

**Capstone Applied Data Science: Battle of neighborhoods Rio de Janeiro**

Business Understanding and Data: December, 2018 – Marcelo Tueiv

## Introduction

Natural, sustainable and transparent consumption will no longer be an exclusive and luxury philosophy to become daily and widespread. Personal and planet care is the strongest trend in response to the stressful and frenetic modern life. And it is what will also determine the new demands of the customers. In the city of Rio de Janeiro this trend will become even stronger, where people have a greater concern with appearance and sports, because their bodies are more exposed on the beaches and because of the heat, where the themperatures reach 45 degrees in the summer.

## Objective

The objective of this project is to explore the possibilities of where to open a healthy food restaurant in the city of Rio de Janeiro and have a greater chance of success due to the characteristics of the place (transportation, proximity to gyms and competition from other restaurants). To do that we will analyze other successful healthy food restaurants characteristics and suggest places to open a new one following these characteristics.

### Stakeholders

Future entrepreneurs who wants to opent healthy food restaurant in Rio de Janeiro

### Selection criteria

The location of the establishment can guarantee the success of sales and, for this, following some behaviors identified in Rio de Janeiro we have, as points indicated for the opening of a restaurant of this type some factors that increase the chance of consumer frequency. We imagine that, some factors to be validated by the result of the approach we will implement can be (to be evaluated by the project):  
- Must be located within 500 meters of the subway  
- High income neighborhood  
- Proximity of more number of gyms  
- Proximity of the smallest possible number of restaurants of this type, with good evaluation

Initially, for this, we must obtain 3 clusters by grouping the neighborhoods as follows:  
- Cluster 0: neighborhoods that indicate a high chance of success of opening a restaurant of this type  
- Cluster1: neighborhoods that indicate little chance of success opening a restaurant of this type  
- Cluster 2: neighborhoods that are neither attractive nor negative for opening a restaurant of this type

## Data

This project will require the following data sources, namely:

**Rio de Janeiro Subway stations**

To identify proximity to subway stations in the city of Rio de Janeiro we will consult page table in Wikipedia with list of all the subway stations of the city and their references of longitude and latitude:

<https://pt.wikipedia.org/wiki/Lista_de_esta%C3%A7%C3%B5es_do_Metr%C3%B4_do_Rio_de_Janeiro>

Sample:



**Neighborhoods and socio-economics**

To identify the neighborhoods and socio-economic data as average income we will consult this other list, also existing in Wikipedia:

<https://pt.wikipedia.org/wiki/Lista_de_bairros_do_Rio_de_Janeiro_por_IDH>

Sample:



**Commercial Establishments Nearby**

Foursquare is a location data provider with information and evaluation about restaurants, hotels, gyms and other venues and events within an area of interest. Such information includes venue names, locations, evaluation, latitude and longitude. Foursquare information of these venues will be obtained via Foursquare API.

**Other Data Source, if necessary**

Another rich source of information is the open data site of the city of Rio de Janeiro, where we can obtain information such as rate of violence and commercial value of real estate:

<http://www.data.rio/>