Searching for suitable location in Central London for a new Music Venue

Course: Coursera Data Science Capstone

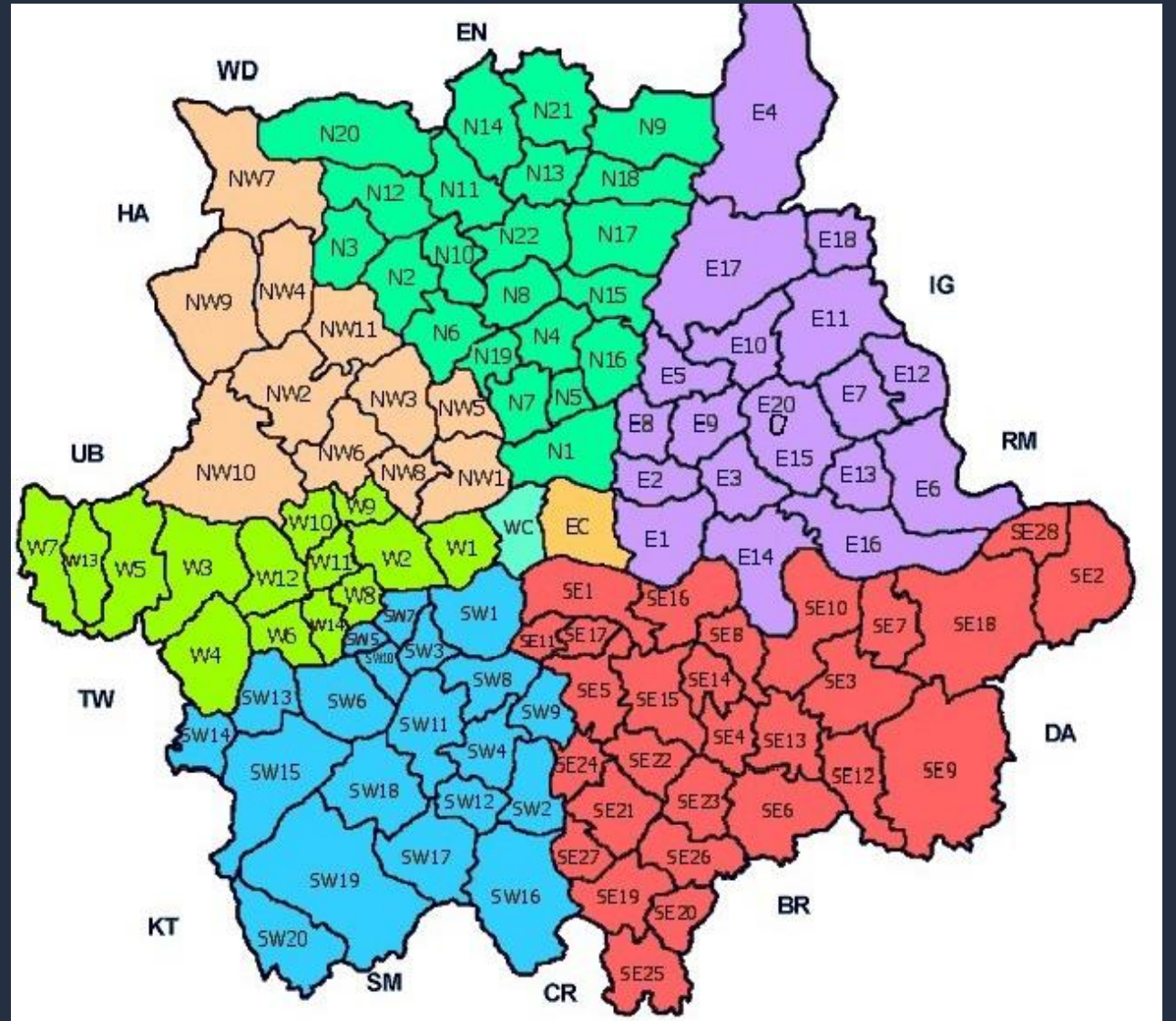Author: Jakob Hegley

Objective

- Find a suitable area of London for our employer to open up a new music venue.

- Want to find an area lacking in music representation and one that already has a decent economic standing or at least room for economic growth.
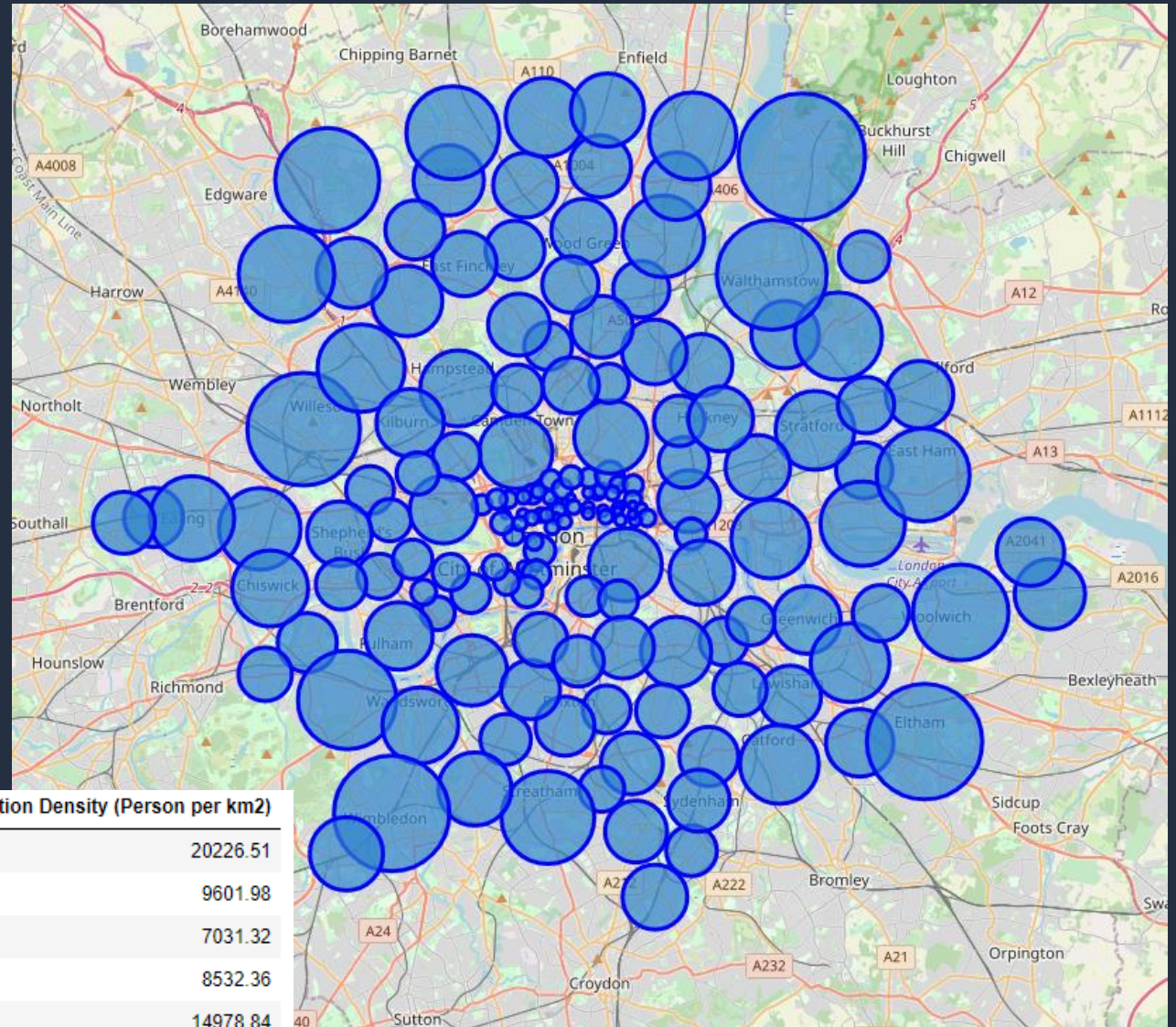
Data Acquisition

- Used London Postal districts to define "areas of interest." Scraped the Wikipedia page: https://en.wikipedia.org/wiki/London_postal_district

- Latitude and Longitude data found from: https://www.freemaptools.com/download-uk-postcode-lat-lng.htm

- Area size and population data found from: https://www.streetlist.co.uk/ and https://www.streetcheck.co.uk/postcode/alldistricts

- Utilized FourSquare API to pull venue data for each postcode district.

Location Data

- Scraped Wikipedia pages for district designations and compiled location data.

- Combined with district area and population data. Also calculated population densities for each area.

- Approximated area of each district using a circular radius.

	Postcode	Borough(s)	Latitude	Longitude	Area (km2)	Population	Population Density (Person per km2)
0	E1	Tower Hamlets, Hackney, City of London	51.51766	-0.05841	4.30	86974	20226.51
1	E10	Waltham Forest, Hackney	51.56814	-0.01153	5.05	48490	9601.98
2	E11	Waltham Forest, Redbridge	51.56769	0.01443	8.78	61735	7031.32
3	E12	Newham, Redbridge	51.54992	0.05404	5.13	43771	8532.36
4	E13	Newham	51.52700	0.02705	3.45	51677	14978.84

Venue Data

- Utilized FourSquare API to pull the top 100 (or less) venues around the latitude and longitude coordinates within the radius defined previously.
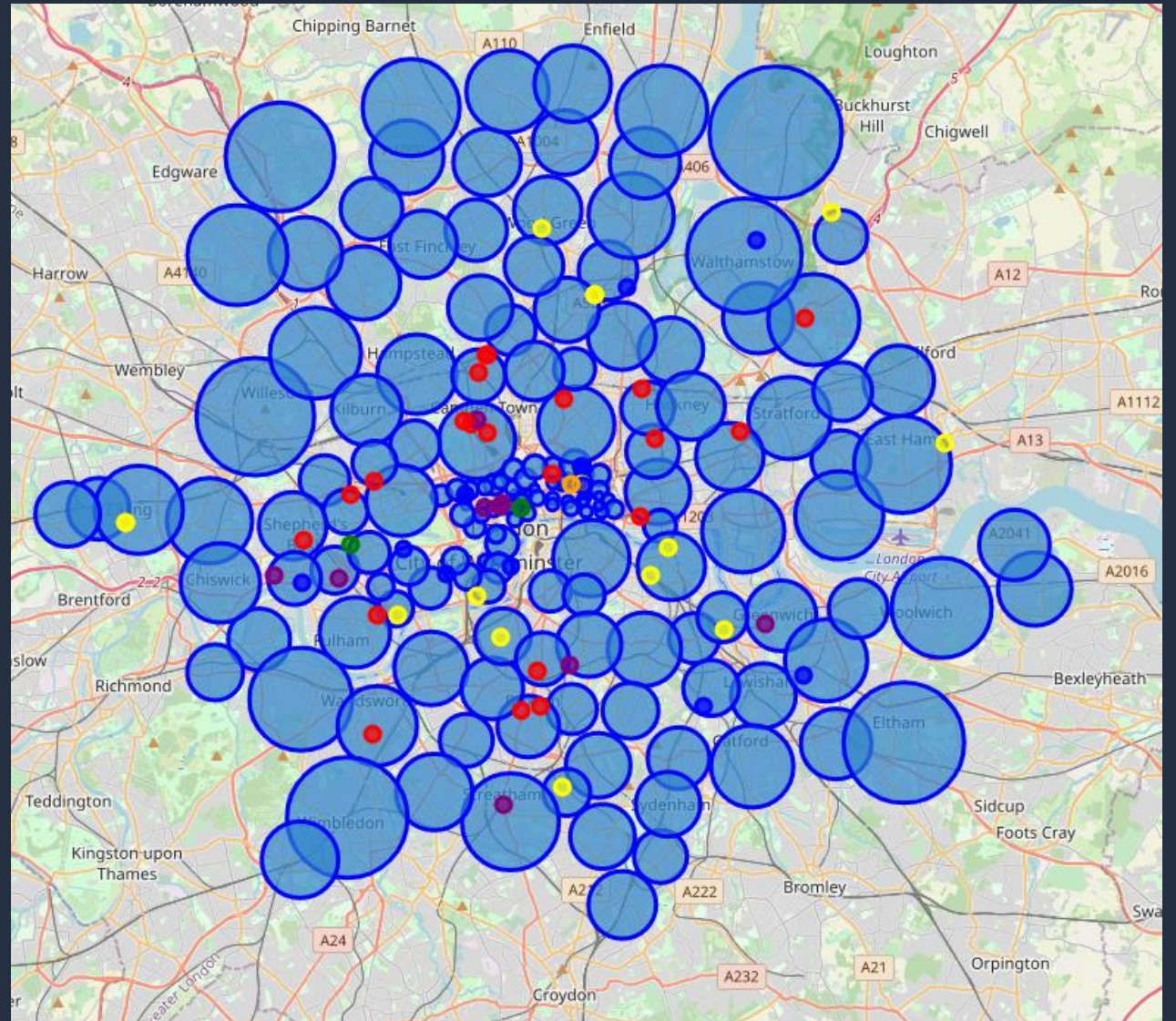
	Postcode	Postcode Latitude	Postcode Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	E1	51.51766	-0.05841	Mouse Tail Coffee Stories	51.519471	-0.058573	Coffee Shop
1	E1	51.51766	-0.05841	Rinkoff's Bakery	51.519964	-0.053238	Bakery
2	E1	51.51766	-0.05841	Tayyabs	51.517240	-0.063476	North Indian Restaurant
3	E1	51.51766	-0.05841	Lahore One	51.514725	-0.059399	Indian Restaurant
4	E1	51.51766	-0.05841	One Mile End	51.520151	-0.056136	Brewery

Existing Music Venues

- Searched for music venue style venue categories and plotted these venues on our map.

Red = Music Venue

Blue = Concert Hall

Green = Opera House

Purple = Jazz Club

Orange = Piano Bar
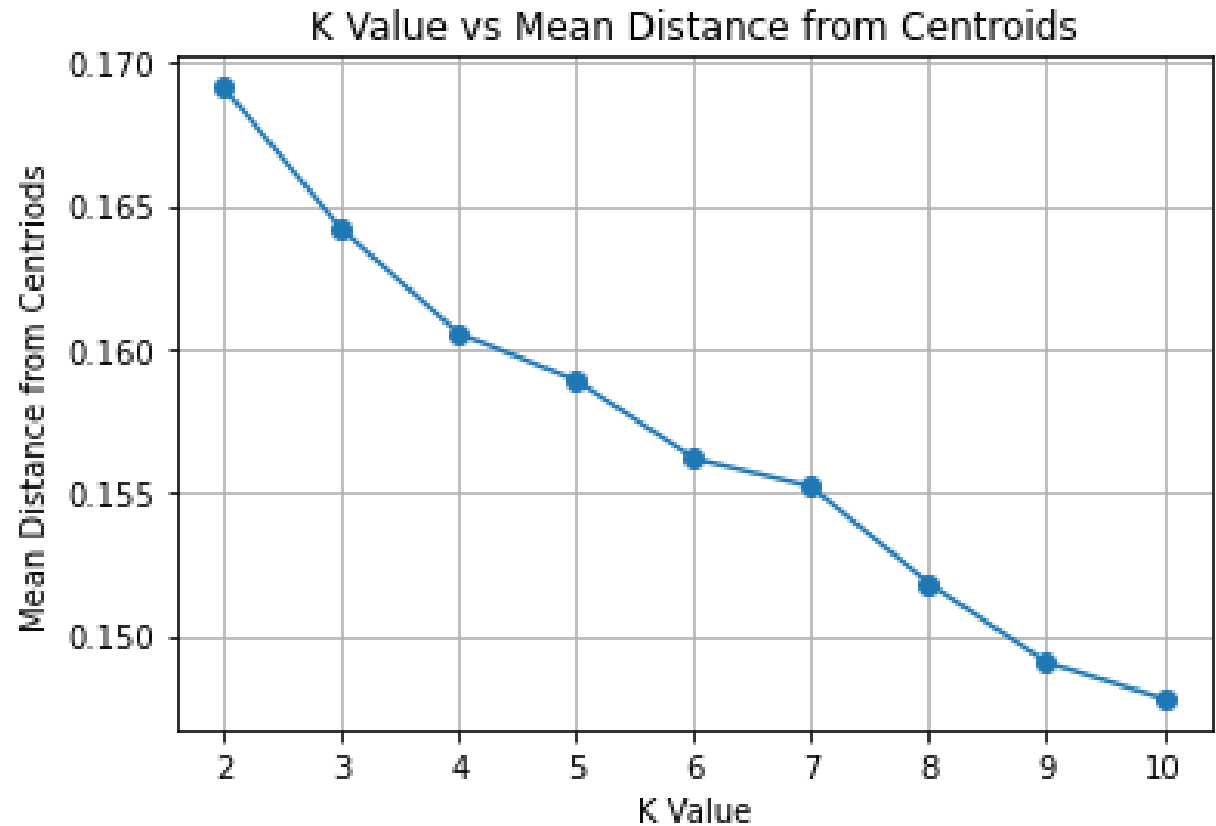
Yellow = Performing Arts Hall

Venue Frequency

- Utilized one-hot encoding to create frequency tables to view the most frequent venue that appears in each district. This will allow us to utilize numerical analysis on the venues.

Postcode	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Aquarium	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	E
E1	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.02	0.0	0.01	
E10	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.00	
E11	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.00	
E12	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.00	
E13	0.0	0.0	0.0	0.05	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.00	

	Postcode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	E1	Coffee Shop	Pub	Hotel	Café	Indian Restaurant
1	E10	Pub	Grocery Store	Café	Park	Chinese Restaurant
2	E11	Pub	Grocery Store	Park	Café	Restaurant
3	E12	Indian Restaurant	Train Station	Restaurant	Gym / Fitness Center	Park
4	E13	Pub	Café	Gym / Fitness Center	Convenience Store	Grocery Store

K-Means Parameter

- Shall use K-Means clustering to gain further insight on district to district venue categorization.

- Utilize the "elbow method" to find the most appropriate k-value (number of clusters to bin data points)

- First upward trend occurs at k=4 so this will be our K-Means parameter of choice.
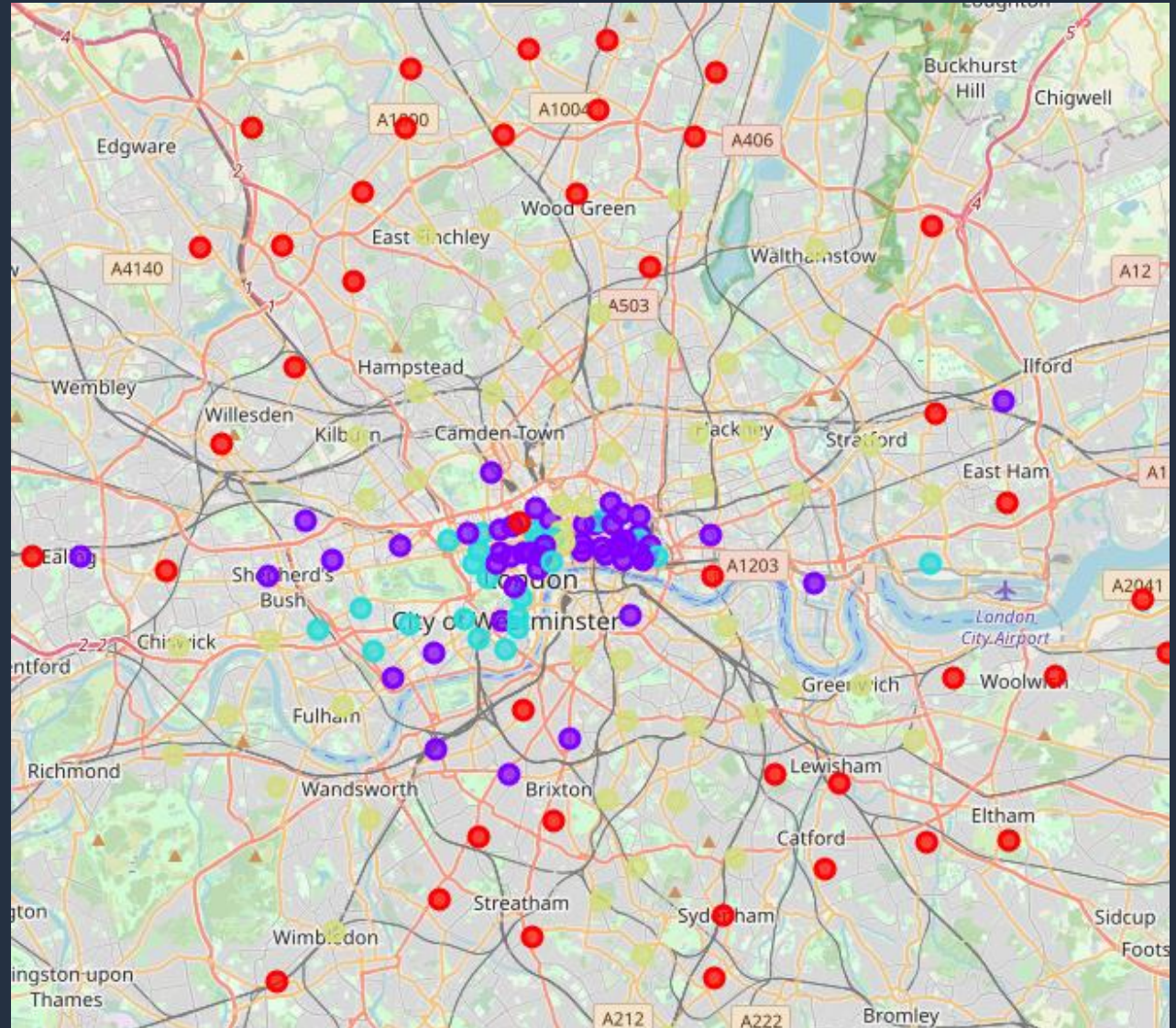
Data Clustering

Cluster 0 (Red):

- Abundance of grocery stores and coffee shops.

- Makes up outer perimeter

- Most likely high volume residential areas

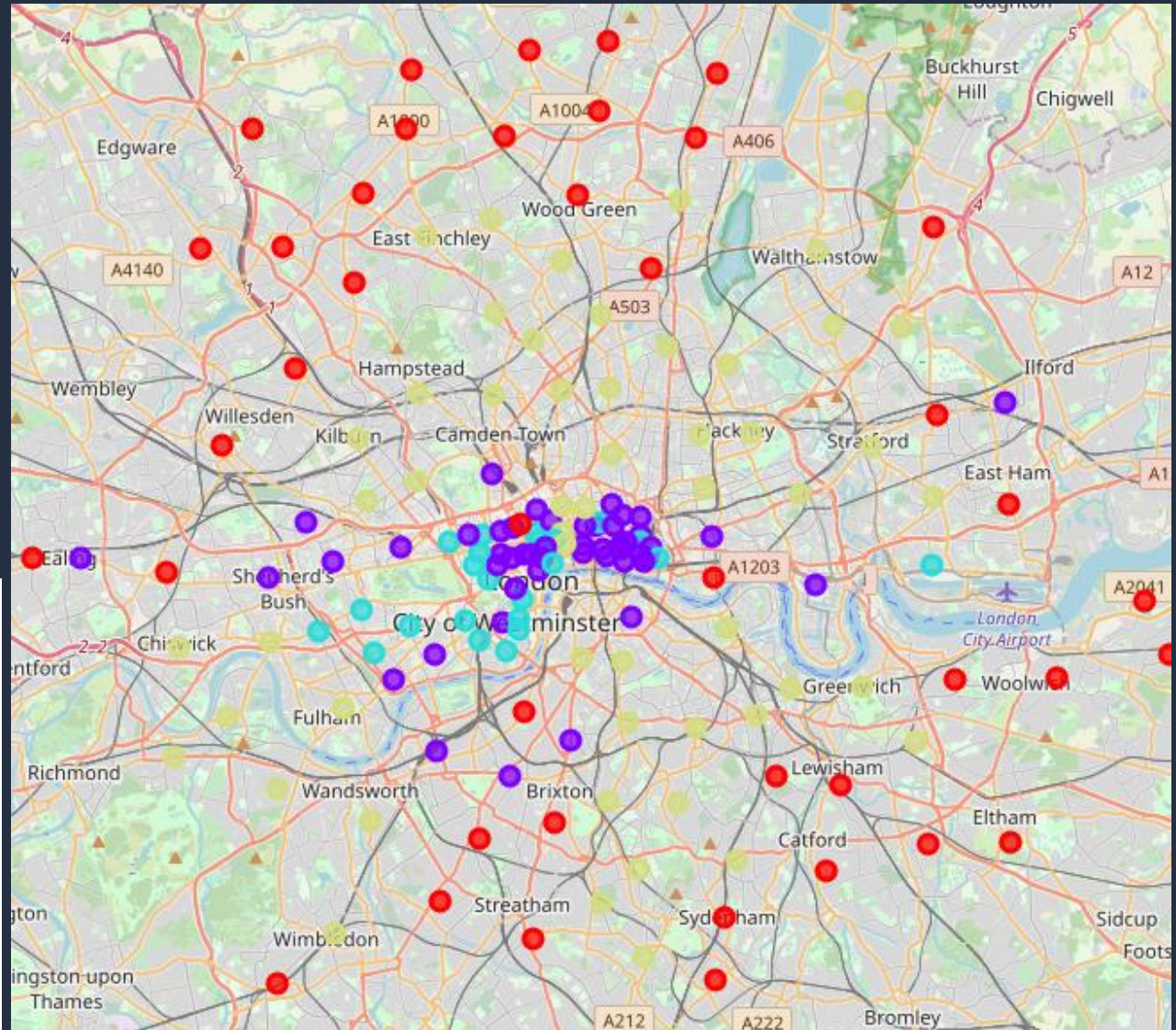|	Postcode	Population Density (Person per km2)	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
9	E18	6814.19	0	Italian Restaurant	Grocery Store	Coffee Shop	Bar	Supermarket
10	E1W	15889.09	0	Coffee Shop	Park	Italian Restaurant	Pub	Pizza Place
15	E6	8483.64	0	Hotel	Supermarket	Coffee Shop	Pub	Grocery Store
16	E7	15765.08	0	Grocery Store	Café	Hotel	Pub	Indian Restaurant
44	N11	7252.74	0	Grocery Store	Hardware Store	Electronics Store	Sporting Goods Shop	Café
45	N12	5971.68	0	Coffee Shop	Grocery Store	Supermarket	Café	Fast Food Restaurant
46	N13	8451.04	0	Grocery Store	Greek Restaurant	Park	Pub	Italian Restaurant
47	N14	5047.63	0	Pub	Grocery Store	Park	Gym / Fitness Center	Café
48	N15	14305.82	0	Café	Grocery Store	Coffee Shop	Bus Stop	Pizza Place
51	N18	6856.95	0	Pub	Turkish Restaurant	Grocery Store	Supermarket	Coffee Shop

Data Clustering

Cluster 1 (Purple):

- Large economic diversity

- Makes up center of London

- Great place for music venue, but already has a large amount

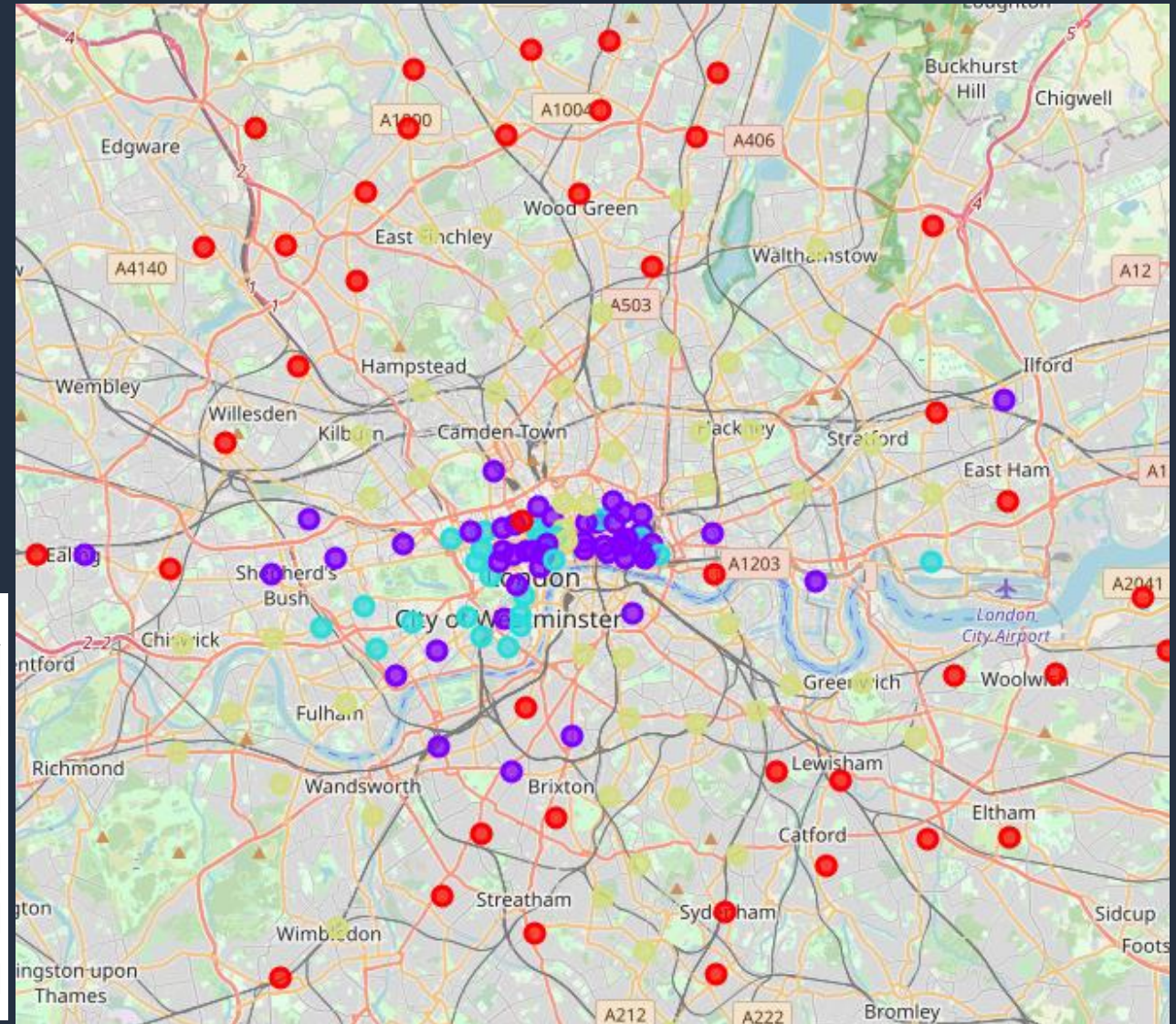|	Postcode	Population Density (Person per km2)	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	E1	20226.51	1	Coffee Shop	Pub	Hotel	Café	Indian Restaurant
3	E12	8532.36	1	Indian Restaurant	Train Station	Restaurant	Gym / Fitness Center	Park
5	E14	12126.99	1	Italian Restaurant	Hotel	Coffee Shop	Park	Indian Restaurant
19	EC1A	11559.26	1	Pub	French Restaurant	Wine Bar	Coffee Shop	Italian Restaurant
21	EC1N	24377.78	1	Coffee Shop	Sandwich Place	Pub	Food Truck	Gym / Fitness Center
23	EC1V	22143.68	1	Coffee Shop	Food Truck	Pub	Italian Restaurant	Café
24	EC1Y	25431.03	1	Food Truck	Coffee Shop	Pub	Hotel	Gym / Fitness Center
25	EC2A	8827.27	1	Gym / Fitness Center	Italian Restaurant	Food Truck	Coffee Shop	English Restaurant
26	EC2M	7393.10	1	Sandwich Place	Coffee Shop	Gym / Fitness Center	Boxing Gym	Café
28	EC2R	9318.18	1	French Restaurant	Wine Bar	Coffee Shop	Modern European Restaurant	Sushi Restaurant
29	EC2V	7911.11	1	Coffee Shop	Steakhouse	Clothing Store	Italian Restaurant	Plaza
30	EC2Y	17125.00	1	Italian Restaurant	Art Gallery	Deli / Bodega	Indie Movie Theater	Café

Data Clustering

Cluster 2 (Blue):

- Dominated by Hotels

- Makes up center of London with an outlier near the airport
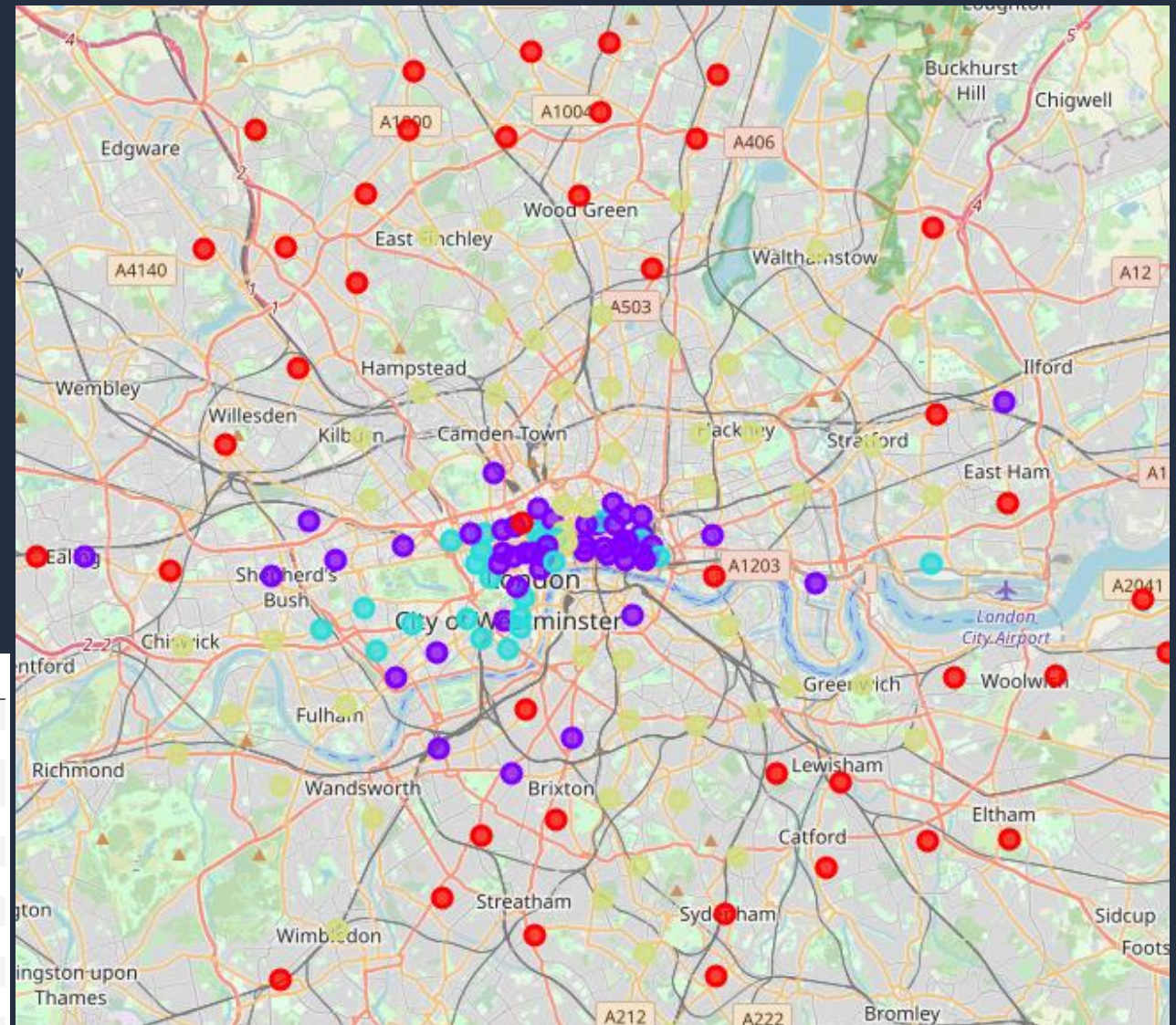
- Most likely high volume tourist areas

	Postcode	Population Density (Person per km2)	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
7	E16	6054.73	2	Hotel	Grocery Store	Coffee Shop	Café	Pub
20	EC1M	22290.91	2	Beer Bar	Hotel	Italian Restaurant	Plaza	Spanish Restaurant
27	EC2N	4711.11	2	Café	Event Space	Hotel	Boxing Gym	German Restaurant
33	EC3N	5814.29	2	Hotel	Sandwich Place	Cocktail Bar	Coffee Shop	Salad Place
113	SW1A	2831.09	2	Outdoor Sculpture	Pub	Plaza	Hotel	Monument / Landmark
115	SW1H	22129.41	2	Coffee Shop	Hotel	Hotel Bar	Juice Bar	Café
116	SW1P	17479.78	2	Hotel	Coffee Shop	Café	Italian Restaurant	Restaurant
117	SW1V	25081.48	2	Hotel	Pub	Pizza Place	Turkish Restaurant	Park
118	SW1W	13237.04	2	Hotel	Italian Restaurant	Pub	Bakery	Coffee Shop
119	SW1X	13427.50	2	Hotel	Café	Pub	Plaza	Italian Restaurant
125	SW5	25092.21	2	Hotel	Pub	Café	Garden	Italian Restaurant
127	SW7	15056.71	2	Café	Hotel	Exhibit	Italian Restaurant	Science Museum

Data Clustering

Cluster 3 (Yellow):

- Dominated by Pubs, Coffee Shops, and Grocery stores, as well as some diverse recreational spaces.

- Makes up middle perimeter of London area.

- High foot traffic could make for

	Postcode	Population Density (Person per km2)	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	E10	9601.98	3	Pub	Grocery Store	Café	Park	Chinese Restaurant
2	E11	7031.32	3	Pub	Grocery Store	Park	Café	Restaurant
4	E13	14978.84	3	Pub	Café	Gym / Fitness Center	Convenience Store	Grocery Store
6	E15	8026.76	3	Pub	Grocery Store	Park	Hotel	Café
8	E17	7783.74	3	Pub	Grocery Store	Coffee Shop	Café	Pizza Place
11	E2	18578.21	3	Coffee Shop	Pub	Café	Wine Bar	Cocktail Bar
12	E3	11816.18	3	Pub	Coffee Shop	Café	Pizza Place	Grocery Store
13	E4	3436.87	3	Pub	Grocery Store	Coffee Shop	Park	Italian Restaurant
14	E5	11985.40	3	Pub	Café	Park	Coffee Shop	Grocery Store
17	E8	17146.57	3	Pub	Café	Coffee Shop	Cocktail Bar	Bakery
18	E9	8553.43	3	Pub	Coffee Shop	Café	Bakery	Italian Restaurant

Conclusions

With all of this anlysis done we can narrow down our options to these several choices, in order of likelyhood of success:

- W1 (in particular W1K, W1J, W1S): This is our best choice. Relatively high population densities, with plenty of surrounding general venues to keep foot traffic flowing to spread interest. While there may be some nearby Jazz Clubs and a Concert Hall, this is a solid lack of a general music venue for quite a distance. Since most of this postcode area works, there is likely many real estate options as well.

- SE3: This ones biggest strength is its distance from any musical venues while still falling into cluster 3, which means that the nearby food options could be an advantage. The surrounding areas are also highly residential, so the music venue would cater to a large amount of people.

- NW (the northwest portion in particular): This is one that has one major flaw and one major strength. The strength is the utter lack of any music venues of any type whatsoever, making the desire for one probably rather high. Its biggest flaw is that it is almost entirely residential, so the music venue would have to create its buzz and support itself all on its own.