# Opening a new Italian restaurant in London, UK

By Aria Kumar

## Introduction/Business Problem

London is one of the world's major cities and is a diverse multicultural hub with over 300 languages being spoken here. The city is world famous for its history, culture, attractions, art, shopping opportunities and food. There are around 15,500 restaurants in London according to the Office of National Statistics, making London one of the most ideal places in the world to open a restaurant. London hosts approximately 30 million tourists each year, so the demand for restaurants is always high and business will never be low if opened in the right place.

Londoners favourite cusine is Italian with approximately 30% saying it is their favourite food to eat out for. According to <https://www.prnewswire.com/news-releases/food-for-thought-the-average-brit-spends-ps700-on-eating-out-per-year-here-s-how-to-do-it-for-less-with-budget-meal-ideas-in-2019-826776701.html>

The objective of this project is to analyse and find the best location to open an Italian food restaurant in London.

The target audience for this project is a businessman who wants to invest in a location in London to open an Italian Restaurant.

## Data

Problem Statement: Where is the best location to open an Italian restaurant in London?

I will be getting data from following sources:

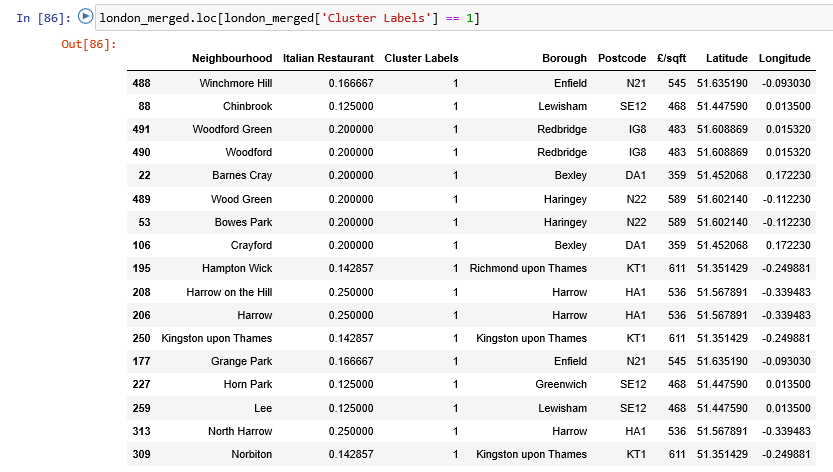
* The Neighbourhoods of London and their postcodes using Wikipedia (<https://en.wikipedia.org/wiki/List_of_areas_of_London>)
* The Locations in each Neighbourhood in London area (<https://developer.foursquare.com/>)
* The average price per square foot (<https://propertydata.co.uk/cities/london> )
* Coordinates of each Neighbourhood/postcode using geocoder python library (<https://developers.arcgis.com/python/guide/using-the-geocode-function/>)

**How will I use the data?**

1. Get the Neighbourhoods data from Wikipedia using Beautiful Soup library.
2. Explore the venues and their category in each Neighbourhood using Foursquare API.
3. Get the latitude and longitude data using Geocoder library.
4. Combine the data above into one dataframe.
5. Group by Neighbourhood and count the occurrences of Venues for each Neighbourhood.
6. Cluster each Neighbourhood based on the venues and add the cluster back into the initial dataframe
7. Identify patterns in each cluster and the most common venue.
8. Data visualisation will be done using Folium maps
9. Will use machine learning clustering techniques to hopefully get an ideal location

## Results

The results were categorised into 3 clusters. These clusters were based on the frequency of the occurrence of Italian Restaurants.

* ​Cluster 0: Neighbourhoods in London with a moderate number of Italian restaurants
* Cluster 1: Neighbourhoods in London with a low number of Italian restaurants
* Cluster 2: Neighbourhoods in London with a high number of Italian restaurants

​

## Discussion

A seen in the results London has a lot of Italian food restaurants and are located all around the city.

If someone wanted to open a new Italian food restaurant, any of the neighbourhoods in Cluster 1 (seen in the image above) would be the best suitable location.

If the customer wanted to extent their food chain restaurant, then cluster 0 would also be suitable. However the customer should avoid cluster 2 due to a high concentration of Italian restaurants already there.

## Conclusion

For this project I had to go through the process of finding a business problem and the data required. With that data and understanding, I had to clean extract and prepare the data so it was ready to perform machine learning by clustering on it. The data was clustered into 3 clusters based on their similarities.

I included the average price per square foot for the customer to consider when making a decision, as costs can sometimes be a limitation when deciding where to invest.

So in conclusion the customer is advised to invest in any of the areas in cluster 1 due to a lower concentration of Italian restaurants there/