

# Capstone Project

Author: Matheus Cammarosano Hidalgo

## Data

As mentioned in the Introduction, socioeconomic data of each district is gathered for the analysis. The initial source of data is the Wikipedia page: [https://pt.wikipedia.org/wiki/Lista\\_dos\\_distritos\\_de\\_S%C3%A3o\\_Paulo\\_por\\_popula%C3%A7%C3%A3o](https://pt.wikipedia.org/wiki/Lista_dos_distritos_de_S%C3%A3o_Paulo_por_popula%C3%A7%C3%A3o) and from it the data of each district is obtained. The data of interest of each district is GDP per capita, which indicates how wealthy it is, the HDI, that shows how developed is the neighborhood. The HDI englobes a number of indexes and indicates how economically and socially developed is a determined place. Also, the population of each district is collected in order to quantify a potential amount of customers in a region.

Also, the geospatial coordinates of the districts with similar socioeconomics characteristics of Pinheiros are gathered. The data source is <https://www.adistanciaentre.com/>. With this information, the distance of the districts from Pinheiros (where the first restaurant is located) is calculated.

From FourSquare Developer API, the Italian restaurants and pizzas places from the districts and its respective price tiers and ratings are collected. In Brasil, it is relatively normal that Italian restaurants have pizzas in their menus, so the classification between Italians and pizzerias might have some intersections, so in this project both categories are analyzed together.