

Data Science Professional Certificate - Capstone Project

Moving Recommendations

From San Diego to London

Aime Lopez Aguilar - Feb 2021

Premise

- The problem is a personal one:

I'm moving from San Diego, USA to London, UK soon.

I want to leverage Data Science to help me select the best Neighborhood for me.

How?

- Using Machine Learning and Location Data:
 1. Webscrape the lists for neighbourhoods in San Diego and London.
 2. Leverage Foursquare location data to get top venues information for both cities.
 3. Train a Recommendation System with my ratings of San Diego neighbourhoods.
 4. Apply my Recommendation System to London neighbourhoods to find the top ranked based on my preferences
 5. Analyze collected data for top neighbourhoods in London and make my choice.

Data Collection: Acquisition and Cleaning

- Source: Wikipedia pages for San Diego and London communities:
 - San Diego: https://en.wikipedia.org/wiki/List_of_communities_and_neighborhoods_of_San_Diego
 - London: https://en.wikipedia.org/wiki/List_of_areas_of_London
- Obtain latitude and longitude for each neighbourhood using Nominatim from GeoPy Geolocator
- Cleanup data (unassigned locations and unnecessary information)
- Outcome: 2 dataframes with Neighborhood, Latitude and Longitude
 - San Diego: 101 Neighborhoods
 - London: 522 Neighborhoods

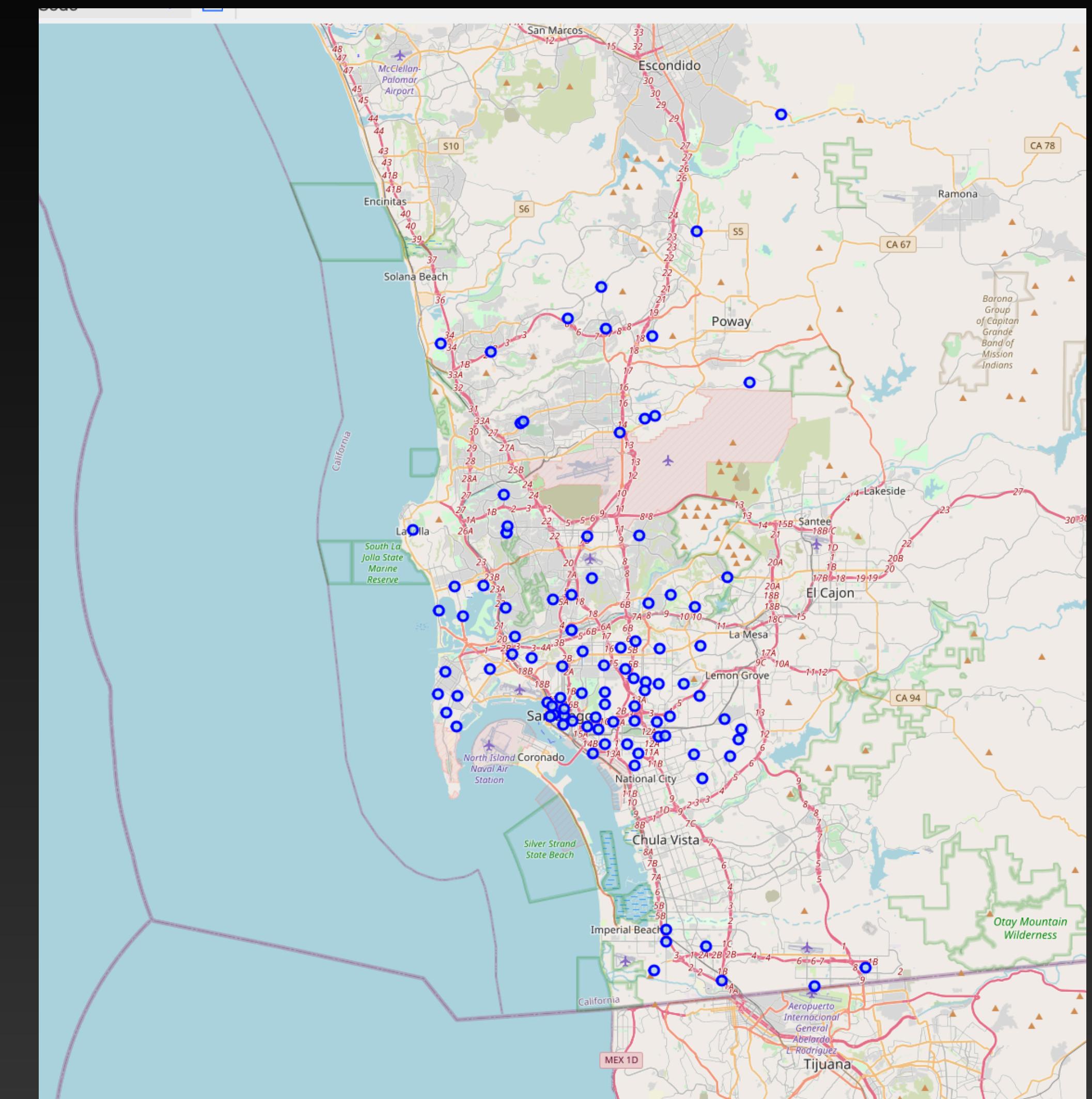
Data Collection: San Diego

Dataframe shape and preview:

(101, 3)

Out[9]:

	Latitude	Longitude	Neighborhood
0	32.731357	-117.146527	Balboa Park
1	32.728293	-117.162105	Bankers Hill
2	32.693886	-117.138007	Barrio Logan
3	32.691914	-117.036631	Bay Terraces
4	32.792333	-117.154206	Birdland



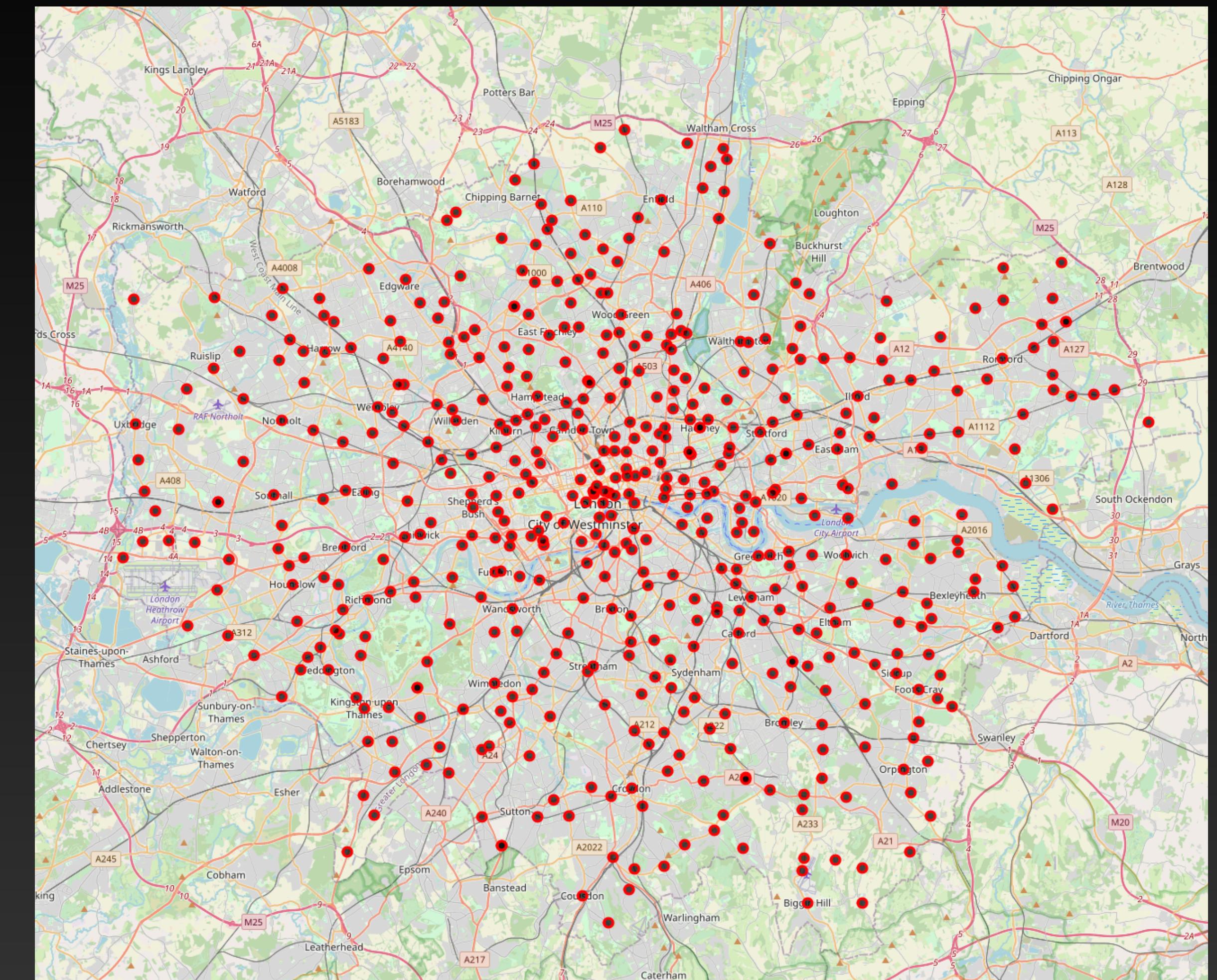
Data Collection: London

Dataframe shape and preview:

(531, 3)

Out[73]:

	Latitude	Longitude	Neighborhood
0	51.4876	0.11405	Abbey Wood
1	51.5081	-0.273261	Acton
2	51.3586	-0.0316347	Addington
3	51.3797	-0.0742821	Addiscombe
4	51.4354	0.125965	Albany Park



Data Processing: Adding top venues information to build Neighborhood profiles

San Diego:

Leverage Foursquare Location data to identify the top **100 Venues** within 800 m of each Neighbourhood

(3463, 7)

Out [83]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Balboa Park	32.731357	-117.146527	Balboa Park Fountain	32.731453	-117.146809	Fountain
1	Balboa Park	32.731357	-117.146527	San Diego Natural History Museum	32.732239	-117.147395	History Museum
2	Balboa Park	32.731357	-117.146527	San Diego Model Railroad Museum	32.731132	-117.148365	Museum
3	Balboa Park	32.731357	-117.146527	Botanical Building & Lily Pond	32.732237	-117.149288	Botanical Garden
4	Balboa Park	32.731357	-117.146527	Spanish Village Art Center	32.733395	-117.147370	Arts & Crafts Store

3463 venues identified, with 324 unique Venue Categories.

Data Processing: Adding top venues information to build Neighborhood profiles

London:

Leverage Foursquare Location data to identify the top **80 Venues** within 500 m of each Neighbourhood

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Elephant and Castle	51.494888	-0.100573	Theo's Pizzeria	51.492847	-0.100349	Pizza Place
1	Elephant and Castle	51.494888	-0.100573	Gymbox	51.494664	-0.097604	Gym
2	Elephant and Castle	51.494888	-0.100573	Corsica Studios	51.493546	-0.098541	Music Venue
3	Elephant and Castle	51.494888	-0.100573	Sabor Peruano	51.492887	-0.101063	Peruvian Restaurant
4	Elephant and Castle	51.494888	-0.100573	Southwark Playhouse	51.497779	-0.098603	Theater

11,314 venues identified, with 413 unique Venue Categories.

Using the Recommendation System:

- Rank my top 5 Best and top 5 Worst Neighborhoods in San Diego:
- And combine with the neighbourhood data to generate my profile (i.e. weight of each venue category)

	Neighborhood	Ranking
0	Barrio Logan	0.1
1	East Village	0.8
2	Gaslamp Quarter	9.0
3	Hillcrest	9.9
4	Kearny Mesa	0.5
5	La Jolla	10.0
6	North Park	9.8
7	San Ysidro	0.2
8	Tijuana River Valley	0.1
9	University Heights	9.5

ATM	Accessories Store	Adult Boutique	African Restaurant	Airport	Airport Lounge	Airport Service	American Restaurant	Amphitheater	Antique Shop	...	Vietnamese Restaurant	Warehouse Store	Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yoga Studio	Zoo	Zoo Exhibit	
0	0.114256	0.098	0.098	0.148438	0.0	0.0	0.0	1.539104	0.0	0.0	...	0.216608	0.0	0.188	0.542438	0.09	0.246438	0.0	0.196	0.0	0.0

Using the Recommendation System: Clean-up and Application

Since some venue types are present in San Diego but not in London, and vice versa, the profile and London dataframe had to be cleaned up to include only the categories included in both.

Use the System:

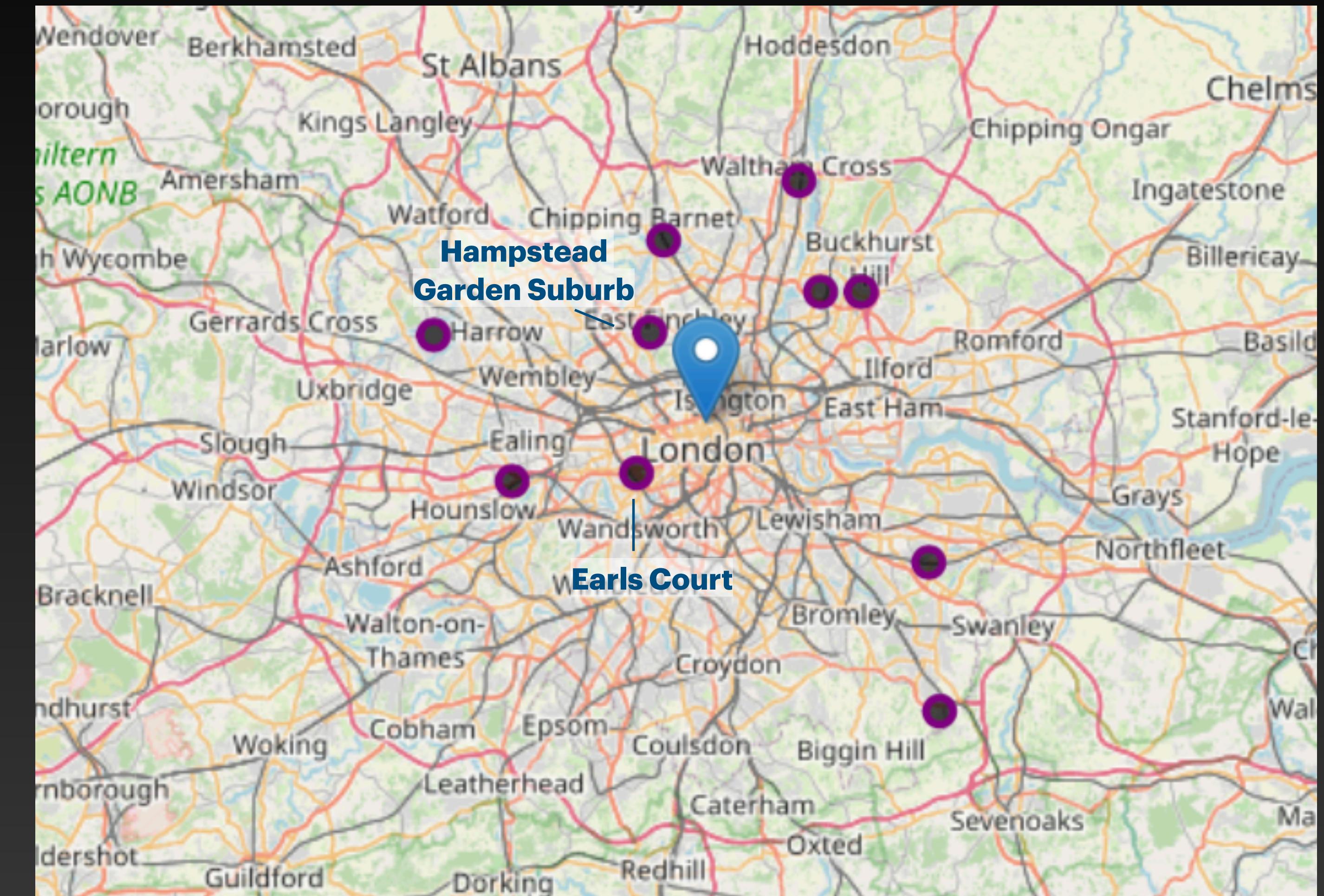
Use my profile to weigh all the venues in London neighbourhoods to generate a ranking based on my preferences

Neighborhood	Rating
Pratt's Bottom	0.039290
Oakleigh Park	0.028367
Freezywater	0.028340
Hampstead Garden Suburb	0.028139
Highams Park	0.023388
Brentford	0.022952
Eastcote	0.021922
Woodford	0.021584
Lamorbey	0.021457
Earls Court	0.021332

**Top 10
Recommended
Neighborhoods**

Top 10 Recommended Neighborhoods in London

Mapping recommended neighbourhoods with a marker for my future job location starts to narrow down the options.



Top 10 Recommended Neighborhoods in London

Most common venues in the recommendations:

	Neighborhood	Rating	Latitude	Longitude	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
0	Pratt's Bottom	0.039290	51.340884	0.111459	Bar	Coffee Shop	Zoo Exhibit	Farmers Market	Ethiopian Restaurant	Event Service	Event Space	Exhibit	Falafel Restaurant	Fast Food Restaurant
1	Oakleigh Park	0.028367	51.637667	-0.166225	Café	Zoo Exhibit	Farmers Market	Ethiopian Restaurant	Event Service	Event Space	Exhibit	Falafel Restaurant	Fast Food Restaurant	Cosmetics Shop
2	Freezywater	0.028340	51.675772	-0.031430	Café	Shoe Store	Pizza Place	Coffee Shop	Falafel Restaurant	Escape Room	Ethiopian Restaurant	Event Service	Event Space	Exhibit
3	Hampstead Garden Suburb	0.028139	51.580508	-0.180616	Park	Coffee Shop	Zoo Exhibit	Falafel Restaurant	Escape Room	Ethiopian Restaurant	Event Service	Event Space	Exhibit	Farmers Market
4	Highams Park	0.023388	51.606544	-0.008336	Gym	Coffee Shop	Zoo Exhibit	Falafel Restaurant	Escape Room	Ethiopian Restaurant	Event Service	Event Space	Exhibit	Farmers Market
5	Brentford	0.022952	51.486396	-0.321662	Coffee Shop	Gym	Convenience Store	Office	Furniture / Home Store	Pizza Place	Deli / Bodega	Sandwich Place	Ethiopian Restaurant	Event Service
6	Eastcote	0.021922	51.579542	-0.401653	Café	Hotel	Burger Joint	Coffee Shop	Sandwich Place	Grocery Store	Pharmacy	Zoo Exhibit	Exhibit	Ethiopian Restaurant
7	Woodford	0.021584	51.606806	0.034027	Coffee Shop	Hotel	Italian Restaurant	Café	Chinese Restaurant	Grocery Store	Restaurant	Bakery	Indian Restaurant	Pizza Place
8	Lamorbey	0.021457	51.435509	0.101805	Coffee Shop	Hotel	Burger Joint	Train Station	Grocery Store	Gym / Fitness Center	Dessert Shop	Restaurant	Mexican Restaurant	Pizza Place
9	Earls Court	0.021332	51.491612	-0.193903	Hotel	Café	Pub	Italian Restaurant	Garden	Grocery Store	Thai Restaurant	Historic Site	Pizza Place	Coffee Shop

... recommendations definitely have thing I like! (it seems the system works)

Conclusion

The *Recommendation System* has done a good job at narrowing down the options and identifying neighbourhoods that have things I like. Added to location data and neighbourhood profile information, I was able to make my decision.

I'm moving to:

Hampstead Garden Suburb

#4 ranked neighbourhood based on my profile, close to my future job and full of things I like (parks, farmers markets, coffee shops, and events venues).