Grouping Boroughs of London based food category

Rohit Biranje

August 2020

Table of contents:

Contents

[1. Introduction 3](#_Toc48654480)

[1.1 Background: 3](#_Toc48654481)

[1.2 Problem: 3](#_Toc48654482)

[1.3 Interest: 3](#_Toc48654483)

[2. Data acquisition and cleaning 3](#_Toc48654484)

[2.1 Data Sources: 3](#_Toc48654485)

[2.2 Data Cleaning: 3](#_Toc48654486)

[3. Methodology 3](#_Toc48654487)

[4. Results 5](#_Toc48654488)

[5. Discussion 7](#_Toc48654489)

[6. Conclusion 7](#_Toc48654490)

[7. References: 7](#_Toc48654491)

# Introduction

## 1.1 Background:

I have been living in London for the past four years and have found that London’s food scene very diverse. There is a wide range of restaurants with varying categories like Indian, Mexican, American, Fast Food, Cafes etc and different types from fine dining to street food. There are many new restaurants opening in London each year, some do well and some not so well.

## 1.2 Problem:

It’s important for restaurant owners to know the specific category and cuisine to target when opening a restaurant in specific Borough or updating menu of existing restaurant to make sure more people visit.

The aim of this project is to group Boroughs of London based on food category to help new / existing restaurant owners.

## 1.3 Interest:

Obviously, new restaurant owners will be able to take advantage of this analysis to determine if they should open restaurant with specific cuisine in a Borough looking at current popularity. Similarly, existing restaurant owners will be able to use this data to improvise or update their menu.

This analysis can also help tourists who have specific taste in food to target a Borough on their visit to London

# Data acquisition and cleaning

## 2.1 Data Sources:

Wikipedia will be used to get list of London Borough’s and their latitude and longitude details(<https://en.wikipedia.org/wiki/List_of_London_boroughs>).

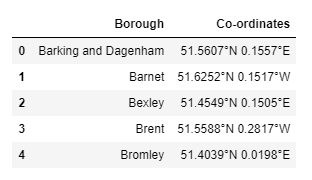
Foursquare APIs Venues Search was used to 100 most popular food venues per Borough.

## 2.2 Data Cleaning:

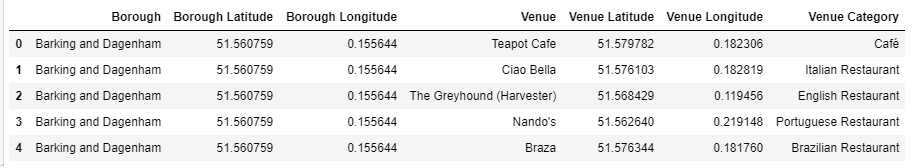
To start off, I decided to check the data for 1 Borough to get an idea of the cleansing that will be required. Then the data was downloaded / scraped from above mentioned sources. Unnecessary columns were dropped from the data. Formatting issues with the Borough names was performed so that I have consistent data.

# Methodology

I utilized pandas read\_csv to extract data from Wiki page. Unnecessary columns were dropped and data was formatted so that this can be passed on the Foursquare API. Here is head of the data that was created:



Using the Foursquare explore API I extracted top 100 food venues in 5000m radius from the Borough’s office. To the API, category id for Food (4d4b7105d754a06374d81259) was passed along with “sortByPopularity = 1” so that 100 most popular food venues were returned for each Borough. Here is head of the data that was created:

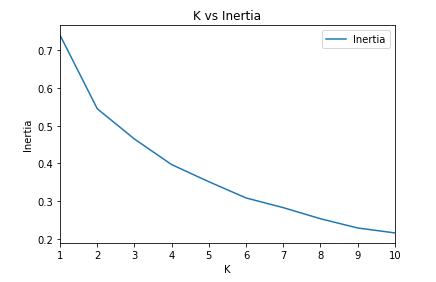


In summary there are 32 London Boroughs and approx. 3000 food venues were returned (i.e. there were Boroughs that did not provide 100 food venues).

There were 108 distinct food categories based on the data extracted and top10 food categories were extracted for each Borough. Here is head of the data that was created:

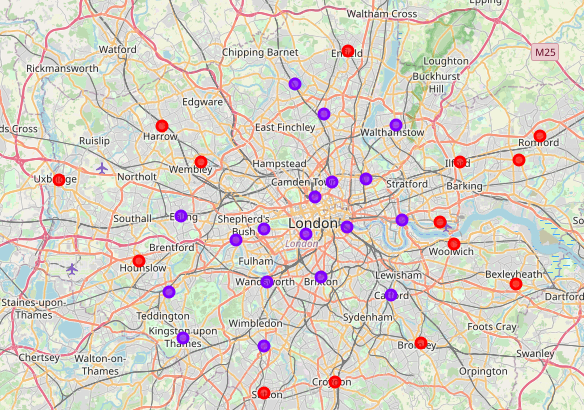


There were some common venue categories in boroughs and hence I used unsupervised learning K-means algorithm to cluster the boroughs. K-Means algorithm is one of the most common cluster methods of unsupervised learning.

I analysed impact of number of clusters on the dataset and found that I can segment the data into 2 clusters (using Elbow method) as there was significant drop in inertia at k=2.

# Results

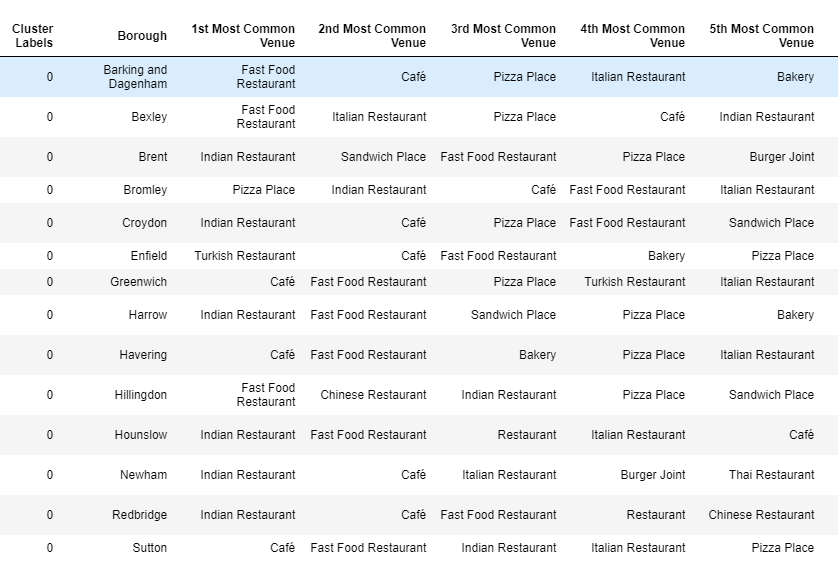
I used python folium library to visualize geographic details of London and its boroughs and I created a map of London with boroughs superimposed on top based on clusters:



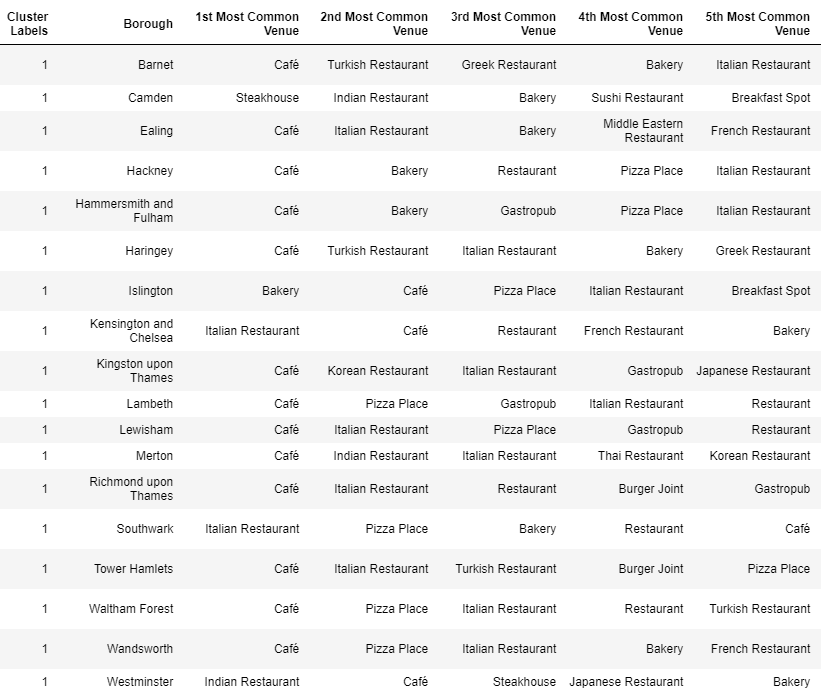
Based on this analysis it seems that the Boroughs near London Main city fall under one group(**cluster 1**) and the Borough’s that are further away fall under another group (**cluster 0**) based on food category.

Analysing the clusters further it was evident that coffee shops and bakeries are more popular in near the city center (cluster 0) whereas Fast Food Restaurants are more popular as we move away from city center(cluster1). Here is the data to support the same:

Cluster 0:



Cluster 1:



# Discussion

The analysis seems to suggest that coffee shops are more popular which may be because people travel from different Boroughs into London City for work and rely on coffee shops for their dose of coffee and quick snacks.

As there is such a complexity, very different approaches can be tried in clustering and classification studies. Moreover, it is obvious that not every classification method can yield the same high-quality results for London. The details were analysed mainly on popularity provided by Foursquare and these may vary considering day of week and recent conditions (like COVID-19).

# Conclusion

For a potential new restaurant owner opening a coffee shop near the city centre gives them a good chance of attracting people in the restaurant. Adding coffee (even take away) may be something that existing restaurant owners can look at if they are close to city centre.

# References:

[**Wiki**](https://en.wikipedia.org/wiki/List_of_London_boroughs)

[**Foursquare**](https://developer.foursquare.com/)