

CRYPTOMONNAIES EN 2024 : ANALYSE DES REACTIONS ET OPINIONS DES UTILISATEURS DE REDDIT

MR01 Ylenia CURCI



18 SEPTEMBER 2024

UTBM

Robert Hakobyan

Contents

I. Introduction	2
II. Scrapping des Subreddits	3
II.1 Fonctionnement reddit	3
II.2 Scrapping des subreddits	3
II.2.1 Technologie utilisée	3
II.2.2 Subreddit scrapé	3
II.2.3 Procédure du Scrapping	4
II.2.4 Limite dans notre procédé	5
II.3 Bilan du scraping	6
III. Analyse des postes et commentaires	7
III.1 Technologie utilisée	7
III.2 Utilisation	7
III.3 Nuage de mot	8
IV. Bilan	10
IV.1 Nos résultats	10
IV.1.1 Diagrammes en boîtes	10
IV.1.2 Nuage de mot :	16
IV.2 Analyse avec des sources externes	18
IV.2.1 Analyse du contexte socio-démographique	18
IV.2.2 Avis médias sur le courant 2024	21
IV.2.3 Comparaison des résultats obtenus sur twitter	23
V. Conclusion	24
VI. Bibliographie	25

I. Introduction

Dans le cadre de mon projet de recherche pour MR01, l'objectif est d'analyser les avis des utilisateurs de Bitcoin sur Reddit au cours de l'année 2024. Cette étude vise à scraper les données de différents subreddits dédiés au Bitcoin afin de comprendre les opinions et sentiments exprimés par les utilisateurs. Les résultats obtenus seront ensuite comparés avec des recherches scientifiques portant sur des périodes antérieures et d'autres plateformes, afin de mettre en lumière les évolutions des perceptions et des discussions autour du Bitcoin.

Pour mener à bien cette analyse, des outils informatiques de traitement de données textuelles et d'analyse des sentiments seront utilisés. En scrappant les posts et les commentaires de Reddit, nous pourrions extraire des informations pertinentes et évaluer les sentiments des utilisateurs, qu'ils soient positifs, négatifs ou neutres. Cette approche nous permettra d'obtenir une vue d'ensemble des tendances et opinions émergentes autour des cryptomonnaies en 2024, tout en les situant dans une perspective plus large à travers la comparaison avec des études réalisées sur d'autres périodes et plateformes.

II. Scrapping des Subreddits

II.1 Fonctionnement reddit

Reddit est une plateforme de discussion communautaire où les utilisateurs partagent des liens, des images ou des textes, et interagissent via des commentaires. Les contenus sont organisés en subreddits, qui sont des communautés thématiques centrées sur un sujet précis (par exemple, r/technology pour la technologie). Chaque post et commentaire peut être voté positivement ou négativement par les utilisateurs, influençant sa visibilité. Les posts ayant plus de votes positifs apparaissent en haut des discussions. Les commentaires permettent aux utilisateurs d'échanger et de débattre sur les posts, créant des discussions hiérarchisées.

II.2 Scrapping des subreddits

II.2.1 Technologie utilisée

Pour scraper les données de Reddit, une bibliothèque officielle utilisant l'API de reddit qui s'appelle PRAW¹ (Python Reddit API Wrapper) qui est une bibliothèque Python qui facilite l'interaction avec l'API de Reddit. Elle permet aux développeurs d'automatiser des tâches sur Reddit, comme extraire des posts, publier des commentaires, ou gérer des comptes. PRAW simplifie l'accès à l'API de Reddit en offrant des méthodes prêtes à l'emploi pour naviguer entre les subreddits, interagir avec des posts (votes, commentaires) et récupérer des données.

Son fonctionnement repose sur l'authentification via OAuth2, et il utilise des objets Python pour représenter les entités de Reddit (comme les posts et les commentaires).

II.2.2 Subreddit scrapé

Pour ce projet, un total de 11 subreddits furent scrapés. Les plus connus furent pris, ainsi que certains un peu moins connus pour avoir un échantillon plus large et pouvoir les comparer entre eux. Le tout a été réalisé le 5 septembre 2024. Dans deux domaines différents, les actualités globales et le marché des traders. Dans la première catégorie nous retrouvons les subreddits suivants :

- Bitcoin : <https://www.reddit.com/r/Bitcoin/> avec 7 millions membres
- CryptoCurrency : <https://www.reddit.com/r/CryptoCurrency/> avec 8.4 millions membres
- BitcoinDiscussion: <https://www.reddit.com/r/BitcoinDiscussion/> avec 34 milles membres
- btc : <https://www.reddit.com/r/btc/> avec 1.1 millions membres
- CryptoCurrencies : <https://www.reddit.com/r/CryptoCurrencies/> avec 394 milles membres
- ethereum : <https://www.reddit.com/r/ethereum/> avec 3.3 millions membres

Pour la catégorie trader nous avons :

- CryptoMarkets : <https://www.reddit.com/r/CryptoMarkets/> avec 1.7 millions membres
- BitcoinMarkets : <https://www.reddit.com/r/BitcoinMarkets/> avec 290 milles membres

¹ <https://praw.readthedocs.io/en/stable/>

- bitcointrading : <https://www.reddit.com/r/bitcointrading/> avec 10 milles membres
- CryptoCurrencyTrading : <https://www.reddit.com/r/CryptoCurrencyTrading/> avec 177 milles membres
- Ethtrader : <https://www.reddit.com/r/ethtrader/> avec 2.3 millions membres

II.2.3 Procédure du Scraping

Tout d'abord définissons le scraping². Il est une technique qui consiste à utiliser des programmes informatiques pour extraire automatiquement des informations à partir de sites web. En parcourant les pages web, ces programmes collectent des données spécifiques, comme du texte, des images ou des chiffres, pour les stocker et les analyser ultérieurement.

Tout le code fut codé en python et pour scrapper une classe RedditScrapper fut créé. En y incorporant toutes mes fonctions puis en l'appelant dans ma fonction main. Au sein de cette classe de nombreuses fonctionnalités sont disponibles.

II.2.3.1 Authentification Reddit

Pour ce faire dans un premier temps un objet Reddit en s'authentifiant et permet créer afin de pouvoir l'utiliser ensuite pour les différentes.

Ci-contre la création de notre objet Reddit afin de l'utiliser plus tard pour scrapper ensuite les Subreddit qui nous permettra de récupérer les données pertinentes.

```
self.reddit = praw.Reddit(
    client_id=client_id,
    client_secret=client_secret,
    password=password,
    user_agent=user_agent,
    username=username,
)
```

Figure 1 - Objet praw Reddit

II.2.3.2 Collecte de posts / commentaires

La collecte de posts et de commentaires dans ce projet suit une méthode structurée pour assurer une extraction efficace.

Scraping des subreddits :

Le processus commence par la récupération des posts de subreddits spécifiques, organisés en deux catégories : *actualités* et *trading*. Trois types de posts sont collectés :

1. **Top posts** : ceux avec le plus de votes sur une période donnée.
2. **Nouveaux posts** : les plus récents, pour capter les tendances.
3. **Posts populaires (hot)** : ceux qui génèrent beaucoup d'engagement.

Ces posts sont filtrés en fonction d'un **timestamp** pour ne récupérer que ceux créés après une date spécifique, dans notre cas après 2024. Pour chaque post, des informations comme le **titre**, le **score**, et l'**auteur** sont extraites et stockées dans une instance de la classe subreddit. Un mécanisme de contrôle des doublons est également mis en place.

² https://fr.wikipedia.org/wiki/Web_scraping

Collecte de commentaires :

Pour chaque post collecté, les **commentaires** sont ensuite extraits, en se concentrant sur ceux ayant reçu le plus de **votes positifs**. Les votes positifs ne signifient pas forcément un avis favorable, mais plutôt un large consensus ou intérêt de la communauté.

Bien que les discussions Reddit soient structurées en **threads imbriqués**, elles sont **aplaties** dans ce projet pour permettre une analyse linéaire plus simple. Chaque commentaire est représenté par une classe dédiée contenant des informations telles que le **texte**, l'**auteur**, et le **score**.

Ainsi, la collecte se fait en plusieurs étapes : identification des posts pertinents, extraction des commentaires les plus votés, et structuration de ces données pour une utilisation future.

II.2.3.3 Enregistrement des données

Pour l'enregistrement des données, le script enregistre en format CSV chaque post et ses commentaires dans des fichiers séparés et classés par catégorie (actualités ou trading).

Fichiers de Posts : Les titres et les scores des posts sont enregistrés dans des fichiers distincts pour chaque sous-reddit, classés sous les répertoires news ou trade selon la catégorie.

Fichiers de Commentaires : Les commentaires de chaque post sont également sauvegardés dans des fichiers séparés. Chaque commentaire est stocké avec son texte nettoyé et son score, dans des fichiers dédiés.

Cette organisation permet une séparation claire entre les données des posts et des commentaires, ainsi qu'une distinction par catégorie, facilitant ainsi l'analyse ultérieure.

II.2.4 Limite dans notre procédé

Le script de scraping ne prend pas en compte les images, GIFs, et « memes », qui sont courants sur Reddit. Leur inclusion nécessiterait des techniques supplémentaires pour télécharger et analyser ces contenus visuels, ce qui augmenterait la complexité et les besoins en stockage. Ainsi, le script actuel se limite aux posts textuels et commentaires, excluant les médias visuels.

Une alternative comme PMAW³, dérivée de PRAW et utilisant le multithreading, aurait permis d'optimiser le processus de scraping. Cependant, cette approche aurait impliqué des concepts techniques plus complexes et un temps de développement plus long, sans garantir des résultats fondamentalement différents, seulement une amélioration des performances.

Il existe également le service API Pushshift, qui propose des fonctionnalités avancées pour accéder aux données Reddit, il présente des problèmes de confidentialité, car il peut inclure des posts supprimés par les utilisateurs. Cependant elle est depuis 2021 seulement utilisable après analyse de dossier pour des sujets spécifiques.

Reddit impose une limite de 1000 requêtes maximum, permettant de récupérer qu'une petite partie des commentaires ou posts, ce qui restreint le volume de données pouvant être récupérées. Bien que les catégories *top*, *new*, et *hot* permettent d'accéder à plus de posts, cette contrainte limite la couverture complète des données.

³ <https://github.com/mattpodolak/pmaw>

La version de PRAW 7.7.1 à un problème quand on veut récolter beaucoup de commentaire et n'a toujours pas été résolu. La version 7.1.1 fonctionne correctement.

II.3 Bilan du scraping

Finalement, malgré quelques problèmes récurrents avec l'API de Reddit, j'ai réussi à collecter un total de 1 587 posts, répartis-en 799 dans la section "news" et 788 dans la section "trade". Cela représente également 53 885 commentaires, dont 30 949 dans la partie "news" et 22 936 dans la partie "trade" tous écrits en 2024.

Ces données ont été récupérées grâce à une fonction de ma classe RedditScraper, qui parse les documents que j'ai extraits.

Cela constitue un échantillon significatif, suffisant pour mener une analyse approfondie et en tirer des conclusions pertinentes.

Ci-dessous un exemple avec le subreddit Bitcoin. Nous y voyons le titre ainsi que les commentaires avec le nombre de vote en dessous.

```
Title: Daily Discussion, September 06, 2024
3
Oh no guis! Bitcoin dumping again... time to sell again and buy back in when line goes up
1
B B B Benny and the jets
1
We just landed at 55.3k usd. What's next?
1
```

Figure 2 - Exemple du scraping

III. Analyse des postes et commentaires

III.1 Technologie utilisée

Pour analyser les différents posts ainsi que leur commentaire j'ai utilisé NLTK⁴ (Natural Language Toolkit) est une bibliothèque Python utilisée pour le traitement du langage naturel. Elle fournit des outils pour manipuler des textes, tels que le tokenization, l'analyse grammaticale, l'étiquetage de parties du discours, l'extraction de mots-clés, et plus encore. NLTK contient également des corpus et des algorithmes pour réaliser des tâches comme l'analyse de sentiments, l'apprentissage automatique et la génération de texte.

III.2 Utilisation

Pour analyser les données scrappées à partir de Reddit, un nouveau fichier a été créé afin d'y appliquer une analyse des sentiments. Ce fichier contient les titres des posts, le nombre de votes, les commentaires associés et leurs votes respectifs. L'objectif est de déterminer si les discussions dans les subreddits sont majoritairement positives, négatives ou neutres, et de générer un rapport sur ces sentiments.

Tout d'abord, les textes des posts et des commentaires sont prétraités. Cette étape consiste à transformer les textes en minuscules, à diviser les phrases en mots individuels (tokenization), et à supprimer les caractères non alphanumériques (comme la ponctuation). Les mots sont ensuite lemmatisés, ce qui signifie qu'ils sont ramenés à leur forme de base. Par exemple, les mots comme "running" ou "ran" sont ramenés à "run". Ce processus permet de mieux analyser le contenu textuel en supprimant les variations grammaticales inutiles.

Une fois les textes prétraités, l'analyse des sentiments est réalisée à l'aide de l'outil VADER (Valence Aware Dictionary for Sentiment Reasoning). Cet outil évalue chaque texte et attribue un score de sentiment qui varie entre -1 (très négatif) et 1 (très positif). Les scores de sentiment permettent de comprendre l'orientation émotionnelle des discussions dans les subreddits. Si un post ou un commentaire reçoit un nombre négatif de votes (c'est-à-dire qu'il est impopulaire), le score de sentiment est ajusté en conséquence en inversant la valeur, afin de refléter ce désaccord dans l'analyse.

Pour chaque post, le programme calcule également une moyenne pondérée des sentiments exprimés dans les commentaires. Cela permet de mesurer non seulement le sentiment du post lui-même, mais aussi celui de l'ensemble des discussions qui en découlent. Le score des commentaires est pondéré par le nombre de votes qu'ils reçoivent, afin de mieux refléter l'importance des opinions exprimées.

Une fois ces analyses effectuées, les résultats sont compilés dans un fichier CSV. Ce fichier contient pour chaque post le titre, le score de sentiment du titre, le nombre de votes du post, la moyenne des sentiments des commentaires et le nombre total de votes des commentaires. Ce fichier permet ainsi d'obtenir une vue d'ensemble des émotions et tendances qui se dégagent des discussions dans différents subreddits.

⁴ <https://www.nltk.org/>

Finalement, les résultats sont enregistrés dans des fichiers CSV, divisés selon les catégories du subreddit (par exemple, "news" ou "trade"). Ces rapports fournissent une interprétation claire et chiffrée des sentiments dans les conversations Reddit, offrant ainsi un outil précieux pour comprendre les dynamiques de chaque communauté.

Ci-dessous un exemple de titre et leur score respectif dans le subreddit Bitcoin. Le premier titre à un score de 0 (neutre) avec un vote positif et une moyenne de score de commentaire de -0.012 (presque neutre) avec un total de 8 votes sur les commentaires.

```
Two Monthly Buys or Daily Purchases?, "Score: 0.0, 1, -0.012749999999999997, 8"
"Daily Discussion, September 06, 2024", "Score: 0.0, 3, 0.03131428571428572, 7"
```

Figure 3 - Exemple de l'analyse

III.3 Nuage de mot

L'objectif ici est de traiter les posts et commentaires déjà scraper, d'identifier les mots les plus fréquemment utilisés, puis de créer des visualisations sous forme de **nuages de mots**, où la taille des mots est proportionnelle à leur fréquence d'apparition.

Fonctionnement général

1. **Lecture et traitement des fichiers** : Le script récupère des textes à partir de plusieurs fichiers, chacun représentant un subreddit spécifique. Nous séparons encore une fois les news des trades.

Pour chaque fichier, le texte est nettoyé en supprimant les caractères inutiles, la ponctuation, les nombres, ainsi que les mots courants (comme "the", "a", "is", etc.). Ce processus vise à ne conserver que les mots significatifs.

2. **Extraction des mots fréquents** : Une fois le texte nettoyé, le script effectue un comptage des occurrences de chaque mot. Il identifie les mots qui apparaissent le plus souvent dans chaque catégorie de subreddits (news et trade), ce qui permet de faire ressortir les termes les plus récurrents dans les discussions.
3. **Création de nuages de mots** : Pour représenter visuellement les résultats, le script génère des **nuages de mots**. Dans un nuage de mots, les mots les plus fréquents apparaissent en plus grand et plus gras, tandis que les mots moins fréquents sont affichés en plus petit. Cela permet de visualiser rapidement les mots dominants dans chaque catégorie. Les nuages de mots sont ensuite sauvegardés sous forme d'images dans des dossiers spécifiques.

Technologies utilisées

1. **Traitement du langage naturel** : Le script utilise des outils de traitement du langage naturel pour gérer le texte brut et effectuer des opérations comme la suppression des mots courants et la normalisation des mots (par exemple, passer tous les mots en minuscules).
2. **Analyse statistique** : La bibliothèque **collections.Counter** est utilisée pour compter les occurrences de chaque mot et générer des statistiques de fréquence.

3. **Visualisation de données** : Les nuages de mots sont créés à l'aide de la bibliothèque WordCloud. Ces visualisations permettent de transformer des statistiques textuelles en graphiques facilement interprétables.
4. **Matplotlib**: Utilisé pour gérer l'affichage et la sauvegarde des nuages de mots sous forme d'images, Matplotlib permet d'enregistrer les résultats dans un format visuel accessible.

IV. Bilan

IV.1 Nos résultats

IV.1.1 Diagrammes en boîtes

Pour l'analyse des sentiments des titres et des commentaires sur Reddit, nous avons utilisé Python avec la bibliothèque Matplotlib pour générer des diagrammes en boîte. Ces graphiques permettent de visualiser efficacement la répartition des données et d'identifier les valeurs aberrantes.

Un **diagramme en boîte** résume la distribution des données en affichant :

- **La Boîte** : Représente les 50 % centraux des données, allant du 1er quartile (Q1) au 3e quartile (Q3), avec la ligne médiane au centre indiquant la valeur médiane.
- **Les Moustaches** : Les lignes qui s'étendent au-delà de la boîte jusqu'à 2 fois l'intervalle interquartile (IQR), délimitant la plage normale des données.
- **Les Points Aberrants** : Les points situés au-delà des moustaches, représentant les valeurs extrêmes ou atypiques.

Ces diagrammes fournissent une vue claire de la dispersion des scores de sentiments pour chaque subreddit et facilitent la comparaison entre les sous-groupes "news" et "trade", tout en mettant en évidence les variations et les anomalies dans les données.

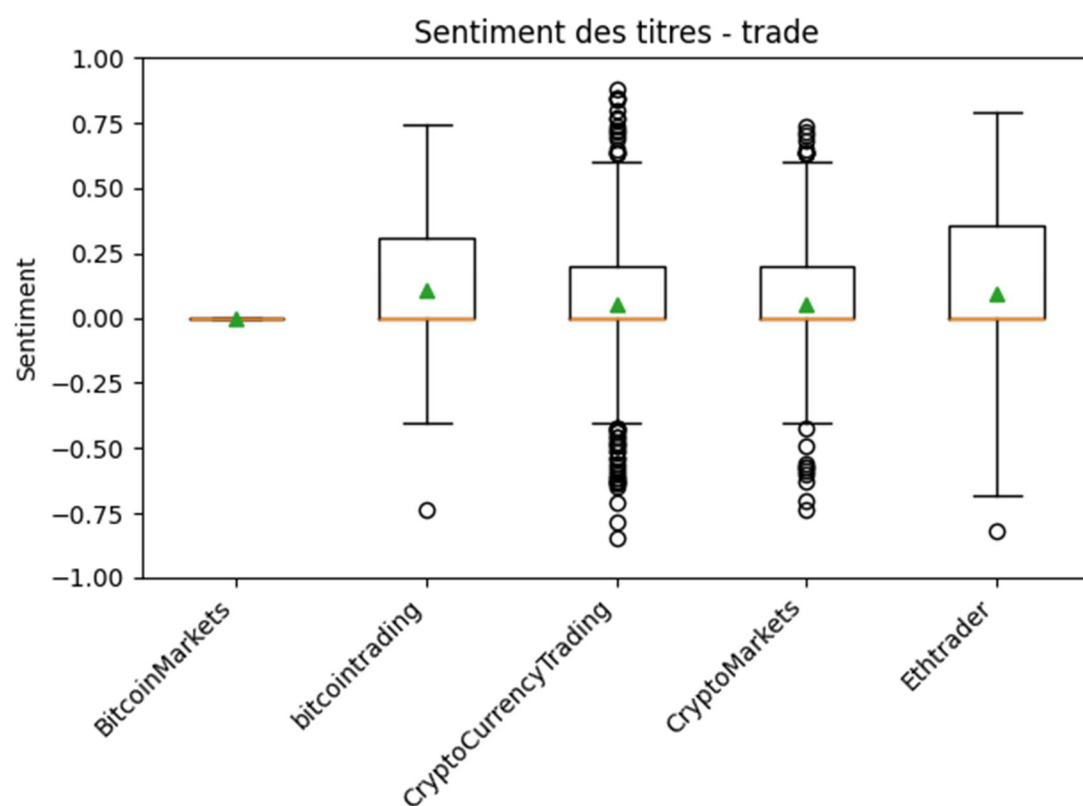


Figure 4 - Diagramme en boîte des titres dans la catégorie trade

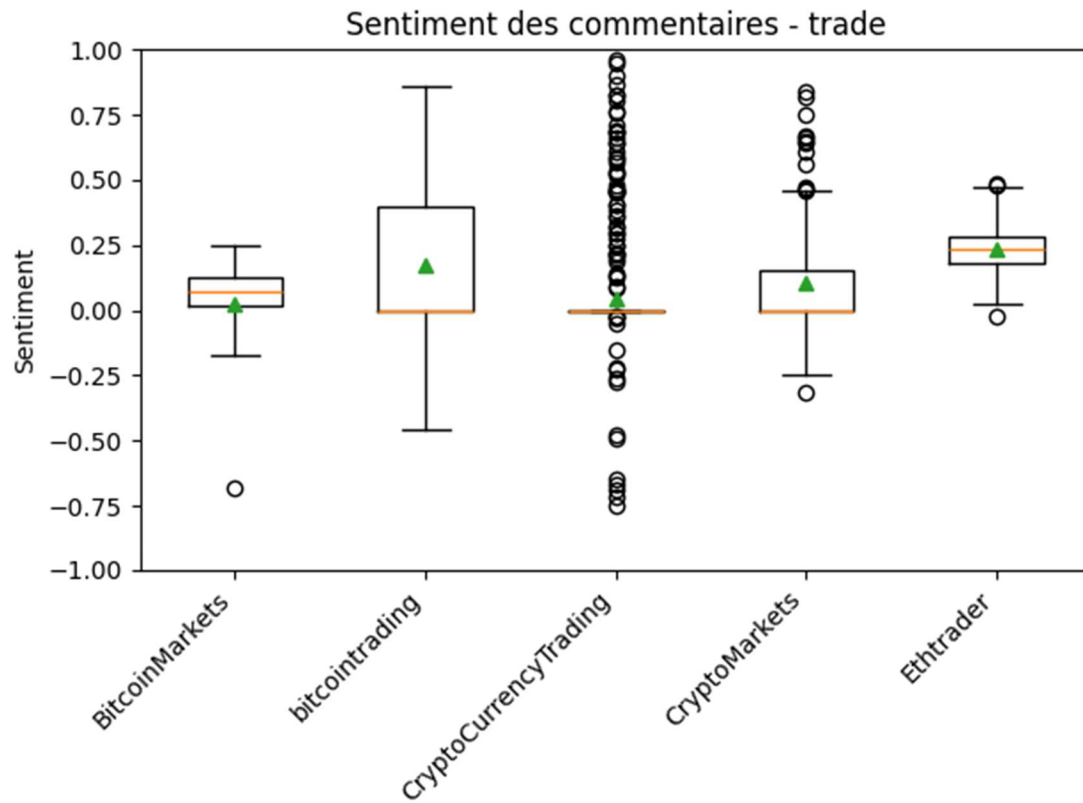


Figure 5 - Diagramme en boîte des com

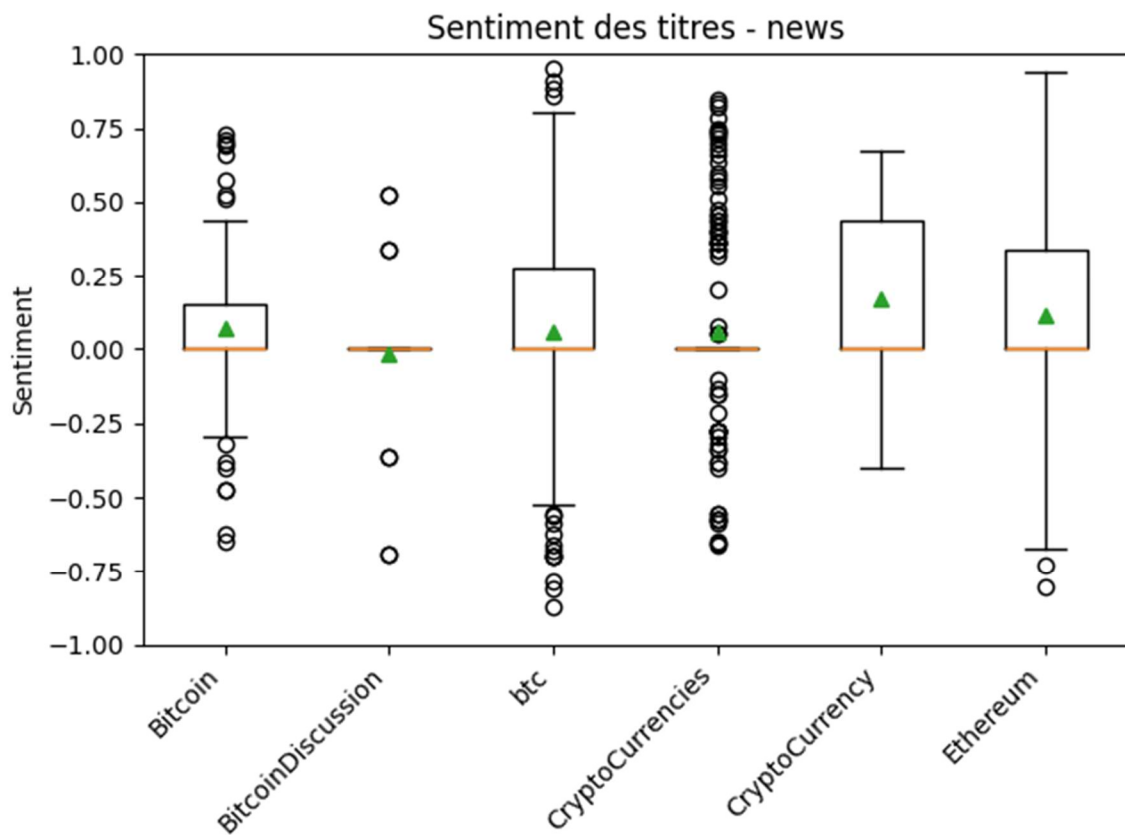


Figure 6 - Diagramme en boîte des titres dans la catégorie news

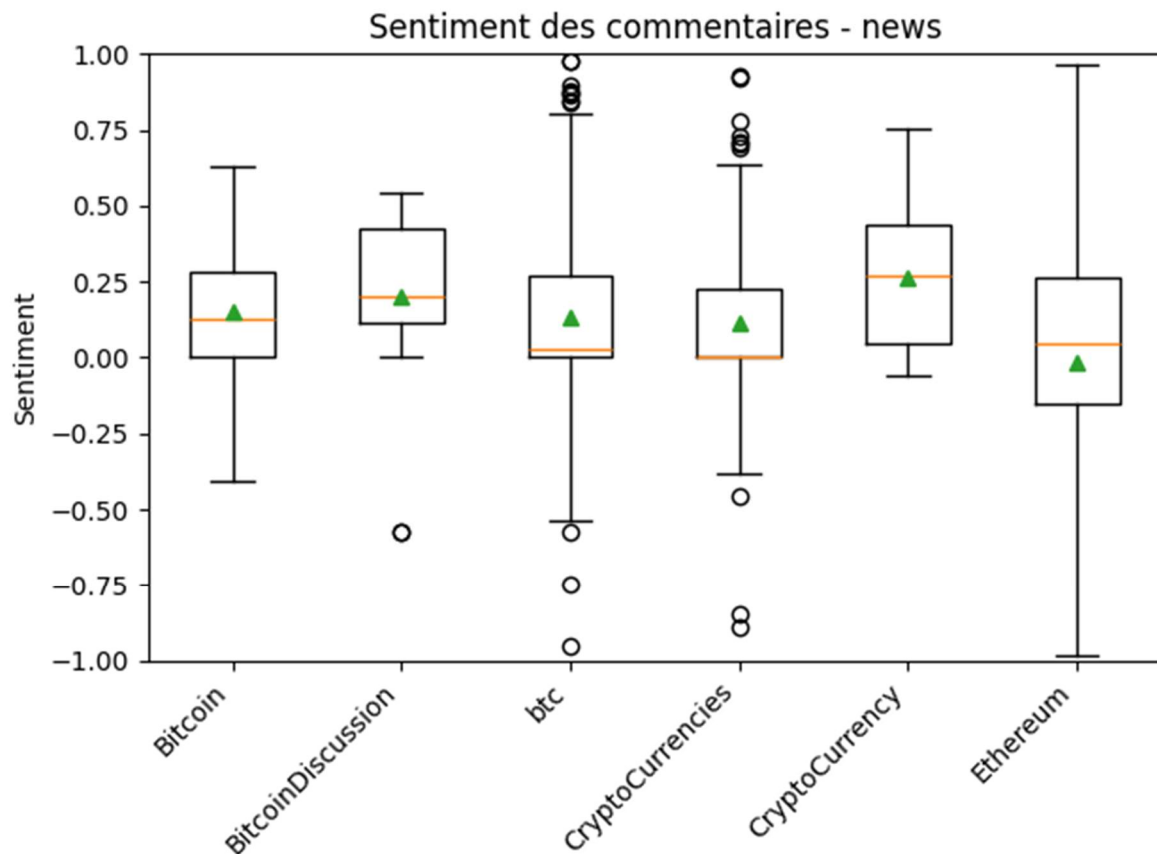


Figure 7 - Diagramme en boîte des commentaires dans la catégorie news

Les quatre diagrammes en boîte permettent d'analyser le sentiment des titres et des commentaires dans différents sous-forums cryptographiques, liés à l'actualité des cryptomonnaies (news) et au trading. Il est important de faire des liens entre ces résultats afin de tirer des conclusions plus approfondies sur la dynamique des discussions en ligne et leur interaction avec les mouvements du marché.

1. Liens entre titres et commentaires – forums de trade

Dans les forums liés au trade, les titres et les commentaires montrent des distributions de sentiment similaires, bien que les commentaires tendent à avoir une plus grande dispersion. Cela reflète une distinction importante entre la façon dont les utilisateurs abordent les sujets en fonction du contexte :

Les titres sont plus neutres et concis, souvent utilisés pour résumer une information ou introduire une idée de manière plus factuelle, ce qui explique leur médiation autour d'une zone positive dans la majorité des sous-forums. La volatilité moindre des sentiments dans les titres pourrait s'expliquer par leur rôle informatif, avec une tendance à éviter les émotions extrêmes.

Les commentaires, en revanche, sont plus polarisés. Par exemple, dans les forums tels que CryptoCurrencyTrading et CryptoMarkets, on note des valeurs aberrantes plus marquées dans les sentiments négatifs. Cela est probablement dû à la nature interactive des commentaires, où les

utilisateurs expriment des opinions plus tranchées, répondent aux autres et débattent. Les forums de trading sont des lieux où l'incertitude et les avis divergents sur les stratégies d'investissement mènent à une diversité émotionnelle plus marquée.

Ce lien entre titres et commentaires peut s'expliquer par l'idée que les discussions plus émotionnelles se développent souvent en réponse à des titres plus modérés. Cela indique que même si le sentiment global est positif, il existe une diversité d'opinions qui s'exprime surtout dans les échanges plus approfondis, comme les commentaires.

2. Liens entre trade et news – titres et commentaires

Les sous-forums axés sur le trade et ceux liés aux news montrent des différences intéressantes dans les sentiments exprimés, à la fois dans les titres et les commentaires :

Dans les forums de news (Bitcoin, BitcoinDiscussion, CryptoCurrencies, etc.), les titres montrent également des médianes positives, mais avec une plus grande dispersion que dans les forums de trade, en particulier pour des sous-forums comme CryptoCurrencies et Ethereum. Cela indique que les actualités captent une plus grande variété d'opinions, notamment à cause de la nature de l'information transmise, qui peut inclure des nouvelles positives (hausse de prix, nouvelle technologie) mais aussi des informations négatives (régulations, piratages, baisse de prix, etc...).

Les commentaires dans les forums de news montrent également une dispersion importante, mais l'optimisme est plus prononcé, avec des médianes positives plus marquées, en particulier pour BitcoinDiscussion et CryptoCurrency. Cela peut s'expliquer par le fait que les nouvelles informations, surtout dans un marché haussier comme en 2024, tendent à engendrer des discussions optimistes, surtout lorsque les actualités confirment ou renforcent des perspectives haussières pour les principales cryptomonnaies.

Dans l'ensemble, les sous-forums de news et de trade partagent un sentiment positif similaire dans les titres et commentaires. Cependant, les discussions sur les news sont légèrement plus polarisées et offrent une plus grande variété d'émotions, car les nouvelles informations affectent directement la perception des utilisateurs sur l'avenir des cryptomonnaies.

3. Corrélation avec la performance du Bitcoin en 2024

Ces résultats peuvent également être fortement corrélés à la performance du Bitcoin en 2024, où son prix n'a jamais baissé en dessous de son niveau de janvier et a même atteint un nouvel ATH (All-Time High). Cette croissance constante et la confiance accrue dans le marché des cryptomonnaies influencent directement les sentiments dans les sous-forums. Ci-dessous la courbe du Bitcoin sur l'année 2024 disponible sur CoinMarketCap.



Figure 8 - Prix du bitcoin en 2024

L'optimisme marqué dans les titres et commentaires peut être perçu comme un réflexe naturel face à un marché haussier. Quand le prix du Bitcoin augmente, les discussions sont naturellement biaisées vers un sentiment positif, avec une majorité des participants exprimant des points de vue optimistes.

Les forums comme CryptoMarkets et CryptoCurrencyTrading, bien qu'ils affichent un grand nombre de valeurs aberrantes, montrent globalement des médianes positives dans les deux cas. Cette polarisation pourrait être le reflet de stratégies d'investissement différentes : les investisseurs à court terme ou ceux cherchant à réaliser des profits rapides peuvent être plus pessimistes dans leurs commentaires, en réponse à une volatilité du marché ou des mouvements de prix spécifiques. En revanche, ceux qui ont une vision plus à long terme (particulièrement dans les forums comme Ethereum) manifestent un optimisme plus stable.

Cette diversité des opinions, bien qu'imbibée d'optimisme général, reflète la complexité des marchés cryptographiques en 2024, où malgré des gains globaux, les opinions divergent en fonction des stratégies et des préférences des utilisateurs.

Pour conclure, les quatre diagrammes montrent une corrélation positive entre le sentiment général des utilisateurs des forums cryptographiques et la performance haussière du marché des cryptomonnaies en 2024. Bien que les titres soient généralement plus modérés, les commentaires révèlent une polarisation plus marquée, avec des valeurs aberrantes qui expriment à la fois des opinions très positives et négatives. Cependant, dans tous les forums, que ce soit pour le trade ou les news, le sentiment médian reste optimiste.

Ce biais vers l'optimisme est probablement soutenu par la montée du prix du Bitcoin, qui influence la perception des utilisateurs sur les forums. Néanmoins, la diversité des émotions exprimées,

notamment dans les commentaires, montre que même dans un marché haussier, les utilisateurs réagissent différemment aux mouvements de prix en fonction de leurs stratégies d'investissement et des nouvelles qu'ils jugent importantes.

Ainsi, les forums cryptographiques sont un baromètre intéressant de la psychologie des investisseurs. Leur interaction avec les événements du marché, les tendances de prix, et les informations ponctuelles (comme les nouvelles régulations ou avancées technologiques) donne une idée claire de l'état d'esprit global des investisseurs, tout en soulignant l'importance des discussions en ligne comme facteur influençant et reflétant les dynamiques du marché.

Cependant, cette interprétation doit être nuancée par plusieurs limites. Le diagramme ne fournit pas de détails sur les raisons spécifiques derrière les opinions positives ou négatives.

- **Manque de contexte** : Les graphiques montrent la répartition des sentiments mais pas les événements ou sujets qui les déclenchent (régulation, technologies, etc.).
- **Biais d'optimisme** : Les subreddits analysés attirent majoritairement des passionnés de cryptomonnaies, ce qui peut créer un biais vers des sentiments excessivement positifs, limitant ainsi la généralisation des résultats.
- **Influence des bots** : Les bots présents sur ces plateformes peuvent fausser les données en publiant massivement des contenus, amplifiant artificiellement des sentiments sans réel fondement humain.
- **Réactivité excessive à l'actualité** : Les discussions sur Reddit réagissent souvent rapidement à des événements récents, créant des pics de sentiments qui peuvent ne pas refléter des tendances de long terme.
- **Absence de distinction entre participants** : Les sentiments sont agrégés sans distinguer entre investisseurs actifs, observateurs ou bots, rendant difficile une évaluation précise des opinions réelles..

Ci-dessous les deux nuages de mot générés.



Nuage des mots pour la catégorie news :

Dans la catégorie "news", qui regroupe les discussions et nouvelles liées aux cryptomonnaies, les mots les plus dominants sont **"bitcoin," "fake," "money," "people,"** et **"like"**. Cela reflète une forte focalisation sur les aspects fondamentaux des cryptomonnaies, en particulier le **Bitcoin**, tout en abordant des sujets de **crédibilité** et de **fraude** (comme le suggère l'occurrence fréquente du mot **"fake"**). Des mots tels que **"ethereum," "btc,"** et **"crypto"** montrent que les discussions s'étendent aussi à d'autres cryptomonnaies majeures et aux marchés en général.

Le vocabulaire associé, comme **"price," "wallet,"** et **"scams,"** indique que les conversations dans ces forums ne sont pas uniquement techniques mais aussi axées sur la sécurité, les fluctuations de prix, et les arnaques potentielles, qui sont des sujets communs dans les forums de discussion sur les cryptomonnaies. Le mot **"would"** reflète aussi une certaine spéculation ou hypothèse commune dans les débats concernant l'avenir des cryptomonnaies.

Nuage des mots pour la catégorie trading :

Dans le domaine du trading, les termes les plus visibles sont très différents. On observe des mots comme **"take," "weight," "donut,"** et **"content,"** qui sont bien moins directement liés aux cryptomonnaies que ceux observés dans la catégorie news. Le mot **"donut"** et des termes techniques comme **"tipped," "minutes"** ou **"learn"** semblent indiquer que les utilisateurs dans ces forums se concentrent davantage sur les outils et mécanismes spécifiques au trading. Cela pourrait aussi suggérer l'utilisation d'un système de **récompense** ou d'outils spécifiques au subreddit qui influencent les discussions.

Les discussions sur le trading montrent donc un aspect plus méthodique et technique, avec un vocabulaire lié à la stratégie ou à des concepts financiers tels que le **"weight"** (peut-être lié à l'allocation de ressources ou de portefeuille) et des échanges de **contenu** technique ou de conseils de trading. Le terme **"learn"** reflète également l'idée que les forums de trading sont utilisés pour l'apprentissage et l'amélioration des compétences des participants.

Pour conclure, les deux nuages de mots montrent clairement des différences dans l'orientation des discussions. La catégorie news est centrée sur la discussion générale des cryptomonnaies, avec des préoccupations concernant leur fiabilité, leur valeur, et les scams dans le marché, tandis que la catégorie trading adopte une approche plus technique et pragmatique, mettant l'accent sur les stratégies de trading, l'utilisation d'outils spécifiques et l'éducation financière des membres.

Ces analyses montrent que les forums "news" sont davantage tournés vers des débats et spéculations globales sur les cryptomonnaies, tandis que ceux axés sur le "trading" mettent en lumière une communauté plus focalisée sur les actions spécifiques, l'exécution d'opérations financières, et le partage de connaissances pratiques. Il serait pertinent d'explorer ces nuances plus en profondeur pour mieux comprendre les besoins et les attentes des communautés dans ces sous-forums.

IV.2 Analyse avec des sources externes

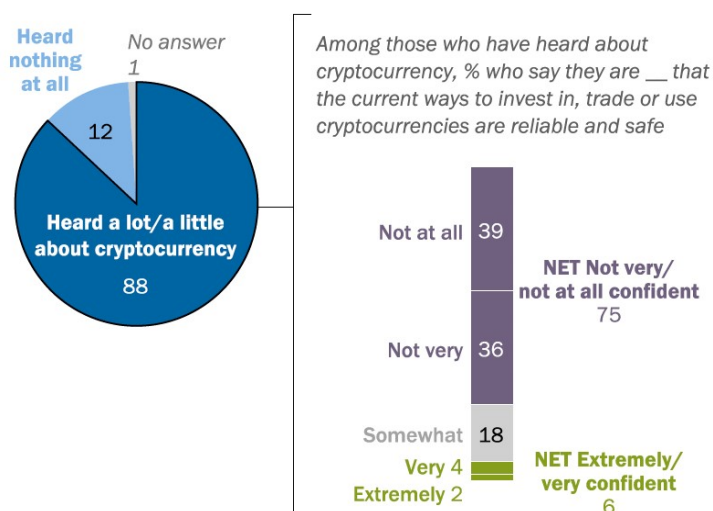
IV.2.1 Analyse du contexte socio-démographique

Nous allons d'abord réaliser une analyse socio-démographique de nos utilisateurs, en commençant par ceux qui utilisent les cryptomonnaies. Selon une étude de Michelle Faverio et Olivia Sidoti intitulée « Majority of Americans aren't confident in the safety and reliability of cryptocurrency »⁵, il apparaît que certains groupes de la population sont plus enclins à parler ou à acheter des cryptomonnaies, tandis que d'autres restent réticents, principalement en raison des préoccupations liées à la sécurité et à la fiabilité.

En parallèle, nous mettrons ces données en corrélation avec celles des utilisateurs de Reddit, telles que décrites dans un article publié par Exploding Topics⁶ en avril 2024. Ce dernier indique que Reddit comptait plus de 500 millions d'utilisateurs en 2024, avec une projection de 550 millions pour 2028. Cette analyse nous permettra d'examiner comment les utilisateurs de Reddit se positionnent par rapport aux cryptomonnaies.

75% of Americans who have heard of cryptocurrencies are not confident in their safety and reliability

% of U.S. adults who say they have ___ about cryptocurrency such as bitcoin or ether



Note: Heard a lot/a little response options are combined. Figures may not add up to the NET values or 100% due to rounding. Those who did not give an answer are not shown.
Source: Survey of U.S. adults conducted March 13-19, 2023.

PEW RESEARCH CENTER

D'après cette figure ci-dessus, les Américains se montrent généralement méfiants à l'égard des cryptomonnaies, en particulier sur les questions de sécurité et de fiabilité en 2023. Cela contraste

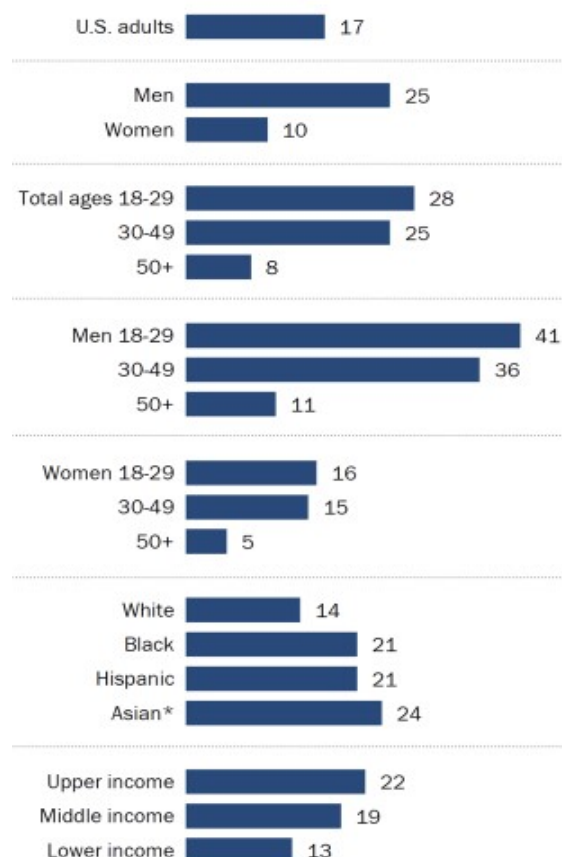
⁵ Michelle Faverio, Olivia Sidoti. Majority of Americans aren't confident in the safety and reliability of cryptocurrency In: Pew Research Center [en ligne] 10/04/2023 disponible sur : <https://www.pewresearch.org/short-reads/2023/04/10/majority-of-americans-arent-confident-in-the-safety-and-reliability-of-cryptocurrency/-reliability-of-cryptocurrency/> (consulté le 01/08/2024)

⁶ Fabio Duarte. Reddit User Age, Gender, & Demographics (2024) In : Exploding Topics [en ligne] 24/04/2024 disponible sur : <https://explodingtopics.com/blog/reddit-users> (consulté le 01/08/2024)

toutefois avec les résultats obtenus sur Reddit, où les discussions sur les cryptomonnaies sont largement positives. Cette différence peut s'expliquer par le fait que les données Reddit sont principalement issues de subreddits spécialisés, attirant un public averti et déjà familiarisé avec ce type d'investissement.

Men under 50 stand out for investing in, trading or using a cryptocurrency

% of U.S. adults who say they have ever invested in, traded or used a cryptocurrency such as bitcoin or ether



* Estimates for Asian adults are representative of English speakers only.
 Note: Family income tiers are based on adjusted 2021 earnings. White, Black and Asian adults include those who report being only one race and are not Hispanic. Hispanic adults are of any race. Those who did not give an answer are not shown.
 Source: Survey of U.S. adults conducted March 13-19, 2023.

PEW RESEARCH CENTER

Les utilisateurs de Reddit partagent des similitudes avec ceux des cryptomonnaies en termes d'âge. La majorité des utilisateurs se situent dans la tranche des 18-29 ans, suivie de près par les 30-49 ans, ce qui est cohérent avec le profil type des détenteurs de cryptomonnaies.

Quant aux utilisateurs de reddit :

Age Group	Proportion of Americans Using Reddit
18 to 29 year olds	44%
30 to 49 year olds	31%
50 to 64 year olds	11%
65+ year olds	3%

Figure 11 - Proportion d'utilisateur de Reddit par âge

Gender	Distribution of Reddit Users
Female	35.1%
Male	63.6%

Figure 12 - Proportion d'utilisateur par genre

En ce qui concerne la répartition par genre, 35,1 % des utilisateurs de Reddit sont des femmes, contre 63,6 % d'hommes. Ces proportions reflètent également les résultats de l'étude sur les cryptomonnaies, où les hommes représentent une majorité.

Il est important de souligner que près de 50 % des utilisateurs de Reddit sont Américains, avec une plus grande proportion d'hommes américains (27 %) que de femmes américaines (17 %). Malheureusement, nous ne disposons pas de données détaillées sur l'ethnicité des utilisateurs de Reddit, ce qui aurait permis une comparaison plus complète avec les études démographiques sur les cryptomonnaies.

D'autre étude on était réalisé courant 2024 comme « 2024 Cryptocurrency Adoption and Sentiment Report »⁷ de Tom Blackstone

La sensibilisation et les taux de détention de cryptomonnaies ont atteint des niveaux record : 40 % des adultes américains possèdent désormais des cryptomonnaies, contre 30 % en 2023. Cela représente potentiellement jusqu'à 93 millions de personnes.

Parmi les détenteurs actuels de cryptomonnaies, environ 63 % prévoient d'acquérir davantage de cryptomonnaies au cours de l'année prochaine. Les devises les plus prisées sont le Bitcoin, l'Ethereum, le Dogecoin et le Cardano.

Le taux de détention de cryptomonnaies chez les femmes a connu une augmentation significative, passant de 18 % il y a un an à 29 % au début de l'année 2024. Coïncidant encore plus avec le nombre de femme présentent sur reddit.

⁷ Tom Blackstone. 2024 Cryptocurrency Adoption and Sentiment Report In security.org [en ligne] disponible sur : <https://www.security.org/digital-security/cryptocurrency-annual-consumer-report/#landscape> (consulté le 01/08/2024)

Parmi les non-détenteurs, 21 % ont déclaré que l'introduction anticipée d'un ETF Bitcoin les rend plus susceptibles d'investir dans les cryptomonnaies. Cela signifie que jusqu'à 29 millions d'Américains supplémentaires pourraient bientôt entrer sur le marché.

En outre, 46 % des Américains estiment que l'approbation des ETF Bitcoin en 2024 aura un impact positif sur l'industrie de la blockchain. Les détenteurs actuels de cryptomonnaies sont encore plus optimistes à ce sujet.

En résumé, les données montrent une adoption croissante des cryptomonnaies aux États-Unis, malgré la persistance de certaines craintes quant à leur sécurité. Les similitudes démographiques entre les utilisateurs de cryptomonnaies et ceux de Reddit, notamment en termes d'âge et de genre, confirment que Reddit constitue une plateforme clé pour échanger et promouvoir l'utilisation des cryptomonnaies. L'évolution notable de la participation féminine, couplée à l'anticipation des ETF Bitcoin, suggère que le marché pourrait connaître une nouvelle vague de croissance dans les prochaines années, renforçant ainsi l'impact des cryptomonnaies dans le paysage financier mondial.

IV.2.2 Avis médias sur le courant 2024

Désormais voici l'avis de certains médias en 2024, allant dans le même sens que mes recherches et ceux des articles scientifiques.

Le rapport du Journalducoin⁸ sur l'adoption des cryptomonnaies en 2024 révèle une croissance rapide, avec 562 millions de détenteurs, soit 6,8 % de la population mondiale, une augmentation de 33 % par rapport à 2023. L'Asie domine avec plus de la moitié des utilisateurs mondiaux, tandis que l'Amérique du Sud, l'Océanie et l'Europe enregistrent les plus fortes progressions.

Les cryptos attirent surtout les 25-44 ans (65 %) et sont majoritairement détenues par des hommes (61 %). L'adoption croissante des cryptos influence les méthodes de paiement, et de plus en plus d'entreprises les acceptent. En parallèle, l'inflation et les régulations plus claires encouragent cette adoption, tandis que les ETF Bitcoin et le halving de 2024 stimulent l'intérêt des investisseurs.

L'article de CoinDesk⁹ met en lumière un sondage montrant que bien que 21 % des électeurs des États décisifs aux États-Unis considèrent les politiques liées aux cryptomonnaies comme un sujet influençant leurs votes, la majorité reste méfiante vis-à-vis des actifs numériques. Seuls 31 % des personnes interrogées ont une opinion favorable à la cryptomonnaies et 14 % en possèdent déjà.

L'industrie crypto tente de capitaliser sur cet intérêt en injectant des fonds dans les campagnes, mais la division entre les électeurs pro et contre la cryptomonnaies pourrait compliquer la position

⁸ Kevin C. Bilan de l'adoption des cryptos dans le monde en 2024 – Rapport Triple-A In Journal du Coin [en ligne] disponible sur : <https://journalducoin.com/actualites/bilan-adoption-cryptos-monde-2024-rapport-triple-a/> (consulté le 01/08/2024)

⁹ Jesse Hamilton. Poll: Most People Cringe About Crypto, But Enough Care to Warrant Politicians' Attention In : CoinDesk [en ligne] disponible sur : <https://www.coindesk.com/policy/2024/05/07/poll-most-people-tinge-about-crypto-but-enough-care-to-warrant-politicians-attention/> (consulté le 01/08/2024)

des candidats politiques. Bien que beaucoup trouvent le système financier actuel dépassé (70 %), ils ne font pas totalement confiance aux autorités pour élaborer des politiques efficaces sur les cryptos.

L'étude montre aussi que les détenteurs de cryptomonnaies pourraient représenter 47 millions d'Américains, mais peu de non-investisseurs (7 %) envisagent d'acheter de la cryptomonnaie sans davantage de garanties, notamment de grandes institutions financières ou de la réglementation gouvernementale.

L'article du Daily Bruin¹⁰ pour finir souligne les risques liés à l'essor des cryptomonnaies, malgré leur adoption croissante, comme par l'Université de Californie qui accepte désormais des dons en crypto. L'arrestation de Sam Bankman-Fried, PDG de FTX, pour fraude en 2022, a accru la méfiance envers ces actifs. Bitcoin, comme d'autres cryptomonnaies, n'est pas régulé et sa valeur dépend entièrement de la demande du marché, ce qui le rend extrêmement volatile – avec des baisses allant jusqu'à 30 % en une seule journée.

Cette volatilité, combinée au manque de réglementation, rend les cryptos risqués pour un usage quotidien. Les experts soulignent que, malgré la technologie prometteuse de la blockchain, des problèmes de confiance persistent, exacerbés par le fait que peu de personnes maîtrisent véritablement ce domaine.

Ces résultats viennent renforcer les tendances observées dans mes analyses de sentiments sur les forums Reddit. L'optimisme affiché par les utilisateurs semble en phase avec la croissance de l'adoption des cryptomonnaies en 2024, particulièrement stimulée par des facteurs tels que les ETF Bitcoin et le halving. Cependant, les préoccupations liées à la réglementation et à la méfiance vis-à-vis des actifs numériques, comme soulignées dans les sondages et articles, sont également présentes dans les discussions sur Reddit, rappelant que l'enthousiasme s'accompagne toujours de prudence, voire de scepticisme, pour certains investisseurs et utilisateurs.

¹⁰ Chiara Grasso. Opinion: Cryptocurrency's risks prove too high to fully trust it as a form of banking In Daily Bruin [en ligne] disponible sur : <https://dailybruin.com/2024/02/08/opinion-cryptocurrencys-risks-prove-too-high-to-fully-trust-it-as-a-form-of-banking> (consulté le 01/08/2024)

IV.2.3 Comparaison des résultats obtenus sur twitter

Pour finir je me suis basé énormément sur l'article scientifique de IEEE, s'intitulant « Sentiment Analysis of Bitcoin data by tweets through Naive Bayes »¹¹ en 2022 qui a fait le même travail mais sur twitter en utilisant le système de classification de Naive Bayes. Qui bien qu'en ayant utilisé une méthode différente, avait le même but que cette recherche sur reddit. Bien que son étude se concentrait sur la fluctuation du prix du bitcoin en fonction des sentiments des utilisateurs en arrivant à ce schéma de conclusion :

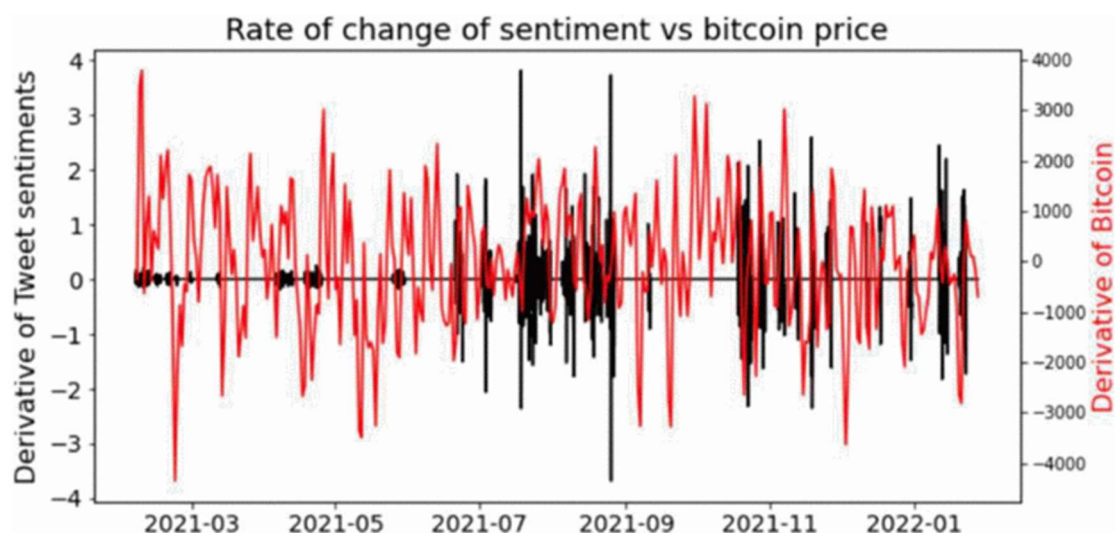


Figure 13 - Comparaison de tendance

En arrivant à la conclusion suivante d'après ses dires : « On peut conclure que la majorité des tweets appartiennent à la catégorie neutre sans montrer d'acceptation ou de rejet total à l'égard de la pièce, ce qui peut être justifié compte tenu de la nature hautement volatile de la pièce. ».

Pour mettre en lien avec notre analyse, qui est globalement positif et non neutre comme celle réalisée sur twitter mais qui dans notre cas, le bitcoin a atteint son ATH (all time high) cette année et est resté à un prix élevé comme nous le montre le schéma ci-dessous sans jamais redescendre plus bas que lors du début 2024 à 45 000\$, ce qui peut expliquer les avis positifs et enthousiastes que ce soit pour les news ou les traders.

Il est également pertinent de souligner que l'usage de différentes plateformes peut influencer la nature des sentiments exprimés. Par exemple, Reddit est souvent utilisé par des communautés plus spécialisées et engagées, où les utilisateurs échangent des opinions détaillées et techniques, alors que Twitter, du fait de son format plus concis, favorise des réactions instantanées et parfois plus neutres. Cela pourrait expliquer pourquoi les résultats sur Reddit tendent à être plus positifs que sur Twitter.

Enfin, il est essentiel de rappeler que bien que les résultats Reddit montrent un enthousiasme marqué, il convient d'intégrer cette tendance dans une perspective globale qui prend en compte la volatilité inhérente au marché du Bitcoin.

¹¹ IEEE. Sentiment Analysis of Bitcoin data by tweets through Naive Bayes In IEEEExplore [en ligne] disponible sur : <https://ieeexplore-ieee-org.ezproxy.utbm.fr/document/9988751> (consulté le 01/08/2024)

V. Conclusion

Cette étude sur les avis des utilisateurs de cryptomonnaie sur Reddit en 2024 nous a permis d'explorer en profondeur les perceptions et les sentiments exprimés par une communauté active et engagée. En scrappant plusieurs subreddits liés aux cryptomonnaies, nous avons pu analyser les posts et commentaires de manière quantitative et qualitative, offrant ainsi un aperçu précis des opinions actuelles autour de cette cryptomonnaie.

Les résultats montrent une tendance globalement positive, notamment en raison de la hausse historique du prix du Bitcoin en 2024. Contrairement aux études précédentes, notamment celle menée sur Twitter en 2022, qui mettaient en avant une attitude plus neutre, nos données révèlent un enthousiasme marqué, probablement influencé par des facteurs comme l'adoption croissante des cryptomonnaies, la clarté accrue des régulations, et l'intérêt des investisseurs institutionnels avec l'arrivée des ETF Bitcoin.

Néanmoins, cette analyse doit être nuancée. Si Reddit reflète un optimisme certain, il convient de souligner que d'autres plateformes, telles que Twitter, ou des revues de presse, montrent un sentiment plus mitigé. De plus, certains biais, notamment liés à la présence de bots ou à l'engagement des utilisateurs les plus actifs, peuvent influencer sur la nature des discussions. Ces limitations doivent être prises en compte pour éviter une surinterprétation des résultats.

En conclusion, l'étude des sentiments sur Reddit apporte une perspective unique sur les tendances et opinions concernant le Bitcoin en 2024. Toutefois, pour avoir une vision plus globale et équilibrée, il est essentiel de compléter ces analyses avec des recherches sur d'autres plateformes et de prendre en compte des éléments contextuels, tels que l'évolution des régulations ou les événements macroéconomiques.

Cette analyse, bien qu'encourageante, montre que les perceptions restent volatiles et sensibles aux fluctuations des prix, mettant en lumière l'importance d'une régulation claire et d'une meilleure compréhension des cryptomonnaies par le grand public pour assurer une adoption durable.

VI. Bibliographie

Chiara Grasso. Opinion: Cryptocurrency's risks prove too high to fully trust it as a form of banking In Daily Bruin [en ligne] disponible sur : <https://dailybruin.com/2024/02/08/opinion-cryptocurrencys-risks-prove-too-high-to-fully-trust-it-as-a-form-of-banking> (consulté le 01/08/2024)

Fabio Duarte. Reddit User Age, Gender, & Demographics (2024) In : Exploding Topics [en ligne] 24/04/2024 disponible sur : <https://explodingtopics.com/blog/reddit-users> (consulté le 01/08/2024)

IEEE. Sentiment Analysis of Bitcoin data by tweets through Naive Bayes In IEEEExplore [en ligne] disponible sur: <https://ieeexplore-ieee-org.ezproxy.utbm.fr/document/9988751> (consulté le 01/08/2024)

Jesse Hamilton. Poll: Most People Cringe About Crypto, But Enough Care to Warrant Politicians' Attention In : CoinDesk [en ligne] disponible sur : <https://www.coindesk.com/policy/2024/05/07/poll-most-people-tinge-about-crypto-but-enough-care-to-warrant-politicians-attention/> (consulté le 01/08/2024)

Kevin C. Bilan de l'adoption des cryptos dans le monde en 2024 – Rapport Triple-A In Journal du Coin [en ligne] disponible sur : <https://journalducoin.com/actualites/bilan-adoption-cryptos-monde-2024-rapport-triple-a/> (consulté le 01/08/2024)

Michelle Faverio, Olivia Sidoti. Majority of Americans aren't confident in the safety and reliability of cryptocurrency In: Pew Research Center [en ligne] 10/04/2023 disponible sur : <https://www.pewresearch.org/short-reads/2023/04/10/majority-of-americans-arent-confident-in-the-safety-and-reliability-of-cryptocurrency/-reliability-of-cryptocurrency/> (consulté le 01/08/2024)

Tom Blackstone. 2024 Cryptocurrency Adoption and Sentiment Report In security.org [en ligne] disponible sur : <https://www.security.org/digital-security/cryptocurrency-annual-consumer-report/#landscape> (consulté le 01/08/2024)