Coursera Capstone Project

The Battle of Neighborhoods

Coffee Shop in Milan

By Carmelo Ettore Viscosi

Summary

Introduction 3
Business Problem 3
Data Description 3
Methodology 4
Results 5
Discussion 6
Conclusion 6

Introduction

Milan is the economic capital of Italy, it is also one of the most beautiful city of world, it newer sleep and there is a lot of things to do every hours. Every day a lot of tourist, after a long walk, would like to take a rest and drink a good Italian espresso.

Therefore a good business opportunity is to open a coffee shop near one of the most famous places to see, but in an location that offer growth possibility.

The question is where would you recommend to open it? The business problem is that in order for a coffee shop to be profitable, there must be enough customers. It is not worth to open a new coffee shop in the immediate proximity of existing ones.

Business Problem

The goal of this project is to analyze and select the best locations in Milan center to open a new coffee shop. Using data science methodology to provide business solution.

Let's also make sure that audience is explicitly defined to be the local entrepreneur in Milan and they care about this problem because the location of the new coffee shop has the significant impact on its revenue and they want this to be a successful one.

Data Description

The data used to solve this problem is geolocation data collected from FourSqure.com. Foursquare has one of the largest Database of 105+ million places and used by over 125,000 developers. Foursquare API will provide many categories of the venue data, it also offers travel tip and feedback of other users.

We focus our analysis on the coffee shop category in order to help us to solve the business problem. Adequate explanation and discussion of the data is the following. Used data - visible in fig.1 - is a single dataframe that containing a list of coffee shop, its location and a short description of what if offer. Every row of dataframe represents a single coffee shop, it is identified by a standard tuple (lat, lng), where lat stands for latitude and lng for longitude. Some other metadata such as name, postal code, and etc., are also collected.

	uid	name	shortname	address	postalcode	lat	lng
0	58b5d3364e31c52f888a0bed	Starbucks Reserve Roastery	Coffee Shop	[Piazza Cordusio 3 (Via Orefici), 20123 Milano...	20123	45.464920	9.186153
1	5485c0d7498e225e0587b79b	Panini Durini	Sandwiches	[Via Mengoni 4, 20121 Milano Lombardia, Italia]	20121	45.465238	9.188590
2	59c2232bd48ec17dd4178369	Lavazza Coffee Design	Coffee Shop	[Piazza San Fedele, 2, 20122 Milano Lombardia,...	20122	45.466274	9.190975
3	576828cc498ec1d7fb395d70	Caffè Napoli	Café	[Via Gaetano Giardino 1, 20123 Milano Lombardi...	20123	45.462816	9.189023
4	4be181c240d676b02de504ee	Princi	Bakery	[Via Speronari, 6, 20123 Milano Lombardia, Ita...	20123	45.463029	9.187968

Fig.1 Sample of coffee shop dataset

Data will be used as follow – by knowing the locations of the already existing coffee shops. It is possible to apply unsupervised learning technique like kernel density estimation (KDE) to determine the area of influence of the existing coffee shops.

Methodology

We will use Foursquare API to get the top 100 venues that are within a radius of 1000 meters. We need to have a Foursquare Developer Account and use our Foursquare ID and Secret Key to obtain the data. We make an API calls to Foursquare passing the type of shop that we want to find in the neighborhoods of center. Foursquare will return the venue data in JSON format and we will extract interested features such as the venue name, venue category, venue address, venue latitude and venue longitude. With the data, we can filter the “Coffee Shop” as venue category for the neighborhoods and save it to pandas dataframe to be passed on folium for map building.

Also, Heatmap-based kernel density estimation was being used, Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample. Heatmap was already implemented as plugin for Folium which was used to visualize data on the map visible in Fig.2.

Fig.2 Obtained map of center and its coffee shop

Results

Based on the map result we can locate as a good location for new coffee shop the south-west side of Duomo Cathedral where which is an entrance to center and where is a lot of shop and office.

We insert a point on folium map where the location of new coffee can be, you'll see it on Fig. 3

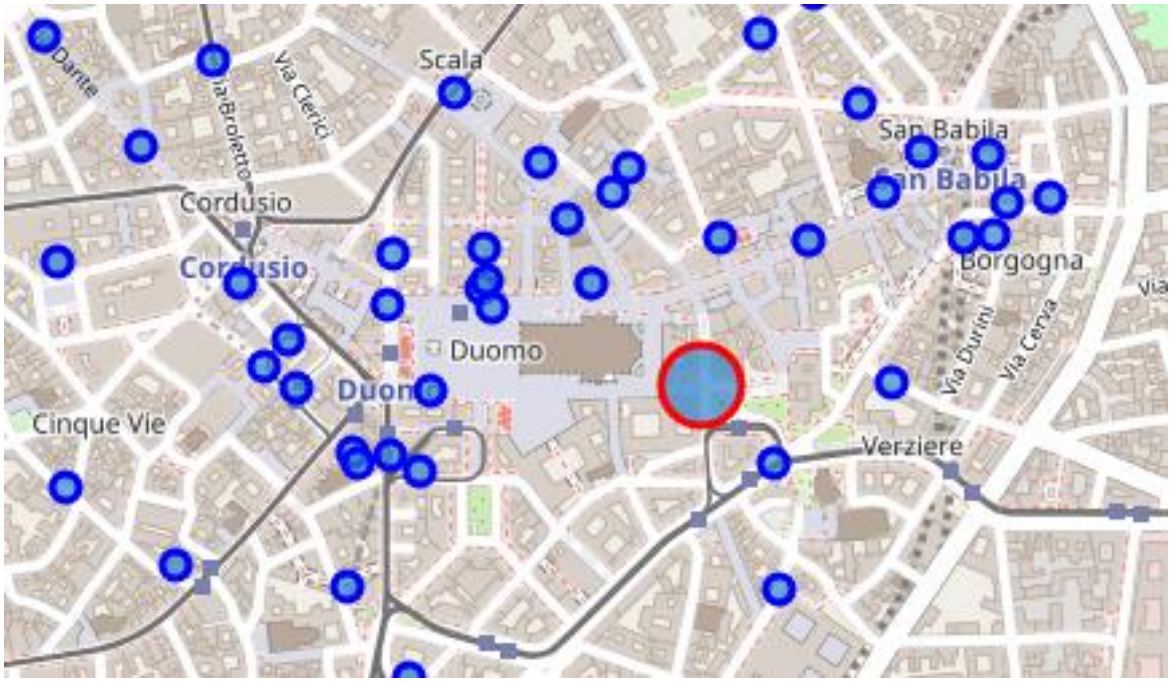

Fig.2 Display a possible location of new coffee shop

Discussion

With the help of the heatmap-based kernel density estimation in the Methodology section, you see that most of the coffee shops are concentrated on north-east side. There is no coffee shop on Piazza Fontana and in the nearby. This represents a great opportunity and high potential area to open a new coffee shop as there is very little to no competition from existing coffee shop. In that location tourist can enjoy a good Italian espresso and relax a bit after a long day of sightseeing in this beautiful place.

Conclusion

The best location for a new coffee shop was estimated based on the data from FourSquare.

It is recommended to the management team to re-run this data science program to get the updated result and use the result into consideration as part of the business growth plan in selecting the new neighborhood to offer their new coffee shop.

This is critical because in a big city like Milan, every day a new coffee shop can open or in a bad lucky situation can close.