

Ατομική Διπλωματική Εργασία

**Εξόρυξη Μέσων Κοινωνικής Δικτύωσης: Ανάλυση του
Συναισθήματος των Επιχειρηματιών στο Twitter**

Κωνσταντίνος Κλαρκ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Μάιος 2023

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Εξόρυξη Μέσων Κοινωνικής Δικτύωσης: Ανάλυση του Συναισθήματος των Επιχειρηματιών στο Twitter

Κωνσταντίνος Κλαρκ

Επιβλέπων Καθηγητής

Δικαιάκος Μάριος

Η Ατομική Διπλωματική Εργασία υποβλήθηκε προς μερική εκπλήρωση των
απαιτήσεων απόκτησης του πτυχίου Πληροφορικής του Τμήματος Πληροφορικής του
Πανεπιστημίου Κύπρου

Μάιος 2023

Ευχαριστίες

Θα ήθελα να εκφράσω τις ειλικρινείς μου ευχαριστίες στον επιβλέπων καθηγητή μου, Δρ.. Μάριος Δικαιάκος, για την πολύτιμη βοήθεια και καθοδήγηση που μου παρείχε κατά τη διάρκεια της πτυχιακής μου εργασίας. Η αφοσίωσή σας, η ικανότητα και η υπομονή σας ήταν πολύτιμα για να με βοηθήσαν να πετύχω τους ακαδημαϊκούς μου στόχους και είμαι ευγνώμων για τον χρόνο που αφιερώσατε μαζί μου.

Θα ήθελα επίσης να εκφράσω τη βαθιά μου εκτίμηση στον κ. Δημοσθένη Στεφανίδη, ο οποίος ήταν πάντα διαθέσιμος να απαντήσει σε οποιεσδήποτε ερωτήσεις είχα και προχώρησε πέρα και πέρα για να διασφαλίσει ότι είχα πρόσβαση στα υλικά που χρειαζόμουν.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου για τη συνεχή υποστήριξη και ενθάρρυνση. Η πίστη τους σε μένα με κράτησε σε όλη τη διάρκεια της διατριβής και δεν θα μπορούσα να την ολοκληρώσω χωρίς την αγάπη και την υποστήριξή τους.

Περίληψη

Η παρούσα μελέτη επιχειρεί να εξετάσει τον τρόπο με τον οποίο οι επιχειρηματίες εκφράζουν τα συναισθήματά τους στο Twitter και να καθορίσει αν έχουν διαφορές, σε σχέση με τους μη επιχειρηματίες. Η μελέτη εξετάζει επίσης πώς άλλαξαν οι στάσεις των επιχειρηματιών κατά τη διάρκεια της κρίσης COVID-19. Χρησιμοποιήσαμε μεθόδους επεξεργασίας φυσικής γλώσσας για να αναλύσουμε σχολαστικά τα συναισθηματικά μοτίβα σε ένα μεγάλο σύνολο δεδομένων 30,1 εκατομμυρίων tweets από 33.825 άτομα.

Τα αποτελέσματα επιβεβαίωσαν την εγκυρότητα προηγούμενων ερευνητικών ευρημάτων και παρείχαν πληροφορίες για τη συναισθηματική δυναμική των επιχειρηματιών. Η έρευνά μας επιβεβαίωσε τα προηγούμενα ευρήματα δείχνοντας ότι οι επιχειρηματίες είχαν πράγματι μεγαλύτερη τάση να εμφανίζουν θετικές απόψεις σε σύγκριση με τους μη επιχειρηματίες. Επιπλέον, η εξέταση των συναισθηματικών μοτίβων κατά τη διάρκεια της πανδημίας COVID-19 παρείχε σημαντικές πληροφορίες σχετικά με τον τρόπο με τον οποίο οι επιχειρηματίες εκφράζανε τις συναισθηματικές τους εκδηλώσεις σε αντίθεση με τους μη επιχειρηματίες υπό δύσκολες συνθήκες.

Τα ευρήματα της παρούσας μελέτης παρέχουν πολύτιμες πληροφορίες σχετικά με τα συναισθήματα που εκφράζουν οι επιχειρηματίες στο Twitter και τις πιθανές επιπτώσεις τους. Παρατηρείται συνοχή σε πολλαπλούς ερευνητικούς στόχους, με τη συμπερίληψη ενός σημαντικού γεγονότος, όπως η πανδημία COVID-19, που υποστηρίζει περαιτέρω την άποψη ότι οι επιχειρηματίες τείνουν να επιδεικνύουν θετική στάση στις πλατφόρμες κοινωνικής δικτύωσης. Τα ευρήματα αυτά έχουν σημαντικές επιπτώσεις στην κατανόηση της επιχειρηματικής συμπεριφοράς και του ρόλου των μέσων κοινωνικής δικτύωσης στη διαμόρφωση της κοινής γνώμης.

Περιεχόμενα

Κεφάλαιο 1 Εισαγωγή	7
1.1 Κίνητρα και Στόχοι	7
1.2 Περίγραμμα της ερευνητικής εργασίας	9
Κεφάλαιο 2 Μεθοδολογία.....	11
2.1 Διασωλήνωση Εργασιών	11
2.2 Εργαλεία	12
2.2.1 Tweepy	12
2.2.2 Python και Βιβλιοθήκες.....	13
2.2.3 VADER.....	15
2.2.4 uClassify.....	15
2.2.5 MongoDB.....	16
2.3 Συλλογή και Επεξήγηση Δεδομένων	17
2.3.1 Twitter	17
2.3.2 Συλλογή Δεδομένων	18
2.3.3 Μεταβλητές.....	20
2.4 Προ επεξεργασία.....	24
2.5 Προκλήσεις ή Περιορισμοί	26
Κεφάλαιο 3 Αποτελέσματα	27
3.1 Σύγκριση Επιχειρηματιών με Μη-Επιχειρηματιών.....	27
3.2 Ανάλυση Θεμάτων	31
3.3 Σύγκριση αποτελεσμάτων με προηγούμενο άρθρο.....	34

3.4 Κορονοϊός (COVID-19)	35
3.4.1 Ορόσημα	37
3.4.2 Περίοδοι Έρευνας.....	37
3.4.3 Αποτελέσματα	39
3.4.4 Σύγκριση με μέσου συναισθήματος με ημερήσιο αριθμών κρουσμάτων....	40
 Κεφάλαιο 4 Συζήτηση – Συμπεράσματα	45
 Βιβλιογραφία.....	48

Κεφάλαιο 1

Εισαγωγή

1.1 Κίνητρα και Στόχοι	7
1.2 Περίγραμμα της ερευνητικής εργασίας	9

1.1 Κίνητρα και Στόχοι

Τα άτομα χρησιμοποιούν όλο και περισσότερο πλατφόρμες μέσω κοινωνικής δικτύωσης όπως το Twitter για να μοιραστούν τις ιδέες και τα συναισθήματά τους σχετικά με μια ποικιλία θεμάτων. Λόγω της ανόδου των μέσω κοινωνικής δικτύωσης, οι ερευνητές έχουν πλέον πρόσβαση σε τεράστιους όγκους δεδομένων που δημιουργούνται από τους χρήστες, τα οποία παρέχουν πληροφορίες για τις πεποιθήσεις και τις δραστηριότητες των ανθρώπων. Η ανάλυση του συναισθήματος των tweets, που μπορεί να αποκαλύψει σημαντικές πληροφορίες σχετικά με τα συναισθήματα των ανθρώπων για συγκεκριμένα θέματα, είναι ένας τομέας έρευνας που έχει προκύψει.

Τα τελευταία χρόνια, έχει αυξηθεί το ενδιαφέρον για την ανάλυση του συναισθήματος των tweets που παράγονται από επιχειρηματίες και μη. Οι ερευνητές ενδιαφέρθηκαν ιδιαίτερα να προσδιορίσουν εάν οι επιχειρηματίες έχουν υψηλότερο θετικό συναίσθημα σε κοινωνικά δίκτυα όπως το Twitter από τους μη επιχειρηματίες. Ο λόγος αυτής της περιέργειας είναι ότι οι επιχειρηματίες γενικά θεωρούνται αισιόδοξα, ριψοκίνδυνα και καινοτόμα άτομα. Ως αποτέλεσμα, είναι λογικό να υποθέσουμε ότι έχουν μια πιο αισιόδοξη άποψη για τη ζωή και εκφράζουν αυτό το συναίσθημα πιο συχνά στα μέσα κοινωνικής δικτύωσης.[6-7]

Ως αποτέλεσμα, ο πρωταρχικός στόχος αυτής της διατριβής είναι να διερευνήσει εάν οι επιχειρηματίες έχουν υψηλότερο θετικό συναίσθημα σε κοινωνικά δίκτυα όπως το Twitter από τους μη επιχειρηματίες. Επιπρόσθετα, ο καθορισμός του εάν οι ιδιοκτήτες επιχειρήσεων έχουν υψηλότερο θετικό συναίσθημα στα κοινωνικά δίκτυα όπως το Twitter από τους μη επιχειρηματίες είναι κρίσιμο για την κατανόηση των στάσεων, των εμπειριών και των αντιλήψεών τους στον ψηφιακό τομέα. Τα ευρήματα αυτής της μελέτης μπορούν να συμβάλουν στην ακαδημαϊκή γνώση, καθώς και να ενημερώσουν πολιτικές και συστήματα υποστήριξης που στοχεύουν στην καλλιέργεια ενός ζωντανού επιχειρηματικού οικοσυστήματος.

Ερευνητικός στόχος 1: Οι επιχειρηματίες εκφράζουν πιο θετικά συναισθήματα στα κοινωνικά δίκτυα από τους μη-επιχειρηματίες;

Επιπλέον, ο δεύτερος στόχος είναι να διερευνήσουμε εάν το συναίσθημα που εκφράζεται μέσα από αυτές τις δύο ομάδες αλλάζει σε περιόδους κρίσης. Η έρευνα για το συναίσθημα των tweets που δημοσιεύονται από ιδιοκτήτες επιχειρήσεων και μη επιχειρηματίες κατά τη διάρκεια κανονικών καταστάσεων και καταστάσεων κρίσης είναι σημαντική για διάφορους λόγους. Αρχικά, η γνώση των διακυμάνσεων των συναισθημάτων μεταξύ αυτών των δύο ομάδων μπορεί να προσφέρει χρήσιμες γνώσεις για τις στάσεις και τις προοπτικές εκείνων με και χωρίς επιχειρηματική εμπειρία απέναντι σε ορισμένες προκλήσεις. Επιπλέον, η μελέτη των αλλαγών στο συναίσθημα σε περιόδους κρίσης μπορεί να δώσει φως στον αντίκτυπο τέτοιων γεγονότων στις στάσεις και τα συναισθήματα των ανθρώπων. Αυτά τα δεδομένα μπορούν να χρησιμοποιηθούν για τη δημιουργία πιο αποτελεσματικών πολιτικών και στρατηγικών διαχείρισης κρίσεων.

Για να απαντήσουμε στα δυο αυτά ερωτήματα, θα κάνουμε ανάλυση συναισθήματος σε ένα μεγάλο σύνολο δεδομένων tweets που συγκεντρώθηκαν από το Twitter από το 2016 μέχρι και το 2022 συγκρίνοντας το συναίσθημα που εκφράζεται από επιχειρηματίες και μη. Συγκεκριμένα για την δεύτερο στόχο, η περίοδος κρίσης η οποία

διαλέξαμε είναι η πανδημία του Κορονοϊού (Covid-19) όπου και πάλι συγκρίναμε τις δυο ομάδες με βάση το συναίσθημα που εκφράζουν στα μέσα κοινωνικής δικτύωσης.

Ερευνητικός στόχος 2: Πώς επηρεάζονται οι δύο ομάδες χρηστών σε περιόδους κρίσης, όπως ο κορονοϊός;

Τέλος, ο τρίτος στόχος της έρευνας είναι να ενημερώσουμε την προηγούμενη έρευνα, με δεδομένα μέχρι και σήμερα, και να αξιολογήσουμε εάν τα προηγούμενα συμπεράσματα εξακολουθούν να ισχύουν ακόμη.

Ερευνητικός στόχος 3: Υπάρχουν αλλαγές σε σχέση με την προηγούμενη έρευνα;

Συνολικά, ο σκοπός αυτής της έρευνας είναι να συνεισφέρει στη βιβλιογραφία αποκαλύπτοντας διαφορές στο συναίσθημα μεταξύ επιχειρηματιών και μη επιχειρηματιών σε κοινωνικά δίκτυα όπως το Twitter. Αυτή η μελέτη θα ρίξει επίσης φως στο πώς το συναίσθημα των tweets που γίνονται από αυτές τις δύο ομάδες ποικίλλει σε περιόδους κρίσης, προσφέροντας χρήσιμες πληροφορίες.

1.2 Περίγραμμα της ερευνητικής εργασίας

Το πρώτο κεφάλαιο χρησιμεύει ως εισαγωγή, περιγράφοντας τα κίνητρα πίσω από την έρευνα και παρουσιάζοντας τους στόχους που στοχεύει να επιτύχει το έργο. Καθορίζει το πλαίσιο για τα επόμενα κεφάλαια.

Το δεύτερο κεφάλαιο βρίσκεται στο επίκεντρο παρουσιάζοντας λεπτομερώς τη θεωρητική βάση και τις μεθόδους του έργου. Περιγράφει διεξοδικά τις τεχνολογίες, τις πλατφόρμες και τα εργαλεία που χρησιμοποιούνται για την επίτευξη των στόχων του έργου. Αυτό το κεφάλαιο εμβαθύνει στις προσεγγίσεις που χρησιμοποιήθηκαν, συζητώντας την αιτιολόγησή τους και δείχνοντας πώς βοηθούν στην επίτευξη των στόχων της έρευνας. Αποτελεί ουσιαστική βάση για την κατανόηση των τεχνικών πτυχών και της γενικής προσέγγισης του έργου.

Το τρίτο κεφάλαιο, η εστίαση μετατοπίζεται στην παρουσίαση των αποτελεσμάτων. Αυτό το κεφάλαιο παρέχει μια λεπτομερή ανάλυση των δεδομένων που συλλέχθηκαν και υποβλήθηκαν σε επεξεργασία κατά τη διάρκεια του έργου. Παρουσιάζει τα ευρήματα και τα συζητά στο πλαίσιο των στόχων της έρευνας. Αυτό το κεφάλαιο περιλαμβάνει οπτικές αναπαραστάσεις, στατιστική ανάλυση και ερμηνείες των αποτελεσμάτων για να παρέχει μια ολοκληρωμένη κατανόηση των αποτελεσμάτων του έργου.

Στη συνέχεια, η διατριβή συνεχίζει στο τελευταίο κεφάλαιο, τα συμπεράσματα. Αυτό το κεφάλαιο συνοψίζει τις σημαντικές ιδέες και συμπεράσματα από ολόκληρο το ερευνητικό έργο. Παρέχει μια πλήρη σύνθεση της διπλωματικής, αναλύοντας τις επιπτώσεις των ευρημάτων, αντιμετωπίζοντας τυχόν περιορισμούς ή εμπόδια που βρέθηκαν και προτείνοντας πιθανές μελλοντικές έρευνες ή πρακτικές εφαρμογές. Το κεφάλαιο συμπερασμάτων δίνει μια ολοκληρωμένη αξιολόγηση των επιτευγμάτων του έργου, υπογραμμίζει τη συμβολή του στην περιοχή και υπογραμμίζει τη σημασία του σε σχέση με το ευρύτερο ακαδημαϊκό ή πρακτικό περιβάλλον.

Κεφάλαιο 2

Μεθοδολογία

2.1 Διασωλήνωση Εργασιών	11
2.2 Εργαλεία	12
2.2.1 Tweepy	12
2.2.2 Python και Βιβλιοθήκες	13
2.2.3 VADER	15
2.2.4 uClassify	16
2.2.5 MongoDB	16
2.3 Συλλογή και Επεξήγηση Δεδομένων	17
2.3.1 Twitter	17
2.3.2 Συλλογή Δεδομένων	18
2.3.3 Μεταβλητές	20
2.4 Προ επεξεργασία	25
2.5 Προκλήσεις ή Περιορισμοί	26

Σε αυτή την ενότητα, θα εμβαθύνουμε στις λεπτομέρειες του τρόπου λήψης των δεδομένων και των διαφόρων μεθόδων επεξεργασίας που εφαρμόστηκαν. Συγκεκριμένα, θα ρίξουμε μια εις βάθος ματιά στα διάφορα εργαλεία που χρησιμοποιήθηκαν κατά την ερευνητική διαδικασία και θα εξετάσουμε τα δεδομένα που συλλέχθηκαν.

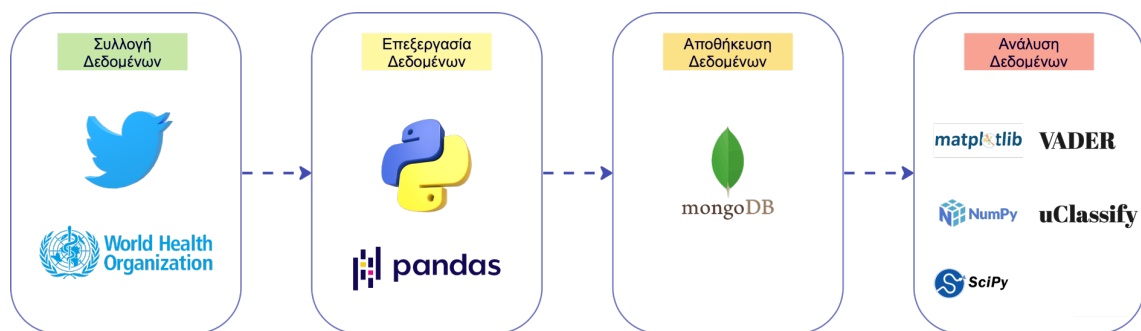
2.1 Διασωλήνωση Εργασιών

Το πρώτο στάδιο της διαδικασίας αφορούσε τη συλλογή δεδομένων, η οποία διεξήχθη μέσω του API (Διεπαφή προγραμματισμού εφαρμογών) του Twitter, καθώς και της βιβλιοθήκης Tweepy. Οι χρήστες του Twitter, πάρθηκαν από το άρθρο το οποίο έχουμε σαν στόχο να εξετάσουμε αν ακόμη τα αποτελέσματα του είναι συνεπή[8]. Ο στόχος της συλλογής δεδομένων ήταν να συγκεντρωθούν tweets για επιχειρηματίες και μη για σκοπούς ανάλυσης.

Μετά τη συλλογή των δεδομένων, το επόμενο στάδιο περιελάμβανε την επεξεργασία τους. Έτσι, χρειαζόταν η αφαίρεση τυχόν στοιχείων που το εργαλείο ανάλυσης συναισθημάτων και θεμάτων δεν μπορούσε να αποδεχτεί ή μπορεί να είχε μπερδέψει. Αυτό έγινε για να διασφαλίσουμε ότι τα αποτελέσματα της ανάλυσής μας ήταν όσο το δυνατόν πιο ακριβή.

Το δεύτερο στάδιο της έρευνας μας περιελάμβανε την ανάλυση του κειμένου κάθε tweet μέσω της ανάλυσης συναισθημάτων και θεμάτων. Για το σκοπό αυτό χρησιμοποιήσαμε τα εργαλεία VADER και uClassify αντίστοιχα.

Τέλος, το τελικό στάδιο της διαδικασίας συλλογής δεδομένων περιελάμβανε την αποθήκευση των δεδομένων στη βάση δεδομένων. Αυτό έγινε για να διασφαλιστεί ότι τα δεδομένα παρέμειναν οργανωμένα και άμεσα διαθέσιμα για μελλοντική αναφορά.



Σχήμα 1.1 Διασωλήνωση Εργασιών

2.2 Εργαλεία

2.2.1 Tweepy[1]

Για την συλλογή των δεδομένων χρησιμοποιήθηκε το πακέτο ανοικτού κώδικα Tweepy. Το Tweepy είναι ένα πακέτο με σκοπό να προσφέρει στον χρήστη έναν εύκολο, άμεσο τρόπο πρόσβασης στο API (Διεπαφή προγραμματισμού εφαρμογών) του Twitter. Όλες οι πιθανές κλήσεις API που προσφέρονται από το Twitter API μπορούν να εκτελεστούν και μέσω εντολών tweepy. Το πακέτο διαχειρίζεται την κωδικοποίηση και την

αποκωδικοποίηση των αρχείων JSON που ανταλλάσσονται μεταξύ του πελάτη και του API και επιτρέπει την απλοποιημένη ανάκτηση συγκεκριμένων στοιχείων από τις απαντήσεις JSON χωρίς να χρειάζεται ο χρήστης να αλληλοεπιδράσει καθόλου με το αρχείο JSON. Το Tweepy μας επέτρεψε να πραγματοποιήσουμε έλεγχο ταυτότητας με το API του Twitter προγραμματιστικά, να ορίσουμε προσαρμοσμένες παραμέτρους ερωτήματος και να ανακτήσουμε tweets με βάση συγκεκριμένα κριτήρια, όπως εύρος ημερομηνιών, πληροφορίες χρήστη και λέξεις-κλειδιά. Ήμασταν σε θέση να συλλέξουμε αποτελεσματικά μεγάλο όγκο δεδομένων Twitter για την ανάλυσή μας αξιοποιώντας τις δυνατότητες του Tweepy. Επιπλέον, επειδή η Python είναι μια ευρέως χρησιμοποιούμενη γλώσσα στην επιστήμη των δεδομένων, η ενσωμάτωση του Tweepy με την Python το κατέστησε μια βολική και κατάλληλη επιλογή για τις ανάγκες συλλογής δεδομένων μας.

2.2.2 Python και Βιβλιοθήκες

Αρχικά, η Python[9] έχει ένα απλό και ευανάγνωστο συντακτικό, καθιστώντας εύκολη τη συγγραφή και την κατανόηση του κώδικα. Δεύτερον, η Python παρέχει ένα τεράστιο οικοσύστημα βιβλιοθηκών και πακέτων που έχουν σχεδιαστεί ειδικά για την ανάλυση δεδομένων, όπως οι pandas[10], numpy[11] και matplotlib[12], οι οποίες επιτάχυναν σημαντικά τη διαδικασία της ανάλυσής μας. Η ευελιξία της Python επιτρέπει επίσης την απρόσκοπτη ενσωμάτωσή της με άλλες γλώσσες και τεχνολογίες, καθιστώντας την εξαιρετική επιλογή για την εργασία. Επιπλέον, η ενεργή και υποστηρικτική κοινότητα της Python παρέχει εκτεταμένη τεκμηρίωση, σεμινάρια και πόρους, καθιστώντας απλή την επίλυση οποιωνδήποτε προβλημάτων αντιμετωπίσαμε. Τέλος, η συμβατότητα της Python με πολλαπλές πλατφόρμες εξασφάλισε ότι ο κώδικάς μας θα μπορούσε να εκτελεστεί σε διάφορα λειτουργικά συστήματα χωρίς σημαντικές αλλαγές. Η Python ήταν η ιδανική γλώσσα προγραμματισμού για το έργο μας ανάλυσης δεδομένων λόγω της απλότητας, του πλούσιου οικοσυστήματος βιβλιοθηκών, της ευελιξίας και της ισχυρής υποστήριξης της κοινότητας.

Pymongo[13]: Για την αποθήκευση και τη διαχείριση των δεδομένων που συλλέξαμε, χρησιμοποιήσαμε τη βιβλιοθήκη pymongo για την αλληλεπίδραση με τη MongoDB, μια

δημοφιλή βάση δεδομένων NoSQL. Το pymongo παρείχε μια απλή διεπαφή για τη σύνδεση με τη βάση δεδομένων, την εισαγωγή και ανάκτηση δεδομένων και την αποτελεσματική εκτέλεση ερωτημάτων.

Numpy: Για αποδοτικούς αριθμητικούς υπολογισμούς και λειτουργίες συστοιχιών, χρησιμοποιήσαμε τη βιβλιοθήκη numpy. Παρείχε μια ποικιλία μαθηματικών συναρτήσεων και βοηθητικών προγραμμάτων που ήταν χρήσιμα στην επεξεργασία και ανάλυση των δεδομένων μας. Επιπλέον, η δομή δεδομένων συστοιχιών της numpy διευκόλυνε τους υπολογισμούς υψηλής απόδοσης και μας επέτρεψε να χειριστούμε αποτελεσματικά μεγάλα σύνολα δεδομένων.

Pandas: Για την επεξεργασία και την ανάλυση των δεδομένων, χρησιμοποιήσαμε τη βιβλιοθήκη pandas. Η pandas παρείχε ισχυρές δομές δεδομένων, όπως τα DataFrames, που μας επέτρεψαν να οργανώσουμε, να φιλτράρουμε και να μετασχηματίσουμε εύκολα τα δεδομένα μας. Το εκτεταμένο σύνολο συναρτήσεων και μεθόδων του μας επέτρεψε να εκτελέσουμε αποτελεσματικά περιγραφικές στατιστικές, να χειριστούμε τις ελλείπουσες τιμές και να συγχωνεύσουμε σύνολα δεδομένων.

Wordcloud[14]: Χρησιμοποιήσαμε τη βιβλιοθήκη wordcloud για να οπτικοποιήσουμε τις πιο συχνά εμφανιζόμενες λέξεις στα δεδομένα κειμένου μας. Παρείχε μια απλή αλλά αποτελεσματική μέθοδο για τη δημιουργία οπτικά ελκυστικών νεφών λέξεων στα οποία το μέγεθος κάθε λέξης αντιστοιχούσε στη συχνότητά της. Χάρη σε αυτή τη βιβλιοθήκη μπορέσαμε να αποκτήσουμε γνώσεις σχετικά με τα εξέχοντα θέματα και μοτίβα στα δεδομένα κειμένου μας.

Scipy.stats[15]: Η βιβλιοθήκη scipy.stats προσφέρει ένα ευρύ φάσμα στατιστικών συναρτήσεων και κατανομών πιθανοτήτων. Αυτή η βιβλιοθήκη χρησιμοποιήθηκε για την εκτέλεση στατιστικής ανάλυσης και ελέγχου υποθέσεων στα δεδομένα μας. Μας επέτρεψε να υπολογίσουμε p-values, να εκτελέσουμε t-tests και να αξιολογήσουμε τη σημαντικότητα των ευρημάτων μας χρησιμοποιώντας καθιερωμένες στατιστικές μεθόδους.

Matplotlib.pyplot: Χρησιμοποιήσαμε τη βιβλιοθήκη matplotlib.pyplot για την οπτικοποίηση δεδομένων, η οποία είναι ένα ισχυρό εργαλείο γραφικών παραστάσεων στην Python. Παρείχε μια ποικιλία συναρτήσεων για τη δημιουργία διαφορετικών τύπων διαγραμμάτων και γραφικών παραστάσεων, όπως γραμμικά διαγράμματα,

ραβδογράμματα και ιστογράμματα. Χρησιμοποιήσαμε αυτή τη βιβλιοθήκη για να παρουσιάσουμε οπτικά τα ευρήματά μας, να διερευνήσουμε τις τάσεις και να επικοινωνήσουμε αποτελεσματικά τα ευρήματά μας.

2.2.3 VADER

Το VADER[2] (Valence Aware Dictionary and SEntiment Reasoner) είναι ένα εργαλείο ανάλυσης συναισθήματος βασισμένο σε λεξικό και κανόνες που είναι ειδικά προσαρμοσμένο στα συναισθήματα που εκφράζονται στα μέσα κοινωνικής δικτύωσης. Το VADER παίρνει ως είσοδο το αναρτημένο κείμενο, τα emoticon και τα hashtags των tweet και επιστρέφει μια βαθμολογία που υποδεικνύει την αρνητικότητα ή τη θετικότητα ενός tweet. Αυτή η βαθμολογία κυμαίνεται από -1 (ακραία αρνητικότητα) έως 1 (ακραία θετικότητα). Συγκεκριμένα θεωρήσαμε αρνητικά τα tweets με βαθμολογία μικρότερη του -0.05, ουδέτερα με βαθμολογία από το -0.05 μέχρι το 0.05 και θετικά με βαθμολογία μεγαλύτερη του 0.05. Η βαθμολογία συναισθήματος ενός tweet στο VADER λαμβάνεται συνοψίζοντας την ένταση κάθε λέξης στο κείμενο. Η αντιστοίχιση κάθε λέξης με τη βαθμολογία συναισθημάτων της κατασκευάζεται εξετάζοντας καλά εδραιωμένες τράπεζες λέξεων συναισθημάτων. Αυτά συνδυάζονται με γενικούς κανόνες που ενσωματώνουν γραμματικές και συντακτικές συμβάσεις για την έκφραση και την έμφαση της έντασης του συναισθήματος. Επιπλέον, ενσωματώνονται πολλά λεξιλογικά χαρακτηριστικά που είναι κοινά στην έκφραση συναισθημάτων, συμπεριλαμβανομένης μιας πλήρους λίστας emoticons (για παράδειγμα, το ":-)") υποδηλώνει ένα "χαμογελαστό πρόσωπο" και γενικά υποδηλώνει θετικό συναίσθημα), ακρωνύμια που σχετίζονται με το συναίσθημα (π.χ., LOL) και συνήθως χρησιμοποιούμενη αργκό με συναισθηματική αξία (π.χ. "nah", "meh" και "giggly"). Επίσης, το VADER μπορεί να χειριστεί τυπικές περιπτώσεις άρνησης χρήσης όπως "δεν ήταν καλό", "δεν ήταν πολύ καλό" κ.λπ., και δεν επηρεάζεται από περιπτώσεις όπου η άρνηση αντιστρέφει το συναίσθημα του κειμένου.

2.2.4 uClassify[3]

Στην συνέχεια, κατηγοριοποιήσαμε τα tweets μας βάση του θέματος που ασχολούνταν. Αρχικά για την εύρεση του θέματος κάθε tweets χρησιμοποιήσαμε το εργαλείο

uClassify πιο συγκεκριμένα το ταξινομητής Topics. Το uClassify είναι μια υπηρεσία Ιστού μηχανικής εκμάθησης ανοιχτού κώδικα που περιέχει αρκετούς προ εκπαιδευμένους ταξινομητές κειμένου. Αυτός ο ταξινομητής κατηγοριοποιεί ένα μη δομημένο κείμενο σε 10 γενικά θέματα, και συγκεκριμένα: τέχνες, επιχειρήσεις, υπολογιστές, παιχνίδια, υγεία, σπίτι, αναψυχή, επιστήμη, κοινωνία και αθλητισμός. Είναι κατάλληλο τόσο για σύντομα όσο και για μεγάλα κείμενα (tweets, καταστάσεις στο Facebook, αναρτήσεις ιστολογίου, κριτικές προϊόντων κ.λπ.).

Από τεχνική άποψη, το uClassify περιλαμβάνει έναν ταξινομητή Naïve Bayes που έχει εκπαιδευτεί σε 2,8 εκατομμύρια έγγραφα με δεδομένα από το Twitter, κριτικές προϊόντων της Amazon και κριτικές ταινιών. Αυτά τα θέματα (κατηγορίες) προέρχονται από το Open Directory Project. Για οποιοδήποτε δεδομένο κείμενο, το uClassify Topics API επιστρέφει την πιθανότητα (εύρος τιμών από 0 έως 1) ότι ένα συγκεκριμένο tweet ανήκει σε ένα συγκεκριμένο θέμα. Έτσι, το θέμα με τη μεγαλύτερη πιθανότητα είναι πιθανότατα το θέμα στο οποίο θα έπρεπε να ανήκει το συγκεκριμένο tweet.

2.2.5 MongoDB[4]

Κυκλοφόρησε τον Φεβρουάριο του 2009, η MongoDB είναι μια βάση δεδομένων NoSQL. Έχει πολλά πλεονεκτήματα σε σχέση με άλλες γνωστές βάσεις δεδομένων, καθιστώντας την μια από τις πιο συχνά χρησιμοποιούμενες βάσεις δεδομένων σε εφαρμογές Big Data και Web σε πραγματικό χρόνο. Ως βάσης δεδομένων NoSQL, δίνει μεγαλύτερη έμφαση στην ανοχή, την ταχύτητα και τη διαθεσιμότητα παρά στη συνοχή της δομής δεδομένων. Αυτό είναι ιδιαίτερα χρήσιμο σε καταστάσεις όπου η διατήρηση της συνέπειας των δεδομένων δεν είναι απαραίτητη, όπως σε αυτήν την προσπάθεια όπου πολλά tweets στερούνται ορισμένων χαρακτηριστικών. Μια SQL δομή δεδομένων θα παρήγαγε πολλές μηδενικές καταχωρήσεις και θα απαιτούσε περισσότερο χώρο αποθήκευσης, ενώ η MongoDB επιτρέπει καταχωρήσεις δεδομένων χωρίς περιορισμούς στα πεδία της, γεγονός που βελτιώνει την ταχύτητα εξοικονομώντας χώρο αποθήκευσης.

Μαζί με τα πλεονεκτήματα των εννοιών NoSQL, η MongoDB παρέχει επίσης μια σειρά από ειδικά πλεονεκτήματα. Ο χειρισμός της βάσης δεδομένων είναι γρήγορος και

απλός, χάρη στις ισχυρές εντολές ερωτημάτων και τα εργαλεία της, τα οποία είναι συγκρίσιμα με τις κοινές εντολές ερωτήματος SQL όσον αφορά τη σύνταξη και τη λειτουργικότητα. Επιπλέον, αντί να χρησιμοποιεί το συμβατικό παράδειγμα γραμμής/στήλης, η MongoDB αποθηκεύει δεδομένα σε έγγραφα JSON. Η μορφή δεδομένων JSON προσφέρει καλύτερη απόδοση και μικρότερα μεγέθη δεδομένων.

2.3 Συλλογή και Επεξήγηση Δεδομένων

2.3.1 Twitter

Αρχικά τα δεδομένα πάρθηκαν από το Twitter, έναν από τα μεγαλύτερα κοινωνικά δίκτυα και η κύρια διαδικτυακή πλατφόρμα δημόσιας έκφρασης. Όπου υπάρχουν καθημερινά περίπου 240 εκατομμύρια ενεργοί χρήστες βάση στο [statista.com](https://www.statista.com). Το Twitter παρέχει ισχυρά εργαλεία API που επιτρέπουν στους προγραμματιστές να αξιοποιήσουν το τεράστιο όγκο δεδομένων από μια παγκόσμια βάση χρηστών που περιλαμβάνει δημόσια πρόσωπα και επιχειρήσεις, εκτός από κανονικούς ανθρώπους. Η μορφή περιεχομένου στην πλατφόρμα είναι επίσης ιδιαίτερα ωφέλιμη για αυτό τον σκοπό, καθώς αποτελείται από κείμενο περιορισμένο σε 280 χαρακτήρες με hashtags που καθορίζουν την εστίαση του περιεχομένου. Η χρήση του Twitter ως πηγή δεδομένων και του API του Twitter προσφέρει πολλά πλεονεκτήματα για τη λήψη δεδομένων. Αρχικά, το Twitter είναι μια δημοφιλής πλατφόρμα κοινωνικής δικτύωσης με μεγάλη βάση χρηστών, γεγονός που το καθιστά πλούσια πηγή ποικίλων πληροφοριών σε πραγματικό χρόνο. Το API του Twitter μας επιτρέπει την προγραμματιστική πρόσβαση σε αυτόν τον πλούτο δεδομένων, επιτρέποντας την αποτελεσματική και αυτοματοποιημένη συλλογή δεδομένων. Επιπλέον, το API του Twitter παρέχει πρόσβαση σε μια ποικιλία τελικών σημείων, επιτρέποντάς μας να ανακτήσουμε όχι μόνο πληροφορίες που αφορούν συγκεκριμένους χρήστες, αλλά και tweets, hashtags, τάσεις και άλλα δεδομένα. Αυτή η προσαρμοστικότητα μας επιτρέπει να προσαρμόζουμε τη συλλογή δεδομένων μας σε συγκεκριμένους ερευνητικούς στόχους. Επιπλέον, το API του Twitter περιλαμβάνει όρια ρυθμού και μηχανισμούς ελέγχου ταυτότητας για την προστασία των δεδομένων και την αποτροπή της κατάχρησης. Αυτά τα χαρακτηριστικά, σε συνδυασμό με την ευκολία χρήσης και την

εκτεταμένη τεκμηρίωση του Twitter API, το καθιστούν εξαιρετική επιλογή για την αποτελεσματική και αξιόπιστη εξαγωγή πολύτιμων πληροφοριών από τα δεδομένα του Twitter.

2.3.2 Συλλογή Δεδομένων

Με την χρήση του Tweepy και της λίστα με χρήστες από το άρθρο Waters J, Nicolaou N, Stefanidis D, Efstathiades H, Pallis G, Dikaiakos M (2021) Exploring the sentiment of entrepreneurs on Twitter, κατεβάσαμε τα δεδομένα για 33,825 χρήστες, συγκεκριμένα 5034 από το Λος Άντζελες, 2626 από το Λονδίνο και 26165 παγκόσμιος.

Συνολικά καταλήξαμε με 30 εκατομμύρια tweets και για τις τρεις τοποθεσίες.

Ο πίνακας 2.1 περιέχει τα δεδομένα του Twitter για την έρευνά μας, τα ονομάζουμε “dataset16-22”, καθώς και δεδομένα από την εργασία που έχουμε σκοπό να επικαιροποιήσουμε και να συγκρίνουμε τα αποτελέσματα της, τα ονομάζουμε “dataset13-15”. Οι περίοδοι συλλογής δεδομένων διαφέρουν, με τα δεδομένα του εγγράφου να εκτείνονται από το 2013 έως το 2015 και τα δικά μας δεδομένα να εκτείνονται από το 2016 έως το 2022. Όσον αφορά το σύνολο των χρηστών, το έγγραφο είχε 60.556 χρήστες, ενώ εμείς έχουμε 33.825 χρήστες. Η μείωση του αριθμού των χρηστών μπορεί να αποδοθεί σε διάφορους παράγοντες, συμπεριλαμβανομένων των διαγραφών λογαριασμών, των ανενεργών λογαριασμών και των αλλαγών στις πολιτικές απορρήτου των λογαριασμών. Όταν συγκρίνουμε τη γεωγραφική κατανομή των χρηστών, παρατηρούμε κάποιες διαφορές μεταξύ των δεδομένων της δημοσίευσης και των δικών μας δεδομένων. Η δημοσίευση είχε 7.838 χρήστες στο Λος Άντζελες, ενώ εμείς έχουμε 5.034 χρήστες. Ακόμη, η δημοσίευση είχε 3.572 χρήστες στο Λονδίνο, ενώ εμείς έχουμε 2.626 χρήστες. Η πιο σημαντική διαφορά παρατηρείται στην παγκόσμια κατηγορία, όπου το άρθρο είχε 49.146 χρήστες σε σύγκριση με τους 26.165 δικούς μας. Επιπλέον, ο πίνακας δείχνει τον αριθμό των επιχειρηματιών και των μη επιχειρηματιών σε κάθε τοποθεσία. Το άρθρο ανέφερε 3.919 επιχειρηματίες στο Λος Άντζελες, 1.786 στο Λονδίνο και 24.573 παγκοσμίως.

Πίνακας 2.1 Ανάλυση χρηστών από προηγούμενη ερευνά

Σύνολο Δεδομένων	Dataset13-15				Dataset16-22			
Χρονολογίες συλλογής δεδομένων	2013-2015				2016-2022			
Συνολικός αριθμός χρηστών	60556				33825			
	Περιοχή		Αριθμός		Περιοχή		Αριθμός	
	Λος Άντζελες		7838		Λος Άντζελες		5034	
	Λονδίνο		3572		Λονδίνο		2626	
	Παγκόσμια		49146		Παγκόσμια		26165	
		Λος Άντζελες	Λονδίνο	Παγκόσμια		Λος Άντζελες	Λονδίνο	Παγκόσμια
Αριθμός επιχειρηματιών και μη-επιχειρηματιών	Επιχειρηματίες	3,919	1,786	24,573	Επιχειρηματίες	775	339	6,103
	Μη-επιχειρηματίες	3,919	1,786	24,573	Μη-επιχειρηματίες	4,259	2,287	20,062

Σύμφωνα με τα δικά μας στοιχεία, υπάρχουν 775 επιχειρηματίες στο Λος Άντζελες, 339 στο Λονδίνο και 6.103 παγκοσμίως. Επιπρόσθετα, το άρθρο ανέφερε 3.919 μη επιχειρηματίες στο Λος Άντζελες, 1.786 στο Λονδίνο και 24.573 παγκοσμίως. Σύμφωνα με τα στοιχεία μας, υπάρχουν 4.259 μη επιχειρηματίες στο Λος Άντζελες, 2.287 στο Λονδίνο και 20.062 παγκοσμίως. Όταν εξετάζουμε τη γεωγραφική κατανομή των χρηστών, παρατηρούμε διαφορές μεταξύ των δεδομένων της δημοσίευσης και των δικών μας δεδομένων. Το Λος Άντζελες αντιπροσώπευε 7.838 χρήστες στην εργασία, αντιπροσωπεύοντας το 13% του συνολικού δείγματός του. Το Λος Άντζελες έχει 5.034 χρήστες στα δικά μας δεδομένα, αντιπροσωπεύοντας περίπου το 14,9% του συνολικού μας δείγματος. Η δημοσίευση περιελάμβανε 3.572 χρήστες από το Λονδίνο, που αντιστοιχούν περίπου στο 5,9% του συνολικού δείγματός τους, ενώ τα δεδομένα μας περιλαμβάνουν 2.626 χρήστες από το Λονδίνο, που αντιστοιχούν περίπου στο 7,8% του συνολικού δείγματός μας. Η παγκόσμια κατηγορία της δημοσίευσης είχε 49.146 χρήστες, που αντιστοιχούν περίπου στο 81,1% του συνολικού δείγματος. Στα δικά μας δεδομένα, έχουμε 26.165 χρήστες παγκοσμίως, που αντιστοιχούν περίπου στο 77,3% του συνολικού μας δείγματος. Οι συγκρίσεις αυτές αναδεικνύουν τις διαφορές στον αριθμό των χρηστών και την κατανομή τους σε διάφορες τοποθεσίες μεταξύ των

δεδομένων του άρθρου και των δικών μας δεδομένων. Λαμβάνοντας υπόψη το ποσοστό των χρηστών από κάθε τοποθεσία, μπορούμε να εκτιμήσουμε τη σχετική εκπροσώπηση αυτών των τοποθεσιών στα δείγματα.

2.3.3 Μεταβλητές

Οι ακόλουθες μεταβλητές υπολογίστηκαν για ολόκληρο το εύρος ημερομηνιών των δεδομένων, το οποίο κυμαίνεται από το 2016 έως το 2022. Αυτό το εκτεταμένο χρονικό πλαίσιο επέτρεψε την ενδελεχή ανάλυση της συμπεριφοράς και των συναισθημάτων των χρηστών για μεγάλο χρονικό διάστημα. Λαμβάνοντας υπόψη το πλήρες φάσμα των tweets και των δραστηριοτήτων των χρηστών, η ανάλυση περιλαμβάνει ένα ευρύτερο πλαίσιο και παρέχει μια πιο ολοκληρωμένη κατανόηση των τάσεων, των μοτίβων και της δυναμικής του συνόλου δεδομένων. Οι μεταβλητές που υπολογίστηκαν χρησιμοποιήθηκαν στη συνέχεια στη φάση της ανάλυσης για την εξαγωγή ουσιαστικών πληροφοριών και την εξαγωγή συμπερασμάτων σχετικά με τη συμπεριφορά, τα συναισθήματα και τη δέσμευση των χρηστών του Twitter.

Πίνακας 2.2 Ονομασίες και Τύποι Μεταβλητών

Αριθμός Ακόλουθων	Ακέραιος αριθμός
Αριθμός Ακολουθούμενων	Ακέραιος αριθμός
Συνολικός αριθμός tweets	Ακέραιος αριθμός
Μέσος όρος Retweet	Δεκαδικός αριθμός [0,1]
Μέσος όρος συναισθηματικής κατάστασης	Δεκαδικός αριθμός [-1,1]
Μέσος όρος αριθμών hashtags	Δεκαδικός αριθμός [0,1]
Επιχειρηματίας	Δυαδική τιμή 1 ή 0

Αριθμός Ακόλουθων (NumberOfFollowers)

Το στοιχείο αυτό πάρθηκε απευθείας μαζί με τα υπόλοιπα στοιχεία του χρήστη και υποδεικνύει τον αριθμό των χρηστών που ακολουθούν τον χρήστη στο προσωπικό του προφίλ στο Twitter. Παρέχει πληροφορίες σχετικά με την εμβέλεια και τη δημοτικότητα

των χρηστών του Twitter σε κάθε τοποθεσία. Υψηλότερες τιμές υποδηλώνουν μεγαλύτερο αριθμό ακολουθών, γεγονός που υποδηλώνει ενδεχομένως σημαντικούς ή δημοφιλείς χρήστες.

Αριθμός Ακολουθούμενων (NumbersOfFollowings)

Επίσης πάρθηκε μέσα από το Twitter API και υποδεικνύει τον αριθμό των χρηστών που ακολουθεί ο χρήστης. Αποκαλύπτει πληροφορίες σχετικά με τα ενδιαφέροντα και τις συνδέσεις του χρήστη. Υψηλότερες τιμές υποδηλώνουν περισσότερες συνδέσεις χρηστών ή ευρύτερο φάσμα ενδιαφερόντων.

Μέσος όρος Retweet (avgRetweet)

Αυτή είναι μια μεταβλητή που δημιουργήθηκε παίρνοντας από όλα τα tweets που έχει δημοσίευση ο συγκεκριμένος χρήστης, την πληροφορία εάν έγινε Retweet. Ένα tweet θεωρείται ως retweet όταν ένας χρήστης στέλνει ξανά ένα μήνυμα που στάλθηκε αρχικά στο Twitter από άλλο χρήστη. Αντιπροσωπεύει το βαθμό εμπλοκής και δημοτικότητας των tweets. Υψηλότερες τιμές υποδηλώνουν ότι τα tweets είναι πιο πιθανό να κοινοποιηθούν και να διαδοθούν.

Μέσος όρος συναισθηματικής κατάστασης (avgSentiment)

Για το μέσο όρο της συναισθηματικής κατάστασης του κάθε χρήστη, χρησιμοποιήθηκαν όλα τα δημοσιευμένα tweets του και ο υπολογισμός του μέσου όρου της συναισθηματικής κατάστασης του tweet. Μας βοηθά να κατανοήσουμε το γενικό συναίσθημα ή τη διάθεση που εκφράζεται στα tweets. Οι υψηλότερες τιμές υποδηλώνουν ότι ο χρήστης είναι πιο θετικός.

Μέσος όρος αριθμών hashtags (avgHashTag)

Αυτή είναι μια μεταβλητή που δημιουργήθηκε παίρνοντας από όλα τα tweets που έχει δημοσίευση ο συγκεκριμένος χρήστης, την πληροφορία πόσα hashtags χρησιμοποιήθηκαν. Η μεταβλητή υποδεικνύει τον συνολικό αριθμό των hashtag διαιρεμένο με τον αριθμό των tweets ανά χρήστη. Συγκεκριμένα, ένα hashtag είναι μια

σύντομη λέξη ή φράση που μπορούν να αναζητήσουν άλλοι χρήστες μέσω της διεπαφής Twitter για να βρίσκουν εύκολα μηνύματα που φέρουν το hashtag. Οι χρήστες χρησιμοποιούν hashtags για να αυξήσουν την ορατότητα των tweet τους.

Επιχειρηματίας (isEntre)

Οι επιχειρηματίες προσδιορίζονται ως χρήστες που έχουν στην προσωπική τους περιγραφή στο Twitter οποιονδήποτε από τους ακόλουθους όρους: “entrepreneur”, “founder”, “co-founder”, “business-owner”, “business owner”, “start-up”, or “start up”. Τα υπόλοιπα προφίλ χρηστών από το Twitter ταξινομήθηκαν ως μη επιχειρηματίες.

Συνολικός αριθμός tweets (NumberOfTweets)

Αυτός είναι ο αριθμός των tweet που έχει δημοσιεύσει ένας χρήστης. Ο αριθμός των tweets που στέλνει ένας χρήστης υποδεικνύει το επίπεδο δραστηριότητας και εμπλοκής του στην πλατφόρμα. Οι χρήστες που έχουν περισσότερα tweets είναι πιο πιθανό να συμμετέχουν και να έχουν μεγαλύτερο ψηφιακό αποτύπωμα. Αυτή η μεταβλητή μπορεί να αποκαλύψει πληροφορίες σχετικά με τη συμβολή του χρήστη στη συζήτηση καθώς και τη δυνατότητα επιρροής του.

Πίνακας 2.3 Πίνακας περιγραφικών στατιστικών και συσχετίσεων ανά τοποθεσία. Με μαυρισμένα η συσχετίσεις μεγαλύτερες του 0.2 και 0.5 όπου θεωρούνται αδύναμες και ισχυρές αντίστοιχα.

Λονδίνο

Variables	mean	std	(1)	(2)	(3)	(4)	(5)	(6)
NumberOfFollowers (1)	3987.61386	24539.44127						
NumbersOfFollowings (2)	1404.00647	2804.089376	0.167406					
NumberOfTweets (3)	12438.1428	26698.57718	0.147651	0.27526				
isEntre (4)	0.129094	0.335367	0.048719	0.02674	0.046549			
avgSentiment (5)	0.204292	0.125323	0.00387	0.074871	0.135861	0.11542		
avgRetweet (6)	0.307528	0.23466	0.003411	0.040919	0.058557	0.014918	0.021167	
avgHashTag (7)	0.425257	0.516171	0.028754	0.003656	0.131088	0.047614	0.219714	0.033954

Παγκόσμια

Variables	mean	std	(1)	(2)	(3)	(4)	(5)	(6)
NumberOfFollowers (1)	6152.72	145094.00						
NumbersOfFollowings (2)	1278.86822	4751.380438	0.052478					
NumberOfTweets (3)	9755.6232	26220.46457	0.140497	0.227585				
isEntre (4)	0.233251	0.422908	0.004088	0.012987	0.059531			
avgSentiment (5)	0.189087	0.150918	0.004911	0.009866	0.101476	0.074979		
avgRetweet (6)	0.284895	0.26562	0.010848	0.001856	0.065176	0.021335	0.040422	
avgHashTag (7)	0.361635	0.484841	0.009301	0.014795	-0.0403	0.021111	0.108003	0.0258

Λος Άντζελες

Variables	mean	std	(1)	(2)	(3)	(4)	(5)	(6)
NumberOfFollowers (1)	12352.22	64995.55						
NumbersOfFollowings (2)	4799.17799	26471.61189	0.534326					
NumberOfTweets (3)	15351.4805	37510.49284	0.258558	0.227784				
isEntre (4)	0.153953	0.36094	0.012266	0.046821	0.020971			
avgSentiment (5)	0.181705	0.142537	0.051672	0.047522	0.072589	0.111825		
avgRetweet (6)	0.292374	0.258803	0.043377	0.048328	0.030282	0.079018	0.124711	
avgHashTag (7)	0.393333	0.535356	0.003799	0.018786	0.046784	0.0875	0.216889	0.118189

Ο πίνακας περιγραφικών στατιστικών και συσχετίσεων παρέχει μια εις βάθος ματιά στους παράγοντες που μετρήθηκαν σε τρεις μεγάλες τοποθεσίες: Λονδίνο, Παγκόσμια και Λος Άντζελες. NumberOfFollowers, NumbersOfFollowings, NumberOfTweets, isEntre, avgSentiment, avgRetweet και avgHashTag είναι μεταξύ των μεταβλητών. Εμφανίζονται ο μέσος όρος και η τυπική απόκλιση κάθε μεταβλητής, καθώς και οι συσχετίσεις ανά ζεύγη μεταξύ τους. Ο μέσος όρος, αντιπροσωπεύει την κεντρική τάση των δεδομένων και παρέχει ένα συνολικό μέτρο της τυπικής ή μέσης τιμής του συνόλου δεδομένων. Ενώ η τυπική απόκλιση, από την άλλη πλευρά, είναι ένα μέτρο της διασποράς ή της μεταβλητότητας των σημείων δεδομένων γύρω από τη μέση τιμή.

Μετρά πόσο πολύ αποκλίνουν τα μεμονωμένα σημεία δεδομένων από τη μέση τιμή. Μια μεγαλύτερη τυπική απόκλιση υποδηλώνει ένα ευρύτερο εύρος τιμών, ενώ μια μικρότερη τυπική απόκλιση υποδηλώνει ότι τα σημεία δεδομένων βρίσκονται πιο κοντά στη μέση τιμή. Η συσχέτιση εκφράζει το βαθμό στον οποίο οι μεταβολές σε μια μεταβλητή σχετίζονται με τις μεταβολές σε μια άλλη. Η συσχέτιση είναι χρήσιμη στην έρευνά μας διότι μας επιτρέπει να αξιολογήσουμε την αλληλεξάρτηση και τα μοτίβα μεταξύ των μεταβλητών στο σύνολο των δεδομένων μας. Μπορούμε να προσδιορίσουμε την ισχύ και την κατεύθυνση των σχέσεων, αν είναι θετικές, αρνητικές ή ουδέτερες, υπολογίζοντας τους συντελεστές συσχέτισης. Τα δεδομένα αυτά είναι κρίσιμα για την κατανόηση των σχέσεων και των επιρροών μεταξύ των μεταβλητών, τον εντοπισμό πιθανών αιτιωδών σχέσεων και την απόκτηση γνώσεων σχετικά με την υποκείμενη δυναμική των δεδομένων. Η ανάλυση συσχέτισης μας βοηθά στην αποκάλυψη κρυφών μοτίβων, στην επικύρωση υποθέσεων και στη λήψη τεκμηριωμένων αποφάσεων με βάση τις παρατηρούμενες σχέσεις στην έρευνά μας.

Η συσχέτιση μεταξύ NumberOfFollowers και NumbersOfFollowings, για το δείγμα στο Λος Άντζελες είναι η μόνη όπου θεωρείται ισχυρή και δείχνει πώς όσοι έχουν περισσότερους ακόλουθους τείνουν να ακολουθούν περισσότερους ανθρώπους οι ίδιοι. Για τις υπόλοιπες συσχετίσεις από το πιο πάνω πίνακα μπορούμε να πούμε ότι είναι ουδέτερες ή αδύναμες.

2.3.4 Προ-επεξεργασία

Πραγματοποιήσαμε διάφορα βήματα προ επεξεργασίας για να καθαρίσουμε και να προετοιμάσουμε τα δεδομένα πριν από τη διεξαγωγή της ανάλυσής μας. Αρχικά, αφαιρέσαμε τις διευθύνσεις URL από το κείμενο για να αφαιρέσουμε τυχόν διευθύνσεις ιστού που δεν ήταν σχετικές με την έρευνά μας. Στη συνέχεια, αφαιρέσαμε τους αριθμούς και το αριθμητικό κείμενο για να αποφύγουμε την παρεμβολή στην ανάλυση των δεδομένων μας.

Αφαιρέσαμε επίσης κοινά σύμβολα όπως τα σύμβολα ισότητας (=), ενώ διατηρήσαμε σημαντικά σημεία στίξης όπως ερωτηματικά, θαυμαστικά και τελείες. Αυτό μας

επέτρεψε να διατηρήσουμε την ακεραιότητα του κειμένου, αφαιρώντας παράλληλα τον περιττό θόρυβο.

Επιπλέον, για βελτίωση του κείμενου και να διασφαλίσουμε τη συνοχή σε όλη την ανάλυσή μας, αφαιρέσαμε σύμβολα ή κείμενο ορολογίας, όπως οι σύνδεσμοι (&). Επιπλέον, αντικαταστήσαμε τις περιπτώσεις τριών ή περισσότερων διπλών χαρακτήρων εντός των λέξεων για να διασφαλίσουμε τη συνοχή. Για παράδειγμα, το "tooool" θα αντικατασταθεί σε "tool" για να διασφαλιστεί η σωστή ανάλυση.

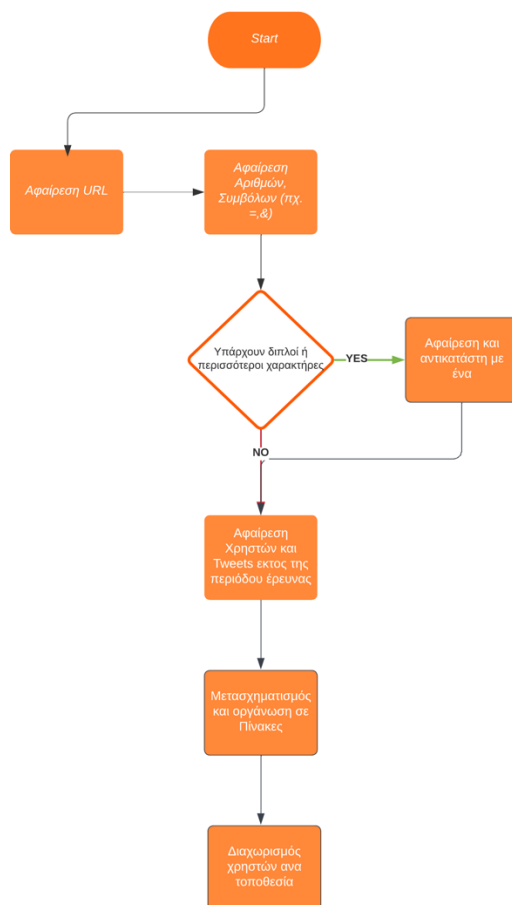
Επίσης, προσθέσαμε ένα βήμα για τον καθαρισμό του συνόλου δεδομένων αφαιρώντας χρήστες και tweets που έπεφταν εκτός του καθορισμένου χρονικού πλαισίου.

Πραγματοποιήσαμε μετασχηματισμό δεδομένων εκτός από τα βήματα προ επεξεργασίας για να διευκολύνουμε την ανάλυσή μας. Οργανώσαμε τα καθαρισμένα δεδομένα σε κατάλληλους πίνακες, διασφαλίζοντας ότι όλες οι σχετικές μεταβλητές και τα χαρακτηριστικά ήταν σωστά δομημένα. Αυτό μας επέτρεψε να έχουμε πρόσβαση και να χειριζόμαστε τα δεδομένα πιο αποτελεσματικά κατά τη φάση της ανάλυσης.

Με τη δημιουργία αυτών των πινάκων θέσαμε μια σταθερή βάση για την ανάλυσή μας. Παρείχαν μια δομημένη και συστηματική προσέγγιση στην οργάνωση των δεδομένων και βοήθησαν τις προσπάθειές μας να εξάγουμε σημαντικές πληροφορίες. Χάρη στους πίνακες μπορέσαμε εύκολα να εξερευνήσουμε και να αναλύσουμε τα δεδομένα, να εντοπίσουμε μοτίβα και να εξάγουμε συμπεράσματα με βάση τις παρατηρούμενες τάσεις.

Πραγματοποιήσαμε επίσης ταξινόμηση των χρηστών με βάση τις διαθέσιμες πληροφορίες κατά το στάδιο της προ επεξεργασίας, ιδίως στις τρεις τοποθεσίες ενδιαφέροντος: Λος Άντζελες, Λονδίνο και σε όλο τον κόσμο. Ωστόσο, είναι σημαντικό να σημειωθεί ότι ένας περιορισμός που αντιμετωπίσαμε ήταν η έλλειψη ακριβών δεδομένων τοποθεσίας του Twitter. Ενώ θα μπορούσαμε να βασιστούμε αποκλειστικά στην περιγραφή του προφίλ των χρηστών και σε άλλα δημόσια διαθέσιμα δεδομένα, αυτή η μέθοδος μπορεί να μην αντικατοπτρίζει πάντα με ακρίβεια την πραγματική τοποθεσία ενός χρήστη. Χρησιμοποιήσαμε δεδομένα από την προηγούμενη μελέτη που παρείχε πληροφορίες για την τοποθεσία για να ξεπεράσουμε αυτόν τον περιορισμό και να βελτιώσουμε την ακρίβεια της ταξινόμησής μας. Με τη χρήση αυτής

της προσέγγισης μπορέσαμε να αξιοποιήσουμε την υπάρχουσα γνώση και να επικυρώσουμε την ακρίβεια της ταξινόμησης θέσης για τους ερευνητικούς μας σκοπούς.



2.3.5 Προκλήσεις ή Περιορισμοί

Αντιμετωπίσαμε έναν περιορισμό κατά την ανάλυση του θέματος των tweets επειδή χρησιμοποιούσαμε μια εμπορική εφαρμογή με ημερήσιο όριο 100.000 αναλύσεων. Αυτό αποτελούσε ένα εμπόδιο, αλλά καταφέραμε να αναλύσουμε ένα εκατομμύριο tweets και να αποκτήσουμε σημαντικές γνώσεις που βοήθησαν πάρα πολύ την έρευνά μας.

Η ανάλυση ενός εκατομμυρίου tweets ήταν μια χρονοβόρα και απαιτητική εργασία. Έπρεπε να επιλέξουμε προσεκτικά τα tweets που θέλαμε να μελετήσουμε παραμένοντας εντός του ημερήσιου ορίου της εμπορικής εφαρμογής. Παρά αυτά τα εμπόδια, ήμασταν αποφασισμένοι να αναλάβουμε μια ολοκληρωμένη έρευνα και να συγκεντρώσουμε όσο το δυνατόν περισσότερες πληροφορίες.

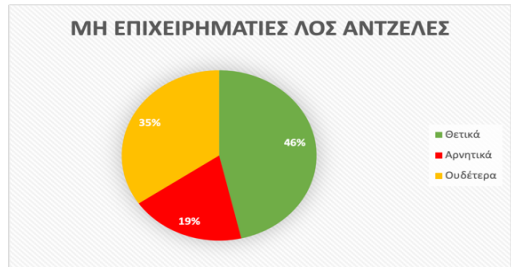
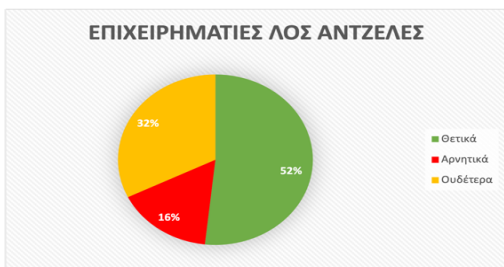
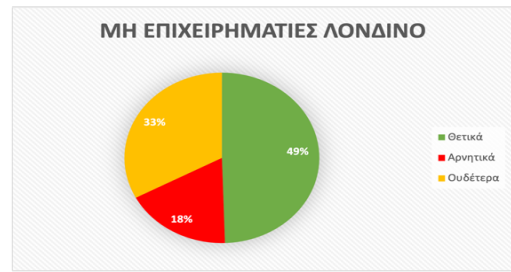
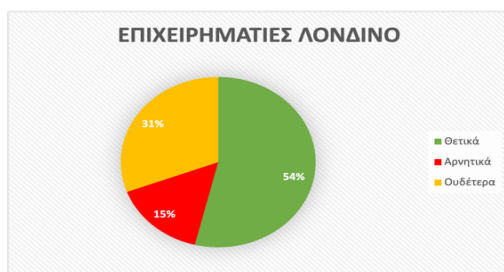
Κεφάλαιο 3

Αποτελέσματα

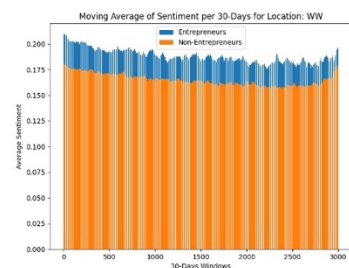
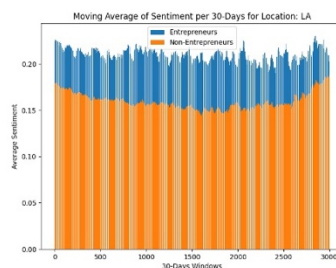
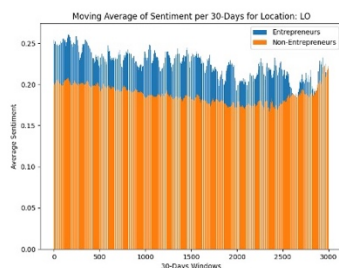
3.1 Σύγκριση Επιχειρηματιών με Μη-Επιχειρηματιών	27
3.2 Ανάλυση Θεμάτων	30
3.3 Σύγκριση αποτελεσμάτων με προηγούμενο άρθρο	34
3.4 Κορονοϊός (COVID-19)	35
3.4.1 Ορόσημα	36
3.4.2 Περίοδοι Έρευνας	37
3.4.3 Αποτελέσματα	39
3.4.4 Σύγκριση με μέσου συναισθήματος με ημερήσιο αριθμών κρουσμάτων	40
3.4.5 Διερεύνηση σημαντικής πτώσης του συναισθήματος	42

3.1 Σύγκριση Επιχειρηματιών με Μη-Επιχειρηματιών

Με βάση τα αποτελέσματα που παρουσιάστηκαν στην εικόνα 3.1, φαίνεται ότι τόσο οι επιχειρηματίες όσο και οι μη επιχειρηματίες στην πόλη του Λονδίνου έχουν γενικά θετικό συναίσθημα στο Twitter. Ωστόσο, η ανάλυση συναισθήματος αποκαλύπτει ότι οι επιχειρηματίες έχουν υψηλότερο ποσοστό θετικού συναισθήματος (54%) σε σύγκριση με τους μη επιχειρηματίες (49%). Επιπρόσθετα τα αποτελέσματα υποδηλώνουν ότι τόσο οι επιχειρηματίες όσο και οι μη επιχειρηματίες εκφράζουν θετικό συναίσθημα στο Twitter σε παγκόσμιο επίπεδο, με το 50% των επιχειρηματιών και το 48% των μη επιχειρηματιών να εκφράζουν θετικά συναισθήματα. Όπως και με βάση τα αποτελέσματα για το Λος Άντζελες, οι επιχειρηματίες έχουν υψηλότερο ποσοστό θετικού συναισθήματος στο Twitter (52%) σε σύγκριση με τους μη επιχειρηματίες (46%). Αυτό δείχνει ότι οι επιχειρηματίες μπορεί να είναι πιο πιθανό να εκφράσουν θετικές απόψεις ή σκέψεις στο Twitter.



Εικόνα 3.1 Μέσο Συναισθήμα για την χρονική περίοδο 2016-2022, Επιχειρηματιών και Μη-Επιχειρηματιών ανά τοποθεσία



Από την άλλη πλευρά, οι μη επιχειρηματίες έχουν υψηλότερο ποσοστό αρνητικού συναισθήματος (18%) σε σύγκριση με τους επιχειρηματίες (15%) στο Λονδίνο, το ίδιο μπορούμε να δούμε και στις δύο άλλες τοποθεσίες. Παγκόσμια οι μη επιχειρηματίες έχουν ποσοστό αρνητικού συναισθήματος (18%) σε σύγκριση με τους επιχειρηματίες (16%) και οι μη επιχειρηματίες στο Λος Άντζελες έχουν υψηλότερο ποσοστό (19%) σε σύγκριση με τους επιχειρηματίες (16%). Αυτό υποδηλώνει ότι οι μη επιχειρηματίες μπορεί να είναι πιο πιθανό να εκφράσουν αρνητικές απόψεις ή σκέψεις στο Twitter. Είναι σημαντικό να σημειωθεί, ωστόσο, ότι και οι δύο ομάδες έχουν παρόμοιο ποσοστό ουδέτερου συναισθήματος, με 34% τόσο για τους επιχειρηματίες όσο και για τους μη

επιχειρηματίες για το Παγκόσμιο δείγμα, 31% για τους επιχειρηματίες και 33% για τους μη επιχειρηματίες στο Λονδίνο και 32% για τους επιχειρηματίες και 35% για τους μη επιχειρηματίες για την τοποθεσία Λος Άντζελες. Αυτό υποδηλώνει ότι ένα σημαντικό ποσοστό των tweets και από τις δύο ομάδες έχουν ουδέτερο τόνο ή εκφράζουν έλλειψη συναισθήματος όπως και στις άλλες δυο περιοχές.

Συγκρίνοντας τα αποτελέσματα του συναισθήματος και στις τρεις τοποθεσίες, οι επιχειρηματίες τείνουν να εκφράζουν ένα γενικά πιο θετικό συναίσθημα στο Twitter, με υψηλότερο ποσοστό θετικού συναισθήματος σε σύγκριση με τους μη επιχειρηματίες. Ωστόσο, υπάρχει μια διαφορά στο ποσοστό του αρνητικού συναισθήματος που εκφράζεται από μη επιχειρηματίες σε σύγκριση με τους επιχειρηματίες και στις τρεις τοποθεσίες. Επιπλέον, και οι δύο ομάδες έχουν παρόμοιο ποσοστό ουδέτερου συναισθήματος.

Πίνακας 3.2 Γραμμικής Παλινδρόμησης για όλα τα δεδομένα. Τα μαυρισμένα είναι οι τιμές όπου δεν θεωρούνται σημαντικές, αφού είναι μεγαλύτερες από $\alpha=0.05$, $p\text{-value}>\alpha$, αποτυγχάνουν να απορρίψουν την μηδενική υπόθεση.

	Λονδίνο	τιμή-p	τυπικό σφάλμα	Λος Άντζελες	τιμή-p	τυπικό σφάλμα	Παγκόσμια	τιμή-p	τυπικό σφάλμα
Intercept	0.1822	0.000	0.005	0.173	0.000	0.004	0.182	0.000	0.002
isEntre	0.0361	0.000	0.007	0.0333	0.000	0.005	0.0235	0.000	0.002
avgHashTag	0.0479	0.000	0.005	0.0519	0.000	0.004	0.0321	0.000	0.002
avgRetweet	-0.0067	0.507	0.01	-0.0491	0.000	0.008	-0.02	0.000	0.003
NumberOfFollowers	3.36E-08	0.731	9.77E-08	1.21E-07	0.178	3.58E-08	8.62E-09	0.001	6.41E-09
NumbersOfFollowings	4.91E-06	0.000	8.80E-07	1.30E-07	0.000	8.72E-08	9.42E-07	0.136	1.99E-07
NumberOfTweets	-6.38E-07	0.000	9.30E-08	-2.99E-07	0.000	5.39E-08	-5.70E-07	0.000	3.65E-08
No. Observations:	2626			5034			26165		
R-squared:	0.082			0.073			0.028		
F-statistic:	38.74			65.53			124.7		
Prob (F-statistic):	2.53E-45			1.06E-78			4.82E-156		

Η πολλαπλή γραμμική παλινδρόμηση είναι μια αποτελεσματική στατιστική τεχνική για την εξέταση της σχέσης μεταξύ πολλαπλών ανεξάρτητων μεταβλητών και μιας ενιαίας εξαρτημένης μεταβλητής. Η πολλαπλή γραμμική παλινδρόμηση χρησιμοποιήθηκε στη μελέτη μας για τη διερεύνηση της σχέσης μεταξύ διαφόρων παραγόντων και του συναισθήματος που εκφράζεται στα μέσα κοινωνικής δικτύωσης από επιχειρηματίες και μη επιχειρηματίες.

Το μοντέλο παλινδρόμησης με k ανεξάρτητες μεταβλητές $X_1, X_2, \dots, X_k$ θα έχει τη μορφή
$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \epsilon_i \quad i=1,2,\dots,n$$

όπου

- Y_i είναι η τιμή της εξαρτημένης μεταβλητής για την i παρατήρηση, συγκεκριμένα το μέσο συναίσθημα.
- $X_{i1}, X_{i2}, \dots, X_{ik}$ είναι οι τιμές των ανεξάρτητων μεταβλητών στην i παρατήρηση, συγκεκριμένα isEntre (εάν ο χρήστης είναι επιχειρηματίας), μέσος αριθμός hashtags, μέσος αριθμός retweets, αριθμός ακόλουθων, αριθμός ακολουθουμένων, αριθμός tweets

Σκοπός της μελέτης αυτής ήταν να δούμε αν οι επιχειρηματίες είναι πιο πιθανό από τους μη επιχειρηματίες να εκφράσουν θετικό συναίσθημα. Ακολουθήσαμε μια κλασική διαδικασία για τη διεξαγωγή της ανάλυσης πολλαπλής γραμμικής παλινδρόμησης. Αρχικά, αντλήσαμε τα απαραίτητα δεδομένα από τη βάση δεδομένων MongoDB και υπολογίστηκαν οι σχετικές μεταβλητές όπως η μέση χρήση hashtag (avgHashTag) και ο μέσος αριθμός retweet (avgRetweet) για κάθε χρήστη με βάση τα δεδομένα των tweets του. Στη συνέχεια, η ανάλυση παλινδρόμησης πραγματοποιήθηκε με τη χρήση του μοντέλου ελαχίστων τετραγώνων (OLS) από τη βιβλιοθήκη statsmodels της Python. Οι συντελεστές και η στατιστική σημαντικότητα κάθε ανεξάρτητης μεταβλητής σε σχέση με την εξαρτημένη μεταβλητή, το συναίσθημα, εκτιμήθηκαν με τη χρήση αυτού του μοντέλου.

Τα αποτελέσματα της ανάλυσης πολλαπλής παλινδρόμησης των δεδομένων Twitter για τρεις διαφορετικές γεωγραφικές τοποθεσίες—Λονδίνο, Λος Άντζελες και ολόκληρο τον κόσμο—εμφανίζονται στον πίνακα 3.2.

Όταν όλοι οι άλλοι παράγοντες διατηρούνται σταθεροί, ο όρος τομής αντιπροσωπεύει την αναμενόμενη μέση διάθεση. Το μέσο συναίσθημα επηρεάζεται από τις μεταβλητές "isEntre", "avgHashTag", "avgRetweet", "NumberOfFollowers", "NumbersOfFollowings" και "NumberOfTweets", σύμφωνα με τις εκτιμήσεις των συντελεστών για αυτές τις μεταβλητές. Τα ευρήματα δείχνουν ότι το να είσαι επιχειρηματίας, έχει αδύναμο αλλά σημαντικό ευεργετικό αντίκτυπο στη γενική ευημερία των ανθρώπων.

Οι χαμηλές τιμές R-τετράγωνο υποδηλώνουν ότι το μοντέλο λαμβάνει υπόψη μόνο εν μέρει τη διακύμανση στο μέσο συναίσθημα. Το μοντέλο στο σύνολό του είναι στατιστικά σημαντικό, σύμφωνα με τη στατιστική F.

3.2 Ανάλυση Θεμάτων

Στόχος μας ήταν να αποκτήσουμε μια εικόνα για τις στάσεις και τα συναισθήματα των επιχειρηματιών και των μη επιχειρηματιών απέναντι σε διαφορετικά θέματα, και πώς συγκρίνονται μεταξύ τους, χρησιμοποιώντας τεχνικές ανάλυσης θεμάτων. Θέλαμε συγκεκριμένα να διερευνήσουμε εάν οι ιδιοκτήτες εταιρειών είναι πιο πιθανό να αισθάνονται θετικά για συγκεκριμένα θέματα όταν δεν μιλούν για τις εταιρείες τους. Ο στόχος αυτής της ανάλυσης ήταν να μας δώσει μια βαθύτερη γνώση των πιθανών διαφορών μεταξύ των συναισθηματικών χαρακτηριστικών των επιχειρηματιών και των μη επιχειρηματιών.

Αρχικά, χρησιμοποιήσαμε το εργαλείο uClassify για να εκτελέσουμε ανάλυση θεμάτων σε ένα τυχαία επιλεγμένο υποσύνολο του μεγάλου συνόλου δεδομένων μας, που αποτελείται από ένα εκατομμύριο tweets. Καταφέραμε να κατηγοριοποιήσουμε τα tweets χρησιμοποιώντας αυτή την ανάλυση σε δέκα διαφορετικές κατηγορίες: Τέχνες, Επιχειρήσεις, Υπολογιστές, Παιχνίδια, Υγεία, Σπίτι, Ψυχαγωγία, Επιστήμη, Κοινωνία και Αθλητισμός.

Χωρίσαμε τα tweets σε δύο ομάδες αφού τα κατηγοριοποιήσαμε σε διαφορετικά θέματα: αυτά που δημοσιεύτηκαν από επιχειρηματίες και αυτά που δημοσιεύτηκαν από μη επιχειρηματίες. Αυτός ο διαχωρισμός ήταν απαραίτητος για τη σύγκριση των συναισθημάτων που εκφράστηκαν από κάθε ομάδα για συγκεκριμένα θέματα.

Χρησιμοποιήσαμε t-tests δύο δειγμάτων για να δούμε αν υπήρχαν σημαντικές διαφορές στα συναισθήματα μεταξύ επιχειρηματιών και μη επιχειρηματιών σε κάθε θέμα. Τα t-test δύο δειγμάτων ήταν κατάλληλα για τη σύγκριση μέσων όρων μεταξύ δύο ανεξάρτητων ομάδων, στην προκειμένη περίπτωση επιχειρηματιών και μη επιχειρηματιών.

Η σημασία της χρήσης αυτών των t-tests δύο δειγμάτων προέκυψε από την ικανότητά τους να εντοπίζουν διαφορές, συγκεκριμένα στα συναισθήματα που εκφράζουν οι επιχειρηματίες και οι μη επιχειρηματίες σε διάφορα θέματα. Συγκρίνοντας τα συναισθήματα, μάθαμε αν οι επιχειρηματίες είχαν πιο θετική ή αρνητική στάση απέναντι σε συγκεκριμένα θέματα από ό,τι οι μη επιχειρηματίες.

Ο απώτερος στόχος αυτής της ανάλυσης ήταν να κατανοήσουμε καλύτερα τις πιθανές διαφορές μεταξύ επιχειρηματιών και μη επιχειρηματιών σε ψυχολογικά και συμπεριφορικά χαρακτηριστικά. Θέλαμε να δούμε αν οι επιχειρηματίες είχαν ξεχωριστές συναισθηματικές αντιδράσεις και στάσεις απέναντι σε διάφορα θέματα εξετάζοντας τα συναισθήματα που εκφράστηκαν από αυτές τις ομάδες, αντί για μια απλή γενικευμένη σύγκριση.

Συνοπτικά, χρησιμοποιήθηκαν t-tests δύο δειγμάτων για τη διερεύνηση και τη σύγκριση των συναισθημάτων που εκφράστηκαν από επιχειρηματίες και μη επιχειρηματίες για συγκεκριμένα θέματα.

Πίνακας 3.3.1 T-tests Δυο δειγμάτων για την τοποθεσία Λονδίνο, διάρκεια 2016-2022

Θέματα	Tweets Μη-Επιχειρηματιών			Tweets Επιχειρηματιών			t	p-value
	N	Mean	SD	N	Mean	SD		
Τέχνες	1782	0.204	0.405	470	0.245	0.384	1.972	0.049
Επιχειρήσεις	427	0.204	0.358	145	0.25	0.35	1.356	0.176
Υπολογιστές	1143	0.188	0.378	332	0.239	0.355	2.173	0.03
Παιχνίδια	282	0.122	0.395	71	0.203	0.363	1.571	0.117
Υγεία	711	0.156	0.422	195	0.176	0.42	0.581	0.562
Σπίτι	350	0.173	0.39	74	0.285	0.357	2.265	0.024
Αναψυχή	475	0.19	0.368	129	0.207	0.364	0.467	0.641
Επιστήμη	351	0.169	0.39	101	0.124	0.36	-1.034	0.302
Κοινωνία	955	0.087	0.422	241	0.111	0.424	0.792	0.429
Αθλητισμός	581	0.229	0.384	243	0.3	0.342	2.501	0.013

Πίνακας 3.3.2 T-tests Δυο δειγμάτων για την τοποθεσία Παγκόσμια, διάρκεια 2016-2022

Θέματα	Tweets Μη-Επιχειρηματιών			Tweets Επιχειρηματιών			t	p-value
	N	Mean	SD	N	Mean	SD		
Τέχνες	155014	0.216	0.414	35776	0.245	0.407	12.203	0
Επιχειρήσεις	41678	0.22	0.359	16684	0.233	0.358	3.902	0
Υπολογιστές	96453	0.182	0.365	30312	0.199	0.364	7.125	0
Παιχνίδια	27777	0.148	0.396	6507	0.168	0.391	3.708	0
Υγεία	67264	0.131	0.426	16604	0.149	0.431	4.702	0
Σπίτι	31451	0.212	0.375	7571	0.23	0.37	3.859	0
Αναψυχή	40928	0.186	0.378	10024	0.2	0.377	3.229	0.001
Επιστήμη	27196	0.138	0.366	8053	0.164	0.371	5.426	0
Κοινωνία	102692	0.055	0.45	25281	0.083	0.442	8.898	0
Αθλητισμός	60945	0.202	0.39	15490	0.238	0.386	10.198	0

Πίνακας 3.3.3 T-tests Δυο δειγμάτων για την τοποθεσία Λος Άντζελες, διάρκεια 2016-2022

Θέματα	Tweets Μη-Επιχειρηματιών			Tweets Επιχειρηματιών			t	p-value
	N	Mean	SD	N	Mean	SD		
Τέχνες	38384	0.206	0.423	6247	0.245	0.413	6.897	0
Επιχειρήσεις	6259	0.214	0.359	2059	0.231	0.361	1.841	0.066
Υπολογιστές	11439	0.188	0.362	2754	0.202	0.362	1.803	0.071
Παιχνίδια	5666	0.125	0.395	912	0.173	0.392	3.439	0.001
Υγεία	15230	0.127	0.434	2463	0.177	0.438	5.282	0
Σπίτι	7913	0.224	0.382	1450	0.278	0.376	4.932	0
Αναψυχή	9774	0.193	0.385	1891	0.206	0.385	1.394	0.163
Επιστήμη	4417	0.128	0.362	922	0.157	0.386	2.148	0.032
Κοινωνία	19586	0.09	0.458	3774	0.089	0.458	-0.13	0.897
Αθλητισμός	17211	0.183	0.39	2891	0.229	0.388	5.951	0

Με λίγες εξαιρέσεις στο σύνολο δεδομένων του Λονδίνου, τα αποτελέσματα των t-τεστ δύο δειγμάτων αποκάλυψαν σημαντικές διαφορές στα συναισθήματα μεταξύ επιχειρηματιών και μη επιχειρηματιών σε διάφορα θέματα και τοποθεσίες.

Τα ευρήματά μας αποκάλυψαν σημαντικές διαφορές στα συναισθήματα που εκφράζουν οι επιχειρηματίες έναντι των μη επιχειρηματιών στην πλειονότητα των θεμάτων και των περιοχών. Τα ευρήματα αυτά υποδεικνύουν την ύπαρξη διακριτών ψυχολογικών και συμπεριφορικών χαρακτηριστικών που σχετίζονται με την επιχειρηματικότητα.

Ωστόσο, υπήρξαν μερικές εξαιρέσεις στο σύνολο δεδομένων του Λονδίνου και του Λος Άντζελες, με μη σημαντικές διαφορές στα συναισθήματα που παρατηρήθηκαν για συγκεκριμένα θέματα, συγκεκριμένα Επιχειρήσεις, Υπολογιστές, Αναψυχή για το δείγμα του Λος Άντζελες και Επιχειρήσεις, Υγεία, Αναψυχή, Επιστήμη, Κοινωνία για το δείγμα του Λονδίνου . Αυτό θα μπορούσε να αποδοθεί σε συγκυριακούς παράγοντες μοναδικούς για την συγκεκριμένη περιοχή ή σε άλλες τοπικές επιρροές που μπορεί να έχουν διαφορετικό αντίκτυπο στα συναισθήματα που εκφράζουν οι επιχειρηματίες και οι μη επιχειρηματίες.

Συνολικά, τα ευρήματά μας παρέχουν εμπειρικές αποδείξεις για τα αποκλίνοντα συναισθηματικά και συμπεριφορικά μοτίβα των επιχειρηματιών και των μη επιχειρηματιών σε διάφορα θέματα και τοποθεσίες.

3.3 Σύγκριση αποτελεσμάτων με προηγούμενο άρθρο

Θα το αντιπαραβάλουμε με τα αποτελέσματα της προηγούμενης ερευνητικής μελέτης, προκειμένου να επικαιροποιήσουμε τα ευρήματα της μελέτης και να αξιολογήσουμε τη συνέπεια των αποτελεσμάτων μέχρι τώρα. Στόχος είναι να διαπιστωθεί αν οι τάσεις που παρατηρήθηκαν στην προηγούμενη μελέτη -όπου οι ιδιοκτήτες επιχειρήσεων ανέφεραν περισσότερα θετικά συναισθήματα στα μέσα κοινωνικής δικτύωσης από ό,τι ο μέσος άνθρωπος- εξακολουθούν να ισχύουν στο σημερινό περιβάλλον.

Η προηγούμενη μελέτη, η οποία διεξήχθη μεταξύ 2013 και 2015, εξέτασε τα συναισθήματα που δημοσιεύονταν στο Twitter από επιχειρηματίες και απλούς ανθρώπους. Συγκέντρωσε tweets από επιχειρηματίες στο Λονδίνο, το Λος Άντζελες και άλλες πόλεις, καθώς και από απλούς ανθρώπους. Οι ερευνητές χρησιμοποίησαν το uClassify για τη θεματική ανάλυση και το εργαλείο VADER για την ανάλυση συναισθήματος. Χρησιμοποιήθηκαν στατιστικές αναλύσεις παλινδρόμησης και t-test δύο δειγμάτων για την εξέταση των ευρημάτων, τα οποία έδειξαν ότι οι επιχειρηματίες ήταν πιο πιθανό να αναφέρουν θετικά συναισθήματα για μια ποικιλία θεμάτων.

Αυτή η ερευνητική μελέτη αναπαρήγαγε τις μεθόδους που χρησιμοποιήθηκαν στην πρώτη μελέτη για να ελεγχθεί η συνέπεια αυτών των αποτελεσμάτων. Ωστόσο, κάλυψε ένα πιο πρόσφατο χρονικό πλαίσιο από το 2016 έως το 2022, προσφέροντας μια πιο επίκαιρη άποψη. Αυτή η μελέτη χρησιμοποίησε τις ίδιες τεχνολογίες, το VADER για την

ανάλυση συναισθήματος και το uClassify για την ανάλυση θεμάτων, ενώ δεν καταφέραμε να αποκτήσουμε πλήρη δεδομένα για όλους τους χρήστες. Τα αποτελέσματα της μελέτης ήταν συνεπή με τα ευρήματα της πρώτης, καθώς έδειξαν ότι οι ιδιοκτήτες επιχειρήσεων ήταν πιο πιθανό να δημοσιεύουν θετικά μηνύματα στα μέσα κοινωνικής δικτύωσης. Τα αποτελέσματα των στατιστικών αναλύσεων που χρησιμοποιήθηκαν και στις δύο έρευνες ήταν παρόμοια, αποδεικνύοντας την εγκυρότητα των συμπερασμάτων. Και στις δύο έρευνες, τα αποτελέσματα των t-tests δύο δειγμάτων και των αναλύσεων παλινδρόμησης εμφάνισαν παρόμοιο επίπεδο συνέπειας.

Η συνέπεια των ευρημάτων δείχνει ότι, παρά το χρονικό διάστημα που μεσολάβησε μεταξύ των δοκιμών, η τάση που εντοπίστηκε στην προηγούμενη μελέτη συνεχίζεται μέχρι σήμερα. Και οι δύο μελέτες δείχνουν ότι οι επιχειρηματίες επιδεικνύουν γενικά πιο θετική συμπεριφορά στα μέσα κοινωνικής δικτύωσης από ό,τι ο μέσος άνθρωπος σε μια σειρά θεμάτων. Αυτή η συνέπεια δείχνει ότι η τάση των επιχειρηματιών να εκφράζουν ευνοϊκές απόψεις σε ιστότοπους κοινωνικής δικτύωσης όπως το Twitter παραμένει σταθερή με την πάροδο του χρόνου, υποστηρίζοντας την εγκυρότητα και την αξιοπιστία των προηγούμενων ευρημάτων.

3.4 Κορονοϊός (COVID-19)

Ο νέος κορονοϊός που ανακαλύφθηκε είναι η αιτία του COVID-19 [5], μιας ιογενούς αναπνευστικής πάθησης. (SARS-CoV-2). Ο ιός ανακαλύφθηκε αρχικά στη Γουχάν της Κίνας τον Δεκέμβριο του 2019 και από τότε εξαπλώθηκε σε όλο τον κόσμο προκαλώντας πανδημία. Τα ήπια έως σοβαρά συμπτώματα του COVID-19 μπορεί να περιλαμβάνουν πυρετό, βήχα και δυσκολία στην αναπνοή. Ο ιός μεταδίδεται κυρίως με σταγονίδια του αναπνευστικού από μολυσμένα άτομα και μπορεί να σταματήσει λαμβάνοντας προφυλάξεις όπως η χρήση μάσκας, η κοινωνική απόσταση και το συχνό πλύσιμο των χεριών. Επιπλέον, έχουν δημιουργηθεί και εγκριθεί εμβόλια για την πρόληψη του COVID-19 για χρήση σε καταστάσεις έκτακτης ανάγκης.

Η απόφαση να χρησιμοποιηθεί η πανδημία COVID-19 ως περίπτωση σύγκρισης μεταξύ επιχειρηματιών και μη επιχειρηματιών είναι σημαντική για διάφορους λόγους. Αρχικά,

η παγκόσμια σημασία της πανδημίας COVID-19 δεν μπορεί να υπερτιμηθεί. Επηρέασε άτομα, επιχειρήσεις και κοινωνίες σε όλο τον κόσμο, καθιστώντας την ένα εύστοχο και σχετικό πλαίσιο για την εξέταση της συναισθηματικής αντίδρασης και του συναισθήματος διαφόρων ομάδων κατά τη διάρκεια μιας κοινής κρίσης.

Επίσης, είχε ως αποτέλεσμα ένα πλήθος παγκόσμιων προκλήσεων, αβεβαιοτήτων και διαταραχών. Μπορούμε να μάθουμε για το πώς τα άτομα, επιχειρηματίες και μη επιχειρηματίες, πλοηγούνται και ανταποκρίνονται σε έκτακτες περιστάσεις μελετώντας το συναίσθημα κατά τη διάρκεια της κρίσης. Προσφέρει μια μοναδική ευκαιρία να διερευνήσουμε τη συναισθηματική ανθεκτικότητα και τις στρατηγικές προσαρμογής που χρησιμοποιούν τα άτομα και οι επιχειρήσεις μπροστά σε τέτοιες έκτακτες προκλήσεις.

Η σύγκριση των συναισθηματικών αντιδράσεων των επιχειρηματιών και των μη επιχειρηματιών κατά τη διάρκεια της πανδημίας COVID-19 παρέχει μια πληρέστερη κατανόηση των συναισθηματικών τους αντιδράσεων. Μπορούμε να εντοπίσουμε τυχόν μοναδικές προκλήσεις, ανησυχίες ή συναισθήματα που μπορεί να έχουν αυτές οι ομάδες αναλύοντας και αντιπαραβάλλοντας τα συναισθήματά τους.

Ακόμη, τα αποτελέσματα της ανάλυσης συναισθήματος που προέκυψαν κατά τη διάρκεια της πανδημίας COVID-19 έχουν πρακτικές επιπτώσεις. Οι υπεύθυνοι χάραξης πολιτικής, οι οργανισμοί και τα ενδιαφερόμενα μέρη που εμπλέκονται στη διαχείριση κρίσεων και τη λήψη αποφάσεων μπορούν να επωφεληθούν από την τεχνογνωσία τους. Η κατανόηση της συναισθηματικής κατάστασης των επιχειρηματιών και των μη επιχειρηματιών κατά τη διάρκεια μιας κρίσης μπορεί να βοηθήσει στην καθοδήγηση στοχευμένων μέτρων υποστήριξης, στην κατανομή των πόρων και στις αλλαγές πολιτικής για την ικανοποίηση των συγκεκριμένων αναγκών τους. Οι υπεύθυνοι χάραξης πολιτικής μπορούν να προωθήσουν ένα πιο ανθεκτικό και υποστηρικτικό περιβάλλον τόσο για τα άτομα όσο και για τις επιχειρήσεις, εντοπίζοντας και αντιμετωπίζοντας τις συγκεκριμένες συναισθηματικές προκλήσεις που αντιμετωπίζουν αυτές οι ομάδες

3.4.1 Ορόσημα

Ακολουθούν ορισμένα ορόσημα[5] και η περίοδος της πρώτης απαγόρευσης κυκλοφορίας (lockdown) που σχετίζεται με τον COVID-19:

- 12 Δεκεμβρίου 2019: Τα πρώτα κρούσματα μιας άγνωστης ασθένειας που μοιάζει με πνευμονία αναφέρονται στη Γουχάν της Κίνας.
- 11 Ιανουαρίου 2020: Η Κίνα αναφέρει τον πρώτο της θάνατο από τον ιό.
- 30 Ιανουαρίου 2020: Ο Παγκόσμιος Οργανισμός Υγείας (ΠΟΥ) κηρύσσει το COVID-19 παγκόσμια έκτακτη ανάγκη για την υγεία.
- 11 Μαρτίου 2020: Ο ΠΟΥ κηρύσσει το COVID-19 πανδημία.
- 23 Μαρτίου 2020: Το Ηνωμένο Βασίλειο ανακοινώνει το πρώτο του lockdown σε εθνικό επίπεδο, το οποίο διαρκεί έως τις 4 Ιουλίου 2020.
- 24 Μαρτίου 2020: Η Ινδία ανακοινώνει ένα εθνικό lockdown, το οποίο διαρκεί έως τις 31 Μαΐου 2020.
- 31 Μαρτίου 2020: Οι Ηνωμένες Πολιτείες ανακοινώνουν κατευθυντήριες γραμμές για εθνικό lockdown 15 ημερών, το οποίο αργότερα παρατείνεται και διαφέρει ανά πολιτεία.

3.4.2 Περίοδοι Έρευνας

Στην έρευνα μας, χρησιμοποιήσαμε δυο περιόδους του κορονοϊού-19 για την σύγκριση του συναισθήματος επιχειρηματιών και μη επιχειρηματιών πριν και κατά την διάρκεια του κορονοϊού-19. Πρώτη περίοδος είναι η πρώτη απαγόρευση κυκλοφορίας (Lockdown). Ορίσαμε αυτή την περίοδο ως 4 μήνες από τον Μάρτιο του 2020 έως τον Ιούνιο του 2020, προκειμένου να συγκεντρώσουμε επαρκή στοιχεία.

Για τη δεύτερη περίοδο που εξετάσαμε, χρησιμοποιήσαμε το μεγαλύτερο μέρος της διάρκειας της πανδημίας. Πιο συγκεκριμένα, εξετάσαμε το χρονοδιάγραμμα σχεδόν δύο ετών που αρχίζει τον Φεβρουάριο του 2020 και λήγει τον Δεκέμβριο του 2021. Στόχος της έρευνας είναι να δούμε εάν υπάρχουν διαφορές στην αλλαγή του συναισθήματος από την πανδημία μεταξύ των επιχειρηματιών και των μη επιχειρηματιών.

Υπάρχει μια κοινή πεποίθηση ότι οι επιχειρηματίες έχουν περισσότερες πιθανότητες όταν η οικονομία βρίσκεται σε ύφεση, επειδή μπορούν να εντοπίσουν και να επωφεληθούν από τις απαιτήσεις των αναδυόμενων αγορών και τις μεταβαλλόμενες τάσεις των καταναλωτών. Πολλοί ιδιοκτήτες επιχειρήσεων μπόρεσαν να αλλάξουν την κατεύθυνση των επιχειρήσεών τους προκειμένου να ανταποκριθούν στη νέα κανονικότητα ως αποτέλεσμα της πανδημίας COVID-19, η οποία ξεκίνησε στα τέλη του 2019. Αυτό έχει αναστατώσει σημαντικά τις οικονομίες σε όλο τον κόσμο.

Λόγω της τάσης τους να είναι πιο ευέλικτοι, οι επιχειρηματίες μπορεί να τα πάνε ιδιαίτερα καλά σε περιόδους οικονομικής κρίσης. Οι μικρές επιχειρήσεις και οι νεοσύστατες επιχειρήσεις είναι συχνά πιο προσαρμόσιμες και γρήγορα αλλάζουν πορεία για να ικανοποιήσουν τις μεταβαλλόμενες απαιτήσεις των πελατών. Οι επιχειρηματίες μπορεί επίσης να είναι πιο ανεκτικοί στον κίνδυνο από άλλους ανθρώπους, κάτι που μπορεί να είναι χρήσιμο σε αβέβαιες στιγμές.

Υπάρχουν πολλές περιπτώσεις εταιρειών που ευημερούσαν εν μέσω της επιδημίας COVID-19. Για παράδειγμα, καθώς περισσότεροι άνθρωποι αγοράζουν online, η ζήτηση για πλατφόρμες ηλεκτρονικού εμπορίου όπως το Amazon και το Shopify έχει αυξηθεί. Καθώς περισσότεροι άνθρωποι ζητούν γεύματα για παράδοση, οι υπηρεσίες παράδοσης φαγητού όπως το DoorDash και το Uber Eats έχουν επίσης σημειώσει σημαντική ανάπτυξη.

Είναι σημαντικό να θυμάστε, ωστόσο, ότι δεν θα κερδίσουν όλοι οι ιδιοκτήτες επιχειρήσεων από την οικονομική ύφεση. Πολλές μικρές επιχειρήσεις έχουν υποφέρει κατά τη διάρκεια της πανδημίας COVID-19 και ορισμένες εταιρείες μπορεί να είναι πιο ευαίσθητες σε οικονομικές πτώσεις. Επιπλέον, οι μακροπρόθεσμες επιπτώσεις της πανδημίας στην παγκόσμια οικονομία είναι ακόμη άγνωστες και δεν είναι σαφές πώς οι εταιρείες θα συνεχίσουν να προσαρμόζονται τα επόμενα χρόνια.

Πίνακας 3.4.1 Σύγκριση μέσου συναισθήματος των δυο ομάδων, για δυο περιόδους

	Επιχειρηματίες	Μη - Επιχειρηματίες
Μέσο συναίσθημα πριν την περίοδο του κορονοϊού:	0.221	0.195
Μέσο συναίσθημα κατά την περίοδο του κορονοϊού (Περίοδος 2 χρονών Φεβ 2020-Δεκ 2021):	0.204	0.172
Μέσο συναίσθημα κατά την περίοδο του κορονοϊού (Περίοδος 4 ^{ων} μηνών Μάρτιος-Ιούνιος 2020):	0.196	0.167

Πίνακας 3.4.2 Σύγκριση μέσης αλλαγής συναισθήματος των δυο ομάδων.

	Επιχειρηματίες	Μη - Επιχειρηματίες
Μέση Αλλαγή Συναισθήματος [Απόλυτη τιμή (Πριν – Μετά)]	0.0967	0.0892

3.4.3 Αποτελέσματα

Οι δύο προσεγγίσεις που χρησιμοποιούνται για τη σύγκριση του συναισθήματος των επιχειρηματιών και των μη επιχειρηματιών πριν και κατά τη διάρκεια της πανδημίας COVID-19 παρέχουν συμπληρωματικές πληροφορίες για τη συναισθηματική δυναμική αυτών των ομάδων

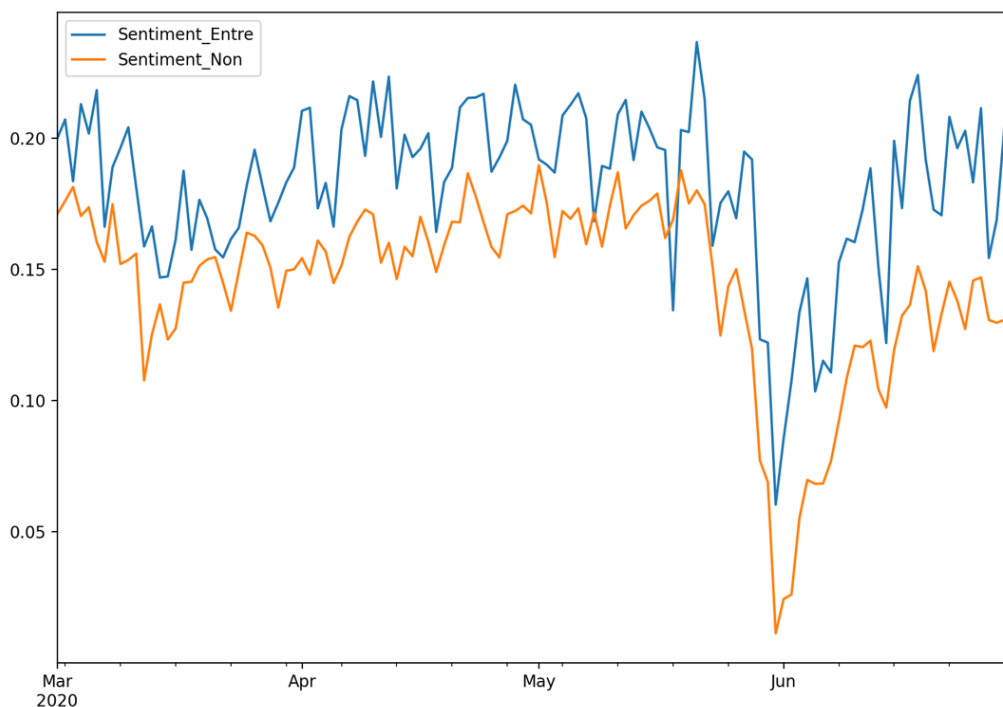
Η πρώτη προσέγγιση αφορούσε τη σύγκριση του μέσου συναισθήματος των επιχειρηματιών και των μη επιχειρηματιών πριν και κατά τη διάρκεια της πανδημίας. Αυτή η προσέγγιση αποκάλυψε ότι και οι δύο ομάδες παρουσίασαν μείωση του συναισθήματος κατά τη διάρκεια της πανδημίας, με τους μη επιχειρηματίες να αντιμετωπίζουν πιο σημαντική πτώση από τους επιχειρηματίες. Τα αποτελέσματα αυτής της προσέγγισης παρέχουν μια ευρεία κατανόηση των συναισθηματικών προτύπων των επιχειρηματιών και των μη επιχειρηματιών κατά τη διάρκεια της πανδημίας.

Η δεύτερη προσέγγιση περιλάμβανε τον υπολογισμό της μέσης αλλαγής στο συναίσθημα για κάθε χρήστη μεταξύ των περιόδων πριν από την COVID-19 και της

πανδημίας και, στη συνέχεια, τη σύγκριση της μέσης αλλαγής στο συναίσθημα μεταξύ επιχειρηματιών και μη επιχειρηματιών. Αυτή η προσέγγιση παρέχει μια πιο αναλυτική εικόνα της συναισθηματικής δυναμικής κάθε μεμονωμένου χρήστη ως απάντηση στην πανδημία. Τα αποτελέσματα αυτής της προσέγγισης υποδηλώνουν ότι τόσο οι επιχειρηματίες όσο και οι μη επιχειρηματίες παρουσίασαν πτώση του συναισθήματος κατά τη διάρκεια της πανδημίας, με τους επιχειρηματίες να παρουσιάζουν ελαφρώς μεγαλύτερη πτώση από τους μη επιχειρηματίες.

Συνολικά, αυτές οι δύο προσεγγίσεις παρέχουν μια ολοκληρωμένη κατανόηση του συναισθηματικού αντίκτυπου της πανδημίας COVID-19 σε επιχειρηματίες και μη. Υποδεικνύουν ότι η πανδημία είχε αρνητικό αντίκτυπο στη συναισθηματική ευημερία των ατόμων και στις δύο ομάδες, με τους μη επιχειρηματίες να αντιμετωπίζουν μια πιο σημαντική πτώση στο συναίσθημα από τους επιχειρηματίες.

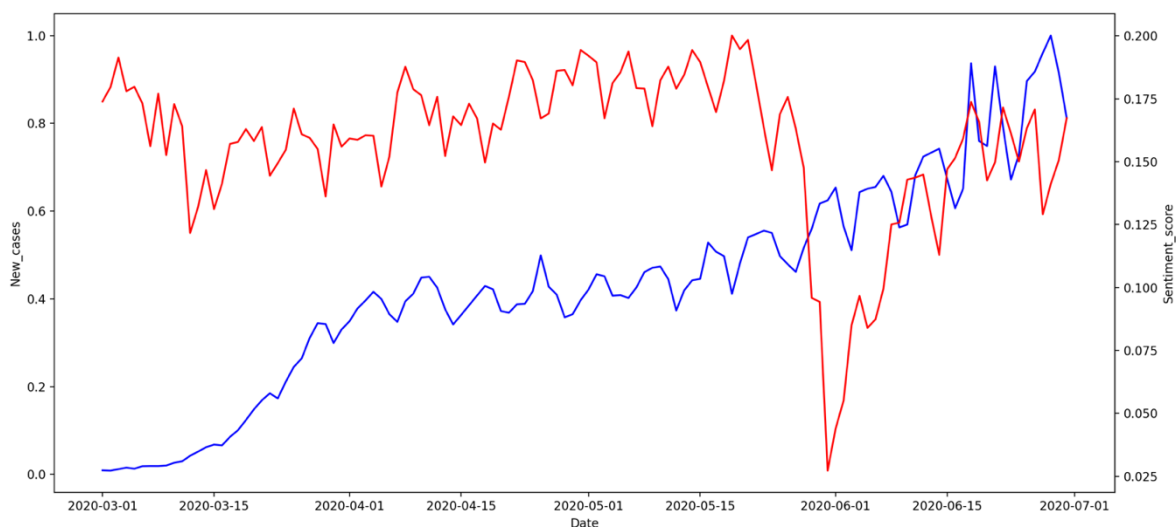
Εικόνα 3.5 Ημερήσιος Μέσος Όρος για τις δυο ομάδες, για την περίοδο του πρώτου lockdown παγκοσμίως



3.4.4 Σύγκριση μέσου συναισθήματος με ημερήσιο αριθμών κρουσμάτων

Σε αυτή την παράγραφο, παρουσιάζουμε μια ανάλυση του συναισθήματος και των περιπτώσεων COVID-19, με στόχο να διερευνήσουμε την πιθανή συσχέτιση μεταξύ του συναισθήματος των επιχειρηματιών και των μη επιχειρηματιών και της εξέλιξης

Εικόνα 2.6 Σύγκριση ημερήσιου μέσου συναισθήματος όλων των χρηστών με ημερήσιο αριθμών κρουσμάτων

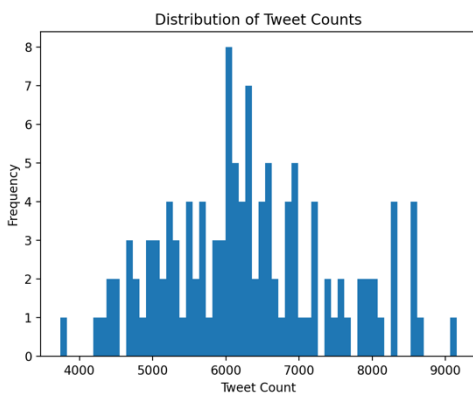


της πανδημίας COVID-19. Η ανάλυση περιλάμβανε τη δημιουργία ενός διαγράμματος που συνέκρινε το μέσο συναίσθημα και των δύο ομάδων με τον κανονικοποιημένο αριθμό των καθημερινά αναφερόμενων θετικών κρουσμάτων COVID-19 κατά τη περίοδο 1 Μαρτίου 2020 μέχρι 30 Ιουνίου 2020. Σκοπός της μελέτης αυτής ήταν να αποκτήσει εικόνα της συναισθηματικής αντίδρασης αυτών των ομάδων κατά τους πρώτους μήνες της πανδημίας και να εξετάσει οποιαδήποτε πιθανή σχέση μεταξύ του συναισθήματος και της σοβαρότητας του ιού.

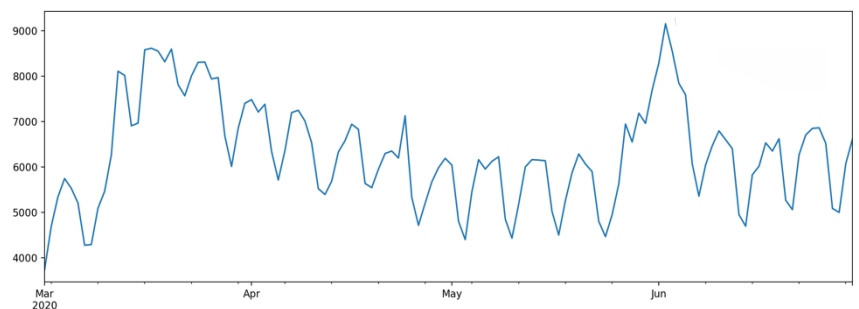
Για τη διεξαγωγή αυτής της ανάλυσης, συλλέξαμε δεδομένα σχετικά με το συναίσθημα των επιχειρηματιών και των μη επιχειρηματιών χρησιμοποιώντας ένα εργαλείο ανάλυσης συναισθήματος. Οι βαθμολογίες συναισθήματος αποδόθηκαν σε κάθε ομάδα για τη συγκεκριμένη χρονική περίοδο. Επιπλέον, ελήφθη ο αριθμός των καθημερινά αναφερόμενων θετικών περιπτώσεων COVID-19 και κανονικοποιήθηκε σε ένα εύρος από 0 έως 1 για να εξασφαλιστεί η συμβατότητα με τις τιμές συναισθήματος. Το γράφημα που προέκυψε απεικόνιζε το μέσο συναίσθημα των επιχειρηματιών και των μη επιχειρηματιών στον άξονα y, ενώ ο άξονας x αντιπροσώπευε τις ημερομηνίες, εστιάζοντας στους τρεις πρώτους μήνες της πανδημίας COVID-19. Ο κανονικοποιημένος αριθμός των καθημερινά αναφερόμενων θετικών κρουσμάτων COVID-19 απεικονίστηκε σε έναν δευτερεύοντα άξονα y.

Κατά την οπτική εξέταση, το διάγραμμα αποκάλυψε διακυμάνσεις στα επίπεδα συναισθήματος και για τις δύο ομάδες με την πάροδο του χρόνου. Ενώ παρατηρήθηκε μια ασθενής αρνητική συσχέτιση -0.25 όπου έχει υπολογιστεί με τη συσχέτιση Pearson, μεταξύ του συναισθήματος και των κρουσμάτων COVID-19, υποδεικνύοντας ότι, κατά μέσο όρο, το συναίσθημα μειώθηκε ελαφρώς καθώς αυξανόταν ο αριθμός των κρουσμάτων, είναι σημαντικό να σημειωθεί ότι η συσχέτιση δεν συνεπάγεται αιτιώδη συνάφεια. Άλλοι παράγοντες μπορεί να συμβάλλουν στις αλλαγές στα επίπεδα του συναισθήματος κατά τη διάρκεια της πανδημίας και απαιτείται περαιτέρω ανάλυση για την πλήρη κατανόηση αυτών των σχέσεων.

3.4.5 Διερεύνηση σημαντικής πτώσης του συναισθήματος



Εικόνα 3.7 Ημερήσια Κατανομή Tweets για την περίοδο πρώτου Lockdown



Εικόνα 3.8 Αριθμός tweets ανά ημέρα, για την περίοδο 4^{ων} μηνών

Μια σε βάθος εξέταση έγινε για να διερευνηθεί μια αξιοσημείωτη πτώση του κλίματος που καταγράφηκε μεταξύ των ημερομηνιών 15 Μαΐου και 1 Ιουνίου. Η αρχική ανακάλυψη έγινε εξετάζοντας ένα γράφημα που έδειχνε το μέσο συναίσθημα των tweets από επιχειρηματίες σε σχέση με μη επιχειρηματίες. Η παρούσα μελέτη επιδιώκει να παράσχει μια πλήρη ανάλυση της έρευνάς μας, συμπεριλαμβανομένης της αξιολόγησης της ποιότητας των δεδομένων, της αξιολόγησης της συχνότητας των tweet και της χρήσης ενός σύννεφου λέξεων για να αντλήσουμε πληροφορίες σχετικά με τα θέματα που καλύφθηκαν κατά τη διάρκεια της συγκεκριμένης περιόδου. Πραγματοποιήθηκε επανεξέταση των πιθανών ασυνεπειών ή περιορισμών των δεδομένων για την επαλήθευση της ακρίβειας και της αξιοπιστίας των δεδομένων.

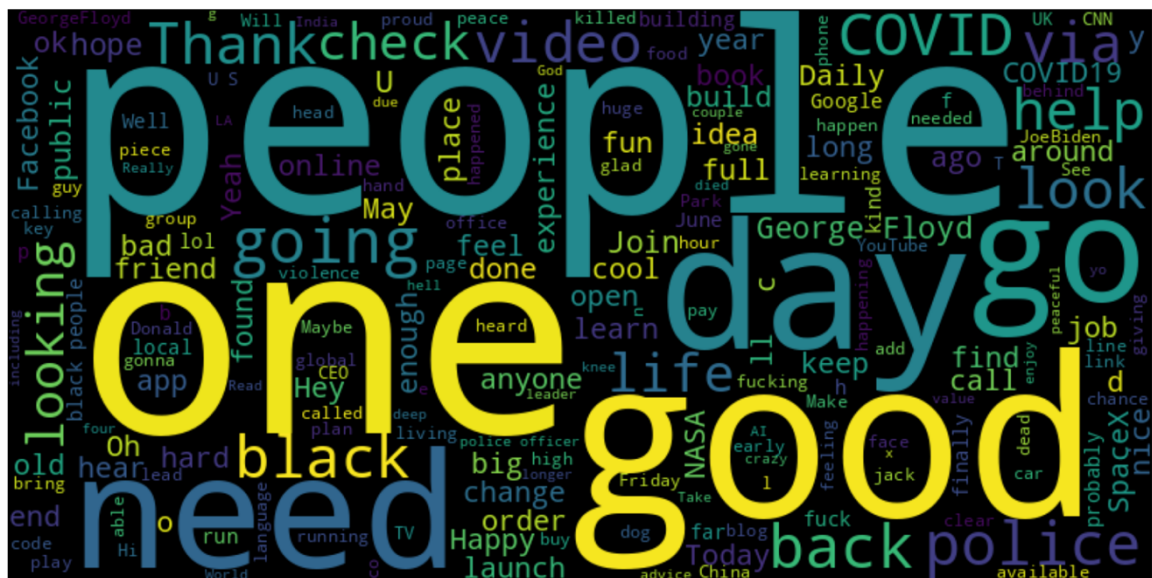
Αρχικά αναρωτηθήκαμε αν η παρατηρούμενη μείωση του συναισθήματος μπορεί να σχετίζεται με δυσκολίες όπως αποτυχίες στη συλλογή δεδομένων ή έλλειψη tweets κατά τη διάρκεια της συγκεκριμένης περιόδου. Η προσεκτική εξέταση της συχνότητας των tweet, από την άλλη πλευρά, έδειξε μια συνεπή και επαρκή αναπαράσταση των δεδομένων καθ' όλη τη διάρκεια. Η αξιολόγηση αυτή μας ικανοποίησε ότι η παρατηρούμενη πτώση του συναισθήματος δεν οφειλόταν σε προβλήματα που σχετίζονται με τα δεδομένα.

Πραγματοποιήθηκε ενδελεχής εξέταση της συχνότητας των tweet προκειμένου να αποκτηθεί μια ολοκληρωμένη εικόνα των μοτίβων του twitter κατά τη διάρκεια της καθορισμένης περιόδου. Η ανάλυση αυτή περιελάμβανε μια ματιά τόσο στο διάγραμμα όσο και στα στατιστικά δεδομένα όπως με τη μέθοδο Shapiro-Wilk[17]. Ειδικότερα, το διάγραμμα και οι στατιστικές μετρήσεις αποκάλυψαν ότι η κατανομή των tweets ανά ημέρα ακολουθούσε κανονική κατανομή. Το εύρημα αυτό υποδηλώνει ότι ο αριθμός των tweets κατά τη διάρκεια της καθορισμένης χρονικής περιόδου ήταν εντός του προβλεπόμενου εύρους. Ως αποτέλεσμα, αυξάνει την αξιοπιστία των δεδομένων που χρησιμοποιούνται για την ανάλυση συναισθήματος.

Πραγματοποιήθηκε ανάλυση νέφους λέξεων για τη μελέτη των θεμάτων και των θεμάτων που ειπώθηκαν κατά τη διάρκεια της περιόδου ενδιαφέροντος, προκειμένου να αποκτηθούν βαθύτερες γνώσεις. Το σύννεφο λέξεων απεικόνισε τους πιο συχνά χρησιμοποιούμενους όρους, αποκαλύπτοντας στοιχεία που επηρεάζουν το συναίσθημα. Ειδικότερα, δύο κύρια γεγονότα ξεχωρίζουν ως ουσιαστικοί παράγοντες που συνέβαλαν στην παρατηρούμενη πτώση του συναισθήματος: η πανδημία COVID-19 και ο θάνατος του George Floyd.

Η παρουσία φράσεων όπως "COVID", "COVID19", "pandemic" και "online" υποδεικνύει ότι οι συζητήσεις σχετικά με τη συνεχιζόμενη παγκόσμια κρίση υγείας ήταν συχνές κατά τη διάρκεια αυτής της χρονικής περιόδου. Ο φόβος, η αβεβαιότητα και οι σημαντικές διαταραχές σε διάφορους τομείς της ζωής προκλήθηκαν από την πανδημία COVID-19. Τα μέτρα αποκλεισμού, η κοινωνική απομόνωση και οι επιπτώσεις του ιού στην υγεία, την απασχόληση και την οικονομία είναι όλοι πιθανοί παράγοντες για το γενικό αρνητικό συναίσθημα που διαπιστώθηκε.

Το συναισθηματικό βάθος και η σημασία αυτού του περιστατικού σίγουρα συνέβαλαν σημαντικά στην πτώση του συναισθήματος κατά τη διάρκεια της υπό μελέτη περιόδου. Εν κατακλείδι, η διερεύνηση της τεράστιας πτώσης του συναισθήματος μεταξύ της 25ης και της 30ής Μαΐου 2020 αποκαλύπτει τον αντίκτυπο δύο σημαντικών γεγονότων: της πανδημίας COVID-19 και του θανάτου του George Floyd. Οι συζητήσεις σχετικά με την πανδημία και την επιρροή της στην καθημερινή ζωή, καθώς και τα επιχειρήματα σχετικά με τη φυλετική αδικία και τις διαμαρτυρίες, συνέβαλαν στην παρατηρούμενη πτώση της στάσης κατά τη διάρκεια αυτής της χρονικής περιόδου.



44

Κεφάλαιο 4

Συζήτηση – Συμπεράσματα

Η έρευνα μας στην επιστήμη των δεδομένων αποσκοπούσε στη διερεύνηση τριών ερευνητικών ερωτημάτων που σχετίζονται με την ανάλυση συναισθήματος στα μέσα κοινωνικής δικτύωσης, συγκεκριμένα στο Twitter, μεταξύ επιχειρηματιών και μη επιχειρηματιών. Αναλύοντας ένα σύνολο δεδομένων 30,1 εκατομμυρίων tweets από τους ίδιους χρήστες με μια προηγούμενη μελέτη, που καλύπτει την περίοδο από το 2016 έως το 2022, χρησιμοποιήσαμε το εργαλείο ανάλυσης συναισθήματος VADER. Επιπλέον, χρησιμοποιήθηκαν στατιστικές μέθοδοι, όπως η πολλαπλή γραμμική παλινδρόμηση, για να προσδιοριστεί η σημαντικότητα των ευρημάτων μας. Πραγματοποιήσαμε επίσης θεματική ανάλυση για να διασφαλίσουμε την απουσία μεροληψίας και εξετάσαμε το συναίσθημα κατά τη διάρκεια της κρίσης COVID-19 φιλτράροντας τα tweets από εκείνη την περίοδο.

Τα κύρια συμπεράσματα της μελέτης μας μπορούν να συνοψιστούν ως εξής:

Κατ' αρχάς, διαπιστώθηκε ότι οι ιδιοκτήτες επιχειρήσεων εξέφρασαν πιο ευνοϊκό συναίσθημα στα μέσα κοινωνικής δικτύωσης από ό,τι οι μη ιδιοκτήτες επιχειρήσεων. Αυτό το αποτέλεσμα ήταν σύμφωνο με άλλες μελέτες, που δείχνουν ένα ξεχωριστό συναισθηματικό μοτίβο που συνδέεται με τις επιχειρήσεις. Δεύτερον, η αναθεώρησή μας μιας προηγούμενης μελέτης που χρησιμοποιούσε δεδομένα από το 2013-2015 παρήγαγε ακριβή και συνεπή αποτελέσματα για το χρονικό διάστημα 2016-2022, επιβεβαιώνοντας τη σταθερότητα των παρατηρούμενων διαφορών συναισθήματος. Τέλος, κατά τη διάρκεια της πανδημίας COVID-19, τόσο οι ιδιοκτήτες επιχειρήσεων όσο και οι μη ιδιοκτήτες επιχειρήσεων είδαν πτώση του συναισθήματος, αλλά οι ιδιοκτήτες επιχειρήσεων έδειξαν υψηλότερο επίπεδο θετικότητας στο Twitter, ακόμη και μπροστά στις δυσκολίες.

Τα ευρήματα της μελέτης μας παρέχουν πολύτιμες πληροφορίες σχετικά με τις συναισθηματικές εκφράσεις των επιχειρηματιών και των μη επιχειρηματιών στις πλατφόρμες κοινωνικής δικτύωσης. Η υψηλότερη τάση των επιχειρηματιών να εκφράζουν θετικό συναίσθημα υποδηλώνει ότι μπορεί να διαθέτουν μια πιο αισιόδοξη

προοπτική, η οποία θα μπορούσε να αποδοθεί στην επιχειρηματική τους νοοτροπία, την ανθεκτικότητα και την ικανότητά τους να αντιμετωπίζουν δύσκολες συνθήκες. Η συνεπής φύση των αποτελεσμάτων σε διαφορετικές χρονικές περιόδους υποστηρίζει περαιτέρω την άποψη ότι οι επιχειρηματίες παρουσιάζουν ένα ξεχωριστό συναισθηματικό μοτίβο στις διαδικτυακές αλληλεπιδράσεις τους.

Η μελέτη μας έχει προεκτάσεις τόσο για τις ακαδημαϊκές όσο και για τις πραγματικές συνθήκες. Η μελέτη μας αναδεικνύει τη συναισθηματική δυναμική που βιώνουν οι επιχειρηματίες, ενώ παράλληλα προσθέτει στο σύνολο των γνώσεων σχετικά με την ανάλυση συναισθήματος. Οι ερευνητές μπορούν να κατανοήσουν καλύτερα τους ψυχολογικούς παράγοντες που επηρεάζουν την επιχειρηματική συμπεριφορά και τη διαδικτυακή παρουσία αυτής της ομάδας, δίνοντας προσοχή στον τρόπο με τον οποίο οι επιχειρηματίες εκφράζουν θετικά συναισθήματα στα μέσα κοινωνικής δικτύωσης. Πρακτικά, τα ευρήματά μας δείχνουν πόσο κρίσιμο είναι να λαμβάνονται υπόψη οι συναισθηματικές αντιδράσεις στους ιστότοπους κοινωνικής δικτύωσης, ιδίως για τους ιδιοκτήτες επιχειρήσεων. Οι επιχειρηματίες μπορούν να επωφεληθούν από τα θετικά συναισθήματά τους για να ενισχύσουν τη διαδικτυακή τους φήμη, να αλληλοεπιδράσουν με το κοινό τους και να δημιουργήσουν ισχυρούς δεσμούς με τα ενδιαφερόμενα μέρη.

Παρόλο που η μελέτη μας προσφέρει χρήσιμες πληροφορίες, είναι ζωτικής σημασίας να αναγνωρίσουμε ορισμένους περιορισμούς. Πρώτον, επειδή στην έρευνά μας εξετάσαμε μόνο το Twitter, δεν μπορούμε απαραίτητα να προεκτείνουμε αυτά τα αποτελέσματα σε άλλες πλατφόρμες κοινωνικής δικτύωσης. Επιπλέον, το λογισμικό ανάλυσης συναισθήματος VADER που χρησιμοποιήσαμε έχει τους δικούς του εγγενείς περιορισμούς, όπως η πιθανότητα ανακριβούς προσδιορισμού της έντασης του συναισθήματος. Επιπλέον, η έρευνά μας επικεντρώθηκε στις διακυμάνσεις στην έκφραση του συναισθήματος και όχι στις υποκείμενες αιτίες αυτών των διακυμάνσεων. Για την καλύτερη κατανόηση των αιτιών των ιδιαίτερων συναισθηματικών μοτίβων που παρατηρούνται μεταξύ των επιχειρηματιών, απαιτείται περισσότερη έρευνα. Η συναισθηματική δυναμική εντός των υποομάδων επιχειρηματιών και μη επιχειρηματιών μπορεί επίσης να γίνει καλύτερα κατανοητή εξετάζοντας τις διακυμάνσεις του συναισθήματος εντός αυτών των ομάδων.

Υπάρχουν διάφορες κατευθύνσεις για πρόσθετη έρευνα ή ανάλυση με βάση τα ευρήματά μας. Πρώτον, για να αποκτήσουμε μια πιο εμπεριστατωμένη γνώση των συναισθηματικών εκφράσεων των επιχειρηματιών, η μελλοντική μελέτη θα μπορούσε να διευρύνει την εστίαση πέρα από το Twitter και να εξετάσει τις τάσεις του συναισθήματος σε πολλαπλές πλατφόρμες κοινωνικής δικτύωσης. Η ακρίβεια και η λεπτομέρεια της διαδικασίας ανάλυσης συναισθήματος θα μπορούσε επίσης να βελτιωθεί με τη χρήση πιο εξελιγμένων εργαλείων ανάλυσης συναισθήματος. Περαιτέρω έρευνα σχετικά με τις υποκείμενες αιτίες των συναισθηματικών μοτίβων που παρατηρούνται, όπως οι ιδιότητες της προσωπικότητας ή τα επιχειρηματικά χαρακτηριστικά, μπορεί να αποκαλύψει περισσότερα για τη συναισθηματική δυναμική των επιχειρηματιών. Τέλος, η ανάλυση των διαφορών στο συναίσθημα εντός συγκεκριμένων υποομάδων, όπως με βάση τον κλάδο ή το μέγεθος της επιχείρησης, μπορεί να αποκαλύψει περαιτέρω περιπλοκές και να δώσει φως στις επιδράσεις του πλαισίου στις συναισθηματικές εκφράσεις.

Οι ερευνητικοί στόχοι αντιμετωπίστηκαν με επιτυχία από το έργο μας στον τομέα της επιστήμης των δεδομένων, το οποίο ανέδειξε τις μοναδικές συναισθηματικές τάσεις των επιχειρηματιών στο Twitter σε σύγκριση με τους μη επιχειρηματίες. Τα αποτελέσματα προσθέτουν στο σύνολο των γνώσεων σχετικά με την ανάλυση συναισθήματος και παρέχουν πληροφορίες για τη συναισθηματική δυναμική των επιχειρηματιών τόσο σε περιόδους ευημερίας όσο και σε περιόδους δυσκολιών.

Βιβλιογραφία

- [1] Harmon, Roesslein, J., & other contributors. Tweepy [Computer software]. <https://doi.org/10.5281/zenodo.7259945>
- [2] Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.
- [3] Kågstroöm J, Karlsson R, Ingridsson E. uClassify. 2008. Available: <https://www.uclassify.com/>
- [4] “MongoDB”, retrieved from: <https://www.mongodb.com>
- [5] World Economic Forum. (2021, December). COVID-19: Coronavirus pandemic, two years, milestones. World Economic Forum. Retrieved from <https://www.weforum.org/agenda/2021/12/covid19-coronavirus-pandemic-two-years-milestones/#:~:text=Two%20years%20after%20the%20first,of%20new%20variants%20of%20coronavirus.>
- [6] Macko, A. and Tyszka, T. (2009), Entrepreneurship and Risk Taking. *Applied Psychology*, 58: 469-487. <https://doi.org/10.1111/j.1464-0597.2009.00402.x>
- [7] A. Rauch and M. Frese, “Let’s put the person back into entrepreneurship research: A meta-analysis on the relationship between business owners’ personality traits, business creation, and success,” *European Journal of Work and Organizational Psychology*, vol. 16, no. 4, pp. 353–385, Dec. 2007, doi: [10.1080/13594320701595438](https://doi.org/10.1080/13594320701595438).
- [8] Waters J, Nicolaou N, Stefanidis D, Efstathiades H, Pallis G, Dikaiakos M (2021) Exploring the sentiment of entrepreneurs on Twitter. *PLoS ONE* 16(7): e0254337. <https://doi.org/10.1371/journal.pone.0254337>
- [9] Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.
- [10] McKinney, W., & others. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51–56).

- [11] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- [12] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95.
- [13] MongoDB. (n.d.). PyMongo 3.12.0 documentation. Retrieved May 28, 2023, from <https://pymongo.readthedocs.io/en/stable/>
- [14] Oesper, L., Merico, D., Isserlin, R., & Bader, G. D. (2011). WordCloud: a Cytoscape plugin to create a visual semantic summary of networks. *Source Code for Biology and Medicine*, 6(1), 7.
- [15] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- [16] *Open Directory Project.org: ODP Web Directory Built With the DMOZ RDF Database*. (n.d.). <http://www.odp.org/homepage.php>
- [17] SHAPIRO, S. S., & WILK, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), 591–611. <https://doi.org/10.1093/biomet/52.3-4.591>