Relocating to London
Finding a good school, easy commute and affordable rent

IBM Data Science Capstone Project

Contents

Introduction
In this project we will look at the problem of relocating to London with a family: finding a good school for the kids, and a place to live that is within budget and has a good commute into Central London. We will consider proximity to top state schools, commuting distance to Central London by tube and average rent near tube stations. For our final shortlist, we will perform a clustering analysis for London boroughs based on venues in order to group similar ones together and get a better feel for what they are like.

Data
Raw data has been obtained from the following websites:
- Top (free) state schools: https://www.homesandproperty.co.uk/property-news/where-to-buy-a-new-home-near-a-good-london-state-school-a126836.html
- Tube stations: commute time into Central London and average weekly rent: https://www.totallymoney.com/rent-vs-tube-journey-time/
- Tube stations: locations: https://wiki.openstreetmap.org/wiki/List_of_London_Underground_stations
- List of London boroughs with locations: https://en.wikipedia.org/wiki/List_of_London_boroughs

For our clustering analysis of London boroughs, we will use the Foursquare API to obtain data on venues.
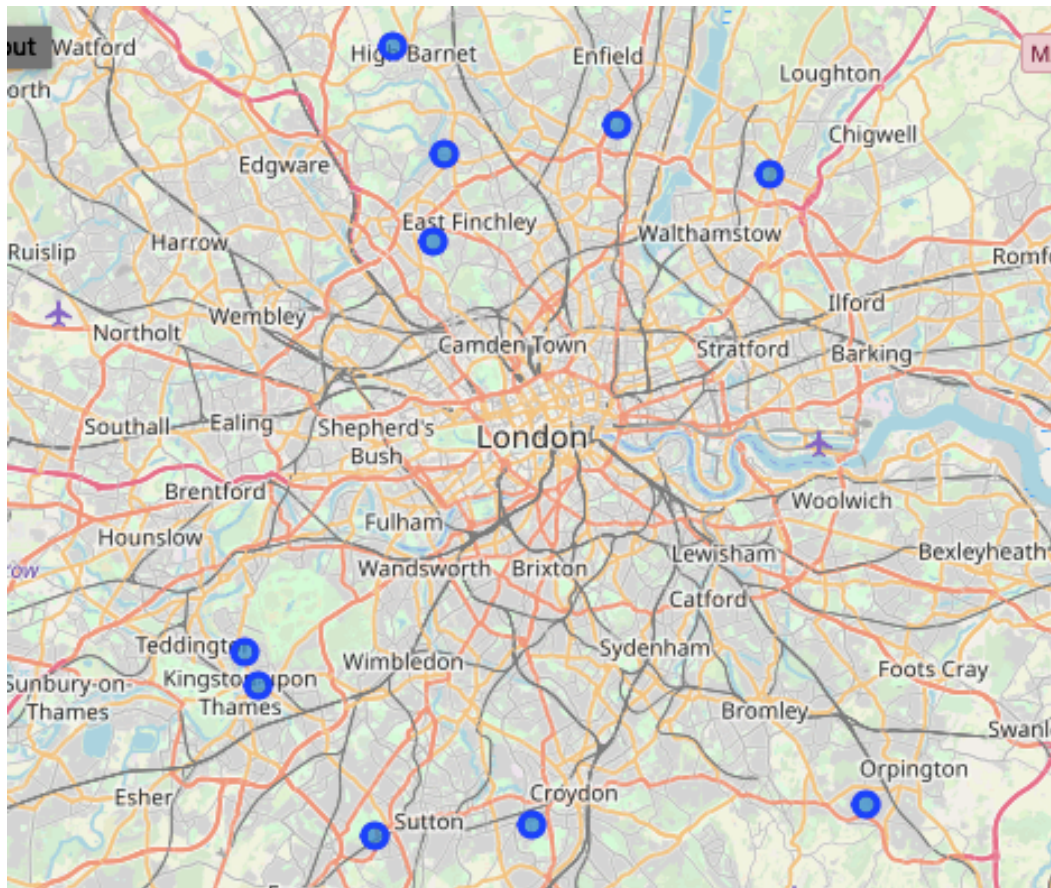
Data has been scraped and stored in excel files for convenience.

Methodology

Top London state schools: schools are ranked based on the percentage of pupils that achieves grade 5 or above in English and maths:

School	Borough	Post code	% of pupils with grade 5 in English and maths
The Henrietta Barnett School	Barnet	NW11 7BN	100
Queen Elizabeth's School	Barnet	EN5 4DQ	100
Wilson's School	Sutton	SM6 9JW	100
St Michael's Catholic Grammar School	Barnet	N12 7NJ	99
Newstead Wood School	Bromley	BR6 9SA	99
The Latymer School	Enfield	N9 9TN	99
The Tiffin Girls' School	Kingston upon Thames	KT2 5PL	99
Tiffin School	Kingston upon Thames	KT2 6RL	99
Woodford County High School	Redbridge	IG8 9LA	99
Nonsuch High School for Girls	Sutton	SM3 8AB	99

We use the Geopy library to obtain latitude and longitude coordinates and to display the schools on a map:

London tube stations

We load the data on London tube stations with commute time to central London and average weekly rent and narrow down the options based on commute time and rent budget. We then add location data and find the nearest borough for each station.
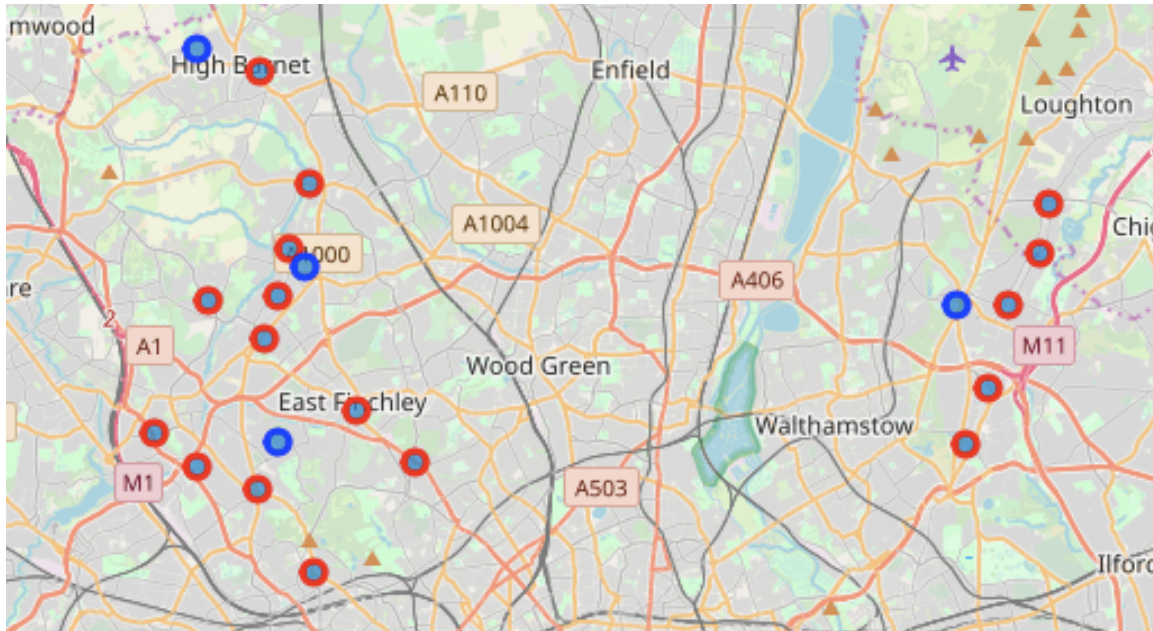
We ignore all the slower connections to Central London from a tube station and only consider tube stations with a commute time into Central London of 45 minutes or less and an average weekly rent of £500 or less.

The following table shows the nearest tube connections for our schools:

School	Nearest tube station	Distance (mi)
The Henrietta Barnett School	Golders Green	0.68
Queen Elizabeth's School	High Barnet	0.87
Wilson's School	Morden	4.16
St Michael's Catholic Grammar School	Woodside Park	0.31
Newstead Wood School	North Greenwich	9.77
The Latymer School	Southgate	2.29
The Tiffin Girls' School	Richmond	2.63
Tiffin School	Richmond	3.57
Woodford County High School	Woodford	0.67
Nonsuch High School for Girls	Morden	3.52

The distance to the nearest tube station ranges from 0.31 to 9.77 miles. We decide that we'll allow for a maximum distance of three miles, which narrows down the number of schools to consider to four: The Henrietta Barnett School, Queen Elizabeth's School, St Michael's Catholic Grammar School, and Woodford Country High School.

Next, we find all the tube stations for those schools that are within 2 miles of a tube station. The schools are shown in blue and tube connections in red in the figure below:

We now have a pretty good idea of our options. The four schools that we selected are all located in the North London boroughs of Barnet and Redbridge. However, the tube stations that are within range add Brent, Camden, Haringey and Waltham Forest to our list of boroughs to consider.

In order to make a final decision, we would like to know what London boroughs are like. We perform a clustering analysis based on the top venues in each borough and use the Foursquare API.

Exploring London boroughs (clustering analysis)

We use the location data on London boroughs to connect to the Foursquare API and find 100 venues within a 500 meter range. We then consider the top ten most common venue categories per borough and group the boroughs into five clusters to get an idea of what they are like.

The following table shows a snapshot of clustering results for ten London boroughs:

	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Barking and Dagenham	51.56	0.16	1	Gym / Fitness Center	Park	Pool	Supermarket	Golf Course
1	Barnet	51.63	-0.15	2	Café	Bus Stop	Yoga Studio	Fast Food Restaurant	Furniture / Home Store
2	Bexley	51.45	0.15	1	Pub	Coffee Shop	Italian Restaurant	Clothing Store	Fast Food Restaurant
3	Brent	51.56	-0.28	1	Coffee Shop	Hotel	Clothing Store	Bar	Sporting Goods Shop
4	Bromley	51.40	0.02	1	Coffee Shop	Clothing Store	Pizza Place	Burger Joint	Gym / Fitness Center
5	Camden	51.53	-0.13	1	Coffee Shop	Hotel	Pub	Café	Burger Joint
6	Croydon	51.37	-0.10	1	Pub	Coffee Shop	Asian Restaurant	Gym / Fitness Center	Gaming Cafe
7	Ealing	51.51	-0.31	1	Coffee Shop	Clothing Store	Italian Restaurant	Park	Bakery
8	Enfield	51.65	-0.08	1	Coffee Shop	Clothing Store	Pharmacy	Mobile Phone Shop	Shopping Mall
9	Greenwich	51.49	0.06	1	Pub	Supermarket	Coffee Shop	Fast Food Restaurant	Hotel

The clustering labels show boroughs that are grouped together. Perhaps unsurprisingly, most boroughs in Greater London have very similar amenities and most boroughs are grouped together in the same cluster (1). Barnet, Harrow, Hounslow, and Newham are a bit more distinct.

Results
On the basis of data collected from several websites, including geo location data, we have been able to create a shortlist of schools that have a good tube connection to Central London and are in areas that fall within our rent budget.

The four schools that we selected are all in North London: The Henrietta Barnett School, Queen Elizabeth's School, St Michael's Catholic Grammar School, and Woodford Country High School. All schools are located in Barnet or Redbridge. However, the tube stations that are within range add Brent, Camden, Haringey and Waltham Forest to our list of boroughs to consider.
In order to make a final decision on where we would like to live, we retrieved venues for each London boroughs using the Foursquare API and performed a clustering analysis based on venue categories. We found that most boroughs in London are hard to distinguish as they offer very similar amenities.

Discussion
Looking at only the top ten state schools, only four schools in London meet our criteria and the remaining six schools in South London are excluded because of poor connectivity to Central London. We could consider a longer list of schools to increase the probability of admission and have not considered school specific admission criteria that are likely to differ somewhat from school to school.

Amenities in London are generally very good and quite similar for most boroughs so we could consider other data sources and explore new features such as house prices, average income levels, average school ratings and private schools.

Conclusion

Based on website and geo location data, we have been able to find the 'sweet spot' of free state schools in areas that fall within our rent budget and provide a good connection to Central London. This provides a useful starting point for possible extensions that could include:

- a wider set of schools,
- selection of multiple schools,
- admission criteria,
- new features that may provide a more distinct grouping of boroughs.