Coursera Capstone Project

IBM Applied Data Science Capstone

Opening a new Fast Food Restaurant in Greater Helsinki

Author: Loc Hoang

September 2019

1. Introduction

Greater Helsinki is the metropolitan area surrounding Helsinki, the capital city of Finland. The Greater Helsinki region is the largest urbanized area in the country with 1,495,271 inhabitants and is by far the most important economic, cultural, and scientific region of Finland. Five out of Finland's 15 universities and most of the headquarters of notable companies and governmental institutions are in Greater Helsinki.

Business problem

As a result, there are many opportunities as well as challenges for stakeholders when coming up with starting their business in this area.

In this project scale, we will analyze and select the best location among Greater Helsinki to open a restaurant. In specify, we are planning to open a Fast Food restaurant.

Clustering technique will be used to solve the question above.

Target audiences

This project is particularly useful to property developers and investors looking to open or invest in new restaurant in the Greater Helsinki.

2. Data

a. List of neighborhoods

The Wikipedia page: https://en.wikipedia.org/wiki/Greater_Helsinki contains a list of neighborhoods in Greater Helsinki, with 13 neighborhoods. We will use web scraping techniques to extract data from web page. In more detail, pandas and requests are 2 modules will be used.

b. Latitude and longitude of each neighborhoods

Python Geocoder package will support us to get the latitude and longitude of each neighborhood.

c. Venues data

After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the fast food restaurant category in order to help us to solve the business problem.

3. Methodology

Firstly, we will do web scraping using Python requests and pandas packages to extract the list of neighborhoods data into a data frame. This data frame needs to be cleaned unnecessary rows as well as columns, then we have a data frame contains list of neighborhoods.

In next step, we need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the wonderful Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the data into a pandas DataFrame and then visualize the neighborhoods in a map using Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in Greater Helsinki area.

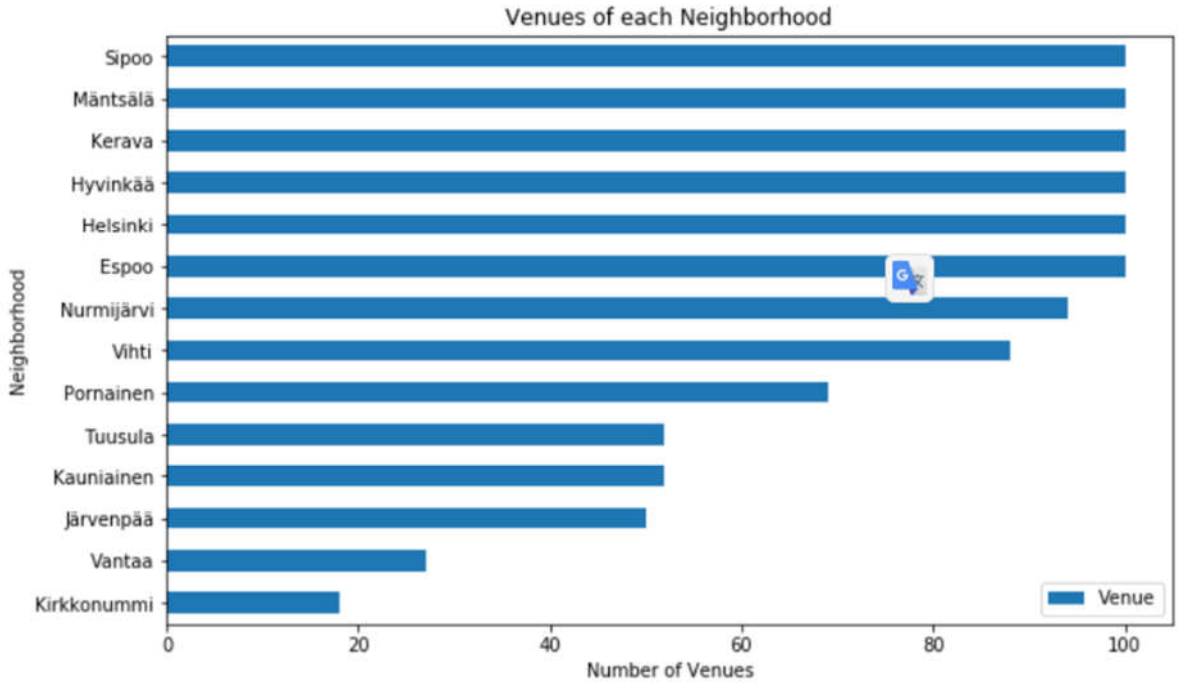
Next, we will make a loop of geographical coordinates in above dataframe. In the loop Foursquare API will be call to get the top 100 venues that are within a radius of 2000 meters.

End of the loop, we got venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each neighborhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyze each neighborhood by grouping the rows by neighborhoods and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering.

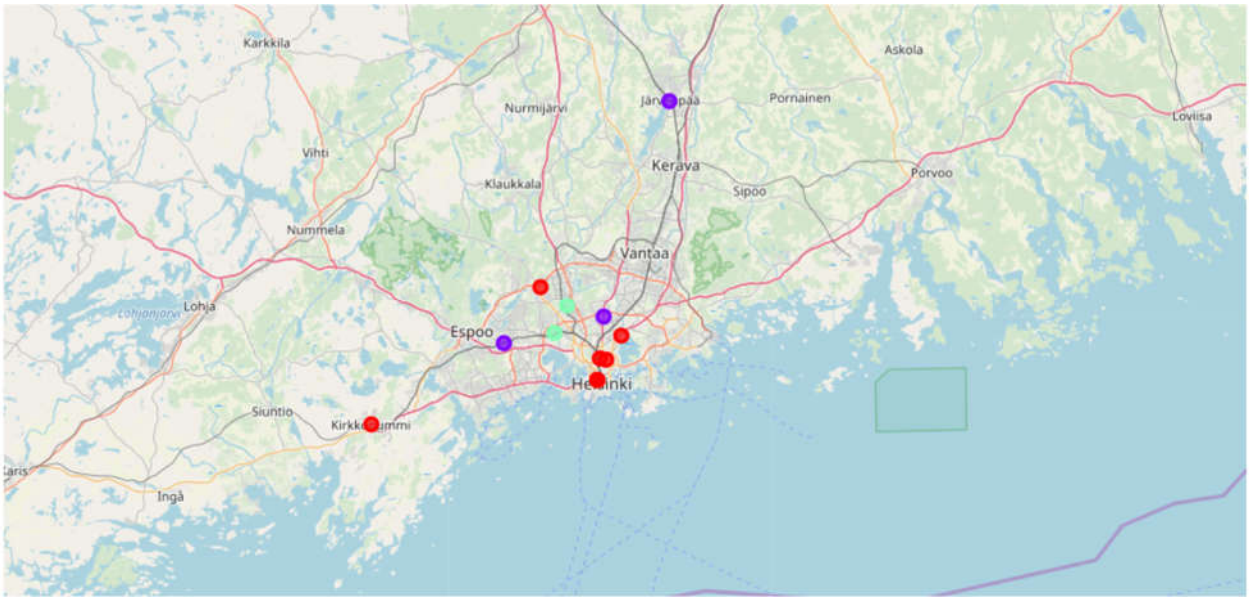
In final step, we will filter the data by 'Fast Food Restaurant' category to get specified data. After that, we will perform clustering on the data using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighborhoods into 3 clusters based on their frequency of occurrence for "Fast Food Restaurant". The results will allow us to identify which neighborhoods have higher concentration of shopping malls while which neighborhoods have fewer number of shopping malls. Based on the occurrence of

shopping malls in different neighborhoods, it will help us to answer the question as to which neighborhoods are most suitable to open new restaurant.

4. Results



Firstly, we can see that most of venues locates in big neighborhood or central of Greater Helsinki.



The results from the k-means clustering show that we can categorize the neighborhoods into 3 clusters based on the frequency of occurrence for “Fast Food Restaurant”:

• Cluster 0: Neighborhoods with low number of shopping malls

• Cluster 1: Neighborhoods with high number to no existence of shopping malls

• Cluster 2: Neighborhoods with moderate concentration of shopping malls

The results of the clustering are visualized in the map below with cluster 0 in red color, cluster 1 in purple color, and cluster 2 in mint green color.

5. Discussion

Most of Fast food restaurants are concentrated at Cluster 1 with the highest number, and the moderate in cluster 2. The most surprised is that both of cluster 1 and 2 is not in the center of Greater Helsinki area, which has the highest population density. This represents a great opportunity to open a fast food restaurant in cluster 1 area which contains many big neighborhoods like Espoo, Helsinki...since there are few or not operating fast food restaurant. In contrast, starting fast food restaurant in Järvenpää, Kauniainen and Tuusula (Cluster 1) will suffer from intense competition due to oversupply.

REFERENCES

Greater Helsinki. *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Greater_Helsinki

Foursquare Developers Documentation. *Foursquare*. Retrieved from https://developer.foursquare.com/docs

APPENDIX

- Cluster 0

	Neighborhood	Fast Food Restaurant	Cluster Labels	Latitude	Longitude
0	Espoo	0.0	0	60.171160	24.932650
1	Helsinki	0.0	0	60.171160	24.932650
2	Hyvinkää	0.0	0	60.171160	24.932650
5	Kerava	0.0	0	60.171160	24.932650
6	Kirkkonummi	0.0	0	60.123680	24.441520
7	Mäntsälä	0.0	0	60.193685	24.952381
9	Pornainen	0.0	0	60.219324	24.986390
10	Sipoo	0.0	0	60.194162	24.939059
12	Vantaa	0.0	0	60.271960	24.808730

- Cluster 1

	Neighborhood	Fast Food Restaurant	Cluster Labels	Latitude	Longitude
3	Järvenpää	0.040000	1	60.471640	25.090750
4	Kauniainen	0.038462	1	60.211810	24.728940
11	Tuusula	0.038462	1	60.240085	24.947802

- Cluster 2

	Neighborhood	Fast Food Restaurant	Cluster Labels	Latitude	Longitude
8	Nurmijärvi	0.010638	2	60.252065	24.867359
13	Vihti	0.011364	2	60.221857	24.840934