**Improving the quality of life and development in a large city**

1.INTRODUCTION

1.1.Problem

São Paulo is the most populous city in the american continent and the richest city in Brazil, but suffers from the inequality of development between the central region and the periphery.

The economic and social development of the peripheries can improve the lives of all, including the inhabitants of the central region.



São Paulo is the capital of the state of São Paulo (r.1)



São Paulo is a state of Brazil (r.1)

1.1.Audience

This report is intended for those responsible for urban development in the city of São Paulo

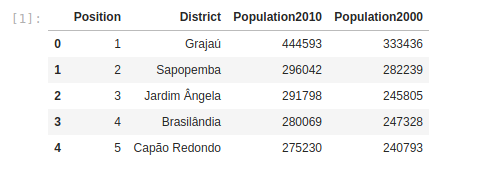
1.2. Purpose

The city of São Paulo is divided into districts (or neighborhoods). Let's compare the central and peripheral districts with regard to nearby venues, the human development index, average monthly salaries and population. Which districts deserve greater attention from the authorities?

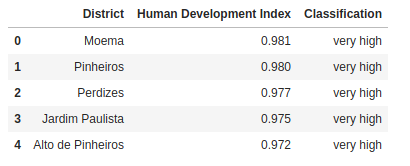
2.DATA DESCRIPTION

To compare São Paulo’s central region and peripheral region, we will study where the largest populations, the best salaries, the stores and the services are concentrated.

Let us scrap some Wikipedia pages “List of districts of São Paulo by population”(r2)

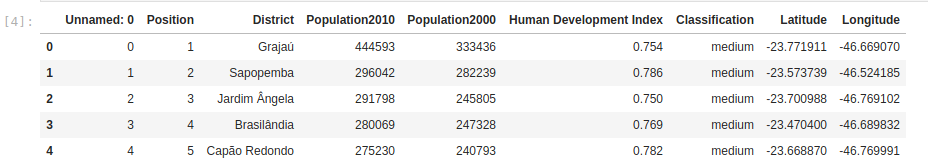


And scrap “List of districts of São Paulo by Human Development Index” (r3)

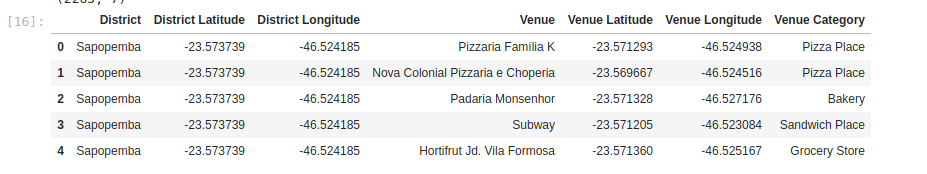


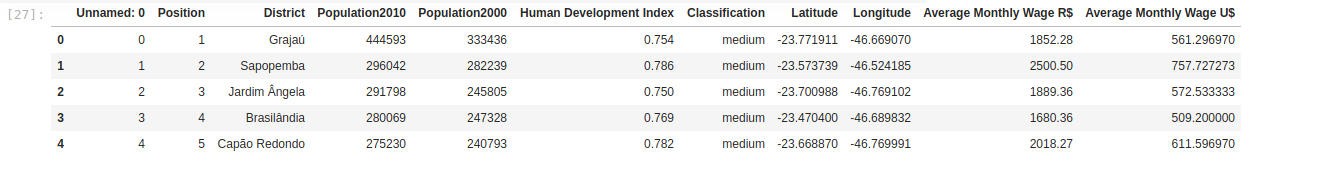
“Human Development Index (HDI) was created to emphasize that people and their capabilities should be the ultimate criteria for assessing the development of a country, not economic growth alone. The HDI can also be used to question national policy choices” (r4). And it is based on long and healthy life, knowledge and descent standard of living.

Let us merge these dataframes and look for districts’ coordinates in Google Maps (r5):



We will explore districts venues with Foursquare API (r6):

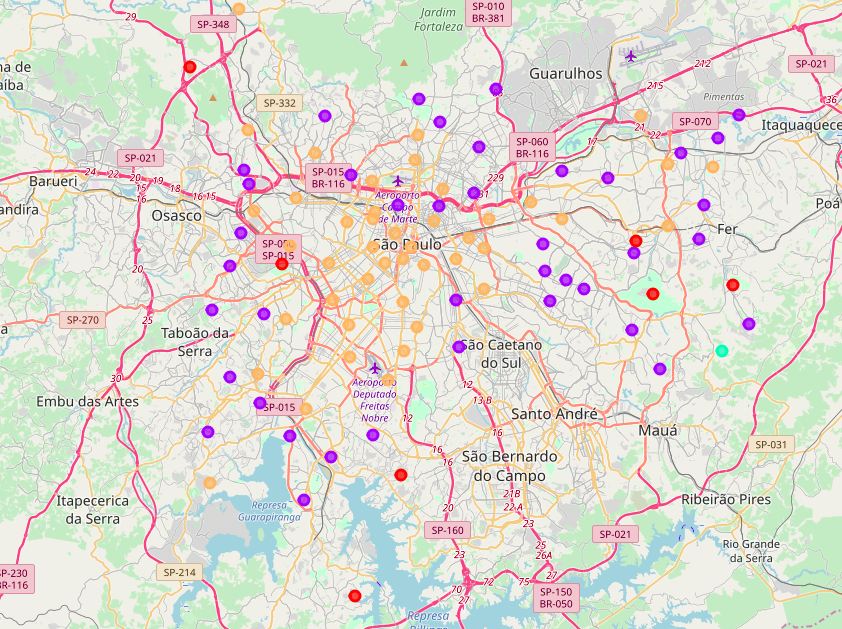
Let us scrap wages by district (r7), convert the values to dollars and merge to other dataframes:



3.METHODOLOGY

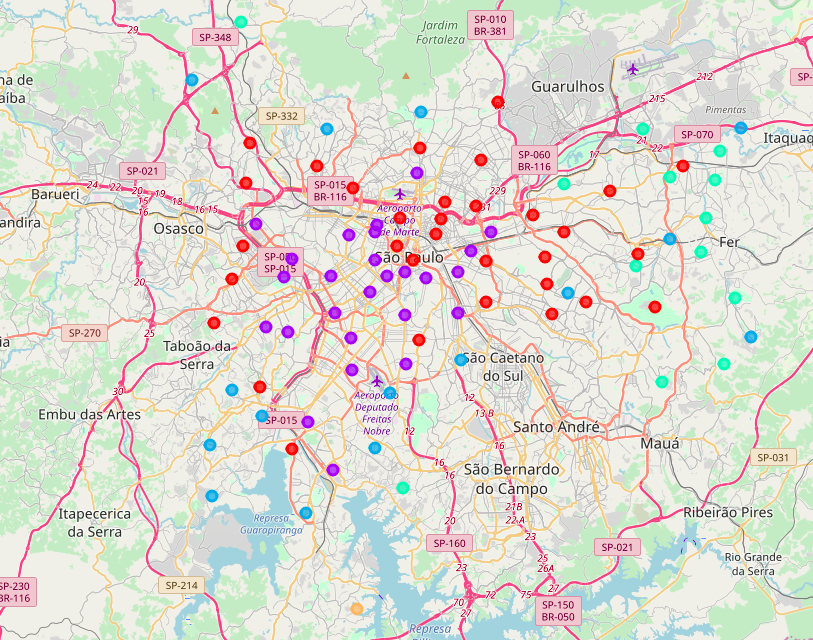
As in week 3 laboratory, we have grouped districts considering nearby venues (from Foursquare). We have used K-Means algorithm in Scikit-learn because we want to compare central districts and peripheral districts.

Here we have the Folium map with the clusters:



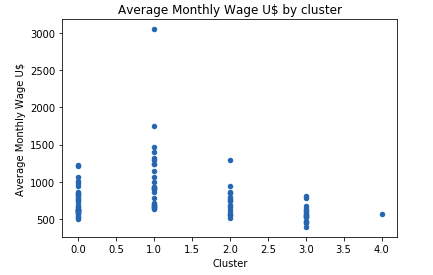
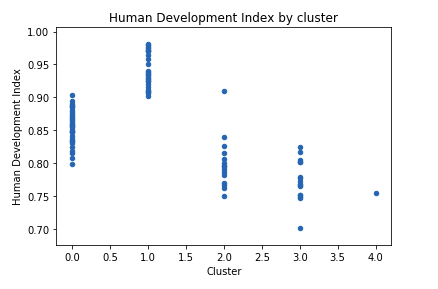
We can see the yellow group in the center of the map and the purple group more distant. And other groups (red and green) even more distant.

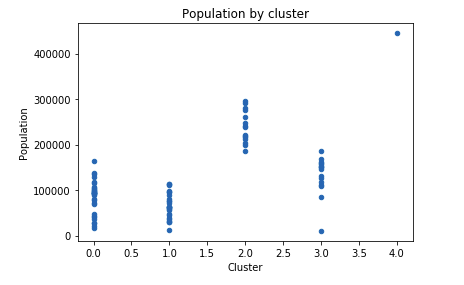
We have tried clustering considering other districts’ attributes like population, HDI, Average Monthly Wage:



We can see a red cluster in the center-north of the city, and we can see a purple cluster in the center-south. And we have blue, yellow and green clusters more distant from the center.

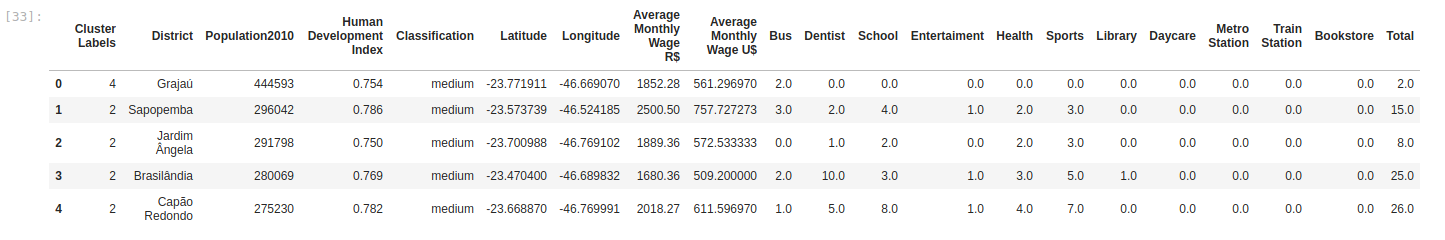
And we plot some scatter graphs with these groups from the second clustering:



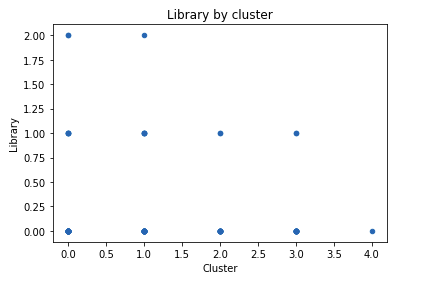
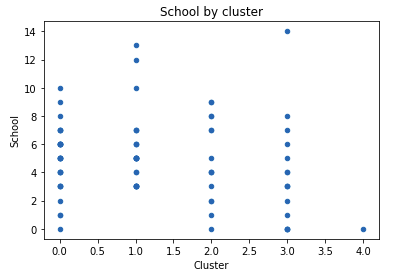
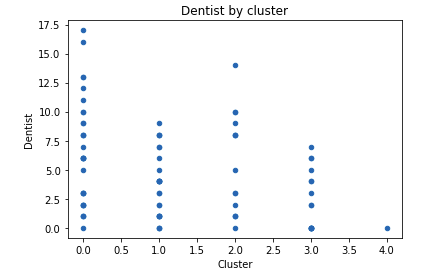
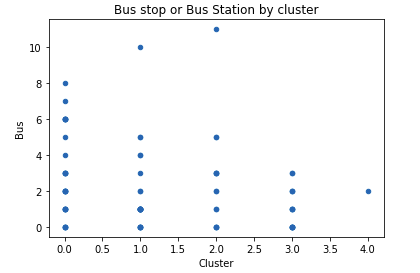


To consider venues and HDI, and wage and population, we have selected some venues categories and we have grouped and renamed the categories returned by Foursquare into a set of new “categories”:

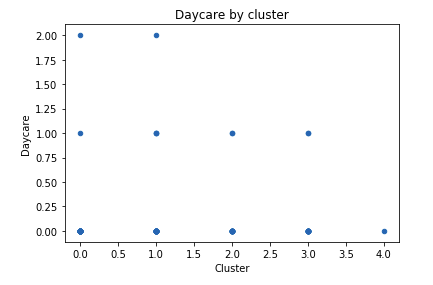
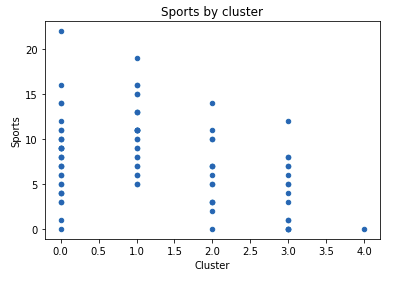


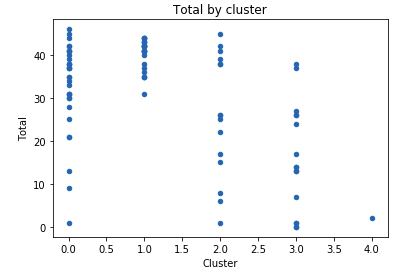
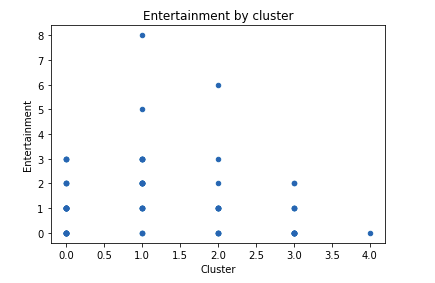


We have compared the existing clusters with the quantity of each venue.







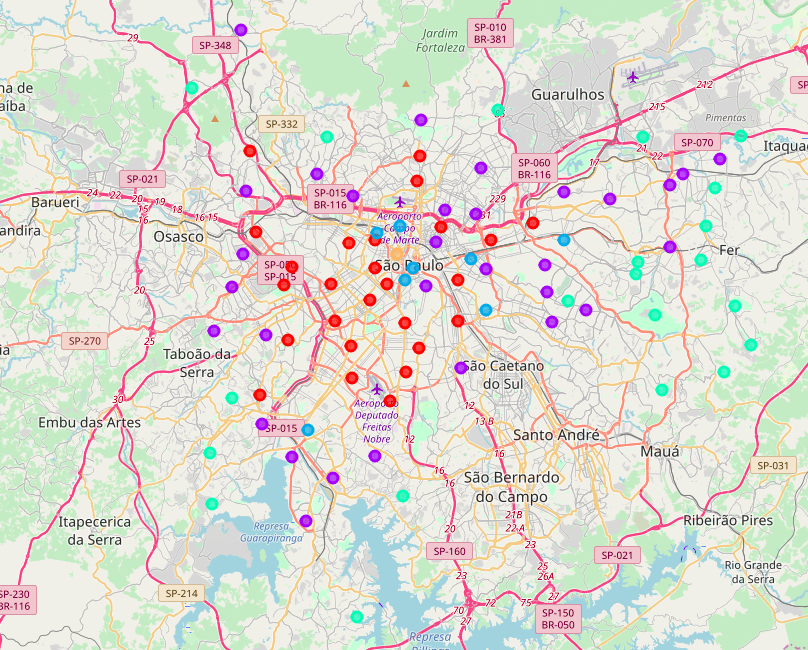


We can see the cluster 4 is very poor. It consists of one district, Grajaú.

And we see the whole city needs more libraries, bookstores and daycares.

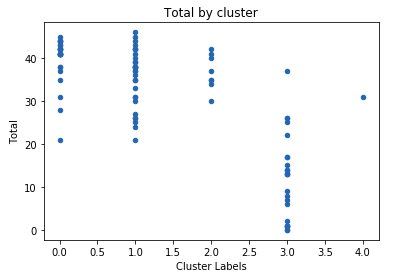
Clusters 0, 1, 2 and 3 have districts with different numbers of venues.

And we have applied k-means again considering HDI, wages, population and venues:

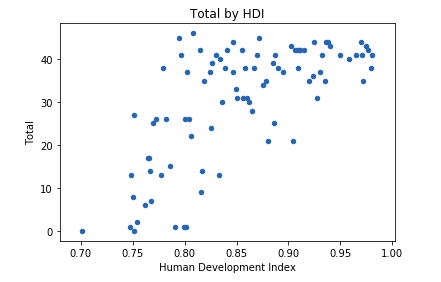


Yellow group is in the center the city (district República) and really has a nice infrastructure. Green group 3 really covers some poor districts.

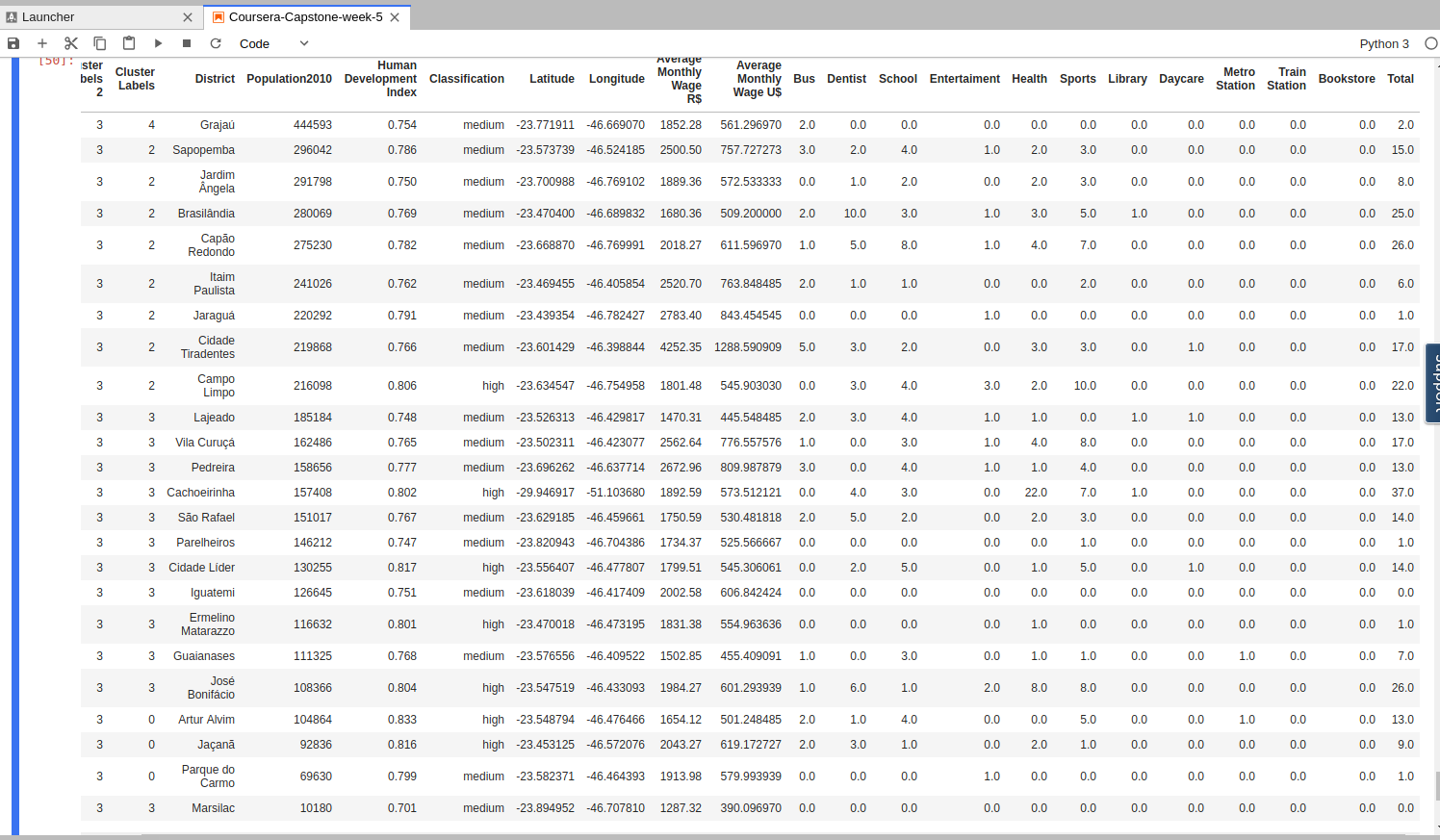
Let us plot some charts (considering the new clusters). Here we have total venues by cluster:



And here we compare total of venues by human development index:

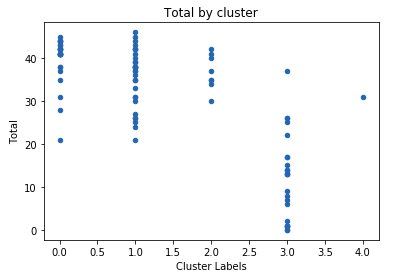


Cluster 3:

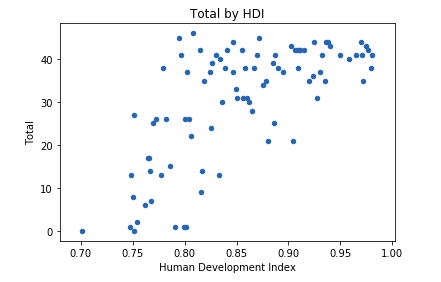


5.RESULTS & DISCUSSION

The last image shows some districts we should pay attention to. The following chart show that clustering 3 covers districts with a low number of selected venues (related to transport, health, education).



And the following chart shows that human development index accompanies the total of the selected venues:



We have limited the number of venues and distance when applying the Foursquare api, and there are libraries in schools and universities, even só, it is surprising the low amount of libraries and bookstores available throughout the city.

6.CONCLUSION

To improve human development index, people needs nice health, transport, education, entertainment. Some districts of São Paulo need improvement in these areas, mainly some districts in the extreme South and many districts of East.

The whole city needs more libraries, bookstores and daycares.

7. REFERENCES

*r1.Wikipedia - São Paulo -* [*https://pt.wikipedia.org/wiki/S%C3%A3o\_Paulo*](https://pt.wikipedia.org/wiki/São_Paulo)

*r2.Wikipedia -* *Lista dos distritos de São Paulo por população -* [*https://pt.wikipedia.org/wiki/Lista\_dos\_distritos\_de\_S%C3%A3o\_Paulo\_por\_popula%C3%A7%C3%A3o*](https://pt.wikipedia.org/wiki/Lista_dos_distritos_de_São_Paulo_por_população)

*r3.Wikipedia -* *Lista dos distritos de São Paulo por Índice de Desenvolvimento Humano -* [*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_São_Paulo_por_Índice_de_Desenvolvimento_Humano)

*r4.Human Development Index (HDI) -* [*http://hdr.undp.org/en/content/human-development-index-hdi*](http://hdr.undp.org/en/content/human-development-index-hdi)

*r5.Google Maps*

*r6.Foursquare*

*r7*.*Média de salário em SP vai de R$ 1,2 mil em Marsilac a R$ 10 mil no Campo Belo -*

[*https://g1.globo.com/sao-paulo/noticia/media-de-salario-em-sp-vai-de-r-12-mil-em-marsilac-a-r-10-mil-no-campo-belo.ghtml*](https://g1.globo.com/sao-paulo/noticia/media-de-salario-em-sp-vai-de-r-12-mil-em-marsilac-a-r-10-mil-no-campo-belo.ghtml)