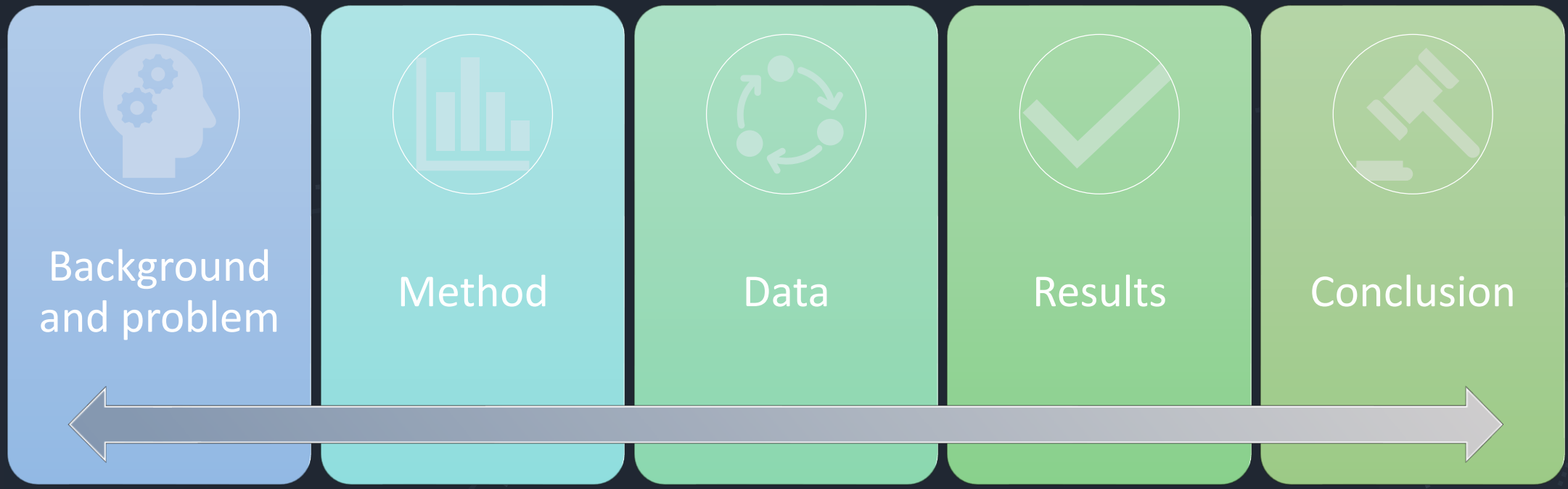
Expansion to Germany

Coursera Capstone Project

Agenda

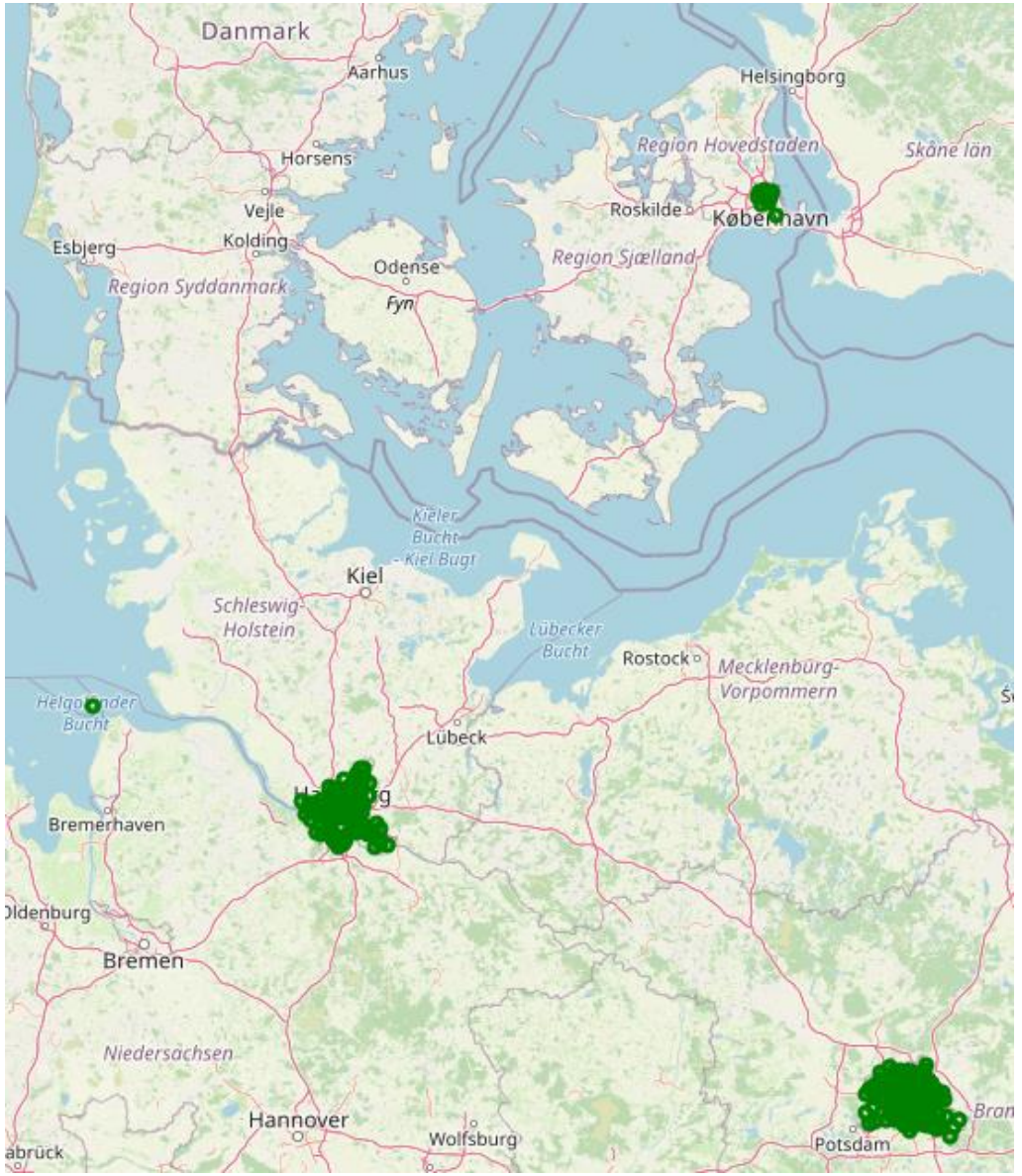


Background and problem

- The client, based in Copenhagen, is an entertainment organizer company who wants to expand into Germany
- Being scarce for resources they can only choose 1 city, and its analysts have narrowed this down to Hamburg or Berlin
- The client then wants to know where they should establish a new office

The **business problem** is then:
“Which of the two cities(Berlin, Hamburg) should the Danish event-organiser company expand to? “

A derived **data-science problem** is then:
“Between Hamburg and Berlin, which city’s neighbourhoods are most similar to those of Copenhagen in terms of entertainment venues?”



Method

- The information the names and populations of the locations were obtained via Citypopulation.de, and Wikipedia.
- Venue types were obtained through foursquare API
- Geospacial data was extracted using GeoPy and Nominatim.

DATA	SOURCE
COPENHAGEN	https://www.citypopulation.de/en/denmark/copenhagen/
HAMBURG	https://www.citypopulation.de/en/germany/hamburg/admin/
BERLIN	https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin
VENUES	https://api.foursquare.com/v2/venues/
GEO SPECIAL DATA(API)	https://nominatim.org/

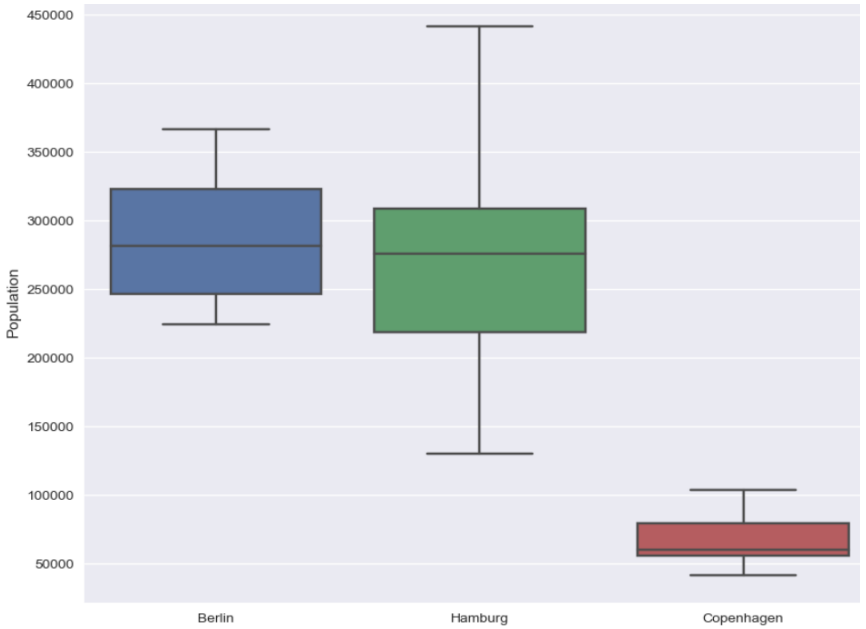
Neighborhoods and Boroughs

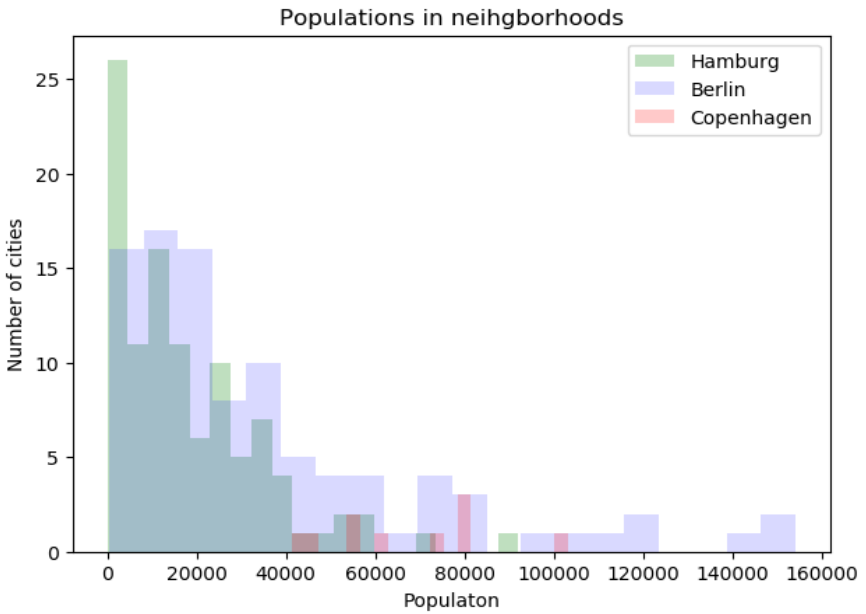
- Copenhagens Boroughs were considered the same as neighborhoods to keep comparability of areas
- Hamburg has 104 quarters ascribed to its 7 boroughs, 2 were not used
- Berlin has 96 officially recongnised neighborhoods in its 12 boroughs.

COPENHAGEN	POP	BERLIN	POP(2010)	HAMBURG	POP
AMAGER ØST	59,803	Charlottenburg-Wilmersdorf	319,628	Altona	275,265
AMAGER VEST	78,973	Friedrichshain-Kreuzberg	268,225	Bergedorf	130,260
BISPEBJERG	55,172	Lichtenberg	259,881	Eimsbüttel	267,053
BRØNSHØJ-HUSUM	44,784	Marzahn-Hellersdorf	248,264	Hamburg-Mitte	301,546
INDRE BY	55,866	Mitte	332,919	Hamburg-Nord	314,595
NØRREBRO	80,254	Neukölln	310,283	Harburg	169,426
ØSTERBRO	79,803	Pankow	366,441	Wandsbek	44,101
VALBY	60,308	Reinickendorf	240,454		
VANLØSE	41,195	Spandau	223,962		
VESTERBRO/KONGENS ENGHAVE	72,688	Steglitz-Zehlendorf	293,989		
FREDERIKSBERG	103,192	Tempelhof-Schöneberg	335,060		
		Treptow-Köpenick	241,335		

Neighborhood borough population distributions

- Understandably the boroughs of berlin and Hamburg are more populoursthan the ones of Copenhagen.
- Looking at the distribution of populations we can see that not breaking Copenhagen to smaller bits is a good idea since most values are right in the middle.





Venue Data

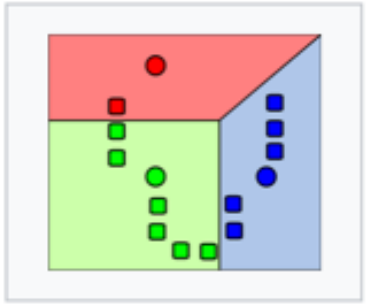
- Venues were extracted from Foursquare API using for each location(using GeoPy to get geospacial coordinates)
- The table to the right shows the data that will be used in the algorithm.
- In total there were 8516 different venues in 416 unique categories

Neighborhood	ATM	Accessories Store	Adult Boutique	Advertising Agency	Afghan Restaurant	African Restaurant	Airport	Airport Lounge	Airport Service	...	Waterfront	Whisky Bar	Windmill	Wine Bar	Wine Shop
Adlershof, Berlin	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000
Alt-Hohenschönhausen, Berlin	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000
Alt-Treptow, Berlin	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000
Altglienicke, Berlin	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000
Baumschulenweg, Berlin	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Wilhelmsburg, Hamburg	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000
Wilstorf, Hamburg	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000
Winterhude, Hamburg	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000
Wohldorf-Ohlstedt, Hamburg	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000
Østerbro, Copenhagen	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.011905

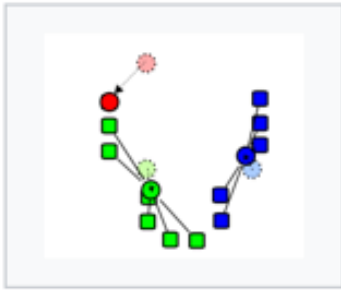
Method

- Kmeans clustering will be used to classify each neighbor
  - 5 clusters
  - Using “Euclidian Distance” for optimisation

1. *k* initial "means" (in this case *k*=3) are randomly generated within the data domain (shown in color).



2. *k* clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means.

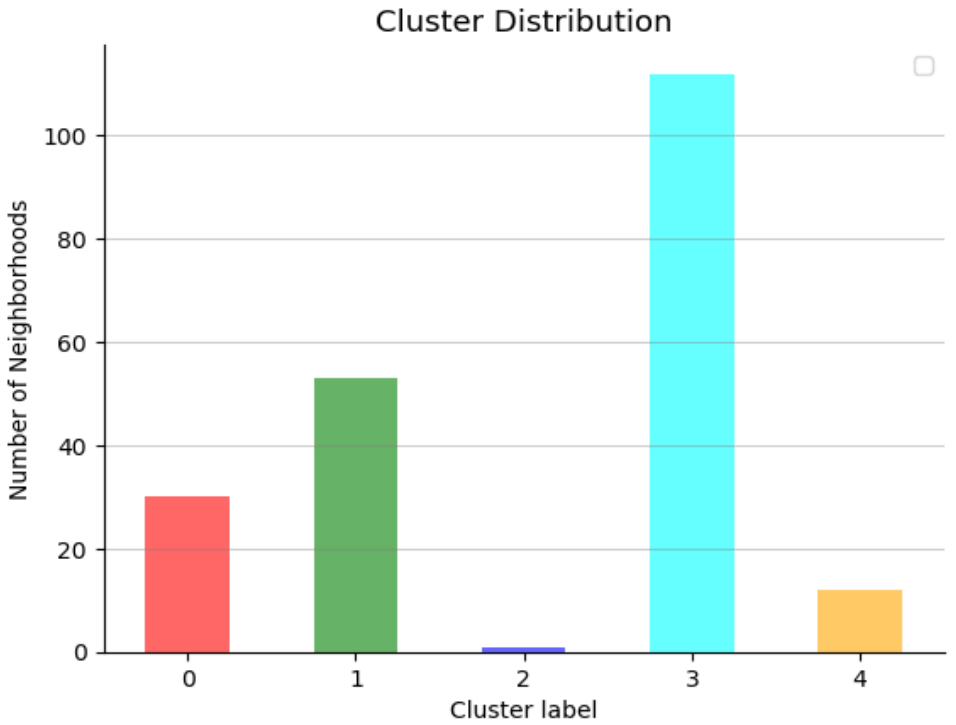


3. The centroid of each of the *k* clusters becomes the new mean.

4. Steps 2 and 3 are repeated until convergence has been reached.

Overall Results

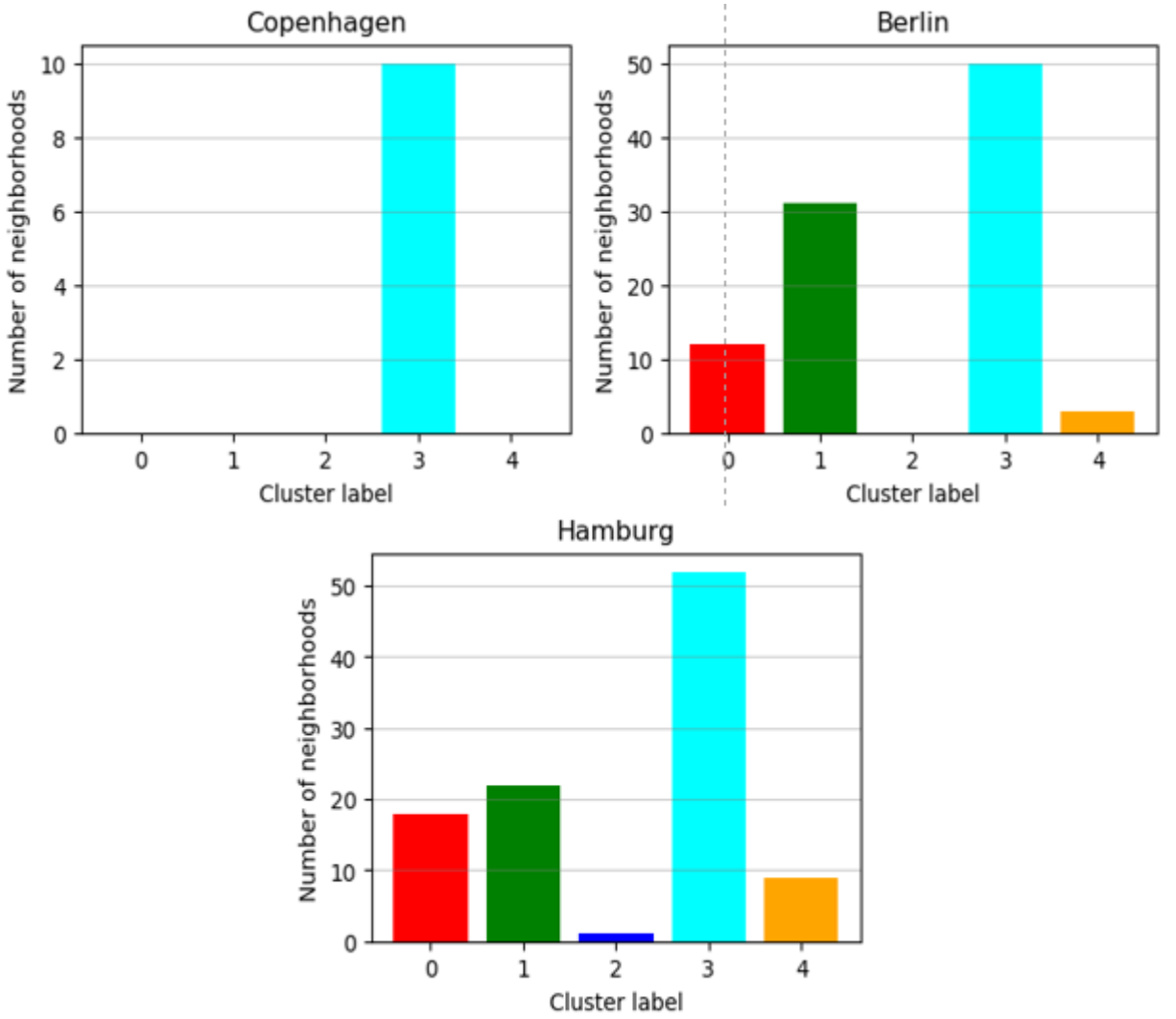
Most of the neighborhoods fell into cluster 3, which represented somewhat central areas of the three cities



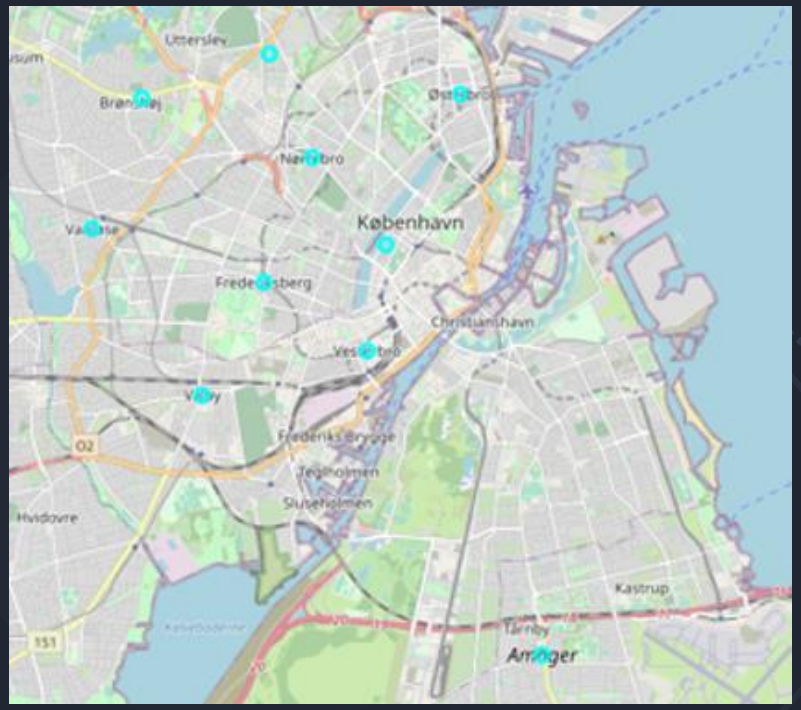
CLUSTER LABEL	NUMBER OF NEIGHBORHOODS
3	112
1	53
0	30
4	12
2	1

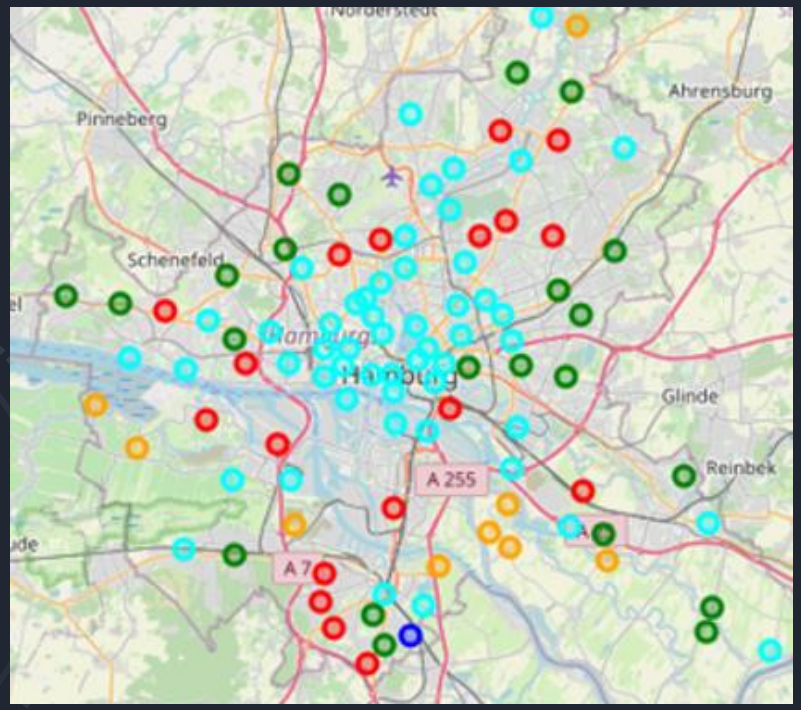
Results by City

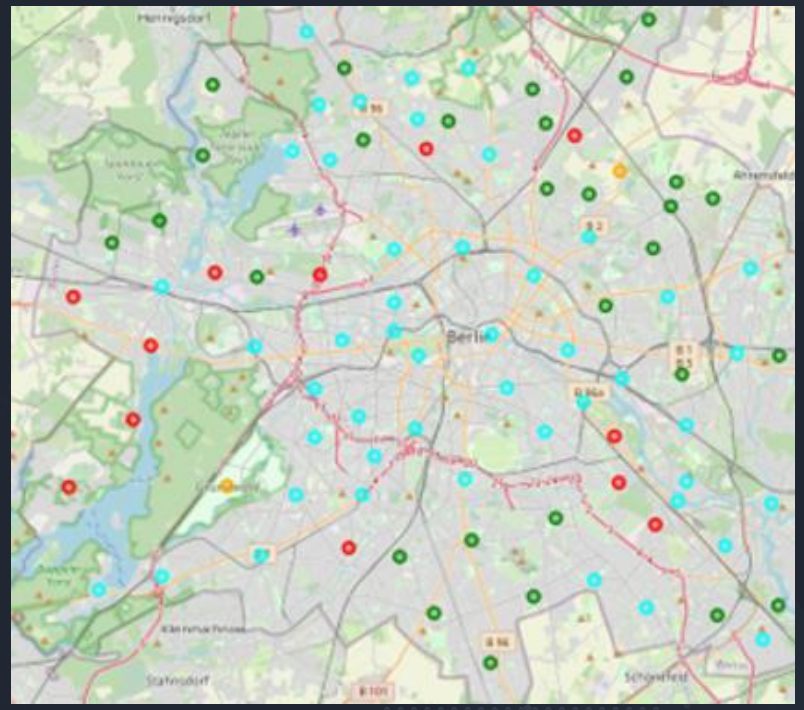
- The chosen neighborhoods of Copenhagen all fell under cluster 3
- Hamurg had the most overalapping neighborhoods with 52 in cluster 3
- Berlin was close by with 50



Results of the analysis on the map







Populations in similar neighborhoods by borough/city

- Berlin overall has over twice as many people in neighborhoods that Similar to Copenhagen
- The most optimal location for an office in Hamburg would be north of the city center
- The most optimal in would West-Central

CITY	POP IN FAVOURABLE CLUSTER
BERLIN	2,267,053
HAMBURG	915,087

HAMBURG		BERLIN	
Borough	Pop	BOROUGH	Pop
Hamburg-Nord	305,311	MITTE	326,474
Altona	175,062	Charlottenburg-Wilmersdorf	283,118
Eimsbüttel	132,980	Friedrichshain-Kreuzberg	261,277
Wandsbek	128,622	Neukölln	231,011
Hamburg-Mitte	73,810	Pankow	216,624

Conclusion

- Overall, there are more people in neighbourhoods of Berlin that are similar to those in Copenhagen, while Hamburg has overall more similar neighbourhoods.
- The best locations(Boroughs) to open the office are Central(Mitte) Berlin, and Northern Berlin.
- Nevertheless, there are plenty of people in either city and the client themselves needs to decide whether to pick Berlin with the potentially harsher competition or Hamburg with potentially less business.
- Given that the client does pick either of the two cities the potentially extended prototype will aid them to a more successful and efficient expansion.