The Battle of Neighborhoods - Curitiba

Introduction:

Curitiba is one of the biggest and most multicultural cities of Brazil. It is regarded the best in which to invest in Brazil and because of its diverse economy and cultural background it is also considered a great place to launch new business ideas, services and even products. The city is divided into 9 boroughs covering the 75 neighborhoods, each with its set of characteristics.

The same features that make Curitiba the great city that it is, also turn the task of choosing a location for a new business into a real challenge. The goal of this project is to figure out the best place to open a new special diets and healthy foods store, taking in consideration the owner's preference for a region and the profile of the neighborhoods being analysed based on local venues and the local population's demographics.

Data discussion:

The first step is using web scraping to read the wikipedia page containing a table for every borough in Curitiba and data about each neighborhood covered by the borough. The data is displayed in the web page as follows:

Bairro Novo

Bairros oficiais de Curitiba - Regional Bairro Novo (IBGE-IPPUC/2010)[3][1][5]						
Bairro	Área (km²)	Habitantes			Domicilios particulares	Rendimento mensal médio por responsáveis dos domicílios (R$)
		Homens	Mulheres	Total		
Ganchinho	11,20	3 667	3 658	7 325	1 921	767,35
Sitio Cercado	11,12	50 631	51 779	102 410	27 914	934,95
Umbará	22,47	7 280	7 315	14 595	17 064	908,70

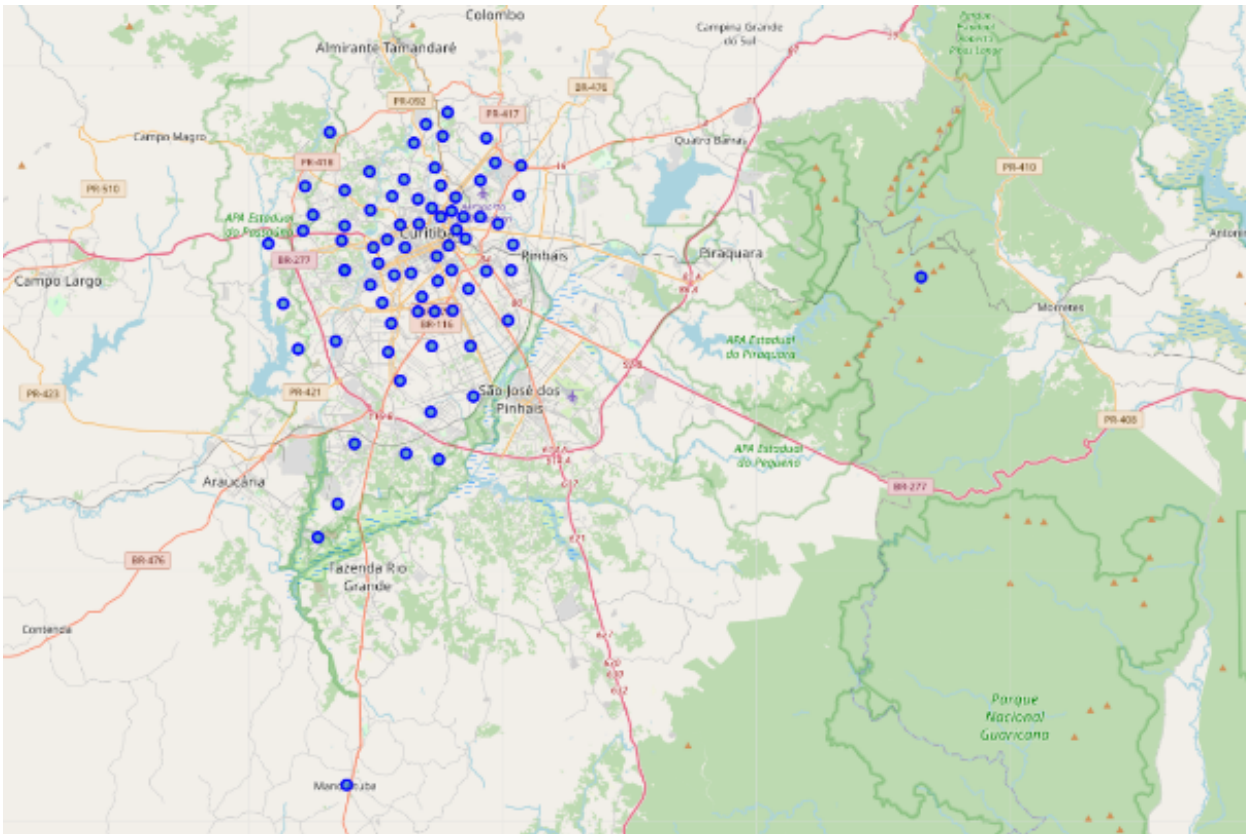
where "Bairro Novo" is the borough, the first column is the name of the neighborhoods in said borough, the second column is the area of the neighborhood in square kilometers, the third column is the number of citizens separated into men, women and total, the fourth column is the number of private homes and the fifth column is the average monthly income per citizen responsible for the home in Reais.

We know that the store is aimed toward the middle class so average income data is very important for future comparison. On the other hand the store is not aimed at a specific gender so the gender data can be dropped, along with the second and fourth columns, keeping only the relevant data on the data frame.

The second step is to add the geographical coordinates of each neighborhood to the data frame using the geopy library and, using the collected data, narrowing down the search to one specific borough.

Next, using the Foursquare location data, a second data frame is created with nearby venues of every neighborhood in the chosen borough. This data will be leveraged for clustering the neighborhoods and determining the neighborhoods' profiles.

Map of Curitiba with its neighborhoods superimposed on top:



Map of the Fazendinha Portão borough with its neighborhoods already clustered in 3 groups: