

**AYRAULT MARIUS**

**Année scolaire 2024 - 2025**

**Numéro Étudiant : 12326806**

**SORBONNE DATA ANALYTICS**

Promotion Mars 2025

## **RAPPORT DE FIN D'ETUDES**

**NOM DE LA STRUCTURE D'ACCUEIL : N/A**

**Analyse des schémas de comportement des utilisateurs  
dans les applications DeFI (Finance Décentralisée).**

**Maître de stage : N/A**

**Unité d'accueil – lieu de stage**

**Ville : N/A**

**Pays : N/A**

**Tuteur pédagogique : N/A**

# Remerciements

Tout d'abord, je tiens à remercier l'université *Paris 1 Panthéon Sorbonne* pour m'avoir accueilli au sein de la formation *Sorbonne Data Analytics* dans le cadre de mon évolution professionnelle.

Je tiens à remercier Mr DIAYE Marc-Arthur, pour nous avoir permis de réaliser cette formation dans des conditions idéales, pour s'être rendu disponible lorsque nécessaire et pour nous avoir transmis ses connaissances associées au domaine complexe que sont les statistiques.

Je tiens à remercier Mr KRASNIQI Dafnis, pour nous avoir permis en toute simplicité, avec passion et enthousiasme, de mettre en œuvre des technologies complexes que sont les réseaux de neurones et les modèles de langages. Je tiens également à le remercier particulièrement pour m'avoir supporté lors de la sélection du sujet du présent mémoire et pour nous avoir présentés les tenants et les aboutissants de l'exercice du métier en tant qu'indépendant.

Je tiens à remercier Mr BOGROFF Alexis, pour nous avoir partagé son expérience et nous avoir transmis une vision analytique ainsi qu'une exigence essentielle à l'emploi de modèles d'apprentissages. Je tiens également à le remercier particulièrement pour ses retours constructifs à l'issue de différents travaux.

Je tiens à remercier Mr GRAND Benoit, pour nous avoir transmis les connaissances nécessaires à la réalisation de la formation que sont l'apprentissage du langage Python et l'apprentissage du langage SQL.

Je tiens à remercier Mr DIATTARA Ibrahim, pour nous avoir transmis de nombreuses connaissances associées aux technologies cloud et aux technologies open-source essentielles dans le domaine. Je tiens également à le remercier pour sa passion et sa joie transmise durant les enseignements.

Je tiens à remercier Mme MOUMMOU Nisrine, pour nous avoir fait découvrir le développement d'applications ainsi que les connaissances nécessaires à la mise en œuvre de méthodes telles que le traitement de données, l'extraction de texte ou la data visualisation.

Je tiens à remercier Mr CISSE Ansoumana, pour nous avoir fait découvrir les bases de données orientées graphe ainsi que le langage Cypher permettant la réalisation d'un système de recommandation complexe.

Je tiens à remercier Mr TAHIROU Ibrahim, pour nous avoir transmis les connaissances nécessaires à l'utilisation de Microsoft Azure et pour s'être rendu disponible pour nous aider à utiliser une machine virtuelle durant le projet.

Je tiens à remercier Mme FAIK Lina, pour nous avoir transmis de nombreuses connaissances liées aux modèles d'apprentissages avec des supports remarquablement bien organisés ainsi que pour nous avoir permis de sélectionner notre propre sujet de projet.

Je tiens à remercier Mr KARE Kamila, pour nous avoir transmis les bonnes pratiques de développement et de mise en production d'un modèle d'apprentissage avec une efficacité remarquable.

# Tables des matières

<b>1. INTRODUCTION GENERALE .....</b>	<b>1</b>
1.1 CONTEXTE ET IMPORTANCE DE LA FINANCE DECENTRALISEE (DEFI) .....	1
1.1.1 Définition de la finance décentralisée .....	1
1.1.2 Évolution de la finance décentralisée .....	2
1.1.3 Importance économique et technologique .....	2
1.2 OBJECTIFS DE L'ÉTUDE .....	2
1.3 PROBLEMATIQUE .....	3
1.4 METHODOLOGIE .....	3
1.5 STRUCTURE .....	4
<b>2. CONTEXTE THEORIQUE ET CADRE CONCEPTUEL .....</b>	<b>4</b>
2.1 FONDEMENTS DE LA FINANCE DECENTRALISEE .....	4
2.2 ÉCOSYSTEME DEFI .....	5
2.3 THEORIES COMPORTEMENTALES .....	5
2.3.1 Théorie de la diffusion des innovations .....	5
2.3.2 Théorie des réseaux et des effets de réseau .....	6
2.4 CADRE CONCEPTUEL DE L'ÉTUDE .....	6
<b>3. REVUE DE LA LITTERATURE SUR LES COMPORTEMENTS DES UTILISATEURS DANS LA DEFI .....</b>	<b>7</b>
3.1 ÉTUDES PRECEDENTES LIEES AUX COMPORTEMENTS DES UTILISATEURS DANS LA DEFI .....	7
3.1.1 DeFi Survival Analysis : Insights into Risks and User Behaviors .....	7
3.1.2 Bridging the adoption gap for cryptocurrencies: understanding the affordances that impact approach-avoidance behavior for potential users and continuation usage for actual users .....	7
3.1.3 User behavior and token adoption on ERC20 .....	7
3.1.4 Behavior-aware Account De-anonymization Ethereum Interaction Graph .....	7
3.1.5 Onchain Analysis : A Comparative Study of Decentralized Exchange (DEX) Activities on Ethereum, Solana, and Binance Smart Chain (BSC) .....	8
3.2 FACTEURS INFLUENÇANT LES COMPORTEMENTS DES UTILISATEURS .....	8
3.2.1 Facteurs psychologiques .....	8
3.2.2 Facteurs économiques .....	8
3.2.3 Facteurs technologiques .....	8
3.3 LACUNES DANS LA LITTERATURE .....	9
3.3.1 Limites méthodologiques des études existantes .....	9
3.3.2 Étude proposée .....	9
3.3.3 Justification de l'étude .....	9
<b>4. METHODOLOGIE DE RECHERCHE .....</b>	<b>9</b>
4.1 APPROCHE METHODOLOGIQUE .....	9
4.2 COLLECTE DE DONNEES .....	10
4.2.1 Définition du périmètre de l'étude .....	10
4.2.2 Constitution du jeu de données .....	10
4.3 TRAITEMENT DES DONNEES COLLECTEES .....	11
4.3.1 Analyse exploratoire .....	11
4.3.1.1 Analyse de corrélation .....	11
4.3.1.2 Analyse des tendances .....	12
4.3.1.3 Analyse des effets de réseau .....	13
4.3.2 Ingénierie des caractéristiques .....	14
4.3.3 Standardisation .....	15
4.3.4 Réduction de la dimensionnalité .....	16
4.4 APPRENTISSAGE NON-SUPERVISE .....	17
4.4.1 Sélection d'un algorithme de clustering .....	17
4.4.2 Optimisation des paramètres d'entraînement .....	18
4.4.2.1 Métriques de performances .....	18

4.4.2.2 Détermination du nombre de cluster .....	18
4.4.2.3 Fonction d'optimisation des paramètres.....	19
4.4.3 Sélection du modèle d'apprentissage.....	19
4.5 EXPLOITATION DU MODELE D'APPRENTISSAGE.....	20
4.5.1 Répartition des utilisateurs par cluster .....	20
4.5.2 Différenciation des utilisateurs par usages .....	20
4.5.3 Différenciation des utilisateurs par comportements.....	21
4.5.4 Caractéristiques principales de distinction.....	22
<b>5. INTERPRETATION DES RESULTATS .....</b>	<b>22</b>
5.1 VALIDITE ET FIABILITE .....	22
5.1.1 Mesures de validité .....	22
5.1.2 Mesures de fiabilité .....	23
5.2 LIMITES METHODOLOGIQUES .....	23
5.2.1 Limites identifiées.....	23
5.2.2 Stratégies d'atténuation.....	24
5.3 COMPARAISON AVEC LES HYPOTHESES .....	24
5.3.1 Première hypothèse.....	24
5.3.2 Seconde hypothèse.....	24
5.3.3 Troisième hypothèse.....	25
5.3.4 Quatrième hypothèse.....	26
5.4 IMPLICATIONS DES RESULTATS.....	26
5.4.1 Implications théoriques .....	26
5.4.2 Implications pratiques .....	27
<b>6. ENRICHISSEMENT DE L'ÉTUDE .....</b>	<b>27</b>
6.1 PERSPECTIVES D'APPROFONDISSEMENT .....	27
6.2 PERSPECTIVES TECHNOLOGIQUES.....	28
6.3 RECOMMANDATIONS POUR LES ACTEURS DE LA DEFI.....	28
6.3.1 Suggestion .....	28
6.3.2 Preuve de concept .....	28
6.3.2.1 Définition de métriques de comparaison .....	29
6.3.2.2 Définition d'un rang utilisateur par métrique.....	30
6.3.2.3 Visualisation des performances d'un utilisateur .....	30
6.3.2.4 Génération de recommandations .....	30
6.3.3 Résultats de la preuve de concept.....	31
<b>7. CONCLUSION GENERALE .....</b>	<b>31</b>
7.1 SYNTHÈSE DES PRINCIPAUX RESULTATS.....	31
7.2 REPONSES AUX QUESTIONS DE RECHERCHE .....	32
7.1.1 Première question de recherche .....	32
7.1.2 Seconde question de recherche .....	32
7.2 SYNTHÈSE DES CONTRIBUTIONS DE L'ÉTUDE .....	32
7.3 PERSPECTIVES .....	33
7.3.1 Avenir de la finance décentralisée.....	33
7.3.2 Rôle des comportements des utilisateurs .....	33
<b>8. CONCLUSION PERSONNELLE.....</b>	<b>34</b>
8.1 APPORTS DE LA FORMATION .....	34
8.2 MISE EN ŒUVRE DES COMPÉTENCES.....	34
8.3 PERSPECTIVES PROFESSIONNELLES .....	35

# Table des illustrations

<i>Figure 1: Mesures de corrélation entre le nombre d'utilisateurs et le nombre de transactions par heure.</i>	11
<i>Figure 2: Mesures de corrélation entre la valeur transférée et le nombre d'utilisateurs par heure.</i>	12
<i>Figure 3: Histogrammes représentant l'évolution du nombre d'utilisateurs mensuels par usage.</i>	12
<i>Figure 4: Histogrammes représentant l'évolution de la valeur transférée mensuelle par usage.</i>	13
<i>Figure 5: Graphe représentant les interactions entre les utilisateurs et les protocoles.</i>	13
<i>Figure 6: Graphe représentant les interactions entre les utilisateurs et les types de protocoles.</i>	14
<i>Figure 7: Réduction de la dimensionnalité des caractéristiques par la méthode PCA.</i>	17
<i>Figure 8: Évolution de l'inertie et du score de silhouette en fonction du nombre de clusters.</i>	18
<i>Figure 9: Répartition des utilisateurs en 4 clusters distincts par comportements similaires.</i>	20
<i>Figure 10: Taux d'utilisation des types de protocoles par cluster.</i>	20
<i>Figure 11: Taux d'utilisation des protocoles spécifiques par cluster.</i>	21
<i>Figure 12 : Caractéristiques d'activités par cluster.</i>	21
<i>Figure 13 : Corrélations entre les caractéristiques et les clusters.</i>	22
<i>Figure 14: Taux d'utilisation par type de protocole</i>	25

# Glossaire

- <sup>1</sup> **DeFi** : Decentralized Finance (En), Finance Décentralisée (Fr).
- <sup>2</sup> **Smart Contract** : Contrat Intelligent (Fr). Programme informatique auto-exécutant sur blockchain respectant des termes contractuels prédéfinis.
- <sup>3</sup> **DEX** : Decentralized Exchanges (En), Échanges Décentralisée (Fr). Plateformes d'échanges permettant aux utilisateurs de négocier des actifs numériques sans passer par un intermédiaire centralisé.
- <sup>4</sup> **Lending Platforms** : Plateformes de prêt.
- <sup>5</sup> **Stablecoins** : Cryptomonnaies stables.
- <sup>6</sup> **Yield Farming** : Exploitation de rendement.
- <sup>7</sup> **NFT** : Non-Fungible Tokens (En), Jeton Non-Fongibles (Fr).
- <sup>8</sup> **dApps** : Decentralized Applications (En), Applications Décentralisées (Fr). Applications qui fonctionnent sur une blockchain ou un réseau décentralisé de pairs, plutôt que sur un serveur centralisé.
- <sup>9</sup> **TVL** : Total Value Locked (En), Valeur Totale Verrouillée (Fr). Représente la valeur totale des actifs déposés dans un protocole ou une plateforme DeFi.
- <sup>10</sup> **MiCA** : Markets in Crypto-Assets (En), Marchés des Cryptoactifs (Fr). Réglementation européenne visant à encadrer les marchés des cryptoactifs.
- <sup>11</sup> **K-means** : Algorithme de partitionnement de données utilisé pour diviser un ensemble de données en k groupes distincts (ou clusters).
- <sup>12</sup> **API** : Application Programming Interface (En), Interface de Programmation d'Applications (Fr). Définit les méthodes et les protocoles utilisés pour échanger des données et des fonctionnalités entre différents systèmes ou composants logiciels.
- <sup>13</sup> **AMF** : Autorité des Marchés Financiers.
- <sup>14</sup> **ACPR** : Autorité de Contrôle Prudentiel et de Résolution.
- <sup>15</sup> **Python** : Langage de programmation orienté objet, interprété, de haut niveau et polyvalent.
- <sup>16</sup> **MongoDB** : Base de données NoSQL orientée documents, conçue pour être évolutive et permettre la gestion de documents non-structurés ou semi-structurés.
- <sup>17</sup> **CoinGecko** : agrégateur de données indépendant qui fournit des informations sur les marchés de la finance décentralisée.
- <sup>18</sup> **Etherscan** : explorateur de blockchain qui fournit des informations sur le réseau Ethereum.
- <sup>19</sup> **YahooFinance** : plateforme qui fournit des informations financières, des actualités économiques et des outils d'analyse pour les investisseurs et traders.
- <sup>20</sup> **GitHub** : plateforme de développement collaboratif permettant aux développeurs de stocker, gérer, contrôler les versions et partager du code source.

<sup>21</sup> **HuggingFace Hub** : plateforme d'hébergement et de partage de modèles de machine learning, de jeux de données et de démonstrations pour faciliter la recherche et le développement en intelligence artificielle.

<sup>22</sup> **Datashader** : bibliothèque Python conçue pour créer des visualisations efficaces de grands ensembles de données en utilisant la rasterisation pour transformer des données vectorielles en images matricielles et l'agrégation pour gérer les données volumineuses.

<sup>23</sup> **PCA** : Principal Component Analysis (En), Analyse en Composantes Principales (Fr). Méthode statistiques utilisée pour réduire la dimensionnalité d'un ensemble de données tout en conservant la variance présente dans les données.

<sup>24</sup> **Optuna** : bibliothèque d'optimisation hyperparamétrique pour automatiser la recherche des meilleurs paramètres dans les modèles de machine learning.

<sup>25</sup> **Prompt** : instruction ou question transmise à un modèle d'intelligence artificielle pour générer une réponse ou effectuer une tâche spécifique.

# Présentation de l'étude

## 1. Introduction Générale

### 1.1 Contexte et importance de la finance décentralisée (DeFi)

#### 1.1.1 Définition de la finance décentralisée

La finance décentralisée (DeFi<sup>1</sup>) désigne un écosystème de services financiers basés sur la technologie blockchain. Elle vise à offrir des alternatives aux systèmes financiers traditionnels en utilisant des contrats intelligents (Smart Contracts<sup>2</sup>) pour automatiser et sécuriser les transactions financières sans intermédiaires. La DeFi permet aux utilisateurs de prêter, emprunter et échanger des actifs de manière autonome et transparente grâce à l'utilisation de protocoles dédiés.

Les types de protocoles dans la DeFi incluent :

- Decentralized Exchanges (DEX<sup>3</sup>) : plateformes permettant aux utilisateurs d'échanger des cryptomonnaies directement entre eux sans passer par un intermédiaire centralisé. Les DEX utilisent des contrats intelligents pour faciliter les transactions et garantir la sécurité des fonds.
- Lending Platforms<sup>4</sup> : plateformes permettant aux utilisateurs de prêter ou d'emprunter des actifs numériques. Les prêteurs peuvent gagner des intérêts sur leurs actifs, tandis que les emprunteurs peuvent accéder à des fonds sans avoir besoin d'une approbation traditionnelle de crédit.
- Stablecoins<sup>5</sup> : cryptomonnaies conçues pour minimiser la volatilité des prix en étant liées à une réserve d'actifs, comme des devises fiduciaires ou des matières premières. Les stablecoins sont souvent utilisés pour les transactions quotidiennes et comme un refuge contre la volatilité des autres cryptomonnaies.
- Yield Farming<sup>6</sup> : pratique consistant à verrouiller des fonds dans un protocole DeFi pour générer des rendements sous forme de récompenses. Les utilisateurs peuvent maximiser leurs rendements en participant à différents pools de liquidité.
- Non-Fungible Tokens (NFT<sup>7</sup>) : jetons cryptographiques uniques qui représentent la propriété d'un actif, qu'il soit numérique ou physique, tels que des œuvres d'art, des objets de collection ou des biens immobiliers. Ils sont utilisés pour prouver la propriété et l'authenticité d'un actif sur la blockchain.

D'autres protocoles et innovations émergent constamment dans l'écosystème DeFi, tels que les protocoles d'assurances décentralisées et les solutions de gouvernances décentralisées. Ces développements continus visent à améliorer l'accessibilité, la sécurité et l'efficacité des services financiers.

---

<sup>1</sup> DeFi : Decentralized Finance (En), Finance Décentralisée (Fr).

<sup>2</sup> Smart Contract : Programme auto-exécutant sur blockchain respectant des termes contractuels prédéfinis.

<sup>3</sup> DEX : Decentralized Exchanges (En), Échanges Décentralisée (Fr).

<sup>4</sup> Lending Platforms : Plateformes de prêt.

<sup>5</sup> Stablecoins : Cryptomonnaies stables.

<sup>6</sup> Yield Farming : Exploitation de rendement.

<sup>7</sup> NFT : Non-Fungible Tokens (En), Jeton Non-Fongibles (Fr).

### 1.1.2 Évolution de la finance décentralisée

L'évolution de la finance décentralisée a été marquée par des innovations technologiques comme l'adoption de la blockchain avec *Bitcoin* en 2009 et les contrats intelligents sur *Ethereum* en 2015, permettant le développement d'applications décentralisées (dApps<sup>8</sup>).

D'après le média *Decrypt*, les protocoles pionniers tels que *Maker DAO* et *Compound*, ont émergés, ouvrant la voie à divers services financiers décentralisés. La croissance exponentielle de la DeFi, notamment en 2020, a attiré des milliards de dollars en valeur totale verrouillée (TVL<sup>9</sup>), diversifiant ainsi les services offerts [1].

Pendant, cette expansion a soulevé des défis en matière de sécurité et de régulation. Aujourd'hui, la DeFi continue de se développer, avec des institutions financières traditionnelles explorant des moyens d'intégrer ces technologies dans leurs opérations, comme le démontre cet article du média *Forbes* [2].

Pour une vue d'ensemble de cette évolution, voir [Annexe 1 : Évolution De La Finance Decentralisée.](#)

### 1.1.3 Importance économique et technologique

D'un point de vue économique, la DeFi offre des alternatives décentralisées aux services financiers traditionnels, rendant l'accès aux crédits et aux investissements possibles pour une population plus large. D'un point de vue technologique, les services financiers décentralisés sont souvent plus rapides, plus fiables et moins coûteux que les systèmes traditionnels.

Ces avantages cumulés amènent les institutions financières traditionnelles à intégrer des éléments de la DeFi dans leurs opérations. Par exemple, selon le média *Cryptonaute*, la Société Générale a été l'une des premières banques européennes à proposer un stablecoin conforme à la réglementation MiCA<sup>10</sup>, démontrant ainsi son intérêt pour l'intégration de la blockchain dans ses services financiers [3]. Cette convergence pourrait donner naissance à de nouveaux produits financiers hybrides, combinant ainsi les avantages des systèmes décentralisés et traditionnels.

## 1.2 Objectifs de l'étude

L'objectif principal de cette étude consiste à analyser les schémas de comportements des utilisateurs dans les applications DeFi. En examinant comment les utilisateurs interagissent avec les protocoles, l'objectif est d'obtenir une compréhension approfondie des dynamiques qui sous-tendent l'adoption et l'utilisation de ces technologies.

Cet objectif principal peut être décomposé en plusieurs sous-objectifs spécifiques :

- Comprendre les différents types d'utilisateurs qui interagissent avec les protocoles : analyser les segments d'utilisateurs en fonction de leur niveau d'engagement, allant des utilisateurs occasionnels aux investisseurs actifs.
- Comprendre les comportements et les avantages recherchés par les utilisateurs : étudier les habitudes d'utilisation des plateformes DeFi, y compris la fréquence des transactions, les types de services utilisés (prêts, échanges, yield farming etc.) et les montants investis.

---

<sup>8</sup> dApps : Decentralized Applications (En), Applications Décentralisées (Fr). Applications qui fonctionnent sur une blockchain ou un réseau décentralisé de pairs, plutôt que sur un serveur centralisé.

<sup>9</sup> TVL : Total Value Locked (En), Valeur Totale Verrouillée (Fr). Représente la valeur totale des actifs déposés dans un protocole ou une plateforme DeFi.

<sup>10</sup> MiCA : Markets in Crypto-Assets (En), Marchés des Cryptoactifs (Fr). Réglementation européenne visant à encadrer les marchés des cryptoactifs.

### 1.3 Problématique

La finance décentralisée (DeFi) a émergé comme une alternative innovante aux systèmes financiers traditionnels, promettant une plus grande accessibilité, transparence et efficacité. Cependant, malgré son potentiel, l'adoption et l'utilisation des applications DeFi restent encore mal comprises. La problématique principale de cette étude est donc de comprendre comment les utilisateurs interagissent avec les applications DeFi et quels sont les schémas de comportement observés.

Pour répondre à cette problématique principale, plusieurs questions de recherche spécifiques seront explorées :

- Quels sont les principaux facteurs influençant les comportements des utilisateurs dans les applications DeFi ? Cette question vise à identifier les motivations et les freins à l'utilisation des services DeFi, en tenant compte des aspects technologiques, économiques et sociaux.
- Comment les usages évoluent-ils au fil du temps ? Il s'agit d'analyser les tendances et les changements dans les habitudes d'utilisation, en fonction des évolutions technologiques, des conditions de marché et des régulations.

### 1.4 Méthodologie

La méthodologie employée pour cette étude repose sur l'apprentissage non-supervisé. A partir d'un jeu de données, l'objectif consiste à identifier des segments d'utilisateurs dont les comportements sont proches ou similaires.

Cette méthodologie implique plusieurs étapes clés :

- Collecte de données : Les données sont extraites de la blockchain *Ethereum* couvrant divers protocoles DeFi tels que les échanges décentralisés (DEX), les plateformes de prêts, les stablecoins, le yield farming et les jetons non fongibles (NFT).
- Ingénierie des caractéristiques : Les données collectées sont traitées pour générer des caractéristiques par adresse utilisateur, en agrégeant les métriques utilisateurs, les données de transactions et les données de marché.
- Clustering : L'algorithme K-means<sup>11</sup> est appliqué pour regrouper les utilisateurs en clusters basés sur leurs caractéristiques et activités transactionnelles.

Le périmètre défini pour cette étude consiste à étudier les 11 protocoles suivants :

- Decentralized Exchanges (DEX): *Uniswap, Curve DAO, Balancer*
- Lending Platforms: *Aave, Maker*
- Stablecoins: *Tether, USD Coin (USDC), Dai*
- Yield Farming: *Yearn Finance, Harvest Finance*
- Non-Fungible Token (NFT): *NFTfi*

Les données transactionnelles utilisées lors de cette analyse sont exclusivement issues de la blockchain *Ethereum* et sont extraites via l'API<sup>12</sup> *Etherscan*. Les données de marchés liées à chaque protocole sont issues de l'API *Yahoo Finance*.

---

<sup>11</sup> K-means : Algorithme de partitionnement de données utilisé pour diviser un ensemble de données en  $k$  groupes distincts (ou clusters).

<sup>12</sup> API : Application Programming Interface (En), Interface de Programmation d'Applications (Fr). Définit les méthodes et les protocoles utilisés pour échanger des données et des fonctionnalités entre différents systèmes ou composants logiciels.

## 1.5 Structure

La première section, intitulée ‘*Contexte Théorique et Cadre Conceptuel*’ traitera des fondements de la finance décentralisée, en expliquant les concepts fondamentaux et les technologies sous-jacentes telles que la blockchain et les contrats-intelligents. Une exploration de l’écosystème DeFi sera réalisée afin d’identifier les principaux acteurs et les types d’applications disponibles. Les théories économiques et comportementales seront abordées, en définissant les concepts clés et en formulant des hypothèses de recherches permettant de définir les modèles de comportement des utilisateurs.

La seconde section, intitulée ‘*Revue de la Littérature sur les Comportements des Utilisateurs dans la DeFi*’ passera en revue les études précédentes sur la DeFi et résumera les principales conclusions. Une comparaison des comportements des utilisateurs dans les environnements financiers traditionnels avec ceux observés dans la DeFi sera également effectué, afin d’identifier les facteurs psychologiques, économiques et technologiques influençant les comportements.

La troisième section, intitulée ‘*Méthodologie de Recherche*’ décrira l’approche méthodologique adoptée, détaillera les sources et les méthodes de collecte de données et présentera les techniques et outils d’analyse utilisés. Une sous-section sera notamment dédiée à la validité des données et aux limites méthodologiques de l’étude.

La quatrième section, intitulée ‘*Interprétation des Résultats*’, présentera les résultats principaux et les visualisations associées. Une analyse des schémas de comportements observés et les facteurs influents déduits seront partagés et l’atteinte des objectifs spécifiques seront mis en perspective.

En cinquième lieu, la section ‘*Enrichissement de l’Étude*’ proposera des perspectives identifiées intéressantes en complément de la présente analyse pour y combler les lacunes et approfondir la compréhension du sujet complexe que représente la finance décentralisée.

La sixième section, intitulée ‘*Conclusion Générale*’ synthétisera les principaux résultats et réponses aux questions de recherches initiales. Une évaluation des contributions sera effectuée et une interprétation des perspectives d’évolution de la finance décentralisée sera proposée.

Enfin, la septième section, intitulée ‘*Conclusion Personnelle*’ clôturera le présent mémoire par la réalisation d’un bilan personnel des apports de la formation et des perspectives professionnelles qui en découlent.

## 2. Contexte Théorique et Cadre Conceptuel

### 2.1 Fondements de la finance décentralisée

D’après le billet de blog n°352 de la Banque de France, la technologie blockchain et les contrats intelligents (smart contracts) sont au cœur de la finance décentralisée :

« La blockchain est une technologie de stockage et de transmission de l’information permettant, entre autres, d’échanger des actifs numériques. Les données des transactions, regroupées au sein de blocs sécurisés par cryptographie, sont enregistrées dans un registre distribué, décentralisé, immuable et transparent.

Ainsi, la blockchain permet de réaliser diverses opérations au moyen de programmes informatiques appelés smart contracts (ou contrats intelligents) exécutant les conditions d’un accord entre des parties prenantes selon des contraintes prédéfinies inscrites dans le code informatique. » [4].

En complément, certains concepts technologiques et idéologiques complètent cet écosystème. La tokenisation permet notamment de convertir des actifs en jetons numériques, facilitant ainsi les échanges et les usages dans l'écosystème numérique que constitue la DeFi. La décentralisation et la transparence sont assurées par l'enregistrement public des transactions, offrant à tous les utilisateurs un accès égal aux informations et aux services financiers, tandis que la gouvernance décentralisée permet aux détenteurs de jetons de participer aux décisions de gestion des protocoles, renforçant ainsi l'autonomie et la responsabilité des utilisateurs.

## 2.2 Écosystème DeFi

L'écosystème DeFi est composé de plusieurs acteurs clés. Les développeurs créent et maintiennent les protocoles et applications décentralisées. Les utilisateurs interagissent avec les applications DeFi pour divers services financiers. Les investisseurs fournissent le capital nécessaire au développement et à la croissance des projets. Enfin, les régulateurs supervisent et encadrent les activités pour assurer la conformité et la protection des utilisateurs. Par exemple, l'AMF<sup>13</sup> et l'ACPR<sup>14</sup> en France mettent régulièrement à jour leurs listes de sites non autorisés proposant des investissements sur les cryptoactifs [5].

D'un point de vue applicatif, la DeFi couvre une variété de services financiers, notamment les prêts et les emprunts, les échanges de devises ou encore le yield farming qui permet de générer des rendements.

L'écosystème est en constante évolution, avec de nouvelles applications et innovations émergentes recherchant l'amélioration des expériences utilisateurs. D'après le média *Kucoin*, l'efficacité opérationnelle en est la clé, notamment via la diminution des frais de transactions, la robustesse des contrats intelligents ou l'interopérabilité inter-chaînes (cross-chain) qui favoriserait une adoption plus large et renforcerait ainsi la résilience de l'écosystème [6].

## 2.3 Théories comportementales

Plusieurs théories liées à l'adoption des usages peuvent être appliquées à la finance décentralisée.

### 2.3.1 Théorie de la diffusion des innovations

La théorie de la diffusion des innovations, proposée par Everett Rogers, explique comment, pourquoi et à quelle vitesse les innovations technologiques se propagent [7]. Elle identifie cinq catégories d'adoptants : les innovateurs, les adeptes précoces, la majorité précoce, la majorité tardive et les retardataires. Au niveau des individus, le mécanisme de l'adoption se décomposerait en cinq phases :

- la connaissance : exposé à l'innovation, l'individu réagit en fonction de son profil personnel et du système social dans lequel il évolue ;
- la persuasion : étape cruciale de la théorie, l'individu amorce une prise de position au sujet de l'innovation en fonction des caractéristiques perçues (avantage relatif d'un point de vue économique et social, compatibilité avec les valeurs du groupe d'appartenance, complexité, possibilité de tester l'innovation et visibilité pour montrer les résultats aux autres individus) ;
- la décision : l'individu s'engage dans des activités d'évaluation lui permettant d'adopter ou de rejeter l'innovation ;

---

<sup>13</sup> AMF : Autorité des Marchés Financiers.

<sup>14</sup> ACPR : Autorité de Contrôle Prudentiel et de Résolution

- l'implantation : l'individu a besoin d'assistance pour réduire les incertitudes sur les conséquences ;
- la confirmation : l'individu tente d'obtenir des informations venant, a posteriori renforcer son choix.

### 2.3.2 Théorie des réseaux et des effets de réseau

La théorie des réseaux et des effets de réseau stipule que la valeur d'un produit ou service augmente avec le nombre d'utilisateurs. Ces effets, positifs ou négatifs, sont présents dans divers segments économiques, y compris la finance décentralisée comme l'exprime cet article de *Binance Academy* [8]. Les réseaux DeFi peuvent être représentés comme des graphes, où les nœuds sont les participants (utilisateurs ou plateformes) et les arêtes sont les interactions (transactions, smart contracts).

D'un point de vue positif, les nouveaux utilisateurs ajoutent de la valeur au réseau. En effet, le principe même de la blockchain induit que la structure et la dynamique des réseaux peuvent influencer la robustesse, l'efficacité et la sécurité des systèmes. Par exemple, une blockchain avec de nombreux nœuds peut être plus résistante aux attaques par rapport à un réseau centralisé. De la même façon, une plateforme de prêt qui intègre un grand nombre d'utilisateurs peut offrir des taux d'intérêts plus compétitifs et une liquidité accrue.

D'un point de vue négatif, la chute d'une plateforme décentralisée peut affaiblir la confiance des utilisateurs dans l'ensemble du réseau. Par exemple, une faille de sécurité dans un protocole pourrait compromettre les fonds des utilisateurs et réduire la confiance dans l'ensemble de l'écosystème. Autre exemple, les frais de transactions de la blockchain *Ethereum* repose sur un système de surenchère. A mesure que l'utilisation de la blockchain augmente, les frais de transactions augmentent en conséquence, ce qui n'est pas viable sur le long terme. Ce phénomène a par exemple induit la gouvernance de la blockchain *Ethereum* à proposer une refonte du système de frais de transactions [8].

Ainsi, bien que les effets de réseau puissent apporter des avantages significatifs, il est crucial pour les plateformes de gérer les risques associés pour garantir la stabilité, la croissance et la résilience de l'écosystème DeFi.

## 2.4 Cadre conceptuel de l'étude

Dans le cadre de cette étude, pour réaliser une analyse des comportements des utilisateurs dans les applications DeFi, les hypothèses suivantes sont formulées.

Hypothèse 1 : Les incitations économiques, telles que les rendements élevés du yield farming, motivent les utilisateurs à s'engager davantage dans les protocoles DeFi.

Hypothèse 2 : Les avantages proposés par la finance décentralisée rendent obsolètes les services financiers traditionnels.

Hypothèse 3 : La courbe d'apprentissage pour comprendre et utiliser les services DeFi peut être raide, décourageant ainsi les utilisateurs moins familiers avec la technologie. La théorie de diffusion des innovations identifie la complexité comme un facteur influençant l'adoption.

Hypothèse 4 : Les effets de réseau positifs renforcent la confiance et l'engagement des utilisateurs et a contrario, les effets de réseau négatifs diminuent l'engagement des utilisateurs dans l'écosystème DeFi.

## 3. Revue de la Littérature sur les Comportements des Utilisateurs dans la DeFi

### 3.1 Études précédentes liées aux comportements des utilisateurs dans la DeFi

#### 3.1.1 DeFi Survival Analysis : Insights into Risks and User Behaviors

Cette étude examine les risques à travers une analyse de survie, les comportements des utilisateurs dans les applications DeFi via le protocole de prêts et d'emprunts *Aave*. Elle révèle les comportements à risque et les facteurs influençant la durée de participation des utilisateurs sur ces plateformes. Selon la recherche, les utilisateurs adoptent des comportements variés, certains étant plus enclins à prendre des risques que d'autres, il serait donc essentiel de comprendre les comportements pour améliorer la conception des plateformes DeFi et réduire les risques pour les utilisateurs [9].

#### 3.1.2 Bridging the adoption gap for cryptocurrencies: understanding the affordances that impact approach-avoidance behavior for potential users and continuation usage for actual users

Cette étude examine les facteurs qui influencent l'adoption des cryptomonnaies, en se concentrant sur les comportements d'approche et d'évitement des utilisateurs potentiels et actuels. Les auteurs ont recueilli des données auprès de 480 utilisateurs potentiels et réels en Allemagne. Selon les résultats de l'étude, les avantages de valeur perçus ont un impact positif sur l'intention comportementale des utilisateurs potentiels d'adopter la crypto-monnaie, et ils influencent l'intention des utilisateurs réels de continuer à l'utiliser. Cependant, la volatilité perçue et la tolérance au risque financier sont les facteurs les plus cruciaux entravant l'adoption de la crypto-monnaie, que ce soit au stade pré-adoption ou post-adoption [10].

#### 3.1.3 User behavior and token adoption on ERC20

Ce papier de recherche analyse les dynamiques sociales et les comportements des utilisateurs sur le réseau de transaction ERC20 (*Ethereum*). Il met en évidence que le réseau ERC20 présente des propriétés similaires à celles des réseaux sociaux, avec des utilisateurs influents qui sont suivis par des utilisateurs moins populaires. La raison de ces similitudes résiderait dans la complexité des systèmes sociaux dans lesquels des groupes cohésifs sont intégrés avec d'autres groupes à grande échelle par le biais de liens faibles.

L'étude démontre des profils d'utilisateurs et une adoption des protocoles hétérogène en termes de nombre et de diversité des transactions. Cependant, seuls quelques utilisateurs diversifient leurs transactions, rendant le système potentiellement vulnérable en cas de défaillance des acteurs clés. En effet, la diversification des portefeuilles améliorerait la robustesse face aux changements du marché grâce au fait que les utilisateurs diversifiés réaliseraient des connexions de communauté, mais sa rareté poserait un risque pour la stabilité globale du réseau [11].

#### 3.1.4 Behavior-aware Account De-anonymization Ethereum Interaction Graph

Cette étude propose un cadre basé sur les réseaux de neurones graphiques pour identifier les comptes *Ethereum* en fonction de leurs comportements et vise à améliorer la sécurité financière en détectant les activités illégales. D'après la recherche, Le cadre proposé est efficace pour caractériser les comportements des comptes (levée de fonds, minage, échanges, hameçonnage) et ainsi identifier les activités suspectes, contribuant ainsi à la régulation du marché et à la sécurité financière [12].

### 3.1.5 Onchain Analysis : A Comparative Study of Decentralized Exchange (DEX) Activities on Ethereum, Solana, and Binance Smart Chain (BSC)

Cette étude compare les activités d'échanges sur les blockchains *Ethereum*, *Solana* et *Binance Smart Chain* (BSC) en termes de volume de transactions, de liquidité et de revenus des protocoles. D'après la recherche, la blockchain *Ethereum* domine en termes de volume de transactions et de liquidité, portée par l'adoption institutionnelle malgré des frais de transaction élevés. La blockchain *Solana* se distingue par son efficacité transactionnelle ce qui attire les traders à haute fréquence, bien que les indicateurs démontrent une rétention de liquidité faible. La blockchain BSC quant à elle, démontre des signes de stagnation. Les auteurs soulignent la nécessité de solutions de scalabilité et de ponts de liquidité inter-chaînes pour atténuer la fragmentation de l'écosystème et améliorer l'efficacité de la DeFi [13].

## 3.2 Facteurs influençant les comportements des utilisateurs

Selon les différentes études évoquées, les comportements des utilisateurs semblent influencés par une multitude de facteurs qui peuvent être regroupés en trois grandes catégories : psychologiques, économiques et technologiques.

### 3.2.1 Facteurs psychologiques

Les comportements des utilisateurs semblent fortement influencés par la perception des risques et des avantages. Par exemple, l'étude '*Bridging the adoption gap for cryptocurrencies*' a révélé que les avantages de valeur perçus ont un impact positif sur l'intention d'adopter ces monnaies, tandis que la tolérance au risque financier et à la volatilité perçue joue un rôle crucial dans la décision d'adoption [10]. De plus, l'étude '*DeFi Survival Analysis*' démontre que les comportements à risque varient parmi les utilisateurs des plateformes DeFi, certains étant plus enclins à prendre des risques que d'autres [9]. Enfin, l'étude '*User behavior and token adoption on ERC20*' démontre qu'il existe des similitudes entre les réseaux blockchains et les réseaux sociaux, dans lesquels les utilisateurs les plus influents impactent les autres utilisateurs moins populaires [11].

### 3.2.2 Facteurs économiques

Les facteurs économiques, tels que la volatilité des marchés et les frais de transaction, jouent également un rôle déterminant dans les comportements des utilisateurs. L'étude '*Bridging the adoption gap for cryptocurrencies*' démontre que la volatilité perçue des cryptomonnaies est un obstacle majeur à leur adoption, tant pour les utilisateurs potentiels que pour les utilisateurs actuels [10]. De plus, l'étude '*Onchain Analysis*' démontre que les frais de transaction élevés sur certaines blockchains peuvent influencer le choix des plateformes par les utilisateurs, malgré une liquidité et un volume de transactions élevés [13].

### 3.2.3 Facteurs technologiques

D'un point de vue technologique, l'étude '*User behavior and token adoption on ERC20*' démontre que dans le système complexe que représente la DeFi, la diversification des transactions est essentielle pour la robustesse du système, bien que peu d'utilisateurs adoptent cette pratique [11]. En complément, l'étude '*Behavior-aware Account De-anonymization Ethereum Interaction Graph*' insiste sur l'importance de la sécurité des systèmes pour une adoption de l'usage et démontre que des solutions technologiques telles que les réseaux de neurones graphiques peuvent améliorer la sécurité en identifiant les activités suspectes sur la blockchain [12].

### 3.3 Lacunes dans la littérature

#### 3.3.1 Limites méthodologiques des études existantes

Les études existantes sur les comportements des utilisateurs dans la DeFi offrent des perspectives précieuses, mais elles présentent également des biais potentiels et des limitations méthodologiques. Par exemple, l'étude '*DeFi Survival Analysis*' se concentre principalement sur le protocole *AAVE*, ce qui peut ne pas refléter la diversité des comportements sur d'autres plateformes DeFi [9]. De plus, les interactions entre différents types de protocoles (DEX, plateformes de prêt, stablecoins, etc.) ne sont pas suffisamment explorées, laissant une lacune dans la compréhension des dynamiques inter-protocoles.

L'étude '*Bridging the adoption gap for cryptocurrencies*' met en lumière les facteurs psychologiques influençant l'adoption des cryptomonnaies, mais elle se limite à un échantillon d'utilisateurs en Allemagne, ce qui peut ne pas être représentatif des comportements globaux [10]. De même, l'analyse des comportements sur le réseau ERC20 dans '*User behavior and token adoption on ERC20*' souligne l'importance des utilisateurs influents, mais elle ne propose pas de solutions concrètes pour encourager la diversification des transactions, essentielle à la robustesse du système [11].

Les facteurs économiques, tels que la volatilité et les frais de transaction, sont bien documentés, mais leurs impacts sur les comportements inter-chaînes restent flous. Par exemple, l'étude '*Onchain Analysis*' compare les activités sur différents blockchains sans explorer comment les utilisateurs naviguent entre ces écosystèmes [13].

#### 3.3.2 Étude proposée

Pour compléter la compréhension des mécanismes dans la finance décentralisée, une analyse approfondie des comportements inter-protocoles sur la blockchain *Ethereum* (réseau ERC20) est proposée. Cette étude se concentrera sur les interactions entre différents types de protocoles, notamment les échanges décentralisés (DEX) comme *Uniswap*, *Curve DAO*, et *Balancer* ; les plateformes de prêt telles qu'*Aave* et *Maker* ; les stablecoins comme *Tether*, *USD Coin (USDC)*, et *Dai* ; les plateformes de yield farming telles que *Yearn Finance* et *Harvest Finance* ; ainsi que les jetons non fongibles (NFT) via *NFTfi*.

#### 3.3.3 Justification de l'étude

Cette étude vise à apporter une contribution supplémentaire à la compréhension des schémas de comportements des utilisateurs dans l'écosystème DeFi. En analysant et en caractérisant les comportements inter-protocoles, cette recherche pourrait offrir des informations complémentaires pour améliorer la conception des plateformes et encourager une adoption plus large et plus diversifiée des technologies DeFi.

## 4. Méthodologie de Recherche

### 4.1 Approche méthodologique

L'approche méthodologique adoptée pour cette étude est de type mixte, combinant des méthodes qualitatives et quantitatives afin de fournir une analyse exhaustive des comportements des utilisateurs dans l'écosystème DeFi.

D'un point de vue quantitatif, l'étude se concentre sur des indicateurs clés tels que les montants de valeurs transférées, les frais de transaction, le nombre d'utilisateurs actifs, ainsi que le nombre de transactions effectuées sur une période donnée. Ces données quantitatives permettent de mesurer l'activité et les tendances au sein des différents protocoles DeFi, notamment les comportements transactionnels des utilisateurs.

D'un point de vue qualitatif, la mise en perspective des poids que représentent les indicateurs clés cités précédemment, attribués aux différentes catégories que sont les protocoles ou les types de protocoles permet d'affiner la compréhension des comportements dans l'écosystème.

## 4.2 Collecte de données

### 4.2.1 Définition du périmètre de l'étude

Dans un premier temps, pour étudier les dynamiques inter-protocoles et disposer d'un jeu de données variées, la blockchain *Ethereum*, considérée comme fondatrice de la finance décentralisée est sélectionnée. Ensuite, dans l'objectif d'étudier les dynamiques entre plusieurs usages (échanges, prêts, rendement etc.), les protocoles suivants sont sélectionnés pour leur popularité :

- Decentralized Exchanges (DEX): *Uniswap, Curve DAO, Balancer*
- Lending Platforms: *Aave, Maker*
- Stablecoins: *Tether, USD Coin (USDC), Dai*
- Yield Farming: *Yearn Finance, Harvest Finance*
- Non-Fungible Token (NFT) : *NFTfi*

### 4.2.2 Constitution du jeu de données

Dans un second temps, le jeu de données est constitué étape après étape en extrayant différents types de données grâce à plusieurs API afin de disposer d'une vue représentative du périmètre. Le script est réalisé en *Python*<sup>15</sup> et les données sont stockées dans une base de données *MongoDB*<sup>16</sup>. Le procédé de constitution est réalisé sous la forme d'un pipeline d'extraction, de transformation et de chargement des données, et suit les étapes suivantes :

- étape 1 : identification du contrat intelligent de chaque protocole sélectionné sur la blockchain *Ethereum* grâce à l'API de *CoinGecko*<sup>17</sup> ;
- étape 2 : extraction des transactions associées aux contrats intelligents sur la blockchain *Ethereum* directement grâce à l'API de *Etherscan*<sup>18</sup> ;
- étape 3 : extraction de profils utilisateurs en agrégeant les données issues des transactions extraites lors de l'étape 2 pour collecter les adresses uniques, les transactions associées et les métadonnées ;
- étape 4 : extraction des données de marchés associées à chaque protocole grâce à l'API de *YahooFinance*<sup>19</sup> ;
- étape 5 : enrichissement des données de marchés en agrégeant les informations des données de transactions dans les plages de temps correspondantes ;
- étape 6 : conversion des collections en données tabulaires et stockage.

---

<sup>15</sup> Python : Langage de programmation orienté objet, interprété, de haut niveau et polyvalent.

<sup>16</sup> MongoDB : Base de données NoSQL orientée documents, conçue pour être évolutive et permettre la gestion de documents non-structurés ou semi-structurés.

<sup>17</sup> CoinGecko : agrégateur de données indépendant qui fournit des informations sur les marchés de la finance décentralisée.

<sup>18</sup> Etherscan : explorateur de blockchain qui fournit des informations sur le réseau *Ethereum*.

<sup>19</sup> YahooFinance : plateforme qui fournit des informations financières, des actualités économiques et des outils d'analyse pour les investisseurs et traders.

La description détaillée du procédé de constitution du jeu de données est documentée dans le repository *GitHub*<sup>20</sup> et les données sont accessibles en open source dans le *HuggingFace Hub*<sup>21</sup>, voir Annexe 2 : Jeu De Données.

### 4.3 Traitement des données collectées

#### 4.3.1 Analyse exploratoire

Pour étudier les informations contenues dans les données, le jeu de transactions est le plus riche puisqu'il offre une temporalité en complément de données quantitatives et qualitatives.

##### 4.3.1.1 Analyse de corrélation

Dans un premier temps, une analyse de corrélation est réalisée. L'analyse est effectuée en agrégeant les données par heure et par catégorie. Chaque graphique est donc dupliqué en deux versions : une première dont les catégories sont l'ensemble des protocoles (ci-dessous à gauche) et une seconde dont les catégories sont les types de protocoles (ci-dessous à droite).

Cette analyse de corrélation permet de déduire qu'il existe une forte corrélation positive entre :

- le nombre de transactions et le nombre d'utilisateurs (1.00, voir figure ci-dessous).
- le nombre d'utilisateurs qui envoient et le nombre d'utilisateurs qui reçoivent (0.89).
- la valeur des frais de transactions et le nombre d'utilisateurs (0.96).

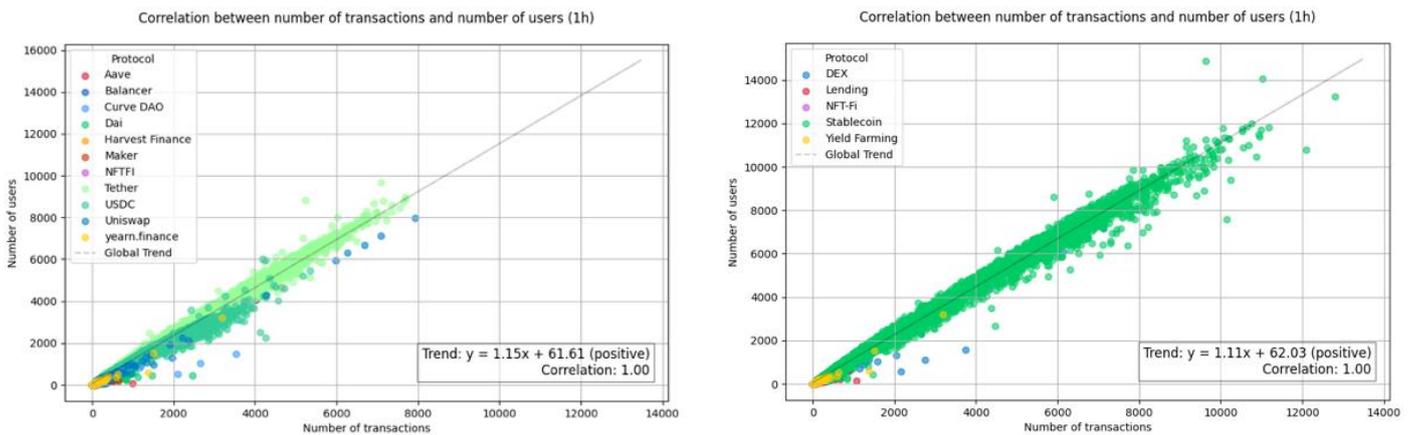


Figure 1: Mesures de corrélation entre le nombre d'utilisateurs et le nombre de transactions par heure.

<sup>20</sup> GitHub : plateforme de développement collaboratif permettant aux développeurs de stocker, gérer, contrôler les versions et partager du code source.

<sup>21</sup> HuggingFace Hub : plateforme d'hébergement et de partage de modèles de machine learning, de jeux de données et de démonstrations pour faciliter la recherche et le développement en intelligence artificielle.

En revanche, il n'existe pas de corrélation spécifique entre la valeur transférée et le nombre d'utilisateurs sur une période, ce qui sous-entend que d'autres facteurs influent sur les valeurs transférées dans l'écosystème (voir figure 2 ci-dessous).

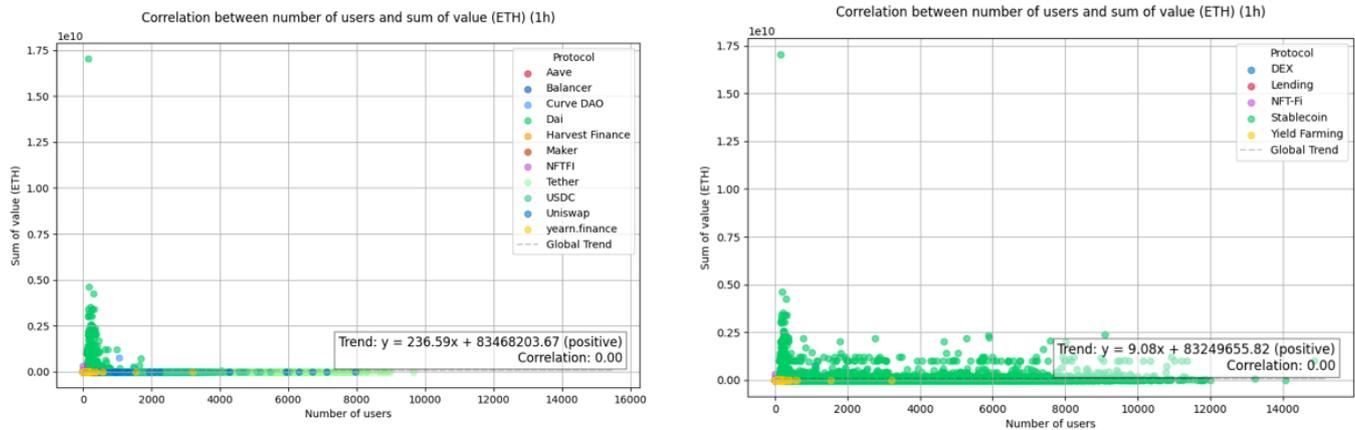


Figure 2: Mesures de corrélation entre la valeur transférée et le nombre d'utilisateurs par heure.

L'ensemble des graphiques de corrélations sont disponibles en Annexe 3 : Analyse De Corrélation.

#### 4.3.1.2 Analyse des tendances

Dans un second temps, une analyse des tendances est réalisée. L'analyse est effectuée en agrégeant les données mensuellement, par catégories. Chaque graphique est également dupliqué en deux versions, une première dont les catégories sont les protocoles (ci-dessous à gauche) et une seconde dont les catégories sont les types de protocoles (ci-dessous à droite).

Cette analyse de tendances permet de déduire qu'il existe une croissance légère de l'usage de l'écosystème DeFi au sens large entre le début de l'année 2023 et la fin de l'année 2024. En effet, le nombre d'utilisateurs (+ 4,84%), le nombre de transactions (+ 18,28%), la valeur transférée (+ 447,16%) ainsi que les frais de transactions (+ 6,80%) ont tendance à croître au fil du temps. Au niveau des comportements, les stablecoins (en vert ci-dessous) représentent la majeure partie de l'usage mais il est également observable que les protocoles d'échanges (en bleu ci-dessous) et d'emprunts (en rouge ci-dessous) ont tendance à être de plus en plus utilisés.

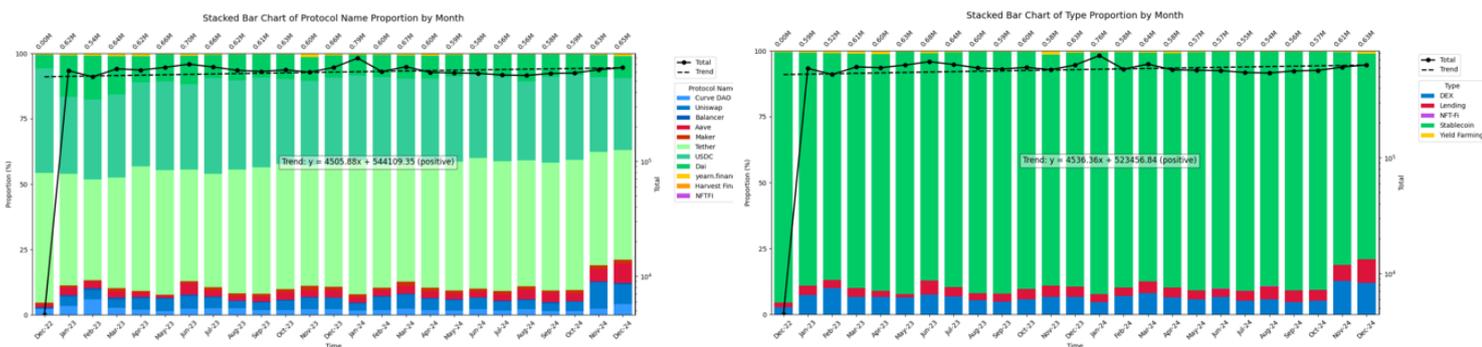


Figure 3: Histogrammes représentant l'évolution du nombre d'utilisateurs mensuels par usage.

En revanche, d'un point de vue valeur transférée, bien que la valeur soit de plus en plus importante avec le temps (voir figure 4 ci-dessous) avec une augmentation significative (+ 447,16%), on observe que les stablecoins représentent une quantité importante (+ 95%) par

rapport aux autres types de protocoles. L'hypothèse associée serait que l'usage des stablecoins en remplacement d'une monnaie traditionnelle serait de plus en plus adopté et que la complexité des autres types de protocoles demanderait un temps d'adoption plus long (voir figure 4 ci-dessous).

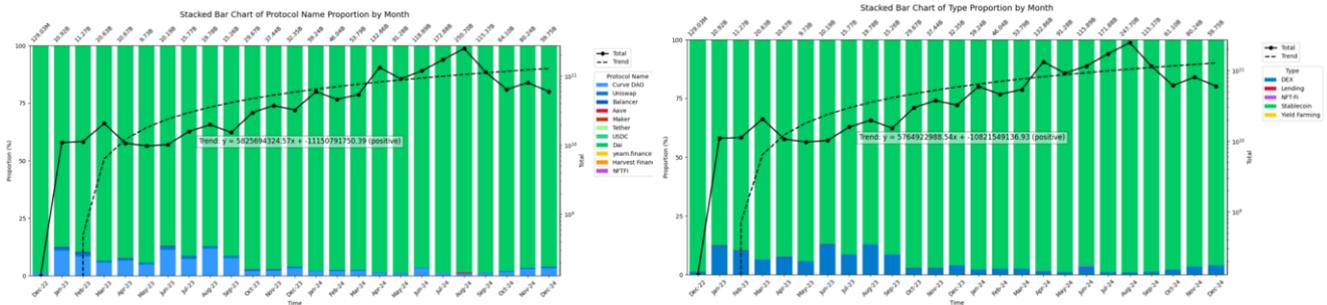


Figure 4: Histogrammes représentant l'évolution de la valeur transférée mensuelle par usage.

L'ensemble des graphiques de tendances sont disponibles en Annexe 4: Analyse De Tendence.

#### 4.3.1.3 Analyse des effets de réseau

Comme évoqué dans le paragraphe '2.3.2 Théorie des réseaux et des effets de réseau', les interactions dans la finance décentralisée peuvent être représentées comme des graphes, où les nœuds sont les participants (utilisateurs ou plateformes) et les arêtes sont les interactions (transactions, smart contracts).

Pour observer visuellement les effets de réseau, deux graphiques ont été générés via la librairie *Datashader*<sup>22</sup> en utilisant l'algorithme *ForceAtlas2* qui simule un système de forces physiques pour positionner les nœuds de manière adéquate et une fonction de regroupement des arêtes (*edge bundling*) pour réduire l'encombrement visuel. Il est important de souligner que pour des raisons de ressources computationnelles nécessaires, ces deux graphiques ne contiennent qu'une représentation de 100 000 utilisateurs aléatoirement sélectionnés.

Le premier graphe représente les connexions entre les utilisateurs et les protocoles spécifiques. Les protocoles sont représentés sous forme de nœuds carrés colorés et les arêtes de connexion sont représentées sous forme de traits bleu (voir figure 5 ci-dessous).

#### Légende :

- Curve Dao
- Aave
- Uniswap
- Maker
- Tether
- Yearn Finance
- Uxdc
- Dai
- Balancer
- Harvest Finance
- Nftfi

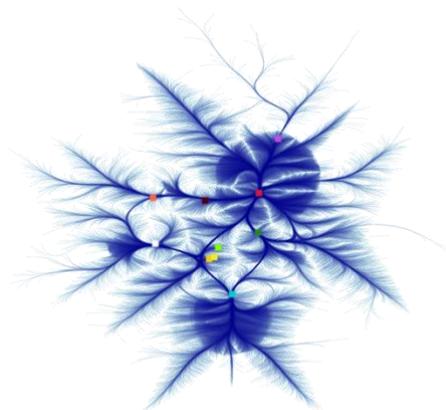


Figure 5: Graphe représentant les interactions entre les utilisateurs et les protocoles.

<sup>22</sup> Datashader : bibliothèque Python conçue pour créer des visualisations efficaces de grands ensembles de données en utilisant la rasterisation pour transformer des données vectorielles en images matricielles et l'agrégation pour gérer les données volumineuses.

Le second graphe représente les connexions entre les utilisateurs et les types de protocoles. Les types de protocoles sont représentés sous forme de nœuds carrés colorés et les arêtes de connexion sont représentées sous forme de traits bleu (voir figure 6 ci-dessous).

Légende :

- DEX
- Lending
- Stablecoin
- Yield farming
- NFT

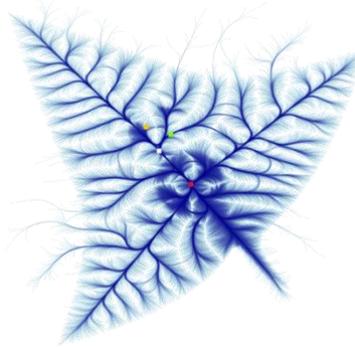


Figure 6: Graphe représentant les interactions entre les utilisateurs et les types de protocoles.

Sur ces graphiques, notamment sur la figure 5, nous pouvons observer que les stablecoins (Tether en rouge, Usdc en bleu) représentent l'épicentre de groupes denses et permettent de connecter les flux avec les autres types de protocoles.

Sur la figure 6, nous pourrions également déduire que les stablecoins représentent le principal point d'entrée ou de sortie en lien avec les protocoles d'échanges (DEX).

Les graphes sont également disponibles en [Annexe 5 : Graphe de Réseau](#).

#### 4.3.2 Ingénierie des caractéristiques

En vue d'exploiter les données pour caractériser les schémas de comportements des utilisateurs dans l'écosystème DeFi, la présente étape a été mise en œuvre pour constituer un jeu de caractéristiques le plus exhaustif possible et entraîner un modèle d'apprentissage. L'un des principaux challenges de cette étape repose sur le fait de constituer un jeu de caractéristiques tabulaire et unique pour chaque utilisateur en agrégeant les données de plusieurs transactions et les données de plusieurs périodes de marché. Pour ce faire, le jeu de caractéristiques a été constitué comme suit :

- étape 1 : chargement du jeu de données 'users.parquet' pour constituer une base tabulaire comprenant une ligne par utilisateur ;
- étape 2 : calcul d'indicateurs tels que la diversité des protocoles et des types de protocoles utilisés, le flux net de valeur transférée ou encore un score de poids par rapport aux autres utilisateurs ;
- étape 3 : à partir du jeu de données 'transactions.parquet', calcul de diverses métriques telles que l'heure de pointe, la fréquence ainsi que des statistiques (minimum, maximum, moyenne, médiane, écart-type) associées aux transactions d'envoi et de réception, aussi bien pour la valeur transférée que pour les frais de transactions ;
- étape 4 : à partir du jeu de données 'market.parquet', calcul de diverses métriques d'exposition au marché pondérées par l'usage des protocoles, telles que l'activité, le volume, la liquidité ou encore la volatilité.

Au total, chaque utilisateur est identifié par une adresse *Ethereum* unique et associé à 62 caractéristiques classables en :

- activités de transactions (4 caractéristiques) ;
- interaction avec les types de protocoles (5 types de protocoles) ;

- interaction avec les protocoles spécifiques (11 protocoles) ;
- diversité et influence (4 caractéristiques) ;
- statistiques de transactions d'envoi (14 caractéristiques) ;
- statistiques de transactions de réception (14 caractéristiques) ;
- exposition au marché (10 caractéristiques).

La description détaillée du procédé de constitution du jeu de caractéristiques est également documentée dans le repository *GitHub* et les données sont également accessibles en open source dans le *HuggingFace Hub*, voir [Annexe 2 : Jeu De Données](#).

### 4.3.3 Standardisation

A la suite de la constitution du jeu de caractéristiques, en vue de disposer d'un jeu exploitable pour réaliser une analyse via un algorithme d'apprentissage non-supervisé, une analyse statistique et une standardisation sont effectuées. L'analyse statistique a pour objectif de qualifier la distribution des données et la standardisation a pour objectif de corriger la distribution pour que chaque caractéristique contribue équitablement dans l'algorithme d'apprentissage.

Ainsi, pour chaque variable, la série d'étape suivante est suivie afin de garantir que les données soient correctement préparées :

- étape 1 : calcul d'un score de normalité (*p\_value*) pour évaluer si les données suivent une distribution normale ;
- étape 2 : calcul de la forme de la distribution avec les métriques d'asymétrie (*skewness*) et d'aplatissement (*kurtosis*) pour comprendre la répartition des données ;
- étape 3 : application d'une méthode de standardisation en fonction des métriques de distribution (*p\_value*, *skewness*, *kurtosis*) selon les règles suivantes :

- Z-score pour les données normalement distribuées (*p\_value* > 0.05). Cette méthode centre les données autour de la moyenne et les échelonne en fonction de l'écart-type.

$$z = \frac{x - \mu}{\sigma}$$

*z* : Valeur standardisée.

*x* : Valeur originale.

$\mu$  : Moyenne des données.

$\sigma$  : Écart-type des données.

- Min-Max est appliqué pour les données présentant une faible asymétrie (*skewness* < 0.5) ou un faible aplatissement (*kurtosis* < 2,5). Cette méthode échelonne les données dans un intervalle fixe [0, 1].

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

*x'* : Valeur normalisée.

*x* : Valeur originale.

*min(x)* : Valeur minimale de la caractéristique.

*max(x)* : Valeur maximale de la caractéristique.

- Transformation logarithmique pour les données présentant une forte asymétrie positive (*skewness* > 1). Cette méthode réduit l'asymétrie et stabilise la variance.

$$x' = \log(x + c)$$

$x'$  : Valeur transformée.

$x$  : Valeur originale.

$c$  : Constante ajoutée pour éviter le logarithme de zéro (ici  $c=1$ ).

- Transformation logarithmique inverse pour les données présentant une forte asymétrie négative (*skewness* < -1). Cette méthode aide à normaliser les distributions fortement inclinées vers la gauche.

$$x' = \log \frac{1}{x}$$

$x'$  : Valeur transformée.

$x$  : Valeur originale.

- Transformation Box-Cox pour les données présentant un fort aplatissement (*kurtosis* > 3.5) ou lorsque les autres méthodes ne sont pas applicables. Cette transformation stabilise la variance et rend les données davantage normales.

$$x' = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \log(x) & \text{si } \lambda = 0 \end{cases}$$

$x'$  : Valeur transformée.

$x$  : Valeur originale.

$\lambda$  : Paramètre de transformation défini pour maximiser la normalité.

Cette méthode permet ainsi de préparer les données pour qu'elles soient mieux adaptées aux algorithmes d'apprentissages automatiques, en réduisant les biais dus à des échelles ou des distributions différentes.

#### 4.3.4 Réduction de la dimensionnalité

Après l'ajustement des caractéristiques en termes de distribution, une réduction de la dimensionnalité est effectuée pour minimiser les ressources computationnelles nécessaires lors de l'apprentissage non-supervisé. Pour cela, la méthode PCA<sup>23</sup> est utilisée pour capturer la variance contenue dans les données initiales. Ainsi, 99,99966 % de la variance des 62 caractéristiques initiales, dont le seuil est défini de façon arbitraire pour équilibrer la réduction de dimensionnalité et la conservation d'informations, est expliquée par un jeu de 28 composants (voir figure 7 page suivante).

---

<sup>23</sup> PCA : Principal Component Analysis (En), Analyse en Composantes Principales (Fr). Méthode statistique utilisée pour réduire la dimensionnalité d'un ensemble de données tout en conservant autant que possible la variance présente dans les données.

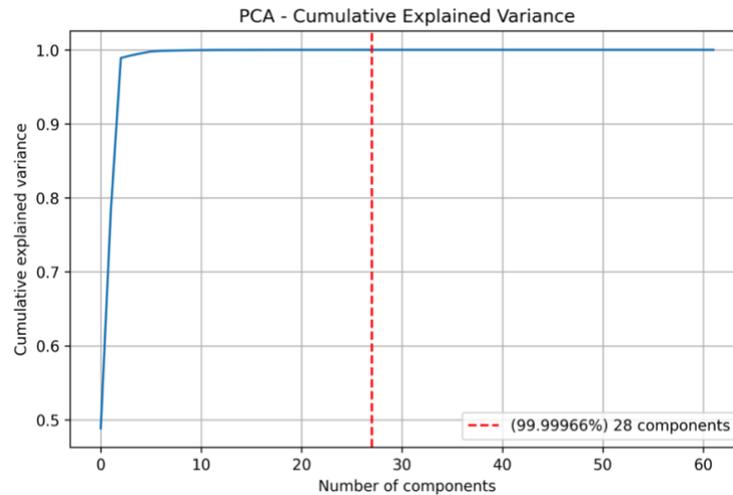


Figure 7: Réduction de la dimensionnalité des caractéristiques par la méthode PCA.

## 4.4 Apprentissage non-supervisé

### 4.4.1 Sélection d'un algorithme de clustering

Une fois le jeu de caractéristiques défini, standardisé et la dimensionnalité réduite, il est nécessaire de sélectionner une méthode de clustering afin d'identifier des groupes de comportements similaires parmi tous les utilisateurs.

Pour cela, plusieurs algorithmes sont considérés en fonction de leurs spécificités et de leur adéquation avec la nature des données.

- *K-MEANS* : rapide et efficace pour les grands ensembles de données, mais nécessite de spécifier le nombre de clusters (k) et reste sensible aux valeurs aberrantes. Existe également avec l'alternative *Mini Batch K-Means*, qui est une variante plus efficace pour les volumes importants en utilisant des sous-ensembles de données (batches) lors de l'entraînement.
- *DBSCAN* : efficace pour identifier des clusters de forme arbitraire et détecter les points aberrants, sans spécifier le nombre de clusters, mais moins performant en haute dimension.
- *HDBSCAN* : extension de *DBSCAN* gérant des clusters de différentes densités et formes, sans spécifier le nombre de clusters, utile pour les données complexes.
- *K-MEDOIDS* : plus robuste aux valeurs aberrantes que *K-Means* utilisant des médoïdes comme points centraux, mais plus coûteux en calcul pour les grands ensembles de données.
- *GMM* : modèles de mélange gaussien avec une approche probabiliste, capturant la structure sous-jacente des données et fournissant des probabilités d'appartenance aux clusters, utile pour les données avec chevauchement, mais sensible à l'initialisation et nécessitant de spécifier le nombre de composants.

En vue de la taille des données, soit 6,8 millions d'observations par 28 dimensions, et des ressources computationnelles nécessaires, les algorithmes *Mini Batch K-MEANS* et *HDBSCAN* ont été testés. Ces deux méthodes ont été sélectionnées pour leur capacité à gérer de grands volumes de données tout en offrant des approches distinctes pour la détection de clusters.

## 4.4.2 Optimisation des paramètres d'entraînement

### 4.4.2.1 Métriques de performances

Pour tester et évaluer la performance des algorithmes sélectionnés précédemment, il est nécessaire de définir les paramètres d'entraînement idéaux. Parmi les mesures de performances pour les algorithmes de clustering, les métriques suivantes sont identifiées :

- *Inertie* : mesure la somme des distances quadratiques entre chaque point et le centre de son cluster (centroïde). Une faible inertie indique que les points sont proches de leur centroïde, ce qui signifie que les clusters sont bien définis et compacts.
- *Silhouette Score* : évalue la qualité des clusters en mesurant à quel point chaque point est similaire aux autres points de son propre cluster par rapport aux points des autres clusters. Un score proche de +1 indique que les points sont bien assignés à leur cluster, tandis qu'un score proche de -1 indique une mauvaise assignation.
- *Davies-Bouldin Index* : mesure la moyenne des similarités entre chaque cluster et le cluster qui lui est le plus similaire. Une valeur faible indique que les clusters sont bien séparés les uns des autres. Une valeur élevée signifie que les clusters ne sont pas bien définis et qu'il y a un chevauchement entre eux.
- *Calinski-Harabasz Index* : représente le rapport entre la dispersion entre les clusters et la dispersion au sein des clusters. Une valeur élevée indique que les clusters sont bien définis et bien séparés. Une valeur faible signifie que les clusters ne sont pas bien définis et qu'il y a peu de différences entre eux.

Ces métriques permettent d'évaluer la qualité des clusters formés par les algorithmes et d'optimiser les paramètres d'entraînement pour obtenir des clusters plus cohérents et significatifs.

### 4.4.2.2 Détermination du nombre de cluster

Dans le cadre de l'utilisation de l'algorithme *K-MEANS*, il est particulièrement nécessaire de définir un nombre de cluster ( $k$ ) idéal. Visuellement, en traçant l'évolution de l'inertie et du score de silhouette en fonction du nombre de cluster, il est possible d'identifier le nombre idéal pour maximiser et minimiser ces métriques de performances. Dans notre cas, on observe que l'inertie et le score de silhouette sont optimaux pour un nombre de cluster égal à 4 (voir figure 8 ci-dessous).

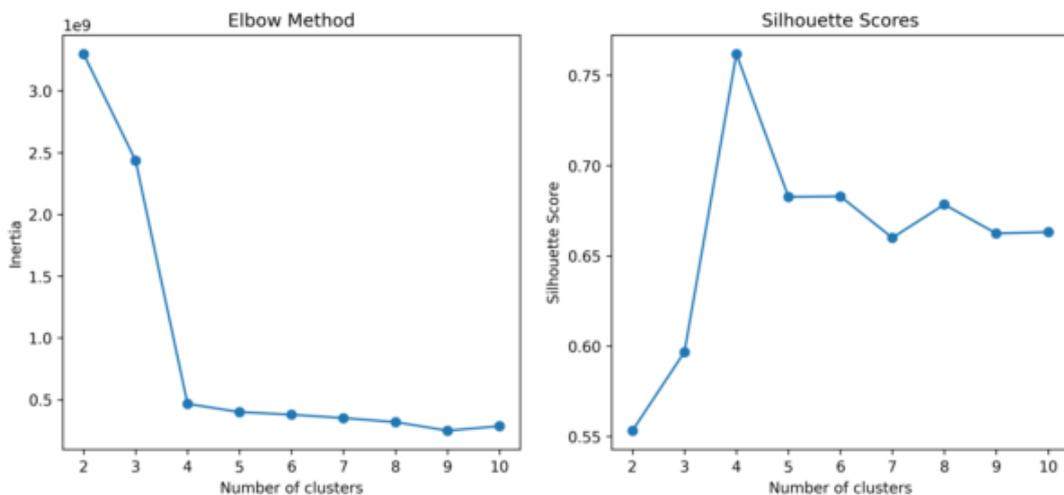


Figure 8: Évolution de l'inertie et du score de silhouette en fonction du nombre de clusters.

#### 4.4.2.3 Fonction d'optimisation des paramètres

En exploitant les métriques de mesure des performances de clustering avec la librairie *Optuna*<sup>24</sup>, une recherche optimale des paramètres d'entraînement du modèle d'apprentissage est effectuée. En effet, la librairie *Optuna*, qui utilise des méthodologies de recherche empiriques optimisées pour réduire l'espace de recherche, permet la réalisation de plusieurs tests consécutifs en variant certains paramètres prédéfinis, tels que le nombre de clusters, la méthode d'initialisation, la taille des lots, le nombre maximal d'itérations et la tolérance.

Chaque essai réalise le calcul d'un score de performance défini par une fonction objective définie pour optimiser les performances en maximisant le *silhouette score* et le *Calinski-Harabasz Index* et en minimisant le *Davies-Bouldin Index*. La pondération de chaque métrique est également recherchée de façon empirique pour identifier la configuration des paramètres qui réalisent le meilleur compromis.

$$score = (x * silhouette\_avg) + (y * ch\_index) - (z * db\_index)$$

$x$  : pondération du silhouette score

$silhouette\_avg$  : silhouette score mesuré

$y$  : pondération du Calinski-Harabasz Index

$ch\_index$  : Calinski-Harabasz Index mesuré

$z$  : pondération du Davies-Bouldin Index

$db\_index$  : Davies-Bouldin Index mesuré

Le code source de la fonction objective est disponible en [Annexe 6 : Fonction Objective](#).

#### 4.4.3 Sélection du modèle d'apprentissage

Après recherche de paramètres optimaux, le modèle *Mini Batch K-MEANS* s'avère être le modèle le plus adapté au contexte de l'étude par rapport au modèle *HDBSCAN*.

En effet, le développement du modèle *HDBSCAN* n'as pas été approfondi puisque vis-à-vis du grand volume de données nécessaire à traiter, le temps d'entraînement s'est révélé conséquent. De plus, les essais sur un sous-échantillon n'ont pas démontré de performances satisfaisantes en termes de *Silhouette Score*, avec une valeur maximale de 0,0856.

En revanche, le modèle *K-MEANS* après de nombreux essais, permet d'obtenir de bonnes performances. En effet, il atteint un *Davies-Bouldin Index* de 0,36 indiquant une relativement faible similarité entre les clusters, ce qui est souhaitable pour une bonne séparation des données. De plus, le *Calinski-Harabasz Index* élevé de 20 025 853,87 souligne une forte cohésion interne des clusters par rapport à leur dispersion globale. Enfin, le *Silhouette Score* moyen de 0,76 témoigne d'une bonne adéquation des points de données à leurs clusters respectifs, confirmant ainsi l'efficacité du modèle *K-MEANS* dans ce contexte d'étude.

Récapitulatif des résultats obtenus avec le modèle *K-MEANS* :

- Davies-Bouldin Index: 0.3619108287721344,
- Calinski-Harabasz Index: 20025853.87041033,
- Silhouette Score: 0.7608796914931638.

---

<sup>24</sup> Optuna : bibliothèque d'optimisation hyperparamétrique pour automatiser la recherche des meilleurs paramètres dans les modèles de machine learning.

## 4.5 Exploitation du modèle d'apprentissage

### 4.5.1 Répartition des utilisateurs par cluster

Après entraînement et optimisation du modèle de clustering, la classification des utilisateurs en quatre clusters distincts et l'analyse de leurs spécificités est réalisée. Chaque utilisateur est donc attribué à l'une des 4 catégories définies automatiquement par des caractéristiques dites similaires.

On observe notamment que la distribution du nombre d'utilisateurs par cluster est variable, ce qui indique des comportements bien spécifiques pour chaque catégorie. Par exemple, le cluster n°1 représente près de 41% des utilisateurs tandis que le cluster n°2 ne représente qu'un peu plus de 2 % des utilisateurs (voir figure 9 ci-dessous).

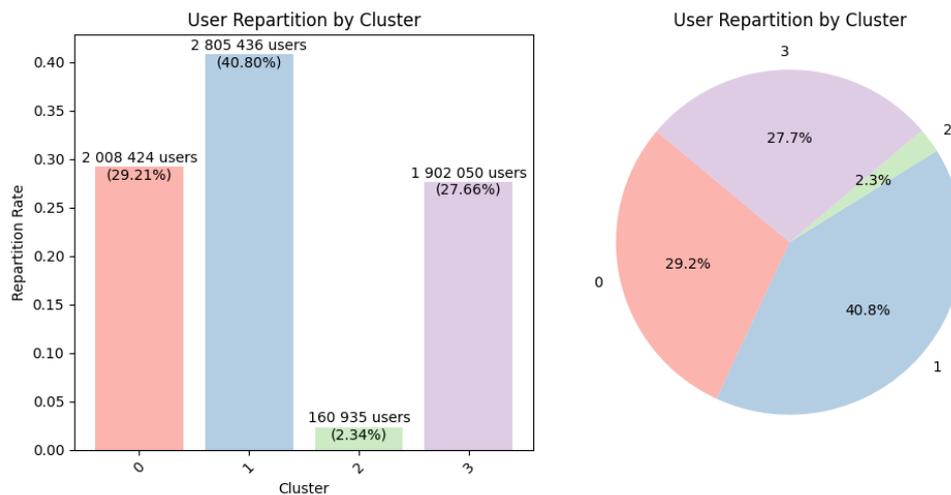


Figure 9: Répartition des utilisateurs en 4 clusters distincts par comportements similaires.

### 4.5.2 Différenciation des utilisateurs par usages

L'un des principaux points de différenciation des clusters repose sur la diversité d'utilisation des protocoles et des types de protocoles. Par exemple, le cluster n°2 se différencie par une utilisation majoritaire de protocoles de prêts (Lending) par rapport aux autres clusters dont l'utilisation principale sont les stablecoins (voir figure 10 ci-dessous).

Il est également observable que le taux d'utilisation du protocole de type NFT est très faible et reste marginal pour chacun des clusters. En revanche, les protocoles de type échanges (DEX) représentent pour chacun des clusters la seconde catégorie majoritaire, ce qui sous-tend qu'il s'agit du second usage principal adopté par les utilisateurs de plateformes de finances décentralisées.

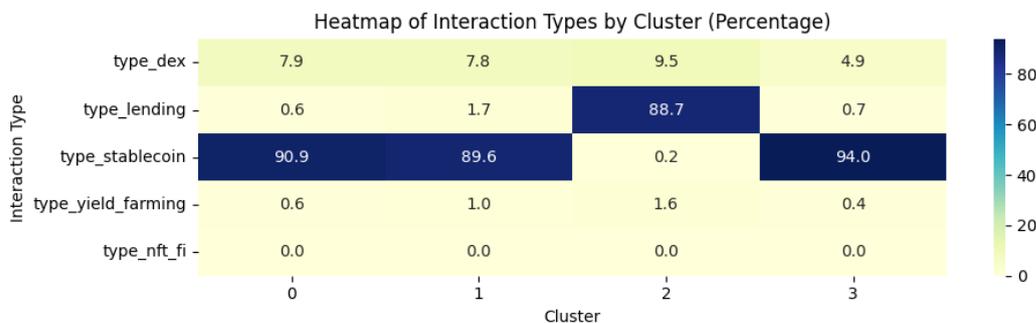


Figure 10: Taux d'utilisation des types de protocoles par cluster.

En regardant plus précisément quels sont les protocoles spécifiquement les plus adoptés, le protocole *Tether* occupe la première position en termes de stablecoins et le protocole *USDC* la seconde, avec un taux d'utilisation approximativement similaire pour les clusters qui utilisent les stablecoins. Le protocole *DAI* en troisième position est plus important pour le cluster 1 ce qui indique un indice de différenciation pour le cluster n°1 parmi les clusters qui utilisent les stablecoins. (voir figure 11 ci-dessous).

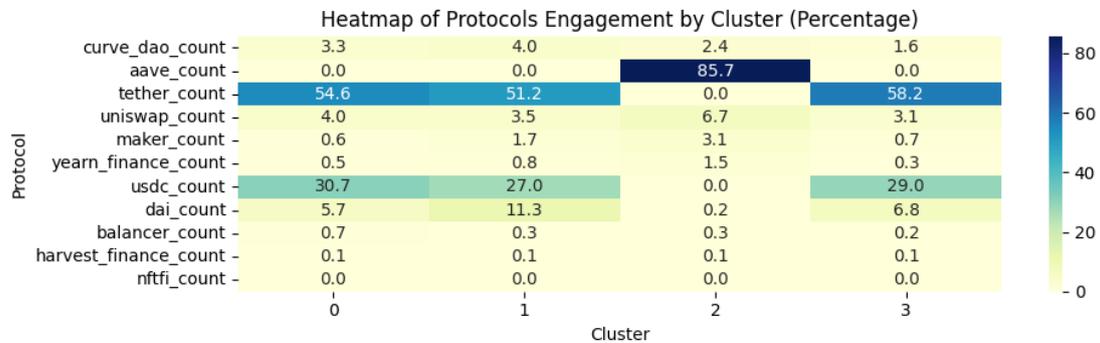


Figure 11: Taux d'utilisation des protocoles spécifiques par cluster.

Pour une représentation visuelle de la répartition des utilisations, voir [Annexe 7 : Interaction par protocoles et types de protocoles par cluster](#).

#### 4.5.3 Différenciation des utilisateurs par comportements

Pour développer l'analyse de différenciation des comportements des utilisateurs dans l'écosystème de la finance décentralisée, plusieurs graphiques mettant en valeur les minimums, maximums, moyennes et médianes par caractéristique permet d'observer les comportements les plus distincts entre clusters.

Par exemple, en observant les caractéristiques d'activités, soit les nombres de transactions et valeurs envoyées et reçues, on observe que le cluster n°0 ne réalise pas d'envoi et que le cluster n°3 ne réalise pas de réception, ce qui indique un point de distinction. De plus, le cluster n°2 se distingue par une valeur transférée nettement plus importante (voir figure 12 ci-dessous).

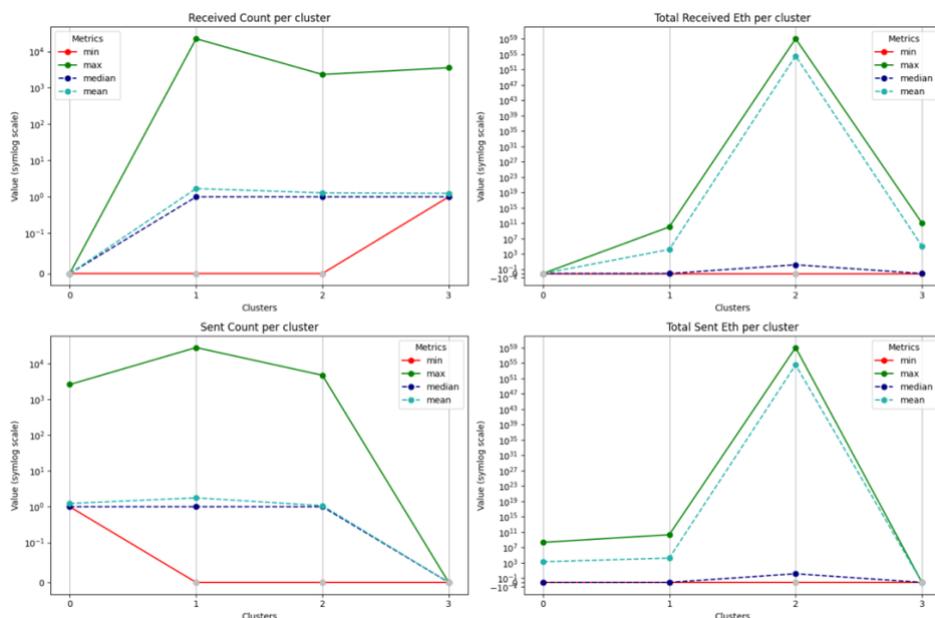


Figure 12 : Caractéristiques d'activités par cluster.

L'ensemble des graphiques de dissociation des comportements sont disponibles dans l'application hébergée sur *HuggingFace* dans la section 'Clustering\_'. L'accès à l'application est disponible en [Annexe 8 : Partage de l'étude en open-source, B\) Accès à la collection hébergée sur HuggingFace, a. Application Streamlit de présentation.](#)

#### 4.5.4 Caractéristiques principales de distinction

En observant les corrélations entre les caractéristiques et les clusters (voir figure 13 ci-dessous), il est possible de déduire quelles sont les caractéristiques les plus influentes dans la détermination de chaque catégorie d'utilisateurs :

- Le cluster n°0 est principalement défini par les frais de transactions des réceptions ainsi que par les heures de pointes (qui sont fortement corrélés par le fonctionnement même de la blockchain) et par le fait que les utilisateurs ne réalisent pas de transactions d'envoi.
- Le cluster n°1 est principalement défini par les frais de transactions également mais il est distinguable des autres clusters par le nombre de transactions envoyées et reçues indiquant une forte activité ainsi que par les schémas d'utilisation de stablecoins.
- Le cluster n°2 est celui dont l'utilisation diffère le plus des autres clusters. Il est distinguable par la diversité de protocole et de types de protocoles utilisés relativement faible, ainsi que par l'exposition à la volatilité et à la liquidité.
- Le cluster n°3 est principalement défini par les mêmes caractéristiques que le cluster n°0 mais ne réalisant pas de transactions de réception, il en est en fait l'opposée symétrique.

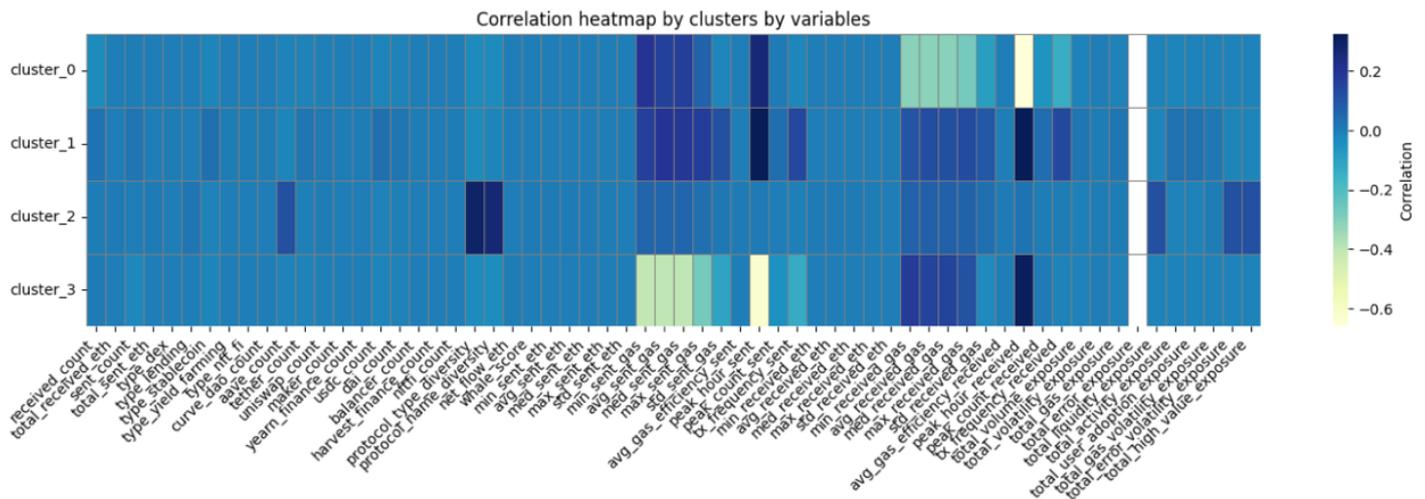


Figure 13 : Corrélation entre les caractéristiques et les clusters.

## 5. Interprétation des Résultats

### 5.1 Validité et fiabilité

L'assurance de la validité et de la fiabilité des résultats est essentielle pour garantir la robustesse des conclusions tirées dans cette étude. Les mesures de validité et de fiabilité mises en place visent à minimiser les biais, à renforcer la rigueur méthodologique et à permettre une vérification indépendante des résultats.

#### 5.1.1 Mesures de validité

Dans un premier temps, la validité des résultats dans cette étude a été recherchée par l'emploi d'une approche méthodologique. Tout d'abord, les données ne sont pas issues d'un jeu de

données préconçu mais ont été extraites directement auprès de sources fiables, notamment *Etherscan*, l'API officielle de la blockchain *Ethereum* pour extraire les données de transactions et l'API *YahooFinance*, reconnue dans le domaine pour extraire les données de marchés.

Dans un second temps, la qualité de l'apprentissage non-supervisé a été contrôlée par le calcul de métriques de performance (*Inertie*, *Silhouette score*, *Davies-Bouldin Index*, *Calinski-Harabasz Index*), ce qui a permis d'évaluer la qualité des clusters. De plus, l'optimisation des hyperparamètres via la librairie *Optuna* a été employée afin de rechercher le meilleur compromis possible de réglage du modèle d'apprentissage.

### 5.1.2 Mesures de fiabilité

Concernant la fiabilité des données, une analyse exploratoire approfondie a été menée pour comprendre les relations entre les variables et l'évolution des ordres de grandeur dans le temps. Par exemple, les valeurs semblant aberrantes ont été vérifiées unitairement dans l'API *Etherscan* pour en assurer la conformité. La distribution des données a également représenté un point important de l'analyse afin d'atténuer du mieux que possible les biais de déduction dus à des échelles différentes, notamment la définition d'une méthode de standardisation sur-mesure.

Travaillant seul sur ce projet, le code source, le processus de constitution du jeu de données, les étapes de traitement ainsi que les résultats obtenus ont été documentés de manière détaillée et sont accessibles en open source, ce qui permet une vérification indépendante des résultats et la réception d'éventuels retours de la communauté.

Les accès au code source, au jeu de données et à l'application de présentation sont accessibles en [Annexe 8 : Partage de l'étude en open-source](#).

## 5.2 Limites méthodologiques

La reconnaissance des limites méthodologiques est cruciale pour une interprétation équilibrée des résultats. Cela permet d'identifier les contraintes et les biais potentiels qui pourraient affecter la validité et la généralisation des conclusions.

### 5.2.1 Limites identifiées

Dans un premier temps, dans l'objectif d'avoir une vue d'ensemble exhaustive de la finance décentralisée, de nombreuses données ont été assemblées pour disposer de plusieurs protocoles sur plusieurs années afin d'observer un premier effet de cycle. Cependant, les cycles dans l'écosystème DeFi peuvent ne pas spécifiquement être réalisés en 2 années mais pourraient être plus long. De plus, l'exploitation d'un tel volume de données (plus de 6,8 millions d'observations contenant plus de 22 millions de transactions) pourrait être approfondie avec des ressources computationnelles plus importantes qui permettraient l'utilisation d'algorithmes différents afin de croiser les résultats.

Dans un second temps, le périmètre défini pour cette étude reste minime et ne représente pas la totalité de l'ensemble immense de l'écosystème de la finance décentralisée. Cette étude se concentre sur des protocoles et des types de protocoles populaires mais pourrait ne pas refléter l'ensemble des cas d'usages de l'écosystème DeFi. De plus, la sélection de la blockchain *Ethereum* seule comme unique source de données pourrait introduire un biais, car d'autres blockchains pourraient avoir des dynamiques différentes étant donné leur logique de fonctionnement.

Enfin, la nature complexe et interconnectée des données DeFi peut rendre difficile l'identification de relations causales claires entre les variables. En effet, le fait que certains utilisateurs dans cette étude ne contiennent pas de métadonnées d'envois ou de réception

pourraient être simplement dû au fait qu'ils n'aient pas interagis avec les protocoles spécifiques de l'étude mais avec d'autres protocoles de l'écosystème.

### 5.2.2 Stratégies d'atténuation

Pour atténuer les problèmes liés au volume de données, des sous-échantillons aléatoires ont été utilisés pour tracer les graphes de réseaux et tester les algorithmes de clustering, notamment l'algorithme *HDBSCAN*. Cette stratégie a permis de confirmer les patterns observables visuellement et permis de rejeter l'algorithme *HDBSCAN* pour des raisons de performance.

En complément, des analyses de sensibilité ont été réalisées pour évaluer l'impact des différentes méthodes de standardisation. En effet, une version du clustering a été effectuée en omettant des valeurs conformes mais dites extrêmes et les métriques se sont avérées moins performantes. Les valeurs extrêmes ont donc été conservées dans le clustering final afin de représenter l'écosystème au plus juste.

## 5.3 Comparaison avec les hypothèses

Après analyse des comportements des utilisateurs dans les applications DeFi, les hypothèses initialement définies dans le paragraphe '2.4 Cadre Conceptuel de l'Étude' sont confirmées ou rejetées.

### 5.3.1 Première hypothèse

**'Les incitations économiques, telles que les rendements élevés du yield farming, motivent les utilisateurs à s'engager davantage dans les protocoles DeFi.'**

Si on observe le second graphe de réseau (figure 5) qui démontre les flux entre les différents types de protocoles, nous pouvons en déduire que les stablecoins représentent le principal point d'entrée ou de sortie de la DeFi qui amène à l'utilisation de protocoles d'échanges (DEX). A la suite de ces protocoles d'échanges, les protocoles de rendement et d'emprunts (yield farming et lending) semblent représenter la troisième hiérarchie logique de la suite de l'adoption des protocoles de finance décentralisée.

En complément, si on s'intéresse aux taux d'utilisation des types de protocoles par cluster (figure 10), on observe que le yield farming est utilisé par chacun des clusters mais qu'il n'occupe que la quatrième position sur cinq des types de protocoles utilisés derrière les emprunts, les échanges et les stablecoins respectivement en troisième, seconde et première position.

Les données de cette étude démontrent donc que le yield farming est un usage adopté mais que les gains sont principalement recherchés par les utilisateurs via les protocoles d'échanges (DEX). De plus, d'après l'analyse de tendance (Annexe 4), son taux d'utilisation est instable et ne démontre pas de tendance croissante, l'hypothèse n°1 est donc rejetée.

### 5.3.2 Seconde hypothèse

**'Les avantages proposés par la finance décentralisée rendent obsolètes les services financiers traditionnels.'**

D'après l'analyse de tendance (Annexe 4) le taux de transactions effectuées sur les années 2023 et 2024 ainsi que le taux d'utilisation de l'usage de stablecoins est croissant. Cela démontre que la finance décentralisée est de plus en plus active et que l'usage de jetons stables indexés aux monnaies traditionnelles est de plus en plus adopté.

Dans un système économique en tension, la confiance envers les institutions semble faiblir au détriment de la confiance en un système algorithmique dit décentralisé. Dans les pays émergents, les paiements rendus possibles grâce aux smartphones décuplent également ce

phénomène. Les institutions semblent donc être obligées de s'adapter face à cette évolution technologique. C'est d'ailleurs ce que démontre un article du média *The Big Whale* qui évoque que la banque française BPCE va permettre à ses clients d'investir dans les cryptoactifs en 2025 [14]. En complément, un article du média *Cryptoast* évoque que le président du marché Nasdaq dévoile un projet d'ouverture 24 heures sur 24 de la bourse américaine pour s'aligner sur les nouveaux usages rendus possibles par l'évolution de la technologie [15].

L'hypothèse de la finance traditionnelle qui apparaît désormais comme obsolète vis-à-vis de la finance décentralisée et des nouveaux usages associés est donc confirmé. Cependant, il est important de souligner qu'il est peu probable que la finance traditionnelle soit totalement remplacée par la DeFi. La finance traditionnelle serait simplement amenée à évoluer afin de former un seul écosystème complet équilibrant de façon complexe les avantages et les inconvénients que sont les frais de transactions, la disponibilité, la sécurité ou encore la confiance.

### 5.3.3 Troisième hypothèse

**‘La courbe d'apprentissage pour comprendre et utiliser les services DeFi peut être raide, décourageant ainsi les utilisateurs moins familiers avec la technologie. La théorie de diffusion des innovations identifie la complexité comme un facteur influençant l'adoption.’**

D'après l'analyse des tendances (Annexe 4), le nombre d'utilisateurs (+ 4,84% en 2 ans), ainsi que le nombre de transactions effectuées (+18,28% en 2 ans) pour les protocoles de type échanges (DEX) et de type emprunts (Lending) ont tendance à croître avec le temps. En revanche, le taux d'utilisation des protocoles de type rendement (yield farming) présente une instabilité sur les deux années de l'étude et ne présente pas de croissance spécifique. Cela pourrait être interprété comme une certaine réticence de la part des utilisateurs qui préféreraient se concentrer sur des protocoles de type échange ou prêt. De la même façon, le protocole de type NFT est très peu utilisé voir ne présente pas d'usage du tout.

Pour compléter cette analyse, en observant les taux d'utilisation par type de protocole (voir figure 14 ci-dessous), on observe que les taux d'utilisation par protocole sont directement induits par la complexité de ceux-ci. Plus le type de protocole est complexe, moins son usage est adopté.

type	users	rate
Stablecoin	6,282,798	91.31
DEX	492,244	7.15
Lending	276,457	4.02
Yield Farming	46,677	0.68
NFT-Fi	390	0.01

Figure 14: Taux d'utilisation par type de protocole

En effet, les stablecoins représentent ce qui est de plus proche d'une monnaie traditionnelle mais sans une institution centrale, les échanges (DEX) représentent ce qui est le plus proche du trading traditionnel mais sans valeur liée à quelque chose de matériel. Le lending représente ce qu'il y a de plus proche des emprunts traditionnels mais sans conseiller bancaire associé. Le yield farming représente ce qu'il y a de plus proche d'une banque coopérative et mutualiste, dont les intérêts sont orientés de façon à satisfaire à la fois les clients et les propriétaires, mais

sans conseil d'administration physique. Enfin, les NFT représentent ce qu'il y a de plus proche d'une œuvre d'art, un élément certes reproductible mais initialement créé par un auteur authentifiable, mais sans expert d'authentification. Ainsi, il pourrait être déduit que plus le niveau d'abstraction nécessaire pour adopter la technologie est important, moins l'usage est adopté, l'hypothèse n°3 est donc confirmée.

#### 5.3.4 Quatrième hypothèse

**‘Les effets de réseau positifs renforcent la confiance et l’engagement des utilisateurs et a contrario, les effets de réseau négatifs diminuent l’engagement des utilisateurs dans l’écosystème DeFi.’**

D'après l'analyse de corrélation (Annexe 3), il est observable qu'il existe une forte corrélation positive entre le nombre d'utilisateurs et le nombre de transactions effectuées. De plus, d'après l'analyse de tendance (Annexe 4), on observe que le nombre d'utilisateurs global sur le périmètre étudié à tendance à croître avec le temps, de la même façon que la valeur transférée, même s'il n'existe pas de corrélation entre ces deux composantes. Ainsi, nous pouvons déduire avec réserve que l'effet de réseau positif est confirmé puisque plus il y a d'utilisateurs, plus l'écosystème est actif mais que d'autres éléments, probablement sociaux interviennent dans cet effet.

Concernant l'effet de réseau négatif, il est important de souligner qu'il existe également une forte corrélation positive entre le montant des frais de transactions et le nombre d'utilisateurs actifs sur une période. Ainsi, nous pouvons également déduire que l'effet de réseau négatif est confirmé puisque plus il y a d'utilisateurs actifs, plus le risque d'abandon de l'usage augmente du fait de l'augmentation des frais de fonctionnement. En effet, sur la période étudiée, le ratio entre augmentation des frais de transaction et augmentation du nombre d'utilisateur est d'environ 37% ce qui signifie qu'une augmentation de 5 % d'utilisateurs engendre une augmentation de 1,85 % des frais de transactions.

L'hypothèse n°3 est donc confirmée avec réserve. En effet, les éléments quantitatifs de l'étude démontrent bien des effets de réseaux mais une approche complémentaire permettrait de déduire avec précision quels en sont facteurs spécifiques.

### 5.4 Implications des résultats

#### 5.4.1 Implications théoriques

Les résultats de cette étude pourraient apporter plusieurs contributions à la théorie existante sur le sujet.

Tout d'abord, si on s'intéresse aux facteurs économiques, psychologiques et technologiques qui influencent l'engagement des utilisateurs, la confirmation partielle des hypothèses initiales souligne l'importance de la complexité perçue par les utilisateurs et les incitations économiques dans le processus d'adoption, ce qui est cohérent avec la théorie de la diffusion des innovations de Rogers [7].

De plus, cette étude met en évidence les effets de réseau positifs et négatifs au sein de l'écosystème DeFi. Les résultats démontrent que l'augmentation du nombre d'utilisateurs peut renforcer la confiance et l'engagement, mais peut également entraîner des coûts de transaction plus élevés, ce qui pourrait décourager certains utilisateurs. Ces observations contribuent à la théorie des réseaux et des effets de réseau en fournissant des preuves spécifiques au domaine de la finance décentralisée [8].

Enfin, cette analyse de comportements inter-protocoles révèle des schémas d'utilisation complexes qui ne peuvent être entièrement expliqués par les théories existantes. Cela ouvre de

nouvelles pistes de recherches pour explorer les interactions entre différents types de protocoles et leurs impacts sur l'adoption et la rétention des utilisateurs.

### 5.4.2 Implications pratiques

Les implications pratiques de cette étude sont multiples et complètent les recherches précédentes évoquées lors de la revue de la littérature.

Tout d'abord, cette étude met en lumière l'importance de la diversification au sein de l'écosystème DeFi pour encourager une adoption plus large. En effet, en complément de l'étude '*Onchain Analysis*' [13], cette analyse démontre qu'au sein d'une même blockchain, les utilisateurs naviguent entre différents types de protocoles (DEX, plateformes de prêt, stablecoins, etc.) ce qui confirme qu'il est essentiel de développer l'interopérabilité afin de réduire les frais de transactions ou la création de pool de liquidité partagés pour améliorer l'attractivité des plateformes. Cet argument est notamment renforcé par l'étude '*User behavior and token adoption on ERC20*' [11] qui évoque l'importance des connexions entre communautés pour renforcer la résilience du marché.

En complément de l'étude '*Bridging the gap for cryptocurrencies*' [10], bien que les points d'attraits partagés soient les incitations économiques et la tolérance au risque perçu, cette présente analyse met en perspective cette fois-ci les points de répulsions que sont la complexité et le niveau d'abstraction nécessaire à l'adoption des usages. Ainsi, la présente étude suggère que les incitations économiques bien que motivantes pour les utilisateurs, sont à relativiser par rapport à l'importance de la mise en avant d'autres atouts de la DeFi que sont la décentralisation et la sécurité, ce qui pourrait notamment renforcer la tolérance au risque perçu par les utilisateurs. Cela suggère que les plateformes DeFi doivent investir dans des interfaces utilisateurs intuitives et des outils éducatifs pour faciliter l'adoption des usages autres que les stablecoins et les échanges (DEX).

Enfin, les effets de réseau positifs et négatifs observés dans cette étude ont des implications directes pour la gestion des plateformes DeFi. Une augmentation du nombre d'utilisateurs peut renforcer la confiance et l'engagement, mais elle peut également entraîner des coûts de transactions plus élevés. Les plateformes doivent donc trouver un équilibre entre croissance et gestion des coûts pour maintenir un écosystème sain et attractif.

En complément des études précédentes, cette recherche offre donc une vision plus holistique des comportements des utilisateurs dans la DeFi, en mettant en évidence les interactions inter-protocoles complexes et les facteurs influençant la navigation dans l'écosystème.

## 6. Enrichissement de l'Étude

### 6.1 Perspectives d'approfondissement

Pour compléter cette étude et comprendre davantage les schémas de comportements des utilisateurs dans les plateformes de finance décentralisée, plusieurs aspects pourraient être approfondis.

Premièrement, pour combler les lacunes de la présente étude, il pourrait être intéressant de réaliser une analyse à la fois inter-protocole mais également inter-chaîne. Les observations réalisées sur un périmètre plus réaliste permettraient en effet de comprendre plus précisément quels sont les types de stratégies adoptées par les utilisateurs et les facteurs influençant leur navigation dans l'écosystème.

Sur un périmètre moins large, pour comprendre plus précisément les avantages et les inconvénients que propose la DeFi, il serait également intéressant d'étudier en profondeur quels sont les tenants et les aboutissants de la stratégie adoptée par le cluster n°2. En effet, ce cluster

représente une portion très faible des utilisateurs et les comportements associés sont très distinctifs des autres clusters. En observant l'amplitude des revenus nets générés par le cluster 2, on observe que ce cluster a l'amplitude la plus faible ce qui démontre une gestion des risques accrue. C'est également le cluster qui contient le plus d'utilisateurs influents sur le marché (dit whales). Comprendre la stratégie des utilisateurs de ce cluster permettrait donc de comprendre les attraits des utilisateurs pour des protocoles plus complexes.

## 6.2 Perspectives technologiques

La DeFi représentant un écosystème vaste et complexe, l'évolution de la technologie au sens large pourrait également influencer les comportements des utilisateurs.

En effet, la création d'outils basés sur l'intelligence artificielle permettrait de capter des signaux faibles dans l'environnement, c'est d'ailleurs ce que démontre l'étude '*Behavior-aware Account De-anonymization Ethereum Interaction Graph*' [12] qui propose un réseau de neurone permettant de détecter des anomalies et ainsi de sécuriser l'écosystème. L'intelligence artificielle pourrait également permettre de proposer aux utilisateurs des fonctionnalités d'analyse et de suggestions.

En complément, la stabilisation de l'informatique quantique pourrait à terme, proposer des ressources computationnelles jusqu'alors inégalables, ce qui pourrait décupler les capacités de surveillance et d'analyse et ainsi permettre de capter des signaux encore plus faibles sur des données massives. Cela représente également un risque de complexification de l'écosystème qui n'apporterait des avantages qu'à certains utilisateurs dédiés et pourrait décourager des utilisateurs plus traditionnels.

## 6.3 Recommandations pour les acteurs de la DeFi

La présente étude en complément des précédentes, confirme que la complexité de compréhension du fonctionnement de la finance décentralisée est un frein majeur à l'adoption de l'usage. Ainsi, accompagner des utilisateurs novices pour les aider à adopter un comportement performant dans l'écosystème permettrait de développer celui-ci. En complément de l'étude, une fonctionnalité de recommandations utilisateur est donc proposée et mise en œuvre.

### 6.3.1 Suggestion

Partant du principe que la perfection n'existe pas, un système de recommandations de comportements basé sur des comparaisons entre utilisateurs sur des caractéristiques clés, pourrait proposer une analyse des forces et faiblesses et ainsi faire tendre les usages vers un comportement plus vertueux. Partant également du principe qu'un cluster est associé à une stratégie spécifique, une comparaison entre tous les utilisateurs et une comparaison entre tous les utilisateurs d'un même cluster permettrait à un utilisateur d'affiner sa propre stratégie.

### 6.3.2 Preuve de concept

Pour réaliser un système de recommandation basé sur une comparaison des forces et des faiblesses d'un utilisateur par rapport aux autres, les étapes suivantes sont mises en œuvre :

- Étape 1 : définition de métriques de comparaison ;
- Étape 2 : définition d'un rang utilisateur par métrique ;
- Étape 3 : génération de graphiques radar pour observer les performances globales ;
- Étape 4 : génération de recommandations.

Il est important de souligner qu'il s'agit d'une preuve de concept pour démontrer la faisabilité d'un système de recommandations. Le système sera donc basé sur le périmètre de l'étude et ne sera donc pas disponible pour des protocoles hors périmètre.

#### 6.3.2.1 Définition de métriques de comparaison

Pour effectuer une comparaison entre utilisateurs, les métriques clés suivantes sont définies :

- **Return on Investment (ROI)** : Le ROI est une mesure de la rentabilité d'un investissement. Une valeur élevée indique une forte rentabilité, tandis qu'une valeur faible suggère une faible rentabilité ou une perte.
- **Activity Score** : Ce score reflète le niveau d'activité de l'utilisateur. Une valeur élevée indique une participation active, tandis qu'une valeur faible suggère une inactivité.
- **Interaction Diversity** : Cette métrique évalue la diversité des interactions de l'utilisateur. Une valeur élevée indique des interactions variées, tandis qu'une valeur faible suggère une diversité limitée.
- **Engagement Diversity** : Cette métrique mesure la diversité des types d'engagement. Une valeur élevée indique un engagement large, tandis qu'une valeur faible suggère un engagement limité.
- **Sending Behavior** : Cette métrique évalue le comportement de transaction d'envoi. Une valeur élevée indique un envoi actif et varié, tandis qu'une valeur faible suggère une inactivité.
- **Sending Fee Efficiency** : Cette métrique mesure l'efficacité des frais de transaction d'envoi. Une valeur élevée indique une utilisation efficace des frais de transactions tandis qu'une valeur faible suggère une inefficacité.
- **Receiving Behavior** : Cette métrique évalue le comportement de transaction de réception. Une valeur élevée indique une réception active et variée, tandis qu'une valeur faible suggère une inactivité.
- **Receiving Fee Efficiency** : Cette métrique mesure l'efficacité des frais de transaction de réception. Une valeur élevée indique une utilisation efficace des frais de de transactions, tandis qu'une valeur faible suggère une inefficacité.
- **Global Fee Efficiency** : Cette métrique évalue l'efficacité globale des frais. Une valeur élevée indique une utilisation efficace des frais de transactions, tandis qu'une valeur faible suggère une inefficacité.
- **Frequency Efficiency** : Cette métrique mesure l'efficacité de la fréquence des transactions. Une valeur élevée indique une fréquence optimale, tandis qu'une valeur faible suggère une irrégularité.
- **Timing Efficiency** : Cette métrique évalue l'efficacité du timing des transactions. Une valeur élevée indique un timing optimal (hors périodes de pointe), tandis qu'une valeur faible suggère une inefficacité (périodes de pointe).
- **Global Market Exposure Score** : Ce score mesure l'exposition globale au marché. Une valeur élevée indique une forte exposition, tandis qu'une valeur faible suggère une exposition limitée.
- **Risk Index** : Cet indice mesure le niveau de risque. Une valeur élevée indique un faible risque, tandis qu'une valeur faible suggère un risque élevé.

- **Opportunity Score** : Ce score évalue les opportunités potentielles. Une valeur élevée indique de nombreuses opportunités, tandis qu'une valeur faible suggère peu d'opportunités.
- **Performance Index** : Cet indice mesure la performance relative au risque. Une valeur élevée indique une haute performance, tandis qu'une valeur faible suggère une faible performance.
- **Adoption Activity Score** : Ce score mesure l'activité d'adoption des utilisateurs. Une valeur élevée indique une forte adoption, tandis qu'une valeur faible suggère une activité limitée.
- **Stability Index** : Cet indice mesure la stabilité. Une valeur élevée indique une forte stabilité, tandis qu'une valeur faible suggère une volatilité.
- **Volatility Exposure** : Cette métrique évalue l'exposition à la volatilité du marché. Une valeur élevée indique une faible exposition, tandis qu'une valeur faible suggère une exposition élevée.
- **Market Influence** : Cette métrique mesure l'influence sur le marché. Une valeur élevée indique une forte influence, tandis qu'une valeur faible suggère une influence limitée.
- **Global Score** : Ce score global évalue la performance globale, il est calculé par pondération à partir des autres métriques. Une valeur élevée indique une forte performance, tandis qu'une valeur faible suggère une faible performance.

#### 6.3.2.2 Définition d'un rang utilisateur par métrique

Pour définir le rang d'un utilisateur afin d'effectuer une comparaison, les métriques décrites précédemment sont d'abords calculées. Pour disposer de données sur une même échelle, les métriques sont par la suite normalisées pour ramener les valeurs dans un intervalle [0,1]. Enfin, chaque métrique de chaque utilisateur est classée par rapport aux autres métriques des autres utilisateurs et le rang est ensuite converti en pourcentage. Ainsi, chaque utilisateur est associé à des métriques de comportement exprimées en pourcentage de performance relative.

#### 6.3.2.3 Visualisation des performances d'un utilisateur

Afin de proposer un visuel intuitif à l'utilisateur, deux graphiques radar sont générés. Le premier permet de comparer les métriques par rapport à tous les utilisateurs et le second permet de comparer les métriques par rapport aux utilisateurs d'un même cluster. Sur les graphiques radars, une courbe représentant la médiane basse est tracée en rouge et une courbe représentant la médiane haute est tracée en vert. Les performances de l'utilisateur sont représentées sous forme d'aire dont l'ampleur dépend de la valeur de chaque métrique.

Un exemple est disponible en [Annexe 9 : Exemples de graphiques de performances](#).

#### 6.3.2.4 Génération de recommandations

Dans l'objectif de suggérer des recommandations personnalisées, une interface utilisateur permettant de renseigner une adresse *Ethereum* permet de lancer l'exécution d'un rapport de performance associé à l'utilisateur de l'adresse. Lors de l'exécution, un prompt<sup>25</sup> est envoyé à un modèle large de langage afin de générer du texte. Pour orienter le modèle de langage à

---

<sup>25</sup> Prompt : instruction ou question transmise à un modèle d'intelligence artificielle pour générer une réponse ou effectuer une tâche spécifique.

générer un rapport de recommandation, le prompt est décomposé en un prompt de système et un prompt utilisateur.

Le prompt système définit les instructions à suivre lors de chaque exécution et instruit une structure logique de contexte. Le modèle est défini comme un analyste de données spécialisé dans l'évaluation des performances utilisateurs et sa tâche consiste à analyser les données de performance afin de produire un rapport détaillé. Il est demandé au modèle de se familiariser avec les métriques, examiner les valeurs et effectuer des comparaisons. Enfin, une structure de rapport est démontrée en exemple.

Le prompt système est disponible en [Annexe 10 : Prompt système de recommandation](#).

Le prompt utilisateur est défini de façon à compléter le prompt système en effectuant une requête de rapport et en intégrant les données associées à l'adresse *Ethereum* renseignée. Les données contiennent le nom, la description, le rang global et le rang de l'utilisateur dans le cluster pour chacune des métriques présentées dans le paragraphe 6.3.2.1 '*Définition de métriques de comparaison*'.

Le prompt utilisateur est disponible en [Annexe 11 : Prompt utilisateur de recommandation](#).

### 6.3.3 Résultats de la preuve de concept

Le système de recommandation décrit précédemment fonctionne correctement, le modèle de langage est capable de générer de façon stable et plus ou moins répétable des recommandations pour un utilisateur.

Les recommandations sont cohérentes, les forces et les faiblesses sont correctement identifiées et le rapport ainsi généré permet d'identifier des axes d'améliorations pour tout type d'utilisateur et de stratégie. Cette fonctionnalité ainsi démontrée pourrait être intégrée à des plateformes de finance décentralisée afin d'accompagner les utilisateurs à développer des comportements plus stables et plus performants.

L'expérience de recommandation est disponible dans l'application hébergée sur *HuggingFace* dans la section '*Performance Report*'. L'accès à l'application est disponible en [Annexe 8 : Partage de l'étude en open-source, B\) Accès à la collection hébergée sur HuggingFace, a. Application Streamlit de présentation](#).

## 7. Conclusion Générale

### 7.1 Synthèse des principaux résultats

Cette étude a permis de mettre en lumière plusieurs aspects fondamentaux concernant les comportements des utilisateurs dans les applications de finance décentralisée (DeFi).

Concernant la validité et la fiabilité des données, les résultats obtenus ont été rigoureusement validés par l'utilisation de données provenant de sources fiables. Des métriques de performances ont été calculées afin d'évaluer la qualité des clusters formés par l'algorithme de clustering et une optimisation des hyperparamètres a été réalisée pour affiner le modèle d'apprentissage.

Une analyse exploratoire a permis de comprendre les relations et les tendances des variables et une approche de standardisation sur-mesure a été mise en œuvre pour atténuer les biais potentiels de déduction.

Bien que l'étude ait couvert un large éventail de données, elle reste limitée par le volume computationnel disponible et par la sélection de la blockchain *Ethereum* uniquement. Les cycles de marchés dans l'écosystème DeFi peuvent être plus longs que les deux années étudiées, et

d'autres blockchains pourraient présenter des dynamiques différentes. De plus, la complexité et l'interconnexion des données rendent difficile l'identification de relations causales claires entre les variables.

Les hypothèses initiales ont été partiellement confirmées ou rejetées en fonction des analyses réalisées. Par exemple, bien que les protocoles de rendements (Yield Farming) soient un usage adopté, les gains sont principalement recherchés par les protocoles d'échanges (DEX). De plus, la finance décentralisée montre des signes de croissance mais il est peu probable qu'elle remplace totalement la finance traditionnelle.

Cette étude présente plusieurs implications pratiques qui enrichissent les recherches antérieures. Elle démontre que les utilisateurs naviguent entre différents types de protocoles au sein d'une même blockchain, ce qui met en évidence la nécessité de développer l'interopérabilité pour réduire les frais de transaction et créer des pools de liquidité partagés pour améliorer l'attractivité des plateformes.

Cette conclusion est corroborée par les études précédentes qui soulignent l'importance des connexions entre communautés pour renforcer la résilience du marché. Par ailleurs, bien que les incitations économiques et la tolérance au risque soient des facteurs d'attraction, cette étude met en lumière les obstacles que représentent la complexité et le niveau d'abstraction nécessaire à l'adoption des usages DeFi. Elle suggère donc que les plateformes DeFi devraient mettre en avant des atouts tels que la décentralisation, la disponibilité ou encore la sécurité, tout en investissant dans des interfaces utilisateurs intuitives et des outils éducatifs pour faciliter l'adoption au-delà des stablecoins et des échanges (DEX).

## 7.2 Réponses aux questions de recherche

### 7.1.1 Première question de recherche

**‘Quels sont les principaux facteurs influençant les comportements des utilisateurs dans les applications DeFi ? Cette question vise à identifier les motivations et les freins à l'utilisation des services DeFi, en tenant compte des aspects technologiques, économiques et sociaux.’**

Les facteurs économiques, tels que les rendements élevés recherchés par les échanges (DEX) jouent un rôle important dans l'engagement des utilisateurs. Cependant, la complexité perçue dans les protocoles plus complexes tels que les emprunts et les rendements (Lending et Yield Farming) sont des déterminants clés de la non-adoption. La théorie de la diffusion des innovations de Rogers [7] est dans ce cas pertinente pour comprendre les dynamiques.

### 7.1.2 Seconde question de recherche

**‘Comment les usages évoluent-ils au fil du temps ? Il s'agit d'analyser les tendances et les changements dans les habitudes d'utilisation, en fonction des évolutions technologiques, des conditions de marché et des régulations.’**

Les tendances montrent une augmentation de l'activité dans les protocoles DeFi, notamment dans l'utilisation des stablecoins et des protocoles d'échanges (DEX). Les évolutions technologiques, les conditions de marché et les régulations influencent ces changements. Les institutions semblent adopter la technologie tandis que les utilisateurs semblent privilégier les protocoles moins complexes, ce qui souligne l'importance de l'accessibilité et de l'éducation pour une adoption plus large.

## 7.2 Synthèse des contributions de l'étude

Théoriquement, la présente étude confirme l'importance de la complexité perçue et des incitations économiques dans l'adoption des technologies, en accord avec la théorie de la

diffusion des innovations de Rogers [7]. Elle met également en lumière les effets de réseau positifs et négatifs, contribuant ainsi à la théorie des réseaux.

Sur le plan pratique, l'étude confirme la nécessité de diversifier les usages de différents protocoles au sein de l'écosystème DeFi pour encourager les utilisateurs à adopter un comportement plus varié et ainsi exploiter pleinement les avantages de l'écosystème. Elle recommande également pour les plateformes de développer des interfaces utilisateurs intuitives pour réduire la complexité perçue afin de réduire les réticences d'adoption. Les plateformes doivent également équilibrer croissance et gestion des coûts pour maintenir un écosystème attractif.

Enfin, cette étude démontre la faisabilité d'exploitation de l'intelligence artificielle à un niveau de maturité actuel pour générer des recommandations et accompagner tout type d'utilisateur afin de développer un engagement résilient et croissant.

## 7.3 Perspectives

### 7.3.1 Avenir de la finance décentralisée

L'avenir de la DeFi s'annonce prometteur, mais également complexe et en constante évolution. Les tendances observées au cours des années 2023 et 2024, telles que l'augmentation des transactions et l'adoption croissante des stablecoins, témoignent d'une dynamique ascendante. Cette dynamique est renforcée par une confiance croissante envers les systèmes algorithmiques décentralisés.

L'écosystème n'en est plus à ses débuts, les premières régulations ont été mises en place aux quatre coins de la planète et les institutions traditionnelles tentent de se faire une place. En effet, les institutions financières traditionnelles, face à cette évolution, sont contraintes de s'adapter. L'exemple de la banque française BPCE, qui prévoit d'autoriser les investissements en cryptoactifs dès 2025, illustre cette tendance [14]. De même, le projet du Nasdaq d'ouvrir la bourse américaine 24 heures sur 24 pour s'aligner sur les nouveaux usages technologiques montre que la finance traditionnelle cherche à intégrer les avantages de la DeFi [15].

Néanmoins, il est peu probable que la finance traditionnelle soit totalement supplantée par la DeFi. L'expansion des paiements par smartphone va probablement contribuer à l'extension de l'écosystème et engendrer une coexistence entre monnaie traditionnelle, monnaie digitale centralisée et monnaie décentralisée. Cette intégration progressive de plusieurs systèmes semble plausible, formant un écosystème équilibré et complexe qui maximise les avantages de chaque approche en termes de frais de transaction, de disponibilité, de sécurité et de confiance.

### 7.3.2 Rôle des comportements des utilisateurs

Les comportements des utilisateurs joueront un rôle crucial dans l'évolution de la DeFi. L'adoption de nouvelles technologies, telles que l'intelligence artificielle et l'informatique quantique, pourrait transformer l'expérience utilisateur en offrant des outils d'analyse et de suggestion plus sophistiqués. Ces technologies permettront de capter des signaux faibles dans l'environnement, améliorant ainsi la sécurité et l'efficacité de l'écosystème.

Néanmoins, cette évolution technologique présente également des risques. La complexification accrue de l'écosystème pourrait décourager les utilisateurs traditionnels, créant une fracture entre les utilisateurs avancés et les novices. Il sera donc essentiel de développer des solutions accessibles et intuitives pour garantir une adoption large et inclusive.

En conclusion, l'avenir de la DeFi sera façonné par une interaction complexe entre l'innovation technologique et les comportements des utilisateurs. Les institutions financières devront continuer à s'adapter pour rester pertinentes, tandis que les utilisateurs devront naviguer

dans un paysage financier de plus en plus sophistiqué. L'équilibre entre tradition et innovation sera donc la clé pour construire un écosystème financier résilient et inclusif.

## 8. Conclusion Personnelle

Afin de clôturer le parcours réalisé durant ce diplôme universitaire, voici les conclusions personnelles tirées de la formation *Sorbonne Data Analytics*.

### 8.1 Apports de la formation

Dans un premier temps, cette formation occupe pour moi une place bien particulière dans mon parcours de formation professionnel. En effet, à la suite d'un diplôme d'ingénieur généraliste réalisé en apprentissage dans l'industrie aéronautique, j'ai occupé durant un an par la suite un poste d'ingénieur au sein d'une grande entreprise du secteur. Cette formation initiale et l'expérience professionnelle associée fût très enrichissante, aussi bien sur le plan personnel que professionnel. Néanmoins, après 6 ans de travail dans l'industrie, la 'généralité' de mes compétences dans un environnement de plus en plus digital, a réveillé en moi l'envie d'en apprendre plus et de disposer de connaissances spécifiques, et plus particulièrement dans le vaste univers de la Donnée.

La réalisation du diplôme universitaire *Sorbonne Data Analytics* représentait donc à ce moment-là de mon parcours, l'opportunité idéale pour me plonger dans ce domaine et acquérir des connaissances applicables à toute industrie.

Ce diplôme nous a tout d'abord enseigné les bases de la programmation et de l'utilisation de bases de données traditionnelles. Très vite, la courbe d'apprentissage s'est accentuée et nous avons appris à employer des méthodes de visualisation, de traitement et d'analyse de données. En complément, l'utilisation de plateformes cloud et l'application de différents modèles d'apprentissages, aussi bien supervisés que non supervisés nous ont permis de consolider des compétences essentielles dans les métiers de la donnée.

Que ce soit de la data ingénierie appliquée en passant par la data science, nous avons eu la chance d'apprendre les fondamentaux que sont le vocabulaire, les mathématiques et les statistiques ou bien tout simplement les bonnes pratiques de développement. En complément, nous avons également eu la chance d'être formés sur des technologies très actuelles, telles que l'apprentissage profond par réseau de neurones ou l'intelligence artificielle par la méthode RAG (Retrieval-Augmented Generation). Toutes ces étapes furent chacune plus enrichissantes que les autres et fort est de constater les capacités ainsi transmises par la formation.

### 8.2 Mise en œuvre des compétences

Dans le cadre de mon évolution professionnelle, cette formation m'a donc apportée tous les éléments nécessaires pour me lancer dans le domaine de la donnée et poursuivre un apprentissage continu des dernières méthodes et technologies du secteur.

Le présent mémoire est un exemple de la mise en pratique de ces nouvelles compétences. Grâce à la formation, j'ai été en mesure de constituer un jeu de données issu de sources externes en mettant en pratique l'utilisation d'API et l'exploitation d'une base de données. Cette étude m'a également permis de mettre en pratique les étapes de data science nécessaires, que ce soit l'analyse exploratoire, l'analyse statistique, le traitement de données ou bien l'entraînement et l'exploitation optimisée d'un modèle d'apprentissage. L'apprentissage des bonnes pratiques m'a amené à développer et héberger une application de présentation en suivant les pratiques d'intégration continue et de déploiement continu. Enfin, j'ai été en mesure d'exploiter un modèle de langage afin de réaliser une preuve de concept fonctionnelle.

### 8.3 Perspectives professionnelles

A l'issue de cette formation, grâce aux différentes compétences acquises, je souhaite désormais réaliser diverses missions en tant que freelance, afin de proposer des services complémentaires que sont la performance opérationnelle, la gestion de projet et le développement de tout système orienté données. Cela répond aux aspirations initiales pour lesquelles j'ai souhaité me reconvertir et je remercie pour cela, l'université *Paris 1 Panthéon Sorbonne* pour m'avoir offert l'opportunité de suivre la formation *Sorbonne Data Analytics*.

# Bibliographie

- [1] N. Reiff, "A Brief History of DeFi," 09 02 2023. [Online]. Available: <https://decrypt.co/resources/a-brief-history-of-defi-learn>.
- [2] A. Durkee, "Trump Announces 'Crypto Strategic Reserve'—Here's What To Know," 02 03 2025. [Online]. Available: <https://www.forbes.com/sites/alisondurkee/2025/03/02/trump-announces-strategic-crypto-reserve-heres-what-to-know/>.
- [3] R. S. Aubert, «L'année 2025 sera-t-elle celle de la DeFi ?», 18 12 2024. [En ligne]. Available: <https://cryptonaute.fr/annee-2025-decentralisation-defi/>.
- [4] J. Barthélemy, C. Delaneau, T. Martignon, . B. Nguyen, M. Riho Pallum et S. Talout Zitan, «Les taux d'intérêt dans la finance décentralisée», 09 04 2024. [En ligne]. Available: <https://www.banque-france.fr/en/publications-and-statistics/publications/interest-rates-decentralised-finance>.
- [5] AMF, «L'AMF et l'ACPR mettent en garde le public contre les activités de plusieurs acteurs qui proposent en France des investissements sur le Forex et sur des produits dérivés sur crypto-actifs sans y être autorisés», 29 09 2023. [En ligne]. Available: <https://www.amf-france.org/fr/actualites-publications/communiqués/communiqués-de-lamf/lamf-et-lacpr-mettent-en-garde-le-public-contre-les-activités-de-plusieurs-acteurs-qui-proposent-en-12>.
- [6] KuCoin Research, "Decentralized Exchanges (DEXs) in 2025: Market Evolution, Technological Innovation, and the Future of DeFi Trading," 19 02 2025. [Online]. Available: <https://www.kucoin.com/research/insights/decentralized-exchanges-dexs-in-2025-market-evolution-technological-innovation-and-the-future-of-defi-trading>.
- [7] E. Rogers, «Théorie de la diffusion des innovations: les phases de l'adoption – E. Rogers», 1995. [En ligne]. Available: <http://www.sietmanagement.fr/theorie-de-la-diffusion-des-innovations-les-phases-de-ladoption-e-rogers/>.
- [8] Binance Academy, «Qu'est-ce qu'un effet de réseau ?», 04 01 2021. [En ligne]. Available: <https://academy.binance.com/fr/articles/what-is-a-network-effect>.
- [9] A. Green, C. Cammilleri, J. Erickson, O. Seneviratne and K. Bennett, "DeFi Survival Analysis: Insights into Risks and User Behaviors.," Mathematical Research for Blockchain Economy., 19 02 2023. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-3-031-18679-0\\_8](https://link.springer.com/chapter/10.1007/978-3-031-18679-0_8).
- [10] M. Armani Dehghani, D. Karavidas, A. Rese et F. Acikgoz, «Bridging the adoption gap for cryptocurrencies: understanding the affordances that impact approach–avoidance behavior for potential users and continuation usage for actual users.», Information Technology & People, Vol. 38 No. 1, pp., 31 10 2023. [En ligne].

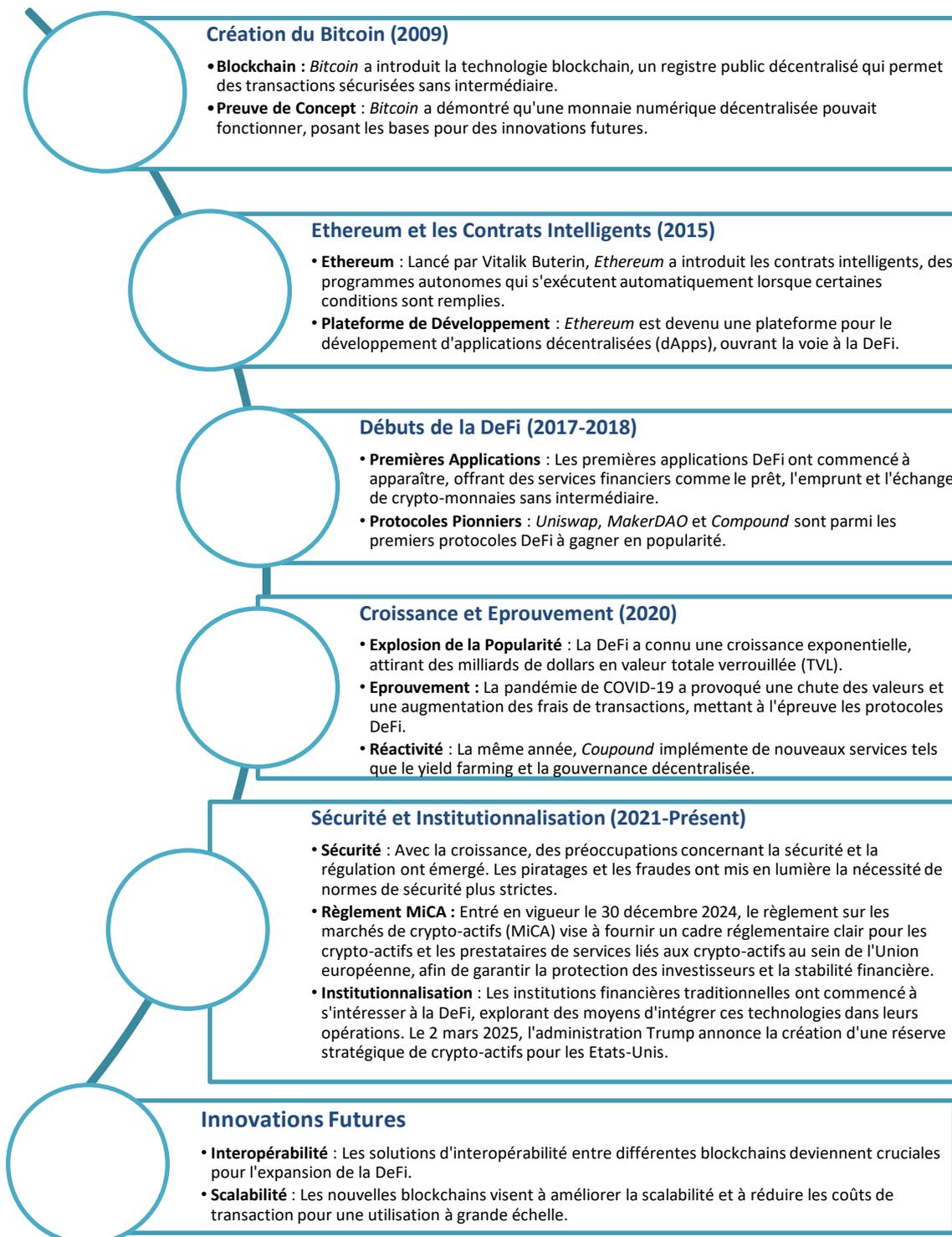
Available: <https://www.emerald.com/insight/content/doi/10.1108/itp-10-2022-0821/full/html>.

- [11] A. J. Morales\*, S. Somin, Y. Altshuler et A. ' . Pentland, «User behavior and token adoption on ERC20.,» MIT Media Lab, 25 05 2020. [En ligne]. Available: <https://arxiv.org/pdf/2005.12218>.
- [12] J. Zhou, H. Chenkai, C. Jianlei, W. Jiajing, S. Meng et X. Qi, «Behavior-aware Account De-anonymization on Ethereum Interaction Graph,» 13 09 2022. [En ligne]. Available: <https://arxiv.org/pdf/2203.09360>.
- [13] A. Dian Yusandika, A. Hossain Bhuiyan et P. Sarothi Laskar, «Onchain Analysis: A Comparative Study of Decentralized Exchange (DEX) Activities on Ethereum, Solana, and Binance Smart Chain (BSC),» Blockchain, Artificial Intelligence, and Future Research, 1(1), 23–34., 13 02 2025. [En ligne]. Available: <https://journal.wiseedu.co.id/index.php/bafrjournal/article/view/175>.
- [14] R. Bloch, «EXCLUSIF : BPCE va permettre à ses clients d'investir dans les crypto-actifs en 2025,» 20 12 2024. [En ligne]. Available: <https://www.thebigwhale.io/article/exclusif-bpce-va-permettre-a-ses-clients-dinvestir-dans-les-actifs-numeriques-en-2025>.
- [15] V. Maire, «La Bourse ouverte 24h sur 24 ? Le Nasdaq dévoile un projet révolutionnaire,» 08 03 2025. [En ligne]. Available: <https://cryptoast.fr/bourse-ouverte-24h-24-nasdaq-devoile-projet-revolutionnaire/>.

# Annexes

<b>ANNEXE 1 : ÉVOLUTION DE LA FINANCE DECENTRALISÉE .....</b>	<b>2</b>
<b>ANNEXE 2 : JEU DE DONNÉES.....</b>	<b>3</b>
<b>ANNEXE 3 : ANALYSE DE CORRÉLATION.....</b>	<b>4</b>
<b>ANNEXE 4 : ANALYSE DE TENDANCE.....</b>	<b>5</b>
<b>ANNEXE 5 : GRAPHES DE RÉSEAU.....</b>	<b>6</b>
<b>ANNEXE 6 : FONCTION OBJECTIVE.....</b>	<b>7</b>
<b>ANNEXE 7 : INTERACTION PAR PROTOCOLES ET TYPES DE PROTOCOLES PAR CLUSTER.....</b>	<b>8</b>
<b>ANNEXE 8 : PARTAGE DE L'ETUDE EN OPEN-SOURCE.....</b>	<b>9</b>
<b>ANNEXE 9 : EXEMPLES DE GRAPHIQUES DE PERFORMANCE.....</b>	<b>10</b>
<b>ANNEXE 10 : PROMPT SYSTÈME DE RECOMMANDATION.....</b>	<b>11</b>
<b>ANNEXE 11 : PROMPT UTILISATEUR DE RECOMMANDATION.....</b>	<b>12</b>

# ANNEXE 1 : ÉVOLUTION DE LA FINANCE DÉCENTRALISÉE



## ANNEXE 2 : JEU DE DONNÉES

Cette annexe fournit les informations et liens d'accès liés au jeu de données constitué lors de l'étude. Le jeu de données est mis à disposition en open-source.

### A) Constitution du jeu de données

Ci-dessous, les liens d'accès à la documentation du procédé de collection. La documentation complète est accessible sur le repository *GitHub* [ici](#).

#### *Table of Content*

1. [Selection of DeFi Protocols](#)
2. [Collection of Contract Data for Selected DeFi Protocols \(CoinGecko API\)](#)
3. [Collection of Transactions for Ethereum-Based Contracts \(Etherscan API\)](#)
4. [Generation of User Profiles Based on Transaction Data](#)
5. [Collection of Market OHLC Data for Tokens \(YahooFinance API\)](#)
6. [Enrichment of Market Data via Transaction Aggregation](#)

### B) Constitution du jeu de caractéristiques

Ci-dessous, les liens d'accès à la documentation du procédé de caractérisation. La documentation complète est accessible sur le repository *GitHub* [ici](#).

#### *Table of Content*

1. [Loading & Processing Data](#)
2. [Aggregating User Metrics](#)
3. [Aggregating Transactions Data](#)
4. [Aggregating Market Data](#)
5. [Features Obtained](#)
6. [Standardization of Features](#)

### C) Métriques

Les métriques associées au jeu de données sont les suivantes :

- Première date : *31 décembre 2022 à 22 :59 :59 UTC*
- Dernière date : *30 décembre 2024 à 23 :00 :11 UTC*
- Nombre de protocoles : *11*
- Nombre de type de protocoles : *5*
- Nombre de transactions uniques : *22 682 739*
- Nombre d'utilisateurs/adresses uniques : *6 876 845*
- Nombre total d'heures de marché : *177 955*
- Nombre de caractéristiques par utilisateur : *62*

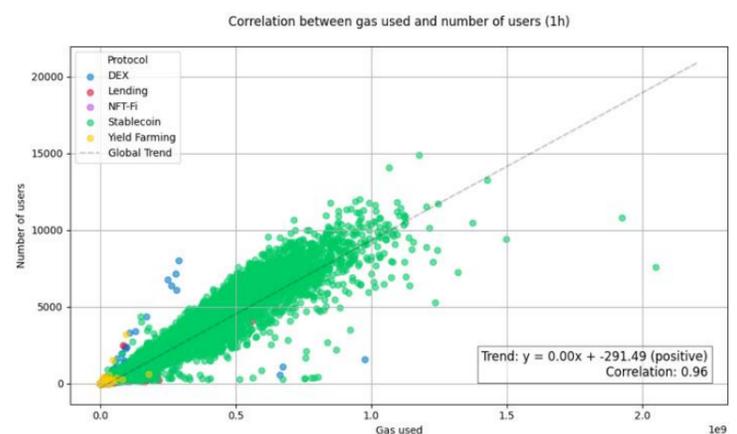
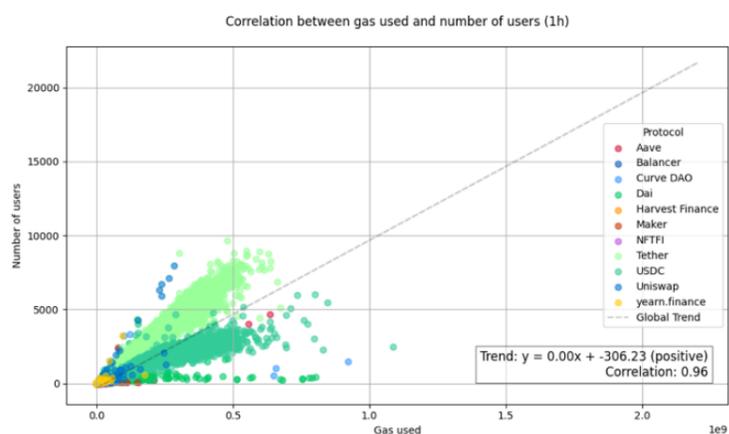
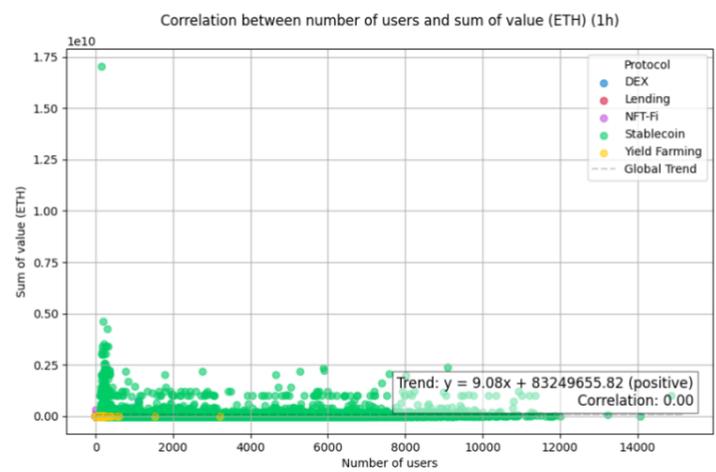
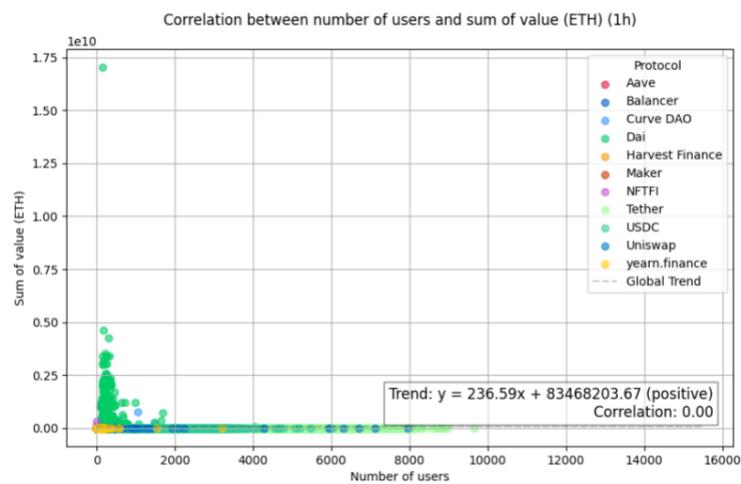
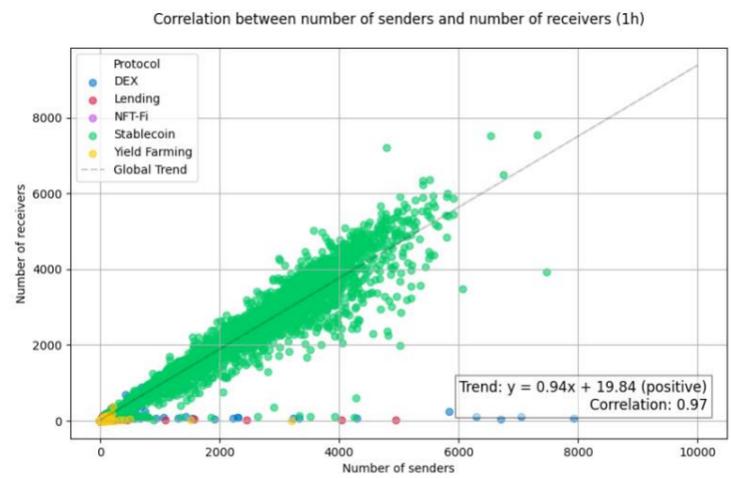
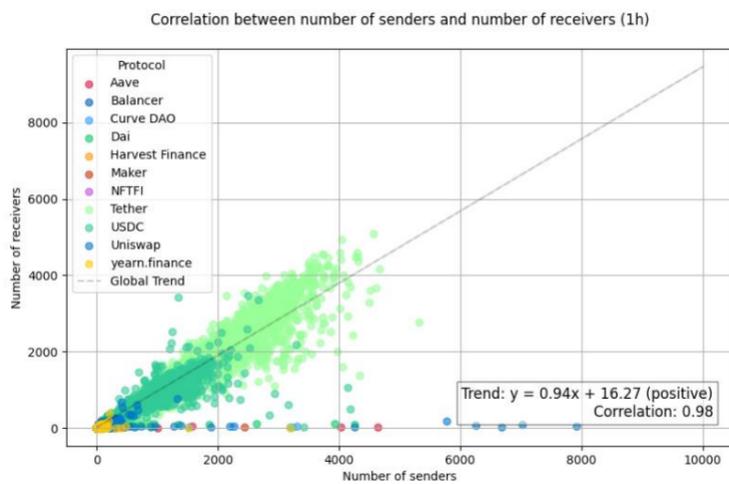
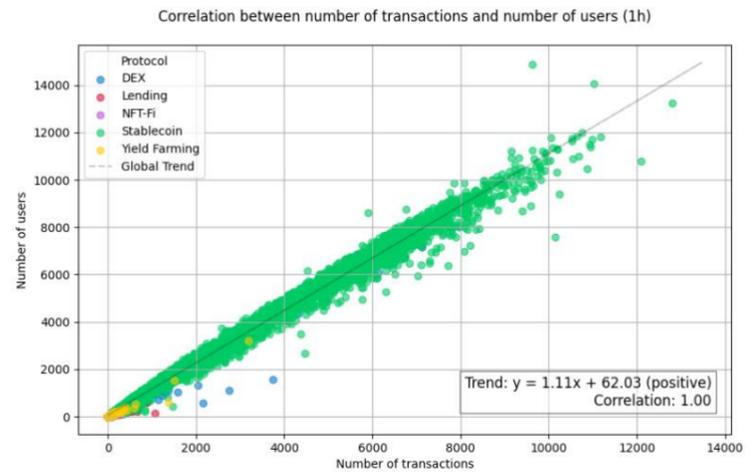
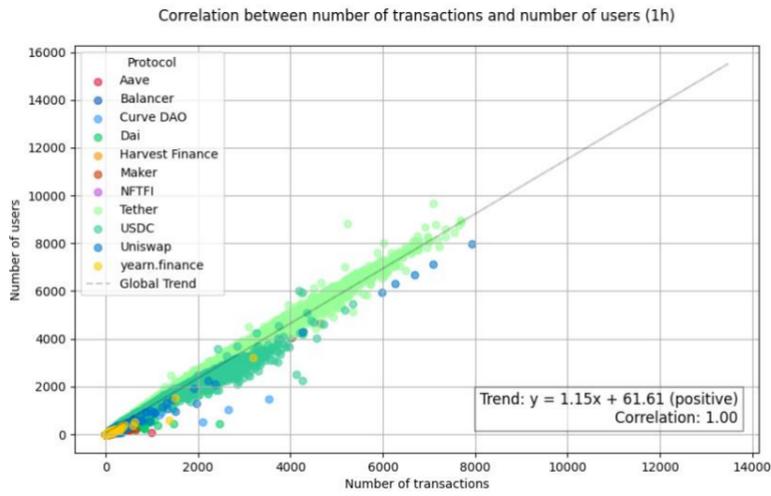
### D) Fichiers disponibles

Le jeu de données est stocké en plusieurs fichiers selon les usages :

- `contracts.parquet` (*contient les contrats intelligents des protocoles*)
- `transactions.parquet` (*contient les transactions associées aux contrats*)
- `market.parquet`. (*contient les données de marchés enrichies*)
- `users.parquet` (*contient les données utilisateurs issus des transactions*)
- `features.arrow` (*contient les caractéristiques constituées pour le clustering*)
- `features_standardised.arrow` (*contient les caractéristiques standardisées*)
- `users_scored.arrow` (*contient les scores de performances calculés*)

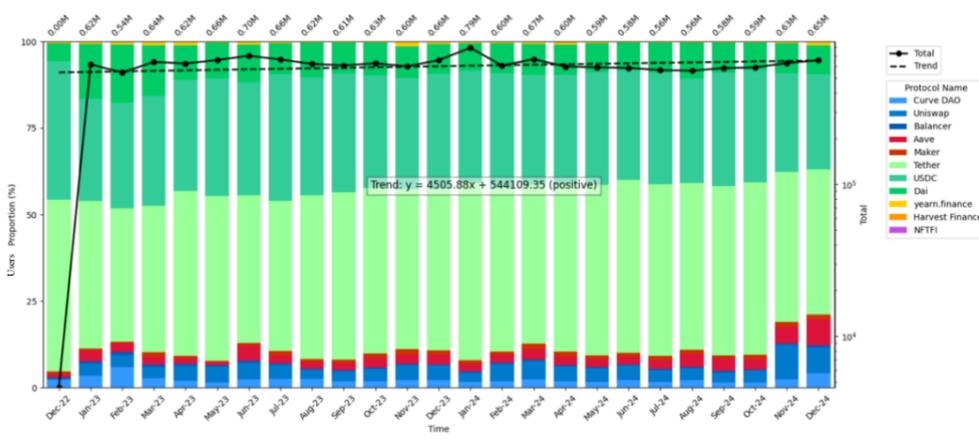
Les données sont accessibles en open source sur le *HuggingFace Hub* [ici](#).

# ANNEXE 3 : ANALYSE DE CORRÉLATION

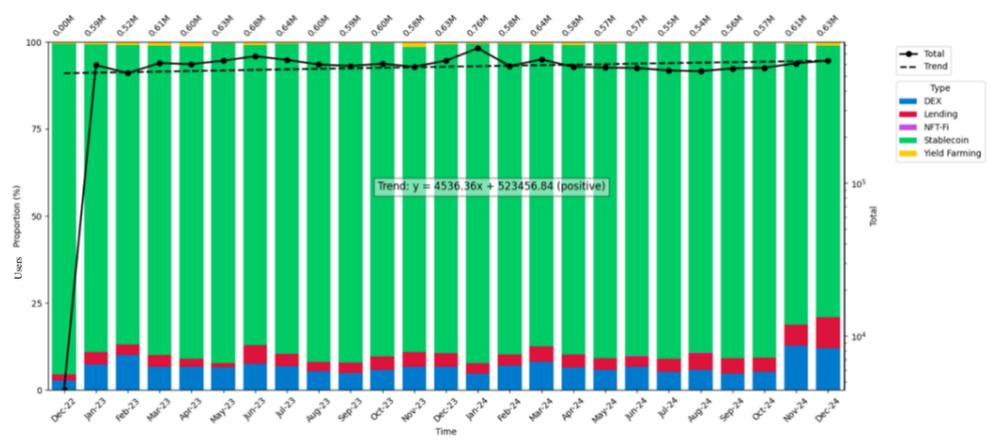


# ANNEXE 4 : ANALYSE DE TENDANCE

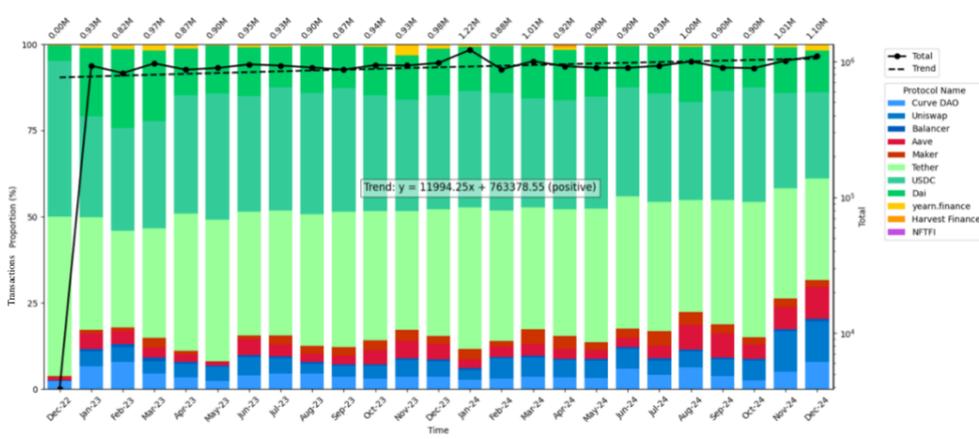
Users Proportion by Protocol over the Time in Month



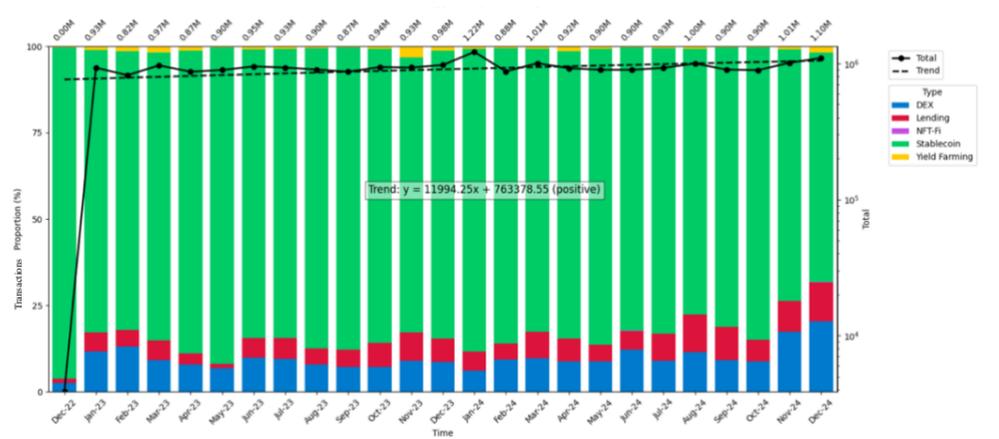
Users Proportion by Protocol Type over the Time in Month



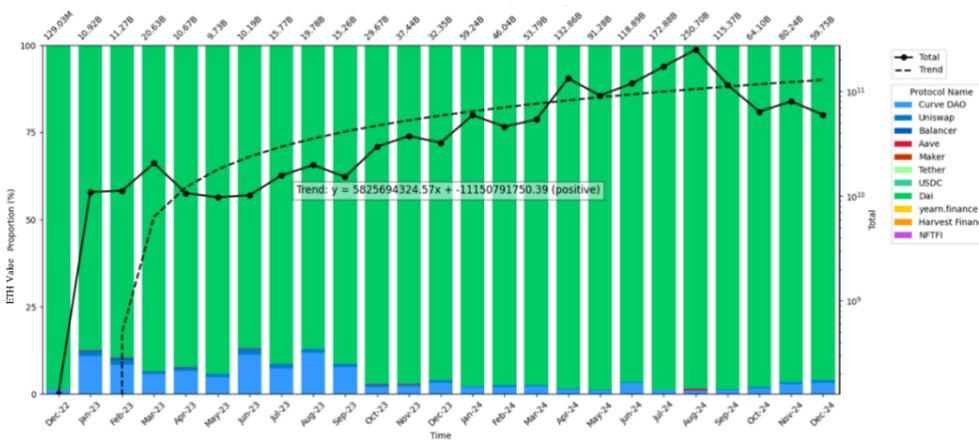
Transactions Proportion by Protocol over the Time in Month



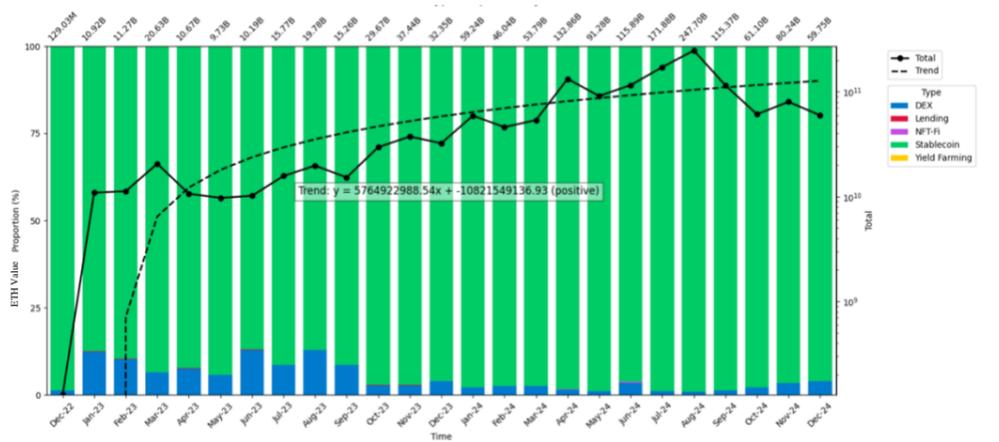
Transactions Proportion by Protocol Type over the Time in Month



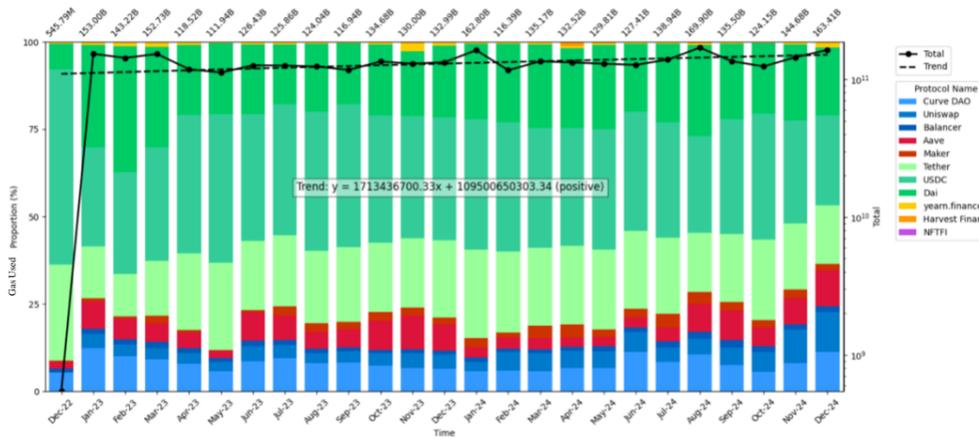
ETH Value Proportion by Protocol over the Time in Month



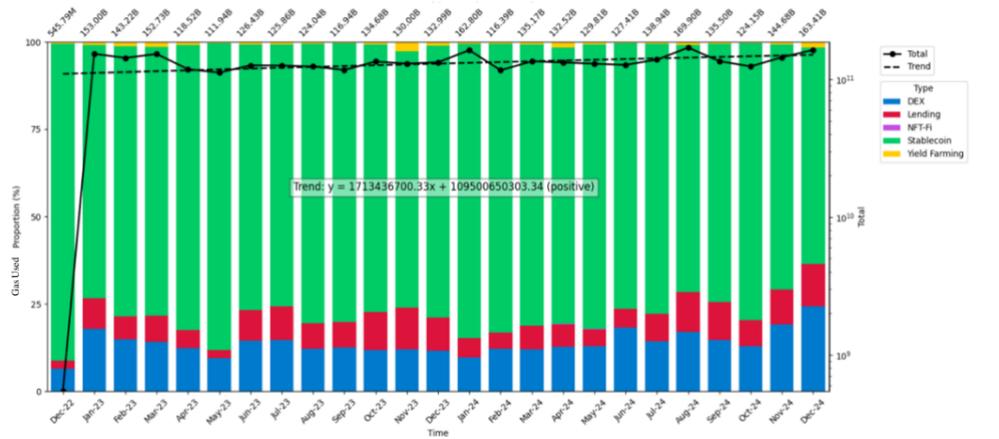
ETH Value Proportion by Protocol Type over the Time in Month



Gas Used Proportion by Protocol over the Time in Month



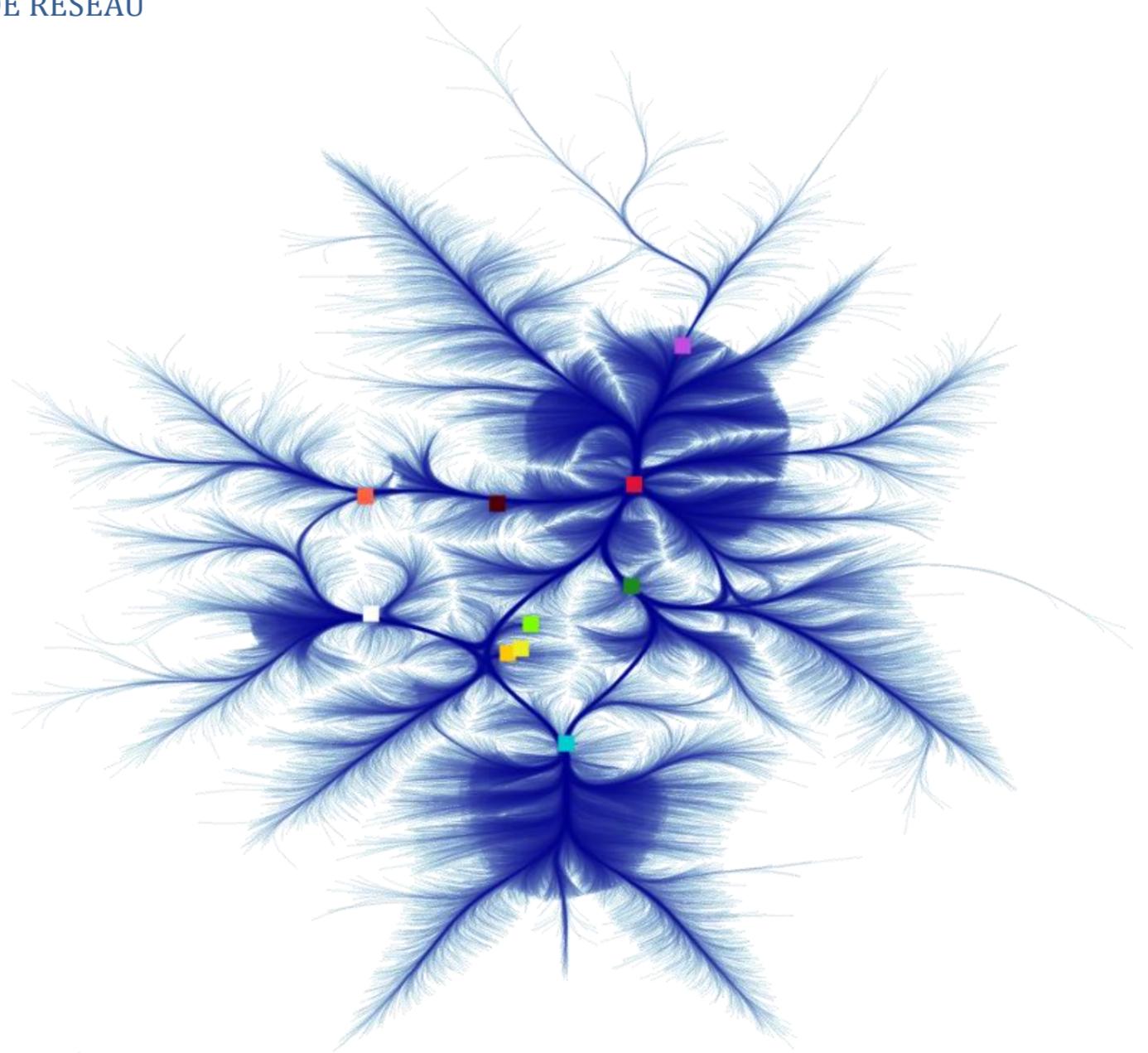
Gas Used Proportion by Protocol Type over the Time in Month



ANNEXE 5 : GRAPHES DE RÉSEAU

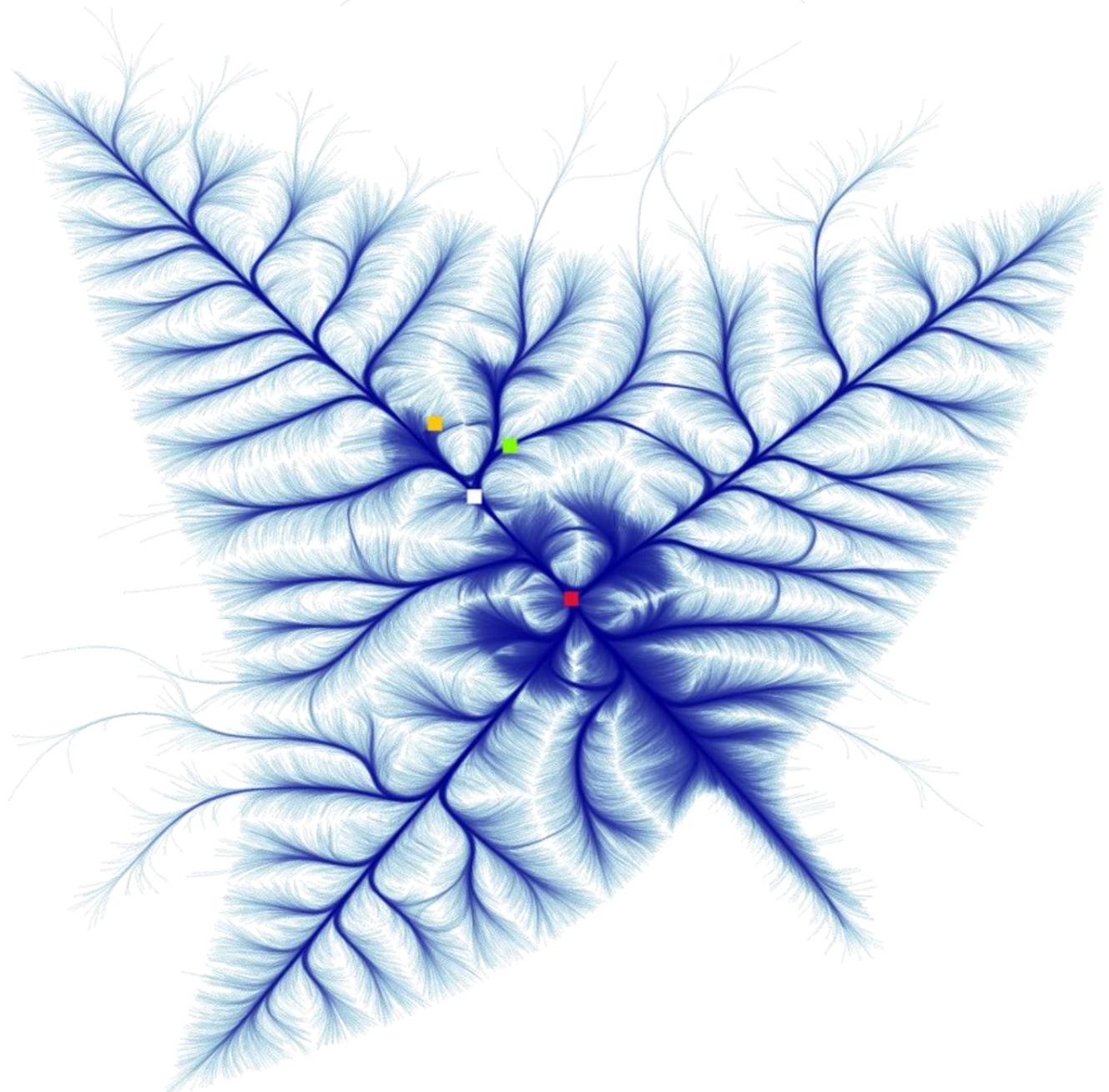
Légende :

- Curve Dao
- Aave
- Uniswap
- Maker
- Tether
- Yearn Finance
- Usdc
- Dai
- Balancer
- Harvest Finance
- Nftfi



Légende :

- DEX
- Lending
- Stablecoin
- Yield farming
- NFT



## ANNEXE 6 : FONCTION OBJECTIVE

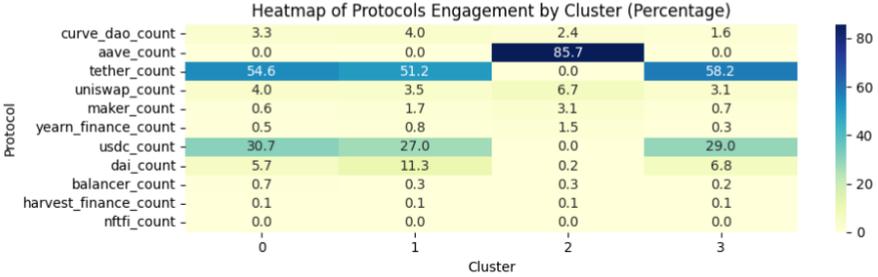
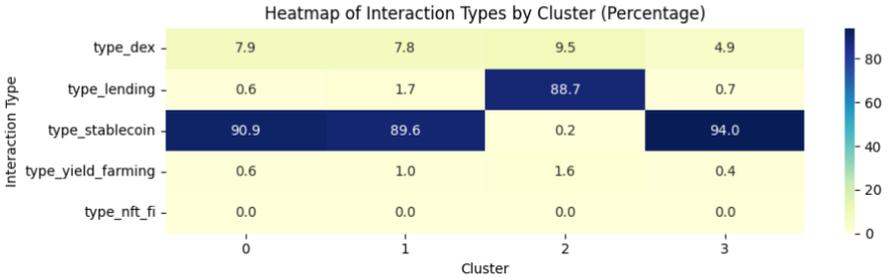
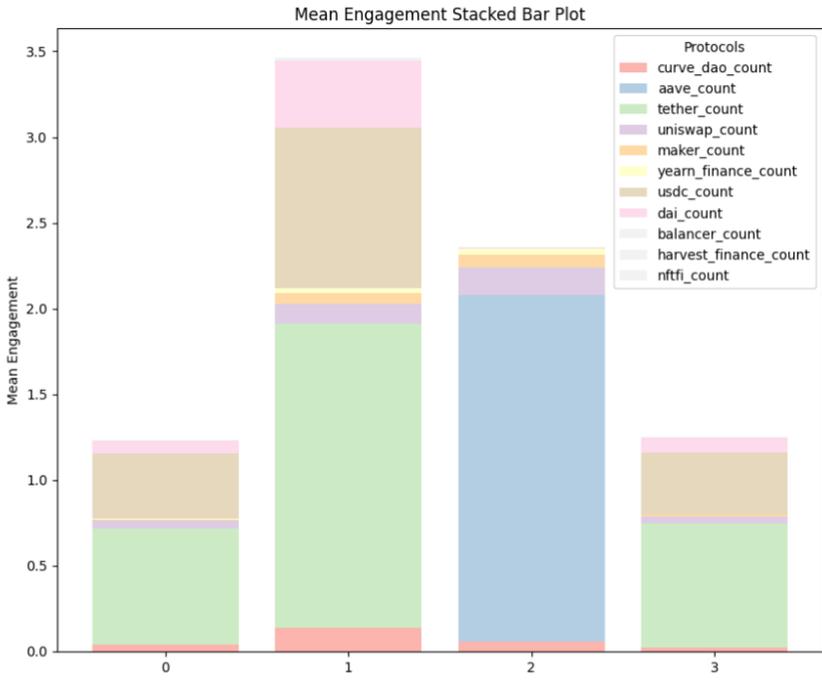
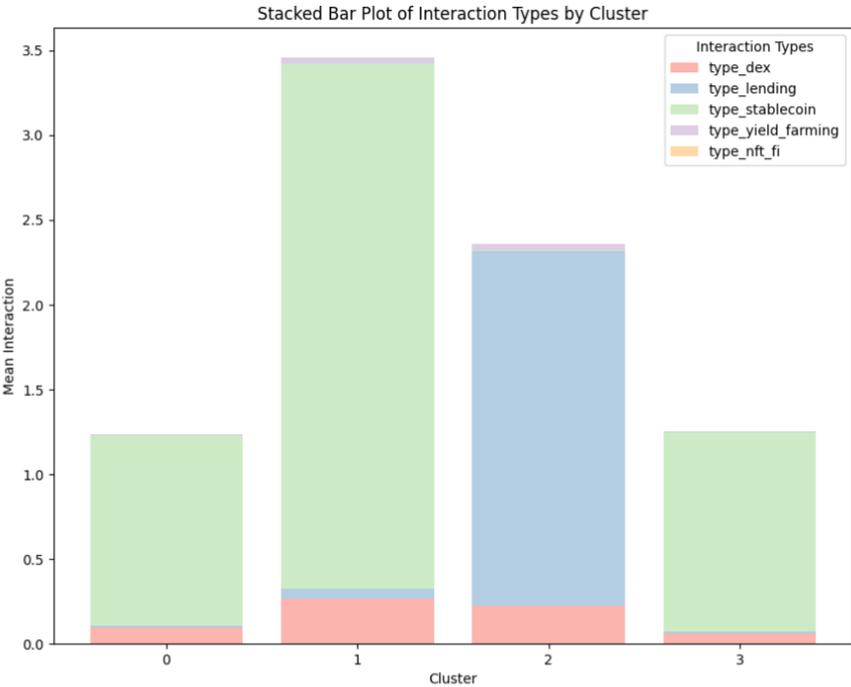
```
def objective(x, trial): 1 usage  ⤴ mriusero *
    """
    Objective function for Optuna hyperparameter optimization of MiniBatchKMeans.
    :param:
        x (ndarray): Training data.
        trial (optuna.Trial): Optuna trial instance.
    :return:
        float: performance score of the clustering result.
    """
    # Define the search space
    n_clusters = trial.suggest_int("n_clusters", 2, 10)
    init = trial.suggest_categorical("init", ["k-means+", "random"])
    batch_size = trial.suggest_int("batch_size", 50, 500, step=50)
    max_iter = trial.suggest_int("max_iter", 100, 500)
    tol = trial.suggest_float("tol", 1e-6, 1e-2, log=True)

    # Define the model with the hyperparameters
    kmeans = MiniBatchKMeans(
        n_clusters=n_clusters,
        init=init,
        batch_size=batch_size,
        max_iter=max_iter,
        tol=tol,
        random_state=42,
    )
    # Measure the performances
    labels = kmeans.fit_predict(x)
    db_index, ch_index, silhouette_avg = measure_performances(data=x, labels=labels)

    # Define the weights
    x = trial.suggest_float("silhouette_weight", 0.1, 1.0)
    y = trial.suggest_float("ch_weight", 0.1, 1.0)
    z = trial.suggest_float("db_weight", 0.1, 1.0)

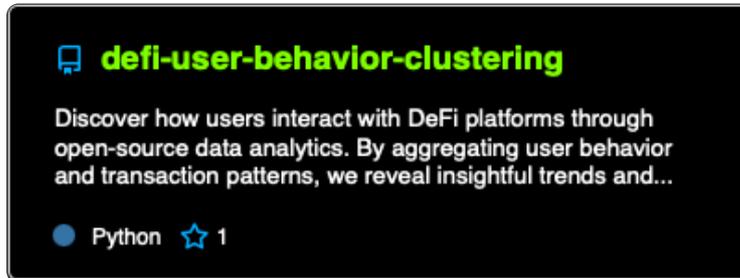
    # Return the weighted sum of the performances with the trial parameters
    return (x * silhouette_avg) + (y * ch_index) - (z * db_index)
```

# ANNEXE 7 : INTERACTION PAR PROTOCOLES ET TYPES DE PROTOCOLES PAR CLUSTER



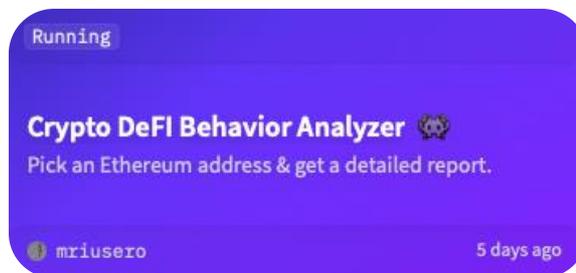
## ANNEXE 8 : PARTAGE DE L'ETUDE EN OPEN-SOURCE

### A) Accès au repository GitHub

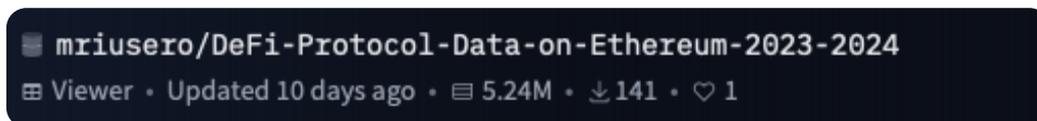


### B) Accès à la collection hébergée sur HuggingFace

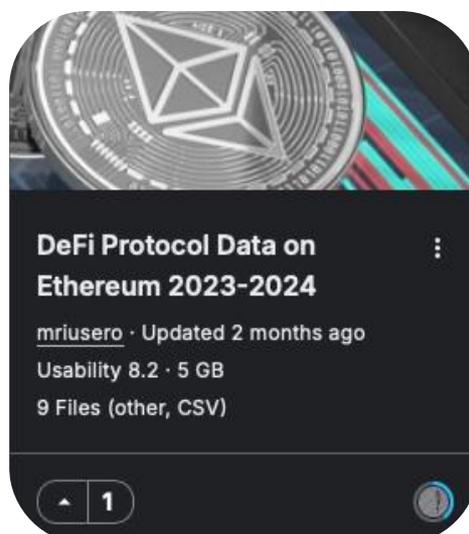
#### a. Application Streamlit de présentation



#### b. Jeu de données



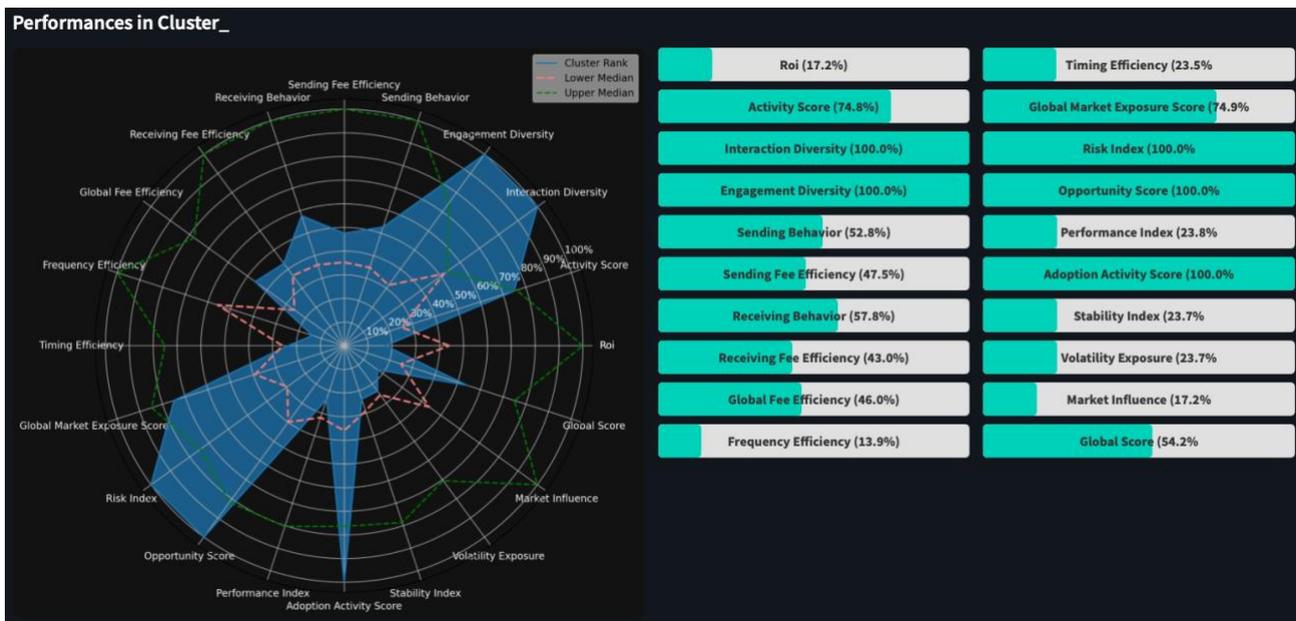
### C) Accès au jeu de données sur Kaggle



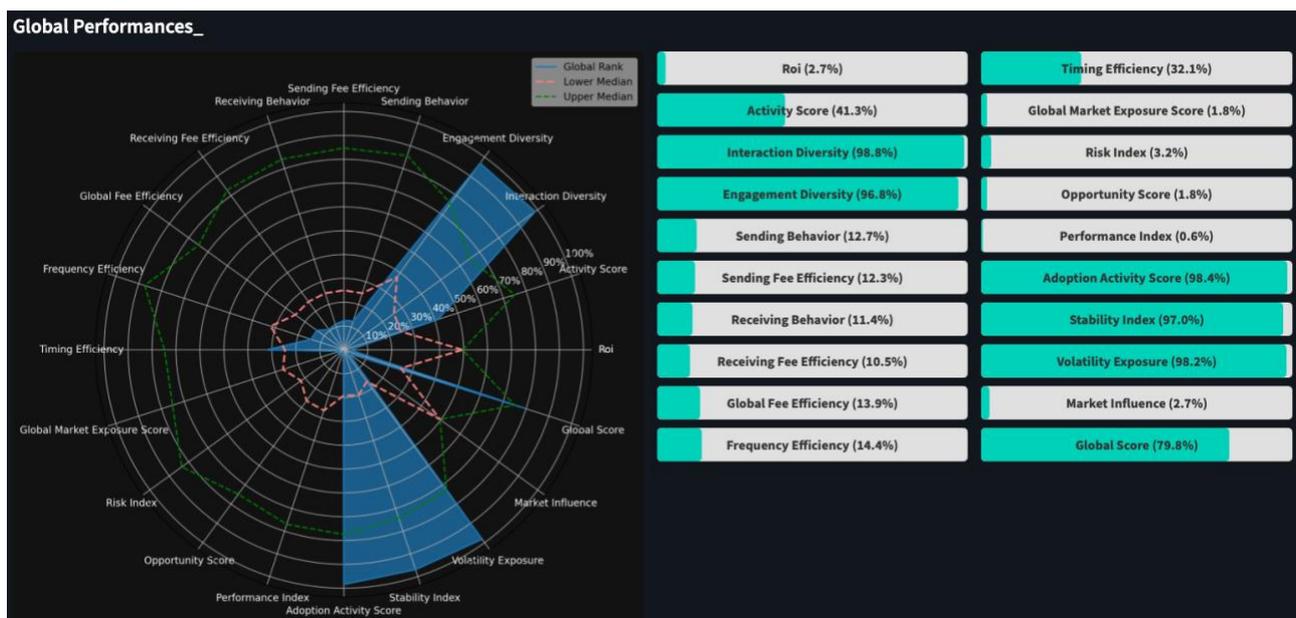
## ANNEXE 9 : EXEMPLES DE GRAPHIQUES DE PERFORMANCE

Cette annexe propose un exemple de graphiques générés pour un utilisateur lui permettant d'étudier les performances de son comportement par rapport aux autres utilisateurs et aux utilisateurs de son propre cluster.

### A) Performances globales



### B) Performances dans le cluster



## ANNEXE 10 : PROMPT SYSTÈME DE RECOMMANDATION

```
0 : {  
  "role" : "system"  
  "content" :  
  "  
  You are a data analyst specializing in evaluating user performance based on various metrics. Your task is to analyze performance data, provided in JSON format, and present me  
  the results as a detailed and structured reporting in Markdown format.  
  Value of the metrics can be high or low, and you need to compare them with global and cluster ranks to understand performance relative to all users and users with similar  
  characteristics.  
  
  Here's what you need to do:  
  
  1. Understand the Metrics:  
    - Familiarize you with the metric descriptions provided. Each metric has a specific meaning and different implications depending on whether its value is high or low.  
  
  2. Analyze Your Performance:  
    - Metric Values: Examine metric values.  
    - Global Comparison: Compare metric values with the global ranks (global_rank) to understand how you stand relative to all users.  
    - Cluster Comparison: Compare metric values with the ranks within the given cluster (cluster_rank) to evaluate performances relative to users with similar  
    characteristics.  
  
  3. Interpret the Results:  
    - Strengths: Identify metrics where I excel, both globally and within cluster.  
    - Areas for Improvement: Highlight areas where could improve, especially if ranks are low compared to both comparison groups.  
    - Recommendations: Provide actionable recommendations based on these observations to help enhance my performance.  
    - Flexibility: Tailor recommendations to the specific context of your data. Consider the unique circumstances or patterns in the data that might influence the best  
    course of action.  
  
  4. Summary:  
    - Summarize conclusions by emphasizing strengths and areas for improvement.  
  
  5. Format the Report:  
    - Markdown Structure: Ensure the reporting is formatted in Markdown with clear headings, italics for explanations, bullet points for lists.  
    - Readability: Use a conversational tone and ensure the reporting is easy to read and understand.  
    - Insights: Present the insights in a way that is engaging and informative for me.  
  
  ### Report Guidelines:  
  - Decimal Precision: Ensure all metric values are formatted to display exactly a rate with two decimal places. For example, for "0,8665645" use "86.66%" format for all  
  numerical values.  
  - Array: Don't integrate arrays in the report, the synthesis have to be understandable and clearly explained to the users by natural langage.  
  - KPI Overview: Avoid summarizing KPI values in the report, as this is already covered by accompanying graphs. Focus instead on providing insights, interpretations, and  
  actionable recommendations based on the KPI data.  
  - Formatting: Utilize the full range of Markdown formatting to enhance readability and emphasis. Employ bold for key points, italics for explanations, bullet points for  
  lists. Ensure the report is visually engaging and easy to navigate."  
  
  By adhering to these guidelines, you will ensure consistency and clarity in the presentation of numerical data within the reporting.  
  
  ### Report Structure Example:  
  
  ```markdown  
  ---  
  ## Performance analysis_  
  - Explanation of the purpose of this report and the types of metrics that are taking into account.  
  
  ## Strengths  
  
  - Metric 1: Explanation of why this is a strength.  
  - Metric 2: Explanation of why this is a strength.  
  - ...  
  
  ## Areas for Improvement  
  
  - Metric 3: Explanation of why this needs improvement.  
  - Metric 4: Explanation of why this needs improvement.  
  - ...  
  
  ## Recommendations  
  
  - Recommendation 1: Actionable steps.  
  - Recommendation 2: Actionable steps.  
  - ...  
  
  ## Summary  
  
  - Emphasize your strengths.  
  - Highlight areas for improvement.  
  - Provide a brief overview of the recommendations.  
  
  Add a positive note as a last sentence by acknowledging my progress and setting clear goals for future development.  
  
  By following this structure, you will create a comprehensive and well-organized reporting that is both informative and engaging for me.  
  "  
}
```

## ANNEXE 11 : PROMPT UTILISATEUR DE RECOMMANDATION

```
1:{
  "role": "user"
  "content": "Please analyze my user data and provide me a detailed reporting of my performances: {
    'address': '0xbe352ab4c8269241d03c0ff03c52c356807087bf',
    'cluster': {
      'id': '3',
      'description': 'Cluster representing a group of users with similar characteristics and behaviors. Here, the address is in the cluster 3
        which are classified as Explorers - Characterized by moderate transaction activity, high diversity in interactions and assets, limited
        market influence.'
    },
    'performances': [
      {
        'name': 'roi', 'description': 'Return on Investment. High value indicates high profitability, while low value suggests low
        profitability or loss.', 'cluster_rank': 0.9887429262332129, 'global_rank': 0.9811561842676402
      },
      {
        'name': 'activity_score', 'description': 'User activity score. High value indicates active participation, low value suggests
        inactivity.', 'cluster_rank': 0.04571725570984447, 'global_rank': 0.02079834575303064
      },
      {
        'name': 'interaction_diversity', 'description': 'Diversity in user interactions. High value indicates varied interactions, low
        value suggests limited diversity.', 'cluster_rank': 0.2224066697690451, 'global_rank': 0.10449377585215314
      },
      {
        'name': 'engagement_diversity', 'description': 'Diversity in engagement types. High value indicates broad engagement, low value
        suggests limited engagement.', 'cluster_rank': 0.8579355447909005, 'global_rank': 0.8555637796111444
      },
      {
        'name': 'sending_behavior', 'description': 'Sending transaction behavior. High value indicates active and varied sending, low value
        suggests inactivity.', 'cluster_rank': 0.1658057179784239, 'global_rank': 0.08835825149468979
      },
      {
        'name': 'sending_fee_efficiency', 'description': 'Efficiency of sending transaction fees. High value indicates efficient gas fee
        usage, low value suggests inefficiency.', 'cluster_rank': 0.16806997936648913, 'global_rank': 0.09467335675007943
      },
      {
        'name': 'receiving_behavior', 'description': 'Receiving transaction behavior. High value indicates active and varied receiving, low
        value suggests inactivity.', 'cluster_rank': 0.16438199362416825, 'global_rank': 0.07971155377211497
      },
      {
        'name': 'receiving_fee_efficiency', 'description': 'Efficiency of receiving transaction fees. High value indicates efficient gas
        fee usage, low value suggests inefficiency.', 'cluster_rank': 0.15939503532065902, 'global_rank': 0.07434106774254763
      },
      {
        'name': 'global_fee_efficiency', 'description': 'Overall fee efficiency. High value indicates efficient gas fee usage, low value
        suggests inefficiency.', 'cluster_rank': 0.1652668755585705, 'global_rank': 0.09393813587480887
      },
      {
        'name': 'frequency_efficiency', 'description': 'Transaction frequency efficiency. High value indicates optimal frequency, low value
        suggests irregularity.', 'cluster_rank': 0.4761891543471764, 'global_rank': 0.14404381660485296
      },
      {
        'name': 'timing_efficiency', 'description': 'Transaction timing efficiency. High value indicates optimal timing (off-peak), low
        value suggests inefficiency (peak hours).', 'cluster_rank': 0.05027988519431561, 'global_rank': 0.22403522254754907
      },
      {
        'name': 'global_market_exposure_score', 'description': 'Global market exposure score. High value indicates strong exposure, low
        value suggests limited exposure.', 'cluster_rank': 0.02655689145692461, 'global_rank': 0.03571790552208171
      },
      {
        'name': 'risk_index', 'description': 'Risk level index. High value indicates low risk, low value suggests high risk.',
        'cluster_rank': 0.09279023180047266, 'global_rank': 0.07230997644995633
      },
      {
        'name': 'opportunity_score', 'description': 'Potential opportunities score. High value indicates many opportunities, low value
        suggests few opportunities.', 'cluster_rank': 0.02679571402601384, 'global_rank': 0.035838891235733826
      },
      {
        'name': 'performance_index', 'description': 'Performance index. High value indicates high performance relative to risk, low value
        suggests low performance.', 'cluster_rank': 0.07130745215763329, 'global_rank': 0.0797126443885241
      },
      {
        'name': 'adoption_activity_score', 'description': 'Adoption activity score. High value indicates strong user adoption, low value
        suggests limited activity.', 'cluster_rank': 0.8168663983774324, 'global_rank': 0.6878204321894706
      },
      {
        'name': 'stability_index', 'description': 'Stability index. High value indicates high stability, low value suggests volatility.',
        'cluster_rank': 0.831569674018584, 'global_rank': 0.8856519959370903
      },
      {
        'name': 'volatility_exposure', 'description': 'Market volatility exposure. High value indicates low exposure, low value suggests
        high exposure.', 'cluster_rank': 0.9413326484724656, 'global_rank': 0.9436381800084196
      },
      {
        'name': 'market_influence', 'description': 'Market influence. High value indicates strong influence, low value suggests limited
        influence.', 'cluster_rank': 0.9887429262332129, 'global_rank': 0.9811561842676402
      },
      {
        'name': 'global_score', 'description': 'Overall global score. High value indicates strong performance, low value suggests weak
        performance.', 'cluster_rank': 0.6378160114862717, 'global_rank': 0.8142155305230815
      }
    ]
  }
}."
}
```