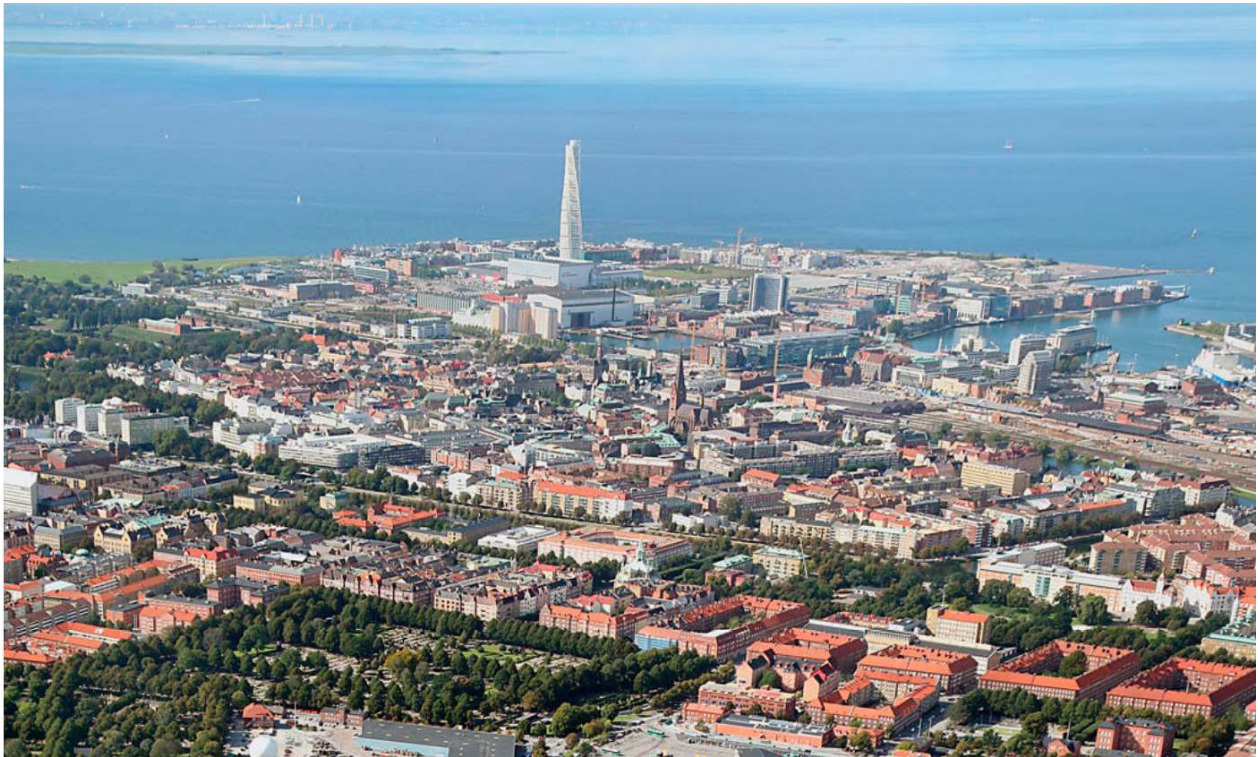


Evaluation of Boroughs in Malmö by Venue categories and apartment prices



Ricardo Marchao

October 10, 2019

1. Introduction

1.1 Background:

I have been living in Malmö, Sweden for the last 15 years. I am not a Swede myself. I came here as an expat due to work reasons. My company offered me a choice of staying in an apartment, and I could choose from several apartments that were available within budget.

I remember my first choice: it was the one that had the nicest furniture. Only later when I started to get to know the neighborhood did I realize that there weren't any great restaurants around or bars, so for me it was not ideal, I was single and did not do any cooking at home. So after 6 months I moved to another apartment that was closer to the office and luckily, with plenty of restaurants and bars around, so I regret a bit the hastily choice I did at first.

I realize that the problem that I faced then is still a valid one today, for people that came today to Malmö

According to Wikipedia: "In 2011, people from 174 countries were represented in Malmö, speaking 150 languages" <https://en.wikipedia.org/wiki/Malm%C3%B6#Demographics>

1.2 Business problem:

For a person that comes to Sweden being single or having a family the problem we are trying to solve is which borough should they move to, given their preferences on certain categories like restaurants vs proximity to playgrounds or shops.

This study will also aim to help choosing the borough given the relative price of the square meter of an apartment versus the percentage of venues of a given category in Malmö. This is a way to help establish the "value for money" of a given borough on a given category. An area may have a high percentage of shops for example but if the price of the square meter is very high, maybe worth instead to go to a borough with lesser shops but a more affordable price per square meter

1.3 Stakeholders interested:

A single person maybe interested in living close to restaurants and bars to socialize, where as a person with a family may prefer proximity to grocery shops, playgrounds.

This study well help both types of stakeholders finding a borough that best suits their preferences

2. Data

2.1 Data Sources

Following data sources will be used:

Description	Purpose
Wikipedia page for Malmö boroughs aka city districts	This data will enable us to identify the list of boroughs. A total of 10 boroughs were identified
Wikipedia page for each borough to identify the neighborhoods	This data will enable us to identify the list of neighborhoods A total of 137 neighborhoods were identified
Geopy API for geo coordinates	For a given neighborhood address we will receive the geo co-ordinates from the API, these data will be later on used to determine which venues are in the proximity
Foursquare API	The API will return the list of venues with a 500 meters range of a given location.

	One thing to note is that: people add venues and can make changes every day so running the notebook at different times can lead to different results
Maklarstatistik.se	This webpage will provide as with the average price, for the last 12 months, with which apartments were sold in Malmö for a given borough
Folium	This package will enable us to produce maps with heatmaps of boroughs for a given category like "food" venues

2.2 Data acquisition

Below is a summary of the data acquisition method by data source:

Source	Method
Wikipedia page for Malmö to identify the list borough aka city districts	Scraping web page using Selenium package and pandas
Wikipedia page for each borough to identify the neighborhoods	Scraping web page using Selenium package and pandas
Geopy API for geo coordinates	API call
Foursquare API	API call
Maklarstatistik.se	Scraping web page using Selenium package and pandas
Folium	Python package with plugins

2.3 Data cleaning & preparation

Below is a summary of the data cleaning activities by data source

Source	Activities
Wikipedia page for Malmö to identify the list borough aka city districts	Converting of object types to numeric
Wikipedia page for each borough to identify the neighborhoods	Ensuring one neighborhood is assigned to only one borough: there was a case where this was not true and had to be correct
Geopy API for geo coordinates	-Creating a neighborhood address -in same cases the neighborhood address was not found by the Geopy API so a new column was created with a and address that could be recognized

	-the API sometimes times out so a while loop was created until a value is retrieved -Checking all neighborhoods have geo coordinates
Foursquare API	Checking for neighborhoods with no venues Checking for duplicate venues since each neighborhood had a radius of 500 meters the same venues would overlap to multiple neighborhoods The rule to clean duplicates was to take was the keep the first one only. Although in a more through exercise a more careful approach would have to be done While doing data preparation it become clear that there are venues missing from the foursquare API database: such as churches, bus stations, etc... Most data seem to be focuses on the “centrum” borough, which is a more touristic area, and focused on food related places.
Maklarstatistik.se	Conversion of object types to numeric

In terms of data preparation, it is also worth mentioning the creation of auxiliary dataframes via the method “create_df_per_categ_pop”. This method creates for a given category (like shops) a relation between the percentage of the total number of venues and any given borough versus the percentage of the total population of the borough. The dataframes this method produces are later on used in the analysis section

2.4 Data Engineering

Apart from using features directly available from scraped tables, following features have been created:

Source	Features
Foursquare API	<p>The study extracts three categories of venues from the venues API: Two categories: "categ_1" and "categ_2" are extracted from the API link "categ_1" is the main category, it contains:</p> <pre> : food 3 shops 2 parks_outdoors 1 travel arts_entertainment building nightlife education </pre> <p>"categ_2" is a sub-category of "categ_1", the most common values are:</p> <pre> busstation default foodgrocery gym playground cafe park scandinavian pizza coffeeshop falafel plaza bakery stadiumsoccer italian sportsoutdoors fastfood hotel furniture asian </pre>
Maklarstatistik.se	<p>For better readability on charts we create a feature of log price of square meters in SEK (Swedish kronors)</p>

Apart from features directly related to one dataframe there are also features combining two dataframes, for example combining categ_1 with population, for example the table below:

```
food_bor_df_t
```

	food	percentage_food	population	perc_population	ratio_per_food_per_pop
borough					
Centrum	154.0	0.469512	47171.0	0.153998	3.048828
Fosie	16.0	0.048780	43889.0	0.143283	0.340449
Husie	6.0	0.018293	20769.0	0.067804	0.269788
Hyllie	24.0	0.073171	32998.0	0.107727	0.679221
Kirseberg	12.0	0.036585	14959.0	0.048836	0.749145
Limhamn-Bunkeflo	11.0	0.033537	42646.0	0.139225	0.240881
Oxie	0.0	0.000000	12453.0	0.040655	0.000000
Rosengård	9.0	0.027439	23563.0	0.076925	0.356697
Södra Innerstaden	49.0	0.149390	34671.0	0.113189	1.319827
Västra Innerstaden	47.0	0.143293	33191.0	0.108358	1.322406
total	328.0	1.000000	306310.0	1.000000	8.327242

Where we are using percentages of the total venues and for each borough and looking at the percentage of the population that lives in the borough

3. Methodology

The previous steps allow us to gather the needed data and prepare it, so with this in place we can start analyzing the data.

In the analysis section we will be exploring the data and the relationships:

We will start by getting an understanding of what are the possible values of the venue categories : “categ_1” and “categ_2”

We will see which are the most frequent values and we will see how they distributed between the different boroughs

We will also display on the map a heat-map that shows which boroughs have a higher concentration of venues of a particular category

A second section of the analysis is to look at relationships, namely the relationship between the percentage of venues within a category, how it is distributed by the boroughs versus the average apartment price in that same borough. The idea is to find “value for money”. If a borough contains a moderate percentage of venues of a given category, for example “shops” , but the price per square meter of an apartment is high, it may not be worth moving into that borough on the basis of this category alone.

To help this analysis we will make use of a linear regression model, the linear regression line represents fair value, values below could be considered cheap boroughs for that category and above expensive ones.

A third and last section in this analysis will be the clustering of neighborhoods by similarity of venues. The model used here is the k-means, which is an unsupervised machine learning model, meaning we don't need to provide any labels. The way clusters are formed will enable us to check if the city planning is unbalanced or not. For example if a cluster that is predominantly related to food venues contains only 1 or 2 boroughs then it means these venues are not evenly spread out and vice versa if we see clusters that contain all types of venues in multiple boroughs means that in different places of the city you can experience a multitude of venues, so the city is well planned

3.1 Exploratory Data Analysis

The exploratory data analysis contains pictures and charts coming from the notebook on github:

https://github.com/ricmarchao/Coursera_Capstone/blob/master/notebook_week_5_capstone-malmo.ipynb

3.1.1 Getting familiar with the data

3.1.1.1 Boroughs

Let's start by identifying the boroughs in Malmö, we have a total of 10, in table "boroughs_df"

```
boroughs_df # 10 boroughs
```

	City district	Population[9]	Area(hectare)[10]	Density(people/km²)	links
0	Centrum	47171	1757	2685	https://en.wikipedia.org/wiki/Centrum,_Malm%C3%B6
1	Fosie	43889	1243	3531	https://en.wikipedia.org/wiki/Fosie
2	Limhamn-Bunkeflo	42646	5147	829	https://en.wikipedia.org/wiki/Limhamn-Bunkeflo
3	Södra Innerstaden	34671	302	11480	https://en.wikipedia.org/wiki/S%C3%B6dra_Inner...
4	Västra Innerstaden	33191	465	7138	https://en.wikipedia.org/wiki/V%C3%A4stra_Inne...
5	Hyllie	32998	901	3662	https://en.wikipedia.org/wiki/Hyllie
6	Rosengård	23563	332	7097	https://en.wikipedia.org/wiki/Roseng%C3%A5rd
7	Husie	20769	2948	705	https://en.wikipedia.org/wiki/Husie
8	Kirseberg	14959	640	2337	https://en.wikipedia.org/wiki/Kirseberg
9	Oxie	12453	2306	540	https://en.wikipedia.org/wiki/Oxie

The table above shows the boroughs sorted by population as of 1st Jan 2013. Largest one was the “Centrum”

3.1.1.2 Neighborhoods

Within these boroughs we have a total of 136 neighborhoods, stored in the table “neigh_df”:

```
neigh_df.head(10)
```

	neigh_name	neigh_link	borough	address	geopy_address	geopy_lat	geopy_long
0	Davidshall	https://en.wikipedia.org/wiki/Davidshall	Centrum	Davidshall ,Malmö	Davidshall ,Malmö	55.598823	12.998539
1	Ellstorp	https://en.wikipedia.org/wiki/Ellstorp	Centrum	Ellstorp ,Malmö	Ellstorp ,Malmö	55.604265	13.032631
2	Frihamnen, Malmö	https://en.wikipedia.org/wiki/Frihamnen,_Malm...	Centrum	Frihamnen, Malmö	Frihamnen, Malmö	55.615554	13.010210
3	Gamla staden, Malmö	https://en.wikipedia.org/wiki/Gamla_staden,_Ma...	Centrum	Gamla staden, Malmö	Gamla staden, Malmö	55.604449	13.002700
4	Inre hamnen	https://en.wikipedia.org/wiki/Inre_hamnen	Centrum	Inre hamnen ,Malmö	Inre hamnen ,Malmö	55.612736	13.010973
5	Katrinelund	https://en.wikipedia.org/wiki/Katrinelund	Centrum	Katrinelund ,Malmö	Katrinelund ,Malmö	55.601224	13.031803
6	Lugnet, Malmö	https://en.wikipedia.org/wiki/Lugnet,_Malm%C3%B6	Centrum	Lugnet, Malmö	Lugnet, Malmö	55.599860	13.004461
7	Malmöhus	https://en.wikipedia.org/wiki/Malm%C3%B6hus	Centrum	Malmöhus ,Malmö	Malmöhus ,Malmö	55.605466	12.986503
8	Mellersta hamnen	https://en.wikipedia.org/wiki/Mellersta_hamnen	Centrum	Mellersta hamnen ,Malmö	Mellersta hamnen ,Malmö	55.619740	13.009006
9	Norra hamnen, Malmö	https://en.wikipedia.org/wiki/Norra_hamnen,_Ma...	Centrum	Norra hamnen, Malmö	Norra hamnen, Malmö	55.630356	12.995961

3.1.1.3 Venues

Please note: I got a different number of venues from the Foursquare API, just by running the notebook on different days. Could be due that venues are being added and removed constantly, could be some restriction on the API. So the numbers may not be the same if you re-run the notebook

We have a 1425 of venues in the “details_venues_nodup” dataframe:

```
details_venues_nodup.head(10)
```

	name	categories	lat	lng	link	categ_1	categ_2	neigh_name	borough
24	400 grader	Italian Restaurant	55.593553	13.006270	https://ss3.4sqi.net/img/categories_v2/food/it...	food	italian	Mollevången	Södra Innerstaden
4	7Eleven	Automotive Shop	55.574993	12.989868	https://ss3.4sqi.net/img/categories_v2/shops/a...	shops	automotive	Ärtholmen	Hyllie
3	AB Småland	Café	55.597621	13.001020	https://ss3.4sqi.net/img/categories_v2/food/cafe_...	food	cafe	Davidshall	Centrum
0	Airec	Construction & Landscaping	55.619478	13.036019	https://ss3.4sqi.net/img/categories_v2/shops/r...	shops	realestate	Östra hamnen	Centrum
27	Akademibokhandeln	Bookstore	55.580252	13.002945	https://ss3.4sqi.net/img/categories_v2/shops/b...	shops	bookstore	Heleneholm	Fosie
1	Akvarie-Leasing AB	Aquarium	55.633145	13.057006	https://ss3.4sqi.net/img/categories_v2/arts_en...	arts_entertainment	aquarium	Spillepengen	Centrum
11	Akvarettjänst	Pet Store	55.595163	12.990471	https://ss3.4sqi.net/img/categories_v2/shops/p...	shops	petstore	Fågelbacken	Västra Innerstaden
0	Allans Korv (Scan)	Fast Food Restaurant	55.579654	12.996462	https://ss3.4sqi.net/img/categories_v2/food/fa...	food	fastfood	Gröndal, Malmö	Hyllie
4	Alm Livs	Grocery Store	55.575985	13.025230	https://ss3.4sqi.net/img/categories_v2/shops/f...	shops	foodgrocery	Augustenborg, Malmö	Fosie
0	Almgården (B)	Bus Stop	55.578025	13.061206	https://ss3.4sqi.net/img/categories_v2/travel/...	travel	busstation	Almgården, Malmö	Husie

The top boroughs in terms of number of venues are:

```
details_venues_nodup["borough"].value_counts().head(10)
```

Centrum	298
Västra Innerstaden	99
Hyllie	98
Södra Innerstaden	95
Fosie	72
Limhamn-Bunkeflo	63
Kirseberg	45
Husie	34
Rosengård	34
Oxie	19

The top 10 neighborhoods in terms of total number of venues are:

```
neigh_name
```

Gamla staden, Malmö	59
Rådmansvången	51
Lugnet, Malmö	43
Hyllievång	35
Malmöhus	34
Västra hamnen	27
Slussen, Malmö	26
Möllevången	25
Svågertorp	19
Davidshall	19

3.1.1.3.1 Understanding the venues categories

The “details_venues_nodup” has three category columns, we will be using mostly “categ_1” and “categ_2”

“categ_1” and “categ_2” have been retrieved from foursquare from the venue link

“categ_2” is a sub category of “categ_1”

The possible values for “categ_1” and their count are:

```
: details_venues_nodup["categ_1"].value_counts().head(20)
: food                333
: shops               209
: parks_outdoors     121
: travel              79
: arts_entertainment  68
: building            39
: nightlife           19
```

So most of the venues in Foursquare API are related to food and shops

Most common values in “categ_2” are:

```
: details_venues_nodup["categ_2"].value_counts().head(20)
: foodgrocery        50
: busstation          50
: default             40
: gym                 35
: playground          29
: pizza               27
: cafe                25
: park                25
: scandinavian        24
: coffeeshop          23
: plaza               21
: falafel              20
: bakery              17
: stadiumsoccer       16
: italian              16
: fastfood             15
: sportsoutdoors      14
: hotel               14
: asian               13
: furniture           12
```

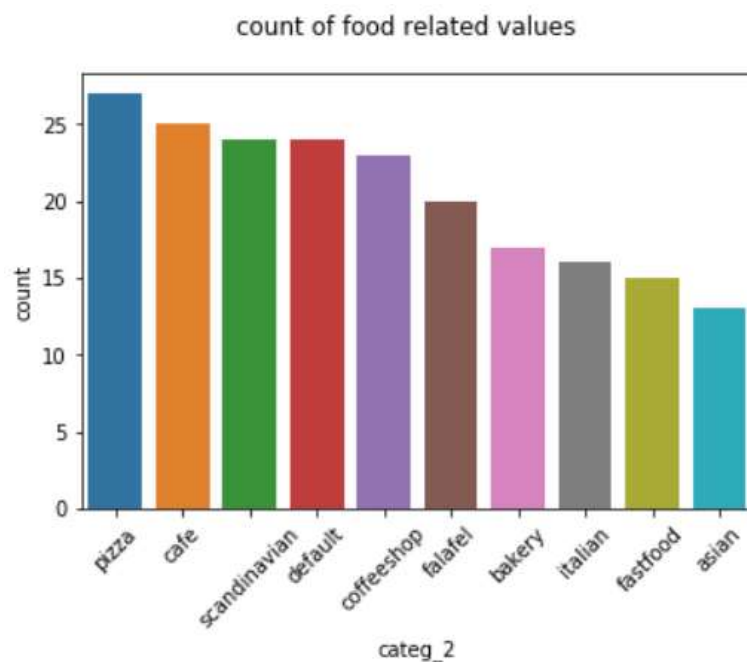
We can see that even though bus station is a top value the data is incomplete as there are more than 49 bus stations in Malmö

We can see that are the most popular food related venues:

```

: details_venues_nodup["categ_2"][details_venues_nodup["categ_1"]=="food"].groupby(details_venues_nodup["categ_2"])
:
categ_2
pizza      27
cafe       25
scandinavian 24
default    24
coffeeshop 23
falafel    20
bakery     17
italian    16
fastfood   15
asian      13

```



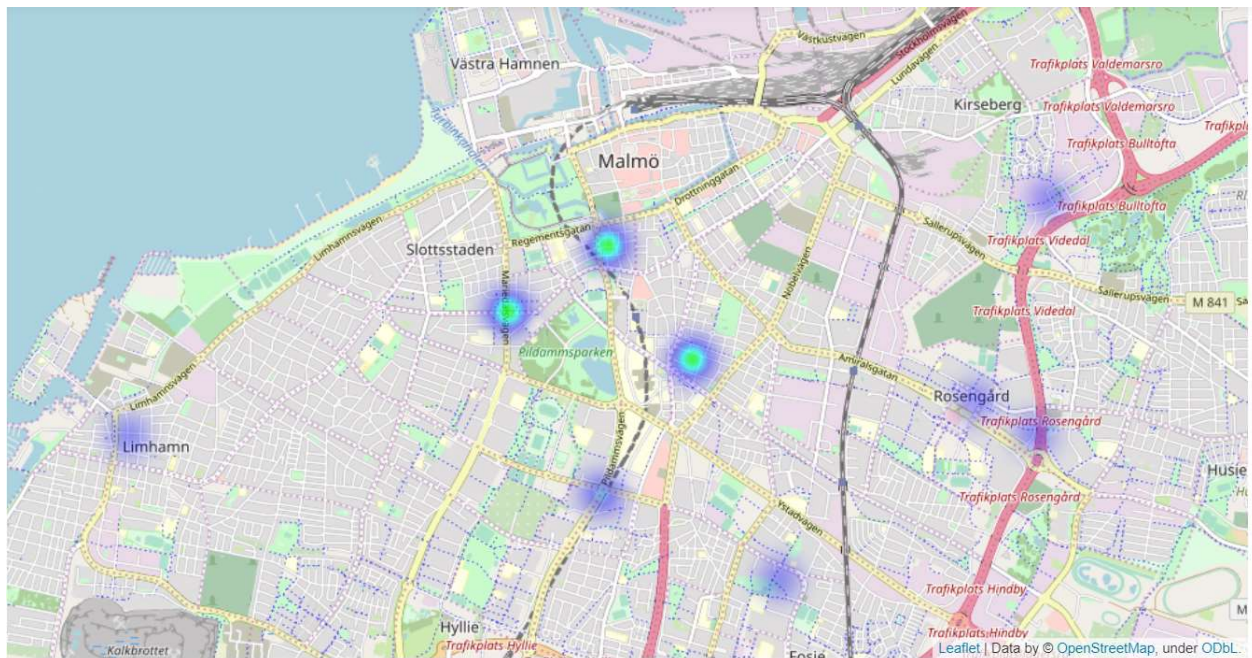
Interesting to see falafel coming in the top list. Confirming the saying that “Malmö is the Swedish capital of falafel”

We can also check the total number of food related venues per borough:

			lat	lng	count
borough					
Centrum	0	55.598697	12.997290	162	
Södra Innerstaden	1	55.590019	13.008885	53	
Västra Innerstaden	0	55.593719	12.983709	45	
Hyllie	5	55.579654	12.996462	22	
Fosie	4	55.573057	13.020688	16	

Centrum has the highest number of food related venues

We can also display the map of Malmö with a heat-map with regards to food venue count:



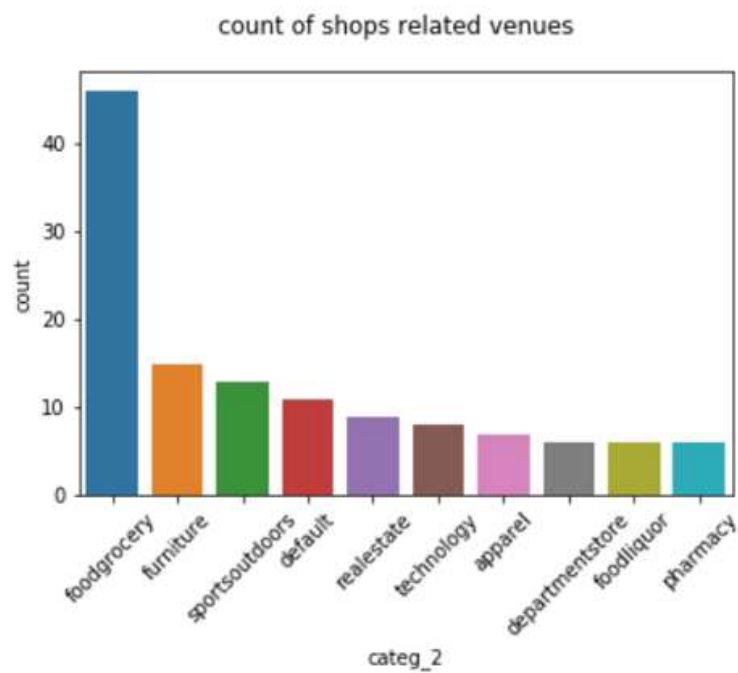
If we look that the shops related venues:

```

categ_2
foodgrocery      50
sportsoutdoors   14
default          12
furniture        12
apparel          10
technology        9
pharmacy          7
departmentstore   6
foodliquor        6
conveniencestore 6

```

And display the bar chart:

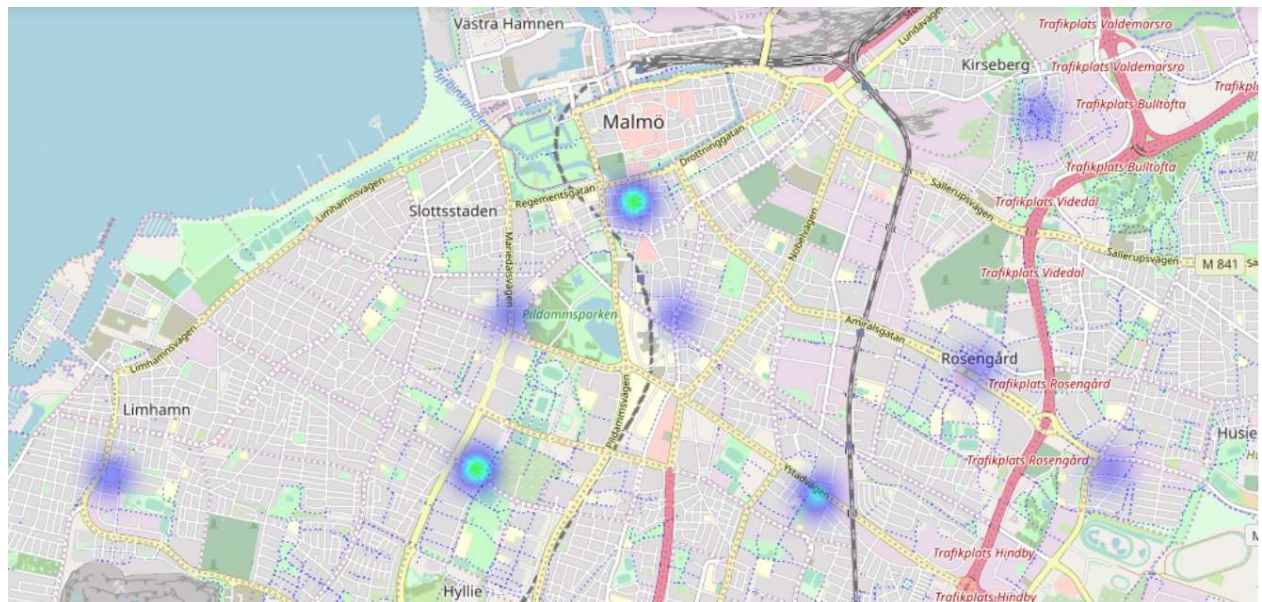


"foodgrocery" are supermarkets and "default" seems to be like a "others" category that doesn't fit into the existing ones.

Looking at the boroughs with the highest number of shop venues:

		lat	lng	count
borough				
Centrum	1	55.599086	13.000199	50
Hyllie	0	55.578703	12.978891	43
Fosie	0	55.576723	13.025019	24
Limhamn-Bunkeflo	1	55.578230	12.929496	21
Södra Innerstaden	0	55.590442	13.005913	18

The Centrum and Hyllie are the leaders, we can also display this on a heat-map:



Families and Singles would then likely appreciate moving to the Centrum as it offers convenient shopping

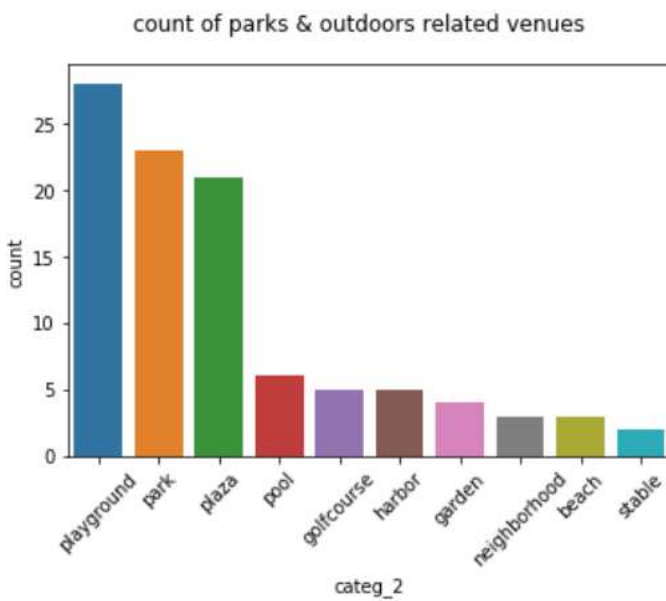
Looking the “parks_outdoors” category, we can see the most popular venues:

```
details_venues_nodup["categ_2"][details_venues_nodup["categ_1"]=="parks_outdoors"].
```

```
categ_2
playground    29
park          25
plaza         21
pool          6
golfcourse    6
harbor        4
sculpture     3
neighborhood  3
garden        3
beach         3
```

The top value is “playground” this could be interesting for families with small children,

We can also visualize it:



We can also check which boroughs have the highest number

			lat	lng	count
borough					
Centrum	22	55.600679	12.998381	31	
Västra Innerstaden	1	55.594715	12.976664	21	
Limhamn-Bunkeflo	4	55.576244	12.930664	17	
Oxie	0	55.524588	13.044890	12	
Södra Innerstaden	7	55.591684	13.007655	10	

Quite interesting to see that Centrum has more playgrounds than any other area, specially since the center of the city is usually an area crowded with business activities, restaurants. The fact that it has more playgrounds than other could be due to Foursquare missing data on other playgrounds or a deliberate effort by the city planners to make it attractive to families to move to the center and create a balanced city centrum.

Looking at the travel category, the top venues are:

```
details_venues_nodup["categ_2"][details_venues_nodup["categ_1"]=="travel"].
```

```

categ_2
busstation      50
hotel           14
trainstation     5
hostel           3
hotelbar         2
default          2
boat             2
movingtarget     1
Name: categ_2, dtype: int64

```

so in travel the most popular item is "bus station", and it has low count I suspect many bus stations have not been added to the venue list. There are several ways to deal with this:

- remove the "busstation" value from the df since it is incomplete
- make no inference based on the "travel" category as it is incomplete and therefore unreliable

I will go for the 2nd option

Looking at the arts& entertainment category, the top venues are:


```
details_venues_nodup["categ_2"][details_venues_nodup["categ_1"]=="arts_entertainment"].g
```

```

categ_2
stadiumsoccer      16
stadium            8
musicvenue         6
movietheater        4
stadiumtennis      4
stadiumhockey      3
performingartstheater 3
performingartsdancestudio 3
museumhistory      3
museumart          3

```

Mostly stadiums and music venues, maybe this area could be more interesting to single persons than families, under the assumption that singles have more free time, and can go watch soccer games or music concerts

In terms of boroughs with the highest count:

			lat	lng	count
borough					
	Centrum	12	55.599143	13.001106	23
	Hyllie	1	55.583859	12.989994	11
	Västra Innerstaden	2	55.596071	12.995151	9
	Södra Innerstaden	3	55.591658	13.027954	6
	Limhamn-Bunkeflo	3	55.576010	12.935554	5

Centrum comes on top again. This could be due to people tagging mostly centrum related venues in the Fourquare API and thus we are missing data. Could also be due Malmö being a relatively small city.

If the data is accurate then a single person would probably want to move to Centrum

Looking at the category "buildings", the top values are:

```
details_venues_nodup["categ_2"][details_venues_nodup["categ_1"]=="building"]
```

categ_2	
gym	35
religiouschurch	1
medicaldoctorsoffice	1
medical	1
default	1

We can see that only one church has been tagged, this is also a clear sign that the foursquare data is not complete, as there are more churches in Malmö.

Last category we look at is “nightlife”, the top values are:

```
details_venues_nodup["categ_2"][details_venues_nodup["categ_1"]=="nightlife"]
```

categ_2	
pub	12
cocktails	3
sportsbar	1
nightclub	1
divebar	1
beergarden	1

This category is mostly pubs and bars, only one nightclub tagged, again seems foursquare list of venues is not complete

3.1.1.4 Apartment prices

the “apartment_prices” dataframe has for each borough the avg price for the last 12 month in SEK (Swedish kronor).

We will be using the log of the prices per square meter

apartment_prices

	num_sold_12_mon	12_mon_sek/sqm	12_mon_price_K_sek	12_mon_price_change	population	log_12_mon_sek/sqm
borough						
Centrum	1497	38429	2760	4.9	47171	10.556568
Hyllie	889	21642	1507	9.9	32998	9.982391
Kirseberg	526	21669	1239	6.6	14959	9.983638
Limhamn-Bunkeflo	642	31129	2305	1.4	42646	10.345895
Södra Innerstaden	1028	26071	1400	6.6	34671	10.168579
Västra Innerstaden	1398	34584	2251	3.5	33191	10.451146
Oxie	463	14475	1125	9.5	12453	9.580178
Fosie	463	14475	1125	9.5	43889	9.580178
Husie	382	17307	1350	1.5	20769	9.758866
Rosengård	382	17307	1350	1.5	23563	9.758866

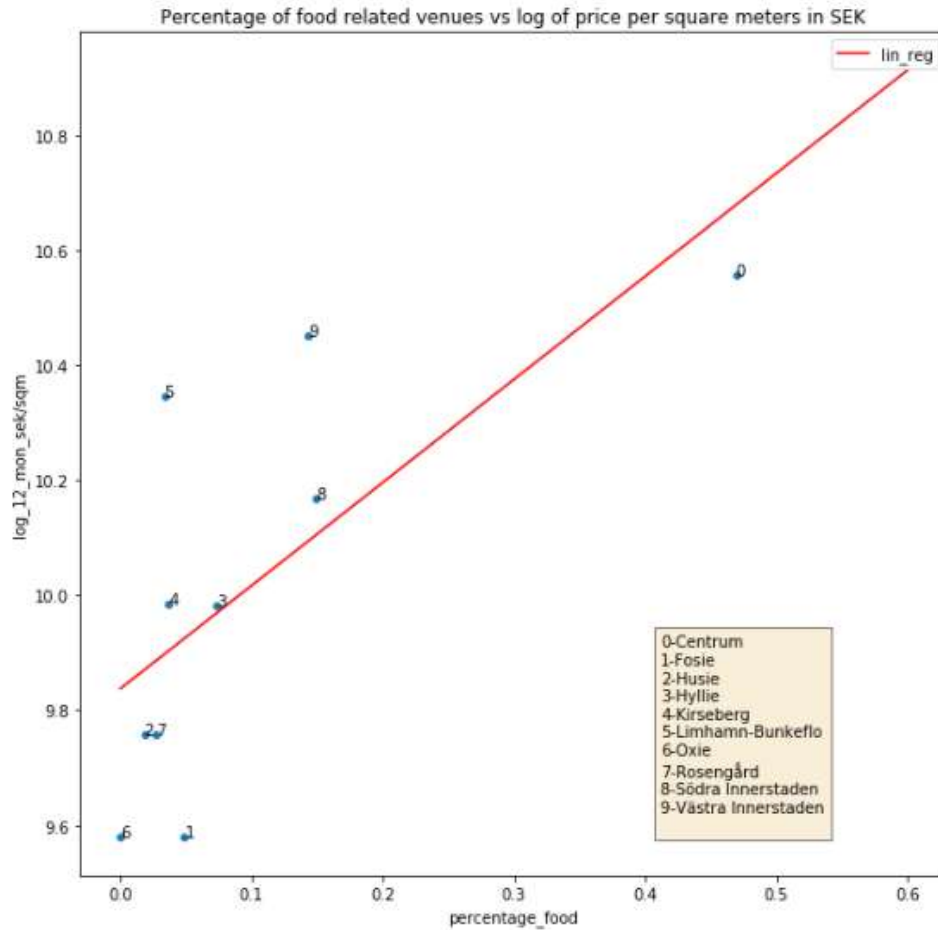
3.1.2 Looking at relationships

Lets look at the relationship between the percentage of venues of a given category that belong to a given borough versus the log price of the square meter in that same borough

We are using here a linear regression where the independent value is the percentage of a count of a given category and the dependent variable is the log price of the square meter

The idea is to find “value for money” , the regression line should give us what is a fair value, point above are over priced and points below under priced

So lets start by looking at the category “food”:

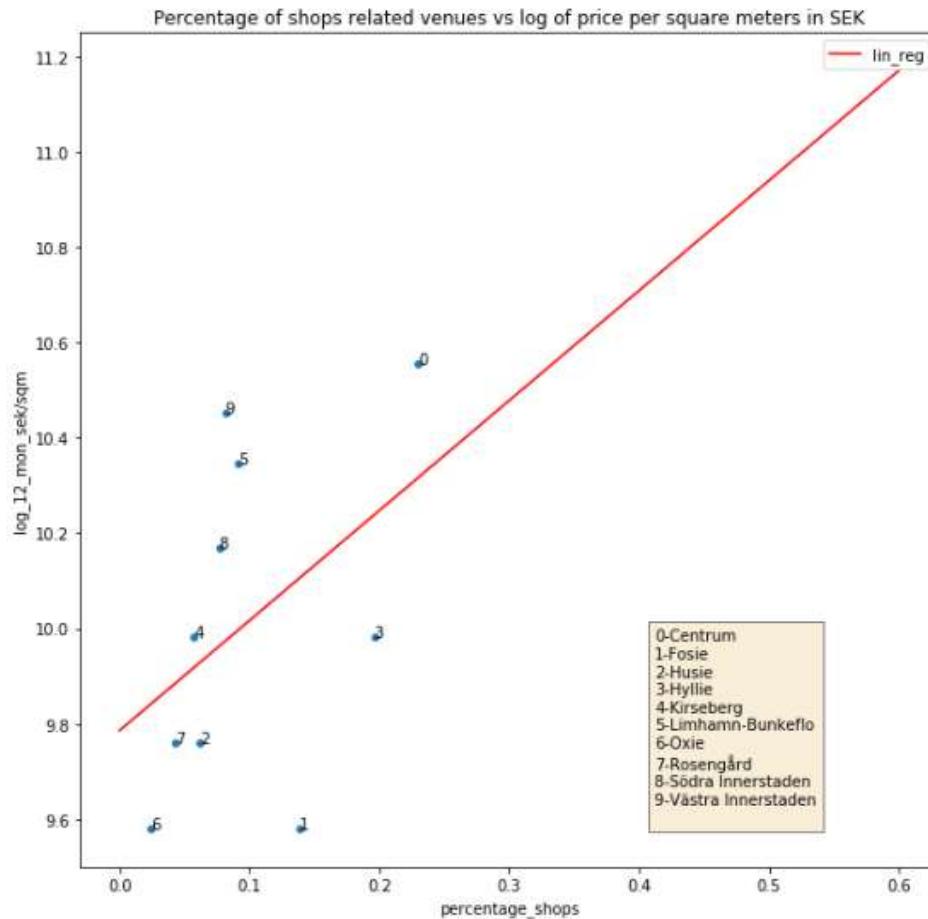


from the figure above it can be seen that 9 "Västra innerstraden" is above the regression line, meaning that given the percentage of food places, in relation to the total for Malmö, the price for square meter is above the average and therefore expensive.

The "Centrum" with label 0 for example even though it has a high percentage of food places with regards to the total food places in Malmö, the price per square meter is below the linear regression line, thus making it relatively cheap.

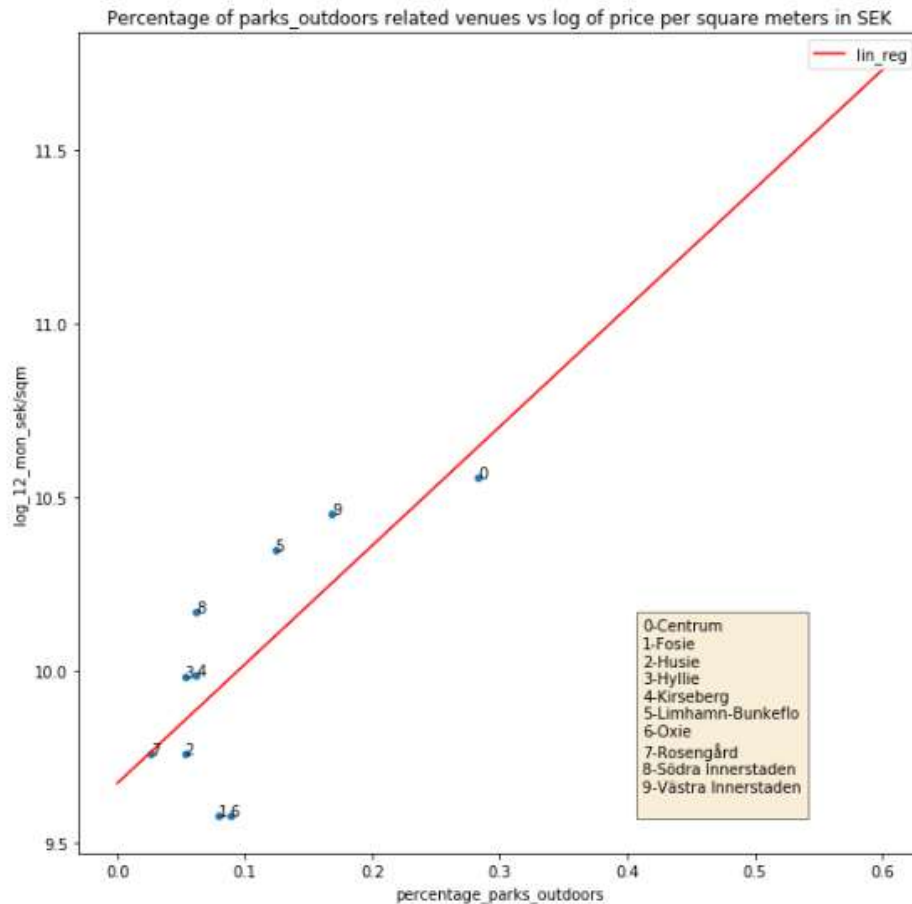
A single person that likes to go out to eat would find the "Centrum" good value for money

Regarding the category "shops":



the picture above shows that "Hyllie" with label 3 has a relatively cheap square meter price versus the number of shops available in the borough. So for families it could be a good area to live as it will be convenient to go around shopping

Looking at the category parks & outdoors:



The picture above shows for example that "Centrum" with label 0, a high number of parks and playgrounds for children and the price for square meter is relatively cheap as it is below the linear regression line, So for families with small children this area could be a good choice according to this criteria

3.1.3 Clustering the neighborhoods

In a similar fashion to what was shown in the course in week 3, we will group neighborhoods by venues to find to ones that are most similar by venues.

We will be using a machine learning model called K-means, clusters together datasets based on their similarity. In this case the similarity of venues categories.

K-means is an unsupervised learning algorithm , meaning we don't need to provide the target labels only the feature values.

We will use one hot encoding to convert the categories into binary values and calculate the average per category for each neighborhood.

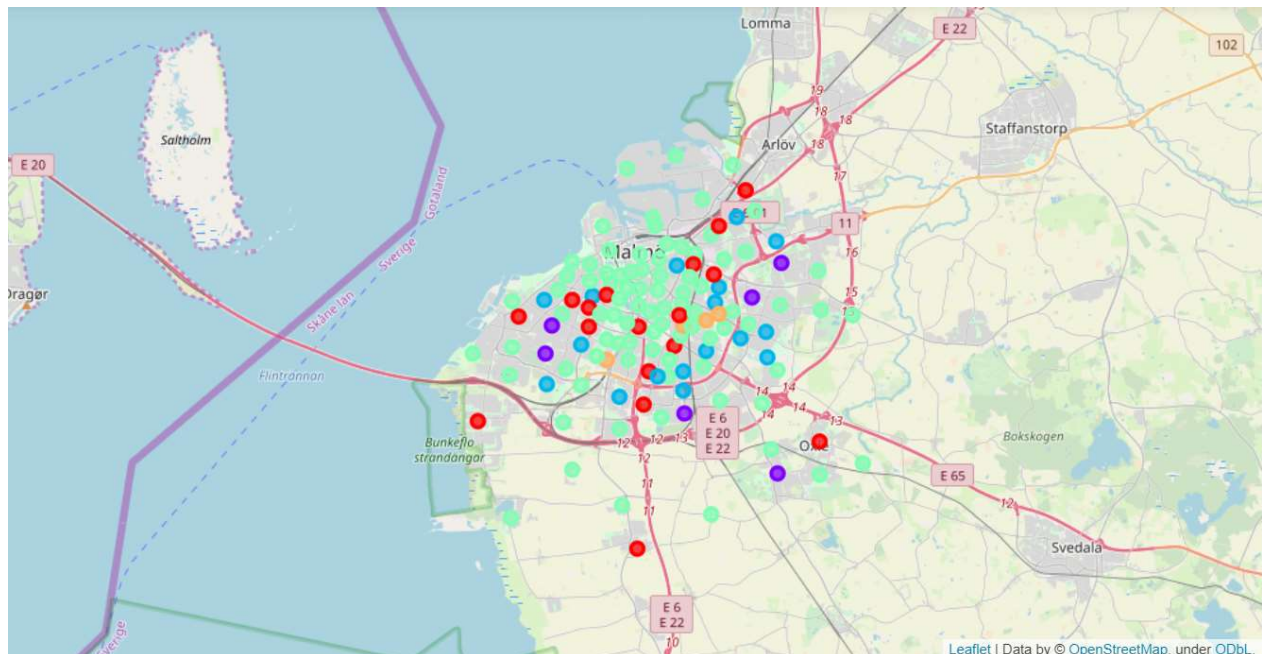
We will be asking the algorithm to group the data in 5 clusters.

With these clusters we can look at their components in terms of neighborhoods and boroughs and see how broad and narrow these clusters are, this will give an indication of how well distributed and broad are the neighborhoods and boroughs

Lets check how the boroughs are distributed by the k-means cluster, last column ("num_borough") has the sum of unique boroughs

	count	borough	num_borough
Cluster Labels			
0	17	[Centrum, Fosie, Limhamn-Bunkeflo, Södra Inner...	8
1	6	[Fosie, Limhamn-Bunkeflo, Husie, Oxie]	4
2	17	[Centrum, Fosie, Limhamn-Bunkeflo, Västra Inne...	8
3	84	[Centrum, Fosie, Limhamn-Bunkeflo, Södra Inner...	10
4	4	[Södra Innerstaden, Hyllie, Rosengård]	3

We can display these clusters on the map:



Let's look at the content of cluster 0:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	Katrinelund	0	Martial Arts Dojo	Gym / Fitness Center	Wine Bar	Eastern European Restaurant	Flower Shop
20	Almhög	0	Pizza Place	Gym / Fitness Center	Grocery Store	Bus Stop	Stadium
21	Almvik, Malmö	0	Grocery Store	Wine Bar	Doner Restaurant	Flea Market	Fish Market
35	Västra Söderkulla	0	Print Shop	Grocery Store	Doner Restaurant	Flea Market	Fish Market
39	Bunkeflostrand	0	Soccer Field	Gym / Fitness Center	Pub	Grocery Store	Bus Station
42	Gamla Limhamn	0	Grocery Store	Liquor Store	Pharmacy	Coffee Shop	Sporting Goods Shop
52	Tygelsjö	0	Candy Store	Food Truck	Grocery Store	Wine Bar	Flea Market
58	Flensburg, Malmö	0	Coffee Shop	Fast Food Restaurant	Currency Exchange	Bus Stop	Grocery Store

as per above cluster 0 seems to be broad by venue types

Cluster 1:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
32	Kastanjegården	1	Playground	Gym / Fitness Center	Wine Bar	Doner Restaurant	Flea Market
40	Djupadal	1	Playground	Park	Camera Store	Flower Shop	Wine Bar
49	Rosenvång	1	Playground	Café	Wine Bar	Eastern European Restaurant	Flea Market
112	Riseberga, Malmö	1	Playground	Wine Bar	Eastern European Restaurant	Flower Shop	Flea Market
116	Videdal	1	Playground	Wine Bar	Eastern European Restaurant	Flower Shop	Flea Market
131	Käglinge	1	Playground	Flower Shop	Wine Bar	Eastern European Restaurant	Flea Market

cluster 1 seems to offer playgrounds so could be good boroughs for families with small children

Cluster 2:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
16	Värnhem	2	Bus Stop	Hot Dog Joint	Garden	Wine Bar	Eastern European Restaurant
24	Fosieby	2	Bus Stop	Music Venue	Wine Bar	Food	Flea Market
28	Gullviksborg	2	Bus Stop	Wine Bar	Currency Exchange	Flower Shop	Flea Market
31	Hindby	2	Bus Stop	Fast Food Restaurant	Grocery Store	Restaurant	Wine Bar
36	Östra Söderkulla	2	Bus Stop	Stadium	Wine Bar	Electronics Store	Flower Shop
38	Bellevue, Malmö	2	Bus Stop	Gym	Golf Course	Wine Bar	Eastern European Restaurant
41	Elinelund	2	Bus Stop	Grocery Store	Wine Bar	Eastern European Restaurant	Flea Market
67	Dammfri	2	Bus Stop	Scandinavian Restaurant	Falafel Restaurant	Food	Flea Market

Cluster 2, seems to be focused on transportation and it covers the peripheral areas of Malmö

Cluster 3:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Davidshall	3	Coffee Shop	Scandinavian Restaurant	Theater	Café	Bike Shop
1	Ellstorp	3	Water Park	Currency Exchange	Food	Flower Shop	Flea Market
2	Frihamnen, Malmö	3	Scandinavian Restaurant	Antique Shop	Harbor / Marina	Wine Bar	Flea Market
3	Gamla staden, Malmö	3	Hotel	Wine Bar	Coffee Shop	Scandinavian Restaurant	Pizza Place
4	Inre hamnen	3	Thrift / Vintage Store	Wine Bar	Doner Restaurant	Flower Shop	Flea Market
6	Lugnet, Malmö	3	Café	Indian Restaurant	Thai Restaurant	Bakery	Falafel Restaurant

Cluster 3 seems to be focused on food venues, maybe interesting for single people who like to eat out

Cluster 4:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
59	Lönngården	4	Food & Drink Shop	Convenience Store	Hawaiian Restaurant	Doner Restaurant	Flea Market
84	Holma, Malmö	4	Convenience Store	Currency Exchange	Flower Shop	Flea Market	Fish Market
100	Rosengård Centrum	4	Falafel Restaurant	Gas Station	Wine Bar	Food	Flea Market
103	Örtagården, Malmö	4	Convenience Store	Bus Stop	Falafel Restaurant	Wine Bar	Eastern European Restaurant

cluster 4 seems to be focused on peripheral areas and food related venues

4. Results

After doing our data exploration we can now advise single persons and families with children to which boroughs of Malmö offer which type of venues and do so at a reasonable price:

Centrum borough seems to be good value in terms of restaurants and also playgrounds, for finding supermarkets nearby value for money can be found in Hyllie as well.

The Centrum borough seems to offer a broad variety of venues having plenty of playgrounds which is unusual for a city centrum

5. Discussion

The main concern with this study is its reliance on the foursquare data, which as we discussed before seems to missing data points for example bus stations, churches, discos.

Foursquare number of venues also changes on a daily basis so reproducibility of this study is an issue.

Most of the foursquare data is focused on restaurants and places to go out in the city centrum, probably tourists to Malmö make many of the updates on the site, so it is more centered on touristic relevant info.

For a more unbiased and accurate study more data sources will be needed with regards to venue details

6. Conclusion

So we scrapped the neighborhoods and boroughs for Malmö from Wikipedia, we got the apartment sales prices and we got the venue details from foursquare API.

While doing data cleansing and preparation and analysis it became clear that foursquare doesn't hold a complete set of data for all categories. For a more serious analysis of the venues in Malmö a much larger effort in data cleansing and data source quality would be required.

We then did some data exploration to see what are the most common values in each category and displayed these as bar charts and heatmaps

We also did an analysis between the percentage of venues in a given category versus the average square meter price, this analysis could help determine if a neighborhood offers value per money for a potential buyer. For this we used linear regression as a tool

We looked at the perspective of value for different stakeholders: children with small families, looking for playground availability and relatively cheap price of square meters, could try the "Centrum" A single person looking to go out for food instead of cooking, could also look at "Centrum" as an area of value for money People that are interested in the convenience of having shops nearby, and have a relatively cheap price per square meter could instead look at "Hyllie"

Finally, we looked at clustering neighborhoods by the similarity of venues, using k-means. We found the two largest clusters are well diversified in terms of venues, which is good it shows that the planning of the city is working so that it offers a balanced and varied set of venues