



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ: ΠΛΗΡΟΦΟΡΙΑΚΑ
ΣΥΣΤΗΜΑΤΑ ΚΑΙ ΥΠΗΡΕΣΙΕΣ

ΕΙΔΙΚΕΥΣΗ: ΜΕΓΑΛΑ ΔΕΔΟΜΕΝΑ ΚΑΙ ΑΝΑΛΥΤΙΚΗ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Optimization of logistics through analytics outcomes»

Παναγιωτακόπουλος Γεώργιος

A.M.: ME2030

Επιβλέπων καθηγητής

Κυριαζής Δημοσθένης

ΠΕΙΡΑΙΑΣ, ΜΑΡΤΙΟΣ 2022

ΠΕΡΙΛΗΨΗ

Μεταφορικό Ισοδύναμο είναι το μέτρο, με τη θέσπιση του οποίου επιδιώκεται η εξίσωση του κόστους μεταφοράς με μέσα μαζικής μεταφοράς από την ηπειρωτική Ελλάδα σε νησί ή από νησί σε νησί με το κόστος που θα ίσχυε στην ηπειρωτική Ελλάδα για την ίδια απόσταση.

Οι νησιωτικές κοινωνίες αντιμετωπίζουν επείγοντα ζητήματα ερημοποίησης και μακροχρόνια ζητήματα απομόνωσης από τις κεντρικές αγορές προϊόντων στις οποίες αναπτύσσονται οι οικονομίες κλίμακας. Ως αποτέλεσμα αυτών είναι αναγκασμένες να επιβαρύνονται με το επιπλέον κόστος θαλάσσιας μεταφοράς ανθρώπων και προϊόντων στα νησιά σε σχέση με τη χερσαία μεταφορά στην ενδοχώρα. Το επιπλέον αυτό κόστος είναι από τις κύριες αιτίες ενδοπεριφερειακών ανισοτήτων, μείωσης του νησιωτικού πληθυσμού και αποτροπής ίδρυσης και παραμονής των επιχειρήσεων στα νησιά. Η εφαρμογή του μεταφορικού ισοδύναμου εναρμονίζει αυτό το κόστος μετακίνησης επιβατών και εμπορευμάτων, από και προς τα νησιά. Η εφαρμογή του μεταφορικού ισοδύναμου εναρμονίζει το κόστος μεταφοράς εμπορευμάτων, από και προς τα νησιά, σε σχέση με το κόστος των χερσαίων μεταφορών. Το κόστος μετακίνησης προϊόντων επιδοτείται από πόρους του Προγράμματος Δημοσίων Επενδύσεων με κανόνες συμβατούς για τις κρατικές ενισχύσεις στην Ε.Ε., σε βαθμό που εναρμονίζει τη μετακίνηση τους στη θάλασσα με το κόστος στη στεριά. Σκοπός είναι η εξασφάλιση της ίσης μεταχείρισης προς τις νησιωτικές κοινότητες αναφορικά με το μεταφορικό κόστος που καλούνται να πληρώνουν για τις μεταφορές τους, το οποίο θα πρέπει να είναι ανάλογο με το τίμημα που καταβάλουν αντίστοιχου μεγέθους κοινότητες της ηπειρωτικής χώρας. Συνεπώς στοχεύει στην ελαχιστοποίηση του οικονομικού μειονεκτήματος που επιβαρύνει τις απομακρυσμένες νησιωτικές κοινότητες έτσι ώστε με τη σειρά τους να συμβάλλουν περισσότερο στην οικονομική ανάπτυξη της χώρας.

Λέξεις – κλειδιά: μεταφορικό ισοδύναμο, ανάλυση εφοδιαστικής αλυσίδας, νησιωτικό εμπόριο, οικονομική δραστηριότητα νησιών, ανάλυση επιδοτήσεων μεταφορικών εξόδων

ABSTRACT

Transport Equivalent is the measure, where its implementation seeks to equalize the cost of transport by public transport from mainland Greece to island or from island to island with the cost that would apply in mainland Greece for the same distance.

Island societies face urgent issues of desertification and long-term isolation from the central product markets in which economies of scale are developing. As a result, they are forced to bear the additional cost of transporting people and goods to the islands in relation to inland transport. This additional cost is one of the main causes of intra-regional disparities, reduction of the island population and prevention of the establishment and stay of businesses on the islands. The implementation of the transport equivalent harmonizes this cost of transporting passengers and goods to and from the islands. The application of the transport equivalent harmonizes the cost of transporting goods, to and from the islands, in relation to the cost of land transport. The cost of transporting products is subsidized by resources of the Public Investment Program with rules compatible with state aid in the EU, to the extent that it harmonizes their movement at sea with the cost on land. The aim is to ensure equal treatment of island communities in terms of the transport costs they are required to pay for their transport, which should be commensurate with the price paid by similarly sized communities on the mainland. It therefore aims to minimize the economic disadvantage that burdens remote island communities so that in turn they contribute more to the country's economic development.

Keywords: transport equivalent, supply chain analysis, island trade, island economic activity, analysis of transport cost subsidies

ΕΥΧΑΡΙΣΤΙΕΣ

Με την ολοκλήρωση της μεταπτυχιακής διπλωματικής μου εργασίας, νιώθω την ανάγκη να εκφράσω τις θερμές μου ευχαριστίες σε όλους όσους συνέβαλλαν στην εκπόνησή της.

Θα ήθελα να ευχαριστήσω τον Επιβλέπων της εργασίας μου, κ. Κυριαζή Δημοσθένη, Καθηγητή του Τμήματος Ψηφιακών Συστημάτων του Πανεπιστημίου Πειραιώς, για την ανάθεση του θέματος και τη συνεχή βοήθειά που μου προσέφερε, για τις πολύτιμες συμβουλές του, για τη συνεργασία και το ενδιαφέρον που έδειξε από την αρχή της εργασίας έως τη περάτωσή της.

Τέλος, θα ήθελα εκφράσω την ευγνωμοσύνη μου και να ευχαριστήσω την οικογένειά μου για όλη τη στήριξη, την συμπαράσταση και την κατανόηση, καθ' όλη τη διάρκεια της διπλωματικής μου εργασίας αλλά και των μεταπτυχιακών μου σπουδών.

Παναγιωτακόπουλος Γεώργιος

Πειραιάς, Μάρτιος 2022

Πίνακας Περιεχομένων

1. Εισαγωγή.....	1
2. Μεθοδολογία, Δεδομένα και Στοιχεία ανάλυσης	1
Μεθοδολογία.....	2
Κύριες φάσεις στην ανάλυση δεδομένων	3
Ανάλυση δεδομένων με χρήση Python	4
Δεδομένα προς ανάλυση.....	4
Ερευνητικά ερωτήματα	7
3. Σχετικές εργασίες.....	9
Αναγκαιότητα ανάλυσης των δεδομένων του Μεταφορικού Ισοδυνάμου.....	9
Τεχνικές / Μέθοδοι Ανάλυσης Δεδομένων	10
Ποσοτικές μεταβλητές.....	10
Συσχέτιση.....	11
Γραμμική παλινδρόμηση	12
Εργαλεία ανάλυσης δεδομένων.....	13
4. Αποτελέσματα μελέτης.....	13
(A) Επιχειρήσεις & οικονομική δραστηριότητα ανά νησί	13
A1. Ποια είναι η κυρίαρχη Νομική Μορφή επιχείρησης ανά ΚΑΔ επιχειρήσεων και ανά νησί;	13
A2. Ποιος είναι ο μέσος όρος ηλικίας επιχείρησης ανά ΚΑΔ επιχειρήσεων και ανά νησί;	14
A3. Ποιος είναι ο μέσος κύκλος εργασιών ανά ΚΑΔ επιχειρήσεων και ανά νησί;.....	15
A4. Ποια είναι η συνολική χρηματοδότηση ανά ΚΑΔ επιχειρήσεων και ανά νησί;.....	15
A5. Ποια είναι η μέση χρονική διάρκεια μεταξύ υποβολής αίτησης και Εγκριτικής Απόφασης, ανά ΚΑΔ και ανά νησί;	20
A6. Ποιο είναι το συνολικό Επιλέξιμο Ποσό ανά εβδομάδα του έτους ανά νησί;	21
A7. Ποια είναι η ποσοστιαία σχέση επιλέξιμου/αιτούμενου ποσού ανά ΚΑΔ και ανά νησί για κάθε εβδομάδα του έτους;.....	24
A8. Σε ποια νησιά παρατηρούνται χωρικές ομαδοποιήσεις συγγενικών ΚΑΔ επιχειρήσεων;	24
A9. Ποια είναι η συνύπαρξη ΚΑΔ ανά νησί	25
(B) Ερωτήματα για Φορτωτικές.....	29
B1. Ποιες είναι οι κυρίαρχες περιοχές που στέλνουν φορτωτικές σε κάθε νησί, ανά ΚΑΔ και ανά συνολική αξία;	29
B2. Ποιες είναι οι κυρίαρχες περιοχές που λαμβάνουν φορτωτικές από κάθε νησί, ανά ΚΑΔ και ανά συνολική αξία;	30
B3. Ποιο είναι το δίκτυο (spatial network) logistics που αναδύεται από τις φορτωτικές ανά νησί;	31

Εμφανίζονται μήπως ακραίες τιμές σε ορισμένα νησιά, στη σχέση απόστασης, κόστους, πλήθους φορτωτικών;	32
(Γ) Γενικός Πληθυσμός.....	35
Γ1. Τι ποσοστό του πληθυσμού χρησιμοποιεί το Μεταφορικό Ισοδύναμο ανά νησί;	35
Υπάρχει μήπως χρονική διακύμανση ανά νησί εντός του έτους και μεταξύ των μηνών και των εποχών (χειμερινοί μήνες - καλοκαιρινοί μήνες);.....	36
Γ2. Ποιοι είναι οι κυρίαρχοι προορισμοί (μέγιστο 25%) της μετακίνησης του πληθυσμού ανά νησί;.....	37
Γ3. Ποια νησιά έχουν σχετικά μεγάλες ενδο-νησιακές μετακινήσεις πληθυσμού (πχ. Διεκπεραίωση διοικητικών υποχρεώσεων);	38
Γ4. Εξαρτήσεις / συνδέσεις μεταξύ διαφορετικών παραμέτρων (π.χ. κύκλος εργασιών επιχειρήσεων, χρήση Μεταφορικού Ισοδυναμού, κύριο νησί αναφοράς, εποχικότητα, κ.α.)	41
Γ5. Κατηγοριοποιήσεις βάσει διαφορετικών (και επικαλυπτόμενων) παραμέτρων.....	42
5. Σύνοψη.....	47
6. Επόμενα βήματα.....	48
Βιβλιογραφία.....	49

Ευρετήριο εικόνων

Εικόνα 1. κυρίαρχη Νομική Μορφή επιχείρησης ανά ΚΑΔ επιχειρήσεων και ανά νησί	14
Εικόνα 2. Μέσος όρος ηλικίας επιχείρησης ανά ΚΑΔ επιχειρήσεων και ανά νησί.....	14
Εικόνα 3. Μέσος κύκλος εργασιών ανά ΚΑΔ επιχειρήσεων και ανά νησί	15
Εικόνα 4. Κύριος ΚΑΔ και το άθροισμα χρηματοδότησης.....	16
Εικόνα 5. Εταιρείες οι οποίες χρηματοδοτήθηκαν.....	17
Εικόνα 6. Ποσά τελικής χρηματοδότησης ανά νησί και ανά ΚΑΔ	17
Εικόνα 7. Συνολική οικονομία του νησιού	18
Εικόνα 8. Συνολική οικονομία του νησιού	18
Εικόνα 9. Ποσοστό ΚΑΔ στη συνολική οικονομία του νησιού	19
Εικόνα 10. Εβδομάδες αίτησης ανά ΚΑΔ που δέχτηκαν χρηματοδότηση	19
Εικόνα 11. Εβδομάδες αίτησης ανά ΚΑΔ και πλήθος αιτήσεων που δέχτηκαν χρηματοδότηση	20
Εικόνα 12. Χρόνος έγκρισης ανά ΚΑΔ.....	20
Εικόνα 13. Μέση χρονική διάρκεια μεταξύ υποβολής αίτησης και Εγκριτικής Απόφασης, ανά ΚΑΔ και ανά νησί.....	21
Εικόνα 14. Συνολικό Επιλέξιμο Ποσό ανά εβδομάδα του έτους ανά νησί	22
Εικόνα 15. Διακύμανση εισιτηρίων 2019	23
Εικόνα 16. Ποσοστιαία σχέση επιλέξιμου/αιτούμενου ποσού ανά ΚΑΔ και ανά νησί για κάθε εβδομάδα του έτους.....	24
Εικόνα 17. Τομέας που ανήκει ο κάθε ΚΑΔ.....	25
Εικόνα 18. Χωρικές ομαδοποιήσεις συγγενικών ΚΑΔ επιχειρήσεων.....	25
Εικόνα 19. Αναλογία των ΚΑΔ επιχειρήσεων σε μια περιοχή, σε σχέση με την κατανομή των ΚΑΔ στο σύνολο του Αιγαίου	26
Εικόνα 20. Αναλογία κάθε ΚΑΔ επιχειρήσεων σε κάθε νησί προς το συνολικό πλήθος επιχειρήσεων στο νησί.....	27
Εικόνα 21. Πιθανότητα εμφάνισης κάθε ΚΑΔ επιχειρήσεων να συνυπάρχει σε μια περιοχή με άλλους τύπους επιχειρήσεων ανά ΚΑΔ επιχειρήσεων	28
Εικόνα 22. Πιθανότητα εμφάνισης ενός ΚΑΔ επιχειρήσεων σε μια περιοχή που υπάρχουν και άλλοι ίδιοι ΚΑΔ, ανά νησί	29
Εικόνα 23. Περιοχές αποστολής εμπορευμάτων προς νησί ανά νησί έδρας και ΚΑΔ.....	30
Εικόνα 24. Περιοχές που λαμβάνουν φορτωτικές από κάθε νησί ανά ΚΑΔ και ανά συνολική αξία...	31
Εικόνα 25. Τόπος αναχώρησης και προορισμού.....	31
Εικόνα 26. Αποστάσεις logistics.....	32
Εικόνα 27. Παράδειγμα εύρεσης ακραίων τιμών (outliers).....	33
Εικόνα 28. Παράδειγμα εντοπισμού ακραίας τιμής (outlier)	33
Εικόνα 29. Παράδειγμα εύρεσης ακραίων τιμών (outliers).....	34
Εικόνα 30. Παράδειγμα εντοπισμού ακραίων τιμών (outlier)	34
Εικόνα 31. Ποσοστό του πληθυσμού χρησιμοποιεί το Μεταφορικό Ισοδύναμο ανά νησί.....	35
Εικόνα 32. Διακύμανση ανά νησί ανά μήνα.....	36
Εικόνα 33. Διακύμανση ανά νησί ανά τουριστική περίοδο	37
Εικόνα 34. Κυρίαρχοι προορισμοί μετακίνησης πληθυσμού ανά νησί (μέγιστο 25%).....	38
Εικόνα 35. Επεξεργασία λιμένων της χώρας	39
Εικόνα 36. Ταξινόμηση αποτελεσμάτων	40
Εικόνα 37. Κωδικοί λιμανιών αναχώρησης και προορισμού	40
Εικόνα 38. Διαδρομές	40
Εικόνα 39. Κατηγορίες Επιδότησης	41
Εικόνα 40. Κατηγορίες και μεγέθη επιδότησης	42

Εικόνα 41. Κατηγορίες επιδότησης	42
Εικόνα 42. Γραμμική παλινδρόμηση	43
Εικόνα 43. Διαγράμματα διασποράς.....	44
Εικόνα 44. Γραμμική παλινδρόμηση	45
Εικόνα 45. Διάγραμμα Διασποράς	47

1. Εισαγωγή

Η θεωρία του Μεταφορικού Ισοδύναμου αναπτύχθηκε το 1974 από τον Pedersen και πρώτο-εφαρμόστηκε στο τιμολογιακό σύστημα της Νορβηγίας. Ο Pedersen παρατήρησε ότι το κράτος επενδύει χρήματα για την κατασκευή του οδικού δικτύου της χώρας, κατασκευάζοντας και συντηρώντας δρόμους και γέφυρες, οι οποίες συνδέουν περιοχές και πόλεις που βρίσκονταν στην ενδοχώρα. Δεν συνέβαινε το ίδιο όμως και για τις νησιωτικές περιοχές. Αυτό είχε ως αποτέλεσμα, ενώ οι κάτοικοι των νησιωτικών περιοχών να πληρώνουν φόρους για αυτές τις κατασκευές να πρέπει να πληρώσουν και επιπλέον ναύλους για τη μεταφορά τους με πλοίο, έτσι ώστε να είναι σε θέση να τις χρησιμοποιήσουν καθώς μόνο μέσω των επιβατηγών και οχηματαγωγών πλοίων μπορούν να συνδεθούν με την ενδοχώρα. Έτσι, έκρινε σκόπιμο να συνδέσει το κόστος μιας διαδρομής μέσω της ακτοπλοϊκής σύνδεσης με το αντίστοιχο οδικό κόστος για την ίδια απόσταση στο χερσαίο δίκτυο.

Με άλλα λόγια, το Μεταφορικό Ισοδύναμο είναι, κατά βάση, μια θεωρία που ορίστηκε ως μέτρο για την αντιμετώπιση των κοινωνικών αναγκών των κατοίκων των νησιωτικών περιοχών, οι οποίοι βρίσκονται σε μειονεκτική θέση σε σύγκριση με τους κατοίκους των ηπειρωτικών περιοχών καθώς δε δύνανται να μετακινηθούν με ευκολία αφού εξαρτώνται από την ποιότητα και την πυκνότητα της ακτοπλοϊκής σύνδεσης που υπάρχει κάθε φορά ενώ είναι αναγκασμένοι να επιφορτίζονται και με τα αυξημένα κόμιστρα.

Έτσι, ως Μεταφορικό Ισοδύναμο, στην Ελλάδα, ορίζεται το μέτρο του Υπουργείου Ναυτιλίας και Νησιωτικής Πολιτικής, με τη θέσπιση του οποίου επιδιώκεται η εξίσωση του κόστους μεταφοράς με μέσα μαζικής μεταφοράς από την ηπειρωτική Ελλάδα σε νησί ή από νησί σε νησί με το κόστος που θα ίσχυε στην ηπειρωτική Ελλάδα για την ίδια απόσταση (Μεταφορικό Ισοδύναμο, 2018). Έχει ως στόχο του την κατά το δυνατόν καλύτερη εναρμόνιση της τιμολογιακής επιβάρυνσης των χρηστών των θαλάσσιων μεταφορών σε σχέση με αυτή των χερσαίων μεταφορών έτσι ώστε να απολαμβάνουν ανάλογες υπηρεσίες αναφορικά με τη συχνότητα, το κόστος, το χρόνο μετακίνησης και την απόσταση που διανύεται αλλά και την ποιότητα τους, ανεξαρτήτως του μεταφορικού μέσου που χρησιμοποιείται.

2. Μεθοδολογία, Δεδομένα και Στοιχεία ανάλυσης

Το Μεταφορικό Ισοδύναμο στην Ελλάδα ξεκίνησε να εφαρμόζεται από το 2018. Υπάρχουν καταγραφές για τη χρήση του από επιχειρήσεις για την επιδότηση φορτωτικών αποστολής εμπορευμάτων σε ηπειρωτικούς προορισμούς αλλά και για τις μετακινήσεις ιδιωτών δικαιούχων είτε προς νησιωτικούς είτε προς ηπειρωτικούς προορισμούς.

Τα δεδομένα αυτά έχουν σημαντικό όγκο και αποτελούν μία πηγή πληροφορίας σχετική με τα χαρακτηριστικά επιχειρήσεων και ιδιωτών που χρησιμοποιούν το Μεταφορικό Ισοδύναμο, τα ποσά που επιχορηγήθηκε το Μεταφορικό Ισοδύναμο καθώς και τις περιοχές οι οποίες χρησιμοποιούν περισσότερο το Μεταφορικό Ισοδύναμο.

Επίσης η επεξεργασία των πρωτογενών δεδομένων του Μεταφορικού Ισοδύναμου προσφέρει σημαντική γνώση σε σχέση με το είδος και την κατανομή των επιχειρήσεων στον Ελληνικό νησιωτικό χώρο.

Η εξαγωγή των πληροφοριών αυτών απαιτεί ανάλυση μεγάλου όγκου δεδομένων με κατάλληλα εργαλεία πληροφορικής, ώστε τα αποτελέσματα να μπορούν να εξαχθούν σε σύντομο χρόνο και επιτρέποντας την παραμετροποίησή τους.

Μεθοδολογία

Τα δεδομένα είναι εκείνα τα ακατέργαστα γεγονότα και αριθμοί χωρίς κατάλληλες πληροφορίες, επομένως πρέπει να υποβληθούν σε επεξεργασία για να ληφθούν οι επιθυμητές πληροφορίες. Επίσης πληροφορίες είναι εκείνα τα αποτελέσματα που λαμβάνουμε μετά την επεξεργασία των ακατέργαστων δεδομένων σε διαφορετικά επίπεδα ή την εξαγωγή συμπερασμάτων από ένα δεδομένο σύνολο δεδομένων μέσω μιας διαδικασίας που ονομάζεται ανάλυση δεδομένων.

Η ανάλυση δεδομένων είναι απλώς η ανάλυση διαφόρων δεδομένων που σημαίνει καθαρισμός των δεδομένων, μετατροπή τους σε κατανοητή μορφή και, στη συνέχεια, μοντελοποίηση δεδομένων για εξαγωγή χρήσιμων πληροφοριών για επαγγελματική ή οργανωτική χρήση. Χρησιμοποιείται κυρίως στη λήψη επιχειρηματικών αποφάσεων. Πολλές βιβλιοθήκες είναι διαθέσιμες για την ανάλυση. Για παράδειγμα, NumPy, Pandas, Seaborn, Matplotlib, Sklearn, κλπ. (Nelli, 2015).

NumPy: Η NumPy είναι μια βιβλιοθήκη γραμμένη σε Python, που χρησιμοποιείται για αριθμητική ανάλυση σε Python. Αποθηκεύει τα δεδομένα με τη μορφή nd-arrays (n-διάστατους πίνακες).

Pandas: Η βιβλιοθήκη Pandas χρησιμοποιείται κυρίως για τη μετατροπή δεδομένων σε μορφή πίνακα και ως εκ τούτου, καθιστούν τα δεδομένα πιο δομημένα και εύκολα στην ανάγνωση.

Matplotlib: Η βιβλιοθήκη Matplotlib είναι ένα πακέτο οπτικοποίησης δεδομένων και σχεδίασης γραφικών για την Python και την αριθμητική επέκταση της βιβλιοθήκης NumPy που εκτελείται σε όλες τις πλατφόρμες.

Seaborn: Η βιβλιοθήκη Seaborn είναι ένα πακέτο οπτικοποίησης δεδομένων Python που βασίζεται στη βιβλιοθήκη matplotlib που είναι στενά συνδεδεμένη με δομές δεδομένων Pandas. Το βασικό συστατικό της Seaborn είναι η οπτικοποίηση, η οποία βοηθά στην εξερεύνηση και την κατανόηση δεδομένων.

Sklearn: Η βιβλιοθήκη Scikit-learn είναι η πιο χρήσιμη βιβλιοθήκη για μηχανική μάθηση στην Python. Περιλαμβάνει πολλά χρήσιμα εργαλεία για ταξινόμηση, παλινδρόμηση, ομαδοποίηση και μείωση διαστάσεων.

Η οπτικοποίηση δεδομένων θα βοηθήσει την ανάλυση των δεδομένων να γίνει πιο κατανοητή και διαδραστική, σχεδιάζοντας ή εμφανίζοντας τα δεδομένα σε εικονογραφική μορφή. Το Pandas, ένα πακέτο ανοιχτού κώδικα Python που ασχολείται με τρεις διαφορετικές δομές δεδομένων: σειρές, πλαίσια δεδομένων και πίνακες, λύνει αυτή την ανάγκη ανάλυσης και οπτικοποίησης δεδομένων (Embarak & Karkal, 2018).

Η ανάλυση δεδομένων με χρήση Python διευκολύνει την εργασία καθώς η γλώσσα προγραμματισμού Python έχει πολλά πλεονεκτήματα σε σχέση με οποιαδήποτε άλλη γλώσσα προγραμματισμού. Έχει εξέχοντα χαρακτηριστικά όπως είναι μια γλώσσα προγραμματισμού υψηλού επιπέδου (οι κωδικοί είναι σε μορφή αναγνώσιμη από τον άνθρωπο) είναι εύκολο να κατανοηθεί και να χρησιμοποιηθεί από οποιονδήποτε προγραμματιστή ή χρήστη. Πολλές βιβλιοθήκες και συναρτήσεις για στατιστική, αριθμητική ανάλυση είναι διαθέσιμες στην Python. Επιπλέον, ο πηγαίος κώδικας είναι ελεύθερα διαθέσιμος σε οποιονδήποτε (ελεύθερος και ανοιχτός κώδικας).

Κύριες φάσεις στην ανάλυση δεδομένων

Απαιτήσεις δεδομένων

Τα δεδομένα είναι η πιο σημαντική μονάδα σε κάθε μελέτη. Τα δεδομένα πρέπει να παρέχονται ως εισροές στην ανάλυση με βάση τις απαιτήσεις ανάλυσης. Ο όρος πειραματική μονάδα αναφέρεται στον τύπο του οργανισμού που θα χρησιμοποιηθεί για τη συλλογή δεδομένων (π.χ. ένα άτομο ή πληθυσμός ατόμων). Είναι δυνατό να εντοπιστούν και να ληφθούν συγκεκριμένες μεταβλητές πληθυσμού (όπως ύψος, βάρος, ηλικία και μισθός). Δεν έχει σημασία αν τα δεδομένα είναι αριθμητικά ή κατηγορικά.

Συλλογή δεδομένων

Η συλλογή δεδομένων είναι η διαδικασία συγκέντρωσης δεδομένων. Τα δεδομένα συλλέγονται από διάφορες πηγές, συμπεριλαμβανομένων των σχεσιακών βάσεων δεδομένων, των βάσεων δεδομένων στο υπολογιστικό νέφος (cloud) και άλλων πηγών, ανάλογα με τις ανάγκες της μελέτης. Αισθητήρες πεδίου, όπως κάμερες κυκλοφορίας, δορυφόροι, συστήματα παρακολούθησης και ούτω καθεξής, μπορούν επίσης να χρησιμοποιηθούν ως πηγές δεδομένων.

Επεξεργασία δεδομένων

Τα δεδομένα που συλλέγονται πρέπει να υποβάλλονται σε επεξεργασία ή να οργανώνονται για ανάλυση. Για παράδειγμα, αυτά μπορεί να περιλαμβάνουν τη διάταξη δεδομένων σε σειρές και στήλες σε μορφή πίνακα (γνωστά ως δομημένα δεδομένα) για περαιτέρω ανάλυση, συχνά μέσω της χρήσης υπολογιστικών φύλλων ή στατιστικού λογισμικού.

Καθαρισμός δεδομένων

Η μέθοδος καθαρισμού δεδομένων μετά την επεξεργασία και την οργάνωσή τους είναι γνωστή ως καθαρισμός δεδομένων. Σαρώνει για ασυνέπειες δεδομένων, διπλότυπα και σφάλματα και στη συνέχεια τα καταργεί. Η διαδικασία καθαρισμού δεδομένων περιλαμβάνει εργασίες όπως αντιστοίχιση εγγραφών, αναγνώριση ανακρίβειας δεδομένων, ταξινόμηση δεδομένων, αναγνώριση ακραίων δεδομένων, ορθογραφικός έλεγχος δεδομένων κειμένου και συντήρηση ποιότητας δεδομένων. Κατά συνέπεια, μας εμποδίζει να έχουμε απροσδόκητα αποτελέσματα και μας βοηθά να παρέχουμε δεδομένα υψηλής ποιότητας, τα οποία είναι απαραίτητα για ένα επιτυχημένο αποτέλεσμα.

Διερευνητική ανάλυση δεδομένων

Μόλις τα σύνολα δεδομένων καθαριστούν και χωρίς σφάλματα, μπορεί στη συνέχεια να αναλυθούν. Μπορεί να εφαρμοστεί μια ποικιλία τεχνικών, όπως διερευνητική ανάλυση δεδομένων- κατανόηση των μηνυμάτων που περιέχονται στα δεδομένα που λαμβάνονται και περιγραφικά στατιστικά στοιχεία- εύρεση μέσου όρου, διάμεσου κ.λπ. Η οπτικοποίηση δεδομένων είναι επίσης μια τεχνική που χρησιμοποιείται, στην οποία τα δεδομένα αναπαρίστανται σε γραφική μορφή προκειμένου να αποκτηθούν πρόσθετες πληροφορίες, σχετικά με τις πληροφορίες εντός των δεδομένων (Jambu, 1991).

Μοντελοποίηση και αλγόριθμοι

Μαθηματικοί τύποι ή μοντέλα (γνωστοί ως αλγόριθμοι), μπορούν να εφαρμοστούν στα δεδομένα προκειμένου να προσδιοριστούν οι σχέσεις μεταξύ των μεταβλητών. για παράδειγμα, χρησιμοποιώντας συσχέτιση ή αιτιότητα.

Προϊόν δεδομένων

Ένα προϊόν δεδομένων, είναι μια εφαρμογή υπολογιστή που λαμβάνει δεδομένα εισόδου και παράγει εξόδους, τροφοδοτώντας τα πίσω στο περιβάλλον. Μπορεί να βασίζεται σε μοντέλο ή αλγόριθμο.

Ανάλυση δεδομένων με χρήση Python

Σε αυτή την ενότητα θα μελετηθεί η ανάλυση δεδομένων με χρήση python. Θα γίνουν κατανοητά τα πιο βασικά πράγματα, όπως το γιατί χρησιμοποιείται η python για ανάλυση δεδομένων. Επιπλέον, θα παρουσιαστεί πώς μπορεί κάποιος να αρχίσει να χρησιμοποιεί python. Θα παρουσιαστούν οι σημαντικές βιβλιοθήκες, οι πλατφόρμες, το σύνολο δεδομένων για την πραγματοποίηση της ανάλυσης. Δίνεται η χρήση διαφόρων συναρτήσεων python για αριθμητική ανάλυση, καθώς και διάφορες μέθοδοι δημιουργίας γραφημάτων ή διαγραμμάτων.

Λόγοι επιλογής της γλώσσας Python

Η Python είναι μια γλώσσα προγραμματισμού υψηλού επιπέδου, ερμηνευμένη, πολλαπλών χρήσεων. Πολλά παραδείγματα προγραμματισμού όπως η διαδικαστική γλώσσα προγραμματισμού, ο αντικειμενοστραφής προγραμματισμός υποστηρίζεται στην python. Μπορεί να χρησιμοποιηθεί για πολλές εφαρμογές, που περιλαμβάνουν στατιστικούς υπολογισμούς με διάφορα πακέτα και συναρτήσεις. Επιπλέον, είναι εύκολο να το μάθεις. Μπορεί να παραληφθεί από οποιονδήποτε, συμπεριλαμβανομένων εκείνων που έχουν λιγότερες δεξιότητες προγραμματισμού (Van Rossum, et al., 2017).

Ορισμένα χαρακτηριστικά της Python αναφέρονται παρακάτω:

1. Ανοιχτός και δωρεάν κώδικας
2. Δυναμική στοιχειοθεσία
3. Πολυάριθμα περιβάλλοντα ανάπτυξης
4. Βιβλιοθήκες που χρησιμοποιούνται: Numpy, Pandas, Seaborn, Matplotlib

Δεδομένα προς ανάλυση

Από το 2018 όταν και ξεκίνησε η εφαρμογή του Μεταφορικού Ισοδύναμου στην Ελλάδα, τα δεδομένα της χρήσης του από πολίτες και επιχειρήσεις έχουν καταγραφεί. Δημιουργήθηκε έτσι ένα σύνολο δεδομένων με καταγραφές για τη χρήση του. Τα δεδομένα χωρίζονται σε δύο κλάσεις: δεδομένα επιχειρήσεων οι οποίες επιδοτούνται για την αποστολή εμπορευμάτων σε εμπορικούς προορισμούς και δεδομένα ιδιωτών, τα οποία αφορούν σε μετακινήσεις ιδιωτών δικαιούχων είτε προς νησιωτικούς είτε προς ηπειρωτικούς προορισμούς.

Όνομα αρχείου δεδομένων που χρησιμοποιήθηκε:

ΕΠΙΧΕΙΡΗΣΕΙΣ	
Αιτήσεις Χρηματοδότησης.xlsx	Αφορά δεδομένα των αιτήσεων επιχειρήσεων για επιδότηση από το Μεταφορικό Ισοδύναμο. Επίσης περιέχει τον κύκλο εργασιών της επιχείρησης που κάνει την αίτηση. Πλήθος εγγραφών: 19.370

Επιχειρήσεις.xlsx	<p>Μητρώο των επιχειρήσεων ανά νησί, οι οποίες εγγράφηκαν στο μέτρο του Μεταφορικού Ισοδυνάμου. Περιέχει πληροφορίες για τη νομική μορφή της επιχείρησης, το μέγεθός της, την ηλικία της, κύκλο εργασιών, τον κύριο ΚΑΔ και τους συμπληρωματικούς ΚΑΔ, οι οποίοι περιγράφουν τις δραστηριότητές της.</p> <p>Πλήθος εγγραφών: 16.057</p>
Ναύλοι Οχημάτων.xlsx	<p>Αρχείο των παραστατικών μεταφοράς επαγγελματικών οχημάτων των επιχειρήσεων που εντάχθηκαν στο Μεταφορικό Ισοδύναμο.</p> <p>Περιέχει μεταξύ άλλων πληροφορίες κωδικού επιχείρησης και κωδικού υποβολής αίτησης, κόστος ναύλου, ημερομηνία, αριθμό παραστατικού, λιμάνι και νησί αναχώρησης, λιμάνι και νησί προορισμού και την χρηματοδότηση που υπολογίζεται βάσει του Μεταφορικού Ισοδυνάμου για κάθε ναύλο.</p> <p>Πλήθος εγγραφών: 44.312</p>
Πληρωμές επιχειρήσεων.xlsx	<p>Περιέχει τις εκκαθαρισμένες πληρωμές προς τις επιχειρήσεις, σύμφωνα με τις προβλέψεις του Μεταφορικού Ισοδυνάμου. Πρόκειται για το ποσό που επιστρέφεται στις επιχειρήσεις, λόγω της επιδότησής τους από το Μεταφορικό Ισοδύναμο.</p> <p>Περιέχει πληροφορίες σχετικές με τον Κωδικό της επιχείρησης, τον Κωδικό Υποβολής Αίτησης, το ποσό που θα πληρωθεί στην επιχείρηση, ο Α/Α Εντολής Πληρωμής και η Ημερομηνία εκτέλεσης της Εντολής Πληρωμής.</p> <p>Πλήθος εγγραφών: 19.019</p>
Φορτωτικές.xlsx	<p>Πρόκειται για το αρχείο στο οποίο βρίσκονται καταχωρημένες οι φορτωτικές, οι οποίες συνοδεύουν τα επαγγελματικά οχήματα τα οποία μεταφέρουν προϊόντα από μία επιχείρηση σε μία άλλη.</p> <p>Μεταξύ άλλων στηλών, περιλαμβάνονται οι ακόλουθες: Κωδικός Παραστατικού, Κωδικός Επιχείρησης, Κωδικός Υποβολής Αίτησης, Τύπος Παραστατικού, Ημ/νία Έκδοσης, Αριθμός Παραστατικού, Αξία χωρίς ΦΠΑ Αξία με ΦΠΑ, Επιλέξιμη αξία χωρίς ΦΠΑ, Υπολογισμένη χρηματοδότηση, Είδος Μεταφερόμενου Φορτίου, Τόπος Αναχώρησης, Τόπος Προορισμού, Απόσταση</p> <p>Πλήθος εγγραφών: 11.047.782</p>

ΙΔΙΩΤΕΣ	
Δικαιούχοι 30-04-20.xlsx	<p>Περιέχει τους δικαιούχους στο Μεταφορικό Ισοδύναμο κατοίκους των νησιών της Ελλάδας μέχρι τις 20/4/2020. Τα πεδία τα οποία περιγράφουν κάθε δικαιούχο είναι τα εξής:</p> <p>Αναγνωριστικό Δικαιούχου, Ωφελούμενη Μονάδα που ανήκει, Χρονικό διάστημα όπου είναι ενεργός στη μονάδα (από), Χρονικό διάστημα όπου είναι ενεργός στη μονάδα (έως), Ενεργός, Είδος Ωφελούμενης Μονάδας, Κατηγορία Υπαλλήλου, Ρόλος στη μονάδα, Κωδικός Νησιού, Νησί, Έτος Γέννησης δικαιούχου, Αριθμός Εισιτηρίων στο πλαίσιο της μονάδας το 2018, Αριθμός Εισιτηρίων στο πλαίσιο της μονάδας το 2019, Αριθμός Εισιτηρίων στο πλαίσιο της μονάδας το 2020.</p> <p>Πλήθος εγγραφών: 410.488</p>
Εισιτήρια 2018.xlsx	<p>Αρχείο εισιτηρίων που εκδόθηκαν σε δικαιούχους του Μεταφορικού Ισοδυναμού το έτος 2018, όταν ξεκίνησε η εφαρμογή του μέτρου. Περιέχει τις εξής πληροφορίες για κάθε εισιτήριο:</p> <p>ID Εισιτηρίου (system-id), Αναγνωριστικό Δικαιούχου που ταξίδεψε, ID Ωφελούμενης Μονάδας στην οποία προσμετράται το εισιτήριο, Είδος Εισιτηρίου (Ακτοπλοϊκό / Αεροπορικό), Κωδικός Λιμανιού Αναχώρησης, Κωδικός Λιμανιού Άφιξης, Είδος πλοίου (Συμβατικό, Ταχύπλοο), Τύπος ακτοπλοϊκού εισιτηρίου, Φύλο επιβάτη, Ηλικιακή ομάδα επιβάτη, Κωδικός Αεροδρομίου Αναχώρησης, Κωδικός Αεροδρομίου Άφιξης, Ημερομηνία και Ώρα αναχώρησης, Τιμή Εισιτηρίου, Ποσοστό Έκπτωσης, ΑΝΗΚΟ που υπολογίστηκε για το εισιτήριο, Κατώφλι Μεταφορικού Ισοδυναμού για τη διαδρομή, Επιλέξιμο, Επιδοτήσιμο.</p> <p>Πλήθος εγγραφών: 68.247</p>
Εισιτήρια 2019 με θερινές τιμές_1.xlsx	<p>Αρχείο εισιτηρίων που εκδόθηκαν σε δικαιούχους του Μεταφορικού Ισοδυναμού από 26/5/2019 έως και 1/9/2019.</p> <p>Έχει ακριβώς την ίδια δομή με το αρχείο Εισιτήρια 2018.xlsx.</p> <p>Πλήθος εγγραφών: 595.676</p>

Εισιτήρια 2019 με θερινές τιμές_2.xlsx	<p>Αρχείο εισιτηρίων που εκδόθηκαν σε δικαιούχους του Μεταφορικού Ισοδυνάμου από 1/10/2019 έως και 28/12/2019.</p> <p>Έχει ακριβώς την ίδια δομή με το αρχείο Εισιτήρια 2018.xlsx.</p> <p>Πλήθος εγγραφών: 568.499</p>
Εισιτήρια 2019 με χειμερινές τιμές.xlsx	<p>Αρχείο εισιτηρίων που εκδόθηκαν σε δικαιούχους του Μεταφορικού Ισοδυνάμου από 1/1/2019 έως και 25/5/2019.</p> <p>Έχει ακριβώς την ίδια δομή με το αρχείο Εισιτήρια 2018.xlsx.</p> <p>Πλήθος εγγραφών: 319.441</p>
Εισιτήρια 2020.xlsx	<p>Αρχείο εισιτηρίων που εκδόθηκαν σε δικαιούχους του Μεταφορικού Ισοδυνάμου από 1/1/2020 έως και 1/5/2022.</p> <p>Έχει ακριβώς την ίδια δομή με το αρχείο Εισιτήρια 2018.xlsx.</p> <p>Πλήθος εγγραφών: 446.954</p>
Ωφελούμενες-Μονάδες 30-04-20.xlsx	<p>Αρχείο εγγεγραμμένων ωφελούμενων πολιτών στο Μεταφορικό Ισοδύναμο έως 30/4/2020. Τα στοιχεία που τηρούνται για κάθε ωφελούμενο είναι:</p> <p>ID Ωφελούμενης Μονάδας, Είδος Ωφελούμενης Μονάδας, Κατηγορία Υπαλλήλου, Κωδικός Νησιού, Νησί, Σύνολο μελών (ενεργά και μη ενεργά), Ενεργά μέλη της ωφελούμενης μονάδας, Αριθμός Διαδρομών που δικαιούται το 2018, Επιλέξιμες Διαδρομές του 2018, Επιδοτούμενες Διαδρομές το 2018, Συνολική επιδότηση για το 2018, Αριθμός Διαδρομών που δικαιούται το 2019, Επιλέξιμες Διαδρομές του 2019, Επιδοτούμενες Διαδρομές το 2019, Συνολική επιδότηση για το 2019, Αριθμός Διαδρομών που δικαιούται το 2020, Επιλέξιμες Διαδρομές του 2020, Επιδοτούμενες Διαδρομές το 2020, Συνολική επιδότηση για το 2020.</p> <p>Πλήθος εγγραφών: 196.233</p>

Ερευνητικά ερωτήματα

Βασιστήκαμε σε ένα σύνολο ερωτημάτων τα οποία έχουν ενδιαφέρον τόσο για τον προσδιορισμό της χρήσης του Μεταφορικού Ισοδυνάμου όσο και των επιπτώσεων που έχει αυτό στους πολίτες και τις επιχειρήσεις της νησιωτικής Ελλάδας.

Τα ερωτήματα αυτά είναι διαχωρισμένα στις εξής κλάσεις:

(Α) Επιχειρήσεις & οικονομική δραστηριότητα ανά νησί

(B) Ανάλυση φορτωτικών

(Γ) Γενικός Πληθυσμός

Ο παρακάτω πίνακας καταγράφει αναλυτικά τα ερωτήματα ανά κλάση, τα οποία θα αναλυθούν στο πλαίσιο της εργασίας αυτής:

(A) Επιχειρήσεις & οικονομική δραστηριότητα ανά νησί

A1. Ποια είναι η κυρίαρχη Νομική Μορφή επιχείρησης ανά ΚΑΔ επιχειρήσεων και ανά νησί

A2. Ποιος είναι ο μέσος όρος ηλικίας επιχείρησης ανά ΚΑΔ επιχειρήσεων και ανά νησί

A3. Ποιος είναι ο μέσος κύκλος εργασιών ανά ΚΑΔ επιχειρήσεων και ανά νησί

A4. Ποια είναι η συνολική χρηματοδότηση ανά ΚΑΔ επιχειρήσεων και ανά νησί

A5. Ποια είναι η μέση χρονική διάρκεια μεταξύ υποβολής αίτησης και Εγκριτικής Απόφασης, ανά ΚΑΔ και ανά νησί

A6. Ποιο είναι το συνολικό Επιλέξιμο Ποσό ανά εβδομάδα του έτους ανά νησί

i. Πότε ξεκινάει η τουριστική περίοδος ανά νησί

ii. Πόσο διαρκεί η τουριστική περίοδος ανά νησί

A7. Ποια είναι η ποσοστιαία σχέση επιλέξιμου/αιτούμενου ποσού ανά ΚΑΔ και ανά νησί για κάθε εβδομάδα του έτους

A8. Σε ποια νησιά παρατηρούνται χωρικές ομαδοποιήσεις συγγενικών ΚΑΔ επιχειρήσεων

A9. Ποια είναι η συνύπαρξη ΚΑΔ ανά νησί

A9.1 LQ: Ποια είναι η αναλογία των ΚΑΔ επιχειρήσεων σε μια περιοχή, σε σχέση με την κατανομή των ΚΑΔ στο σύνολο του Αιγαίου

A9.2 Dissimilarity Index: Ποια είναι η αναλογία κάθε ΚΑΔ επιχειρήσεων ξεχωριστά σε κάθε νησί προς τον συνολικό πλήθος επιχειρήσεων στο νησί

A9.3 Interaction Index: Ποια είναι η πιθανότητα εμφάνισης κάθε ΚΑΔ επιχειρήσεων να συνυπάρχει σε μια περιοχή με άλλους τύπους επιχειρήσεων ανά ΚΑΔ επιχειρήσεων

A9.4 Isolation Index: Ποια είναι η πιθανότητα εμφάνισης ενός ΚΑΔ επιχειρήσεων σε μια περιοχή που υπάρχουν και άλλοι ίδιοι ΚΑΔ, ανά νησί

(B) Ερωτήματα για Φορτωτικές

B1. Ποιες είναι οι κυρίαρχες περιοχές που στέλνουν φορτωτικές σε κάθε νησί, ανά ΚΑΔ και ανά συνολική αξία

B2. Ποιες είναι οι κυρίαρχες περιοχές που λαμβάνουν φορτωτικές από κάθε νησί, ανά ΚΑΔ και ανά συνολική αξία

B3. Ποιο είναι το δίκτυο (spatial network) logistics που αναδύεται από τις φορτωτικές ανά νησί

Εμφανίζονται μήπως ακραίες τιμές σε ορισμένα νησιά, στη σχέση απόστασης, κόστους, πλήθους φορτωτικών

- i. Ακραίες τιμές στη σχέση απόστασης
- ii. Ακραίες τιμές στη σχέση κόστους
- iii. Ακραίες τιμές στη σχέση πλήθους φορτωτικών

(Γ) Γενικός Πληθυσμός

Γ1. Τι ποσοστό του πληθυσμού χρησιμοποιεί το Μεταφορικό Ισοδύναμο ανά νησί

Υπάρχει μήπως χρονική διακύμανση ανά νησί εντός του έτους και μεταξύ των μηνών και των εποχών (χειμερινοί μήνες - καλοκαιρινοί μήνες)

Γ2. Ποιοι είναι οι κυρίαρχοι προορισμοί (μέγιστο 25%) της μετακίνησης του πληθυσμού ανά νησί

Γ3. Ποια νησιά έχουν σχετικά μεγάλες ενδο-νησιακές μετακινήσεις πληθυσμού (πχ. Διεκπεραίωση διοικητικών υποχρεώσεων)

Γ4. Εξαρτήσεις / συνδέσεις μεταξύ διαφορετικών παραμέτρων (π.χ. κύκλος εργασιών επιχειρήσεων, χρήση Μεταφορικού Ισοδυναμού, κύριο νησί αναφοράς, εποχικότητα, κ.α.)

Γ5. Κατηγοριοποιήσεις βάσει διαφορετικών (και επικαλυπτόμενων) παραμέτρων.

- i. Ανάλυση Παλινδρόμησης
- ii. Συσχέτιση πλήθους παραστατικών με το μέγεθος εταιρίας

3. Σχετικές εργασίες

Έγινε σε βάθος αναζήτηση για την εύρεση βιβλιογραφίας σχετικής με επιστημονικά συγγράμματα, τα οποία αναλύουν τα αποτελέσματα εφαρμογής του Μεταφορικού Ισοδυναμού στην Ελλάδα και σε άλλες χώρες της Ευρώπης, οι οποίες έχουν εφαρμόσει αντίστοιχες πολιτικές. Η εφαρμογή του Μεταφορικού Ισοδυναμού πραγματοποιείται ήδη σε πολλές ευρωπαϊκές χώρες όσο και διεθνώς, είναι όμως περιορισμένη. Τα κράτη μέλη της Ευρωπαϊκής Ένωσης δεν διαθέτουν συγκεκριμένο τρόπο με τον οποίο να καθορίζουν τα ακτοπολιτικά κόμιστρα. Συνήθως, η πολιτική που χρησιμοποιείται σχετίζεται με τους κυβερνητικούς στόχους σχετικά με τη διατήρηση και την ανάπτυξη των τοπικών κοινωνιών, του εμπορίου, την προώθηση του τουρισμού και τη διατήρηση της ευαίσθητης πληθυσμιακής πυκνότητας σε απομακρυσμένα νησιά (Ντουμή & Βουκελάτου, 2020).

Η περιορισμένη εφαρμογή του Μεταφορικού Ισοδυναμού συνεπάγεται και τις περιορισμένες αναλύσεις και μελέτες περίπτωσης στη βιβλιογραφία. Δεν έγινε δυνατό να βρεθούν κάποιες αναλύσεις δεδομένων της εφαρμογής του μεταφορικού ισοδυναμού είτε σε άλλες ευρωπαϊκές χώρες είτε στην Ελλάδα.

Αναγκαιότητα ανάλυσης των δεδομένων του Μεταφορικού Ισοδυναμού

Τα νησιά αποτελούν σημαντικό πόρο για την ελληνική οικονομία (OECD, 2020).

Οι θάλασσες και οι παράκτιες περιοχές διαδραματίζουν πολύ σημαντικό οικονομικό και στρατηγικό ρόλο για την Ελλάδα και η θαλάσσια και γαλάζια οικονομία της Ελλάδας έχουν

μεγάλες δυνατότητες καινοτομίας και ανάπτυξης. Το 2018, η γαλάζια οικονομία στην Ελλάδα μετρούσε το 14,2% του συνόλου των θέσεων εργασίας και περίπου το 5,2% του ΑΕΠ. Ο παράκτιος τουρισμός και οι θαλάσσιες μεταφορές ήταν οι μεγαλύτεροι συνεισφέροντές με 13% του ΑΕΠ και 3,8% της απασχόλησης, ενώ οι θαλάσσιοι έμβιοι πόροι δημιούργησαν περίπου το 7% των θέσεων εργασίας και του ΑΕΠ (European Commission, 2020). Βρισκόμενη στο σταυροδρόμι Αφρικής, Ασίας και Ευρώπης, η Ελλάδα κατέχει στρατηγική θέση στις παγκόσμιες εμπορευματικές μεταφορές. Το μεγαλύτερο μέρος του εμπορίου αγαθών κατ' όγκο και αξία ανταλλάσσεται με φορτίο. Το λιμάνι του Πειραιά στην Αθήνα είναι ένα από τα δέκα κορυφαία ευρωπαϊκά λιμάνια σε αριθμό μονάδων φορτίου έως το 2017. Επιπλέον, είναι ένα από τα ταχύτερα αναπτυσσόμενα μεγάλα ευρωπαϊκά λιμάνια. ο αριθμός των εμπορευμάτων που διακινήθηκαν υπερδιπλασιάστηκε μεταξύ 2011 και 2017, σύμφωνα με την αύξηση των εξαγωγών την ίδια περίοδο. Τα επιβατικά λιμάνια είναι επίσης απασχολημένα. Το 2015 περισσότεροι από 8 εκατομμύρια επιβάτες πέρασαν από το λιμάνι του Πειραιά, όγκος που ξεπέρασε μόνο τα λιμάνια του Καλαί, του Ντόβερ, του Ελσίνκι, της Στοκχόλμης και του Ταλίν στην Ευρώπη. Ωστόσο, το 2015, το λιμάνι του Πειραιά εξυπηρετούσε 25% λιγότερους επιβάτες από το 2010. Το λιμάνι του Περάματος, ο δυτικός τερματικός σταθμός του λιμανιού της Αθήνας, επιβίβασε και αποβίβασε πάνω από 7 εκατομμύρια επιβάτες, σημειώνοντας πτώση κατά 45% μεταξύ 2010 και 2015 (Eurostat, 2015)

Το Μεταφορικό Ισοδύναμο δημιουργήθηκε και εφαρμόστηκε για να εξομαλυνθούν οι ανισότητες μεταξύ των πολιτών της ηπειρωτικής και νησιωτικής Ελλάδας και να αυξηθεί η ανταγωνιστικότητα των επιχειρήσεων με έδρα τα νησιά, λόγω της επιβάρυνσης του κόστους των προϊόντων τους από τα αυξημένα θαλάσσια ναύλα.

Το Μεταφορικό Ισοδύναμο ξεκίνησε να εφαρμόζεται το 2018. Έχουν προκύψει συνεπώς σημαντικά ιστορικά δεδομένα η ανάλυση των οποίων μπορεί να εξάγει σημαντικά συμπεράσματα ως προς την επίδραση του Μεταφορικού Ισοδυναμίου στους πολίτες και τις επιχειρήσεις της νησιωτικής χώρας.

Τεχνικές / Μέθοδοι Ανάλυσης Δεδομένων

Τα τελευταία χρόνια, η ευρεία χρήση των ηλεκτρονικών υπολογιστών και των στατιστικών προγραμμάτων, όπως SPSS (SPSS Inc, Chicago, Illinois, USA), Python, R, κ.λπ., έχει διευκολύνει σημαντικά την ανάλυση δεδομένων. Τα εργαλεία αυτά θα πρέπει να συνδυάζονται με την κατάλληλη μέθοδο, ώστε τα αποτελέσματα να είναι ορθά και τεκμηριωμένα. Ο κατάλληλος μεθοδολογικός σχεδιασμός και ο περιορισμός στο μεγαλύτερο δυνατό βαθμό του τυχαίου και του συστηματικού σφάλματος –και ιδιαίτερα των συγχυτών– αποτελούν το πρώτο και πλέον καθοριστικό βήμα στη διεξαγωγή μιας μελέτης και την εξαγωγή ασφαλών συμπερασμάτων. Εξίσου σημαντική βεβαίως είναι και η στατιστική επεξεργασία των δεδομένων με την επιλογή της κατάλληλης μεθόδου ανάλογα με το είδος τόσο της ερευνητικής υπόθεσης όσο και των δεδομένων.

Ποσοτικές μεταβλητές

Στην περίπτωση των ποσοτικών μεταβλητών απαιτείται καταρχήν η διεξαγωγή ελέγχου κανονικότητας, όπου στην πλειονότητα των περιπτώσεων είναι ο έλεγχος Kolmogorov-Smirnov. (Boslaugh & Watters, 2008) (Foster, et al., 2006) Σημειώνεται ότι στην περίπτωση όπου ο αριθμός των παρατηρήσεων είναι μικρός, συνιστάται η εφαρμογή του ελέγχου κανονικότητας Shapiro-Wilk. Με τους ελέγχους κανονικότητας εκτιμάται εάν η κατανομή των

πληθυσμών από τους οποίους προέρχονται τα δεδομένα των «δειγμάτων» είναι η κανονική κατανομή ή αν την προσεγγίζει.

Στην περίπτωση όπου η κατανομή των πληθυσμών από τους οποίους προέρχονται τα μελετώμενα «δείγματα» είναι η κανονική κατανομή (normal distribution), τότε χρησιμοποιούνται οι παραμετρικές μέθοδοι (parametric methods), ενώ στην περίπτωση που δεν είναι η κανονική κατανομή χρησιμοποιούνται οι μη παραμετρικές μέθοδοι. (Boslaugh & Watters, 2008) (Foster, et al., 2006) το σημαντικότερο πλεονέκτημα των μη παραμετρικών μεθόδων είναι ότι μπορούν να εφαρμοστούν στην περίπτωση όπου η κατανομή των πληθυσμών δεν είναι η κανονική κατανομή. Επιπλέον, στις μη παραμετρικές μεθόδους χρησιμοποιούνται οι διατάξεις των παρατηρήσεων και όχι οι πραγματικές τιμές των παρατηρήσεων, με αποτέλεσμα να είναι δυνατή η εφαρμογή τους και στην περίπτωση «δειγμάτων» με μικρό αριθμό παρατηρήσεων. Η χρήση, εξάλλου, των διατάξεων δίνει τη δυνατότητα εφαρμογής των μη παραμετρικών μεθόδων και στην περίπτωση διατάξιμων δεδομένων και όχι μόνο ποσοτικών. Επίσης, οι μη παραμετρικές μέθοδοι είναι λιγότερο ευαίσθητες στο σφάλμα μέτρησης εξαιτίας της χρήσης των διατάξεων. Οι μη παραμετρικές μέθοδοι, ωστόσο, παρουσιάζουν και ορισμένα μειονεκτήματα έναντι των αντίστοιχων παραμετρικών, όπως μικρότερη στατιστική ισχύ, λιγότερο ειδικές υποθέσεις και μη χρήση όλης της πληροφορίας που είναι γνωστή για μια κατανομή, καθώς χρησιμοποιούνται οι διατάξεις των παρατηρήσεων και όχι οι πραγματικές τους τιμές. Τέλος, στην περίπτωση των παραμετρικών μεθόδων, για τον έλεγχο της μηδενικής υπόθεσης χρησιμοποιούνται οι μέσες τιμές και η μεταβλητότητα των παρατηρήσεων εκτιμάται με τη χρήση της τυπικής απόκλισης, ενώ στην περίπτωση των μη παραμετρικών μεθόδων χρησιμοποιούνται οι διάμεσοι και το ενδοτεταρτημοριακό εύρος, αντίστοιχα.

Συσχέτιση

Το είδος και το μέγεθος της γραμμικής σχέσης μεταξύ δύο ποσοτικών μεταβλητών που ακολουθούν την κανονική κατανομή εκτιμάται με το συντελεστή συσχέτισης του Pearson (Pearson's coefficient of correlation), ενώ όταν έστω και μία από τις δύο ποσοτικές μεταβλητές δεν ακολουθεί την κανονική κατανομή, τότε χρησιμοποιείται ο συντελεστής συσχέτισης διατάξεων του Spearman (Spearman's rank correlation coefficient). (Foster, et al., 2006) Επιπλέον, ο συντελεστής συσχέτισης διατάξεων του Spearman χρησιμοποιείται και στην περίπτωση όπου η μία ή και οι δύο μεταβλητές είναι σε διατεταγμένη κλίμακα.

Ο συντελεστής συσχέτισης του Pearson συμβολίζεται με r όταν αφορά σε ολόκληρο τον πληθυσμό και με r όταν υπολογίζεται για ένα συγκεκριμένο «δείγμα» (αποτελώντας έτσι εκτίμηση για τον πληθυσμό από τον οποίο προέρχεται το «δείγμα»). Ο συντελεστής συσχέτισης του Pearson δεν έχει μονάδες μέτρησης και λαμβάνει τιμές από -1 έως +1. Οι τιμές -1 έως +1 συμβαίνουν όταν υπάρχει μια τέλεια (αρνητική ή θετική, αντίστοιχα) γραμμική σχέση μεταξύ των δύο μεταβλητών. Όταν ο συντελεστής συσχέτισης λαμβάνει την τιμή 0 σημαίνει ότι δεν υπάρχει γραμμική σχέση μεταξύ των δύο μεταβλητών, αλλά δεν αποκλείεται η ύπαρξη μιας άλλης μη γραμμικής σχέσης. Εάν το μέγεθος της μιας μεταβλητής τείνει να αυξάνεται όπως αυξάνεται το μέγεθος και της άλλης μεταβλητής, τότε υπάρχει θετική συσχέτιση (positive correlation) μεταξύ των δύο μεταβλητών και ο συντελεστής συσχέτισης είναι >0 . Αντίθετα, εάν το μέγεθος της μιας μεταβλητής τείνει να αυξάνεται όπως ελαττώνεται το μέγεθος της άλλης μεταβλητής, τότε υπάρχει αρνητική συσχέτιση (negative correlation) μεταξύ των δύο μεταβλητών και ο συντελεστής συσχέτισης είναι <0 . Ο συντελεστής συσχέτισης του Pearson παρουσιάζει ορισμένα μειονεκτήματα, καθώς (α) ποσοτικοποιεί τη σχέση μεταξύ δύο μεταβλητών μόνο στην περίπτωση κατά την οποία η

σχέση αυτή είναι γραμμική, (β) είναι πολύ ευαίσθητος σε ακραίες τιμές και η ύπαρξή τους σε αρκετές περιπτώσεις οδηγεί σε λανθασμένα συμπεράσματα, (γ) η εκτιμηθείσα συσχέτιση δεν μπορεί να επεκταθεί πέρα από το παρατηρούμενο εύρος των μεταβλητών και (δ) η ύπαρξη μιας ισχυρής συσχέτισης μεταξύ δύο μεταβλητών δεν υποδηλώνει απαραίτητα και την ύπαρξη μιας σχέσης μεταξύ αιτίας και έκβασης.

Ο συντελεστής συσχέτισης διατάξεων του Spearman όταν υπολογίζεται σε «δείγματα» συμβολίζεται με r_s και λαμβάνει τιμές από -1 έως +1, με την ερμηνεία να είναι ίδια με εκείνη στην περίπτωση του συντελεστή συσχέτισης του Pearson. Ο συντελεστής συσχέτισης διατάξεων του Spearman είναι πολύ λιγότερο ευαίσθητος σε ακραίες τιμές από το συντελεστή συσχέτισης του Pearson και μπορεί να χρησιμοποιηθεί όταν η μία ή και οι δύο μεταβλητές είναι σε διατεταγμένη κλίμακα, αλλά επειδή ως μη παραμετρική μέθοδος χρησιμοποιεί τις διατάξεις και όχι τις παρατηρούμενες τιμές, δεν λαμβάνει υπόψη όλη την πληροφορία που είναι γνωστή για μια κατανομή.

Γραμμική παλινδρόμηση

Η απλή γραμμική παλινδρόμηση (simple linear regression), καθώς και η ανάλυση συσχέτισης, αποτελεί μια μέθοδο διερεύνησης της φύσης της σχέσης μεταξύ δύο ποσοτικών μεταβλητών που ακολουθούν την κανονική κατανομή. Η σημαντικότερη διαφορά ανάμεσα στην ανάλυση παλινδρόμησης και την ανάλυση συσχέτισης είναι ότι η πρώτη παρέχει τη δυνατότητα να διερευνηθεί η αλλαγή σε μια ποσοτική μεταβλητή, που ονομάζεται εξαρτημένη μεταβλητή (dependent variable) ή, αλλιώς, απόκριση (response), η οποία οφείλεται σε μια δεδομένη αλλαγή στη δεύτερη ποσοτική μεταβλητή, γνωστή ως ανεξάρτητη μεταβλητή (independent variable) ή, αλλιώς, επεξηγηματική μεταβλητή (explanatory variable). Η ανάλυση συσχέτισης δεν έχει τη δυνατότητα να πραγματοποιήσει τη διαφοροποίηση αυτή αντιμετωπίζοντας συμμετρικά τις δύο ποσοτικές μεταβλητές, όπως προαναφέρθηκε εκτενώς. Η ανάλυση παλινδρόμησης αποβλέπει στην πρόβλεψη ή την εκτίμηση της τιμής της εξαρτημένης μεταβλητής που αντιστοιχεί σε μια συγκεκριμένη τιμή της επεξηγηματικής μεταβλητής.

Το μοντέλο της πολλαπλής γραμμικής παλινδρόμησης (multiple linear regression) και γενικότερα τα μοντέλα που περιλαμβάνουν περισσότερες από μία ανεξάρτητες μεταβλητές ταυτόχρονα χρησιμοποιούνται για την εξουδετέρωση των συγχυτών. Στα πολυμεταβλητά μαθηματικά μοντέλα περιλαμβάνονται περισσότεροι από δύο προσδιοριστές, με το αποτέλεσμα του καθενός να μη συγχέεται από τη δράση των υπολοίπων, καθιστώντας έτσι τα μοντέλα αυτά ιδανικά για την εξουδετέρωση των συγχυτών μιας μελέτης. Παραδείγματος χάρη, εάν διερευνάται η σχέση μεταξύ του μεγέθους της επιχείρησης με το πλήθος των παραστατικών που εκδίδει και το ποσό που επιδοτήθηκε από το Μεταφορικό Ισοδύναμο, τότε το μοντέλο της πολλαπλής γραμμικής παλινδρόμησης θα περιλαμβάνει δύο ανεξάρτητες μεταβλητές. Οι δύο αυτές μεταβλητές συσχετίζονται μεταξύ τους, οπότε αναμένεται καθένα από τα τρία αυτά χαρακτηριστικά να είναι συγχυτής της σχέσης του άλλου με την εξαρτημένη μεταβλητή, που στη συγκεκριμένη περίπτωση είναι το μέγεθος της επιχείρησης.

Μαθηματικά, τουλάχιστον, δεν υπάρχει περιορισμός στον αριθμό των ανεξάρτητων μεταβλητών που μπορούν να χρησιμοποιηθούν στην πολλαπλή γραμμική παλινδρόμηση, αν και μικρός αριθμός δεδομένων επιτρέπει να χρησιμοποιούνται λίγες μόνο ανεξάρτητες μεταβλητές. Ορισμένοι ερευνητές θεωρούν ότι για κάθε ανεξάρτητη μεταβλητή πρέπει να υπάρχουν τουλάχιστον 10 παρατηρήσεις στο μελετώμενο «δείγμα», δηλαδή στην περίπτωση πέντε ανεξάρτητων μεταβλητών θα πρέπει να υπάρχουν τουλάχιστον 50 παρατηρήσεις, ενώ

ορισμένοι άλλοι θεωρούν ότι για κάθε ανεξάρτητη μεταβλητή θα πρέπει να υπάρχουν τουλάχιστον 40 παρατηρήσεις.

Εργαλεία ανάλυση δεδομένων

Η τεχνολογία του λογισμικού σήμερα έχει δημιουργήσει πολλές εφαρμογές, οι οποίες μπορούν να χρησιμοποιηθούν για συγκεκριμένους σκοπούς, σε σχέση με την ανάλυση δεδομένων. Για το σκοπό αυτό έχουν προταθεί διάφορες κατηγοριοποιήσεις των εργαλείων αυτών, σύμφωνα με τη λειτουργικότητά τους και τις μεθόδους ανάλυσης δεδομένων που μπορούν να υποστηρίξουν. Μία κατηγοριοποίησή τους είναι η ακόλουθη: (Bhuvanaga, et al., 2018)

Διαχείριση μεγάλων δεδομένων

- SQL
- MongoDB
- Spark
- Hadoop (MapReduce)

Διερευνητική ανάλυση δεδομένων / Μηχανική μάθηση

- Python
- R
- SAS
- Weka
- RapidMiner
- MatLab

4. Αποτελέσματα μελέτης

(Α) Επιχειρήσεις & οικονομική δραστηριότητα ανά νησί

A1. Ποια είναι η κυρίαρχη Νομική Μορφή επιχείρησης ανά ΚΑΔ επιχειρήσεων και ανά νησί;

Αρχικά, κανονικοποιούμε τα δεδομένα. Δημιουργούμε ένα νέο dataframe, το οποίο έχει δύο στήλες: τον κωδικό της επιχείρησης και τον ΚΑΔ. Όταν μία επιχείρηση έχει περισσότερους ΚΑΔ, τότε δημιουργούνται περισσότερες γραμμές για την επιχείρηση. Αυτό γίνεται κάνοντας split στην στήλη «Λίστα όλων των ΚΑΔ της επιχείρησης (κωδικοί)».

Στη συνέχεια, δημιουργείται ένα ακόμη dataframe με τα δεδομένα των επιχειρήσεων (αρχείο «Επιχειρήσεις.xlsx»).

Τα δύο dataframes γίνονται join στο πεδίο «Κωδικός της επιχείρησης» και παράγεται ένα νέο dataframe, στο οποίο θα γίνει τελικά η ομαδοποίηση «ΚΑΔ -> Νησί -> Νομική μορφή», καταμετρώντας το πλήθος των εταιριών που ανήκουν σε κάθε νομική μορφή και κρατώντας στο αποτέλεσμα τη μεγαλύτερη τιμή. Εάν υπάρχουν περισσότερες από μία νομικές μορφές με το ίδιο πλήθος επιχειρήσεων (το οποίο είναι το μεγαλύτερο για κάποιο νησί), τότε θα εμφανιστούν όλες οι νομικές αυτές μορφές στο αποτέλεσμα.

```
result = grouped.groupby(level=['ΚΑΔ', 'Νησί που έχει έδρα η επιχείρηση']).nlargest(1).reset_index(level=0, drop=True)
```


ΚΑΔ	Νησί που έχει έδρα η επιχείρηση	Νομική μορφή επιχείρησης	
10000	ΚΕΡΚΥΡΑ	ΟΕ	1
		ΦΥΣΙΚΟ ΠΡΟΣΩΠΟ	1
	ΚΩΣ	ΑΕ	1
	ΡΟΔΟΣ	ΟΕ	1
	ΤΗΝΟΣ	ΦΥΣΙΚΟ ΠΡΟΣΩΠΟ	1
			..
96091917	ΡΟΔΟΣ	ΟΕ	1
		ΦΥΣΙΚΟ ΠΡΟΣΩΠΟ	1
9900000	ΚΙΜΩΛΟΣ	ΑΕ	1
9901900	ΛΕΡΟΣ	ΦΥΣΙΚΟ ΠΡΟΣΩΠΟ	1
	ΜΗΛΟΣ	ΑΕ	1

Name: Κωδικός της επιχείρησης, Length: 67421, dtype: int64

Εικόνα 1. κυρίαρχη Νομική Μορφή επιχείρησης ανά ΚΑΔ επιχειρήσεων και ανά νησί

Η εφαρμογή δημιουργεί ένα πλαίσιο για την ανάλυση των νομικών μορφών ανά ΚΑΔ. Δεδομένου ότι οι ΚΑΔ είναι εκατοντάδες στο πλήθος, ο ερευνητής μπορεί να καθορίσει τον ΚΑΔ τον οποίο επιθυμεί να μελετήσει. Στο παράδειγμα της Εικόνας 1, ο ερευνητής διαπιστώνει ότι ο ΚΑΔ 10000 εμφανίζεται μόνο σε τέσσερα νησιά του Αιγαίου και εμφανίζεται μεγάλη ποικιλία ως προς τις νομικές μορφές που κυριαρχούν για αυτόν. Παρατηρούμε ότι ανάλογα με το νησί, κυριαρχούν επιχειρήσεις ΟΕ, ΑΕ και Φυσικά πρόσωπα.

A2. Ποιος είναι ο μέσος όρος ηλικίας επιχείρησης ανά ΚΑΔ επιχειρήσεων και ανά νησί; Αρχικά δημιουργήθηκε μία νέα στήλη στο dataframe “companies”, στην οποία υπολογίστηκε η ηλικία κάθε επιχείρησης ως η διαφορά του έτους ίδρυσής της από το τρέχον έτος.

Στη συνέχεια έγινε join το dataframe “companies” με το κανονικοποιημένο dataframe των ΚΑΔ ανά επιχείρηση (“kadnorm”). Στο αποτέλεσμα εφαρμόστηκε ομαδοποίηση ανά 'ΚΑΔ' και 'Νησί που έχει έδρα η επιχείρηση', υπολογίζοντας το μέσο όρο της στήλης 'Ηλικία' για κάθε ομάδα.

ΚΑΔ	Νησί που έχει έδρα η επιχείρηση	
10000	ΚΕΡΚΥΡΑ	28.00
	ΚΩΣ	31.00
	ΡΟΔΟΣ	16.00
	ΤΗΝΟΣ	20.00
1000000	ΑΝΔΡΟΣ	14.25
		...
96091917	ΜΥΚΟΝΟΣ	9.00
	ΡΟΔΟΣ	8.50
9900000	ΚΙΜΩΛΟΣ	28.00
9901900	ΛΕΡΟΣ	17.00
	ΜΗΛΟΣ	121.00

Εικόνα 2. Μέσος όρος ηλικίας επιχείρησης ανά ΚΑΔ επιχειρήσεων και ανά νησί

A3. Ποιος είναι ο μέσος κύκλος εργασιών ανά ΚΑΔ επιχειρήσεων και ανά νησί;

Εξετάζοντας το dataset, διαπιστώνουμε ότι περίπου 5.500 επιχειρήσεις δεν παρέχουν την πληροφορία του Κύκλου εργασιών τους. Συνεπώς, οι επιχειρήσεις θα πρέπει να εξαιρεθούν από τον υπολογισμό του μέσου όρου κύκλου εργασιών, γιατί θα μείωναν δραματικά το τελικό αποτέλεσμα.

Για το λόγο αυτό, από το αρχικό dataframe companies, δημιουργείται ένα νέο dataframe με όνομα “companiesWithTurnaround”, στο οποίο εισάγονται μόνο οι επιχειρήσεις με μη κενή τιμή στη στήλη 'Κύκλος Εργασιών'.

Στη συνέχεια έγινε join το dataframe “ companiesWithTurnaround ” με το κανονικοποιημένο dataframe των ΚΑΔ ανά επιχείρηση (“kadnorm”). Στο αποτέλεσμα εφαρμόστηκε ομαδοποίηση ανά 'ΚΑΔ' και 'Νησί που έχει έδρα η επιχείρηση', υπολογίζοντας το μέσο όρο της στήλης 'Κύκλος Εργασιών' για κάθε ομάδα.

ΚΑΔ	Νησί που έχει έδρα η επιχείρηση	
10000	ΚΕΡΚΥΡΑ	47888.67
	ΚΩΣ	5238633.99
	ΡΟΔΟΣ	362498.64
	ΤΗΝΟΣ	47004.07
1000000	ΑΝΔΡΟΣ	14183.88
	...	
96091917	ΚΩΣ	105533.19
	ΜΥΚΟΝΟΣ	15738.17
9900000	ΚΙΜΩΛΟΣ	1332315.68
9901900	ΛΕΡΟΣ	1292777.96
	ΜΗΛΟΣ	4441136.87

Εικόνα 3. Μέσος κύκλος εργασιών ανά ΚΑΔ επιχειρήσεων και ανά νησί

Βάσει της ανάλυσης αυτής, παρατηρούμε για τον ΚΑΔ 10000, ότι υπάρχει τεράστια ανομοιομορφία μεταξύ των τεσσάρων νησιών στα οποία δραστηριοποιούνται επιχειρήσεις με αυτόν. Η Κέρκυρα και η Τήνος έχουν μέσο κύκλο εργασιών επιχειρήσεων του ΚΑΔ 10000 στην περιοχή των 47.000€, ενώ η Κως της τάξης των 5,2 εκατομμυρίων €.

A4. Ποια είναι η συνολική χρηματοδότηση ανά ΚΑΔ επιχειρήσεων και ανά νησί;

- Ποιοι ΚΑΔ χρηματοδοτήθηκαν περισσότερο ανά νησί;

Λαμβάνουμε την πληροφορία για το ύψος της χρηματοδότησης από το dataset «Αιτήσεις χρηματοδότησης» και συγκεκριμένα τις στήλες «Εγκεκριμένη Χρηματοδότηση από τον Έλεγχο» και «Εγκεκριμένη Χρηματοδότηση από το Επανεέλεγχο». Συγκεκριμένα, εάν υπάρχει τιμή στη στήλη «Εγκεκριμένη Χρηματοδότηση από το Επανεέλεγχο», τότε λαμβάνεται αυτή υπόψη στο σύνολο της χρηματοδότησης. Διαφορετικά λαμβάνεται υπόψη η τιμή της στήλης «Εγκεκριμένη Χρηματοδότηση από τον Έλεγχο».

Επίσης, για να υπολογιστεί η χρηματοδότηση ανά ΚΑΔ, λαμβάνεται υπόψη η στήλη «Κύριος ΚΑΔ – κωδικός» του dataset «Επιχειρήσεις».

Δημιουργούνται δύο dataframes:

1. Companies. Περιέχει τις πληροφορίες του dataset «Επιχειρήσεις», από το οποίο διατηρούνται μόνο οι στήλες 'Κωδικός της επιχείρησης', 'Κύριος ΚΑΔ – κωδικός'.
2. aggregatedFunding. Περιλαμβάνει τις στήλες 'Κωδικός επιχείρησης', 'Νησί Έδρας' και τις συναθροιστικές στήλες 'Αιτούμενη Χρηματοδότηση' και 'Τελική χρηματοδότηση', οι οποίες περιλαμβάνουν το άθροισμα των αντίστοιχων ποσών ανά κωδικό επιχείρησης και ανά νησί.

Στη συνέχεια συγχωνεύονται τα δύο dataframes (companies και aggregatedFunding), ώστε να προκύψει ο κύριος ΚΑΔ για κάθε κωδικό επιχείρησης, για τον οποίο υπολογίστηκε το άθροισμα της χρηματοδότησης.

Νησί Έδρας	Κύριος ΚΑΔ - κωδικός	
ΑΓΑΘΟΝΗΣΙ	47111000	357.32
	47111005	0.00
ΑΓΙΟΣ ΕΥΣΤΡΑΤΙΟΣ	1114900	0.00
	46731301	0.00
	56101109	0.00
ΑΓΚΙΣΤΡΙ	43121200	5291.47
	47616100	0.00
ΑΙΓΙΝΑ	1253402	0.00
	1301015	0.00
	1472100	0.00
	8121200	956.50
	10510000	0.00
	10710000	0.00

Εικόνα 4. Κύριος ΚΑΔ και το άθροισμα χρηματοδότησης

- Τι ποσοστό αποτελούν αυτοί οι ΚΑΔ στην οικονομία κάθε νησιού;

```
ONLY COMPANIES THAT GOT FUNDING (['Τελική χρηματοδότηση'].sum())>0)
```

Κωδικός επιχείρησης	Νησί Έδρας	Κύριος ΚΑΔ - κωδικός	
38	ΧΙΟΣ	11050000	21437.41
40	ΚΑΛΥΜΝΟΣ	47111001	109271.82
49	ΠΑΤΜΟΣ	47111001	96427.64
52	ΜΥΤΙΛΗΝΗ	46731600	7141.10
59	ΧΙΟΣ	10392202	62263.02
64	ΚΑΡΠΑΘΟΣ	46691500	1738.29
69	ΜΥΤΙΛΗΝΗ	11011000	4614.42
70	ΧΙΟΣ	47111005	50498.04
73	ΜΥΤΙΛΗΝΗ	10412301	6170.40
80	ΧΙΟΣ	47595500	5321.56
81	ΜΥΤΙΛΗΝΗ	46390000	30089.38
84	ΜΥΤΙΛΗΝΗ	47910000	10.87
89	ΚΑΛΥΜΝΟΣ	47656700	3824.23

Εικόνα 5. Εταιρείες οι οποίες χρηματοδοτήθηκαν

Στο επόμενο dataframe (fundedKADperIsland), έχουν αθροιστεί τα ποσά τελικής χρηματοδότησης ανά νησί και ανά ΚΑΔ.

	Νησί Έδρας	Κύριος ΚΑΔ - κωδικός	Κωδικός επιχείρησης	Τελική χρηματοδότηση \
0	ΑΓΑΘΟΝΗΣΙ	47111000	14725	357.32
1	ΑΓΚΙΣΤΡΙ	43121200	5923	5291.47
2	ΑΙΓΙΝΑ	8121200	9865	956.50
3	ΑΙΓΙΝΑ	47211108	11772	5259.87
4	ΑΙΓΙΝΑ	49421101	3318	4230.39
5	ΑΛΟΝΝΗΣΟΣ	41200000	11717	997.13
6	ΑΛΟΝΝΗΣΟΣ	43320000	10938	201.35
7	ΑΛΟΝΝΗΣΟΣ	47211000	4129	7813.03
8	ΑΛΟΝΝΗΣΟΣ	47231511	4229	4136.35
9	ΑΛΟΝΝΗΣΟΣ	47788400	4241	2029.95

Εικόνα 6. Ποσά τελικής χρηματοδότησης ανά νησί και ανά ΚΑΔ

Στη συνέχεια θα συγχωνεύσουμε τον πίνακα fundedKADperIsland με τον πίνακα ο οποίος περιέχει το σύνολο των εσόδων (συνολική οικονομία κάθε νησιού) (islandGDP). Η εικόνα του dataframe islandGDP είναι η ακόλουθη:

```

GDP OF EACH ISLAND
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 78 entries, 0 to 77
Data columns (total 2 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Νησί που έχει έδρα η επιχείρηση          78 non-null     object
1   Οικονομία του νησιού                     78 non-null     float64
dtypes: float64(1), object(1)
memory usage: 1.3+ KB
None
      Νησί που έχει έδρα η επιχείρηση  Οικονομία του νησιού
0                ΑΓΑΘΟΝΗΣΙ                114914.94
1                ΑΓΙΟΣ ΕΥΣΤΡΑΤΙΟΣ           78918.63
2                ΑΓΚΙΣΤΡΙ                   818562.32
3                ΑΙΓΙΝΑ                     25718086.07
4                ΑΛΟΝΝΗΣΟΣ                 11607194.30
5                ΑΜΜΟΥΛΙΑΝΗ                777660.18
6                ΑΜΟΡΓΟΣ                   11667163.44
7                ΑΝΑΦΗ                      1528110.67
8                ΑΝΔΡΟΣ                     38816381.75
9                ΑΝΤΙΠΑΞΟΙ                   81898.28
10               ΑΝΤΙΠΑΡΟΣ                 13486722.66

```

Εικόνα 7. Συνολική οικονομία του νησιού

Στη συνέχεια τα δύο dataframes (fundedKADperIsland και islandGDP) θα συγχωνευθούν χρησιμοποιώντας ως κοινό γνώρισμα το όνομα του νησιού. Με τον τρόπο αυτό, το νέο dataframe που θα προκύψει θα περιέχει μία ακόμη στήλη, η οποία θα είναι η οικονομία του νησιού. Η στήλη «Τελική χρηματοδότηση» θα διαιρεθεί με τη στήλη «Κύκλος Εργασιών», ώστε να προκύψει το ποσοστό της χρηματοδότησης στο σύνολο της οικονομίας.

Νησί Έδρας	Κύριος ΚΑΔ - κωδικός	Τελική χρηματοδότηση	Κύκλος Εργασιών	Οικονομία του νησιού	
0	ΑΓΑΘΟΝΗΣΙ	47111000	357.32	0.00	114914.94
1	ΑΓΚΙΣΤΡΙ	43121200	5291.47	488567.75	818562.32
2	ΑΙΓΙΝΑ	8121200	956.50	253631.29	25718086.07
3	ΑΙΓΙΝΑ	47211108	5259.87	771936.17	25718086.07
4	ΑΙΓΙΝΑ	49421101	4230.39	41407.40	25718086.07
5	ΑΛΟΝΝΗΣΟΣ	41200000	997.13	181268.83	11607194.30
6	ΑΛΟΝΝΗΣΟΣ	43320000	201.35	117365.82	11607194.30
7	ΑΛΟΝΝΗΣΟΣ	47211000	7813.03	504067.75	11607194.30
8	ΑΛΟΝΝΗΣΟΣ	47231511	4136.35	434221.96	11607194.30
9	ΑΛΟΝΝΗΣΟΣ	47788400	2029.95	85451.44	11607194.30

Εικόνα 8. Συνολική οικονομία του νησιού

Στη συνέχεια δημιουργούμε μία νέα στήλη («Ποσοστό ΚΑΔ στην οικονομία»), η οποία προκύπτει από το πηλίκο της στήλης «Κύκλος εργασιών»/«Οικονομία του νησιού». Το αποτέλεσμα είναι το ποσοστό κατά το οποίο συμμετέχει ο συγκεκριμένος ΚΑΔ στη συνολική οικονομία του νησιού.

Νησί Έδρας	Κύριος ΚΑΔ - κωδικός	Τελική χρηματοδότηση	Κύκλος Εργασιών	Οικονομία του νησιού	Ποσοστό ΚΑΔ στην οικονομία	
0	ΑΓΑΘΟΝΗΣΙ	47111000	357.32	0.00	114914.94	0.00
1	ΑΓΚΙΣΤΡΙ	43121200	5291.47	488567.75	818562.32	59.69
2	ΑΙΓΙΝΑ	8121200	956.50	253631.29	25718086.07	0.99
3	ΑΙΓΙΝΑ	47211108	5259.87	771936.17	25718086.07	3.00
4	ΑΙΓΙΝΑ	49421101	4230.39	41407.40	25718086.07	0.16
5	ΑΛΟΝΝΗΣΟΣ	41200000	997.13	181268.83	11607194.30	1.56
6	ΑΛΟΝΝΗΣΟΣ	43320000	201.35	117365.82	11607194.30	1.01
7	ΑΛΟΝΝΗΣΟΣ	47211000	7813.03	504067.75	11607194.30	4.34
8	ΑΛΟΝΝΗΣΟΣ	47231511	4136.35	434221.96	11607194.30	3.74
9	ΑΛΟΝΝΗΣΟΣ	47788400	2029.95	85451.44	11607194.30	0.74

Εικόνα 9. Ποσοστό ΚΑΔ στη συνολική οικονομία του νησιού

Η ανάλυση αυτή δημιουργεί ένα εποπτικό πλαίσιο, ώστε ο ερευνητής να μπορεί να προσδιορίσει τη σημαντικότητα ενός ΚΑΔ (δραστηριότητας) σε κάθε νησί ξεχωριστά.

- Ποιες εβδομάδες του έτους αιτήθηκαν χρηματοδότηση ανά ΚΑΔ επιχειρήσεων;

Για την εξαγωγή του συγκεκριμένου στατιστικού μεγέθους, πρέπει να διεκπεραιωθούν τα εξής βήματα:

1. Δημιουργία dataframe με στοιχεία των αιτήσεων χρηματοδότησης. Το dataframe αυτό προέρχεται από τη φόρτωση του αρχείου Αιτήσεις Χρηματοδότησης.xlsx, και από το οποίο διατηρούνται μόνο οι χρήσιμες στήλες: 'Κωδικός επιχείρησης', 'Νησί Έδρας', 'Ημ/νία υποβολής της Αίτησης'. Το dataframe αυτό ονομάζεται "applications".
2. Δημιουργία νέας στήλης στο dataframe "applications", η οποία προκύπτει από τη στήλη 'Ημ/νία υποβολής της Αίτησης', ώστε να εξαχθεί ο αριθμός της εβδομάδας του έτους, στην οποία ανήκει η αντίστοιχη ημερομηνία υποβολής αίτησης. Η νέα στήλη ονομάζεται «Εβδομάδα αίτησης».
3. Συγχώνευση του dataframe "applications" με το dataframe "companies", ώστε στο dataframe που θα προκύψει να περιέχεται η στήλη του ΚΑΔ της κάθε επιχείρησης.
4. Στο dataframe που προκύπτει εφαρμόζεται συναθροιστική συνάρτηση count στα δεδομένα του, ομαδοποιημένα ανά ΚΑΔ και εβδομάδα που υποβλήθηκε η αίτηση.

Τα αποτελέσματα της επεξεργασίας, εμφανίζονται στον πίνακα που ακολουθεί:

Νησί Έδρας	Ημ/νία υποβολής της Αίτησης	Κωδικός της επιχείρησης	Κύριος ΚΑΔ - κωδικός	Εβδομάδα αίτησης	
0	ΜΥΤΙΛΗΝΗ	2020-01-30	37	10730000	5
1	ΜΥΤΙΛΗΝΗ	2019-10-03	37	10730000	40
2	ΜΥΤΙΛΗΝΗ	2019-04-02	37	10730000	14
3	ΜΥΤΙΛΗΝΗ	2018-10-10	37	10730000	41
4	ΧΙΟΣ	2020-02-17	38	11050000	8
...
19369	ΚΩΣ	2020-02-17	18177	11021200	8
19370	ΜΥΤΙΛΗΝΗ	2020-02-17	18178	47111005	8
19371	ΚΥΘΟΣ	2020-02-17	18179	56301007	8
19372	ΚΩΣ	2020-02-17	18180	69200000	8
19373	ΚΥΘΟΣ	2020-02-17	18181	56101111	8

Εικόνα 10. Εβδομάδες αίτησης ανά ΚΑΔ που δέχτηκαν χρηματοδότηση

	Κύριος ΚΑΔ - κωδικός	Εβδομάδα αίτησης	Πλήθος αιτήσεων
0	1000000	1	2
1	1000000	6	1
2	1000000	7	2
3	1000000	38	1
4	1000000	41	1
5	1113000	6	1
6	1114900	6	1
7	1130000	5	1
8	1130000	6	1
9	1130000	7	2

Εικόνα 11. Εβδομάδες αίτησης ανά ΚΑΔ και πλήθος αιτήσεων που δέχτηκαν χρηματοδότηση

Από την ανάλυση αυτή προκύπτει ότι οι αιτήσεις για χρηματοδότηση γίνονται κατά κύριο λόγο στην αρχή και το τέλος του έτους. Ενδεικτικά, για τον ΚΑΔ 1000000 παρατηρούμε ότι οι αιτήσεις έγιναν τις εβδομάδες 1,6, 7 (πρώτες εβδομάδες του έτους) και 38, 41 (εβδομάδες του τέταρτου τριμήνου του έτους).

A5. Ποια είναι η μέση χρονική διάρκεια μεταξύ υποβολής αίτησης και Εγκριτικής Απόφασης, ανά ΚΑΔ και ανά νησί;

Αρχικά δημιουργήσαμε ένα dataframe με όνομα “applications”, στο οποίο συμπεριλήφθηκαν οι εξής στήλες από το αρχείο «Αιτήσεις Χρηματοδότησης.xls»: 'Κωδικός επιχείρησης', 'Νησί Έδρας', 'Ημ/νία υποβολής της Αίτησης', 'Ημ/νία Εγκριτικής Απόφασης'

Στη συνέχεια δημιουργήθηκε μία νέα υπολογιζόμενη στήλη με όνομα 'Χρόνος έγκρισης', η οποία υπολογίζει τη διαφορά της ημερομηνίας Εγκριτικής Απόφασης από την Ημερομηνία υποβολής της αίτησης. Το αποτέλεσμα είναι το πλήθος των ημερών που μεσολαβούν από την αίτηση στην έγκριση.

Στη συνέχεια γίνεται συγχώνευση με το dataframe “companies”, ώστε να ενσωματωθεί στο dataframe “applications” η στήλη ΚΑΔ της επιχείρησης. Το τελικό dataframe που προκύπτει έχει τις εξής στήλες: 'Νησί Έδρας', 'Ημ/νία υποβολής της Αίτησης', 'Ημ/νία Εγκριτικής Απόφασης', 'Κωδικός της επιχείρησης', 'Κύριος ΚΑΔ – κωδικός', 'Χρόνος έγκρισης'.

	Νησί Έδρας	Ημ/νία υποβολής της Αίτησης	Ημ/νία Εγκριτικής Απόφασης	Κωδικός της επιχείρησης	Κύριος ΚΑΔ - κωδικός	Χρόνος έγκρισης
0	ΜΥΤΙΛΗΝΗ	2020-01-30	2020-04-03	37	10730000	64
1	ΜΥΤΙΛΗΝΗ	2019-10-03	2019-11-06	37	10730000	34
2	ΜΥΤΙΛΗΝΗ	2019-04-02	2019-05-14	37	10730000	42
3	ΜΥΤΙΛΗΝΗ	2018-10-10	2018-12-07	37	10730000	58
4	ΧΙΟΣ	2020-02-17	2020-04-03	38	11050000	46
...
19321	ΚΩΣ	2020-02-17	2020-04-03	18177	11021200	46
19322	ΜΥΤΙΛΗΝΗ	2020-02-17	2020-04-03	18178	47111005	46
19323	ΚΥΘΟΣ	2020-02-17	2020-04-03	18179	56301007	46
19324	ΚΩΣ	2020-02-17	2020-04-03	18180	69200000	46
19325	ΚΥΘΟΣ	2020-02-17	2020-04-03	18181	56101111	46

Εικόνα 12. Χρόνος έγκρισης ανά ΚΑΔ

Τέλος, προκειμένου να γίνει ο υπολογισμός της μέσης διάρκειας έγκρισης, ομαδοποιούνται οι εγγραφές ανά ΚΑΔ και ανά Νησί, υπολογίζοντας τη μέση τιμή (mean) της στήλης 'Χρόνος έγκρισης' για κάθε ομάδα εγγραφών. Το αποτέλεσμα αποτυπώνεται στον ακόλουθο πίνακα:

	Κύριος ΚΑΔ - κωδικός	Νησί Έδρας	Μέσος χρόνος έγκρισης
0	1000000	ΑΝΔΡΟΣ	57.333333
1	1000000	ΚΑΛΥΜΝΟΣ	10.000000
2	1000000	ΦΟΛΕΓΑΝΔΡΟΣ	90.000000
3	1113000	ΚΑΣΟΣ	56.000000
4	1114900	ΑΓΙΟΣ ΕΥΣΤΡΑΤΙΟΣ	4.000000
...
5338	96091102	ΣΑΜΟΣ	62.000000
5339	96091902	ΥΔΡΑ	46.000000
5340	96091904	ΜΗΛΟΣ	39.000000
5341	96091909	ΜΥΚΟΝΟΣ	33.000000
5342	96091916	ΚΩΣ	49.000000

Εικόνα 13. Μέση χρονική διάρκεια μεταξύ υποβολής αίτησης και Εγκριτικής Απόφασης, ανά ΚΑΔ και ανά νησί

Η παρατήρηση των μεγεθών της ανάλυσης αυτής δείχνει ότι για τη χορήγηση Εγκριτικής Απόφασης μεσολαβεί χρόνος μεγαλύτερος από 1 μήνα και μπορεί να φτάσει και τους δύο μήνες. Η λεπτομερέστερη ανάλυση του μεγέθους αυτού μπορεί να βοηθήσει τη Διοίκηση, ώστε να αναπροσαρμόσει τις διαδικασίες έγκρισης, ώστε οι επιχειρήσεις να ικανοποιούνται σε συντομότερο χρόνο.

Α6. Ποιο είναι το συνολικό Επιλέξιμο Ποσό ανά εβδομάδα του έτους ανά νησί;

Για τον υπολογισμό του επιλέξιμου ποσού ανά εβδομάδα ανά νησί, χρησιμοποιείται μόνο το dataset «Αιτήσεις Χρηματοδότησης.xls». Από δημιουργείται ένα dataframe με τίτλο "applications", στο οποίο διατηρούνται μόνο οι στήλες 'Νησί Έδρας', 'Ημ/νία υποβολής της Αίτησης', 'Επιλέξιμο ποσό', οι οποίες είναι απαραίτητες για τον υπολογισμό.

Στο dataframe προστίθεται μία νέα στήλη με τίτλο 'Εβδομάδα αίτησης', η οποία προκύπτει από τη στήλη "Ημ/νία υποβολής της Αίτησης", εφαρμόζοντας σε αυτή συνάρτηση μετατροπής της στον αριθμό της εβδομάδας του έτους που αυτή ανήκει, σύμφωνα με το πρότυπο ISO.

Τέλος, προκειμένου να γίνει ο υπολογισμός του συνολικού επιλέξιμου ποσού ανά εβδομάδα, ομαδοποιούνται οι εγγραφές ανά 'Εβδομάδα αίτησης' και 'Νησί Έδρας', υπολογίζοντας το άθροισμα (sum) της στήλης 'Επιλέξιμο ποσό' για κάθε ομάδα εγγραφών. Το αποτέλεσμα αποτυπώνεται στον ακόλουθο πίνακα:

Εβδομάδα αίτησης	Νησί Έδρας	Συνολικό επιλέξιμο ποσό
0	1 ΑΙΓΙΝΑ	3007.58
1	1 ΑΛΟΝΝΗΣΟΣ	39092.66
2	1 ΑΝΔΡΟΣ	5610.19
3	1 ΖΑΚΥΝΘΟΣ	2030.48
4	1 ΘΑΣΟΣ	64.48
...
1019	52 ΣΙΦΝΟΣ	2873.69
1020	52 ΣΚΙΑΘΟΣ	1146.06
1021	52 ΣΚΟΠΕΛΟΣ	101.59
1022	52 ΣΥΡΟΣ	3443.52
1023	52 ΧΙΟΣ	8427.71

Εικόνα 14. Συνολικό Επιλέξιμο Ποσό ανά εβδομάδα του έτους ανά νησί

Πότε ξεκινάει η τουριστική περίοδος ανά νησί;

Προκειμένου να υπολογιστεί η έναρξη και η λήξη της τουριστικής περιόδου, θα πρέπει να προσδιοριστεί η περίοδος στην οποία υπάρχει απότομη αύξηση των εισιτηρίων για άφιξη σε ένα νησί, διατήρηση των αφίξεων σε υψηλό επίπεδο και στο τέλος της τουριστικής περιόδου, απότομη μείωση των αφίξεων και κατάληξη σε μικρό αριθμό τους.

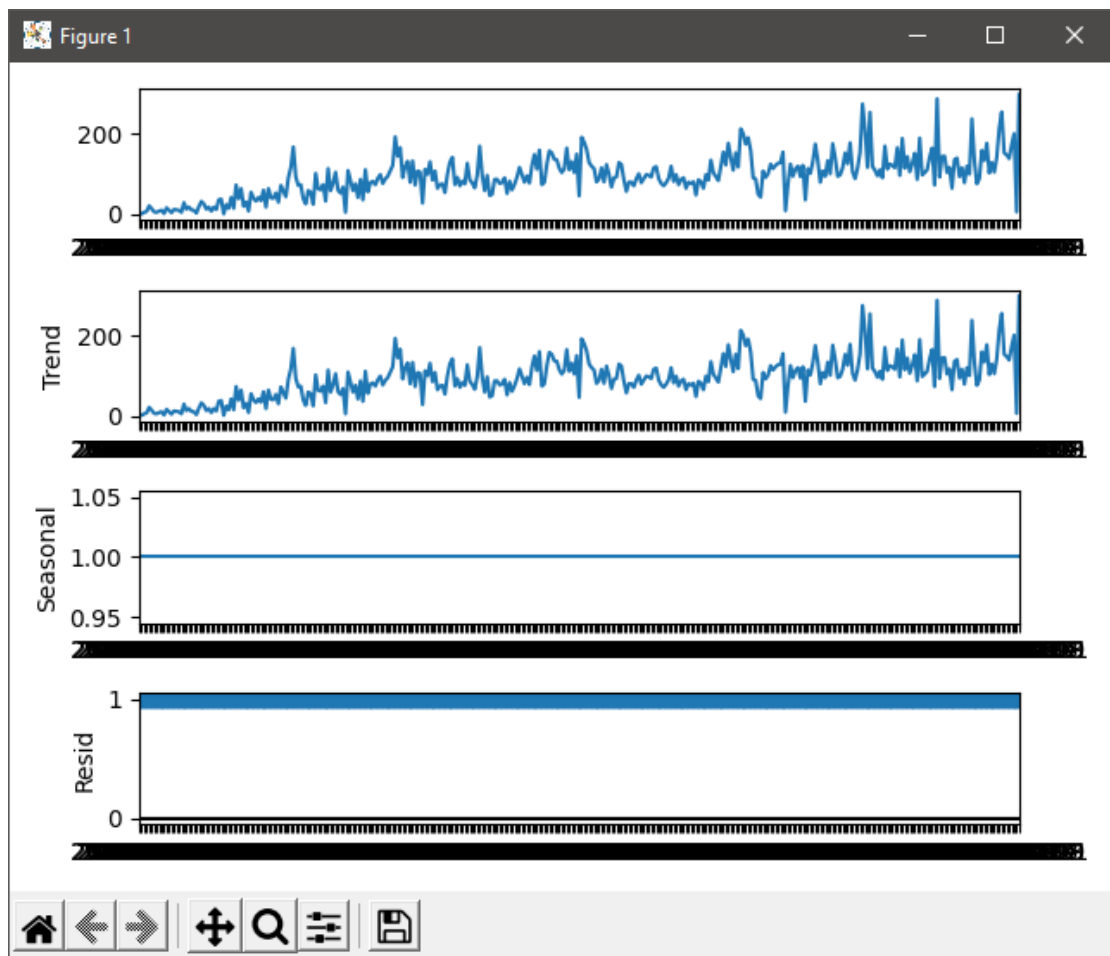
Για να γίνει ο υπολογισμός της εποχικότητας του μεγέθους «πλήθος ημερήσιων εισιτηρίων» θα πρέπει να υπάρχουν επαρκή δεδομένα, για περισσότερα από ένα έτη, ώστε να είναι δυνατή η σύγκριση των περιόδων μεταξύ τους.

Με βάση τα δεδομένα του dataset Εισιτήρια 2018, Εισιτήρια 2019 με θερινές τιμές_1, Εισιτήρια 2019 με θερινές τιμές_2, Εισιτήρια 2019 με χειμερινές τιμές, Εισιτήρια 2020, προκύπτουν τα εξής:

1. Τα δεδομένα είναι διαθέσιμα μόνο για τους μήνες 7^ο έως 12^ο του 2018, 1^ο έως και 12^ο του 2019 και 1^ο έως 4^ο του 2020, πράγμα το οποίο δεν επιτρέπει τη σύγκριση όμοιων περιόδων μεταξύ των τριών ετών.
2. Ενδεικτικά, για το λιμάνι της Νάξου (κωδικός λιμανιού: JNX) προκύπτουν τα ακόλουθα αποτελέσματα για τις ημερήσιες αφίξεις του 2019:

Ιαν	351
Φεβ	863
Μαρ	2060
Απρ	2812
Μαϊ	2750
Ιουν	3225
Ιουλ	3220
Αυγ	2868
Σεπ	3614
Οκτ	4061
Νοε	3869

Επίσης, βάσει των δεδομένων για το έτος 2019, η διακύμανση των ημερήσιων εισιτηρίων, φαίνεται παραστατικά στο επόμενο διάγραμμα:



Εικόνα 15. Διακύμανση εισιτηρίων 2019

Επίσης, το διάγραμμα αυτό προέκυψε από τη μελέτη της εποχικότητας του πλήθους των ημερήσιων εισιτηρίων του λιμανιού με κωδικό JNX (Νάξος). Επειδή όμως δεν υπάρχουν δεδομένα περισσότερων ετών ώστε να προκύψει η όποια περιοδικότητα στο πλήθος των αφίξεων στο συγκεκριμένο λιμάνι, τα μεγέθη “Seasonal” και “Resid” εμφανίζονται ως ευθείες γραμμές. Το μέγεθος “Trend” ακολουθεί επακριβώς την καμπύλη του πλήθους των ημερήσιων αφίξεων, δεδομένου ότι δεν υπάρχουν συγκριτικά στοιχεία άλλων ετών, ώστε να διαμορφωθεί η τάση.

Πόσο διαρκεί η τουριστική περίοδος ανά νησί;

Μελετώντας τα datasets, μπορούμε να θεωρήσουμε ότι η τουριστική περίοδος ενός νησιού προσδιορίζεται από την περίοδο κατά την οποία οι αφίξεις στο νησί εμφανίζουν μεγάλη άνοδο (απότομη άνοδο) σε σχέση με την προηγούμενη περίοδο, διατηρούνται σε ένα υψηλό επίπεδο και στο τέλος της τουριστικής περιόδου, εμφανίζουν απότομη πτώση.

A7. Ποια είναι η ποσοστιαία σχέση επιλέξιμου/αιτούμενου ποσού ανά ΚΑΔ και ανά νησί για κάθε εβδομάδα του έτους;

Για να γίνει ο υπολογισμός, απαιτείται το dataset «Αιτήσεις Χρηματοδότησης.xls». Στο dataset αρχικά δημιουργείται μία νέα στήλη, στην οποία υπολογίζεται η εβδομάδα και το έτος στην οποία ανήκει η αντίστοιχη αίτηση. Ο σκοπός είναι να μπορεί να υπολογιστεί το άθροισμα των αιτούμενων και επιλέξιμων ποσών ανά εβδομάδα.

Επειδή το συγκεκριμένο dataset δεν περιλαμβάνει στήλη με τον κύριο ΚΑΔ της επιχείρησης, αυτό γίνεται merge με το dataset «Επιχειρήσεις.xls», ώστε να δημιουργηθεί η στήλη με τον κύριο ΚΑΔ της αντίστοιχης επιχείρησης. Η συγχώνευση των δύο datasets γίνεται στη στήλη του Κωδικού της επιχείρησης. Προκύπτει τελικά ένα dataframe, το οποίο περιλαμβάνει όλες τις απαραίτητες, προκειμένου να είναι δυνατή η ομαδοποίηση και συνάθροιση των αιτούμενων και επιλέξιμων ποσών ανά ΚΑΔ, ανά νησί και εβδομάδα αίτησης.

Το αποτέλεσμα της συνάθροισης απεικονίζεται στον ακόλουθο πίνακα:

Ποσοστιαία σχέση επιλέξιμου/αιτούμενου ποσού ανά ΚΑΔ και ανά νησί για κάθε εβδομάδα του έτους	Κύριος ΚΑΔ - κωδικός	Νησί	Εβδομάδα αίτησης	Επιλέξιμο ποσό \	Αιτούμενη Χρηματοδότηση	ποσοστιαία σχέση επιλέξιμου/αιτούμενου ποσού
0	1000000	ΑΝΑΡΟΣ	2019-41	80.64	55.73	69.11
1	1000000	ΑΝΑΡΟΣ	2020-1	138.27	95.13	68.80
2	1000000	ΑΝΑΡΟΣ	2020-6	64.52	43.96	68.13
3	1000000	ΑΝΑΡΟΣ	2020-7	558.34	300.44	53.81
4	1000000	ΚΑΛΥΜΝΟΣ	2019-38	1140.40	725.30	63.60
...
16450	96091902	ΥΔΡΑ	2020-8	5434.82	3281.15	60.37
16451	96091904	ΜΗΛΟΣ	2019-41	1330.80	765.63	57.53
16452	96091904	ΜΗΛΟΣ	2020-7	128.00	70.46	55.05
16453	96091909	ΜΥΚΟΝΟΣ	2019-40	36.67	17.77	48.46
16454	96091916	ΚΩΣ	2020-7	1479.93	379.57	25.65

Εικόνα 16. Ποσοστιαία σχέση επιλέξιμου/αιτούμενου ποσού ανά ΚΑΔ και ανά νησί για κάθε εβδομάδα του έτους

Η ανάλυση αυτή δείχνει ότι το επιλέξιμο ποσό για χρηματοδότηση κυμαίνεται στην περιοχή του 60% του ποσού που αιτείται κάθε εταιρεία.

A8. Σε ποια νησιά παρατηρούνται χωρικές ομαδοποιήσεις συγγενικών ΚΑΔ επιχειρήσεων;

Βάσει της κωδικοποίησης των ΚΑΔ, τα δύο πρώτα ψηφία προσδιορίζουν τον Τομέα / Κλάδο στον οποίο δραστηριοποιείται η επιχείρηση. Για κάθε ΚΑΔ στον οποίο ανήκει η επιχείρηση, έγινε ομαδοποίηση σύμφωνα με τα δύο πρώτα ψηφία του ΚΑΔ και το νησί στο οποίο δραστηριοποιείται η επιχείρηση. Η συνάθροιση έγινε με καταμέτρηση του κωδικού της επιχείρησης, ώστε να προκύψει το σύνολο των επιχειρήσεων που ανήκουν σε κάθε Τομέα / Κλάδο οικονομικής δραστηριότητας σε κάθε νησί.

Για να εμφανίζεται ο τίτλος του Τομέα/Κλάδου που ανήκει κάθε ΚΑΔ, δημιουργήθηκε ένα νέο dataframe (kad_sectors), το οποίο περιέχει το πρώτο διψήφιο τμήμα των ΚΑΔ (το οποίο και αντιστοιχεί στον τομέα οικονομικής δραστηριότητας) και τον αντίστοιχο τίτλο του τομέα. Η μορφή αυτό του dataframe είναι η ακόλουθη:

Τομείς οικονομικής δραστηριότητας σύμφωνα με ΚΑΔ

Τομέας	Τίτλος
0 1	Φυτική και ζωική παραγωγή, θήρα και συναφείς δ...
1 2	Δασοκομία και υλοτομία
2 3	Αλιεία και υδατοκαλλιέργεια
3 5	Εξόρυξη άνθρακα και λιγνίτη
4 6	Άντληση αργού πετρελαίου και φυσικού αερίου
..
84 96	Άλλες δραστηριότητες παροχής προσωπικών υπηρεσιών
85 97	Δραστηριότητες νοικοκυριών ως εργοδοτών οικιακ...
86 98	Δραστηριότητες ιδιωτικών νοικοκυριών, που αφορ...
87 99	Δραστηριότητες εξωχώριων οργανισμών και φορέων
88 0	Έλλειψη δραστηριότητας

Εικόνα 17. Τομείς που ανήκει ο κάθε ΚΑΔ

Το αποτέλεσμα της αρχικής επεξεργασίας έγινε merge με το dataframe “kad_sectors”, ώστε η στήλη “Τίτλος” να ενσωματωθεί στο τελικό αποτέλεσμα.

Το αποτέλεσμα απεικονίζεται στον πίνακα που ακολουθεί και είναι αποτέλεσμα της εκτέλεσης των παραπάνω υπολογισμών.

Χωρικές ομαδοποιήσεις συγγενικών ΚΑΔ επιχειρήσεων	Τίτλος	Πλήθος επιχειρήσεων
Νησί που έχει έδρα η επιχείρηση Τομέας / Κλάδος		
0 ΑΓΑΘΟΝΗΣΙ 14	Κατασκευή ειδών ένδυσης	1
1 ΑΓΙΟΣ ΕΥΣΤΡΑΤΙΟΣ 14	Κατασκευή ειδών ένδυσης	1
2 ΑΙΓΙΝΑ 14	Κατασκευή ειδών ένδυσης	8
3 ΑΛΟΝΝΗΣΟΣ 14	Κατασκευή ειδών ένδυσης	2
4 ΑΜΟΡΓΟΣ 14	Κατασκευή ειδών ένδυσης	1
...
2615 ΜΗΛΟΣ 21	Παραγωγή βασικών φαρμακευτικών προϊόντων και φ...	1
2616 ΝΑΞΟΣ 21	Παραγωγή βασικών φαρμακευτικών προϊόντων και φ...	4
2617 ΡΟΔΟΣ 21	Παραγωγή βασικών φαρμακευτικών προϊόντων και φ...	1
2618 ΣΥΡΟΣ 21	Παραγωγή βασικών φαρμακευτικών προϊόντων και φ...	1
2619 ΧΙΟΣ 21	Παραγωγή βασικών φαρμακευτικών προϊόντων και φ...	1

Εικόνα 18. Χωρικές ομαδοποιήσεις συγγενικών ΚΑΔ επιχειρήσεων

Η ανάλυση αυτή κατέδειξε το πλήθος των επιχειρήσεων ανά νησί, οι οποίες ανήκουν στην ίδια ομάδα ΚΑΔ, η οποία προσδιορίζεται από τα δύο πρώτα ψηφία του ΚΑΔ. Ο ερευνητής μπορεί να εξειδικεύσει την ανάλυσή του σε συγκεκριμένο νησί, ώστε να διαμορφώσει την εικόνα της οικονομίας των συναφών ΚΑΔ.

A9. Ποια είναι η συνύπαρξη ΚΑΔ ανά νησί

A9.1 Ποια είναι η αναλογία των ΚΑΔ επιχειρήσεων σε μια περιοχή, σε σχέση με την κατανομή των ΚΑΔ στο σύνολο του Αιγαίου

Μία μελέτη του dataset των επιχειρήσεων καταλήγει στο ότι περιλαμβάνει δεδομένα και για νησιά εκτός του Αιγαίου. Για το λόγο αυτό εξαιρούμε τα νησιά που δεν ανήκουν στην περιοχή του Αιγαίου, εξαιρώντας από το dataframe τα νησιά: 'ΚΕΦΑΛΛΟΝΙΑ', 'ΚΕΡΚΥΡΑ', 'ΙΘΑΚΗ', 'ΖΑΚΥΝΘΟΣ', 'ΑΝΤΙΠΑΞΟΙ', 'ΜΕΓΑΝΗΣΙ', 'ΚΑΛΑΜΟΣ', 'ΤΡΙΖΟΝΙΑ'.

Η επεξεργασία αυτή καταλήγει στη δημιουργία ενός dataframe επιχειρήσεων μόνο των νησιών του Αιγαίου. Το dataframe αυτό ονομάζεται “aegeancompanies”.

Για κάθε επιχείρηση του dataframe aegeancompanies, δημιουργείται ένα νέο dataframe, στο οποίο αναλύονται οι επιμέρους ΚΑΔ στους οποίους ανήκει η επιχείρηση. Το dataframe αυτό

έχει για κάθε επιχείρηση τόσες γραμμές, όσες και οι ΚΑΔ που έχει. Το dataframe ονομάζεται “kadnorm”. Το dataframe kadnorm έχει δύο στήλες, τον κωδικό της επιχείρησης και τον ΚΑΔ.

Συναθροίζουμε το dataframe kadnorm ώστε να προκύψει το σύνολο των επιχειρήσεων του Αιγαίου ανά κωδικό ΚΑΔ.

Συγχωνεύουμε το dataframe των ΚΑΔ των επιχειρήσεων του Αιγαίου με τον πίνακα των επιχειρήσεων, ώστε στον πίνακα των ΚΑΔ να συμπεριληφθεί το νησί, και να γίνει έτσι η καταμέτρηση των ΚΑΔ ανά νησί.

Συγχωνεύουμε τον πίνακα των ΚΑΔ των επιχειρήσεων του Αιγαίου με τον πίνακα ο οποίος έχει υπολογίσει την καταμέτρηση ανά ΚΑΔ στο σύνολο του Αιγαίου (aegeankad).

Ο πίνακας αυτός περιέχει σε κάθε γραμμή το σύνολο των επιχειρήσεων του Αιγαίου για τον ΚΑΔ της αντίστοιχης γραμμής και τον κωδικό της επιχείρησης.

Το αποτέλεσμα της επεξεργασίας φαίνεται στον παρακάτω πίνακα.

Αναλογία των ΚΑΔ επιχειρήσεων Αιγαίου σε σχέση με την κατανομή των ΚΑΔ στο σύνολο του Αιγαίου						
	Νησί που έχει έδρα η επιχείρηση	ΚΑΔ	Επιχειρήσεις Αιγαίου	\	Πλήθος επιχειρήσεων	Ποσοστό
0	ΑΓΑΘΟΝΗΣΙ	1491000	4		1	25.00
1	ΑΓΑΘΟΝΗΣΙ	3109000	32		1	3.12
2	ΑΓΑΘΟΝΗΣΙ	33121100	10		1	10.00
3	ΑΓΑΘΟΝΗΣΙ	33122100	50		1	2.00
4	ΑΓΑΘΟΝΗΣΙ	33141101	13		1	7.69
...
41209	ΨΕΡΙΜΟΣ	47788904	1131		1	0.09
41210	ΨΕΡΙΜΟΣ	47996900	5		1	20.00
41211	ΨΕΡΙΜΟΣ	56101104	458		2	0.44
41212	ΨΕΡΙΜΟΣ	56101105	281		2	0.71
41213	ΨΕΡΙΜΟΣ	56101109	208		1	0.48

Εικόνα 19. Αναλογία των ΚΑΔ επιχειρήσεων σε μια περιοχή, σε σχέση με την κατανομή των ΚΑΔ στο σύνολο του Αιγαίου

Η ανάλυση αυτή μπορεί να αναδείξει τα νησιά στα οποία υπάρχει μεγάλη συγκέντρωση κάποιου ή κάποιων ΚΑΔ, σε σχέση με το σύνολο των επιχειρήσεων του ίδιου ΚΑΔ στο Αιγαίο. Τα αποτελέσματα αυτά μπορούν να χρησιμοποιηθούν ώστε να αναπτυχθούν στοχευμένες πολιτικές και δράσεις σε συγκεκριμένο νησιωτικό χώρο, ώστε να στηριχτούν οι επιχειρήσεις αυτών των κλάδων αλλά και να δημιουργηθούν συνέργειες επιχειρήσεων που βρίσκονται χωρικά κοντά.

A9.2 Dissimilarity Index: Ποια είναι η αναλογία κάθε ΚΑΔ επιχειρήσεων ξεχωριστά σε κάθε νησί προς το συνολικό πλήθος επιχειρήσεων στο νησί;

Μία μελέτη του dataset των επιχειρήσεων καταλήγει στο ότι περιλαμβάνει δεδομένα και για νησιά εκτός του Αιγαίου. Για το λόγο αυτό εξαιρούμε τα νησιά που δεν ανήκουν στην περιοχή του Αιγαίου, εξαιρώντας από το dataframe τα νησιά: 'ΚΕΦΑΛΛΟΝΙΑ', 'ΚΕΡΚΥΡΑ', 'ΙΘΑΚΗ', 'ΖΑΚΥΝΘΟΣ', 'ΑΝΤΙΠΑΞΟΙ', 'ΜΕΓΑΝΗΣΙ', 'ΚΑΛΑΜΟΣ', 'ΤΡΙΖΟΝΙΑ'.

Η επεξεργασία αυτή καταλήγει στη δημιουργία ενός dataframe επιχειρήσεων μόνο των νησιών του Αιγαίου. Το dataframe αυτό ονομάζεται “aegeancompanies”.

Για κάθε επιχείρηση του dataframe aegeancompanies, δημιουργείται ένα νέο dataframe, στο οποίο αναλύονται οι επιμέρους ΚΑΔ στους οποίους ανήκει η επιχείρηση. Το dataframe αυτό

έχει για κάθε επιχείρηση τόσες γραμμές, όσες και οι ΚΑΔ που έχει. Το dataframe ονομάζεται “kadnorm”. Το dataframe kadnorm έχει δύο στήλες, τον κωδικό της επιχείρησης και τον ΚΑΔ.

Συναθροίζουμε τον πίνακα των επιχειρήσεων του Αιγαίου (aegeancompanies), ώστε να προκύψει το σύνολο των επιχειρήσεων ανά νησί του Αιγαίου. Το dataframe στο οποίο καταλήγει ο υπολογισμός ονομάστηκε “companies_per_island”.

Συγχωνεύουμε το dataframe των ΚΑΔ των επιχειρήσεων του Αιγαίου με τον πίνακα των επιχειρήσεων, ώστε στον πίνακα των ΚΑΔ να συμπεριληφθεί το νησί, και να γίνει έτσι η καταμέτρηση των ΚΑΔ ανά νησί. Το dataframe με το αποτέλεσμα του υπολογισμού ονομάστηκε “aegean_island_kad”. Ο πίνακας aegean_island_kad συναθροίστηκε, ώστε να προκύψει το σύνολο των επιχειρήσεων ανά νησί και ανά ΚΑΔ.

Στη συνέχεια γίνεται συγχώνευση των πινάκων companies_per_kad_per_island και aegean_island_kad, ώστε στον πίνακα companies_per_kad_per_island να προστεθεί μία στήλη με το πλήθος των επιχειρήσεων ανά νησί. Επίσης, στο αποτέλεσμα προστίθεται μία επιπλέον στήλη στην οποία υπολογίζεται το ποσοστό των επιχειρήσεων που ανήκει σε έναν ΚΑΔ ανά νησί προς το σύνολο των επιχειρήσεων του νησιού.

Το αποτέλεσμα της επεξεργασίας εμφανίζεται στον επόμενο πίνακα.

Dissimilarity Index: Αναλογία κάθε ΚΑΔ επιχειρήσεων σε κάθε νησί προς τον συνολικό πλήθος επιχειρήσεων στο νησί

Νησί που έχει έδρα η επιχείρηση	ΚΑΔ	Πλήθος επιχειρήσεων ανά ΚΑΔ στο νησί	Πλήθος επιχειρήσεων νησιού	ποσοστό
0	ΑΓΑΘΟΝΗΣΙ	1491000	1	3 0.33
1	ΑΓΑΘΟΝΗΣΙ	31090000	1	3 0.33
2	ΑΓΑΘΟΝΗΣΙ	33121100	1	3 0.33
3	ΑΓΑΘΟΝΗΣΙ	33122100	1	3 0.33
4	ΑΓΑΘΟΝΗΣΙ	33141101	1	3 0.33
...
41209	ΨΕΡΙΜΟΣ	47788904	1	5 0.20
41210	ΨΕΡΙΜΟΣ	47996900	1	5 0.20
41211	ΨΕΡΙΜΟΣ	56101104	2	5 0.40
41212	ΨΕΡΙΜΟΣ	56101105	2	5 0.40
41213	ΨΕΡΙΜΟΣ	56101109	1	5 0.20

Εικόνα 20. Αναλογία κάθε ΚΑΔ επιχειρήσεων σε κάθε νησί προς το συνολικό πλήθος επιχειρήσεων στο νησί

Η ανάλυση αυτή δημιουργεί μία καθαρή εικόνα του είδους των επιχειρήσεων που αναπτύσσονται σε κάθε νησί. Μπορούν μέσω αυτών των μετρήσεων να προσδιοριστούν οι λόγοι που αναπτύσσεται ένας συγκεκριμένος κλάδος επιχειρήσεων σε ένα νησί και να προκύψουν πολιτικές για τη στήριξή του.

A9.3 Interaction Index: Ποια είναι η πιθανότητα εμφάνισης κάθε ΚΑΔ επιχειρήσεων να συνυπάρχει σε μια περιοχή με άλλους τύπους επιχειρήσεων ανά ΚΑΔ επιχειρήσεων;

Για να υπολογίσουμε την πιθανότητα συνύπαρξης ενός ΚΑΔ με τους υπόλοιπους ΚΑΔ ενός νησιού, πρέπει να μετρήσουμε τις επιχειρήσεις οι οποίες ανήκουν σε κάθε έναν ΚΑΔ και το πλήθος των διαφορετικών ΚΑΔ που εμφανίζονται στο κάθε νησί.

Συγχωνεύουμε το dataframe των ΚΑΔ των επιχειρήσεων του Αιγαίου με τον πίνακα των επιχειρήσεων, ώστε στον πίνακα των ΚΑΔ να συμπεριληφθεί το νησί, και να γίνει έτσι η καταμέτρηση των ΚΑΔ ανά νησί. Το dataframe με το αποτέλεσμα του υπολογισμού ονομάστηκε “ island_kad”. Ο πίνακας island_kad συναθροίστηκε, ώστε να προκύψει το σύνολο των επιχειρήσεων ανά νησί και ανά ΚΑΔ.

Στη συνέχεια δημιουργήθηκε ένας νέος πίνακας, ο οποίος καταμετρά το πλήθος των διαφορετικών ΚΑΔ ανά νησί. Ο πίνακας αυτός ονομάστηκε “companies_per_kad_per_island”.

Ο πίνακας “companies_per_kad_per_island” συγχωνεύθηκε με τον πίνακα island_kad, ώστε σε κάθε γραμμή να περιλαμβάνεται το όνομα του νησιού, ο ΚΑΔ, το πλήθος των επιχειρήσεων ανά ΚΑΔ στο κάθε νησί και το πλήθος των διαφορετικών ΚΑΔ ανά νησί.

Στο αποτέλεσμα προστέθηκε μία επιπλέον στήλη, η οποία προκύπτει από τη διαίρεση του πλήθους των επιχειρήσεων ανά ΚΑΔ διά το πλήθος των διαφορετικών ΚΑΔ στο νησί, ώστε να προκύψει η πιθανότητα συνύπαρξης ενός ΚΑΔ με τους υπόλοιπους στο νησί.

Το αποτέλεσμα της επεξεργασίας εμφανίζεται στον επόμενο πίνακα.

Interaction Index					
	Νησί που έχει έδρα η επιχείρηση	ΚΑΔ	Πλήθος επιχειρήσεων ανά ΚΑΔ στο νησί	Πλήθος διαφορετικών ΚΑΔ	Interaction index
0	ΑΓΑΘΟΝΗΣΙ	1491000	1	23	0.04
1	ΑΓΑΘΟΝΗΣΙ	31090000	1	23	0.04
2	ΑΓΑΘΟΝΗΣΙ	33121100	1	23	0.04
3	ΑΓΑΘΟΝΗΣΙ	33122100	1	23	0.04
4	ΑΓΑΘΟΝΗΣΙ	33141101	1	23	0.04
...
46019	ΨΕΡΙΜΟΣ	47788904	1	8	0.12
46020	ΨΕΡΙΜΟΣ	47996900	1	8	0.12
46021	ΨΕΡΙΜΟΣ	56101104	2	8	0.25
46022	ΨΕΡΙΜΟΣ	56101105	2	8	0.25
46023	ΨΕΡΙΜΟΣ	56101109	1	8	0.12

Εικόνα 21. Πιθανότητα εμφάνισης κάθε ΚΑΔ επιχειρήσεων να συνυπάρχει σε μια περιοχή με άλλους τύπους επιχειρήσεων ανά ΚΑΔ επιχειρήσεων

Το συγκεκριμένο μέγεθος απεικονίζει το ποσοστό των επιχειρήσεων που ανήκουν σε έναν συγκεκριμένο ΚΑΔ στο πλήθος των διαφορετικών ΚΑΔ που δραστηριοποιούνται στο κάθε νησί. Το μέγεθος αυτό εκφράζει την πιθανότητα να δημιουργηθεί μία νέα επιχείρηση του συγκεκριμένου ΚΑΔ στο νησί που εξετάζεται.

A9.4 Isolation Index: Ποια είναι η πιθανότητα εμφάνισης ενός ΚΑΔ επιχειρήσεων σε μια περιοχή που υπάρχουν και άλλοι ίδιοι ΚΑΔ, ανά νησί;

Για να υπολογίσουμε την πιθανότητα εμφάνισης ενός ΚΑΔ σε μία περιοχή με άλλους όμοιους ΚΑΔ, θα διαιρούμε το πλήθος των επιχειρήσεων που έχουν τον ίδιο ΚΑΔ δια το σύνολο των επιχειρήσεων στο νησί.

Για να γίνει αυτό, πρώτα υπολογίζουμε το πλήθος των επιχειρήσεων ανά νησί. Ομαδοποιούμε το dataframe “companies” ως προς τη στήλη “Νησί που έχει έδρα η επιχείρηση”, εφαρμόζοντας τη συνάρτηση count στη στήλη “Κωδικός της επιχείρησης”, ώστε να γίνει η καταμέτρηση. Προκύπτει ένα νέο dataframe με όνομα “companies_per_island”, το οποίο έχει δύο στήλες (“Νησί που έχει έδρα η επιχείρηση” και “Σύνολο επιχειρήσεων στο νησί”).

Το επόμενο βήμα περιλαμβάνει την καταμέτρηση των κωδικών επιχειρήσεων οι οποίες ανήκουν στον ίδιο ΚΑΔ (ανά νησί) και την απόρριψη εκείνων για τις οποίες το πλήθος των επιχειρήσεων στον ίδιο ΚΑΔ είναι μόνο 1. Αρχικά συγχωνεύουμε τον πίνακα των ΚΑΔ των

επιχειρήσεων (“kadnorm”) με τον πίνακα “companies”, ώστε να συμπεριλάβουμε στον πίνακα “kadnorm” τη στήλη “Νησί που έχει έδρα η επιχείρηση”. Το αποτέλεσμα περιέχεται στον πίνακα με όνομα “companies_per_kad”. Επί αυτού του πίνακα γίνεται ομαδοποίηση ανά “Νησί που έχει έδρα η επιχείρηση” και “ΚΑΔ”. Το αποτέλεσμα της επεξεργασίας είναι ένας νέος πίνακας με τρεις (3) στήλες (“Νησί που έχει έδρα η επιχείρηση”, “ΚΑΔ”, “Πλήθος επιχειρήσεων στον ίδιο ΚΑΔ”).

Από τον πίνακα αυτόν, εξαιρούμε τις γραμμές για τις οποίες η στήλη “Πλήθος επιχειρήσεων στον ίδιο ΚΑΔ” έχει την τιμή 1. Αυτό σημαίνει ότι δεν υπάρχουν άλλες επιχειρήσεις με τον ίδιο ΚΑΔ στο ίδιο νησί. Το αποτέλεσμα της επεξεργασίας καταχωρείται στον πίνακα “multiple_kad”.

Τελευταίο βήμα είναι η συγχώνευση των δύο πινάκων, με κοινό στοιχείο τη στήλη “Νησί που έχει έδρα η επιχείρηση”, ώστε να προκύψει ένας νέος πίνακας με στήλες: “Νησί που έχει έδρα η επιχείρηση”, “ΚΑΔ”, “Πλήθος επιχειρήσεων στον ίδιο ΚΑΔ”, “Σύνολο επιχειρήσεων στο νησί”, “Πιθανότητα εμφάνισης ΚΑΔ με όμοιους ΚΑΔ”.

Η τελευταία στήλη προκύπτει διαιρώντας τη στήλη “Πλήθος επιχειρήσεων στον ίδιο ΚΑΔ” με τη στήλη “Σύνολο επιχειρήσεων στο νησί”. Το αποτέλεσμα της επεξεργασίας φαίνεται στην παρακάτω εικόνα.

Νησί που έχει έδρα η επιχείρηση	ΚΑΔ	Πλήθος επιχειρήσεων στον ίδιο ΚΑΔ	Σύνολο επιχειρήσεων στο νησί	Πιθανότητα εμφάνισης ΚΑΔ με όμοιους ΚΑΔ
0	ΑΓΑΘΟΝΗΣΙ 47111005	2	3	0.67
1	ΑΓΚΙΣΤΡΙ 43111000	2	9	0.22
2	ΑΓΚΙΣΤΡΙ 43121200	2	9	0.22
3	ΑΓΚΙΣΤΡΙ 46211400	2	9	0.22
4	ΑΓΚΙΣΤΡΙ 46731600	2	9	0.22
...
19171	ΨΑΡΑ 56301004	3	12	0.25
19172	ΨΑΡΑ 56301007	2	12	0.17
19173	ΨΑΡΑ 56301014	2	12	0.17
19174	ΨΕΡΙΝΟΣ 56101104	2	5	0.40
19175	ΨΕΡΙΝΟΣ 56101105	2	5	0.40

Εικόνα 22. Πιθανότητα εμφάνισης ενός ΚΑΔ επιχειρήσεων σε μια περιοχή που υπάρχουν και άλλοι ίδιοι ΚΑΔ, ανά νησί

Εφαρμόζοντας την ανάλυση αυτή, προκύπτει για κάθε νησί η πιθανότητα να εμφανιστούν ομοειδείς επιχειρήσεις (επιχειρήσεις στον ίδιο ΚΑΔ) στο σύνολο των επιχειρήσεων που δραστηριοποιούνται στο νησί. Στην περίπτωση του Αγαθονησίου, τα 2/3 των επιχειρήσεων (ποσοστό 67%) ανήκουν στον ΚΑΔ 47111005.

(B) Ερωτήματα για Φορτωτικές

B1. Ποιες είναι οι κυρίαρχες περιοχές που στέλνουν φορτωτικές σε κάθε νησί, ανά ΚΑΔ και ανά συνολική αξία;

Για τον προσδιορισμό των περιοχών που αποστέλλουν εμπορεύματα σε κάποιο νησί, χρησιμοποιήθηκε το αρχείο Φορτωτικες.xlsx, από το οποίο δημιουργήθηκε ένα dataframe με τις στήλες: 'Κωδικός Επιχείρησης', 'Τόπος Αναχώρησης', 'Αξία με ΦΠΑ', οι οποίες είναι οι απαραίτητες για την εξαγωγή του αποτελέσματος.

Το αποτέλεσμα (dataframe “fortotikes”) συγχωνεύθηκε με το dataframe “companies”, ώστε να συμπεριληφθούν οι στήλες 'Νησί που έχει έδρα η επιχείρηση', 'Κύριος ΚΑΔ – κωδικός'. Το αποτέλεσμα ομαδοποιήθηκε ως προς τις στήλες 'Τόπος Αναχώρησης', 'Νησί που έχει έδρα η επιχείρηση', 'Κύριος ΚΑΔ – κωδικός', ώστε να προκύψουν οι περιοχές οι οποίες στέλνουν εμπορεύματα ανά νησί και ανά κύριο ΚΑΔ επιχείρησης. Η στήλη συνάθροισης (άθροισμα)

είναι η στήλη 'Αξία με ΦΠΑ', η οποία προσδιορίζει την αξία των εμπορευμάτων που αποστέλλονται από κάθε περιοχή.

Το αποτέλεσμα απεικονίζεται στον επόμενο πίνακα:

Περιοχές αποστολής εμπορευμάτων προς νησί ανά νησί έδρας και ΚΑΔ				
	Τόπος Αναχώρησης Νησί που έχει έδρα η επιχείρηση	\	Κύριος ΚΑΔ - κωδικός	Αξία με ΦΠΑ
0	ΑΙΤΩΛΟΑΚΑΡΝΑΝΙΑΣ (Π.Ε.)	ΜΥΤΙΛΗΝΗ	10514000	18.60
1	ΑΙΤΩΛΟΑΚΑΡΝΑΝΙΑΣ (Π.Ε.)	ΜΥΤΙΛΗΝΗ	13922202	149.61
2	ΑΙΤΩΛΟΑΚΑΡΝΑΝΙΑΣ (Π.Ε.)	ΜΥΤΙΛΗΝΗ	46731200	1336.08
3	ΑΙΤΩΛΟΑΚΑΡΝΑΝΙΑΣ (Π.Ε.)	ΧΙΟΣ	47545438	8.93
4	ΑΡΓΟΛΙΔΑΣ (Π.Ε.)	ΧΙΟΣ	10321000	23873.75
..
261	ΧΙΟΣ	ΜΥΤΙΛΗΝΗ	10514000	6.39
262	ΧΙΟΣ	ΧΙΟΣ	10321000	21742.32
263	ΧΙΟΣ	ΧΙΟΣ	10392202	36782.97
264	ΧΙΟΣ	ΧΙΟΣ	47524115	43.65
265	ΧΙΟΣ	ΧΙΟΣ	56101113	201.95

Εικόνα 23. Περιοχές αποστολής εμπορευμάτων προς νησί ανά νησί έδρας και ΚΑΔ

Αναλύοντας το αποτέλεσμα, προκύπτουν οι περιοχές από τις οποίες φορτωτικές στάλθηκαν προς νησιά και παραλήφθηκαν από επιχειρήσεις συγκεκριμένου ΚΑΔ. Στην περίπτωση της Μυτιλήνης, διαπιστώνουμε ότι μεταξύ άλλων αποστολές εμπορευμάτων έγιναν από την Αιτωλοακαρνανία (Π.Ε.) και τη Χίο, ενώ παραλήπτες ήταν οι ΚΑΔ 10514000, 13922202, 46731200.

B2. Ποιες είναι οι κυρίαρχες περιοχές που λαμβάνουν φορτωτικές από κάθε νησί, ανά ΚΑΔ και ανά συνολική αξία;

Για τον προσδιορισμό των περιοχών που αποστέλλονται εμπορεύματα από κάποιο νησί, χρησιμοποιήθηκε το αρχείο `Fortotikes.xlsx`, από το οποίο δημιουργήθηκε ένα dataframe με τις στήλες: 'Κωδικός Επιχείρησης', 'Τόπος Αναχώρησης', 'Τόπος Προορισμού', 'Αξία με ΦΠΑ', οι οποίες είναι οι απαραίτητες για την εξαγωγή του αποτελέσματος.

Μετά έγινε εξαγωγή των μοναδικών τιμών της στήλης 'Τόπος Αναχώρησης'. Οι τιμές αυτές στη συνέχεια επεξεργάστηκαν, ώστε να διαγραφούν εκείνοι οι τόποι αναχώρησης οι οποίοι δεν είναι νησιά. Το αποτέλεσμα αποθηκεύεται σε ένα .csv αρχείο (`shipping_areas.csv`) και αυτό φορτώνεται στη συνέχεια από το πρόγραμμα.

Το αποτέλεσμα (dataframe "fortotikes") συγχωνεύθηκε με το dataframe "companies", ώστε να συμπεριληφθούν οι στήλες 'Νησί που έχει έδρα η επιχείρηση', 'Κύριος ΚΑΔ - κωδικός'.

Από το αποτέλεσμα της συγχώνευσης, προέκυψε ο πίνακας "fortotikesMerged", από τον οποίο διαγράφονται οι γραμμές εκείνες στις οποίες η τιμή της στήλης "Τόπος Αναχώρησης" δεν ανήκει στις τιμές που περιέχονται στο αρχείο `shipping_areas.csv`. Αυτό σημαίνει ότι από την περαιτέρω επεξεργασία εξαιρούνται οι φορτωτικές που δεν έχουν ως τόπο αποστολής κάποιο νησί.

Το αποτέλεσμα ομαδοποιήθηκε ως προς τις στήλες 'Κύριος ΚΑΔ - κωδικός', 'Τόπος Προορισμού', ώστε να προκύψουν τα νησιά τα οποία στέλνουν εμπορεύματα σε άλλες περιοχές ανά κύριο ΚΑΔ επιχείρησης. Η στήλη συνάθροισης (άθροισμα) είναι η στήλη 'Αξία με ΦΠΑ', η οποία προσδιορίζει την αξία των εμπορευμάτων που αποστέλλονται από κάθε νησί.

Το αποτέλεσμα της επεξεργασίας απεικονίζεται στον επόμενο πίνακα:

Περιοχές που λαμβάνουν φορτωτικές από κάθε νησί, ανά ΚΑΔ και ανά συνολική αξία				
	Τόπος Αναχώρησης	Κύριος ΚΑΔ - κωδικός	Τόπος Προορισμού	Αξία με ΦΠΑ
0	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	10412302	ΣΑΜΟΣ	186.19
1	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	33151001	ΧΙΟΣ	48.80
2	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	46731200	ΜΥΤΙΛΗΝΗ	1003.19
3	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	47595500	ΚΩΣ	499.17
4	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	55101003	ΚΩΣ	9.10
..
68	ΧΙΟΣ	10392202	ΑΤΤΙΚΗΣ (ΠΕΡΙΦΕΡΕΙΑ)	36782.97
69	ΧΙΟΣ	10514000	ΜΥΤΙΛΗΝΗ	6.39
70	ΧΙΟΣ	47524115	ΑΤΤΙΚΗΣ (ΠΕΡΙΦΕΡΕΙΑ)	43.65
71	ΧΙΟΣ	56101113	ΑΤΤΙΚΗΣ (ΠΕΡΙΦΕΡΕΙΑ)	3.17
72	ΧΙΟΣ	56101113	ΘΕΣΣΑΛΟΝΙΚΗ (Π.Ε.)	198.78

Εικόνα 24. Περιοχές που λαμβάνουν φορτωτικές από κάθε νησί ανά ΚΑΔ και ανά συνολική αξία

Η ανάλυση αυτή συμπεραίνει ποιες είναι οι σημαντικότερες περιοχές που λαμβάνουν εμπορεύματα ανά νησί και αντικείμενο της επιχείρησης που αποστέλλει τα προϊόντα της. Ενδεικτικά παρατηρούμε ότι για την περίπτωση της Χίου, η Περιφέρεια Αττικής αποτελεί σημαντικό προορισμό των προϊόντων του ΚΑΔ 10392202.

B3. Ποιο είναι το δίκτυο (spatial network) logistics που αναδύεται από τις φορτωτικές ανά νησί;

Υπολογίστηκαν οι περιοχές οι οποίες δέχονται εμπορεύματα από κάθε νησί. Αυτός ο υπολογισμός βασίστηκε στο αποτέλεσμα του προηγούμενου ερωτήματος (B2), από το οποίο διαγράφηκαν όλες οι στήλες εκτός από τις 'Τόπος Αναχώρησης' και 'Τόπος Προορισμού'.

Στη συνέχεια ομαδοποιήθηκαν οι στήλες του πίνακα ως προς τον 'Τόπο Αναχώρησης' και τον 'Τόπο Προορισμού'. Το αποτέλεσμα περιλαμβάνει το όνομα του εκάστοτε νησιού και τα ονόματα των περιοχών στις οποίες αποστέλλει εμπορεύματα. Η εικόνα του αποτελέσματος είναι η ακόλουθη:

	Τόπος Αναχώρησης	Τόπος Προορισμού
0	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	ΚΩΣ
1	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	ΜΥΤΙΛΗΝΗ
2	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	ΣΑΜΟΣ
3	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	ΧΙΟΣ
4	ΚΑΛΥΜΝΟΣ	ΑΝΑΦΗ
5	ΚΑΛΥΜΝΟΣ	ΑΤΤΙΚΗΣ (ΠΕΡΙΦΕΡΕΙΑ)
6	ΚΑΛΥΜΝΟΣ	ΛΕΙΨΟΙ
7	ΚΑΛΥΜΝΟΣ	ΠΑΤΜΟΣ
8	ΚΑΛΥΜΝΟΣ	ΡΟΔΟΣ

Εικόνα 25. Τόπος αναχώρησης και προορισμού

Στην ανάλυση αυτή δημιουργήθηκαν τα ζεύγη περιοχών αποστολής και παραλαβής εμπορευμάτων. Δημιουργείται έτσι μία σχέση «συνεργασίας» των δύο περιοχών, ώστε να προκύψουν οι εμπορικές διασυνδέσεις μεταξύ τους. Ενδεικτικά και με βάση τα διαθέσιμα

δεδομένα, οι φορτωτικές του Ηρακλείου (Π.Ε.) αφορούσαν μόνο τα νησιά Κω, Μυτιλήνη, Σάμο και Χίο.

Εμφανίζονται μήπως ακραίες τιμές σε ορισμένα νησιά, στη σχέση απόστασης, κόστους, πλήθους φορτωτικών;

Ακραίες τιμές στη σχέση απόστασης

Για να βρεθούν ακραίες τιμές (outliers), σε σχέση με την απόσταση των προορισμών από κάθε νησί, στον υπολογισμό του προηγούμενου ερωτήματος, θα προστεθεί η στήλη της απόστασης του νησιού που αποστέλλει φορτωτικές από την περιοχή που παραλαμβάνει τα προϊόντα της φορτωτικής.

Στο dataframe 'fortotikes' προσθέτουμε τη στήλη 'Απόσταση', κατά την ανάγνωση του αρχείου εισόδου ('Φορτωτικές.xlsx'). Η στήλη 'Απόσταση' χρησιμοποιήθηκε στην ομαδοποίηση των φορτωτικών που εκδόθηκαν από κάθε νησί, ώστε το αποτέλεσμα να περιλαμβάνει και την απόσταση του νησιού από την περιοχή που παραλαμβάνονται τα εμπορεύματα. Το αποτέλεσμα της ομαδοποίησης ταξινομήθηκε κατά αύξουσα αλφαβητική σειρά νησιού αποστολής και φθίνουσα σειρά απόστασης (εμφανίζεται πρώτος ο προορισμός που απέχει τη μεγαλύτερη απόσταση από το νησί αποστολής). Το αποτέλεσμα εμφανίζεται στον επόμενο πίνακα:

Δίκτυο logistics

	Τύπος Αναχώρησης	Τύπος Προορισμού	Απόσταση
1	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	ΜΥΤΙΛΗΝΗ	370
2	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	ΣΑΜΟΣ	311
0	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	ΚΩΣ	267
3	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	ΧΙΟΣ	267
5	ΚΑΛΥΜΝΟΣ	ΑΤΤΙΚΗΣ (ΠΕΡΙΦΕΡΕΙΑ)	322
4	ΚΑΛΥΜΝΟΣ	ΑΝΑΦΗ	157
8	ΚΑΛΥΜΝΟΣ	ΡΟΔΟΣ	131
7	ΚΑΛΥΜΝΟΣ	ΠΑΤΜΟΣ	72
6	ΚΑΛΥΜΝΟΣ	ΛΕΙΨΟΙ	54

Εικόνα 26. Αποστάσεις logistics

Στη συνέχεια, υπολογίστηκε το zscore για τη στήλη 'Απόσταση' και αποδείχτηκε ότι καμία τιμή του z δεν είναι μεγαλύτερη από 3, συνεπώς το σύνολο των δεδομένων δεν περιλαμβάνει ακραίες τιμές ως προς την απόσταση.

Αυτό σημαίνει ότι οι αποστολές εμπορευμάτων έχουν παγιωθεί και δεν υπάρχουν αποστολές σε ιδιαίτερα απομακρυσμένες περιοχές. Οι αποστολές γίνονται προς προορισμούς εντός της Ελλάδα και στις συναλλαγές αυτές δεν περιλαμβάνονται εξαγωγές προς άλλες χώρες.

Ακραίες τιμές στη σχέση κόστους

Στο dataframe 'fortotikes' προσθέτουμε τη στήλη 'Αξία με ΦΠΑ', κατά την ανάγνωση του αρχείου εισόδου ('Φορτωτικές.xlsx'). Η στήλη 'Αξία με ΦΠΑ' χρησιμοποιήθηκε στην ομαδοποίηση των φορτωτικών που εκδόθηκαν από κάθε νησί, ώστε το αποτέλεσμα να

περιλαμβάνει το σύνολο της αξίας των φορτωτικών που κατευθύνονται από κάθε νησί σε κάθε προορισμό. Το αποτέλεσμα της ομαδοποίησης ταξινομήθηκε κατά αύξουσα αλφαβητική σειρά νησιού αποστολής και φθίνουσα σειρά «Αξία με ΦΠΑ» (εμφανίζεται πρώτος ο προορισμός που απέστειλε τη μεγαλύτερη συνολική αξία φορτωτικών).

Προσθέτουμε στο dataframe αυτό μία επιπλέον στήλη, στην οποία υπολογίζεται το z-score της αντίστοιχης τιμής της αξίας φορτωτικών, ώστε να μπορέσουν στη συνέχεια να απομονωθούν τα outliers.

Το αποτέλεσμα εμφανίζεται στον επόμενο πίνακα:

Δίκτυο logistics

	Τόπος Αναχώρησης	Τόπος Προορισμού	Αξία με ΦΠΑ	Z-SCORE
1	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	ΜΥΤΙΛΗΝΗ	1003.19	0.167686
0	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	ΚΩΣ	508.27	0.228941
2	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	ΣΑΜΟΣ	186.19	0.268805
3	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	ΧΙΟΣ	48.80	0.285809
5	ΚΑΛΥΜΝΟΣ	ΑΤΤΙΚΗΣ (ΠΕΡΙΦΕΡΕΙΑ)	806.00	0.192091
4	ΚΑΛΥΜΝΟΣ	ΑΝΑΦΗ	632.40	0.213578
8	ΚΑΛΥΜΝΟΣ	ΡΟΔΟΣ	152.00	0.273036
6	ΚΑΛΥΜΝΟΣ	ΛΕΙΨΟΙ	24.00	0.288879
7	ΚΑΛΥΜΝΟΣ	ΠΑΤΜΟΣ	3.00	0.291478
9	ΚΩΣ	ΑΤΤΙΚΗΣ (ΠΕΡΙΦΕΡΕΙΑ)	4376.11	0.249776
12	ΚΩΣ	ΝΙΣΥΡΟΣ	593.00	0.218454
14	ΚΩΣ	ΡΟΔΟΣ	74.40	0.282641
13	ΚΩΣ	ΠΑΤΜΟΣ	42.00	0.286651

Εικόνα 27. Παράδειγμα εύρεσης ακραίων τιμών (outliers)

Ο παραπάνω πίνακας φιλτραρίστηκε, ώστε να παραμείνουν μόνο οι γραμμές για τις οποίες το z-score είναι μεγαλύτερο ή ίσο της τιμής 3. Το αποτέλεσμα δείχνει ότι ακραία τιμή για την αξία των φορτωτικών αποτελεί μόνο η διαδρομή Χίος προς Αττική (Περιφέρεια), όπως φαίνεται στον παρακάτω πίνακα.

Outliers συνολικής αξίας φορτωτικών

	Τόπος Αναχώρησης	Τόπος Προορισμού	Αξία με ΦΠΑ	Z-SCORE
42	ΧΙΟΣ	ΑΤΤΙΚΗΣ (ΠΕΡΙΦΕΡΕΙΑ)	54750.16	6.484504

Εικόνα 28. Παράδειγμα εντοπισμού ακραίας τιμής (outlier)

Όσον αφορά την αξία των φορτωτικών, παρατηρούμε ότι οι συνολικές αξίες των φορτωτικών μεταξύ προορισμών κινούνται σε ένα εύρος το οποίο δεν εμφανίζει τιμές που να βρίσκονται υπερβολικά έξω από τα όρια αυτού του συνόλου. Η έρευνα κατέδειξε ότι η μοναδική περίπτωση που η αξία των φορτωτικών βρίσκεται υπερβολικά εκτός ορίων είναι στις αποστολές εμπορευμάτων από τη Χίο στην Αττική (Περιφέρεια).

Ακραίες τιμές στη σχέση πλήθους φορτωτικών

Το dataframe 'fortotikes' ομαδοποιήθηκε ανά νησί αποστολής και τόπο προορισμού και για κάθε ομάδα υπολογίστηκε το πλήθος των φορτωτικών που καταγράφηκαν. Το αποτέλεσμα της ομαδοποίησης ταξινομήθηκε κατά αύξουσα αλφαβητική σειρά νησιού αποστολής και φθίνουσα σειρά πλήθους φορτωτικών (εμφανίζεται πρώτος ο προορισμός που δέχτηκε το μεγαλύτερο πλήθος φορτωτικών από το εκάστοτε νησί).

Προσθέτουμε στο dataframe αυτό μία επιπλέον στήλη, στην οποία υπολογίζεται το z-score της αντίστοιχης τιμής του πλήθους φορτωτικών, ώστε να μπορέσουν στη συνέχεια να απομονωθούν τα outliers.

Το αποτέλεσμα εμφανίζεται στον επόμενο πίνακα:

Δίκτυο logistics

	Τόπος Αναχώρησης	Τόπος Προορισμού	Πλήθος φορτωτικών	Z-score
0	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	ΚΩΣ	2	0.320114
1	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	ΜΥΤΙΛΗΝΗ	1	0.396332
2	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	ΣΑΜΟΣ	1	0.396332
3	ΗΡΑΚΛΕΙΟΥ (Π.Ε.)	ΧΙΟΣ	1	0.396332
4	ΚΑΛΥΜΝΟΣ	ΑΝΑΦΗ	1	0.396332
learn All	ΚΑΛΥΜΝΟΣ	ΑΤΤΙΚΗΣ (ΠΕΡΙΦΕΡΕΙΑ)	1	0.396332
6	ΚΑΛΥΜΝΟΣ	ΛΕΙΨΟΙ	1	0.396332
7	ΚΑΛΥΜΝΟΣ	ΠΑΤΜΟΣ	1	0.396332
8	ΚΑΛΥΜΝΟΣ	ΡΟΔΟΣ	1	0.396332
9	ΚΩΣ	ΑΤΤΙΚΗΣ (ΠΕΡΙΦΕΡΕΙΑ)	7	3.902346
12	ΚΩΣ	ΝΙΣΥΡΟΣ	2	0.320114
10	ΚΩΣ	ΚΑΛΥΜΝΟΣ	1	0.396332
11	ΚΩΣ	ΛΕΡΟΣ	1	0.396332
13	ΚΩΣ	ΠΑΤΜΟΣ	1	0.396332
14	ΚΩΣ	ΡΟΔΟΣ	1	0.396332
15	ΚΩΣ	ΤΗΛΟΣ	1	0.396332

Εικόνα 29. Παράδειγμα εύρεσης ακραίων τιμών (outliers)

Ο παραπάνω πίνακας φιλτραρίστηκε, ώστε να παραμείνουν μόνο οι γραμμές για τις οποίες το z-score είναι μεγαλύτερο ή ίσο της τιμής 3. Το αποτέλεσμα δείχνει ότι ακραία τιμή για το πλήθος των φορτωτικών αποτελεί η διαδρομή Κως προς Αττική (Περιφέρεια) και Μυτιλήνη προς Αττική (Περιφέρεια), όπως φαίνεται στον παρακάτω πίνακα.

Outliers πλήθους φορτωτικών

	Τόπος Αναχώρησης	Τόπος Προορισμού	Πλήθος φορτωτικών	Z-score
9	ΚΩΣ	ΑΤΤΙΚΗΣ (ΠΕΡΙΦΕΡΕΙΑ)	7	3.902346
23	ΜΥΤΙΛΗΝΗ	ΑΤΤΙΚΗΣ (ΠΕΡΙΦΕΡΕΙΑ)	7	3.902346

Εικόνα 30. Παράδειγμα εντοπισμού ακραίων τιμών (outlier)

Τα μεγέθη που προέκυψαν δείχνουν ότι και πάλι δύο νησιωτικές αφετηρίες στέλνουν υπερβολικά μεγάλες ποσότητες φορτωτικών προς τον ίδιο προορισμό (Αττική (Περιφέρεια)). Οι υπόλοιπες αποστολές κινούνται σε συγκεκριμένα όρια του συνόλου του πλήθους των φορτωτικών, χωρίς μεγάλες αποκλίσεις από αυτό.

(Γ) Γενικός Πληθυσμός

Γ1. Τι ποσοστό του πληθυσμού χρησιμοποιεί το Μεταφορικό Ισοδύναμο ανά νησί;
Αρχικά δημιουργήθηκε ένα dataframe (tickets) με το σύνολο των εισιτηρίων για τα έτη 2018, 2019 και 2020. Από το dataframe αφαιρέθηκαν οι διπλότυπες εμφανίσεις των ίδιων αναγνωριστικών δικαιούχων, ώστε να προκύψουν οι μοναδικοί δικαιούχοι του Μεταφορικού Ισοδυνάμου.

Στη συνέχεια φορτώθηκε το dataset Δικαιούχοι 30-04-20.xlsx, ώστε να προσδιοριστεί το νησί στο οποίο έχει καταχωρηθεί ο κάθε δικαιούχος. Το dataframe ονομάστηκε 'δικαιούχοι'.

Στη συνέχεια τα δύο dataframes συγχωνεύθηκαν, ώστε να προκύψει μία νέα στήλη για κάθε δικαιούχο, η οποία αντιστοιχεί στο νησί που έχει αυτός καταχωρηθεί. Το αποτέλεσμα της συγχώνευσης καταχωρείται στο dataframe 'tickets_merged'.

Το dataframe αυτό ομαδοποιείται ως προς το νησί για να γίνει απαρίθμηση των μοναδικών δικαιούχων ανά νησί που χρησιμοποιούν το Μεταφορικό Ισοδύναμο. Προκύπτει έτσι ένα dataframe με στήλες το όνομα του νησιού και το πλήθος των δικαιούχων του Μεταφορικού Ισοδυνάμου που έχουν ταξιδέψει επωφελούμενοι το Μεταφορικό Ισοδύναμο (το dataframe ονομάστηκε 'mi_users_per_island'). Για να υπολογιστεί όμως το ποσοστό των ωφελούμενων που χρησιμοποίησαν το Μεταφορικό Ισοδύναμο επί του πληθυσμού του κάθε νησιού, έπρεπε να βρούμε την πληροφορία για τον πληθυσμό κάθε νησιού. Οι πληροφορίες αυτές αντλήθηκαν από την απογραφή πληθυσμού 2011 και ενσωματώθηκαν σε κατάλληλο αρχείο Excel.

Για να υπολογιστεί το ποσοστό χρηστών του Μεταφορικού Ισοδυνάμου επί του πληθυσμού, το dataframe 'mi_users_per_island' συγχωνεύθηκε με το αρχείο απογραφής και προέκυψε μία επιπλέον στήλη με το συνολικό πληθυσμό του κάθε νησιού. Για να υπολογιστεί το ποσοστό, έγινε διαίρεση της στήλης του συνόλου των ωφελούμενων που ταξίδεψαν με Μεταφορικό Ισοδύναμο ανά νησί διά το σύνολο του πληθυσμού του αντίστοιχου νησιού. Το αποτέλεσμα αποτυπώνεται στον επόμενο πίνακα:

	Νησί	Χρήστες MI ανά νησί	Πληθυσμός (2011)	Ποσοστό πληθυσμού που χρησιμοποιεί MI
0	ΑΓΑΘΟΝΗΣΙ	1	186	0.01
1	ΑΓΙΟΣ ΕΥΣΤΡΑΤΙΟΣ	1	249	0.00
2	ΑΣΤΥΠΑΛΑΙΑ	4	1270	0.00
3	ΙΚΑΡΙΑ	29	8431	0.00
4	ΚΑΛΥΜΝΟΣ	51	15863	0.00
5	ΚΩΣ	39	33388	0.00
6	ΛΕΙΨΟΙ	2	784	0.00
7	ΛΕΡΟΣ	37	7915	0.00
8	ΛΕΣΒΟΣ	315	86312	0.00
9	ΛΗΜΝΟΣ	60	16743	0.00
10	ΝΙΣΥΡΟΣ	3	982	0.00
11	ΠΑΤΜΟΣ	4	3429	0.00
12	ΣΑΜΟΘΡΑΚΗ	49	2773	0.02

Εικόνα 31. Ποσοστό του πληθυσμού χρησιμοποιεί το Μεταφορικό Ισοδύναμο ανά νησί

Η εικόνα που προκύπτει από τη χρήση του μεταφορικού ισοδυνάμου είναι ότι πολύ μικρό ποσοστό του γενικού πληθυσμού των ελληνικών νησιών χρησιμοποιεί το Μεταφορικό Ισοδύναμο. Το ποσοστό αυτό έχει υπολογιστεί στο σύνολο των μοναδικών δικαιούχων του Μεταφορικού Ισοδυνάμου οι οποίοι έκαναν τουλάχιστον ένα ταξίδι μέσα στο χρονικό διάστημα εφαρμογής του μέτρου.

Υπάρχει μήπως χρονική διακύμανση ανά νησί εντός του έτους και μεταξύ των μηνών και των εποχών (χειμερινοί μήνες - καλοκαιρινοί μήνες);

Αρχικά δημιουργήθηκε ένα dataframe (tickets) με το σύνολο των εισιτηρίων για τα έτη 2018, 2019 και 2020. Από το dataframe αφαιρέθηκαν οι διπλότυπες εμφανίσεις των ίδιων αναγνωριστικών δικαιούχων, ώστε να προκύψουν οι μοναδικοί δικαιούχοι του Μεταφορικού Ισοδυνάμου.

Στη συνέχεια φορτώθηκε το dataset Δικαιούχοι 30-04-20.xlsx, ώστε να προσδιοριστεί το νησί στο οποίο έχει καταχωρηθεί ο κάθε δικαιούχος. Το dataframe ονομάστηκε 'δικαιούχοι'.

Στη συνέχεια τα δύο dataframes συγχωνεύθηκαν, ώστε να προκύψει μία νέα στήλη για κάθε δικαιούχο, η οποία αντιστοιχεί στο νησί που έχει αυτός καταχωρηθεί. Το αποτέλεσμα της συγχώνευσης καταχωρείται στο dataframe 'tickets_merged'.

Στο dataframe 'tickets_merged' δημιουργούνται δύο στήλες με το έτος και το μήνα του ταξιδιού, ώστε να είναι δυνατή η περαιτέρω ομαδοποίηση ανά έτος και μήνα. Ακολουθεί ομαδοποίηση ανά στήλες 'Νησί', 'Έτος' και 'Μήνας', ώστε να προκύψει το σύνολο εισιτηρίων Μεταφορικού Ισοδυνάμου ανά νησί, έτος και μήνα.

Εφαρμόζεται στη συνέχεια η συνάρτηση var για τον υπολογισμό της διακύμανσης κάθε γραμμής του dataframe tickets_merged. Το αποτέλεσμα παρουσιάζεται στον παρακάτω πίνακα.

Διακύμανση ανά νησί ανά μήνα					
	Νησί	Έτος	Μήνας	Ταξίδια MI ανά νησί	Variance
0	ΑΓΑΘΟΝΗΣΙ	2018	7	2	1351400.33333
1	ΑΓΙΟΣ ΕΥΣΤΡΑΤΙΟΣ	2018	7	1	1352074.33333
2	ΑΓΙΟΣ ΕΥΣΤΡΑΤΙΟΣ	2018	8	1	1351406.33333
3	ΑΣΤΥΠΑΛΑΙΑ	2018	7	7	1348040.33333
4	ΙΚΑΡΙΑ	2018	7	32	1331490.33333
5	ΙΚΑΡΙΑ	2018	8	8	1346700.00000
6	ΚΑΛΥΜΝΟΣ	2018	7	60	1313449.00000
7	ΚΑΛΥΜΝΟΣ	2018	8	28	1333433.33333
8	ΚΩΣ	2018	7	24	1336741.00000
9	ΚΩΣ	2018	8	26	1334748.00000

Εικόνα 32. Διακύμανση ανά νησί ανά μήνα

Για να υπολογιστεί η διακύμανση μεταξύ τουριστικών περιόδων, δημιουργήθηκε μία νέα στήλη στο dataframe 'tickets_merged', με τίτλο 'Περίοδος'. Η στήλη αυτή παίρνει τιμή 'Χειμερινή' εάν ο μήνας είναι Νοέμβριος έως και Μάρτιος και 'Θερινή' για όλους του άλλους μήνες.

Εφαρμόζεται στη συνέχεια η συνάρτηση var για τον υπολογισμό της διακύμανσης κάθε γραμμής του dataframe tickets_merged. Το αποτέλεσμα παρουσιάζεται στον παρακάτω πίνακα.

Διακύμανση ανά νησί ανά τουριστική Περίοδο				
	Νησί	Περίοδος	Ταξίδια MI ανά νησί	Variance
0	ΑΓΑΘΟΝΗΣΙ	θερινή	2	NaN
1	ΑΓΙΟΣ ΕΥΣΤΡΑΤΙΟΣ	θερινή	2	NaN
2	ΑΣΤΥΠΑΛΛΑΙΑ	θερινή	7	NaN
3	ΙΚΑΡΙΑ	θερινή	40	NaN
4	ΚΑΛΥΜΝΟΣ	θερινή	88	NaN
5	ΚΩΣ	θερινή	50	NaN
6	ΛΕΙΨΟΙ	θερινή	5	NaN
7	ΛΕΡΟΣ	θερινή	74	NaN
8	ΛΕΣΒΟΣ	θερινή	502	NaN
9	ΛΗΜΝΟΣ	θερινή	94	NaN
10	ΝΙΣΥΡΟΣ	θερινή	3	NaN
11	ΠΑΤΜΟΣ	θερινή	5	NaN

Εικόνα 33. Διακύμανση ανά νησί ανά τουριστική περίοδο

Γ2. Ποιοι είναι οι κυρίαρχοι προορισμοί (μέγιστο 25%) της μετακίνησης του πληθυσμού ανά νησί;

Για να βρεθούν οι κυρίαρχοι προορισμοί ανά νησί, θα χρησιμοποιηθεί το dataset των εισιτηρίων για τα έτη 2018, 2019 και 2020. Το dataset αυτό ομαδοποιείται ανά αφετηρία και προορισμό, υπολογίζοντας το πλήθος των ταξιδιών που έγιναν από μία αφετηρία στον κάθε προορισμό.

Προκειμένου να υπολογιστεί το 25% των κορυφαίων προορισμών, στο παραπάνω αποτέλεσμα εφαρμόστηκε η συνάρτηση `nlargest`, η οποία για την περίπτωση σαν όρισμα δέχθηκε μία συνάρτηση `lambda`:

```
lambda x: x.nlargest(int(len(x) * 0.25)
```

Η συνάρτηση αυτή υπολογίζει το 25% του συνολικού πλήθους των γραμμών που έχουν προκύψει για κάθε ομάδα δεδομένων και κρατάει στο αποτέλεσμα το πλήθος αυτών των γραμμών.

Το αποτέλεσμα του υπολογισμού απεικονίζεται στον ακόλουθο πίνακα:

Κυρίαρχοι προορισμοί (μέγιστο 25%) μετακίνησης πληθυσμού ανά νησί

	Κωδικός Λιμανιού Αναχώρησης	Κωδικός Λιμανιού Αφίξης	Πλήθος
5	AKI	PIR	5
12	EYD	KAR	38
24	JKH	MJT	334
33	JSY	KAR	12
46	KAR	PIR	78
54	KGS	PIR	38
58	KGS	RHO	31
48	KGS	AOK	3
65	KMI	PIR	31
69	KMI	RHO	25
76	KVA	MJT	121
83	LIP	PIR	4
97	MJT	PIR	483
92	MJT	JKH	335
95	MJT	MYR	169

Εικόνα 34. Κυρίαρχοι προορισμοί μετακίνησης πληθυσμού ανά νησί (μέγιστο 25%)

Ο πίνακας του αποτελέσματος προσφέρει μια καθαρή εικόνα του κορυφαίου 25% των προορισμών από κάθε λιμάνι. Το αποτέλεσμα ταξινομεί ανά λιμάνι αναχώρησης το πλήθος αναχωρήσεων σε φθίνουσα σειρά. Άμεσα προκύπτει το δημοφιλέστερο λιμάνι προορισμού δεδομένου ενός λιμανιού αναχώρησης.

Το αποτέλεσμα υπολογίζεται ανά λιμάνι και όχι ανά νησί, δεδομένου ότι δεν είναι διαθέσιμη η αντιστοίχιση λιμένων με νησί, ώστε να γίνει περαιτέρω ομαδοποίηση.

Γ3. Ποια νησιά έχουν σχετικά μεγάλες ενδο-νησιακές μετακινήσεις πληθυσμού (πχ. Διεκπεραίωση διοικητικών υποχρεώσεων);

Προκειμένου να συμπεράνουμε ποιες μετακινήσεις είναι ενδονησιακές, κάνουμε την παραδοχή ότι ένας δικαιούχος μετέβη σε ένα άλλο νησί για τη διεκπεραίωση κάποιας υπόθεσής του εάν ισχύουν τα παρακάτω:

1. Ο προορισμός είναι ένα επίσης νησιωτικό λιμάνι και όχι λιμάνι της ηπειρωτικής χώρας.
2. Ο δικαιούχος, στο τελευταίο χρονικά ταξίδι του επιστρέφει στο λιμάνι της αφετηρίας του (δηλαδή της πρώτης αναχώρησης).

Για να είναι δυνατοί οι υπολογισμοί αυτοί, αρχικά έγινε επεξεργασία του συνόλου των κωδικών των λιμένων της χώρας, ώστε να καθοριστούν οι νησιωτικοί λιμένες. Η μορφή που κατέληξε μετά την επεξεργασία του είναι η ακόλουθη:

A	B	C	D
Κωδικός	Όνομασία	Λατινική ονομασία	Νησί
KIM	Κύμη Εύβοιας	Kimi (Kyimi)	N
KIN	Κοιλάδα Ερμιονίδας	Koilas Ermionidos	N
KIO	Κιάτο	Kiatio	N
KIP	Κυπαρισσία	Kiparissia	N
KIS	Καστέλι Κισσάμου Χανίων	Kastelli Kissamou	Y
KLD	Κάλαθος Ρόδου	Kalathos Rodou	Y
KLI	Κλήμα Σκοπέλου	Klima	Y
KLL	Καλοί Λιμένες Ηρακλείου	Kalilimenes	Y
KLM	Καλαμάκι Ισθμίων	Kalamaki	N
KLX	Καλαμάτα	Kalamata	N
KMI	Κάλυμνος	Kalymnos	Y
KMS	Κίμωλος	Kimolos	Y
KNS	Κοντιάς Λήμνου	Kontias Limnou	Y
KOF	Κουφονήσια Κυκλάδων	Koufonissia Kykladon	Y
KRE	Καμάρες Σίφνου	Kamares Sifnos	Y
JTY	Αστυπάλαια	Astypalaia	Y
KAG	Καργκινάρι Ικαρίας	Karkinagri Ikarias	N
KAH	Κάτω Αχαΐα	Kato Achaia	N
KAI	Κάλαμος (Λευκάδας)	Kalamos Island (Lefkada)	N
KAK	Κατάκολο	Katakolon	N
KAP	Καψάλι Κυθήρων	Kapsalion (Kythira)	N

Εικόνα 35. Επεξεργασία λιμένων της χώρας

Στο σύνολο δεδομένων προστέθηκε μία στήλη, η οποία περιέχει τις εξής τιμές:

- Y: εάν ο λιμένας είναι νησιωτικός
- N: εάν ο λιμένας ανήκει στην ηπειρωτική χώρα

Στη συνέχεια δημιουργήθηκε ένα dataframe (all_tickets) με το σύνολο των εισιτηρίων για τα έτη 2018, 2019 και 2020 που είναι διαθέσιμα. Το dataframe αυτό περιλαμβάνει τις στήλες:

1. 'Αναγνωριστικό Δικαιούχου που ταξίδεψε'
2. 'Είδος Εισιτηρίου (Ακτοπλοϊκό / Αεροπορικό)'
3. 'Κωδικός Λιμανιού Αναχώρησης'
4. 'Κωδικός Λιμανιού Άφιξης'
5. 'Κωδικός Αεροδρομίου Αναχώρησης'
6. 'Κωδικός Αεροδρομίου Άφιξης'
7. 'Ημερομηνία και Ώρα αναχώρησης'

Ο σκοπός είναι σε πρώτη φάση να απομονωθούν εκείνες οι διαδρομές οι οποίες αφορούν αναχωρήσεις από νησιωτικούς λιμένες και αφίξεις σε νησιωτικούς επίσης λιμένες αποκλειστικά. Για το σκοπό αυτό έγινε join του dataframe all_tickets με το dataframe των λιμένων, ώστε στο αποτέλεσμα να συμπεριληφθούν μόνο οι διαδρομές του δικαιούχου προς νησιωτικούς προορισμούς και επιστροφή.

Τα αποτελέσματα ταξινομήθηκαν ανά κωδικό δικαιούχου και ημερομηνία μετακίνησης, ώστε να ομαδοποιηθούν οι διαδρομές και να ταξινομηθούν χρονικά, και να είναι γνωστή η ημερομηνία της αναχώρησης από το νησί του δικαιούχου και η ημερομηνία επιστροφής του σε αυτό. Το αποτέλεσμα αυτής της επεξεργασίας φαίνεται στην παρακάτω εικόνα:

Αναγνωριστικό Δικαιούχου που ταξίδεψε	Ημερομηνία και Ώρα αναχώρησης	Κωδικός Λιμανιού Αναχώρησης	Κωδικός Λιμανιού Αφίξης	Διαδρομή
186 0040DC00-6853-4981-96A0-736432F89DF3	2018-08-08 17:45:00	MJT	MYR	MJTMYR
139 00A4D7CB-97F1-451D-A228-4CA1C6B247D4	2018-07-29 12:45:00	MJT	MYR	MJTMYR
623 00A4D7CB-97F1-451D-A228-4CA1C6B247D4	2018-08-02 21:40:00	MYR	MJT	MYRMJT
309 01D23905-F1AF-4DB4-8748-C7FE3847A516	2018-07-28 09:00:00	VLC	PAA	VLC PAA
830 01D23905-F1AF-4DB4-8748-C7FE3847A516	2018-07-29 18:30:00	PAA	VLC	PAAVLC
...
672 FEF24FD8-C0D2-4B4A-AC5F-3BDF365D9A1B	2018-07-31 13:40:00	EYD	MJT	EYDMJT
1 FF606E2E-1799-413A-89DD-1AD84967BF11	2018-07-14 20:30:00	MJT	VTH	MJTVTH
497 FF606E2E-1799-413A-89DD-1AD84967BF11	2018-07-17 16:45:00	VTH	MJT	VTHMJT
715 FFC75071-2C99-46D9-916A-3DDCE7066628	2018-08-02 04:00:00	PKK	RHO	PKKRHO
444 FFC75071-2C99-46D9-916A-3DDCE7066628	2018-08-02 23:59:00	RHO	PKK	RHOPKK

Εικόνα 36. Ταξινόμηση αποτελεσμάτων

Στη συνέχεια εφαρμόστηκε μία τεχνική συγχώνευσης των γραμμών του dataframe που αφορούν στις μετακινήσεις ενός δικαιούχου σε μία, όπου οι διαδοχικές διαδρομές του εισάγονται σε μία δομή λίστας, κάθε στοιχείο της οποίας είναι οι κωδικοί λιμανιών αναχώρησης και προορισμού. Το αποτέλεσμα φαίνεται στον παρακάτω πίνακα:

Αναγνωριστικό Δικαιούχου που ταξίδεψε	Διαδρομή
0040DC00-6853-4981-96A0-736432F89DF3	[MJTMYR]
00A4D7CB-97F1-451D-A228-4CA1C6B247D4	[MJTMYR, MYRMJT]
01D23905-F1AF-4DB4-8748-C7FE3847A516	[VLC PAA, PAAVLC]
0233C9D2-0A24-4092-8139-9EAA94628EE8	[MJTVTH, VTHMJT]
0292EB10-2012-41C1-BDB1-1E0B3BDD117B	[PAA MJT]
...	...
FC41FDA0-C292-4DB7-B189-E4598BB4266E	[MJTVTH]
FDB8722D-DE2C-4F8C-AA6F-791A54FAA4F2	[MJTAKI]
FEF24FD8-C0D2-4B4A-AC5F-3BDF365D9A1B	[MJTEYD, EYDMJT]
FF606E2E-1799-413A-89DD-1AD84967BF11	[MJTVTH, VTHMJT]
FFC75071-2C99-46D9-916A-3DDCE7066628	[PKKRHO, RHOPKK]

Εικόνα 37. Κωδικοί λιμανιών αναχώρησης και προορισμού

Από τη λίστα της στήλης 'Διαδρομή', εντοπίζονται εκείνες οι διαδρομές που στο πρώτο στοιχείο της λίστας έχουν τον ίδιο τριψήφιο κωδικό λιμανιού με το τελευταίο τριψήφιο τμήμα του τελευταίου στοιχείου της λίστας.

Τέλος, τα αποτελέσματα ομαδοποιούνται ανά τον τριψήφιο κωδικό του λιμανιού αναχώρησης και υπολογίζεται το πλήθος τους. Το αποτέλεσμα απεικονίζεται στον παρακάτω πίνακα:

Διαδρομή	Αριθμός
KMI	22
KRM	1
LIP	2
MJT	119
MXK	4
MYR	21
NIS	5
PKK	16
PMS	1
PYT	5
RHO	4
SYM	11
TIL	8
VLC	15
VTH	9

Εικόνα 38. Διαδρομές

Η ομαδοποίηση γίνεται ανά λιμάνι, δεδομένου ότι δεν είναι διαθέσιμη κάποια ομαδοποίηση των λιμανιών ανά νησί, ώστε να προκύψουν συνολικά οι αναχωρήσεις και επιστροφές σε επίπεδο νησιού.

Γ4. Εξαρτήσεις / συνδέσεις μεταξύ διαφορετικών παραμέτρων (π.χ. κύκλος εργασιών επιχειρήσεων, χρήση Μεταφορικού Ισοδυνάμου, κύριο νησί αναφοράς, εποχικότητα, κ.α.)

Αναλύθηκε η σχέση μεταξύ μεγέθους επιχείρησης και ποσού επιδότησης Μεταφορικού Ισοδυνάμου που έλαβαν. Για το σκοπό αυτό, το ποσό της επιδότησης εντάχθηκε σε 11 κατηγορίες, ανά 10.000€ ληφθείσας επιδότησης.

Δημιουργήθηκε ένα dataframe με στήλες το μέγεθος της επιχείρησης (<8.000.000 και >=8.000.000 σε κύκλο εργασιών) και την κατηγορία ληφθείσας επιδότησης. Το dataframe αυτό ομαδοποιήθηκε, ώστε να προκύψει το πλήθος των επιχειρήσεων ανά μέγεθος και κατηγορία επιδότησης. Το αποτέλεσμα του υπολογισμού αυτού φαίνεται στον παρακάτω πίνακα:

	Μέγεθος	Κατηγορία επιδότησης	Πλήθος επιχειρήσεων
0	8.000.000 - 40.000.000	(0, 10000]	97
1	8.000.000 - 40.000.000	(10000, 20000]	21
2	8.000.000 - 40.000.000	(20000, 30000]	18
3	8.000.000 - 40.000.000	(30000, 40000]	6
4	8.000.000 - 40.000.000	(40000, 50000]	1
5	8.000.000 - 40.000.000	(50000, 60000]	0
6	8.000.000 - 40.000.000	(60000, 70000]	0
7	8.000.000 - 40.000.000	(70000, 80000]	1
8	8.000.000 - 40.000.000	(80000, 90000]	0
9	8.000.000 - 40.000.000	(90000, 100000]	0
10	8.000.000 - 40.000.000	(100000, 110000]	0
11	<8.000.000	(0, 10000]	18650
12	<8.000.000	(10000, 20000]	151
13	<8.000.000	(20000, 30000]	37
14	<8.000.000	(30000, 40000]	22
15	<8.000.000	(40000, 50000]	2
16	<8.000.000	(50000, 60000]	3
17	<8.000.000	(60000, 70000]	3
18	<8.000.000	(70000, 80000]	3
19	<8.000.000	(80000, 90000]	1
20	<8.000.000	(90000, 100000]	0
21	<8.000.000	(100000, 110000]	1

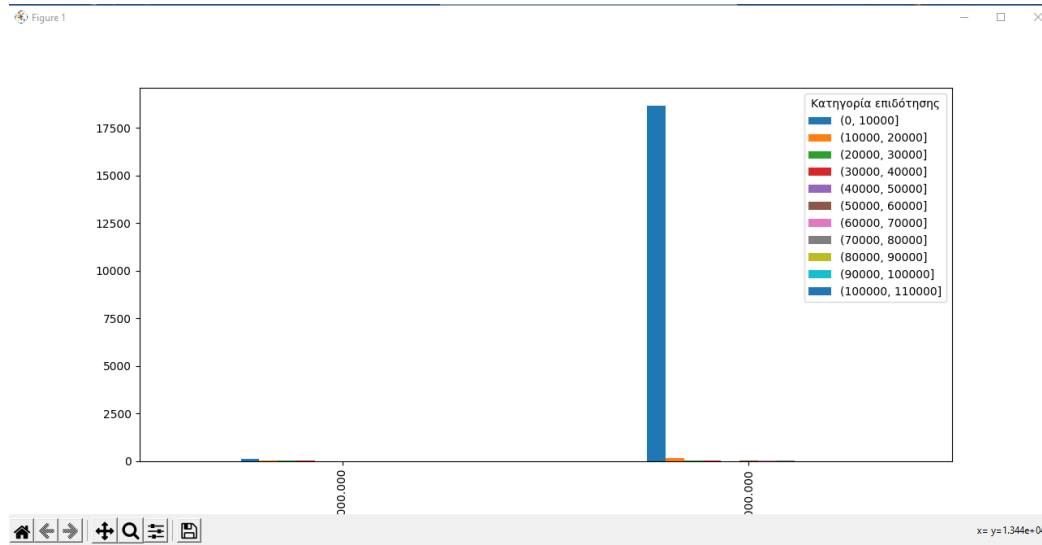
Εικόνα 39. Κατηγορίες Επιδότησης

Προκειμένου τα δεδομένα αυτά να απεικονιστούν σε διάγραμμα στηλών, χρησιμοποιήθηκε η συνάρτηση .pivot, η οποία δημιούργησε τον συγκεντρωτικό πίνακα που ακολουθεί:

Κατηγορία επιδότησης	(0, 10000]	(10000, 20000]	(20000, 30000]
Μέγεθος			
8.000.000 - 40.000.000	97	21	18
<8.000.000	18650	151	37

Εικόνα 40. Κατηγορίες και μεγέθη επιδότησης

Η απεικόνιση του συγκεντρωτικού πίνακα σε γράφημα είναι η ακόλουθη:



Εικόνα 41. Κατηγορίες επιδότησης

Συμπεραίνουμε ότι το μεγάλο πλήθος επιδοτήσεων χορηγείται σε επιχειρήσεις με κύκλο εργασιών < 8.000.000€. Οι μεγαλύτερες επιχειρήσεις (οι οποίες είναι και λιγότερες σε πλήθος). Επίσης, τα περισσότερα ποσά των επιχορηγήσεων κινούνται στην περιοχή έως 10.000€

Γ5. Κατηγοριοποιήσεις βάσει διαφορετικών (και επικαλυπτόμενων) παραμέτρων.

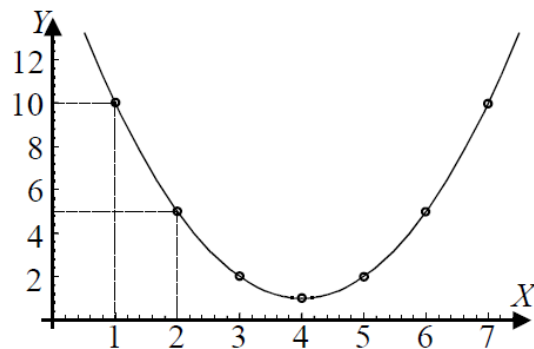
Ανάλυση Παλινδρόμησης

Με την *ανάλυση παλινδρόμησης (regression analysis)* εξετάζουμε τη σχέση μεταξύ δύο ή περισσότερων μεταβλητών με σκοπό την πρόβλεψη των τιμών της μιας, μέσω των τιμών της άλλης (ή των άλλων). Σε κάθε πρόβλημα παλινδρόμησης διακρίνουμε δύο είδη μεταβλητών: τις **ανεξάρτητες ή ελεγχόμενες ή επεξηγηματικές (independent, predictor, casual, input, explanatory variables)** και τις **εξαρτημένες ή απόκρισης (dependent, response variables)**. Σε **πειραματικές έρευνες**, ανεξάρτητη μεταβλητή X είναι εκείνη την οποία μπορούμε να ελέγξουμε, δηλαδή, να καθορίσουμε τις τιμές της (π.χ. το ύψος της διαφημιστικής δαπάνης ενός προϊόντος, ο αριθμός των λειτουργούντων ταμείων σε ένα υποκατάστημα τραπεζής, η ποσότητα λιπάσματος, η θερμοκρασία επεξεργασίας ενός προϊόντος). **Εξαρτημένη μεταβλητή** Y είναι εκείνη στην οποία αντανακλάται το αποτέλεσμα των μεταβολών στις ανεξάρτητες μεταβλητές (π.χ. η ζήτηση ενός προϊόντος, ο χρόνος αναμονής των πελατών ενός υποκαταστήματος τραπεζής, η απόδοση μιας καλλιέργειας, η αντοχή ενός υλικού). Σε **μη πειραματικές έρευνες (δειγματοληψίες)** η διάκριση μεταξύ ανεξάρτητων και εξαρτημένων μεταβλητών δεν είναι πάντοτε σαφής γιατί καμία μεταβλητή δεν είναι ελεγχόμενη αλλά όλες

είναι τυχαίες (π.χ. το ύψος και το βάρος των φοιτητών, οι ώρες μελέτης των φοιτητών ενός πανεπιστημιακού τμήματος και η απόδοση τους σε ένα τεστ, οι εβδομάδες εμπειρίας ενός εργάτη σε μια επιχείρηση και ο αριθμός των ελαττωματικών προϊόντων που παράγει, η κατάταξη δέκα προϊόντων από έναν κριτή και η κατάταξη των ιδίων προϊόντων από έναν άλλο κριτή, ο αριθμός των πωλήσεων μουσικών CD σε μια περιοχή και ο αριθμός των νέων στην ίδια περιοχή) (Schneider, et al., 2010).

Ας θεωρήσουμε δύο μεταβλητές X, Y . Αν οι μεταβλητές αυτές συνδέονται με μια σχέση της μορφής $Y = f(X)$ μέσω της οποίας για κάθε τιμή της X μπορούμε να προβλέψουμε ακριβώς την τιμή της Y , δηλαδή, αν οι τιμές της Y δεν υπόκεινται σε σφάλματα, τότε λέμε ότι οι δύο μεταβλητές συνδέονται με τη *συναρτησιακή-προσδιοριστική (deterministic) σχέση* $Y = f(X)$. Για παράδειγμα, το ρεύμα που καταναλώνει μια οικογένεια σε ένα δίμηνο και το ποσό που πληρώνει για την κατανάλωση αυτή συνδέονται με συναρτησιακή-προσδιοριστική σχέση⁶. Επίσης, το ποσό που καταθέτει κάποιος στο Ταμειυτήριο και ο τόκος που παίρνει για το ποσό αυτό, συνδέονται με συναρτησιακή-προσδιοριστική σχέση. Σε αυτές τις περιπτώσεις τα σημεία του διαγράμματος διασποράς βρίσκονται όλα πάνω στην καμπύλη που έχει εξίσωση $Y = f(X)$ και όσες φορές και αν επαναλάβουμε το πείραμα θέτοντας το X στο ίδιο επίπεδο $X = x_i$, θα παίρνουμε πάντα την ίδια τιμή για το Y . Για παράδειγμα, η εξίσωση $Y = (X - 4)^2 + 1$ (που παριστάνει μια παραβολή) περιγράφει προσδιοριστικά τη σχέση μεταξύ των X και Y του παρακάτω πίνακα:

x_i	y_i
1	10
2	5
3	2
4	1
5	2
6	5
7	10

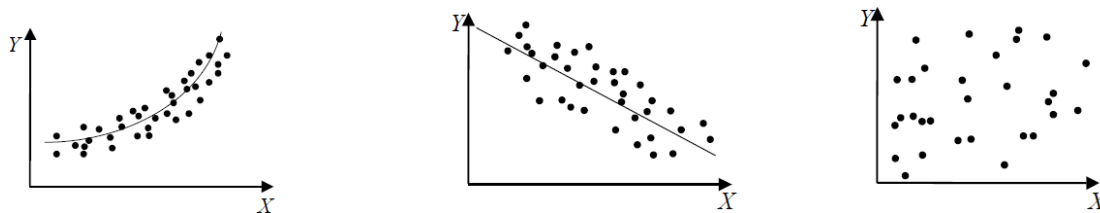


Εικόνα 42. Γραμμική παλινδρόμηση

Οι μη προσδιοριστικές σχέσεις μεταξύ μεταβλητών ονομάζονται *στοχαστικές – στατιστικές (stochastic, probabilistic) σχέσεις*. Στην περίπτωση αυτή, αν επαναλάβουμε το πείραμα πολλές φορές θέτοντας το X στο ίδιο επίπεδο $X = x_i$ τότε στην τιμή x_i της X δεν αντιστοιχεί μια μόνο τιμή y_i της Y αλλά, γενικά, αντιστοιχεί ένα πλήθος διαφορετικών τιμών της Y . Για παράδειγμα, αν X είναι η τιμή ενός προϊόντος και Y είναι η ζήτησή του, η Y βρίσκεται σε στοχαστική σχέση-εξάρτηση από τη X , γιατί η ζήτηση ενός προϊόντος επηρεάζεται και από άλλους παράγοντες όπως είναι το ύψος του εισοδήματος των καταναλωτών, οι τιμές ομοειδών προϊόντων, οι καταναλωτικές συνήθειες, κ.ά.

Σε μια στοχαστική σχέση το διάγραμμα διασποράς είναι, γενικά, ένα *νέφος σημείων* το οποίο πολλές φορές καθορίζει μια ιδεατή γραμμή η οποία δίνει μια πρώτη εικόνα της σχέσης που συνδέει τις δύο μεταβλητές. Η σχέση μάλιστα μεταξύ των δύο μεταβλητών είναι τόσο περισσότερο ισχυρή όσο πιο κοντά στην ιδεατή γραμμή βρίσκονται τα σημεία του διαγράμματος διασποράς. Στο πρώτο από τα παρακάτω σχήματα έχουμε το διάγραμμα διασποράς μιας ισχυρής σχέσης στην οποία όταν αυξάνουν οι τιμές της X αυξάνουν γενικά

και οι τιμές της Y , ενώ στο δεύτερο σχήμα έχουμε μια λιγότερο ισχυρή σχέση στην οποία όταν αυξάνουν οι τιμές της X ελαττώνονται γενικά και οι τιμές της Y . Τέλος, στην περίπτωση του τρίτου σχήματος δε φαίνεται να υπάρχει κάποια σχέση μεταξύ των X και Y .



Εικόνα 43. Διαγράμματα διασποράς
 Πηγή: Εργαστήριο Μαθηματικών & Στατιστικής / Γ. Παπαδόπουλος
 (www.aua.gr/gpapadopoulos)

Γενικά, δύο μεταβλητές που συνδέονται είτε με συναρτησιακή-προσδιοριστική σχέση είτε με στοχαστική σχέση λέγονται «εξαρτημένες». Αν υπάρχει εξάρτηση μεταξύ δύο μεταβλητών, τότε μπορούμε τη μια από αυτές να τη χαρακτηρίσουμε ως «αιτία» και την άλλη ως «αποτέλεσμα». Αυτό όμως, μόνο στην περίπτωση που η εξάρτηση οφείλεται σε σχέση αιτιότητας των δύο μεταβλητών και όχι σε μια απλή συμμεταβολή η οποία μπορεί να οφείλεται σε εξάρτηση των δύο μεταβλητών από μια τρίτη μεταβλητή. Αν, για παράδειγμα, X είναι το ετήσιο εισόδημα μιας οικογένειας και Y , Z είναι τα ποσά που ξοδεύει η οικογένεια αυτή σε ένα έτος για κρέας και για αγορά λογοτεχνικών βιβλίων, τότε: αν διαπιστώσουμε σε ένα σύνολο οικογενειών σχέση μεταξύ των X και Y (ή μεταξύ των X και Z) δεχόμαστε ότι υπάρχει εξάρτηση μεταξύ των δύο μεταβλητών και τότε μπορούμε να χαρακτηρίσουμε τη X ως «αιτία» και την Y (ή τη Z) ως «αποτέλεσμα». Αν όμως διαπιστωθεί σχέση μεταξύ των Y και Z (που είναι πολύ πιθανό, αφού και οι δύο μεταβάλλονται με το ετήσιο εισόδημα X) ασφαλώς θα πρόκειται για «νόθα» εξάρτηση.

Για να περιγράψουμε τη στοχαστική εξάρτηση δύο μεταβλητών X και Y προσπαθούμε να βρούμε, όπως και στην προσδιοριστική εξάρτηση, μια σχέση μεταξύ των X και Y η οποία όμως τώρα δε θα δίνει ακριβή αλλά προσεγγιστική μόνο εικόνα της εξάρτησης των X και Y και τα σημεία του διαγράμματος διασποράς των X και Y δε θα βρίσκονται πάνω, αλλά, γύρω από μια καμπύλη.

Συσχέτιση πλήθους παραστατικών με το μέγεθος εταιρίας

Προκειμένου να μελετηθεί η εξάρτηση του πλήθους των παραστατικών που επιδοτούνται στο πλαίσιο του μεταφορικού ισοδύναμου, εφαρμόστηκε μέθοδος γραμμικής παλινδρόμησης (linear regression). Η επιλογή των μεταβλητών έγινε ως εξής:

1. Εξαρτημένη μεταβλητή: αριθμός παραστατικών
2. Ανεξάρτητη μεταβλητή: κύκλος εργασιών

Τα βήματα για την εφαρμογή της γραμμικής παλινδρόμησης είναι:

1. Το dataset των αιτήσεων χρηματοδότησης φορτώθηκε σε κατάλληλο dataframe (με όνομα «fundingApplications»).
2. Από το dataframe αυτό δημιουργήθηκαν δύο επιμέρους dataframes για την εξαρτημένη και ανεξάρτητη μεταβλητή αντίστοιχα (με ονόματα αντίστοιχα “dependent” και “independent”).

Η βιβλιοθήκη που χρησιμοποιήθηκε για την εφαρμογή της γραμμικής παλινδρόμησης είναι η statsmodels. Η εφαρμογή του μοντέλου στην python έγινε με την εξής κλήση:

```
model = sm.OLS(dependent, independent).fit()
```

OLS Regression Results						
Dep. Variable:	Αριθμός Παραστατικών	R-squared (uncentered):	0.104			
Model:	OLS	Adj. R-squared (uncentered):	0.104			
Method:	Least Squares	F-statistic:	2257.			
Date:	Wed, 26 Jan 2022	Prob (F-statistic):	0.00			
Time:	13:21:01	Log-Likelihood:	-1.1799e+05			
No. Observations:	19370	AIC:	2.360e+05			
Df Residuals:	19369	BIC:	2.360e+05			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Κύκλος Εργασιών	1.874e-05	3.94e-07	47.511	0.000	1.8e-05	1.95e-05
Omnibus:	21796.145	Durbin-Watson:	1.198			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	7888706.324			
Skew:	5.313	Prob(JB):	0.00			
Kurtosis:	101.293	Cond. No.	1.00			

Εικόνα 44. Γραμμική παλινδρόμηση

Μελετώντας τον παραπάνω πίνακα, τα βασικά συμπεράσματα που εξαγονται είναι:

Η παράμετρος $P = 0.000 < 0.05$, πράγμα το οποίο σημαίνει ότι υπάρχει ισχυρή συσχέτιση μεταξύ της εξαρτημένης μεταβλητής «Αριθμός παραστατικών» με την ανεξάρτητη μεταβλητή «Κύκλος εργασιών» της επιχείρησης.

Η παράμετρος $coef = 1.874e-05$ δηλώνει ότι η αύξηση της εξαρτημένης μεταβλητής κατά 1 μονάδα (δηλαδή η έκδοση ενός επιπλέον παραστατικού), σημαίνει αύξηση του κύκλου εργασιών της επιχείρησης κατά $1.874e-05\text{€}$.

Η παράμετρος R^2 (coefficient of determination ή R-squared). αξιολογεί το πόσο καλά το μοντέλο της γραμμικής παλινδρόμησης που εφαρμόστηκε εξηγεί τη διακύμανση της τιμής της εξαρτημένης μεταβλητής. Ο συντελεστής αυτός για ένα μοντέλο γραμμικής παλινδρόμησης λαμβάνει τιμές από το 0 έως 1 και εκφράζει το ποσοστό της διακύμανσης που κατορθώνει να εξηγήσει το γραμμικό μοντέλο παλινδρόμησης. Ο συντελεστής R^2 δεν έχει μονάδες μέτρησης καθότι εκφράζει ποσοστό. Στην εφαρμογή η τιμή του R^2 είναι 0.104, το οποίο σημαίνει ότι το γραμμικό μοντέλο παλινδρόμησης είναι ικανό να εξηγήσει το 10.4% της διακύμανσης της εξαρτημένης μεταβλητής στο σύνολο των δεδομένων. Η τιμή του συντελεστή R^2 είναι αρκετά χαμηλή και δεν επιτυγχάνει επαρκώς να εξηγήσει τη διακύμανση της ανεξάρτητης μεταβλητής.

Στη συνέχεια έγινε δοκιμή του κατά πόσο το μοντέλο μπορεί να προβλέψει με ακρίβεια τις τιμές του πλήθους των παραστατικών από τον κύκλο εργασιών μιας επιχείρησης. Για το

σκοπό αυτό δημιουργήθηκε ένα train set και ένα test set από τα διαθέσιμα δεδομένα και σε αυτά τα σύνολα εκτελέστηκαν οι εντολές:

```
linreg = sl.LinearRegression()
```

```
linreg.fit(X_train, y_train)
```

Με η εντολή fit εφαρμόζει το μοντέλο linear regression στα δεδομένα εκπαίδευσης.

Ο υπολογισμός του R squared στα σύνολα εκπαίδευσης και ελέγχου γίνεται με τις εντολές:

```
linreg.score(X_train, y_train)
```

```
linreg.score(X_test, y_test)
```

Τα αποτελέσματα του R squared είναι αντίστοιχα:

```
R squared of the Linear Regression on training set: 0.067
```

```
R squared of the Linear Regression on test set: -0.01
```

Στη συνέχεια, με την εντολή γίνεται η πρόβλεψη της εξαρτημένης μεταβλητής στα test δεδομένα, προκειμένου να υπολογιστεί το Root mean square error.

```
linreg.predict(X_test)
```

Το αποτέλεσμα είναι:

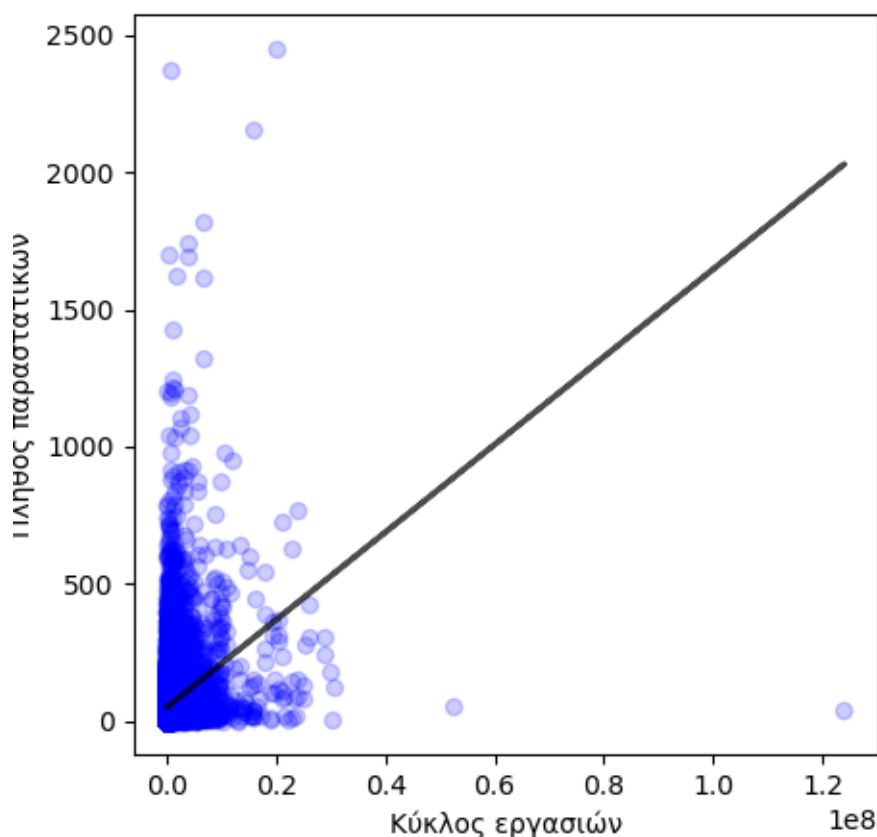
```
Root mean squared error (RMSE): 90.52051191709667
```

Η μελέτη των παραπάνω μετρικών (R Squared και RMSE) μας οδηγεί στα εξής συμπεράσματα:

1. Οι χαμηλές τιμές του R Squared τόσο στο training όσο και στο test set σημαίνουν ότι η ανεξάρτητη μεταβλητή που χρησιμοποιήθηκε για να γίνει η πρόβλεψη του πλήθους των παραστατικών που υποβάλλει μία επιχείρησης στο Μεταφορικό ισοδύναμο, μπορεί να κάνει την πρόβλεψη μόνο σε περίπου 6% των περιπτώσεων.
2. Η υψηλή τιμή του RMSE σημαίνει ότι υπάρχει μία μέση απόκλιση της προβλεφθείσας από την πραγματική τιμή παραστατικών της τάξης των 90 μονάδων.

Συμπερασματικά μπορούμε να πούμε ότι η μεταβλητή «Κύκλος εργασιών» δεν είναι μία αξιόπιστη παράμετρος για πρόβλεψη του πλήθους των παραστατικών μεταφορικού ισοδύναμου που εκδίδει μία εταιρεία.

Το διάγραμμα διασποράς (scatter plot) της εξαρτημένης σε σχέση με την ανεξάρτητη μεταβλητή είναι το παρακάτω:



Εικόνα 45. Διάγραμμα Διασποράς

Η παρατήρηση της γραμμής κλίσης μας οδηγεί στο συμπέρασμα ότι υπάρχει γραμμικά θετική συσχέτιση της εξαρτημένης από την ανεξάρτητη μεταβλητή, άρα αύξηση του κύκλου εργασιών μιας επιχείρησης προκαλεί γραμμική αύξηση του αριθμού των παραστατικών μεταφορικού ισοδύναμου που υποβάλλει για χρηματοδότηση.

5. Σύνοψη

Η συγκεκριμένη έρευνα βασίστηκε στα δεδομένα τα οποία προέκυψαν από την εφαρμογή του μέτρου του Μεταφορικού Ισοδυναμού στα ελληνικά νησιά. Η εφαρμογή του μέτρου ξεκίνησε από το 2018 και αντίστοιχα τα δεδομένα που είναι διαθέσιμα για επεξεργασία αναφέρονται στο διάστημα 2018 έως 2020.

Το μέτρο έχει εφαρμοστεί εκτός από την Ελλάδα και σε ορισμένες χώρες της Ευρωπαϊκής Ένωσης, οι οποίες έχουν νησιωτικά συμπλέγματα (Σουηδία, Σκωτία, Νορβηγία, Κορσική). Εντούτοις, σε καμία χώρα από αυτές δεν έχει δημοσιευτεί κάποια μελέτη σχετική με τα αποτελέσματα της εφαρμογής αυτού του μέτρου.

Τα ερευνητικά ερωτήματα που διατυπώθηκαν στο πλαίσιο αυτής της μελέτης αφορούσαν τόσο τα οφέλη των επιχειρήσεων σε σχέση με την επιβάρυνσή τους λόγω των μεταφορικών εξόδων της ναυσιπλοΐας όσο και των ιδιωτών κατοίκων των νησιών, οι οποίοι επιδοτούνται

για τις θαλάσσιες διαδρομές που κάνουν κατά τη διάρκεια του έτους. Η έρευνα είναι πρωτότυπη και αποκάλυψε σημαντικά στοιχεία σε σχέση με την αξιοποίηση του μέτρου του Μεταφορικού Ισοδυνάμου στη νησιωτική Ελλάδα.

Η ποιότητα των προς επεξεργασία δεδομένων ήταν καλή, και η ανάγκη για καθαρισμό τους ήταν πολύ περιορισμένη. Στην εξέλιξη της έρευνας απαραίτητο ήταν να συνδυαστούν αρκετοί από τους πίνακες των δεδομένων, ώστε να παραχθούν τα επιθυμητά αποτελέσματα.

6. Επόμενα βήματα

Δεδομένου ότι το μέτρο του Μεταφορικού Ισοδυνάμου θα συνεχίσει να εφαρμόζεται, τα δεδομένα που θα παράγονται από τη χρήση του θα αυξάνονται διαρκώς. Αυτό σημαίνει ότι θα πρέπει να δημιουργηθεί μία πλατφόρμα, η οποία θα είναι σε θέση να ενημερώνεται ανά τακτά χρονικά διαστήματα με τα νέα δεδομένα, ώστε οι αναλύσεις και οι υπολογισμοί να γίνονται πάνω σε επικαιροποιημένα δεδομένα. Αυτό προϋποθέτει την ανάπτυξη ενός μηχανισμού για την αυτοματοποιημένη εισαγωγή των στοιχείων αυτών.

Η εργασία αυτή παρήγαγε ένα σημαντικό πλήθος αποτελεσμάτων, τα οποία θα πρέπει να είναι δυνατόν να εξειδικευτούν από τους αρμόδιους, οι οποίοι θα παρακολουθούν την εξέλιξη της αποδοτικότητας του μέτρου του Μεταφορικού Ισοδυνάμου. Αυτό σημαίνει ότι θα πρέπει να δημιουργηθεί μία εφαρμογή, στην οποία ο χρήστης θα μπορεί να επιλέγει το ή τα νησιά τα οποία θέλει να μελετήσει και να παράγει αποτελέσματα για τα νησιά αυτά και μόνο. Αυτό προϋποθέτει την δημιουργία περιβάλλοντος υποβολής ερωτήσεων και εφαρμογής φίλτρων πάνω στα αποτελέσματα της ανάλυσης αυτής της εργασίας.

Περαιτέρω εξέλιξη μπορεί να αποτελέσει η διασύνδεση της εφαρμογής με λογισμικά οπτικοποίησης των αποτελεσμάτων (visualization), όπως το Microsoft Power BI και το Tableau. Στο πλαίσιο της εξέλιξης της εφαρμογής αυτής, τα αποτελέσματα της ανάλυσης των δεδομένων θα διαβιβάζονται στις εφαρμογές αυτές μέσω πρότυπων αρχείων (.csv), ώστε ο μελετητής να παράγει προσαρμοσμένες οπτικοποιήσεις των αποτελεσμάτων.

Βιβλιογραφία

- Bhuvanya, R., Sagunthala, D. & Matheswaran, P., 2018. An Overview: Classification of Big Data Tools. *International Journal of Engineering & Technology*, 6(7).
- Boslaugh, S. & Watters, P., 2008. *Statistics in a nutshell*. Cambridge: O'Reilly Media Inc.
- Embarak, O. & Karkal, 2018. *Data analysis and visualization using python*. s.l.:Springer.
- European Commission, 2020. *The EU Blue Economy Report 2020*, s.l.: European Commission.
- Eurostat, 2015. *Top 20 Passenger Ports in 2015 - On the Basis of Number of Passengers Embarked and Disembarked, Statistics Explained*, s.l.: Eurostat.
- Foster, J., Barkus, E. & Yavorsky, C., 2006. *Understanding and using advanced statistics*. London: SAGE Publications.
- Jambu, M., 1991. *Exploratory and multivariate data analysis*. s.l.:Elsevier.
- Nelli, F., 2015. *Python data analytics: Data analysis and science using PANDAs, Matplotlib and the Python Programming Language*. s.l.:Apress.
- OECD, 2020. *OECD Territorial Reviews - Regional Policy for Greece Post-2020*, Paris: OECD Publishing.
- Schneider, A., Hommel, G. & Blettner, M., 2010. Linear Regression Analysis. *Dtsch Arztebl*, 107(44).
- Van Rossum, et al., 2017. *Python programming language*. s.l., USENIX annual technical conference.
- Ντουμή, Μ. & Βουκελάτου, Ε., 2020. *Μεταφορικό Ισοδύναμο - Μία πρώτη αποτίμηση της εφαρμογής του. Η περίπτωση της Χίου*, Χίος: Πανεπιστήμιο Αιγαίου, Πολυτεχνική Σχολή, Τμήμα Μηχανικών Οικονομίας και Διοίκησης.