



The Battle of Neighborhoods

Helsinki (Finland) vs. Tallinn (Estonia)

Prepared by: Norbert Juhász

For: Applied Data Science Capstone by IBM/Coursera

On: May 30th, 2020



Introduction / Study Problem

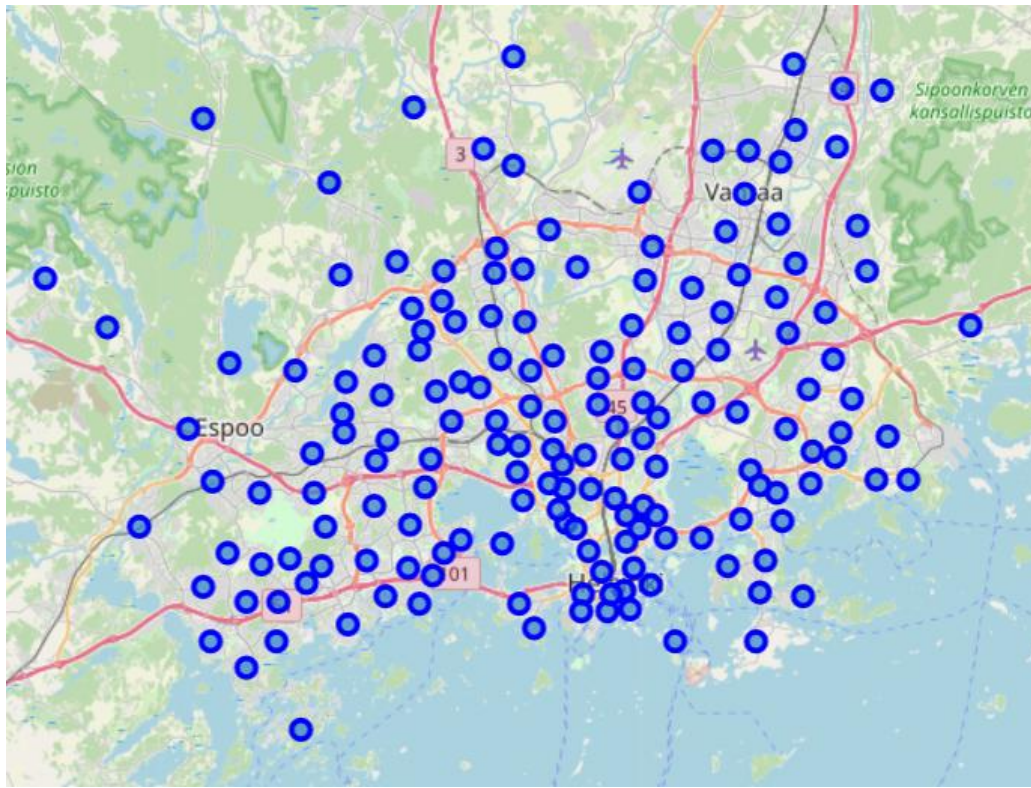
- The neighborhoods of two cities will be compared:
 - Helsinki, Finland (where I live)
 - Tallinn, Estonia (the closest other EU capital, just on the other side of the Gulf of Finland)
- Target:
 - To find clusters of similar/different neighborhoods
- Target audience:
 - To help people who consider moving from one city to the other
 - For someone who, for instance, would be looking for a similar neighborhood to live in Tallinn as compared to where they live in Helsinki, the results of my analysis may be helpful.

Data

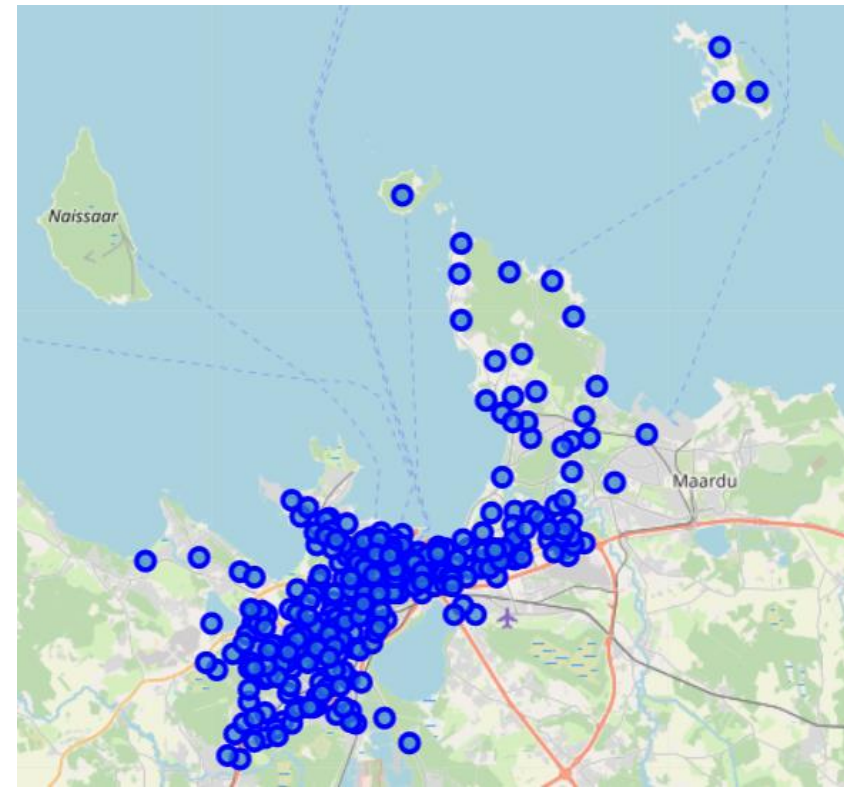
- The neighborhoods of Helsinki and Tallinn are defined based on postal codes from **Opendatasoft.com's API**
 - Dataset name: "geonames-postal-code%40public-us"
 - Data of interest: postal code, area name, latitude, longitude, administrative structure
 - **Helsinki** is defined as Greater Helsinki area, including the administrative (but not geographically) separate cities of Espoo, Kauniainen and Vantaa
 - **Tallinn** city is also extended by the region of Viimsi vald which encompasses Tallinn towards the direction of Helsinki
- The central coordinates of Helsinki and Tallinn are retrieved from **geocoders Nominatim**
 - In order to center the folium map and make both cities visible, the coordinates of the two cities are averaged
- The venues, their type and location in every neighborhood will be obtained from the **Foursquare API**

Postal codes / City definitions

- 169 postal codes for **Helsinki** region



- 310 postal codes for **Tallinn**

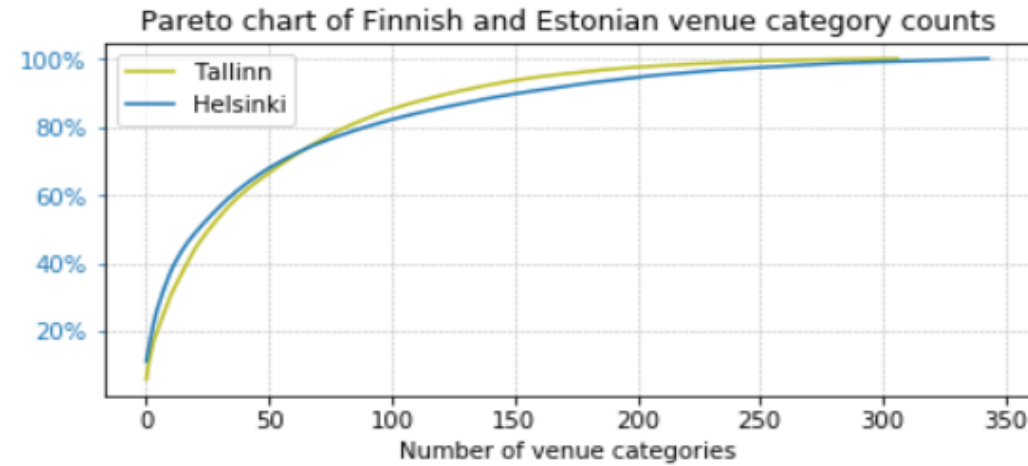


Location venue data

- The Foursquare API request resulted in 17 538 **venue records** (sample to the right):

postal_code	latitude	longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
02600	60.2122	24.8066	Mezza	60.215323	24.811755	Middle Eastern Restaurant
02600	60.2122	24.8066	Jungle Juice Bar	60.218400	24.813196	Juice Bar
02600	60.2122	24.8066	Caffi	60.218603	24.812755	Coffee Shop
02600	60.2122	24.8066	Fressi	60.216913	24.818744	Gym / Fitness Center
02600	60.2122	24.8066	ELIXIA Sello	60.218130	24.812810	Gym / Fitness Center

- 417 unique **venue categories** (Approx. 75-80 of them cover 80% of the total venue count)



Location venue data (cont'd)

- 110 venue categories are unique for Helsinki (26.4%)
- 73 venue categories are unique for Tallinn (17.5%)
- **234 venue categories shared** between the two cities (56.1%), representing a 90.5% share of the total venue count

country_code	EE	FI
Venue Category		
Himalayan Restaurant	NaN	38.0
Waterfront	NaN	17.0
Taxi Stand	NaN	17.0
Platform	NaN	15.0
Karaoke Bar	NaN	13.0
...
Molecular Gastronomy Restaurant	NaN	1.0
Outdoor Gym	NaN	1.0
Canal Lock	NaN	1.0
Carpet Store	NaN	1.0
Road	NaN	1.0

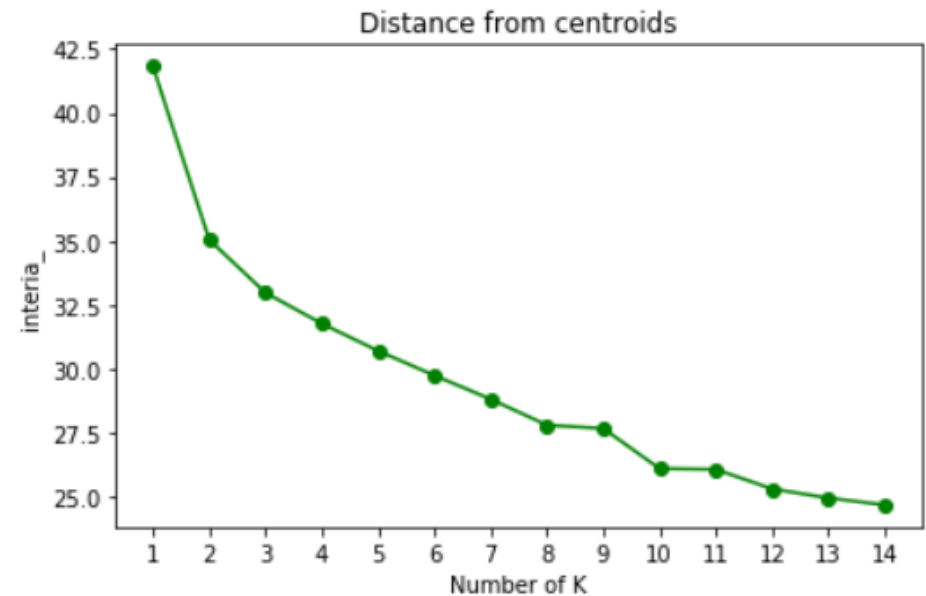
[110 rows x 2 columns]
Total venues in unique Finnish venue categories: 406

country_code	EE	FI
Venue Category		
Eastern European Restaurant	258.0	NaN
Bus Line	111.0	NaN
Market	59.0	NaN
Church	55.0	NaN
Shoe Store	53.0	NaN
...
Ski Chairlift	1.0	NaN
Residential Building (Apartment / Condo)	1.0	NaN
Fruit & Vegetable Store	1.0	NaN
Animal Shelter	1.0	NaN
Veterinarian	1.0	NaN

[73 rows x 2 columns]
Total venues in unique Estonian venue categories: 1265

Selecting k for the k-Means clustering algorithm

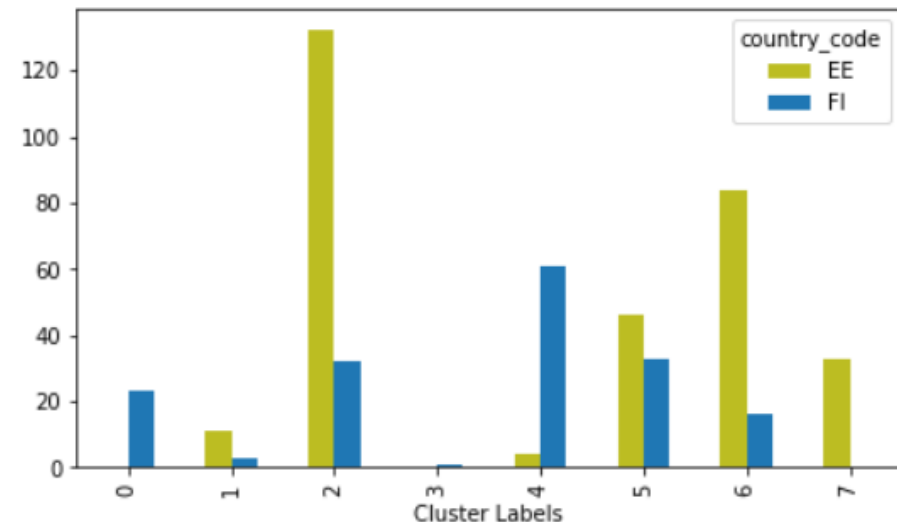
- The **k-Means clustering algorithm** is run in a loop
- In order to try various number of clusters (2 to 15) on the formatted dataset
- The inertia is extracted from each model (as plotted to the right)
- k = 2 was not found to be useful
- **k= 8 was selected** as the next (with elbow rule recommendation)



k-Means results (k = 8 clusters)

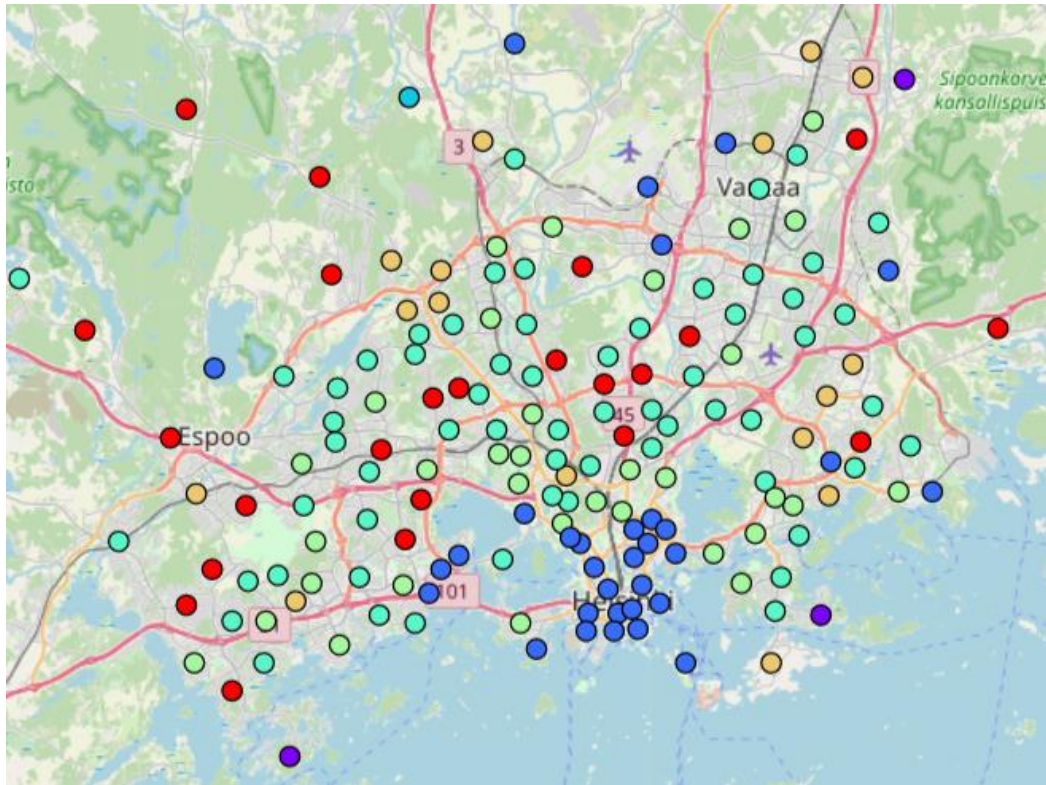
- 7 clusters for Helsinki
- 6 clusters for Tallinn
- **5 overlapping clusters**
- 2 clusters unique for Helsinki
 - Cluster Label 3 only contains one postal code (a construction venue)
- 1 cluster unique for Tallinn

The number of neighborhoods (postal codes) by cluster

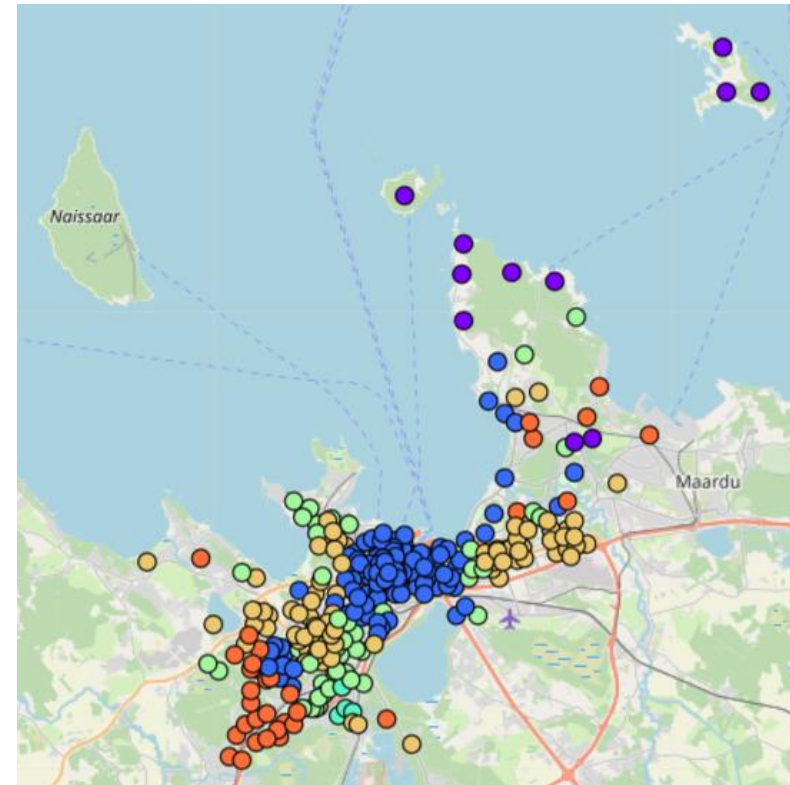


Clusters mapped and visualized

- **Helsinki region**



- **Tallinn**



- **Cluster 1**
- **Cluster 2**
- **Cluster 3**
- **Cluster 4**
- **Cluster 5**
- **Cluster 6**
- **Cluster 7**
- **Cluster 0**

Clusters described (ones that clearly overlap)

Cluster 1 (purple on maps)

- small cluster
- represents neighborhoods close to seashore, nature, historical sites

Labels	Venue Category	Freq
1	Beach	0.254762
1	Historic Site	0.107143
1	Harbor / Marina	0.079762
1	Boat or Ferry	0.053571
1	Bar	0.047619
1	History Museum	0.035714
1	Food Court	0.032143
1	Playground	0.026190
1	Campground	0.026190
1	Forest	0.023810

Cluster 2 (dark blue on maps)

- large cluster
- clearly describes the city centres and a few other central neighborhoods in other parts of the cities

Labels	Venue Category	Freq
2	Café	0.062043
2	Restaurant	0.050510
2	Park	0.030481
2	Hotel	0.030082
2	Coffee Shop	0.024791
2	Pizza Place	0.016718
2	Eastern European Restaurant	0.016291
2	Bar	0.016153
2	Burger Joint	0.015776
2	Scenic Lookout	0.015748

Clusters described (diverging profiles towards the suburbs - Helsinki)

Cluster 5 (light green)

- well represented in both cities
- often borders Cluster 2 (city centres)

Labels	Venue Category	Freq
5	Bus Stop	0.092222
5	Park	0.077513
5	Pizza Place	0.038554
5	Café	0.029687
5	Supermarket	0.024467
5	Restaurant	0.023945
5	Trail	0.022948
5	Gym / Fitness Center	0.021179
5	Grocery Store	0.018181
5	Diner	0.016424

Cluster 4 (cyan)

- strong in Helsinki, very small in Tallinn
- increasing profile weight for bus stops

Labels	Venue Category	Freq
4	Bus Stop	0.264373
4	Grocery Store	0.051638
4	Park	0.041050
4	Pizza Place	0.039794
4	Soccer Field	0.039355
4	Supermarket	0.036313
4	Trail	0.021081
4	Playground	0.019473
4	Beach	0.018395
4	Café	0.018376

Cluster 0 (red)

- only in Helsinki
- strong profile weight for bus stops
- more nature and sport

Labels	Venue Category	Freq
0	Bus Stop	0.510887
0	Grocery Store	0.067734
0	Pizza Place	0.058091
0	Beach	0.019928
0	Soccer Field	0.018711
0	Playground	0.017475
0	Gym / Fitness Center	0.015259
0	Park	0.014612
0	Convenience Store	0.013686
0	Trail	0.011702

Clusters described (diverging profiles towards the suburbs - Tallinn)

Cluster 5 (light green)

- well represented in both cities
- often borders Cluster 2 (city centres)

Labels	Venue Category	Freq
5	Bus Stop	0.092222
5	Park	0.077513
5	Pizza Place	0.038554
5	Café	0.029687
5	Supermarket	0.024467
5	Restaurant	0.023945
5	Trail	0.022948
5	Gym / Fitness Center	0.021179
5	Grocery Store	0.018181
5	Diner	0.016424

Cluster 6 (light orange)

- strong in Tallinn, very small in Helsinki
- intensive in stores and places to buy food

Labels	Venue Category	Freq
6	Grocery Store	0.066632
6	Supermarket	0.059283
6	Bus Station	0.040550
6	Bus Stop	0.038769
6	Fast Food Restaurant	0.038513
6	Pizza Place	0.032923
6	Café	0.031210
6	Gym	0.027997
6	Convenience Store	0.027521
6	Shopping Mall	0.027121

Cluster 7 (orange)

- only in Tallinn
- strong profile weight for public transportation

Labels	Venue Category	Freq
7	Bus Station	0.251918
7	Trail	0.058700
7	Grocery Store	0.041574
7	Train Station	0.035991
7	Park	0.033008
7	Bus Line	0.025504
7	Cable Car	0.020297
7	Restaurant	0.019938
7	Skate Park	0.019459
7	Café	0.015480

Conclusions and future directions

- The clusters identified in this study provide useful insights into the structural neighbourhood profiles of Helsinki capital region and Tallinn
- The gained insights may help people who consider moving to the other city
 - To understand which neighborhoods are similar or different from the one where they currently live
- Possible future direction could be to continue the study by taking assumptions about the target group preferences and group venue categories, such as:
 - Restaurants: general "Restaurants" and types ("Himalayan", "Eastern European" etc.)
 - Bus related categories: "Bus stop", "Bus station", "Bus line"
 - Etc.