

Applied Data Science Capstone Project:

Battle of the neighborhoods

Introduction/Brief

Business Problem:

This project will attempt to help a client looking to expand their retail business into Copenhagen. The client is a high-end cafe and roasterie which also bakes and sells their own pastries. They currently have a small but strong business in the US and are looking to expand their presence, possibly internationally.

My goal is to provide them with some insights to help them focus their expansion and understand where in Copenhagen they should scout for locations. I plan to look at which neighborhoods have the highest concentration of coffee shops and cafes and where they might have a particular opportunity. I will look at distance to the city center and any other characteristics that might pop up during the exploration.

Data to be used:

I have compiled some data on Copenhagen from a few websites. The names of the neighborhoods are from google searches and wikipedia. GPS coordinates for the neighborhoods I got from latitude.to. Since I live in Copenhagen I was able to correct or guide the locations a bit to better capture the neighborhoods. Unfortunately, I was not able to find any data to scrape so I had to input all the locations manually.

Venue data was all collected from Foursquare through their API the same way we did in the other exercises.

Background facts about Copenhagen

Copenhagen is an historic Nordic city. It is the capital of Denmark and is a big global attraction for tourists. It is located on the Danish side of the straits between Denmark and Sweden giving it a beautiful waterfront. Together with Malmö in Sweden and a few smaller cities, Copenhagen is part of the business region known as the Øresund region which seeks to foster collaboration between businesses in cities on both sides of the border. As a result, significant investment has been made by the local governments to sponsor infrastructure and cross-border trade.

Copenhagen has a population of about 1.3 million people. It is governed by 10 Municipal districts with the exception of Fredriksberg which is located with Copenhagen's districts. Fredriksberg is technically its own city but is included in our data because it is geographically part of Copenhagen.

Methodology

All of the analysis was done in Python 3 in Jupyter Notebooks.

Libraries used: pandas, numpy, requests, json, geopy.geocoders, matplotlib, sklearn.cluster, folium, geocoder

Starting data set:

Figure 1:

	District	District Population	Neighborhood	Latitude	Longitude
0	Indre By	55493	Christianshavn	55.67358	12.59325
1	Indre By	55493	Kongens Nytorv	55.67837	12.58358
2	Indre By	55493	Nørreport	55.68345	12.56908
3	Indre By	55493	Fredriksstaden	55.68693	12.58702
4	Indre By	55493	Fredriksholm/Nyholm	55.68346	12.60662
5	Vesterbro/Kongens Enghave	67844	Kødbyen/Saxopark	55.67033	12.55717
6	Nørrebro	80572	København N	55.68967	12.55585
7	Nørrebro	80572	Nørrebro	55.69645	12.54561
8	Østerbro	79276	Trianglen	55.69937	12.57750
9	Østerbro	79276	København Øst	55.70934	12.57716
10	Amager Øst	57673	Amagerbro	55.66562	12.60095
11	Amager Øst	57673	Sundbyøster	55.65455	12.61739
12	Amager Vest	71755	Sundby	55.63772	12.58295
13	Valby	56274	Valby	55.66166	12.51678
14	Bispebjerg	55239	Nordvest	55.70865	12.53185
15	Vanløse	40843	Vanløse	55.69117	12.48118
16	Brønshøj-Husum	44741	Brønshøj	55.70387	12.49922
17	Brønshøj-Husum	44741	Husum	55.71015	12.47079
18	Fredriksberg	103960	Fredriksberg C	55.67567	12.54501

After collecting our data, we started with an exploratory analysis. Foursquare returned 1042 venues spread across 19 neighborhoods.

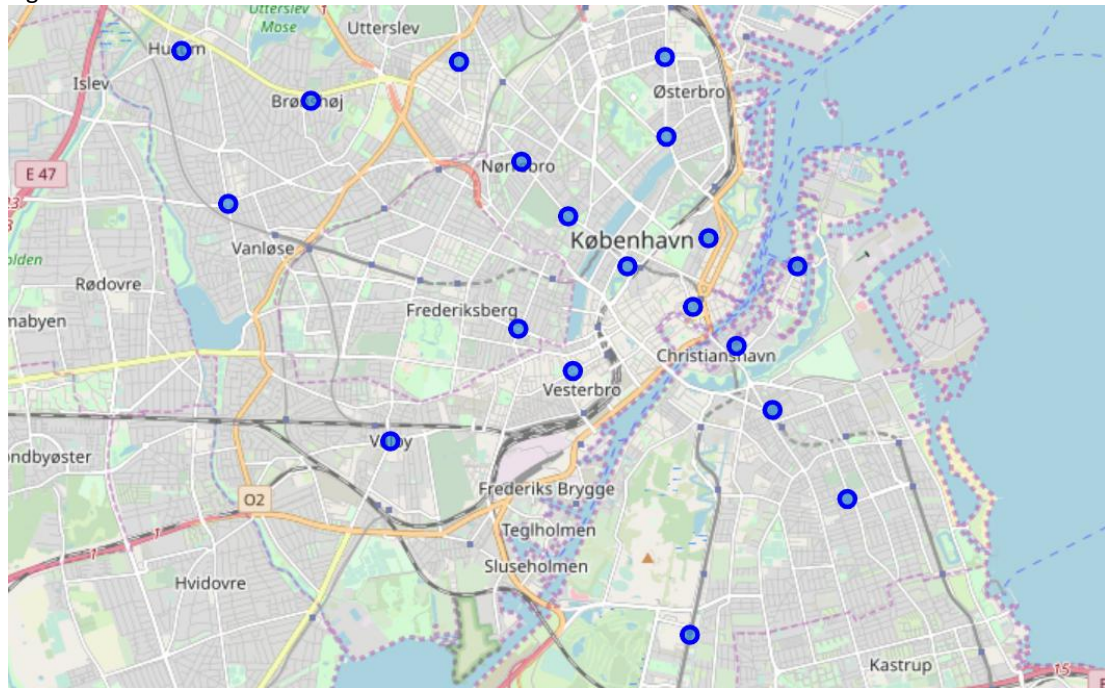
We start by taking a closer look at cafes and coffee shops. We will include bakeries in our coffee related venues. Bakeries could also be considered a competitor. Our client also serves fresh pastries and many bakeries serve coffee as well as provide a cafe experience.

We will then look at number of venues per neighborhood and distance of the venue from the city center.

We will also use kmeans to explore any other relationships there might be in the total data including all venues by creating clusters which tell us some characteristics of some of the neighborhoods

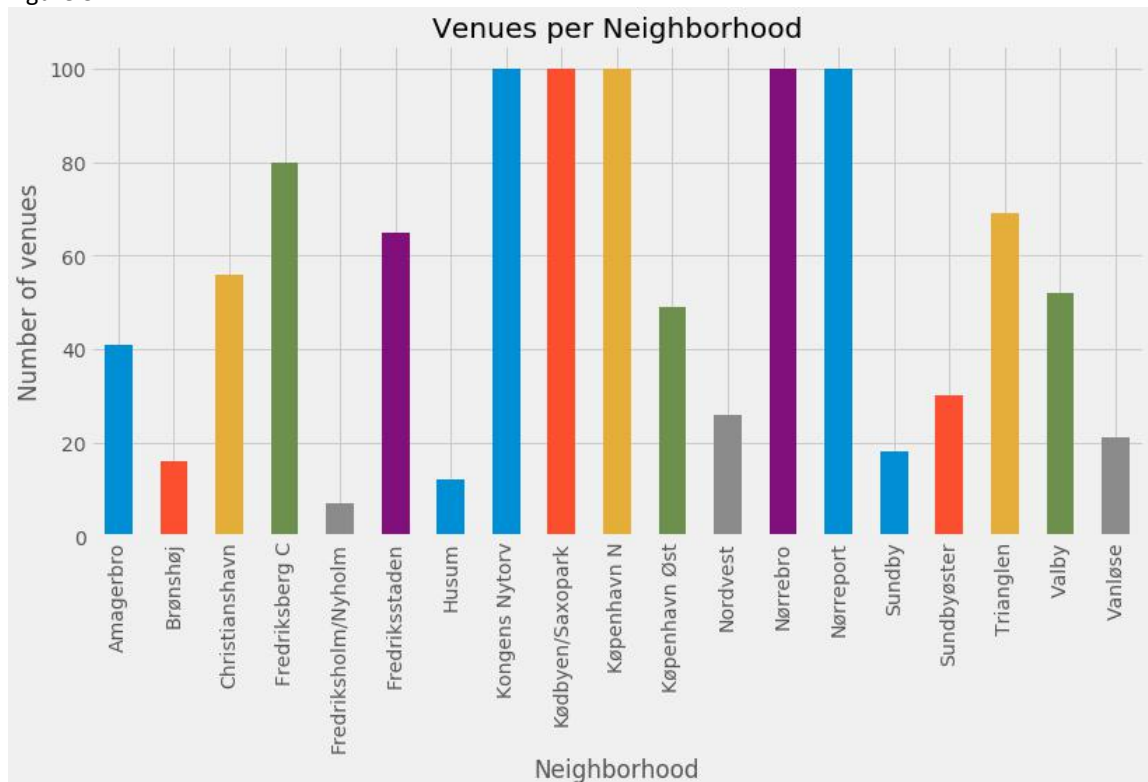
Starting with folium, we generate a quick overview of the city showing our neighborhoods.

Figure 2:



Foursquare returns our venue search showing 1042 venues in 181 categories. The search was done in a 600 meter range from each neighborhood center, while controlling for duplication in case of overlap.

Figure 3:

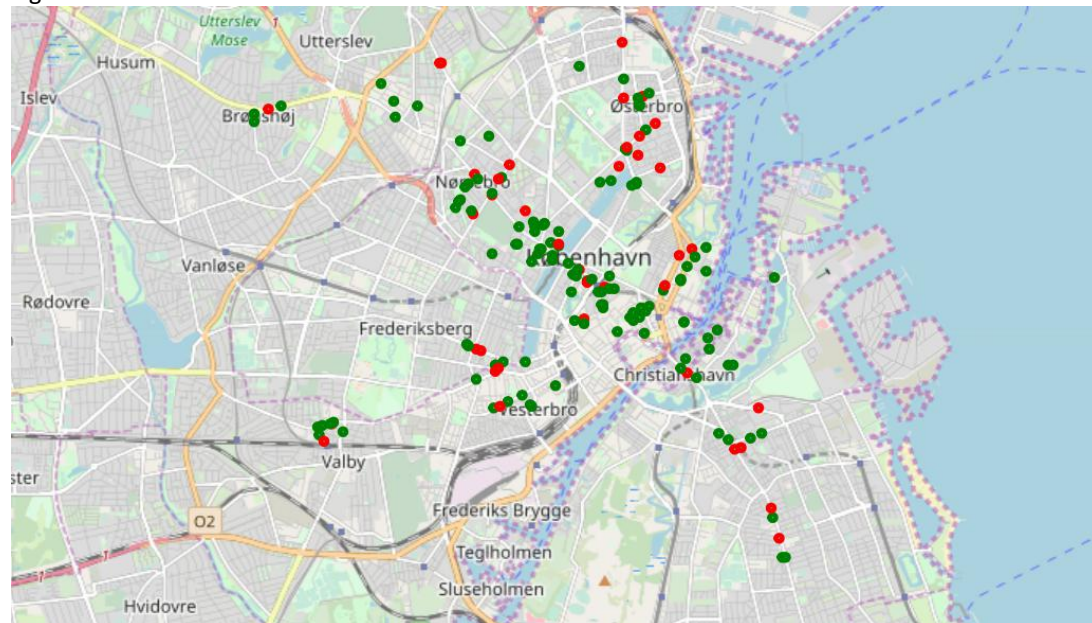


After extracting just the coffee related venues (see figure 4) and calculating the distance to the center we can see the breakdown by neighborhood. Neighborhoods have an average of 10.125 coffee related venues with an average distance of 2.85 km from the center. There are 4 neighborhoods with an above average number of coffee venues: København N, Nørrebro, Nørreport, and Trianglen.

Figure 4:

	Neighborhood	Dist from center	Venue count
0	Amagerbro	1.175714	7
1	Brønshøj	6.772500	4
2	Christianshavn	0.348000	10
3	Fredriksberg C	2.862000	10
4	Fredriksholm/Nyholm	1.590000	1
5	Fredriksstaden	1.327000	10
6	Kongens Nytorv	0.989000	10
7	Kødbyen/Saxopark	2.452222	9
8	København N	2.742857	21
9	København Øst	3.870000	9
10	Nordvest	5.380000	6
11	Nørrebro	3.846875	16
12	Nørreport	1.763182	22
13	Sundbyøster	2.572000	5
14	Trianglen	2.907857	14
15	Valby	5.066250	8

Figure 5:

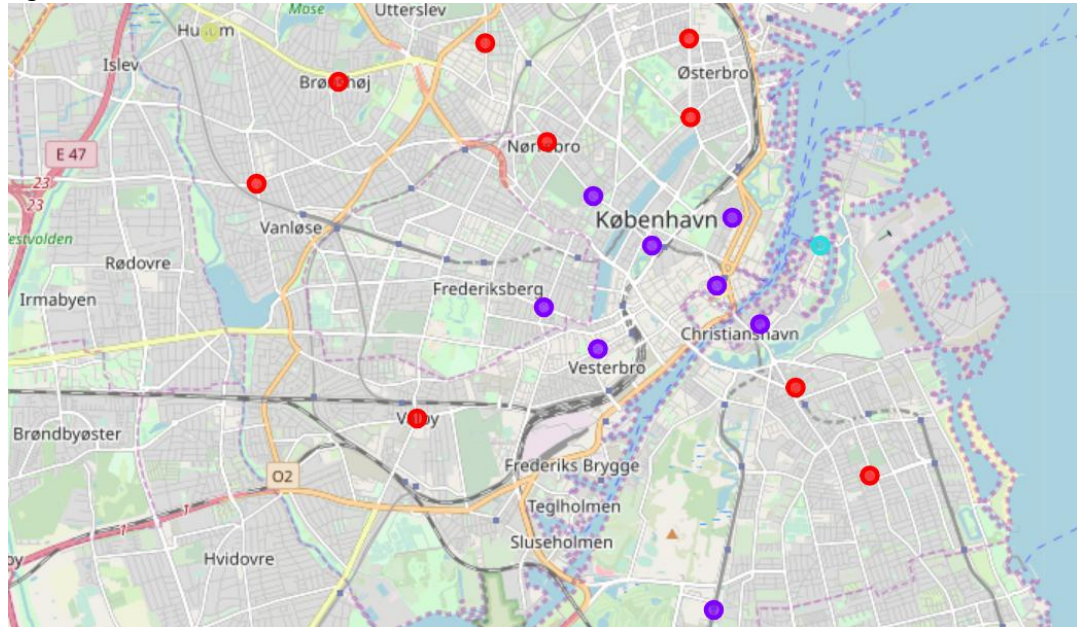


When displayed on a map we see our cafes and coffee shops, shown in green, make up about 75% of

the total (about 12% of all venues). Bakeries shown in red make up 25% of coffee related venues (4% of all venues).

From here we go back to the original data and prepare it for analysis with kmeans to see if we can find any more information about the neighborhoods. Using a K of 4 and running the algorithm 10 times, we get two main clusters seen below.

Figure 6:



Results

From clusters 0 and 1 we can see that the inner neighborhoods around the city center form one cluster and the outer neighborhoods form another. Since we didn't use any location with kmeans we can see there is a difference in the mix of venues in these neighborhoods.

Cluster 0: This cluster surrounds the city center, shown in red on the map above. It is characterized by a very high frequency of cafe, coffee shops, and bakeries and pizza places. In 6 out of 9 of the neighborhoods in this cluster, cafe, coffee shop, or bakery is the first or second most frequent venue. In 4 of the neighborhoods, pizza place is the most frequent venue.

This cluster also has a high frequency and variety of restaurants including Thai, Indian, Italian, Sushi, and Chinese. This makes sense as these neighborhoods are located close to, or in, the suburbs and represent more diverse communities.

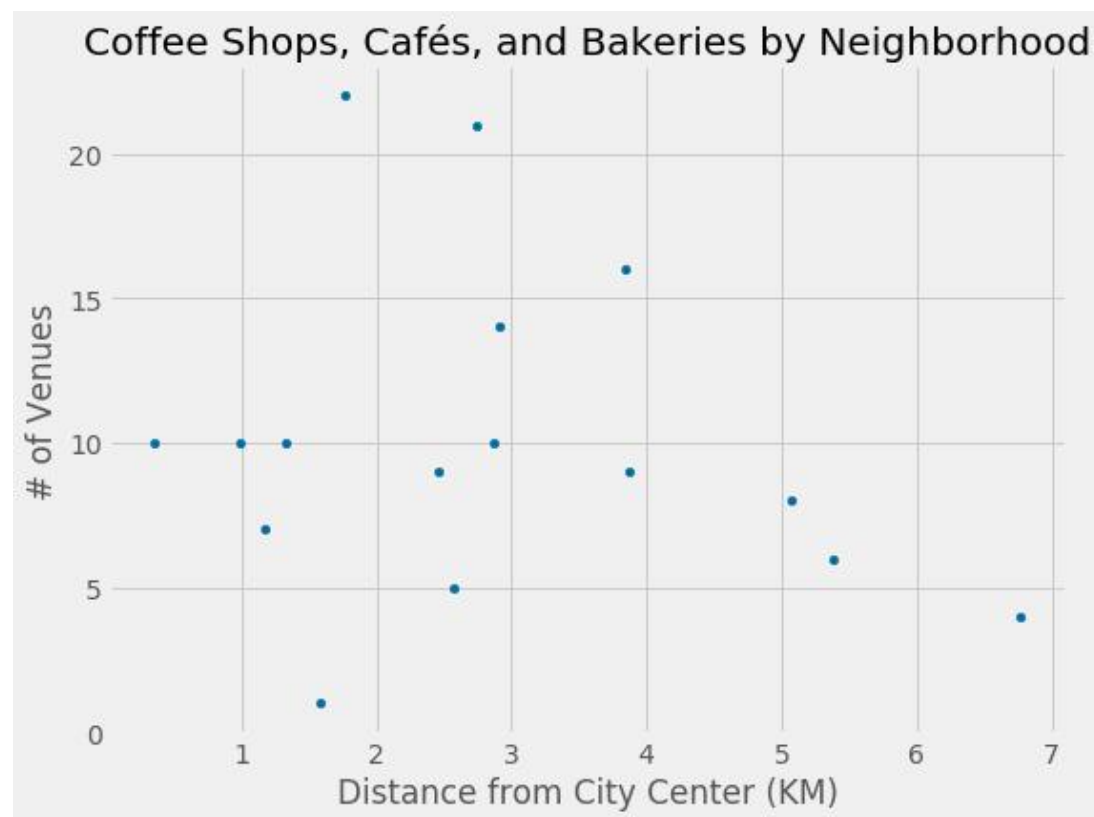
Cluster 1: This cluster also has a high frequency of coffee shops, but not to the extent of cluster 0. Instead we see hotels, hotel bars, and cocktail bars among the top venue. Also interesting is the much higher frequency of Scandinavian restaurants. All but one neighborhood has Scandinavian restaurant in its top 10 venue categories. We also see some more diversity in the venues overall but less diversity in the type of restaurant than we saw in cluster 0. These neighborhoods are near or in the city center which is heavily trafficked by tourists. That would explain, not only the hotels, but the higher frequency of Scandinavian restaurants as tourists would likely prefer to try local cuisine.

Figure 7:

	Neighborhood	Dist from center	Venue count	Cluster Label
0	Amagerbro	1.175714	7	0
1	Brønshøj	6.772500	4	0
4	Fredriksholm/Nyholm	1.590000	1	0
9	København Øst	3.870000	9	0
10	Nordvest	5.380000	6	0
11	Nørrebro	3.846875	16	0
13	Sundbyøster	2.572000	5	0
14	Trianglen	2.907857	14	0
15	Valby	5.066250	8	0
2	Christianshavn	0.348000	10	1
3	Fredriksberg C	2.862000	10	1
5	Fredriksstaden	1.327000	10	1
6	Kongens Nytorv	0.989000	10	1
7	Kødbyen/Saxopark	2.452222	9	1
8	København N	2.742857	21	1
12	Nørreport	1.763182	22	1

Discussion

Figure 8:



From this scatter plot above, we can see that the average number of venues tends to be around 10 venues per neighborhood. There are 4 neighborhoods that stand out with an above the average: Nørrebro, Triangeln, København N, and Nørreport. These 4 neighborhoods sit between about 1.7km and 3.9km from the city center.

One other neighborhood, Kødbyen/Saxopark (also known as the former meatpacking district) has a slightly less than average number of coffee shops, just 9. But it is located within the 1.7-3.9km radius which the most popular neighborhoods are located in. It was one of the 5 city neighborhoods with 100+ venues and was clustered in cluster 1, making it more similar to the downtown neighborhoods.

One final note is Fredriksholm/Nyholm neighborhood, shown on as the lowest # of venues on the scatter plot above, is a bit of an outlier. While the neighborhood is close to the city center it is located on an island with very few shops. It does have tourist attractions such as an opera house and some museums, but it is largely a sight seeing location and not a place with restaurants, cafes, or nightlife. This explains its unusually low number of venues.

Conclusion

København N and Nørreport are densely populated hip areas with a lot of competition (the highest number of coffee related venues). But Nørrebro and Triangeln are still up and coming neighborhoods and may have room for growth. These two neighborhoods would be good places to scout for possible locations and get a feel for their potential.

The meatpacking district is one of the most popular areas in the city, housing a lot of bars, clubs, and restaurants as well as small businesses. This is likely one of the best places to scout for a location as it appears to be overlooked by coffee shops and bakeries.

This study was intended as an exploratory analysis to be followed up by more in-depth analysis of the high-prospect neighborhoods: Nørrebro, Trianglen, and Kødbyen/Saxopark.

Sources:

<https://www.citypopulation.de/php/denmark-admin.php?adm2id=147>

<https://www.citypopulation.de/php/denmark-copenhagen.php>

https://en.wikipedia.org/wiki/Districts_of_Copenhagen