

# The Battle of the Neighborhoods

Course 9: Applied Data Science Capstone by IBM/Coursera

# Introduction: Business Problem

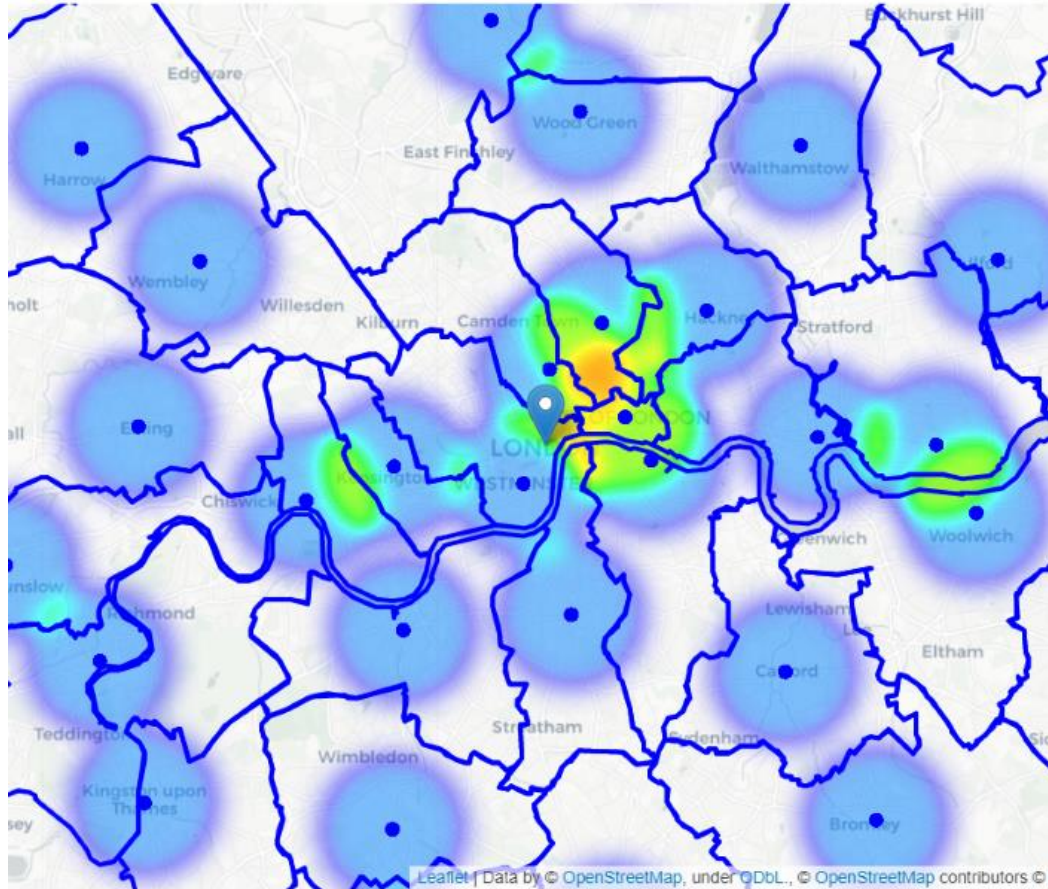
- ▶ In this project, we will try to build a model to find the optimal neighborhood for opening a new business. As an example, we will specify the business type to be an **Italian restaurant**.
- ▶ Since London is huge city with many restaurants, we will try to identify the optimal neighborhood to open 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**.

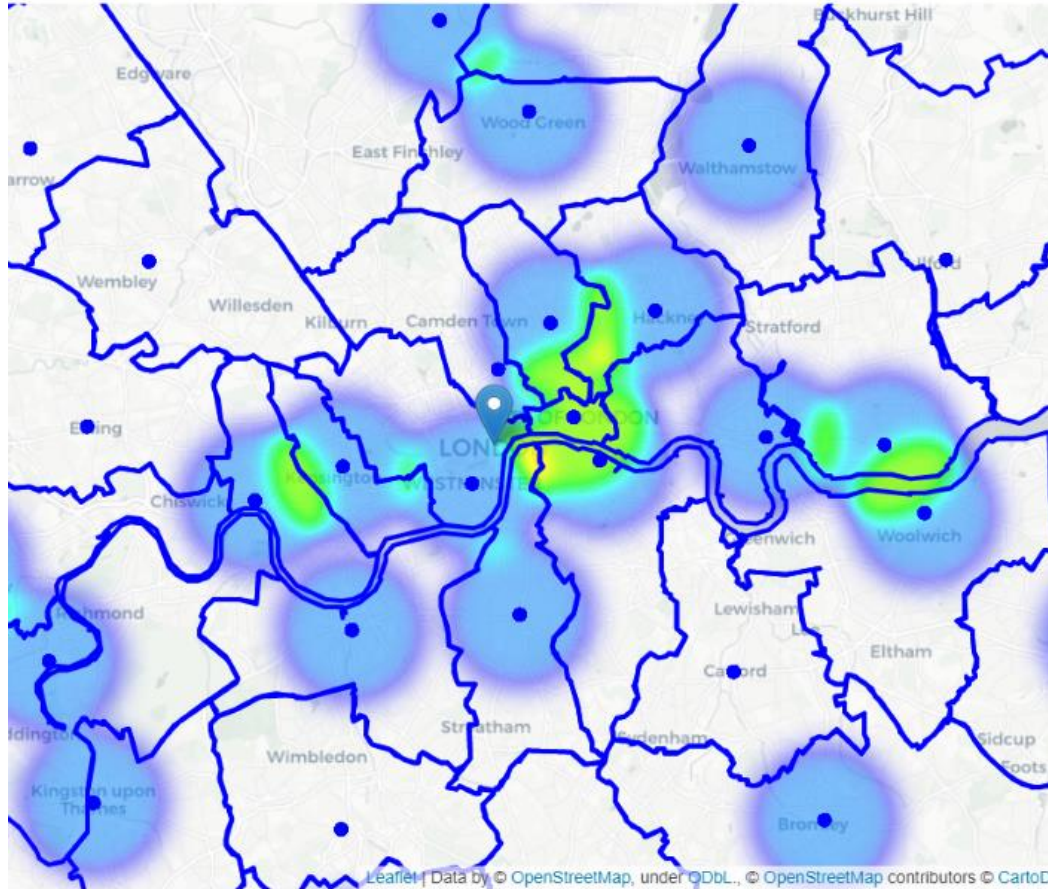
# Methodology

- ▶ 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.



# All Restaurants Count

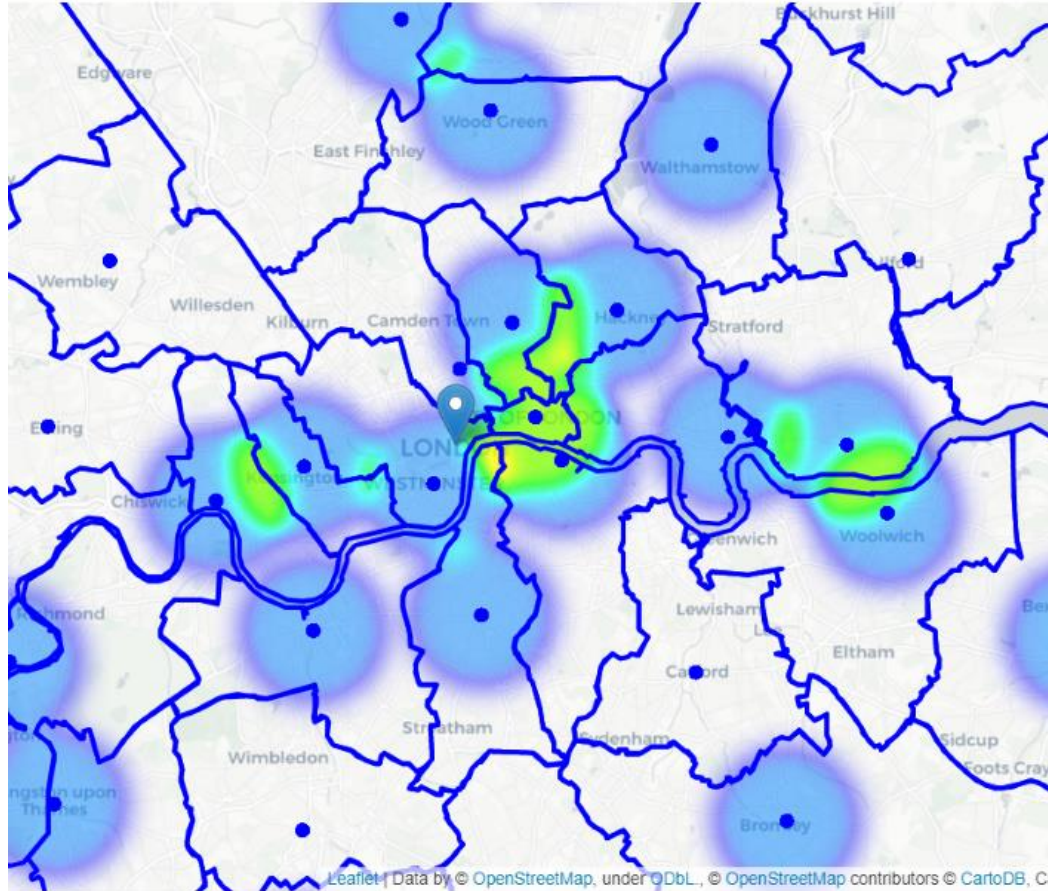
Here we can see the  
distribution of restaurants  
across London neighborhoods



# Italian Restaurants Count

Here we can see the distribution of Italian restaurants across London neighborhoods.





# Italian Restaurants Rating

The average rating of the  
Italian restaurants in each  
neighborhood.

# Modeling

- ▶ As we gathered the required data, we will apply unsupervised learning to group similar neighborhoods.
- ▶ We will use the ***K-Means*** algorithm to cluster our neighborhood into 5 clusters.
- ▶ This will help us identify which neighborhood(s) with less number of restaurants and especially Italian Restaurants.



# 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**.
- ▶ 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**.

	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

# 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.