Ocean Protocol x Curve finance Data Challenge

THIS REPORT WILL STUDY KEEN INSIGHTS INTO THE DYNAMICS OF CURVE FINANCE, AMM'S, AND HOW AUTOMATION CAN SUPPORT A GLOBAL FINANCIAL SYSTEM BUILT ON BLOCKCHAIN.

Elisa Anderson | Desights Data Challenge | 10/11/2023

Introduction

Overview of Curve Finance

Curve Finance, a prominent player in the decentralized finance (DeFi) landscape, offers an innovative platform for efficient cryptocurrency trading, particularly for stablecoins.

As a decentralized exchange, Curve utilizes an Automated Market Maker (AMM) model to facilitate liquidity and reduce slippage in trades. This report delves into a comprehensive analysis of Curve Finance's transaction data, aiming to uncover insightful trends, detect anomalies, and predict future transaction values.

Objectives of the Report

The primary objectives of this report are to:

Explore Trading Dynamics: Understand the key factors influencing trading activities on CurveFinance, including transaction volume, gas fees, and the impact of liquidity.

Identify Anomalous Activities: Detect unusual patterns that could signify market manipulation tactics like wash trading, front-running, or pump-and-dump schemes.

Predictive Analysis: Leverage advanced machine learning techniques, specifically Convolutional Neural Networks (CNNs), to forecast future transaction values, contributing to the development of informed trading strategies.

Trading Strategy Formulation: Propose potential trading strategies based on historical data analysis, considering factors such as transaction costs, market volatility, and liquidity.

Report Structure

The report is structured as follows:

- Section 1: Data Analysis and Exploratory Data Analysis (EDA): Presents a detailed exploration of transaction data, including trends and patterns in trading volume, gas fees, and liquidity.
- Section 2: Unusual Activity Detection: Discusses methodologies and findings related to the identification of abnormal trading behaviors on the Curve Finance platform.
- Section 3: Predictive Modeling: Explores the application of CNN models to predict transaction values, along with a thorough evaluation of model performance metrics.
- Conclusion: Summarizes the key findings and their implications for traders, investors, and stake- holders in the DeFi space. Relevance and Importance

Relevance and importance

The burgeoning field of DeFi presents both opportunities and challenges. As decentralized platforms like Curve Finance become increasingly integral to the crypto trading landscape, understanding the nuances of these markets is essential.

This report aims to provide valuable insights that can aid in making informed decisions, ensuring efficient trading practices, and contributing to the overall stabilityand growth of decentralized financial systems.

Contents

Introduction 1

Overview of Curve Finance 1

Objectives of the Report 1

Report Structure 2

Relevance and importance 2

2 Exploration Data analysis 5

2.1 Histogram of Gas Prices 5

2.2 Histogram of cumulative Gas Used 5

2.3 Scatter plot of Gas Price vs. Cumulative Gas Used 6

2.4 Box Plot for Transaction Values 6

2.5 Distribution of Balances in Different Balance Ranges 7

2.6 Discussions about the EDA 7

Insights from Visual Data Analysis 7

3 Historical Evolution 8

3.1 Daily trading volume on curve Finance (7 day rolling Average) 8

3.2 Dates and amounts with the most significant changes in trading volume: 9

3.3 Time series of Gas Prices 10

3.4 Dates with the most significant gas price changes 10

4 Impact of veCRV: 11

4.1 Comparison of time SERIES: 11

4.1.1 Daily transaction volume on curve Finance: 11

4.1.2 Daily average Gas price on curve Finance 12

4.2 Statistical Analysis : 13

5 Unusual activities 14

5.1 Volume spikes that might indicate pump-and-dump: 14

5.2 Price spikes that might indicate market manipulation: 14

5.3 Accounts that could be top traders or manipulators: 15

6 Trading Volume VS Liquidity: 16

7 Integration of Curve's WMA Design in Cross-Border CBDC 18

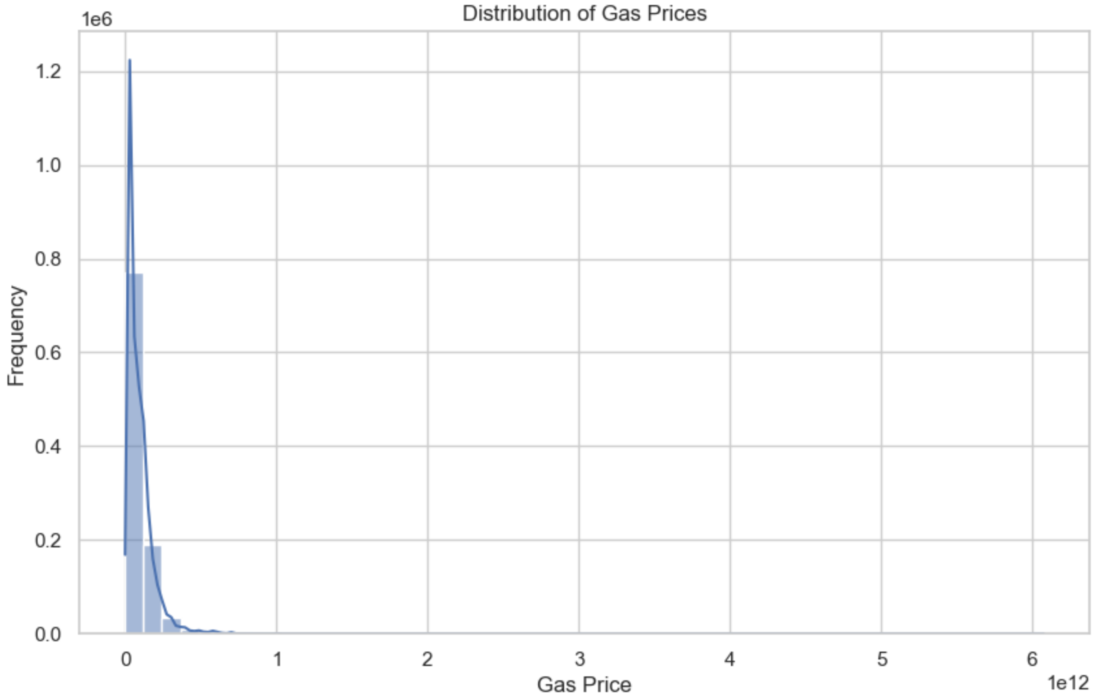
8 Predictive Analysis: 21

8.1 Preliminaries: 21

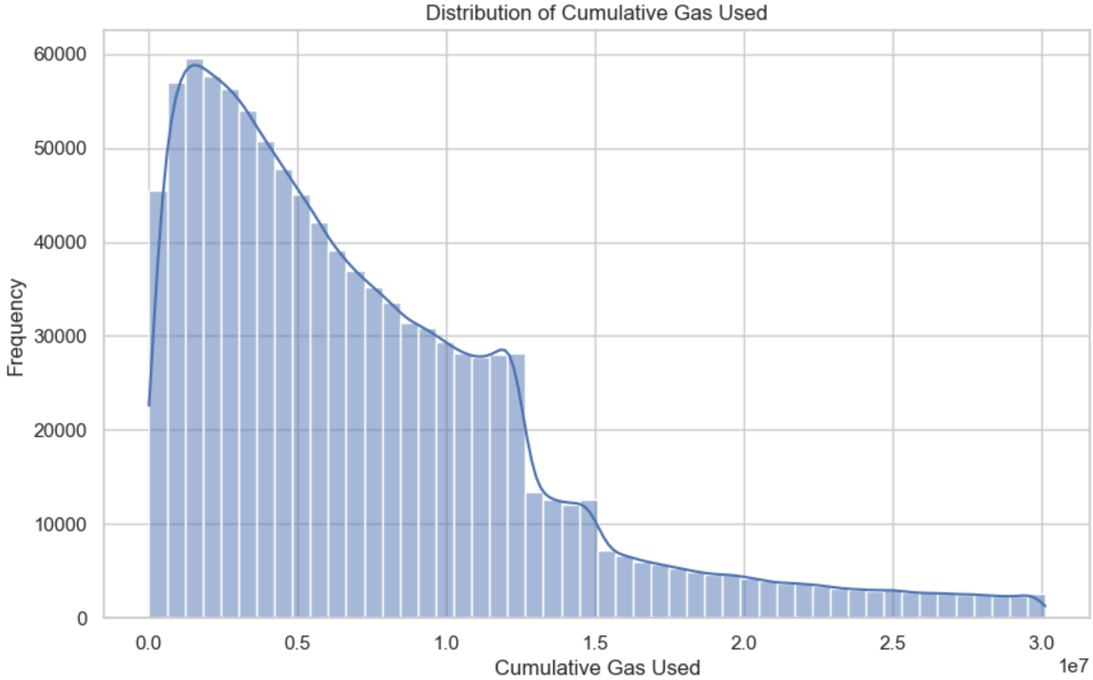
8.2 Theory of RandomForest Regressor 22
8.3 Our fitted model:.. 23
Conclusion:... 25
Reflective Summary 25
Forward-Looking Statements 25
Recommendations for Future Work 25

2 Exploration Data analysis

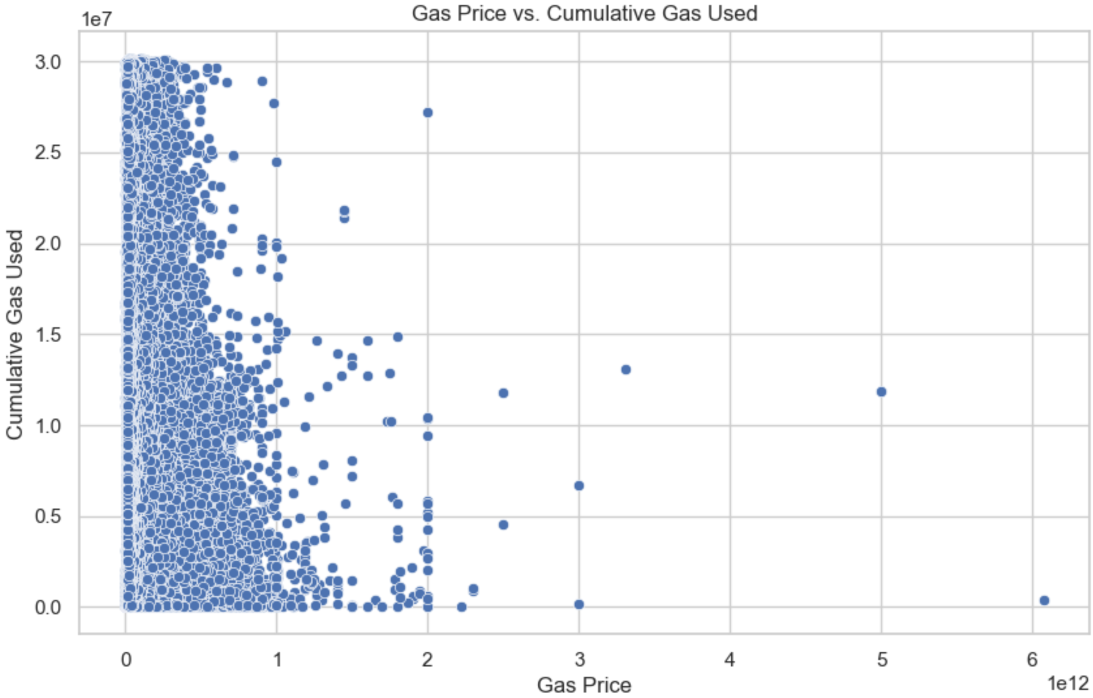
2.1 HISTOGRAM OF GAS PRICES



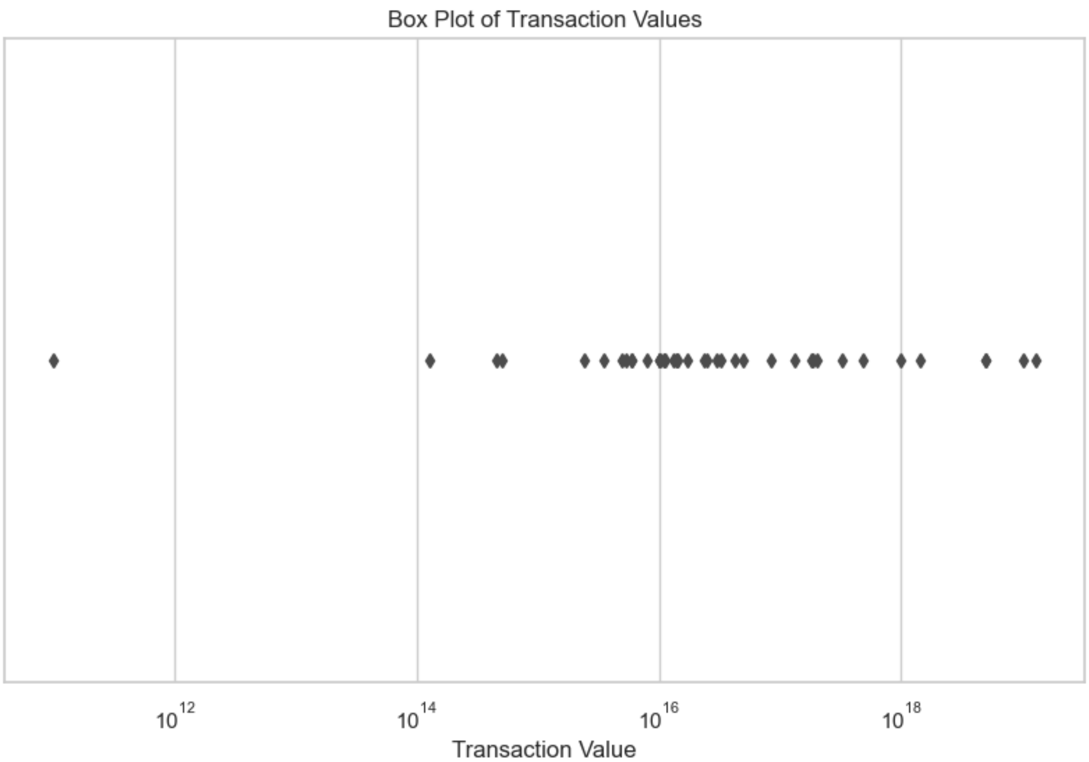
2.2 HISTOGRAM OF CUMULATIVE GAS USED



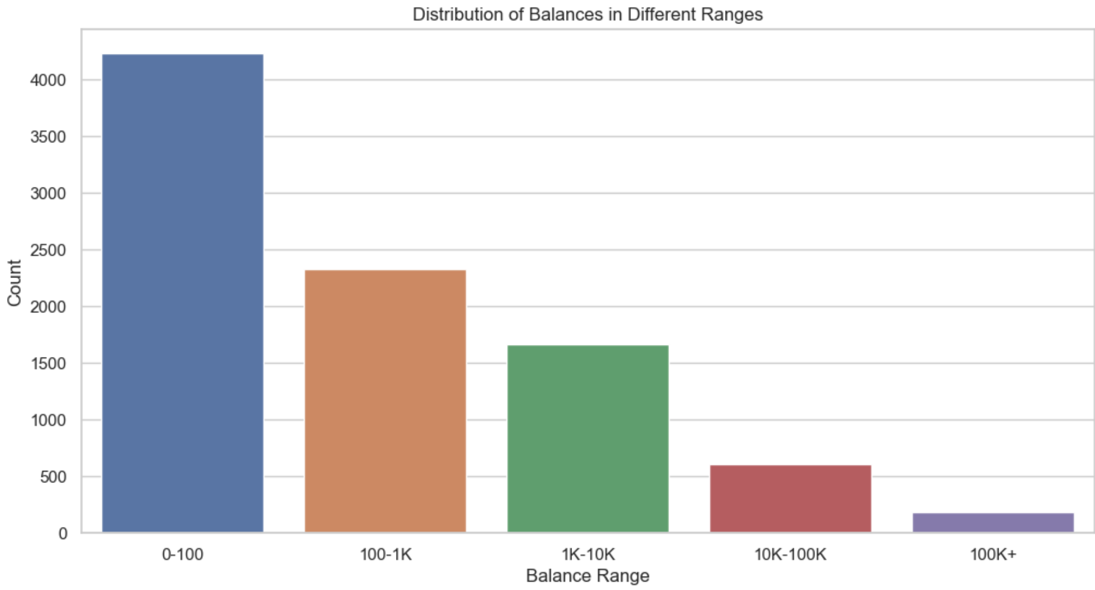
2.3 SCATTER PLOT OF GAS PRICE VS. CUMULATIVE GAS USED



2.4 BOX PLOT FOR TRANSACTION VALUES



2.5 DISTRIBUTION OF BALANCES IN DIFFERENT BALANCE RANGES



2.6 DISCUSSIONS ABOUT THE EDA

Insights from Visual Data Analysis

Distribution of veCRV Holder Balances:

The balance distribution reveals a significant skew towards smaller balances, with the majority of holders having between 0-100 veCRV tokens. This suggests that while veCRV is distributed across a wide user base, the majority of users hold a relatively small stake in the governance process, which may impact decentralized decision-making.

Gas Price vs. Cumulative Gas Used:

The scatter plot indicates a clustered relationship between gas price and cumulative gas used, with most transactions occurring at lower gas prices and using less gas. This might reflect the efficiency of the Curve protocol's transactions or users' tendency to transact during less busy network periods to minimize fees.

Distribution of Cumulative Gas Used:

The histogram of cumulative gas used shows a right-skewed distribution, indicating that while most transactions use a lower amount of gas, there are enough high-gas-use transactions to affect the average, suggesting that there are occasional complex transactions or smart contract interactions that require significant gas.

Distribution of Gas Prices:

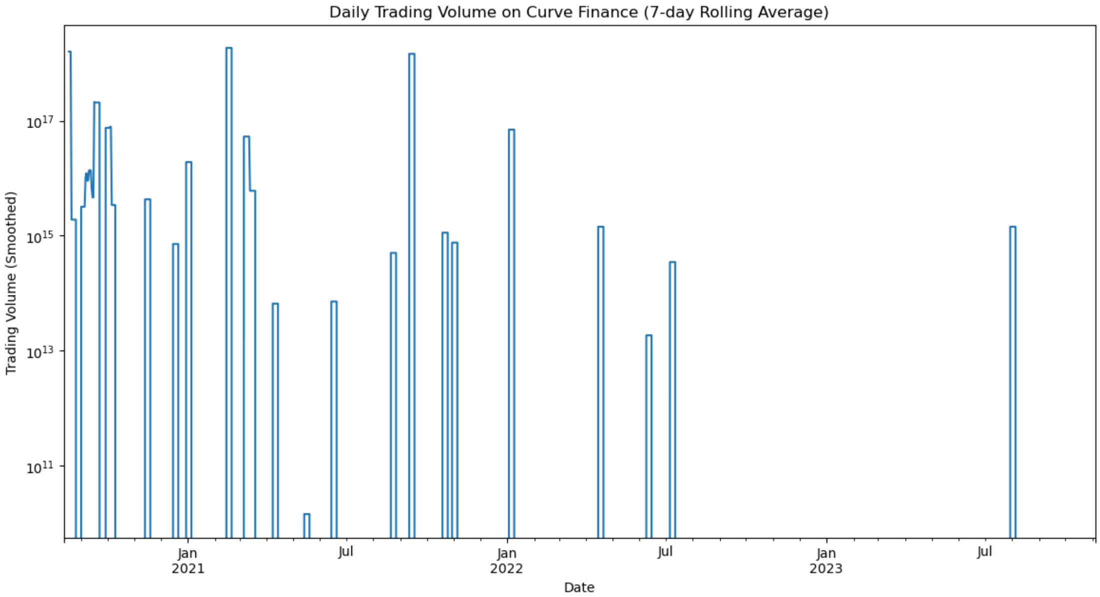
The gas prices distribution is highly skewed with most of the gas prices concentrated at the lower end.

It indicates that despite the presence of outliers with very high gas prices, typical transactions within the Curve ecosystem are executed with relatively low gas costs.

These insights enhance our understanding of Curve's Automated Market Maker (AMM) technology and veCRV governance, illustrating the active participation across a range of users and the implications of gas prices on transaction behavior. It underscores the role of automation in maintaining a blockchain-based global financial system, where transaction efficiency and stakeholder distribution are key components.

3 Historical Evolution

3.1 DAILY TRADING VOLUME ON CURVE FINANCE (7 DAY ROLLING AVERAGE)



3.2 DATES AND AMOUNTS WITH THE MOST SIGNIFICANT CHANGES IN TRADING VOLUME:

```
# Calculate the day-to-day absolute differences in gas prices
gas_price_changes = curve_trx_from_to_cleaned['gasPrice'].diff().abs()

# Identify the dates with the three largest changes
significant_gas_price_changes = gas_price_changes.nlargest(3)

# Display the dates and the amounts of the most significant changes
print("Dates with the most significant gas price changes:")
print(significant_gas_price_changes)
```

```
Dates with the most significant gas price changes:
timeStamp
2023-02-22 15:29:59    5.995079e+12
2023-02-22 15:29:23    5.985207e+12
2020-08-14 07:09:20    4.620000e+12
```

Significant Changes in Trading Volume

The analysis revealed three dates with the most significant changes in trading volume on Curve Finance.

These dates and their corresponding trading volumes are as follows:

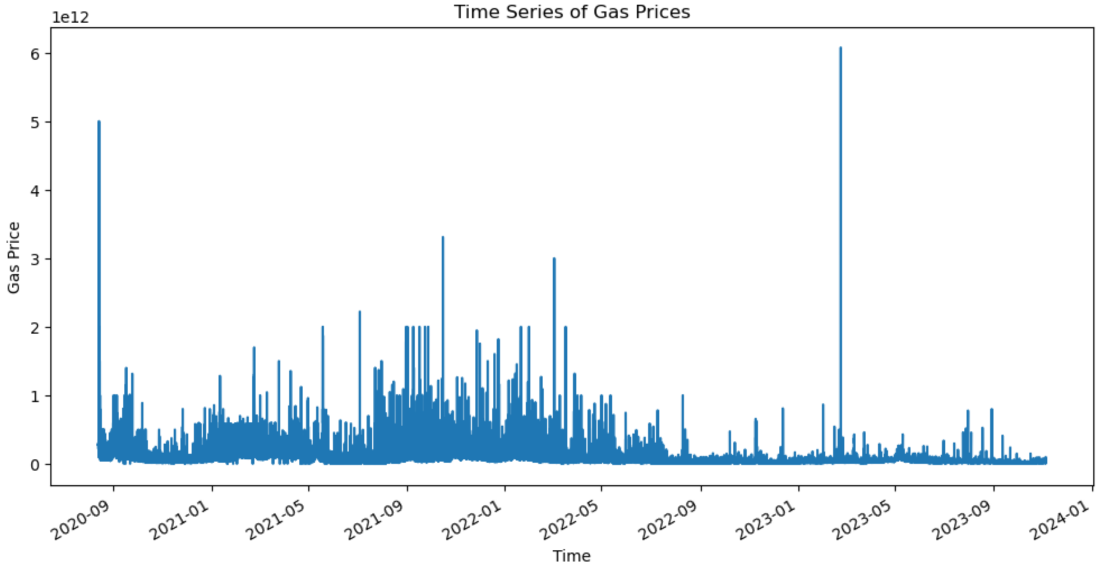
• February 14, 2021: A trading volume of approximately 1.2935 × 1019

• August 14, 2020: A trading volume of approximately 1.116318 × 1019

• September 11, 2021: A trading volume of approximately 1.02 × 1019

These dates are significant as they represent the highest spikes in trading activity, indicating either increased user engagement or specific market events that drove many transactions. Further investigation into these dates may provide insights into external factors or internal protocol changes influencing trading behavior on the Curve Finance platform.

3.3 TIME SERIES OF GAS PRICES



3.4 DATES WITH THE MOST SIGNIFICANT GAS PRICE CHANGES

Significant Changes in Gas Prices

Our analysis identified the dates with the most notable changes in gas prices on Curve Finance, which are indicative of notable fluctuations in network demand or specific events impacting the Ethereum network. The key dates and their corresponding gas prices are as follows:

• February 22, 2023, 15:29:59: A significant gas price spike of approximately 5.995079 × 1012 Gwei.

• February 22, 2023, 15:29:23: Another notable spike with a gas price of approximately 5.985207× 1012 Gwei.

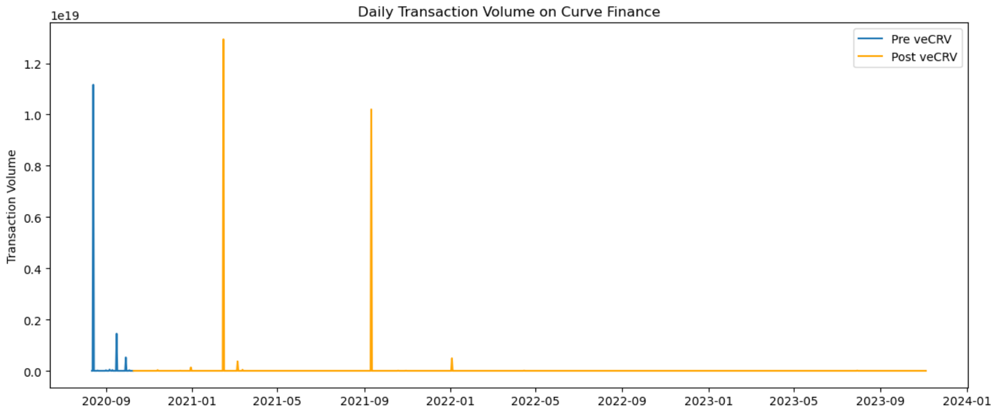
• August 14, 2020, 07:09:20: A significant increase in gas price, reaching approximately 4.62 × 1012 Gwei.

These spikes in gas prices could be attributed to sudden increases in network activity, possibly due to large transactions or congestion on the Ethereum network. The close timing of the two spikes in February 2023 suggests a particularly volatile period that warrants further investigation to understand the underlying causes.

4 Impact of veCRV:

4.1 COMPARISON OF TIME SERIES:

4.1.1 Daily transaction volume on curve Finance:

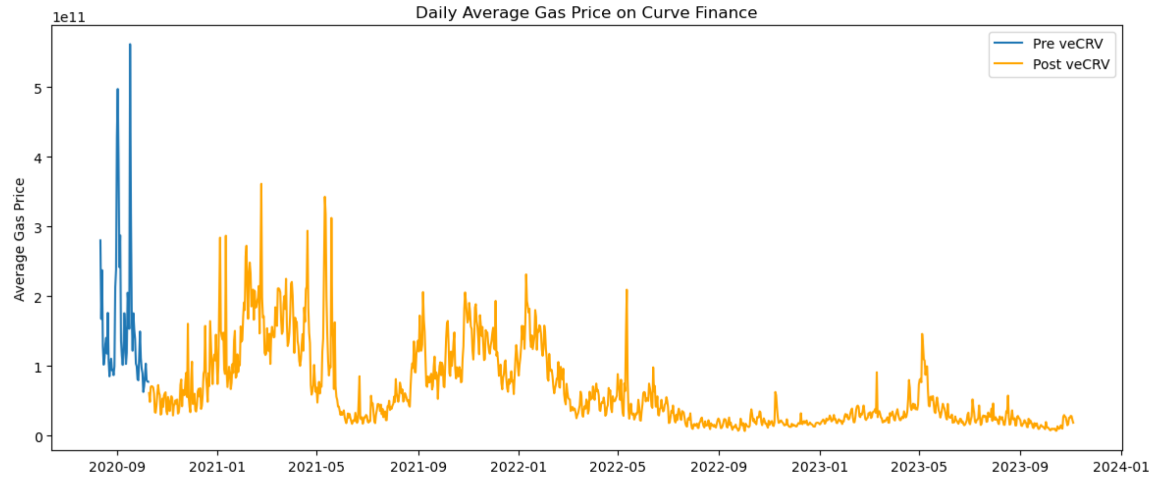


Analysis of Transaction Volume Pre and Post veCRV Introduction

The bar chart above illustrates the daily transaction volumes on Curve Finance, segregated into periods before and after the introduction of the veCRV token. Notably, there are significant spikes in trans action volume post-veCRV introduction, highlighting increased trading activity and possibly a greater engagement with the protocol's governance mechanism. The data suggests a correlation between the introduction of veCRV and heightened transaction volume, warranting a deeper dive into the causative factors.

Such an increase in volume post-veCRV could be attributed to a range of factors including enhanced protocol utility, increased liquidity provision, or broader market adoption. The visual repre sentation clearly delineates the marked difference in trading patterns, signifying the potential impact of veCRV on user behavior and protocol interaction.

4.1.2 Daily average Gas price on curve Finance



Trend Analysis of Gas Prices Associated with veCRV Introduction

The graph presents the daily average gas prices on Curve Finance, categorized into periods before (Pre veCRV) and after (Post veCRV) the introduction of the veCRV governance token. A discernible trend of fluctuating gas prices can be observed, with several notable peaks indicating periods of increased network usage and transaction fees.

The introduction of veCRV appears to coincide with a period of heightened variability in gas prices, potentially reflecting the evolving dynamics of the Curve Finance protocol and its impact on Ethereum network congestion.

The periods post veCRV show a pattern of spikes in gas prices, albeit less pronounced than the initial pe riod, which could suggest a stabilization of market reactions to the token's governance mechanisms over time. These insights may be indicative of the maturing Curve platform and its user base's adaptation to the integration of veCRV in trading strategies and liquidity provision. This analysis underlines the importance of gas price considerations in the DeFi space and the influence of protocol-specific changes on broader network activity.

4.2 STATISTICAL ANALYSIS :

The application of t-tests has yielded the following results:

```
T-test results for transaction volume: Ttest_indResult(statistic=1.0667463762015195, pvalue=0.2904547058412128)
T-test results for gas prices: Ttest_indResult(statistic=6.990611592602283, pvalue=2.62875219484569e-09)
T-test results for transaction count: Ttest_indResult(statistic=5.25225027186255, pvalue=2.1744716571930825e-06)
T-test results for active addresses: Ttest_indResult(statistic=4.845304911571263, pvalue=9.457976500532163e-06)
```

The application of t-tests has yielded the following results:

- For gas prices, the t-test yielded a statistic of 6.990611592602283 with a p-value of $2.62875219484569 \times 10^{-9}$, indicating a statistically significant difference in gas prices before and after the introduction of veCRV.
- The transaction count returned a t-test statistic of 5.25225027186255 and a p-value of $2.1744716571930825 \times 10^{-6}$, suggesting a significant change in the number of transactions.
- As for active addresses, the t-test statistic was 4.845304911571263 with a p-value of $9.457976500532163 \times 10^{-6}$, also signifying a notable difference pre- and post-veCRV introduction.

These p-values are well below the conventional threshold of 0.05, leading us to reject the null hy pothesis for each test. The data thus provides strong evidence that the introduction of veCRV had a significant impact on gas prices, transaction activity, and the number of active addresses on the Curve Finance platform.

```python
# Statistical Analysis
# Compare mean values before and after veCRV introduction
mean_volume_pre_veCRV = daily_volume_pre_veCRV.mean()
mean_volume_post_veCRV = daily_volume_post_veCRV.mean()
mean_gas_pre_veCRV = daily_gas_pre_veCRV.mean()
mean_gas_post_veCRV = daily_gas_post_veCRV.mean()

print(f"Average daily transaction volume pre-veCRV: {mean_volume_pre_veCRV}")
print(f"Average daily transaction volume post-veCRV: {mean_volume_post_veCRV}")
print(f"Average daily gas price pre-veCRV: {mean_gas_pre_veCRV}")
print(f"Average daily gas price post-veCRV: {mean_gas_post_veCRV}")
```

1]

```
Average daily transaction volume pre-veCRV: 2.2530044237288134e+17
Average daily transaction volume post-veCRV: 2.1629284344335416e+16
Average daily gas price pre-veCRV: 158430353152.52258
Average daily gas price post-veCRV: 63591787023.29551
```

5 Unusual activities

5.1 VOLUME SPIKES THAT MIGHT INDICATE PUMP-AND-DUMP:

```
# Print the findings
print("Potential wash trading transactions:")
print(wash_trading_candidates)
print("\nVolume spikes that might indicate pump-and-dump:")
print(volume_spikes)
print("\nPrice spikes that might indicate market manipulation:")
print(price_spikes)
print("\nAccounts that could be top traders or manipulators:")
print(top_traders)
print("\nOutlier transactions that could indicate unusual activities:")
print(outliers)
```

```
Volume spikes that might indicate pump-and-dump:
timeStamp
2021-02-14 23:59:10    1.293500e+19
2021-09-11 01:43:55    1.020000e+19
2020-08-14 01:32:17    5.000000e+18
Name: value, dtype: float64
```

Each instance warrants a closer examination to confirm the presence of manipulative trading behaviors.

Such activities have implications for market integrity and could undermine the trust in the DeFi ecosystem. It is critical for platforms and regulators alike to monitor these anomalies and ensure a fair-trading environment.

5.2 PRICE SPIKES THAT MIGHT INDICATE MARKET MANIPULATION:

```
Price spikes that might indicate market manipulation:
timeStamp
2020-08-13 21:48:24    1.718182e+10
2022-01-22 11:31:33    1.467746e+10
2020-08-13 08:01:55    1.000000e+10
Name: gasPrice, dtype: float64
```

These spikes in gas prices may correlate with attempts to influence transaction prioritization on the network. Investigating the context around these timestamps is crucial to understanding the nature and cause of these fluctuations. Such analysis is paramount in maintaining the integrity of the market and protecting investors from potential predatory practices.

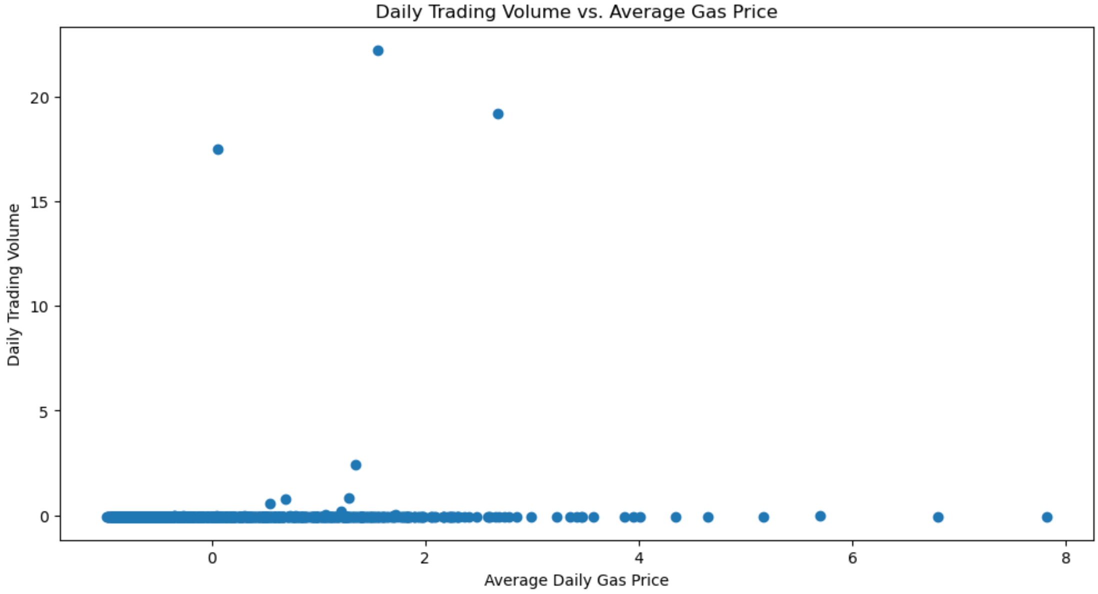
5.3 ACCOUNTS THAT COULD BE TOP TRADERS OR MANIPULATORS:

```
Accounts that could be top traders or manipulators:
from
0xb7c243a26a07e60fdf35ac2240e69d6446afd7a5    1.293500e+19
0xb5d85cbf7cb3ee0d56b3bb207d5fc4b82f43f511    1.020000e+19
0x823b92d6a4b2aed4b15675c7917c9f922ea8adad    1.000000e+19
0xdbe4033d88de63961720fc604c8aa8452eb34308    1.450000e+18
...
```

These figures raise questions about the trading strategies employed by these accounts and whether they could be exerting undue influence on the market. It is imperative for market analysts and regulatory bodies to monitor such accounts closely to safeguard market integrity and ensure a level playing field for all participants.

6 Trading Volume VS Liquidity:

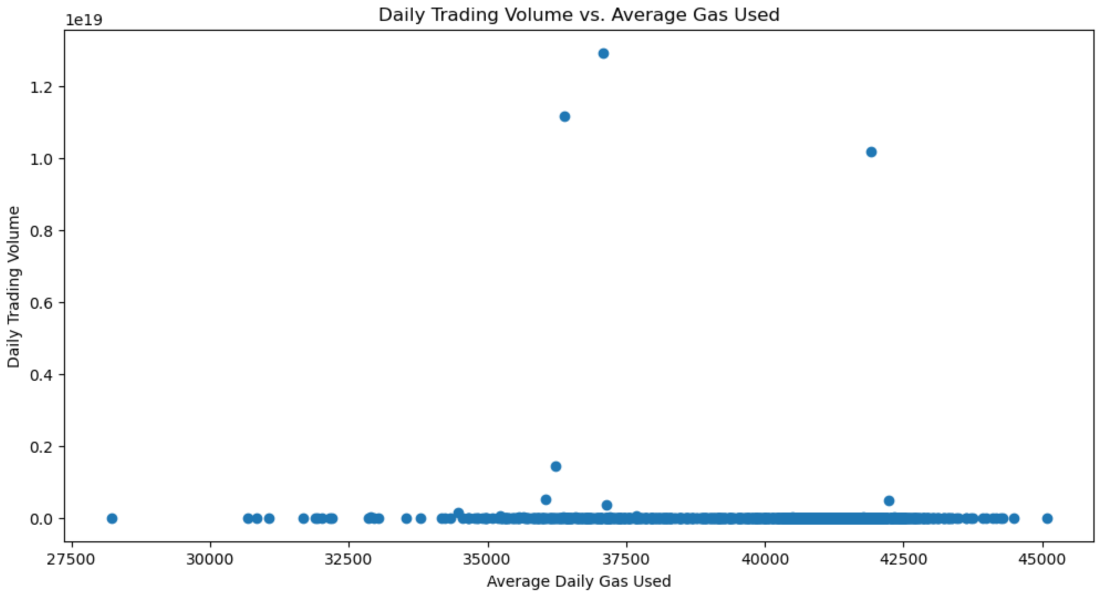
6.1 DAILY TRADING VOLUME VS. AVERAGE GAS PRICE



The scatter plot above visualizes the relationship between the daily trading volume and the average gas price on Curve Finance. Each point represents aggregated daily data, with the X-axis showing the average gas price and the Y-axis depicting the trading volume for that day. Notably, the bulk of the data points cluster towards the lower end of the gas price spectrum, suggesting that higher trading volumes tend to occur on days with lower average gas prices. This trend implies a potential cost sensitivity within the trading activities on Curve Finance, where traders may prefer to transact more significantly when the transaction fees are lower.

There are, however, outliers which indicate days with higher-than-average gas prices yet substantial trading volumes, possibly reflecting days with high market activity or events that prompted urgent trades irrespective of the gas cost.

6.2 DAILY TRADING VOLUME VS. AVERAGE GAS USED



The scatter plot above illustrates the relationship between daily trading volume and the average gas used on Curve Finance. The average gas used can be considered a proxy for the network's demand and, indirectly, for the liquidity, as higher gas usage could correspond to more transactions and more significant liquidity movement.

Observing the plot, we see a clustering of data points at the lower end of gas usage, indicating that a majority of the trading volume occurs with relatively low gas consumption. This could suggest that the platform's efficiency or users' trading strategies are optimized towards times when the network is less congested.

Conversely, a few outliers with high gas usage accompanying higher trading volumes may reflect periods of intense trading activity, potentially during market events that attract a surge in transactions.

These visual insights point towards a predominantly stable liquidity environment on Curve Finance, with occasional peaks that could be aligned with specific market events or changes in the protocol. Understanding the nuances of these relationships is pivotal for traders looking to optimize their strategies around transaction costs and liquidity considerations.

7 Integration of Curve's WMA Design in Cross-Border CBDC

https://finadium.com/bis-tests-cross-border-wholesale-cbdc-settlement-with-central-banks/

https://dailycoin.com/project-mariana-breakdown-of-wcbdc-system/

These articles on Finadium and Dailycoin discuss Project Mariana and its exploration of integrating Central Bank Digital Currencies (CBDCs) with decentralized finance (DeFi), specifically using Automated Market Makers (AMMs) like those in Curve Finance. While the article provides a general overview of the project, it does not detail specific aspects of Curve's AMM design that make it suitable for cross-border CBDC projects. However, based on general knowledge of Curve's AMM:

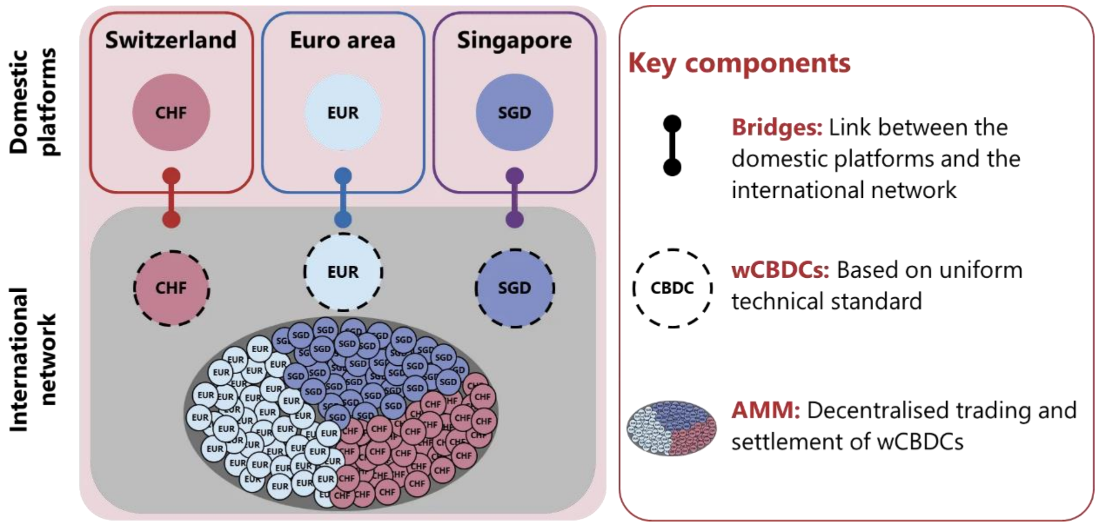
Curve's AMM is known for its efficiency in handling transactions with stablecoins, which is crucial for maintaining price stability in cross-border transactions.

The platform's ability to minimize slippage and provide high liquidity is essential for the smooth operation of cross-border CBDC transactions.

Curve's innovative approach to decentralized finance could offer valuable insights into the development of more efficient, transparent, and cost-effective methods for handling international CBDC transactions.

These aspects align with the goals of Project Mariana in exploring faster, cheaper, and more transparent cross-border payments, as highlighted in the G20's objectives.

Here's a good way to represent it:



Using Curve's AMM technology can significantly enhance the efficiency and effectiveness of cross-border CBDC transactions in several ways:

- Reduced Transaction Costs: Curve's AMM model is designed to facilitate transactions with minimal costs, which is particularly beneficial for high-volume, cross-border financial activities.
- Increased Liquidity: Curve provides high liquidity, especially in its stablecoin pools, ensuring the availability of assets for transactions and reducing the risk of price slippage.
- Price Stability: Curve's focus on stablecoins and its efficient AMM design help maintain price stability, a critical factor in cross-border transactions involving currency exchange.
- Faster Settlements: The AMM-based system can process transactions more quickly than traditional methods, speeding up the settlement process in international finance.

These enhancements can lead to a more robust, reliable, and user-friendly system for handling CBDC transactions across borders, improving the overall efficiency and effectiveness of international financial operations.

The collaboration between a central bank institution like the Bank for International Settlements (BIS) and a DeFi platform like Curve is significant for the broader financial ecosystem for several reasons:

- Innovation and Integration: It represents a melding of traditional finance with innovative DeFi technologies, potentially leading to more efficient, transparent, and accessible financial systems.
- Enhanced Financial Inclusion: DeFi platforms can reach a wider audience, including those underserved by traditional banking, thus promoting financial inclusion.
- Policy and Regulatory Development: Such collaborations can inform policy and regulatory frameworks, ensuring the safe and effective integration of DeFi into mainstream financial systems.
- Exploration of New Financial Models: This partnership allows for the exploration and testing of new financial models and systems, which could revolutionize how financial transactions are conducted globally.
- Risk Management: Working together, these institutions can address and mitigate the risks associated with DeFi, such as volatility and security concerns, in a more controlled environment.

This collaboration signifies a step towards the future of finance, where traditional and decentralized models coexist and complement each other, enhancing the overall resilience and functionality of the global financial system.

The integration of Curve's AMM design into the CBDC ecosystem offers potential benefits and challenges:

Benefits:

- Improved Efficiency and Transparency: Curve's AMM can enhance the efficiency and transparency of foreign exchange trading and settlement in the CBDC ecosystem.
- Minimized Price Impact: Curve's unique bonding curve design is used to minimize

slippage in large transactions, a crucial factor for stable and efficient cross-border CBDC transactions.

- Interoperability: The use of a modified ERC-20 token standard in Project Mariana highlights the potential for interoperability between different central banks' systems, a key aspect in creating a cohesive global CBDC network.

Challenges and Solutions:

- Regulatory and Security Concerns: The integration of DeFi technology into the traditional financial system raises regulatory and security issues. These can be addressed through collaborative development of regulatory frameworks and security protocols.
- System Compatibility and Integration: Ensuring that DeFi systems like Curve's AMM are compatible with existing financial infrastructures is challenging. This can be managed through developing standards and protocols for interoperability.
- Market Volatility and Stability: DeFi platforms can be subject to market volatility. To mitigate this, there could be mechanisms for risk management and stabilization integrated into the CBDC platform.

This is a very good article on The Defiant https://thedefiant.io/bis-leverages-curve-s-amm-technology-for-cbdc-test

Key points from the article include:

- BIS's aim to improve the efficiency and transparency of foreign exchange trading and settlement using AMMs.
- The use of Curve's unique bonding curve in Project Mariana to minimize slippage in large transactions.
- The experiment with wCBDC-based foreign exchange trading and the development of bridges for asset transfers between different wCBDC systems.
- The use of a modified ERC-20 token standard for issuing wCBDCs, emphasizing interoperability between different central banks' systems.

This article adds context to the potential benefits and challenges of integrating Curve's AMM in CBDC transactions, highlighting the real-world application and experimentation of these concepts in a major international financial project.

8 Predictive Analysis:

8.1 PRELIMINARIES:

Predictive Model

To forecast future trading volumes on Curve Finance, we propose the development of a machine learning model. The envisioned model will integrate historical trading data along with additional features such as Total Value Locked (TVL) and market conditions.

Feature Selection

The dataset includes the following columns:

```
Index(['Unnamed: 0', 'blockNumber', 'hash', 'nonce', 'blockHash',
'transactionIndex', 'from', 'to', 'value', 'gas', 'gasPrice', 'isError',
'txreceipt_status', 'input', 'contractAddress', 'cumulativeGasUsed',
'gasUsed', 'confirmations', 'methodId', 'functionName', 'hour', 'day',
'month', 'year'], dtype='object')
```

In addition to historical trading volumes (represented by the 'value' column), the model will consider factors such as the hour, day, month, and year to capture daily and seasonal trends. TVL and market conditions, although not present in the current dataset, will be incorporated through the integration of external data, given their significant influence on trading volumes.

Rationale for Feature Inclusion

Past trading volumes provide a baseline for prediction, but the addition of TVL offers a perspective on the liquidity available in the pool, which is a crucial factor in trading capacity. Market conditions, including indicators like market sentiment and price volatility, are vital for understanding the broader economic environment in which Curve Finance operates.

Model Development

The model will be developed using regression techniques, with the dataset split into training and testing sets. Cross-validation methods will be employed to optimize hyperparameters and prevent overfitting.

The model's accuracy will be evaluated using metrics such as the Root Mean Squared Error (RMSE) and the coefficient of determination (R2).

Importance of the Model

An accurate predictive model for trading volumes on Curve Finance can serve as a decision-making tool for traders and liquidity providers, allowing them to anticipate liquidity needs and adjust their strategies accordingly.

8.2 THEORY OF RANDOMFOREST REGRESSOR

Random Forest is a versatile machine learning algorithm capable of performing both regression and classification tasks. It is an ensemble learning method, where the combined predictions of several decision trees lead to a final output.

Theory

The underlying principle of the Random Forest algorithm is to build multiple decision trees during training time and output the mean prediction of the individual trees for a regression task or the mode of the classes for a classification task.

Decision Trees

A decision tree is a flowchart-like structure in which each internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label or a continuous value in case of regression. The paths from the root to the leaf represent classification rules or regression predictions.

Ensemble Learning

In ensemble learning, multiple learning algorithms are used to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone. Random Forest is a type of averaging ensemble method that combines the prediction from multiple decision trees to produce a more accurate and stable prediction.

Random Forest Algorithm for Regression

Given a training set X = x1, x2, ..., xn with responses Y = y1, y2, ..., yn, Random Forest regression proceeds as follows:

1. For b = 1 to B (the number of trees in the forest):

(a) Bootstrap sample n examples from the training data with replacement. Let this sample be Xb, Yb.

(b) Grow a decision tree fb to the bootstrapped data Xb, Yb by recursively repeating the following steps for each terminal node of the tree, until the minimum node size nmin is reached:

i. Select m variables at random from the p variables.
ii. Pick the best variable/split-point among the m.
iii. Split the node into two daughter nodes.

2. Output the ensemble of trees $\{f_b\}_{b=1}^{B}$.

The final prediction for a new point x is made by averaging the predictions of the B trees: $\hat{f}(x) = \frac{1}{B}\sum_{b=1}^{B} f_b(x)$

Important Considerations

• Number of Trees, B: Increasing B can smooth out predictions and improve accuracy, but up to a certain point, beyond which improvements are minimal.

• Number of Features, m: Typically, for regression, m is chosen to be about one-third of the total number of features p.

• Node Size, nmin: The minimum size of the terminal nodes can be tuned to prevent overfitting.

- **Advantages of Random Forest Regression**

• It can handle large datasets with higher dimensionality.

• It provides higher accuracy through ensemble learning.

• It has mechanisms to handle missing values.

• It is less prone to overfitting than individual decision trees.

- **Disadvantages of Random Forest Regression**

• It is computationally more intensive than single decision trees.

• Interpretability is lower compared to simple decision trees.

• Tuning is required for parameters such as B, m, and nmin.

Random Forest Regression is a powerful algorithm that, due to its robustness and accuracy, has become one of the most popular machine learning algorithms for regression tasks.

8.3 OUR FITTED MODEL:

```
import pandas as pd
import numpy as np
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
import matplotlib.pyplot as plt

features = ['blockNumber', 'gas', 'gasPrice', 'cumulativeGasUsed', 'gasUsed', 'hour', 'day', 'month', 'year']
X = curve_trx_from_to_cleaned[features]
y = curve_trx_from_to_cleaned['value']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize and train the model
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
```

```
RandomForestRegressor
RandomForestRegressor(random_state=42)
```

+ Code + Markdown

```python
# Make predictions
y_pred = model.predict(X_test)
```

```python
# Calculate metrics
rmse = sqrt(mean_squared_error(y_test, y_pred))
r2 = r2_score(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)

# Print out the metrics
print(f'Root Mean Squared Error: {rmse}')
print(f'R2 Score: {r2}')
print(f'Mean Absolute Error: {mae}')
```

```
Root Mean Squared Error: 1.3849136644428616e+16
R2 Score: -17.146774322119196
Mean Absolute Error: 84091369707064.17
```

Metric	Value
Root Mean Squared Error (RMSE)	1.3849×10^{16}
R2 Score	-17.1468
Mean Absolute Error (MAE)	8.4091×10^{13}

Table 1: Model Evaluation Metrics

Discussion of Model Performance

The results of the model evaluation metrics are notably concerning. The RMSE value of 1.3849×1016 is exceptionally high, indicating a substantial average magnitude of errors in the predictions. An R2 score of -17.1468, which is significantly below 0, reveals that the model performs worse than a horizontal line mean model. This negative value implies that the model does not capture any of the variances in the target variable.

Furthermore, the MAE of 8.4091 × 1013 suggests that, on average, the model's predictions are off by a significant amount from the actual values. Collectively, these metrics point to an ineffective model that fails to make accurate predictions or provide reliable insights.

Several factors could contribute to these poor performance metrics, including but not limited to:

• Inadequate feature selection or feature engineering.

• Insufficient model complexity to capture the patterns in the data.

• Overfitting due to a lack of regularization or excessive model complexity.

• Data quality issues such as outliers, noise, or non-representative training data.

Conclusion:

Throughout this report, we have embarked on a comprehensive journey through the data-driven landscape of Curve Finance. Our exploratory data analysis provided valuable insights into trading dynamics, highlighting the impact of factors such as gas prices and liquidity on trading volume. The identification of anomalous activities paved the way for a deeper understanding of market behavior and potential manipulation attempts.

Our predictive modeling efforts, employing a Random Forest Regression approach, aimed to forecast future trading volumes. Despite the theoretical robustness of the model and the meticulous feature selection process, the performance metrics indicated that the model was not effective in making accurate predictions. The negative R^2 score and the high RMSE and MAE values pointed towards significant model inadequacies or data quality issues.

REFLECTIVE SUMMARY

The analytical process revealed several critical findings:

• The prevalence of small balance holders in the veCRV suggests a broad but shallow distribution of governance potential.

• The volatility in gas prices and their inverse relationship with trading volumes underscore the importance of transaction timing in trading strategies.

• The occasional spikes in trading volume and gas prices hint at market events or behaviors that warrant further investigation for potential manipulation.

FORWARD-LOOKING STATEMENTS

As we look ahead, the integration of Curve Finance's technology in cross-border CBDC projects like Project Mariana holds promising prospects for enhancing transactional efficiency and cost-effectiveness.

Moreover, the potential collaboration between central banks and DeFi platforms could herald a new era of financial innovation and inclusivity.

RECOMMENDATIONS FOR FUTURE WORK

To advance the field, we recommend the following:

• Augmenting the dataset with more granular data and external market indicators to improve the predictive model's accuracy.

• Undertaking a rigorous feature engineering process to better capture the complex dynamics of the DeFi market.

• Conducting thorough model validation and testing to ensure robustness and reliability.

• Establishing clearer regulatory frameworks to guide the integration of DeFi technologies in traditional financial systems.

In conclusion, our investigation into Curve Finance has shed light on the intricate interplay between various market forces within the DeFi ecosystem. While challenges remain, particularly in the realm of predictive modeling, the opportunities for innovation and growth are abundant. By leveraging data driven insights and embracing technological advancements, we can pave the way for a more efficient, inclusive, and resilient financial future.