ref is scraped from Wikipedia.

The centre of each London areas is located by converting OS Grid Reference number in geodetic coordinates (latitude and longitude).

SoHo, New York centre is located and added to the dataset

A list of venues around a radius of 500mt around every area centre is retrieved using the Foursquare API. Venue categories hierarchies are scraped from the Foursquare documentation

For every venue retrieved using the Foursquare API, the number of category keywords available is expanded using the category hierarchies.

**Few assumptions are made:**

Characteristics of an area can be inferred by the characteristics of the venues present in said area.

A theme is a specific mix of venue characteristics which are somehow related together. (e.g. the theme “Sport & Fitness” refers to all the venues related to sport and fitness, such as gyms, fitness centres, SPAs, golf courses, etc.).

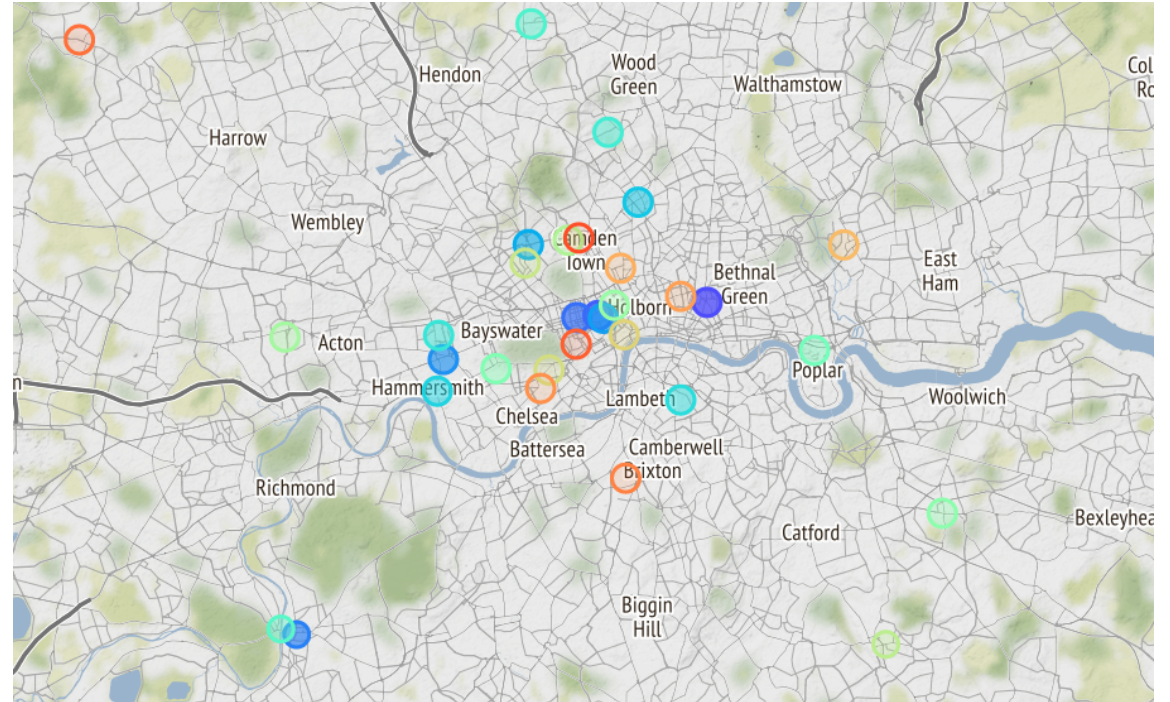
An area is defined by a combination of themes (e.g. an area could be 0.8 Sport & Fitness and 0.2 Residential)

Similarity between two areas is calculated based on the proportion of the same themes they share.

Using geographical data and venues dataset from the Foursquare API, themes composition of all the London areas were derived, and similarities were calculated using cosine similarity.

London most similar areas (similarity  $\geq 0.9$ ) (top)

borough	area	similarity
Hackney	Shoreditch	0.973975
Camden	Fitzrovia	0.972802
Westminster	Marylebone (also St Marylebone)	0.965995
Kingston upon Thames	Kingston upon Thames	0.962092
Hammersmith and Fulham	Shepherd's Bush	0.960121
Westminster	Soho	0.956681
Camden	Swiss Cottage	0.952429
Bromley	Orpington	0.952374
Islington	Holloway	0.947718
Hammersmith and Fulham	Hammersmith	0.947667
Southwark	Walworth	0.942715



Geographically closer areas were grouped into clusters.

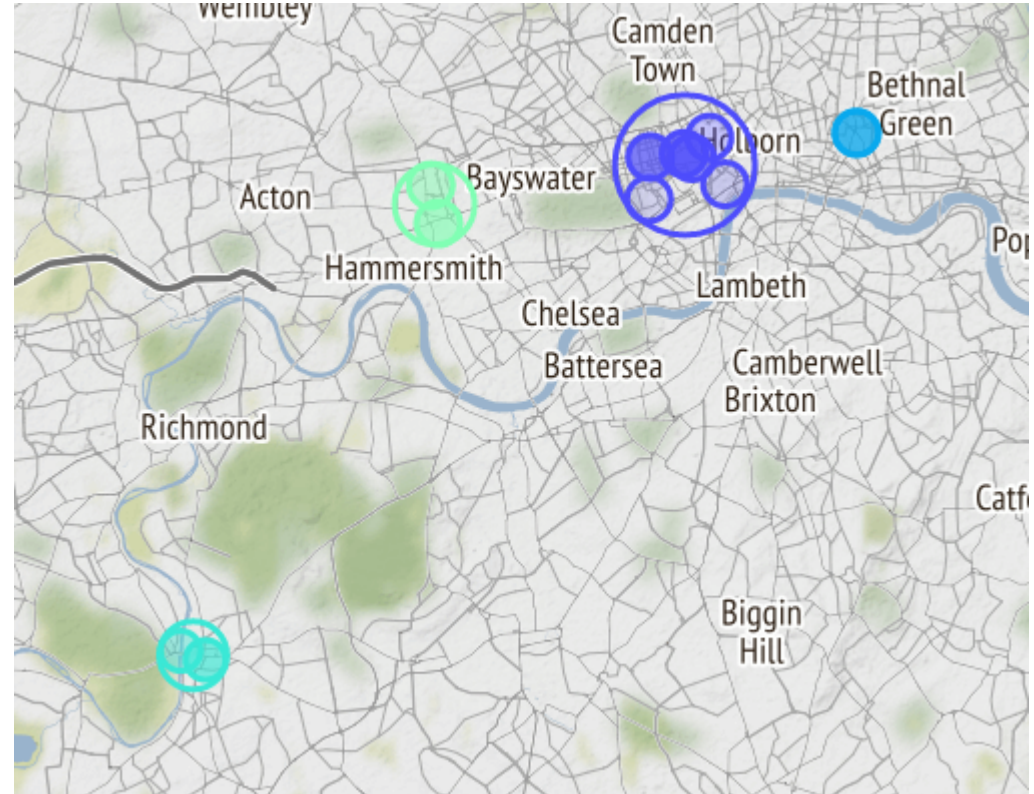
4 main optimal zones are identified, by clustering the most similar areas of London and grouping them by proximity.

**Cluster 0**, a cluster composed only by the areas of Shoreditch, a popular and fashionable part of London, particularly associated with the creative industries. Art galleries, bars, restaurants, media businesses are common in the area.

**Cluster 1**, a large cluster composed by London areas belonging to the very central boroughs of Westminster and part of Camden. This cluster is dense (boundaries of the areas overlapping) and covers the biggest area in meters. This zone famous for the reputation as a major entertainment district of London. It is filled with galleries, bars, restaurants and major theatres.

**Cluster 2**, a cluster composed by two areas belonging to the borough of Kingston Upon Thames. Although far from London's centre, Kingston is identified as a metropolitan area and is today a major retail centre, one of the biggest in the UK, receiving 18 million visitors a year.

**Cluster 3**, a cluster composed by two areas hosting a major luxury retail centre and campus of universities



Several additional external factor could be taken in account to choose which cluster is optimal (population of the areas, percentage of tourists, proximity to tourist attraction, average cost of rental). Anyway, for this project only the previously defined internal metrics are used: mean similarity of the cluster and cluster density/areas coverage.

A Cluster composed by the central areas of London (Fitzrovia, Marylebone, Soho, Bloomsbury, Covent Garden, Mayfair) was selected as the optimal cluster; having the optimal qualities of being composed by areas the most similar areas to SoHo and covering the biggest area in meters, giving gives the client a more options on where to open a venue.