**OpenStreetMap Project**

**Content**

[1. Map area 2](#_Toc502934500)

[2. Issues in the map 3](#_Toc502934501)

[3. Data Overview 4](#_Toc502934502)

[4. Conclusion 7](#_Toc502934503)

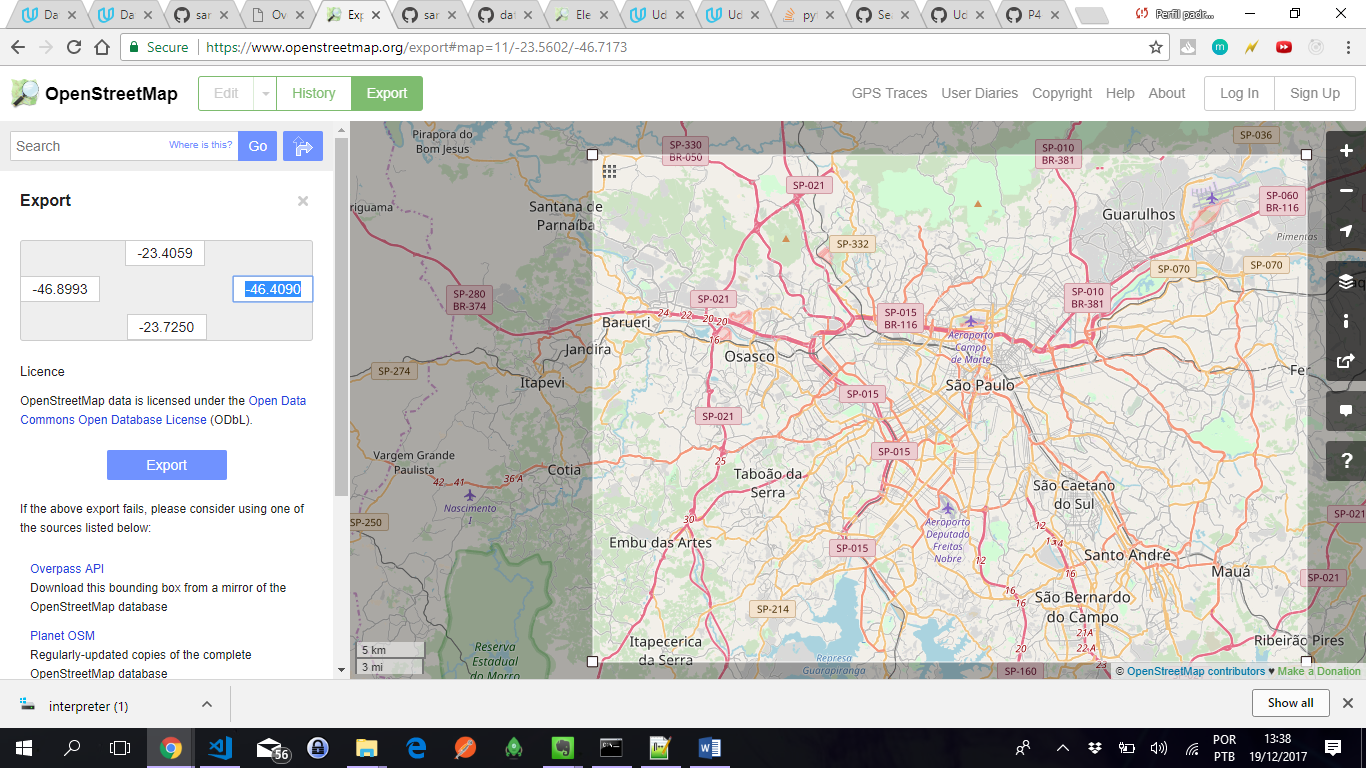
[5. References 9](#_Toc502934504)

# Map area

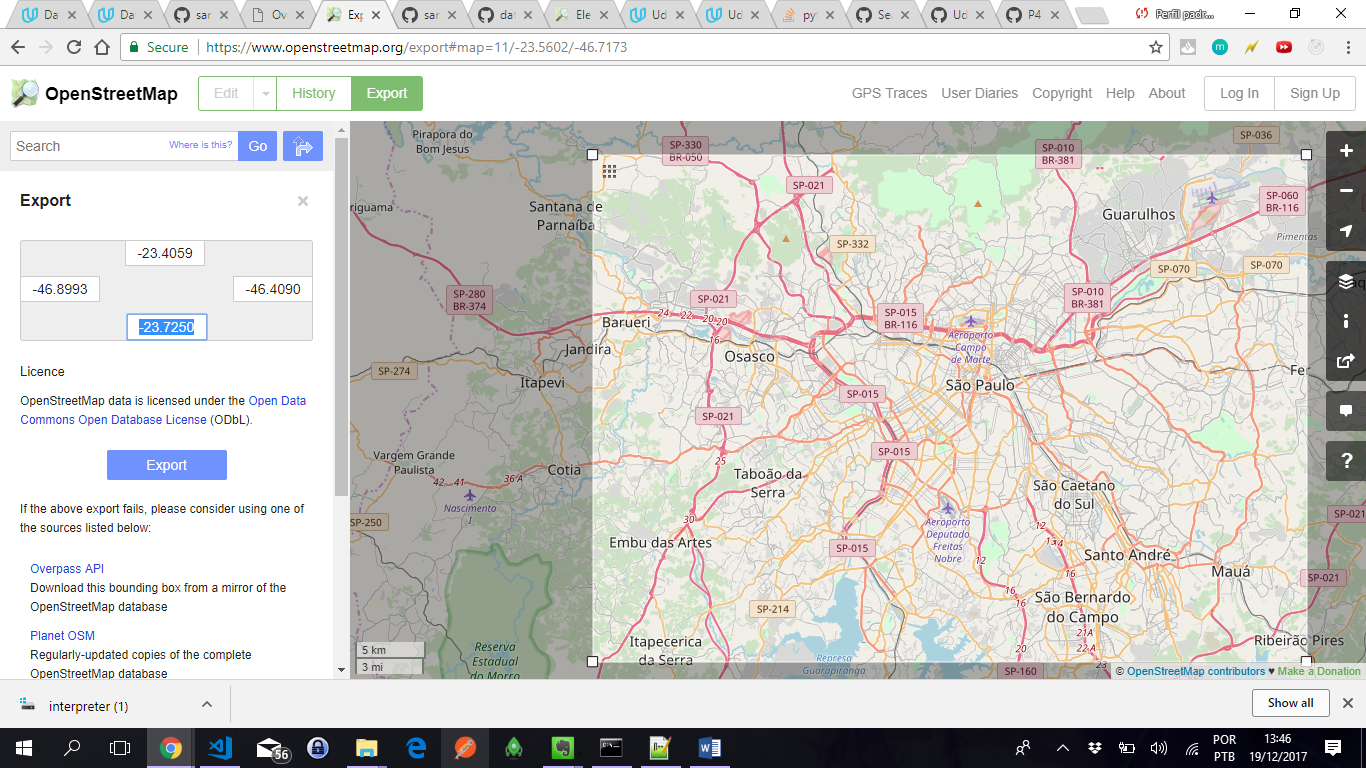
São Paulo, SP, Brazil.

São Paulo central area was the sample for this analyze and its coordinates is minimum latitude: -23.5595, minimum longitude: -46.6534, maximum latitude: -23.5375, maximun longitude: -46.6202

Size file:  39293 KB.



São Paulo complete map has this coordinates: minimum latitude: -23.725, minimum longitude: -46.8993, maximum latitude: -23.4059, maximum longitude: -46.4090.



The city where I live does not have enough content to analyze, and then, I decide for São Paulo because there were stores enough for a good understanding, and analyzing.

# Issues in the map

This section contain information about the issues found during analyze.

Based on the sample file, it was found these issues:

* Overabbreviated and misspelled name:

The map has 46923 streets and 98 errors, less than 1% of errors. Some of them was abbreviation error, and it was fixed mapping the errors and their correct value.

MAPSTREET = {

"Av":"Avenida",

"Estr":"Estrada",

"Pç":"Praça",

"Al":"Alameda",

"Alamedas":"Alameda",

"Rue":"Rua",

"Rúa":"Rua",

"R":"Rua"

}

After the fixes, the errors felt down to 65.

* Postal code inconsistence

The audit method returned postcode did not follow the Brazilian pattern: 5 numbers plus dash plus 3 numbers

Some of the postcode returned for the audit method as error was possible to fix. The number of digits was correct but without the dash. For cases like these, it was developed a method to fix it. Follow an example for this issue:

  <node id="3063117943" lat="-23.5568747" lon="-46.6498269" version="1" timestamp="2014-09-06T14:07:36Z" changeset="25267529" uid="2315838" user="Ana Paula1608">

    ...

    <tag k="addr:postcode" v="01330000"/>

    ...

  </node>

Follow the routine to fix these cases. When it returns None, the value is not insert in the csv file.

if is\_postalcode(kvalue):

if auditory\_postalcode(value) == False:

value = fix\_postalcode(value)

if value == None:

continue

def fix\_postalcode(value):

"""Return the postal code when it is possible to fix or None when it is not"""

if len(value) == 8:

nvalue = value[0:5] +'-'+value[5:]

# verify if the new code is valid

if auditory\_postalcode(nvalue):

return nvalue

else:

return None

else:

return None

# Data Overview

This section contains information about the files, basic statistics about the dataset and queries used to gather them.

File size:

sample.osm: 392MB

SaoPaulo.osm: 849MB

Nodes.csv 314MB

nodes\_tags 11MB

Ways.csv 32MB

ways\_nodes.csv 122MB

ways\_tags.csv 49MB

Number of nodes select count(\*) from nodes; 3653282

Number of ways select count(\*) from ways; 529372

Number of unique users: 1865

select count(distinct(uid)) from (select uid from nodes union select uid usway from ways);

Number of traffic\_signals: 6618

select (count(\*) + qtd) As qtdtrafficsignals from nodes nd, (select count(\*) as qtd from ways wy inner join ways\_tags wt on wy.id = wt.id where wt.VALUE = 'traffic\_signals') inner join nodes\_tags nt on nd.id = nt.id where nt.VALUE = 'traffic\_signals';

Number of positions you can cycle: 2591

select (count(\*) + qtd) As qtdcicleplace from nodes nd, (select count(\*) as qtd from ways wy inner join ways\_tags wt on wy.id = wt.id where wt.value = 'yes' and wt.key='cicle') inner join nodes\_tags nt on nd.id = nt.id where nt.VALUE = 'yes' and nt.key='cicle';

Number of nodes with tags: 131783

SELECT COUNT(DISTINCT(ID)) FROM NODES\_TAGS;

Most 5 popular cuisine

SELECT tags.value AS CUISINE, COUNT(\*) as count FROM (SELECT \* FROM nodes\_tags where nodes\_tags.key='cuisine' UNION ALL SELECT \* FROM ways\_tags where ways\_tags.key='cuisine') tags WHERE tags.key='cuisine' GROUP BY tags.value ORDER BY count DESC limit 5;

CUISINE | count

regional|254

burger|172

pizza|136

japanese|62

sandwich|56

Rank of the top 5 neighborhood by the amount of pharmacies

SELECT tags.'value:1',count(\*) as qtd

FROM (

SELECT \*

FROM nodes\_tags nt

INNER JOIN (

SELECT nt1.id, nt1.value

FROM nodes\_tags nt1 WHERE nt1.key = 'suburb'

) nt2 on nt2.id = nt.id

UNION ALL

SELECT \*

FROM ways\_tags wt

INNER JOIN (

SELECT wt1.id, wt1.value

FROM ways\_tags wt1 WHERE wt1.key = 'suburb'

) wt2 on wt2.id = wt.id

) tags where

tags.key = 'amenity'

and tags.value = 'pharmacy'

group by tags.'value:1'

order by qtd desc

limit 5;

Results:

Neighborhood|Qtd

Centro|17

Piraporinha|5

Planalto|5

Assunção|4

Rudge Ramos|4

The top 10 contributors

SELECT USER,COUNT(\*) FROM NODES GROUP BY USER ORDER BY COUNT(\*) DESC LIMIT 10

Bonix-Mapper|2454573

Bonix-Importer|434301

AjBelnuovo|135497

cxs|73426

O Fim|72681

ygorre|46031

MCPicoli|33776

naoliv|28221

D4RK-L3G10N|28119

patodiez|24726

# Conclusion

The most important asset for the OpenStreetMap is its data and it get the data from the user contribution. People need to be engaged with OpenStreetMap project to become users from the project and start to contribute.

The chart 1 shows the number of registered users by date. Until 2009, there were less than 200,000 users and in 2017, there are more than 4,200,000 users. The last 8 years, the amount of users raised at least 20 times.

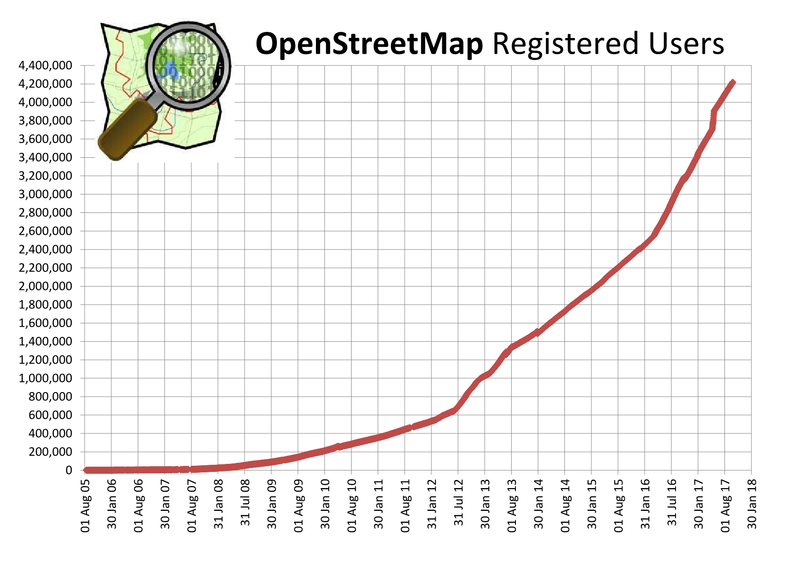


Chart 1 – This chart was found at the <http://wiki.openstreetmap.org/wiki/Stats> internet address.

When I analyzed Sao Paulo map, there were 3653282 nodes in the XML file, and there are 131783 nodes with tags, which represents 3.60% of nodes with tags. There are 1865 different users’ contributors and only one user inserted about 67% (2454573) of these insertions as you can see in the last chapter, the top 10 contributor.

Even the amount of users raising at all, for São Paulo city, the amount of users represents less than 0.04%. It is necessary to have more users knowing and engaged with this project to increase the amount of data about São Paulo area. My suggestion to increase the amount of user are:

* Work with universities project: the students could make researches and add data to map, like for example, map the place in Sao Paulo where there are access to wheelchairs;
* Encourage users to invite others users;
* Gamification: Create a reputation strategy for users who most contribute with the community;
* Mobile application: a simple interface informing a list of changes needed around the place where the device is located.

The map data have to be trustworthy. However, the challenge is that the users who add data to the map does not need to have training to add information in it and consequently the increase of users can increase the amount of mistakes. My suggestion to avoid mistake are:

* Software with simple interface and easy to work. For example, if the user add a name street for a node, make it easier to add the same name for the other node as well.
* The reputation strategy could allow users with good reputation be more trustworthy and verify when users who initial reputation are adding places correctly.
* Identification and exclusion of people who add bad data or remove correct data.
* Suggestion for street names when the user is typing the street name.
* Cross data added to the map with others maps.

The implementation from the suggestion are going to have benefits as well as some challenges. Analyzing each suggestion, the benefits and challenges are:

* Work with universities project:

Benefits

* + Many users can have interest for this subject and continue the project even after the graduation;
  + Students can influence others students;
  + Many startup can show up based on these projects.

Challenges:

* + How to influence the universities professors to develop projects in this area;
* Encourage users to invite others users:

Benefits

* + Usually, the users are going to invite the people who believe they are going to engage with the project. Therefore, the interest rate for the project can be higher than some marketing strategy.

Challenges

* + How to create motivation for the people who receive the invite to contribute to the map.
* Gamification:

Benefits

* + Gamification usually increase engagement and enthusiasm. They are going to do their best to reach the best.

Challenges

* + The rewards that call the attention of the people to continue adding places and streets name to the map.
* Mobile application:

Benefits

* + People are going to have interest for the project if they find easy tool to contribute.

In conclusion, the challenge with worldwide project is how to involve so many people and how to guarantee the integrity of the information. São Paulo is a big city where people really need maps to localize places and address. However, the most part of the Brazilian cities are small and people does not need maps to localize stores and streets. The people know where the streets and the places are. Nevertheless, these cities should have to be included to the map, as well their stores.

# References

<https://github.com/SamMorrowDrums/Udacity-OpenStreetmap/blob/master/submission.pdf>

<https://github.com/ollkorrect/P4-Streetmap-Data-Wrangling/blob/master/Report.pdf>

<https://gist.github.com/carlward/54ec1c91b62a5f911c42#file-sample_project-md>

<https://www.openstreetmap.org/stats/data_stats.html>

<http://wiki.openstreetmap.org/wiki/Stats>