Applied Data Mining

1st part – Preliminary Data Analysis, SQL queries

Alevizopoulou Sofia 2022201704002

Avgeros Giannis 2022201704003

Tsiatsios George 2022201704024

Athens 2018

Contents

[Entity Relation model 3](#_Toc528573827)

[Create tables 4](#_Toc528573828)

[We have added all the above tables at the database. 4](#_Toc528573829)

[Preliminary data analytics 4](#_Toc528573830)

[SQL commands: 4](#_Toc528573831)

[How many ships there are based on shipname? 4](#_Toc528573832)

[How many ships have imo number? 4](#_Toc528573833)

[How many ships there are based on mmsi( dynamic table)? 4](#_Toc528573834)

[How many ships there are based on mmsi (static table)? 4](#_Toc528573835)

[How many ships have mouthermmsi (ships that other ships use them as start point)? 4](#_Toc528573836)

[Display 5 ships that have mouthermmsi? 5](#_Toc528573837)

[Display the ship which is the ship that defined as mothershipmmsi ship most of the times. 5](#_Toc528573838)

[How many ship types there are at the static table? 6](#_Toc528573839)

[For each ship type how many vessels there are? 6](#_Toc528573840)

[Find the shipnames that have no ship type defined 6](#_Toc528573841)

[Display the number of vessels per country. The first 3 digits of MMSI declare the region country of the ship. 7](#_Toc528573842)

[Display the number of ships and their type per country 7](#_Toc528573843)

[Find the ship types for French vessels. 8](#_Toc528573844)

[Display the most popular destinations 9](#_Toc528573845)

[Display the number of ships based on ship type 9](#_Toc528573846)

[Display the type of messages that are sent from ship with shiptype=0. 9](#_Toc528573847)

[Display the number of messages that are sent from ships based on their type. 10](#_Toc528573848)

[Display the number of fishing messages per month 10](#_Toc528573849)

[Display the number of messages per month 11](#_Toc528573850)

[Display the types of messages inside a fishing area 11](#_Toc528573851)

[Display the number of fishing messages inside a fishing area 12](#_Toc528573852)

[Display the number of ships per type that send fishing messages 13](#_Toc528573853)

[Display the months with increased fishing activities based on ships status. 13](#_Toc528573854)

[Display the traffic per month 14](#_Toc528573855)

[Average speed per Month 14](#_Toc528573856)

[Display average speed, draught, length and width per ship type 14](#_Toc528573857)

[Traffic of Brest Port based on total trafffic 14](#_Toc528573858)

[Average of draught,width per month for each ship type 14](#_Toc528573859)

[Usage of SAR vessels every month 15](#_Toc528573860)

[Type of atons 15](#_Toc528573861)

[Make a plot for a specific ship according its ais messages 15](#_Toc528573862)

[Plot constrain fishing areas 15](#_Toc528573863)

[Plot fishing areas 16](#_Toc528573864)

[Plot natura areas 17](#_Toc528573865)

[Plot ports of brittany 17](#_Toc528573866)

[Plot the ports of the europe 17](#_Toc528573867)

[Postgres version 18](#_Toc528573868)

# Entity Relation model

The tables that we are going to be used in this assignment are listed below. They contain real-world data from the maritime domain [https://zenodo.org/record/1167595#.W9BgcFUzapo] which monitor, analyze and visualize the sea movements. There are tables that contain info of the whole maritime activities and their impact on the environment, tables that contain a set of complementary data having spatial and temporal information and tables with information about ships positions within Celtic sea, the Channel and Bay of Biscay (France).

More specifically, there are four categories of data: Navigation data, vessel-oriented data, geographic data, and environmental data. It covers a time span of six months, from October 1st, 2015 to March 31st, 2016.

Tables are:

* aton
* pg\_catalog
* country\_codes
* pg\_temp\_1
* geographic\_features
* pg\_toast.
* geography\_columns
* pg\_toast\_temp\_1
* geometry\_columns
* ports.
* information\_schema
* public.
* raster\_overviews
* receiver
* ship\_types\_detailled\_list
* natura2000
* ship\_types\_list
* navigational\_status
* spatial\_ref\_sys
* raster\_columns
* nari\_ais\_static
* nari\_dynamic
* nari\_dynamic\_aton
* nari\_dynamic\_sar

# Create tables

# We have added all the above tables at the database.

# Preliminary data analytics

Preliminary data analysis refers to some simple and basic analysis by running some SQL queries. This analysis will helpful in understanding the data.

Note : we have memory issues so most of the queries have been executed with LIMIT.

## SQL commands:

### How many ships there are based on shipname?

***SELECT COUNT (DISTINCT shipname) FROM nari\_ais\_static;***

Count: 4824

Note: Extract this info from static information about ships

### How many ships have imo number?

***SELECT COUNT (DISTINCT imo) FROM nari\_ais\_static;***

Count: 4033

### How many ships there are based on mmsi( dynamic table)?

***SELECT COUNT (DISTINCT mmsi) FROM nari\_dynamic;***

Count: 5055

Note: Extract this info from dynamic information about ships (ais messages).

Here we can see that there are ships which send ais messages but there is no info about them at the static table of ships (nari\_static)

### How many ships there are based on mmsi (static table)?

***SELECT COUNT (DISTINCT sourcemmsi) FROM nari\_ais\_static;***

Count: 4842

### How many ships have mouthermmsi (ships that other ships use them as start point)?

***SELECT COUNT (DISTINCT mothershipmmsi) FROM nari\_ais\_static;***

Count: 228

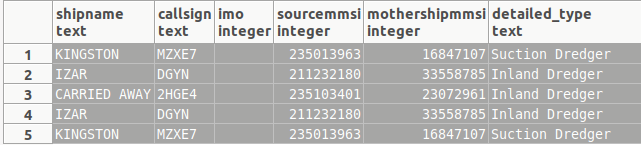
### Display 5 ships that have mouthermmsi?

***SELECT shipname, callsign, imo, sourcemmsi, mothershipmmsi,detailed\_type***

***FROM public.nari\_ais\_static, public.ship\_types\_list, public.ship\_types\_detailled\_list***

***WHERE nari\_ais\_static.shiptype = ship\_types\_detailled\_list.id\_detailedtype AND***

***ship\_types\_detailled\_list.id\_shiptype = ship\_types\_list.id\_shiptype AND nari\_ais\_static.mothershipmmsi>0 LIMIT 5;***



### Display the ship which is the ship that defined as mothershipmmsi ship most of the times.

***SELECT mothershipmmsi, count(DISTINCT sourcemmsi) AS num FROM nari\_ais\_static WHERE mothershipmmsi>0 GROUP BY mothershipmmsi ORDER BY num DESC;***

Mothershipmmsi | count

6320258 | 6

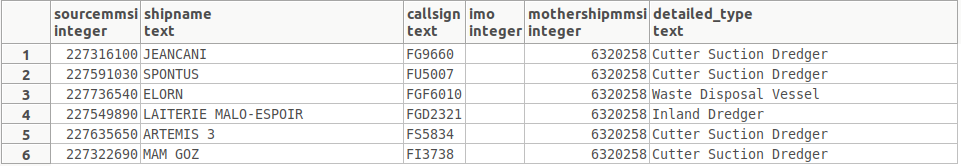
4223106 | 5

***SELECT DISTINCT sourcemmsi, shipname, callsign, imo, mothershipmmsi,detailed\_type***

***FROM public.nari\_ais\_static, public.ship\_types\_list, public.ship\_types\_detailled\_list***

***WHERE nari\_ais\_static.shiptype = ship\_types\_detailled\_list.id\_detailedtype AND***

***ship\_types\_detailled\_list.id\_shiptype = ship\_types\_list.id\_shiptype AND nari\_ais\_static.mothershipmmsi=6320258;***



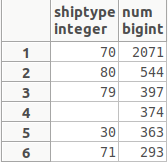
### How many ship types there are at the static table?

***SELECT COUNT (DISTINCT shiptype) FROM nari\_ais\_static;***

Count: 45

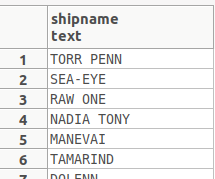
### For each ship type how many vessels there are?

***SELECT shiptype, count(DISTINCT shipname) AS num FROM nari\_ais\_static GROUP BY shiptype ORDER BY num DESC;***



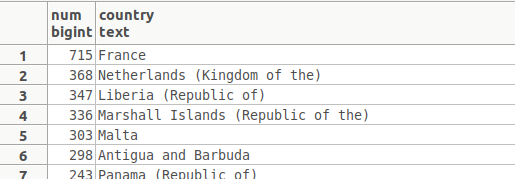
### Find the shipnames that have no ship type defined

***SELECT DISTINCT shipname FROM nari\_ais\_static WHERE nari\_ais\_static.shiptype is null;***



### Display the number of vessels per country. The first 3 digits of MMSI declare the region country of the ship.

***SELECT count(DISTINCT sourcemmsi),country FROM public.country\_codes, public.nari\_ais\_static WHERE LEFT(nari\_ais\_static.sourcemmsi::text,3)::integer = country\_codes.mmsi\_country\_code group by country;***



### Display the number of ships and their type per country

***SELECT count(Distinct(sourcemmsi)),country,ais\_type\_summary***

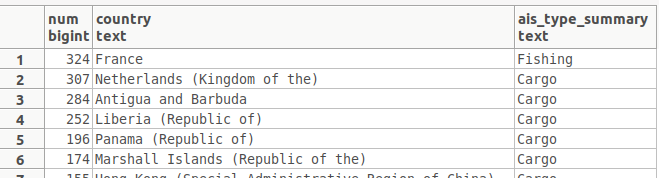
***FROM public.country\_codes, public.nari\_ais\_static, public.ship\_types\_list***

***WHERE LEFT(nari\_ais\_static.sourcemmsi::text,3)::integer = country\_codes.mmsi\_country\_code AND***

***ship\_types\_list.shiptype\_min<=nari\_ais\_static.shiptype AND***

***ship\_types\_list.shiptype\_max>=nari\_ais\_static.shiptype***

***group by country,ais\_type\_summary;***



### Find the ship types for French vessels.

***SELECT DISTINCT mmsi\_country\_codeFROM public.country\_codes***

***WHERE country LIKE 'France';***

Note: mmsi codes for France are: 226, 227,228

***SELECT sourcemmsi, shipname, imo, mothershipmmsi,detailed\_type***

***SELECT shiptype,detailed\_type,COUNT(distinct shipname)***

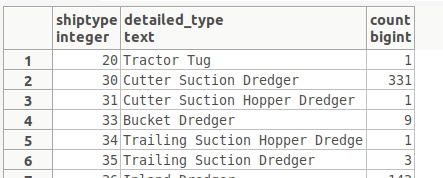
***FROM public.nari\_ais\_static, public.ship\_types\_list, public.ship\_types\_detailled\_list***

***WHERE nari\_ais\_static.shiptype = ship\_types\_detailled\_list.id\_detailedtype AND***

***ship\_types\_detailled\_list.id\_shiptype = ship\_types\_list.id\_shiptype AND (LEFT(nari\_ais\_static.sourcemmsi::text,3)::integer=226 OR***

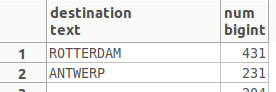
***LEFT(nari\_ais\_static.sourcemmsi::text,3)::integer=227 OR LEFT(nari\_ais\_static.sourcemmsi::text,3)::integer=228)***

***GROUP BY shiptype,detailed\_type ;***



### Display the most popular destinations

***SELECT destination, count(DISTINCT sourcemmsi) AS num FROM nari\_ais\_static WHERE length(destination)>0 GROUP BY destination ORDER BY num DESC;***



### Display the number of ships based on ship type

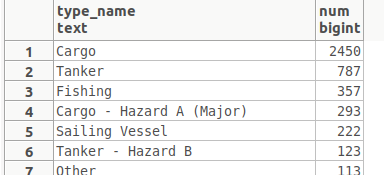
***SELECT type\_name, count(DISTINCT sourcemmsi) as num***

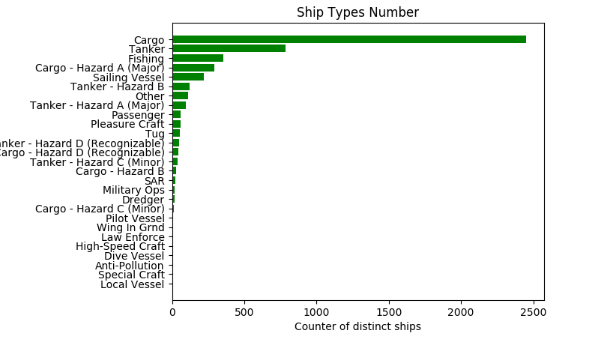
***FROM public.nari\_ais\_static, public.ship\_types\_list***

***WHERE ship\_types\_list.shiptype\_min<=nari\_ais\_static.shiptype AND***

***ship\_types\_list.shiptype\_max>=nari\_ais\_static.shiptype***

***GROUP BY type\_name ORDER BY num DESC ;***





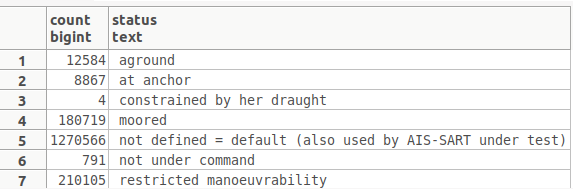
### Display the type of messages that are sent from ship with shiptype=0.

***SELECT count(\*),n.status FROM***

***(SELECT mmsi,status FROM public.nari\_dynamic) a,***

***(SELECT distinct(sourcemmsi),shiptype from nari\_ais\_static) b,***

***navigational\_status as n mmsi=sourcemmsi and b.shiptype=0 and n.code=a.status group by n.status;***



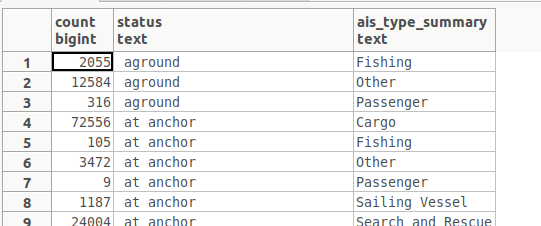
### Display the number of messages that are sent from ships based on their type.

***SELECT count(\*),n.status,shl.ais\_type\_summary FROM***

***(SELECT mmsi,status FROM public.nari\_dynamic) a,***

***(SELECT distinct(sourcemmsi),shiptype from nari\_ais\_static) b,***

***ship\_types\_list as shl, navigational\_status as n where mmsi=sourcemmsi and shl.shiptype\_min<=b.shiptype AND shl.shiptype\_max>=b.shiptype and b.shiptype!=0 and n.code=a.status group by n.status,shl.ais\_type\_summary;***

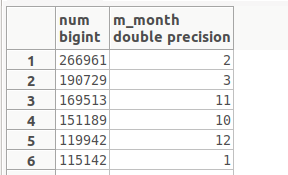


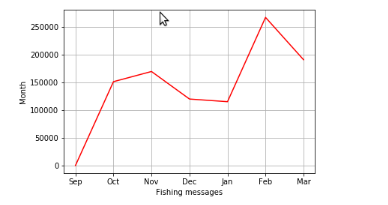
### Display the number of fishing messages per month

***select count(\*) as num,m\_month***

***FROM (select status,date\_part('month',TIMESTAMP 'epoch' + t \* INTERVAL '1 second') as m\_month from nari\_dynamic) m***

***where m.status=7 group by m\_month order by num DESC;***



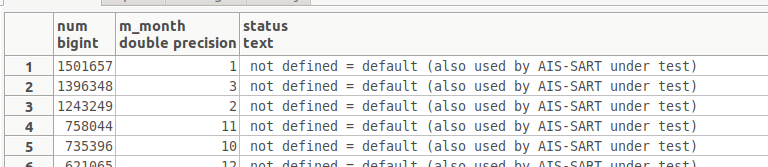


### Display the number of messages per month

***select count(\*) as num,m\_month,n.status***

***FROM (select status,date\_part('month',TIMESTAMP 'epoch' + t \* INTERVAL '1 second') as m\_month from nari\_dynamic) m,***

***navigational\_status as n where m.status=n.code group by m\_month,n.status order by num DESC;***



***Kanena plot me arithmo mnm ana mhna***

### Display the types of messages inside a fishing area

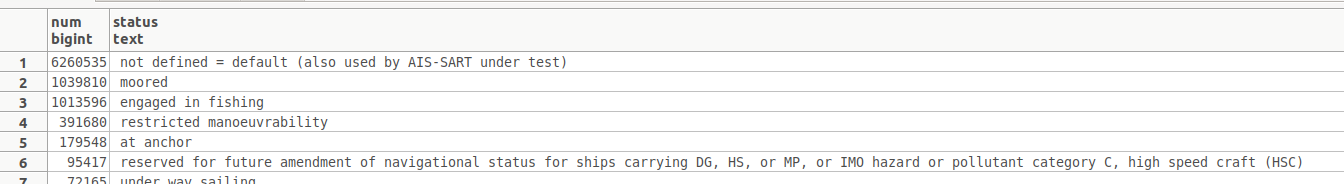
***SELECT count(\*) as num,n.status***

***FROM (select distinct maxlong,maxlat,minlong,minlat from geographic\_features.fishing\_areas\_eu) g,***

***(select lat,lon,status from nari\_dynamic) d,***

***navigational\_status as n where maxlong>=lon and minlong<=lon***

***and maxlat>=lat and minlat<=lat and n.code=d.status group by n.status order by num DESC;***



### Display the number of fishing messages inside a fishing area

***SELECT count(\*)***

***FROM (select distinct maxlong,maxlat,minlong,minlat from geographic\_features.fishing\_areas\_eu) g,***

***(select lat,lon,status from nari\_dynamic) d,***

***navigational\_status as n where maxlong>=lon and minlong<=lon***

***and maxlat>=lat and minlat<=lat and n.code=d.status and n.code=7;***

Count: 1013596

Whereas the total number of messages inside this area is:

***SELECT count(\*)***

***FROM (select distinct maxlong,maxlat,minlong,minlat from geographic\_features.fishing\_areas\_eu) g,***

***(select lat,lon,status from nari\_dynamic) d,***

***navigational\_status as n where maxlong>=lon and minlong<=lon***

***and maxlat>=lat and minlat<=lat and n.code=d.status;***

Count: 9092197

Note: fishing messages is 11,147977% of the total messages

### Display the number of ships per type that send fishing messages

***SELECT type\_name, count(DISTINCT sourcemmsi) as num***

***FROM public.nari\_ais\_static, public.ship\_types\_list, public.nari\_dynamic***

***WHERE ship\_types\_list.shiptype\_min<=nari\_ais\_static.shiptype AND***

***ship\_types\_list.shiptype\_max>=nari\_ais\_static.shiptype AND status=7***

***GROUP BY type\_name ORDER BY num DESC ;***

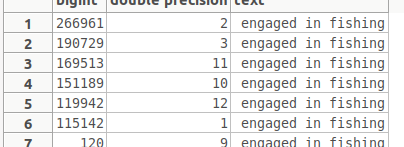
Note: out of memory

### Display the months with increased fishing activities based on ships status.

***select count(\*) as num,m\_month,n.status***

***FROM (select status,date\_part('month',TIMESTAMP 'epoch' + t \* INTERVAL '1 second') as m\_month from nari\_dynamic) m,***

***navigational\_status as n where m.status=n.code and n.code=7 group by m\_month,n.status order by num DESC;***



### Display the traffic per month

***select count(mmsi),extract(MONTH from to\_timestamp(t)) a from nari\_dynamic group by a order by count(mmsi) DESC;***

### Average speed per Month

***select avg(speed),extract(MONTH from to\_timestamp(t)) a from nari\_dynamic group by a ;***

### Display average speed, draught, length and width per ship type

***select b.shiptype,type\_name,avg(speed),avg(b.draught),avg(b.to\_stern),avg(b.to\_starboard) from (select mmsi,speed from nari\_dynamic) a,(select distinct(sourcemmsi),shiptype,draught,to\_stern,to\_starboard from nari\_ais\_static) b, (select type\_name,shiptype\_min from ship\_types\_list where shiptype\_min>0 and shiptype\_max<100) c where a.mmsi = b.sourcemmsi and c.shiptype\_min = b.shiptype group by b.shiptype,type\_name;***

### Traffic of Brest Port based on total trafffic

***SELECT COUNT(\*) from nari\_ais\_static WHERE destination LIKE '%BREST%' or destination LIKE '%BES' or destination LIKE ' ';***

### Average of draught,width per month for each ship type

***select b.shiptype,type\_name,avg(b.draught) avg\_draught,avg(b.to\_starboard) avg\_len,a.c from (select mmsi,speed,extract(MONTH from to\_timestamp(t)) c from nari\_dynamic) a,(select distinct(sourcemmsi),shiptype,draught,to\_starboard from nari\_ais\_static) b, (select type\_name,shiptype\_min from ship\_types\_list where shiptype\_min>0 and shiptype\_max<=100) c where a.mmsi = b.sourcemmsi and c.shiptype\_min = b.shiptype group by b.shiptype,type\_name,a.c,type\_name order by a.c;***

## Usage of SAR vessels every month

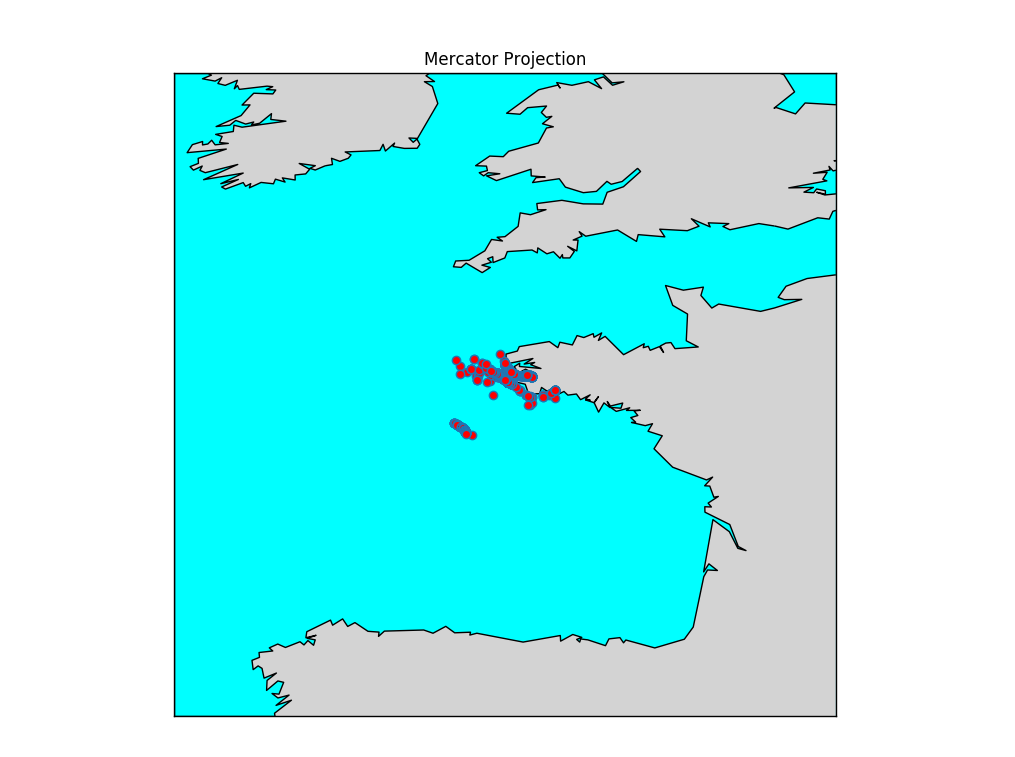
**select count(mmsi),extract(MONTH from to\_timestamp(t)) a,avg(speed) from nari\_dynamic\_sar group by a order by count(mmsi) DESC;**

### Type of atons

***select typeofaid,count(typeofaid),at.definition,virtual from nari\_dynamic\_aton a,aton at where a.typeofaid=at.code group by typeofaid,at.definition,virtual;***

### Make a plot for a specific ship according its ais messages

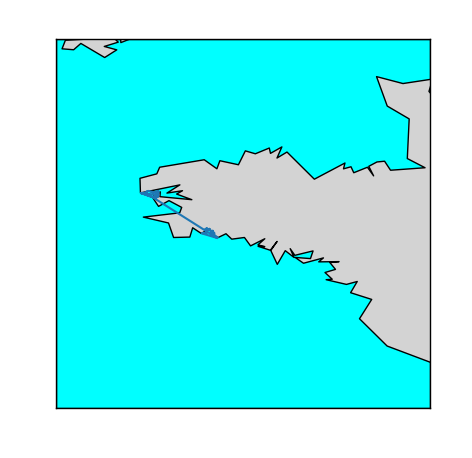
Source code: spatial.py



Na to sxoliasoume h na vroume sugkekrimena ploia p thelouem na deixouem thn kinhsh ts

### Plot constrain fishing areas

Source code: constrain\_fishing/ constrain\_fish.py



Na to sxoliasoume kai na plottaroume kati panw se auto

### Plot fishing areas

Source code:fishing\_area/ fishing.py



Na to sxoliasoume kai na plottaroume kati panw se auto

### Plot natura areas

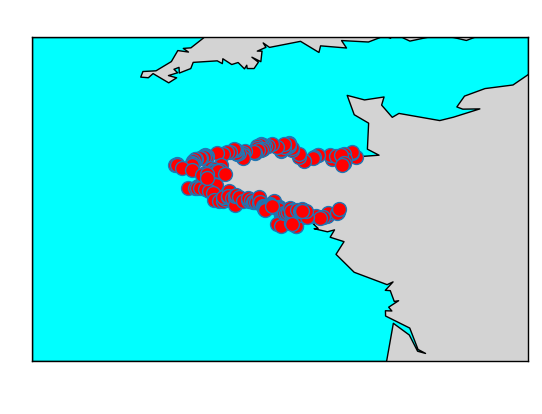
Source code:natura/ fishing.py

Under construction

Na to sxoliasoume kai na plottaroume kati panw se auto

### Plot ports of brittany

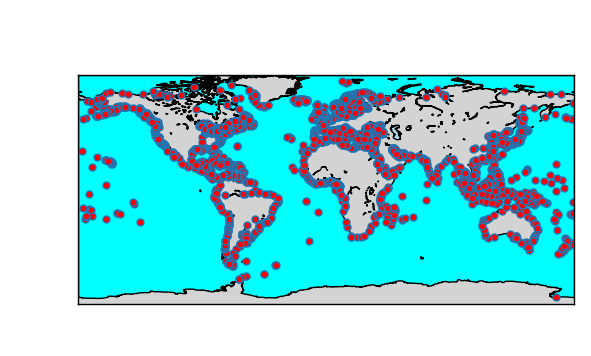
Source code:fishing\_area/ port\_britany.py



Na to sxoliasoume kai na plottaroume kati panw se auto

### Plot the ports of the europe

Source code:world\_port\_index/wpi.py



Na to sxoliasoume kai na plottaroume kati panw se auto

# Postgres version

* postgres=# SELECT version();

-----------------------------------------------------------------------------------------------------------------

PostgreSQL 10.5 (Ubuntu 10.5-0ubuntu0.18.04) on x86\_64-pc-linux-gnu, compiled by gcc (Ubuntu 7.3.0-16ubuntu3) 7.3.0, 64-bit