

Ανοικτό Πανεπιστήμιο Κύπρου
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΚΑΙ ΕΦΑΡΜΟΣΜΕΝΩΝ ΕΠΙΣΤΗΜΩΝ

Μεταπτυχιακό Πρόγραμμα Σπουδών *ΠΛΗΡΟΦΟΡΙΑΚΑ ΚΑΙ
ΕΠΙΚΟΙΝΩΝΙΑΚΑ ΣΥΣΤΗΜΑΤΑ*

Μεταπτυχιακή Διατριβή



Συσχετισμός Τιμών Κρυπτονομισμάτων με Εξωγενή Φαινόμενα

Χρήστος Γκλίναβος

Επιβλέπουσα Καθηγήτρια
Δρ. Ευπραξία Ζαμάνη

Δεκέμβριος 2018

Ακαδημαϊκοί Υπεύθυνοι: Δρ. Σιαηλής Σταύρος - Δρ. Σωκράτης Κάτσικας
Επιβλέπουσα: Δρ. Ευπραξία Ζαμάνη



ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΚΑΙ ΕΦΑΡΜΟΣΜΕΝΩΝ ΕΠΙΣΤΗΜΩΝ

**ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΑ
ΣΥΣΤΗΜΑΤΑ**

Μεταπτυχιακή Διατριβή

Συσχετισμός Τιμών Κρυπτονομισμάτων με Εξωγενή Φαινόμενα

Χρήστος Γκλίναβος

Επιβλέπουσα Καθηγήτρια: Δρ. Ευπραξία Ζαμάνη



Η παρούσα μεταπτυχιακή διατριβή υποβλήθηκε προς μερική εκπλήρωση των απαιτήσεων για απόκτηση μεταπτυχιακού τίτλου σπουδών στα Πληροφοριακά και Επικοινωνιακά Συστήματα από τη Σχολή Θετικών και Εφαρμοσμένων Επιστημών του Ανοικτού Πανεπιστημίου Κύπρου.

Δεκέμβριος 2018



**Πρόγραμμα Σπουδών: Πληροφοριακά και
Επικοινωνιακά Συστήματα**

ΠΕΣ700

**Μεταπτυχιακή Διατριβή στον Συσχετισμό Τιμών
Κρυπτονομισμάτων με Εξωγενή Φαινόμενα**

Χρήστος Γκλίναβος
ΑΡ. ΦΟΙΤ. 11300812
christos.gklinavos@st.ouc.ac.cy
gklinavos@gmail.com

Δεκέμβριος 2018

Ακαδημαϊκοί Υπεύθυνοι: Δρ. Σιαηλής Σταύρος - Δρ. Σωκράτης Κάτσικας
Επιβλέπουσα: Δρ. Ευπραξία Ζαμάνη

1 Περίληψη

1.1 Γενικά

Η παρούσα μελέτη επικεντρώνεται στα ψηφιακά νομίσματα και διερευνά αν και κατά πόσο εξωγενή φαινόμενα, δηλ. φαινόμενα που δεν έχουν σχέση με την εσωτερική αξία των ψηφιακών νομισμάτων, επιδρούν στην διαμόρφωση των ισοτιμιών τους με τα πραγματικά εθνικά νομίσματα (fiat currency). Ο λόγος που γίνεται αυτό είναι ότι είναι ότι τα ψηφιακά νομίσματα, ή αλλιώς κρυπτονομίσματα στερούνται ουσιαστικής σύνδεσης με θεμελιώδη μεγέθη όπως ΑΕΠ, επιτόκια, πληθωρισμό, ανεργία, επενδύσεις και κατανάλωση (Kristoufek 2013, 1) και ως εκ τούτου ο υπολογισμός της αξίας τους με βάση την προσφορά και την ζήτηση παρουσιάζει εγγενή προβλήματα.

Επίσης λόγω της ψηφιακής τους φύσης και της στενής τους σχέσης με την επιστήμη της πληροφορικής διαθέτουν μοναδικά χαρακτηριστικά όπως σταθερή καμπύλη προσφοράς, ή εξελισσόμενη με βάση δημόσιους αλγόριθμους, και ζήτηση που είναι επιρρεπής στα φαινόμενα δικτύου (network effect) και σε εξωγενείς παράγοντες που εμφανίζονται στα κοινωνικά δίκτυα.

Συγκεκριμένα για να εξάγει συμπεράσματα εξετάζει την διακύμανση της τιμής 3 σημαντικών κρυπτονομισμάτων, του Bitcoin, του Ethereum και του Ripple σε συνάρτηση με εξωγενή φαινόμενα που αφορούν το πλήθος των σχετικών με αυτά αναζητήσεων στην κυριότερη μηχανή αναζήτησης <https://www.google.com/>, την τάση-διαμόρφωση γνώμης των σχετικών αναρτήσεων στο κοινωνικό δίκτυο Twitter <https://twitter.com/> και στα σχετικά με το καθένα, φόρουμ ανταλλαγής απόψεων όπως το <https://bitcointalk.org/>, το <https://www.xrpchat.com/> και το <https://www.reddit.com/r/ethereum/>.

1.2 Κύριοι στόχοι της έρευνας

Ο κύριος στόχος της έρευνας είναι να καταδείξει την ύπαρξη μια συσχέτισης μεταξύ της τάσης-γνώμης του κοινού, όπως αυτή διαμορφώνεται και εκφράζεται στα σύγχρονα εργαλεία επικοινωνίας (Google Search, Twitter και γνωστά φόρουμ ανταλλαγής απόψεων) και της διαμόρφωσης των τιμών (εβδομαδιαίων, ημερήσιων, ωριαίων) των αντίστοιχων κρυπτονομισμάτων. Η συσχέτιση αυτή εξετάζεται στην βάση 3 ερευνητικών ερωτημάτων (με 3 αντίστοιχα υποερωτήματα το καθένα, που αντιστοιχούν στα 3 ψηφιακά νομίσματα):

- 1ο Ερευνητικό ερώτημα: Υπάρχει σχέση μεταξύ εβδομαδιαίων τιμών τάσης αναζητήσεων στο Google Trend και εβδομαδιαίων τιμών (μέσος όρος) των 3 κρυπτονομισμάτων αντιστοίχως;
- 2ο Ερευνητικό ερώτημα: Υπάρχει σχέση μεταξύ ωριαίων τιμών τάσης αναρτήσεων (μέσος όρος) στο Twitter και ωριαίων τιμών των 3 κρυπτονομισμάτων αντιστοίχως;
- 3ο Ερευνητικό ερώτημα: Υπάρχει σχέση μεταξύ ημερήσιων τιμών τάσης αναρτήσεων (μέσος όρος) σε φόρουμ ανταλλαγής απόψεων (Bitcointalk, Reddit, Xrpchat) και ημερήσιων τιμών των 3 κρυπτονομισμάτων αντιστοίχως;

1.3 Σχέδιο έρευνας

Η έρευνα κινείται στο εμπειρικό επίπεδο, καθώς η πρόθεση είναι να ελεγχθεί η θεωρία ότι εξωγενείς παράγοντες (εκδήλωση σχετικού ενδιαφέροντος και διαμόρφωση σχετικής γνώμης σε μηχανές αναζήτησης και κοινωνικά δίκτυα) μπορούν να σχετίζονται με την διαμόρφωση της τιμής και της αξίας σημαντικών

ψηφιακών νομισμάτων. Συλλέγοντας λοιπόν εμπειρικά δεδομένα από βάσεις δεδομένων και μέσω scrapping and opinion mining tools, διεξάγεται μια παραγωγική έρευνα(theory-testing) με στόχο την επαλήθευση ή μη των αρχικών υποθέσεων.

Η επιλογή των ψηφιακών νομισμάτων έγινε με βάση την δημοφιλία τους, αλλά και την ύπαρξη μοναδικών χαρακτηριστικών, που τα διαφοροποιούν τόσο μεταξύ τους, όσο και από τα υπόλοιπα ψηφιακά νομίσματα. Τα κυριότερα είναι η μακροβιότητα του Bitcoin και οι καινοτομίες που εισήγαγε, η χρήση του Ethereum σε εφαρμογές δημιουργίας έξυπνων συμβολαίων και η πληρότητα του (Turing complete), η ταχύτητα συναλλαγών μέσω του Ripple και η υιοθέτηση της τεχνολογίας του για διατραπεζικές συναλλαγές.

Η επιλογή της των εργαλείων αναζήτησης και των κοινωνικών δικτύων βασίστηκε επίσης, στην έκταση της χρήσης τους από το ευρύ κοινό και στην σταθερή παρουσία τους κατά την διάρκεια της χρήσης των εν λόγω νομισμάτων.

1.4 Κυριότερα ευρήματα και συμπεράσματα

Διαπιστώνεται από τα ευρήματα ότι υπάρχει μια ισχυρή συσχέτιση μεταξύ των εβδομαδιαίων αναζητήσεων στην μηχανή Google search και την διαμορφωθείσα τιμή των κρυπτονομισμάτων. Η συσχέτιση αυτή ισχύει και για τα τρία ψηφιακά νομίσματα (Bitcoin, Ethereum, Ripple).

Παρόμοια συσχέτιση διαπιστώνεται και μεταξύ διαμορφωθείσας τάσης-γνώμης(ωριαία) στις αναρτήσεις στο Twitter και την τιμή του Bitcoin. Αντιθέτως δεν διαπιστώνεται συσχέτιση για τα άλλα 2 νομίσματα και τις αναρτήσεις στο Twitter.

Επίσης δεν διαπιστώνεται συσχέτιση μεταξύ των 3 ψηφιακών νομισμάτων και των αναρτήσεων στα επιλεγέντα φόρουμ ανταλλαγής απόψεων.

Ωστόσο τα μεγέθη του δείγματος των υπολοίπων κρυπτονομισμάτων είναι σαφώς μικρότερα, γεγονός που επιδεινώνεται από τις διαδικασίες φιλτραρίσματος των μηνυμάτων και από τις αναπόφευκτές αδυναμίες των τεχνικών της ανάλυσης γνώμης-τάσης.

2 Abstract

2.1 In general

The present study focuses on digital coins and investigates whether and to what extent external effects, i.e. phenomena unrelated to the intrinsic value of digital currencies, affect the formation of their exchange rates with the actual currencies (fiat currency). The reason for this is, that digital currencies or otherwise crypto bonds lack substantial connectivity with fundamentals such as GDP, interest rates, inflation, unemployment, investment and consumption(Kristoufek 2013) and therefore their monetary valuation based on supply and demand has inherent problems.

Also, due to their digital nature and their close relationship with information science, they have unique features such as a steady supply curve, or an evolution based on public algorithms, and a demand that is prone to network effects and to extrinsic factors appearing in social networks.

In particular, in order to draw conclusions, it examines the fluctuation of the value of 3 major crypto coins, Bitcoin, Ethereum and Ripple, in response to exogenous phenomena related to the number of related searches on the main search engine <https://www.google.com/>, the trend-setting opinion of the relevant posts on the Twitter <https://twitter.com/> and of the relevant posts on exchange view forums such

as <https://bitcointalk.org/>, <https://www.xrpchat.com/>,
<https://www.reddit.com/r/ethereum/>.

2.2 Main objectives of research

The primary objective of the survey is to demonstrate the correlation between the public opinion trend as it is shaped and expressed in the modern communication tools (Google Search, Twitter, and popular exchange forums) and price formation (weekly, daily, hourly) of the corresponding crypto bonds. This correlation is examined on the basis of 3 research questions (with 3 corresponding sub-queries each corresponding to the 3 digital coins):

- 1st Research Question: Is there a relationship between Weekly trend values in Google Trend and weekly (average) values of the three crypto bonds respectively?
- 2nd Research Question: Is there a relationship between hourly posting trend values (average) on Twitter and hourly values of 3 crypto bonds respectively?
- 3rd Research Question: Is there a relationship between daily posting trend values (average) in forums for the exchange of views (Bitcointalk, Reddit, Xrpchat) and daily values of 3 crypto bonds respectively.

2.3 Plan of research

Research is moving on the empirical level as the intention is to test the theory that external factors (expression of relative interest and the formation of relative opinion in search engines and social networks) may be related to the formation of the value and value of significant digital currencies. Thus, by collecting empirical data from databases and through scrapping and opinion mining tools, we conduct a theoretical test with the aim of verifying the original case or not.

The choice of digital coins was based on their popularity and the existence of unique features that differentiate them both from each other and from other digital coins. The most important are Bitcoin's longevity and the innovations it introduced, the use of Ethereum in smart contracts applications and its completeness (Turing complete), the speed of trading through Ripple and the adoption of its technology for interbank transactions.

The choice of search tools and social networks has also been based on the extent of its use by the general public and on their consistent presence during the use of these currencies.

2.4 Main findings and conclusions

It can be seen from the findings that there is a strong correlation between weekly searches on the google search engine and the configured value of cryptos. This correlation applies to all three digital coins (Bitcoin, Ethereum, Ripple).

A similar correlation is also found between the trend-consensus (hourly) on Twitter posts and the Bitcoin price. Instead, there is no correlation for the other 2 currencies and posts on Twitter.

There is also no correlation between the 3 digital coins and the posts in the selected forums.

However, the sample sizes of the other crypto bonds are significantly smaller, which is exacerbated by the messages filtering processes and the inevitable weaknesses of the opinion-trend analysis.

Περιεχόμενα

Κεφάλαιο 1	3
Εισαγωγή	3
1.1 Σημασία της έρευνας	3
1.2 Σχετικές έρευνες και εναλλακτικές θεωρητικές προσεγγίσεις	4
1.3 Ερευνητικά ερωτήματα	4
1.4 Σημαντικότητα της έρευνας	5
1.5 Διάρθρωση της έρευνας	5
Κεφάλαιο 2	7
Βιβλιογραφική επισκόπηση	7
2.1 Εισαγωγή	7
2.2 Ανασκόπηση Βιβλιογραφίας	7
Κεφάλαιο 3	12
Έρευνα	12
3.1 Ερευνητικά ερωτήματα αναλυτικά	12
3.2 Μεθοδολογικό πλαίσιο	13
3.3 Συλλογή δεδομένων και προσέγγιση ανάλυσης	14
3.3.1 Ανάλυση Google trends και συσχέτιση με τιμές	14
3.3.2 Ανάλυση τάσης-γνώμης στο Twitter και συσχέτιση με τιμές	15
3.3.3 Ανάλυση τάσης-γνώμης σε Forums και συσχέτιση με τιμές	19
3.4 Ερευνητικά δεδομένα	22
3.4.1 Google trends και τιμές νομισμάτων	22
1ο Ερευνητικό Ερώτημα-Google trends και τιμές νομισμάτων:	22
3.4.1.1 Google trends και τιμές του Bitcoin	22
• Υποερώτημα 1A-Google trends και τιμές Bitcoin:	22
3.4.1.2 Google trends και τιμές του Ethereum	28
• Υποερώτημα 1B-Google trends και τιμές Ethereum:	28
3.4.1.3 Google trends και τιμές του Ripple	33
• Υποερώτημα 1Γ-Google trends και τιμές Ripple::	33

3.4.2	Ανάλυση τάσης-γνώμης στο Twitter και τιμές νομισμάτων	38
	2ο Ερευνητικό ερώτημα-Twitter και τιμές νομισμάτων:	38
3.4.2.1	Ανάλυση τάσης-γνώμης στο Twitter και τιμές του Bitcoin	38
	• Υποερώτημα 2Α-Twitter και τιμές Bitcoin:	38
3.4.2.2	Ανάλυση τάσης-γνώμης στο Twitter και τιμές του Ethereum	42
	• Υποερώτημα 2Β-Twitter και τιμές Ethereum:	42
3.4.2.3	Ανάλυση τάσης-γνώμης στο Twitter και τιμές του Ripple	46
	• Υποερώτημα 2Γ-Twitter και τιμές Ripple:	46
3.4.3	Ανάλυση τάσης-γνώμης σε Forums και τιμές νομισμάτων	51
	3ο Ερευνητικό ερώτημα-Forums και τιμές νομισμάτων:	51
3.4.3.1	Ανάλυση τάσης-γνώμης στο Bitcointalk και τιμές του Bitcoin	51
	• Υποερώτημα 3Α-Bitcointalk και τιμές Bitcoin:	51
3.4.3.2	Ανάλυση τάσης-γνώμης στο Reddit και τιμές του Ethereum	55
	• Υποερώτημα 3Β-Reddit και τιμές Ethereum:	55
3.4.3.3	Ανάλυση τάσης-γνώμης στο Xrchat και τιμές του Ripple	58
	• Υποερώτημα 3Γ-Xrchat και τιμές Ripple:	58
	Κεφάλαιο 4	62
	Επίλογος	62
4.1	Γενικά Συμπεράσματα	62
4.2	Περιορισμοί της έρευνας	64
4.3	Προτάσεις	65
	Κεφάλαιο 5	66
	Βιβλιογραφία	66
	Κεφάλαιο 6	68
	Παράρτημα	68
6.1	Πίνακας εικόνων και γραφημάτων	68
6.2	Αρχεία Δεδομένων	70

Κεφάλαιο 1

Εισαγωγή

1.1 Σημασία της έρευνας

Από την εμφάνιση του πρώτου ψηφιακού νομίσματος (Bitcoin) με αποκεντρωμένα χαρακτηριστικά και ισχυρή ασφάλεια ενσωματωμένη στον αλγόριθμό του, υπάρχει ένα συνεχές, αμείωτο και εξαιρετικά αυξανόμενο ενδιαφέρον για την υιοθέτηση των ψηφιακών νομισμάτων, ως μέσα συναλλαγών και όχι μόνο. Στην σημερινή εποχή αυτό το ενδιαφέρον έχει δώσει ώθηση σε περισσότερα από 100 νέα ψηφιακά νομίσματα με πληθώρα διαφορετικών χαρακτηριστικών. Επίσης έχει οδηγήσει στην συγκέντρωση μεγάλων χρηματικών ποσών που επενδύονται σε αγοραπωλησίες τόσο εδραιωμένων ψηφιακών νομισμάτων όσο και σε αρχικές προσφορές νέο-εμφανισθέντων νομισμάτων (Initial Coin Offers – ICO) και δικαιωμάτων επί αυτών (Tokens). Το όλο ζήτημα έχει εξαιρετικό ενδιαφέρον από τρεις διαφορετικές οπτικές: την τεχνολογική, την οικονομική και την κοινωνικό-πολιτική.

Από τεχνολογική σκοπιά, αξία έχει η καινοτομία και η ποιότητα των τεχνικών τους χαρακτηριστικών: ο βαθμός ασφάλειας των αλγόριθμων κρυπτογράφησης που χρησιμοποιούν, οι δυνατότητες και οι αδυναμίες του αποκεντρωμένου συστήματος καταγραφής των συναλλαγών (blockchain), η ανωνυμία που παρέχουν, η ταχύτητα των συναλλαγών, η τεχνολογία των ψηφιακών υπογραφών και η ανθεκτικότητα των συστημάτων σε κυβερνο-επιθέσεις.

Από οικονομική σκοπιά, σημασία έχει αν τα ψηφιακά νομίσματα μπορούν να επιτελέσουν επαρκώς τις τρεις βασικές λειτουργίες του χρήματος, δηλ. αν μπορούν να χρησιμοποιηθούν ως μέσα καθημερινών συναλλαγών, ως αποθήκες αξίας και ως μονάδες μέτρησης και σύγκρισης.

Από κοινωνικό-πολιτική σκοπιά, ενδιαφέρον έχουν τα οφέλη που προσφέρουν στους πολίτες, όπως ανωνυμία, απεξάρτηση από κυβερνητικές παρεμβάσεις, μείωση του κόστους συναλλαγών, διακίνηση κεφαλαίων χωρίς φραγμούς, αλλά και τα προβλήματα που επιφέρουν όπως ενίσχυση εγκληματικών συμπεριφορών, ξέπλυμα κεφαλαίων από παράνομες δραστηριότητες, φόρο-αποφυγή και φοροδιαφυγή.

Εκτός όμως από το γενικό ενδιαφέρον, μεγαλύτερο ειδικό ενδιαφέρον ενσωματώνει η παρούσα έρευνα καθώς, αποπειράται να απαντήσει αν εξωγενή κοινωνικό-τεχνολογικά φαινόμενα-όπως η τάση αναζήτησεων σε μηχανές αναζήτησης και η έκφραση απόψεων σε κοινωνικά δίκτυα μπορεί να συσχετιστεί με την διαμόρφωση της αξίας των ψηφιακών νομισμάτων. Αν η απάντηση είναι καταφατική σημαίνει ότι ο τρόπος λειτουργίας της οικονομίας των κρυπτονομισμάτων γίνεται πιο κατανοητός και πιθανός πιο προβλέψιμος. Αντίθετα, αν η απάντηση είναι αρνητική, η ενασχόληση με τα ψηφιακά νομίσματα γίνεται πιο αβέβαιη, και με δεδομένη την «μη συμμόρφωση» τους με παραδοσιακές οικονομικές θεωρίες, υπόκειται σε υψηλότερο ρίσκο.

Να σημειωθεί ότι η έρευνα αποτελεί ένα πρώτο βήμα σε αυτή την πιθανή σύνδεση και δεν αποσκοπεί να απαντήσει αν αυτά τα εξωγενή φαινόμενα μπορούν να προβλέψουν ή να επηρεάσουν την διαμόρφωση των αξιών των κρυπτονομισμάτων. Δεν επιχειρεί δηλ. να αποκαλύψει σχέσεις αιτίου και αποτελέσματος, πάρα μόνο να ανιχνεύσει ενδείξεις συσχέτισης μεταξύ αυτών.

Αντλεί λοιπόν την σημαντικότητα της τόσο από το γενικότερο ενδιαφέρον για τα ψηφιακά νομίσματα όσο και από το ειδικότερο ενδιαφέρον για τον τρόπο διαμόρφωσης της αξίας τους που ξεπερνά την παραδοσιακή οικονομική ανάλυση. Επίσης καταδεικνύει εμμέσως αλλά εμφανώς την σύγκλιση μεταξύ των τομέων της τεχνολογίας, της επιστήμης των δεδομένων(big data) και της οικονομίας, επιβεβαιώνοντας την ανάδυση του νέου τομέα του financial technology, ευρύτερα γνωστού ως FinTech.

1.2 Σχετικές έρευνες και εναλλακτικές θεωρητικές προσεγγίσεις

Εξαιτίας των παραπάνω λόγων, υπάρχει μια πληθώρα ερευνών και ερευνητικών προσεγγίσεων που προσπαθούν να συσχετίσουν εξωγενή κοινωνικά φαινόμενα με την διαμόρφωση των τιμών των ψηφιακών νομισμάτων. Οι περισσότερες εξ' αυτών εστιάζουν αποκλειστικά στο bitcoin ή γενικότερα σε ένα συγκεκριμένο ψηφιακό νόμισμα.

Επίσης συνήθως επικεντρώνονται στην ανάλυση μέσω ενός συγκεκριμένου χώρου, είτε αυτός είναι το Google trends, είτε το Twitter, είτε το Facebook, είτε κάποιο από τα πολλά φόρουμ ανταλλαγής απόψεων.

Μια καθιερωμένη προσέγγιση είναι η σύνδεση μεταξύ της διακύμανσης των τιμών ενός κρυπτονομίσματος (Bitcoin) και του όγκου των αναζητήσεων στο google trends(Matta, Lunesu, and Marchesi, n.d.) και στην Wikipedia(Kristoufek 2013). Επίσης η ανάλυση τάσης(sentiment analysis) χρησιμοποιείται ευρέως για τη συσχέτιση της διαμορφωθείσας γνώμης σε φόρουμ(Kim et al. 2016) και εφαρμογές ανταλλαγής μηνυμάτων (Bollen, Mao, and Zeng, n.d.) (Xie, Rambow, and Passonneau, n.d.) (Kaminski, n.d.) με τις εναλλαγές της τιμής κάποιου κρυπτονομίσματος, ή άλλου χρηματοοικονομικού αγαθού.

1.3 Ερευνητικά ερωτήματα

Σκοπός της έρευνας είναι, όπως και των προαναφερθέντων ερευνών, να διερευνήσει και να αποκαλύψει την ύπαρξη συσχέτισης μεταξύ εξωγενών κοινωνικών φαινομένων και της διαμόρφωσης της τιμής των ψηφιακών νομισμάτων. Για να το επιτύχει αυτό χρησιμοποιεί μια σειρά μεθόδων και στόχων που περιλαμβάνονται στις προηγούμενες έρευνες, συνθέτοντας όμως τρεις διαφορετικές προσεγγίσεις και τρία διαφορετικά αλλά βασικά κρυπτονομίσματα. Αυτό αποτυπώνεται στα παρακάτω ερευνητικά ερωτήματα:

1ο Ερευνητικό ερώτημα: Οι τιμές των κρυπτονομισμάτων σχετίζονται με τις αναζητήσεις στην μηχανή αναζήτησης Google, όπως αυτές απεικονίζονται στο εργαλείο Google Trends;

2ο Ερευνητικό ερώτημα: Οι τιμές των κρυπτονομισμάτων σχετίζονται με τις στάσεις-γνώμες στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter;

3ο Ερευνητικό ερώτημα: Οι τιμές των κρυπτονομισμάτων σχετίζονται με τις στάσεις-γνώμες στις αναρτήσεις μηνυμάτων στο αντίστοιχα forum bitcointalk, reddit και xrpchat;

1.4 Σημαντικότητα της έρευνας

Η εργασία αυτή μπορεί να θεωρηθεί σημαντική γιατί όπως προαναφέρθηκε χρησιμοποιεί τρεις διαφορετικές προσεγγίσεις(ανάλυση Google trends, ανάλυση γνώμης στο Twitter, ανάλυση γνώμης σε τρία εξειδικευμένα φόρουμ ανταλλαγής απόψεων) σε συνδυασμό με τρία διαφορετικά από άποψη λειτουργικότητας και τεχνολογίας ψηφιακά νομίσματα (Bitcoin, Ethereum, Ripple), δίνοντας έτσι μια περιεκτικότερη εικόνα(έναντι πιο εξειδικευμένων ερευνών) του θέματος προς εξέταση.

Επίσης ενώ υιοθετεί κάποιες από τις μεθόδους των προ υπαρχουσών ερευνών (π.χ. σύνδεση εβδομαδιαίας τάσης Google trends και τιμών νομισμάτων), αποφεύγει να χρησιμοποιήσει άλλες(π.χ. σύνδεση όγκου μηνυμάτων σε κοινωνικά δίκτυα και διακύμανσης τιμών νομισμάτων), δοκιμάζοντας νέες προσεγγίσεις (π.χ. σύνδεση μεταβολής του μέσου όρου των θετικών/αρνητικών γνώμων ανά ώρα με την ωριαία μεταβολή της τιμής των κρυπτονομισμάτων).

Άρα οποιαδήποτε ευρήματα έχουν ιδιαίτερη στατιστική σημασία επειδή η χρήση 9 διαφορετικών πηγών δεδομένων (Google trends, Twitter, Bitcointalk, Reddit, Xrpchat, kraken, CoinDesk, Etherscan, Yahoo finance), 3 διαφορετικών νομισμάτων, 3 διαφορετικών χρονικών πλαισίων (εβδομαδιαίο, ημερήσιο και ωριαίο), και 3 διαφορετικών προσεγγίσεων στην ανάλυσή τους (τάση αναζητήσεων στο Google, τάση μηνυμάτων Twitter, τάση συζητήσεων σε φόρουμ ανταλλαγής απόψεων) εξασφαλίζει υψηλή διαφοροποίηση και απομακρύνει σε μεγάλο βαθμό την τυχαιότητα των οποιοδήποτε εμφανισθέντων συσχετισμών.

1.5 Διάρθρωση της έρευνας

Η έρευνα χωρίζεται σε 3 μέρη που αφορούν τα τρία ερευνητικά ερωτήματα. Στο πρώτο μέρος διερευνάται η σχέση μεταξύ google trends και τιμών ψηφιακών νομισμάτων, στο δεύτερο μέρος διερευνάται η σχέση μεταξύ διαμόρφωσης γνώμης στα μηνύματα του twitter και τις τιμές των ψηφιακών νομισμάτων και στο τρίτο μέρος αναλύεται η σχέση μεταξύ γνώμων σε διαφορετικά φόρουμ και τις τιμές των κρυπτονομισμάτων.

Και στις τρεις περιπτώσεις, αρχικά συγκεντρώνονται τα δεδομένα που αφορούν κάποια συγκεκριμένη χρονική περίοδο. Αυτό γίνεται είτε με την μεταφόρτωση

αρχείων και βάσεων δεδομένων από διάφορες πηγές και σχετικές ιστοσελίδες, είτε με την χρήση εργαλείων εξόρυξης δεδομένων από τις αντίστοιχες πηγές.

Στην συνέχεια γίνεται η εισαγωγή δεδομένων σε αρχεία excel όπου γίνεται η διαμόρφωση τους και ο «καθαρισμός» τους από διπλότυπες τιμές και ελλειπίς ή αδιάφορες καταχωρήσεις. Κάποια από αυτά αναλύονται με εργαλεία ανάλυσης τάσης (sentiment analysis) και τελικά συγκρίνονται μεταξύ τους για την πιθανή ύπαρξη στατιστικής σχέσης χρησιμοποιώντας το στατιστικό εργαλείο SPSS και διερευνώντας τους συντελεστές συσχέτισης, όπως αυτός του Pearson. Η σειρά ανάλυσης είναι πάντα η ίδια, δηλ. για κάθε ερευνητικό ερώτημα πρώτα αναλύεται το Bitcoin, ύστερα το Ethereum και τέλος το Ripple.

Κεφάλαιο 2

Βιβλιογραφική επισκόπηση

2.1 Εισαγωγή

Η αναζήτηση και εύρεση των μελετών έγινε με βάση τις παρακάτω φράσεις και λέξεις κλειδιά:

Bitcoin/Ethereum/Ripple price prediction, Twitter sentiment analysis, Bitcoin/Ethereum/Ripple and Google trends, Bitcoin/Ethereum/Ripple forums, Cryptocurrency Fluctuations, Cryptocurrency price/value Fluctuation prediction, Cryptocurrency price evaluation based on web searches trend, digital currencies, number of cryptocurrency searches, social media posts, οι συντομογραφίες των κρυπτονομισμάτων BTC, ETH, XRP, value, price, τιμή, αξία, κρυπτονομίσματα, ψηφιακά νομίσματα, διακυμάνσεις τιμών.

Τα αποτελέσματα διακρίνονται σε 4 κατηγορίες:

1. Στις μελέτες που εξετάζουν την αναγκαιότητα και την σκοπιμότητα της διερεύνησης της διαμόρφωσης της τιμής των κρυπτονομισμάτων μέσω εξωγενών παραγόντων.
2. Στις μελέτες που διερευνούν την σχέση μεταξύ των σχετικών αναζητήσεων στο Google και της διακύμανσης της τιμής των κρυπτονομισμάτων.
3. Στις μελέτες που διερευνούν την σχέση μεταξύ των σχετικών δημοσιεύσεων στο Twitter και της διακύμανσης της τιμής των κρυπτονομισμάτων.
4. Στις μελέτες που διερευνούν την σχέση μεταξύ των σχετικών ανακοινώσεων σε κοινωνικά forum και της διακύμανσης της τιμής των κρυπτονομισμάτων.

2.2 Ανασκόπηση Βιβλιογραφίας

Στην 1η κατηγορία παρατηρούμε ότι οι περισσότερες μελέτες συγκλίνουν στην ανάγκη χρήσης μιας διαφορετικής προσέγγισης για την πρόβλεψη της διαμόρφωσης των τιμών πέρα από τα εργαλεία των κλασικών οικονομικών θεωριών.

Πολλές έρευνες, μελέτες και δημοσιεύσεις εστιάζουν στην αδυναμία των κρυπτονομισμάτων να επιτελέσουν κάποιες ή και όλες τις λειτουργίες του χρήματος: την χρήση του ως μέσου ανταλλαγής, αποθήκευσης της αξίας και μονάδας μέτρησης. Όπως αναφέρεται πολύ εύστοχα (Gantori et al. 2017) η δυνατότητα πληρωμής φόρων που είναι κύρια οικονομική λειτουργία, μέσω κρυπτονομισμάτων είναι αδύνατη, υπονομεύοντας την χρήση τους ως μέσα συναλλαγής. Επίσης το γεγονός ότι η προσφορά νομισμάτων είναι πάντα συγκεκριμένη, ενώ η ζήτηση μπορεί να παρουσιάσει διακυμάνσεις οδηγεί σε απότομες μεταβολές της αξίας τους, καταργώντας στην πράξη την λειτουργία τους ως μέσα διατήρησης της αξίας.

Υπάρχει επίσης η άποψη(Li and Wang 2017) ότι η τιμή ανταλλαγής των κρυπτονομισμάτων (exchange rate) βραχυπρόθεσμα και μεσοπρόθεσμα επηρεάζεται από τεχνολογικούς παράγοντες όπως η δυσκολία εξόρυξης και η δημόσια αναγνώριση στα κοινωνικά μέσα, αλλά τελικά μακροπρόθεσμα συγκλίνει με τα θεμελιώδη οικονομικά μεγέθη. Για το τελευταίο όμως υπάρχουν έντονες διαφωνίες, που ξεκινάνε από την διαπίστωση ότι μέσω οικονομετρικών μοντέλων(Cheah and Fry 2015) η αποδιδόμενη θεμελιώδης αξία νομισμάτων όπως για παράδειγμα το bitcoin είναι μηδέν και φτάνουν στο συμπέρασμα ότι παραβιάζουν(Harvey, n.d.) όλους τους θεμελιώδης κανόνες των οικονομικών.

Ως εκ τούτου, προτείνεται μια νέα προσέγγιση(Kristoufek 2013) που ξεφεύγει από οικονομικές θεωρήσεις όπως μελλοντικές χρηματοροές, ισοτιμίες αγοραστικής δύναμης και ισοτιμίες επιτοκίων και εστιάζει στην επίδραση της τεχνολογίας και ειδικότερα των κοινωνικών δικτύων και των τάσεων που διαμορφώνονται σε αυτά. Η δεύτερη κατηγορία λοιπόν αναφέρεται σε αυτή την προσέγγιση και συγκεκριμένα εστιάζει στην σχέση της τάσης των αναζητήσεων στο εργαλείο Google Trends κυρίως και των τιμών των ψηφιακών νομισμάτων.

Εδώ συναντώνται μεθοδολογίες που χρησιμοποιούν εβδομαδιαία δεδομένα όσον αφορά το google trend και ημερήσια όσον αφορά την Wikipedia, διαπιστώνοντας ισχυρή συσχέτιση μεταξύ αυτών και των τιμών του bitcoin(Kristoufek 2013) με επίπεδα γραμμικής συσχέτισης βαθμού 0.8786 (με $t(111)=19.3850[<0.01]$, p-value) και 0.8271 (με $t(786) = 41.2587[<0.01]$ p-value) για το Google Trends και την Wikipedia, αντιστοίχως. Οι εβδομαδιαίες τιμές του bitcoin ορίζονται ως ο μέσος όρος των τιμών κλεισίματος για τις συγκεκριμένες εβδομάδες, διαδικασία που ακολουθείται και στην παρούσα έρευνα.

Η σύγκριση του όγκου των συναλλαγών σε συγκεκριμένα ανταλλακτήρια με την διαμόρφωση των τιμών των νομισμάτων(Balcilar et al. 2017) δεν χρησιμοποιείται εδώ, καθώς το μέγεθος του όγκου συναλλαγών σε κάθε ανταλλακτήριο δεν μπορεί να είναι ενδεικτικό της γενικότερης κατάστασης, λόγω της σχετικά σύντομης ζωής πολλών ανταλλακτηρίων και του ποσοστού αγοράς που κατέχουν κάθε φορά και το οποίο είναι ευμετάβλητο.

Μια εναλλακτική οδό προσφέρει το Google Trends με την δυνατότητα να συγκρίνονται ταυτόχρονα μέχρι 5 διαφορετικά θέματα-λέξεις κλειδιά ως προς την σχετική δημοτικότητα τους(Matta, Lunesu, and Marchesi, n.d.) αναδεικνύοντας εκείνο που κερδίζει έδαφος. Έτσι η άνοδος της δημοτικότητας ενός νομίσματος θα μπορούσε να συνδεθεί με πιθανή άνοδο της τιμής του. Παρατηρείται όμως το τελευταίο διάστημα μια γενικότερη αύξηση του ενδιαφέροντος για το σύνολο των ψηφιακών νομισμάτων, οπότε η μέθοδος δεν προσφέρεται για την έρευνα μας.

Επίσης, η σύνδεση μεταξύ μιας πιθανής κατηγοριοποίησης των χρηστών του Google Trends σε προγραμματιστές, φιλελευθέρους, κερδοσκόπους και εγκληματίες και της αύξησης του θετικού ή αρνητικού ενδιαφέροντος (προγραμματιστές/εγκληματίες - θετικό ενδιαφέρον, φιλελεύθεροι/κερδοσκόποι - αρνητικό ενδιαφέρον) για το

bitcoin(Wilson and Yelowitz 2014) είναι ενδιαφέρουσα, αλλά με σχετικά μικρό και περιορισμένο πρακτικό ενδιαφέρον.

Το γενικότερο πρόβλημα με τις αναζητήσεις στο Google Trends είναι ότι οι αναζητήσεις γενικά φανερώνουν με έμμεσο τρόπο, κατά κύριο λόγο, τις προθέσεις των χρηστών, τις προτιμήσεις, τα ενδιαφέροντα και τις επιθυμίες τους. Δεν καταγράφουν όμως τις πράξεις και τις σκέψεις των χρηστών που συμβαίνουν και υλοποιούνται σε πραγματικό χρόνο(Schonfeld 2009), στερώντας πολύτιμες ενδείξεις και στοιχεία από τις έρευνες. Έτσι γίνεται αντιληπτή η σημασία πιο άμεσων κοινωνικών δικτύων, γεγονός που μας οδηγεί στην τρίτη κατηγορία μελετών που ασχολούνται με το Twitter.

Στην τρίτη κατηγορία υπάρχει πληθώρα δημοσιεύσεων που οφείλεται εν μέρει στις προκλήσεις της συλλογής δεδομένων από το Twitter(Bifet and Frank 2010). Όπως αναφέρεται: «τον Απρίλιο 2010, το Twitter είχε 106 εκατομμύρια εγγεγραμμένους χρήστες, και 180 εκατομμύρια μοναδικούς επισκέπτες κάθε μήνα. Η εταιρεία αποκάλυψε ότι 300,000 νέοι χρήστες εγγράφονταν κάθε μέρα και λάμβανε... συνολικά 3 δισεκατομμύρια requests την ημέρα μέσω του API που χρησιμοποιεί».

Σε αυτό αξίζει να προστεθεί η ιδιαίτερη δομή των μηνυμάτων που εκτός από το καθαρό κείμενο περιλαμβάνουν ειδικούς χαρακτήρες όπως, το @ για αναφορές σε συγκεκριμένους χρήστες, το # για την δήλωση θεμάτων και κατηγοριών και το RT για την επισήμανση των retweets.

Τα ίδια προβλήματα αντιμετώπισε και η παρούσα έρευνα και για το σκοπό αυτό χρησιμοποίησε ειδικά εργαλεία όπως το Chorus Tweetcatcher το οποίο δίνει την δυνατότητα για υπέρβαση των περιορισμών του Twitter API και αυτόματη αφαίρεση των retweets.

Πιο μεγάλη δυσκολία παρουσιάζει όμως η ερμηνεία του κειμένου αυτού καθ' αυτού και η κατάταξη της διαπιστωθείσας τάσης-γνώμης(classification) είτε σε θετική είτε σε αρνητική. Ένα κείμενο μπορεί να περιέχει σε πολύ συμπυκνόμενη μορφή αρνητικά και θετικά συναισθήματα ταυτόχρονα, ή και ειρωνικά σχόλια(Bifet and Frank 2010). Για την αποφυγή τέτοιων παρεξηγήσεων μπορεί να χρησιμοποιηθούν training data, τα οποία εφαρμόζονται στους κατάλληλους αλγόριθμους μάθησης. Στην εργασία αυτή χρησιμοποιήθηκε το γενικό μοντέλο της meaning cloud με παραμέτρους για το φιλτράρισμα των ειρωνικών σχολίων και με βαθμούς εμπιστοσύνης άνω του 90% για κάθε κείμενο και αποκλεισμό των ουδέτερων σχολίων.

Φυσικά υπάρχουν πολλά μοντέλα ανάλυσης τάσης-γνώμης και γενικότερα της διάθεσης του εκάστοτε υποκειμένου σε σχέση με κάποιο ζήτημα. Κάποια όπως το προαναφερθέν και το OpinionFinder μετράνε την διάθεση σε δύο διαστάσεις(θετική ή αρνητική), ενώ άλλα όπως το Google-Profile of Mood(Bollen, Mao, and Zeng, n.d.) είναι πιο πολύπλοκα και χρησιμοποιούν έως και 6 διαστάσεις(Calm, Alert, Sure, Vital, Kind, and Happy).

Γενικά όμως, η χρήση τους δικαιολογείται από τα βασικά ευρήματα των συμπεριφορικών οικονομικών και πλήθος σχετικών ερευνών που δείχνουν ότι οι οικονομικές αποφάσεις επηρεάζονται σε μεγάλο βαθμό από την διάθεση και το

συναίσθημα(Nofsinger 2005), σε τέτοιο βαθμό μάλιστα που είναι ασφαλές να συμπεράνουμε ότι η διάθεση και η τάση του κοινού είναι σε θέση να επηρεάσει τις τιμές οικονομικών αγαθών.

Έτσι δικαιολογείται και η επιλογή της διερεύνησής μας προς την ανάλυση της τάσης που αναδύεται στα μηνύματα των διάφορων κοινωνικών δικτύων, αν και πρέπει να ξανα-επισημανθεί ότι η διερεύνηση αφορά την ύπαρξη πιθανής συσχέτισης και όχι την κατεύθυνση αυτής.

Βέβαια υπάρχουν πολλές παρόμοιες έρευνες που διερευνούν την τάση-γνώμη αναλύοντας όρους κειμένου όπως "happy", "love", "fun", "good", "bad", "sad" and "unhappy" και συνδέοντας τους με θετικά και αρνητικά σήματα διερευνούν την συσχέτιση με τις τιμές κλεισίματος συγκεκριμένων κρυπτονομισμάτων(Kaminski, n.d.).

Ωστόσο εδώ επιχειρείται μια πιο ολοκληρωμένη προσέγγιση χρησιμοποιώντας 3 διαφορετικά ψηφιακά νομίσματα και μικρότερα χρονικά πλαίσια, χωρίς όμως να στερείται και αυτή προβλημάτων που οφείλονται στις εγγενείς αδυναμίες του εκάστοτε λεξικού που χρησιμοποιείται(Kouloumpis, Wilson, and Moore, n.d.) και οι οποίες εντοπίζονται στη κατάταξη της πολικότητάς τους σε θετική, αρνητική και ουδέτερη.

Τα ίδια περίπου προβλήματα εμφανίζονται και στην τέταρτη κατηγορία όπου επιχειρείται η σύνδεση μεταξύ θετικών/αρνητικών/ουδέτερων σχετικών σχολίων σε διαδικτυακές κοινότητες χρηστών-forums με την διακύμανση των τιμών και των αριθμό των συναλλαγών των ψηφιακών νομισμάτων.

Σε παρόμοια με την δική μας έρευνα(Kim et al. 2016), εξετάζονται τα ίδια 3 ψηφιακά νομίσματα, χρησιμοποιείται η Python ως εργαλείο περιήγησης και εξόρυξης δεδομένων και το πρόβλημα της κατάταξης των σχολίων αντιμετωπίζεται με την χρήση μηχανικής μάθησης(machine learning) και ειδικών αλγόριθμων όπως του VADER(Hutto and Gilbert, n.d.). Αυτό γίνεται ώστε να αντιμετωπιστούν επιτυχώς φαινόμενα νεολογισμών, χρήσης ιδιωτικών διαλέκτων(αργκό) και εικονιδίων(emoticons). Το μοντέλο VADER(Valence Aware Dictionary for sEntiment Reasoning) χρησιμοποιεί ειδικό λεξικό προσαρμοσμένο σε micro-blogging σε συνδυασμό με 5 γενικούς συντακτικούς και γραμματικούς κανόνες, επιτυγχάνοντας ακρίβεια κατάταξης (correlation coefficient $r = 0.881$) εφάμιλλής των ανθρώπων βαθμολογητών ($r = 0.888$). Η κατάταξη των σχολίων της έρευνας(Kim et al. 2016) σε very negative, negative, positive, and very positive, είναι τελικά παρόμοια με την δική μας(N+, N, NEU, NONEN, P, P+), δημιουργώντας ένα μοντέλο πρόβλεψης των τιμών με ακρίβεια 74% (μέση σταθμισμένη).

Γενικά στην ανάλυση των μηνυμάτων (sentiment analysis or opinion mining) των κοινωνικών μέσων και ευρύτερα οποιοδήποτε κειμένου, κυριαρχούν δύο προσεγγίσεις: Η μία που αποδίδεται ως **learning-based approaches**(Pang, Lee, and Vaithyanathan 2002), περιλαμβάνει πληθώρα μεθόδων όπως [SVM](#) (Support Vector Machines), [Naive Bayes](#) and [KNN](#) (K-Nearest Neighbor), που βασίζονται στην βασίζεται στην επιβλεπόμενη μάθηση (supervised learning), και άλλες πιο

πρόσφατες και εξελιγμένες μη επιβλεπόμενης μάθησης (unsupervised learning) όπως [LSA](#) (Latent Semantic Analysis) and [Deep Learning](#). Η άλλη προσέγγιση που αποδίδεται ως **Semantic approaches** (Turney 2001) χαρακτηρίζεται από την χρήση λεξικών (*lexicons*).

Το μοντέλο AYLIE Text Analysis ("<https://aylien.com/research/>," n.d.) που είναι αρκετά διαδεδομένο και θα αποτελούσε μια καλή επιλογή, ανήκει στην πρώτη κατηγορία και χρησιμοποιεί αλγόριθμους επεξεργασίας φυσικής γλώσσας NLP (Natural Language Processing) που εκπαιδεύονται με την χρήση ερμηνευτικών μοντέλων (representation models), όπως προτείνεται από διάφορες μελέτες (Howard and Ruder 2018) και (Kulkarni and Sattikar 2012) για την ερμηνεία κειμένων και σχολίων, έχει όμως πολλούς περιορισμούς όσον αφορά την ποσότητα και τον ρυθμό των κλήσεων στο API του και στην υπολογιστική ισχύ και χρόνο που απαιτεί.

Αντ' αυτού το μοντέλο της Meaning cloud ("<https://www.meaningcloud.com/products/sentiment-analysis/>," n.d.) που χρησιμοποιήθηκε εδώ, είναι πιο γρήγορο και ευέλικτο. Όπως τα περισσότερα μοντέλα που ακολουθούν την δεύτερη προσέγγιση (semantic approaches) είναι πιο εύκολο να διορθωθεί προσθέτοντας νέες λέξεις και να βελτιωθεί προσθέτοντας στο λεξικό του λεξιλόγιο από εξειδικευμένα domains.

Εν κατακλείδι οι περισσότερες έρευνες καταδεικνύουν ότι η συλλογική διάθεση (Cohen-Charash et al. 2013) των επενδυτών και η τάση στα μαζικά μέσα επικοινωνίας και κοινωνικής δικτύωσης (Garcia et al. 2014), (Lamon, Nielsen, and Redondo, n.d.) μπορούν να επηρεάσουν την εξέλιξη του όγκου συναλλαγών και των τιμών διάφορων οικονομικών αγαθών. Αυτό δεν σημαίνει ότι η συσχέτιση θα είναι πάντα θετική (DiFonzo and Bordia 1997), ούτε ότι η συσχέτιση απαραίτητως εκφράζεται μόνο από την τάση-γνώμη όπως διαμορφώνεται στα κοινωνικά δίκτυα (ο όγκος των μηνυμάτων μπορεί και αυτός να είναι ενδεικτικός (Abraham et al. 2018)). Ωστόσο παρέχουν βάσιμα στοιχεία ότι η διερεύνηση αυτών των εξωγενών φαινομένων αξίζει περαιτέρω προσπάθειες, οι οποίες μέσα από διαφορετικές οπτικές (ανάλυση όγκου μηνυμάτων/σχολίων, ανάλυση θετικών/αρνητικών τάσεων, αναζήτηση συσχετίσεων με μέσους όρους τιμών και μεταβολές αυτών κ.α.), θα αποδώσουν καλύτερα το μέγεθος και το είδος της διασύνδεσης μεταξύ αυτών και της αξίας των ψηφιακών νομισμάτων.

Κεφάλαιο 3

Έρευνα

3.1 Ερευνητικά ερωτήματα αναλυτικά

Όπως γίνεται ορατό από την βιβλιογραφική επισκόπηση πολλά από τα ερωτήματα που θέτει η παρούσα έρευνα έχουν ήδη τεθεί, αλλά είτε εστιάζουν σε ένα νόμισμα είτε σε μια συγκεκριμένη μεθοδολογία, είτε σε ένα συγκεκριμένο φαινόμενο. Εδώ επιχειρείται μια διερεύνηση που καλύπτει ταυτόχρονα τα 3 σημαντικότερα ψηφιακά νομίσματα σε σχέση με 3 διαφορετικές εκφάνσεις εξωγενών κοινωνικών φαινομένων με 3 διαφορετικούς τρόπους.

Έτσι στο **1ο Ερευνητικό Ερώτημα-Google trends και τιμές νομισμάτων** διερωτόμαστε: **οι τιμές των κρυπτονομισμάτων σχετίζονται με τις αναζητήσεις στην μηχανή αναζήτησης Google, όπως αυτές απεικονίζονται στο εργαλείο Google Trends;**

Το ερώτημα αναλύεται σε 3 υποερωτήματα (*Υποερώτημα 1A-Google trends και τιμές Bitcoin, Υποερώτημα 1B-Google trends και τιμές Ethereum, Υποερώτημα 1Γ-Google trends και τιμές Ripple*) που αφορούν το κάθε νόμισμα ξεχωριστά.

Στο **2ο Ερευνητικό ερώτημα-Twitter και τιμές νομισμάτων** διερωτόμαστε: **οι τιμές των κρυπτονομισμάτων σχετίζονται με τις στάσεις-γνώμες στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter;**

Το ερώτημα αναλύεται σε 3 υποερωτήματα (*Υποερώτημα 2A-Twitter και τιμές Bitcoin, Υποερώτημα 2B-Twitter και τιμές Ethereum, Υποερώτημα 2Γ-Twitter και τιμές Ripple*) που αφορούν το κάθε νόμισμα ξεχωριστά.

Τέλος στο **3ο Ερευνητικό ερώτημα- Forums και τιμές νομισμάτων** διερωτόμαστε: **οι τιμές των κρυπτονομισμάτων σχετίζονται με τις στάσεις-γνώμες στις αναρτήσεις μηνυμάτων στο αντίστοιχα forum bitcointalk, reddit και xrpchat;**

Το ερώτημα αναλύεται σε 3 υποερωτήματα (*Υποερώτημα 3A-Bitcointalk και τιμές Bitcoin, Υποερώτημα 3B-Reddit και τιμές Ethereum, Υποερώτημα 3Γ-Xrpchat και τιμές Ripple*) που αφορούν το κάθε νόμισμα ξεχωριστά.

3.2 Μεθοδολογικό πλαίσιο

Το σκεπτικό της έρευνας είναι να συγκρίνει τις τιμές διαφορετικών κρυπτονομισμάτων σε διαφορετικά χρονικά πλαίσια με τιμές τάσης-ανάλυσης γνώμης από διαφορετικά εργαλεία αναζήτησης και κοινωνικής δικτύωσης του διαδικτύου.

Για το σκοπό αυτό συλλέγονται πρωτογενή δεδομένα, που αφορούν την διαμορφωθείσα τάση στα κοινωνικά μέσα, μέσω εξειδικευμένων εργαλείων εξόρυξης (scrapping tools όπως Webhose.io, Chorus Tweetcatcher) και δευτερογενή δεδομένα που αφορούν τις τιμές των κρυπτονομισμάτων από γνωστά ανταλλακτήρια (Kraken, Coindesk, Etherscan, Yahoo finance. Επίσης δευτερογενή δεδομένα αποτελούν και τα αρχεία των τιμών τάσης που αντλούνται από το google trends.

Η έρευνα είναι κυρίως ποσοτική. Χρησιμοποιεί μέσους όρους, σύγκριση μέσων όρων και συσχετίσεις αριθμητικών αξιών. Οι συσχετίσεις βασίζονται σε δείκτες όπως του Pearson και Biserial και χρησιμοποιούν σε μεγάλο βαθμό ποσοτικά δεδομένα, όπως οι τιμές των κρυπτονομισμάτων και τάσεων.

Χρησιμοποιούνται όμως και ποιοτικά δεδομένα, όπως η κατεύθυνση της τάσης-γνώμης. Η ορθότητα του αποτελέσματος εξαρτάται από το εργαλείο ανάλυσης γνώμης (sentiment analysis) και το μοντέλο-λεξικό που χρησιμοποιεί. Το μοντέλο αυτό είναι ένα γενικής χρήσης μοντέλο που δεν βασίζεται σε εξειδικευμένους οικονομικούς όρους. Τα αποτελέσματα μεταφράζονται σε μια διχοτομική μεταβλητή που παίρνει τιμές 1 και 0 για θετικές και αρνητικές γνώμες αντιστοίχως.

Η διαδικασία επεξεργασίας των δεδομένων έχει ως εξής:

- Αρχικά άντληση τους είτε πρωτογενώς μέσω ειδικών εργαλείων, είτε δευτερογενώς από αντίστοιχες βάσεις δεδομένων.
- Εισαγωγή τους σε αρχεία excel όπου τα δεδομένα φιλτράρονται για διπλότυπες τιμές, ταξινομούνται χρονολογικά και διαχωρίζονται με συγκεκριμένα κριτήρια (βαθμοί βεβαιότητας, κατεύθυνση τάσης κ.α.).
- Εισαγωγή των ζευγών τιμών της αξίας των κρυπτονομισμάτων και της ανάλυσης τάσης-γνώμης στο στατιστικό εργαλείο SPSS και στατιστικός έλεγχος για συσχετίσεις.
- Εξαγωγή γραφικών και αριθμητικών αποτελεσμάτων σε αρχεία μορφής .sav και .sprn και τελικός σχολιασμός τους στο βασικό αρχείο word.

3.3 Συλλογή δεδομένων και προσέγγιση ανάλυσης

3.3.1 Ανάλυση Google trends και συσχέτιση με τιμές

Η πρώτη διερεύνηση αφορά τον συσχετισμό των τάσεων στις αναζητήσεις χρηστών με συγκεκριμένους όρους, μέσω της μηχανής αναζήτησης Google search και τις αντίστοιχες διακυμάνσεις των τιμών των 3 κρυπτονομισμάτων όπως απεικονίζονται σε γνωστά ανταλλακτήρια κρυπτονομισμάτων (crypto-exchanges). Πιο συγκεκριμένα με όρους αναζήτησης Bitcoin, Ethereum και Ripple ζητήθηκε η εξέλιξη της τάσης για συγκεκριμένα χρονικά διαστήματα, που εξαρτώνται από την πρώτη παρουσία του κάθε κρυπτονομίσματος και από την ουσιαστική έναρξη συναλλαγών με αυτό στην παγκόσμια κοινότητα. Πρέπει να σημειωθεί ότι για το σκοπό αυτό έγιναν δύο παραδοχές:

1. Η τάση που απεικονίζεται στο Google trend είναι σε εβδομαδιαία βάση, ενώ οι τιμές των κρυπτονομισμάτων (αξία σε δολάρια Αμερικής) από τα ανταλλακτήρια (crypto-exchanges) απεικονίζονται σε καθημερινή βάση. Για τον σκοπό αυτό σε κάθε εβδομαδιαία τιμή μονάδων τάσης αντιστοιχήθηκε ο μέσος όρος των τιμών 7 ημερών του εκάστοτε κρυπτονομίσματος.
2. Η τιμή των μονάδων τάσης αφορά την εβδομάδα που πέρασε (εβδομαδιαία βάση) ενώ ο μέσος όρος των καθημερινών τιμών του κρυπτονομίσματος αφορά την εβδομάδα που ακολουθεί. Η επιλογή αυτή έγινε πρώτον λόγω ανάγκης, λόγω του διαφορετικού τρόπου υπολογισμού της τάσης του εργαλείου Google trends και δεύτερον, λόγω του ότι ο χρονικός υπολογισμός του μέσου όρου της τιμής του κρυπτονομίσματος των επόμενων ημερών είχε περισσότερη νοηματική και υπολογιστική αξία, αφού διερευνούμε τον συσχετισμό μεταξύ τους και το αν οι τιμές διακυμαίνονται σε συσχετισμό με τις τάσεις αναζήτησης (είναι λογικό να παρατηρούμε δηλ. αν οι τιμές ακολουθούν τις τάσεις, έστω και αν αυτό δεν αποδεικνύει από μόνο του προβλεπτική ικανότητα ούτε σχέση αιτίας και αιτιατού).

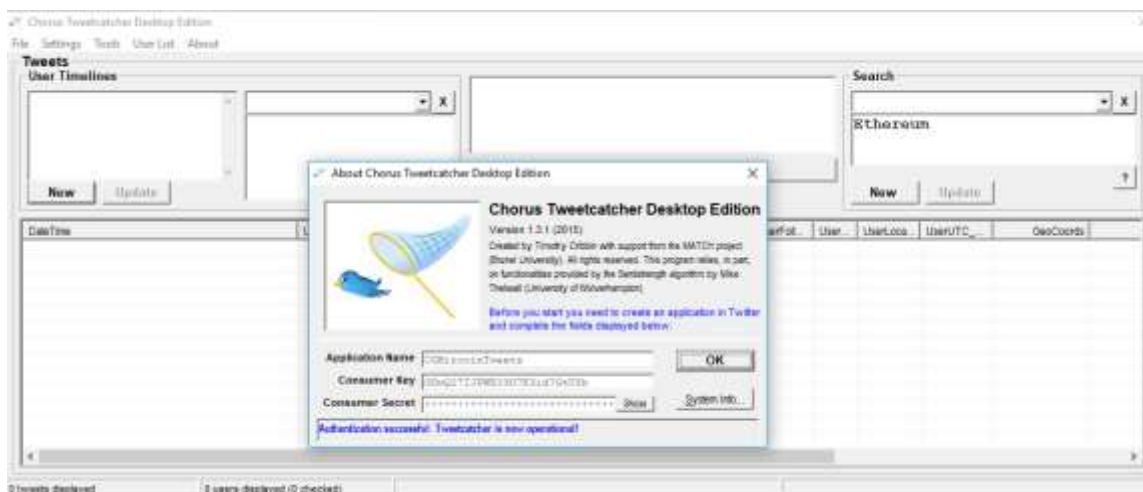
Στην συνέχεια οι τιμές αυτές (τάσης-αξίας σε δολάρια Αμερικής) απεικονίστηκαν σε αλληλεπικαλυπτόμενα γραφήματα μέσω excel, ώστε να υπάρχει μια πρώτη οπτική ένδειξη του συσχετισμού τους ή μη. Οι γραφικές παραστάσεις όπως είναι εμφανές παρακάτω, δείχνουν ισχυρή ταύτιση σε πρώτη φάση.

Το επόμενο βήμα ήταν ο έλεγχος ύπαρξης στατιστικά σημαντικού συσχετισμού των ζευγών των τιμών (correlation analysis) μέσω του στατιστικού πακέτου SPSS. Όπως επίσης είναι εμφανές παρακάτω και για τα 3 κρυπτονομίσματα υπάρχει σημαντική συσχέτιση με την τάση στις αναζητήσεις. Παρακάτω, στην παρουσίαση των ερευνητικών δεδομένων εμφανίζονται αναλυτικά οι χρονικοί περίοδοι ανάλυσης, τα γραφήματα και τα αποτελέσματα της στατιστικής ανάλυσης για κάθε κρυπτονομίσμα ξεχωριστά.

3.3.2 Ανάλυση τάσης-γνώμης στο Twitter και συσχέτιση με τιμές

Η δεύτερη διερεύνηση αφορά τον συσχετισμό του γενικότερου συναισθήματος—τάσης-γνώμης(sentiment) στις αναρτήσεις χρηστών με συγκεκριμένο θέμα, μέσω του κοινωνικού δικτύου Twitter και τις αντίστοιχες διακυμάνσεις των τιμών των 3 κρυπτονομισμάτων όπως απεικονίζονται σε γνωστό ανταλλακτήριο κρυπτονομισμάτων (<http://www.cryptodatadownload.com/>, Kraken). Πιο συγκεκριμένα με όρους θέματος Bitcoin, Ethereum και Ripple συγκεντρώθηκε πληθώρα σχετικών μηνυμάτων(tweets) για συγκεκριμένα χρονικά διαστήματα, που εξαρτώνται από την διαθεσιμότητα στοιχείων από το Twitter και την δυνατότητα επεξεργασίας των scrapping tools, καθώς ο όγκος των δεδομένων και των μηνυμάτων κειμένου είναι τεράστιος για τα συγκεκριμένα θέματα. Σαν βασικό εργαλείο scrapping αρχικά χρησιμοποιήθηκε ο operator του RapidMiner Studio v8.2 που ονομάζεται Search Twitter. Δυστυχώς τα δεδομένα που μπορούσε να αντλήσει ήταν λιγοστά εξαιτίας περιορισμών στο API του Twitter. Εντέλει χρησιμοποιήθηκε το freeware εργαλείο Chorus Tweetcatcher(Figure 0-1) με το οποίο έγινε δυνατή η άντληση tweets διάρκειας μέχρι και 10 ημερών.

Figure 0-1: Chorus Tweetcatcher Desktop Edition v.1.3.1(2015)

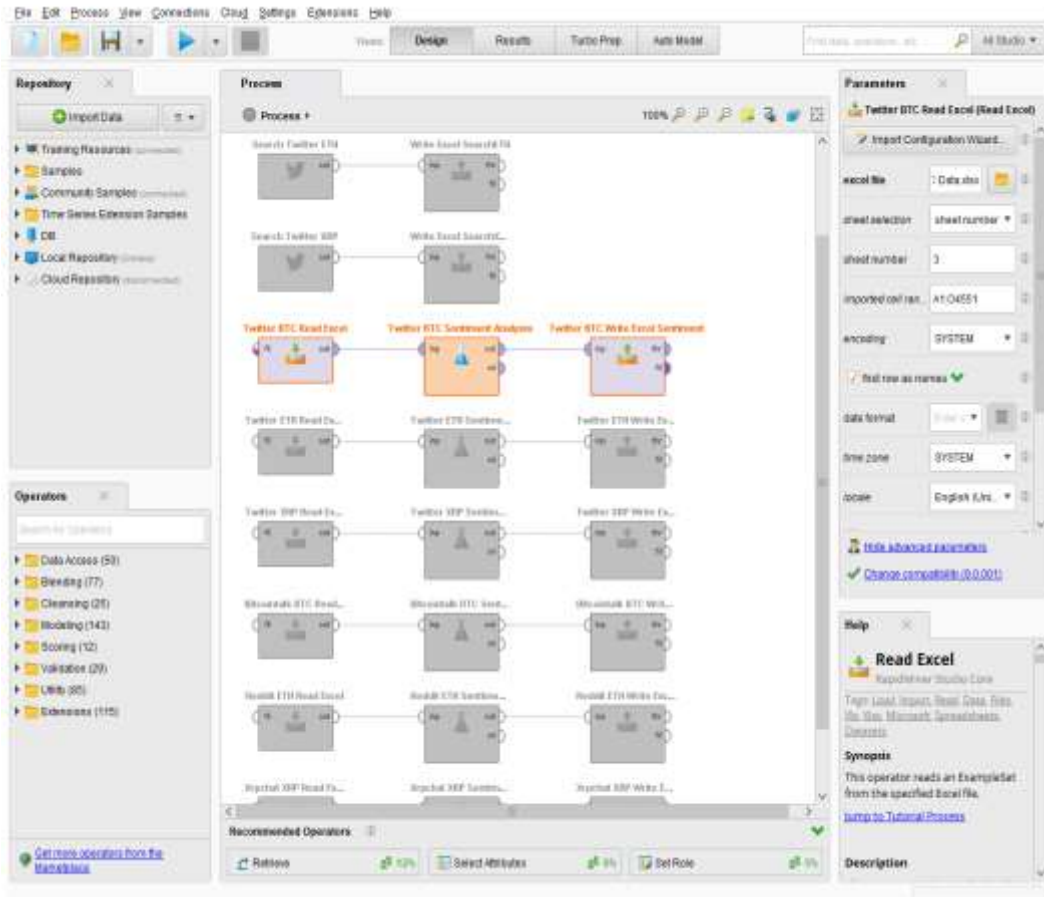


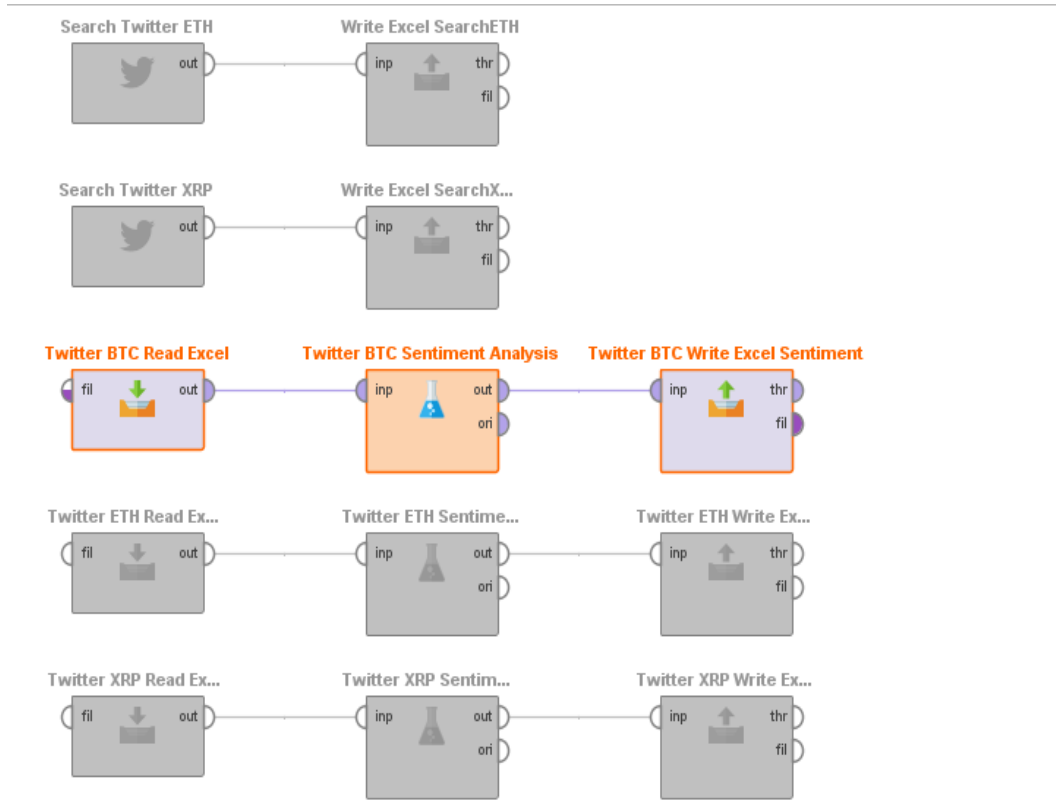
Γενικά τα βήματα που ακολουθήθηκαν για κάθε ψηφιακό νόμισμα έχουν ως εξής:

1. Άντληση tweets διάρκειας 7-10 ημερών με όρο αναζήτησης την ονομασία του κάθε νομίσματος και την συντομογραφία του.
2. Αφαίρεση των retweets και προσαρμογή των ημερομηνιών σε αριθμητική και τοπική μορφή.
3. Εισαγωγή τους σε αρχείο excel με όλα τα στοιχεία τους, όπως ημερομηνία, ώρα, tweetID, username, url κ.λ.
4. Από κάθε ημέρα έγινε φιλτράρισμα των tweets του πρώτου λεπτού κάθε ολόκληρης ώρας, ώστε να μειωθεί ο όγκος τους, σε μέγεθος εφικτό για την διενέργεια sentiment analysis σε εύλογο χρόνο.

5. Διενεργήθηκε sentiment analysis με το εργαλείο sentiment analysis της Cloud Mining για το RapidMiner(Figure 0-2) και του μοντέλου της όπως αυτό περιγράφεται στο <https://www.meaningcloud.com/developer/sentiment-analysis/doc/request#sentiment-models-language>.

Figure 0-2: RapidMiner Studio v9.0-Twitter Sentiment Analysis





6. Στην συνέχεια έγινε περαιτέρω επεξεργασία με βάση τις ακόλουθες ιδιότητες που προέκυψαν από την ανάλυση τάσης:

`polarity(Tweet)` `polarity value` `confidence(Tweet)` `agreement(Tweet)` `subjectivity(Tweet)` `irony(Tweet)` `Acception`

7. Αρχικά επιλέχθηκαν (Acception) τα μηνύματα με POLARITY<>(NONE OR NEU) & IRONY=NONIRONIC.
8. Επίσης για να είναι πιο ακριβής η ανάλυση, επιλέχθηκαν (με text filter) τα μηνύματα που περιείχαν απαραίτητα την ονομασία ή την συντομογραφία κάθε νομίσματος, ώστε να αποφευχθούν κρίσεις που αφορούσαν γενικότερα την πορεία των κρυπτονομισμάτων.
9. Για κάθε ολόκληρη ώρα τα μηνύματα ομαδοποιήθηκαν και εφόσον ο μέσος όρος της τάσης τους(P[1] για θετική Polarity, N[0] για αρνητική Polarity) ήταν μεγαλύτερος του συνολικού μέσου όρου, θεωρήθηκε ότι η γνώμη είναι POSITIVE[1] για την συγκεκριμένη ώρα, διαφορετικά NEGATIVE[0].
10. Στην συνέχεια η δυαδική αυτή μεταβλητή POSITIVE[1]-NEGATIVE[0] για κάθε ώρα αντιπαρατέθηκε με την τιμή του νομίσματος την αντίστοιχη ώρα.
11. Το ζεύγος τιμών υποβλήθηκε σε ανάλυση συσχέτισης Biserial μέσω του SPSS.

Πρέπει να σημειωθεί επίσης, ότι για το σκοπό αυτό έγιναν οι εξής παραδοχές:

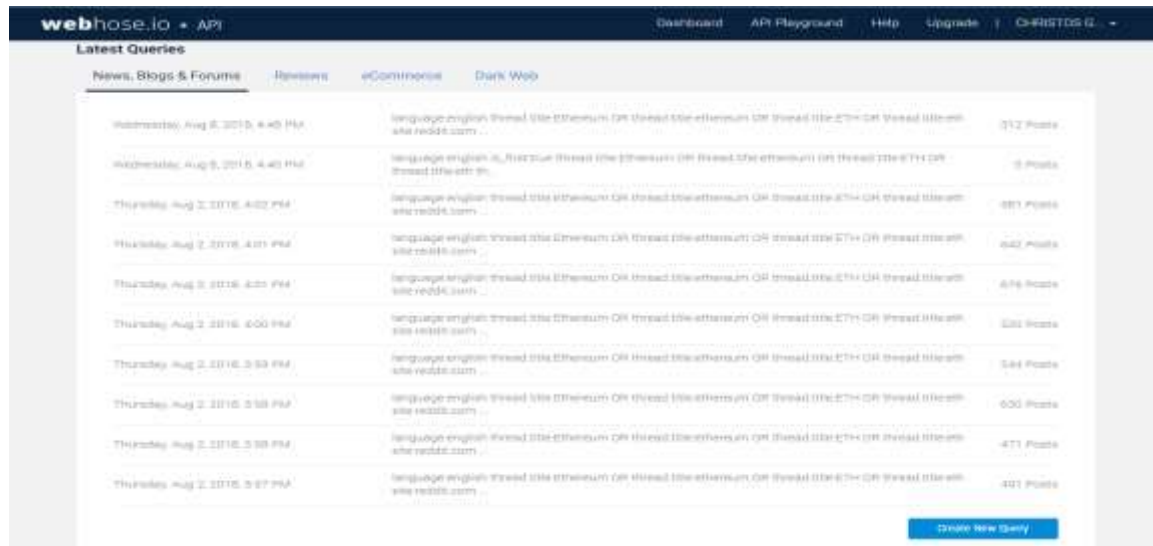
1. Όπως προαναφέρθηκε επιλέχθηκαν τα μηνύματα που δεν περιείχαν ειρωνεία ή αμφιλεγόμενη ερμηνεία (NON IRONIC) και POLARITY<>(NONE OR NEUTRAL), γιατί στην ουσία δεν διατύπωναν πιθανή γνώμη, αλλά απλή παράθεση τιμών ή γεγονότος, ή η τάση ήταν ανάμεικτα θετική και αρνητική.
2. Η τάση-γνώμη που διέπει κάθε tweet διαχωρίζεται σε Polarity=P,P+,P++ και Polarity=N,N+,N++ σύμφωνα με το λεξικό του sentiment analysis. Τα πρώτα χαρακτηρίζονται ως P[1] και τα δεύτερα ως N[0].
3. Στην συνέχεια υπολογίζεται ο μέσος όρος των P[1] και N[0] για όλα τα tweets μιας συγκεκριμένης ώρας.
4. Αν ο μέσος όρος των P[1] και N[0] είναι μεγαλύτερος ή ίσος του γενικού μέσου όρου τότε αποδίδεται η τιμή POSITIVE[1], διαφορετικά η τιμή NEGATIVE[0].
5. Υποθέτουμε ότι ο χαρακτηρισμός POSITIVE[1] θα αντιστοιχεί σε ανοδικές τιμές, ενώ ο χαρακτηρισμός Negative[0] θα αντιστοιχεί σε καθοδικές τιμές, αποκαλύπτοντας μια συσχέτιση μεταξύ των μεταβλητών της γνώμης και της αξίας του νομίσματος.
6. Στην περίπτωση του Ethereum και ειδικότερα του Ripple υπήρξε διαθέσιμος μικρότερος όγκος δεδομένων, επηρεάζοντας αρνητικά την συνέχεια των ωριαίων ζευγών γνώμης-αξίας νομίσματος.
7. Όλες οι σειρές τιμών της αξίας των νομισμάτων υποβλήθηκαν σε έλεγχο κανονικής κατανομής και όπου αυτό δεν ίσχυε έγινε μετατροπή των τιμών με βάση τον δεκαδικό λογάριθμο κάθε τιμής.

Το τελικό βήμα ήταν ο έλεγχος ύπαρξης στατιστικά σημαντικού συσχετισμού των ζευγών των τιμών (Biserial analysis) μέσω του στατιστικού πακέτου SPSS. Διαπιστώνεται παρακάτω ότι για το Bitcoin υπάρχει σημαντική συσχέτιση με την τάση στα tweets. Αντίστοιχη συσχέτιση δεν διαπιστώνεται για το Ethereum και το Ripple, όπως όμως προαναφέρθηκε, είναι δυνατό να αποδοθεί στην μη ομαλή εξέλιξη των ωριαίων ζευγών γνώμης-αξίας νομίσματος λόγω των επιλεγμένων φίλτρων και του μικρότερου σετ δεδομένων. Παρακάτω στην παρουσίαση των ερευνητικών δεδομένων εμφανίζονται αναλυτικά οι χρονικοί περίοδοι ανάλυσης, τα γραφήματα και τα αποτελέσματα της στατιστικής ανάλυσης για κάθε κρυπτονόμισμα ξεχωριστά.

3.3.3 Ανάλυση τάσης-γνώμης σε Forums και συσχέτιση με τιμές

Η τρίτη διερεύνηση αφορά τον συσχετισμό του γενικότερου συναισθήματος—τάσης-γνώμης(sentiment) στις αναρτήσεις χρηστών με συγκεκριμένο θέμα(τα 3 ψηφιακά νομίσματα), σε 3 διαφορετικά forums (<https://bitcointalk.org/>, <https://www.reddit.com/r/ethereum/>, <https://www.xrpchat.com/>) και τις αντίστοιχες διακυμάνσεις των τιμών των 3 κρυπτονομισμάτων όπως απεικονίζονται σε γνωστό ανταλλακτήριο κρυπτονομισμάτων (<http://www.cryptodatadownload.com/>, Kraken). Πιο συγκεκριμένα με όρους θέματος Bitcoin, Ethereum και Ripple και τις αντίστοιχες συντομογραφίες τους, συγκεντρώθηκε πληθώρα σχετικών αναρτήσεων-posts για συγκεκριμένα χρονικά διαστήματα, που εξαρτώνται από την διαθεσιμότητα στοιχείων από τα αντίστοιχα forums και την δυνατότητα επεξεργασίας του scrapping tool, καθώς ο όγκος των δεδομένων και των μηνυμάτων κειμένου είναι τεράστιος για τα συγκεκριμένα θέματα. Σαν βασικό εργαλείο scrapping αυτή την φορά το online εργαλείο <https://webhose.io/> (Figure 0-3) με το οποίο έγινε δυνατή η άντληση posts διάρκειας μέχρι και 15 ημερών.

Figure 0-3: Webhose.io

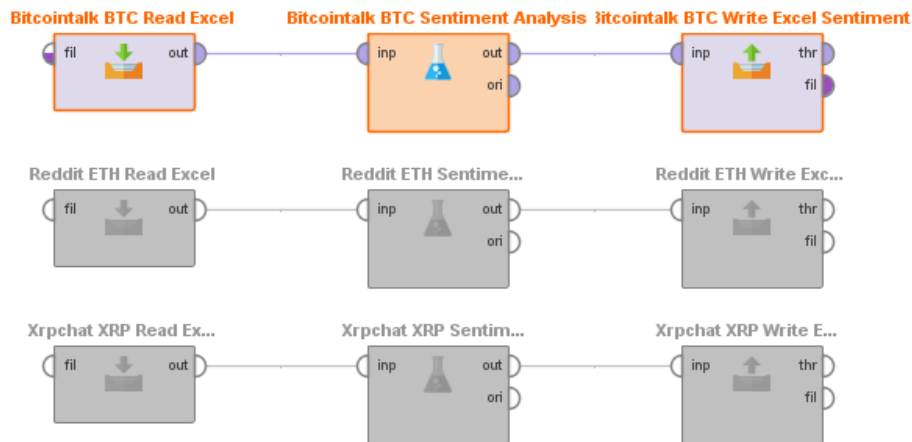
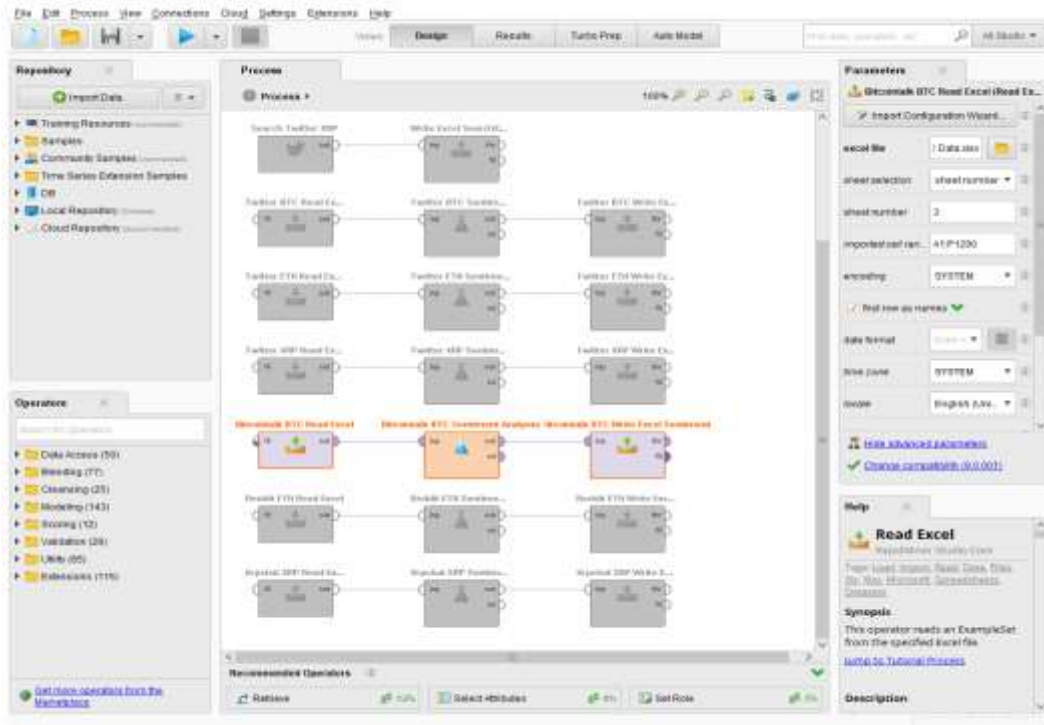


Latest Queries	News, Blogs & Forums	Reviews	eCommerce	Dark Web
Wednesday, Aug 8, 2018, 4:40 PM	language:english Thread title:Bitcoin OR Thread title:Ethereum OR Thread title:ETN OR Thread title:Bitcoin OR Thread title:ETN			
Wednesday, Aug 8, 2018, 4:40 PM	language:english & Bitcoin Thread title:Bitcoin OR Thread title:Ethereum OR Thread title:ETN OR Thread title:ETN OR Thread title:ETN			
Thursday, Aug 2, 2018, 4:02 PM	language:english Thread title:Bitcoin OR Thread title:Ethereum OR Thread title:ETN OR Thread title:ETN OR Thread title:ETN			
Thursday, Aug 2, 2018, 4:01 PM	language:english Thread title:Bitcoin OR Thread title:Ethereum OR Thread title:ETN OR Thread title:ETN OR Thread title:ETN			
Thursday, Aug 2, 2018, 4:00 PM	language:english Thread title:Bitcoin OR Thread title:Ethereum OR Thread title:ETN OR Thread title:ETN OR Thread title:ETN			
Thursday, Aug 2, 2018, 4:00 PM	language:english Thread title:Bitcoin OR Thread title:Ethereum OR Thread title:ETN OR Thread title:ETN OR Thread title:ETN			
Thursday, Aug 2, 2018, 3:59 PM	language:english Thread title:Bitcoin OR Thread title:Ethereum OR Thread title:ETN OR Thread title:ETN OR Thread title:ETN			
Thursday, Aug 2, 2018, 3:58 PM	language:english Thread title:Bitcoin OR Thread title:Ethereum OR Thread title:ETN OR Thread title:ETN OR Thread title:ETN			
Thursday, Aug 2, 2018, 3:58 PM	language:english Thread title:Bitcoin OR Thread title:Ethereum OR Thread title:ETN OR Thread title:ETN OR Thread title:ETN			
Thursday, Aug 2, 2018, 3:57 PM	language:english Thread title:Bitcoin OR Thread title:Ethereum OR Thread title:ETN OR Thread title:ETN OR Thread title:ETN			

Γενικά τα βήματα που ακολουθήθηκαν για κάθε ψηφιακό νόμισμα έχουν ως εξής:

1. Άντληση posts διάρκειας 14-15 ημερών με όρο αναζήτησης την ονομασία του κάθε νομίσματος και την συντομογραφία του.
2. Εισαγωγή τους σε αρχείο excel με όλα τα στοιχεία τους, όπως ημερομηνία, ώρα, postID, username, url κ.λ.
3. Από κάθε ημέρα έγινε φιλτράρισμα των posts για διπλότυπες τιμές.
4. Διενεργήθηκε sentiment analysis με το εργαλείο sentiment analysis της Cloud Mining για το RapidMiner(Figure 0-4) και του μοντέλου της όπως αυτό περιγράφεται στο <https://www.meaningcloud.com/developer/sentiment-analysis/doc/request#sentiment-models-language>.

Figure 0-4: RapidMiner Studio v9.0-Forums Sentiment Analysis



5. Στην συνέχεια έγινε περαιτέρω επεξεργασία με βάση τις ακόλουθες ιδιότητες που προέκυψαν από την ανάλυση τάσης:

`polarity(Text)` `polarity value` `confidence(Text)` `agreement(Text)` `subjectivity(Text)` `irony(Text)` `Acception`

6. Αρχικά επιλέχθηκαν (Acception) τα μηνύματα με POLARITY<>(NONE OR NEU) & IRONY=NONIRONIC.

7. Επίσης για να είναι πιο ακριβής η ανάλυση επιλέχθηκαν (με text filter) τα μηνύματα που περιείχαν απαραίτητα την ονομασία ή την συντομογραφία κάθε νομίσματος, ώστε να αποφευχθούν κρίσεις που αφορούσαν γενικότερα την πορεία των κρυπτονομισμάτων.
8. Για κάθε ημέρα τα μηνύματα ομαδοποιήθηκαν και εφόσον ο μέσος όρος της τάσης τους (P[1] για θετική Polarity, N[0] για αρνητική Polarity) ήταν μεγαλύτερος του συνολικού μέσου όρου, θεωρήθηκε ότι η γνώμη είναι POSITIVE[1] για την συγκεκριμένη ημέρα, διαφορετικά NEGATIVE[0].
9. Στην συνέχεια η δυαδική αυτή μεταβλητή POSITIVE[1]-NEGATIVE[0] για κάθε ημέρα αντιπαρατέθηκε με την τιμή του νομίσματος την αντίστοιχη ημέρα.
10. Το ζεύγος τιμών υποβλήθηκε σε ανάλυση συσχέτισης Biserial μέσω του SPSS.

Πρέπει να σημειωθεί επίσης, ότι για το σκοπό αυτό έγιναν οι εξής παραδοχές:

1. Όπως προαναφέρθηκε επιλέχθηκαν τα μηνύματα που δεν περιείχαν ειρωνεία ή αμφιλεγόμενη ερμηνεία (NON IRONIC) και POLARITY<>(NONE OR NEUTRAL), γιατί στην ουσία δεν διατύπωναν πιθανή γνώμη, αλλά απλή παράθεση τιμών ή γεγονότος, ή η τάση ήταν ανάμεικτα θετική και αρνητική.
2. Η τάση-γνώμη που διέπει κάθε post διαχωρίζεται σε Polarity=P,P+,P++ και Polarity=N,N+,N++ σύμφωνα με το λεξικό του sentiment analysis. Τα πρώτα χαρακτηρίζονται ως P[1] και τα δεύτερα ως N[0].
3. Στην συνέχεια υπολογίζεται ο μέσος όρος των P[1] και N[0] για όλα τα posts μιας συγκεκριμένης ημέρας.
4. Αν ο μέσος όρος των P[1] και N[0] είναι μεγαλύτερος ή ίσος του γενικού μέσου όρου τότε αποδίδεται η τιμή POSITIVE[1], διαφορετικά η τιμή NEGATIVE[0].
5. Υποθέτουμε ότι ο χαρακτηρισμός POSITIVE[1] θα αντιστοιχεί σε ανοδικές τιμές, ενώ ο χαρακτηρισμός Negative[0] θα αντιστοιχεί σε καθοδικές τιμές, αποκαλύπτοντας μια συσχέτιση μεταξύ των μεταβλητών της γνώμης και της αξίας του νομίσματος.
6. Όλες οι σειρές τιμών της αξίας των νομισμάτων υποβλήθηκαν σε έλεγχο κανονικής κατανομής και όπου αυτό δεν ίσχυε έγινε μετατροπή των τιμών με βάση τον δεκαδικό λογάριθμο κάθε τιμής.

Το τελικό βήμα ήταν ο έλεγχος ύπαρξης στατιστικά σημαντικού συσχετισμού των ζευγών των τιμών (Biserial analysis) μέσω του στατιστικού πακέτου SPSS. Διαπιστώνεται παρακάτω ότι γενικά δεν υπάρχει σημαντική συσχέτιση με την τάση στα posts. Παρακάτω στην παρουσίαση των ερευνητικών δεδομένων, εμφανίζονται αναλυτικά οι χρονικοί περίοδοι ανάλυσης, τα γραφήματα και τα αποτελέσματα της στατιστικής ανάλυσης για κάθε κρυπτονομίσμα ξεχωριστά.

3.4 Ερευνητικά δεδομένα

3.4.1 Google trends και τιμές νομισμάτων

Η πρώτη διερεύνηση όπως προαναφέρθηκε, αφορά τον συσχετισμό των τάσεων στις αναζητήσεις χρηστών με συγκεκριμένους όρους, μέσω της μηχανής αναζήτησης google search και τις αντίστοιχες διακυμάνσεις των τιμών των 3 κρυπτονομισμάτων όπως απεικονίζονται σε γνωστά ανταλλακτήρια κρυπτονομισμάτων (crypto-exchanges). Πιο συγκεκριμένα με όρους αναζήτησης Bitcoin, Ethereum και Ripple ζητήθηκε η εξέλιξη της τάσης για συγκεκριμένα χρονικά διαστήματα όπως ορίζονται παρακάτω. Υπενθυμίζεται το:

1ο Ερευνητικό Ερώτημα-Google trends και τιμές νομισμάτων: Οι τιμές των κρυπτονομισμάτων σχετίζονται με τις αναζητήσεις στην μηχανή αναζήτησης Google, όπως αυτές απεικονίζονται στο εργαλείο Google Trends;

3.4.1.1 Google trends και τιμές του Bitcoin

Η αναζήτηση στο εργαλείο google trends έγινε με μοναδικό όρο το “Bitcoin” και αφορούσε την περίοδο 04/01/15 έως 31/12/17, μια επαρκής περίοδος εξέτασης που περιλαμβάνει τις σχετικά πρόσφατες σημαντικές εξελίξεις στην τιμή του νομίσματος, αλλά και μεγάλο μέρος της επέκτασης και αποδοχής του. Οι τιμές της αξίας του σε δολάρια Αμερικής αντλήθηκαν από το coindesk (<https://www.coindesk.com/price/>) και αφορούσαν το διάστημα 04/01/15 έως 06/01/18

- **Υποερώτημα 1A-Google trends και τιμές Bitcoin:** Η τιμή του κρυπτονομίσματος Bitcoin σχετίζεται με τις αναζητήσεις στην μηχανή αναζήτησης Google, όπως αυτές απεικονίζονται στο εργαλείο Google Trends;

Πρόταση H0: Η μέση εβδομαδιαία τιμή του κρυπτονομίσματος Bitcoin δεν σχετίζεται με την εβδομαδιαία τιμή της τάσης αναζήτησης στο εργαλείο Google Trends.

Πρόταση H1: Η μέση εβδομαδιαία τιμή του κρυπτονομίσματος Bitcoin σχετίζεται με την εβδομαδιαία τιμή της τάσης αναζήτησης στο εργαλείο Google Trends.

Τα δεδομένα από τα αρχικά γραφήματα τάσης-τιμών(Figure 0-5 και Figure 0-6) μετατρέπονται σε γραφήματα Excel(Figure 0-7 και Figure 0-8) και στην συνέχεια επιχειρείται μια πρώτη οπτική διερεύνηση του συσχετισμού τους, με την προβολή τους σε αλληλοεπικάλυψη(Figure 0-9). Παρατηρούμε την ταύτιση των γραφικών παραστάσεων και όπως φαίνεται η τάση και οι τιμές κινούνται παράλληλα μαρτυρώντας έντονο συσχετισμό.

Figure 0-5: Γράφημα Bitcoin από google trends 01/01/2015-01/01/2018

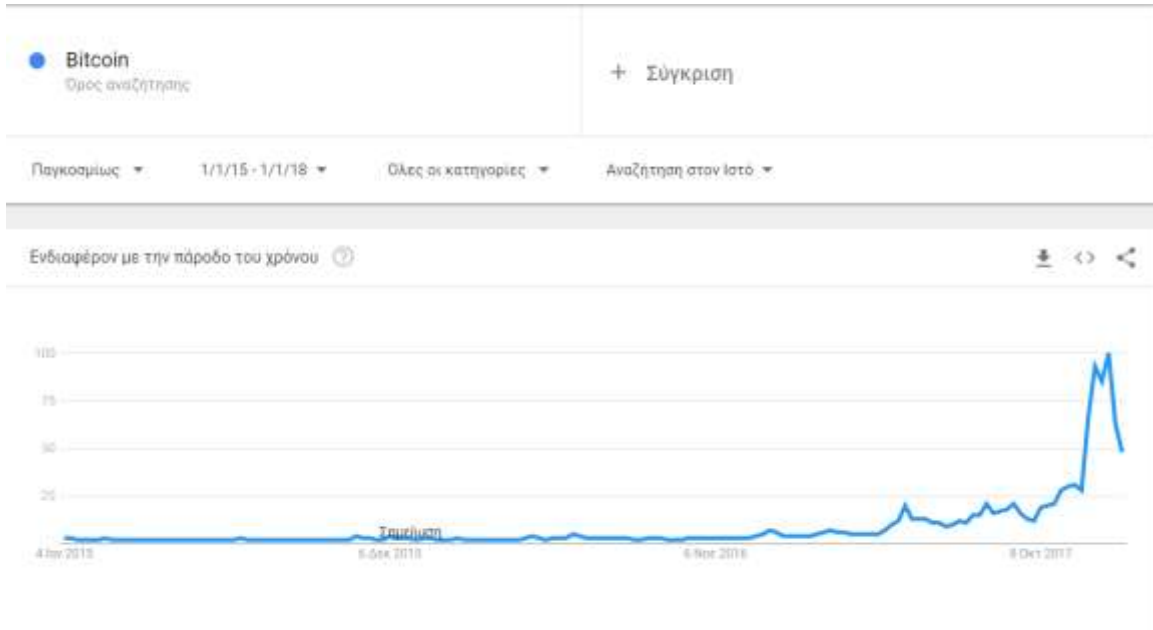


Figure 0-6: Γράφημα BitcoinUSD Price από Coindesk



Figure 0-7: Γράφημα Bitcoin από Google trends σε Excel

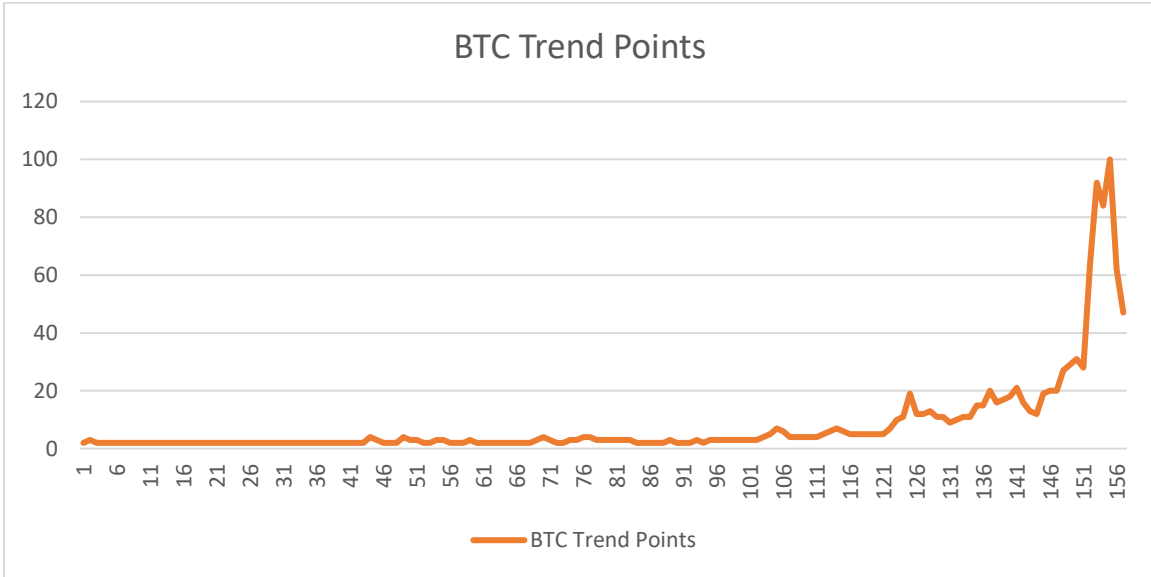


Figure 0-8: Γράφημα BitcoinUSD Price από Coindesk σε Excel

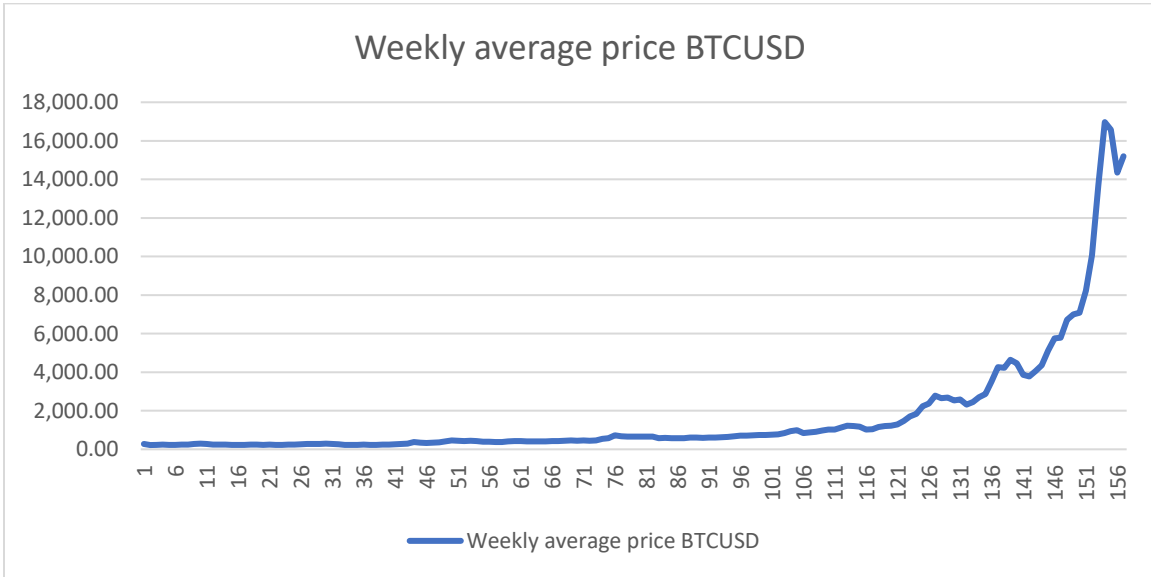
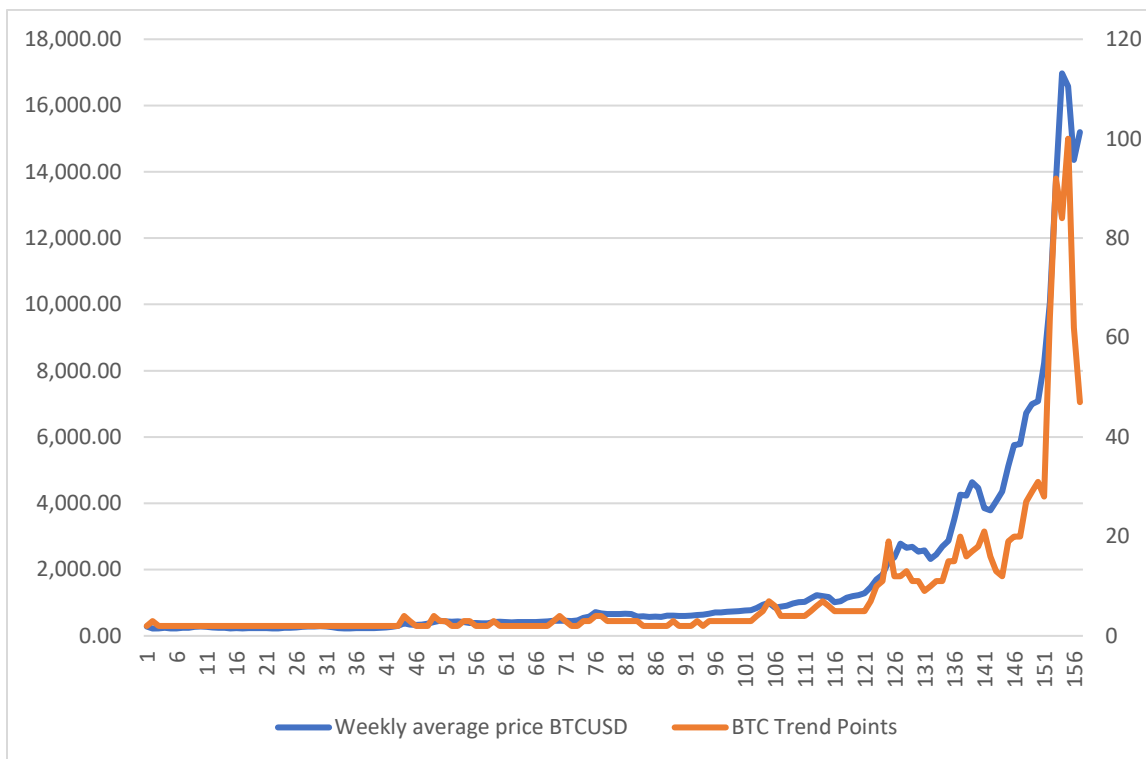


Figure 0-9: Αλληλεπικαλυπτόμενα γραφήματα Bitcoin από Google trends και BitcoinUSD Price από Coindesk σε Excel



Στην στατιστική ανάλυση που ακολουθεί με τα ίδια ζεύγη τιμών παρατηρούμε οπτικά στο διάγραμμα διασποράς(Figure 0-10) ότι οι μέσες εβδομαδιαίες τιμές του νομίσματος σε σχέση με τις τιμές τάσης του Google trend ομαδοποιούνται ως προς την περιοχή και παρουσιάζουν εικόνα αναλογικότητας (αυξάνουν ταυτόχρονα σε μεγάλο βαθμό και οι δύο), γεγονός που καταδεικνύει συσχέτιση. Αλλά και αριθμητικά μέσω των 3 διαφορετικών (παραμετρικών και μη) συντελεστών συσχέτισης (Pearson's, Kendall's, Spearman's-Figure 0-11), λόγω του ότι η τιμή **p-value είναι μικρότερη του επιπέδου σημαντικότητας $\alpha=0.05$** που έχουμε θέσει ως όριο και οι αντίστοιχοι συντελεστές συσχέτισης μεγαλύτεροι του **+0.5**, επιβεβαιώνεται ότι υπάρχει στατιστικά σημαντική συσχέτιση και μάλιστα ισχυρά θετική, μεταξύ των μεταβλητών της τάσης των αναζητήσεων και της τιμής του κρυπτονομίσματος BTC.

Figure 0-10: SCATTERPLOT(BIVAR)=BTC_Trend_Points WITH Weekly_average_price_BTCUSD

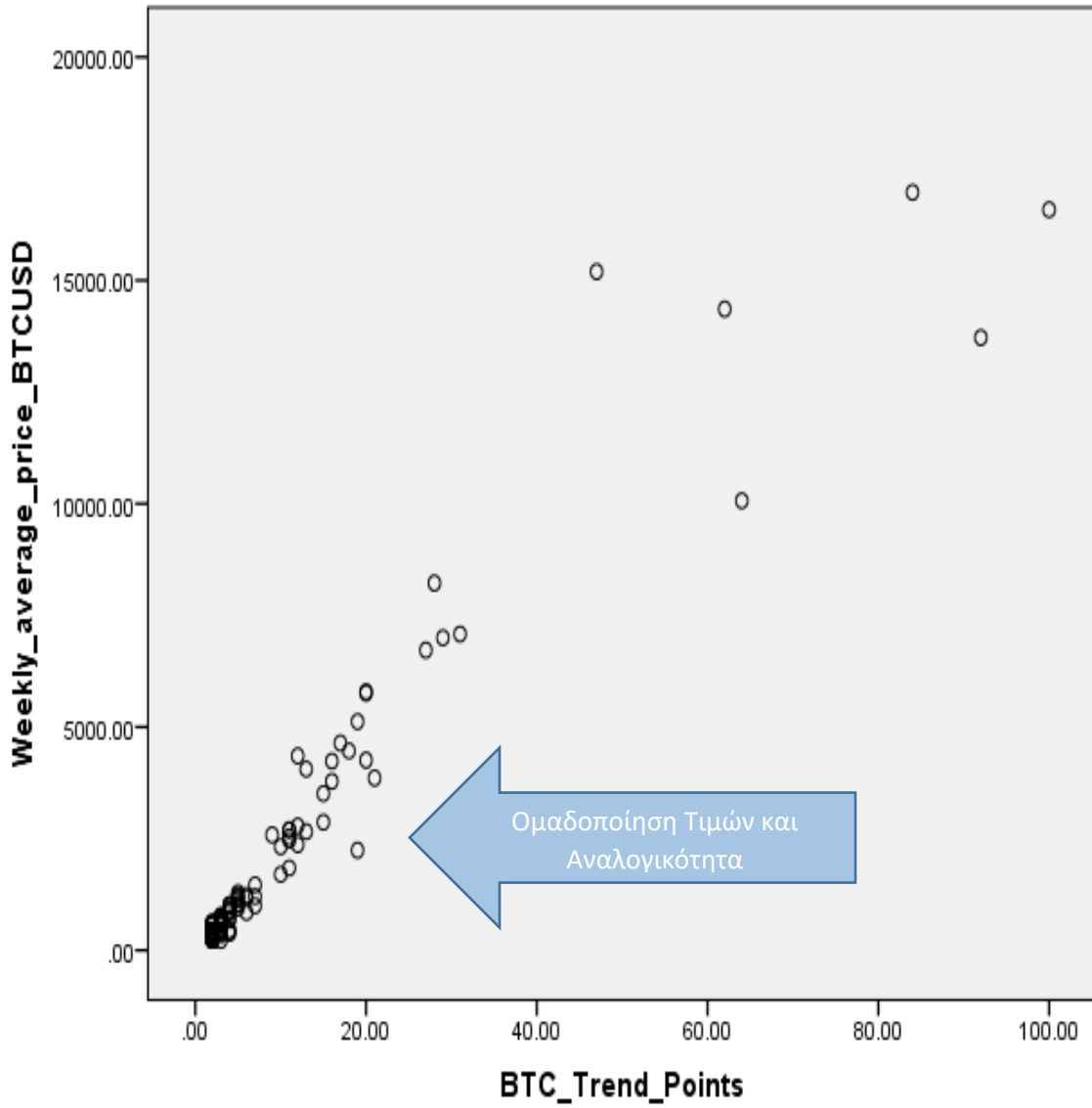


Figure 0-11: CORRELATIONS=BTC_Trend_Points Weekly_average_price_BTCUSD

Correlations			BTC_Trend_Points	Weekly_average_price_BTCUSD
BTC_Trend_Points	Pearson Correlation		1	
	Sig. (2-tailed)			.000
Weekly_average_price_BTCUSD	Pearson Correlation		.960**	1
	Sig. (2-tailed)		.000	

Non-Parametric Correlations			BTC_Trend_Points	Weekly_average_price_BTCUSD
Kendall's tau_b	BTC_Trend_Points	Correlation Coefficient	1.000	
		Sig. (2-tailed)	.000	.790**
	Weekly_average_price_BTCUSD	Correlation Coefficient	.790**	1.000
		Sig. (2-tailed)	.000	.000
Spearman's rho	BTC_Trend_Points	Correlation Coefficient	1.000	
		Sig. (2-tailed)	.000	.898**
	Weekly_average_price_BTCUSD	Correlation Coefficient	.898**	1.000
		Sig. (2-tailed)	.000	.000

** . Correlation is significant at the 0.05 level (2-tailed).

Συμπερασματικά μπορούμε να πούμε ότι απορρίπτεται η **Πρόταση H0**: και αντίθετα ισχύει η **Πρόταση H1**: Η μέση εβδομαδιαία τιμή του κρυπτονομίσματος Bitcoin σχετίζεται με την εβδομαδιαία τιμή της τάσης αναζήτησης στο εργαλείο Google Trends.

3.4.1.2 Google trends και τιμές του Ethereum

Η αναζήτηση στο εργαλείο google trends έγινε με μοναδικό όρο το “Ethereum” και αφορούσε την περίοδο 06/09/15 έως 31/12/17, μια επαρκής περίοδος εξέτασης που περιλαμβάνει τις σχετικά πρόσφατες σημαντικές εξελίξεις στην τιμή του νομίσματος αλλά και μεγάλο μέρος της επέκτασης και αποδοχής του, καθώς και την διαθεσιμότητα των στοιχείων. Οι τιμές της αξίας του σε δολάρια Αμερικής αντλήθηκαν από το Etherscan (<https://etherscan.io/chart/etherprice>) και αφορούσαν το διάστημα 06/09/15 έως 09/01/18

- **Υποερώτημα 1B-Google trends και τιμές Ethereum:** Η τιμή του κρυπτονομίσματος Ethereum σχετίζεται με τις αναζητήσεις στην μηχανή αναζήτησης Google, όπως αυτές απεικονίζονται στο εργαλείο Google Trends;

Πρόταση H0: Η μέση εβδομαδιαία τιμή του κρυπτονομίσματος Ethereum δεν σχετίζεται με την εβδομαδιαία τιμή της τάσης αναζήτησης στο εργαλείο Google Trends.

Πρόταση H1: Η μέση εβδομαδιαία τιμή του κρυπτονομίσματος Ethereum σχετίζεται με την εβδομαδιαία τιμή της τάσης αναζήτησης στο εργαλείο Google Trends.

Τα δεδομένα από τα αρχικά γραφήματα τάσης-τιμών(Figure 0-12 και Figure 0-13) μετατρέπονται σε γραφήματα Excel(Figure 0-14 και Figure 0-15) και στην συνέχεια επιχειρείται μια πρώτη οπτική διερεύνηση του συσχετισμού τους με την προβολή τους σε αλληλοεπικάλυψη(Figure 0-16). Παρατηρούμε την ταύτιση των γραφικών παραστάσεων και όπως φαίνεται η τάση και οι τιμές κινούνται παράλληλα μαρτυρώντας έντονο συσχετισμό.

Figure 0-12: Γράφημα Ethereum από Google trends 06/09/2015-01/01/2018

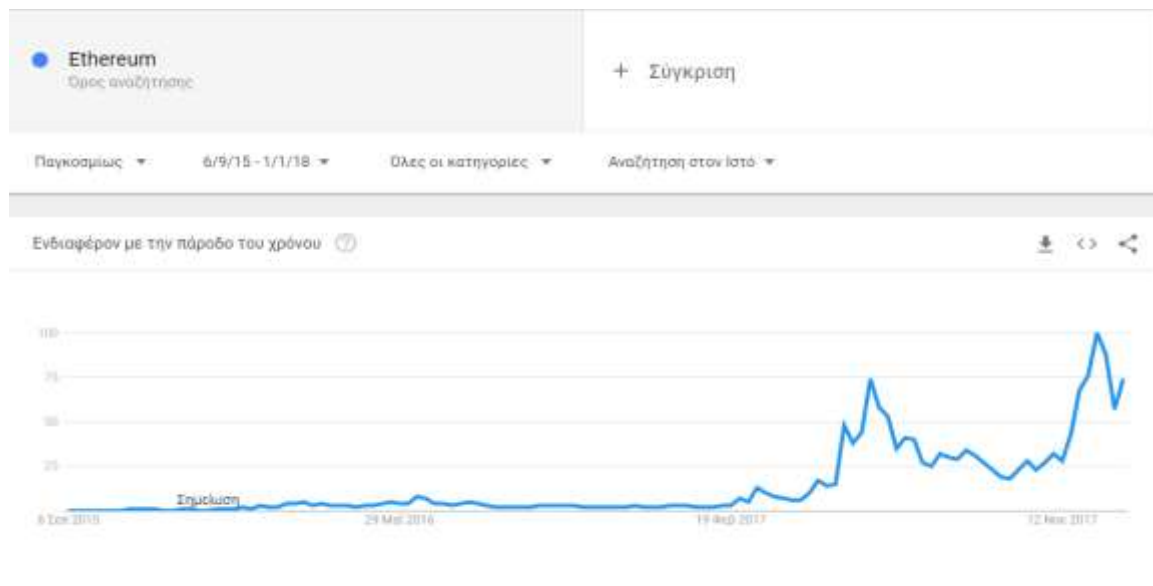


Figure 0-13: Γράφημα EthereumUSD Price από Etherscan

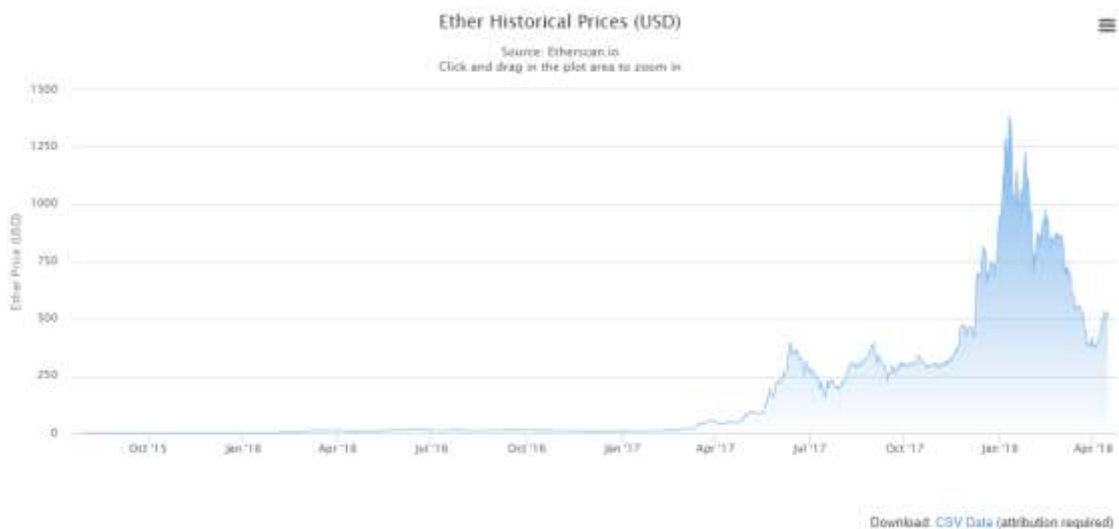


Figure 0-14: Γράφημα Ethereum από Google trends σε Excel

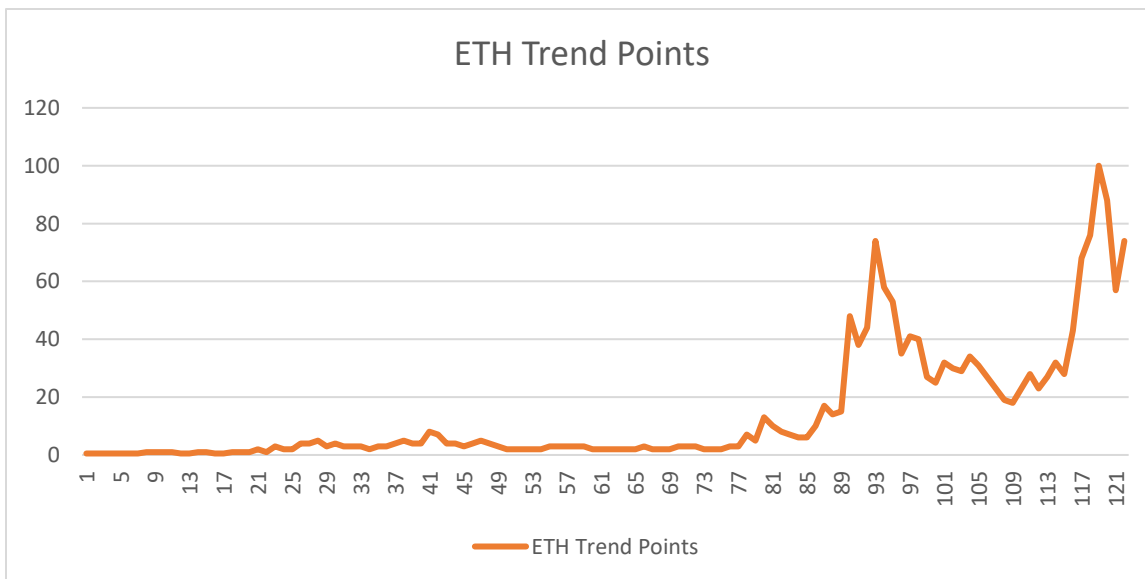


Figure 0-15: Γράφημα EthereumUSD Price από Etherscan σε Excel

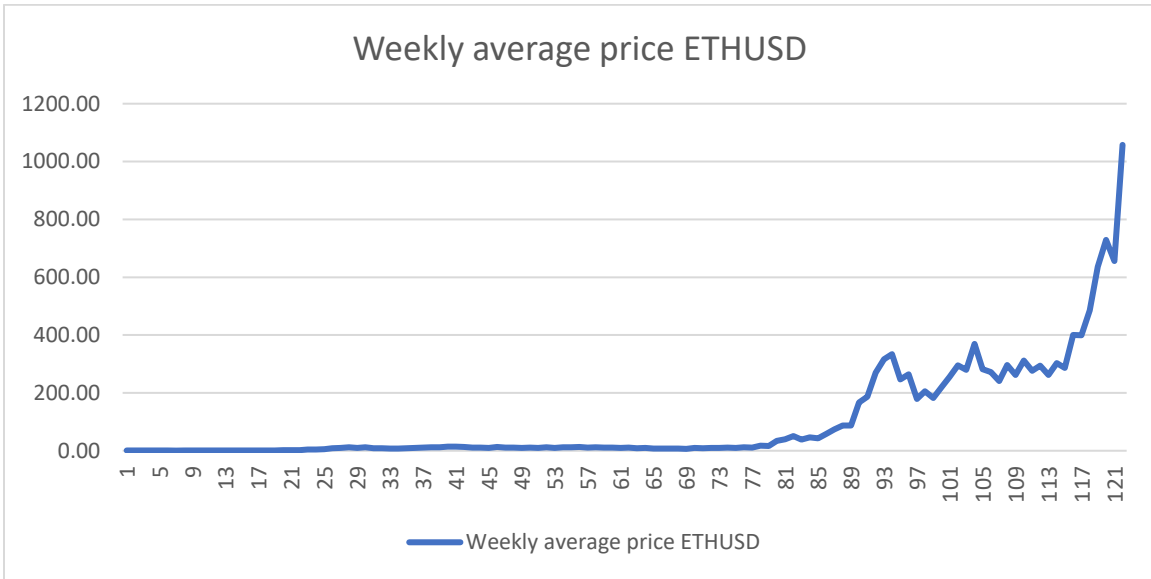
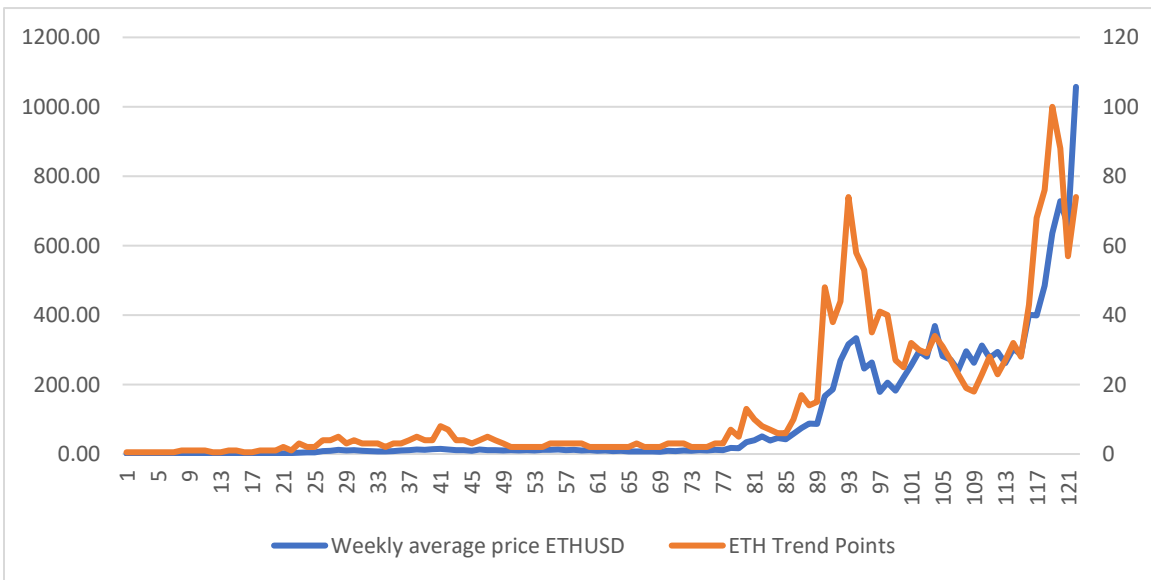


Figure 0-16: Αλληλεπικαλυπτόμενα γραφήματα Ethereum από google trends και EthereumUSD Price από Etherscan σε Excel



Στην στατιστική ανάλυση που ακολουθεί με τα ίδια ζεύγη τιμών παρατηρούμε οπτικά στο διάγραμμα διασποράς(

Figure 0-17) ότι οι μέσες εβδομαδιαίες τιμές του νομίσματος σε σχέση με τις τιμές τάσης του Google trend ομαδοποιούνται ως προς την περιοχή και παρουσιάζουν εικόνα αναλογικότητας (αυξάνουν ταυτόχρονα σε μεγάλο βαθμό και οι δύο), γεγονός που καταδεικνύει συσχέτιση. Αλλά και αριθμητικά μέσω των 3 διαφορετικών (παραμετρικών και μη) συντελεστών συσχέτισης (Pearson's, Kendall's, Spearman's- Figure 0-18), λόγω του ότι η τιμή **p-value είναι μικρότερη του επιπέδου σημαντικότητας $\alpha=0.05$** που έχουμε θέσει ως όριο και οι **αντίστοιχοι συντελεστές μεγαλύτεροι του +0.5**, επιβεβαιώνεται ότι **υπάρχει στατιστικά σημαντική συσχέτιση και μάλιστα ισχυρά θετική, μεταξύ των μεταβλητών της τάσης των αναζητήσεων και της τιμής του κρυπτονομίσματος ETH.**

Figure 0-17: SCATTERPLOT(BIVAR)=ETH_Trend_Points WITH Weekly_average_price_ETHUSD

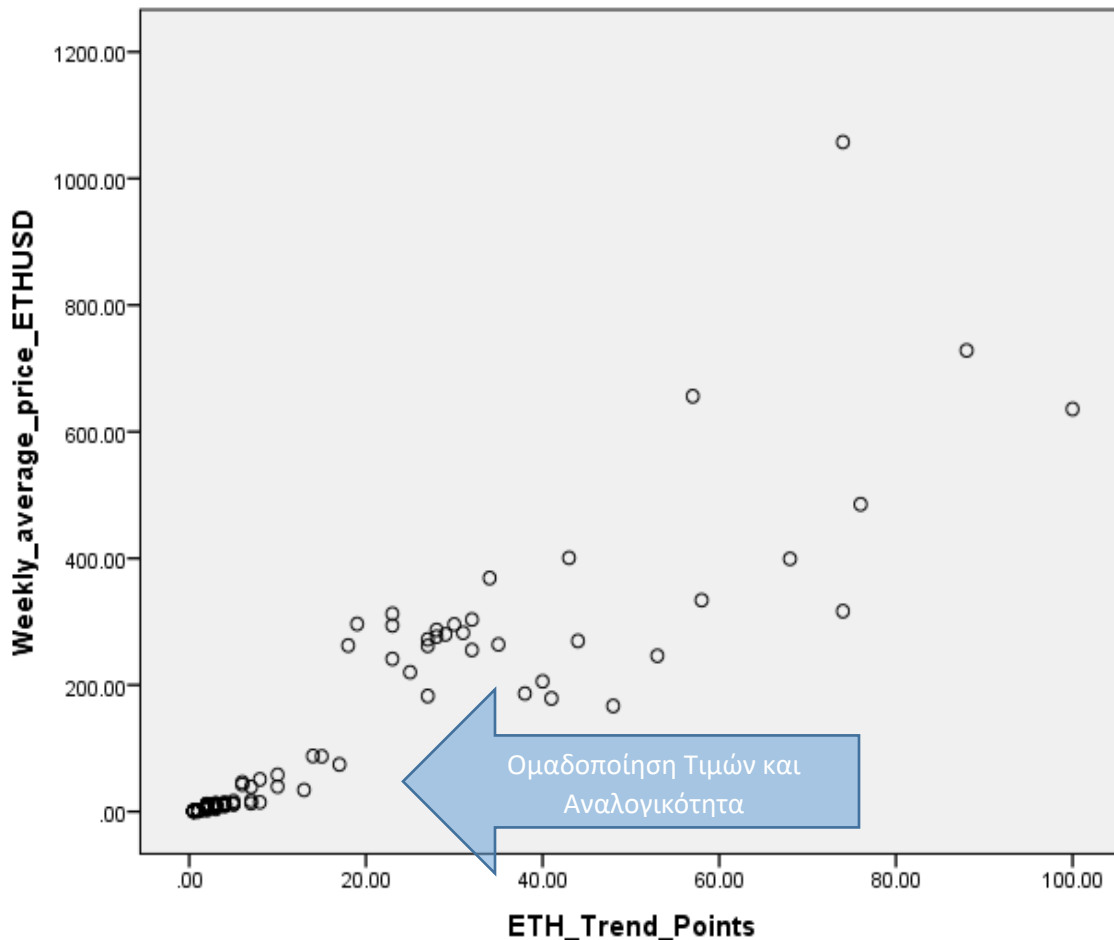


Figure 0-18: CORRELATIONS=ETH_Trend_Points Weekly_average_price_ETHUSD

Correlations			ETH_Trend_Points	Weekly_average_price_ETHUSD
ETH_Trend_Points	Pearson Correlation		1	
	Sig. (2-tailed)			.000
Weekly_average_price_ETHUSD	Pearson Correlation		.906**	1
	Sig. (2-tailed)		.000	

Non-parametric Correlations			ETH_Trend_Points	Weekly_average_price_ETHUSD
Kendall's tau_b	ETH_Trend_Points	Correlation Coefficient	1.000	
		Sig. (2-tailed)	.	.000
	Weekly_average_price_ETHUSD	Correlation Coefficient	.817**	1.000
		Sig. (2-tailed)	.000	.
Spearman's rho	ETH_Trend_Points	Correlation Coefficient	1.000	
		Sig. (2-tailed)	.	.000
	Weekly_average_price_ETHUSD	Correlation Coefficient	.941**	1.000
		Sig. (2-tailed)	.000	.

** . Correlation is significant at the 0.05 level (2-tailed).

Συμπερασματικά μπορούμε να πούμε ότι απορρίπτεται η **Πρόταση H0**: και αντίθετα ισχύει η **Πρόταση H1**: Η μέση εβδομαδιαία τιμή του κρυπτονομίσματος Ethereum σχετίζεται με την εβδομαδιαία τιμή της τάσης αναζήτησης στο εργαλείο Google Trends.

3.4.1.3 Google trends και τιμές του Ripple

Η αναζήτηση στο εργαλείο google trends έγινε με μοναδικό όρο το “Ripple” και αφορούσε την περίοδο 01/02/2015 έως 31/12/2017, μια επαρκής περίοδος εξέτασης που περιλαμβάνει τις σχετικά πρόσφατες σημαντικές εξελίξεις στην τιμή του νομίσματος αλλά και μεγάλο μέρος της επέκτασης και αποδοχής του, καθώς και την διαθεσιμότητα των στοιχείων. Οι τιμές της αξίας του σε δολάρια Αμερικής αντλήθηκαν από το Yahoo Finance (<https://finance.yahoo.com/quote/XRP-USD/history?p=XRP-USD>) και αφορούσαν το διάστημα 01/02/15 έως 06/01/18.

- Υποερώτημα 1Γ-Google trends και τιμές Ripple:: Η τιμή του κρυπτονομίσματος Ripple σχετίζεται με τις αναζητήσεις στην μηχανή αναζήτησης Google, όπως αυτές απεικονίζονται στο εργαλείο Google Trends;

Πρόταση Η0: Η μέση εβδομαδιαία τιμή του κρυπτονομίσματος Ripple δεν σχετίζεται με την εβδομαδιαία τιμή της τάσης αναζήτησης στο εργαλείο Google Trends.

Πρόταση Η1: Η μέση εβδομαδιαία τιμή του κρυπτονομίσματος Ripple σχετίζεται με την εβδομαδιαία τιμή της τάσης αναζήτησης στο εργαλείο Google Trends.

Τα δεδομένα από τα αρχικά γραφήματα τάσης-τιμών(Figure 0-19 και Figure 0-20) μετατρέπονται σε γραφήματα Excel(Figure 0-21 και Figure 0-22) και στην συνέχεια επιχειρείται μια πρώτη οπτική διερεύνηση του συσχετισμού τους με την προβολή τους σε αλληλοεπικάλυψη(Figure 0-23). Παρατηρούμε την ταύτιση των γραφικών παραστάσεων και όπως φαίνεται η τάση και οι τιμές κινούνται παράλληλα μαρτυρώντας έντονο συσχετισμό.

Figure 0-19: Γράφημα Ripple από Google trends 01/02/2015-01/01/2018

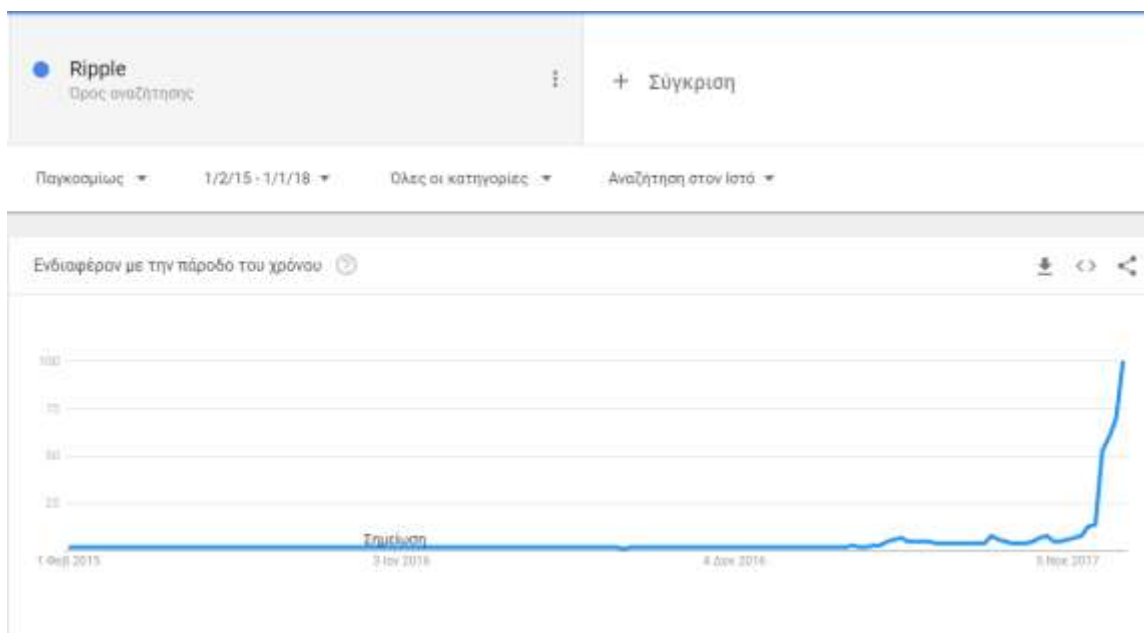


Figure 0-20: Γράφημα RippleUSD Price από Yahoo Finance



Figure 0-21: Γράφημα Ripple από Google trends σε Excel

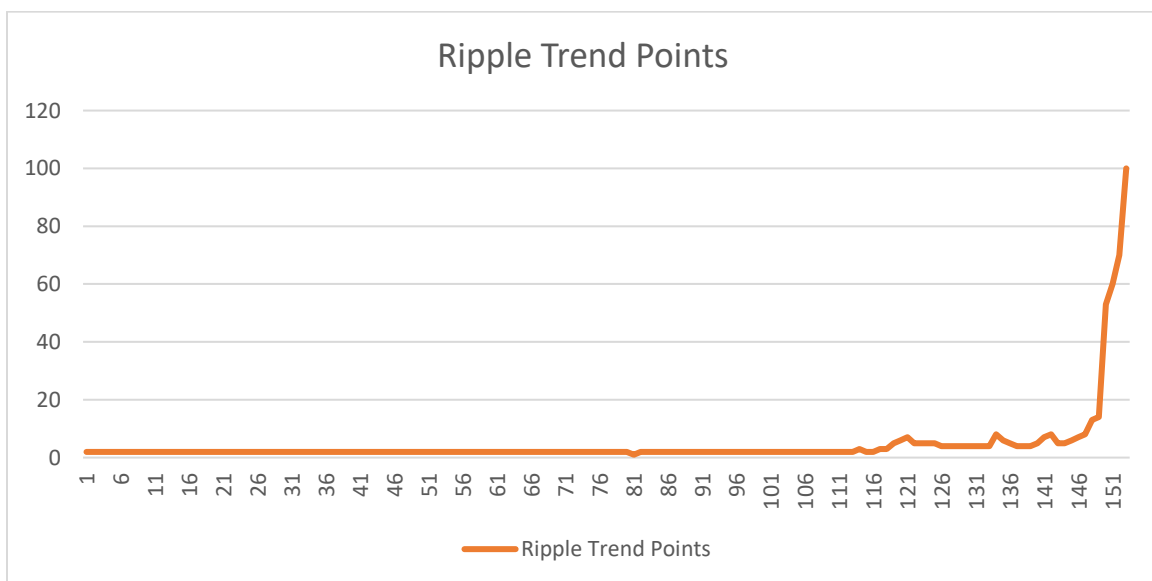


Figure 0-22: Γράφημα RippleUSD Price από Yahoo Finance σε Excel

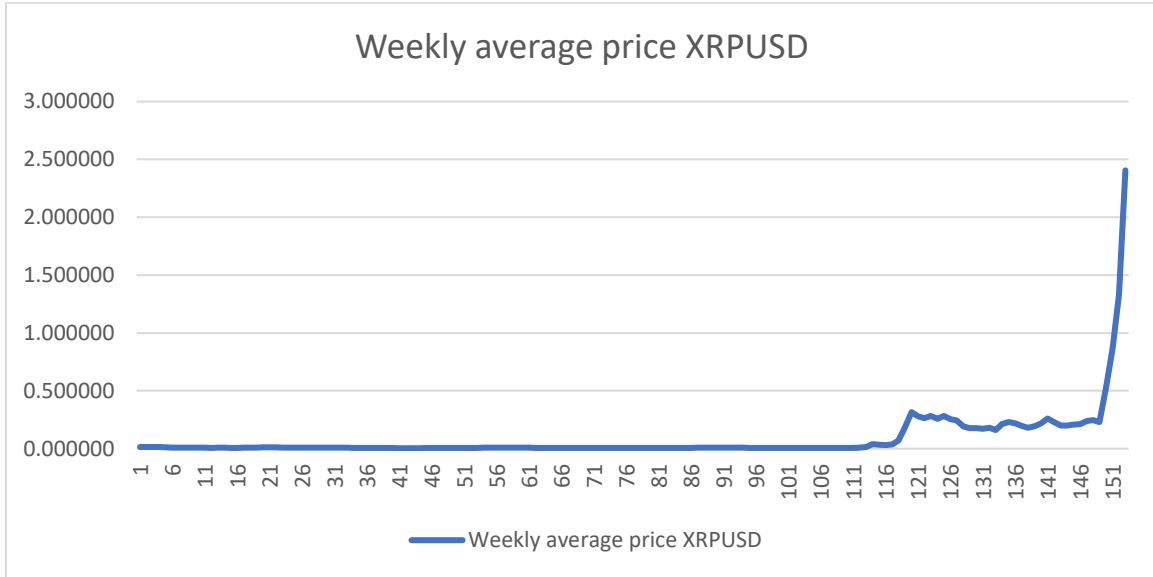
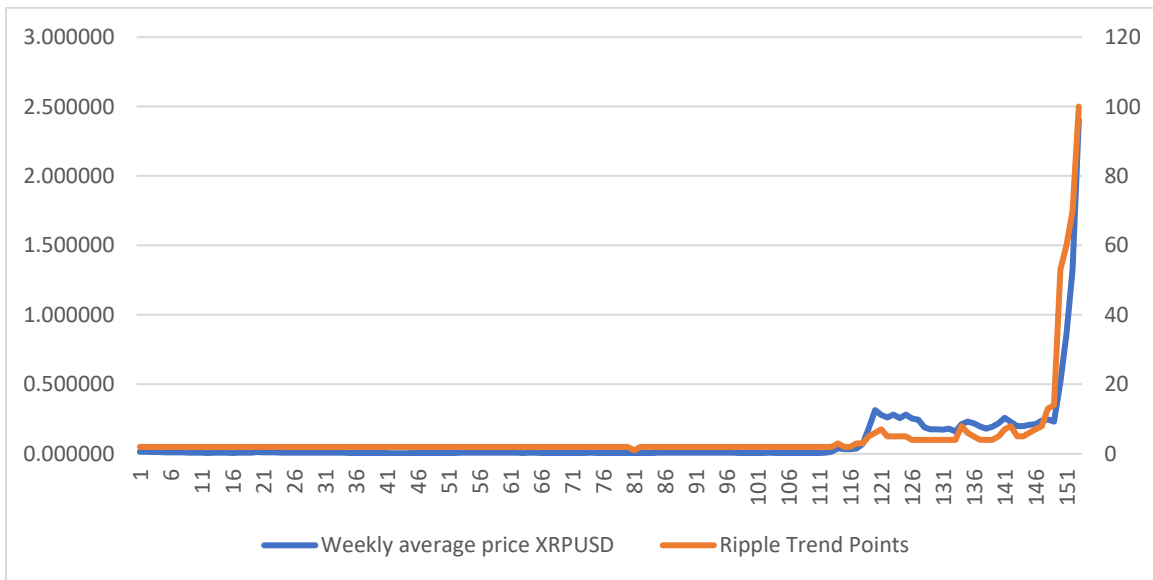


Figure 0-23: Αλληλεπικαλυπτόμενα γραφήματα Ripple από google trends και RippleUSD Price από Yahoo σε Excel



Παρατηρούμε την ταύτιση των γραφικών παραστάσεων.

Στην στατιστική ανάλυση που ακολουθεί με τα ίδια ζεύγη τιμών παρατηρούμε οπτικά στο διάγραμμα διασποράς(Figure 0-24) ότι οι μέσες εβδομαδιαίες τιμές του νομίσματος σε σχέση με τις τιμές τάσης του Google trend ομαδοποιούνται ως προς την περιοχή και παρουσιάζουν εικόνα αναλογικότητας (αυξάνουν ταυτόχρονα σε μεγάλο βαθμό και οι δύο), γεγονός που καταδεικνύει συσχέτιση. Αλλά και αριθμητικά μέσω των 3 διαφορετικών (παραμετρικών και μη) συντελεστών συσχέτισης (Pearson's, Kendall's, Spearman's-Figure 0-25), λόγω του ότι η **τιμή p-value είναι μικρότερη του επιπέδου σημαντικότητας $\alpha=0.05$** που έχουμε θέσει ως όριο και οι **αντίστοιχοι συντελεστές μεγαλύτεροι του +0.5**, επιβεβαιώνεται ότι **υπάρχει στατιστικά σημαντική συσχέτιση και μάλιστα ισχυρά θετική, μεταξύ των μεταβλητών της τάσης των αναζητήσεων και της τιμής του κρυπτονομίσματος XRP.**

Figure 0-24: SCATTERPLOT(BIVAR)=Ripple_Trend_Points WITH Weekly_average_price_XRPUSD

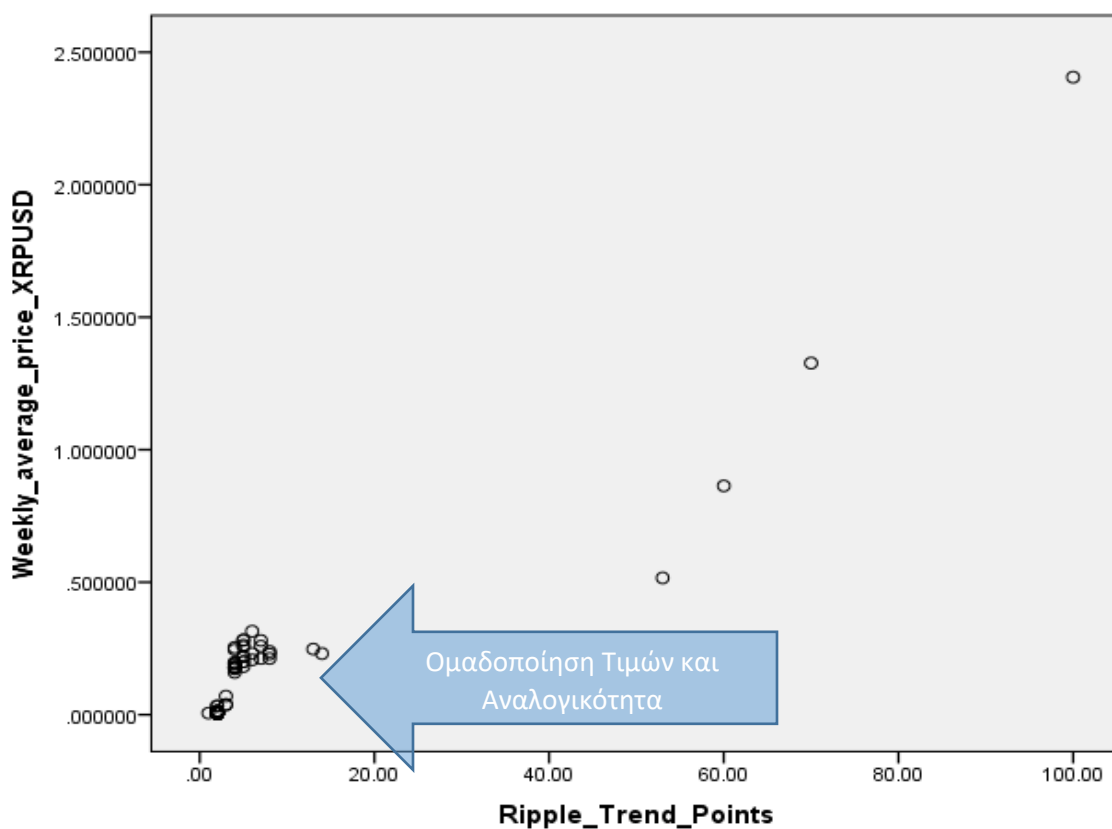
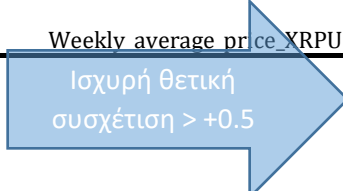
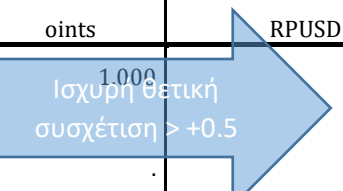
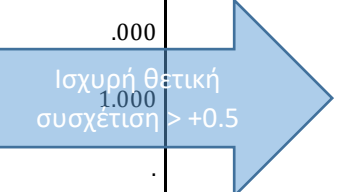


Figure 0-25: CORRELATIONS=Ripple_Trend_Points Weekly_average_price_XRPUSD

Correlations			Ripple_Trend_Points	Weekly average price_XRPUSD
Ripple_Trend_Points	Pearson Correlation		1	
	Sig. (2-tailed)			
Weekly_average_price_XRPUSD	Pearson Correlation		.935**	1
	Sig. (2-tailed)		.000	

Non-parametric Correlations			Ripple_Trend_Points	Weekly_average_price_XRPUSD
Kendall's tau_b	Ripple_Trend_Points	Correlation Coefficient	1.000	
		Sig. (2-tailed)	.000	
	Weekly_average_price_XRPUSD	Correlation Coefficient	.629**	1.000
		Sig. (2-tailed)	.000	
Spearman's rho	Ripple_Trend_Points	Correlation Coefficient	1.000	
		Sig. (2-tailed)	.000	
	Weekly_average_price_XRPUSD	Correlation Coefficient	.753**	1.000
		Sig. (2-tailed)	.000	

** . Correlation is significant at the 0.05 level (2-tailed).

Συμπερασματικά μπορούμε να πούμε ότι απορρίπτεται η **Πρόταση H0:** και αντίθετα ισχύει η **Πρόταση H1:** Η μέση εβδομαδιαία τιμή του κρυπτονομίσματος Ripple σχετίζεται με την εβδομαδιαία τιμή της τάσης αναζήτησης στο εργαλείο Google Trends.

Εν κατακλείδι, μπορούμε να πούμε ότι όσον αφορά το **1^ο Ερευνητικό Ερώτημα,** πράγματι οι τιμές των κρυπτονομισμάτων σχετίζονται με τις αναζητήσεις στην μηχανή αναζήτησης Google, όπως αυτές απεικονίζονται στο εργαλείο Google Trends.

3.4.2 Ανάλυση τάσης-γνώμης στο Twitter και τιμές νομισμάτων

Η δεύτερη διερεύνηση αφορά τον συσχετισμό του γενικότερου συναισθήματος—τάσης-γνώμης(sentiment) στις αναρτήσεις χρηστών με συγκεκριμένο θέμα, μέσω του κοινωνικού δικτύου Twitter και τις αντίστοιχες διακυμάνσεις των τιμών των 3 κρυπτονομισμάτων όπως απεικονίζονται σε γνωστό ανταλλακτήριο κρυπτονομισμάτων (<http://www.cryptodatadownload.com/>, Kraken). Πιο συγκεκριμένα με όρους θέματος Bitcoin, Ethereum και Ripple συγκεντρώθηκε πληθώρα σχετικών μηνυμάτων(tweets) για συγκεκριμένα χρονικά διαστήματα όπως ορίζονται παρακάτω. Υπενθυμίζεται το:

2ο Ερευνητικό ερώτημα-Twitter και τιμές νομισμάτων: Οι τιμές των κρυπτονομισμάτων σχετίζονται με τις στάσεις-γνώμες στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter;

3.4.2.1 Ανάλυση τάσης-γνώμης στο Twitter και τιμές του Bitcoin

Η αναζήτηση στο εργαλείο Chorus Tweetcatcher Desktop Edition έγινε με τους όρους “Bitcoin”, “bitcoin”, “BTC”, “btc” και αφορούσε την περίοδο 10/07/2018 έως 17/07/2018, μια περίοδος εξέτασης που απέδωσε αρχικά 100073 tweets, τα οποία μειώθηκαν σε 46116 μετά την αφαίρεση των retweets. Από αυτά επιλέχθηκαν εκείνα αφορούσαν το πρώτο λεπτό κάθε ολόκληρης ώρας, με συνολικό αριθμό 4550. Ακολούθησε ανάλυση στάσης για 4550 μηνύματα με το εργαλείο sentiment analysis της Cloud Mining για το RapidMiner. Στην συνέχεια έγιναν αποδεκτά με τα κριτήρια που προαναφέρθηκαν 804 μηνύματα, τα οποία ομαδοποιήθηκαν και αντιστοιχήθηκαν με τις ωριαίες αξίες του Bitcoin. Τέλος τα ζεύγη τάσης-γνώμης και αξίας νομίσματος αναλύθηκαν μέσω του SPSS για πιθανή συσχέτιση.

- **Υποερώτημα 2A-Twitter και τιμές Bitcoin:** Η τιμή του κρυπτονομίσματος Bitcoin σχετίζεται με τις στάσεις-γνώμες στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter;

Πρόταση H0: Η ωριαία τιμή του κρυπτονομίσματος Bitcoin δεν σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter για την αντίστοιχη ώρα.

Πρόταση H1: Η ωριαία τιμή του κρυπτονομίσματος Bitcoin σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter για την αντίστοιχη ώρα.

Αρχικά λόγω της επιλογής δεδομένων από το πρώτο λεπτό κάθε ώρας, ως καθοριστικού για τον χαρακτηρισμό της υπόλοιπης ώρας ως θετικής ή αρνητικής ως προς την αξία του νομίσματος γίνεται έλεγχος για κανονικότητα των τιμών(Figure 0-26, Figure 0-27, Figure 0-28). Οπτικά από το ιστόγραμμα παρατηρούμε ότι δεν έχει

το γνωστό σχήμα καμπάνας της κανονικής κατανομής, οι τιμές στο διάγραμμα Q-Q δεν είναι κοντά στην διαγώνιο και τα τεστ κανονικότητας δείχνουν **p-value μικρότερη του επιπέδου σημαντικότητας 0.05**, που σημαίνει ότι η αρχική υπόθεση(null hypothesis) για κανονική κατανομή απορρίπτεται. Στην συνέχεια λοιπόν εξομαλύνονται οι τιμές(TR_PRICES) χρησιμοποιώντας τον δεκαδικό λογάριθμό τους, αλλά και πάλι η κατανομή δεν πλησιάζει την κανονική(Figure 0-29, Figure 0-30).

Figure 0-26: EXAMINE VARIABLES=BTC_HOURLY_PRICES NORMALITY TEST

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
BTC_HOURLY_PRICES	.246	167	.000	.822	167	.000

Figure 0-27: BTC_HOURLY_PRICES HISTOGRAM BEFORE NORMALIZATION

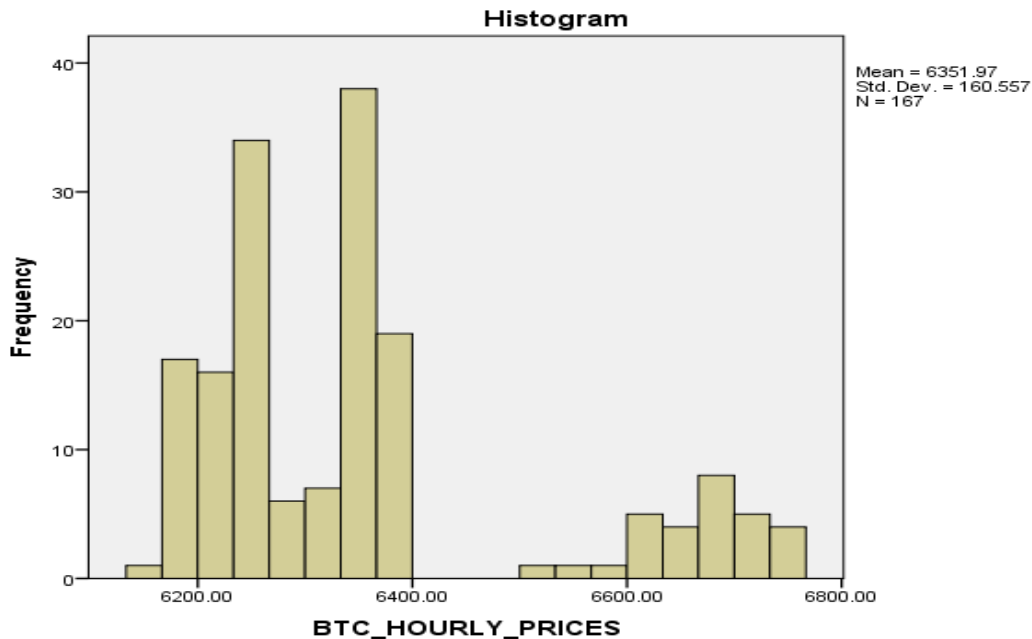


Figure 0-28: Normal Q-Q Plot of BTC_HOURLY_PRICES

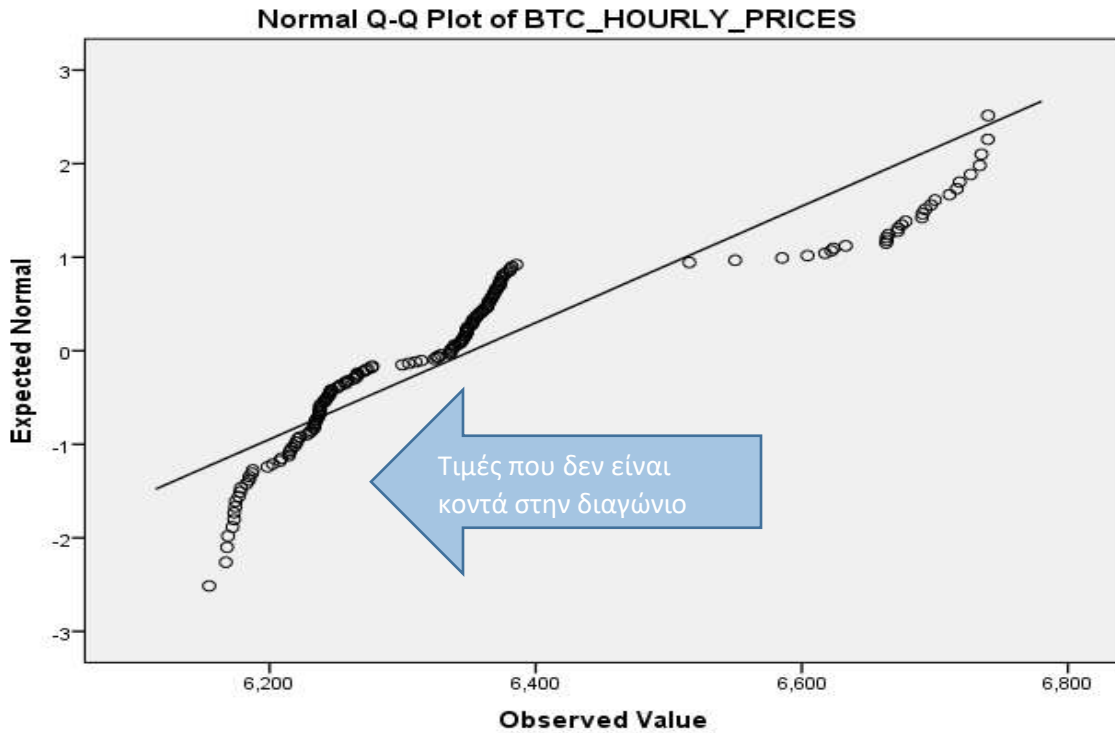


Figure 0-29: TR_BTC_HOURLY_PRICES=LG10(BTC_HOURLY_PRICES) HISTOGRAM

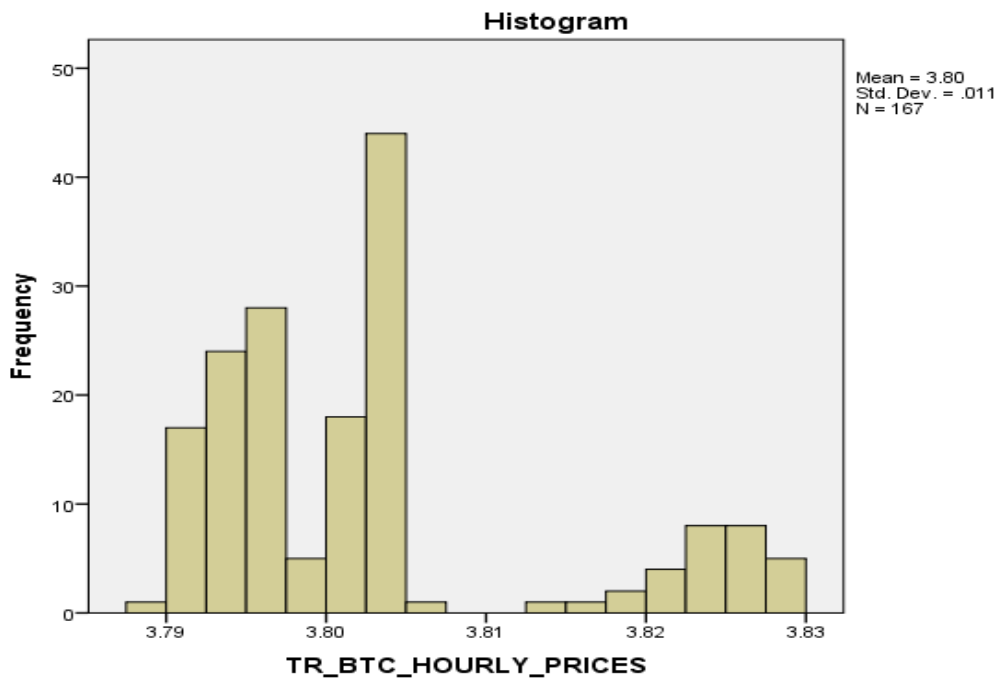
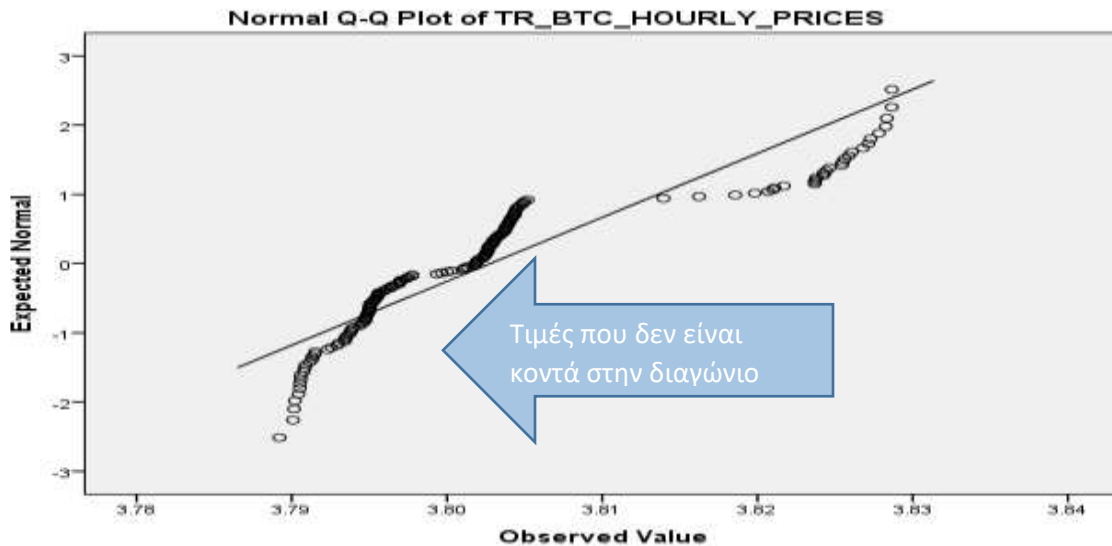


Figure 0-30: Normal Q-Q Plot of TR_BTC_HOURLY_PRICES



Στην στατιστική ανάλυση (Biserial Analysis για συσχέτιση μιας δυαδικής και μια συνεχούς ποσοτικής μεταβλητής), που ακολουθεί παρατηρούμε μέσω του συντελεστή συσχέτισης (Pearson's-Figure 0-31), λόγω του ότι η **τιμή p-value είναι μικρότερη του επιπέδου σημαντικότητας $\alpha=0.05$** που έχουμε θέσει ως όριο και **ο αντίστοιχος συντελεστής συσχέτισης μεγαλύτερος του +0.1, αλλά μικρότερος του +0.3**, πως επιβεβαιώνεται ότι υπάρχει στατιστικά σημαντική συσχέτιση και μάλιστα ασθενής θετική μεταξύ των μεταβλητών της στάσης-γνώμης και της τιμής του κρυπτονομίσματος BTC.

Figure 0-31: CORRELATIONS= POSITIVE_NEGATIVE_VALUE TR_BTC_HOURLY_PRICES

		POSITIVE_NEGATIVE_VALUE	TR_BTC_HOURLY_PRICES
POSITIVE_NEGATIVE_VALUE	Pearson Correlation		.159*
	Sig. (2-tailed)		.040
TR_BTC_HOURLY_PRICES	Pearson Correlation	.159*	1
	Sig. (2-tailed)	.040	

*. Correlation is significant at the 0.05 level (2-tailed).

Συμπερασματικά μπορούμε να πούμε ότι απορρίπτεται η **Πρόταση H0**: και αντίθετα ισχύει η **Πρόταση H1**: Η ωριαία τιμή του κρυπτονομίσματος Bitcoin σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter για την αντίστοιχη ώρα.

3.4.2.2 Ανάλυση τάσης-γνώμης στο Twitter και τιμές του Ethereum

Η αναζήτηση στο εργαλείο Chorus Tweetcatcher Desktop Edition έγινε με τους όρους “Ethereum”, “ethereum”, “ETH”, “eth” και αφορούσε την περίοδο 07/07/2018 έως 17/07/2018, μια περίοδος εξέτασης που απέδωσε αρχικά 71808 tweets, τα οποία μειώθηκαν σε 21209, μετά την αφαίρεση των retweets. Από αυτά επιλέχθηκαν εκείνα αφορούσαν το πρώτο λεπτό κάθε ολόκληρης ώρας, με συνολικό αριθμό 1921. Ακολούθησε ανάλυση στάσης για 1921 μηνύματα με το εργαλείο sentiment analysis της Cloud Mining για το RapidMiner. Στην συνέχεια έγιναν αποδεκτά με τα κριτήρια που προαναφέρθηκαν 397 μηνύματα, τα οποία ομαδοποιήθηκαν και αντιστοιχήθηκαν με τις ωριαίες αξίες του Ethereum. Τέλος τα ζεύγη τάσης-γνώμης και αξίας νομίσματος αναλύθηκαν μέσω του SPSS για πιθανή συσχέτιση.

- **Υποερώτημα 2B-Twitter και τιμές Ethereum:** Η τιμή του κρυπτονομίσματος Ethereum σχετίζεται με τις στάσεις-γνώμες στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter;

Πρόταση H0: Η ωριαία τιμή του κρυπτονομίσματος Ethereum δεν σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter για την αντίστοιχη ώρα.

Πρόταση H1: Η ωριαία τιμή του κρυπτονομίσματος Ethereum σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter για την αντίστοιχη ώρα.

Αρχικά λόγω της επιλογής δεδομένων από το πρώτο λεπτό κάθε ώρας ως καθοριστικού για τον χαρακτηρισμό της υπόλοιπης ώρας ως θετικής ή αρνητικής ως προς την αξία του νομίσματος γίνεται έλεγχος για κανονικότητα των τιμών (Figure 0-32, Figure 0-33, Figure 0-34). Οπτικά από το ιστόγραμμα παρατηρούμε ότι δεν έχει το γνωστό σχήμα καμπάνας της κανονικής κατανομής, οι τιμές στο διάγραμμα Q-Q δεν είναι κοντά στην διαγώνιο και τα τεστ κανονικότητας δείχνουν **p-value μικρότερη του επιπέδου σημαντικότητας 0.05**, που σημαίνει ότι η αρχική υπόθεση (null hypothesis) για κανονική κατανομή απορρίπτεται. Στην συνέχεια λοιπόν εξομαλύνονται οι τιμές (TR_PRICES) χρησιμοποιώντας τον δεκαδικό λογάριθμό τους, αλλά και πάλι η κατανομή δεν πλησιάζει την κανονική (Figure 0-35, Figure 0-36).

Figure 0-32: EXAMINE VARIABLES=ETH_HOURLY_PRICES NORMALITY TEST

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ETH_HOURLY_PRICES	.193	238	.000	.861	238	.000

Figure 0-33: ETH_HOURLY_PRICES BEFORE NORMALIZATION

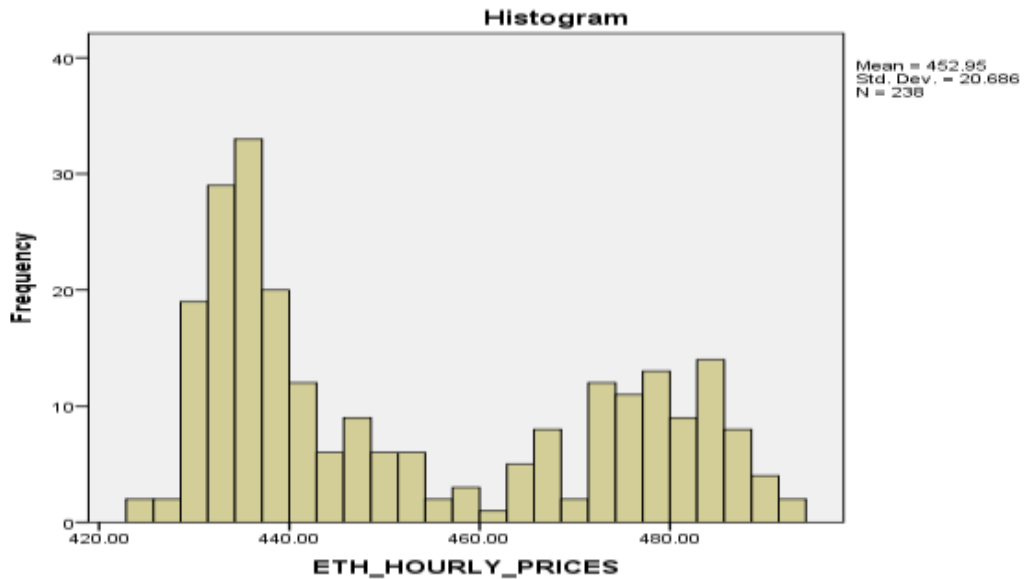


Figure 0-34: Normal Q-Q Plot of ETH_HOURLY_PRICES

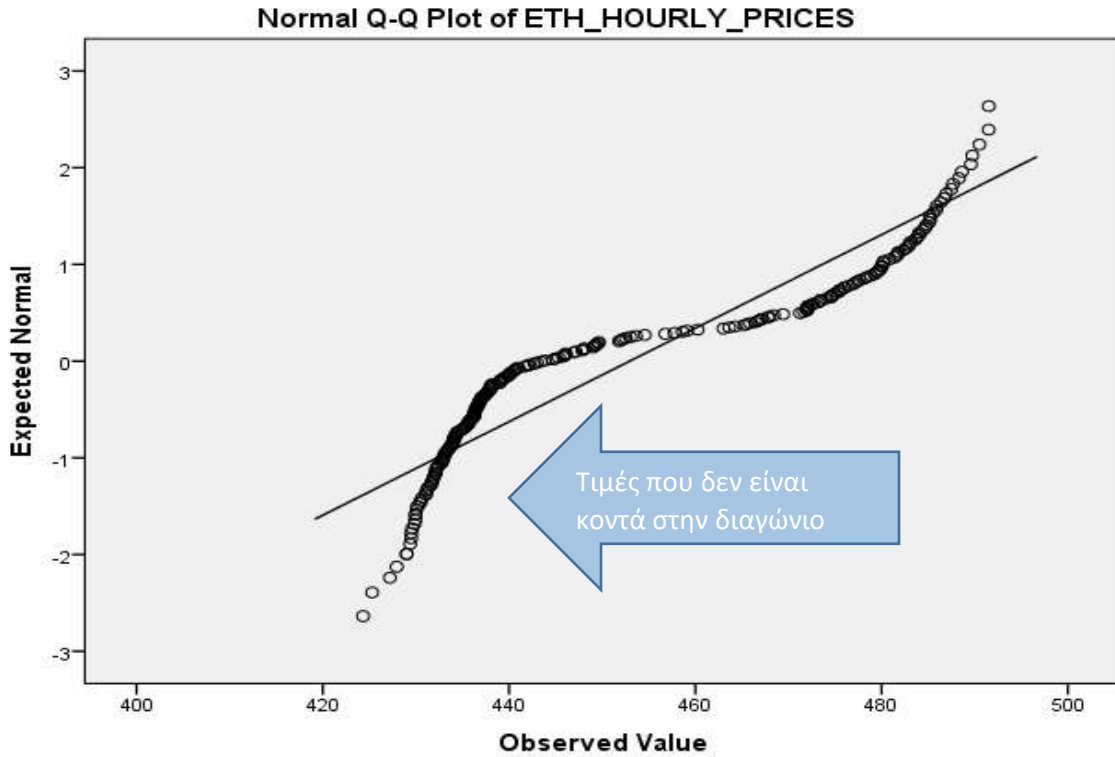


Figure 0-35: TR_ETH_HOURLY_PRICES=LG10(ETH_HOURLY_PRICES) HISTOGRAM

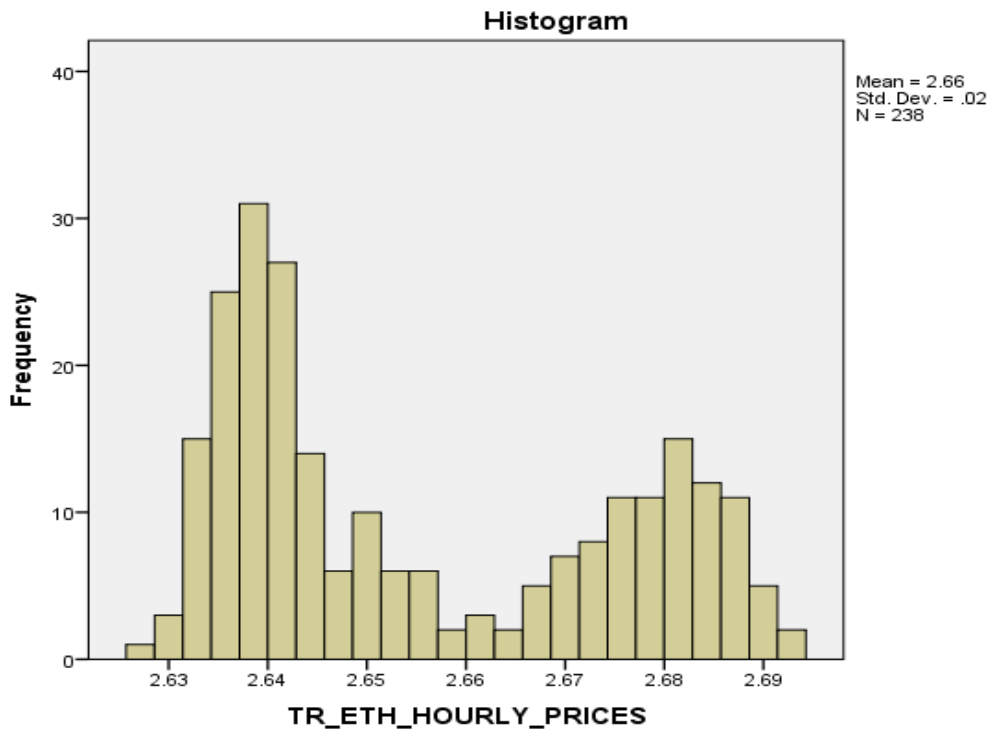
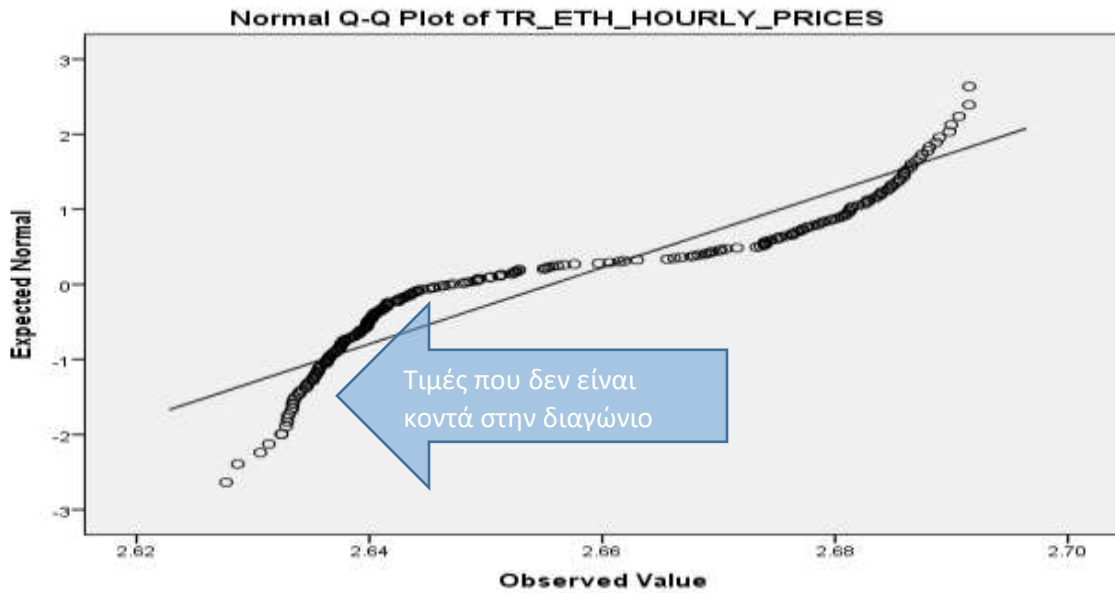


Figure 0-36: Normal Q-Q Plot of TR_ETH_HOYRLY_PRICES



Στην στατιστική ανάλυση (Biserial Analysis για συσχέτισμό μιας δυαδικής και μια συνεχούς ποσοτικής μεταβλητής) που ακολουθεί παρατηρούμε μέσω του συντελεστή συσχέτισης (Pearson's-Figure 0-37), λόγω του ότι η **τιμή p-value είναι μεγαλύτερη του επιπέδου σημαντικότητας $\alpha=0.05$** που έχουμε θέσει ως όριο, πως επιβεβαιώνεται ότι **δεν υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ των μεταβλητών της στάσης-γνώμης και της τιμής του κρυπτονομίσματος ETH.**

Figure 0-37=CORRELATIONS: POSITIVE_NEGATIVE_VALUE TR_ETH_HOURLY_PRICES

		POSITIVE_NEGATIVE_VALUE	TR_ETH_HOURLY_PRICES
POSITIVE_NEGATIVE_VALUE	Pearson Correlation	1	-.048
	Sig. (2-tailed)		.461
TR_ETH_HOURLY_PRICES	Pearson Correlation	-.048	1
	Sig. (2-tailed)	.461	

Καμία συσχέτιση
 $p > 0.05$

Συμπερασματικά μπορούμε να πούμε ότι απορρίπτεται η **Πρόταση H1:** και αντίθετα ισχύει η **Πρόταση H0:** Η ωριαία τιμή του κρυπτονομίσματος Ethereum δεν σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter για την αντίστοιχη ώρα.

3.4.2.3 Ανάλυση τάσης-γνώμης στο Twitter και τιμές του Ripple

Η αναζήτηση στο εργαλείο Chorus Tweetcatcher Desktop Edition έγινε με τους όρους “Ripple”, “ripple”, “XRP”, “xrp” και αφορούσε την περίοδο 07/07/2018 έως 17/07/2018, μια περίοδος εξέτασης που απέδωσε αρχικά 28486 tweets, τα οποία μειώθηκαν σε 12025 μετά την αφαίρεση των retweets. Από αυτά επιλέχθηκαν εκείνα που αφορούσαν το πρώτο λεπτό κάθε ολόκληρης ώρας, με συνολικό αριθμό 418. Ακολούθησε ανάλυση στάσης για 418 μηνύματα με το εργαλείο sentiment analysis της Cloud Mining για το RapidMiner. Στην συνέχεια έγιναν αποδεκτά με τα κριτήρια που προαναφέρθηκαν 143 μηνύματα, τα οποία ομαδοποιήθηκαν και αντιστοιχήθηκαν με τις ωριαίες αξίες του Ripple. Τέλος τα ζεύγη τάσης-γνώμης και αξίας νομίσματος αναλύθηκαν μέσω του SPSS για πιθανή συσχέτιση.

- **Υποερώτημα 2Γ-Twitter και τιμές Ripple:** Η τιμή του κρυπτονομίσματος Ripple σχετίζεται με τις στάσεις-γνώμες στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter;

Πρόταση H0: Η ωριαία τιμή του κρυπτονομίσματος Ripple δεν σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter για την αντίστοιχη ώρα.

Πρόταση H1: Η ωριαία τιμή του κρυπτονομίσματος Ripple σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter για την αντίστοιχη ώρα.

Αρχικά λόγω της επιλογής δεδομένων από το πρώτο λεπτό κάθε ώρας ως καθοριστικού για τον χαρακτηρισμό της υπόλοιπης ώρας ως θετικής ή αρνητικής ως προς την αξία του νομίσματος, γίνεται έλεγχος για κανονικότητα των τιμών (Figure 0-38, Figure 0-39, Figure 0-40). Οπτικά από το ιστόγραμμα παρατηρούμε ότι δεν έχει το γνωστό σχήμα καμπάνας της κανονικής κατανομής, οι τιμές στο διάγραμμα Q-Q δεν είναι κοντά στην διαγώνιο και τα τεστ κανονικότητας δείχνουν **p-value μικρότερη του επιπέδου σημαντικότητας 0.05**, που σημαίνει ότι η αρχική υπόθεση (null hypothesis) για κανονική κατανομή απορρίπτεται. Στην συνέχεια λοιπόν εξομαλύνονται οι τιμές (TR_PRICES) χρησιμοποιώντας τον δεκαδικό λογάριθμό τους, αλλά και πάλι η κατανομή δεν πλησιάζει την κανονική (Figure 0-41, Figure 0-42).

Figure 0-38: EXAMINE VARIABLES=XRP_HOURLY_PRICES NORMALITY TEST

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
XRP_HOURLY_PRICES	.217	93	.000	.853	93	.000

Figure 0-39: XRP_HOURLY_PRICES BEFORE NORMALIZATION

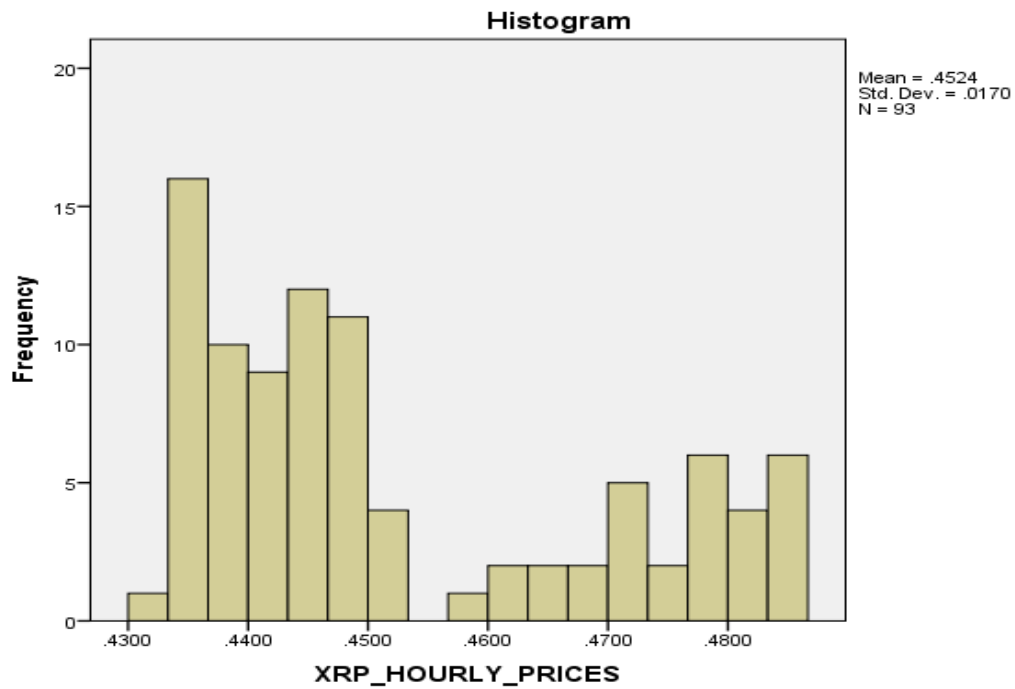


Figure 0-40: Normal Q-Q Plot of XRP_HOURLY_PRICES

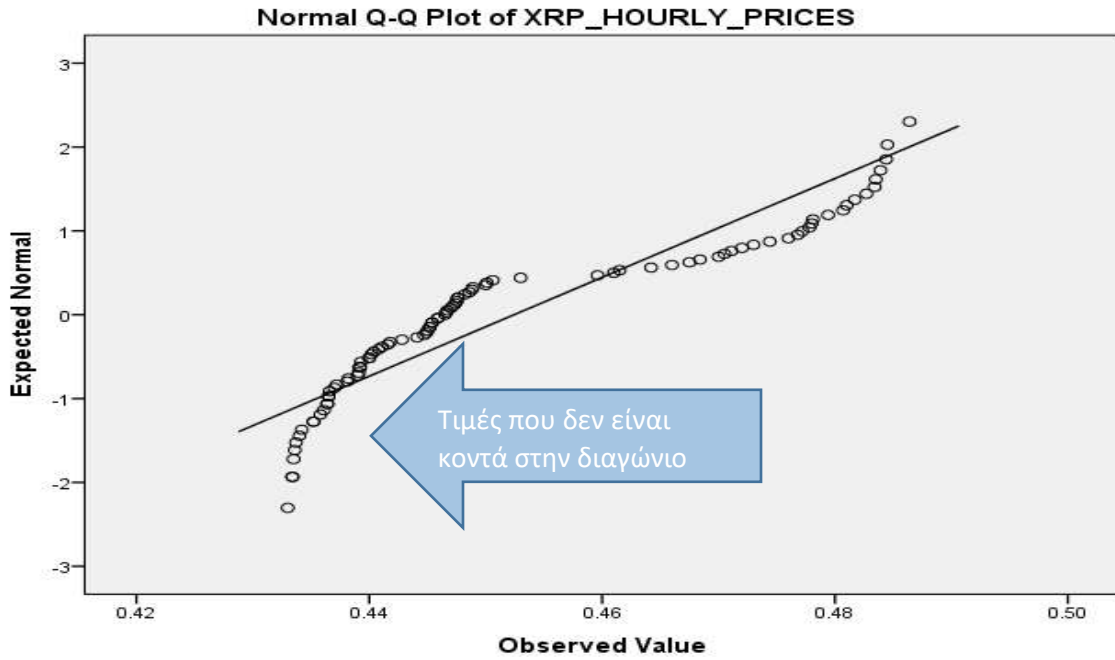


Figure 0-41: TR_XRP_HOURLY_PRICES=LG10(XRP_HOURLY_PRICES) HISTOGRAM

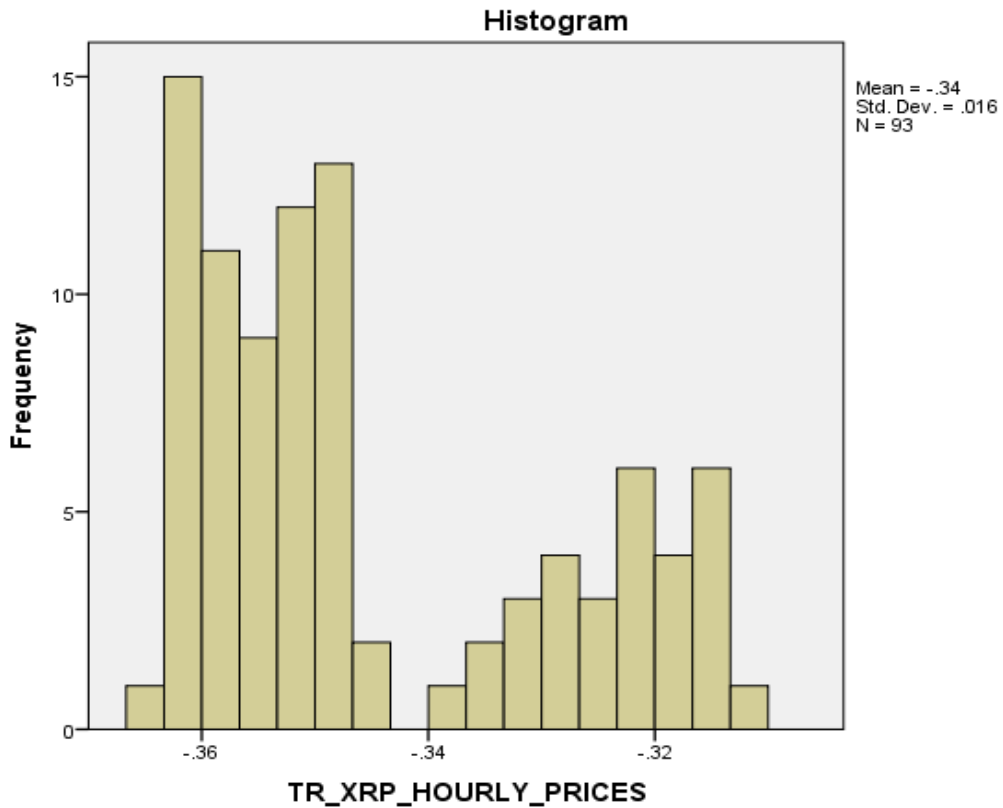
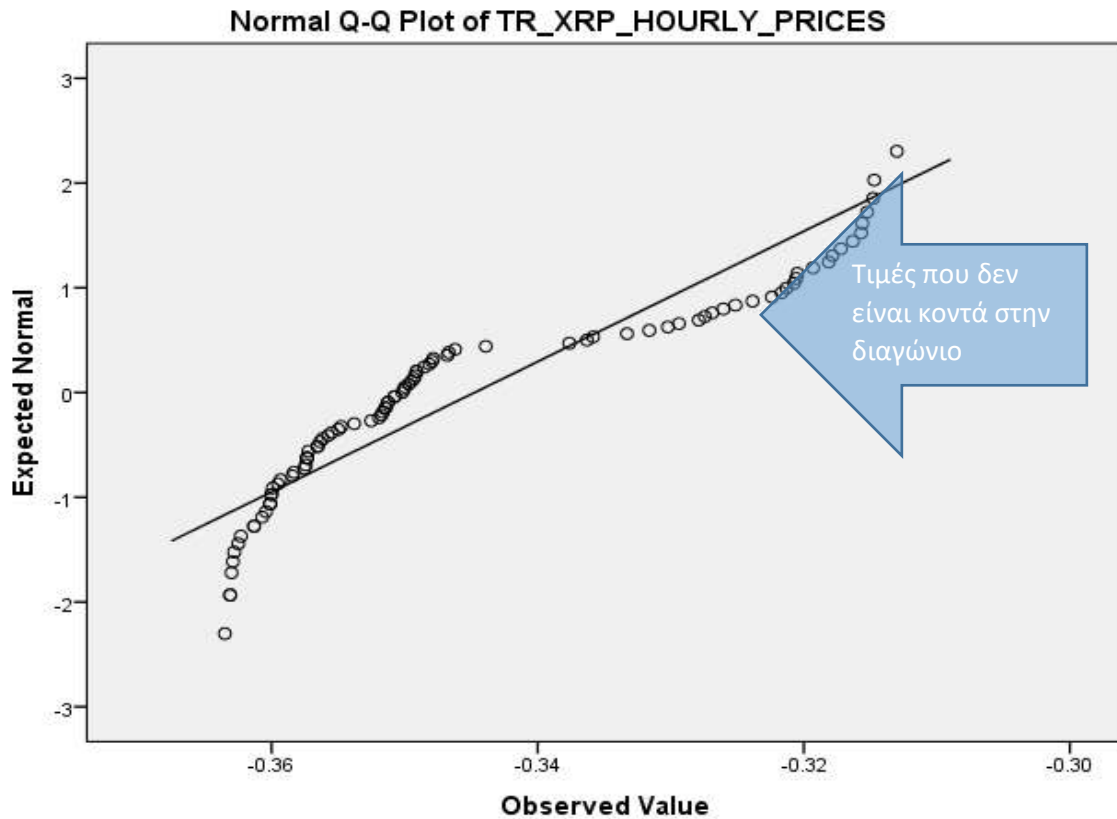


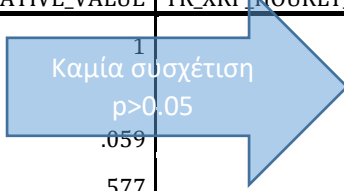
Figure 0-42: Normal Q-Q Plot of TR_XRP_HOURLY_PRICES



Στην στατιστική ανάλυση (Biserial Analysis για συσχέτισμό μιας δυαδικής και μια συνεχούς ποσοτικής μεταβλητής) που ακολουθεί, παρατηρούμε μέσω του συντελεστή συσχέτισης (Pearson's-Figure 0-43), λόγω του ότι η **τιμή p-value είναι μεγαλύτερη του επιπέδου σημαντικότητας $\alpha=0.05$** που έχουμε θέσει ως όριο, πως επιβεβαιώνεται ότι **δεν υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ των μεταβλητών της στάσης-γνώμης και της τιμής του κρυπτονομίσματος XRP.**

Figure 0-43: CORRELATIONS=POSITIVE_NEGATIVE_VALUE TR_XRP_HOURLY_PRICES

		POSITIVE_NEGATIVE_VALUE	TR_XRP_HOURLY_PRICES
POSITIVE_NEGATIVE_VALUE	Pearson Correlation	1	.059
	Sig. (2-tailed)		.577
TR_XRP_HOURLY_PRICES	Pearson Correlation	.059	1
	Sig. (2-tailed)	.577	



Συμπερασματικά μπορούμε να πούμε ότι απορρίπτεται η **Πρόταση H1**: και αντίθετα ισχύει η **Πρόταση H0**: Η ωριαία τιμή του κρυπτονομίσματος Ripple δεν σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter για την αντίστοιχη ώρα.

Εν κατακλείδι, μπορούμε να πούμε ότι όσον αφορά το 2^ο Ερευνητικό Ερώτημα, μόνο η τιμή του Bitcoin σχετίζεται με τις στάσεις-γνώμες στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter.

3.4.3 Ανάλυση τάσης-γνώμης σε Forums και τιμές νομισμάτων

Η τρίτη διερεύνηση αφορά τον συσχετισμό του γενικότερου συναισθήματος—τάσης-γνώμης(sentiment) στις αναρτήσεις χρηστών με συγκεκριμένο θέμα(τα 3 ψηφιακά νομίσματα), σε 3 διαφορετικά forums (<https://bitcointalk.org/>, <https://www.reddit.com/r/ethereum/>, <https://www.xrpchat.com/>) και τις αντίστοιχες διακυμάνσεις των τιμών των 3 κρυπτονομισμάτων όπως απεικονίζονται σε γνωστό ανταλλακτήριο κρυπτονομισμάτων (<http://www.cryptodatadownload.com/>, Kraken). Πιο συγκεκριμένα με όρους θέματος Bitcoin, Ethereum και Ripple και τις αντίστοιχες συντομογραφίες τους συγκεντρώθηκε πληθώρα σχετικών αναρτήσεων-posts για συγκεκριμένα χρονικά διαστήματα όπως ορίζονται παρακάτω. Υπενθυμίζεται το:

3ο Ερευνητικό ερώτημα-Forums και τιμές νομισμάτων: Οι τιμές των κρυπτονομισμάτων σχετίζονται με τις στάσεις-γνώμες στις αναρτήσεις μηνυμάτων στα αντίστοιχα forum bitcointalk, reddit και xrpchat;

3.4.3.1 Ανάλυση τάσης-γνώμης στο Bitcointalk και τιμές του Bitcoin

Η αναζήτηση στο εργαλείο <https://webhose.io/> έγινε με τους όρους “Bitcoin”, “bitcoin”, “BTC”, “btc” και αφορούσε την περίοδο 14/07/2018 έως και 28/07/2018, μια περίοδος εξέτασης που απέδωσε αρχικά 1400 posts, τα οποία μειώθηκαν σε 1289 μετά την αφαίρεση των διπλότυπων. Ακολούθησε ανάλυση στάσης για 1289 αναρτήσεις με το εργαλείο sentiment analysis της Cloud Mining για το RapidMiner. Στην συνέχεια έγιναν αποδεκτά με τα κριτήρια που προαναφέρθηκαν 825 μηνύματα, τα οποία ομαδοποιήθηκαν και αντιστοιχήθηκαν με τις ημερήσιες αξίες του Bitcoin. Τέλος τα ζεύγη τάσης-γνώμης και αξίας νομίσματος αναλύθηκαν μέσω του SPPS για πιθανή συσχέτιση.

- **Υποερώτημα 3Α-Bitcointalk και τιμές Bitcoin:** Η τιμή του κρυπτονομίσματος Bitcoin σχετίζεται με τις στάσεις-γνώμες στις αναρτήσεις μηνυμάτων στο forum bitcointalk;

Πρόταση H0: Η ημερήσια τιμή του κρυπτονομίσματος Bitcoin δεν σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο forum bitcointalk για την αντίστοιχη ημέρα.

Πρόταση H1: Η ημερήσια τιμή του κρυπτονομίσματος Bitcoin σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο forum bitcointalk για την αντίστοιχη ημέρα.

Λόγω του μικρού αριθμού ημερών που χρησιμοποιούνται για τον χαρακτηρισμό ως θετικών ή αρνητικών(αν και ο χαρακτηρισμός προκύπτει ως μέσος όρος πλήθους παρατηρήσεων-σχολίων) ως προς την αξία του νομίσματος, γίνεται έλεγχος για κανονικότητα των τιμών(Figure 0-44, Figure 0-45, Figure 0-46). Οπτικά από το

ιστόγραμμα παρατηρούμε ότι πλησιάζει το σχήμα καμπάνας της κανονικής κατανομής, οι τιμές στο διάγραμμα Q-Q, είναι αρκετά κοντά στην διαγώνιο και τα τεστ κανονικότητας δείχνουν **p-value μεγαλύτερη του επιπέδου σημαντικότητας 0.05**, που σημαίνει ότι η αρχική υπόθεση(null hypothesis) για κανονική κατανομή μπορεί οριακά να υιοθετηθεί. Στην συνέχεια λοιπόν εξομαλύνονται οι τιμές(TR_PRICES) χρησιμοποιώντας τον δεκαδικό λογάριθμό τους, ώστε η κατανομή να πλησιάζει ακόμα περισσότερο την κανονική (Figure 0-47, Figure 0-48).

Figure 0-44: EXAMINE VARIABLES=BTC_DAILY_PRICES NORMALITY TEST

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
BTC_DAILY_PRICES	.202	15	.101	.923	15	.218

Figure 0-45: BTC_DAILY_PRICES BEFORE NORMALIZATION

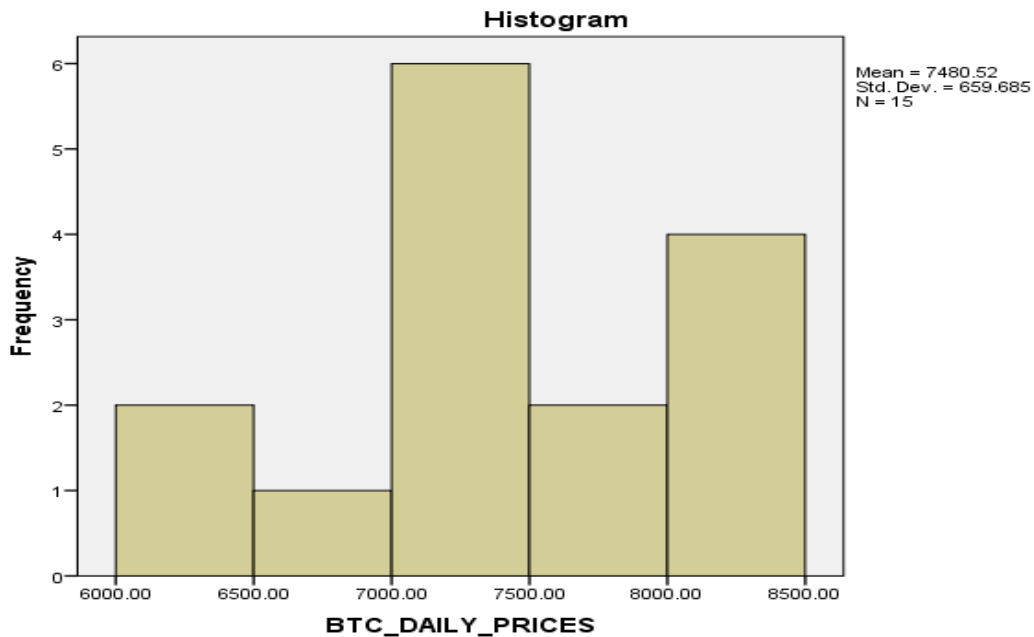


Figure 0-46: Normal Q-Q Plot of BTC_DAILY_PRICES

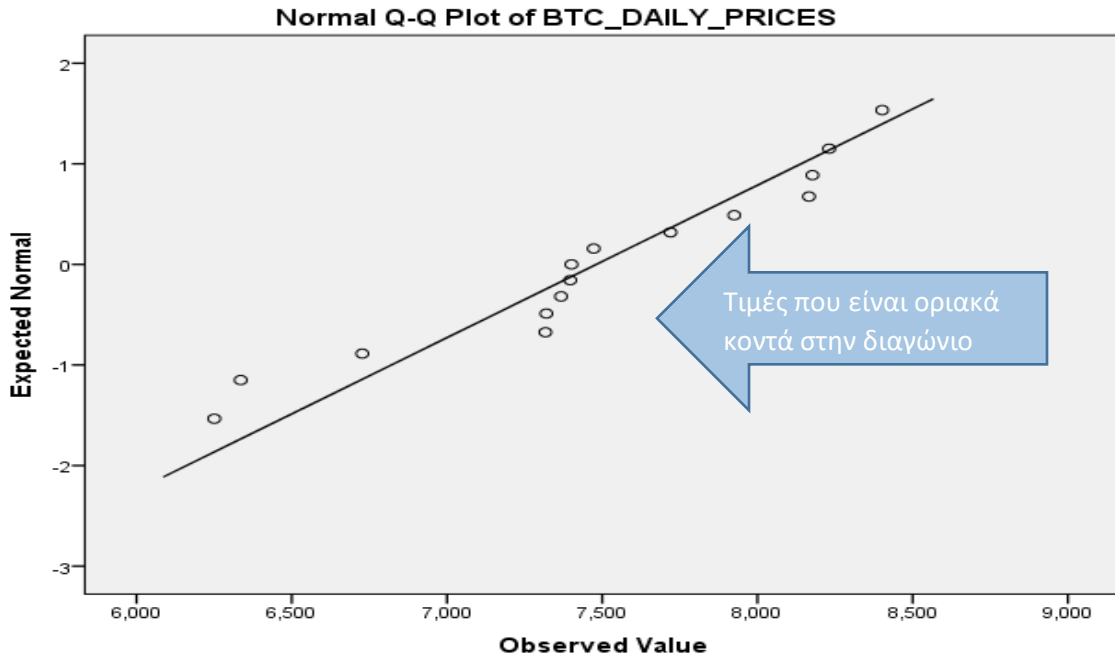


Figure 0-47: TR_BTC_DAILY_PRICES=LG10(BTC_DAILY_PRICES) HISTOGRAM

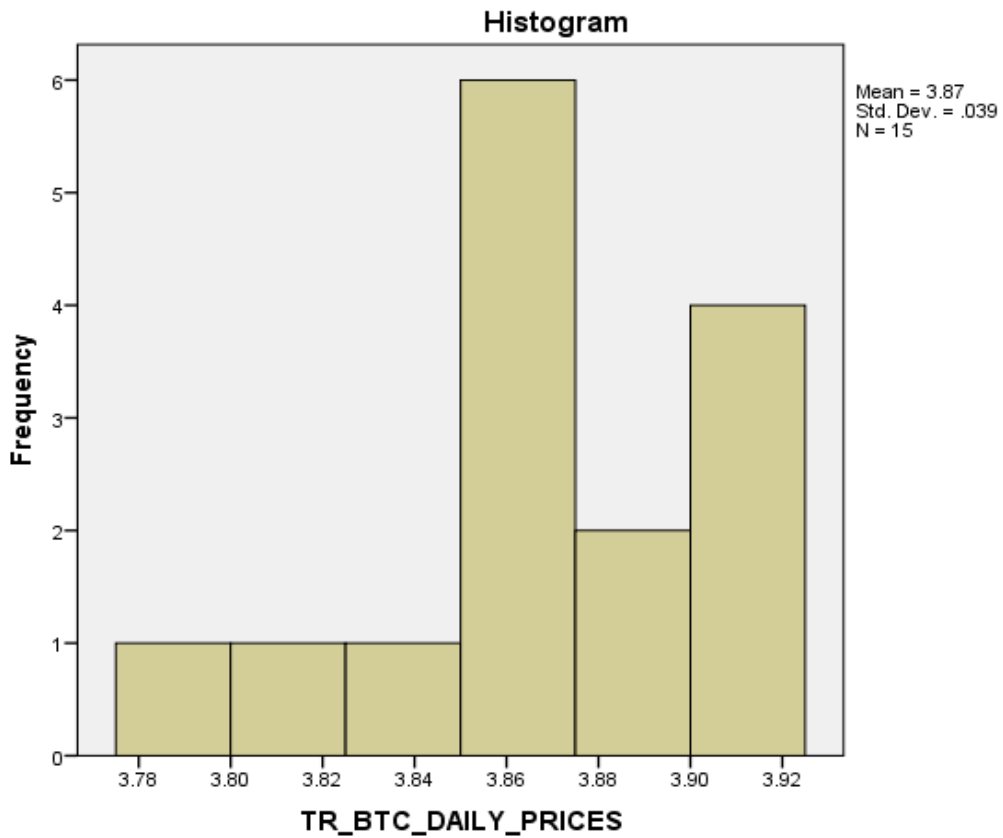
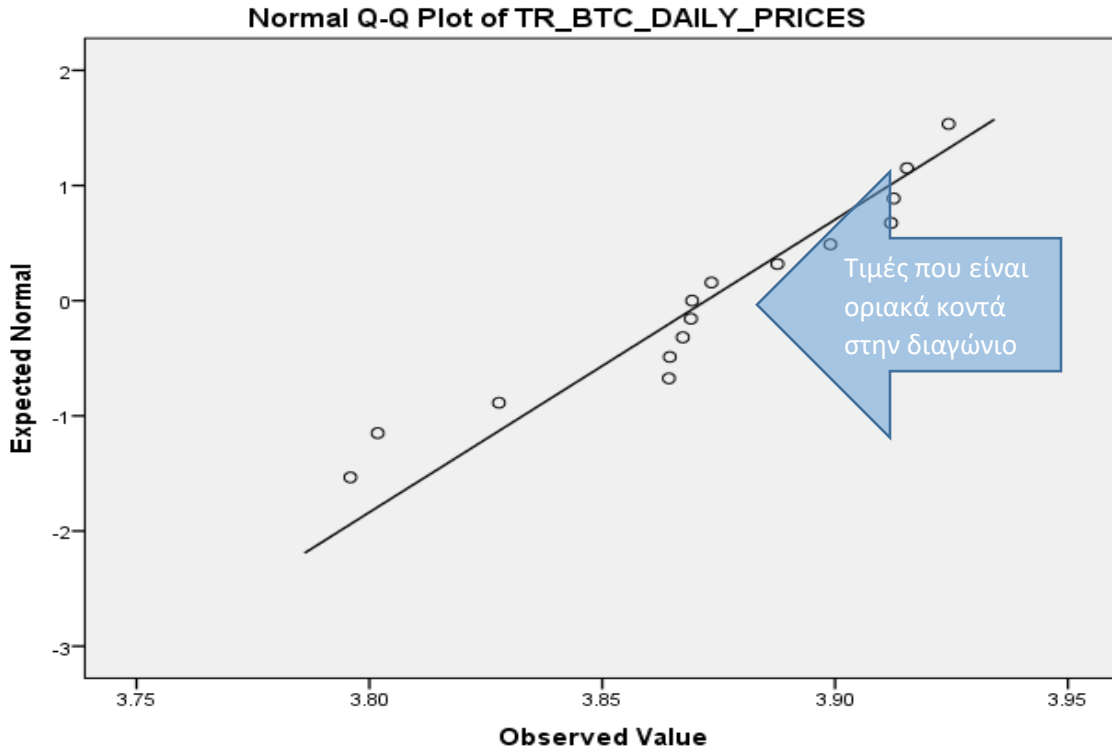


Figure 0-48: Normal Q-Q Plot of TR_BTC_DAILY_PRICES



Στην στατιστική ανάλυση (Biserial Analysis για συσχέτισμό μιας δυαδικής και μια συνεχούς ποσοτικής μεταβλητής) που ακολουθεί παρατηρούμε μέσω του συντελεστή συσχέτισης (Pearson's-Figure 0-49), λόγω του ότι η **τιμή p-value είναι μεγαλύτερη του επιπέδου σημαντικότητας $\alpha=0.05$** που έχουμε θέσει ως όριο, πως επιβεβαιώνεται **ότι δεν υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ των μεταβλητών της στάσης-γνώμης και της τιμής του κρυπτονομίσματος BTC.**

Figure 0-49: CORRELATIONS: POSITIVE_NEGATIVE_VALUE TR_BTC_DAILY_PRICES

		Correlations	
		POSITIVE_NEGATIVE_VALUE	TR_BTC_DAILY_PRICES
POSITIVE_NEGATIVE_VALUE	Pearson Correlation	1	.104
	Sig. (2-tailed)		.711
TR_BTC_DAILY_PRICES	Pearson Correlation	.104	1
	Sig. (2-tailed)	.711	

Καμία συσχέτιση $p > 0.05$

Συμπερασματικά μπορούμε να πούμε ότι απορρίπτεται η **Πρόταση H1:** και αντίθετα ισχύει η **Πρόταση H0:** Η ημερήσια τιμή του κρυπτονομίσματος Bitcoin δεν σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο forum bitcointalk για την αντίστοιχη ημέρα.

3.4.3.2 Ανάλυση τάσης-γνώμης στο Reddit και τιμές του Ethereum

Η αναζήτηση στο εργαλείο <https://webhose.io/> έγινε με τους όρους “Ethereum”, “ethereum”, “ETH”, “eth” και αφορούσε την περίοδο 14/07/2018 έως και 27/07/2018, μια περίοδος εξέτασης που απέδωσε αρχικά 1400 posts. Ακολούθησε ανάλυση στάσης για 1400 αναρτήσεις με το εργαλείο sentiment analysis της Cloud Mining για το RapidMiner. Στην συνέχεια έγιναν αποδεκτά με τα κριτήρια που προαναφέρθηκαν 278 μηνύματα, τα οποία ομαδοποιήθηκαν και αντιστοιχήθηκαν με τις ημερήσιες αξίες του Ethereum. Τέλος τα ζεύγη τάσης-γνώμης και αξίας νομίσματος αναλύθηκαν μέσω του SPSS για πιθανή συσχέτιση.

- **Υποερώτημα 3B-Reddit και τιμές Ethereum:** Η τιμή του κρυπτονομίσματος Ethereum σχετίζεται με τις στάσεις-γνώμες στις αναρτήσεις μηνυμάτων στο forum reddit;

Πρόταση H0: Η ημερήσια τιμή του κρυπτονομίσματος Ethereum δεν σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο forum reddit για την αντίστοιχη ημέρα.

Πρόταση H1: Η ημερήσια τιμή του κρυπτονομίσματος Ethereum σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο forum reddit για την αντίστοιχη ημέρα.

Λόγω του μικρού αριθμού ημερών που χρησιμοποιούνται για τον χαρακτηρισμό ως θετικών ή αρνητικών (αν και ο χαρακτηρισμός προκύπτει ως μέσος όρος πλήθους παρατηρήσεων-σχολίων) ως προς την αξία του νομίσματος γίνεται έλεγχος για κανονικότητα των τιμών (Figure 0-50, Figure 0-51, Figure 0-52). Οπτικά από το ιστόγραμμα παρατηρούμε ότι έχει το σχήμα καμπάνας της κανονικής κατανομής, οι τιμές στο διάγραμμα Q-Q, είναι κοντά στην διαγώνιο και τα τεστ κανονικότητας δείχνουν **p-value μεγαλύτερη του επιπέδου σημαντικότητας 0.05**, που σημαίνει ότι η αρχική υπόθεση (null hypothesis) για κανονική κατανομή μπορεί να υιοθετηθεί. Στην συνέχεια χρησιμοποιούνται ως έχουν, σε αντίθεση με τα προηγούμενα σετ τιμών, καθώς δεν υπάρχει ανάγκη εξομάλυνσης, χρησιμοποιώντας τον δεκαδικό λογάριθμό τους.

Figure 0-50: EXAMINE VARIABLES=ETH_DAILY_PRICES NORMALITY TEST

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ETH_DAILY_PRICES	.103	14	.200*	.981	14	.982

Figure 0-51: ETH_DAILY_PRICES NO NEED FOR NORMALIZATION

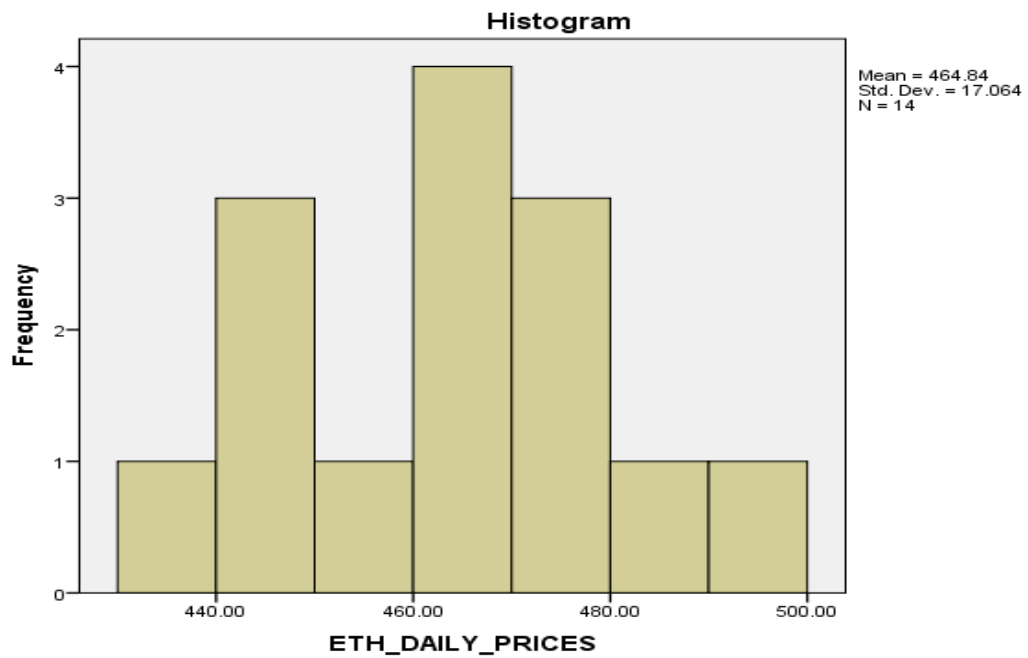
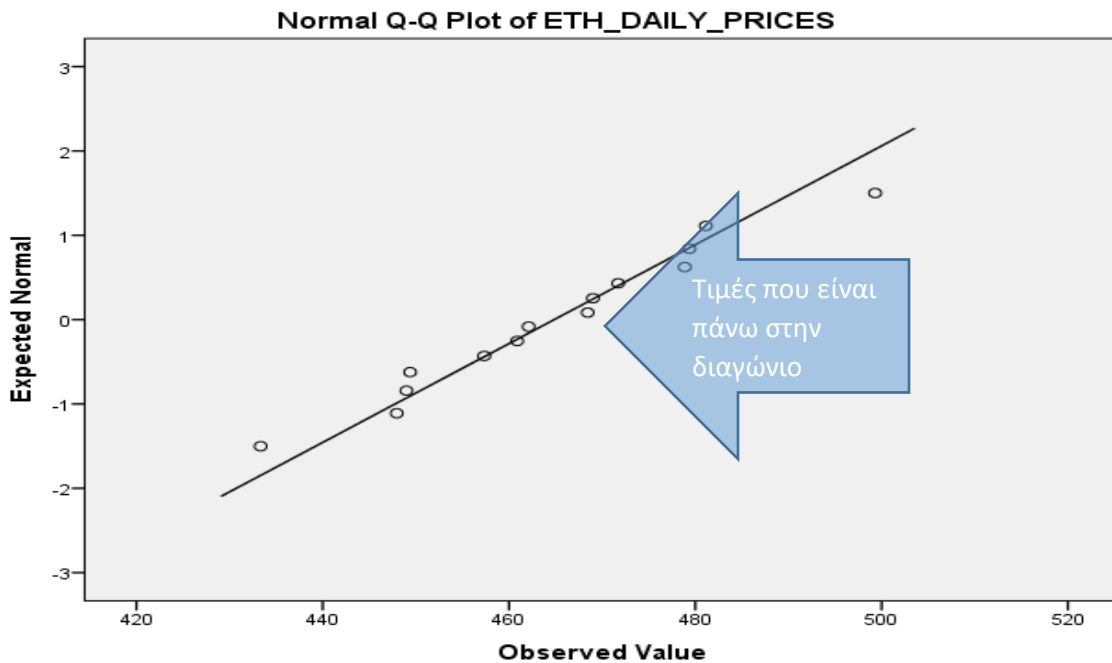


Figure 0-52: Normal Q-Q Plot of ETH_DAILY_PRICES



Στην στατιστική ανάλυση (Biserial Analysis για συσχέτισμό μιας δυαδικής και μια συνεχούς ποσοτικής μεταβλητής) που ακολουθεί, παρατηρούμε μέσω του συντελεστή συσχέτισης (Pearson's-Figure 0-53), λόγω του ότι η **τιμή p-value είναι μεγαλύτερη του επιπέδου σημαντικότητας $\alpha=0.05$** που έχουμε θέσει ως όριο, πως επιβεβαιώνεται ότι **δεν υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ των μεταβλητών της στάσης-γνώμης και της τιμής του κρυπτονομίσματος ETH.**

Figure 0-53: CORRELATIONS=POSITIVE_NEGATIVE_VALUE ETH_DAILY_PRICES

		Correlations	
		POSITIVE_NEGATIVE_VALUE	ETH_DAILY_PRICES
POSITIVE_NEGATIVE_VALUE	Pearson Correlation	1	.036
	Sig. (2-tailed)		.903
ETH_DAILY_PRICES	Pearson Correlation	.036	1
	Sig. (2-tailed)	.903	

Καμία συσχέτιση
p>0.05

Συμπερασματικά μπορούμε να πούμε ότι απορρίπτεται η **Πρόταση H1:** και αντίθετα ισχύει η **Πρόταση H0:** Η ημερήσια τιμή του κρυπτονομίσματος Ethereum δεν σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο forum reddit για την αντίστοιχη ημέρα.

3.4.3.3 Ανάλυση τάσης-γνώμης στο Xrpchat και τιμές του Ripple

Η αναζήτηση στο εργαλείο <https://webhose.io/> έγινε με τους όρους “Ripple”, “ripple”, “XRP”, “xrp” και αφορούσε την περίοδο 14/07/2018 έως και 27/07/2018, μια περίοδος εξέτασης που απέδωσε αρχικά 627 posts, τα οποία μειώθηκαν σε 337 μετά την αφαίρεση των διπλότυπων. Ακολούθησε ανάλυση στάσης για 337 αναρτήσεις με το εργαλείο sentiment analysis της Cloud Mining για το RapidMiner. Στην συνέχεια έγιναν αποδεκτά με τα κριτήρια που προαναφέρθηκαν 132 μηνύματα, τα οποία ομαδοποιήθηκαν και αντιστοιχήθηκαν με τις ημερήσιες αξίες του Ripple. Τέλος τα ζεύγη τάσης-γνώμης και αξίας νομίσματος αναλύθηκαν μέσω του SPSS για πιθανή συσχέτιση.

- **Υποερώτημα 3Γ-Xrpchat και τιμές Ripple:** Η τιμή του κρυπτονομίσματος Ripple σχετίζεται με τις στάσεις-γνώμες στις αναρτήσεις μηνυμάτων στο forum xrpchat;

Πρόταση H0: Η ημερήσια τιμή του κρυπτονομίσματος Ripple δεν σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο forum xrpchat για την αντίστοιχη ημέρα.

Πρόταση H1: Η ημερήσια τιμή του κρυπτονομίσματος Ripple σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο forum xrpchat για την αντίστοιχη ημέρα.

Λόγω του μικρού αριθμού ημερών που χρησιμοποιούνται για τον χαρακτηρισμό ως θετικών ή αρνητικών (αν και ο χαρακτηρισμός προκύπτει ως μέσος όρος πλήθους παρατηρήσεων-σχολίων) ως προς την αξία του νομίσματος, γίνεται έλεγχος για κανονικότητα των τιμών (Figure 0-54, Figure 0-55, Figure 0-56). Οπτικά από το ιστόγραμμα παρατηρούμε ότι δεν έχει το γνωστό σχήμα καμπάνας της κανονικής κατανομής, οι τιμές στο διάγραμμα Q-Q δεν είναι κοντά στην διαγώνιο και τα τεστ κανονικότητας δείχνουν **p-value μικρότερη του επιπέδου σημαντικότητας 0.05**, που σημαίνει ότι η αρχική υπόθεση (null hypothesis) για κανονική κατανομή απορρίπτεται. Στην συνέχεια λοιπόν εξομαλύνονται οι τιμές (TR_PRICES) χρησιμοποιώντας τον δεκαδικό λογάριθμό τους, αλλά και πάλι η κατανομή δεν πλησιάζει την κανονική (Figure 0-57, Figure 0-58).

Figure 0-54: EXAMINE VARIABLES=XRP_DAILY_PRICES NORMALITY TEST

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
XRP_DAILY_PRICES	.266	15	.005	.852	15	.019

Figure 0-55: XRP_DAILY_PRICES BEFORE NORMALIZATION

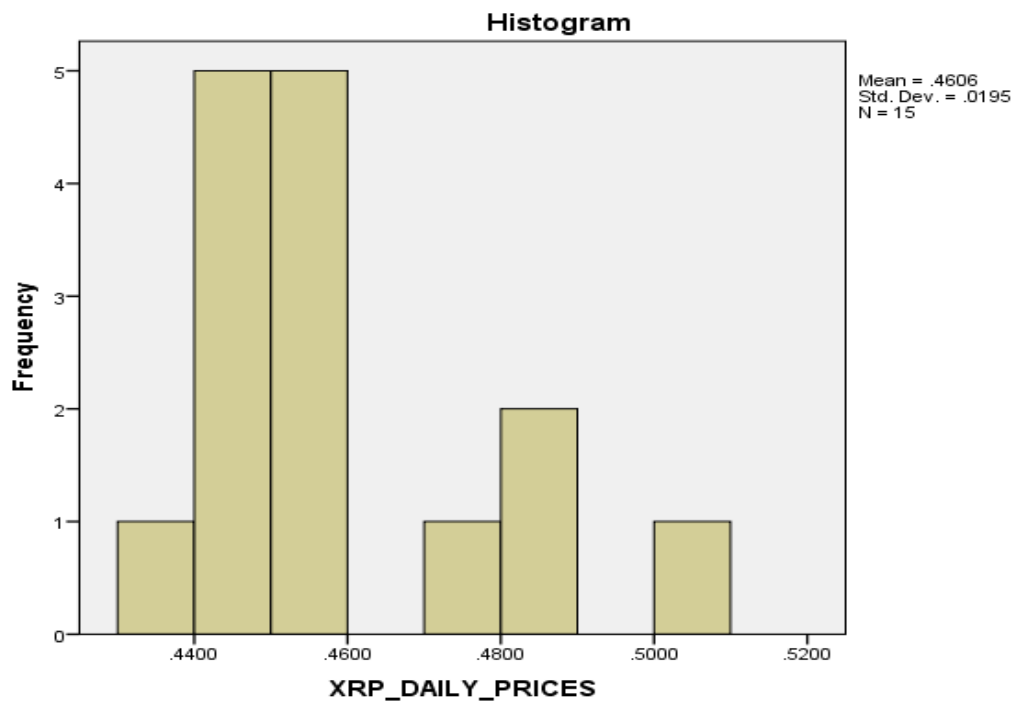


Figure 0-56: Normal Q-Q Plot of XRP_DAILY_PRICES

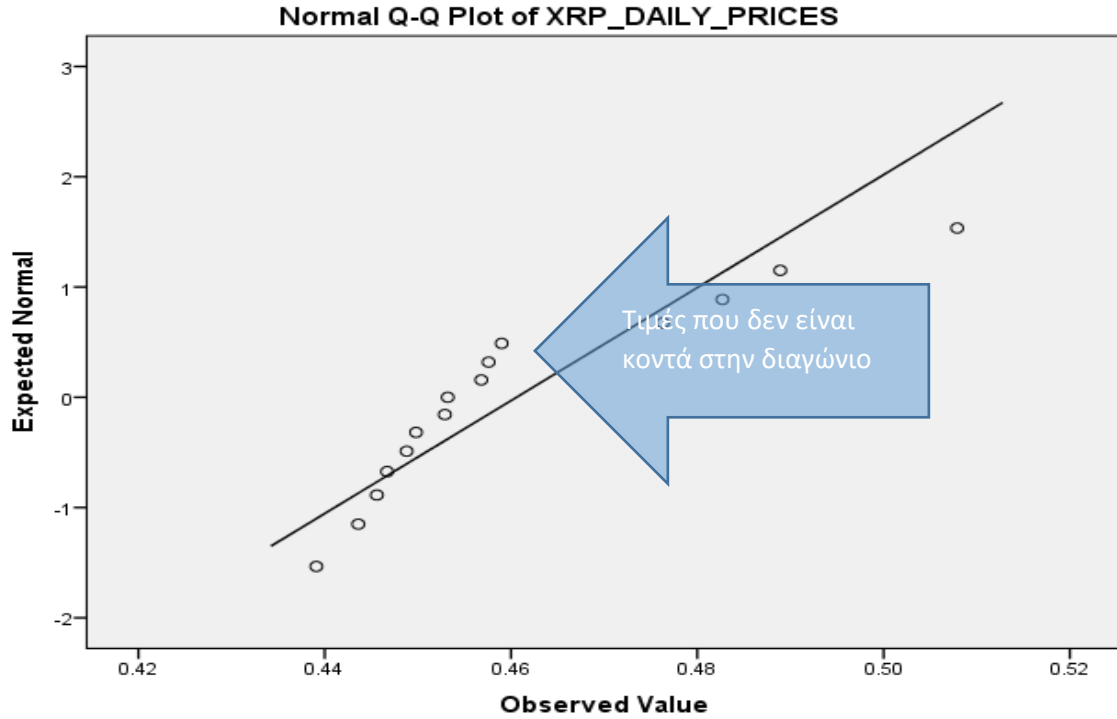


Figure 0-57: TR_XRP_DAILY_PRICES=LG10(XRP_DAILY_PRICES) HISTOGRAM

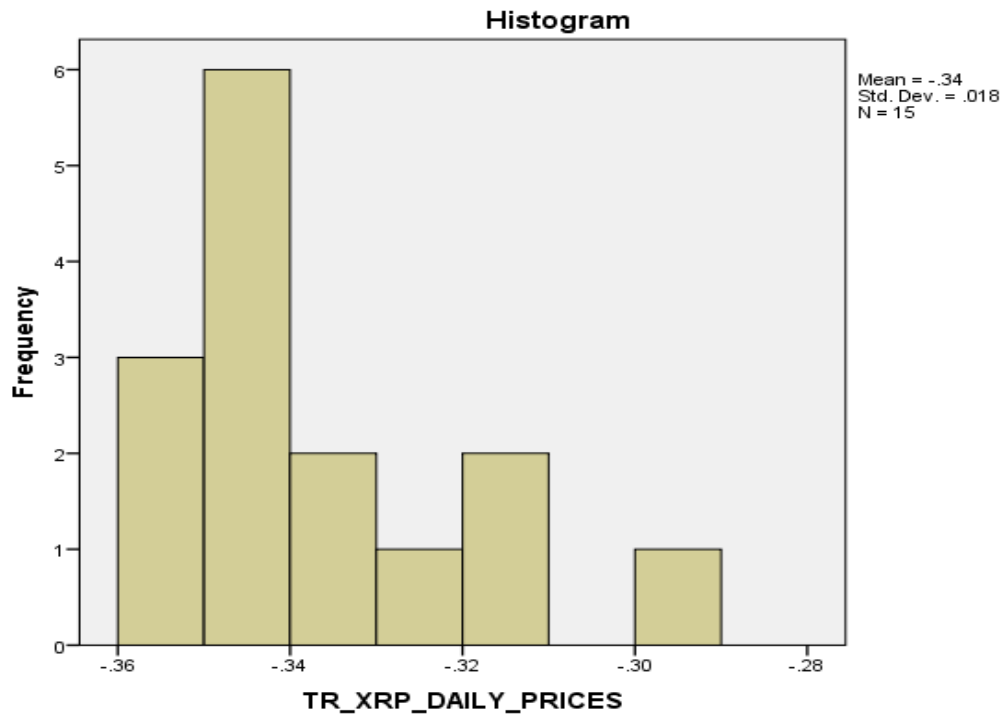
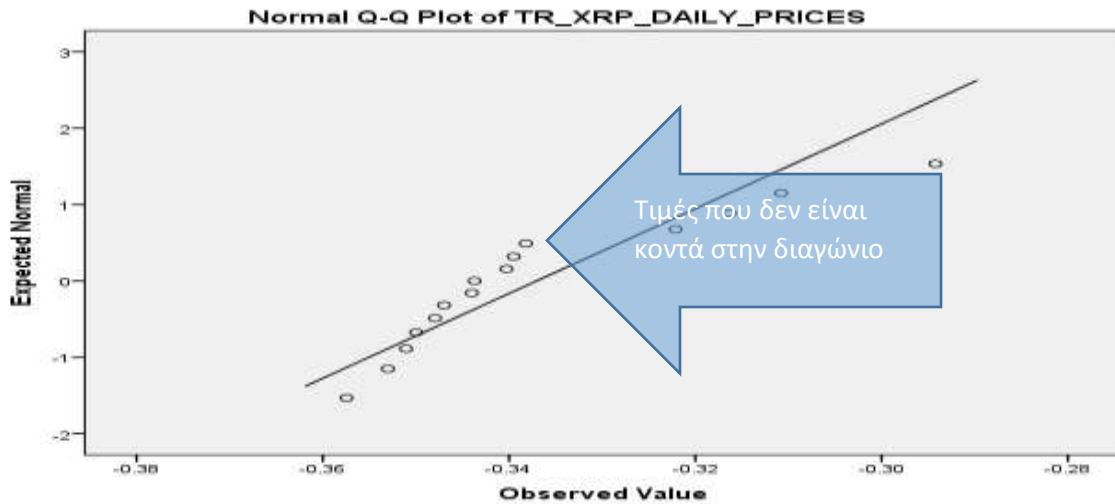


Figure 0-58: Normal Q-Q Plot of TR_XRP_DAILY_PRICES



Στην στατιστική ανάλυση (Biserial Analysis για συσχέτισμό μιας δυαδικής και μια συνεχούς ποσοτικής μεταβλητής) που ακολουθεί παρατηρούμε μέσω του συντελεστή συσχέτισης (Pearson's-Figure 0-59), λόγω του ότι η τιμή **p-value** είναι **μεγαλύτερη του επιπέδου σημαντικότητας $\alpha=0.05$** που έχουμε θέσει ως όριο, πως επιβεβαιώνεται **ότι δεν υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ των μεταβλητών της στάσης-γνώμης και της τιμής του κρυπτονομίσματος XRP.**

Figure 0-59: CORRELATIONS: POSITIVE_NEGATIVE_VALUE TR_XRP_DAILY_PRICES

		POSITIVE_NEGATIVE_VALUE	TR_XRP_DAILY_PRICES
POSITIVE_NEGATIVE_VALUE	Pearson Correlation	1	-.059
	Sig. (2-tailed)		.835
TR_XRP_DAILY_PRICES	Pearson Correlation	-.059	1
	Sig. (2-tailed)	.835	

Καμία συσχέτιση
p>0.05

Συμπερασματικά μπορούμε να πούμε ότι απορρίπτεται η **Πρόταση H1:** και αντίθετα ισχύει η **Πρόταση H0:** Η ημερήσια τιμή του κρυπτονομίσματος Ripple δεν σχετίζεται με την μέση θετική/αρνητική στάση-γνώμη στις αναρτήσεις μηνυμάτων στο forum xrchat για την αντίστοιχη ημέρα.

Εν κατακλείδι, μπορούμε να πούμε ότι όσον αφορά το **3^ο Ερευνητικό Ερώτημα,** οι τιμές των κρυπτονομισμάτων δεν σχετίζονται με τις στάσεις-γνώμες στις αναρτήσεις μηνυμάτων στα αντίστοιχα forum **bitcointalk, reddit** και **xrchat.**

Κεφάλαιο 4

Επίλογος

4.1 Γενικά Συμπεράσματα

Διαπιστώνεται τελικά από τα ευρήματα, ότι υπάρχει μια ισχυρή συσχέτιση μεταξύ των εβδομαδιαίων αναζητήσεων στην μηχανή Google Search και την διαμορφωθείσα τιμή των κρυπτονομισμάτων. Η συσχέτιση αυτή ισχύει και για τα τρία ψηφιακά νομίσματα (Bitcoin, Ethereum, Ripple). Άρα στο 1ο Ερευνητικό ερώτημα(Οι τιμές των κρυπτονομισμάτων σχετίζονται με τις αναζητήσεις στην μηχανή αναζήτησης Google, όπως αυτές απεικονίζονται στο εργαλείο Google Trends;) η απάντηση είναι σαφώς καταφατική.

Συσχέτιση διαπιστώνεται και μεταξύ διαμορφωθείσας τάσης-γνώμης(ωριαία) στις αναρτήσεις στο Twitter και την τιμή του Bitcoin. Αντιθέτως δεν διαπιστώνεται συσχέτιση για τα άλλα 2 νομίσματα και τις αναρτήσεις στο Twitter. Συνεπώς η απάντηση στο 2ο Ερευνητικό ερώτημα (Οι τιμές των κρυπτονομισμάτων σχετίζονται με τις στάσεις-γνώμες στις αναρτήσεις μηνυμάτων στο κοινωνικό δίκτυο Twitter;) είναι μόνο μερικώς καταφατική. Ωστόσο τα μεγέθη του δείγματος των υπολοίπων κρυπτονομισμάτων είναι σαφώς μικρότερα, γεγονός που επιδεινώνεται από τις διαδικασίες φιλτραρίσματος των μηνυμάτων και από τις αναπόφευκτές αδυναμίες των τεχνικών της ανάλυσης γνώμης-τάσης.

Τέλος δεν διαπιστώνεται συσχέτιση μεταξύ των 3 ψηφιακών νομισμάτων και των αναρτήσεων στα επιλεγέντα φόρουμ ανταλλαγής απόψεων. Έτσι στο 3ο Ερευνητικό ερώτημα (Οι τιμές των κρυπτονομισμάτων σχετίζονται με τις στάσεις-γνώμες στις αναρτήσεις μηνυμάτων στο αντίστοιχα forum bitcointalk, reddit και xrpchat;) η απάντηση είναι αρνητική.

Η ανακάλυψη αυτών των συσχετίσεων στο 1ο και 2ο ερευνητικό ερώτημα συνάδει με τα αποτελέσματα των περισσότερων ερευνών που θεωρούν ότι οι τάσεις στις μηχανές αναζήτησης και στα κοινωνικά δίκτυα κινούνται σε αρμονία με την διαμόρφωση των τιμών των ψηφιακών νομισμάτων. Φαίνεται λοιπόν να δικαιώνεται ο αρχικός προσανατολισμός της παρούσας έρευνας στην διερεύνηση εξωγενών φαινομένων ως σχετικών με την αξία των νομισμάτων. Ωστόσο οι απαντήσεις σε μέρος του 2ου ερωτήματος και ολόκληρου του 3ου κινούνται σε αντίθετη κατεύθυνση. Είναι όμως πεποίθηση αυτής της έρευνας ότι αυτό δεν σημαίνει απόρριψη των αρχικών υποθέσεων αλλά ανάγκη για:

- Βελτίωση των μοντέλων ανάλυσης τάσης και των λεξικών τους
- Μεγαλύτερο όγκο αρχικών δεδομένων, ώστε η ανάλυση τάσης-γνώμης να παρέχει περισσότερες ερμηνείες στο φάσμα των θετικών-αρνητικών κατηγοριών και λιγότερες ουδέτερες ή αόριστες ερμηνείες.

- Μεγαλύτερα δείγματα για την εγκυρότερη αναζήτηση συσχέτισης μεταξύ τάσης και τιμών
- Εισαγωγή προβλέψεων για τις περιπτώσεις όπου υπάρχει ασυνέχεια των δεδομένων, όπως αργίες, Σαββατοκύριακα, τοπικές εορτές.

Η πεποίθηση αυτή επίσης ενισχύεται, όπως προαναφέραμε, και από τα συμπεράσματα πλήθους άλλων ερευνών και από την ύπαρξη περιορισμών στην παρούσα, όπως αυτοί αναφέρονται στην παρακάτω ενότητα.

Εν κατακλείδι, γίνεται κατανοητό ότι οι τιμές των ψηφιακών νομισμάτων συσχετίζονται σε μεγάλο βαθμό με εξωγενή φαινόμενα, όπως αναζητήσεις σε μηχανές αναζήτησης με σχετικούς όρους και συζητήσεις που αφορούν τα κρυπτονομίσματα σε κοινωνικά δίκτυα. Η συσχέτιση αυτή φαίνεται να είναι θετική, δηλαδή περισσότερες αναζητήσεις δείχνουν εντονότερο ενδιαφέρον των χρηστών και συμβαδίζουν με αυξήσεις της αξίας των νομισμάτων, παράλληλα η επικράτηση θετικών γνώμων στα κοινωνικά δίκτυα συνδέεται επίσης με ενίσχυση της αξίας των νομισμάτων και το αντίστροφο.

Θεωρητικά, αυτή η πληροφορία θα μπορούσε να λειτουργήσει ως ένας δείκτης πρόβλεψης για την πορεία της τιμής ενός κρυπτονομίσματος. Θα ήταν ωστόσο παρακινδυνευμένο να χρησιμοποιηθεί ως μοναδικός δείκτης, αλλά σε συνδυασμό με άλλους δείκτες τεχνικής ή θεμελιώδους οικονομικής ανάλυσης θα μπορούσε να βοηθήσει σε επενδυτικές αποφάσεις. Ένα πρακτικό παράδειγμα του προηγούμενου θα μπορούσε να είναι το εξής: Υποψήφιος επενδυτής εξετάζει την εβδομαδιαία τάση του Bitcoin στο Google trends και βρίσκοντάς την ανοδική, σε σχέση με προηγούμενες εβδομάδες, θεωρεί ότι την επερχόμενη εβδομάδα το νόμισμα θα κινηθεί επίσης ανοδικά, άρα μπορεί να προβεί με σχετική ασφάλεια στην αγορά του. Ή διαφορετικά αναλύοντας την ωριαία τάση των σχολίων στο Twitter και βρίσκοντάς την θετική, να υποθέσει άνοδο του νομίσματος στην επόμενη ώρα και να προβεί σε αγορά. Ακόμα καλύτερα θα μπορούσε να χρησιμοποιήσει συνδυαστικά τις δύο αυτές αναλύσεις και εφόσον βρίσκονται σε συμφωνία να μειώσει τον κίνδυνο της επενδυτικής του απόφασης.

4.2 Περιορισμοί της έρευνας

Οι περιορισμοί της έρευνας είναι αρκετοί:

- Χρονικοί, που αφορούν το βάθος χρόνου άντλησης των μηνυμάτων και των αναρτήσεων στο διάφορα μέσα κοινωνικής δικτύωσης.
- Ποσοτικοί, που αφορούν την δυνατότητα άντλησης πληροφοριών από τα διάφορα μέσα, η οποία εξαρτάται από τους περιορισμούς που θέτουν τα ίδια στην παροχή και διάθεση δεδομένων.
- Τεχνικοί, που αφορούν τις δυνατότητες των εργαλείων εξόρυξης και επεξεργασίας των δεδομένων.
- Μεθοδολογικοί, οι οποίοι αφορούν τα κριτήρια επιλογής και το μέγεθος των τελικών δειγμάτων.

Αναλυτικότερα:

- Ενώ για τις τιμές των ψηφιακών νομισμάτων υπάρχει πληθώρα δεδομένων και σε βάθος χρόνου, υπάρχουν χρονικοί περιορισμοί για την άντληση μηνυμάτων τόσο στο twitter (λόγω των περιορισμών από το API του twitter-) όσο και στα διάφορα φόρουμ (λόγω των αδειών των scrapping tools).
- Επίσης για τους ίδιους λόγους υπάρχουν ποσοτικοί περιορισμοί στην άντληση μηνυμάτων και αναρτήσεων που σχετίζονται με την δυνατότητα κλήσεων και αιτημάτων προς τους εξυπηρετητές των αντίστοιχων υπηρεσιών.
- Αντίστοιχοι περιορισμοί υπάρχουν και στα εργαλεία sentiment analysis και στα πρόσθετα-επεκτάσεις τους.
- Ενδεικτικά αναφέρουμε:
 - Twitter API: 15 calls every 15 minutes, and 180 calls every 15 minutes.
 - Search Twitter add-on: 100 tweets maximum per search.
 - MeaningCloud : requests*/month= 20000, requests/second= 2, classification models= 1, dictionaries= 1, entries per sentiment model= 30.
*A MeaningCloud request is equal to the analysis of any text up to 500 words for public APIs.
- Χρόνος ανάλυσης γνώμης-τάσης στον Υ/Η (CPU i3,RAM 8GB, VGA Intel integrated graphics card, HDD 1T) για κάθε αρχείο δεδομένων περίπου 1 ώρα.
- Webhose scrapping tool: 100 responses per day.
- Το μοντέλο sentiment analysis της MeaningCloud είναι το γενικό, χωρίς εξειδίκευση σε οικονομικούς όρους και νέα.
- Σε πολλές περιπτώσεις, όπως για την ωριαία ανάλυση τα διαθέσιμα δεδομένα που πληρούν τα ερευνητικά κριτήρια είναι ελάχιστα.
- Επίσης τα τελικά επεξεργασμένα δεδομένα σε περιπτώσεις όπως της ημερήσιας ανάλυσης είναι λίγα σε πλήθος, παράγοντας μικρό τελικό δείγμα, που δημιουργεί προβληματισμούς οι οποίοι αναλύονται στην μεθοδολογία της έρευνας.

4.3 Προτάσεις

Τα αποτελέσματα της έρευνας σε συνδυασμό με τους προαναφερόμενους περιορισμούς κάνουν φανερή την ανάγκη για:

- Ανάπτυξη και χρήση μοντέλων ανάλυσης τάσης-γνώμης πιο προσαρμοσμένων σε οικονομικά και τεχνολογικά θέματα και πιο εξειδικευμένων λεξικών.
- Ανάπτυξη και χρήση ταχύτερων και αποδοτικότερων εργαλείων εξόρυξης δεδομένων.
- Χρήση μεγαλύτερων μεγεθών δειγμάτων ώστε να υπάρχουν περισσότερα διαθέσιμες πληροφορίες μετά το φιλτράρισμα των αρχικών ακατέργαστων δεδομένων.
- Ανοιχτή πρόσβαση στα δεδομένα των εργαλείων κοινωνικής δικτύωσης για ερευνητικούς σκοπούς.
- Δημιουργία βάσεων δεδομένων με ιστορικά στοιχεία για έλεγχο back-testing των εκάστοτε θεωριών, ώστε να αποφεύγεται η επανάληψη της εξόρυξης στοιχείων.
- Συνειδητοποίηση της πολυπλοκότητας των διαστάσεων που εμπεριέχει η ανάλυση εξωγενών κοινωνικών φαινομένων και των παραγόντων που πρέπει να λαμβάνονται υπόψη, ώστε να δημιουργηθεί ένα πλαίσιο για την επιλογή των κατάλληλων δεδομένων και τους βέλτιστους τρόπους χρήσης τους και επεξεργασίας τους.

Κεφάλαιο 5

Βιβλιογραφία

- Abraham, Jethin, Daniel Higdon, John Nelson, and Juan Ibarra. 2018. "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis" 1 (3): 22.
- Balcilar, Mehmet, Elie Bouri, Rangan Gupta, and David Roubaud. 2017. "Can Volume Predict Bitcoin Returns and Volatility? A Quantiles-Based Approach." *Economic Modelling* 64 (August): 74–81. <https://doi.org/10.1016/j.econmod.2017.03.019>.
- Bifet, Albert, and Eibe Frank. 2010. "Sentiment Knowledge Discovery in Twitter Streaming Data." In *Discovery Science*, edited by Bernhard Pfahringer, Geoff Holmes, and Achim Hoffmann, 6332:1–15. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-16184-1_1.
- Bollen, Johan, Huina Mao, and Xiao-Jun Zeng. n.d. "Twitter Mood Predicts the Stock Market," 8.
- Cheah, Eng-Tuck, and John Fry. 2015. "Speculative Bubbles in Bitcoin Markets? An Empirical Investigation into the Fundamental Value of Bitcoin." *Economics Letters* 130 (May): 32–36. <https://doi.org/10.1016/j.econlet.2015.02.029>.
- Cohen-Charash, Yochi, Charles A. Scherbaum, John D. Kammeyer-Mueller, and Barry M. Staw. 2013. "Mood and the Market: Can Press Reports of Investors' Mood Predict Stock Prices?" Edited by Alejandro Raul Hernandez Montoya. *PLoS ONE* 8 (8): e72031. <https://doi.org/10.1371/journal.pone.0072031>.
- DiFonzo, Nicholas, and Prashant Bordia. 1997. "Rumor and Prediction: Making Sense (but Losing Dollars) in the Stock Market." *Organizational Behavior and Human Decision Processes* 71 (3): 329–53. <https://doi.org/10.1006/obhd.1997.2724>.
- Gantori, Sundeep, Paul Donovan, Kiran Ganesh, Matthew DeMichie, Kevin Dennean, Fabio Trussardi, and Michael Klien. 2017. "Cryptocurrencies - Beneath the Bubble." UBS AG and UBS Financial Services Inc. (UBS FS) and UBS Switzerland AG. <https://www.ubs.com/content/dam/WealthManagementAmericas/cio-impact/cryptocurrencies.pdf>.
- Garcia, D., C. J. Tessone, P. Mavrodiev, and N. Perony. 2014. "The Digital Traces of Bubbles: Feedback Cycles between Socio-Economic Signals in the Bitcoin Economy." *Journal of The Royal Society Interface* 11 (99): 20140623–20140623. <https://doi.org/10.1098/rsif.2014.0623>.
- Harvey, Campbell R. n.d. "Do Cryptocurrencies Such as Bitcoin Have a Future? NO As a Currency, Bitcoin Violates All the Rules of Finance," 5.
- Howard, Jeremy, and Sebastian Ruder. 2018. "Universal Language Model Fine-Tuning for Text Classification." *ArXiv:1801.06146* [Cs, Stat], January. <http://arxiv.org/abs/1801.06146>.
- "<https://Ayllen.Com/Research/>." n.d.
- "<https://Www.Meaningcloud.Com/Products/Sentiment-Analysis/>." n.d.
- Hutto, C J, and Eric Gilbert. n.d. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," 11.
- Kaminski, Jermain C. n.d. "Nowcasting the Bitcoin Market with Twitter Signals," 16.

- Kim, Young Bin, Jun Gi Kim, Wook Kim, Jae Ho Im, Tae Hyeong Kim, Shin Jin Kang, and Chang Hun Kim. 2016. "Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies." Edited by Wei-Xing Zhou. *PLOS ONE* 11 (8): e0161197. <https://doi.org/10.1371/journal.pone.0161197>.
- Kouloumpis, Efthymios, Theresa Wilson, and Johanna Moore. n.d. "Twitter Sentiment Analysis: The Good the Bad and the OMG!," 4.
- Kristoufek, Ladislav. 2013. "Bitcoin Meets Google Trends and Wikipedia: Quantifying the Relationship between Phenomena of the Internet Era." *Scientific Reports* 3 (1). <https://doi.org/10.1038/srep03415>.
- Kulkarni, R. V., and A. A. Sattikar. 2012. "Natural Language Processing For Content Analysis in Social Networking." *International Journal of Engineering Inventions* 1 (4): 6–9.
- Lamon, Connor, Eric Nielsen, and Eric Redondo. n.d. "Cryptocurrency Price Prediction Using News and Social Media Sentiment," 6.
- Li, Xin, and Chong Alex Wang. 2017. "The Technology and Economic Determinants of Cryptocurrency Exchange Rates: The Case of Bitcoin." *Decision Support Systems* 95 (March): 49–60. <https://doi.org/10.1016/j.dss.2016.12.001>.
- Matta, Martina, Iliaria Lunesu, and Michele Marchesi. n.d. "Bitcoin Spread Prediction Using Social And Web Search Media," 10.
- Nofsinger, John R. 2005. "Social Mood and Financial Economics." *Journal of Behavioral Finance* 6 (3): 144–60. https://doi.org/10.1207/s15427579jpfm0603_4.
- Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. "Thumbs up?: Sentiment Classification Using Machine Learning Techniques." In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - EMNLP '02*, 10:79–86. Not Known: Association for Computational Linguistics. <https://doi.org/10.3115/1118693.1118704>.
- Schonfeld, Erick. 2009. "Mining The Thought Stream." *TechCrunch Weblog Article*, 3.
- Turney, Peter D. 2001. "Thumbs up or Thumbs down?: Semantic Orientation Applied to Unsupervised Classification of Reviews." In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, 417. Philadelphia, Pennsylvania: Association for Computational Linguistics. <https://doi.org/10.3115/1073083.1073153>.
- Wilson, Matthew Graham, and Aaron Yelowitz. 2014. "Characteristics of Bitcoin Users: An Analysis of Google Search Data." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2518603>.
- Xie, Apoorv Agarwal Boyi, Ilia Vovsha Owen Rambow, and Rebecca Passonneau. n.d. "Sentiment Analysis of Twitter Data," 9.

Κεφάλαιο 6

Παράρτημα

6.1 Πίνακας εικόνων και γραφημάτων

Figure 0-1: Chorus Tweetcatcher Desktop Edition v.1.3.1(2015)	15
Figure 0-2: RapidMiner Studio v9.0-Twitter Sentiment Analysis.....	16
Figure 0-3: Webhose.io.....	19
Figure 0-4: RapidMiner Studio v9.0-Forums Sentiment Analysis	20
Figure 0-5: Γράφημα Bitcoin από google trends 01/01/2015-01/01/2018	23
Figure 0-6: Γράφημα BitcoinUSD Price από Coindesk.....	23
Figure 0-7: Γράφημα Bitcoin από Google trends σε Excel.....	24
Figure 0-8: Γράφημα BitcoinUSD Price από Coindesk σε Excel.....	24
Figure 0-9: Αλληλεπικαλυπτόμενα γραφήματα Bitcoin από Google trends και BitcoinUSD Price από Coindesk σε Excel.....	25
Figure 0-10: SCATTERPLOT(BIVAR)=BTC_Trend_Points WITH Weekly_average_price_BTCUSD	26
Figure 0-11: CORRELATIONS=BTC_Trend_Points Weekly_average_price_BTCUSD.....	27
Figure 0-12: Γράφημα Ethereum από Google trends 06/09/2015-01/01/2018.....	28
Figure 0-13: Γράφημα EthereumUSD Price από Etherscan	29
Figure 0-14: Γράφημα Ethereum από Google trends σε Excel	29
Figure 0-15: Γράφημα EthereumUSD Price από Etherscan σε Excel.....	30
Figure 0-16: Αλληλεπικαλυπτόμενα γραφήματα Ethereum από google trends και EthereumUSD Price από Etherscan σε Excel.....	30
Figure 0-17: SCATTERPLOT(BIVAR)=ETH_Trend_Points WITH Weekly_average_price_ETHUSD	31
Figure 0-18: CORRELATIONS=ETH_Trend_Points Weekly_average_price_ETHUSD	32
Figure 0-19: Γράφημα Ripple από Google trends 01/02/2015-01/01/2018.....	33
Figure 0-20: Γράφημα RippleUSD Price από Yahoo Finance	34
Figure 0-21: Γράφημα Ripple από Google trends σε Excel	34
Figure 0-22: Γράφημα RippleUSD Price από Yahoo Finance σε Excel	35
Figure 0-23: Αλληλεπικαλυπτόμενα γραφήματα Ripple από google trends και RippleUSD Price από Yahoo σε Excel	35
Figure 0-24: SCATTERPLOT(BIVAR)=Ripple_Trend_Points WITH Weekly_average_price_XRPUSD	36
Figure 0-25: CORRELATIONS=Ripple_Trend_Points Weekly_average_price_XRPUSD.....	37
Figure 0-26: EXAMINE VARIABLES=BTC_HOURLY_PRICES NORMALITY TEST.....	39
Figure 0-27: BTC_HOURLY_PRICES HISTOGRAM BEFORE NORMALIZATION.....	39
Figure 0-28: Normal Q-Q Plot of BTC_HOURLY_PRICES.....	40
Figure 0-29: TR_BTC_HOURLY_PRICES=LG10(BTC_HOURLY_PRICES) HISTOGRAM	40
Figure 0-30: Normal Q-Q Plot of TR_BTC_HOURLY_PRICES.....	41

Figure 0-31: CORRELATIONS= POSITIVE_NEGATIVE_VALUE TR_BTC_HOURLY_PRICES	41
Figure 0-32: EXAMINE VARIABLES=ETH_HOURLY_PRICES NORMALITY TEST	43
Figure 0-33: ETH_HOURLY_PRICES BEFORE NORMALIZATION	43
Figure 0-34: Normal Q-Q Plot of ETH_HOURLY_PRICES.....	44
Figure 0-35: TR_ETH_HOURLY_PRICES=LG10(ETH_HOURLY_PRICES) HISTOGRAM.....	44
Figure 0-36: Normal Q-Q Plot of TR_ETH_HOURLY_PRICES	45
Figure 0-37=CORRELATIONS: POSITIVE_NEGATIVE_VALUE TR_ETH_HOURLY_PRICES.....	45
Figure 0-38: EXAMINE VARIABLES=XRP_HOURLY_PRICES NORMALITY TEST	47
Figure 0-39: XRP_HOURLY_PRICES BEFORE NORMALIZATION	47
Figure 0-40: Normal Q-Q Plot of XRP_HOURLY_PRICES.....	48
Figure 0-41: TR_XRP_HOURLY_PRICES=LG10(XRP_HOURLY_PRICES) HISTOGRAM	48
Figure 0-42: Normal Q-Q Plot of TR_XRP_HOURLY_PRICES.....	49
Figure 0-43: CORRELATIONS=POSITIVE_NEGATIVE_VALUE TR_XRP_HOURLY_PRICES	50
Figure 0-44: EXAMINE VARIABLES=BTC_DAILY_PRICES NORMALITY TEST	52
Figure 0-45: BTC_DAILY_PRICES BEFORE NORMALIZATION	52
Figure 0-46: Normal Q-Q Plot of BTC_DAILY_PRICES.....	53
Figure 0-47: TR_BTC_DAILY_PRICES=LG10(BTC_DAILY_PRICES) HISTOGRAM.....	53
Figure 0-48: Normal Q-Q Plot of TR_BTC_DAILY_PRICES	54
Figure 0-49: CORRELATIONS: POSITIVE_NEGATIVE_VALUE TR_BTC_DAILY_PRICES	54
Figure 0-50: EXAMINE VARIABLES=ETH_DAILY_PRICES NORMALITY TEST	56
Figure 0-51: ETH_DAILY_PRICES NO NEED FOR NORMALIZATION	56
Figure 0-52: Normal Q-Q Plot of ETH_DAILY_PRICES	57
Figure 0-53: CORRELATIONS=POSITIVE_NEGATIVE_VALUE ETH_DAILY_PRICES	57
Figure 0-54: EXAMINE VARIABLES=XRP_DAILY_PRICES NORMALITY TEST	59
Figure 0-55: XRP_DAILY_PRICES BEFORE NORMALIZATION	59
Figure 0-56: Normal Q-Q Plot of XRP_DAILY_PRICES.....	60
Figure 0-57: TR_XRP_DAILY_PRICES=LG10(XRP_DAILY_PRICES) HISTOGRAM	60
Figure 0-58: Normal Q-Q Plot of TR_XRP_DAILY_PRICES.....	61
Figure 0-59: CORRELATIONS: POSITIVE_NEGATIVE_VALUE TR_XRP_DAILY_PRICES	61

6.2 Αρχεία Δεδομένων

Google Data	Twitter Data	Forums Data	SPSS Data
Ripple Google Trends 2015-18.csv yahoo-XRP-USD-close_data-2015-02-01_2018-01-06.csv BTC Google Trends 2015-18.csv coindesk-BTC-USD-close_data-2015-01-04_2018-01-06.csv ETH Google Trends 2015-18.csv etherscan-ETH-USD-close_data-2015-09-06_2018-01-06.csv Google Trends-Prices.xlsx	Twitter ETH Sentiment.xlsx Twitter ETH Sentiment-Prices.xlsx Twitter XRP Data.xlsx Twitter XRP Sentiment.xlsx Twitter XRP Sentiment-Prices.xlsx XRPRawData0.txt BTCRawData0.txt BTCRawData1.txt ETHRawData0.txt ETHRawData1.txt Kraken_BTCUSD_1h.csv Kraken_BTCUSD_1h.xls x Kraken_ETHUSD_1h.csv Kraken_ETHUSD_1h.xls x Kraken_XRPUSD_1h.csv Kraken_XRPUSD_1h.xls x Kraken_XRPUSD_1h.csv Twitter BTC Data.xlsx Twitter BTC Sentiment-Prices.xlsx Twitter ETH Data.xlsx	Xrpchat XRP Data.xlsx Xrpchat XRP Sentiment.xlsx Xrpchat XRP Sentiment-Prices.xlsx Bitcointalk BTC Data.xlsx Bitcointalk BTC Sentiment.xlsx Bitcointalk BTC Sentiment-Prices.xlsx Kraken_BTCUSD_d.csv Kraken_BTCUSD_d.xls x Kraken_ETHUSD_d.csv Kraken_ETHUSD_d.xls x Kraken_XRPUSD_d.csv Kraken_XRPUSD_d.xls x Reddit ETH Data.xlsx Reddit ETH Sentiment.xlsx Reddit ETH Sentiment-Prices.xlsx	POINT_BISERIAL_CORRELATION TWITTER_XRP_SENTIMENT_PRICES.spv POINT_BISERIAL_CORRELATION XRPCHAT_XRP_SENTIMENT_PRICES.sav POINT_BISERIAL_CORRELATION XRPCHAT_XRP_SENTIMENT_PRICES.spv SCATTERPLOT BTC_Trend_Points WITH Weekly_average_price_BTCUSD.spv SCATTERPLOT ETH_Trend_Points WITH Weekly_average_price_ETHUSD.spv SCATTERPLOT XRP_Trend_Points WITH Weekly_average_price_XRPUSD.spv CORRELATIONS BTC_Trend_Points WITH Weekly_average_price_BTCUSD.spv CORRELATIONS ETH_Trend_Points WITH Weekly_average_price_ETHUSD.spv CORRELATIONS XRP_Trend_Points WITH Weekly_average_price_XRPUSD.spv DATA Google_Trend_Points WITH Weekly_average_prices.sav NORMALITY_XRPCHAT_XRP_PRICES.spv POINT_BISERIAL_CORRELATION BITCOINTALK_BTC_SENTIMENT_PRICES.sav POINT_BISERIAL_CORRELATION BITCOINTALK_BTC_SENTIMENT_PRICES.spv POINT_BISERIAL_CORRELATION REDDIT_ETH_SENTIMENT_PRICES.sav POINT_BISERIAL_CORRELATION REDDIT_ETH_SENTIMENT_PRICES.spv POINT_BISERIAL_CORRELATION TWITTER_BTC_SENTIMENT_PRICES.sav POINT_BISERIAL_CORRELATION TWITTER_BTC_SENTIMENT_PRICES.spv POINT_BISERIAL_CORRELATION TWITTER_ETH_SENTIMENT_PRICES.sav POINT_BISERIAL_CORRELATION TWITTER_ETH_SENTIMENT_PRICES.spv POINT_BISERIAL_CORRELATION TWITTER_XRP_SENTIMENT_PRICES.sav NORMALITY_TWITTER_TR_BTC_HOURLY_PRICES.spv NORMALITY_TWITTER_TR_ETH_HOURLY_PRICES.spv NORMALITY_TWITTER_TR_XRP_HOURLY_PRICES.spv NORMALITY_BITCOINTALK_TR_BTC_DAILY_PRICES.spv v NORMALITY_XRPCHAT.TR_XRP_DAILY_PRICES.spv