Applied Data Science Capstone

A New Bike Store Location in London, England

Alan Smale

August 2019

Business Problem

In congested cities, bicycles are an efficient, environmentally friendly method of transport. There is a growing demand for the infrastructure to support people who choose to take up cycling as a means of transport. Bike shops are required as part of this infrastructure, providing bicycles, maintenance services, spare parts and safety equipment. Investors have recognised this opportunity and desire to establish a new bike shop in London.

This project aims to seek desired locations within London with which to locate a new bike shop. The search will be concentrated in Central London districts.

The business question to be answered is:

"Where would you recommend in Central London to open a new bike shop?"

Data Sets and APIs

The following data sets and APIs will be used:

Wikipedia

Provides information on Central London Postal Districts. The tables in the following URLs will be scraped to extract Central London Postal District information:

https://en.wikipedia.org/wiki/WC_postcode_area
https://en.wikipedia.org/wiki/EC_postcode_area

ArcGIS

Provides geocoding through the GeoPy Geocoders Python library. Latitude and longitude coordinates for Central London and for each of the Central London Districts will be obtained.

Foursquare API

Provides the ability find data about venues of a given type for a specified latitude and longitude. Bike Venues for each of the Central London Districts will be geographically located.

Methodology

Wikipedia was scraped to acquire Central London Districts using the "Postcode district" and "Coverage" columns for both the WC and EC districts.

	District	Neighbourhood
0	WC1A	New Oxford Street
1	WC1B	Bloomsbury, British Museum, Southampton Row
2	WC1E	University College London, SOAS
3	WC1H	St Pancras, UCL Institute of Education
4	WC1N	Russell Square, Great Ormond Street

Table 1. Districts of central London scraped from Wikipedia. (Only the first 5 are shown).

Latitude and longitude information was obtained from ArcGIS and combined.

	District	Neighbourhood	Latitude	Longitude
0	WC1A	New Oxford Street	51.517165	-0.126811
1	WC1B	Bloomsbury, British Museum, Southampton Row	51.519140	-0.127759
2	WC1E	University College London, SOAS	51.522420	-0.133671
3	WC1H	St Pancras, UCL Institute of Education	51.524755	-0.131390
4	WC1N	Russell Square, Great Ormond Street	51.524460	-0.123583

Table 2. Districts with location data added. (Only the first 5 are shown).

The resulting frame was overlaid on a map of central London, the coordinates of which were again obtained from ArcGIS.
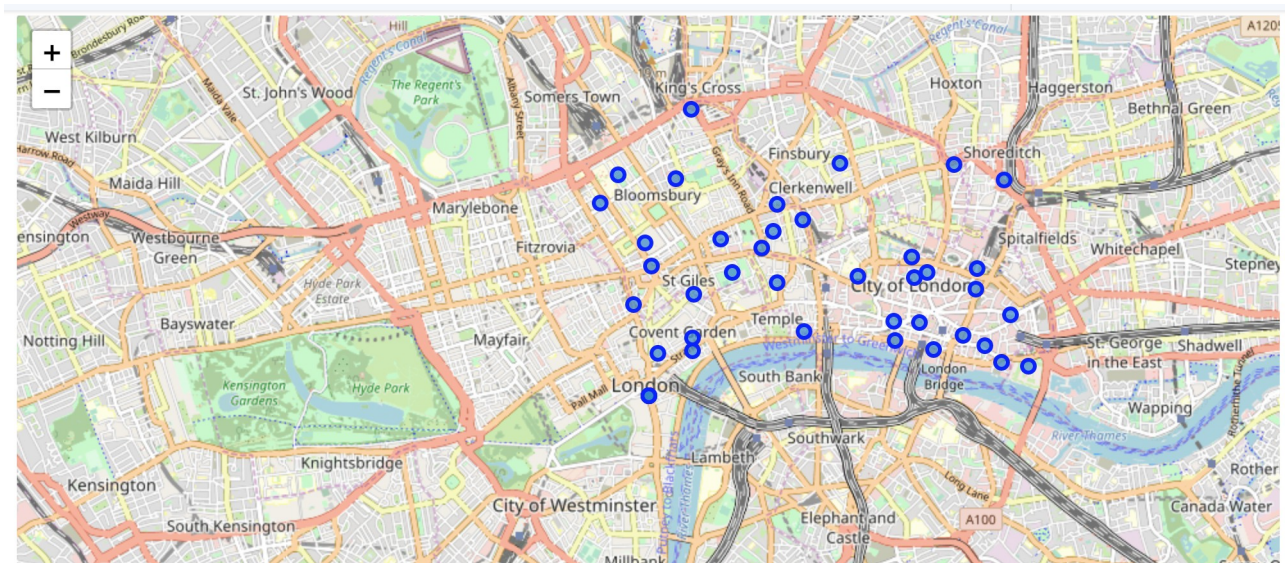

Image 1. Central London districts.

Each district was provided as input to the Foursquare API and a request was made to return venue information for bike shops in the surrounding 500 meters.

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	WC1A	51.517165	-0.126811	Brompton Junction	51.514481	-0.122230	Bike Shop
1	WC1A	51.517165	-0.126811	B1866	51.513732	-0.126321	Bike Shop
2	WC1A	51.517165	-0.126811	Cloud 9 Cycles	51.520209	-0.131023	Bike Shop
3	WC1A	51.517165	-0.126811	Cycle Republic	51.517525	-0.124207	Bike Shop
4	WC1A	51.517165	-0.126811	Vanmoof	51.514126	-0.126456	Bike Shop
5	WC1A	51.517165	-0.126811	Specialized Concept Store	51.512979	-0.125918	Bike Shop
6	WC1B	51.519140	-0.127759	Cloud 9 Cycles	51.520209	-0.131023	Bike Shop
7	WC1B	51.519140	-0.127759	Cycle Republic	51.517525	-0.124207	Bike Shop
8	WC1E	51.522420	-0.133671	Cloud 9 Cycles	51.520209	-0.131023	Bike Shop

Table 3. Bike shop locations added to the data. (Only the first 9 are shown).

To ensure that bike shop location data had been retrieved, counts were made of the number of shops returned for each district.

District	Count		District	Count
EC1A	5		EC4M	1
EC1M	3		EC4N	1
EC1N	4		EC4P	1
EC1P	1		EC4R	3
EC1R	4		EC4V	1
EC1V	2		EC4Y	1
EC1Y	3		EC50	1
EC2A	4		WC1A	6
EC2M	4		WC1B	2
EC2N	1		WC1E	3
EC2P	1		WC1N	2
EC2R	2		WC1R	4
EC2V	3		WC1V	4
EC2Y	3		WC1X	2
EC3A	2		WC2A	4
EC3M	1		WC2B	7
EC3P	1		WC2E	5
EC3R	1		WC2H	4
EC3V	1		WC2N	4
EC4A	1		WC2R	3

Table 4. Bike shop counts in Central London districts.

One hot encoding was performed to prepare the data for clustering and to ensure that no other venue types has been retrieved. It was discovered that the Foursquare API had returned a cafe, a juice bar and motorcycle shop. Since we are not interested in venues that aren't bike shops we removed this data from the data set. The data was clustered by way of the k-means clustering technique with a target of three clusters. The clustered data was merged back in to the original data to provide a view of clustered districts.

Results

The clusters were assigned unique colours and plotted onto a map of central London.
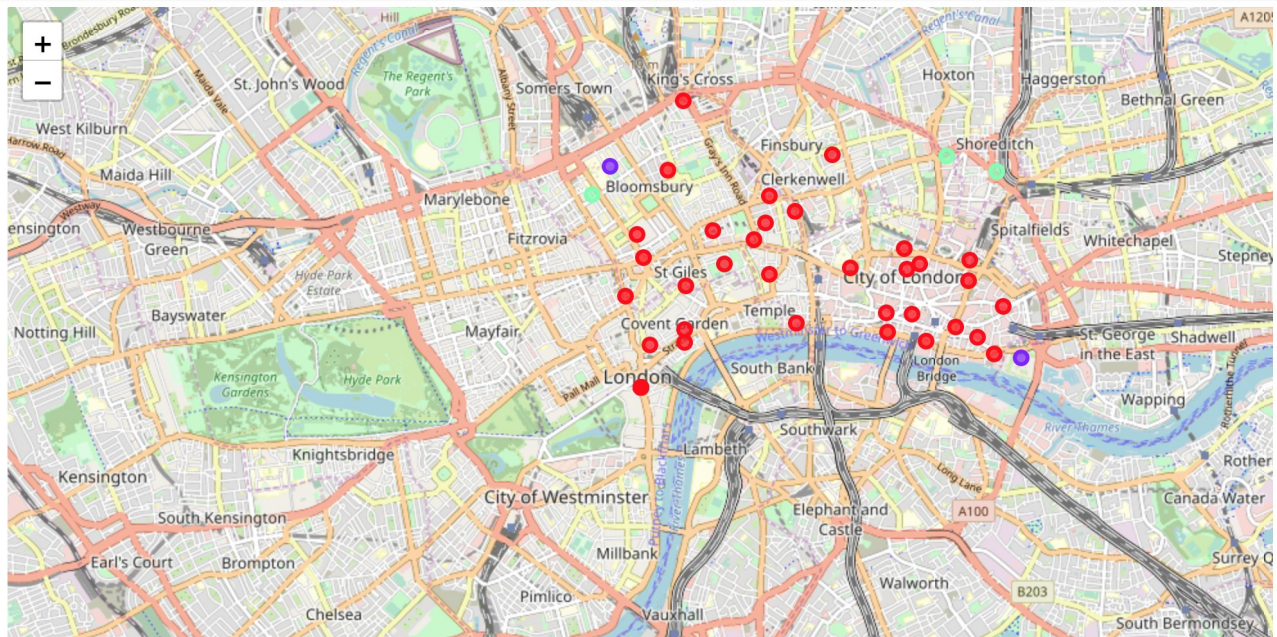

Image 2. Clusters of bike shops in Central London districts.

The clusters have the following attributes:

Cluster 0 – Red, a large number of bike shops near

Cluster 1 – Purple, no bike shops near

Cluster 2 – Green, few bike shops near

Discussion

Based on the clustered data, it is preferable to locate the new bike shop in a district where there are no bike shops nearby; this corresponds with locations in cluster 1:

	District	Neighbourhood	Latitude	Longitude	Cluster Labels	Bike Shop
3	WC1H	St Pancras, UCL Institute of Education	51.524755	-0.131390	1	0.0
30	EC3N	Tower Hill, Tower of London	51.508825	-0.076407	1	0.0

Table 5. Central London districts with no bike shops nearby.

Conclusion

During this project we have been able to successfully locate two places of interest where a new bike shop can be located that does not have other established bike shops nearby.

Links

Jupyter Notebook:

https://github.com/alan-j-smale/ibm-ds-capstone/blob/master/Battle%20of%20the%20Neighbourhoods.ipynb