Find Location to Shop for a new Coffee Shop in Brazil

Juliano Magalhães da Silva

August 3, 2019

1. Introduction

1.1 Background

A Coffee Shop Group will open stores in Brazil and they choose start for the two biggest cities: **São Paulo and Rio de Janeiro**, they request the list of neighborhoods that are better to open new stores.

The characteristic from the Group: the principal product are coffee and drinks based in coffee. In addition, teas, milk and chocolate. All products with high quality. They sell snacks and fast food during lunch and dinner. Every store are have comfortable tables a nice ambience and decoration.

This group in other countries are compared with Starbucks, Tim Hortons, McDonald's Coffee and Second Cup.

1.2 Problem

Where to open new shops?

Coffee shops are very common and already there are various group operating in Brazil.

To archive the answer I will analyze two approaches, crossing the competition, their location and characteristics, with the neighborhood, and get:**Strategic places.** The principal places the competition exist and have more shops; this place can be strategic places to open a store, with a high flow of people that consume coffee.

The approach is to identify places that already have shops alike the client or business correlated, and can support this kind of business, that demand a specific public, special in Brazil where coffee shop generally are more simple.

1.3 Interest

The client that request this analysis and other company in the same business.

2. Data acquisition and cleaning

2.1   Data sources

To generate this information, is necessary a few information:

- Cities Neighborhood for São Paulo and Rio de Janeiro;
- Neighborhood geolocation;
- Categories correlated with client business;
- Venues (shops) from each Neighborhood, filtered by category;
- Venues details;

Cities Neighborhood get from https://www.estadoecapitaisdobrasil.com that have a list of Cities neighborhoods.

With Neighborhood name, query the service: Nominatin – Open Street Map (https://wiki.openstreetmap.org/wiki/Nominatim) and discover the Neighborhood geolocation.

Foursquare API (https://pt.foursquare.com/developers/apps):

- Discovery categories that the Client can consider as direct competition.
- Query the neighborhood's geolocation, associate the categories selected to get the venues from each region.
- Query venue's detail for price of venues.

2.2   Data cleaning

Neighborhood and geolocation have a good result and did not need a cleaning. Just parsed to be used in future.

Foursquare impose some difficult to get the venues location with category, and five result of the same venue in more than one answer. To clean this information, calculate the distance from each neighborhood that appear in the answer, and remove duplicate venues ID, keeping the one with small distance.

Other problems is that border neighborhood get venues from other cities, and this are solved dropping this information.

A few Venues Group have more than one categories, it is normalized with the categories that have more matches. A few venues that was form the same group have some wrong name that just overwrite with the right information.

The Venue price is available only in premium query that have a limited query (500). This case collected a sample, and extrapolating the value to other venues in the same group.

2.3   Feature selection

There are 4938 venues in seven categories.

Groups of company are bigger company or franchise groups that have procedures, rules and analyses the place where the brand / shop will have better chances of prosper and locations that are good for the business. With this prerogative, create this information in the database, a feature that identify the group size.

Considering venues with same name are part of a group or same company, since there are:

```
Total of Venues:   4938
Total of Venues that area part of a Group: 1823 (36.92%)
Number of Groups: 224
```

The feature *v_total* represent percentage when the venue is part of a group and how big this group are compare with others. The bigger group have 100%, and it regard the number of venues this group have.

The feature *top_n* identify with value 1 the venues that are between the 15 venue's groups with more shop in each city.

v_price are extra information get from foursquare, range from 0 to 3, that represent the price for each venue. As this query are limited to get this information, this value are extrapolated to entire groups, and focus in query for groups.

Feature *cat_score* is score based in how the Client Group operate in this market. Each venue will received this value according the following table:

	v_cat	cat_score
0	Coffee Shop	5
1	Tea Room	5
2	Snack Place	4
3	Chocolate Shop	4
4	Sandwich Place	3
5	Bakery	2
6	Breakfast Spot	4

There are two features to identify the places with more stores, which are part of group, and with high price.

During the analysis, realize that the neighborhood are huge and the result was not precise. To avoid this problem, create a *Cluster* with Venues, by distance between. Find the most accurate locations, and will use the area of this cluster as a variable that improve small clusters.

Venue Score:

Venue Score:

```
* 45% cat_score - Score according Venue Categorie. High value, high
score.
* 5% top_n - Venue is part of the top 15 bigger Group from each city.
* 15% v_total - Number of Venues in that Group.
* 35% v_price - Price acquiring from Venue Detail from FourSquare.
More expensive high score.
```

The feature Cluster that represent the location of the venue, have a score calculate with the formula. This formula identify places more concentrated.

Cluster Score:

```
* 70% Sum Venue Score
* 30% Cluster Area - Small cluster have a high score.
```

3. Exploratory Data Analysis

3.1   Categories

Choose key words to identify categories in Foursquare.

This words was choose how describe the client group area of actuation, and are:

Café, Cafe, Coffee, Snack, Chocolate, Bakery, Tea, Breakfast, Fast Food and Sandwich

From Foursquare find the following Categories:

	categories	subcat	id	filter
38	College & University	College Cafeteria	4bf58dd8d48988d1a1941735	Cafe
77	Food	Bakery	4bf58dd8d48988d16a941735	Bakery
80	Food	Breakfast Spot	4bf58dd8d48988d143941735	Breakfast
81	Food	Bubble Tea Shop	52e81612bcbc57f1066b7a0c	Tea
84	Food	Cafeteria	4bf58dd8d48988d128941735	Cafe
85	Food	Café	4bf58dd8d48988d16d941735	Café
89	Food	Coffee Shop	4bf58dd8d48988d1e0931735	Coffee
102	Food	Fast Food Restaurant	4bf58dd8d48988d16e941735	Fast Food
131	Food	Pet Café	56aa371be4b08b9a8d573508	Café
138	Food	Sandwich Place	4bf58dd8d48988d1c5941735	Sandwich
143	Food	Snack Place	4bf58dd8d48988d1c7941735	Snack
150	Food	Tea Room	4bf58dd8d48988d1dc931735	Tea
296	Shop & Service	Chocolate Shop	52f2ab2ebcbc57f1066b8b31	Chocolate
327	Shop & Service	Gaming Cafe	4bf58dd8d48988d18d941735	Cafe
340	Shop & Service	Internet Cafe	4bf58dd8d48988d1f0941735	Cafe

After query FourSquare filtering by this categories, find the venues, and after normalize and clean up, this is the result:

	Category	Total_Venues	Unique_Venues	Total_Perc	Unique_Perc
0	Coffee Shop	1610	1121	22.70%	69.63%
1	Bakery	1215	1140	23.09%	93.83%
2	Snack Place	794	723	14.64%	91.06%
3	Chocolate Shop	537	59	1.19%	10.99%
4	Sandwich Place	411	190	3.85%	46.23%
5	Tea Room	264	36	0.73%	13.64%
6	Breakfast Spot	107	101	2.05%	94.39%

The *Unique_perc* represent if the venue are part of groups or individuals shops. In this table is visible that Bankery, Snack Place and Breakfast Spot are all most individual companies. In other point that Chocolate Shop and Tea Room are most compose by groups. In the middle, there are Coffee Shoop and Sandwich Place that are a mixed market.

```
          Top  5  Venues per Categories - Identify Competition

Categorie:  Coffee Shop
Total venues: 1610
Unique venues: 1121  ( 69.63% )
                            Qnt
venue
Starbucks                    81
Casa do Pão de Queijo        59
Grão Espresso                31
California Coffee            21
Fran's Café                  21
Sum Top  5 : 213
----------------------------------------
Categorie:  Bakery
Total venues: 1215
Unique venues: 1140  ( 93.83% )
                            Qnt
venue
Benjamin A Padaria           8
Dona Deôla                   6
Le Pain Quotidien            6
Padaria Michelli             5
Casa do Pão                  4
Sum Top  5 : 29
----------------------------------------
Categorie:  Snack Place
Total venues: 794
Unique venues: 723  ( 91.06% )
                            Qnt
venue
Mr. Pretzels                11
Bibi Sucos                  10
Croasonho                    8
Big Bi                       7
Kroke Salgados               6
Sum Top  5 : 42
----------------------------------------
```

```
Categorie:  Chocolate Shop
Total venues: 537
Unique venues: 59  ( 10.99% )
                                Qnt
venue
Cacau Show                      253
Kopenhagen                      117
Chocolates Brasil Cacau          60
Lindt                            20
Katz Chocolates                   8
Sum Top  5 : 458
----------------------------------------
Categorie:  Sandwich Place
Total venues: 411
Unique venues: 190  ( 46.23% )
                                Qnt
venue
Subway                          179
Porto do Sabor                   15
Balada Mix                        5
Bolados Sucos                     4
Lanchonete da Cidade              4
Sum Top  5 : 207
----------------------------------------
Categorie:  Tea Room
Total venues: 264
Unique venues: 36  ( 13.64% )
                                    Qnt
venue
Rei do Mate                         110
MegaMatte                            71
EVS - Espaço Vida Saudável Herbalife  47
Talchá                                4
Açai com Mate                         1
Sum Top  5 : 233
----------------------------------------
Categorie:  Breakfast Spot
Total venues: 107
Unique venues: 101  ( 94.39% )
                                Qnt
venue
Café da Manhã                     5
Ibis Kitchen                      2
Restaurante                       2
A Bruncheria                      1
Padaria Nova Karina               1
Sum Top  5 : 11
----------------------------------------
```
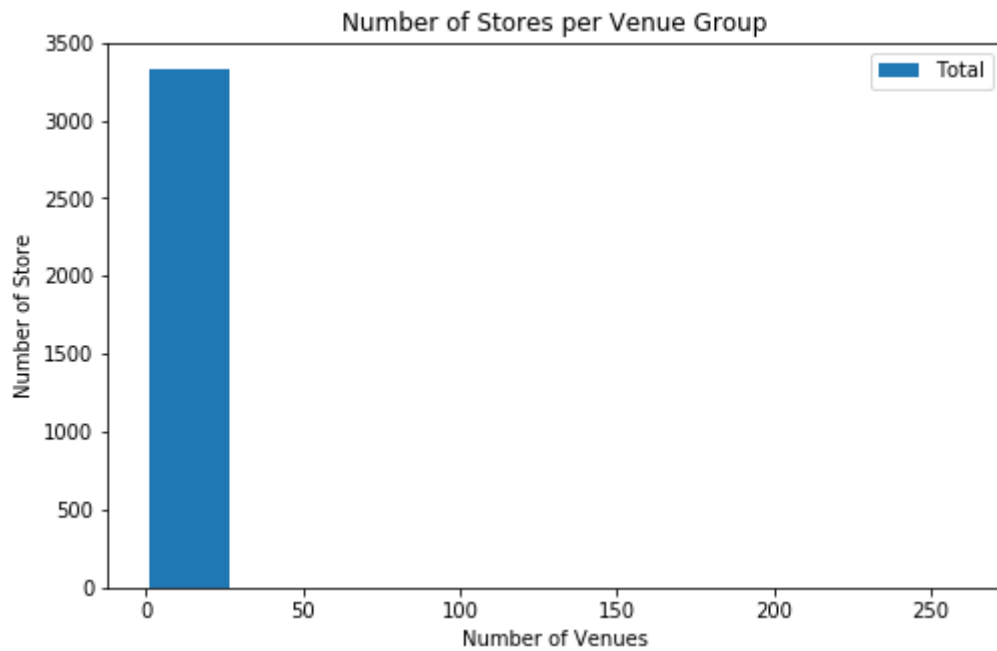
3.2   Venues

Found 4938 venues, and

The information show that 67% of the venues are individual / unique.

```
Total of Venues:  4938
Total of Venues that area part of a Group: 1823 (36.92%)
```
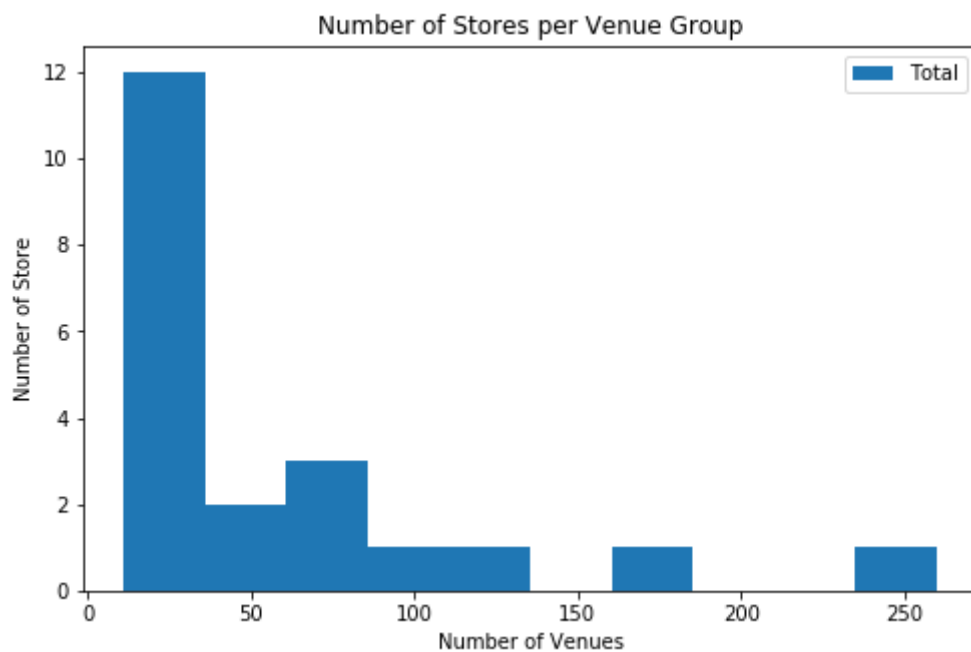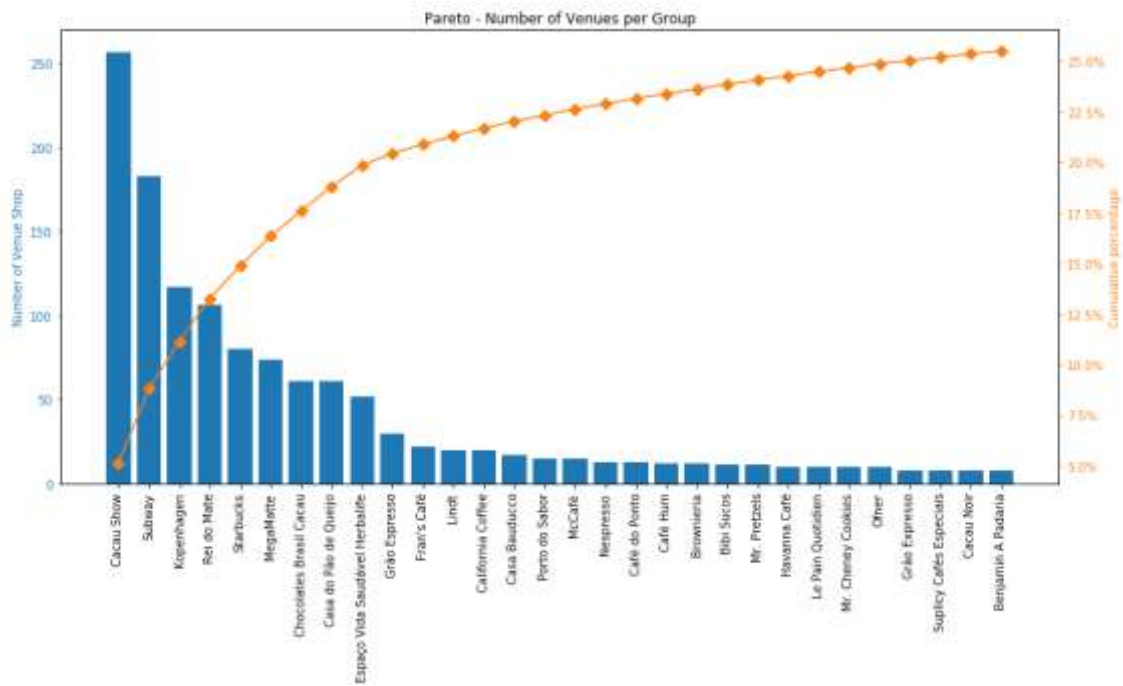
Number of Stores per Venue Group

3.3   Groups of Venues:

Observing the groups, there are 25% of Venues in these groups that are direct competition.
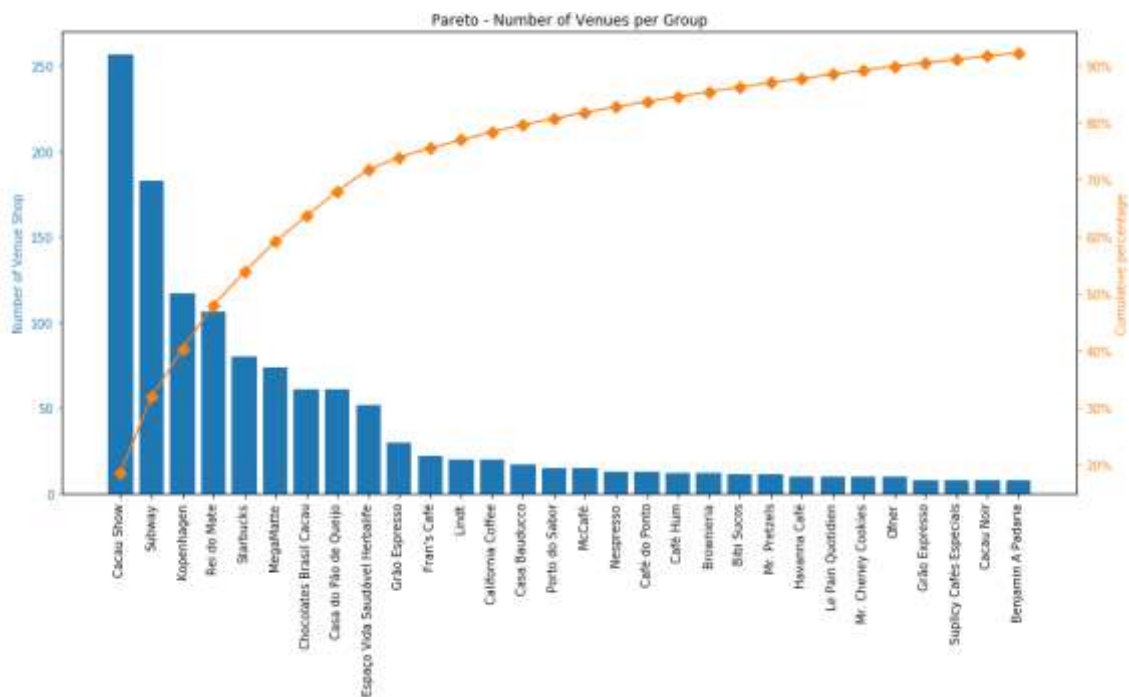
Considering that this groups have rules, process to choose places that have a high flow of people, and with the correct characteristics, this venues are in places that the client have interest.

Number of Stores per Venue Group

The groups represent 25% of the venues:

Pareto - Number of Venues per Group

The venues with more stores:

Pareto - Number of Venues per Group

3.4   Venues by Category

Identify the top group of venues, and the number of Stores that this group have.

A manual exploratory in each venue group identified that most are franchise, and have a process to choose location to new stores.

```
Venue Groups by Category:
                               Venue_Groups          Category   Qnt
Top
1                               Cacau Show     Chocolate Shop    252
2                                  Subway     Sandwich Place    183
3                              Kopenhagen     Chocolate Shop    117
4                              Rei do Mate          Tea Room    106
5                               Starbucks        Coffee Shop     80
6                               MegaMatte          Tea Room     74
7                     Casa do Pão de Queijo       Coffee Shop     61
8                 Chocolates Brasil Cacau     Chocolate Shop     56
9    EVS - Espaço Vida Saudável Herbalife          Tea Room     52
10                           Grão Espresso        Coffee Shop     30
11                             Fran's Café        Coffee Shop     22
12                       California Coffee        Coffee Shop     20
13                                   Lindt     Chocolate Shop     20
14                           Casa Bauducco        Coffee Shop     17
15                                  McCafé        Coffee Shop     15
16                          Porto do Sabor     Sandwich Place     15
17                           Café do Ponto        Coffee Shop     13
18                               Nespresso        Coffee Shop     13
19                               Brownieria        Coffee Shop     12
20                                Café Hum        Coffee Shop     12


        City:  São Paulo
        Total venues: 2941
        Unique venues: 2053
                                          Total
        venue
        Cacau Show                          188
        Subway                               97
        Kopenhagen                           72
        Starbucks                            53
        Casa do Pão de Queijo                39
        EVS - Espaço Vida Saudável Herbalife 39
        Rei do Mate                          35
        Chocolates Brasil Cacau              34
        Grão Espresso                        26
        Fran's Café                          18
        Casa Bauducco                        16
        Lindt                                14
        Ofner                                10
        Café do Ponto                        10
        Mr. Pretzels                         10
        Sum Top  15 : 661
```

```
City:  Rio de Janeiro
Total venues: 2053
Unique venues: 1374
                                          Total
venue
Subway                                       86
MegaMatte                                    74
Rei do Mate                                  71
Cacau Show                                   69
Kopenhagen                                   45
Starbucks                                    27
Chocolates Brasil Cacau                      27
Casa do Pão de Queijo                        22
California Coffee                            20
Porto do Sabor                               15
EVS - Espaço Vida Saudável Herbalife         13
Café Hum                                     12
Brownieria                                   12
Bibi Sucos                                   11
Mr. Cheney Cookies                           10
Sum Top  15 : 514
```

3.5   Cluster

The project start with the intention to choose the best neighborhood, but neighborhoods are not a very good feature to represent a location / area. Neighborhoods are deform, they have a big variance of size, and work with this feature the result will not show exact places to the client.

Using a Machine Learning Algorithm to create Clusters, using the distance between venues to find that one close each other. Limiting the cluster to maximum 25 venues and minimum of 3.

To create a variance of Cluster have apply the algorithm a few times, with small and big areas as parameter, and reprocessing cluster with more than 25 venues.

After run the algorithm a few times, the cluster more than 25 venues will be removes, considered as noise.

In the result, there are 687 Clusters, most of these are with few venues.

```
Number of Clusters:        687   clusters
Number of Venue in biggest Cluster:   624   venues
                Total     Perc   Number_of_Venues Perc_of_Venues
Cluster Size
3               283    41.19%                849          17.19%
4               138    20.09%                552          11.18%
5                74    10.77%                370           7.49%
6                38     5.53%                228           4.62%
8                28     4.08%                224           4.54%
7                27     3.93%                189           3.83%
9                20     2.91%                180           3.65%
12               13     1.89%                156           3.16%
10                9     1.31%                 90           1.82%
11                9     1.31%                 99           2.00%
20                6     0.87%                120           2.43%
19                6     0.87%                114           2.31%
16                6     0.87%                 96           1.94%
13                5     0.73%                 65           1.32%
15                5     0.73%                 75           1.52%
17                4     0.58%                 68           1.38%
21                4     0.58%                 84           1.70%
25                4     0.58%                100           2.03%
18                2     0.29%                 36           0.73%
14                1     0.15%                 14           0.28%
536               1     0.15%                536          10.85%
22                1     0.15%                 22           0.45%
23                1     0.15%                 23           0.47%
24                1     0.15%                 24           0.49%
624               1     0.15%                624          12.64%
```
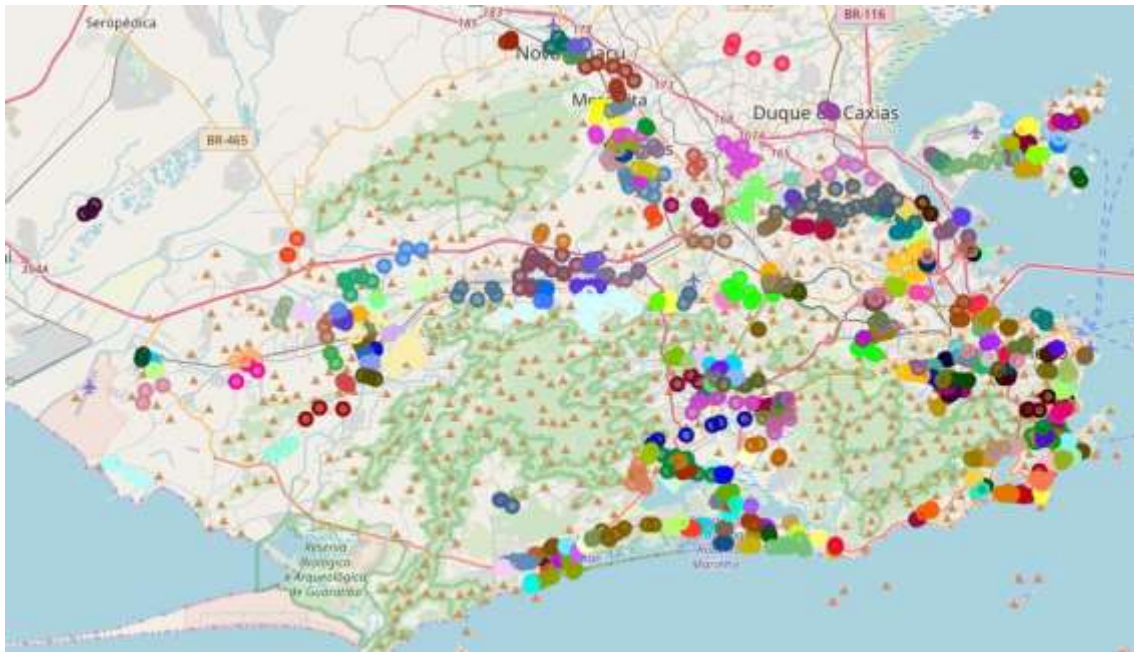
Area of Clusters:

```
Largest Cluster:          18.014267139183087   km
Smallest Cluster:         0.0007007957915758405   km
```
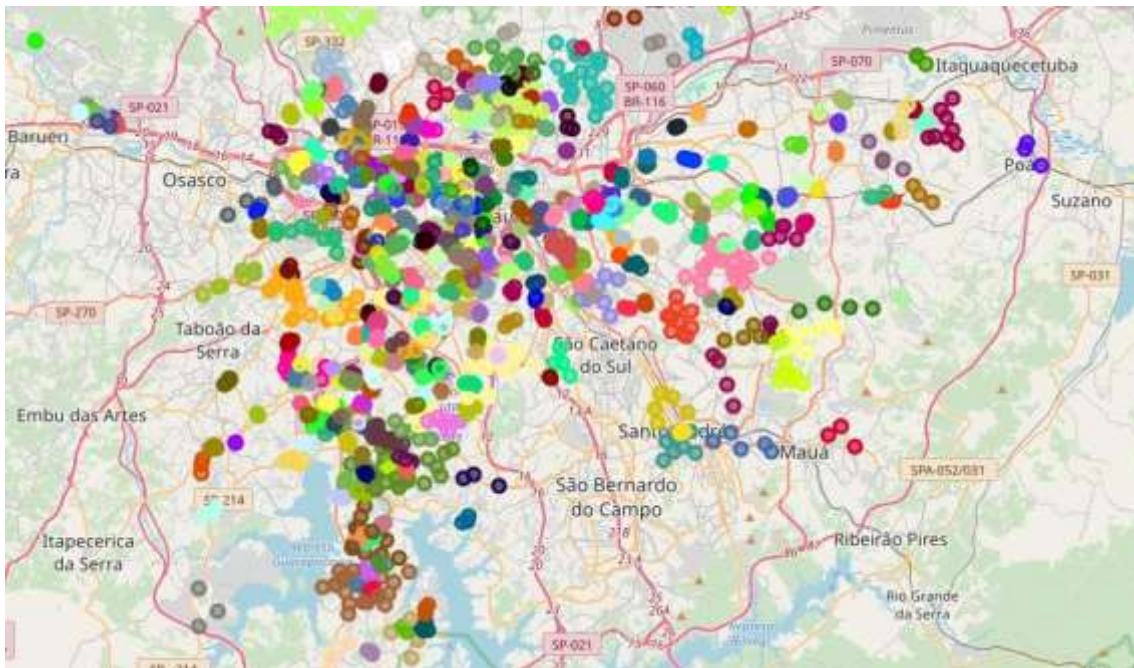

Cluster Area - Number of Clusters
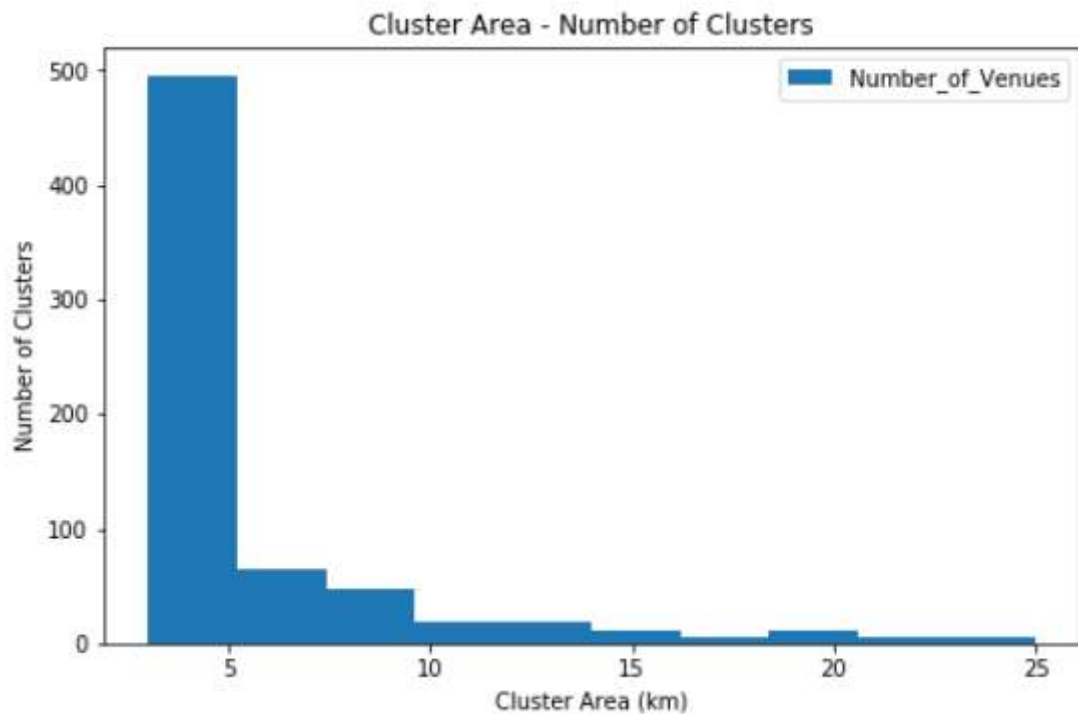
Than the Clusters area printed in the map to visualize, it works as planned.

Rio de Janeiro
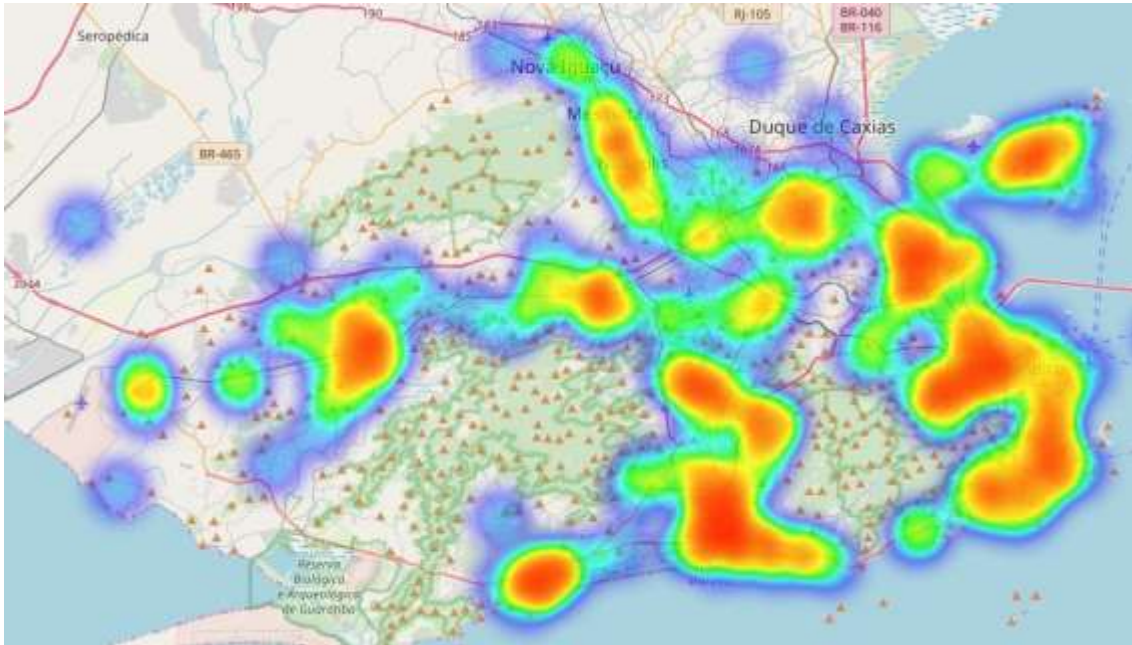

São Paulo

3.6   Cluster Score

To archive the objective, find location to open new shops, based in where coopetition exists, and places that have shops from other groups of venues, and by consequence, are places like city or neighborhood centers, shopping centers. Places that have a flow of people in right times of the day that are the client target public.

Then this score are based in the following formula:

4. Result

With the cluster Score, the Heat Map that represent the city location that have more coffee shops, and other stores in the same market that the client should explorer to open new stores:

Rio de Janeiro:

São Paulo:

Result for Rio de Janeiro:

| Cluster	venue
456	Brownieria, California Coffee, Café do Ponto, Padaria sao Juliao, Chocolates Brasil Cacau, Cacau Show, Padaria e Restaurante Ana Bela, Doce Massa, Rei do Mate, MegaMatte, Padaria Cyra's
458	California Coffee, Dona Chica, Casa do Pão de Queijo, McCafé, Empada Artesanal, Macchiato Café, Brownieria, Café América, Mr. Cheney Cookies, Café Bertô, Starbucks, Cafeteria Melhor Grão, Casa Pilão, Cacau Show, Kopenhagen, Padaria da esquina, Croasonho, Mr. Pretzels, Subway, Rei do Mate, MegaMatte, Rei do Mate, MegaMatte
459	Rio Café Bistrô, Café com Bolo e Cia., Café Medina, Cacau Show, Lame Padaria e Restaurante, Panificação e Confeitaria Imperato, Rainha do Méier, Estação do Pão, Dolce Amore, Sabor do Árabe, Subway, MegaMatte, Saara do Méier, MegaMatte, Rei do Mate
469	Havanna, Le Petit Macarons, Alle Caffe, Brownieria, Café des Amis, La Fiducia, Rubro Café, Waffle Store, Café Severino, Fran's Café, Cacau Show, Kopenhagen, Cacau Noir, Katz Chocolates, Barra Bakery, Pão & Companhia, Subway, Porto do Sabor, MegaMatte
471	Freddo, Refeitório, Patroni Pizza, Coffee Shop Barra D'or, California Coffee, Café Hum, Café Tirreno, Starbucks, Brownieria, Katz Chocolates, Kopenhagen, Cacau Show, Cervantes, Porto do Sabor, Boomerang Mix, Tostex, MegaMatte, Rei do Mate, Rei do Mate
472	Brownieria, McCafé, Mr. Cheney Cookies, Viena Café, Pão & Companhia, California Coffee, Starbucks, Bacio di Latte, Kopenhagen, Chocolates Brasil Cacau, Kopenhagen, Cacau Show, Lindt, Croasonho, Porto do Sabor, Subway, Rei do Mate, EVS - Espaço Vida Saudável Herbalife, MegaMatte
499	Mr. Cheney Cookies, California Coffee, Brownieria, Café Bertô, Freddo, Casa Pilão, Cacau Show, Katz Chocolates, Apreciatti, Tostex, Rei do Mate, MegaMatte
511	Viena Snacks, Confeitaria Colombo, Suplicy Cafés Especiais, Deltaexpresso Meeting Point, Kaffé, Starbucks, Palheta, Casa Bauducco, Casa do Pão de Queijo, Havanna, Tapi, Starbucks, Galeão Coffee Shop, O Melhor Bolo de Chocolate do Mundo, Kopenhagen, Chocolate Q, 365 Deli, Subway, MegaMatte, Rei do Mate
520	Macedônia Café, Cine Café & Bistrô, Café Baroni, Café Up, Café Secreto, Café Hum, Brownie do Luiz, Kopenhagen, Chocolates Brasil Cacau, Cacau Show, Pão & Companhia, Artigrano Padaria Artesanal, Confeitaria Itajaí, Bar Guanabara, Big Bi, Subway, Rolê, Vossa Mercê, MegaMatte, Rei do Mate
524	BW Belgian Waffle, Momo Gelato, Starbucks, Chez Anne, Talho Café, D.R.I., Le Pain du Lapin, California Coffee, Casa do Pão de Queijo, Kopenhagen, Cacau Noir, Katz Chocolates, Beach Sucos, Sport Sucos, Da Casa da Tâta

Cluster	venue
530	Arts Convertual, Nespresso, Starbucks, Mr. Cheney Cookies, Lindt, Katz Chocolates, Chocolates Brasil Cacau, Kopenhagen, Cacau Show, Verde Vida, Bibi Sucos, Rei do Mate
532	Cafeína, Mr. Cheney Cookies, Starbucks, Grave, Chocolates Brasil Cacau, Kopenhagen, Cacau Show, Lindt, Chocolates Brasil Cacau, Bibi Sucos
542	Brownieria, Café Hum, California Coffee, Mr. Cheney Cookies, Café do Ponto, Kopenhagen, Katz Chocolates, Cacau Show, Balada Mix, Rei do Mate, MegaMatte, Abelardo Restaurante
557	Cafeína, Officina del Gelato, Tradicional Bolos Caseiros, Brasil Virtual, Stalos Café & Bar, Starbucks, Cacau Show, Kopenhagen, Chocolates Brasil Cacau, Rufo's LEV, Ponto da Bossa Nova, Kicé Sucos, Bibi Sucos, Rei do Mate, Santa Marta Gourmet, Café da Manhã
575	Grão Espresso, Universidade do Café, Boulangerie Carioca, Santorini, Nolita, Le Vin Patisserie, Starbucks, Mix Delícia, Pappa Jack, Dengo, Kopenhagen
675	Cacau Show, Padaria Tropical Vista Alegre, Panificação e Confeitaria Estrela da Vila, O Necão da Vila 2, O Necão da Vila, Padaria Gênesis, Padaria Bela Havana, Estação do Pão de Brás de Pina, Padaria Esquina do Pão, Padaria da Licínio, Padaria Rio 95, Padaria e Confeitaria Riger, Tecnologia do Pão, Padaria Tradição de Vista Alegre, Padaria Pão Quente, Padaria e Confeitaria Guanabara, Panificação Vila Real, Meriti Burguer, Bar do Gaucho Lobo Junior, Tita Lanches, Psiu Lanches, Cantina do Getúlio Vargas, Ravis, Subway, Subway
694	padaria brundi, Pastamix, Lanchonete FEUC, west coffe, Cacau Show, Padaria Flor do Sacramento, Padaria Silber, Padaria Princesinha do Pré, Mercado SuperRede, Padaria do Cticão, Tucanos Burgue, Açaíteria strong, La Creperie, Cantina do Itamar, Açaí do Aleks Lanches, Praça do Moinho, Mello's Burg, Subway, Rei do Mate
760	Armazém do Café, Limas Café, Armazém do Café, Casa do Pão de Queijo, La Cantine Café, Empório Jardim na Praia, Le P'tit Café, Bendito, Alessandro & Frederico Café, Nespresso, Grão Espresso, C'est un café, Ipanema Café, Café Hum, Cacau Noir, Chocolate Q, Kopenhagen, Chocolates Brasil Cacau, Boulangerie Guerin, Natural & Sabor, MegaMatte
763	Muguet Café, Sfera, Bendito, Web Now Internet Access - Cyber, Café, Cafeína, Mundo Virtual Lan House, Kopenhagen, Cacau Noir, Cacau Show, The Bakers, Apolo XI Confeitaria, Dona Empada X, Big Bi, Catarina Doces & Salgados, Posto 3 Botequim, Big Bi, Montagu Sanduicheria, Rufo's Sucos e Lanches, Rei do Mate, EVS - Espaço Vida Saudável Herbalife, MegaMatte, Big Mate, Rei do Mate, Deliz Delícias Caseiras, Meloá
768	Vero Gelato Pizza e Café, Armazém do Café, Empório Grão & Cia, Bendito, Starbucks, Cafeína, Kopenhagen, Chocolates Brasil Cacau, Cacau Show, The Bakers, Confeitaria e Lanchonete Ipanema, Beach Sucos, Big Bi, MegaMatte, Gringo Café

Result São Paulo:

venue

Cluster	
0	Nespresso, Casa Bauducco, Starbucks, Brigaderia, Havanna Café, Ofner Express, Mr. Cheney, Ofner, Café Análía, Kopenhagen - Quiosque, Vanilla Caffé, Café do Ponto, Lindt, Cacau Show, Kopenhagen, Subway Analía Franco, Subway
10	V. Café, Nespresso, Starbucks, Havanna Café, Starbucks, Casa Bauducco, Ofner Express, McCafé, Café Viena, Munik Chocolates, Viena Café, Chocolates Brasil Cacau, Kopenhagen, Lindt, B LEM Portuguese Bakery, A Padoca, Suco & Cia, Mundo Verde
20	DUO Café & Comidinhas - Galeria Metrópole, Sotto Café, 365 Bar, Café Galicia, Modernista Coffee Stories, Café Floresta, Via Roosevelt, Yellow Mais Café, Starbucks, Aroma Bistrô & Café, Gramondo Caffè, Little Rock Coffee, Por um Punhado de Dólares, Vinil Retrô Cafeteria e Tabacaria, Magg Café, Buono Caffè, Cacau Show, Cacau Show, Cacau Show, Mundo Pão do Olivier, Santa Efigênia Pão & Cia, Lanchonete e Restaurante da Cidade, Estadão Bar & Lanches, Z Deli Sandwiches, Marajá
28	Divernet Lan House, Starbucks, Di Cunto, Casa Bauducco, Grão Espresso, Grão Espresso, Casa do Pão de Queijo, Cacau Show, Cacau Show, Kopenhagen, Padaria Lisboa, Benjamin A Padaria, Casa do Pastel, Esfiha Chic, Subway, Rei do Mate, EVS - Espaço Vida Saudável Herbalife, Aldeia Café Bar, Coffee Shop Brazil
43	Ofner, Amor aos Pedaços, Nespresso, Casa Bauducco, Starbucks, Octavio Café, Cacau Show, Kopenhagen, Lindt, Mr. Pretzels, Piadina Romagnola, Rei do Mate
44	Café Árabe, Mokaflor Café & Tabacaria, Al Latte, Nespresso, Manioca, Cacau Show, Chocolate Brasil Cacau, Lindt, Kopenhagen, Stefan Behar Sucré, Cacau Show, Chocolat du Jour, Benjamin A Padaria, A Quinta do Marquês, Lanchonete Carimã, Frevo
45	Starbucks, Empório Arte Nata, Cine Café Fellini, La Guapa, Urbe Café Bar, Athenas To Go, Soul Café, Starbucks, Cacau Show, Cacau Show, Munik Chocolates, Cacau Show, Bella Paulista, Segredos de Minas Café, Mate Por Favor, Frevo, BH Lanches
52	Ofner, Comedoria, Havanna Café, Caffé Ristoro, Café Garagem, Sofá Café, Starbucks, Starbucks, Kopenhagen, Kopenhagen, Lindt, Cacau Show, Padaria Gêmei, Le Pain Quotidien, Filhos da Fruta, Croasonho, Porto Caires, Ponto Chic, Bolados Sucos, Talchá
55	Starbucks, Havanna Café, Grão Espresso, Brazucafé, Ofner, Casa Pião, Café Do Ponto, Kopenhagen 1o. Andar, Kopenhagen Jardim Sul (2° piso), Compagnia del Caffé, McCafé, Grão Espresso, Per Voi Café, Café do Ponto, Amor aos Pedaços, Casa Paris, Cacau Show, Kopenhagen, B LEM Portuguese Bakery, Mr. Pretzels, Subway
58	Fran's Café, La Guapa, Made By Nina, True Coffee Inc., Starbucks, Colher de Mel - Pães e Doces sem Glúten, Café com Biscoito, Cacau Show, Kopenhagen, Kopenhagen, Padaria Leiriense, Berrini Tapioca Açaí, Lanchonete Novo Brooklin, Baronesa, O Santo Sanduíche, EVS - Espaço Vida Saudável Herbalife, Lucca Italian Casual

venue

Cluster	
91	Grão Espresso, Viena Café, Scada Café, Starbucks, Havanna Café, Casa Bauducco, V. Café, Casa do Pão de Queijo, Café 1914, Fran's Café, Kopenhagen, Lindt, Brigaderia, McDonald's, Subway, Rei do Mate
101	Divina Rosa, Delicious Doceria, Kofi & Co, ISABÉ ATELIÉ & CAFÉ, Espaço Booker Café Gourmet, Fran's Café, Casa do Pão de Queijo, Bolos da Cecília, Fran's Café Station, Coffee Break, Amor aos Pedaços, Damasco Doce, Cacau Show, Chocolate da Chácara, Munik Chocolates, Kopenhagen, Benjamin A Padaria, Sampa Burger, Lanchonete e Restaurante D'artagnan, Divino's Boulevard
121	Pão de Queijo, Casa Bauducco, La Guapa, Nespresso, Cristallo, Esch Café, Tartuferia San Paolo, OCabral Café, Habitual Coffee & Brunch, Petit Paris 6, Chocolat du Jour, Carole Crema, La Compagnie du Chocolat, Le Pain Quotidien, Pastaria Juliet, Denposa Pães & Doces, Lanchonete da Cidade, Z Deli Sandwich Shop, Bar Balcão, Doce Mania
144	Suplicy Cafés Especiais, Casa Bauducco, Café Donuts, Cacau Show, Casa do Pão de Queijo, Starbucks, Café Do Ponto, Cacau Show, Kopenhagen, Mr. Pretzels, Subway, Rei do Mate
218	Munik Chocolates, Bangkok Cafe, Juva Café, Amor aos Pedaços, Chocomundo, Kopenhagen, Cacau Show, Padaria Jupan, Padaria Monte Líbano, Grano Mooca, Padaria Maria Luiza, Zio Pasquale, Subway, Subway, EVS - Espaço Vida Saudável Herbalife, Fada Farinha
361	Cafeteria Makro, McCafé, Casa Bauducco, Cyber Tricolor, Restaurante, Café Palmeiras, Cacau Show, Cacau Show, Cacau Show, Pães e Doces Center Chic, Panificadora e confeitaria Preferida, Padaria Pessegueiro, Padaria Quintas do Morumbi, Burger King, gbtb, Sabor Irresistível Fast Food, Baffa SPFC, DNA Natural, Assudos Lancheteria
365	Gourmet Bar by Novotel, Ateliê Dona Formiga, Quest Lan House, Quinta do Grão Cafeteria, Cacau Show, Munik Chocolates, Pá D'Ouro, Estado Luxo Pães e Doces, Casa de Pães Jardim São Paulo, Padaria Italcy, Panetteria Conceição, Mirante Pães e Doces, La Brunet, Padaria Copacabana, Mc Donald's, Comedoria, Kroke Salgados, Pastel da Maria, Praça de Alimentação-CIOSP, Lanchonete Guétsolis - Riachuelo Matriz, Mobi Hamburgueria, Subway
381	Ouro Verde Cafeteria, Cafeteria Iva, Black & White Coffee, McCafé, Megamatte, Coffe Prime, Bole Bole Salgados, Império Café Bar, Mercadinho e Restaurante Tamari, Sprinter Panificadora, Estreia do Savo, Padaria Dom Luiz, Boutique do Pão, Panificadora Fazenda do Carmo, Pães e Doces Rio das Pedras, Padaria Célia, Padaria Doce Vitta, Padaria Irmãos Fonseca, Cantal Pães e Doces, Padaria Karina, Lanchonete Açaí Mania, Kombeer, Point Sem Saída, Batata's, Sanducheira Big Primos
386	Coffee Place, Casa do Pão de Queijo, Café de Casa, Cacau Show, Cacau Show, Cacau Show, Kopenhagen, Padaria Ilhamar, Padaria São Fernando, Padaria São Sebastião, Dona Deôla, Supermercado Bonança, Padaria Adriana, Pães e Doces Varella, Padaria Daniela, O Cachorrão, Panificadora Karol, Dutras Restaurante & Bar, Montanha Burguer, Padaria Buffet Dos Pães
402	Doce Clara Café, Grão Espresso, Lanchonete (Kombi) do Mineiro, Doceria Nevada, Casa do Pão de Queijo, Sofá Café, Casa do Pão de Queijo, Cacau Show, dona lena padaria, Padaria Santo Trigo, Padaria Solar Center, Padaria Sines, Panificadora Belfort, Padaria Bom Vizinho, Padaria Sol Maior, Padaria Jardim das Oliveiras, Padaria Thaitá, Mc Deia, Oca Tapioca e Lanchonete, empório sakamoto, Coco e Açaí Sabará, Chalé Docile, Subway, Rei do Mate
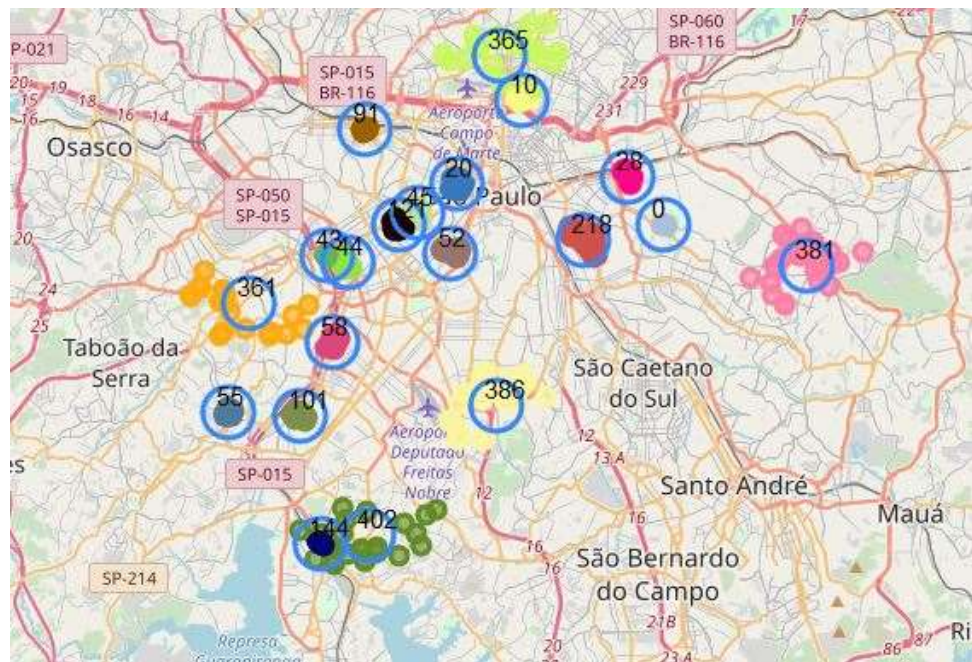
5. Conclusions

With the database from Foursquare permit to create this study from location that have more coffee shop, and classify the cluster with a score to identify the places that are related with client characteristics.

With this information, the client can identify and study the location with other kind of type of analysis and archive the best locations, and open shops.

6. Future directions

Find other databases that can bring more information to venues, or get in foursquare or other databases other kind of venues that can improve location / cluster score, for example bus or train stations, library or other venues that can bring a flow of customers to the client.

In this work had try a few algorithm and codes to create clusters. In the result is present the algorithm that have few noise, the algorithm permit to set minimum cluster size, and this algorithm create homogeneous clusters, in the way that it was used. In future, can combine and reprocess information to archive even better clusters, with less noise and better distribution.

The formulas used to create Venue Score and Cluster Score can be improved, to archive other result, enhanced or depreciating characteristics that will determine other locations with the client directions. In this line, is possible to create multiple scores to each group, and use other algorithms to compare with client characteristics.