

# Finding the best neighbourhoods in Rio de Janeiro to help a Retail Pharmacy business

Victor Burgardt

April 14, 2020

## Contents

<b>1</b>	<b>Introduction/Business Problem</b>	<b>2</b>
1.1	Background . . . . .	2
1.2	Problem . . . . .	3
<b>2</b>	<b>Data Section</b>	<b>3</b>
2.1	Data Gathering . . . . .	3
2.2	Data Cleaning . . . . .	4
2.2.1	Neighbourhood Dataset . . . . .	5
2.2.2	Population Density Dataset . . . . .	5
2.2.3	Human Development Index (IDH) Dataset . . . . .	6
2.2.4	Unic Health System (SUS) Dataset . . . . .	6
2.3	Result Data set . . . . .	7
<b>3</b>	<b>Methodology</b>	<b>7</b>
3.1	Exploratory Analysis . . . . .	7
3.1.1	Rio de Janeiro City Analysis . . . . .	7
3.1.2	Rio de Janeiro Regions Analysis . . . . .	8
3.1.3	Rio de Janeiro Subprefecture Analysis . . . . .	10
3.1.4	Rio de Janeiro Neighbourhoods Analysis . . . . .	12
3.2	Clustering . . . . .	13
<b>4</b>	<b>Results</b>	<b>14</b>
4.1	Regions Rio . . . . .	14
4.2	Subprefecture Rio . . . . .	15
4.3	Clusters Rio . . . . .	15
4.4	Selected Neighbourhoods from Clusters . . . . .	16
<b>5</b>	<b>Conclusion</b>	<b>16</b>

# 1 Introduction/Business Problem

## 1.1 Background

In recent years, Brazil went through a severe political-economic crisis due a serie of events but mainly this period is attributed to a strong economic recession and a corrupt government that was involved with corruption schemes and money laundering. In this time period, Brazil's economy was declining and politics were unstable making investors scared and more conservative with their investments. This scenario resulted in one of highest unemployment rates recorded in the country, with low purchasing power of families and consequently small incomes for companies and even bankruptcy in some cases. Until this day, Brazil economy is still recovering from the impacts caused by this time. Although this crisis have impacted all the industry sectors of Brazil some have recovered and are back in their tracks growing faster than before, one notorious case is the pharmacy retailers of Brazil. The Brazilian pharmacy retail business have grown immensely in these recent years turning it into a very lucrative and, nowadays, a competitive business investment. In this period, most of the Brazilian cities have experienced a big increase of the number of pharmacies in their neighbourhoods. Even with this competitive market, investors still want to invest, because Brazil is gradually recovering from the crisis, families are spending more money with medication, the growing number of medication sold, improvement in citizens life expectancy and the ageing of the population. Some interesting media headlines and statistics that reinforce this perspective are:

- Projection says medications sales in Brazil will be between US\$ 38 billions e US\$ 42 billions in 2022.
- In 2017, Brazil became the sixth largest pharmacy market in the world and it is expected to surpass France and becoming the fifth in 2022.
- Aging of the Brazilian population and growth of life expectancy.
- The largest pharmacy retail growth was in the southeast region of Brazil above the average of Brazil Growth 38,5%.
- Coronavirus triggers the sale of pharmacies and supermarkets on the internet.

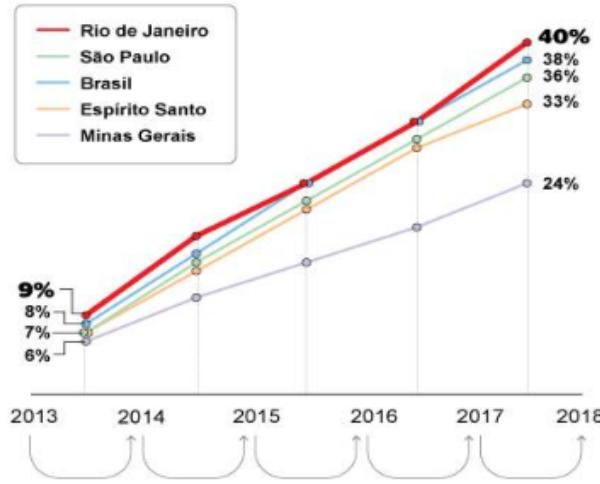


Figure 1: Number of pharmacies in the states of southeast region of Brazil [1]

Our IBM Data Science team was hired by Drogasil S.A, one of the largest retailers of medicine in Brazil, to help them with their next decision. This company is investing in new strategies to improve their customer relations as well as intelligent marketing and fast product delivery in order to make them more competitive and attractive to the population. But recently this year, their biggest competitor Drogaria SP invested

heavily in the city of Rio de Janeiro, opening new pharmacies around the city of Rio with similar strategies. With this in mind and the possibility of losing market share in one of the largest cities of Brazil, Drogasil S.A wants as soon possible to establish new pharmacies around the city combined with their new strategies. Because of the urgency of the matter they hired our expert Data Science team to find the best places to establish their new pharmacies.

## 1.2 Problem

The company Drogasil S.A hired our Data Science Team to make a project to help them clarify the question **In which neighbourhoods of Rio de Janeiro should they establish their new pharmacies ?**

After an intense research on articles and discussions with our clients we captured some key aspects that our team should have in mind while building this project. In order to build a new lucrative pharmacy, we should always look into 3 important pillars : The local market that corresponds to the Location, The numbers that relates to business Revenues and expenses and The business that relates to the strategies used and how is the management of the business. Two of these pillars are out scope of this project so the IBM team should focus on extracting as much information as it can about the local markets of each neighbourhood in Rio.

## 2 Data Section

### 2.1 Data Gathering

In order to find the best neighbourhoods in Rio de Janeiro for Pharmacy Retail business, our Data Science team will take an in-depth look in the local market of each neighbourhood. That means finding the neighbourhoods demographics, number of hospitals/clinics in the area, total population, life expectancy, average income and the amount of pharmacies. In our process, there will be 4 datasets that will be merged into one at the end of extraction:

- The first dataset consists of data about all official neighbourhoods in Rio de Janeiro. This dataset is composed of the names of each neighbourhood, region, subprefecture, latitude and longitude. This data will be gathered through web scraping of this link on Wikipedia:

The screenshot shows a table titled 'Bairros oficiais' (Official Neighborhoods) from the Wikipedia page. The table has three columns: 'Zona' (Zone), 'Subprefeitura(s) da bairros' (Subprefecture(s) of neighborhoods), and 'Bairros' (Neighborhoods). The 'Zona' column has two rows: 'Central' and 'Sul'. The 'Central' row lists 'Centro Histórico e Zona Portaria (117)' and includes neighborhoods like 'Bairro Imperial de São Cristóvão', 'Botafogo', 'Copacabana', 'Cidade Velha', 'Flamengo', 'Gávea', 'Glória', 'Humaitá', 'Ipanema', 'Jardim Botânico', 'Lagoa', 'Leblon', 'Leme', 'São Conrado', 'Urca', and 'Vigia'. The 'Sul' row lists 'Zona Sul (173)' and includes neighborhoods like 'Anil', 'Barra da Tijuca', 'Cachambi', 'Cidade das Artes', 'Curicica', 'Freguesia de Jacarepaguá', 'Gávea', 'Gloria', 'Humaitá', 'Ipanema', 'Jardim Botânico', 'Lagoa', 'Leblon', 'Leme', 'São Conrado', 'Urca', and 'Vigia'. There are also sections for 'Bairros da Tijuca e Barra da Tijuca' (20) and 'Bairros da Barra da Tijuca' (20).

Figure 2: "[https://pt.wikipedia.org/wiki/Lista\\_de\\_bairros\\_da\\_cidade\\_do\\_Rio\\_de\\_Janeiro](https://pt.wikipedia.org/wiki/Lista_de_bairros_da_cidade_do_Rio_de_Janeiro)"

- The second dataset consist of data about the demographics of each neighbourhood. The dataset is composed of Total Population, Total Area and Population Density. This data will be gathered through web scraping of all the LINKs of each neighbourhood of the previous wikipedia link. One example of scraped page :



Figure 3: "[https://pt.wikipedia.org/wiki/G%C3%A1vea\\_\(bairro\\_do\\_Rio\\_de\\_Janeiro\)](https://pt.wikipedia.org/wiki/G%C3%A1vea_(bairro_do_Rio_de_Janeiro))"

- The third dataset consists of data about Health, Education and Income about each neighbourhood in Rio. The data is composed of life expectancy, income, education index and human development index of each neighbourhood. This data will be gathered through web scraping of this link on Wikipedia:

Critérios		Classificação									
		Índice de Desenvolvimento Humano (IDH)	Evolução da expectativa de vida (em anos)	Renda per capita (em R\$)	Índice de cidadania urbana	Índice de Economia	Índice de Inovação				
1º	Bairro ou grupo de bairros										
2	Jardim Guanabara	0,97	98,92	111.159 <sup>3</sup>	0,924	0,930	0,972	0,963			
3	Gávea	0,95	98,08	118.159 <sup>3</sup>	0,924	0,957	1,000	0,970			
4	Leme	0,97	99,01	105.109 <sup>3</sup>	0,941	0,930	0,950	1,000	0,967		
5	Laranjeiras	0,97	99,43	107.959 <sup>3</sup>	0,947	0,950	0,950	0,950	0,962		
6	Botafogo	0,97	98,96	113.209 <sup>3</sup>	0,953	0,950	0,950	0,950	0,950		
7	Morumbi, Urca	75,28	95,46	112,509 <sup>3</sup>	0,947	0,888	0,990	0,979	1,000		
8	Lagoa	77,91	99,46	115.209 <sup>3</sup>	0,955	0,882	0,996	1,000	0,959		
9	Flamengo	77,91	99,28	119.009 <sup>3</sup>	1,001 <sup>7</sup>	0,882	0,995	1,000	0,959		
10	Jurujuba	77,91	99,20	122.209 <sup>3</sup>	183,059 <sup>3</sup>	0,882	0,995	1,000	0,959		
11	Maracanã	77,91	98,91	113.509 <sup>3</sup>	120,73	0,882	0,993	0,957	0,944		
12	Vila Isabel	77,91	98,53	102.509 <sup>3</sup>	0,953	0,888	0,990	1,000	0,959		
13	Bonsucesso, Jardim	77,91	99,58	110.009 <sup>3</sup>	248,47 <sup>3</sup>	0,881	0,996	1,000	0,959		
14	Laranjeiras	77,91	98,74	115.509 <sup>3</sup>	167,22 <sup>3</sup>	0,881	0,992	1,000	0,957		
15	Jardim Botânico	77,91	98,71	104.809 <sup>3</sup>	192,77 <sup>3</sup>	0,881	0,991	1,000	0,957		
16	Jardim Sulacap	77,91	97,90	107.009 <sup>3</sup>	112,93	0,888	0,989	0,957	0,944		

Figure 4: "[https://pt.wikipedia.org/wiki/Lista\\_de\\_bairros\\_do\\_Rio\\_de\\_Janeiro\\_por\\_IDH](https://pt.wikipedia.org/wiki/Lista_de_bairros_do_Rio_de_Janeiro_por_IDH)"

- The Fourth Dataset consists of data about the location of all Clinics and Hospitals of SUS “Health Unic System” . The data is composed by types of clinics, latitude, longitude, neighborhood. This dataset is downloaded at the website data.rio that is a website with open data about the city of Rio de janeiro:

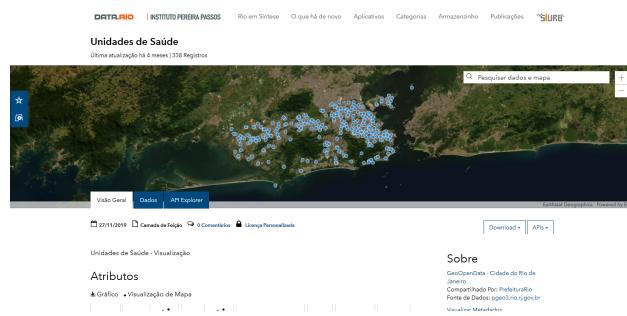


Figure 5: "<http://www.data.rio/datasets/unidades-de-sa%C3%BAde>"

## 2.2 Data Cleaning

After the Data extraction has been complete our team will focus on cleaning each Dataset Extracted separately. In this process we will treat missing values, standardize extracted data and remove unwanted and

invalid information from the Datasets.

### 2.2.1 Neighbourhood Dataset

In Neighbourhood Dataset our cleaning process was simple. Firstly we used uppercase in all extracted data, after that, we removed accent mark from Portuguese Language and at last added Latitude and Longitude from each neighbourhood to be used later in visualizations. The result dataframe is the following.

	BAIRRO	SUB	ZONA	LAT	LON
0	SAO CRISTOVAO	CENTRO HISTÓRICO E ZONA PORTUÁRIA	CENTRAL	-22.900200	-43.230240
1	BENFICA	CENTRO HISTÓRICO E ZONA PORTUÁRIA	CENTRAL	-22.889890	-43.245620
2	CAJU	CENTRO HISTÓRICO E ZONA PORTUÁRIA	CENTRAL	-22.879920	-43.222180
3	CATUMBI	CENTRO HISTÓRICO E ZONA PORTUÁRIA	CENTRAL	-22.917590	-43.197040
4	CENTRO	CENTRO HISTÓRICO E ZONA PORTUÁRIA	CENTRAL	-22.908076	-43.182324

Figure 6: The first five rows in the Neighbourhood Dataset after cleaning process

The meaning of each Feature in the result Neighbourhood Data Set

- Bairro = Neighbourhood in Rio de Janeiro
- Sub = Subprefecture in Rio de Janeiro
- Zona = Region of Rio de Janeiro
- LAT = Latitude
- LON = Longitude

### 2.2.2 Population Density Dataset

In Population Density Dataset our cleaning process was the following. Firstly we used uppercase in all extracted data, after that, we removed data discrepancies, after that we converted numeric string values to float and format theses values from hectares to square kilometers. The result dataframe is the following.

	BAIRRO	Hab	Area	Den
0	SAO CRISTOVAO	26510.0	4.1056	6457.034295
1	BENFICA	25081.0	1.7364	14444.252476
2	CAJU	20477.0	5.3475	3829.266012
3	CATUMBI	12556.0	0.5395	23273.401297
4	CENTRO	41142.0	5.7231	7188.761336

Figure 7: The first five rows in the Population Density Dataset after cleaning process

The meaning of each Feature in the result Population Density Data set

- Bairro = Neighbourhood in Rio de Janeiro.
- Hab = Total population in the neighbourhood.
- Area = Total Area in square kilometers in the Neighbourhood.
- Den = Number of Habitants in square kilometers in the Neighbourhood.

### 2.2.3 Human Development Index (IDH) Dataset

In Human Development Index Dataset our cleaning process was the following. Firstly we used uppercase in all extracted data, after that, we removed null values, after that we converted numeric string values to float. The result dataframe is the following.

	<b>BAIRRO</b>	<b>ESPER</b>	<b>RENDA</b>	<b>IDHL</b>	<b>IDHE</b>	<b>IDHR</b>	<b>IDH</b>
0	SAO CRISTOVAO	72.27	412.39	0.788	0.933	0.778	0.833
1	BENFICA	73.59	376.65	0.810	0.901	0.763	0.825
2	CAJU	68.90	236.59	0.732	0.843	0.685	0.753
3	CATUMBI	69.60	324.83	0.743	0.923	0.738	0.802
4	CENTRO	76.12	633.36	0.852	0.981	0.850	0.894

Figure 8: The first five rows in the Human Development Index (IDH) Dataset after cleaning process

The meaning of each Feature in the result Human Development Index (IDH) Data set

- Bairro = Neighbourhood in Rio de Janeiro.
- Esper = Life expectancy of Neighbourhood
- Renda = Average Income of Neighbourhood
- IDHL = Longevity Index
- IDHE = Education Index
- IDHR = Income Index
- IDH = Human Development Index

### 2.2.4 Unic Health System (SUS) Dataset

In The SUS Data set our cleaning process was the following. Firstly we removed unwanted columns such as telephone numbers, inauguration date, etc. After that we uppercase all extracted data, remove null values and create a feature for the number hospitals and clinics in the neighbourhood. The result data frame is the following.

	<b>BAIRRO</b>	<b>HOS</b>
0	ABOLICAO	1
1	ACARI	3
2	ALTO DA BOA VISTA	1
3	ANIL	1
4	BANCARIOS	1

Figure 9: The first five rows in the SUS Data set after cleaning process

The meaning of each Feature in the result Unic Health System (SUS) Data set

- Bairro = Neighbourhood in Rio de Janeiro.
- Hos = Total number of Health Establishments in the Neighbourhood.

## 2.3 Result Data set

After cleaning all four Data sets separately we are going to merge all four into one data set to be later used in machine learning algorithms, exploratory analysis and data visualizations techniques.

	BAIRRO	SUB	ZONA	LAT	LON	Hab	Area	Den	ESPER	REND	IDHL	IDHE	IDHR	IDH	HOS
0	SAO CRISTOVAO	CENTRO HISTÓRICO E ZONA PORTUÁRIA	CENTRAL	-22.90020	-43.23024	26510.0	4.1056	6457.034295	72.27	412.39	0.788	0.933	0.778	0.833	2.0
1	BENFICA	CENTRO HISTÓRICO E ZONA PORTUÁRIA	CENTRAL	-22.88989	-43.24562	25081.0	1.7364	14444.252476	73.59	376.65	0.810	0.901	0.763	0.825	3.0
2	CAJU	CENTRO HISTÓRICO E ZONA PORTUÁRIA	CENTRAL	-22.87992	-43.22218	20477.0	5.3475	3829.266012	68.90	236.59	0.732	0.843	0.685	0.753	1.0

Figure 10: The first three rows in the Result Data set after merging all four data sets

## 3 Methodology

### 3.1 Exploratory Analysis

Before jumping into statistics about the Result Dataset, lets first see some facts about the wonderful city of Rio de Janeiro. Rio de Janeiro City or just RIO as most people say is a city located in the Rio de Janeiro State in the southeast region of Brazil. This city is one of the largest cities in Brazil and is well known for their famous beaches and touristic sights such as 'Christ the redeemer' and 'Sugar loaf'. Sadly Rio is also known for being one of cities with the highest social contrast, in one hand there are very wealth neighbourhoods and in other hand we have the 'Favelas' that are informal settlement neighborhood that are undeveloped and dangerous.

Other important facts to notice about Rio de Janeiro is that RIO is the largest tourist destination in the country and in Latin America and In the southern hemisphere. Also RIO is the second largest metropolis and industrial complex of Brazil.

Having all these facts in mind we can see that Rio de Janeiro is one of the most important cities in Brazil because of it's economy, culture and industry.

In the following sections we are going to briefly analyse the Result Dataset and Statistics related to Rio de Janeiro City, Regions, Sub prefectures and Neighbourhoods.

#### 3.1.1 Rio de Janeiro City Analysis

After analysing the merged dataset our team extracted a set of informations about Rio de Janeiro. RIO has 335 public Hospitals and clinics, 6344418.0 habitants and it's area is about 1227 square kilometers. To visualize better the distributions of the important values in RIO we used Box plots, distribution histograms and Distribution bins. We can see this representations below :

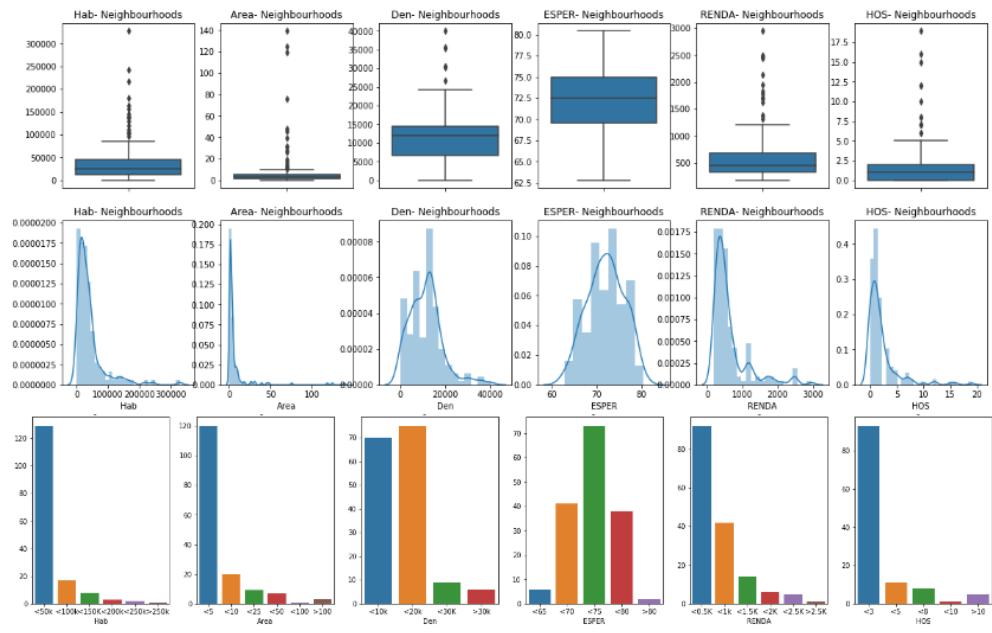


Figure 11: The Distributions of Values in Rio de Janeiro City

Analysing the distributions we can make few observation :

- Habitants : The average Hab is 40.000 and more than 80% of neighbourhood is below 50.000.
- Area : The average is 7.6 Km2 and Most of the neighbourhoods are below 5 Km2.
- Density : The average is 11.000 People/Km2 and most of neighbourhoods are not heavily populated.
- Life expectancy : The average 72 years and most of the neighbourhoods are between 70-75 years.
- Income : The average income in RIO is 622.0 Reais monthly and Almost 80% of RIO earns below 1.000 Reais monthly.
- Hospitals and Clinics : The average amount of Hospitals and Clinics is 2 per neighbourhood

### 3.1.2 Rio de Janeiro Regions Analysis

Rio de Janeiro is Divided in four region West (Oeste), South (Sul), Central (Central) and North (Norte).The neighbourhoods are divided in the following : West 39, South 17, Central 16 and North 88 Neighbourhoods. Below we can see the map with all the neighbourhoods in RIO and colored in order with each region.

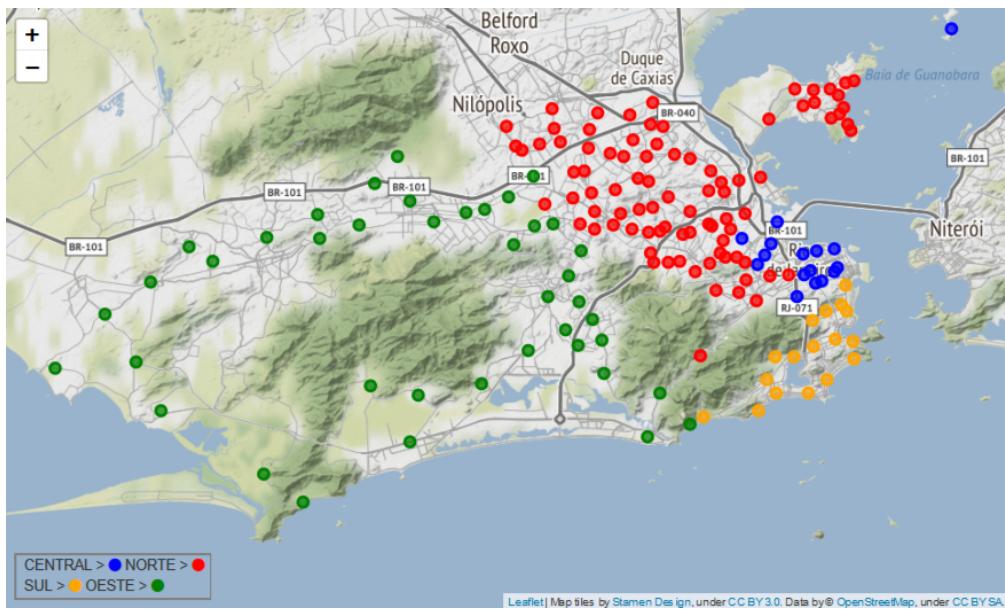


Figure 12: Regions of Rio de Janeiro

We used the same methodology as with Rio de Janeiro city to see the distribution, But this time we used colors to easily identify each region: West Green, South Orange, Central Blue and North Red.

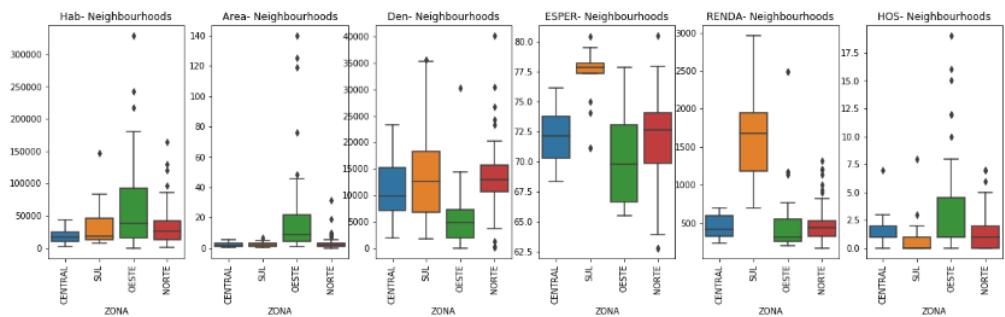


Figure 13: Regions of Rio de Janeiro BoxPlots

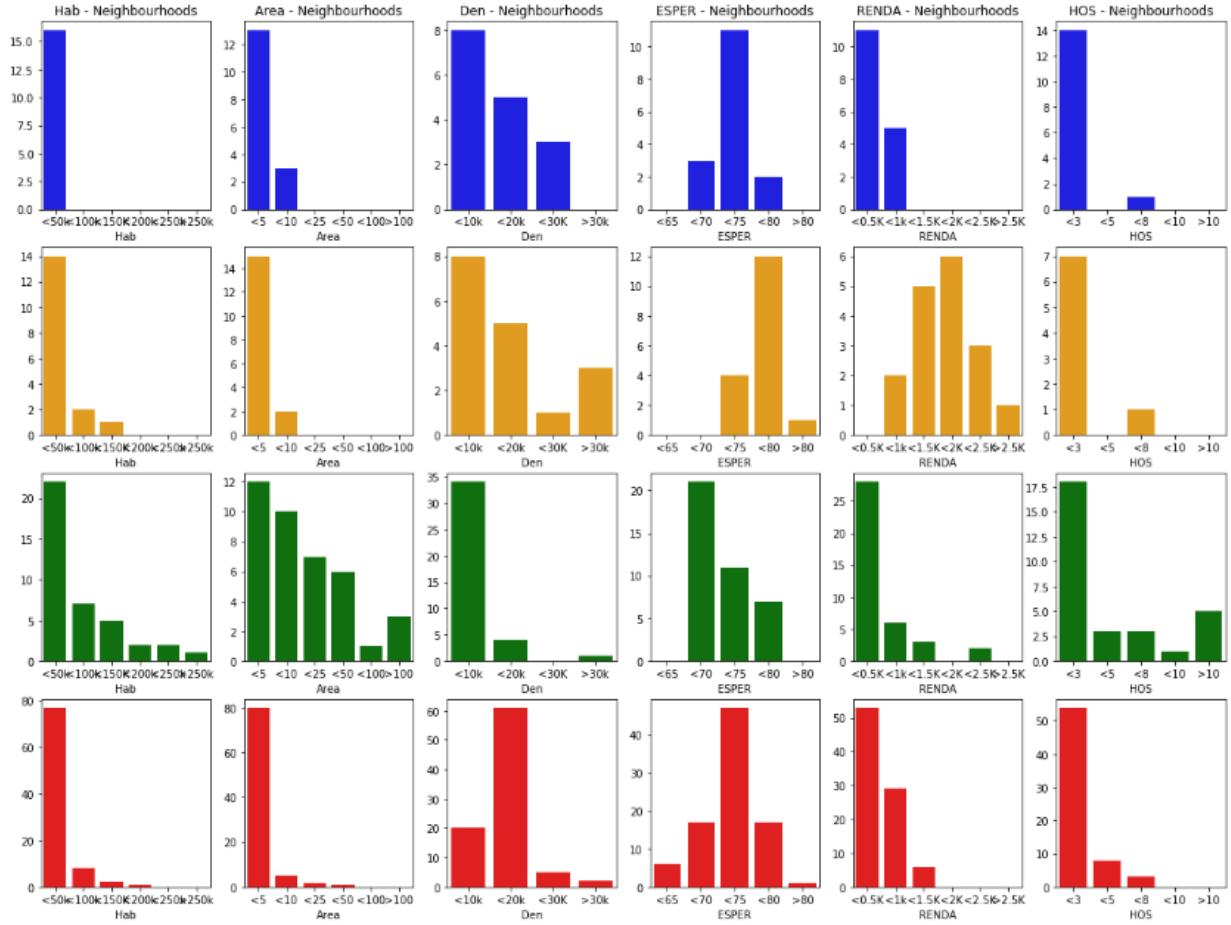


Figure 14: Regions of Rio de Janeiro Distribution Bins

Analysing the distributions we can make few observations :

- West : The West region is the largest Region, it has most populated neighbourhoods but not the most densely populated, it has the lowest average Income and life expectancy but it has the most number of hospitals.
- Central : The Central region is the smallest region, it is least populated region, second worst in life expectancy and income.
- South : The South region is the most densely populated Region with the highest income and life expectancy.
- North : The North region is the second with most habitants, second most densely populated, second largest income and life expectancy

### 3.1.3 Rio de Janeiro Subprefecture Analysis

In Rio de Janeiro there are nine subprefectures they are : 'Zona sul', 'Centro Historico', 'Baixada de Jacarepagua', 'Grande Bangu', 'Zona oeste', 'Grande Tijuca', 'Grande Meier', 'Ilha governador', 'Zona Norte'. Below is all the neighbourhoods grouped by their subprefecture.

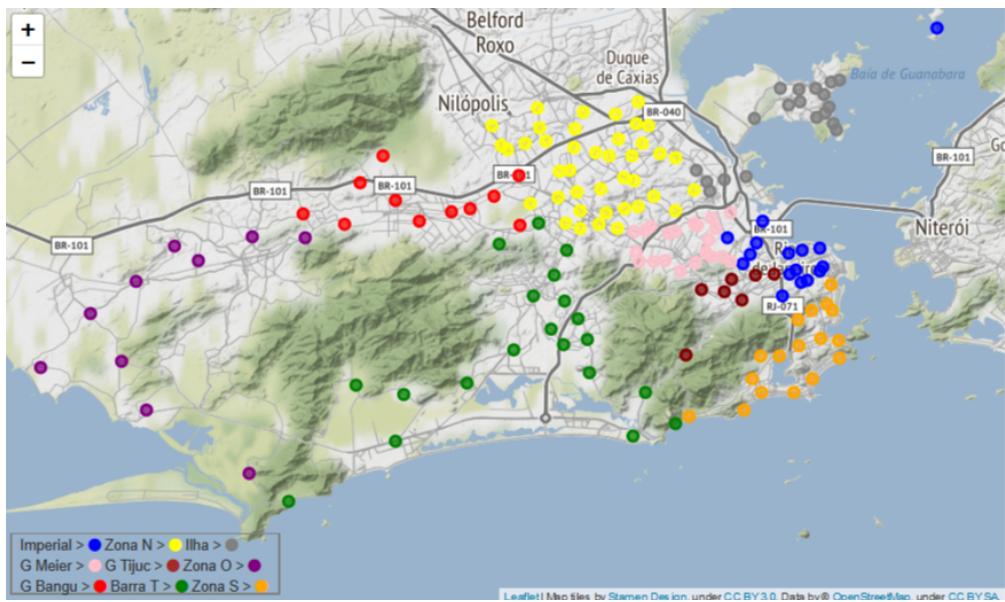


Figure 15: Regions of Rio de Janeiro BoxPlots

Applying the same methodology our team used in Rio de Janeiro Regions to see the distributions we can finding some interesting patterns.

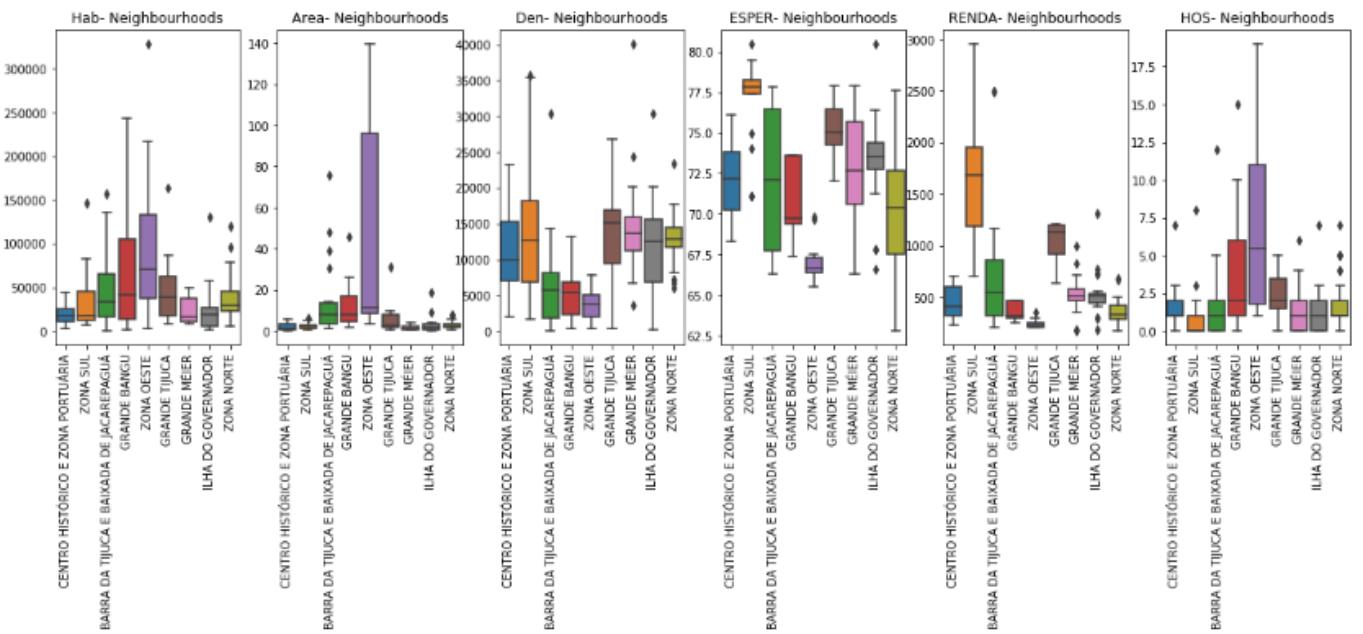


Figure 16: Subprefectures of Rio de Janeiro BoxPlots

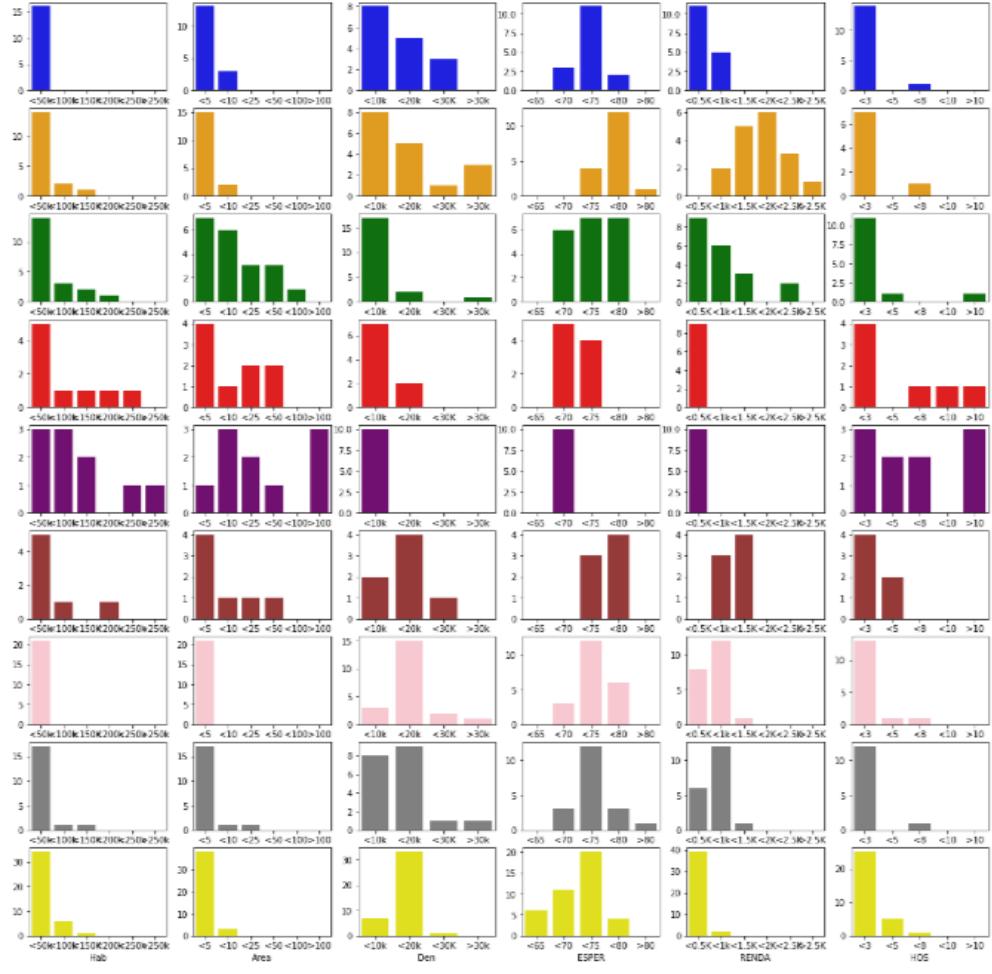


Figure 17: Subprefectures of Rio de Janeiro Distribution Bins

Analysing the distributions we can make few observation about the most notable subprefectures:

- Zona Oeste : This Subprefecture has the largest in Habitans, Area and Hospitals but the lowest in life expectancy and income.
- Zona Sul : This subprefecture is one of the most densely populated subprefecture, it has the best life expectancy and income but lowest in hospital numbers.
- Grande Tijuca : Grande Tijuca is a subprefecture that correspond wealthiest part of the North Region and it has the second best values in life expectancy, income and is the most densely populated area.

### 3.1.4 Rio de Janeiro Neighbourhoods Analysis

Since we are analysing a very large number of Neighbourhoods and there is a limitation of number of requests our team can make in Foursquare API. The team found a way to analyse only the relevant neighbourhoods, consequently it was decided to analyse only the neighbourhoods that are in the top values of the important features. To better understand these values our team made a bar chart with the Top neighbourhoods and their values. Our team also kept the color of their regions to show their representation in the Top neighbourhoods.

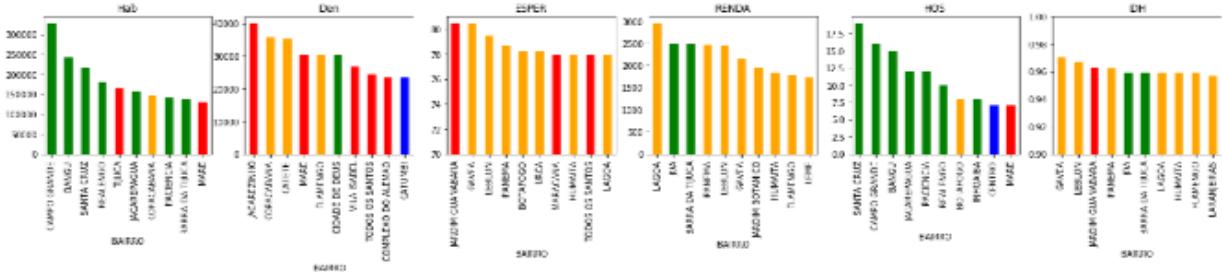


Figure 18: Top Neighbourhoods Rio de Janeiro Bar Chart

### 3.2 Clustering

To facilitate our analysis in the choice of the best neighbourhoods in Rio de Janeiro our team decided to use machine learning models to aid our problem resolution. The model created in this project use Kmeans that is an algorithm that create clusters based on patterns in the data. After some parameter tuning in the model the result map is the following :



Figure 19: Cluster Map Rio de Janeiro

Our model resulted in 7 clusters that are well distinguishable between each other. After that our team have examined each cluster, here is the following results:

- Cluster 0 : This cluster consist in a group of neighbourhood that has good values of Life expectancy, average Income, Human Development index and low competition in pharmacies.
- Cluster 1 : This cluster has low population density, bad life expectancy and income but low competition in pharmacies and good amount of hospitals near by.

- Cluster 2 : This cluster has good values in population density, life expectancy and human development index but the highest competition between pharmacies.
- Cluster 3 : This cluster has good values in population density, income and human development index but high competition between pharmacies.
- Cluster 4 : This cluster has good values in population density, income and human development index but high competition between pharmacies..
- Cluster 5 : This cluster has good values in population density and overall bad values about life expectancy, income and IDH but good values about hospitals near by and low competition
- Cluster 6 : This cluster has good values in life expectancy, income and human development index but average values of competition and low number of hospitals near by.

## 4 Results

In this section we are going to resume all results throughout the project:

### 4.1 Regions Rio

	<b>South</b>	<b>North</b>	<b>East</b>	<b>Central</b>
<b>Avg Den</b>	<b>14512</b>	<b>13194</b>	<b>5626</b>	<b>11429</b>
<b>Avg Esper</b>	<b>76.9</b>	<b>71.7</b>	<b>70.5</b>	<b>72.0</b>
<b>Avg Renda</b>	<b>1665</b>	<b>490</b>	<b>536</b>	<b>446</b>
<b>Avg IDH</b>	<b>0.93</b>	<b>0.83</b>	<b>0.81</b>	<b>0.82</b>
<b>Avg Hos</b>	<b>1.17</b>	<b>1.63</b>	<b>3.69</b>	<b>1.75</b>

Figure 20: Region Results

If we were to choose the best location to place a new pharmacy based only on the data extracted from the regions. The best choice would be the southern region of Rio de Janeiro. Because they have a better number of Average Income, Population Density, Life expectancy and Human Development Index.

## 4.2 Subprefecture Rio

	Zona portuaria	Zona sul	Barra da Tijuca e Baixada jacarepagua	grande bangu	zona oeste	grande tijuca	grande meier	ilha do governador	Zona Norte
Avg Den	11429	14512	6444	5876	3765	13519	14848	12096	12801
Avg Esper	72.0	76.9	72.1	70.6	67.1	75.1	72.7	73.3	70.0
Avg Renda	446	1665	767	348	244	1031	530	544	351
Avg IDH	0.82	0.93	0.84	0.81	0.76	0.91	0.85	0.85	0.80
Avg Hos	1.7	1.1	1.7	4.1	7.3	2.4	1.5	1.3	1.6

Figure 21: Subprefecture Results

If we were to choose the best location to place a new pharmacy based on the Subprefectures data extracted from the wiki pages. We would have two options to choose "Zona Sul" and "Grande Tijuca". "Zona Sul" has the great values on : Average Income, Life expectancy and Human Development Index but the worst value on the number of hospitals in the area. In other hand 'Grande Tijuca' have lower values than 'Zona sul' but has more hospitals. If the number of hospitals has a big impact on the choice of pharmacy placement 'Grande Tijuca' is the best choice.

## 4.3 Clusters Rio

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Avg Den	3538	8689	14875	29910	26027	3403	11197
Avg Esper	77.8	69.3	77.2	76.7	69.8	67.8	78.5
Avg Renda	2471	428	1003	1519	377	277	1651
Avg IDH	0.95	0.81	0.92	0.93	0.79	0.77	0.95
Avg Hos	1.2	6.1	6.3	2.2	2.6	14.8	0.7
Avg Phar	8.0	8.1	48.3	45.2	8.2	5.0	27.5

Figure 22: Cluster Results

If we were to choose the best location to place a new pharmacy based on data extracted from our clusters. We would have two options, the first one Cluster 0 have great values on : Average Income, Human Index values and Number of Pharmacies (Low Competition). The second option Cluster that has better life expectancy and population density.

#### 4.4 Selected Neighbourhoods from Clusters

	Avg Den	Avg Esper	Avg Renda	Avg IDH	Avg Hos	Avg Phar
JARDIM BOTANICO	6696	77.8	1952	0.95	0	5
LAGOA	4148	77.9	2955	0.95	0	10
BARRA DA TIJUCA	2822	77.8	2488	0.95	5	17
JOA	484	77.8	2488	0.95	0	0
GAVEA	6774	80.4	2139	0.97	1	28
HUMAITA	12598	77.9	1830	0.95	0	36
IPANEMA	13855	78.6	2465	0.96	0	40
LARANJEIRAS	18269	77.8	1679	0.95	0	30
LEME	15144	77.4	1713	0.95	1	15
URCA	3044	78.2	1376	0.95	0	31
GRAJAU	6738	77.8	1134	0.93	2	16
MARACANA	15147	77.9	1206	0.94	3	22
JARDIM GUANABARA	9204	80.4	1316	0.96	0	30

Figure 23: Selected Neighbourhoods Results

Cluster 0 corresponds a small set of neighbourhoods that have highest Income and Low number of Pharmacies in the area. Cluster 6 corresponds a small set of neighbourhoods that have highest Human Development Index and Life expectancy. The neighbourhoods that have shown the best results were : Jardim Botanico, Lagoa, Barra da Tijuca, Joa, Gavea and Leme.

## 5 Conclusion

This project helps Drogasil to choose more wisely their next retail pharmacy placement based on exploring a series of data such as: Life expectancy, Number of Hospitals, Average income, Human development index, population density and most importantly how many pharmacies there is in each neighbourhood. With these results our team was able to find a set of neighbourhoods that could be a good place to establish a retail pharmacy.

In conclusion the best neighbourhoods to establish a new pharmacy of Drogasil S.A in Rio de Janeiro are:

	Region	Subprefecture	Population	Area	Avg Den	Avg Esper	Avg Renda	Avg IDH	Avg Hos	Avg Phar
LAGOA	Sul	Zona Sul	21198	5.1	4148	77.9	2955	0.95	0	10
GAVEA	Sul	Zona Sul	17475	2.5	6774	80.4	2139	0.97	1	28
LE ME	Sul	Zona Sul	14799	0.9	15144	77.4	1713	0.95	1	15
JARDIM BOTANICO	Sul	Zona Sul	18009	2.6	6696	77.8	1952	0.95	0	5
BARRA DA TIJUCA	Oeste	Barra da tijuca baixada jacarepagua	135924	48.1	2822	77.8	2488	0.95	5	17
JOA	Oeste	Barra da tijuca baixada jacarepagua	818	1.6	484	77.8	2488	0.95	0	0

Figure 24: Best neighbourhoods from selected Clusters



Figure 25: Neighbourhoods of choice Map Rio de Janeiro

The Neighbourhoods selected to establish a new pharmacy from Drogasil S.A are composed of two groups : The first one is a small group of neighbourhoods from the south region of RIO that is that is wealthiest region of RIO, in this small group we selected neighbourhoods with low pharmacy competition. The second group is similar to the first, it is a small set of neighbourhoods from the East region that has good wealth and population of age older people and most of importance low competition in the retail pharmacy business. The other neighbourhoods haven't been chosen because of low average income, low life expectancy or Too much pharmacy competition.