



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ
ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Ανάλυση δεδομένων προβλέψεων και κυβερνητικών
μέτρων προστασίας και στήριξης κατά την πανδημία
της COVID-19

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Εμμανουήλ Κ. Βεκράκης

Επιβλέπων: Βασίλειος Ασημακόπουλος
Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2021



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρικών Βιομηχανικών Διατάξεων και
Συστημάτων Αποφάσεων

Ανάλυση δεδομένων προβλέψεων και κυβερνητικών
μέτρων προστασίας και στήριξης κατά την πανδημία
της COVID-19

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Εμμανουήλ Κ. Βεκράκης

Επιβλέπων: Βασίλειος Ασημακόπουλος
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 3η Νοεμβρίου 2021.

.....
Βασίλειος Ασημακόπουλος
Καθηγητής Ε.Μ.Π.

.....
Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.

.....
Ιωάννης Ψαρράς
Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2021

.....
Εμμανουήλ Κ. Βεκράκης
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Εμμανουήλ Κ. Βεκράκης, 2021.
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Η πανδημία της νόσου COVID-19, που προκαλείται από τον ιό SARS-CoV-2, ξεκίνησε από την πόλη Γουχάν της Κίνας στα τέλη του 2019. Έκτοτε, έχει προκαλέσει περισσότερους από 4.5 εκατομμύρια θανάτους παγκοσμίως, τα μέτρα προστασίας έχουν πλήξει την παγκόσμια οικονομία και αγορά εργασίας σφοδρότερα από την κρίση του 2008, ενώ ο εγκλεισμός και το άγχος απειλούν με ανεπανόρθωτες κοινωνικές και ψυχολογικές συνέπειες το κοινωνικό σύνολο.

Αντικείμενο της παρούσας εργασίας αποτελεί ο καθορισμός των παραγόντων που, σε συνδυασμό με τα μέτρα προστασίας, οδήγησαν ορισμένες χώρες σε μια πιο ομαλή διαχείριση της πανδημίας σε σχέση με άλλες. Στο πλαίσιο αυτό, εξετάζεται η επίδραση 8 παραγόντων σε 33 χώρες, ενώ διενεργείται επαλήθευση 7 εξ αυτών για 48 Πολιτείες των Ηνωμένων Πολιτειών της Αμερικής. Ακόμη, αναλύεται η διαδικασία παραγωγής προβλέψεων των εβδομαδιαίων θανάτων για τις ΗΠΑ, η επίδραση των εξωτερικών παραγόντων σε αυτές και η εξέλιξη της ακρίβειας τους με την πάροδο του χρόνου. Σκοπός όλων αυτών, είναι η εξαγωγή συμπερασμάτων που θα βοηθήσουν στην πιο εμπεριστατωμένη διαδικασία λήψης αποφάσεων κατά τον σχεδιασμό των μέτρων προστασίας από την ίδια ή παρόμοιες πανδημίες στο μέλλον.

Η διαδικασία που ακολουθείται για το πρώτο σκέλος της εργασίας, αποτελείται από την αποκλιμάκωση των δεδομένων και τη συσταδοποίηση των χωρών με βάση τα χαρακτηριστικά τους. Έτσι, εξετάζεται η εξέλιξη της πανδημίας για τις χώρες της κάθε συστάδας. Η ίδια διαδικασία επαναλαμβάνεται για την περίπτωση των ΗΠΑ για την επαλήθευση των συμπερασμάτων που προκύπτουν. Στο δεύτερο σκέλος, εξετάζονται σε επίπεδο χρονοσειρών τα μέτρα προστασίας σε συνδυασμό με τους θανάτους, καθώς επίσης και οι προβλέψεις που παρήχθησαν για τους θανάτους στις ΗΠΑ.

Τα σημαντικότερα συμπεράσματα που προέκυψαν από το πρώτο σκέλος της εργασίας είναι ότι τα μεγάλα αστικά κέντρα αποτέλεσαν εστία υπερμετάδοσης του κορωνοϊού στην αρχή της πανδημίας, ενώ επιζήμια έδρασαν και οι τουριστικές αφίξεις. Ακόμη, χώρες με καλύτερο βιοτικό επίπεδο κατάφεραν να θέσουν υπό έλεγχο την πανδημία πολύ πιο αποτελεσματικά, σε αντίθεση με χώρες χαμηλού βιοτικού επιπέδου. Η μέση ετήσια θερμοκρασία δεν φάνηκε να επηρεάζει τους θανάτους, κάνοντας τον παράγοντα του κλίματος αντικείμενο μελλοντικής έρευνας. Σχετικά με τα μοντέλα προβλέψεων, ο βασικός παράγοντας που επηρεάζει τις προβλέψεις τους είναι οι θάνατοι της προηγούμενης εβδομάδας, ενώ τα μοντέλα που χρησιμοποιήθηκαν φάνηκε να είναι απαισιόδοξα.

Λέξεις κλειδιά

Πανδημία, COVID-19, Ανάλυση Δεδομένων, Συσταδοποίηση, Μέτρα Προστασίας, Προβλέψεις

Abstract

The COVID-19 pandemic, caused by SARS-CoV-2, originated in Wuhan, China in late 2019. Since then, more than 4.5 million people around the world have died and protective measures have crippled the global economy and labour market even worse than the 2008 crisis. At the same time, the public faces irreparable social and psychological consequences from confinement and stress.

The subject of this study, is to determine the factors which, coupled with protective measures, lead to some countries containing the pandemic more smoothly than others. In this context, the impact of 8 factors in 33 countries is studied, while 7 of them are validated in 48 states of the United States of America. Furthermore, the forecasting methods used to predict weekly deaths in the USA, the impact of external factors on them and their accuracy over time are analyzed. The ultimate goal is to draw conclusions that will aid to more informed decision making regarding the government response in this or future pandemics.

The first part of the study, consists of scaling the data set and clustering the countries based on their features. Thus, it is analyzed how the pandemic progressed in the countries of each cluster. Then, the same process is repeated for the USA, in order to validate the results. In the second part, the timeseries of protective measures in relation to deaths, as well as the forecasted deaths in the USA are examined.

The most notable results of the first part of this study are that metropolitan areas facilitate the super spread of the coronavirus at the start of the pandemic, while tourist arrivals also helped in that direction. In addition to that, countries with better quality of life managed to control the pandemic much more effectively, than countries with low quality of life. Average annual temperature does not seem to affect COVID-19 deaths, leaving the examination of the climate factor a subject for a future study. Regarding the forecasting models, the driving factor of the forecasts are previous week's deaths, while the models themselves were found to be pessimistic.

Keywords

Pandemic, COVID-19, Data Analysis, Clustering, Protective Measures, Forecasting

Ευχαριστίες

Αρχικά θέλω να ευχαριστήσω τον καθηγητή Βασίλειο Ασημακόπουλο για την ευκαιρία και την εμπιστοσύνη που μου έδειξε να εκπονήσω τη διπλωματική μου εργασία στη Μονάδα Προβλέψεων και Στρατηγικής. Στη συνέχεια οφείλω να ευχαριστήσω θερμά τον ερευνητικό συνεργάτη της μονάδας, δρ. Ευάγγελο Σπηλιώτη, ο οποίος με καθοδηγούσε και με συμβούλευε όλο αυτό το διάστημα με μεγάλη επιτυχία.

Με την παρουσίαση της διπλωματικής μου εργασίας, κλείνει για εμένα το κεφάλαιο της τριτοβάθμιας εκπαίδευσης, γεγονός που μου δίνει την ευκαιρία να αναπολήσω όλη αυτή την εμπειρία ζωής. Αρχικά, το μεγαλύτερο μερίδιο επαίνων για την μέχρι τώρα πορεία μου, αξίζει στην οικογένεια μου, τους γονείς μου Κωνσταντίνο και Ειρήνη και την αδερφή μου, Αθηνά, για την αμέριστη φροντίδα, την αγάπη, τη στήριξη τους και τη μέριμνα για την εκπαίδευση μου όλα αυτά τα χρόνια. Οφείλω επίσης ένα πολύ μεγάλο ευχαριστώ και στη γιαγιά μου, Κατερίνα, για την αγάπη της και για όλα εκείνα τα βράδια που η διαμονή κοντά στο Πολυτεχνείο ήταν μονόδρομος.

Τέλος, δεν θα μπορούσα να μην αναφερθώ στις σχέσεις που σύναψα κατά τη διάρκεια των σπουδών μου, και που ελπίζω να με ακολουθούν και στο πολύ μακρινό μέλλον. Την Σταματία, τον Γιάννη, τον Γιώργο, το Δημήτρη, τον Κωνσταντίνο, τον Χρήστο και τον Γιώργο, καθώς επίσης και τον παιδικό μου φίλο, Θοδωρή. Χωρίς αυτούς, την στήριξη τους σε κάθε δυσκολία και την παρουσία τους σε κάθε ευχάριστη στιγμή, η ζωή μου αυτά τα χρόνια θα είχε ένα μεγάλο κενό. Για αυτό και τους ευχαριστώ βαθύτατα.

Περιεχόμενα

Περίληψη	5
Abstract	7
Ευχαριστίες	9
Περιεχόμενα	12
Κατάλογος Σχημάτων	14
Κατάλογος Πινάκων	15
1 Εισαγωγή	16
1.1 Κίνητρο	16
1.1.1 Πανδημία COVID-19	16
1.1.2 Μέτρα Προστασίας και Στήριξης	17
1.2 Σκοπός	18
1.3 Σχετικές Έρευνες και Μελέτες	18
1.4 Δομή Εργασίας	19
2 Μέθοδοι και Τεχνικές Ανάλυσης Δεδομένων	21
2.1 Συσταδοποίηση	21
2.1.1 Εισαγωγικές Έννοιες	21
2.1.2 Αλγόριθμος K-means	22
2.1.3 Αλγόριθμοι ιεραρχικής συσταδοποίησης	23
2.1.4 Αλγόριθμος DBSCAN	24
2.1.5 Επιλογή αλγορίθμου	25
2.2 Παλινδρόμηση	27
2.2.1 Εισαγωγικές Έννοιες	27
2.2.2 Απλή Γραμμική Παλινδρόμηση	27
2.3 Ανάλυση Κύριων Συνιστωσών (PCA)	29

3	Ανάλυση Συγκεντρωτικών Δεδομένων	32
3.1	Καθορισμός Συνόλου Δεδομένων	32
3.1.1	Θάνατοι που ευθύνονται στην COVID-19	32
3.1.2	Κλιματικά Στοιχεία	34
3.1.3	Μέτρα Προστασίας και Στήριξης	35
3.1.4	Δημογραφικά και Οικονομικά Στοιχεία	45
3.1.5	Τουριστικές Αφίξεις	48
3.2	Επίδραση Δεδομένων	48
3.3	Ανάλυση Συγκεντρωτικών Δεδομένων	50
3.3.1	Περιγραφή Διαδικασίας	50
3.3.2	Ανάλυση σε Κύριες Συνιστώσες (PCA)	53
3.3.3	Συσταδοποίηση	54
3.3.4	Αποτελέσματα Συσταδοποίησης	57
3.4	Συμπεράσματα	64
3.5	Η περίπτωση των ΗΠΑ	65
3.5.1	Καθορισμός Συνόλου Δεδομένων	66
3.5.2	Επίδραση Δεδομένων	67
3.5.3	Ανάλυση Δεδομένων	70
3.5.4	Συμπεράσματα	73
4	Ανάλυση Χρονοσειρών	78
4.1	Επίδραση Μέτρων Προστασίας σε Επίπεδο Χρόνου	78
4.1.1	Επίδραση Θανάτων στη Λήψη Μέτρων Προστασίας	78
4.1.2	Επίδραση Μέτρων Προστασίας στη Μείωση των Θανάτων	81
4.2	Επίδραση Μεταβλητών στις Προβλέψεις	83
4.3	Εξέλιξη Ακρίβειας Προβλέψεων	92
5	Συμπεράσματα και μελλοντικές επεκτάσεις	94

Κατάλογος Σχημάτων

1.1	Παγκόσμια αύξηση του ΑΕΠ ανά τα χρόνια	17
2.1	Ενδεικτικό παράδειγμα συσταδοποίησης.	22
2.2	Παράδειγμα εκτέλεσης αλγορίθμου K-Means	23
2.3	Ενδεικτικό παράδειγμα ιεραρχικής συσταδοποίησης με το αντίστοιχο δενδρόγραμμα.	25
2.4	Χαρακτηρισμός σημείων δεδομένων με τον αλγόριθμο DBSCAN	26
2.5	Ενδεικτικό παράδειγμα απλής γραμμικής παλινδρόμησης	28
2.6	Ενδεικτικά παραδείγματα συντελεστή συσχέτισης Pearson (r)	28
2.7	Ενδεικτικό παράδειγμα εφαρμογής της Ανάλυσης Κύριων Συνιστωσών (PCA).	31
3.1	Κρούσματα COVID-19 προς θανάτους ανά χώρα	33
3.2	Κατανομή συνολικών θανάτων ανα εκατομμύριο των χωρών	34
3.3	Κατανομή μέσης θερμοκρασίας των χωρών	35
3.4	Θάνατοι COVID-19 ανά εκατομμύριο προς Containment and Health Index ανά χώρα	44
3.5	Διακύμανση Containment and Health Index ανά χώρα	46
3.6	Εξέλιξη Containment and Health Index κατά τη διάρκεια της πανδημίας	47
3.7	Κατανομή δημογραφικών και οικονομικών στοιχείων ανά χώρα	49
3.8	Κατανομή τουριστικών αφίξεων για κάθε χώρα	50
3.9	Διαγράμματα σκέδασης θανάτων ανά εκατομμύριο με τα υπόλοιπα χαρακτηριστικά των χωρών	51
3.9	Διαγράμματα σκέδασης θανάτων ανά εκατομμύριο με τα υπόλοιπα χαρακτηριστικά των χωρών (συνέχεια)	52
3.10	Διαδικασία ανάλυσης συγκεντρωτικών δεδομένων	53
3.11	Διακύμανση που διατηρείται ανάλογα το πλήθος Κυρίων Συνιστωσών	54
3.12	Διάγραμμα χωρών μετά την εφαρμογή PCA	55
3.13	Δενδρόγραμμα συσσωρευτικής ιεραρχικής συσταδοποίησης χωρών	57
3.14	Κατανομή χωρών σε συστάδες	58
3.15	Διάγραμμα χωρών μετά την εφαρμογή συσταδοποίησης	59
3.16	Γεωγραφική κατανομή χωρών σε συστάδες	60
3.17	Εβδομαδιαία εξέλιξη της πανδημίας όλων των χωρών	61
3.18	Εβδομαδιαία εξέλιξη της πανδημίας για κάθε συστάδα	62

3.19	Εβδομαδιαία εξέλιξη της πανδημίας για κάθε συστάδα χωρίς συνάντρωση	63
3.20	Κατανομή χωρών κάθε συστάδας στα χαρακτηριστικά	65
3.20	Κατανομή χωρών κάθε συστάδας στα χαρακτηριστικά (συνέχεια) . .	66
3.21	Διαγράμματα σκέδασης θανάτων ανά εκατομμύριο με τα υπόλοιπα χαρακτηριστικά των Πολιτειών	68
3.21	Διαγράμματα σκέδασης θανάτων ανά εκατομμύριο με τα υπόλοιπα χαρακτηριστικά των Πολιτειών (συνέχεια)	69
3.22	Δενδρόγραμμα συσσωρευτικής ιεραρχικής συσταδοποίησης Πολιτειών	70
3.23	Κατανομή Πολιτειών σε συστάδες	71
3.24	Εβδομαδιαία εξέλιξη της πανδημίας των 48 Πολιτειών	72
3.25	Εβδομαδιαία εξέλιξη της πανδημίας για κάθε συστάδα	73
3.26	Εβδομαδιαία εξέλιξη της πανδημίας για κάθε συστάδα χωρίς συνάντρωση	74
3.27	Κατανομή Πολιτειών κάθε συστάδας στα χαρακτηριστικά	75
3.27	Κατανομή Πολιτειών κάθε συστάδας στα χαρακτηριστικά (συνέχεια) .	76
3.28	Γεωγραφική κατανομή Πολιτειών σε συστάδες	77
4.1	Εξέλιξη του εβδομαδιαίου CHI για όλες τις χώρες	79
4.2	Εξέλιξη του εβδομαδιαίου CHI και των εβδομαδιαίων θανάτων για όλες τις χώρες	80
4.3	Συσχέτιση θανάτων-μέτρων των χωρών κάθε συστάδας	81
4.4	Εξέλιξη του εβδομαδιαίου CHI για κάθε συστάδα χωρών	82
4.5	Χρονοσειρές θανάτων ανά εκατομμύριο και CHI για τη Δανία και τη Γερμανία	84
4.6	Χρονοσειρές θανάτων ανά εκατομμύριο και CHI για την Ελλάδα και τη Πολωνία	85
4.7	Μετατοπισμένες χρονοσειρές προβλέψεων μία εβδομάδα πίσω	87
4.7	Μετατοπισμένες χρονοσειρές προβλέψεων μία εβδομάδα πίσω (συνέχεια 1)	88
4.7	Μετατοπισμένες χρονοσειρές προβλέψεων μία εβδομάδα πίσω (συνέχεια 2)	89
4.8	Διαγράμματα σκέδασης sMAPE των προβλέψεων με τα υπόλοιπα χαρακτηριστικά	90
4.8	Διαγράμματα σκέδασης sMAPE των προβλέψεων με τα υπόλοιπα χαρακτηριστικά (συνέχεια)	91
4.9	Κατανομή Πολιτειών κάθε συστάδας στο sMAPE των προβλέψεων των θανάτων	92
4.10	Εξέλιξη ποσοστιαίου σφάλματος προβλέψεων με την πάροδο του χρόνου	93

Κατάλογος Πινάκων

3.1	Επεξήγηση συντελεστών που καταγράφει το OxCGRΤ.	41
3.2	Συντελεστές των δεικτών του OxCGRΤ.	42
3.3	Θάνατοι και χρούσματα ανά εκατομμύριο για κάθε εύρος του μέσου CHI	45
3.4	Στατιστικά μεγέθη συσχέτισης χαρακτηριστικών με θανάτους ανά εκατομμύριο των χωρών	51
3.5	Διατηρούμενη διακύμανση ανάλογα το πλήθος Κυρίων Συνιστωσών .	56
3.6	Συντελεστές βαρύτητας χαρακτηριστικών για κάθε κύρια συνιστώσα .	56
3.7	Μέση τιμή χαρακτηριστικών ανά συστάδα	60
3.8	Στατιστικά μεγέθη συσχέτισης χαρακτηριστικών με θανάτους ανά εκατομμύριο των Πολιτειών	67
3.9	Μέση τιμή χαρακτηριστικών ανά συστάδα για τις Πολιτείες	74
4.1	Συσχέτιση και sMAPE προβλέψεων θανάτων πριν και μετά τη μετατόπιση τους μία εβδομάδα πίσω	89

Κεφάλαιο 1

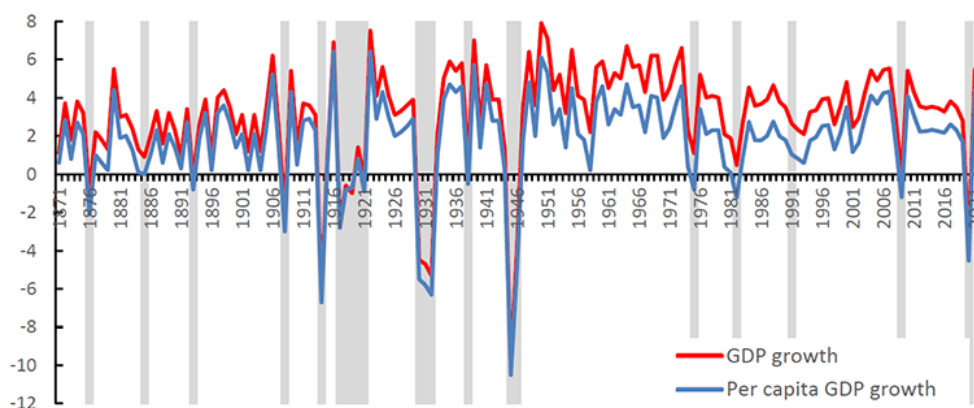
Εισαγωγή

1.1 Κίνητρο

1.1.1 Πανδημία COVID-19

Η νόσος που προκαλείται από τον ιό SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2), γνωστή και ως COVID-19 (Coronavirus Disease 2019) εμφανίστηκε στα τέλη του 2019 στην πόλη Γουχάν της Κίνας και στις 11 Μαρτίου 2020 χαρακτηρίστηκε πανδημία από τον Παγκόσμιο Οργανισμό Υγείας [8]. Τα συμπτώματα της νόσου ποικίλουν, με τα συχνότερα να είναι ο πυρετός, ο ξηρός βήχας και η σωματική εξάντληση, αλλά δεν είναι σπάνιες και οι επιπλοκές, οδηγώντας ένα 20% των νοσούντων να χρειαστεί νοσηλεία [34]. Η θνητότητα των ασθενών της COVID-19 που νοσηλεύθηκαν σε Μονάδα Εντατικής Θεραπείας κυμαίνεται μεταξύ 39% και 72%, ανάλογα με την εκάστοτε έρευνα και τα χαρακτηριστικά των νοσούντων [6]. Η μετάδοση της νόσου από άτομο σε άτομο συμβαίνει όταν αυτά έρθουν σε επαφή μεταξύ τους. Συγκεκριμένα, όταν ένας φορέας του ιού φτερνιστεί, βήξει, μιλήσει ή αναπνεύσει έντονα, τον μεταδίδει μέσω αναπνευστικών σταγονιδίων και αερολυμάτων, τα οποία μπορεί να εισέλθουν στο στόμα, τη μύτη και τα μάτια ενός κοντινού ατόμου. Ο ιός μπορεί επίσης να μεταδοθεί μέσω της επαφής κάποιας μολυσμένης επιφάνειας με τα παραπάνω σημεία εισόδου του ανθρώπινου σώματος. Η ευκολία μετάδοσης του ιού, σε συνδυασμό με την έλλειψη μέσων διάγνωσης και θεραπείας στην αρχή της πανδημίας, ανάγκασε τις περισσότερες χώρες να προσφύγουν σε μέτρα προστασίας, με στόχο την αναστολή της εξάπλωσης του. Παρά τα μέτρα αυτά, η νόσος COVID-19 έχει οδηγήσει σε παραπάνω από 4.5 εκατομμύρια θανάτους παγκοσμίως, ενώ στη χώρα μας ο αριθμός των νεκρών υπερβαίνει τις 14 χιλιάδες. Εκτός αυτού, τα μέτρα προστασίας που ελήφθησαν απειλούν όλους τους πολίτες με ανεπανόρθωτες κοινωνικές και ψυχολογικές συνέπειες λόγω του εγκλεισμού και του εκτεταμένου άγχους [21]. Τέλος, προκάλεσαν πλήγμα στην παγκόσμια οικονομία, αντίστοιχου μεγέθους με τις μεγάλες κρίσεις του παρελθόντος (Σχήμα 1.1), ενώ εκτιμάται ότι το 2020 χάθηκαν 255 εκατομμύρια θέσεις πλήρους απασχόλησης, δηλαδή σχεδόν τέσσερις φορές

περισσότερο από την κρίση το 2009 [18]. Οι τεράστιες αυτές συνέπειες καθιστούν αναγκαία τη μελέτη των παραγόντων που συνέβαλαν στην επιδείνωση της πανδημίας σε κάποιες χώρες σε σχέση με άλλες, προκειμένου να μπορέσουν να αποφευχθούν στο μέλλον, στην πανδημία COVID-19 και σε τυχούσες επόμενες.



Σχήμα 1.1: Παγκόσμια αύξηση του ΑΕΠ ανά τα χρόνια [9]

1.1.2 Μέτρα Προστασίας και Στήριξης

Τα μέτρα προστασίας και στήριξης των πολιτών από τη νόσο COVID-19 χωρίζονται σε δύο κατηγορίες, τα φαρμακευτικά και τα μη φαρμακευτικά. Τα φαρμακευτικά μέσα εμφανίστηκαν στο τέλος του 2020, με την έγκριση εμβολίων τα οποία στοχεύουν στη δημιουργία ενός τοίχους ανοσίας ενάντια στον κορωνοϊό. Η αναγκαστική καθυστέρηση μέχρι την ανάπτυξη και τη διάθεση των εμβολίων, ωστόσο, οδήγησε τις χώρες να στραφούν στα μη φαρμακευτικά μέτρα προστασίας. Τα λιγότερο περιοριστικά εξ αυτών είναι η παρότρυνση του περιορισμού των κοινωνικών επαφών, η τήρηση αποστάσεων σε χώρους όπου δημιουργείται συνωστισμός και η ανίχνευση επαφών σε περίπτωση επιβεβαιωμένου κρούσματος. Πέραν αυτών, οι περισσότερες χώρες έλαβαν και πιο περιοριστικά μέτρα, τα οποία ενδεικτικά περιλαμβάνουν τον περιορισμό των μετακινήσεων - τόσο εγχώριων όσο και διακρατικών -, την απαγόρευση δημόσιων συναθροίσεων και εκδηλώσεων, την επιβεβλημένη χρήση μάσκας προστασίας και την καθιέρωση της τηλεκαίδευσης και τηλεργασίας. Οι επιλογές, λοιπόν, της έντασης, του πεδίου εφαρμογής και της διάρκειας των μέτρων που λαμβάνει η κάθε χώρα είναι αμέτρητες, επομένως η καθεμία έχει χαράξει τη δική της πορεία στον αγώνα προστασίας από την πανδημία. Έπειτα, όπως αναφέρθηκε και προηγουμένως, τα περισσότερα από αυτά τα μέτρα - άλλα περισσότερο και άλλα λιγότερο - έχουν αντίκτυπο και στην οικονομία της κάθε χώρας. Το γεγονός ότι κάθε χώρα υπολογίζει διαφορετικά το ισοζύγιο προστασίας των πολιτών και της οικονομίας, έχει εγείρει ερωτήματα περί της αναγκαιότητας και της αποτελεσματικότητας των μέτρων αυτών. Έρευνες έχουν δείξει, μάλιστα, ότι τα πιο περιοριστικά μέτρα προστασίας, τα οποία είναι αυτά που έχουν

και το μεγαλύτερο οικονομικό και κοινωνικό αντίκτυπο, δεν αρουσιάζουν στατιστικά σημαντικά αποτελέσματα ως προς την ωφέλεια τους στην αντιμετώπιση της πανδημίας έναντι των λιγότερο αυστηρών [10].

1.2 Σκοπός

Ο σκοπός της παρούσας εργασίας είναι η εξερεύνηση και η κατανόηση του πολύπλοκου περιβάλλοντος των παραγόντων που, σε συνδυασμό με τη λήψη ορισμένων μέτρων προστασίας, οδήγησαν την πανδημία COVID-19 να εξελιχτεί διαφορετικά σε ορισμένες χώρες σε σχέση με άλλες. Γίνεται προσπάθεια να καθοριστούν αρχικά οι παράγοντες που μπορούν να επιδράσουν στην πορεία της πανδημίας σε μία χώρα και να βρεθεί ένας κατάλληλος τρόπος ποσοτικοποίησης τόσο των μέτρων προστασίας και στήριξης που ελήφθησαν, όσο και της εξέλιξης της πανδημίας στη χώρα αυτή. Στη συνέχεια, εξετάζεται η σημαντικότητα καθενός από τα παραπάνω στην πορεία της πανδημίας και κατά πόσο χώρες με παρόμοια χαρακτηριστικά και συμπεριφορά, οδηγήθηκαν σε αντίστοιχα παρόμοια αποτελέσματα. Τέλος, αναλύεται ο τρόπος παραγωγής προβλέψεων για τις Ηνωμένες Πολιτείες της Αμερικής σχετικά με τους θανάτους λόγω COVID-19 και η επίδραση των παραπάνω παραγόντων σε αυτές. Όλα αυτά, με στόχο τον καθορισμό των παραγόντων που πράγματι επηρεάζουν την εξέλιξη μιας πανδημίας με παρόμοια συμπεριφορά με αυτή του κορωνοϊού SARS-CoV-2 και τη συμπερίληψη των συμπερασμάτων της μελέτης σε μία πιο εμπειριστατωμένη διαδικασία λήψης αποφάσεων για το σχεδιασμό της πρόληψης και της προστασίας στο μέλλον.

1.3 Σχετικές Έρευνες και Μελέτες

Οι υπάρχουσες έρευνες με σκοπό παρόμοιο με αυτόν που περιγράφηκε παραπάνω, υστερούν από την τρέχουσα σε κάποιον από τους δύο παράγοντες. Στο εύρος των χωρών στο οποίο εστιάζουν και στον τρόπο ποσοτικοποίησης της πορείας της πανδημίας. Σχετικά με το πρώτο, κάποιες έρευνες έχουν εστιάσει περισσότερο σε μικρό αριθμό χωρών, προκειμένου να αναλύσουν τους παράγοντες που επηρεάζουν την εξέλιξη της πανδημίας σε επίπεδο περιοχών εντός των χωρών αυτών [35]. Αξίζει να σημειωθεί ότι τα συμπεράσματα που προκύπτουν από την πιο τοπική ανάλυση συμβαδίζουν σε μεγάλο βαθμό με αυτά που προέκυψαν από την τρέχουσα, παγκόσμια ανάλυση. Ο δεύτερος παράγοντας στον οποίο υστερούν οι υπάρχουσες έρευνες, είναι το μέτρο αξιολόγησης της εξέλιξης της πανδημίας, που είναι το πλήθος κρουσμάτων COVID-19 [19]. Συγκεκριμένα, όπως εξηγείται και παρακάτω, η χρήση των κρουσμάτων COVID-19 κρίθηκε ότι δεν αποτελεί αντικειμενικό μέτρο όταν η έρευνα εκτείνεται σε πολλές χώρες, καθώς το νούμερο αυτό εξαρτάται σε μεγάλο βαθμό από την πολιτική των ελέγχων που εφαρμόζει η κάθε χώρα. Δηλαδή, ορισμένες χώρες διενεργούν μαζικότερα διαγνωστικούς ελέγχους (είτε κατά την ιχνηλάτηση επαφών είτε προληπτικά), ενώ άλλες μόνο σε ανθρώπους που εμφανίζουν συμπτώματα COVID-19. Είναι φυσικό οι δεύτερες να εμφανίζουν πολύ λιγότερα κρούσματα από ο,τι αν διενεργούσαν

περισσότερους ελέγχους.

Στο κομμάτι των προβλέψεων, οι σχετικές έρευνες επικεντρώθηκαν στην παραγωγή προβλέψεων για την πανδημία μέσω ενός ή περισσότερων μοντέλων και στην αξιολόγηση της επίδοσης αυτών. Ορισμένες εξ αυτών, εστίασαν στη χρήση ενός μοντέλου για την πρόβλεψη των νέων κρουσμάτων COVID-19 [28], των θανάτων και των αναρρώσεων [14], ενώ άλλες εξέτασαν διαφορετικά μοντέλα προβλέψεων σε πολλαπλούς τομείς που φτάνουν μέχρι την εξέλιξη της ζήτησης προϊόντων και υπηρεσιών κατά τη διάρκεια της πανδημίας και την προσομοίωση της επίδρασης των κυβερνητικών αποφάσεων [22]. Ωστόσο, οι έρευνες αυτές βασίζονται οι ίδιες στην παραγωγή προβλέψεων, με αποτέλεσμα την ύπαρξη πολλών ερευνητικών ομάδων που παράγουν προβλέψεις για την COVID-19, χωρίς να συναθροίζονται με κάποιον τρόπο οι διαφορετικές αυτές προσεγγίσεις και να αξιολογούνται στο σύνολο τους.

1.4 Δομή Εργασίας

Η παρούσα εργασία ακολουθεί την παρακάτω δομή:

- **Κεφάλαιο 1 - Εισαγωγή.** Στο πρώτο κεφάλαιο γίνεται εισαγωγή στο κίνητρο της μελέτης, δηλαδή την πανδημία COVID-19, ενώ γίνεται και μία αναφορά στο περιβάλλον των μέτρων προστασίας και στήριξης που ελήφθησαν. Στη συνέχεια παρουσιάζεται ο σκοπός της εργασίας, καθώς επίσης και σχετικές έρευνες και μελέτες.
- **Κεφάλαιο 2 - Μέθοδοι και Τεχνικές Ανάλυσης Δεδομένων.** Στο δεύτερο κεφάλαιο, παρουσιάζονται οι μέθοδοι και οι τεχνικές ανάλυσης δεδομένων που χρησιμοποιήθηκαν για την επεξεργασία των δεδομένων της μελέτης και την εξαγωγή συμπερασμάτων. Συγκεκριμένα, παρουσιάζονται τρεις αλγόριθμοι συσταδοποίησης εκ των οποίων επιλέγεται ο ένας, η μέθοδος της απλής γραμμικής παλινδρόμησης και της ανάλυσης στις κύριες συνιστώσες (PCA). Τέλος γίνεται αναφορά στους τρόπους αξιολόγησης προβλέψεων και εξηγείται ο λόγος επιλογής του επικρατέστερου.
- **Κεφάλαιο 3 - Ανάλυση Συγκεντρωτικών Δεδομένων.** Στο τρίτο κεφάλαιο της πανδημίας παρουσιάζονται αρχικά τα χαρακτηριστικά που απαρτίζουν το σύνολο δεδομένων μαζί με τις πηγές τους και η επίδραση τους στους θανάτους λόγω COVID-19. Στη συνέχεια, με βάση τα χαρακτηριστικά αυτά οι χώρες χωρίζονται σε συστάδες, οι οποίες βοηθούν στην εξαγωγή συμπερασμάτων. Τέλος, γίνεται επαλήθευση των συμπερασμάτων για την περίπτωση των Ηνωμένων Πολιτειών της Αμερικής.
- **Κεφάλαιο 4 - Ανάλυση Χρονοσειρών.** Στο τέταρτο κεφάλαιο αναλύονται τα δεδομένα σε επίπεδο χρονοσειρών. Συγκεκριμένα, μελετάται η επίδραση των θανάτων λόγω COVID-19 στα μέτρα προστασίας που λαμβάνονται, καθώς επίσης και το πώς αυτά με τη σειρά τους οδηγούν στη μείωση των θανάτων.

Ακόμη, μελετάται η επίδραση των χαρακτηριστικών του συνόλου δεδομένων στις προβλέψεις των θανάτων για τις Ηνωμένες Πολιτείες της Αμερικής και η εξέλιξη της ακρίβειας των προβλέψεων αυτών στην πορεία του χρόνου.

- **Κεφάλαιο 5 - Συμπεράσματα.** Στο πέμπτο και τελευταίο κεφάλαιο της εργασίας, συνοψίζονται τα συμπεράσματα που προέκυψαν από τα προηγούμενα κεφάλαια και αναφέρονται τρόποι επέκτασης της παρούσας εργασίας στο μέλλον.

Κεφάλαιο 2

Μέθοδοι και Τεχνικές Ανάλυσης Δεδομένων

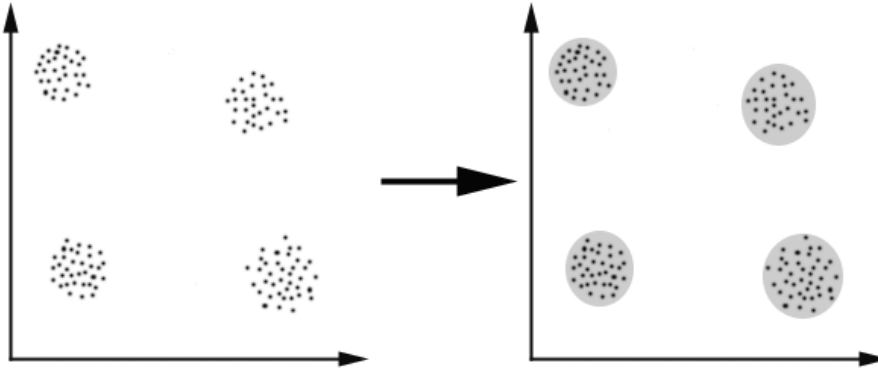
2.1 Συσταδοποίηση

2.1.1 Εισαγωγικές Έννοιες

Οι αλγόριθμοι συσταδοποίησης ανήκουν στην κατηγορία της μη επιβλεπόμενης μάθησης και έχουν να κάνουν με τον χωρισμό ενός χώρου δεδομένων σε ομάδες, τις συστάδες. Τα σημεία δεδομένων του συνόλου κατατάσσονται στις συστάδες, με βάση τον εκάστοτε αλγόριθμο και τη μετρική ομοιότητας που χρησιμοποιούνται. Η επιλογή τόσο του αλγορίθμου συσταδοποίησης όσο και της μετρικής ομοιότητας εξαρτάται από τη φύση του εκάστοτε προβλήματος. Για το λόγο αυτό, η συσταδοποίηση απαιτεί την κατάλληλη προεπεξεργασία των δεδομένων ώστε να αναγνωριστούν τα χαρακτηριστικά γνωρίσματα που θα αποδώσουν τη βέλτιστη ομοιογένεια μεταξύ των σημείων της ομάδας. Ούσα μία μέθοδος μη επιβλεπόμενης μάθησης, απαιτείται η εκ των υστέρων επικύρωση και ερμηνεία των συστάδων που προκύπτουν με στόχο την εξαγωγή νέας γνώσης. Οι αλγόριθμοι συσταδοποίησης κρίνονται ιδιαίτερα χρήσιμοι στην παρούσα εργασία, καθώς καθιστούν ευκολότερη την κατανόηση του συνόλου δεδομένων, το οποίο αποτελείται από μεγάλο πλήθος χωρών. Η οργάνωση των χωρών αυτών σε συστάδες, με βάση ορισμένα χαρακτηριστικά τους, βοηθάει την εξέταση του κατά πόσο τα χαρακτηριστικά αυτά επηρέασαν την πορεία της πανδημίας και την εξαγωγή γενικότερων συμπερασμάτων.

Υπάρχουν τρεις γενικές κατηγορίες αλγορίθμων συσταδοποίησης [36]:

Αλγόριθμοι βασισμένοι σε διαχωρισμούς (partition based) οι οποίοι χωρίζουν επαναληπτικά το χώρο ενός συνόλου δεδομένων σε μικρότερες ομάδες, είτε μέχρι να φτάσουν σε ένα συγκεκριμένο πλήθος συστάδων (π.χ. K-means, K-Medoids), είτε ανακαλύπτοντας τις ομάδες ανάλογα με την πυκνότητα των δεδομένων στο χώρο (π.χ. DBSCAN).



Σχήμα 2.1: Ενδεικτικό παράδειγμα συσταδοποίησης.

Ιεραρχικοί (hierarchical) αλγόριθμοι οι οποίοι επιχειρούν να ανακαλύψουν το πλήθος και τη δομή των ομάδων με ιεραρχικό τρόπο, είτε συγχωνευτικά (agglomerative), είτε διαιρετικά (divisive).

Πιθανοτικοί (partition based) αλγόριθμοι που χρησιμοποιούν μοντέλα πιθανοτήτων για να περιγράψουν σε ποια ομάδα ανήκει κάποιο σημείο, αντί να το κατατάξουν αποκλειστικά σε μία (π.χ. Fuzzy K-Means).

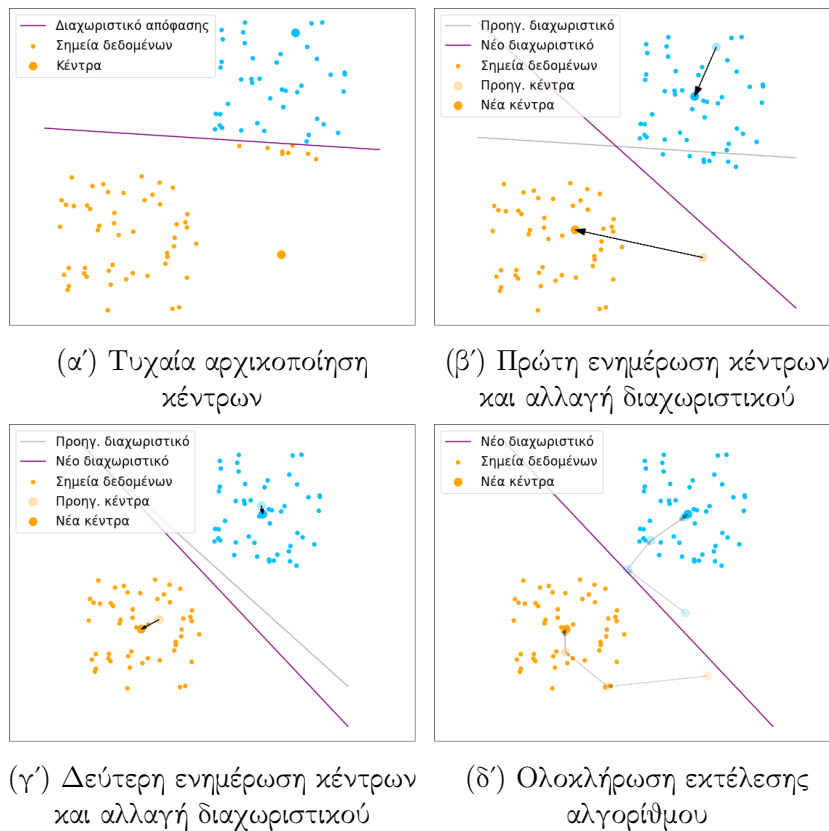
Στη συνέχεια παρουσιάζονται αναλυτικότερα αλγόριθμοι των δύο πρώτων κατηγοριών, καθώς επίσης εξηγείται και ο τρόπος επιλογής του καταλληλότερου για το συγκεκριμένο πρόβλημα.

2.1.2 Αλγόριθμος K-means

Ο αλγόριθμος K-Means είναι από τους πιο γνωστούς αλγορίθμους συσταδοποίησης, εξαιτίας της απλότητας του. Το πλήθος (K) των συστάδων καθορίζεται πριν την εκτέλεση του αλγορίθμου, η οποία έχει ως εξής:

- Αρχικά επιλέγονται τυχαία K σημεία τα οποία αποτελούν τα κέντρα των ομάδων.
- Στη συνέχεια, κάθε σημείο δεδομένων ανατίθεται στην ομάδα της οποίας το κέντρο είναι πιο κοντά του, δηλαδή έχει τη μικρότερη απόσταση.
- Έπειτα για κάθε ομάδα υπολογίζεται το νέο κέντρο της, το οποίο έχει συντεταγμένες τη μέση τιμή των συντεταγμένων των στοιχείων της εκάστοτε ομάδας.
- Τέλος, τα δύο προηγούμενα βήματα επαναλαμβάνονται μέχρις ότου να μην υπάρξει κάποια αλλαγή στη σύνθεση καμίας ομάδας.

Ένα παράδειγμα της εκτέλεσης του αλγορίθμου K-Means φαίνεται στο σχήμα 2.2.



Σχήμα 2.2: Παράδειγμα εκτέλεσης αλγορίθμου K-Means.

Ο αλγόριθμος K-Means διαχωρίζει τα δεδομένα σε έναν προκαθορισμένο αριθμό συστάδων, ακόμη κι αν αυτός δεν είναι ο βέλτιστος. Για τον ορθότερο προσδιορισμό του πλήθους των συστάδων, χρησιμοποιείται η μέθοδος elbow, κατά την οποία εκτελείται ο αλγόριθμος K-Means για κάποιο εύρος τιμών K και υπολογίζεται για κάθε τιμή το άθροισμα των τετραγώνων των αποστάσεων SSE (Sum of Squares Error) των σημείων κάθε ομάδας. Αν c_i το κέντρο της ομάδας i , τότε:

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} dist(c_i, x)^2$$

Στο τέλος επιλέγεται η τιμή του K για την οποία ελαχιστοποιείται το SSE.

2.1.3 Αλγόριθμοι ιεραρχικής συσταδοποίησης

Οι αλγόριθμοι ιεραρχικής συσταδοποίησης, όπως δηλώνει και το όνομα τους, βασίζονται στη δημιουργία μιας ιεραρχίας εμφωλιασμένων συσταδοποιήσεων. Στο ένα άκρο της ιεραρχίας αυτής βρίσκεται μία ομάδα που περιέχει όλα τα δεδομένα, ενώ

στο άλλο τόσες ομάδες όσο και το πλήθος των δεδομένων. Οι αλγόριθμοι αυτοί χωρίζονται στους αλγορίθμους συγχώνευσης (agglomerative) και στους αλγορίθμους διαίρεσης (divisive).

Οι αλγόριθμοι συγχώνευσης είναι οι πιο διαδεδομένοι από τους δύο. Δεδομένης της αρχικής συσταδοποίησης, όπου κάθε σημείο αποτελεί μία ομάδα, οι αλγόριθμοι αυτοί συγχωνεύουν τις δύο πιο κοντινές ομάδες σε μία. Η διαδικασία αυτή επαναλαμβάνεται όσπου να συγχωνευτούν όλες οι ομάδες σε μία, που θα περιέχει όλα τα δεδομένα. Υπάρχουν διάφοροι μέθοδοι εντοπισμού των δύο κοντινότερων ομάδων:

- Η μέθοδος της απλής μοναδιαίας σύνδεσης (single linkage), που είναι η απόσταση μεταξύ των δύο κοντινότερων σημείων.
- Η μέθοδος της πλήρους σύνδεσης (complete linkage), που είναι η απόσταση μεταξύ των δύο πιο απομακρυσμένων σημείων των ομάδων.
- Ο μέσος όρος των αποστάσεων όλων των σημείων των ομάδων (average link).
- Η απόσταση μεταξύ των κέντρων των ομάδων.
- Η μέθοδος ward, που ελέγχει όλα τα δυνατά ζεύγη και επιλέγει αυτά που όταν ενωθούν θα έχουν το μικρότερο άθροισμα τετραγωνικού σφάλματος.

Οι ιεραρχίες που προκύπτουν από τους αλγορίθμους ιεραρχικής συσταδοποίησης συνήθως απεικονίζονται με τη χρήση του λεγόμενου δενδρογράμματος, στο οποίο φαίνεται τόσο τα επιμέρους επίπεδα συσταδοποίησης, όσο και οι αποστάσεις μεταξύ των ομάδων του κάθε επιπέδου. Ένα παράδειγμα ιεραρχικά συσταδοποιημένων δεδομένων με το αντίστοιχο δενδρόγραμμα τους φαίνεται στο σχήμα 2.3.

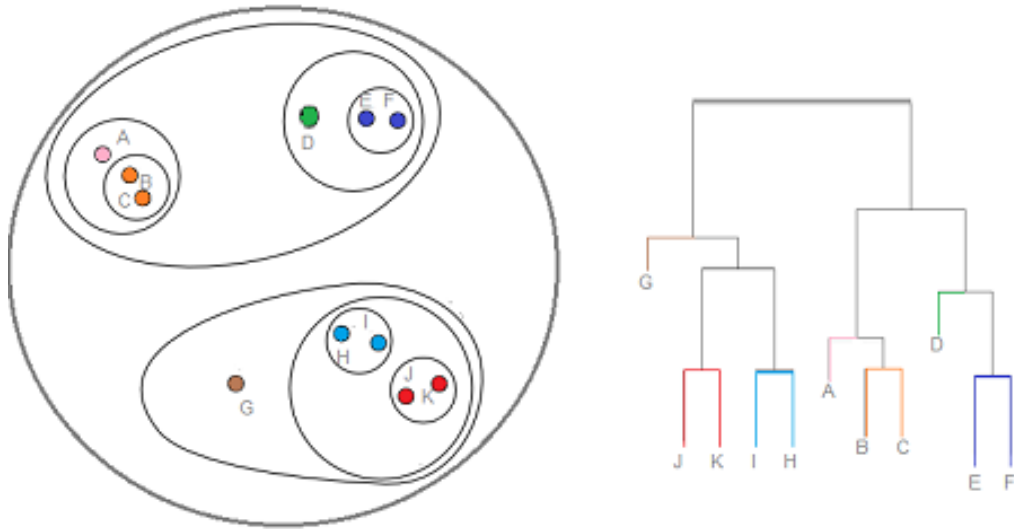
2.1.4 Αλγόριθμος DBSCAN

Ο αλγόριθμος DBSCAN ομαδοποιεί τα δεδομένα βασιζόμενος στην πυκνότητα τους στο χώρο. Η πυκνότητα ενός σημείου καθορίζεται από το πλήθος των γειτονικών του σημείων. Για το λόγο αυτό ορίζονται δύο παράμετροι πριν την εκτέλεση του αλγορίθμου:

- Η ακτίνα ελέγχου γύρω από ένα σημείο (eps).
- Το ελάχιστο πλήθος γειτονικών σημείων εντός eps (minPts).

Ο DBSCAN ταξινομεί σε πρώτη φάση όλα τα σημεία δεδομένων σε μία από τις παρακάτω κατηγορίες:

- Σημεία πυρήνα (core points), που έχουν σε ακτίνα eps τουλάχιστον άλλα minPts σημεία.
- Σημεία ορίου (border points), τα οποία δεν είναι σημεία πυρήνα, αλλά ανήκουν στη γειτονιά κάποιου σημείου πυρήνα.



Σχήμα 2.3: Ενδεικτικό παράδειγμα ιεραρχικής συσταδοποίησης με το αντίστοιχο δένδρογραμμα.

- Σημεία θορύβου (outliers ή points), τα οποία δεν είναι ούτε σημεία πυρήνα, ούτε σημεία ορίου.

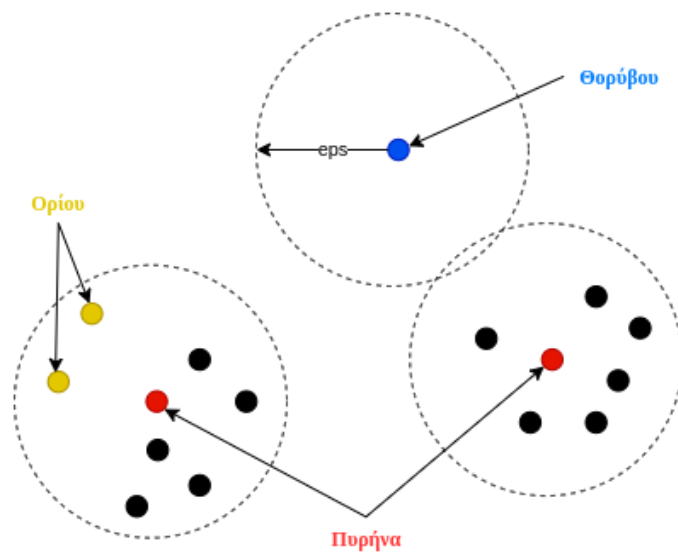
Τα σημεία ελέγχονται πρώτα για το αν είναι πυρήνες, όσα περισσέψουν ελέγχονται για το αν είναι σημεία ορίου και τα υπόλοιπα είναι σημεία θορύβου.

Μετά τον χαρακτηρισμό όλων των σημείων, ο αλγόριθμος δημιουργεί τις συστάδες επιλέγοντας έναν τυχαίο πυρήνα που δεν ανήκει ήδη σε κάποια ομάδα και προσθέτοντας στην ομάδα του όλους τους υπόλοιπους πυρήνες που απέχουν μέχρι eps από αυτόν. Στην ίδια ομάδα προσθέτει αναδρομικά και τους πυρήνες που απέχουν μέχρι eps από αυτούς που μόλις προσέθεσε. Μόλις ολοκληρωθεί αυτή η διαδικασία, επιλέγει τυχαία τον πυρήνα που θα σχηματίσει την επόμενη ομάδα, ώσπου να μην υπάρχει πια πυρήνας που δεν ανήκει σε κάποια ομάδα. Τέλος, προσθέτει τα σημεία ορίου στην ομάδα του κοντινότερου σε αυτά πυρήνα, ενώ αγνοεί τα σημεία θορύβου.

Στο σχήμα 2.4 φαίνεται ένα παράδειγμα χαρακτηρισμού σημείων δεδομένων όπως περιγράφηκε παραπάνω.

2.1.5 Επιλογή αλγορίθμου

Η επιλογή του κατάλληλου αλγορίθμου συσταδοποίησης δεν αποτελεί απλή διαδικασία. Στην πραγματικότητα, κανένας αλγόριθμος συσταδοποίησης δεν αποδίδει πάντοτε καλύτερα από κάποιον άλλον, αλλά η αποτελεσματικότητά της συσταδοποίησης εξαρτάται από την εκάστοτε κατανομή των σημείων στο χώρο. Για παράδειγμα, σε σύνολα δεδομένων όπου τα σημεία κατανέμονται σε μη γραμμικά διαχωρίσιμες συστάδες, η ευκλείδεια απόσταση που χρησιμοποιεί ο K-means δεν θα ανακαλύψει ποτέ



Σχήμα 2.4: Χαρακτηρισμός σημείων δεδομένων με τον αλγόριθμο DBSCAN. Στο συγκεκριμένο παράδειγμα η ακτίνα ελέγχου είναι ίση με ϵ και το \minPts είναι ίσο με 7.

τις κατάλληλες συστάδες. Για την επιλογή του αλγορίθμου συσταδοποίησης που θα χρησιμοποιηθεί στην εργασία, χρειάζεται η εξέταση των χαρακτηριστικών του καθενός από τους προαναφερθέντες. Ο αλγόριθμος K-Means, εκτός από το παραπάνω παράδειγμα, υστερεί των υπολοίπων και ως προς δύο ακόμη βασικά χαρακτηριστικά. Πρώτον, το πλήθος των συστάδων απαιτείται να καθοριστεί από πριν, γεγονός που καθίσταται δύσκολο στη συγκεκριμένη περίπτωση, όπου ο ρόλος της συσταδοποίησης είναι διερευνητικός για την κατανόηση των χωρών ανάλογα με τις συνθήκες. Δεύτερον, η τυχαία αρχικοποίηση των κέντρων στην αρχή του αλγορίθμου οδηγεί σε μία μη ντετερμινιστική συμπεριφορά, η οποία κρίθηκε μη αποδεκτή. Ο αλγόριθμος DBSCAN υπερτερεί των άλλων δύο στο γεγονός ότι δεν θεωρεί όλα τα σημεία μέρος κάποιας ομάδας, αλλά αγνοεί όσα χαρακτηρίζει σημεία θορύβου. Ωστόσο, η διερευνητική χρήση της συσταδοποίησης που αναφέρθηκε παραπάνω οδήγησε στην προτίμηση του συγχωνευτικού αλγορίθμου ιεραρχικής συσταδοποίησης, ώστε να φανερωθούν όποιες κρυφές δομές περιέχονται στα δεδομένα.

2.2 Παλινδρόμηση

2.2.1 Εισαγωγικές Έννοιες

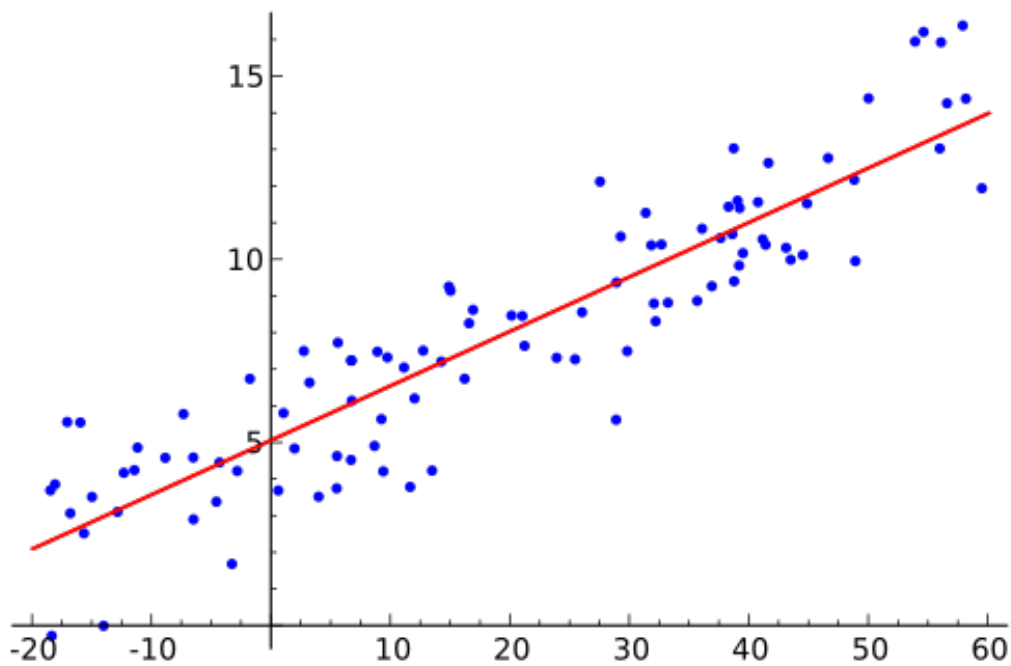
Η παλινδρόμηση είναι μία τεχνική μοντελοποίησης, η οποία χρησιμοποιείται ευρέως για την διερεύνηση της συσχέτισης κάποιας εξαρτημένης μεταβλητής με μία ή περισσότερες ανεξάρτητες μεταβλητές. Η ανάλυση της παλινδρόμησης χρησιμοποιήθηκε αρχικά για την εξαγωγή συσχετίσεων μεταξύ διαφόρων δεδομένων, καθώς και τη μορφή αυτής της συσχέτισης. Αξιοσημείωτη προσπάθεια είναι αυτή του Francis Galton στο βιβλίο *Hereditary Genius*, στο οποίο γίνεται προσπάθεια απόδειξης της κληρονομικότητας της ιδιοφυΐας μέσω στατιστικής ανάλυσης της υπεροχής ανθρώπων σε επαγγέλματα στα οποία διέπρεψαν και οι γονείς τους, έναντι αυτών που οι γονείς τους δεν είχαν σχέση με τον τομέα εργασίας τους [16]. Τα μοντέλα παλινδρόμησης ωστόσο, έχουν βρει εύφορο έδαφος και στον τομέα των προβλέψεων, για την εκτίμηση της εξαρτημένης μεταβλητής, δεδομένων των τιμών των ανεξάρτητων μεταβλητών. Η εκτιμώμενη τιμή της εξαρτημένης μεταβλητής προκύπτει από μία συνάρτηση των ανεξάρτητων μεταβλητών, η οποία καλείται *εξίσωση της παλινδρόμησης*. Στη συνέχεια γίνεται αναφορά στη βασική ιδέα της απλής γραμμικής παλινδρόμησης και σε μοντέλα παλινδρόμησης που χρησιμοποιούνται αργότερα στην εργασία.

2.2.2 Απλή Γραμμική Παλινδρόμηση

Η απλή γραμμική παλινδρόμηση στοχεύει στην εύρεση μιας συνάρτησης που περιγράφει τη σχέση μεταξύ δύο μεταβλητών, της ανεξάρτητης και της εξαρτημένης. Η συνάρτηση αυτή, όπως προδίδει το όνομα, είναι μία ευθεία γραμμή της μορφής $Y = \beta_0 + \beta_1 \cdot X$. Ο συντελεστής a αποτελεί την τεταγμένη του σημείου τομής της ευθείας με τον κατακόρυφο άξονα, ενώ ο συντελεστής b δηλώνει την κλίση της ευθείας, δηλαδή τον ρυθμό μεταβολής της εξαρτημένης μεταβλητής ανα μοναδιαία αύξηση της ανεξάρτητης. Για την εύρεση των συντελεστών a και b χρησιμοποιείται η μέθοδος ελαχίστων τετραγώνων, η οποία ελαχιστοποιεί το άθροισμα $\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$, όπου Y_i η πραγματική τιμή της εξαρτημένης μεταβλητής στο i -οστό δείγμα και \hat{Y}_i η προβλεπόμενη από την εξίσωση παλινδρόμησης. Μετά τον υπολογισμό της εξίσωσης παλινδρόμησης, από την επόμενη τιμή της ανεξάρτητης μεταβλητής, είναι εύκολη η εκτίμηση της τιμής της εξαρτημένης.

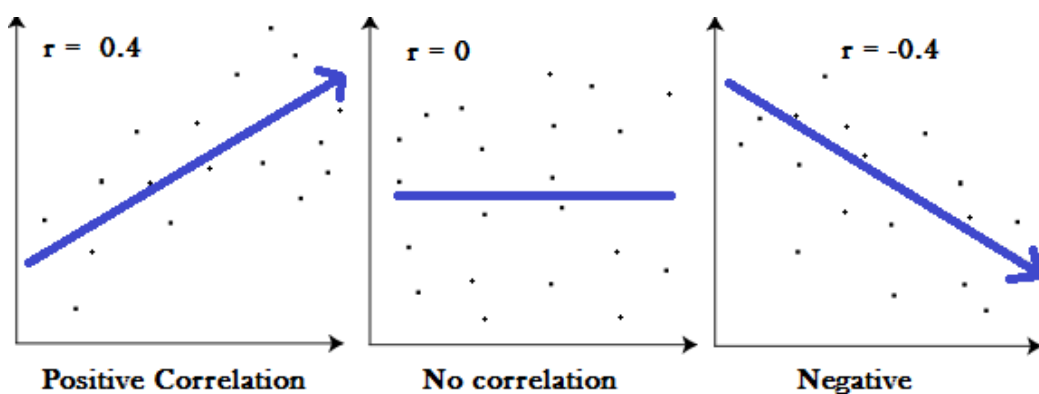
Το μοντέλο της απλής γραμμικής παλινδρόμησης επεκτείνεται στην περίπτωση που υπάρχουν περισσότερες από μία ανεξάρτητες μεταβλητές (X_1, X_2, \dots). Η επέκταση αυτή ονομάζεται πολλαπλή γραμμική παλινδρόμηση και προκύπτει από τη γενίκευση της απλής. Δηλαδή, η εξίσωση της γραμμής του μοντέλου δίνεται από τον τύπο $Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_n \cdot X_n$. Οι τιμές των συντελεστών υπολογίζονται και πάλι με τη μέθοδο των ελαχίστων τετραγώνων και οι προβλέψεις προκύπτουν από την αντικατάσταση των τιμών των ανεξάρτητων μεταβλητών στην εξίσωση.

Η κλίση της ευθείας που προκύπτει από την απλή γραμμική παλινδρόμηση, ωστόσο, αποτελεί ένα σημαντικό μέγεθος στον τομέα της στατιστικής, που ονομάζεται συντελεστής συσχέτισης Pearson (r). Ο συντελεστής συσχέτισης Pearson λαμβάνει



Σχήμα 2.5: Ενδεικτικό παράδειγμα ευθείας που προκύπτει από εφαρμογή της απλής γραμμικής παλινδρόμησης.

τιμές στο διάστημα $[-1, 1]$ και ισούται με την κλίση της ευθείας που προκύπτει από την απλή γραμμική παλινδρόμηση. Όπως φαίνεται στο σχήμα 2.6, όσο πιο θετικά συσχετισμένες είναι οι δύο μεταβλητές, τόσο το r πλησιάζει στο 1, ενώ όσο πιο αρνητικά συσχετισμένες είναι, τόσο αυτό πλησιάζει στο -1.



Σχήμα 2.6: Ενδεικτικά παραδείγματα συντελεστή συσχέτισης Pearson (r).

Ο συντελεστής συσχέτισης r είναι ιδιαίτερα χρήσιμος στις στατιστικές μελέτες για τον υπολογισμό των t και p τιμών. Τα δύο μεγέθη αυτά, χρησιμοποιούνται για

να αποδειχτεί κατά πόσο μία φαινομενική συσχέτιση των δεδομένων βασίζεται σε κάποιο πραγματικό γεγονός, ή αν ευθύνεται στην τύχη. Συγκεκριμένα, το t-value υπολογίζεται από τον τύπο $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$, όπου r ο συντελεστής συσχέτισης Pearson των δύο μεγεθών και n το μέγεθος του δείγματος. Το t-value δείχνει στην ουσία τη διαφορά μεταξύ των δύο μεγεθών, σε σχέση με τη διακύμανση εντός του καθενός μεγέθους. Όσο πιο μικρό είναι το t-value τόσο πιο πολύ μοιάζουν μεταξύ τους τα δύο μεγέθη, ενώ όσο πιο μεγάλο είναι, τόσο πιο πολύ διαφέρουν.

Το t-value χρησιμεύει για την απόρριψη της υπόθεσης ότι τα δεδομένα προέκυψαν κατά τύχη, μέσω του p-value. Το p-value αποτελεί την τιμή της πιθανότητας να προέκυψαν πράγματι τυχαία τα δεδομένα της μελέτης. Όσο πιο μικρή είναι η πιθανότητα αυτή, τόσο πιο έγκυρα είναι τα συμπεράσματα που εξαγώνται από τη συσχέτιση των δεδομένων. Μία κοινώς αποδεκτή τιμή της πιθανότητας αυτής (του p-value), ώστε να απορριφθεί η λεγόμενη κενή υπόθεση είναι το 0.05, δηλαδή 5%. Παλαιότερα, η διεξαγωγή του ελέγχου αυτού γινόταν μέσω πινάκων, οι οποίοι στον κάθετο άξονα περιείχαν τιμές για τα διάφορα μεγέθη του συνόλου δεδομένων και στον οριζόντιο άξονα περιείχαν διάφορες τιμές για το p-value. Το κελί που αντιστοιχεί στο μέγεθος του δείγματος και το μέγιστο αποδεκτό p-value περιέχει την ελάχιστη τιμή που θα πρέπει να έχει το t-value, προκειμένου να θεωρηθεί η συσχέτιση των δεδομένων *στατιστικά σημαντική*. Τα τελευταία χρόνια, ο υπολογισμός του p-value για το εκάστοτε σύνολο δεδομένων και t-value γίνεται μέσω λογισμικού, οπότε ελέγχεται αμέσως το κατά πόσο αυτό είναι μικρότερο από το ανώτατο όριο. Αν κάτι τέτοιο δεν ισχύει, δηλαδή αν το p-value είναι μεγαλύτερο του 0.05 (συνήθως), τότε δεν μπορεί να αποκλιστεί ότι η οποιαδήποτε συσχέτιση των μεγεθών προέκυψε τυχαία.

2.3 Ανάλυση Κύριων Συνιστωσών (PCA)

Η Ανάλυση Κύριων Συνιστωσών (Primary Component Analysis ή PCA) είναι μία διαδικασία μείωσης της διαστατικότητας ενός συνόλου πιθανώς συσχετισμένων μεταβλητών, διατηρώντας όσο το δυνατόν περισσότερη πληροφορία. Αυτό επιτυγχάνεται μέσω τις προβολής των σημείων δεδομένων σε κάποιες ασυσχέτιστες μεταξύ τους κύριες συνιστώσες. Στην πράξη, παράγονται νέα χαρακτηριστικά, τα οποία αποτελούν γραμμικό συνδυασμό των αρχικών. Η PCA ως μέθοδος εμφανίστηκε το 1901 από τον K. Pearson [24], ενώ έλαβε το όνομα της από τον H. Hotelling το 1936 [17]. Η χρησιμότητα της εκτείνεται από την ανάλυση δεδομένων, επιτρέποντας την οπτικοποίηση πολυδιάστατων συνόλων δεδομένων προβάλλοντας τα σε λιγότερες, κατανοητές από τον άνθρωπο, διαστάσεις (συνήθως δύο ή τρεις), ως την εξαγωγή χαρακτηριστικών και τη μηχανική μάθηση, επιταχύνοντας την εκπαίδευση μοντέλων σε μεγάλα σύνολα δεδομένων.

Το πρώτο βήμα πριν την ανάλυση του συνόλου δεδομένων στις κύριες συνιστώσες είναι η αποκλιμάκωση των χαρακτηριστικών του. Το βήμα αυτό είναι απαραίτητο για την επιτυχημένη ανάλυση κύριων συνιστωσών, καθώς στην περίπτωση που ένα χαρακτηριστικό λαμβάνει τιμές σε πολύ μεγάλο σχετικά εύρος, θα μονοπωλήσει έναντι

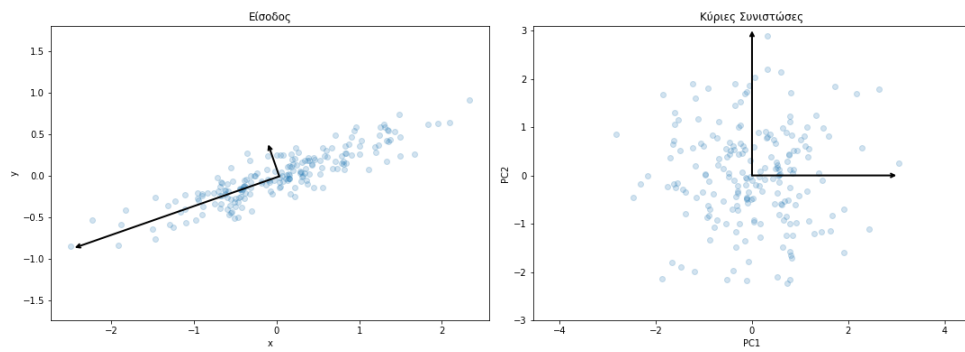
των υπολοίπων στην ανάλυση. Ο τρόπος για να γίνει η αποκλιμάκωση αυτή είναι η λεγόμενη *κανονικοποίηση*, η οποία στην ουσία είναι η αφαίρεση του μέσου όρου των τιμών ενός χαρακτηριστικού από την καθεμία τιμή και η διαίρεση της με την τυπική απόκλιση ($z = \frac{\text{value} - \text{mean}}{\text{standard deviation}}$). Με αυτό τον τρόπο, όλα τα χαρακτηριστικά του συνόλου δεδομένων λαμβάνουν τιμές στο εύρος $[-1, 1]$.

Στη συνέχεια, για την εύρεση των κύριων συνιστωσών, χρησιμοποιείται ο *πίνακας συνδιακύμανσης*, ο οποίος απαρτίζεται από τις ανά δύο συνδιακυμάνσεις των χαρακτηριστικών του συνόλου δεδομένων. Για παράδειγμα, για ένα σύνολο δεδομένων με τρία χαρακτηριστικά x, y, z , ο πίνακας συνδιακύμανσης θα είναι ο εξής:

$$\begin{bmatrix} \text{Cov}(x, x) & \text{Cov}(x, y) & \text{Cov}(x, z) \\ \text{Cov}(y, x) & \text{Cov}(y, y) & \text{Cov}(y, z) \\ \text{Cov}(z, x) & \text{Cov}(z, y) & \text{Cov}(z, z) \end{bmatrix}$$

όπου $\text{Cov}(x, y) = E[(x - E[x])(y - E[y])]$. Ο πίνακας συνδιακύμανσης ενός συνόλου δεδομένων με n χαρακτηριστικά, εξ ορισμού είναι τετραγωνικός με $n \times n$ διαστάσεις και έχει n ιδιοδιανύσματα και ιδιοτιμές. Η εύρεση των ιδιοδιανυσμάτων και των ιδιοτιμών αυτών αποτελεί το βασικότερο βήμα της PCA. Το κάθε ιδιοδιάνυσμα αποτελεί τον πίνακα των συντελεστών με τους οποίους συνεισφέρει το αντίστοιχο χαρακτηριστικό του συνόλου δεδομένων στην αντίστοιχη κύρια συνιστώσα. Ο τρόπος κατάταξης των ιδιοδιανυσμάτων στις κύριες συνιστώσες, γίνεται κατά φθίνουσα σειρά των ιδιοτιμών που αντιστοιχούν σε αυτά. Δηλαδή, το ιδιοδιάνυσμα που αντιστοιχεί στην μεγαλύτερη ιδιοτιμή, αποτελεί τους συντελεστές των χαρακτηριστικών του αρχικού συνόλου δεδομένων για την πρώτη κύρια συνιστώσα, αυτό με τη δεύτερη μεγαλύτερη ιδιοτιμή για τη δεύτερη κύρια συνιστώσα και ούτω καθ' εξής. Το ποσοστό της διακύμανσης του αρχικού συνόλου δεδομένων που διατηρεί η κάθε κύρια συνιστώσα υπολογίζεται διαιρώντας την αντίστοιχη ιδιοτιμή της με το άθροισμα όλων των ιδιοτιμών ($\frac{\lambda_i}{\sum_{j=1}^n \lambda_j}$).

Το ποσοστό της διακύμανσης που διατηρείται με την κάθε κύρια συνιστώσα βοηθάει στον καθορισμό της διαστατικότητας του νέου συνόλου δεδομένων. Ωστόσο, όταν η PCA χρησιμοποιείται για τη μείωση της διαστατικότητας του συνόλου δεδομένων ώστε να αυτό να μπορεί να απεικονιστεί, τότε η επιλογή αυτή περιορίζεται στις δύο ή τρεις κύριες συνιστώσες. Το τελικό σύνολο δεδομένων, προκύπτει από τον πολλαπλασιασμό του αρχικού, με τον πίνακα των ιδιοδιανυσμάτων των όσων κύριων συνιστωσών επιλέχθηκαν. Το αποτέλεσμα που προκύπτει είναι μία γραμμική απεικόνιση των αρχικών δεδομένων σε ένα νέο ορθογωνικό σύστημα αξόνων με την επιθυμητή διαστατικότητα, διατηρώντας όσο το δυνατόν περισσότερη πληροφορία.



Σχήμα 2.7: Ενδεικτικό παράδειγμα εφαρμογής της Ανάλυσης Κύριων Συνιστωσών (PCA) σε δεδομένα τριών διαστάσεων. Στο αρχικό διάγραμμα, η τρίτη διάσταση αναπαρίστανται από τη διαφάνεια των σημείων, ενώ δίπλα φαίνεται το αποτέλεσμα της ανάλυσης σε δύο κύριες συνιστώσες.

Κεφάλαιο 3

Ανάλυση Συγκεντρωτικών Δεδομένων

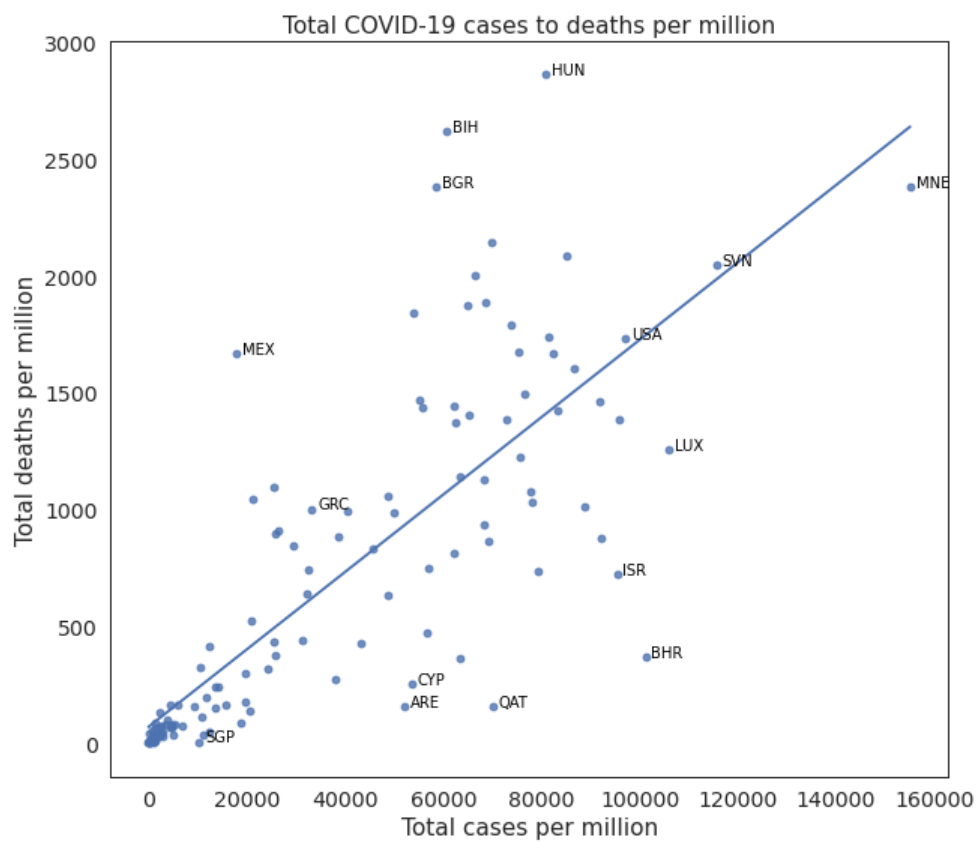
3.1 Καθορισμός Συνόλου Δεδομένων

Όπως αναφέρθηκε στην Εισαγωγή, σκοπός της παρούσας εργασίας είναι η μελέτη των δεδομένων που αφορούν την πανδημία COVID-19 για τον καθορισμό των παραγόντων που συμβάλλουν ή εμποδίζουν την εξάπλωση της.

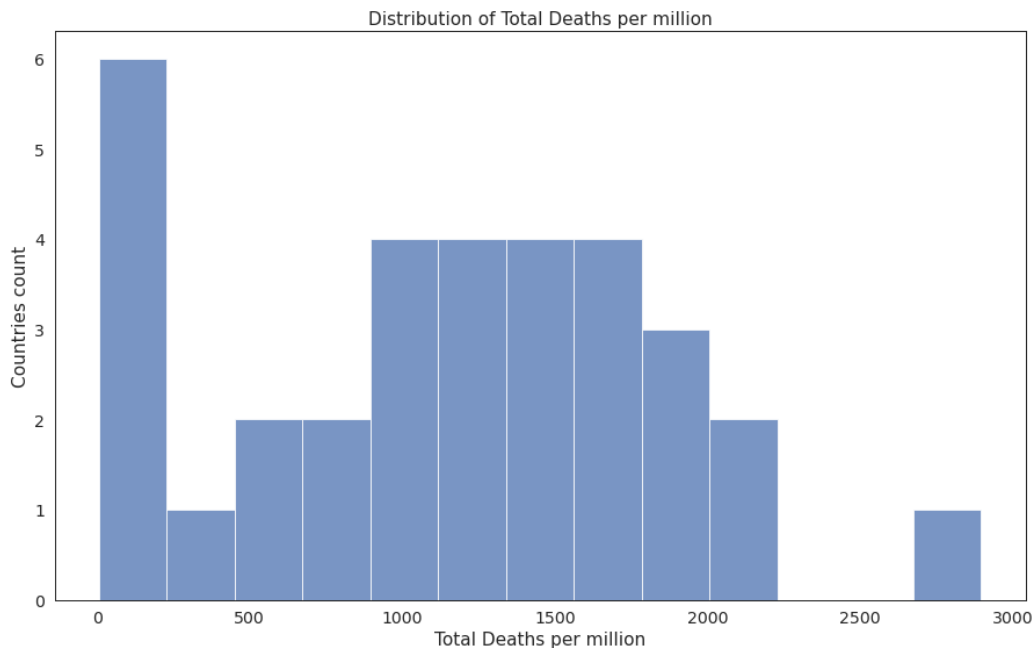
3.1.1 Θάνατοι που ευθύνονται στην COVID-19

Ενδεικτικά μεγέθη για την κατάσταση της πανδημίας σε επίπεδο χώρας μπορούν να θεωρηθούν τα κρούσματα COVID-19 και οι θάνατοι που ευθύνονται στην COVID-19. Από τα δύο, το δεύτερο κρίθηκε το πιο αντικειμενικό, καθώς τα κρούσματα COVID-19 εξαρτώνται σε μεγάλο βαθμό από την πολιτική ελέγχου κάθε χώρας (π.χ. προληπτικοί έλεγχοι, ανίχνευση επαφών), οπότε εμπεριέχεται θόρυβος. Ο παραπάνω ισχυρισμός επιβεβαιώνεται και από το διάγραμμα 3.1, στο οποίο φαίνεται ότι αν και τα κρούσματα COVID-19 είναι συσχετισμένα με τους θανάτους που ευθύνονται στην COVID-19, δεν είναι λίγες οι περιπτώσεις χωρών που με λιγότερα κρούσματα έχουν περισσότερους θανάτους. Αυτό αποτελεί ένδειξη ότι τα συνολικά κρούσματα, αν και σημαντικός, δεν είναι ο μοναδικός παράγοντας που καθορίζει την εξέλιξη της πανδημίας σε μία χώρα. Και τα δύο μεγέθη αναπαριστώνται αφού αναχθούν ανά εκατομμύριο κατοίκους, για να εξομαλυνθεί η μεγάλη διαφορά των πολυπληθών χωρών έναντι αυτών με μικρότερο πληθυσμό. Τα συγκεκριμένα δεδομένα τόσο για τα κρούσματα όσο και για τους θανάτους συλλέχθηκαν από το αποθετήριο του OxCGRT στο GitHub [3].

Όπως είναι γνωστό και φαίνεται και στο παραπάνω διάγραμμα, οι χώρες παρουσίασαν πολύ διαφορετικές εικόνες σχετικά με τους θανάτους που ευθύνονται στην COVID-19, ανεξάρτητα από το μέγεθος του πληθυσμού τους. Για το λόγο αυτό κρίθηκε απαραίτητη η διερεύνηση των χαρακτηριστικών αυτών που οδήγησαν στην τόσο μεγάλη διακύμανση ($Var = 505781.6$). Στο διάγραμμα 3.2 φαίνεται ο τρόπος



Σχήμα 3.1: Κρούσματα COVID-19 ανά εκατομμύριο προς θανάτους που ευθύνονται στην COVID-19 ανά εκατομμύριο ανά χώρα. Για λόγους αναγνωσιμότητας, σημειώνονται λίγες μόνο χώρες που ανήκουν στα άκρα του συνόλου.

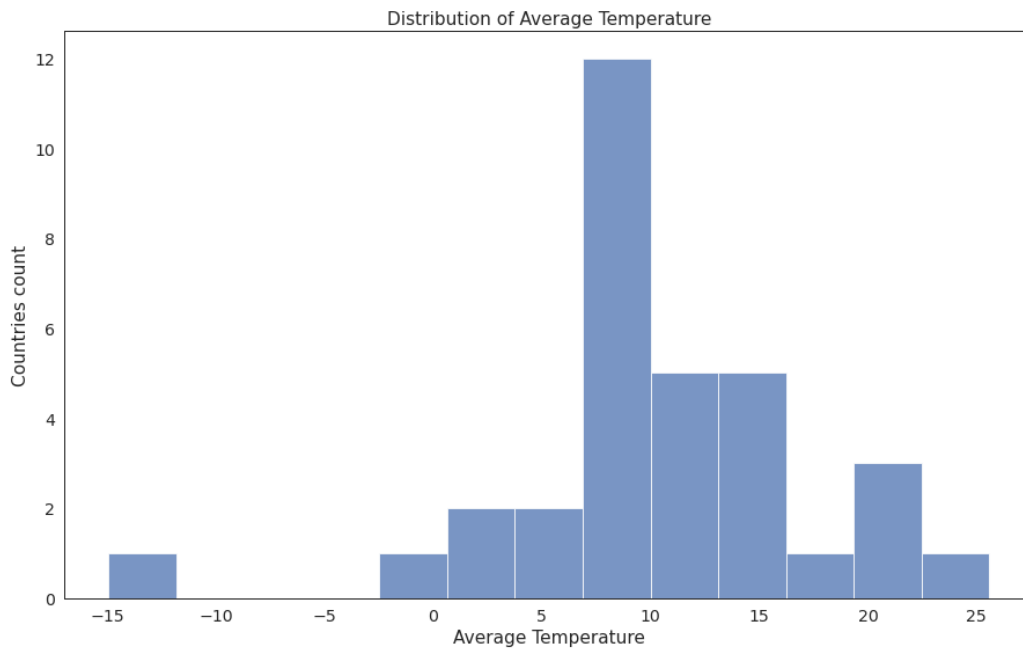


Σχήμα 3.2: Κατανομή συνολικών θανάτων ανα εκατομμύριο που ευθύνονται στην COVID-19 των χωρών.

κατανομής των συνολικών θανάτων ανά εκατομμύριο ενός υποσυνόλου των χωρών. Το υποσύνολο αυτό προκλήθηκε από την έλλειψη στοιχείων που θα περιγραφούν στη συνέχεια από τις υπόλοιπες χώρες και αποτελείται από τις εξής (33) χώρες: Αυστραλία, Αυστρία, Βέλγιο, Γαλλία, Γερμανία, Δανία, Ελβετία, Ελλάδα, Εσθονία, Ηνωμένο Βασίλειο, Ιαπωνία, Ιρλανδία, Ισλανδία, Ισπανία, Ισραήλ, Ιταλία, Καναδάς, Κολομβία, Λετονία, Λιθουανία, Λουξεμβούργο, Μεξικό, Νέα Ζηλανδία, Νορβηγία, Ολλανδία, Ουγγαρία, Πολωνία, Πορτογαλία, Σλοβενία, Σουηδία, Τουρκία, Φινλανδία και Χιλή. Οι Ηνωμένες Πολιτείες της Αμερικής δεν συμπεριλήφθηκαν στις χώρες αυτές, διότι οι Πολιτείες διαφέρουν μεταξύ τους σε μεγάλο βαθμό σε πλήθος χαρακτηριστικών. Επομένως, η συγχώνευση τους σε μία χώρα δεν θα παρήγαγε αντιπροσωπευτικά αποτελέσματα για τις περισσότερες εξ αυτών. Ωστόσο, τα συμπεράσματα που θα προκύψουν από την ανάλυση, θα επαληθευτούν στη συνέχεια για την περίπτωση των 50 Πολιτειών.

3.1.2 Κλιματικά Στοιχεία

Ένας παράγοντας που έχει παρατηρηθεί ότι συμβάλλει στην επιβίωση του SARS κορωνοϊού σε επιφάνειες είναι η θερμοκρασία και η σχετική υγρασία. Συγκεκριμένα, ο ιός φαίνεται να διατηρείται σε επιφάνειες έως και πέντε ημέρες σε θερμοκρασίες 22-25°C και 40-50% σχετική υγρασία. Η ύπαρξη των ανωτέρω συνθηκών κατά την άνοιξη στις υποτροπικές περιοχές και σε κλιματιζόμενους χώρους επιταχύνει τη διασπορά του



Σχήμα 3.3: Κατανομή μέσης θερμοκρασίας σε °C για το έτος 2013 των χωρών.

κορωνοϊού [7].

Στην παρούσα εργασία χρησιμοποιήθηκαν δεδομένα από το Berkley Earth[2]. Τα δεδομένα αυτά βρίσκονται στη μορφή της μέσης θερμοκρασίας ανά μήνα για κάθε χώρα για κάθε έτος μέχρι και το 2013. Επομένως χρειάστηκε επεξεργασία ώστε να εξαχθεί η μέση θερμοκρασία της κάθε χώρας ενδιαφέροντος για το έτος του 2013, που είναι το πιο πρόσφατο διαθέσιμο. Οι τιμές της μέσης θερμοκρασίας κυμαίνονται από τους -15°C έως τους 25°C (για τη Δανία και την Κολομβία αντίστοιχα), αλλά οι περισσότερες χώρες εμφανίζουν μέση θερμοκρασία στο εύρος $7-15^{\circ}\text{C}$ όπως φαίνεται και στο διάγραμμα 3.3.

3.1.3 Μέτρα Προστασίας και Στήριξης

Η ραγδαία εξέλιξη της πανδημίας COVID-19 οδήγησε τις κυβερνήσεις ανά τον κόσμο να λάβουν πληθώρα μέτρων για την προστασία και την στήριξη των πολιτών. Τα μέτρα αυτά εκτείνονται από το κλείσιμο σχολείων και εργασιακών χώρων και την απαγόρευση μετακινήσεων, ως την έκτακτη ενίσχυση των συστημάτων υγείας, την εφαρμογή πολιτικών ανίχνευσης κρουσμάτων και πιο πρόσφατα το σχεδιασμό εμβολιασμών. Εξαιτίας της μεγάλης ποικιλίας διαφορετικών προσεγγίσεων για την προστασία και τη στήριξη από τον κορωνοϊό, αναπτύχθηκαν πολλά συστήματα παρακολούθησης και ερμηνείας των προσεγγίσεων αυτών, με σκοπό την εξαγωγή συμπερασμάτων για να βοηθήσουν τις κυβερνήσεις να σχεδιάσουν πιο εμπειριστατωμένα την πολιτική τους. Τρία από τα πιο γνωστά συστήματα αυτά είναι το Oxford COVID-19 Government

Response Tracker (OxCGRT)[27], το CoronaNet[4] και το Complexity Science Hub COVID-19 Control Strategies List (CCCSL)[1]. Στην παρούσα εργασία χρησιμοποιήθηκε το OxCGRT, καθώς υπερτερεί του CCCSL στο πλήθος των χωρών που παρακολουθεί (180 έναντι σε 56) και του CoronaNet καθώς οι δείκτες του φαίνεται να περιέχουν επιπλέον πληροφορίες για σύγκριση και ανάλυση των χωρών. Επιπλέον, το OxCGRT είναι το μοναδικό σύνολο δεδομένων που παρέχει συνδέσμους σε όλες τις επίσημες πηγές των δεδομένων του, παρέχοντας έτσι μεγαλύτερη αξιοπιστία.

Το OxCGRT συλλέγει και διατηρεί τιμές για 20 συντελεστές κυβερνητικής ανταπόκρισης σε εθνικό επίπεδο για 180 χώρες, καθώς επίσης και σε υποεθνικό επίπεδο για τη Βραζιλία, τις Ηνωμένες Πολιτείες της Αμερικής, το Ηνωμένο Βασίλειο και τον Καναδά. Οι συντελεστές αυτοί ενημερώνονται σε καθημερινή βάση και ανάλογα με τη σημασία τους λαμβάνουν σειριακές τιμές που υποδεικνύουν τη σοβαρότητα ή την ένταση των μέτρων, αριθμητικές τιμές για τη μέτρηση ορισμένων μεγεθών ή κείμενο για την καταγραφή χρήσιμων πληροφοριών. Οι συντελεστές χωρίζονται σε τέσσερις βασικές κατηγορίες:

- C - containment and closure policies (μέτρα περιορισμών και κλεισίματος)
- E - economic policies (οικονομικά μέτρα)
- H - health system policies (μέτρα ενίσχυσης του συστήματος υγείας)
- M - miscellaneous (διάφορα μέτρα)

Στον πίνακα 3.1 παρουσιάζονται οι 20 αυτοί συντελεστές μαζί με μία σύντομη περιγραφή της σημασίας τους και των τιμών που λαμβάνουν.

Όνομα	Περιγραφή	Τιμές
C1	Κλείσιμο σχολείων	<p>0 - κανένα μέτρο</p> <p>1 - προτεινόμενο κλείσιμο ή όλα τα σχολεία ανοιχτά με σημαντικές τροποποιήσεις στον τρόπο λειτουργίας</p> <p>2 - απαιτούμενο κλείσιμο κάποιων βαθμίδων</p> <p>3 - απαιτούμενο κλείσιμο όλων των βαθμίδων</p>

C2	Κλείσιμο χώρων εργασίας	<p>0 - κανένα μέτρο</p> <p>1 - προτεινόμενο κλείσιμο και τηλεργασία ή όλοι οι χώροι εργασίας ανοιχτοί με σημαντικές τροποποιήσεις στον τρόπο λειτουργίας</p> <p>2 - απαιτούμενο κλείσιμο και τηλεργασία για κάποιους τομείς</p> <p>3 - απαιτούμενο κλείσιμο και τηλεργασία για όλους τους τομείς εκτός από τους βασικούς εργαζόμενους</p>
C3	Ακύρωση δημοσίων εκδηλώσεων	<p>0 - κανένα μέτρο</p> <p>1 - προτεινόμενη ακύρωση</p> <p>2 - απαιτούμενη ακύρωση</p>
C4	Περιορισμός συναθροίσεων	<p>0 - κανένα μέτρο</p> <p>1 - περιορισμός πολύ μεγάλων συναθροίσεων (>1000 ατόμων)</p> <p>2 - περιορισμός συναθροίσεων μεταξύ 101-1000 ατόμων</p> <p>3 - περιορισμός συναθροίσεων μεταξύ 11-100 ατόμων</p> <p>4 - περιορισμός μικρών συναθροίσεων (<10 ατόμων)</p>
C5	Κλείσιμο Μέσων Μαζικής Μεταφοράς	<p>0 - κανένα μέτρο</p> <p>1 - προτεινόμενο κλείσιμο ή αραιώση δρομολογίων και μέσων</p> <p>2 - απαιτούμενη κλείσιμο ή απαγόρευση χρήσης τους από την πλειοψηφία των πολιτών</p>

C6	Κατ' οίκον περιορισμός	<ul style="list-style-type: none"> 0 - κανένα μέτρο 1 - προτεινόμενος κατ' οίκον περιορισμός 2 - απαιτούμενος κατ' οίκον περιορισμός εκτός από περιπτώσης άθλησης, αγοράς τροφίμων και αναγκαίων εξόδων 3 - απαιτούμενος κατ' οίκον περιορισμός με ελάχιστες εξαιρέσεις (π.χ. μία φορά τη βδομάδα ή ένα άτομο τη φορά)
C7	Περιορισμός εγχώριων μετακινήσεων	<ul style="list-style-type: none"> 0 - κανένα μέτρο 1 - προτεινόμενος περιορισμός μετακινήσεων μεταξύ περιοχών/πόλεων 2 - εφαρμογή περιορισμού εγχώριων μετακινήσεων
C8	Έλεγχος διεθνών ταξιδιών	<ul style="list-style-type: none"> 0 - κανένα μέτρο 1 - διαλογή αφίξεων 2 - καραντίνα αφίξεων από όλες τις περιοχές 3 - απαγόρευση αφίξεων από κάποιες περιοχές 4 - απαγόρευση αφίξεων από όλες τις περιοχές ή πλήρες κλείσιμο συνόρων
E1	Οικονομική υποστήριξη σε άτομα που έχασαν τη δουλειά τους ή δεν μπορούν να εργαστούν	<ul style="list-style-type: none"> 0 - καμία υποστήριξη 1 - παροχή 50% του χαμένου μισθού 2 - παροχή 50% του χαμένου μισθού

E2	Ελάφρυνση οικονομικών υποχρεώσεων νοικοκυριών	<p>0 - καμία ελάφρυνση</p> <p>1 - ασθενής ελάφρυνση, σε ένα είδος εξόδων</p> <p>2 - ευρεία ελάφρυνση</p>
E3	Οικονομική ενίσχυση	Καταγράφεται το ύψος της οικονομικής ενίσχυσης σε USD
E4	Υποστήριξη σε άλλες χώρες σχετική με COVID-19	Καταγράφεται το ύψος της οικονομικής υποστήριξης σε USD
H1	Εκστρατείες ενημέρωσης κοινού	<p>0 - καμία εκστρατεία ενημέρωσης</p> <p>1 - επίσημη προσοχής από κυβερνητικές αρχές</p> <p>2 - συντονισμένη εκστρατεία ενημέρωσης (σε παραδοσιακά και κοινωνικά μέσα)</p>
H2	Πολιτική διενέργησης ελέγχων	<p>0 - καμία πολιτική διενέργησης ελέγχων</p> <p>1 - έλεγχος μόνο σε όσους έχουν συμπτώματα και πληρούν ορισμένες προϋποθέσεις (π.χ. σημαντικοί εργαζόμενοι, νοσηλευμένοι, όσοι ήρθαν σε επαφή με επιβεβαιωμένο κρούσμα)</p> <p>2 - έλεγχος όσων παρουσιάζουν συμπτώματα COVID-19</p> <p>3 - ανοιχτοί δημόσιοι έλεγχοι (διαθέσιμοι και σε ασυμπτωματικά άτομα)</p>
H3	Ιχνηλάτηση επαφών	<p>0 - καμία ιχνηλάτηση</p> <p>1 - περιορισμένη ιχνηλάτηση, όχι για όλα τα κρούσματα</p> <p>2 - εξωνυχιστική ιχνηλάτηση για όλα τα επιβεβαιωμένα κρούσματα</p>
H4	Έκτακτη ενίσχυση συστήματος υγείας	Καταγράφεται το ύψος της οικονομικής ενίσχυσης σε USD

H5	Επένδυση στην ανάπτυξη εμβολίων	Καταγράφεται το ύψος της επένδυσης σε USD
H6	Μάσκες προστασίας	<ul style="list-style-type: none"> 0 - καμία πρόβλεψη 1 - πρόταση χρήσης τους 2 - υποχρεωτική χρήση τους σε συγκεκριμένα μέρη/περιπτώσεις όπου η τήρηση αποστάσεων είναι αδύνατη 3 - υποχρεωτική χρήση τους σε όλα τα μέρη/περιπτώσεις όπου η τήρηση αποστάσεων είναι αδύνατη 4 - υποχρεωτική χρήση τους σε κάθε περίπτωση εκτός του σπιτιού
H7	Πολιτική εμβολιασμών	<ul style="list-style-type: none"> 0 - κανένας εμβολιασμός 1 - εμβολιασμός ΜΙΑΣ από τις ομάδες: σημαντικοί εργαζόμενοι/ ευάλωτες ομάδες/ ηλικιωμένοι 2 - εμβολιασμός ΔΥΟ από τις ομάδες: σημαντικοί εργαζόμενοι/ ευάλωτες ομάδες/ ηλικιωμένοι 3 - εμβολιασμός ΟΛΩΝ από τις ομάδες: σημαντικοί εργαζόμενοι/ ευάλωτες ομάδες/ ηλικιωμένοι 4 - εμβολιασμός για όλες τις ομάδες και επιπλέον διαθεσιμότητα για κάποιες ακόμη 5 - ολική διαθεσιμότητα

H8	Προστασία ηλικιωμένων ομάδων	<p>0 - καμία προστασία</p> <p>1 - προτεινόμενη αποστασιοποίηση, προσωπική υγιεινή και περιορισμός επισκέψεων στους οίκους ευγηρίας ή/και στα σπίτια των ηλικιωμένων</p> <p>2 - μερικός περιορισμός για αποστασιοποίηση, προσωπική υγιεινή και περιορισμό επισκέψεων στους οίκους ευγηρίας ή/και στα σπίτια των ηλικιωμένων</p> <p>2 - εκτεταμένος περιορισμός για αποστασιοποίηση, προσωπική υγιεινή και περιορισμό επισκέψεων στους οίκους ευγηρίας ή/και στα σπίτια των ηλικιωμένων, όπως απαγόρευση εξόδου πλην ελαχίστων εξαιρέσεων και απαγόρευση εξωτερικών επισκεπτών</p>
M1	Καταγραφή μέτρων που δεν ταιριάζουν στις υπόλοιπες κατηγορίες	Κείμενο που περιγράφει τα εκάστοτε μέτρα

Πίνακας 3.1: Επεξήγηση συντελεστών που καταγράφει το OxCGRIT.

Για την ευκολότερη παρακολούθηση των κυβερνητικών αντιδράσεων, το OxCGRIT υπολογίζει τέσσερις δείκτες οι οποίοι συνδυάζουν τις τιμές μιας συγκεκριμένης οικογένειας συντελεστών και παρέχουν μία γενικότερη εικόνα της αντίδρασης για τον εκάστοτε τομέα. Οι δείκτες αυτοί είναι:

- General Response Index (όλες οι κατηγορίες συντελεστών)
- Stringency Index (μέτρα κλεισίματος και περιορισμών "lockdown")
- Containment and Health Index (μέτρα κλεισίματος, περιορισμών και ενίσχυσης του συστήματος υγείας)
- Economic Support Index (μέτρα οικονομικής ενίσχυσης)

Οι συντελεστές από τους οποίους καθορίζεται ο κάθε δείκτης παρουσιάζονται στον πίνακα 3.2.

Οι συντελεστές του κάθε δείκτη επιδρούν προσθετικά και με ίσο βάρος σε αυτόν. Η τιμή του δείκτη υπολογίζεται παίρνοντας την τιμή του κάθε συντελεστή, αφαιρώντας μισό πόντο σε περίπτωση που το μέτρο εφαρμόζεται στοχευμένα αντί για γενικευμένα και έπειτα αλλάζει κλίμακα ώστε να βρεθεί μεταξύ του 0 και του 100. Οι διαφορετικοί

Συντελεστές	GRI	CHI	SI	EI
C1	✓	✓	✓	
C2	✓	✓	✓	
C3	✓	✓	✓	
C4	✓	✓	✓	
C5	✓	✓	✓	
C6	✓	✓	✓	
C7	✓	✓	✓	
C8	✓	✓	✓	
E1	✓			✓
E2	✓			✓
E3				
E4				
H1	✓	✓	✓	
H2	✓	✓		
H3	✓	✓		
H4				
H5				
H6	✓	✓		
H7	✓	✓		
H8	✓	✓		
M1				

Πίνακας 3.2: Συντελεστές των δεικτών του ΟxCGRT.

συντελεστές του δείκτη στη συνέχεια αθροίζονται και ο μέσος όρος τους λαμβάνεται ως η τιμή του δείκτη για τη συγκεκριμένη χώρα τη συγκεκριμένη ημερομηνία. Ο υπολογισμός αυτός περιγράφεται και με τον παρακάτω τύπο, όπου k το πλήθος των συντελεστών που απαρτίζουν το δείκτη.

$$Index = \frac{1}{k} \sum_{j=1}^k I_j$$

I_j είναι το σκορ του κάθε συντελεστή, το οποίο προκύπτει με τον τρόπο που περιγράψαμε από τον τύπο

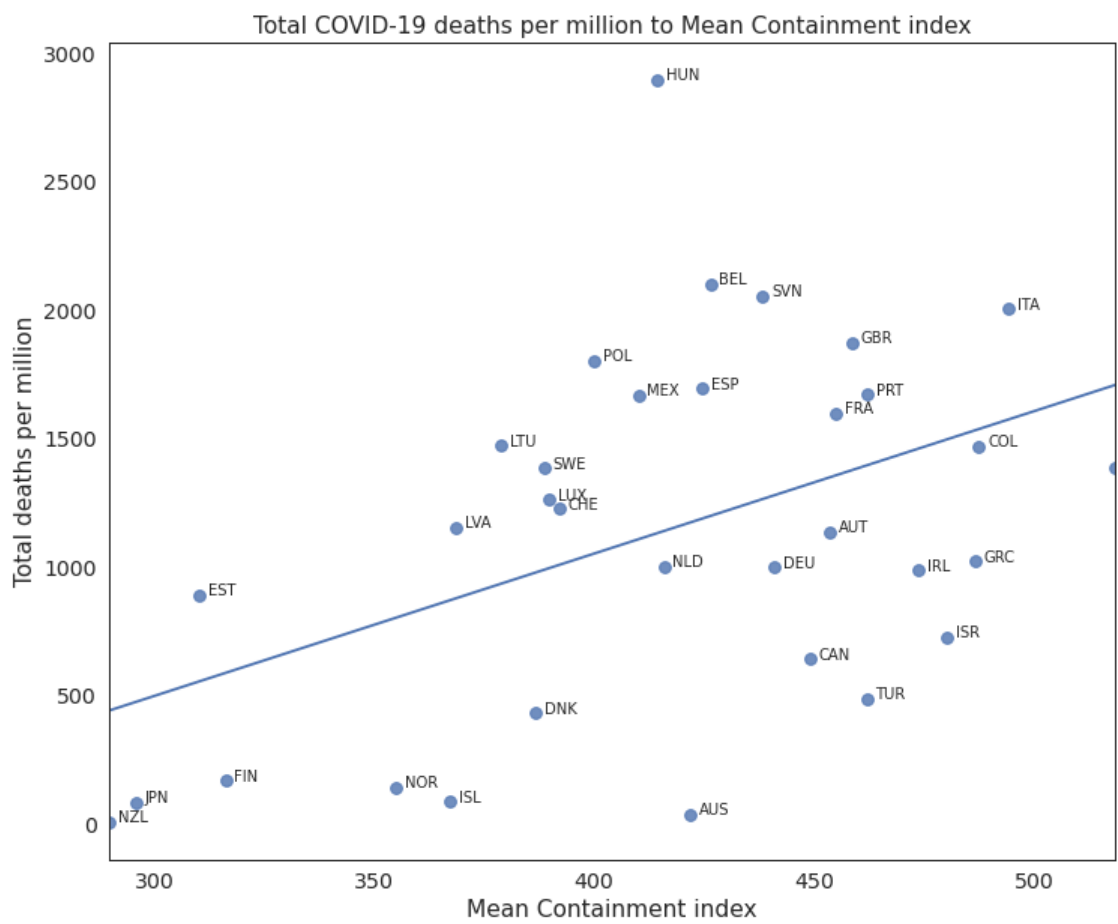
$$I_j = 100 \frac{v_{j,t} - 0.5(F_j - f_{j,t})}{N_j}$$

όπου $v_{j,t}$ είναι η τιμή του συντελεστή j την t ημερομηνία, το F_j παίρνει τιμή 1 αν ο συντελεστής έχει μεταβλητή σημαίας (flag) και $f_{j,t}$ η τιμή της μεταβλητής αυτής την t ημερομηνία, ενώ N_j είναι η μέγιστη τιμή που παίρνει ο συντελεστής.

Η απλοϊκή αυτή προσέγγιση στον υπολογισμό των δεικτών δημιουργεί την υποψία ότι χάνεται σημαντική πληροφορία, ωστόσο παρέχεται μεγαλύτερη διαφάνεια και τα δεδομένα γίνονται ευκολότερα στην κατανόηση τους. Επιπλέον, δείκτες παραγόμενοι με εφαρμογή της θεωρίας απόκρισης ερωτήματος (IRT) παρουσίασαν πολύ υψηλή συσχέτιση με τους απλοϊκούς δείκτες ($r = 0.98$), επομένως καταρρίπτεται η παραπάνω υποψία.

Από τους δείκτες που προσφέρει το OxCGRT, την μεγαλύτερη συσχέτιση με τους θανάτους από COVID-19 εμφανίζει η μέση τιμή του Stringency Index ($r = 0.51$). Ωστόσο, παρόμοια συσχέτιση εμφανίζει και ο Containment and Health Index ($r = 0.46$) και θεωρείται πλέον κατάλληλος καθώς οι επιπλέον συντελεστές του σχετίζονται με την προστασία κατά της διάδοσης της COVID-19 και των θανάτων που σχετίζονται με αυτή. Η συσχέτιση αυτή φαίνεται και στο διάγραμμα σκέδασης 3.4.

Οι χώρες έλαβαν διαφορετικές προσεγγίσεις κατά τη διάρκεια της πανδημίας, όσον αφορά τα μέτρα περιορισμών και ενίσχυσης του συστήματος υγείας. Το γεγονός αυτό αντικατοπτρίζεται στο διάγραμμα 3.5. Το διάγραμμα αυτό εμφανίζει ενδιαφέρον περαιτέρω μελέτης, για να εξεταστεί το κατά πόσο οι διαφορετικές προσεγγίσεις των χωρών στα μέτρα προστασίας αντικατοπτρίζονται στην πορεία της πανδημίας, δηλαδή στους θανάτους και τα χρούσματα COVID-19 ανά εκατομμύριο κατοίκων. Στον πίνακα 3.3 φαίνεται η μέση τιμή των μεγεθών αυτών για τέσσερις ζώνες στον άξονα του μέσου Containment and Health Index, μικρότερο από 45, από 45 μέχρι 55, από 55 μέχρι 65 και μεγαλύτερο από 65. Από τον πίνακα αυτό, φαίνεται ότι οι χώρες με χαμηλότερο Containment and Health Index εμφάνισαν αρκετά λιγότερα χρούσματα και θανάτους από COVID-19, σε σχέση με τις χώρες που βρίσκονταν στο μέσο, ενώ εκείνες που είχαν μεγάλο Containment and Health Index είχαν περισσότερα. Το γεγονός αυτό, ωστόσο, είναι λογικό, εφόσον ως επί το πλείστον, οι χώρες αυστηροποίησαν τα μέτρα προστασίας όταν αρχίζει να παρουσιάζεται κάποια έξαρση στο ρυθμό μετάδοσης του



Σχήμα 3.4: Θάνατοι COVID-19 ανά εκατομμύριο προς μέσο εβδομαδιαίο Containment and Health Index ανά χώρα.

Mean CHI	Mean Deaths p.m.	Mean Cases p.m.
< 45	281.77	28241.70
45 – 55	1296.21	69144.21
55 – 65	1125.75	57898.56
> 65	1686.43	64502.27

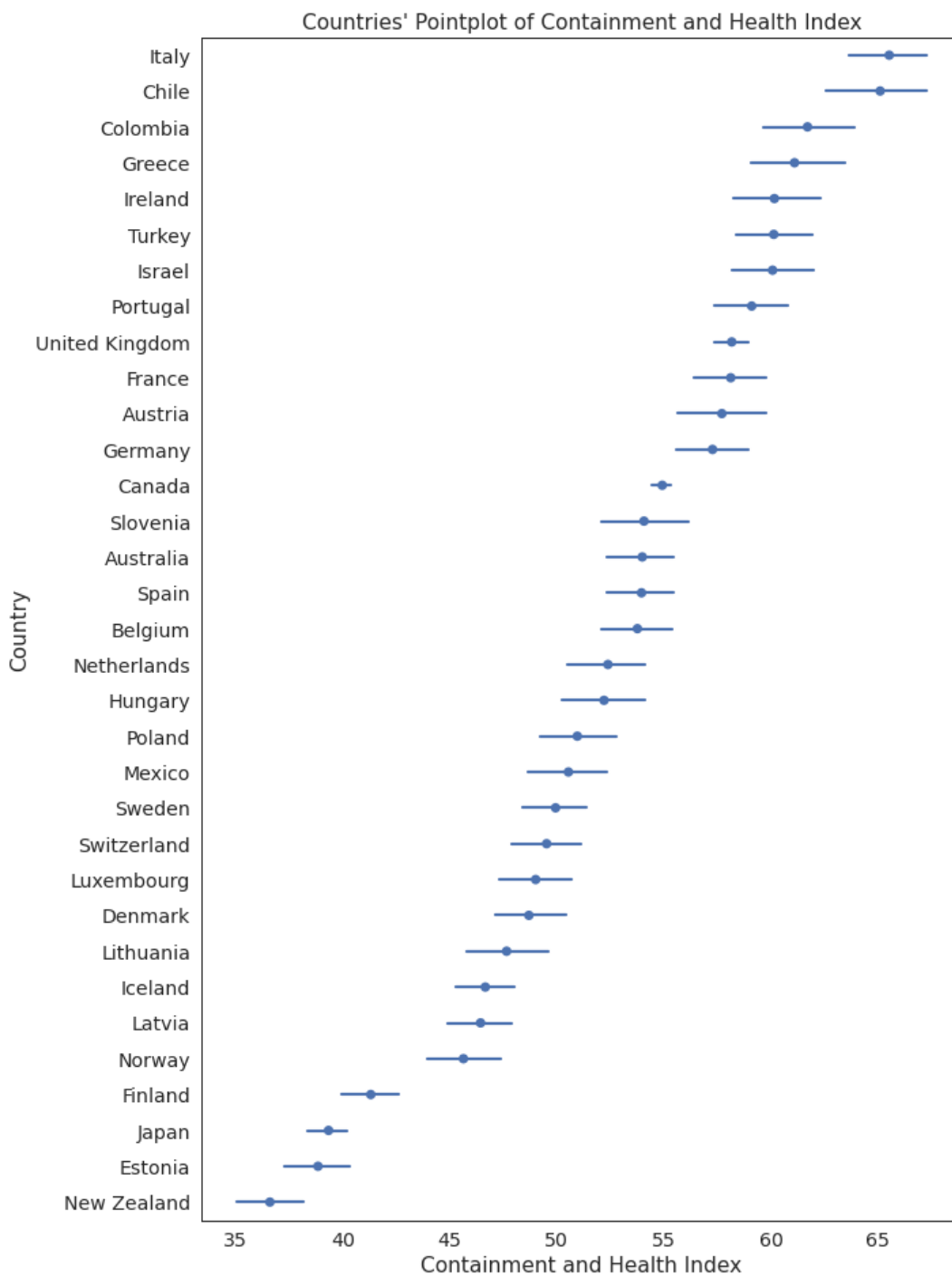
Πίνακας 3.3: Μέσοι θάνατοι και κρούσματα ανά εκατομμύριο για κάθε εύρος του μέσου Containment and Health Index.

κορωνοϊού. Επομένως, κατά κάποιο τρόπο, όσο μεγαλύτερη είναι η έξαρση των κρουσμάτων (άρα και των θανάτων) COVID-19, τόσο πιο αυστηρά μέτρα λαμβάνονται για τον περιορισμό της.

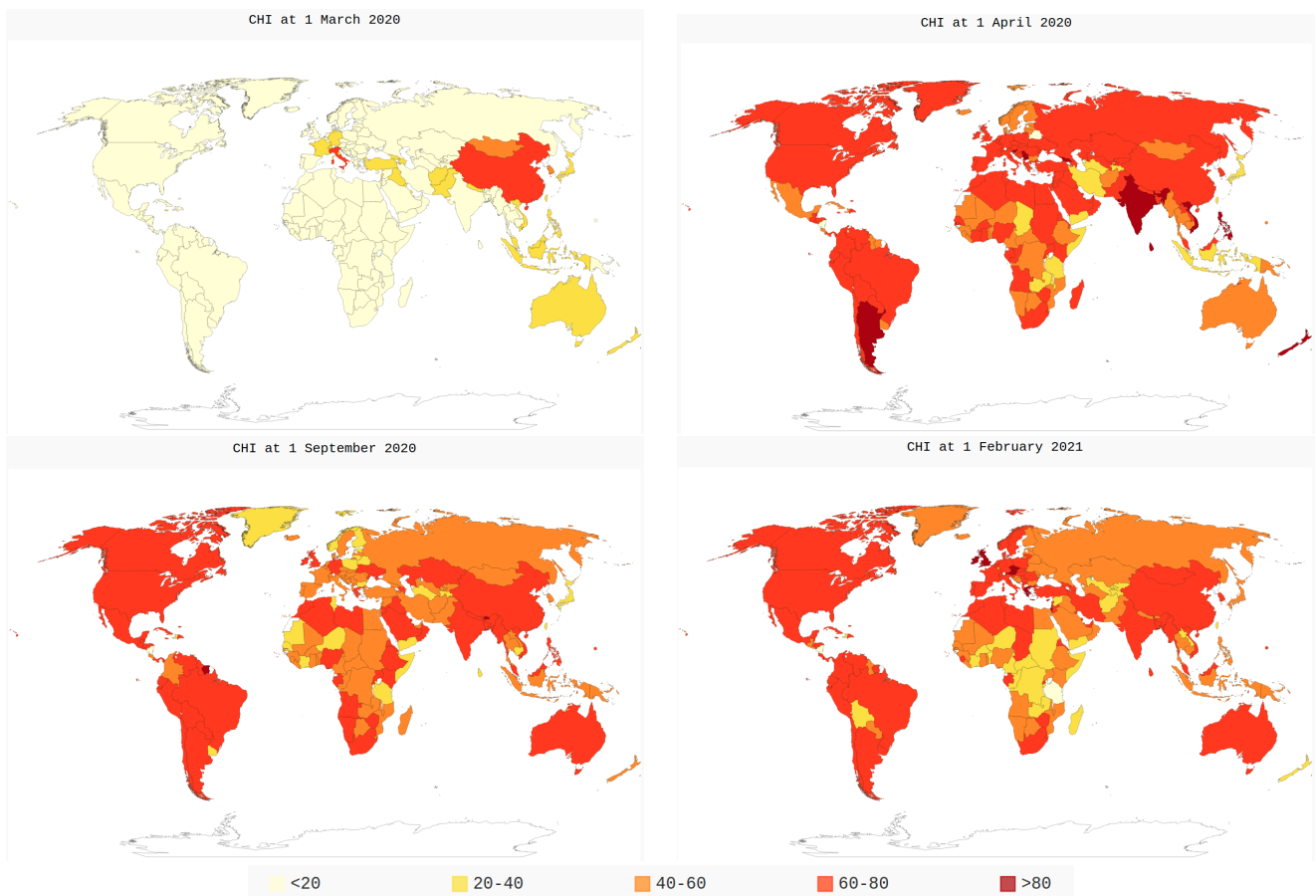
Η παραπάνω παρατήρηση, συμπληρώνεται από τους χάρτες του σχήματος 3.6, οι οποίοι δείχνουν την ύπαρξη ενός σχετικού συγχρονισμού στην λήψη και την χαλάρωση των μέτρων προστασίας. Ενώ την 1η Μαρτίου του 2020 ελάχιστες χώρες με κυριότερες την Κίνα και την Ιταλία είχαν αρχίσει να λαμβάνουν μέτρα, μόλις ένα μήνα μετά οι περισσότερες χώρες είχαν λάβει πολύ αυστηρά μέτρα, κατά περιπτώσεις αυστηρότερα από τις παραπάνω χώρες. Αυτό βέβαια μπορεί να αιτιολογηθεί, καθώς στις 11 Μαρτίου ο Παγκόσμιος Οργανισμός Υγείας χαρακτήρισε την νόσο COVID-19 πανδημία. Κοινή φαίνεται και η αποσυμπίεση των μέτρων αυτών από τις Ευρωπαϊκές χώρες μετά το τέλος του καλοκαιριού, την 1η Σεπτεμβρίου του 2020, καθώς επίσης και η εφαρμογή τους ξανά τον Φεβρουάριο του 2021, όταν κορυφωνόταν παγκοσμίως το δεύτερο κύμα της πανδημίας. Στο διάστημα αυτό, κάποιες χώρες παρέμειναν σταθερά ψηλά στον συγκεκριμένο δείκτη, άλλες σταθερά χαμηλά, ενώ άλλες εμφάνισαν παρόμοιες τάσεις χαλάρωσης και ισχυροποίησης. Η μεγάλη αυτή ομοιομορφία αποτελεί αφορμή για περαιτέρω αναζήτηση της επίδρασης των μέτρων στην εξέλιξη των θανάτων που ευθύνονται στην COVID-19.

3.1.4 Δημογραφικά και Οικονομικά Στοιχεία

Από τους παράγοντες που επηρέασαν την πορεία της πανδημίας COVID-19 σε μία χώρα κρίθηκε ότι είναι και δημογραφικά στοιχεία, όπως ο πληθυσμός της χώρας και η πυκνότητα αυτού, η διάμεση ηλικία των πολιτών της, αλλά και οικονομικά στοιχεία αντιπροσωπευτικά του επιπέδου διαβίωσης όπως το διάμεσο εισόδημα ανά νοικοκυριό και το κατά κεφαλήν κόστος ιατροφαρμακευτικής περίθαλψης. Μάλιστα, η πυκνότητα του πληθυσμού έχει επιβεβαιωθεί ότι επηρεάζει τη διάδοση και την περίοδο υποχώρησης της COVID-19 [35], ενώ έχει μελετηθεί ότι η ηλικία είναι ο πιο επικίνδυνος παράγοντας για να οδηγήσει στο θάνατο κάποιον που νοσεί με COVID-19 [13]. Η πυκνότητα πληθυσμού, ο συνολικός πληθυσμός και η διάμεση ηλικία των χωρών συλλέχθηκαν από τα δημοσιευμένα δεδομένα των Ηνωμένων Εθνών [29], το διάμεσο εισόδημα ανά νοικοκυριό των χωρών από το Gallup [15] και το κατά κεφαλήν κόστος ιατροφαρμακευτικής περίθαλψης από το Peterson-KFF Health System Tracker



Σχήμα 3.5: Διακύμανση Containment and Health Index ανά χώρα. Το μήκος της γραμμής αντιστοιχεί στο 95% της διακύμανσης, ενώ τα σημεία δείχνουν τη μέση τιμή του δείκτη.



Σχήμα 3.6: Εξέλιξη Containment and Health Index κατά τη διάρκεια της πανδημίας μέσα από στιγμιότυπα την πρώτη ημέρα του Μαρτίου, Απριλίου και Σεπτεμβρίου του 2020 και του Φεβρουαρίου του 2021.

[25]. Από το σχήμα 3.7 φαίνεται ότι οι χώρες της μελέτης καταναίμονται επαρκώς για όλα τα μεγέθη, προκειμένου να μπορεί να φανεί η συνεισφορά τους στην πορεία της πανδημίας, με εξαίρεση το συνολικό πληθυσμό των χωρών.

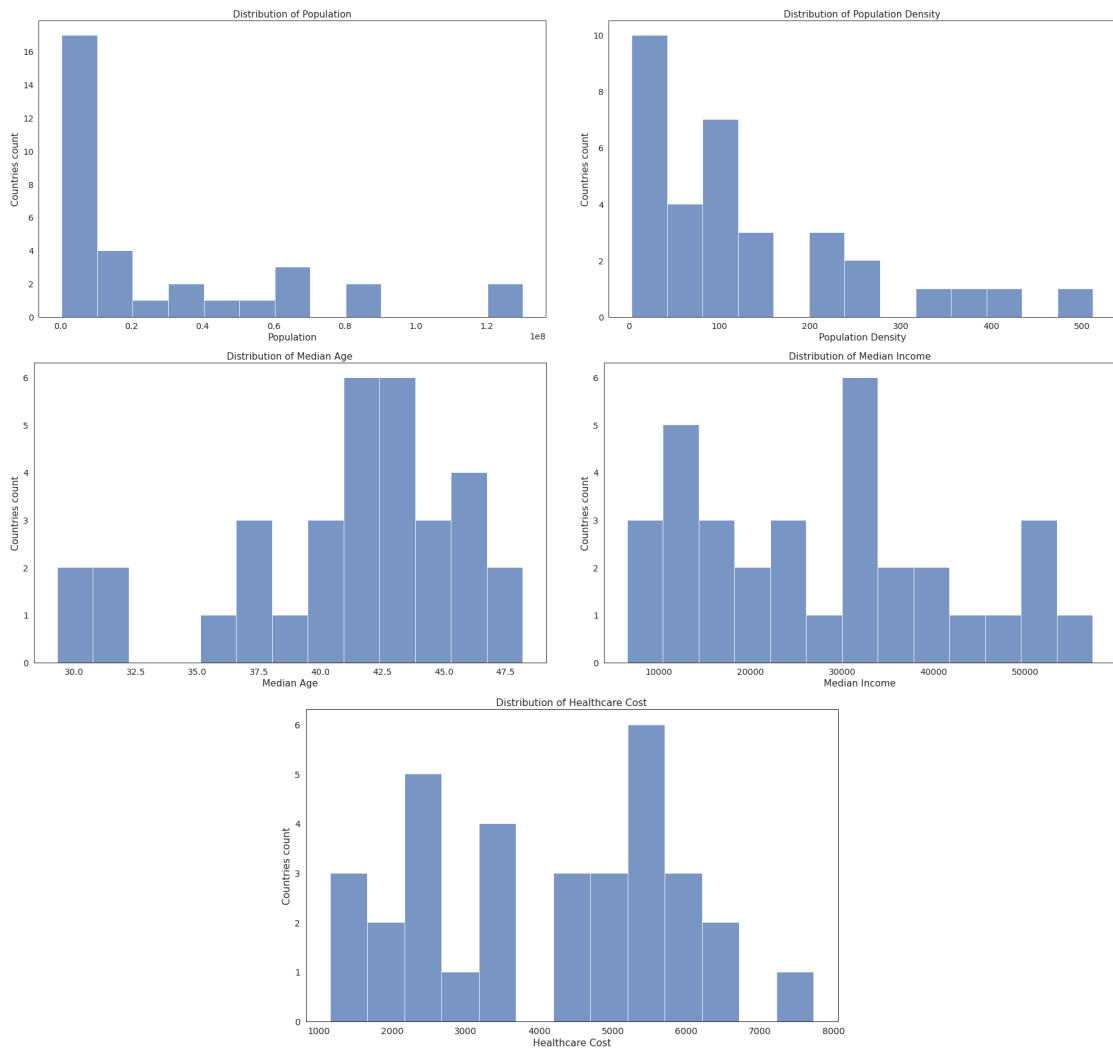
3.1.5 Τουριστικές Αφίξεις

Ένας τελευταίος παράγοντας για την κατάσταση της πανδημίας που κρίθηκε χρήσιμο να προσμετρηθεί είναι οι τουριστικές αφίξεις σε κάθε χώρα. Οι αφίξεις τουριστών σε μία χώρα μπορούν να χειροτερέψουν την επιδημιολογική της εικόνα βοηθώντας στη διασπορά της COVID-19, αλλά και επιβαρύνοντας επιπλέον το σύστημα υγείας της. Τα δεδομένα συλλέχθηκαν από τον Οργανισμό Παγκόσμιου Τουρισμού των Ηνωμένων Εθνών και αφορούν στοιχεία για το έτος 2016 [30]. Επομένως, τα δεδομένα δεν αποτελούν καθαρή ένδειξη του τουρισμού κατά τη διάρκεια της πανδημίας, αλλά μία γενικότερη εικόνα του μεγέθους της βιομηχανίας του τουρισμού κάθε χώρας. Στο διάγραμμα 3.8 φαίνεται η κατανομή των χωρών όσον αφορά τις αφίξεις τουριστών το 2016, στο οποίο παρατηρείται ότι ενώ για τις περισσότερες χώρες αυτές δεν ξεπερνούν τα 20 εκατομμύρια, υπάρχουν 10 χώρες οι οποίες ξεπερνούν αυτό το φράγμα.

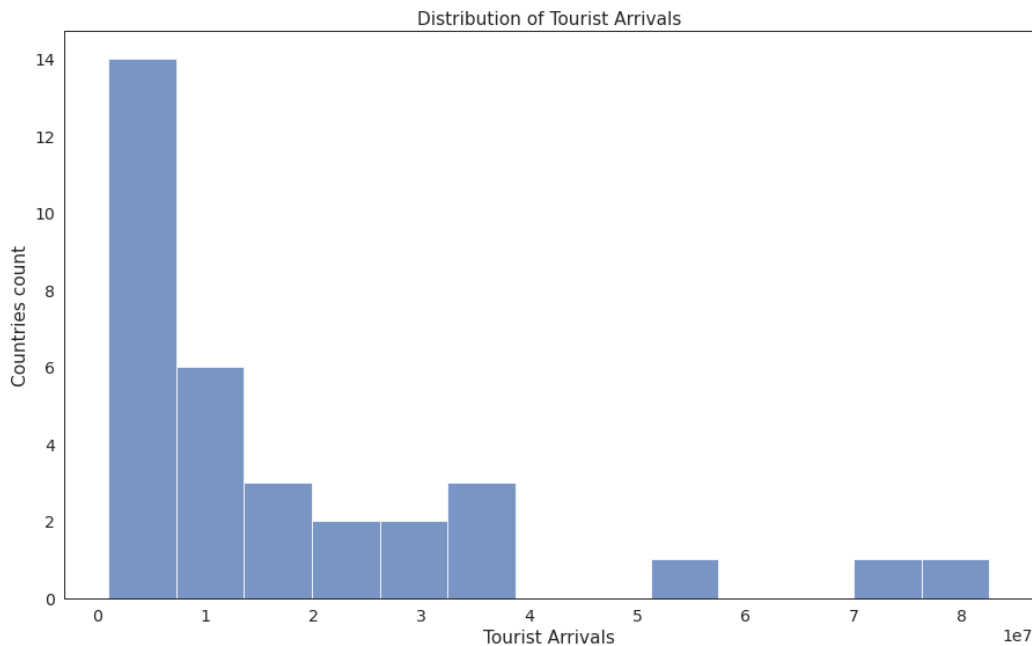
3.2 Επίδραση Δεδομένων

Το χαρακτηριστικό που θα χρησιμοποιηθεί για την αξιολόγηση της πορείας της πανδημίας σε μία χώρα, όπως αναφέρθηκε προηγουμένως, είναι οι θάνατοι ανά εκατομμύριο που ευθύνονται στην COVID-19. Επομένως, είναι χρήσιμη η μελέτη της συσχέτισης των θανάτων ανά εκατομμύριο με τα υπόλοιπα χαρακτηριστικά, προκειμένου να σχηματιστεί μία εικόνα ως προς το ποια συσχετίζονται θετικά με αυτούς, ποια αρνητικά και ποια φαίνεται να μην έχουν σχέση. Για το λόγο αυτό, παρουσιάζονται στο σχήμα 3.9 τα διαγράμματα διασποράς όλων των χαρακτηριστικών με τους θανάτους ανά εκατομμύριο, εκτός από το μέσο εβδομαδιαίο Containment and Health Index που καλύφθηκε στο σχήμα 3.4. Στον πίνακα 3.4 παρουσιάζονται ο συντελεστής συσχέτισης, και οι t , p τιμές των αντίστοιχων χαρακτηριστικών με τους θανάτους ανά εκατομμύριο. Με αστερίσκο και έντονη γραμματοσειρά σημειώνονται τα χαρακτηριστικά εκείνα που εμφανίζουν στατιστικά σημαντική συσχέτιση με τους θανάτους, με συντελεστή συσχέτισης Pearson μεγαλύτερο ή ίσο του 0.4 κατά απόλυτη τιμή.

Τα χαρακτηριστικά αυτά, είναι το μέσο Containment and Health Index, το οποίο συσχετίζεται θετικά με τους θανάτους με $r = 0.46$ και το διάμεσο εισόδημα, το οποίο σχετίζεται αρνητικά με $r = -0.40$. Το πρώτο, επιβεβαιώνει από στατιστική σκοπιά τον πίνακα 3.3, σύμφωνα με τον οποίο οι χώρες που έλαβαν τα πιο αυστηρά μέτρα για την προστασία από την πανδημία ήταν αυτές που παρουσίασαν τους περισσότερους θανάτους ανά εκατομμύριο. Η παρατήρηση αυτή φαίνεται λογική, από την άποψη ότι η κάθε χώρα προσπαθεί εντατικά να περιορίσει την περαιτέρω αύξηση των θανάτων, αφού αυτοί έχουν φτάσει πάνω από κάποια ανησυχητικά επίπεδα. Η αρνητική συσχέτιση του διαμέσου εισοδήματος μπορεί να ευθύνεται σε δύο παράγοντες. Πρώτον, στην



Σχήμα 3.7: Κατανομή δημογραφικών και οικονομικών στοιχείων ανά χώρα.



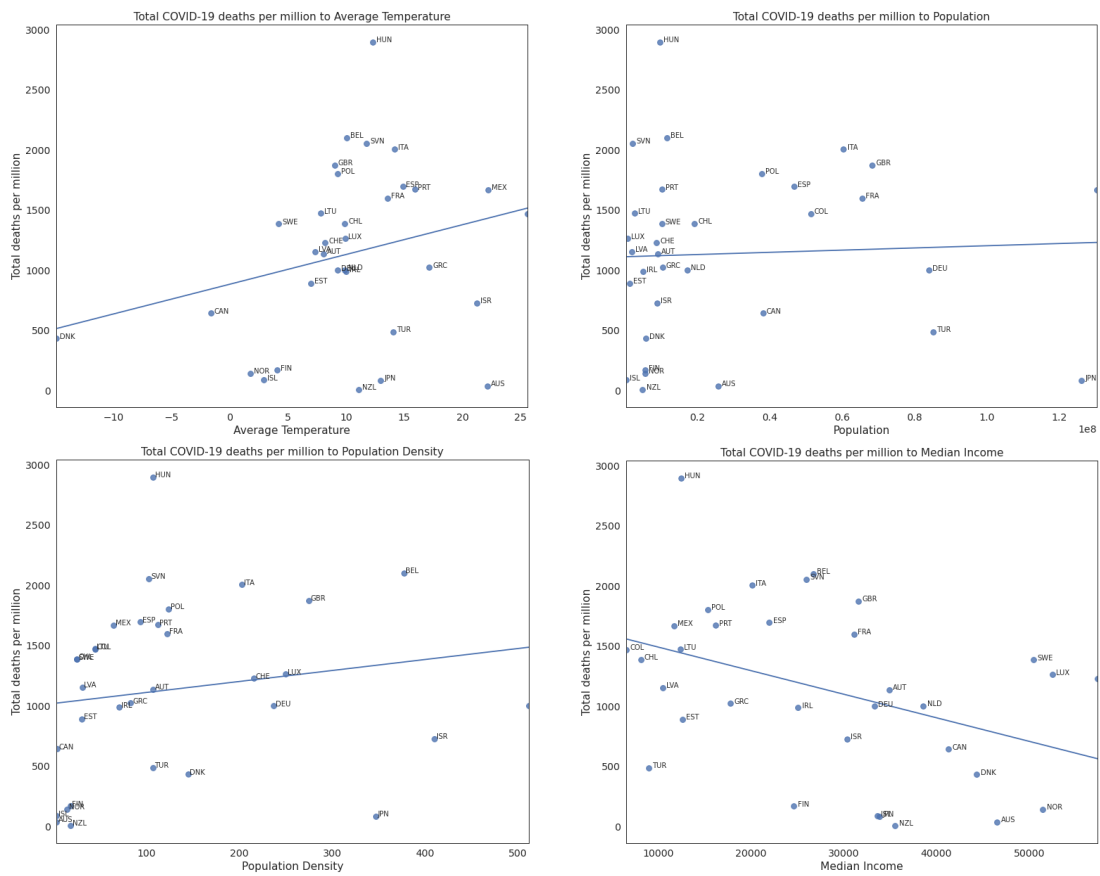
Σχήμα 3.8: Κατανομή τουριστικών αφίξεων για κάθε χώρα για το έτος 2016.

ύπαρξη καλύτερου και πιο προσιτού εθνικού συστήματος υγείας, στο οποίο οι ασθενείς έχουν ευκολότερη πρόσβαση και είναι το ίδιο πιο καλά καταρτισμένο για την ίαση των νοσούντων. Στην υπόθεση αυτή συμβάλει και το κόστος ιατροφαρμακευτικής περίθαλψης που εμφανίζει συντελεστή συσχέτισης Pearson ίσο με $r = -0.30$ με τους θανάτους ανά εκατομμύριο. Ο δεύτερος παράγοντας στον οποίο μπορεί να ευθύνεται η αρνητική συσχέτιση του διάμεσου εισοδήματος με τους θανάτους ανά εκατομμύριο είναι η ύπαρξη καλύτερου μορφωτικού επιπέδου, που οδηγεί στην κατανόηση της κρισιμότητας της πανδημίας από το γενικότερο πληθυσμό και επομένως στην καλύτερη τήρηση των μέτρων προστασίας.

3.3 Ανάλυση Συγκεντρωτικών Δεδομένων

3.3.1 Περιγραφή Διαδικασίας

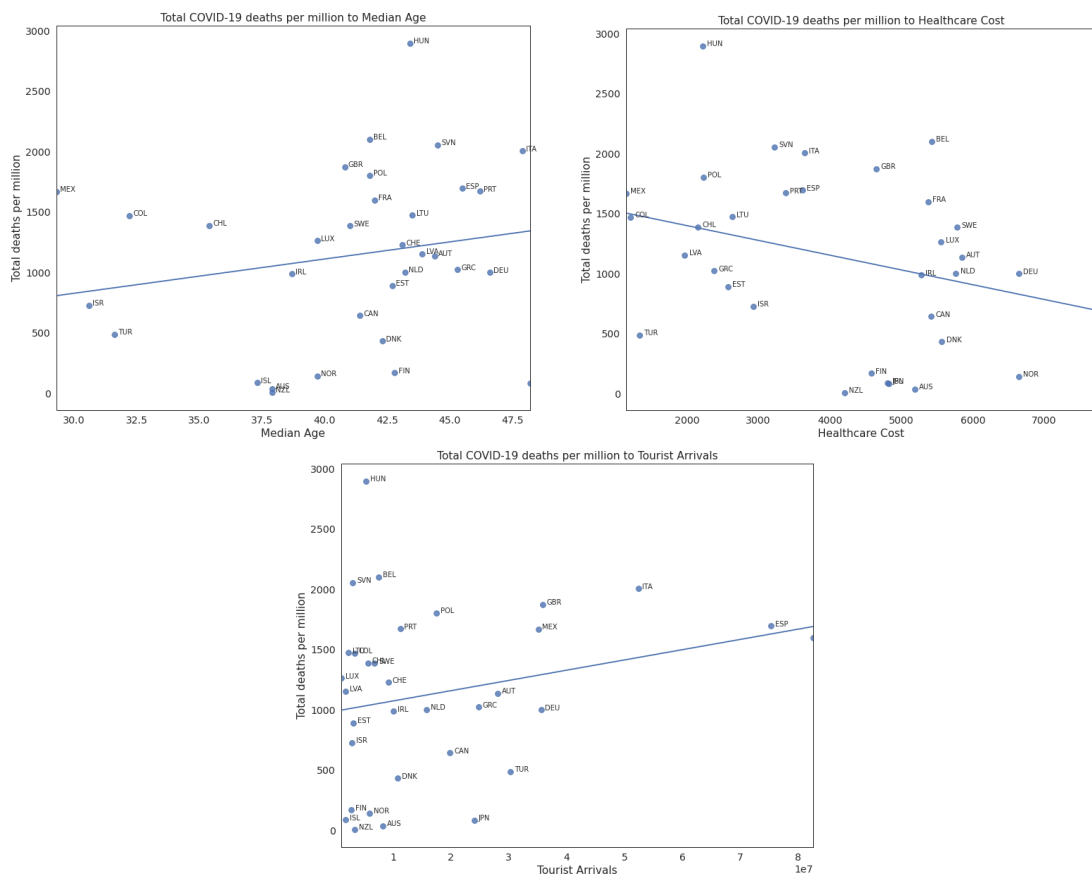
Στο στάδιο αυτό μελετούνται οι 33 χώρες που αναφέρθηκαν προηγουμένως ως προς τα παραπάνω σύνολα δεδομένων, εξετάζοντας αν αυτά μπορούν να εξηγήσουν τις διαφορετικές πορείες που έλαβε η πανδημία για τις χώρες αυτές. Για την εξαγωγή μιας συγκεντρωτικής τιμής των μέτρων του Containment and Health Index που έλαβαν οι χώρες, επιλέχθηκε ο μέσος όρος των εβδομαδιαίων τιμών (άθροισμα τιμών για κάθε ημέρα της εβδομάδας) του δείκτη αυτού. Οι θάνατοι ανά εκατομμύριο αφήθηκαν εκτός του συνόλου δεδομένων και χρησιμοποιήθηκαν ως κριτήριο αξιολόγησης του κατά



Σχήμα 3.9: Διαγράμματα σχέδασης θανάτων ανά εκατομμύριο με τα υπόλοιπα χαρακτηριστικά των χωρών.

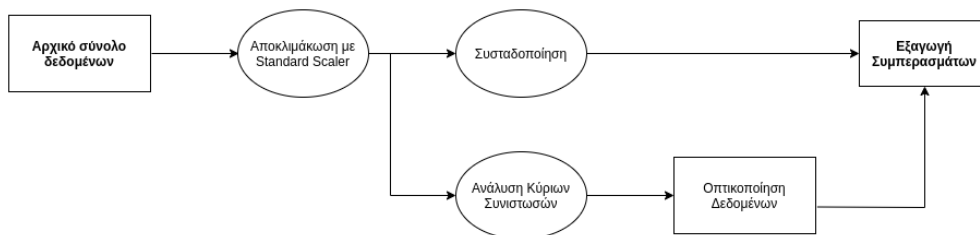
Feature	Correlation	T-Value	P-Value
Μέση Θερμοκρασία	0.26	-9.10	$2.17 \cdot 10^{-10}$
* Μέσο CHI	0.46	-5.81	$1.82 \cdot 10^{-6}$
Πληθυσμός	0.05	4.70	$4.69 \cdot 10^{-5}$
Πυκνότητα πληθυσμού	0.17	-7.99	$2.53 \cdot 10^{-9}$
* Διάμεσο Εισόδημα	-0.40	10.65	$4.44 \cdot 10^{-12}$
Διάμεση Ηλικία	0.19	-8.85	$4.11 \cdot 10^{-10}$
Κόστος Περίθαλψης	-0.30	9.04	$1.98 \cdot 10^{-11}$
Τουριστικές αφίξεις	0.24	4.98	$2.12 \cdot 10^{-5}$

Πίνακας 3.4: Στατιστικά μεγέθη συσχέτισης χαρακτηριστικών με θανάτους ανά εκατομμύριο των χωρών. Με αστερίσκο και έντονη γραμματοσειρά σημειώνονται τα χαρακτηριστικά που εμφανίζουν στατιστικά σημαντική συσχέτιση με τους θανάτους.



Σχήμα 3.9: Διαγράμματα σχέδασης θανάτων ανά εκατομμύριο με τα υπόλοιπα χαρακτηριστικά των χωρών (συνέχεια).

πόσο τα υπόλοιπα χαρακτηριστικά μπορούν να ερμηνεύσουν την πορεία των χωρών. Η διαδικασία που ακολουθήθηκε περιγράφεται συνοπτικά στο σχήμα 3.10 και αποτελείται από αποκλιμάκωση αρχικά των δεδομένων μέσω ενός Standard Scaler, προκειμένου να μην υπερισχύουν ορισμένα χαρακτηριστικά λόγω των εν γένει μεγάλων τιμών τους, εφαρμογή Ανάλυσης Κυρίων Συνιστωσών (PCA) για να είναι δυνατή η οπτικοποίηση των δεδομένων και τέλος από συσταδοποίηση των χωρών ως προς τα αποκλιμακωμένα χαρακτηριστικά ώστε να φανεί αν αυτά καθορίζουν την πορεία των χωρών κατά τη διάρκεια της πανδημίας.



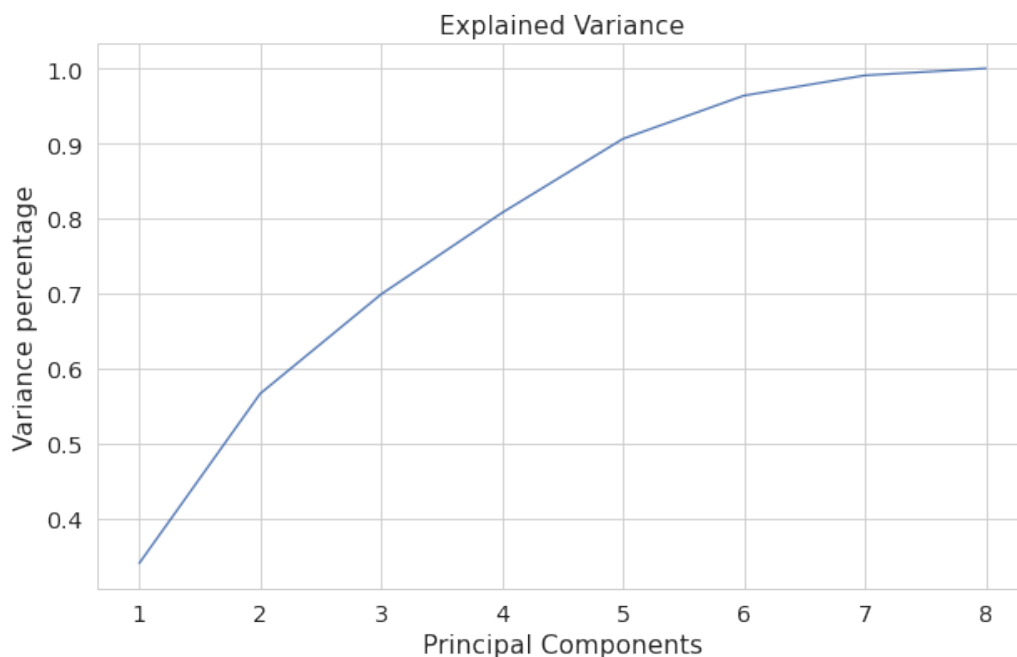
Σχήμα 3.10: Διαδικασία ανάλυσης συγκεντρωτικών δεδομένων.

3.3.2 Ανάλυση σε Κύριες Συνιστώσες (PCA)

Πρωτού εφαρμοστεί η Ανάλυση σε Κύριες Συνιστώσες, τα δεδομένα αποκλιμακώθηκαν χρησιμοποιώντας τον Standard Scaler που παρέχεται από τη βιβλιοθήκη scikit-learn της Python, ο οποίος αφαιρεί το μέσο όρο από τα δεδομένα και τα κλιμακώνει διαιρώντας με την τυπική απόκλιση. Στη συνέχεια, με τη χρήση της συνάρτησης PCA από την ίδια βιβλιοθήκη εξετάστηκε η διακύμανση που διατηρείται για κάθε πλήθος κυρίων συνιστωσών. Τα αποτελέσματα φαίνονται διαγραμματικά στο σχήμα 3.11 και στον πίνακα 3.5.

Από τα νούμερα αυτά φαίνεται ότι τα χαρακτηριστικά δεν εμφανίζουν μεγάλες ομοιότητες μεταξύ τους, καθώς η διατηρούμενη διακύμανση αυξάνεται σημαντικά μέχρι και την πέμπτη κύρια συνιστώσα. Το φαινόμενο αυτό, υπενθυμίζει τη μεγάλη δυσκολία της διαχείρισης της πανδημίας, καθώς υπάρχουν πολλαπλοί παράγοντες που επιδρούν με διαφορετικό τρόπο ο καθένας. Ωστόσο, προκειμένου να είναι δυνατή η απεικόνιση των χωρών σε κάποιο χώρο αντιληπτό από τον άνθρωπο, επιλέχθηκε η χρήση των τριών πρώτων κυρίων συνιστωσών, οι οποίες διατηρούν σχεδόν το 70% της συνολικής διακύμανσης.

Στον πίνακα 3.6 παρουσιάζονται οι συντελεστές βαρύτητας κάθε κύριας συνιστώσας για κάθε χαρακτηριστικό του συνόλου δεδομένων. Για κάθε συνιστώσα φαίνονται με έντονα γράμματα οι τρεις σημαντικότεροι συντελεστές της. Όπως φαίνεται, όλα τα χαρακτηριστικά έχουν παρόμοια βάρη αθροιστικά στις τρεις κύριες συνιστώσες, και μάλιστα τα περισσότερα από αυτά έχουν μεγάλο βάρος και στις τρεις κύριες συνιστώσες. Από την πρώτη κύρια συνιστώσα, τα αντίθετα πρόσημα της μέσης θερμο-



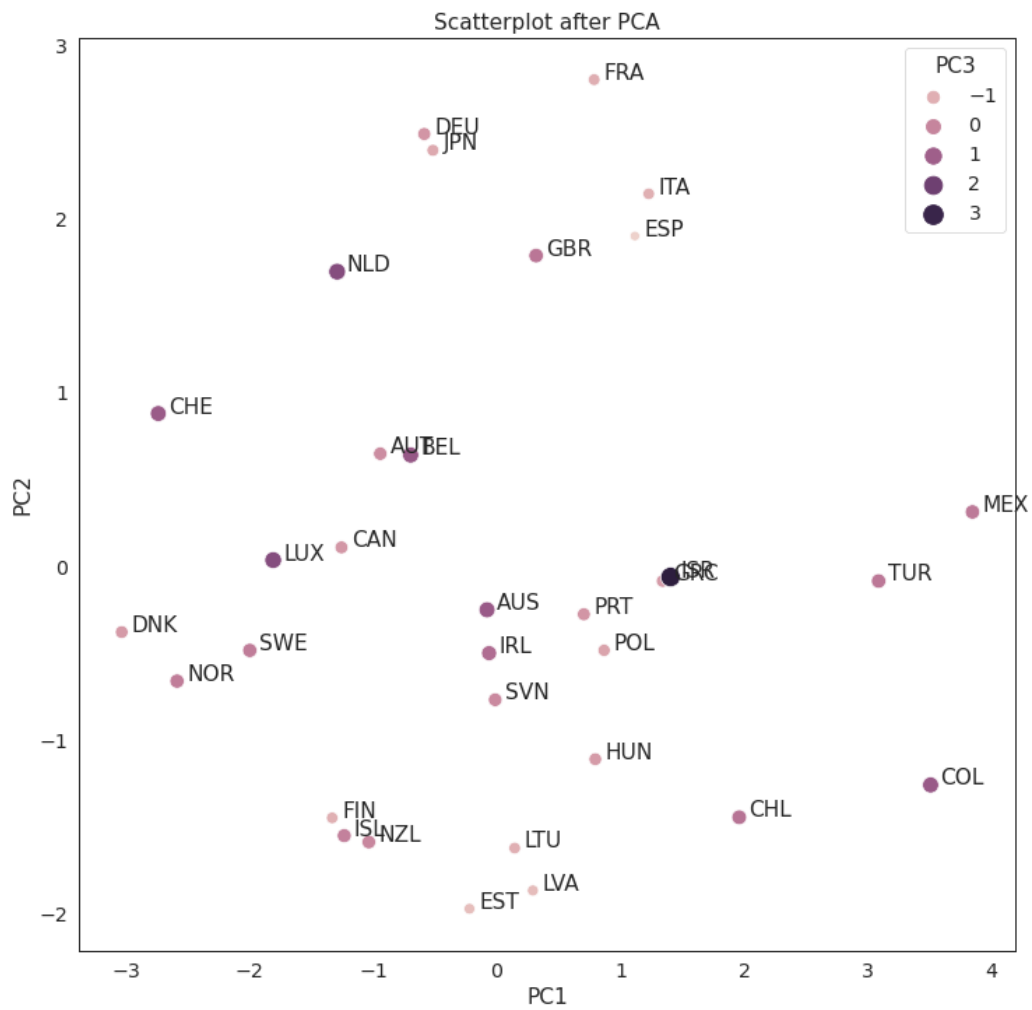
Σχήμα 3.11: Διακύμανση που διατηρείται ανάλογα το πλήθος Κυρίων Συνιστωσών.

κρασίας με το διάμεσο εισόδημα και το κόστος περίθαλψης οφείλονται μάλλον στην ύπαρξη περισσότερου πλούτου και καλύτερου εθνικού συστήματος υγείας στις βόρειες χώρες σε σχέση με τις νότιες. Υπενθυμίζεται ότι οι παράγοντες αυτοί εμφάνισαν αρνητική συσχέτιση με τους θανάτους ανά εκατομμύριο στο σχήμα 3.9, γεγονός που σχετίζεται άμεσα από την ποιότητα και την προσβασιμότητα της ιατροφαρμακευτικής περίθαλψης του πληθυσμού. Από τους συντελεστές της δεύτερης κύριας συνιστώσας, φαίνεται ότι οι χώρες με μεγάλο πληθυσμό διαθέτουν συνήθως και μεγάλα αστικά κέντρα, τα οποία τείνουν να προσελκύουν περισσότερες τουριστικές αφίξεις από τις μικρότερες. Το τελευταίο συμπέρασμα, ωστόσο, φαίνεται να διαψεύδεται εν μέρει από την τρίτη κύρια συνιστώσα, όπου η πυκνότητα πληθυσμού και οι τουριστικές αφίξεις εμφανίζουν αντίθετα πρόσημα.

Στο σχήμα 3.12 φαίνεται το διάγραμμα των χωρών που προκύπτει μετά την ανάλυση στις τρεις κύριες συνιστώσες. Η τρίτη κύρια συνιστώσα απεικονίζεται μέσω του χρώματος και του μεγέθους των σημείων.

3.3.3 Συσταδοποίηση

Το επόμενο βήμα είναι η οργάνωση των χωρών σε συστάδες, προκειμένου να εξεταστεί αν χώρες με κοινά χαρακτηριστικά παρουσίασαν κοινή πορεία κατά τη διάρκεια της πανδημίας. Για το σκοπό αυτό χρησιμοποιήθηκε ο συσσωρευτικός αλγόριθμος ιεραρχικής συσταδοποίησης που παρέχεται στη βιβλιοθήκη scikit-learn της Python, για



Σχήμα 3.12: Διάγραμμα χωρών μετά την ανάλυση σε τρεις κύριες συνιστώσες. Η τρίτη διάσταση απεικονίζεται μέσω του χρώματος και του μεγέθους των σημείων.

Πλήθος Κυρίων Συνιστωσών	Διατηρούμενη Διακύμανση
1	34.07%
2	56.67%
3	69.90%
4	80.76%
5	90.62%
6	96.37%
7	99.07%
8	100%

Πίνακας 3.5: Διατηρούμενη διακύμανση ανάλογα το πλήθος Κυρίων Συνιστωσών.

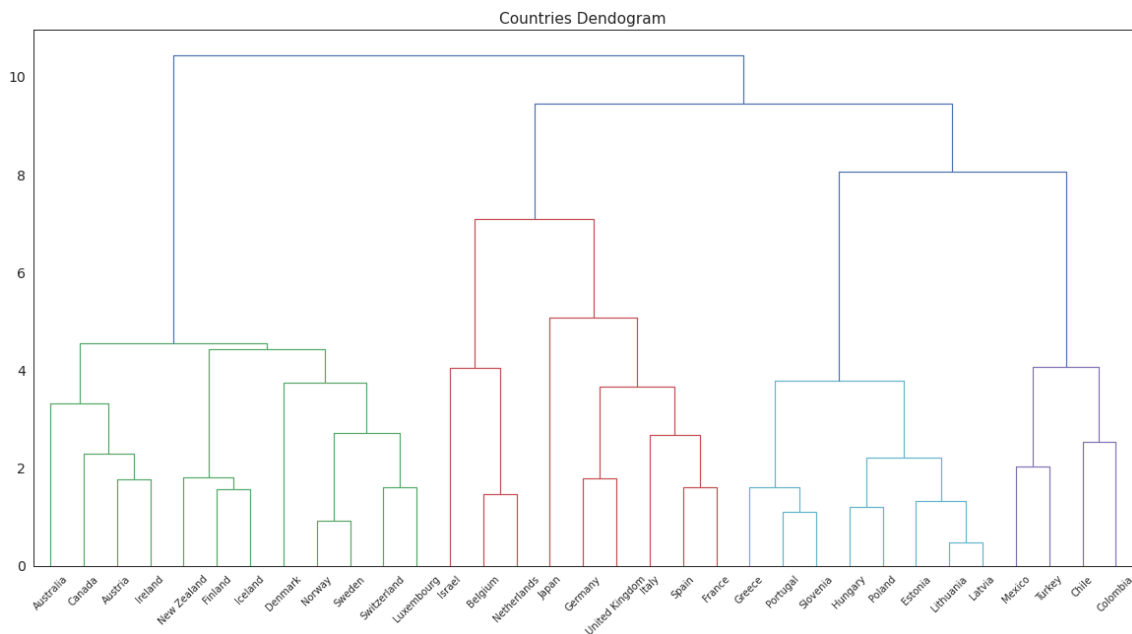
Κύρια Συνιστώσα	1	2	3
Μέση Θερμοκρασία	0.45	0.18	0.29
Μέσο CHI	0.31	0.20	0.28
Πληθυσμός	0.30	0.45	-0.13
Πυκνότητα Πληθυσμού	-0.06	0.45	0.46
Διάμεσο Εισόδημα	-0.49	0.21	0.32
Διάμεση Ηλικία	-0.27	0.25	-0.58
Κόστος Περίθαλψης	-0.51	0.31	0.16
Τουριστικές Αφίξεις	0.18	0.57	-0.37

Πίνακας 3.6: Συντελεστές βαρύτητας χαρακτηριστικών για κάθε κύρια συνιστώσα. Με έντονα γράμματα φαίνονται οι τρεις σημαντικότεροι συντελεστές κάθε συνιστώσας.

το λόγο που αναφέρθηκε στο Κεφάλαιο 2.1.5. Με βάση το δένδρογραμμα του σχήματος 3.13, φαίνεται ότι οι χώρες καταναίμονται σε τέσσερις επαρκώς διαχωρισμένες συστάδες.

Η σύνθεση των συστάδων με τη σειρά που εμφανίζονται στο δένδρογραμμα είναι:

1. Αυστραλία, Καναδάς, Αυστρία, Ιρλανδία, Νέα Ζηλανδία, Φινλανδία, Ισλανδία, Δανία, Νορβηγία, Σουηδία, Ελβετία και Λουξεμβούργο.
2. Ισραήλ, Βέλγιο, Ολλανδία, Ιαπωνία, Γερμανία, Ηνωμένο Βασίλειο, Ιταλία, Ισπανία και Γαλλία.
3. Σλοβενία, Ελλάδα, Πορτογαλία, Ουγγαρία, Πολωνία, Εσθονία, Λιθουανία και Λετονία.
4. Μεξικό, Τουρκία, Χιλή και Κολομβία.



Σχήμα 3.13: Δενδρόγραμμα συσσωρευτικής ιεραρχικής συσταδοποίησης χωρών.

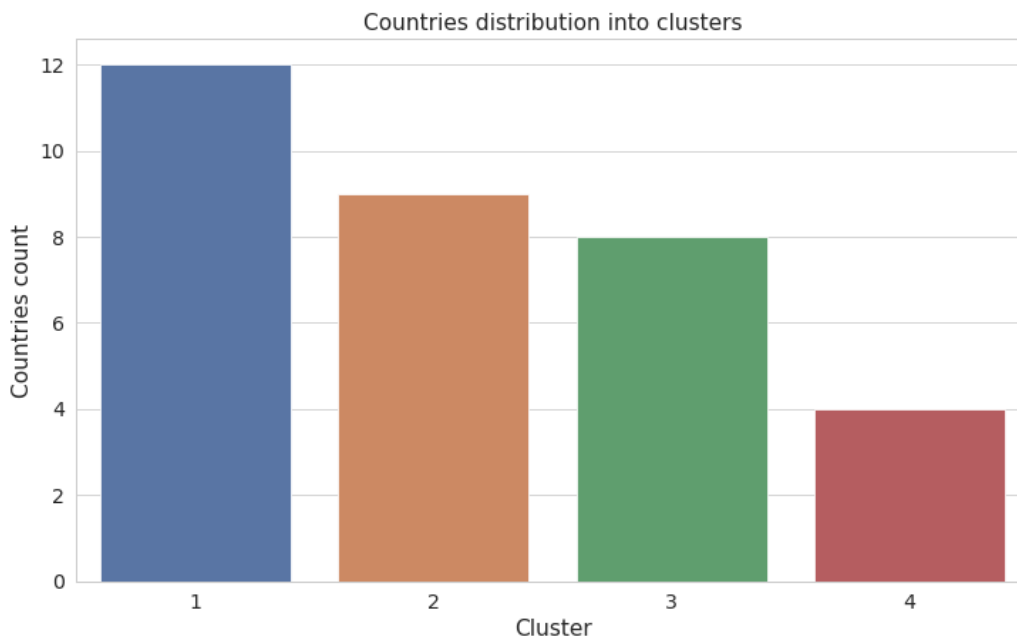
Στο διάγραμμα του σχήματος 3.14 φαίνεται η κατανομή των χωρών στις τέσσερις συστάδες. Η μεγαλύτερη συστάδα είναι η πρώτη με 12 χώρες, αλλά κάθε συστάδα εκπροσωπείται από αρκετές χώρες ώστε να μπορέσουν να εξαχθούν συμπεράσματα για αυτές.

Στο διάγραμμα του σχήματος 3.15 φαίνεται η κατανομή των χωρών στις τρεις κύριες συνιστώσες μετά τη συσταδοποίηση, ενώ στον πίνακα 3.7 παρατίθεται η μέση τιμή των χαρακτηριστικών των χωρών κάθε συστάδας.

Τέλος, στο διάγραμμα 3.16 φαίνεται η γεωγραφική κατανομή των χωρών στις συστάδες.

3.3.4 Αποτελέσματα Συσταδοποίησης

Για την ανάλυση των αποτελεσμάτων της συσταδοποίησης απαιτείται η εξέταση της πορείας των χωρών κατά τη διάρκεια της πανδημίας. Στο διάγραμμα 3.17 φαίνονται οι θάνατοι ανά εκατομμύριο που ευθύνονται σε COVID-19 για τις 33 χώρες για κάθε βδομάδα της πανδημίας. Η γραμμή αντιπροσωπεύει τη μέση τιμή των θανάτων ανά εκατομμύριο των χωρών ενώ το διάστημα αντιστοιχεί σε διάστημα εμπιστοσύνης 95% των τιμών κάθε χώρας. Στο διάγραμμα αυτό φαίνονται ξεκάθαρα τα δύο λεγόμενα κύματα της πανδημίας, το πρώτο τον Μάρτιο-Απρίλιο του 2020 και το δεύτερο από το Νοέμβριο του 2020 με μία αποκλιμάκωση τον Φεβρουάριο του 2021. Σημαντική ένδειξη ότι ορισμένες χώρες εμφάνισαν καλύτερη επιδημιολογική εικόνα από κάποιες

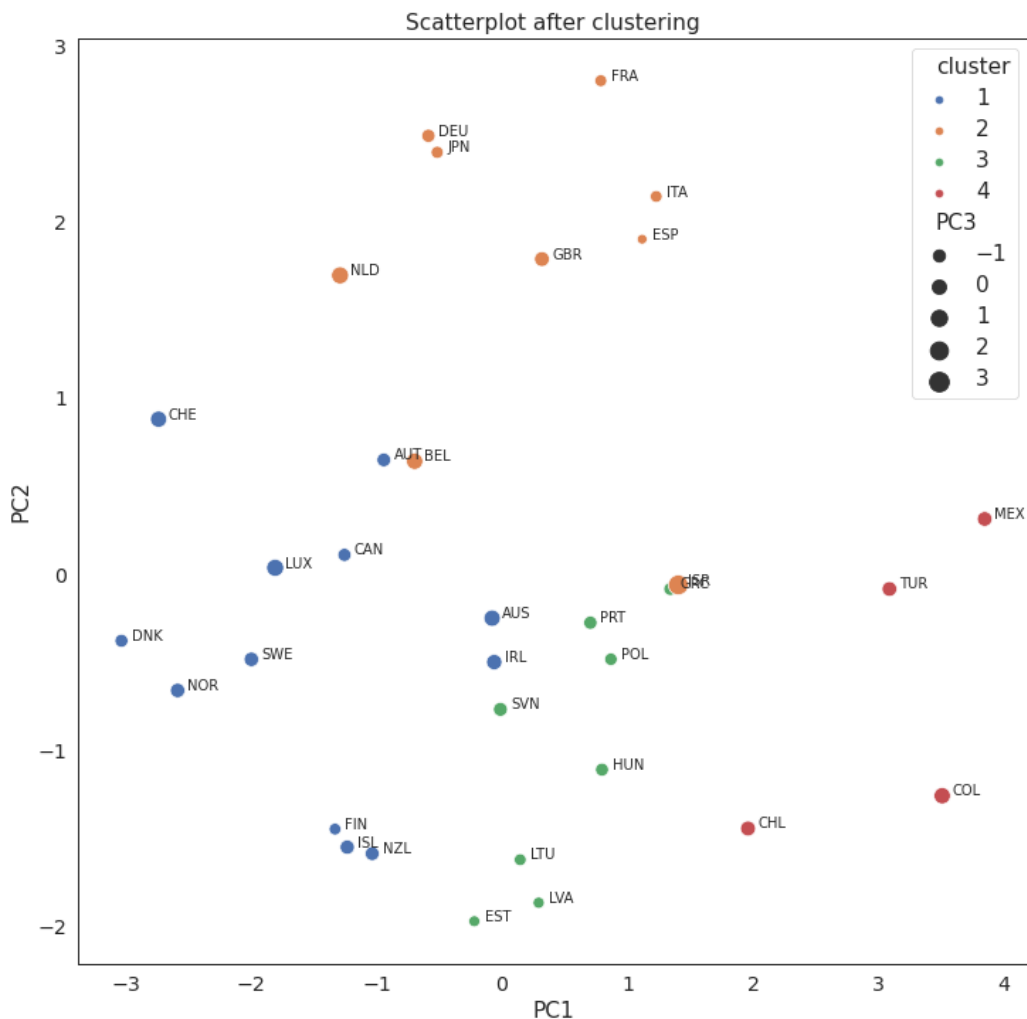


Σχήμα 3.14: Κατανομή χωρών σε συστάδες.

άλλες αποτελεί το μεγάλο πάχος του διαστήματος εμπιστοσύνης κατά τα δύο κύματα.

Στο σχήμα 3.18 φαίνεται το αντίστοιχο διάγραμμα με ξεχωριστή γραμμή για κάθε συστάδα, η οποία ορίζεται από το αντίστοιχο χρώμα. Το διάγραμμα αυτό, σε συνδυασμό με τον πίνακα 3.7, οδηγούν στις παρακάτω παρατηρήσεις, σχετικά με την εξέλιξη της πανδημίας για τις χώρες κάθε συστάδας:

- Οι χώρες του cluster 1 παρουσίασαν σχετικά λίγους θανάτους ανά εκατομμύριο καθ' όλη τη διάρκεια της πανδημίας, με εξαίρεση το πρώτο κύμα στο οποίο ήταν η συστάδα με τους δεύτερους περισσότερους θανάτους ανά εκατομμύριο. Ωστόσο, συνολικά εμφάνισαν πολύ λιγότερους θανάτους ανά εκατομμύριο σε σχέση με τις υπόλοιπες συστάδες.
- Οι χώρες του cluster 2 παρουσίασαν τους περισσότερους θανάτους ανά εκατομμύριο κατά το πρώτο κύμα της πανδημίας ενώ πέρασαν πιο ήπια το δεύτερο κύμα.
- Οι χώρες του cluster 3 παρόλο που εμφάνισαν ελάχιστους θανάτους κατά το πρώτο κύμα της πανδημίας είναι αυτές που υπέφεραν περισσότερο από το δεύτερο κύμα, με πολλές από αυτές να μην έχουν καταφέρει ακόμη να το θέσουν σε ύφεση. Αυτό είχε ως αποτέλεσμα να αποτελούν τελικά τη συστάδα με τους περισσότερους θανάτους ανά εκατομμύριο.
- Οι χώρες του cluster 4 εμφανίζουν ιδιόμορφη συμπεριφορά, καθώς η αύξηση των θανάτων συνέβη στις αρχές του καλοκαιριού του 2020 όταν οι χώρες των



Σχήμα 3.15: Διάγραμμα χωρών μετά την ανάλυση σε τρεις κύριες συνιστώσες και το διαμοιρασμό τους σε συστάδες. Η συστάδα στην οποία ανήκει η κάθε χώρα απεικονίζεται μέσω του χρώματος των σημείων.

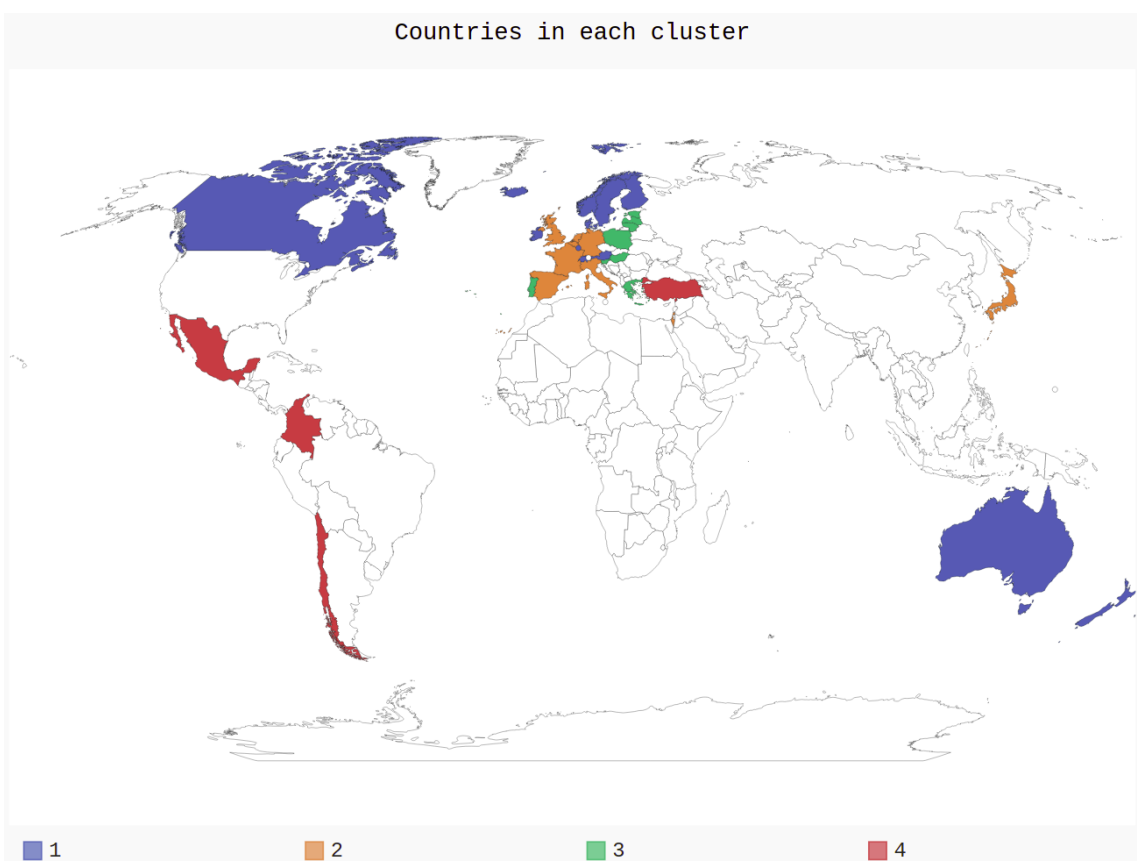
υπόλοιπων συστάδων είχαν αντιμετωπίσει το πρώτο κύμα, και τα νούμερα αυτά διατηρήθηκαν στα ίδια επίπεδα καθ' όλη τη διάρκεια της πανδημίας.

Το γεγονός ότι οι χώρες κάθε συστάδας παρουσίασαν διαφορετική πορεία από αυτές των υπόλοιπων συστάδων, αποτελεί επιβεβαίωση ότι η επιλογή των χαρακτηριστικών και της επεξεργασίας τους ήταν σωστή.

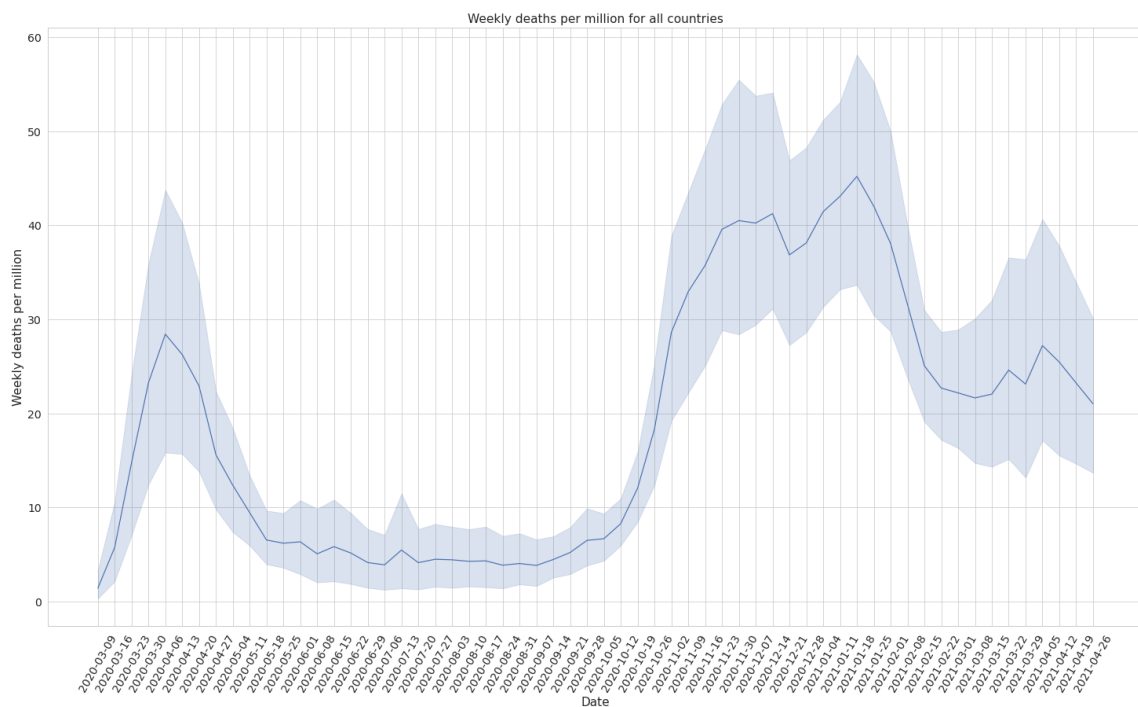
Για λόγους πληρότητας, παρατίθεται στο σχήμα 3.19 η πορεία των εβδομαδιαίων θανάτων ανά εκατομμύριο κάθε χώρας χωρίς τη συνάνθρωπισή τους σε συστάδες.

Cluster	1	2	3	4
Θάνατοι ανά εκατομμύριο	622.67	1340.63	1619.53	1251.66
Μέσο εβδομαδιαίο CHI	391.03	431.48	406.97	469.36
Μέση Θερμοκρασία	5.47	12.77	11.05	17.93
Πληθυσμός (εκατομ.)	9.95	54.25	9.49	71.45
Πυκνότητα Πληθυσμού	72.93	286.41	79.48	60.46
Διάμεσο Εισόδημα	41485.83	29731.00	15391.00	8819.25
Διάμεση Ηλικία	40.52	42.96	43.91	32.13
Κόστος Περίθαλψης	5551.16	4765.33	2578.48	1465.68
Τουριστικές Αφίξεις (εκατομ.)	9.00	36.88	8.63	18.58

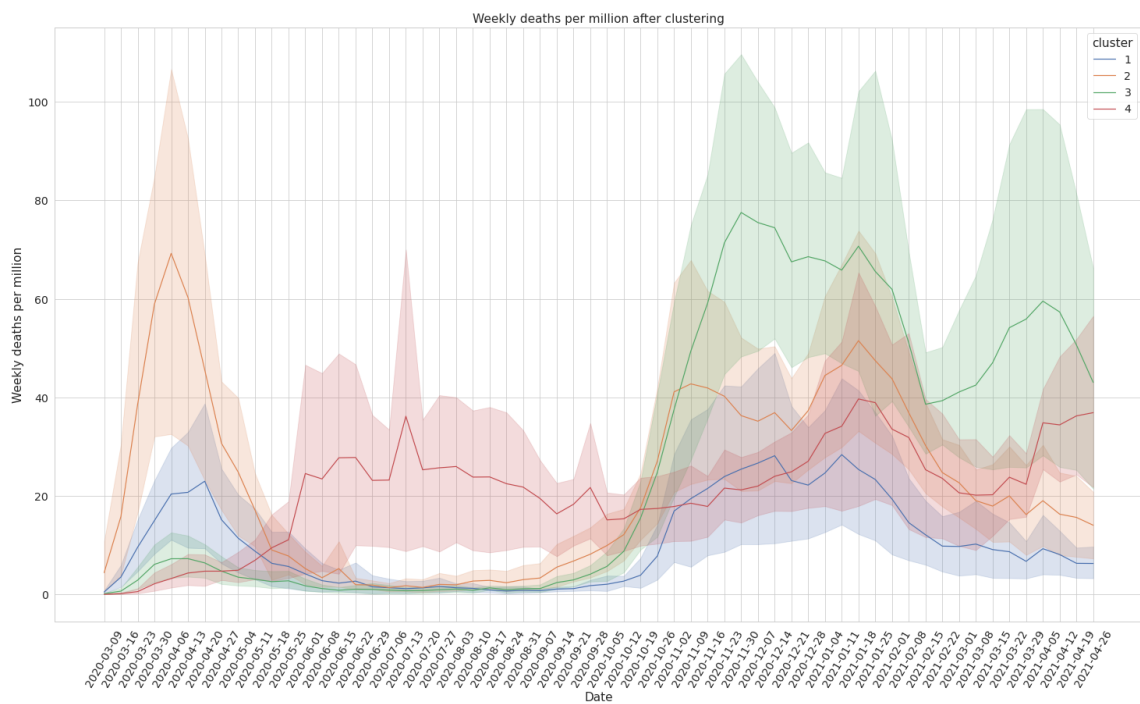
Πίνακας 3.7: Μέση τιμή χαρακτηριστικών ανά συστάδα.



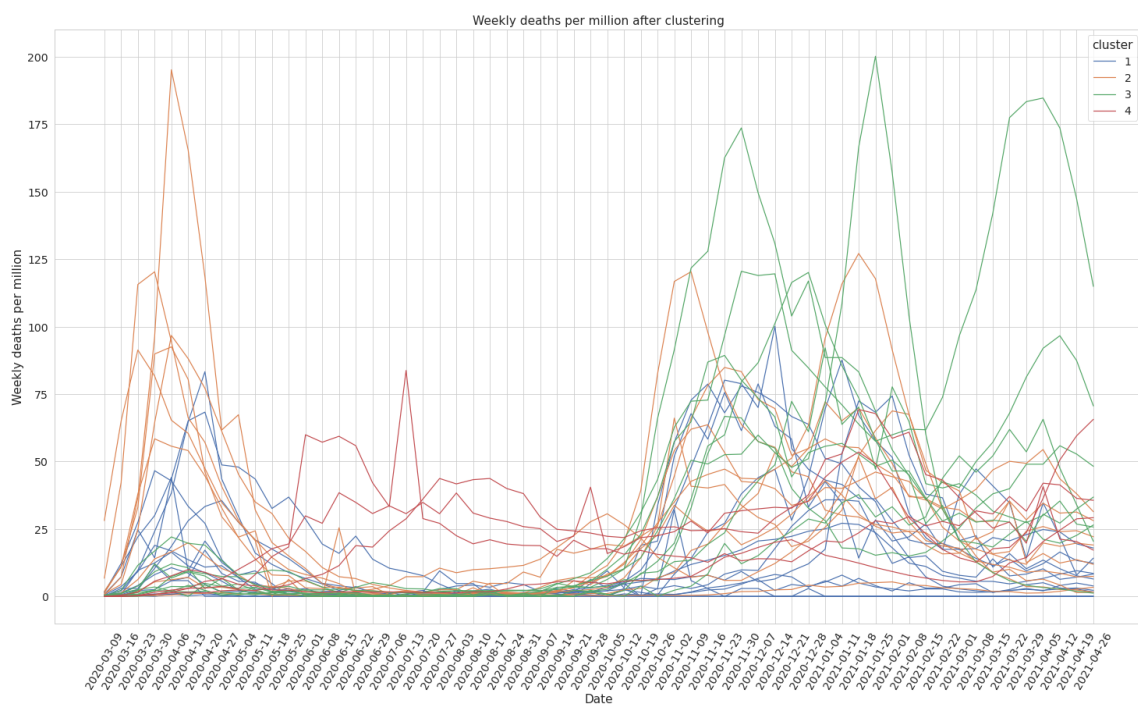
Σχήμα 3.16: Γεωγραφική κατανομή χωρών σε συστάδες.



Σχήμα 3.17: Εβδομαδιαίοι θάνατοι ανά εκατομμύριο που ευθύνονται σε COVID-19 για τις 33 χώρες. Η γραμμή αντιπροσωπεύει τη μέση τιμή των θανάτων ανά εκατομμύριο των χωρών ενώ το διάστημα αντιστοιχεί σε διάστημα εμπιστοσύνης 95% των τιμών κάθε χώρας.



Σχήμα 3.18: Εβδομαδιαίοι θάνατοι ανά εκατομμύριο που ευθύνονται σε COVID-19 για κάθε συστάδα χωρών.



Σχήμα 3.19: Εβδομαδιαίοι θάνατοι ανά εκατομμύριο που ευθύνονται σε COVID-19 των χωρών χωρίς τη συνάθροιση τους σε συστάδες. Το χρώμα της κάθε γραμμής αντιστοιχεί στη συστάδα στην οποία ανήκει η χώρα.

3.4 Συμπεράσματα

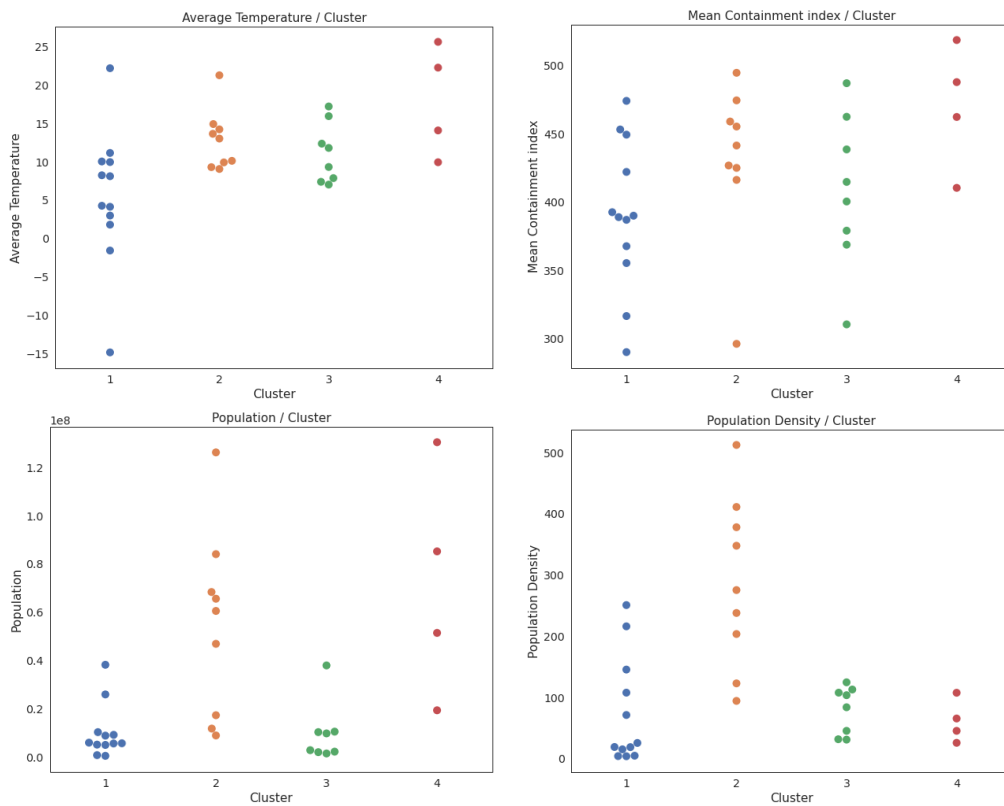
Για την εξαγωγή συμπερασμάτων, απαραίτητη είναι η ανάλυση της κατανομής των χωρών της κάθε συστάδας σε καθένα από τα χαρακτηριστικά, ώστε να εξηγηθεί το πώς αυτά οδηγούν στις συμπεριφορές που παρουσιάστηκαν παραπάνω. Για το σκοπό αυτό, εκτός από τον πίνακα 3.7, βοηθητικό είναι και το σχήμα 3.20, στο οποίο παρουσιάζονται οι τιμές των χωρών κάθε συστάδας για τα οκτώ διαφορετικά χαρακτηριστικά.

Από τα δεδομένα αυτά, ενδιαφέρον παρουσιάζει ότι οι χώρες των συστάδων 2 και 4 αποτελούν τις χώρες με τον μεγαλύτερο πληθυσμό, ωστόσο αυτές του cluster 2 έχουν ταυτόχρονα και τη μεγαλύτερη πυκνότητα πληθυσμού. Η μεγάλη πυκνότητα πληθυσμού πιθανόν να ευθύνεται σε μεγάλο βαθμό στην έκταση του πρώτου κύματος στις χώρες του cluster 2, καθώς τα μεγάλα αστικά κέντρα αποτέλεσαν εύφορο έδαφος για την γρήγορη εξάπλωση της COVID-19 προτού προσαρμοστούν κατάλληλα οι κυβερνητικές πολιτικές και η ανθρώπινη συμπεριφορά. Ακόμη, οι χώρες και των δύο συστάδων υποδέχονται τους περισσότερους τουρίστες, με τις χώρες του cluster 2 να υποδέχονται κατά μέσο όρο όσους τουρίστες υποδέχονται οι χώρες των υπόλοιπων clusters συνολικά. Το γεγονός αυτό, επιδεινώνει τη διασπορά της COVID-19, λόγω της συμπεριφοράς των τουριστών, αλλά και της πιθανότητας γέννησης νέων μεταλλάξεων του κορωνοϊού.

Οι χώρες του cluster 4 εμφανίζουν τη μικρότερη διάμεση ηλικία, διάμεσο εισόδημα ανά νοικοκυριό και κατα κεφαλήν κόστος ιατροφαρμακευτικής περίθαλψης. Τα τρία αυτά χαρακτηριστικά αποτελούν ενδείξεις χωρών χαμηλού βιοτικού επιπέδου, γεγονός που πιθανόν συντέλεσε στην αδυναμία τους να θέσουν την COVID-19 υπό έλεγχο καθ' όλη τη διάρκεια της πανδημίας.

Ένα επιπλέον φαινόμενο που αξίζει να μελετηθεί είναι η μεγάλη διαφορά θανάτων κατά το δεύτερο κύμα της πανδημίας των χωρών των συστάδων 1 και 3, όπου για τις χώρες του cluster 3 αυτό ήταν πολύ πιο έντονο και παρουσιάζει επανέξαρση ως τρίτο κύμα. Οι χώρες των δύο αυτών συστάδων παρουσιάζουν αρκετά κοινά χαρακτηριστικά με εξαίρεση το διάμεσο εισόδημα και το κόστος ιατροφαρμακευτικής περίθαλψης, που είναι χαμηλότερο στις χώρες του cluster 3. Οι ενδείξεις αυτές οδηγούν στο συμπέρασμα ότι οι χώρες με καλύτερο βιοτικό επίπεδο κατάφεραν να κατευάσουν την έξαρση της πανδημίας πολύ αποτελεσματικότερα από χώρες με κατά τα άλλα παρόμοια χαρακτηριστικά, τόσο δημογραφικά όσο και στα μέτρα προστασίας που πάρθηκαν.

Ενδιαφέρον παρουσιάζει και η γεωγραφική κατανομή των χωρών στις συστάδες, η οποία φαίνεται στο σχήμα 3.16. Από αυτό, φαίνεται πως οι χώρες των clusters 2, 3 και 4 βρίσκονται γεωγραφικά κοντά με τις υπόλοιπες χώρες της συστάδας τους. Συγκεκριμένα, οι χώρες του cluster 2 καταλαμβάνουν την κεντρική και νότια Ευρώπη, οι χώρες του cluster 3 εκτείνονται στο ανατολικό κομμάτι της Ευρώπης ενώ οι χώρες του cluster 4 βρίσκονται κυρίως στη Λατινική και Νότια Αμερική. Η παρατήρηση αυτή, σε συνδυασμό με το γεγονός ότι οι διαφορετικές συστάδες παρουσίασαν διαφορετική εξέλιξη της πανδημίας, αποτελεί ένδειξη ότι ο τρόπος αντιμετώπισης της πανδημίας και η αποτελεσματικότητα αυτού ενέχουν και γεωγραφικό παράγοντα. Ο παράγοντας αυτός πιθανόν βασίζεται στο κλίμα (μπορεί η μέση θερμοκρασία να μην διαφέρει πολύ

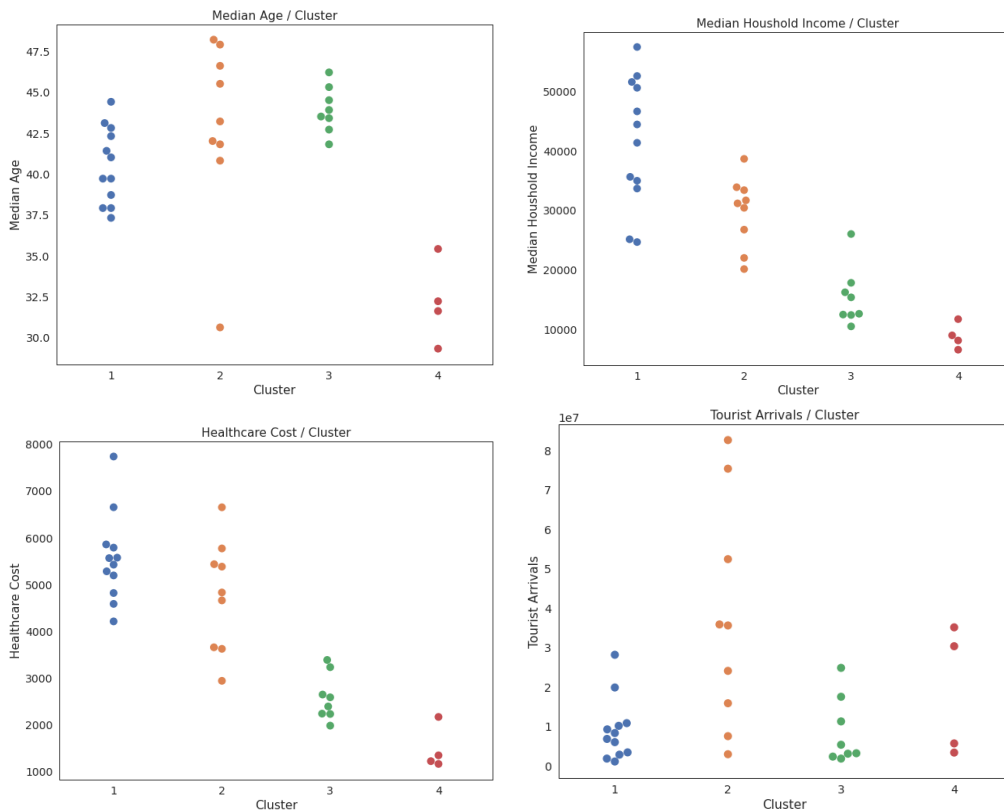


Σχήμα 3.20: Κατανομή χωρών κάθε συστάδας στα χαρακτηριστικά.

από συστάδα σε συστάδα, αλλά δεν περιλαμβάνει στοιχεία όπως τη διάρκεια και την ένταση των εποχών και τη σχετική υγρασία) ή σε κοινωνικά χαρακτηριστικά και συμπεριφορές των κατοίκων των αντίστοιχων γεωγραφικών περιοχών. Για παράδειγμα, το τροπικό κλίμα των χωρών της Λατινικής και Νοτίου Αμερικής, με σχετικά σταθερές θερμοκρασίες γύρω στους 30°C , αποτελούν αρκετά πιο εύφορο έδαφος για τη διασπορά του κορωνοϊού σε σχέση με την ψυχρή Σκανδιναβική χερσόνησο. Η αδυναμία πλήρους αποτύπωσης των κλιματικών στοιχείων μιας χώρας μέσω της μέσης θερμοκρασίας πιθανόν οφείλεται στη σχετικά μικρή συσχέτιση της με τους θανάτους ανά εκατομμύριο στον πίνακα 3.4.

3.5 Η περίπτωση των ΗΠΑ

Όπως αναφέρθηκε προηγουμένως, οι Ηνωμένες Πολιτείες της Αμερικής δεν περιλήφθηκαν στο παραπάνω σύνολο δεδομένων λόγω της μεγάλης ετερογένειας των μέτρων και της γενικότερης κατάστασης της πανδημίας σε κάθε Πολιτεία. Το γεγονός αυτό ωστόσο, αποτελεί ταυτόχρονα ευκαιρία για την επικύρωση των παραπάνω συμπερασμάτων των χωρών στις Πολιτείες. Στο παρακάτω κεφάλαιο παρουσιάζεται η



Σχήμα 3.20: Κατανομή χωρών κάθε συστάδας στα χαρακτηριστικά (συνέχεια).

επανάληψη της ίδιας διαδικασίας για τις 48 Πολιτείες των ΗΠΑ. Οι δύο Πολιτείες που λείπουν από το σύνολο δεδομένων είναι το Wisconsin και το Wyoming, καθώς για αυτές υπήρχαν ελλείψεις στα σύνολα δεδομένων που χρησιμοποιήθηκαν.

3.5.1 Καθορισμός Συνόλου Δεδομένων

Τα δεδομένα που συλλέχθηκαν για την ανάλυση της περίπτωσης των ΗΠΑ είναι τα ίδια με αυτά των χωρών, απλώς προέρχονται από διαφορετικές πηγές. Η μόνη διαφορά στο σύνολο δεδομένων έγκειται στην έλλειψη των τουριστικών αφίξεων σε κάθε Πολιτεία, καθώς δεν βρέθηκε αντίστοιχη πηγή. Οι θάνατοι που ευθύνονται στην COVID-19 και τα μέτρα προστασίας και στήριξης που έλαβαν οι Πολιτείες κατά τη διάρκεια της πανδημίας παρέχονται και αυτά από το αποθετήριο του OxCGRIT στο GitHub[3].

Τα δεδομένα για τη μέση θερμοκρασία κάθε Πολιτείας ελήφθησαν από το National Climatic Data Center[23]. Η αρχική τους μορφή ήταν σε βαθμούς Fahrenheit, ωστόσο μετατράπηκαν σε βαθμούς Κελσίου προκειμένου να υπάρχει κοινό μέτρο σύγκρισης με τις θερμοκρασίες των υπόλοιπων χωρών.

Στοιχεία για τον πληθυσμό κάθε Πολιτείας και την πυκνότητα αυτού, τη διάμεση

Feature	Correlation	T-Value	P-Value
Μέση Θερμοκρασία	0.26	-17.91	$1.23 \cdot 10^{-22}$
Μέσο CHI	0.03	-13.38	$1.01 \cdot 10^{-17}$
Πληθυσμός	0.18	6.19	$1.40 \cdot 10^{-7}$
* Πυκνότητα πληθυσμού	0.48	-14.41	$8.96 \cdot 10^{-22}$
Διάμεσο Εισόδημα	-0.18	40.17	$3.15 \cdot 10^{-38}$
Διάμεση Ηλικία	0.01	-17.61	$2.48 \cdot 10^{-22}$
Κόστος Περίθαλψης	-0.11	36.19	$5.70 \cdot 10^{-49}$

Πίνακας 3.8: Στατιστικά μεγέθη συσχέτισης χαρακτηριστικών με θανάτους ανά εκατομμύριο των Πολιτειών. Με αστερίσκο και έντονη γραμματοσειρά σημειώνονται τα χαρακτηριστικά που εμφανίζουν στατιστικά σημαντική συσχέτιση με τους θανάτους.

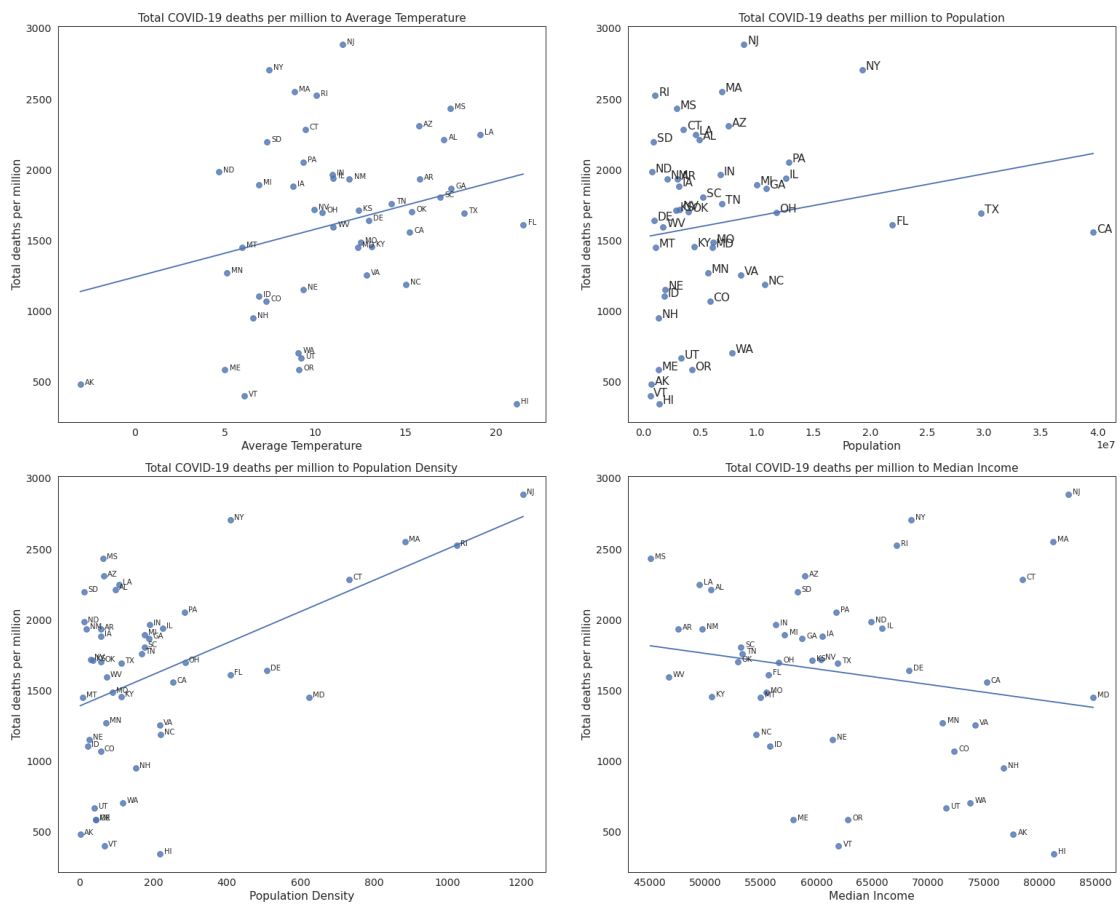
ηλικία και το διάμεσο εισόδημα ανά νοικοκυριό για κάθε Πολιτεία παρέχονται από το United States Census Bureau[33, 31, 32].

Τέλος, το κατά κεφαλήν κόστος ιατροφαρμακευτικής περίθαλψης για κάθε Πολιτεία παρέχεται από το KFF (Kaiser Family Foundation)[20].

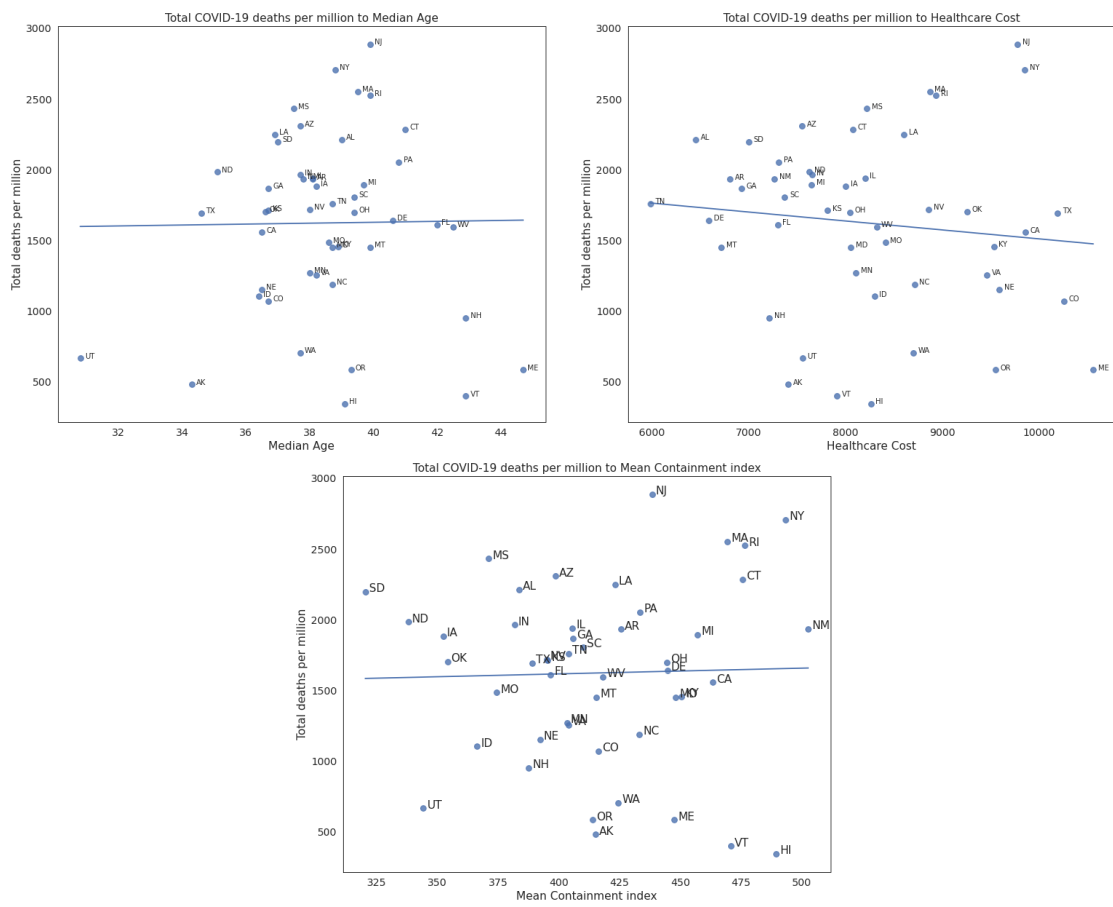
3.5.2 Επίδραση Δεδομένων

Για την αξιολόγηση της πορείας της πανδημίας στις Ηνωμένες Πολιτείες της Αμερικής, όπως και στην περίπτωση των χωρών, θα χρησιμοποιηθούν οι θάνατοι ανά εκατομμύριο που ευθύνονται στην COVID-19. Στο υποκεφάλαιο αυτό, μελετάται η συσχέτιση των θανάτων ανά εκατομμύριο με τα υπόλοιπα χαρακτηριστικά, προκειμένου για την κατανόηση της επίδρασης των δευτέρων στην εξέλιξη της πανδημίας. Στο σχήμα 3.21 παρατίθενται τα διαγράμματα διασποράς όλων των χαρακτηριστικών με τους θανάτους ανά εκατομμύριο για τις Πολιτείες. Στη συνέχεια, στον πίνακα 3.8 παρουσιάζονται ο συντελεστής συσχέτισης, και οι t, p τιμές των αντίστοιχων χαρακτηριστικών με τους θανάτους ανά εκατομμύριο. Με αστερίσκο και έντονη γραμματοσειρά σημειώνονται τα χαρακτηριστικά εκείνα που εμφανίζουν στατιστικά σημαντική συσχέτιση με τους θανάτους.

Στην περίπτωση των Ηνωμένων Πολιτειών, παρατηρείται διαφορετική συμπεριφορά από αυτή των χωρών. Ο μοναδικός παράγοντας που παρουσιάζει στατιστικά σημαντική συσχέτιση με τους θανάτους ανά εκατομμύριο είναι η πυκνότητα πληθυσμού της κάθε Πολιτείας. Μάλιστα, όπως είναι αναμενόμενο, η συσχέτιση αυτή είναι θετική, καθώς οι Πολιτείες με μεγαλύτερη πυκνότητα πληθυσμού, έχουν μεγαλύτερα αστικά κέντρα με περισσότερη επαφή ανθρώπων μεταξύ τους. Επομένως, είναι περισσότερη η διασπορά του κορωνοϊού. Το γεγονός ότι το μέσο Containment and Health Index και το διάμεσο εισόδημα δεν εμφανίζουν σημαντική συσχέτιση με τους θανάτους ανά εκατομμύριο, σε αντίθεση με την περίπτωση των χωρών, μπορεί να ερμηνευτεί ως...



Σχήμα 3.21: Διαγράμματα σχέδασης θανάτων ανά εκατομμύριο με τα υπόλοιπα χαρακτηριστικά των Πολιτειών.

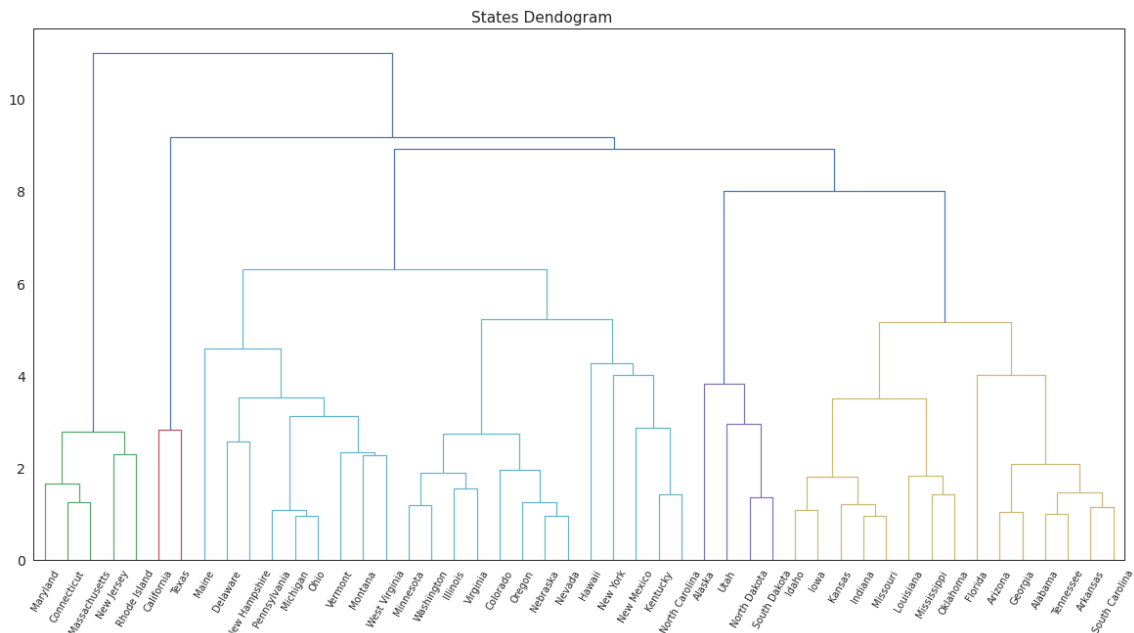


Σχήμα 3.21: Διαγράμματα σχέδασης θανάτων ανά εκατομμύριο με τα υπόλοιπα χαρακτηριστικά των Πολιτειών (συνέχεια).

3.5.3 Ανάλυση Δεδομένων

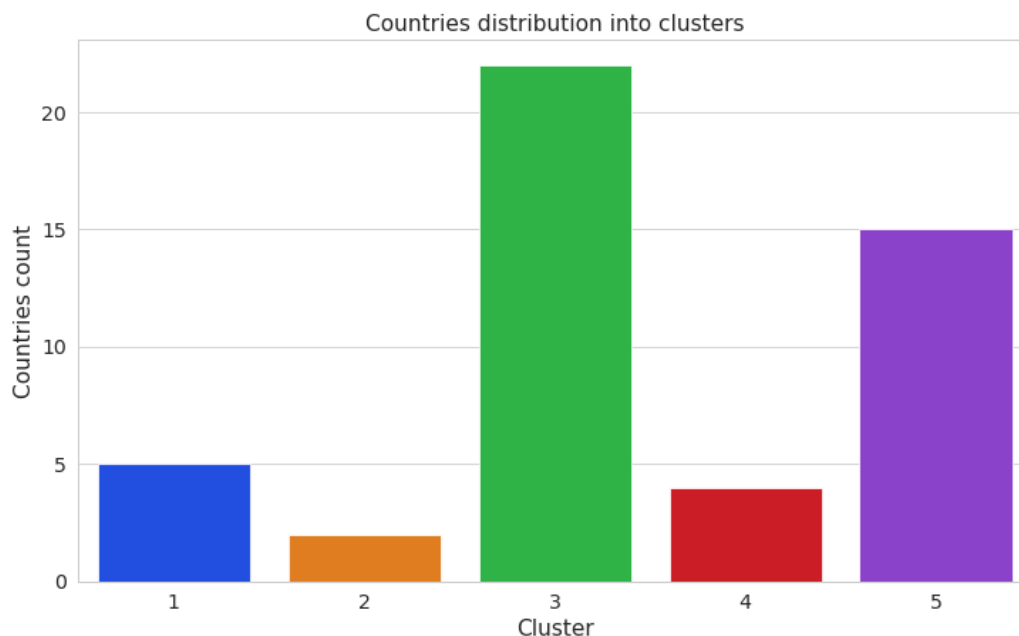
Στο σχήμα 3.22 φαίνεται το δενδρόγραμμα της συσσωρευτικής ιεραρχικής συσταδοποίησης των Πολιτειών για τα χαρακτηριστικά του συνόλου δεδομένων που περιγράφηκαν παραπάνω. Υπενθυμίζεται ότι οι θάνατοι ανά εκατομμύριο δεν συμπεριλαμβάνονται στο σύνολο δεδομένων, αλλά χρησιμεύουν ως μέσο ερμηνείας της επίδοσης των χωρών κάθε συστάδας κατά της πανδημίας. Σύμφωνα με το δενδρόγραμμα, οι Πολιτείες χωρίζονται στις εξής πέντε διακριτές συστάδες:

1. Maryland, Connecticut, Massachusetts, New Jersey, Rhode Island.
2. California, Texas.
3. Maine, Delaware, New Hampshire, Pennsylvania, Michigan, Ohio, Vermont, Montana, West Virginia, Minnesota, Washington, Illinois, Virginia, Colorado, Oregon, Nebraska, Nevada, Hawaii, New York, New Mexico, Kentucky, North Carolina.
4. Alaska, Utah, North Dakota, South Dakota.
5. Idaho, Iowa, Kansas, Indiana, Missouri, Louisiana, Mississippi, Oklahoma, Florida, Arizona, Georgia, Alabama, Tennessee, Arkansas, South Carolina.



Σχήμα 3.22: Δενδρόγραμμα συσσωρευτικής ιεραρχικής συσταδοποίησης Πολιτειών.

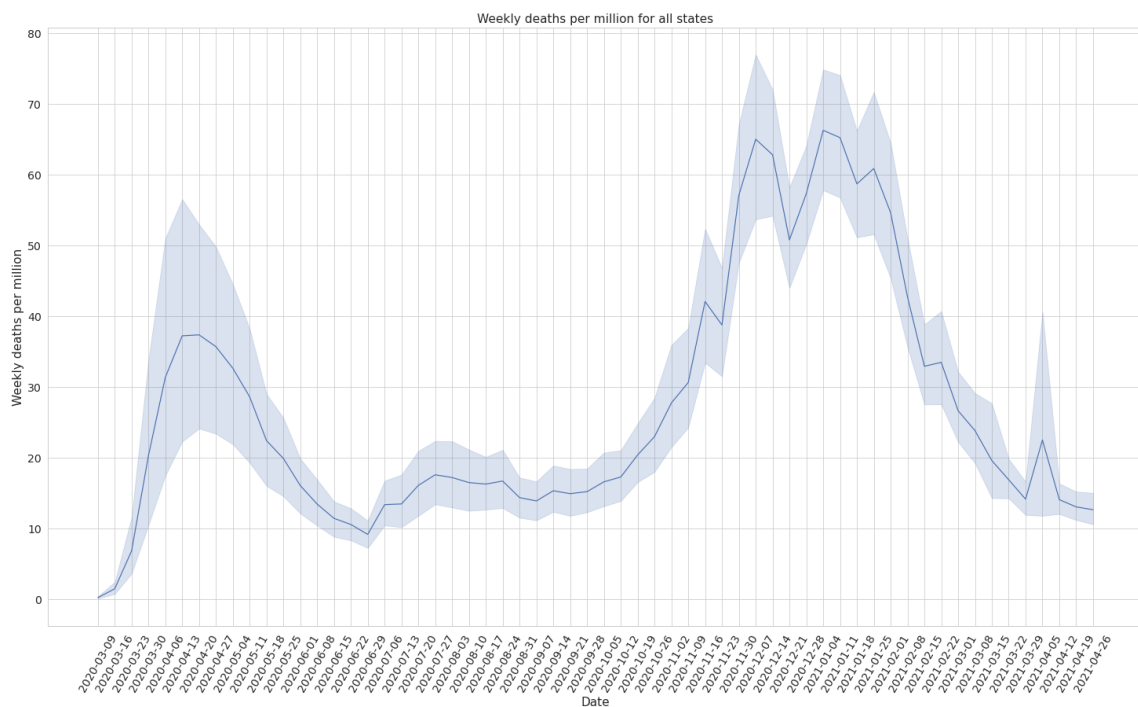
Στο διάγραμμα του σχήματος 3.23 φαίνεται η κατανομή των Πολιτειών στις πέντε συστάδες. Η μεγαλύτερη συστάδα είναι η τρίτη με 22 Πολιτείες, ενώ η μικρότερη είναι η δεύτερη με μόνο δύο Πολιτείες.



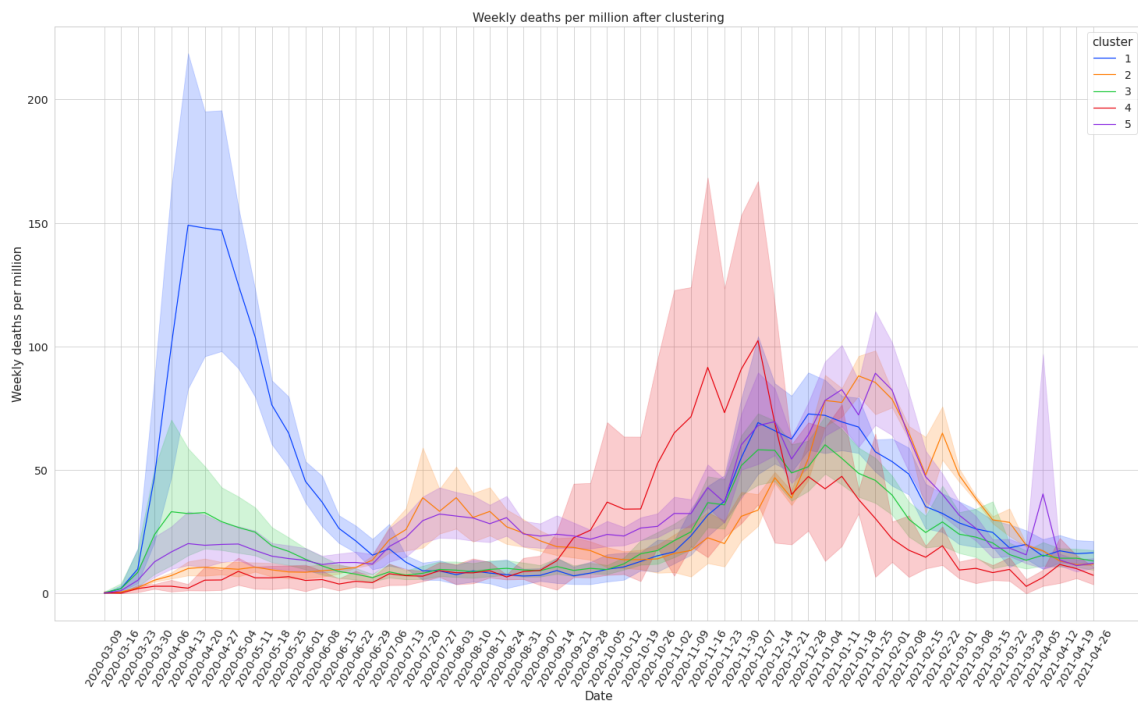
Σχήμα 3.23: Κατανομή Πολιτειών σε συστάδες.

Στο διάγραμμα 3.24 φαίνεται η πορεία της πανδημίας σχετικά με τους θανάτους ανά εκατομμύριο συνολικά για τις Ηνωμένες Πολιτείες της Αμερικής. Η γραμμή αντιπροσωπεύει τη μέση τιμή των Πολιτειών ενώ το διάστημα αντιστοιχεί σε διάστημα εμπιστοσύνης 95%. Το διάγραμμα αυτό θυμίζει αρκετά το αντίστοιχο των χωρών (3.17) σε επίπεδο ύπαρξης των δύο κυμάτων της πανδημίας, με τη διαφορά ότι σε αυτή την περίπτωση φαίνεται όλες οι Πολιτείες να κατάφεραν να θέσουν υπό έλεγχο το δεύτερο κύμα στις αρχές του 2021.

Στο σχήμα 3.25 φαίνεται το αντίστοιχο διάγραμμα με ξεχωριστή γραμμή για κάθε συστάδα. Η πρώτη παρατήρηση που εξάγεται από το διάγραμμα αυτό είναι ότι οι Πολιτείες του cluster 1 ήταν αυτές που υπέφεραν περισσότερο κατά το πρώτο κύμα της πανδημίας. Μάλιστα, οι θάνατοι ανά εκατομμύριο των Πολιτειών της συστάδας αυτής ήταν πολύ περισσότεροι από τις υπόλοιπες Πολιτείες, αφού ο μέσος όρος τους ξεπέρασε τους 150, τη στιγμή που το 95% διάστημα εμπιστοσύνης των θανάτων ανά εκατομμύριο όλων των Πολιτειών δεν ξεπέρασε τους 60. Μία ακόμη παρατήρηση είναι το γεγονός ότι οι Πολιτείες όλων των συστάδων παρουσίασαν σχετικά παρόμοια εικόνα κατά το δεύτερο κύμα της πανδημίας, δυσκολεύοντας την εξαγωγή ξεκάθαρων συμπερασμάτων για αυτή τη χρονική περίοδο. Μικρή εξαίρεση αποτελούν οι Πολιτείες του cluster 4, οι οποίες παρουσίασαν την έξαρση του δεύτερου κύματος σχετικά νωρίτερα και με μεγαλύτερη ένταση από τις υπόλοιπες συστάδες στην περίοδο του



Σχήμα 3.24: Εβδομαδιαίοι θάνατοι ανά εκατομμύριο που ευθύνονται σε COVID-19 για τις 48 Πολιτείες. Η γραμμή αντιπροσωπεύει τη μέση τιμή των θανάτων ανά εκατομμύριο των Πολιτειών ενώ το διάστημα αντιστοιχεί σε διάστημα εμπιστοσύνης 95% των τιμών κάθε Πολιτείας.



Σχήμα 3.25: Εβδομαδιαίοι θάνατοι ανά εκατομμύριο που ευθύνονται σε COVID-19 για κάθε συστάδα Πολιτειών.

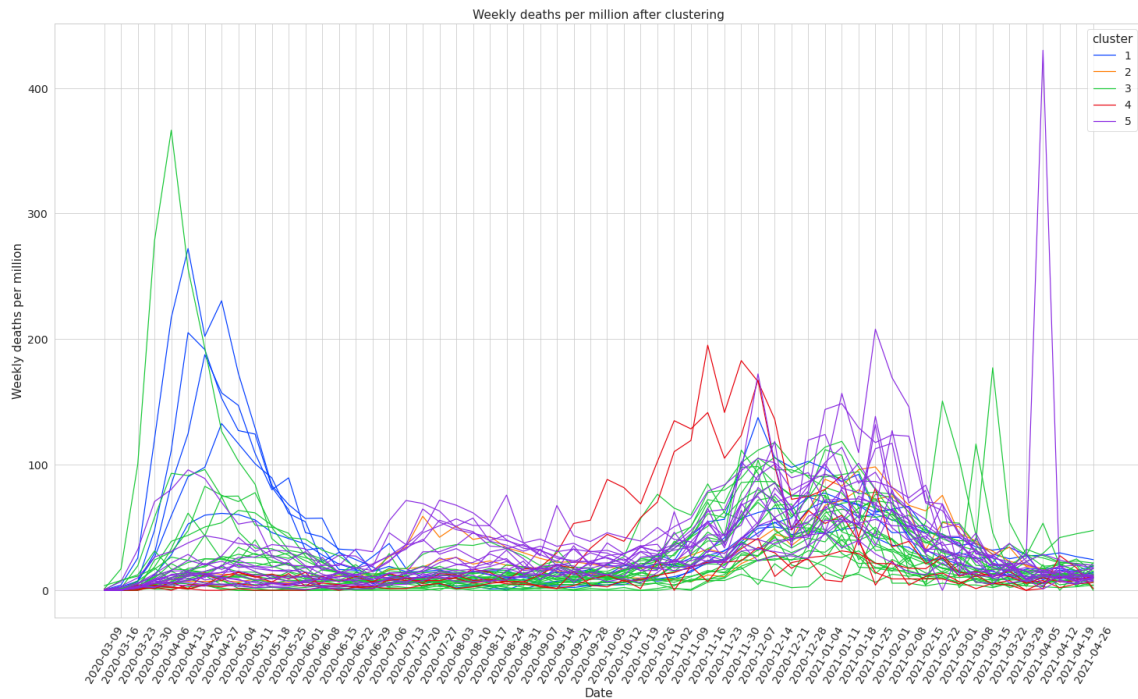
Οκτωβρίου - Δεκεμβρίου του 2020, αλλά κατάφεραν να το θέσουν υπό ύφεση αμέσως μετά.

Για λόγους πληρότητας, στο σχήμα 3.26 παρουσιάζεται η πορεία των Πολιτειών της κάθε συστάδας, με ξεχωριστή γραμμή για κάθε χώρα. Από το διάγραμμα αυτό επιβεβαιώνεται η παρόμοια εικόνα που εμφάνισαν όλες οι Πολιτείες κατά τη διάρκεια του δεύτερου κύματος της πανδημίας.

3.5.4 Συμπεράσματα

Για την εξαγωγή συμπερασμάτων σχετικά με τα αίτια της πορείας της πανδημίας που παρατηρήθηκε στο προηγούμενο κεφάλαιο, όπως και στην περίπτωση των χωρών, είναι απαραίτητη η εξέταση της κατανομής των Πολιτειών της κάθε συστάδας σε καθένα από τα χαρακτηριστικά του συνόλου δεδομένων. Στον πίνακα 3.9 παρατίθεται η μέση τιμή των χαρακτηριστικών των Πολιτειών κάθε συστάδας, ενώ στο σχήμα 3.27 οπτικοποιούνται οι τιμές των Πολιτειών κάθε συστάδας για καθένα από τα επτά χαρακτηριστικά.

Ο πίνακας και τα διαγράμματα αυτά επιβεβαιώνουν την υπόθεση που είχε σχηματιστεί στην περίπτωση των χωρών και τη συσχέτιση του πίνακα 3.8, ότι οι περιοχές με μεγάλη πυκνότητα πληθυσμού αποτέλεσαν πηγή υπερμετάδοσης του κορωνοϊού στο πρώτο κύμα προτού προσαρμοστούν οι κυβερνήσεις και η ανθρώπινη συμπεριφορά. Οι



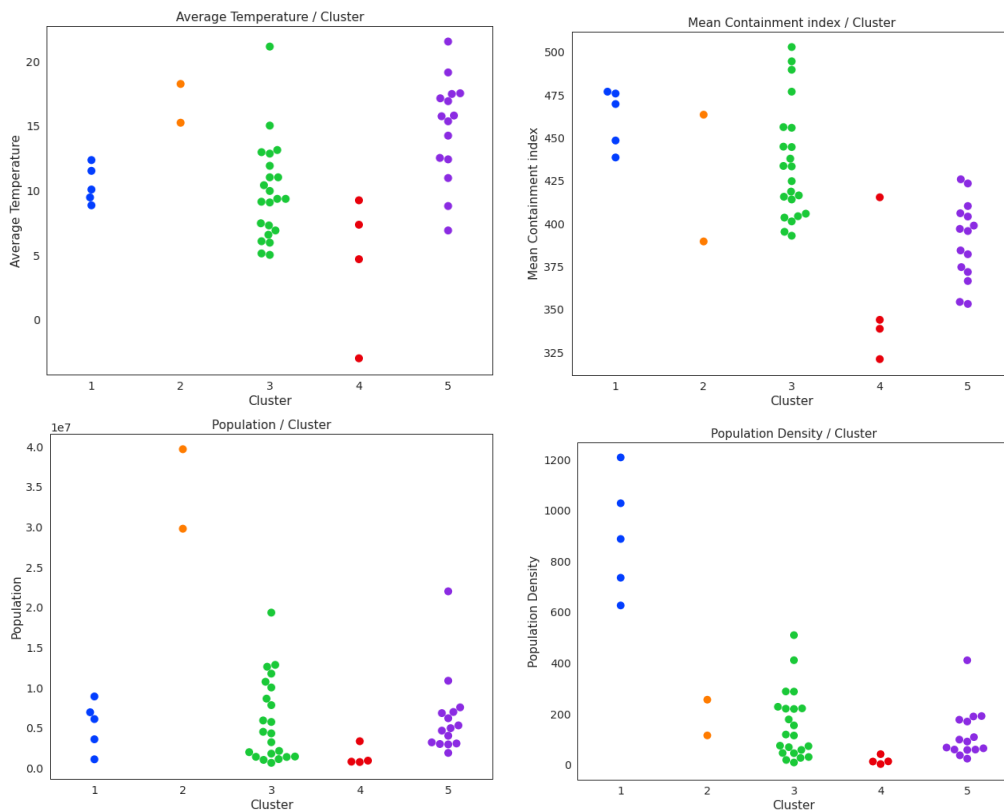
Σχήμα 3.26: Εβδομαδιαίοι θάνατοι ανά εκατομμύριο που ευθύνονται σε COVID-19 των Πολιτειών χωρίς τη συνάθροιση τους σε συστάδες. Το χρώμα της κάθε γραμμής αντιστοιχεί στη συστάδα στην οποία ανήκει η Πολιτεία.

Cluster	1	2	3	4	5
Θάνατοι ανά εκατομμύριο	2399.84	1624.44	1362.75	1436.32	1721.30
Μέσο εβδομαδιαίο CHI	466.92	426.20	435.44	382.53	426.93
Μέση Θερμοκρασία	9.94	16.72	8.23	9.27	16.58
Πληθυσμός (εκατομ.)	7.63	34.67	4.63	4.40	6.98
Πυκνότητα Πληθυσμού	814.62	184.05	178.17	72.09	153.37
Διάμεσο Εισόδημα	77110.33	68554.50	60235.22	63170.58	55303.83
Διάμεση Ηλικία	39.63	35.55	41.49	36.92	38.58
Κόστος Περίθαλψης	8927.33	10024.50	7812.00	8419.68	7563.00

Πίνακας 3.9: Μέση τιμή χαρακτηριστικών ανά συστάδα για τις Πολιτείες.

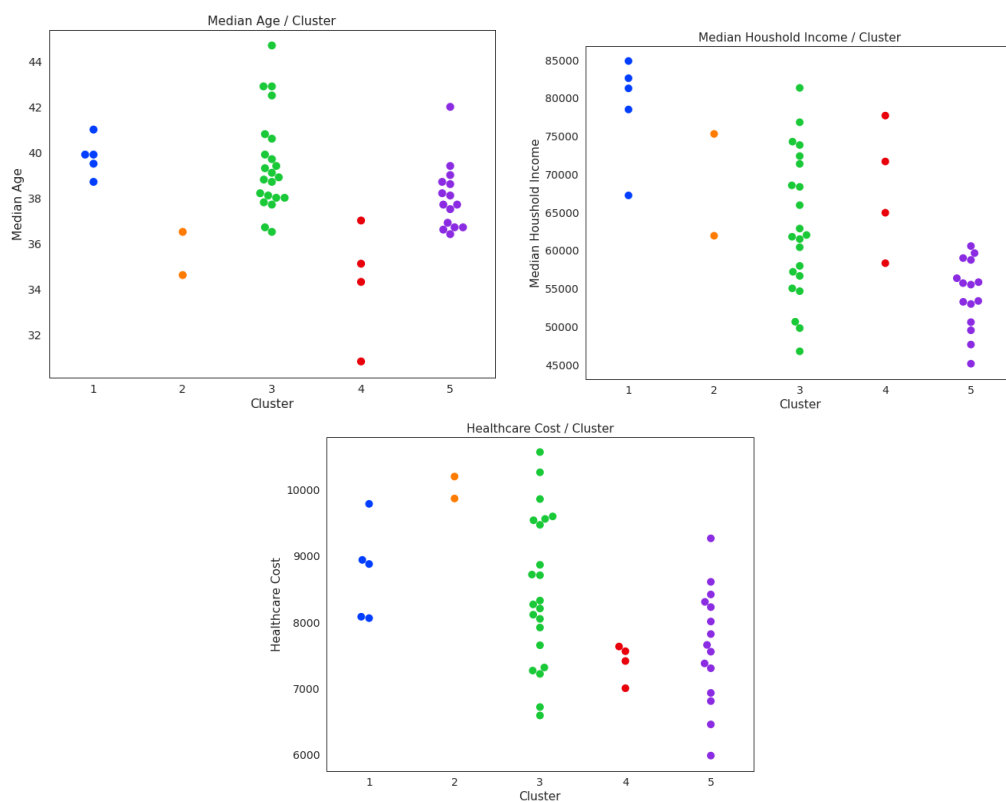
Πολιτείες του cluster 1, οι οποίες όπως αναφέρθηκε προηγουμένως παρουσίασαν με διαφορά τους περισσότερους θανάτους ανά εκατομμύριο στο πρώτο κύμα της πανδημίας, είναι οι Πολιτείες με τη μακράν μεγαλύτερη πυκνότητα πληθυσμού. Ωστόσο, οι Πολιτείες αυτές δεν ξεχωρίζουν ως προς τον πληθυσμό σε σχέση με τις υπόλοιπες συστάδες, γεγονός που αποτελεί ένδειξη ότι η πυκνότητα πληθυσμού ευθύνεται αποκλειστικά για τη συμπεριφορά αυτή, ανεξάρτητα από το συνολικό πληθυσμό.

Η άφιξη του δεύτερου κύματος της πανδημίας νωρίς το φθινόπωρο του 2020 για τις Πολιτείες του cluster 4, ενδεχομένως να ευθύνεται στο ότι οι Πολιτείες της συστάδας αυτής είχαν το χαμηλότερο μέσο Containment and Health Index κατά τη διάρκεια της πανδημίας. Ωστόσο, στην περίπτωση των χωρών δεν υπήρξαν επαρκή στοιχεία που να επιβεβαιώνουν μία τέτοια υπόθεση.



Σχήμα 3.27: Κατανομή Πολιτειών κάθε συστάδας στα χαρακτηριστικά.

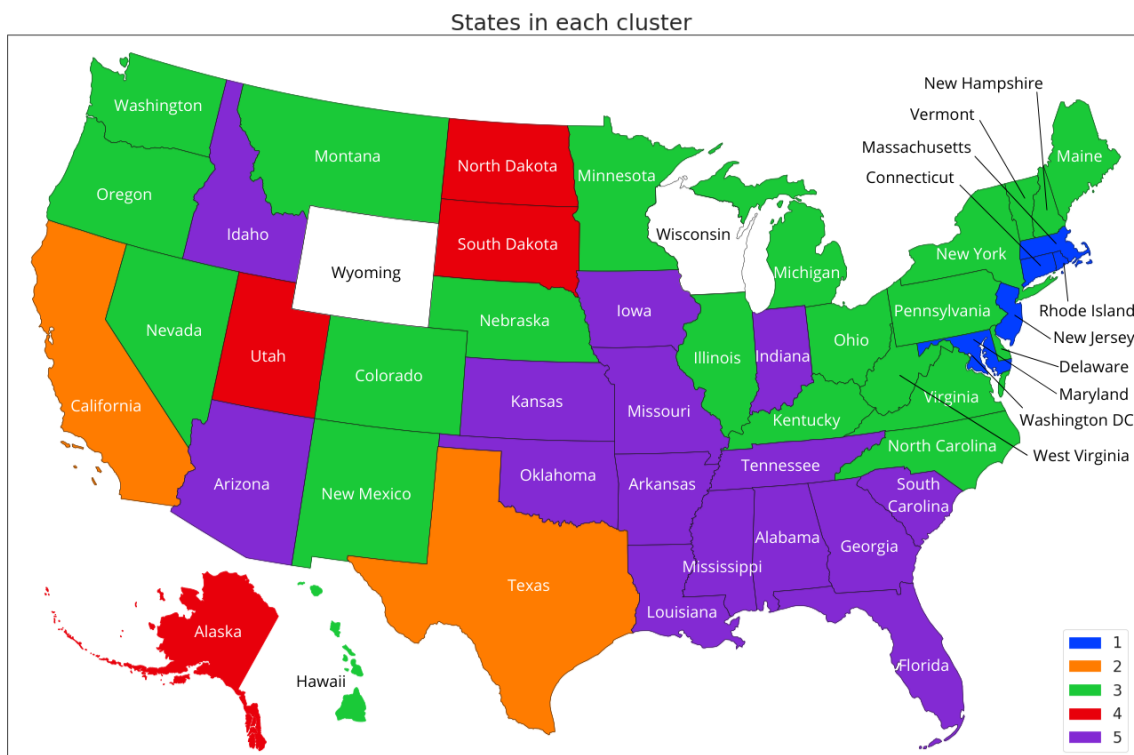
Ενδιαφέρον παρουσιάζει και σε αυτή την περίπτωση η γεωγραφική κατανομή των Πολιτειών σε κάθε συστάδα. Η κατανομή αυτή φαίνεται στον χάρτη του σχήματος 3.28, στον οποίο η κάθε Πολιτεία είναι χρωματισμένη ανάλογα με τη συστάδα στην οποία ανήκει. Το Wyoming και το Wisconsin δεν ανήκουν σε καμία συστάδα καθώς έχουν εξαιρεθεί από το σύνολο δεδομένων όπως αναφέρθηκε προηγουμένως. Από το χάρτη αυτό φαίνεται αρχικά ότι οι χώρες του cluster 1, οι οποίες είναι και αυτές που



Σχήμα 3.27: Κατανομή Πολιτειών κάθε συστάδας στα χαρακτηριστικά (συνέχεια).

παρουσίασαν το σφοδρότερο πρώτο κύμα βρίσκονται όλες στη βορειοανατολική ακτή των ΗΠΑ. Δεδομένου ότι το κύριο χαρακτηριστικό των Πολιτειών αυτών είναι η μεγάλη πυκνότητα πληθυσμού τους, σημαίνει ότι είτε η γεωγραφική αυτή εγκύτητα είναι τυχαία, είτε ότι υπάρχει κάποιος ακόμη παράγοντας που σχετίζεται με τα χαρακτηριστικά της συγκεκριμένης γεωγραφικής περιοχής και συμβάλλει είτε στην αυξημένη πυκνότητα πληθυσμού, είτε στην ραγδαία εξάπλωση του κορωνοϊού στο πρώτο κύμα της πανδημίας.

Επιπλέον, οι Πολιτείες του cluster 5 βρίσκονται ως επί το πλείστον στα νοτιοανατολικά της χώρας. Το συγκεκριμένο συμβάν πιθανόν να οφείλεται στο γεγονός ότι οι Πολιτείες της συστάδας αυτής χαρακτηρίζονται από σχετικά υψηλές μέσες θερμοκρασίες και χαμηλή πυκνότητα πληθυσμού και συνολικό πληθυσμό. Από γεωγραφική σκοπιά, το κλίμα της περιοχής αυτής χαρακτηρίζεται από θερμές και με υψηλή υγρασία συνθήκες, οι οποίες όπως αναφέρθηκε νωρίτερα ευνοούν την επιβίωση του SARS-CoV-2 σε επιφάνειες.



Σχήμα 3.28: Γεωγραφική κατανομή Πολιτειών σε συστάδες.

Κεφάλαιο 4

Ανάλυση Χρονοσειρών

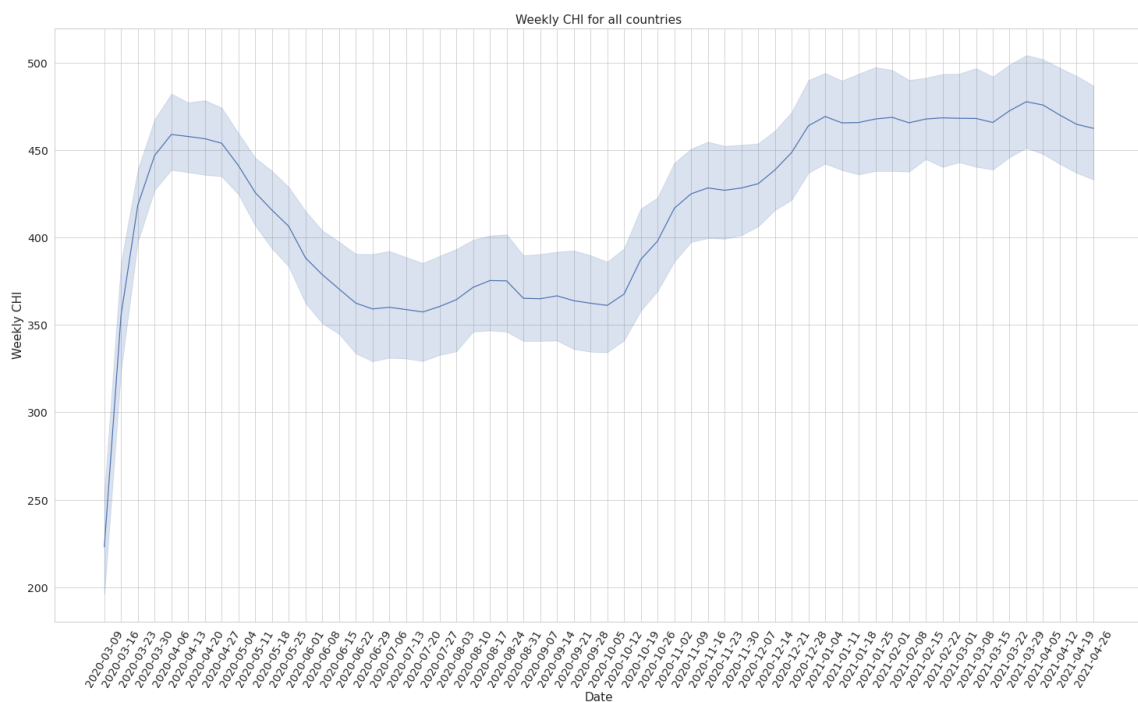
Στο προηγούμενο κεφάλαιο μελετήθηκε η πορεία των χωρών και Πολιτειών κατά τη διάρκεια της πανδημίας COVID-19 ως προς συγκεντρωτικά μεγέθη, από τα οποία απουσίαζε η έννοια του χρόνου. Στο κεφάλαιο αυτό, θα εξετασθεί πώς οι θάνατοι ανά εκατομμύριο των χωρών και Πολιτειών που μελετήθηκαν επηρεάστηκαν και επηρέασαν διάφορα μεγέθη εν μέσω της πανδημίας. Τα μεγέθη είναι τα μέτρα προστασίας και στήριξης που έλαβαν οι κυβερνήσεις, υπό τη μορφή του Containment and Health Index που παρουσιάστηκε στο προηγούμενο κεφάλαιο και οι συγκεντρωτικές εβδομαδιαίες προβλέψεις θανάτων διαφόρων ερευνητικών ομάδων για τις ΗΠΑ.

4.1 Επίδραση Μέτρων Προστασίας σε Επίπεδο Χρόνου

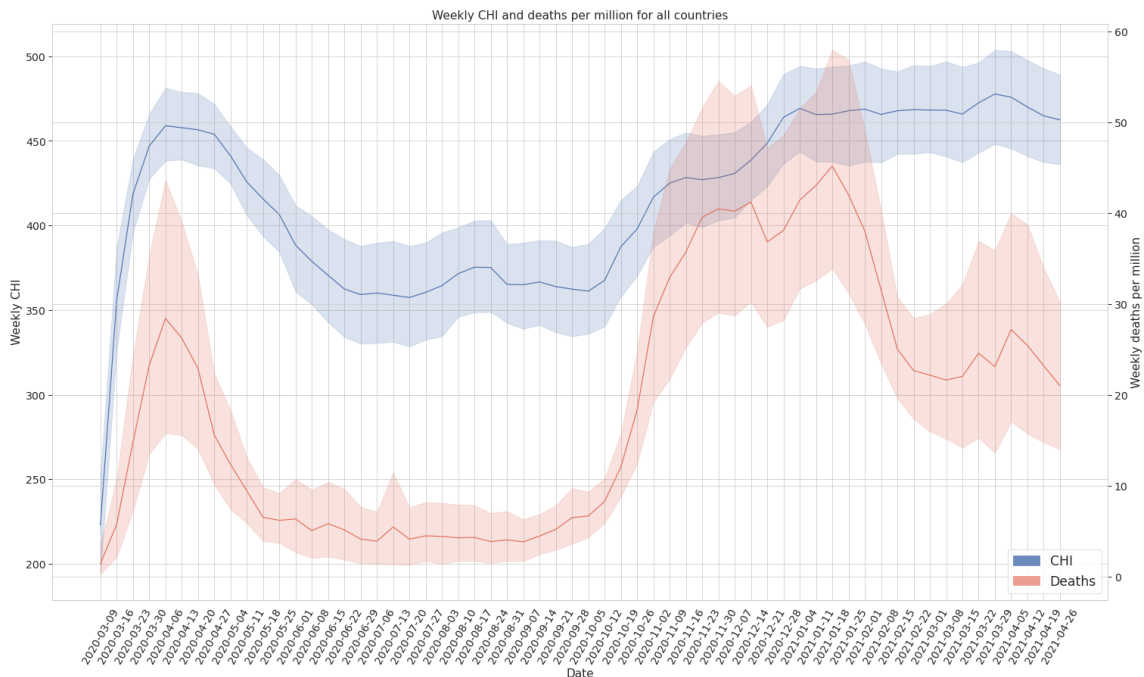
4.1.1 Επίδραση Θανάτων στη Λήψη Μέτρων Προστασίας

Ενδιαφέρον παρουσιάζει η μελέτη της επίδρασης των θανάτων που ευθύνονται στην COVID-19 της κάθε χώρας στη λήψη ή τη χαλάρωση των μέτρων προστασίας και η επίδραση των εν λόγω μέτρων με τη σειρά τους στην ύφεση ή την έξαρση της πανδημίας. Στο σχήμα 4.1 φαίνεται η γενική πορεία των μέτρων προστασίας που ελήφθησαν από τις χώρες στην πορεία του χρόνου. Από το διάγραμμα αυτό φαίνεται η πολύ ξαφνική ανάληψη μέτρων στην αρχή του πρώτου κύματος της πανδημίας το Μάρτιο του 2020, η μερική χαλάρωση τους το καλοκαίρι και το Σεπτέμβριο του ίδιου έτους και τέλος η επανισχυροποίηση τους από το φθινόπωρο του 2020 μέχρι και το τέλος του Απριλίου του 2021.

Η μορφή του διαγράμματος θυμίζει αρκετά τη μορφή των θανάτων ανά εκατομμύριο των χωρών. Μάλιστα, στο σχήμα 4.2 παρουσιάζονται τα δύο αυτά μεγέθη μαζί, προκειμένου να φανεί καλύτερα η συσχέτιση τους. Οι γραμμές των μέσων τιμών, όπως είναι εμφανές, συσχετίζονται σημαντικά θετικά μεταξύ τους ($r = 0.75$, με



Σχήμα 4.1: Εξέλιξη του εβδομαδιαίου Containment and Health Index για όλες τις χώρες. Η γραμμή αντιπροσωπεύει τη μέση τιμή του δείκτη την αντίστοιχη εβδομάδα, ενώ το διάστημα αντιστοιχεί σε διάστημα εμπιστοσύνης 95% των τιμών κάθε χώρας.

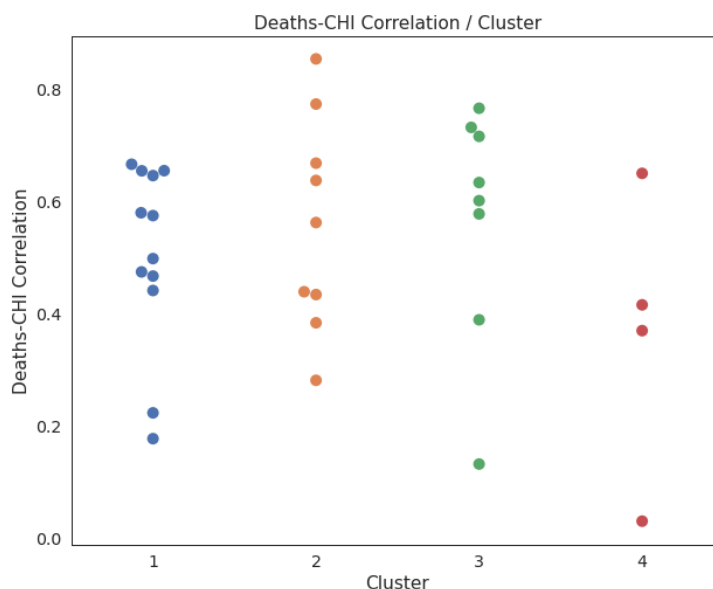


Σχήμα 4.2: Εξέλιξη του εβδομαδιαίου Containment and Health Index και των εβδομαδιαίων θανάτων ανά εκατομμύριο για όλες τις χώρες. Η γραμμή αντιπροσωπεύει τη μέση τιμή του δείκτη την αντίστοιχη εβδομάδα, ενώ το διάστημα αντιστοιχεί σε διάστημα εμπιστοσύνης 95% των τιμών κάθε χώρας.

$p = 1.08 \cdot 10^{-59}$).

Τα παραπάνω στοιχεία δείχνουν ότι οι κυβερνήσεις σε γενικές γραμμές έλαβαν και χαλάρωσαν τα μέτρα προστασίας και στήριξης ανάλογα με την κατάσταση των θανάτων λόγω COVID-19 που επικρατούσε την εκάστοτε περίοδο. Ωστόσο, ενδιαφέρον παρουσιάζει η μελέτη του κατά πόσο χώρες από τις ίδιες συστάδες, όπως αυτές προέκυψαν στο προηγούμενο κεφάλαιο, παρουσίασαν διαφορετική αντιμετώπιση σε σχέση με αυτές των υπόλοιπων συστάδων. Στο σχήμα 4.3 φαίνεται η κατανομή της συσχέτισης εβδομαδιαίων θανάτων ανά εκατομμύριο λόγω COVID-19 με το εβδομαδιαίο Containment and Health Index των χωρών της κάθε συστάδας. Κάθε συστάδα αποτελείται ταυτόχρονα από χώρες με υψηλή αλλά και χαμηλή (θετική) συσχέτιση των θανάτων με το Containment and Health Index. Επομένως, η πολιτική των χωρών όλων των συστάδων φαίνεται να ακολουθεί την πορεία των θανάτων στην εκάστοτε χώρα.

Ο ισχυρισμός αυτός επιβεβαιώνεται και από το διάγραμμα του σχήματος 4.4, στο οποίο φαίνεται η μέση τιμή και το διάστημα εμπιστοσύνης 95% του εβδομαδιαίου Containment and Health Index για τις χώρες της κάθε συστάδας. Στο διάγραμμα αυτό, φαίνονται οι χώρες των τριών πρώτων συστάδων να αποσύρουν κάποια μέτρα στην περίοδο μεταξύ των δύο κυμάτων της πανδημίας και να τα επανισχυροποιούν με την

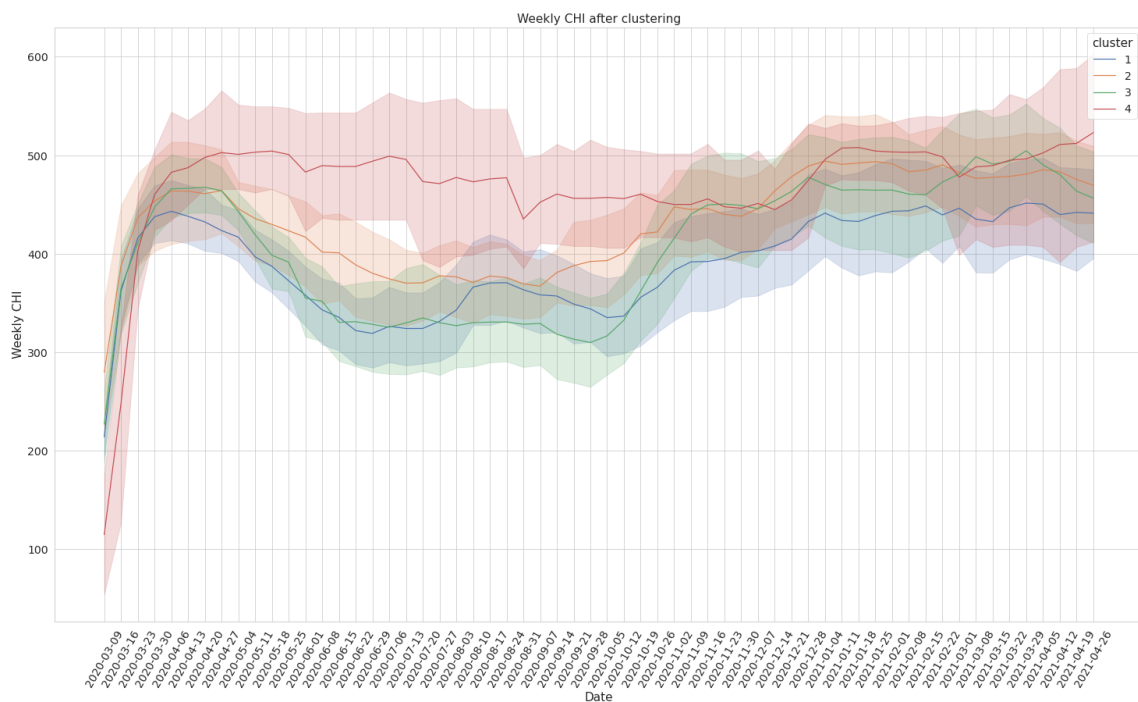


Σχήμα 4.3: Συσχέτιση θανάτων ανά εκατομμύριο - Containment and Health Index των χωρών κάθε συστάδας.

εμφάνιση του δεύτερου κύματος. Αντιθέτως, οι χώρες του cluster 4, στις οποίες οι επιπτώσεις της πανδημίας ήταν συνεχόμενες και δεν ακολούθησαν τη μορφή των κυμάτων, έλαβαν αυστηρά μέτρα τα οποία κράτησαν σχετικά σταθερά καθ' όλη τη διάρκεια της πανδημίας. Αξιοσημείωτο είναι ακόμη το γεγονός ότι οι χώρες των τριών πρώτων συστάδων διατήρησαν σε σχετικά υψηλό επίπεδο τα μέτρα κατά τη διάρκεια της ύφεσης της πανδημίας, καθώς επίσης και ότι με την επανισχυροποίηση τους κατά το δεύτερο κύμα, δεν ξεπέρασαν τα επίπεδα των μέτρων των χωρών του cluster 4, τα οποία παρέμειναν συνεχώς στο αυστηρό αυτό επίπεδο.

4.1.2 Επίδραση Μέτρων Προστασίας στη Μείωση των Θανάτων

Από το προηγούμενο κεφάλαιο, φαίνεται ότι οι θάνατοι από COVID-19 οδηγούν τις χώρες να λάβουν επιπλέον μέτρα προστασίας και να αυστηροποιήσουν τα υπάρχοντα. Είναι όμως σημαντική και η μελέτη του τρόπου με τον οποίο τα μέτρα αυτά επιδρούν στη μείωση των εβδομαδιαίων θανάτων. Όπως είναι αναμενόμενο, η αύξηση του Containment and Health Index οδηγεί στη μείωση των θανάτων στις επόμενες εβδομάδες. Η ταχύτητα με την οποία συμβαίνει η μείωση αυτή, δεν φαίνεται να μπορεί να προσδιοριστεί από τους παράγοντες που λαμβάνονται υπόψη στην παρούσα εργασία. Για παράδειγμα, στο σχήμα 4.5, φαίνονται οι εβδομαδιαίοι θάνατοι ανά εκατομμύριο σε συνδυασμό με το Containment and Health Index για τη Δανία και για τη Γερμανία. Στην περίπτωση της Δανίας, η μείωση των θανάτων φαίνεται να συμπίπτει



Σχήμα 4.4: Εξέλιξη του εβδομαδιαίου Containment and Health Index για κάθε συστάδα χωρών. Η γραμμή αντιπροσωπεύει τη μέση τιμή του δείκτη την αντίστοιχη εβδομάδα, ενώ το διάστημα αντιστοιχεί σε διάστημα εμπιστοσύνης 95% των τιμών των χωρών κάθε συστάδας.

με την κορύφωση της αυστηρότητας των μέτρων, ενώ στην περίπτωση της Γερμανίας, για να συμβεί αυτό πρέπει να μετατοπιστεί η χρονοσειρά των μέτρων τρεις εβδομάδες στο μέλλον. Αυτό σημαίνει ότι ενώ στην περίπτωση της Δανίας, τα μέτρα είχαν αρχίσει να δρουν προτού κορυφωθούν, στην περίπτωση της Γερμανίας χρειάστηκε να μεσολαβήσουν τρεις εβδομάδες προτού αυτά φέρουν αποτέλεσμα.

Ωστόσο, αξίζει να σημειωθεί το γεγονός ότι η παραμονή των μέτρων που κατάφεραν αρχικά να θέσουν σε ύφεση την έξαρση της πανδημίας, δεν συνεπάγεται πάντα τη διατήρηση των θανάτων σε χαμηλά νούμερα. Χαρακτηριστικό παράδειγμα αποτελούν η Ελλάδα και η Πολωνία, οι οποίες όπως φαίνεται στο σχήμα 4.6, ενώ κατάφεραν να αντιμετωπίσουν το δεύτερο κύμα της πανδημίας αυξάνοντας τα υπάρχοντα μέτρα, εμφάνισαν νέο κύμα θανάτων όσο διατηρούσαν το Containment and Health Index στο ίδιο επίπεδο. Το γεγονός αυτό, υποδηλώνει την ύπαρξη επιπλέον παραγόντων που οδηγούν ορισμένες χώρες σε τέτοια συμπεριφορά και δεν περιλαμβάνονται στην παρούσα εργασία.

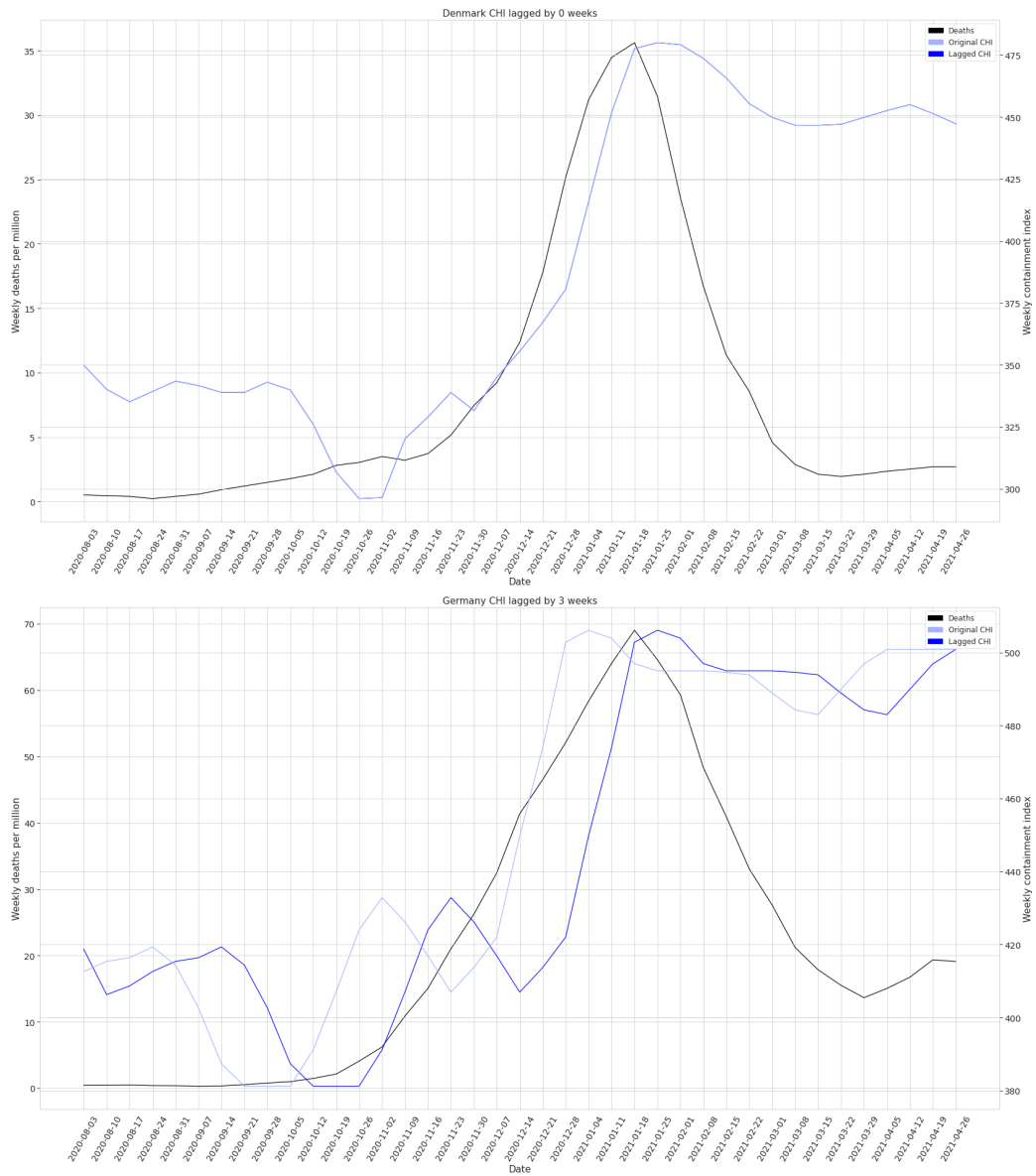
4.2 Επίδραση Μεταβλητών στις Προβλέψεις

Οι προβλέψεις που χρησιμοποιήθηκαν προέρχονται από το COVID-19 Forecast Hub, το οποίο αποτελεί το κέντρο προβλέψεων του Centers for Disease Control and Prevention (CDC)[5]. Το COVID-19 Forecast Hub συγκεντρώνει εβδομαδιαίες προβλέψεις από 32 ερευνητικά κέντρα και ιδιωτικές επιχειρήσεις σχετικά με προβλέψεις για τους θανάτους, τα κρούσματα και τις εισαγωγές στα νοσοκομεία που ευθύνονται στην COVID-19 για παράθυρα μέχρι και 4 εβδομάδες στο μέλλον. Οι προβλέψεις αυτές συγκεντρώνονται και μαζί με ένα συνδυαστικό μοντέλο, δημοσιεύονται στο αποθετήριο GitHub του Hub[26].

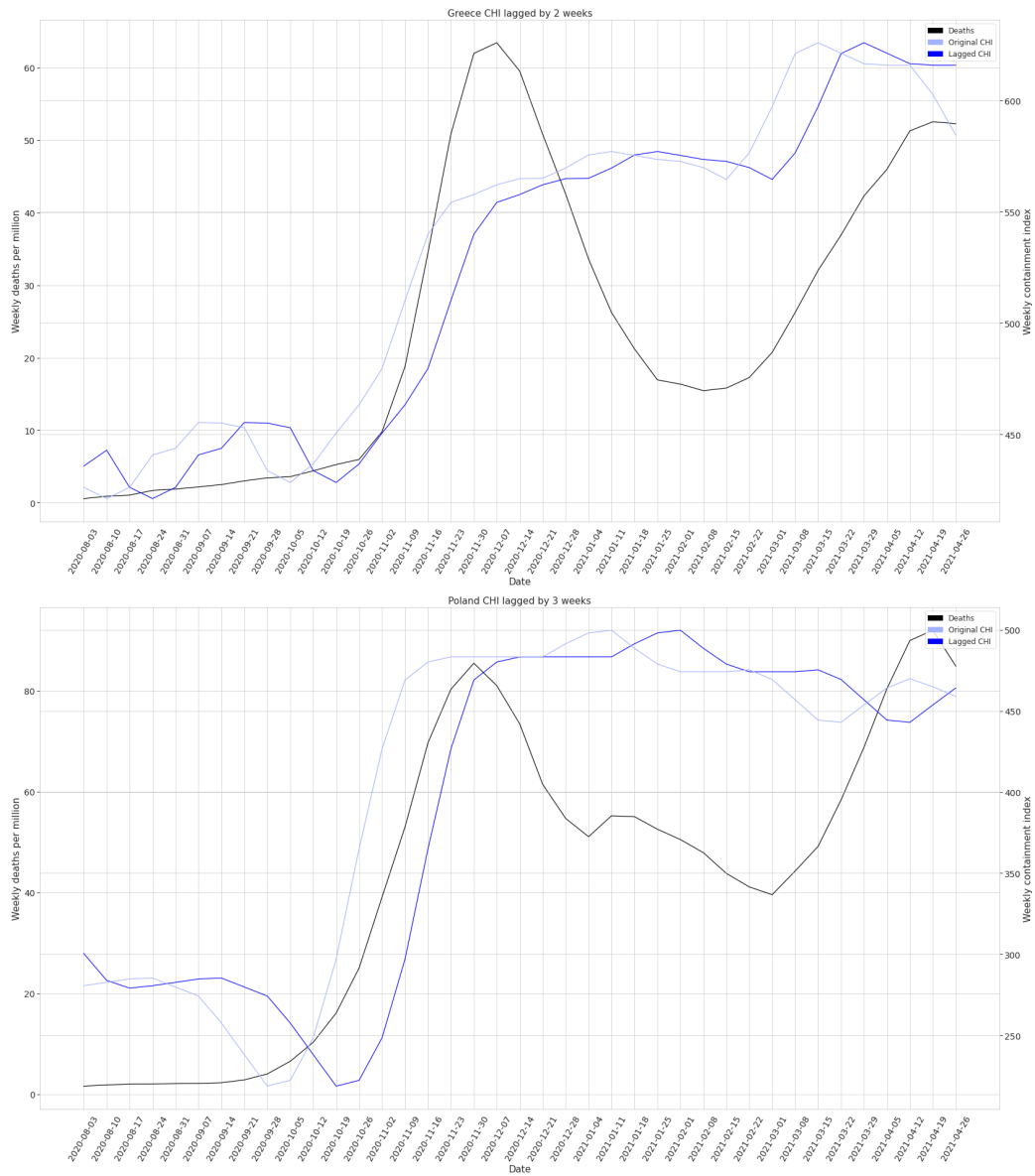
Η κάθε ερευνητική ομάδα χρησιμοποιεί ξεχωριστό συνδυασμό πηγών δεδομένων, μεθόδων και εκτιμήσεων της επίδρασης των διαφόρων μέτρων. Ενδεικτικό παράδειγμα αποτελεί το γεγονός ότι ορισμένες ομάδες θεωρούν ότι τα μέτρα προστασίας των Πολιτειών θα παραμείνουν ως έχουν για το διάστημα του παραθύρου πρόβλεψης, ενώ άλλες επιχειρούν να προβλέψουν το πώς αυτά θα αλλάξουν. Η ποικιλομορφία αυτή, καθιστά εξαιρετικά δύσκολη, αν όχι αδύνατη, την εξαγωγή συμπερασμάτων σχετικά με τους παράγοντες που επηρεάζουν τις μεμονωμένες προβλέψεις.

Για το λόγο αυτό, στην παρακάτω ανάλυση θα χρησιμοποιηθεί το συνδυαστικό μοντέλο του COVID-19 Forecast Hub, το οποίο συνδυάζει όλα τα παραπάνω παράγοντας μία τιμή για κάθε παράθυρο πρόβλεψης. Η συγκεντρωτική αυτή τιμή προκύπτει από έναν ισοβαρή μέσο όρο των κατάλληλων προβλέψεων. Κατάλληλες κρίνονται οι υποβολές που περιέχουν προβλέψεις για παράθυρα μίας, δύο, τριών και τεσσάρων εβδομάδων [12]. Επίσης, δεδομένου ότι οι θάνατοι ανά εκατομμύριο έχουν ήδη αναχθεί σε εβδομαδιαία βάση, για καλύτερη κατανόηση θα χρησιμοποιηθούν οι σημειακές προβλέψεις με παράθυρο πρόβλεψης μίας εβδομάδας.

Για την αξιολόγηση της ακρίβειας των προβλέψεων, επιλέχθηκε το συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα (Symmetric Mean Absolute Percentage Error ή



Σχήμα 4.5: Χρονοσειρές θανάτων ανά εκατομμύριο και CHI για τη Δανία και τη Γερμανία. Στην περίπτωση της Γερμανίας, η χρονοσειρά του CHI έχει μετατοπιστεί κατά τρεις εβδομάδες στο μέλλον προκειμένου να συμπέσει η κορύφωση των μέτρων με αυτή των θανάτων.



Σχήμα 4.6: Χρονοσειρές θανάτων ανά εκατομμύριο και CHI για την Ελλάδα και τη Πολωνία. Στην περίπτωση της Ελλάδας, η χρονοσειρά του CHI έχει μετατοπιστεί κατά δύο εβδομάδες στο μέλλον προκειμένου να συμπέσει η κορύφωση των μέτρων με αυτή των θανάτων, ενώ της Πολωνίας κατά τρεις.

sMAPE), το οποίο υπολογίζεται ως εξής:

$$sMAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - F_i}{\left(\frac{Y_i + F_i}{2}\right)} \right|$$

Ο λόγος επιλογής του sMAPE είναι ότι είναι ποσοστιαίο, επομένως δεν επηρεάζεται από το διαφορετικό επίπεδο των θανάτων στις διάφορες Πολιτείες, και σε αντίθεση με το μέσο απόλυτο ποσοστιαίο σφάλμα, μπορεί να οριστεί στην περίπτωση που οι εβδομαδιαίοι θάνατοι είναι μηδενικοί. Ωστόσο, παρά το όνομα του, το sMAPE δεν είναι απολύτως συμμετρικό, δίνοντας μικρότερη τιμή σφάλματος στην περίπτωση που οι προβλεπόμενοι θάνατοι είναι περισσότεροι από τους πραγματικούς. Στη συγκεκριμένη περίπτωση, το φαινόμενο αυτό δεν αποτελεί μεγάλο πρόβλημα, καθώς είναι προτιμότερη η έγκαιρη λήψη μέτρων για την αποφυγή περισσότερων θανάτων, από ό,τι η βεβιασμένη λήψη τους έπειτα από εμφάνιση περισσότερων θανάτων από τους προβλεπόμενους.

Από τις προβλέψεις του συνδυαστικού μοντέλου, γίνεται ξεκάθαρο ότι η μέση εκτίμηση των μοντέλων για τους εβδομαδιαίους θανάτους ανά εκατομμύριο για κάθε Πολιτεία εξαρτώνται σε πολύ μεγάλο βαθμό από τους θανάτους της προηγούμενης εβδομάδας. Χαρακτηριστικά, οι προβλέψεις του συνδυαστικού μοντέλου εμφανίζουν κατά μέσο όρο sMAPE ίσο με 0.3 σε σχέση με τις πραγματικές τιμές των θανάτων ανά εκατομμύριο, ενώ η μετατοπισμένη χρονοσειρά κατά μία εβδομάδα πίσω παρουσιάζει sMAPE κατά μέσο όρο ίσο με 0.19. Ακόμη, η μετατοπισμένη χρονοσειρά προβλέψεων παρουσιάζει με την χρονοσειρά των θανάτων συντελεστή συσχέτισης $r = 0.91$, σημαντικά υψηλότερο από το 0.82 της αρχικής χρονοσειράς.

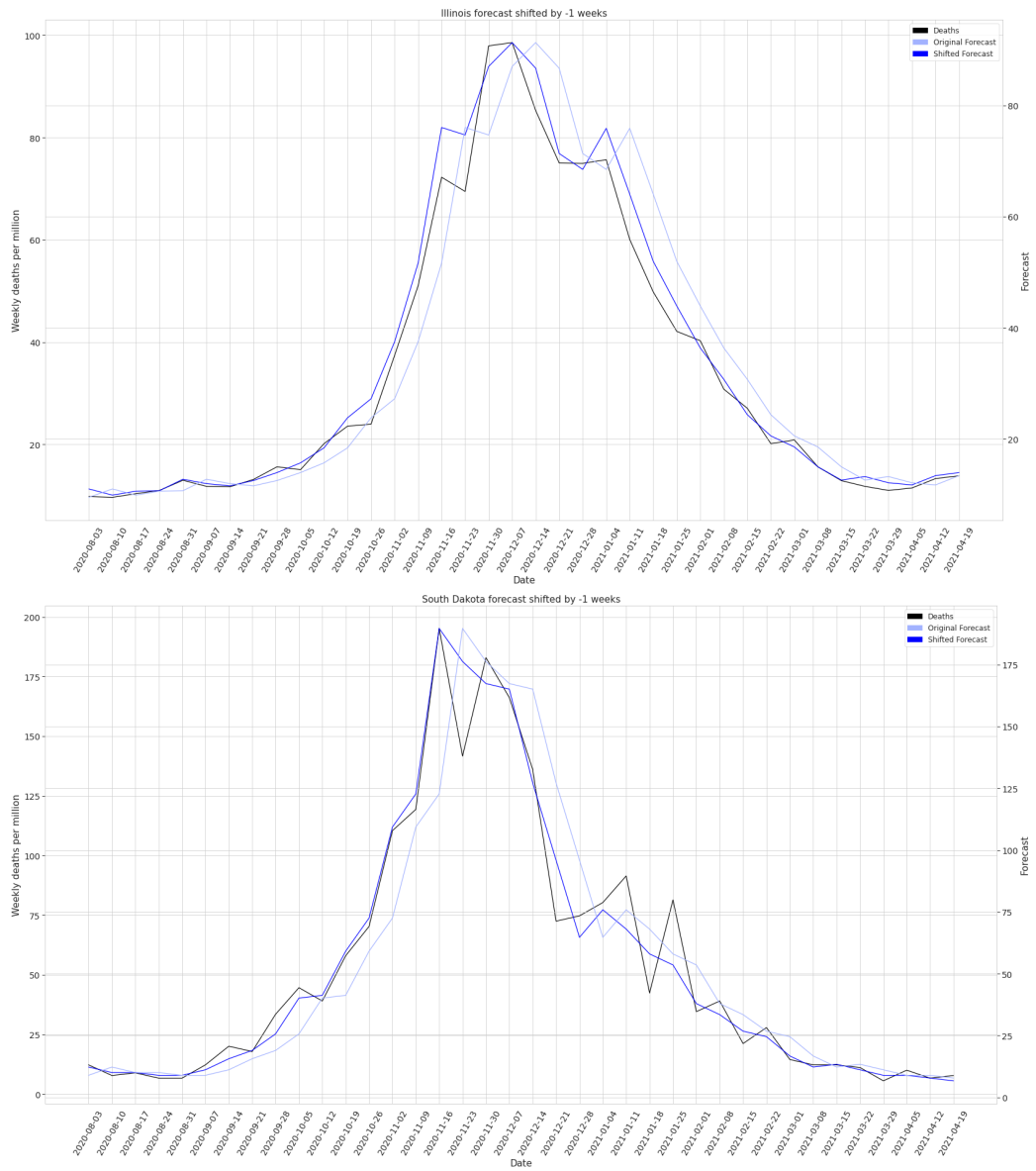
Για να γίνει καλύτερα αντιληπτή η παρατήρηση αυτή, στο σχήμα 4.7 παρουσιάζονται ενδεικτικά παραδείγματα Πολιτειών, ένα για κάθε συστάδα, στα οποία φαίνονται με αχνή μπλε γραμμή οι αρχικές προβλέψεις του συνδυαστικού μοντέλου και με έντονη μπλε γραμμή οι προβλέψεις αυτές μετατοπισμένες κατά μία εβδομάδα πίσω. Είναι εμφανές ότι η μετατοπισμένη χρονοσειρά μοιάζει πολύ με τη μαύρη γραμμή που αντιστοιχεί στους θανάτους ανά εκατομμύριο την εκάστοτε εβδομάδα. Στον πίνακα 4.1 παρουσιάζονται οι τιμές του συντελεστή συσχέτισης Pearson r και του στατιστικού σφάλματος sMAPE για τις εν λόγω Πολιτείες πριν και μετά τη μετατόπιση της χρονοσειράς των προβλέψεων.

Από τα παραπάνω φαίνεται, λοιπόν, ότι ο βασικός παράγοντας που επηρεάζει τις προβλέψεις των μοντέλων των ερευνητικών ομάδων είναι οι θάνατοι της προηγούμενης εβδομάδας. Το φαινόμενο αυτό είναι δικαιολογημένο, καθώς τα μεμονωμένα μοντέλα τα οποία είναι τα πιο ακριβή, δέχονται ως είσοδο απλά σύνολα δεδομένων. Χαρακτηριστικά, από τα πέντε ακριβέστερα, μόνο τα δύο μοντέλα χρησιμοποιούν δεδομένα εισόδου πέρα από τα χρούσματα και τους θανάτους των Πολιτειών [11].

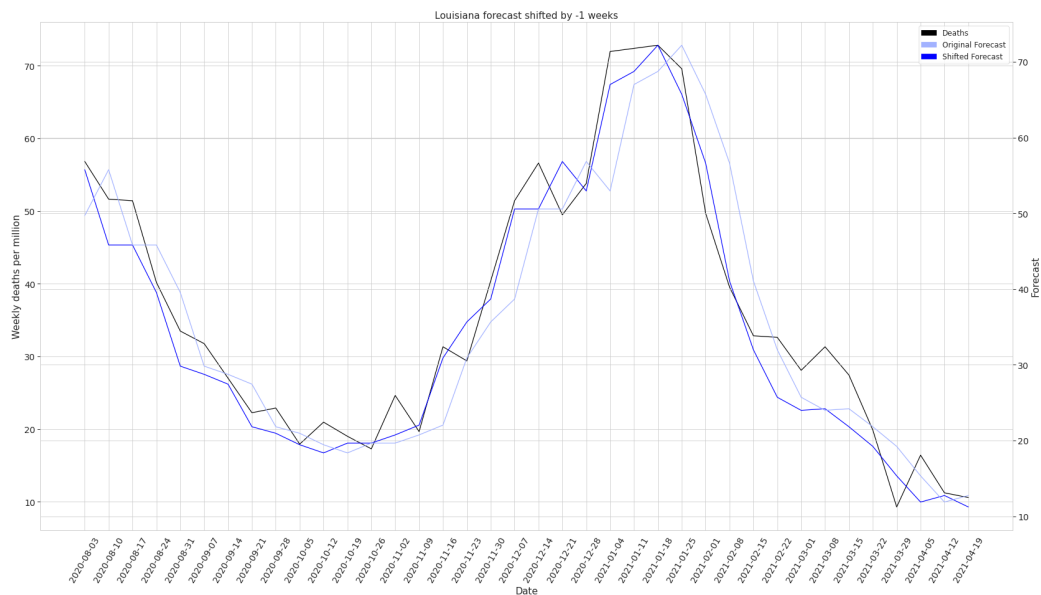
Στο σχήμα 4.8 παρουσιάζονται τα διαγράμματα σκέδασης του sMAPE των προβλέψεων των θανάτων για κάθε Πολιτεία με τα υπόλοιπα χαρακτηριστικά του συνόλου δεδομένων. Από αυτά, ξεχωρίζουν οι περιπτώσεις του πληθυσμού και της μέσης θερμοκρασίας, στις οποίες ο συντελεστής συσχέτισης r είναι ίσος με -0.44 και στις δύο. Το γεγονός ότι τα μοντέλα τείνουν να προβλέπουν με μεγαλύτερη ακρίβεια τους



Σχήμα 4.7: Μετατοπισμένες χρονοσειρές προβλέψεων μία εβδομάδα πίσω.



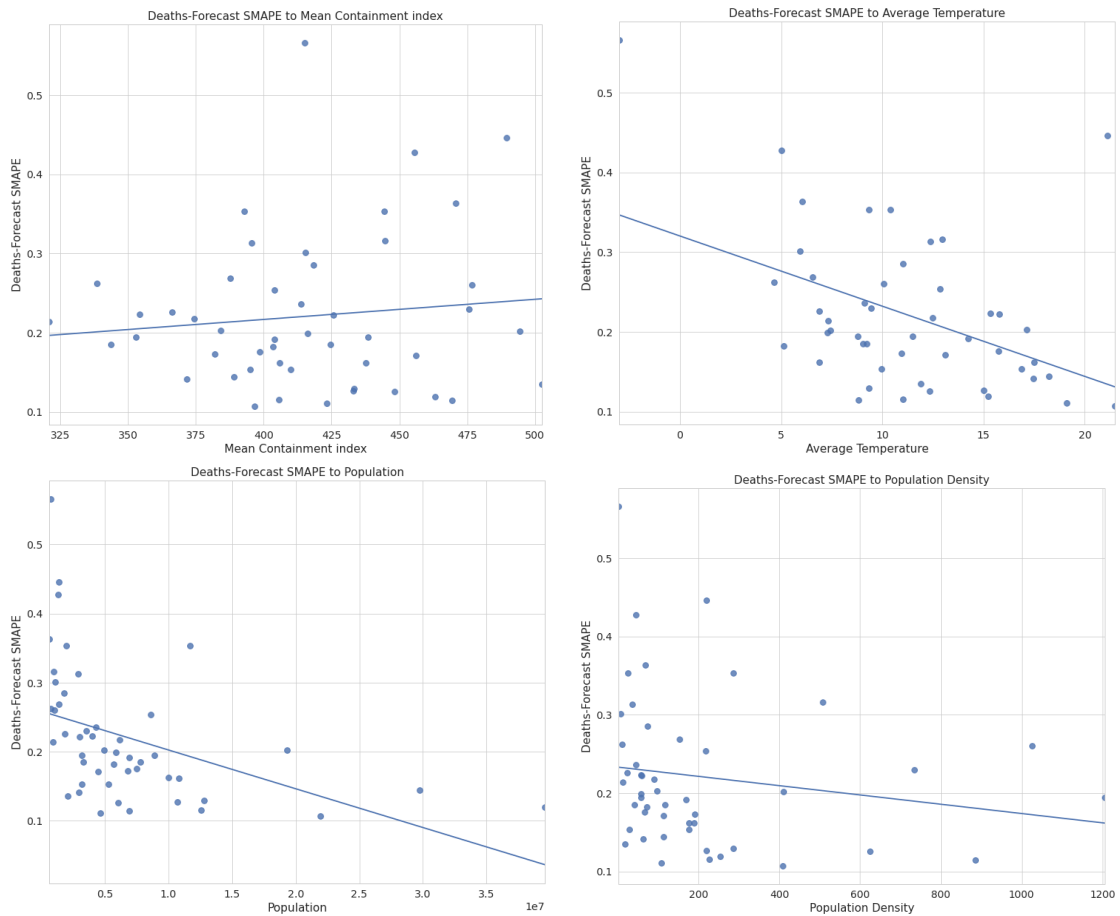
Σχήμα 4.7: Μετατοπισμένες χρονοσειρές προβλέψεων μία εβδομάδα πίσω (συνέχεια 1).



Σχήμα 4.7: Μετατοπισμένες χρονοσειρές προβλέψεων μία εβδομάδα πίσω (συνέχεια 2).

Πολιτεία	r (πριν)	r (μετά)	$sMAPE$ (πριν)	$sMAPE$ (μετά)
Massachusetts (1)	0.96	0.98	0.17	0.09
Texas (2)	0.88	0.98	0.16	0.07
Illinois (3)	0.96	0.99	0.14	0.05
South Dakota (4)	0.92	0.98	0.28	0.14
Louisiana (5)	0.93	0.98	0.15	0.11

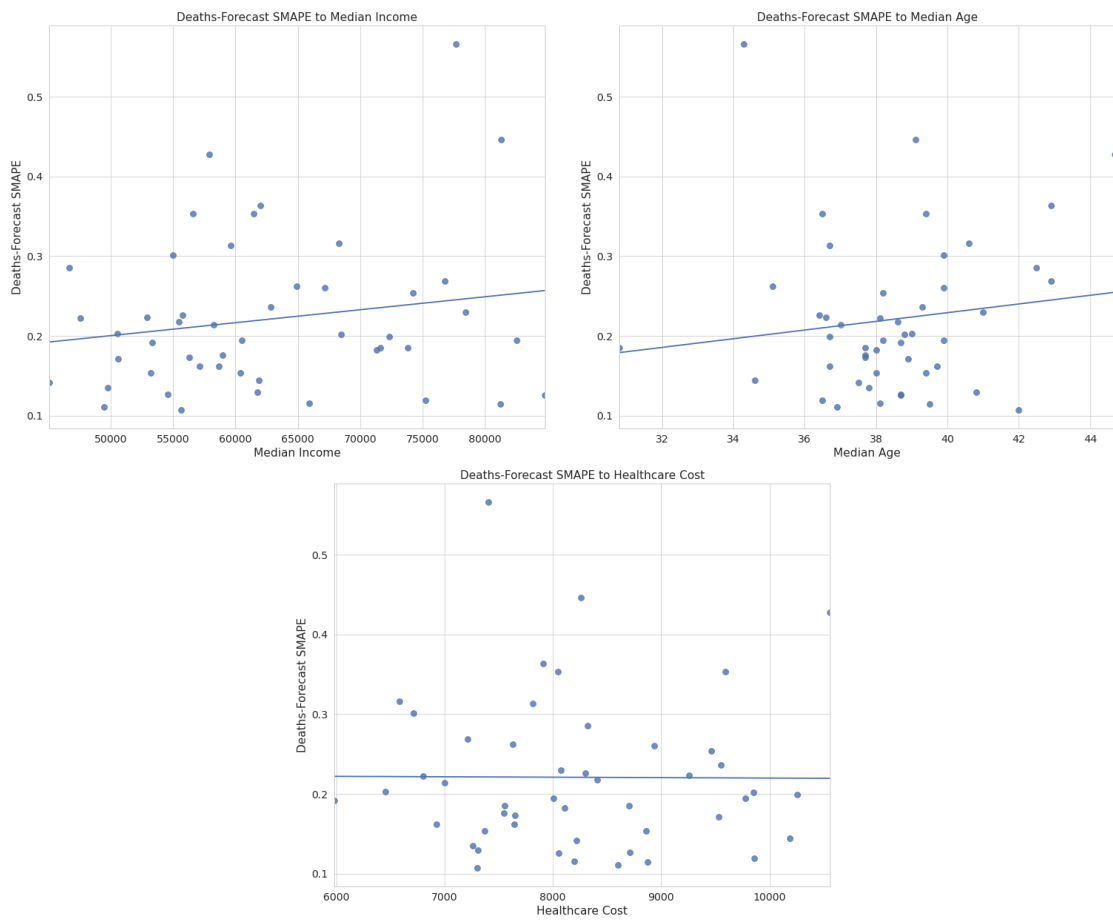
Πίνακας 4.1: Συσχέτιση και $sMAPE$ προβλέψεων θανάτων πριν και μετά τη μετατόπιση τους μία εβδομάδα πίσω.



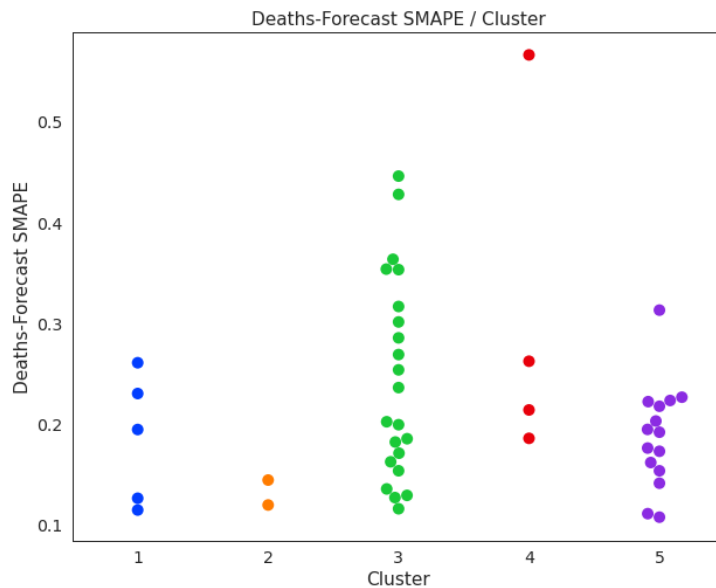
Σχήμα 4.8: Διαγράμματα σχέδωσης sMAPE των προβλέψεων με τα υπόλοιπα χαρακτηριστικά.

θανάτους για τις Πολιτείες με το μεγαλύτερο πληθυσμό, μπορεί να εξηγηθεί καθώς οι Πολιτείες αυτές περιέχουν μεγάλα αστικά κέντρα, στα οποία χρειάζονται ακριβείς χειρισμοί για την προστασία από την ραγδαία εξάπλωση του κορωνοϊού. Επομένως, ορισμένες ερευνητικές ομάδες ενδέχεται να δίνουν μεγαλύτερο βάρος στις ακριβείς προβλέψεις για τις συγκεκριμένες Πολιτείες στα μοντέλα τους, ώστε να προλάβουν τυχούσες εξάρσεις.

Αναλύοντας την κατανομή των Πολιτειών στο sMAPE των προβλέψεων των θανάτων λόγω COVID-19 σε επίπεδο συστάδων στο σχήμα 4.9, προκύπτει ότι σε γενικές γραμμές οι προβλέψεις παρουσίασαν παρόμοιες επιδόσεις για τις Πολιτείες όλων των συστάδων. Μικρή εξαίρεση αποτελούν ορισμένες Πολιτείες του cluster 3, το οποίο υπενθυμίζεται ότι περιέχει σχεδόν τις μισές Πολιτείες, καθώς επίσης και τις Πολιτείες με τον μικρότερο πληθυσμό, όπου τείνουν να παρουσιάζουν λιγότερο ακριβείς προβλέψεις.



Σχήμα 4.8: Διαγράμματα σχέδασης sMAPE των προβλέψεων με τα υπόλοιπα χαρακτηριστικά (συνέχεια).

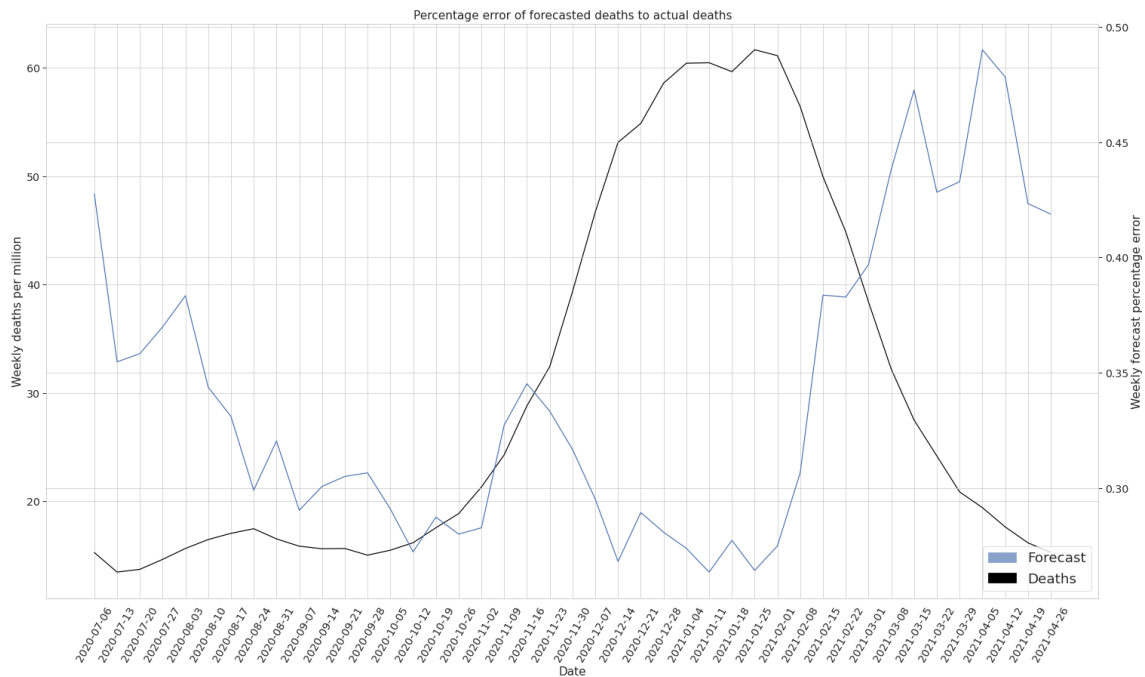


Σχήμα 4.9: Κατανομή Πολιτειών κάθε συστάδας στο sMAPE των προβλέψεων των θανάτων.

4.3 Εξέλιξη Ακρίβειας Προβλέψεων

Στο κεφάλαιο αυτό μελετάται η εξέλιξη της ακρίβειας των προβλέψεων με την πάροδο του χρόνου. Για κάθε Πολιτεία, υπολογίζεται το ποσοστιαίο σφάλμα της εβδομαδιαίας πρόβλεψης, σύμφωνα με τον τύπο $\frac{\text{actual deaths} - \text{forecast}}{\text{actual deaths}}$. Στο σχήμα 4.10 φαίνεται η μέση τιμή των ποσοστιαίων σφαλμάτων αυτών για όλες τις Πολιτείες (μπλε γραμμή). Στο ίδιο διάγραμμα φαίνονται και οι μέσοι εβδομαδιαίοι θάνατοι ανά εκατομμύριο λόγω COVID-19 στις Πολιτείες, για την χρονική τοποθέτηση των προβλέψεων στην ιστορία της πανδημίας. Και οι δύο γραμμές του διαγράμματος έχουν περάσει από επεξεργασία κινητού μέσου όρου 5 εβδομάδων, προκειμένου να αφαιρεθεί το μεγαλύτερο μέρος του θορύβου και να φανεί καθαρότερα η πορεία τους.

Από το διάγραμμα, φαίνεται ότι το ποσοστιαίο σφάλμα των προβλέψεων από 43% αρχικά μειώνεται και φτάνει έως και 28%. Όταν όμως ξεκινάει το δεύτερο κύμα της πανδημίας, τότε τα σφάλματα αυξάνονται για δύο εβδομάδες και μετά μειώνονται όταν το κύμα κορυφώνεται, με ελάχιστη τιμή το 26%. Στη συνέχεια, όταν υποχωρεί το δεύτερο κύμα της πανδημίας, η ακρίβεια των προβλέψεων αυξάνεται ραγδαία, φτάνοντας το 49%. Η αστάθεια στην ακρίβεια των προβλέψεων κάθε φορά που αλλάζει η πορεία της πανδημίας, συμφωνεί με το συμπέρασμα του προηγούμενου κεφαλαίου, ότι ο κυριότερος παράγοντας που επηρεάζει τα μοντέλα προβλέψεων είναι οι τιμές των θανάτων για τις προηγούμενες εβδομάδες. Επομένως, μέχρι να αντιληφθούν τα μοντέλα μία αλλαγή στην τάση της χρονοσειράς των θανάτων, γίνονται μερικές λανθασμένες προβλέψεις. Ωστόσο, ενδιαφέρον παρουσιάζει το γεγονός ότι τα μοντέλα προσαρμόστηκαν πολύ πιο γρήγορα στην αύξηση των θανάτων, παρά στη μείωση τους στη



Σχήμα 4.10: Εξέλιξη ποσοστιαίου σφάλματος προβλέψεων με την πάροδο του χρόνου.

συνέχεια. Από τις 43 εβδομάδες προβλέψεων που εξετάστηκαν, οι 32 (74.4%) πρόβλεψαν περισσότερους θανάτους κατά μέσο όρο, από αυτούς που συνέβησαν στην πραγματικότητα. Το γεγονός αυτό, πιθανότατα οφείλεται στο ότι τα μοντέλα που χρησιμοποιήθηκαν για την παραγωγή προβλέψεων είναι απαισιόδοξα. Όταν οι προβλέψεις έχουν να κάνουν με ανθρώπινες απώλειες και βάσει αυτών λαμβάνονται αποφάσεις που επηρεάζουν την πορεία της πανδημίας, τότε από ο,τι φαίνεται, επιλέγεται η ασφάλεια έναντι της ναρτίτερης δυνατής χαλάρωσης των μέτρων και η προστασία από ένα απαισιόδοξο σενάριο.

Κεφάλαιο 5

Συμπεράσματα και μελλοντικές επεκτάσεις

Στο τελευταίο κεφάλαιο της εργασίας γίνεται μία σύνοψη των συμπερασμάτων που εξήχθησαν από την ανάλυση που προηγήθηκε.

Σχετικά με τα μέτρα προστασίας και στήριξης από την πανδημία, επιβεβαιώθηκε το λογικό, ότι όσο μεγαλύτερη είναι η έξαρση των θανάτων λόγω COVID-19, οι χώρες λαμβάνουν πιο αυστηρά μέτρα για τον περιορισμό της, καθώς επίσης και ότι μετά από λίγες εβδομάδες, τα μέτρα αυτά αντικατοπτρίζονται στη μείωση της αρχικής έξαρσης. Ωστόσο, ορισμένες εξ αυτών παρουσίασαν έκ νέου αύξηση των θανάτων παρά τη διατήρηση των πιο αυστηρών μέτρων προστασίας.

Όσον αφορά τους παράγοντες που επιδρούν στην εξέλιξη της πανδημίας, επαληθεύτηκαν τα πορίσματα σχετικών ερευνών, ότι τα μεγάλα αστικά κέντρα αποτελούν εστία υπερμετάδοσης του κορωνοϊού, γεγονός που κατέστη επιζήμιο στην αρχή της πανδημίας, όταν οι κυβερνήσεις και ο πληθυσμός δεν είχαν ακόμη προσαρμοστεί στα νέα δεδομένα. Παρόμοια επίδραση φάνηκε ότι έχουν και οι τουριστικές αφίξεις, όπως ήταν αναμενόμενο, καθώς η συμπεριφορά των τουριστών ευνοεί τη διασπορά του κορωνοϊού και η μεταφορά πληθυσμών αυξάνει τις πιθανότητες γέννησης νέων μεταλλάξεων του. Ακόμη, η μελέτη των Ηνωμένων Πολιτειών της Αμερικής επιβεβαίωσε το πόρισμα ότι οι θερμές και υψηλές σε υγρασία συνθήκες αυξάνουν το χρόνο ζωής του SARS-CoV-2 σε επιφάνειες, διευκολύνοντας συνεπώς τη διασπορά του.

Παρουσιάστηκαν όμως και ενδιαφέροντα συμπεράσματα, από την άποψη ότι δεν βρέθηκε στη βιβλιογραφία κάτι που να τα επαληθεύει ή να τα διαψεύδει. Συγκεκριμένα, φάνηκε ότι μεταξύ χωρών με παρόμοια δημογραφικά χαρακτηριστικά και παρόμοια πολιτική μέτρων προστασίας, οι χώρες με καλύτερο βιοτικό επίπεδο κατάφεραν να κατευνάσουν την έξαρση της πανδημίας πολύ πιο αποτελεσματικά. Το γεγονός αυτό, ενισχύει την πεποίθηση ότι η ύπαρξη καλύτερου μορφωτικού επιπέδου, οδηγεί στην κατανόηση της χρισιμότητας της πανδημίας από το γενικότερο πληθυσμό και επομένως στην καλύτερη τήρηση των μέτρων προστασίας. Από την άλλη πλευρά, φάνηκε ότι οι χώρες χαμηλού βιοτικού επιπέδου δεν κατάφεραν να θέσουν την έξαρση της

COVID-19 υπό έλεγχο καθ' όλη τη διάρκεια της πανδημίας. Το γεγονός αυτό, πιθανόν οφείλεται στην ύπαρξη υποστελεχωμένων και δυσπρόσιτων εθνικών συστημάτων υγείας, με περιορισμένη δυναμικότητα στην ίαση των ασθενών.

Επιπλέον, φάνηκε ότι τόσο ο τρόπος αντιμετώπισης της πανδημίας, όσο και η αποτελεσματικότητα αυτού ενέχουν και γεωγραφικό παράγοντα. Πέρα από τη μέση ετήσια θερμοκρασία, η οποία δεν φαίνεται να επαρκεί, δεν βρέθηκαν πηγές δεδομένων με χαρακτηριστικά που να περιγράφουν αναλυτικότερα το κλίμα ή τα κοινωνικά χαρακτηριστικά και τις συμπεριφορές των κατοίκων των αντίστοιχων γεωγραφικών περιοχών. Επομένως δεν μπόρεσε να επαληθευτεί το κατά πόσο βασίζεται ο γεωγραφικός παράγοντας αυτός σε τέτοιου είδους χαρακτηριστικά.

Στο κομμάτι των προβλέψεων των εβδομαδιαίων θανάτων για τις Ηνωμένες Πολιτείες της Αμερικής, έγινε ξεκάθαρο ότι ο βασικός παράγοντας που επηρεάζει τις προβλέψεις των μοντέλων των ερευνητικών ομάδων, είναι οι θάνατοι της προηγούμενης εβδομάδας. Έτσι δικαιολογείται το γεγονός ότι τα πιο ακριβή από τα μοντέλα που παρακολουθούνται δέχονται ως είσοδο απλά σύνολα δεδομένων. Παράλληλα, φάνηκε ότι κάποια μοντέλα δίνουν μεγαλύτερο βάρος στις ακριβείς προβλέψεις για Πολιτείες με μεγάλο πληθυσμό, πιθανόν γιατί σε αυτά είναι επικίνδυνο να υπάρξει μεγαλύτερη διασπορά της COVID-19. Ως αποτέλεσμα, οι Πολιτείες με τον μικρότερο πληθυσμό παρουσιάζουν λιγότερο ακριβείς προβλέψεις. Ένα τελευταίο σημαντικό εύρημα, αποτελεί το γεγονός ότι τα μοντέλα που χρησιμοποιήθηκαν για την παραγωγή προβλέψεων είναι απαισιόδοξα. Αυτό είναι λογικό, καθώς ο σκοπός των προβλέψεων είναι η προειδοποίηση της κάθε Πολιτείας για την εκτιμώμενη έκταση της πανδημίας την επόμενη εβδομάδα, με σκοπό αυτή να προετοιμαστεί κατάλληλα. Έτσι, φαίνεται επόμενο τα μοντέλα να προσπαθούν να καλύψουν το κακό σενάριο κάθε εβδομάδα, γεγονός που πετυχαίνουν κρίνοντας από το μεγάλο πλήθος αρνητικών σφαλμάτων (74.4%). Η χρήση του όρου απαισιόδοξια, σε αντίθεση με τη συνηθισμένη ερμηνεία στις προβλέψεις, αναφέρεται στην πρόβλεψη μεγαλύτερης τιμής από την πραγματική, λόγω του ότι η προβλεπόμενη τιμή είναι απώλειες ανθρώπινων ζώων.

Όσον αφορά τις μελλοντικές επεκτάσεις της παρούσας εργασίας, οι περισσότερες αποτελούν τρόπους εμπλουτισμού των παραγόντων που απαρτίζουν το σύνολο δεδομένων. Συγκεκριμένα, ιδιαίτερο ενδιαφέρον παρουσιάζει η ανάλυση των επιμέρους μέτρων προστασίας και στήριξης από την COVID-19, προκειμένου να μελετηθεί η συμβολή του καθενός ξεχωριστά και να μπορέσει να γίνει ακόμα πιο λεπτομερής συσταδοποίηση των χωρών και Πολιτειών. Το σύνολο δεδομένων μπορεί να διανθιστεί περαιτέρω με κλιματικά στοιχεία και χαρακτηριστικά κοινωνικής φύσεως, όπως ήθη και έθιμα, εμπιστοσύνη στην κυβέρνηση, εξωστρέφεια και άλλα πολλά. Παράγοντες σαν κι αυτούς, μπορούν να επιβεβαιώσουν ή να διαψεύσουν ορισμένα από τα συμπεράσματα που προέκυψαν από τη μελέτη. Ωστόσο, η μεγάλη δυσκολία συλλογής και ποσοτικοποίησης τους, οδήγησε στην παράλειψη τους από την τρέχουσα μελέτη. Όσο για τα μέτρα προστασίας, κρίθηκε ότι η περαιτέρω ανάλυση τους, θα οδηγούσε στο να μονοπωλήσουν έναντι άλλων πτυχών της εργασίας. Τέλος, ενδιαφέρον παρουσιάζει η μεμονωμένη μελέτη χωρών οι οποίες εμφάνισαν πολύ θετική ή αρνητική εξέλιξη της πανδημίας, σε σχέση με αυτή που υποδηλώνουν τα χαρακτηριστικά τους και η συστάδα

στην οποία ανήκουν.

Βιβλιογραφία

- [1] A. Desvars-Larrive, E. Dervic, N. Haug et al: *A structured open dataset of government interventions in response to covid-19*. 2020. <https://doi.org/10.1038/s41597-020-00609-9>.
- [2] Berkley Earth: *Global Warming Data Overview*. <http://berkeleyearth.org/data/>.
- [3] Blavatnik School of Government, University of Oxford: *Oxford COVID-19 Government Response Tracker*. <https://github.com/OxCGRT/covid-policy-tracker>.
- [4] C. Cheng, J. Barceló, A.S. Hartnett, et al: *Covid-19 government response event dataset (coronamet v.1.0)*. 2020. <https://doi.org/10.1038/s41562-020-0909-7>.
- [5] Centers for Disease Control and Prevention: *COVID-19 Mathematical Modeling*. <https://www.cdc.gov/coronavirus/2019-ncov/science/forecasting/mathematical-modeling.html>.
- [6] Centers for Disease Control and Prevention: *Interim Clinical Guidance for Management of Patients with Confirmed Coronavirus Disease (COVID-19)*. <https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html>.
- [7] Chan, K.H., J.S. Malik Peiris, L.L. M. Poon S.Y. Lam, K.Y. Yuen, και W.H. Seto: *The effects of temperature and relative humidity on the viability of the sars coronavirus*. 2011. <https://doi.org/10.1155/2011/734690>.
- [8] Cucinotta D, Vanelli M: *Who declares covid-19 a pandemic*. 2020. <https://doi.org/10.23750/abm.v9i1i1.9397>.
- [9] Eduardo Levy Yeyati, Federico Filippini: *Social and economic impact of COVID-19*. <https://www.brookings.edu/research/social-and-economic-impact-of-covid-19/>.

- [10] Eran Bendavid, Christopher Oh, Jay Bhattacharya, John P.A. Ioannidis: *Assessing mandatory stay-at-home and business closure effects on the spread of covid-19*. 2021. <https://doi.org/10.1111/ECI.13484>.
- [11] Estee Y Cramer, Evan L. Ray, Velma K. Lopez et al: *Evaluation of individual and ensemble probabilistic forecasts of covid-19 mortality in the us*. 2021. <https://www.medrxiv.org/content/10.1101/2021.02.03.21250974v2>.
- [12] Evan L Ray, Nutch Wattanachit, Jarad Niemi et al: *Ensemble forecasts of coronavirus disease 2019 (covid-19) in the u.s.* 2020. <https://doi.org/10.1101/2020.08.19.20177493>.
- [13] F. Caramelo, N. Ferreira, B. Oliveiros: *Estimation of risk factors for covid-19 mortality - preliminary results*. 2020. <https://doi.org/10.1101/2020.02.24.20027268>.
- [14] Fotios Petropoulos, Spyros Makridakis: *Forecasting the novel coronavirus covid-19*. 2020. <https://doi.org/10.1371/journal.pone.0231236>.
- [15] Gallup: *Worldwide, Median Household Income About \$10,000*. <https://news.gallup.com/poll/166211/worldwide-median-household-income-000.aspx>.
- [16] Galton, Francis: *Hereditary genius: An inquiry into its laws and consequences*. 1869. <https://doi.org/10.1037/13474-000>.
- [17] Hotelling, Harold: *Relations between two sets of variates*. 1936. <https://doi.org/10.2307/2333955>.
- [18] International Labour Organization: *COVID-19 and the world of work*. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/documents/briefingnote/wcms_767028.pdf.
- [19] Kawther Aabed, Maha M.A. Lashin: *An analytical study of the factors that influence covid-19 spread*. 2021. <https://doi.org/10.1016/j.sjbs.2020.11.067>.
- [20] KFF (Kaiser Family Foundation): *Health Care Expenditures per Capita by State of Residence*. <https://www.kff.org/other/state-indicator/health-spending-per-capita>.
- [21] Konstantinos Kontoangelos, Marina Economou, Charalambos Papageorgiou: *Mental health effects of covid-19 pandemia: A review of clinical and psychological traits*. 2020. <https://doi.org/10.30773/pi.2020.0161>.
- [22] Nikolopoulos, Konstantinos, Sushil Punia, Andreas Schäfers, Christos Tsinopoulos, και Chrysovalantis Vasilakis: *Forecasting and planning during a pandemic: Covid-19 growth rates, supply chain disruptions, and governmental decisions*. 2020. <https://doi.org/10.1016/j.ejor.2020.08.001>.

- [23] NOAA National Climatic Data Center: *Average Annual Temperature for Each US State*. <https://www.currentresults.com/Weather/US/average-annual-state-temperatures.php>.
- [24] Pearson, Karl: *Liii. on lines and planes of closest fit to systems of points in space*. 1901. <https://doi.org/10.1080/14786440109462720>.
- [25] Peterson-KFF Health System Tracker: *How does health spending in the U.S. compare to other countries?* <https://www.healthsystemtracker.org/chart-collection/health-spending-u-s-compare-countries/#item-start/>.
- [26] Reich Lab, University of Massachusetts Amherst: *Covid-19 forecast hub*. 2005. <https://doi.org/10.5281/zenodo.4305938>.
- [27] T., Hale, Angrist N., and Goldszmidt R. et al: *A global panel database of pandemic policies (oxford covid-19 government response tracker)*. 2021. <https://doi.org/10.1038/s41562-021-01079-8>.
- [28] Tahir Mumtaz Awan, Faheem Aslam: *Prediction of daily covid-19 cases in european countries using automaticarima model*. 2020. <https://doi.org/10.4081/jphr.2020.1765>.
- [29] United Nations: *World Population Prospects 2019*. <https://population.un.org/wpp/Download/Standard/Population/>.
- [30] United Nations World Tourism Organization: *International arrivals by country*. <https://ourworldindata.org/tourism#international-arrivals-by-country>.
- [31] United States Census Bureau: *2019 Median Household Income in the United States*. <https://www.census.gov/library/visualizations/interactive/2019-median-household-income.html>.
- [32] United States Census Bureau: *American Community Survey (ACS)*. <https://worldpopulationreview.com/state-rankings/median-age-by-state>.
- [33] United States Census Bureau: *US Census 2017 State Estimates*. <https://worldpopulationreview.com/states>.
- [34] World Health Organization: *Coronavirus Disease (COVID-19)*. <https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19>.
- [35] Yinliang Diao, Sachiko Kodera, Daisuke Anzai, Jose Gomez-Tames, Essam A. Rashed, Akimasa Hirata: *Influence of population density, temperature, and absolute humidity on spread and decay durations of covid-19: A comparative study of scenarios in china, england, germany, and japan*. 2021. <https://doi.org/10.1016/j.onehlt.2020.100203>.

- [36] Βλαχαβάς, Ι., Π. Κεφαλάς, Ν. Βασιλειάδης, Φ. Κόκκορας, και Η. Σακελλαρίου: *Τεχνητή Νοημοσύνη*. σελίδα 530–539. Εκδόσεις Πανεπιστημίου Μακεδονίας, 2020, ISBN 978-618-5196-44-8.