Coursera Capstone Project
The Battle of Neighborhoods

Aditya Rao

March 06, 2020

Abstract

In this project, machine learning was used to estimate a good place for a new coffee shop in New York City.

Contents

1
Introduction

In New York City, 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 for a café to be profitable, it is not wise setting up a café in the immediate proximity of existing ones.

Let’s also make sure that audience is explicitly defined to be the local restaurant entrepreneurs in New York, and they should care about this problem because the location of the new café is very important in terms of profits.

2
Dataset

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 1.

	uid	name	shortname	postalcode	lat	lng
0	49dbfd0ff964a520385f1fe3	La Maison du Macaron	Bakery	10011	40.743427	-73.993997
1	5b819467598e64002c8b3db7	Gotan	Café	10011	40.740699	-73.993268
2	5a5e29ae0868a24e8b9a8f1e	Maman	Café	10010	40.743626	-73.990211
3	3fd66200f964a52084e51ee3	Eisenberg's Sandwich Shop	Sandwiches	10010	40.741013	-73.989983
4	4a85d797f964a5206aff1fe3	Johny's Luncheonette	Diner	10001	40.744626	-73.992913

Table 1: Five first rows of data used in the machine learning algorithm.

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.

3

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 1.

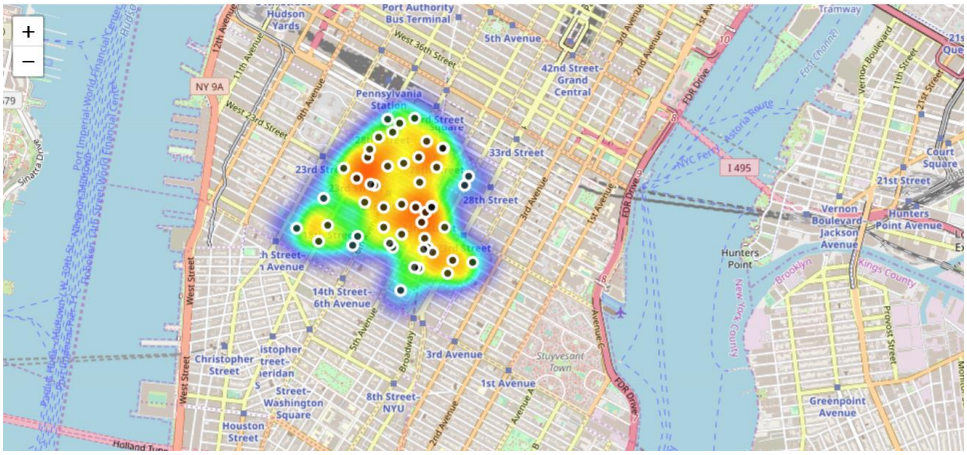

Figure 1: Data visualized to the map of New York, including heatmap-based kernel density estimation.

4

Results

Based on the preliminary results, one possibly good location for new Cafe would be on West 22nd Street, New York City.

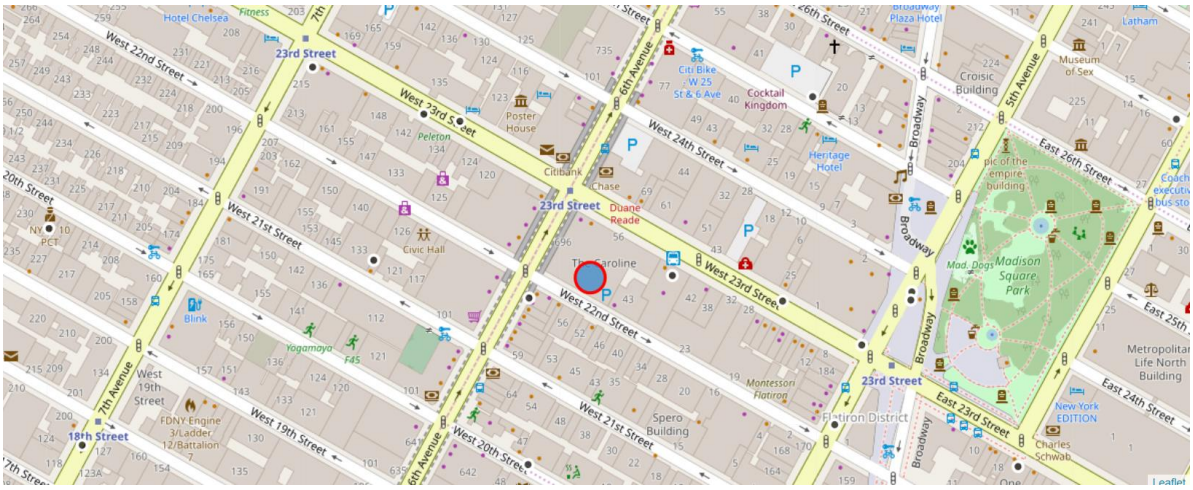

Figure 2: Proposed location for a new café restaurant.

5

Discussion

Before starting a business, some further data analysis of the optimal location of shop may be required.

6
Conclusions

Optimal location for a new coffee shop in New York was estimated based on data from Foursquare.