Vigneshkumar
IBM Data Science Capstone Project

IBM Final Capstone Project Report

Introduction

The current world crisis revolves around the invisible enemy known as the coronavirus (COVID-19). This disease has its name written in history by forcing the WHO to state it as a pandemic state as the number of infections and death tolls has been exponentially increasing. Some countries such as Italy have even gone to the stage of lockdown. There are others that are yet to be under lockdown. Based on the articles from dailymail [1] and mirror [2], it is evident that people are highly risk of being exposed to the virus with all the drastic measures such as closing of restaurants and schools are taking place.

With this in mind, I will like be doing an analysis of neighborhoods specifically in Enfield, London to identify the availability of essential amenities in the neighbourhoods than others in this crucial time. This project will also identify the neighborhoods that do not have access or have access to some amenities. This can aid Government or social volunteers to focus on the efficient resource distribution in these areas in times of need.

So, the next question is what are the essential amenities?

With reference to recommended list from the business insider article [3], the essential items are food sources that are available in grocery stores or any type of markets and medicines from pharmacies for self-care and hygiene purposes like Panadol and sanitizers.

Business Problem

Therefore, the **business problem is**: To identify and explore the similarities and differences in food availability and medical needs among the neighborhoods in Enfield, London based on the number of amenities (markets, pharmacies and etc.) in the times of COVID-19 outbreak.

Data Collection

The data required are the neighbourhoods in Enfield which can be can obtained scraped from the Wikipedia table as shown below using BeautifulSoup.

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

https://en.wikipedia.org/wiki/List_of_areas_of_London

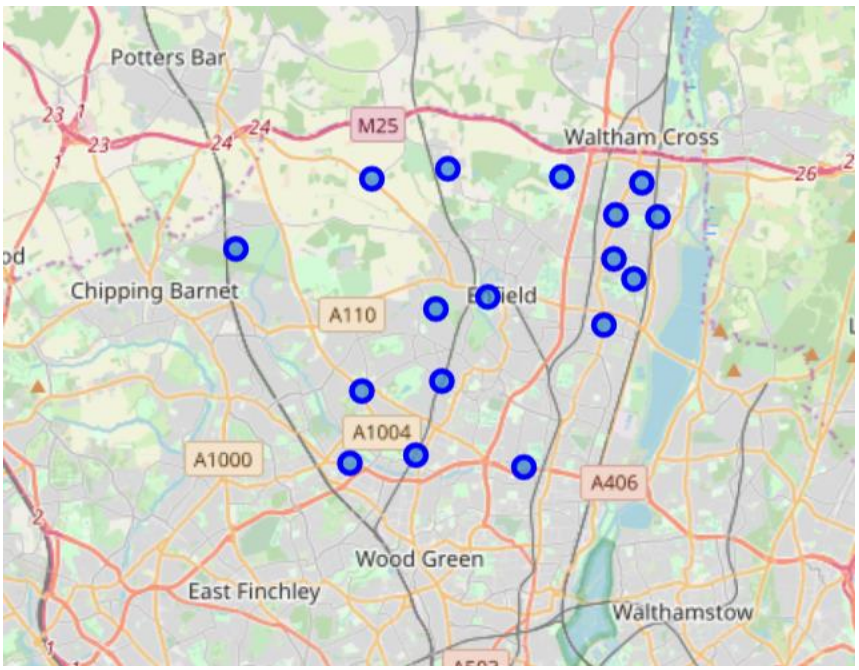
The neighbourhoods are considered to be various locations of the Enfield corresponding to the same postcode which can be obtained from the first column in the table above. The coordinates of these neighbourhoods are obtained manually by individually searching for neighbourhood coordinates.

The preview of the outcome of the data set is as follows:

	Postcode	Borough	Neighbourhood	Grid	Latitude	Longitude
0	EN1, EN2	Enfield	Enfield Town	TQ325965	51.651685	-0.086136
1	EN1, EN3	Enfield	Ponders End	TQ355955	51.645627	-0.045920
2	EN2	Enfield	Botany Bay	TQ297992	51.676601	-0.125585
3	EN2	Enfield	Crews Hill	TQ315995	51.678879	-0.099455
4	EN2, EN3	Enfield	Bulls Cross	TQ342994	51.677342	-0.060466

There are a total of 17 neighbourhoods namely Enfield Town, Ponders End, Botany Bay, Crews Hill, Bulls Cross, Brimsdown, Enfield Highway, Enfield Lock, Enfield Wash, Freezywater, Hadley Wood, Arnos Grove, Palmers Green, Southgate, Grange Park, Winchmore Hill and Edmonton.

The location of all 17 neighbourhoods in Enfield are observed in the map below.



For each neighbourhood in Enfield, the venues can be obtained by connecting to Foursquare API surrounding 1500m radius. This distance is approximated as it is not considered safe to be travelling for long distances especially under lockdown.

The preview of data set with venues is as follows:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Enfield Town	51.651685	-0.086136	Enfield Town Park	51.649998	-0.083855	Park
1	Enfield Town	51.651685	-0.086136	Waitrose & Partners	51.651602	-0.084114	Supermarket
2	Enfield Town	51.651685	-0.086136	Goodlooking Optics	51.652223	-0.079490	Optical Shop
3	Enfield Town	51.651685	-0.086136	PizzaExpress	51.652475	-0.080832	Pizza Place
4	Enfield Town	51.651685	-0.086136	Caffè Nero	51.652061	-0.081949	Coffee Shop

Methodology

Collection of input data is done according to the methods mentioned in section Data Collection. This includes gathering location data through Wikipedia through manual work and web scraping and also using Foursquare to obtain a total of 200 venues in the neighbourhoods within a radius of 1500m. Here 1500m as radius is assumed to be safe distance that one can travel during lockdown amid the crisis.

After obtaining the venues, it is essential to filter out the venues by "Venue's Category" from the data obtained through Foursquare API. Here, only the venues associated to grocery stores/markets and pharmacies are needed. The specific types of venue categories are identified from the Foursqaure website by looking for stores and markets. After filtering, the data can be then clustered through k-means to explore how similar or how different the neighbourhoods are.

Based on the code below,

```
#Total number unique venues
print('There are {} uniques categories.'.format(len(Enfield_venues['Venue Category'].unique())))
```

```
There are 101 uniques categories.
```

there are 101 venue categories. Based on the list of venue cateogories, the following will be focus in this project:

1) For medical needs: 'Pharmacy'
2) For groceries/food/basic items: 'Supermarket', 'Grocery Store','Convenience Store','Department Store'

Taking a quick look on the number of times the neighbourhood name will show an overview of how each neighbourhood is doing.

For medical purposes based on the table below:

```
Enfield_MedicalVenues.groupby('Neighborhood').count()
```

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Enfield Highway	1	1	1	1	1	1
Enfield Town	2	2	2	2	2	2
Grange Park	1	1	1	1	1	1
Palmers Green	1	1	1	1	1	1
Ponders End	1	1	1	1	1	1
Southgate	1	1	1	1	1	1

Overall, only 7 out of 17 neighbourhoods contain minimum of 1 pharmacy within 1500m.

For other basic needs:

```
Enfield_BasicNeedVenues.groupby('Neighborhood').count()
```

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Arnos Grove	5	5	5	5	5	5
Brimsdown	4	4	4	4	4	4
Bulls Cross	1	1	1	1	1	1
Edmonton	7	7	7	7	7	7
Enfield Highway	7	7	7	7	7	7
Enfield Lock	7	7	7	7	7	7
Enfield Town	8	8	8	8	8	8
Enfield Wash	5	5	5	5	5	5
Freezywater	9	9	9	9	9	9
Grange Park	5	5	5	5	5	5
Palmers Green	9	9	9	9	9	9
Ponders End	7	7	7	7	7	7
Southgate	3	3	3	3	3	3
Winchmore Hill	5	5	5	5	5	5

Overall, 14 out of 17 neighbourhoods contain minimum of 1 basic need shop within 1000m. Neighbourhoods Freezywater and Palmers Green contains the highest number of such categories.

From these 2 tables, it is clear that the neighbourhoods require more medical attention that basic items. Nevertheless, it will be great if we are able to cluster the neighbourhoods for better resource allocation rather than identifying for each neighbourhood which is time consuming.

For clustering, the model used is K-Means. The input data to this model is the total number of each selected venue categories each neighbourhood has. Through coding, a sample of the table data will look as follows:

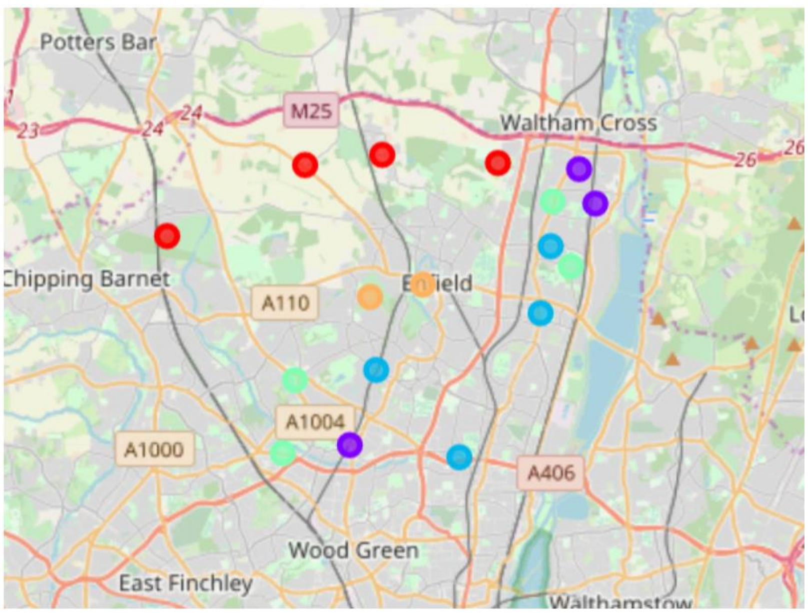
```python
# Find the total number of each venue category per neighbourhood
Enfield_grouped = Enfield_onehot.groupby('Neighborhood').sum().reset_index()
Enfield_grouped.head()
```

	Neighborhood	Pharmacy	Supermarket	Grocery Store	Convenience Store	Department Store
0	Arnos Grove	0	0	4	1	0
1	Botany Bay	0	0	0	0	0
2	Brimsdown	0	2	2	0	0
3	Bulls Cross	0	0	1	0	0
4	Crews Hill	0	0	0	0	0

Through the K-Means model imported in python, the data clusters are generated. The final result of 5 clusters shows will be discussed in the next session.

Results

The final result of the clusters is shown below.



Among the 5 clusters, the highest number of neighbourhoods found similar is 4 while the lowest number of neighbourhoods found is 2.

The following shows the individual cluster components.

Vigneshkumar
IBM Data Science Capstone Project

Cluster 0:

	Borough	Neighbourhood	Cluster Labels	Pharmacy	Supermarket	Grocery Store	Convenience Store	Department Store
2	Enfield	Botany Bay	0	0	0	0	0	0
3	Enfield	Crews Hill	0	0	0	0	0	0
4	Enfield	Bulls Cross	0	0	0	1	0	0
10	Enfield	Hadley Wood	0	0	0	0	0	0

Cluster 1:

	Borough	Neighbourhood	Cluster Labels	Pharmacy	Supermarket	Grocery Store	Convenience Store	Department Store
7	Enfield	Enfield Lock	1	0	1	6	0	0
9	Enfield	Freezywater	1	0	3	6	0	0
12	Enfield	Palmers Green	1	1	2	6	1	0

Cluster 2:

	Borough	Neighbourhood	Cluster Labels	Pharmacy	Supermarket	Grocery Store	Convenience Store	Department Store
1	Enfield	Ponders End	2	1	5	2	0	0
6	Enfield	Enfield Highway	2	1	4	3	0	0
15	Enfield	Winchmore Hill	2	0	3	2	0	0
16	Enfield	Edmonton	2	0	4	2	1	0

Cluster 3:

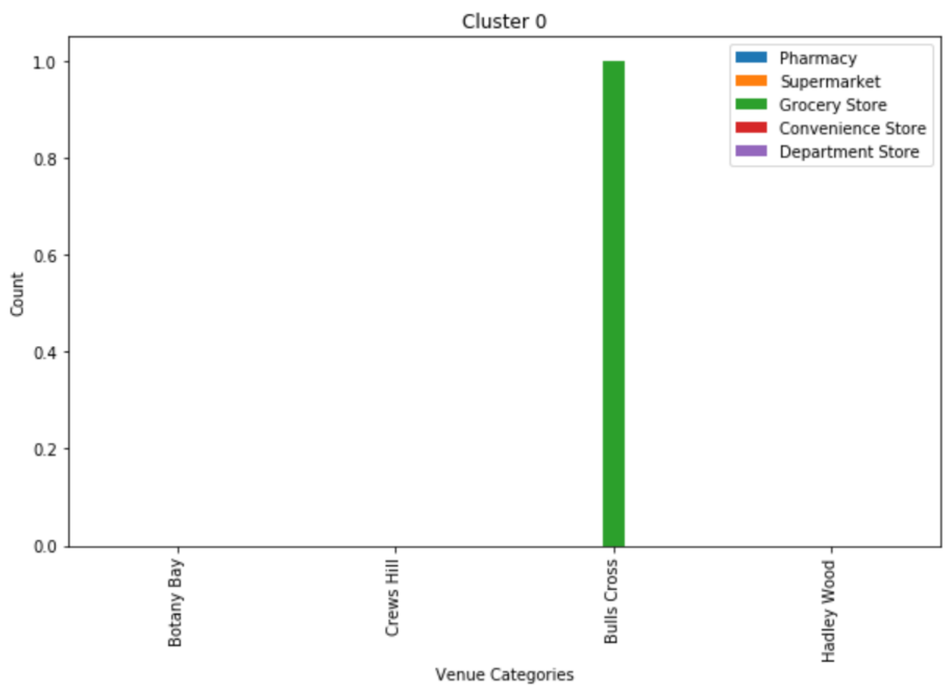
	Borough	Neighbourhood	Cluster Labels	Pharmacy	Supermarket	Grocery Store	Convenience Store	Department Store
5	Enfield	Brimsdown	3	0	2	2	0	0
8	Enfield	Enfield Wash	3	0	1	4	0	0
11	Enfield	Arnos Grove	3	0	0	4	1	0
13	Enfield	Southgate	3	1	1	2	0	0

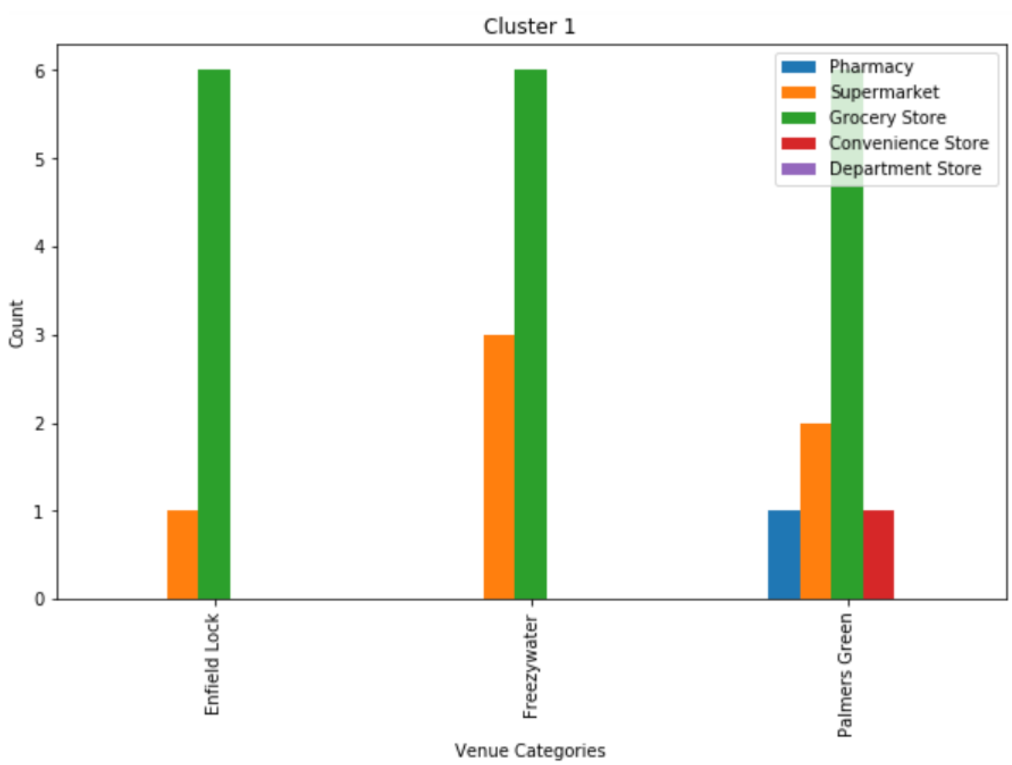
Cluster 4:

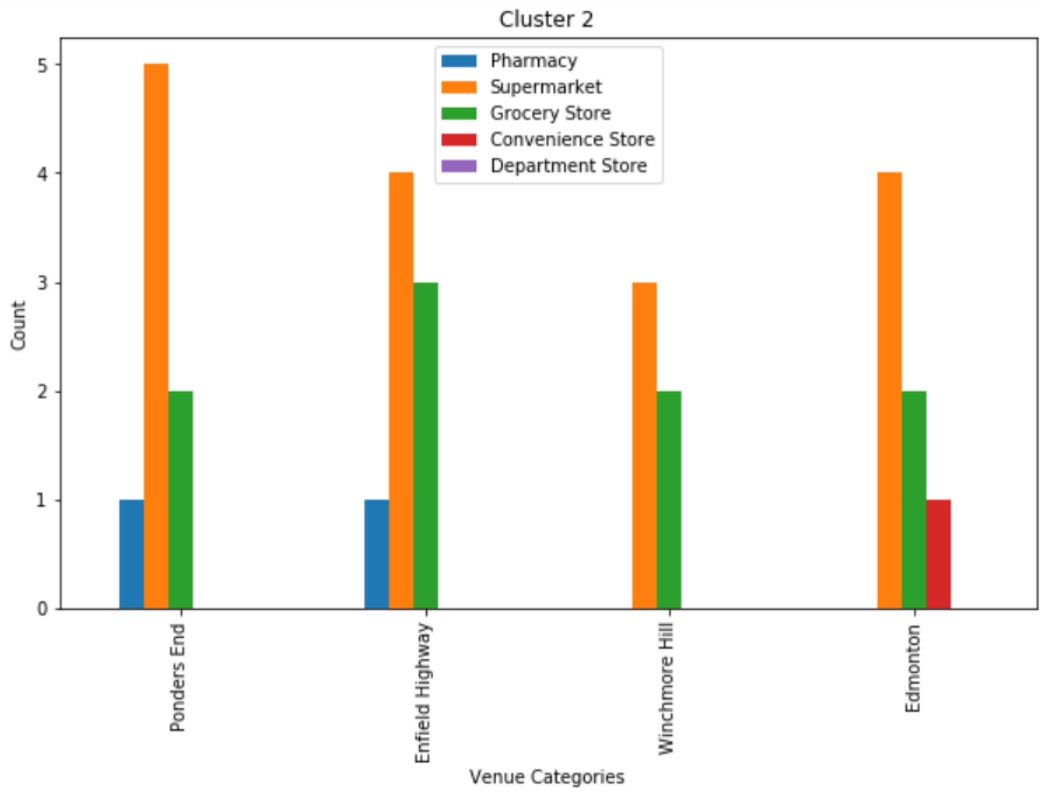
	Borough	Neighbourhood	Cluster Labels	Pharmacy	Supermarket	Grocery Store	Convenience Store	Department Store
0	Enfield	Enfield Town	4	2	3	3	0	2
14	Enfield	Grange Park	4	1	2	1	0	2

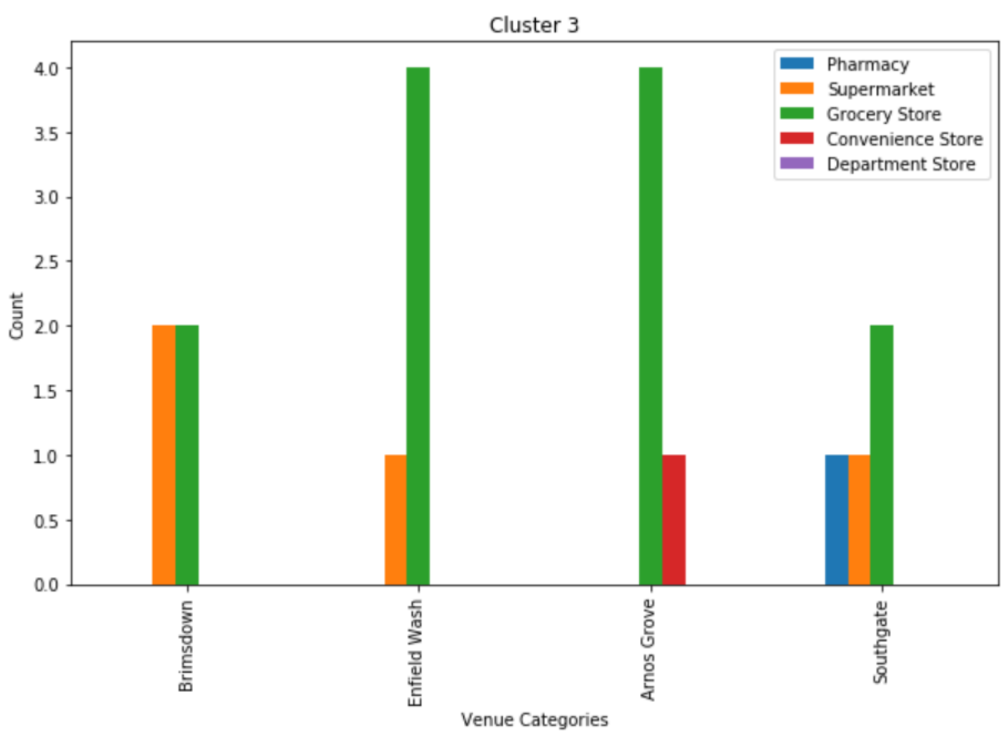
Discussions & Observations

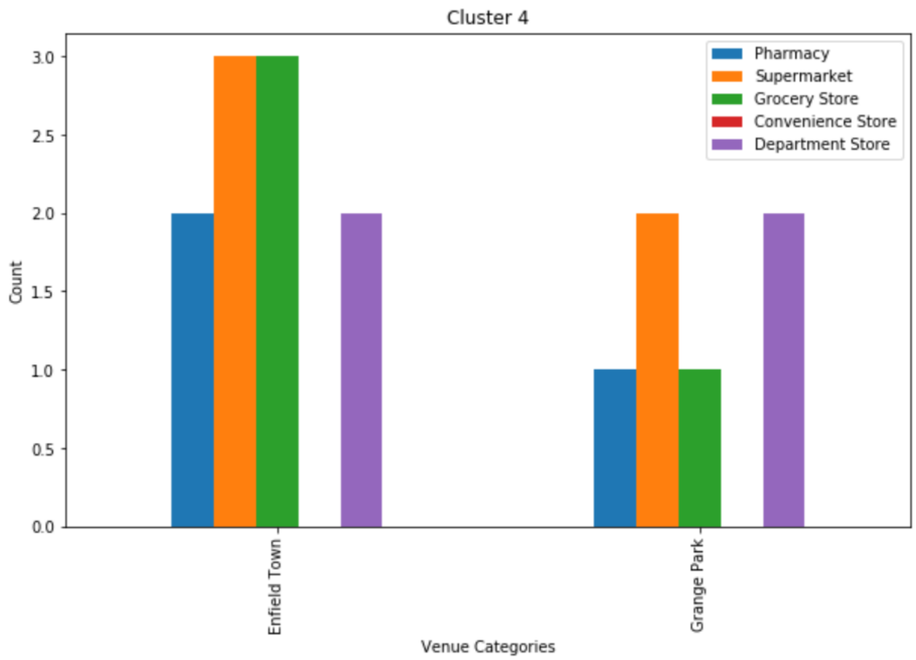
This section will explore on the similarities within each cluster group.






Cluster 2
Pharmacy
Supermarket
Grocery Store
Convenience Store
Department Store
Count
0
1
2
3
4
5
Ponders End
Enfield Highway
Winchmore Hill
Edmonton
Venue Categories


Cluster 3
Pharmacy
Supermarket
Grocery Store
Convenience Store
Department Store
Count
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Brimsdown
Enfield Wash
Arnos Grove
Southgate
Venue Categories



Based on thee above plots, we can easily find the similarity based on the colour and heights of the bars in plots.

Cluster 0:

The similarity is that these neighbourhoods lack bascic amenities. Only 1 of the neighbourhoods contains 1 grocery store. Hence, more focus on resource allocation is required for the neighbourhoods in this cluster.

Cluster 1:

In this cluster, the neighbourhoods have grocery store as the highest number of amenities followed by supermarket. In fact, these places have the maximum number of grocery stores.

Cluster 2:

On the contrast to cluster 1, the neighbourhoods in cluster 2 contains more number of supermarkets than grocery stores.

Cluster 3:

Similar to cluster 1, the neighbourhoods contain grocery stores as highest number of amentities but lower when compares to cluster 1.

Cluster 4:

There is a wide range of amenities in each neighbourhood and their count is approximately same as compared to other clusters. In fact, this is the only cluster where all neighbourhoods have pharmacies.

Analysing each cluster, Cluster 4 seems to have most variety of amenities. Cluster 0 has approximately none. Hence, maximum attention is required for monitoring and the resource allocation to Cluster 0. Though all neighbourhoods in Custer 4 contains pharmacies, all clusters require more medical attention has the highest number of pharmacies observed within one neighbourhood is 2.

Obstacles

There was one major obstacle when implementing this project and it was to collect the coordinates of the neighbourhoods in Enfield. Geocoder and web scraping of many webpages were not able to provide coordinates for all neighbourhoods. Therefore, they have been manually obtained by using GeoHack in Wikipedia for each neighbourhood. Advance web scraping can be used to obtain the coordinates by automating the process of having to extract data from multiple web pages.

Improvements

One of the improvements that can be done in this project is having more factors apart from the type of venues. One such factor is the population density of the neighbourhoods. Taking that into consideration, the total number of each type of venue categories (amenities) can be standardized by dividing by the population density in each neighbourhood. This will produce a more accurate data allowing better decisions to be made for resource distribution. For neighbourhoods with lower population density, it is sufficient to have lower number of amenities and it does not mean that these places require more attention by the Government officials for resources.

Another improvement that can be made is further analyzing the clustering. Further analysis can be done by clustering the neighbourhoods with the distance range of the amenities. This leads to exploration of the distribution of venue categories (amenities) within each neighbourhood. This helps to refine resource allocation as the officials with the knowledge on such data.

Last but not least, it does takes time to identify most suitable number of clusters required. Perhaps, a better model will be desity based modelling which can filter out the anomalies and cluster those that are really similar.

Conclusion

Based on the analysis, it has been found that the neighbourhoods in Cluster 0 requires most amount of resources. All clusters require medical attention even for Cluster 4 as it is observed that there are not many of such amenities. Though the result is acceptable, it is limited as there is no consideration of population to see number of amenities per head. An advanced analysis can include the number of amenities within intervals of distances to understand distribution of existing resources.

References

[1] M. Robinson, "Prince George and Princess Charlotte's school becomes the latest to send pupils home for coronavirus isolation - as NINE shut - despite advice it's 'unnecessary' - offices close and sports fixtures are cancelled across UK," Mailonline, 27 02 2020. [Online]. Available: https://www.dailymail.co.uk/news/article-8047195/Is-Britain-heading-coronavirus-lockdown-amid-confusion-dangers-Europe.html. [Accessed 14 03 2020].

[2] G. Diebelius, "Coronavirus: Entire UK cities could be placed on lockdown in bid to contain spread," Mirror, 02 03 2020. [Online]. Available: https://www.mirror.co.uk/news/uk-news/coronavirus-entire-uk-cities-could-21612813. [Accessed 14 03 2020].

[3] K. Warren, "What to buy if you’re quarantined at home during the coronavirus pandemic," Business Insider US, 11 03 2020. [Online]. Available: https://www.businessinsider.sg/what-to-buy-for-home-quarantine-coronavirus-2020-3?r=US&IR=T. [Accessed 15 03 2020].