Segmenting a London Neighborhood to Discover Suitable Streets to Pedestrianise During Coronavirus

1. Introduction

1.1 Background

The UK government has introduced an 'Eat out to Help Out' scheme to encourage the UK public to eat out in restaurants and help get the economy moving again after coronavirus. One of the issues with this is that far fewer people are able to eat at a restaurant at a given time because of social distancing measures.

This can be mitigated against by closing roads and allowing restaurants to set out tables on the road as well as inside. This has happened in some streets in London, where a high volume of restaurants on a given road has made this a feasible solution. However, there are far more areas where this might be applicable.

1.2 Problem

Can we identify areas within a given London neighbourhood that would be suitable for road closures based on the type of businesses that operate in the given area?

1.3 Stakeholders

The stakeholders for this project would be the UK government, London Mayor and London citizens.

2. Data

2.1 Data Sources

The neighbourhood analysed for this project is the SW12 postcode area. The reason being; this neighborhood already has had some streets already pedestrianised, therefore the criteria for doing so can be better understood.

https://crystalroof.co.uk/postcodes/SW12 has been crawled, to build a dataframe consisting of all the postcodes in the SW12 neighbourhood.

The google maps API has been used to obtain latitude and longitude data for all of the postcodes in SW12.

The Foursquare API has been used to search for venues within a radius of each of these postcode locations.

2.2 Data Cleaning

The postcode data was crawled from the web using the beautiful soup package and then transformed into a data frame:

	postCode
0	SW12 0AA
1	SW12 0AB
2	SW12 0AF
3	SW12 0AH
4	SW12 0AJ

The latitude and longitude data was pulled from the foursquare API, where the results set of the api call was reduced to just latitude and longitude columns:

	postCode	latitude	longitude
0	SW12 0AA	51.448100	-0.142647
1	SW12 0AB	51.447770	-0.143745
2	SW12 0AF	51.447115	-0.142642
3	SW12 0AH	51.448185	-0.139805
4	SW12 0AJ	51.447860	-0.136869

Then the foursquare API was used to obtain all of the venues within a 150m radius of each postcode location. This was then created as a dataframe with the following columns:

	postCode	Postcode Latitude	Postcode Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	SW12 0AA	51.448100	-0.142647	La Retraite RC School	51.448123	-0.143380	High School
1	SW12 0AA	51.448100	-0.142647	St Bernadette's	51.447086	-0.143057	School
2	SW12 0AA	51.448100	-0.142647	Big Yellow Self Storage Company	51.446458	-0.144089	Storage Facility
3	SW12 0AA	51.448100	-0.142647	Pocklington Resource Centre	51.449378	-0.147123	Building
4	SW12 0AA	51.448100	-0.142647	BBC (Balham Banter Centre)	51.449820	-0.144897	Comedy Club
...	...	...	...	...	...	...	...
15570	SW12 9ZS	51.446374	-0.151477	Hob	51.445194	-0.151075	Café
15571	SW12 9ZS	51.446374	-0.151477	Little Dessert Shop	51.445732	-0.150623	Dessert Shop
15572	SW12 9ZS	51.446374	-0.151477	Charlotte Cave Store & Services	51.445299	-0.150931	Health & Beauty Service
15573	SW12 9ZS	51.446374	-0.151477	Escape The Daily Grind	51.444892	-0.150731	Café
15574	SW12 9ZS	51.446374	-0.151477	The Umi	51.444867	-0.151144	Japanese Restaurant

3. Methodology

A k-means clustering algorithm was used to group the postcodes into 5 groups, based on the similarity in frequency of the most common venues within a 150m radius of each postcode.

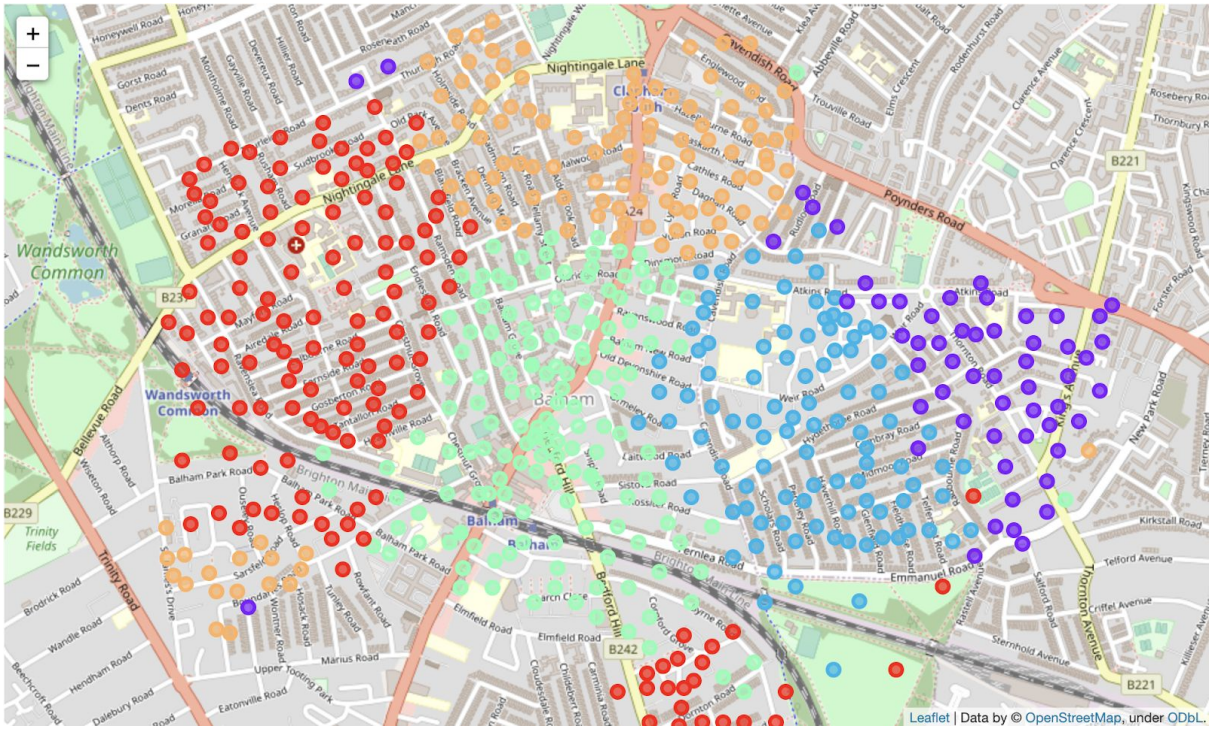
The cluster labels were then applied to each postcode in the dataframe:

	postCode	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 Mos Commo Venu
0	SW12 0AA	51.448100	-0.142647	2	Office	Building	School	Hardware Store	Italian Restaurant	Grocery Store	Pet Service	Convenience Store	Comedy Club	Storag Facilit
1	SW12 0AB	51.447770	-0.143745	2	Office	Hardware Store	Building	Pub	Italian Restaurant	Medical Center	Event Space	Pet Service	Recruiting Agency	Convenienc Stor
2	SW12 0AF	51.447115	-0.142642	2	Office	Hardware Store	Building	Italian Restaurant	Tech Startup	High School	Grocery Store	Meeting Room	Pet Service	Event Spac
3	SW12 0AH	51.448185	-0.139805	1	Bus Stop	Building	Playground	School	Grocery Store	Daycare	Elementary School	Nursery School	Church	Frenc Restauran
4	SW12 0AJ	51.447860	-0.136869	1	Bus Stop	Building	Grocery Store	Park	School	Playground	Elementary School	Warehouse	Storage Facility	Distributio Cente

A clustering algorithm was used, since there is pre-existing knowledge concerning streets that have already been shut down. Therefore, a cluster that shares characteristics with these streets, must be the one we are looking for.

4. Results

The SW12 area was successfully segmented into 5 separate clusters, where each postcode within the area has been assigned a cluster. The breakdown of clusters can be visualised in the map below:



5. Discussion

The segmentation of the SW12 neighborhood appears to divide the area into 5 distinct sections, where the sections are similar based on the venues within them.

Looking at the most common venues within these sections, we can propose a 'type' for each section - which would indicate which types of venues one would expect to be in and therefore allow us to identify the area suitable for street shutdown.

Cluster 0 - Red

Most common venues:

- Church
- Office
- Doctor's

Type: Personal services

Cluster 1 - Purple

Most common venues:

- Bus stop
- Tennis court
- School

Type: Residential

Cluster 2 - Blue

Most common venues:

- Office
- School
- Hardware store

Type: Functional services

Cluster 3 - Green

Most common venues:

- Cafe
- Restaurant
- Pub
- Barbershop

Type: Eating out

Cluster 4 - Orange

Most common venues:

- Grocery Store
- Supermarket

Type:

- Retail

Therefore, we would recommend that postcodes within cluster 3 be considered for pedestrianisation.

5. Conclusion

This study analysed the venues within a given London postcode, to assess the feasibility of using machine learning algorithms to predict the suitability of sub-sections of this area for pedestrianisation during coronavirus.

Through the implementation of a k-means clustering algorithm, the SW12 areas was segmented into 5 main clusters, where one cluster was highlighted as being suitable for pedestrianisation. This was cluster three, where the most common venues were cafes, restaurants pubs and barbershops.