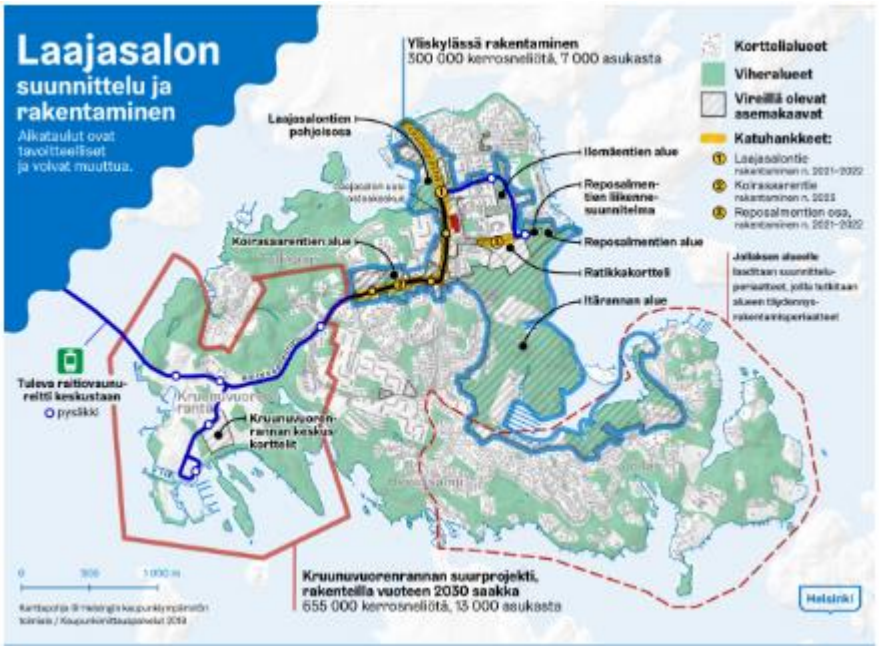
Coursera Capstone - The Battle of Neighborhood



Introduction/Business Problem:

My cousin lives in Finland in a quite small town which has approximately only 10000 inhabitants. He has there a restaurant which has operated mainly as a lunch restaurant. My cousin has several times told that he would like to open a restaurant in Helsinki where are many more inhabitants.

Up to this point he has succeeded with his restaurant in that small town. When Finnish government decided to close all restaurants until the end of May because of COVID-19, my cousin also had to close his restaurant. Now at this moment it seems that this kind of business is slowly recovering but in that small town where my cousin lives, recovering will take too long time. My cousin decided that now it is time to make that big decision to open a restaurant in Helsinki. He knows that the competition will be fierce, when the restaurants which are now closed, will open at about the same time.

My cousin is certainly not the only one who wants to start new restaurant business. This analysis is

a good addition for decision making.

To succeed the restaurant should be in a neighborhood where population will grow at least 10000 inhabitants until year 2030, not too many competitors in the same neighborhood at this time and it would also help if the transport connections are good.

Data

Based on definition of my cousin's problem, factors that will help him to decide are:

- All venues of neighborhood area through Foursquare API.
- Top venue categories in neighborhoods
- The amount of competitors
- Growth of population is minimum 10000

Following data sources will be needed to generate the required information:

- Wikipedia page of Helsinki neighborhoods https://fi.wikipedia.org/wiki/Helsingin_alueellinen_jako
- Geolocator to get coordinates of neighborhoods.
- A sample of a study of Helsinki population growth from year 2019 to year 2030, Helsinki population growth research https://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/18_10_30_Tilastoja_18_Vuori_Kaasila.pdf

The explore function is used to get the most common venue categories in each neighborhood of Helsinki. Also clustering neighborhoods is used to give similarity information to my cousin.

Methodology

The methodology section describes how the analysis is conducted.

At first we need the Helsinki neighborhoods and their longitudes and latitudes. With that table we merge a sample of a study of Helsinki population growth. We calculate the percentage and sum of the population growth. Only the largest increases are taken into account.

Foursquare API service is used to explore the neighborhoods to analyze and segment them.

The neighborhoods were clustered into 5 clusters by k-means and the values and a map is produced.

Finally, the competing venues were identified and calculated and the result was found out. The growth of population from 2019 to 2030 must be at least 10000 inhabitants, the amount of competitors must be under 5 and at least 3000 customer must be available for a venue.

Methodology

We are providing characteristic information about Helsinki neighborhoods combining venue and pricing information and making clusters of neighborhoods.

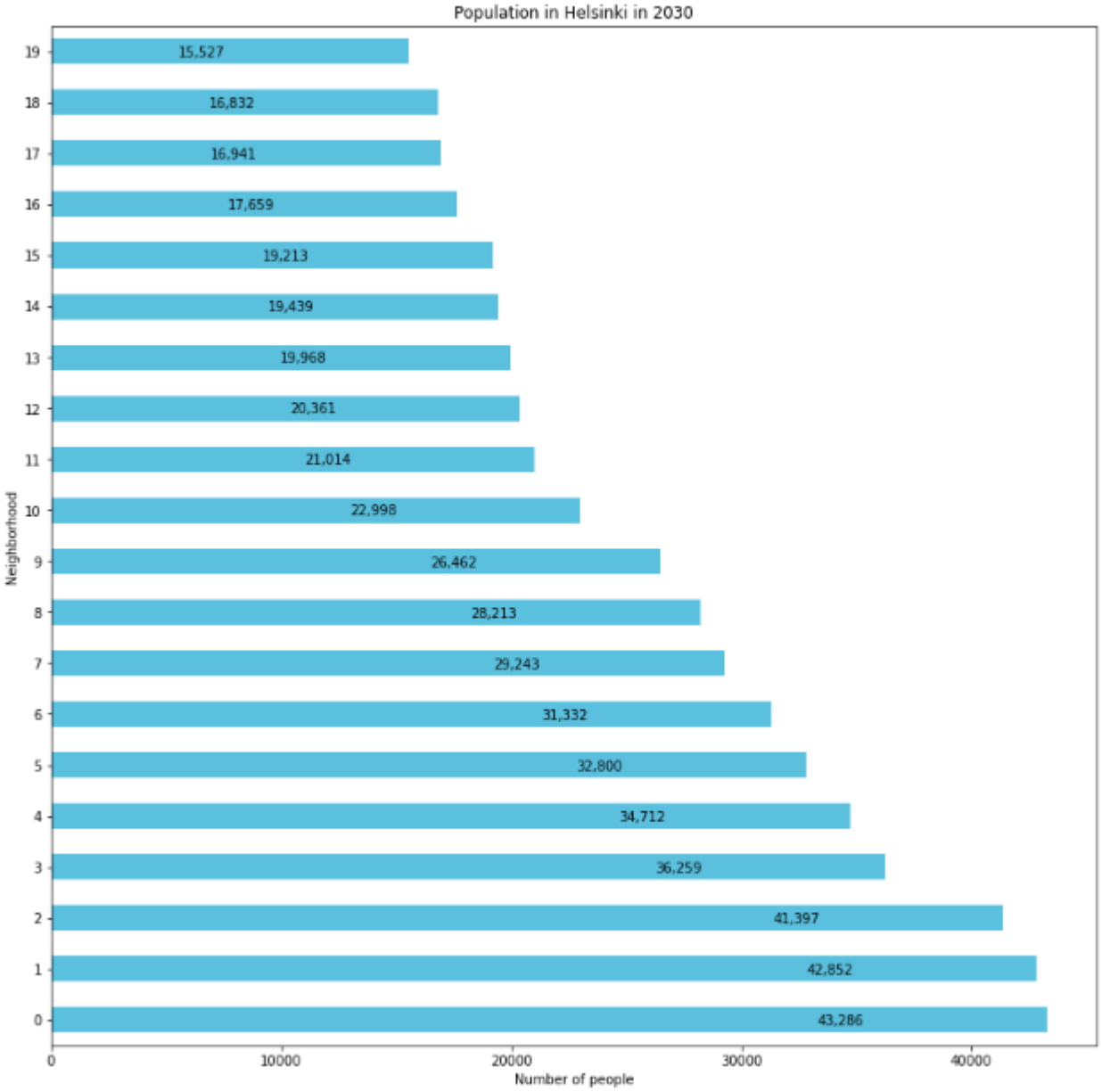
First phase for project was that:

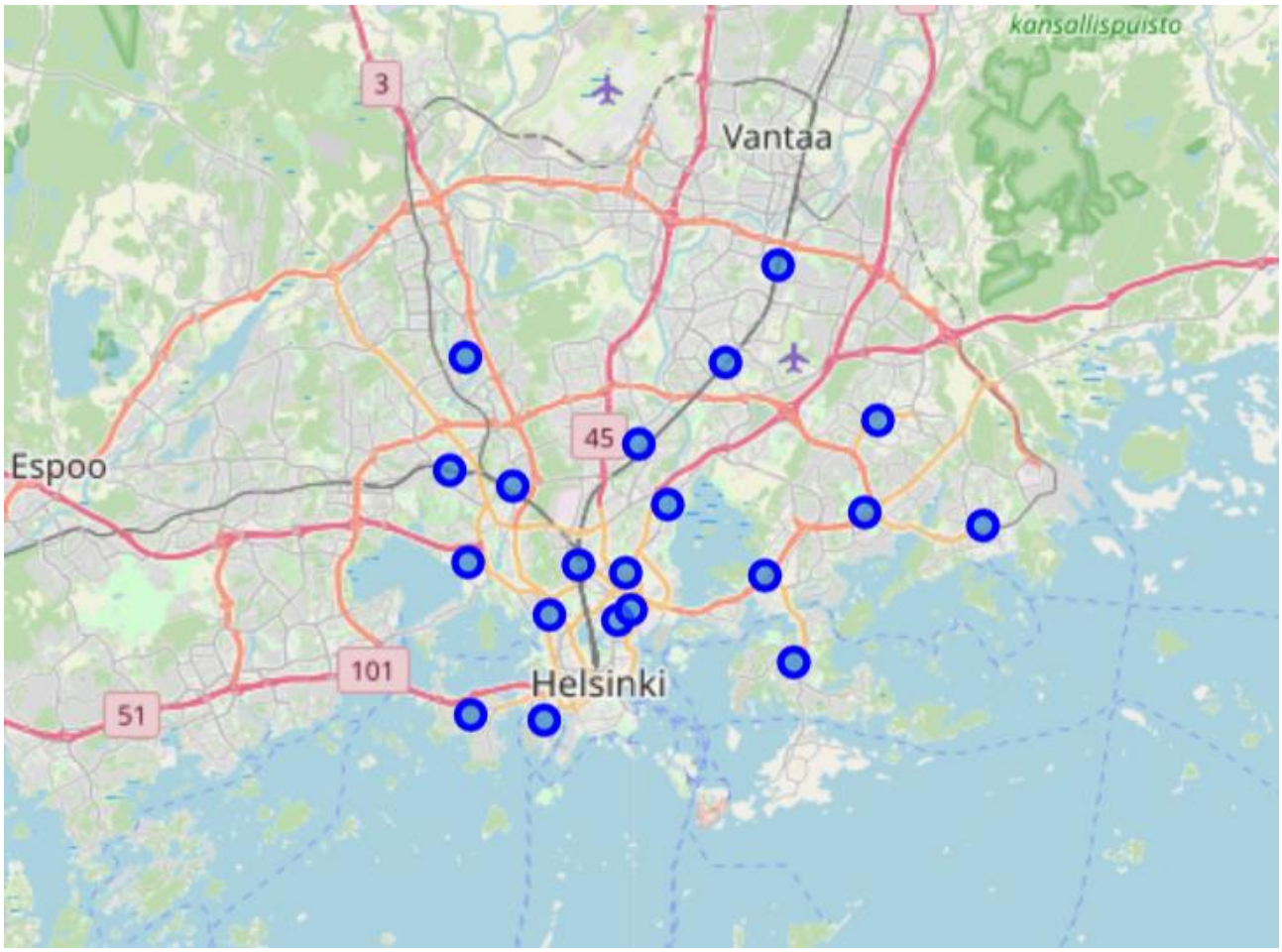
- We collected all neighborhoods with subneighborhoods.
- Added coordinates to all neighborhoods
- Added average m2 pricing to all neighborhoods, when data was available

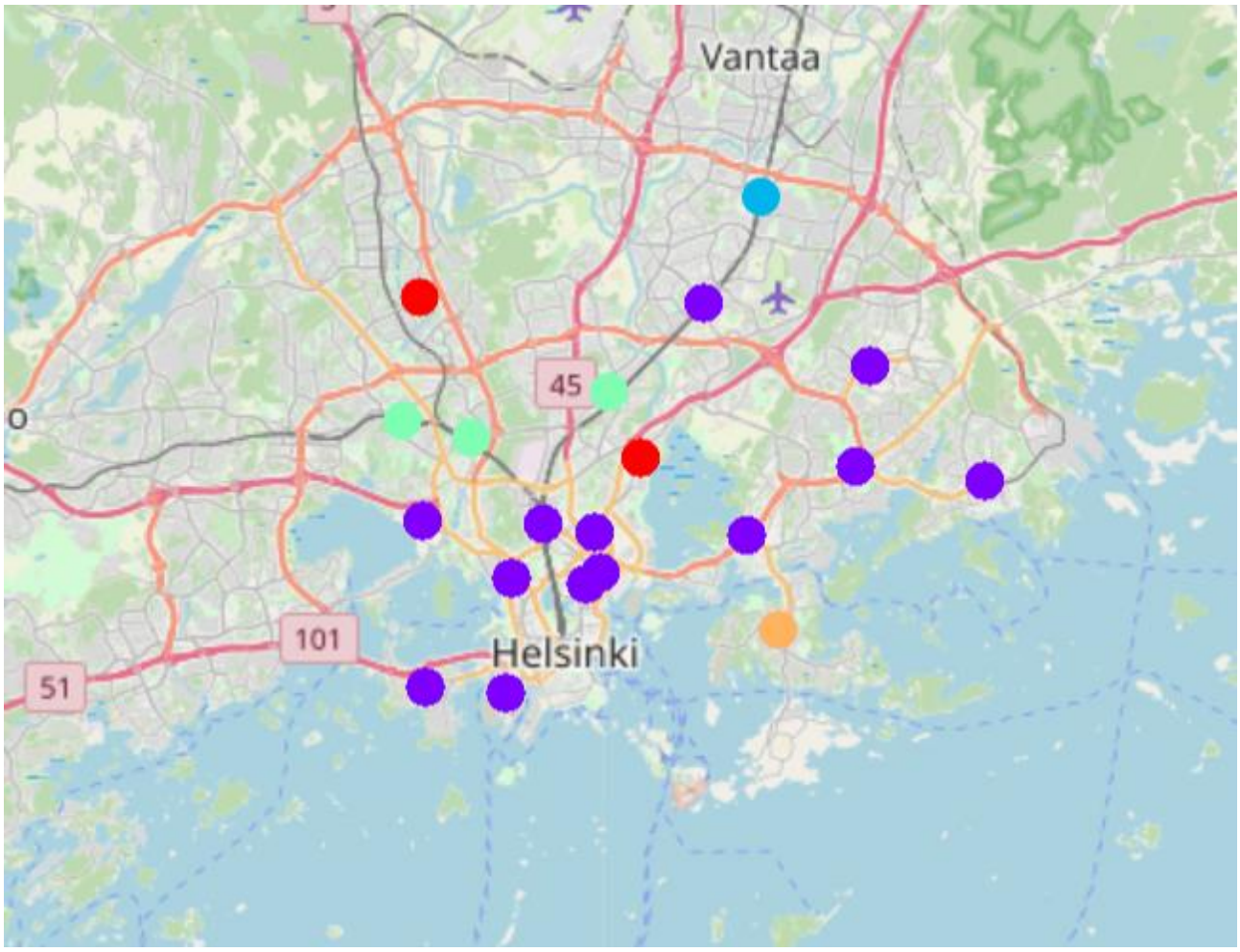
Second phase was make analysis

- We

Results


Population in Helsinki in 2030
Neighborhood
Number of people
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
15,527
16,832
16,941
17,659
19,213
19,439
19,968
20,361
21,014
22,998
26,462
28,213
29,243
31,332
32,800
34,712
36,259
41,397
42,852
43,286
0
10000
20000
30000
40000


kansallispuisto
3
Vantaa
45
Espoo
101
51
Helsinki


Vantaa
45
101
Helsinki
51

Discussion

Helsinki City is relatively small city so for effective data analysis we would need even more venues in Foursquare. Lot of neighborhoods have small amount of many kind of venues to offer people who want to move there. With adjusting k-means clustering we could have even better results. There is a big difference in average m2 housing price inside Helsinki and look's that relates more other things that venues in neighborhood.

Conclusion

This kind of data analysis can be a helpful for a real estate agent and provides competitive advance quickly. Data should be used from app and Python code of this analysis should be provided as a microservice

Juhan (tee samanlainen selostus omastasi)!!!:

Helsinki neighborhoods https://fi.wikipedia.org/wiki/Helsingin_alueellinen_jako

Helsinki population growth research https://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/18_10_30_Tilastoja_18_Vuori_Kaasila.pdf