

Deciding where to move in Helsinki

Introduction

Helsinki is a small but diverse city. With 650 000 inhabitants, the capital of Finland is also the most populous city in Finland. Its districts are significantly different from each other and there is a district for every flavor.

When people decide where to move, they often search the internet for information about the areas they are interested in. The features of the area or district can be an important criterion when deciding on a location. For example, outgoing people might want cafeterias, bars and restaurants nearby, while families might want parks and recreational centers near their home.

However, researching the surroundings can be time-consuming to do by hand. There are several factors and an abundance of information to take into account. My aim is to cluster the districts of Helsinki based on the most popular venues in that district and compare the obtained clusters. Based on the clusters and the comparison, people can include district information in their decisions about where to move in Helsinki.

Data

The data are from the Helsinki Region Infoshare (HRI) service¹, whose aims are to make regional information quickly and easily accessible to all. The data are in two parts. The first part consists of all the addresses, postal codes and coordinates in Helsinki. The second part consists of postal codes and postal area names. These data were merged into one datafile with addresses, postal codes, postal area names and the coordinates related to the postal areas (districts). The available coordinates were in the ETRS-GK25 coordinate system, based on the Finnish version EUREF-FIN of the European ETRS89 coordinate system. However, Foursquare does not recognize this system so I had to install the pyproj package and import Proj and transform from that package. With it, I was able to obtain the corresponding WSG84 coordinates.

See Table 1 for a snapshot of the final data. There were a total of 84 neighbourhoods in the data.

Table 1. Five first neighbourhoods in the data with corresponding WSG84 coordinates.

	Area	newLon	newLat
0	Pajamäki	24.862298	60.218664
1	Kivihaka	24.903746	60.209473
2	Meilahden sairaala-alue	24.910364	60.189822
3	Herttoniemi	25.024531	60.189653
4	Ruskeasuo	24.901674	60.202758

Methods and results

After the datasets were merged and cleaned, and the coordinates transformed to the correct form, I ran some exploratory analyses. First, I checked the coordinates of Helsinki with the Nominatim package. The coordinates are 60.167 (latitude) and 24.943 (longitude).

Next, I contacted Foursquare and did a test with a random neighbourhood. Neighbourhood at index 33 was "Taka-Töölö" and a snapshot of the venues associated with that neighbourhood are shown in Table 2.

Table 2. Five first venues associated with neighbourhood “Taka-Töölö”.

	name	categories	lat	lng
0	Daddy Greens Pizzabar	Pizza Place	60.184982	24.920734
1	Korjaamo Summer Garden	Beer Garden	60.184098	24.920041
2	Korjaamo Culture Factory	Event Space	60.184097	24.919950
3	Korjaamo Kino	Indie Movie Theater	60.184262	24.919897
4	Linko	Pizza Place	60.181740	24.925074

Next, I built a loop to retrieve the venues for all 84 neighbourhoods. There were 1556 rows of data and 240 unique venue categories in the obtained dataframe. To check if I am able to pick top venues from the data, I then chose the 5 top venues in each neighbourhood based on frequency. After that worked, I picked the top 10 venues in each neighbourhood into a dataframe. A snapshot is shown in Table 3.

Table 3. Five first rows of the dataframe with top 10 most common venues for each neighbourhood in Helsinki.

	Area	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Aurinkolahti	Beach	Grocery Store	Park	Playground	Bridge	Bus Stop	Canal	Gym / Fitness Center	Sri Lankan Restaurant	Restaurant
1	Eira - Hernesaari	Park	Scandinavian Restaurant	French Restaurant	Bakery	Pizza Place	Ice Cream Shop	Italian Restaurant	Boat or Ferry	Beach	Beer Bar
2	Etelä-Haaga	Bus Stop	Café	Supermarket	Sushi Restaurant	Pizza Place	Grocery Store	Zoo	Food	Flower Shop	Flea Market
3	Etelä-Lajasalo	Canal	Convenience Store	Bus Stop	Café	Zoo	Forest	Food & Drink Shop	Food	Flower Shop	Flea Market
4	Etelä-Vuosaari	Port	ATM	Business Service	Event Space	Forest	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish Market

Finally, I used K-means to cluster the data. K-means clusters the data based on similar characteristics. Because I have the top 10 most common venues, clustering the data by hand would be almost impossible. K-means finds the similarities and gives you the clusters, which you can then name yourself.

After testing different numbers of clusters, I concluded that six clusters gives the best result. The clusters and the neighbourhoods are shown in Table 4.

Table 4. Clusters of neighbourhoods based on similarity in venue frequency.

Cluster 1: Close to bus stops and ethnic foods	Cluster 2: Bars, restaurants and coffee shops	Cluster 3: Quiet neighbourhood	Cluster 4: Outdoor activities and food	Cluster 5: Close to a lot of activities, foods and bus stops; busy	Cluster 6: Markets and foods
Puotila	Meilahden sairaala-alue	Myllypuro	Itäsalmi	Pajamäki	Ilmala
Veräjämäki	Herttoniemi			Kivihaka	
Tuomarinkylä-Torppar	Ruskeasuo			Koskela-Helsinki	
Itä-Pakila	Pohjois-Meilahti			Lassila	
Metsäla-Etelä-Oulunkylä	Suomenlinna			Pohjois-Haaga	
Vartiokharju	Santahamina			Kulosaari	
	Pikku Huopalahti			Roihuvuori	
	Kaivopuisto - Ullanlinna			Viikki	
	Kalasatama			Roihupellon teollisuusalue	
	Pohjois-Vuosaari			Pihlajamäki	
	Länsi-Pasila			Maunanneva	
	Vattuniemi			Tapaninvainio	
	Vallila			Mellunmäki	
	Itä-Pasila			Konala	
	Kallio			Kontula - Vesala	
	Toukola-Vanhakaupunki			Paloheinä	
	Sörnäinen			Länsi-Pakila	
	Pitäjänmäen teollisuusalue			Tammisalo	
	Siltamäki			Länsi-Herttoniemi	
	Jätkäsaari			Etelä-Haaga	
	Eira - Hernesaari			Etelä-Laaajasalo	
	Kannelmäki			Jakomäki - Alppikylä	
	Taka-Töölö			Maunula-Suursuo	
	Munkkivuori-Niemenmäki			Etelä-Vuosaari	
	Kuusisaari-Lehtisaari				
	Kamppi - Ruoholahti				
	Laaajasalo				
	Kruununkhaka				
	Kaitalahti				
	Munkkiniemi				
	Malminkartano				
	Punavuori				
	Itäkeskus-Marjaniemi				
	Helsinki Keskusta - Etu-Töölö				
	Jollas				
	Aurinkolahti				
	Puotinkharju				
	Kaartinkaupunki				
	Puistola				
	Pukimäki-Savela				
	Verkkosaari				
	Oulunkylä-Patola				
	Lauttasaari				
	Etu-Vallila - Alppila				
	Katajanokka				
	Suurmetsä				
	Malmi				
	Reimaria				
	Käpylä				
	Keski-Töölö				
	Tapanila				

Discussion

The results of the K-means clustering showed that there are many neighbourhoods with a lot of different activities around and then there are some more quiet neighbourhoods. The K-means clustering was not that effective in making the clusters manageable and some of the obtained clusters contained only one neighbourhood. However, based on the results, some recommendations can be made. For people who want lots people and happening around them, there are several choices (clusters 2 and 5 in Table 4). I did not include housing prices in this analysis but based on my knowledge of the prices for apartments in Helsinki, the two “busy neighbourhood” clusters are different in that regard. People could therefore choose, if they want to live in the city center (expensive) close to services and restaurants (cluster 2 in Table 4), or if they want to live a bit outside the city center (cheaper) but still be close to activities and restaurants (cluster 5 in Table 4). More information about the venues in the clusters can be found in the Notebook².

Conclusions

There are differences in venue popularity in the neighbourhoods in Helsinki. People who are figuring out where to move in Helsinki can use the clusters I created to examine the popularity of venues in different neighbourhoods. They can then combine this information with housing prices, information about schools and other relevant information to make an informed decision about where to move.

References

1. Helsinki region inforshare. https://hri.fi/en_gb/
2. https://github.com/mariatornroos/Coursera_Capstone/blob/master/Capstone%20project_Helsinki.ipynb