# Data: Sources and Concerns

Brazil is composed of 5,570 municipalities. In order to collect the information related to them, collection was performed from several different sources, indicated below:

Human Development Index: UNDP-United Nations Development Programme

Per Capita Income : IBGE – Instituto Brasieiro de Geografia e Estatística (Brazilian Institute for Geographics and Statistics)

Population: IBGE – Instituto Brasieiro de Geografia e Estatística (Brazilian Institute for Geographics and Statistics)

Radio Base Stations: ANATEL – Agência Nacional de Telecomunicações (Brazilian National Telecommunications Agency)

Venues: Foursquare

Population will be used directly as an attribute, as well as for feature engineering. For example, it will be used to normalize the RBS and Venues data. Therefore, the RBS and Venues data will be converted to ‘per capita’ values, as (intuitively) it is expected that one city with twice the number of inhabitants tends to have two times more venues and RBS deployed to serve its population.

The data will be filtered to identify outliers. E.g. São Paulo city is the largest city in the south hemisphere, and certainly will distort the numbers. Similarly, cities on the lower limit (e.g. under 1,000 habitants) will be reviewed if shall be included.

Foursquare limits the number of venues per request to 100. This cap will create a distortion in any linearity analysis, therefore it will be verified if the cities hitting this cap shall be removed from the analysis to avoid its implications.