Price of living in Helsinki, Finland – Neighborhood characteristics or geographical location?

Emil Kronberg, 2021

1 Introduction

The price of living in Helsinki, the capital of Finland, is steadily increasing both for renting and buying apartments. At a population of approximately 650 000 people (1 200 000 if counting the neighboring cities of Espoo and Vantaa), Helsinki is the largest city in the country and is steadily growing in population. While many people move from other areas of Finland to the capital to work, the apartment prices may sometimes be double of what they paid for a similarly sized apartment back home. Prices in Helsinki, however, may vary depending on location and neighborhood.

This research will explore if there are differences between the price and location depending on the characteristics of the neighborhood, or if the price of living is solely tied to the geographic location of the neighborhood. The research is aimed towards giving insight of the differences in prices for those moving or planning to move to Helsinki.

2 Description of data

The data that will be used in the research are location data of each neighborhood, such as coordinates, name, and postal code. Other neighborhood specific data include average price (€) per square meter (for buying homes). Finding the characteristics of the neighborhood requires more specific data of the areas, such as information about venues in the neighborhood.

2.1 Neighborhoods and price

The average price data was extracted from the website of Real estate firm 'Blok' [1]. This also included the name and postal code of each neighborhood. Using the postal codes, I managed to extract the coordinates for each neighborhood using OpenStreetMap data and Nominatim in python. Some coordinates that were completely wrong were manually corrected using Google maps [3].

2.2 Foursquare

The venue data was extracted using Foursquare API [2]. The data included up to 100 venues within 500 meters of each of the neighborhood coordinates. It may be worth mentioning that this method can be somewhat unreliable since

all venues may not necessarily be registered and available through Foursquare, but it does offer a good overview of the surrounding areas.

2.3 Postal code areas

The boundaries of each postal code area were extracted through HSY (Environmental services of the region of Helsinki) [4].

3 Methodology

After the initial data extraction of neighborhood data, each neighborhood was mapped using the folium library in python. This ensured that the neighborhoods both were in the correct location, as well as offering an overview of the spread of the neighborhoods. The spread was the deciding factor to use a radius of 500 meters, since while some neighborhoods are closer together, there are others that have a lot more distance between them. Figure 1 shows the initial mapping after extracting the neighborhood data.

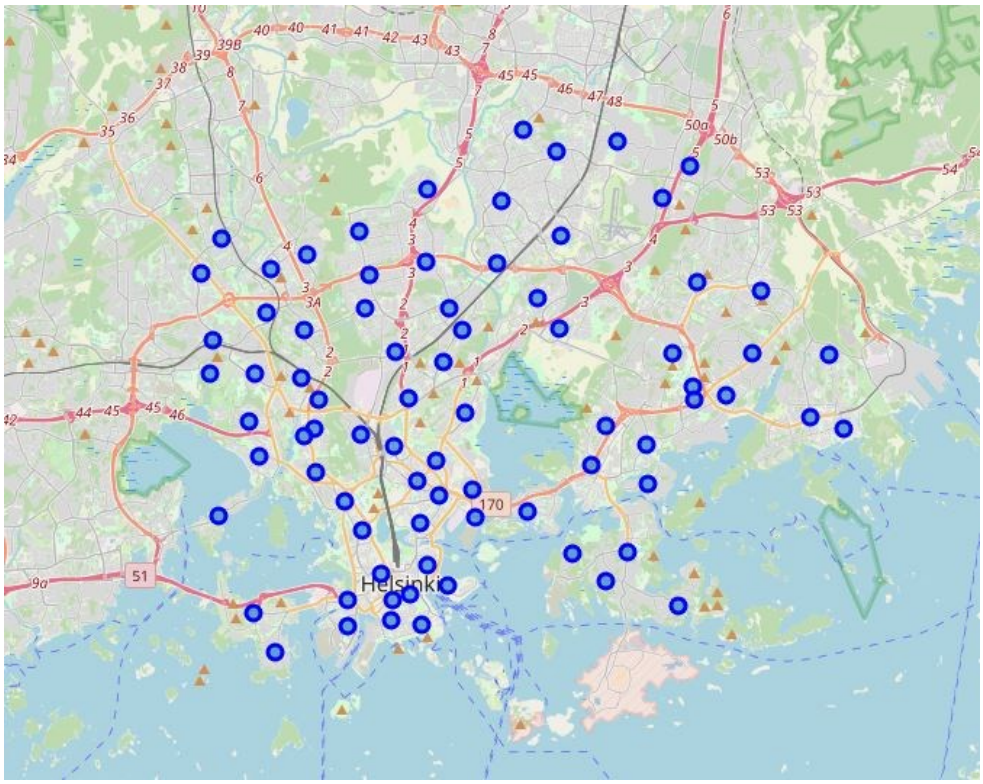

Figure 1 Initial mapping of neighborhoods

Table 1 represents the first few rows of the venue data extracted using Foursquare API. There were a total of 1841 venues across the 78 neighborhoods extracted for this research. It is worth noting that the number of neighborhoods used for the research may be somewhat lower that the actual number of neighborhoods in Helsinki, as will be visible later on. This is due to not having complete data of a few postal codes. This should not however affect the outcome of the research.

Table 1 Initial venue extraction

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Kaivopuisto - Ullanlinna	60.157935	24.952702	Kaivopuisto / Brunnsparken (Kaivopuisto)	60.156810	24.956732	Park
1	Kaivopuisto - Ullanlinna	60.157935	24.952702	Rams Roasters	60.158002	24.948908	Coffee Roaster
2	Kaivopuisto - Ullanlinna	60.157935	24.952702	Helsingin Jäätelötehdas	60.155643	24.950891	Ice Cream Shop
3	Kaivopuisto - Ullanlinna	60.157935	24.952702	Paulette	60.158475	24.946111	French Restaurant
4	Kaivopuisto - Ullanlinna	60.157935	24.952702	Cafe Compass	60.155131	24.951614	Coffee Shop

The venue data was grouped with their respective neighborhood and transformed into a table of dummy variables, followed by giving each variable the value of frequency for every neighborhood as seen in Table 2. This transformation allowed further statistical analysis to be performed on the dataset.

Table 2 Venue data with dummies

	Neighborhood	ATM	Accessories Store	American Restaurant	Antique Shop	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auditorium	Auto Dealership	Auto Garage	Auto Workshop	Automotive Shop	BBQ Joint	Badminton Court	Bagel Shop	Bakery	Bar	Baseball Field
0	Aurinkolahti	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.0
1	Eira - Hernesaari	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.022727	0.068182	0.0	0.0
2	Etelä-Haaga	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.0
3	Etelä-Laajasalo	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.0
4	Etelä-Vuosaari	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.040000	0.0	0.0

I also created a table visualizing the top 10 most popular sort of venues per neighborhood (as seen in Table 3) which was used at a later stage for giving a better description of the differences between the neighborhoods.

Table 3 Top 10 venues per neighborhood

	Neighborhood	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	Grocery Store	Ice Cream Shop	Restaurant	Canal	Sri Lankan Restaurant	Gym / Fitness Center	Beer Bar	Coffee Shop	Theater	Beach
1	Eira - Hernesaari	Coffee Shop	Pizza Place	Park	Bakery	Scandinavian Restaurant	Café	Turkish Restaurant	French Restaurant	Modern European Restaurant	Beer Garden
2	Etelä-Haaga	Café	Shoe Repair	Supermarket	Sushi Restaurant	Bus Stop	Park	Pizza Place	Zoo	Filipino Restaurant	Film Studio
3	Etelä-Laajasalo	Bus Stop	Playground	Zoo	Flower Shop	Filipino Restaurant	Film Studio	Fish & Chips Shop	Fish Market	Flea Market	Food
4	Etelä-Vuosaari	Bus Stop	Discount Store	Park	Health Food Store	Gym / Fitness Center	Gym	Grocery Store	Men's Store	Flower Shop	Bike Shop
5	Etu-Vallila - Alppila	Bar	Dive Bar	Park	Restaurant	Beer Garden	Café	Thai Restaurant	Gym	Beer Bar	Sushi Restaurant
6	Helsinki Keskusta - Etu-Töölö	Scandinavian Restaurant	Wine Bar	Art Museum	Clothing Store	Food Court	Sushi Restaurant	Beer Bar	Beer Garden	Tea Room	Middle Eastern Restaurant
7	Herttoniemi	Bus Stop	Gym / Fitness Center	Supermarket	Pizza Place	Convenience Store	Chinese Restaurant	Café	Gastropub	Scandinavian Restaurant	Kebab Restaurant
8	Itä-Pakila	Bus Stop	Grocery Store	Zoo	Food & Drink Shop	Film Studio	Fish & Chips Shop	Fish Market	Flea Market	Flower Shop	Food
9	Itä-Pasila	Tram Station	Bus Stop	Park	Italian Restaurant	Gym / Fitness Center	Gym	Restaurant	Bar	Thai Restaurant	Pool

3.1 K-means

To analyze the differences in neighborhood I used unsupervised learning in the form of K-means clustering, to sort the neighborhoods into different clusters depending on their characteristics of popular venues, using the data visualized in Table 2. To find the optimal number of K, or clusters, I used the Elbow Method. As can be seen in Figure 2, the optimal number of clusters according to the Elbow method was 4.

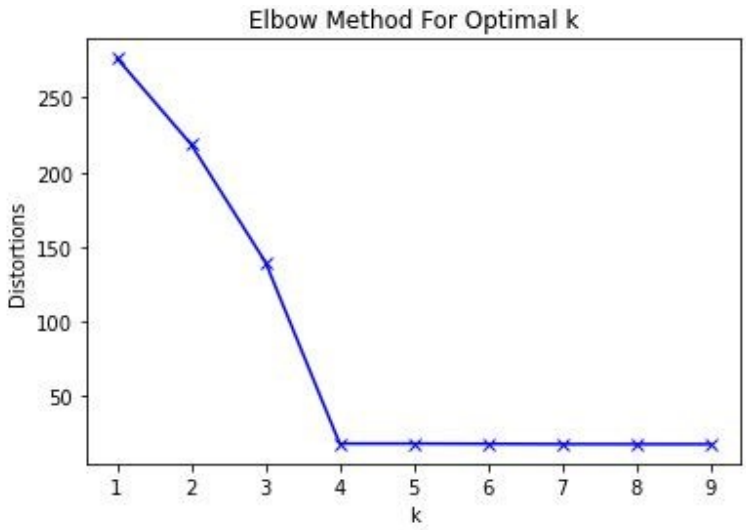

Figure 2 Elbow method

I double-checked the silhouette score for optimal K as well and found that either 2 or 4 would work. I decided to stay consistent and use K=4. The silhouette score can be seen in Figure 3.

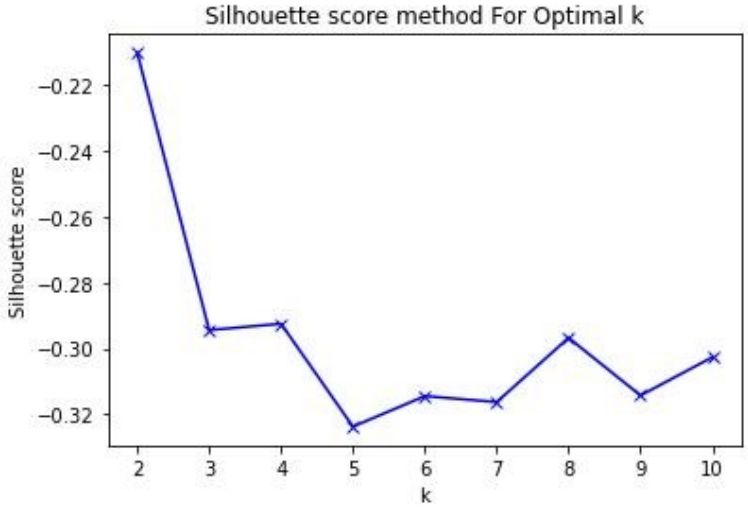

Figure 3 Silhouette score method

I merged the cluster label for each data entry, as well as the most common venues to create one complete data set for each neighborhood, as seen in Table 4.

Table 4 Complete data of neighborhood

	Postcode	Neighborhood	Average price	Latitude	Longitude	Cluster Labels	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	00140	Kaivopuisto - Ullanlinna	8713	60.157935	24.952702	0	Coffee Shop	Grocery Store	Park	Ice Cream Shop	Scandinavian Restaurant	Recreation Center	Beer Garden	Thai Restaurant	Nightclub	Pharmacy
1	00150	Eira - Hernesaari	8367	60.158939	24.938014	0	Coffee Shop	Pizza Place	Park	Bakery	Scandinavian Restaurant	Café	Turkish Restaurant	French Restaurant	Modern European Restaurant	Beer Garden
2	00120	Punavuori	8160	60.163562	24.939202	2	Scandinavian Restaurant	Beer Bar	Bar	Hotel	Cocktail Bar	Pizza Place	Bakery	Coffee Shop	Sushi Restaurant	Park
3	00180	Kamppi - Ruoholahti	8023	60.163576	24.917557	1	Restaurant	Gym	Hotel	Beer Bar	Grocery Store	Liquor Store	Scandinavian Restaurant	Escape Room	Sauna / Steam Room	Park
4	00220	Jätkäsaari	7871	60.157433	24.917381	0	Bar	Electronics Store	Grocery Store	BBQ Joint	Food & Drink Shop	Canal	Pool	Supermarket	Falafel Restaurant	Boat or Ferry
5	00170	Kruununhaka	7841	60.171866	24.955321	0	Café	Boat or Ferry	Pizza Place	Bar	Scandinavian Restaurant	Grocery Store	Theater	History Museum	Coffee Shop	Indie Movie Theater
6	00130	Kaartinkaupunki	7825	60.165009	24.947547	0	Hotel	Scandinavian Restaurant	Bar	Park	Cocktail Bar	Pizza Place	Coffee Shop	Café	Beer Bar	Modern European Restaurant
7	00580	Verkkosaari	7609	60.189562	24.976701	2	Gym / Fitness Center	Supermarket	Event Space	Food Court	Italian Restaurant	Scandinavian Restaurant	Burger Joint	Bar	Café	Grocery Store
8	00100	Helsinki Keskusta - Etu-Töölö	7575	60.169989	24.933727	1	Scandinavian Restaurant	Wine Bar	Art Museum	Clothing Store	Food Court	Sushi Restaurant	Beer Bar	Beer Garden	Tea Room	Middle Eastern Restaurant
9	00260	Keski-Töölö	7384	60.180014	24.924237	0	Sushi Restaurant	Scandinavian Restaurant	Park	Hotel	French Restaurant	Coffee Shop	Italian Restaurant	Bakery	Plaza	Gym / Fitness Center

With the cluster labels for each neighborhood, I returned to the previously created map but now included the cluster labels as colors of each marked neighborhood. This is visualized in Figure 4, each color corresponding to a cluster:

- **Cluster 0: Yellow**
- **Cluster 1: Red**
- **Cluster 2: Green**
- **Cluster 3: Blue**

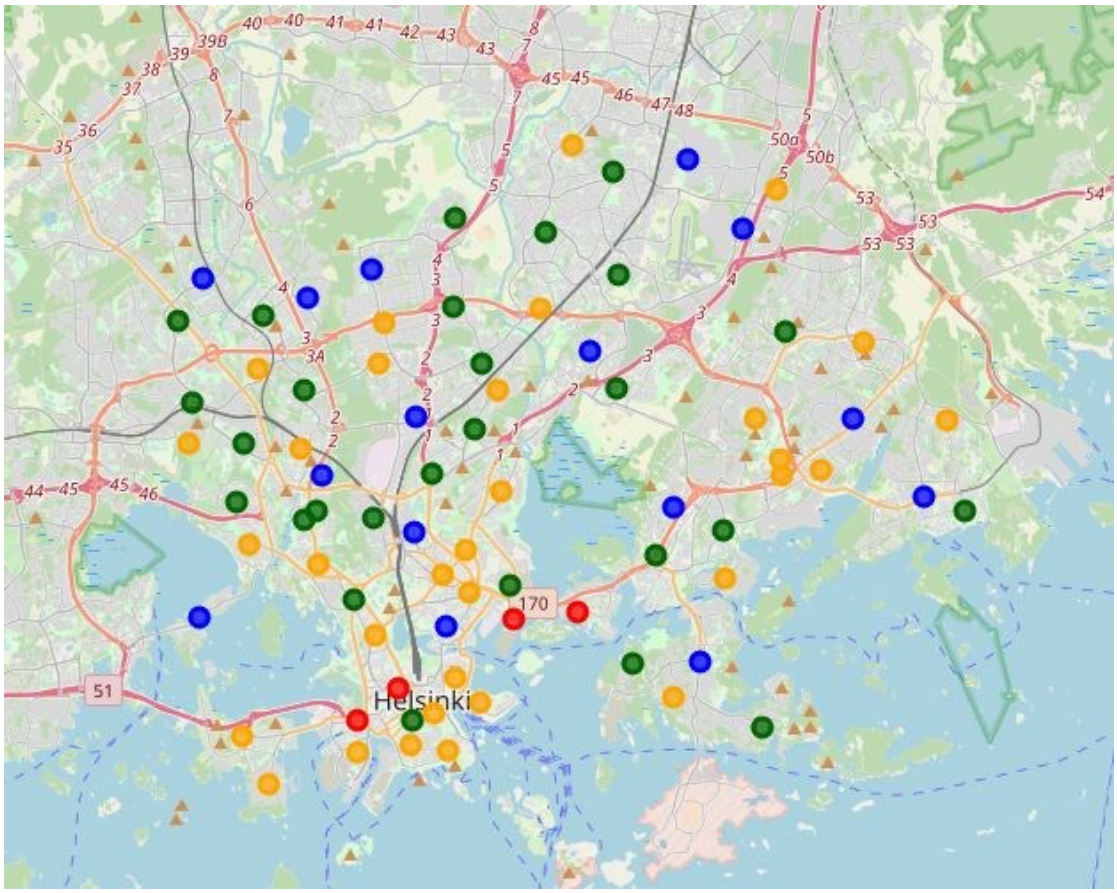

Figure 4 Map with clusters

As we can see, there does not seem to be a significant amount of geographical effect on the clusters, except that Cluster 0 & 1 is somewhat more dominant in downtown Helsinki, and Cluster 3 is further away from the city center.

I took a more detailed look at each cluster by creating a bar graph representing the number 1 most common venues in each cluster, the number of neighborhoods that have that kind of venue as its number one, as well as what kind of venue it is. The initial bar graph can be seen I Figure 5. Since there were a wide variety of venues both the legend, the colors and the bars are difficult to interpret.

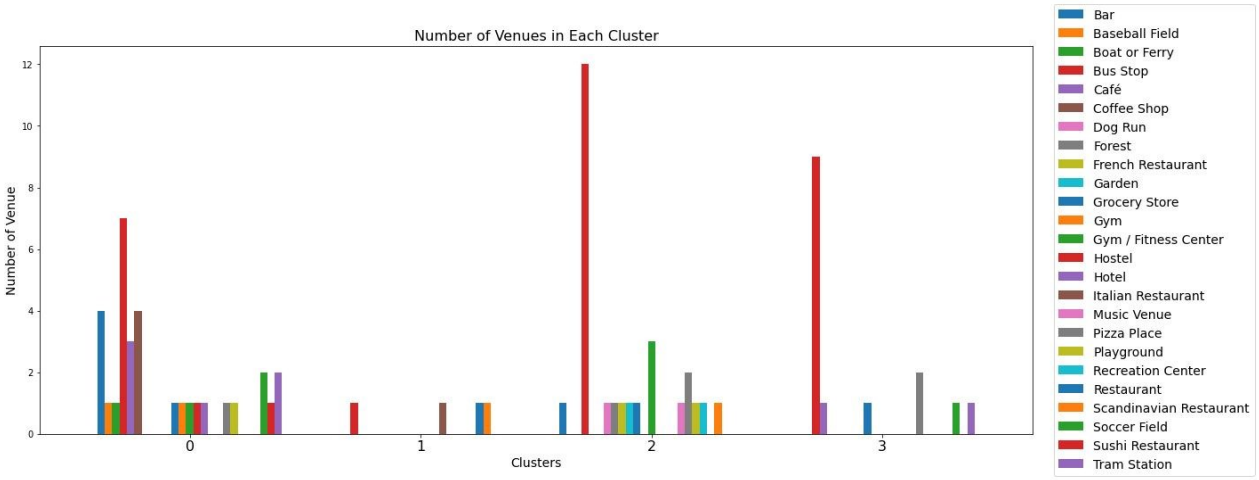

Figure 5 Bar chart with every venue

To make the bar chart more readable and more descriptive, I merged certain venues into sub-categories. The specifics of this merger can be seen in the code. As seen in Figure 6, we now have a better picture of the differences in the clusters.

Cluster 0 both has the most amount and the most variations of venues. Cafés, bars and hotels can also be found here more frequently compared to other clusters.

Cluster 1 clearly has the least number of registered venues. This was also a fairly small cluster, but still pretty central geographically which indicates more downtown/urban residential neighborhoods.

Cluster 2 also has a fair number of venues but is dominated by transport stations and outdoor areas compared to the other clusters. The transport stations are mainly bus stops, as can also be seen in Figure 5. Geographically, many neighborhoods in this cluster are located further away from the city center, which does explain both of these observations, yet interestingly the trendy and central district of Punavuori in the center of Helsinki is also included in this cluster.

Cluster 3 has less variation and fewer venues than cluster 2, while also mainly consisting of neighborhoods located further away from central Helsinki. The closest neighborhood to downtown Helsinki is Kallio to the north-north-east. Transportation venues, mainly bus stops, are also dominant in these neighborhoods.

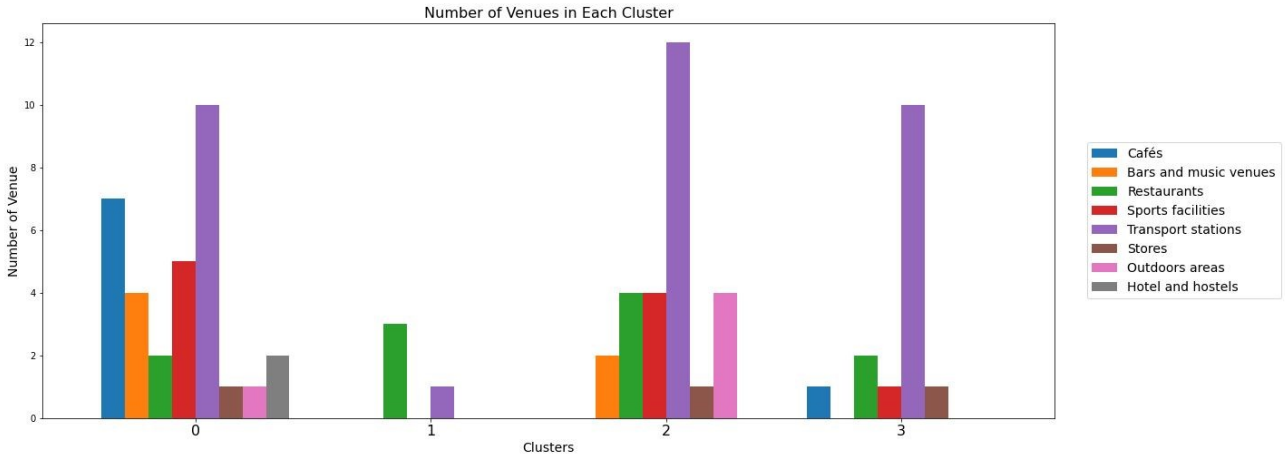

Figure 6 Bar chart with merged venues

It is no surprise that transportation tops most neighborhoods in number of venues, since public transport is both common and effective in most cities in Finland. The ratio of transport stations to other venues does however give us some insight into the differences between the clusters. The clusters are still not descriptive enough for accurate analysis. We can label the general characteristics of each cluster as the following:

- **Cluster 0**: Populous, residential, and social venues
- **Cluster 1**: Central and residential
- **Cluster 2**: Populous and outlying
- **Cluster 3**: Residential and outlying

3.2 Geographical position and average price

If we take a look at the differences in prices as seen in Table 5, we can note that there is some variety between the clusters. The one that contains the by far most expensive neighborhoods in Cluster 1. The neighborhoods in this cluster are as previously mentioned located quite close to the center of Helsinki but did not stand out as having a lot of venues except for restaurants. Cluster 2 and 3 has several neighborhoods further away from the city center which may explain them having a lower average price. Cluster 2 being more expensive, but also consisting of neighborhoods with more social venues nearby. Cluster 0, being the most spread out cluster of the 4 still has a fairly high average price which could be due to the amount of central neighborhoods raising the average price of the cluster.

Table 5 Average price per cluster

Cluster Labels	Average price
0	5196.43750
1	7138.50000
2	4640.62963
3	3998.80000

Since it is fairly difficult to draw any conclusion over the average price and difference in price between the clusters, let us take a look at the geographical distribution of average price per square meter in Helsinki.

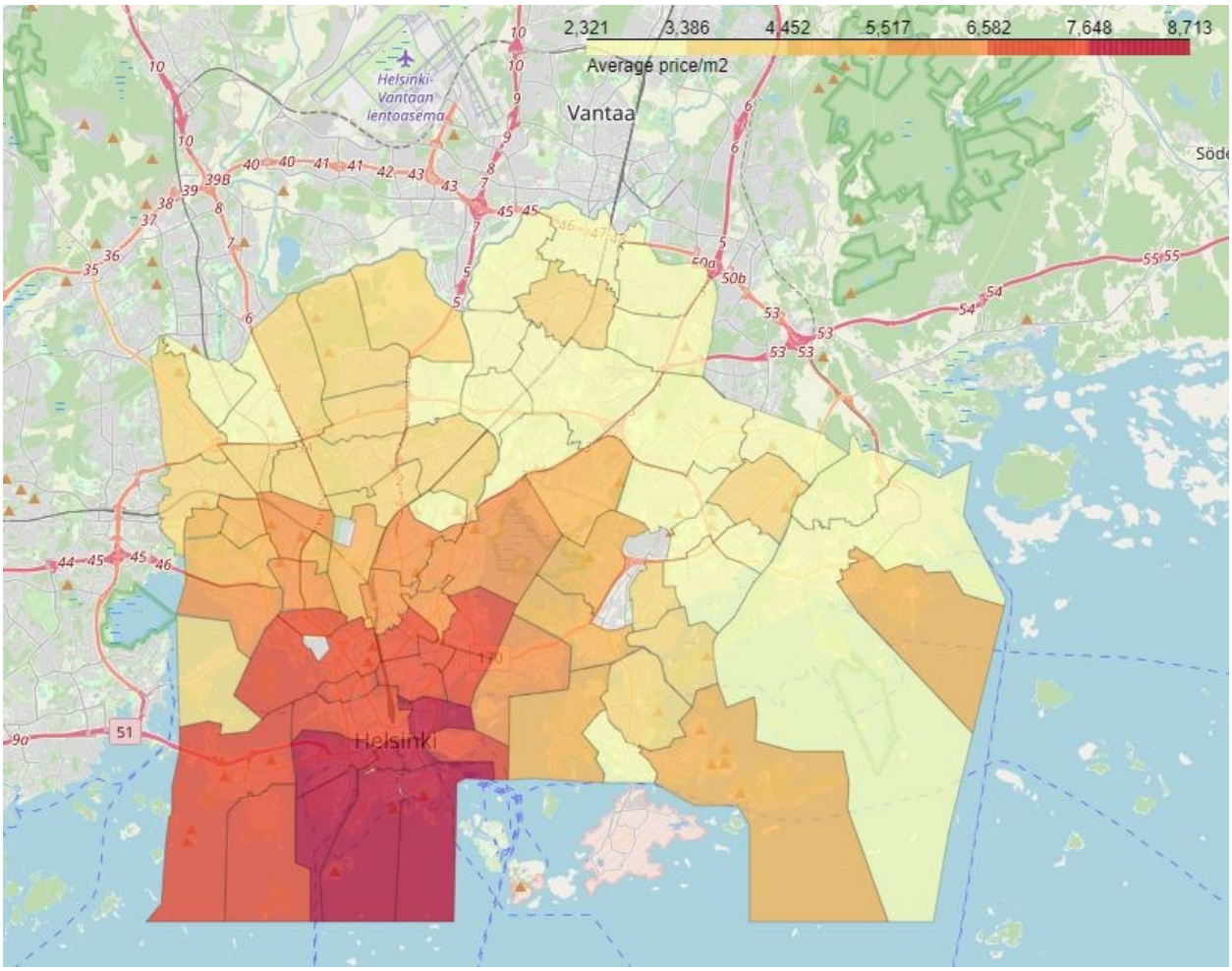

Figure 7 Average price per square meter in each neighborhood

I created a choropleth map of each neighborhood based on their average price per square meter. Note that there are three neighborhoods missing due to lack of data. As we can see in Figure 7, the price of living is very much affected by how close to the city center the neighborhood is. We can therefore note that the actual characteristics of the neighborhood may not contribute significantly to the price of living. This, however, also means that there may be neighborhoods further away from the city with similar characteristics as those more central, but which are much cheaper.

4 Results

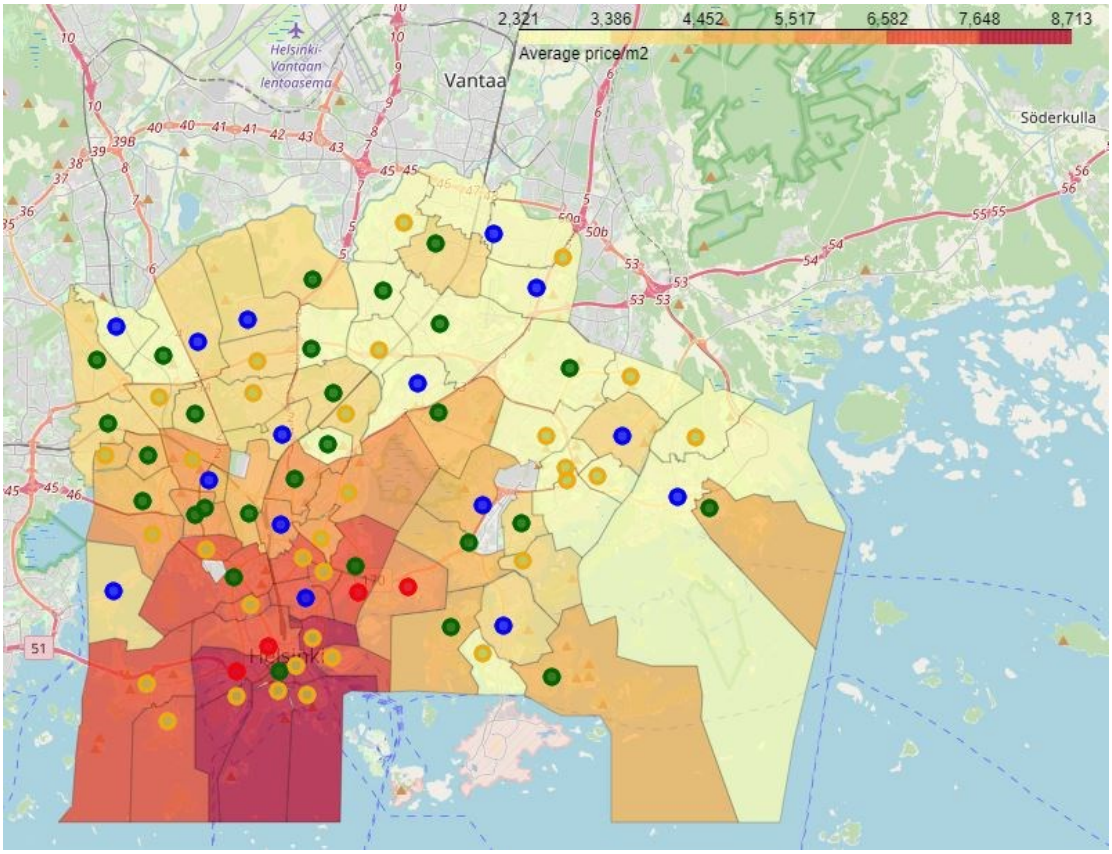

Figure 8 Merged average price and cluster map

I merged the previously created cluster and choropleth map to more accurately visualize prices and neighborhood clusters. As can bee seen in Figure 8, there are few neighborhoods that belong to clusters with a much higher average price per square meter than they themselves have. In the notebook code, there additional information, such as the name, cluster and price are available by clicking on the markers.

5 Discussion

The overall results of the study were as expected that the price is heavily tied to the geographical location, meaning if you wish to live close to the center of Helsinki, there is little choice but to pay more. Despite this fact, some observations are quite surprising, such as that Punavuori and Kallio, both which have become quite "hip" or "trendy" neighborhoods, fall into categories of neighborhoods mainly located outside of the center of Helsinki, while still being fairly expensive and central. This is something that people who know Helsinki may actually expect but is still something that I have never personally witnessed through analyzing data.

There are still some things that could be improved in this research. The distribution of venues was quite large, and the merger that was preformed for the second bar graph, could actually have been done in the beginning with all data, since many categories were practically the same but still under different names, which may have affected the results. This would, however, have been an extremely tedious task since it could only be done manually and there were 283 entries (different types of venues). If further analysis would be done, this would be more accurate since some categories are almost the same.

I also suspect that the data gathered through the Foursquare API is hardly enough to make any accurate conclusions and that there are many venues that were not actually registered. It would be a good idea to also find other variables that may help identify the characteristics of the neighborhoods, since 'Bus Stops' as a venue essentially dominated most clusters.

6 Conclusion

While data may help analyze the which part of a city both has the right characteristics as well as an acceptable price of living, it ultimately depends on what aspects one value the most. We can conclude that there are neighborhoods outside the central areas of Helsinki that shares their characteristics with the most hip, trendy and expensive ones, but while the prices may be cheaper, the distance to work may be longer. Data in this situation should, therefore, not be deterministic, but may indeed assist a newcomer with finding the perfect neighborhood.

7 References

[1] Blok (2021), House Prices, Retrieved July 15, 2021 from: https://blok.ai/en/neighbourhoods/

[2] Foursquare API, https://developer.foursquare.com/

[3] Google maps, https://www.google.com/maps

[4] HSY (2021), Retrieved July 16, 2021 from: https://kartta.hsy.fi/geoserver/wfs