# Capstone Project: The Battle of the Neighborhoods (Week 2)

Course 9: Applied Data Science Capstone by IBM/Coursera

# Table of Contents

# Introduction: Business Problem

In this project, we will try to build a model to find the optimal neighborhood for opening a new business. As an example, we will specify the business type to be an **Italian restaurant**.

Since London is huge city with many restaurants, we will try to identify the optimal neighborhood to open a new Italian restaurant. We will try to identify the optimal neighborhood based on:

- Less number of **high rating** Italian restaurants.
- Less number of **Italian restaurants**.
- Less number of restaurants.

# Data

Based on the description of the Business Problem, we will need to get data of:

- List of neighborhoods in London with their latitudes, longitudes and area.
- Number of restaurants (any type) in each neighborhood.
- Number of Italian restaurants in the neighborhood.
- Rating of each Italian restaurants in the neighborhood.

We will extract the required data as follows:

- We will get the list of neighborhoods in London and their geo-location from Wikipedia page.
- We will get the restaurants information for each neighborhood from **Foursquare API**.
- Also, for the map visualization we will get the location of London using **geopy API**.

## Neighborhood Data

We will retrieve the list of London neighborhoods from Wikipedia page List of London boroughs which contains London neighborhoods and their Geo-Location information and other irrelevant information which will be excluded.

| | Neighbourhood | Latitude | Longitude | Area | Area-Radius |
|---|---|---|---|---|---|
| 0 | Barking and Dagenham | 51.5607 | 0.1557 | 36078700.0 | 3388.835625 |
| 1 | Barnet | 51.6252 | -0.1517 | 86739100.0 | 5254.513588 |
| 2 | Bexley | 51.4549 | 0.1505 | 60554200.0 | 4390.330342 |
| 3 | Brent | 51.5588 | -0.2817 | 43253000.0 | 3710.506368 |
| 4 | Bromley | 51.4039 | 0.0198 | 150142300.0 | 6913.159800 |

Each neighborhood in London has different area. We already consider/assume the Longitude and Latitude we got from Wikipedia to be the center of each neighborhood.

As we are going to provide the **radius** in our Foursquare APIs calls, we associated each neighborhood to a radius value of a circle whose area equals the neighborhood area.

## London City Location

We will use **geopy API** to get the Latitude and Longitude of London City.

Coordinate of London, UK: [51.5073219, -0.1276474]

Now we can visualize the London maps and its neighborhood. Each neighborhood center will be surrounded with circles of radius proportional to its area.
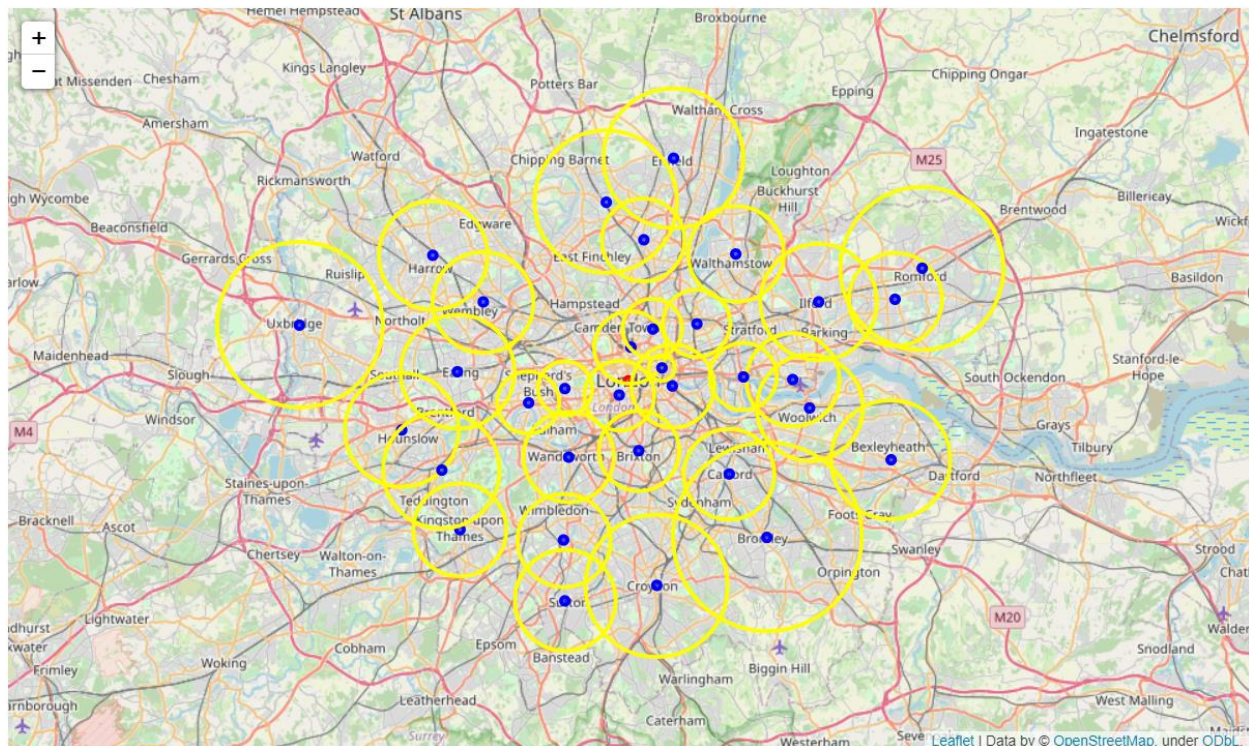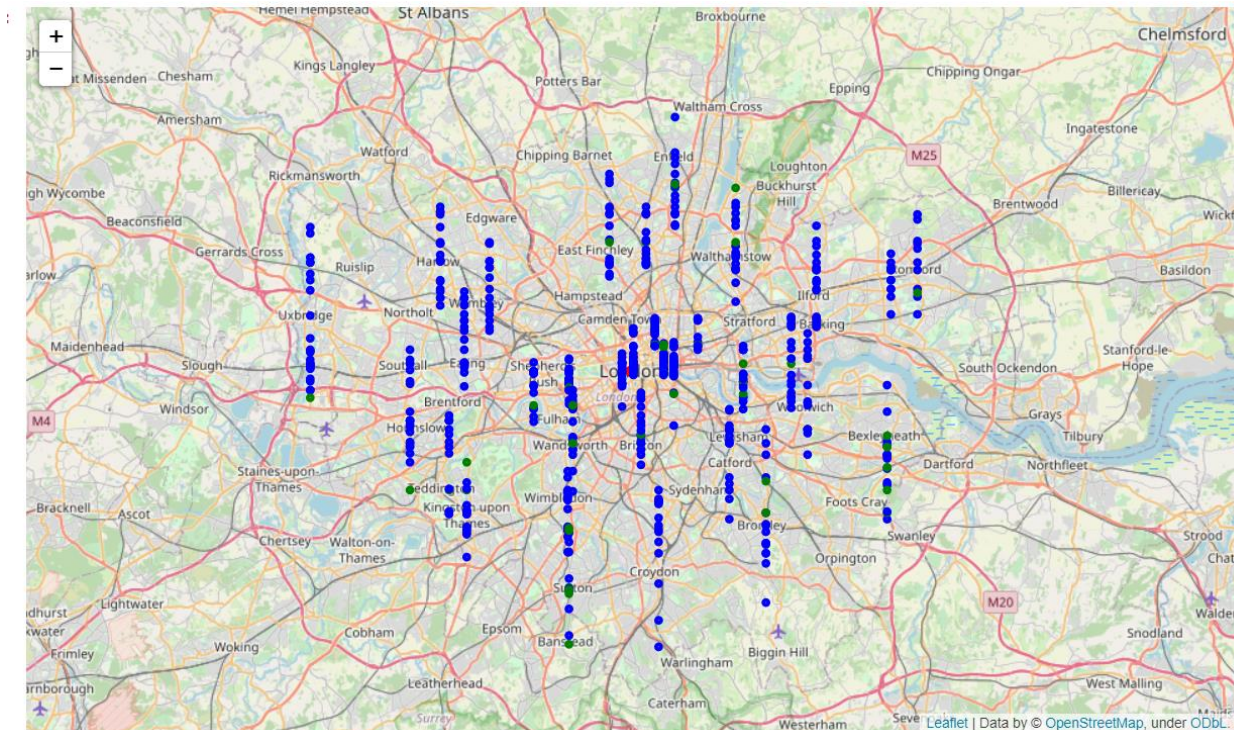


## Restaurants Data

Now we need to build a data set of restaurants in each neighborhood. Our focus will be in restaurants in general and especially Italian restaurants. For this purpose, we will use **Foursquare APIs**.

| Restaurant_ID | Restaurant_Name | Category_ID | Category_Name | Restaurant_Latitude | Restaurant_Longitude | Is_Italian | Rating |
|---|---|---|---|---|---|---|---|
| 4bb9e6183db7b713faa0229a | Moby Dick | 4bf58dd8d48988d1c4941735 | Restaurant | 51.580585 | 0.140618 | False | 0.0 |
| 4eb064fb775bbddde7b4f546 | Cosmo | 4bf58dd8d48988d1c4941735 | Restaurant | 51.575445 | 0.180456 | False | 0.0 |
| 50baf653183fee49a8b0b4cc | Lara Grill | 4f04af1f2fb6e1c99f3db0bb | Turkish Restaurant | 51.562445 | 0.147178 | False | 0.0 |
| 4e7a1f5414954a343fb58258 | The Pipe Major | 4bf58dd8d48988d1c4941735 | Restaurant | 51.545795 | 0.165834 | False | 0.0 |
| 4c3218fa3896e21e07f2e790 | The Compasses (Harvester) | 52e81612bcbc57f1066b7a05 | English Restaurant | 51.557806 | 0.209121 | False | 0.0 |

Now let's visualize the results on the map of London.



This concludes our Data Section.

Now we have the list of restaurants and their locations. Also, we know which restaurant is Italian or not.

## Methodology

In this project we will try to detect neighborhoods in London with low number of restaurants especially with low number of Italian restaurants. Also, since London contains many restaurants, we will consider the rating of the Italian restaurants so we pick neighborhoods with low average rating of Italian restaurants.

In the first section of the report, we gathered the data that will be used in our report:

- List of neighborhoods of London.

- The restaurants in each neighborhood.
- The Italian restaurants in each neighborhood.
- The rating of each Italian restaurant.

In the second section of the report, we will perform some analysis on the data we gathered so we could have better understanding of the distribution of restaurants in London and -if needed- we will enrich the data with any extra information to provide better decision.

In the third section, we will use the *K-Means* clustering to split London neighborhoods into similar clusters based on the criteria defined above and accordingly, we can decide which neighborhood(s) best to opening a new Italian restaurant.

## Analysis

Let's perform some basic analysis on the data.

### All Restaurants Count

Let's check the total count of restaurants per neighborhood.

|  | Neighbourhood | count |
|---|---|---|
| 6 | City of London | 27 |
| 16 | Hillingdon | 26 |
| 5 | Camden | 24 |
| 27 | Southwark | 23 |
| 19 | Kensington and Chelsea | 23 |
| 17 | Hounslow | 21 |
| 32 | Westminster | 21 |
| 29 | Tower Hamlets | 20 |
| 14 | Harrow | 20 |
| 21 | Lambeth | 20 |
| 24 | Newham | 19 |
| 31 | Wandsworth | 19 |
| 2 | Bexley | 19 |
| 18 | Islington | 19 |
| 8 | Ealing | 19 |
| 30 | Waltham Forest | 18 |
| 25 | Redbridge | 18 |
| 7 | Croydon | 17 |
| 9 | Enfield | 17 |
| 12 | Hammersmith and Fulham | 17 |
| 1 | Barnet | 17 |
| 4 | Bromley | 17 |
| 3 | Brent | 17 |
| 20 | Kingston upon Thames | 16 |
| 23 | Merton | 16 |
| 13 | Haringey | 16 |
| 22 | Lewisham | 15 |
| 10 | Greenwich | 15 |
| 26 | Richmond upon Thames | 14 |
| 28 | Sutton | 14 |
| 0 | Barking and Dagenham | 12 |
| 15 | Havering | 11 |
| 11 | Hackney | 11 |
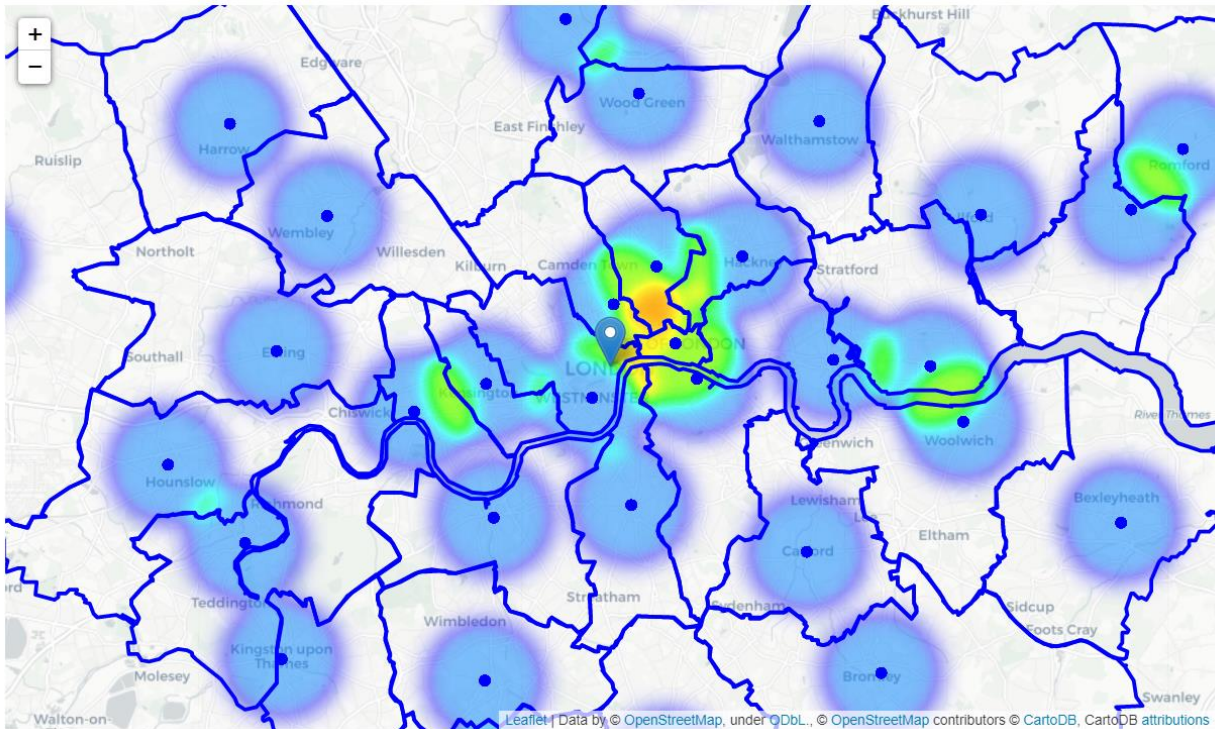
## Italian Restaurants Count

Let's check the Italian Restaurants in each neighborhood.

| | Neighbourhood | count |
|---|---|---|
| **19** | Sutton | 6 |
| **1** | Bexley | 5 |
| **3** | City of London | 4 |
| **20** | Tower Hamlets | 4 |
| **21** | Waltham Forest | 3 |
| **12** | Islington | 2 |
| **22** | Wandsworth | 2 |
| **18** | Southwark | 2 |
| **16** | Newham | 2 |
| **15** | Lambeth | 2 |
| **14** | Kingston upon Thames | 2 |
| **13** | Kensington and Chelsea | 2 |
| **23** | Westminster | 2 |
| **4** | Enfield | 2 |
| **2** | Bromley | 2 |
| **11** | Hounslow | 1 |
| **10** | Hillingdon | 1 |
| **9** | Havering | 1 |
| **8** | Haringey | 1 |
| **17** | Richmond upon Thames | 1 |
| **7** | Hammersmith and Fulham | 1 |
| **6** | Hackney | 1 |
| **5** | Greenwich | 1 |
| **0** | Barnet | 1 |
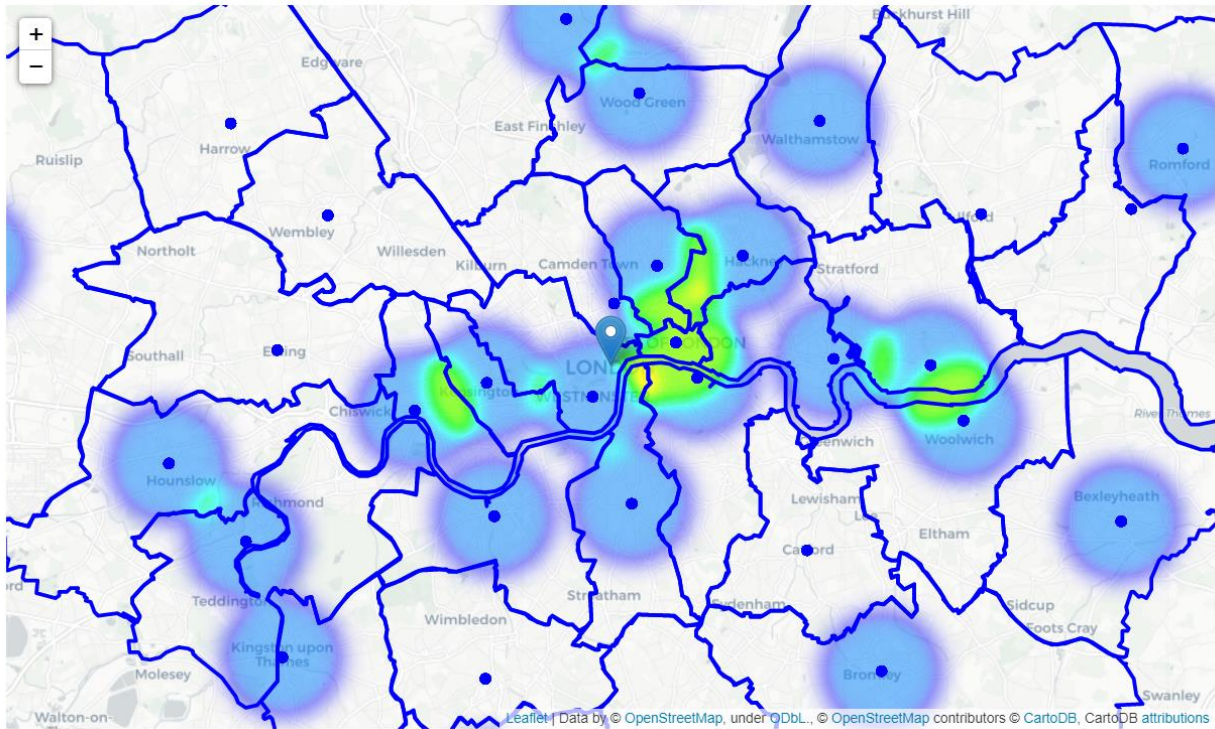
## Italian Restaurants Rating

Let's check the average rating of the Italian restaurants in each neighborhood.

| | Neighbourhood | Rating |
|---|---|---|
| 18 | Southwark | 8.900000 |
| 23 | Westminster | 8.850000 |
| 13 | Kensington and Chelsea | 8.850000 |
| 3 | City of London | 8.500000 |
| 7 | Hammersmith and Fulham | 8.400000 |
| 8 | Haringey | 8.400000 |
| 0 | Barnet | 8.400000 |
| 22 | Wandsworth | 8.200000 |
| 6 | Hackney | 8.100000 |
| 12 | Islington | 8.100000 |
| 21 | Waltham Forest | 7.800000 |
| 15 | Lambeth | 7.650000 |
| 20 | Tower Hamlets | 7.425000 |
| 11 | Hounslow | 7.300000 |
| 4 | Enfield | 7.300000 |
| 17 | Richmond upon Thames | 7.300000 |
| 5 | Greenwich | 6.700000 |
| 9 | Havering | 6.700000 |
| 16 | Newham | 6.550000 |
| 19 | Sutton | 5.766667 |
| 1 | Bexley | 5.500000 |
| 14 | Kingston upon Thames | 3.650000 |
| 10 | Hillingdon | 0.000000 |
| 2 | Bromley | 0.000000 |

As we can see from the analysis and visualization, we can see that **City of London** is a hot place for restaurants. This is a typical for most of the cities.

Based on that, we can include calculate the distance to the **City of London** as an attribute to each neighborhood.

| Restaurant_Name | Category_ID | Category_Name | Restaurant_Latitude | Restaurant_Longitude | Is_Italian | Rating | count | Distance_To_Center |
|---|---|---|---|---|---|---|---|---|
| Moby Dick | 4bf58dd8d48988d1c4941735 | Restaurant | 51.580585 | 0.140618 | False | 0.0 | 1 | 17920.030817 |
| Cosmo | 4bf58dd8d48988d1c4941735 | Restaurant | 51.575445 | 0.180456 | False | 0.0 | 1 | 17920.030817 |
| Lara Grill | 4f04af1f2fb6e1c99f3db0bb | Turkish Restaurant | 51.562445 | 0.147178 | False | 0.0 | 1 | 17920.030817 |
| The Pipe Major | 4bf58dd8d48988d1c4941735 | Restaurant | 51.545795 | 0.165834 | False | 0.0 | 1 | 17920.030817 |
| The Compasses (Harvester) | 52e81612bcbc57f1066b7a05 | English Restaurant | 51.557806 | 0.209121 | False | 0.0 | 1 | 17920.030817 |

Now we will prepare the dataset that will be used for our modeling.

For each neighborhood, we will aggregate:

1. Count of ALL restaurants.
2. Count of Italian restaurants.
3. Average rating of Italian restaurants.
4. Also, we will include the distance to the City Center.

## Modeling

We will use the **K-Means** algorithm to cluster our neighborhood into 5 clusters.

After clustering the dataset, we get the below result:

| | Neighbourhood | Restaurant_Count | Distance_To_Center | Italian_Count | Average_Ratings | Labels |
|---|---|---|---|---|---|---|
| 0 | Barking and Dagenham | 12 | 17920.030817 | 0.0 | 0.000000 | 0 |
| 22 | Lewisham | 15 | 9257.906442 | 0.0 | 0.000000 | 0 |
| 25 | Redbridge | 18 | 12512.433585 | 0.0 | 0.000000 | 0 |
| 3 | Brent | 17 | 14003.045824 | 0.0 | 0.000000 | 0 |
| 4 | Bromley | 17 | 14654.616692 | 2.0 | 0.000000 | 0 |
| 5 | Camden | 24 | 2756.412095 | 0.0 | 0.000000 | 0 |
| 7 | Croydon | 17 | 16036.658074 | 0.0 | 0.000000 | 0 |
| 8 | Ealing | 19 | 15045.632474 | 0.0 | 0.000000 | 0 |
| 23 | Merton | 16 | 14594.407020 | 0.0 | 0.000000 | 0 |
| 14 | Harrow | 20 | 18735.296447 | 0.0 | 0.000000 | 0 |
| 20 | Kingston upon Thames | 16 | 19061.098358 | 2.0 | 3.650000 | 1 |
| 26 | Richmond upon Thames | 14 | 17898.540544 | 1.0 | 7.300000 | 1 |
| 17 | Hounslow | 21 | 19686.870384 | 1.0 | 7.300000 | 1 |
| 31 | Wandsworth | 19 | 9481.306493 | 2.0 | 8.200000 | 1 |
| 15 | Havering | 11 | 20486.772699 | 1.0 | 6.700000 | 1 |
| 12 | Hammersmith and Fulham | 17 | 10160.583093 | 1.0 | 8.400000 | 1 |
| 11 | Hackney | 11 | 4162.847895 | 1.0 | 8.100000 | 1 |
| 10 | Greenwich | 15 | 11287.461679 | 1.0 | 6.700000 | 1 |
| 9 | Enfield | 17 | 15410.781065 | 2.0 | 7.300000 | 1 |
| 30 | Waltham Forest | 18 | 10002.976870 | 3.0 | 7.800000 | 1 |
| 1 | Barnet | 17 | 12883.465693 | 1.0 | 8.400000 | 1 |
| 13 | Haringey | 16 | 9500.135755 | 1.0 | 8.400000 | 1 |
| 24 | Newham | 19 | 9695.637179 | 2.0 | 6.550000 | 1 |
| 27 | Southwark | 23 | 1566.402203 | 2.0 | 8.900000 | 2 |
| 29 | Tower Hamlets | 20 | 6023.360469 | 4.0 | 7.425000 | 2 |
| 32 | Westminster | 21 | 3723.168447 | 2.0 | 8.850000 | 2 |
| 19 | Kensington and Chelsea | 23 | 7273.068499 | 2.0 | 8.850000 | 2 |
| 18 | Islington | 19 | 2985.616195 | 2.0 | 8.100000 | 2 |
| 6 | City of London | 27 | 0.000000 | 4.0 | 8.500000 | 2 |
| 21 | Lambeth | 20 | 6322.539825 | 2.0 | 7.650000 | 2 |
| 28 | Sutton | 14 | 18520.635266 | 6.0 | 5.766667 | 3 |
| 2 | Bexley | 19 | 18156.865207 | 5.0 | 5.500000 | 3 |
| 16 | Hillingdon | 26 | 26823.258913 | 1.0 | 0.000000 | 4 |

## Results and Discussion

As we can see from the results of our Analysis, London neighborhoods of **Labels 0** are the most interesting neighborhoods to open new Italian restaurant because they have the least Italian restaurants with low rating.

Let's check the neighborhoods in **Cluster 0** and order the result based on the **Distance to the City Center**.

|  | Neighbourhood | Restaurant_Count | Distance_To_Center | Italian_Count | Average_Ratings | Labels |
|---|---|---|---|---|---|---|
| 5 | Camden | 24 | 2756.412095 | 0.0 | 0.0 | 0 |
| 22 | Lewisham | 15 | 9257.906442 | 0.0 | 0.0 | 0 |
| 25 | Redbridge | 18 | 12512.433585 | 0.0 | 0.0 | 0 |
| 3 | Brent | 17 | 14003.045824 | 0.0 | 0.0 | 0 |
| 23 | Merton | 16 | 14594.407020 | 0.0 | 0.0 | 0 |
| 4 | Bromley | 17 | 14654.616692 | 2.0 | 0.0 | 0 |
| 8 | Ealing | 19 | 15045.632474 | 0.0 | 0.0 | 0 |
| 7 | Croydon | 17 | 16036.658074 | 0.0 | 0.0 | 0 |
| 0 | Barking and Dagenham | 12 | 17920.030817 | 0.0 | 0.0 | 0 |
| 14 | Harrow | 20 | 18735.296447 | 0.0 | 0.0 | 0 |

As we can see, neighborhood *Camden* is the nearest neighborhood to the City Center. However, it has many restaurants which might be a downside because of competition for other restaurants serving other cuisines.

Second in the list is neighborhood *Lewisham* which might be better option as it has a smaller number of restaurants despite being little far from the City Center compared to **Camden**.

## Conclusion

The objective of this report is to identify the neighborhoods in London with least number of restaurants especially Italian restaurants and close to the city center. This should help the report's stakeholders to narrow down the neighborhoods to search for opening a new Italian restaurant.

In order to achieve this objective, we gathered geo-data of London neighborhoods and used Foursquare endpoints to gather information about the restaurants in each neighborhood. Also, we used maps visualization to help the user with the data exploration stage.

Based on the neighborhoods clustering we did, we provided 2 recommendations to the end-user to choose from.

For sure, there are many factors that might influence the end-user decision (i.e., other business in the area, easy access and transportation, etc.) which is out of the scope of this report.