The Battle of Neighborhoods

Author: John Njeri

Title: Identifying Best Possible Neighborhood for a MealKit Business

1.1 Introduction

According to a marketing research report by Hexa Research, the Global Meal kit Delivery Service Market will be valued at USD 8.94 billion by 2025 (Hexa Research, 2019). Companies like Kroger, Blue Apron, and HelloFresh have seen a significant increase in their sales based on the market demands (Hexa Research, 2019). Meal kit delivery and sales have become popular because people are increasingly becoming busy with day to day lives, meaning that convenient packaged food delivery has become a norm for them. London is one of the cities that has been considered for the study and considering that the city is a major global financial hub, the city has numerous working-class populace who need meal kit sales and delivery services on a daily frequency. For this reason, London seems to be an appropriate city to establish a meal kit sales and delivery business to serve the increasing demand. For a potential investor, it is significant for them to identify the most viable physical location for the business depending on factors such as demand, market volume, source of materials and infrastructure. This project aims at analyzing the London neighborhoods based on the city's postcodes. The objective is to identify the most common businesses close to these postal codes and help identify the possibility of the existence of a market gap. The project will enlighten potential meal kit service investors on locations devoid of such services to invest in such areas.

1.2 Problem Statement and Description

The demand for meal kit sales and delivery has been increasing and potential investors need to identify the most viable physical positions to invest. London has become a busy city and the demand for meal kit services has significantly increased. This project will identify popular businesses close to the major London postcodes and help identify possible market gaps in the city for meal kit services investors.

1.3 Interest

New and existing investors would be interested in this research because it is a chance for the to start a business or expand their franchise in a marketable city. For this reason, the possible interested parties are existing meal kit service business and newcomers who want to kickstart their investments in a busy city.

2 Data Acquisition and Cleaning

2.1 Data Sources

The data for analysis came from three sources. The first one entailed all the London post codes from the site SpareRoom that connects possible tenants to vacant residential places. This site is up to date, meaning that obtaining the post codes from it leads to the most up to date data. The site can be accessed here. The second source of data was a csv file that has all the latitude and longitudes of all the postcodes in the United Kingdom. The file can be accessed here. The third source of data was FourSquare, a site that keeps global geographical data of all venues in all locations. This site updates regularly, meaning that the data was up to date. With the help of FourSquare API, I was able to get the most popular businesses close to the identified post codes.

2.2 Data Cleaning

I scrapped data from the pos codes data using BeautifulSoup and request to get all the postcodes and their names. With that, I had two separate files, on with the post codes and their names and the other with the longitudes and latitudes of all post codes in the United Kingdom. The common attribute between the two files was the post code ID, meaning that I could easily join the two files using inner join to get the dataframe that has the post code id, post code name, their latitudes and longitudes. The combined dataframe looked as follows. The data still had some unwanted columns, so I filtered it to retain critical columns such as Postcode ID, post code name, latitude, and longitude. The final dataframe looked as follows:

```
In [49]: final_df.head(10)
```

Out[49]:

	postcode	postcode_name	latitude	longitude
0	EC1	Aldersgate, Finsbury, Holborn	51.52286	-0.10144
1	EC2	Bishopsgate, Cheapside	51.51995	-0.08859
2	EC3	Aldgate	51.51357	-0.08309
3	EC4	St. Pauls	51.51475	-0.10034
4	E1	Whitechapel, Shoreditch	51.51766	-0.05841
5	E2	Bethnal Green	51.52939	-0.06080
6	E3	Bow	51.52789	-0.02482
7	E4	Chingford	51.62196	-0.00339
8	E5	Clapton, Homerton	51.55893	-0.05233
9	E6	East Ham	51.52560	0.05583

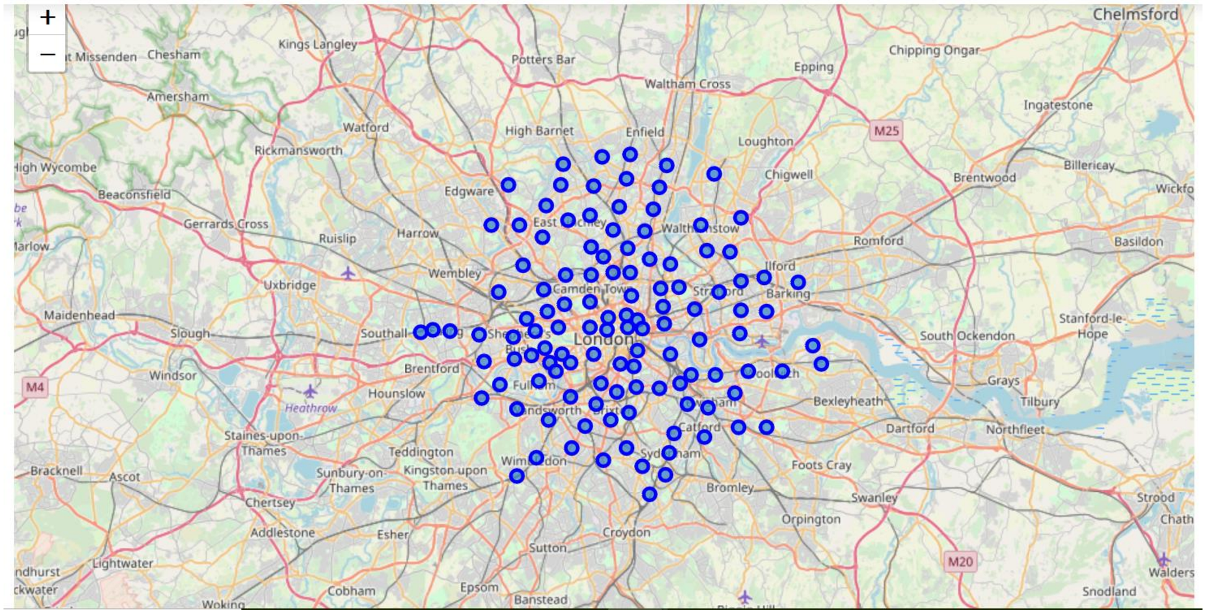
The following table identifies the columns in the combined dataframe and those that I kept.

Dropped Columns	**Kept Columns**	**Reasons for dropping**
Postcode, postcode_name, eastings, northings, latitude, longitude, town, region, uk_region, country, country_string	Postcode, postcode_name, latitude, longitude	The columns I got rid of were not essential because the data only needed was the longitude and latitude for the post codes.

3 Exploratory Data Analysis

3.1 Mapping the Data

After importing the relevant libraries I was ready to map the points in the dataframe to folium maps using geopy. I used the London Latitude and longitude as the central locations and then mapped the post codes into the maps. The London maps with the post codes superimposed looked as follows:



3.2 Integrating FourSquare API to identify the closes neighborhoods

To get the closest venues in all neighborhoods, I wrote a function named getVenues with post codes names, latitudes, longitudes, and radius as the arguments. I used the FourSquare API to get the data and specified that locations had to be within a 500-meter radius. I used the requests.get method to collect the venue, neighborhood, its latitude and longitude, and the category to a json file. After cleaning and formatting the data, the venues table was as follows:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Aldersgate, Finsbury, Holborn	51.52286	-0.10144	The Zetter Townhouse	51.522849	-0.103658	Hotel
1	Aldersgate, Finsbury, Holborn	51.52286	-0.10144	BrewDog Clerkenwell	51.522401	-0.103835	Beer Bar
2	Aldersgate, Finsbury, Holborn	51.52286	-0.10144	Foxlow Clerkenwell	51.521488	-0.101767	Steakhouse
3	Aldersgate, Finsbury, Holborn	51.52286	-0.10144	Sushi Tetsu	51.523348	-0.104015	Sushi Restaurant
4	Aldersgate, Finsbury, Holborn	51.52286	-0.10144	Luca	51.522017	-0.101703	Italian Restaurant

I then organized the data to get the venues returned for each neighborhood and the table was as follows:

Out[67]:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Abbey Wood	6	6	6	6	6	6
Acton	23	23	23	23	23	23
Aldersgate, Finsbury, Holborn	67	67	67	67	67	67
Aldgate	100	100	100	100	100	100
Angel	68	68	68	68	68	68
Balham	54	54	54	54	54	54
Barnes	25	25	25	25	25	25
Battersea	40	40	40	40	40	40
Bayswater, Paddington	100	100	100	100	100	100
Bermondsey, Rotherhithe	35	35	35	35	35	35

3.3 Analyzing Each Neighborhood

I used one hot encoding to get a table of dummies with the venue categories being the columns of the onehot table. The snippet of the table is as follows:

	Yoga Studio	Accessories Store	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Austr Restau
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

After adding the neighborhood category to the onehot table, I grouped the rows by neighborhoods and the mean of the frequency of the occurrence of each category and got the following table:

	Neighborhood	Yoga Studio	Accessories Store	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	As Restaur
0	Abbey Wood	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.00	0.0	0.0	0.000000	0
1	Acton	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.00	0.0	0.0	0.000000	0
2	Aldersgate, Finsbury, Holborn	0.014925	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.00	0.0	0.0	0.000000	0
3	Aldgate	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.01	0.0	0.0	0.000000	0
4	Angel	0.014706	0.0	0.014706	0.0	0.0	0.014706	0.0	0.0	0.00	0.0	0.0	0.014706	0

After getting the table above, I wrote a for loop code that iterated through all the neighborhoods and the new table was as follows:

	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	Convenience Store	Supermarket	Campground	Train Station	Coffee Shop	Grocery Store	Food & Drink Shop	Ethiopian Restaurant	Event Space	Exhibit
1	Acton	Pub	Gym / Fitness Center	Park	Fast Food Restaurant	Turkish Restaurant	Chinese Restaurant	Bakery	Shopping Mall	Grocery Store	Coffee Shop
2	Aldersgate, Finsbury, Holborn	Pub	Coffee Shop	Hotel	Café	Italian Restaurant	Bar	French Restaurant	Vietnamese Restaurant	Gym / Fitness Center	Steakhouse
3	Aldgate	Coffee Shop	Restaurant	Hotel	Gym / Fitness Center	French Restaurant	Sandwich Place	Cocktail Bar	Italian Restaurant	Pub	Salad Place
4	Angel	Pub	Coffee Shop	Gastropub	Park	Café	French Restaurant	Mediterranean Restaurant	Japanese Restaurant	Bakery	Burger Joint

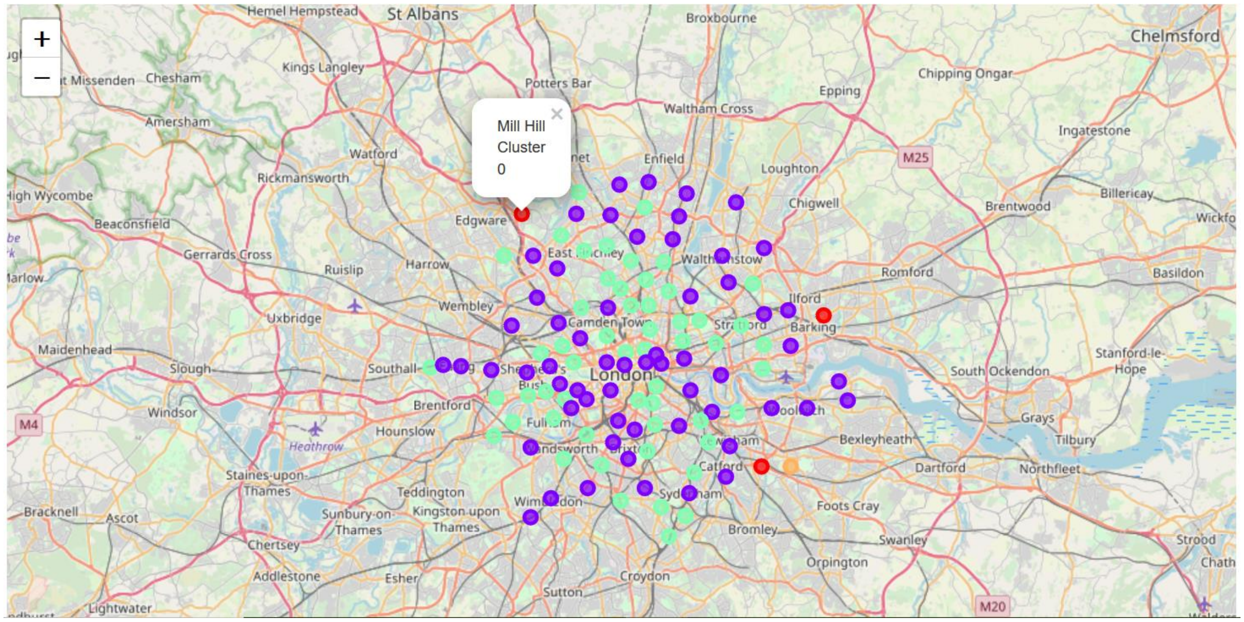
I then clustered the neighborhoods into 5 clusters and added the cluster labels column as follows:

	postcode	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 M Comr Ve
0	EC1	Aldersgate, Finsbury, Holborn	51.52286	-0.10144	3	Pub	Coffee Shop	Hotel	Café	Italian Restaurant	Bar	French Restaurant	Vietnamese Restaurant	Steakhc
1	EC2	Bishopsgate, Cheapside	51.51995	-0.08859	1	Coffee Shop	Food Truck	Hotel	Sushi Restaurant	Gym / Fitness Center	Bar	Café	Boxing Gym	(
2	EC3	Aldgate	51.51357	-0.08309	1	Coffee Shop	Restaurant	Hotel	Gym / Fitness Center	Sandwich Place	French Restaurant	Salad Place	Cocktail Bar	Ita Restau
3	EC4	St. Pauls	51.51475	-0.10034	1	Coffee Shop	Italian Restaurant	Pub	Gym / Fitness Center	Japanese Restaurant	Vietnamese Restaurant	Falafel Restaurant	Modern European Restaurant	Restau
4	E1	Whitechapel, Shoreditch	51.51766	-0.05841	1	Pub	Hotel	Indian Restaurant	Grocery Store	Burger Joint	Coffee Shop	Sandwich Place	Ice Cream Shop	Tur Restau
5	E2	Bethnal Green	51.52939	-0.06080	3	Coffee Shop	Café	Cocktail Bar	Pub	Hotel	Flower Shop	Bar	Restaurant	Pizza P
6	E3	Bow	51.52789	-0.02482	3	Pub	Grocery Store	Hotel	Bar	Rental Car Location	Gym	Burger Joint	Locksmith	Light Sta
7	E4	Chingford	51.62196	-0.00339	1	Museum	American Restaurant	Flea Market	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Film Studio	Fis Chips S
8	E5	Clapton, Homerton	51.55893	-0.05233	1	Grocery Store	Park	Coffee Shop	Breakfast Spot	Train Station	Bus Stop	Café	Garden	Fis Chips S

4 Results

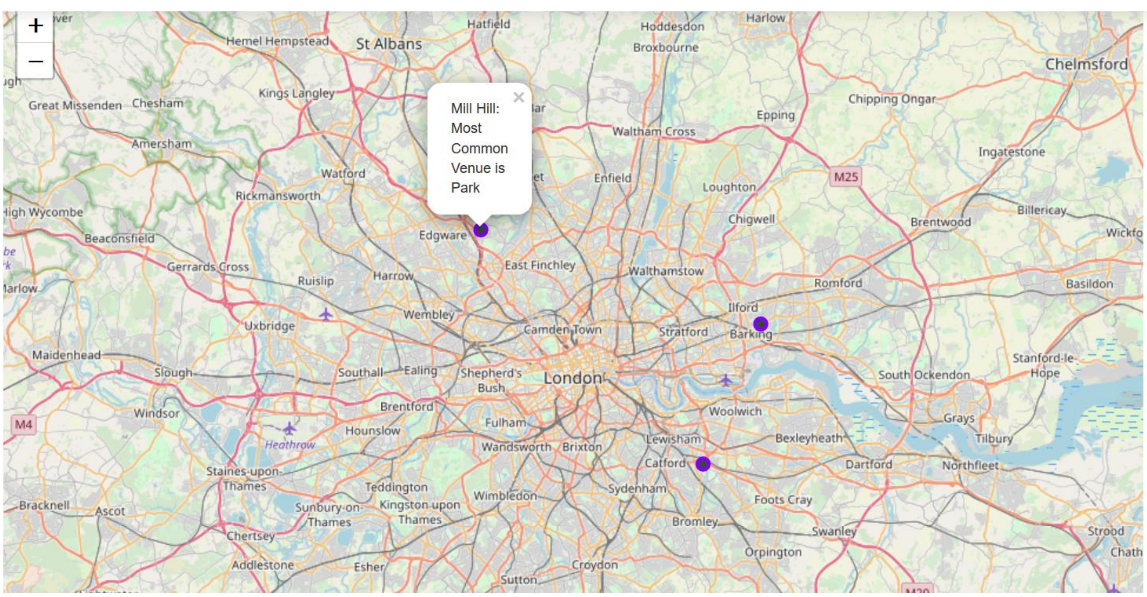
The data was ready for observation in maps and I used folium maps to map these venues as below.

Total Clusters

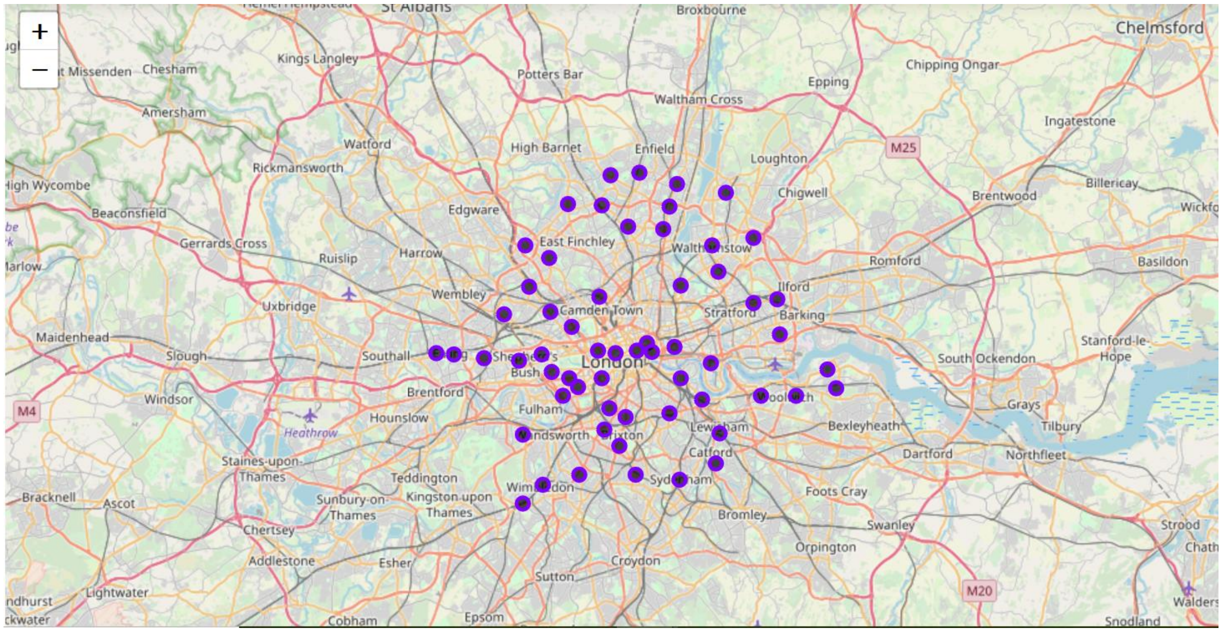


The map above contains all the neighborhoods clustered into 5 groups. I mapped each cluster with the labels indicating the most common neighborhood.

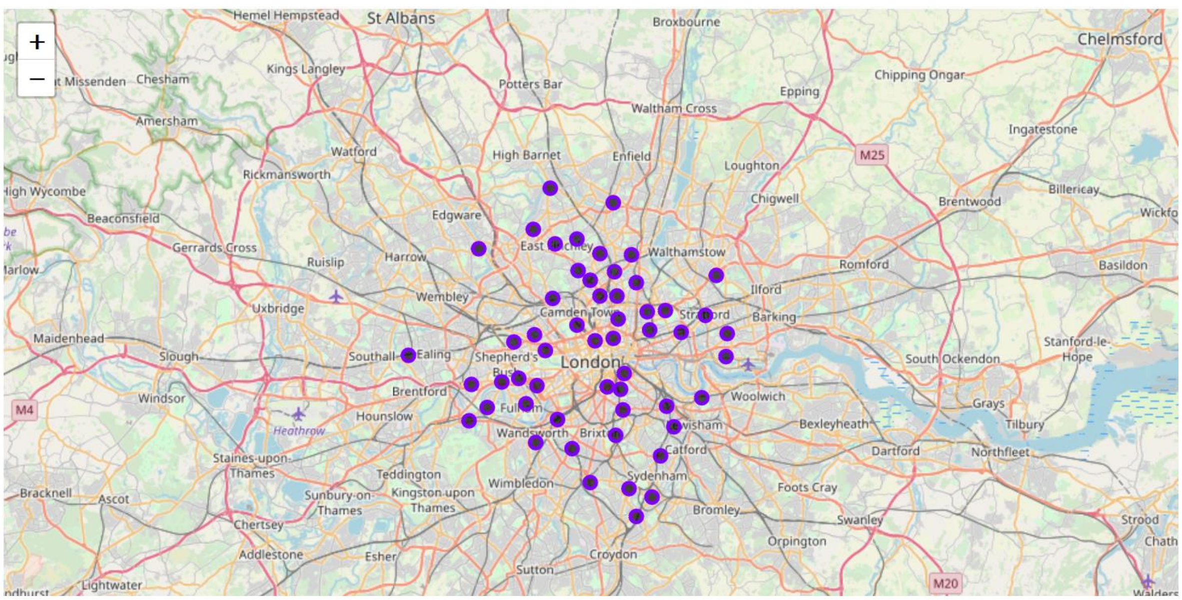
First Cluster



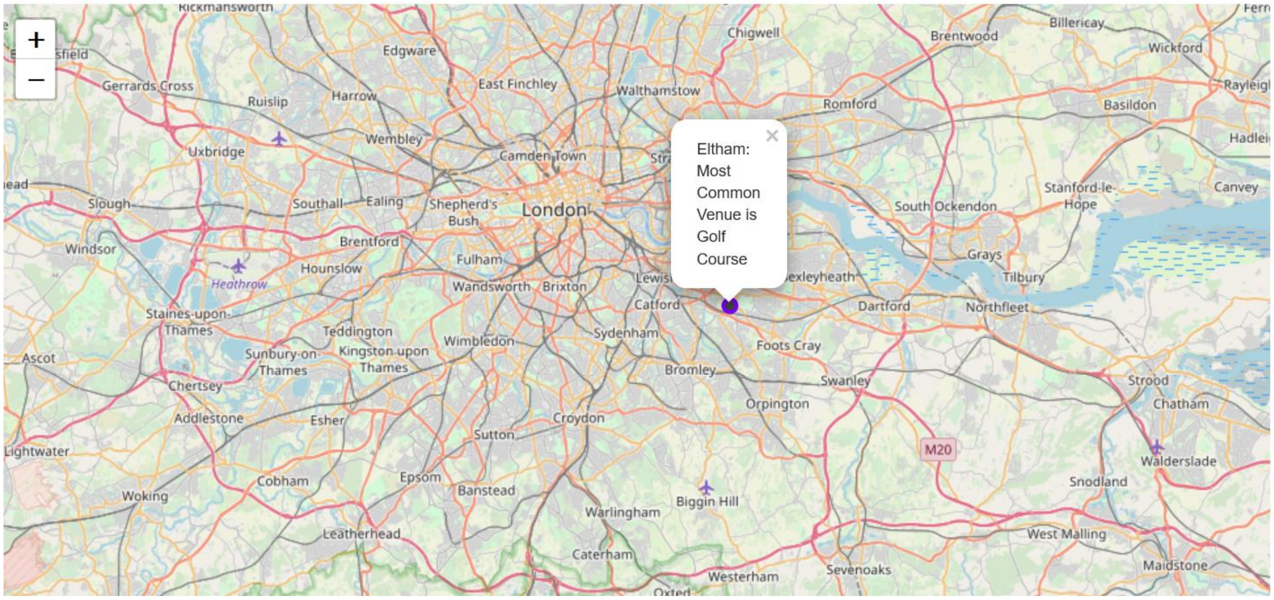
Second Cluster



Fourth Cluster



Fifth Cluster



5 Analysis of results

The following table represents the clusters and the most common venues

Cluster Number	Number of Postcodes	Most Notable Venues
1	3	Park, Pharmacy, Fish Market, Restaurant
2	59	Coffee Shop, Grocery Store, Pub, Pizza Place
3	0	None
4	54	Pub, Restaurant, Café, Bakery
5	1	Golf Course, Farmers Market, Fish Market, Farm

Based on the above results, it shows that most neighborhoods are concentrated in clusters 2 and 4 and the most common venues for these clusters are restaurants, pubs, café, and other eateries. Other clusters have few neighborhoods with most common venues being parks, farms and restaurants. The indication is that London has the majority of the population in clusters 2 and 4 thus explaining the availability of all these businesses. The rest of the clusters, due to the popularity of farms, signifies that the population is low. For this reason, a Meal Kit service business investor should consider clusters 2 and 4 because the market is huge in the area. The competition is high, but the area has busy people meaning that the gap exists. For this reason, based on the observation of this project, the most viable locations to choose to invest are clusters 2 and 4.

References

Hexa Research. (2019, March 13). Global Meal Kit Delivery Service Market Worth USD 8.94 Billion by 2025: Hexa Research. Retrieved from https://www.prnewswire.com/news-releases/global-meal-kit-delivery-service-market-worth-usd-8-94-billion-by-2025-hexa-research-300811555.html

Reuters. (n.d.). 2018 IFZ Global FinTech Rankings. Retrieved from https://innovation.thomsonreuters.com/en/labs/portfolio/global-fintech-rankings.html#/