Battle of the Neighbourhoods – Coffee Stops

An analysis into the ideal location for a new automated coffee vendor.

Sam Hyland – 03/08/21

Introduction

Coffee has been big business from the very onset of its discovery and export. All through its lifecycle from production to distribution and retail there are a variety of large multinationals who exclusively trade in the business of coffee. Only in recent decades has the retail aspect of coffee began to outpace the growth of other areas of coffee business. A decline in property ownership particularly among younger generations has led to a shared economy with many amenities being purchased on demand rather than owned. Coffee machines sales have come under pressure by ubiquitous coffee chains at multiple locations within most any urban city worldwide. Starbucks, once a small coffee chain just 50 years ago has now exploded into the foremost coffee chain in the world being the dominant market competitor in both its home market of the US as well as other large markets in China and Japan. The Coffee chain market is still considered quite competitive with Starbucks facing pressure in each of these markets such as Dunkin Donuts and McDonalds (recently pivoting heavily to focus on coffee sales in advertisements) in the US and new competitors such as Luckin and Doutor in China and Japan respectively. This competitive market has encouraged new ways of retailing coffee such as Starbucks $7.1bn deal with Nestlé to market their ready-made coffee in convenience stores rather than purchasing them to order in a Starbucks location.

This deal highlights the potential for new ways of distributing coffee to an ever-growing consumer appetite for quick, affordable, high-quality coffee drinks. It presents an opportunity to the likes of Breville, DeLonghi and Lavazza among others who both sell large quantities of manufactured coffee suitable to be used in one of their own coffee machines. With access to both the technological know-how for creating a large, self-contained coffee vendor as well as the supply arrangements for keeping a large stock of suitable coffee both on site and in reserve for resupplying, these companies should be well equipped to take cater to this demand via an on-demand automated coffee vendor with minimal staffing or premises and requiring only maintenance. The obvious benefit from this approach versus the incumbent competitors comes via the reduced operating costs and overheads from the removal of most wages and leases that brick and mortar retails stores like Starbucks face across their 32,000+ stores worldwide.

This analysis takes the approach of catering to a perceived target persona who would be most sensitive to the possible reduction in cost efficiencies that may be achieved, namely a university student. The analysis will also focus on the capital cities of 4 of the 5 biggest coffee drinking nations per capita in the world (Norway, Finland, Sweden, Denmark) to limit the exploratory costs involved in preparing this analysis. The aim is to find an ideal location to pilot this new method of distributing retail coffee. With the automated coffee vendor not being restricted to pre-established buildings or requiring high construction costs there is little other blockers in place from a financial perspective making the project a relatively inexpensive endeavour. The probable ideal location should be nearby to local universities to increase exposure and footfall with as few as possible local competitors.

Data – Acquisition and Preparation

Data was acquired from a variety of sources with the primary data purpose being that of location data. Principally, the data required will be that of:

- The capital cities and their universities
- The Universities and their nearby Cafés

With the capital cities decided in advance courtesy of their coffee consumption habits, a listing was found containing a list of registered universities in each of the capital cities (https://www.unipage.net/en/universities). The geographical coordinates for the capital cities were found manually via google Maps but these merely served as launch points when observing the map. With university names the PositionStack API (https://positionstack.com) was utilised to try and find a set of geographic coordinates for each university. With coordinates for each university serving as our focal points, FourSquare API (https://foursquare.com) was used to get 10 nearby venues under the café and coffee shop venue categories within 2km of the university.

With regards to data cleaning, due to linguistic differences and incomplete location data, some universities had to be dropped as no set of coordinates could be found against them by PositionStack. There was also a feature within FourSquare that would bring in venues of different category types (such as restaurant or park) if not enough café/coffee shop venues could be found based on the limit criteria set.

As most data was gathered via API geo location data there was little additional features provided by the APIs that would have been of use in this analysis, mitigating the need for any feature selection. Feature scaling however was used once the distance between the venue and its respective university had been found so that the K-means clustering algorithm would run as desired.

Analysis - Methodology

As a starting point, a list of universities would provide us with focal points about which to begin the search for cafés and local venues. The Unipage site provided a considerable list of possible universities – ranging from the main universities with tens of thousands of students to smaller community colleges. The results of the university page can be seen below with Stockholm providing the largest number of universities and Helsinki providing the lowest:

	City	University
0	Copenhagen	13
1	Helsinki	8
2	Oslo	14
3	Stockholm	20
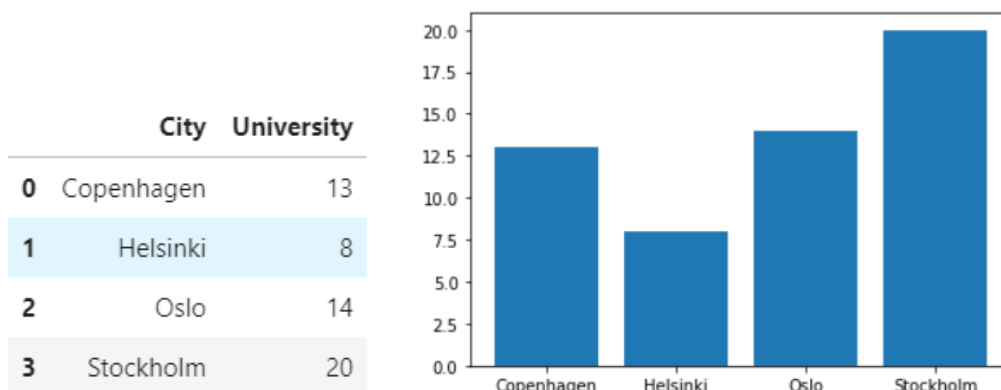
After this initial population was run though the PositionStack API several universities could not be found with a matching set of geographic coordinates and so had to be dropped from the analysis.

This drop-off in university population was most pronounced in the two previously most populous states, Oslo and Stockholm. After the dropping of unfound universities, the final population of universities in use was significantly smaller – losing a total of 26 universities of just under half of its original 55 with Copenhagen now possessing the most universities in scope of the analysis.

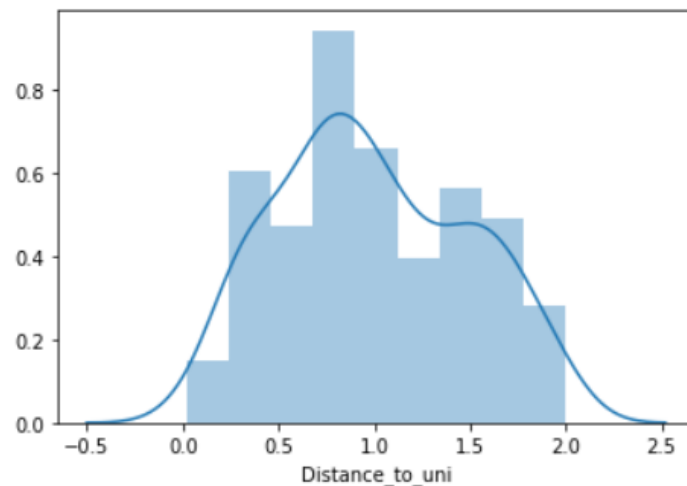|	City	Universities
0	Copenhagen	10
1	Helsinki	5
2	Oslo	5
3	Stockholm	9
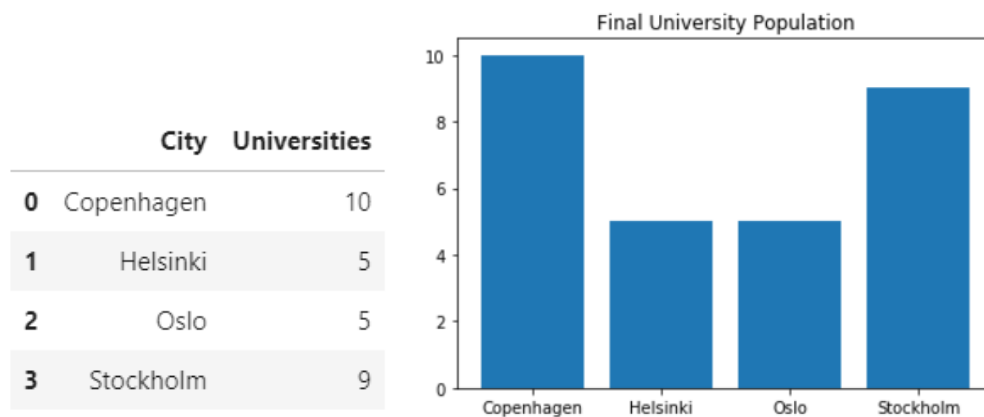
With the universities and the location data gathered, FourSquare then provided the first 10 venue neighbours matching the category of café or coffee shop within 2km of each university. Several non-coffee related venues were pulled in and so these were subsequently dropped so the venues and their locations wholly reflected coffee venues. The drop off was very low losing only 8 venues and going from a total of 252 to 244 split by 239 "Cafés" and 5 "Coffee Shops". This resulted in an average of just over 8 venues per university which was within acceptable bounds to proceed with

As some universities were quite close to each other, particularly in the case of Copenhagen and Oslo, some venues appeared more than once – being within the 2m radius of more than just 1 university. This was acceptable for the analysis as the focus was on the university locations and their proximity to nearby cafés. That some cafés had several nearby universities was noted as a factor and fed into the K-means clustering algorithm as a feature alongside the mean, max and min distance to a university for that particular venue.

The clustering was run against the more populous list of venues rather than that of the universities as we preferred to define the worth of a university location or proximity based on the characteristics of the universities in its vicinity. If a university had many of a particular cluster that was not favourable, then the value of the university as a possible location would be reduced.

The distances from university to venue were equitably distributed with the majority falling more than 250m and just less than 1250m away from their respective university. The relatively large number of venues that were less than 500m away from the university posed a challenge as it would suggest that the café could well fall within range of the extended university campus or indeed in amongst its student accommodation.

Distribution of Distances to University

Also observed was the number of universities within 2km to any venue. This metric was important as any venue within the range of multiple universities was likely to be very competitive for the same target market identified by this report. The below plot shows the observed relationship between the number of universities within 2km of a particular venue.

Venues Mean Distance vs Number of Proximate Universities
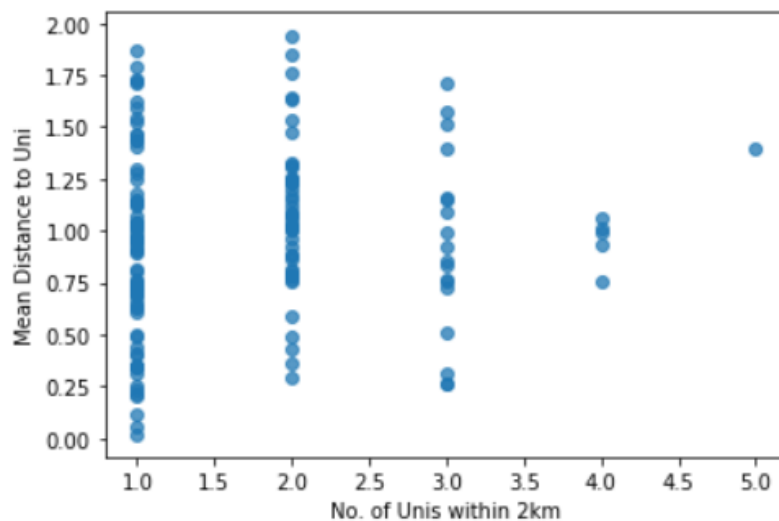
Analysis - K-Means Clustering

With the data gathered and trimmed down by removing non-existent values as well as engaging in feature scaling, a K-means clustering algorithm was run on the data with a given input of 4 clusters. These clusters would inform the colour scheme on the map providing a visual cue as to which university locations were better than others based on our knowledge of favourable vs unfavourable clusters.

The output and characteristics of this algorithm can be seen in the below table. It can clearly be observed that cluster 3 is the most favourable under our conditions possessing the highest "Mean Distance" as well as the greatest "Lowest Distance". These venues were typically in range of just under 2 universities, however they were typically located in the outer belt of the university's 2km radius so quite unlikely to be situated on campus. Cluster 0 can be observed as being the least

favourable cluster, being mainly comprised of venues in very close proximity to their respective university with a very narrow range of distances from 348m to 365m. Any universities containing a larger number of venues from this cluster should be discounted as a possible location for the future coffee vendor.

Cluster Labels	Number of Café Members	Mean Distance	<2km Unis	Lowest Distance	Highest Distance
0	39	0.357292	1.871795	0.348838	0.365746
1	65	0.904139	1.400000	0.885852	0.922426
2	92	1.022107	2.869565	0.605962	1.420544
3	48	1.590452	1.916667	1.574738	1.606166

Results

With the characteristics of the clusters determined the data was then inserted into a map to visually represent the locations under scrutiny and compare them with a "calculated" shortlist of good possible locations determined purely by the clustering output whereby the amount of cluster 3 venues would be maximised and the number of cluster 0 venues would be minimised. This approach produced the below top 5 candidates under this criterion.
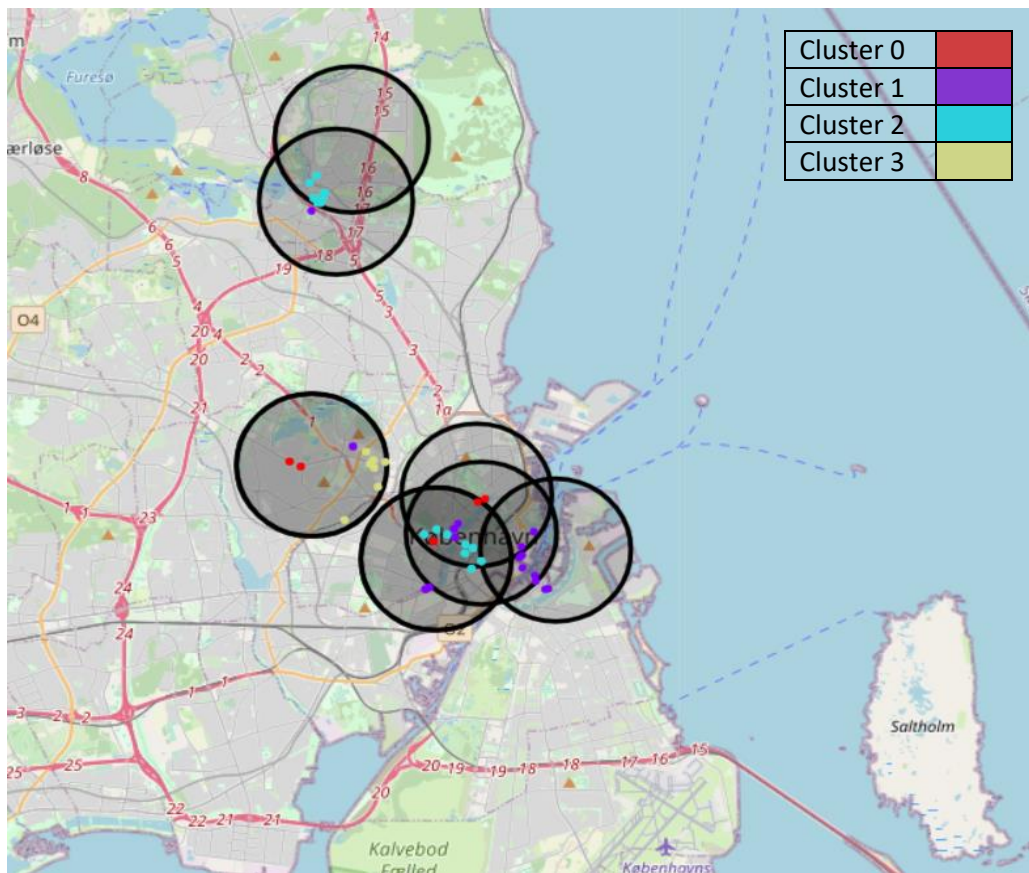
Top 5 Candidates

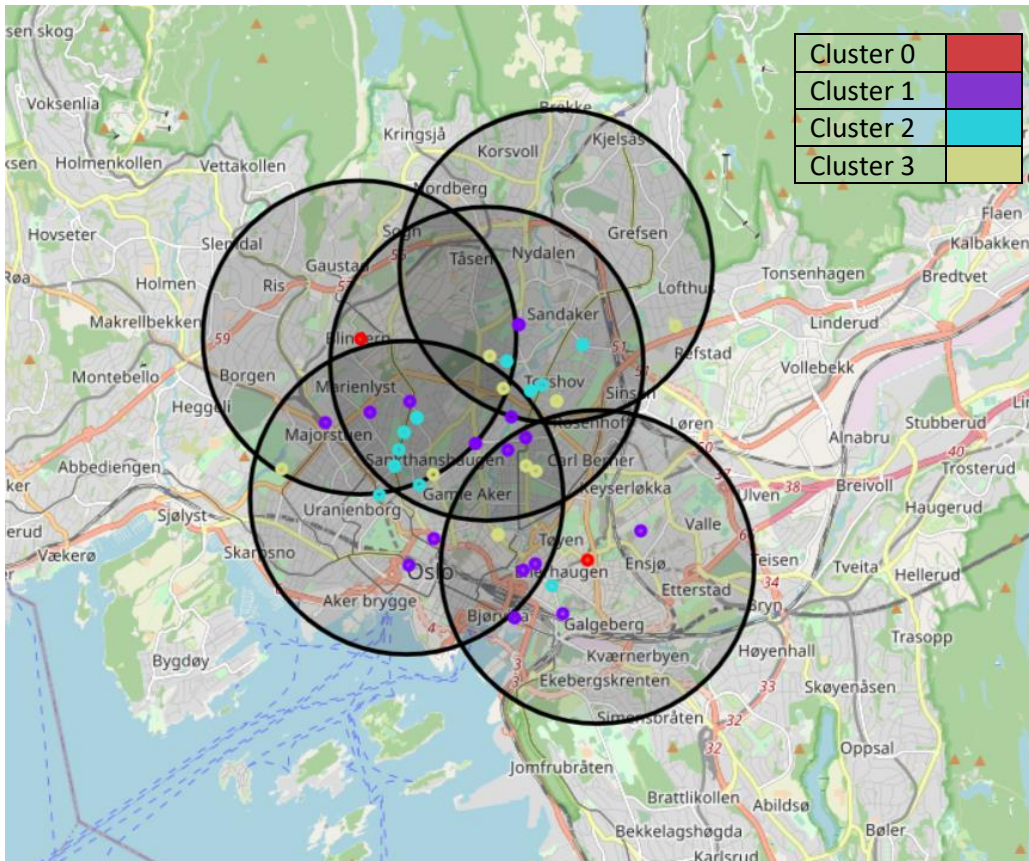| University	City	Cluster Labels_0	Cluster Labels_1	Cluster Labels_2	Cluster Labels_3
BI Norwegian Business School	Oslo	0	1	4	5
Stockholm School of Economics	Stockholm	0	0	1	4
Stockholm School of Theology	Stockholm	0	0	1	4
Stockholm University	Stockholm	0	0	1	4
OsloMet - Oslo Metropolitan University	Oslo	0	2	5	1

Oslo and Stockholm quickly emerge as strong candidate cities highlighted by their low count of cluster 0 venues and relatively high distribution of cluster 3 venues. It can also be observed that Oslo typically has more venues in range of its 2 universities seen above with 10 and 8 respectively whereas the universities within Stockholm each only have 5 venues. This presents an added view to the analysis that can be expanded upon in a further deep dive into the results of this analysis around whether having less venues could be indicative of lower footfall or possibly due to leasing costs or local regulations around building premises for the surrounding areas.

This dimension can be further explored by observing the views seen on the maps of the respective cities to see both the venue and university density within the city and the surrounding areas of the universities to observe whether they sit within a largely urban environment or isolated within a large university campus. The below images depict a view of each of the 4 cities currently under consideration with the black circles representing the 2km radius of a central university while each of the coloured dots represents a coffee venue.
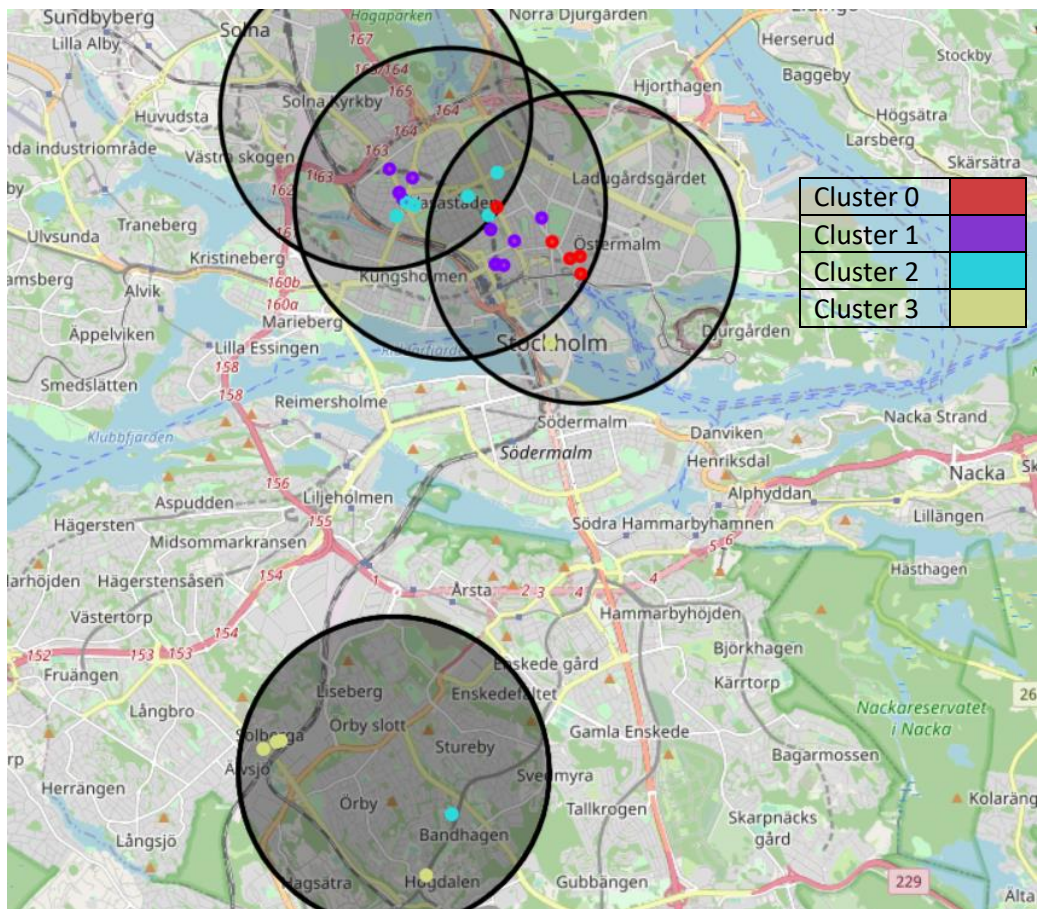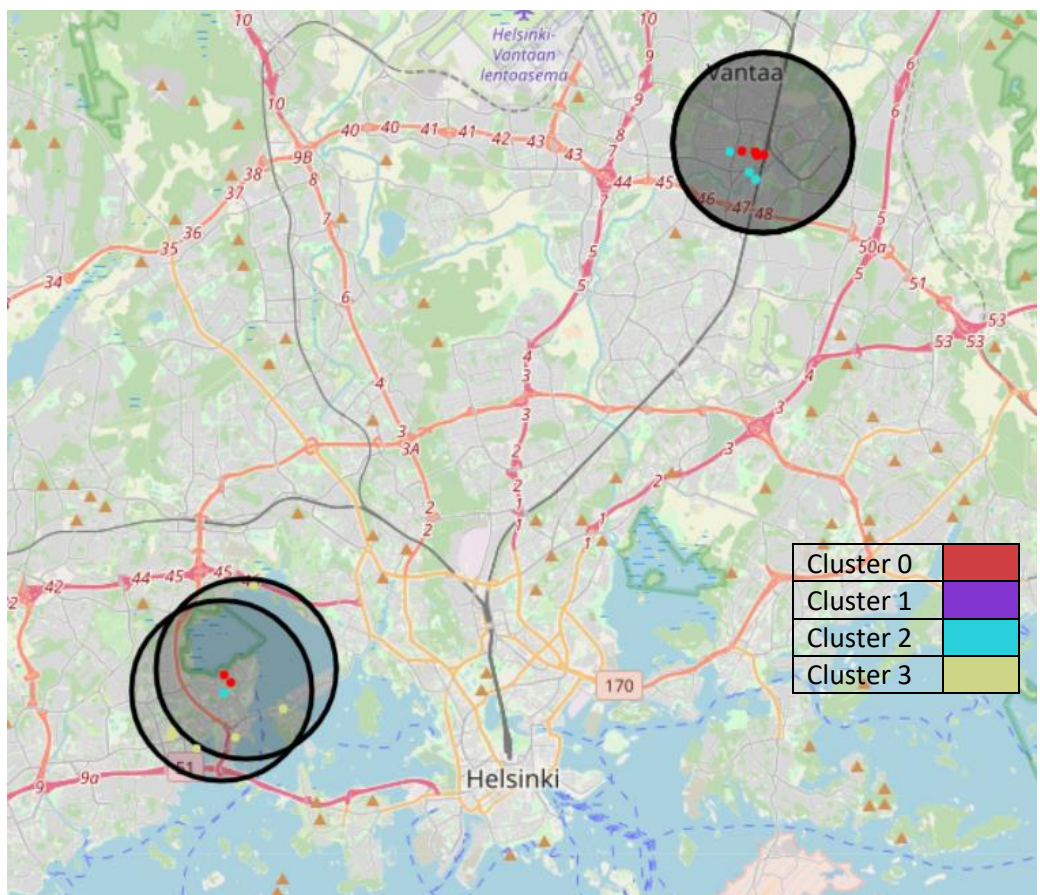
Copenhagen

Oslo

Stockholm

Cluster 0	
Cluster 1	
Cluster 2	
Cluster 3	

Helsinki

Cluster 0	
Cluster 1	
Cluster 2	
Cluster 3	

Discussion

The analysis raised some issues that should be further explored by any subsequent iterations of this analysis seeking to build upon it. The PositionStack API failed to find a number of universities (24 of possible 55) likely due to translation issues so a translation exercise or someone of similar experience should be sought out to further increase the population size. Several universities were also listed either in close proximity (being identified as being different buildings on the same campus) or simply stacked on top of one another with the same geographical coordinates. Part of this issue is unavoidable due to the varied organisational structure that come colleges could avail of but it could be mitigated in future analysis by placing focal points at university campuses with radiuses in proportion with their alumni numbers so as to better reflect the contrasting sizes of some of these universities.

It should also be noted that this analysis was performed without input from a local source of expertise from any of these cities. There may be a variety of factors that influence a person's coffee drinking habits and the nature of the coffee retail in their locale. A quick background check into the 4 cities showed that Starbucks, the industry leader worldwide, has an extremely small presence in these cities. This may simply indicate that the region is a target for expansion for Starbucks or could also signify an inherent aversion to the types of retail coffee that Starbucks specialises in producing or could be the result of industry specific cultural heritage laws that might seek to prohibit large coffee chains from displacing family run cafés. This report can only speak to the locational suitability of areas within these cities excluding these factors.

Conclusions

Upon examining all results, the maps provide additional evidence supporting the idea of using either Oslo or Stockholm as the pilot city as a starting point. Their cities present quite a centralised format with plenty of urban landscape available likely indicative of well-established public transport links which can only support increased footfall in these areas. The red dots provide an easy visual flag of areas to avoid – most pertinent in the disparate city layout of Helsinki with its universities sitting on opposite sides of the city with multiple cluster 0 venues each but also visible to a lesser extent in Copenhagen and Stockholm.

It can be stated that based on the results of this analysis that a possible ideal location for a pilot project of this coffee vendor would be in Oslo, with one such example being in Northern Oslo just West of Nydalen and East of Tåsen lying close to BI Norwegian Business School while also falling just within range of 2 other universities: University of Oslo and The Oslo School of Architecture and Design. This can be observed below by the yellow circled area.

Possible Candidate Location
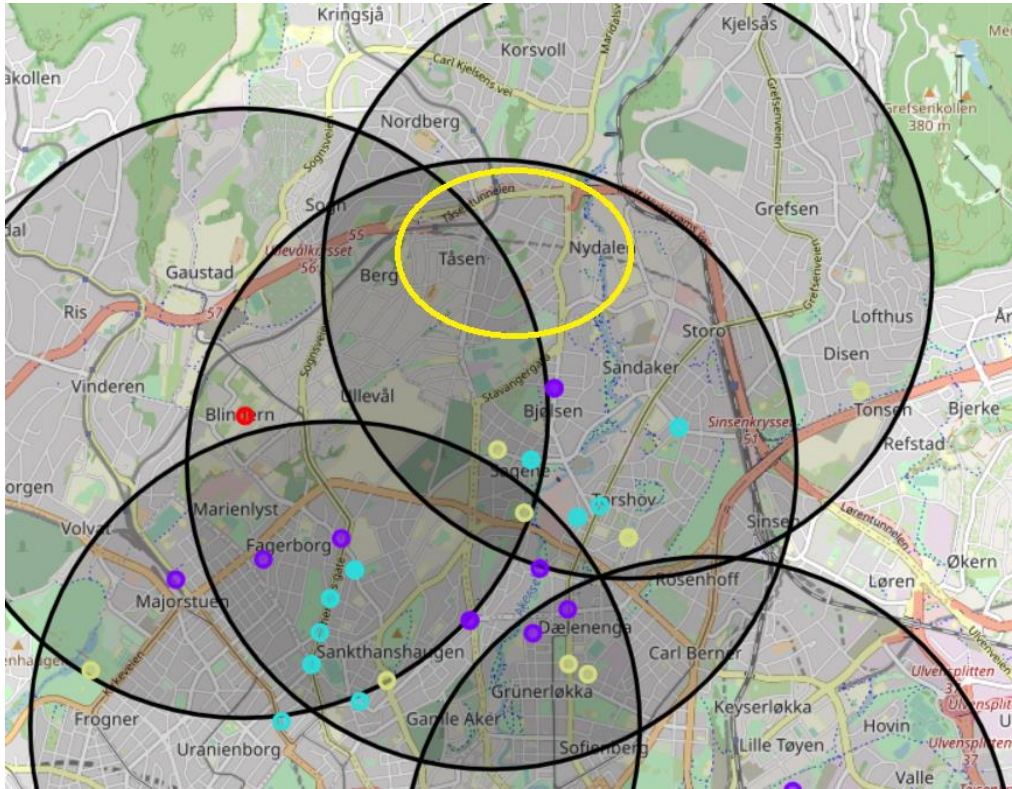
Overall, Oslo provides a number of promising locations as it possesses a number of universities with favourable characteristics and a highly centralised density distribution with multiple universities clustered around one another providing an ample supply of university students. This approach to location determination could also be replicated using alternate parameters such as a different target consumer or demographic and differing focal points to universities such as residential complexes or large size grocery stores.