**Introduction**

Stockholm is the capital and most populous urban area of Sweden as well as in Scandinavia. Around **1** million people live in the municipality, approximately **1.6** million in the urban area, and **2.4** million in the metropolitan area. The city stretches across fourteen islands, and it's one of the five fastest-growing cities in Europe.

Geographically, the Stockholm city area only covers **188** squared kilometers; it stretches to **6,519** square kilometers for the metropolitan area. The disparity between the city center's density and the metropolitan area is more than ten folds. The city center's density is **5200** per sqr/km, the density in the metropolitan area is only **370** per sqr/km. There are more than **3000** restaurants in Stockholm. The biggest food retailers in Sweden collectively have around **460** stores in Stockholm.

**Problem**

The food supply in Stockholm metropolitan area concentrates mainly within the center, while the peripheries have less coverage. Opening a new food venue might be more profitable if the new establishments are located in higher food demand areas. To tackle this problem, we need to estimate the food demand in each neighborhood.

1. Define Stockholm's neighborhoods
2. Estimate the food demand in each neighborhood

**Target market**

This study can benefit both food retailers and new investors in Stockholm's food and beverage market, pointing out the regions with the least food coverage can open the path for various investment opportunities.

**Data**

**The data is obtained from various sources. The issue of data availability is overcome by creating our own sources.**

To answer the question of which neighborhood in Stockholm has the least coverage of food venues, we should look into two main variables: the number of venues in each neighborhood. The Second variable is concerning the number of inhabitants in the neighborhood. In the list below, I include the three sources of data used in this study.

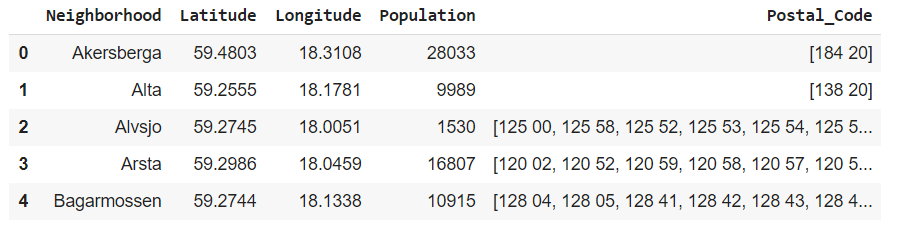
1. The name and the postal code for each neighborhood collected and scrapped from the following website: [https://worldpostalcode.com](https://worldpostalcode.com/)
2. The Latitude, the Longitude, and the Population for each neighborhood in Stockholm were manually collected and gathered in one CSV file.
3. Name, location, latitude, longitude for each food venue in the neighborhoods collected from [https://api.foursquare.com](https://api.foursquare.com/) website.

**Assumptions**

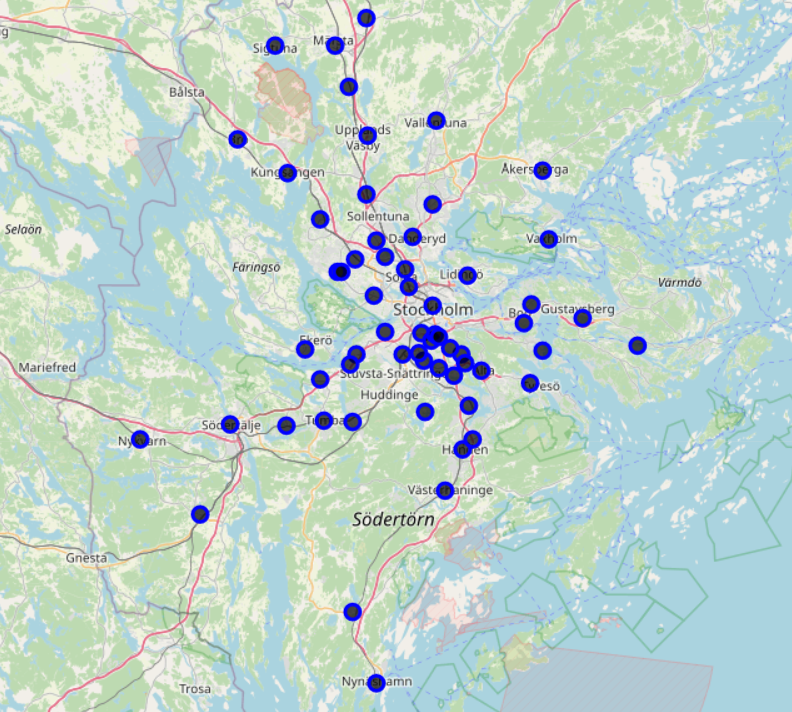
Due to the difference in each neighborhood's characteristics, there should be a difference in the food supply. The number of food venues per inhabitants, Ceteris paribus, can reflect the food supply.

**Methodology**

To achieve the goal above, few steps must be taken. First, creating a dataset that contains the information needed. The table below gives an idea about the data structure. The name of the neighborhood, coordinates, and each one's population are listed within the table

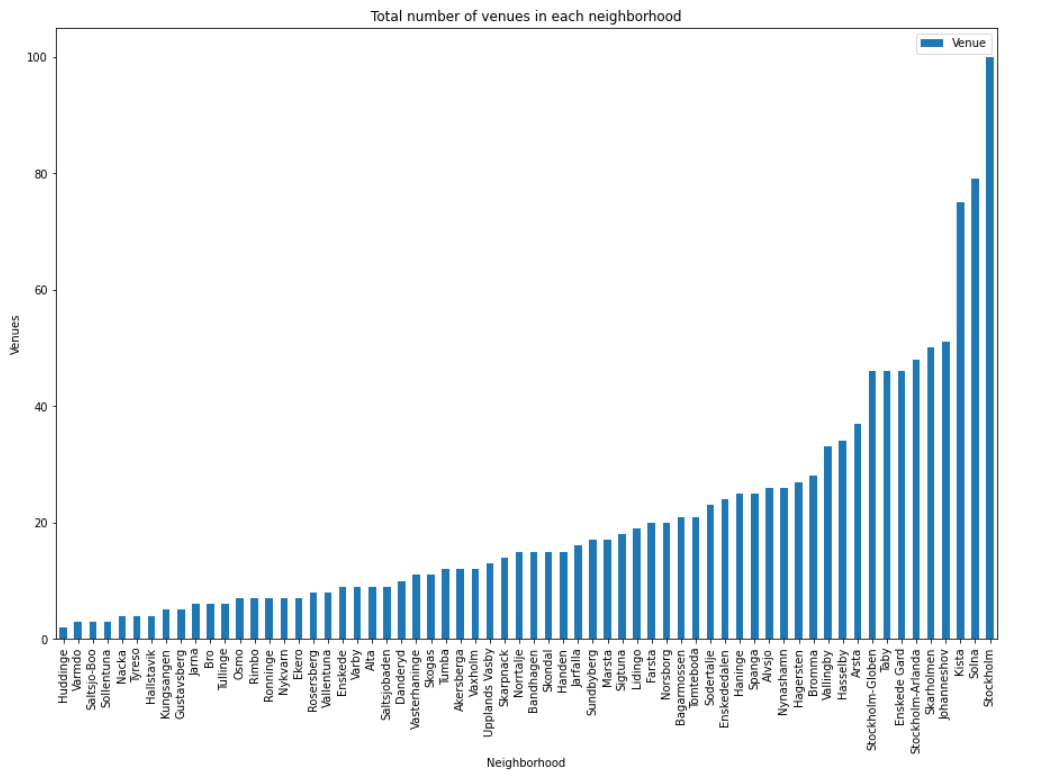


Understanding the data requires visualization, to illustrate our data set I use the geospatial Analysis tool called Folium, Folium is a Python library used for visualizing geospatial data. It is easy to use and yet a powerful library. Folium is a Python wrapper for Leaflet.js which is a leading open-source JavaScript library for plotting interactive maps. As we can see in the map below the neighborhoods agglomerate in the center of Stockholm and despair moving towards the peripheries.



As I mentioned above, the data collection continues by obtaining names, coordinates, and the type of each food venue within the Stockholm metropolitan area. Utilizing Foursquare API leads to obtain the information needed. **1311**venues are retrieved

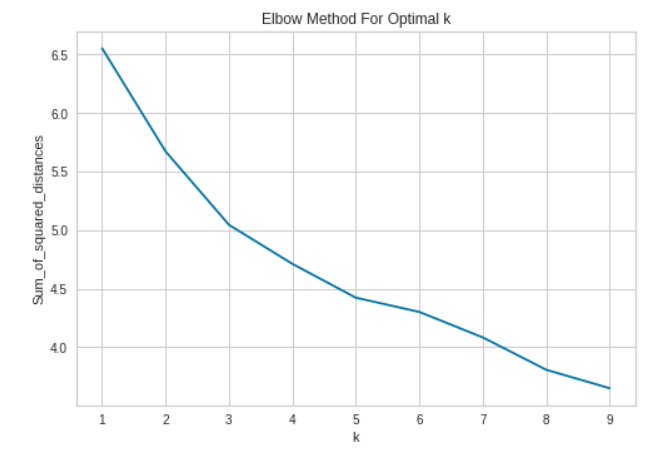
In the graph below, we can see the total number of venues in each neighborhood.



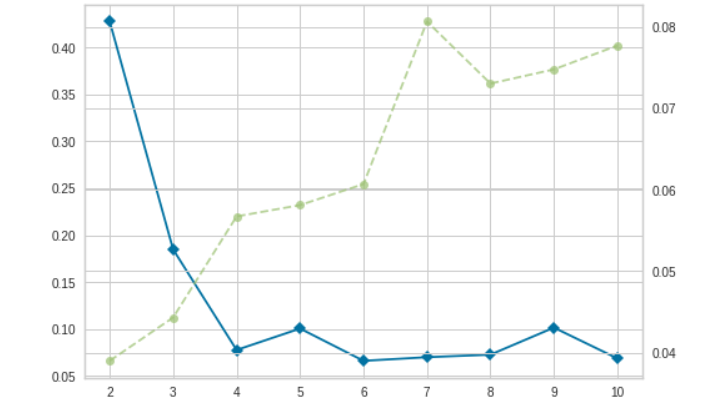
As we can see, there is a huge disparity in the number of venues across the neighborhood, this short-sighted approach ignores the fact that those neighborhoods also differ greatly with the number of inhabitants. Also, the type of venues also differ a lot, where those venues fall under 208 different categories. Inspecting the data shows that pizza places (Pizzeria) rank as the most common venues more than any other venues, other trends emerge by inspecting the data thoroughly. The table below is an example of the regularity of venues in some neighborhoods.



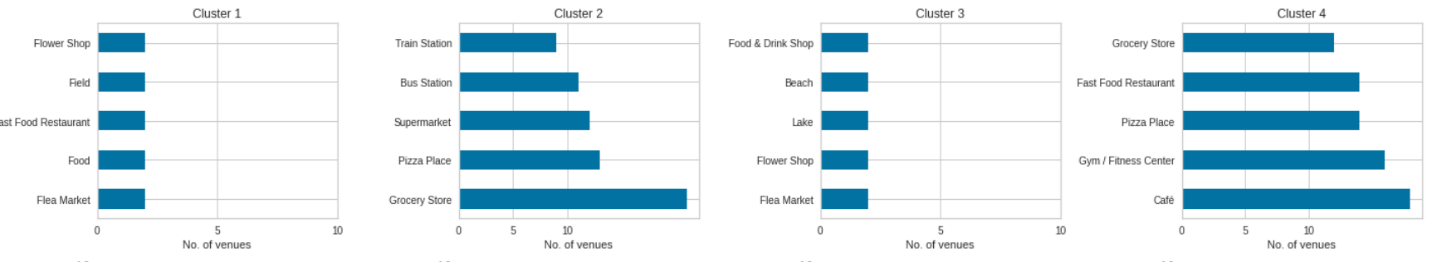
The commonality of specific venue categories in the neighborhoods lead us to use an unsupervised learning **K-means algorithm** to cluster the neighborhoods. K-Means is a method of vector quantization that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centroid). **K-means algorithm** is one of the most common cluster methods of unsupervised learning. The first step in clustering the data using the algorithm mentioned above is to determine the optimal number of clusters. Using the K-elbow shows a K of four clusters as a good fit, even though the result is inconclusive.



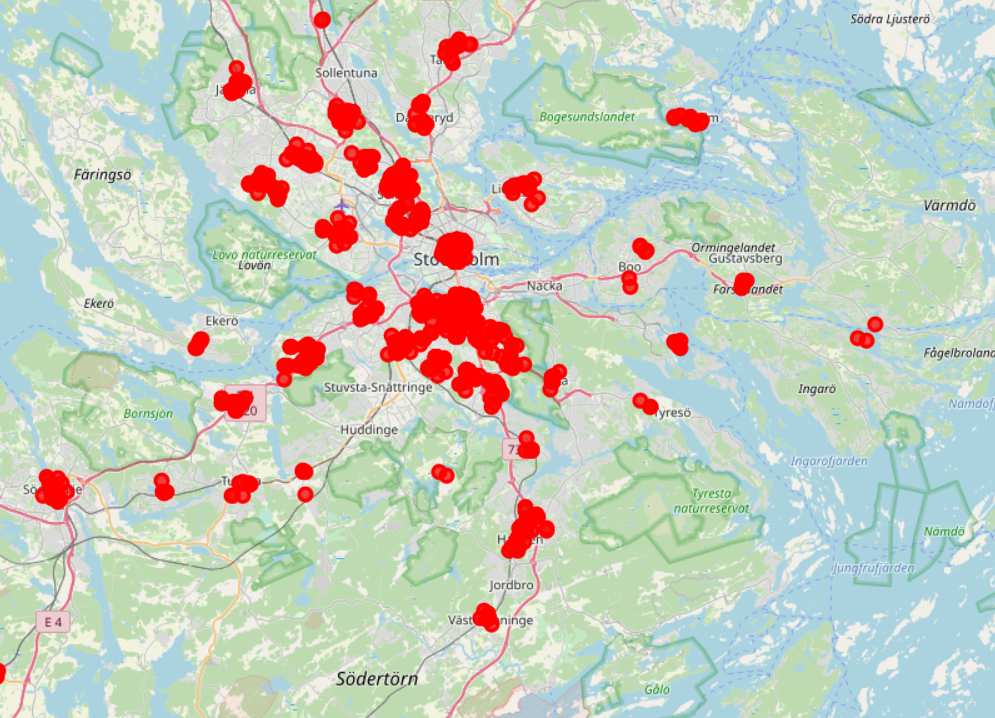
To overcome this inconclusive result, a more advanced method to calculate the K-elbow is used with different scoring metrics to evaluate the clusters. in the graph below a mean ratio of intra-cluster and nearest-cluster distance is used to evaluate the clusters.



Two of the four clusters contains most of the neighborhoods with 22 and 37 neighborhoods in those clusters respectively, while the other two clusters consist of only two neighborhood each. The distribution of the most common venues in each cluster is represented in the bar chart below



As we can see both grocery stores and pizza places account for a huge chunk of neighborhood venues. to illustrate the clustering of venues in Stockholm I draw the following map; this map shows again the agglomeration of venues in the center and the disparity in the peripheries.



To achieve the study goal, I create a new variable to represent the density of food venues in each neighborhood with respect to the number of inhabitants. this new variable shows that some neighborhoods lack the proper coverage of food venues; this opens an opportunity for new investors to investigate the matter in-depth and fill this gap with new venues and services.



As shown above, there is a huge disparity in the supply levels across Stockholm. Some neighborhoods might benefit immediately from opening any food venues in it. Huddinge, Nacka and Sollentuna rank as least covered neighborhoods. Further investigation will help in understanding the need of each neighborhood; the type of food venues might differ according to the need of the respective neighborhood