Ερώτημα 1

Προ-επεξεργασία Δεδομένων

Για την προ-επεξεργασία των δεδομένων αρχικά κάναμε import τις βιβλιοθήκες που θα χρειαστούμε

A picture containing font, text, screenshot, graphics

Description automatically generated

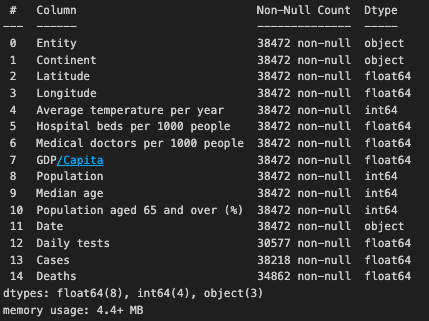
Την pandas για την επεξεργασία των .csv αρχείων ως dataframes, την matplotlib.pyplot για την σχεδίαση των γραφικών παραστάσεων την seaborn για τον υπολογισμό του Correlation Heatmap. Έπειτα φορτώνουμε τα δεδομένα και τυπώνουμε ενδεικτικά τα πρώτα rows του dataset όπως και πληροφορίες για το dataset (columns,mean, std, min, max, 25%, 50%, 75%, non-null count και data type of each για κάθε στήλη )

A picture containing text, font, screenshot

Description automatically generated







Παρατηρούμε ότι ορισμένες στήλες έχουν null τιμές

A screenshot of a computer

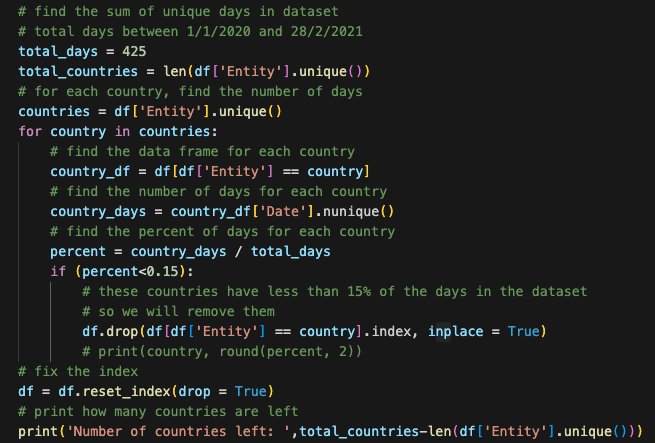
Description automatically generated

Επομένως, αφαιρούμε τις στήλες που έχουν null καθώς και τις αρνητικές τιμές στη στήλη Daily Tests.

A picture containing text, screenshot, font

Description automatically generated

Έπειτα, υπολογίζουμε το πλήθος των δεδομένων που έχουμε ανά χώρα και αφαιρούμε όποια χώρα έχει λιγότερο από το 5% από το συνολικό πλήθος ημερών (1/1/2020-28/2/2021).



Προκύπτει ότι 5 χώρες έχουν λιγότερο από το 5%.

Ακόμη, θεωρούμε ότι πριν την εμφάνιση του πρώτου θετικού τεστ και θανάτου ανά χώρα το πλήθος θετικών τεστ, και το πλήθος θανάτων είναι μηδέν, επομένως αντικαθιστούμε τις null τιμές με μηδενικά στις θέσεις αυτές. Μετά από και αυτήν την αλλαγή παρατηρούμε ότι δεν υπάρχουν άλλες Null τιμές στο dataset.

A picture containing text, screenshot, font

Description automatically generated

A screenshot of a computer

Description automatically generated

Έπειτα, μέσω των στηλών Cases και Deaths δημιουργούμε και υπολογίζουμε δύο στήλες για τα Daily cases και Daily deaths. Ωστόσο, παρατηρήσαμε ότι υπήρχαν αρνητικές τιμές στα Daily cases / Daily deaths πράγμα αδύνατο. Αυτή η παρανόηση προήρθε από το γεγονός ότι υπήρχαν χώρες που ο αριθμός των Cases/Deaths μειωνόταν, το οποία πρακτικά είναι αδύνατον. Για τον λόγο αυτό πριν υπολογίσουμε τις νέες στήλες μετασχηματίσαμε τις στήλες Cases και Deaths, έτσι ώστε οι μέρες που είχαν μικρότερο πλήθος Cases/Deaths (σε σχέση με την προηγούμενη μέρα) να πάρουν τον αριθμό Cases/Deaths της προηγούμενης μέρας.

A picture containing text, screenshot, font

Description automatically generated

A picture containing text, screenshot, font

Description automatically generated

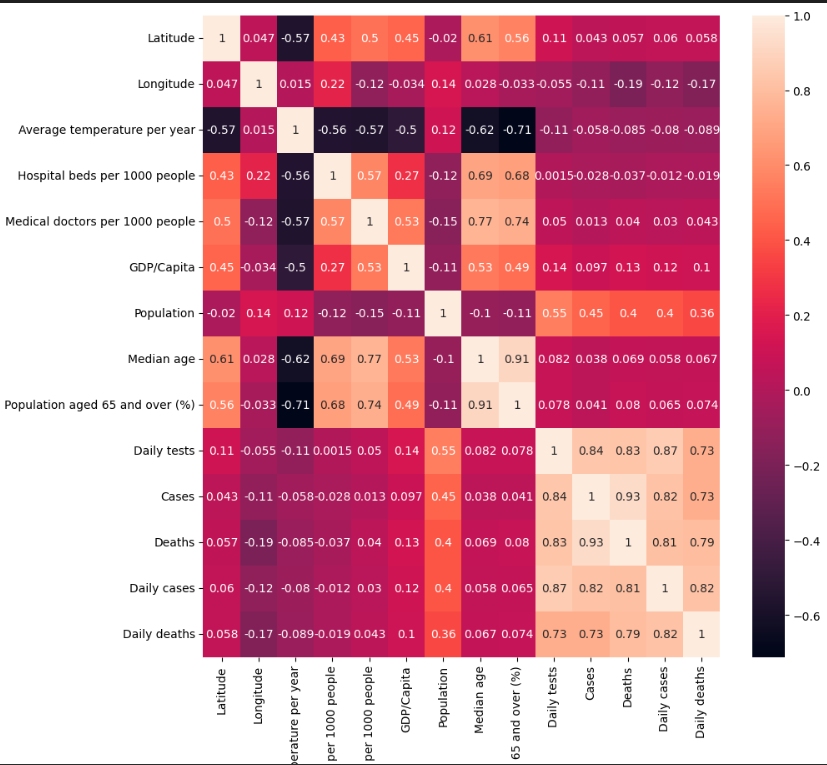
Επίσης, μετά τον υπολογισμό παρατηρήσαμε ότι υπάρχουν μέρες που ο αριθμός Daily Cases ήταν μεγαλύτερος από τον αριθμό Daily tests. Θεωρήσαμε ότι ήταν καλύτερο να αφαιρέσουμε αυτές τις γραμμές.

Ακόμη, σε κάθε γραμμή προσθέσαμε μια στήλη που προσδιορίζει την εποχή (χειμώνας, άνοιξη, καλοκαίρι και φθινόπωρο)

A picture containing screenshot, text, font

Description automatically generated

Δημιουργήσαμε το Correlation Heatmap για να δούμε εάν υπάρχουν συσχετίσεις μεταξύ χαρακτηριστικών ώστε να μην τα παραλείψουμε στις γραφικές παραστάσεις.



Τέλος, αποθηκεύουμε το νέο dataset σε ένα νέο αρχείο csv.



Γραφικές Παραστάσεις

Για την δημιουργία γραφικών παραστάσεων φορτώνουμε τις απαραίτητες βιβλιοθήκες (pandas, matplotlib.pyplot) καθώς και το καινούργιο αρχείο csv.

A picture containing text, font, screenshot, graphics

Description automatically generated

Και πάλι ενδεικτικά τυπώνουμε τις αρχικές τιμές του dataset, όπου βλέπουμε ότι έχουν προστεθεί οι στήλες Daily cases, Daily deaths και Season.



1. Cases by Date

Για κάθε μέρα του dataset αθροίζουμε όλα τα Cases από κάθε χώρα για να δούμε την συνολική πορεία του ιού.

A picture containing text, screenshot, plot, line

Description automatically generated

1. Deaths by Date

Αντιστοίχως για τα Deaths. Παρατηρούμε, όπως είναι λογικό ότι τα Deaths ακολουθούν τα Cases.

A picture containing text, screenshot, plot, line

Description automatically generated

1. Deaths by Season

A picture containing text, screenshot, diagram, display

Description automatically generated

Όπως είδαμε οι παραπάνω γραφικές παραστάσεις δείχνουν την συνολική εικόνα του ιού. Οι παρακάτω γραφικές παραστάσεις στοχεύουν στο να αναδείξουν την πορεία κάθε χώρας ώστε να έχουμε μια πρώτη εικόνα για το ποιες χώρες ξεχώρισαν θετικά και ποιες αρνητικά.

1. Mortality rate (συνολικό πλήθος θανάτων δια το πληθυσμό της χώρας) by country

A picture containing screenshot, plot, line

Description automatically generated

1. Positivity Rate (συνολικό πλήθος Cases δια το συνολικό πλήθος tests) by Country

A picture containing screenshot, line, plot

Description automatically generated

1. Death Rate (συνολικό πλήθος Deaths δια συνολικό πλήθος Cases) by Country

A picture containing screenshot, line, plot, diagram

Description automatically generated

1. Average Temperature by Country

A picture containing screenshot, plot, line

Description automatically generated

1. Hospital beds per 1000 people by Country

A picture containing screenshot, plot, line, diagram

Description automatically generated

1. Medical doctors per 1000 people by Country

A picture containing screenshot, plot, line

Description automatically generated

1. Population aged 65 more and over (%) by Country

A picture containing screenshot, plot, line, diagram

Description automatically generated

Από τις παραπάνω γραφικές παραστάσεις διακρίνεται η μη επιτυχημένη πορεία του Μεξικού το οποίο έχει το μεγαλύτερο δείκτη Positivity Rate καθώς και Death Rate. Κατά αντιστοιχία το Μεξικό φαίνεται να έχει χαμηλό δείκτη Hospital beds per 1000 people καθώς και μια τιμή κοντά στη μέση για τον δείκτη Medical Doctors per 1000 people.