ORIE 5255 Final Report

Github: liolio66666/ORIE5255_Crypto_Project

Introduction

The rapid evolution of blockchain technology fostered innovative approaches to optimize performance, enhance user experience, and predict key behaviors in decentralized networks. Through dimensionality reduction using Principal Component Analysis (PCA) and clustering via K-Mean, this report focuses on leveraging Uniswap Universal Router transaction data obtained via Etherscan to provide predictive insights that align with Monad Labs' mission to build a high-performance EVM Layer-1 blockchain.

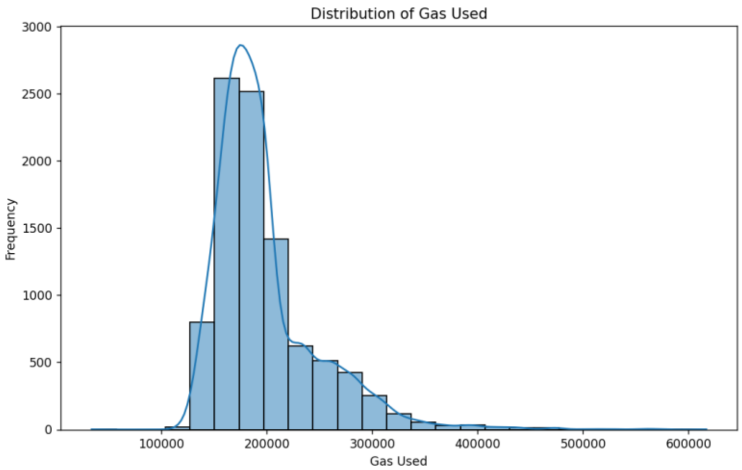
Specifically, the results of this study could inform Monad Labs' efforts to optimize transaction costs, improve predictive capabilities for user activity, and enhance the economic efficiency of its blockchain infrastructure. By applying these cutting-edge techniques to real-world transaction data, this report contributes to Monad Labs' objective of leveraging data science and machine learning to solve open-ended, impactful problems in the blockchain space.

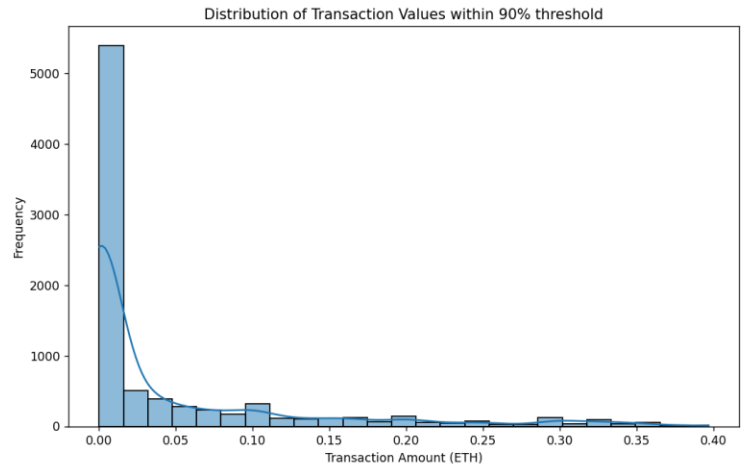
Exploratory Data Analysis

We first downloaded most recent 10000 Uniswap Universal Router transactions using Etherscan API ranging from 2024-11-20 22:37:11 to 2024-11-21 02:24:24:59. The transaction data have 20 columns including but not limited to timeStamp, transactionIndex (the position of transaction within specific block), gas (the maximum of gas sending willing to pay), gasPrice (price of the gas in Wei), gasUsed (actual amount of gas used), value (amount of Ether transferred in Wei), etc.

Out of 10000 transactions, 529 transactions have errors indicated by isError and txreceipt_status columns. We will delete those columns in our analysis. All transactions have NA value in contractAddress column, this means that the transaction does not involve deployment of smart contracts, we can ignore this column in our transaction.

For all valid transactions, figures below shows the dstribution of gasUsed for all transaction and the value of the transaction within 90% threhold in ETH (1ETH = 1e18 Wei). The value of the transaction varies significantly with the average 0.2 ETH while maximum transaction is 46.43 ETH. Most of transaction have 100000 to 300000 gas used.

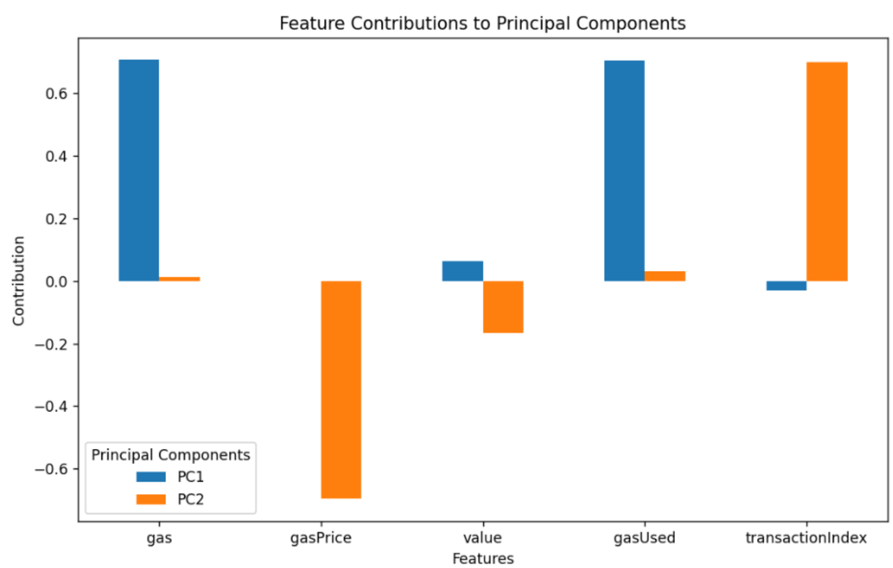




PCA

We then conducted Principal Component Analysis (PCA) on five key transaction features: gas, gasPrice, value, gasUsed, and transactionIndex. These features represent essential aspects of blockchain transaction behavior. PCA reduces the dimensionality of the dataset, simplifying the analysis of complex relationships while retaining the most significant variance. The first two principal components (PC1 and PC2) explain 62.0% of the total variance, with PC1 capturing 37.7% and PC2 accounting for 24.3%, highlighting that a significant portion of the data's variability can be understood through these two dimensions.

The histogram on the right shows features contributions to PC1 and PC2. PC1 is dominated by gas (0.706) and gasUsed (0.705), indicating that it primarily reflects transaction complexity, where higher gas limits and higher actual gas usage contribute to a transaction's computational effort. Value, representing the Ether amount transferred, has a small positive contribution to PC1 (0.064), suggesting it



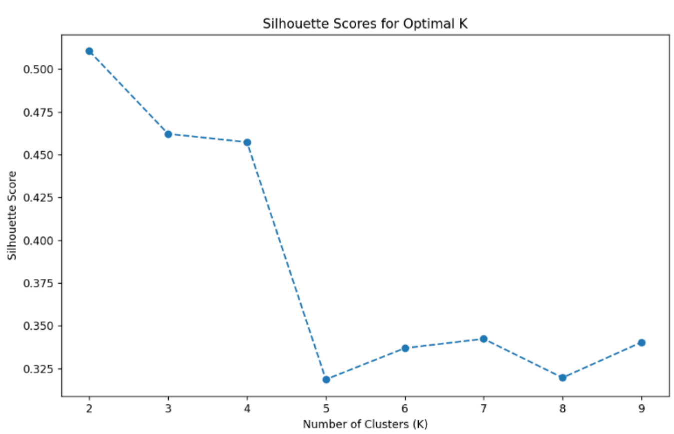
plays a relatively minor role in defining transaction complexity. On the other hand, PC2 captures a distinct aspect of transaction behavior, driven by gasPrice (-0.696) and transactionIndex (0.698). This component reflects the trade-off between economic incentives and transaction timing, where higher gas prices are associated with earlier inclusion in a block (lower transaction index), illustrating the prioritization dynamics in the blockchain network.

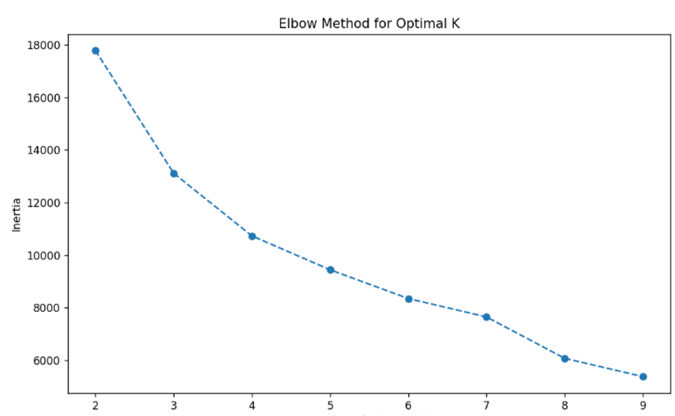
Through dimensionality reduction, PCA provides a clearer picture of transaction behavior by isolating the key drivers of variance. Transactions are primarily differentiated by their computational demands (PC1) and economic dynamics (PC2). This simplification allows for the identification of behavioral patterns and potential inefficiencies. The results can be used for clustering transactions, detecting outliers, and building predictive models, offering actionable insights into optimizing transaction costs and blockchain performance. By focusing on these two principal components, we retain the most critical information while reducing complexity, paving the way for further analysis and decision-making.

K means and clustering

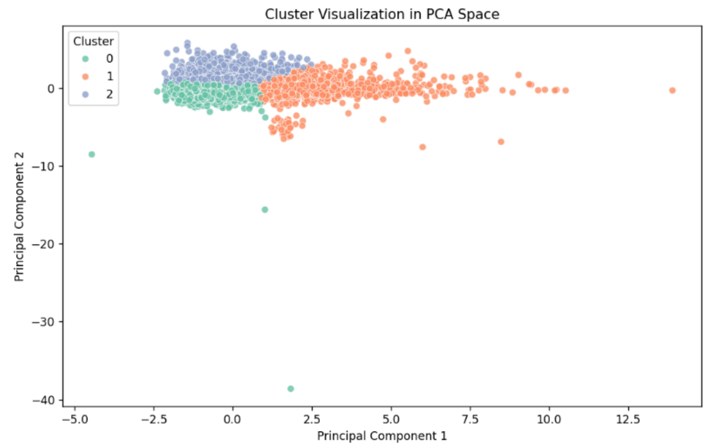
To further explore patterns in transaction behavior, we applied K-Means clustering to the reduced dataset obtained from Principal Component Analysis (PCA). By clustering the transactions in PCA space, the aim was to identify distinct groups of transactions based on their shared characteristics. This approach provided deeper insights into the data, revealing key behavioral patterns and potential inefficiencies within the transactions.

To determine the optimal number of clusters, we relied on two methods: the Elbow Method and the Silhouette Score. The Elbow Method evaluated the within-cluster sum of squares (inertia) for different number of clusters, and the Silhouette Score, which measures the separation and compactness of clusters. From our results, we selected 3 cluster as the optimal number of clusters for our analysis.





After applying the K-Means algorithm with 3 clusters, each transaction was assigned to one of the three clusters. The clusters were visualized in PCA space, demonstrating distinct groupings of transactions. The separation between the clusters indicated that the algorithm effectively identified transactions with similar behavioral characteristics.

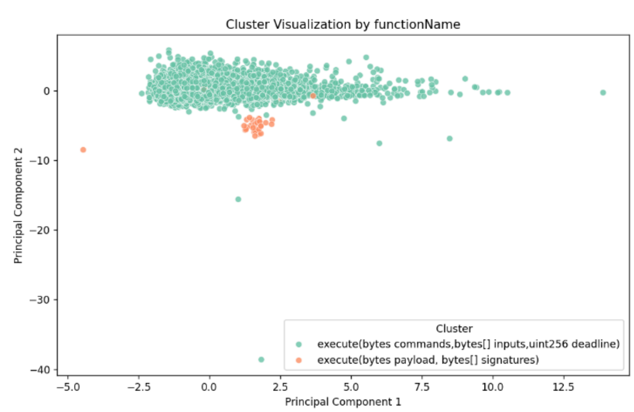


Furthermore, to better understand the underlying differences between the clusters, we calculated the mean values of the key transaction features—gas, gasPrice, value, gasUsed, and transactionIndex for each cluster. Cluster 1 contained transactions with higher gas and gasUsed values, representing transactions with significant computational demands. These transactions may indicate periods of higher network activity or users engaging in more complex operations. Cluster 0 was defined by higher gas prices and lower transaction indices, highlighting transactions that users prioritized for earlier inclusion in a block by offering higher fees. In contrast, Cluster 2 was characterized by lower gas prices and higher transaction indices.

Average	gas	gasPrice(Wei)	Value(ETH)	gasUsed	TransactionIndex
Cluster 0	2.52e5	1.10e10	0.18	1.78e5	28.8
Cluster 1	2.94e5	1.09e10	0.31	2.78e05	33.5
Cluster 2	2.74e5	0.96e10	0.12	1.85e5	120.7

Cluster with functionName

In the clustering analysis, a distinct group of transactions was identified, primarily associated with the function execute(bytes payload, bytes[] signatures). This function is commonly utilized in multisignature (multisig) wallets, which require multiple approvals before executing a transaction. Notably, over half of these transactions had their gas limit set to 50,000 units. This



uniformity suggests a standardized approach to estimating the gas required for multi-sign transactions, reflecting the routine nature of their operations. Additionally, these transactions exhibited higher gas prices compared to other clusters, indicating that users were willing to pay a premium to prioritize these transactions for faster inclusion in blocks. The elevated gas prices suggest that these transactions were time-sensitive, necessitating prompt processing.

The clustering of these transactions, based on their function name, underscores the influence of transaction type on gas usage and pricing strategies. Transactions involving multi-sign functions tend to have predictable gas requirements and are often processed with higher gas prices, reflecting their standardized yet time-sensitive nature. This insight is valuable for understanding user behavior and optimizing transaction processing within blockchain networks.

Conclusion

In conclusion, this report’s analysis of Uniswap's Universal Router transaction data offers valuable insights that can significantly aid Monad Labs in enhancing its EVM Layer-1 blockchain. By employing Principal Component Analysis (PCA) and K-Means clustering, we identified key factors influencing transaction behavior and categorized transactions into distinct clusters. Further study can be done by cluster transaction with transaction value and frequencies as well as predictive modelling for gas price and gas usage. By integrating these insights, Monad Labs can enhance the performance and economic efficiency of its blockchain infrastructure, aligning with its mission to leverage data science and machine learning in addressing complex challenges within the blockchain ecosystem.