Capstone Project - The Battle of Neighborhoods (Week 2)

Table of contents

Introduction / Business Problem
knowledge of neighborhood for some real estate agent

Clearly define a problem or an idea of your choice, where you would need to leverage the Foursquare location data to solve or execute. Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem.

This submission will eventually become your Introduction/Business Problem section in your final report. So I recommend that you push the report (having your Introduction/Business Problem section only for now) to your Github repository and submit a link to it:

- A real estate agent is a natural person who is dedicated to providing mediation, advice and management services in real estate transactions related to: the sale, rental, exchange or transfer of real estate and their corresponding rights, including the constitution of these rights. In each country, the activity is governed by a particular law, so far there is no law that regulates real estate issues worldwide, despite the fact that many real estate agents carry out transactions in countries other than their headquarters.
- No one can remember or know all the places in the Helsinki, Finland area, so they cannot promote all the places and categories that can be found through the Foursquare API. We will provide specific information near the property sold for real estate agents. We would also group and classify habitable areas to quickly tell which category of property it belongs to and what are the unique characteristics of that area, for example, good parks and coffee shops. This makes a difference when the agent has a sales meeting with the owner. This information can even be crucial when families decide where they will move and buy a new house.

Data

Describe the data that you will be using to solve the problem or execute your idea. Remember that you will need to use the Foursquare location data to solve the problem or execute your idea. You can absolutely use other datasets in combination with the Foursquare location data. So make sure that you provide adequate explanation and discussion, with examples, of the data that you will be using, even if it is only Foursquare location data.

This submission will eventually become your Data section in your final report. So I recommend that you push the report (having your Data section) to your Github repository and submit a link to it.
Based on definition of our problem, factors that will help real estate agent are:

- All venues of neighborhood
- Top venue categoeries in neighborhood
- Overall style for example cafes and parks

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

- Wikipedia page of Helsinki neighborhood including
- All venues or neigborhood area through Foursqueare API
- Geolocator to get coordinates of neighborhoods

We will use the explore function to get the most common venue categories in each neighborhood of Helsinki. We will also cluster neighborhoods to give similarity information to end customer.

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.

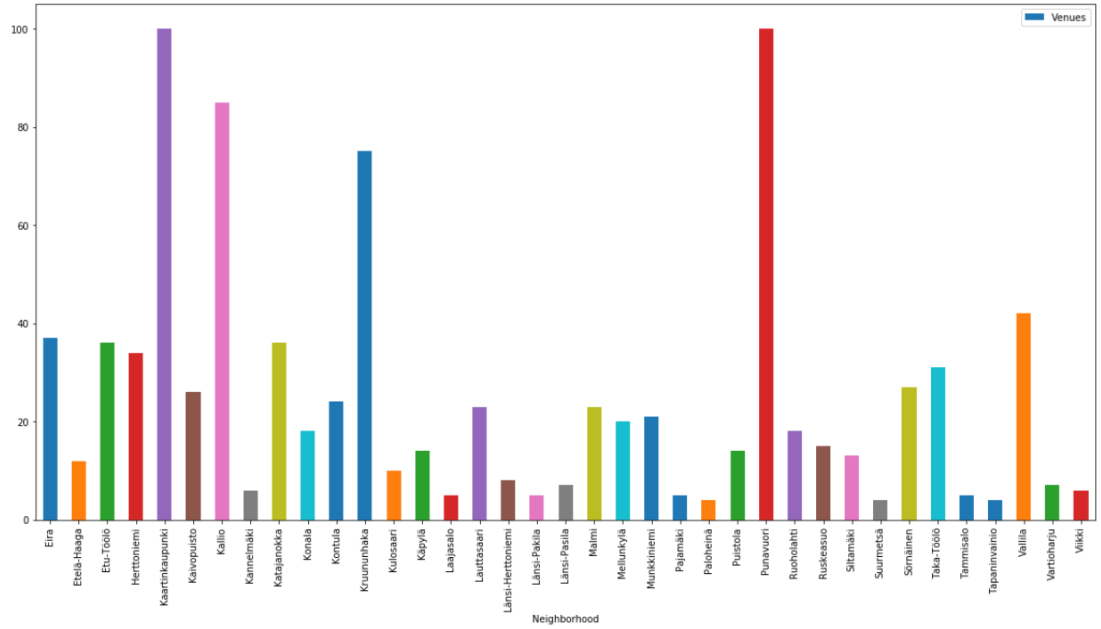
Second phase for project was that:

- We cluster all neighborhoods and venues correlated.
- Then we got that number down to the top 10 unique venues in that neighborhoods.
- Then we down to the top 5 based on the most frequented venues.

Out[107]:

	name	categories	lat	lng
0	Restaurant Ask	Scandinavian Restaurant	60.172366	24.955864
1	Korea House	Korean Restaurant	60.172910	24.956436
2	Cafe LOV	Café	60.171284	24.956623
3	Bei Fang	Chinese Restaurant	60.171602	24.953990
4	Papu Cafe	Café	60.173040	24.956453

Out[113]: <matplotlib.axes._subplots.AxesSubplot at 0x1a26aaba58>



Out[121]:

	Neighborhood	lat	lng	Price	Venues	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6t C Ve
0	Kruununhaka	60.172870	24.954733	6946	75	4	Scandinavian Restaurant	Zoo	Flower Shop	Garden Center	Garden	Fu / H St
1	Kaartinkaupunki	60.165214	24.947222	8205	100	0	Korean Restaurant	Zoo	Flower Shop	Gas Station	Garden Center	Ga
2	Punavuori	60.161237	24.936505	7481	100	3	Café	Zoo	Gastropub	Garden Center	Garden	Fu / H St
3	Eira	60.156191	24.938375	8051	37	0	Chinese Restaurant	Zoo	Gastropub	Garden Center	Garden	Fu / H St
4	Katajanokka	60.166975	24.968151	6134	36	3	Café	Zoo	Gastropub	Garden Center	Garden	Fu / H St

Discussion

Helsinki City is a relatively small city, so for an effective data analysis we would need even more places provided by Foursquare. Many neighborhoods have a small number of many types of places to offer to people who want to move there. With the adjustment of the k-means grouping we could have even better results. There is a big difference in the average price of housing in square meters within Helsinki and aspects related to other things found in neighborhoods.

Conclusion

This type of data analysis can be useful for any real estate agent and provides rapid competitive progress at the time of home sales in the neighborhood. The data should be used from the application and the Python code of this analysis should be provided as a microservice with support from Foursquare.