Application of Foursquare Data to an Educational Consultancy Business

Contents

Introduction

This project demonstrates how Foursquare data combined with machine learning could be used to offer enhanced services for overseas clients in an international education consultancy business.

Large educational consultancies tend to offer additional services to clients besides academic tutoring, for example, admissions interview preparation, institution insights, academic trends etc. Most if not all of these services focus on helping the students navigate through the admissions processes of major universities, however, once an offer is attained, little help is given to help the students adjust to the new environment. This can be an especially daunting experience for students who have never lived abroad before.

In this analysis, we aim to help students ease into their new environment by using Foursquare data to provide a more informed decision on the best area to seek accommodation for the academic year.

Business Problem

The target audience for this problem is for educational consultancies who wish to see how programmatic access to location data could potentially be used to offer bespoke services to clients in the area of identifying ideal areas for renting. This could also be used for computer savvy students with the same aim of filtering for areas to live.

As a proof of concept, let's assume that a client has attained an unconditional offer to study at Imperial College London, United Kingdom, and has asked us, the educational consultants, to help with identifying the best areas to seek accommodation. The client has also specified that due to moderate budget constraints, they'd like to find accommodation in zone 2 regions and because of the client's athletic nature; they'd like to be within close proximity to parks or gyms. Since the client is new to the area, they would like to live near tube stations.

In summary:

1. Zone 2 area
2. Close to parks or gyms
3. Close to tube stations

London is the capital and largest city of England and the United Kingdom with a population of 8.982 million according to the 2019 census. London is divided into 9 zones but most of it fits into zones 1-6. Central London, which is the most expensive area, is zone 1 and zone 2 is the ring around zone 1. According to geopy's Nominatim, the geographical coordinate of London, United Kingdom are 51.5073219, -0.1276474.

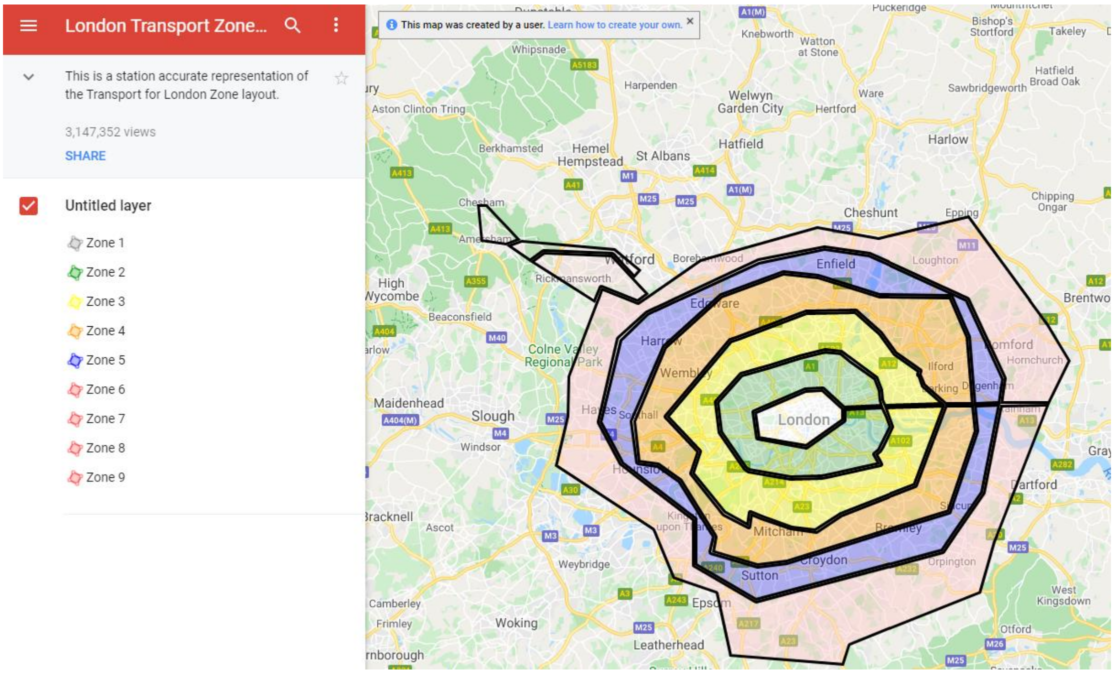

Fig 1: Image source: [Google Maps]()

Data Description

The data which we will use to solve this problem will come from two sources:

1. Since the client has expressed an interest in living near tube stations, we will use this information as a starting point for our analysis. Tube station information can be found from many sources, the one which we've chosen is https://www.doogal.co.uk/london_stations.php since it conveniently provides not only longitude and latitude information, but also the zones that each station is in. We will use the zone information to filter for stations which are in zone 2 and longitude and latitude information for identifying venues around the area.

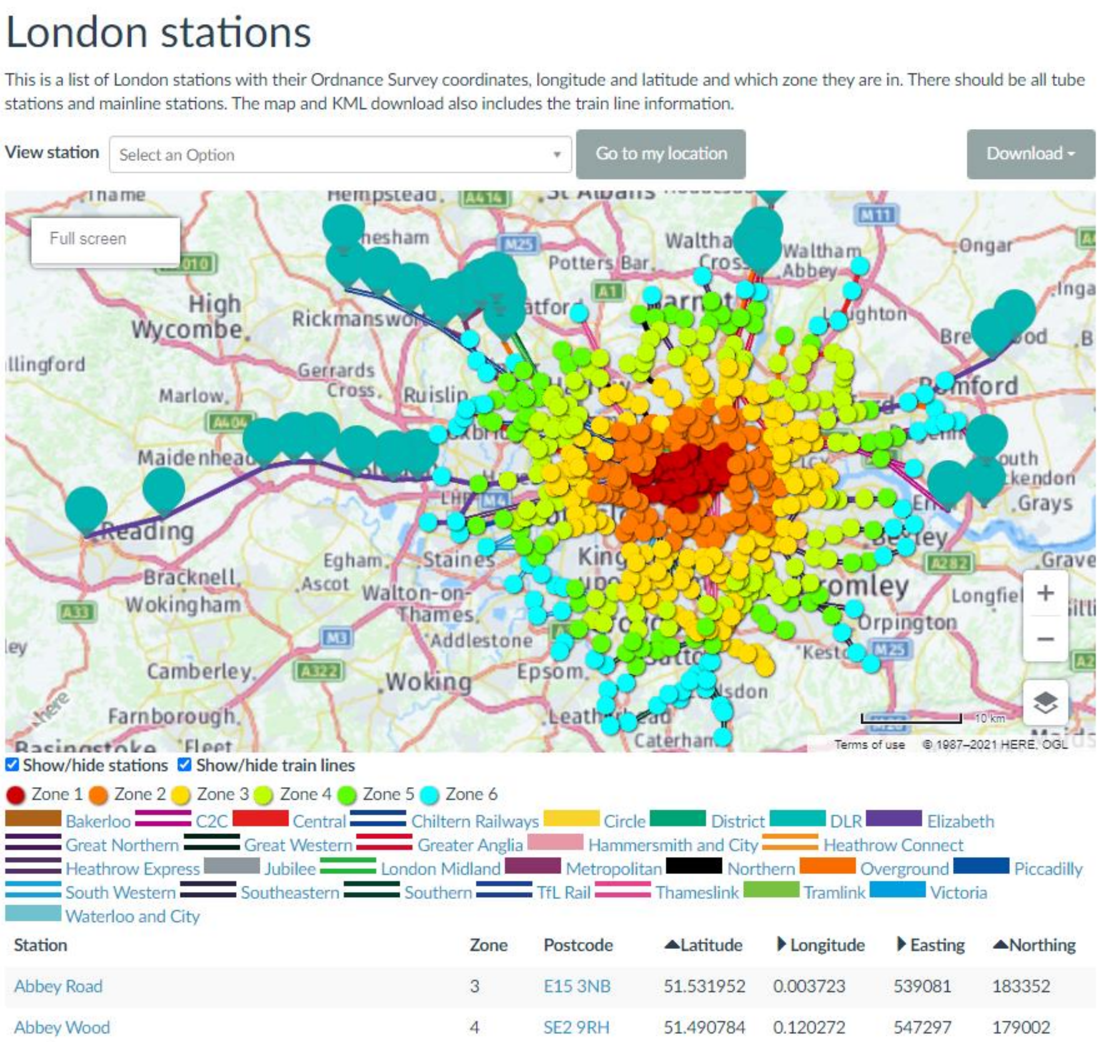

Fig 2: Image Source: Doogal

2. We will use Foursquare data to get information about the venues in the surrounding areas for each station. The client has expressed an interest of living in the areas close to parks or gyms. Before any analysis South Kensington would be an excellent area to live in as it's near Hyde Park but it's in zone 1 where rent would exceed the client's budget.

Methodology

The first step is to scrape the Doogal website so that we have data on all the tube stations. This is done using Python's `request` library for extracting from the site and `BeautifulSoup` library for parsing into a data frame.

Let's have a look at the first 10 records

	Neighborhood	zone	latitude	longitude
0	Abbey Road	3	51.531952	0.003723
1	Abbey Wood	4	51.490784	0.120272
2	Acton Central	2	51.508758	-0.263430
3	Acton Main Line	3	51.516887	-0.267690
4	Acton Town	3	51.503071	-0.280303
5	Addington Village	3,4,5,6	51.356239	-0.032665
6	Addiscombe	3,4,5,6	51.379808	-0.073213
7	Albany Park	5	51.435816	0.126445
8	Aldgate	1	51.514342	-0.075627
9	Aldgate East	1	51.515082	-0.073001

The client is only interested in zone 2 so we use the second column to filter for such records. Note that an area could within multiple zones so we adjust our query accordingly.

The student is only interested in living in zone 2, records before: 653, records after: 148

	Neighborhood	zone	latitude	longitude
2	Acton Central	2	51.508758	-0.263430
11	All Saints	2	51.510477	-0.012625
18	Archway	2,3	51.565491	-0.135122
21	Arsenal	2	51.558541	-0.105500
33	Barons Court	2	51.490229	-0.213430
34	Battersea Park	2	51.477318	-0.148124
47	Belsize Park	2	51.550191	-0.163974
49	Bermondsey	2	51.497961	-0.064330
51	Bethnal Green	2	51.527192	-0.055392
52	Bethnal Green Rail	2	51.524283	-0.060014

Next, we conduct some exploratory data analysis by examining visually the areas identified using the `folium` library.

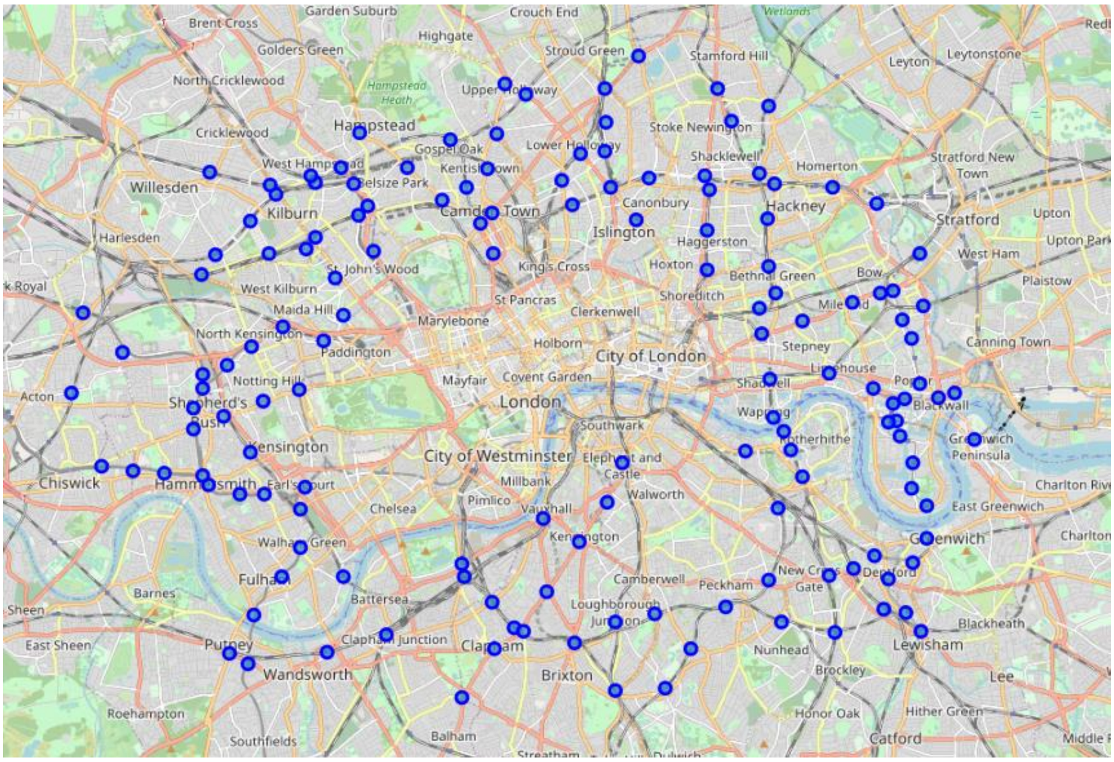

Fig 3: From Folium, see Python notebook for the source code.

Visually we can see that this is extracting the correct locations since zone 2 forms a concentric circle around Central London. We then use the Foursquare API to extract data on venues which are within 500 meters from each of the train stations and collect them into a `london_venues` data frame. Let's examine the first 10 records:

```
[67]: print(london_venues.shape)
      london_venues.head()
```

```
(5938, 7)
```

[67]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Acton Central	51.508758	-0.26343	The Station House	51.508877	-0.263076	Pub
1	Acton Central	51.508758	-0.26343	The Rocket	51.508772	-0.263787	Pub
2	Acton Central	51.508758	-0.26343	Acton Park	51.508595	-0.261573	Park
3	Acton Central	51.508758	-0.26343	Acton Centre	51.506608	-0.266878	Gym / Fitness Center
4	Acton Central	51.508758	-0.26343	MrBakeme	51.508452	-0.268543	Creperie

We employ the one hot encoding technique and pandas manipulation to create a data frame which outlines the frequency of various venue types by tube station:

```
london_grouped = london_onehot.groupby('Neighborhood').mean().reset_index()
london_grouped
```

	Neighborhood	ATM	African Restaurant	American Restaurant	Animal Shelter	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store
0	Acton Central	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000
1	All Saints	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000
2	Archway	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000
3	Arsenal	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000
4	Barons Court	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000
5	Battersea Park	0.000000	0.000000	0.000000	0.050000	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Belsize Park	0.000000	0.000000	0.025000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000
7	Bermondsey	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000
8	Bethnal Green	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.018519	0.000000	0.037037	0.000000	0.000000
9	Bethnal Green Rail	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.017241	0.000000	0.000000
10	Blackwall	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000
11	Bow Church	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.071429	0.000000	0.000000
12	Bow Road	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000
13	Brixton	0.000000	0.010000	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.010000	0.000000	0.000000
14	Brockley	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000

	Neighborhood	Arts & Entertainment	Asian Restaurant	Athletics & Sports	Australian Restaurant	Austrian Restaurant	Auto Garage	BBQ Joint	Badminton Court	Bagel Shop	Bakery	Bar	Baseball Field
0	Acton Central	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.052632	0.000000	0.00
1	All Saints	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.090909	0.00
2	Archway	0.000000	0.030303	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.030303	0.00
3	Arsenal	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
4	Barons Court	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
5	Battersea Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.150000	0.00
6	Belsize Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.075000	0.000000	0.00
7	Bermondsey	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.035714	0.000000	0.00
8	Bethnal Green	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.018519	0.00
9	Bethnal Green Rail	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.034483	0.017241	0.00
10	Blackwall	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.055556	0.00
11	Bow Church	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.071429	0.00
12	Bow Road	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.062500	0.00
13	Brixton	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.020000	0.000000	0.000000	0.010000	0.010000	0.00
14	Brockley	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00

This matrix like structure allows us to see for example that 9% of venues around the All Saints station are bars, which may be suitable for students who wish to frequent such venues. The entire data frame is still too big to act as a summary, it has 148 rows and 328 columns, so we will, so we compute high level statistics for each tube station by extracting the top 5 venue types for each area, this offers an easier to read summary than the above.

```
----Acton Central----
                venue  freq
0                 Pub  0.21
1  Gym / Fitness Center  0.21
2         Coffee Shop  0.11
3       Train Station  0.11
4                Park  0.05
```

```
----All Saints----
               venue  freq
0      Grocery Store  0.09
1 Chinese Restaurant  0.09
2         Steakhouse  0.09
3                Bar  0.09
4               Park  0.09
```

```
----Archway----
                venue  freq
0         Coffee Shop  0.18
1       Grocery Store  0.15
2                 Pub  0.12
3         Pizza Place  0.09
4  Italian Restaurant  0.06
```

```
----Arsenal----
                  venue  freq
0                   Pub  0.09
1                  Café  0.09
2  Gym / Fitness Center  0.09
3           Pizza Place  0.09
4        Soccer Stadium  0.09
```

There are still 148 stations from which we need to choose from, too much information to sift through. Hence, we utilize the k-means clustering unsupervised machine learning technique to help identify clusters within the data and select the cluster which best matches the client's requirements.

Results

We applied the K-means clustering technique to the data frame of tube station with their nearby venues to segregate the areas into one of five categories. The interpretation of each of these categories requires further interpretation, but let's visually inspect the clusters using `folium` again.

	Neighborhood	zone	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
2	Acton Central	2	51.508758	-0.263430	3	Gym / Fitness Center	Pub	Coffee Shop	Creperie	Mini Golf	Park	Bakery	Grocery Store	Train Station	Supermarket
11	All Saints	2	51.510477	-0.012625	0	Chinese Restaurant	Park	Grocery Store	Sandwich Place	English Restaurant	Coffee Shop	Fish Market	Café	Outdoor Sculpture	Film Studio
18	Archway	2,3	51.565491	-0.135122	4	Coffee Shop	Grocery Store	Pub	Pizza Place	Café	Italian Restaurant	Sandwich Place	Asian Restaurant	Gym / Fitness Center	Farmers Market
21	Arsenal	2	51.558541	-0.105500	2	Gym / Fitness Center	Pub	Soccer Stadium	Café	Pizza Place	Sports Bar	Park	Clothing Store	Burger Joint	Bus Stop
33	Barons Court	2	51.490229	-0.213430	0	Pub	Café	Coffee Shop	Hotel	Grocery Store	Thai Restaurant	Tennis Court	Chinese Restaurant	Sandwich Place	Cocktail Bar

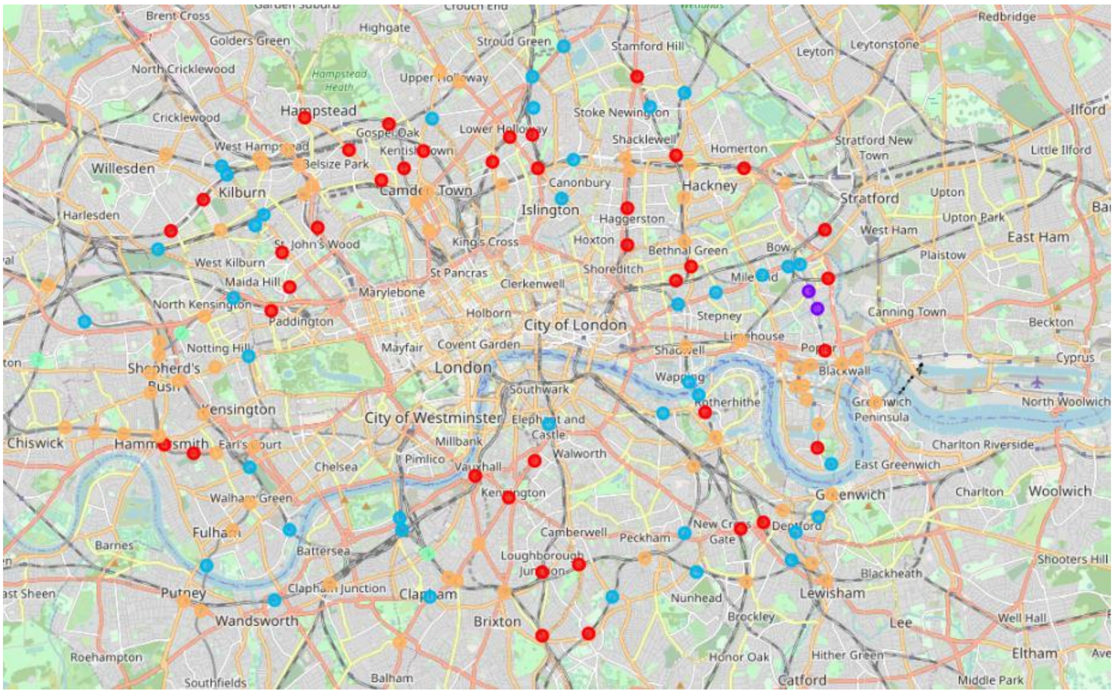

Fig 4: From Folium, see Python notebook for the source code.

We will discuss the interpretation of each cluster in the next segment.

Discussion

The first cluster, coloured red in the map above, appears to have a lot of pubs and restaurants nearby, this is more suited for a foody but not our client.

	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
21	Arsenal	Café	Pizza Place	Pub	Gym / Fitness Center	Soccer Stadium	Sports Bar	Park	Train Station	Mediterranean Restaurant	Burger Joint
33	Barons Court	Pub	Café	Coffee Shop	Hotel	Grocery Store	Thai Restaurant	Indian Restaurant	Performing Arts Venue	Garden	Chinese Restaurant
47	Belsize Park	Café	Pub	Bakery	Deli / Bodega	Pizza Place	Hotel	Coffee Shop	Bookstore	Thai Restaurant	Grocery Store
49	Bermondsey	Pub	Coffee Shop	Grocery Store	Café	Park	Farmers Market	Brazilian Restaurant	Gym / Fitness Center	Beer Bar	Gelato Shop
51	Bethnal Green	Pub	Coffee Shop	Café	Cocktail Bar	Park	Art Gallery	Beer Bar	Gym	Grocery Store	Hotel
52	Bethnal Green Rail	Pub	Café	Coffee Shop	Grocery Store	Pizza Place	Fast Food Restaurant	Park	Turkish Restaurant	Bakery	Gym
77	Bromley-by-Bow	Café	Historic Site	Supermarket	Metro Station	Pub	Waterfront	Bus Stop	Light Rail Station	Bike Rental / Bike Share	Canal Lock
78	Brondesbury	Coffee Shop	Café	Pub	Pizza Place	Park	Portuguese Restaurant	Theater	Deli / Bodega	Convenience Store	Doner Restaurant
79	Brondesbury Park	Park	Pub	Gym / Fitness Center	Coffee Shop	Farmers Market	Café	Japanese Restaurant	Flower Shop	Flea Market	Fishing Spot
109	Chalk Farm	Café	Bar	Italian Restaurant	Pub	Coffee Shop	Bakery	Pizza Place	Restaurant	Music Venue	French Restaurant

The second cluster, coloured purple in the map above, appears to have a lot of bars and pubs nearby. This may be more suited to students who wish to partake in pub crawls.

	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
34	Battersea Park	Bar	Pub	Coffee Shop	Sports Bar	Furniture / Home Store	Café	Gastropub	Chinese Restaurant	Grocery Store	Track Stadium
66	Bow Church	Pub	Convenience Store	Bus Stop	Metro Station	Grocery Store	Gym	Park	Burger Joint	Coffee Shop	Bar
67	Bow Road	Pub	Bus Stop	Park	Metro Station	Chinese Restaurant	Locksmith	Burger Joint	Hotel	Bar	Coffee Shop
96	Canonbury	Pub	Bus Stop	Café	Coffee Shop	Park	Thai Restaurant	Platform	Organic Grocery	Fish & Chips Shop	Modern European Restaurant
133	Clapton	Pub	Grocery Store	Chinese Restaurant	Pizza Place	Gym / Fitness Center	Park	Train Station	Turkish Restaurant	Breakfast Spot	Café
161	Deptford Bridge	Pub	Gym / Fitness Center	Grocery Store	Hotel	Japanese Restaurant	River	Light Rail Station	Park	Athletics & Sports	Vietnamese Restaurant
162	Devons Road	Pub	Light Rail Station	Fish & Chips Shop	Gym	Fried Chicken Joint	Dance Studio	Event Space	Exhibit	Falafel Restaurant	Farm

The third cluster, coloured blue in the map above, is an isolated cluster with pub, train station and bakery as its 1st, 2nd and 3rd nearest venue types respectively.

	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
303	Kensal Green	Pub	Train Station	Bakery	Portuguese Restaurant	Indian Chinese Restaurant	Film Studio	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant

The fourth cluster, coloured turquoise in the map above, seems to be predominated by coffee shops which doesn't fit our client profile.

	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
11	All Saints	Bar	Café	Park	Sandwich Place	English Restaurant	Outdoor Sculpture	Coffee Shop	Grocery Store	Chinese Restaurant	Fish Market
18	Archway	Coffee Shop	Grocery Store	Pub	Pizza Place	Italian Restaurant	Asian Restaurant	Café	Gastropub	Farmers Market	Sandwich Place
61	Blackwall	Hotel	Sandwich Place	Italian Restaurant	Coffee Shop	Outdoor Sculpture	Gym / Fitness Center	Chinese Restaurant	Bar	Steakhouse	Café
73	Brixton	Caribbean Restaurant	Market	Pizza Place	Coffee Shop	Pub	Cocktail Bar	Beer Bar	Fried Chicken Joint	Sandwich Place	Italian Restaurant
74	Brockley	Coffee Shop	Gastropub	Beer Store	Grocery Store	Indian Restaurant	Train Station	Deli / Bodega	Pizza Place	Pub	Convenience Store
87	Caledonian Road	Café	Grocery Store	Rental Car Location	Yoga Studio	Park	Coffee Shop	Theater	Breakfast Spot	Restaurant	Convenience Store
88	Caledonian Road and Barnsbury	Park	Coffee Shop	Grocery Store	Brewery	Rental Car Location	Fast Food Restaurant	Greek Restaurant	Bar	Gastropub	Caucasian Restaurant
89	Cambridge Heath	Coffee Shop	Pub	Hotel	Cocktail Bar	Flower Shop	Music Venue	Breakfast Spot	Park	Canal	Pizza Place
90	Camden Road	Pub	Coffee Shop	Café	Market	Burger Joint	Italian Restaurant	Ice Cream Shop	Beer Bar	Greek Restaurant	Music Venue
91	Camden Town	Coffee Shop	Pub	Greek Restaurant	Italian Restaurant	Pizza Place	Bar	Market	Tea Room	Beer Bar	Supermarket
92	Canada Water	Café	Pizza Place	Pharmacy	Pub	Coffee Shop	Sporting Goods Shop	Bar	Szechuan Restaurant	Street Food Gathering	Furniture / Home Store
93	Canary Wharf	Coffee Shop	Burger Joint	Sandwich Place	Italian Restaurant	Plaza	Hotel	Indian Restaurant	Bakery	Steakhouse	Turkish Restaurant

The fifth cluster, coloured yellow in the map above, has gyms and fitness centres as the most common venue. Parks are also within the vicinity but not as high up on the list as you can see from the table below. This cluster also only has three areas which significantly cuts down the number the 148 venues to search from.

	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
2	Acton Central	Gym / Fitness Center	Pub	Coffee Shop	Train Station	Bakery	Mini Golf	Park	Supermarket	Wine Shop	Creperie
331	Latimer Road	Gym / Fitness Center	Indoor Play Area	Latin American Restaurant	Coffee Shop	Park	Food & Drink Shop	Farm	Event Space	French Restaurant	Exhibit
590	Wandsworth Road	Gym / Fitness Center	Pub	Deli / Bodega	Grocery Store	Restaurant	Brewery	Dance Studio	Cupcake Shop	Exhibit	Falafel Restaurant

Thus, the student can focus their efforts in sourcing accommodation in Acton Central, Latimer Road or Wandsworth Road.

Conclusion

The purpose of this project was to provide a proof of concept of how a user with programmatic access to location data and a machine learning toolkit can use it to make informed geographically based decisions. Through our analysis, we have demonstrated how we can use such information to visualize various areas of the region, programmatically explore the area and apply machine learning techniques to act as a smart filter. The searching mechanism can definitely be improved. In this analysis, we simply used zones to filter for ideal regions which mean we could potentially be missing regions which might be deemed suitable in other zones. We could enhance this functionality by augmenting the filtering mechanism with rent prices, crime rate, student density etc.