Applied Data Mining

1st part – Preliminary Data Analysis, SQL queries

Alevizopoulou Sofia 2022201704002

Avgeros Giannis 2022201704003

Tsiatsios George 2022201704024

Athens 2018

Contents

[Table of figures 3](#_Toc528622413)

[Entity Relation model 4](#_Toc528622414)

[Create tables 5](#_Toc528622415)

[We have added all the above tables at the database. 5](#_Toc528622416)

[Preliminary data analytics 5](#_Toc528622417)

[Null values 5](#_Toc528622418)

[SQL commands: 5](#_Toc528622419)

[1. How many ships there are based on shipname? 5](#_Toc528622420)

[2. How many ships have imo number? 5](#_Toc528622421)

[3. How many ships there are based on mmsi( dynamic table)? 5](#_Toc528622422)

[4. How many ships there are based on mmsi (static table)? 6](#_Toc528622423)

[5. How many ships have mouthermmsi (ships that other ships use them as start point)? 6](#_Toc528622424)

[6. Display 5 ships that have mouthermmsi? 6](#_Toc528622425)

[7. Display the ship which is the ship that defined as mothershipmmsi ship most of the times. 7](#_Toc528622426)

[8. How many ship types there are at the static table? 8](#_Toc528622427)

[9. Display the number of ships based on ship type 8](#_Toc528622428)

[10. Find the shipnames that have no ship type defined 9](#_Toc528622429)

[11. Display the number of vessels per country. The first 3 digits of MMSI declare the region country of the ship. 10](#_Toc528622430)

[12. Display the number of ships and their type per country 10](#_Toc528622431)

[13. Find the ship types for French vessels. 12](#_Toc528622432)

[14. Display the most popular destinations 13](#_Toc528622433)

[15. Display the type of messages that are sent from ship with shiptype=0. 13](#_Toc528622434)

[16. Display the number of messages that are sent from ships based on their type. 14](#_Toc528622435)

[17. Display the number of fishing messages per month 15](#_Toc528622436)

[18. Display the number of messages per month 16](#_Toc528622437)

[19. Display the types of messages inside a fishing area 17](#_Toc528622438)

[20. Display the number of fishing messages inside a fishing area 18](#_Toc528622439)

[21. Display the number of ships per type that send fishing messages 19](#_Toc528622440)

[22. Display the months with increased fishing activities based on ships status. 20](#_Toc528622441)

[23. Display the traffic per month 21](#_Toc528622442)

[24. Average speed per Month 22](#_Toc528622443)

[25. Display average speed, draught, length and width per ship type 23](#_Toc528622444)

[26. Traffic of most popular Port based on total traffic 24](#_Toc528622445)

[27. Average of draught, width per month for each ship type 25](#_Toc528622446)

[28. Usage of SAR vessels every month 26](#_Toc528622447)

[29. Type of atons 27](#_Toc528622448)

[30. Make a plot for a specific ship according its ais messages 28](#_Toc528622449)

[31. Plot the route of a specific vessel inside a fishing area 29](#_Toc528622450)

[32. Plot constrain fishing areas 30](#_Toc528622451)

[33. Plot fishing areas 31](#_Toc528622452)

[34. Plot natura areas 31](#_Toc528622453)

[35. Plot ports of Brittany 32](#_Toc528622454)

[36. Plot the ports of the Europe 33](#_Toc528622455)

[Postgres version 34](#_Toc528622456)

# Table of figures

[Figure 1: Ships with mothermmsi 7](#_Toc528622378)

[Figure 2:Ships with the same mothermmmsi 7](#_Toc528622379)

[Figure 3: Ships per type 8](#_Toc528622380)

[Figure 4 Visualization of number of each vessel type. 9](#_Toc528622381)

[Figure 5:Ships with no type defined 9](#_Toc528622382)

[Figure 6:Ships per counrty 10](#_Toc528622383)

[Figure 7:Detailed info for ships per counrty 11](#_Toc528622384)

[Figure 8:Ships of France 12](#_Toc528622385)

[Figure 9:Most popular destinations 13](#_Toc528622386)

[Figure 10:Messages sent from ships with type=0 13](#_Toc528622387)

[Figure 11:Number of messages sent based on ship type 14](#_Toc528622388)

[Figure 12:Fishing messages per month 15](#_Toc528622389)

[Figure 13:Visualization of fishing messages per month 15](#_Toc528622390)

[Figure 14;Number of messages per month 16](#_Toc528622391)

[Figure 15:Type of messages inside a fishing area 17](#_Toc528622392)

[Figure 16:Visualization of total and fishing messages inside a fishing area 19](#_Toc528622393)

[Figure 17:Fishing messages per month 20](#_Toc528622394)

[Figure 18:Visualization of fishing messages per month 20](#_Toc528622395)

[Figure 19:Traffic per month 21](#_Toc528622396)

[Figure 20:Visualization of traffic per month 21](#_Toc528622397)

[Figure 21:Average speed of vessels per month 22](#_Toc528622398)

[Figure 22: Visualization of Average speed of vessels per month 22](#_Toc528622399)

[Figure 23:Average speed, draught, length and width per ship type 23](#_Toc528622400)

[Figure 24:Visualization of average speed, draught, length and width per ship type 24](#_Toc528622401)

[Figure 25:Visualization of the percentage for the most popular port 24](#_Toc528622402)

[Figure 26:Average of draught,width per month for each ship type 25](#_Toc528622403)

[Figure 27:SAR vessels per month 26](#_Toc528622404)

[Figure 28:Visualization of SAR vessels per month 26](#_Toc528622405)

[Figure 29: Visualization of type of atons 27](#_Toc528622406)

[Figure 30:Route of a specific vessel with mmsi= 228231800 28](#_Toc528622407)

[Figure 31:Specific route inside a fishing area for vessel with mmsi= 227741610 29](#_Toc528622408)

[Figure 32: Fishing constraint areas 30](#_Toc528622409)

[Figure 33:Fishing areas 31](#_Toc528622410)

[Figure 34:Ports of Brittany 32](#_Toc528622411)

[Figure 35:Ports all over the world 33](#_Toc528622412)

# 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

## Null values

By observing the data, we discovered many columns that contain multiple null values. We decided it would be a good practice not to remove them in this step, but include them in the dataset and decide in one of the next steps how to handle those cases.

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

## 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;

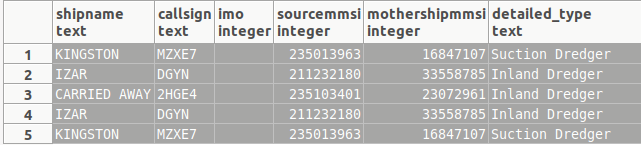


Figure 1: Ships with mothermmsi

### 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;

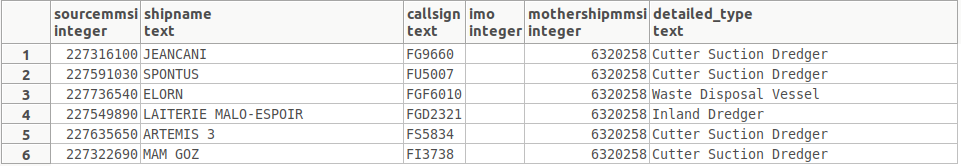


Figure 2:Ships with the same mothermmmsi

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

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

Count: 45

### 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;

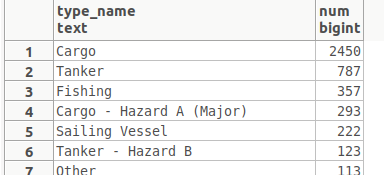


Figure 3: Ships per type

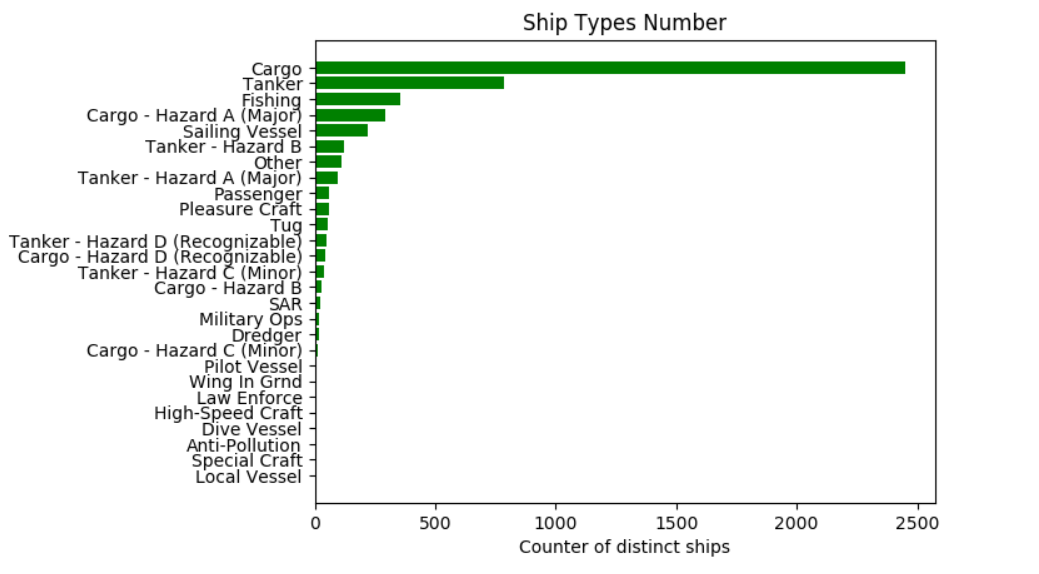


Figure 4 Visualization of number of each vessel type.

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

SELECT DISTINCT shipname   
FROM   nari\_ais\_static   
WHERE  nari\_ais\_static.shiptype IS NULL;

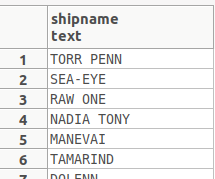


Figure 5:Ships with no type defined

### 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;

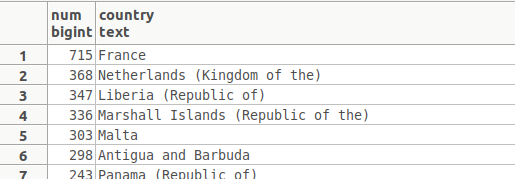


Figure 6:Ships per counrty

### 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;

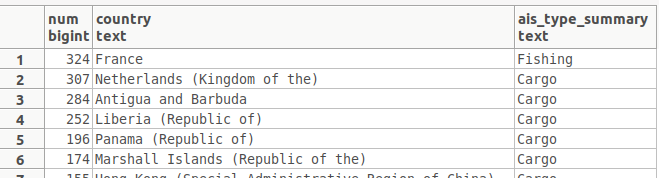


Figure 7:Detailed info for ships per counrty

### Find the ship types for French vessels.

SELECT DISTINCT mmsi\_country\_code   
FROM   PUBLIC.country\_codes   
WHERE  country LIKE 'France';

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

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;

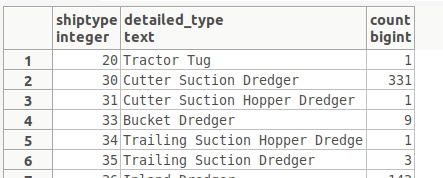


Figure 8:Ships of France

### 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;

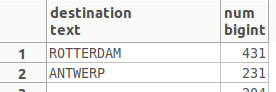


Figure 9:Most popular destinations

### 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;

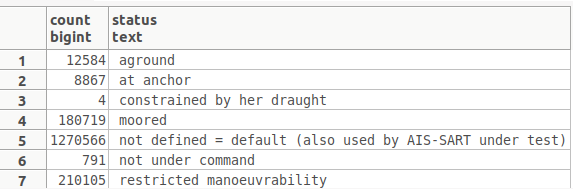


Figure 10:Messages sent from ships with type=0

### 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;

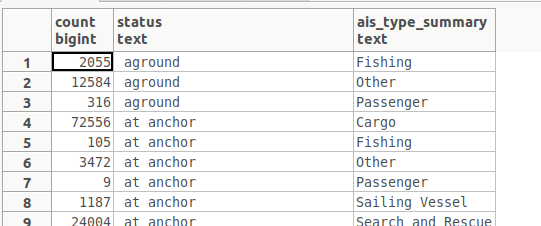


Figure 11:Number of messages sent based on ship type

### 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;

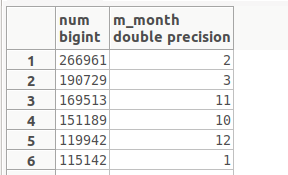


Figure 12:Fishing messages per month

Figure 13:Visualization of fishing messages per month

### 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;

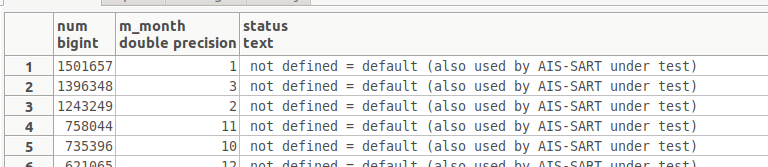


Figure 14;Number of messages per month

### 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;

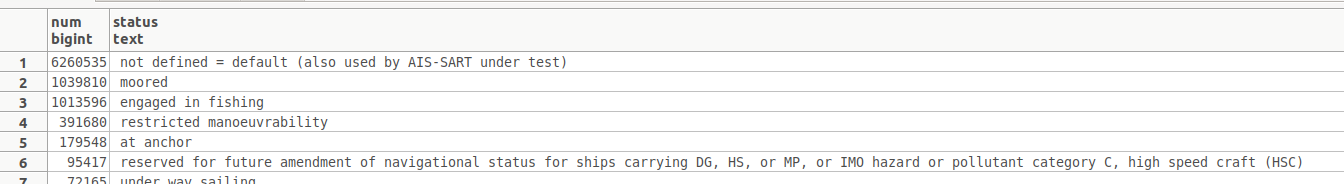


Figure 15:Type of messages inside a fishing area

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

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

Figure 16:Visualization of total and fishing messages inside a fishing area

### 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;

### 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;

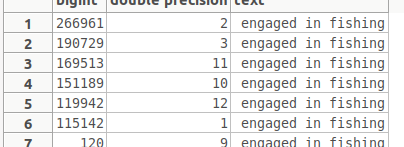


Figure 17:Fishing messages per month

Figure 18:Visualization of fishing messages per month

### 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;

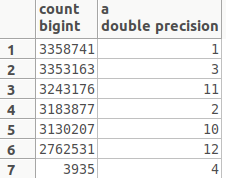


Figure 19:Traffic per month

Figure 20:Visualization of traffic per month

### Average speed per Month

SELECT *Avg*(speed),   
       **Extract**(month FROM **To\_timestamp**(t)) a   
FROM   nari\_dynamic   
GROUP  BY a;

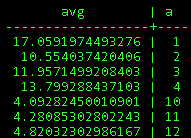


Figure 21:Average speed of vessels per month

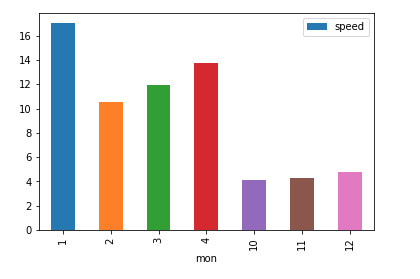


Figure 22: Visualization of Average speed of vessels per month

### 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;

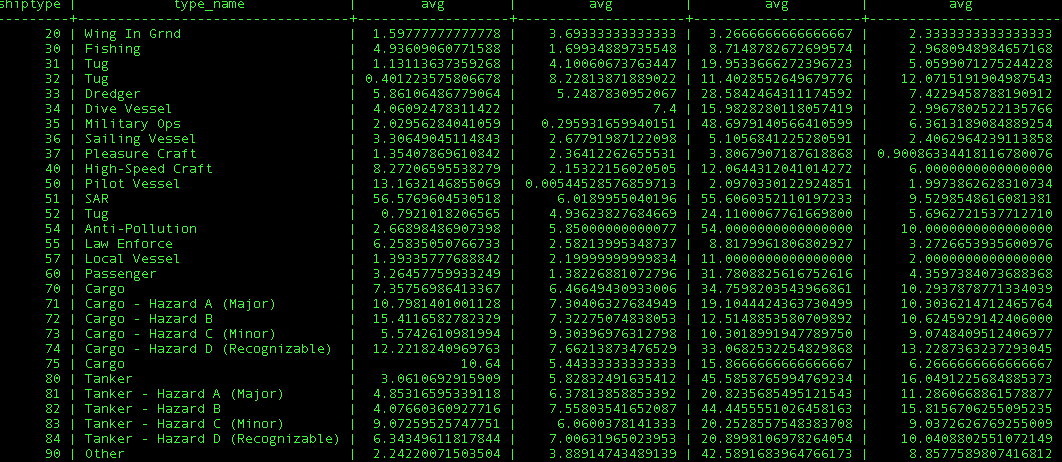
******

Figure 23:Average speed, draught, length and width per ship type

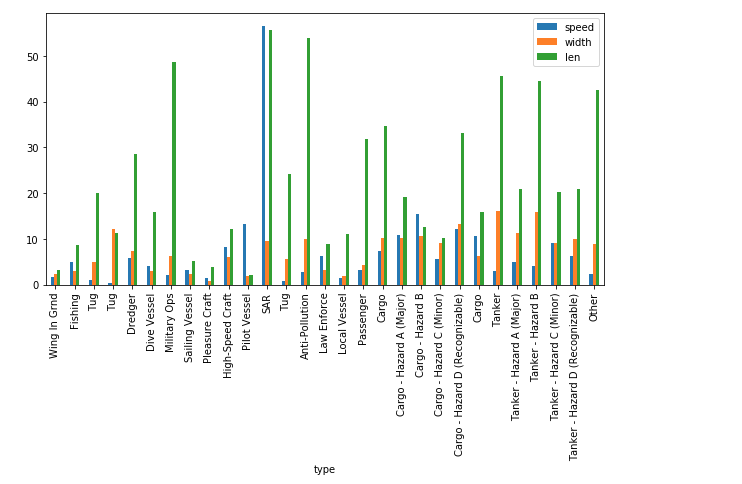
******

Figure 24:Visualization of average speed, draught, length and width per ship type

### Traffic of most popular Port based on total traffic

SELECT *Count*(\*)   
FROM   nari\_ais\_static   
WHERE  destination LIKE '%BREST%'   
        OR destination LIKE '%BES'   
        OR destination LIKE ' ';

*Count:640060*

Count of all 1078617

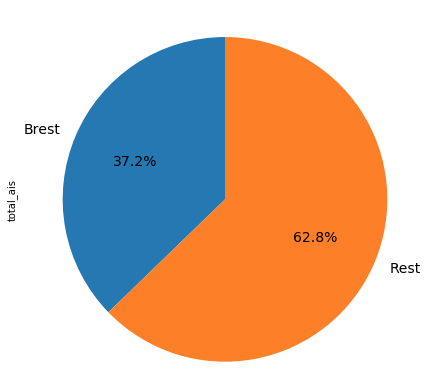


Figure 25:Visualization of the percentage for the most popular port

### 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;

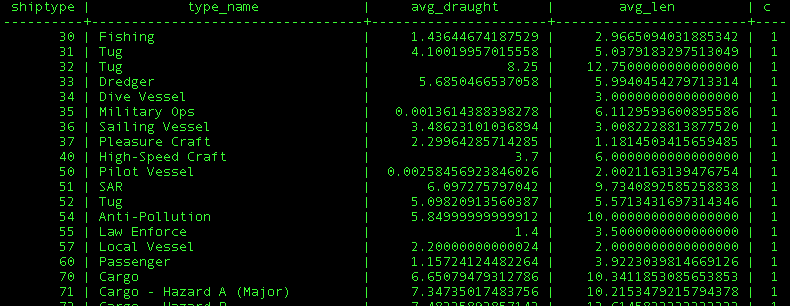
******

Figure 26:Average of draught,width per month for each ship type

### 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;

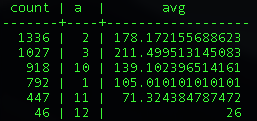
****

Figure 27:SAR vessels per month

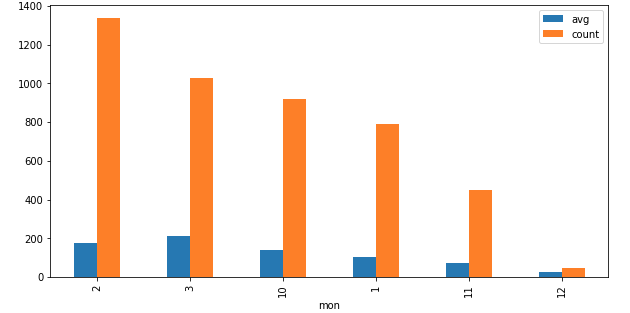
****

Figure 28:Visualization of SAR vessels per month

### 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;

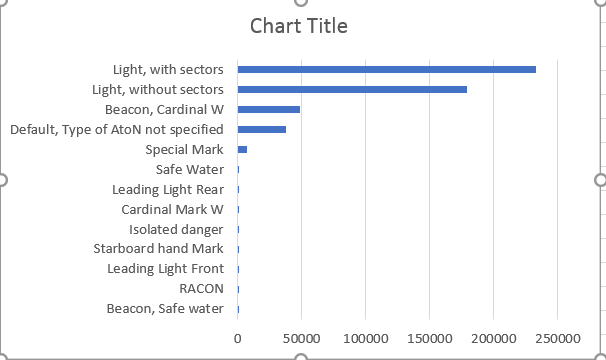


Figure 29: Visualization of type of atons

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

Source code: spatial.py

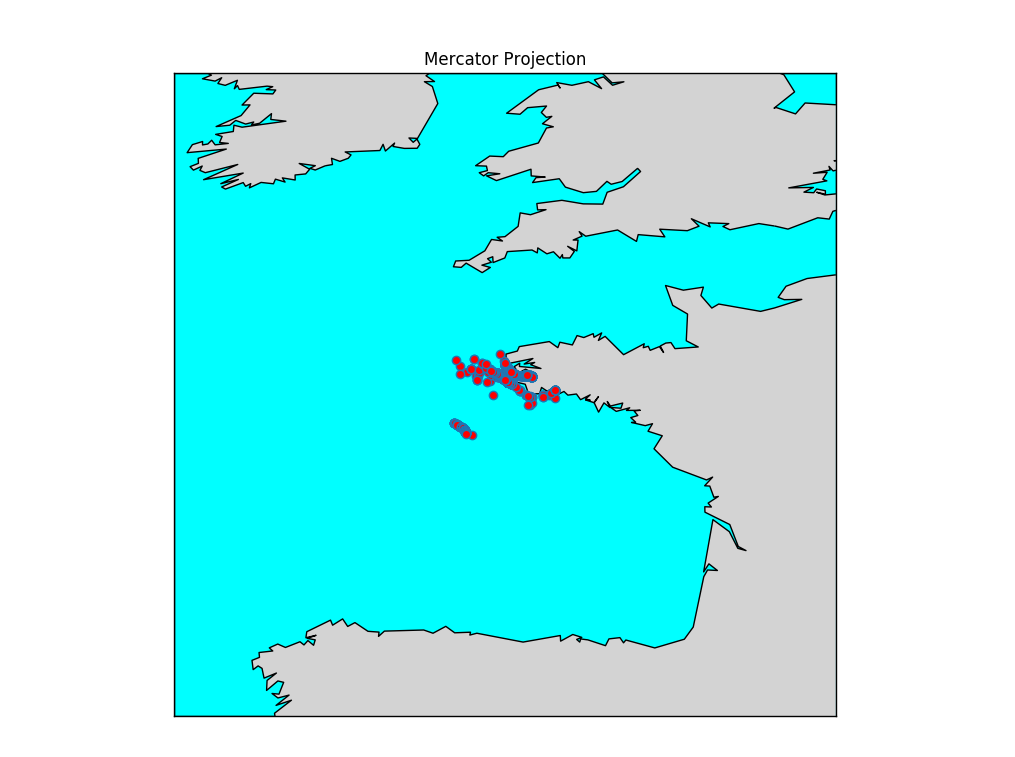


Figure 30:Route of a specific vessel with mmsi= 228231800

### Plot the route of a specific vessel inside a fishing area

Source code: vessel\_at\_fishing\_area.py

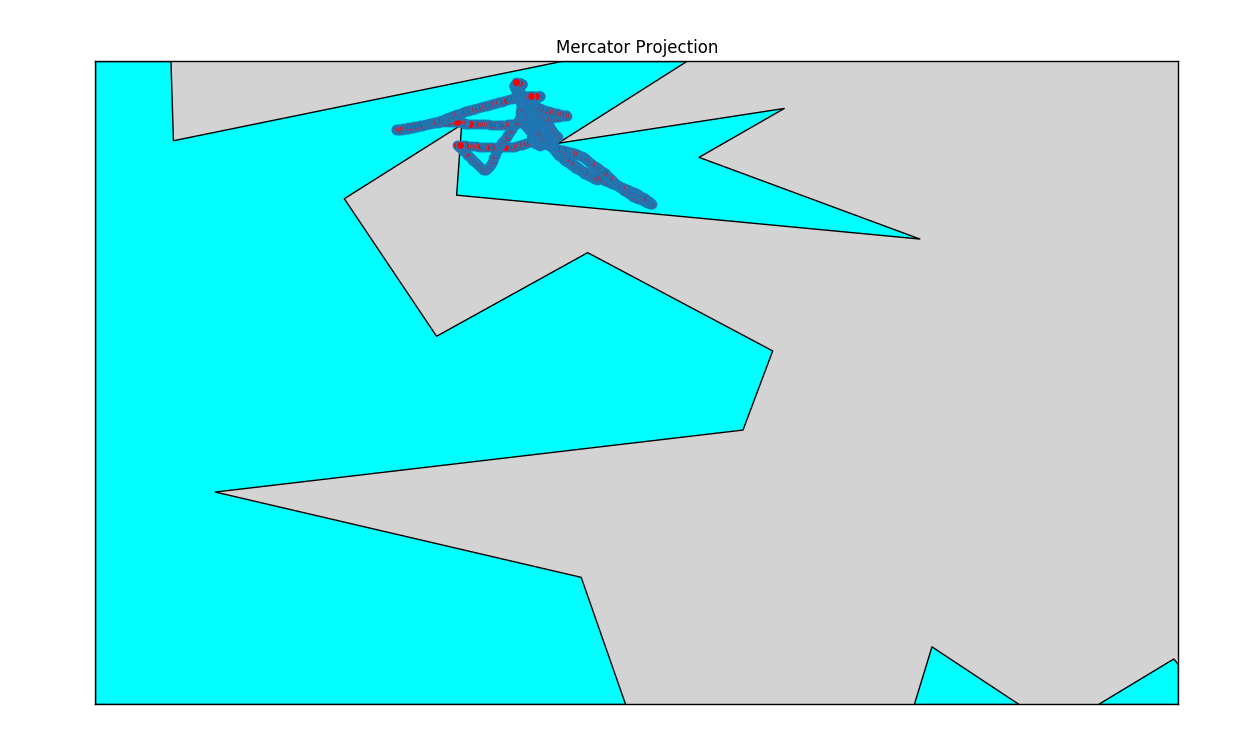


Figure 31:Specific route inside a fishing area for vessel with mmsi= 227741610

### Plot constrain fishing areas

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

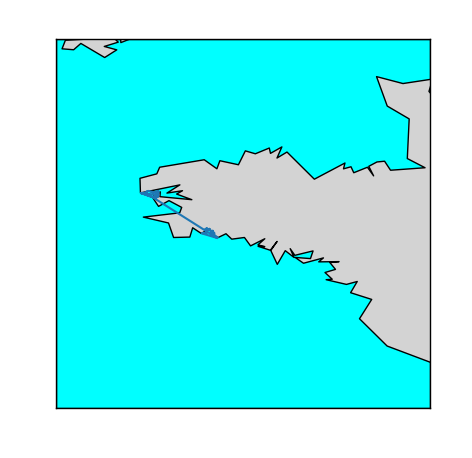


Figure 32: Fishing constraint areas

### Plot fishing areas

Source code:fishing\_area/ fishing.py



Figure 33:Fishing areas

### 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

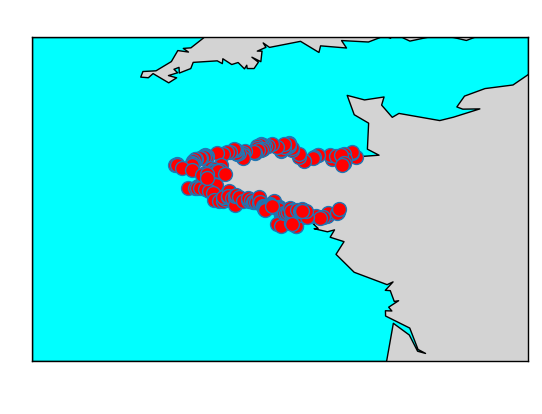


Figure 34:Ports of Brittany

### Plot the ports of the Europe

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

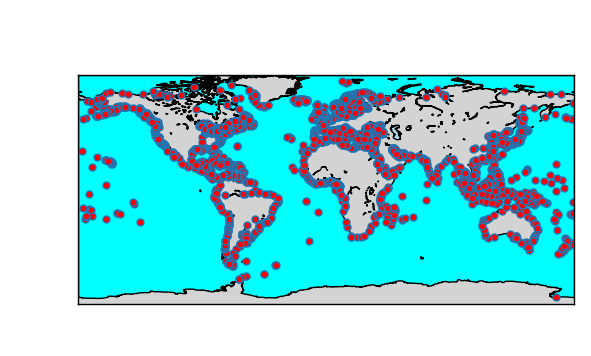


Figure 35:Ports all over the world

QGIS

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

# 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