Projeto Openstreetmap

Área: São Paulo – central

# Problemas encontrados

1. Inicialmente um problema que compromete as análises são os nomes com e sem acentuação, minha opção foi retirar a acentuação de todos os nomes para uma melhor análise
2. Segundo tinham problemas na nomenclatura dos tipos de ruas, tratei os tipos mais comuns, removi as abreviações e aqueles que não tinha indicação tratei como rua
3. Os ceps foram formatados para 00000-000, facilitando análises
4. Nos nomes das cidades maiúsculas e minúsculas traziam problemas, optei por colocar todas com as primeiras letras em maiúsculas
5. Vários itens não possuíam localização

# Análise dos dados

1. Primeiramente os dados contidos no OSM foram tratados, foi gerado um json que foi importado para o MongoDB no banco dbopen e collection sp.

## Análise por cep

1. Inicialmente verifiquei os ceps com maior quantidade de pontos

pipeline = [ {"$match":{"address.postcode":{"$exists":1}}},

{"$group":{"\_id":"$address.postcode", "count":{"$sum":1}}},

{"$sort": {"count":-1}},

{"$limit": 3}]

CEPs com maiores pontos:

03017-000: 247

03026-000: 229

03011-001: 226

## Análise por cidade

1. Verificação da distribuição por cidade, São Paulo na área analisada concentra maior parte dos ponto

pipeline = [{"$match":{"address.city":{"$exists":1}}},

{"$group":{"\_id":"$address.city", "count":{"$sum":1}}},

{"$sort": {"count":-1}},

{"$limit": 10}]

Cidades com maiores pontos:

Sao Paulo: 18416

Mooca: 1

## Análise dos usuários

1. Verifiquei que os contribuidores são poucos, e há uma concentração em alguns poucos:

pipeline = [{"$group":{"\_id":"$created.user", "count":{"$sum":1}}},

{"$sort": {"count":-1}}]

Usuarios com maiores publicacoes:

Bonix-Mapper: 992969

O Fim: 30279

naoliv: 2791

AjBelnuovo: 2747

MCPicoli: 2638

Wololo: 1882

ygorre: 1847

doenza: 1652

D4RK-L3G10N: 1313

Arthur R Gonçalves: 964

Distribuicao das publicaces por Usuarios %:

Bonix-Mapper: 94.41%

O Fim: 2.88%

naoliv: 0.27%

AjBelnuovo: 0.26%

MCPicoli: 0.25%

Wololo: 0.18%

ygorre: 0.18%

doenza: 0.16%

D4RK-L3G10N: 0.12%

Arthur R Gonçalves: 0.09%

1. Existe uma concentração muito forte em um usuário com quase 95% das publicações.
2. Os 2 maiores representam: 97.29%
3. Os 10 maiores representam: 98.79%
4. Total de usuários que contribuem menos de 1%: 417
5. Existem 155 usuários com uma única publicação:

pipeline = [{"$group":{"\_id":"$created.user", "count":{"$sum":1}}},

{"$group":{"\_id":"$count", "num\_users":{"$sum":1}}},

{"$sort":{"\_id":1}},

{"$limit":1}]

## Outras análises

1. Anos, cadastros por anos, o ano de 2016 foi que mais teve cadastros, metade de 2017 já é maior que os demais anos:

pipeline = [ { "$project" : { "ano" : {"$substr": ["$created.timestamp", 0, 4] }}}, { "$group" : { "\_id" : "$ano", "count" : { "$sum" : 1 }}},   
{ "$sort" : { "count" : -1 }} ]

2016: 738285

2017: 294579

2013: 6007

2015: 5370

2014: 3098

2012: 2257

2011: 1634

2010: 331

2009: 133

2007: 56

1. Facilidades, quantitativos por tipo de facilidade

pipeline = [{"$match":{"amenity":{"$exists":1}}},

{"$group":{"\_id":"$amenity","count":{"$sum":1}}},

{"$sort":{"count": -1}},

{"$limit":10}]

Facilidades:

parking: 1471

restaurant: 488

bank: 255

fast\_food: 197

fuel: 180

pub: 162

school: 103

pharmacy: 91

bicycle\_rental: 88

bar: 88

1. Religião, quantitativos por Religião

pipeline = [{"$match":{"amenity":{"$exists":1}, amenity":"place\_of\_worship"}},

{"$group":{"\_id":"$religion", "count":{"$sum":1}}},

{"$sort":{"count": -1}},

{"$limit":10}]

Religiao:

christian: 69

Nao Identificado: 9

jewish: 5

muslim: 2

1. Cozinha, quantitativos por tipo de cozinha

pipeline = [{"$match":{"amenity":{"$exists":1}, "amenity":"restaurant"}},

{"$group":{"\_id":"$cuisine", "count":{"$sum":1}}},

{"$sort":{"count":-1}},

{"$limit":10}]

Cozinhas:

Nao Identificado: 272

regional: 68

japanese: 17

pizza: 15

arab: 12

italian: 11

burger: 7

brazilian: 5

vegetarian: 5

international: 4

## Números gerais

1. Tamanho dos arquivos:
   1. Osm: 224,5 MB
   2. Json: 323,5 MB
2. Documentos: 1.051.798
3. Nos: 915.046
4. Ways: 136.743
5. Usuarios unicos: 419

# Sugestões e observações

1. Inicialmente acho importante que sejam disponibilizados mecanismos de divulgação dessa ferramenta, como ela pode ser usada e como contribuir, utilizo de forma constante a internet e não tinha conhecimento até este curso
   1. Benefícios
      1. Ampliação das informações disponíveis, tanto em detalhamento quanto em quantidades
   2. Problemas esperados
      1. Informações imprecisas
      2. Aumento da necessidade de validações das informações cadastradas
2. Poderia ser adicionado um tratamento para internacionalização, as facilidades, cozinha entre outros são apresentados em inglês
   1. Benefícios
      1. Ter os dados em outras línguas diferentes do inglês poderia ampliar o uso
   2. Problemas esperados
      1. Porém esta opção poderia cair em desuso ja que a grande maioria dos usuários teriam que preencher informações de forma duplicada
3. A quantidade de nós com endereço é bem pequena, seria importante o encaminhamento/fomento para o cadastro de pelo menos a cidade
   1. Benefício
      1. Facilidade de análises por cidade
4. Uma possibilidade seria separar no endereço o tipo de rua
   1. Benefício
      1. Facilidade de análises por tipo de rua
5. Um ponto importante para ressaltar é a quantidade de pontos com informações completas, ele é bem inferior ao numero total.