Δημοκρίτειο Πανεπιστήμιο Θράκης

Πολυτεχνική Σχολή Ξάνθης

Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Η/Υ

Τομές Λογισμικού

**Ανάκτηση Πληροφοριών και Εξόρυξη Δεδομένων**

Εργασία ΙI

Ομάδα 5

Καρασακαλίδης Αλέξανδρος – Ηλίας (57448)

Μασλάρης Ιωάννης – Αριστείδης (57348)

Χανής Ηλίας (57337)

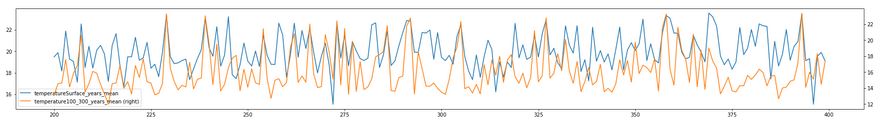
Ξάνθη

Μάιος 2020

# Μέρος Α

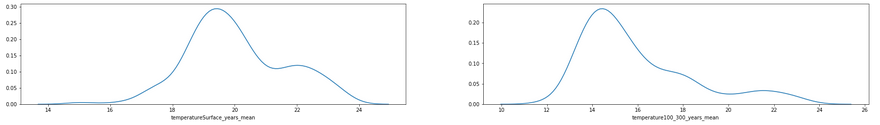
Ξεκινάμε αναλύοντας τα δεδομένα ώστε να αποκτήσουμε μία εικόνα της πληροφορίας που περιέχουν. Για να το κάνουμε αυτό κατασκευάζουμε κάποια γραφήματα ώστε να εντοπίσουμε συσχετίσεις και πρότυπα στα δεδομένα. Περιορίζουμε τον αριθμό των features, ώστε να εντοπίσουμε πιο εύκολα τα γραφήματα που έχουν νόημα. Δοκιμάζουμε το dataset σε διάφορους αλγορίθμους και κρατάμε για τον καθένα τα features τα οποία είναι πιο σχετικά με την έξοδο. Τέλος, από το πλήθος όλων των πιο χρήσιμων features από κάθε αλγόριθμο κρατάμε αυτά που γενικά φαίνεται να είναι πιο σχετικά με την έξοδο.

Ξεκινάμε την ανάλυση με ένα γράφημα που απεικονίζει την θερμοκρασία στην επιφάνεια και την θερμοκρασία σε βάθος 100-300m για ένα έτος.

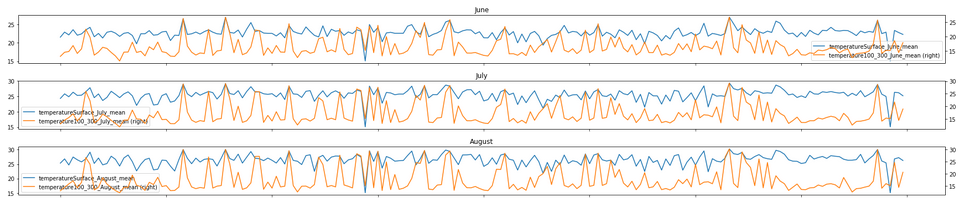


temperatureSurface\_years\_mean temperatureSurface100\_300\_years\_mean

Παρατηρούμε ότι σε γενικές γραμμές η θερμοκρασία στην επιφάνεια είναι ίδια με αυτήν σε βάθος 100-300m. Το ίδιο μοτίβο παρατηρείται και για την τυπική απόκλιση των αντίστοιχων μετρήσεων.

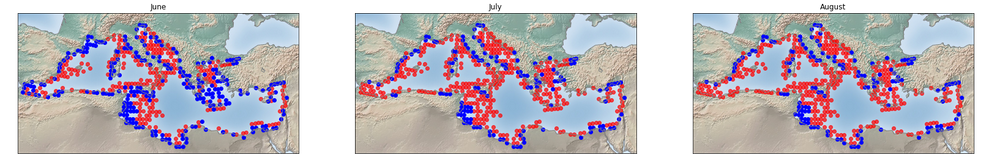


Ο ίδιος κανόνας βλέπουμε ότι ισχύει αν εξετάσουμε τα προηγούμενα δύο χαρακτηριστικά σε μηνιαία βάση. Απόκλιση παρατηρείται στους καλοκαιρινούς μήνες. Αυτό είναι λογικό αφού τότε η θερμοκρασία στην επιφάνεια διατηρείτε σε υψηλότερα επίπεδα.



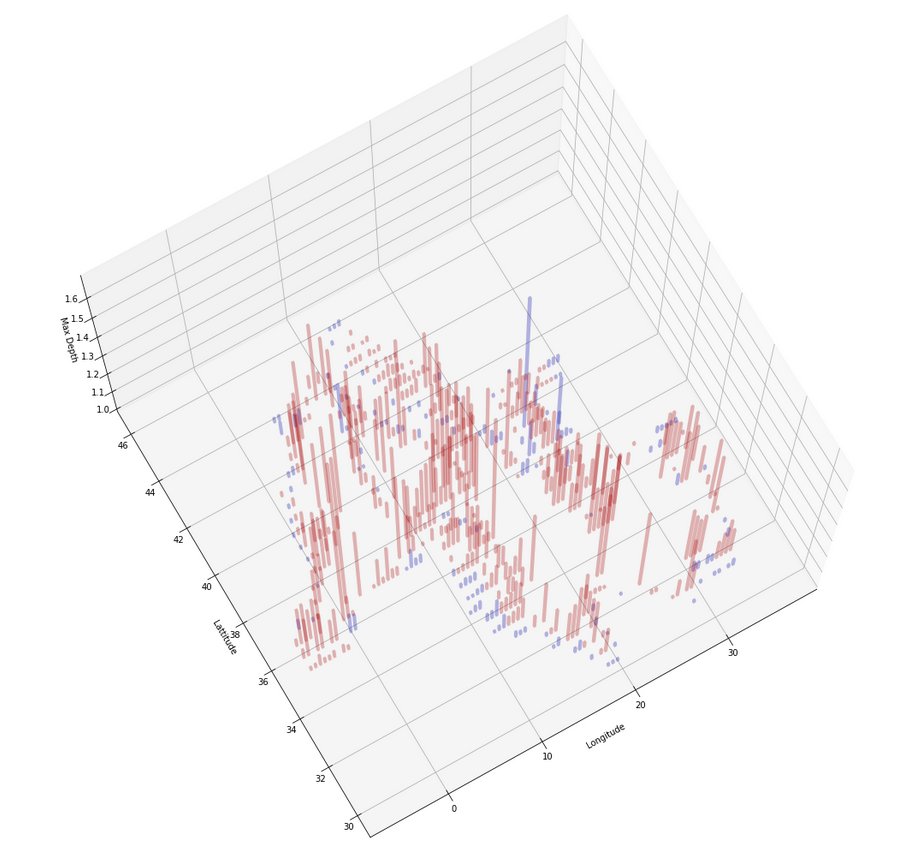
June, July, August

Στη συνέχεια βλέπουμε σε ποια σημεία της μεσογείου υπάρχει μεγάλη διαφορά θερμοκρασίας μεταξύ επιφάνειας και 100-300m τους καλοκαιρινούς μήνες.



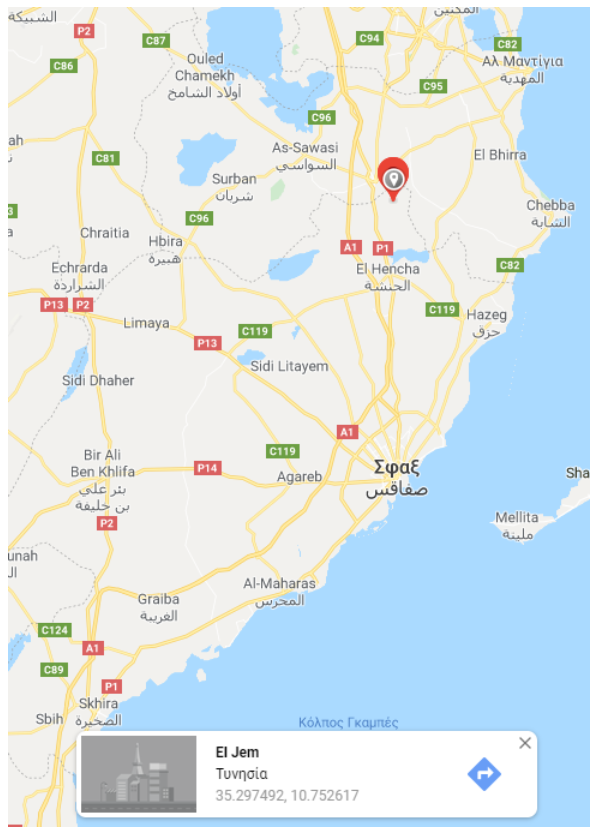
June, July, August

Στο προηγούμενο γράφημα απεικονίζονται τα σημεία στα οποία η διαφορά θερμοκρασίας είναι μεγαλύτερή ίση των 7.5 ℃ με κόκκινο χρώμα, ενώ αυτά που η διαφορά είναι μικρότερη των 7.5 ℃ με μπλε χρώμα. Βλέπουμε ότι τα σημεία με μεγάλη διαφορά θερμοκρασίας είναι στα ανοιχτά της Μεσογείου, ενώ αυτά στα οποία η διαφορά θερμοκρασίας είναι μικρή βρίσκονται κοντά στις ακτές. Αυτή η παρατήρηση φαίνεται λογική αφού σύμφωνα και με το επόμενο γράφημα η διαφορά θερμοκρασίας είναι ανάλογη του βάθους μέγιστου βάθους στο σημείο.



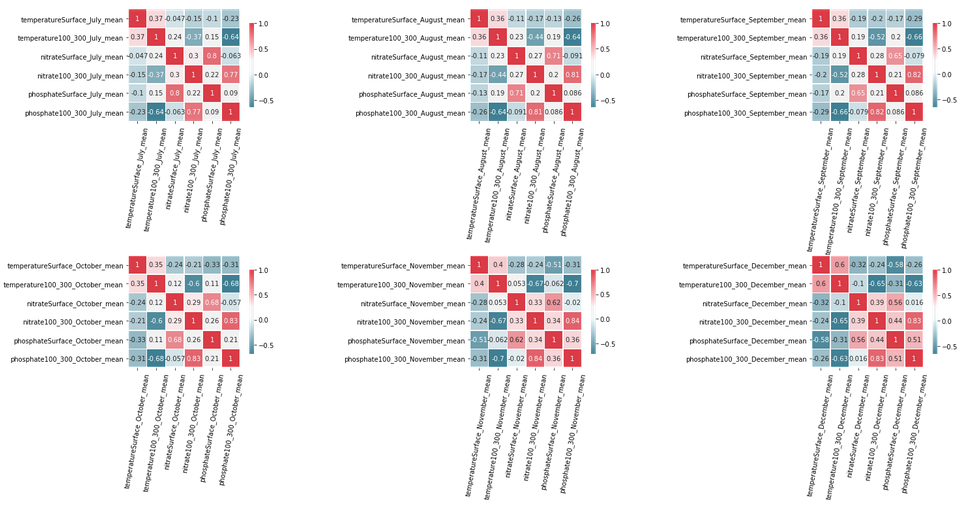
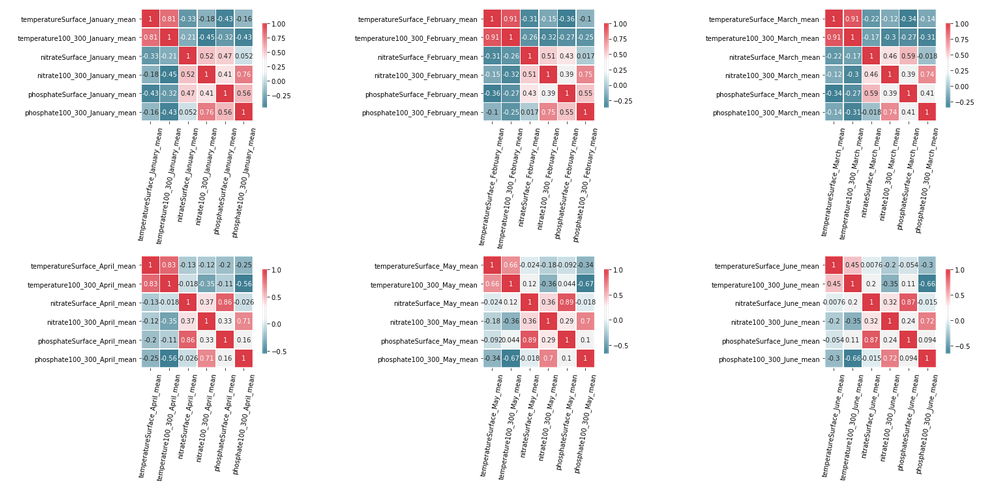
Βλέποντας την κατανομή του χαρακτηριστικού bathymetry παρατηρούμε ότι υπάρχουν σημεία όπου έχει μετρηθεί θετικό ‘βάθος’, δηλαδή είναι χερσαίες περιοχές.

Την παρατήρησή μας αυτή επιβεβαιώνουμε ψάχνοντας τις συντεταγμένες τω σημείωνα αυτών στο Google maps. Ένα από αυτά φαίνεται παρακάτω.

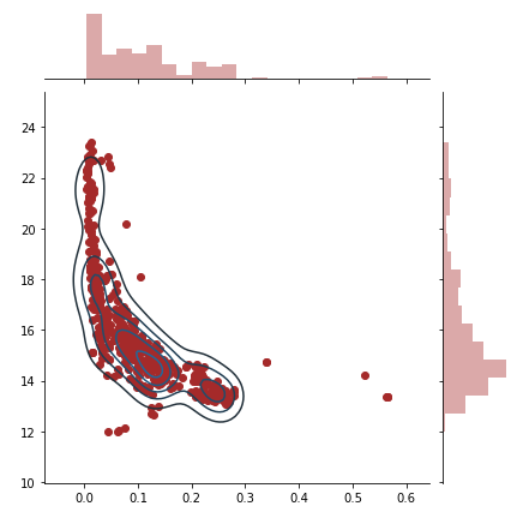
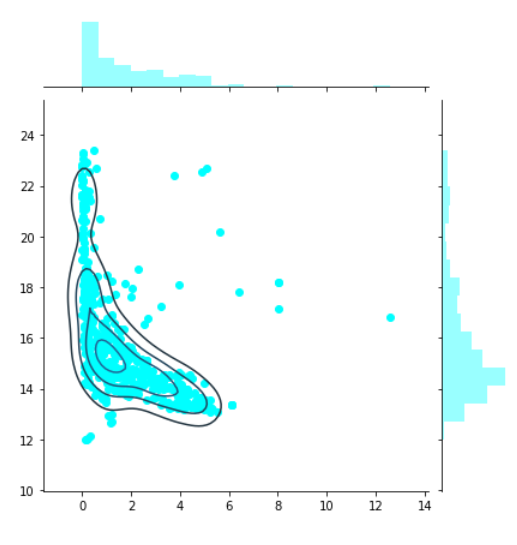


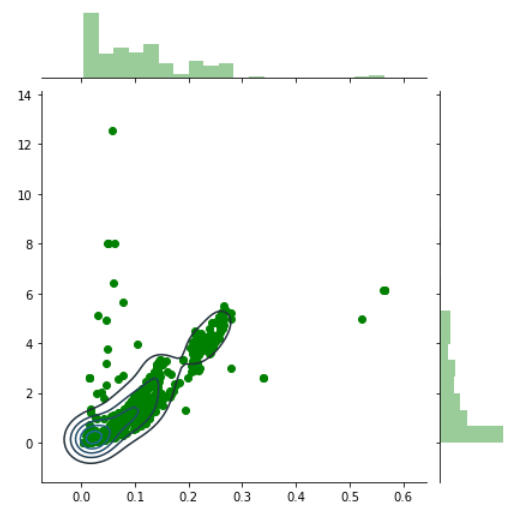
Καταλήγουμε ότι αυτά τα σημεία προκύπτουν λόγο λανθασμένων συνταραγμένων από κάποιες σημαδούρες. Ακόμα και μερικά δεκαδικά απόκλιση μπορεί να δώσουν ένα τελείως διαφορετικό σημείο από αυτό που περιμένουμε.

Για να δούμε την συσχέτιση μεταξύ των features temperature, phosphate and nitrate levels, φτιάχνουμε γραφήματα συσχέτισης (Heatmap) για τις μηνιαίες μετρήσεις αυτών των μεταβλητών στην επιφάνεια και σε βάθος 100-300m. Φαίνεται πως υπάρχει ισχυρή αρνητική συσχέτιση μεταξύ temperature και phosphate στα 100-300m όλο το χρόνο. Επίσης υπάρχει καλή αρνητική συσχέτιση μεταξύ temperature και nitrate στα 100-300m όλο το χρόνο, παρουσιάζοντας κάποια μείωση τους καλοκαιρινούς μήνες. Τέλος τα phosphate και nitrate φαίνεται ότι έχουν θετική συσχέτιση όλο το χρόνο.



Τα αντίστοιχα Joinplots μας δείχνουν πως οι συσχετίσεις temperature/phosphate και temperature/nitrate @100 – 300m είναι αντιστρόφως ανάλογες, ενώ η συσχέτισης του nitrate με το phosphate είναι ανάλογη.





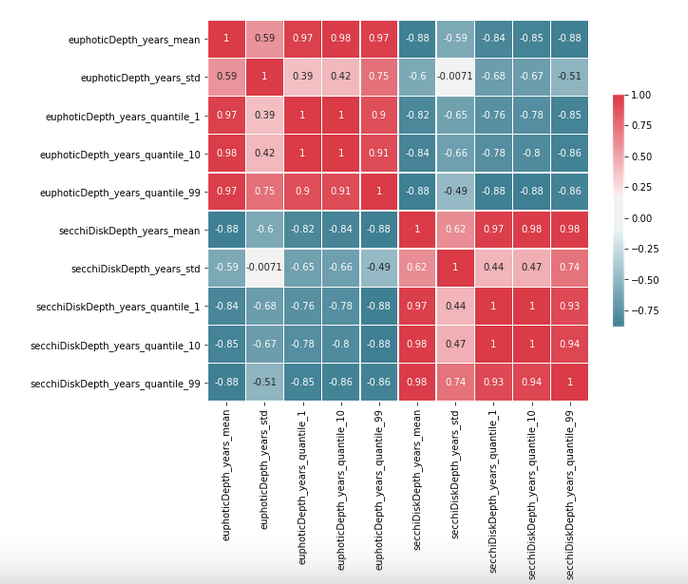
Φτιάχνοντας έναν πίνακα συσχέτισης για τα όλα τα year\_mean features παρατηρούμε τα εξής. Εκτός της συσχέτισης μεταξύ temperature, phosphate, nitrate υπάρχουν συσχετίσεις μεταξύ των:

euphoticDepth και secchiDiskDepth (αρνητική συσχέτιση)

salinity και temperature (θετική συσχέτιση)

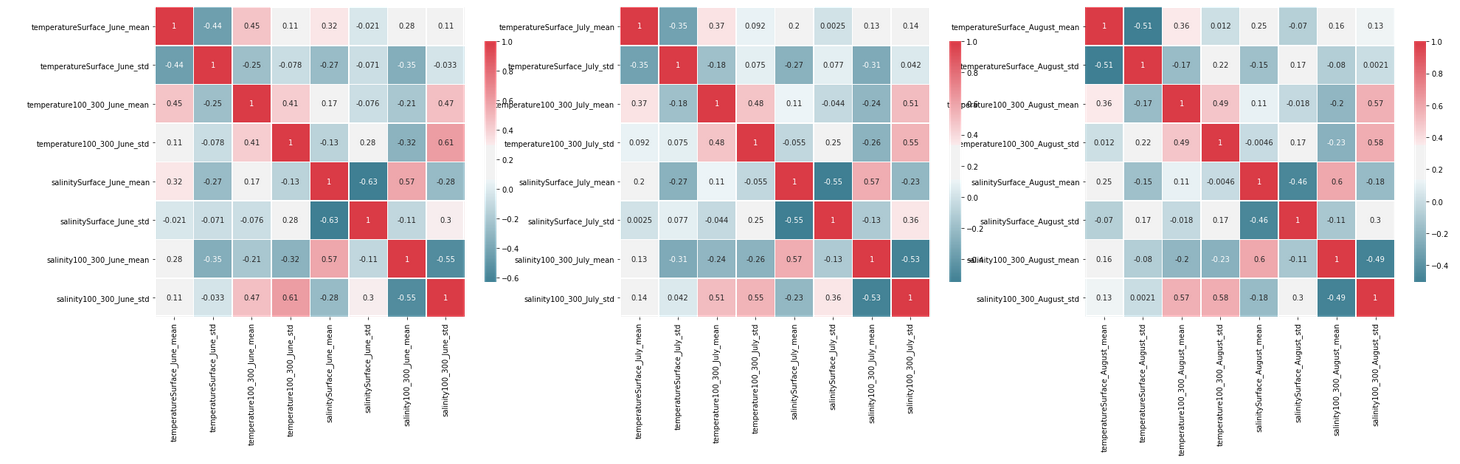
disolvedOxygen και temperature (αρνητική συσχέτιση)

Πιο συγκεκριμένα για τις προηγούμενες συσχετίσεις:

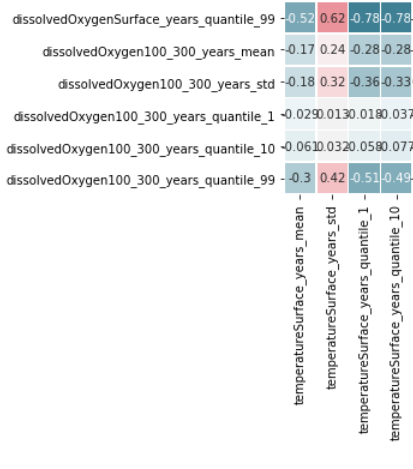


Τα features secchi Disk (το βάθος στο οποίο το φως αντανακλάται σε μεγάλο βαθμό) και euphotic Depth (επίπεδο φωτός ) έχουν αντιστρόφως ανάλογη συσχέτιση, το οποίο είναι λογικό.

Από έναν πίνακα συσχέτισης μεταξύ τιμών temperature και salinity, φαίνεται ότι υπάρχει μία μικρή θετική συσχέτιση μεταξύ αυτών των features. Η συσχέτιση αυτή φαίνεται λογική αφού όσο μεγαλύτερη θερμοκρασία έχει τόσο περισσότερη εξάτμιση υπάρχει άρα και η περιεκτικότατη σε αλάτι αυξάνεται. Τους καλοκαιρινούς μήνες όμως που η θερμοκρασία είναι πολύ υψηλή αυτός ο κανόνας φαίνεται να μην ισχύει. Άρα ίσως η προηγούμενη παρατήρηση είναι απλά τυχαία.



Επίσης φαίνεται ότι υπάρχει μία καλή συσχέτιση μεταξύ της τυπικής απόκλισης των temperature και dissolved oxygen. Αυτό σημαίνει ότι όπου υπάρχουν μεγάλες ή μικρές διαφοροποιήσεις στην θερμοκρασία υπάρχουν αντίστοιχες διαφοροποιήσεις και στο οξυγόνο των υδάτων.

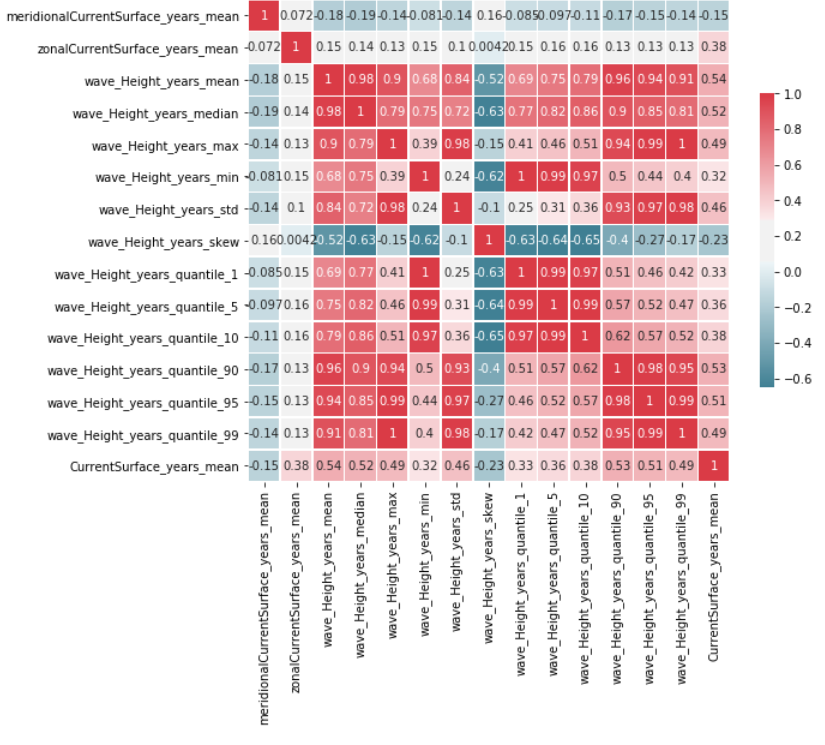


Στη συνέχεια ελέγχουμε αν υπάρχει κάποια συσχέτιση των ρευμάτων με κάποιο άλλο feature. Απεικονίζουμε την ένταση και την κατεύθυνση των ρευμάτων με βέλη των οποίων το χρώμα αναπαριστά την μέση θερμοκρασία στην περιοχή. Με μαύρα βέλη φαίνεται η μέση κατεύθυνση των ρευμάτων κάθε περιοχής σε ένα χρόνο.

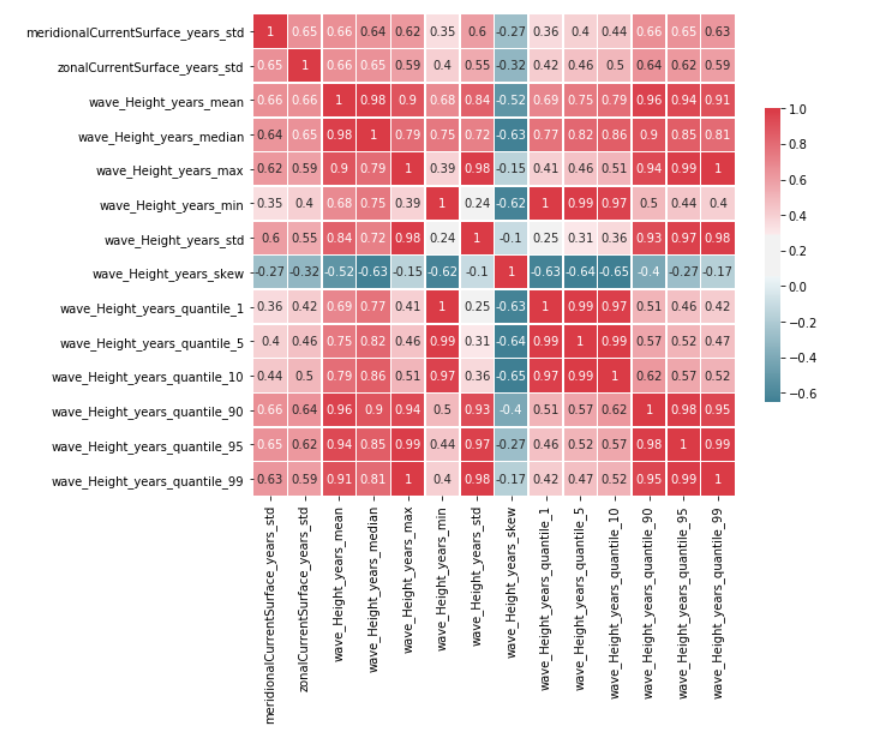


Ενδεικτικά παραθέτουμε το γράφημα για τον μήνα Ιανουάριο. Αντίστοιχα γραφήματα για όλους τους μήνες βρίσκονται στο notebook.

Τελικά υπάρχει συσχέτιση μεταξύ της μέσης τιμής των επιφανειακών ρευμάτων και όλων των μετρικών που αφορούν το ύψος των κυμάτων.



Επίσης φαίνεται ότι όσο ένταση των ρευμάτων με κατεύθυνση Βορά-Νότου, Νότου-Βορά διαφοροποιείτε τόσο διαφοροποιείτε και η το ύψος τον κυμάτων στο έτος.



Εξετάζοντας τώρα την πιθανότητα κάθε περιοχής να φιλοξενεί το Engraulis Encrasicolus βλέπουμε πως αυτή είναι μεγαλύτερη σε περιοχές της νοτιοανατολικής Μεσογείου. Επίσης φαίνεται πως η παρουσία κάποιου ποταμού δεν παίζει μεγάλο ρόλο στην πιθανότητα εμφάνισης του ψαριού, αφού υπάρχουν περιοχές με μεγάλη πιθανότητα εμφάνισης χωρίς όμως να έχουν κάποιο ποτάμι κοντά και το αντίστροφο.

