

# Data preparation

## Getting Helsinki Neighbourhoods Information

Since we are going to clustering the data, having the neighbourhood information helps us in grouping the data into each district in Helsinki. Using Google and BeautifulSoup4 library in Python, a data-frame contains the name of 51 subdivisions in the capital of Finland is created.

Out[2]:		27	Tapatur
Neighborhood		28	Haaga
0	Kamppi	29	Eira
1	Kallio	30	Paloheinä
2	Herttoniemi	31	Maunula
3	Vuosaari	32	Arabianranta
4	Töölö	33	Sompasaari
5	Malmi	34	Jakomäki
6	Tali	35	Mellunmäki
7	Pasila	36	Munkkivuori
8	Munkkiniemi	37	Kluuvi
9	Itäkeskus	38	Sörnäinen
10	Ullanlinna	39	Tapulikaupunki
11	Vallila	40	Kumpula
12	Kalasatama	41	Pikku Huopalahti
13	Kontula	42	Merihaka
14	Malminkartano	43	Pihlajamäki
15	Kuninkaantammi	44	Pakila
16	Pitäjänmäki	45	Pukinmäki
17	Viikki	46	Kivikko
18	Myllypuro	47	Ruskeasuo
19	Katajanokka	48	Machine Shop
20	Oulunkylä	49	Jollas
21	Kannelmäki	50	Hermann

Using Geopy Client, a python library that helps us getting the latitudes and longitudes of all districts. Combing with the list of subdivisions above, we can have a completed data frame as below:

	Neighborhood	Latitude	Longitude
0	Kamppi	60.168535	24.930494
1	Suomenlinna	60.145706	24.988860
2	Munkkiniemi	60.199094	24.874358
3	Herttoniemi	60.195525	25.029063
4	Vuosaari	60.208348	25.143554
5	Itäkeskus	60.208652	25.084760
6	Hermesaari	60.148235	25.096343
7	Ullanlinna	60.158715	24.949404
8	Kallio	60.183982	24.952451
9	Vallila	60.196167	24.956710
10	Töölö	13.744717	-1.964599
11	Kalasatama	60.185295	24.980187
12	Katajanokka	60.166975	24.968151
13	Malmi	60.250761	25.008574
14	Kontula	60.238786	25.080050
15	Kuninkaantammi	60.265462	24.892970
16	Pitäjänmäki	60.222711	24.864356
17	Oulunkylä	60.229443	24.963583
18	Kannelmäki	60.243608	24.883289
19	Myllypuro	60.224010	25.066383
20	Malminkartano	60.249340	24.859959

## Collect Venues Locations using Foursquare:

Using Foursquare API, we can get a very large database of nearby venues for each neighborhood in JSON format. Here in this table, we are able to choose 100 popular locations maximum for each district within a 1 km radius.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Kamppi	60.168535	24.930494	Cafe Rouge	60.168711	24.933027	Middle Eastern Restaurant
3	Kamppi	60.168535	24.930494	Pobre	60.169500	24.933484	Filipino Restaurant
12	Kamppi	60.168535	24.930494	Singapore Hot Wok	60.160163	24.933890	Malay Restaurant
14	Kamppi	60.168535	24.930494	Sushibar + Wine	60.167187	24.932170	Sushi Restaurant
15	Kamppi	60.168535	24.930494	Hoku	60.168490	24.933543	Hawaiian Restaurant
...	...	...	...	...	...	...	...
1162	Kaisaniemi	60.173182	24.946400	Momotoko	60.169834	24.946957	Ramen Restaurant
1169	Kaisaniemi	60.173182	24.946400	Doner Harju City	60.170082	24.941644	Doner Restaurant
1173	Kaisaniemi	60.173182	24.946400	Kiinalainen Ravintola Family	60.174245	24.950121	Chinese Restaurant
1178	Kaisaniemi	60.173182	24.946400	Fazer Food & Co Oodi	60.173648	24.937799	Restaurant
1195	Kaisaniemi	60.173182	24.946400	Ravintola Kaisaniemi	60.176665	24.941657	Restaurant