

Renato de Oliveira Souza

Exploring new spots for Pilates Studio in São Paulo - Brazil

IBM Data Science Capstone Project

São Paulo – SP – Brazil

2021

INTRODUCTION

São Paulo is one of the largest cities in the world, with more than 12 million inhabitants (5% of the Brazilian population) and occupies the first place in the ranking of municipal GDP in Brazil, with 10% of the Brazilian GDP.

Thus, São Paulo is considered the main Brazilian market for several consumer items, including services focused on well-being and health.

Among the services focused on health and well-being are fitness centers, spas, yoga studios and, in growing relevance, pilates studios.

Pilates is an exercise method developed by Joseph Pilates in the 1920s, being a recognized technique for treating and preventing postural problems. The method works on concentration, body centralization, precision of movement, breathing, motor control and fluidity of movement.

A client intends to expand its operations within the city of São Paulo, opening a new studio and defining some parameters, based on the registration of clients of the existing units.

The criteria are:

- People aged 40 or over;
- Higher education, such as minimum schooling;
- Live up to 900 meters away from the studio;
- High family income.

To support this decision, we will use indicators at the neighborhood level, such as the human development indicator (HDI), to capture schooling and health concerns. The population density of people aged 40 and over to capture public concentration and monthly family income to capture neighborhoods with a high family income.

These indicators will build an index that added to a survey of the geographic distribution of gyms, spas and pilates studios in the city of São Paulo, carried out through the API of the Foursquare website, will indicate the best location to open the new studio, using algorithms of clustering like K-means.

DATA

The data to be used comes from public sources. The Human Development Index (HDI) is a summary measure of long-term progress in three basic dimensions of human development: income, education and health. website: [List of Sao Paulo's Boroughs by Human Development Index](#).

Neighborhood	HDI
Moema	0.961
Pinheiros	0.960
Perdizes	0.957
Jardim Paulista	0.957
Alto de Pinheiros	0.955

To calculate the demographic density of people over 40 years of age, the area of each neighborhood in the city of São Paulo will be collected on the website: [Demographic data of the city of São Paulo](#) and the population of the neighborhoods by age group on the website: [Population by age group](#), which provides the data in CSV format.

$$\text{Density} = \frac{\text{Population over 40}}{\text{Area (Km}^2\text{)}}$$

Neighborhood	Area (Km2)	Population over 40	Density
Aricanduva	6.6	32288	0.42
Carrão	7.5	34431	0.40
Vila Formosa	7.4	36903	0.43
Butantã	12.5	21827	0.15
Morumbi	11.4	20644	0.16

The data on the average monthly family income will be taken from the annual report “Map of inequality”, available on the website: www.nossasaopaulo.org.br.

Neighborhood	Monthly family income(R\$)
Alto de Pinheiros	9591.93
Perdizes	9348.58
Jardim Paulista	9327.12
Moema	9248.43
Santo Amaro	9159.73

These variables will be normalized by min max, for a scale of 0 to 1 and the average value between them calculated, to create the index of place of interest (HDDII), as a reference.

$$\text{HDDII} = \frac{\text{IDH} + \text{Density}^{(\text{min max})} + \text{Family Income}^{(\text{min max})}}{3}$$

Neighborhood	HDDII
Perdizes	0.90
Jardim Paulista	0.86
Moema	0.81
Alto de Pinheiros	0.79
Consolação	0.76

Geographical location of São Paulo neighborhoods will be gotten from Geopy Python library that converts geographical searches by name into coordinates – latitudes and longitudes. The neighborhood venues will be retrieved within 900 meters radius from each neighborhood geographical center using the Foursquare API. This radius was selected because it can select the perfect distance to move from home to new site by foot. After retrieving the venues, the API also returns the latitude and longitude of each venue, to complete the analysis for the best location.

Neighborhood	Latitude	Longitude
Alto de Pinheiros	-23.549.461	-46.712.293
Anhanguera	-23.432.908	-46.788.534
Aricanduva	-23.578.024	-46.511.454
Artur Alvim	-23.539.221	-46.485.265
Barra Funda	-23.525.462	-46.667.513

Neighborhood	Venue	Venue Latitude	Venue Longitude	Venue Category
Moema	Arte de Viver	-23.596.409	-46.666.665	Yoga Studio
Moema	Amadí SPA	-23.599.936	-46.660.857	Spa
Moema	Smart Fit	-23.601.100	-46.665.530	Gym / Fitness Center
Moema	Race Bootcamp	-23.595.136	-46.671.162	Gym / Fitness Center
Moema	Needs Academia	-23.602.955	-46.657.411	Gym