Baixei o arquivo em uma planilha Excel, para analisar a qualidade dos atributos.

**Observação**: quantidade de registros se refere a registros com valores não nulos

**created\_on**

-dados de2013 a 2016

**property\_type** - apartment, house, STORE, 1,4% lixo com urls de....

\*\*\*\*\* **place\_name**

- pode ser um estado, município ou bairro o menor nível hierárquico informado no place\_with\_parent\_name, ou lixo

\*\*\*\*\***place\_with\_parent\_name** - dividi o atributo em quatro:

- País, estado, município, bairro

\*\*\*\*\***país**

- 2 valores Brasil e Miami (140 registros)

**estado**

- Proporção de infos para estado de SP - 611k/872k - 70%

- estados na base 26 e DF

**município**

**- São Paulo –** 385 municípios na base do total de 645 no estado

**- Rio de Janeiro -** 67 municípios na base do total de 92 no estado

**bairro**

**-** 575550 registros preenchidos com bairro para um total de 868285 registros de município

**-** 292745 registros sem bairro – aproximadamente 33,7%

**geonames\_id, lat, lon, description** se encontram em tabela em anexo

**price**

- 819299 registros

**currency**

- duas moedas BRL e USD (235 registros)

**price\_aprox\_local\_currency**

- 819227 registros

**price\_aprox\_usd**

- 819172 registros

**price\_usd\_per\_m2 e price\_per\_m2**

- valores estranhos - 589139 registros

**Ponto de atenção**: para fazermos valores médios por estado ou município temos que levar em conta que temos preços de 3 ANOS DIFERENTES e precisamos separar a variável **created\_on**.