

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

DEPARTMENT OF INFORMATICS

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΜΕΡΟΣ Β ΕΡΓΑΣΙΑΣ Γ.Π.Σ

ON/NYMO:MAPKO ΠΛΑΚΟΥ Α.Μ:Π17107 HM/NIA:06/09/2021



ΓΕΩΓΡΑΦΙΚΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

ΠΡΟΛΟΓΟΣ





Στην εργασία αυτή:



Παρουσιάζονται τα μέρη 1,2,3 της εργασίας.

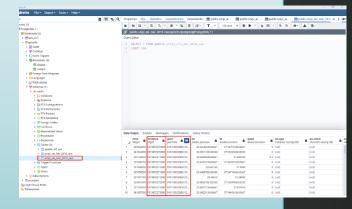
Επεξηγείται η απάντηση κάθε ερωτήματος. Χρησιμοποιήθηκαν τα εργαλία Postgres, Qgis, Python3 (Jupyter NB)

1. Φόρτωση δεδομένων (data loading)

Χρησιμοποιήθηκε η βάση δεδομένων Saronic Gulf Maritime AIS Dataset.

Η εισαγωγή της στήλης point έγινε μέσω της Postgres με
 SIRD 4326. Τα σήματα αποθηκεύτηκαν σε διαφορετικούς

πίνακες για κάθε μήνα.



mar_geo	df.head()										
	mmsi	timestmp	goom	lon	lat	enood	aie typo	ale etatue	ais heading	aie turn	ale course
	minisi	umesump	geom	1011	ıaı	speed	ais_type	ais_status	ais_neading	ais_tuiii	ais_course
239569	636013190	1519855200000	POINT (23.53934 37.88567)	23.539338	37.885675	1.4	None	None	44.0	None	15.5
239571	237991700	1519855201000	POINT (23.54615 37.94995)	23.546150	37.949950	0.0	None	None	NaN	None	268.0
239572	239722800	1519855201000	POINT (23.63220 37.94397)	23.632205	37.943968	0.0	None	None	NaN	None	0.0
239573	239550200	1519855201000	POINT (23.64085 37.94716)	23.640850	37.947157	0.0	None	None	264.0	None	142.1
239570	241024000	1519855201000	POINT (23.65010 37.93166)	23.650097	37.931657	0.0	None	None	332.0	None	178.1

1.ΣΧΕΔΙΑΓΡΑΜΜΑ ΤΗΣ ΒΑΣΗΣ

PART B: B1.GEOINFO UML DATABASE SCHEMA







2.Γνωριμία με τα δεδομένα και προετοιμασία (data acquaintance and preprocessing)

2.1 ΚΑΘΑΡΙΣΜΟΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

- Αφαιρέθηκαν οι διπλότυπες εγγραφές, και αυτές που είχαν NaN τιμές στις στήλες timestmp, speed.
- Αφαιρέθηκε ο χωρικός θόρυβος.

Έμειναν μόνο οι εγγραφές που το mmsi τους υπήρχε στα στατικά δεδομένα και αυτές των οποίων

η ταχύτητα είναι μικρότερη των 50 κόμβων.

Υπολογίστηκε η επιτάχυνση.



ΜΕΤΑ ΤΟΝ ΚΑΘΑΡΙΣΜΟ

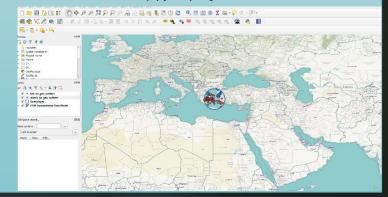
Το πλήθος δεδομένων του Φεβρουαρίου μειώθηκε κατά 1.145.481 εγγραφές.

len(final_feb_geodf) #the final length of the February dataset.
4960765

Το πλήθος του Μαρτίου μειώθηκε κατά 1.587.841.

 ${\tt len(final_mar_geodf)\#the\ final\ length\ of\ the\ March\ dataset.}$

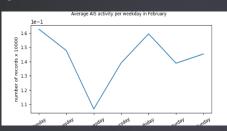
5116283

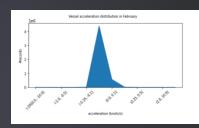


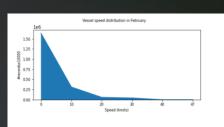
2.2-2.4

 2.2-2.3 Τα διαγράμματα δειγματοληψίας και ταχύτητας για τον μήνα Φεβρουάριο:



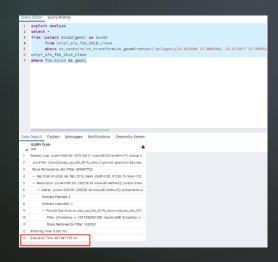






2.4 Η ὑπαρξη του R-tree ευρετηρίου μείωσε κατά πολύ τον χρόνο εκτέλεσης :

ПРІN(**421147.105** ms)



META(613.951 ms)

1	explain analyze		
2	select *		
3	from (select box2d(geom) as box2d		
4	from unipi_ais_feb_2018_clean		
5	where st_contains(st_transform(st_geomfromtext('polygon((2	3.423538 37.868298,	23.472977 37.856
6 7	unipi_ais_feb_2018_clean where foo.box2d && geom:		
	where too.box2d as geon;		
Data	Output Explain Messages Notifications Geometry Viewer		
	QUERY PLAN		
4	QUERY PLAN text		
1	QUERY PLAN (ext. Gather (cost-1020.15.799841.95 rows-704429 width-511) (actual time-304.284.613.888 rows		
4	QUERY PLAN test Settler (2011-2020 15.799841 95 rows-704429 width-511) (actual time-304 284, 613 888 rows Workers Planner: 2		
1 2 3	OUERY PLAN Game (cost-1920 15.79981 93 ross-70429 width-311) (actual time-304.284.613.888 rosss Widness Fanance 2 Widness Exercise 2		
.d 1 2	QUERY PLAN test Settler (2011-2020 15.799841 95 rows-704429 width-511) (actual time-304 284, 613 888 rows Workers Planner: 2		
1 2 3	OUERY PLAN Game (cost-1920 15.79981 93 ross-70429 width-311) (actual time-304.284.613.888 rosss Widness Fanance 2 Widness Exercise 2		
1 2 3 4	OUEN PLAN Test Gene (cost-1920 15, 799841 95 roses-704429 width-511) (actual time-104 284, 513.888 roses- Woolsen Flances 2 Woolsen Flances 2 **Neeled Loss (cost-2015, 728999 65 roses-789512 width-511) (actual time-288 202, 527 16.		
1 2 3 4 5	OSEP FLAM Gather (cost-1203.5, 799841.95 ross-764429 width-511) (actual time-304.294.513.886 ross-760409 width-511) (actual time-304.294.513.886 ross-760409 width-511) (actual time-304.294.513.886 ross-760409 width-511) (actual time-204.294.513.886 ross-760409 width-511) (actual time-208.202.527.604.004) **Partial Res Central Osep (cost-60.004.294.904.004) (actual time-208.202.527.604.004) **Partial Res Central Osep (cost-60.004.294.904.004) (actual time-208.202.527.604.004) **Partial Res Central Osep (cost-60.004.004.004) **Partial Res Central Osep (cost-60.004.004.004.004) **Partial Res Central Osep (cost-60.004.004.004) **Partial Res Central Osep (cost-60.004.004.004.004) **Partial Res Central Osep (cost-60.004.004.004.004) **Partial Res Central Osep (cost-60.004.004.004.004.004) **Partial Res Central Osep (cost-60.004.004.004.004.004.004.004.004.004.0		
1 2 3 4 5 6	OUEN FLAN test Outnite (cost-1020.15.799819 5 rose-764429 width-511) (actual time-164.284.613.888 roses- Worksen Flanching Worksen Flanching - Needer Loop (cost-020.15.728999 0.0 rose-299522 width-511) (actual time-288.202.527.16. Flanching See Cost on unity, List Not. 2018, celes unity, List Not. 2018, Celes 1.1 (cost-020.05.4. Flanching See Cost on unity, List Not. 2018, celes unity, List Not. 2018, Celes 1.1 (cost-020.05.4. Flanching See Cost on unity, List Not. 2018, celes 1.1 (cost-020.05.4. Flanching See Cost on unity, List Not. 2018, celes 1.1 (cost-020.05.4. Flanching See Cost on unity, List Not. 2018, celes 1.1 (cost-020.05.4.)		
al 1 2 3 4 5 6 7	Octor PLAN Gather (cont-12021.5.799841.95 rose-764429 wieth-911) (actual trans-304.294.013.888 roses- Workers Laucrine 2 Workers Laucrine 2 Workers Laucrine 2 **Needed Loop (cont-0.21.5.728999.05 rose-709512 wieth-911) (actual trans-288.202.027.16. **Partial		
al 1 2 3 4 5 6 7 8	Query PLAN Gather (cost-1020.15.79981.95 ross-764429 wicth-511) (actual time-304.284.513.888 ross- Worsers Faunch 2 Worsers Faunch 2 **Needer Cost-1020.15.79981.95 ross-764429 wicth-511) (actual time-204.284.513.888 ross- Worsers Laurchine 2 **Needer Cost-1020.15.72999.95 ross-793912 wicth-511) (actual time-288.292.527 lo. **Plantic Rec Cost-1020.15.72999.95 ross-793912 wicth-511) (actual time-288.292.527 lo. **Plantic Rec Cost-1020.15.72999.95 ross-793912 wicth-511) (actual time-288.292.527 lo. *Plantic Rec Cost-1020.15.7299.95 ross-793912 wicth-511) (actual time-288.292.527 lo. **Plantic Rec Cost-1020.15.7299.95 ross-793912 wicth-511) (actual time-288.292.527 lo. **Plantic Rec Cost-1020.15.7299.95 ross-793912 wicth-511) (actual time-288.292.527 lo. **Plantic Rec Cost-1020.15.72999.95 ross-793912 wicth-511) (actual time-298.292.527 lo. **Plantic Rec Cost-1020.15.72		
1 2 3 4 5 6 7 8 9	Octor FLAV Gather (cord-12021.5.799841.95 rose-706429 wieth-911) (actual trans-304.294.613.888 roses- Woodwar Flander 2 Woodwar Flander 2 Needer Loop (cord-021.5.789991.05 rose-709522 wieth-911) (actual trans-304.294.613.888 roses- Woodwar Loop (cord-021.5.78999.05 rose-709522 wieth-911) (actual trans-308.202.527.16. * Planting lead Color out play, as Jac. 2015, cean unity, as Jac. 2018, cean.1 (cord-020.5.47. Plant (illimenting ~ 1519126001600 bugs) abd ((increasing ~ 15179600100 bugs) file. * Planting lead Color out play, as Jac. 2018, cean unity, as Jac. 2018, cean.1 * Planting lead Color out play, as Jac. 2018, cean. (soci. 2013. 1.312.80 rose-469 wich-44. * Planting lead Color out play, as Jac. 2018, cean. (soci. 2013. 1.312.80 rose-469 wich-44. * Planting lead Color out play, as Jac. 2018, cean. (soci. 2013. 1.312.80 rose-469 wich-44. * Planting lead Color out play, as Jac. 2018, cean. (soci. 2013. 1.312.80 rose-469 wich-44. * Planting lead Color out play, as Jac. 2018, cean. (soci. 2013. 1.312.80 rose-469 wich-44. * Planting lead Color out play, as Jac. 2018, cean. (soci. 2013. 1.312.80 rose-469 wich-44. * Planting lead Color out play, as Jac. 2018, cean. (soci. 2013. 1.312.80 rose-469 wich-44. * Planting lead Color out play, as Jac. 2018, cean. (soci. 2013. 1.312.80 rose-469 wich-44. * Planting lead Color out play, as Jac. 2018, cean. (soci. 2013. 1.312.80 rose-469 wich-44. * Planting lead Color out play, as Jac. 2018, cean. (soci. 2013. 1.312.80 rose-469 wich-44. * Planting lead Color out play, as Jac. 2018, cean. (soci. 2013. 1.312.80 rose-469 wich-44. * Planting lead Color out play, as Jac. 2018, cean. (soci. 2013. 1.312.80 rose-469 wich-44. * Planting lead Color out play, as Jac. 2018, cean. (soci. 2013. 1.312.80 rose-469 wich-44. * Planting lead Color out play, as Jac. 2018, cean. (soci. 2013. 1.312.80 rose-469 wich-44. * Planting lead Color out play, as Jac. 2018, cean. (soci. 2013. 1.312.80 rose-469 wich-44. * Planting lead Color out play, as Jac. 2018, cean. (soci. 2013. 1.		
1 2 3 4 5 6 7 8 9 10	Query PLAN deficer (cost-1020.15.799841.95 trees-784429 wicth-911) (actual time-304.294.513.885 more- Workson Favorine 2 **Nether Long (cost-0120.15.729991.95 trees-784429 wicth-911) (inclusi time-204.294.513.885 more- Workson Favorine 2 **Nether Long (cost-02.15.72999.95 more-293912 wicth-911) (inclusi time-298.292.597.16. **Plantic Reg Cost con on plantin 34.76.2012, ceasuring sign 74.2012, ceasuring secondary sign 84.2012, ceasuring secondary sign 94.2012, ceasuring secondary sign 94.2012		
1 2 3 4 5 6 7 8 9 10 11	Other (Inch. 1201.5.79981.95 rose-764429 wieth-911) (actual time-304.294.613.88 rose-764429 wieth-911) (actual time-304.294.613.88 rose-764429 wieth-911) (actual time-304.294.613.88 rose-764429 wieth-911) (actual time-204.295.02 rose-764429 wieth-911) (actual time-204.205.02 rose-76429 205.02 rose-7		
al 1 2 3 4 5 6 7 8 9 10 11 12 13	Query PLAN Outher (pcn-17031 5, 799841 95 tross-764429 wieth-911) (actual tross-304 284. \$13.885 mose-, Workers Lancone 2 Workers Lancone 2 Workers Lancone 2 Workers Lancone 2 **Neether Long (cost-02 13, 72899 05 ross-793912 wieth-911) (actual tross-288.292 5.97 16 **Parative Ede Cost on orange, 18, Pcn, 2015, ceens unique, 18, Pcn, 2015, ceens 1, 10014-00 0.0.44. **Filled (Rinesting **-181744001002 Signil) 4AD ((insesting **-181749501000) Signil) - **Dross Rinerood on Orange, 18, Pcn, 2016, ceens ((insesting **-181749501000) Signil) - **Bress Plans (Soot on unique, 18, Pcn, 2016, ceens ((insesting **-181749501000) Signil) - **Bress Plans (Soot on unique, 18, Pcn, 2016, ceens ((insesting **-181749501000) Signil) - **Bress (Cost ((insesting **-181749501000) Signil) - **Bress (Cost ((insesting **-18174950100) Signil) Signil) Signil Signi		

3.Επεξεργασία και αναλυτική των δεδομένων (data processing and analytics)

> 3.1 Για το ερώτημα αυτό εκτελέστηκε ένα query στο QGIS (συνδεδεμένο με Postgres). Οι τροχιές κατηγοριοποιήθηκαν ανά mmsi και ανά μήνα. Το segmentation έγινε με την helper.

```
| The state | The
```

```
(Initial) Number of segments: 1
(Final-Useful) Number of port-based segments produced: 1
(Initial) Number of port-based segments: 1
(Intermediate) Number of temporal-gap-based segments: 6
(Final-Useful) Number of trips produced: 5
```

3.2 Διάφορα χωροχρονικά ερωτήματα πραγματοποιήθηκαν:





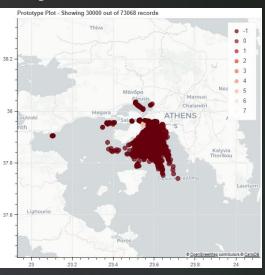
> 3.3 Το temporal alignment και η αναδειγματοληψία έγινε με την helper.

3.3 Temporal alignment and Resampling Using Helper In [8]: #TEMPORAL ALIGNMENT AND RESAMPLING resampled_and_aligned=helper2.temporal_alignment_v2(trajectories_seg) In [9]: resampled_and_aligned=resampled_and_aligned.rename(columns={'datetime':'ts'}) In [10]: len(resampled_and_aligned) #the length of the resampled dataset. Out[10]: 83459



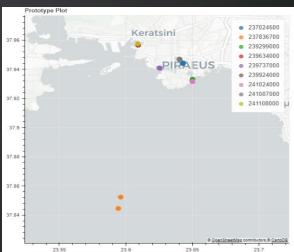
3.4-3.5

3.4 Υλοποιήθηκε ο OPTICS και βρήκε 9 συστάδες:



3.5 Υλοποιήθηκε ο Evolving Clusters και βρήκε 14 evolving cluster αποτελούμενα από 1 πλοίο:

```
Implementing Evolving Clusters Algorithm
In [12]: [res_mcs, res_mc] = evolving_clusters(final, distance_threshold=1000, min_cardinality=1, time_threshold=1, disable_progress_bar=1
                    71910/71910 [06:03<00:00, 197.94it/s]
In [13]: for row in res mcs.itertuples(): # display the clusters found
             print(row)
         Pandas(Index=0, clusters=(241024000,), st=Timestamp('2018-02-01 01:07:00'), et=Timestamp('2018-02-01 01:08:00'))
         Pandas(Index=0, clusters=(241024000), st=Timestamp('2018-02-06 01:23:00'), et=Timestamp('2018-02-06 01:24:00'))
         Pandas(Index=0, clusters=(241087000,), st=Timestamp('2018-02-14 18:37:00'), et=Timestamp('2018-02-14 18:38:00'))
         Pandas(Index=0, clusters=(241024000)), st=Timestamp('2018-02-17 00:24:00'), et=Timestamp('2018-02-17 00:25:00'))
         Pandas(Index=0, clusters=(241024000,), st=Timestamp('2018-02-17 18:06:00'), et=Timestamp('2018-02-17 18:07:00'))
         Pandas(Index=0, clusters=(237836700,), st=Timestamp('2018-02-18 13:57:00'), et=Timestamp('2018-02-18 13:58:00'))
         Pandas(Index=0, clusters=(239299000,), st=Timestamp('2018-03-03 14:13:00'), et=Timestamp('2018-03-03 14:14:00'))
         Pandas(Index=0, clusters=(241024000,), st=Timestamp('2018-03-06 18:26:00'), et=Timestamp('2018-03-06 18:27:00'))
         Pandas(Index=0, clusters=(239634000,), st=Timestamp('2018-03-13 00:34:00'), et=Timestamp('2018-03-13 00:36:00'))
         Pandas(Index=0, clusters=(241108000), st=Timestamp('2018-03-15 07:56:00'), et=Timestamp('2018-03-15 07:57:00'))
         Pandas(Index=0, clusters=(239737000)), st=Timestamp('2018-03-19 22:08:00'), et=Timestamp('2018-03-19 22:09:00'))
         Pandas(Index=0, clusters=(239737000,), st=Timestamp('2018-03-20 01:34:00'), et=Timestamp('2018-03-20 01:35:00'))
         Pandas(Index=0, clusters=(237024500,), st=Timestamp('2018-03-28 09:52:00'), et=Timestamp('2018-03-28 09:53:00'))
         Pandas(Index=0, clusters=(239924000,), st=Timestamp('2018-03-29 03:54:00'), et=Timestamp('2018-03-29 03:56:00'))
```





Data Story και Κατακλείδα

Η εργασία αυτή ήταν εξαιρετικά ενδιαφέρουσα αφού έθεσε πραγματικά ερωτήματα και επίσης χρησιμοποιήθηκαν αληθινά δεδομένα.

Τα συμπεράσματα πού εξήχθησαν και 'διηγούνται' την ιστορία των δεδομένων είναι:

- Μειωμένη κίνηση των πλοίων τις Τετάρτες του Φεβρουαρίου και τις Κυριακές του Μαρτίου.
- Δεν υπήρχαν πλοία με ακριβώς κοινή τροχιά.
- Από την εύρεση των hot-spots, είναι εμφανές ότι η πλειοψηφία των σημάτων λαμβάνεται σε κοντινή απόσταση από το λιμάνι του Πειραιά.

Εν κατακλείδι.

Έχοντας ως 'πρώτη ύλη' τα ακατέργαστα δεδομένα με θόρυβο που δεν παρέχουν σχεδόν καμία χρήσιμη πληροφορία και γνώση, με την κατάλληλο 'καθαρισμό' και προ επεξεργασία και με την χρήση τεχνικών αναλυτικής δεδομένων καταλήγουμε σε χρήσιμη πληροφορία που μπορεί να αξιοποιηθεί κατάλληλα από τους ειδικούς (domain experts).