Coursera Capstone Project The Battle of Neighborhoods

Youssef Ali

07, 2020

In this project, machine learning was used to estimate a good place for a new coffee business in center of Helsinki.

Contents 1 Introduction 1 2 Data 2 3 Methodology 3 4 Results 4 5 Discussion 5 6 Conclusions

Introduction

In a city of Helsinki, if someone is looking to open a café restaurant, the question is, where would you recommend that they open it? The background of the problem is that in order for a café to be profitable, there must be enough customers, and in order to have enough customers, it is not worth setting up a café in the immediate promixity of existing ones. Let's also make sure that audience is explicitly defined to be the local restaurant entrepreuners in Helsinki and they should care about this problem because the location of the new café has a significant impact on the expected returns.

Data

 A description of the data: the data used to solve this problem is geolocation data collected from FourSquare. Adequate explanation and discussion, with examples, of the data is the following. Data is a single dataframe, containing at least a location of the café. Explanation of the location data is a standard tuple (lat, lng), where lat stands for latitude and lng for longitude. Some other metadata like name, postal code and so on is also collected, but let us discuss that they are not absolutely necessary for the analysis. Example of the data used in analysis is shown in table

Identifier	Name	Shortname	Address	Postalcode	Latitude	Longitude
1	Patisserie Teemu & Markus	Bakery	Yrjönkatu 25	00100	60.167899	24.938190
2	Kaffecentralen	Coffee Shop	Fredrikinkatu 59	00100	60.167580	24.932526
3	La Torrefazione	Café	Mannerheimintie 22	00100	60.170721	24.936158
4	The Ounce	Tea Room	Fredrikinkatu 55	00100	60.167182	24.932993
5	La Torrefazione	Coffee Shop	Aleksanterinkatu 50	00100	60.168877	24.943845

Data will be used in the following way: by knowing the locations of already existing cafes, it's possible to apply unsupervised learning technique like kernel density estimation (KDE) to determine the area of influence of the existing cafes, and start up new café which is not in the area of influence.

Methodology

Heatmap-based kernel density estimation was used. Heatmap was already implemented as plugin for Folium, which was used to visualize data to map. Visualization is shown in figure
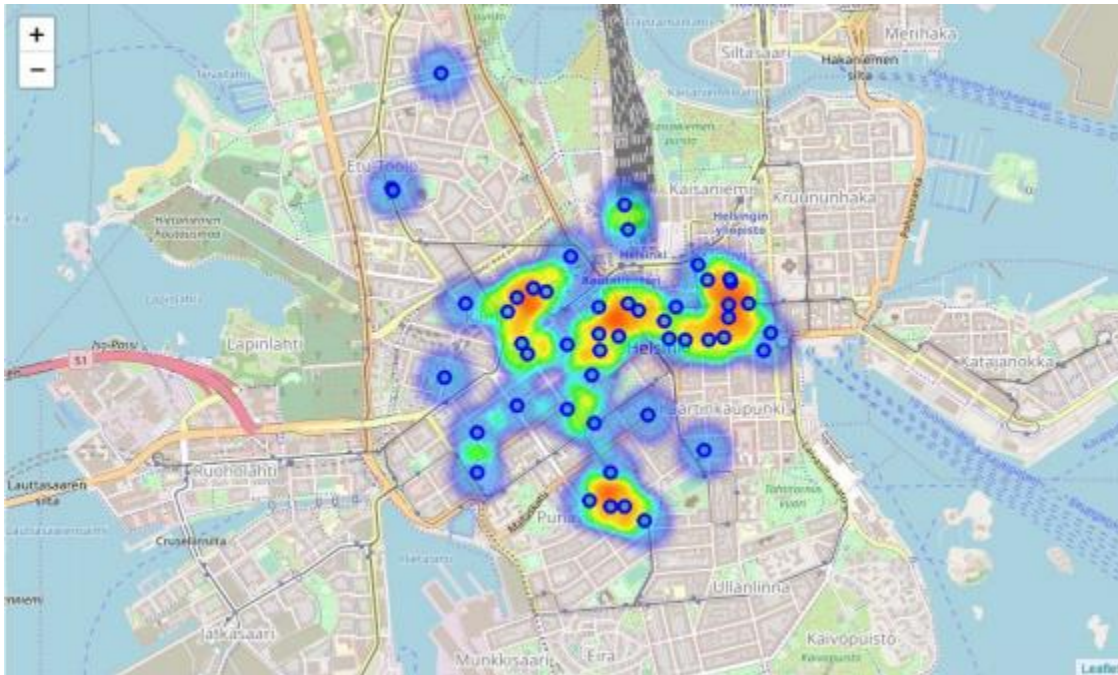
Results

Based on the preliminary results, one possibly good location for new Cafe would be in crossroad of Aleksanterinkatu and Mikonkatu, shown in figure

Conclusions

Optimal location for a new coffee shop in center of Helsinki was estimated based on data from FourSquare.