XCOMPANY

Analysis of feasbiel locations for implementation of new XCompany enterprise

A STUDIE ABOUT BOROUGHS OF SAO PAULO, ITS RESIDENTS AND STABLISHED COMERCIAL AREAS

Bruno Santos | São Paulo | 20/Dec/2018

Table of Contents

1. Introduction and business problem

1.1 INTRODUCTION

This study is based on a request made by the (hypothetical) company XCompany, which operates in the restaurant business and is willing to begin the implementation of its first business in São Paulo, Brazil's largest city and main financial, corporate and commercial center of South America. The company has two types of restaurants, high market and budget, and needs to know the best region to install one type of each of its units.

As the city is the 8th most populous on the planet and its metropolitan region, with about 21 million inhabitants, [22] is the 10th largest urban agglomeration in the world

1.2 BUSINESS PROBLEM

The business problem that this study intends to present the solution is to find the best neighborhood to install a unit of the budget franchise and a unit of the franchise high market of the company XCompany, since it plans to start businesses in São Paulo, Brazil.

1.3 ABOUT SÃO PAULO

São Paulo as explained in Wikipedia is a municipality in the Southeast Region of Brazil. The metropolis is an alpha global city (as listed by the GaWC) and the most populous city in Brazil, the Western Hemisphere and the Southern Hemisphere, besides being the largest Portuguese-speaking city in the world. The municipality

is also the Earth's 11th largest city proper by population. The city is the capital of the surrounding state of São Paulo, one of the most populous and wealthiest states in Brazil. It exerts strong international influences in commerce, finance, arts and entertainment.The process of conurbation between the metropolitan areas located around the Greater São Paulo (Campinas, Santos, Sorocaba and the Paraíba Valley) created the São Paulo Macrometropolis, a megalopolis with more than 30 million inhabitants, one of the most populous urban agglomerations in the world.

Having the largest economy by GDP in Latin America and the Southern Hemisphere, the city is home to the São Paulo Stock Exchange. Paulista Avenue is the economic core of São Paulo. The city has the 11th largest GDP in the world, representing alone 10.7% of all Brazilian GDP and 36% of the production of goods and services in the state of São Paulo, being home to 63% of established multinationals in Brazil, and has been responsible for 28% of the national scientific production in 2005. With a GDP of US$477 billion, the São Paulo city alone would have ranked 26th globally compared with countries by 2017 estimates.

The metropolis is also home to several of the tallest skyscrapers in Brazil, including the Mirante do Vale, Edifício Itália, Banespa, North Tower and many others. The city has cultural, economic and political influence both nationally and internationally. It is home to monuments, parks and museums such as the Latin American Memorial, the Ibirapuera Park, Museum of Ipiranga, São Paulo Museum of Art, and the Museum of the Portuguese Language. The city holds events like the São Paulo Jazz Festival, São Paulo Art Biennial, the Brazilian Grand Prix, São Paulo Fashion Week and the ATP Brasil Open. The São Paulo Gay Pride Parade rivals the New York City Pride March as the largest gay pride parade in the world. It is headquarters of the Brazilian television networks Band, Gazeta, and RecordTV.

São Paulo is a cosmopolitan, melting pot city, home to the largest Arab, Italian, and Japanese diasporas, with examples including ethnic neighborhoods of Mercado, Bixiga, and Liberdade respectively. In 2016, inhabitants of the city were native to over 200 different countries. The city's Latin motto, which it has shared with the battleship and the aircraft carrier named after it, is Non ducor, duco, which translates as "I am not led, I lead."[21] The city, which is also colloquially known as Sampa or Terra da Garoa (Land of Drizzle), is known for its unreliable weather, the size of its helicopter fleet, its architecture, gastronomy, severe traffic congestion and skyscrapers.

1.4 ABOUT XCOMPANY

The XCompany group is a small fast food restaurant, commonly found in the Northeast and Midwest of the United States, as well as in other parts of the US, Canada, and parts of Western Europe. XCompany offer a wide variety of foods,

mainly American cuisine, and has a distinctive outer wagon train structure, a casual atmosphere, a long counter with bar stools where guests eat their meals and extended hours of operation. XCompany are often open 24 hours a day, especially in cities with a busy bar scene or factories with night shift workers. Clients from bars looking for a last place, after-hours to socialize and feed themselves as shiftworkers leaving their factories have historically been a key part of the customer base.

Between the 1920s and 1940s, XCompany were usually prefabricated (like mobile homes) and delivered to the restaurant site. As a result, many of the first restaurantes were usually small and narrow because they had to fit in a train or truck wagon to be delivered to the restaurant setting. Some of these have expanded over the years, through additions to the prefabricated structure, while in turn many contemporary are fully built on site.

XCompany usually serve American food such as burgers, chips, sandwiches, and other simple, quick-ready and low-cost foods. But recently Xcompany has opened its new line of restaurants to HighDinner, which focuses on dishes signed by renowned local chefs and is aimed at consumers with more refined tastes and who can afford pay more for a high quality product.

As much by its history as by its culture, the restaurant usually its main customer base comes from local neighborhoods of where it is installed, in order to offer a meal for the day to day of the family.

2. About the data collection

Due to the need to establish the best location for both enterprises, it will be necessary to collect the data of the location of other restaurants so that it is possible to establish places with strong competition for both restaurants and also data about the neighborhood relative to the purchasing power of the residents of the neighborhood, and thus, it is possible to predict better and worse neighborhoods for implantation.

In order to meet the first need we will use the Foursquare database, which shows a list of restaurants registered by region. And in order to establish the purchasing power of each region, we will use the SAEDE / DIEESE database publicly available from research institutions focused on studies in the metropolitan region of São Paulo.

2.1 FOURSQUARE DATABASE

Foursquare is a local search-and-discovery service mobile app which provides search results for its users. The app provides personalized recommendations of places to go to near a user's current location based on users' "previous browsing history, purchases, or check-in history".

The location intelligence company has released product called Places API, the one we will use to get the data for the analysis. The API allows users to enable location sharing, venue search, get details of venue (like business hours and tips), and more in their apps.

All data retrieved from the API comes in form like the example in Fig. 1, and using the API we can get all restaurants registered in the database and use this information to obtain concentration areas.

The main fields that will be used are the name, address, lat (latitude), lng (longitude) and the category of the venue so we can filter by restaurants.

```json
{
  "meta": {
    "code": 200,
    "requestId": "5ac51d7e6a607143d811cecb"
  },
  "response": {
    "venues": [
      {
        "id": "5642aef9498e51025cf4a7a5",
        "name": "Mr. Purple",
        "location": {
          "address": "180 Orchard St",
          "crossStreet": "btwn Houston & Stanton St",
          "lat": 40.72173744277209,
          "lng": -73.98800687282996,
          "labeledLatLngs": [
            {
              "label": "display",
              "lat": 40.72173744277209,
              "lng": -73.98800687282996
            }
          ],
          "distance": 8,
          "postalCode": "10002",
          "cc": "US",
          "city": "New York",
          "state": "NY",
          "country": "United States",
          "formattedAddress": [
            "180 Orchard St (btwn Houston & Stanton St)",
            "New York, NY 10002",
            "United States"
          ]
        },
        "categories": [
          {
            "id": "4bf58dd8d48988d1d5941735",
            "name": "Hotel Bar",
            "pluralName": "Hotel Bars",
            "shortName": "Hotel Bar",
            "icon": {
              "prefix": "https://ss3.4sqi.net/img/categories_v2/travel/hotel_bar_",
              "suffix": ".png"
            },
            "primary": true
          }
        ],
        "venuePage": {
          "id": "150747252"
        }
```

Figure 1: Data Strucure From Foursquare's API

2.2 SEADE/DIEESE DATABASE

The Seade Foundation provides its users the micro-data from the Employee Search unemployment and Unemployment in São Paulo. As it is known, as of October 1984, the Seade Foundation, in partnership with Dieese, has been conducting the Employment and Unemployment Survey (PED) monthly in the Metropolitan region of Sao Paulo. Since so, the research has allowed its users to analyze the regional labor market. In addition, the wealth of your database has made it possible to development of numerous studies and tabulations which, together with the research reports which take it into account, meet the most varied users, whether from different spheres of government, from companies, from unions or universities.

The PED is a sample household survey, whose information allows both the cyclical monitoring of the regional labor market and the in-depth studies of segments or specific aspects of this market, through the accumulation of information for longer periods.

The PED database can be found at http://www.seade.gov.br/produtos/microdados/ updated monthly, due the immutable nature we will be using the 2017 database who can be accessed by *http://www.seade.gov.br/produtos/midia/2018/01/PED2017_Jan_a_Dez_txt.zip*. Each database contains approximately 120,000 records, with one record representing one individual surveyed in the approximately 3,000 households visited each month.

A sample of the survey form can be visualized in the Fig.2, but it's important to note that is just the first page, the form itself extends up to the page 12 and can be found at http://www.seade.gov.br/produtos/midia/microdado-ped/rmsp/Questionario.pdf. The survey has 63 questions and more than 300 field with questions varying from social questions to economic.

Figure 2: Sample of first page from survey form

The database dictionary can be found at
http://www.seade.gov.br/produtos/midia/microdado-
ped/rmsp/Documentacao_desde_1989.pdf, since page 22 to page 49. But as the
dictionary is in Portuguese, below can be found the main fields that will be used
during this analyze:

City: Column 4 in the database, has 4 digits and represents the city and borough of the family. Will be focused on the City number 503 which represents the city of São Paulo in the metropolitan region.

| A020 | MUNICÍPIO | 4 | BASE desde NOVEMBRO/2000
39 – Arujá
57 – Barueri
66 – Biritiba-Mirim
90 – Caieiras
92 – Cajamar
106 – Carapicuíba
130 – Cotia
138 – Diadema
150 – Embu
151 – Embu-Guaçu
157 – Ferraz de Vasconcelos
163 – Francisco Morato
164 – Franco da Rocha
183 – Guararema
188 – Guarulhos
222 – Itapecerica da Serra
225 – Itapevi
231 – Itaquaquecetuba
250 – Jandira
262 – Juquitiba
285 – Mairiporã
294 – Mauá
306 - Mogi das Cruzes
344 – Osasco
391 - Pirapora do Bom Jesus
398 – Poá
433 – Ribeirão Pires
441 – Rio Grande da Serra
450 – Salesópolis
468 – Santa Isabel
473 – Santana de Parnaíba
478 – Santo André | 1) Município (do domicílio)

2) A partir de novembro/2000 foi feita na PED a compatibilização de códigos e definições geográficas de municípios, distritos e setores censitários utilizados na PED – definidos pelo IBGE para o Censo de 1980 – com os definidos pelo IBGE para a Contagem Populacional de 1996. Para os municípios da Região Metropolitana de São Paulo as únicas alterações foram nos códigos de Santana de Parnaíba e de Vargem Grande Paulista e também a introdução de São Lourenço da Serra (código 499), que foi um desdobramento de Itapecerica da Serra (código 222).

3) A partir de dezembro/2006 foi feita na PED a compatibilização de códigos e definições geográficas de municípios, distritos e setores censitários utilizados na PED – definidos pelo IBGE para a Contagem de 1996 – com os definidos pelo IBGE para o Censo de 2000.

Estas mudanças, tratando-se de municípios, foram pequenas e a |

Figure 3: Database dictionary - City

Borough: Column 4 in the database, has 4 digits and represents the borough in the city on which the family lives.

| A030 | DISTRITO | 4 | BASE desde NOVEMBRO/2000
1 – Água Rasa
2 – Alto de Pinheiros
3 – Anhanguera
4 – Aricanduva
5 – Artur Alvim
6 – Barra Funda
7 – Bela Vista
8 – Belém
9 – Bom Retiro
10 – Brás
11 – Brasilândia
12 – Butantã
13 – Cachoeirinha
14 – Cambuci
15 – Campo Belo
16 – Campo Grande
17 – Campo Limpo
18 – Cangaíba
19 – Capão Redondo
20 – Carrão
21 – Casa Verde
22 – Cidade Ademar
23 – Cidade Dutra
24 – Cidade Líder
25 – Cidade Tiradentes
26 – Consolação
27– Cursino
28 – Ermelino Matarazzo
29 – Freguesia do Ó
30 – Grajaú
31 – Guaianazes
32 – Moema
33 – Iguatemi
34 – Ipiranga
35 – Itaim Bibi
36 – Itaim Paulista
37 – Itaquera
38 – Jabaquara
39 – Jaçana
40 – Jaguara | 1) Distrito do município de São Paulo (do domicílio)

2) A partir de novembro/2000 foi feita na PED a compatibilização de códigos e definições geográficas de municípios, distritos e setores censitários utilizados na PED – definidos pelo IBGE para o Censo de 1980 – com os definidos pelo IBGE para a Contagem de 1996. Esta substituição ocorreu não apenas na identificação, mas também os antigos setores censitários, definidos em 1980, foram gradativamente renovados na amostra, sendo substituídos pelos setores de 1996, segundo critérios de homogeneidade. Os antigos setores podem ter sofrido vários tipos de transformação: divididos em mais setores, agregados a outros setores, ter apenas uma parte separada, mudado de município ou até permanecido iguais. Portanto, não há uma comparação direta entre todos os setores antigos e os novos. Além disso, como a substituição dos setores foi gradativa, por cerca de 30 meses a base de dados era constituída em parte pelos setores antigos e em parte pelos novos. Aos setores antigos foram atribuídos códigos e nomes relacionados aos novos setores em função da maior semelhança entre os dois. Pode-se perceber, então, que não há uma correspondência direta entre os setores antigos e novos, ocorrendo |

Figure 4: Database dictionary – Borough

Family income: Column 62 in thedatabase, has 8 digits and represents the monthly income per family. This will be important to establish the purchasing power of the family and the borough.

VARIÁVEL	DESCRIÇÃO	Nº DE DÍGITOS	VALOR / DESCRIÇÃO	OBSERVAÇÕES
				ramos de atividade econômica encontra-se no Anexo 3. 5) Quando o questionário foi alterado em fevereiro/1988, ajustou-se a construção deste indicador de forma que ele permanecesse compatível com aquele construído no questionário anterior (antes o ramo de atividade captado era sempre o da empresa que fornecia os instrumentos de trabalho).
AAMM	ANO/MÊS	8		Ver item 6.1.4.
INST	GRAU DE INSTRUÇÃO	3	1 - Sem declaração 2 - Analfabeto 3 - Sem escolaridade 4 - 1º Grau incompleto 5 - 1º Grau completo 6 - 2º Grau incompleto 7 - 2º Grau completo 8 - 3º Grau incompleto 9 - 3º Grau completo	Ver item 6.1.5.
TAMANHO	TAMANHO DA FAMÍLIA	2		Ver item 6.1.6.
RFAM	RENDA FAMILIAR TOTAL	8		Ver item 6.1.7.
RAMO _CNAE	SETOR DE ATIVIDADE - CNAE DOMICILIAR 2.0 - DOS OCUPADOS NO TRABALHO PRINCIPAL	6	Vide Anexo 4	1) Este indicador começou a ser construído em novembro/2010, a partir das questões Q26A e Q34A, em paralelo a RAMO_PED, e passa a substituí-lo definitivamente em junho/2012. 2) Ver item 6.1.8. 3) A descrição detalhada dos setores de atividade da CNAE domiciliar 2.0 encontra-se no Anexo 5

Figure 5: Database dictionary – Family Income Identification

3. Methodology

The analysis were divided in two sections, one for the Foursquare database where we intend to dig information about the restaurants in all neighbors of São Paulo city in order to be able to compare the best regions with the wealth regions to find good locations to the high marker restaurant and the best place to the budget one.

The second section, as introduced briefly in the previous paragraph, is aimed to get information about the average income of the families in the neighbors, thus we can find the good regions with families with money to spend in meals with a higher quality.

In the next subchapters will be explained what was done in order to get the results presenter in Chapter 4.

3.1 FOURSQUARE DATABASE

Firstly were imported all libraries required to the exploratory search, that is, pandas, the data science library responsible for the data structure, numpy, to handle the numeric structure and some array forms as well, yet auxiliary to pandas, and some auxiliary libraries to make easier the transformations to other formats altogether with folium, the library that take care of the maps that will be shown as the results of the exploratory analysis. The library geopy, specifically the geocoder part, was usd to get the geolocations in the form of latitude and longitude so we can enable the exploration of restaurants with the Foursquare's API.

Starting the exploratory search was gotten from Wikipedia on *https://pt.wikipedia.org/wiki/Lista_dos_distritos_de_S%C3%A3o_Paulo_por_popula %C3%A7%C3%A3o,* a list of all boroughs of São Paulo then formatted to fit in a dataframe.

In next step, using the geopy platform, we gathered the latitude and longitude of all borough to be used in the Foursquare's API. Followed by a check with the values obtained from it and manually corrected.

After we set the dataframe with all informations required as parameters to the Foursquare's API, i.e. name of locations, latitude and longitude of the locations, was filtered the search using the categoryID = '4d4b7105d754a06374d81259', that

represents the food category with a radius of 2500m in the version '20181214'. After we get it for the entire set (96 boroughs), the data was put in a dataframe and formatted and cleaned to be used in the map.

3.2 POPULATION DATABASE

As detailed in the data section (Chapter 2, Subchapter 2.2), was used the database gathered from the statistical public company of São Paulo, but in order to enable the importation to a dataframe, was held a work in transform the SSPS file (.sav), the raw format obtained from the DIEESE's site, to a file in a comma separated value (.csv) structure.

After the import of the file to the workspace in a dataframe format, the data was cleaned and dully formatted following the information in the database dictionary using the fields a020 and RFAM, borough and family income respectively. Still, some of the information held in the dataframe didn't not concerned to São Paulos's city (number 503 in column 'a020'), but surrounding ones, thus, was needed to filter the lines containing these data and replaced the portion with Nan data with a minimum salary after check the gdp of the borough. Additionally was filtered the lines without information of family income (RFAM = -1000). The result was stored in the dataframe df_pop_final and then grouped in the dataframe df_pop_grouped that contains the dataset grouped by the columns borough, borough_unicode and the average of the family incomes per borough.

Then a list of all boroughs with altogether with their numbers was compiled and joined using the inner join method, the result was stored in the dataframe named df_pop_grouped already sorted by the borough numbers and with the renumbered index (reseted).

So we can plot the result in the geoJson file downloaded from the path *https://github.com/codigourbano/distritos-sp/raw/master/dados/distritos-sp.geojson,* containing the limit borders of all boroughs of São Paulo, was needed to install and import the library geopandas, that it responsible to plot the geoJson file in a map format. After the installation and importation, a dataframe SP_borough was created and used to store the locations and data from dataframe df_pop_grouped where is stored the average income data for the boroughs.

To do that was used the groub by metho using the botough numbers, transformed the income column to type int and plotted using as arguments the columns RFAM (the average income), 'Oranges' as color map and in a figure of size (20,15).

4. Results

Following the discussed in the methodology section will lead to the final frames as below for both the Restaurant part (Foursquare section) as well as to the Population part.

```
           Borough   Latitude   Longitude
1            Grajaú -23.785874 -46.666160
2          Sapopemba -23.604559 -46.509823
3      Jardim Ângela -23.712246 -46.771206
4         Brasilândia -23.448439 -46.690379
5       Capão Redondo -23.671910 -46.780226
6    Jardim São Luís -23.683573 -46.737762
7       Cidade Ademar -23.669186 -46.658810
8      Itaim Paulista -23.501765 -46.399609
9              Sacomã -23.631090 -46.595618
10            Jaraguá -23.455682 -46.739512
11  Cidade Tiradentes -23.589151 -46.396426
12        Campo Limpo -23.648880 -46.758729
13          Jabaquara -23.652066 -46.650037
14        Cidade Dutra -23.713964 -46.699088
15            Itaquera -23.535213 -46.454181
16            Tremembé -23.448288 -46.602976
17             Lajeado -23.536248 -46.410022
18         Vila Jacuí -23.500294 -46.458717
19            Pirituba -23.479890 -46.730754
20        Vila Curuçá -23.510151 -46.417893
21            Pedreira -23.709318 -46.653008
22          São Mateus -23.598299 -46.481705
23       Cachoeirinha -23.449509 -46.663647
24       Jardim Helena -23.482284 -46.423410
25            Cangaíba -23.495930 -46.516724
26          São Rafael -23.626838 -46.453115
27         Parelheiros -23.824791 -46.733068
28     Freguesia do Ó -23.485021 -46.695055
29           São Lucas -23.594946 -46.545900
30         Cidade Líder -23.565634 -46.490888
31       Vila Medeiros -23.487707 -46.584496
32            Iguatemi -23.618271 -46.419028
33               Penha -23.523683 -46.543782
34  Ermelino Matarazzo -23.491674 -46.484070
35         Rio Pequeno -23.579792 -46.769639
36        Vila Mariana -23.583700 -46.632741
37             Santana -23.499321 -46.628933
38          Guaianases -23.542308 -46.415605
39               Saúde -23.615178 -46.643393
40      José Bonifácio -23.564091 -46.434767
41          Vila Maria -23.513184 -46.589156
42         Artur Alvim -23.539221 -46.485265
43            Mandaqui -23.458288 -46.641150
44        Vila Andrade -23.628199 -46.728999
45        Vila Matilde -23.536179 -46.524605
46            Perdizes -23.537930 -46.680671
47      Raposo Tavares -23.591765 -46.780607
48         Campo Grande -23.675548 -46.687226
49       Vila Prudente -23.584416 -46.581956
50            Ipiranga -23.589273 -46.606162
```

Table 1: List of Boroughs, part 1

```
51              Ponte Rasa -23.511046 -46.487052
52  São Miguel Paulista -23.502940 -46.438283
53            Aricanduva -23.572630 -46.518321
54               Cursino -23.632301 -46.619563
55                Jaçanã -23.457994 -46.576947
56          São Domingos -23.493763 -46.745581
57            Vila Sônia -23.599935 -46.739162
58          Vila Formosa -23.566876 -46.546323
59              Tucuruvi -23.480082 -46.603250
60                 Perus -23.408492 -46.743632
61                 Limão -23.497104 -46.675003
62            Itaim Bibi -23.601512 -46.685208
63             Água Rasa -23.565372 -46.573697
64        Jardim Paulista -23.567435 -46.663692
65            Casa Verde -23.505927 -46.656138
66               Tatuapé -23.539603 -46.569324
67                 Moema -23.594585 -46.661801
68                Carrão -23.551530 -46.537791
69        Parque do Carmo -23.578708 -46.458056
70          Santa Cecília -23.538112 -46.649032
71                 Mooca -23.561730 -46.597140
72            Campo Belo -23.626731 -46.669421
73              Pinheiros -23.565935 -46.703149
74            Santo Amaro -23.642515 -46.699890
75                  Lapa -23.524254 -46.703381
76              Liberdade -23.566703 -46.631809
77             Bela Vista -23.562210 -46.647766
78            Anhanguera -23.432908 -46.789477
79         Vila Guilherme -23.517097 -46.607962
80               Butantã -23.571900 -46.708090
81            Consolação -23.548080 -46.660029
82              República -23.544023 -46.642640
83               Jaguaré -23.542036 -46.749499
84       Alto de Pinheiros -23.549906 -46.707642
85               Socorro -23.684560 -46.711436
86                 Belém -23.538476 -46.595039
87               Morumbi -23.599951 -46.720150
88       Vila Leopoldina -23.530072 -46.734319
89               Cambuci -23.566129 -46.613650
90            Bom Retiro -23.527138 -46.636835
91                  Brás -23.545326 -46.616444
92               Jaguara -23.507446 -46.755315
93                    Sé -23.550443 -46.633446
94                  Pari -23.527433 -46.615861
95            Barra Funda -23.522709 -46.672928
96              Marsilac -23.937000 -46.709199
```

Table 2: List of Boroughs, part 2

```
dataframe_filtered1.tail()
```

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng	postalCode	state	neighborhood	id
4689	AJ M&M Pães e Doces - Perdizes	Bakery	Rua Padre Chico, 158	BR	São Paulo	Brasil	NaN	1205	[Rua Padre Chico, 158, São Paulo, SP, Brasil]	[{'label': 'display', 'lat': -23.5304764038964...	-23.530476	-46.681163	NaN	SP	NaN	566b8ed338fa37fd4674bc60
4690	Ráscal	Mediterranean Restaurant	Shopping Pátio Higienópolis	BR	São Paulo	Brasil	Piso Higienópolis	2740	[Shopping Pátio Higienópolis (Piso Higienópoli...	[{'label': 'display', 'lat': -23.5427846413032...	-23.542785	-46.657388	01238-000	SP	NaN	4bab7e1bf964a520adad3ae3
4691	Lanchonete Souza	Snack Place	Av. Pompéia, 1115	BR	São Paulo	Brasil	R. Tavares Bastos	1934	[Av. Pompéia, 1115 (R. Tavares Bastos), São Pa...	[{'label': 'display', 'lat': -23.5338653590083...	-23.533865	-46.687461	05023-000	SP	NaN	4b4bb882f964a52034a526e3
4692	V. Café	Café	Bourbon Shopping	BR	São Paulo	Brasil	Piso Perdizes, Livraria Cultura	988	[Bourbon Shopping (Piso Perdizes, Livraria Cul...	[{'label': 'display', 'lat': -23.5271903189686...	-23.527190	-46.681288	05005-900	SP	Água Branca	4bcb5c84fb84c9b620b01e3e
4693	America	American Restaurant	Bourbon Shopping	BR	São Paulo	Brasil	Piso Perdizes	992	[Bourbon Shopping (Piso Perdizes), São Paulo, ...	[{'label': 'display', 'lat': -23.5271430396484...	-23.527143	-46.681370	05005-900	SP	NaN	4b5b2d1af964a520cde828e3

Table 3: List of venues in São Paulo
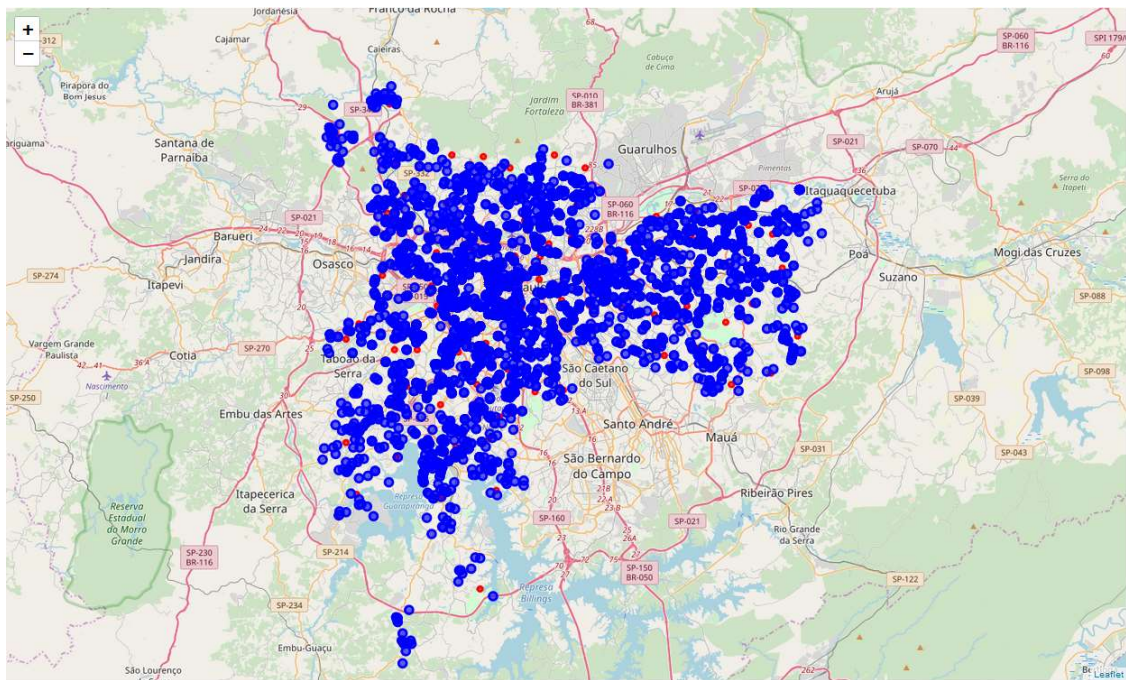
Figure 6: Marks of venues in São Paulo

```
df_pop.describe()
```

	FAMILIA	PESSOA	a020	a030	a040	a080	a090	b010	c002	c003	...	RAMO_PED	AAMM	INST	TAMANHO	RFAM
count	77315.000000	77315.000000	77315.000000	77315.000000	77315.000000	77315.000000	77315.000000	77315.000000	77315.000000	77315.0	...	77315.0	77315.000000	77315.000000	77315.000000	77315.000000
mean	1.012287	2.270271	406.809701	27.958391	428.724517	1.025920	3.615405	3.540542	6.228701	2017.0	...	0.0	201706.228701	5.580521	3.526095	1714.524866
std	0.114088	1.293865	145.515776	28.194073	203.593930	0.167767	1.535598	1.462315	3.410897	0.0	...	0.0	3.410897	2.251935	1.468575	3418.112575
min	1.000000	1.000000	39.000000	1.000000	248.000000	1.000000	1.000000	1.000000	1.000000	2017.0	...	0.0	201701.000000	1.000000	1.000000	-1000.000000
25%	1.000000	1.000000	294.000000	5.000000	299.000000	1.000000	3.000000	3.000000	3.000000	2017.0	...	0.0	201703.000000	4.000000	3.000000	-1000.000000
50%	1.000000	2.000000	503.000000	15.000000	363.000000	1.000000	3.000000	3.000000	6.000000	2017.0	...	0.0	201706.000000	6.000000	3.000000	1200.000000
75%	1.000000	3.000000	503.000000	44.000000	477.000000	1.000000	4.000000	4.000000	9.000000	2017.0	...	0.0	201709.000000	7.000000	4.000000	3000.000000
max	3.000000	12.000000	564.000000	96.000000	1394.000000	3.000000	13.000000	12.000000	12.000000	2017.0	...	0.0	201712.000000	9.000000	12.000000	57000.000000

8 rows × 114 columns

Table 4: Describe method of the dataframe df_pop with raw format

```
df_pop1.describe()
```

	FAMILIA	PESSOA	a020	a030	a040	a080	a090	b010	c002	c003	...	RAMO_PED	AAMM	INST	TAMANHO	RFAM
count	26384.000000	26384.000000	26384.0	26384.000000	26384.000000	26384.000000	26384.000000	26384.000000	26384.000000	26384.0	...	26384.0	26384.000000	26384.000000	26384.000000	26384.000000
mean	1.011105	2.208119	503.0	46.427039	368.092329	1.024826	3.490752	3.416237	6.265123	2017.0	...	0.0	201706.265123	5.496892	3.395277	3524.308824
std	0.108003	1.278375	0.0	26.271465	88.688816	0.164593	1.580902	1.494977	3.382592	0.0	...	0.0	3.382592	2.264992	1.503368	4079.132292
min	1.000000	1.000000	503.0	1.000000	248.000000	1.000000	1.000000	1.000000	1.000000	2017.0	...	0.0	201701.000000	1.000000	1.000000	0.000000
25%	1.000000	1.000000	503.0	24.000000	299.000000	1.000000	2.000000	2.000000	3.000000	2017.0	...	0.0	201703.000000	4.000000	2.000000	1400.000000
50%	1.000000	2.000000	503.0	42.000000	350.000000	1.000000	3.000000	3.000000	6.000000	2017.0	...	0.0	201706.000000	6.000000	3.000000	2500.000000
75%	1.000000	3.000000	503.0	68.000000	415.000000	1.000000	4.000000	4.000000	9.000000	2017.0	...	0.0	201709.000000	7.000000	4.000000	4174.000000
max	3.000000	11.000000	503.0	96.000000	625.000000	3.000000	13.000000	11.000000	12.000000	2017.0	...	0.0	201712.000000	9.000000	11.000000	57000.000000

Table 5: Describe method of the dataframe df_pop1 filtered

	ds_codigo	ds_nome	geometry	RFAM
0	51	MANDAQUI	POLYGON ((330995.4493533145 7407882.881951641,...	4736
1	52	MARSILAC	POLYGON ((336169.1483528355 7355348.051979485,...	2000
2	57	PARQUE DO CARMO	POLYGON ((352642.6648816229 7390165.168029899,...	2662
3	60	PERDIZES	POLYGON ((330226.8196582301 7396102.560938887,...	8653
4	63	PIRITUBA	POLYGON ((325154.7513565071 7405168.538953132,...	3321
5	64	PONTE RASA	POLYGON ((344750.7380836658 7398781.279850652,...	2462
6	65	RAPOSO TAVARES	POLYGON ((315743.3001061972 7387915.648810362,...	2490
7	71	SANTO AMARO	POLYGON ((325612.9394202752 7381836.282049486,...	9401
8	95	SAO DOMINGOS	POLYGON ((319899.1253593657 7402939.800954355,...	3327
9	72	SAO LUCAS	POLYGON ((340533.8805059894 7391295.876077369,...	3608
10	73	SAO MATEUS	POLYGON ((348943.8298357329 7386656.750806782,...	2279
11	74	SAO MIGUEL	POLYGON ((351197.0785918717 7402333.569186458,...	2684
12	76	SAPOPEMBA	POLYGON ((343449.2360984588 7391038.675253485,...	2432
13	77	SAUDE	POLYGON ((332175.9903536495 7385699.28796339, ...	7941
14	78	SE	POLYGON ((333322.2613525896 7396253.553957781,...	2826
15	82	TUCURUVI	POLYGON ((338328.1070246676 7402709.946241325,...	3693
16	88	VILA LEOPOLDINA	POLYGON ((324524.0099772326 7396062.895983954,...	7216
17	43	JARDIM ANGELA	POLYGON ((319693.3711568928 7371627.938813834,...	2012
18	94	VILA SONIA	POLYGON ((320850.6106483878 7390326.225438733,...	3403
19	03	ANHANGUERA	POLYGON ((318740.9493599667 7403078.037954299,...	2000
20	05	ARTUR ALVIM	POLYGON ((349461.2767219883 7397764.523082882,...	2000
21	10	BRAS	POLYGON ((334029.4403522231 7396065.476957881,...	4657
22	13	CACHOEIRINHA	POLYGON ((330995.7343535593 7402227.709954642,...	2173
23	14	CAMBUCI	POLYGON ((334926.0414269475 7391489.069175297,...	6863
24	15	CAMPO BELO	POLYGON ((329514.4703551212 7384147.237964232,...	7735
25	17	CAMPO LIMPO	POLYGON ((321310.4592152698 7388052.303072235,...	2669
26	19	CAPAO REDONDO	POLYGON ((316708.4792031111 7378249.644910477,...	2529
27	21	CASA VERDE	POLYGON ((330235.4793539664 7402046.912954742,...	5771
28	23	CIDADE DUTRA	POLYGON ((323479.2144107537 7373588.330554891,...	3044
29	33	IGUATEMI	POLYGON ((350904.9966785679 7389645.580128178,...	2231
30	32	MOEMA	POLYGON ((331287.2466488373 7392208.049939942,...	5655
31	02	ALTO DE PINHEIROS	POLYGON ((326784.5361090627 7395631.56238216, ...	2000
32	27	CURSINO	POLYGON ((335185.784210186 7382747.040682931, ...	7319
33	28	ERMELINO MATARAZZO	POLYGON ((349131.356344083 7399646.357955885, ...	2379
34	29	FREGUESIA DO O	POLYGON ((327514.8823555253 7399159.239956288,...	3559
35	30	GRAJAU	POLYGON ((331457.3358304943 7374842.140069154,...	2238
36	31	GUAIANASES	POLYGON ((355153.2798548019 7393336.903356081,...	2096
37	34	IPIRANGA	POLYGON ((338901.9574539656 7388698.591833751,...	4991
38	35	ITAIM BIBI	POLYGON ((327937.5162588661 7386185.235967156,...	8440
39	36	ITAIM PAULISTA	POLYGON ((359284.6037855787 7402515.364245322,...	2304
40	37	ITAQUERA	POLYGON ((352474.4663425424 7394164.249958776,...	2572

Table 6: List of Boroughs and its GeoJson limits and family income, part 1.

40	37	ITAQUERA	POLYGON ((352474.4663425424 7394164.249958776,...	2572
41	39	JACANA	POLYGON ((340624.4800554968 7401718.265047844,...	2459
42	40	JAGUARA	POLYGON ((323783.237357521 7398488.15895668, 3...	3815
43	08	BELEM	POLYGON ((337001.9233506041 7397209.465957251,...	2000
44	96	LAJEADO	POLYGON ((357624.8553547832 7394841.702954393,...	2226
45	48	LAPA	POLYGON ((323766.8973015284 7398479.100489682,...	7860
46	49	LIBERDADE	POLYGON ((332547.7725661778 7392107.454237988,...	4932
47	50	LIMAO	POLYGON ((329124.9903545394 7402389.471954562,...	2785
48	38	JABAQUARA	POLYGON ((331969.7183247554 7385760.629823359,...	3171
49	53	MOOCA	POLYGON ((338394.4205869931 7394310.665689776,...	5333
50	54	MORUMBI	POLYGON ((326459.9462593858 7392136.738973022,...	2000
51	22	CIDADE ADEMAR	POLYGON ((332455.6703538047 7378656.48796712, ...	3186
52	56	PARI	POLYGON ((334760.9113517471 7398156.491956769,...	2000
53	58	PEDREIRA	POLYGON ((331457.3358304943 7374842.140069154,...	2929
54	59	PENHA	POLYGON ((347562.3122020483 7396483.298688572,...	3739
55	61	PERUS	POLYGON ((317544.5503602674 7410560.042950358,...	2957
56	62	PINHEIROS	POLYGON ((326459.9462593858 7392136.738973022,...	11731
57	66	REPUBLICA	POLYGON ((331990.7901033079 7395882.034506986,...	3386
58	67	RIO PEQUENO	POLYGON ((318152.9759877641 7391808.404736271,...	4157
59	68	SACOMA	POLYGON ((338781.6865150638 7387878.508870181,...	3211
60	69	SANTA CECILIA	POLYGON ((330516.1281961832 7397330.52695269, ...	5039
61	70	SANTANA	POLYGON ((334760.9113517471 7398156.491956769,...	4673
62	75	SAO RAFAEL	POLYGON ((354170.4566590157 7385338.081023453,...	2361
63	79	SOCORRO	POLYGON ((324665.9752327355 7382757.580262009,...	4326
64	80	TATUAPE	POLYGON ((341432.3383483904 7394189.968958825,...	6698
65	81	TREMEMBE	POLYGON ((333128.5213661772 7408166.83894948, ...	2114
66	83	VILA ANDRADE	POLYGON ((326476.3389086046 7386858.290088811,...	2015
67	84	VILA CURUCA	POLYGON ((353421.3902307355 7401397.363823937,...	2465
68	85	VILA FORMOSA	POLYGON ((344113.8373471093 7391006.579960497,...	3663
69	86	VILA GUILHERME	POLYGON ((335968.8903511308 7397702.513956996,...	3645
70	87	VILA JACUI	POLYGON ((349986.8804435085 7402506.094028369,...	2689
71	89	VILA MARIA	POLYGON ((341226.5383482852 7399204.662956165,...	3270
72	90	VILA MARIANA	POLYGON ((334156.872631352 7391572.920592255, ...	12568
73	91	VILA MATILDE	POLYGON ((341994.9933479975 7396441.617957629,...	3219
74	92	VILA MEDEIROS	POLYGON ((340624.4800554968 7401718.265047844,...	2461
75	93	VILA PRUDENTE	POLYGON ((341776.3540084544 7388505.222995834,...	3487
76	55	PARELHEIROS	POLYGON ((333606.2023538684 7362933.638975473,...	2045
77	01	AGUA RASA	POLYGON ((341138.4406296417 7391958.215325007,...	2000
78	04	ARICANDUVA	POLYGON ((344113.8373471093 7391006.579960497,...	2000
79	06	BARRA FUNDA	POLYGON ((330226.8196582301 7396102.560938887,...	2000
80	07	BELA VISTA	POLYGON ((332547.7725661778 7392107.454237988,...	2000
81	09	BOM RETIRO	POLYGON ((334029.4403522231 7396065.476957881,...	2000
82	11	BRASILANDIA	POLYGON ((324940.1743566582 7404303.680953588,...	2295
83	12	BUTANTA	POLYGON ((322222.1511237745 7393130.728996397,...	7710
84	16	CAMPO GRANDE	POLYGON ((325612.9394202752 7381836.282049486,...	3086
85	18	CANGAIBA	POLYGON ((342208.3295162857 7398747.696811901,...	2372
86	20	CARRAO	POLYGON ((343628.4878784406 7392996.291127594,...	3510
87	24	CIDADE LIDER	POLYGON ((345226.3443463795 7394250.385958767,...	3372
88	25	CIDADE TIRADENTES	POLYGON ((358857.1192615093 7392606.612629535,...	2108
89	26	CONSOLACAO	POLYGON ((330220.6094902321 7396091.959604891,...	7937
90	41	JAGUARE	POLYGON ((322997.5993030824 7395024.342091518,...	2594
91	42	JARAGUA	POLYGON ((319217.494359625 7405224.915953157, ...	2580
92	44	JARDIM HELENA	POLYGON ((352960.6685789257 7402666.757417267,...	2356
93	45	JARDIM PAULISTA	POLYGON ((329650.5351547573 7390883.216611659,...	9486
94	46	JARDIM SAO LUIS	POLYGON ((321673.0860525215 7377826.369338656,...	2656
95	47	JOSE BONIFACIO	POLYGON ((354136.2479657517 7396748.686230991,...	2671

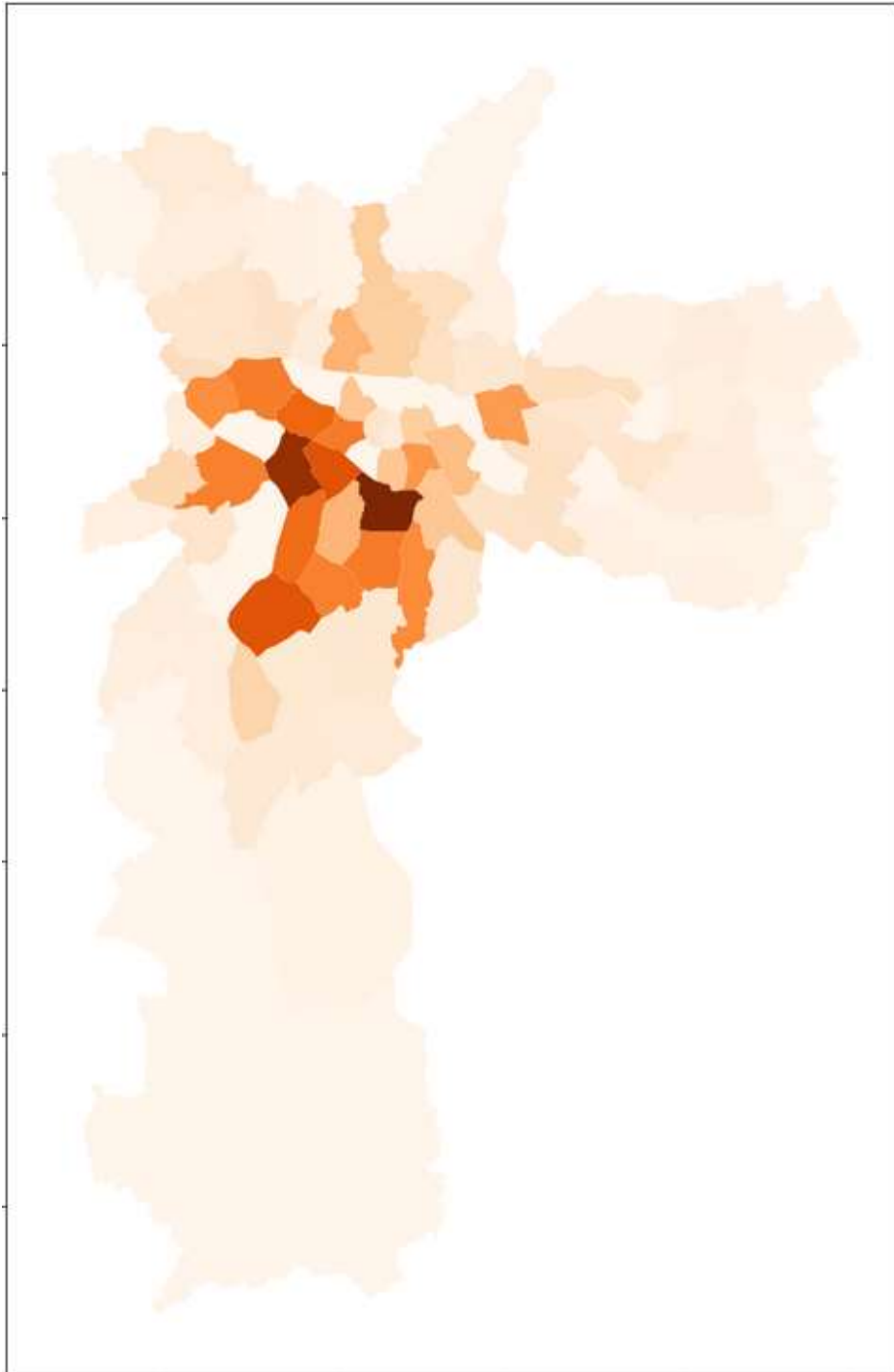Table 7: List of Boroughs and its GeoJson limits and family income, part 2.

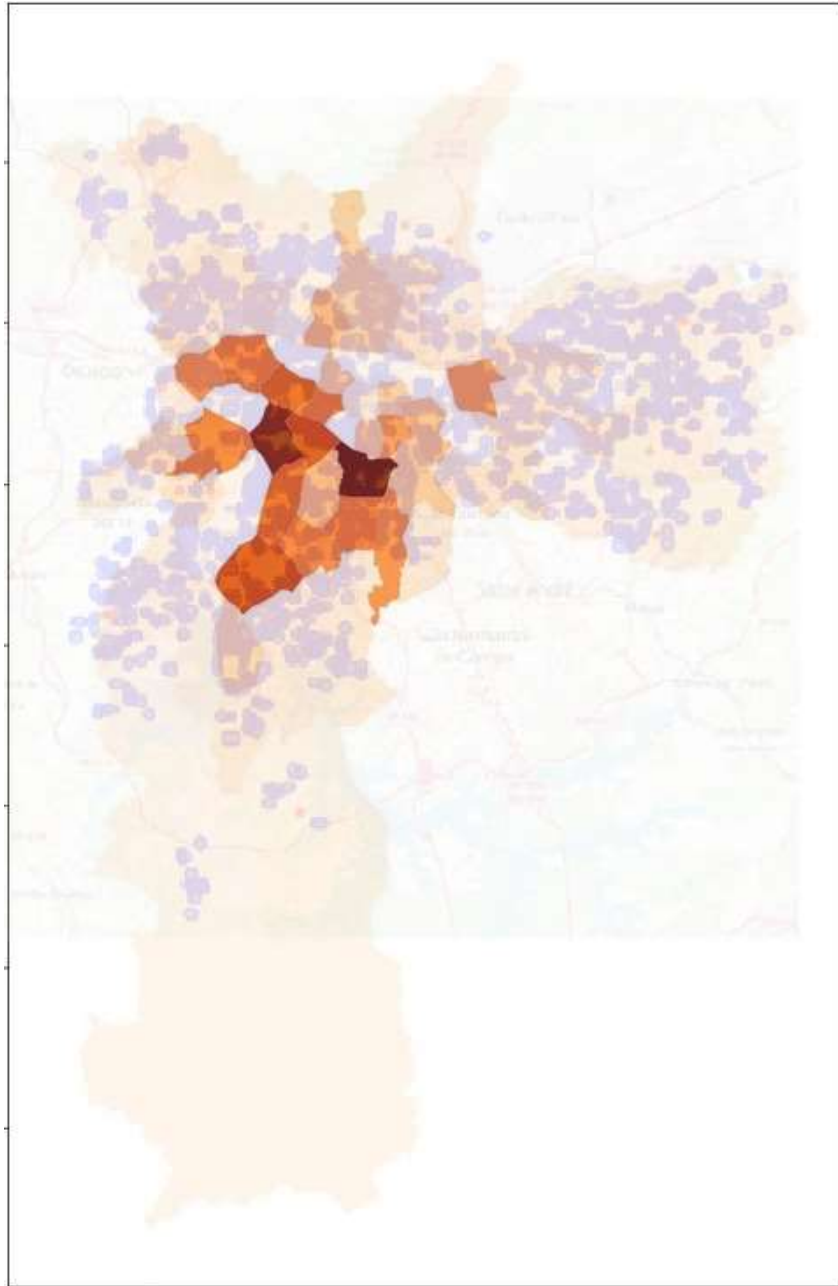Figure 7: Choroplet map with colors in red indicating a higher family income average by Borough.

Figure 8: Overlapping image between figures 6 and 7.

5. Discussion

As can be seen in figure 6, there are no spaces in the neighborhoods of São Paulo where there is a well defined area of empty spaces of restaurants, with this, it is important to note that whatever the chosen area this will dispute space and competition with other restaurants of the same or different types.

In Figure 7, it is possible to establish a well-defined region of the most income-earning areas, where a more specific survey can be carried out of which districts in this region could be considered for the establishment of the high quality restaurant of the XCompany group.

With the development of figure 8 from figures 6 and 7 it is possible to establish less dense regions of restaurants where the budget restaurant of the group could be implanted.

6. Conclusion

For the implementation of the restaurant income X was possible to establish 4 possible areas that would have less competition. In this way a more detailed study within these areas is indicated to establish the public contained in these and thus to develop the business plan for a greater probability of success of the enterprise. It is important to note that some of these areas have higher income areas as a peripheral region, which would be preferable in case of a joint implementation of both projects. The 4 regions is shown in Figure 9 below.
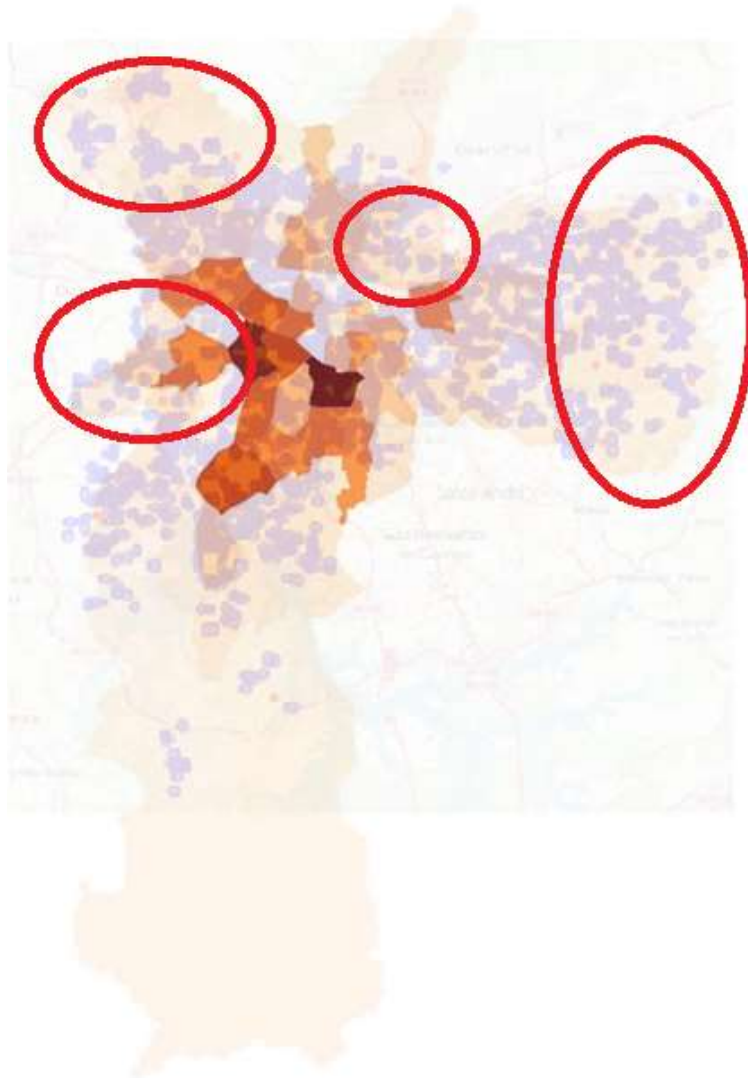

Figure 9: Less dense regions with established venues

And for the development and implementation of the high-quality restaurant, two preferable and peripheral regions were identified X in high income regions (in red) if the cost of implementation should be low which, otherwise if the location should be the driver of the enterprise, it is identified in the blue region in figure 10 below.
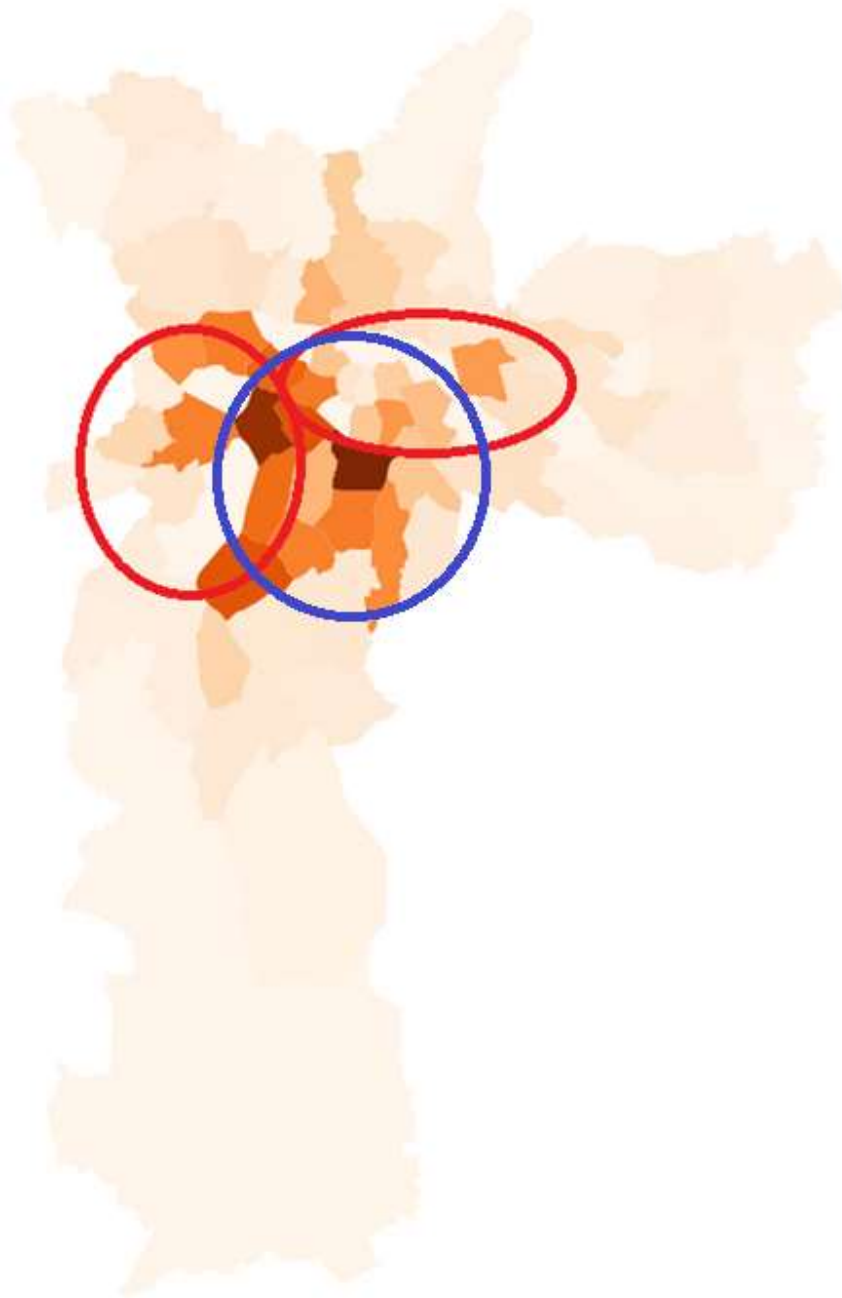

Figure 10: Areas with high income residents