

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 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	<ul style="list-style-type: none"> -Creating a neighborhood address -in some 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	<p>Checking for neighborhoods with no venues</p> <p>Checking for duplicate venues since each neighborhood had a radius of 500 meters the same venues would overlap to multiple neighborhoods</p> <p>The rule to clean duplicates was to take was the keep the first one only. Although in a more thorough exercise a more careful approach would have to be done</p> <p>While doing data preparation it become clear that there are venues missing from the foursquare API database: such as churches, bus stations, etc...</p>

	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>

	busstation default foodgrocery gym playground cafe park scandinavian pizza coffeeshop falafel plaza bakery stadiumsoccer italian sportsoutdoors fastfood hotel furniture asian
Maklarstatistik.se	For better readability on charts we create a feature of log price of square meters in SEK (Swedish kronors)

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