IBM Capstone Project

Samuel Hartwell

Project Description

The city used for the project will be Hamburg. The problem will be looking at types of restaurant in the city, such as Italian, German, Korean, Chinese etc, and trying to cluster the different neighbourhoods by which type of restaurants are most populous in which areas, such that the client, a new restaurateur who wants to open a Korean restaurant, can see areas of potentially high and low competition.

Data Description

The data will be the locations and types of restaurants in each city borough, as well as their average rating. To be used to determine how much competition they could provide. The data will be used to look at areas of high and low restaurant density (of all types of cuisine), as well as look at the frequency of korean restaurants present in each city sector. The data will be sourced from Foursquare if I can ever get the site to let me make an account. Else I will find a CSV dataset somewhere.

The location data will be taken from https://www.citypopulation.de/en/germany/hamburg/admin/

Raw data scraped from Webpage

	Name	Status	PopulationEstimate2002-12-31	PopulationEstimate2007-12-31	PopulationEstimate2012-12-31	PopulationEstimate2017-12-31	PopulationEstimate2019-12-31	Unnamed: 7
0	Altona	Borough	...	257299	257412	270263	275265	→
1	Altona-Altstadt	Quarter	27550	26918	28174	29034	29170	→
2	Altona-Nord	Quarter	21662	22521	21759	22137	24536	→
3	Bahrenfeld	Quarter	26343	27193	27165	31047	30203	→
4	Blankenese	Quarter	13452	13435	13016	13407	13730	→
...	...	...	...	...	...	...	...	...
108	Volksdorf	Quarter	18818	19362	20115	20625	20978	→
109	Wandsbek	Quarter	32583	33034	33131	34469	36444	→
110	Wellingsbüttel	Quarter	9432	9829	10263	10506	10848	→
111	Wohldorf-Ohlstedt	Quarter	4189	4405	4423	4656	4650	→
112	Hamburg	City	1728806	1770629	1775659	1860759	1899160	NaN

113 rows × 8 columns

Reduced to useable data

	Name	Status	PopulationEstimate2019-12-31
0	Altona	Borough	275265
1	Altona-Altstadt	Quarter	29170
2	Altona-Nord	Quarter	24536
3	Bahrenfeld	Quarter	30203
4	Blankenese	Quarter	13730
...	...	...	...
108	Volksdorf	Quarter	20978
109	Wandsbek	Quarter	36444
110	Wellingsbüttel	Quarter	10848
111	Wohldorf-Ohlstedt	Quarter	4650
112	Hamburg	City	1899160

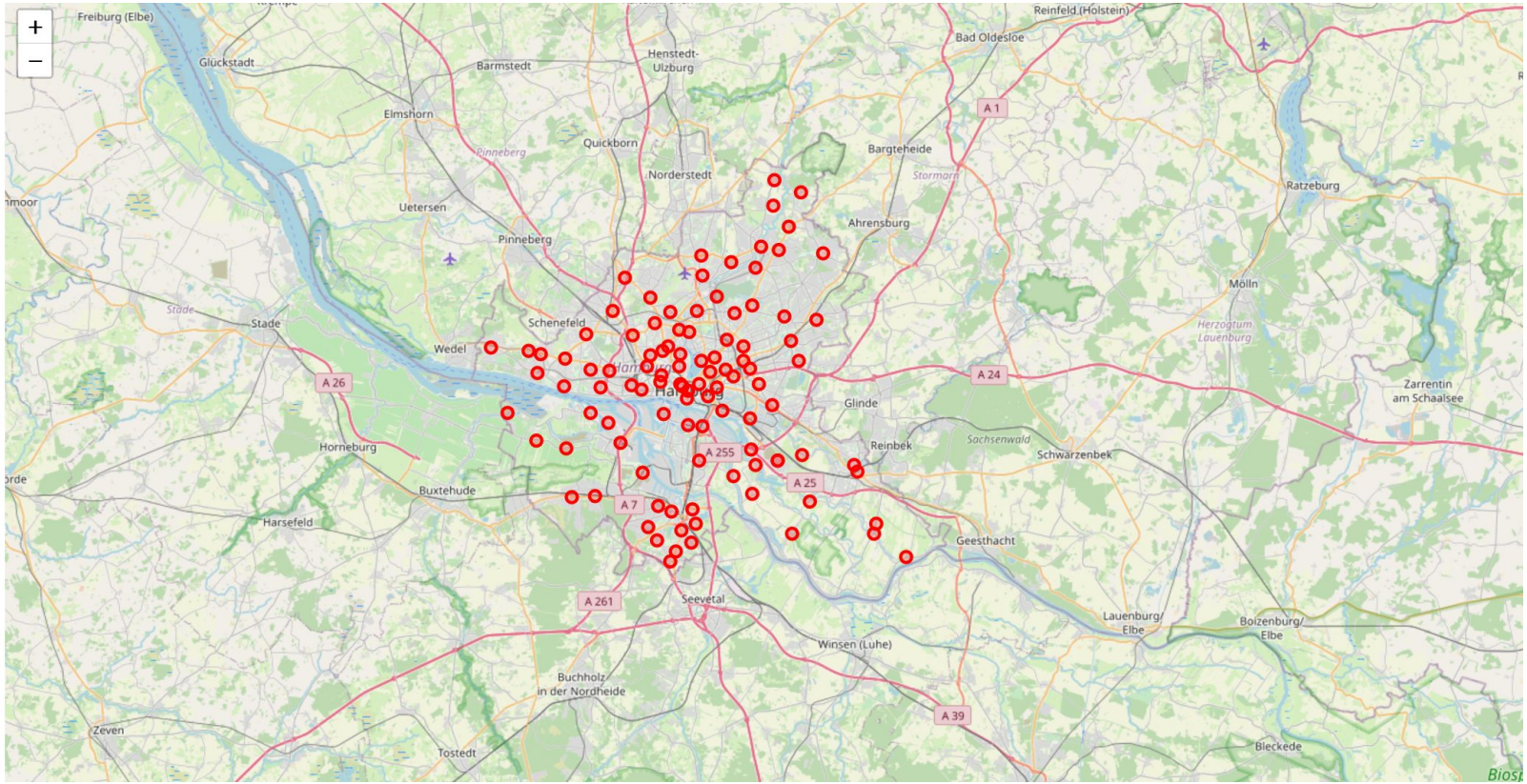
113 rows × 3 columns

Cleaned the data up and found the latitude and longitude values for each location

	Name	Status	PopulationEstimate2019-12-31	latitude	longitude
1	Altona-Altstadt	Quarter	29170	53.55078	9.94071
2	Altona-Nord	Quarter	24536	53.56804	9.94814
3	Bahrenfeld	Quarter	30203	53.56466	9.90166
4	Blankenese	Quarter	13730	53.56314	9.81232
5	Groß Flottbek	Quarter	11111	53.56564	9.87784
...	...	...	...	...	...
107	Tonndorf	Quarter	15148	53.58697	10.12713
108	Volksdorf	Quarter	20978	53.65131	10.16703
109	Wandsbek	Quarter	36444	53.57205	10.06707
110	Wellingsbüttel	Quarter	10848	53.64064	10.08302
111	Wohldorf-Ohlstedt	Quarter	4650	53.69557	10.13900

105 rows × 5 columns

Plotted the boroughs on a map using folium



Next steps

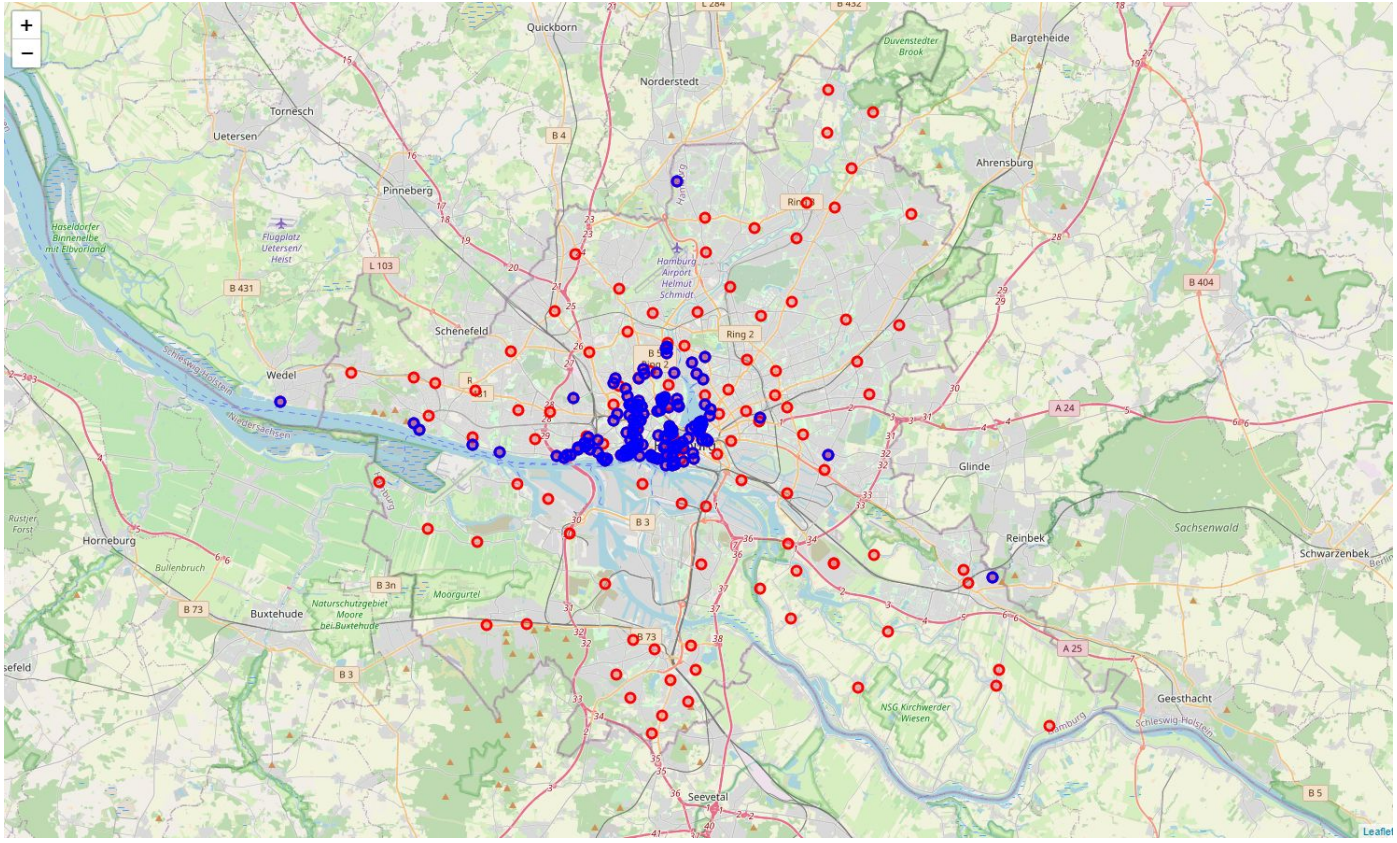
Foursquare doesn’t want to work, so instead I have retrieved a CSV file with different restaurant data from the web, https://www.europeandataportal.eu/data/datasets/32d2e295-08c0-446e-8bf8-5a7d3b315d51?locale=de

Using this I can retrieve the latitude and longitude data and plot the restaurants also on the map.

Eventual dataframe.

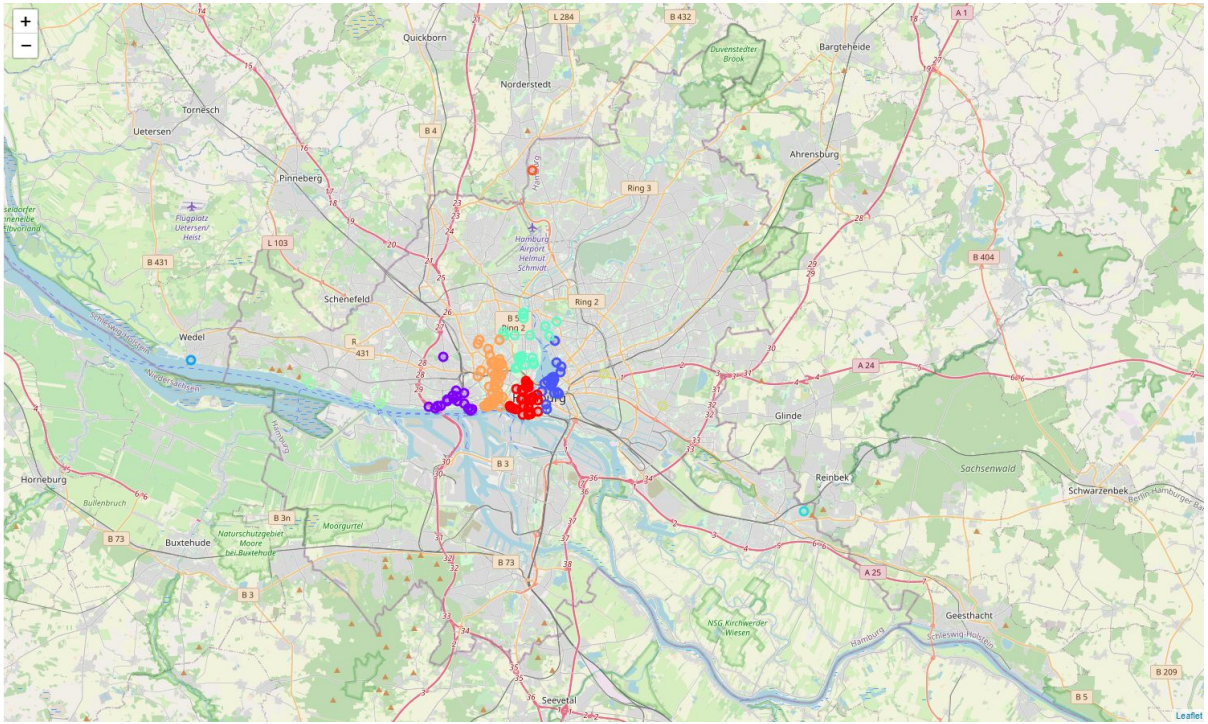
	Restaurant	Adresse	latitude	longitude
0	Abendmahl	Hein-Köllisch-Platz 6, 20359 Hamburg	53.547597	9.956741
1	Al Volo	Eppendorfer Weg 211, 20253 Hamburg	53.581407	9.970589
2	Alsterperle	Eduard-Rhein-Ufer 1, 22087 Hamburg	53.567994	10.015723
3	Alt Hamburger Aalspeicher	Deichstraße 43, 20459 Hamburg	53.545489	9.986939
4	An Khang Quan	Hoheluftchaussee 86, 20253 Hamburg	53.583590	9.970379
...	...	...	...	...
149	Williamine	Kleiner Schäferkamp 16, 20357 Hamburg	53.567116	9.968927
150	Zipang	Eppendorfer Weg 171, 20253 Hamburg	53.579587	9.966488
151	Zum Alten Lotsenhaus	Övelgönne 13, 22605 Hamburg	53.544837	9.912018
152	Zum Schiffchen	Großer Grasbrook 9, 20457 Hamburg	53.541788	9.993654
153	Zur Traube	Karl-Theodor-Straße 4, 22765 Hamburg	53.548572	9.931221

Map of Hamburg with boroughs in red, and restaurants in blue



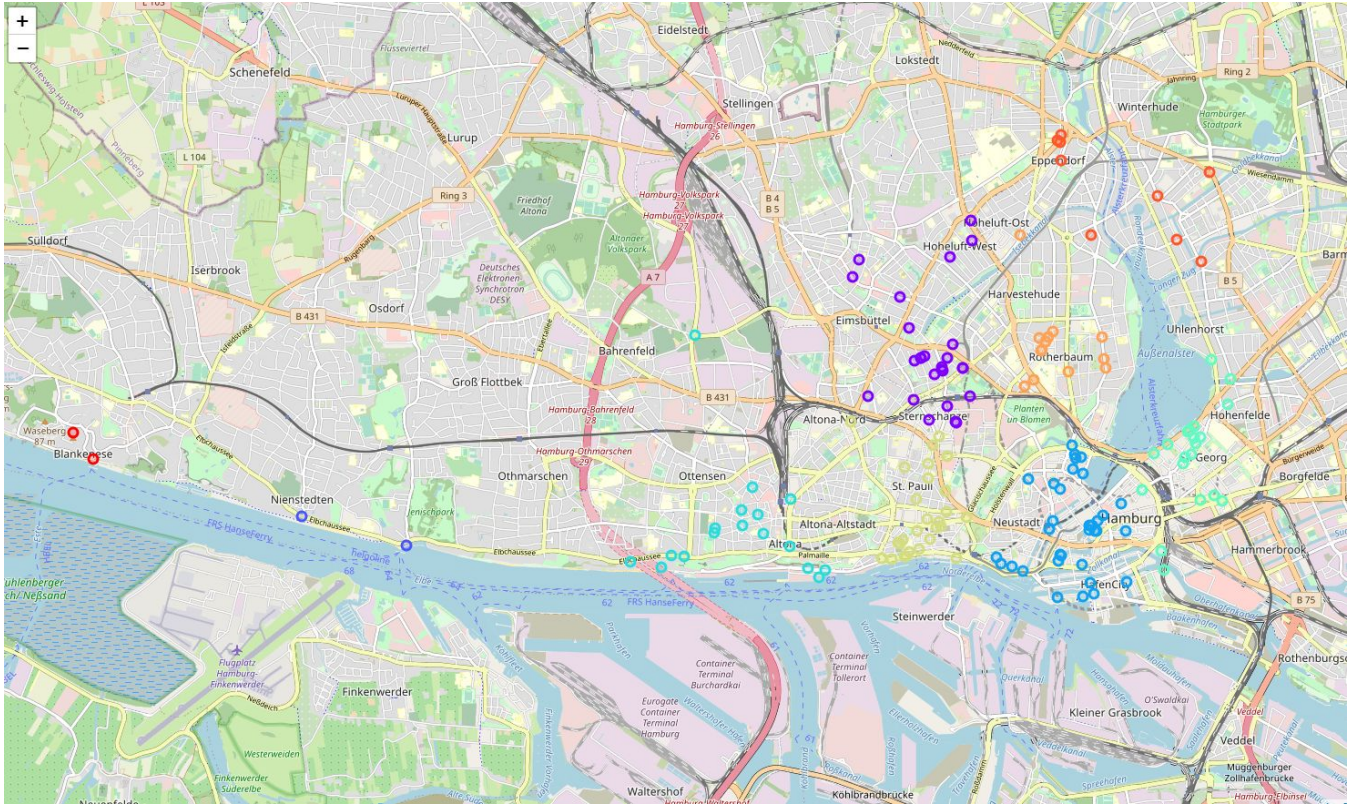
Clustering

Using 10 clusters we can see there are a few outliers in the data. Let’s remove those and try to recluster the data.



Clustered restaurants

Here we have ten clusters of restaurants in the city.



Conclusion

From the data is is apparent that there is a lower density of restaurants in the areas of eppendorf, and in Rotherbaum.

As such the recommendation for the client would be to open their new restaurant in one of these locations.