Individual Coursework Submission Form

Specialist Masters Programme

<table>
<tr><td>Please list all names of group members:
1. Bhasin, Ruhanijot Kaur
2. Raghuvanshi, Pawni
3. Rawal, Nikhil
4. Verma, Shubham</td><td>GROUP NUMBER: 12</td></tr>
<tr><td colspan="2">MSc in: Actuarial Science and Business Analytics</td></tr>
<tr><td colspan="2">Module Code: SMM695</td></tr>
<tr><td colspan="2">Module Title: Data Management Systems</td></tr>
<tr><td>Lecturer: Matteo Devigili</td><td>Submission Date: 19 July, 2024</td></tr>
<tr><td colspan="2">Declaration:
By submitting this work, we declare that this work is entirely our own except those parts duly identified and referenced in my submission. It complies with any specified word limits and the requirements and regulations detailed in the coursework instructions and any other relevant programme and module documentation. In submitting this work we acknowledge that we have read and understood the regulations and code regarding academic misconduct, including that relating to plagiarism, as specified in the Programme Handbook. We also acknowledge that this work will be subject to a variety of checks for academic misconduct.

We acknowledge that work submitted late without a granted extension will be subject to penalties, as outlined in the Programme Handbook. Penalties will be applied for a maximum of five days lateness, after which a mark of zero will be awarded.</td></tr>
<tr><td colspan="2">Marker’s Comments (if not being marked on-line):</td></tr>
</table>

Deduction for Late Submission:

Final Mark: %

"Analysing Governance and Operational Efficiency in Decentralized Autonomous Organizations (DAOs)"

1. Introduction

Decentralized Autonomous Organizations (DAOs) revolutionize organizational governance by enabling distributed, transparent decision-making without a central authority. Using blockchain technology and smart contracts, DAOs rely on members holding native tokens to make collective decisions through platforms like Snapshot. This decentralized model democratizes governance but introduces complexities in data management and analysis.

DAOs have significant implications for information systems, management, and sociology. They streamline operations by reducing administrative overhead, enhancing transparency and trust, and facilitating global collaboration without physical offices. DAOs also offer unique compensation structures, allowing contributors to earn tokens for specific tasks, fostering democratic, flexible, and fulfilling work environments.

Uniswap, a decentralized exchange, exemplifies a successful DAO. It allows users to trade cryptocurrencies directly using automated smart contracts. Uniswap's DAO framework empowers token holders to vote on critical decisions, ensuring community-driven development. This success showcases DAOs' potential to efficiently manage and grow complex projects, providing valuable insights for researchers and practitioners.

This project focuses on data from 15 DAOs, particularly during the Temperature Check (Snapshot Poll) governance step. Data from Snapshot, Polygonscan, and Etherscan underwent extensive cleaning and structuring. MongoDB was chosen for its flexibility in handling complex, nested data structures, enabling a comprehensive analysis of governance and operational efficiencies within these DAOs.

2. Methodology

a) Data Loading

The initial phase involved loading extensive data from multiple pickle files into MongoDB collections. This data encompassed DAO-level information, proposals, votes, follows, and wallet-level transactions. The process required careful handling of data types, particularly large integers, to ensure compatibility with MongoDB's storage capabilities. For instance, integers that exceeded the maximum size were converted to strings to prevent overflow issues. This step was crucial to maintain data integrity and avoid potential errors during data storage and retrieval.

b) Data Cleaning, Manipulation, and Structuring

i. Created New Collections

Rationale: Separating different components into respective collections makes the data easier to manage and query. This modular approach aligns with MongoDB's strengths in handling hierarchical data.

Components: This step enhance data manageability and querying efficiency, components are:

- Strategies - Contains details about different voting strategies.
- Treasuries - Contains treasury addresses.
- Validation - Contains validation rules and parameters.
- Admins - List of admin addresses.
- Members - List of member addresses.
- DAOs - Contains the main data structure with comprehensive details about the DAOs, including id, name, rank, twitter, website, about, network, symbol, strategies, filters, plugins, treasuries, voting, validation, votesCount, activeProposals, proposalsCount, proposalsCount7d, and other fields.
- Spaces - Each DAO can have multiple spaces.

This segregation ensured that each component of the DAOs was independently accessible and manageable.

ii. Cleaning Each Document and Inserting Components

- *Redundant Fields* - Removed consistently null or empty fields (e.g., email, github, website) to retain only valid data.
- *Normalization* - Ensured data consistency by converting addresses to lowercase, converting large numbers to strings, and correctly formatting numeric fields.
- *Type Inconsistencies* - Addressed inconsistencies to ensure uniform data formats, enhancing accuracy and reliability in data analysis.
- *Duplicates* - Identified and removed duplicate records to maintain dataset integrity and reduce storage costs.
- *Nesting* - Unnested data to simplify access to related information, enhancing query performance.

iii. Restructuring

Implemented a comprehensive cleaning and restructuring process for each collection to ensure data integrity and usability. For example, DAO-level data was decomposed into multiple collections for better organization and access.

iv. Detailed Operations

To enhance data quality and organization, the following steps were undertaken:

- *Null Fields* - Fields with null or empty values were systematically removed from all documents. This process retained only relevant information, reducing data redundancy and enhancing overall quality. For example, consistently empty fields like email, GitHub, and website were removed.
- *DAO Information Restructuring* - Initially stored in a single collection, DAO-level data was decomposed into multiple collections such as daos, spaces, strategies, treasuries, validation, admins, and members. This modular approach facilitated better organization and more efficient access to related data segments. For instance, each DAO's space details were moved to a separate spaces collection, enabling targeted queries about specific spaces without needing to parse through the entire DAO document.
- *Proposals Cleaning* - Proposals data underwent thorough cleaning, involving the removal of null fields and storage in a dedicated collection named cleaned_proposals. This ensured the integrity and usability of the proposals data for subsequent analysis, reducing noise and focusing on critical information.
- *Unnesting Follows* - Nested follows data was unwrapped and stored in a new collection, unnested_follows. This involved extracting individual follow records, normalizing critical fields like space_id and follower, and removing duplicates. Data normalization, such as converting space_id and follower to lowercase, ensured consistency across the dataset, simplifying data merging and querying.
- *Unnesting and Cleaning Transactions* - Transaction data from Polygonscan and Etherscan required unnesting to transform nested transaction arrays into individual documents. The data was cleaned by removing duplicates, normalizing addresses, and ensuring numeric fields were correctly formatted. Cleaned transaction data was stored in unnested_ether and unnested_polygon collections, with addresses converted to lowercase and null address transactions removed to maintain integrity.
- *Votes Unnesting*- Votes data was unnested and cleaned, ensuring all relevant voting information was accurately represented in the unnested_votes collection. This process allowed direct access to individual vote details, streamlining analysis without additional parsing.

These detailed operations ensured the dataset was well-organized, clean, and ready for efficient querying and analysis.

v. Processed Dataset

The outcome of this extensive cleaning and structuring process was a well-designed dataset that complies with MongoDB’s document-oriented approach. The dataset is characterized by:

- *Modularity* - Data is organized into multiple, interrelated collections, allowing for efficient querying and analysis. Each collection serves a specific purpose, reducing redundancy and enhancing clarity. For instance, separating DAO data into spaces, strategies, and other collections allows for focused queries and better data management.

- *Normalized Data* - Key fields across all collections are normalized, ensuring consistency and reliability in data representation. For example, converting all addresses to lowercase and ensuring numeric fields are correctly formatted prevent inconsistencies and errors during data processing.

- *Reduced Redundancy* - By removing null and duplicate entries, the dataset is streamlined, containing only valuable and relevant information. This not only saves storage space but also improves query performance.

- *Enhanced Accessibility* - The dataset's structure facilitates easy access to interconnected data points, supporting complex queries and analytical tasks. For example, the separation of DAO details into various collections means queries can be more specific and efficient.

3. Results

This section provides descriptive insights from the cleaned and processed dataset, demonstrating trends, participation dynamics, and governance activities within the DAOs. These insights are critical for understanding DAO functions and identifying areas for improvement in governance models.

A) Popularity and Engagement in DAOs

i. Top Followed DAOs

The analysis of follows within different DAO spaces revealed which DAOs are most attractive to the community. Figure 1 below presents top followed DAOs from our dataset. These DAOs draw the largest number of community members, indicating strong interest and potential influence within the blockchain ecosystem. Understanding which DAOs are most followed helps in identifying which projects are gaining traction and could potentially influence future developments in the DAO space.

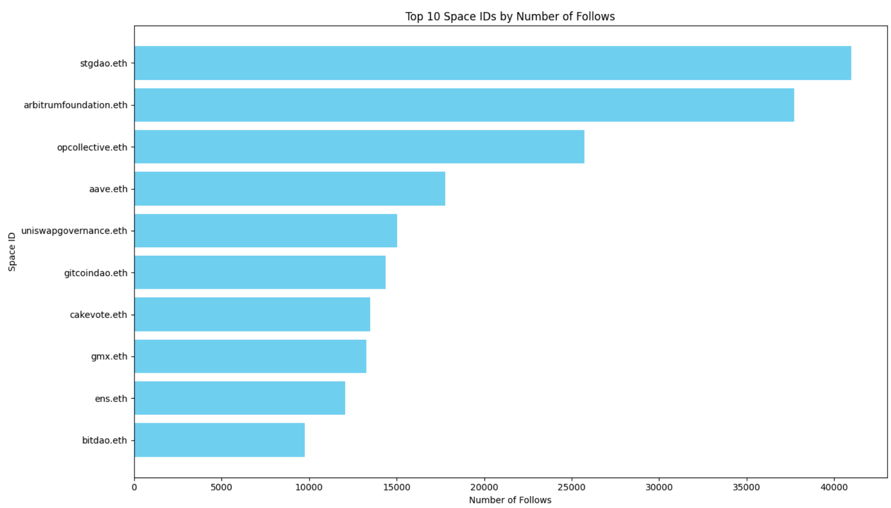

Figure 1. Top followed DAOs presented by Space IDs

ii. Member VS Admin Dynamics

We examined the distribution of admins and members to understand the structure of participation within DAOs. The pie chart in Figure 2 shows that the majority of participants are members only, with a smaller portion holding administrative roles and an even smaller group performing both functions. This indicates a hierarchical structure within DAOs with distinct roles and responsibilities. Identifying these dynamics is essential for understanding the internal governance and operational efficiency of DAOs.

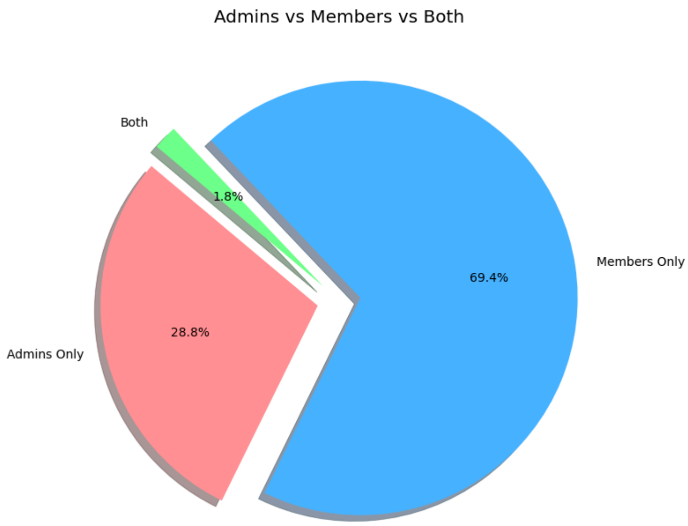

Figure 2. Pie Chart for Admin and Members

iii. Proposals and Votes

To gauge the level of engagement within these DAOs, we analysed the number of proposals made and votes cast per space. The chart in Figure 3 clearly indicates that certain DAOs are highly active in decision-making processes. For instance:

- arbitrumfoundation.eth and aave.eth showed high numbers of both proposals and votes, indicating strong participation and engagement from their communities.
- opcollective.eth and cakevote.eth also showed significant activity, though to a lesser extent.

This highlights the DAOs where community governance is most active, reflecting a robust participatory culture. Understanding the level of engagement helps in assessing the effectiveness of governance mechanisms and the overall health of the DAO community.

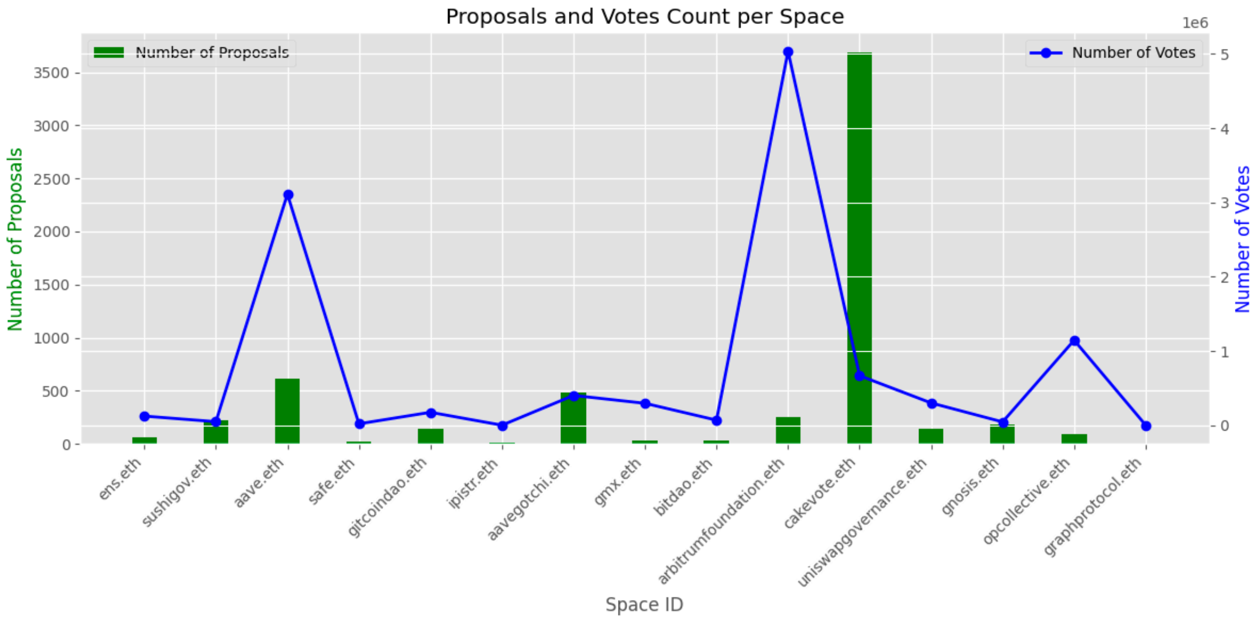

Figure 3. Proposals and Votes Count per Space ID

B) DAO Governance Process

i. Common Voting Strategies

The bar chart on common voting strategies in Figure 4 reveals various methods DAOs use to facilitate governance and decision-making.

- Pagination and erc20-balance-of are the most common strategies, employed by 8 and 7 DAOs respectively. These strategies allow for flexible and scalable decision-making processes, essential for managing large and active communities.

- Other prevalent strategies include contract-call, delegation, and erc20-votes, each used by multiple DAOs. These methods offer different levels of automation, delegation, and token-based voting, catering to various governance needs and member preferences.

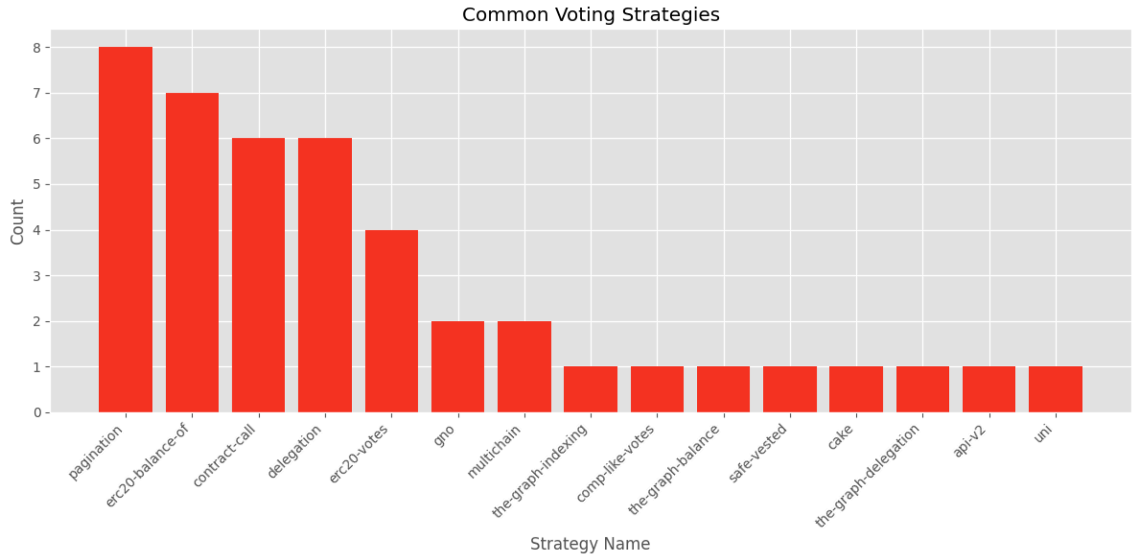

Figure 4. Common Voting Strategies

These strategies demonstrate the diverse approaches DAOs employ to ensure effective and inclusive governance. The choice of strategy can significantly impact the efficiency of proposal voting, the inclusiveness of member participation, and the overall governance experience. Understanding these strategies helps in evaluating the governance frameworks and identifying best practices for different types of DAOs.

ii. Voting Power Distribution

The analysis of voting power distribution highlighted the concentration of influence within DAOs as shown in Figure 5.

- The top 10 voters hold significant voting power, with the highest having over 4 billion votes. This suggests a high level of centralization in voting power, where a small number of participants can have a substantial impact on governance outcomes.

- The distribution showed a notable disparity, with the top few voters holding substantial influence while the majority of voters have comparatively less power. This could lead to scenarios where the interests of a few major stakeholders dominate the decision-making process, potentially marginalizing smaller holders.

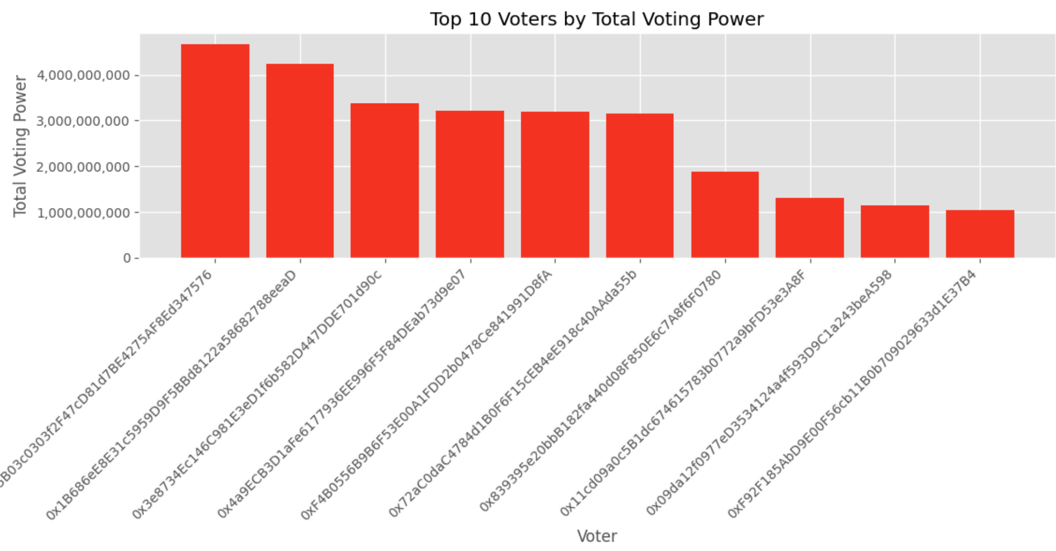

Figure 5. Top Voters by Power Distribution

This concentration of voting power can impact the democratic nature of DAO governance. It underscores the importance of mechanisms to distribute voting power more evenly or incentivize broader participation to ensure decisions reflect the diverse interests of the community. Addressing these issues is crucial for maintaining the legitimacy and inclusiveness of DAO governance.

C) Block Chain Transaction Analysis

i. Transaction Volume and Value Over Time

- The analysis of transaction volumes and values over time revealed significant periods of activity within the blockchain network. Figure 6 demonstrates that the number of transactions per timestamp steadily increased, with noticeable spikes. These spikes could correspond to major events such as network upgrades, market volatility, or significant protocol changes, highlighting moments of intense activity in the blockchain ecosystem.
- Figure 7 shows a similar trend, with the total value transferred (in Ether) increasing over time and peaking at specific intervals. These peaks in transaction value may indicate large transactions or a surge in high-value transactions, potentially correlating with the same events that caused spikes in transaction volume.

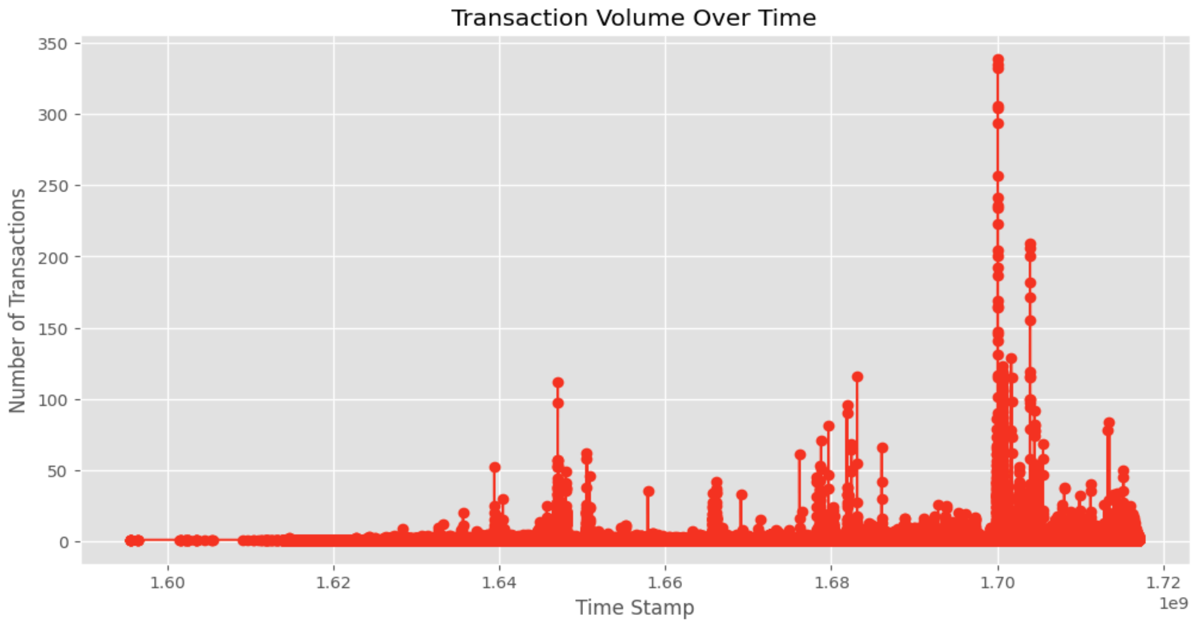

Figure 6. Transactions Volume over time

- Identifying these periods of high activity helps in understanding the factors driving network usage and can provide insights into market behaviour and the impacts of specific events on the blockchain ecosystem.

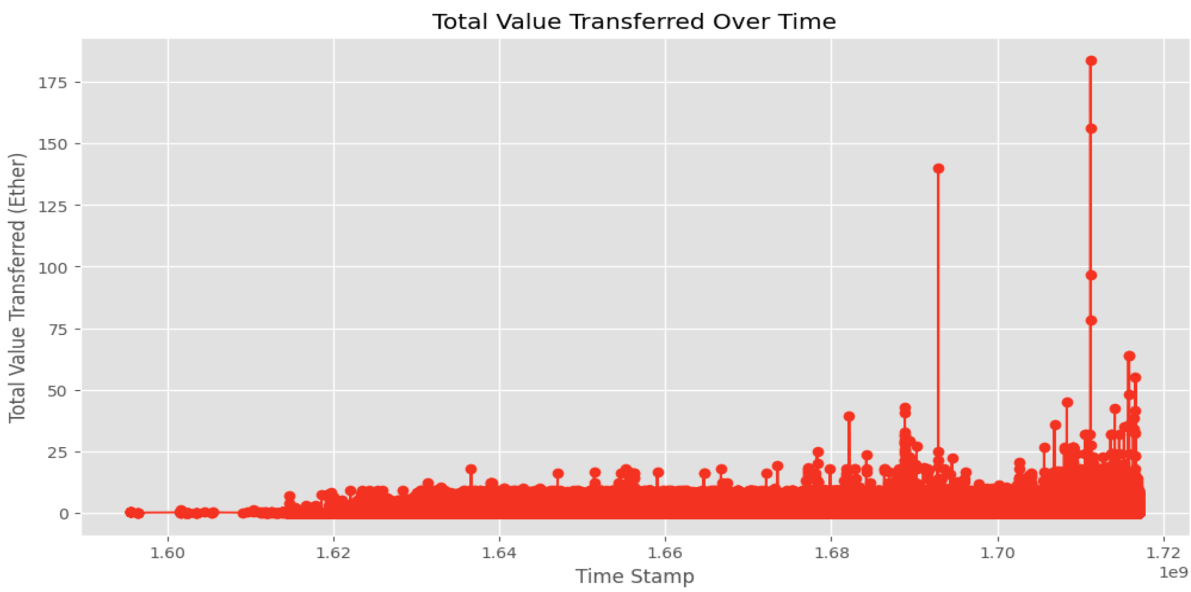

Figure 7. Total Value over time

ii. Top Participants in Transactions

Figure 8 identifies the most active participants in the network based on the number of transactions they have sent. The data revealed that a small number of addresses are responsible for a significant portion of the transactions. The leading address stood out with an exceptionally high number of transactions, indicating either a highly active user or an automated system/bot conducting numerous transactions. Understanding who the key players are in the transaction ecosystem can provide insights into network dynamics and potential centralization risks.

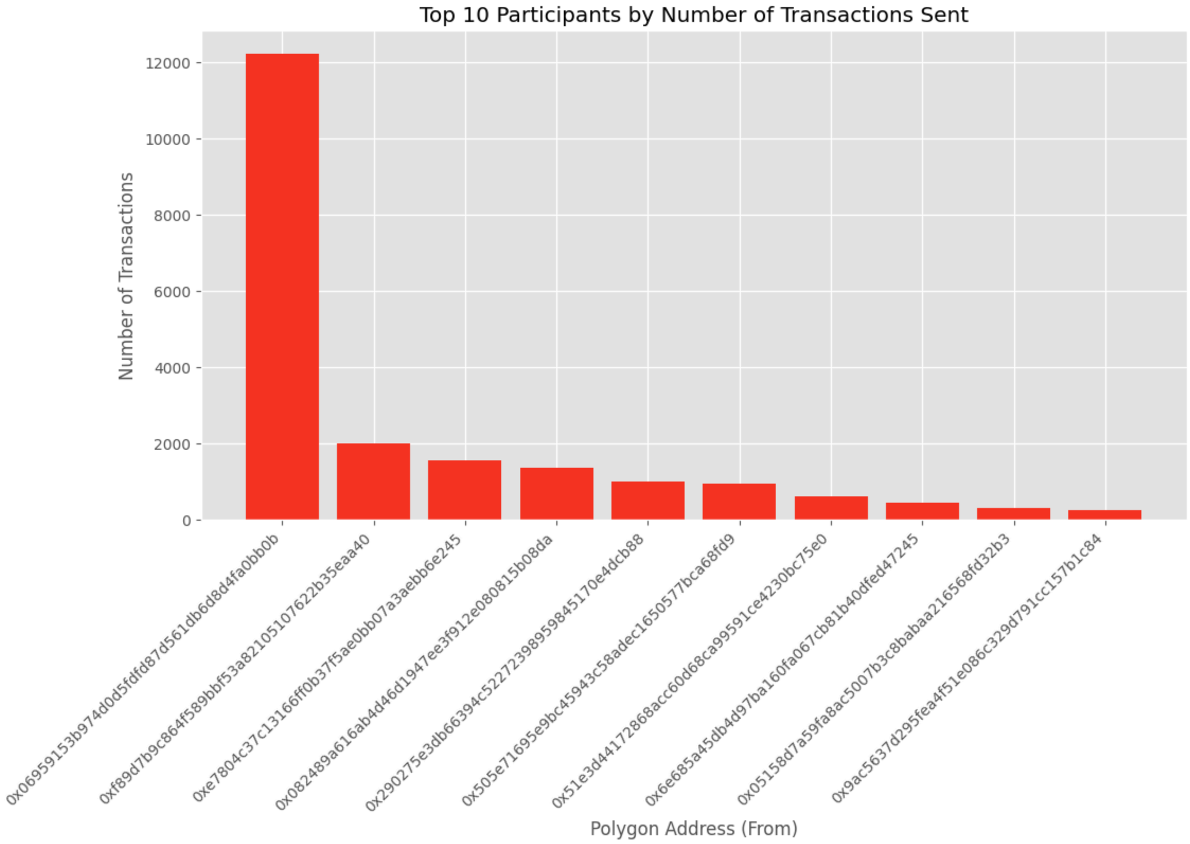

Figure 8. Top Participants by Number of Transactions

- **Periods of High Activity:** Both transaction volume and value charts indicated specific periods where blockchain activity surged. These periods may align with significant events in the blockchain world, such as market booms, regulatory announcements, or major updates to the blockchain protocol. Identifying these periods can help stakeholders understand the impact of such events on network activity.

- **High-Value Transactions:** The spikes in total value transferred suggested moments when large amounts of Ether were moved across the network. These could be indicative of significant investments, large-scale trades, or major transactions between entities. Monitoring these spikes can provide insights into market behaviour and investor sentiment.

- **Key Players in the Ecosystem:** The analysis of top participants by transaction count highlighted the concentration of activity among a few addresses. This concentration suggested that a small number of participants or automated systems dominate transaction activity, which could have implications for network decentralization and participant behaviour.

By analysing these insights, stakeholders can gain a better understanding of network dynamics, the impact of significant events on blockchain activity, and the potential centralization of transaction activities. This information is crucial for making informed decisions in the blockchain ecosystem.

D) Operational Efficiency

i. Gas Fee Analysis for Ether and Polygon

In decentralized autonomous organizations (DAOs) and blockchain transactions, understanding operational costs, particularly gas fees, is crucial. Gas fees are the costs required to conduct transactions on blockchain networks.

Total Gas Fees

The analysis revealed the following total gas fees for two prominent blockchain networks as also represented in Figure 9.

Ether - The total gas fees for Ether transactions amounted to 873,458,230,403,084,421,096 Wei, equivalent to approximately 873.46 Ether.

Polygon - For Polygon, the total gas fees were significantly higher, reaching 17,815,747,514,356,339,002,419 Wei, which is about 17,815.75 Ether.

Total Transactions

The number of transactions processed on each network also highlights the activity level:

Ether - A total of 586,170 transactions were recorded.

Polygon - The Polygon network processed 1,064,437 transactions.

Average Gas Fees per Transaction

Evaluating the average gas fee per transaction provides insights into the cost-efficiency of each network.

Ether - The average gas fee per transaction was 1,490,110,770,600,823 Wei, which translates to approximately 0.00149 Ether.

Polygon - On Polygon, the average gas fee per transaction was substantially higher at 16,737,249,376,295,956 Wei, or about 0.0167 Ether.

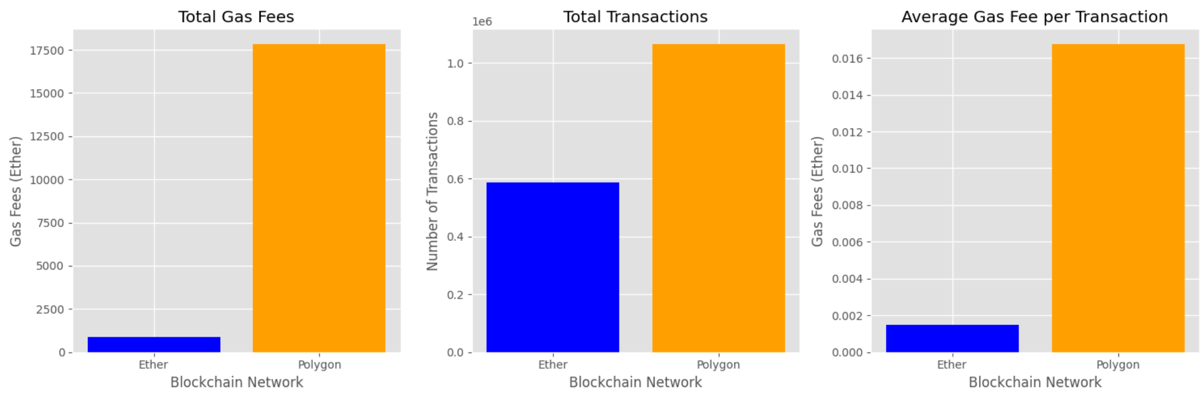

Figure 9. Gas Fee Analysis

The data points to several key insights regarding the operational efficiency of these blockchain networks:

- **Higher Total Gas Fees on Polygon:** The total gas fees on the Polygon network are substantially higher than those on Ether. This difference is driven by the higher transaction volume on Polygon.

- **Transaction Volume Disparity:** Polygon's higher transaction count indicates more extensive usage and possibly greater adoption for various applications. However, this increased activity comes with higher total operational costs.

- **Cost per Transaction:** The average gas fee per transaction on Polygon is significantly higher than on Ether. This suggests that conducting transactions on Polygon is more expensive on a per-transaction basis. The reasons could include network congestion, higher gas prices, or more complex transactions.

- **Implications for DAOs:** For DAOs operating on these networks, the higher gas fees on Polygon might affect smaller participants or frequent micro-transactions. This could lead to a preference for networks with lower transaction costs to maximize participation and reduce overhead expenses.

Understanding gas fees is essential for evaluating the cost-efficiency of blockchain networks and planning operational strategies for DAOs. The significant difference in gas fees underscores the need for efficient transaction processing and cost management. Optimizing for lower gas fees could enhance participation and engagement by making transactions more affordable, supporting the sustainability and scalability of blockchain networks and DAOs. Efforts to reduce gas fees, such as improving network efficiency, adopting layer-2 solutions, or using more cost-effective blockchain platforms, could drive broader adoption and foster a more inclusive ecosystem.

4. Conclusion

The comprehensive process of cleaning, manipulating, and structuring DAO governance data has resulted in a robust and analysable dataset, offering valuable insights into DAO functioning and dynamics. Leveraging MongoDB's document-oriented architecture, we achieved a modular, normalized, and accessible dataset supporting detailed analysis and complex queries.

Our analysis uncovered key aspects of DAO popularity, engagement, governance processes, and operational efficiency. We identified the most followed DAOs, examined member vs admin dynamics, analysed proposals and voting patterns, and evaluated common voting strategies and voting power distributions. Additionally, we conducted an in-depth analysis of blockchain transactions, highlighting periods of high activity, key participants, and gas fee implications.

These insights provide a deeper understanding of DAO governance and participation, revealing strengths and challenges of current models. Addressing issues such as concentration of voting power and high operational costs can help DAOs evolve towards more democratic and cost-effective governance structures. This project demonstrates DAOs' potential to revolutionize organizational governance through decentralized, transparent, and community-driven processes. Insights gained can guide future research, inform the design of new DAOs, and help existing ones optimize governance practices for better participation and sustainability.