**Predicting the Possible Location for Opening a Restaurant in Stockholm, Sweden.**

*- by Md. Nazmus Sakib*

**1.Introduction**

**1.1 Background**

In this project, we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening a Coffee Shop/Restaurant in Stockholm, Sweden.

**1.2 Problem**

Since there are lots of restaurants in Stockholm we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no coffee shop in the vicinity. We would also prefer locations as close to the city center as possible, assuming that the first two conditions are met.

**1.3 Interest**

We will use our data science powers to generate a few most promising neighborhoods based on these criteria. Advantages of each area will then be clearly expressed so that the best possible final location can be chosen by stakeholders.

**2. Data acquisition and cleaning**

**2.1 Data sources**

Based on definition of our problem, factors that will influence our decision are:

* number of existing coffee shop in the neighborhood (including restaurant offer coffee)
* number of and distance to other coffee shop/restaurant in the neighborhood, if any
* distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

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

* centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding
* number of coffee shop and their type and location in every neighborhood will be obtained using Foursquare API
* coordinate of Stockholm center will be obtained using Google Maps API geocoding

**3. Data Requirements and Methodology**

**Data Acquisition Approach**

To acquire the data mentioned above, I will need to do the following:

* Get geolocator lat and long coordinates for Stockholm, Sweden
* Use Foursquare API to get a list of all venues in Stockholm
  + Get venue name, venue ID, location, category, and likes

**4. Methodology**

The thought process behind this is that likes are a proxy for quality. The more likes there are, the better the restaurant is. This might be incorrect but API call issues (how many I can use for free) holds me back from getting price / rating data. I will then bin this data into a quality categorical variables so we can cluster appropriately.

I am also going to create new categorical variables for the restaurants to better group them based on type of cuisine. This way you can look for good Mexican food or now what type of food might– be best to eat in Stockholm if you are new to the area.

I will take the gathered data (see above in Data Acquisition Approach and Data Required sections) and will create a k-means clustering algorithm that groups restaurants into 4-5 clusters so that people looking to eat in Stockholm can easily see which restaurants are the best to eat at, what cuisine is available and where in Stockholm they can look to eat.

**5. Results**

Our analysis shows that although there is a great number of restaurants in Stockholm, there are pockets of low restaurant density fairly close to city center.Another borough was identified as potentially interesting but our attention was focused on offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics *and* a number of pockets of low restaurant density.

After directing our attention to this more narrow area of interest we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with a restaurant closer than 400m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors

Result of all this is all zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Italian restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to Stockholm center but not crowded with existing restaurants - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

Running my clustering algorithm, I was able to generate four clusters of restaurants. These are as follows:

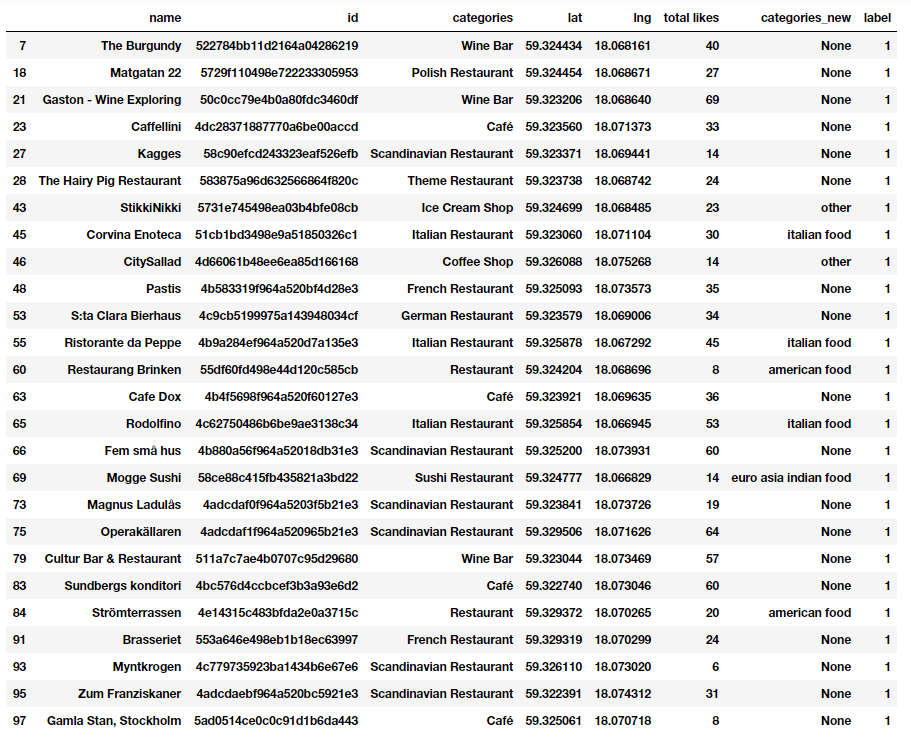
**Cluster 1**

* Characteristics
  + Good quality food
  + Mostly Bar or Cafe



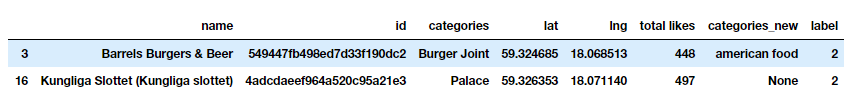
**Cluster 2**

* Characteristics
  + Below average quality food
  + Mostly Scandinavian food



**Cluster 3**

* Characteristics
  + High quality food
  + American inspired food

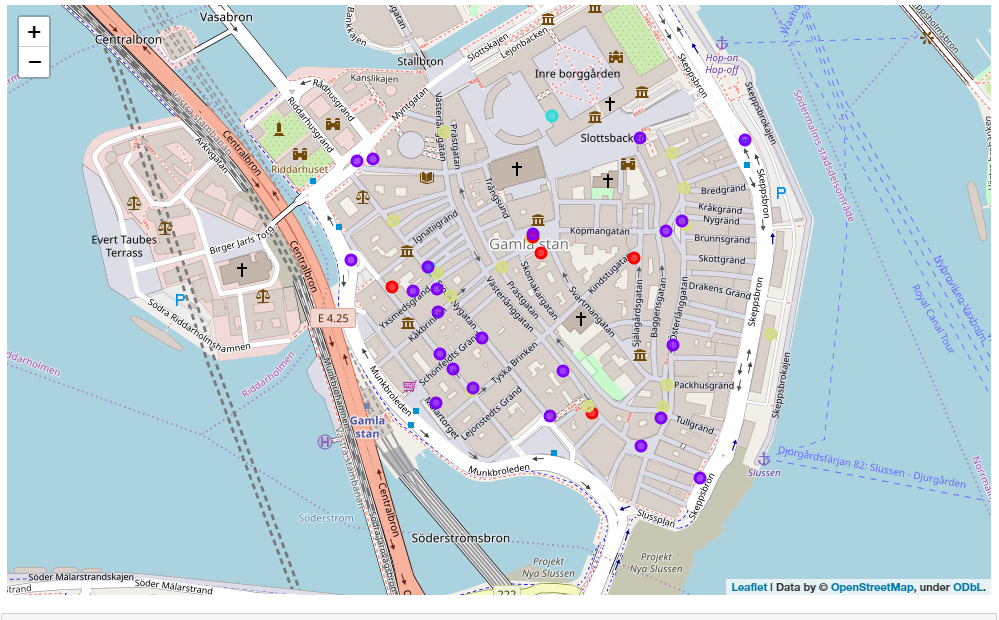


**Cluster 4**

* Characteristics
  + Above average quality food
  + Mostly Bars / Cafe



**Map of Clusters for Users**



**6. Conclusion**

Purpose of this project was to identify Stockholm areas close to center with low number of restaurants in order to aid stakeholders in narrowing down the search for optimal location for a new restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.