**The Battle of Neighborhoods**

**In SAO PAULO-BR**

Etienne Vanhaecke

26 May 2020

1. Introduction
2. Background

The city of Sao Paulo, Sao Paulo state in Brazil, is one of the bigger world city with more of twenty millions of habitants.

The city has multiple territory division, some of which are not compatible. This incompatibility causes doubts about spatial references, for both the population and the state representatives.

For this work, it will have use of three territory divisions:

* First level: The 32 sub prefectures.]
* Second level: The 96 districts (each district belongs to one sub prefecture).]
* Third level: The hundreds of neighborhoods. Unlike the sub prefectures and the districts, whose list is fixe, being administrative structures, there is no single list of neighborhood in Sao Paulo. For this work, we will consider 525 neighborhoods (the richest in information) on more than 1700 listed.



The clustering of these neighborhoods, in function of the more represented venues categories and in function of socio-economic indicators, should give interesting insights to help deciding where open a new business.

1. Problem

The data might help to visualize the neighborhoods with similarities about socio-economic indicators and about categories venues.

1. Interest

The executives should have interest to visualize neighborhood distributions of the São Paulo city, economic capital of Brazil and of the South America continent, to help them to define the best locals to open new business.

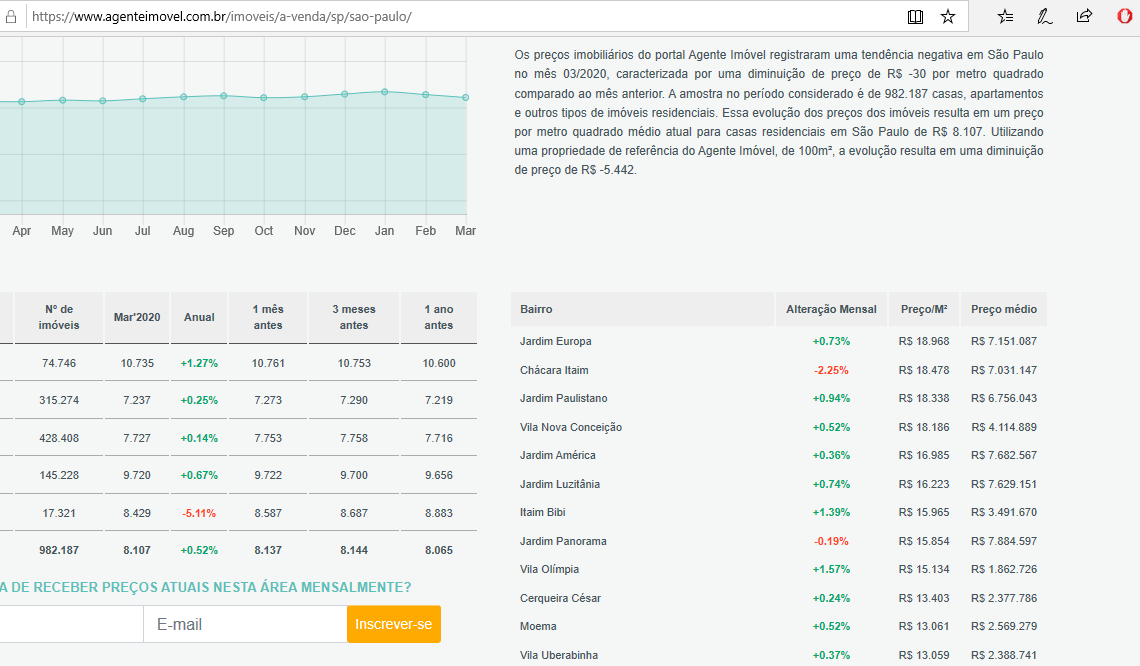
The general public, interested in geography, should also be interested in this study.

1. Data acquisition and cleaning
2. Data Source

The data have been recovered from different sources:

* The list of the neighborhood by the scrap of one table in the Web Site page:
* [**https://www.agenteimovel.com.br/imoveis/a-venda/sp/sao-paulo/**](https://www.agenteimovel.com.br/imoveis/a-venda/sp/sao-paulo/)

At most of the name of the neighborhood, it has been recovered information about the real state at march 2020, like the price at m2, but limited at 525 neighborhood.

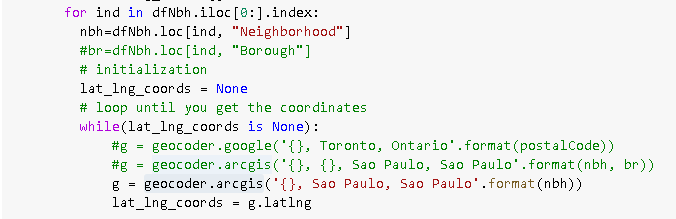


* <https://www.tudoaquisaopaulo.com.br/bairros/sao-paulo-sp>

The name of more of 1700 neighborhood referenced in Sao Paulo city.



* The latitude and longitude coordinates of the center of each neighborhood calling the API of the provider ARCGIS with the Python library Geocoder.



* The limits of the Sao Paulo city sub prefectures (32) and districts (96), downloading shape files from the official Sao Paulo web site [**http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/\_SBC.aspx**](http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx)



* The IDH (Human Development Index) by district of Sao Paulo City by scrap of the Brazilian Wiki page

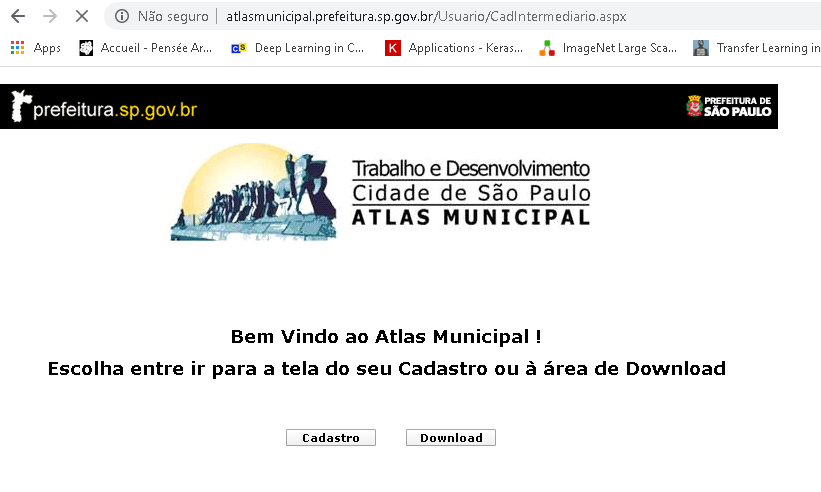
[**https://pt.wikipedia.org/wiki/Lista\_dos\_distritos\_de\_S%C3%A3o\_Paulo\_por\_%C3%8Dndice\_de\_Desenvolvimento\_Humano**](https://pt.wikipedia.org/wiki/Lista_dos_distritos_de_S%C3%A3o_Paulo_por_%C3%8Dndice_de_Desenvolvimento_Humano)

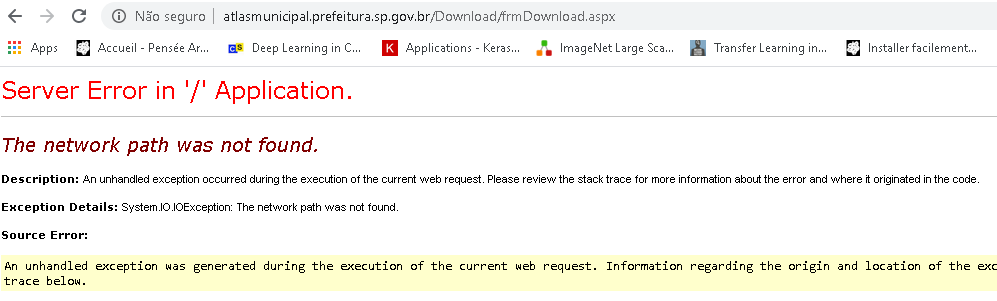


The origin of this data IDH is the *“Atlas do Trabalho de Desenvolvimento do Município de São Paulo 2007”*, available, after creating a login, with the official Web Site:

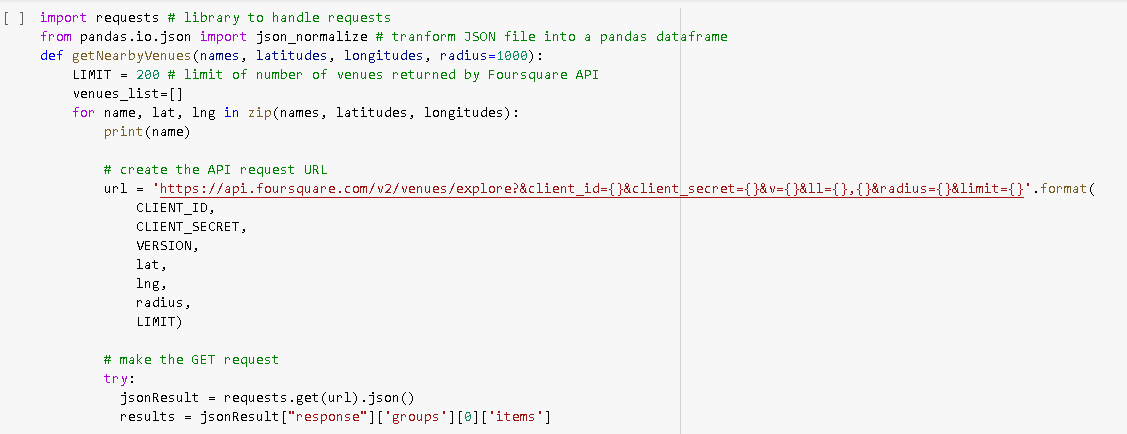
<http://atlasmunicipal.prefeitura.sp.gov.br/Login/Login.aspx>

This source should be excellent to enrich the data to use for this study, but unfortunately, the download of this archive does not currently work:





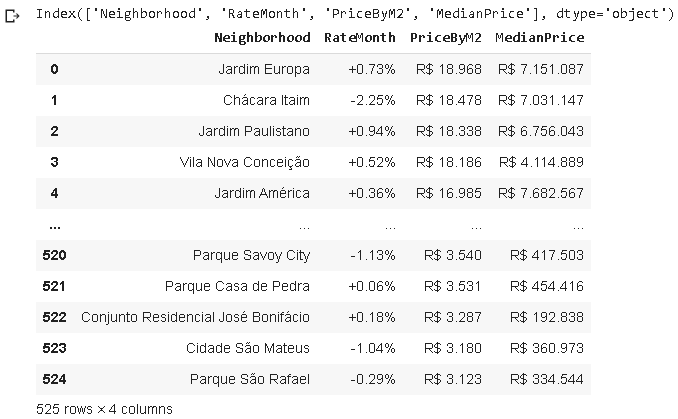
* The venues of each neighborhood has been recovered calling a API of the Foursquare site:



1. Data Transformation/Cleaning

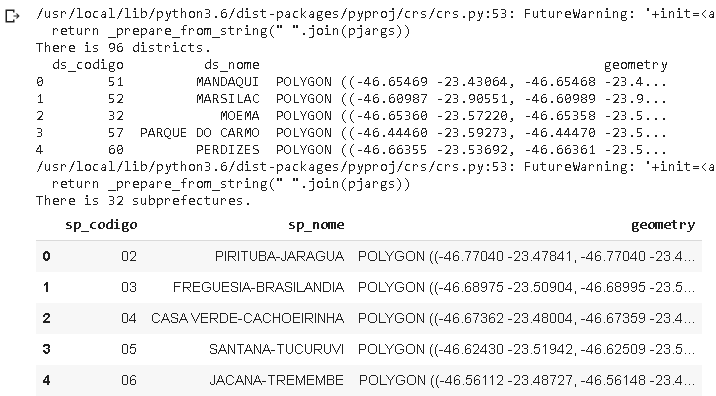
* List of the neighborhood

Like it has not be possible to find information about the complete list of the neighborhood (more of 1700), it has been used the limited list of 525 neighborhood, scraped with some information about the real state:



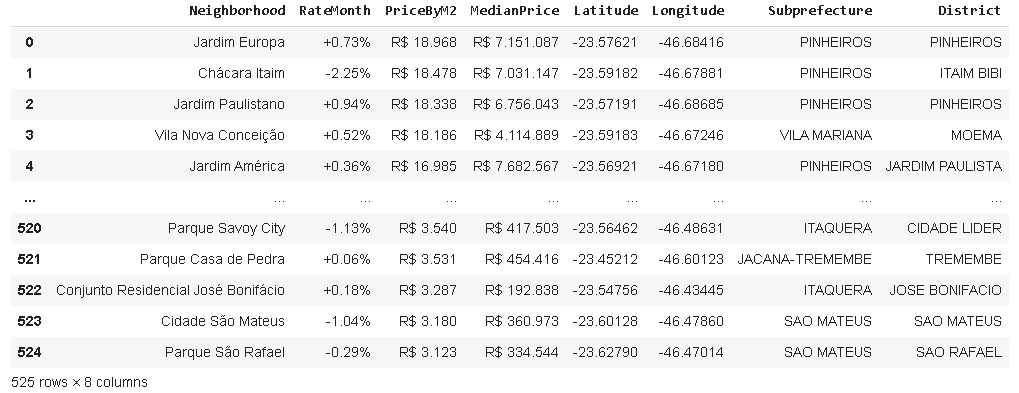
* Coordinates of the 32 sub prefectures and 96 districts:

Like the library folium, used to create map of the Sao Paulo city with his division, runs with the coordinates defined in spherical system (latitude/longitude degrees **EPSG:4326**), it has been necessary to convert, to this unit, the coordinates of the administrative limits (in form of polygon) of the sub prefecture and district, present in the shape file with the unit UTM Cartesian (**EPSG:29193)**:



* Sub prefecture and district of each neighborhood

Like it hasn't been found a web site with the indications, easy to recover, of the sub prefecture and district of each Sao Paulo city neighborhood, it has been used a geometric approach to define these: for each neighborhood loop on all the sub prefectures and all districts to find these whose the polygon contains the neighborhood central point.



1. Feature Selections

* The sub prefecture and district limits are used to show these administrative limits in a Sao Paulo map.
* The center neighborhood coordinates are showed in the same map to have an idea of the selected neighborhoods distribution between the different sub prefectures and districts of Sao Paulo city.
* The first cluster of these neighborhoods has been mounted with the top ten venues categories of each neighborhood.
* The second cluster of these neighborhoods has been mounted with the normalized data about the IDH and price/m2 of each neighborhood.