Vegetarian Restaurant in Bologna

Introduction

A new company interested in investment on a vegetarian restaurant is searching a zone in Bologna, a town of about 400.000 inhabitants in Northern Italy, for opening it.

The characteristics of the restaurant in part will depend on the characteristics of the zone.

The problem this reports attempts to solve is: which characteristics should a vegetarian have in order to meet the requirements different customers in different zones in Bologna?

Data

In order to provide an answer, we have collected relevant data from the Bologna Municipality Open Data Site (http://dati.comune.bologna.it/).

Data we have collected are:

- Statistical Areas - Three levels: Borough, Neighbourhood, Stat area
  Source: http://dati.comune.bologna.it/download/file/fid/1275 (headers are translated)

	name	id	id2	sqm	borough_id	borough_name	neighborhood_id	neighborhood_name	type
0	AEROPORTO	4	23	4402575,4660000000	1	BORGO PANIGALE	C	Borgo Panigale	Poli funzionali, aree produttive e infrastrutture
1	AGUCCHI	39	20	1092352,0940000000	4	RENO	S	Santa Viola	Tessuto consolidato e aree della trasformazion...
2	ARCOVEGGIO	14	41	961087,3560000000	2	NAVILE	B	Bolognina	Tessuto consolidato e aree della trasformazion...
3	BARGELLINO	3	6	696718,8840000000	1	BORGO PANIGALE	C	Borgo Panigale	Poli funzionali, aree produttive e infrastrutture
4	BATTINDARNO	37	35	936825,9900000000	4	RENO	A	Barca	Tessuto consolidato e aree della trasformazion...

- Income in 2017 per Statistical Area
  Source: http://dati.comune.bologna.it/download/file/fid/5066

	stat_area_name	tax_payer_nr	total_income_for_state_taxation	total_income_for_coucil_taxation
0	Aeroporto	324	7306571	7130548
1	Agucchi	3203	68382532	65810791
2	Arcoveggio	7819	165245894	159182761
3	Bargellino	68	1483171	1477482
4	Battindarno	6835	150912113	145989868

- Age per Statistical Area - 2018 (12/31?)
  Source: http://dati.comune.bologna.it/download/file/fid/5094

	year	census_2011_section_id	stat_area_id	stat_area_name	borough_name	neighborood_name	sex	age_different_intervals	age_equal_intervals	citizenship	nr_of_residents
0	2018	1	2	Via del Vivaio	Borgo Panigale - Reno	Borgo Panigale	F	02 and before	00-04	Italian	2
1	2018	1	2	Via del Vivaio	Borgo Panigale - Reno	Borgo Panigale	F	03-05	00-04	Italian	1
2	2018	1	2	Via del Vivaio	Borgo Panigale - Reno	Borgo Panigale	F	03-05	00-04	Foreign	1
3	2018	1	2	Via del Vivaio	Borgo Panigale - Reno	Borgo Panigale	F	03-05	05-09	Italian	2
4	2018	1	2	Via del Vivaio	Borgo Panigale - Reno	Borgo Panigale	F	06-14	05-09	Italian	3

- Statistical area cartography
  Source: http://dati.comune.bologna.it/download/file/fid/5161
  This is a shp file. We converted it in geojson using
  https://mygeodata.cloud/converter/shp-to-geojson
  The result on the map is
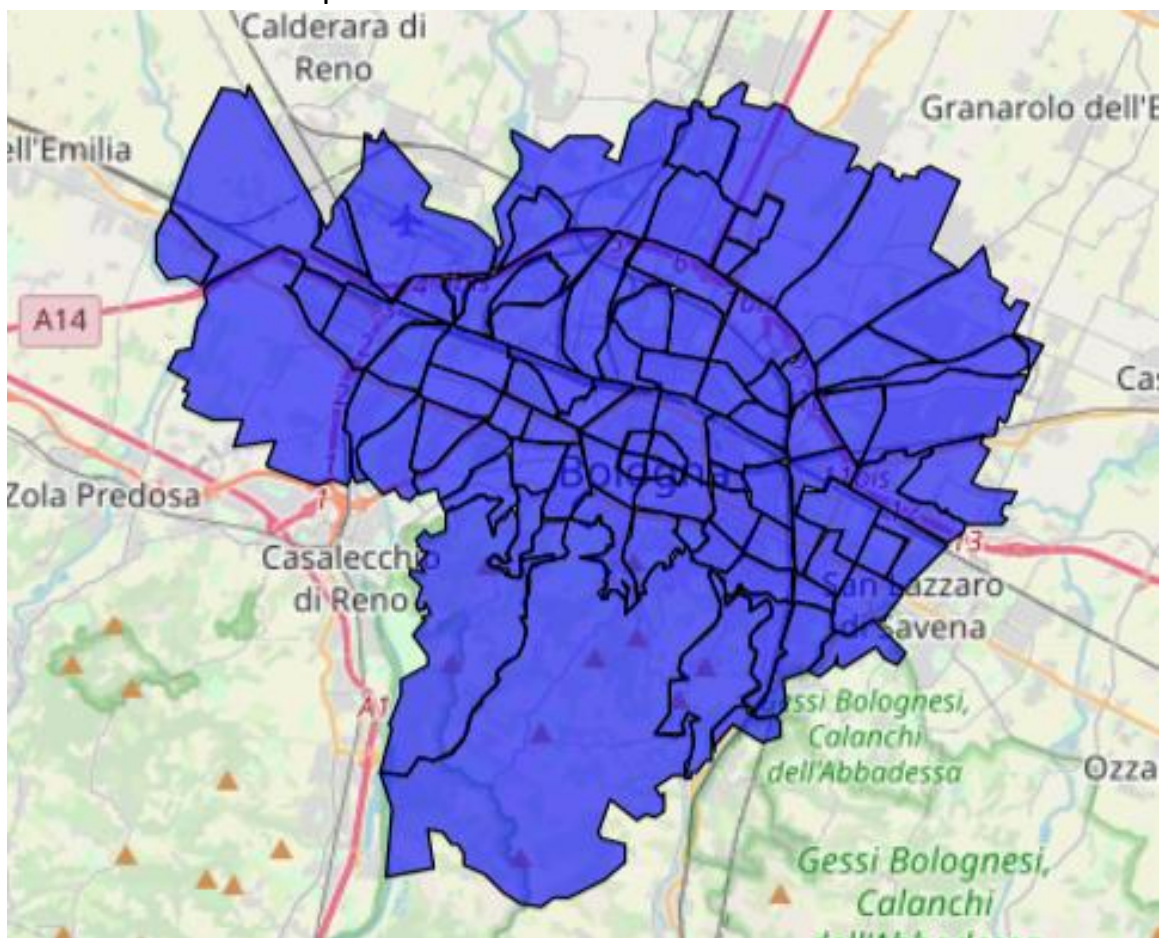
From the map I take the coordinates of the polygons of each area in order to query Foursquare:

	stat_area	coords
0	VILLAGGIO DELLA BARCA	[[11.28887243687619, 44.493839678171874], [11....
1	BATTINDARNO	[[11.304354670908275, 44.502231153585114], [11...
2	CANALE DI RENO	[[11.301304394939708, 44.490825007457346], [11...
3	CASERME ROSSE-MANIFATTURA	[[11.364140469163491, 44.52279866578613], [11....
4	CNR	[[11.34781923054498, 44.524232532194105], [11....

As we said we use Foursquare for collecting venues querying the by categories considered relevant:

Vegetarian/Vegan Restaurants: 4bf58dd8d48988d1d3941735

School: 4bf58dd8d48988d13b941735

College & University: 4d4b7105d754a06372d81259

Office: 4bf58dd8d48988d124941735

Museum: 4bf58dd8d48988d181941735.

Methodology

From tabular data we obtained a dataset with these columns:

- Area name
- Income x taxpayer
- Density
- Percent of people aged 15-29
- Percent of people aged 30-44
- Percent of people aged 45-64
- Percent of people aged 65-74
- We searched Foursquare by category and by rectangle (max sw min ne) for each area. From for each set of result took away the venues outside the polygon.

We then counted the number of venue and got:

- Number of vegetarian restaurants
- Number of offices
- Number of schools
- Number of universities
- Number of museums

Final dataset looks like:

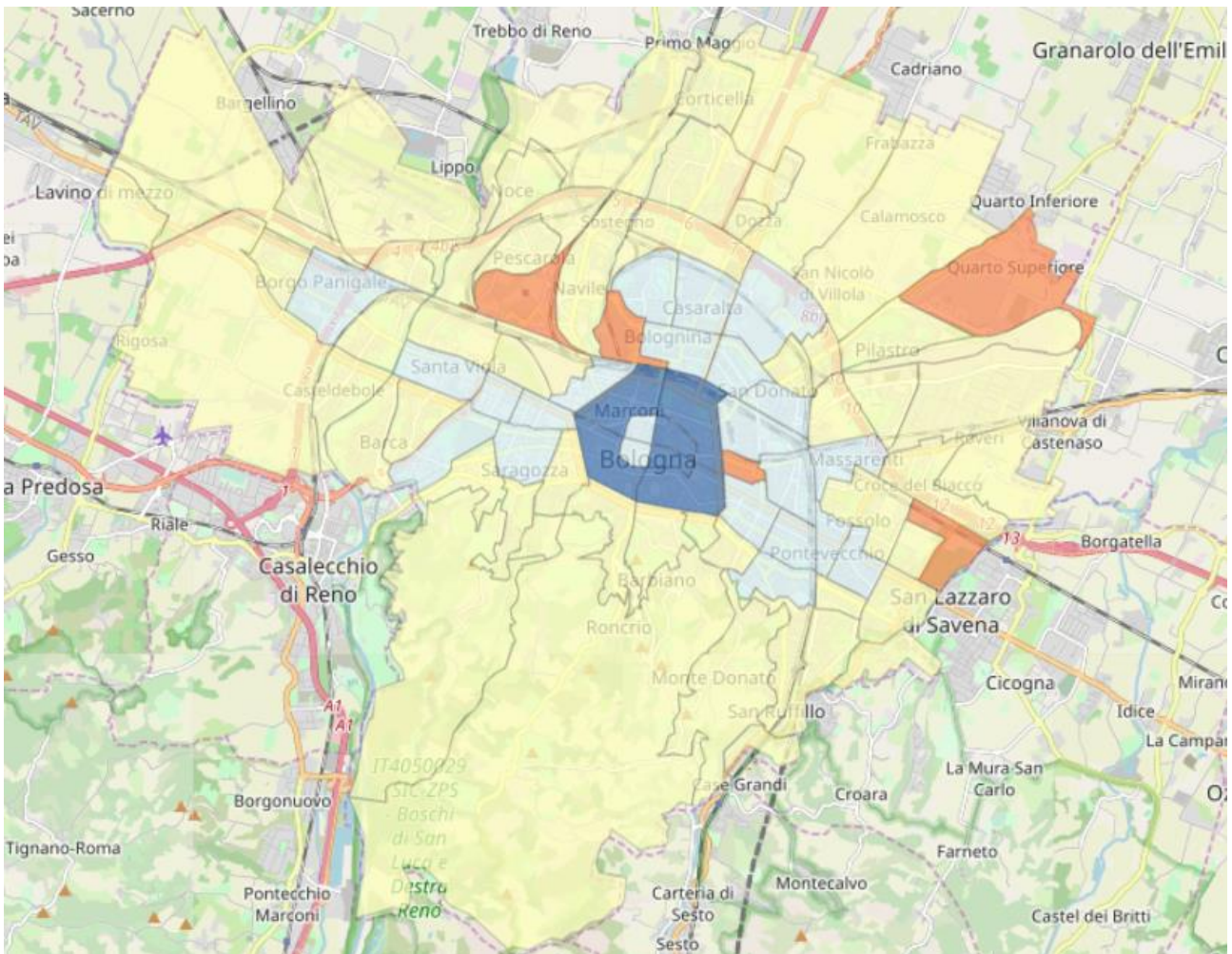|	name	income_x_taxpayer	density	15-29_perc	30-44_perc	45-64_perc	65-74_perc	veg_restaurant_nr	office_nr	school_nr	university_nr	museum_nr
0	BITONE	21837.937219	17958.249550	16.662391	31.618779	38.045151	13.673679	1.0	10.0	5.0	3.0	0.0
1	XXI APRILE	26021.767837	15635.233271	17.612917	31.205150	38.001266	13.180667	0.0	27.0	4.0	8.0	0.0
2	DAGNINI	25910.428492	15264.883873	17.029993	30.446233	39.019751	13.504023	0.0	15.0	8.0	7.0	0.0
3	MARCONI-2	30567.398289	13941.735177	17.519031	32.454165	36.346092	13.680712	2.0	18.0	18.0	21.0	3.0
4	VIA MONDO	20246.937130	13908.284476	18.589744	32.591093	35.307018	13.512146	0.0	7.0	4.0	5.0	0.0

We then apply KMeans algorithm with 4 clusters and calculate the mean of values in each cluster obtaining the following result

	cluster	income_x_taxpayer	density	15-29_perc	30-44_perc	45-64_perc	65-74_perc	veg_restaurant_nr	office_nr	school_nr	university_nr	museum_nr
0	0	20071.165994	373.997183	34.120594	31.774981	22.989069	11.115355	0.800000	9.600000	2.000000	8.800000	0.200000
1	1	27752.985664	3152.284373	18.796627	25.463121	41.236637	14.503615	0.196429	6.910714	1.892857	2.160714	0.375000
2	2	25265.118308	11212.663601	17.258685	31.688826	37.941220	13.111269	0.652174	16.434783	4.000000	6.217391	0.913043
3	3	36936.794307	11403.793382	18.268855	29.993880	37.682399	14.054867	4.500000	18.333333	12.000000	19.333333	8.666667

Results

Where you discuss the results.

Here are the different areas based on cluster:

For each cluster we can have a summary based on this table:

	cluster	income_x_taxpayer_bin	density_bin	15-29_perc_bin	30-44_perc_bin	45-64_perc_bin	65-74_perc_bin	veg_restaurant_nr_bin	office_nr_bin	school_nr_bin	university_nr_bin	museum_nr_bin
0	0	Medium	Low	Low	Low	High	High	Low	Low	Low	Low	Low
1	1	Low	High	Low	High	High	Medium	Low	High	Low	Low	Low
2	2	High	High	Low	High	High	High	High	High	High	High	High
3	3	Low	Low	High	High	Low	Low	Low	Low	Low	Medium	Low

Discussion

Now we give a description for each cluster taking into accounts the relevant features and, on its basis, we can recommend the type of restaurant for each cluster.

Cluster	Colour on map	Notes
0	Light Yellow	Few veg restaurants, medium income, high aged people, few activities
1	Light Blue	Few veg restaurants, high density, middle aged people,
2	Blue	Many veg restaurants, high income, high density, many activities
3	Orange	Few veg restaurants, many young people, few activities

Our recommendations are:

Cluster 0: In this cluster the opening of a new veg restaurant is not so much recommended.

Cluster 1: Not so many offices, not a big income, but high density and middle-high aged people. So people seems to work away from home. We recommend a low cost-take away veg restaurant, for people who get home after work and don't want to make dinner.

Cluster 2: High income of residents and high activity. Here we recommend a veg restaurant opened at midday with lower prices and at evening with higher prices and better service, at midday for people from offices and at evening for residents. Of course there are already many veg restaurants in this cluster. But if we go on detail we see that there are areas in this cluster where the veg restaurants are relatively few, e.g. MARCONI-2 and GALVANI-1:

name	income_x_taxpayer	cluster	density	15-29_perc	30-44_perc	45-64_perc	65-74_perc	veg_restaurant_nr	office_nr	school_nr	university_nr	museum_nr
MARCONI-2	30567.398289	2	13941.735177	17.519031	32.454165	36.346092	13.680712	2.0	15.0	18.0	20.0	3.0
GALVANI-1	55681.036126	2	8471.420230	19.060967	25.227751	41.485634	14.225648	3.0	17.0	5.0	17.0	12.0
GALVANI-2	37403.042157	2	12416.741189	18.108170	28.046554	39.262894	14.582382	4.0	18.0	13.0	26.0	14.0
IRNERIO-1	36250.047223	2	11546.030765	18.739235	29.417844	36.823975	15.018946	4.0	20.0	9.0	18.0	5.0
MALPIGHI-2	29785.946051	2	12600.483171	15.965001	31.904641	38.638093	13.492265	6.0	15.0	15.0	9.0	6.0
IRNERIO-2	31933.295992	2	9446.349759	20.220723	32.912324	33.537707	13.329246	8.0	17.0	12.0	26.0	12.0

Cluster 3: Low income, not high density, many young people, not many activities. For many aspects it is similar to Cluster 1. People here are younger, so we recommend take away restaurant with fast service and low prices.

Conclusion

Good place to open a veg restaurant is still cluster 2, but it is the most challenging cluster. Cluster 1 and cluster 3 offer good possibilities for starting a new and different type of restaurant, low cost and faster in service.