# Exploring Stockholm, Sweden: The battle of neighborhoods!


## Ekta Vats


## November 7, 2020

## Table of contents:



## Introduction: Business Problem

Stockholm is the capital and most populous city of Sweden, and the city stretches across 14 islands where Lake Mälaren flows into the Baltic Sea. Stockholm is also the cultural, media, political, and economic centre of Sweden. Stockholm is a vibrant and exciting place, with rich cultural diversity and friendly people all around.

Want to explore the Stockholm city? Let's do some fun with data science!

In this project, we will explore the city of Stockholm, and find out the most popular places (such as museums, parks, cafe, restaurants, etc.), to help guide the tourists, investors and general public interested in visiting Stockholm. We would also prefer locations that are close to the city centre.

We will use the data science skills acquired as part of the IBM Data Science Professional Certification (by Coursera), and generate a few most popular neighbourhoods with information about interesting venues to visit. The details about each venue will then be suggested for the best possible exploration of the Stockholm city.

## Data

Based on the problem definition, following factors will contribute in deciding which neighborhood is popular:

- number of museums and attractions in the neighborhood.
- number of cafe and restaurants in the neighborhood.
- whether the neighborhood is close to the city centre.
- population density.

We will focus around the Stockholm city centre to define our neighborhoods.

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

- **Wikipedia**, (https://en.wikipedia.org/wiki/Stockholm_City_Centre, to obtain the information related to the districts/boroughs in the Stockholm city centre.
- **Foursquare API** to obtain the location/venue data (i.e. number of museums, attractions, cafe, restaurants, etc.)
- **BeautifulSoup** to parse the area information from a HTML table.
- **Geocoder** to obtain the coordinates of the Stockholm city center.
- **Folium** to visualize the maps.

### Neighborhood Candidates

So let's get started!

Let's first install and import the required libraries, and then proceed with finding the latitude and longitude coordinates for the centroids of our candidate neighborhoods around the Stockholm city centre.

Following table represents the data extracted from the Wikipedia page of the Stockholm City Centre:

```
In [5]: print(df.shape)
        df
```

```
(39, 6)
```

Out[5]:

| | District | Area | Population | Density | Borough | Province |
|---|---|---|---|---|---|---|
| 0 | Djurgården | 290 | 788 | 3 | Östermalm | Upland |
| 1 | Fredhäll | 38 | 4,958 | 130 | Kungsholmen | Upland |
| 2 | Gustav Vasa | 80 | 12,911 | 161 | Norrmalm | Upland |
| 3 | Gärdet | 459 | 18,158 | 40 | Östermalm | Upland |
| 4 | Hedvig Eleonora | 58 | 10,387 | 179 | Östermalm | Upland |
| 5 | Hjorthagen-Värtahamnen | 174 | 2,225 | 13 | Östermalm | Upland |
| 6 | Jakobi | 92 | 201 | 3 | Norrmalm | Upland |
| 7 | Klara | 71 | 1,597 | 22 | Norrmalm | Upland |
| 8 | Kristineberg | 63 | 5,572 | 88 | Kungsholmen | Upland |
| 9 | Kungsholm | 107 | 18,465 | 173 | Kungsholmen | Upland |
| 10 | Lilla Essingen | 23 | 4,519 | 196 | Kungsholmen | Upland |
| 11 | Marieberg | 62 | 14,099 | 227 | Maria-Gamla stan | Södermanland |
| 12 | Mariaberg | 66 | 2,700 | 41 | Kungsholmen | Upland |
| 13 | Mellersta Högalid | 54 | 9,914 | 184 | Maria-Gamla stan | Södermanland |
| 14 | Norra Adolf Fredrik | 19 | 3,816 | 201 | Norrmalm | Upland |
| 15 | Norra Johannes | 66 | 9,543 | 137 | Norrmalm | Upland |
| 16 | Norra Högalid | 65 | 13,166 | 203 | Maria-Gamla stan | Södermanland |
| 17 | Norra Södra | 58 | 7,721 | 133 | Katarina-Södra | Södermanland |
| 18 | Oscars Kyrka | 116 | 15,271 | 132 | Östermalm | Upland |
| 19 | Reimarholme-Långholmen | 52 | 2,349 | 45 | Maria-Gamla stan | Södermanland |
| 20 | Roslagstull | 9 | 2,713 | 301 | Norrmalm | Upland |
| 21 | Stadshagen | 63 | 2,258 | 36 | Kungsholmen | Upland |
| 22 | Stora Essingen | 72 | 3,954 | 55 | Kungsholmen | Upland |
| 23 | Stortorget | 42 | 3,017 | 72 | Maria-Gamla stan | Södermanland |
| 24 | Stureplan-Lärkstaden | 71 | 8,104 | 114 | Östermalm | Upland |
| 25 | Södra Adolf Fredrik | 35 | 3,703 | 106 | Norrmalm | Upland |
| 26 | Södra Hammarbyhamnen | 125 | 10,815 | 85 | Katarina-Södra | Södermanland |
| 27 | Södra Högalid | 72 | 4,155 | 58 | Katarina-Södra | Södermanland |
| 28 | Södra Johannes | 15 | 2,011 | 134 | Norrmalm | Upland |
| 29 | Södra Sofia | 74 | 11,015 | 148 | Katarina-Södra | Södermanland |
| 30 | Södra Station | 18 | 4,844 | 269 | Katarina-Södra | Södermanland |
| 31 | Tavricka Högskolan | 156 | 3,442 | 22 | Östermalm | Upland |
| 32 | Universitetet | 476 | 3,131 | 7 | Östermalm | Upland |
| 33 | Västra Katarina | 85 | 13,220 | 156 | Katarina-Södra | Södermanland |
| 34 | Västra Mathaus | 94 | 14,272 | 170 | Kungsholmen | Upland |
| 35 | Östra Katarina | 85 | 19,855 | 234 | Katarina-Södra | Södermanland |
| 36 | Östra Mathaus | 51 | 12,325 | 242 | Kungsholmen | Upland |
| 37 | Östra Sankt Göran | 93 | 14,579 | 156 | Kungsholmen | Upland |
| 38 | Total | 3577 | 296,323 | 83 | | |

Furthermore, using the geocoder, we obtained the latitude and longitude coordinates of the districts in the Stockholm city centre. These are represented as follows:

```
In [10]: df.head(15)
```

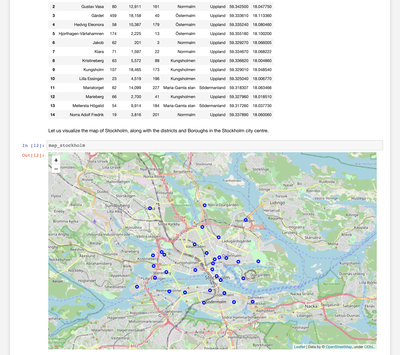
Out[10]:

| | District | Area | Population | Density | Borough | Province | Latitude | Longitude |
|---|---|---|---|---|---|---|---|---|
| 0 | Djurgården | 290 | 788 | 3 | Östermalm | Upland | 59.324820 | 18.097800 |
| 1 | Fredhäll | 38 | 4,958 | 130 | Kungsholmen | Upland | 59.331330 | 18.005450 |
| 2 | Gustav Vasa | 80 | 12,911 | 161 | Norrmalm | Upland | 59.342500 | 18.047750 |
| 3 | Gärdet | 459 | 18,158 | 40 | Östermalm | Upland | 59.333610 | 18.113560 |
| 4 | Hedvig Eleonora | 58 | 10,387 | 179 | Östermalm | Upland | 59.335240 | 18.080480 |
| 5 | Hjorthagen-Värtahamnen | 174 | 2,225 | 13 | Östermalm | Upland | 59.350180 | 18.102200 |
| 6 | Jakobi | 92 | 201 | 3 | Norrmalm | Upland | 59.329270 | 18.066500 |
| 7 | Klara | 71 | 1,597 | 22 | Norrmalm | Upland | 59.334070 | 18.066222 |
| 8 | Kristineberg | 63 | 5,572 | 88 | Kungsholmen | Upland | 59.336620 | 18.004860 |
| 9 | Kungsholm | 107 | 18,465 | 173 | Kungsholmen | Upland | 59.329010 | 18.038540 |
| 10 | Lilla Essingen | 23 | 4,519 | 196 | Kungsholmen | Upland | 59.325040 | 18.006770 |
| 11 | Marieberg | 62 | 14,099 | 227 | Maria-Gamla stan | Södermanland | 59.318307 | 18.063466 |
| 12 | Mariaberg | 66 | 2,700 | 41 | Kungsholmen | Upland | 59.329010 | 18.018110 |
| 13 | Mellersta Högalid | 54 | 9,914 | 184 | Maria-Gamla stan | Södermanland | 59.317260 | 18.037730 |
| 14 | Norra Adolf Fredrik | 19 | 3,816 | 201 | Norrmalm | Upland | 59.337890 | 18.060360 |

Let us visualize the map of Stockholm, along with the districts and Boroughs in the Stockholm city centre.

```
In [12]: map_stockholm
```

Out[12]:




## Methodology

After successfully extracting the information about Stockholm city centre from its Wikipedia page, and visualizing it on the city map, we used the Foursquare API to explore venues around the area. Let us explore the 5 most popular/common venues in a given neighborhood.

Following table represents the details with respect to a neighborhood, that includes the area, population, density, Boroughs, Province, lastitude, longitude coordinates, and 5 most common venues.

```
In [27]: stockholm_merged.head()
```

Out[27]:

| | Neighborhood | Area | Population | Density | Borough | Province | Latitude | Longitude | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Djurgården | 290 | 788 | 3 | Östermalm | Upland | 59.32482 | 18.09780 | Theme Park Ride / Attraction | Scandinavian Restaurant | Museum | Tram Station | Music Venue |
| 1 | Fredhäll | 38 | 4,958 | 130 | Kungsholmen | Upland | 59.33103 | 18.00545 | Café | Indian Restaurant | Thai Restaurant | Park | Mini Golf |
| 2 | Gustav Vasa | 80 | 12,911 | 161 | Norrmalm | Upland | 59.34250 | 18.04775 | Café | Pizza Place | Bakery | Italian Restaurant | Indian Restaurant |
| 3 | Gärdet | 459 | 18,158 | 40 | Östermalm | Upland | 59.33361 | 18.11356 | Science Museum | Middle Eastern Restaurant | History Museum | Harbor / Marina | Museum |
| 4 | Hedvig Eleonora | 58 | 10,387 | 179 | Östermalm | Upland | 59.33524 | 18.08048 | Scandinavian Restaurant | Hotel | Italian Restaurant | Seafood Restaurant | Clothing Store |

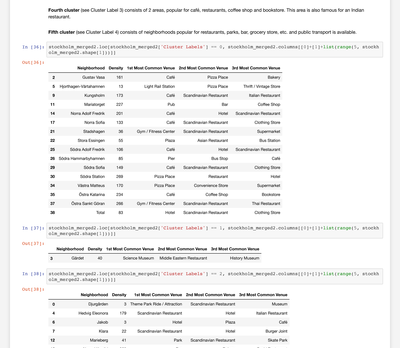
## Clustering the Neighborhoods using K-means

In order cluster the neighborhoods based on their similar characteristics, K-means clustering approach was used. K-means clustering is one of the simplest and popular unsupervised machine learning algorithms, where we specified total number of clusters as 5.

One can visualize these clusters on a map as follows:

```
In [35]: map_clusters
```

Out[35]:




## Results and Discussion

As a result, the neighborhoods around the Stockholm city centre were segmented into 5 clusters using the K-means clustering approach (unsupervised machine learning, where each cluster represent a group of similar neighborhoods. In the following data frames, one can observe the details about the venues in each of the 5 clusters.

**First cluster** (see Cluster Label 0) represents the neighborhoods sharing similar venues, such as café, pub, pizza place, restaurants, gym etc. being the most popular. This is a nice neighborhood to visit if you want to enjoy a nice variety of food, that includes Scandanavian, Thai, and Italian choices, to mention a few. This neighborhood is densely populated and quite popular amongst the locals and the visitors. It is also well connected with public transport, making it easier to commute around the area. Also, there are hotels nearby, making it a perfect neighborhood for tourists to stay and explore. Other facilities include supermarket, clothing store, bookstore and shopping plaza.

**Second cluster** (see Cluster Label 1) consists of a less crowded neighborhood "Gärdet", which is popular for Science and History museums and, authentic middle eastern restaurant. This is a nice choice of neighborhood if you enjoy visiting museums, and prefer less crowded places.

**Third cluster** (see Cluster Label 2) is dominated by neighborhoods with tourist attractions such as theme parks, skate park, beaches, lake, museums, and is also popular for the nightlife. There are a variety of restaurants in this neighborhood, including Sushi restaurant. Also, there are hotels and hostels in the vicinity, and the neighborhood is well connected with the public transport.

**Fourth cluster** (see Cluster Label 3) consists of 2 areas, popular for café, restaurants, coffee shop and bookstore. This area is also famous for an Indian restaurant.

**Fifth cluster** (see Cluster Label 4) consists of neighborhoods popular for restaurants, parks, bar, grocery store, etc. and public transport is available.

```
In [36]: stockholm_merged2.loc[stockholm_merged2['Cluster Labels'] == 0, stockholm_merged2.columns[[1]+[3]+list(range(5, stockholm_merged2.shape[1]))]]
```

Out[36]:

| | Neighborhood | Density | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue |
|---|---|---|---|---|---|
| 2 | Gustav Vasa | 161 | Café | Pizza Place | Bakery |
| 5 | Hjorthagen-Värtahamnen | 13 | Light Rail Station | Pizza Place | Thrift / Vintage Store |
| 9 | Kungsholm | 173 | Café | Scandinavian Restaurant | Italian Restaurant |
| 11 | Marieberg | 227 | Pub | Bar | Coffee Shop |
| 14 | Norra Adolf Fredrik | 201 | Café | Hotel | Scandinavian Restaurant |
| 17 | Norra Södra | 133 | Café | Scandinavian Restaurant | Clothing Store |
| 21 | Stadshagen | 36 | Gym / Fitness Center | Scandinavian Restaurant | Supermarket |
| 22 | Stora Essingen | 55 | Plaza | Asian Restaurant | Bus Station |
| 25 | Södra Adolf Fredrik | 106 | Café | Hotel | Scandinavian Restaurant |
| 26 | Södra Hammarbyhamnen | 85 | Pier | Bus Stop | Café |
| 29 | Södra Sofia | 148 | Café | Scandinavian Restaurant | Clothing Store |
| 30 | Södra Station | 269 | Pizza Place | Restaurant | Hotel |
| 34 | Västra Mathaus | 170 | Pizza Place | Convenience Store | Supermarket |
| 35 | Östra Katarina | 234 | Café | Coffee Shop | Bookstore |
| 37 | Östra Sankt Göran | 156 | Gym / Fitness Center | Scandinavian Restaurant | Thai Restaurant |
| 38 | Total | 83 | Hotel | Scandinavian Restaurant | Clothing Store |

```
In [37]: stockholm_merged2.loc[stockholm_merged2['Cluster Labels'] == 1, stockholm_merged2.columns[[1]+[3]+list(range(5, stockholm_merged2.shape[1]))]]
```

Out[37]:

| | Neighborhood | Density | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue |
|---|---|---|---|---|---|
| 3 | Gärdet | 40 | Science Museum | Middle Eastern Restaurant | History Museum |

```
In [38]: stockholm_merged2.loc[stockholm_merged2['Cluster Labels'] == 2, stockholm_merged2.columns[[1]+[3]+list(range(5, stockholm_merged2.shape[1]))]]
```

Out[38]:

| | Neighborhood | Density | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue |
|---|---|---|---|---|---|
| 0 | Djurgården | 3 | Theme Park Ride / Attraction | Scandinavian Restaurant | Museum |
| 4 | Hedvig Eleonora | 179 | Scandinavian Restaurant | Hotel | Italian Restaurant |
| 6 | Jakobi | 3 | Hotel | Plaza | Café |
| 7 | Klara | 22 | Scandinavian Restaurant | Hotel | Burger Joint |
| 12 | Mariaberg | 41 | Park | Scandinavian Restaurant | Skate Park |
| 15 | Norra Högalid | 203 | Bar | Bakery | Sushi Restaurant |
| 18 | Oscars Kyrka | 132 | Scandinavian Restaurant | Park | Italian Restaurant |
| 19 | Reimarholme-Långholmen | 45 | Hotel | Beach | Other Nightlife |
| 20 | Roslagstull | 301 | Sushi Restaurant | Miscellaneous Shop | Lake |
| 23 | Stortorget | 72 | Scandinavian Restaurant | Plaza | Café |
| 24 | Stureplan-Lärkstaden | 114 | Scandinavian Restaurant | Hotel | Asian Restaurant |
| 28 | Södra Johannes | 134 | Scandinavian Restaurant | Café | Hostel |
| 32 | Universitetet | 7 | Café | Convenience Store | Bus Station |

```
In [39]: stockholm_merged2.loc[stockholm_merged2['Cluster Labels'] == 3, stockholm_merged2.columns[[1]+[3]+list(range(5, stockholm_merged2.shape[1]))]]
```

Out[39]:

| | Neighborhood | Density | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue |
|---|---|---|---|---|---|
| 1 | Fredhäll | 130 | Café | Indian Restaurant | Thai Restaurant |
| 33 | Västra Katarina | 156 | Café | Coffee Shop | Bookstore |

```
In [40]: stockholm_merged2.loc[stockholm_merged2['Cluster Labels'] == 4, stockholm_merged2.columns[[1]+[3]+list(range(5, stockholm_merged2.shape[1]))]]
```

Out[40]:

| | Neighborhood | Density | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue |
|---|---|---|---|---|---|
| 8 | Kristineberg | 88 | Park | Middle Eastern Restaurant | Bus Station |
| 10 | Lilla Essingen | 196 | Grocery Store | Restaurant | Thai Restaurant |
| 13 | Mellersta Högalid | 184 | Bar | Bakery | Sushi Restaurant |
| 16 | Norra Johannes | 137 | Scandinavian Restaurant | Café | Hostel |
| 27 | Södra Högalid | 58 | Bar | Bakery | Sushi Restaurant |
| 31 | Tavricka Högskolan | 22 | Sushi Restaurant | Italian Restaurant | Gym / Fitness Center |

## Conclusion

The purpose of this project was to explore the Stockholm city centre and find out the most popular places to visit. Using the Foursquare API, we gathered the location data about the popular venues around the Stockholm city centre. Using unsupervised machine learning approach, K-means clustering, we clustered the locations to generate the major areas of interest containing the most popular places to explore.

Final decision on the best neighborhood to visit will be made by the tourists, visitors or investors, based on their personal preferences about the specific characteristics they like in a neighborhood in every recommended zone (or cluster).

For example:

- If you are a tourist visiting the Stockholm city center, the neighborhoods in the third cluster are a must visit! The venues suggested in the first cluster are also very interesting.
- If you are an investor, interested in the restaurant business, do check the neighborhoods and venues in the first, third and fifth cluster.
- If you want to visit museums, check out the first and third cluster.
- Do check out the map above to explore the neighborhoods, and refer to the tables for information about the venues that interests you.

Thank you for stopping by!
This page will be further updated, as I am still exploring the power of data science!