Coursera capstone project

The Battle of Neighborhoods

Manish Chauhan

May 06, 2019

Abstract

In this project, machine learning was used to estimate a good place for a new coffee business in the center of Noida, Uttar Pradesh, India.

Contents

1. Introduction

In the city of Noida, 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 café in the immediate proximity of existing ones.

Let’s also make sure that audience is explicitly define to be the local restaurant entrepreneurs in Noida and they should care about this problem because the location of the new café has significant impact on the expect returns.

2. Data

A description of the data: the data used to solve this problem is geolocation data collected from FourSquare. Adequate explanation and discussion, with example, of the data the following. Data is a single data frame, containing at least a location of the café. Explanation of the location data is a standard tuple (lag ,lng), where lat stands for latitude and lng for longitude .Some other meta 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 the analysis is shown in table 1.

Identifier	Name	Shortname	Address	Poastalcode	Latitude	Londitude
4ebfa65377c885a64e5f9052	Abhishe	Café	TGIP	201301	28.570080	77.3237
510d017be4b0dc1da493a09e	kaffiiaa	Café	Sector-18	201301	28.568715	77.3242
4c13d982127f9521d8c02425	teasta	Tea room	16 complex	201301	28.565530	77.3400
5b5203c4b9a5a8002ce00945	starbuck	Coffee shop	DLF	201301	28.567397	77.3207
4c6d43bf1585ef3bac000a9e	Café coffee day	Coffee shop	GIP-mall	201301	28.567525	77.3252

Table1. Five first rows of data used in the machine learning algorithm.

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

3. Methodology

Heat map –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.

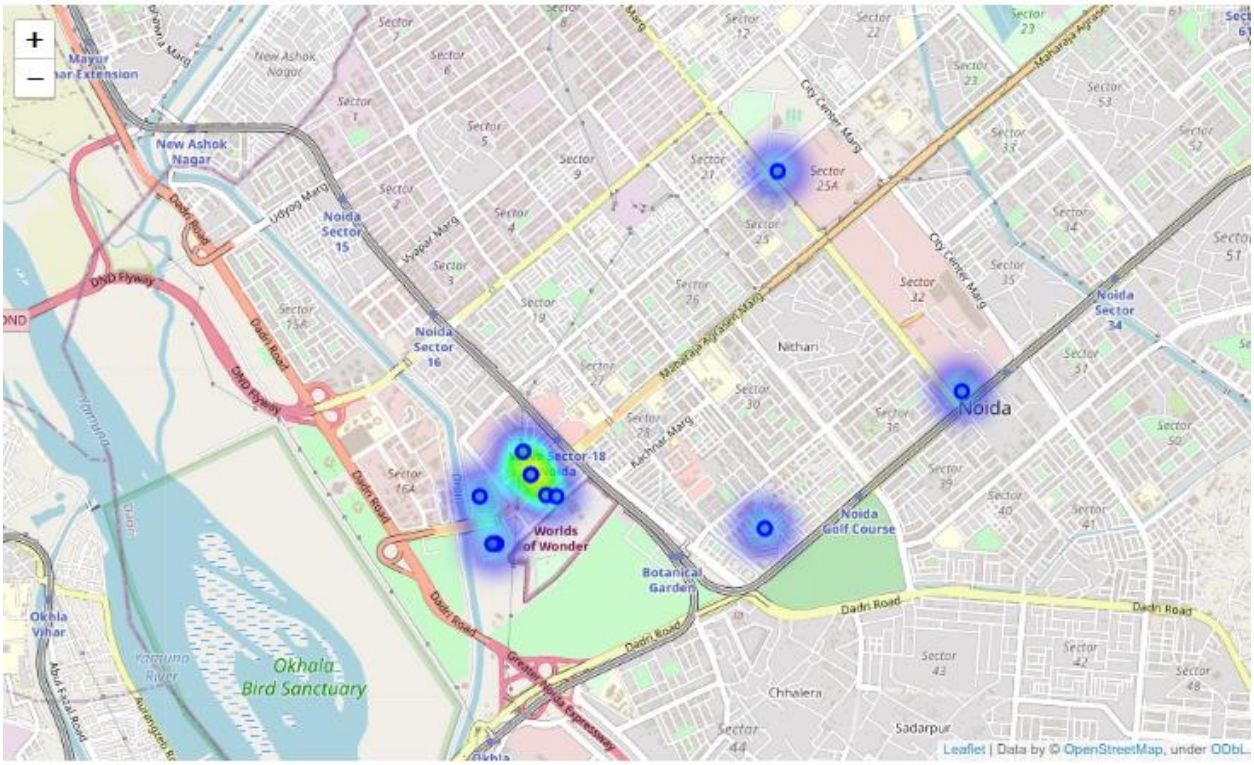

Figure 1: Data visualized to the map of Noida, including heat map -based kernel density estimation.

4. Result

Based on the preliminary result ,one possibly good location for new café would be in crossroad of Maharaja Agrasen Marg. As shown in the figure 2.

Figure 2: purposed location for a new café restaurant.

5. Discussion

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

6. Conclusion

Optimal location for a new coffee shop in the center of Noida was estimated based on data from Foresquare.