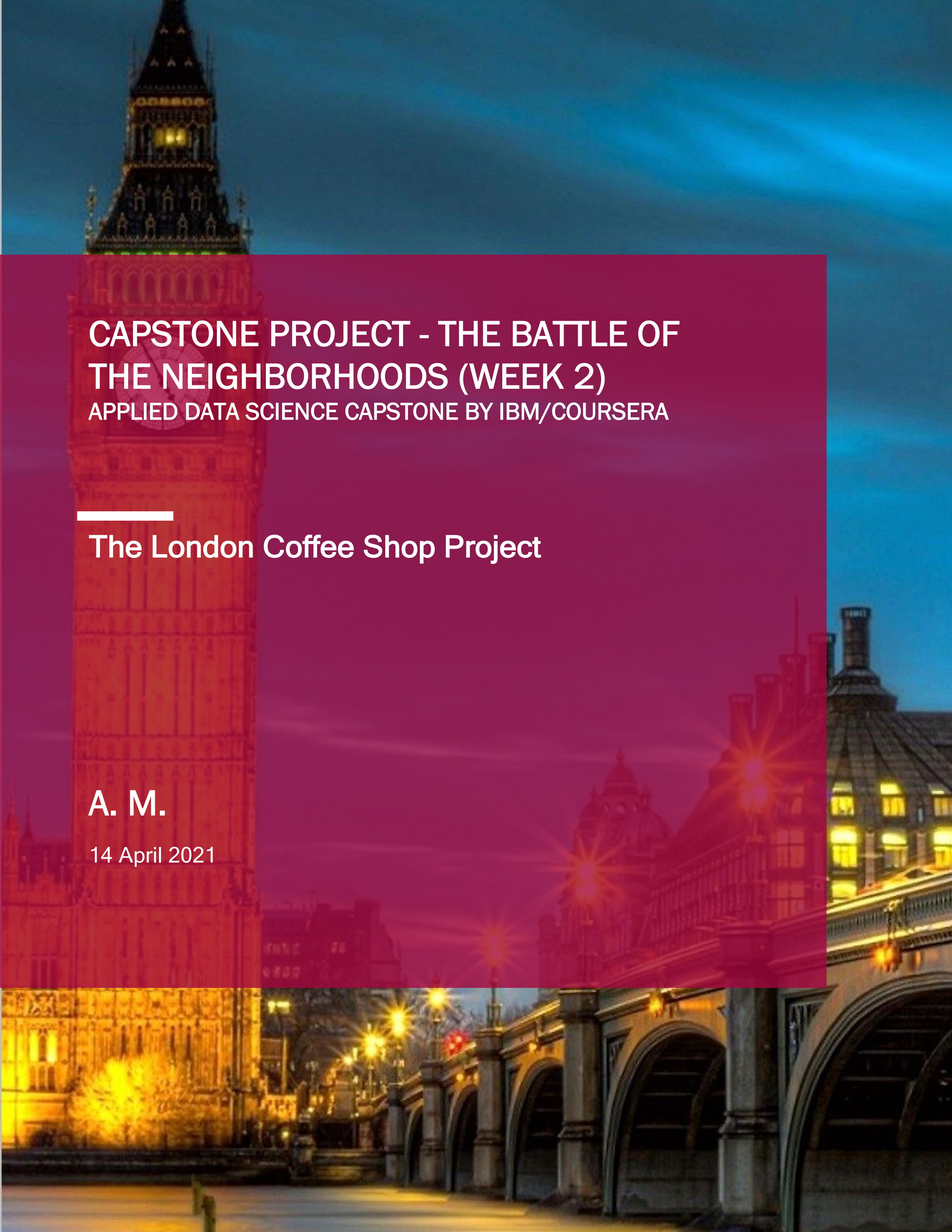


CAPSTONE PROJECT - THE BATTLE OF THE NEIGHBORHOODS (WEEK 2)

APPLIED DATA SCIENCE CAPSTONE BY IBM/COURSERA

The London Coffee Shop Project

A. M.

14 April 2021

1. Introduction

1.1. Business Problem

A young entrepreneur has decided to open a coffee shop in London. Coffee shop serve multiple purposes for example chatting, meeting, eating, or studying [4]. Four main attributes that can reflect coffee shop characteristics: atmosphere, employee attitude, IT service, and taste [4]. The project is financially viable, but the location of the shop has not been selected yet. To make the best decision, the entrepreneur needs to know which borough and neighborhoods in London are the most profitable. We will need to explore the venues in London's boroughs to understand where the competition is located. This will help in determining the best location for the new shop. Also, we need to know which customers these competing venues serve. To this end, we will explore the boroughs further to predict how the new shop can be supported by nearby venues such as university campuses and other cultural venues. Ideally, the new coffee shop will be located in the borough with the highest concentration of potential customers and the lowest competing venues' concentration.

Potential customers are very important in any business. Therefore before opening this shop we have to search which are is best suitable where there is large amount of customers. Another important factor is competitors. If there are more than one shop in that area that customers attraction will be divided. We aim to choose that place where number of competitors are less. So ideal area would be that has low number of competitors and high number of customers.

2. Data

Data is very important while making any decisions. It enables companies to create new business opportunities. That will help business to grow. Data also helps in setting future trends. Data is crucial to business success. Social networking sites are getting our data. Things we clicked, have been added in our favorite list. After getting user's interest from user's action. These sites displayed products that matches user's choice. This is one of the example that shows importance of data. As our project is to search best place for coffee shop. Data related to this project would be location of different streets that have high number of customers and less numbers of competitors. There are few steps that will be used for data acquisition.

2.1. Data acquisition

The data required for this project can be collected from multiple sources. The project relies primarily on the foursquare API to collect data about the venues in London and their location. API stands for application programmable interface. API is used to access data and interact with external components. The Greater London Area will be covered in

this project. The data is publicly available and is collected and organized from the Wikipedia *List of areas of London* page which can be found [HERE]. Wikipedia is free encyclopedia. Anybody can edit this. The data is then complemented with Latitude and Longitude data for the different postcodes which are collected using geocoder and ArcGIS. Longitude and Latitudes are used to locate any position in earth. Geocoder is used to convert any location to its longitude and latitude.
Libraries are predefined functionalities that a program used to perform its functions. Our project used following libraries pandas, requests, bs4, plotly, folium, geocoder,and numpy. Each library has different functions for examples panda library is used for data manipulation and analysis. Data preparation is next step after this.

2.2. Data preparation

This data is obtained from Wikipedia. Wikipedia is basically an encyclopedic website. It contains information about everything. Anybody can add or delete any information from this site. The data contains information about the location of the neighborhoods, their boroughs, towns, postcodes, dial codes, and OS grid ref. This project utilizes the columns Location, *London borough*, *Post town*, *Postcode district*; the other columns are not required and dropped from the dataframe. The data also requires cleaning, especially the data related to boroughs as they contain unnecessary references. *The Postcode district* column contains cells that have multiple postcodes that have been split into two rows to standardize the data and prepare it for geocoding.
In data preparation website scrapping is used. Scrapping is a process of extracting content of site using bots. It is a powerful tool. After scrapping we get following columns e.g London, London borough, Post town, Postcode district, Dial code and OS grid.

	Location	London borough	Post town	Postcode district	Dial code	OS grid ref
0	Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
2	Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728

Table 1.

2.3. Data cleaning

Before cleaning of data, we need to rename columns. London is converted into Neighborhood. Post town is converted into town.

	Neighborhood	Borough	Town	Postcode	Dialcode	OSgridRef
0	Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
2	Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728

Table 2.

Now after renaming columns. Next step is to remove unnecessary information e.g., borough column contains references name. We replaced these symbols with space. This is shown in diagram below.

	Neighborhood	Borough	Town	Postcode	Dialcode	OSgridRef
0	Abbey Wood	Bexley, Greenwich	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham	LONDON	W3, W4	020	TQ205805
2	Addington	Croydon	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728

Table 3.

After examining of data, we have seen that there are unnecessary data present e.g. Dialcode and OSgridRef. We should remove those columns as they have nothing to do with our analysis. Final columns that are left Neighborhood, Borough, Town and Postcode. This is shown in table below.

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

Table 4.

In postcode column there are multiple postcodes present in a single row. We need to spread it into multiple columns. This is shown in table below.

	Neighborhood	Borough	Town	Postcode
0	Abbey Wood	Bexley, Greenwich	LONDON	SE2
1	Acton	Ealing, Hammersmith and Fulham	LONDON	W3
2	Acton	Ealing, Hammersmith and Fulham	LONDON	W4
3	Addington	Croydon	CROYDON	CR0
4	Addiscombe	Croydon	CROYDON	CR0

Table 5.

After that we removed white space from Postcode and convert Town to title case. This is shown in table below.

	Neighborhood	Borough	Town	Postcode
0	Abbey Wood	Bexley, Greenwich	London	SE2
1	Acton	Ealing, Hammersmith and Fulham	London	W3
2	Acton	Ealing, Hammersmith and Fulham	London	W4
3	Addington	Croydon	Croydon	CR0
4	Addiscombe	Croydon	Croydon	CR0

Table 6.

Now we used subet to only include postcodes with London as Town. This is shown in table below.

	Neighborhood	Borough	Town	Postcode
0	Abbey Wood	Bexley, Greenwich	London	SE2
1	Acton	Ealing, Hammersmith and Fulham	London	W3
2	Acton	Ealing, Hammersmith and Fulham	London	W4
3	Aldgate	City	London	EC3
4	Aldwych	Westminster	London	WC2

Table 7.

Two API's geocoder and ArcGIS are used to get coordinates for different postcodes. Longitudes and latitudes are returned after this line. New columns are Neighborhood, Borough,
Town, Postcode, Latitude and Longitude. This is shown in table below.

	Neighborhood	Borough	Town	Postcode	Latitude	Longitude
0	Abbey Wood	Bexley, Greenwich	London	SE2	51.49245	0.12127
1	Acton	Ealing, Hammersmith and Fulham	London	W3	51.51324	-0.26746
2	Acton	Ealing, Hammersmith and Fulham	London	W4	51.48944	-0.26194
3	Aldgate	City	London	EC3	51.51200	-0.08058
4	Aldwych	Westminster	London	WC2	51.51651	-0.11968

Table 8

.

3. Methodology

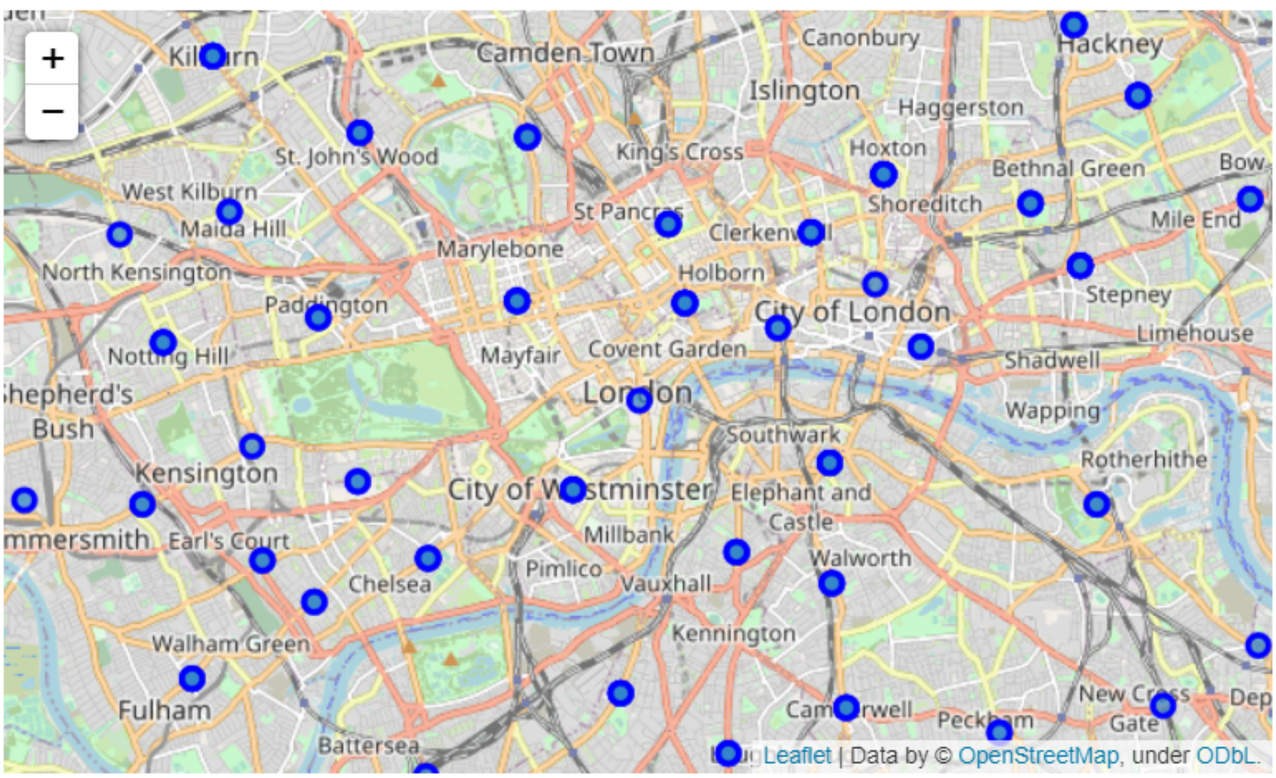
In this project we will direct our efforts on detecting areas of London that have a high concentration of commercial activity, particularly those with high numbers of venues that can support our new business. We will limit our analysis to London borough and exclude the greater Lonodn area.

In first step we have collected the required data: georaphical coordinates of london boroughs and the locaition of its venues with their category.

Second step in our analysis will be calculation and exploration of 'venues across different areas of London. Once collected, we will move to the third step of the analysis. We will focus on most promising areas by creating clusters of location to identify where most of the commercial activity is located. We will use maps to visualise those areas and focus our analysis on the most promising cluster in the fourth step. Finally, Within the most promosing cluster, we will proceed by elemination and define further clusters to explore optimal venue location by supporting venues such as museums and hotels in order to choose a location within these areas.

4. Analysis

Checking data if there is any missing value. Library geopy is used to get longitude and latitude of London. After getting longitude and latitude values, we created map that has markers showing neighborhoods in London.



Next step is to explore nearby venues. Parameters that are giving name, longitude, latitude within 500 meter radius. This will gives us all places that are in 500 meter radius. After getting all places we need to get exact location of these. Location has got in two parts longitude and latitude. This is shown in table below.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbey Wood	51.49245	0.12127	Lesnes Abbey	51.489526	0.125839	Historic Site
1	Abbey Wood	51.49245	0.12127	Sainsbury's	51.492826	0.120524	Supermarket
2	Abbey Wood	51.49245	0.12127	Lidl	51.496152	0.118417	Supermarket
3	Abbey Wood	51.49245	0.12127	Abbey Wood Railway Station (ABW)	51.491097	0.121334	Train Station
4	Abbey Wood	51.49245	0.12127	Bean @ Work	51.491172	0.120649	Coffee Shop

Table 10.

We have seen that most of the venues are repeated so we need to group those venues and show their count e.g Acton is repeated 47 times. We don't need repetitive values. Therefore we only choose unique values. After we have unique places, Next step is to explore each place. This is shown in table below.

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Abbey Wood	6	6	6	6	6	6
Acton	47	47	47	47	47	47
Aldgate	89	89	89	89	89	89
Aldwych	87	87	87	87	87	87
Anerley	6	6	6	6	6	6
...	...	...	...	...	...	...
Wood Green	8	8	8	8	8	8
Woodford	75	75	75	75	75	75
Woodside Park	33	33	33	33	33	33
Woolwich	8	8	8	8	8	8
Wormwood Scrubs	29	29	29	29	29	29

Table 11.

We do this by getting venues one by one and search weather there exits Accessories Store, Adult Boutique, African Restaurant, American Restaurant, Antique Shop, Arcade, Arcade, Arepa Restaurant, Argentinian Restaurant, Art Gallery, Art Museum, Whisky Bar, Windmill, Wine Bar, Wine Shop, Wings Joint, Women's Store, Xinjiang Restaurant, Yoga Studio and Zoo Exhibit. Now we have to get top 5 menus for example in case of Abbey Wood have venues Supermarket, Platform, Train Station, Historic Site and Historic Site.

Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Table 12.

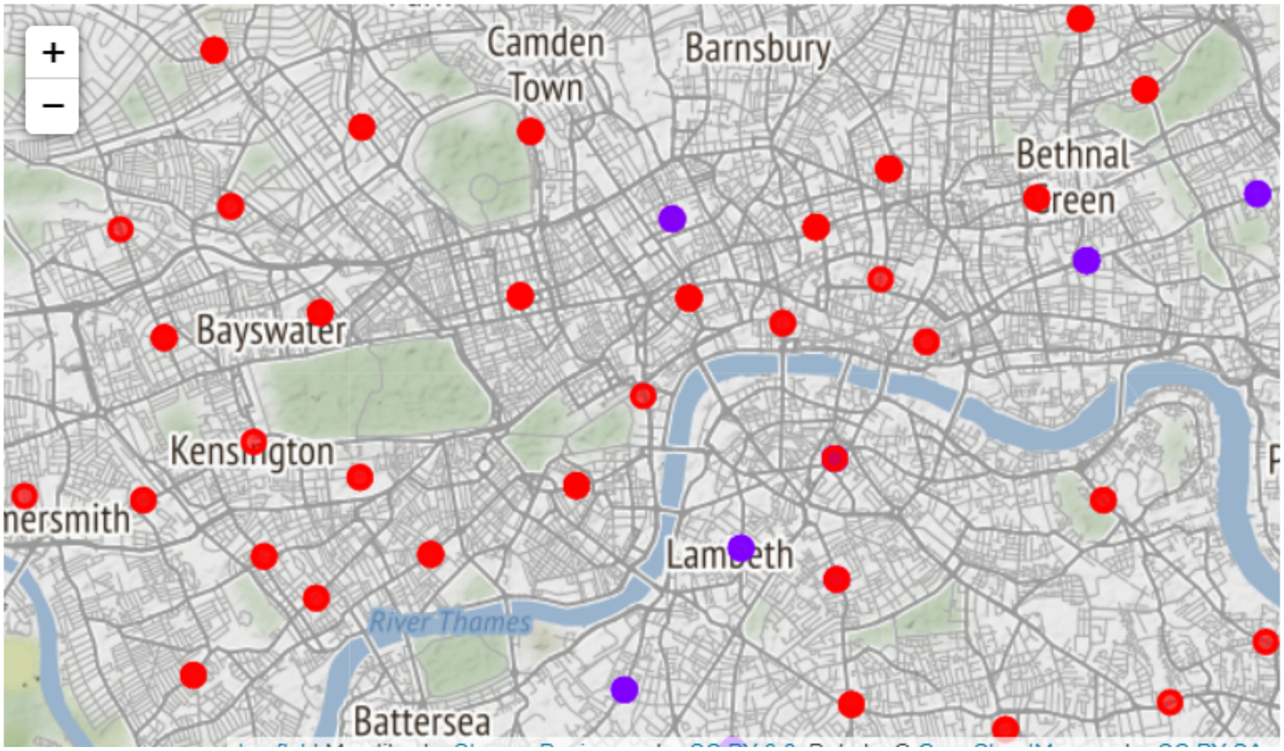
This step involves getting top 10 venues of each neighborhood for example Abbey Wood’s 1st Most Common Venue is Supermarket, 2nd Most Common Venue is Train Station, 3rd Most Common Venue is Coffee Shop, 4th Most Common Venue is Historic Site, 5th Most Common Venue is 5th Most Common Venue, 6th Most Common Venue is Zoo Exhibit, 7th Most Common Venue Fish Market, 8th Most Common Venue is Farmers Market, 9th Most Common Venue is Fast Food Restaurant and 10th Most Common Venue Filipino Restaurant.
This is shown in table 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
0	Abbey Wood	Supermarket	Train Station	Coffee Shop	Historic Site	Platform	Zoo Exhibit	Fish Market
1	Acton	Pub	Café	Grocery Store	Bakery	Coffee Shop	Italian Restaurant	Bookstore
2	Aldgate	Hotel	Gym / Fitness Center	Restaurant	Coffee Shop	Salad Place	Cocktail Bar	Garden
3	Aldwych	Café	Pub	Sandwich Place	Hotel	Theater	Bookstore	Japanese Restaurant
4	Anerley	Supermarket	Grocery Store	Fast Food Restaurant	Convenience Store	Hotel	Zoo Exhibit	Flea Market

Table 13.

After getting top 10 venues of neighborhood we will used k mean clustering algorithm. It is type of unsupervised learning used for unlabeled data. This algorithm works in iterations. New column named Cluster Labels will be added after this step. Now total columns are Neighborhood, Borough, Town, Postcode, Latitude, Cluster Labels, and top 10 most common venues.

New map has been created by using this new data that has been processed. This is shown below.



Exploring 3 most common venues per cluster. New columns are Neighborhood, Cluster Labels, 1st Most Common Venue, 2nd Most Common Venue and 3rd Most Common Venue. Now we have top 3 menus per cluster for example Abbey Wood has Supermarket, Train Station and Coffee Shop. This is shown below.

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Abbey Wood	4	Supermarket	Train Station	Coffee Shop
1	Acton	0	Pub	Café	Grocery Store
2	Acton	0	Pub	Café	Grocery Store
3	Aldgate	0	Hotel	Gym / Fitness Center	Restaurant
4	Aldwych	0	Café	Pub	Sandwich Place
5	Anerley	4	Supermarket	Grocery Store	Fast Food Restaurant
6	Angel	0	Coffee Shop	Pub	Food Truck
7	Angel	0	Coffee Shop	Pub	Food Truck
8	Archway	4	Grocery Store	Pizza Place	Coffee Shop
9	Arnos Grove	4	Grocery Store	Bus Stop	Fast Food Restaurant
10	Arnos Grove	4	Grocery Store	Bus Stop	Fast Food Restaurant
11	Balham	0	Coffee Shop	Grocery Store	Pub
12	Bankside	1	Pub	Coffee Shop	Park
13	Barbican	0	Food Truck	Coffee Shop	Pub
14	Barbican	0	Food Truck	Coffee Shop	Pub

Table 14.

This step involves exploring venue of each cluster individually. We do this by getting each venue from cluster and counts number of times specific place occurred nearby each cluster for example Angel has cluster label 0. In label 0 Pub venue is occurred 66 times in first most common venue. Café is repeated 60 in second most common venue and pub is repeated 60 times in third most common venue.

Now we have venue and number of occurrences of that venue for example Bakery is repeated 372 times. We plot this data. Plotting this in histogram gives us a clear idea of which venue is repeated most frequently.

We represent each cluster individually for example cluster 1 has columns Coffee Shop, Coffee Shop, 1st Most Common Venue, 1st Most Common Venue and 3rd Most Common Venue. Make histogram of each cluster after getting data.

Detailed analysis of cluster zero represents that it has most commercial activities. This is the place where our new coffee shop should be open. This will attracts many customers. In this way this startup will generate great revenue. This is shown in table below.

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
1	Acton	0	Pub	Café	Grocery Store
2	Acton	0	Pub	Café	Grocery Store
3	Aldgate	0	Hotel	Gym / Fitness Center	Restaurant
4	Aldwych	0	Café	Pub	Sandwich Place
6	Angel	0	Coffee Shop	Pub	Food Truck

Table 15.

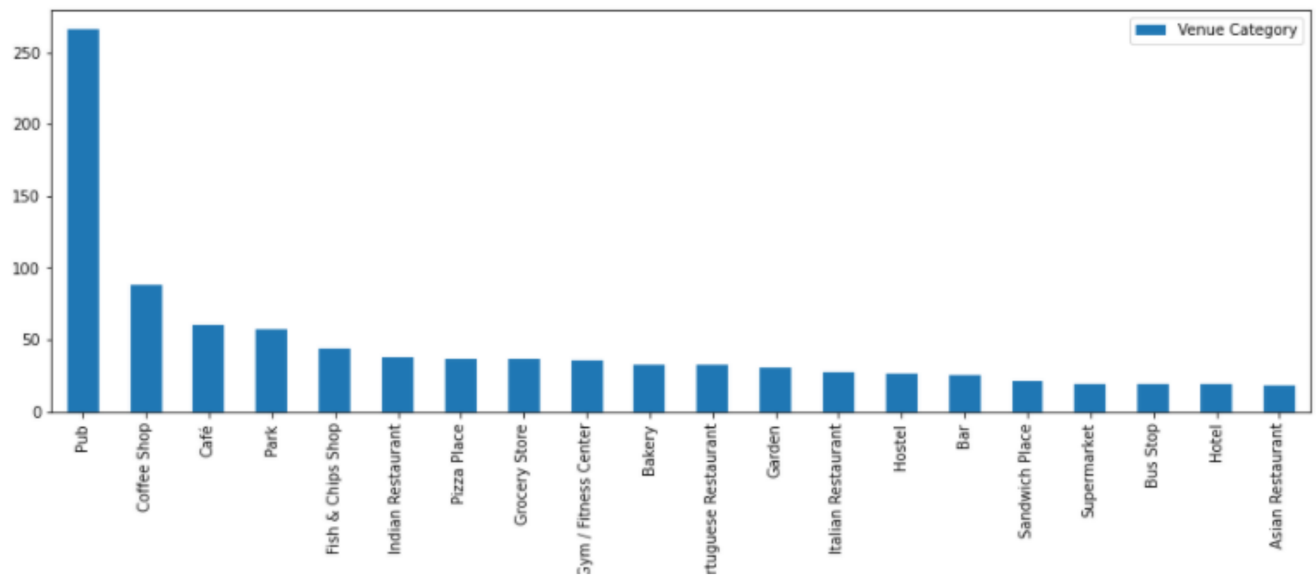
We will run another cluster analysis to make sure that our choice is correct. Analysis involved selecting top 10 most common venues, Applying clustering, Adding coordinates, Creating Maps, Making graphs, counting occurrence of each venue, discussing each cluster.

After detailed analysis of all clusters we got two clusters that fulfills our criteria. Cluster 0and Cluster 3. Table of cluster 1 is shown below.

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
12	Bankside	1	Pub	Coffee Shop	Park
23	Bermondsey	1	Pub	Coffee Shop	Park
26	Blackheath	1	Pub	Photography Studio	Art Gallery
27	Blackheath Royal Standard	1	Pub	Art Gallery	Café
28	Blackheath Royal Standard	1	Pub	Art Gallery	Café

Table 16.

Graph of cluster 1 is shown below.



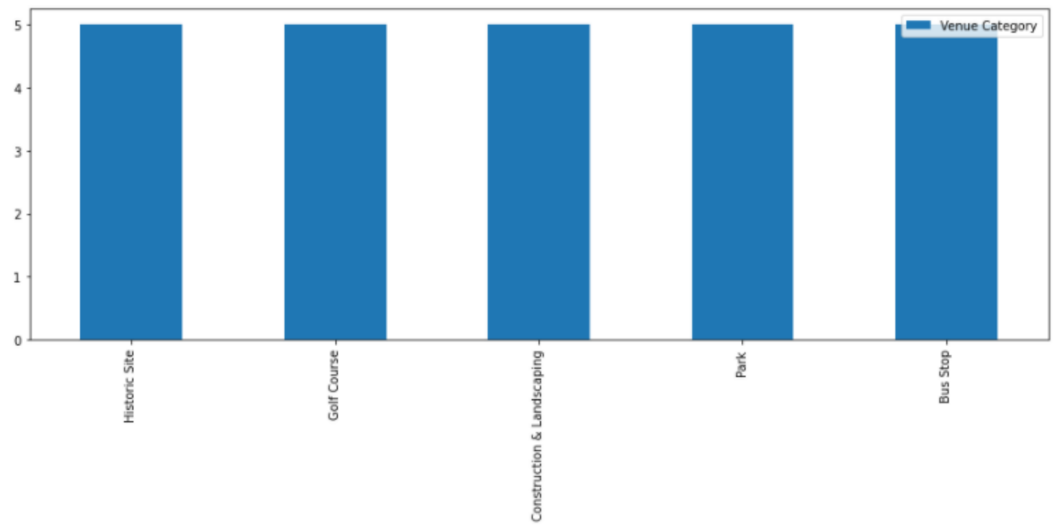
Graph of Cluster 1

Table of cluster 2 is shown below.

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
103	Eltham	2	Park	Bus Stop	Historic Site
201	Middle Park	2	Park	Bus Stop	Historic Site
207	Mottingham	2	Park	Bus Stop	Historic Site
212	New Eltham	2	Park	Bus Stop	Historic Site
328	Well Hall	2	Park	Bus Stop	Historic Site

Table 17.

Graph of cluster 2 is shown below.



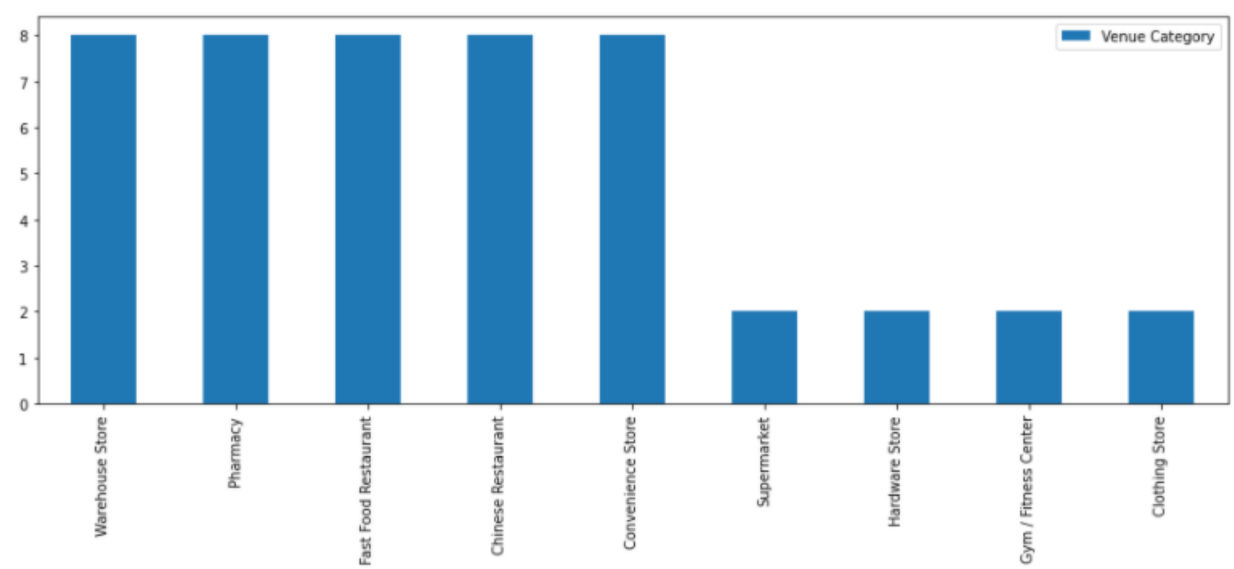
Graph of Cluster 2

Table of cluster 3 is shown below.

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
37	Brent Park	3	Chinese Restaurant	Pharmacy	Fast Food Restaurant
142	Harlesden	3	Chinese Restaurant	Pharmacy	Fast Food Restaurant
209	Neasden	3	Supermarket	Convenience Store	Gym / Fitness Center
210	Neasden	3	Supermarket	Convenience Store	Gym / Fitness Center
226	Old Oak Common	3	Chinese Restaurant	Pharmacy	Fast Food Restaurant

Table 18.

Graph of cluster 3 is shown below



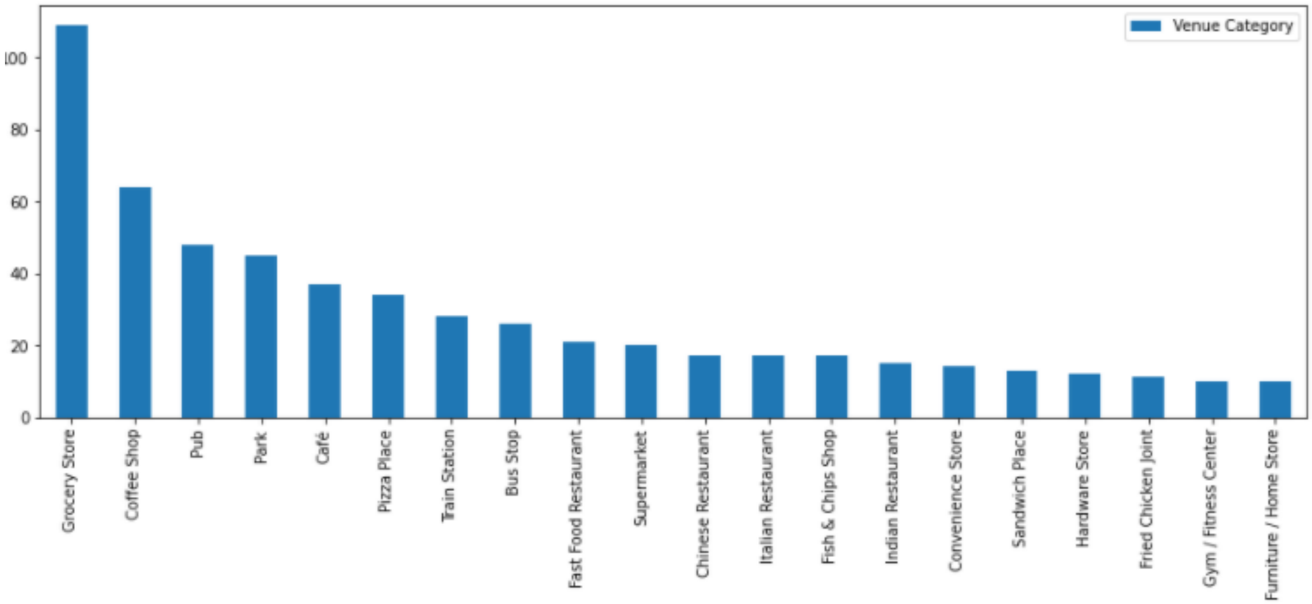
Graph of Cluster 3

Table of cluster 4 is shown below.

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Abbey Wood	4	Supermarket	Train Station	Coffee Shop
5	Anerley	4	Supermarket	Grocery Store	Fast Food Restaurant
8	Archway	4	Grocery Store	Pizza Place	Coffee Shop
9	Arnos Grove	4	Grocery Store	Bus Stop	Fast Food Restaurant
10	Arnos Grove	4	Grocery Store	Bus Stop	Fast Food Restaurant

Table 19.

Graph of cluster 4 is shown below



Graph of Cluster 4

Detailed analysis of cluster 0 is shown in table below.

Neighborhood	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Acton	0	51.51324	-0.26746	Sainsbury's Local	51.514966	-0.268902	Grocery Store
Acton	0	51.51324	-0.26746	Acton Main Line Railway Station (AML)	51.517077	-0.267317	Train Station
Acton	0	51.51324	-0.26746	Co-op Food	51.515960	-0.267735	Grocery Store
Acton	0	51.51324	-0.26746	The Balti House	51.516627	-0.267307	Indian Restaurant
Acton	0	51.51324	-0.26746	Springfield Gardens	51.510826	-0.271955	Park

Table 20.

We explore supporting venues of these two clusters. Cluster 0 and 4 data includes Cluster Label, 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, 8th Most Common Venue, 8th Most Common Venue, 10th Most Common Venue, Neighborhood Latitude, Neighborhood Longitude, Venue, Venue Latitude, Venue Longitude and Venue Category. We will specify each venue number occurrence. This is shown in below table.

Neighborhood	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Arge Res
Acton	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.
Aldgate	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0
Aldwych	0.011494	0.000000	0.000000	0.0	0.011494	0.011494	0.0	0
Angel	0.000000	0.007937	0.000000	0.0	0.000000	0.000000	0.0	0.
Balham	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.
...	...	...	...	...	...	...	...	
White City	0.000000	0.000000	0.034483	0.0	0.000000	0.000000	0.0	0.
Wimbledon	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.
Woodford	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.
Woodside Park	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.
Wormwood Scrubs	0.000000	0.000000	0.034483	0.0	0.000000	0.000000	0.0	0.

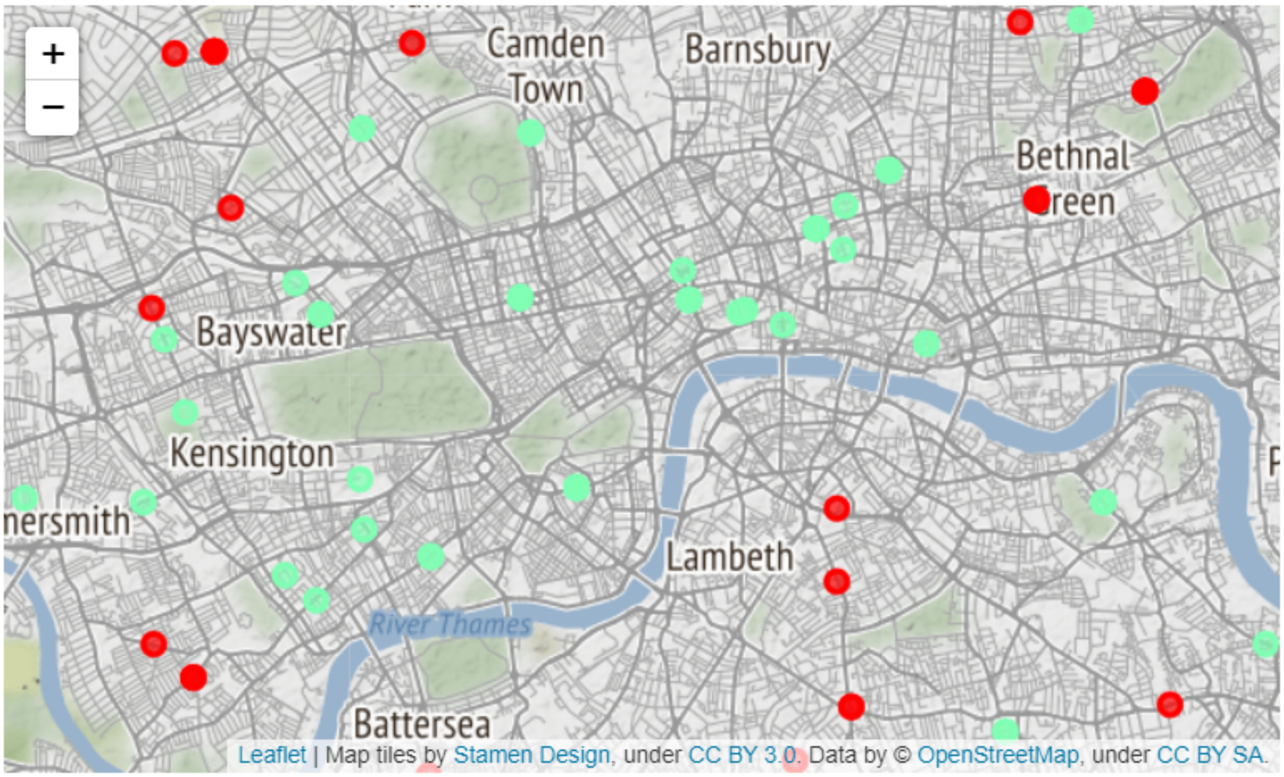
Table 21.

K-means clustering is applied. Data is shown below.

	Cluster Labels	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
0	0	Acton	Café	Pub	Coffee Shop	Bookstore	Bakery	Grocery Store
1	3	Aldgate	Hotel	Gym / Fitness Center	Restaurant	Coffee Shop	Wine Bar	Salad Place
2	3	Aldwych	Café	Pub	Sandwich Place	Hotel	Theater	Bookstore
3	3	Angel	Coffee Shop	Pub	Food Truck	Vietnamese Restaurant	Cocktail Bar	Hotel
4	3	Balham	Coffee Shop	Grocery Store	Pub	Fast Food Restaurant	Indian Restaurant	Bakery
5	3	Barbican	Food Truck	Coffee Shop	Pub	Gym / Fitness Center	Hotel	Italian Restaurant
6	0	Barnes	Pub	Park	Farmers	Indie Movie	French	Thai

Table 22.

Last step is to visualize this in map. Map includes markers that shows nearby places. This is shown below.

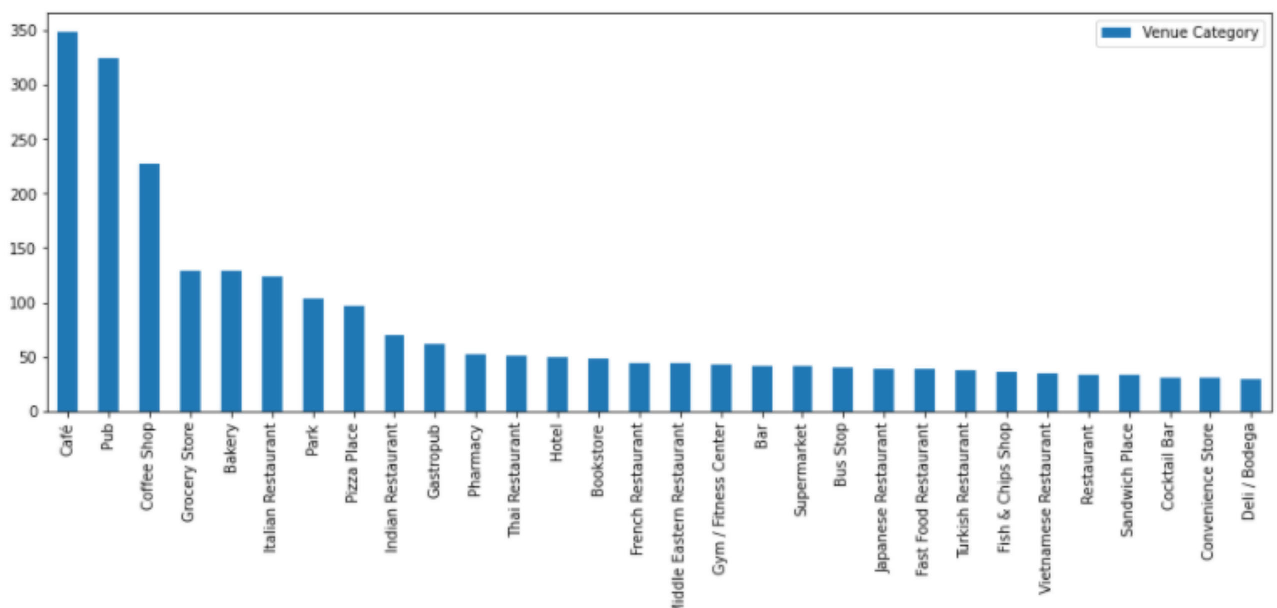


Cluster 0 after k mean is shown below.

Cluster Labels	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
0	Acton	Café	Pub	Coffee Shop	Bookstore	Bakery	Grocery Store	Italian Restaurant
0	Barnes	Pub	Park	Farmers Market	Indie Movie Theater	French Restaurant	Thai Restaurant	Gastropub
0	Battersea	Café	Indian Restaurant	Pub	Bar	Breakfast Spot	Restaurant	Supermarket
0	Bedford Park	Pub	Café	Coffee Shop	Bookstore	Bakery	Italian Restaurant	Creperie
0	Belsize Park	Café	Pub	Bakery	Ice Cream Shop	Italian Restaurant	Museum	Deli / Bodega

Table 23.

Graph after k mean is shown below.



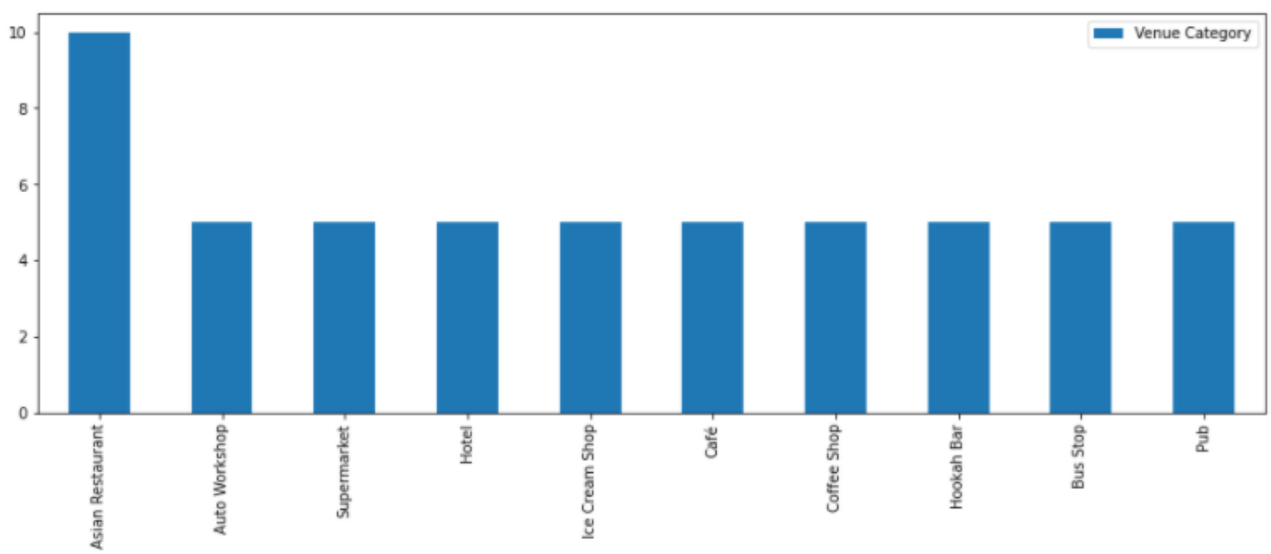
Graph after K mean

Cluster 1 after k mean is shown below.

Cluster Labels	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
1	Colindale	Asian Restaurant	Ice Cream Shop	Coffee Shop	Auto Workshop	Supermarket	Hookah Bar	Hotel
1	Grahame Park	Asian Restaurant	Ice Cream Shop	Coffee Shop	Auto Workshop	Supermarket	Hookah Bar	Hotel
1	Kingsbury	Asian Restaurant	Ice Cream Shop	Coffee Shop	Auto Workshop	Supermarket	Hookah Bar	Hotel
1	The Hyde	Asian Restaurant	Ice Cream Shop	Coffee Shop	Auto Workshop	Supermarket	Hookah Bar	Hotel
1	West Hendon	Asian Restaurant	Ice Cream Shop	Coffee Shop	Auto Workshop	Supermarket	Hookah Bar	Hotel

Table 24.

Graph after k mean is shown below.



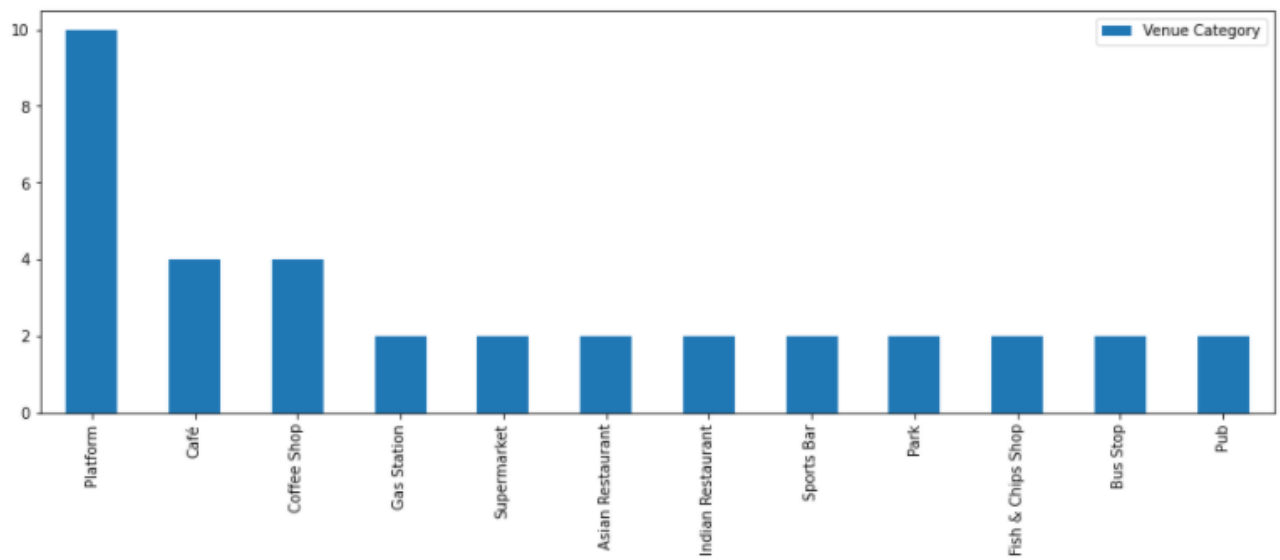
Graph of cluster 1 after K mean

Cluster 2 after k mean is shown below.

Cluster Labels	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
2	Selhurst	Platform	Café	Coffee Shop	Sports Bar	Pub	Park	Indian Restaurant
2	South Norwood	Platform	Café	Coffee Shop	Sports Bar	Pub	Park	Indian Restaurant

Table 25.

Graph after k mean is shown below.



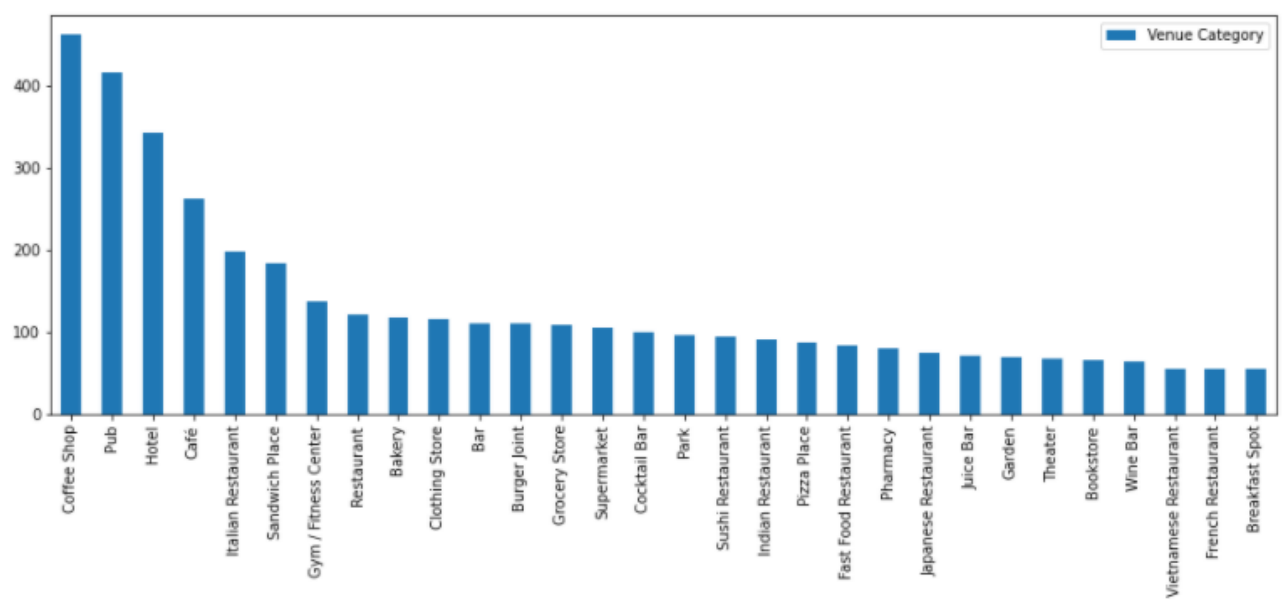
Graph of cluster 2 after K mean

Cluster 3 after k mean is shown below.

Cluster Labels	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 Mos Commo Venu
3	Aldgate	Hotel	Gym / Fitness Center	Restaurant	Coffee Shop	Wine Bar	Salad Place	Garde
3	Aldwych	Café	Pub	Sandwich Place	Hotel	Theater	Bookstore	Japanes Restaurai
3	Angel	Coffee Shop	Pub	Food Truck	Vietnamese Restaurant	Cocktail Bar	Hotel	Pai
3	Balham	Coffee Shop	Grocery Store	Pub	Fast Food Restaurant	Indian Restaurant	Bakery	Italia Restaurai
3	Barbican	Food Truck	Coffee Shop	Pub	Gym / Fitness Center	Hotel	Italian Restaurant	Vietnames Restaurai

Table 26.

Graph after k mean is shown below.



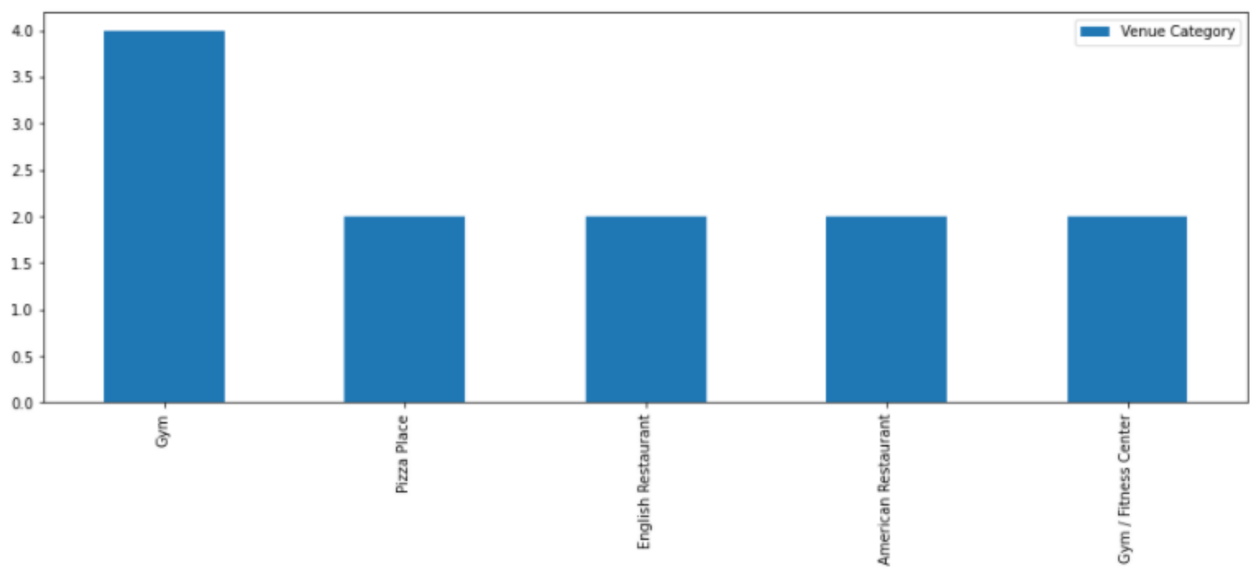
Graph of cluster 3 after K mean

Cluster 4 after k mean is shown below.

Cluster Labels	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
4	Chingford	Gym	Gym / Fitness Center	American Restaurant	English Restaurant	Pizza Place	Zoo Exhibit	Flea Market
4	Highams Park	Gym	Gym / Fitness Center	American Restaurant	English Restaurant	Pizza Place	Zoo Exhibit	Flea Market

Table 27.

Graph after k mean is shown below.



Graph of cluster 4 after K mean

5. Results and Discussion

Starting a business is not only investing money. It is more than that. If you are going to start a new business that a complete and detailed analysis is required. It is time consuming effort. This effort will have long lasting fruits. In this case study a businessman in London has decided to open a new café in London. This man is looking for a place that has large amount of customers and low number of competitors. This is not an easy task. We have chosen places that are in 500 radius. After choosing places we got their location. We performed data cleaning operation. This step will give required data that is necessary for this analysis. Than finding most top 10 most common places after that we applied k mean algorithm to make cluster. Our aim is to find common places in each cluster. After performing all these steps we found cluster0 is most suitable for this business because it has large potential customers and low competitors. We have cluster 3 also that have shown somehow similar results as cluster 0. But after detailed analysis we choose cluster 0.

This analysis will gives us area that is most suitable for this business. Majority of the offices are besides this area so there are chances that this business will established in a limited time.

Our analysis used python language and many libraries that we have already discussed in above area of this project. Multiple libraries are used for performing analysis. Two API's are used.

One is geocoder and second is ArcGis. API's are mostly paid. They provide easy access of the data. Data is very important in business decision. Geocoder API gives location of any place. It gives two coordinates of location e.g longitude and latitude. All these analysis is performed in a python language. Therefore we need a tool that supports python language and provides all libraries that we needed. This model used unsupervised learning algorithm. Due to this we cannot deduce which area is suitable for this shop in one iteration. We run this algorithm multiple iteration and found that cluster 0 is best suitable for this case study. Because it gives that area that has maximum number of customers and less number of competetors.

We used k-mean algorithm to find clusters in our data. This s unsupervised algorithm. Different cluster has made during our experiment. Out of all clusters, cluster 0 and 3 performed better. It filters out areas that would prove beneficial for this project. On further iteration cluster 0 is proved best. Therefore area that should be chosen for coffee shop would belong to cluster 0.

Only venue does not matter. Other important factors for business success would be customer satisfaction, customer loyalty, customer retention, market share, and the firm's ability to charge a premium price [1]. Quality is another metric for business success in case of food business [1]. Brand loyalty is another important factor for business success [2]. Majority of customers are loyal to specific brands. Atmosphere is added as one of the quality factors.

Another important factor for customer satisfaction is quality on food being served [5]. Service quality means how a customer would be treated at restaurants. Service quality is another important feature for business success [5]. If there are more than one transaction channel for making payment that would also attracts more customers [5]. Because now a days people like to pay bills through card. If your business does not support cash payment method other than traditional way than it will not be attracted much customers. If you have good service, atmosphere but price is not optimal than customers would be distracted to other shops. It is necessary to decide optimal prices for your products. Prices of products can be deiced by making a survey about that product.

Employees working on your company affects customer's satisfaction greatly [6]. If your employee is not satisfied than they will not treat customers rightly. This will result in loss of customers.

6. Conclusion

In this project we focused on area selection for new business related to coffee. We have performed detailed analysis that will give area that has maximum number of potential customers and less competitors. We used unsupervised algorithm e,g K- mean to find clusters. We used this algorithm in more than one iteration to find best area for this business. Cluster 0 is proved to be best among all clusters. We used paid API;s in our project. Geo coder API is used to get location that are in 500 meter radius. We then chose 10 best common places that are near places that are in 500 radius. After performing this analysis we have found two clusters that are suitable for this business cluster 0 and cluster 3. After running one more iteration we found that area suitable for this business should belong to cluster 0.

7. References

[1]. Ko, W. H., & Chiu, C. P. (2006). A new coffee shop location planning for customer satisfaction in Taiwan. *International Journal of the Information Systems for Logistics and Management*, *2*(1), 55-62.

[2]. Susanty, A., & Kenny, E. (2015). The relationship between brand equity, customer satisfaction, and brand loyalty on coffee shop: Study of Excelso and Starbucks. *ASEAN Marketing Journal*, 14-27.

[3]. Shin, C. S., Hwang, G. S., Lee, H. W., & Cho, S. R. (2015). The impact of Korean franchise coffee shop service quality and atmosphere on customer satisfaction and loyalty. *The Journal of Business Economics and Environmental Studies*, *5*(4), 47-57.

[4]. Lee, W. S., Moon, J., & Song, M. (2018). Attributes of the coffee shop business related to customer satisfaction. *Journal of Foodservice Business Research*, *21*(6), 628-641.

[5]. Suhud, U., Allan, M., Wibowo, S. F., Sabrina, E., & Willson, G. (2020). Measuring customer satisfaction of a café and coffee shop colony at a traditional market. *Journal of Foodservice Business Research*, *23*(1), 78-94.

[6]. Sadeli, A. H., Utami, H. N., & Rahmanisa, R. (2016). Does Customer Satisfaction on Product Quality Illustrates Loyalty of Agricultural Product. *International Journal of Applied Business and Economics Research*, *4*(1), 223-232.