



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ**

**ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ**

**ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

---

**ΑΝΑΛΥΣΗ ΤΗΣ ΠΟΙΟΤΗΤΑΣ ΤΟΥ ΑΕΡΑ**

**ΜΕ ΧΡΗΣΗ**

**ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ**

---

Διπλωματική Εργασία

**Ματραπάζης Αναστάσιος**

Επιτροπή:

**Dr. Δασκαλοπούλου Ασπασία (επιβλέπουσα)**

**Dr. Σταμπουλής Γεώργιος**

**Dr. Τσουκαλάς Ελευθέριος**

Βόλος, Σεπτέμβριος 2019

## ΕΥΧΑΡΙΣΤΙΕΣ

Πρώτα από όλα θα ήθελα να ευχαριστήσω την επιβλέπουσα καθηγήτρια κ. Ασπασία Δασκαλοπούλου που μου έδωσε την ευκαιρία να εργαστώ πάνω σε ένα πρόβλημα επίκαιρο και σημαντικό. Η καθοδήγηση και η διαθεσιμότητα της ήταν πολύ σημαντική για την περάτωση αυτής της διπλωματικής εργασίας.

Επιπλέον, θα ήθελα να ευχαριστήσω από καρδιάς την οικογένειά μου για την αγάπη και την αμέριστη στήριξη που μου παρείχαν, παρά τις δυσκολίες, όλα αυτά τα χρόνια. Αισθάνομαι πολύ τυχερός που σας έχω δίπλα μου και με στηρίζετε σε κάθε επιλογή και όνειρο που κάνω και αποτελείτε υγιή πρότυπα για εμένα κάνοντάς καθημερινά καλύτερο άνθρωπο.

Ευχαριστώ επίσης όλους του ανθρώπους που με βοήθησαν και με ενέπνευσαν να εξελιχθώ μέσα σε αυτά τα χρόνια. Μέσα από αυτούς αισθάνομαι πιο δυνατός στις νέες προκλήσεις που θα μου φέρει η μετέπειτα σταδιοδρομία μου. Η συμβολή τους ήταν και θα είναι καταλυτική.

Τέλος, ευχαριστώ τον Δημήτρη Μάμαλη και τον Στράτο Ζολώτα που με εμπιστεύθηκαν και μου παρείχαν τα δεδομένα για την πραγματοποίηση της παρούσας διπλωματικής εργασίας. Εύχομαι το έργο τους να βοηθήσει στην ποιότητα ζωής της Σύρου και με υπομονή και επιμονή να ξεπεράσουν κάθε πρόβλημα που θα τους σταθεί εμπόδιο στην προσπάθεια αυτή.

Για τον Μάνο και τον Γιώργο

## **ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΠΕΡΙ ΑΚΑΔΗΜΑΪΚΗΣ ΔΕΟΝΤΟΛΟΓΙΑΣ ΚΑΙ ΠΝΕΥΜΑΤΙΚΩΝ ΔΙΚΑΙΩΜΑΤΩΝ**

«Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ρητά ότι η παρούσα διπλωματική εργασία, καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας, αποτελεί αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή/και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής».

Ο Δηλών

(Υπογραφή)

Ματραπάζης Αναστάσιος  
20/9/2019

## ΠΕΡΙΛΗΨΗ

Μία από τις σημαντικότερες απειλές που οφείλει η κοινωνία να αντιμετωπίσει είναι αυτή της μόλυνσης της ατμόσφαιρας. Στην παρούσα εργασία, ασχολούμαστε με δεδομένα ποιότητας αέρα που παρατηρήθηκαν από αισθητήρες της Σύρου. Αρχικά, παρουσιάζονται οι βασικές μετρικές που καθιστούν την ποιότητα του αέρα μολυσμένη και ποιες τιμές επηρεάζουν αρνητικά τόσο τον ανθρώπινο οργανισμό όσο και το περιβάλλον. Αυτές οι μετρικές αναλύονται έτσι ώστε να αναδειχθεί το πρόβλημα της ποιότητας του αέρα στη Σύρο. Τα δεδομένα αυτά χωρίζονται σε εποχές (φθινόπωρο, χειμώνας, άνοιξη, καλοκαίρι) και στην συνέχεια αναλύονται και σχολιάζονται τα αποτελέσματα. Παράλληλα γίνεται η υπόθεση ότι ο ανθρώπινος παράγοντας επηρεάζει την ποιότητα του αέρα και συγκρίνονται οι τιμές του δείκτη AQI πάνω σε κάθε ζώνη ώρας (πρωί, μεσημέρι, απόγευμα, βράδυ) με την ανθρώπινη δραστηριότητα. Η ανάδειξη του προβλήματος, μας ωθεί να αναπτύξουμε μοντέλα μηχανικής μάθησης με στόχο την παροχή προβλέψεων της ποιότητας του αέρα. Τέλος, συγκρίνεται η απόδοση αυτών των μοντέλων με χρήση δεικτών αξιολόγησης και παρουσιάζονται αδυναμίες και δυνατότητες για την συρρίκνωση της απειλής της ποιότητας του αέρα.

## **ABSTRACT**

One of the greatest threats that our society must face is the problem of air pollution. In this thesis, we tackle air quality data based on Syros installed sensors. Firstly, we present the basic measurements that make air quality disastrous and what are the negative ranges that affect both human and environmental health. Those air quality data spits into seasons (autumn, winter, spring, summer) and after making of analysis the results are commented on. After that, we assume that the human factor affects air quality and we compare the AQI values with human activity in each time zone. This problem pushes us to develop machine learning models to predict the air quality index. Finally, the performance of those models are compared with some main rating indices and we present the weaknesses and opportunities to reduce the threat of air quality.

# ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

|                                                     |            |
|-----------------------------------------------------|------------|
| <b>ΠΕΡΙΛΗΨΗ.....</b>                                | <b>v</b>   |
| <b>ABSTRACT.....</b>                                | <b>vi</b>  |
| <b>ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ .....</b>                   | <b>vii</b> |
| <b>ΕΙΣΑΓΩΓΗ .....</b>                               | <b>1</b>   |
| <b>ΚΕΦΑΛΑΙΟ 1 .....</b>                             | <b>5</b>   |
| Περιγραφή Δεδομένων ποιότητας του αέρα:.....        | 5          |
| <b>ΚΕΦΑΛΑΙΟ 2 .....</b>                             | <b>12</b>  |
| <b>2.1 Εισαγωγή.....</b>                            | <b>12</b>  |
| <b>2.2 Συναρτήσεις .....</b>                        | <b>13</b>  |
| 2.2.1 Καθαρισμός Δεδομένων: .....                   | 13         |
| 2.2.2 Χωρισμός δεδομένων για ανάλυση:.....          | 14         |
| 2.2.3 Συνάρτηση εύρεσης πιθανότητας .....           | 15         |
| 2.2.4 Οπτικές Συναρτήσεις (Visuals): .....          | 15         |
| 2.2.5 Συναρτήσεις για μέτρηση της τιμής AQI.....    | 16         |
| 2.2.6 Επιπλέον Συναρτήσεις .....                    | 17         |
| <b>Κεφάλαιο 3 .....</b>                             | <b>18</b>  |
| <b>3.1 Μελέτη περίπτωσης - Σύρος.....</b>           | <b>18</b>  |
| <b>3.2 Ανάλυση σωματιδίων pm 25 ανά εποχή .....</b> | <b>20</b>  |
| 3.2.1 Sensor 8200015E (Περιοχή: Κάτω Μάννα) .....   | 20         |
| 3.2.2 Sensor 8200015F (Περιοχή: Λιμάνι) .....       | 25         |
| <b>3.3 Ανάλυση AQI τιμής ανά εποχή.....</b>         | <b>29</b>  |
| 3.3.1 Sensor 820015E (Περιοχή: Κάτω Μάννα) .....    | 29         |
| 3.3.2 Sensor 8200015F (Περιοχή: Λιμάνι) .....       | 35         |
| <b>ΚΕΦΑΛΑΙΟ 4 .....</b>                             | <b>39</b>  |
| <b>4.1 Εισαγωγή.....</b>                            | <b>39</b>  |
| <b>4.2 Ανάλυση AQI ανά εποχή και ζώνη ώρας.....</b> | <b>40</b>  |
| 4.2.1 sensor: 8200015E (Περιοχή: Κάτω Μάννα).....   | 40         |
| 4.2.2 sensor: 8200015F (Περιοχή: Λιμάνι).....       | 50         |
| 4.2.2 Συμπεράσματα .....                            | 58         |
| <b>ΚΕΦΑΛΑΙΟ 5 .....</b>                             | <b>59</b>  |
| <b>5.1 Εισαγωγή.....</b>                            | <b>59</b>  |
| <b>5.2 Multiple Linear Regression .....</b>         | <b>60</b>  |
| <b>5.3 Support Vector Regression (SVR).....</b>     | <b>62</b>  |
| <b>5.4 Regression Tree .....</b>                    | <b>63</b>  |

|                           |                                        |           |
|---------------------------|----------------------------------------|-----------|
| 5.5                       | Random Forest .....                    | 64        |
| 5.6                       | Σχολιασμός και σύγκριση μοντέλων ..... | 65        |
| <b>ΚΕΦΑΛΑΙΟ 6 .....</b>   |                                        | <b>70</b> |
| 6.1                       | Συμπεράσματα .....                     | 70        |
| 6.2                       | Αδυναμίες .....                        | 70        |
| 6.3                       | Προτάσεις Για Περαιτέρω Έρευνα .....   | 71        |
| <b>ΒΙΒΛΙΟΓΡΑΦΙΑ .....</b> |                                        | <b>73</b> |

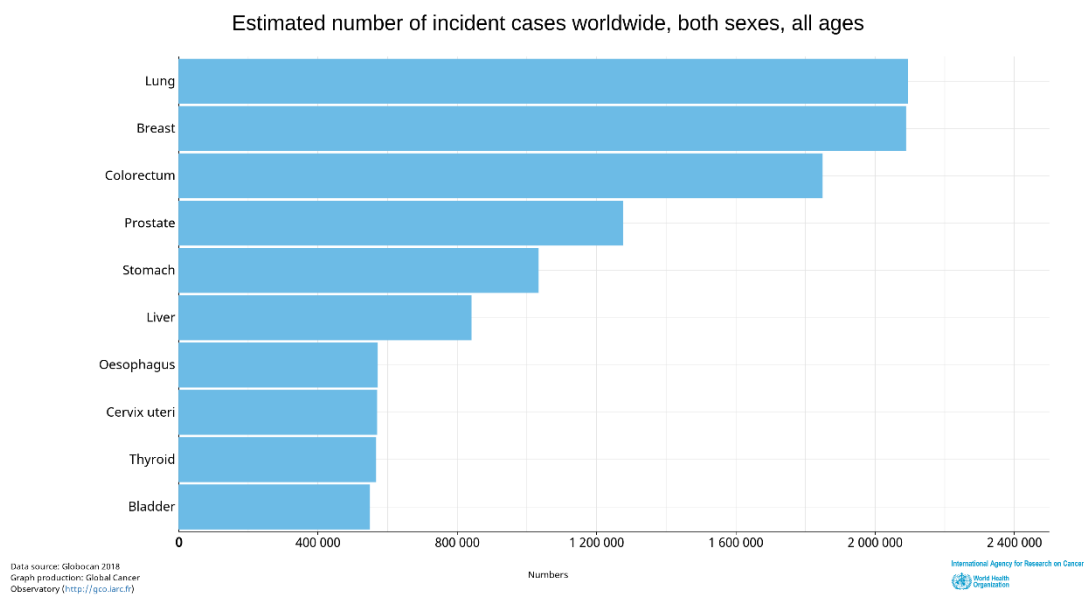
## ΕΙΣΑΓΩΓΗ

### Πρόβλημα: Ατμοσφαιρική ρύπανση

Κατά τον 21<sup>ο</sup> αιώνα η διάχυση των νέων τεχνολογιών δημιούργησε περισσότερες ανάγκες στους ανθρώπους. Αυτές οι ανάγκες δημιουργούν περισσότερη ζήτηση και η επιχειρηματική δραστηριότητα δράττεται της ευκαιρίας με σκοπό να καλύψει αυτή την ζήτηση. Από την δημιουργία, την παραγωγή, την μεταφορά και την κατανάλωση προϊόντων δημιουργούνται ρύποι που επιβαρύνουν την ατμόσφαιρα και το περιβάλλον. Με αυτήν την ανθρώπινη δραστηριότητα σύμφωνα με τον παγκόσμιο οργανισμό υγείας [1] ο κόσμος γίνεται πιο ζεστός και πιο συνωστισμένος. Οι εκπομπές ρύπων της βιομηχανίας, η ανάγκη για φθηνά καύσιμα από ανθρώπους που χρησιμοποιούν και η χρήση αναχρονιστικής τεχνολογίας για θέρμανση είναι μερικοί παράγοντες που καθιστούν τον αέρα μολυσμένο και επικίνδυνο.

Οι επιπτώσεις της δραστηριότητας αυτής στην ατμόσφαιρα δημιουργούν σοβαρά προβλήματα όχι μόνο στο περιβάλλον αλλά και στην υγεία του ανθρώπου. Όπως αναφέρεται στον παγκόσμιο οργανισμό υγείας, το ένα τρίτο των θανάτων από εγκεφαλικό επεισόδιο, καρκίνο του πνεύμονα, καρδιοπάθειες οφείλεται στην ρύπανση του αέρα [2]. Έχει υπολογιστεί ότι οι εκπομπές ρύπων του αέρα προκαλούν 7 εκατομμύρια θανάτους παγκοσμίως. Στην **εικόνα 1** φαίνεται ότι ο καρκίνος του πνεύμονα ταλανίζει περισσότερο τους ανθρώπους και η ρύπανσης της ατμόσφαιρας είναι μία σημαντική αιτία στην καταγραφή αυτή.

Η ρύπανση του αέρα αποτελεί παγκόσμια απειλή όσον αφορά την κλιματική αλλαγή καθώς επιφέρει διαρκή αύξηση της θερμοκρασίας κατά περίπου δύο βαθμούς κελσίου πάνω από τα προβιομηχανικά επίπεδα. Αυτό οδήγησε σε συμφωνία 195 κρατών μελών της UNFCC, γνωστή ως η «Συμφωνία του Παρισιού» με σκοπό να πάρουν δραστικά μέτρα και να συρρικνώσουν την απειλή αυτή. Αξίζει να σημειωθεί ότι στόχοι που τέθηκαν στο Παρίσι πρόκειται να κοστίσουν σχεδόν το 1% του παγκόσμιου ακαθάριστου εγχώριου προϊόντος [3].



Εικόνα 1: International Agency for Research on Cancer / Cancer Today [2]

Παγκοσμίως εκφράζεται η ανάγκη για παρακολούθηση αυτού του φαινομένου, ανάλυσης των δεδομένων που λαμβάνουμε αλλά και δημιουργία μοντέλων πρόγνωσης και πρόβλεψης με σκοπό τα χημικά στοιχεία των ρύπων του αέρα να βρίσκονται σε επιθυμητά επίπεδα. Στην Ευρώπη υπάρχουν πάνω από 7000 ιστοσελίδες παρακολούθησης [4] ανάλογων δεδομένων και αυτό επιβεβαιώνεται και από την αναφορά της ευρωπαϊκής επιτροπής προς το ευρωπαϊκό κοινοβούλιο [5] κάνοντας σαφές την ανάγκη για χρήση της τεχνολογίας με σκοπό να επωφεληθεί η κοινωνία και το περιβάλλον από την ενημέρωση ή/και την μείωση των ατμοσφαιρικών ρύπων. Στον παγκόσμιο οργανισμό υγείας [6] αναφέρονται πολλά παραδείγματα επιτυχημένων πολιτικών. Οι τομείς των μεταφορών, του πολεοδομικού σχεδιασμού, της παραγωγής ενέργειας και της βιομηχανίας αποτελούν μερικά από τα παραδείγματα που επιδεινώνει την ατμοσφαιρική ρύπανση και αναδεικνύουν την ανάγκη επίλυσης των προβλημάτων αυτών με «έξυπνους» τρόπους.

Σήμερα, η ανάπτυξη της τεχνολογίας βελτίωσε την υπολογιστική ισχύ και αυτό επιτρέπει τα δεδομένα ποιότητας αέρα που συλλέγονται από αισθητήρες να αναλύονται και να επεξεργάζονται γρήγορα με σκοπό την εξόρυξη γνώσης για την λήψη αποφάσεων. Η στατιστική ανάλυση μας επιτρέπει να ταξινομήσουμε, να επεξεργαστούμε, να αναλύσουμε και να ερμηνεύσουμε τα δεδομένα τα ήδη υπάρχοντα δεδομένα. Η ερμηνεία

των δεδομένων αποτελεί σημαντικό εργαλείο αλλά εξίσου σημαντική καθίσταται η δράση που πρέπει να γίνει έχοντας αυτή την γνώση.

Αρχικά, η πρόβλεψη είναι μία από τις σημαντικότερες δράσεις που μπορούν να γίνουν. Η ανάπτυξη αλγορίθμων μηχανικής μάθησης που μπορούν να μαθαίνουν από δεδομένα ποιότητας του αέρα και να κάνουν τις ανάλογες προβλέψεις. Έχοντας πρόγνωση για την ποιότητα του αέρα, η κοινωνία, ο επιχειρηματικός κόσμος και το κράτος μπορούν να λάβουν κατάλληλα μέτρα προστασίας. Κάνοντας πρόβλεψη, με κάποιο βαθμό αβεβαιότητας και έχοντας, όσο πιο άμεσα γίνεται, αυτή τη πρόβλεψη, τότε ο άνθρωπος μπορεί να προχωρήσει σε αναγκαίες στρατηγικές δράσεις υπολογίζοντας μάλιστα και το κόστος που θα χρειαστεί για να έρθει εις πέρας μία τέτοια στρατηγική.

Αξίζει να σημειωθεί σε αυτό το σημείο ότι, ο παγκόσμιος οργανισμός υγείας προτείνει δράσεις που μπορούν να κάνουν οι άνθρωποι για να περιορίσουν την εισπνοή μολυσμένου αέρα [7] και να προστατεύσουν έγκαιρα τους εαυτούς τους. Είναι σημαντικό για την κοινωνία να λάβει γνώση αυτής της σημαντικής απειλής και να εκπαιδευτεί κατάλληλα με σκοπό να προστατεύσει τον εαυτό του. Ειδικά, άτομα με αναπνευστικά προβλήματα θα πρέπει να ενημερώνονται καθημερινά για τις τιμές της ποιότητας του αέρα και με βάση την εκπαίδευση που έχουν λάβει, να προφυλαχτούν καταλλήλως.

Η παρούσα εργασία έχει στόχο να παρουσιάσει και να αναδείξει το πρόβλημα της ρύπανσης χρησιμοποιώντας δεδομένα ποιότητας αέρα που έχουν καταγραφεί στις Κυκλάδες και συγκεκριμένα στο νησί της Σύρου.

Στο Κεφάλαιο 1 αναδεικνύονται οι βασικές μετρικές που χρησιμοποιούνται για τον προσδιορισμό της ρύπανσης του αέρα. Κάθε δείκτης, ανάλογα με την τιμή του δείχνει το εύρος που θα πρέπει να κυμαίνονται και καθορίζει το εύρος των υγιεινών και των ανθυγιεινών τιμών.

Στο Κεφάλαιο 2 εισάγεται η μελέτη περίπτωσης στο νησί της Σύρου και παρουσιάζονται οι συναρτήσεις που δημιουργήθηκαν με στόχο την οργάνωση των δεδομένων, το χωρισμό του ανά περιοχή που είναι τοποθετημένοι οι αισθητήρες, την οπτικοποίηση των δεδομένων σε διαγράμματα που παρέχουν σημαντική πληροφορία κ.α.

Στο Κεφάλαιο 3, ξεκινάει η ανάλυση των δεδομένων της Σύρου ανά εποχή. Συγκεκριμένα παρουσιάζονται: η συμπεριφορά των σωματιδίων (particular matter-PM2.5) και οι τιμές

Air Quality Index (AQI) που έχουν καταγραφεί την περίοδο 6 Οκτωβρίου 2018 έως 12 Ιουλίου 2019.

Στο Κεφάλαιο 4, επεκτείνεται η ανάλυση προσδιορίζοντας τις τιμές AQI για κάθε εποχή ανά ζώνη ώρας (πρωί, μεσημέρι, απόγευμα και βράδυ). Στόχος του συγκεκριμένου κεφαλαίου είναι αναδείξει ότι το πρόβλημα της ρύπανσης είναι ανάλογο με την ένταση της ανθρώπινης δραστηριότητας.

Στο Κεφάλαιο 5, εισάγονται τέσσερις μέθοδοι μηχανικής μάθησης. Αυτά τα μοντέλα δημιουργούνται και εκπαιδεύονται από τα δεδομένα ποιότητας αέρα και έχουν στόχο την προσέγγιση των τιμών AQI. Στόχος είναι ναδειχθεί ότι τα μοντέλα πρόβλεψης μηχανικής μάθησης είναι ικανά να αποδώσουν ικανοποιητικά στην αντιμετώπιση του προβλήματος της ρύπανσης το αέρα. Στο τέλος του κεφαλαίου σχολιάζονται τα μοντέλα για την απόδοσή τους.

Κλείνοντας, στο Κεφάλαιο 6, γίνεται μια ανασκόπηση της έρευνα στην μελέτη περίπτωσης της Σύρου. Παρουσιάζονται δυνατότητες και αδυναμίες και γίνεται πρόταση για περαιτέρω έρευνα και ανάπτυξη που στόχο έχει την αποτελεσματική μείωση των ρύπων που ταλανίζουν την κοινωνία και το περιβάλλον.

## ΚΕΦΑΛΑΙΟ 1

### Περιγραφή Δεδομένων ποιότητας του αέρα:

Για την επίτευξη της ανάλυσης των δεδομένων ποιότητας αέρα και την κατανόηση του μεγέθους του προβλήματος, αξίζει να παρουσιαστούν μερικές από τις μετρικές ποιότητας του αέρα καθώς και το εύρος τιμών που τις χαρακτηρίζουν. Στόχος αυτού του κεφαλαίου είναι να παρουσιάσει τα στοιχεία αυτά καθώς ποιες τιμές είναι αποδεκτές και κατάλληλες για τον ανθρώπινο οργανισμό, την κοινωνία και το περιβάλλον και ποιες όχι.

- **Δείκτης AQI**

Το AQI είναι ένας δείκτης ο οποίος αναφέρεται στην ημερήσια ποιότητα του αέρα. Βασικός σκοπός είναι η κατανόηση για το πόσο μολυσμένη είναι η ατμόσφαιρα αλλά και τι συνέπειες επιφέρει στην υγεία μία τέτοια μέτρηση. Παίρνει τιμές, φυσικούς αριθμούς  $N$  από 1 έως 500 και ο **πίνακας 1** δείχνει τις κατηγορίες που υπάρχουν ανάλογα με τις τιμές του AQI (Good, Moderate, Unhealthy for Sensitive Groups, Unhealthy, Very Unhealthy, Hazardous) [9]. Ο δείκτης αυτός αποτελεί μια τμηματική γραμμική συνάρτηση της συγκέντρωσης των ρύπων με πεδίο ορισμού το κλειστό διάστημα των φυσικών αριθμών  $[1,500]$ . Το όριο μεταξύ των κατηγοριών AQI, είναι ένα ασυνεχές άλμα μιας μονάδας AQI. Η συνάρτηση υπολογισμού που χρησιμοποιείται είναι:

$$I = \frac{I_{high} - I_{low}}{C_{high} - C_{low}} (C - C_{low}) + I_{low}$$

Όπου:

- $I$  = Δείκτης Ποιότητας Αέρα,
- $C$  = Η συγκέντρωση ρύπων,
- $C_{low}$  = Το σημείο διακοπής της συγκέντρωσης που είναι  $\leq C$ ,
- $C_{high}$  = Το σημείο διακοπής της συγκέντρωσης που είναι  $\geq C$ ,
- $I_{low}$  = Το σημείο διακοπής του δείκτη που αντιστοιχεί στο  $C_{low}$ ,
- $I_{high}$  = Το σημείο διακοπής του δείκτη που αντιστοιχεί στο  $C_{high}$ ,

Το πρακτορείο περιβαλλοντικής προστασίας (EPA) υπολογίζει το AQI για πέντε κύριους ατμοσφαιρικούς ρύπους όπως ορίζεται από τον ομοσπονδιακό νόμο για τον καθαρό αέρα. Αυτοί είναι: Τα σωματίδια PM, το όζον εδάφους (ground-level ozone- $O_3$ ), το διοξείδιο του

αζώτου (NO<sub>2</sub>), διοξείδιο του θείου (SO<sub>2</sub>) και το μονοξείδιο του άνθρακα (CO). Ο πίνακας 2 δείχνει τα σημεία διακοπής των ατμοσφαιρικών ρύπων.

| Air Quality Index Levels of Health Concern | Numerical Value | Meaning                                                                                                                                                                        |
|--------------------------------------------|-----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Good                                       | 0 to 50         | Air quality is considered satisfactory, and air pollution poses little or no risk.                                                                                             |
| Moderate                                   | 51 to 100       | Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution. |
| Unhealthy for Sensitive Groups             | 101 to 150      | Members of sensitive groups may experience health effects. The general public is not likely to be affected.                                                                    |
| Unhealthy                                  | 151 to 200      | Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects.                                                       |
| Very Unhealthy                             | 201 to 300      | Health alert: everyone may experience more serious health effects.                                                                                                             |
| Hazardous                                  | 301 to 500      | Health warnings of emergency conditions. The entire population is more likely to be affected.                                                                                  |

Πίνακας 1: AQI Classification values developed by EPA [8]

| O <sub>3</sub> (ppb)                            | O <sub>3</sub> (ppb)                            | PM <sub>2.5</sub> (μg/m <sup>3</sup> )          | PM <sub>10</sub> (μg/m <sup>3</sup> )           | CO (ppm)                                        | SO <sub>2</sub> (ppb)                           | NO <sub>2</sub> (ppb)                           | AQI                                       | AQI                            |
|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------|--------------------------------|
| <i>C<sub>low</sub> - C<sub>high</sub> (avg)</i> | <i>C<sub>low</sub> - C<sub>high</sub> (avg)</i> | <i>C<sub>low</sub> - C<sub>high</sub> (avg)</i> | <i>C<sub>low</sub> - C<sub>high</sub> (avg)</i> | <i>C<sub>low</sub> - C<sub>high</sub> (avg)</i> | <i>C<sub>low</sub> - C<sub>high</sub> (avg)</i> | <i>C<sub>low</sub> - C<sub>high</sub> (avg)</i> | <i>I<sub>low</sub> - I<sub>high</sub></i> | Category                       |
| 0-54 (8-hr)                                     | -                                               | 0.0-12.0 (24-hr)                                | 0-54 (24-hr)                                    | 0.0-4.4 (8-hr)                                  | 0-35 (1-hr)                                     | 0-53 (1-hr)                                     | 0-50                                      | Good                           |
| 55-70 (8-hr)                                    | -                                               | 12.1-35.4 (24-hr)                               | 55-154 (24-hr)                                  | 4.5-9.4 (8-hr)                                  | 36-75 (1-hr)                                    | 54-100 (1-hr)                                   | 51-100                                    | Moderate                       |
| 71-85 (8-hr)                                    | 125-164 (1-hr)                                  | 35.5-55.4 (24-hr)                               | 155-254 (24-hr)                                 | 9.5-12.4 (8-hr)                                 | 76-185 (1-hr)                                   | 101-360 (1-hr)                                  | 101-150                                   | Unhealthy for Sensitive Groups |
| 86-105 (8-hr)                                   | 165-204 (1-hr)                                  | 55.5-150.4 (24-hr)                              | 255-354 (24-hr)                                 | 12.5-15.4 (8-hr)                                | 186-304 (1-hr)                                  | 361-649 (1-hr)                                  | 151-200                                   | Unhealthy                      |
| 106-200 (8-hr)                                  | 205-404 (1-hr)                                  | 150.5-250.4 (24-hr)                             | 355-424 (24-hr)                                 | 15.5-30.4 (8-hr)                                | 305-604 (24-hr)                                 | 650-1249 (1-hr)                                 | 201-300                                   | Very Unhealthy                 |
| -                                               | 405-504 (1-hr)                                  | 250.5-350.4 (24-hr)                             | 425-504 (24-hr)                                 | 30.5-40.4 (8-hr)                                | 605-804 (24-hr)                                 | 1250-1649 (1-hr)                                | 301-400                                   | Hazardous                      |
| -                                               | 505-604 (1-hr)                                  | 350.5-500.4 (24-hr)                             | 505-604 (24-hr)                                 | 40.5-50.4 (8-hr)                                | 805-1004 (24-hr)                                | 1650-2049 (1-hr)                                | 401-500                                   |                                |

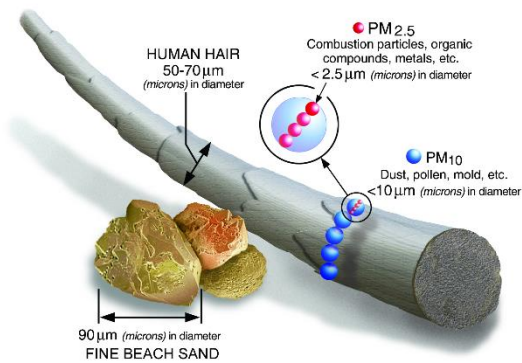
Πίνακας 2: EPA's table of breakpoints [8]

- **Σωματίδια (Particulate matter - PM)**

Τα PM αντιπροσωπεύουν την σωματιδιακή ύλη. Πρόκειται για μείγμα στερεών σωματιδίων και υγρών σταγονιδίων τα οποία βρίσκονται στον αέρα. Αποτελούν μία σημαντική μορφή ατμοσφαιρικής ρύπανσης. Τέτοιου είδους σωματίδια είναι η σκόνη, η βρωμιά και ο καπνός και συχνά είναι δύσκολο να τα παρατηρήσει ο άνθρωπος με γυμνό μάτι [10]. Γενικά περιλαμβάνουν άμεσες εκπομπές από μία πηγή όπως ένα εργοτάξιο, μία καπνοδόχο φούρνου ή και από συνθέσεις χημικών αντιδράσεων που εκπέμπονται από σταθμούς ηλεκτροπαραγωγής, βιομηχανικής παραγωγής [11].

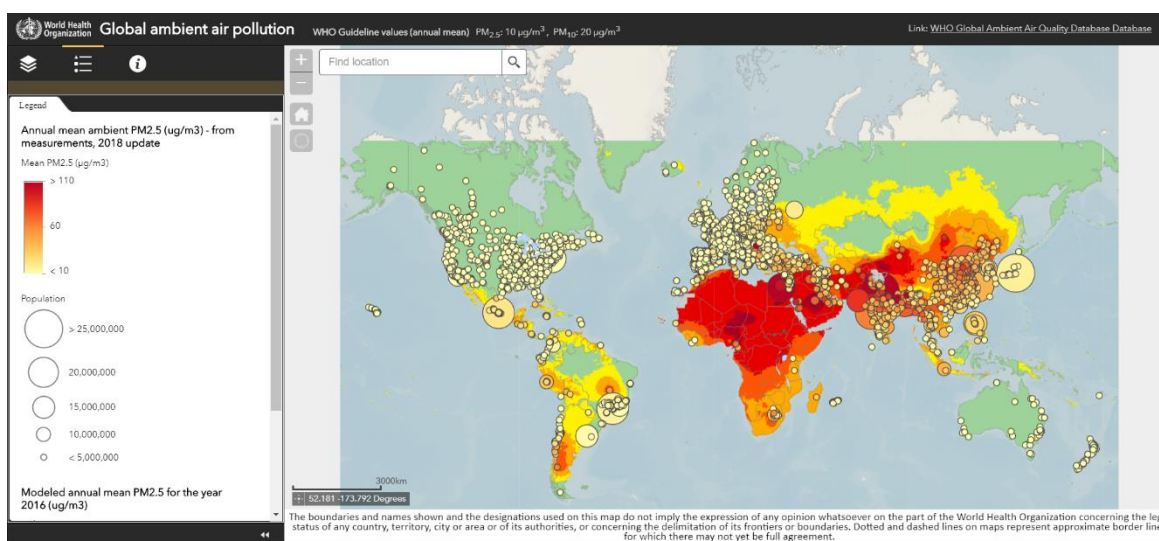
Τα σωματίδια αυτά ομαδοποιούνται ανάλογα με την διάμετρό τους όπως φαίνονται στην **εικόνα 2**. Τα πιο κοινά σωματίδια είναι:

- $PM_{10}$ : Ο δείκτης χαρακτηρίζει την διάμετρο (εδώ  $10\mu m/m^3$  ή μικρότερη). Προκαλεί σοβαρά πνευμονολογικά προβλήματα και είναι ικανά να εισέλθουν στην κυκλοφορία του αίματος. Οι κατάλληλες τιμές [13] που αναδεικνύουν τα όρια των τιμών, δίνονται με βάση τις οδηγίες του παγκόσμιου οργανισμού υγείας. Συγκεκριμένα:
  - $10\mu g/m^3$  ετήσιος μέσος όρος
  - $25\mu g/m^3$  μέσος όρος 24-ωρών
- $PM_{2.5}$ : Λεπτά εισπνεύσιμα αιωρούμενα σωματίδια με διάμετρο μικρότερη ή ίση των  $10\mu m/m^3$ . Η έκθεση σε τέτοιου είδους σωματίδια μπορούν να δημιουργήσουν, σύμφωνα με τον WHO, καρκίνο. Επίσης είναι ικανά να προκαλέσουν μειωμένη ορατότητα, όπως έχει φανεί σε μέρη όπως η Ηνωμένες Πολιτείες, πολύτιμα εθνικά πάρκα και περιοχές άγριας φύσης [13]. Επιτρεπτές τιμές:
  - $20\mu g/m^3$  ετήσιος μέσος όρος
  - $50\mu g/m^3$  μέσος όρος 24-ωρών



Εικόνα 2: Σύγκριση μεγέθους PM σωματιδίου με το μέγεθος μιας τρίχας ενός ανθρώπου

Στην **Εικόνα 3** μπορούμε να δούμε μία απεικόνιση της συγκέντρωσης σωματιδίων  $PM_{2.5}$  στον παγκόσμιο χάρτη που δημιούργησε ο παγκόσμιος οργανισμός υγείας με σκοπό να αναδείξουν το πρόβλημα της ρύπανσης του αέρα. Αξίζει να σημειωθεί ότι η **Εικόνα 3** αποτελεί τον ετήσιο μέσο όρο των τιμών  $PM_{2.5}$ . Επιβεβαιώνεται με αυτόν τον τρόπο η ανάγκη της επαγρύπνησης της κοινωνίας με σκοπό την αντιμετώπιση αυτής της απειλής αλλά και η ανάγκη για μελέτη ατμοσφαιρικών μοντέλων και παροχή κανονιστικών και συμβουλευτικών οδηγιών για τους κατοίκους, τις εταιρίες μεταφοράς και τις βιομηχανίες.



Εικόνα 3: Annual mean ambient PM<sub>2.5</sub> (µg/m<sub>3</sub>) - from measurements, 2018 update (WHO)

- **Ozone (O<sub>3</sub>)**

Το όζον σχηματίζεται στην ατμόσφαιρα με φωτοχημικές αντιδράσεις με παρουσία ηλιακού φωτός και πρόδρομων ρύπων όπως είναι τα οξείδια του αζώτου (NO<sub>x</sub>) και οι πτητικές οργανικές ενώσεις (Volatile Organic Compounds – VOCs). Μελέτες έχουν δείξει ότι καθώς οι συγκεντρώσεις του όζοντος αυξάνονται πάνω από την κατευθυντήρια τιμή που έχει ορίσει ο παγκόσμιος οργανισμός υγείας - WHO, οι επιπτώσεις στην υγεία στο επίπεδο του πληθυσμού γίνονται πολυάριθμες και σοβαρές [13].

Σύμφωνα με παρατηρήσεις που έχουν γίνει, φαίνεται ότι αργά το απόγευμα οι τιμές φτάνουν στην κορυφή τους. Η συσχέτιση του όζοντος με την θερμοκρασία και τον αέρα είναι άμεση, καθώς οι μέγιστες τιμές φαίνονται όταν υπάρχει παράλληλα υψηλή θερμοκρασία και ο αέρας είναι στάσιμος.

Οι επιτρεπόμενες τιμές του όζοντος (O<sub>3</sub>) σύμφωνα με την κατευθυντήρια οδηγία του WHO είναι 100µg/m<sub>3</sub> στην κάτω ατμόσφαιρα.

Η έκθεση σε μακροπρόθεσμη έκθεση σε O<sub>3</sub> υψηλών τιμών μπορεί να προκαλέσει προβλήματα στην ανάπτυξη των πνευμόνων των παιδιών [12].

- **Διοξείδιο του αζώτου (NO<sub>2</sub>)**

Αποτελεί τοξικό αέριο με σημαντικές επιπτώσεις στην υγεία. Τοξικολογικές μελέτες σε ζώα υποδεικνύουν επίσης ότι η μακροχρόνια έκθεση σε NO<sub>2</sub> σε συγκεντρώσεις πάνω από τις τρέχουσες συγκεντρώσεις στο περιβάλλον έχει δυσμενείς επιπτώσεις. Οι κατευθυντήριες οδηγίες θέτουν τις τιμές του NO<sub>2</sub> σε:

- 40 µg/m<sub>3</sub> ετήσιο μέσο όρο και
- 200 µg/m<sub>3</sub> ωριαίο μέσο όρο [13].

Έχει άμεση συμβολή στις τιμές του όζοντος O<sub>3</sub> καθώς και των σωματιδίων PM [12].

- **Διοξείδιο του θείου (Sulfur dioxide SO<sub>2</sub>)**

Υπάρχει σημαντική αβεβαιότητα ως προς το αν το SO<sub>2</sub> είναι ο ρύπος που είναι υπεύθυνος για τις παρατηρούμενες δυσμενείς επιπτώσεις ή αν είναι υποκατάστατο για πολύ λεπτά σωματίδια PM<sub>2.5</sub> μικρότερης διαμέτρου των 2.5µm ή κάποια άλλη συσχετιζόμενη ουσία [12]. Παρ'όλα αυτά οι οδηγίες του WHO σχετικά με τις επιτρεπόμενες τιμές του διοξειδίου του θείου είναι:

- 20 µg/m<sub>3</sub>, μέσος όρος 24 ωρών
- 500 µg/m<sub>3</sub>, μέσος όρος 10 λεπτών

Οι παραπάνω τιμές αποτελούνε προληπτικές οδηγίες καθώς το διοξείδιο του θείου δεν σχετίζονται με τα σωματίδια (PM)

- **Πτητική οργανική ένωση (VOC)**

Είναι οργανικές ενώσεις που έχουν υψηλή τάση ατμών σε θερμοκρασία δωματίου. Οι Faisal I. Khan και Alope Kr. Ghoshal αναγνωρίζουν ότι, από περιβαλλοντικής απόψεως, είναι απαραίτητο αυτές οι ενώσεις να περιοριστούν και να ελέγχονται οι εκπομπές ατμών, λόγω του ότι επηρεάζουν την αλλαγή του κλίματος, την ανάπτυξη της χλωρίδας και της πανίδας της γης, εξετάζοντας τις διαθέσιμες τεχνικές για τον έλεγχο των VOCs [14]. Οι ενώσεις αυτές αφορούν την μόλυνση του αέρα τόσο εξωτερικού όσο και εσωτερικού χώρου. Παραδείγματα τέτοιων ενώσεων είναι η φορμαλδεΐδη, η ακετόνη, η αιθανόλη (αιθυλική αλκοόλη), η 2-προπανόλη (ισοπροπυλική αλκοόλη) και άλλα [15]. Μερικά από αυτά τα VOCs - για παράδειγμα, η φορμαλδεΐδη και το βενζόλιο - θεωρούνται από πολλές αρχές αποδεδειγμένα καρκινογόνα.

- **Φορμαλδεΐδη (CH<sub>2</sub>O)**

Η μεθανάλη ή φορμαλδεΐδη είναι οργανική χημική ένωση άνθρακα, οξυγόνου και υδρογόνου, με χημικό τύπο CH<sub>2</sub>O. Εξαιτίας της ευρύτατης χρήσης της για την δημιουργία νέων υλικών, όπως επίπλων, της σημαντικής τοξικότητας και ευφλεκτότητας, η έκθεση στη μεθανάλη αποτελεί ένα σημαντικό κίνδυνο για την ανθρώπινη υγεία με καρκινογόνες συνέπειες. Αρχικά ο Παγκόσμιος Οργανισμός Υγείας θεώρησε την φορμαλδεΐδη ως ρύπος εσωτερικού χώρου θέτοντας ως κατευθυντήρια τιμή τα 0,1mg/m<sup>3</sup> (0.08ppm)[16]. Το 2010, το WHO επιβεβαίωσε την τιμή αυτή [17]. Σημαντική παράμετρος για την τιμή της φορμαλδεΐδης είναι η υγρασία και η θερμοκρασία.

- **Διοξείδιο του Άνθρακα (CO<sub>2</sub>)**

Άχρωμο, άοσμο, μη εύφλεκτο αέριο το οποίο υπάρχει εκ φύσεως στην ατμόσφαιρα. Παράγεται από το μεταβολισμό του σώματος και είναι φυσιολογικό συστατικό της ζωής όπως η εκπνοή. Μεγαλύτερες ποσότητες CO<sub>2</sub> εκπίπτει και από ανθρώπινη δραστηριότητα όπως για παράδειγμα η καύση ορυκτών καυσίμων.

Σύμφωνα με το υπουργείο γεωργίας των Ηνωμένων Πολιτειών, τα επίπεδα CO<sub>2</sub> στον εξωτερικό αέρα κυμαίνονται συνήθως από 300 έως 400ppm. Σε μητροπολιτικές περιοχές τα επίπεδα CO<sub>2</sub> μπορεί να φτάσουν ακόμα και σε τιμές 600 έως 900 ppm. [18]

Οι επιδράσεις στον ανθρώπινο οργανισμό δεν είναι ιδιαίτερα σημαντικές καθώς το CO<sub>2</sub> θεωρείται ελάχιστα τοξικό. Συμπτώματα σε ήπια έκθεση είναι ο πονοκέφαλος και η υπνηλία. Ενώ για υψηλότερα επίπεδα τα συμπτώματα είναι η ταχεία αναπνοή, σύγχυση, αυξημένη αρτηριακή πίεση και αυξημένες καρδιακές αρρυθμίες. Θα ήταν σημαντικό και θα βοηθούσε πολλούς ανθρώπους με αναπνευστικά, πνευμονολογικά και καρδιακά προβλήματα, να ενημερώνονται για τις τιμές του διοξειδίου του άνθρακα που υπάρχουν στην ατμόσφαιρα καθώς και στα σημεία που .

- **Μετεωρολογικά δεδομένα**

Οι μετεωρολογικές καταστάσεις ενός σημείου παρατήρησης, αποτελούν ισχυρά δεδομένα για την ανάλυση της ποιότητας του αέρα [19]. Η χαμηλή θερμοκρασία του περιβάλλοντος και εισερχόμενη ακτινοβολία του ήλιου επιβραδύνουν και μειώνουν τους ατμοσφαιρικούς ρύπους [20]. Η ταχύτητα του αέρα παίζει σημαντικό ρόλο στην στιγμιαία κατάσταση της ποιότητας του αέρα.

- Ο στάσιμος αέρα επηρεάζει άμεσα μετρικές όπως αυτή του όζοντος που σε συνδυασμό με την πιθανή υψηλή θερμοκρασία μπορεί να δούμε πολύ υψηλές τιμές.
- Όταν υπάρχει χαμηλή ταχύτητα του αέρα, τότε οι ρύποι του αέρα πιθανώς να απομακρυνθούν από την περιοχή που παρατηρούμε [21].
- Οι ισχυρές ταχύτητες του ανέμου μπορεί να προκαλέσουν καταιγίδες σκόνης καθώς ανυψώνουν τα σωματίδια του εδάφους [22].

Εξίσου σημαντικά είναι τα περιβαλλοντικά δεδομένα που επηρεάζουν τις τιμές της φορμαλδεΐδης.

## ΚΕΦΑΛΑΙΟ 2

### 2.1 Εισαγωγή

Στην παρούσα εργασία γίνεται μια προσπάθεια μελέτης των δεδομένων ποιότητας του αέρα. Πιο συγκεκριμένα, αυτά τα δεδομένα αφορούν την ποιότητα του αέρα της Σύρου. Τα δεδομένα που υπάρχουν, συλλέγονται μέσα από τρεις αισθητήρες και θα τους αναγνωρίζουμε με το id του κάθε αισθητήρα. Και οι τρεις αισθητήρες έχουν ξεκινήσει την συλλογή δεδομένων σχετικά πρόσφατα, συγκριτικά με την ημερομηνία που υλοποιείται η παρούσα εργασία, καθώς δεν έχει ολοκληρωθεί ένας χρόνος από την έναρξή τους. Αυτό έχει ως αποτέλεσμα ότι οι αναλύσεις, οι πιθανότητες που προκύπτουν και τα μοντέλα πρόβλεψης που θα εξάγουμε δεν μπορούν να γενικοποιηθούν, να βγάλουμε άμεσα και ασφαλή συμπεράσματα και να προχωρήσουμε σε δράσεις βελτίωσης της ποιότητας του αέρα στην Σύρο.

Κάνοντας ανασκόπηση σε προηγούμενες αντίστοιχες μελέτες, μπορούμε να δούμε ότι το σύνολο των δεδομένων που χρησιμοποιήθηκε αφορούν μία συλλογή τουλάχιστον τριών ετών. Οι San Martini, F. M., Hasenkopf, C. A., και Roberts, D. C. [11] το 2015 κάνουν μια συνολική έρευνα σε παγκόσμια κλίμακα των παρατηρούμενων σωματιδίων  $pm_{2.5}$  σε παγκόσμια κλίμακα κάνοντας λόγο για συλλογή δεδομένων πολλών ετών. Αξίζει να σημειωθεί ότι στην Κίνα συλλέγονται δεδομένα για πάνω από δέκα χρόνια και σε πολλά γεωγραφικά τμήματα της χώρας από επίσημους κρατικούς φορείς.

Γεννάται λοιπόν το εύλογο ερώτημα: «Γιατί να προχωρήσουμε σε μία έρευνα όπου τα δεδομένα είναι σχετικά λίγα και δεν μπορούν άμεσα να εξαχθούν σαφή συμπεράσματα;» Είμαστε σε θέση να κάνουμε μία ανασκόπηση της χρονιάς που διανύουμε και να ορίσουμε το πρόβλημα και να τονίσουμε την σημαντικότητά της.

Σε συζητήσεις με ανθρώπους που διαμένουν στην Σύρο και ενδιαφέρονται για το πρόβλημα της ποιότητας του αέρα, γίνεται αντιληπτό ότι η Σύρος έχει μια βαριά βιομηχανία στον αστικό ιστό της. Επίσης, η Ερμούπολη, η πρωτεύουσα της Σύρου, είναι αρκετά πυκνοκατοικημένη και αυτό έχει ως αποτέλεσμα να υπάρχουν πολλά οχήματα για το μέγεθός της. Επιπροσθέτως, έχει ένα αρκετά δημοφιλές λιμάνι όπου πολλά πλοία, κρουαζιερόπλοια και άλλα μηχανοκίνητα θαλάσσια οχήματα δραστηριοποιούνται σε αυτό. Άξιο αναφοράς είναι ότι, αρκετά συχνά, πολλά πλοία καίνε καύσιμα κατά τη

διάρκεια της νύχτας για να ηλεκτροδοτούνται. Τέτοιου είδους δραστηριότητες επηρεάζουν αρκετά το περιβάλλον και την ποιότητα του αέρα έχοντας ως αποτέλεσμα η κοινωνία της Ερμούπολης να επιβαρύνεται ακόμα περισσότερο από το πρόβλημα αυτό. Η επιβάρυνση αυτή στην κοινωνία φαίνεται τα τελευταία χρόνια καθώς τουλάχιστον ένας άνθρωπος χάνει την μάχη με τον καρκίνο. Έτσι η κοινωνία αρχίζει να αφυπνίζεται και να δηλώνει ενδιαφέρον για την απειλή που αντιμετωπίζουν.

## **2.2     Συναρτήσεις**

Στο κεφάλαιο 3 γίνεται η ανάλυση του προβλήματος της ποιότητας του αέρα της Σύρου. Η ανάλυση γίνεται με χρήση του προγράμματος R-studio και ο κώδικας είναι γραμμένος στη γλώσσα R. Για λόγους ευκολίας προγραμματισμού και καλής γραφής, οι μέθοδοι που χρησιμοποιήθηκαν είναι σε συναρτήσεις με σκοπό να χρησιμοποιηθούν για τα δεδομένα κάθε αισθητήρα αλλά και ανάλυση δεδομένων ποιότητας αέρα από μελλοντικούς αισθητήρες που θα ενεργοποιηθούν στην Σύρο.

### **2.2.1   Καθαρισμός Δεδομένων:**

- `org.data(sensor)`
  - `input`: ο αρχικός πίνακας δεδομένων ποιότητας αέρα (`sensor`).
  - `output`: επιστρέφει έναν οργανωμένο πίνακα όπου τα δεδομένα ποιότητας αέρα είναι οργανωμένα ανά τέταρτο της ώρας.
  - Η συνάρτηση `org.data(sensor)` χρησιμοποιεί την συνάρτηση `hour.fun(subsensor)` δίνοντάς της έναν πίνακα μίας μέρας και παίρνοντας από αυτήν έναν πίνακα τον καλά οργανωμένο πίνακα που επιθυμούμε και τον ενώνει με τον νέο πίνακα τον οποίο επιστρέφει (`newsensor`).
- `hour.fun(subsensor)`
  - `input`: ένα υποσύνολο του αρχικού πίνακα (`sensor`).
  - `output`: ένας πίνακας ο οποίος επιστρέφει τον μέσο όρο των δεδομένων ανά τέταρτο της ώρας. Για παράδειγμα, τον μέσο όρο των δεδομένων `pm2.5`, `voc`, `vocaqi` από την ώρα 14:00 μέχρι 14:15
  - Στην συνάρτηση `org.data(sensor)` δίνει έναν πίνακα δεδομένων που αφορά την ημέρα την οποία ζητάει η `org.data()` να οργανώσει

- Σημείωση: Τυχόν χαμένα δεδομένα αντικαθίστανται ανάλογα με την προηγούμενη εγγραφή λόγω του API του προγράμματος που χρησιμοποιείται από τους αισθητήρες. Ακραίες τιμές λαμβάνονται υπόψιν αλλά μορφοποιούνται ανάλογα με τον μέσο όρο των προηγούμενων δεδομένων.

### 2.2.2 Χωρισμός δεδομένων για ανάλυση:

- epoch (sensor,epc)
  - input: Πίνακας (sensor) και μια δήλωση epc όπου αναφέρεται στην εποχή που θέλουμε να εξετάσουμε και να αναλύσουμε στην συνέχεια. Epc παίρνει μόνο τις τιμές “autumn”, “winter”, “spring” και “summer”.
  - output: Πίνακας που περιέχει δεδομένα που αφορά μόνο την εποχή την οποία δήλωσε ο χρήστης.
- epccsummary(sensor)
  - input: Πίνακας δεδομένων ποιότητας αέρα
  - Χρησιμοποιεί την συνάρτηση summary της γλώσσας R, η οποία δίνει μία σύνοψη των δεδομένων που ανάλογα με την εποχή που επιθυμεί ο χρήστης.
- timezones(sensor, timez)
  - input: Πίνακας (sensor) και μια δήλωση timez όπου αναφέρεται στο διάστημα ζώνης ώρας που θέλουμε να εξετάσουμε και να αναλύσουμε στην συνέχεια. Timez παίρνει μόνο τις τιμές “morning”, “afternoon”, “evening” και “night”.
  - output: Πίνακας που περιέχει δεδομένα που αφορά μόνο την ζώνη ώρας την οποία δήλωσε ο χρήστης.
- hrsummary(sensor)
  - input: Πίνακας δεδομένων ποιότητας αέρα
  - Χρησιμοποιεί την συνάρτηση summary της γλώσσας R, η οποία δίνει μία σύνοψη των δεδομένων που ανάλογα με την εποχή και την ζώνη ώρας που επιθυμεί ο χρήστης.
- mean\_month(sensor,month)
  - input: Πίνακας δεδομένων ποιότητας αέρα, μήνας για εύρεση μέσων τιμών ανά ημέρα.
  - output: Ένας νέος πίνακας που έχει τις μέσες τιμές (mean) για τον συγκεκριμένο μήνα που δόθηκε σαν είσοδο.

### 2.2.3 Συνάρτηση εύρεσης πιθανότητας

- pz(vector, value)
  - input: ένα διάνυσμα που αφορά τις τιμές ποιότητας αέρα (pm10, pm2.5, voc, vocaqi κ.α)
  - output: Η πιθανότητα η τιμή “value” να συμβεί.
  - Στόχος: Να βρούμε την πιθανότητα μία συγκεκριμένη τιμή να λάβει χώρα.
  - Απαντά στην ερώτηση «ποια η πιθανότητα να είναι “καλή” η ποιότητα του αέρα;»

### 2.2.4 Οπτικές Συναρτήσεις (Visuals)

- hist\_ep\_PM25(sensor)
  - input: Πίνακας Δεδομένων
  - output: Δημιουργούμε ιστογράμματα ανά εποχή για να δούμε τις διακυμάνσεις των τιμών pm25. Μια καμπύλη πυκνότητας είναι ένα γράφημα που δείχνει την πιθανότητα. Η περιοχή κάτω από την καμπύλη ισούται με το 100% όλων των πιθανοτήτων. Συνήθως αντιπροσωπεύονται ως δεκαδικά ψηφία, μπορούμε επίσης να πούμε ότι η περιοχή είναι ίση με 1 (100% ως δεκαδικό του 1)
- hist\_hour\_PM25(sensor)
  - input: Πίνακας Δεδομένων
  - output: Δημιουργούμε ιστογράμματα ανά εποχή και ανά ζώνη ωρών για να δούμε τις διακυμάνσεις των τιμών pm25
- hist\_ep\_vocaqi (sensor)
  - input: Πίνακας Δεδομένων
  - output: Δημιουργεί ιστόγραμμα ανά εποχή για να δει ο αναγνώστης τις διακυμάνσεις αλλά και την καμπύλη πυκνότητας. Η καμπύλη πυκνότητας είναι ένα γράφημα που δείχνει την πιθανότητα. Η περιοχή κάτω από την καμπύλη ισούται με το 100% όλων των πιθανοτήτων. Συνήθως αντιπροσωπεύονται ως δεκαδικά ψηφία, μπορούμε επίσης να πούμε ότι η περιοχή είναι ίση με 1 (100% ως δεκαδικό του 1)
- hist\_hour\_vocaqi (sensor)
  - input: Πίνακας Δεδομένων
  - output: Δημιουργούμε histograms ανά εποχή και ανά ζώνη ωρών για να δούμε τις διακυμάνσεις των τιμών vocaqi

- vis\_table(sensor, matrix)
  - input: Πίνακας Δεδομένων, Πίνακα πιθανοτήτων
  - output: Οπτικοποίηση του πίνακα πιθανοτήτων για καλύτερη ανάγνωση από τον αναγνώστη
- graph\_pm25(sensor, season)
  - input: πίνακας δεδομένων ποιότητας αέρα, εποχή για οπτικοποίηση μέσων τιμών (mean)
  - output: Εμφάνιση γραφήματος των μέσων pm25 χαρακτηριστικών ανά ημέρα.
  - Στόχος: Να δούμε τις διαφορές που παρουσιάζουν ανά μήνα οι τιμές pm25.
- graph\_aqi(sensor, season)
  - input: πίνακας δεδομένων ποιότητας αέρα, εποχή για οπτικοποίηση μέσων τιμών (mean) .
  - output: Εμφάνιση γραφήματος των μέσων aqi χαρακτηριστικών ανά ημέρα.
  - Στόχος: Να δούμε τις διαφορές που παρουσιάζουν ανά μήνα οι τιμές aqi.

## 2.2.5 Συναρτήσεις για μέτρηση της τιμής AQI

- aqi1(sensor, min, max)
  - input: πίνακας δεδομένων ποιότητας αέρα, ελάχιστη και μέγιστη τιμή των vocaqi χαρακτηριστικών
  - output: Επιστρέφεται η πιθανότητα να λάβει χώρα μία τιμή στο διάστημα [min,max]
- aqi2(sensor)
  - input: πίνακας δεδομένων ποιότητας αέρα
  - output: επιστρέφει ένα διάνυσμα με τις πιθανότητες ύπαρξης των vocaqi τιμών ανάλογα με τα επίπεδα ανησυχίας που έχει ορίσει ο EPA [8].
  - Απαντάμε στα ερωτήματα
    - Ποια η πιθανότητα να είναι “GOOD” οι τιμές του vocaqi την εποχή x την ώρα γ;
    - Ποια η πιθανότητα να είναι “UNHEALTHY” οι τιμές του vocaqi την εποχή x την ώρα γ;
  - Η Συνάρτηση aqi2() χρησιμοποιεί την aqi1() για να δημιουργήσει το διάνυσμα που θέλουμε

- `aqi_ep_table(sensor)`
  - input: πίνακας δεδομένων ποιότητας αέρα
  - output: Δημιουργεί μία οπτικοποίηση του πίνακα πιθανοτήτων aqi ανάλογα με την εποχή.
- `aqi_hz_table(sensor)`
  - input: πίνακας δεδομένων ποιότητας αέρα
  - output: Δημιουργεί μία οπτικοποίηση του πίνακα πιθανοτήτων aqi ανάλογα με την ζώνη ωρών.

### 2.2.6 Επιπλέον Συναρτήσεις

- `clc_matrix(sensor)`
  - input: πίνακας δεδομένων ποιότητας αέρα
  - output: Επιστρέφει έναν πίνακα που διατηρεί μόνο τα σημαντικά δεδομένα ποιότητας αέρα και όχι τα γνωρίσματα που χαρακτηρίζουν χρόνο, id, γεωγραφικό ύψος, πλάτος κ.α.
- `Backward_Elimination(training_set, sl)`
  - Input: training dataset, significance level value
  - Output: Μοντέλο multiple linear regression το οποίο είναι απαλλαγμένος από παραμέτρους με  $P\_value < Significance\ level$

## Κεφάλαιο 3

### 3.1 Μελέτη περίπτωσης - Σύρος

Στην παρακάτω ανάλυση γίνεται μια προσπάθεια να δούμε και να ερμηνεύσουμε τα διαγράμματα και τους πίνακες που δημιουργήσαμε με τα δεδομένα που έχουμε.

Αρχικά χωρίζουμε το σύνολο των raw data (dt2) που έχουμε σε τρεις πίνακες όπου αντιπροσωπεύουν τα δεδομένα κάθε αισθητήρα με το id τους. Τα id μπορούμε να τα δούμε με στην γλώσσα της R με την συνάρτηση `unique(data$id)` όπως φαίνεται παρακάτω.

Κάθε id αναφέρεται σε αισθητήρα που έχει τοποθετηθεί σε διαφορετικά σημεία στην Σύρο.

```
unique(dt2$id)
[1] "8200015E" "8200015F" "82000160"
```

- id: "8200015E" τοποθετημένος στην περιοχή Κάτω Μάνα.



Η περιοχή της Κάτω Μάννας, δεν είναι ιδιαίτερα πυκνοκατοικημένη αλλά είναι μία περιοχή κοντά στην Ερμούπολη όπου είναι αρκετά βιομηχανική. Για παράδειγμα, υπάρχει εταιρία προμήθειας πετρελαιοειδών, αποσταγμάτων και λιπαντικών. Σε αρκετά κοντινή απόσταση μπορεί κανείς να βρει και το αεροδρόμιο της Σύρου. Ακόμα, στην περιοχή αυτή, στεγάζονται και σχολεία δευτεροβάθμιας εκπαίδευσης του νησιού. (Γενικό Λύκειο, ΕΠΑΛ) Αυτά τα χαρακτηριστικά της περιοχής αποτελούν σημαντικοί λόγοι για την εγκατάσταση του αισθητήρα 8200015E.

- id: "8200015F" τοποθετημένος στην περιοχή του Λιμανιού της Ερμούπολης.

Σύμφωνα με πληροφορίες από το Wikipedia [3], βρίσκεται στη ανατολική πλευρά του νησιού και είναι γνωστή και για το ναυπηγείο της και τον σημαντικό λιμένα της. Παλιότερα υπήρξε βιομηχανικό, ναυτιλιακό και πολιτιστικό κέντρο της Ελλάδας, ενώ πλέον αποτελεί γραφικό θέρετρο.

Η Ερμούπολη βρίσκεται στην ανατολική πλευρά του νησιού της Σύρου και συνδέεται ακτοποϊκά όλο τον χρόνο με τον Πειραιά (καθημερινά), με όλες τις υπόλοιπες Κυκλάδες, τα Δωδεκάνησα, τη Χίο και τη Μυτιλήνη.

- id: "82000160" τοποθετημένος στην περιοχή Ιλίσια της Σύρου.

Στην παρούσα εργασία δεν πραγματοποιείται ανάλυση των δεδομένων του συγκεκριμένου αισθητήρα καθώς αφενός τα δεδομένα είναι λίγα και αφετέρου ο συγκεκριμένος αισθητήρας έχει πάψει να λειτουργεί με το τέλος του έτους 2018.

Παρακάτω δημιουργούμε τρεις πίνακες για κάθε id με την συνάρτηση `filter(data,data$id == id)` της βιβλιοθήκης `dplyr` που υπάρχει στην R

```
> sensor1 = filter(dt2,dt2$id=="8200015E")  
> sensor2 = filter(dt2,dt2$id=="8200015F")  
> sensor3 = filter(dt2,dt2$id=="82000160")
```

Έχοντας απομονώσει τις μετρήσεις, αναγνωρίζουμε πλέον πού αναφέρονται οι εγγραφές, μπορούμε να χρησιμοποιήσουμε τις συναρτήσεις που δημιουργήσαμε και περιγράφονται παραπάνω.

Σημείωση: Οι εγγραφές ξεκινάνε από 6 Οκτωβρίου 2018 για κάθε αισθητήρα. Για τους `sensor1` και `sensor2` τα δεδομένα μας φτάνουν μέχρι τις 12 Ιουλίου 2019 ενώ για τον `sensor3`, η συλλογή των δεδομένων σταματάει στις 28 Δεκεμβρίου 2018. Η παρακάτω ανάλυση θα σταθούνε στα αποτελέσματα που θα παραχθούν από τους αισθητήρες 1 και 2 ("8200015E" και "8200015F").

Στο παρόν κεφάλαιο γίνεται ανάλυση των παρατηρούμενων σωματιδίων `pm25` ανά εποχή με στόχο να φανούν οι κατανομές τους ανά εποχή. Έπειτα, συνεχίζεται η ανάλυση του δείκτη ποιότητας αέρα AQI στο πνεύμα της ανάλυσης των σωματιδίων. Παράλληλα υπολογίζονται και παρουσιάζονται οι πιθανότητες ύπαρξης τιμών AQI ανά εποχή ανάλογα με το εύρος τιμών που τις χαρακτηρίζει ο πίνακας 1.

### 3.2 Ανάλυση σωματιδίων pm 25 ανά εποχή

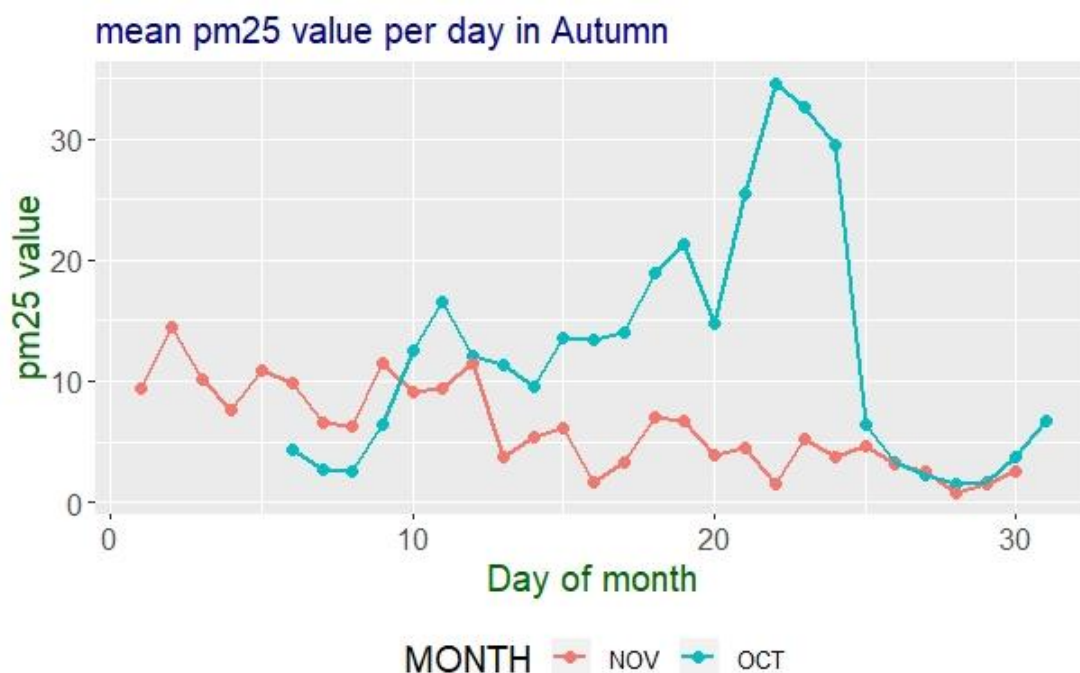
Αρχικά γίνεται ένας διαχωρισμός του χρόνου ανά εποχή (autumn, winter, spring, summer). Στόχος είναι να δούμε πώς κατανέμονται τα χαρακτηριστικά (attributes) της κάθε εγγραφής ανά εποχή. Σε αυτόν τον διαχωρισμό, σημειώνεται ότι το σύνολο των δεδομένων για τον χειμώνα και την άνοιξη είναι ολοκληρωμένο σε εγγραφές ενώ για το φθινόπωρο δεν έχουμε εγγραφές για τον μήνα Σεπτέμβριο και για το καλοκαίρι δεν έχουμε εγγραφές για τον μισό μήνα του Ιουλίου και για τον Αύγουστο.

Με την συνάρτηση `hist_ep_pm25(sensor1)` παίρνουμε πίνακα γραφημάτων των σωματιδίων pm25 και βλέπουμε την κατανομή τους. Στην **Εικόνα 1**: βλέπουμε ότι οι τιμές των σωματιδίων pm25 δεν φτάνουν σε ακραία επίπεδα. Τα σωματίδια αυτά έχουν την ικανότητα να γίνονται αισθητά στην ατμόσφαιρα για έναν παρατηρητή με γυμνό μάτι.

#### 3.2.1 Sensor 8200015E (Περιοχή: Κάτω Μάννα)

Παρακάτω γίνεται η ανάλυση των τιμών pm25 για τον αισθητήρα **8200015E** που φαίνεται στην **εικόνα 1**:

- Το φθινόπωρο φαίνεται να έχει τις πιο υψηλές τιμές φτάνοντας μάλιστα και σε μία ακραία τιμή. Οι τιμές του φθινοπώρου δεν μπορούν να μας δώσουν, για την ώρα, ένα καθαρό συμπέρασμα λόγω του όγκου των δεδομένων.



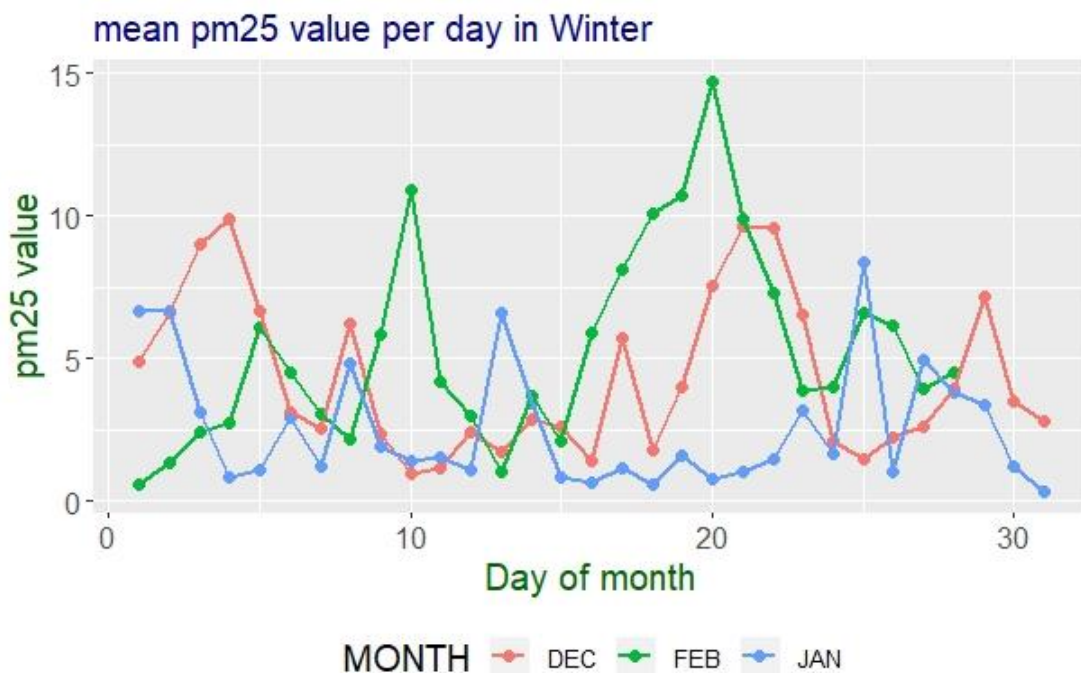
Γράφημα 1: Μέσες τιμές ανά ημέρα για τον αισθητήρα 8200015E την εποχή του φθινοπώρου

```
> summary(epoch(sensor = sensor1,"autumn")$pm25)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000  3.000   7.000   9.196  12.000   56.000
```

- Το χειμώνα οι τιμές δείχνουν να είναι λιγότερο ανησυχητικές. Καθώς, το μεγαλύτερο ποσοστό των τιμών είναι κάτω του 10. Η παρατήρηση της μέγιστης τιμής που ξεπερνάει τα επιτρεπτά όρια των σωματιδίων οφείλεται είτε σε κάποιο τεχνικό λάθος είτε από ένα μεμονωμένο περιστατικό, λάθος καταγραφή, το οποίο δεν πρέπει να προκαλεί ανησυχία σε κάποιον ευαίσθητο δέκτη.

```
> summary(epoch(sensor = sensor1,"winter")$pm25)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000  1.000   3.000   3.945  6.000   72.000
```

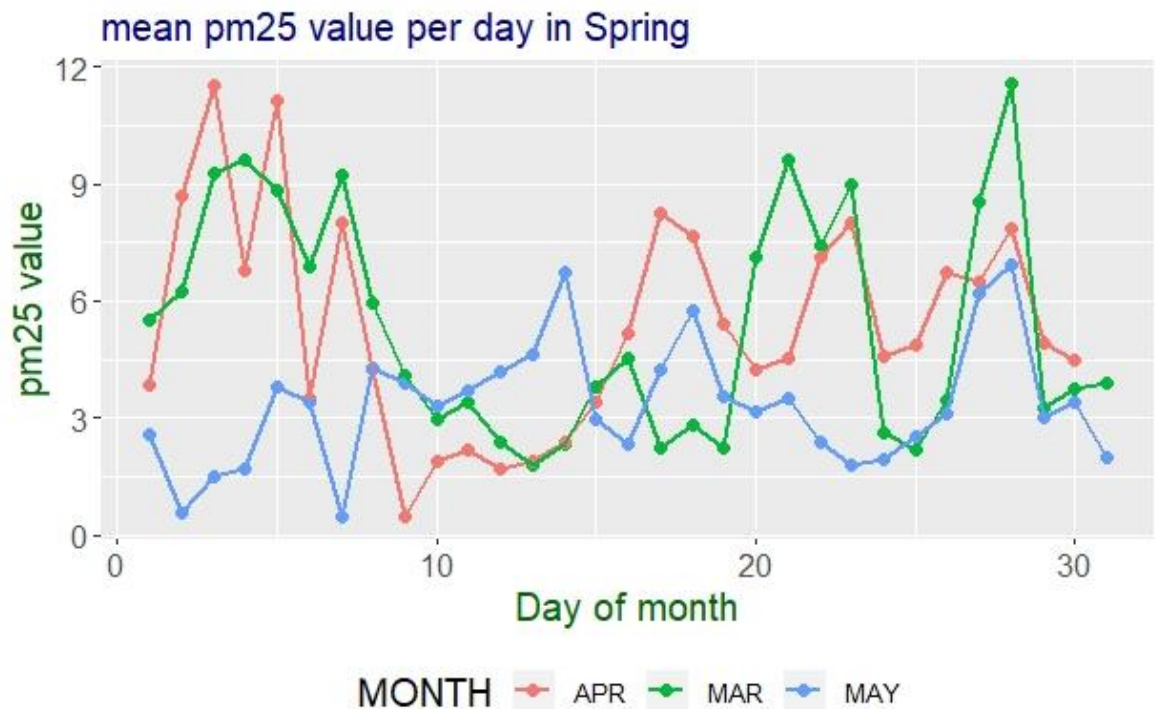
- Την άνοιξη σημειώθηκε αύξηση της μέσης τιμής και οι πιο συχνές τιμές έχουν



Γράφημα 2: Μέσες τιμές ανά ημέρα για τον αισθητήρα 8200015E την εποχή του χειμώνα

μετακινηθεί προς στα δεξιά, σύμφωνα με την **Εικόνα 4**, αυξάνοντας τις τιμές των σωματιδίων pm25 σε σύγκριση με την χειμερινή περίοδο. Ωστόσο οι τιμές δεν έφτασαν σε επικίνδυνα επίπεδα

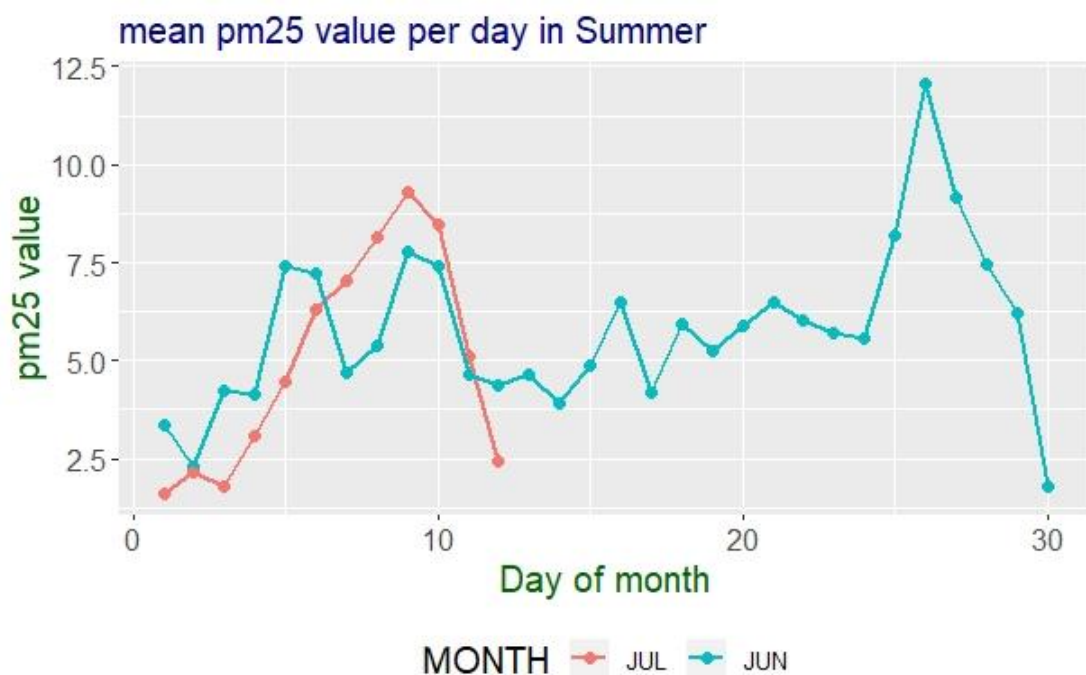
```
> summary(epoch(sensor = sensor1,"spring")$pm25)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000  2.000   4.000   4.471  6.000   28.000
```



Γράφημα 3: Μέσες τιμές ανά ημέρα για τον αισθητήρα 8200015E την εποχή του άνοιξη.

- Το καλοκαίρι, όπως και στην άνοιξη, φαίνεται εκ νέου αύξηση των ρύπων. Αξίζει να σχολιαστεί η κορυφή που υπάρχει στο **Γράφημα 4** συγκεκριμένα στις 26 Ιουλίου. Φαίνεται ότι μία επαναλαμβανόμενη δραστηριότητα υπάρχει κοντά στην περιοχή προκαλώντας αρκετή αύξηση των ρύπων.

```
> summary(epoch(sensor = sensor1,"summer")$pm25)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.000  4.000   5.000   5.586  7.000  22.000
```

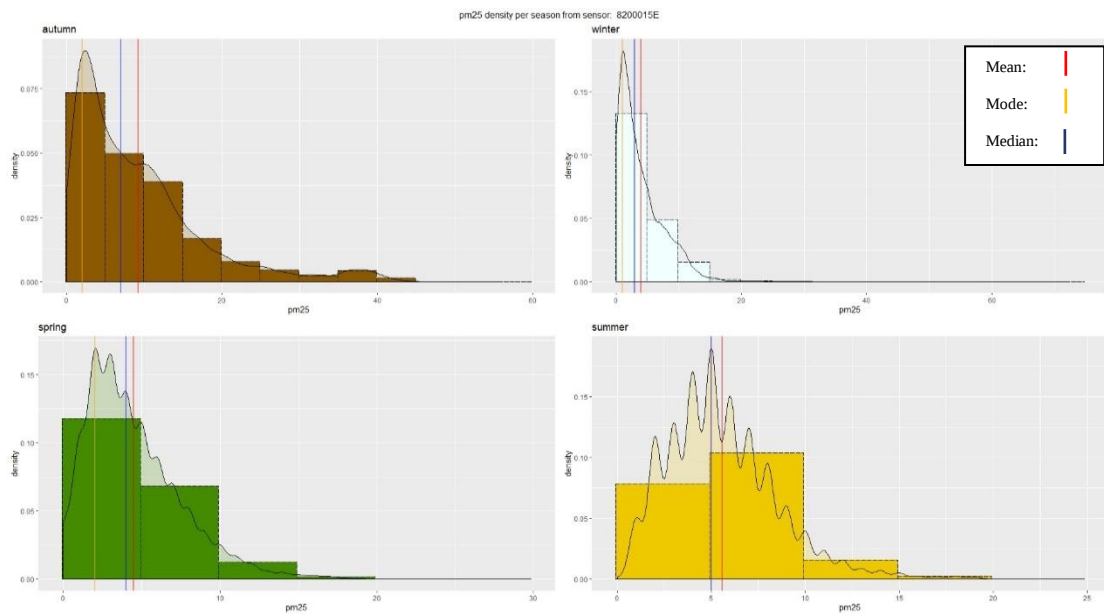


Γράφημα 4: Μέσες τιμές ανά ημέρα για τον αισθητήρα 8200015E την εποχή του καλοκαιριού.

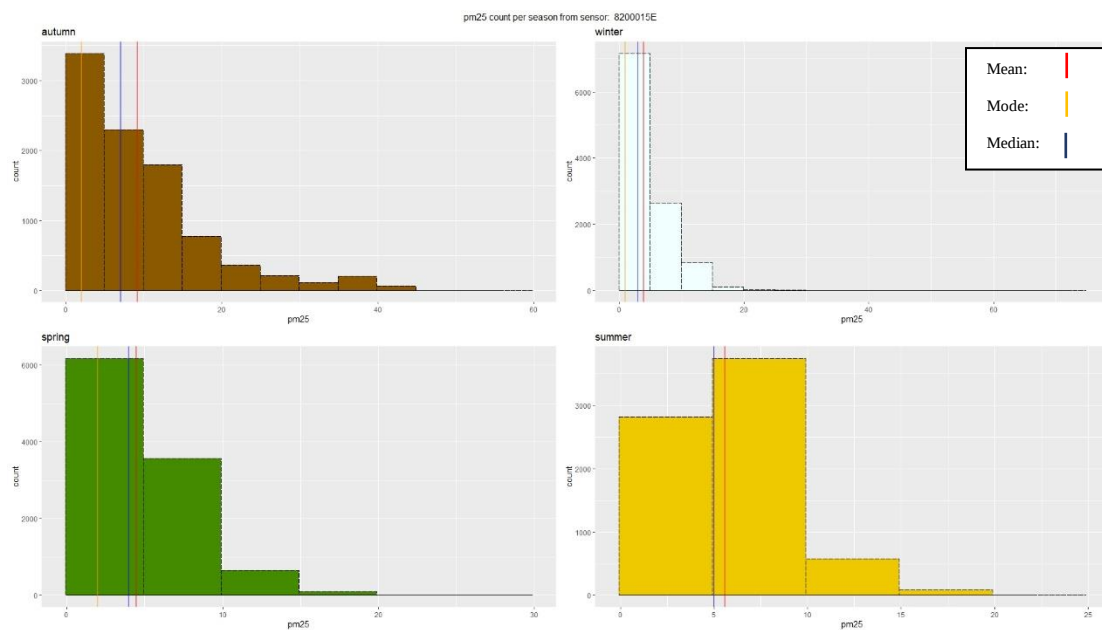
Συνολικά θα μπορούσε κανείς να συμπεράνει, εξαιρώντας την εποχή του φθινοπώρου, ότι οι τιμές αυξάνονται. Η πιο πιθανή αιτία για την αύξηση των τιμών μπορεί να είναι και η μεγαλύτερη ανθρώπινη δραστηριότητα και κινείται από την περιοχή Κάτω Μάνα προς την Ερμούπολη, όσο περισσότερο αυξάνεται η θερμοκρασία, με αποτέλεσμα την αύξηση των τιμών. Παράλληλα, υπάρχει και η τουριστική καλοκαιρινή περίοδος που επηρεάζει άμεσα τις τιμές των ρύπων καθώς η κίνηση σε χερσαίους και θαλάσσιους δρόμους αυξάνεται. Παρουσιάζει ιδιαίτερο ενδιαφέρον η μελέτη και ανάλυση με την ολοκλήρωση της τουριστικής περιόδου και να υπολογίσουμε και να ερμηνεύσουμε τη διαφορά που υπάρχει από εποχή σε εποχή.

Στα **Γράφημα 1, Γράφημα 2, Γράφημα 3 και Γράφημα 4** βλέπουμε τις μέσες τιμές των χαρακτηριστικών pm25 ανά ημέρα την εποχή που αναγράφεται στον τίτλο του πίνακα, για τον αισθητήρα 8200015E. Με αυτά τα γραφήματα γίνεται πιο ευδιάκριτο ποιοι μήνες έχουν τις μεγαλύτερες τιμές, κατά μέσο όρο, κατά τη διάρκειά τους. Το φθινόπωρο φαίνεται καθαρά ότι οι τιμές των pm25 είναι μεγαλύτερες κατά τον μήνα του Οκτωβρίου. Τον χειμώνα οι μέσες τιμές ανά ημέρα είναι ικανοποιητικά χαμηλές, κυρίως για τον μήνα του Ιανουαρίου. Τον Φεβρουάριο βλέπουμε τις μεγαλύτερες τιμές της μέσης τιμής όπου αυτό σημαίνει ταυτόχρονα ότι πιθανώς υπάρχουν και τιμές, επικίνδυνες για την υγεία. Την άνοιξη, οι τιμές παραμένουν σε ικανοποιητικά επίπεδα και δεν ξεπερνάνε τα όρια των υγιεινών τιμών. Το καλοκαίρι δεν μπορούμε να βγάλουμε και πολλά συμπεράσματα, λόγω της έλλειψης των δεδομένων.

Στις **Εικόνα 4 και Εικόνα 5** φαίνεται η κατανομή των παρατηρούμενων σωματιδίων ανά εποχή και παρατηρείται ότι οι βλαβερές τιμές είναι αρκετά σπάνιες.



Εικόνα 4: Ιστογράμματα δημιουργημένα ανά εποχή για την τα σωματίδια  $pm_{25}$  για τον αισθητήρα 8200015E. Στον κάθετο άξονα βλέπουμε την πιθανότητα της ύπαρξης των τιμών. Ικανοποιητικές τιμές για την μέτρηση των σωματιδίων είναι τα  $50\mu g/m^3$  κατά μέσο όρο σε 24 ώρες.



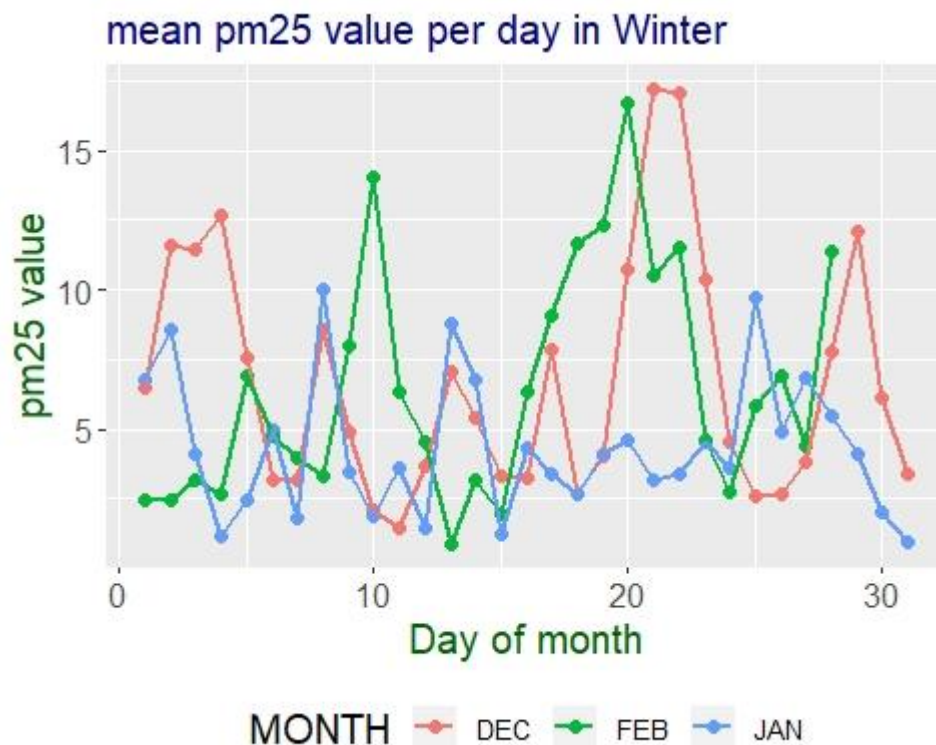
Εικόνα 5: Ιστόγραμμα που δείχνει το σύνολο των τιμών ανά εποχή για τον αισθητήρα 8200015E

### 3.2.2 Sensor 8200015F (Περιοχή: Λιμάνι)

Με την ίδια λογική, συνεχίζουμε την ανάλυση για τον αισθητήρα **8200015F**. Στις **Εικόνα 6** και **Εικόνα 7** φαίνονται τα ιστογράμματα που αφορούν την χρονιά που διανύουμε, ανά εποχή.

Συγκρίνοντας τις **Εικόνα 5** και **Εικόνα 7**, την οποία θα αναλύσουμε παρακάτω, γίνεται άμεσα αντιληπτή η διαφορά με τον προηγούμενο αισθητήρα. Ο λόγος είναι ότι ο αισθητήρας είναι τοποθετημένος στο κέντρο της Ερμούπολης. Βρίσκεται σε ένα σημείο με αυξημένη ανθρώπινη δραστηριότητα, κοντά στον κεντρικό δρόμο όπου υπάρχει μεγάλη κίνηση και παράλληλα κοντά στο κεντρικό λιμάνι.

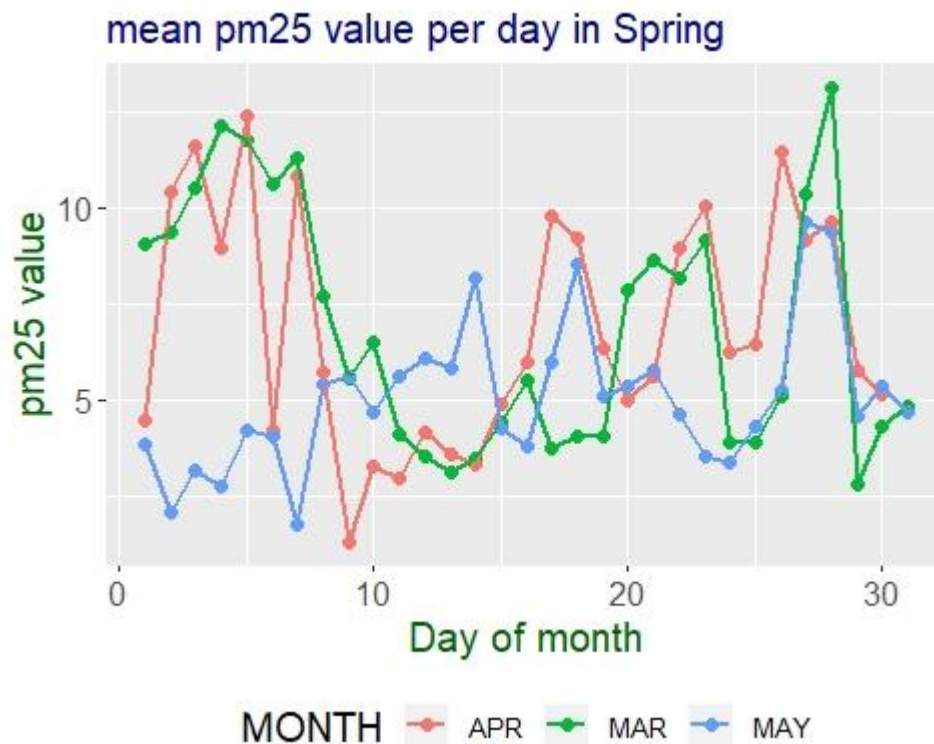
- Τον χειμώνα, ενώ κατά μέσο όρο οι τιμές είναι χαμηλές με την τάση των τιμών να κινούνται σε χαμηλά επίπεδα, παρατηρούνται όμως μέσες τιμές οι οποίες βρίσκονται κοντά στα επιτρεπτά όρια. Αυτό σημαίνει ότι τις συγκεκριμένες μέρες υπάρχουν σίγουρα τιμές που είναι ανθυγιεινές.



Γράφημα 5: Μέσες τιμές ανά ημέρα για τον αισθητήρα 8200015F την εποχή του χειμώνα.

```
> summary(epoch(sensor2,"winter")$pm25)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.000  2.000   4.000   5.983  8.000 107.000
```

- Την άνοιξη, οι μέσες τιμές ανά ημέρα όπως φαίνονται στο **Γράφημα 6** κυμαίνονται σε ικανοποιητικά επίπεδα. Παράλληλα, υπάρχουν μεμονωμένες εμφανίσεις ακραίων τιμών που όμως δεν έχουνε διάρκεια οι τιμές τους.

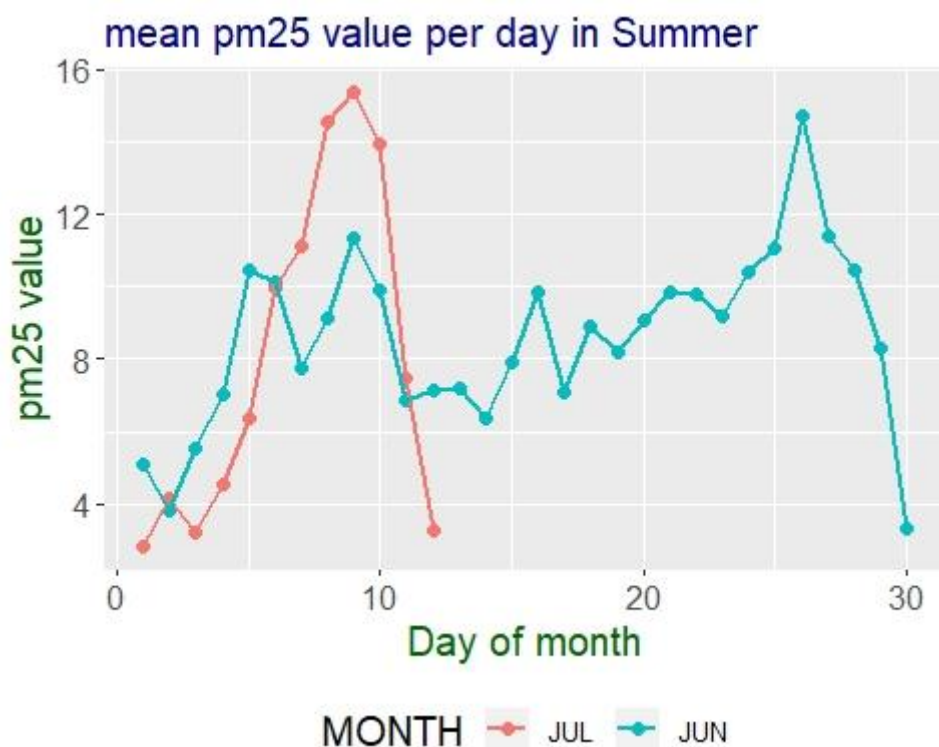


Γράφημα 6: Μέσες τιμές ανά ημέρα για τον αισθητήρα 8200015F την εποχή της άνοιξης.

```
> summary(epoch(sensor2,"spring")$pm25)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.000  3.000   5.000   6.096  8.000   60.000
```

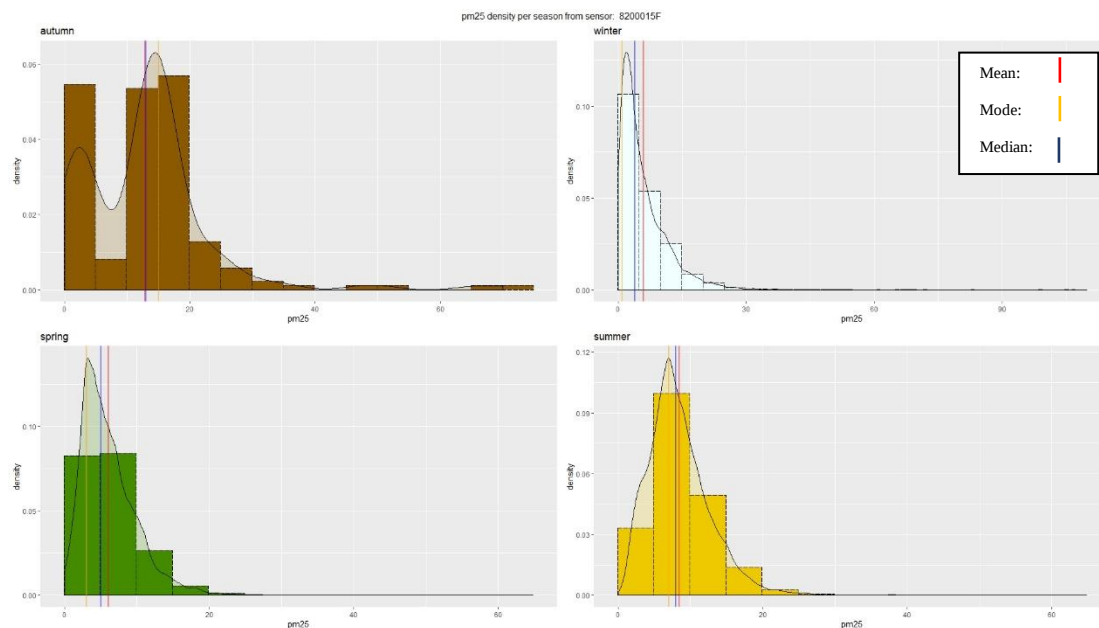
- Το καλοκαίρι, παρουσιάζονται πιο συχνές ακραίες τιμές pm25. Πιθανώς η μεγαλύτερη συχνότητα προσέλευσης πλοίων στο λιμάνι, η αύξηση του αριθμού των οχημάτων και η κίνηση αυτών στην Ερμούπολη, προκαλεί την αύξηση των τιμών αυτών.

```
> summary(epoch(sensor2,"summer")$pm25)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.000  6.000   8.000   8.473 11.000   62.000
```

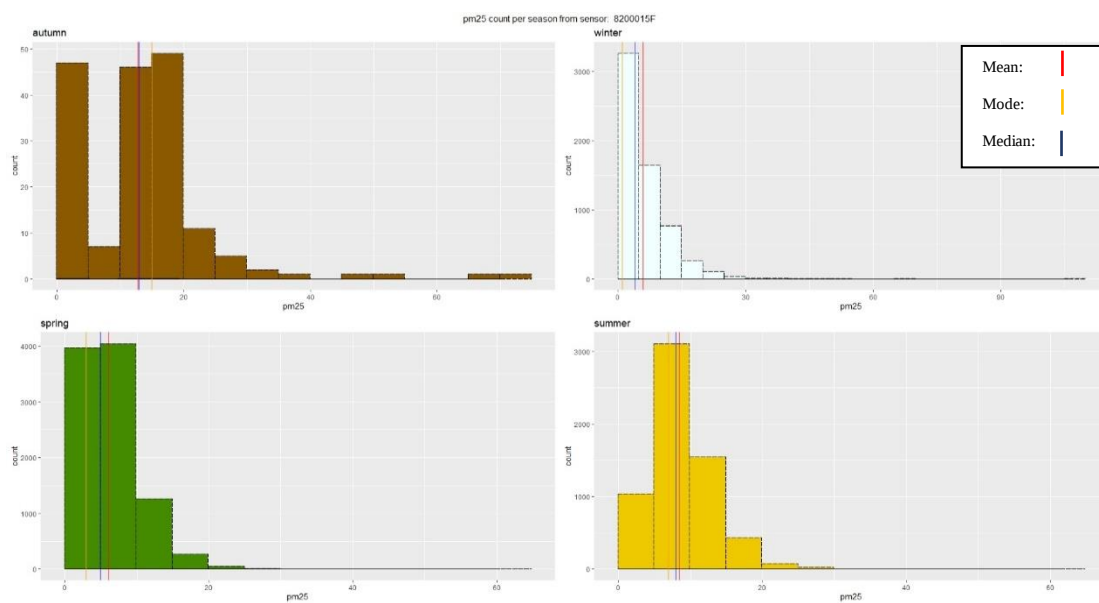


Γράφημα 7: Μέσες τιμές ανά ημέρα για τον αισθητήρα 8200015F την εποχή του καλοκαιριού.

Στα **Γράφημα 5**, **Γράφημα 6** και **Γράφημα 7** φαίνονται οι μέσες τιμές των ποσοτήτων pm25 ανά ημέρα, για κάθε μήνα, την εποχή που μας ενδιαφέρει. Τον χειμώνα φαίνονται κορυφές στο γράφημα για τους μήνες Δεκέμβριος-Φεβρουάριος όπου οι μέσες τιμές ξεπερνούν τα 15  $\mu\text{g}/\text{m}^3$  και υπάρχει μεγάλη πιθανότητα οι τιμές να έφτασαν άνω των 20  $\mu\text{g}/\text{m}^3$ . Την άνοιξη, γενικά οι μέσες τιμές ανά ημέρα υποχωρούνε βρισκόμενες σε αρκετά ικανοποιητικά. Το καλοκαίρι, αρχικά τα ποσά φαίνεται να συνεχίζουν τις ικανοποιητικές τους τιμές της άνοιξης, εμφανίζονται όμως κατά τόπους ακρότατα που ξεπερνούν τις μέγιστες μέσες τιμές της άνοιξης. Με την ολοκλήρωση του καλοκαιριού του έτους 2019 αναμένουμε να δούμε μεγαλύτερες τιμές λόγω της έλευσης της τουριστικής περιόδου και κατ' επέκταση, της αύξησης της ανθρώπινης δραστηριότητας.



Εικόνα 6: Ιστογράμματα δημιουργημένα ανά εποχή για την τα σωματίδια  $pm_{25}$  για τον αισθητήρα 8200015F. Στον κάθετο άξονα βλέπουμε την πιθανότητα της ύπαρξης των τιμών.



Εικόνα 7: Ιστογράμματα δημιουργημένα ανά εποχή για την τα σωματίδια  $pm_{25}$  με βάση τις μετρήσεις για τον αισθητήρα 8200015F.

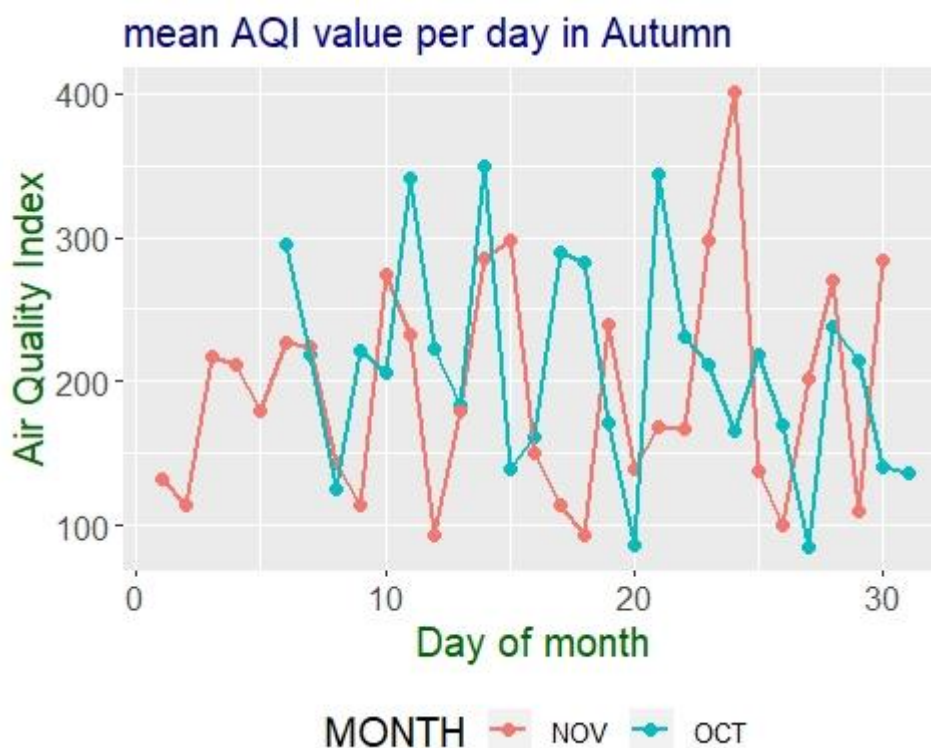
### 3.3 Ανάλυση AQI τιμής ανά εποχή

Στο κεφάλαιο 1 αναφέραμε στον **Πίνακα 1** την σημασία της μετρικής AQI ανάλογα με την τιμή. Υπενθυμίζουμε ωστόσο ότι οι τιμές από 0 έως 100 είναι οι αρκετά καλές τιμές και αποδεκτές-μέτριες τιμές για την ποιότητα του αέρα. Από την τιμή 101 και μετά υπάρχει ανάλογος βαθμός επικινδυνότητας.

Στην συνέχεια θα εξετάσουμε, στο πνεύμα της ανάλυσης των σωματιδίων pm25, τις τιμές AQI ανά εποχή για τους αισθητήρες 1 και 2.

#### 3.3.1 Sensor 820015E (Περιοχή: Κάτω Μάννα)

Ξεκινώντας με την εποχή του Φθινοπώρου και συμβουλευόμενοι τον **πίνακα 1** AQI Classification values της EPA [8] οι «ικανοποιητικές» τιμές από 0 έως 50 βρίσκονται σε πολύ χαμηλά επίπεδα. Συγκεκριμένα, η πιθανότητα να είναι «καλή» η ποιότητα του αέρα την εποχή του φθινοπώρου είναι πολύ μικρή. Στον **πίνακα 3** βλέπουμε ότι η πιθανότητα αυτών των τιμών είναι 6%. Στο 10% είναι η πιθανότητα η ποιότητα του αέρα να βρίσκεται



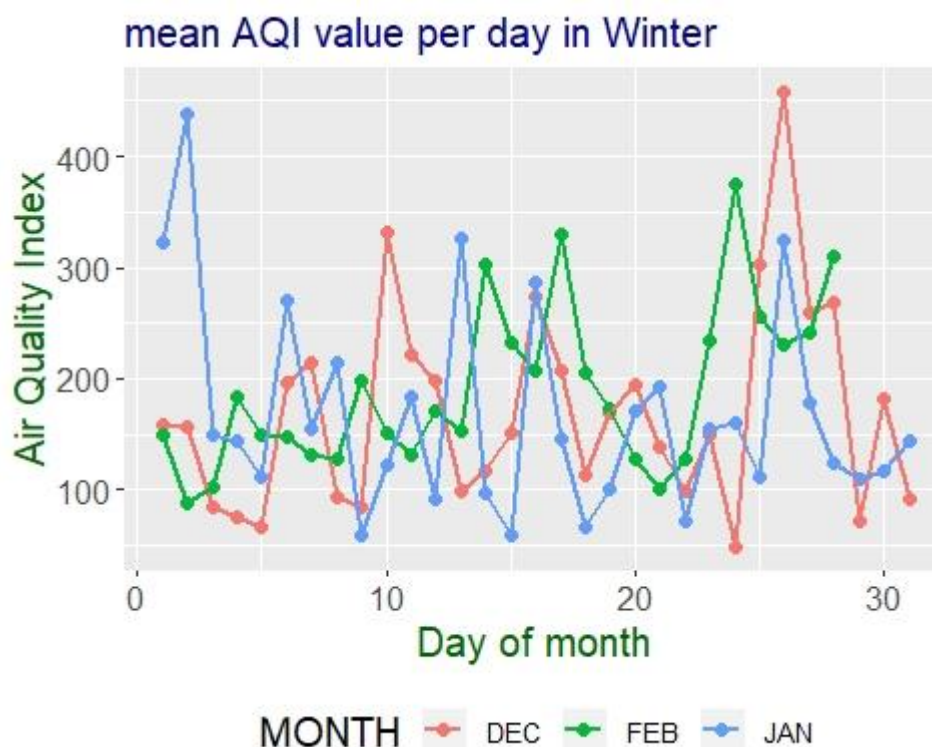
Γράφημα 8: Μέσες τιμές των εγγραφών AQI ανά ημέρα για τον αισθητήρα 8200015E την εποχή του φθινοπώρου.

σε μέτρια επίπεδα. 13% είναι η πιθανότητα η ποιότητα να είναι ανθυγιεινή για ευαίσθητες ομάδες ανθρώπων. Ανησυχία προκαλεί το γεγονός ότι υπάρχει μεγάλη πιθανότητα της τάξεως του 62% η ποιότητα του αέρα να είναι ιδιαίτερα ανθυγιεινή (14% η ποιότητα να

χαρακτηρίζεται ανθυγιεινή, 27% πολύ ανθυγιεινή, 21% επιβλαβής) Στο **Γράφημα 8** φαίνεται ότι η μέση τιμή για κάθε ημέρα τείνει να είναι πάνω από 200 που θεωρείται «πολύ ανθυγιεινή» για πλήθος ημερών. Το γράφημα αυτό αποδεικνύει και τις πιθανότητες που φαίνονται στον **Πίνακα 3**.

```
> summary(epoch(sensor = sensor1, "autumn")$vocaqi)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.00  90.25  168.00  198.66  286.00  500.00
```

- Τον **Χειμώνα** ωστόσο, η τάση των τιμών AQI δείχνει να δέχεται μείωση και αυτό φαίνεται τόσο από τις μέσες τιμές που βλέπουμε στο **Γράφημα 9** όσο και στον **Πίνακα 3**. Μόλις 19% είναι η πιθανότητα οι τιμές να είναι αποδεκτές (για καλή ή μέτρια ποιότητα). Το υπόλοιπο ποσοστό εξακολουθεί να βρίσκεται σε μεγάλα επίπεδα καθώς το 40% των τιμών AQI χαρακτηρίζει το επικίνδυνο φάσμα της ποιότητας του αέρα. Καθημερινά,

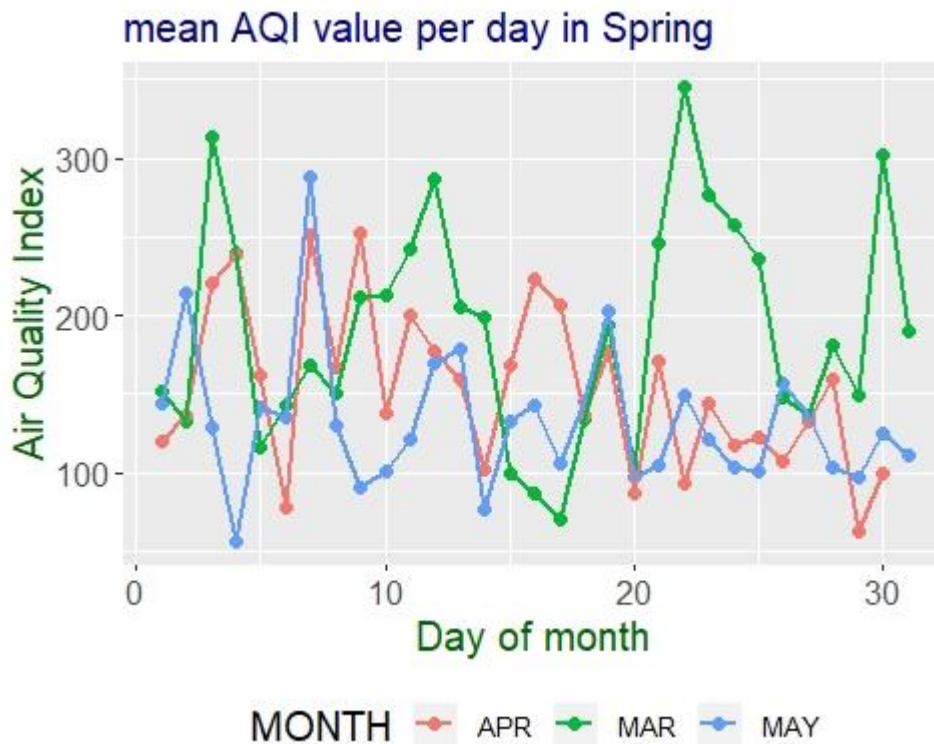


Γράφημα 9: Μέσες τιμές των εγγραφών AQI ανά ημέρα για τον αισθητήρα 8200015E την εποχή του χειμώνα.

βλέπουμε ότι οι τιμές για τον χειμώνα είναι πιο πυκνές στο διάστημα από 100 μέχρι 200, ωστόσο το γεγονός ότι υπάρχουν μέρες που έχουν μέση τιμή πάνω από 300 θα πρέπει να μελετηθεί περεταίρω.

```
> summary(epoch(sensor = sensor1, "winter")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  62.0  142.0 172.7  250.0  500.0
```

- Την **Άνοιξη** καθώς τα ποσοστά επικίνδυνων ρύπων φαίνεται να υποχωρούν. Οι «καλές» και «μέτριες» τιμές της ποιότητας του αέρα είναι 10% και 14% αντίστοιχα. Σε σύγκριση με τον χειμώνα, οι πολύ επικίνδυνες τιμές υποχωρούν κατά 8% την άνοιξη, χωρίς αυτό να μειώνει το μέγεθος του ποσοστού του 32% για πολύ ανθυγιεινά και επικίνδυνα επίπεδα ποιότητας. Στο **Γράφημα 10**, φαίνεται η αραίωση στο διάστημα 100-200 αλλά επίσης φαίνεται και μείωση των ακραίων τιμών που υπάρχουν πλέον κάτω από την τιμή 350.

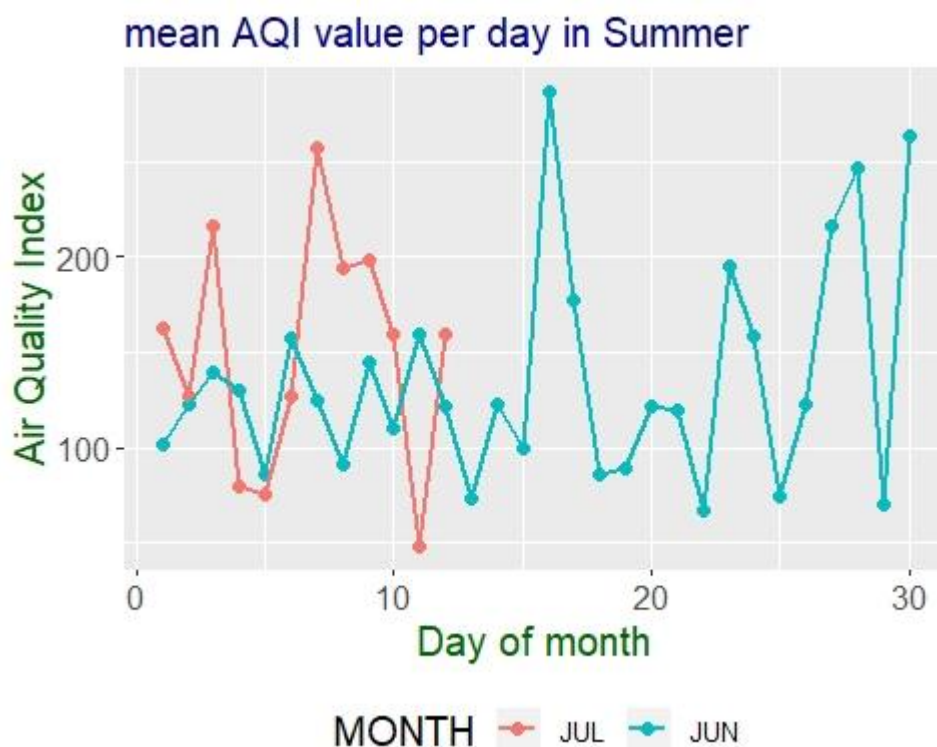


Γράφημα 10: Μέσες τιμές των εγγραφών AQI ανά ημέρα για τον αισθητήρα 8200015E την εποχή του φθινοπώρου.

```
> summary(epoch(sensor = sensor1, "spring")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  50.0  109.0 144.6  207.0  500.0
```

- Το **Καλοκαίρι** παρατηρούμε εξίσου μειωμένα τα επιβλαβή ποσοστά. Τα «καλά» και «μέτρια» ποσοστά έχουν άθροισμα 26%. Τα επικίνδυνα, από την άλλη μεριά, βρίσκονται στο 29%. Αξίζει να αναφερθεί σε αυτό το σημείο ότι τα δεδομένα που έχουμε συλλέξει για το καλοκαίρι του 2019 όλα τα δεδομένα. Οπότε δεν μπορούμε να έχουμε καθαρή, γενική εικόνα για την ποιότητα του αέρα του καλοκαιριού.

Το καλοκαίρι φαίνεται να συνεχίζει τις αρχικά, ενθαρρυντικές του Μαΐου, από τα μέσα του Ιουνίου ωστόσο οι τιμές απομακρύνονται σε επίπεδα πιο επικίνδυνα που ανά μέρα διαφέρουν αρκετά πολύ.



Γράφημα 11: Ιστογράμματα δημιουργημένα ανά εποχή για την τα σωματίδια AQI για τον αισθητήρα 8200015E. Στον κάθετο άξονα βλέπουμε τη συχνότητα των τιμών.

```
> summary(epoch(sensor = sensor1, "summer")$vocaqi)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.0   56.0   107.0   139.7   191.8   500.0
```

Παρατηρώντας τις μέσες τιμές ανά εποχή για τον αισθητήρα 8200015E βλέπουμε μια υποχώρηση των τιμών AQI από εποχή σε εποχή. Ωστόσο, ενώ η τάση των τιμών δείχνει να υποχωρεί στις πιο χαμηλές και ταυτόχρονα ικανοποιητικές τιμές για το AQI, οι τιμές στο

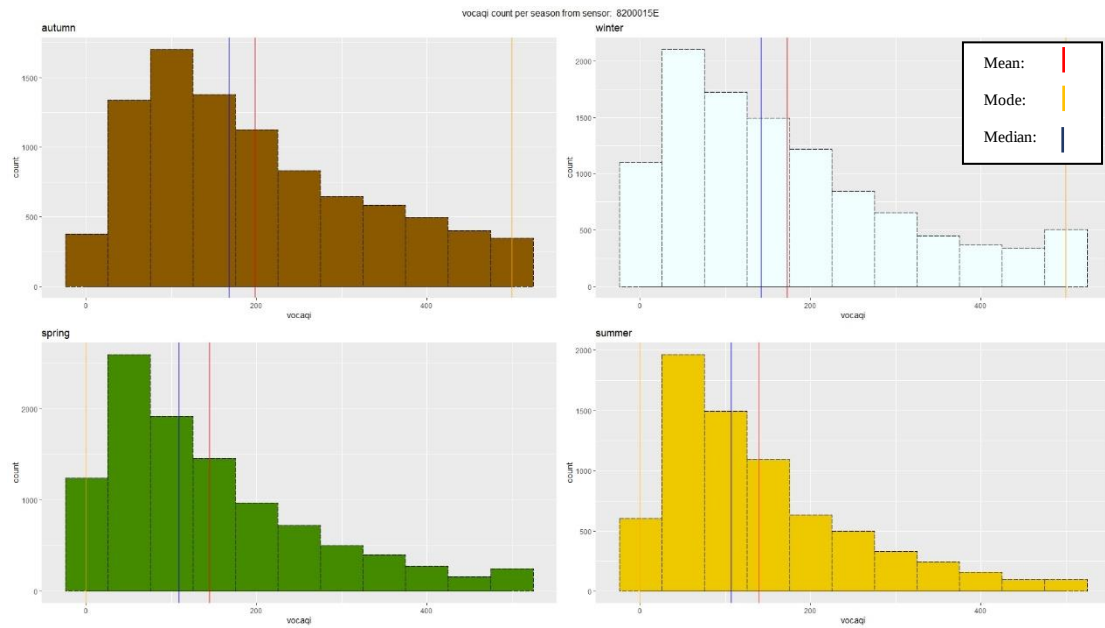
πρώτο τεταρτημόριο της κατανομής πέφτουμε από εποχή σε εποχή και έτσι η συχνότητα εμφάνισης των πιο κακόβουλων τιμών αυξάνεται πηγαίνοντας πιο κοντά στις τιμές του μέσου όρου.

Από την άλλη πλευρά της κατανομής, παρόλο που η συχνότητα είναι μικρότερη, η πιθανότητα εμφάνισης αυτής των «επικίνδυνων» τιμών είναι αρκετά υψηλή. Σε αυτό συμβάλλουν σε φθινόπωρο και χειμώνα οι mode τιμές (οι τιμές με την μεγαλύτερη συχνότητα ο οποίες ενδέχεται να αποτελούνε λάθος παρατηρήσεις). Σε άνοιξη και καλοκαίρι οι mode τιμές είναι οι πιο χαμηλές (ενδέχεται εξίσου να έχουν γίνει λάθος παρατηρήσεις). Έτσι οι πιθανότητες των βλαβερών τιμών, της δεξιάς ουράς της κατανομής υποχωρούνε.

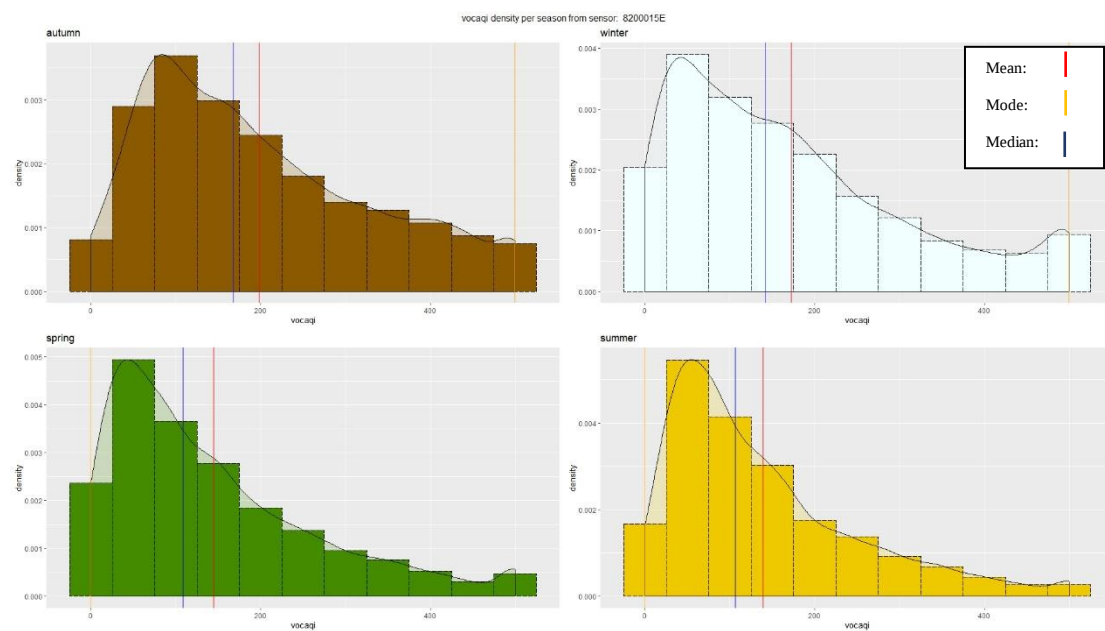
Στην παραπάνω διαπίστωση, πιθανώς παίζουν σημαντικό ρόλο και λανθασμένες ή ακραίες τιμές (outliers) που ο αισθητήρας έχει καταγράψει λάθος.

| AQI total sensor: 8200015E |               |               |               |               |
|----------------------------|---------------|---------------|---------------|---------------|
|                            | <i>autumn</i> | <i>winter</i> | <i>spring</i> | <i>summer</i> |
| <b>0-50</b>                | 0.06          | 0.08          | 0.1           | 0.11          |
| <b>51-100</b>              | 0.1           | 0.11          | 0.14          | 0.15          |
| <b>101-150</b>             | 0.13          | 0.13          | 0.16          | 0.17          |
| <b>151-200</b>             | 0.14          | 0.14          | 0.15          | 0.17          |
| <b>201-300</b>             | 0.27          | 0.24          | 0.22          | 0.22          |
| <b>301-500</b>             | 0.21          | 0.16          | 0.1           | 0.07          |

Πίνακας 3: Πιθανότητα ύπαρξης εύρους AQI τιμής ανά εποχή για τον αισθητήρα 8200015E



Εικόνα 8: Ιστογράμματα δημιουργημένα ανά εποχή για την τα σωματίδια AQI για τον αισθητήρα 8200015E. Στον κάθετο άξονα βλέπουμε τη συχνότητα των τιμών.

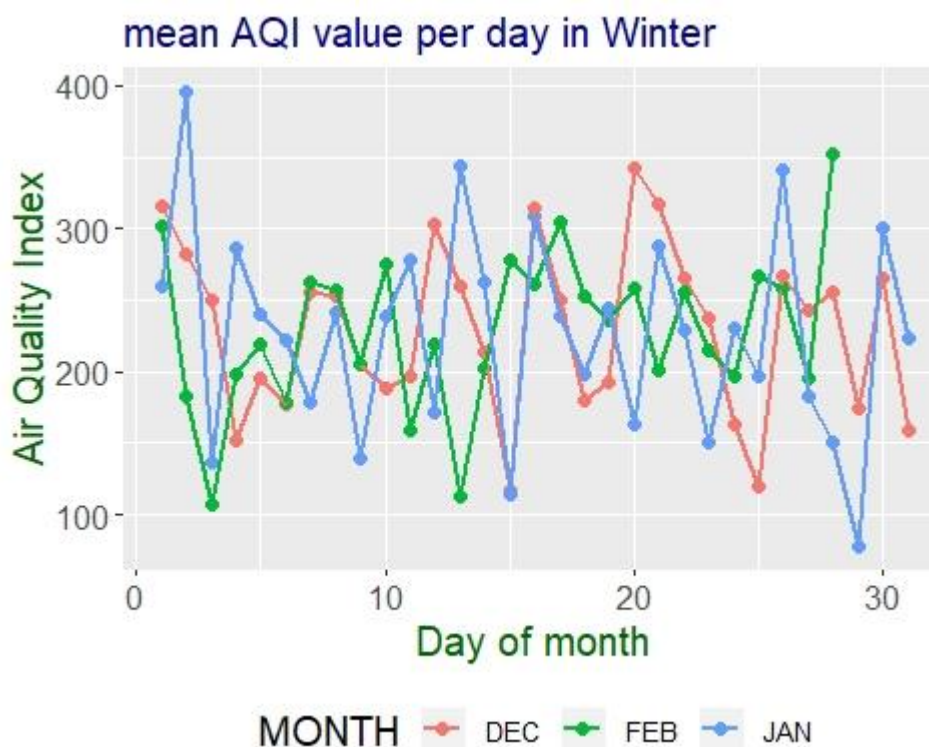


Εικόνα 9: Ιστογράμματα δημιουργημένα ανά εποχή για την τα σωματίδια AQI για τον αισθητήρα 8200015E. Στον κάθετο άξονα βλέπουμε την πιθανότητα της ύπαρξης των τιμών.

### 3.3.2 Sensor 8200015F (Περιοχή: Λιμάνι)

Αν ένας παρατηρητής, δει τα ιστογράμματα του αισθητήρα 8200015F στην **Εικόνα 10**, θα διαπιστώσει ότι πρόκειται ότι τα δεδομένα εδώ είναι πιο ομοιόμορφα κατανεμημένα σε σχέση με αυτά του αισθητήρα 8200015E με μία εξαίρεση αυτή του φθινοπώρου όπου δεν είναι ολοκληρωμένη η συλλογή δεδομένων. Έτσι λοιπόν, για να μην βγάλουμε βιαστικά συμπεράσματα η ανάλυση για τον αισθητήρα του Λιμανιού ξεκινάει από τον χειμώνα.

- Τον **χειμώνα** ο μέσος όρος των καθημερινών τιμών βρίσκει πολύ ανθυγιεινές τιμές πάνω από τιμή 200 όπως φαίνεται στο **Γράφημα 12**. Κοιτώντας τον πίνακα 4 επιβεβαιώνουμε ότι υπάρχει μεγαλύτερη πιθανότητα τον χειμώνα η ποιότητα του αέρα να είναι ανθυγιεινή καθώς για καλές και μέτριες τιμές υπάρχει πιθανότητα μόλις 12%.

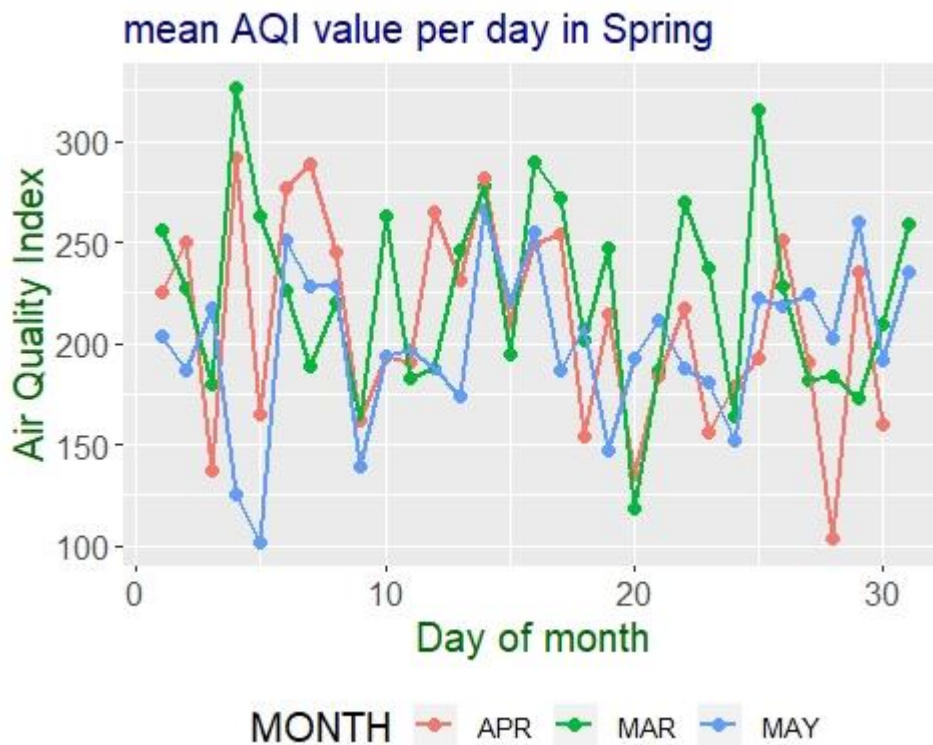


Γράφημα 12: Μέσες τιμές των εγγραφών AQI ανά ημέρα για τον αισθητήρα 8200015F την εποχή του χειμώνα.

```
> summary(epoch(sensor = sensor2, "winter")$vocaqi)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.0   130.0   215.0   227.7   318.0   500.0
```

- Συνεχίζοντας με την **άνοιξη**, βλέπουμε ότι ενώ ο μέσος όρος των τιμών υποχωρεί σε σχέση με τον χειμώνα διανέμοντας τις καθημερινές τιμές του σε μεγαλύτερο εύρος,

παραμένει σε πολύ ανθυγιεινά επίπεδα. Παρατηρώντας τον πίνακα 4, βλέπουμε μία μικρή μείωση στην πιθανότητας των βλαβερών επιπέδων. Αυτή η μείωση ωστόσο, μετακινήθηκε στις πιο κεντρικές τιμές κάνοντας την διαφορά (mean-median) ακόμα πιο μικρή. Στα θετικά αποτελέσματα βλέπουμε μια πολύ μικρή αύξηση της πιθανότητας για να έχουμε πιο υγιεινές τιμές.

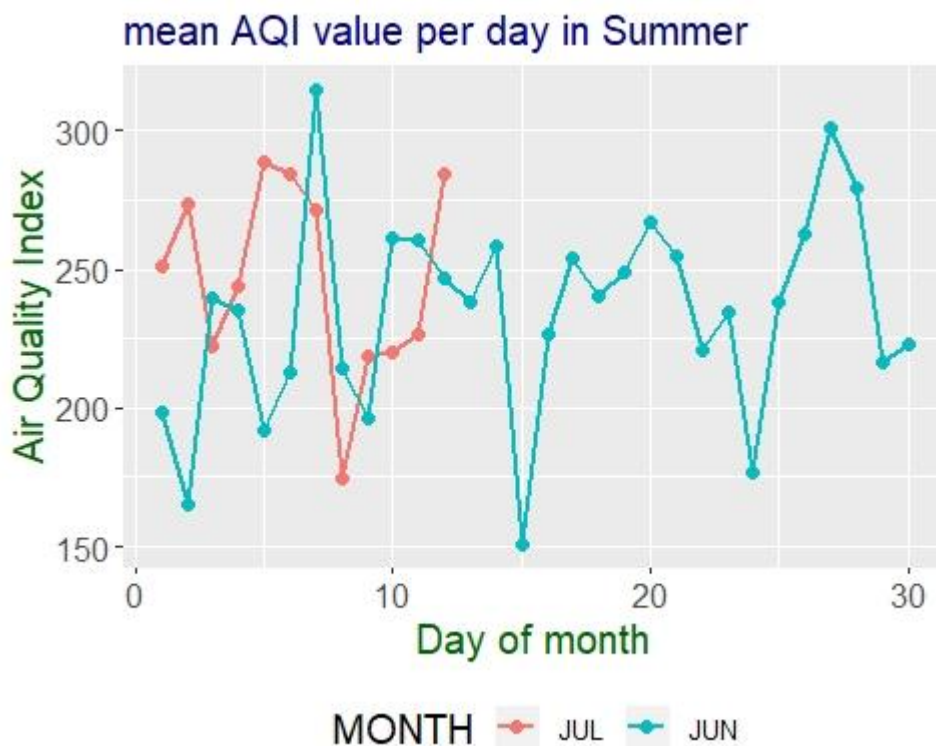


Γράφημα 13: Μέσες τιμές των εγγραφών AQI ανά ημέρα για τον αισθητήρα 8200015F την εποχή της άνοιξης.

```
> summary(epoch(sensor = sensor2, "spring")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  113.0  194.0 202.1  280.0  500.0
```

- Με την έλευση του **καλοκαιριού**, παρατηρούμε ότι οι τιμές μετακινήθηκαν πιο δεξιά κάνοντας τα δεδομένα μας να έχουν μία πιο κανονικά κατανεμημένη μορφή. Υπάρχει αύξηση κατά 15% στην μέση τιμή των καταγραφών AQI ενισχύοντας τις πιο βλαβερές τιμές. Ανησυχία προκαλεί ότι το καλοκαίρι η πιθανότητα η ποιότητα του αέρα να είναι στα πολύ ανθυγιεινά έως και βλαβερά επίπεδα είναι 67%.

Φαίνεται λοιπόν ότι παρόλο που η καλοκαιρινή - τουριστική περίοδος πιθανώς οφελεί οικονομικά το νησί, η αύξηση του πληθυσμού και της ανθρώπινης δραστηριότητας επιδρά αρνητικά ως προς την ποιότητα του αέρα, το περιβάλλον και την υγεία της κοινωνίας.



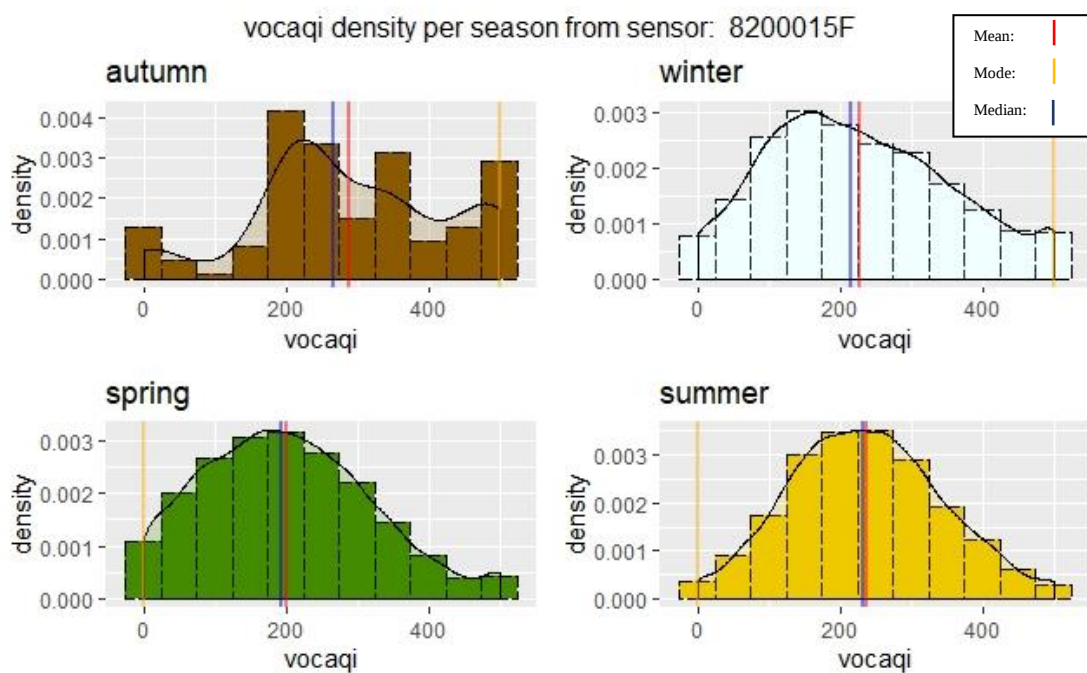
Γράφημα 14: Μέσες τιμές των εγγραφών AQI ανά ημέρα για τον αισθητήρα 8200015F την εποχή του καλοκαιριού.

```
> summary(epoch(sensor = sensor2, "summer")$vocaqi)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.0   161.0   232.0   236.6   309.0   500.0
```

## AQI total sensor: 8200015F

|                | autumn | winter | spring | summer |
|----------------|--------|--------|--------|--------|
| <b>0-50</b>    | 0.02   | 0.04   | 0.05   | 0.03   |
| <b>51-100</b>  | 0.01   | 0.08   | 0.09   | 0.06   |
| <b>101-150</b> | 0.06   | 0.11   | 0.13   | 0.11   |
| <b>151-200</b> | 0.07   | 0.14   | 0.16   | 0.16   |
| <b>201-300</b> | 0.27   | 0.3    | 0.3    | 0.36   |
| <b>301-500</b> | 0.4    | 0.27   | 0.19   | 0.27   |

Πίνακας 4:Πιθανότητα ύπαρξης εύρους AQI τιμής ανά εποχή για τον αισθητήρα 8200015F



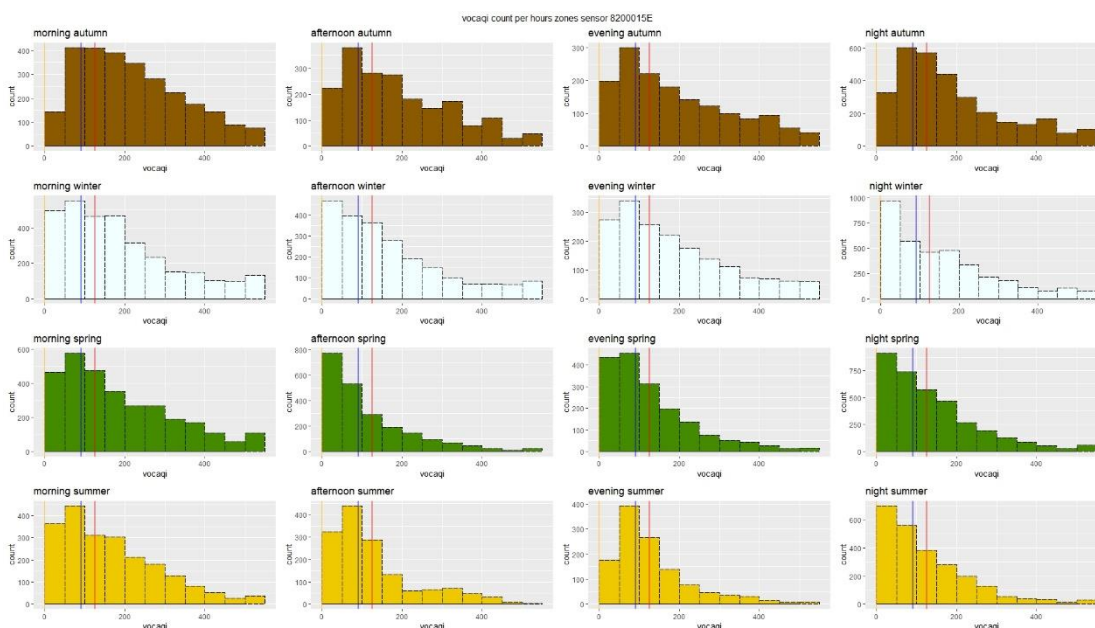
Εικόνα 10: Ιστογράμματα δημιουργημένα ανά εποχή για την τα σωματίδια AQI για τον αισθητήρα 8200015F. Στον κάθετο άξονα βλέπουμε την πιθανότητα της ύπαρξης των τιμών.

## ΚΕΦΑΛΑΙΟ 4

### 4.1 Εισαγωγή

Στο κεφάλαιο 3 αναλύσαμε την «συμπεριφορά» των σωματιδίων pm25 αλλά και της βασικής μέτρησης ποιότητας του αέρα – AQI ανά εποχή βλέποντας και την πιθανότητα από τους πίνακες, πότε η ποιότητα του αέρα είναι καλή, μέτρια, βλαβερή κοκ. Στο παρόν κεφάλαιο επεκτείνεται η ανάλυση σε ένα παραπάνω επίπεδο. Πιο συγκεκριμένα ανά ζώνη ώρας (πρωί, μεσημέρι, απόγευμα και βράδυ) την ποιότητα του αέρα σε κάθε αισθητήρα και να αναγνωριστούν τυχόν πρότυπα που πιθανόν αναπαράγονται ανά εποχή.

Στην **Εικόνα 13** φαίνεται ένας πίνακας ιστογραμμάτων όπου κάθε γραμμή αναφέρεται στην εποχή και κάθε στήλη αναφέρεται στην ζώνη ώρας. Στην παρακάτω ανάλυση γίνεται σύγκριση των τιμών AQI ανά εποχή σε συγκεκριμένη ζώνη ώρας. Συγκεκριμένα, γίνεται σύγκριση κάθε στήλης της Εικόνας 13. Για παράδειγμα, για την ζώνη ώρας «πρωί» συγκρίνονται οι τιμές και οι διακυμάνσεις σε κάθε εποχή. Με τον ίδιο τρόπο συνεχίζεται η ανάλυση για κάθε εποχή.



Εικόνα 13: Πίνακας Ιστογραμμάτων για τις μετρήσεις των AQI. Οριζόντια ανά εποχή, κάθετα ανά ζώνη ώρας.

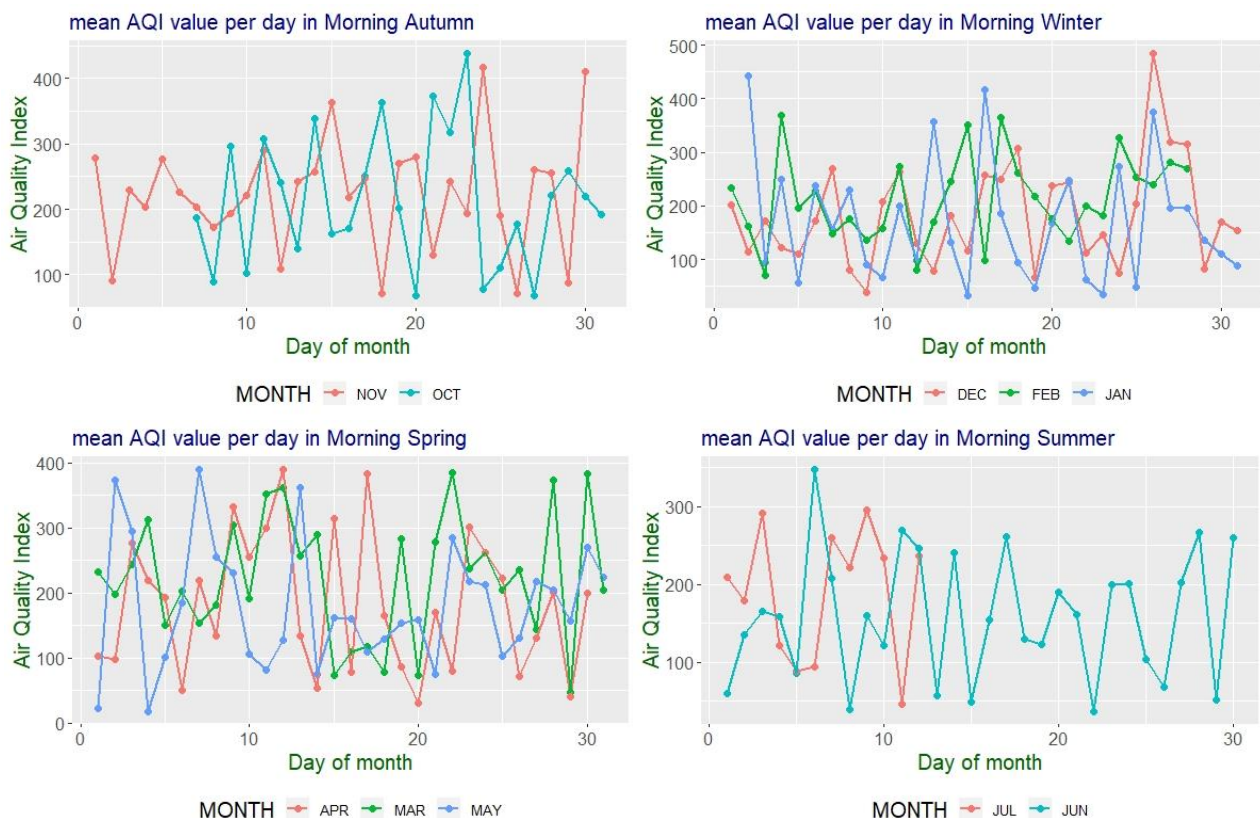
## 4.2 Ανάλυση AQI ανά εποχή και ζώνη ώρας

Σε παράλληλη έρευνα που πραγματοποιήθηκε με την παρούσα εργασία για τα δεδομένα της Σύρου, ο Constantine Hatziaodoniu μελετάει και δείχνει πόσο ο ανθρώπινος παράγοντας επηρεάζει τους δείκτες ποιότητας αέρα [24]. Το παρόν κεφάλαιο αναλύονται τα δεδομένα με βάση την ζώνη ώρας σε κάθε εποχή με στόχο να φανούν ομοιότητες με την μέση ανθρώπινη δραστηριότητα.

### 4.2.1 sensor: 8200015E (Περιοχή: Κάτω Μάννα)

- Πρώι ανά εποχή:

Ξεκινώντας την ανάλυση, παρατηρούμε τις μέσες τιμές των AQI για τις πρωινές ώρες που αναφέρονται από τις 5:00 μέχρι τις 11:59. Βλέπουμε ότι τις μεγαλύτερες τιμές τις βλέπουμε το φθινόπωρο αλλά όπως και στο **κεφάλαιο 3** δεν μπορούμε να μιλήσουμε με σιγουριά για την ποιότητα αυτού του συνόλου δεδομένων λόγω του μικρού όγκου. Ωστόσο, στο **Γράφημα 15** βλέπουμε την κατανομή των μέσων τιμών ανά ημέρα για το φθινόπωρο, τις πρωινές ώρες και παρατηρούμε ότι οι περισσότερες τιμές βρίσκονται στο



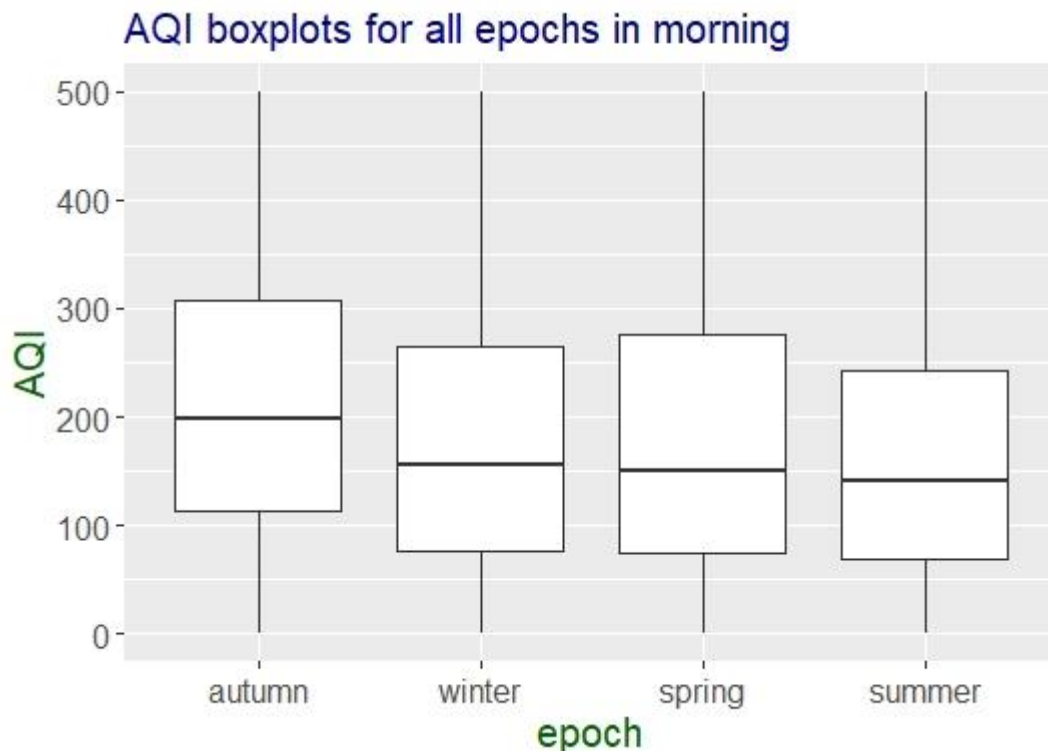
Γράφημα 15: Γραφήματα μέσων τιμών AQI τις πρωινές ώρες ανά εποχή.

εύρος των ανθυγιεινών τιμών. Μάλιστα το 26,3% των καταγεγραμμένων τιμών βρίσκονται στο εύρος των «βλαβερών» τιμών σύμφωνα με τον **Πίνακα 1** της ΕΡΑ.

Στη συνέχεια βλέπουμε τις τιμές που έχουμε λάβει το χειμώνα. Η μέση τιμή βρίσκεται στις «ανθυγιεινές» τιμές. Το 66,1% των τιμών ανήκουν σε αυτό το εύρος των ανθυγιεινών τιμών και το 20% των τιμών AQI βρίσκεται σε επικίνδυνα επίπεδα. Σε σύγκριση με το φθινόπωρο έχουμε υποχώρηση των βλαβερών τιμών αλλά σε αυτό το ποσοστό των βλαβερών τιμών παραμένει μεγάλο. Στο **Γράφημα 15**, φαίνεται ότι αρκετές μέρες του χειμώνα η ποιότητα του αέρα χαρακτηρίζεται «καλή» και «μέτρια»

Ομοίως, την άνοιξη παρατηρούμε ότι η μέση τιμή κυμαίνεται τις πρωινές ώρες στα ίδια επίπεδα με τον χειμώνα. Ο αναγνώστης συμβουλευμένος το **Γράφημα 16** θα μπορούσε να παρατηρήσει ότι τόσο τον χειμώνα όσο και την άνοιξη η δραστηριότητα ήταν κοινή. Παρατηρώντας όμως το **Γράφημα 15** φαίνεται ότι οι μέσες τιμές της άνοιξης σε καθημερινή βάση, σε σύγκριση με αυτές του χειμώνα, συνηθίζουν να ξεπερνάνε το εύρος των «καλών» και «μέτριων» τιμών AQI.

Τέλος, το καλοκαίρι παρατηρούμε πτώση της μέσης τιμής AQI. Στα **Γράφημα 15** και **Γράφημα 16** φαίνεται αυτή η μείωση τόσο σε καθημερινή βάση όσο και γενικά το καλοκαίρι. Λόγω έλλειψης καταγεγραμμένων δεδομένων το καλοκαίρι δεν μπορούμε να



Γράφημα 16: Boxplots φαίνεται η κατανομή των τιμών AQI τις πρωινές ώρες ανά εποχή

κάνουμε απόλυτες συγκρίσεις με τους προηγούμενους μήνες. Μία πρώτη παρατήρηση που θα μπορούσε να γίνει είναι ότι με την αύξηση της θερμοκρασίας μειώθηκε η ανθρώπινη δραστηριότητα στην περιοχή Κάτω Μάννα ή μετακινήθηκε σε άλλες περιοχές του νησιού.

```
> summary(timezones(epoch(sensor1,epc = "autumn"),"morning")$vocaqi)
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0   113.5   198.0   218.3   308.0   500.0
> summary(timezones(epoch(sensor1,epc = "winter"),"morning")$vocaqi)
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0    76.0   155.0   185.7   265.2   500.0
> summary(timezones(epoch(sensor1,epc = "spring"),"morning")$vocaqi)
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0    74.0   150.0   184.9   275.8   500.0
> summary(timezones(epoch(sensor1,epc = "summer"),"morning")$vocaqi)
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.00   68.75   142.00   166.76   242.00   500.00
```

- **Μεσημέρι ανά εποχή:**

Στο ίδιο πνεύμα συνεχίζεται η ανάλυση και για τις μεσημεριανές ώρες 12:00 – 16:59. Με μία γρήγορη ματιά στα **Γράφημα 17** και **Γράφημα 18** φαίνεται η συμπεριφορά και η κατανομή των τιμών ανά εποχή. Ο αναγνώστης μπορεί, παρατηρώντας τα γραφήματα, να διαπιστώσει ότι για τις μεσημεριανές ώρες ισχύει η ανάλυση που έγινε με τις πρωινές ώρες παρατηρώντας ωστόσο τη μείωση που υπάρχει στις τιμές.

Ενδεικτικά θα αναφέρουμε ότι για το φθινόπωρο το 42% των τιμών βρίσκεται πάνω από το μέσο όρο ενώ το 20% βρίσκεται στο εύρος των βλαβερών επιπέδων. Τον χειμώνα το 40% των τιμών βρίσκεται πάνω από το μέσο όρο ενώ το 17% βρίσκεται στο εύρος των βλαβερών επιπέδων. Όμοια, την άνοιξη, το 36% των τιμών βρίσκεται πάνω από το μέσο όρο και μόλις το 7% βρίσκεται στο εύρος των βλαβερών επιπέδων. Τέλος, το καλοκαίρι, το 33,4% των τιμών βρίσκεται πάνω από το μέσο όρο ενώ το 11,1% βρίσκεται στο εύρος των βλαβερών επιπέδων. Είναι εμφανές ότι υπάρχει μία σταθερή μείωση και από τα ποσοστά που αντιστοιχούν στις ανθυγιεινές τιμές ενώ αξίζει να σημειωθεί ότι την άνοιξη και το καλοκαίρι τα ποσοστά των υγιεινών τιμών είναι 59,91% και 52,07 αντίστοιχα.

Φαίνεται λοιπόν ότι η ανθρώπινη δραστηριότητα πέφτει το μεσημέρι σε σύγκριση με τις πρωινές ώρες σε κάθε μήνα και αυτό μπορεί να φανεί από τα γραφήματα και την παραπάνω επιχειρηματολογία.

```

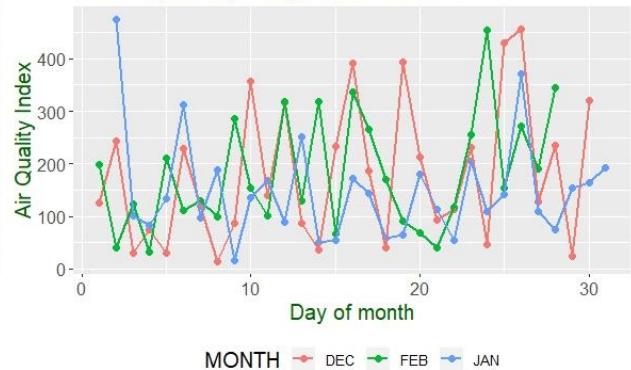
> summary(timezones(epoch(sensor1,"autumn"), "afternoon")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  82.5  163.0  190.3  281.5  500.0
> summary(timezones(epoch(sensor1,"winter"), "afternoon")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.00  60.75  132.50  169.59  241.00  500.00
> summary(timezones(epoch(sensor1,"spring"), "afternoon")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  35.0  78.0  112.4  157.0  500.0
> summary(timezones(epoch(sensor1,"summer"), "afternoon")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  56.0  96.0  130.2  168.0  500.0

```

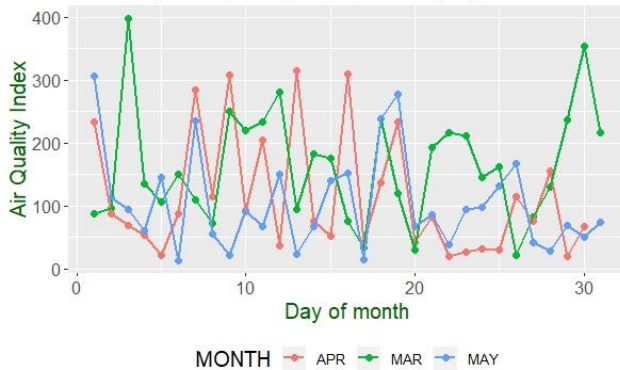
mean AQI value per day in Afternoon Autumn



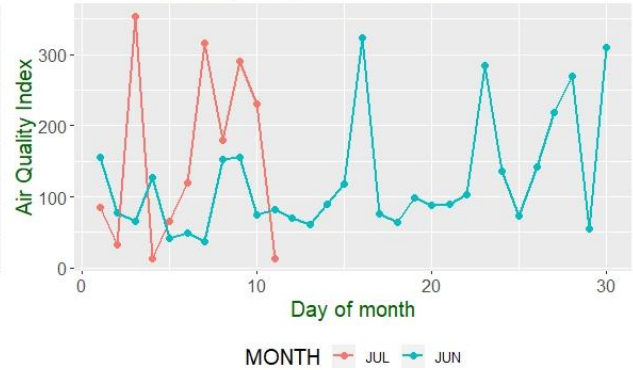
mean AQI value per day in Afternoon Winter



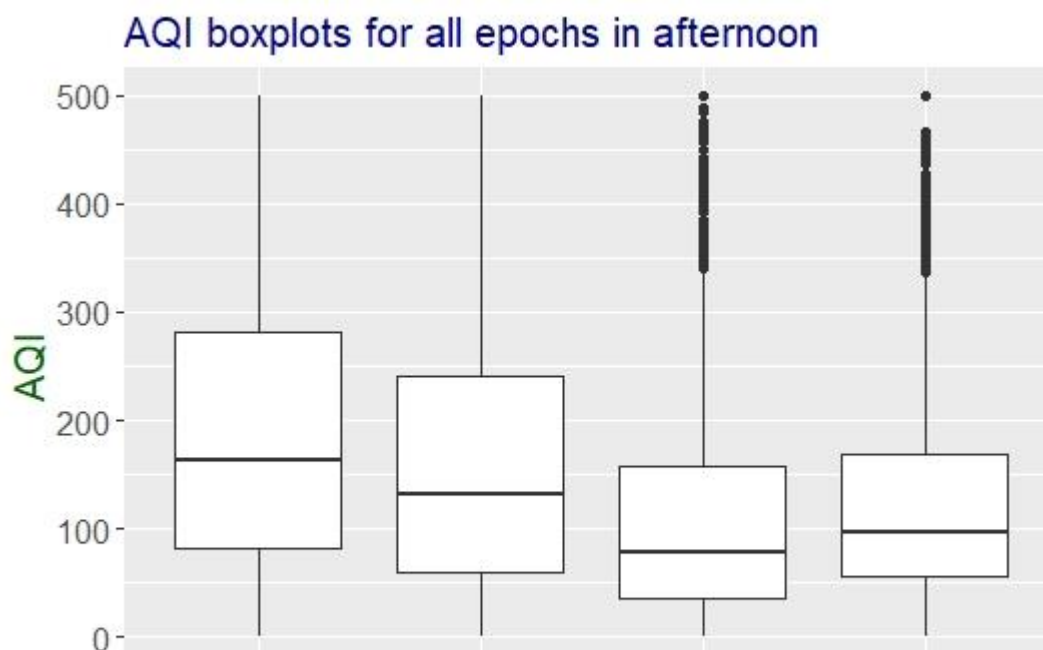
mean AQI value per day in Afternoon Spring



mean AQI value per day in Afternoon Summer



Γράφημα 17: Γραφήματα μέσω των τιμών AQI τις μεσημεριανές ώρες ανά εποχή.



Γράφημα 18: Boxplots φαίνεται η κατανομή των τιμών AQI τις μεσημεριανές ώρες ανά εποχή

- **Απόγευμα ανά εποχή:**

Οι ώρες 17:00 – 20:59 αναφέρονται στις απογευματινές ώρες. Οι συγκεκριμένες ώρες συνηθίζεται η ανθρώπινη δραστηριότητα να είναι πιο έντονη σε σχέση με αυτές των μεσημεριανών ωρών. Παρακάτω θα δούμε πώς κυμάνθηκαν οι τιμές AQI για τις συγκεκριμένες ώρες.

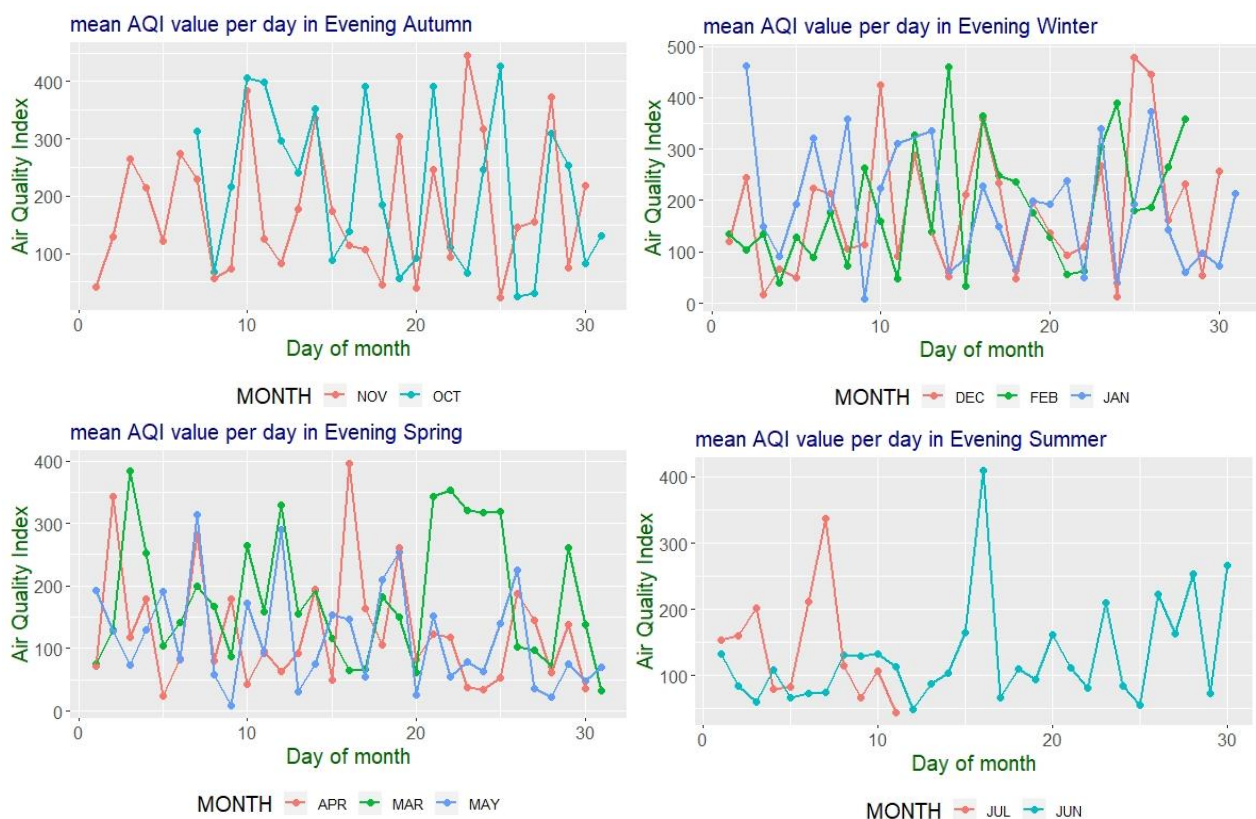
Το 42,1% των δεδομένων για τις απογευματινές ώρες του φθινοπώρου βρίσκονται πάνω από το μέσο όρο (>195). Το 42,1% με βάση των πίνακα 1 του EPA βρίσκεται αφορά ανθυγιεινές τιμές με το 24,08% είναι οι καταγεγραμμένες τιμές που αναφέρονται στην βλαβερή ποιότητα της ατμοσφαιράς για την περιοχή Κάτω Μάννα. Βλέπουμε λοιπόν μία αύξηση των ποσοστών κατά 7% σε σύγκριση με τις μεσημεριανές ώρες για το φθινόπωρο. Το 33,2% των δεδομένων για τις απογευματινές μέρες του φθινοπώρου αναφέρονται στη «καλή» και «μέτρια» ποιότητα αέρα του φθινοπώρου (<101).

Τον χειμώνα, 42% των δεδομένων για τις απογευματινές ώρες βρίσκονται πάνω από τον μέσο όρο (>185) δηλαδή στις ανθυγιεινές τιμές. Στο «βλαβερό» φάσμα της ποιότητας αέρα (>300) το ποσοστό των δεδομένων είναι 20,8%. Παρατηρούμε μείωση σε σύγκριση με τα ποσοστά του φθινοπώρου αλλά σημαντική αύξηση σε σχέση με τα ποσοστά που καταγράφηκαν τις μεσημεριανές ώρες του χειμώνα.

Συνεχίζοντας με τις απογευματινές ώρες της άνοιξης, βλέπουμε σημαντική μείωση του μέσου όρου (129) αλλά και μειωμένο ποσοστό των τιμών που βρίσκονται πάνω από αυτόν (38,1%). Εξίσου σημαντική μείωση βλέπουμε την άνοιξη και στα ποσοστά που αναφέρονται στις βλαβερές τιμές της ποιότητας του αέρα (>300). Μόλις το 8,5% αναφέρονται σε αυτές τις τιμές. 12% μείωση σε σύγκριση με τον χειμώνα. Ωστόσο, σε σύγκριση με τις μεσημεριανές ώρες της άνοιξης, υπάρχει μία αύξηση της τάξεως το 1,5%. Άξιο αναφοράς είναι ότι την άνοιξη το 51% των δεδομένων αντιστοιχούν στην «καλή» και «μέτρια» ποιότητα του αέρα.

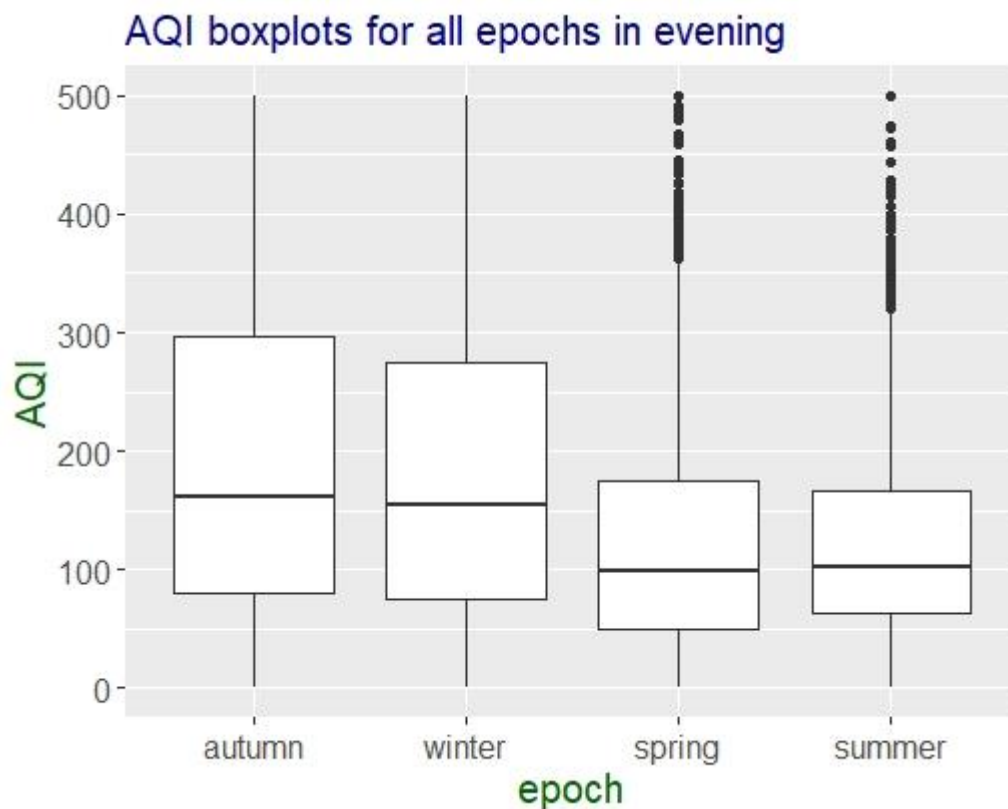
Τέλος, το καλοκαίρι, το 36,1% των δεδομένων βρίσκονται πάνω από το μέσο όρο (>132) με 8% των δεδομένων να αναφέρονται στο βλαβερό εύρος. Βλέπουμε μείωση σε σχέση με τα ποσοστά των απογευματινών ωρών της άνοιξης καθώς και μείωση των βλαβερών ποσοστών σε σχέση με τις μεσημεριανές ώρες κατά 3,1%. Στο εύρος των «καλών» και «μέτριων» τιμών ποιότητας αέρα αναφέρεται το 48,3% των δεδομένων.

Στο **Γράφημα 19** φαίνονται πώς κυμαίνονται οι μέσες τιμές AQI τις απογευματινές ώρες ανά ημέρα και για τις τέσσερις εποχές. Στο **Γράφημα 20** φαίνεται πώς κατανέμονται οι



Γράφημα 19: Γραφήματα μέσων τιμών AQI τις απογευματινές ώρες ανά εποχή.

τιμές AQI τις απογευματινές ώρες ανά ημέρα. Υπενθυμίζεται ότι τους φθινοπωρινούς και τους καλοκαιρινούς μήνες τα δεδομένα δεν είναι πλήρη οπότε δεν μπορούμε να βγάλουμε επαρκή συμπεράσματα. Φαίνεται όμως η διαφορά μεταξύ των χειμερινών και των ανοιξιάτικων μηνών.



Γράφημα 20: Boxplots φαίνεται η κατανομή των τιμών AQI τις απογευματινές ώρες ανά εποχή

```
> summary(timezones(epoch(sensor1,"autumn"),"evening")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.00 79.75 161.00 195.35 296.00 500.00
> summary(timezones(epoch(sensor1,"winter"),"evening")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.0 74.0 155.0 185.4 274.0 500.0
> summary(timezones(epoch(sensor1,"spring"),"evening")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.0 50.0 98.0 128.9 175.0 500.0
> summary(timezones(epoch(sensor1,"summer"),"evening")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.0 63.0 102.0 132.2 165.5 500.0
```

- **Βράδυ ανά εποχή:**

Οι ώρες 21:00 – 4:59 αντιστοιχούν στις βραδινές ώρες. Παρακάτω θα αναφέρονται τα ίδια στατιστικά από τα οποία μπορούμε να βγάλουμε σημαντικά συμπεράσματα για τις τιμές AQI.

Το 39,2% των δεδομένων για τις απογευματινές ώρες του φθινοπώρου βρίσκονται πάνω από το μέσο όρο (>188). Αυτό το ποσοστό, με βάση των πίνακα 1 του EPA, βρίσκεται αφορά ανθυγιεινές τιμές και το 20,3% είναι οι καταγεγραμμένες τιμές που αναφέρονται στην βλαβερή ποιότητα της ατμοσφαιράς. Βλέπουμε λοιπόν μία μείωση των ποσοστών κατά 3,7% σε σύγκριση με τις απογευματινές ώρες για το φθινόπωρο. Το 30,9% των δεδομένων για τις απογευματινές μέρες του φθινοπώρου αναφέρονται στη «καλή» και «μέτρια» ποιότητα αέρα του φθινοπώρου (<101).

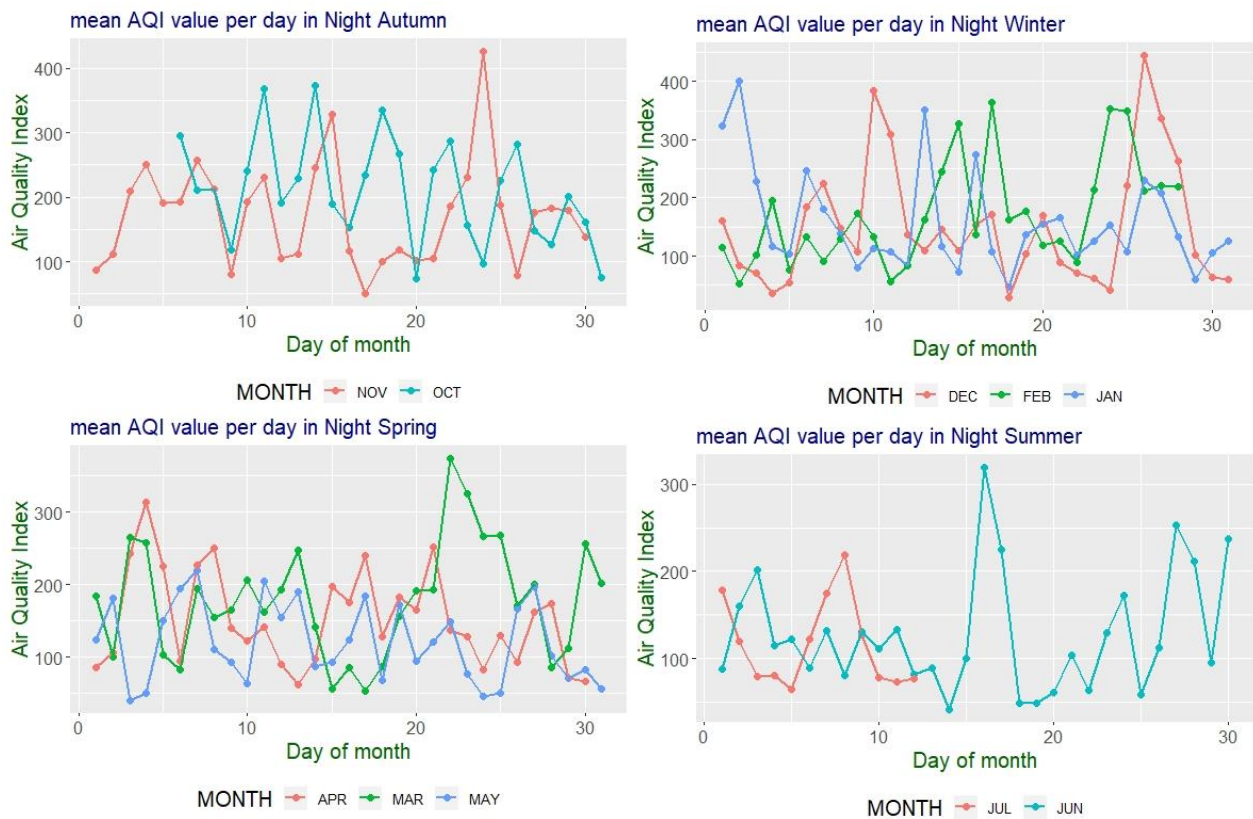
Τον χειμώνα, 42,2% των δεδομένων για τις απογευματινές ώρες βρίσκονται πάνω από τον μέσο όρο (>156). Στο «βλαβερό» φάσμα της ποιότητας αέρα (>300) το ποσοστό των δεδομένων είναι 15,2%. Βλέπουμε μία μείωση σε σύγκριση με τα ποσοστά του φθινοπώρου. Επίσης παρατηρούμε μείωση 5,6 ποσοστιαίων μονάδων σε σχέση με τα ποσοστά που καταγράφηκαν τις απογευματινές ώρες του χειμώνα.

Συνεχίζοντας με τις απογευματινές ώρες της άνοιξης, βλέπουμε σημαντική μείωση του μέσου όρου (137) αλλά και μειωμένο ποσοστό των τιμών που βρίσκονται πάνω από αυτόν (40,8%). Ακόμα, μείωση παρατηρούμε την άνοιξη και στα ποσοστά που αναφέρονται στις βλαβερές τιμές της ποιότητας του αέρα (>300) με ποσοστό 10,2%, μειωμένο δηλαδή κατά 5% σε σύγκριση με τον χειμώνα. Ωστόσο, σε σύγκριση με τις απογευματινές ώρες της άνοιξης, υπάρχει μία αύξηση της τάξεως του 1,7%. Τέλος, αξίζει να αναφέρουμε ότι την άνοιξη το 47,1% των δεδομένων αντιστοιχούν στην «καλή» και «μέτρια» ποιότητα του αέρα.

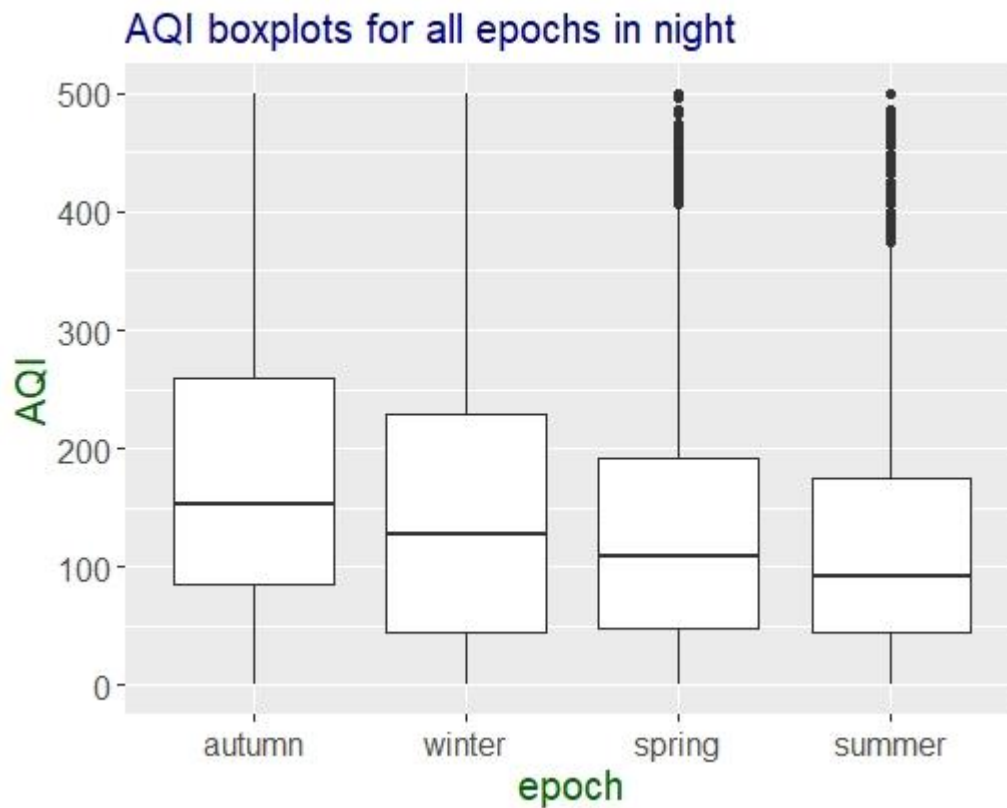
Κλείνοντας με τον αισθητήρα της περιοχής Κάτω Μάννα, τη καλοκαιρινή περίοδο, το 40,8% των δεδομένων βρίσκονται πάνω από το μέσο όρο (>125) με 7,1% των δεδομένων να αναφέρονται στο βλαβερό εύρος. Βλέπουμε λοιπόν εκ νέου μείωση των ποσοστών των βλαβερών δεδομένων σε σχέση με τα ποσοστά των βραδινών ωρών της άνοιξης καθώς και μείωση των βλαβερών ποσοστών σε σχέση με τις απογευματινές ώρες κατά 0,9%. Στο

εύρος των «καλών» και «μέτριων» τιμών ποιότητας αέρα αναφέρεται το 52,5% των δεδομένων.

Τα παραπάνω μπορούν να φανούν και στο **Γράφημα 22** περιμένοντας όμως την συλλογή δεδομένων όλου του καλοκαιριού για να κάνουμε συγκρίσεις. Στο **Γράφημα 21**, ο αναγνώστης μπορεί να παρατηρήσει πώς κυμάνθηκαν κατά μέσο όρο οι τιμές AQI κάθε βράδυ ανά εποχή.



Γράφημα 21: Γραφήματα μέσων τιμών AQI τις βραδινές ώρες ανά εποχή.



Γράφημα 22: Boxplots φαίνεται η κατανομή των τιμών AQI τις βραδινές ώρες ανά εποχή

```
> summary(timezones(epoch(sensor1,"autumn"),"night")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  85.0  153.0 188.4 260.0 500.0
> summary(timezones(epoch(sensor1,"winter"),"night")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  45.0  127.0 156.9 229.0 500.0
> summary(timezones(epoch(sensor1,"spring"),"night")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  48.0  108.0 137.9 191.0 500.0
> summary(timezones(epoch(sensor1,"summer"),"night")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  44.0   91.0 125.1 175.5 500.0
```

#### 4.2.2 sensor: 8200015F (Περιοχή: Λιμάνι)

Ο αισθητήρας αυτός βρίσκεται σε κεντρικό σημείο της Σύρου. Συγκεκριμένα, είναι τοποθετημένος τόσο κοντά στο λιμάνι όσο και κοντά σε κεντρική οδό της Ερμούπολης όπου υπάρχει έντονη ανθρώπινη δραστηριότητα. Επίσης, η ποιότητα του αέρα επηρεάζεται και από τις αφίξεις και τις διαμονές πλοίων στο λιμάνι. Αυτή η επιρροή εξαρτάται από τον αριθμό των πλοίων που έρχονται καθημερινά στο λιμάνι καθώς και πόση ώρα κάθονται ενεργά σε αυτό. Η κεντρική οδός επηρεάζει και αυτή την ποιότητα του αέρα. Τόσο ο μεγάλος πληθυσμός όσο και τα πολλά οχήματα καθιστούν την ανθρώπινη δραστηριότητα υψηλή με αποτέλεσμα η ποιότητα του αέρα να αγγίζει ανθυγιεινά, για τον ανθρώπινο οργανισμό, επίπεδα.

Στο παρόν κεφάλαιο, σχολιάζεται η «συμπεριφορά» των AQI τιμών στο λιμάνι στο πνεύμα του κεφαλαίου 4.2.1. Ο αισθητήρας του Λιμανιού έχει έλλειψη δεδομένων για την εποχή του φθινοπώρου οπότε και γίνεται σχολιασμός μόνο των εποχών του χειμώνα, της άνοιξης και το μισό καλοκαίρι.

- **Πρωί ανά εποχή:**

Ξεκινώντας με τους χειμερινούς μήνες παρατηρούμε υψηλό μέσο όρο για τις πρωινές ώρες 5:00 – 11:59. Συγκεκριμένα η μέση τιμή AQI είναι 278 και βρίσκεται στις πολύ ανθυγιεινές τιμές σύμφωνα με τον **Πίνακα 1** της EPA. Ανησυχία θα πρέπει να προκαλεί το γεγονός ότι τον χειμώνα το 50,5% των τιμών ανήκει πάνω από τον μέσο όρο και το 43,7% υπάρχει στο φάσμα των βλαβερών τιμών. Μόλις το 9,4% των εγγραφών ανήκει στο εύρος των «καλών» και «μέτριων» τιμών.

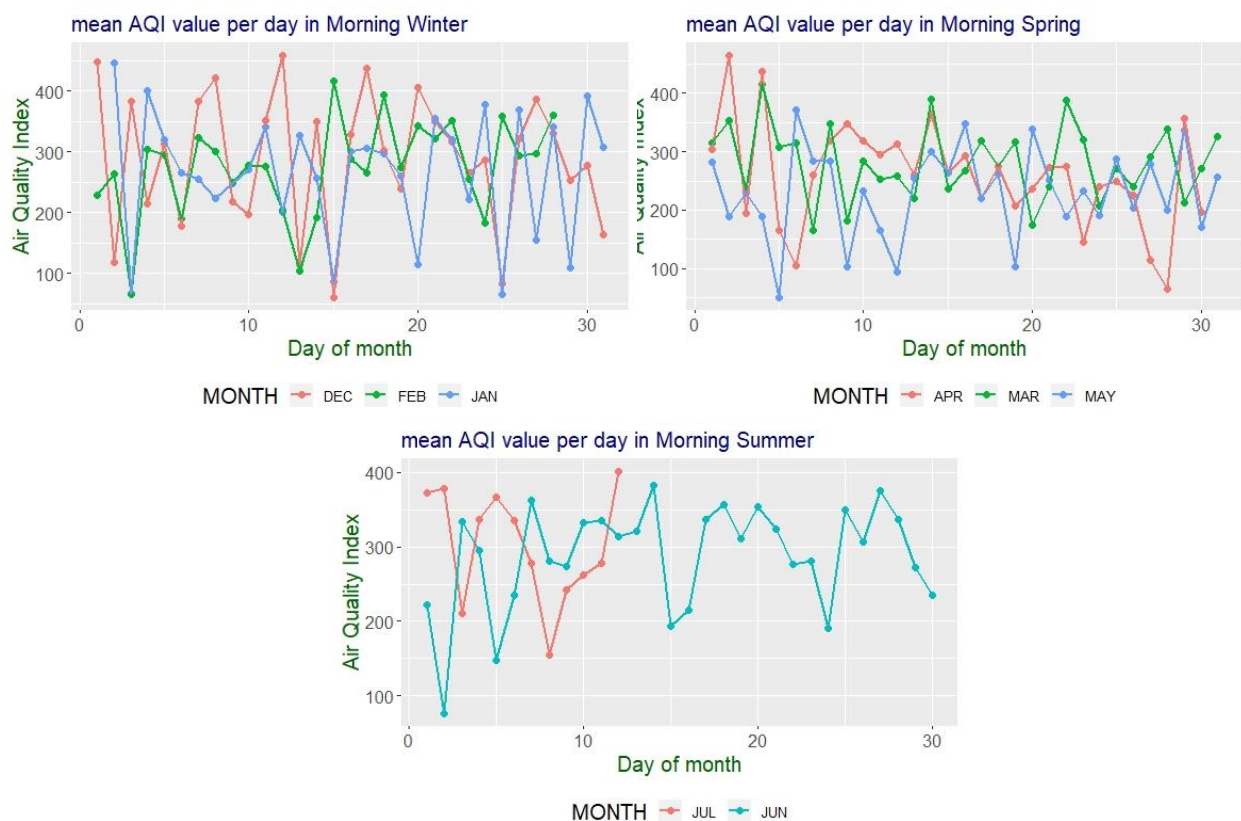
Συνεχίζοντας με την άνοιξη, παρατηρούμε ότι ο μέσος όρος υποχωρεί αλλά παραμένει σε υψηλά «πολύ ανθυγιεινά» επίπεδα με το 49,1% των τιμών να βρίσκεται πάνω του μέσου όρου. Τα ποσοστά των «βλαβερών» επιπέδων υποχωρούν σε σχέση με αυτές του χειμώνα κατά 14,7% δηλαδή το 29% των εγγραφών βρίσκεται σε τιμές AQI μεγαλύτερες του 300. Μικρή αύξηση υπάρχει και στα ποσοστά για τα «καλά» και «μέτρια» επίπεδα της ποιότητας του αέρα (12,3%).

Τέλος, τους καλοκαιρινούς μήνες η πρώτη ανάγνωση των δεδομένων εξακολουθεί να έχει αρνητικό πρόσημο. Ο μέσος όρος ανεβαίνει πιο ψηλά και από αυτόν του χειμώνα (288) και το 51,5% των τιμών βρίσκεται πάνω από τον μέσο όρο. Στο «βλαβερό» εύρος ανήκουν

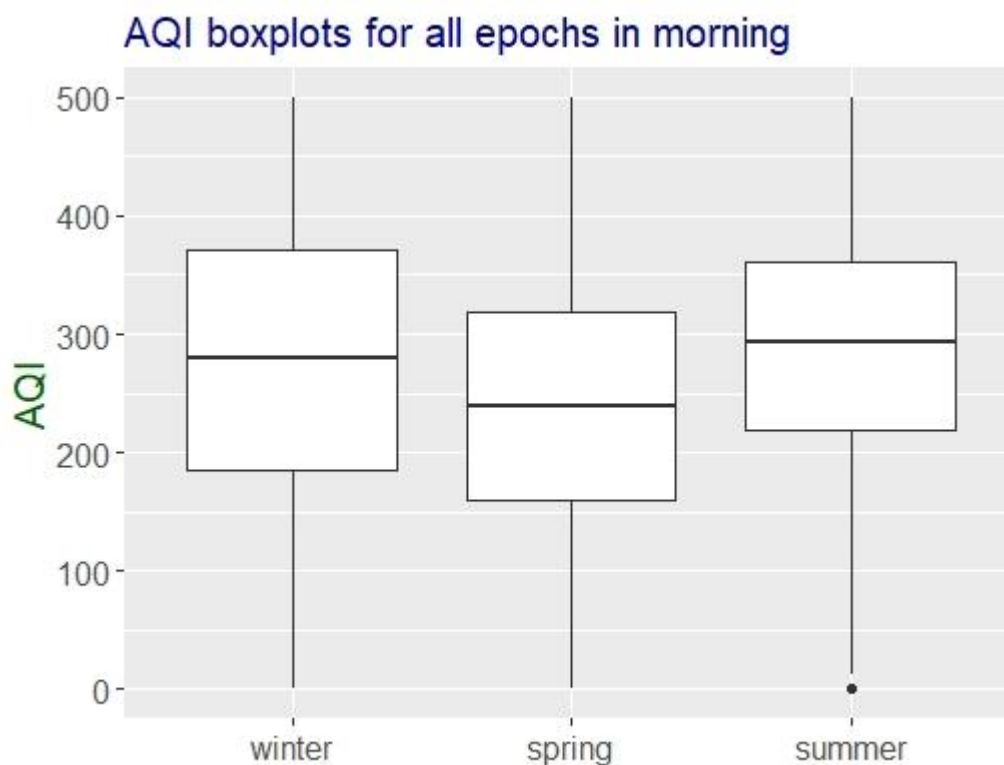
το 46.8% των εγγραφών των δεδομένο. Δυστυχώς, μόλις το 4% των εγγραφών βρίσκεται στο εύρος των «καλών» και «μέτριων» τιμών AQI.

Τα παραπάνω μπορούν να παρατηρηθούν και από τα **Γράφημα 23** και **Γράφημα 24**. Οι μέσες τιμές σε καθημερινή βάση φαίνεται να αγγίζουν υψηλές τιμές καθώς και η κατανομή των δεδομένων και για τις τέσσερις εποχές όπως φαίνεται από τα boxplots βρίσκονται σε αρκετά υψηλά επίπεδα. Αρκετά υψηλότερα σε σύγκριση με τα δεδομένα που μας παρείχε ο αισθητήρας στην περιοχή Κάτω Μάννα.

```
> summary(timezones(epoch(sensor2,"winter"),"morning")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  185.0  280.0 278.3  371.0 500.0
> summary(timezones(epoch(sensor2,"spring"),"morning")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  160.0  239.0 240.9  318.0 500.0
> summary(timezones(epoch(sensor2,"summer"),"morning")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  219.0  293.0 288.1  360.2 500.0
```



Γράφημα 23: Γραφήματα μέσων τιμών AQI τις πρωινές ώρες ανά εποχή.



Γράφημα 24: Boxplots φαίνεται η κατανομή των τιμών AQI τις πρωινές ώρες ανά εποχή

- **Μεσημέρι ανά εποχή:**

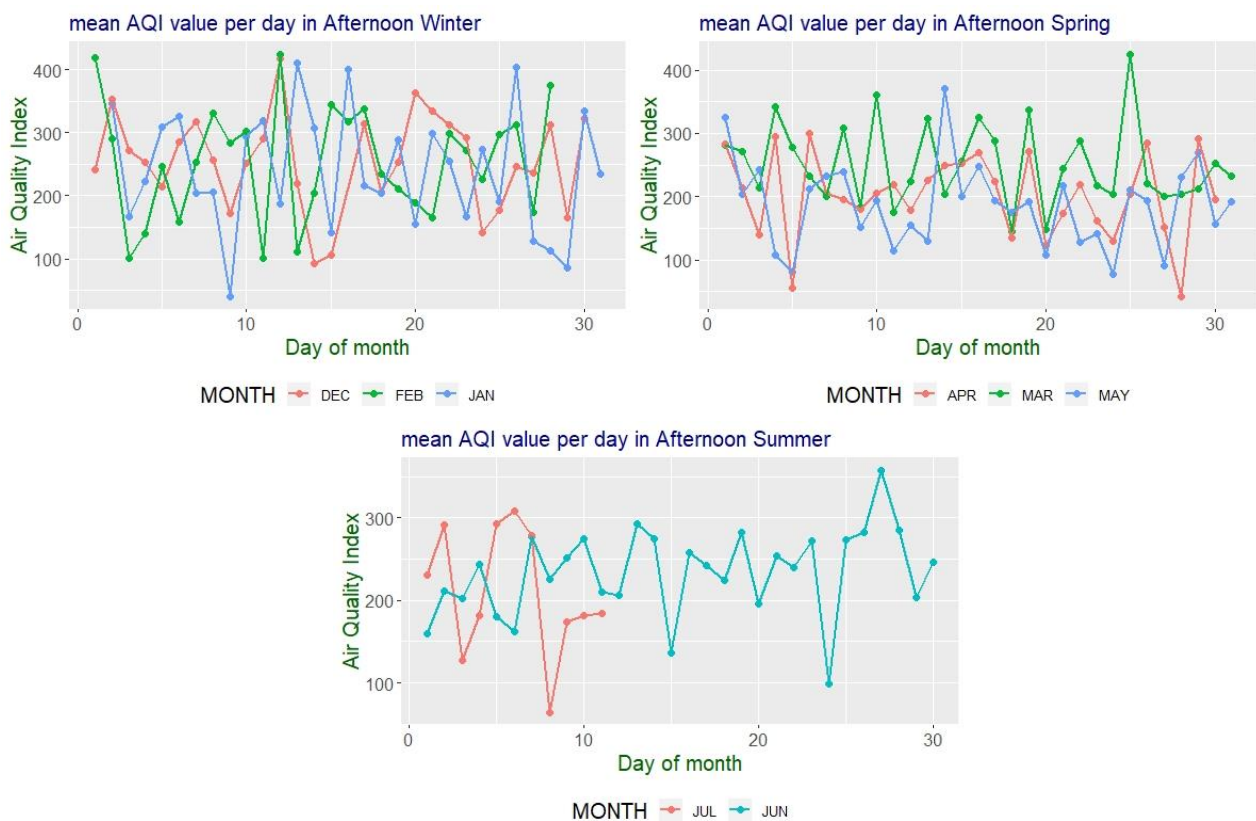
Συνεχίζοντας με τις μεσημεριανές ώρες, παρατηρούμε ότι είναι εξίσου υψηλός είναι ο μέσος όρος τον χειμώνα με τιμή AQI 251 και το 48,8% των τιμών βρίσκεται πάνω από το μέσο όρο. Το 34% των εγγραφών βρίσκεται στα «βλαβερά» επίπεδα, μειωμένο κατά 9% σε σχέση με τις πρωινές ώρες του χειμώνα. Το 9,8% των τιμών βρίσκεται στα «καλά» και «μέτρια» επίπεδα ελάχιστα αυξημένο ποσοστό σε σχέση με τις πρωινές ώρες.

Την άνοιξη ο μέσος όρος έχει τιμή 200 με το 46,9% των τιμών να ανήκει πάνω από το μέσο όρο. Αυτό μεταφράζεται ότι το 30% των τιμών βρίσκεται στα «πολύ βλαβερά» επίπεδα και το 16,8% βρίσκεται στα «βλαβερά» επίπεδα. Αρκετά μειωμένα τα βλαβερά επίπεδα σε σχέση με τις πρωινές ώρες της άνοιξης. Το 18,4% των τιμών βρίσκεται υπό το φάσμα των «καλών» και «μέτριων» τιμών. Σε σύγκριση με τις πρωινές ώρες της άνοιξης βλέπουμε αύξηση.

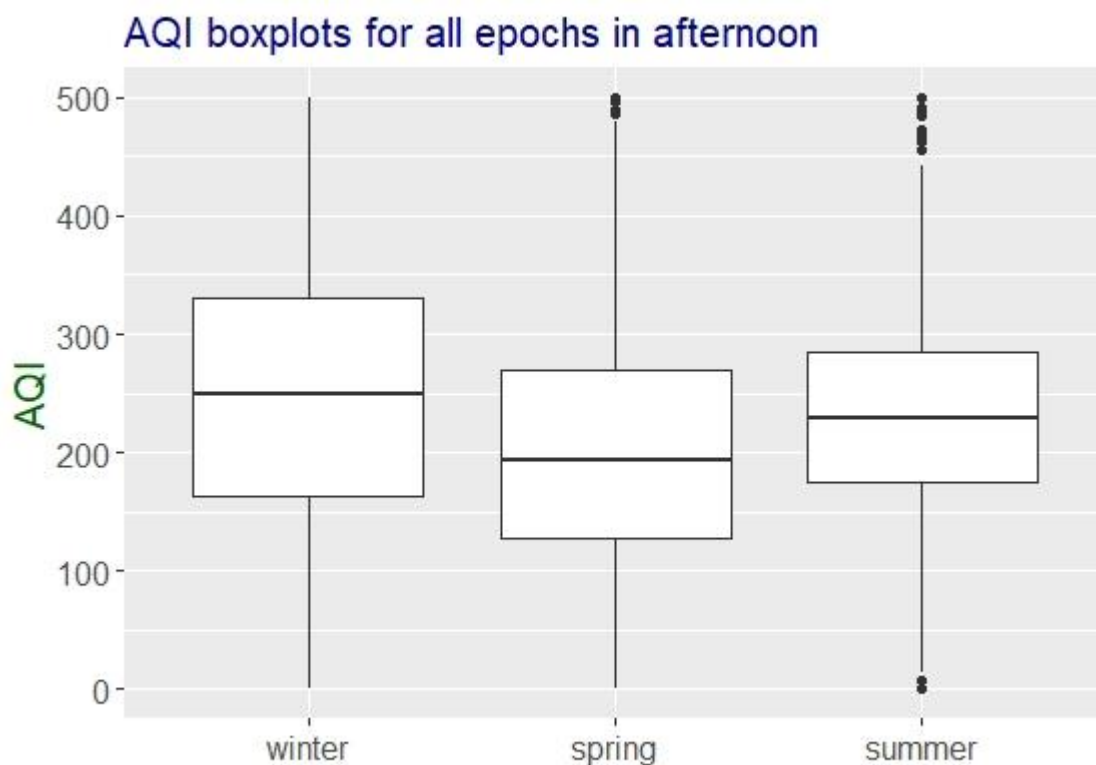
Κλείνοντας με τους καλοκαιρινούς μήνες βλέπουμε ότι ο μέσος όρος βρίσκεται στο επίπεδο των «πολύ βλαβερών» τιμών AQI (230). Το 49,8% βρίσκεται άνω του μέσου όρου με 20% των δεδομένων να βρίσκεται στα «βλαβερά επίπεδα. Βλέπουμε μία σημαντική μείωση των βλαβερών αυτών τιμών AQI κατά 26,8% σε σχέση με τις πρωινές ώρες του καλοκαιριού. Το 7,3% των εγγραφών ανήκουν στα «καλά» και «μέτρια» επίπεδα.

Παρατηρώντας τα **Γράφημα 25** και **Γράφημα 26** και συγκρίνοντας τα με τα αντίστοιχα γραφήματα των πρωινών ωρών βλέπουμε ότι οι τιμές υποχωρούνε αλλά το πρόβλημα παραμένει εξίσου σοβαρό για τον ανθρώπινο οργανισμό. Πιθανώς αυτή η μείωση να οφείλεται στην μειωμένη ανθρώπινη δραστηριότητα τις μεσημεριανές ώρες για κάθε εποχή.

```
> summary(timezones(epoch(sensor2,"winter"),"afternoon")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  163.0  249.0 251.5  331.0 500.0
> summary(timezones(epoch(sensor2,"spring"),"afternoon")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  127.0  193.0 200.3  269.0 500.0
> summary(timezones(epoch(sensor2,"summer"),"afternoon")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  174.5  229.0 230.5  285.0 500.0
```



Γράφημα 25: Γραφήματα μέσω των τιμών AQI τις μεσημεριανές ώρες ανά εποχή.



Γράφημα 26: Boxplots φαίνεται η κατανομή των τιμών AQI τις μεσημεριανές ώρες ανά εποχή

- **Απόγευμα ανά εποχή:**

Συνεχίζοντας με τις απογευματινές ώρες που αντιστοιχούν στις ώρες 17:00-20:59 βλέπουμε ότι την χειμερινή περίοδο ο μέσος όρος των τιμών AQI είναι αρκετά υψηλό με τιμή 253. Το 48,2% των τιμών αυτών βρίσκεται πάνω από το μέσο όρο με το 37% των τιμών να βρίσκονται στο εύρος των «βλαβερών» τιμών. «Καλή» και «μέτρια» ποιότητα αέρα χαρακτηρίζεται το 12,7% των εγγραφών. Ποσοστό μεγαλύτερο τόσο των πρωινών όσο και των μεσημεριανών ωρών.

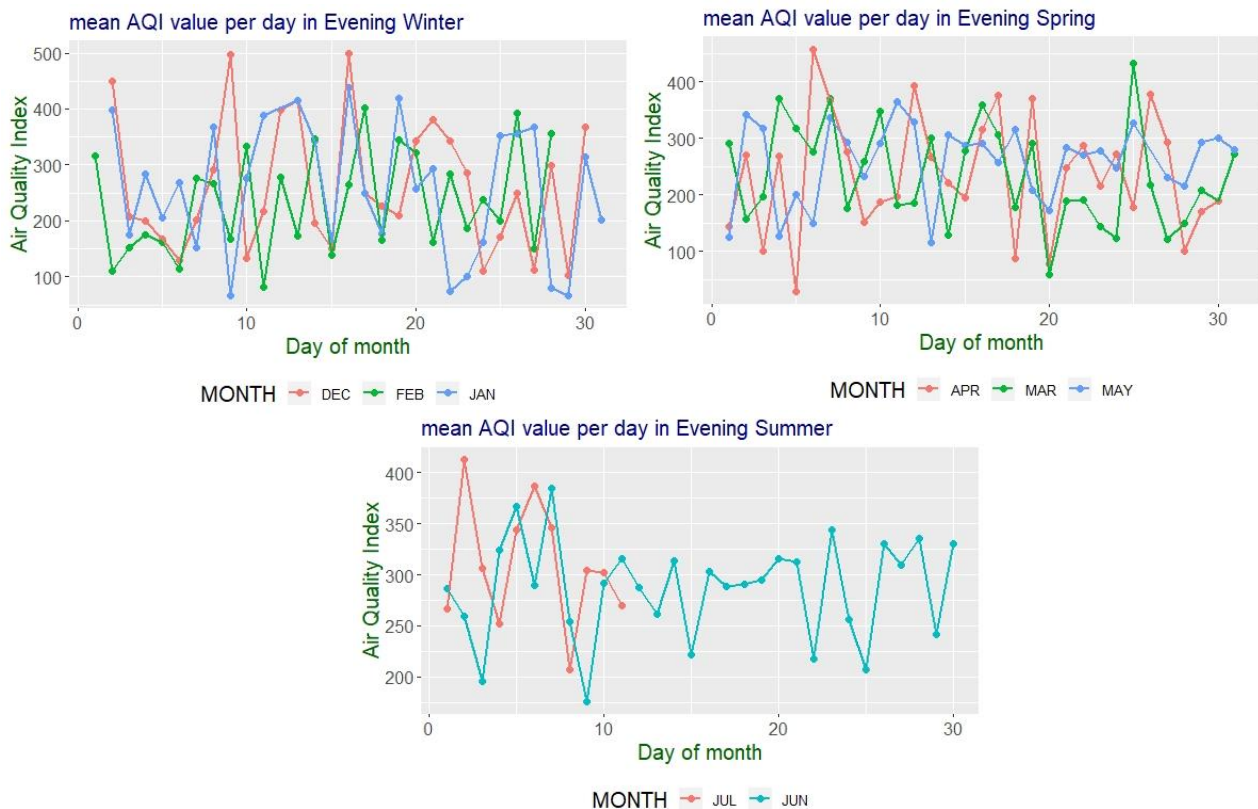
Την περίοδο της άνοιξης ο μέσος όρος βρίσκεται αρκετά κοντά με αυτόν του χειμώνα (249) με το 49,2% να βρίσκεται πάνω από αυτή την τιμή. Το 32,5% των εγγραφών βρίσκεται στο «βλαβερό» εύρος ποιότητας αέρα. Ποσοστό αυξημένο σε σύγκριση με τις δύο προηγούμενες ζώνες ωρών. Το 9,6% των δεδομένων χαρακτηρίζεται «καλή» και «μέτρια» ποιότητα αέρα.

Κλείνοντας με την απογευματινή ζώνη ώρας, την καλοκαιρινή περίοδο ο μέσος όρος των τιμών AQI είναι 291. Αριθμός δηλαδή πολύ κοντά στα «βλαβερά» επίπεδα. Το ποσοστό των δεδομένων ποιότητας αέρα που ανήκει στα «βλαβερά» επίπεδα το καλοκαίρι είναι

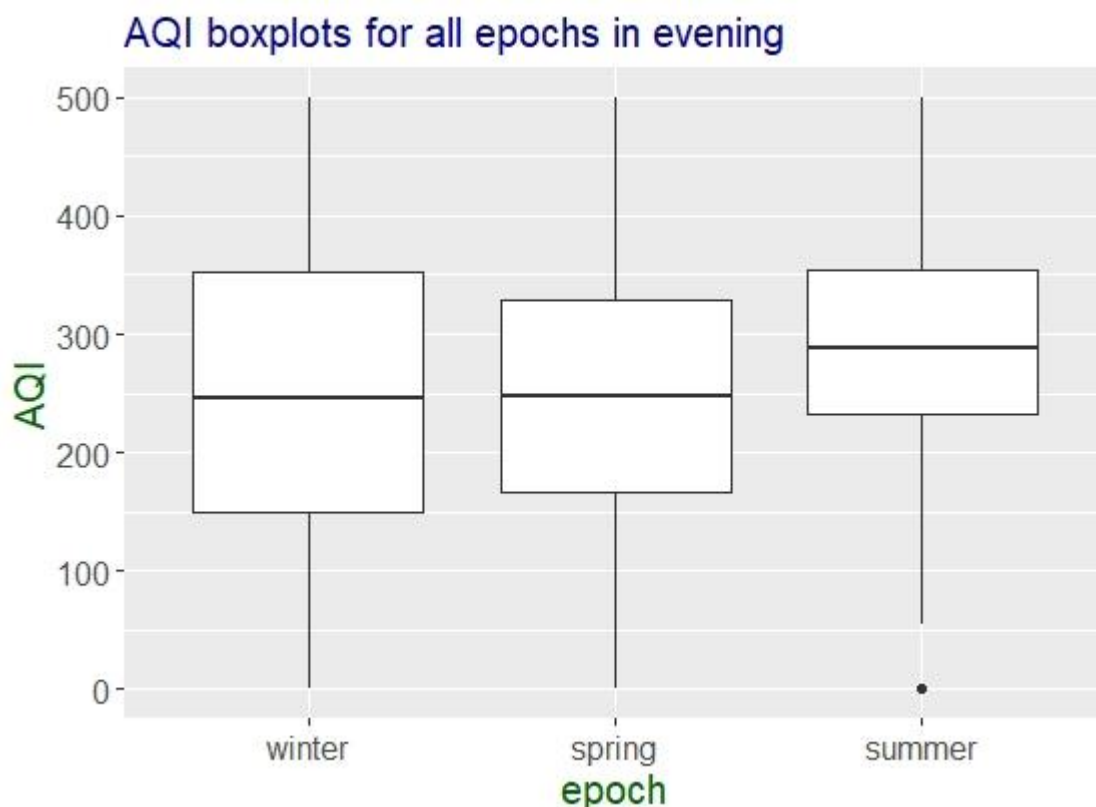
45%, ποσοστό αντίστοιχο με αυτό που παρατηρούμε τις πρωινές καλοκαιρινές ώρες. Μόλις το 1,8% των εγγεγραμμένων τιμών AQI μπορεί να χαρακτηριστεί «καλή» ή «μέτρια» ποιότητα του αέρα.

Στα **Γράφημα 27** και **Γράφημα 28** μπορεί και φαίνεται αυτή η αύξηση των τιμών που περιεγράφηκε παραπάνω. Συγκεκριμένα παρατηρείται ότι, ενώ υπάρχει μείωση τις μεσημεριανές ώρες σε κάθε εποχή, στις απογευματινές ώρες υπάρχει εκ νέου αύξηση των τιμών αντίστοιχες με αυτές των πρωινών ωρών.

```
> summary(timezones(epoch(sensor2,"winter"),"evening")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  148.8  246.0 253.3  353.2 500.0
> summary(timezones(epoch(sensor2,"spring"),"evening")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  167.0  248.0 249.1  329.0 500.0
> summary(timezones(epoch(sensor2,"summer"),"evening")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  232.0  288.0 291.7  354.0 500.0
```



Γράφημα 27: Γραφήματα μέσων τιμών AQI τις απογευματινές ώρες ανά εποχή.



Γράφημα 28: Boxplots φαίνεται η κατανομή των τιμών AQI τις απογευματινές ώρες ανά εποχή

- **Βράδυ ανά εποχή:**

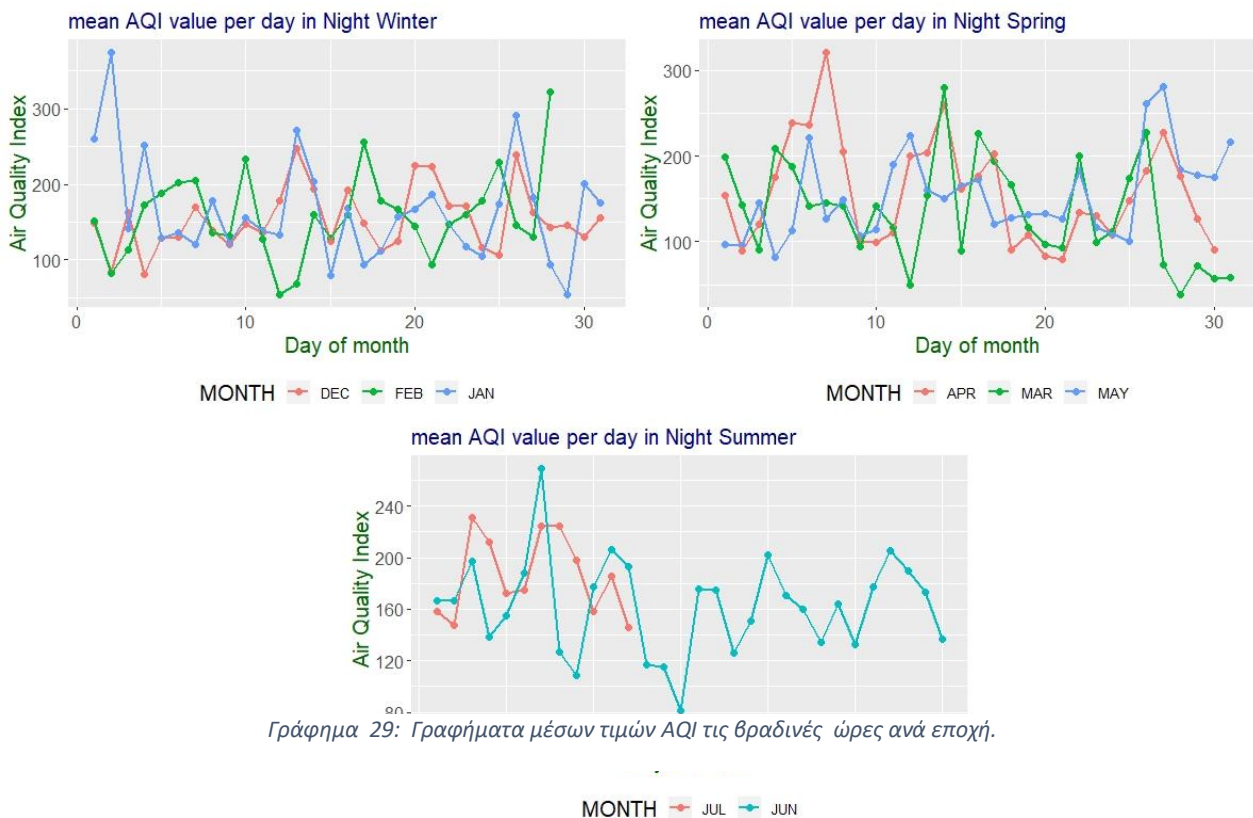
Οι βραδινές ώρες αντιστοιχούν στις 21:00 – 4:59. Την χειμερινή περίοδο η μέση τιμή είναι αισθητά χαμηλότερη σε σχέση με τις προηγούμενες ζώνες ώρας (161). Το 28,5% των τιμών βρίσκεται στα «βλαβερά» επίπεδα με τιμές AQI μεγαλύτερες του 300. Σε σύγκριση με τις προηγούμενες ζώνες ώρας και εποχές η ποιότητα του αέρα είναι σε πολύ καλύτερα επίπεδα. Συγκεκριμένα το 28,5% βρίσκεται στα επιθυμητά επίπεδα ποιότητας του αέρα.

Την άνοιξη οι τιμές είναι εξίσου σημαντικά κάτω σε σύγκριση με τις προηγούμενες ζώνες ώρας. Ο μέσος όρος την άνοιξη είναι 147 και μόλις το 8% των τιμών αντιστοιχεί στο «βλαβερό» επίπεδο ποιότητας αέρα. Σε σύγκριση με την χειμερινή περίοδο υπάρχει αύξηση των ποσοστών της «καλής» και «μέτριας» ποιότητας αέρα φτάνοντας το 38,4%.

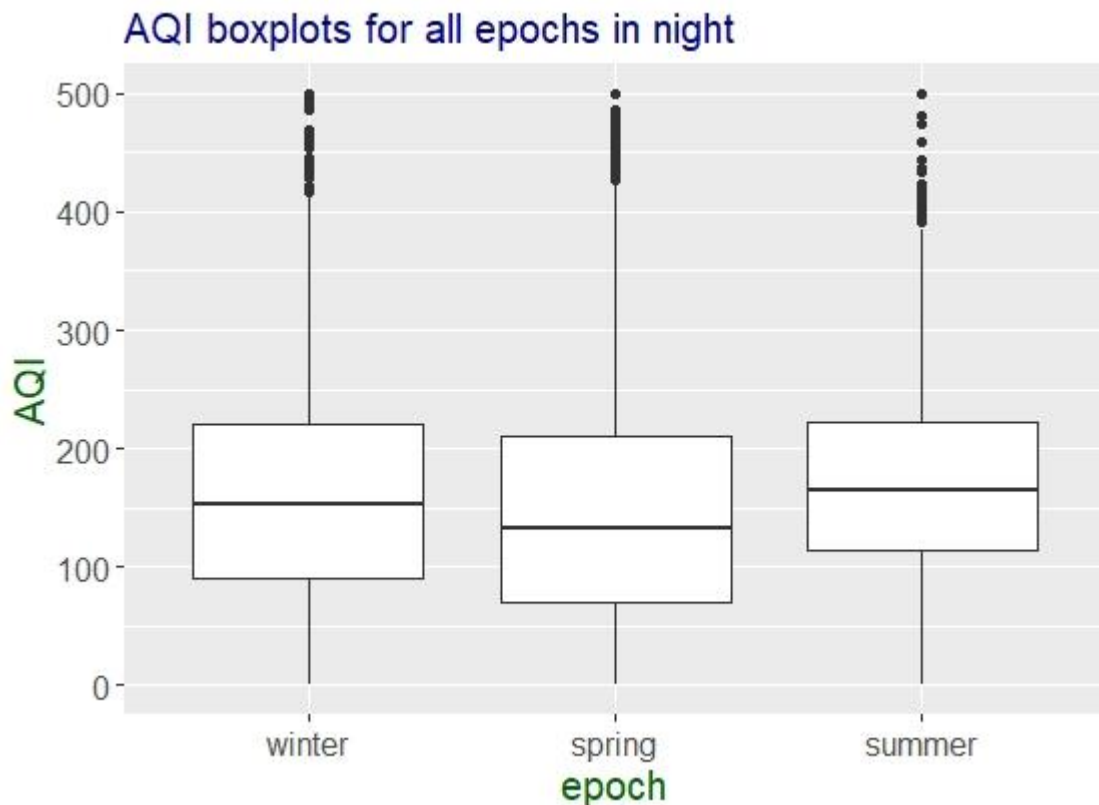
Τέλος, την καλοκαιρινή περίοδο ο μέσος όρος βρίσκει μία μικρή αύξηση της τιμής AQI του μέσου όρου (169). Το 6,5% των τιμών τις βραδινές ώρες αντιστοιχεί στην «βλαβερή» ποιότητα αέρα όπου σε σύγκριση με τα ποσοστά του καλοκαιριού στις απογευματινές ώρες φαίνεται ραγδαία μείωση. Το 19,8% των καταγεγραμμένων AQI τιμών αντιστοιχούν στην «καλή» και «μέτρια» ποιότητα αέρα. Περίπου 9% μειωμένο από τις βραδινές ώρες της άνοιξης και αλλά αυξημένο κατά 18% σε σχέση με τις απογευματινές ώρες της άνοιξης.

Παρατηρώντας τα **Γράφημα 29** και **Γράφημα 30** και συγκρίνοντας με τις προηγούμενα αντίστοιχα γραφήματα φαίνεται σημαντική μείωση των τιμών AQI στις βραδινές ώρες. Σε αυτό συμβάλει και το γεγονός ότι οι ώρες στη βραδινή περίοδο είναι περισσότερες σε σχέση με τις άλλες ζώνες ώρας.

```
> summary(timezones(epoch(sensor2,"winter"),"night")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0   90.0   152.0 161.4 220.0   500.0
> summary(timezones(epoch(sensor2,"spring"),"night")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0   69.0   132.0 147.4 211.0   500.0
> summary(timezones(epoch(sensor2,"summer"),"night")$vocaqi)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 0.0  113.0   164.0 169.4 222.0   500.0
```



Γράφημα 29: Γραφήματα μέσων τιμών AQI τις βραδινές ώρες ανά εποχή.



Γράφημα 30: Boxplots φαίνεται η κατανομή των τιμών AQI τις βραδινές ώρες ανά εποχή.

#### 4.2.2 Συμπεράσματα

Συνοψίζοντας λοιπόν τα παραπάνω παρατηρούμε ότι οι τιμές AQI ποικίλλει ανάλογα με την περίοδο. Οι Jiang, S., Ferreira, J., & González, M. C. [25] αναφέρουν ότι η ανθρώπινη δραστηριότητα συνηθίζεται να ξεκινάει από τις 6:00 έως τις 24:00 το βράδυ. Σε αυτό το διάστημα υπάρχουν δύο κορυφές στην ανθρώπινη δραστηριότητα και συγκεκριμένα η μία βρίσκεται στις πρωινές ώρες και η δεύτερη στις απογευματινές προς βραδινές ώρες. Με βάση αυτό και παρατηρώντας τις τιμές AQI ανά ζώνη ώρας και πώς αυτές κυμαίνονται θα μπορούσαμε εύκολα να βγάλουμε το συμπέρασμα ότι πράγματι αυτές οι AQI τιμές συσχετίζονται και είναι ανάλογες με την ανθρώπινη δραστηριότητα. Τα παραπάνω αποτελούν ένα ακόμα επιχείρημα ότι η ποιότητα του αέρα βρίσκεται συναρτήσει της ανθρώπινης συμπεριφοράς.

## ΚΕΦΑΛΑΙΟ 5

### 5.1 Εισαγωγή

Στα κεφάλαιο 3 και κεφάλαιο 4 έγινε μια ανάλυση των μετρήσεων από τους αισθητήρες που συλλέγουν δεδομένα ποιότητας του αέρα στην Σύρο. Σύμφωνα με τις παρατηρήσεις που έγιναν και συμπεράσματα που προέκυψαν, γίνεται γρήγορα αντιληπτό το πρόβλημα της μόλυνσης του αέρα στη Σύρο. Κρίνεται, λοιπόν, άμεση η ανάγκη δράσεων για την αντιμετώπιση της απειλής αυτής. Αρχικά, η παρακολούθηση των ποσοτήτων (monitoring) είναι το πρώτο στάδιο και η ενεργοποίηση δύο αισθητήρων από τους οποίους αναλύθηκαν τα δεδομένα σε προηγούμενα κεφάλαια, ήταν ένα πρώτο βήμα όπου και επιβεβαίωσαν το πρόβλημα. Αναμένεται στο μέλλον να τοποθετηθούν περισσότεροι αισθητήρες με σκοπό την περαιτέρω παρακολούθηση των ποσοτήτων ποιότητας του αέρα. Ωστόσο, κρίνεται άμεση και η πρόβλεψη της ποιότητας του αέρα με σκοπό την λήψη άμεσων δράσεων από την τοπική κοινωνία και κυρίως από άτομα με ευαισθησία στους ρύπους που πολλές φορές επικρατούν στην ατμόσφαιρα.

Στο παρόν κεφάλαιο, εισάγονται τεχνικές μηχανικής μάθησης με σκοπό την πρόβλεψη (regression) της ποιότητας του αέρα. Με την χρήση της γλώσσας προγραμματισμού R στο περιβάλλον του R-studio εκπαιδεύονται μοντέλα με στόχο να μας εξάγουν γνώση χρήσιμη για την ποιότητα του αέρα. Στο τέλος του κεφαλαίου, γίνεται σύγκριση της αποτελεσματικότητας αυτών των μοντέλων και διαφαίνεται ποιο μοντέλο παράγει τα πιο χρήσιμα αποτελέσματα.

Σημαντικό μειονέκτημα στην εκπαίδευση αυτών των μοντέλων είναι τόσο ο όγκος των δεδομένων όσο και η έλλειψη μετεωρολογικών δεδομένων που έχουν άμεση επίδραση στο τελικό αποτέλεσμα της ποιότητας του αέρα (όπως αναφέρθηκε στο Κεφάλαιο 1).

Σημειώνεται ότι τα παρακάτω μοντέλα θα εκπαιδευτούν και θα προσδιορίσουν τα αποτελέσματά τους μόνο για τον αισθητήρα 820015E και για την εποχή του χειμώνα.

## 5.2 Multiple Linear Regression

Η μέθοδος της πολλαπλής γραμμικής παλινδρόμησης δίνεται από τον τύπο (1) πιο γνωστή ως linear regression analysis. Χρησιμοποιείται για να εξηγήσει τις σχέσεις ανάμεσα σε μία συνεχή εξαρτημένη μεταβλητή και δύο ή περισσότερες ανεξάρτητες μεταβλητές. Υπάρχουν τρεις κύριες χρήσεις για ανάλυση multiple linear regression. Πρώτον, μπορεί να χρησιμοποιηθεί για να προσδιορίσει τη δύναμη του αποτελέσματος που έχουν οι ανεξάρτητες μεταβλητές σε μια εξαρτημένη μεταβλητή. Δεύτερον, μπορεί να χρησιμοποιηθεί για την πρόβλεψη των επιπτώσεων των αλλαγών. Δηλαδή, η ανάλυση multiple linear regression μας βοηθά να κατανοήσουμε πόσο θα μεταβληθεί η εξαρτημένη μεταβλητή όταν αλλάζουμε τις ανεξάρτητες μεταβλητές. Τρίτον, η ανάλυση multiple linear regression προβλέπει τάσεις και μελλοντικές τιμές. Η ανάλυση πολλαπλής γραμμικής παλινδρόμησης μπορεί να χρησιμοποιηθεί για να ληφθούν εκτιμήσεις σημείων.

Στην σχέση (1)  $y$  είναι η εξαρτημένη μεταβλητή,  $x_i$  οι ανεξάρτητες μεταβλητές και  $n$  το σύνολο των εξαρτώμενων μεταβλητών.

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n \quad (1)$$

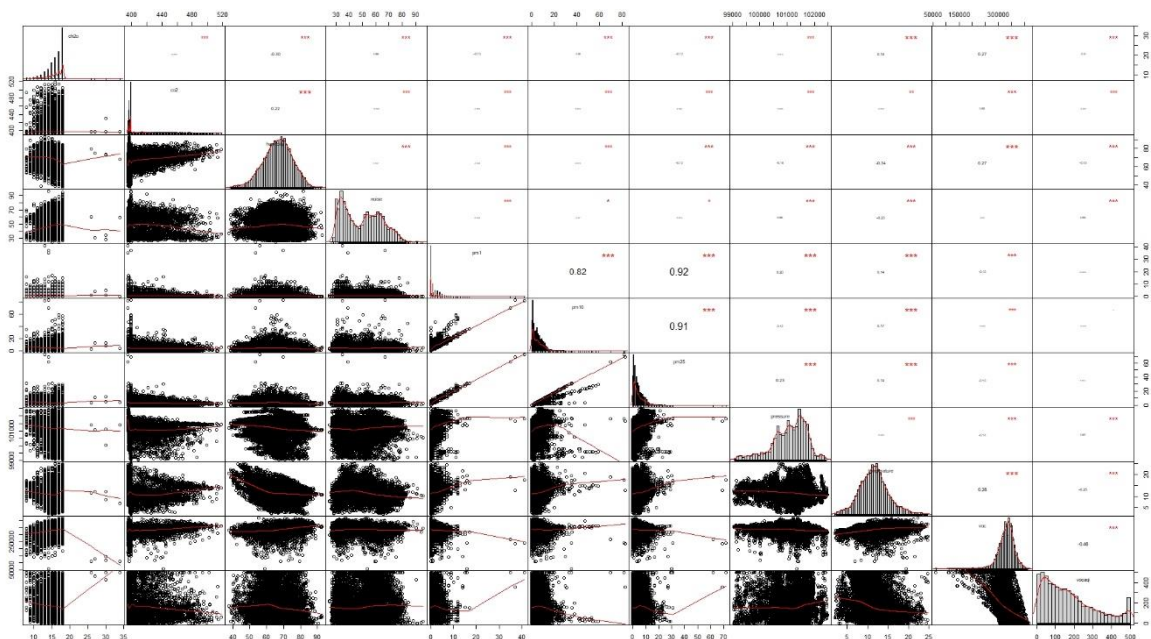
Κατά την επιλογή του μοντέλου για την ανάλυση πολλαπλής γραμμικής παλινδρόμησης, ένα άλλο σημαντικό στοιχείο είναι η προσαρμογή του μοντέλου. Η προσθήκη ανεξάρτητων μεταβλητών σε ένα μοντέλο πολλαπλής γραμμικής παλινδρόμησης θα αυξάνει πάντοτε την ποσότητα εξηγηθείσας διακύμανσης στην εξαρτώμενη μεταβλητή (εκφραζόμενη  $R^2$ ). Επομένως, η προσθήκη πάρα πολλών ανεξάρτητων μεταβλητών χωρίς οποιαδήποτε θεωρητική δικαιολόγηση μπορεί να οδηγήσει σε ένα overfit μοντέλο.

Στόχος μας είναι, σύμφωνα με τα παραπάνω να εφαρμόζουμε την μέθοδο της multiple linear regression στα δεδομένα της Σύρου και να εξάγουμε από αυτήν την πληροφορία που χρειαζόμαστε. Αρχικά, χωρίζουμε τα δεδομένα σε δύο ξεχωριστά datasets. Το ένα είναι το training dataset με το οποίο θα «εκπαιδεύσουμε» το μοντέλο μας και το δεύτερο data set είναι το test-set το οποίο θα χρησιμοποιήσουμε για να ελέγξουμε την αποτελεσματικότητα του μοντέλου που εκπαιδεύσαμε.

Στην σχέση (1) η εξαρτημένη τιμή που χρησιμοποιούμε είναι  $y$ , δηλαδή αυτή που θέλουμε να παρεμβάλουμε και η  $x_i$  ανεξάρτητες μεταβλητές που αντιστοιχούν στις υπόλοιπες παραμέτρους. Είναι όμως όλες οι παράμετροι σημαντικοί για να τους

συμπεριλάβουμε στο μοντέλο; Στην **Εικόνα 14** βλέπουμε μία γραφική απεικόνιση του πίνακα συσχέτισης (correlation matrix). Στον άνω τετραγωνικό πίνακα της εικόνας βλέπουμε ποιες παράμετροι είναι «σημαντικές» για την κάθε παράμετρο. Αυτό προκύπτει από το P\_value. Η P\_value μας βοηθάει να αποφασίσουμε ποιες παράμετροι είναι στατιστικά σημαντικές και παίρνει τιμές από 0 έως 1. Στην Εικόνα 14 αυτό εκφράζεται με τον αριθμό των αστερίσκων στον άνω τριγωνικό πίνακα. Όσο περισσότεροι οι αστερίσκοι, τόσο σημαντικές είναι οι παράμετροι για την εκάστοτε εξαρτημένη μεταβλητή.

Στο μοντέλο που θα δημιουργήσουμε στην συνέχεια, εξαρτημένη μεταβλητή είναι η παράμετρος vocaqi. Με βάση τα παραπάνω, δημιουργήθηκε η συνάρτηση



Εικόνα 14: Γραφική αναπαράσταση του πίνακα συσχέτισης (correlation matrix)

backward\_Elimination\_regr() με στόχο την επαναπροσδιορισμό του μοντέλου multiple linear regression. Σε κάθε επανάληψη γίνεται οπισθοδρόμηση και υπολογισμός των P\_values με στόχο τη διαγραφή των μη σημαντικών παραμέτρων για την εξαρτημένη μεταβλητή vocaqi.

Έτσι χρησιμοποιούμε τον παρακάτω κώδικα για να εκπαιδεύσουμε το μοντέλο.

```
> regressor = backward_Elimination_regr(training_set,0.05)
```

Μετά την δημιουργία του μοντέλου παλινδρόμησης είμαστε σε θέση να τρέξουμε τον παρακάτω κώδικα και να υπολογίσουμε τις προβλέψεις που γίνονται για την εξαρτώμενη μεταβλητή `vocaqi`.

```
> y_pred = predict(regressor, newdata = test_set)
```

Στο τέλος του κεφάλαιο 5 σχολιάζεται η αποδοτικότητα των προβλέψεων αυτών και απόδοση του μοντέλου multiple linear regression.

### 5.3 Support Vector Regression (SVR)

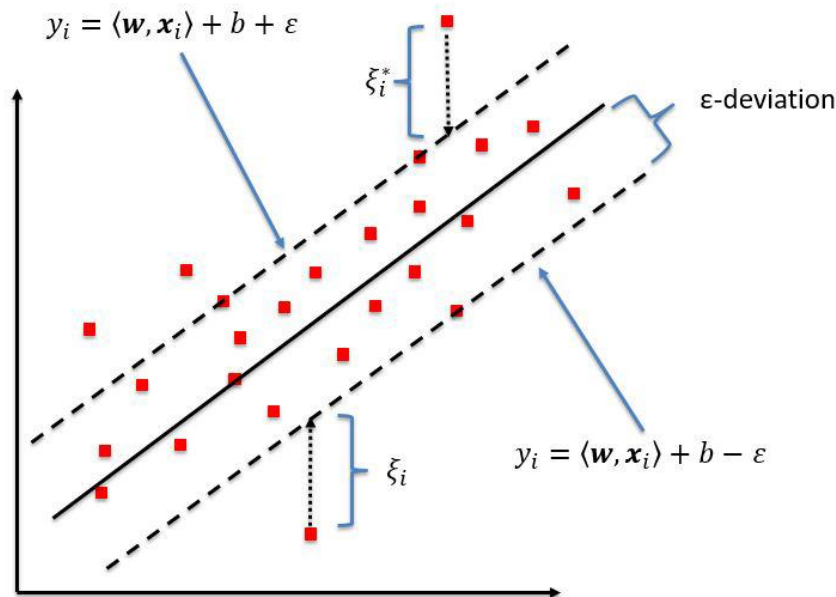
Όροι οι οποίοι θα χρησιμοποιηθούν σε αυτή την μέθοδο περιγράφονται παρακάτω.

1. Kernel: Είναι μία συνάρτηση που στόχο έχει την χαρτογράφηση των δεδομένων «χαμηλότερων» διαστάσεων σε δεδομένα «υψηλότερων» διαστάσεων.
2. Hyper Plane: Είναι η γραμμή διαχωρισμού μεταξύ των κατηγοριών δεδομένων. Στην μέθοδο που θα ακολουθήσουμε, την ορίζουμε ως τη γραμμή που θα μας βοηθήσει να προβλέψουμε τη συνεχή στοχευμένη τιμή.
3. Boundary line: Στην μέθοδο κατηγοριοποίησης SVM υπάρχουν δύο επιπλέον γραμμές εκτός του Hyper Plane που δημιουργούν ένα περιθώριο. Τα support vectors μπορούν να βρίσκονται σε αυτές τις οριακές γραμμές ή εκτός από αυτές. Οι οριακές γραμμές έχουν στόχο να διαχωρίσουν τις δύο κατηγορίες. Η μέθοδος SVR βασίζεται στην ίδια ιδέα.
4. Support Vectors: Τα διανύσματα προσδιορίζουν τα σημεία των δεδομένων που βρίσκονται πιο κοντά στο όριο. Η απόσταση των σημείων είναι η ελάχιστη-μικρότερη δυνατή.

Στόχος μας είναι να βρούμε αυτά τα boundary lines με απόσταση 'ε' από την hyper plane τέτοια ώστε τα σημεία δεδομένα να βρίσκονται πιο κοντά στη hyper plane ή τα support vectors να βρίσκονται εντός αυτής της οριακής γραμμής. Στην Εικόνα 15 δίνεται απεικόνιση των παραπάνω που δημιούργησαν οι Tania Kleynhans, Matthew Montanaro, Aaron Gerace and Christopher Kanan [26] με στόχο την πρόβλεψη των κύριων θερμικών ακτινοβολιών της ατμόσφαιρας.

Στην R η βιβλιοθήκη για την μέθοδο SVR είναι η «`e1071`». Σε αντίθεση με την μέθοδο του κεφαλαίου 5.2, εδώ όλες οι παράμετροι χρησιμοποιούνται για την εκπαίδευση του μοντέλου.

```
> regressor_svr = svm(formula = vocaqi~.,
```



Εικόνα 15: [https://www.researchgate.net/figure/Schematic-of-the-one-dimensional-support-vector-regression-SVR-model-Only-the-points\\_fig5\\_320916953](https://www.researchgate.net/figure/Schematic-of-the-one-dimensional-support-vector-regression-SVR-model-Only-the-points_fig5_320916953) [26]

```
+ data = epoch(sensor1,"winter"),
+ type = "eps-regression")
```

Τέλος, προσεγγίζουμε τα δεδομένα του test-set με στόχο να δούμε τα αποτελέσματα του μοντέλου.

```
> y_pred_svr = predict(regressor_svr,test_set)
```

Στο τέλος του κεφαλαίου σχολιάζεται η αποδοτικότητα των προβλέψεων αυτών και απόδοση του μοντέλου SVR.

#### 5.4 Regression Tree

Τα δέντρα παλινδρόμησης χρησιμοποιούνται όταν η εξαρτημένη μεταβλητή είναι συνεχής. Το δέντρο χωρίζει τον χώρο πρόβλεψης σε ξεχωριστές και μη επικαλυπτόμενες περιοχές. Ακολουθεί μία προσέγγιση top-down, δηλαδή ξεκινάει από την κορυφή του δέντρου όταν όλες οι παρατηρήσεις είναι διαθέσιμες σε μια περιοχή και διαδοχικά διαχωρίζουν το εύρος πρόβλεψης σε κλάδους κάτω από το δέντρο. Πρόκειται για μία άπληστη προσέγγιση επειδή ο αλγόριθμος αναζητά τη βέλτιστη διαθέσιμη μεταβλητή στον παρόντα διαχωρισμό, χωρίς να προβλέπει για μελλοντικούς διαχωρισμούς που πιθανώς να οδηγήσουν σε ένα καλύτερο δέντρο. Ο αλγόριθμος σταματάει με έναν κριτήριο το οποίο το ορίζει ο προγραμματιστής.

Όπως και στα προηγούμενα μοντέλα, έτσι και στην παρούσα μέθοδο η εξαρτημένη μεταβλητή που ψάχνουμε είναι η voca*q*i. Στην περίπτωση των regression trees στην R

χρησιμοποιούμε την βιβλιοθήκη `rpart` για να εκπαιδεύσουμε το μοντέλο μας και να αναπτύξουμε το δέντρο. Για την εκπαίδευση του μοντέλου χρησιμοποιούμε το `training_set` που χρησιμοποιήσαμε στο κεφάλαιο 5.1 και θα χρησιμοποιήσουμε τις ανεξάρτητες μεταβλητές του συνόλου δεδομένων `test_set` για να προβλέψουμε την τιμή `vocaqi`.

```
> regressor_tree = rpart(formula = vocaqi~., data = training_set,  
+ control = rpart.control(cp = 0,maxdepth = 4))
```

Αφού εκπαιδεύσαμε το μοντέλο `regressor_tree`, στην συνέχεια είμαστε σε θέση να χρησιμοποιήσουμε το `test_set` για να προβλέψουμε τις τιμές `vocaqi`. Στον παραπάνω κώδικα βλέπουμε την συνάρτηση `rpart.control(cp,maxdepth)`. `Cp` είναι η παράμετρος πολυπλοκότητας. Στόχος αυτού είναι η εξοικονόμηση χρόνου υπολογισμού στην περικοπή των χωρισμάτων. Με την παράμετρο `maxdepth` ορίζεται το μέγιστο βάθος οποιουδήποτε κόμβου του τελικού δέντρου. Ο πρώτος κόμβος έχει βάρος μηδέν. Στο συγκεκριμένο κώδικα οδηγούμε τον αλγόριθμο να αναπτύξει δέντρο με μέγιστο βάθος 4.

```
> y_pred_tree = predict(regressor_tree, test_set)
```

Με τον παραπάνω κώδικα τα αποτελέσματα του μοντέλου `regression tree`. Στο τέλος του κεφάλαιο 5 σχολιάζεται η αποδοτικότητα των προβλέψεων αυτών, απόδοση του μοντέλου και συγκρίνεται με την απόδοση των υπολοίπων μεθόδων.

## 5.5 Random Forest

Η μέθοδος `Random Forest` είναι μία πολύπλευρη μέθοδος μηχανικής μάθησης ικανή να εφαρμόσει τεχνικές τόσο παλινδρόμησης όσο και κατηγοριοποίησης. Στην `Random Forest` αναπτύσσονται πολλαπλά δέντρα, όπως αυτά του προηγούμενου κεφαλαίου. Αναπτύσσονται λοιπόν πολλαπλά `CART` (`Classification and Regression Trees`) μοντέλα με διαφορετικό δείγμα και διαφορετικές αρχικές μεταβλητές. Για παράδειγμα, θα λάβει ένα

τυχαίο δείγμα 200 παρατηρήσεων και 5 από τις 11 μεταβλητές για να δημιουργήσει ένα μοντέλο CART. Ο αριθμός των επαναλήψεων είναι ο αριθμός των δέντρων που θα αναπτυχθούν και ορίζονται στην συνάρτηση που θα χρησιμοποιήσουμε. Η τελική πρόβλεψη είναι συνάρτηση κάθε πρόβλεψης. Τέτοια, μπορεί να είναι ο μέσος όρος κάθε πρόβλεψης.

Για την δημιουργία του μοντέλου αυτού, χρησιμοποιούμε την βιβλιοθήκη “randomForst” της R. Παρακάτω, εκπαιδεύουμε το μοντέλο regressor\_rForest με το training\_set συμπεριλαμβανομένων όλων των εξαρτώμενων μεταβλητών. Ο αριθμός που ορίσαμε για τον αριθμό των δέντρων είναι 100 και επιλέχθηκε τυχαία.

```
> regressor_rForest = randomForest(vocabulary~.,  
+                                 data = training_set,  
+                                 ntree = 100)
```

Παρακάτω η y\_pred\_rForest είναι οι προβλέψεις οι οποίες έχουν εξαχθεί από το μοντέλο regressor\_rForest.

```
> y_pred_rForest = predict(regressor_rForest, test_set)
```

Στο κεφάλαιο 5.6 ελέγχεται η αποδοτικότητα των προβλέψεων αυτών, απόδοση του μοντέλου και συγκρίνεται με την απόδοση των υπολοίπων μεθόδων.

## 5.6 Σχολιασμός και σύγκριση μοντέλων

Στα προηγούμενα κεφάλαια έγινε η «εκπαίδευση» των μοντέλων με στόχο την πρόβλεψη των τιμών vocabulary από το σύνολο δεδομένων test\_set. Στο παρόν κεφάλαιο σχολιάζονται και συγκρίνονται οι προβλέψεις των μοντέλων μεταξύ τους. Για την αξιολόγηση αυτή θα χρησιμοποιηθούν δύο δείκτες απόδοσης, στο πνεύμα της αξιολόγησης των μοντέλων όπως προτείνουν οι Li, X., Peng, L., Hu, Y., Shao, J., & Chi, T. [27] για την απόδοση των δικών τους μοντέλων. Αυτές οι δύο είναι οι root-mean-square error (RMSE) και η mean absolute

error (MAE). Επιπλέον υπολογίζεται και ερμηνεύεται ο δείκτης R-squared. Αυτοί οι δείκτες υπολογίζονται ως εξής:

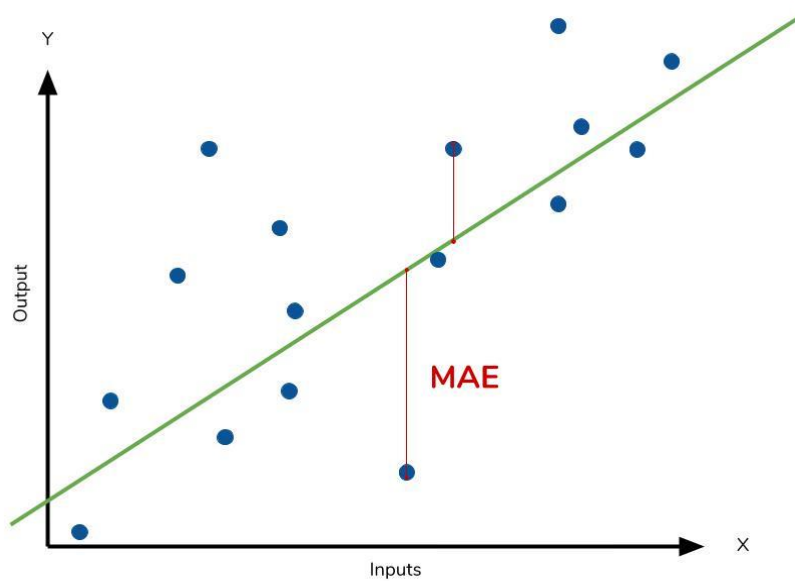
$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - P_i)^2}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |O_i - P_i|$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O}_i)^2}$$

Όπου:

- $O_i$  είναι οι παρατηρούμενες τιμές της εξαρτημένης μεταβλητής (εδώ test\_set που αντιστοιχούν στην παράμετρο voca*q*i)
- $\bar{O}_i$  είναι η μέση τιμή των παρατηρούμενων εξαρτημένων μεταβλητών.
- $P_i$  είναι οι προβλέψεις που έχουν προκύψει από την εκπαίδευση των μοντέλων για την εξαρτώμενη μεταβλητή.
- $N$  ο αριθμός των δειγμάτων αξιολόγησης.



Εικόνα 16: Παράδειγμα όπου απεικονίζεται το Mean Absolute Error. Από κάθε data point υπολογίζεται η απόσταση από την ευθεία γραμμή σε απόλυτη τιμή. Αθροίζονται όλες αυτές οι αποστάσεις και διαιρούνται με τον αριθμό των data points

Στον **πίνακα 3** φαίνονται τα αποτελέσματα των δεικτών για κάθε μέθοδο μηχανικής μάθησης που εκπαιδεύσαμε στο κεφάλαιο 5.

Αξία έχει ωστόσο να ερμηνεύσουμε και να καταλάβουμε τους αριθμούς που προκύπτουν από τους παραπάνω δείκτες.

| Method                     | RMSE     | MAE      | $R^2$     |
|----------------------------|----------|----------|-----------|
| Multiple Linear Regression | 120.2258 | 94.35186 | 0.2203581 |
| Support Vector Regressor   | 95.54546 | 65.58721 | 0.5075978 |
| CART (Regression Tree)     | 115.0214 | 89.2997  | 0.2863964 |
| Random Forest              | 67.3594  | 48.25294 | 0.7552617 |

Πίνακας 3: Δείκτες απόδοσης των μοντέλων που εκπαιδεύτηκαν από τα δεδομένα του αισθητήρα 8200015F την εποχή του χειμώνα

Ξεκινώντας με τον δείκτη RMSE βλέπουμε το σφάλμα της πρόβλεψης κατά μέσο όρο. Συγκεκριμένα μας λέει πόσο συγκεντρωμένα τα δεδομένα είναι γύρω από τη γραμμή που ταιριάζει τα δεδομένα. Συνήθως, η μονάδα μέτρησης είναι η αντίστοιχη με αυτή της παρατηρούμενης μεταβλητής (εδώ μονάδες AQI).

Το MAE είναι ο πιο απλός δείκτης σφάλματος παλινδρόμησης για κατανόηση. Υπολογίζουμε την απόσταση για κάθε σημείο δεδομένων (data point), λαμβάνοντας μόνο την απόλυτη τιμή του καθενός έτσι ώστε οι αρνητικές και θετικές αποστάσεις να μην ακυρώνονται. Στη συνέχεια λαμβάνουμε τον μέσο όρο όλων αυτών των αποστάσεων που υπολογίσαμε. Το αποτέλεσμα φαίνεται στην **Εικόνα 16**. Στην πράξη, ο MAE περιγράφει το τυπικό μέγεθος των αποστάσεων από τα σημεία δεδομένων.

Τέλος, ο δείκτης R-squared είναι ένας αριθμός ο οποίος παίρνει τιμές από 0-1 ή 0% και 100%. Πρόκειται για μία μέτρηση η οποία μας λέει πόσο ικανοποιητικά το μοντέλο προσεγγίζει τα σημεία.

Σύμφωνα με τα παραπάνω και τον **πίνακα 3** παρατηρούμε ότι το μοντέλο Random Forest προσεγγίζει καλύτερα τα πραγματικά δεδομένα σε έναν πολύ καλό βαθμό ικανό να μην τα ταιριάζει απόλυτα (overfitting). Αυτό δεν σημαίνει ότι τα υπόλοιπα μοντέλα δεν παρέχουν σημαντική πληροφορία.

Στη μέθοδο multiple linear regression βλέπουμε υψηλές τιμές RMSE, MAE. Αυτές οι τιμές είναι αρκετά μεγάλες και μπορεί να απέχουν μακριά από την πραγματικότητα. Για παράδειγμα, μία παρατήρηση να έχει τιμή στις «υγιεινά» επίπεδα AQI, ωστόσο το μοντέλο να προβλέψει για ανθυγιεινές τιμές. Το μικρό R-squared μας επιβεβαιώνει ότι δεν πρόκειται για ικανοποιητικό μοντέλο ή μας ενημερώνει ότι το μοντέλο εξαρτάται από επιπλέον παραμέτρους.

Όμοια, το Regression Tree δεν φαίνεται να έχει καλή απόδοση. Κάνοντας περισσότερους ελέγχους τόσο στο μέγιστο βάθος του δέντρου όσο και με την παράμετρο πολυπλοκότητας, πιθανόν το μοντέλο να βελτιστοποιηθεί και να ταιριάζει περισσότερο στα δεδομένα.

Η μέθοδος SVR παρουσιάζει ικανοποιητική απόδοση ταιριάζοντας τα δεδομένα ικανά να μας δώσουν αποτελέσματα που έχουν αξία. Η μέθοδος SVR επιδέχεται βελτιστοποιήσεις

εξετάζοντας και επαναπροσδιορίζοντας τόσο την  $eps$  απόστασή μεταξύ hyperplane και boundary line όσο και την cost function.

Σύμφωνα με τα παραπάνω, είμαστε σε θέση να δημιουργήσουμε αξιόπιστα μοντέλα πρόβλεψης της ποιότητας του αέρα τροποποιώντας ή εκπαιδεύοντας τα μοντέλα με περισσότερα δεδομένα από αυτά που έχουμε μέχρι στιγμής.

## **ΚΕΦΑΛΑΙΟ 6**

### **6.1 Συμπεράσματα**

Στην παρούσα εργασία έγινε ο προσδιορισμός του προβλήματος που ταλανίζει την εποχή μας, αυτό της ποιότητας του αέρα εξετάζεται η μελέτη περίπτωσης στο νησί της Σύρου. Τα δεδομένα που συλλέγονται στην Σύρο εξετάζονται ανά εποχή και στην συνέχεια ελέγχονται και τα δεδομένα ανά ζώνη ώρας σε κάθε εποχή. Με αυτό τον τρόπο, φαίνεται καθαρά το πρόβλημα της ρύπανσης του αέρα που επικράτησε στην Σύρο. Υπάρχει ανάγκη για περαιτέρω έρευνα στο πρόβλημα με σκοπό την λήψη αποφάσεων και εύρεση δράσεων με σκοπό την μείωση της απειλής αυτής τόσο για την ανθρώπινη υγεία όσο και για το περιβάλλον.

Με τα δεδομένα που συλλέχτηκαν, εκπαιδεύτηκαν μοντέλα μηχανικής μάθησης που στόχο έχουν την πρόβλεψη της παραμέτρου που χαρακτηρίζει την ποιότητα του αέρα AQI. Τα μοντέλα αυτά δεν είχαν στόχο να είναι βελτιστοποιημένα και να μας παρέχουν τα καλύτερα δυνατά αποτελέσματα για τα δεδομένα του χειμώνα στην περιοχή Κάτω Μάννα. Στόχος, στην παρούσα εργασία, είναι να δείξουν ότι είναι ικανά να προσδιορίσουν ικανοποιητικά τα δεδομένα και να κάνουν προβλέψεις με στόχο την λήψη αποφάσεων.

### **6.2 Αδυναμίες**

Η συλλογή μεγάλου όγκου δεδομένων για την ανάπτυξη και εκπαίδευση μοντέλων μηχανικής μάθησης είναι απαραίτητη για να προσδιοριστούν ικανοποιητικά οι προβλέψεις που θα εξαχθούν. Η παρούσα εργασία πραγματοποιείται χωρίς να έχει συμπληρωθεί ένας χρόνος συλλογής καθαρών δεδομένων. Συγκεκριμένα υπάρχει έλλειψη σε δεδομένα από μήνες του φθινοπώρου και του καλοκαιριού. Με την συλλογή περισσότερων δεδομένων, θα φαίνεται ότι τα δεδομένα θα παρουσιάζουν όλο και περισσότερο πιο κανονική κατανομή. Αυτό θα αποτελέσει αρωγός στην προσπάθεια για τη δημιουργία αποτελεσματικών μοντέλων μηχανικής μάθησης.

Στα παραπάνω αξίζει να προσθέσουμε και την έλλειψη μετεωρολογικών δεδομένων. Δεδομένα όπως η ένταση (beaufort) και η ταχύτητα του αέρα και άλλα μετεωρολογικά δεδομένα είναι ικανά να επηρεάσουν την ποιότητα του αέρα και να αποτελέσουν σημαντικοί παράμετροι στην ανάπτυξη μοντέλων ποιότητας αέρα με σκοπό την πρόβλεψη.

Επιπλέον αδυναμία που υπάρχει στην παρούσα εργασία είναι η ύπαρξη πολλών ακραίων σημείων (outliers) στις περισσότερες παραμέτρους. Αυτό το πρόβλημα μπορεί να συμβάλει στην αλλοίωση της απόδοσης των μοντέλων πρόβλεψης. Μία μέθοδος που μπορεί να χρησιμοποιηθεί είναι η διαγραφή των παρατηρήσεων αυτών, ωστόσο αυτό θα μείωνε πολύ το σύνολο των δεδομένων κάτι που δεν επιθυμούμε έχοντας ήδη λίγα δεδομένα για να χρησιμοποιήσουμε. Μία ακόμα μέθοδος είναι η αντικατάσταση των ακραίων τιμών με την μέση τιμή των παρατηρήσεων. Αυτή η μέθοδος πιθανότατα θα συνέκλινε αρκετά τα δεδομένα προς την μέση τιμή και ο όγκος των δεδομένων δεν μας επιτρέπει την χρήση της μεθόδου αυτής.

### **6.3 Προτάσεις Για Περαιτέρω Έρευνα**

Στο κεφάλαιο 4 αναφέραμε ότι η ανθρώπινη δραστηριότητα παίζει σημαντικό ρόλο στον καθορισμό της ποιότητας του αέρα. Υπάρχει η ανάγκη να παρθούν αποφάσεις για βελτίωση της ποιότητας του αέρα με στόχο τόσο την υγεία του ανθρώπου όσο και του περιβάλλοντος. Στο κεφάλαιο 5 αναπτύχθηκαν μοντέλα μηχανικής μάθησης με στόχο την πρόβλεψη του δείκτη AQI. Προσθέτοντας επιπλέον σημαντικά δεδομένα όπως είναι τα μετεωρολογικά δεδομένα τα μοντέλα μηχανικής μάθησης θα παρέχουν ακόμα πιο καλά αποτελέσματα και αξιοπιστία, καθώς πρόκειται για δεδομένα που στις μέρες μας παρέχονται ικανοποιητικά από μετεωρολογικές υπηρεσίες.

Για μελλοντική έρευνα προτείνεται να δημιουργηθούν μοντέλα κατηγοριοποίησης μηχανικής μάθησης (Classification) που στόχο έχουν να κατηγοριοποιήσουν νέες παρατηρήσεις. Πιο συγκεκριμένα, μπορεί να χρησιμοποιηθεί ο **πίνακας 1** και οι νέες παρατηρήσεις να κατηγοριοποιηθούν με βάση αυτόν. Να γίνεται γνωστό δηλαδή ότι μία νέα παρατήρηση ανήκει στην κατηγορία ποιότητας αέρα «καλή», «μέτρια», «ανθυγιεινή», «βλαβερή» κ.α.

Εντούτοις, τόσο μέθοδοι πρόβλεψης όσο και κατηγοριοποίησης αποτελούν εργαλεία για την αντιμετώπιση του προβλήματος της ρύπανσης της ατμόσφαιρας. Υπάρχει ανάγκη για περεταίρω έρευνα που θα μπορούσε να συρρικνώσει την απειλή αυτή.

Ας φανταστούμε το νησί της Σύρου ως ένα σύστημα το οποίο έχει τα δικά της πρότυπα ανθρώπινης δραστηριότητας που επηρεάζουν στην ποιότητα του αέρα. Αυτή η δραστηριότητα συνεπάγεται με την κινητικότητα των μηχανοκίνητων οχημάτων, με τη

βιομηχανική δραστηριότητα, με τη καύση πετρελαίου θέρμανσης και την προσέλευση πλοίων στο λιμάνι. Κάθε μία από τις παραπάνω δραστηριότητες-πρότυπα θα μπορούσε να παρατηρηθεί και δειχθεί σε τι ποσοστό επηρεάζει την ποιότητα του αέρα του συστήματος. Με αυτόν τον τρόπο θα μπορεί να γίνει μία πρόβλεψη, με την χρήση των μοντέλων μηχανικής μάθησης, της ποιότητας του αέρα δείχνοντας για παράδειγμα το όριο των πλοίων που μπορούν να διαμένουν ενεργά στο λιμάνι της Σύρου χωρίς αυτά να επηρεάζουν αρνητικά την ποιότητα του αέρα της Σύρου. Έτσι προκύπτει ένα σύστημα που μπορεί να θέτει όρια και να παρέχει πληροφορία στην λήψη των αποφάσεων από την τοπική αυτοδιοίκηση.

Σε επέκταση του παραπάνω συστήματος, θα μπορούσαν να δημιουργηθούν servers και να τοποθετηθούν σε κάθε νησί των Κυκλάδων. Αυτοί οι servers θα αναλύουν τα δεδομένα ποιότητας του αέρα και θα επικοινωνούν με τους υπολοίπους. Με χρήση Hadoop μπορεί να γίνει κατανεμημένη ανάλυση αυτών των δεδομένων και να εξαχθεί επιπλέον πληροφορία που απαντάει στο ερώτημα “πόσο επηρεάζει η δραστηριότητα ενός νησιού, την ποιότητα αέρα ενός γειτονικού νησιού”.

Για την αντιμετώπιση της απειλής της ρύπανσης του αέρα είναι ανάγκη να γίνουν δράσεις που στόχο έχουν την προστασία της κοινωνίας και του περιβάλλοντος. Στην συνθήκη του Παρισιού για την αντιμετώπιση της κλιματικής αλλαγής έχει προϋπολογιστεί ότι το κόστος θα είναι το 1% του παγκόσμιου ΑΕΠ. Για την αντιμετώπιση του προβλήματος, θα πρέπει να υπάρχει κατάλληλη ενημέρωση της κοινωνίας προωθώντας μία φιλοσοφία προστασίας του περιβάλλοντος. Παράλληλα, οι επιχειρήσεις και οι βιομηχανίες θα πρέπει να προσαρμόστουν στις οδηγίες που θα δίνει η τοπική αυτοδιοίκηση με στόχο την μείωση των ρύπων που απελευθερώνουν με την δραστηριότητά τους. Η πειθαρχία των ανθρώπων σε υγιείς και δίκαιους κανόνες είναι το κλειδί για την επίλυση ή την συρρίκνωση κάθε προβλήματος και απειλής.

## BIBΛΙΟΓΡΑΦΙΑ

- [1] World Health Organization, 2018, *How air pollution is destroying our health*, world health organization, 17/6/2019, <<https://www.who.int/air-pollution/news-and-events/how-air-pollution-is-destroying-our-health>>
- [2] World Health Organization, 2018, *Estimated age-standardized incidence rates (World) in 2018, worldwide, both sexes, all ages*, 17/6/2019, world health organization, <<http://gco.iarc.fr/today/online-analysis-multi-bars>>
- [3] World Health Organization, 2018, *Health benefits far outweigh the costs of meeting climate change goals*, world health organization, 17/6/2019, <<https://www.who.int/news-room/detail/05-12-2018-health-benefits-far-outweigh-the-costs-of-meeting-climate-change-goals>>
- [4] European Environment Agency, 2017, *Validated monitoring data and air quality maps*, European Environment Agency, 17/6/2019, <<http://www.eea.europa.eu/themes/air/airbase>>
- [5] European Commission, Brussels, 9.6.2017, *report from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions*
- [6] World Health Organization, May 2018. *Ambient Air Quality and Health*, 17/6/2019, world health organization, <<http://www.who.int/mediacentre/factsheets/fs313/en/>>
- [7] World Health Organization, *10 ways you can fight air pollution*, 17/6/2019, world health organization, <<https://www.who.int/air-pollution/news-and-events/how-air-pollution-is-destroying-our-health/10-ways-you-can-fight-air-pollution>>
- [8] Environmental Protection Agency, 18 June 2019, *Air Quality Index (AQI) Basics*, AirNow, 19/6/2019, <<https://www.airnow.gov/index.cfm?action=aqibasics.aqi#unh>>
- [9] Environmental Protection Agency, February 2014, *AIR QUALITY INDEX A Guide to Air Quality and your Health*, State of North Carolina

- [10] Environmental Protection Agency, 14/11/2018, *Particulate Matter (PM) Basics*, United States Environmental Protection Agency, 19/6/2019, <<https://www.epa.gov/pm-pollution/particulate-matter-pm-basics#PM>>
- [11] San Martini, F. M., Hasenkopf, C. A., & Roberts, D. C. (2015). *Statistical analysis of PM<sub>2.5</sub> observations from diplomatic facilities in China*. *Atmospheric Environment*, 110, 174–185. doi:10.1016/j.atmosenv.2015.03.060
- [12] World Health Organization Working Group, Bonn, Germany, 13–15 January 2003, *Health Aspects of Air Pollution with Particulate Matter, Ozone and Nitrogen Dioxide*, Report on a WHO Working Group
- [13] World Health Organization, 2006, *WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide*, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland
- [14] Khan, F. I., & Kr. Ghoshal, A. (2000). *Removal of Volatile Organic Compounds from polluted air*. *Journal of Loss Prevention in the Process Industries*, 13(6), 527–545. doi:10.1016/s0950-4230(00)00007-3
- [15] Environmental Protection Agency, 2017, *Technical Overview of Volatile Organic Compounds*, Environmental Protection Agency, 25/6/2019, <<https://www.epa.gov/indoor-air-quality-iaq/technical-overview-volatile-organic-compounds#definition>>
- [16] World Health Organization, 1987, *Air Quality Guidelines for Europe*, 2<sup>nd</sup> Edition, WHO Regional Office for Europe, Copenhagen
- [17] World Health Organization, 2010, *WHO Guidelines for Indoor Air Quality: Selected Pollutants*, WHO Regional Office for Europe, Copenhagen.
- [18] Safety and Health Group, *Carbon Dioxide Health Hazard Information Sheet*, United States Department of Agriculture, 25/6/2019, <<https://www.fsis.usda.gov/wps/wcm/connect/bf97edac-77be-4442-aea4-9d2615f376e0/Carbon-Dioxide.pdf?MOD=AJPERES>>

- [19] Kalkstein, L. S., & Corrigan, P. (1986). *A Synoptic Climatological Approach For Geographical Analysis: Assessment of Sulfur Dioxide Concentrations*. *Annals of the Association of American Geographers*, 76(3), 381–395. doi:10.1111/j.1467-8306.1986.tb00126.x
- [20] Akbari, H. (2002). *Shade trees reduce building energy use and CO2 emissions from power plants*. *Environmental pollution*, 116, S119-S126.
- [21] DeGaetano, A. T., & Doherty, O. M. (2004). *Temporal, spatial and meteorological variations in hourly PM2.5 concentration extremes in New York City*. *Atmospheric Environment*, 38(11), 1547-1558.
- [22] Natsagdorj, L., Jugder, D., & Chung, Y. S. (2003). *Analysis of dust storms observed in Mongolia during 1937–1999*. *Atmospheric Environment*, 37(9-10), 1401-1411.
- [24] Constantine Hatziaioniu, *Prospecting the Data from One of the Air-Quality Monitor Sites on Syros*, 15/08/2019, <<https://env.kernelit.gr/category/analysis/>>
- [25] Jiang, S., Ferreira, J., & González, M. C. (2012). *Clustering daily patterns of human activities in the city*. *Data Mining and Knowledge Discovery*, 25(3), 478–510. doi:10.1007/s10618-012-0264-z
- [26] Kleynhans, T., Montanaro, M., Gerace, A., & Kanan, C. (2017). *Predicting Top-of-Atmosphere Thermal Radiance Using MERRA-2 Atmospheric Data with Deep Learning*. *Remote Sensing*, 9(11), 1133. doi:10.3390/rs9111133
- [27] Li, X., Peng, L., Hu, Y., Shao, J., & Chi, T. (2016). *Deep learning architecture for air quality predictions*. *Environmental Science and Pollution Research*, 23(22), 22408–22417. doi:10.1007/s11356-016-7812-9
- [28] Zhu, D., Cai, C., Yang, T., & Zhou, X. (2018). *A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization*. *Big Data and Cognitive Computing*, 2(1), 5. doi:10.3390/bdcc2010005