Opening a new restaurant in London

Jazz Bhangu

3 July 2020

Introduction

The client for this project is a restaurant owner in West Hampstead, a busy borough in central London. Here, the client runs a successful Vegetarian restaurant. He now plans to open a second restaurant with the same offering and needs data driven insights on the best location. He is clear about the business strategy that he believes have led to his success:

- Busy, cosmopolitan area of London. The next location must be similar to West Hampstead in terms of overall amenity availability.
- Vegetarian food is an upcoming health trend but it is still not the norm in British society. Hence the owner believes people will only give this a try if there are limited other restaurant options in the vicinity. Once they try it, they love it – having discovered it tastes as good as meat….and hence they will keep coming back for more.
- Location:
    - Only interested in locations within very quick (500m max) walking distance of the main tube station.
    - Can open anywhere in London, but all else being equal, it would be ideal if the restaurant was on the same side of London as West Hampstead – for ease of managing both simultaneously.

Data acquisition and cleaning

There are distinct requirements for data provision and cleaning. Sources and methods have been identified to address these.

Geo-referencing

The primary geo-reference will be the postcode area eg. W1. This offers a reasonable segmentation of London; approx. 120 postcode areas. Alternative was to look at each Borough but there are only approx.. 30 so each covers too large an area for meaningful analysis.

- **London postcodes and associated borough names**: available from https://www.milesfaster.co.uk/london-postcodes-list.htm

- **Geo-spatial coordinates for centre of each postcode area in London**: available from https://www.freemaptools.com/download-uk-postcode-lat-lng.htm

London postcodes/borough names list and postcode geo-spatial coordinates were downloaded from the internet, into CSV files. These were then uploaded to Watson Studio which helpfully provides the Python code to read the data into the primary project Jupyter notebook

We do not have geo-locations of each tube-station. However, we can use the postcode area centre coordinates as an initial proxy, given that tube stations tend to be in the centre of the districts, from historical legacy. This can then be confirmed for the final choices using **Folium maps**.

Location data for local amenities

Location data is available using the 'explore' endpoint of Foursquare API. A standard sandbox account is sufficient for this. Client wants a max distance of 500m walking from the tube station (which we have assumed to be the postal area centre). Accordingly, a radius of 300m 'as the crow flies' will be used in the 'get' queries.

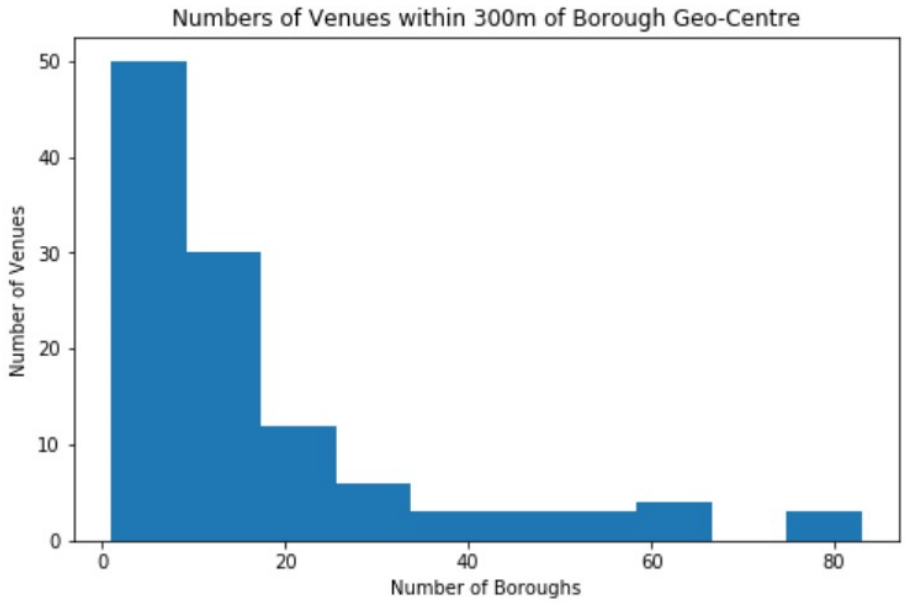
Methodology

Merging the files with names of boroughs/districts with the postcode geo-spatial coordinates file provides us with the master data set on which we can perform exploratory analysis.

	Postcode	Borough	latitude	longitude
0	E1	Whitechapel, Stepney, Mile End	51.51766	-0.05841
1	E1W	Wapping	51.50775	-0.05739
2	E2	Bethnal Green, Shoreditch	51.52939	-0.06080
3	E3	Bow, Bromley-by-Bow	51.52789	-0.02482
4	E4	Chingford, Highams Park	51.62196	-0.00339

Foursquare 'explore' endpoint is used to extract all venues within 300m of the postcode centre. These can then be viewed at an aggregated 'count' level for each postcode area.

Matplotlib allows to view a histogram of all the postcode areas in London:



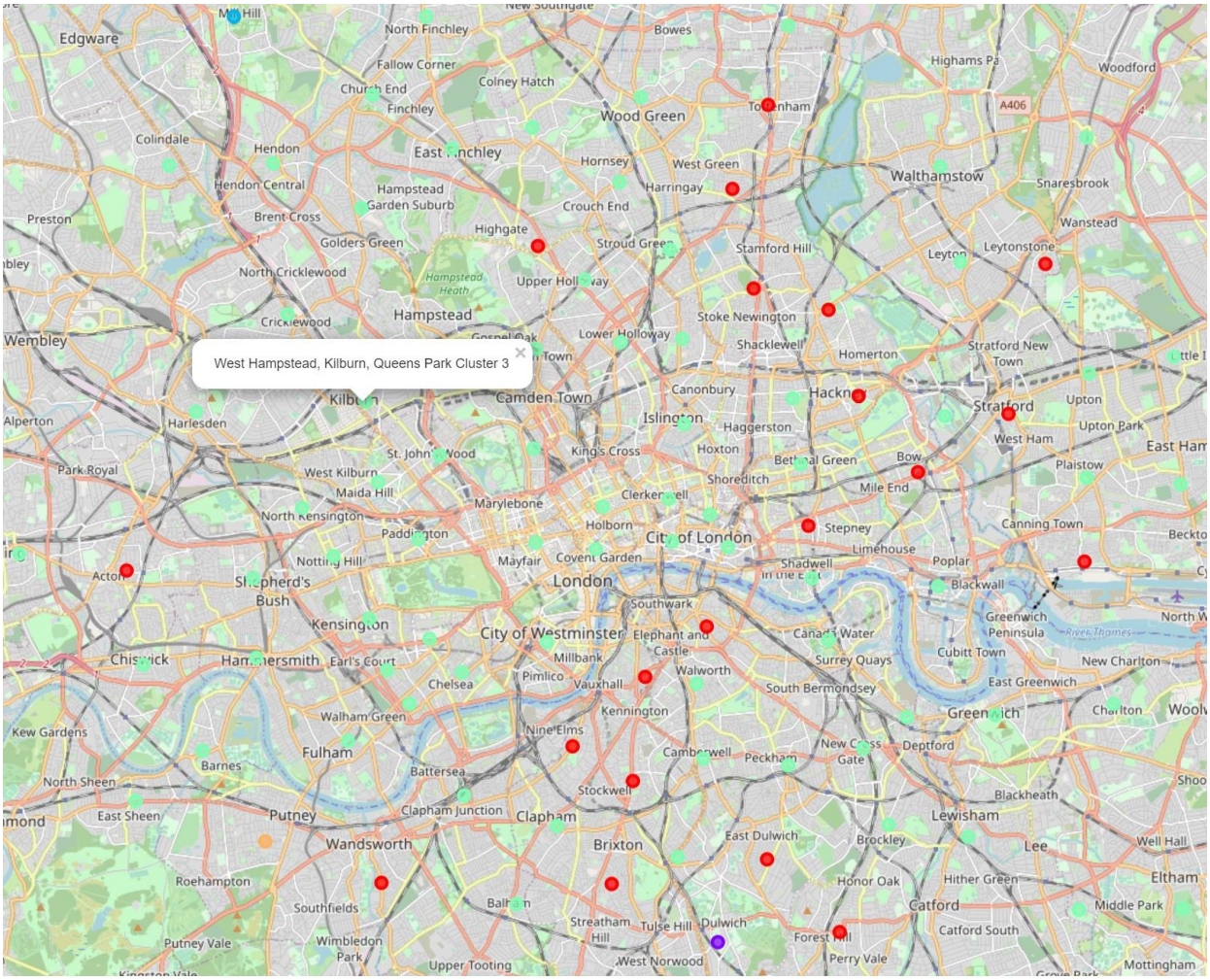
This informs us that there are significant differences between ‘Boroughs’ and hence we need to use a machine learning clustering algorithm to group similar Boroughs together and create an initial subset of those akin to West Hampstead, where the first restaurant is located.

After ‘One-Hot’ encoding and normalising using the mean values of counts, we can run K-means clustering algorithm. Trial and error showed that k=5 clusters provides an optimised level of granularity. Given further time, ‘elbow analysis’ could be used to calculate mean distance from centre of clusters to its constituent points and prove that k=5 is optimal.

Now, we have sufficient information to merge the cluster labels information with a dataframe that we created showing the top 10 venues for each ‘Borough’.

	Postcode	Borough	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Mo
0	E1	Whitechapel, Stepney, Mile End	51.517660	-0.058410	0	Pub	Coffee Shop	Asian Restaurant	Gym / Fitness Center	
1	E1W	Wapping	51.507750	-0.057390	3	Indian Restaurant	Bar	Convenience Store	Event Space	
2	E2	Bethnal Green, Shoreditch	51.529390	-0.060800	3	Grocery Store	Coffee Shop	Fast Food Restaurant	Sandwich Place	
3	E3	Bow, Bromley-by-Bow	51.527890	-0.024820	0	Burger Joint	Metro Station	Rental Car Location	Coffee Shop	
4	E4	Chingford, Highams Park	51.621960	-0.003390	3	Home Service	Yoga Studio	Exhibit	French Restaurant	
5	E5	Clapton	51.558930	-0.052330	0	Intersection	Café	Gym / Fitness Center	Pub	
6	E6	East Ham	51.525600	0.055830	3	Grocery Store	Café	Sandwich Place	Yoga Studio	
7	E7	Forest Gate, Upton Park	51.546780	0.027930	3	Fast Food Restaurant	Grocery Store	Bus Stop	Fish & Chips Shop	
8	E8	Hackney, Dalston	51.542020	-0.063150	3	Pool	Park	Yoga Studio	Falafel Restaurant	
9	E9	Hackney, Homerton	51.542580	-0.042930	0	Pub	Park	Deli / Bodega	French Restaurant	
10	E10	Leyton	51.568140	-0.011530	3	Cricket Ground	Convenience Store	Park	Farm	

This allows us to plot the clusters on a map of London, and we can immediately note that West Hampstead is in cluster 3.



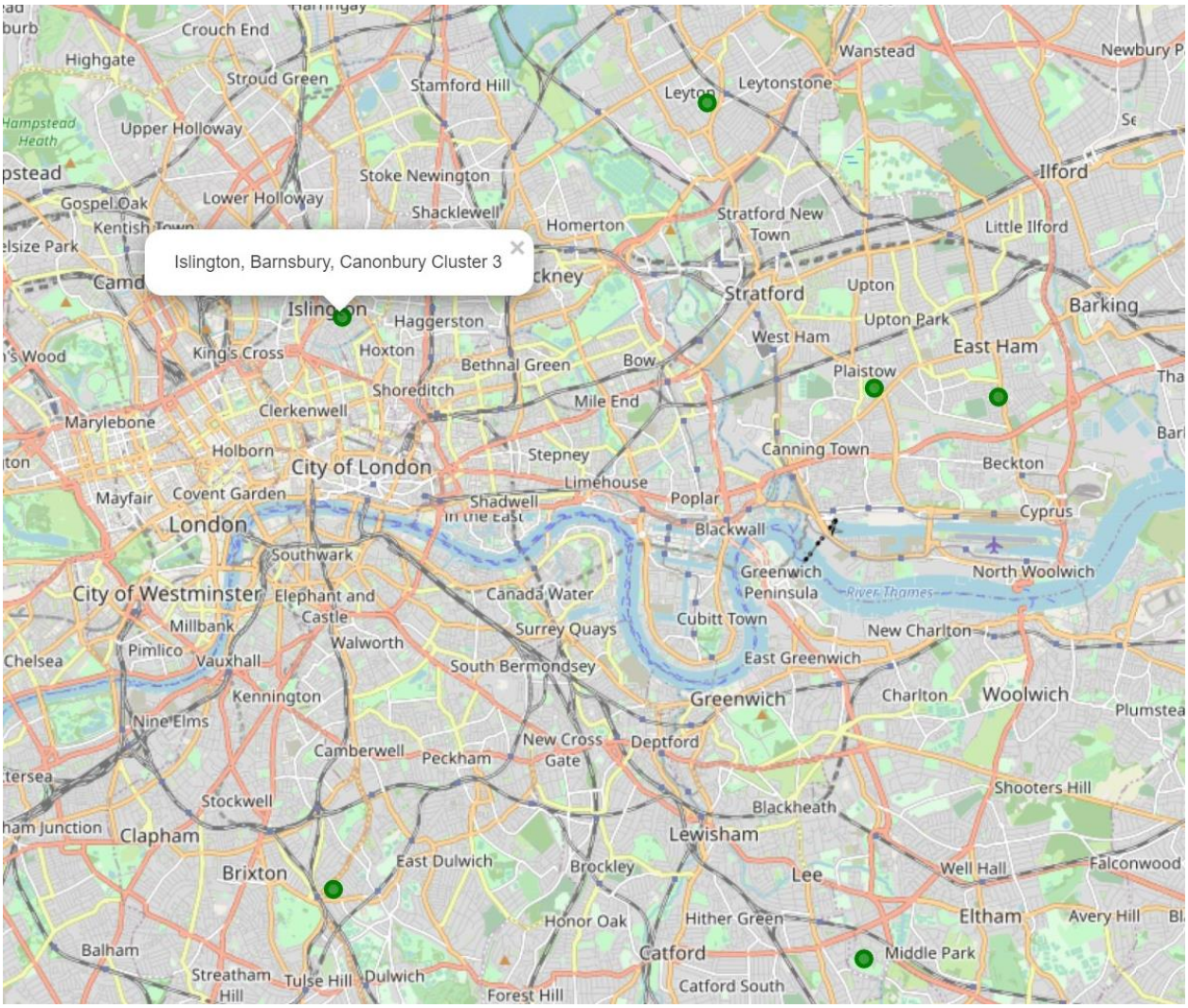
Results

The task is then to filter down the results using the client's strategic criteria.

The most likely competitor in the client's target space is Indian cuisine, which has a significant vegetarian menu offering'. So, any 'Cluster 3' venues with 'Indian Restaurant' as any of the top 10 venues was filtered out.

Further filtering was then applied to omit any 'Boroughs' with any type of 'Restaurant' in the top 10 venues list. This is in line with client's requirement to focus on areas with limited other sit-down dining options. Fast-food is not considered a competitor.

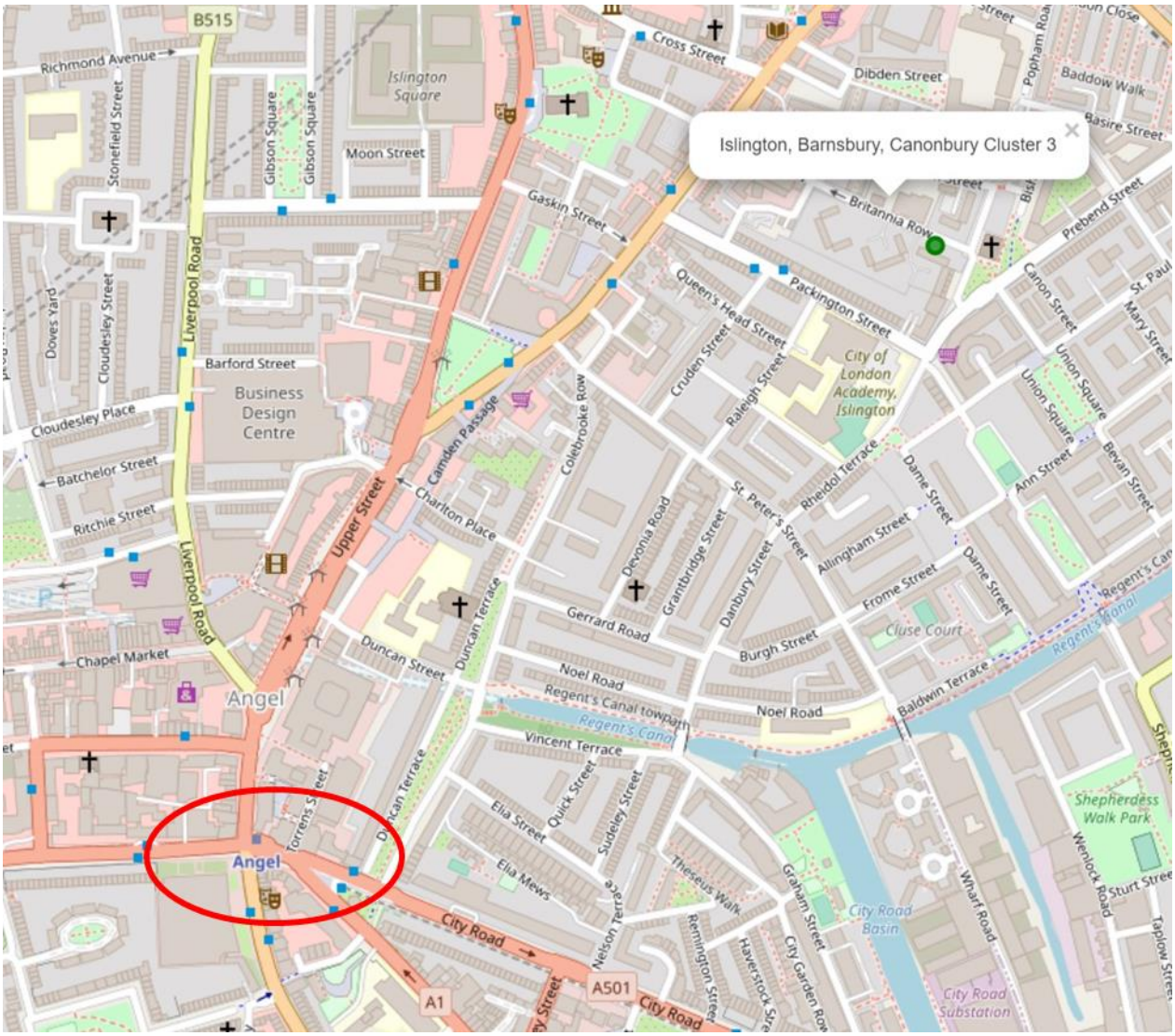
This filtering yielded 6 possible locations:



Discussion

Given the client's final criteria to locate the second restaurant on the same side of London as the first (West Hampstead), if possible, it is clear that there is only one logical choice; '**Islington, Barnsbury, Canonbury'** postcode area: **N1**.

Closer examination of the map confirms that this particular geo-location is very close to 'Angle' tube station.



In this manner, all the client's requirements have been satisfied.

Conclusion

Based on the client's business strategy, insights from machine learning algorithms and leveraging of location data, the best location in London for the client to open his/her second Vegetarian restaurant is **Islington**, close to **Angel tube** station.