Coursera Capstone Final Project

Amer Nureddin

10 Aug 2020

Contents

Abstract

This project is about utilizing Data Science and Machine learning tools to estimate a good location to open a new coffee shop in the heart of the city of Istanbul in Turkey.

Introduction

When someone is planning to open a café restaurant in the city of Istanbul, the question is, where would you suggest that they open it? The backdrop to the issue is that in order for a café to be competitive, there must be enough customers and it is not worth opening up a café in the immediate proximity of existing ones to have enough customers.

Let us also make sure that the audience is clearly identified as Istanbul's local restaurant businessmen and they should be concerned about this problem as the position of the new café has a big effect on the planned returns.

Data

Data description: Geolocation data obtained from FourSquare is the data used to solve this problem. Adequate description and summary of the results is as follows, with illustrations. Data is a single data frame which includes at least one café location. The positioning data is described as a regular tuple (lat, lng), where lat stands for latitude and lng for longitude.

This also gathers some other metadata such as name, postal code and so on, but let us discuss that they are not relevant for the study. Table 1 shows example of data used in the analysis.

	uid	name	shortname	address	postalcode	lat	lng
0	5d9268cd69aaf30008fff5c1	Dükkan Galata	Café	Müeyyedzade Mah. Tatarbeyi Sok. 15/B Galata-Be...	34425	41.025881	28.975390
1	4ba63359f964a5200b3b39e3	Vefa Bozacısı	Café	Mollahüsrev Mah. Vefa Cad. No:66	34134	41.015260	28.958504
2	524a795011d20ce6648c5579	Velvet Cafe	Café	Kule Çıkmazı 7A Galata Beyoğlu	34420	41.025631	28.973714
3	559922f8498e7e8184a87be5	Cumbalı Kahve	Coffee Shop	Ayvansaray mahallesi Kürkçü Çeşmesi Sokak No:1...	34087	41.032992	28.945756
4	59b8dea63149b905ced9edc5	Hamur İşi Cafe & Pastane	Coffee Shop	Haraççıbaşı Sok. No:6A	34083	41.026882	28.955411

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

Data can be used as follows: by understanding where already existing cafes are located, it is possible to use unsupervised learning techniques such as kernel density estimation (KDE) to assess the area of influence of existing cafes and to launch new cafés that are not in the area of influence.

Methodology

Estimation of kernel density based on heatmap was employed. Heatmap has already been deployed as a Folium plugin and was used to visualize mapping data. Visualization as shown in Figure 1.

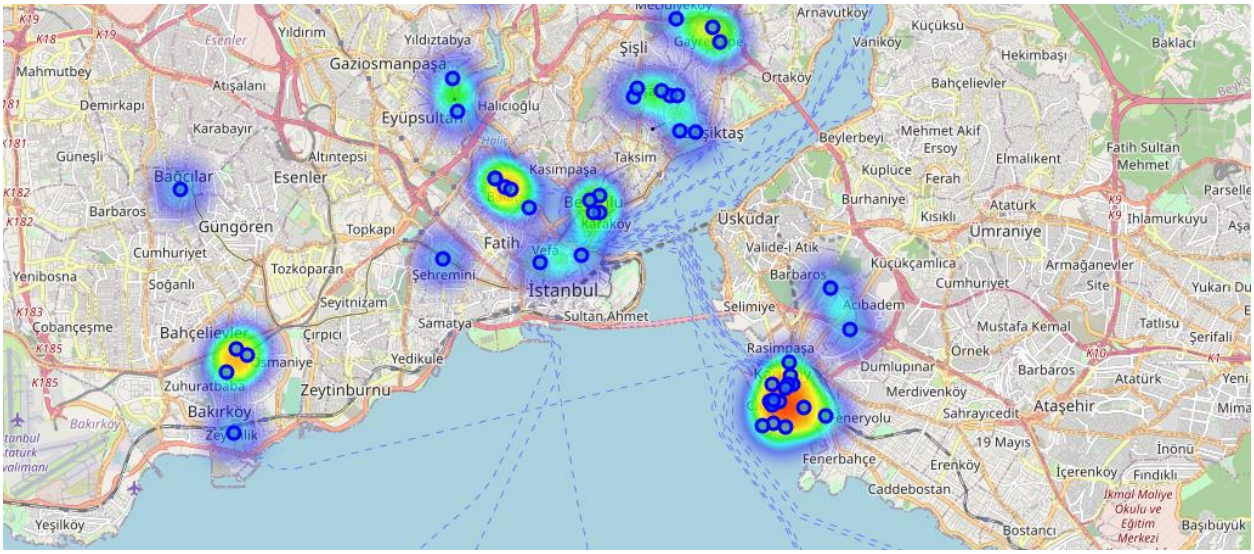

Figure 1: The Data visualized to the map of Istanbul, including heatmap-based kernel density estimation

Results

Based on the preliminary results, on the seaside of Kdikoy area, as seen in Figure 2, could be a good place for the new Cafe.

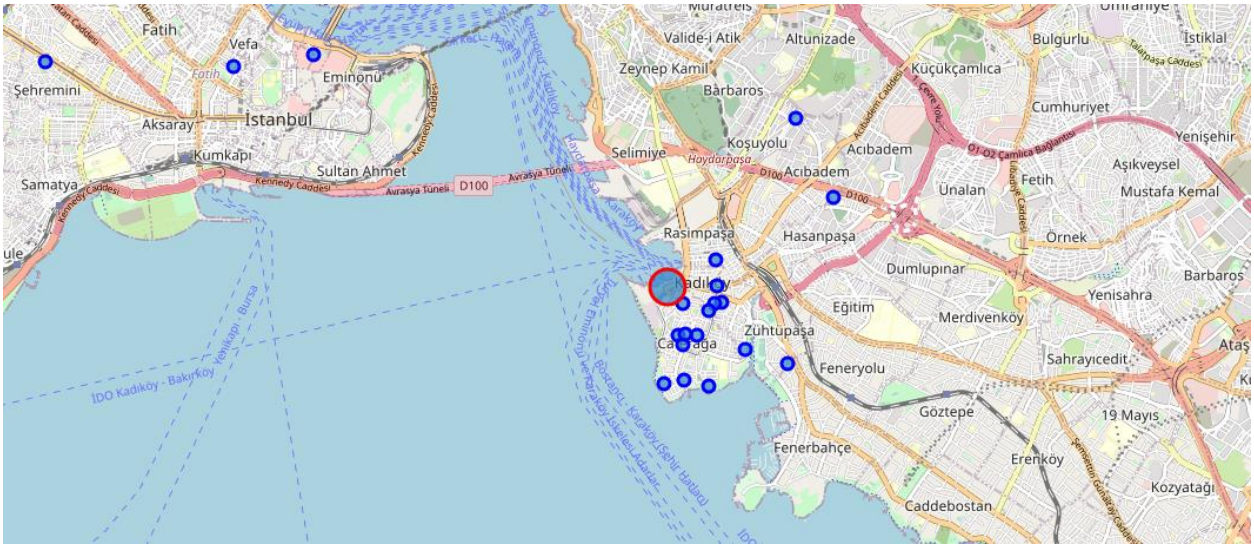

Figure 2: Proposed site for a new café restaurant.

Discussion

Some more data analysis of the optimum position of the shop may be needed before starting the business. Because choosing the best location is a crucial factor for the success of the business.

Conclusion

Based on data from FourSquare, optimal location for a new coffee shop in central Istanbul was estimated.