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

- by Aditya Kulgod

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

	name	id	categories	lat	lng	total likes	categories_new	label
1	Tweed	51434b65e4b00cf8d47795f1	Cocktail Bar	59.324471	18.067696	198	bars	0
14	Pharmarium	51ccb742498e64f625f8ce57	Cocktail Bar	59.324851	18.070891	178	bars	0
22	AIFUR I krog & bar	4ece875cb634dd298f5ea304	Scandinavian Restaurant	59.323097	18.071993	271	None	0
39	Under Kastanjen	4e198d6e18a8166f7396ffaa	Café	59.324794	18.072893	217	None	0
42	Chokladkoppen	4adcdaebf964a520e35921e3	Café	59.325020	18.070727	221	None	0

Cluster 2
- Characteristics
    - Below average quality food
    - Mostly Scandinavian food

	name	id	categories	lat	lng	total likes	categories_new	label
7	The Burgundy	522784bb11d2164a04286219	Wine Bar	59.324434	18.068161	40	None	1
18	Matgatan 22	5729f110498e722233305953	Polish Restaurant	59.324454	18.068671	27	None	1
21	Gaston - Wine Exploring	50c0cc79e4b0a80fdc3460df	Wine Bar	59.323206	18.068640	69	None	1
23	Caffellini	4dc28371887770a6be00accd	Café	59.323560	18.071373	33	None	1
27	Kagges	58c90efcd243323eaf526efb	Scandinavian Restaurant	59.323371	18.069441	14	None	1
28	The Hairy Pig Restaurant	583875a96d632566864f820c	Theme Restaurant	59.323738	18.068742	24	None	1
43	StikkiNikki	5731e745498ea03b4bfe08cb	Ice Cream Shop	59.324699	18.068485	23	other	1
45	Corvina Enoteca	51cb1bd3498e9a51850326c1	Italian Restaurant	59.323060	18.071104	30	italian food	1
46	CitySallad	4d66061b48ee6ea85d166168	Coffee Shop	59.326088	18.075268	14	other	1
48	Pastis	4b583319f964a520bf4d28e3	French Restaurant	59.325093	18.073573	35	None	1
53	S:ta Clara Bierhaus	4c9cb5199975a143948034cf	German Restaurant	59.323579	18.069006	34	None	1
55	Ristorante da Peppe	4b9a284ef964a520d7a135e3	Italian Restaurant	59.325878	18.067292	45	italian food	1
60	Restaurang Brinken	55df60fd498e44d120c585cb	Restaurant	59.324204	18.068696	8	american food	1
63	Cafe Dox	4b4f5698f964a520f60127e3	Café	59.323921	18.069635	36	None	1
65	Rodolfino	4c62750486b6be9ae3138c34	Italian Restaurant	59.325854	18.066945	53	italian food	1
66	Fem små hus	4b880a56f964a52018db31e3	Scandinavian Restaurant	59.325200	18.073931	60	None	1
69	Mogge Sushi	58ce88c415fb435821a3bd22	Sushi Restaurant	59.324777	18.066829	14	euro asia indian food	1
73	Magnus Ladulås	4adcdaf0f964a5203f5b21e3	Scandinavian Restaurant	59.323841	18.073726	19	None	1
75	Operakällaren	4adcdaf1f964a520965b21e3	Scandinavian Restaurant	59.329506	18.071626	64	None	1
79	Cultur Bar & Restaurant	511a7c7ae4b0707c95d29680	Wine Bar	59.323044	18.073469	57	None	1
83	Sundbergs konditori	4bc576d4ccbcef3b3a93e6d2	Café	59.322740	18.073046	60	None	1
84	Strömterrassen	4e14315c483bfda2e0a3715c	Restaurant	59.329372	18.070265	20	american food	1
91	Brasseriet	553a646e498eb1b18ec63997	French Restaurant	59.329319	18.070299	24	None	1
93	Myntkrogen	4c779735923ba1434b6e67e6	Scandinavian Restaurant	59.326110	18.073020	6	None	1
95	Zum Franziskaner	4adcdaebf964a520bc5921e3	Scandinavian Restaurant	59.322391	18.074312	31	None	1
97	Gamla Stan, Stockholm	5ad0514ce0c0c91d1b6da443	Café	59.325061	18.070718	8	None	1

Cluster 3
- Characteristics
  - High quality food
  - American inspired food

	name	id	categories	lat	lng	total likes	categories_new	label
3	Barrels Burgers & Beer	549447fb498ed7d33f190dc2	Burger Joint	59.324685	18.068513	448	american food	2
16	Kungliga Slottet (Kungliga slottet)	4adcdaeef964a520c95a21e3	Palace	59.326353	18.071140	497	None	2

Cluster 4
- Characteristics
  - Above average quality food
  - Mostly Bars / Cafe

	name	id	categories	lat	lng	total likes	categories_new	label
5	Tradition	534d0a34498ef782062ac89c	Scandinavian Restaurant	59.325940	18.073710	96	None	3
6	Corner Club	514c0868e4b07a3d6af2fca3	Cocktail Bar	59.323342	18.069431	101	bars	3
8	Slingerbulten	4ba0c261f964a520a77b37e3	Scandinavian Restaurant	59.324376	18.068967	128	None	3
9	The Hairy Pig Deli	537f7e60498e54e075c24907	Deli / Bodega	59.325555	18.073958	100	None	3
10	19 Glas Bar & Matsal	4b9a5e8ff964a52072af35e3	Wine Bar	59.324633	18.068670	109	None	3
19	Kryp In	4c78322f93ef236a9772ab0f	Scandinavian Restaurant	59.324694	18.070059	120	None	3
20	Geronimo's FGT	550867d7498e86f0d8c9a61d	Mexican Restaurant	59.325198	18.067725	75	mex southam food	3
24	Ardbeg Embassy	4b63079df964a520945e2ae3	Scandinavian Restaurant	59.323170	18.071904	144	None	3
29	Café Schweizer Konditori	4d381d443ffba143aed55856	Café	59.326174	18.068824	104	None	3
30	Café Stiernan	4c63fd4479d1e21eacddd915	Noodle House	59.323406	18.073597	75	None	3
34	Österlånggatan 17	4b180045f964a52039cb23e3	Scandinavian Restaurant	59.325151	18.074016	112	None	3
37	Kaffekoppen	4adcdaedf964a520555a21e3	Café	59.325016	18.070693	99	None	3
41	Den Gyldene Freden	4adcdaf0f964a520655b21e3	Scandinavian Restaurant	59.323159	18.073512	114	None	3
62	Mr. French	51829a91498e80962fe0411e	Seafood Restaurant	59.323966	18.075842	152	italian food	3

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.