Applied Data Science Capstone

The Battle of Neighborhoods

Final Report

Navid Karimian Pour

Date: 13.12.2020

1. Introduction

People relocate to new cities all the time for a wide variety of reasons such as work, studies, family, weather, and so on. Relocating to a new city whether in your country of residence or a new country can be both exciting and challenging. One of the main challenges that people face when relocating is finding a new apartment to buy/rent in a nice neighborhood. But in many cases, we are unfamiliar with the city we are relocating to and we need to do extensive googling to find reliable information. Besides that, many people would like to move to a neighborhood that has a lot of similarities with their old neighborhood. This is where data science can jump in and help in finding similar neighborhoods in any city we are relocating to. In this assignment, I introduce a fictional character called "John Anderson" who has been facing challenges in finding a good neighborhood in a new city he is relocating to.

John is a 35-year-old business consultant working at a fortune 500 consulting firm. John lives in a rental apartment in the "Central Bay Street" neighborhood, Toronto, Canada. He recently got a promotion at his job which requires him to relocate to Helsinki, Finland. John is very excited about the promotion, but he is a bit worried because he does not know much about Helsinki. He enjoys living at "Central Bay Street" and hopes to find a similar neighborhood in Helsinki to rent an apartment. John's friend Mark is a data scientist and when he became aware of his problem, he offered to help. John told Mark that he has two main criteria for selecting a neighborhood in Helsinki:

- He wishes to find a neighborhood that has very similar characteristics to "Central Bay Street". Especially in terms of venues available in the area.
- He only wants to live in a neighborhood which is a maximum of 3km away from Helsinki city center so the apartment is not far away from the city center and also, he can explore the city easier during his free time.

Mark promised John to give him a list of neighborhoods in Helsinki that meet his criteria.

From now on, I will be taking the role of Mark and use the data science knowledge acquired throughout this course to come up with a good solution for this challenge. Let us dive into the challenge. I hope you find it interesting.

2. Data

Now that we understand the challenge and selection criteria, it is time to figure out what datasets are required and how we are getting the data. To solve this challenge, we need to find neighborhood data for both Helsinki and Toronto, coordinates of the neighborhoods, and venues data.

Neighborhood data is easily accessible from the following Wikipedia pages:

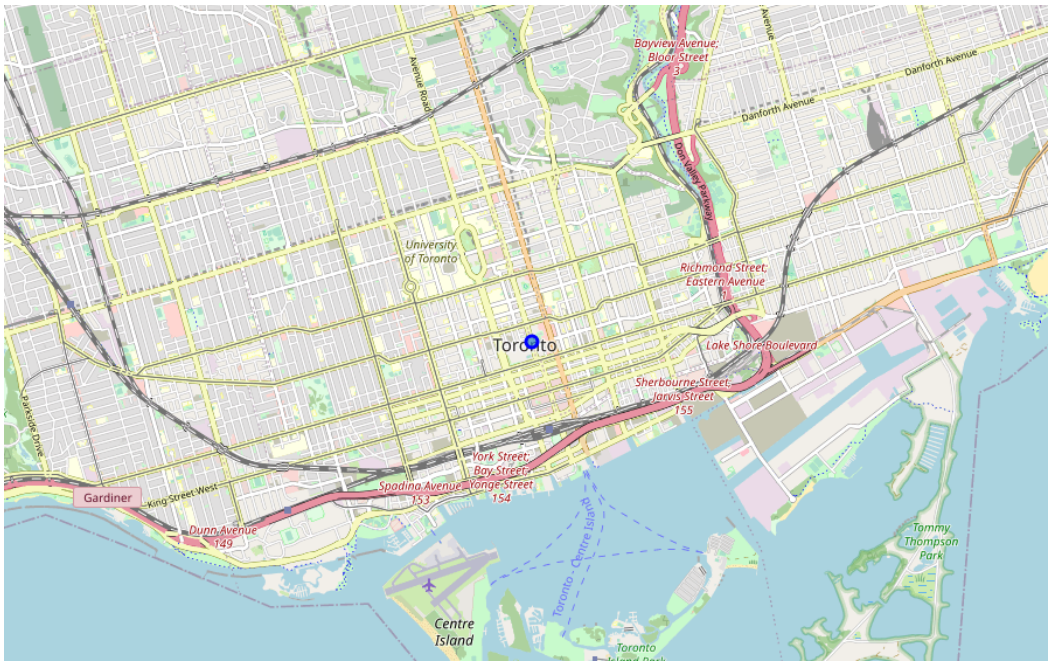
- Subdivisions of Helsinki
- List of postal codes of Canada

To fetch neighborhoods' geographical coordinates, we use **GeoPy** library. And to get the data for venues in each neighborhood, we use **Foursquare API**.

3. Methodology

In the methodology section we first perform exploratory data analysis to get a better understanding of the data and the challenge. Next, we use the data to build a model which will help us in selecting the right neighborhoods.

First let's take a look at the neighborhood that Joe Anderson lives: "Central Bay Street" neighborhood, Toronto, Canada

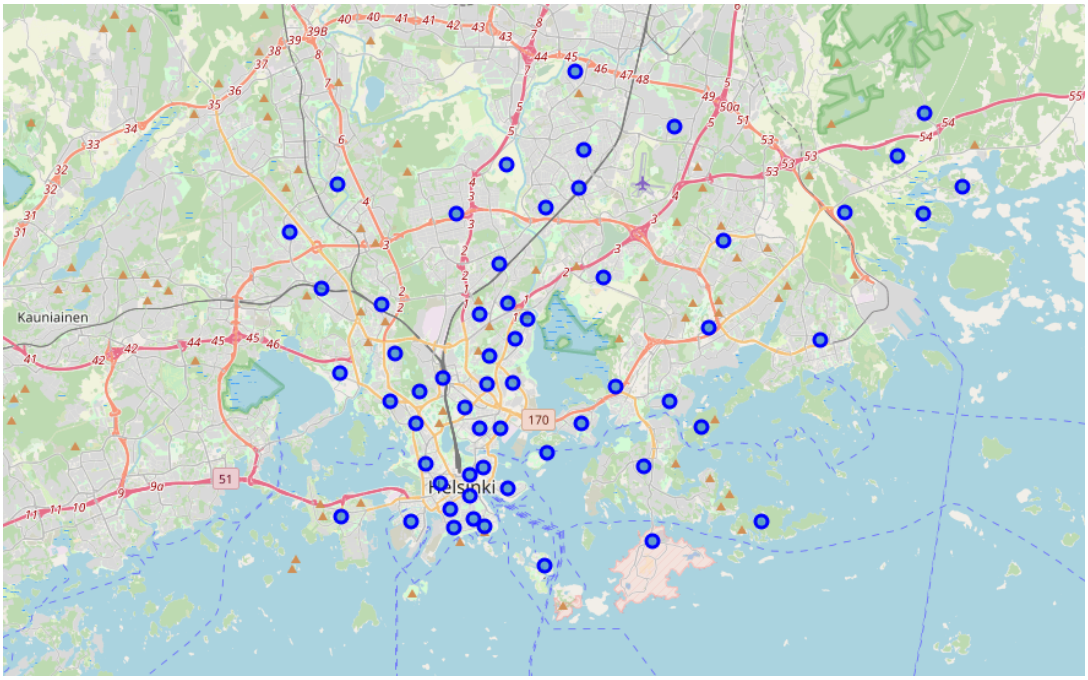


Now that we highlighted the location of the neighborhood on the map, let's take a look at venues which are available in this neighborhood.

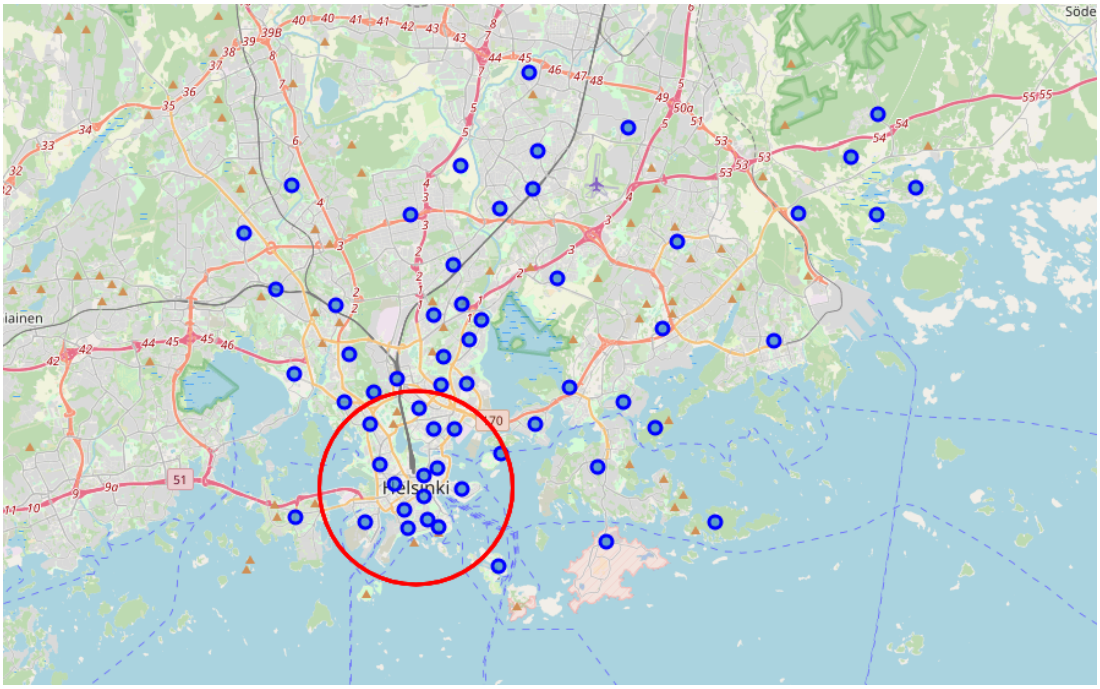
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
829	Central Bay Street	43.653779	-79.382944	Downtown Toronto	43.653232	-79.385296	Neighborhood
830	Central Bay Street	43.653779	-79.382944	Nathan Phillips Square	43.652270	-79.383516	Plaza
831	Central Bay Street	43.653779	-79.382944	Indigo	43.653515	-79.380696	Bookstore
832	Central Bay Street	43.653779	-79.382944	CF Toronto Eaton Centre	43.654447	-79.380952	Shopping Mall
833	Central Bay Street	43.653779	-79.382944	LUSH	43.653557	-79.380400	Cosmetics Shop
...	...	...	...	...	...	...	...
912	Central Bay Street	43.653779	-79.382944	Pantages Hotel & Spa	43.654498	-79.379035	Hotel
913	Central Bay Street	43.653779	-79.382944	Tim Hortons	43.655212	-79.380063	Coffee Shop
914	Central Bay Street	43.653779	-79.382944	Pantages Lounge & Bar	43.654493	-79.379000	Cocktail Bar
915	Central Bay Street	43.653779	-79.382944	Imperial Pub	43.656254	-79.378955	Pub
916	Central Bay Street	43.653779	-79.382944	Tim Hortons	43.653625	-79.377732	Coffee Shop

As we can see from the above table, "Central Bay Street" is a neighborhood which is located in the center of Toronto and there are wide variety of venues available in this neighborhood.

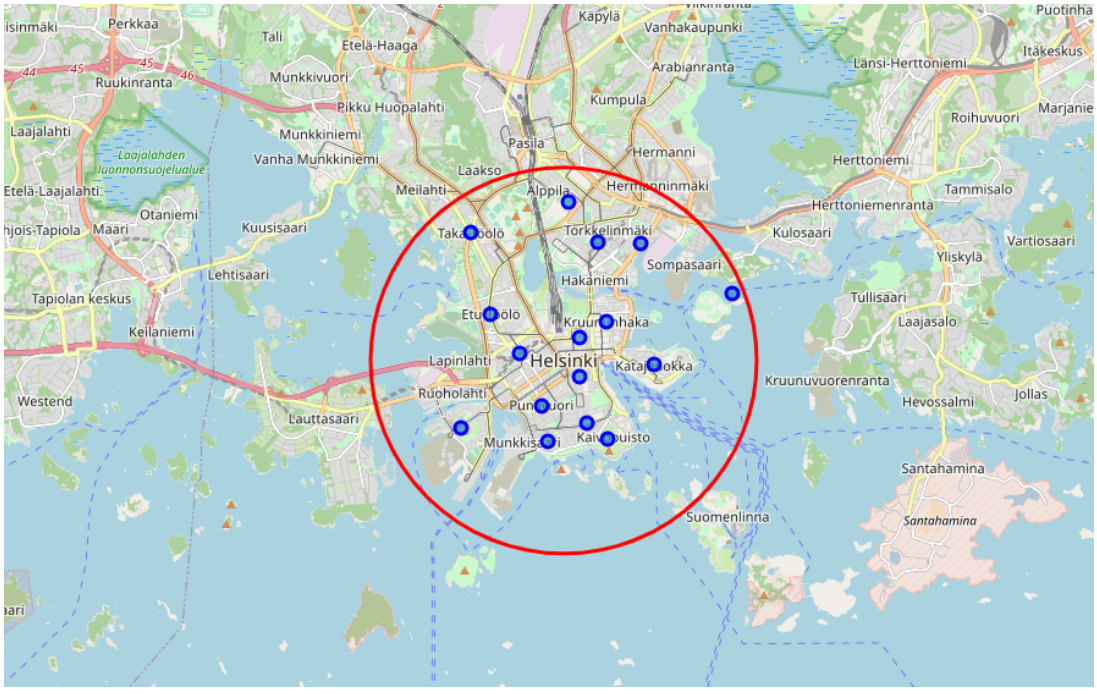
Now let's take a closer look at Helsinki neighborhoods by visualizing them on the map.



As mentioned in the Introduction section, Joe is only interested in neighborhoods which are maximum 3km away from Helsinki city center. Let's visualize that on the map with a red circle.



Now we can remove other neighborhoods from the map and take a closer look only at neighborhoods which are in the 3km radius.



Now that we limited number of neighborhoods based on their proximity to Helsinki city center, it is time to start building a model which can help us compare venue similarities between "Central Bay Street" and Helsinki neighborhoods. To build the model, we go through the following steps:

- Transforming venue data from categorical to numerical
- Calculate on average how many venue categories are in every neighborhood
- Select top 20 most popular venues in each neighborhood
- Use K means clustering algorithm to group similar neighborhoods (5 clusters)

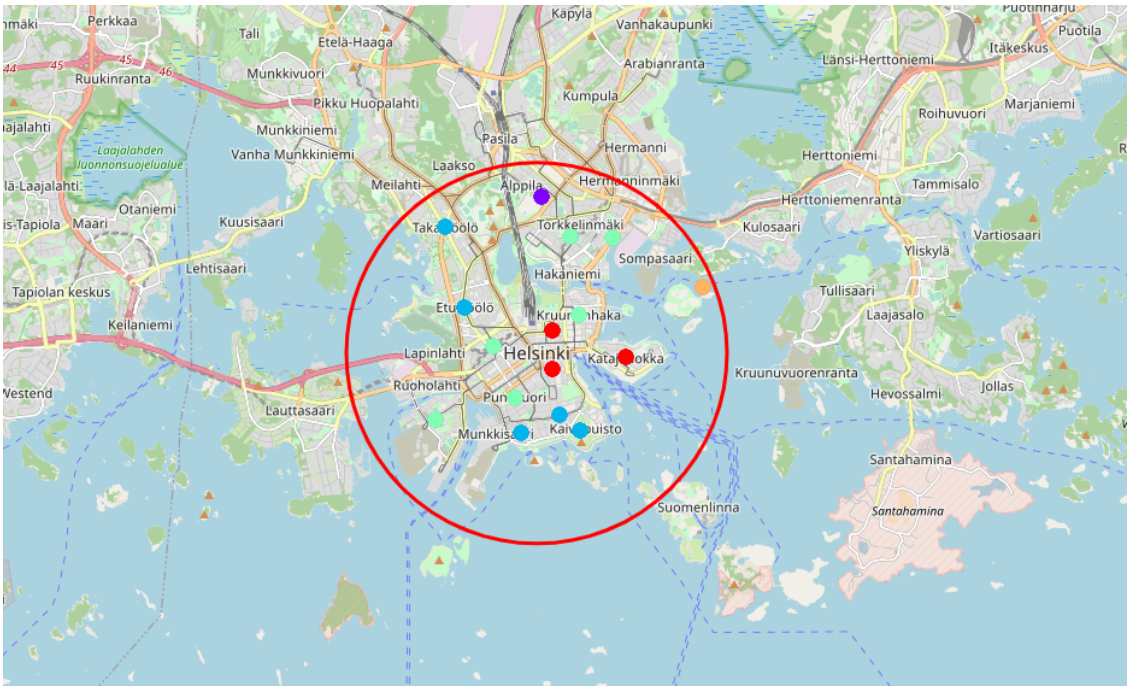
The result of clustering is shown in the below table:

	Neighborhood	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	...	11th Most Common Venue	12th Most Common Venue
0	Kruununhaka	60.172870	24.954733	3	Boat or Ferry	History Museum	Bar	Theater	Scandinavian Restaurant	Grocery Store	...	Indie Movie Theater	Beer Bar
1	Kluuvi	60.170778	24.947329	0	Coffee Shop	Café	Scandinavian Restaurant	Gym / Fitness Center	Theater	Park	...	Bar	Music Venue
2	Kaartinkaupunki	60.165214	24.947222	0	Scandinavian Restaurant	Hotel	Coffee Shop	Café	Cocktail Bar	Hotel Bar	...	Dance Studio	Plaza
3	Kamppi	60.168535	24.930494	3	Wine Bar	Beer Bar	Scandinavian Restaurant	Art Museum	Bar	Japanese Restaurant	...	Sushi Restaurant	Food Court
4	Punavuori	60.161237	24.936505	3	Scandinavian Restaurant	Restaurant	Bakery	Pizza Place	Coffee Shop	Park	...	Sandwich Place	Pub

In this section we did an expletory analysis on the data and built a clustering model to help in grouping neighborhoods according to the venue data. In the next section, we will review the clusters to see which one is like the "Central Bay Street" neighborhood.

4. Analysis

In the previous section we used K means clustering algorithm to group neighborhoods into 5 clusters based on their similarities in terms of venues. Below figure demonstrates these clusters on Helsinki map.



The above visualization shows us how the clusters are color coded and located in Helsinki, but we cannot see which cluster "Central Bay Street" belongs to. So, lets filter out the map further and only show neighborhoods in the same cluster as the "Central Bay Street".

	Neighborhood	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	...	11th Most Common Venue	12th Most Common Venue
16	Central Bay Street	43.653779	-79.382944	0	Coffee Shop	Clothing Store	Hotel	Restaurant	Theater	Electronics Store	...	Plaza	Seafood Restaurant

As we can see from the above table, "Central Bay Street" cluster label is 0. So, lets create a data frame which contains data for all cluster 0 neighborhoods and then visualize them on Helsinki map.

	Neighborhood	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	...	11th Most Common Venue	12th Most Common Venue	13th Most Common Venue
1	Kluuvi	60.170778	24.947329	0	Coffee Shop	Café	Scandinavian Restaurant	Gym / Fitness Center	Theater	Park	...	Bar	Music Venue	Burger Joint
2	Kaartinkaupunki	60.165214	24.947222	0	Scandinavian Restaurant	Hotel	Coffee Shop	Café	Cocktail Bar	Hotel Bar	...	Dance Studio	Plaza	Pizza Place
7	Katajanokka	60.166975	24.968151	0	Park	Hotel	Scandinavian Restaurant	Restaurant	Bar	Boat or Ferry	...	Plaza	Escape Room	Piano Bar
16	Central Bay Street	43.653779	-79.382944	0	Coffee Shop	Clothing Store	Hotel	Restaurant	Theater	Electronics Store	...	Plaza	Seafood Restaurant	Breakfast Spot

So, based on this analysis Kluuvi, Kaartinkaupunki, and Katajanokka are three neighborhoods which are both similar to "Central Bay Street" and are located within 3km radius of Helsinki city center.

5. Discussion

Nowadays, it is very common that people relocate to new cities for various reasons. One of the challenges of relocating to new cities is selecting a neighborhood to rent an apartment. The objective of this study was to help people find desirable neighborhoods in the new city they are relocating. To make the case more relatable, I introduced a fictional character called Joe Anderson who was planning to move from Toronto to Helsinki and sought data driven help to make a better decision.

We used retrieved neighborhood data for Helsinki and Toronto from Wikipedia and used Geopy and Foursquare to fetch location and venue data for all the neighborhoods. Then I performed exploratory analysis to better understand the challenge.

In order to select the best neighborhoods in Helsinki, I first filtered out the neighborhoods based on their proximity to the Helsinki city center. Then I used the

K means algorithm for clustering neighborhoods. I set K value to 5 and used top 20 venues in each neighborhood. Finally, I summarized the results in the form of table and visualization on the Helsinki map.

6. Conclusion

The result of the clustering revealed that there are three neighborhoods in Helsinki namely Kluuvi, Kaartinkaupunki, and Katajanokka which had a lot of similarities to Central Bay street neighborhood. Also, they were within 3km radius of Helsinki city center which was also a deciding factor for the fictional character.

In the future, we can add more criteria to find even better matching neighborhoods in cities we are relocating to.