How to choose the best London's borough for a restaurant

Samuel Beau

July 15, 2019

Contents

1 Introduction **2**

2 Data **2**
2.1 Data source 2
2.2 Data description 3

3 Methodology **3**

4 Discussion **4**

5 Conclusion **4**

1 Introduction

A restaurant company want to develop a branch in london. But as the rent in london is very expensive we want to be sure the chosen place will be the most interesting for this kind of activity. Thus the goal here is to choose the best place amid all the london's borough.

London is divided in 33 Borough with differents venues. Each borough is defined by a set of venues. This is what we will use in order to sort the borough into different groups. And find out which of this borough's group is the best option for a restaurant.

2 Data

In this part we describe our data source, the data needed for the analysis and some caracteristics of the data.

2.1 Data source

We will need the london's borough list, their latitude and longitude and the top 100 venues from each Borough.

We can get the london's borough list from wikipedia: `https://en.wikipedia.org/wiki/London_boroughs` To pull data out of the HTML and XML files we will use BeautifulSoup.

	Borough	Designation
0	Greenwich	Inner
1	Hackney	Inner
2	Hammersmith	Inner
3	Islington	Inner
4	Kensington and Chelsea	Inner

Figure 1: london's borough

Once we get the london's borough list we use geopy library to get the latitude and longitude values.

	Borough	latitude	longitude
0	Greenwich	51.482084	-0.004542
1	Hackney	51.543240	-0.049362
2	Hammersmith	51.492038	-0.223640
3	Islington	51.538429	-0.099905
4	Kensington and Chelsea	51.498995	-0.199123

Figure 2: borough's latitude and longitude

And then we eventually with Foursquare we get the top 100 venues from each Borough. All the data are merged in a simgle table.

	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Greenwich	51.482084	-0.004542	Old Royal Naval College	51.483234	-0.005579	Historic Site
1	Greenwich	51.482084	-0.004542	Painted Hall	51.482889	-0.006420	Museum
2	Greenwich	51.482084	-0.004542	National Maritime Museum	51.481329	-0.005581	History Museum
3	Greenwich	51.482084	-0.004542	Greenwich Naval College Gardens	51.483007	-0.008362	Garden
4	Greenwich	51.482084	-0.004542	The Plume of Feathers	51.481945	-0.001126	Pub

Figure 3: Venues Table

2.2 Data description

There are 220 uniques categories that can be curated from all the returned venues.

Also it's interesting for each borough to sort the venue by the most commons. We get this table.

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Barking	Platform	Grocery Store	Fast Food Restaurant	Supermarket	Theater	Coffee Shop	Pizza Place	Plaza	Portuguese Restaurant	Pub
1	Barnet	Pub	Supermarket	Train Station	Gym	Yoga Studio	Electronics Store	Fish Market	Fish & Chips Shop	Filipino Restaurant	Fast Food Restaurant
2	Bexley	Pub	Fast Food Restaurant	Italian Restaurant	Greek Restaurant	Breakfast Spot	Indian Restaurant	Toy / Game Store	Chinese Restaurant	Train Station	Yoga Studio
3	Brent	Pub	Indian Restaurant	Cupcake Shop	Fast Food Restaurant	Nightclub	Gym / Fitness Center	Yoga Studio	English Restaurant	Fish Market	Fish & Chips Shop
4	Bromley	Coffee Shop	Clothing Store	Gym / Fitness Center	Pub	Burger Joint	Pizza Place	Furniture / Home Store	Sushi Restaurant	Supermarket	Portuguese Restaurant

Figure 4: Most commons venues

3 Methodology

As we are working with unlabeled data we will choose k-means models for clustering. Using the Venue Category, we will separate the london's borough into

groups that have similar characteristics. And then from each group we will find the most suitable borough for a restaurant.

The k-means models groups the borough in five sets. We can see thoses clusters on the following london map.

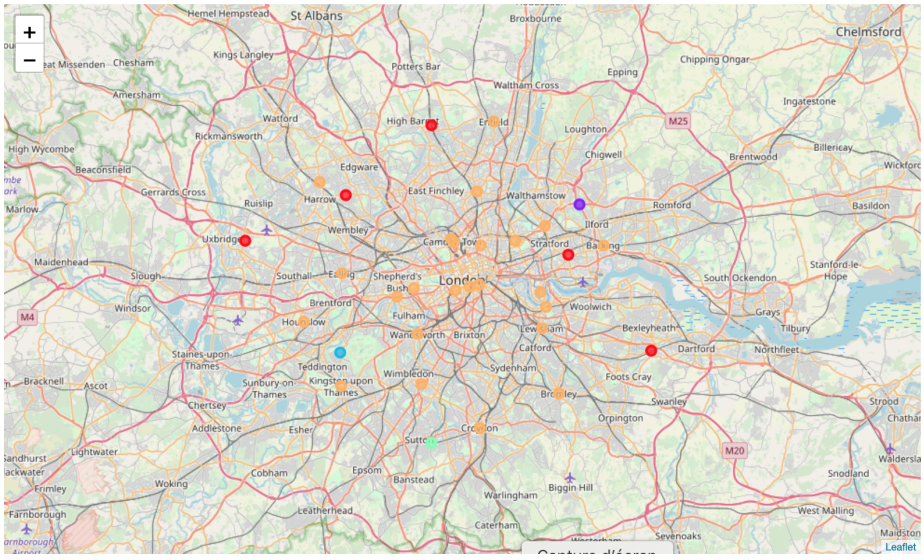

Figure 5: London map

Now, you can examine each cluster and determine the discriminating venue categories that distinguish each cluster.

4 Discussion

We choose to group the Borough into five clusters. It appear that two clusters over five Cluster 1 (red one), Cluster 5 (orange one) get most of the Borough. In the cluster 1 the most commun venue is the Pub followed by Fast Food Restaurant. Also all the Borough belonging to this cluster are far from central london.

The cluter 5 globally close to central london contain Borough with coffee shop or Pub as first commun venue and restaurant as second or third most commun venue. For example : Islington or Kensington and Chelsea could be a good choice for a restaurant.

5 Conclusion

In conclusion, I think this script is good enough to have a first view on the issue. But for a better analysis we may add some more information like the average rent in each Borough.