

2. Data Acquisition and Cleaning

2.1 Data Acquisition

The data acquired for this project is a combination of data from three sources. The first data source of the project uses a OPEN DATA ISP that shows the crime per Citys in Rio de Janeiro/BR.

The dataset contains many itens but we wil use the following columns:

- **fmun_cod**: location of cities.
- **fmun**: name for city borough.
- **year** – that crime occour 2014 - 2019
- **roubo_comercio - trade theft**
- **Roubo_carga** - Cargo theft

The second source of data is scraped from a wikipedia page that contains the list of Cities in boroughs. This page contains additional information about the boroughs, the following are the columns:

- **Borough**: The names of the 21 Rio de Janeiro boroughs.
- **Population**: Quantiti of cityzens.
- **Territorial area**. Meter square of cities
- **Densiti**: The political party that control the borough.
- **Population (2013 est)[1]**: The population in the borough recorded during the year 2013.

The third data source is the <http://astro.if.ufrgs.br/br.htm> . This data set contain:

Latitude – Longitude – Nome