

# Projeto OpenStreetMap Data Wragling com MongoDB

Jader Martins Camboim de Sá

26 de Abril, 2018

## 1 Problemas Encontrados no Mapa

A primeira vista observei que existiam:

- Códigos postais em diferentes formatações: 71733-105, 7289255, 71.727-127.
- Endereços errados ou inconsistentes como `{city:GuaraII}`, `{suburb:GuaraII, city:Brasilia}`.
- Atributo “place” quando deveria estar como “city”.

### 1.1 Códigos Postais

Durante algumas rápidas checagens foram encontradas má formatação, é importante que os dados sigam o mesmo padrão de apresentação para um mesmo campo, porém os ceps (codigos postais) estão despadronizados, com o python cada elemento de cep teve a formatação removida e repadronizado para o formato mais usual (12.345-678), por fim reinserida com um update.

```
13 for x in db.mapbsb.find():
14     if 'postcode' in x:
15         newcep = cepformat (x['postcode'])
16         db.mapbsb.update_one({'_id ': x['_id']},
17                               {'$set':{'postcode': newcep }}))
```

### 1.2 Endereços

Endereços estavam em variáveis incorretas, cidades como subúrbios, bairros e regiões administrativas sem posição certa, parte disso se deve a natureza da própria região, que segue uma lógica diferente do resto do Brasil para delimitação de territórios, também estavam em categorias completamente fora de

sentido como uma cidade com atributo place. Uma parte significativa dessas correções tiveram de ser manuais.

## 2 Visão Geral do Dados

Table 1: Descrição dos dados.

Tamanho	Arquivo
76M	docs.json
83M	map.osm

### 2.1 Número de Documentos

```
mongo> db.mapbsb.find().count()
349396
```

### 2.2 Número de Tags por Tag

```
mongo> db.mapbsb.find({type:'node'}).count()
349396
mongo> db.mapbsb.find({type:'way'}).count()
77351
mongo> db.mapbsb.find({type:'relation'}).count()
795
```

### 2.3 Número de Usuários Únicos

```
> db.mapbsb.distinct('created.user').length
538
```

### 2.4 Maiores Contribuidores

```
mongo> db.mapbsb.aggregate([{"$group":{"_id":"$created.user", "count":{"$sum":1}}},
                             {"$sort":{"count": -1}},{"$limit":5}])
{ "_id" : " erickdeoliveiraleal ", "count" : 105393 }
{ "_id" : " MAPconciierge ", "count" : 38126 }
{ "_id" : " teste18 ", "count" : 27945 }
{ "_id" : " Linhares ", "count" : 24825 }
{ "_id" : "wille", "count" : 17899 }
```

### 3 Ideias Adicionais

Mesmo com a grande quantidade de ciclovias em Brasília, poucas são demarcadas no mapa, o que cria seu desconhecimento e desuso, seria interessante se o governo do Distrito Federal disponibilizasse tal informação no OpenStreetMap.

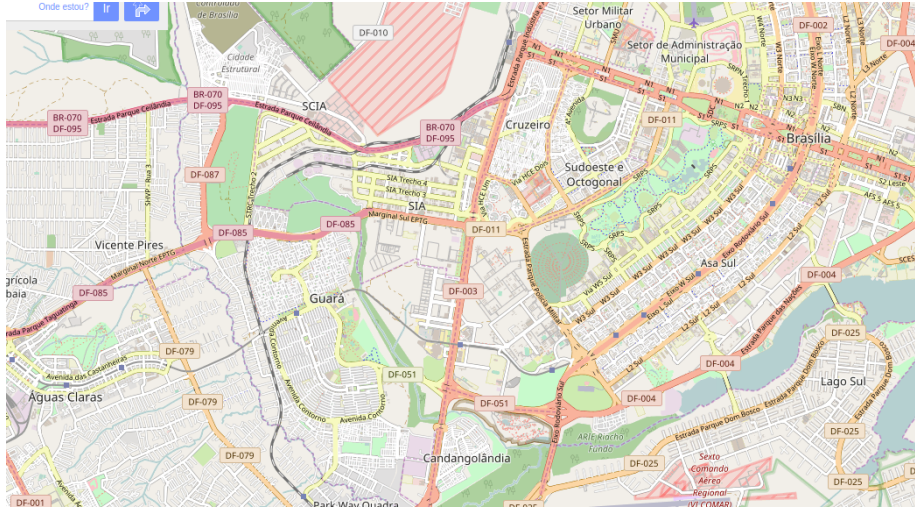


Figure 1: Entorno, Brasília, Distrito Federal - Brasil

### 4 Referências

[https://docs.google.com/document/d/1F0Vs14oNEs2idFJR3C\\_OPxwS6L0HPliOii-QpbmrMo4/pub](https://docs.google.com/document/d/1F0Vs14oNEs2idFJR3C_OPxwS6L0HPliOii-QpbmrMo4/pub)

[https://wiki.openstreetmap.org/wiki/Main\\_Page](https://wiki.openstreetmap.org/wiki/Main_Page)