IBM Data Science Course

Capstone Project

Business Problem

Evaluate how German venues (e.g. restaurants, culture, museums, shops, etc.) search the proximity of German, Austrian and Swiss companies /embassies/schools abroad.

For this study the city of London is evaluated and how German venues aggregate next to foreign offices of traditional, world leading German/Austrian/Swiss companies and also around embassies and schools of those 3 countries.

Conclusion

A strong relationship was found between German venues in London which are located to German/Austrian/Swiss enterprises and institutions.

An opportunity would exist to filter further for restaurants and use the results as guidance to evaluate the opportunity to open a 'German' style restaurant in these areas.

Data Used

Wikipedia to find the largest companies of Germany, Austria and Switzerland
https://en.wikipedia.org/wiki/List_of_largest_German_companies
https://en.wikipedia.org/wiki/List_of_largest_Austrian_companies
https://en.wikipedia.org/wiki/List_of_Swiss_companies_by_revenue

Company webpages to find the London office locations
various

Embassy websites to find the London address of the German, Austrian and Swiss representation
https://uk.diplo.de/uk-en/01/embassy
https://www.bmeia.gv.at/en/austrian-embassy-london
https://www.eda.admin.ch/london

School webpage of German School in London
https://www.dslondon.org.uk

Areas of London
https://en.wikipedia.org/wiki/List_of_areas_of_London

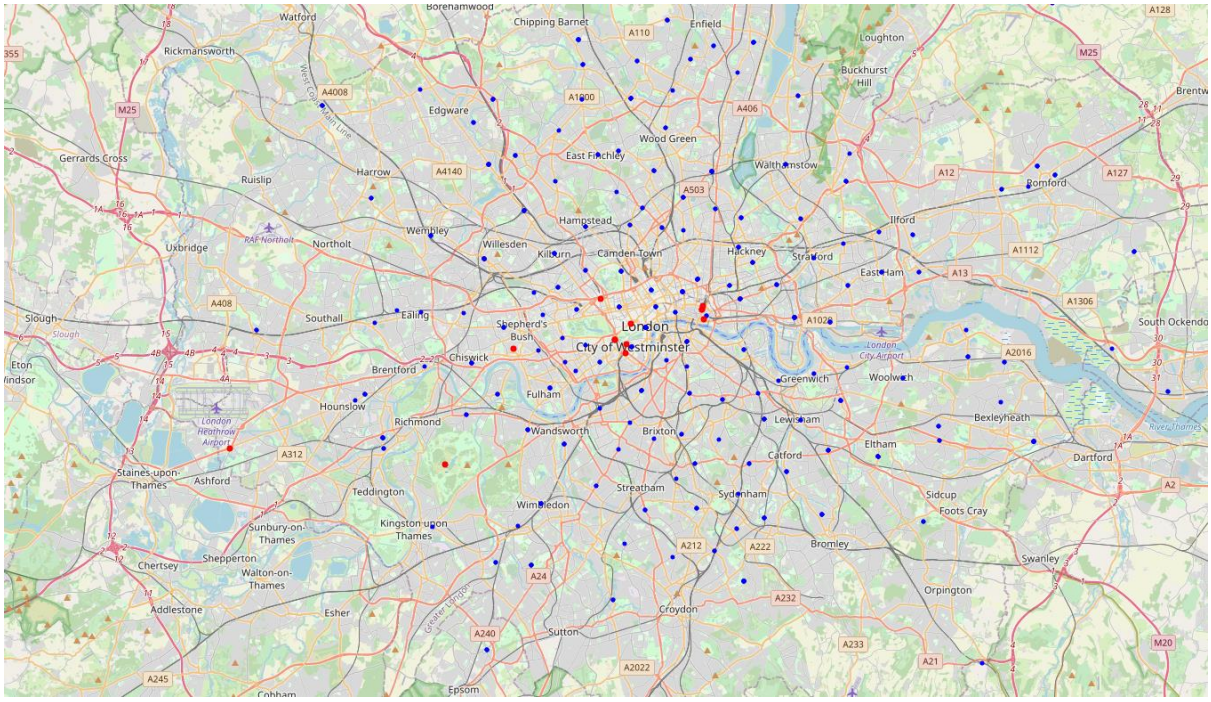
With the information from above ('Areas of London') a data frame was generated showing various areas of London. As a next step this data frame was cleaned up. In order to plot those areas on a map latitude and longitude values were generated and merged with the table above.

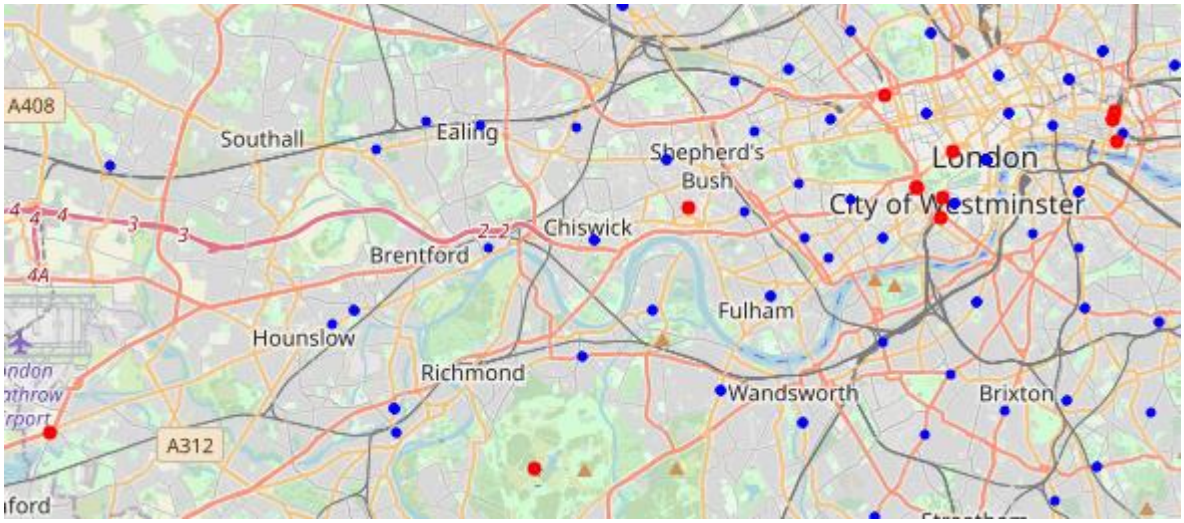
	Location	London Borough	Post town	Postcode district	Latitude	Longitude
0	Abbey Wood	Bexley	LONDON	SE2	51.492450	0.121270
1	Acton	Ealing	LONDON	W3	51.513240	-0.267460
2	Addington	Croydon	CROYDON	CR0	51.384755	-0.051498
3	Addiscombe	Croydon	CROYDON	CR0	51.384755	-0.051498
4	Albany Park	Bexley	BEXLEY	DA5	51.506420	-0.127210
...	...	...	...	...	...	...
526	Woolwich	Greenwich	LONDON	SE18	51.482070	0.071430
527	Worcester Park	Sutton	WORCESTER PARK	KT4	51.506420	-0.127210
528	Wormwood Scrubs	Hammersmith and Fulham	LONDON	W12	51.506450	-0.236910
529	Yeading	Hillingdon	HAYES	UB4	51.506420	-0.127210
530	Yiewsley	Hillingdon	WEST DRAYTON	UB7	51.506420	-0.127210

The next step was to bring information from various webpages under 'Data Used' into a data frame. A selection of the top 3 companies, all embassies and schools was considered sufficient enough. Like before the addresses got matched with latitude and longitude values.

	Entity	Location	Postcode district	Latitude	Longitude
0	German Embassy	Belgravia	SW1X	51.500645	-0.150956
1	Austrian Embassy	Belgravia	SW1X	51.500645	-0.150956
2	Swiss Embassy	Marylebone	W1H	51.520290	-0.161740
3	German School	Richmond	TW10	51.440750	-0.282092
4	BASF	London	EC2N	51.515305	-0.083495
5	Munich RE	London	EC3M	51.510555	-0.082258
6	E.ON	London	SW1V	51.494219	-0.142810
7	SAP	Feltham	TW14	51.448448	-0.448491
8	Erste Group	London	EC2N	51.515305	-0.083495
9	OMV	Westminster	SW1E	51.498400	-0.142026
10	VoestAlpine	Hammersmith	W6	51.496170	-0.229350
11	Glencore	Mayfair	W1J	51.508380	-0.138806
12	UBS	London	EC2M	51.516950	-0.083340

Both data sets got plotted on the map below to show distribution of companies, embassies and schools (red dots) in relation to remaining areas in London (blue dots).



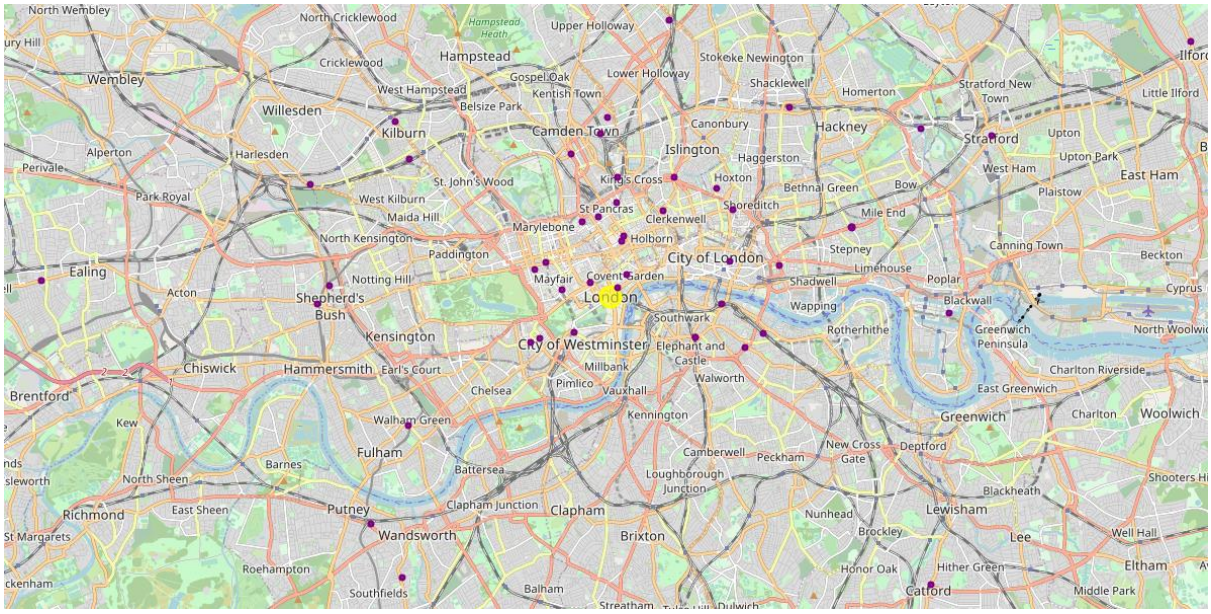


Based on the plots from above one can see a heavy concentration in the city centre stretching towards the western part of London.

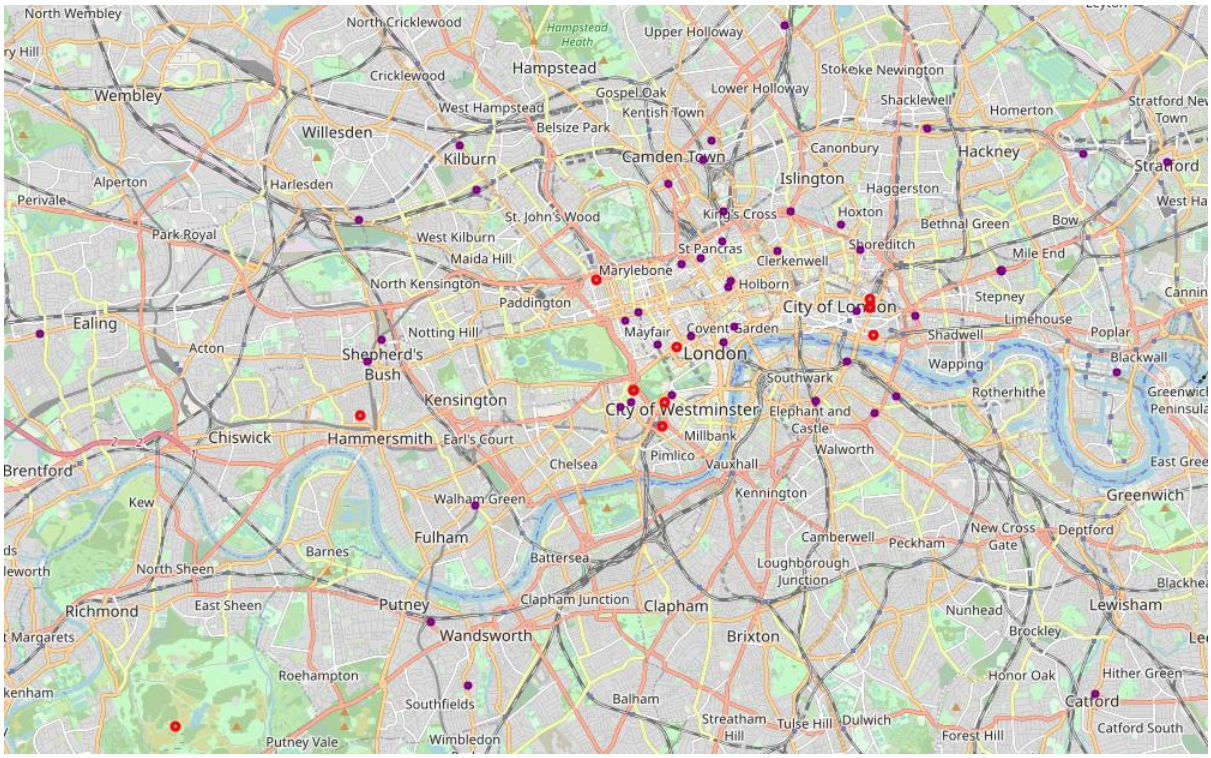
By having now an overview of the location of German/Austrian/Swiss entities in London the ‘Foursquare’ API could be used to list ‘German’ venues against them and check for proximity.

	name	categories	location.lat	location.lng
0	German Historical Institute	[{'id': '4bf58dd8d48988d190941735', 'name': 'H...	51.520094	-0.123265
1	German Doner Kebab	[{'id': '5283c7b4e4b094cb91ec88d8', 'name': 'D...	51.532408	-0.106242
2	German Embassy London	[{'id': '4bf58dd8d48988d12c951735', 'name': 'E...	51.497723	-0.154989
3	German Deli (Food Stall)	[{'id': '4bf58dd8d48988d10d941735', 'name': 'G...	51.505739	-0.089977
4	Buy German Driver License Online	[{'id': '52e81612bcbc57f1066b7a42', 'name': 'D...	51.509143	-0.125259
5	German Gymnasium	[{'id': '52e81612bcbc57f1066b79f9', 'name': 'M...	51.532392	-0.125267
6	Jost German Design	[{'id': '4bf58dd8d48988d1e1931735', 'name': 'A...	51.510170	-0.134421
7	Hooray's German Hot Dogs	[{'id': '4bf58dd8d48988d16f941735', 'name': 'H...	51.511919	-0.122090
8	German Doner Kebab	[{'id': '5283c7b4e4b094cb91ec88d8', 'name': 'D...	51.480196	-0.196337
9	German Doner Kebab	[{'id': '5283c7b4e4b094cb91ec88d8', 'name': 'D...	51.544297	-0.200867

Plotting German places (in purple) from ‘Foursquare’ gives following map:



Combining now both data sets, German/Austrian/Swiss institutions (red dots) and German venues from ‘Foursquare’ (purple dots) give following picture.



Results Interpretation

It seems that there is a strong correlation between ‘German’ venues and overseas offices. The reason might be that a higher expatriate population form those native countries work in those offices and therefore enjoy some ‘feeling at home’ entertainment.

Opportunities

The ‘Foursquare’ data could be filtered further and check for which venues are restaurants. This filtered set could then be used to evaluate the potential to open a German style restaurant in those areas.

Currently the red dots are based on a list of a few entries from an individual web search. This could be improved by getting data from the German/Austrian/Swiss embassies where their countrymen/women live and therefore this would give a much better resolution.

Potentially this brings a hot topic of data sensitivity as giving such sensitive information away might not be allowed which I guess applies to many application in the ‘Data Science’ that the technology could give great results but could be in contradiction of individual rights.