Milan, Where open a pizza shop?

Introduction

Pizza is one of the most famous Italian dishes. And there are so many good reasons why! As can see from previous research (i.e. Turrini et.al. 2001, Di Vita et.al. 2016), pizza is an important food for Italian's diet. And due to the famous of the pizza, almost all the tourist that came to Italy want to taste "The true Italian Pizza".

Milan is the second most populated city in Italy, with almost 1,300,000 habitants and have the biggest airport in Italy. Is also, the Italian city with the best public transport and a great transport connection network. All these, make that Milan offers then a huge market for selling pizza. But, of course, this is not something new for anyone in Italy.

Never the less, choose the correct place to open a Pizza shop is quite difficult, science there are many variables that can affect the success or failed of a food business (i.e., quality, price, location). Based on the free available information this problem will address this problem in order to have different cluster neighborhood and see the characteristics of the same to define which is the best option to open the pizza shop.

This project is oriented to help an investment, that will decide in which Milan district open a new pizza shop.

Data

Based on definition of our problem, factors that will influence our decision are:

* number of existing restaurants in the district (any type of restaurant)

* number of and distance to Pizza Shop in the district, if any

* distance of neighborhood from city center

Data sources:

To choose the best place for open a pizza shop we will consider, each district (148) and according to the division that can be found in https://it.wikipedia.org/wiki/Municipi_di_Milano#cite_note-3.

#	Denominazione	Superficie (km²)	Abitanti (31.12.2015)	Densità (ab/km²)	Quartieri compresi[4]
Municipio 1	Centro storico	9,67	97.403	10.072,70	Cordusio, Cinque Vie, Carrobbio, Verziere, Pasquirolo, Borgonuovo, Crocetta, Quadronno, Vetra, Sant'Ambrogio, Brera, Guastalla, Porta Tenaglia, Porta Sempione, Conca del Naviglio
Municipio 2	Stazione Centrale, Gorla, Turro, Greco, Crescenzago	12,58	159.134	12.649,76	Stazione Centrale, Loreto, Turro, Crescenzago, Quartiere Adriano, Gorla, Precotto, Ponte Seveso, Quartiere Maggiolina, Mirabello, Villaggio dei Giornalisti, Greco
Municipio 3	Città Studi, Lambrate, Venezia	14,23	142.939	10.044,91	Porta Venezia, Porta Monforte, Casoretto, Rottole, Quartiere Feltre, Cimiano, Città Studi, Lambrate, Ortica
Municipio 4	Vittoria, Forlanini	20,95	159.750	7.625,29	Porta Vittoria, Porta Romana, Cavriano, Quartiere Forlanini, Monluè, La Trecca, Taliedo, Morsenchio, Ponte Lambro, Calvairate, Gamboloita o Corvetto, Quartiere Omero, Nosedo, Castagnedo, Rogoredo, Santa Giulia, San Luigi, Triulzo Superiore
Municipio 5	Vigentino, Chiaravalle, Gratosoglio	29,87	124.903	4.181,55	Porta Vigentina, Porta Lodovica, San Gottardo, Morivione, Vigentino, Vaiano Valle, Chiaravalle, Macconago, Stadera, Chiesa Rossa, Quartiere Le Terrazze, Case Nuove, Quartiere Missaglia, Gratosoglio, Selvanesco, Quintosole, Ronchetto delle Rane, Quartiere Torretta, Conca Fallata, Quartiere Basmetto, Quartiere Cantalupa
Municipio 6	Barona, Lorenteggio	18,28	150.356	8.225,16	Porta Ticinese, Porta Genova, Conchetta, Moncucco, Barona, Quartiere Sant'Ambrogio, Quartiere Cascina Bianca, Boffalora, Quartiere Teramo, San Cristoforo, Quartiere Lodovico il Moro, Ronchetto sul Naviglio, Quartiere Villa Magentino, Arzaga, Giambellino, Lorenteggio, Villaggio dei Fiori, Creta
Municipio 7	Baggio, De Angeli, San Siro primaticcio	31,34	173.643	5.540,62	Porta Magenta, Quartiere De Angeli - Frua, San Siro, Quartiere Harar, Quartiere Vercellese, Quarto Cagnino, Quinto Romano, Figino, Molinazzo, Sella Nuova, Baggio, Quartiere Valsesia, Quartiere degli Olmi, Assiano, Muggiano
Municipio 8	Fiera, Quartiere Gallaratese, Quarto Oggiaro	23,72	186.179	7.849,03	Porta Volta, Bullona, Ghisolfa, Portello, Cagnola, Quartiere Campo dei Fiori, Villapizzone, Quartiere Varesina, Boldinasco, Garegnano, Certosa, Musocco, Quarto Oggiaro, Vialba, Roserio, Cascina Triulza, Q.T.8, Lampugnano, Quartiere Gallaratese, Quartiere San Leonardo, Trenno, Cascina Merlata.
Municipio 9	Stazione Garibaldi, Niguarda	21,12	186.566	8.833,62	Porta Garibaldi, Porta Nuova, Centro Direzionale, Isola, La Fontana, Montalbino, Segnano, Bicocca, Fulvio Testi, Ca' Granda, Pratocentenaro, Niguarda, Dergano, Bovisa, Affori, Bruzzano, Quartiere Comasina, Quartiere Bovisasca
	Totale comune	**181,76**	**1 380 873**	**7 597,23**	

Screen shot of Wikipedia information.

Then we will obtain the coordinates of each using MapQuest Open APIs, that offers a free API key to obtain address information based on coordinates or coordinates based on address specification, like the districts https://developer.mapquest.com/documentation/open/.

Then we will obtain the number of pizza shops from Foursquare venues and all the type of restaurants per district. This will allow us to related the number of habitants with the amount of pizza shops and restaurants. Also, to know if the pizza is the venue category that is most repeated in that district or in which category it is.

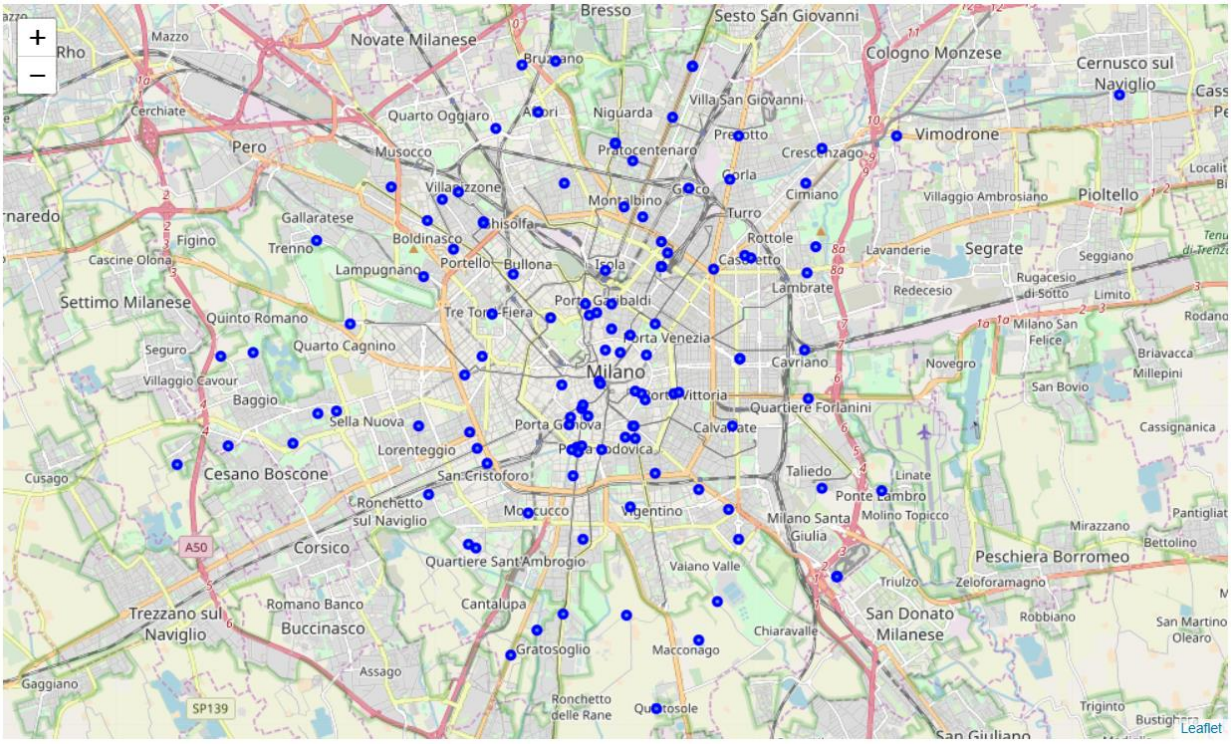
All this will allow us to make the cluster of the districts and analyzed which district is more susceptible to the open of a new pizza shop.

Districts Candidates

Let’s choose search for all the Milan's districts from Wikipedia

	index	Borough	Names	Area_Km2	Population	Pop_Density_hab_km2	Districts
0	0	Municipio 1	Centro storico	9.67	97403	10072.70	Cordusio
1	1	Municipio 2	Stazione Centrale, Gorla, Turro, Greco, Cresce...	12.58	159134	12649.76	Stazione Centrale
2	2	Municipio 3	Città Studi, Lambrate, Venezia	14.23	142939	10044.91	Porta Venezia
3	3	Municipio 4	Vittoria, Forlanini	20.95	159750	7625.29	Porta Vittoria
4	4	Municipio 5	Vigentino, Chiaravalle, Gratosoglio	29.87	124903	4181.55	Porta Vigentina

Getting the coordinates from MapQuest and plot the data.



Methodology

In this project we will direct our efforts on detecting districts of Milan that have low restaurant density, particularly those with low number of pizza shop, and see how many habitants we have per pizza shop. We will limit our analysis to the Milan districts according, to Wikipedia information.

In first step we have collected the data: All the Milan's districts, their location and their population density, using the Wikipedia information Beautiful Soup, MapQuest, pandas and folium.

Second, we will calculate the population per district, get location and type (category) of every restaurant type (category) of every restaurant per district. We have also identified pizza shop (according to Foursquare categorization).

Third step in our analysis will be calculation and exploration of restaurant density across different districts of Milan, we will use heatmaps to identify a few promising areas with low number of pizza shop and an important number of costumers and focus our attention on those areas.

In finally we will focus on most promising districts and within those create clusters of locations that meet some basic requirements: we will take into consideration locations with less restaurants in radius of 250 meters, and we want locations without Italian restaurants in pizza shops of 500 meters. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

Analysis

Let's see how many Districts we have per Borough:

	index	values
0	Municipio 8	22
1	Municipio 5	21
2	Municipio 4	18
3	Municipio 6	18
4	Municipio 9	18
5	Municipio 7	15
6	Municipio 1	15
7	Municipio 2	12
8	Municipio 3	9

We assume that each district has the same size inside each Borough, and according to this we can calculate the population inside each district.

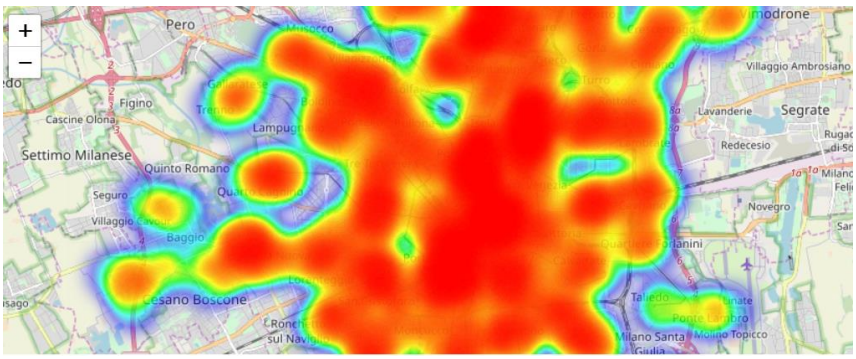
	index	Borough	Names	Area_Km2	Population	Pop_Density_hab_km2	Districts	Latitude	Longitude
0	0	Municipio 1	Centro storico	0.644667	6493	10072.70	Cordusio	45.4649	9.18589
1	1	Municipio 2	Stazione Centrale, Gorla, Turro, Greco, Cresce...	1.048333	13261	12649.76	Stazione Centrale	45.4645	9.18644
2	2	Municipio 3	Città Studi, Lambrate, Venezia	1.581111	15882	10044.91	Porta Venezia	45.4759	9.20127
3	3	Municipio 4	Vittoria, Forlanini	1.163889	8874	7625.29	Porta Vittoria	45.4625	9.19753
4	4	Municipio 5	Vigentino, Chiaravalle, Gratosoglio	1.422381	5947	4181.55	Porta Vigentina	45.4539	9.19604

Foursquare

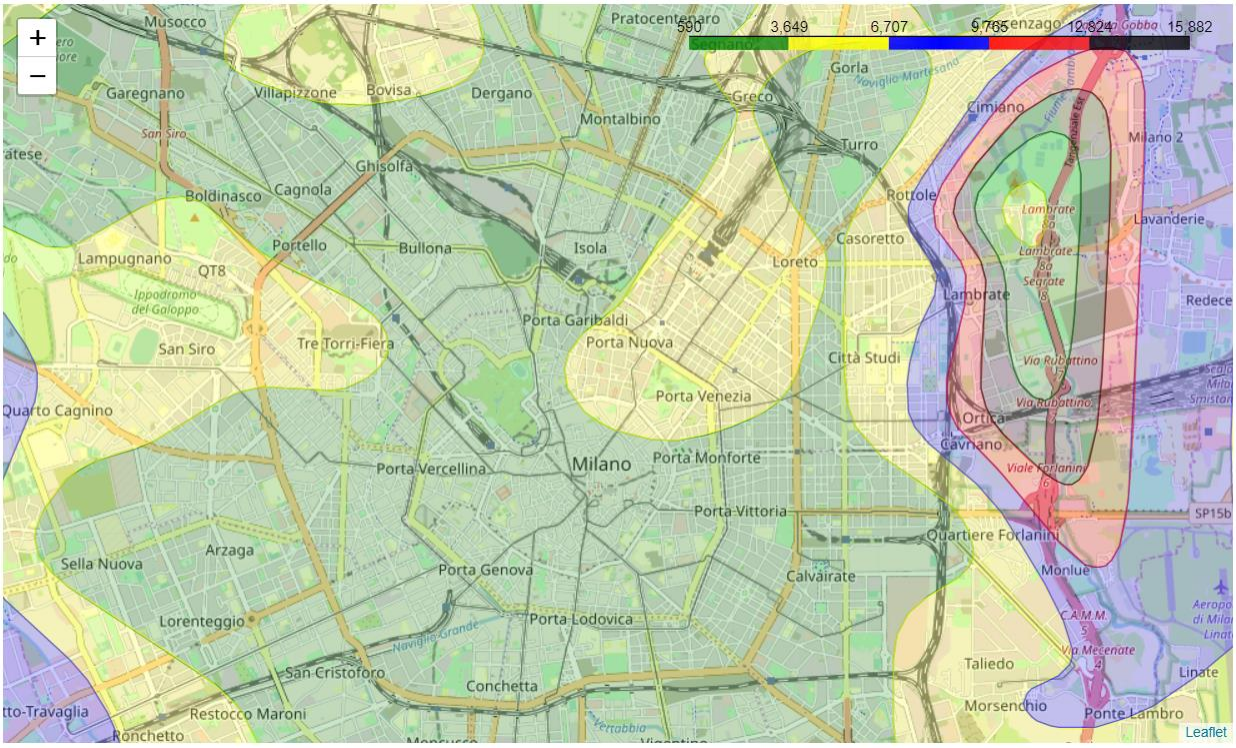
We're interested in venues in 'food' category, obtaining:

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Short Category
0	Cordusio	45.46491	9.18589	Starbucks Reserve Roastery	45.464920	9.186153	Coffee Shop	Coffee Shop
1	Cordusio	45.46491	9.18589	Ciacco. Gelato senz'altro	45.463704	9.186796	Ice Cream Shop	Ice Cream
2	Cordusio	45.46491	9.18589	KFC	45.464281	9.187572	Fast Food Restaurant	Fast Food
3	Cordusio	45.46491	9.18589	Panini Durini	45.464299	9.187460	Sandwich Place	Sandwiches
4	Cordusio	45.46491	9.18589	Granaio Caffé e Cucina	45.465692	9.188366	Italian Restaurant	Italian
5	Cordusio	45.46491	9.18589	McDonald's	45.464408	9.188457	Fast Food Restaurant	Fast Food
6	Cordusio	45.46491	9.18589	Granaio Caffè e Cucina	45.465794	9.185562	Italian Restaurant	Italian
7	Cordusio	45.46491	9.18589	Motta Caffè Bar Milano 1928	45.464657	9.190182	Café	Café
8	Cordusio	45.46491	9.18589	B Café	45.462640	9.183381	Café	Café

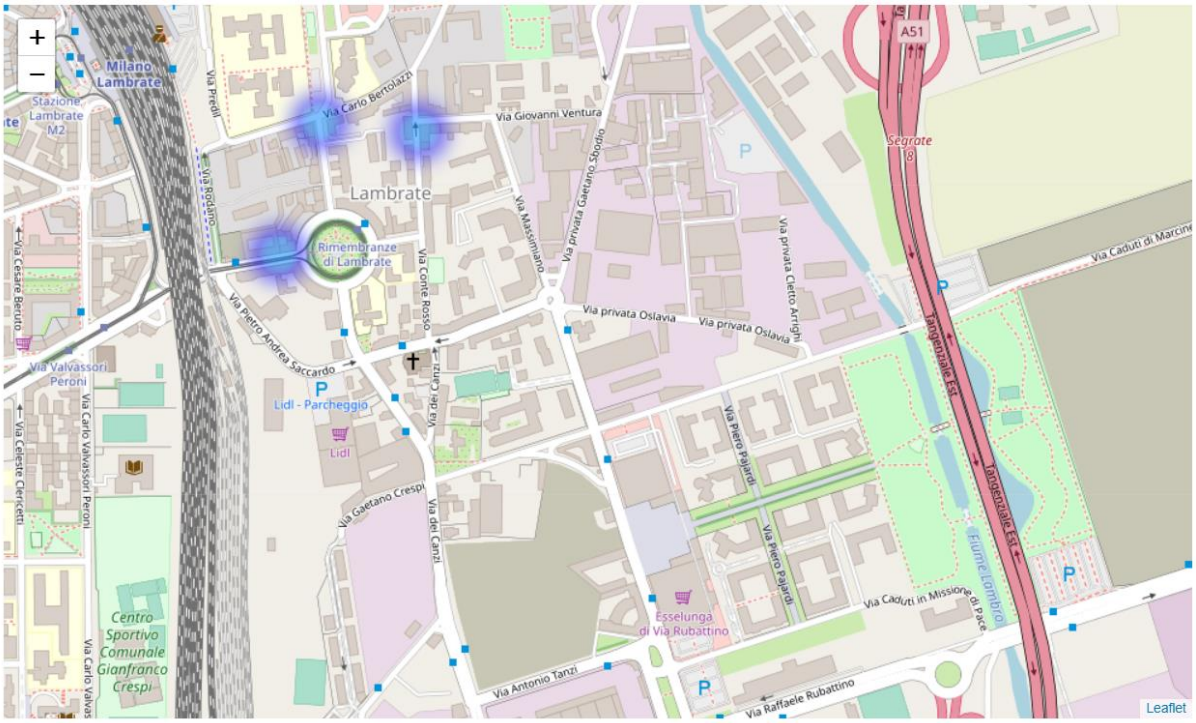
The heat map shoe us:



Never the less this is not completely clear, so let's see how many costumers could we have per districts according to the quantity of pizza shops per district:



From these, we let's focus in the area of Lambrate which have an important train station and a lot of people living there, defining a new area. The heat map shows us:

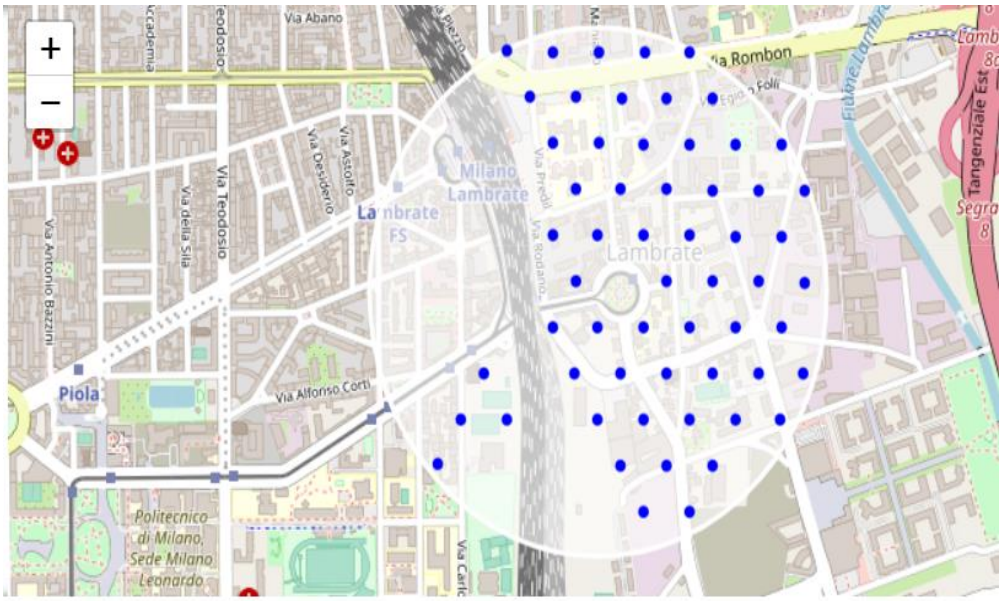


Let's also create new, denser grid of location candidates restricted to our new region of interest (let's make our location candidates 100m apart). For this we need to change our coordinate system to UTM, so let's start with the Lambrate venues.

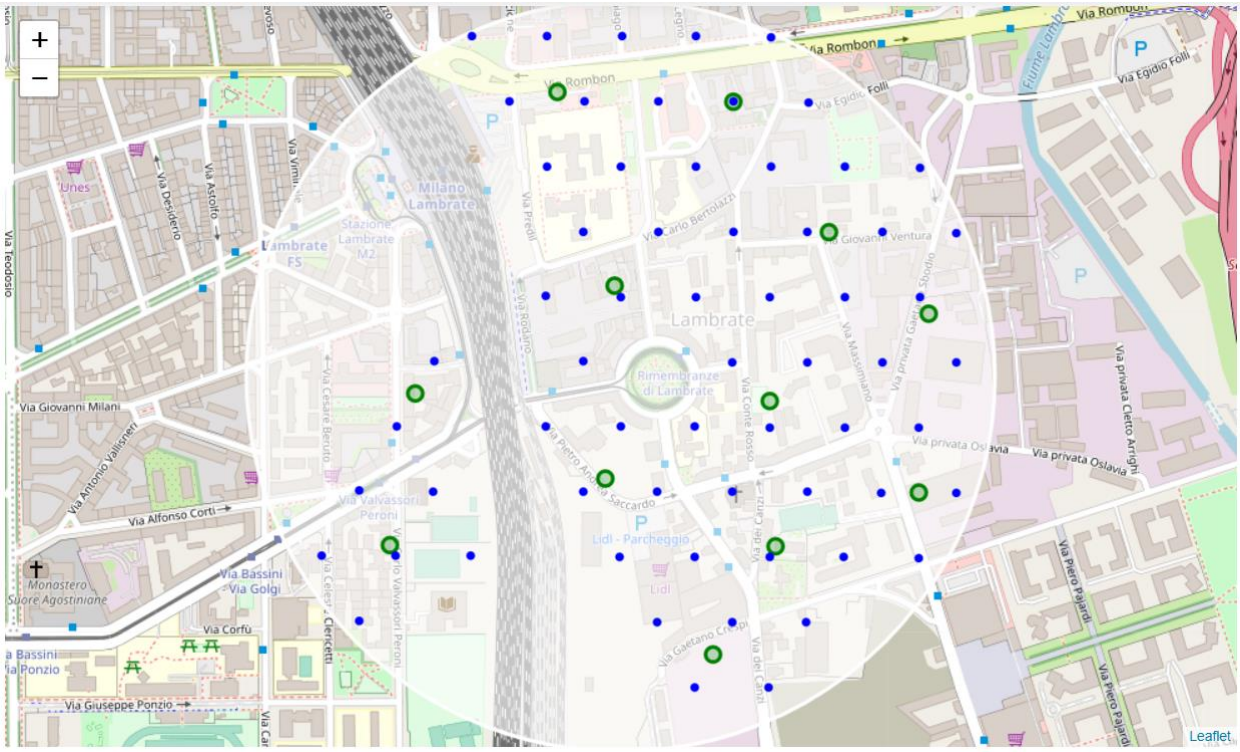
	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Short Category	District_X	District_y	Venue_X	Venue_Y
24	Rimembranze di Lambrate, Milano, Italia	45.48246	9.24042	Kebap Pizza.it	45.483186	9.231525	Pizza Place	Pizza	518789	5.03658e+06	518094	5.03665e+06
31	Rimembranze di Lambrate, Milano, Italia	45.48246	9.24042	Pizzeria Spontini	45.482009	9.213112	Pizza Place	Pizza	518789	5.03658e+06	516655	5.03652e+06
36	Rimembranze di Lambrate, Milano, Italia	45.48246	9.24042	'A Tarantella	45.490889	9.233899	Pizza Place	Pizza	518789	5.03658e+06	518277	5.03751e+06
38	Rimembranze di Lambrate, Milano, Italia	45.48246	9.24042	La Cuccuma	45.482123	9.228962	Pizza Place	Pizza	518789	5.03658e+06	517894	5.03654e+06
41	Rimembranze di Lambrate, Milano, Italia	45.48246	9.24042	Piccola Ischia	45.481748	9.217081	Pizza Place	Pizza	518789	5.03658e+06	516965	5.03649e+06
42	Rimembranze di Lambrate, Milano, Italia	45.48246	9.24042	Pizzeria Piccolo Borgo	45.482743	9.230598	Pizza Place	Pizza	518789	5.03658e+06	518021	5.03661e+06
44	Rimembranze di Lambrate, Milano, Italia	45.48246	9.24042	Piccola Ischia	45.477781	9.211707	Pizza Place	Pizza	518789	5.03658e+06	516546	5.03605e+06

Let us now filter those locations: we're interested only in locations with no more than two food places in radius of 250 meters, and no Pizza Shop in radius of 300 meters.

- Locations with no more than two restaurants nearby: 90
- Locations with no Pizza Shops within 500m: 74
- Locations with both conditions met: 74



Lest make a cluster of the points in order to ger 12 different zones:



- Finally, let get the addresses which can be presented to stakeholders.
- Via Carlo Bertolazzi Milan IT => 127.1m from Rimembranze di Lambrate
- Via Conte Rosso Milan IT => 167.8m from Rimembranze di Lambrate
- Via privata Giovanni Ventura Milan IT => 372.2m from Rimembranze di Lambrate
- Via dei Canzi Milan IT => 340.0m from Rimembranze di Lambrate
- Via Carlo Valvassori Peroni Milan IT => 407.7m from Rimembranze di Lambrate
- Via Riccardo Pitteri Milan IT => 382.2m from Rimembranze di Lambrate
- Via Tommaso Pini Milan IT => 388.0m from Rimembranze di Lambrate
- Via privata Gaetano Sbodio Milan IT => 400.0m from Rimembranze di Lambrate
- Via Rombon Milan IT => 401.7m from Rimembranze di Lambrate
- Via Pietro Andrea Saccardo Milan IT => 137.7m from Rimembranze di Lambrate
- Piazza Enrico Bottini Milan IT => 362.5m from Rimembranze di Lambrate
- Milan Milan IT => 300.9m from Rimembranze di Lambrate

This concludes our analysis. We have created 12 addresses representing centers of zones containing locations with low number of restaurants and no pizza shop nearby.

Results and Discussion

Our analysis shows that although there is a great number of pizza shops in Milan's districts, there are pockets of low pizza shop density. We focused our attention to Lambrate area. Our attention focused to this area due to the low density of pizza shops and the high density of potential customers.

After directing our attention to this narrower area of interest we first created a dense grid of location candidates (spaced 100m apart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with a Pizza Shop closer than 500m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 12 zones containing largest number of potential new Pizza Shop locations based on the district population, number of and distance to existing venues - both restaurants in general and Pizza Shop particularly. This, of course, does not imply that those zones are actually optimal locations for a new Pizza Shop! Purpose of this analysis was to only provide info on areas in Milan not crowded with existing restaurants (particularly Pizza Shops) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

Conclusion

Purpose of this project was to identify Berlin areas close to center with low number of restaurants (particularly Pizza Shops) in order to aid stakeholders in narrowing down the search for optimal location for a new Pizza Shop. First, we obtained the districts and population from Milan city, after that we get the coordinates of each district, then, by calculating restaurant density distribution from Foursquare data and correlated to the population of each district we choose Lambrate area. Then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.

Bibliography

G. Di Vita, G. De Salvo, S. Bracco, G. Gulisano, M. D'Amico. Future Market of Pizza: Which Attributes Do They Matter?, Agris on-line Papers in Economics and Informatics, Volume VIII, Number 4, pp59-71 2016 https://www.researchgate.net/publication/311966640_Future_Market_of_Pizza_Which_Attributes_Do_They_Matter

A. Turrini, A. Saba, D. Perrone2, E. Cialfa and A. D'Amicis. Original Communication Food consumption patterns in Italy: the INN-CA Study 1994 – 1996. European Journal of Clinical Nutrition (2001) 55, 571-588 https://dietistica.campusnet.unito.it/didattica/att/c8b7.3449.file.pdf