



Where to open a Bakery in Curitiba (Brazil)

Fernando Tauscheck

The Battle of the Neighborhoods

Applied Data Science Capstone by
IBM/Coursera



Business Problem

What defines the success of a commercial business? Can we predict if a point is good enough to open a profitable bakery?

Although the analysis can, in theory, be replicated for any type of business, this report will be targeted to stakeholders interested in opening a bakery in Curitiba, Brazil. We will use geographic and socioeconomic data from existing bakeries to define a list of possible locations to open a bakery.

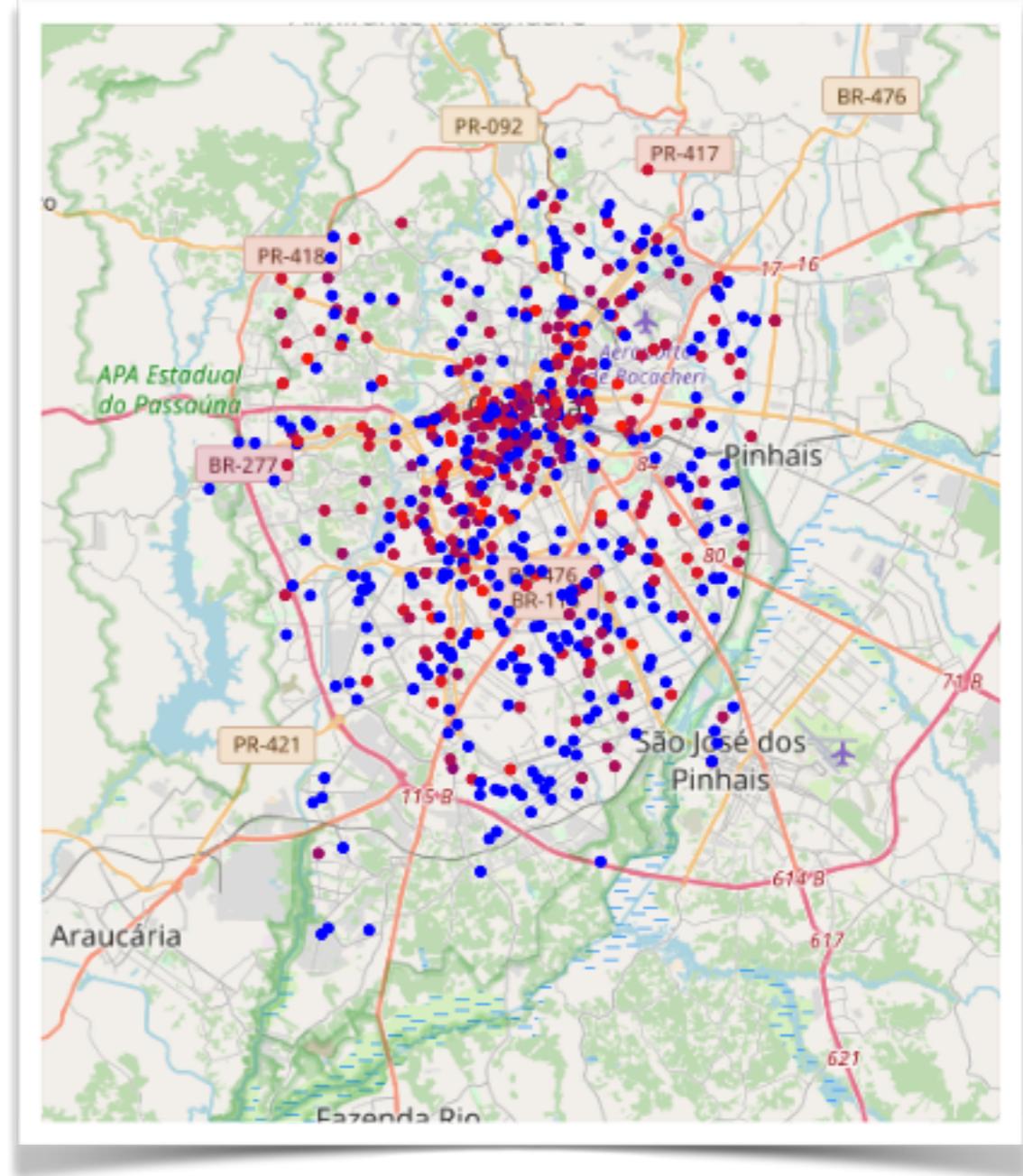
Data:

Some factors will influence our analysis:

- Number of existing bakeries in the neighborhood;
- Socioeconomic data of the neighborhoods (Per capita income, population density, ...);
- Zones from City Master Plan;
- Proximity to parks, public square, boardwalk, main streets, and avenues of great circulation;
- If possible, we will try to compare the Foursquare Rating of each bakery and understand if the location and Rating have correlation

Foursquare API:

This project uses the Foursquare API as its main data gathering source as it has a database of millions of venues.



	id	name	lat	long	address	categories	tipcount	tier	likes	rating
0	4b69efebf964a5201bbd2be3	Confeitoria das Famílias	-25.430643	-49.270212	R. Quinze de Novembro, 374	[["Dessert Shop", "4bf58dd8d48988d1d0941735"], ...]	114	1	235	6.7
1	4b75d4fcf964a520ee272ee3	Panetteria Maiochi	-25.472368	-49.288013	R. Maranhão, 1730	[["Bakery", "4bf58dd8d48988d16a941735"], ["Con...]	22	1	34	6.3
2	4b7c57d1f964a5209f8d2fe3	La Patisserie	-25.442422	-49.279188	Av. Sete de Setembro, 4194	[["Bakery", "4bf58dd8d48988d16a941735"], ["Cof...]	71	2	111	6.3
3	4b8abddbf964a520c07d32e3	Saint Germain	-25.432826	-49.290227	Al. Prca. Izabel, 1347	[["Bakery", "4bf58dd8d48988d16a941735"], ["Del...]	57	3	221	6.9
4	4ba29a89f964a520680838e3	Saint Germain	-25.444152	-49.287664	Av. Visc. de Guarapuava, 4882	[["Bakery", "4bf58dd8d48988d16a941735"], ["Bre...]	180	3	617	7.8
5	4b88083025643616862e4	180	3	719	8.1
6	4b88083025643616862e4	21	3	551	8.0

Geographic Data:

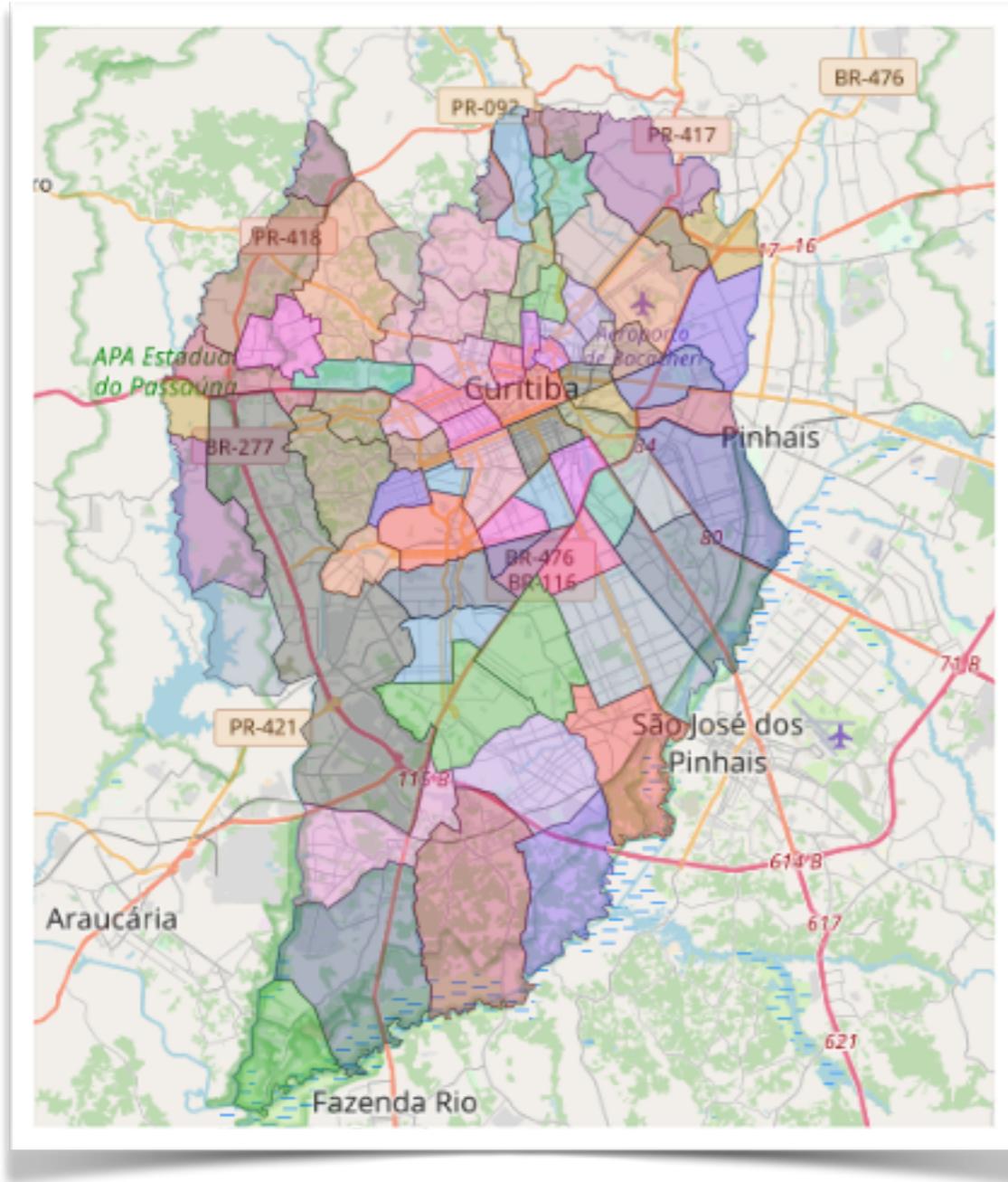
We've got geographic information from Curitiba at the website of the "*Instituto de Pesquisa e Planejamento Urbano de Curitiba*"* (Institute of Urban Planning and Research of Curitiba also know as IPPUC). The Institute provides all sorts of maps of Curitiba. We will use:

- Zones of City Master Plan;
- Neighborhoods;
- Mains streets;
- Boardwalks, public squares, and parks

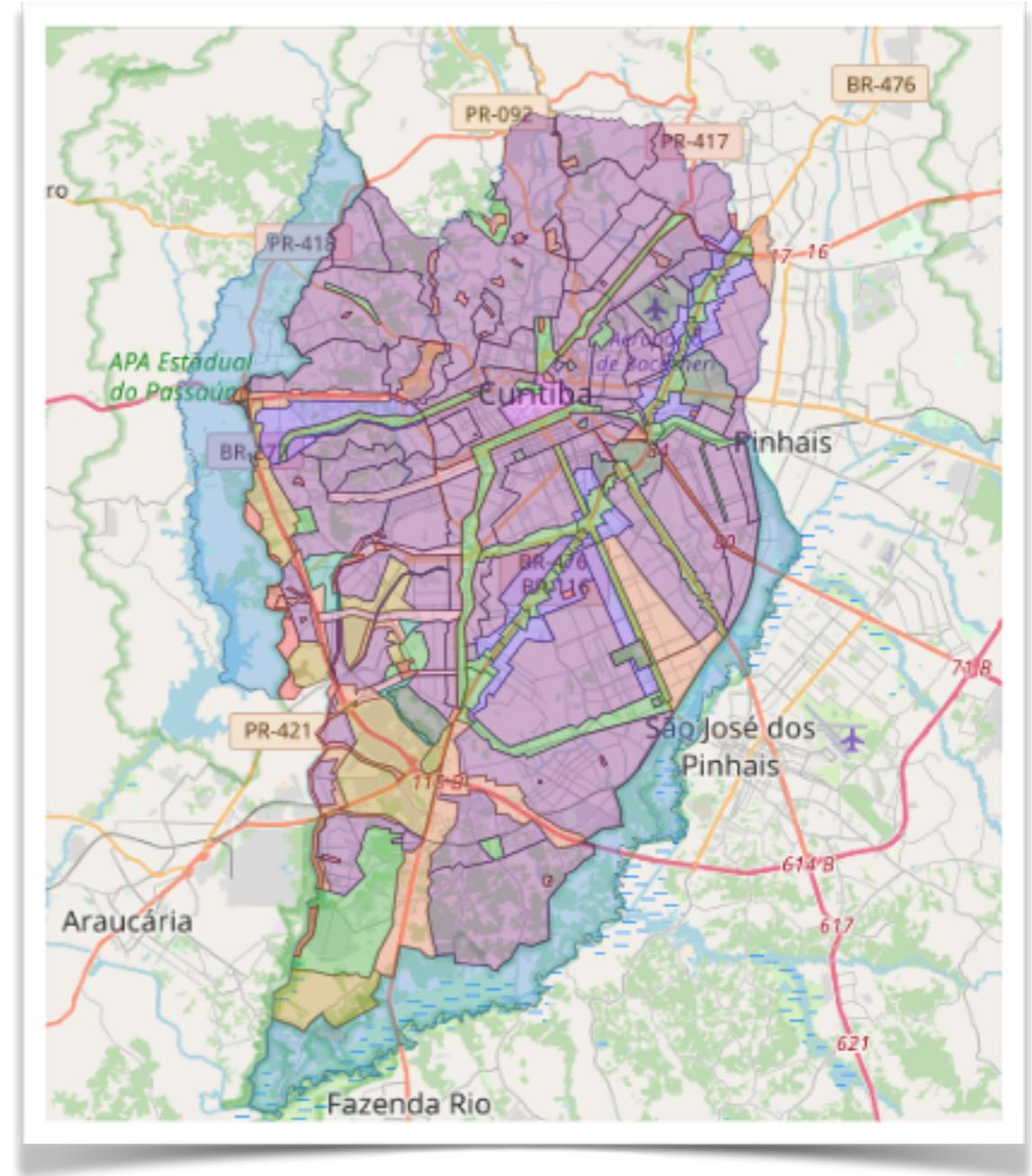
These maps are provided in SHP format (ESRI). Posteriorly they were converted to GeoJSON in a proper representation (WGS84). The GeoJSON files was inserted in an RDMBS (PostgreSQL), where will be used the PostGIS extension to analyze.

*: <https://ippuc.org.br/geodownloads/geo.htm>

Neighborhoods:



MasterPlan:



Socioeconomic data of the neighborhoods:

The socioeconomic data of the municipality was collected from the Wikipedia article* "Lista de bairros de Curitiba".

	id	neighbourhood	norm_neighbourhood	area	men	women	total	households	avg_income	area_sqm	personpersqm	incomepersqm
0	1	Ganchinho	GANCHINHO	11.20	3667	3658	7325	1921	767.35	1.135980e+07	0.000645	0.000068
1	2	Sitio Cercado	SITIO CERCADO	11.12	50631	51779	102410	27914	934.95	1.120015e+07	0.009144	0.000083
2	3	Umbará	UMBARA	22.47	7280	7315	14595	17064	908.70	2.245102e+07	0.000650	0.000040
3	4	Abranches	ABRANCHES	4.32	5463	5702	11165	3154	1009.67	4.316511e+06	0.002587	0.000234
4	5	Atuba	ATUBA	4.27	6156	6476	12632	3627	1211.60	4.159346e+06	0.003037	0.000291
5	6	eduta	eduta	12.4	1249	1251	2500	6251	903.72	0.994682e+07	0.008000	0.000068
6	7	SECHINASIDA	SECHINASIDA	4.35	2823	2805	5628	1428	18.8000	0.997328e+07	0.005281	0.000534

*: https://pt.wikipedia.org/wiki/Lista_de_bairros_de_Curitiba

Methodology:

The objective of this project is to find regions in Curitiba with the best conditions for opening a well rated bakery.

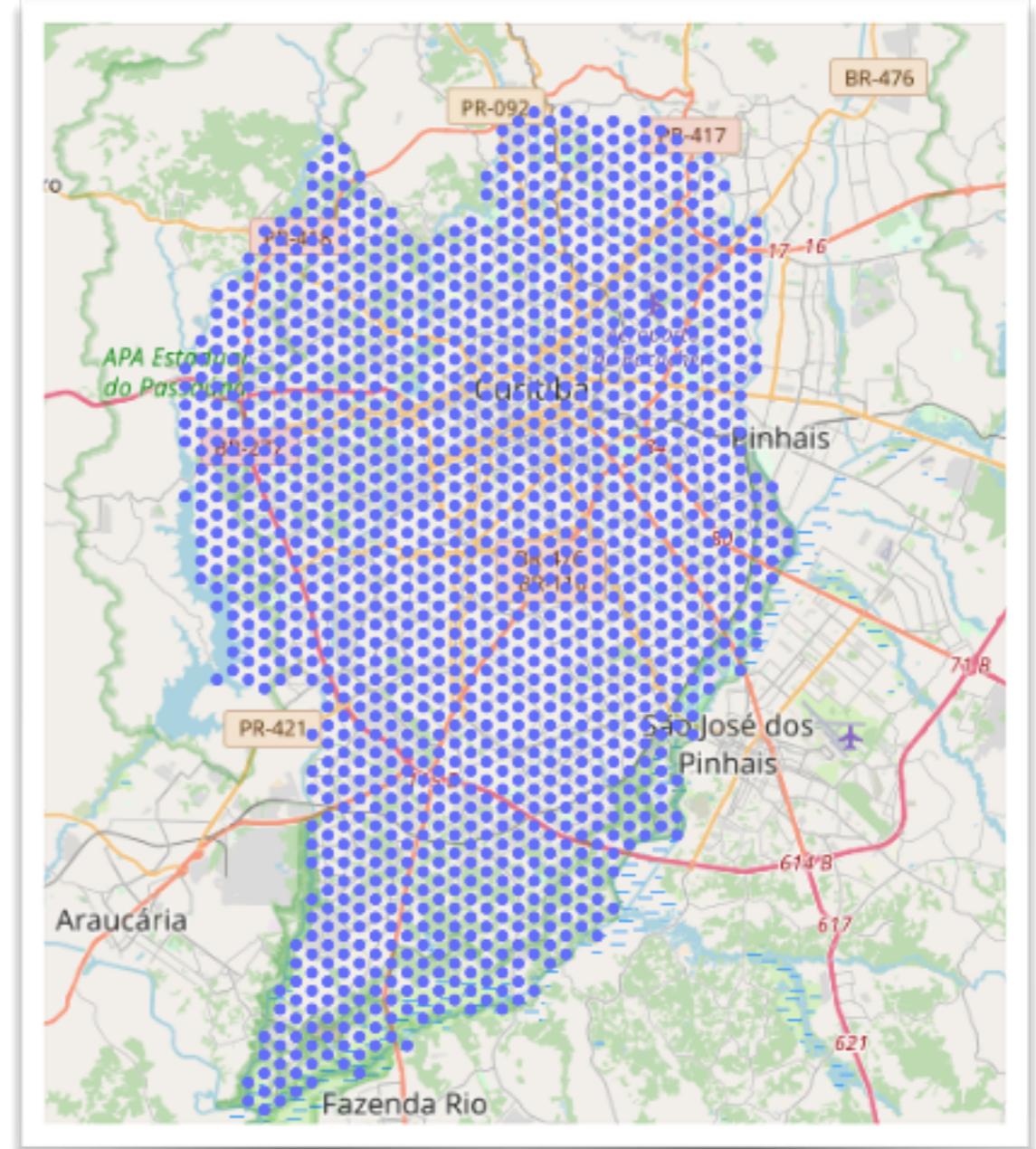
In a first step, we collect all relevant data. Geographical data provided by the city of Curitiba (through IPPUC), socioeconomic data (collected on Wikipedia) and location and classification data of current bakeries were considered. For this, we use the Foursquare API. All data were submitted to tables in PostgreSQL database.

In a second step, there will be data exploration. For this purpose, the city will be divided into hexagons with a radius of 300m. For each of these 'areas', geographic and socioeconomic data will be added that will allow the application of the K algorithm - Nearest Neighbors (KNN).

Only in the third stage of the project the Foursquare data from current venues will be added to the study. With this, we will be able to define the regions with the greatest potential and classify them for our stakeholders.

Hexagons of the study:

A total of 1406 hexagons were created, covering the entire urban area of the municipality.



Hexagons vs Geographic data:

Using PostGIS functions, we will add to each hexagon some characteristics of the region:

- Calculating the overlapping area of each neighborhood with the hexagons and applying the proportionality of population and income to the overlapping area (about the neighborhood area), we will have income and population for each hexagon.
- Area of each zone of the Master Plan;
- Length of each type of main streets;
- Area of Parks, BoardWalks, Gardens, Public Square...

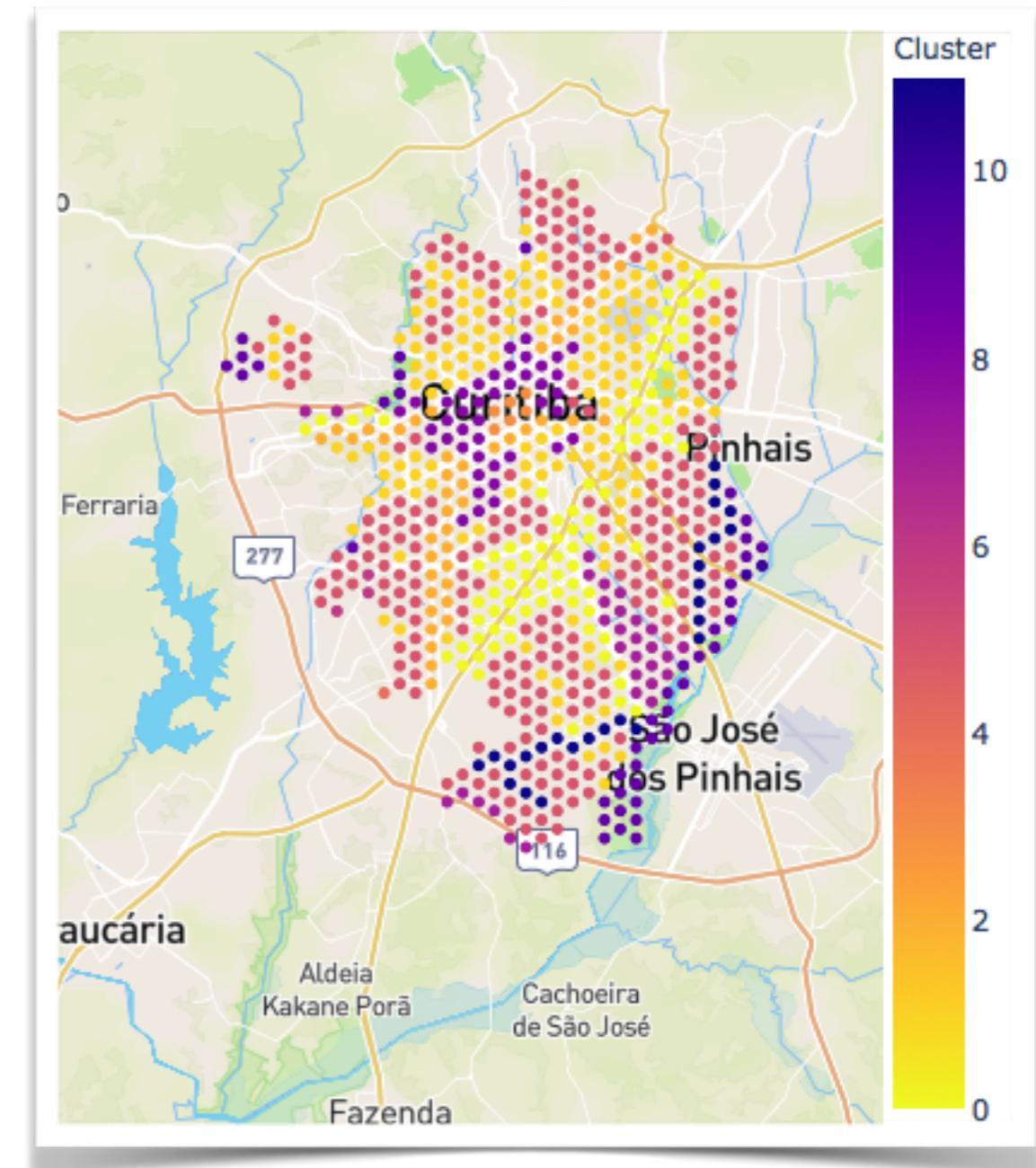
	lat	long	persons	income	zc	zum	conec	zs	zt	apa	...	collector	external	main_line	other_routes	priority	sectorial	Park	Sport Center	Garden	Public Square
id																					
481	-25.429425	-49.301793	2914.15	11052448.76	0.0	0.0	0.0	0.0	0.0	0.0	...	0.00	85.38	0.0	4.99	0.00	0.0	0.0	0.0	0.00	
408	-25.432123	-49.296619	2914.09	11052206.70	0.0	0.0	0.0	0.0	0.0	0.0	...	0.00	1172.82	0.0	806.03	531.83	0.0	0.0	0.0	0.00	
340	-25.434821	-49.291444	2896.61	11046487.33	0.0	0.0	0.0	0.0	0.0	0.0	...	217.78	0.00	0.0	0.00	1238.67	0.0	0.0	0.0	170.65	4258.76
410	-25.434821	-49.301794	2911.16	11036708.26	0.0	0.0	0.0	0.0	0.0	0.0	...	65.29	1030.96	0.0	1159.63	0.00	0.0	0.0	0.0	0.00	162.54
342	-25.437519	-49.296619	2841.38	11028934.74	0.0	0.0	0.0	0.0	0.0	0.0	...	0.00	0.00	0.0	0.00	0.00	0.0	0.0	0.0	75.39	2704.50
245	-25.437519	-49.296619	2841.38	11028934.74	0.0	0.0	0.0	0.0	0.0	0.0	...	0.00	0.00	0.0	0.00	0.00	0.0	0.0	0.0	0.00	02.40115

Clustering the data:

Using the K Means algorithm, the points were grouped into 12 clusters.

By analyzing the clusters together with the Foursquare information, we determined the clusters most likely to have a high-rated bakery (rating ≥ 7.5).

	points	backerys	mean
2	31.0	13.0	0.684211
1	30.0	10.0	0.357143
3	6.0	1.0	0.333333
7	45.0	8.0	0.181818
5	174.0	31.0	0.179191
11	8.0	1.0	0.125000
6	11.0	0.0	0.000000
9	0.0	0.0	0.000000
8	0.8	0.1	0.025000

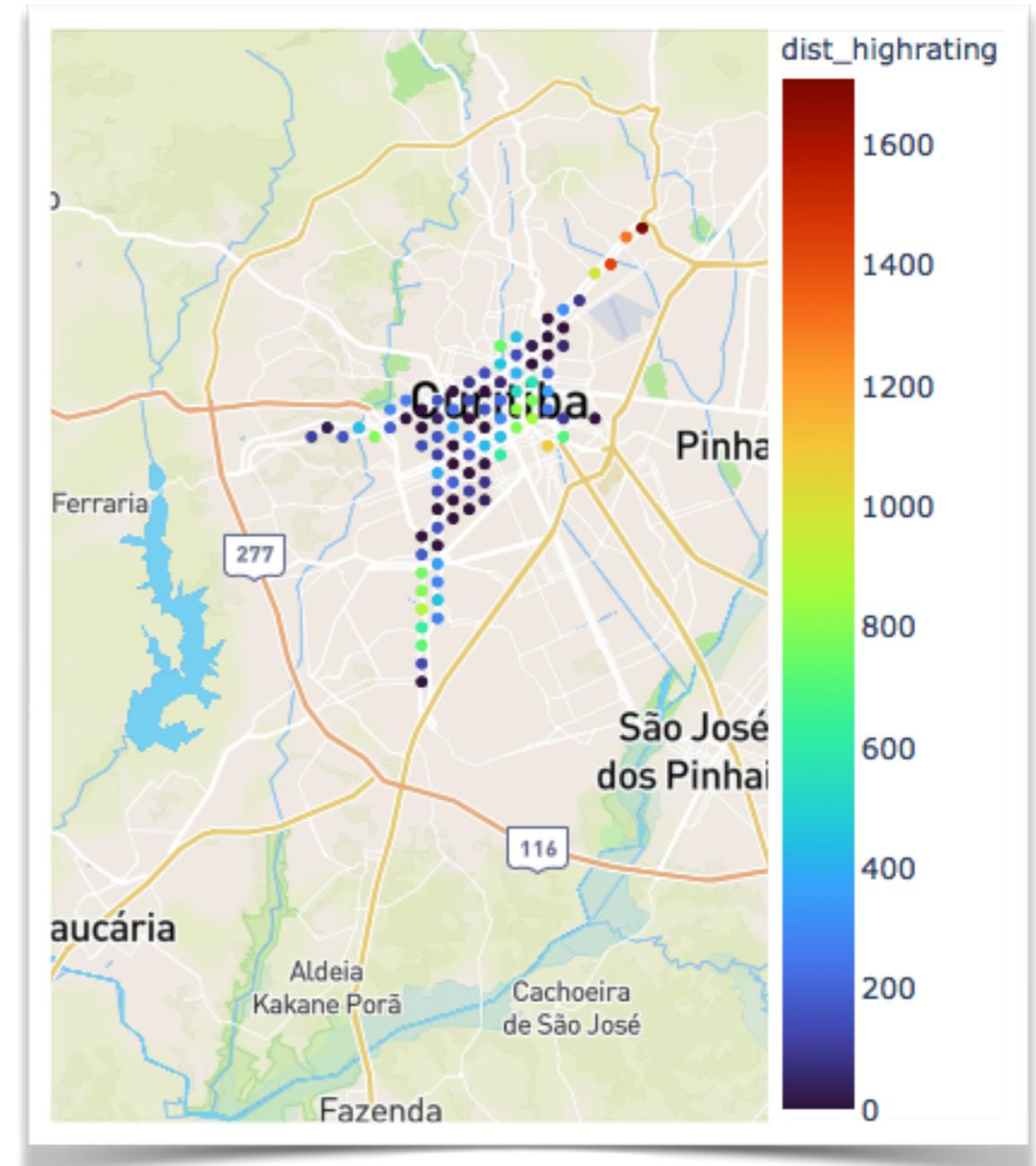


A peak in the data:

Analyzing only the distance between a hexagon and the closest high-income bakery, the choice would fall to the extreme points (greater distance).

But we have more information that can help us in the analysis:

- What is the population in these hexagons?
- What is the average income on these hexagons?

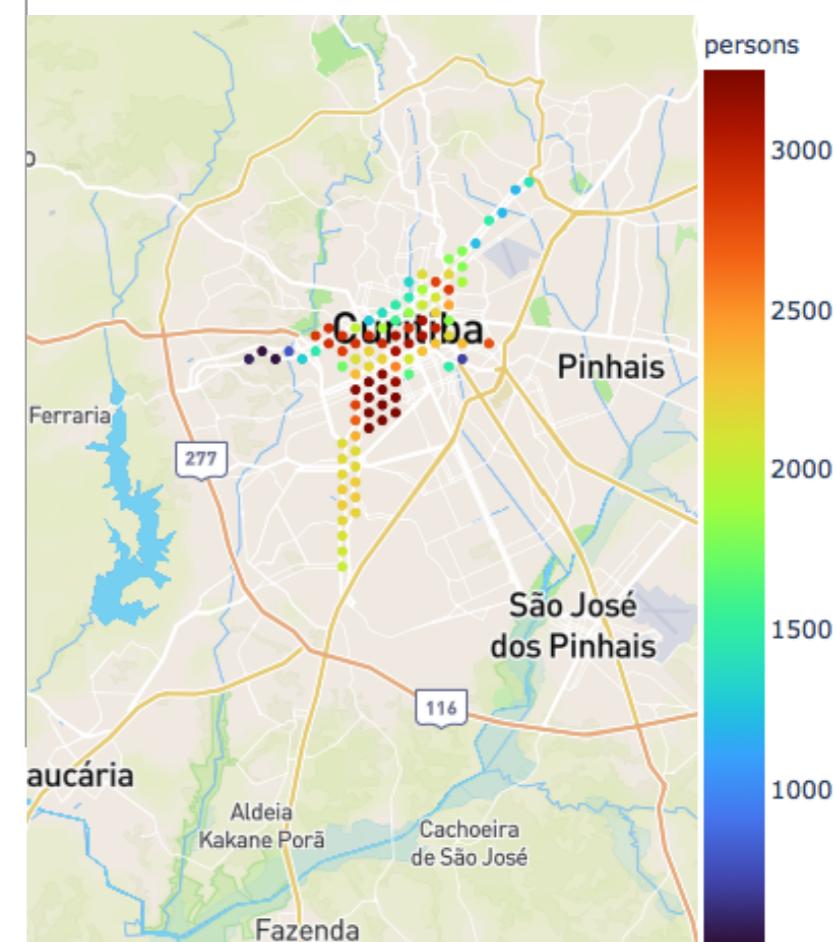
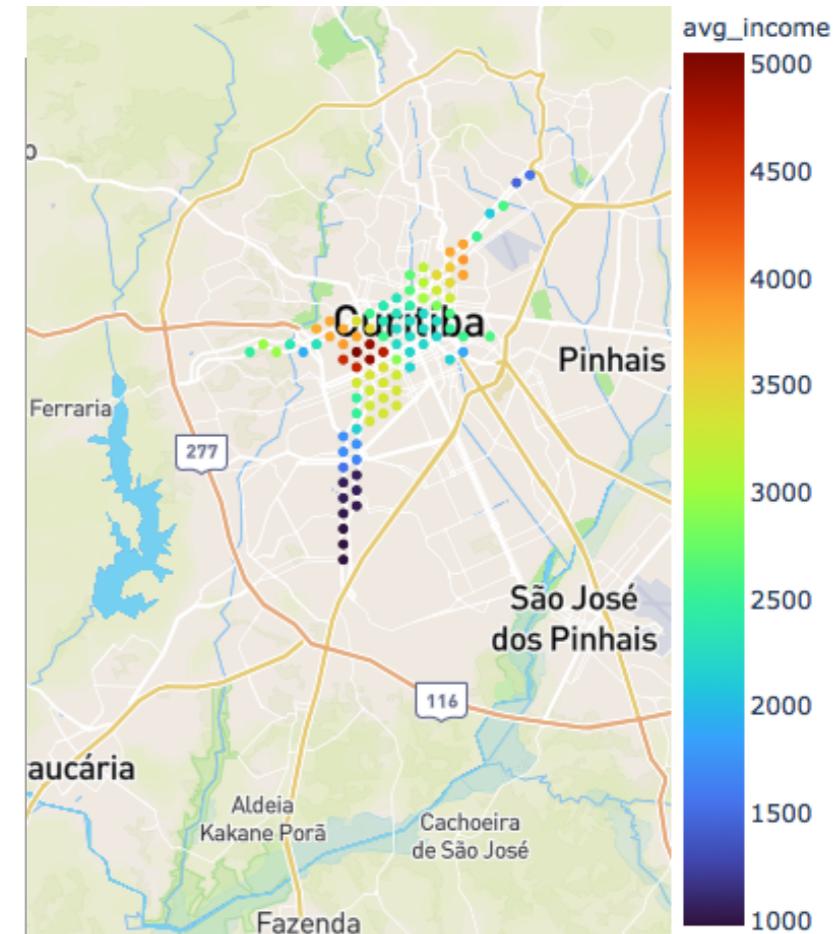


AVG Income and Persons per hexagon:

Analyzing the income and population distribution maps, we have a clear population cleavage, without a direct correlation between population and income (characteristics of the Brazilian population).

According to data from IBGE (Brazilian Institute of Geography and Statistics) in 2019 the GINI index of socioeconomic inequality in Brazil was 0.539, placing the country in 159th position*.

"In 2019, the 10% share of people with the lowest per capita household income received a share of 0.8% of the total income. Half of the Brazilian population corresponded to 15.6% of the observed income, while the 10% with the highest earnings 42.9% of all income received by people in 2019."



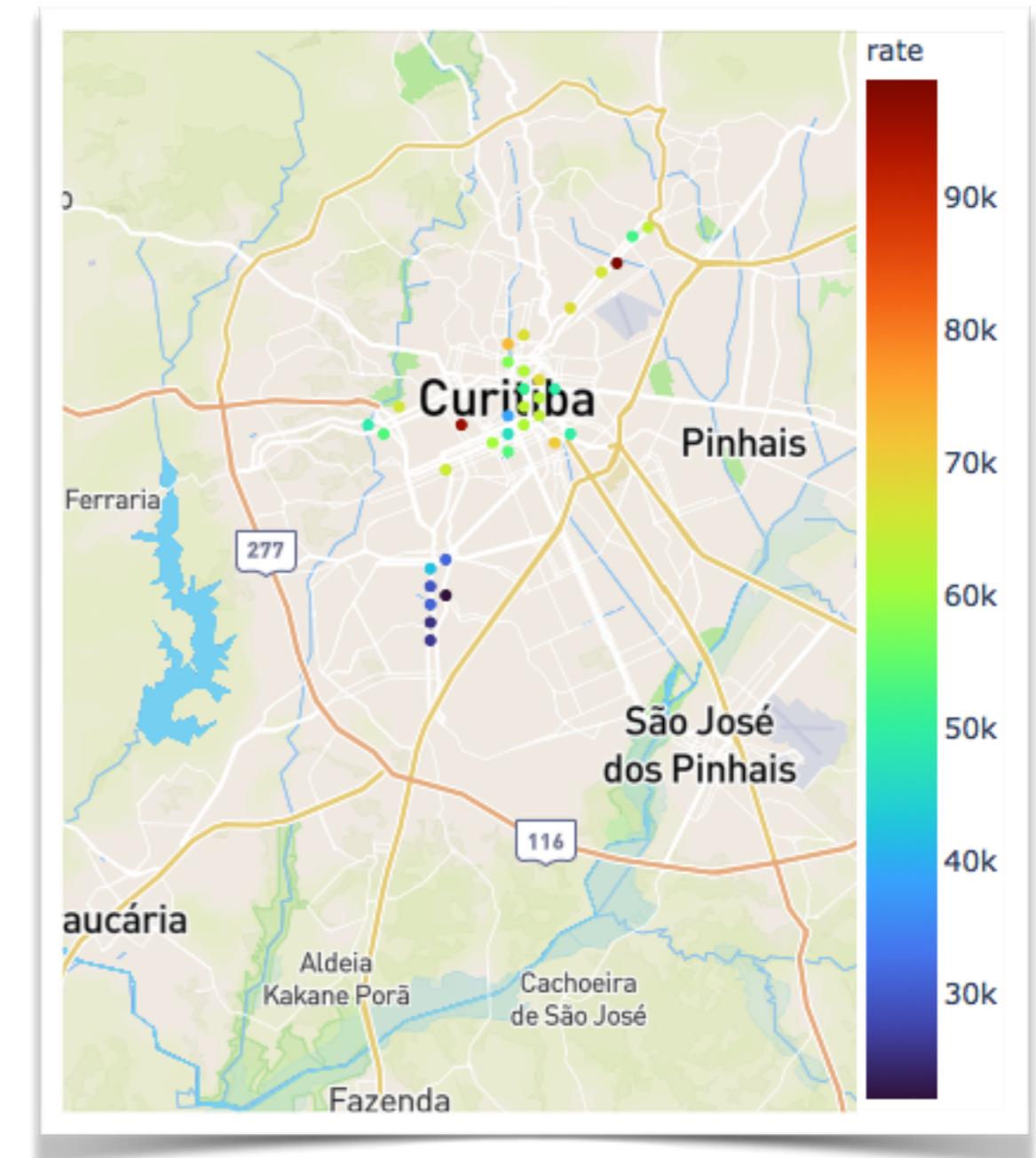
*: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101760.pdf>

Relevance:

In an attempt to focus on regions with an adequate balance between population, income and distance to another well-rated bakery (rating greater than 7.5), we propose a *Relevance* metric considering all these variables.

$$\text{Per Capita Income (PCI)} = \frac{\text{Income}}{\text{Persons}}$$

$$\text{Relevance} = \text{PCI} * \sqrt{\text{Distance to Bakery}}$$



To avoid distortions with hexagons with very small distances, we are filtering cases with less than 300m of distance for a bakery.

Final Ranking:

Rank	Address	Persons	Income	Avg Income	Max Rating	Dist High Rating	Rate	Lat	Long
1	Rua Cleide Iurk, 65 - Bacacheri, Curitiba - PR, 82510-440	1191,63	R\$3.110.706,45	R\$2610,46	0,0	1436,69	98946,22	-25,388955	-49,234533
2	Av. Vicente Machado, 1039 - Centro, Curitiba - PR, 80420-011	2207,75	R\$10.830.081,16	R\$4905,48	6,1	382,30	95914,98	-25,437519	-49,286269
3	R. Prof. Benedito Nicolau dos Santos, 522 - Centro Cívico, Curitiba - PR, 80530-150	1333,71	R\$3.589.838,74	R\$2691,62		738,38	73139,79	-25,413237	-49,270744
4	R. Imac, Conceição, 247 - Rebouças, Curitiba - PR, 80215-030	1489,91	R\$3.216.532,29	R\$2158,88	0,0	1071,77	70677,12	-25,442915	-49,255219
5	R. Ubaldino do Amaral, 334 - Alto da Glória, Curitiba - PR, 80060-195	2208,14	R\$6.287.645,07	R\$2847,48	0,0	573,50	68191,45	-25,424029	-49,260396
6	R. Belém, 892 - Cabral, Curitiba - PR, 80035-170	1794,35	R\$6.896.679,00	R\$3843,55		312,12	67903,50	-25,402445	-49,25005
7	R. Mal, Hermes, 910 - Ahú, Curitiba - PR, 80530-230	2148,41	R\$6.820.238,58	R\$3174,55		452,54	67532,32	-25,410539	-49,26557
8	R. Luiz Leitner, 221 - Bigorrilho, Curitiba - PR, 80710-390	2870,19	R\$10.818.638,78	R\$3769,31		309,48	66309,37	-25,432123	-49,306969
9	Rua Lodovico Geronazzo, 159 - Boa Vista, Curitiba - PR, 82560-040	1493,24	R\$3.148.672,48	R\$2108,62		986,62	66232,90	-25,391653	-49,239706
10	Rua Cândido Xavier, 533 - Água Verde, Curitiba - PR, 80240-280	3210,09	R\$10.643.640,24	R\$3315,68		389,88	65469,68	-25,451009	-49,291446
11	R. Mariano Torres, 1007-957 - Centro, Curitiba - PR, 80060-120	2300,8	R\$5.049.581,19	R\$2194,71	6,3	850,86	64018,53	-25,434821	-49,260394
12	R. Calixto de Jesus, 95 - Tingui, Curitiba - PR, 82620-270	1428,9	R\$2.202.716,54	R\$1541,55		1710,65	63758,36	-25,378163	-49,22419
13	R. João Negrão, 363 - Centro, Curitiba - PR, 80010-200	3092,03	R\$6.868.974,69	R\$2221,51	5,7	818,80	63567,75	-25,432123	-49,265569
14	R. Gen. Carneiro, 784 - Alto da Glória, Curitiba - PR, 80060-150	2980,16	R\$6.682.908,94	R\$2242,47	5,6	781,60	62693,04	-25,429425	-49,260395
15	Av. Sete de Setembro, 1865 - Rebouças, Curitiba - PR, 80060-070	2292,09	R\$5.054.900,05	R\$2205,37	0,0	794,36	62157,11	-25,437519	-49,265569
16	Av. João Gualberto, 570 - Alto da Glória, Curitiba - PR, 80030-000	2049,18	R\$6.183.966,50	R\$3017,78		419,27	61792,05	-25,421331	-49,26557
17	R. Brg, Franco, 2580 - Batel, Curitiba - PR, 80230-050	2557,89	R\$7.389.412,45	R\$2888,87	6,0	441,35	60690,08	-25,442915	-49,275919
18	R. Aristides Teixeira, 155 - Centro Cívico, Curitiba - PR, 80530-110	1460,82	R\$3.876.225,13	R\$2653,46		470,86	57578,02	-25,418633	-49,270744
19	R. André Cecone, 16 - Campina do Siqueira, Curitiba - PR, 80740-530	1302,76	R\$2.480.459,95	R\$1904,00	7,3	793,43	53631,78	-25,440217	-49,312146
20	Rua 24 de Maio, 1012 - Rebouças, Curitiba - PR, 80220-060	1638,24	R\$3.572.141,34	R\$2180,47		604,04	53590,22	-25,445613	-49,270744
21	Rua Coronel Wallace Scott Murray, 9 - Santa Cândida, Curitiba - PR, 82640-170	1196,48	R\$1.731.580,07	R\$1447,23	6,6	1307,03	52321,39	-25,380861	-49,229362
22	R. Luiz Leão, 33 - Centro, Curitiba - PR, 80030-010	3091,57	R\$6.867.948,15	R\$2221,51	6,6	531,01	51191,76	-25,426727	-49,26557
23	Av. Prefeito Omar Sabbag, 614 - Jardim Botânico, Curitiba - PR, 80210-000	694,44	R\$1.309.549,30	R\$1885,76	6,6	692,48	49623,99	-25,440217	-49,250044
24	R. XV de Novembro, 1948 - Alto da XV, Curitiba - PR, 80045-125	1818,59	R\$4.750.300,60	R\$2612,08	5,6	357,92	49417,31	-25,426727	-49,255221
25	R. Prof. Pedro Viriato Parigot de Souza, 999 - Campina do Siqueira, Curitiba - PR, 81200-100	790,46	R\$1.866.436,80	R\$2361,20		432,84	49124,23	-25,437519	-49,317321
26	Av. Silva Jardim, 994 - Rebouças, Curitiba - PR, 80230-000	2087,63	R\$4.591.237,21	R\$2199,26	5,6	447,25	46510,71	-25,440217	-49,270744
27	R. São Mateus, 605 - Portão, Curitiba - PR, 81070-080	2187,36	R\$3.340.562,28	R\$1527,21	0,0	759,28	42082,31	-25,480687	-49,296629
28	Alameda Dr. Muricy, N° 270 - Sala 105 - Centro, Curitiba - PR, 80010-120	3091,96	R\$6.868.824,23	R\$2221,51	6,3	304,20	38746,12	-25,434821	-49,270744
29	Av. República Argentina, 4121 - Portão, Curitiba - PR, 81050-000	2171,21	R\$3.595.777,24	R\$1656,12	0,0	356,33	31261,84	-25,477989	-49,291452
30	R. Mario Chaubald Biscaia, 381 - Novo Mundo, Curitiba - PR, 81050-240	2248,87	R\$2.339.728,69	R\$1040,40		882,74	30911,35	-25,491479	-49,296629
31	R. Visc. do Serro Frio, 229 - Novo Mundo, Curitiba - PR, 81050-080	2248,97	R\$2.339.831,45	R\$1040,40		788,75	29219,36	-25,486083	-49,296629
32	Av. Winston Churchill, 670 - Capão Raso, Curitiba - PR, 81130-070	2118,18	R\$2.069.862,52	R\$977,19	6,3	712,42	26082,38	-25,502271	-49,296629
33	R. Pedro Gusso, 3305 - Novo Mundo, Curitiba - PR, 81315-000	2175,47	R\$2.188.393,16	R\$1005,94	7,4	635,91	25367,04	-25,496875	-49,296629
34	Av. Brasília, 4925 - Novo Mundo, Curitiba - PR, 81020-010	2248,92	R\$2.339.780,07	R\$1040,40		447,21	22001,73	-25,488781	-49,291452

Conclusion:

Purpose of this project was to identify Curitiba areas in order to aid stakeholders in narrowing down the search for optimal location for a new Bakery. By crossing socioeconomic and geographic information with the distribution os Bakerys from Foursquare data, we create a list of zones of interest and addresses to be used as starting points for final exploration by stakeholders.

Final decision on optimal bakery 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.

Thank You!

