COURSERA CAPSTONE PROJECT GEOLOCATION DATA ANALYSIS

ANALYZING NEIGHBORHOOD IN THE CITY OF BUENOS AIRES TO HELP STRATEGIC BUSINESS EXPANSION FOR A COFFEE SHOP FRANCHISE CHAIN

Introduction

- Argentina has a well-known culture of meeting with a friend or relative to have a small *cortado* coffee over a nice, warm talk. As such, in the City of Buenos Aires, its capital city, there are many small coffee shops and coffee chains. Some neighborhoods seem to be overpopulated with venues, and many are struggling to make their business work. On the contrary, other neighborhoods have grown in population over the last years, and there are business opportunities for new coffee shops to open.

Introduction

- **Problem**

Data that might help determine which of the almost 50 *Barrios Porteños* (Buenos Aires neighborhoods) have too many coffee venues, and which of the lack of good places where to have a *cortado*, would help brands thrive and take data-driven decision on their expansion strategies.

- **Interest**

One of Argentina's biggest Coffee Franchise Chains, is analyzing where to expand and open new brunches within the City of Buenos Aires. They need to relate population density with coffee shop density of each of the city's neighborhoods. This would definitely help them avoid wrong locations and increase chance of success for the new franchises.

Data Gathering and Processing

- **Data Sources**

A list and census of all the neighborhoods of the city is available as annex for Buenos Aires's article.

https://es.wikipedia.org/wiki/Anexo:Barrios_de_la_ciudad_de_Buenos_Aires

The features included are: Name, area, population, district.

From Geopy Library for Python, we locate the coordinates for each neighborhood.

Finally, Foursquare Database, gives us access to the already existing coffee venues in the City and allows us to establish a measure of density for this kind of stores.

Data Gathering and Processing

- **Data Wrangling**

First, we create a dataframe with all the neighborhoods and the needed features to establish the population density. We read the Wikipedia article CSV file and drop the unwanted features. Then, we use that list to iterate a process in Geopy's Nominatin and get Longitude and Latitude for each neighborhood. We add this values to the above dataframe.

Connecting to Foursquare's API with our credentials, we get all the top venues ina a specific radius for each neighborhood.

	name	categories	lat	lng
0	Megatlon	Gym	-34.579216	-58.426123
1	Nicolo Helados	Ice Cream Shop	-34.581652	-58.426735
2	La Carnicería	Argentinian Restaurant	-34.582814	-58.423778
3	Distrito Arcos	Shopping Mall	-34.580494	-58.427621
4	Vain Boutique Hotel Buenos Aires	Hotel	-34.583672	-58.424980

Exploratory Data Analysis

- **Statistical Data**

Once we count with all the listed venues, we identify and keep only the categories of coffee places that could be a business competitor to our new potential stores. We take the average occurrence for these kinds of venues (coffee shops, tea house, pie store, etc) and add them to establish a unique value.

Doing simple arithmetic, we use are and population in our former dataframe, to calculate the density of population for each neighborhood.

Finally, we combine the two density indexes (for citizens and coffee stores), and generate a new unique variable that could determine if there enough venues for the density of people or if there is a business opportunity in each neighborhood.

Conclusions and Further Analysis

- **Conclusions**

With this study, we could identify those neighborhoods with better opportunities for expansion of coffee shops. Those in the highest part of the table, have a low density of venues to satisfy the population living in that specific area. Some of those neighborhoods have really few places and they have grown a lot in its density. As people moved from more classic downtown neighborhoods to new areas in the outskirts, new opportunities arise. Neighborhoods as Nueva Pompeya, Barracas, Villa del Parque, Floresta, to name a few, are examples of this.

On the contrary, there are many neighborhoods, like Puerto Madero or Villa Riachuelo, which seem to be saturated with places and not a good opportunity for an expansion.

	DISTRICT	NEIGHBORHOOD
0	19	Nueva Pompeya
1	3	Barracas
2	43	Villa Real
3	39	Villa Lugano
4	1	Almagro
5	36	Villa del Parque
6	12	Floresta
7	32	San Telmo
8	2	Balvanera
9	30	San Cristóbal
10	27	Recoleta
11	22	Parque Avellaneda
12	41	Villa Ortúzar
13	15	Liniers
14	20	Núñez
15	9	Colegiales
16	21	Palermo
17	13	La Boca
18	0	Agronomía
19	37	Villa Devoto
20	33	Vélez Sarsfield
21	24	Parque Chas
22	5	Boedo
23	17	Montserrat
24	35	Villa Crespo
25	45	Villa Santa Rita
26	10	Constitución
27	8	Coghlan
28	6	Caballito
29	4	Belgrano
30	40	Villa Luro
31	11	Flores
32	29	Saavedra
33	28	Retiro
34	18	Monte Castro
35	25	Parque Patricios
36	14	La Paternal
37	42	Villa Pueyrredón
38	31	San Nicolás

Further Analysis

- This final table is a good starting point for a better reallocation of resources. Other variables as average income or average renting price for commercial properties, is also an information worth considering.

	Neighbourhoods	Area	Population	District	Longitude	Latitude	COFFEES	PopDensity	CoffeeDensity
19	Nueva Pompeya, Ciudad de Buenos Aires, Argentina	6.2	63276	4	-58.414710	-34.652630	0.000000	0.010206	0.00
3	Barracas, Ciudad de Buenos Aires, Argentina	7.6	77474	4	-58.387562	-34.645285	0.000000	0.010194	0.00
43	Villa Real, Ciudad de Buenos Aires, Argentina	1.3	14278	10	-58.525877	-34.618943	0.000000	0.010983	0.00
39	Villa Lugano, Ciudad de Buenos Aires, Argentina	9.0	114253	8	-58.477689	-34.674153	0.000000	0.012695	0.00
1	Almagro, Ciudad de Buenos Aires, Argentina	4.1	139262	5	-58.422233	-34.609988	0.100000	0.033966	2.94
36	Villa del Parque, Ciudad de Buenos Aires, Arge...	3.4	58573	11	-58.493821	-34.604797	0.076923	0.017227	4.47
12	Floresta, Ciudad de Buenos Aires, Argentina	2.3	39473	10	-58.483791	-34.628105	0.076923	0.017162	4.48
32	San Telmo, Ciudad de Buenos Aires, Argentina	1.2	25969	1	-58.373750	-34.621401	0.100000	0.021641	4.62
2	Balvanera, Ciudad de Buenos Aires, Argentina	4.4	152198	3	-58.403140	-34.609215	0.160000	0.034590	4.63

THANK YOU

IGNACIO KANTOR