Capstone Project - The Battle of Neighborhoods Report

on

"London vs. New York? What is your choice? Let the data decide!"

Submitted By:

Chirag Sable

June 05, 2020

A) <u>Introduction</u>

1) <u>Background:</u>

London is the capital and the largest city of England and the United Kingdom. Standing on the River Thames in the south-east of England, London has been a major settlement for two millennia. London is one of the leading global cities in art, education, entertainment, finance, commerce, and media. It is considered the largest financial center in the world and among the top five largest metropolitan areas by GDP in the world. It also receives the highest number of international visitors of any city in the world. London is also a preferred investment destination, boasting more international businesses and billionaires than any other city in the world. London has a diverse population, and over 300 languages are spoken in the region.

New York City (**NYC**), often called **The City** or simply **New York** (**NY**), is the most populous city in the United States. It is diverse and the financial capital of the USA. It is multicultural. It provides a lot of business opportunities and a business friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce. The city is a major center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in the United States.

2) <u>Problem Description:</u>

NYC and **London** are the two most important metro cities and the financial centers of their respective country in the world. So, there is a competition among the cities as in which city is the most perfect city. Also the tourism sector of both these cities' is very strong. There has been a war for supremacy in terms of quality of

life, jobs, education, entertainment and recreational facilities that these cities have to offer to its residents and also for the tourists.

The deciding factor for most would be on how lively, supportive, vibrant and unique each of the cities can be when compared to each other. Now, say you are a non-resident of these cities'. Some day you plan to visit these cities. It is quite obvious that if you are planning to visit any new city you would first consider what does this city has to offer me? Alright. So, you would like to know about the famous places to visit, the best foods to try, etc. This is the problem which will be addressed in this project wherein we will be comparing the two cities (NYC and London) and determine how similar or dissimilar they are. This project attempts to analyze the neighborhoods in each of these two cities and tries to understand what is popular in them and what they have to offer to someone who is contemplating to make a choice on visiting or seeking a life in either of these metro cities.

3) **<u>Target Audience:</u>**

The business problem in this study assumes that people who would be interested in this study are those who would like to create a projection of potential life and activities in these metro city neighborhoods if the subject moves to either visit or live in one of them. The decision to choose one over the other would depend on popular venues in the neighborhoods in each of these metro cities.

B) **Data Section**

1) **Data Acquisition:**

For any "data science project" data is of paramount importance. For this study, we will need data about the neighborhoods in each of these metro cities.

Cities which will be analysed in this project: **London** and **New York City**.

- ***Dataset 1 (London Dataset):***

  The London Area consists of 32 Boroughs and the "City of London". The dataset is not readily available on the web. Instead there is a Wikipedia page that exists that has all the information we need to explore and cluster the neighborhoods. Hence, it will be required to scrape the Wikipedia page and wrangle the data, clean it, and then read it into a pandas dataframe so that it is in a structured format.

  The link for the Wikipedia page is as follows: https://en.wikipedia.org/wiki/List_of_areas_of_London

  Now, the data in the Wikipedia page has the following columns: Location, London Borough, Post town, Postcode district, Dial code and OS grid ref. We won't be requiring the complete list of columns. The columns which would be needed are:

  - **Location**
  - **London Borough**

- ***Dataset 2 (New York Dataset):***

  The New York neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the

neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

This dataset exists for free on the web. Link to the dataset is: *https://geo.nyu.edu/catalog/nyu_2451_34572* . For convenience, the data is downloaded and the files are placed on the server. The following link has it: *https://cocl.us/new_york_dataset*

The data present on the website is a **JSON** format data consisting of the attributes like type, id, coordinates (latitudes and longitudes), name, borough, neighborhood, stacked, annoline 1, annoline 2, annoline 3, etc. Since it contains many attributes we would not be extracting all of them. The attributes which we would be using/extracting from the dataset will be:

- **Borough**
- **Neighborhood**
- **Latitude**
- **Longitude**

2) **Data Cleaning:**

Dataset 1 (London):

For scraping the Wikipedia page we will use the ***Urllib*** and ***Beautiful Soup packages***.

To clean the HTML file, locate the 'table' tag in the Wikipedia page and extract it. After doing the extraction the data will look like the following:

	PostalCode	Borough	Neighborhood
0	SE2	Bexley, Greenwich	Abbey Wood
1	W3, W4	Ealing, Hammersmith and Fulham	Acton
2	CR0	Croydon	Addington
3	CR0	Croydon	Addiscombe
4	DA5, DA14	Bexley	Albany Park

Now, the dataset has rows which has the same Postal Code and Neighborhood which needs to be combined. Also, for multiple postcodes in one column, we need to split the postcodes to multiple rows and assign the same value from the other columns. After performing the above procedure our dataset *df0* will look like the following:

	PostalCode	Borough	Neighborhood
0	BR1	Bromley	Bromley, Plaistow, Sundridge, Widmore
1	BR1	Lewisham	Downham
2	BR2	Bromley	Hayes, Keston, Leaves Green, Southborough
3	BR3	Bromley	Bickley, Bromley Common, Eden Park, Elmers End
4	BR3	Bromley	Beckenham
5	SE20	Bromley	Beckenham
6	BR4	Bromley	Coney Hall, West Wickham
7	BR5	Bromley	Derry Downs, Petts Wood, St Mary Cray, St Paul...
8	BR5	Bromley	Orpington
9	BR6	Bromley	Orpington

The latitudes and longitudes are not readily available on the Wikipedia page. So we need to generate it. In order to obtain the location data of the locations, the *Geocoder* package is used with the *arcgis_geocoder* to obtain the latitude and longitude of the needed locations.

Initially we will create a function named *get_latlng* for generating the latitudes and longitudes using *arcgis_geocoder* and during the call of the function we

will pass the column 'PostalCode' during the function call. The output dataframe is shown below:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	BR1	Bromley	Bromley, Plaistow, Sundridge, Widmore	51.416710	0.009042
1	BR1	Lewisham	Downham	51.416710	0.009042
2	BR2	Bromley	Hayes, Keston, Leaves Green, Southborough	51.506420	-0.127210
3	BR3	Bromley	Bickley, Bromley Common, Eden Park, Elmers End	51.415095	-0.035403
4	BR3	Bromley	Beckenham	51.415095	-0.035403
5	SE20	Bromley	Beckenham	51.410090	-0.056830
6	BR4	Bromley	Coney Hall, West Wickham	51.506420	-0.127210
7	BR5	Bromley	Derry Downs, Petts Wood, St Mary Cray, St Paul...	51.506420	-0.127210
8	BR5	Bromley	Orpington	51.506420	-0.127210
9	BR6	Bromley	Orpington	51.506420	-0.127210

Size of dataframe *df_london_loc* is calculated by *df_london_loc.shape* which is given by (418, 5)

Dataset 2 (New York):

To prepare a dataframe we will loop through the JSON data and store every parameter into their respective cells. The dataframe will look like the following:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Size of dataframe *neighborhoods* is calculated by *neighborhoods.shape* viz (306,4).

C) <u>Methodology</u>

So, now we have both the datasets ready. We will now visualize the neighborhoods for both the cities on a *Folium* map.

❏ London neighborhoods overview
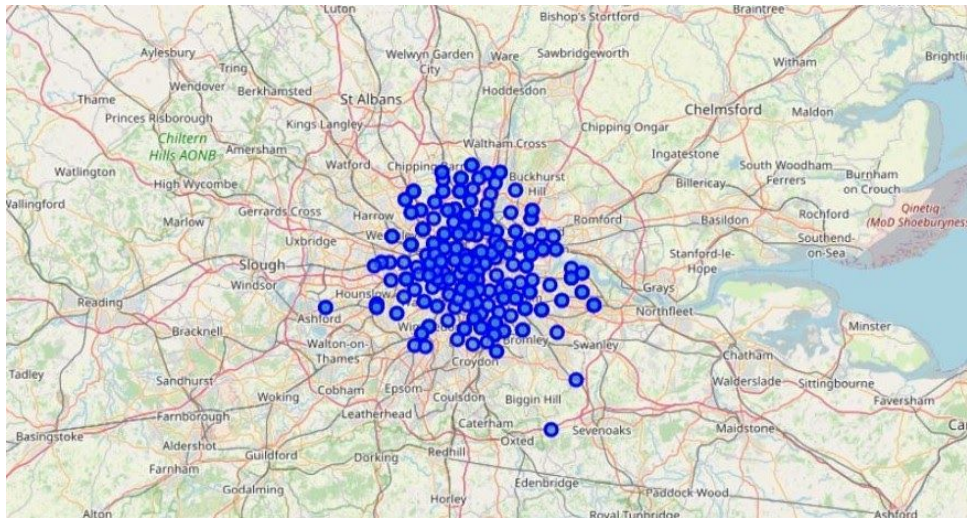
❏ New York neighborhoods overview

The Foursquare API will be used to obtain the neighborhood venues for the geographical location data . These will be used to explore the neighbourhoods of London and New York accordingly.

To use the **Foursquare API**, we have to define the **Foursquare Credentials and Version** in the following format:

CLIENT_ID = 'your Foursquare ID'
CLIENT_SECRET = 'your Foursquare Secret'
VERSION = '20180605'

Exploring and Analysing Neighborhoods:

● *Finding top venues near London Neighborhoods:*

We will use the Foursquare API to find the top venues in the neighbourhoods of London. This will help us understand the nature of life London neighborhoods have to offer. We will iteratively make Foursquare API calls for each of the London neighborhoods in our dataset.

The limit is designed as **100 venues** and the radius **2000 meters** for each borough from their given latitude and longitude information.

Next, we will employ statistical and analytical methods to find the unique venues/venue categories in the London neighborhoods and we will build a dataframe that calibrates each of the neighborhoods with the frequency of occurrence for each of the venue categories.

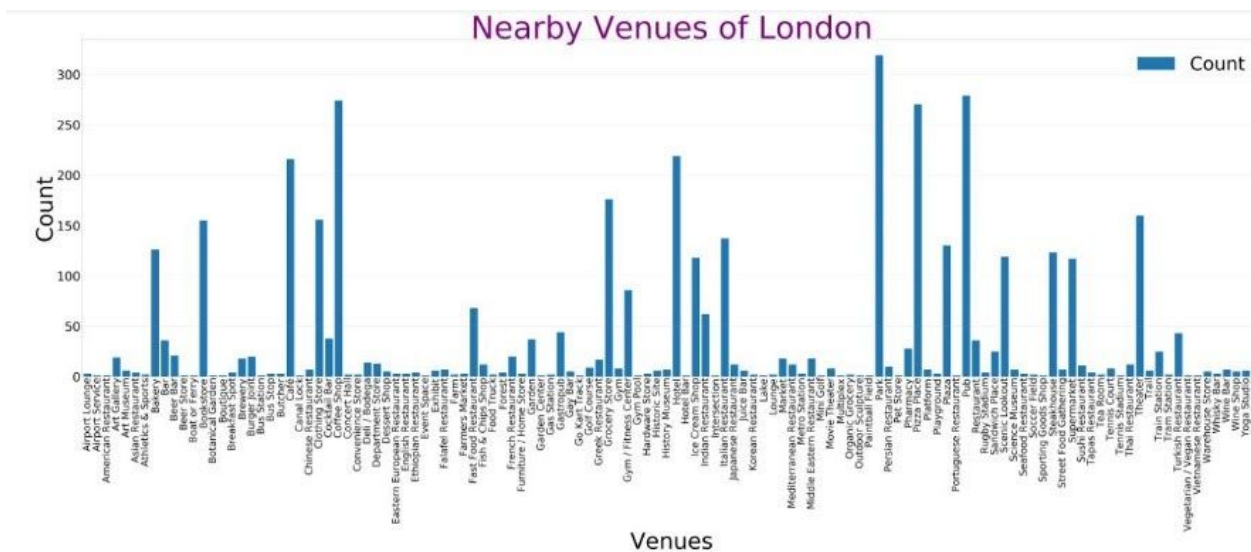|	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bromley, Plaistow, Sundridge, Widmore	51.41671	0.009042	Bromley Picturehouse	51.406740	0.012303	Indie Movie Theater
1	Bromley, Plaistow, Sundridge, Widmore	51.41671	0.009042	Cinnamon Culture	51.414196	0.020883	Indian Restaurant
2	Bromley, Plaistow, Sundridge, Widmore	51.41671	0.009042	Barrel & Horn	51.405887	0.013701	Bar
3	Bromley, Plaistow, Sundridge, Widmore	51.41671	0.009042	Beckenham Place Park	51.418044	-0.012489	Park
4	Bromley, Plaistow, Sundridge, Widmore	51.41671	0.009042	Nando's	51.406128	0.016046	Portuguese Restaurant

From our analysis, **357** unique venue categories were returned in the London neighbourhoods. Pub, Cafe, Park, Hotel, Italian Restaurant, Coffee Shop, Grocery

Store, Theatre, Scenic Lookout, Ice Cream Shop, Bakery, Pizza Place, Clothing Store, Indian Restaurant, Gym/Fitness Center, Supermarket being some of them.

We then create a dataset that lists the top 10 common venues against each of the neighbourhoods in London. We get a representation such as below for all the neighbourhoods in London.

	Neighborhood	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	Abbey Wood	Supermarket	Grocery Store	Fast Food Restaurant	Train Station	Pharmacy	Hardware Store	Historic Site	Clothing Store	Warehouse Store	Gym / Fitness Center
1	Acton	Pub	Coffee Shop	Grocery Store	Park	Gym / Fitness Center	Café	Bakery	Gastropub	Middle Eastern Restaurant	Pizza Place
2	Addington, Addiscombe, Coombe, Croydon, Forest...	Grocery Store	Park	Pub	Gym / Fitness Center	Fast Food Restaurant	Coffee Shop	Chinese Restaurant	Bakery	Golf Course	Mini Golf
3	Albany Park	Pub	Hotel	Clothing Store	Theater	Grocery Store	Plaza	Pizza Place	Supermarket	Steakhouse	Coffee Shop
4	Aldborough Hatch, Gants Hill, Newbury Park	Hotel	Theater	Bookstore	Plaza	Steakhouse	Clothing Store	Scenic Lookout	Ice Cream Shop	Park	Pizza Place
5	Aldgate	Hotel	Coffee Shop	Art Gallery	Cocktail Bar	Gym / Fitness Center	Grocery Store	French Restaurant	Seafood Restaurant	Italian Restaurant	Theater
6	Aldwych, Charing Cross, Covent Garden	Coffee Shop	Theater	Hotel	History Museum	Steakhouse	Wine Bar	Pub	Ice Cream Shop	Bookstore	Restaurant
7	Alperton	Hotel	Theater	Bookstore	Plaza	Steakhouse	Clothing Store	Scenic Lookout	Ice Cream Shop	Park	Pizza Place
8	Anerley, Penge	Pub	Coffee Shop	Supermarket	Café	Pizza Place	Gym / Fitness Center	Park	Italian Restaurant	Grocery Store	Train Station
9	Angel	Coffee Shop	Pub	Café	Hotel	Cocktail Bar	Bakery	Art Gallery	Theater	Bookstore	Concert Hall

The most common venues around London has been visualized using the bar plot as follows:

Let's see the top 20 most common venues in London using a bar plot.
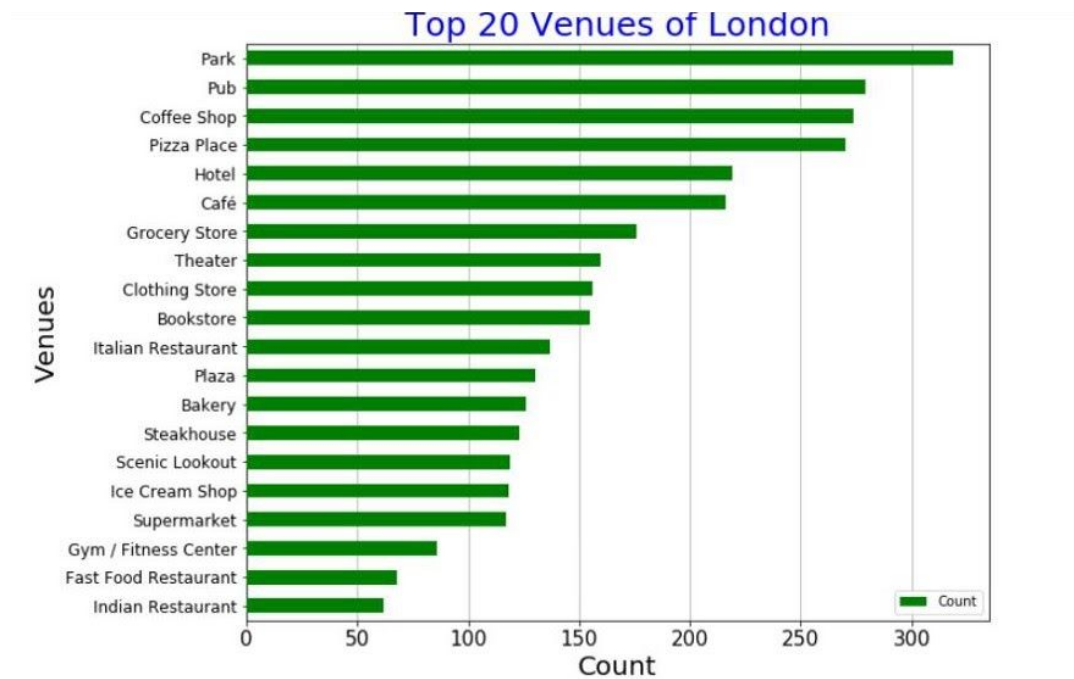
Cluster the neighbourhoods in <u>London</u> based on the similarity of top common venues:

For this, we will use the K-Means clustering algorithm. The reason for choosing this algorithm is that we have to group those neighborhoods which have similarity for the top common venues.

However, a drawback of k-means is that we have to specify the number of clusters, k, before we know what the optimal k is. The number of clusters to choose may not always be so obvious in real-world applications, especially if we are working with a higher dimensional dataset that cannot be visualized.

The **elbow method** is a useful graphical tool to estimate the optimal number of clusters k for a given task. Intuitively, we can say that, if k increases, the within-cluster SSE ("**distortion**") will decrease. This is because the samples will be closer to the centroids they are assigned to.

The idea behind the **elbow method** is to identify the value of k where the distortion begins to decrease most rapidly, which will become clearer if we plot the distortion for different values of k:

Elbow Method For Optimal k
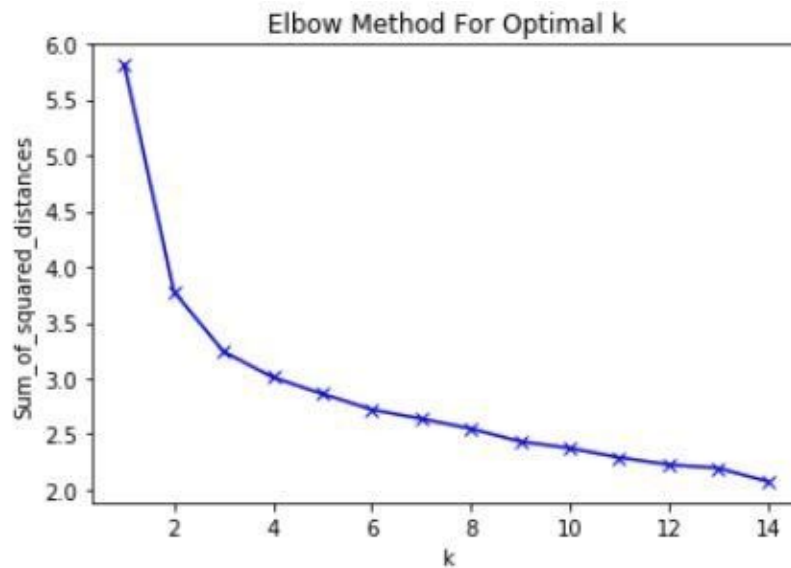
As we can see in the resulting plot, the elbow is located at $k = 4$, which is evidence that $k = 4$ is indeed a good choice for this dataset.

Now, we'll apply this to our dataset. Each of the neighborhoods gets a Cluster Label assigned.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	BR1	Bromley	Bromley, Plaistow, Sundridge, Widmore	51.416710	0.009042	1.0	Grocery Store	Pub	Coffee Shop	Clothing Store	Park	Café	Indian Restaurant	Pizza Place	Train Station	Burger Joint
1	BR1	Lewisham	Downham	51.416710	0.009042	1.0	Grocery Store	Pub	Coffee Shop	Clothing Store	Park	Café	Indian Restaurant	Pizza Place	Train Station	Burger Joint
2	BR2	Bromley	Hayes, Keston, Leaves Green, Southborough	51.506420	-0.127210	0.0	Hotel	Theater	Bookstore	Plaza	Steakhouse	Clothing Store	Scenic Lookout	Ice Cream Shop	Park	Pizza Place
3	BR3	Bromley	Bickley, Bromley Common, Eden Park, Elmers End	51.415095	-0.035403	1.0	Supermarket	Pub	Coffee Shop	Park	Café	Pizza Place	Grocery Store	Train Station	Pharmacy	Italian Restaurant

● *Finding top venues near New York Neighborhoods:*

We will use the Foursquare API to find the top venues in the neighbourhoods of New York.

The limit is designed as **100 venues** and the radius **2000 meters** for each borough from their given latitude and longitude information.

We will build a dataframe that calibrates each of the neighborhoods with the frequency of occurrence for each of the venue categories.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
2	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
3	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop

From our analysis, **426** unique venue categories were returned in the New York neighbourhoods. Pizza Place, Coffee Shop, Sandwich Place, Chinese Restaurant, Italian Restaurant, Donut Shop, Bar, River, Gym, Juice Bar, Bus Station, Baseball Field, Caribbean Restaurant being some of them.

We then create a dataset that lists the top 10 common venues against each of the neighbourhoods in New York. We get a representation such as below:

	Neighborhood	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	Allerton	Pizza Place	Chinese Restaurant	Deli / Bodega	Supermarket	Intersection	Fried Chicken Joint	Bus Station	Fast Food Restaurant	Check Cashing Service	Electronics Store
1	Annadale	American Restaurant	Pub	Pizza Place	Sports Bar	Food	Bakery	Train Station	Diner	Restaurant	Liquor Store
2	Arden Heights	Pharmacy	Deli / Bodega	Coffee Shop	Pizza Place	Women's Store	Farm	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant
3	Arlington	Intersection	Boat or Ferry	American Restaurant	Deli / Bodega	Fast Food Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Service	Event Space
4	Arrochar	Bus Stop	Italian Restaurant	Deli / Bodega	Liquor Store	Mediterranean Restaurant	Bagel Shop	Sandwich Place	Middle Eastern Restaurant	Pharmacy	Athletics & Sports

The most common venues around New York has been visualised using the bar plot as follows:
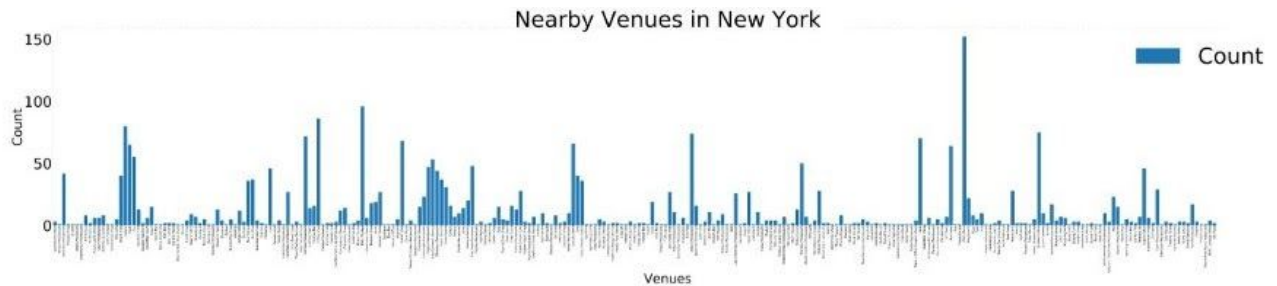
Let's see the top 20 most common venues in London using a bar plot.
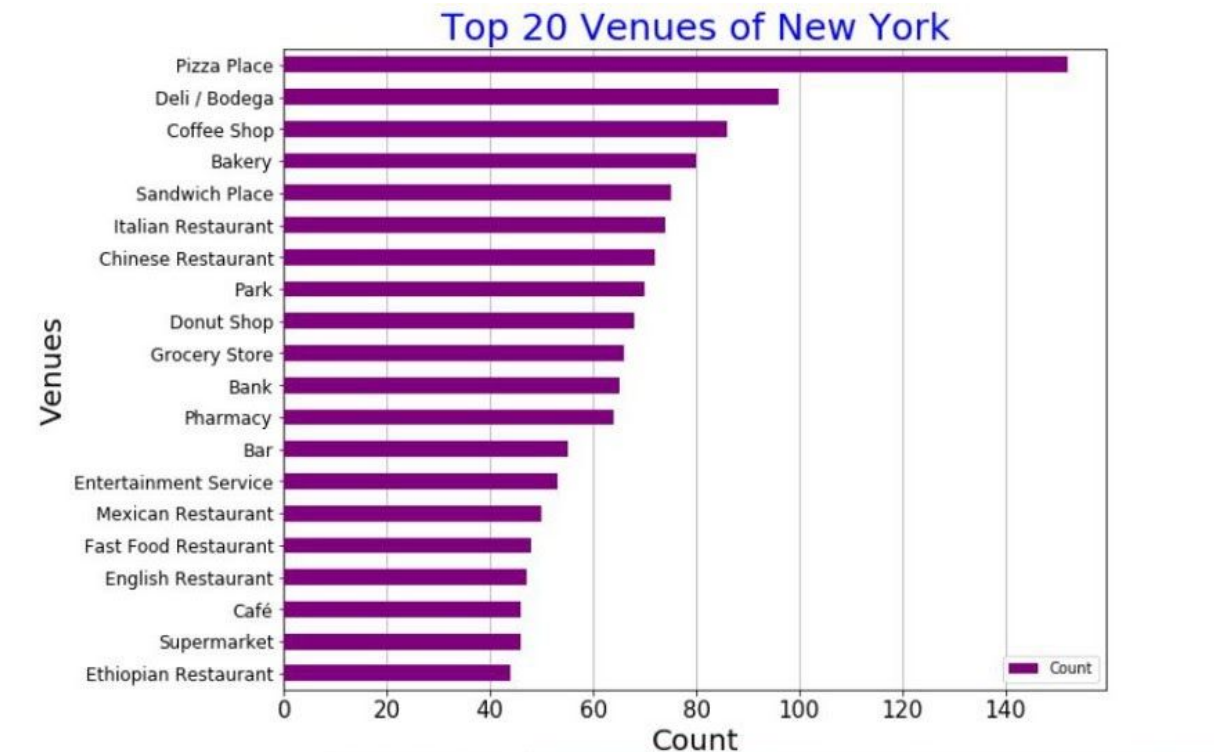
Cluster the neighbourhoods in <u>New York</u> based on the similarity of top common venues:

For this, we will use the K-Means clustering algorithm. The number of clusters have been set to 3.

Now, we'll apply the k-means algorithm to our dataset. Each of the neighborhoods gets a Cluster Label assigned. The dataset is as shown below:

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	Bronx	Wakefield	40.894705	-73.847201	1.0	Pharmacy	Laundromat	Deli / Bodega	Gas Station	Sandwich Place	Dessert Shop	Ice Cream Shop	Donut Shop	Factory	Women's Store
1	Bronx	Co-op City	40.874294	-73.829939	1.0	Fast Food Restaurant	Basketball Court	Park	Pharmacy	Restaurant	Discount Store	Mattress Store	Pizza Place	Bagel Shop	Bus Station
2	Bronx	Eastchester	40.887556	-73.827806	1.0	Caribbean Restaurant	Deli / Bodega	Bus Station	Diner	Intersection	Pizza Place	Fast Food Restaurant	Donut Shop	Bowling Alley	Juice Bar
3	Bronx	Fieldston	40.895437	-73.905643	1.0	Medical Supply Store	Plaza	Bus Station	River	Women's Store	Farm	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant
4	Bronx	Riverdale	40.890834	-73.912585	0.0	Park	Bus Station	Baseball Field	Gym	Plaza	Playground	Bank	Women's Store	Falafel Restaurant	English Restaurant

D) Result

Let us visualize the London dataset with cluster labels assigned in a *Folium* map.
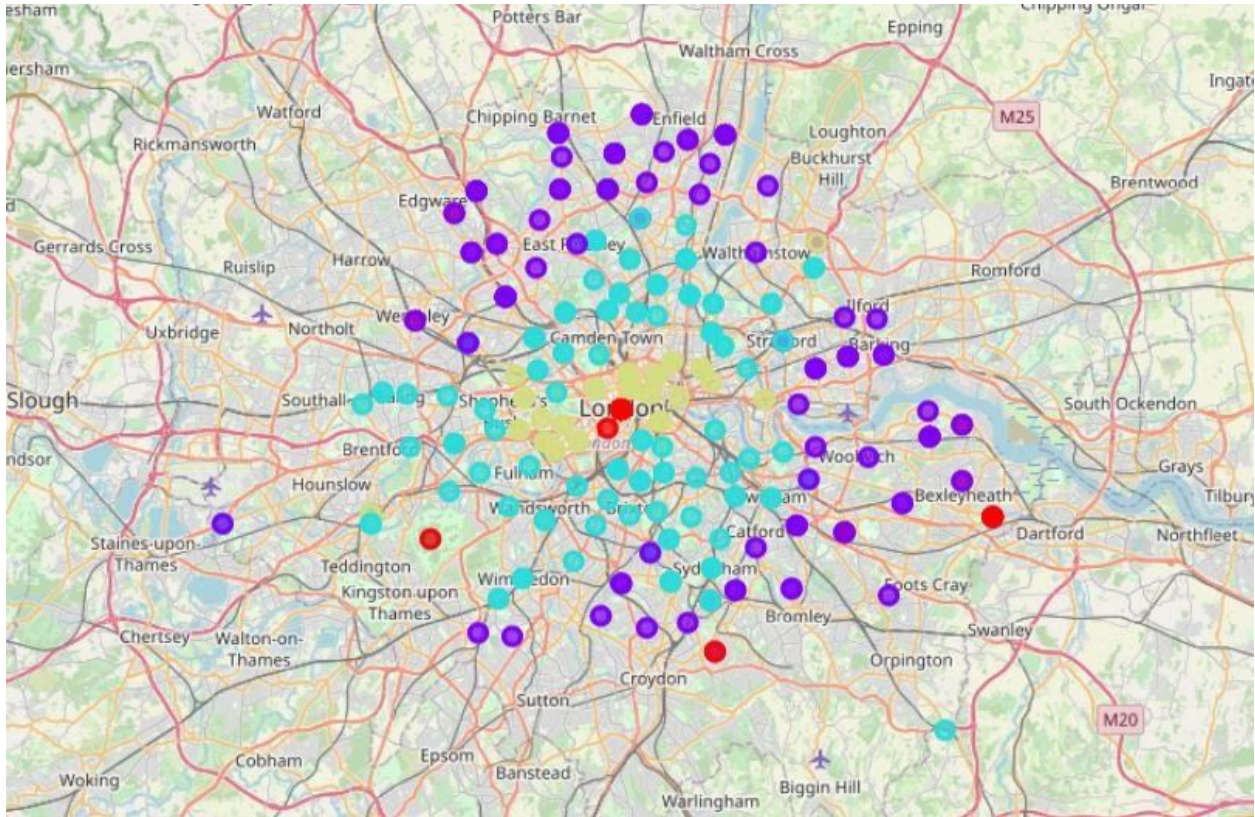
A piece of important information this map provides is that many neighborhoods in London are of similar nature concerning the venues they have around, indicated by the clusters marked in light blue, purple, red and yellow green.

Let us now dig a little deeper into how the neighborhoods are clustered and what is the characteristic of the cluster that is very common across most neighborhoods in London.

Cluster Label 0:

	Borough	PostalCode	Cluster Labels	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
2	Bromley	BR2	0.0	Hotel	Theater	Bookstore	Plaza	Steakhouse	Clothing Store	Scenic Lookout	Ice Cream Shop	Park	Pizza Place
6	Bromley	BR4	0.0	Hotel	Theater	Bookstore	Plaza	Steakhouse	Clothing Store	Scenic Lookout	Ice Cream Shop	Park	Pizza Place
7	Bromley	BR5	0.0	Hotel	Theater	Bookstore	Plaza	Steakhouse	Clothing Store	Scenic Lookout	Ice Cream Shop	Park	Pizza Place
8	Bromley	BR5	0.0	Hotel	Theater	Bookstore	Plaza	Steakhouse	Clothing Store	Scenic Lookout	Ice Cream Shop	Park	Pizza Place
9	Bromley	BR6	0.0	Hotel	Theater	Bookstore	Plaza	Steakhouse	Clothing Store	Scenic Lookout	Ice Cream Shop	Park	Pizza Place
10	Bromley	BR6	0.0	Hotel	Theater	Bookstore	Plaza	Steakhouse	Clothing Store	Scenic Lookout	Ice Cream Shop	Park	Pizza Place
13	Croydon	CR2	0.0	Hotel	Theater	Bookstore	Plaza	Steakhouse	Clothing Store	Scenic Lookout	Ice Cream Shop	Park	Pizza Place

The neighborhoods belonging to this cluster are popular for having Hotels, Theatre, Bookstore, Plaza, Clothing, Park, Pizza Place. We see that this neighborhood would be preferable for the one who wants different categories in a nearby place right from food to entertainment.

We can also estimate the number of **1st Most Common Venues** in this cluster.

Top Venues in Cluster 1 (London)
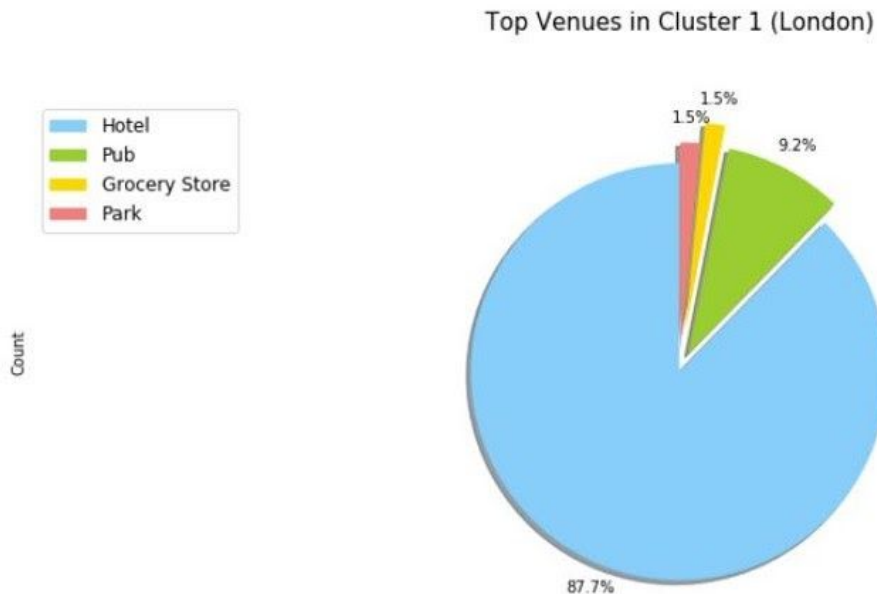
As the pie chart indicates, the topmost venue for this cluster is Hotel followed by Pub, Grocery Store and Park.

Cluster Label 1:

	Borough	PostalCode	Cluster Labels	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	Bromley	BR1	1.0	Grocery Store	Pub	Coffee Shop	Clothing Store	Park	Café	Indian Restaurant	Pizza Place	Train Station	Burger Joint
1	Lewisham	BR1	1.0	Grocery Store	Pub	Coffee Shop	Clothing Store	Park	Café	Indian Restaurant	Pizza Place	Train Station	Burger Joint
3	Bromley	BR3	1.0	Supermarket	Pub	Coffee Shop	Park	Café	Pizza Place	Grocery Store	Train Station	Pharmacy	Italian Restaurant
4	Bromley	BR3	1.0	Pub	Supermarket	Coffee Shop	Café	Park	Pizza Place	Grocery Store	Italian Restaurant	Train Station	Gastropub
5	Bromley	SE20	1.0	Pub	Supermarket	Coffee Shop	Café	Park	Pizza Place	Grocery Store	Italian Restaurant	Train Station	Gastropub
11	Bromley	BR7	1.0	Pub	Coffee Shop	Pizza Place	Grocery Store	Italian Restaurant	Fast Food Restaurant	Pharmacy	Indian Restaurant	French Restaurant	Fish & Chips Shop

The neighborhoods belonging to this cluster are popular for having Pub, Coffee Shop, Cafe, Park and Supermarket. This cluster will be more familiar for the ones who want to have some kind of refreshments and fast food. Also, it is for the one who wants to go shopping. The cluster is also home to train stations for travelling.

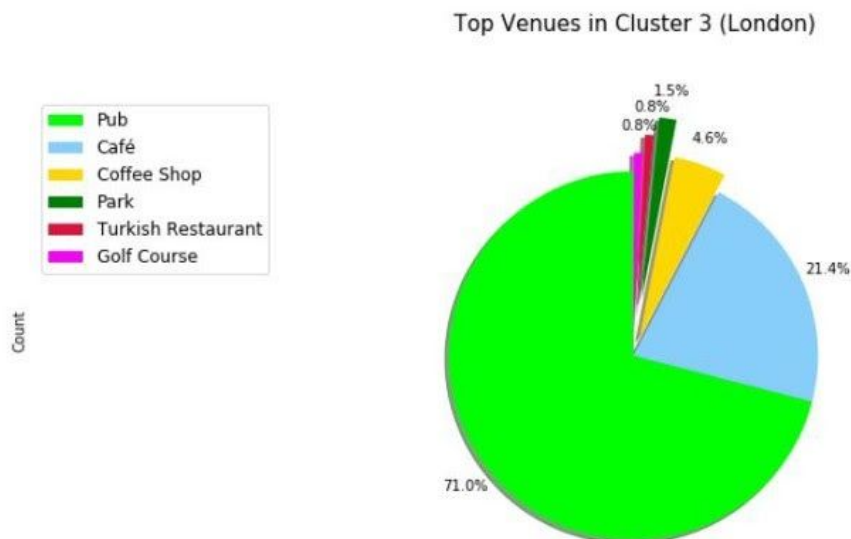Let's estimate the number of **1st Most Common Venues** in this cluster.

Top Venues in Cluster 2 (London)
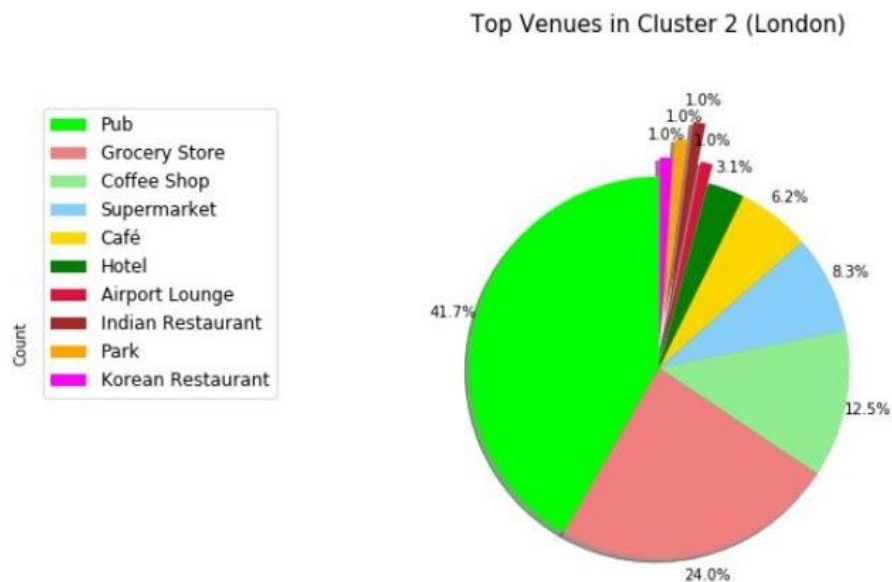
As the pie chart indicates, the topmost venue for this cluster is Pub followed by Grocery Store, Coffee Shop, Supermarket and others.

Cluster Label 2:

	Borough	PostalCode	Cluster Labels	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
41	Hackney	E10	2.0	Pub	Café	Coffee Shop	Park	Restaurant	Grocery Store	Bar	Italian Restaurant	Bakery	Juice Bar
42	Waltham Forest	E10	2.0	Pub	Café	Park	Coffee Shop	Grocery Store	Bar	Restaurant	Art Gallery	Fish & Chips Shop	Department Store
43	Waltham Forest	E15	2.0	Pub	Café	Park	Coffee Shop	Grocery Store	Bar	Restaurant	Art Gallery	Fish & Chips Shop	Department Store
44	Redbridge	E11	2.0	Pub	Café	Park	Mediterranean Restaurant	Pizza Place	Bakery	Grocery Store	Coffee Shop	Greek Restaurant	Turkish Restaurant
45	Redbridge, Waltham Forest	E11	2.0	Pub	Café	Park	Mediterranean Restaurant	Pizza Place	Bakery	Grocery Store	Coffee Shop	Greek Restaurant	Turkish Restaurant
46	Waltham Forest	E11	2.0	Pub	Café	Park	Mediterranean Restaurant	Pizza Place	Bakery	Grocery Store	Coffee Shop	Greek Restaurant	Turkish Restaurant

The neighborhoods belonging to this cluster are popular for having Pub, Cafe, Medierranean Restaurant, Greek and Turkish Restaurants. We see that this neighborhood would be preferable for the one who is foody as mediterranean food is also available here along with Greek and Turkish food. It is also for someone who loves golf.

Let's estimate the number of **1st Most Common Venues** in this cluster.

Top Venues in Cluster 3 (London)

As the pie chart indicates, the topmost venue for this cluster is Pub followed by Cafe, Coffee Shop, Park, Turkish Restaurant and Golf Course.

Cluster Label 3:

	Borough	PostalCode	Cluster Labels	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
40	Tower Hamlets	E1	3.0	Coffee Shop	Café	Pub	Italian Restaurant	Indian Restaurant	Hotel	Cocktail Bar	Restaurant	Middle Eastern Restaurant	Bar
51	Tower Hamlets	E14	3.0	Coffee Shop	Hotel	Pub	Park	Italian Restaurant	Burger Joint	Plaza	Gym / Fitness Center	Café	Bar
53	Newham	E16	3.0	Hotel	Coffee Shop	Pub	Grocery Store	Park	Bar	Lounge	Burger Joint	Hotel Bar	Chinese Restaurant
56	Hackney	E2	3.0	Coffee Shop	Pub	Bakery	Café	Bookstore	Pizza Place	Hotel	Beer Bar	Cocktail Bar	Indian Restaurant
57	Tower Hamlets	E2	3.0	Coffee Shop	Pub	Bakery	Café	Bookstore	Pizza Place	Hotel	Beer Bar	Cocktail Bar	Indian Restaurant
77	City	EC1	3.0	Coffee Shop	Pub	Hotel	Cocktail Bar	Falafel Restaurant	Café	Bookstore	Pizza Place	Whisky Bar	Beer Bar

The neighborhoods belonging to this cluster are popular for having Coffee Shop, Hotel, Pub, Cafe, Bakery, Italian Restaurants and others. This would be perfect for someone wanting certain refreshments along with food of Italian type. Also someone who loves cakes would be more familiar here.
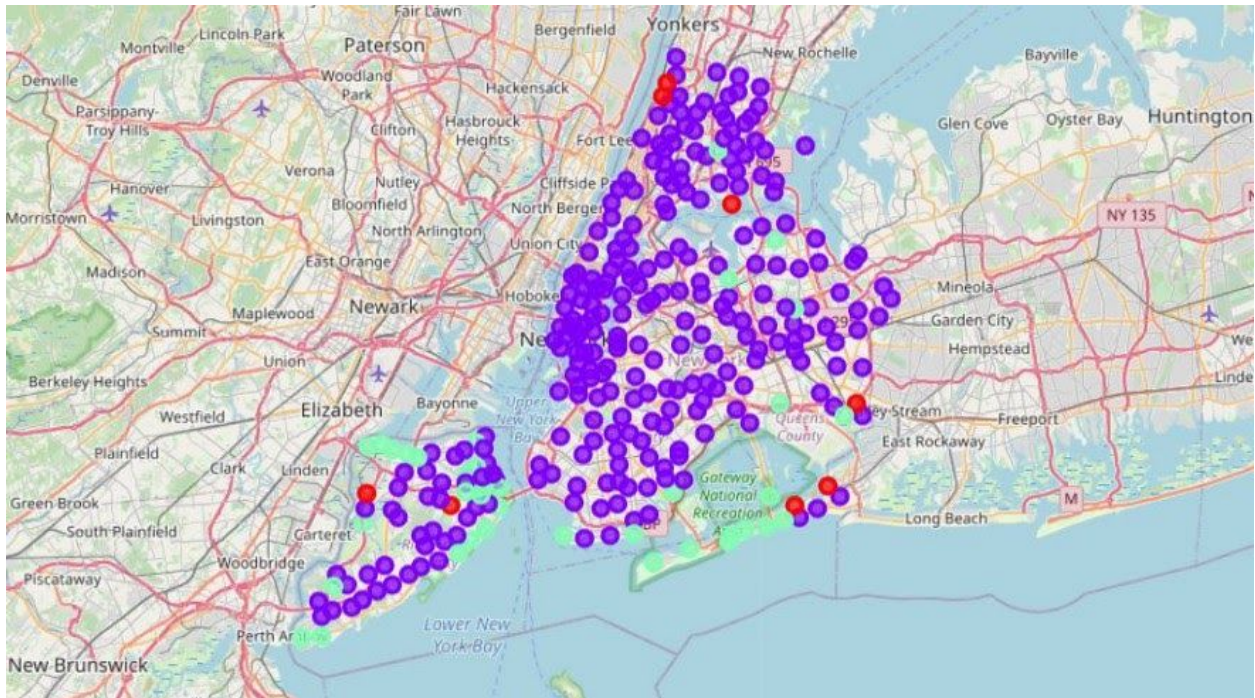
Let's estimate the number of **1st Most Common Venues** in this cluster.

Top Venues in Cluster 4 (London)
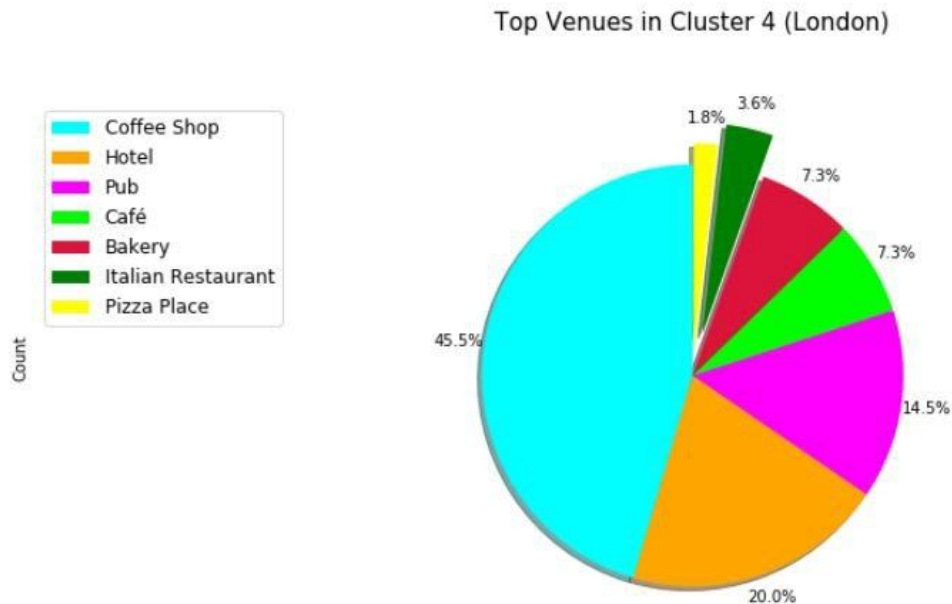
As the pie chart indicates, the topmost venue for this cluster is Coffee Shop followed by Hotel, Pub, Bakery and others.

Dataset 2 (New York):

Let us visualize the New York dataset with cluster labels assigned in a *Folium* map.

A piece of important information this map provides is that many neighborhoods in NYC are of similar nature concerning the venues they have around, indicated by the clusters marked in light blue, purple and red.

Let us now dig a little deeper into how the neighborhoods are clustered and what is the characteristic of the cluster that is very common across most neighborhoods in London.

Cluster Label 0:

	Neighborhood	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
4	Riverdale	Park	Bus Station	Baseball Field	Gym	Plaza	Playground	Bank	Women's Store	Falafel Restaurant	English Restaurant
27	Clason Point	Park	Home Service	Moving Target	Grocery Store	Pool	South American Restaurant	Boat or Ferry	Bus Stop	Exhibit	Falafel Restaurant
35	Spuyten Duyvil	Park	Bank	Thai Restaurant	Scenic Lookout	Tennis Stadium	Tennis Court	Pharmacy	Egyptian Restaurant	Electronics Store	Empanada Restaurant
188	Laurelton	Caribbean Restaurant	Cosmetics Shop	Train Station	Park	Farm	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Service
192	Somerville	Park	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Service	Event Space	Exhibit	Factory
203	Todt Hill	Park	Trail	Farm	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Service	Event Space
245	Bloomfield	Park	Discount Store	Theme Park	Recreation Center	Bus Stop	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Service
303	Bayswater	Park	Playground	Farmers Market	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Service	Event Space

The neighborhoods belonging to this cluster are popular for having Park, Bus Station, Baseball Field, Cosmetics Shop, Theme Park, Restaurants and many more. We see that this neighborhood would be preferable for the one who wants entertainment and some food. Also someone who wants to go shopping could fit here.

We can also estimate the number of **1st Most Common Venues** in this cluster with the help of a bar chart.
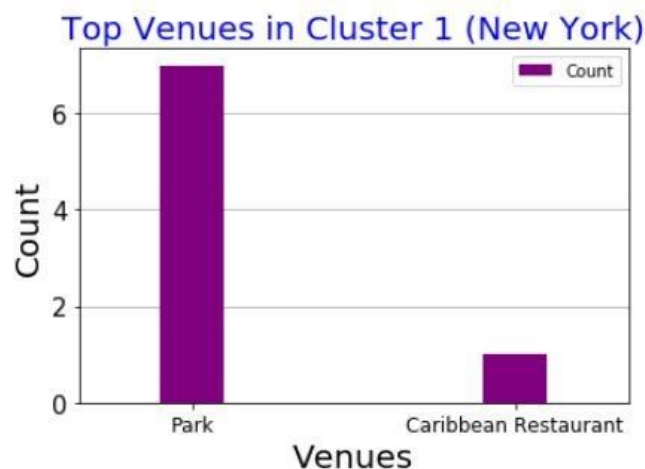
As the bar chart indicates, the topmost venue for this cluster is Park followed by Caribbean Restaurant.

Cluster Label 1:

	Neighborhood	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	Wakefield	Pharmacy	Laundromat	Deli / Bodega	Gas Station	Sandwich Place	Dessert Shop	Ice Cream Shop	Donut Shop	Factory	Women's Store
1	Co-op City	Fast Food Restaurant	Basketball Court	Park	Pharmacy	Restaurant	Discount Store	Mattress Store	Pizza Place	Bagel Shop	Bus Station
2	Eastchester	Caribbean Restaurant	Deli / Bodega	Bus Station	Diner	Intersection	Pizza Place	Fast Food Restaurant	Donut Shop	Bowling Alley	Juice Bar
3	Fieldston	Medical Supply Store	Plaza	Bus Station	River	Women's Store	Farm	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant
5	Kingsbridge	Pizza Place	Bar	Sandwich Place	Mexican Restaurant	Supermarket	Latin American Restaurant	Spanish Restaurant	Bakery	Donut Shop	Fried Chicken Joint
6	Marble Hill	Sandwich Place	Gym	Coffee Shop	Pharmacy	Steakhouse	Miscellaneous Shop	Supplement Shop	Shopping Mall	Seafood Restaurant	Tennis Stadium
7	Woodlawn	Pizza Place	Deli / Bodega	Food & Drink Shop	Pub	Bar	Playground	Food Truck	Grocery Store	Park	Donut Shop

The neighborhoods belonging to this cluster are popular for having Fast Food Restaurant, Pizza Place, Bar, Gym, Basketball court, Food and many more. This cluster will be more familiar for the ones who love pizza as you'll find here pizza shops more often. A foodie will definitely love this place.

Let's estimate the number of **1st Most Common Venues** in this cluster.
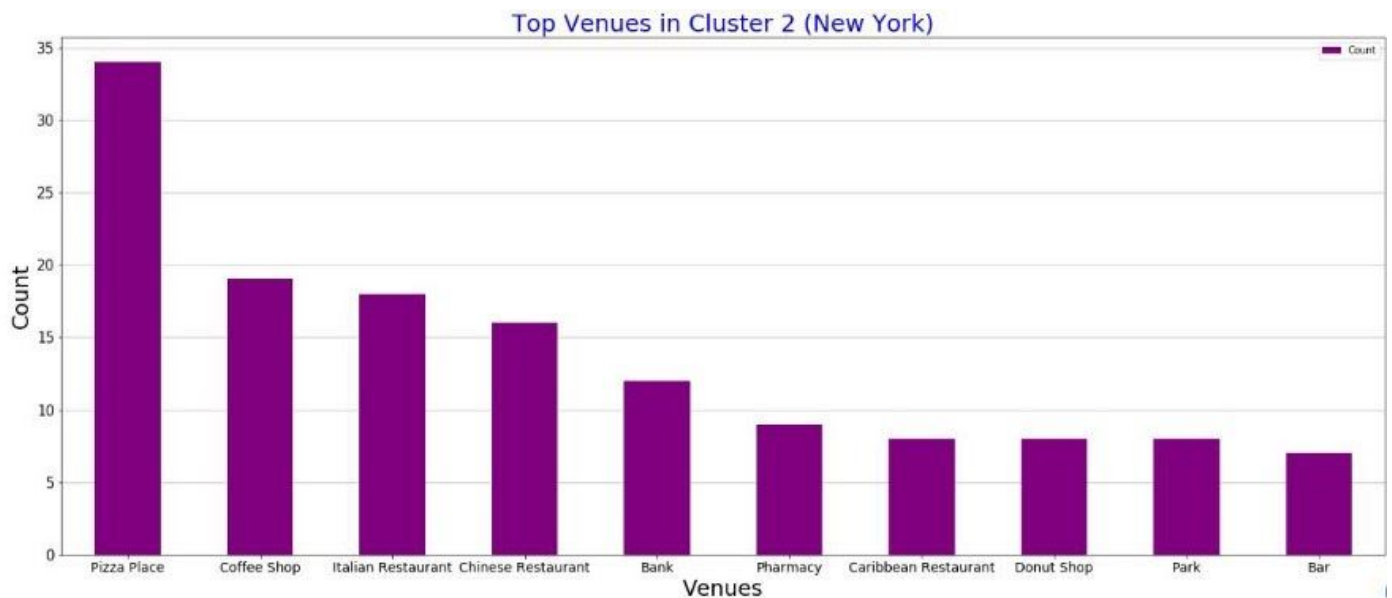
As the bar chart indicates, the topmost venue for this cluster is Pizza Place followed by Coffee Shop, Restaurants and more.

Cluster Label 2:

	Neighborhood	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
32	Van Nest	Pizza Place	Deli / Bodega	Middle Eastern Restaurant	Donut Shop	Hookah Bar	Discount Store	Bus Stop	Bus Station	Bakery	Supermarket
76	Mill Island	Pool	Women's Store	Fast Food Restaurant	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Service	Event Space	Exhibit
77	Manhattan Beach	Café	Bus Stop	Pizza Place	Food	Harbor / Marina	Ice Cream Shop	Beach	Sandwich Place	Playground	Women's Store
85	Sea Gate	Sports Club	Spa	Bus Station	Beach	Women's Store	Farmers Market	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Service
148	South Ozone Park	Park	Deli / Bodega	Fast Food Restaurant	Donut Shop	Sandwich Place	Bar	Hotel	Food Truck	Fishing Store	Fish Market
150	Whitestone	Bubble Tea Shop	Dance Studio	Candy Store	Deli / Bodega	Women's Store	Fast Food Restaurant	Entertainment Service	Ethiopian Restaurant	Event Service	Event Space
171	Broad Channel	Sporting Goods Shop	Deli / Bodega	Other Nightlife	Pizza Place	Dive Bar	Women's Store	Farm	English Restaurant	Entertainment Service	Ethiopian Restaurant
172	Breezy Point	Beach	Trail	Monument / Landmark	Bus Stop	Women's Store	Farmers Market	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Service
178	Rockaway Beach	Beach	Ice Cream Shop	Deli / Bodega	BBQ Joint	Metro Station	Latin American Restaurant	Food Truck	Bagel Shop	Seafood Restaurant	Pizza Place

The neighborhoods belonging to this cluster are popular for having Beach, Pool, Sports Club, Park, Spa, Bar, Hotel and many more. One who wants recreation activities would definitely love this as it has Beaches and Pools. Also, if you get hungry you have many restaurants. This is also home for travelling people via bus and train.

Let's estimate the number of **1st Most Common Venues** in this cluster.
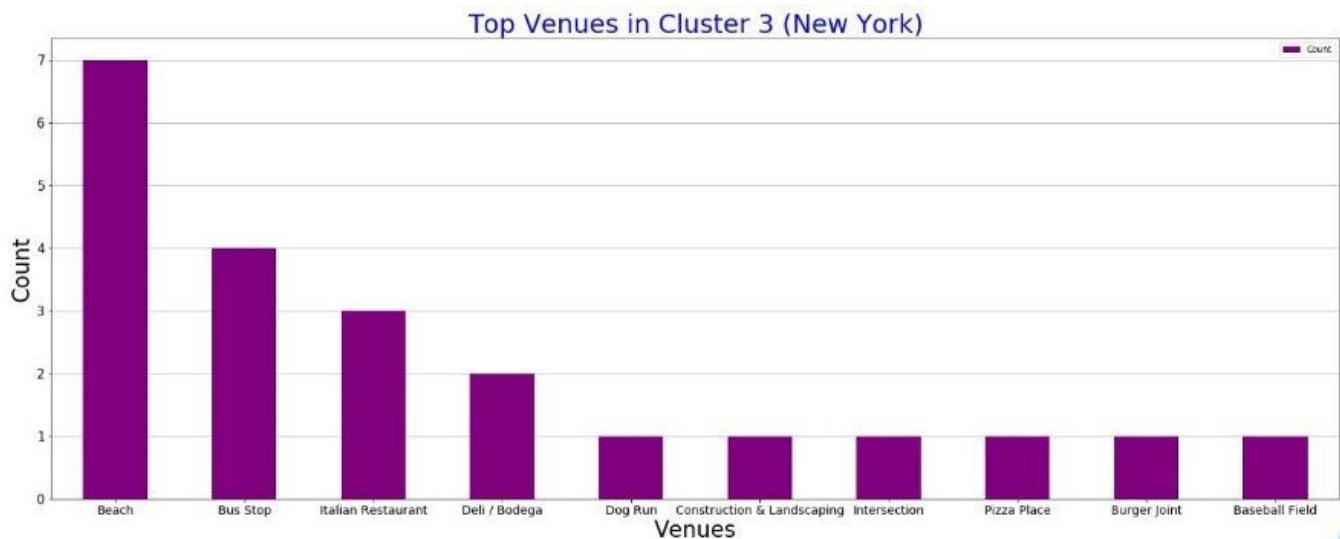
As the bar chart indicates, the topmost venue for this cluster is Beach followed by Bus Stop, Italian Restaurant and more.

E) <u>Discussion</u>

Given our cluster information for both London and New York, we see that New York and its neighborhoods are a great place for a foodie as it has varieties of restaurants of different cultures and categories. It is also the home for varieties of things such as Parks, Baseball courts, Beaches, Pools, Basketball courts, Fast food, and many more. It also has a number of transportation methods nearby. On the other hand, London is for the one who is keen in exploring a variety of places. You'll get an overall place for entertainment right from movie theatres, park, museums, concert halls to shopping, grocery stores, fast foods, restaurants and many more.

The two cities have almost the same culture when compared to entertainment, refreshments, recreation, travelling, shopping, but, for food, the result would be slightly inclined towards **New York** as it has many different kinds of varieties of food for different cultures available nearby.

<p align="center">"<i><u>New York would be a choice if you are a foodie!</u></i>"</p>

With the data, *New York* wins this battle of metros!

F) <u>Conclusion</u>

In this project, we have attempted to load the dataset for two of the prime metro cities and have tried to analyze the neighborhood regions in these metro cities based on the type of popular and top venues they have. We have clustered the neighborhoods based on the most common top venues in each of the neighborhoods. Our intention with this project was to analyze and understand the difference in the type of life in these metros, which can offer decision points for anybody who is considering to settle or plan a tour in either of the metro cities and can get a peek into what type of experience and facilities he will be provided with.