The Battle of Neighborhoods.

Which neighborhoods in the Stockholm area are better suited for relocation given activities of interest and accommodation prices

Introduction / Background

1.1 Background

Although Sweden is quite an unpopulated country, with around 10 million people overall, the country is quite extensive in area. The metropolitan area of Stockholm is the largest metropolitan area in Sweden and can be compared in size with cities like Paris, Barcelona and Milan. This means that there are over 150 neighborhoods where people can establish residence and go around their daily routines. Having lived in Stockholm for more than 5 years in several of these neighborhoods, i have noticed 4 main things that Swedish people, in general, really enjoy doing. The number one is to 'Fika'. Fika is a concept and is immerse within the culture of the country. It means to gather up with friends or colleagues for a coffee or tea and something small to eat, like a bun or a cake. The number two is outdoor activities such as hiking, skiing or bathing in the lakes. Being an archipelago, Stockholm is fully surrounded by nature which is highly appreciated it by locals, especially in the summer. Number three is to drink and dance. Although this is especially true for younger generations, people from all ages enjoy a lot having a drink in their local pubs and are starting to get more attracted to local breweries around Stockholm. The number four is eating out in restaurants around the neighborhood. Whether is a lazy weekend day with no cooking, a socializing event or a curiosity for new gastronomy, locals like to eat out.

1.2 Problem description

Part of the criteria for people in Stockholm to move to a specific neighborhood, apart from accommodation prices, is the type of activities one can do around the neighborhood. Can you grab a cup of coffee nearby with a friend? is there nature around where you could go for a swim in the summer and maybe be a little bit isolated from people? Is it possible to go drinking with colleagues and be able to walk back home? Or is there a neighborhood offering all these possibilities? To answer these questions, one could of course manually investigate each neighborhood, visit them or check online what activities are available around the localities. Given Stockholm's vast area and the big number of neighborhoods, this would be a tedious and time-consuming process. Instead, one could automatically create a sort of profile for each neighborhood, group them, and make this information available and visualizable for people to decide which options are more suitable.This project aims at identifying clusters of similar neighborhoods in Stockholm taking into account the availability to carry on one or more of swedes preferred activities (generally speaking), namely: to Fika, outdoor activities, to drink/party and to eat out in restaurants. Additionally, the report will try to contrast these clusters with the average municipality accommodation prices. By doing so, the goal is to help

people living in Stockholm to identify which neighborhood better match their profiles, and which ones would be a better option for them if they decide to relocate within Stockholm.

1.3 Target Audience

This project would be of interest to real state companies or real state websites looking to help its clients when relocating / searching for new places to live within Stockholm. By having a clear segmentation of neighborhoods based on popular swedish activities and having average price information, the agencies could match a customer profile to filter out best suited neighborhoods

Data Collection

Sweden's territory is divided into Counties. Counties itself are divided into municipalities / Boroughs or what is known in Sweden as 'Komuns'. If one zooms in even further, we find that each komun is divided into several districts or neighborhoods. In this project, data from neighborhoods within Stockholm county will be considered.

Following data sources will be needed to extract/generate the required information:

- Source file(s) containing information on Stockholm Counties and komuns [1]
- Wikipedia for scraping neighborhood names for a given komun [2]
- Google Maps API geocoding for finding the locations of each neighborhood [3]
- Foursquare API for finding venues and venue categories around a given point [4]
- Hemnet website for scraping house prices within komuns [5]

2.1 Stockholm Komuns (Municipalities or Boroughs)

Stockholm has a total of 26 komuns. Since we want to analyze neighborhoods, we need to obtain which are Stockholm's Komuns so we can later obtain its corresponding neighborhoods. Information about Sweden's Counties and Komuns is available to download on the web for free. There are many sources for this information. The repo in reference [1] both County level and Municipality level information within Sweden. This information includes the geographical limits and borders of each County / Municipality

Note: we use the previous source for obtaining existing komun names in Sweden. However, for visualizing border limits between komuns in a map, the referenced topojson was not possible to use. For this reason, we use an additional data source for that. We use a geojson public data set containing the delimiting borders between Komuns in Sweden, which is used for visualizing Avg. house prices per komun [6]. After extracting komun names we get:

```
['Österåker',
 'Botkyrka',
 'Danderyd',
 'Ekerö',
 'Haninge',
 'Huddinge',
 'Järfälla',
 'Lidingö',
 'Nacka',
 'Norrtälje',
 'Nykvarn',
 'Nynäshamn',
 'Södertälje',
 'Salem',
 'Sigtuna',
 'Sollentuna',
 'Solna',
 'Stockholm',
 'Sundbyberg',
 'Täby',
 'Tyresö',
 'Upplands-Bro',
 'Upplands-Väsby',
 'Värmdö',
 'Vallentuna',
 'Vaxholm']
```

2.2 Stockholm Districts (Neighborhoods)

Once we have the name of the Komuns, we need to obtain their corresponding districts. There is no open dataset in the web containing this information. So, the procedure to obtain this data is to scrape the information from wikipedia. If we look at the main wikipedia page for Stockholm County: https://www.en.wikipedia.org/wiki/Stockholm_County we can observe a table with all Komun names, with a link to its corresponding wikipedia page. These individual pages of each Komun are used to obtain the corresponding neighborhoods. Additionally, Google's Geolocation API is used to obtain the center location point of each gathered neighborhood.

Each komun has a corresponding page in wikipedia where the names of the corresponding districts / neighborhoods can be scrapped. The wikipedia pages follows the pattern: https://en.wikipedia.org/wiki/[komun_name]_Municipality

To keep this notebook with relevant information and no so extent, the code for getting all neighborhoods from wikipedia is hidden. A Json file 'stockholm_neighborhoods.json' containing the information and output is generated and stored in the repo. A snapshot of neighborhoods with locations is shown below

	Neighborhood	Municipality	County	Latitude	Longitude
0	Rydbo	Österåker	Stockholm	59.464170	18.182120
1	Skärgårdsstad	Österåker	Stockholm	59.483144	18.407275
2	Svinninge	Österåker	Stockholm	59.454130	18.277120
3	Åkersberga	Österåker	Stockholm	59.480277	18.310783
4	Österskär	Österåker	Stockholm	59.466667	18.300000
5	Fittja	Botkyrka	Stockholm	59.250138	17.864999
6	Alby	Botkyrka	Stockholm	59.240555	17.858093
7	Hallunda	Botkyrka	Stockholm	59.246070	17.823123
8	Norsborg	Botkyrka	Stockholm	59.246070	17.823123
9	Eriksberg	Botkyrka	Stockholm	59.236018	17.838196

2.3 Venue data

The Foursquare API is used to obtain information related to the venues around a given neighborhood.

2.3.1 Explore API for fetching categories

The explore endpoint of the Foursquare API is used to obtain venue categories related to a specified location and query. By doing this, we can obtain all related venue categories to a specific activity such as 'Fika'. Consequently, we can group together all venues that belong to the categories found and have a rough idea of the weight of 'Fika' venues around the neighborhood area. A snapshot of filtered categories will be shown under data exploration.

2.3.2 Explore API for obtaining all venues

The explore endpoint of the Foursquare API is used to obtain all venues around a neighborhood within a certain radio range. Once we have the venues, the categories obtained in the previous step are used to group certain venues and proceed with neighborhood grouping NOTE: Using the free version of Foursquare API there is a limitation of a 100 venues that can be obtained in a given query

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rydbo	59.464170	18.182120	Rydbo (L)	59.465490	18.184010	Light Rail Station
1	Rydbo	59.464170	18.182120	Subway	59.477838	18.164180	Sandwich Place
2	Rydbo	59.464170	18.182120	Kocktorp	59.454735	18.152776	Bus Station
3	Skärgårdsstad	59.483144	18.407275	Skärgårdsstads Hamn	59.480879	18.405921	Harbor / Marina
4	Skärgårdsstad	59.483144	18.407275	Skärgårdsstads VVS	59.487729	18.396879	Construction & Landscaping
...	...	...	...	...	...	...	...
4368	Vaxholm	59.403297	18.326380	Engarn (B)	59.417583	18.311293	Bus Stop
4369	Ytterby	59.428524	18.334887	Ytterbyviks Brygga	59.426177	18.344378	Pier
4370	Ytterby	59.428524	18.334887	Storängsbryggan	59.416897	18.328939	Pier
4371	Ytterby	59.428524	18.334887	Trälhavet	59.432105	18.359594	Lake
4372	Ytterby	59.428524	18.334887	Engarn (B)	59.417583	18.311293	Bus Stop

2.4 House prices

In order to obtain average housing prices for each Komun and display it as reference for the cluster groups, the website www.hemnet.se is used. This is the main real state website used in Sweden for searching apartments to buy. Hemnet provides an API to make queries and requests. However, given that it was a little cumbersome to set it up and there is a time limitation for the project, the data is scrapped instead from URLs. This is not the most elegant approach but is quite fast to set up and obtain the desired results.

Scrap komun average house prices from hemnet

In order to scrap average house prices per komun from hemnet, a komun id is needed to know what to search for. These ids have been added to a dictionary with the komun name as key and the ids as values. We will iterate through the komun names fetched at the beginning of this notebook in the Stockholm Komuns section. A snapshot of avg. price per komun is shown below

```
[20]: {'Österåker': 39884.2962962963,
       'Botkyrka': 31635.722222222223,
       'Danderyd': 71638.87804878049,
       'Ekerö': 39909.307692307695,
       'Haninge': 30223.263157894737,
       'Huddinge': 39981.67567567567,
       'Järfälla': 37048.0,
       'Lidingö': 73215.88461538461,
       'Nacka': 53190.70212765958,
       'Norrtälje': 26666.333333333332,
       'Nykvarn': 37662.11111111111,
       'Nynäshamn': 23523.21052631579,
       'Södertälje': 30704.620689655174,
       'Salem': 31302.85714285714,
       'Sigtuna': 30853.236842105263,
       'Sollentuna': 45110.5,
       'Solna': 54041.3,
       'Stockholm': 63208.186046511626,
       'Sundbyberg': 51913.84,
       'Täby': 46900.84444444445,
       'Tyresö': 42329.45714285714,
       'Upplands-Bro': 31530.3125,
       'Upplands-Väsby': 36355.170731707316,
       'Värmdö': 44191.142857142855,
       'Vallentuna': 41120.95238095238,
       'Vaxholm': 39054.5625}
```

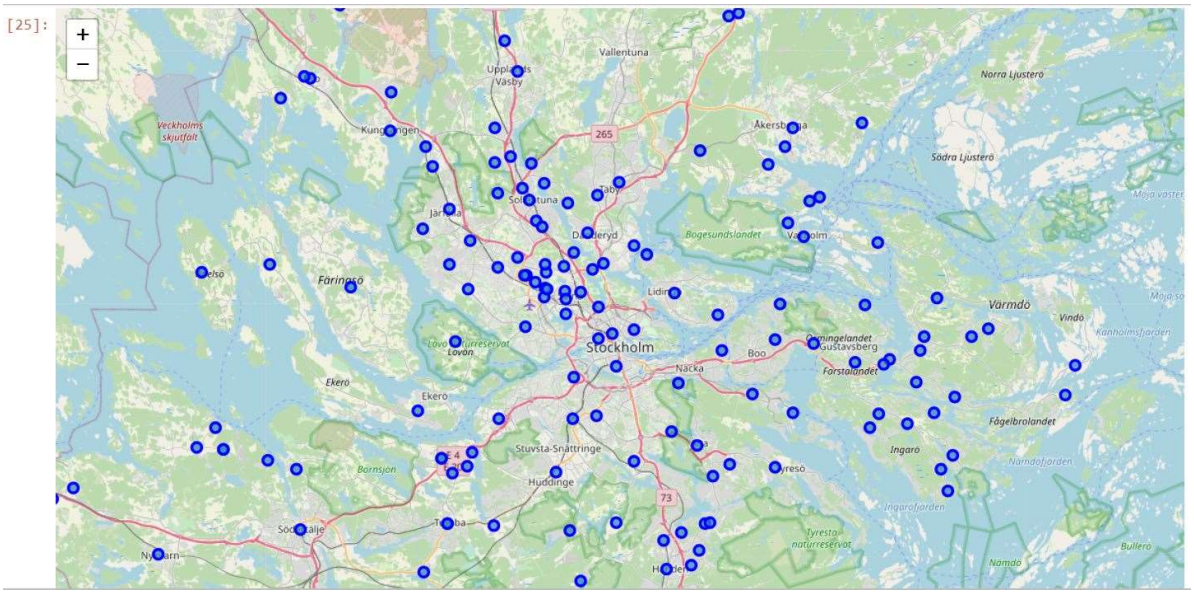
Methodology

In general, the methodology used in this report consist in 3 overall steps: First an exploratory data analysis is carried on, where we visualize the most important aspects of the raw data collected in the previous section. We continue doing data analysis and preparing the data for applying machine learning on it. We lastly perform machine learning clustering on the data to try to draw some conclusions towards our problem statement.

3.1 Exploratory data analysis

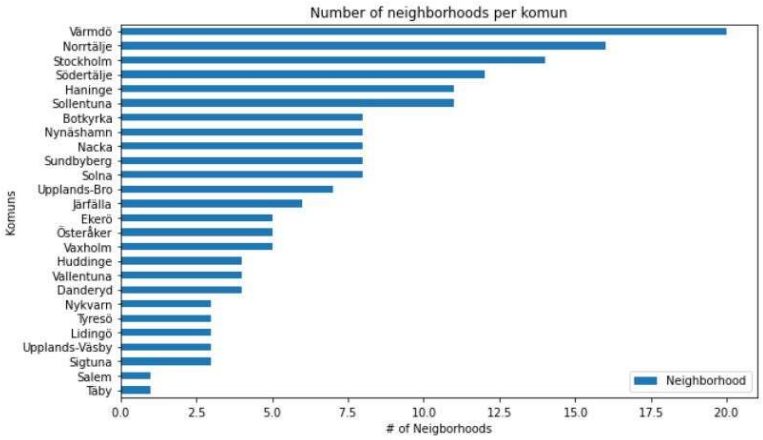
First, we want to visualize how are neighborhoods spread within Stockholm county
Folium library is used for map visualization.
The Nominatum class from geocoder is used to obtain coordinates for Stockholm center point without doing a request to GOOGLE's geolocation API. We can see in the below images how neighborhoods are spread within Stockholm County

[25]:



Let's explore how many neighborhoods we have per komun. This would give us an idea of how big the different municipalities/komuns are.

```
[25]: Text(0.5, 1.0, 'Number of neighborhoods per komun')
```



Explored filtered categories for activities

In Section 2.3.1 'Explore API for fetching categories' venue categories related to all 4 activities (fika,outdoor,drink&dance,restaurants) were gathered.

These categories were manually curated to remove categories that are not so relevant. For example, we got the category 'museum' when searching for Fika. Although you can indeed have fika in some museums, it is not a category that well represent the activity, generally speaking. We define four list of categories, one for each activity. Bellow, a snapshot of the different activities is shown.

Show a snapshot of how many venues were obtained, grouped by category:

[27]:

Venue Category	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Pizza Place	216	216	216	216	216	216
Café	211	211	211	211	211	211
Grocery Store	198	198	198	198	198	198
Supermarket	158	158	158	158	158	158
Scandinavian Restaurant	148	148	148	148	148	148
Hotel	132	132	132	132	132	132
Gym / Fitness Center	113	113	113	113	113	113
Park	92	92	92	92	92	92
Fast Food Restaurant	91	91	91	91	91	91
Bus Station	89	89	89	89	89	89
Shopping Mall	89	89	89	89	89	89
Restaurant	88	88	88	88	88	88
Sushi Restaurant	87	87	87	87	87	87
Coffee Shop	85	85	85	85	85	85
Bus Stop	85	85	85	85	85	85
Asian Restaurant	82	82	82	82	82	82
Burger Joint	82	82	82	82	82	82
Bakery	78	78	78	78	78	78
Department Store	64	64	64	64	64	64
Thai Restaurant	62	62	62	62	62	62

3.2 Data analysis

One hot encoding for venue categories

One hot encoding is performed to convert categorical values into a form better suited for ML algorithms. In this case, each 'venue category' will be a column and a 1 or a 0 will be written to that column for each row in the data frame (depending on whether the category is present in the venue row). An example snapshot of the generated dataframe is shown below

[28]:

	NeighborhoodS	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Amphitheater	Antique Shop	Art Gallery	Art Museum	...	Trail	Train Station	Tram Station	Tunnel	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant	Warehouse	Wine Bar	Women's Store
0	Rydbo	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	Rydbo	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	Rydbo	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	Skärgårdsstad	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	Skärgårdsstad	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

5 rows × 273 columns

If we now visualize the shape of the data frame, we see we have 4373 rows and 273 columns (venues categories)

(4373, 273)

Let's now group the dataframe by neighborhood, using the sum of occurrences for each category. A snapshot is shown below

[32]:

	NeighborhoodS	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Amphitheater	Antique Shop	Art Gallery	Art Museum	...	Trail	Train Station	Tram Station	Tunnel	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant	Warehouse	Wine Bar	Women's Store
0	Adelsö	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	Alby	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	Barkarby	0	0	0	0	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	0
3	Bergshamra	0	0	0	0	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	0
4	Björnömalmen and Klacknäset	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
5	Bollmora	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
6	Boo	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
7	Brandbergen	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
8	Brevik	0	0	0	0	0	0	0	0	0	...	1	0	0	0	0	0	0	0	0	0
9	Bro	0	0	0	0	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	0
10	Bromma	1	1	3	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
11	Brottby	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
12	Brunn	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
13	Brunna	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
14	Central Sundbyberg	0	0	1	0	0	0	0	0	0	...	1	0	0	0	0	0	1	0	0	0

We should now have less rows since we have grouped the dataframe by neighborhoods. Indeed, if we see the shape of the dataframe, there are now 178 rows. The number of columns remain the same, as we have not touched the venue categories

(178, 273)

Group categories by activities of interest

This time we use the extracted categories from section 2.3.1 for grouping all the categories that are related to a specific activity type. A snapshot is shown below.

[33]:

NeighborhoodS	Airport Gate	Airport Service	Airport Terminal	Amphitheater	Antique Shop	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment	...	Train Station	Tram Station	Tunnel	Video Game Store	Warehouse	Women's Store	Fika	Restaurants	Outdoor Places	Drinks/Dance
Adelsö	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	1	0	0
Alby	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	2	7	4	0
Barkarby	0	0	0	0	0	0	0	1	0	...	1	0	0	0	0	0	2	7	0	0
Bergshamra	0	0	0	0	0	0	0	0	0	...	1	0	0	0	0	0	8	10	3	3
Björnömalmen and Klacknäset	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	1	3	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Östermalm	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	1	8	37	9	5
Österskär	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	4	4	1
Östhamra	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	2	2	1	0
Östorp	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	2	5	1	0
Östra Kallfors	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

In the provided snapshot we can see for example new 'Fika' 'Outdoor Places' and 'Drinks/Dance' categories. Also, since some categories were grouped into new categories, the columns should have decreased. If we look at the shape of the dataframe. There are now 188 columns after the grouping of categories

(178, 188)

Sort and store the 10 most common venues per neighborhood in a dataframe
After sorting and selecting the 10 most common venue categories, the dataframe looks as the snapshot below

[37]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelsö	Pizza Place	Restaurants	Dog Run	Fish Market	Field	Farmers Market	Farm	Factory	Exhibit	Event Space
1	Alby	Restaurants	Outdoor Places	Supermarket	Metro Station	Bus Station	Pizza Place	Fika	Shopping Mall	Department Store	Electronics Store
2	Barkarby	Restaurants	Sporting Goods Shop	Clothing Store	Department Store	Pizza Place	Furniture / Home Store	Shopping Mall	Electronics Store	Toy / Game Store	Supermarket
3	Bergshamra	Restaurants	Fika	Hotel	Drinks/Dance	Grocery Store	Outdoor Places	Toy / Game Store	Botanical Garden	Garden	Multiplex
4	Björnömalmen and Klacknäset	Outdoor Places	Restaurants	Duty-free Shop	Flea Market	Fish Market	Field	Farmers Market	Farm	Factory	Exhibit

Construct a new Dataframe containing municipality/komun name and the avg. house price of that municipality/komun.

We will use here the dictionary obtained in the data collection section with the komun house price information. A snapshot of the new dataframe is shown below

[38]:

	Municipality	Value
0	Österåker	41201.107143
1	Botkyrka	31884.805556
2	Danderyd	72044.425000
3	Ekerö	40979.461538
4	Haninge	30046.051282

3.3 Machine learnings

We use unsupervised machine learning for making some sense out of the data. Specifically, we use the K-means clustering algorithm with 5 k clusters. Neighnorhoods are clustered based on venue categories. By doing this, we will identify neigborhoods with similar profiles in terms of which activities can one carry on there

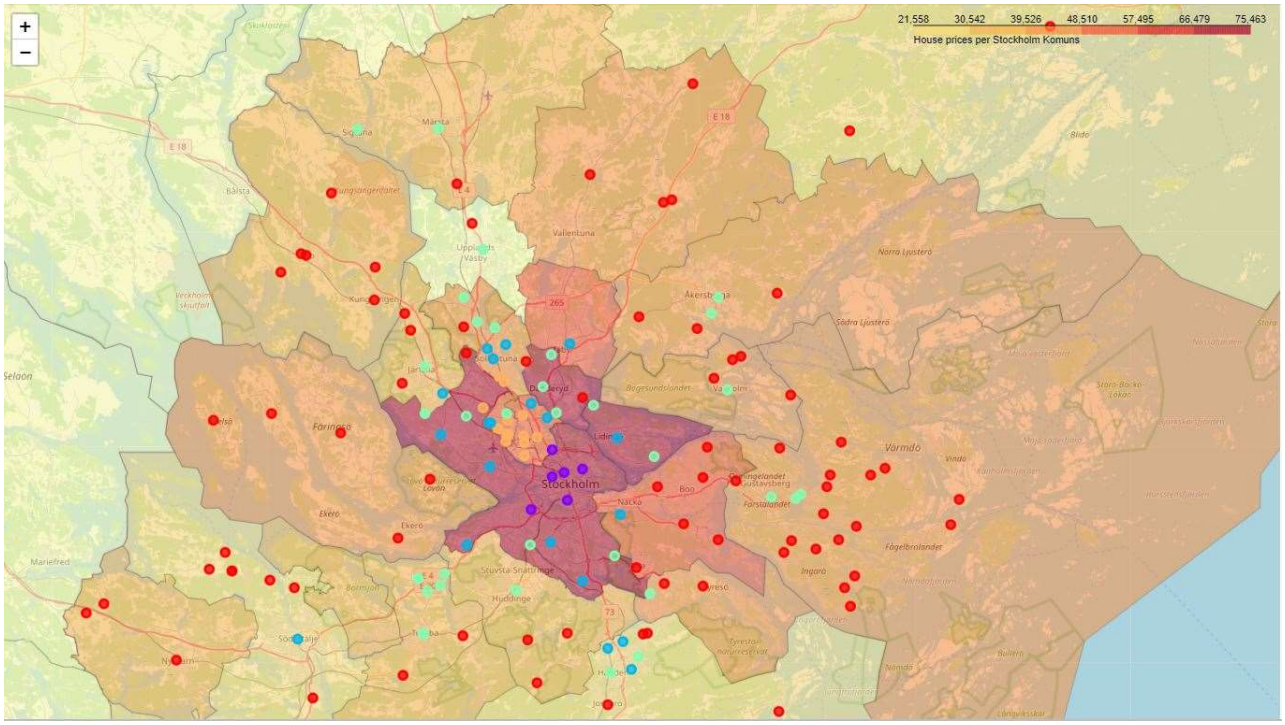
After training the model, we obtain the cluster label for each neighborhood. Each neighborhood should now have one of the 5 possible cluster labels. We add this label back into our sorted neighborhood dataframe. The new dataframe looks like the snapshot below

[41]:

	Neighborhood	Municipality	County	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Rydbo	Österåker	Stockholm	59.464170	18.182120	0	Light Rail Station	Bus Station	Sandwich Place	Drinks/Dance	Flea Market	Field	Farmers Market	Farm	Factory	Exhibit
1	Skärgårdsstad	Österåker	Stockholm	59.483144	18.407275	0	Outdoor Places	Construction & Landscaping	Gym / Fitness Center	Flower Shop	Fish Market	Field	Farmers Market	Farm	Factory	Exhibit
2	Svinninge	Österåker	Stockholm	59.454130	18.277120	0	Outdoor Places	Restaurants	Convention Center	Duty-free Shop	Flea Market	Fish Market	Field	Farmers Market	Farm	Factory
3	Åkersberga	Österåker	Stockholm	59.480277	18.310783	3	Restaurants	Pizza Place	Outdoor Places	Grocery Store	Athletics & Sports	Gym / Fitness Center	Forest	Liquor Store	Department Store	Canal Lock
4	Österskär	Österåker	Stockholm	59.466667	18.300000	3	Restaurants	Outdoor Places	Grocery Store	Pizza Place	Drinks/Dance	Burger Joint	Liquor Store	Department Store	Gym / Fitness Center	Arts & Crafts Store

Results

Let's Visualize and describe the result of performing clustering on all neighborhoods. Using folium, we create a choropleth map using the avg house prices in Stockholm. By doing this, we can visualize how expensive is housing in the municipalities/komuns where the clusters are located



Examine clusters

Cluster 0 [in red]

We can see that in almost every neighborhood, the most common category is 'outdoor activities', which suggests these neighborhoods have many open spaces around with outdoor venues to utilize

	Neighborhood	Municipality	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Rydbo	Österåker	Light Rail Station	Bus Station	Sandwich Place	Drinks/Dance	Flea Market	Field	Farmers Market	Farm	Factory	Exhibit
1	Skärgårdsstad	Österåker	Outdoor Places	Construction & Landscaping	Gym / Fitness Center	Flower Shop	Fish Market	Field	Farmers Market	Farm	Factory	Exhibit
2	Svinninge	Österåker	Outdoor Places	Restaurants	Convention Center	Duty-free Shop	Flea Market	Fish Market	Field	Farmers Market	Farm	Factory
11	Tullinge	Botkyrka	Fika	Grocery Store	Pizza Place	Shopping Mall	Train Station	Drinks/Dance	Discount Store	Farm	Factory	Exhibit
12	Vårsta	Botkyrka	Pizza Place	Grocery Store	Golf Course	Bus Stop	Athletics & Sports	Dog Run	Field	Farmers Market	Farm	Factory
...	...	...	...	...	...	...	...	...	...	...	...	...
175	Lindholmen	Vallentuna	Outdoor Places	Grocery Store	Golf Course	Construction & Landscaping	Fika	Flea Market	Fish Market	Field	Farmers Market	Farm
176	Kullö	Vaxholm	Outdoor Places	Bus Stop	Restaurants	Fika	Grocery Store	Skating Rink	Soccer Field	Drinks/Dance	Duty-free Shop	Field
177	Oskar-Fredriksborg	Vaxholm	Outdoor Places	Historic Site	Boat or Ferry	Lighthouse	Diner	Hotel	Drinks/Dance	Electronics Store	Fish Market	Field
178	Resarö	Vaxholm	Outdoor Places	Grocery Store	Duty-free Shop	Flea Market	Fish Market	Field	Farmers Market	Farm	Factory	Exhibit
180	Ytterby	Vaxholm	Outdoor Places	Bus Stop	Electronics Store	Flea Market	Fish Market	Field	Farmers Market	Farm	Factory	Exhibit

Cluster 1 [in purple]

A key differentiator for this cluster of neighborhoods is the fact that in almost all of them, the four most common venues are venues within the 4 activities being analyzed in this report: restaurants (eat out), Fika, Outdoor places and Drinks/Dance. This suggests that you have many venue options within all these activities in the neighborhoods of this cluster.

	Neighborhood	Municipality	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
109	Haga	Solna	Restaurants	Fika	Outdoor Places	Pizza Place	Drinks/Dance	Gym / Fitness Center	Burger Joint	Breakfast Spot	Sandwich Place	Hotel
116	Kungsholmen	Stockholm	Restaurants	Fika	Outdoor Places	Drinks/Dance	Hotel	Burger Joint	Pizza Place	Clothing Store	Sandwich Place	Salad Place
117	Norrmalm	Stockholm	Restaurants	Fika	Drinks/Dance	Hotel	Burger Joint	Outdoor Places	Gym / Fitness Center	Clothing Store	Pizza Place	Breakfast Spot
118	Södermalm	Stockholm	Restaurants	Fika	Drinks/Dance	Outdoor Places	Bookstore	Burger Joint	Hotel	Dance Studio	Candy Store	Breakfast Spot
119	Östermalm	Stockholm	Restaurants	Outdoor Places	Fika	Hotel	Drinks/Dance	Burger Joint	Museum	Clothing Store	Pizza Place	History Museum
121	Hägersten-Liljeholmen	Stockholm	Restaurants	Fika	Outdoor Places	Drinks/Dance	Pizza Place	Burger Joint	Grocery Store	Supermarket	Liquor Store	Gym / Fitness Center

Cluster 2 [in light blue]

The neighborhoods in this cluster have a clear differentiator. The two most common venues are constantly restaurants and fika venues. Another thing to notice is that there doesn't seem to be many places to drink and dance within venues in these neighborhoods

	Neighborhood	Municipality	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
22	Brandbergen	Haninge	Restaurants	Gym / Fitness Center	Supermarket	Fika	Department Store	Pizza Place	Grocery Store	Bus Station	Shoe Store	Shopping Mall
26	Norrby	Haninge	Restaurants	Fika	Pizza Place	Supermarket	Shopping Mall	Sporting Goods Shop	Grocery Store	Department Store	Furniture / Home Store	Clothing Store
28	Vega	Haninge	Restaurants	Fika	Shopping Mall	Sporting Goods Shop	Supermarket	Pizza Place	Grocery Store	Department Store	Furniture / Home Store	Auto Garage
39	Barkarby	Järfälla	Restaurants	Sporting Goods Shop	Clothing Store	Department Store	Pizza Place	Furniture / Home Store	Shopping Mall	Electronics Store	Toy / Game Store	Supermarket
43	Lidingö	Lidingö	Restaurants	Fika	Grocery Store	Shopping Mall	Hotel	Light Rail Station	Plaza	Pizza Place	Flower Shop	Art Museum
51	Hästhagen	Nacka	Restaurants	Fika	Outdoor Places	Supermarket	Electronics Store	Light Rail Station	Bus Stop	Shopping Mall	Steakhouse	Food Truck
89	Södertälje	Södertälje	Restaurants	Fika	Grocery Store	Supermarket	Outdoor Places	Hotel	Electronics Store	Department Store	Pool	Canal Lock
97	Tureberg	Sollentuna	Restaurants	Fika	Pizza Place	Grocery Store	Bus Station	Stadium	Outdoor Places	Supermarket	Hotel	Liquor Store
100	Edsberg	Sollentuna	Restaurants	Fika	Outdoor Places	Pizza Place	Grocery Store	Shopping Mall	Pharmacy	Bus Station	Gym / Fitness Center	Supermarket
103	Häggvik	Sollentuna	Restaurants	Fika	Grocery Store	Bus Station	Outdoor Places	Pizza Place	Supermarket	Diner	Shopping Mall	Pharmacy
108	Bergshamra	Solna	Restaurants	Fika	Hotel	Drinks/Dance	Grocery Store	Outdoor Places	Toy / Game Store	Botanical Garden	Garden	Multiplex
115	Ulriksdal	Solna	Restaurants	Pizza Place	Outdoor Places	Gym / Fitness Center	Fika	Hotel	Bus Stop	Grocery Store	Plaza	Bus Station

Cluster 3 [in light green]

This cluster is probably the least clear at simple sight. Restaurants seem to be the most common venues within the neighborhoods. Fika venues seem to be present in the top 10 most common venues every time but are more disperse than in other clusters. Some neighborhoods show them in the top 3 and some in the bottom 10. Outdoor places are quite mixed between neighborhoods of this cluster, either being in the top 5 most common venues or not being there at all. Places to drink and dance and outdoor places, are noticeably less common within this cluster.

	Neighborhood	Municipality	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Åkersberga	Österåker	Restaurants	Pizza Place	Outdoor Places	Grocery Store	Athletics & Sports	Gym / Fitness Center	Forest	Liquor Store	Department Store	Canal Lock
4	Österskär	Österåker	Restaurants	Outdoor Places	Grocery Store	Pizza Place	Drinks/Dance	Burger Joint	Liquor Store	Department Store	Gym / Fitness Center	Arts & Crafts Store
5	Fittja	Botkyrka	Restaurants	Metro Station	Pizza Place	Athletics & Sports	Gym Pool	Garden Center	Indoor Play Area	Electronics Store	Performing Arts Venue	Bus Station
6	Alby	Botkyrka	Restaurants	Outdoor Places	Supermarket	Metro Station	Bus Station	Pizza Place	Fika	Shopping Mall	Department Store	Electronics Store
7	Hallunda	Botkyrka	Restaurants	Bus Station	Shopping Mall	Bus Stop	Metro Station	Athletics & Sports	Food	Stadium	Supermarket	Department Store
8	Norsborg	Botkyrka	Restaurants	Bus Station	Shopping Mall	Bus Stop	Metro Station	Athletics & Sports	Food	Stadium	Supermarket	Department Store
9	Eriksberg	Botkyrka	Bus Station	Pizza Place	Supermarket	Metro Station	Shopping Mall	Restaurants	Fika	Outdoor Places	Food	Department Store
10	Tumba	Botkyrka	Restaurants	Bus Stop	Fika	Pizza Place	Train Station	Discount Store	Farmers Market	Farm	Factory	Exhibit
13	Danderyd	Danderyd	Bus Station	Fika	Golf Course	Outdoor Places	Restaurants	Light Rail Station	Gym / Fitness Center	Hotel	Pool	Castle
15	Stocksund	Danderyd	Outdoor Places	Restaurants	Fika	Gym / Fitness Center	Hotel	Drinks/Dance	Movie Theater	Botanical Garden	Breakfast Spot	Bus Stop
16	Enebyberg	Danderyd	Restaurants	Pizza Place	Supermarket	Golf Course	Skating Rink	Grocery Store	Outdoor Places	Light Rail Station	Garden Center	Flower Shop
24	Handen	Haninge	Restaurants	Fika	Gym / Fitness Center	Department Store	Shopping Mall	Outdoor Places	Supermarket	Bookstore	Shoe Store	Bus Station
30	Vendelsömalm	Haninge	Restaurants	Supermarket	Pizza Place	Shoe Store	Department Store	Furniture / Home Store	Liquor Store	Burger Joint	Flower Shop	Bus Stop
34	Östorp	Huddinge	Restaurants	Pizza Place	Supermarket	Fika	Gym / Fitness Center	Outdoor Places	Liquor Store	Farm	Electronics Store	Plaza

Cluster 4 [in yellow]

This cluster clearly has Restaurants and fika venues as the most common types of venues within the neighborhoods. One can notice that Outdoor places and Drink/Dance places are almost always present in the top 10 most common venues. This suggests that within all 4 activities we can find several venue options in the neighborhoods of this cluster. A key difference between this cluster and cluster 1 is that in cluster 1, venues within the 4 activities described in the report were almost all the time in the top 4 most common venues. In cluster 4, Outdoor places and Drink/Dance places are, although always present in the top 10, not always in the top 5 most common venues. This suggest that although you can drink and perform outdoor activities within these neighborhoods, there are not as many venues of this types as in cluster 1

	Neighborhood	Municipality	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
99	Helenelund	Sollentuna	Restaurants	Fika	Gym / Fitness Center	Hotel	Grocery Store	Pizza Place	Shopping Mall	Steakhouse	Movie Theater	Outdoor Places
107	Rest	Sollentuna	Restaurants	Fika	Hotel	Gym / Fitness Center	Pizza Place	Outdoor Places	Grocery Store	Sandwich Place	Deli / Bodega	Electronics Store
110	Hagalund	Solna	Restaurants	Fika	Hotel	Outdoor Places	Drinks/Dance	Supermarket	Toy / Game Store	Botanical Garden	Department Store	Shopping Mall
111	Huvudsta	Solna	Restaurants	Fika	Outdoor Places	Drinks/Dance	Supermarket	Pizza Place	Hotel	Gym / Fitness Center	Grocery Store	Burger Joint
112	Järva	Solna	Restaurants	Fika	Outdoor Places	Bus Stop	Pizza Place	Drinks/Dance	Toy / Game Store	Electronics Store	Hotel	Supermarket
113	Råsunda	Solna	Restaurants	Fika	Pizza Place	Drinks/Dance	Gym / Fitness Center	Hotel	Burger Joint	Grocery Store	Outdoor Places	Liquor Store
114	Skytteholm	Solna	Restaurants	Fika	Drinks/Dance	Pizza Place	Gym / Fitness Center	Burger Joint	Hotel	Grocery Store	Department Store	Liquor Store
127	Rinkeby-Kista	Stockholm	Restaurants	Fika	Grocery Store	Gym / Fitness Center	Pizza Place	Hotel	Salad Place	Plaza	Metro Station	Noodle House
130	Central Sundbyberg	Sundbyberg	Restaurants	Fika	Grocery Store	Outdoor Places	Pizza Place	Burger Joint	Gym / Fitness Center	Supermarket	Liquor Store	Steakhouse
131	Duvbo	Sundbyberg	Restaurants	Fika	Pizza Place	Grocery Store	Outdoor Places	Supermarket	Gym / Fitness Center	Burger Joint	Liquor Store	Hotel
132	Hallonbergen	Sundbyberg	Restaurants	Fika	Outdoor Places	Pizza Place	Grocery Store	Gym / Fitness Center	Drinks/Dance	Burger Joint	Bistro	Metro Station
133	Lilla Alby	Sundbyberg	Restaurants	Fika	Pizza Place	Supermarket	Drinks/Dance	Burger Joint	Grocery Store	Hotel	Outdoor Places	Liquor Store
135	Storskogen	Sundbyberg	Restaurants	Fika	Grocery Store	Outdoor Places	Pizza Place	Burger Joint	Supermarket	Drinks/Dance	Gym / Fitness Center	Liquor Store

Discussion

Based on the results visualized in the previous sections, we can draw several observations.

We can clearly see that most of the neighborhoods in the outskirts of the Stockholm area belong to cluster 0, where the most common activity is outdoor activities. This seems logical since there is a lot of nature and water around those areas. Also, the average price for accommodation in those areas is low. Agencies could recommend these neighborhoods for relocation if someone likes outdoor activities and nature and has a limited monthly income.

If a person is really interested in relocating to a neighborhood where there are many options available for restaurants and fika, but do not have a very high income, agencies could recommend neighborhoods belonging to cluster 2. Most of them are around areas with medium - low average house prices. One thing to notice here is that drinking - dancing places are not very common around these neighborhoods, so agencies could filter them out when dealing with younger clients looking for a more active neighborhood.

There are persons that are interested to have a little of everything around, meaning that they would like to have restaurants to eat out, places to fika, outdoor activities and availability to go drinking / Dancing. There is a clear cluster of neighborhoods (cluster 1) where venues for these activities are almost always the four most common overall. This seems ideal except for the fact that average house prices in these neighborhoods are super high. For a person with an average income, it could be unfeasible to relocate to these areas. If the client's income is super high and is interested in having venues around for all types of activities, the neighborhoods within cluster 1 are ideal.

There is an alternative, however, for people with interest in all four activities and an average income. Neighborhoods within cluster 4 have a lower house average price but they still fulfill the requirement of having venues around for all four types of activities. The difference here is that when it comes to Drink/Dance and outdoor activities, the related venues are not always in the top 5 most common venues, which mean that although there are some present around, they are not as common as in cluster 1. This in turn means that you have more limited options and might have to walk farther distances, for example, to reach a venue where you can drink or dance. People might be willing to accept that trade off, however, given the reduced housing prices in these areas.

Conclusion

The goal of this project was to try to identify similar neighborhoods based on the possibility to eat out, fika, drink/dance and carry out outdoor activities in outdoor venues. Based on this, together with average house prices in the neighborhoods, the project aimed to serve as a tool for agencies to recommend relocation possibilities. Using unsupervised machine learning with k-means clustering, we were able to differentiate and group together neighborhoods offering one, two or all four activities together, with both medium and high average housing prices.

References

[1] Sweden's Counties and komuns

[2] Sweden's Neighborhoods per komun

[3] Google Maps API geocoding

[4] Foursquare API

[5] Hemnet for Sweden's house prices

[6] Sweden's komuns geojson spatial data