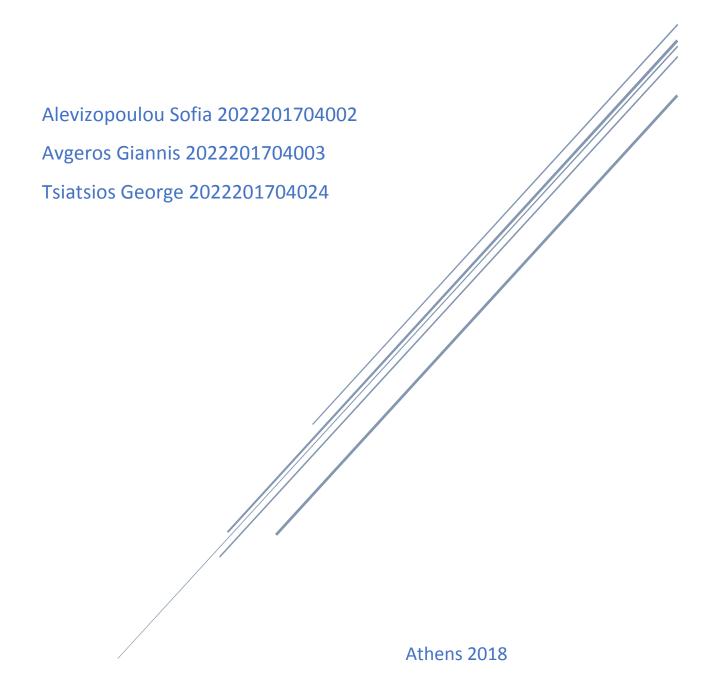
# **Applied Data Mining**

1st part – Preliminary Data Analysis, SQL queries



# Contents

Tak	ole of	figures	3
Ent	ity Re	lation model	4
Cre	ate ta	ables	5
We	have	added all the above tables at the database	5
Pre	limina	ary data analytics	5
ſ	Null va	alues	5
9	SQL co	ommands:	5
	1.	How many ships there are based on shipname?	5
	2.	How many ships have imo number?	5
	3.	How many ships there are based on mmsi( dynamic table)?	6
	4.	How many ships there are based on mmsi (static table)?	6
	5. poir	How many ships have mouthermmsi (ships that other ships use them as start nt)?	6
	6.	Display 5 ships that have mouthermmsi?	7
	7.	Display the most popular mothershipmmsi ship	7
	8. ship	Display the ships that have as mothershipmmsi the most popular mothershipm	
	9.	How many ship types there are at the static table?	8
	10.	Display the number of ships based on ship type	9
	11.	Find the shipnames that have no ship type defined	10
	12. regi	Display the number of vessels per country. The first 3 digits of MMSI declare on country of the ship.	
	13.	Display the number of ships and their type per country	11
	14.	Find the ship types for French vessels	12
	15.	Display the most popular destinations	13
	16.	Display the type of messages that are sent from ship with shiptype=0	13
	17.	Display the number of messages that are sent from ships based on their type	e. 14
	18.	Display the number of fishing messages per month	15
	19.	Display the number of messages per month	16
	20.	Display the types of messages inside a fishing area	17
	21.	Display the number of fishing messages inside a fishing area	18
	22. fish	Display the number of ships that have navigational status "fishing" but out o	
	23.	Display the number of ships per type that send fishing messages	
	24.	Display the traffic per month	
	25.	Display the average speed per Month	
		. ,	_

	26.	Display average speed, draught, length and width per ship type	. 23
	27.	Display the traffic of Brest port	. 24
	28.	Display the average of draught, width per month for each ship type	. 25
	29.	Display the usage of SAR vessels every month	. 26
	30.	Display the type of atons	. 27
	31.	Make a plot for a specific ship according its ais messages (python)	. 28
	32. messa	Make a plot for a specific ship according its fishing ais messages. All these ges are inside a fishing area (QJIS)	. 29
	33. messa	Make a plot for a specific ship according its fishing ais messages. All these ges are inside a fishing area. Fishing areas are presented. (QJIS)	. 30
		Make a plot for a specific ship according its fishing ais messages. All these ges are inside a fishing area. Fishing areas and constraint fishing areas are nted. (QJIS)	31
	35.	Plot the route of a specific vessel inside a fishing area (python)	. 32
	36.	Plot constrain fishing areas (python)	. 33
	37.	Plot fishing areas(python)	. 34
	38.	Plot ports of Brittany(python)	. 35
	39.	Plot the ports of the Europe (python)	. 36
Ru	nning en	vironment	. 37
(	OS versio	on	. 37
١	Postgres	version	. 37
(	QGIS ver	sion	. 37
١	Python v	rersion	. 37

# Table of figures

Figure 1: Ships with mothermmsi	/
Figure 2:Ships with the same mothermmmsi	8
Figure 3: Ships per type	9
Figure 4 Visualization of number of each vessel type	9
Figure 5:Ships with no type defined	10
Figure 6:Ships per counrty	10
Figure 7:Detailed info for ships per counrty	11
Figure 8:Ships of France	12
Figure 9:Most popular destinations	13
Figure 10:Messages sent from ships with type=0	13
Figure 11:Number of messages sent based on ship type	14
Figure 12:Fishing messages per month	15
Figure 13:Visualization of fishing messages per month	15
Figure 14; Number of messages per month	16
Figure 15:Type of messages inside a fishing area	17
Figure 16:Visualization of total and fishing messages inside a fishing area	19
Figure 17:Traffic per month	21
Figure 18:Visualization of traffic per month	21
Figure 19:Average speed of vessels per month	22
Figure 20: Visualization of Average speed of vessels per month	22
Figure 21:Average speed, draught, length and width per ship type	23
Figure 22:Visualization of average speed, draught, length and width per ship type	24
Figure 23:Visualization with the vessels that are connected with Brest Port	24
Figure 24:Average of draught, width per month for each ship type	25
Figure 25:SAR vessels per month	26
Figure 26:Visualization of SAR vessels per month	26
Figure 27: Visualization of type of atons	27
Figure 28:Route of a specific vessel with mmsi= 228931000	28
Figure 29: Route of a specific vessel with mmsi= 228931000 with fishing messages	29
Figure 30: Route of a specific vessel with mmsi= 228931000 with fishing messages, fishing	
areas are presented	30
Figure 31:Route of a specific vessel with mmsi= 228931000 with fishing messages, fishing	
areas and constrained fishing areas are presented	31
Figure 32:Specific route inside a fishing area for vessel with mmsi= 227741610	32
Figure 33: Fishing constraint areas	33
Figure 34:Fishing areas	34
Figure 35:Ports of Brittany	35
Figure 36:Ports all over the world	36

# **Entity Relation model**

The tables that we are going to use 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

We are going to use most of the tables at the next section at which we will try to understand the data.

# **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:

1. 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

2. How many ships have imo number?

```
SELECT Count (DISTINCT imo)
FROM nari_ais_static;
```

Count: 4033

Note: There are ships with no imo number

3. 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)

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

```
SELECT Count (DISTINCT sourcemmsi)
FROM nari ais static;
```

Count: 4842

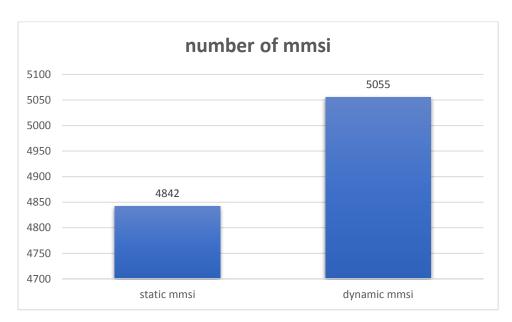


Figure 1:Number of mmsis at 2 tables (static, dynamic)

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

```
SELECT Count (DISTINCT mothershipmmsi)
FROM nari ais static;
```

Count: 228

## 6. 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;
```

	shipname text	callsign text	sourcemmsi integer	mothershipmmsi integer	detailed_type text
1	KINGSTON	MZXE7	235013963	16847107	Suction Dredger
2	IZAR	DGYN	211232180	33558785	Inland Dredger
3	CARRIED AWAY	2HGE4	235103401	23072961	Inland Dredger
4	IZAR	DGYN	211232180	33558785	Inland Dredger
5	KINGSTON	MZXE7	235013963	16847107	Suction Dredger

Figure 2: Ships with mothermmsi

7. Display the most popular mothershipmmsi ship.

**8.** Display the ships that have as mothershipmmsi the most popular mothershipmmsi ship

	sourcemmsi integer	shipname text	callsign text	mothershipmmsi integer	detailed_type text
1	227316100	JEANCANI	FG9660	6320258	Cutter Suction Dredger
2	227591030	SPONTUS	FU5007	6320258	Cutter Suction Dredger
3	227736540	ELORN	FGF6010	6320258	Waste Disposal Vessel
4	227549890	LAITERIE MALO-ESPOIR	FGD2321	6320258	Inland Dredger
5	227635650	ARTEMIS 3	FS5834	6320258	Cutter Suction Dredger
6	227322690	MAM GOZ	FI3738	6320258	Cutter Suction Dredger

Figure 3:Ships with the same mothermmmsi

9. How many ship types there are at the static table?

```
SELECT Count (DISTINCT shiptype)
FROM nari_ais_static;
```

Count: 45

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

	type_name text	num bigint
1	Cargo	2450
2	Tanker	787
3	Fishing	357
4	Cargo - Hazard A (Major)	293
5	Sailing Vessel	222
6	Tanker - Hazard B	123
7	0ther	113

Figure 4: Ships per type

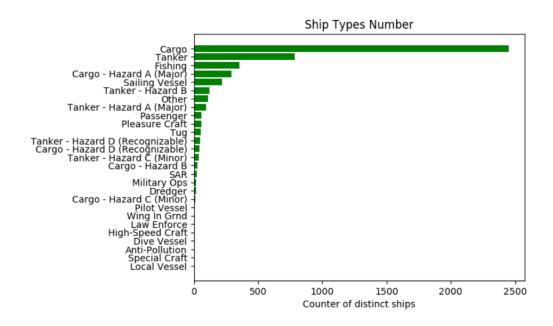


Figure 5 Visualization of number of each vessel type.

11. Find the shipnames that have no ship type defined

```
SELECT DISTINCT shipname
FROM nari_ais_static
WHERE nari_ais_static.shiptype IS NULL;
```

	shipname text
1	TORR PENN
2	SEA-EYE
3	RAW ONE
4	NADIA TONY
5	MANEVAI
6	TAMARIND
7	DOI ENN

Figure 6:Ships with no type defined

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

	num bigint	country text		
1	715	715 France		
2	2 368 Netherlands (Kingdom of the)			
3	3 347 Liberia (Republic of)			
4 336 Marshall Islands (Republic of the)		Marshall Islands (Republic of the)		
5	5 303 Malta			
6 298 Antigua and Barbuda		Antigua and Barbuda		
7	7 243 Panama (Republic of)			

Figure 7:Ships per counrty

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

	num bigint	country text	ais_type_summary text
1	324	France	Fishing
2	307	Netherlands (Kingdom of the)	Cargo
3	284	Antigua and Barbuda	Cargo
4	4 252 Liberia (Republic of) Cargo 5 196 Panama (Republic of) Cargo		Cargo
5	196	Panama (Republic of)	Cargo
6	174	Marshall Islands (Republic of the)	Cargo
-	100	Hone Mane (Consist Administrative Design of China)	Caraa

Figure 8:Detailed info for ships per counrty

## 14. 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) :: IN
TEGER = 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;
```

	shiptype integer	detailed_type text	count bigint
1	20	Tractor Tug	1
2	30	Cutter Suction Dredger	331
3	31	Cutter Suction Hopper Dredger	1
4	33	Bucket Dredger	9
5	34	Trailing Suction Hopper Dredge	1
6	35	Trailing Suction Dredger	3
_	3.0	Taland Bandana	145

Figure 9:Ships of France

## 15. Display the most popular destinations

	destination text	num bigint
1	ROTTERDAM	431
2	ANTWERP	231
-		204

Figure 10:Most popular destinations

16. 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
         b.shiptype=0
AND
         n.code=a.status
AND
GROUP BY n.status;
```

	count bigint	status text			
1	12584	aground			
2	8867	8867 at anchor			
3	4	4 constrained by her draught			
4 180719 moored					
5	1270566	<pre>not defined = default (also used by AIS-SART under test)</pre>			
6	6 791 not under command				
7	210105	restricted manoeuvrability			

Figure 11:Messages sent from ships with type=0

17. 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;
```

	count status bigint text		ais_type_summary text	
1	2055	aground	Fishing	
2	12584	aground	Other	
3	316	aground	Passenger	
4	72556	at anchor	Cargo	
5	105	at anchor	Fishing	
6	3472	at anchor	Other	
7	9	at anchor	Passenger	
8	1187	at anchor	Sailing Vessel	
9	24004	at anchor	Search and Rescue	

Figure 12:Number of messages sent based on ship type

# 18. Display the number of fishing messages per month

	num bigint	m_month double precision
1	266961	2
2	190729	3
3	169513	11
4	151189	10
5	119942	12
6	115142	1

Figure 13:Fishing messages per month

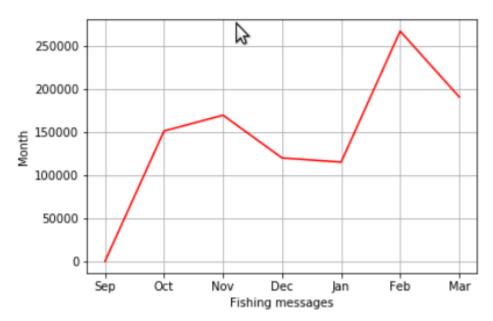


Figure 14:Visualization of fishing messages per month

# 19. Display the number of messages per month

J									
	num m_month bigint double precision			status text					
ľ	1	1501657	1	not defined = default (also used by AIS-SART under test)					
	2	1396348	3	not defined = default (also used by AIS-SART under test)					
	3	1243249	2	not defined = default (also used by AIS-SART under test)					
	4	758044	11	not defined = default (also used by AIS-SART under test)					
	5	735396	10	not defined = default (also used by AIS-SART under test)					
	6	621065	12	not defined - default (also used by ATS-SART under test)					

Figure 15; Number of messages per month

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

```
SELECT Count(*) AS num,
      n.status
      (SELECT DISTINCT maxlong,
FROM
                     maxlat,
                     minlong,
                     minlat
            geographic features fishing areas eu) g,
       FROM
      (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 16:Type of messages inside a fishing area

## 21. 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

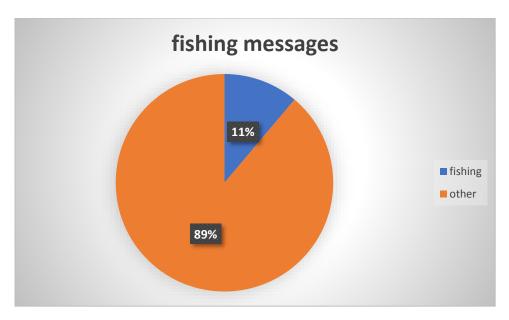


Figure 17:Visualization of total and fishing messages inside a fishing area

22. Display the number of ships that have navigational status "fishing" but out of fishing areas.

```
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
       OR minlong > lon )
       AND (maxlat < lat
       OR minlat > lat)
       AND n.code = d.status
       AND n.code = 7
GROUP BY n.status
ORDER BY num DESC;
```

#### Count:0

Note: All ships that have navigational status "fishing" are inside fishing areas.

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

# 24. Display the traffic per month

	count bigint	a double precision
1	3358741	1
2	3353163	3
3	3243176	11
4	3183877	2
5	3130207	10
6	2762531	12
7	3935	4

Figure 18:Traffic per month

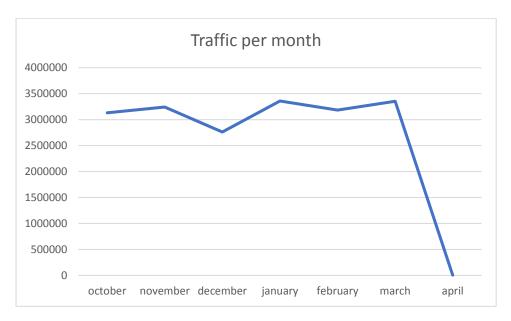


Figure 19:Visualization of traffic per month

# 25. Display the average speed per Month

avg	a
17.0591974493276	1
10.554037420406	2
11.9571499208403	3
13.799288437103	4
4.09282450010901	10
4.28085302802243	11
4.82032302986167	12

Figure 20:Average speed of vessels per month

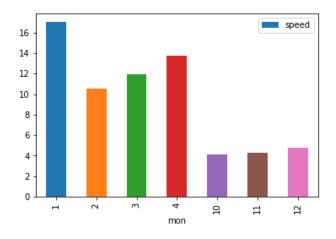


Figure 21: Visualization of Average speed of vessels per month

## 26. 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
               nari dynamic) a,
        FROM
       (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
       a.mmsi = b.sourcemmsi
WHERE
       AND c.shiptype min = b.shiptype
GROUP
       BY b.shiptype,
          type name;
```

Figure 22:Average speed, draught, length and width per ship type

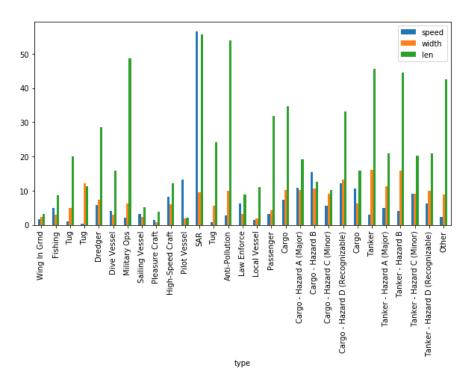


Figure 23: Visualization of average speed, draught, length and width per ship type

## 27. Display the traffic of Brest port

Count of vessels connected with Brest port :640060

Count of all vessels; 1078617

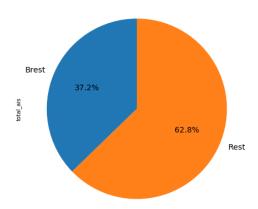


Figure 24:Visualization with the vessels that are connected with Brest Port

## 28. Display the 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,
      (SELECT mmsi,
FROM
               speed,
               Extract(month FROM To timestamp(t)) c
              nari dynamic) a,
       FROM
       (SELECT DISTINCT ( sourcemmsi ),
                       shiptype,
                      draught,
                      to_starboard
       FROM nari_ais_static) b,
       (SELECT type name,
              shiptype_min
              ship types list
        FROM
       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;
```

shiptype	type_name	avg_draught	avg_len	<	
30	Fishing	1.43644674187529	2.9665094031885342	1 1	1
31	Tug	4.10019957015558	5.0379183297513049	1	1
32	Tug	8.25	12.7500000000000000	1	1
33	Dredger	5.6850466537058	5.9940454279713314	1	1
34	Dive Vessel		3.0000000000000000	1	1
35	Military Ops	0.0013614388398278	6.1129593600895586	1	1
36	Sailing Vessel	3.48623101036894	3.0082228813877520	1	1
37	Pleasure Craft	2.29964285714285	1.1814503415659485	1	1
40	High-Speed Craft		6.0000000000000000	1	1
50	Pilot Vessel	0.00258456923846026	2.0021163139476754	1	1
51	SAR	[ 6.097275797042	9.7340892585258838	1	1
52	Tug	5.09820913560387	5.5713431697314346	1	1
54	Anti-Pollution	5.84999999999912	10.00000000000000	1	1
55	Law Enforce	1.4	3.5000000000000000	1	1
57	Local Vessel	2.20000000000024	2.0000000000000000	1	1
60	Passenger	1.15724124482264	3.9223039814669126	1	1
70	Cargo	6.65079479312786	10.3411853085653853	1	1
71	Cargo - Hazard A (Major)	7.34735017483756	10.2153479215794378	1	1

Figure 25:Average of draught, width per month for each ship type

# 29. Display the usage of SAR vessels every month

count				avg
1336	Ĺ	2	i	178.172155688623
1027		3		211.499513145083
918		10		139.102396514161
792		1		105.010101010101
447		11		71.324384787472
46		12	1	26

Figure 26:SAR vessels per month

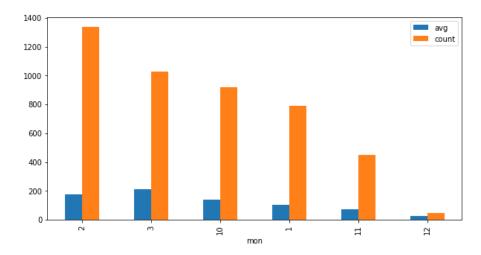


Figure 27:Visualization of SAR vessels per month

## 30. Display the type of atons

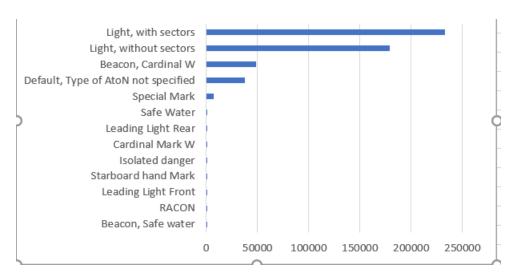


Figure 28: Visualization of type of atons

# 31. Make a plot for a specific ship according its ais messages (python) Source code: vessel\_route.py

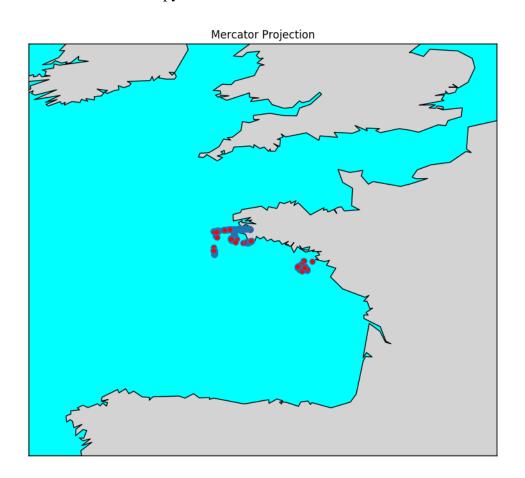


Figure 29:Route of a specific vessel with mmsi= 228931000

**32.** Make a plot for a specific ship according its fishing ais messages. All these messages are inside a fishing area (QJIS)

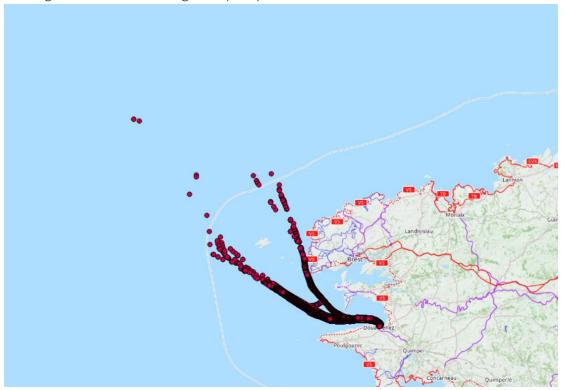


Figure 30: Route of a specific vessel with mmsi= 228931000 with fishing messages

33. Make a plot for a specific ship according its fishing ais messages. All these messages are inside a fishing area. Fishing areas are presented. (QJIS)

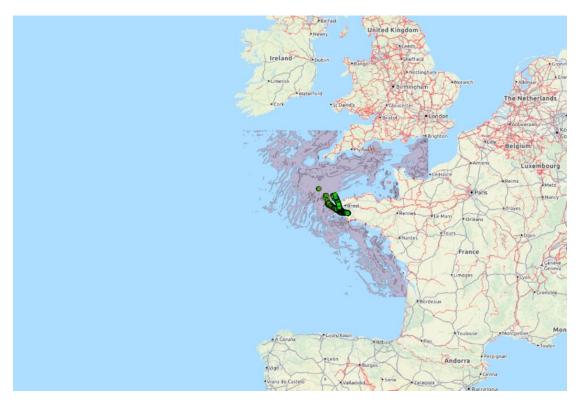


Figure 31: Route of a specific vessel with mmsi= 228931000 with fishing messages, fishing areas are presented

34. Make a plot for a specific ship according its fishing ais messages. All these messages are inside a fishing area. Fishing areas and constraint fishing areas are presented. (QJIS)

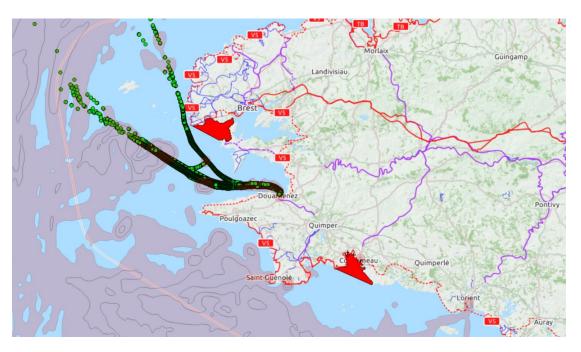


Figure 32:Route of a specific vessel with mmsi= 228931000 with fishing messages, fishing areas and constrained fishing areas are presented (fishing: purple, constraint: red)

# 35. Plot the route of a specific vessel inside a fishing area (python)

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

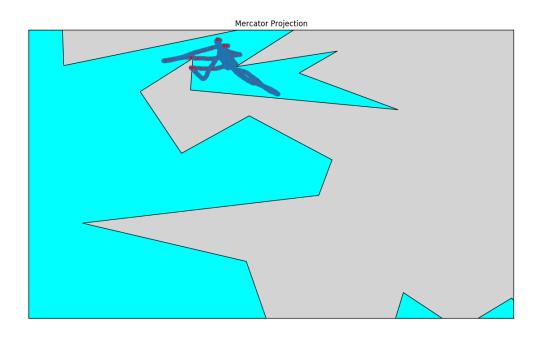


Figure 33:Specific route inside a fishing area for vessel with mmsi= 227741610

# 36. Plot constrain fishing areas (python)

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

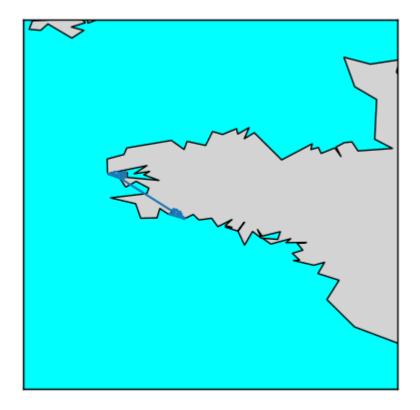


Figure 34: Fishing constraint areas

# 37. Plot fishing areas(python)

# Source code: fishing\_area/fishing.py

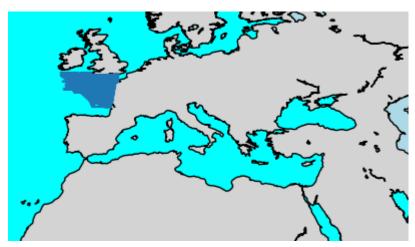


Figure 35:Fishing areas

# 38. Plot ports of Brittany(python)

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

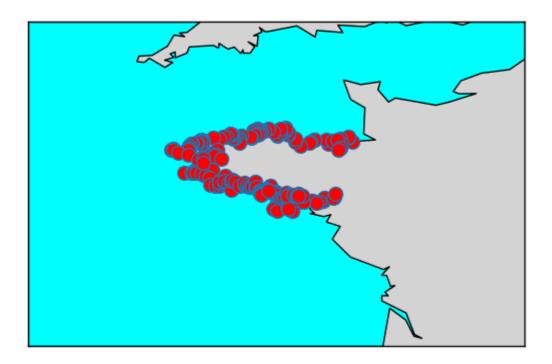


Figure 36:Ports of Brittany

# 39. Plot the ports of the Europe (python)

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

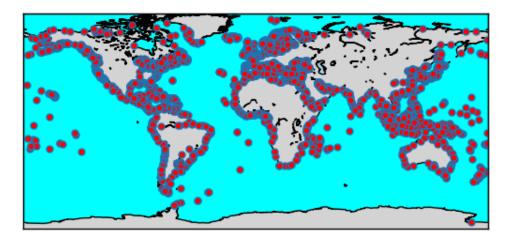


Figure 37:Ports all over the world

# **Running environment**

## OS version

- PRETTY\_NAME="Ubuntu 18.04.1 LTS"
- VERSION\_ID="18.04"

•

## 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

#### QGIS version

• version 2.18.0

# Python version

• Python 2.7.15