Application of Foursquare Data to an Educational Consultancy Business

Applied Data Science Capstone, Coursera

Introduction

- Large educational consultancies tend to offer additional services to clients besides academic tutoring, e.g.:
    - Admissions interview preparation
    - Institution insights
    - Academic trends etc.
- We use Foursquare data to provide a more informed decision on the best area to seek accommodation for the academic year.
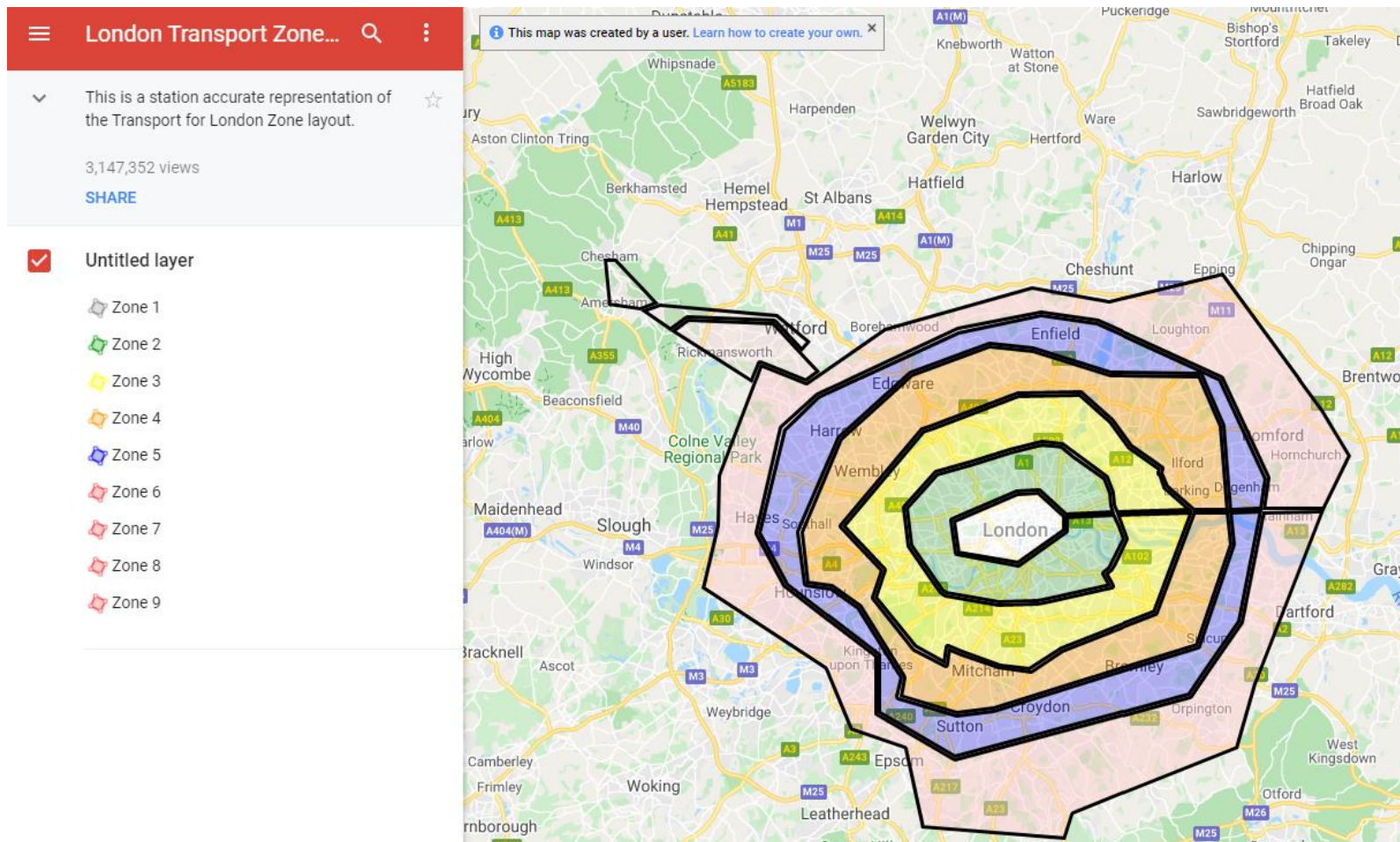
Business Problem

- A client has attained an unconditional offer to study at Imperial College London, United Kingdom and needs help identifying the best areas to seek accommodation.

- The client has provided the following specifications:

  – Accommodation should be in zone 2 regions

  – They'd like to be within close proximity to parks or gyms

  – They'd like to live near tube stations

Business Problem

- 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.

- The geographical coordinate of London, United Kingdom are 51.5073219, -0.1276474.

Business Problem

Data Description

- The client has expressed an interest in living near tube stations so 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 [Doogal](#)

- 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.

Data Description
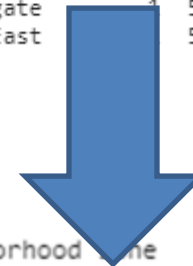
Data Description

- 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

- Scrape the Doogal website using request and BeautifulSoup

- The client is only interested in zone 2 so we use the second column to filter for such records.

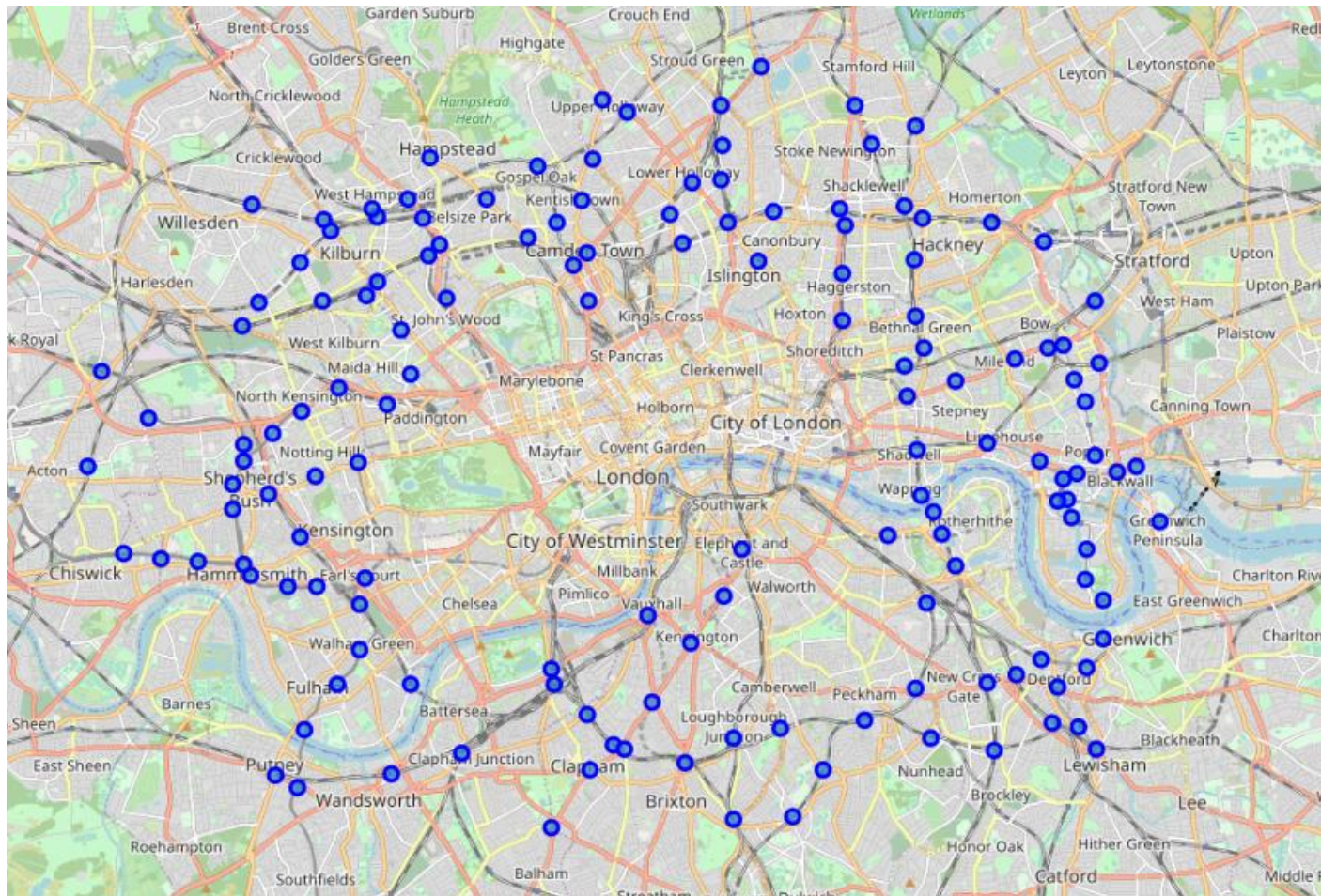- Note that an area could within multiple zones so we adjust our query accordingly.

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                   51.514342  -0.075627
9    Aldgate East              51.515082  -0.073001
```

```
     Neighborhood       ne   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
```

Methodology

Methodology

- 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.

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

       (5938, 7)
```
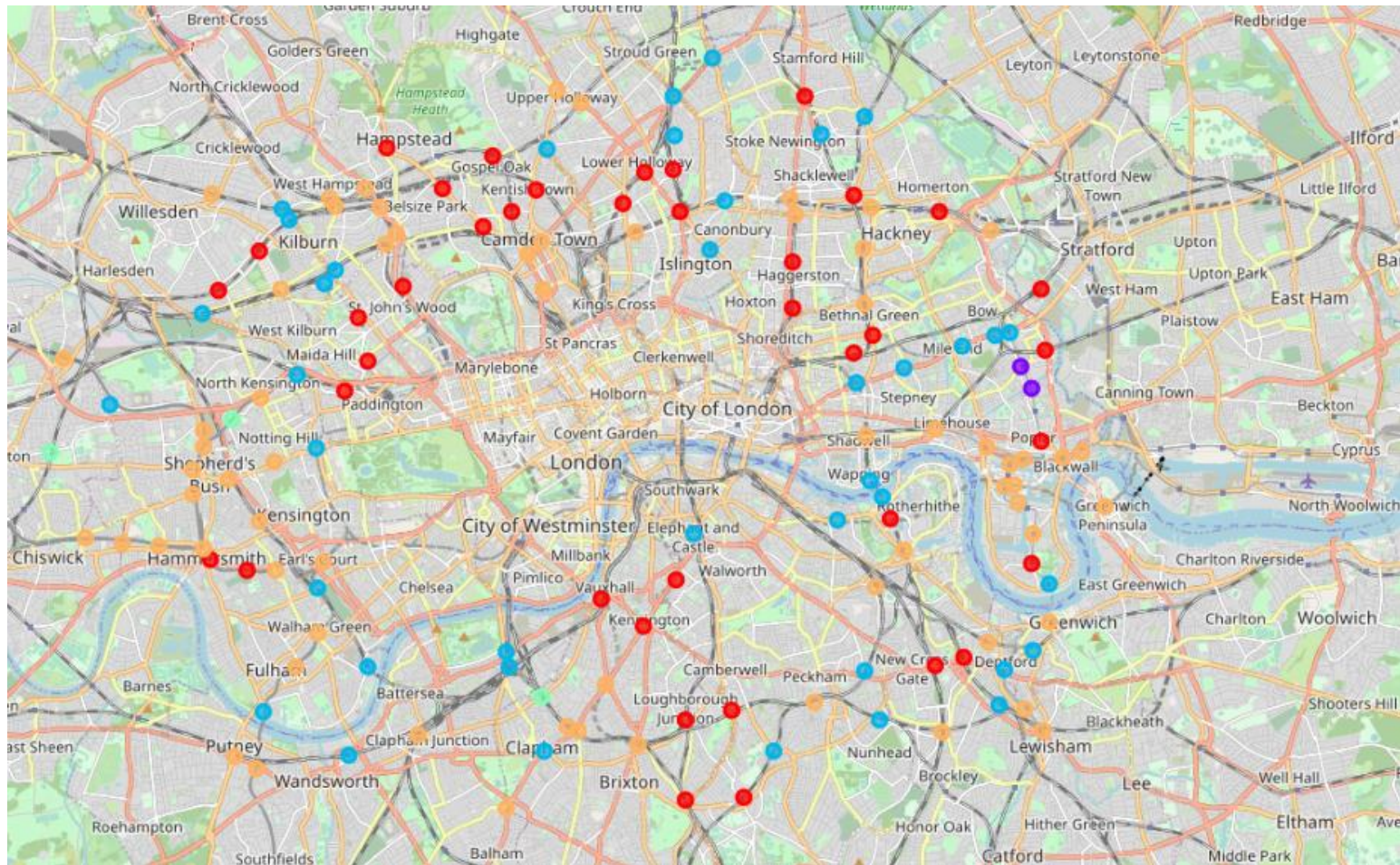
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Acton Central	51.508758	-0.26343	The Station House	51.508877	-0.263076	Pub
1	Acton Central	51.508758	-0.26343	The Rocket	51.508772	-0.263787	Pub
2	Acton Central	51.508758	-0.26343	Acton Park	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

Methodology

- 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

Methodology

Results

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

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

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

Results

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

- The fifth cluster, coloured **yellow** in the map in the previous slide, 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

Results

- Our K-means clustering analysis indicates that the following areas will suit the clients requirements:
  - Acton Central
  - Latimer Road
  - Wandsworth Road
- These three are tube stations which satisfy the requirement of being near the London transportation system.
- They also have gyms and recreational centres as the most common venues nearby which satisfies the client's requirement of being near parks or gyms.
- Finally, these are all in zone 2, which should help narrow down the searches for regions which fall within the client's budget constraint.
- This produces actionable insights and we can proceed by going to areas like RightMove to search for accommodation.

Results