Economics Design

Data Analysis HW

Organising of Wallet Addresses

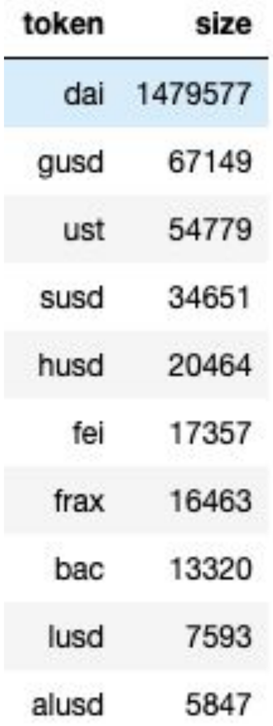
Executive Summary

- **Tasks and findings**
  - Qualify wallet address
    - More human address (EOA) than smart contracts (1,440,022 vs 200,101)
    - A total of 129 CEX related addresses
    - K-Means shrunk the search space and found various groups of users from VCs/Funds, exchanges to market markers
  - Understand the network of these wallets
    - CEX are likely to work with stable coins with larger market cap and are fiat collateralised
    - Stablecoins listed on CEX are likely to have higher degree of centrality as entering DeFi is easy
    - CEX don’t work with a range of stable coins but DEXes do
- **Methodology**
  - Used ZMOK (Alternative to Infura to circumvent API limits) to classify address types and retrieve account balance
  - Used Etherscan to retrieve token contract, exchange addresses, and possible DEX smart contracts

Files and data overview

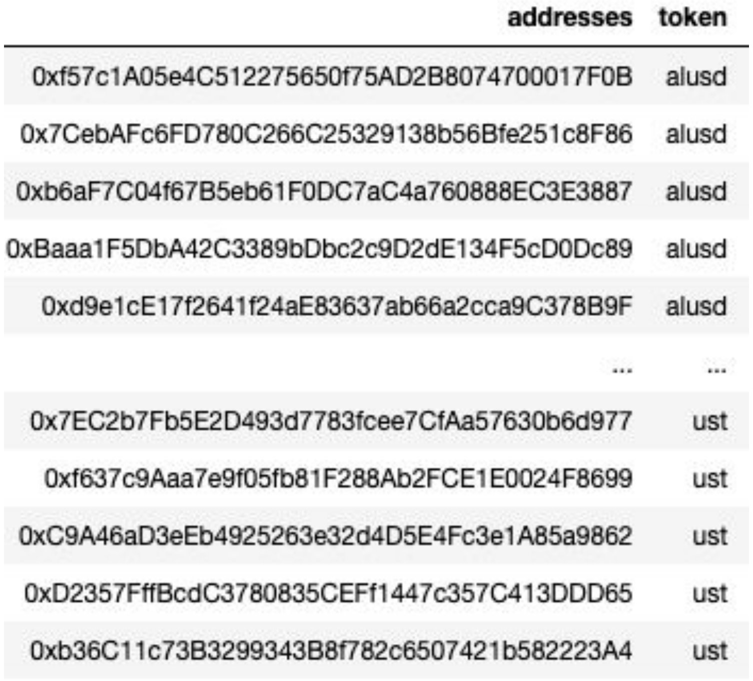
Properties of file received are

- 10 Stablecoins: ALUSD, BAC, DAI, FEI, FRAX, GUSD, HUSD, LUSD, SUSD,UST
- Each file contains unique addresses. Not unique across files.
  - ○ Total addresses: 1,717,200
  - ○ Unique addresses: 1,640,123
- No information on:
  - ○ Balance, transaction history, transaction count, type of address (to find out)
- DAI had the most unique addresses while ALUSD had the least



token	size
dai	1479577
gusd	67149
ust	54779
susd	34651
husd	20464
fei	17357
frax	16463
bac	13320
lusd	7593
alusd	5847

No. of unique addresses in each stablecoin file



addresses	token
0xf57c1A05e4C512275650f75AD2B8074700017F0B	alusd
0x7CebAFc6FD780C266C25329138b56Bfe251c8F86	alusd
0xb6aF7C04f67B5eb61F0DC7aC4a760888EC3E3887	alusd
0xBaaa1F5DbA42C3389bDbc2c9D2dE134F5cD0Dc89	alusd
0xd9e1cE17f2641f24aE83637ab66a2cca9C378B9F	alusd
...	...
0x7EC2b7Fb5E2D493d7783fcee7CfAa57630b6d977	ust
0xf637c9Aaa7e9f05fb81F288Ab2FCE1E0024F8699	ust
0xC9A46aD3eEb4925263e32d4D5E4Fc3e1A85a9862	ust
0xD2357FffBcdC3780835CEFf1447c357C413DDD65	ust
0xb36C11c73B3299343B8f782c6507421b582223A4	ust

Combining all addresses across files for analysis

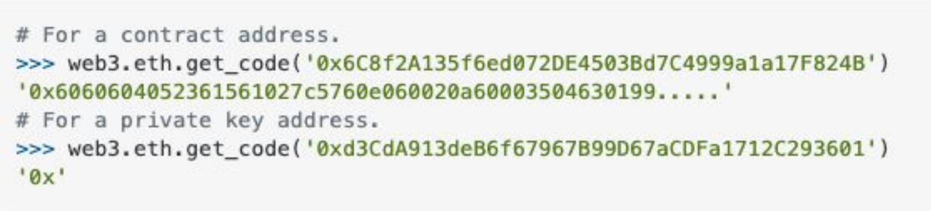
Assumptions and Unknowns

- **Assumptions**
  - Weak time ordering, addresses may be sequential in a log but lack of transaction data means the exact time and order of transactions is not known
    - Therefore, time order is ignored
- **Unknowns**
  - Do not know when the addresses are retrieved. Assumed to be 28th Feb (possible time of file generation)
- **Constraints**
  - Work mainly within the data in the CSV
  - $0 budget (can’t get richer data)
  - Time constrained and hence would have to work fast (can’t wait for API rate limits)
  - Computing power.
    - 1.6mil addresses had crashed my computer a few times. This makes working with transactional data or network related analysis difficult

Characterization of wallet types

Wallet address (EOA) or contract?

- Assumed that “human address’ is an address with a private key. This makes the address known as an *Externally Owned Address* (EOA).
- Smart contracts have no private keys and contains code
- We can test an address by running *web3.eth.get_code().* If there is code it is likely a smart contract, if not it is likely an EOA
- **Currently not robust due to Zmok**



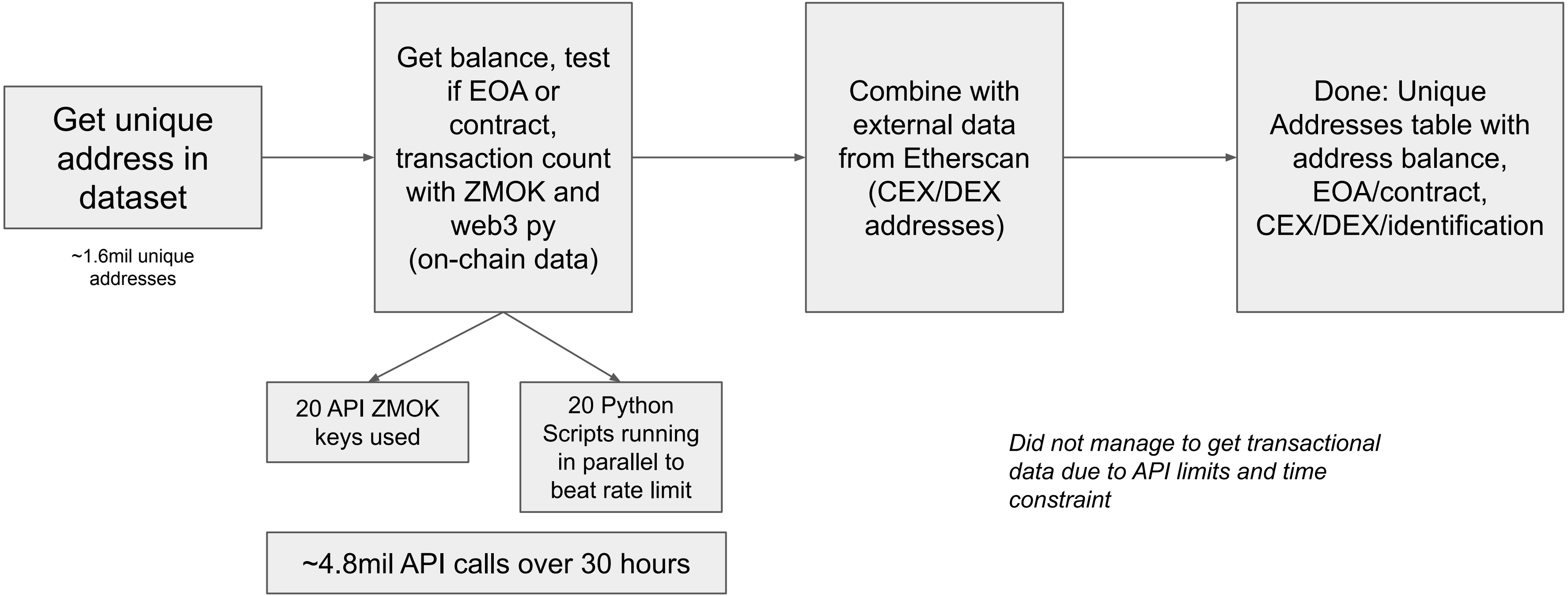
```
# For a contract address.
>>> web3.eth.get_code('0x6C8f2A135f6ed072DE4503Bd7C4999a1a17F824B')
'0x6060604052361561027c5760e060020a60003504630199.....'
# For a private key address.
>>> web3.eth.get_code('0xd3CdA913deB6f67967B99D67aCDFa1712C293601')
'0x'
```

How to get EOA labels? (CEX? DEX? Funds?)

- Statistical studies by looking at more data
- Crowdsourcing (Etherscan/Nansen)
- Recommended to filter by address characteristics to reduce search space

Explore & Navigate Etherscan's Label Word Cloud

Approach to getting more data to characterize wallets



Get unique address in dataset

~1.6mil unique addresses

Get balance, test if EOA or contract, transaction count with ZMOK and web3 py (on-chain data)

20 API ZMOK keys used

20 Python Scripts running in parallel to beat rate limit

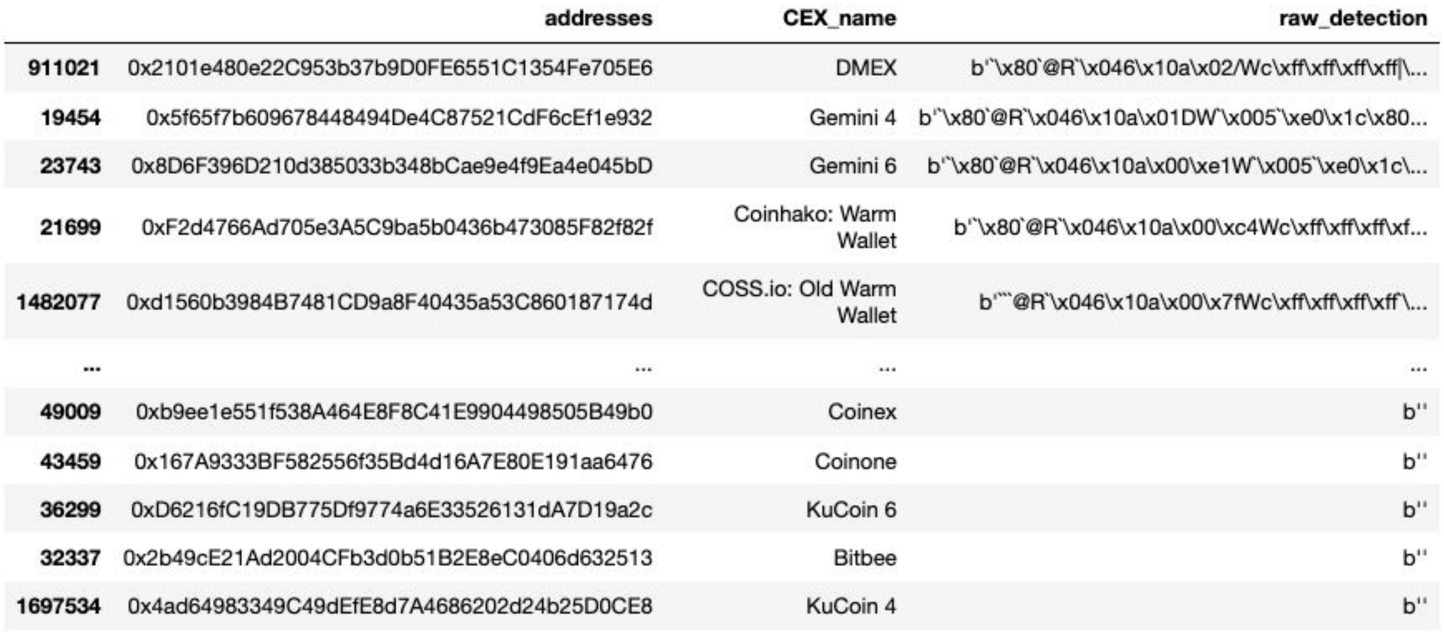
~4.8mil API calls over 30 hours

Combine with external data from Etherscan (CEX/DEX addresses)

Done: Unique Addresses table with address balance, EOA/contract, CEX/DEX/identification

Did not manage to get transactional data due to API limits and time constraint

Example: Raw return of Hexbyte to determine contract or EOA



```
                                               addresses                 CEX_name                                       raw_detection
911021   0x2101e480e22C953b37b9D0FE6551C1354Fe705E6                     DMEX    b'`\x80`@R`\x046\x10a\x02/Wc\xff\xff\xff\xff|\...
19454     0x5f65f7b609678448494De4C87521CdF6cEf1e932                 Gemini 4  b'`\x80`@R`\x046\x10a\x01DW`\x005`\xe0\x1c\x80...
23743    0x8D6F396D210d385033b348bCae9e4f9Ea4e045bD                 Gemini 6   b'`\x80`@R`\x046\x10a\x00\xe1W`\x005`\xe0\x1c\...
21699     0xF2d4766Ad705e3A5C9ba5b0436b473085F82f82f          Coinhako: Warm   b'`\x80`@R`\x046\x10a\x00\xc4Wc\xff\xff\xff\xf...
                                                                    Wallet
1482077  0xd1560b3984B7481CD9a8F40435a53C860187174d  COSS.io: Old Warm     b'```@R`\x046\x10a\x00\x7fWc\xff\xff\xff\xff`\...
                                                                    Wallet
...                                                ...                  ...                                                 ...
49009    0xb9ee1e551f538A464E8F8C41E9904498505B49b0                   Coinex                                                 b''
43459    0x167A9333BF582556f35Bd4d16A7E80E191aa6476                  Coinone                                                 b''
36299    0xD6216fC19DB775Df9774a6E33526131dA7D19a2c                KuCoin 6                                                 b''
32337    0x2b49cE21Ad2004CFb3d0b51B2E8eC0406d632513                   Bitbee                                                 b''
1697534  0x4ad64983349C49dEfE8d7A4686202d24b25D0CE8                KuCoin 4                                                 b''
```

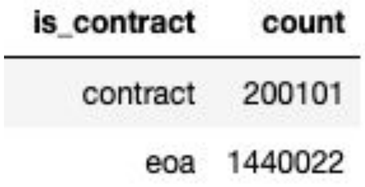
Long gibberish text are codes in hexbytes and are therefore smart contracts, ***b”*** are empty code and are therefore EOAs

A glimpse at the enhanced dataset

addresses	Is contract?	On chain balance in eth	On chain balance in usd	CEX name	CEX txn count etherscan	DEX name	DEX txn count etherscan	txn_count_on_chain
0xFBb1b73C4f0BDa4f67dcA266ce6Ef42f520fBB98	eoa	14816.9006	36227766.4	Bittrex	11273371			11040400
0x3f5CE5FBFe3E9af3971dD833D26bA9b5C936f0bE	eoa	2.37573611	5808.74606	Binance	17017319			7735967
0x3cD751E6b0078Be393132286c442345e5DC49699	eoa	5669.25399	13861496.1	Coinbase 4	8198305			7529991

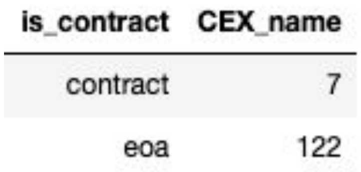
Tabulation for contract and number of CEX

Dataset contained 200,101 smart contracts and 1,440,027 EOAs



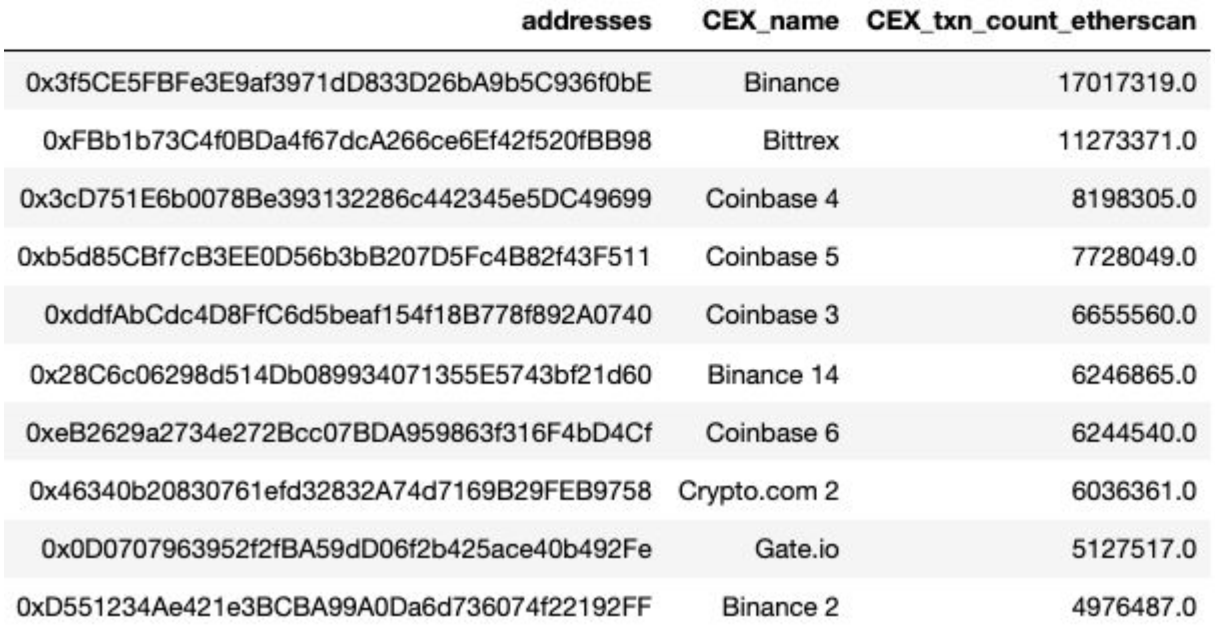
is_contract	count
contract	200101
eoa	1440022

With best efforts, there are 122 CEX addresses that are EOAs and 7 are contracts



is_contract	CEX_name
contract	7
eoa	122

Top 10 CEX by transaction count



	addresses	CEX_name	CEX_txn_count_etherscan
	0x3f5CE5FBFe3E9af3971dD833D26bA9b5C936f0bE	Binance	17017319.0
	0xFBb1b73C4f0BDa4f67dcA266ce6Ef42f520fBB98	Bittrex	11273371.0
	0x3cD751E6b0078Be393132286c442345e5DC49699	Coinbase 4	8198305.0
	0xb5d85CBf7cB3EE0D56b3bB207D5Fc4B82f43F511	Coinbase 5	7728049.0
	0xddfAbCdc4D8FfC6d5beaf154f18B778f892A0740	Coinbase 3	6655560.0
	0x28C6c06298d514Db089934071355E5743bf21d60	Binance 14	6246865.0
	0xeB2629a2734e272Bcc07BDA959863f316F4bD4Cf	Coinbase 6	6244540.0
	0x46340b20830761efd32832A74d7169B29FEB9758	Crypto.com 2	6036361.0
	0x0D0707963952f2fBA59dD06f2b425ace40b492Fe	Gate.io	5127517.0
	0xD551234Ae421e3BCBA99A0Da6d736074f22192FF	Binance 2	4976487.0

Analysis on EOA (“Human” wallets)

Performed K-means, a clustering technique to find groupings of data on its own. We would check if the groupings make sense and learn something about the data. (Similar work here [1] [2])

After some trial and error, 6 groups are found. Numbers are averaged except cluster size

Group	Balance (in ETH)	No. Txn. On Chain	Tokens Interacted	Cluster Size
1	1.8 ± 58.82	111.69±2,767.3	1	671,238
2	6.91±134.23	759.08±11,385.75	2	34,717
3	13.82±186.67	1855.60±12,523.62	3.47	10,498
4	23,404.58±11,657.22	1,490.27±2,951.86	1.42	33
5	332,647.15	6103.0	2	1
6	1,282	42,617,593	1	1

EOA (“Human” wallets)

mean±SD

Group	Balance (in ETH)	No. Txn. On Chain	Tokens Interacted	Cluster Size
1	1.8 ± 58.82	111.69±2,767.3	1	671,238
2	6.91±134.23	759.08±11,385.75	2	34,717
3	13.82±186.67	1855.60±12,523.62	3.47	10,498
4	23,404.58±11,657.22	1,490.27±2,951.86	1.42	33
5	332,647.15	6103.0	2	1
6	1,282	42,617,593	1	1

- Group 1: Highly likely to be mainly normal users with probably a few traders
- Group 2: Could be specialized traders/investors
- Group 3: Could be larger traders/investors as they trade more token
- Group 4: Likely untagged exchanges or whales due to high balance and transaction
- Group 5: Likely untagged exchanges due to high transaction and high ETH balance
- Group 6: Apparently a miner (Ethermine)

Interesting findings in Group 3 and 4

Group 3

- Three Arrows Capital
  - 0x4862733B5FdDFd35f35ea8CCf08F5045e57388B3
- Alameda
  - 0x0F4ee9631f4be0a63756515141281A3E2B293Bbe
- Analytico
  - 0xa0f75491720835b36edC92D06DDc468D201e9b73
  - A Singapore Crypto market maker
- Paul Veradittakit
  - 0x1333c53A798547126Ca04647BA925485A6FA7Aad
  - Partner of Pantera Capital

Group 4

- Blockchain bandit
  - 0x957cD4Ff9b3894FC78b5134A8DC72b032fFbC464
  - Guesses weak private keys and steals money
- Binance US 2
  - 0x34ea4138580435B5A521E460035edb19Df1938c1
  - Exchange wallet
- Patricio Worthalter
  - 0x57757e3d981446d585af0d9ae4d7df6d64647806
  - 0xb1e9D641249A2033C37CF1C241a01E717c2F6c76
  - Founder of Pixel Vault, PUNKS Comic, and MetaHero Universe.

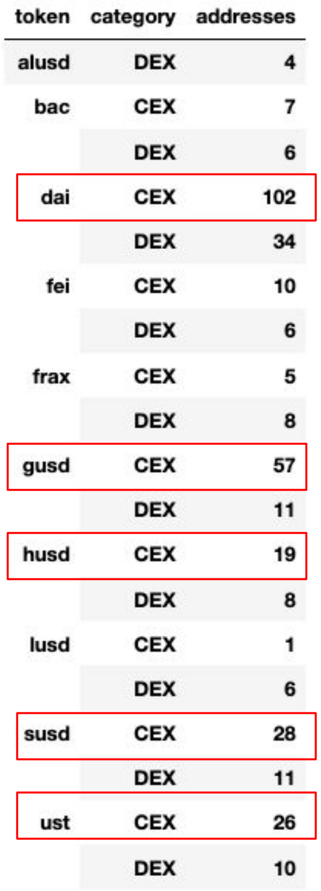
Network Analysis

A look at how addresses interact with the stablecoins

Intro: Network Analysis

- Limited analysis due to lack of transactional data and therefore unable to show movement of money and interaction
  - Data availability is the biggest constraint
    - Balances, DEX/CEX classification, no.of transactions are the only extra data available
  - Unable to show path of transaction and therefore unable to show:
    - A few whales trading between each other?
    - Real utility vs wash trading based on the various user types
- Temporal (time series) data requires significantly more data or external data that aggregated which is outside the scope of the CSV

Token interaction by CEX or DEX



token	category	addresses
alusd	DEX	4
bac	CEX	7
	DEX	6
dai	CEX	102
	DEX	34
fei	CEX	10
	DEX	6
frax	CEX	5
	DEX	8
gusd	CEX	57
	DEX	11
husd	CEX	19
	DEX	8
lusd	CEX	1
	DEX	6
susd	CEX	28
	DEX	11
ust	CEX	26
	DEX	10

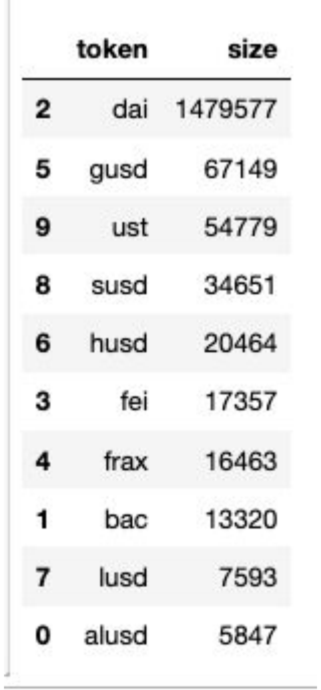
Stablecoin	Mechanism	Market Cap
UST	Algo	$16,245,500,957
DAI	Crypto-Collateral	$7,382,347,255
FRAX	Algo	$2,638,211,121
LUSD	Crypto-Collateral	$719,901,571
FEI	Algo	$419,098,934
HUST	HUSD	$389,665,812
ALUSD	Crypto-Collateral	$237,771,849
GUSD	Fiat Collateral	$198,313,044
SUSD	Crypto-Collateral	$118,065,451
BAC	Algo	$437,462

It appears that

- More CEX engage with the tokens when they are
  - High in market cap (UST, DAI)
  - Fiat Collateralized (GUSD,HUSD)
  - Allows easier onboarding from CEX users and may increase adoption on-chain
- Notable stablecoins
  - sUSD has a high adoption rate likely due to its Chainlink oracle mechanism
  - BAC has lost its peg. Had a low CEX interaction (do they know something?)

Degree Centrality

Stablecoin	Degree Centrality
DAI	0.90211
GUSD	0.04094
UST	0.0334
SUSD	0.02113
HUSD	0.01248
FEI	0.01058
FRAX	0.01004
BAC	0.00812
LUSD	0.00463
ALUSD	0.00356



```
   token     size
2    dai  1479577
5   gusd    67149
9    ust    54779
8   susd    34651
6   husd    20464
3    fei    17357
4   frax    16463
1    bac    13320
7   lusd     7593
0  alusd     5847
```

Degree Centrality if a node has a larger than average number of connections for that graph

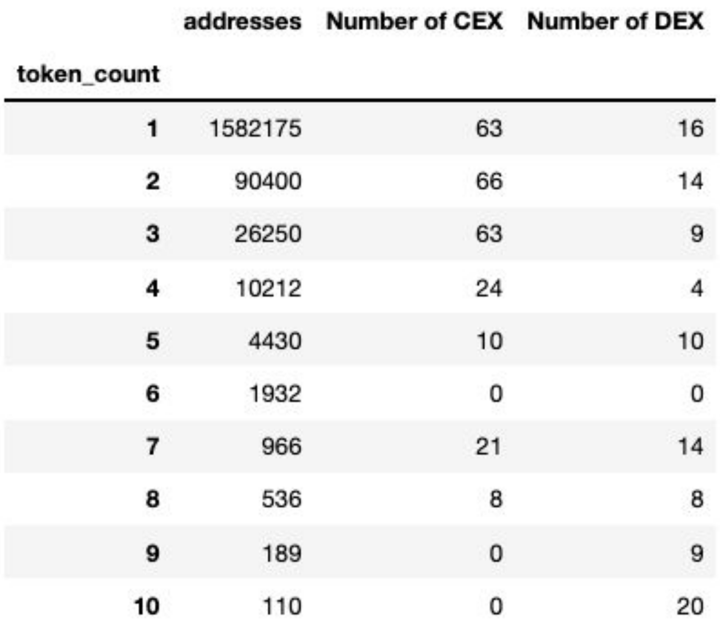
Is the same as counting the number of address transacting with a token

More connections between token and address signify higher adoption

Unable to determine if token is held by whales or adopted by many users due to lack of transactional data

Degree Centrality = Number of edges for a node / All nodes - 1

Link between token count and type of address



	addresses	Number of CEX	Number of DEX
token_count			
1	1582175	63	16
2	90400	66	14
3	26250	63	9
4	10212	24	4
5	4430	10	10
6	1932	0	0
7	966	21	14
8	536	8	8
9	189	0	9
10	110	0	20

In this particular dataset:

- The more tokens are traded per address, the number of address decreases
  - The number of CEX address decreases as well
- Conversely, the higher the amount unique tokens are traded, the more DEX related addresses are present
- May suggest that at this point of time where the data is downloaded, CEX are not processing a wide range of stablecoins
- CEX offerings are relatively limited despite stablecoins being known for its stable non speculative nature

Further Project Improvements

Address Identification

- More dimensions to look at like transaction size, time between transactions, rate of transactions within a time frame (a month?), variety of tokens, can help in better clustering and thus ability in finding new wallet categories in an automated manner
- Improve data collection methods (using Etherscan API)

Network Analysis

- A relatively computationally heavy work. Would need to scope out the task and determine the data needed early.
- Would be interesting to look at VC transaction activity or activity in DEXes
    - High activity suggests opportunity or a storm brewing…
- B2C product for consumers to check if sender is fraudulent?

Learnings

- On-chain data acquisition are expensive and time consuming
  - Spent over 30 hours running 20 scripts with 20 different API keys to get wallet balance for over 1.6 million addresses
  - At Etherscan’s highest tier at 1,000,000 API calls a day, it would cost $399/mo

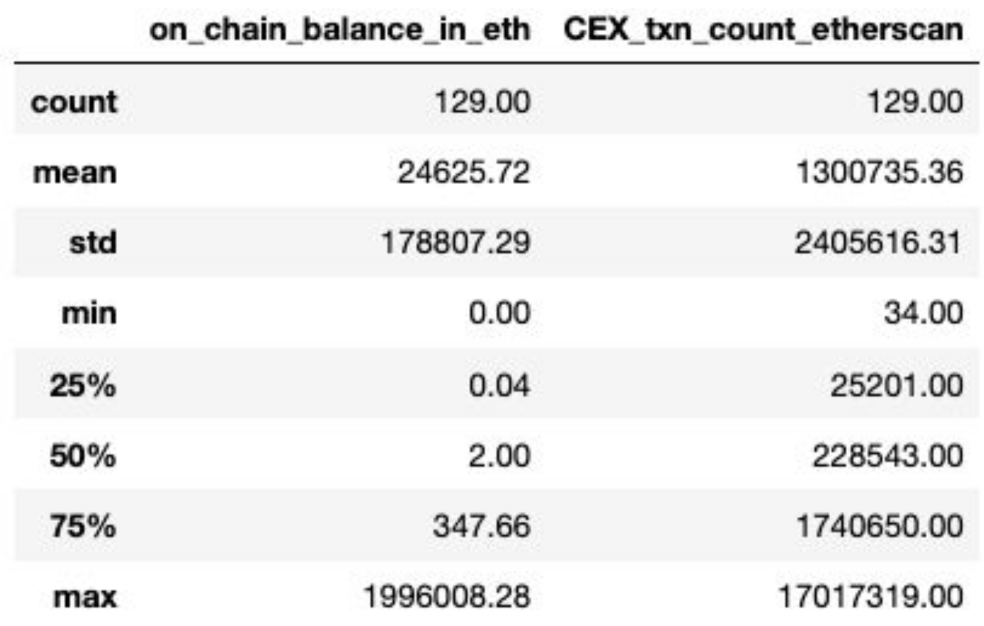
- When using new databases, test data quality early and compare against well-known databases before investing significant time. Zmok(alternative to Infura) may have sunk this project

- Keeping a private database of address labels and tracking their activity provides a competitive advantage over competitors as it is expensive, time-consuming, and challenging to curate them. However, this depends if it suitable for corporate strategy

Thank you and feedback are welcomed!

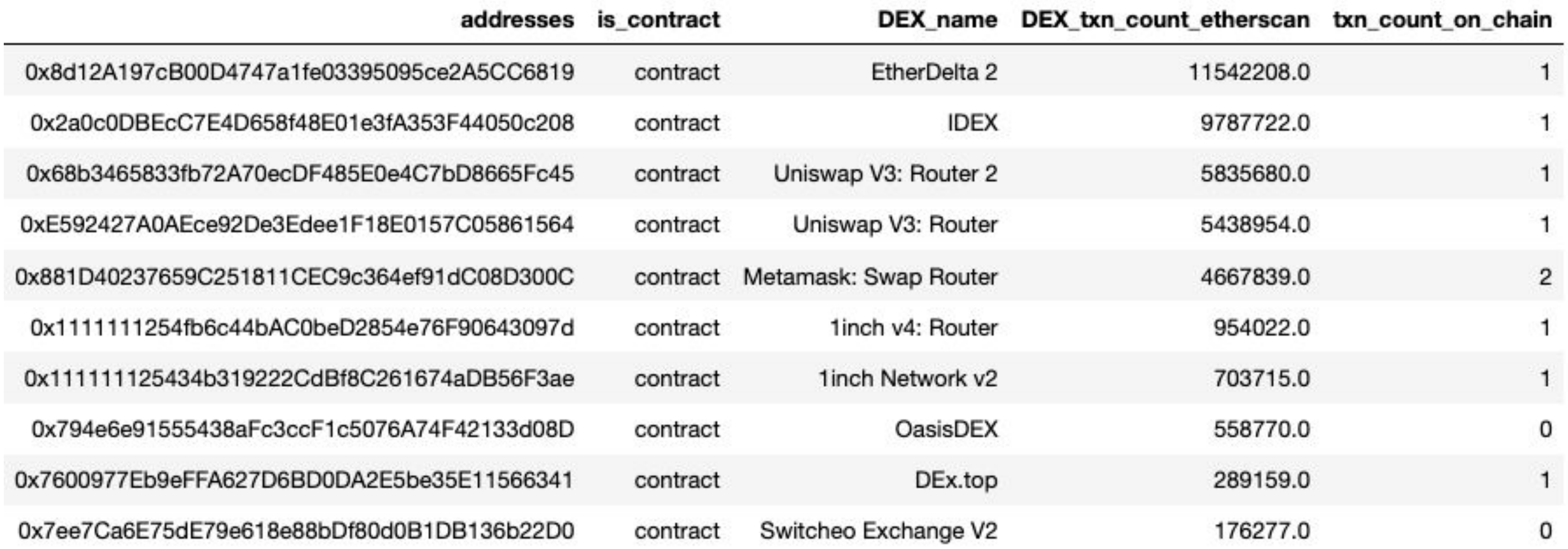
Appendix

Appendix: Properties of CEX Wallets



	on_chain_balance_in_eth	CEX_txn_count_etherscan
count	129.00	129.00
mean	24625.72	1300735.36
std	178807.29	2405616.31
min	0.00	34.00
25%	0.04	25201.00
50%	2.00	228543.00
75%	347.66	1740650.00
max	1996008.28	17017319.00

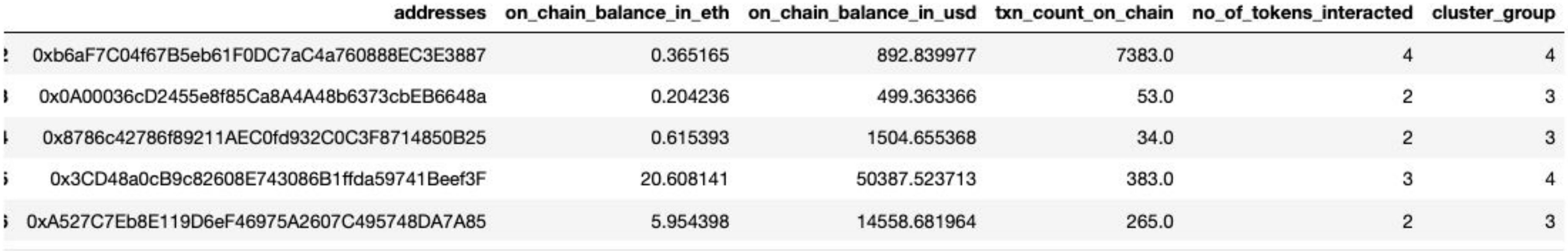
Appendix: Top 10 DEX contracts in the list



	addresses	is_contract	DEX_name	DEX_txn_count_etherscan	txn_count_on_chain
	0x8d12A197cB00D4747a1fe03395095ce2A5CC6819	contract	EtherDelta 2	11542208.0	1
	0x2a0c0DBEcC7E4D658f48E01e3fA353F44050c208	contract	IDEX	9787722.0	1
	0x68b3465833fb72A70ecDF485E0e4C7bD8665Fc45	contract	Uniswap V3: Router 2	5835680.0	1
	0xE592427A0AEce92De3Edee1F18E0157C05861564	contract	Uniswap V3: Router	5438954.0	1
	0x881D40237659C251811CEC9c364ef91dC08D300C	contract	Metamask: Swap Router	4667839.0	2
	0x1111111254fb6c44bAC0beD2854e76F90643097d	contract	1inch v4: Router	954022.0	1
	0x111111125434b319222CdBf8C261674aDB56F3ae	contract	1inch Network v2	703715.0	1
	0x794e6e91555438aFc3ccF1c5076A74F42133d08D	contract	OasisDEX	558770.0	0
	0x7600977Eb9eFFA627D6BD0DA2E5be35E11566341	contract	DEx.top	289159.0	1
	0x7ee7Ca6E75dE79e618e88bDf80d0B1DB136b22D0	contract	Switcheo Exchange V2	176277.0	0

In txn_count_on_chain shows erroneous data from ZMOK

Appendix: Raw Clustering Result



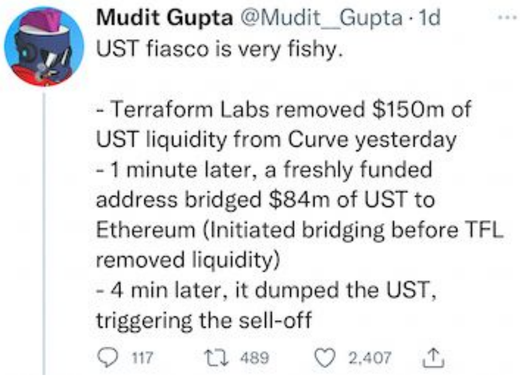
	addresses	on_chain_balance_in_eth	on_chain_balance_in_usd	txn_count_on_chain	no_of_tokens_interacted	cluster_group
2	0xb6aF7C04f67B5eb61F0DC7aC4a760888EC3E3887	0.365165	892.839977	7383.0	4	4
3	0x0A00036cD2455e8f85Ca8A4A48b6373cbEB6648a	0.204236	499.363366	53.0	2	3
4	0x8786c42786f89211AEC0fd932C0C3F8714850B25	0.615393	1504.655368	34.0	2	3
5	0x3CD48a0cB9c82608E743086B1ffda59741Beef3F	20.608141	50387.523713	383.0	3	4
6	0xA527C7Eb8E119D6eF46975A2607C495748DA7A85	5.954398	14558.681964	265.0	2	3

Appendix: Gotchas (Idiosyncrasies that disrupt operational processes)

- Etherscan’s addresses on the website are in lowercase which can result in wrong joins (wrong wallet joined)

- Discrepancies can exist
    - 0x74de5d4FCbf63E00296fd95d33236B9794016631 appears as a contract in Etherscan Web3.py (using ZMOK) classifies as EOA.
    - Therefore cross checking data quality is paramount. May need to have more than 1 data source

Appendix: Reiterating why classifying and tracking addresses are important



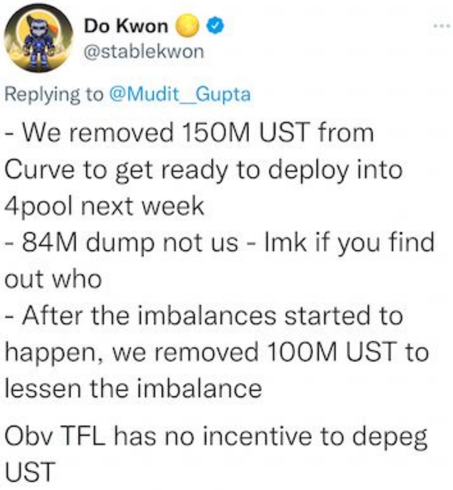
Mudit Gupta @Mudit_Gupta · 1d

UST fiasco is very fishy.

- Terraform Labs removed $150m of UST liquidity from Curve yesterday
- 1 minute later, a freshly funded address bridged $84m of UST to Ethereum (Initiated bridging before TFL removed liquidity)
- 4 min later, it dumped the UST, triggering the sell-off

117 489 2,407

CT speculating a coordinated FUD attack (although IMO not an intentional coordinated attack but just FUD in general)

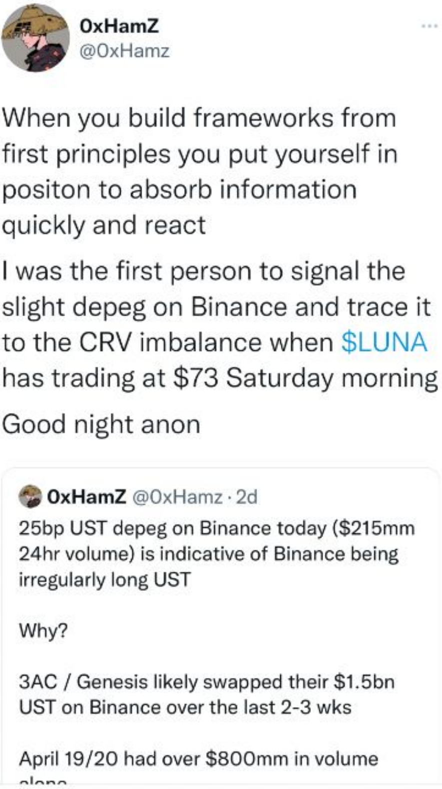


Do Kwon 🌕 @stablekwon

Replying to @Mudit_Gupta

- We removed 150M UST from Curve to get ready to deploy into 4pool next week
- 84M dump not us - lmk if you find out who
- After the imbalances started to happen, we removed 100M UST to lessen the imbalance

Obv TFL has no incentive to depeg UST



0xHamZ @0xHamz

When you build frameworks from first principles you put yourself in positon to absorb information quickly and react

I was the first person to signal the slight depeg on Binance and trace it to the CRV imbalance when $LUNA has trading at $73 Saturday morning

Good night anon

0xHamZ @0xHamz · 2d

25bp UST depeg on Binance today ($215mm 24hr volume) is indicative of Binance being irregularly long UST

Why?

3AC / Genesis likely swapped their $1.5bn UST on Binance over the last 2-3 wks

April 19/20 had over $800mm in volume alone

Evidence of pool imbalance in Curve as foreshadow?

94.7% of my personal wealth is in UST during the depeg 🥲