OpenStreetMap Data Case Study

Map Area

Balneario Camboriu, Santa Catarina, Brazil.

Link to Open Street Maps:
https://www.openstreetmap.org/export#map=12/-27.0869/-48.5616

This is the city I live in for the last 4.5 years, where my baby boy was born, so it is quite important to me. It is also a touristic city with many skyscrapers and nice turquoise sea:

However, due to the file size restrictions (being greater than 50 MBs), I had to pick a larger area that included 20 neighbour cities.

Problems encountered in your map

First problem every foreigner always find when handling data is encoding. First time running the code already got the error when running the prints for debugging:

```
UnicodeEncodeError: 'ascii' codec can't encode character u'\xed' in
position 5: ordinal not in range(128)
```

After solving that issue with UTF-8 decoding, I got started to verify the addresses.

First identified problem was repeated data: one street had several entries with different speed limits, lane numbers, etc. But the problem for writing a function to correct that is: since our section of the map contains more than one city, it is common in Brazil for many cities have the same street names, specially for famous historical figures and historical dates. Also, there are repeated building names in different cities, and since there is no “city” field, one must find all the related “nodes” to each “way” entry in order to know to which city each street belong, so I gave up of this task.

No entry had the postal code, so it is impossible to perform any search using this field.

There were only 2 entries with the name “street” abbreviated in portuguese, so I renamed those manually.

Overview of the Data

I chose MongoDB for the Database since I already worked with SQL previously. As a support to some scripts in Python, I have used the software Compass, available in MongoDB website. I also mixed way tags and node tags into one database, and saved only entries which had “name” values, since it did not made much sense to me to analyze data which contained only GPS positions.

Overall, we have the following file sizes:

```
Camboriu.osm                                  256MB
Entries on MongoDB                            37.6K
Size on MongoDB                               6.1MB
Avg size on MongoDB                           171Bytes
```

Count of Cities

So, our first task was to get a count of the field addr:city and count how many each entry appeared. For this, I wrote a function that receives the mongodb collection as a parameter:

```
def list_cities(camboriu_entries_db):
      pipeline = [
     {"$unwind": "$addr:city"},
     {"$group": {"_id": "$addr:city", "count": {"$sum": 1}}},
     {"$sort": SON([("count", -1), ("_id", -1)])}
      ]
      entries_with_cities =
list(camboriu_entries_db.aggregate(pipeline))
      for entry in entries_with_cities:
            print unicode(entry['_id']+'   ').encode('utf-8'),
            print unicode(entry['count'])
```

Our result were:

```
Jaraguá do Sul   1036
Guaramirim   70
Itapema   56
Massaranduba   44
Corupá   28
Barra Velha   24
Balneário Piçarras   22
Doutor Pedrinho   17
São João do Itaperiú   13
Blumenau   13
Itajaí   12
Balneário Camboriú   12
Pomerode   11
Schroeder   7
Penha   7
Brusque   6
Ilhota   4
Balneário Barra do Sul   4
Rio dos Cedros   3
Gaspar   3
Araquari   3
Navegantes   2
Porto Belo   1
Piçarras   1
Indaial   1
Camboriú   1
Apiúna   1
```

There is a large difference in number between the first city Jaraguá do Sul and all the rest, much more than one would expect.

Top 10 contributors

```
def top10_user_contribution(camboriu_entries_db):
      pipeline = [
     {"$unwind": "$author"},
     {"$group": {"_id": "$author", "count": {"$sum": 1}}},
     {"$sort": SON([("count", -1), ("_id", -1)])},
     {"$limit":10}
      ]
      author_contribution_list =
list(camboriu_entries_db.aggregate(pipeline))
      for contrib in author_contribution_list:
            print unicode(contrib['_id']+'  ').encode('utf-8'),
            print unicode(contrib['count'])
```

After running the code above we get the following result:

```
adrianojbr         18505
Victor 2015        8117
André Alvarenga    2925
Tomio              2686
Geomir             486
Cladimir Luis Lang 445
patodiez           443
Corujão            412
poeiradasestrelas  406
portalaventura     251
```

I guess it would not be a surprise if user adrianojbr lived in Jaragua do Sul.

Unique Users

```
def count_authors(camboriu_entries_db):
      print(len(camboriu_entries_db.distinct('author')))
```

Gives us the number of 279 unique users.

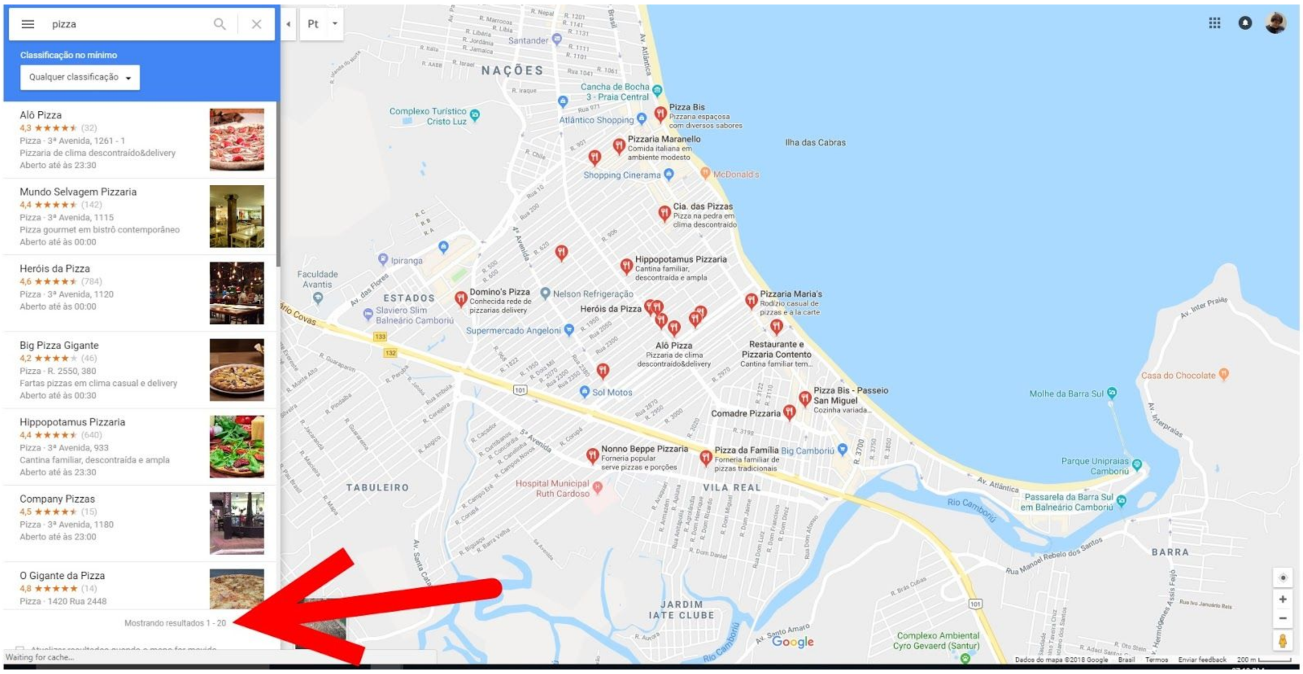
Other ideas about the datasets

Contributor Statistics

Following the suggestions from the model which I used to make this document (available in https://gist.github.com/carlward/54ec1c91b62a5f911c42#file-sample_project-md), the distribution I found on this map is also very skewed. One city has more entries than all other cities together and one user has more entries than all other user together. For this reason, gamification would definitely be a useful tool to motivate user contribution to this mapping project. Of course, as a downside, it would require more programming, maybe a major change in the OSM coding since I did not found many open-source gamification solutions which have a large community (best I could find is https://code.google.com/archive/p/userinfuser/).

Another idea would be to make a page per user contribution available for display as a public portfolio, just like github does. This could need a smaller change in terms of software development, and could motivate rookies to spend some extra hours doing volunteer work for a neat curriculum.

Last but not least, there is the possibility of importing data from other map sources such as Google maps. Let's take a look, for an example, at what our search in 27 cities brought for restaurants a few topics ahead: 46 pizza parlors. Making a quick search on my city, we have the following results:



pizza

Classificação no mínimo

Qualquer classificação

Alô Pizza
4,3 ★★★★★ (32)
Pizza · 3ª Avenida, 1261 - 1
Pizzaria de clima descontraído&delivery
Aberto até às 23:30

Mundo Selvagem Pizzaria
4,4 ★★★★★ (142)
Pizza · 3ª Avenida, 1115
Pizza gourmet em bistrô contemporâneo
Aberto até às 00:00

Heróis da Pizza
4,6 ★★★★★ (784)
Pizza · 3ª Avenida, 1120
Aberto até às 00:00

Big Pizza Gigante
4,2 ★★★★☆ (46)
Pizza · R. 2550, 380
Fartas pizzas em clima casual e delivery
Aberto até às 00:30

Hippopotamus Pizzaria
4,4 ★★★★★ (640)
Pizza · 3ª Avenida, 933
Cantina familiar, descontraída e ampla
Aberto até às 23:30

Company Pizzas
4,5 ★★★★★ (15)
Pizza · 3ª Avenida, 1180
Aberto até às 23:00

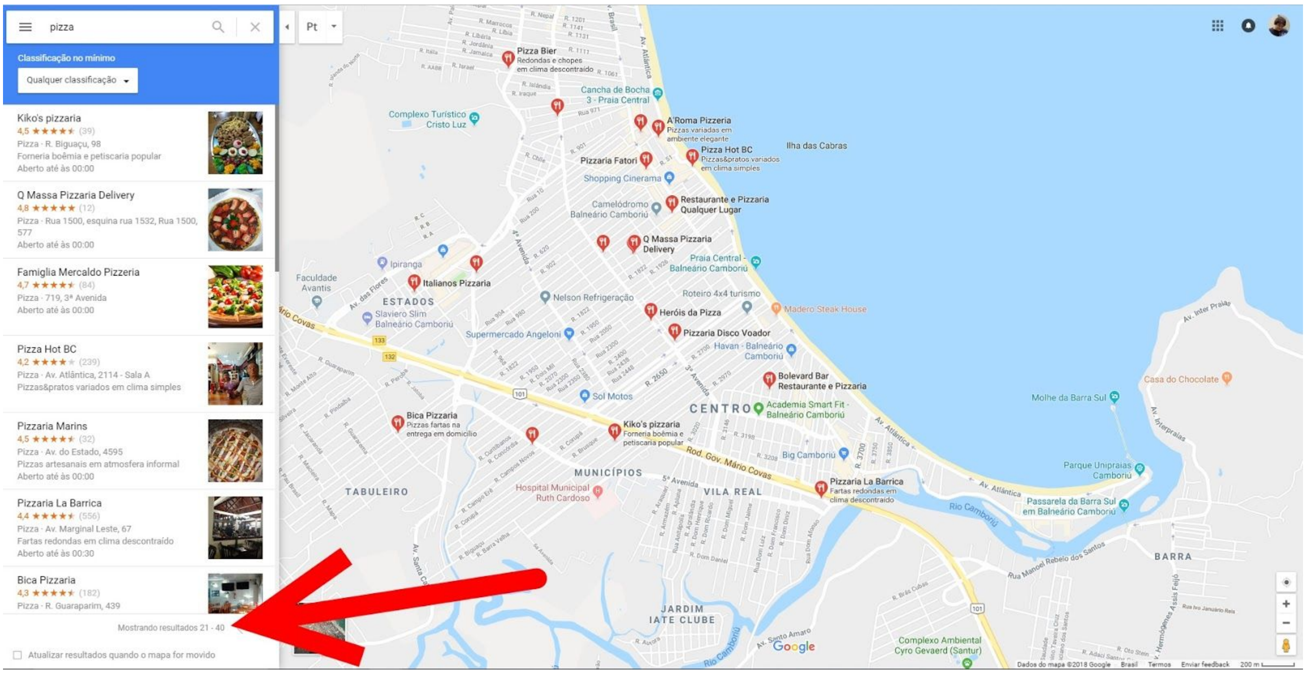
O Gigante da Pizza
4,8 ★★★★★ (14)
Pizza · 1420 Rua 2448

Mostrando resultados 1 - 20

Waiting for cache...

Pt
NAÇÕES
Santander
Cancha de Bocha 3 - Praia Central
Complexo Turístico Cristo Luz
Atlântico Shopping
Pizza Bis
Pizzaria espaçosa com diversos sabores
Pizzaria Maranello
Comida italiana em ambiente modesto
Ilha das Cabras
Shopping Cinerama
McDonald's
Cia. das Pizzas
Pizza na pedra em clima descontraído
Ipiranga
Hippopotamus Pizzaria
Cantina familiar, descontraída e ampla
Faculdade Avantis
ESTADOS
Slaviero Slim Balneário Camboriú
Domino's Pizza
Conhecida rede de pizzarias delivery
Nelson Refrigeração
Heróis da Pizza
Pizzaria Maria's
Rodízio casual de pizzas e à la carte
Supermercado Angeloni
Alô Pizza
Pizzaria de clima descontraído&delivery
Restaurante e Pizzaria Contento
Cantina familiar tem...
Casa do Chocolate
Sol Motos
Pizza Bis - Passeio San Miguel
Cozinha variada...
Molhe da Barra Sul
Comadre Pizzaria
Nonno Beppe Pizzaria
Fornerìa popular serve pizzas e porções
Pizza da Família Big Camboriú
Fornerìa familiar de pizzas tradicionais
Parque Unipraias Camboriú
TABULEIRO
Hospital Municipal Ruth Cardoso
VILA REAL
Rio Camboriú
Passarela da Barra Sul em Balneário Camboriú
BARRA
JARDIM IATE CLUBE
Google
Complexo Ambiental Cyro Gevaerd (Santur)
Dados do mapa ©2018 Google Brasil Termos Enviar feedback 200 m

Image 1: displaying results 1 - 20



pizza

Classificação no mínimo

Qualquer classificação

Kiko's pizzaria
4,5 ★★★★★ (39)
Pizza · R. Biguaçu, 98
Fornería boêmia e petiscaria popular
Aberto até às 00:00

Q Massa Pizzaria Delivery
4,8 ★★★★★ (12)
Pizza · Rua 1500, esquina rua 1532, Rua 1500, 577
Aberto até às 00:00

Famiglia Mercaldo Pizzeria
4,7 ★★★★★ (84)
Pizza · 719, 3ª Avenida
Aberto até às 00:00

Pizza Hot BC
4,2 ★★★★★ (239)
Pizza · Av. Atlântica, 2114 - Sala A
Pizzas&pratos variados em clima simples

Pizzaria Marins
4,5 ★★★★★ (32)
Pizza · Av. do Estado, 4595
Pizzas artesanais em atmosfera informal
Aberto até às 00:00

Pizzaria La Barrica
4,4 ★★★★★ (556)
Pizza · Av. Marginal Leste, 67
Fartas redondas em clima descontraído
Aberto até às 00:30

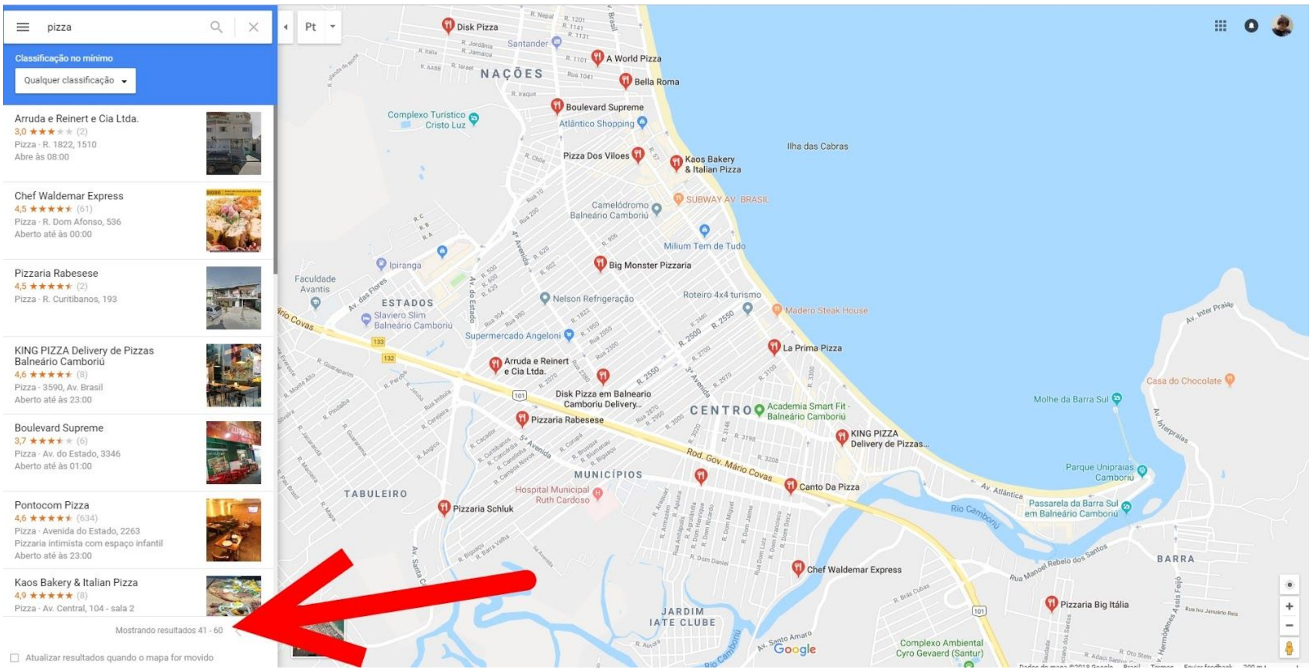
Bica Pizzaria
4,3 ★★★★★ (182)
Pizza · R. Guaraparim, 439

Mostrando resultados 21 - 40

□ Atualizar resultados quando o mapa for movido

Pt
Pizza Bier
Redondas e chopes em clima descontraído
Cancha de Bocha 3 - Praia Central
Complexo Turístico Cristo Luz
A'Roma Pizzeria
Pizzas variadas em ambiente elegante
Pizza Hot BC
Pizzas&pratos variados em clima simples
Ilha das Cabras
Pizzaria Fatori
Shopping Cinerama
Camelódromo Balneário Camboriú
Restaurante e Pizzaria Qualquer Lugar
Q Massa Pizzaria Delivery
Praia Central - Balneário Camboriú
Ipiranga
Italianos Pizzaria
Faculdade Avantis
ESTADOS
Slaviero Slim Balneário Camboriú
Nelson Refrigeração
Roteiro 4x4 turismo
Madero Steak House
Heróis da Pizza
Supermercado Angeloni
Pizzaria Disco Voador
Havan - Balneário Camboriú
Bolevard Bar Restaurante e Pizzaria
Sol Motos
CENTRO
Academia Smart Fit - Balneário Camboriú
Casa do Chocolate
Molhe da Barra Sul
Bica Pizzaria
Pizzas fartas na entrega em domicílio
Kiko's pizzaria
Fornería boêmia e petiscaria popular
Big Camboriú
Rod. Gov. Mário Covas
MUNICÍPIOS
TABULEIRO
Hospital Municipal Ruth Cardoso
VILA REAL
Pizzaria La Barrica
Fartas redondas em clima descontraído
Parque Unipraias Camboriú
Passarela da Barra Sul em Balneário Camboriú
Rio Camboriú
Av. Atlântica
BARRA
JARDIM IATE CLUBE
Google
Complexo Ambiental Cyro Gevaerd (Santur)
Dados do mapa ©2018 Google
Brasil
Termos
Enviar feedback
200 m

Image 2: displaying results 21 - 40



pizza

Classificação no mínimo

Qualquer classificação

Arruda e Reinert e Cia Ltda.
3,0 ★★★★★ (2)
Pizza · R. 1822, 1510
Abre às 08:00

Chef Waldemar Express
4,5 ★★★★★ (61)
Pizza · R. Dom Afonso, 536
Aberto até às 00:00

Pizzaria Rabesese
4,5 ★★★★★ (2)
Pizza · R. Curitibanos, 193

KING PIZZA Delivery de Pizzas Balneário Camboriú
4,6 ★★★★★ (8)
Pizza · 3590, Av. Brasil
Aberto até às 23:00

Boulevard Supreme
3,7 ★★★★★ (6)
Pizza · Av. do Estado, 3346
Aberto até às 01:00

Pontocom Pizza
4,6 ★★★★★ (634)
Pizza · Avenida do Estado, 2263
Pizzaria intimista com espaço infantil
Aberto até às 23:00

Kaos Bakery & Italian Pizza
4,9 ★★★★★ (8)
Pizza · Av. Central, 104 - sala 2

Mostrando resultados 41 - 60

□ Atualizar resultados quando o mapa for movido

Image 3: displaying results 41 - 60

Those indicate that only in Balneario Camboriu there are more than 60 different pizza restaurants (Link available here for reference: https://www.google.com.br/maps/search/pizza/@-26.9971824,-48.6338856,15z/data=!3m1!4b1), and the maps api is also available for usage on python (link: https://github.com/googlemaps/google-maps-services-python) . Main problem with googlemaps is the limit of accesses for non-paying users: this would still require that many users perform the task of updating OSM.

Additional Data Exploration

Top 10 Amenities

```
def top10_amenities(camboriu_entries_db):
      pipeline = [
     {"$unwind": "$amenity"},
     {"$group": {"_id": "$amenity", "count": {"$sum": 1}}},
     {"$sort": SON([("count", -1), ("_id", -1)])},
     {"$limit":10}
      ]
      author_contribution_list =
list(camboriu_entries_db.aggregate(pipeline))
      for contrib in author_contribution_list:
            print unicode(contrib['_id']+'  ').encode('utf-8'),
            print unicode(contrib['count'])
```

```
restaurant   406
place_of_worship   391
school   240
fuel   212
fast_food   207
clinic   206
pharmacy   186
community_centre   141
bank   138
parking   123
```

Top 3 Religions

```
def top3_religion(camboriu_entries_db):
      pipeline = [
     {"$unwind": "$religion"},
     {"$group": {"_id": "$religion", "count": {"$sum": 1}}},
     {"$sort": SON([("count", -1), ("_id", -1)])},
     {"$limit":3}
]
      author_contribution_list =
list(camboriu_entries_db.aggregate(pipeline))
      for contrib in author_contribution_list:
            print unicode(contrib['_id']+'  ').encode('utf-8'),
            print unicode(contrib['count'])
```

```
christian   416
spiritualist   3
umbanda   1
```

Top 10 Cuisines

```
def top10_cuisines(camboriu_entries_db):
      pipeline = [
     {"$unwind": "$cuisine"},
     {"$group": {"_id": "$cuisine", "count": {"$sum": 1}}},
     {"$sort": SON([("count", -1), ("_id", -1)])},
     {"$limit":10}
      ]
      author_contribution_list =
list(camboriu_entries_db.aggregate(pipeline))
      for contrib in author_contribution_list:
            print unicode(contrib['_id']+'  ').encode('utf-8'),
            print unicode(contrib['count'])
```

```
pizza   46
burger   45
regional   31
steak_house   18
sandwich   15
empanada   14
japanese   11
sushi   9
italian   9
sausage   7
```

Conclusion

Given the size of the area in square kilometers and population, it would be expected to have much more data available, but that was not the case. We noticed that the vast majority of work seems to have been done by one single user, and given the open source nature of the website (openmap), that does appears to be the truth. There is in fact a structural problem in order to feed precise data to openstreetmap because even government websites do not provide nice databases to the public as of 2018.