



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS



D.I.M.A

ΠΜΣ “Ψηφιακή Καινοτομία
και Διοίκηση”
MSc in Digital Innovation
and Management

Σχολή Οικονομικών Επιστημών και Διοίκησης
Επιχειρήσεων

Τμήμα Διοικητικής Επιστήμης και Τεχνολογίας

**ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΨΗΦΙΑΚΗ
ΚΑΙΝΟΤΟΜΙΑ ΚΑΙ ΔΙΟΙΚΗΣΗ**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**«Sentiment Analysis στο Twitter.
Η περίπτωση της Αγοράς των Κρυπτονομισμάτων»**

Ντεμίρη Μαρία-Αγγελική

Επιτροπή Επίβλεψης Διπλωματικής Εργασίας

Επιβλέπων Καθηγητής	
Παπαδόπουλος Δημήτρης	
Α΄ Συν-Επιβλέπων κ. Γιωτόπουλος Κωνσταντίνος	Β΄ Συν-Επιβλέπων κ. Χαλκιοπούλου Κωνσταντίνος

Πάτρα 2022

© Copyright συγγραφέως Ντεμίρη Μαρία-Αγγελική 2022

© Copyright θέματος Παπαδόπουλος Δημήτρης

Με την επιφύλαξη παντός δικαιώματος. All rights reserved.

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Διοικητικής Επιστήμης και Τεχνολογίας
δεν συνεπάγεται απαραίτητως και αποδοχή των απόψεων του συγγραφέως εκ μέρους του τμήματος.

ΠΡΟΛΟΓΟΣ

Η παρούσα διπλωματική εργασία εκπονήθηκε κατά τη διάρκεια του χειμερινού ακαδημαϊκού έτους 2021-2022, στα πλαίσια του Μεταπτυχιακού Προγράμματος Σπουδών «Ψηφιακή Καινοτομία και Διοίκηση» του Πανεπιστημίου Πατρών.

Η εργασία πραγματοποιήθηκε υπό την επίβλεψη του κ. Δημήτρη Παπαδόπουλου, Επίκουρο Καθηγητή του Τμήματος Διοικητικής Επιστήμης και Τεχνολογίας του Πανεπιστημίου Πατρών.

Βασικό αντικείμενο της εργασίας αποτελεί η έρευνα και η ανάλυση Συναισθήματος (Sentiment Analysis) σε σχόλια του Twitter που αφορούν την αγορά των Κρυπτονομισμάτων. Η ανάλυση θα γίνει με τη χρήση κάποιων βασικών NLP (Natural Language Processing) μεθόδων με τη χρήση της Γλώσσας Προγραμματισμού Python.

Σε αυτό το σημείο, θα ήθελα να ευχαριστήσω τον επιβλέπον καθηγητή μου, κ. Δημήτρη Παπαδόπουλο καθώς και την οικογένειά μου.

Περίληψη

Όλο και περισσότεροι χρήστες του διαδικτύου έχουν αρχίσει και αλληλοεπιδρούν μεταξύ τους μέσω των Μέσων Κοινωνικής Δικτύωσης (Social Media). Με αυτόν τον τρόπο μεταδίδονται απόψεις, ιδέες και πληροφορίες οι οποίες είναι χρήσιμες και αναγκαίες προς αποθήκευση, επεξεργασία και ανάλυση.

Σε αυτή την εργασία, σκοπός είναι η εκμετάλλευση αυτής της Βάσης Δεδομένων των σχολίων στο Twitter, από τους χρήστες, και να γίνει εστίαση στην «Ανάλυση Συναισθήματος» τους (Εξόρυξη Γνώμης). Επιπλέον, το θέμα της έρευνας θα επικεντρωθεί στα σχόλια που αφορούν την Αγορά των Κρυπτονομισμάτων. Η διαδικασία συλλογής των δεδομένων θα ακολουθήσει συγκεκριμένη μεθοδολογία με χρήση του Twitter Application Program Interface (API). Στη συνέχεια, τα δεδομένα που συλλέχθηκαν θα επεξεργασθούν και θα αναλυθούν με την χρήση της Γλώσσας Προγραμματισμού Python.

Επιπρόσθετα, κατά της διάρκεια της εργασίας θα παρουσιαστεί η σημαντική, για την Ανάλυση Συναισθήματος, έννοια της «Επεξεργασίας της Φυσικής Ανθρώπινης Γλώσσας» (Natural Language Processing -NLP). Έννοια η οποία εστιάζει στο πώς κάθε άνθρωπος βιώνει, αντιλαμβάνεται και δίνει νόημα στις εμπειρίες που απολαμβάνει τόσο στον πραγματικό κόσμο όσο και στον κόσμο του διαδικτύου.

Επιπλέον, θα παρουσιαστεί η εξέλιξη του Παγκόσμιου Ιστού (World Wide Web) από την αρχική του μορφή (Web 1.0) έως σήμερα (Web 3.0), καθώς η σημασία του είναι πολύ σημαντική για την κατανόηση της αγοράς των Κρυπτονομισμάτων. Σε αυτό το κεφάλαιο θα αναλυθεί και το σύστημα του Blockchain, με το οποίο λειτουργεί η Αγορά καθώς θεωρείται από τα σημαντικότερα τεχνολογικά επιτεύγματα του Web 3.0.

Συνοψίζοντας, στην παρούσα διπλωματική εργασία θα γίνει τόσο η θεωρητική όσο και η πρακτική παρουσίαση της «Ανάλυσης Συναισθήματος». Θα συλλεχθούν και θα επεξεργασθούν τα σχόλια στο Twitter που αφορούν της Αγορά των Κρυπτονομισμάτων στις μέρες μας, καθώς επίσης θα γίνει η ανάλυση και η οπτική παρουσίαση των αποτελεσμάτων τους. Τέλος, θα προτείνουμε την μελλοντική ανάγκη για παρόμοιες έρευνες.

Abstract

More and more internet users have started interacting with each other on Social Media. This is a significant way for their ideas, emotions and information to be transmitted. These data are useful and necessary, so they should be stored, processed and analyzed in order to be a useful tool of enterprises.

In this thesis, the aim is to exploit this database, that comes from users' comments on the Twitter Platform and then we will study their Sentiment Analysis (Opinion Mining). In addition, the research topic will focus on the comments concerning the Cryptocurrency Market nowadays. Thus, the process of the data collection will follow a specific methodology by using the Twitter Application Program Interface (API). After that, the collected data will be edited and analyzed by using the Python Programming Language.

Furthermore, during this study there will be presented the concept of Natural Language Processing -NLP, which is important for the Sentiment Analysis' field. So, this is a connotation that focuses on how each human being experiences, perceives and gives meaning to every action they have in both real and online world.

Additionally, the evolution of the World Wide Web from its original form (Web 1.0) to today (Web 3.0) will be presented, as it is very important for understanding the Cryptocurrency Market. In this chapter, there will also be analyzed the Blockchain system, with which the Market operates. Blockchain is considered one of the most important technological achievements of the Web 3.0.

To sum up in this dissertation, there will not only be made the theoretical but also the practical presentation of the Sentiment Analysis. The comments on Twitter regarding the Cryptocurrency Market will be collected and processed, as well as their analysis and their results will be visualized.

Finally, we will give suggestions for future needs of similar researches based on the Opinion Mining topic.

Πίνακας Περιεχομένων

ΠΡΟΛΟΓΟΣ	2
Περίληψη	3
Abstract	4
Πίνακας Περιεχομένων Εικόνων	7
Πίνακας Περιεχομένων Κώδικα	7
Πίνακας Περιεχομένων Πινάκων	8
Πίνακας Περιεχομένων WordCloud	8
Πίνακας Περιεχομένων Scatter Plot	8
Πίνακας Περιεχομένων Pie Chart	9
Πίνακας Περιεχομένων Chart	9
Κεφάλαιο 1ο : Εισαγωγικές έννοιες	10
A. Ορισμός Sentiment Analysis	10
B. Εισαγωγή στην Python	11
C.Περίληψη Κεφαλαίων	12
Κεφάλαιο 2ο: Θεωρητικές Έννοιες της έρευνας	13
A.Η σημασία της Εξέλιξης του Παγκόσμιου Ιστού.	13
Τί είναι ο Παγκόσμιος Ιστός (World Wide Web).....	13
Η διαφορά του Διαδικτύου και του Παγκόσμιου Ιστού (Internet vs Web).	13
Από το Web 1.0 έως το Web 3.0	14
Η Αγορά των Κρυπτονομισμάτων και η Θέση της στο Web 3.0	16
B.Διαδίκτυο και Social Media	17
C. Εξόρυξη Δεδομένων ή Επιστήμη των δεδομένων (Data Science)	18
Ορισμοί:	18
Βιβλιοθήκες Python για την Ανάλυση Συναισθήματος:	19
D. Natural Language Processing (NLP)	21
Ορισμός και μεθοδολογία NLP	21
NLP και Sentiment Analysis.....	22
NLP και Python.....	23
Κεφάλαιο 3ο Εφαρμογή της έρευνας	24
Twitter	24
Twitter for Developers- Twitter API	25
Τί είναι το Twitter API	25
Πρόσβαση στα δεδομένα του Twitter.	26
Sentiment Analysis σε Tweets με χρήση της Γλώσσας Προγραμματισμού Python	27
Διαδικασία έρευνας	27
Τα Βασικά Βήματα για Sentiment Analysis σε Tweet με συγκεκριμένο Hashtag	29
Τα βασικά βήματα για Sentiment Analysis σε Tweet συγκεκριμένου χρήστη-influencer στον χώρο των Κρυπτονομισμάτων	40

Κεφάλαιο 4^ο: Αποτελέσματα Έρευνας	44
Case Studies	44
i. Case Study for Sentiment Analysis in Tweets with the #cryptocurrency	44
ii. Case Study for Sentiment Analysis in Tweets with the #trading	47
iii. Case Study for Sentiment Analysis in Tweets with the #blockchain	49
iv. Case Study for Sentiment Analysis in Tweets with the #eth	50
v. Case Study for Sentiment Analysis in Tweets with the #bitcoin.....	52
vi. Case Study for Sentiment Analysis in Twitter user @rogerkver	54
vii. Case Study for Sentiment Analysis in Twitter user @adam3us.....	55
viii. Case Study for Sentiment Analysis in Twitter user @aantonop	57
ix. Case Study for Sentiment Analysis in Twitter user @cz_binance.....	58
Κεφάλαιο 5^ο: Σύνοψη και Συμπεράσματα Έρευνας.....	60
Μελλοντική Ανάγκη για Υλοποίηση Ερευνών πάνω στη Συναισθηματική Ανάλυση.....	62
Bibliography.....	63
Πηγές Φωτογραφιών.....	65
https://colab.research.google.com/?utm_source=scs-index	65
Παραρτήματα	66
Εντολές μέσω R στο Google Colab για κάθε μελέτη περίπτωση	66

Πίνακας Περιεχομένων Εικόνων

Εικόνα 1 Συναισθηματική Ανάλυση	10
Εικόνα 2 Βασικά Στοιχεία της Python	11
Εικόνα 3 Η πορεία του Διαδικτύου	14
Εικόνα 4 Η Επιστήμη των Δεδομένων	18
Εικόνα 5 Word Cloud της Επιστήμης των Δεδομένων	20
Εικόνα 6 Διάγραμμα Natural Language Processing	22
Εικόνα 7 Στατιστικά Στοιχεία του Twitter	24
Εικόνα 8 Logo Twitter API Developer	25
Εικόνα 9 Logo του Google Colab	27

Πίνακας Περιεχομένων Κώδικα

Κώδικας 1 Εισαγωγή Βιβλιοθηκών	29
Κώδικας 2 Εισαγωγή Κωδικών του Twitter API App	29
Κώδικας 3 Εξόρυξη 2000 Tweet με #crypto	30
Κώδικας 4 Αποθήκευση των Δεδομένων σε ένα Dataframe (#crypto)	31
Κώδικας 5 «Καθαρισμός» των Tweet (#crypto)	31
Κώδικας 6 «Καθαρισμός» των Tweet (#crypto)	32
Κώδικας 7 Εύρεση της «Πόλωσης» (Polarity) και της «Υποκειμενικότητας» (Subjectivity) των Tweet. (#crypto)	33
Κώδικας 8 Οπτικοποίηση αποτελεσμάτων με την μέθοδο του Word Cloud.	34
Κώδικας 9 Υπολογισμός του βαθμού «Συναισθήματος»	35
Κώδικας 10 Δημιουργία ενός Scatter Plot	36
Κώδικας 11 Εμφάνιση όλων των θετικών Tweet #crypto	37
Κώδικας 12 Εμφάνιση όλων των αρνητικών Tweet (#crypto)	37
Κώδικας 13 Εμφάνιση όλων των ουδέτερων Tweet (#crypto)	38
Κώδικας 14 Στατιστική Ανάλυση Συναισθήματος (#crypto)	38
Κώδικας 15 Δημιουργία Διαγράμματος	39
Κώδικας 16 Εισαγωγή Βιβλιοθηκών	40
Κώδικας 17 Εισαγωγή Κωδικών του Twitter API app	40
Κώδικας 18 Εξόρυξη 2000 Tweet από έναν user	41
Κώδικας 19 «Καθαρισμός» των Tweet	41
Κώδικας 20 Δημιουργία διαγράμματος Top Words Used	42
Κώδικας 21 Δημιουργία WordCloud	42
Κώδικας 22 Υπολογισμός του Βαθμού Συναισθήματος	43
Κώδικας 23 Στατιστική Ανάλυση	43
Κώδικας 24 Στατιστική Ανάλυση #cryptocurrency	46

Πίνακας Περιεχομένων Πινάκων

Πίνακας 1 Subjectivity and Polarity (#crypto)	33
Πίνακας 2 Υπολογισμός του βαθμού «Συναισθήματος» (#crypto)	35
Πίνακας 3 Tweet @VitalikButerin	41
Πίνακας 4 Sentiment Analysis για @VitalikButerin	43
Πίνακας 5 Καθαρισμός Tweet #cryptocurrency	44
Πίνακας 6 Polarity and Subjectivity #cryptocurrency	44
Πίνακας 7 Sentiment Analysis #cryptocurrency	45
Πίνακας 8 Sentiment Analysis #trading	48
Πίνακας 9 Tweets by @rogerkver	54
Πίνακας 10 Tweets by @adam3us	55
Πίνακας 11 Tweets by @aantonop	57
Πίνακας 12 Tweets by @cz_binance	58
Πίνακας 13 Λέξεις που χρησιμοποιούν συχνότερα στα Tweet τους	61

Πίνακας Περιεχομένων WordCloud

WordCloud 1 #crypto	34
WordCloud 2 @VitalikButerin	42
WordCloud 3 #cryptocurrency	45
WordCloud 4 #trading	47
WordCloud 5 #blockchain	49
WordCloud 6 #eth	50
WordCloud 7 #bitcoin	52
WordCloud 8 @rogerkver	54
WordCloud 9 @adam3us	56
WordCloud 10 @aantonop	57
WordCloud 11 @cz_binance	59

Πίνακας Περιεχομένων Scatter Plot

ScatterPlot 1 #crypto	36
ScatterPlot 2 #cryptocurrency	46
ScatterPlot 3 #trading	48
ScatterPlot 4 #blockchain	49
ScatterPlot 5 #eth	51
ScatterPlot 6 #bitcoin	52

Πίνακας Περιεχομένων Pie Chart

<i>Pie Chart 1 @VitalikButerin</i>	43
<i>Pie Chart 2 #cryptocurrency</i>	47
<i>Pie Chart 3 #trading</i>	48
<i>Pie Chart 4 #blockchain</i>	50
<i>Pie Chart 5 #eth</i>	51
<i>Pie Chart 6 #bitcoin</i>	53
<i>Pie Chart 7 @rogerkver</i>	55
<i>Pie Chart 8 @adam3us</i>	56
<i>Pie Chart 9 @aantonop</i>	58
<i>Pie Chart 10 @cz_binance</i>	59

Πίνακας Περιεχομένων Chart

<i>Chart 1 #crypto</i>	39
<i>Chart 2 Top Words Used by @VitalikButerin</i>	42
<i>Chart 3 @VitalikButerin</i>	43
<i>Chart 4 Sentiment Analysis #cryptocurrency</i>	47
<i>Chart 5 Sentiment Analysis #trading</i>	48
<i>Chart 6 Sentiment Analysis #blockchain</i>	50
<i>Chart 7 Sentiment Analysis #eth</i>	51
<i>Chart 8 Sentiment Analysis #bitcoin</i>	53
<i>Chart 9 Top Words used by @rogerkver</i>	54
<i>Chart 10 Sentiment Analysis @rogerkver</i>	55
<i>Chart 11 Top Words Used by @adam3us</i>	56
<i>Chart 12 Sentiment Analysis @adam3us</i>	56
<i>Chart 13 Top Words used by @aantonop</i>	57
<i>Chart 14 Sentiment Analysis @aantonop</i>	58
<i>Chart 15 Top Words used by @cz_binance</i>	59
<i>Chart 16 Sentiment Analysis @cz_binance</i>	59
<i>Chart 17 Σύγκριση αποτελεσμάτων στην Ανάλυση Συναισθήματος με συγκεκριμένο Hashtag</i>	60

Κεφάλαιο 1ο : Εισαγωγικές έννοιες

A. Ορισμός Sentiment Analysis

Η Ανάλυση Συναισθήματος (ή Εξόρυξη Γνώμης), γνωστή ως Sentiment Analysis ορίζεται ως:

«το πεδίο της έρευνας που αναλύει τις γνώμες των ανθρώπων, τα συναισθήματα, τις εκτιμήσεις και τις στάσεις απέναντι σε οντότητες όπως τα προϊόντα, υπηρεσίες, ανθρώπους και σημαντικά γεγονότα».

Μέσω της Ανάλυσης Συναισθήματος τα αποτελέσματα επικεντρώνονται κυρίως στις θετικές ή αρνητικές επιδράσεις των χρηστών απέναντι στο ζητούμενο θέμα που ερευνάται. Με αυτή την διαδικασία απορρέονται σημαντικά στοιχεία τα οποία είναι στην διάθεση κάθε ερευνητή ή εταιρίας για επεξεργασία και χρήση. (Liu Bing, 2010)

Παρακάτω παρουσιάζονται κάποια ονόματα με τα οποία είναι γνωστή η Ανάλυση Συναισθήματος στη βιβλιογραφία, χωρίς όμως να είναι συνήθης η χρήση τους:

- Opinion Extraction
- Sentiment Mining
- Emotion Analysis
- Affect Analysis
- Review Mining



Εικόνα 1 Συναισθηματική Ανάλυση

Η Ανάλυση Συναισθήματος είναι μια διαδικασία η οποία επιλύει ένα μεγάλο αριθμό προβληματισμών στις επιχειρήσεις, καθώς με αυτόν τον τρόπο μπορούν να ανακαλύψουν τις στάσεις των ενδιαφερόμενων απέναντι στην αγορά. Ειδικότερα, σε μια εποχή που οι καταναλωτές αλληλοεπιδρούν μεταξύ τους και μεταξύ των επιχειρήσεων μέσω των εφαρμογών Κοινωνικής Δικτύωσης, καθίσταται η ανάγκη για περαιτέρω έρευνα πάνω στη στάση και τη γνώμη των χρηστών. Για αυτό το λόγο υπάρχει αύξηση της ζήτησης από τις εταιρίες για αξιοποίηση του εργαλείου της Εξόρυξης Γνώμης ώστε μέσω των μοντέλων ανάλυσης των δεδομένων να έχουν πλήρη εικόνα για τα προϊόντα τους, της ζήτησης και της μελλοντικής πρόβλεψης των αναγκών τους.

Τέλος, η Ανάλυση Συναισθήματος σχετίζεται άμεσα με την έννοια της Επεξεργασίας της Φυσικής Ανθρώπινης Γλώσσας (Natural Language Processing -NLP) καθώς αποτελεί αναπόσπαστο κομμάτι του. Ως NLP θεωρείται η προσπάθεια να εξάγουμε ένα πλήρες, αντιπροσωπευτικό νόημα για το τί αισθάνεται ή τί στάση έχει κάποιος, μέσω των κειμένων του.

Περαιτέρω ανάλυση της έννοιας του NLP θα γίνει στο κεφάλαιο 2.

B. Εισαγωγή στην Python

Η Python είναι μια γλώσσα δυναμικής τοποθέτησης (*dynamic typing*), που αναπτύχθηκε από τον Ολλανδό προγραμματιστή Guido van Rossum τη δεκαετία του '90 (Rajaraman A., Ullman J., 2013). Πιο συγκεκριμένα, η Python χαρακτηρίζεται από την απλή σύνταξή της η οποία βοηθάει στην ανάπτυξη ευανάγνωστου κώδικα, καθώς επίσης και από το διαδραστικό περιβάλλον της. Στοιχεία που την καθιστούν εύκολη στην εκμάθησή της. Επιπλέον, ως μια γλώσσα δυναμικής τοποθέτησης, η Python, μπορεί να αναπτύξει και να υποστηρίξει μεγάλα προγραμματιστικά έργα.

Χρήση της γλώσσας Python πλέον βρίσκουμε σε πολλούς κλάδους της τεχνολογίας και της καθημερινότητας. Παραδείγματος χάριν, μπορεί να βοηθήσει στην ανάπτυξη γραφικών και παιχνιδιών, σε εφαρμογές και σενάρια του διαδικτύου, σε επιστημονικούς τομείς όπως των μαθηματικών καθώς επίσης και στην επεξεργασία των βάσεων δεδομένων. (Rajaraman A, & Ullman J, 2020)

Είναι αξιοσημείωτο ότι μεγάλες εταιρίες και όμιλοι (Google, Dropbox, Quora κλπ.) εφαρμόζουν την Python στα συστήματά τους, καθώς αποτελεί μια σεναριογλώσσα (*scripting language*).

FUNDAMENTALS				
Datatypes	Flow Control	Functions	File Handling	Object & Class
Numbers Strings Lists Dictionaries	If Else For While Continue	Definition Function Call Docstring Return	Reading Writing Editing	Variables Functions

Εικόνα 2 Βασικά Στοιχεία της Python

Όταν μια γλώσσα προγραμματισμού κατατάσσεται στην κατηγορία «γλώσσα σεναρίων» ή «σεναριογλώσσα» ή «γλώσσα επέκτασης», σημαίνει ότι χρησιμοποιείται για την αυτοματοποίηση επαναλαμβανόμενων εργασιών. Επιπλέον, υποστηρίζει τη δυνατότητα σύνταξης σεναρίων (*scripts*) τα οποία αξιολογούνται απευθείας από ένα εκτελεστικό περιβάλλον (*runtime environment*). Λόγω ότι η Python είναι μια δυναμική γλώσσα, δίνει την δυνατότητα στον προγραμματιστή να δημιουργεί νέες μεταβλητές χωρίς να χρειαστεί να δηλώσει τον τύπο τους, εκτελώντας τις εντολές αυτόματα.

Όλα τα στοιχεία που υποστηρίζει η Python είναι αντικείμενα, όπου το καθένα περιλαμβάνει έναν τύπο, εσωτερική παράσταση και μια σειρά από δυνατές λειτουργίες-μεθόδους. Επίσης, κάθε αντικείμενο αποτελεί ένα στιγμιότυπο μιας αρχέτυπης κλάσης.

Πιο συγκεκριμένα, οι τύποι των δεδομένων που εφαρμόζονται στην Python είναι:

- Αριθμοί (*numbers*) - πχ. 2021
- Συμβολοσειρές (*strings*) – πχ. “University of Patras”
- Λίστες (*lists*) – πχ. [1,2,3]
- Λεξικά (*dictionaries*) – { ‘surname’ : ‘Ntemiri’, ‘firstname’ : ‘Mariangela’ }
- Πλειάδες (*tuples*) : χρησιμοποιούν λιγότερο χώρο από τις «λίστες» και έχουν ταχύτερη επεξεργασία– πχ. (1,2,3)
- Σύνολα (*sets*)- πχ. `set_of_numbers = {10,20,30}`
- Αρχεία (*files*)- πχ. `File_name = open('test.txt', 'r')`

Τα τελευταία χρόνια λόγω της αυξημένης χρήσης της και διάδοσής της στην κοινωνία, μπορεί να ενταχθεί εύκολα σε πολλά περιβάλλοντα από Windows, MacOS μέχρι και μικροσυσκευές (Καλαφούτης Σ, & Σταμούλης Γ., 2018).

Σ.Περίληψη Κεφαλαίων

Η παρούσα διπλωματική εργασία χωρίζεται σε πέντε επιμέρους κεφάλαια.

Όπως είδαμε, στο πρώτο κεφάλαιο έγινε η εισαγωγή στον βασικό κορμό που θα ερευνήσουμε, την Συναισθηματική Ανάλυση, καθώς ακόμα και τη μεθοδολογία της έρευνας η οποία θα βασιστεί στη γλώσσα Python.

Στη συνέχεια, στο δεύτερο κεφάλαιο θα γίνει εκτεταμένη αναφορά στις έννοιες που θα συναντήσουμε στην υπόλοιπη εργασία. Λόγω ότι η έρευνα αφορά ένα ιδιαίτερο θέμα, τα Κρυπτονομίσματα, θα χρειαστεί να παρουσιαστούν όλες οι έννοιες που θα μας βοηθήσουν να κατανοήσουμε την σημαντικότητά του θέματος αυτού.

Ξεκινώντας, θα αναλυθεί η σημασία του Παγκόσμιου ιστού και γιατί είναι τόσο σημαντικό το Web 3.0 στη λειτουργία των Κρυπτονομισμάτων. Επιπλέον, θα δοθεί μεγάλη βάση στην έννοια και την κρισιμότητα της εξόρυξης των δεδομένων (Data Science) διότι σε αυτήν την επιστήμη βασίζεται η έρευνα. Τέλος, στο δεύτερο κεφάλαιο θα γίνει επίσης ανάλυση του όρου NLP (Natural Language Processing) διότι είναι πολύ σημαντικός για την Ανάλυση Συναισθήματος.

Ακολουθώντας, στο τρίτο κεφάλαιο θα γίνει η εφαρμογή της έρευνας στο Twitter. Ακόμα, θα παρουσιαστεί η έννοια του Twitter API, μέσω του οποίου δίνεται η δυνατότητα για λήψη των ζητούμενων δεδομένων από την πλατφόρμα. Η διαδικασία που θα ακολουθήσει η έρευνα της εργασίας θα παρουσιαστεί σε δύο σκέλη, βήμα-βήμα. Στην πρώτη κατηγορία μελέτης περιπτώσεων θα γίνει η ανάλυση συναισθήματος σε ομάδες σχολίων που περιέχουν συγκεκριμένη αναφορά λέξης (Hashtag-#) ενώ στη δεύτερη κατηγορία θα γίνει η ανάλυση πάνω σε σχόλια από χρήστες που έχουν άμεση επίδραση στην Αγορά των Κρυπτονομισμάτων (influencers).

Εφόσον έχουν διερευνηθεί όλες οι έννοιες που είναι σημαντικές για την εργασία, στο τέταρτο κεφάλαιο θα παρουσιαστούν τα αποτελέσματα για κάθε μελέτη περίπτωση ξεχωριστά. Να σημειωθεί ότι στο τέλος της εργασίας, στο παράρτημα θα υπάρχουν εκτεταμένα οι κώδικες κάθε περίπτωσης.

Τέλος, στο πέμπτο κεφάλαιο, θα γίνει μια μικρή σύνοψη της έρευνας. Επιπλέον, θα παρουσιαστούν τα συμπεράσματα καθώς και οι λόγοι που χρειάζεται να γίνονται έρευνες στην Ανάλυση Συναισθήματος.

Κεφάλαιο 2^ο: Θεωρητικές Έννοιες της έρευνας

Α.Η σημασία της Εξέλιξης του Παγκόσμιου Ιστού.

Τί είναι ο Παγκόσμιος Ιστός (World Wide Web)

Ο παγκόσμιος Ιστός (World Wide Web ή Web) είναι ένα υποσύνολο του διαδικτύου που αποτελείται από σελίδες στις οποίες κάθε χρήστης έχει πρόσβαση μέσω ενός προγράμματος περιήγησης (*web browser*).

Η διαφορά του Διαδικτύου και του Παγκόσμιου Ιστού (Internet vs Web).

Το Διαδίκτυο

Ξεκινώντας, το διαδίκτυο (internet) ιστορικά , θέτει τις βάσεις του από τις δεκαετίες του '60 και του '70. Τότε η Αμερικάνικη Κυβέρνηση σκόπευε να διευκολύνει τον διαμοιρασμό των πληροφοριών μεταξύ των υπολογιστών. Είναι αξιοσημείωτο το γεγονός ότι οι πρώτοι ηλεκτρονικοί υπολογιστές εκείνη την εποχή ήταν αρκετά μεγάλοι σε όγκο και η μεταφορά των αρχείων τους γινόταν με μαγνητικές ταινίες μέσω ενός συμβατικού, για αυτούς, ταχυδρομικού συστήματος (Βλαχοπούλου Μ, & Δημητριάδης Σ., 2014).

Η «γέννηση» του διαδικτύου όμως θεωρείται η 1^η Ιανουαρίου του 1983. Μέχρι τότε η επικοινωνία μεταξύ των υπολογιστών διέφερε από σύστημα σε σύστημα. Το 1983 ήταν η χρονιά που δημιουργήθηκε ένα νέο σύστημα/ πρωτόκολλο επικοινωνίας το οποίο ονομάστηκε «Πρωτόκολλο Ελέγχου Μεταφοράς» (Transmission Control Protocol- TCP), ή αλλιώς «Πρωτόκολλο Διαδικτύου» (Internet Protocol- IP). Με τη συγκεκριμένη τεχνολογία όλα τα δίκτυα (networks) μπορούσαν να επικοινωνήσουν μέσω μιας κοινής γλώσσας (Universal Language). (University System of Georgia, n.d.)

Ο Παγκόσμιος Ιστός

Από την άλλη πλευρά, ο Παγκόσμιος Ιστός είναι μια υποκατηγορία του διαδικτύου στην οποία αναπτύσσονται πολλές ιστοσελίδες (webpages).

Αρχικά, οι ιστοσελίδες σχηματίζονται μέσω της γλώσσας HTML (Hyper Text Markup Language), η οποία «αποτελεί την κύρια γλώσσα σήμανσης για τις ιστοσελίδες και τα στοιχεία της είναι τα βασικά δομικά στοιχεία τους». (Wikipedia)

Στην πράξη, μέσω της HTML, επιτρέπεται στους χρήστες να πλοηγηθούν σε ιστοσελίδες πατώντας απλά σε συνδέσμους (links). Η μεταφορά των δεδομένων μέσα στον Παγκόσμιο Ιστό γίνεται βάσει του HTTP- HyperText Transfer Protocol Πρωτόκολλου (Πρωτόκολλο Μεταφοράς Υπερκειμένου). Μέσω αυτής της διαδικασίας , όλες οι ζητούμενες πληροφορίες μεταφέρονται από έναν διακομιστή (server) σε έναν πελάτη (client). (Wikipedia)

Τέλος, κάποιες ακόμα μορφές μεταφοράς δεδομένων στο Web είναι τα μηνύματα στα μέσα κοινωνικής δικτύωσης, τα email, η μεταφορά αρχείων και το cloud.

Από το Web 1.0 έως το Web 3.0 .

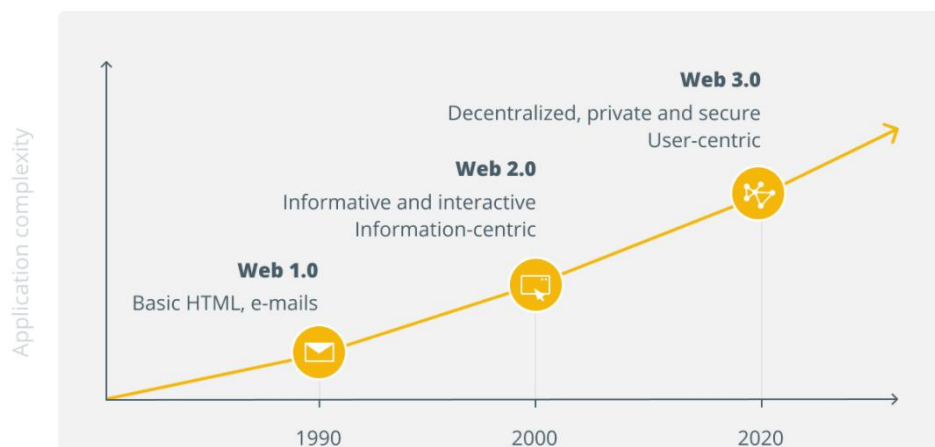
Web 1.0

Η πρώτη μορφή του Παγκόσμιου Ιστού είναι το Web 1.0. Στα πρωταρχικά στάδιά του, το περιεχόμενο ήταν αρκετά περιορισμένο και λίγοι ήταν οι χρήστες που μπορούσαν να έχουν πρόσβαση και ήταν ικανοί να το διαχειριστούν. Επίσης, οι ιστοσελίδες αποτελούνταν από στατικές σελίδες (Read-only Static). (Sharma M, 2022)

Τα βασικά χαρακτηριστικά του Web 1.0 είναι:

- i. Οι στατικές σελίδες (static pages),
- ii. Το περιεχόμενο της ιστοσελίδας δίνεται από το σύστημα αρχείων του διακομιστή (server),
- iii. Οι σελίδες έχουν δημιουργηθεί με Server-Side-Includes ή με Common-Gateway-Interface (CGI). Δηλαδή, στην πρώτη περίπτωση οι εργασίες γίνονται όλες από τον διακομιστή (server), πχ να εμφανιστούν κάποια ζητούμενα αρχεία. Στην δεύτερη περίπτωση, δίνεται η δυνατότητα να υλοποιηθούν σενάρια στον εξυπηρετητή (server) αντί της απλής αποστολής στατικών σελίδων, (Wikipedia)
- iv. Οι πίνακες (tables) και τα πλαίσια (frames) σε μια ιστοσελίδα χρησιμοποιούνται για την ευθυγράμμιση των στοιχείων της σελίδας.

The history of the internet



Εικόνα 3 Η πορεία του Διαδικτύου

source: 101 Blockchains

Web 2.0

Στην συνέχεια της εξέλιξης του Παγκόσμιου Ιστού είναι το Web 2.0 .

Σε αυτή την νέα μορφή Web , δίνεται έμφαση στον χρήστη και στο περιεχόμενο που του προσφέρει η ιστοσελίδα. Οι λέξεις κλειδιά που χαρακτηρίζουν το Web 2.0 είναι η «Διαδραστικότητα» και η «Αλληλεπίδραση». (Kasireddy P, 2021)

Πλέον οι ιστοσελίδες δεν είναι στατικές καθώς δίνεται έμφαση στην κατασκευή τους και στην διαμόρφωσή τους. Μπορούν, δηλαδή, να τροποποιηθούν και να αναπτυχθούν με ευκολία καθώς ακόμα είναι πιο προσιτές οι τιμές «φιλοξενίας» των ιστότοπων (web host).

Είναι φανερό ότι αυτή η «Δυναμική» σχεδίαση των ιστοσελίδων (Read-Write Dynamic) είναι η αφετηρία της ανάπτυξης των Μέσω Μαζικής Δικτύωσης , που έχουμε μέχρι σήμερα. Οι χρήστες στο Web 2.0 είναι σε θέση να αλληλοεπιδρούνε μεταξύ τους με εικόνες, βίντεο, κοινοποιήσεις και οποιαδήποτε άλλης μορφής περιεχομένου βρίσκουν ενδιαφέρον. (Nader "Diamondhands" Al-Naji, 2021)

Τα βασικά χαρακτηριστικά του Web 2.0 είναι:

- i. Οι ιστοσελίδες αναπτύσσονται κυρίως με JavaScript Frameworks,
- ii. Οι πληροφορίες είναι δωρεάν για τους χρήστες, οι οποίοι μπορούν να τις ανακτούν και τις ταξινομούν οποιαδήποτε στιγμή,
- iii. Το περιεχόμενο της ιστοσελίδας είναι δυναμικό,
- iv. Αναπτύσσονται API- Application Programming Interface, τα οποία επιτρέπουν την αυτοχρήση πληροφοριών και λήψη δεδομένων. (Pragati V, 2021)

Web 3.0

Το τρίτο στάδιο της εξέλιξης του Παγκόσμιου Ιστού είναι το Web 3.0 ή αλλιώς Sematic Web (Σημασιολογικός Ιστός).

Σε αυτή τη μορφή Web, δίνεται βάση στην σημασία του περιεχομένου της ιστοσελίδας. Σκοπός είναι η εξέλιξη του Παγκόσμιου Ιστού και η διευκόλυνση του διαμοιρασμού της πληροφορίας και των δεδομένων μεταξύ των χρηστών. Μέσω του Web 3.0 οι υπολογιστές γίνονται πιο «έξυπνοι» και κατανοούν πολυπλοκότερες έννοιες, εντολές και ενέργειες των ανθρώπων. Δηλαδή παραδείγματος χάριν , μια δημοσίευση στο Twitter περιέχει μεταδεδομένα (metadata) , τα οποία θα περιγράφουν άλλα σύνολα δεδομένων. Παράλληλα, αυτά τα μεταδεδομένα καθιστούν πηγή πληροφοριών για τους υπολογιστές, οι οποίοι τα συλλέγουν και θα επεξεργάζονται με μεγαλύτερη ευκολία και ταχύτητα από ότι γινόταν παλιότερα. (Γεωργούλη Κ, 2015)

Τα -5- βασικά χαρακτηριστικά του Web 3.0 είναι:

- i. **Είναι Σημασιολογικός Ιστός (Sematic Web).**
Βελτιώνεται ο τρόπος αναζήτησης του περιεχομένου στον ιστό. Η ικανότητα να κατανοούν πολύπλοκες έννοιες οι μηχανές όσο αφορά την σημασία τους και όχι βάσει μια λέξης-κλειδί, φέρει μεγαλύτερη αξιοπιστία στα αποτελέσματα που δίνονται.
- ii. **Τεχνητή Νοημοσύνη (Artificial Intelligence- AI)**
Με τη βοήθεια του NLP (Natural Language Processing), οι υπολογιστές έχουν την δυνατότητα να επεξεργάζονται τη φυσική γλώσσα και να κατανοούν τις έννοιες όπως οι άνθρωποι. Επίσης, αυτή η ικανότητα φέρει όχι μόνο ταχύτερα αποτελέσματα αλλά και πιο αξιόπιστα για τον χρήστη ώστε να ικανοποιηθούν οι απαιτήσεις του. Σε αυτή την περίπτωση τονίζεται η εξέλιξη της ευφυΐας των υπολογιστών (Winston P, 1993).
- iii. **Τρισδιάστατα Γραφικά (3D Graphics)**
Οι περισσότερες ιστοσελίδες πλέον επικεντρώνονται στην παρουσίασή τους και όχι μόνο στο περιεχόμενό τους. Ο γραφικός σχεδιασμός τους και κυρίως τα πολυδιάστατα γραφικά είναι υπηρεσίες του Web 3.0 που έχουν κερδίσει μεγάλο έδαφος στην μορφοποίηση των ιστοτόπων. Ένα παράδειγμα τέτοιας ανάπτυξης είναι η σχεδίαση των online παιχνιδιών.

iv. **Συνδεσιμότητα (Connectivity)**

στην γενιά του Web 3.0 , ο χρήστης μπορεί να αξιοποιήσει όλες τις πληροφορίες που είναι διαθέσιμες στο διαδίκτυο. Αυτό συμβαίνει διότι πλέον οι πληροφορίες συνδέονται περισσότερο μεταξύ τους χάρη στα σημασιολογικά μεταδεδομένα (semantic metadata).

v. **Πανταχού Παρουσία (Ubiquity)**

Κάθε συσκευή που είναι συνδεδεμένη στο διαδίκτυο καθιστά ικανό το περιεχόμενο του Παγκόσμιου Ιστού να είναι προσβάσιμο και διαθέσιμο από πολλές εφαρμογές ταυτόχρονα, ανεξαρτήτου γεωγραφικής τοποθεσίας. (Mersch M, & Muirhead R., 2019)

Η Αγορά των Κρυπτονομισμάτων και η Θέση της στο Web 3.0 .

Στην εργασία αυτή θα γίνει η Συναισθηματική Ανάλυση σχολίων στην πλατφόρμα του Twitter που αφορούν την Αγορά των Κρυπτονομισμάτων.

Ξεκινώντας, η Αγορά των Κρυπτονομισμάτων και γενικότερα το σύστημα του Blockchain είναι οι ακρογωνιαίοι λίθοι που συγκροτούν τον Παγκόσμιο Ιστό στις μέρες μας. Το Web 3.0 θεωρείται η αποκεντρωμένη έκδοση (decentralized) ενός εικονικού κόσμου, στον οποίο οι χρήστες αλληλοεπιδρούν μεταξύ τους.

Μέσα σε αυτόν τον εικονικό κόσμο λειτουργεί ένα σύστημα ανταλλαγής πληροφοριών ή και ακόμα μπορεί να θεωρηθεί ως ψηφιακό βιβλίο συναλλαγών, το Blockchain. Οι πληροφορίες που βρίσκονται εκεί μεταδίδονται και αντιγράφονται μόνο στους υπολογιστές που βρίσκονται στην ίδια αλυσίδα του Blockchain. Η τεχνολογία με την οποία λειτουργεί αυτό το σύστημα καθιστά αδύνατη την αλλαγή των αποθηκευμένων πληροφοριών , την επεξεργασία τους καθώς ακόμα και την κλοπή τους από εξωτερικούς εισβολείς (hackers). Επιπρόσθετα, κάθε νέα συναλλαγή που πραγματοποιείται μέσα στο σύστημα , προστίθεται στο καθολικό του κάθε συμμετέχοντα. Αυτή η αποκεντρωμένη Βάση Δεδομένων (Decentralized Data Base) που διαχειρίζονται οι συμμετέχοντες είναι γνωστή ως «Κατακερματισμένο Καθολικό» (DLT- Distributed Ledger Technology). (Wikipedia)

Στην περίπτωση του Blockchain , οι συναλλαγές καταγράφονται με μια αμετάβλητη κρυπτογραφική υπογραφή, η οποία ονομάζεται «Κατακερματισμός» (Hash).

Επιπλέον, τα πιο γνωστά και πιο διαπεράστα συστήματα Blockchain είναι το Bitcoin και το Ethereum. Στην περίπτωση αυτών των δύο παραδειγμάτων, όσο μεγαλώνουν τα “blocks” στην αλυσίδα τους, τόσο δυσκολότερη είναι η παραβίαση των πληροφοριών τους. Επομένως αυξάνεται η ασφάλεια του καθολικού τους (DLT).

B. Διαδίκτυο και Social Media

Στην εργασία αυτή θα γίνει χρήση της εξόρυξης δεδομένων από Γράφους Κοινωνικών Δικτύων. Και συγκεκριμένα από την πλατφόρμα του Twitter.

Ξεκινώντας, η χρήση των Social Media αποτελεί μεγάλο μέρος της καθημερινότητας των ανθρώπων, ανεξαρτήτου ηλικίας. Άτομα από κάθε κοινωνική τάξη, χώρα και μορφωτικό επίπεδο έχουν πρόσβαση στο διαδίκτυο και συνεπώς σε κάποιο Μέσο Κοινωνικής Δικτύωσης. Μέσω της χρήσης τους και της πλοήγησής τους, αλληλοεπιδρούν, ανταλλάσσουν ιδέες, μηνύματα, πολυμέσα (όπως φωτογραφίες , βίντεο, αρχεία) και εμπειρίες. (Michael Dewing, 2010)

Όταν αυτές οι κοινοποιήσεις είναι δημόσιες σε προβολή, αποτελούν πηγή για άντληση πληροφοριών είτε για οργανισμούς και επιχειρήσεις είτε για μεμονωμένους ερευνητές. Δηλαδή, είναι κοινώς προσβάσιμες για συλλογή και επεξεργασία ώστε στη συνέχεια να αναλυθούν.

Πιο συγκεκριμένα, τα social media αποτελούν βασικό παράδειγμα του Web 2.0 (Διαδραστικός Ιστός) και οι βασικές υπηρεσίες που προσφέρει είναι:

a. Ιστολόγια-Blogs:

Τα ιστολόγια (ή Blogs , συντομογραφία το “web log”) είναι συνήθως ιστοσελίδες ή και ακόμα μέρη από ιστοσελίδες. Σε ένα blog παρέχονται κείμενα που αφορούν διάφορα θέματα και το περιεχόμενο του είναι σε μορφή άρθρου καθώς επίσης η πρόσβαση σε αυτό είναι εύκολη από κάθε ενδιαφερόμενο. Επιπλέον, τα ιστολόγια είναι σαν ένα «προσωπικό ημερολόγιο» του διαχειριστή διότι μοιράζεται ενδιαφέροντα και προσωπικές του στιγμές. Παραδείγματος χάρη , μπορεί ένα blog να αναφέρεται στις εξορμήσεις και τα ταξίδια του συγγραφέα. Η πιο γνωστή πλατφόρμα για δημιουργία ενός ιστολογίου είναι το “Blogger” , η οποία είναι η παλαιότερη και ανήκει στην Google. (Maisha R, 2021)

b. Online Εγκυκλοπαίδειες-Wikis:

Είναι οι ιστοσελίδες όπου ο καθένας μπορεί να έχει πρόσβαση στην αναδιαμόρφωσή τους , να προσθέσει ή και να αφαιρέσει πληροφορίες. Είναι οι γνωστές online εγκυκλοπαίδειες, όπως η “Wikipedia” , στην οποία η πρόσβαση είναι δωρεάν.

c. Social Network:

Είναι ιστοσελίδες διαμόρφωσης περιεχομένου. Εκεί οι χρήστες , αφού έχουν δημιουργήσει το προφίλ του, μπορούν να αλληλοεπιδρούν και να επικοινωνούν με άλλες χρήστες ή ομάδες χρηστών. Επίσης, στα Μέσα Κοινωνικής Δικτύωσης δημιουργείται μια ψηφιακή κοινωνία. (Wikipedia)

C. Εξόρυξη Δεδομένων ή Επιστήμη των δεδομένων (Data Science)

Ορισμοί:

Ως εξόρυξη των δεδομένων ορίζεται η χρήση του ισχυρότερου υλικού σε συνδυασμό με αποτελεσματικούς αλγορίθμους ώστε να επιλυθούν σημαντικά προβλήματα στο χώρο της επιστήμης, της υγείας, του εμπορίου και γενικότερα σε διάφορους κλάδους της κοινωνίας. (Berry M, & Kogan J., 2010)

Επιπρόσθετα, ο τομέας της επιστήμης των δεδομένων μπορεί να θεωρηθεί η δημιουργία ενός μοντέλου χρησιμοποιώντας τη Μηχανική Μάθηση (Machine Learning). Συγκεκριμένα ως Μηχανική Μάθηση βάσει του Mitchell (1997) έχει δοθεί ο παρακάτω γενικός ορισμός:

«Ένα πρόγραμμα υπολογιστή λέμε ότι μαθαίνει από την εμπειρία E ως προς κάποια κλάση εργασιών T και μέτρο απόδοσης P , αν η απόδοσή του σε εργασίες από το T , όπως μετριέται από το P , βελτιώνεται μέσω της εμπειρίας E ».



Εικόνα 4 Η Επιστήμη των Δεδομένων

Εναλλακτικά, η Μηχανική Μάθηση μπορεί να οριστεί πιο απλοϊκά και κατανοητά ως:

«το φαινόμενο κατά το οποίο ένα σύστημα βελτιώνει την απόδοσή του κατά την εκτέλεση μιας συγκεκριμένης εργασίας, χωρίς να υπάρχει ανάγκη να προγραμματιστεί εκ νέου».

(Animikh Aich, 2019)

Στόχος της εξόρυξης δεδομένων είναι η δημιουργία ενός μοντέλου και έπειτα ενός αλγορίθμου επίλυσης του προβλήματος.

Οι πρώτοι που ένταξαν στην καθημερινότητά τους και στην εργασία τους την επιστήμη των Δεδομένων ήταν οι Μαθηματικοί και κυρίως οι στατιστικοί, οι οποίοι απορρέουν ορατά δεδομένα μέσω ενός στατιστικού μοντέλου, μοντελοποιώντας τα δεδομένα.

Για να γίνει πιο προσιτή η έννοια της μοντελοποίησης των δεδομένων, έχουν τεθεί διάφορες προσεγγίσεις, όπως ό,τι είναι:

- η κατασκευή μιας στατιστικής διαδικασίας με την οποία τα δεδομένα θα μπορούσαν να δημιουργηθούν,
- η σύνοψη των δεδομένων με σύντομο και προσεγγιστικό τρόπο και τέλος,
- η εξαγωγή του πλέον εξέχοντος γνωρίσματος των δεδομένων αγνοώντας τα υπόλοιπα.

(Luellen E, 2018)

Βιβλιοθήκες Python για την Ανάλυση Συναισθήματος:

Όπως έγινε αναφορά στον *Κεφάλαιο 1*, για να συλλεχθούν και να μοντελοποιηθούν τα δεδομένα θα γίνει χρήση της Γλώσσας Προγραμματισμού Python.

Για να χτιστεί ένα ολοκληρωμένο σύστημα ανάλυσης θα πρέπει να δοθεί μεγάλη βάση στην αρχιτεκτονική του μοντέλου και του τελικού έργου ώστε να αποδώσει τα ακριβή αποτελέσματα. (Jayendran, 2021)

Λόγω ότι η επιστήμη της Ανάλυσης Συναισθήματος βρίσκεται σε αναδυόμενο στάδιο , η χρήση της κατάλληλης Βιβλιοθήκης της Γλώσσας Python (Libraries) θα μας δώσει και την αναμενόμενη ζητούμενη απόδοση. Με απώτερο σκοπό την πλήρη εικόνα του μοντέλου και των περιορισμών του, ο ερευνητής χρειάζεται να γνωρίζει τις βιβλιοθήκες οι οποίες να είναι και κατάλληλες για NLP ανάλυση.

Επιπλέον, αν και υπάρχουν αρκετές διαθέσιμες βιβλιοθήκες για NLP Analysis, πριν γίνει χρήση τους θα χρειαστεί να ελεγχθούν ώστε να είναι ικανές και για την Ανάλυση Συναισθήματος (Sentiment Analysis). (Hellmann D, 2011)

Οι κυριότερες Python-based NLP Βιβλιοθήκες είναι οι εξείς:

1. NLTK (Natural Language Toolkit)
2. SpaCy,
3. TextBlob,
4. Stanford CoreNLP και
5. Gensim.

Για να γίνει όμως Ανάλυση Συναισθήματος μέσω της πλατφόρμας του Twitter, είναι προτιμότερη η χρήση των παρακάτω Library:

1. Tweepy,
2. TextBlob
3. Pandas
4. Numpy
5. Wordcloud
6. Re και
7. Matplotlib.

(Anderson Martin, 2019)

Στην έρευνα που θα ακολουθήσει έχει γίνει χρήση και των επτά βιβλιοθηκών οι οποίες είναι χρήσιμες για να βγάλουμε αξιόπιστα αποτελέσματα όσο αφορά την Sentiment Analysis.

Συγκεκριμένα κάθε βιβλιοθήκη που χρησιμοποιήσαμε λειτουργεί ως εξής:

1. Tweepy

Είναι ένα πακέτο ανοιχτού κώδικα για την Python ώστε να δίνει τη δυνατότητα στον χρήστη να έχει πρόσβαση στο Twitter API. Επίσης περιέχει μεθόδους και κλάσεις που αντιπροσωπεύουν τα μοντέλα του Twitter και τα API endpoint. (JP Hwang, 2020)

2. TextBlob

Η βιβλιοθήκη TextBlob είναι από τις βασικότερες βιβλιοθήκες στην Python που βασίζονται στο Natural Language Processing (NLP). Πιο συγκεκριμένα, δίνει τη δυνατότητα να υπάρχει πρόσβαση σε διάφορες εργασίες του NLP όπως η εξαγωγή φράσεων, ανάλυση συναισθημάτων κειμένου, ταξινόμηση και μετάφραση. (Shubham J, 2018)

3. Pandas

Η Pandas , είναι μια από τις πιο γνωστές βιβλιοθήκες για ανάλυση δεδομένων στην γλώσσα Python. Δημιουργήθηκε το 2008 και θεωρείται από το πιο δυνατό και ευέλικτο εργαλείο για ανάλυση. Επιπλέον, έχει χτιστεί βάσει άλλων δυο βιβλιοθηκών της Python, την matplotlib και την Numpy που θα τις εξηγήσουμε αναλυτικά στη συνέχεια. (Python Libraries)

4. Numpy

Είναι μια επιστημονική υπολογιστική βιβλιοθήκη για την Python. Χαρακτηριστικό της είναι ότι προσφέρει στον προγραμματιστή μαθηματικές συναρτήσεις υψηλού επιπέδου και μια πολυδιάστατη δομή (γνωστή ως **ndarray**) ώστε να χειρίζεται μεγάλα σύνολα δεδομένων. Είναι αξιοσημείωτο ότι πολλές άλλες βιβλιοθήκες ανάλυσης δεδομένων έχουν χτιστεί πάνω στην δομή της Numpy. (Python Libraries)

5. Wordcloud

Η Wordcloud βιβλιοθήκη προσφέρει τεχνικές οπτικοποίησης. Πιο συγκεκριμένα, παρουσιάζει μια λέξη ανάλογα με την συχνότητα εμφάνισης ή την σημαντικότητά της στο αντίστοιχο μέγεθος γραμματοσειράς. Η χρήση της Wordcloud γίνεται συνήθως για ανάλυση δεδομένων από τα μέσα κοινωνικής δικτύωσης. (Python Libraries)

Ένα παράδειγμα είναι:



Εικόνα 5 Word Cloud της Επιστήμης των Δεδομένων

6. RE (Regular Expressions)

Η RE βιβλιοθήκη προσφέρει μια ειδική ακολουθία χαρακτήρων που βοηθάει τον αναλυτή να ταιριάζει ή να βρει άλλες συμβολοσειρές χρησιμοποιώντας μια εξειδικευμένη σύνταξη ώστε να ταιριάζει στο μοτίβο που επιθυμεί. (Python Libraries)

7. Matplotlib

Είναι από τις παλιότερες και πιο ευρέως χρησιμοποιούμενη βιβλιοθήκη της Python για οπτικοποίηση των δεδομένων. Αρχικά, έχει δημιουργηθεί από τον επιστήμονα και νευρολόγο John D. Hunter με σκοπό να σχεδιάζει γραφικά τα δεδομένα που καταγράφει από τους ασθενείς του. Στο χώρο του Data Science , οι προγραμματιστές χρησιμοποιούν την Matplotlib παράλληλα με άλλες αντίστοιχες γλώσσες τις Python. (Python Libraries)

D. Natural Language Processing (NLP)

Ορισμός και μεθοδολογία NLP

Στην ελληνική γλώσσα ο όρος Natural Language Processing (NLP) μπορεί να μεταφραστεί ως η «Επεξεργασία της Φυσικής Ανθρώπινης Γλώσσας». Εν συντομία στην εργασία αυτή θα παρουσιάζεται ως 'NLP'.

Ξεκινάμε θέτοντας το τί κάνει το NLP στην πράξη. Αρχικά, προσπαθεί να δημιουργήσει μηχανές που καταλαβαίνουν και ανταποκρίνονται σε δεδομένα κειμένου ή φωνής (text and voice data) και στη συνέχεια αποδίδουν κάποια απάντηση με ένα αντίστοιχο γραπτό κείμενο ή φωνητικό μήνυμα. Όπως ακριβώς γίνεται σε μια ανθρώπινη συνομιλία.

Συγκεκριμένα, το NLP υπάγεται στον κλάδο της επιστήμης των Ηλεκτρονικών Υπολογιστών και κυρίως στον τομέα που ασχολείται με την Τεχνητή Νοημοσύνη (Kao A, & Poteet S., 2007) (Artificial Intelligence- AI). Ως τεχνητή νοημοσύνη ορίζεται:

«εκείνος ο κλάδος της επιστήμης των υπολογιστών που ασχολείται με το σχεδιασμό ευφρών υπολογιστικών συστημάτων, δηλαδή συστημάτων με χαρακτηριστικά τα οποία σχετίζονται με την ευφυΐα στην ανθρώπινη συμπεριφορά (μάθηση, αιτίαση, επίλυση προβλημάτων, κατανόηση φυσικής γλώσσας, αναγνώριση αντικειμένων κτλ.)».

Όπως αναφέρει στο βιβλίο του ο καθηγητής του MIT, Patrick Henry Winston, Η Τεχνητή Νοημοσύνη μας βοηθάει να γίνουμε πιο έξυπνοι και κατά συνέπεια να γίνουν και οι υπολογιστές.

Επιπρόσθετα, η διαδικασία που χρησιμοποιεί η τεχνολογία NLP, είναι να συνδυάσει διάφορα μοντέλα, όπως της μηχανική μάθηση (machine learning), την Μοντελοποίηση της ανθρώπινης γλώσσας και την «εκ βάθους» μάθησης (deep learning). Ο συνδυασμός αυτός δίνει τη δυνατότητα στους υπολογιστές να **επεξεργαστούν** την ανθρώπινη γλώσσα είτε αυτή είναι σε γραπτή μορφή είτε σε μορφή φωνητικών δεδομένων και τελικώς να **«κατανοήσουν»** το πλήρες νόημά της. Όλη η παραπάνω διαδικασία πρωτίστως βασίζεται στην συναισθηματική στάση και πρόθεση του συγγραφέα του κειμένου ή του ομιλητή. (Özgür Genç, 2019)

Εν συντομία, πολλά υπολογιστικά προγράμματα που μεταφράζουν κείμενα που μια γλώσσα σε άλλη ή που ανταποκρίνονται σε γραπτές ή προφορικές εντολές, συνήθως σε πραγματικό χρόνο, λειτουργούν με την καθοδήγηση ενός μοντέλου NLP.

Οι πιο γνωστές εφαρμογές που λειτουργούν με την τεχνογνωσία του NLP και τις βρίσκουμε στην καθημερινότητά μας είναι η Siri (για συσκευές με λειτουργικό iOS), η Alexa (συσκευές της Amazon) και η Google Assistant. Και οι τρεις αυτές «κυρίες» δέχονται εντολές από τον χρήστη της συσκευής είτε γραπτά είτε προφορικά. Για την ενεργοποίηση τους πρέπει ο χρήστης να περάσει κάποια «δοκιμαστικά» στάδια, ώστε η ίδια βοηθός να αναγνωρίζει στο μέλλον την φωνή του.

Ένα παράδειγμα μιας εντολής στην Siri είναι:

“Hey Siri, set a timer for 10 minutes”

Δίνοντας άμεσα την απάντηση:

“10 minutes counting down” – ανοίγοντας ταυτόχρονα την αντίστροφη μέτρηση της συσκευής.

Ας αναλύσουμε όμως την παραπάνω εντολή.

- “Hey Siri” είναι η εντολή που ΠΑΝΤΑ θα «πυροδοτεί» την εκκίνηση της Siri, έχοντας αυτόματα αναγνωρίσει την φωνή σου, ώστε στη συνέχεια να είναι έτοιμη να λάβει τα δεδομένα προς επεξεργασία.
- “set a timer for 10 minutes” είναι η εντολή που θα λάβει η Siri, θα την επεξεργαστεί, θα αναζητήσει λύση και στο τέλος θα την εκτελέσει, προς όφελος του χρήστη.

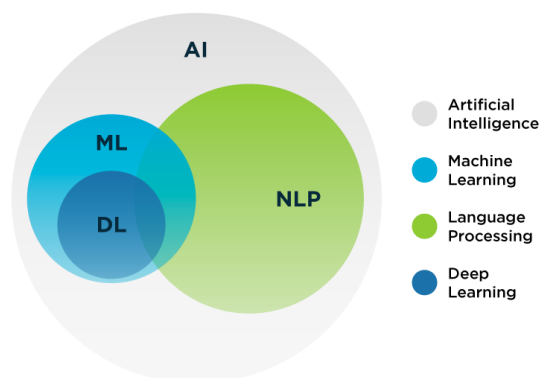
NLP και Sentiment Analysis

Σε αυτή τη διπλωματική εργασία θα γίνει η Συναισθηματική Ανάλυση (Sentiment Analysis) σε σχόλια στο Twitter, τα οποία σχετίζονται με μια συγκεκριμένη θεματολογία.

Όπως ήδη αναφέρθηκε παραπάνω, σκοπός του NLP είναι να δημιουργήσει υπολογιστές που κατανοούν την ανθρώπινη γλώσσα και συμπεριφορά, εξίσου καλά όπως ένα άτομο. Μια βασική εργασία της διαδικασίας αυτής είναι ότι οι μηχανές πλέον κατανοούν τις συναισθηματικές ενδείξεις ενός κειμένου.

Με αυτόν τον τρόπο εξάγονται συμπεράσματα για το αν μια πρόταση είναι θετική/ αρνητική αν έχει χιούμορ ή είναι συναισθηματικά φορτισμένη.

Αυτή η διαδικασία βρίσκει έδαφος για συνεχή ανάπτυξη και αποδοχή τα τελευταία χρόνια από πολλούς προγραμματιστές και κυρίως επιχειρήσεις που θέλουν να μαθαίνουν την στάση των πελατών για τις υπηρεσίες ή προϊόντα τους βάσει των κριτικών ή σχολίων στα Social Media.



Εικόνα 6 Διάγραμμα Natural Language Processing

Η γλώσσα προγραμματισμού Python παρέχει στους χρήστες της πολλές βιβλιοθήκες (libraries) που είναι χρήσιμες στην επεξεργασία της φυσικής γλώσσας- NLP. Αρκετές από αυτές τις βιβλιοθήκες υπάγονται στην «εργαλειοθήκη» του NLP, που ονομάζεται **NLTK** .

Εκτός από τις βασικές βιβλιοθήκες για πρωτεύοντες εργασίες όπως τη Συναισθηματική Ανάλυση, η NLTK περιέχει και βιβλιοθήκες για δευτερεύοντες εργασίες όπως:

- i. Τμηματοποίηση λέξεων (word segmentation),
- ii. Απλή ανάλυση προτάσεων (sentence parsing),
- iii. Περικοπή λέξεων βάσει της ρίζας τους (stemming and lemmatization),
- iv. Διάσπαση φράσεων ή αποσπασμάτων σε διακριτά τμήματα ώστε να είναι ευκολότερο κατανοητά από τον υπολογιστή (tokenization) και
- v. Εξαγωγή λογικών συμπερασμάτων βάσει γεγονότων που παρουσιάζονται στο κείμενο (semantic reasoning).

Το NLTK θεωρείται πλέον η πιο εύχρηστη και αποδοτικότερη πλατφόρμα με βιβλιοθήκες για NLP ανάλυση με την Python. (NLTK Python Libraries)

Κεφάλαιο 3^ο Εφαρμογή της έρευνας

Twitter

Το Twitter είναι μια από τις μεγαλύτερες πλατφόρμες παγκοσμίως που ανήκουν στην κατηγορία των Social Media. Σε αυτόν τον ιστότοπο οι χρήστες μπορούν να μοιράζονται μεταξύ τους πολυμέσα, ιδέες, μηνύματα και κοινοποιήσεις, τα γνωστά “Tweets”. Το ιδιαίτερο χαρακτηριστικό είναι το Microblogging χαρακτηριστικό των δημοσιεύσεων των χρηστών. Πιο συγκεκριμένα ως Microblogging ορίζεται:

Μια δημοσίευση που υπάρχει ως μια συγκεκριμένη μορφή blogging. Ένα micro-blog διαφέρει από ένα παραδοσιακό blog στο ότι το περιεχόμενό του είναι συνήθως μικρότερο τόσο σε πραγματικό όσο και σε συγκεντρωτικό μέγεθος αρχείου.

*Τα micro-blog «επιτρέπουν στους χρήστες να ανταλλάσσουν μικρά στοιχεία περιεχομένου, όπως σύντομες προτάσεις, μεμονωμένες εικόνες ή συνδέσμους βίντεο» ,που μπορεί να είναι ο κύριος λόγος για τη δημοτικότητά τους. Αυτά τα μικρά μηνύματα ονομάζονται μερικές φορές και **μικρο-αναρτήσεις (micro post)**. (Wikipedia)*

Επιπλέον, το Twitter ιδρύθηκε στις 21 Μαρτίου 2006 από τον Jack Dorsey, Noah Glass, Biz Stone και τον Evan Williams, στο Σαν Φρανσίσκο των Ηνωμένων Πολιτειών της Αμερικής. Στο ευρύ κοινό όμως ήταν διαθέσιμο για χρήση τον Ιούλιο του 2006. (Wikipedia)

Μέχρι τώρα καταμετράει πάνω από 330 εκατομμύρια ενεργούς χρήστες και καθημερινώς καταγράφονται περισσότερα από 140 εκατομμύρια tweets .



Εικόνα 7 Στατιστικά Στοιχεία του Twitter

Χαρακτηριστικό στοιχείο του Twitter είναι το ύφος και η δομή των δημοσιεύσεων. Στο Twitter μπορείς να μάθεις αμέσως (on time) κάθε γεγονός που συμβαίνει στο κόσμο και τι πιστεύουν οι άνθρωποι για αυτό.

Twitter for Developers- Twitter API

Εμβαθύνοντας στην συνεισφορά του Twitter στην κοινωνία, εκτός από την άμεση ενημέρωση και την ελευθερία του λόγου για τους χρήστες, η πλατφόρμα προσφέρει σημαντική βοήθεια και πρόσβαση στα δεδομένα της στους προγραμματιστές και στους ερευνητές.

Developer Platform

Εικόνα 8 Logo Twitter API Developer

Τί είναι το Twitter API

Αρχικά, το ακρωνύμιο API αναφέρεται στο “Application Programming Interface” (Διεπαφή Προγραμματισμού Εφαρμογών), το οποίο είναι η σύνδεση μεταξύ υπολογιστών ή προγραμμάτων. Βασική εργασία ενός API είναι να *ορίζει το σύνολο των λειτουργιών που μπορεί να παρέχει μια βιβλιοθήκη ή ένα λειτουργικό σύστημα σε άλλα προγράμματα*. Επιπλέον, ένα API αποτελείται κυρίως από μέρη που λειτουργούν ως διαθέσιμα εργαλεία για έναν προγραμματιστή, ο οποίος μπορεί οποιαδήποτε στιγμή να τα χρησιμοποιήσει / «καλέσει» (calls). Αυτά τα “calls” είναι γνωστά ως μέθοδοι ή endpoints. (Wikipedia)

Στην περίπτωση του Twitter API, η εταιρία προσφέρει δωρεάν πρόσβαση στα δεδομένα τις στους ερευνητές και στους προγραμματιστές ακόμα και στις ενδιαφερόμενες εταιρίες. Λόγω της μορφής των tweets, τα δεδομένα που δίνονται είναι πολύ χρήσιμα καθώς είναι δημόσια και παρουσιάζουν την στάση των ατόμων πάνω στο ζητούμενο θέμα. Σε κάποιες περιπτώσεις όμως δίνεται η δυνατότητα πρόσβασης και σε μη-δημόσια δεδομένα, όπως τα προσωπικά μηνύματα (Direct Messages). Για την προστασία των δεδομένα η πρόσβαση αυτή γίνεται σε προγραμματιστές που έχουν πάρει την άδεια να προβούν στην εξόρυξη αυτών των προσωπικών στοιχείων. (Twitter API)

Πιο συγκεκριμένα, υπάρχουν τρεις κατηγορίες εξουσιοδότησης για δωρεάν πρόσβαση στο Twitter API.

Οι κατηγορίες με τα χαρακτηριστικά τους είναι οι παρακάτω:

1. Essential:

- Μπορείς να κάνεις εξόρυξη Tweet έως 500.000 τον μήνα,
- Επιτρέπεται η δημιουργία ενός μόνο project,
- Επιτρέπεται η δημιουργία ενός μόνο App για κάθε project,
- Δεν υπάρχει δυνατότητα πρόσβασης στο standard v1.1, premium v1.1 ή στο enterprise.

2. Elevated:

- Έχεις επιπλέον πρόσβαση σε endpoints και δεδομένα καθώς επίσης και σε διάφορα περιβάλλοντα App,
- Δίνεται η δυνατότητα εξόρυξης έως 2 εκατομμυρίων Tweet τον μήνα,
- Επιτρέπεται η δημιουργία ενός μόνο project,
- Επιτρέπεται η δημιουργία τριών App για κάθε project,
- Επιτρέπεται η πρόσβαση στο standard v1.1, premium v1.1 ή στο enterprise.

3. Academic Research:

- Με την παροχή των ακαδημαϊκών στοιχείων, ένας φοιτητής ή ακαδημαϊκός ερευνητής έχει πρόσβαση σε περισσότερα δεδομένα και ανεπτυγμένα endpoint.
- Υπάρχει δυνατότητα εξόρυξης έως 10 εκατομμυρίων Tweet κάθε μήνα,
- Πρόσβαση σε όλα τα αρχεία και δεδομένα του Twitter,
- Πρόσβαση σε όλους τους προηγμένους τελεστές αναζήτησης.

Εκτός των βασικών αυτών κατηγοριών, το Twitter API προσφέρει ακόμα τρεις κατηγορίες πρόσβασης στα δεδομένα. Αυτές είναι οι ακόλουθες:

4. Enterprise Gnip 2.0 :

- Ιδανικό πακέτο για εταιρίες που βασίζονται στις γνώμες των χρηστών του twitter και θέλουν περαιτέρω πρόσβαση και ευχέρεια από ότι προσφέρει το Standard και Premium πακέτο. Παρέχει αξιοπιστία και αποτελείται από αφοσιωμένους βοηθούς και τεχνικούς υποστήριξης.

5. Premium v1.1:

- Δίνει τη δυνατότητα στους προγραμματιστές να έχουν πρόσβαση σε προηγμένα χαρακτηριστικά όπως η βελτιωμένη αναζήτηση ώστε να αναπτυχθούν ταχύτερα.

6. Standard v1.1:

- Είναι από τα πρώτα δωρεάν Twitter API και δημιουργήθηκε το 2012. Είναι η βασική πρόσβαση ενός προγραμματιστή το API ώστε να συλλέξει δεδομένα.

Πρόσβαση στα δεδομένα του Twitter.

Εφόσον ο προγραμματιστής ή η εταιρία ή ο ερευνητής προβεί στην αίτηση αδειοδότησης για πρόσβαση στο Twitter API και τα δεδομένα του στο εκάστοτε πακέτο που τον ενδιαφέρει θα χρειαστεί να απαντήσει σε κάποιες βασικές ερωτήσεις. Οι ερωτήσεις αυτές αφορούν τους λόγους που θέλει ο ενδιαφερόμενος να έχει πρόσβαση στα δεδομένα και ποια η χρήση τους. Όταν η διαδικασία ολοκληρωθεί, μέσω email η πλατφόρμα ενημερώνει αν η αίτηση έγινε δεκτή ή απορρίφθηκε.

Στη περίπτωση που γίνει δεκτή η αίτηση απευθείας ο χρήστης έχει πρόσβαση στο Twitter API και μπορεί να δημιουργήσει το Project του και το App του.

Τέλος, κάθε project έχει κάποια μοναδικά «κλειδιά» - keys που δίνουν πρόσβαση στον χρήστη στα δεδομένα.

Τα κλειδιά αυτά είναι:

- a) API KEY και API SECRET: θεωρούνται ως το username και ο κωδικός του APP ώστε να έχει πρόσβαση το πρόγραμμα στην αναζήτηση των δεδομένων
- b) ACCESS TOKEN και ACCESS TOKEN SECRET: είναι οι μοναδικοί κωδικοί του χρήστη για επαλήθευση και έλεγχο των στοιχείων πριν δοθεί πρόσβαση στα δεδομένα του Twitter για το APP.

Sentiment Analysis σε Tweets με χρήση της Γλώσσας Προγραμματισμού Python

Όπως αναφέρθηκε στο κεφάλαιο 2, η επεξεργασία και η ανάλυση των δεδομένων που θα συλλεχθούν από το Twitter θα γίνουν με τη χρήση της Γλώσσας Προγραμματισμού Python.

Στην εργασία αυτή, οι εντολές θα γραφτούν στο εξειδικευμένο περιβάλλον της Google Research, το Colaboratory (ή Colab).



Εικόνα 9 Logo του Google Colab

Το Google Colab είναι ένα δωρεάν προϊόν που δίνει την δυνατότητα στους προγραμματιστές να γράψουν κώδικα σε γλώσσα Python σε ένα εύκολο περιβάλλον. Επίσης, χαρακτηριστικό του Google Colab είναι η αποθήκευση του κώδικα σε Cloud και συγκεκριμένα στο Google Drive του χρήστη.

Διαδικασία έρευνας

Θεωρήθηκε αξιοσημείωτο να γίνει ανάλυση σε δύο διαφορετικά στάδια καθώς στο Twitter είναι σημαντικές οι δημοσιεύσεις τόσο όλων των χρηστών όσο και συγκεκριμένων ατόμων που επηρεάζουν σημαντικά την κοινή γνώμη (influencers).

Οι κατηγορίες τι οποίες θα ερευνήσουμε είναι οι παρακάτω.

Sentiment Analysis σε Tweet με συγκεκριμένο Hashtag

Η γνώμη και η στάση των χρηστών είναι ο κυριότερος λόγος υλοποίησης μιας έρευνας Συναισθηματικής Ανάλυσης. Για να βρούμε ευκολότερα την στάση των χρηστών απέναντι στην Αγορά των Κρυπτονομισμάτων θα αναζητήσουμε κάποια από τα κρισιμότερα Hashtag (#) του διαδικτύου που αφορούν την αγορά.

Ιδιαίτερη έμφαση θα δώσουμε στα παρακάτω:

- i. #crypto
- ii. #cryptocurrency
- iii. #trading
- iv. #blockchain
- v. #eth - αφορά το κρυπτονόμισμα “ETH” το οποίο γεννάται από την πλατφόρμα Ethereum η οποία λειτουργεί με blockchain τεχνολογία. Το ETH είναι το δεύτερο σε αξία κρυπτονόμισμα και θεωρείται από τα πιο ασφαλή στην αγορά. (Frankenliefeld J, 2022)
- vi. #bitcoin – αφορά το μεγαλύτερο σε αξία αποκεντρωμένο κρυπτονόμισμα σήμερα. Δημιουργήθηκε το 2009 και θεωρείται το ασφαλέστερο κρυπτονόμισμα καθώς ο αλγόριθμος που κρύβεται πίσω από τη λειτουργία του είναι κρυφός και αδιαπέραστος. (Frankenliefeld J, 2021)

Sentiment Analysis σε Tweet συγκεκριμένου χρήστη-influencer στον χώρο των Κρυπτονομισμάτων.

Στην εποχή που η πληροφορία και η επιρροή είναι άμεση, χρίζει μεγάλο ενδιαφέρον να γίνει Ανάλυση Συναισθήματος σε σημαντικά άτομα που καθορίζουν την κοινωνία σε μεγάλο βαθμό. Πιο συγκεκριμένα, βάσει της εργασίας αυτής, κρίνεται αναγκαίο να αναλύσουμε την στάση κάποιων ανθρώπων επιρροής στην Αγορά των Κρυπτονομισμάτων.

Αυτοί είναι οι εξείς:

- i. **Vitalik Buterin (@VitalikButerin)** – προγραμματιστής και συνιδρυτής ενός από τα μεγαλύτερα και σημαντικότερα κρυπτονομίσματα , το Ethereum.
- ii. **Roger Ver (@rogerkver)**- είναι από τους πρώτους επενδυτές του Bitcoin από το 2011. Μέχρι σήμερα θεωρείται ως “ο Θεός του Bitcoin”
- iii. **Adam Back (@adam3us)**- είναι από τους πιο γνωστούς και καταξιωμένους κρυπτογράφους και ιδρυτής της blockchain εταιρίας “Blockstream”.
- iv. **Andreas Antonopoulos (@aantonop)**- Ελληνοβρετανός συγγραφέας και επιχειρηματίας στο χώρο της τεχνολογίας. Είναι γνωστός για τις ομιλίες του σχετικά με τα Κρυπτονομίσματα και κυρίως για τη θέση του απέναντι στο Bitcoin.
- v. **Changpeng Zhao (@cz_binance)**- είναι ο ιδρυτής της Binance, του μεγαλύτερου παγκοσμίως ανταλλακτηρίου Κρυπτονομισμάτων, βάσει του ημερήσιου όγκου συναλλαγών.

Προτού ξεκινήσει η παρουσίαση της έρευνας και των αποτελεσμάτων της θα παρουσιάσουμε την διαδικασία που χρειάστηκε να τεθεί.

Για να γίνει κατανοητή έρευνα θα παρουσιαστούν τα βασικά βήματα με το παράδειγμα του Case Study για το “#crypto”. Αντίστοιχα θα παρουσιαστούν και τα βήματα για την εξόρυξη γνώμης από τα προφίλ κάποιων συγκεκριμένων χρηστών που έχουν επιρροή στην Αγορά των Κρυπτονομισμάτων, με το παράδειγμα του χρήστη “@VitalikButerin”.

Τέλος, στο επόμενο κεφάλαιο θα γίνει παρουσίαση των αποτελεσμάτων όλων των περιπτώσεων.

Τα Βασικά Βήματα για Sentiment Analysis σε Tweet με συγκεκριμένο Hashtag

Βήμα 1^ο :Εισαγωγή Βιβλιοθηκών – Import Libraries

Ξεκινώντας, στο notebook του Colab δημιουργήσαμε κώδικα εισαγωγής των βιβλιοθηκών που είναι της παρακάτω μορφής:

```
[ ] #Import the libraries
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
```

Κώδικας 1 Εισαγωγή Βιβλιοθηκών

Για να εισάγουμε οποιαδήποτε βιβλιοθήκη στην Python , αρκεί να γράψουμε την λέξη **import**.

Βήμα 2^ο: Εισαγωγή Κωδικών του Twitter API app

Το δεύτερο βήμα που ακολουθούμε σε κάθε στάδιο είναι να γράψουμε στο πρόγραμμά μας τους «κωδικούς» και την άδεια για να έχει πρόσβαση στο Twitter API. Οπότε, δίνουμε τα Keys και την εξουσιοδότηση που μας είχε δώσει το Twitter Developer για το Project μας.

```
[ ] #Get the Twitter API credentials
consumerKey = "aY535XcqGFyTgqhjL5H7AeASw"
consumerSecret = "MmBTgxZuiHQb7uSIMNfoZZTXdE7ZwcjozsqaK2t1212u13RGIU"
accessToken = "324268217-psnscsQoCEmlPEi50kwNlqTHFrrQlgPVZJz39uvBe"
accessTokenSecret = "fXFHG9Qk1X8vg04iFbSkzJW67dvsnoQgABajER8tFOImD"
```

```
#Create the authentication object
authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
#Set the access token and access token secret
authenticate.set_access_token(accessToken, accessTokenSecret)
#Create the API object
api = tweepy.API(authenticate,wait_on_rate_limit=True)
```

Κώδικας 2 Εισαγωγή Κωδικών του Twitter API App

Σε κάθε μια μεταβλητή εισχωρούμε το αντίστοιχο μοναδικό κλειδί. Ως “consumerKey” θεωρείται το “username” του app μας στο Twitter Developer και ως “consumerSecret” το “password”. Στις μεταβλητές “accessToken” και “accesssTokenSecret” έχουμε δώσει τους κωδικούς ταυτοποίησης.

Βήμα 3^ο: Εξόρυξη 2000 Tweet

Συνεχίζοντας, θα ακολουθήσει η παρουσίαση του κώδικα για κάθε μελέτη περίπτωσης ώστε να γίνει κατανοητή η Συναισθηματική Ανάλυση των χρηστών.

```
#Gather 2000 Tweets about Cryptocurrency Market
search_term= '#crypto -filter:retweets'
#Create a cursor object
tweets= tweepy.Cursor(api.search, q=search_term, lang='en', tweet_mode='extended').items(2000)
#Store the tweets in a variable and get the full text
all_tweets= [tweet.full_text for tweet in tweets]
```

Κώδικας 3 Εξόρυξη 2000 Tweet με #crypto

Ζητάμε από το πρόγραμμα να μας εξορύξει 2000 Tweet που φέρουν μέσα στο κείμενό τους το Hashtag #crypto . Επιπλέον για να μην γίνεται επανάληψη των δημοσιεύσεων, αφαιρούμε τη δυνατότητα εξόρυξης των “retweet” με την εντολή “-filter:retweets”.

Επιπρόσθετα, ζητήσαμε από το πρόγραμμα να μας φέρει όλα τα κείμενα που είναι στην αγγλική γλώσσα. Αυτό το επιλέξαμε ώστε μην γίνεται υπερφόρτωση στα δεδομένα με στοιχεία που δεν μπορούμε να κατανοήσουμε.

Αυτή η διαδικασία θα ακολουθηθεί σε κάθε περίπτωση αυτής της κατηγορίας Συναισθηματικής Ανάλυσης και το μόνο που θα διαφέρει είναι η μεταβλητή “search_term” διότι θα λαμβάνει άλλο Hashtag κάθε φορά.

Βήμα 4^ο: Αποθήκευση των Δεδομένων σε ένα Dataframe

Για να μπορέσουμε να δούμε τα Tweet που συλλέχθηκαν θα τα αποθηκεύσουμε σε ένα Dataframe , στην στήλη με όνομα Tweets.

```
#Create a Dataframe to store the tweets with columns called "Tweets"
df= pd.DataFrame(all_tweets, columns=['Tweets'])
#Show the first 5 rows of data
df.head(5)
```

Κώδικας 4 Αποθήκευση των Δεδομένων σε ένα Dataframe (#crypto)

	Tweets
0	Trade and make more #Bitcoin on autopilot. Or ...
1	@SHIBAMETA_BSC @GoodProject #SHIBAMETA #Airdro...
2	@CryptoTownEU I really believe that this proje...
3	@Crypto_Buddy2 Partners are very important for...
4	@airdropinspect Wow, Talkaboat is an amazing p...

Για να ελέγξουμε αν έγινε σωστά η διαδικασία, ζητήσαμε να μας δείξει μόνο τις πρώτες πέντε σειρές.

Βήμα 5^ο: «Καθαρισμός» των Tweet

Λόγω ότι τα Tweet είναι υπερφορτωμένα με στοιχεία όπως υπερ-συνδέσμους , είτε αναφορές (@) και hashtag (#), θα χρειαστεί να γίνει μια εκκαθάριση αυτών ώστε το κείμενο- Tweet να γίνει πιο ευανάγνωστο.

Ο κώδικας είναι ο εξής:

```
[ ] #Create a function to clean the data
def cleanTxt(text):
    text= re.sub('@[A-Za-z0-9]+', '', text)
    text= re.sub('#[A-Za-z0-9]+', '', text)
    text= re.sub('https?:\/\/\S+', '', text)
    text= re.sub('\n', '', text)
    text= re.sub('[A-Za-z0-9]#[A-Za-z0-9] +', '', text)
    return text
```

Κώδικας 5 «Καθαρισμός» των Tweet (#crypto)

Για να γίνει η εκκαθάριση των δεδομένων, δημιουργήσαμε μια συνάρτηση-function, με την οποία αντικαθιστούμε κάποια στοιχεία με την εντολή « ' ' ». Σημασιολογικά, αντικαθιστούμε κάποιες περιττές φράσεις, χαρακτήρες ή αντικείμενα ένα «κενό» κείμενο. Είναι δηλαδή, μια εντολή για να «διαγράψουμε» αυτές τις φράσεις.

Πιο συγκεκριμένα, με την εντολή «`text= re.sub('@[A-Za-z0-9]+', "", text)`» αντικαθιστούμε οτιδήποτε ξεκινάει με τον τελεστή «@» και ακολουθείται από γράμματα τις αλφαβήτα είτε αριθμούς , με μια κενή φράση.

Ομοίως ισχύει για τον τελεστή «#».

Στην περίπτωση των υπερ-συνδεσμων (hyperlinks), δίνουμε την εντολή «`text= re.sub('https?:\\S+', "", text)`» με την οποία, κάθε σύνδεσμος που ξεκινάει με “http” είτε με το “https” και ακολουθείται από οποιαδήποτε φράση , αντικαθίσταται επίσης από το κενό « ' ' ».

Αφού δώσαμε τις εντολές να καθαριστούν τα κείμενα, θα πρέπει να τεθεί σε εφαρμογή στο Dataframe μας, που συμπεριλαμβάνει τα Tweet μας.

```
#Cleaning the Text
df['Tweets'] = df['Tweets'].apply(cleanTxt)

#Show the Cleaned Text
df
```

	Tweets
0	Trade and make more on autopilot. Or make more...
1	_BSC \$SHIBA ❤️🚀🚀🚀_lal Excellent project an...
2	I really believe that this project will reall...
3	_Buddy2 Partners are very important for every ...
4	Wow, Talkaboat is an amazing project. It is v...
...	...
1995	. \$ftm to the moon! 🚀 For more Visuals 🙌 Stay ...
1996	Almost a quarter of small businesses in nine c...
1997	Good project 🙌 I'm sure this project will grow...
1998	The project is implemented professionally and...
1999	I still trust this one

2000 rows x 1 columns

Όπως φαίνεται στο παράδειγμα, έχουν αφαιρεθεί όλες οι ετικέτες (#), οι αναφορές (@) και οι υπερ-συνδέσμοι από τα Tweet μας. Μέσω αυτής της διαδικασίας θα είναι ευκολότερη και γρηγορότερη η επεξεργασία τους.

Κώδικας 6 «Καθαρισμός» των Tweet (#crypto)

Βήμα 6^ο: Εύρεση της «Πόλωσης» (Polarity) και της «Υποκειμενικότητας» (Subjectivity) των Tweet.

Μέσω της βιβλιοθήκης TextBlob και της συνάρτησης συναισθήματός της, επιστρέφονται δύο σημαντικοί παράγοντες, το Subjectivity και το Polarity. Και οι δυο αυτοί παράγοντες καθορίζουν την Ανάλυση Συναισθήματος .

Αρχικά, από την μία πλευρά είναι η «Υποκειμενικότητα» , Subjectivity, μέσω της οποίας αναφέρεται η προσωπική γνώμη, τα συναισθήματα και η κρίση του χρήστη. Ακολουθεί τη μέθοδο float() δηλαδή μετατρέπεται από έναν αριθμό που είναι αποθηκευμένος σε μια συμβολοσειρά (string) ή σε ακέραια μορφή (integer) σε αριθμό κινητής υποδιαστολής. Τέλος, το εύρος του Subjectivity είναι [0,1].

Από την άλλη , έχουμε την «Πολικότητα» , Polarity, η οποία επίσης ακολουθεί τη μέθοδο float() και με εύρος τιμών [-1,1]. Δηλαδή, με τιμή '-1' η δήλωση είναι αρνητική και αντίστοιχα η τιμή '1' είναι θετική.

Ο κώδικας που υπολογίζει το Polarity και το Subjectivity είναι:

```
#Create a function to get Subjectivity
def getSubjectivity(text):
    return TextBlob(text).sentiment.subjectivity

#Create a function to get Polarity
def getPolarity(text):
    return TextBlob(text).sentiment.polarity

#Create two columns called "Subjectivity" and "Polarity"
df['Subjectivity'] = df['Tweets'].apply(getSubjectivity)
df['Polarity'] = df['Tweets'].apply(getPolarity)

#Show data
df.head()
```

Κώδικας 7 Εύρεση της «Πόλωσης» (Polarity) και της «Υποκειμενικότητας» (Subjectivity) των Tweet. (#crypto)

Σαν παράδειγμα, τα αποτελέσματα που επιστρέφονται είναι της παρακάτω μορφής. Επιπλέον, δημιουργήσαμε ξεχωριστές στήλες στις οποίες καταχωρήθηκαν αντίστοιχα οι τιμές των δύο μεταβλητών.

	Tweets	Subjectivity	Polarity
0	Singapore has announced the closure of cryptoc...	0.000000	0.0
1	Hey, Please check out my oil painting itemFe...	0.800000	0.4
2	wonderful project sit,,join must all. _red \$...	0.500000	0.5
3	Hey, Please check out my oil painting itemFe...	0.800000	0.4
4	What are you waiting for, guys!Hop in early an...	0.633333	0.5

Πίνακας 1 Subjectivity and Polarity (#crypto)

Βήμα 7^ο: Οπτικοποίηση αποτελεσμάτων με την μέθοδο του Word Cloud.

Τα Tweets στο σύνολό τους αναφέρονται στο ίδιο θέμα και σχετίζονται μεταξύ τους με κοινές λέξεις.

Η βιβλιοθήκη του Word cloud μας δίνει την δυνατότητα να δούμε σε σχήμα ποιες είναι αυτές οι λέξεις που εμφανίζονται συχνότερα στα δεδομένα μας.

Ακόμα, κάθε μια από αυτές τις λέξεις θα έχει διαφορετικό μέγεθος γραμματοσειράς βάσει της συχνότητας εμφάνισής της. Δηλαδή, η λέξη που εμφανίζεται περισσότερες φορές στα Tweet θα έχει το μεγαλύτερο μέγεθος. Η διαδικασία που ακολουθούμε λέγεται “PLOT”.

Ο Κώδικας στην Python που μας δίνει τα οπτικοποιημένα αποτελέσματα είναι:

```
#How well the sentiments are distributed
#PLOT the word cloud

allWords = ' '.join( [twts for twts in df['Tweets']] )
wordCloud = WordCloud(width= 500, height= 300, random_state= 21, max_font_size= 119).generate(allWords)

plt.imshow(wordCloud, interpolation = "bilinear")
plt.axis('off')
plt.show()
```

Κώδικας 8 Οπτικοποίηση αποτελεσμάτων με την μέθοδο του Word Cloud.

Σαν Αποτέλεσμα στην περίπτωση αυτή έχουμε την παρακάτω εικόνα.



WordCloud 1 #crypto

Παρατηρούμε ότι η λέξη “crypto” έχει τη μεγαλύτερη συχνότητα στα tweet μας. Να σημειωθεί ότι δεν καταμετράται το “#crypto” καθώς στο βήμα 5 είχαμε «καθαρίσει» τα δεδομένα μας από όλες τις αναφορές (@), ετικέτες (#) και περιττά σύμβολα που μπορεί να επηρέαζαν την επεξεργασία και ανάλυση τους.

Βήμα 8^ο: Υπολογισμός του βαθμού «Συναισθήματος»

Σε αυτό το βήμα δημιουργούμε μια μεταβλητή με την οποία υπολογίζουμε τον βαθμό Συναισθήματος του κειμένου. Ειδικότερα, αυτό είναι το βασικό στάδιο της Sentiment Analysis καθώς βρίσκουμε την στάση του χρήστη απέναντι στο ζητούμενό μας.

Για να γίνει Συναισθηματική Ανάλυση στα Tweet μας δώσαμε τις παρακάτω εντολές και λάβαμε τα εξής αποτελέσματα.

```
#Create a function to get the text sentiment score

def getAnalysis (score):
    if score<0:
        return 'Negative Sentiment'
    elif score ==0:
        return 'Neutral Sentiment'
    else:
        return 'Positive Sentiment'

#Create a column for the Sentiment

df['Analysis'] = df['Polarity'].apply(getAnalysis)

#Show data
df
```

Κώδικας 9 Υπολογισμός του βαθμού «Συναισθήματος»

	Tweets	Subjectivity	Polarity	Analysis
0	Singapore has announced the closure of cryptoc...	0.000000	0.000000	Neutral Sentiment
1	Hey, Please check out my oil painting itemFe...	0.800000	0.400000	Positive Sentiment
2	wonderful project sit,,join must all. _red \$...	0.500000	0.500000	Positive Sentiment
3	Hey, Please check out my oil painting itemFe...	0.800000	0.400000	Positive Sentiment
4	What are you waiting for, guys!Hop in early an...	0.633333	0.500000	Positive Sentiment
...
1995	Looking for \$gem tokens of 2022, then \$ANGEL ...	0.700000	0.300000	Positive Sentiment
1996	🔥 UNITY THE BEST #UTILITY/PROJECT OF THIS YE...	0.150000	0.500000	Positive Sentiment
1997	👉 Do you read on platform? ! Learn how to effec...	0.433333	0.366667	Positive Sentiment
1998	🐳 \$ADA Whale ❤️ laced!💰 Transaction Output: 1.5...	0.000000	0.000000	Neutral Sentiment
1999	This is what do. Isn't that cool ? Let's make a...	0.000000	0.000000	Neutral Sentiment

2000 rows x 4 columns

Πίνακας 2 Υπολογισμός του βαθμού «Συναισθήματος» (#crypto)

Υπάρχουν τρεις κατηγορίες στις οποίες κατατάσσεται η στάση του χρήστη όσο αφορά την Συναισθηματική Ανάλυση.

Η πρώτη είναι κατηγορία είναι το αρνητικό συναίσθημα (Negative Sentiment) και παρουσιάζεται όταν η Πολικότητα (Polarity) είναι μικρότερη από το μηδέν. Η δεύτερη περίπτωση είναι όταν το Polarity είναι ίσο με το μηδέν, όπου θεωρείται ως ουδέτερο συναίσθημα (Neutral Sentiment). Τέλος αν η ανάλυση του Polarity. Είναι μεγαλύτερη από το μηδέν τότε η στάση του χρήστη θεωρείται θετική (Positive Sentiment).

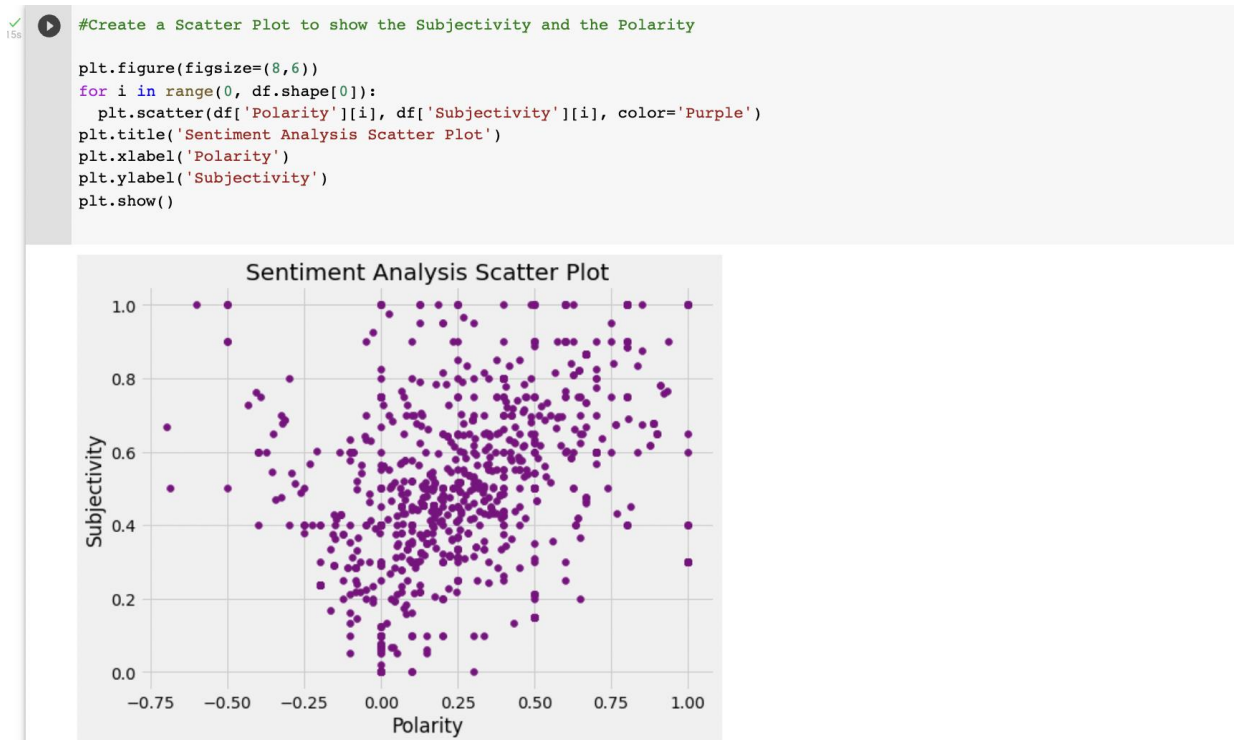
Στο παράδειγμα έχουμε καταχωρήσει σε νέα στήλη το αποτέλεσμα της ανάλυσης.

Βήμα 9^ο: Δημιουργία ενός Scatter Plot για να δείξουμε το Subjectivity και το Polarity

Για έναν ερευνητή ή επιχειρήση είναι πολύ σημαντικό να έχει μια εικόνα των αποτελεσμάτων του. Στην περίπτωση της Συναισθηματικής Ανάλυσης δημιουργούμε ένα scatter plot για να δούμε ως κυμαίνονται τα δεδομένα.

Πιο συγκεκριμένα, στο Scatter Plot μας θα υπάρχουν δύο άξονες . ένας θα έχει αναφορές για το Subjectivity και ο άλλος για το Polarity του κάθε Tweet.

Ο Κώδικας που θα μας δώσει αυτή τη λύση είναι ο παρακάτω.



Κώδικας 10 Δημιουργία ενός Scatter Plot

ScatterPlot 1 #crypto

Αναλυτικότερα, όσα tweet έχουν αρνητικό score στο polarity, δηλαδή έχουν αρνητικό Sentiment , βρίσκονται αριστερά της τιμής 0.00 . Στο συγκεκριμένο σχήμα φαίνεται ότι οι τιμές κυμαίνονται δεξιά του 0.00 , το οποίο προσδίδει μια θετική Συναισθηματική στάση των χρηστών.

Βήμα 10^ο: Εμφάνιση όλων των θετικών Tweet

Εκτός από τα σχεδιαγράμματα και τα οπτικοποιημένα αποτελέσματα, χρήσιμα είναι και τα ίδια τα Tweet. Μια επιχείρηση που ενδιαφέρεται για την γνώμη των χρηστών, καταγράφει και διαχωρίζει τα θετικά από τα αρνητικά.

Επομένως, ο κώδικας που θα μας δώσει τα θετικά Tweet είναι:

Με μια συνάρτηση for, ζητάμε από το πρόγραμμα να βρει όσα Tweet έχουν “Positive Sentiment” και να μας το δείξει στην οθόνη μας.

```
#Print all the positive tweets
j=1
sortedDF= df.sort_values(by=['Polarity'])
for i in range(0, sortedDF.shape[0]):
    if(sortedDF['Analysis'][i] == 'Positive Sentiment'):
        print(str(j) + ' ' +sortedDF['Tweets'][i])
        print()
        j= j+1
```

- 1) Hey, Please check out my oil painting itemFeel free to place your bid and check my collection
- 2) wonderful project sit,,join must all. _red \$ASTAR
- 3) Hey, Please check out my oil painting itemFeel free to place your bid and check my collection
- 4) What are you waiting for, guys!Hop in early and secure your \$SPO tokens and get awesome rewards!90% APY staking pool is just mind blowing 🤖🚀
- 5) The X99 Game Team of this task is very hardworking and promising. I aid this undertaking due to the fact I be aware that in the close to future
- 6) \$BTC \$ETH \$ATOM \$NEO \$ONE \$VET \$SRM \$MIR \$FIL \$SUSHI \$FTT \$FTM \$BNB \$LUNA \$SHIB \$DOGE \$THETA \$HNT \$ALICE \$ZEC \$FLUX \$NEAR \$THETA \$AUDIO
- 7) Have you ever heard of the Fidelis Crypto Loyalty Points platform? This is a tool that offers rewards every time a customer consumes at your es
- 8) Get \$25 today by joining one of the best Crypto Exchanges to date using this referral link 🚀
- 9) Look at the percent of profits hahaha👉👉👉👉👉👉👉👉👉👉 This is so amazing to watch as it pumps constantly! The Fastest profit yielding project!🔥
- 10) 🚀 ADA rallied ahead of the coin's Sundaeswap launch, taking the top spot in the C20 index!Click on the link in our bio to learn more.
- 11) Don't Buy Cryptocurrency; Buy These 3 Stocks Instead - Motley Fool
- 12) Don't Buy Cryptocurrency; Buy These 3 Stocks Instead - Motley Fool Buy BTC:
- 13) You should consider buying today. Best project to date. \$AXN #Metaverse

Κώδικας 11 Εμφάνιση όλων των θετικών Tweet #crypto

Με τον ίδιο τρόπο το πρόγραμμα μας επέστρεψε τα αρνητικά και ουδέτερα Tweet.

```
#Print all the negative tweets
j=1
sortedDF= df.sort_values(by=['Polarity'], ascending='False')
for i in range(0, sortedDF.shape[0]):
    if (sortedDF['Analysis'][i]=='Negative Sentiment'):
        print(str(j)+' ' + sortedDF['Tweets'][i])
        print()
        j=j+1
```

- 31) alright... i m addicted to my moonmen series...
- 32) _ is a collection of 8,000 NFTs NFT holder's benefits - Ticket to enter a regular giveaway lottery - Way to get in the whitelist of our blockc
- 33) GM, Has money stopped exiting the crypto markets?We look at what is to come if the down trend holds on the next rally attemptLIVE IN 25 Minute
- 34) shown a but if strength near the support , looks like a small retracement only. Currently, Binance Launchpad event is about to end which might
- 35) Approximately 5.7 million BTC are now underwater (-30% of circulating supply).#BTCUSD
- 36) A reminder is both a pyramid type scheme & a method to launder money & secret enable criminal activity from human tracking all the way
- 37) Cryptocurrencies are a key focus for ASIC in 2022 due to the number of people exposed to violent market fluctuations and scams from lack of un
- 38) Notice how the previous dump is similar to our current one? Only difference is that we haven't seen any capitulation yet.. Waiting for big vol
- 39) Blockchain Firm Animoca Brands Raises \$358 Million to Enhance Web3 and the Metaverse - Finance Bitcoin News #worldnews
- 40) Drop your address. Randomly drop 🚀 🗳️ Just Retweet + Follow
- 41) Another rare card is out! "CRAZY MONKEY #0015" 🐵_eth
- 42) What's happening Why all this massive \$OOE Getting worst than a Bear Run 🐻🐻#BSC

Κώδικας 12 Εμφάνιση όλων των αρνητικών Tweet (#crypto)

```

#Print all the Neutral tweets
j=1
sortedDF= df.sort_values(by=['Polarity'])
for i in range(0, sortedDF.shape[0]):
    if(sortedDF['Analysis'][i] == 'Neutral Sentiment'):
        print(str(j) + ' ' +sortedDF['Tweets'][i])
        print()
        j= j+1

169) BitYard: Cryptocurrency Trading in a Licensed Environment - CryptoPotato Buy BTC:
170) BitYard: Cryptocurrency Trading in a Licensed Environment - CryptoPotato
171) $FDBL Surpasses Operational and Growth Milestones, Providing Shareholders with a 2021 Year in Review Update and 2022 Strategy for Artist, Ser
172) LTO Network is a hybrid blockchain for securing, verifying and exchanging business critical information. Choose LTO Network!
173) Rand App - Earning Passive Income Has Never Been Easier. Sign up for our waitlist and get your share of $50,000 in prizes! via
174) _ is a collection of 8,000 NFTs NFT holder's benefits - _phap
175) Today's DropTreasure Hunt Art - Minnaar Money No.00499 - "InvisibleMinnaar" Portrait SeriesDetails at
176) Cryptocurrency Bitcoin To The Moon Bars--
177) Because the blockchain puts your crypto to work, it generates incentives (form of passive income) while it is staked.$doge

```

Κώδικας 13 Εμφάνιση όλων των ουδέτερων Tweet (#crypto)

Βήμα 11^ο: Στατιστική Ανάλυση Συναισθήματος

Σε κάθε περίπτωση, κρίνεται αναγκαία η στατιστική ανάλυση των αποτελεσμάτων της έρευνας. Ποιο είναι δηλαδή, το ποσοστό όλων των θετικών, αρνητικών και ουδέτερων Tweet και πως αυτά παρουσιάζονται σε ένα απλό διάγραμμα για να γίνει κατανοητό.

Αρχικά, θα ζητήσουμε από το πρόγραμμα να μας δώσει το ποσοστό των Tweet για κάθε κατηγορία αντίστοιχα.

```

[16] #Get the percentage(%) of the positive Tweets
ptweets=df[df.Analysis=='Positive Sentiment']
ptweets=ptweets['Tweets']

round( (ptweets.shape[0]/ df.shape[0])*100,1)

57.9

```

```

[17] #Get the percentage (%) of the negative Tweets
ptweets=df[df.Analysis=='Negative Sentiment']
ptweets=ptweets['Tweets']

round( (ptweets.shape[0]/ df.shape[0])*100,1)

7.0

```

```

[18] #Get the percentage (%) of the neutral Tweets
ptweets=df[df.Analysis=='Neutral Sentiment']
ptweets=ptweets['Tweets']

round( (ptweets.shape[0]/ df.shape[0])*100,1)

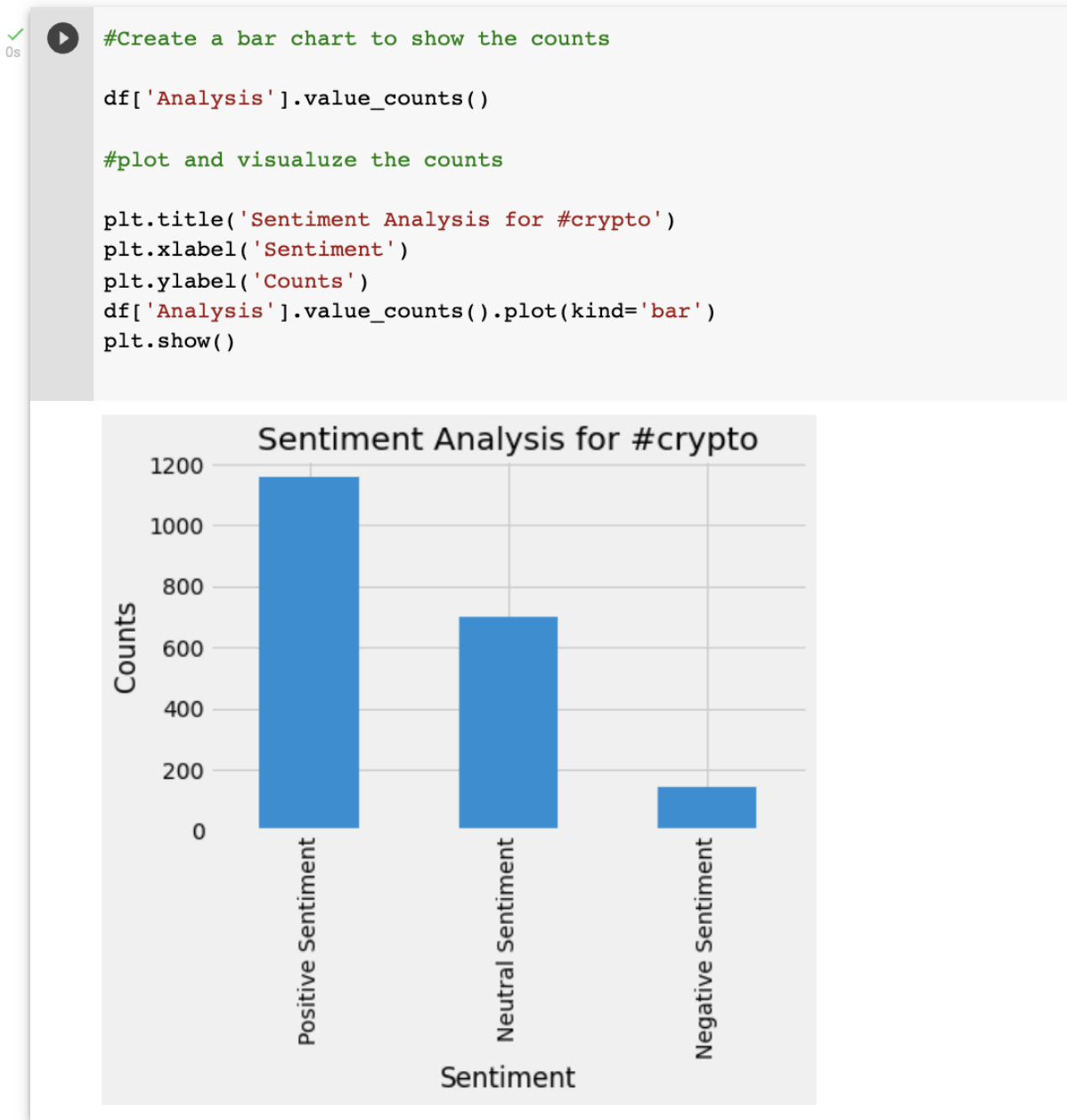
35.1

```

Παρατηρούμε ότι το 57.9% των Tweet που περιέχουν το hashtag #crypto έχουν θετική στάση. Μόλις το 7% παρουσιάζει αρνητικό Sentiment. Και το υπόλοιπο 35.1% φαίνεται να είναι ουδέτερο. Σε πολλές περιπτώσεις, το πρόγραμμα κατά τη διάρκεια της ανάλυσης «επηρεάζεται» από λέξεις κλειδιά. Επίσης αρκετά Tweet που βρίσκονται στην κατηγορία του Neutral Sentiment είναι και ανεπιθύμητα ή spam. Για αυτό το λόγο είναι προτιμότερο τα Neutral σε Sentiment, Tweet μια επιχείρηση να τα παρουσιάζει για να έχει πλήρη εικόνα.

Κώδικας 14 Στατιστική Ανάλυση Συναισθήματος (#crypto)

Τέλος, το Chart που δημιουργήσαμε για να δούμε πως παρουσιάζεται το πλήθος των ευρημάτων μας δίνεται με τον κώδικα:



Κώδικας 15 Δημιουργία Διαγράμματος

Chart 1 #crypto

Από τα 2000 Tweets που συλλέχθηκαν με το #crypto, αυτή είναι η κατανομή της συναισθηματικής ανάλυσής τους .

Τα βασικά βήματα για Sentiment Analysis σε Tweet συγκεκριμένου χρήστη-influencer στον χώρο των Κρυπτονομισμάτων.

Βήμα 1^ο :Εισαγωγή Βιβλιοθηκών – Import Libraries

Η εισαγωγή των βιβλιοθηκών είναι το πρώτο και κυριότερο βήμα για να ξεκινήσει η εξόρυξη των δεδομένων.

```
3 ▶ #Import the libraries
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
# pip install -U spacy
# python -m spacy download en_core_web_sm
import spacy
nlp = spacy.load("en_core_web_sm")
import seaborn as sns
from nltk.stem.snowball import SnowballStemmer
```

Κώδικας 16 Εισαγωγή Βιβλιοθηκών

Σε αυτήν την περίπτωση εισάγουμε τρεις επιπλέον βιβλιοθήκες, την “SpaCy”, την “Seaborn” και την “SnowballStemmer”, οι οποίες είναι NLP βιβλιοθήκες, και θα βοηθήσουν να εξάγουμε συμπεράσματα από τα κείμενα των χρηστών.

Βήμα 2^ο: Εισαγωγή Κωδικών του Twitter API app

Για να έχουμε πρόσβαση στα δεδομένα του Twitter , θα πρέπει να δώσουμε τα απαραίτητα στοιχεία για να δοθεί η άδεια από το Twitter API. Επομένως, τα ίδια KEYS που χρησιμοποιήσαμε και στη πρώτη περίπτωση μας δίνεται η εξουσιοδότηση για να λάβουμε τα Tweets που ζητάμε.

```
[40] #Get the Twitter API credentials
consumerKey = "aY535XcqGFyTgqhJL5H7AeASw"
consumerSecret = "MmBTgxZuiHQB7uSIMNfoZZTXdE7ZwcjozsqaK2t1212u13RGIU"
accessToken = "324268217-psnCsQoCEmlPEi5OkwNlqTHFrrQlgPVZJz39uvBe"
accessTokenSecret = "fXFHG9Qk1X8vgO4iFbSkZJW67dvsnoQgABajER8tFOImD"
```

```
[41] #Create the authentication object
authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
#Set the access token and access token secret
authenticate.set_access_token(accessToken, accessTokenSecret)
#Create the API object
api= tweepy.API(authenticate,wait on rate limit=True)
```

Κώδικας 17 Εισαγωγή Κωδικών του Twitter API app

Βήμα 3ο: Εξόρυξη 2000 Tweet από έναν user

Σαν παράδειγμα θα λάβουμε τον χρήστη @VitalikButerin.

Αφού γίνει εξόρυξη των Tweet του από τον ζητούμενο user, έπειτα τα δεδομένα τα αποθηκεύουμε σε ένα Dataframe, στο οποίο παρουσιάζονται τόσο τα κείμενα όσο και τα Like που έχει λάβει το κάθε Tweet.

```
[59] number_of_tweets=2000
tweets= []
likes= []

for i in tweepy.Cursor(api.user_timeline, id="VitalikButerin",tweet_mode="extended").items(number_of_tweets):
    tweets.append(i.full_text)
    likes.append(i.favorite_count)

df= pd.DataFrame({'Tweets': tweets, 'Likes': likes})
df
```

	Tweets	Likes
0	I'm speaking in 3 min!	5778
1	@zengjiajun_eth 其实人类文明的很多部分是去中心化的。一个很明显的例子: 语言\...	573
2	@DeepCipher @shl @elonmusk Disparities in econ...	4005
3	@owocki @ivanmolto retro funding for public go...	87
4	@MacaesBruno Latin America seems to me the mos...	47
...
1995	@ArthurB @QiaochuYuan (sorry about the confusi...	3
1996	@ArthurB @QiaochuYuan Likely the most realisti...	9
1997	@ArthurB @QiaochuYuan You live on twitter (and...	3
1998	@WgRSSbxbdbNN1cU @zooko 他应该在美国吧, 如果你在深圳你的早上是他的晚上	0
1999	@MihailoBjelic I'm assuming the worst case: la...	2

2000 rows x 2 columns

Κώδικας 18 Εξόρυξη 2000 Tweet από έναν user
Πίνακας 3 Tweet @VitalikButerin

Βήμα 4ο: «Καθαρισμός» των Tweet

```
[98] #Clean the Texts
def cleanTxt(text):
    text= re.sub('[A-Za-z0-9]+', '', text)
    text= re.sub('[A-Za-z0-9]+', '', text)
    text= re.sub('https?:\\/\n', '', text)
    return text

[99] df=df[-df.Tweets.str.contains("RT")]

[100] #Cleaning the Text
df['Tweets'] = df['Tweets'].apply(cleanTxt)

#Show the Cleaned Text
df
```

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

	Tweets	Likes
0	I'm speaking in 3 min!	5780
1	_eth 其实人类文明的很多部分是去中心化的。一个很明显的例子: 语言几乎所有的...	573
2	Disparities in economic success between men...	4005
3	retro funding for public goods + project tok...	87
4	Latin America seems to me the most clearly Ep...	47
...
1995	(sorry about the confusion from the referent...	3
1996	Likely the most realistic way that tech elit...	9
1997	You live on twitter (and other provinces of ...	3

Για να είναι ευδιάκριτα και έτοιμα για επεξεργασία τα δεδομένα μας, θα πρέπει να «καθαριστούν». Με αυτόν τον τρόπο αφαιρούνται περιττά στοιχεία από τα κείμενα που μπορεί να δώσουν λάθος αποτελέσματα στη συνέχεια.

Κώδικας 19 «Καθαρισμός» των Tweet

Βήμα 5^ο: Δημιουργία ενός διαγράμματος και του Word Cloud του, που θα παρουσιάζει τις λέξεις που χρησιμοποιεί συχνότερα ο χρήστης

```
[79] list_of_sentences= [sentence for sentence in df.Tweets]

lines=[]
for sentence in list_of_sentences:
    words = sentence.split()
    for w in words:
        lines.append(w)
```

```
[80] lines= [re.sub(r'[^A-Za-z0-9]+', '', x) for x in lines]

lines2=[]

for word in lines:
    if word != '':
        lines2.append(word)
```

```
[81] s_stemmer= SnowballStemmer (language='english')

stem=[]
for word in lines2:
    stem.append(s_stemmer.stem(word))
```

```
[82] stem2=[]
for word in stem:
    if word not in nlp.Defaults.stop_words:
        stem2.append(word)
```

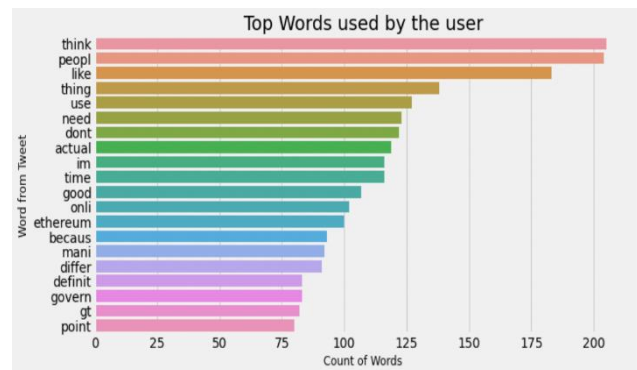
```
[83] df2= pd.DataFrame(stem2)

df2=df2[0].value_counts()
```

```
[84] df2= df2[:20,]
plt.figure(figsize=(10,5))
sns.barplot(df2.values, df2.index, alpha=1)
plt.title('Top Words used by the user')
plt.ylabel('Word from Tweet',fontsize=12)
plt.xlabel('Count of Words',fontsize=12)
plt.show()
```

Από τα 2000 Tweets που συλλέχθηκαν, φαίνεται και από το διάγραμμα και από το word Cloud ότι συχνότερα χρησιμοποιούνται οι λέξεις:

- > People
- > Think
- > Time
- > ethereum

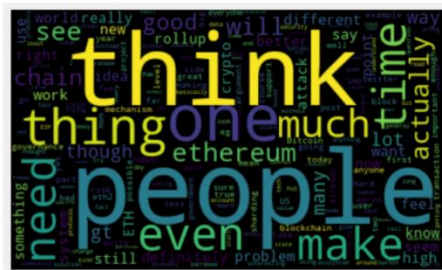


Κώδικας 21 Δημιουργία διαγράμματος Top Words Used

Chart 2 Top Words Used by @VitalikButerin

```
[86] allWords = ' '.join([twts for twts in df['Tweets']])
wordCloud = WordCloud(width= 500, height= 300, random_state= 21, max_font_size= 119).generate(allWords)

plt.imshow(wordCloud, interpolation = "bilinear")
plt.axis('off')
plt.show()
```



Κώδικας 20 Δημιουργία WordCloud

WordCloud 2 @VitalikButerin

Βήμα 6^ο: Υπολογισμός του βαθμού «Συναισθήματος»

Οι μεταβλητές Subjectivity και Polarity είναι αυτές που θα μας οδηγήσουν στο συμπέρασμα της ερευνάς μας για Συναισθηματική Ανάλυση.

Θα μας δείξουν αν οι χρήστες στο κάθε Tweet, έχουν θετική (Positive), αρνητική (Negative) ή ουδέτερη (Neutral) στάση.

```
[89] def getAnalysis (score):
      if score<0:
          return 'Negative Sentiment'
      elif score ==0:
          return 'Neutral Sentiment'
      else:
          return 'Positive Sentiment'

      #Create a column for the Sentiment
      df['Analysis'] = df['Polarity'].apply(getAnalysis)

      #Show data
      df

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:11: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
# This is added back by InteractiveShellApp.init_path()


```

	Tweets	Likes	Subjectivity	Polarity	Analysis
0	I'm speaking in 3 min!	5778	0.000000	0.000000	Neutral Sentiment
1	_eth 其实人类文明的很多部分是去中心化的。一个很明显的例子：语言几乎所有的事情不是二选...	573	0.000000	0.000000	Neutral Sentiment
2	Disparities in economic success between men...	4005	0.519167	0.189167	Positive Sentiment
3	retro funding for public goods + project tok...	87	0.066667	0.000000	Neutral Sentiment
4	Latin America seems to me the most clearly Ep...	47	0.576667	0.120000	Positive Sentiment
...
1995	(sorry about the confusion from the referent...	3	0.602056	0.075081	Positive Sentiment
1996	Likely the most realistic way that tech elit...	9	0.557576	0.210606	Positive Sentiment
1997	You live on twitter (and other provinces of ...	3	0.404167	0.109754	Positive Sentiment
1998	他应该在英国吧，如果你在深圳你的早上是他的晚上	0	0.000000	0.000000	Neutral Sentiment
1999	I'm assuming the worst case: large rollups wh...	2	0.609524	-0.278571	Negative Sentiment

Κώδικας 22 Υπολογισμός του Βαθμού Συναισθήματος
Πίνακας 4 Sentiment Analysis για @VitalikButerin

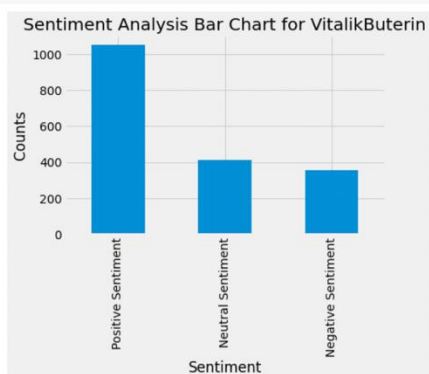
Βήμα 7^ο: Στατιστική Ανάλυση Συναισθήματος

Έχοντας υπολογίσει τις τιμές για το Subjectivity και το Polarity, θα δημιουργήσουμε ένα Chart Bar και ένα Pie Chart για να παρουσιαστούν οπτικοποιημένα τα αποτελέσματα.

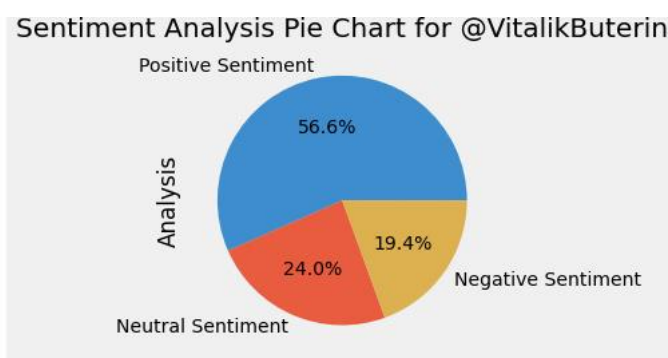
```
[91] #Create a bar chart to show the counts
      df['Analysis'].value_counts()

      #plot and visualize the counts

      plt.title('Sentiment Analysis Bar Chart for VitalikButerin')
      plt.xlabel('Sentiment')
      plt.ylabel('Counts')
      df['Analysis'].value_counts().plot(kind='bar')
      plt.show()
```



Κώδικας 23 Στατιστική Ανάλυση
Chart 3 @VitalikButerin



Pie Chart 1 @VitalikButerin

Κεφάλαιο 4^ο: Αποτελέσματα Έρευνας

Case Studies

Sentiment Analysis σε 2000 Tweet με συγκεκριμένο Hashtag

Σε όλες τις περιπτώσεις ακολουθούνται τα βήματα που αναφέρθηκαν στο προηγούμενο κεφάλαιο.

Επομένως, σε αυτό το κεφάλαιο θα εξηγήσουμε τα αποτελέσματα που φέρνουν σε βαθμό συναισθήματος κάθε tweet με συγκεκριμένο hashtag χωρίς να παρουσιάζουμε ολοκληρωμένα την διαδικασία του κώδικα.

i. Case Study for Sentiment Analysis in Tweets with the #cryptocurrency

Αρχικά, έγινε καθαρισμός και προβολή των Tweet.

```
#Cleaning the Text
df['Tweets'] = df['Tweets'].apply(cleanTxt)

#Show the Cleaned Text
df
```

	Tweets
0	of the dayquick look at-1 crypto players seek ...
1	A specialised form of cryptography, known as a...
2	Things change. And friends leave. Life doesn't...
3	[Spot] By 2022-01-19 14:00 GMT, DOGE total tra...
4	Falcore found in a User vault at this location...
...	...
1995	Check out my new NFT on OpenSea! via #NFT
1996	[Spot] By 2022-01-19 14:00 GMT, ATOM had a pri...
1997	Best project ever. I am very happy to be part...
1998	_pay It's very good project and strong team. I...
1999	[Futures] By 2022-01-18 22:00 GMT, \$XLM future...

2000 rows x 1 columns

Πίνακας 5 Καθαρισμός Tweet #cryptocurrency

```
[8] #Create a function to get Subjectivity

def getSubjectivity(text):
    return TextBlob(text).sentiment.subjectivity

#eate a function to get Polarity

def getPolarity(text):
    return TextBlob(text).sentiment.polarity

#Create two columns called "Subjectivity" and "Polarity"

df['Subjectivity'] = df['Tweets'].apply(getSubjectivity)
df['Polarity'] = df['Tweets'].apply(getPolarity)

#Show data

df.head()
```

Στη συνέχεια υπολογίσαμε τις μεταβλητές Subjectivity και Polarity ώστε να υπολογιστεί η Ανάλυση συναισθήματος.

	Tweets	Subjectivity	Polarity
0	of the dayquick look at-1 crypto players seek ...	0.466667	0.166667
1	A specialised form of cryptography, known as a...	0.458333	-0.041667
2	Things change. And friends leave. Life doesn't...	0.000000	0.000000
3	[Spot] By 2022-01-19 14:00 GMT, DOGE total tra...	0.272222	0.000000
4	Falcore found in a User vault at this location...	1.000000	1.000000

Πίνακας 6 Polarity and Subjectivity #cryptocurrency

Έπειτα δημιουργήσαμε το Word Cloud για να δούμε ποιες λέξεις παρουσιάζονται συχνότερα. Στην περίπτωση αυτή η λέξεις “cryptocurrency” και “project” και “spot” κατέχουν τις πρώτες θέσεις. Σαν αποτέλεσμα είναι λογικό καθώς είναι όροι που χρησιμοποιούνται από τα άτομα που ασχολούνται με την αγορά των Κρυπτονομισμάτων.

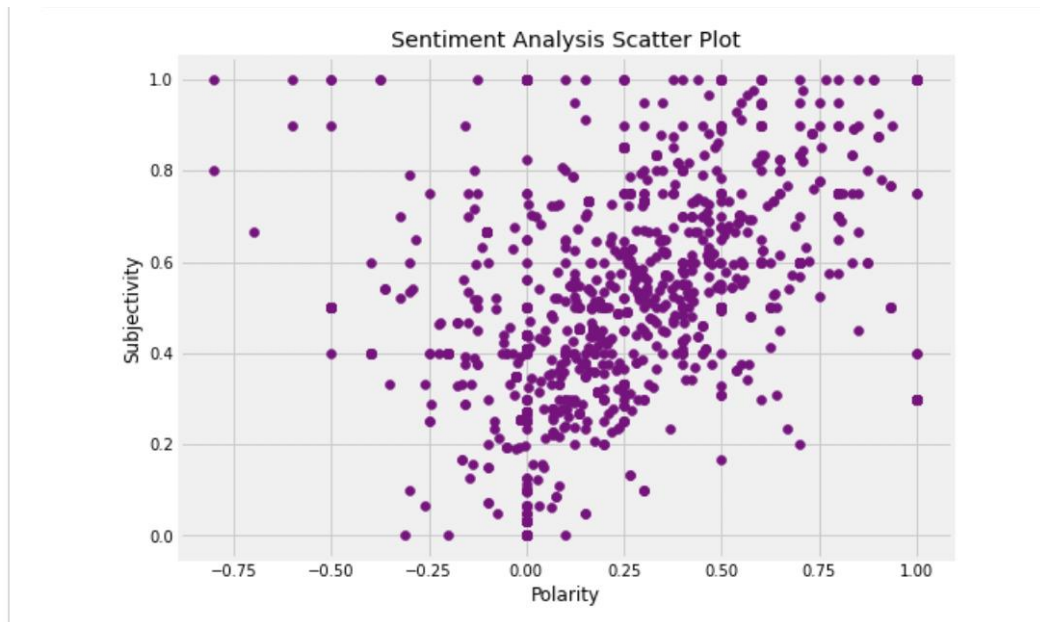


WordCloud 3 #cryptocurrency

Επιπλέον, ορίσαμε αν η στάση του χρήστη είναι θετική, αρνητική ή ουδέτερη.

index	Tweets	Subjectivity	Polarity	Analysis
0	of the dayquick look at-1 crypto players seek clarity on taxes from budget.2 google pay hires paypal exec to head up crypto payments push.3 individual miner hits jackpot with \$540 block reward.follow us for more such updates.	0.4666666666666666	0.1666666666666666	Positive Sentiment
1	A specialised form of cryptography, known as an Elliptical Curve Digital Signature Algorithm (ECDSA), underpins and other systems as a way of providing uneaten security and ensuring that funds can only be used by their rightful owners.	0.4583333333333333	-0.04166666666666666	Negative Sentiment
2	Things change. And friends leave. Life doesn't stop for anybodyHave you looked into \$TAF #cryptocurrency	0.0	0.0	Neutral Sentiment
3	[Spot] By 2022-01-19 14:00 GMT, DOGE total trading volume of 2,606,460,544DOGE and net cap flow of \$35M in last 24 hrs. \$DOGE	0.2722222222222222	0.0	Neutral Sentiment
4	Falcore found in a User vault at this location! Join me playing #coinhuntworld, It's awesome!	1.0	1.0	Positive Sentiment
5	Good luck team! I believe this is a faithful project.The projector has a lot of attractions so hopefully the project will be better in the future and will be the best_nil #Airdropstario	0.4083333333333333	0.4583333333333333	Positive Sentiment
6	Oh shit!! FINALLY adding \$SCRT/ \$USDT pair 🙌👉👈	0.9	-0.15625	Negative Sentiment
7	Volt Inu Now Available On BKEX as it prepares to List on LBank the next day _cro	0.2	0.2	Positive Sentiment

Πίνακας 7 Sentiment Analysis #cryptocurrency



ScatterPlot 2 #cryptocurrency

Από το Scatter Plot για τα Tweet με hashtag #cryptocurrency βλέπουμε μια θετική στάση των χρηστών. Ακόμα, είναι ευδιάκριτο ότι οι χρήστες με Negative Sentiment τείνουν να έχουν μεγαλύτερο Subjectivity, δηλαδή είναι έντονη η προσωπική τους γνώμη.

Τα στατιστικά στοιχεία των αποτελεσμάτων είναι ότι σχεδόν το μισό του δείγματος έχει θετικό Sentiment. Το 8.3% είναι αρνητικοί και το 35.8% παρουσιάζει Neutral στάση.

✓ [17] #Get the percentage(%) of the positive Tweets

```
ptweets=df[df.Analysis=='Positive Sentiment']
ptweets=ptweets['Tweets']

round( (ptweets.shape[0]/ df.shape[0])*100,1)
```

55.9

✓ [18] #Get the percentage (%) of the negative Tweets

```
ptweets=df[df.Analysis=='Negative Sentiment']
ptweets=ptweets['Tweets']

round( (ptweets.shape[0]/ df.shape[0])*100,1)
```

8.3

✓ [19] #Get the percentage (%) of the neutral Tweets

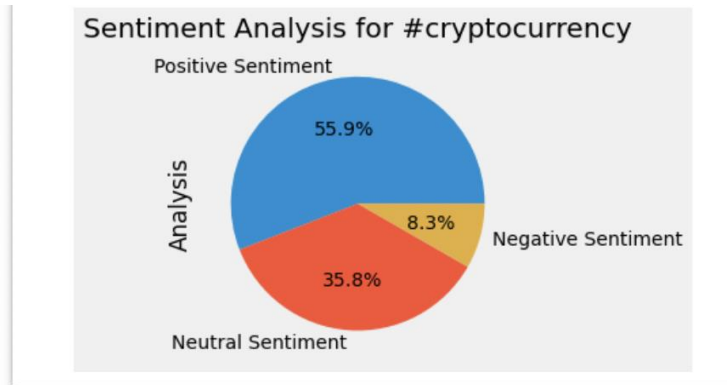
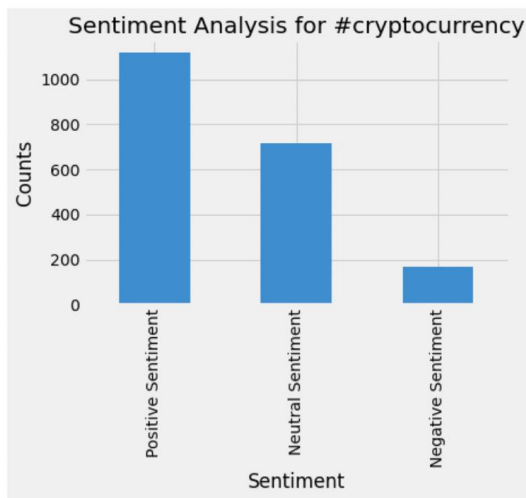
```
ptweets=df[df.Analysis=='Neutral Sentiment']
ptweets=ptweets['Tweets']

round( (ptweets.shape[0]/ df.shape[0])*100,1)
```

35.8

Κώδικας 24 Στατιστική Ανάλυση #cryptocurrency

Τέλος, δημιουργήσαμε τα σχεδιαγράμματα για την οπτική παρουσίαση των αποτελεσμάτων.



Pie Chart 2 #cryptocurrency

Chart 4 Sentiment Analysis #cryptocurrency

Ομοίως και στις υπόλοιπες περιπτώσεις:

- ii. Case Study for Sentiment Analysis in Tweets with the #trading

Το Word Cloud για τις σημαντικότερες λέξεις είναι:



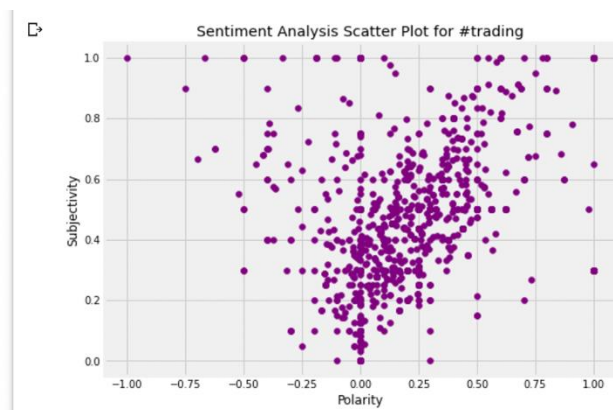
WordCloud 4 #trading

Ένα δείγμα των Tweet και των αποτελεσμάτων σε βαθμό Subjectivity και Polarity για το #trading είναι η παρακάτω λίστα.

ID	Tweet Text	Subjectivity	Polarity	Sentiment
23	Happy Sunday 🌞🌧️	1.0	1.0	Positive Sentiment
24	When markets open 300 points against your position but you chose to remain strong.	0.616666666666667	0.216666666666667	Positive Sentiment
25	🔥 best offer for all 🔥 open free demat account and get 📄 link to open demat and joint our community.....🌱🌿	0.533333333333333	0.466666666666667	Positive Sentiment
26	"The in comes in the preparation.The actual process of #trading, however, should be effortless." -	0.1	0.0	Neutral Sentiment
27	Don't try to fix the chart. Fix yourself in such a way that whatever chart does, you'll be fine. #Trading	0.5	0.208333333333334	Positive Sentiment
28	Trade of the week:Bought: QQQ put - strike \$366 1/24 \$692 on 1/20.Sold: next day for \$1,934.179% gain.\$QQQ \$SPY #options	0.0	0.0	Neutral Sentiment
29	Crypto Tax Trends in 2022: Increased Reporting, Updated Rules, and a Wealth Tax Debate #Cryptocurrency	0.0	0.0	Neutral Sentiment
30	Top Notch Charting Software for Traders	0.5	0.5	Positive Sentiment
31	5 Chinese Stocks Into Traditional Medicine	0.375	0.0	Neutral Sentiment

Πίνακας 8 Sentiment Analysis #trading

Η εικόνα του scatterplot για το #trading μας δείχνει ότι υπάρχει μεγαλύτερο πλήθος “Neutral” Tweet. Επιπλέον, τα «Θετικά» tweet κυμαίνονται με τις τιμές του Polarity σε ουδέτερα επίπεδα. Συνεπώς, είτε το “#trading” χρησιμοποιείται από χρήστες για θετικά Tweets , ή για να παρουσιάσουν μια καλή «αγορά» τους.



ScatterPlot 3 #trading

Τα διαγράμματα για την αναζήτηση με #trading μας επιβεβαιώνουν το scatter plot , δηλαδή ότι η συναισθηματική κρίση των χρηστών είναι θετική ή με ουδέτερη κύμανση.

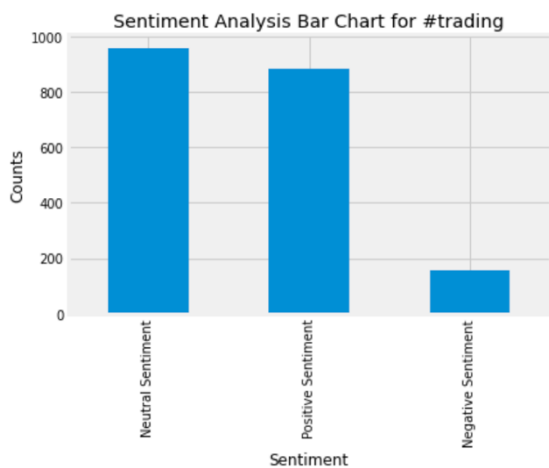
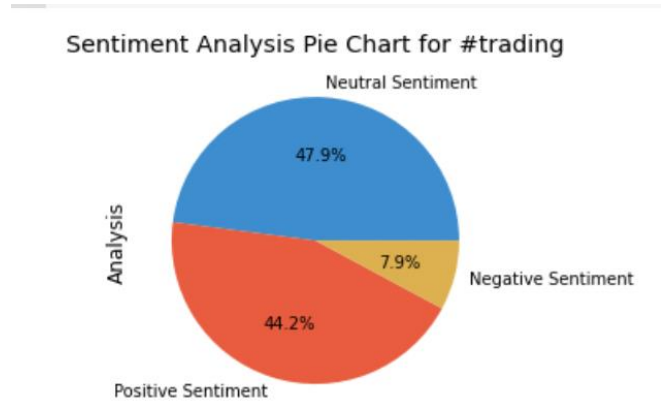


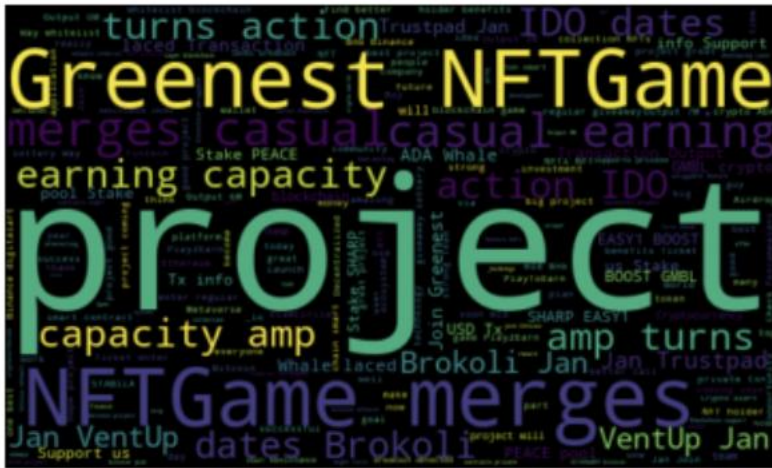
Chart 5 Sentiment Analysis #trading



Pie Chart 3 #trading

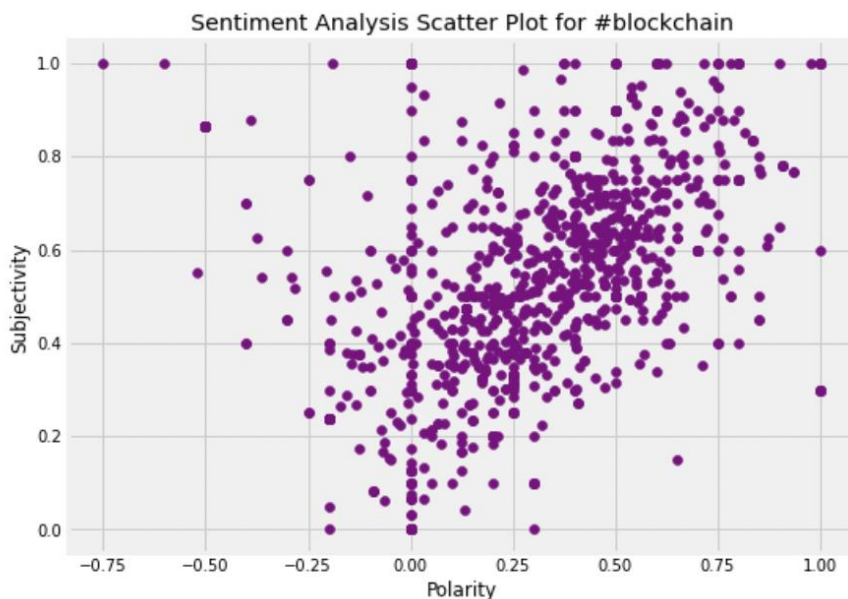
iii. Case Study for Sentiment Analysis in Tweets with the #blockchain

Το Word cloud που μας δείχνει ότι οι χρήστες που δημοσίευσαν Tweet με το #blockchain χρησιμοποιούν στα κείμενά τους περισσότερο τις λέξεις “project” και “NFTGame”. Και οι δύο αυτές λέξεις είναι keywords στην αγορά των Κρυπτονομισμάτων. Επίσης τον τελευταίο καιρό υπάρχει η τάση περιβαλλοντικής συνείδησης στην Αγορά, οπότε πολλά Tweet αναφέρονται στα «πράσινα» και “power friendly” project. Για αυτόν τον λόγο η λέξη “Greenest” τονίζεται τόσο συχνά.



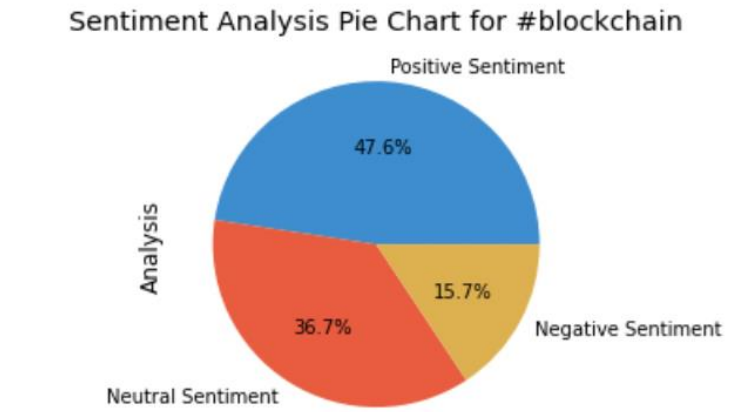
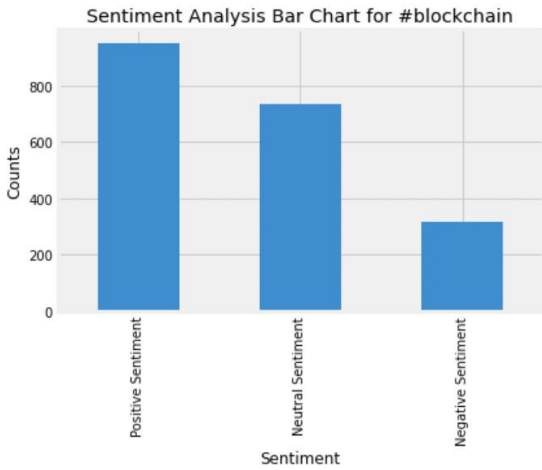
WordCloud 5 #blockchain

Μια εικόνα των tweet και πώς κυμαίνονται στον άξονα του Subjectivity και Polarity είναι η παρακάτω:



ScatterPlot 4 #blockchain

Τέλος, τα στατιστικά σχεδιαγράμματα για την Ανάλυση συναισθήματος των Tweet με #blockchain είναι:



Pie Chart 4 #blockchain

Chart 6 Sentiment Analysis #blockchain

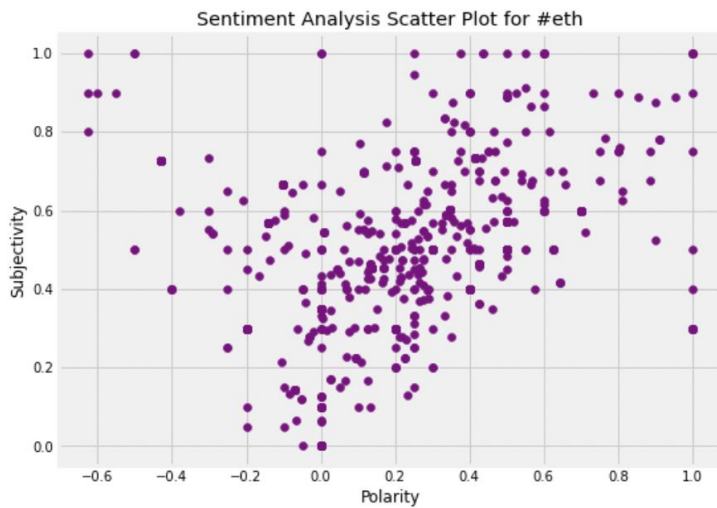
iv. Case Study for Sentiment Analysis in Tweets with the #eth

Το Word cloud για τα Tweet που αναφέρονται στο #eth έχει την παρακάτω μορφή.



WordCloud 6 #eth

Συνεχίζουμε με το Scatter Plot, και τα στατιστικά σχεδιαγράμματα για να γίνει πιο κατανοητή η Sentiment Analysis των σχολίων με #eth.



ScatterPlot 5 #eth

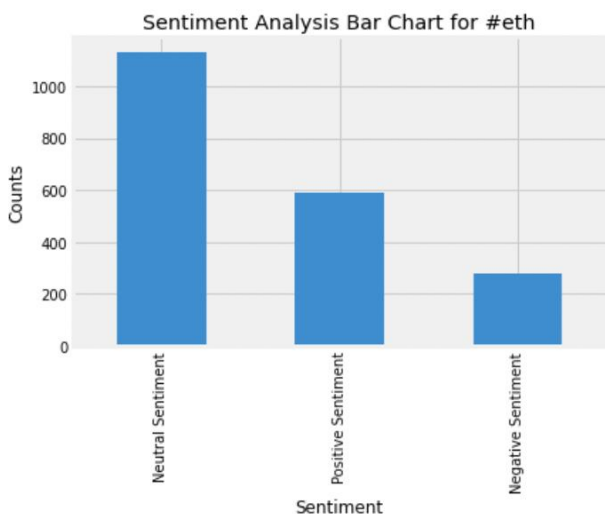
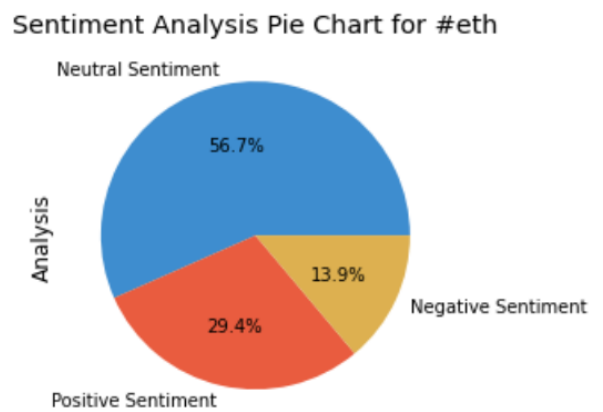


Chart 7 Sentiment Analysis #eth



Pie Chart 5 #eth

Όπως φαίνεται και στα διαγράμματα που ακολουθούν, τα Tweet με #bitcoin έχουν θετική στάση. Μόλις το 8.5% αυτών που συλλέχθηκαν ήταν αρνητικά.

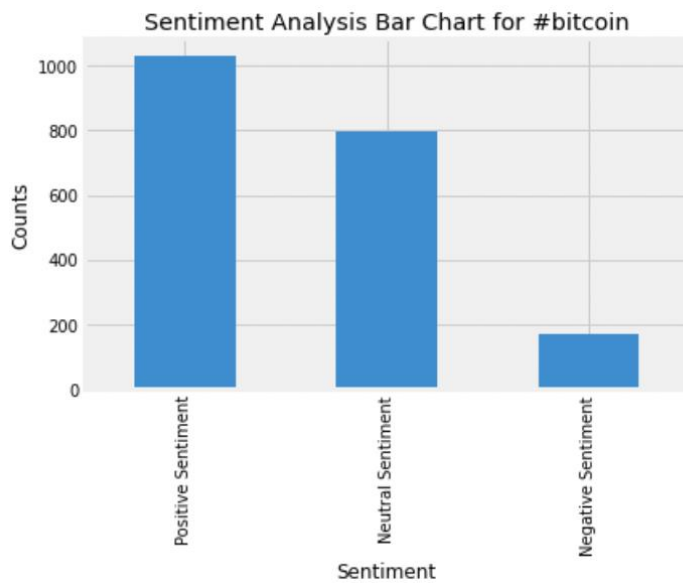
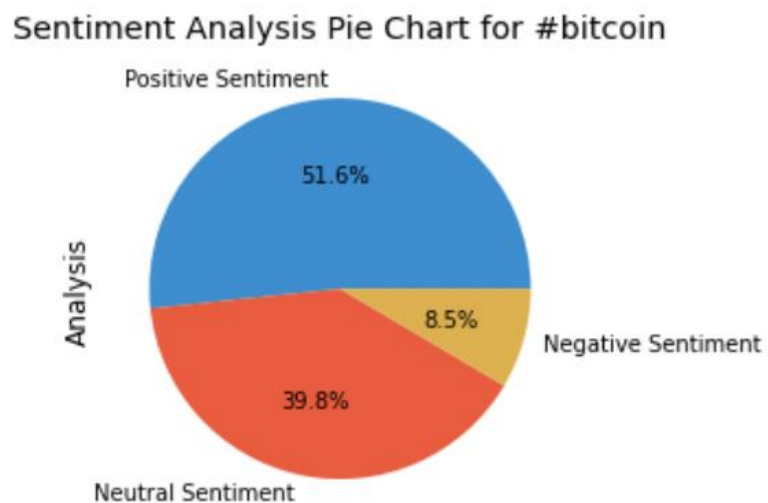


Chart 8 Sentiment Analysis #bitcoin



Pie Chart 6 #bitcoin

B. Sentiment Analysis σε Tweet συγκεκριμένου χρήστη-influencer στον χώρο των Κρυπτονομισμάτων.

vi. Case Study for Sentiment Analysis in Twitter user @rogerkver

Τα Tweets που λάβαμε από τον χρήστη @rogerkver αφού επεξεργάστηκαν είναι τα εξής:

index	Tweets	Likes
0	Today marks the 10 year anniversary of my first public mention of Bitcoin, and of the day Satoshi disappeared.	4368
1	Securities laws are outdated. Needing a license to issue a security token in 2021 is just as dumb as needing a licence to setup a website in 2001.	1704
2	It took about a decade for crypto to grow to a market cap of \$1 trillion, but took just one quarter to add another trillion.	2081
3	There is no inflation they say....	2180
4	Soon everyone will realize that the one megabyte block size limit on BTC was a mistake. Once everyone realizes it, everyone will pretend that they knew it the whole time. #BitcoinCash	1291
5	Today's and fees:	757
6	"when you give people freedom, you don't know what they'll do with it." - _Ross	413
7	Rather than following sheep, think for yourself.	1115
8	Renouncing your US citizenship is the most American thing you can do:	311
10	Utility trumps speculation in the long run. That's why	842
11	Join me in next month for the Nomad Capitalist Live conference.	282
12	It all started with a desire for a peer to peer electronic cash system to bring more economic freedom to the world. Congratulations on helping make it happen!	576
13	My 1,000 stickers I bought with just arrived. Buy yours at Add your business at	607
14	If you believe voluntary charity should replace government coercion, here is your chance to show it. I know Nathan and his family personally.	173
15	is the best way to allow #India's work force to participate in the global economy. Reply with your address from your wallet to see how it works.	315
16	It's time for BTC to give up their claim to "Bitcoin". It no longer resembles Bitcoin and it hasn't for years. The title of the whitepaper explains it all.	611
17	Voluntary trade is the foundation of human civilization.	860
18	Imagine being so brainwashed that you believe some humans need permission from other humans to transmit money. #cryptocurrency	2183
19	is not for sale.	776
20	Better cash makes a better store of value.	728
21	Rejecting the religion of Statism is more difficult than other religions because almost everyone currently believes in it. #Atheism	410
22	Meanwhile in California thousands of Bitcoin Cash stickers are being printed.	784
23	The only thing Bitcoin Maximalists like more than BTC is CENSORSHIP.	305
24	is accepted at more physical locations than any other Check for places near you.	466
25	Holding dollars feels riskier than holding cryptocurrency.	1456

Show [25] per page

Πίνακας 9 Tweets by @rogerkver

Στη συνέχεια δημιουργήσαμε το διάγραμμα και το Word Cloud για να δούμε την συχνότητα των λέξεων που χρησιμοποιεί ο χρήστης.

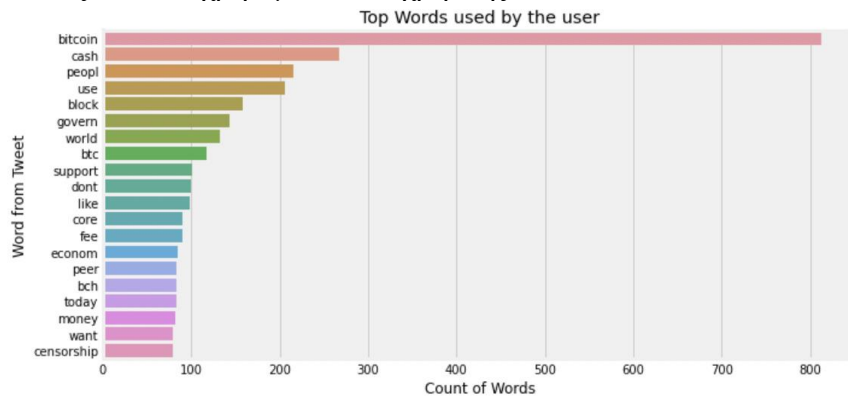
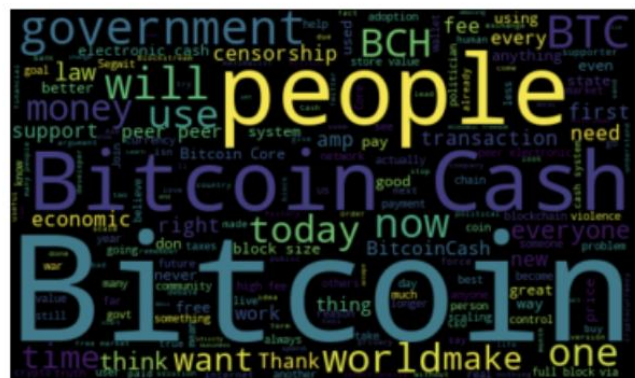


Chart 9 Top Wordst used by @rogerkver



WordCloud 8 @rogerkver

Τέλος, αφού υπολογίσαμε το Subjectivity και το Polarity για κάθε Tweet, δημιουργήσαμε τα διαγράμματα για την οπτικοποίησή τους. Όπως φαίνεται, πάνω από το 50% του δείγματος έχουν θετική στάση και μόλις το 14.5% αρνητική.

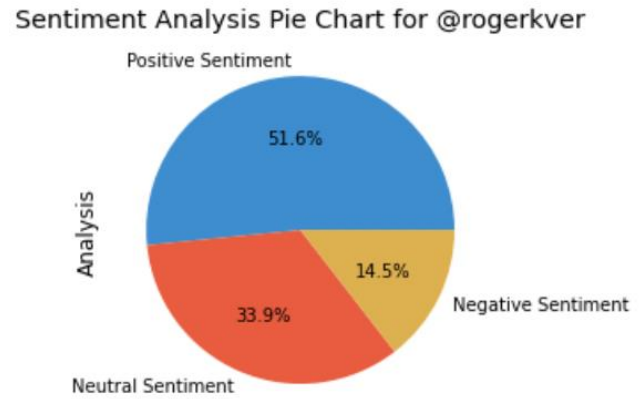
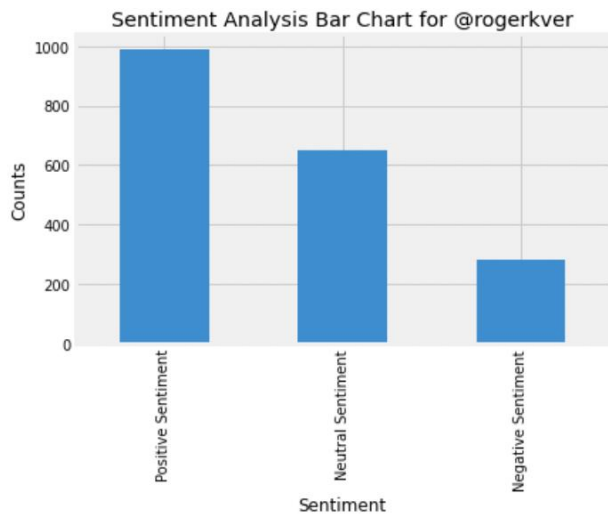


Chart 10 Sentiment Analysis @rogerkver

Pie Chart 7 @rogerkver

vii. Case Study for Sentiment Analysis in Twitter user @adam3us

Για τον χρήστη Adam Back, τα επεξεργασμένα δεδομένα που λάβαμε είναι:

	Tweets	Likes
0	and then the economy collapses, and they have...	4
2	{ 2^(224-d) > H(n,#bitcoin) } = 600mc^2	88
4	this is good! the plebs are taking coins from...	43
6	pun on FOMC coming tues/weds, US interest p...	10
7	Weekends are scams. (Low volume markets)	33
...
1995	You are maybe "selling" awareness and adopt...	2
1996	i think you maybe mixing simplicity with ...	3
1997	i didn't say stacks was operating it's c...	1
1998	liquid has fees. we're working on it. so...	12
1999	ah right that's the confusion.	9

1601 rows x 2 columns

Πίνακας 10 Tweets by @adam3us

Έπειτα, για να δούμε ποιες λέξεις χρησιμοποιήθηκαν περισσότερο μέσα στα κείμενα, φτιάξαμε το Word Cloud και το αντίστοιχο bar chart.

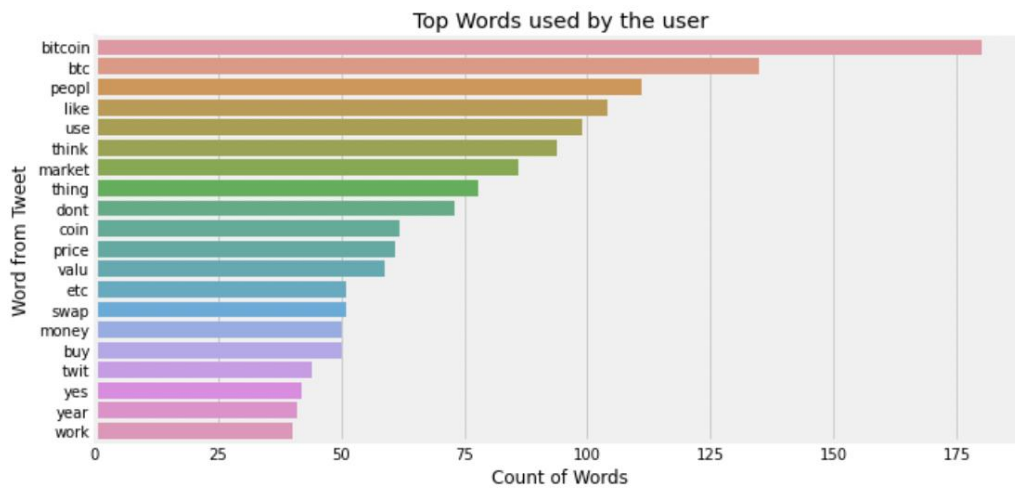


Chart 11 Top Words Used by @adam3us

Συμπεραίνουμε, ότι οι λέξεις “bitcoin”, “btc”, “market”, “think” και “people” είναι από τις πιο χρησιμοποιούμενες στα Tweet.



WordCloud 9 @adam3us

Τέλος, τα διαγράμματα που μας δείχνουν την συναισθηματική ανάλυση για τον @adam3us είναι τα παρακάτω και μας δείχνουν την θετική στάση του.

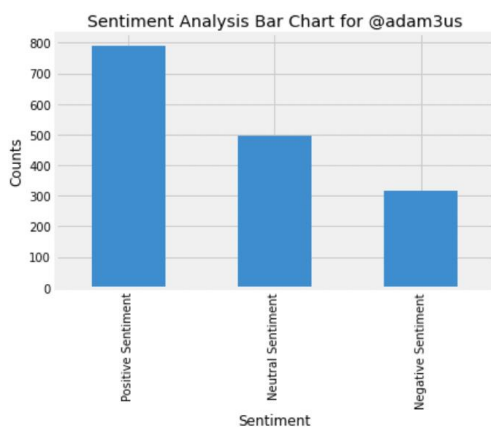
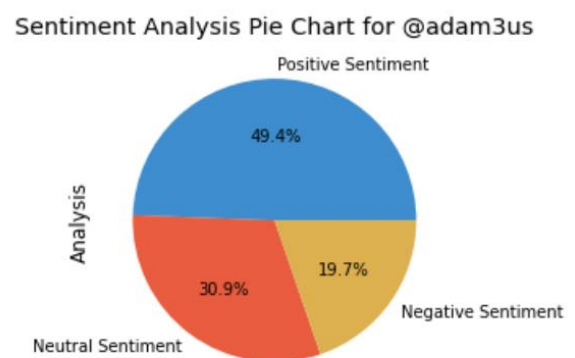


Chart 12 Sentiment Analysis @adam3us



Pie Chart 8 @adam3us

viii. Case Study for Sentiment Analysis in Twitter user @aantonop

Από τον κ. Αντωνόπουλο τα δεδομένα που λάβαμε είναι:

Index	Tweets	Likes
0	As enthusiasm for grows. have banks moved into the bargaining stage of grief? As more banks launch their own digital tokens, trying to jump on the train, they miss the point entirely. **	261
1	Aprende a planificar la herencia de tus #criptomonedas. Obten el libro de y comienza en tan solo 30 minutos. Disponible en tapa blanda y en formato electrónico	44
2	Learn how to do inheritance planning for your #cryptocurrency. It's easier than you think. Get the book by & get started in just 30 min. You'll be glad you did. Available in paperback and ebook	97
3	The Lightning Network can be layered on top of any Blockchain-based cryptocurrency, allowing for immediate transactions and low fees. Find out more about how it can solve scalability issues in Mastering the Network.	359
4	_bob Q4'2022 is the planned release for the third edition	5
5	** OPEN. BORDERLESS. DECENTRALIZED. CENSORSHIP-RESISTANT. Bitcoin isn't money for the internet, it's The Internet of Money. Learn why is more than what it seems at first glance 🔥	329
6	Bitcoin is not asking for permission. It doesn't need to. If you're new to #crypto, this video about is a great place to start. 🍷	572
7	Happy 2022! Moving forward, this account is no longer a personal account, but a team account. My team will be making announcements about what we're working on. Read more about this change:	170
8	Are Ethereum and Bitcoin competing with each other? and are both the apex predators of their own environments. 🤖📺 Learn more in our latest video:	264
9	Your private key doesn't leave a hardware wallet, so why do you need to update the firmware? In this video, Andreas explains how firmware updates improve security:	330
10	For years people have been asking me about inheritance planning and #crypto. Now there's a workshop that helps you start to build your own plan, taught by , the author of Cryptoasset Inheritance Planning book. Learn more:	269
11	For the last time, stop using paper wallets! Learn why in this video:	212
12	Get struck by a bolt of inspiration. Mastering the Lightning Network is now available through Amazon and O'Reilly Media. Get your copy:	291
13	ICYMI: , and I are excited to let you know that Mastering the Lightning Network, which is 450+ pages long, is finally available to buy through Amazon or O'Reilly Media. Get yours:	323

Πίνακας 11 Tweets by @aantonop

Στη συνέχεια δημιουργήσαμε το Word Cloud και το Bar Chart με τις λέξεις που χρησιμοποιήθηκαν περισσότερες φορές στα Tweet του χρήστη.

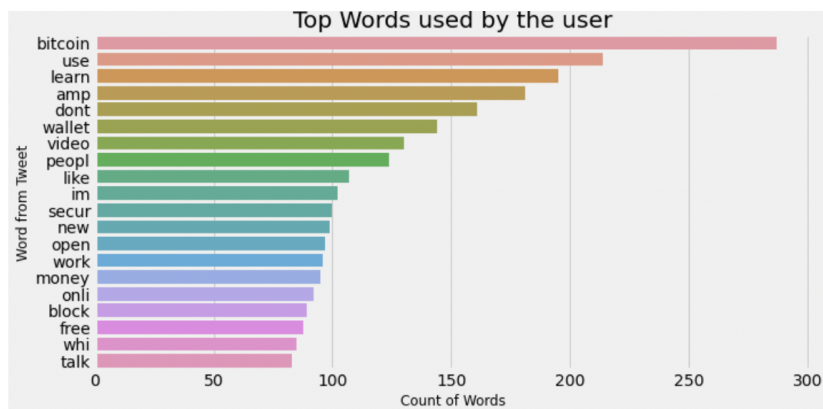
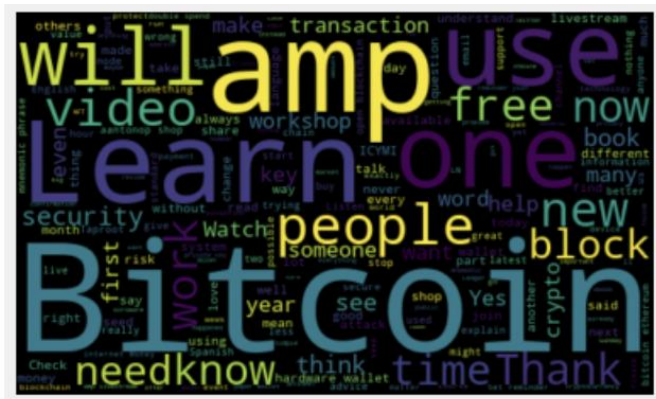


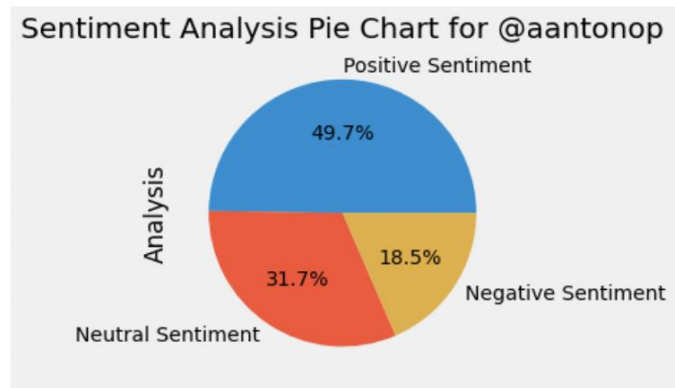
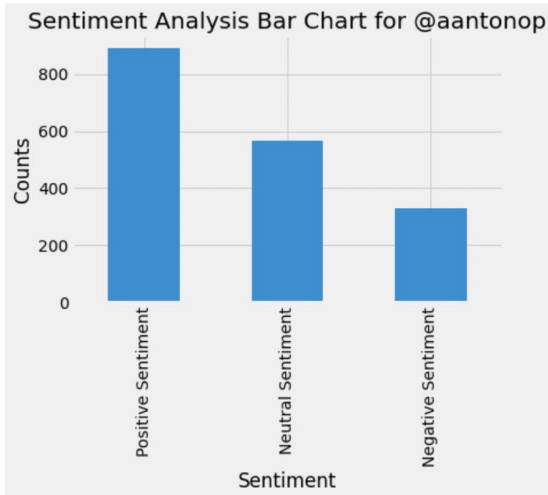
Chart 13 Top Words used by @aantonop



WordCloud 10 @aantonop

Στο τελευταίο στάδιο, αφού υπολογίστηκαν οι τιμές για το Subjective και το Polarity , οπτικοποιήσαμε τα αποτελέσματα για να δούμε την Συναισθηματική τάση των Tweet.

Όπως φαίνεται στα σχεδιαγράμματα, τα σχόλια τείνουν να είναι θετικά με ποσοστό 49.7%.



Pie Chart 9 @aantonop

Chart 14 Sentiment Analysis @aantonop

ix. Case Study for Sentiment Analysis in Twitter user @cz_binance

Από τον χρήστη @cz_binance , έγινε εξόρυξη και καθαρισμός των παρακάτω Tweet.

	Tweets	Likes
0	One of the rare times we disagree, haha. But ...	602
1	No drama. Just execute.	12173
2	Crypto doesn't need a spokesperson. is king. ...	281
3	_voyage No celebration needed. Seeing you guys...	328
4	Thank you for your support all along the way....	158
...
1753	I like all of the above. :)	551
1754	_Dep _swap Most CEX are missing out on...	775
1755	is not a new asset class; it is a new technology.	16263
1756	many thanks!	202

Πίνακας 12 Tweets by @cz_binance

Στη συνέχεια οπτικοποιήσαμε τις λέξεις με τη μεγαλύτερη συχνότητα.

Όπως φαίνεται και στα διαγράμματα οι λέξη “crypto” και “binance” εμφανίζονται περισσότερες φορές. Ως ιδρυτής της Binance είναι λογικό αποτέλεσμα η συχνότερη εμφάνιση των λέξεων αυτών στα κείμενά του.

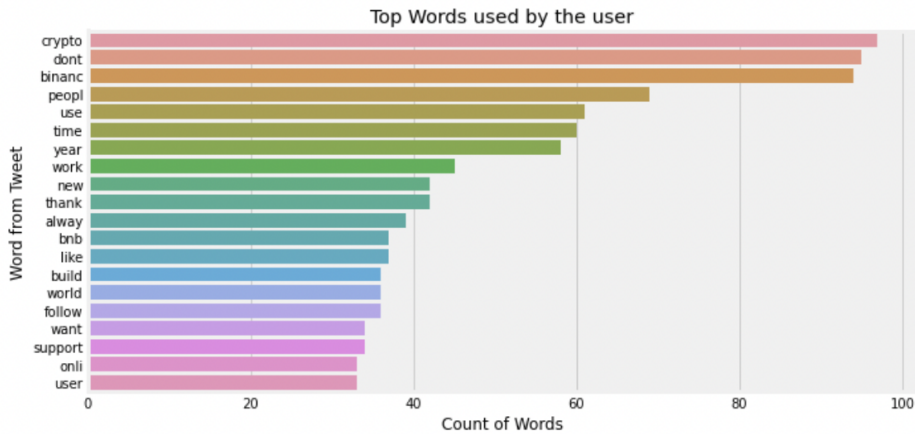


Chart 15 Top Words used by @cz_binance



WordCloud 11 @cz_binance

Στο τελευταίο βήμα, υπολογίστηκε ο βαθμός Subjectivity και Polarity ώστε να γίνει κατανοητή η συναισθηματική ανάλυση. Στην περίπτωση αυτή φαίνεται η στάση του χρήστη να τείνει στο “Neutral”. Αυτό συνεπάγεται από το είδος των Tweet που δεν έχουν έντονες λέξεις ή φράσεις που να δείχνουν μια θετική ή αρνητική κατεύθυνση. Οπότε ο κώδικας των κατατάσσει σε ουδέτερη συναισθηματική στάση.

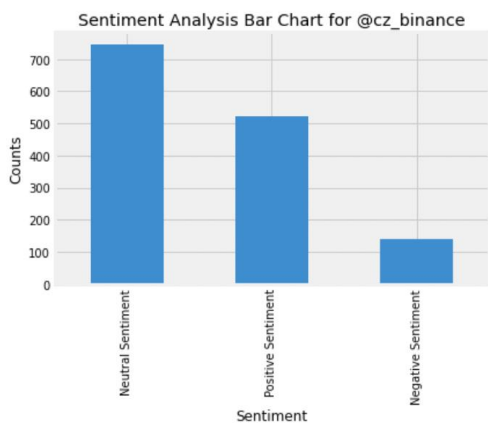
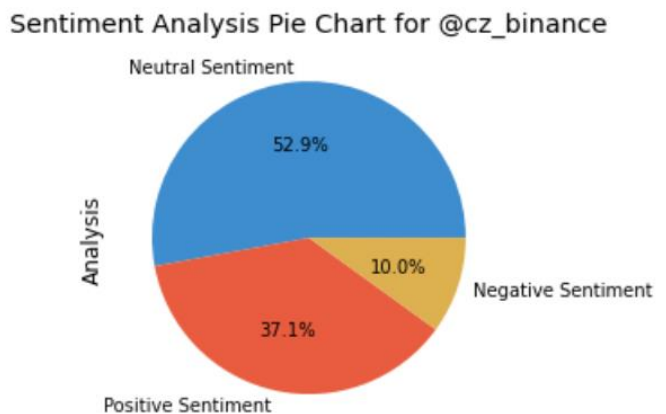


Chart 16 Sentiment Analysis @cz_binance



Pie Chart 10 @cz_binance

Κεφάλαιο 5^ο: Σύνοψη και Συμπεράσματα Έρευνας

Η Συναισθηματική Ανάλυση αποτελεί από τα σημαντικότερα εργαλεία μάρκετινγκ για πολλές εταιρίες, οργανισμούς και ερευνητικές ομάδες, καθώς εξάγει αποτελέσματα για τη γνώμη, την στάση και τα συναισθήματα των καταναλωτών και χρηστών. Η έρευνα που έλαβε μέρος αυτή Διπλωματική εργασία αφορά την Ανάλυση του Συναισθήματος των χρηστών του Twitter που αφορούν την αγορά των Κρυπτονομισμάτων.

Ξεκινώντας, θα παρουσιαστούν συνοπτικά και συγκριτικά κάποια αξιοσημείωτα στοιχεία της έρευνας.

Όσο αφορά την πρώτη κατηγορία μελέτης περίπτωσης, η έρευνα επικεντρώθηκε στην ανάλυση συναισθήματος σε κάποιες «αναφορές»- “hashtag”. Αυτές οι αναφορές είναι οι λέξεις που χρησιμοποιούν οι χρήστες συχνότερα όταν δημοσιεύουν τη γνώμη τους για κάποια γεγονότα που συμβαίνουν στην αγορά των Κρυπτονομισμάτων.

Οι λέξεις-κλειδιά είναι:

- i. Crypto,
- ii. Cryptocurrency,
- iii. Trading,
- iv. Blockchain,
- v. Eth και
- vi. Bitcoin.

Στη συνέχεια, συλλέχθηκαν τα στατιστικά αποτελέσματα και συγκεντρώθηκαν σε ένα διάγραμμα.

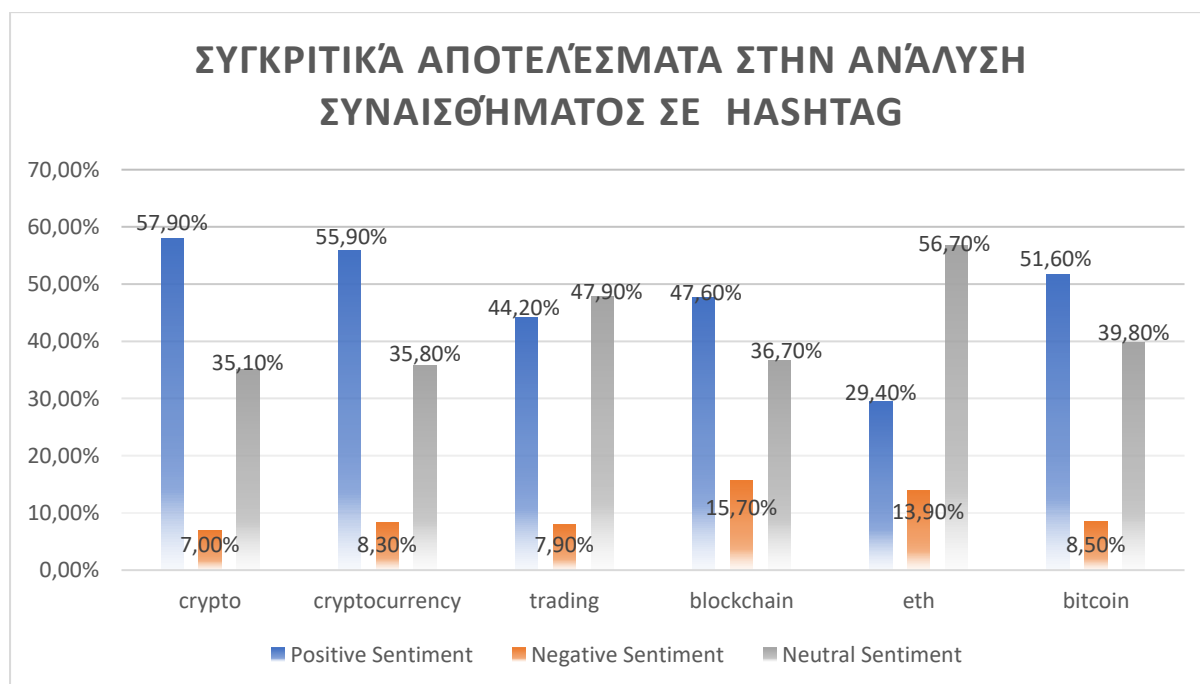


Chart 17 Σύγκριση αποτελεσμάτων στην Ανάλυση Συναισθήματος με συγκεκριμένο Hashtag

Σε κάθε περίπτωση φαίνεται ότι η στάση των χρηστών είναι κατά μεγάλο ποσοστό είτε θετική είτε ουδέτερη. Αυτό μπορεί να συμβεί στις παρακάτω περιπτώσεις:

- A. Όταν ο αλγόριθμος «διαβάζει» τα σχόλια στο Twitter, επεξεργάζεται κάθε λέξη αυτόνομα και όχι στο σύνολο του κειμένου. Δηλαδή, πολλές λέξεις σε μια πρόταση μπορεί να έχουν θετική υπόσταση όμως το κείμενο να έχει αρνητικό χαρακτήρα. Αυτό το χαρακτηριστικό είναι ένα πρόβλημα που παρατηρείται συχνά, για αυτόν το λόγο είναι χρήσιμο παράλληλα ο ερευνητής να βλέπει και τα σχόλια.
- B. Αν ο χρήστης χρησιμοποιεί πολλά σημεία στίξης μπορεί να επηρεάσει το αποτέλεσμα την ανάλυσης. Η χρήση πολλών σημείων στίξης μπορεί να οδηγήσει ένα σχόλιο να χαρακτηριστεί ως «θετικό» ή «αρνητικό» ανάλογα με των υπολοίπων λέξεων.
- C. Η χρήση κάποιου χαρούμενου εικονιδίου (emoticon), μπορεί να χαρακτηρίσει ένα σχόλιο θετικό και αντίστοιχα ένα θυμωμένο εικονίδιο να δημιουργήσει αρνητικό sentiment.
- D. Σημαντικό στοιχείο που μπορεί να επηρεάσει τα αποτελέσματα μια έρευνας συναισθήματος είναι η χρονική περίοδος που λήφθηκαν τα αποτελέσματα. Λόγω ότι στο Twitter API η πρόσβαση στα δεδομένα ισχύει μόνο για σχόλια που δημιουργήθηκαν έως 30 μέρες πριν την αναζήτηση. Στην έρευνα αυτή τα σχόλια λήφθηκαν αρχές του Δεκέμβρη 2021, μια σχετικά καλή περίοδος για τους «παίκτες» της αγοράς, λίγο όμως πριν αρχίσει η μεγάλη πτώση (dip) της.

Από την άλλη πλευρά, στην δεύτερη ερευνητική κατηγορία, τα στοιχεία που εξορίστηκαν ήταν από πέντε σημαντικά άτομα που κρίνουν την αγορά των Κρυπτονομισμάτων. Οι V.Buterin, R.Ver, A.Back, A.Antonopoulos και C.Zhao αποτελούν μέσο επιρροής στο Twitter. Σημαντική παράμετρος είναι οι λέξεις που χρησιμοποιούν περισσότερο στα κείμενά τους. Παρακάτω θα παρουσιαστούν οι τρεις επικρατέστερες και συχνότερα χρησιμοποιούμενες λέξεις από τον καθένα.

Όνομα	Λέξεις που χρησιμοποιούνται περισσότερο στα Tweet		
<i>V.Buterin</i>	Think	People	Ethereum
<i>R.Ver</i>	Bitcoin	Cash	People
<i>A.Back</i>	Bitcoin	People	Market
<i>A.Antonopoulos</i>	Bitcoin	Use	Learn
<i>C.Zhao</i>	Crypto	Binance	People

Πίνακας 13 Λέξεις που χρησιμοποιούν συχνότερα στα Tweet τους

Συμπερασματικά, οι λέξεις που παρατηρήθηκαν να χρησιμοποιούνται περισσότερο στα Tweet των χρηστών αυτών έχουν να κάνουν άμεσα με την επιρροή στους υπόλοιπους χρήστες που τους ακολουθούν. Συγκεκριμένα, οι λέξεις “people” και “learn” βρίσκονταν μέσα στα αποτελέσματα τις έρευνας. Στον παραπάνω πίνακα, παρατηρείται η λέξη “bitcoin” να κατέχει την πρώτη θέση σε χρήση καθώς αποτελεί το κρυπτονόμισμα με το μεγαλύτερο κοινό.

Μελλοντική Ανάγκη για Υλοποίηση Ερευνών πάνω στη Συναισθηματική Ανάλυση

Η ανάγκη των επιχειρήσεων να γνωρίζουν τη στάση και την άποψη των ανθρώπων για το εκάστοτε προϊόν, υπηρεσία ή και γεγονός φέρει την επιστήμη της Εξόρυξης Γνώμης στο προσκήνιο. Όλο και συχνότερα κρίνεται αναγκαίο να αναλυθεί το συναίσθημα των χρηστών σε κάποια πλατφόρμα των μέσων κοινωνικής δικτύωσης καθώς σε αυτόν τον χώρο οι χρήστες εκφράζονται ελεύθερα. Αν παραδείγματος χάριν η εταιρία «Χ» θέλει να εισάγει στην Ελλάδα ένα προϊόν από την Αμερική, μπορεί να κάνει Συναισθηματική Ανάλυση σε σχόλια ή δημοσιεύσεις στο Instagram ή το Twitter για να δουν ποιο είναι το κοινό ενδιαφέροντος και ποια η στάση των χρηστών που το χρησιμοποίησαν (θετική, αρνητική, ουδέτερη).

Επιπλέον, λόγω ότι η Συναισθηματική Ανάλυση υλοποιείται με τη χρήση Γλώσσας Προγραμματισμού και NLP εργαλείων, είναι σημαντικό να υπάρχουν βελτιώσεις στις εντολές ώστε να μειώνονται οι αποκλίσεις. Πιο συγκεκριμένα, πολλές φορές τα κείμενα που συναντάμε στο διαδίκτυο δεν ακολουθούν συγκεκριμένα σύνταξη και γραμματική, έτσι ο κώδικας διαβάζει τις λέξεις, και τα σύμβολα ξεχωριστά και βγάζει ένα συμπέρασμα. Ως παράδειγμα μπορεί να δοθεί το Tweet :

«Πώς την έλυσες αυτή την άσκηση; είσαι τρελός!».

Σε αυτή την περίπτωση οι εντολές για ανάλυση συναισθήματος πολύ πιθανόν να αποδώσουν “Negative Sentiment” καθώς η λέξη «τρελός» μπορεί να θεωρηθεί αρνητική.

Τέλος, είναι σημαντικό να σημειωθεί ότι η επιστήμη του Μάρκετινγκ επικεντρώνεται στην ψυχολογία του καταναλωτή και βασίζεται στη δημιουργία αξίας τόσο για τον πελάτη όσο και για την κοινωνία στο σύνολο. Για αυτόν τον λόγο είναι ανάγκη να αναπτυχθεί και να εδραιωθεί η Ανάλυση του Συναισθήματος στην σημερινή εποχή,

Bibliography

- Hellmann D, 2011. *The Python Standard Library by Example*. New York, USA: Addison-Wesley.
- Γεωργούλη Κ, 2015. *Τεχνητή Νοημοσύνη: Μια Εισαγωγική Προσέγγιση*. Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών (ΣΕΑΒ).
- Winston P, 1993. *Artificial Intelligence*. 3rd Edition ed. USA: Addison-Wesley Publishers.
- Wikipedia, n.d. [Online]
Available at: <https://el.wikipedia.org/wiki/HTML>
- University System of Georgia, n.d. [Online]
Available at: https://www.usg.edu/galileo/skills/unit07/internet07_02.phtml
- Sharma M, 2022. [Online]
Available at: <https://www.geeksforgeeks.org/web-1-0-web-2-0-and-web-3-0-with-their-difference/>
- Kasireddy P, 2021. [Online]
Available at: <https://www.preethikasireddy.com/post/what-do-web-2-0-and-web-3-0-mean-which-one-is-better>
- Pragati V, 2021. [Online]
Available at: <https://dev.to/pragativerma18/evolution-of-web-42eh>
- Nader "Diamondhands" Al-Naji, 2021. [Online]
Available at: <https://www.coindesk.com/tech/2021/11/17/web-3-social-media-needs-dedicated-blockchains/>
- Mersch M, & Muirhead R., 2019. [Online]
Available at: <https://medium.com/fabric-ventures/what-is-web-3-0-why-it-matters-934eb07f3d2b>
- Rajaraman A., Ullman J., 2013. *Εξόρυξη από Μεγάλα Σύνολα Δεδομένων*. 1η Έκδοση ed. Αθήνα: Εκδόσεις Νέων Τεχνολογιών.
- Rajaraman A, & Ullman J, 2020. *Εξόρυξη από Μεγάλα Σύνολα Δεδομένων*. 3η Έκδοση ed. Αθήνα: Εκδόσεις Νέων Τεχνολογιών.
- Kao A, & Poteet S., 2007. *Natural Language Processing and Text Mining*. Washington, USA: Springer Science & Business Media.
- Berry M, & Kogan J., 2010. *Text Mining: Applications and Theory*. Sussex, UK: John Wiley & Sons Ltd.
- Βλαχοπούλου Μ, & Δημητριάδης Σ., 2014. *e-λεκτρονικό επιχειρείν & Μάρκετινγκ*. Αθήνα: Rosili.
- Liu Bing, 2010. *Sentiment Analysis and Opinion Mining*. Toronto, Canada: Morgan & Claypool Publishers.
- Michael Dewing, 2010. *Social Media an Introduction*. Ontario, Canada: Library of Parliament.
- Καλαφούτης Σ, & Σταμούλης Γ., 2018. *Προγραμματισμός με την Python*. Αθήνα: Εκδόσεις Νέων Τεχνολογιών.
- Maisha R, 2021. [Online]
Available at: https://www.hostinger.com/tutorials/what-is-a-blog#What_Is_a_Blog
- Wikipedia, n.d. [Online]
Available at:
https://el.wikipedia.org/wiki/%CE%9C%CE%AD%CF%83%CE%B1_%CE%BA%CE%BF%CE%B9%CE%BD%CF%89%CE%BD%CE%B9%CE%BA%CE%AE%CF%82_%CE%B4%CE%B9%CE%BA%CF%84%CF%8D%CF%89%CF%83%CE%B7%CF%82

Animikh Aich, 2019. [Online]
 Available at: <https://www.zeolearn.com/magazine/what-is-data-science>

Luellen E, 2018. [Online]
 Available at: <https://towardsdatascience.com/why-data-science-succeeds-or-fails-c24edd2d2f9>

Jayendran, 2021. [Online]
 Available at: <https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/>

JP Hwang, 2020. [Online]
 Available at: <https://towardsdatascience.com/what-python-package-is-best-for-getting-data-from-twitter-comparing-tweepy-and-twint-f481005eccc9>

Anon., n.d. [Online]
 Available at: <https://mode.com/python-tutorial/libraries/pandas/>

Shubham J, 2018. [Online]
 Available at: <https://www.analyticsvidhya.com/blog/2018/02/natural-language-processing-for-beginners-using-textblob/>

Python Libraries, n.d. [Online]
 Available at: <https://mode.com/python-tutorial/libraries/pandas/>

Python Libraries, n.d. [Online]
 Available at: <https://mode.com/python-tutorial/libraries/numpy/>

Python Libraries, n.d. [Online]
 Available at: <https://www.geeksforgeeks.org/generating-word-cloud-python/>

Python Libraries, n.d. [Online]
 Available at: https://www.tutorialspoint.com/python/python_reg_expressions.htm

Python Libraries, n.d. [Online]
 Available at: <https://mode.com/python-tutorial/libraries/matplotlib/>

NLTK Python Libraries, n.d. [Online]
 Available at: https://en.wikipedia.org/wiki/Natural_Language_Toolkit

Anderson Martin, 2019. [Online]
 Available at: <https://www.iflexion.com/blog/sentiment-analysis-python>

Özgür Genç, 2019. [Online]
 Available at: <https://towardsdatascience.com/real-time-sentiment-analysis-on-social-media-with-opensource-tools-f864ca239afe>

Wikipedia, n.d. [Online]
 Available at: <https://en.wikipedia.org/wiki/Microblogging>

Wikipedia, n.d. [Online]
 Available at: <https://en.wikipedia.org/wiki/Twitter#2019%E2%80%93present>

Wikipedia, n.d. [Online]
 Available at: <https://en.wikipedia.org/wiki/API>

Twitter API, n.d. [Online]
 Available at: <https://developer.twitter.com/en/docs/twitter-api>

Farhadloo M, 2016. [Online]
 Available at:
https://www.researchgate.net/publication/300965436_Fundamentals_of_Sentiment_Analysis_and_Its_Applications

Akash S, 2021. [Online]
 Available at: <https://www.analyticsvidhya.com/blog/2021/10/making-natural-language-processing-easy-with-textblob/>

Federico Pascual, 2019. [Online]
 Available at: <https://monkeylearn.com/blog/sentiment-analysis-of-twitter/>

Aayushi Johari, 2021. [Online]

Available at: <https://www.edureka.co/blog/python-programming-language>

Frankenliefeld J, 2021. [Online]

Available at: <https://www.investopedia.com/terms/b/bitcoin.asp>

Frankenliefeld J, 2022. [Online]

Available at: <https://www.investopedia.com/terms/e/ethereum.asp>

Wikipedia, n.d. [Online]

Available at: https://en.wikipedia.org/wiki/Distributed_ledger

Wikipedia, n.d. [Online]

Available at:

<https://el.wikipedia.org/wiki/%CE%A0%CF%81%CF%89%CF%84%CF%8C%CE%BA%CE%BF%CE%BB%CE%BB%CE%BF%CE%9C%CE%B5%CF%84%CE%B1%CF%86%CE%BF%CF%81%CE%AC%CF%82%CE%A5%CF%80%CE%B5%CF%81%CE%BA%CE%B5%CE%B9%CE%BC%CE%AD%CE%BD%CE%BF%CF%85>

Wikipedia, n.d. [Online]

Available at: https://el.wikipedia.org/wiki/Common_Gateway_Interface

Πηγές Φωτογραφιών

<https://monkeylearn.com/sentiment-analysis/>

<https://zephyrnet.com/el/%CE%B3%CE%B9%CE%B1%CF%84%CE%AF-%CE%B7-%CE%B1%CF%80%CE%BF%CE%BA%CE%AD%CE%BD%CF%84%CF%81%CF%89%CF%83%CE%B7-%CE%B4%CE%B5%CE%BD-%CE%B5%CE%AF%CE%BD%CE%B1%CE%B9-%CE%BF-%CE%B1%CF%80%CF%8E%CF%84%CE%B5%CF%81%CE%BF%CF%82-%CF%83%CF%84%CF%8C%CF%87%CE%BF%CF%82-%CF%84%CE%BF%CF%85-web3/>

<https://guides.co/g/python-programming-language/151355>

<https://www.zeolearn.com/magazine/what-is-data-science>

<https://towardsdatascience.com/why-data-science-succeeds-or-fails-c24edd2d2f9>

https://www.researchgate.net/figure/Relationship-between-AI-ML-DL-and-NLP-7_fig8_343079524

<https://ps5098252.medium.com/how-we-can-store-huge-data-which-is-beyond-our-capacity-ab51243b3dda>

<https://developer.twitter.com/en>

https://colab.research.google.com/?utm_source=scs-index

Παραρτήματα

Εντολές μέσω R στο Google Colab για κάθε μελέτη περίπτωση

i. Cryptocurrency

```
## Import the libraries
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')

## Get the Twitter API credentials
consumerKey = "aY535KcQFytgghjL5H7AeASw"
consumerSecret = "MmBTqxZuiHQb7uSIHNfoZ2TXdE74wcj0zsqak2t1212ul38GIU"
accessToken = "324268217-pnscsQocEmlPEi50kwnlqTHFrrQlgPVzJz39uvBe"
accessTokenSecret = "FXPHG9QkIX8vgO41FbSkzJW67dvsnoQgAbaJER8tFOImD"

## Create the authentication object
authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
## Set the access token and access token secret
authenticate.set_access_token(accessToken, accessTokenSecret)
## Create the API object
api = tweepy.API(authenticate, wait_on_rate_limit=True)

## Gather 2000 Tweets about Cryptocurrency Market
search_term = "#cryptocurrency -filter:retweets"
## Create a cursor object
tweets = tweepy.Cursor(api.search, q=search_term, lang='en', tweet_mode='extended').items(2000)
## Store the tweets in a variable and get the full text
all_tweets = [tweet.full_text for tweet in tweets]
```

```
## Create a DataFrame to store the tweets with columns called "Tweets"
df = pd.DataFrame(all_tweets, columns=['Tweets'])
## Show the first 5 rows of data
df.head(5)
```

Tweets

```
0    Reef-World were not contacted by, nor did the ...
1    #Saitama #SaitamaWolfPack #SaitamaInu #SaitaMa...
2    @TheMoonCarl @weyuofficial\n@LaunchMyNFT \nThi...
3    In 2 days my NFT collection is available on op...
4    @hasten_clark\n What will be the future of Rus...
```

```
## Create a function to clean the data
def cleanTxt(text):
    text = re.sub('[A-Za-z0-9]+', '', text)
    text = re.sub('#[A-Za-z0-9]+', '', text)
    text = re.sub('https://\\S+', '', text)
    text = re.sub('\\n', '', text)
    text = re.sub('[A-Za-z0-9][A-Za-z0-9] +', '', text)
    return text
```

```
## Cleaning the Text
df['Tweets'] = df['Tweets'].apply(cleanTxt)

## Show the Cleaned Text
df
```

```
#Print all the positive tweets
```

```
j=1
sortedDF= df.sort_values(by=['Polarity'])
for i in range(0, sortedDF.shape[0]):
    if(sortedDF['Analysis'][i] == 'Positive Sentiment'):
        print(str(j) + ' ' +sortedDF['Tweets'][i])
        print()
        j= j+1
```

47) Which are you accumulating more of during these dips?

48) ASD Token Pre Sale ONLY 0,008\$! 🚀With minimum purchase of 5,000 tokens, you can get coins with fabulous projects.Get yours now

49) This is just amazing! _OFFICIAL

50) Buy more \$LUFFY and hold 💙 i believe luffy waiting momen to bullish again after NFTs Marketplace done, market settled and mass

51) #babydoge#BabyDogeArmy#cryptocrash#cryptocurrency#Bitcoin#Crypto#Ethereum has very strong fundamentals and will remove 3 zeros s

52) Over 49% of the supply is staked. And it launches on 2 new chains today! Ape in before it blows UP!

53) The scope of work of this project is staggering, they will succeed. Good idea. Good luck in the implementation.

54) #cryptoThis is one of the best project in 2022. I hope one day this project will be the top project in cryptocurrency.

55) The markets falling, perfect buying opportunity. And guess what hasn't fell whatsoever? Baseline value, that's right. very own

56) #cryptocurrencyWonderful project. This project is top listed. This project is perfect from every angle. I recommend you to see t

57) #babydoge#BabyDogeArmy#cryptocrash#cryptocurrency#Bitcoin#Crypto#Ethereum has very strong fundamentals and will remove 3 zeros s

58) _info This project is looks so innovative and impactful, happy to take participate in such huge project. You guys are very hard

59) Do you need a pavia ESTATE? Mint some panda fighters and have a chance to become _io newest 📈degen baller📈 ONLY 400 chances lef

60) I am go for \$TAF. This is very promising project with usecases and utilities. Great team behind this. Buy it before too late.

61) _info This project is looks so innovative and impactful, happy to take participate in such huge project. You guys are very hard

62) _sajet Buy more \$LUFFY and hold 💙 i believe luffy waiting momen to bullish again after NFTs Marketplace done, market settled an

63) That read me, first be careful, and in the second follow me on Facebook and Twitter. #WitLinkisGlobalMarketplace

```
#Print all the negative tweets
```

```
j=1
sortedDF= df.sort_values(by=['Polarity'], ascending='False')
for i in range(0, sortedDF.shape[0]):
    if (sortedDF['Analysis'][i]=='Negative Sentiment'):
        print(str(j)+' ' + sortedDF['Tweets'][i])
        print()
        j=j+1
```

1) Top 10 Live Prices: (08:15 AM . Jan 23,2022)\$35778.50\$2492.86\$383.80\$0.14\$0.0000223414\$52.89\$2.10\$11.56\$0.0000458214\$0.000000019

2) cryptocurrency news.Crypto Miners: Cheap, or Dangerous to Buy Now? - Nasdaq #money

3) [Futures] By 2022-01-22 16:00 GMT, \$ALGO futures Long Rate is 54.69% and Short Rate is 45.31%.

4) Without all of this VOLATILITY there would be no OPPORTUNITY. So... Don't complain, be thankful you get extra opportunities to pu

5) [Futures] By 2022-01-22 17:00 GMT, \$IOTA futures had a liquidiation of Short Positions of \$186.38K and of Long Positions of \$1400

6) Securities and Exchange Commission SEC reprimands ... #binance

7) Top 10 Live Prices: (08:10 AM . Jan 23,2022)\$35754.45\$2491.49\$383.49\$0.14\$0.0000223654\$52.82\$2.10\$11.56\$0.0000460556\$0.000000019

8) 🚀Frodo Tech ICO Stage 19 is running! 📈Stage 19: (90% is completed📈) 📈1 FRDX = 0.000235 USD19th stage intended target:\$19,631

9) broke the Support break the previous swing low and made new ATL (All - Time - Low). Anticipating anything at the moment can be ri

10) #Ether, the world's second-largest by market cap, was down about 13% over the past 24 hours, compared with a 14% drop in AVAX an

11) [Futures] By 2022-01-22 16:00 GMT, \$ZRX futures Long Rate is 59.95% and Short Rate is 40.05%.

12) What is your plan for Pakistan ?I voted you for a But so much disappointed.You should regulate In Pakistan for better future.

13) 🚀Be afraid, be very afraid...

14) [Futures] By 2022-01-22 17:00 GMT, \$QTUM futures had a liquidiation of Short Positions of \$28.13K and of Long Positions of \$652.

15) Crypto Signals Update Ada/Usdt Buy 1.1 Sell 1.3 Long Term : 2\$ 5 \$ USDT

16) Top 10 Live Prices: (08:05 AM . Jan 23,2022)\$35793.83\$2494.82\$383.50\$0.14\$0.0000224038\$52.87\$2.11\$11.57\$0.0000459913\$0.00000001

17) Next sale effort of empty promises? - Bulgaria pla #binance

ii. Trading,

```
#Import the libraries
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')

#Get the Twitter API credentials
consumerKey = "aY535XcqGFytgqhjL5H7AeASw"
consumerSecret = "MmBTgxZuiHQb7uSIMNfoZZTXdE7ZwcjozsqaK2t1212u13RGIU"
accessToken = "324268217-psncsQoCEmlPEi5OkwNlqTHFrRqLgPVZJz39uvBe"
accessTokenSecret = "fXPHG9Qk1X8vg04iPbSkZJW67dvsnoQgABajER8tFOImD"

#Create the authentication object
authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
#Set the access token and access token secret
authenticate.set_access_token(accessToken, accessTokenSecret)
#Create the API object
api= tweepy.API(authenticate,wait_on_rate_limit=True)

#Gather 2000 Tweets about Cryptocurrency Market
search_term= '#trading -filter:retweets'
#Create a cursor object
tweets= tweepy.Cursor(api.search, q=search_term, lang='en', tweet_mode='extended').items(2000)
#Store the tweets in a variable and get the full text
all_tweets= [tweet.full_text for tweet in tweets]

#Create a Dataframe to store the tweets with columns called "Tweets"
df= pd.DataFrame(all_tweets, columns=['Tweets'])
#Show the first 5 rows of data
df.head(5)
```

Tweets

```
0 #BTCUSDTPERP - New Technical Analysis \n\nChar...
1 #AUDUSD - New Technical Analysis \n\nChart on ...
2 #spy 2022 #market leaders\n\nRanking update 📈\...
3 Don't stay in bed, unless you can make money i...
4 t42#Forex #FX #Trading #Trader\nl make 100% in...
```

```
[ ] #Create a function to clean the data
def cleanTxt(text):
    text= re.sub('[A-Za-z0-9]+', '', text)
    text= re.sub('[A-Za-z0-9]+', '', text)
    text= re.sub('https?:\\/\s+', '', text)
    text= re.sub('\n', '', text)
    text= re.sub('[A-Za-z0-9][A-Za-z0-9] +', '', text)
    return text
```

```
[ ] #Cleaning the Text
df['Tweets']= df['Tweets'].apply(cleanTxt)

#Show the Cleaned Text
df
```

Tweets

```
0 - New Technical Analysis Chart on TradingView ...
1 - New Technical Analysis Chart on TradingView
2 2022 leadersRanking update 📈
3 Don't stay in bed, unless you can make money i...
4 t42#Traderl make 100% in Forex Trading. PM for...
...
1995 Is Cryptocurrency Selling Out to Centralizatio...
1996 Trading princess alexis for best offer or 30 d...
1997 Sign up on WhiteBIT through this link so you ...
1998 If you don't have problems in front of you you...
1999 Machine Learning and Its Application in Forex ...
```

2000 rows x 1 columns

iii. Blockchain

```

#Import the libraries
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')

#Get the Twitter API credentials
consumerKey = "aY535XcqGFyTgghjLSH7AeASw"
consumerSecret = "MmBTgxZuiHQB7uSIMNfoZZTXde7ZwcjozsqaK2t1212u13RGIU"
accessToken = "324268217-psnccQoCEmlPEi50kwNlqTHFrq1gPVZJz39uvBe"
accessTokenSecret = "fXPHG9Qk1X8vgO4iFbSkZJW67dvsnOQgABajER8tFOImD"

#Create the authentication object
authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
#Set the access token and access token secret
authenticate.set_access_token(accessToken, accessTokenSecret)
#Create the API object
api= tweepy.API(authenticate,wait_on_rate_limit=True)

#Gather 2000 Tweets about Cryptocurrency Market
search_term= '#blockchain -filter:retweets'
#Create a cursor object
tweets= tweepy.Cursor(api.search, q=search_term, lang='en', tweet_mode='extended').items(2000)
#Store the tweets in a variable and get the full text
all_tweets= [tweet.full_text for tweet in tweets]

#Create a Dataframe to store the tweets with columns called "Tweets"
df= pd.DataFrame(all_tweets, columns=['Tweets'])
#Show the first 5 rows of data
df.head(5)

```

	Tweets
0	@Sentinel_co @BPSAA_Official The Power Friendl...
1	Russia's Central Bank Proposes New Token Frame...
2	🤖📢 GIVEAWAY 📢🤖 5 prizes of 2 EGLD each!!!📢...
3	🐋 #Cardano \$ADA Whale ❤️laced!📢 Transaction ...
4	@NetworkSandwich Very good and strong project ...

```

#Create a function to clean the data
def cleanTxt(text):
    text= re.sub('@[A-Za-z0-9]+', '', text)
    text= re.sub('#[A-Za-z0-9]+', '', text)
    text= re.sub('https?:\/\/\S+', '', text)
    text= re.sub('\n', '', text)
    text= re.sub('[A-Za-z0-9]#[A-Za-z0-9] +', '', text)
    return text

```

```

#Create a function to get Subjectivity
def getSubjectivity(text):
    return TextBlob(text).sentiment.subjectivity

#Create a function to get Polarity
def getPolarity(text):
    return TextBlob(text).sentiment.polarity

#Create two columns called "Subjectivity" and "Polarity"
df['Subjectivity']= df['Tweets'].apply(getSubjectivity)
df['Polarity']= df['Tweets'].apply(getPolarity)

#Show data
df.head()

```

	Tweets	Subjectivity	Polarity
0	_co_Official The Power Friendly Blockchain @ ...	0.500000	0.375000
1	Russia's Central Bank Proposes New Token Frame...	0.352273	0.068182
2	🤖📢 GIVEAWAY 📢🤖 5 prizes of 2 EGLD each!!!📢 #Web3	0.000000	0.000000
3	🐋 \$ADA Whale ❤️laced!📢 Transaction Output: 1.6...	0.000000	0.000000
4	Very good and strong project i like this proj...	0.756667	0.671667

iv. Eth

```

#Import the libraries
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')

```

```

#Get the Twitter API credentials
consumerKey = "aY535XcqFPyTgghjL5H7AeASw"
consumerSecret = "MmBTgxZuiHQB7uSIMNfoZZTXdE7ZwcjozsqaK2t1212u13RGIU"
accessToken = "324268217-psncsQoCEmlPEi50kwNlqTHFrRq1gPVZJz39uvBe"
accessTokenSecret = "fXPHG9Qk1X8vg04iFbSkZJW67dvsn0OqABajER8tFOImD"

```

```

#Create the authentication object
authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
#Set the access token and access token secret
authenticate.set_access_token(accessToken, accessTokenSecret)
#Create the API object
api = tweepy.API(authenticate, wait_on_rate_limit=True)

```

```

#Gather 2000 Tweets about Cryptocurrency Market
search_term = '#eth -filter:retweets'
#Create a cursor object
tweets = tweepy.Cursor(api.search, q=search_term, lang='en', tweet_mode='extended').items(2000)
#Store the tweets in a variable and get the full text
all_tweets = [tweet.full_text for tweet in tweets]

```

```

#Create a Dataframe to store the tweets with columns called "Tweets"
df = pd.DataFrame(all_tweets, columns=['Tweets'])
#Show the first 5 rows of data
df.head(5)

```

Tweets

```

0 @altcryptocom @binance In the crypto market co...
1 Blue Star by anJoani\n\n1 of 1 · Now Available...
2 @altcryptocom @binance In the crypto market co...
3 @AirdropStario It's a wonderful project 🥰 I'm ...
4 @altcryptocom @binance In the crypto market co...

```

```

#Create a function to clean the data
def cleanTxt(text):
    text = re.sub('[A-Za-z0-9]+', '', text)
    text = re.sub('#[A-Za-z0-9]+', '', text)
    text = re.sub('https?:\/\/\S+', '', text)
    text = re.sub('\n', '', text)
    text = re.sub('[A-Za-z0-9]+#[A-Za-z0-9]+', '', text)

```

```

#Cleaning the Text
df['Tweets'] = df['Tweets'].apply(cleanTxt)

```

```

#Show the Cleaned Text
df

```

Tweets

```

0 In the crypto market competition, no compaig...
1 Blue Star by anJoani1 of 1 · Now Available for...
2 In the crypto market competition, no compaig...
3 It's a wonderful project 🥰 I'm so interested ...
4 In the crypto market competition, no compaig...
...
1995 1. 100% liquidity locked 🔥 2. Ownership Renoun...
1996 1. 100% liquidity locked 🔥 2. Ownership Renoun...
1997 1. 100% liquidity locked 🔥 2. Ownership Renoun...
1998 🐰🐰🐰🐰_/_/🐰🐰.*) #littlerabbit🐰🐰/ ...
1999 "Flamingos of the Galapagos" from the collect...

```

2000 rows x 1 columns

```

#Create a function to get the text sentiment score

def getAnalysis (score):
    if score<0:
        return 'Negative Sentiment'
    elif score ==0:
        return 'Neutral Sentiment'
    else:
        return 'Positive Sentiment'

#Create a column for the Sentiment

df['Analysis'] = df['Polarity'].apply(getAnalysis)

#Show data
df

```

	Tweets	Subjectivity	Polarity	Analysis
0	In the crypto market competition, no compaig...	0.30	-0.200000	Negative Sentiment
1	Blue Star by anJoani1 of 1 • Now Available for...	0.25	0.200000	Positive Sentiment
2	In the crypto market competition, no compaig...	0.30	-0.200000	Negative Sentiment
3	It's a wonderful project 😊 I'm so interested ...	0.80	0.616667	Positive Sentiment
4	In the crypto market competition, no compaig...	0.30	-0.200000	Negative Sentiment
...
1995	1. 100% liquidity locked 🔥 2. Ownership Renoun...	0.00	0.000000	Neutral Sentiment
1996	1. 100% liquidity locked 🔥 2. Ownership Renoun...	0.00	0.000000	Neutral Sentiment
1997	1. 100% liquidity locked 🔥 2. Ownership Renoun...	0.00	0.000000	Neutral Sentiment
1998	🔗🔗🔗🔗_🔗🔗🔗🔗)🔗🔗🔗🔗.#fittlerabbit🔗🔗🔗🔗/ ...	0.00	0.000000	Neutral Sentiment
1999	"Flamingos of the Galapagos" from the collect...	0.40	0.400000	Positive Sentiment

2000 rows x 4 columns

v. Bitcoin.

```
#Import the libraries
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')

#Get the Twitter API credentials
consumerKey = "aY535XcqGfYtgqjL5H7AeASw"
consumerSecret = "MmBTgxZuiRQB7uSIMNfoZTXde7ZwcjozsqaK2t1212u13RGIU"
accessToken = "324268217-psnacsQoCEmlPEi50kwNlqTHFrqLgPVZJz39uvBe"
accessTokenSecret = "fXFBG9QkIX8vgO4iFbSkZJW67dvsnOQgABajER8tFOImD"

#Create the authentication object
authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
#Set the access token and access token secret
authenticate.set_access_token(accessToken, accessTokenSecret)
#Create the API object
api = tweepy.API(authenticate, wait_on_rate_limit=True)

#Gather 2000 Tweets about Cryptocurrency Market
search_term = '#bitcoin -filter:retweets'
#Create a cursor object
tweets = tweepy.Cursor(api.search, q=search_term, lang='en', tweet_mode='extended').items(2000)
#Store the tweets in a variable and get the full text
all_tweets = [tweet.full_text for tweet in tweets]

#Create a Dataframe to store the tweets with columns called "Tweets"
df = pd.DataFrame(all_tweets, columns=['Tweets'])
#Show the first 5 rows of data
df.head(5)
```

	Tweets
0	I only believe in one project for the long ter...
1	Less than 0.3% of the world will ever be able ...
2	Btc for BYBIT:BTCUSD by Erictaylor #bitcoin #...
3	Free #Airdrop Are You Ready ? 🤔🤔🤔\nClaim Your ...
4	📈 +5% profit of \$IOTX-USD in less than 1 hours...

```
#Create a function to clean the data
def cleanTxt(text):
    text = re.sub('[A-Za-z0-9]+', '', text)
    text = re.sub('#[A-Za-z0-9]+', '', text)
    text = re.sub('https?:\\/\S+', '', text)
    text = re.sub('\n', '', text)
    text = re.sub('[A-Za-z0-9]#[A-Za-z0-9] +', '', text)
    return text
```

```
#Cleaning the Text
df['Tweets'] = df['Tweets'].apply(cleanTxt)

#Show the Cleaned Text
df
```

	Tweets
0	I only believe in one project for the long ter...
1	Less than 0.3% of the world will ever be able ...
2	Btc for BYBIT:BTCUSD by Erictaylor
3	Free Are You Ready ? 🤔🤔🤔Claim Your Tokens Befo...
4	📈 +5% profit of \$IOTX-USD in less than 1 hours...
...	...
1995	👏 A new block was found on the network. We're ...
1996	Good luck to all the joiners of this giveaway...
1997	Thanks to Robinhood for Supporting the Shib Fa...
1998	Don't miss this wonderful opportunity to get ...
1999	From Reddit! Just a request. Please make sure ...

2000 rows x 1 columns

```

#Create a function to get the text sentiment score

def getAnalysis (score):
    if score<0:
        return 'Negative Sentiment'
    elif score ==0:
        return 'Neutral Sentiment'
    else:
        return 'Positive Sentiment'

#Create a column for the Sentiment

df['Analysis'] = df['Polarity'].apply(getAnalysis)

#Show data
df

```

	Tweets	Subjectivity	Polarity	Analysis
0	I only believe in one project for the long ter...	0.520000	-0.062727	Negative Sentiment
1	Less than 0.3% of the world will ever be able ...	0.560417	0.342708	Positive Sentiment
2	Btc for BYBIT:BTCUSD by EricTaylor	0.000000	0.000000	Neutral Sentiment
3	Free Are You Ready ? 🤔🤔🤔Claim Your Tokens Befo...	0.633333	0.075000	Positive Sentiment
4	📈 +5% profit of \$IOTX-USD in less than 1 hours...	0.158333	0.041667	Positive Sentiment
...
1995	🏆 A new block was found on the network. We're ...	0.284848	0.045455	Positive Sentiment
1996	Good luck to all the joiners of this giveaway...	0.600000	0.700000	Positive Sentiment
1997	Thanks to Robinhood for Supporting the Shib Fa...	0.225000	0.225000	Positive Sentiment
1998	Don't miss this wonderful opportunity to get ...	0.750000	0.625000	Positive Sentiment
1999	From Reddit! Just a request. Please make sure ...	0.888889	0.500000	Positive Sentiment

2000 rows x 4 columns

i. V.Buterin,

```
#Import the libraries
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
# pip install -U spacy
# python -m spacy download en_core_web_sm
import spacy
nlp = spacy.load("en_core_web_sm")
import seaborn as sns
from nltk.stem.snowball import SnowballStemmer

#Get the Twitter API credentials
consumerKey = "aY5J5KcQGFyTgghjL5H7AeAsw"
consumerSecret = "MmB7gxZuiBQB7uSfMNFoZ2TXdE7ZwcjozsqK2t1212u13RGIU"
accessToken = "324268217-pancsQoCEmlPE150kwNlqTHFrr0lqPVIz39uvBe"
accessTokenSecret = "FXFH9YQk1X8vg04iFbSkZJw67dvanOQgAbaJER8tFOImb"

#Create the authentication object
authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
#Set the access token and access token secret
authenticate.set_access_token(accessToken, accessTokenSecret)
#Create the API object
api = tweepy.API(authenticate,wait_on_rate_limit=True)

number_of_tweets=2000
tweets= []
likes= []

for i in tweepy.Cursor(api.user_timeline, id="VitalikButerin",tweet_mode="extended").items(number_of_tweets):
    tweets.append(i.full_text)
    likes.append(i.favorite_count)

df= pd.DataFrame({'Tweets': tweets, 'Likes': likes})
df
```

	Tweets	Likes
0	I'm speaking in 3 min https://t.co/DynHSWNPPE	5780
1	@zengjajun_eth 其实人类文明的很多部分是去中心化的。一个很明显的例子：语言...	573
2	@DeepCipher @sh @elonmusk Disparities in econ...	4005
3	@owocki @ivanmolto retro funding for public go...	87
4	@MacaesBruno Latin America seems to me the mos...	47
...
1995	@ArthurB @QiaochuYuan (sorry about the confus...	3
1996	@ArthurB @QiaochuYuan Likely the most realist...	9
1997	@ArthurB @QiaochuYuan You live on twitter (and...	3
1998	@WgRSBxbkdbNN1cU @zooko 他应该在英国吧，如果你在深圳你的早上是他的晚上	0
1999	@MhaloBjelic I'm assuming the worst case: la...	2

2000 rows x 2 columns

```
list_of_sentences= [sentence for sentence in df.Tweets]

lines=[]
for sentence in list_of_sentences:
    words = sentence.split()
    for w in words:
        lines.append(w)

lines= [re.sub(r'^A-Za-z0-9+', '', x) for x in lines]

lines2=[]

for word in lines:
    if word != '':
        lines2.append(word)

s_stemmer= SnowballStemmer (language='english')

stem=[]
for word in lines2:
    stem.append(s_stemmer.stem(word))

stem2=[]
for word in stem:
    if word not in nlp.Defaults.stop_words:
        stem2.append(word)

df2= pd.DataFrame(stem2)

df2=df2[0].value_counts()

df2= df2[:20,]
plt.figure(figsize=(10,5))
sns.barplot(df2.values, df2.index, alpha=1)
plt.title('Top Words used by the user')
plt.ylabel('Word from Tweet',fontsize=12)
plt.xlabel('Count of Words',fontsize=12)
plt.show()
```

ii. R.Ver,

```

#Import the libraries
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
# pip install -U spacy
# python -m spacy download en_core_web_sm
import spacy
nlp = spacy.load("en_core_web_sm")
import seaborn as sns
from nltk.stem.snowball import SnowballStemmer

#Get the Twitter API credentials
consumerKey = "aY535XcqGFyTgghjL5H7AeASw"
consumerSecret = "MmBTgxZuiHQB7uS1MNfoZZTXdE7ZwcjozsqaK2t1212u13RGIU"
accessToken = "324268217-psnscQoCEmlPEi5OkwNlqTHFrrQlgPVZJz39uvBe"
accessTokenSecret = "fXPHG9Qk1X8vgO4iFbSkZJW67dvsn00gABajER8tFOImD"

#Create the authentication object
authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
#Set the access token and access token secret
authenticate.set_access_token(accessToken, accessTokenSecret)
#Create the API object
api= tweepy.API(authenticate,wait_on_rate_limit=True)

number_of_tweets=2000
tweets= []
likes= []

for i in tweepy.Cursor(api.user_timeline, id="rogerkver",tweet_mode="extended").items(number_of_tweets):
    tweets.append(i.full_text)
    likes.append(i.favorite_count)

df= pd.DataFrame({'Tweets': tweets, 'Likes': likes})
df

```

	Tweets	Likes
0	Today marks the 10 year anniversary of my firs...	4368
1	Securities laws are outdated. \nNeeding a lice...	1704
2	It took about a decade for crypto to grow to a...	2081
3	There is no inflation they say...\nhttps://t...	2180
4	Soon everyone will realize that the one megaby...	1291
...
1995	Opening a Bank account\nApplication: 3 hour l...	97
1996	Being called a soldier does not separate you f...	84
1997	"Lawmakers" should more accurately be called "...	32
1998	Buy the book on how to use Bitcoin Anonymously...	29
1999	"Laws are simply rules made by politicians, an...	18

2000 rows x 2 columns

```

#Clean the Texts
def cleanText(text):
    text = re.sub('[A-Za-z0-9]+', '', text)
    text = re.sub('[A-Za-z0-9]+', '', text)
    text = re.sub('http://\V/\V*', '', text)
    text = re.sub('\n', '', text)
    return text

df=df[~df.Tweets.str.contains("RT")]

#Cleaning the Text
df['Tweets']= df['Tweets'].apply(cleanText)

#Show the Cleaned Text
df

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy

```

	Tweets	Likes	Subjectivity	Polarity	Analysis
0	Today marks the 10 year anniversary of my fir...	4368	0.200000	0.1250	Positive Sentiment
1	Securities laws are outdated. Needing a licens...	1704	0.566667	-0.3875	Negative Sentiment
2	It took about a decade for crypto to grow to a...	2081	0.000000	0.0000	Neutral Sentiment
3	There is no inflation they say...	2180	0.000000	0.0000	Neutral Sentiment
4	Soon everyone will realize that the one megaby...	1291	0.400000	0.2000	Positive Sentiment
...
1995	Opening a Bank account:Application: 3 hour int...	97	1.000000	0.5000	Positive Sentiment
1996	Being called a soldier does not separate you f...	84	0.000000	0.0000	Neutral Sentiment
1997	"Lawmakers" should more accurately be called "...	32	0.566667	0.4500	Positive Sentiment
1998	Buy the book on how to use Bitcoin Anonymously...	29	0.000000	0.0000	Neutral Sentiment
1999	"Laws are simple rules made by politicians, an...	18	0.303571	0.0000	Neutral Sentiment

```

list_of_sentences= [sentence for sentence in df.Tweets]
lines=[]
for sentence in list_of_sentences:
    words = sentence.split()
    for w in words:
        lines.append(w)

```

iii. A.Back,

```

#Import the libraries
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
# pip install -U spacy
# python -m spacy download en_core_web_sm
import spacy
nlp = spacy.load("en_core_web_sm")
import seaborn as sns
from nltk.stem.snowball import SnowballStemmer

#Get the Twitter API credentials
consumerKey = "aY535XcQGFtqghJL5H7AeASw"
consumerSecret = "Mm87gxZuiHQ87uS1MmfoZTXdE7zwcjozsqaK2t1212u13RGIU"
accessToken = "324268217-pnscsQoCEmLPE150kwNlqTHFrQlgPvZJz39uvBe"
accessTokenSecret = "FXPHG9Qk1X8vgO4iFbSKZJW67dvan00gABajER8tFOImd"

#Create the authentication object
authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
#Set the access token and access token secret
authenticate.set_access_token(accessToken, accessTokenSecret)
#Create the API object
api = tweepy.API(authenticate, wait_on_rate_limit=True)

number_of_tweets=2000
tweets= []
likes= []

for i in tweepy.Cursor(api.user_timeline, id="adam3us", tweet_mode="extended").items(number_of_tweets):
    tweets.append(i.full_text)
    likes.append(i.favorite_count)

df= pd.DataFrame({'Tweets': tweets, 'Likes': likes})
df

```

	Tweets	Likes
0	@TheStalwart and then the economy collapses, a...	4
1	RT @DigiKapitalist: @adam3us Money=energy * time	0
2	{ 2^(224-d) > H(n, #bitcoin) } = 600mc*2vnh...	88
3	RT @100trillionUSD: How long before #bitcoin l...	0
4	@Leishman this is good! the plebs are taking c...	43
...
1995	@CORNivore @BitcoinsSaving @christinenews You...	2
1996	@lylepratt @PrestonPysh @muneeb @fluidvoice @S...	3
1997	@rationalaussie @JohnMichaelKuhn @muneeb @Pres...	1
1998	@udiWertheimer @lylepratt @PrestonPysh @muneeb...	12
1999	@noigrubies @lylepratt @PrestonPysh @muneeb @f...	9

2000 rows x 2 columns

```
] list_of_sentences= [sentence for sentence in df.Tweets]

lines=[]
for sentence in list_of_sentences:
    words = sentence.split()
    for w in words:
        lines.append(w)

] lines= [re.sub(r'[^\A-Za-z0-9]+', '', x) for x in lines]

lines2=[]

for word in lines:
    if word != '':
        lines2.append(word)

] s_stemmer= SnowballStemmer (language='english')

stem=[]
for word in lines2:
    stem.append(s_stemmer.stem(word))

] stem2=[]
for word in stem:
    if word not in nlp.Defaults.stop_words:
        stem2.append(word)

] df2= pd.DataFrame(stem2)

df2=df2[0].value_counts()

] df2= df2[:20,]
plt.figure(figsize=(10,5))
sns.barplot(df2.values, df2.index, alpha=1)
plt.title('Top Words used by the user')
plt.ylabel('Word from Tweet',fontsize=12)
plt.xlabel('Count of Words',fontsize=12)
plt.show()
```

iv. A.Antonopoulos

```
#Import the libraries
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
# pip install -U spacy
# python -m spacy download en_core_web_sm
import spacy
nlp = spacy.load("en_core_web_sm")
import seaborn as sns
from nltk.stem.snowball import SnowballStemmer

#Get the Twitter API credentials
consumerKey = "aY535XcGFyTgghjL5H7AeASw"
consumerSecret = "MmbTgqxZuHQ87uSIMNfoZ2TXdE7ZwcjozsqaK2t1212u13RGIU"
accessToken = "324268217-psnc80cEmLPEi50kwNlqTHFrrQlgPVzJz39uvBe"
accessTokenSecret = "fXPHG9Qk1X8vg04iFbSkZJW67dvn00qABajER8tFOImD"

#Create the authentication object
authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
#Set the access token and access token secret
authenticate.set_access_token(accessToken, accessTokenSecret)
#Create the API object
api = tweepy.API(authenticate, wait_on_rate_limit=True)

number_of_tweets=2000
tweets= []
likes= []

for i in tweepy.Cursor(api.user_timeline, id="aantonop", tweet_mode="extended").items(number_of_tweets):
    tweets.append(i.full_text)
    likes.append(i.favorite_count)

df= pd.DataFrame({'Tweets': tweets, 'Likes': likes})
df
```

	Tweets	Likes
0	As enthusiasm for #Bitcoin grows, have banks m...	261
1	Aprende a planificar la herencia de tus #cript...	44
2	Learn how to do inheritance planning for your ...	97
3	The Lightning Network can be layered on top of...	359
4	@tack_bob Q4'2022 is the planned release for t...	5
...
1995	and then they put a flag in it and said: WE OW...	223
1996	They took from open source software, open lice...	208
1997	Their infrastructure runs on top of hundreds o...	176
1998	They run a DEX (didn't invent it), based on P2...	187
1999	Their business is based on atomic swaps (didn'...	200

2000 rows x 2 columns

```
#Clean the Tweets
def cleanText(text):
    text= re.sub('#[A-Za-z0-9]+', '', text)
    text= re.sub('#[A-Za-z0-9]+', '', text)
    text= re.sub('http?://\S+', '', text)
    text= re.sub('\n', '', text)
    return text

df[df[-df.Tweets.str.contains("RT")]]

#Cleaning the Text
df['Tweets'] = df['Tweets'].apply(cleanText)

#Show the Cleaned Text
df
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
```

	Tweets	Likes
0	As enthusiasm for grows, have banks moved into...	261
1	Aprende a planificar la herencia de tus #cript...	44
2	Learn how to do inheritance planning for your ...	97
3	The Lightning Network can be layered on top of...	359
4	_bob Q4'2022 is the planned release for the th...	5
...
1995	and then they put a flag in it and said: WE OW...	223
1996	They took from open source software, open lice...	208
1997	Their infrastructure runs on top of hundreds o...	176
1998	They run a DEX (didn't invent it), based on P2...	187
1999	Their business is based on atomic swaps (didn'...	200

1789 rows x 2 columns

```

list_of_sentences= [sentence for sentence in df.Tweets]

lines=[]
for sentence in list_of_sentences:
    words = sentence.split()
    for w in words:
        lines.append(w)

lines= [re.sub(r'[^\A-Za-z0-9]+', '', x) for x in lines]

lines2=[]
for word in lines:
    if word != '':
        lines2.append(word)

s_stemmer= SnowballStemmer (language='english')

stem=[]
for word in lines2:
    stem.append(s_stemmer.stem(word))

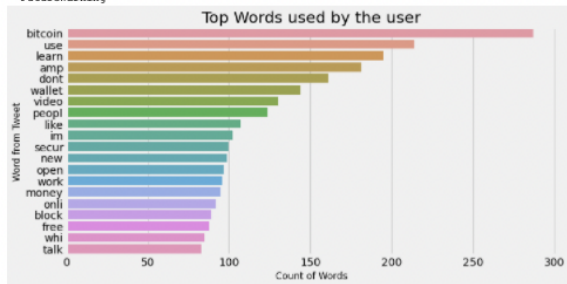
stem2=[]
for word in stem:
    if word not in nlp.Defaults.stop_words:
        stem2.append(word)

df2= pd.DataFrame(stem2)
df2=df2[0].value_counts()

df2= df2[:20,]
plt.figure(figsize=(10,5))
sns.barplot(df2.values, df2.index, alpha=1)
plt.title('Top Words used by the user')
plt.ylabel('Word from Tweet',fontsize=12)
plt.xlabel('Count of Words',fontsize=12)
plt.show()

/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keyword args: x, y. From version
FutureWarning

```



v. C.Zhao

```

#Import the libraries
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
# pip install -U spacy
# python -m spacy download en_core_web_sm
import spacy
nlp = spacy.load("en_core_web_sm")
import seaborn as sns
from nltk.stem.snowball import SnowballStemmer

```

```

#Get the Twitter API credentials
consumerKey = "aY535XcQGFyTgghjL5H7AeASw"
consumerSecret = "Mn9FqzZuIHQB7u5IMNfo8ZTXde7ZwejozsqR2t1212u13RGIU"
accessToken = "324268217-pancsOocEm1PE150kNlqTHFrrQ1gPVZJz39uvBe"
accessTokenSecret = "fXFBG9QkIX8vg041FbSkZJW67dvsn00qABajER8tFOImD"

```

```

#Create the authentication object
authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
#Set the access token and access token secret
authenticate.set_access_token(accessToken, accessTokenSecret)
#Create the API object
api = tweepy.API(authenticate, wait_on_rate_limit=True)

```

```

number_of_tweets=2000
tweets = []
likes = []

for i in tweepy.Cursor(api.user_timeline, id="cz_binance", tweet_mode="extended").items(number_of_tweets):
    tweets.append(i.full_text)
    likes.append(i.favorite_count)

df = pd.DataFrame({'Tweets': tweets, 'Likes': likes})
df

```

	Tweets	Likes
0	@Eljaboom One of the rare times we disagree, h...	602
1	No drama. Just execute.	12173
2	@antepilyarasa Crypto doesn't need a spokesper...	281
3	@artists_voyage No celebration needed. Seeing ...	328
4	@MustacheTommy Thank you for your support all ...	158
...
1753	@dinhshoo12345 I like all of the above. :)	551
1754	@Crypto_Dep @PancakeSwap @bakery_swap @VenusPr...	775
1755	#bitcoin is not a new asset class; it is a new...	16263
1756	@xbitbabe many thanks!	202
1757	@CapitalRolings @binance nice car, dude!	1757

```

#Clean the Texts
def cleanTxt(text):
    text = re.sub('[A-Za-z0-9]+', '', text)
    text = re.sub('[A-Za-z0-9]+', '', text)
    text = re.sub('https?://\S+', '', text)
    text = re.sub('\n', '', text)
    return text

```

```
df[df['Tweets'].str.contains("RT")]
```

```

#Cleaning the Text
df['Tweets'] = df['Tweets'].apply(cleanTxt)

#Show the Cleaned Text
df

```

```

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

	Tweets	Likes
0	One of the rare times we disagree, haha. But ...	602
1	No drama. Just execute.	12173
2	Crypto doesn't need a spokesperson. is king. ...	281
3	_voyage No celebration needed. Seeing you guys...	328
4	Thank you for your support all along the way...	158
...
1753	I like all of the above. :)	551
1754	_Dep _swap Most CEX are missing out on...	775
1755	is not a new asset class; it is a new technology.	16263
1756	many thanks!	202
1757	nice car, dude!	1757

1412 rows x 2 columns

```

list_of_sentences = [sentence for sentence in df.Tweets]
lines = []
for sentence in list_of_sentences:
    words = sentence.split()
    for w in words:
        lines.append(w)

```

```

lines= [re.sub(r'[A-Za-z0-9]+', '', x) for x in lines]
lines2=[]
for word in lines:
    if word != '':
        lines2.append(word)

s_stemmer= SnowballStemmer (language='english')
stem=[]
for word in lines2:
    stem.append(s_stemmer.stem(word))

stem2=[]
for word in stem:
    if word not in nlp.Defaults.stop_words:
        stem2.append(word)

df2= pd.DataFrame(stem2)
df2=df2[0].value_counts()

df2= df2[:20,]
plt.figure(figsize=(10,5))
sns.barplot(df2.values, df2.index, alpha=1)
plt.title('Top Words used by the user')
plt.ylabel('Word from Tweet',fontsize=12)
plt.xlabel('Count of Words',fontsize=12)
plt.show()

/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keyword args: x, y. From vers
FutureWarning

```

