

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 2](#_Toc532814758)

[1.1 Introduction 2](#_Toc532814759)

[1.2 Business Problem 2](#_Toc532814760)

[1.3 About São Paulo 2](#_Toc532814761)

[1.4 About XCompany 3](#_Toc532814762)

[2. About the data collection 5](#_Toc532814763)

[2.1 Foursquare database 5](#_Toc532814764)

[2.2 seade/dieese database 7](#_Toc532814765)

[3. Methodology 11](#_Toc532814766)

**(This study was developed for the Capstone project from the course “Applied Data Science Capstone” by Coursera/IBM and does not represent in any way a study comissioned for a real company, its main purpose being only the execution of the tasks required for the fulfillment of the project. Despite this, the data, as much as its sources, represents real numbers from trusted sources and institutions).**

# Introduction and business problem

## 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

## 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.

## About São Paulo

São Paulo as explained in Wikipedia is a [municipality](https://en.wikipedia.org/wiki/Municipality) in the [Southeast Region](https://en.wikipedia.org/wiki/Southeast_Region,_Brazil) of [Brazil](https://en.wikipedia.org/wiki/Brazil). The [metropolis](https://en.wikipedia.org/wiki/Metropolis) is an [alpha global city](https://en.wikipedia.org/wiki/Global_city) (as listed by the [GaWC](https://en.wikipedia.org/wiki/Globalization_and_World_Cities_Research_Network)) and the [most populous city](https://en.wikipedia.org/wiki/List_of_largest_cities_in_Brazil) in Brazil, the [Western Hemisphere](https://en.wikipedia.org/wiki/Western_Hemisphere) and the [Southern Hemisphere](https://en.wikipedia.org/wiki/Southern_Hemisphere), besides being the largest [Portuguese-speaking](https://en.wikipedia.org/wiki/Portuguese_language) city in the world. The municipality is also the [Earth's 11th largest city proper by population](https://en.wikipedia.org/wiki/List_of_cities_proper_by_population). The city is the [capital](https://en.wikipedia.org/wiki/Capital_city) of the surrounding [state](https://en.wikipedia.org/wiki/States_of_Brazil) of [São Paulo](https://en.wikipedia.org/wiki/S%C3%A3o_Paulo_(state)), one of the [most populous](https://en.wikipedia.org/wiki/List_of_Brazilian_states_by_population) and [wealthiest](https://en.wikipedia.org/wiki/List_of_Brazilian_federative_units_by_gross_domestic_product) states in Brazil. It exerts strong international influences in commerce, finance, arts and entertainment.The process of [conurbation](https://en.wikipedia.org/wiki/Conurbation) between the metropolitan areas located around the Greater São Paulo ([Campinas](https://en.wikipedia.org/wiki/Metropolitan_Region_of_Campinas), [Santos](https://en.wikipedia.org/wiki/Baixada_Santista), [Sorocaba](https://en.wikipedia.org/wiki/Sorocaba) and the [Paraíba Valley](https://en.wikipedia.org/wiki/Para%C3%ADba_Valley)) created the [São Paulo Macrometropolis](https://en.wikipedia.org/wiki/S%C3%A3o_Paulo_Macrometropolis), a [megalopolis](https://en.wikipedia.org/wiki/Megalopolis_(city_type)) with more than 30 million inhabitants, one of the most populous urban agglomerations in the world.

Having the largest economy by [GDP](https://en.wikipedia.org/wiki/Gross_domestic_product) in [Latin America](https://en.wikipedia.org/wiki/Latin_America) and the [Southern Hemisphere](https://en.wikipedia.org/wiki/Southern_Hemisphere), the city is home to the [São Paulo Stock Exchange](https://en.wikipedia.org/wiki/BM%26F_Bovespa). [Paulista Avenue](https://en.wikipedia.org/wiki/Paulista_Avenue) is the economic core of São Paulo. The city has the [11th largest GDP in the world](https://en.wikipedia.org/wiki/List_of_cities_by_GDP), representing alone 10.7% of all [Brazilian GDP](https://en.wikipedia.org/wiki/Economy_of_Brazil) and 36% of the production of goods and services in the [state of São Paulo](https://en.wikipedia.org/wiki/S%C3%A3o_Paulo_(state)), being home to 63% of established [multinationals](https://en.wikipedia.org/wiki/Multinational_corporation) 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](https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)) by 2017 estimates.

The metropolis is also home to several of [the tallest skyscrapers in Brazil](https://en.wikipedia.org/wiki/List_of_tallest_buildings_in_Brazil), including the [Mirante do Vale](https://en.wikipedia.org/wiki/Mirante_do_Vale), [Edifício Itália](https://en.wikipedia.org/wiki/Edif%C3%ADcio_It%C3%A1lia), [Banespa](https://en.wikipedia.org/wiki/Altino_Arantes_Building), [North Tower](https://en.wikipedia.org/wiki/Centro_Empresarial_Na%C3%A7%C3%B5es_Unidas) 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](https://en.wikipedia.org/wiki/Latin_America_Memorial), the [Ibirapuera Park](https://en.wikipedia.org/wiki/Ibirapuera_Park), [Museum of Ipiranga](https://en.wikipedia.org/wiki/Museu_Paulista), [São Paulo Museum of Art](https://en.wikipedia.org/wiki/S%C3%A3o_Paulo_Museum_of_Art), and the [Museum of the Portuguese Language](https://en.wikipedia.org/wiki/Museum_of_the_Portuguese_Language). The city holds events like the [São Paulo Jazz Festival](https://en.wikipedia.org/wiki/Sao_Paulo_Jazz_Festival), [São Paulo Art Biennial](https://en.wikipedia.org/wiki/S%C3%A3o_Paulo_Art_Biennial), the [Brazilian Grand Prix](https://en.wikipedia.org/wiki/Brazilian_Grand_Prix), [São Paulo Fashion Week](https://en.wikipedia.org/wiki/S%C3%A3o_Paulo_Fashion_Week) and the [ATP Brasil Open](https://en.wikipedia.org/wiki/Brasil_Open). The [São Paulo Gay Pride Parade](https://en.wikipedia.org/wiki/S%C3%A3o_Paulo_Gay_Pride_Parade) rivals the [New York City](https://en.wikipedia.org/wiki/New_York_City) [Pride March](https://en.wikipedia.org/wiki/New_York_City_Pride_March) as the largest [gay pride](https://en.wikipedia.org/wiki/Gay_pride) [parade](https://en.wikipedia.org/wiki/Pride_march) in the world. It is headquarters of the [Brazilian television](https://en.wikipedia.org/wiki/Television_in_Brazil) networks [Band](https://en.wikipedia.org/wiki/Rede_Bandeirantes), [Gazeta](https://en.wikipedia.org/wiki/Rede_Gazeta), and [RecordTV](https://en.wikipedia.org/wiki/RecordTV).

São Paulo is a [cosmopolitan](https://en.wikipedia.org/wiki/Multiculturalism), [melting pot](https://en.wikipedia.org/wiki/Melting_pot) city, home to the largest [Arab](https://en.wikipedia.org/wiki/Arab_Brazilians), [Italian](https://en.wikipedia.org/wiki/Italian_Brazilians), and [Japanese](https://en.wikipedia.org/wiki/Japanese_Brazilians) diasporas, with examples including [ethnic neighborhoods](https://en.wikipedia.org/wiki/Ethnic_enclave) of [Mercado](https://pt.wikipedia.org/wiki/Mercado_(bairro_de_S%C3%A3o_Paulo)), [Bixiga](https://en.wikipedia.org/wiki/Bixiga), and [Liberdade](https://en.wikipedia.org/wiki/Liberdade_(district_of_S%C3%A3o_Paulo)) respectively. In 2016, inhabitants of the city were native to over 200 different countries. The city's [Latin](https://en.wikipedia.org/wiki/Latin) motto, which it has shared with [the battleship](https://en.wikipedia.org/wiki/Brazilian_battleship_S%C3%A3o_Paulo) and [the aircraft carrier](https://en.wikipedia.org/wiki/Brazilian_aircraft_carrier_S%C3%A3o_Paulo_(A12)) named after it, is Non ducor, duco, which translates as "I am not led, I lead."[[21]](https://en.wikipedia.org/wiki/S%C3%A3o_Paulo#cite_note-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](https://en.wikipedia.org/wiki/List_of_tallest_buildings_in_S%C3%A3o_Paulo).

## 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.

# 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.

## Foursquare database

Foursquare is a local search-and-discovery service [mobile app](https://en.wikipedia.org/wiki/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](https://developer.foursquare.com/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.





Figure 1: Data Strucure From Foursquare’s API

## 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.

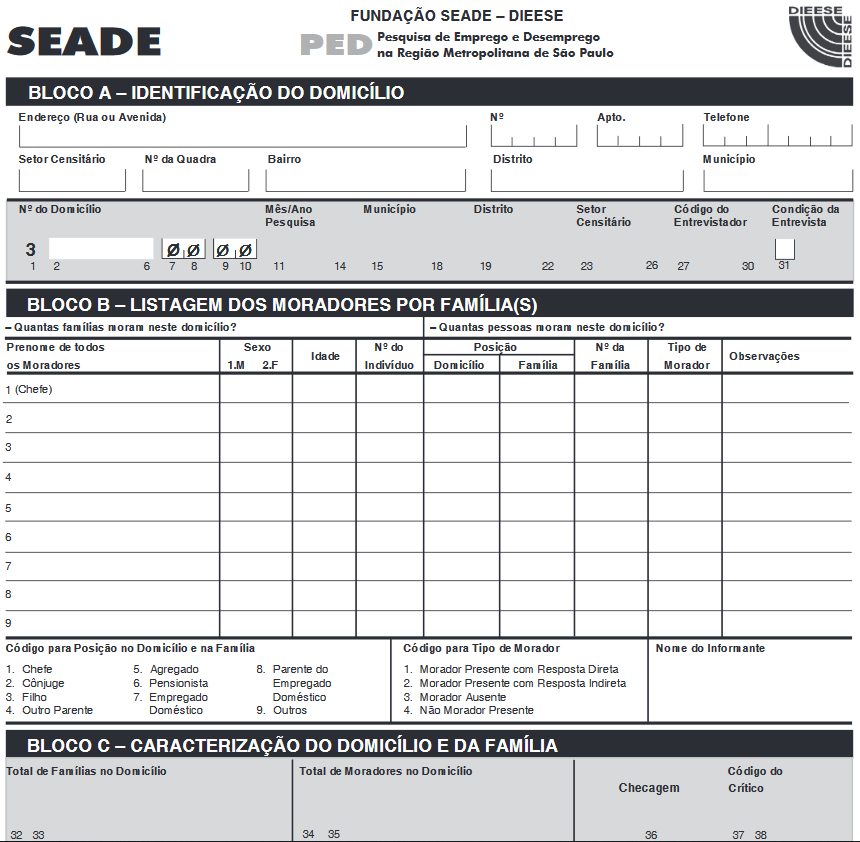


Figure 2: Sample of first page from survey form

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.



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.

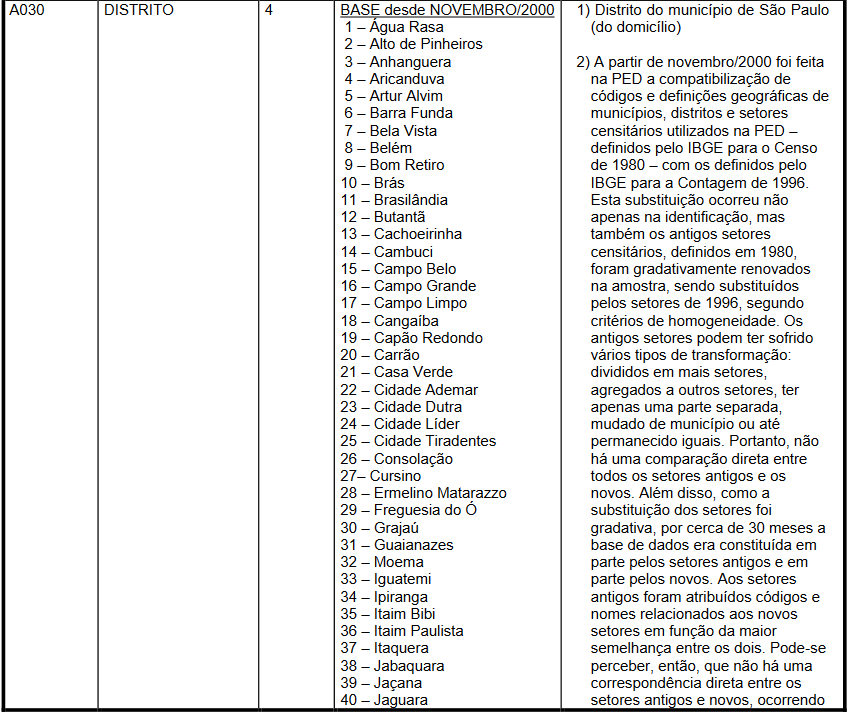
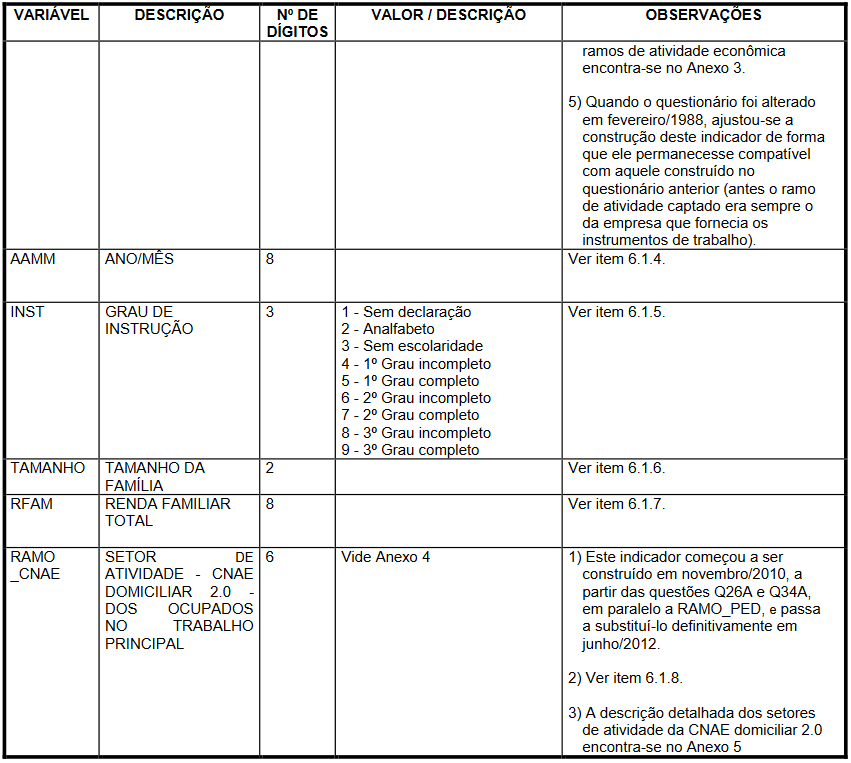


Figure 4: Database dictionary – Borough

**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.



# Methodology