Capstone Project
Japanese Restaurants in São Paulo

Applied Data Science Capstone by IBM/Coursera

Manuela Carvalho

1. Introduction

About São Paulo

São Paulo is the biggest city in Brazil. As all metropolis around the world, it has a long history of international immigration and it is home to nearly 12.2 million people in 2018, accounting for over 27% of the population of São Paulo State.

The city continues to be the leading metropolitan gateway for legal immigrants admitted into Brazil.

São Paulo is known for its strong Asian culture, not only the one that came with Japanese immigrants but also Korean, Chinese, and many others.

In 1908, the Japanese began to immigrate to Brazil, becoming the place with the highest number of Japanese people outside Japan. The Brazilian census agency calculated in 2017 there were around 1.5 million Japanese descendants living in Brazil.

For many years, the Japanese culture was concentrated in São Paulo, in a neighborhood called 'Liberdade' ('Freedom' in Portuguese). There, the majority of the shops sell Asian related products.

Nowadays, Japanese cooking is a Brazilian passion, and it gained lots of local adaptations.

2. Business Problem

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an **Japanese restaurant** in **São Paulo, Brazil**.

Since there are lots of japanese restaurants in São Paulo we will try to detect **locations that are already crowded with japanese restaurants** and **locations that are not already crowded with japanese restaurants**. We are also particularly interested in **areas with no Japanese restaurants in vicinity**. We would also prefer locations **as close to city center as possible**, assuming that first two conditions are met.

We will use our data science powers to generate a few most promissing neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

3. Data Collection

Based on definition of our problem, factors that will influence our decission are:

- number of existing restaurants in the neighborhood (any type of restaurant)
- number of and distance to Japanese restaurants in the neighborhood, if any
- distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

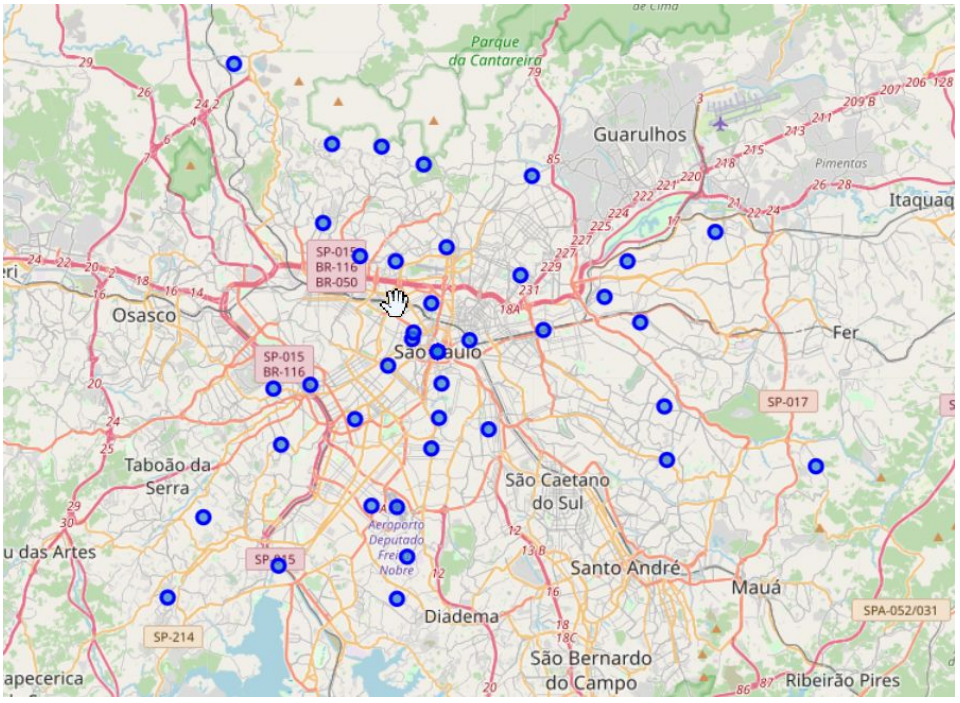
Following data sources will be needed to extract/generate the required information:

São Paulo's information about neighborhoods and boroughs will be obtained from the data source: https://en.wikipedia.org/wiki/List_of_postal_codes_in_Brazil

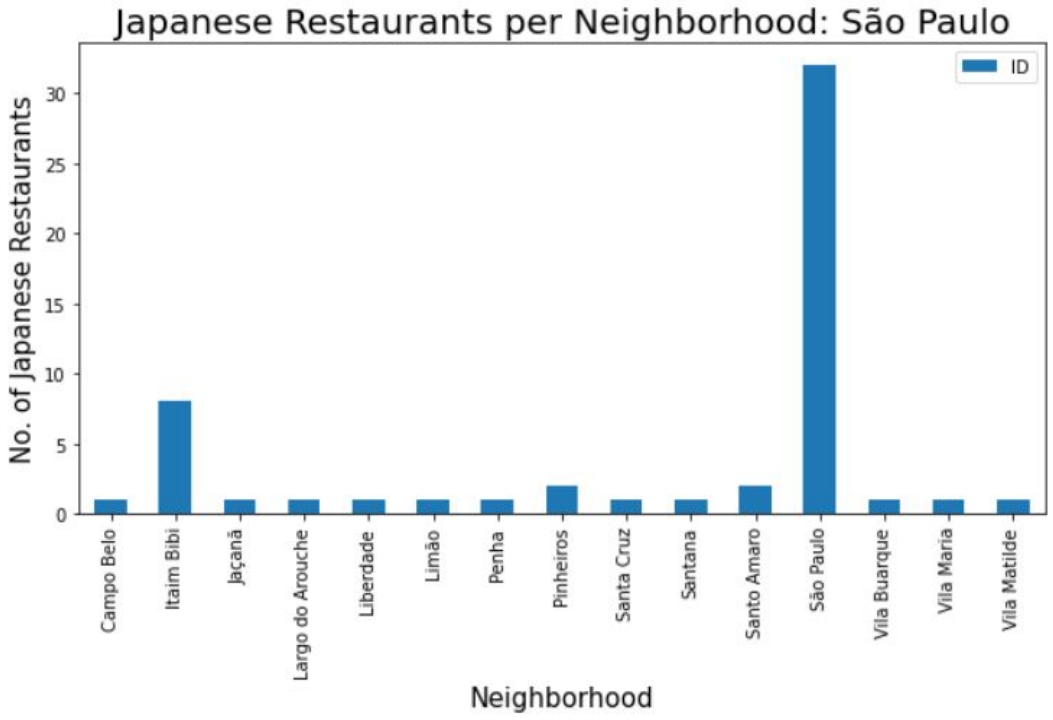
	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	01010	São Paulo	São Paulo	-23.550651	-46.633382
1	01060	São Paulo	São Paulo	-23.550651	-46.633382
2	01070	São Paulo	São Paulo	-23.550651	-46.633382
3	01080	São Paulo	São Paulo	-23.550651	-46.633382
4	01090	São Paulo	São Paulo	-23.550651	-46.633382
...	...	...	...	...	...
58	05500	São Paulo	Butantã	-23.569131	-46.721874
59	05600	São Paulo	Morumbi	-23.596499	-46.717845
60	05700	São Paulo	Campo Limpo	-23.632558	-46.759666
61	05800	São Paulo	Capão Redondo	-23.671903	-46.779435
62	08383	São Paulo	3ª Divisão	-23.607466	-46.429690

63 rows × 5 columns

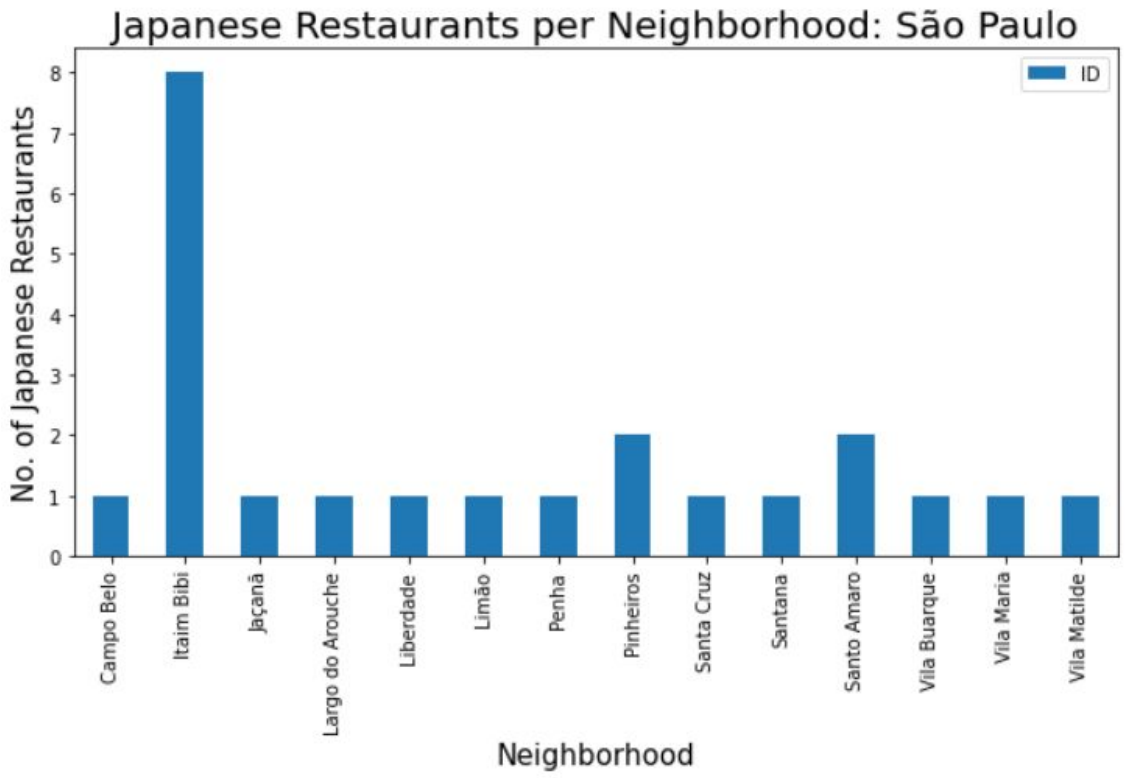
All data related to locations and quality of Japanese restaurants will be obtained via the FourSquare API utilized via the Request library in Python.



We got the following number of restaurants



We have a problem with our data, because majority have neighborood 'São Paulo'. Let's drop this values



4. Methodology

- Data fom São Paulo will be collected, cleaned and processed into a dataframe.
- FourSquare be used to locate all venues and then filtered by Japanese restaurants. Ratings, tips, and likes by users will be counted and added to the dataframe.
- Data will be sorted based on rankings.
- Finally, the data will be visually assessed using graphing from Python libraries.

5. Business Analysis

Areas with the tradition of the same kind of restaurants can be a two-edged sword. In these areas, people already expect to find those kinds of restaurants. Therefore, marketing is automatic. On the other side, competition is fierce.

Places with little or no tradition with this kind of restaurant can be an excellent place to start because there is no competition. But maybe the neighborhood is not used for the type of food.

In our analysis, the majority of restaurants are in Itaim Bibi. But I think it is underreported. I believe a more in-depth analysis should be done. Liberdade is a well know place with Japanese culture, and the search returned only one restaurant.

In São Paulo, I believe any place is an excellent place to open a Japanese restaurant. The population loves this kind of food. But I think in the nearby areas of Liberdade Neighborhood would be the best place.

6. Conclusion

The purpose of this project was to identify São Paulo areas close to center with low restaurants (particularly Japanese restaurants) to aid stakeholders in narrowing down the search for the optimal location for a new Japanese restaurant. By calculating restaurant density distribution from Foursquare data, we first identified general boroughs that justify further analysis (Liberdade) and then generated an extensive collection of locations that satisfy some basic requirements regarding existing nearby restaurants. The clustering of those locations was then performed to create significant zones of interest (containing the most significant number of potential areas). The addresses of those zone centers were designed to be used as starting points for final exploration by stakeholders.

The final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and places in every recommended zone. I am considering factors like the attractiveness of each position (proximity to Liberdade neighborhood), levels of noise/proximity to major roads, real estate availability, prices, social and economic dynamics of every area.