

IBM Applied Data Science Capstone Project

Battle of the Neighborhoods

Introduction

As a Data Scientist I have been approached by a client who is planning on setting up a “walk-in” Creche business in London. Her requirements are to maximize the number of children attending the Creche and the duration that the children are left at the Creche. Depending on the success of her business she would be interested in extending the relationship and looking at setting up Creches in other areas in London with similar conditions.

Data

To address the business problem, we need to understand what data is necessary to determine the best location for the creche. We can do this by gathering information for different areas in London and checking the amount of businesses that reside in or around the area.

Technically we can use wiki to obtain a list of postal districts of London. We can then use the list of postcodes within each district to obtain the centroid of each one, again from Wiki. Some of these Postcodes are classed as non-geographic and these should be omitted. From the centroids we can then use FourSquare to identify the various businesses within a set radius.

The expectation being that people will make use of the Creche for an hour or more, allowing them to; grab something to eat; have a coffee with friends; do some shopping, while their children are being looked after for them.

Methodology

Firstly I identified the various postcodes for the regions of Greater London using wiki https://en.wikipedia.org/wiki/{}_postcode_area passing in the 8 values ['E','EC','N','NW','SE','SW','W','WC'] to pick up the postcodes for that region.

I then used google geocoder to get the latitudes and longitudes for these postcodes.

	Postal Region	Postcode	Post town	Coverage	Local authority area	Latitude	Longitude
0	E	E1	LONDON	Eastern head district: Aldgate (part), Bishops...	Tower Hamlets, Hackney, City of London	51.517623	-0.055755
1	E	E1W	LONDON	Wapping, St Katharine Docks, Stepney (part), S...	Tower Hamlets	51.505796	-0.057199
2	E	E2	LONDON	Bethnal Green district: Bethnal Green (mostly)...	Tower Hamlets, Hackney	51.528134	-0.061651
3	E	E3	LONDON	Bow district: Bow, Bow Common, Bromley-by-Bow,...	Tower Hamlets, Newham	51.528148	-0.020542
4	E	E4	LONDON	Chingford district: Chingford, Sewardstone, Hi...	Waltham Forest, Enfield, Epping Forest (Essex)	51.630750	-0.011780

I then loop through these postcodes and request venues using the Foursquare API that are nearby.

	Neighborhood Postcode	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Postcode	Venue Postal Outcode	Venue Postal Region
0	E1	51.517623	-0.055755	Mouse Tail Coffee Stories	51.519471	-0.058573	Coffee Shop	E1 5DG	E1	E
1	E1	51.517623	-0.055755	Rinkoff's Bakery	51.519964	-0.053238	Bakery	E1 3AR	E1	E
2	E1	51.517623	-0.055755	Genesis Cinema	51.521036	-0.051073	Movie Theater	E1 4UH	E1	E
3	E1	51.517623	-0.055755	The White Hart	51.520245	-0.056385	Pub	E1 5QH	E1	E
4	E1	51.517623	-0.055755	Dirty Burger/Chicken Shop	51.520253	-0.055112	Burger Joint	E1 4TL	E1	E

The next step is to split out the venue and neighbourhood data

	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Postcode	Venue Postal Outcode	Venue Postal Region
0	Mouse Tail Coffee Stories	51.519471	-0.058573	Coffee Shop	E1 5DG	E1	E
1	Rinkoff's Bakery	51.519964	-0.053238	Bakery	E1 3AR	E1	E
2	Genesis Cinema	51.521036	-0.051073	Movie Theater	E1 4UH	E1	E
3	The White Hart	51.520245	-0.056385	Pub	E1 5QH	E1	E
4	Dirty Burger/Chicken Shop	51.520253	-0.055112	Burger Joint	E1 4TL	E1	E

	Postal Region	Postcode	Post town	Coverage	Local authority area	Latitude	Longitude
0	E	E1	LONDON	Eastern head district: Aldgate (part), Bishops...	Tower Hamlets, Hackney, City of London	51.517623	-0.055755
1	E	E1W	LONDON	Wapping, St Katharine Docks, Stepney (part), S...	Tower Hamlets	51.505796	-0.057199
2	E	E2	LONDON	Bethnal Green district: Bethnal Green (mostly)...	Tower Hamlets, Hackney	51.528134	-0.061651
3	E	E3	LONDON	Bow district: Bow, Bow Common, Bromley-by-Bow,...	Tower Hamlets, Newham	51.528148	-0.020542
4	E	E4	LONDON	Chingford district: Chingford, Sewardstone, Hi...	Waltham Forest, Enfield, Epping Forest (Essex)	51.630750	-0.011780

I then create a matrix of venues x postcodes and calculate the geodesic distance between the neighbourhood coordinates and the venue coordinates

	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Postcode	Venue Postal Outcode	Venue Postal Region	E1	E1W	E2	E3	E4	E5
0	Mouse Tail Coffee Stories	51.519471	-0.058573	Coffee Shop	E1 5DG	E1	E	283.790058	1524.497955	987.204411	2810.526519	12798.708543	4514.640314
1	Rinkoff's Bakery	51.519964	-0.053238	Bakery	E1 3AR	E1	E	313.616195	1600.106846	1080.384607	2445.164954	12656.641756	4425.967678
2	Genesis Cinema	51.521036	-0.051073	Movie Theater	E1 4UH	E1	E	499.781710	1748.069321	1078.279641	2261.942843	12506.984483	4301.559963
3	The White Hart	51.520245	-0.056385	Pub	E1 5QH	E1	E	294.984347	1608.582189	950.787728	2638.496861	12677.610803	4411.420058
4	Dirty Burger/Chicken Shop	51.520253	-0.055112	Burger Joint	E1 4TL	E1	E	296.043420	1615.028004	987.306424	2555.037167	12655.472668	4402.446949

I then applied k-means clustering to see if there would be any benefit of filtering the data based on cluster. I attempted a variety of cluster counts but none really helped with a lot of occurrences of pubs, hotels and restaurants.

Cluster 0

	Venue Postal Outcode	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
121	SW1H	0	Hotel	Sandwich Place	Sporting Goods Shop	Restaurant	Café	Hotel Bar	Zoo Exhibit	Fish Market	Factory	Falafel Restaurant

Cluster 1

	Venue Postal Outcode	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
45	IG8	1	Supermarket	Zoo Exhibit	Flower Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Filipino Restaurant	Film Studio	Fish & Chips Shop

Cluster 2

	Venue Postal Outcode	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	E10	2	Grocery Store	Park	Café	Coffee Shop	Pub	Hotel	Restaurant	Soccer Stadium	Train Station	Fast Food Restaurant
5	E13	2	Grocery Store	Metro Station	Pub	Bus Stop	Hostel	Gym / Fitness Center	Park	Bakery	Pharmacy	Fish & Chips Shop
14	E3	2	Pub	Café	Grocery Store	Bus Stop	Art Gallery	Coffee Shop	Park	Light Rail Station	Hotel	Historic Site
15	E4	2	Pub	Park	Italian Restaurant	Grocery Store	Fast Food Restaurant	Sandwich Place	Fish Market	Falafel Restaurant	Farm	Farmers Market
60	N20	2	Park	Cricket Ground	Pub	Brazilian Restaurant	Soccer Field	Golf Course	Zoo Exhibit	Fish Market	Farm	Farmers Market

Cluster 3

	Venue Postal Outcode	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
6	E14	3	Coffee Shop	Italian Restaurant	Park	Burger Joint	Hotel	Gym / Fitness Center	Plaza	Trail	Sandwich Place	Café
13	E20	3	Café	Italian Restaurant	Toy / Game Store	Juice Bar	Grocery Store	Park	Pizza Place	Restaurant	Clothing Store	Sandwich Place
22	EC1M	3	Italian Restaurant	Pub	Hotel	Sandwich Place	French Restaurant	Nightclub	Museum	Steakhouse	Café	Fish & Chips Shop
28	EC2M	3	Hotel	Coffee Shop	Boxing Gym	Breakfast Spot	Historic Site	Hawaiian Restaurant	Salad Place	Colombian Restaurant	Gym / Fitness Center	Organic Grocery
41	EC4R	3	Trail	Wine Bar	Gym	Beer Bar	Seafood Restaurant	Asian Restaurant	Nightclub	Dessert Shop	Flower Shop	Farmers Market

Cluster 7

	Venue Postal Outcode	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
1	E1	7	Pub	Coffee Shop	Hotel	Indian Restaurant	Pizza Place	Café	Grocery Store	Turkish Restaurant	Park	Thai Restaurant
3	E11	7	Pub	Café	Coffee Shop	Pharmacy	Bakery	Restaurant	Fish & Chips Shop	Fast Food Restaurant	Metro Station	Supermarket
7	E15	7	Pub	Café	Hotel	Grocery Store	General Entertainment	Bus Stop	Bar	Park	Coffee Shop	Fast Food Restaurant
9	E17	7	Pub	Café	Coffee Shop	Restaurant	Grocery Store	Pizza Place	Italian Restaurant	Fast Food Restaurant	Brewery	Bar
11	E1W	7	Pub	Coffee Shop	Park	Restaurant	Italian Restaurant	Breakfast Spot	Street Food Gathering	Supermarket	Tapas Restaurant	Event Space
12	E2	7	Pub	Coffee Shop	Café	Wine Bar	Italian Restaurant	Park	Restaurant	Art Gallery	Cocktail Bar	Pizza Place

I then decided to look for Postcodes with a high number of locations within 250 metres and within 1000 metres

	Postcode	Very Close	Fairly Close	Total
0	W1F	48	365	413
0	WC2E	52	271	323

I then checked for other locations with a high number of total venues within 1000 metres (> 400)

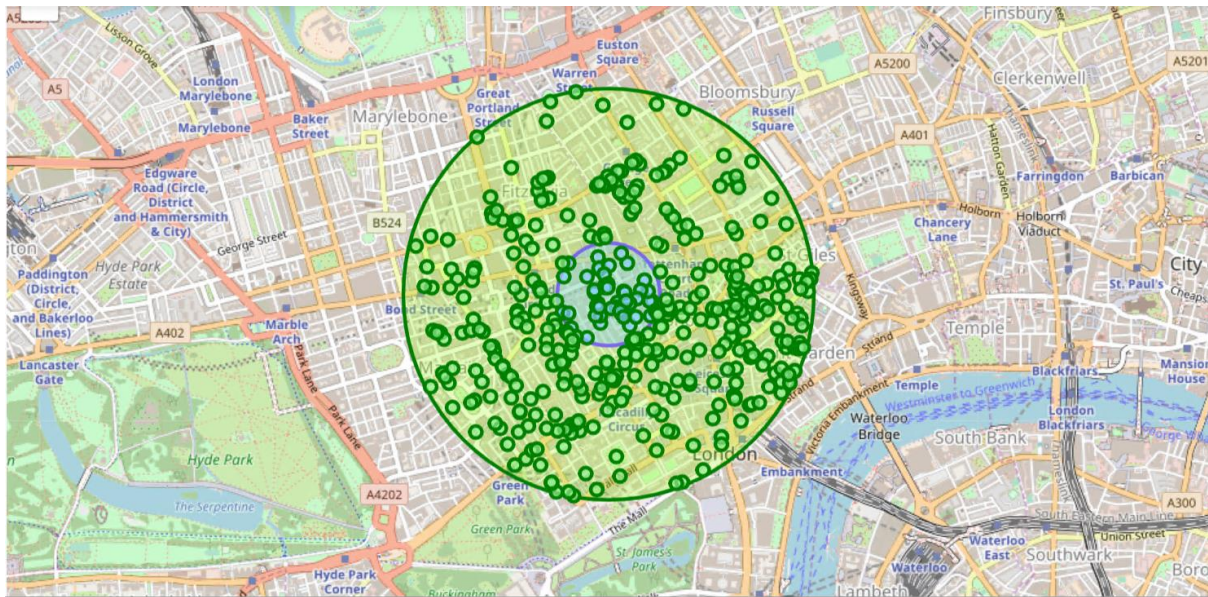
	Postcode	Very Close	Fairly Close	Total
0	W1D	46	370	416
0	W1F	48	365	413

To validate the process I also looked for postcodes with very few venues within 1000 metres.

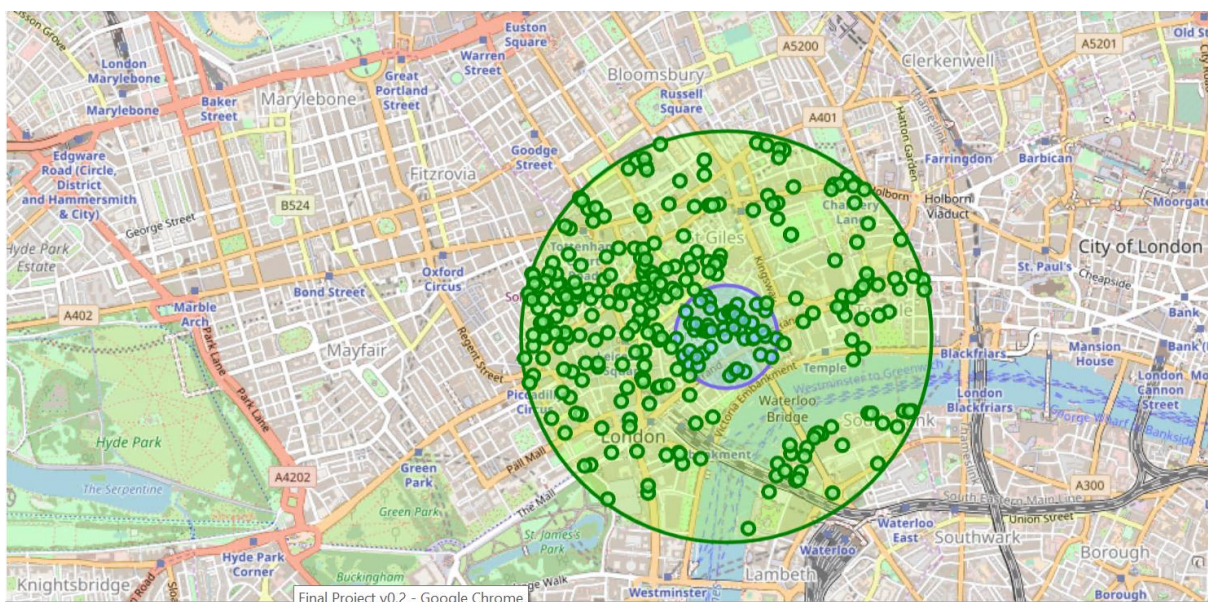
	Postcode	Very Close	Fairly Close	Total
0	E4	0	9	9
0	E6	1	14	15
0	E12	1	11	12
0	N20	0	5	5
0	N21	1	17	18
0	NW7	0	7	7
0	SE2	1	7	8
0	SE9	0	6	6
0	SE12	0	18	18
0	SE18	0	15	15
0	SE28	0	11	11

Results

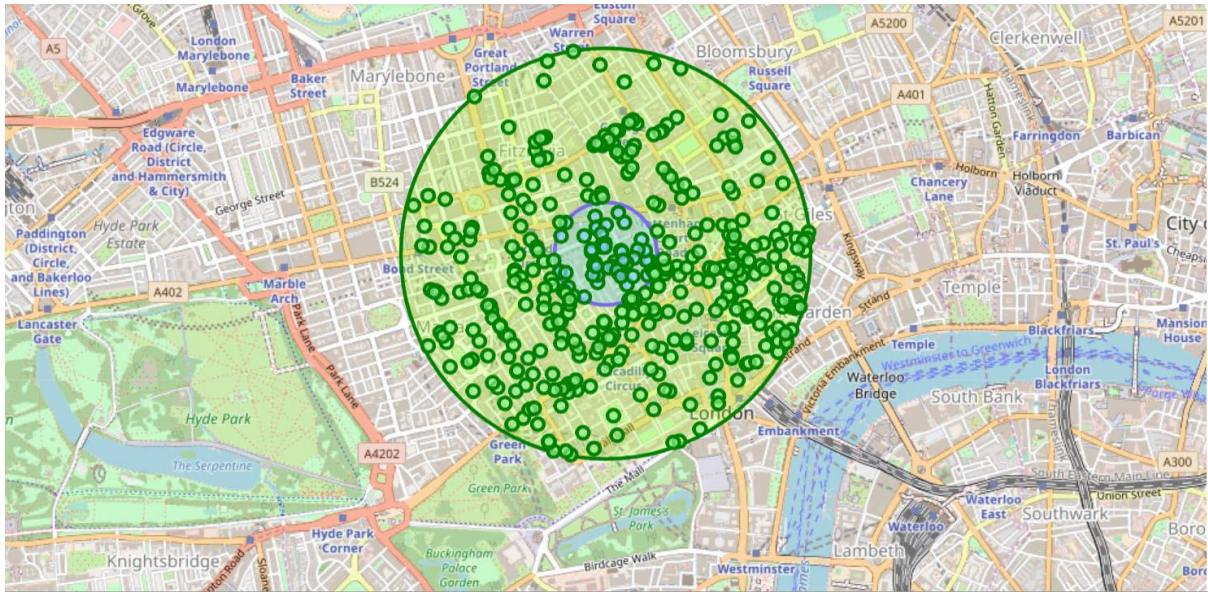
The map below shows the location W1F with nearby venues



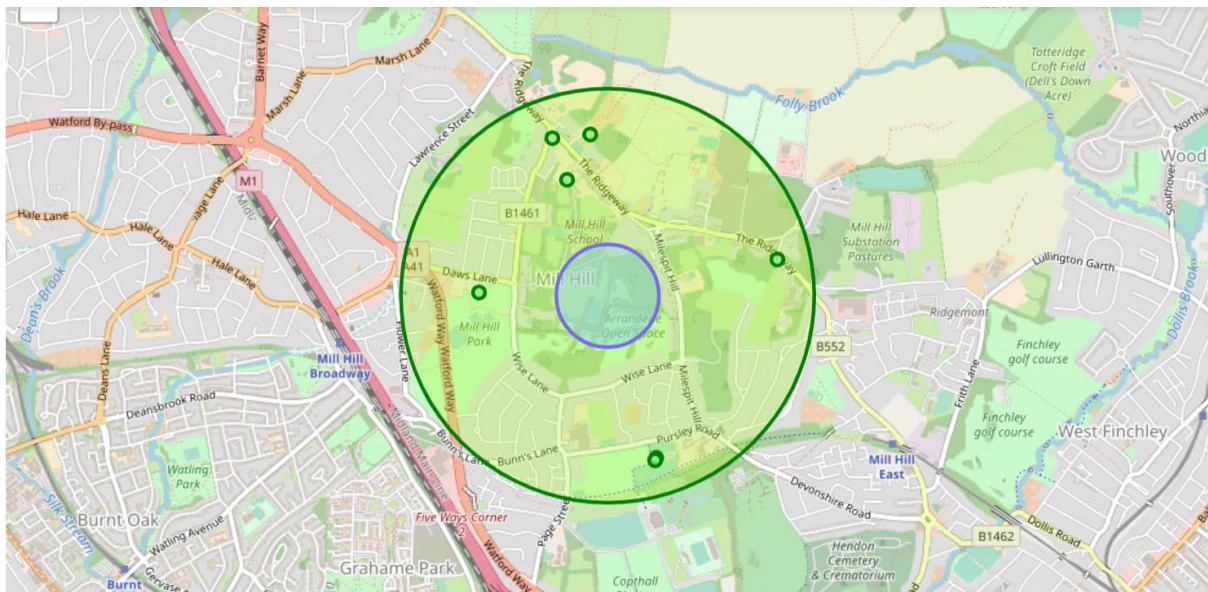
The map below is WC2E with venues nearby



Although not appearing in the top two postcodes with very close venues W1D does appear top for venues within 1000 metres of the postcode centroid.



NW7 is an example of a poor choice, having no very close venues and a small number of venues within 1000 metres.



Conclusion

Unfortunately, there is no clear winner however given that WC2E spans the Thames we could rule this one out as crossing the river could be inconvenient. W1F could be a good choice containing a bar, a coffee shop and 4 restaurants in the top 10.

	Venue Postal Outcode	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
145	W1D	Theater	Hotel	Ice Cream Shop	Liquor Store	Lounge	Park	Plaza	Tapas Restaurant	Steakhouse	Sri Lankan Restaurant
146	W1F	Cocktail Bar	Coffee Shop	Italian Restaurant	Art Gallery	Burger Joint	BBQ Joint	Chocolate Shop	Cupcake Shop	Tapas Restaurant	Lingerie Store
173	WC2E	Dessert Shop	Italian Restaurant	Burger Joint	French Restaurant	Clothing Store	Pub	Theater	Cupcake Shop	Sushi Restaurant	Boutique