FRAUD DETECTION & DEFENSE

Getting Started!

<Data and Motivation>

/// Introduction

The **Fraud Detection & Defense (FDD) workstream** and the current **Open Data Community** have been open-source funding's first line of defense against malicious sybil attackers who attempt to manipulate Gitcoin's donor reward system. Consisting of high caliber developers, data scientists, and digital detectives, the team worked tirelessly to both proactively prevent and reactively thwart active threats to web3's flagship public funding protocol, while continuously improving fraud defense detection processes and contributors' behavior and tendencies analyses. Through their experience defending the Gitcoin protocol, the team has developed advanced analytical approaches for the analysis and understanding of on-chain behavior of wallet accounts. To this end, the team proposed a wallet segmentation analysis that would help Aave better understand the behaviors of their user base & characterize platform utilization. To that end, this report and presentation are a culmination of those efforts.

Our goal is to enhance Aave's understanding of their current and future user base by performing a wallet segmentation analysis that will unearth insights into how groups of users borrow, lend, and interact with the Aave platform. Successful completion of this project will produce personas of wallet addresses that will:

1. Provide insight into Aave platform usage and group behavior.

2. Guide and direct marketing and product development.

3. Produce reproducible analysis that can be used to leverage and identify user behavioral pattern (usage, attitudes, loyalty).

/// Data

Before data collection could begin, **a discovery step had to be initiated to identify and track each of the contracts deployed by Aave**. **Contracts for the platform** have been well **documented across multiple sources**, where the most relevant can be found for V1, V2, and V3 here, here, and here. It should be noted that not every callable method corresponds to a relevant behavior of interest. To that end, **we focused only on events that mapped directly to behaviors to be used within the segmentation model**. These contract methods were:

- **deposit()**
- **supply()**
- **borrow()**
- **repay()**
- **withdraw()**
- **redeemUnderlying()**
- **liquidationCall)**
- **flashLoan()**

Demographic comparisons across profiles will be of interest and to support this task, **additional data was collected from other on chain sources such as DeBank**, the web3 social media platform **Lens Protocol**, off-chain governance platform **Snapshot**, **Gitcoin's proof of personhood Passport** and the experimental sybil score created by **Trustalabs**. Since most of these platforms exist only on the Ethereum mainnet, we were forced to constrain our analysis to only wallet addresses that interacted with Aave on that blockchain. **This resulted in a final dataset of 114,915 wallet addresses with complete records suitable for segmentation through an unsupervised machine learning routine.**

/// Data

address	CV_polygon_v2_calls	CE_fantom_Repay_calls	VE_v2_Repay_calls	CVE_ethereum_v1_Borrow_calls	CVE_optimism_v3_LiquidationCall_calls	aave_debt_usd
C_arbitrum_calls	CV_polygon_v3_calls	CE_fantom_Supply_calls	VE_v2_Withdraw_calls	CVE_ethereum_v1_Deposit_calls	CVE_optimism_v3_Repay_calls	aave_net_usd
C_avalanche_calls	CE_arbitrum_Borrow_calls	CE_fantom_Withdraw_calls	VE_v3_Borrow_calls	CVE_ethereum_v1_FlashLoan_calls	CVE_optimism_v3_Supply_calls	if_lens_data
C_ethereum_calls	CE_arbitrum_FlashLoan_calls	CE_optimism_Borrow_calls	VE_v3_FlashLoan_calls	CVE_ethereum_v1_LiquidationCall_calls	CVE_optimism_v3_Withdraw_calls	lens_prof_count
C_fantom_calls	CE_arbitrum_LiquidationCall_calls	CE_optimism_FlashLoan_calls	VE_v3_LiquidationCall_calls	CVE_ethereum_v1_RedeemUnderlying_calls	CVE_polygon_v2_Borrow_calls	lens_id
C_optimism_calls	CE_arbitrum_Repay_calls	CE_optimism_LiquidationCall_calls	VE_v3_Repay_calls	CVE_ethereum_v1_Repay_calls	CVE_polygon_v2_Deposit_calls	lens_name
C_polygon_calls	CE_arbitrum_Supply_calls	CE_optimism_Repay_calls	VE_v3_Supply_calls	CVE_ethereum_v2_Borrow_calls	CVE_polygon_v2_FlashLoan_calls	lens_isDefault
E_Borrow_calls	CE_arbitrum_Withdraw_calls	CE_optimism_Supply_calls	VE_v3_Withdraw_calls	CVE_ethereum_v2_Deposit_calls	CVE_polygon_v2_LiquidationCall_calls	lens_followers
E_Deposit_calls	CE_avalanche_Borrow_calls	CE_optimism_Withdraw_calls	CVE_arbitrum_v3_Borrow_calls	CVE_ethereum_v2_FlashLoan_calls	CVE_polygon_v2_Repay_calls	lens_following
E_FlashLoan_calls	CE_avalanche_Deposit_calls	CE_polygon_Borrow_calls	CVE_arbitrum_v3_FlashLoan_calls	CVE_ethereum_v2_LiquidationCall_calls	CVE_polygon_v2_Withdraw_calls	lens_posts
E_LiquidationCall_calls	CE_avalanche_FlashLoan_calls	CE_polygon_Deposit_calls	CVE_arbitrum_v3_LiquidationCall_calls	CVE_ethereum_v2_Repay_calls	CVE_polygon_v3_Borrow_calls	lens_comments
E_RedeemUnderlying_calls	CE_avalanche_LiquidationCall_calls	CE_polygon_FlashLoan_calls	CVE_arbitrum_v3_Repay_calls	CVE_ethereum_v2_Withdraw_calls	CVE_polygon_v3_FlashLoan_calls	lens_mirrors
E_Repay_calls	CE_avalanche_Repay_calls	CE_polygon_LiquidationCall_calls	CVE_arbitrum_v3_Supply_calls	CVE_ethereum_v3_Borrow_calls	CVE_polygon_v3_LiquidationCall_calls	lens_publications
E_Supply_calls	CE_avalanche_Supply_calls	CE_polygon_Repay_calls	CVE_arbitrum_v3_Withdraw_calls	CVE_ethereum_v3_FlashLoan_calls	CVE_polygon_v3_Repay_calls	lens_collects
E_Withdraw_calls	CE_avalanche_Withdraw_calls	CE_polygon_Supply_calls	CVE_avalanche_v2_Borrow_calls	CVE_ethereum_v3_LiquidationCall_calls	CVE_polygon_v3_Supply_calls	snap_voted_aave
V_1_calls	CE_ethereum_Borrow_calls	CE_polygon_Withdraw_calls	CVE_avalanche_v2_Deposit_calls	CVE_ethereum_v3_Repay_calls	CVE_polygon_v3_Withdraw_calls	snap_voted_aave_numprop
V_2_calls	CE_ethereum_Deposit_calls	VE_v1_Borrow_calls	CVE_avalanche_v2_FlashLoan_calls	CVE_ethereum_v3_Supply_calls	if_debank_data	trustalabs_is
V_3_calls	CE_ethereum_FlashLoan_calls	VE_v1_Deposit_calls	CVE_avalanche_v2_LiquidationCall_calls	CVE_ethereum_v3_Withdraw_calls	agedays	trustalabs_score
CV_arbitrum_v3_calls	CE_ethereum_LiquidationCall_calls	VE_v1_FlashLoan_calls	CVE_avalanche_v2_Repay_calls	CVE_fantom_v3_Borrow_calls	iscontract	trustalabs_label
CV_avalanche_v2_calls	CE_ethereum_RedeemUnderlying_calls	VE_v1_LiquidationCall_calls	CVE_avalanche_v2_Withdraw_calls	CVE_fantom_v3_FlashLoan_calls	ismultisig	
CV_avalanche_v3_calls	CE_ethereum_Repay_calls	VE_v1_RedeemUnderlying_calls	CVE_avalanche_v3_Borrow_calls	CVE_fantom_v3_LiquidationCall_calls	num_chains_active	
CV_ethereum_v1_calls	CE_ethereum_Supply_calls	VE_v1_Repay_calls	CVE_avalanche_v3_FlashLoan_calls	CVE_fantom_v3_Repay_calls	num_tokens	
CV_ethereum_v2_calls	CE_ethereum_Withdraw_calls	VE_v2_Borrow_calls	CVE_avalanche_v3_LiquidationCall_calls	CVE_fantom_v3_Supply_calls	aave_amt	
CV_ethereum_v3_calls	CE_fantom_Borrow_calls	VE_v2_Deposit_calls	CVE_avalanche_v3_Repay_calls	CVE_fantom_v3_Withdraw_calls	bal_usd	
CV_fantom_v3_calls	CE_fantom_FlashLoan_calls	VE_v2_FlashLoan_calls	CVE_avalanche_v3_Supply_calls	CVE_optimism_v3_Borrow_calls	eth_bal_usd_	
CV_optimism_v3_calls	CE_fantom_LiquidationCall_calls	VE_v2_LiquidationCall_calls	CVE_avalanche_v3_Withdraw_calls	CVE_optimism_v3_FlashLoan_calls	aave_asset_usd	

Figure: Columns of the full superset of data collected across all data sources, blockchains and versions.

/// Data Dictionary

num_tokens_eth: Number of tokens held on the Ethereum blockchain.
bal_usd_eth: Total balance, in USD held by that wallet on the Ethereum blockchain.
aave_asset_usd_eth: Total amount of assets, in USD, supplied to the Aave platform.
aave_debt_usd_eth: Total amount of debt, in USD, borrowed from the Aave platform.
CE_ethereum_Deposit_calls: Total number of deposit events recorded for this address on the Ethereum blockchain.
CE_ethereum_Withdraw_calls: Total number of withdraw and/or redeem events recorded for this address on the Ethereum blockchain.
CE_ethereum_Borrow_calls: Total number of borrow events recorded for this address on the Ethereum blockchain.
CE_ethereum_FlashLoan_calls: Total number of flash loan events recorded for this address on the Ethereum blockchain.
CE_ethereum_Repay_calls: Total number of repay events recorded for this address on the Ethereum blockchain.
CE_ethereum_LiquidationCall_calls: Total number of liquidations events recorded for this address on the Ethereum blockchain.
if_debank_data: An indicator variable for whether we have Debank data for that particular address.
if_lens_data: An indicator variable of whether an address has a lens profile. If not, an NA is recorded.
snap_voted_aave: An indicator variable if whether an address has voted on any Aave snapshot proposals.
snap_voted_aave_numprop: Total number of proposals voted on by that wallet address on the Snapshot platform. Zeros are represented as NAs.
trustalabs_is: An indicator of whether a wallet address has a TrustaLabs sybil score.
trustalabs_score: A quantitative variable recording the TrustaLabs sybil risk score. Scores range from 0 to 100.
passport_is: An indicator of whether a wallet address has a Gitcoin Passport.
passport_numstamps: A count variable recording the number of stamps associated with that user's Gitcoin Passport. num_chains_active: Number of chains this wallet has interacted with.

Data dictionary for all variables included in the profiling analysis.

Searching the Data!

<Exploratory Analysis>

/// Per Chain Exploratory Analysis

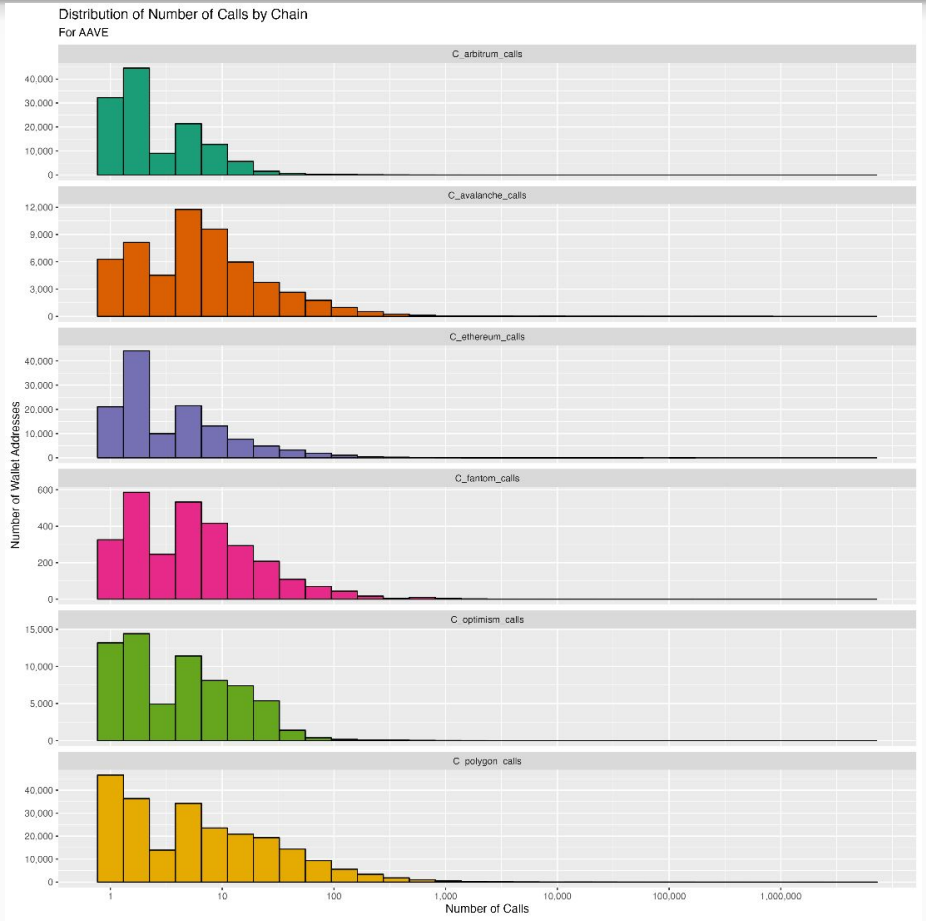


Figure: The distribution of the number of contract calls within each chain with an Aave deployment.

Activity on each chain follow similar poisson distributions, with most activity occurring on Ethereum and Arbitrum.

/// Per Version Exploratory Analysis

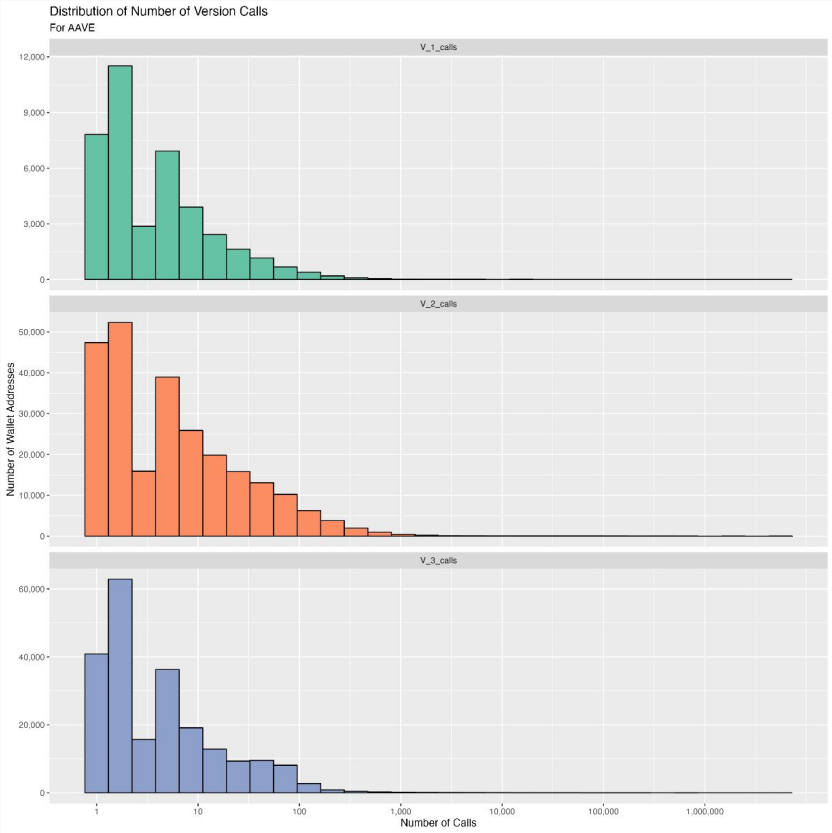


Figure: The Aave v2 deployment, which rests on Avalanche, Ethereum and Polygon, had the most contract interactions across all summary statistics.

The predominant amount of contract interactions come from the v2 and v3 protocol versions.

/// Per Call Exploratory Analysis

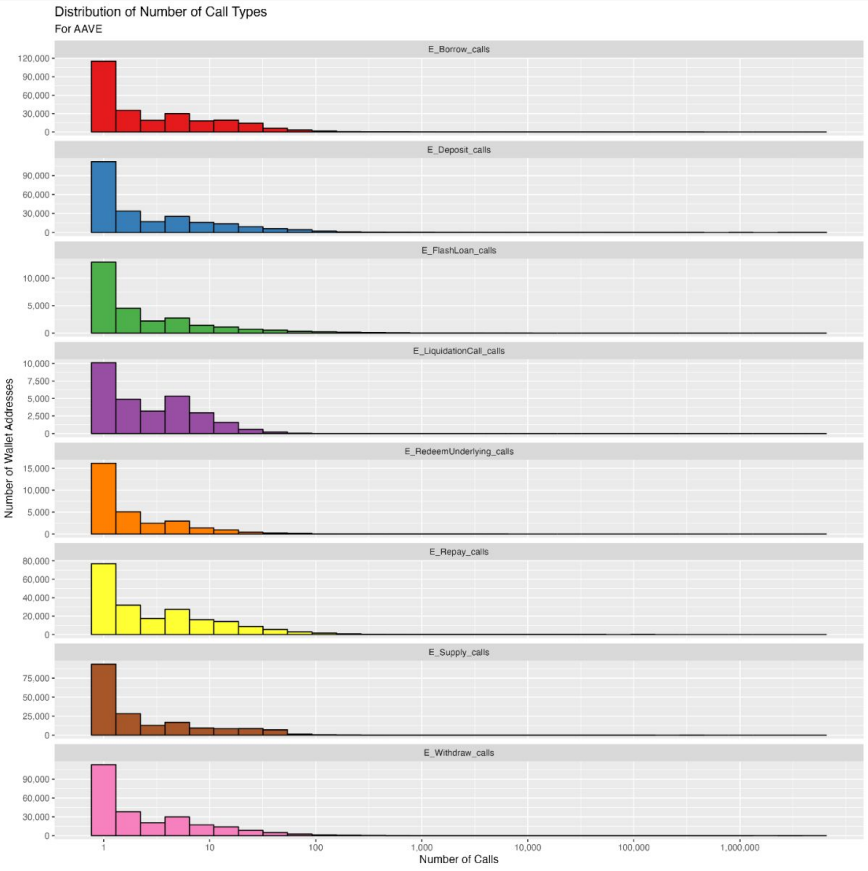


Figure: Distribution of events emitted from the Ethereum based mainnet Aave contracts.

These are all long tail distributions with most activity occurring within the deposit(), borrow(), withdraw(), supply() and repay() events.

/// Variable Cross Correlations

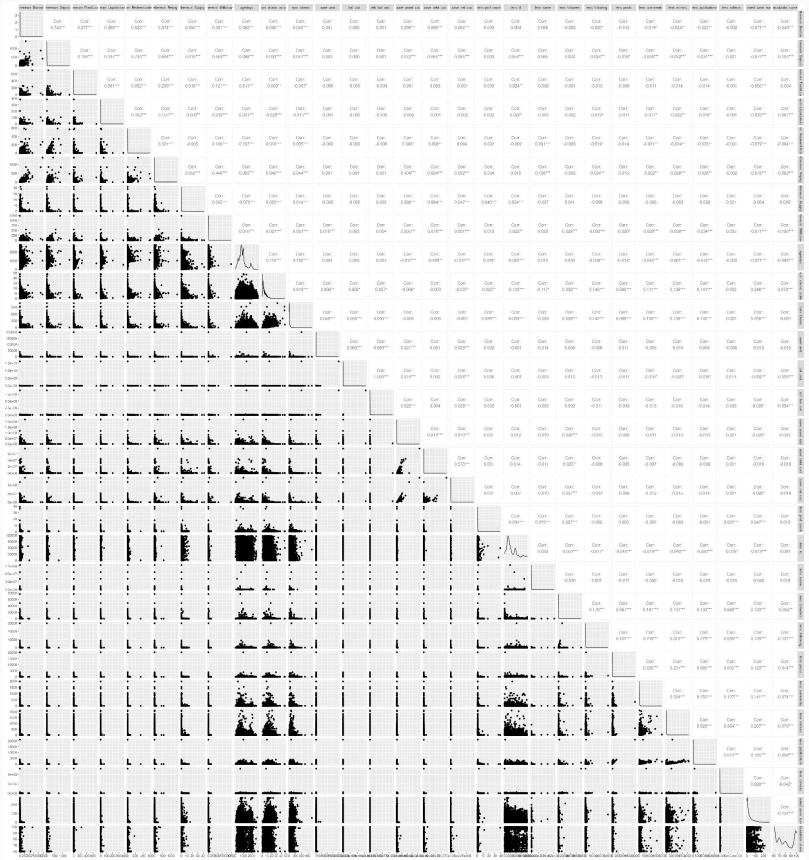


Figure: Zoomed out view of the scatterplot matrix calculated for all variables endogenous and exogenous to the segmentation model.

The highlight here is that very few of the variables are correlated outside of the event calls that go hand in hand like:

borrow() <-> repay() & deposit() <-> withdraw().

As an appeal to Occum's Razer, we merged the relevant calls according to the following rubric:
Deposit = Deposit (v1 and v2) and Supply (v3) while
Withdraw = Redeem (v1) and Withdraw (v2 and v3).

/// EDA Inspired Insights

Outlier analysis identified several interesting wallets. The most extreme 3 were:

Binance's Hot Wallets at 0xf977814e90da44bfa03b6295a0616a897441acec
MakerDAO's Vault at 0x7d6149ad9a573a6e2ca6ebf7d4897c1b766841b4
patricioworthalter.eth at 0x57757E3D981446D585Af0D9Ae4d7DF6D64647806.

Given their networth's are **clear outliers and institutions** are their own special class of users with usage likely to differ, **we removed** them from the dataset, **yet kept individual whale accounts within the dataset.**

The exploratory analysis journey culminated into a set of facts, justifications and verifications for a few of the assumptions and design choices for this study. **Specifically, the analysis backs the removal of contracts and multi-signature wallets from the sample to reduce bias, the merging of function calls to reduce complexity, the removal of outliers to further reduce bias, and the intentional focus on Ethereum wallet addresses due to data availability.**

How did you do it?!

<Methodology>

/// Multi-Dimensional Scaling

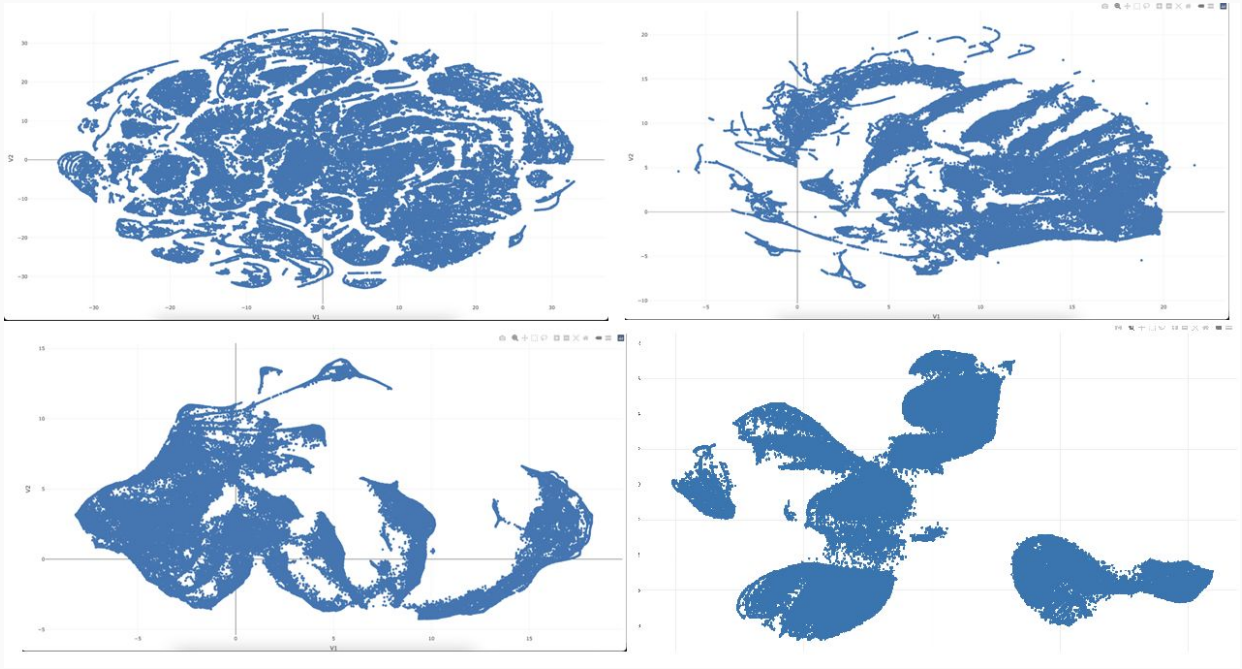


Figure: A visualization of the UMAP parameter search that culminated into the final solution on the bottom right.

Like t-SNE, Uniform Manifold Approximation and Projection (UMAP) is a non-linear dimension reduction technique that must be parameterized during initialization. As part of the search routine, we looked across both t-SNE and Uniform Manifold Approximation and Projections at various parameter configurations to find 2D projections with meaningful separation. The following figure shows four such iterations where, going from left to right, top to bottom we see more distinct groups forming in the 2D space. As mentioned previously, the assumption is that groups of points are related and if those groups are sufficiently distant from other groups there will be some quantifiable difference in their values.

/// Multi-Dimensional Scaling

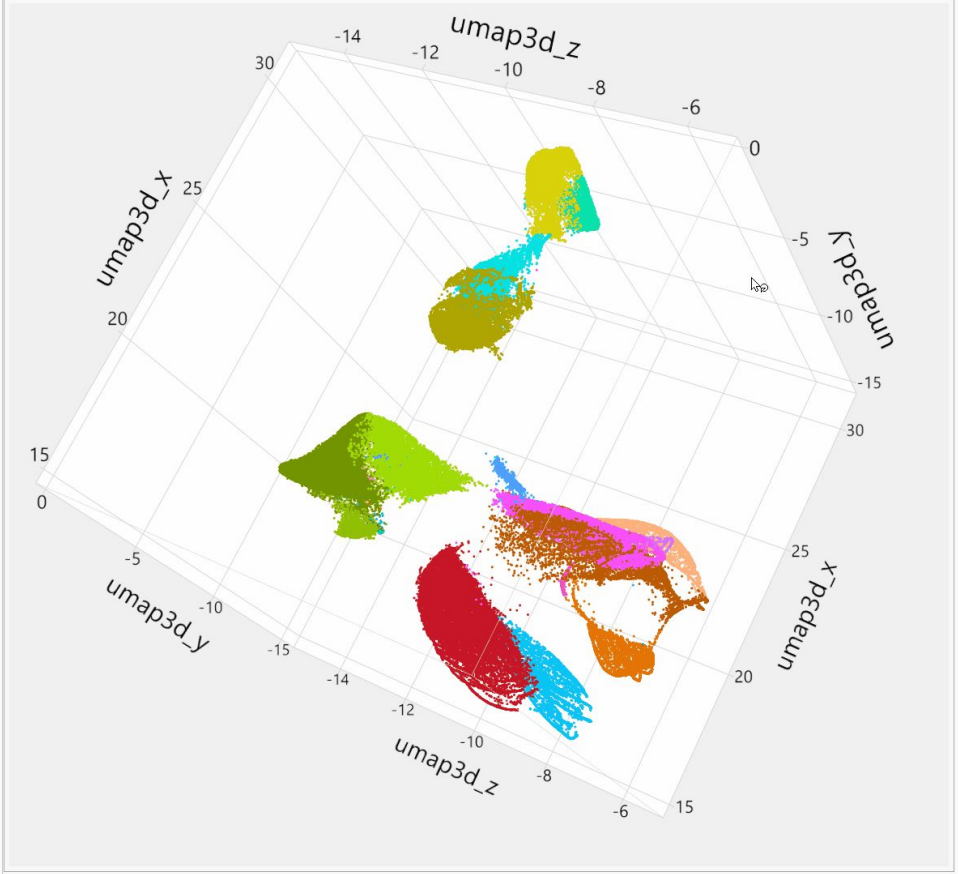


Figure: An alternative 3D view of the wallet addresses within the UMAP derived feature space. Notice that the distance between specific clusters becomes more pronounced at various perspectives.

Step 1. Plot the 3-Dimensional coordinate system of the UMAP projection and investigate the feature space visually.

/// Multi-Dimensional Scaling

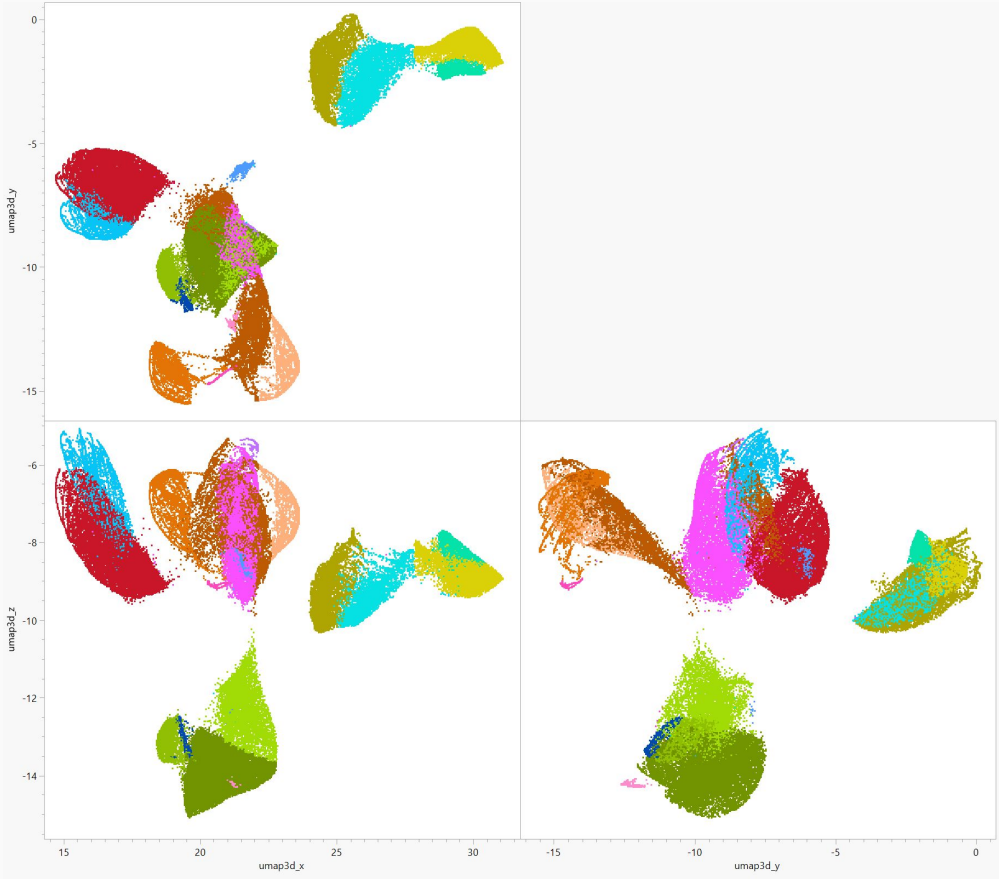


Figure: Multiple 2D projections of the UMAP solution used to manually select specific regions of the graph and those corresponding wallets.

Step 2. Plot multiple 2D projections of the UMAP space.

/// Interactive Graphics and Tables

		OmniUMAP																	
		6	9	12	19	22	25	28	37	39	40	42	43	44	45	46	53	70	78
num_chains_active	Mean	1.0001083306	3.4260262258	4.4912280702	5.777823169	2.6291020365	9.1807228916	5.0169079516	10.962562396	3.3687607573	2.4119318182	9.3173453123	4.2563875786	5.8509079811	2.0239529547	1	1.0398505604	1	0
num_tokens_eth	Mean	2.0892644351	6.0453249715	5.230387289	9.465013131	4.0683760684	6.4728915663	7.7578327571	17.496672213	6.3700516351	0	8.9113172004	8.5753734277	12.128643369	3.0556511761	0	1.8356164384	1.005409304	0
bal_usd_eth	Mean	25261.719419	74507.171695	37586.198069	40563.290721	25687.067471	283.42226454	45412.981217	2980.2797141	6611.8717602	0	2908.3042299	965264.12124	159242.1815	17280.292715	0	25097.357886	71.348640845	0
aave_asset_usd_eth	Mean	0.8473344166	528765.3369	0.3246613704	0.1697130435	72554.513268	1.8810773785	1.139809745	16.497499168	2021.8195693	63.566417614	42.946320243	24767.369726	64404.148974	0.1328066552	160.47364691	0.1374694894	397.92230472	0
aave_debt_usd_eth	Mean	0.0012901094	233845.69232	0.0182601787	0.0020355413	0.0637357814	0.5904722684	0.1553795916	0.5293427621	941.52039415	0	15.830370091	0.0467963836	2.469214253	0.0013304647	0.0115953608	0.0165716065	0.0076696718	0
CE_ethereum_Borrow_calls	Mean	0.0012999675	6.9680729761	1.9321416749	0.0025094835	0.0262741374	1.031262983	6.6136019879	0.2179700499	8.2267641997	0	1.0641935951	0.0045204403	7.1706088814	0.0014343087	0.0154639175	1.2826899128	0.0036062027	0.4968492487
CE_ethereum_Deposit_calls	Mean	1.6758747698	5.9017958951	0.0215160543	1.8320396849	1.5790862087	1.0628375571	6.9345289542	7.7911813644	5.7728055077	1.2528409091	1.1407042458	1.8701847484	6.8173355715	1.6936316695	1.1610824742	0.0062266501	1.1831950956	1.7285506544
CE_ethereum_FlashLoan_calls	Mean	0.0111580544	0.7373147092	0.3174445548	0.008578932	0.0097077134	0.0001038637	0.6166810717	0.0008319468	1.2082616179	0.0056818182	0.0031879438	0.0119889937	0.6003357241	0.0159208262	0.0012886598	0.2540473225	0.0068517851	0.1342704799
CE_ethereum_LiquidationCall_calls	Mean	0.0003249919	0.268386545	0.0642171466	0.0008170411	0.0013717421	0.0018695472	0.4244814175	0	7.571858864	0	0.0134763078	0.0003930818	0.4181291012	0.0001434309	0	0.0361145704	0	0.0489578284
CE_ethereum_Repay_calls	Mean	0.0146246344	4.7558437856	1.8828202582	0.0387510943	0.0167774612	0.021603656	5.6273228176	0.0291181364	3.0753012048	0.0085227273	0.0427474279	0.014740566	6.8683045933	0.0240963855	0.0077319588	1.5417185554	0.0036062027	0.4847309743
CE_ethereum_Withdraw_calls	Mean	1.4741631459	2.8021664766	0.0390599139	1.7168368836	0.4727234357	0.0165143332	5.0192307692	0.0424292845	1.7594664372	0.1079545455	0.0685407912	0.7468553459	4.9389592553	1.5149168101	0	0.0174346202	0.0003606203	1.4648570044
N		9231	7016	3021	17135	9477	9628	18512	1202	2324	352	6901	10176	6553	6972	776	803	2773	2063

Figure : An interactive table of cluster means that is automatically updated when colored regions are added, removed or modified in the 2D map.

Step 3. Link the graphs with an interactive table that tracks the cluster averages.

/// Brushing in the MDS Space

Figure : An animation showing the individual groups after the manual brushing process has been completed.

Step 4. Brush then Investigate pairs of clusters for cohesion, compactness and quantitative distinctness.

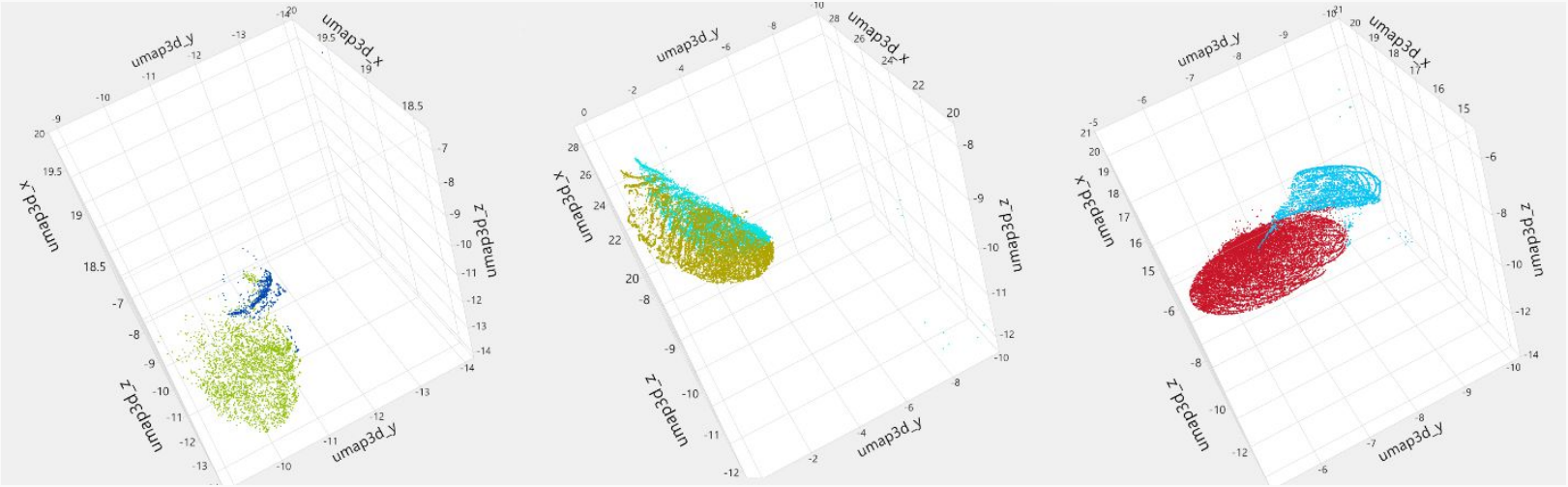


Figure: Sample visualizations of pairwise comparisons done to assess whether groups should be merged or continue to exist separated from one another.

Step 5. Review the final cluster solution in 2D and 3D space.

/// Brushing in the MDS Space

Figure : The final clustering solution after manually color brushing groups that were "interesting", relatively distinct and reasonably partitionable.

Utilizing a **manual brushing technique** within an interactive graphical framework to color code partitions inside of a 3-Dimensional non-linear feature space yielded a final segmentation solution **consisting of 13 distinct groupings of wallet addresses**.

Segmentation Fun!

<Results>

/// Aave Wallet Profiles

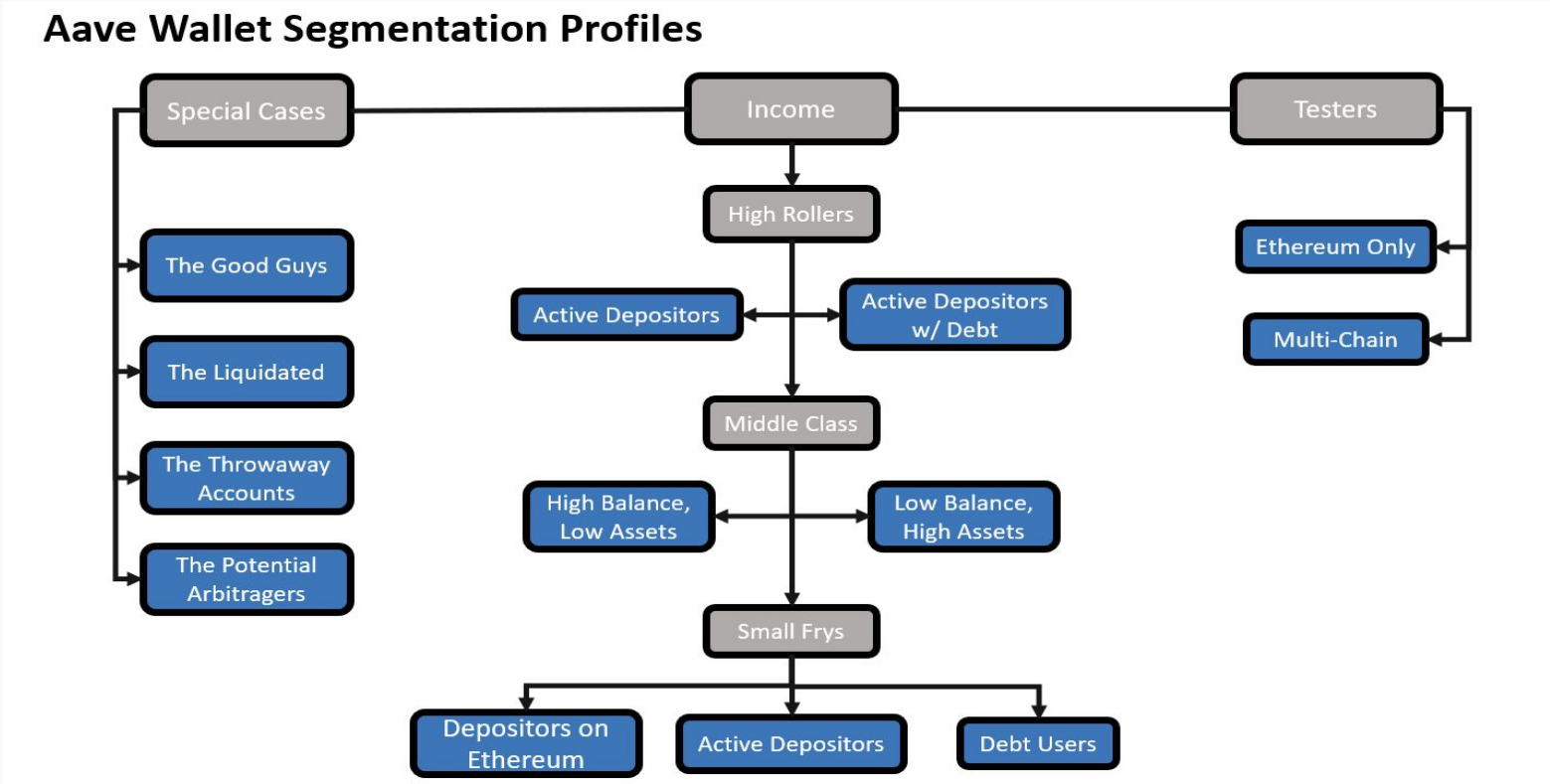


Figure: High level break down of one such categorization of the Aave wallet profiles.

/// Aave Wallet Profiles

FinalClusters	Counts	Proportions	num_chains_active	num_tokens_eth	bal_usd_eth	aave_asset_usd_eth	aave_debt_usd_eth	CE_ethereum_Borrow_calls	CE_ethereum_Deposit_calls	CE_ethereum_FlashLoan_calls	CE_ethereum_LiquidationCall_calls	CE_ethereum_Repay_calls	CE_ethereum_Withdraw_calls
ETH Small Frys Depositooors	2773	2.41%	1.000	1.005	$71.35	$397.92	$0.01	0.004	1.183	0.007	0.000000	0.004	0.0004
ETH Testers	9231	8.03%	1.000	2.089	$25,261.72	$0.85	$0.00	0.001	1.676	0.011	0.000325	0.015	1.4742
High Rollers AD	6553	5.70%	5.851	12.129	$159,242.18	$64,404.15	$2.47	7.171	6.817	0.600	0.418129	6.868	4.9390
High Rollers AD w/Debt	7016	6.11%	3.426	6.045	$74,507.17	$528,765.34	$233,845.69	6.968	5.902	0.737	0.268387	4.756	2.8022
Middle-Class HC-LS	10174	8.85%	4.255	8.545	$67,959.50	$24,205.10	$0.05	0.005	1.869	0.012	0.000393	0.015	0.7441
Middle-Class LC-HS	9477	8.25%	2.629	4.068	$25,687.07	$72,554.51	$0.06	0.026	1.579	0.010	0.001372	0.017	0.4727
Multi-Chain Testers	24107	20.98%	4.692	7.611	$33,829.60	$0.16	$0.00	0.002	1.792	0.011	0.000622	0.035	1.6584
Potential Arbitragers	3824	3.33%	3.766	4.518	$34,963.67	$0.29	$0.02	1.796	0.018	0.304	0.058316	1.811	0.0345
Small Fry Degen Active Depositoors	1202	1.05%	10.963	17.497	$2,980.28	$16.50	$0.53	0.218	7.791	0.001	0.000000	0.029	0.0424
Small Fry Degen Debt Users	16529	14.38%	9.238	7.491	$1,379.33	$19.03	$6.95	1.045	1.095	0.001	0.006715	0.030	0.0382
The Good Guys	18512	16.11%	5.017	7.758	$45,412.98	$1.14	$0.16	6.614	6.935	0.617	0.424481	5.627	5.0192
The Liquidated	2324	2.02%	3.369	6.370	$6,611.87	$2,021.82	$941.52	8.227	5.773	1.208	7.571859	3.075	1.7595
Throw Away Accounts	3191	2.78%	0.509	0.000	$0.00	$46.04	$0.00	0.325	1.538	0.088	0.031652	0.316	0.9589

Figure: A quantitative representation of the group mean vectors also shown in the parallel coordinate plots. Values have been placed along a color gradient for easier examination.

/// Aave Wallet Profiles

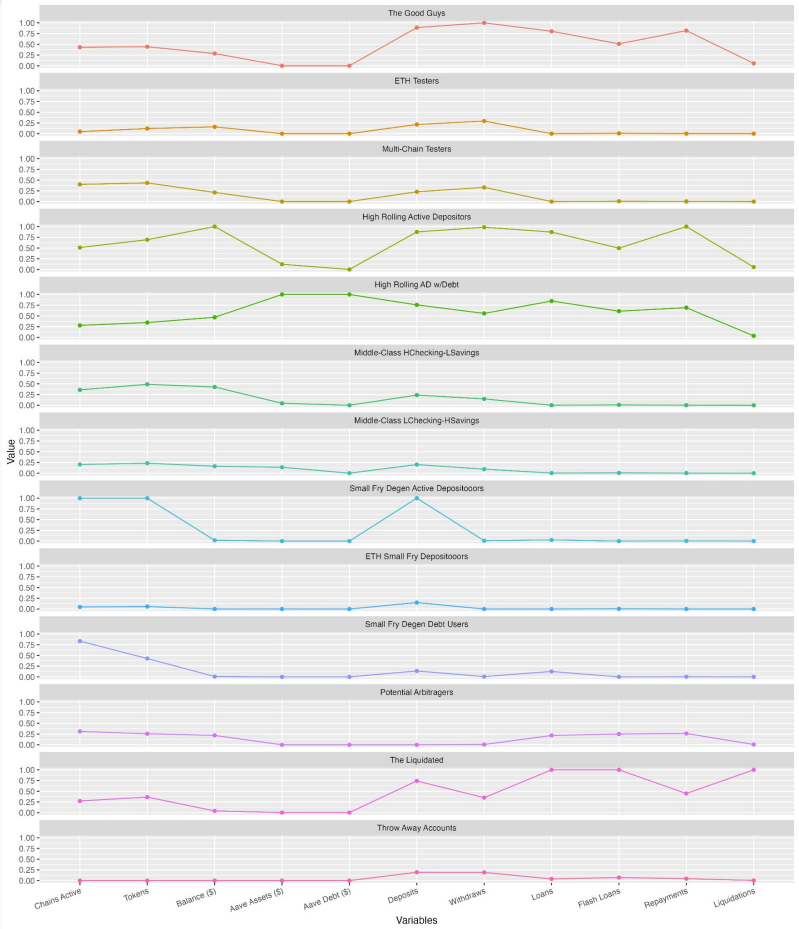
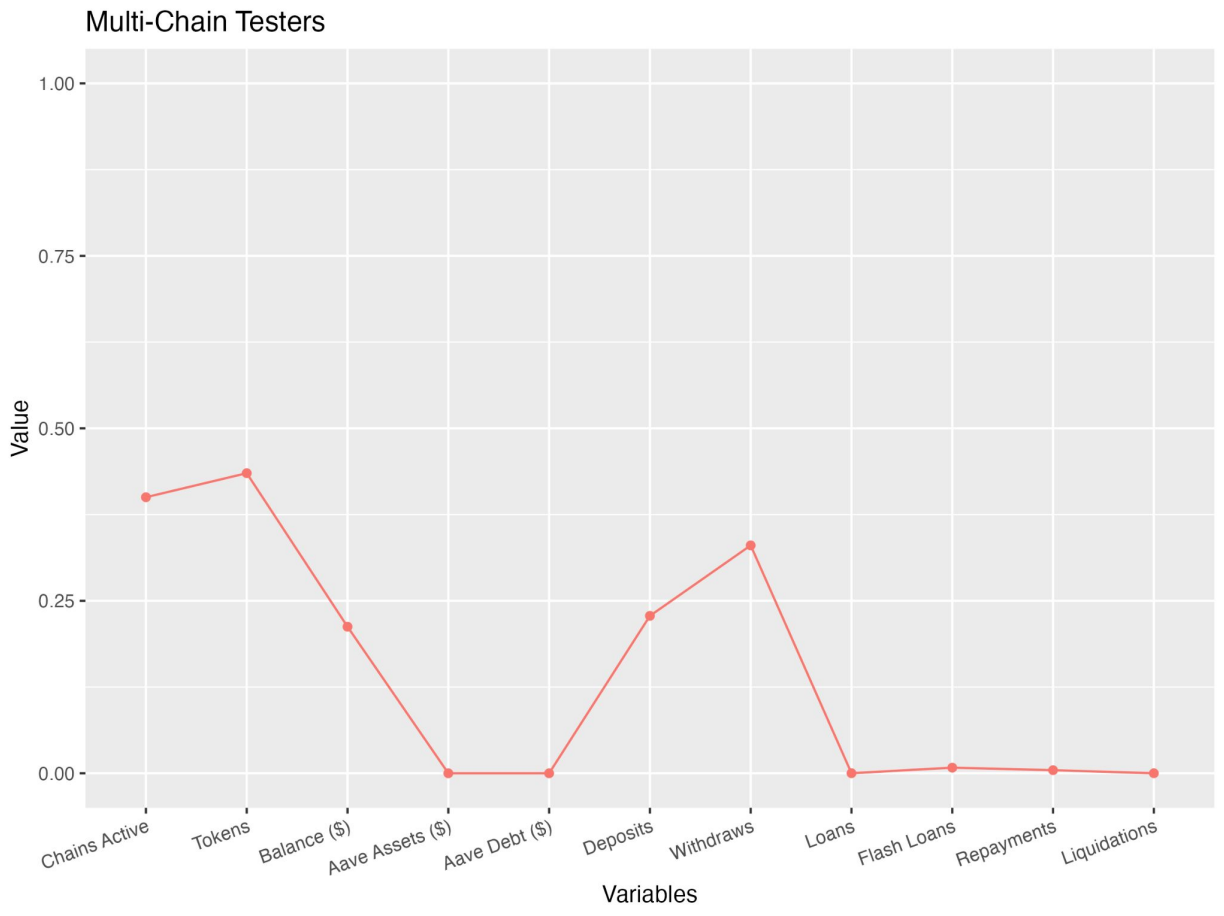


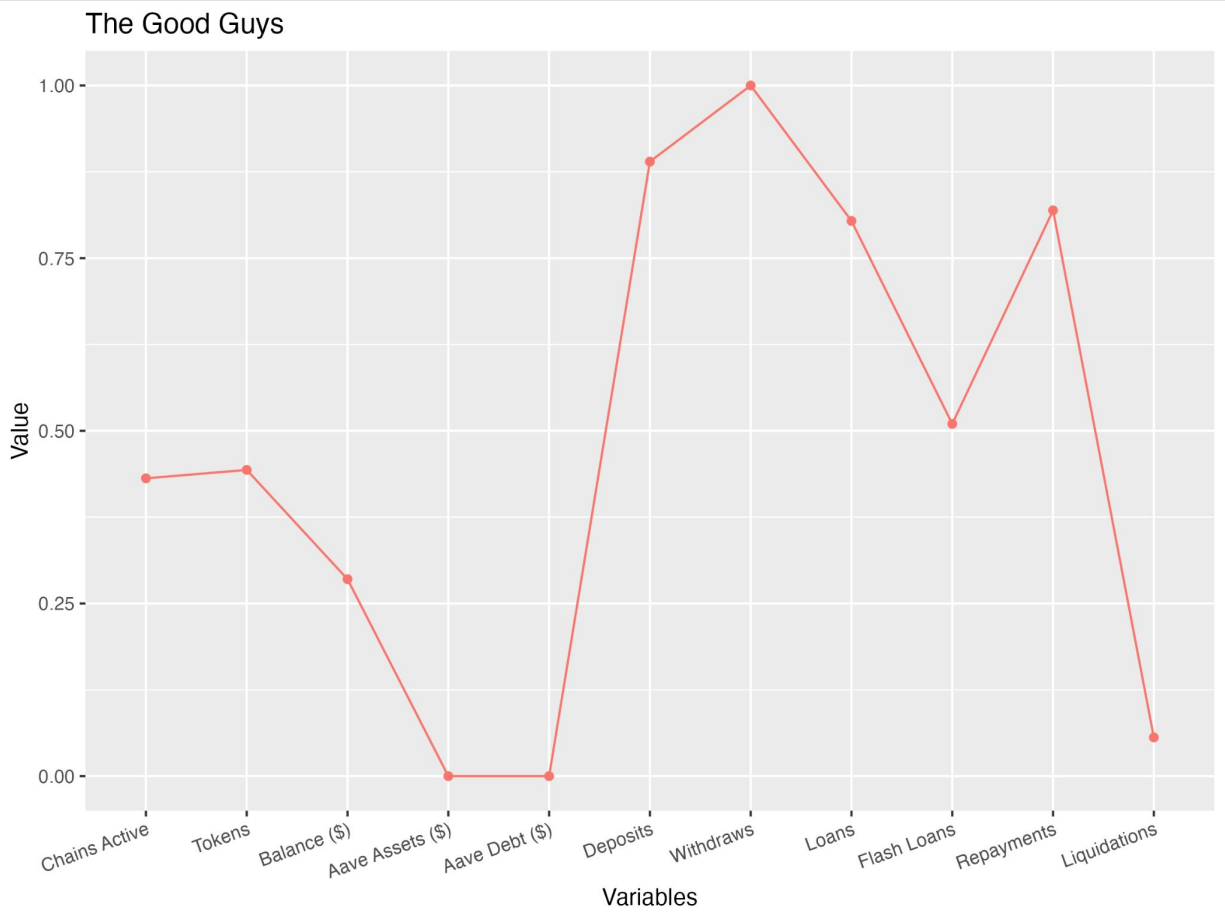
Figure: Parallel coordinate plots are the standard way user segments are visualized in multiple dimensions. The x-axis corresponds to a set of variables whose values are shown on the y-axis. The profile name in the title references the group.

/// Aave Wallet Profiles



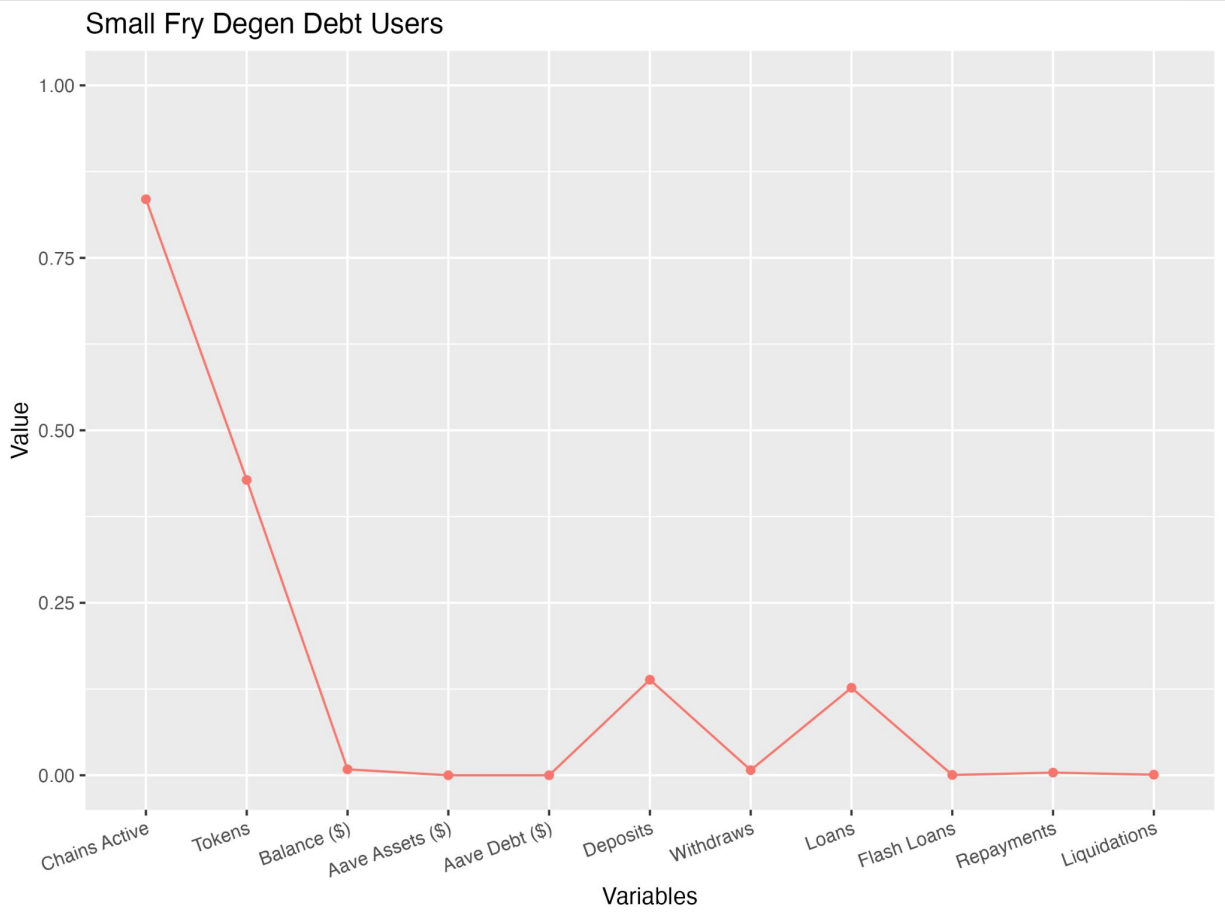
Multi-Chain Testers (24,107 | 20.98%) - These wallets are active on more than one chain, yet interacted with the ETH Aave contract in the same way as the other testers, by refraining from borrowing, but making and withdrawing one or two deposits.

/// Aave Wallet Profiles



The Good Guys (18,512 | 16.11%) - The credit industry maliciously calls these users "deadbeats" but they are shining examples of users that make deposits, borrow money, repay their loans and withdrawn when necessary. They are rarely liquidated or call for liquidations and keep a large balance in their wallets.

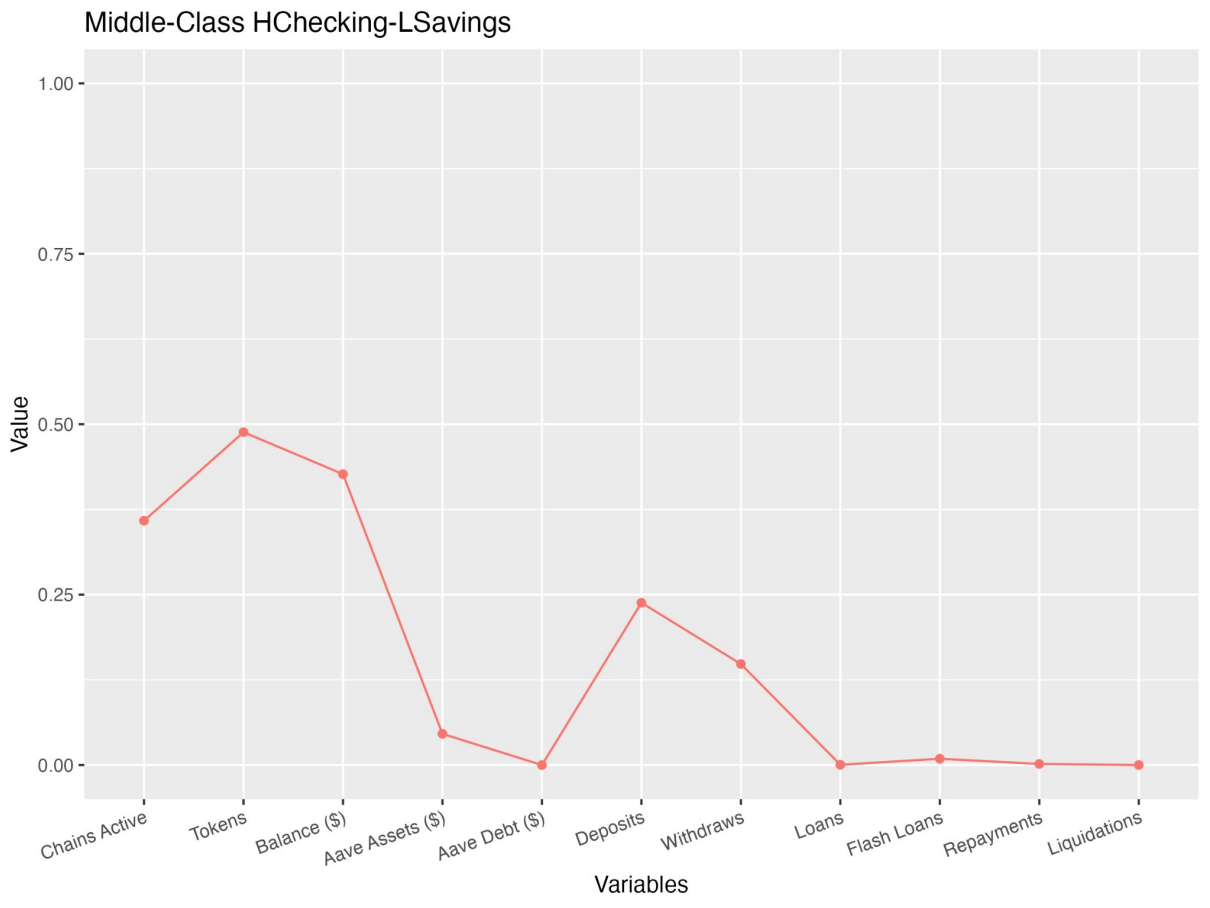
/// Aave Wallet Profiles



Small Fry Degen Debt Users (16,529 | 14.38%)

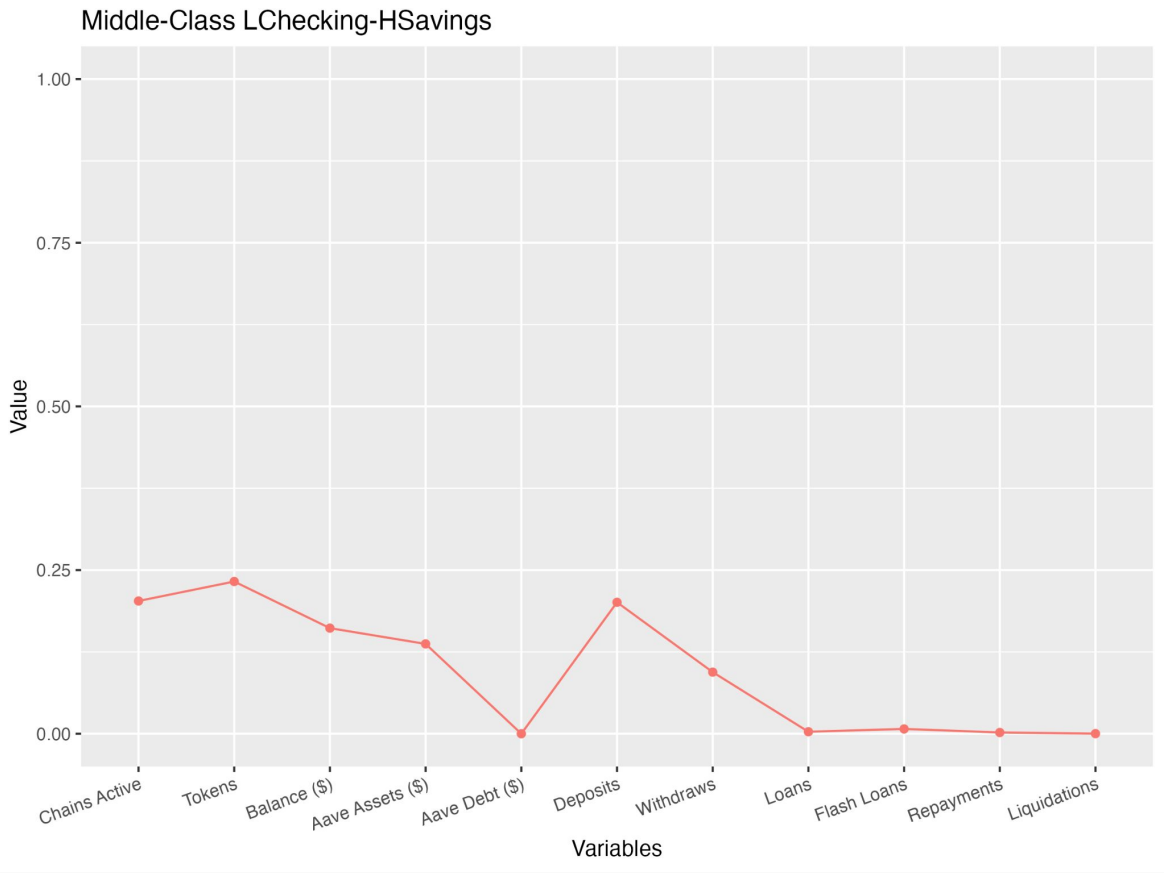
- These small fries deposit, but also try out the platform's borrowing functionality with a small amount of funds that they keep outstanding.

/// Aave Wallet Profiles



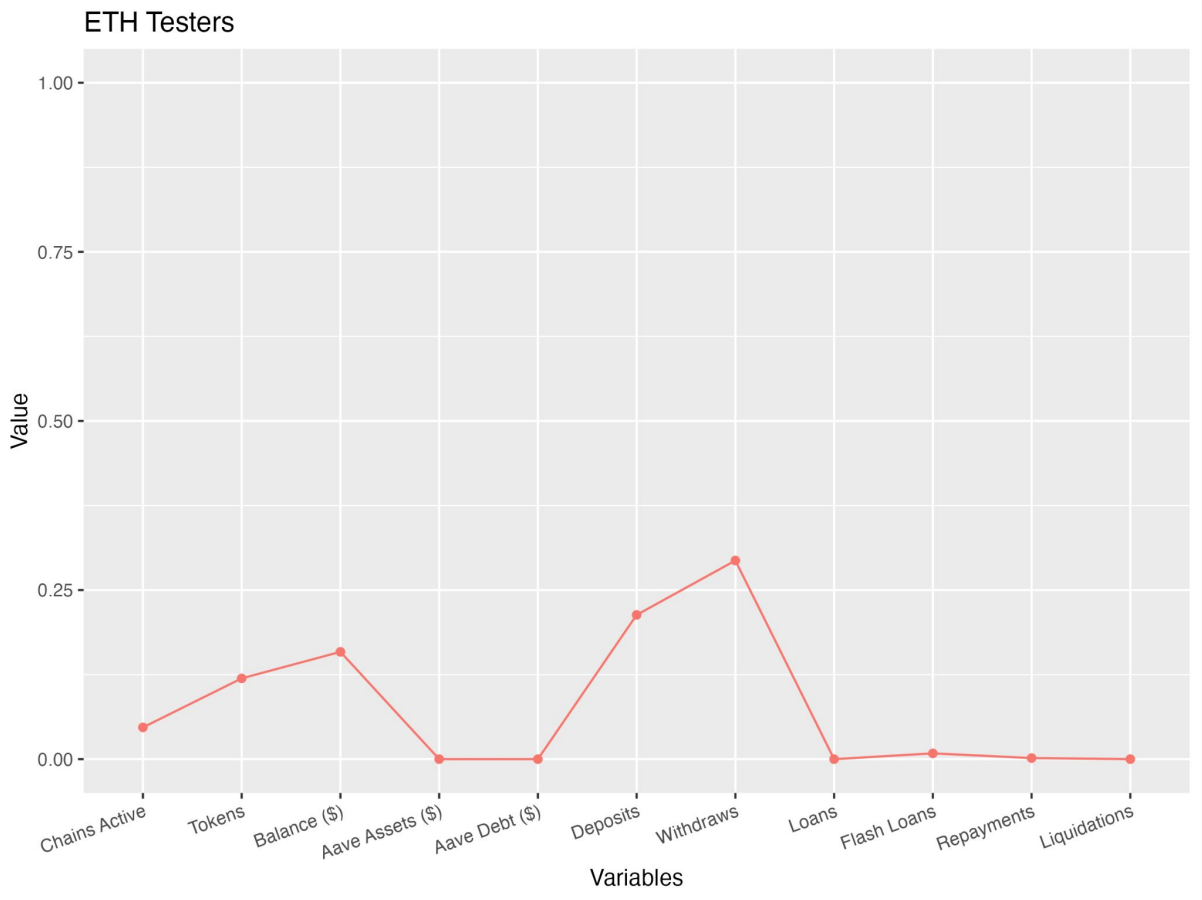
Middle-Class High Checking - Low Savings (10,174 | 8.85%) - This cluster of users keep a high wallet balance (checking), but a lower deposit balance (savings). Interestingly, neither of the middle class profiles use debt.

/// Aave Wallet Profiles



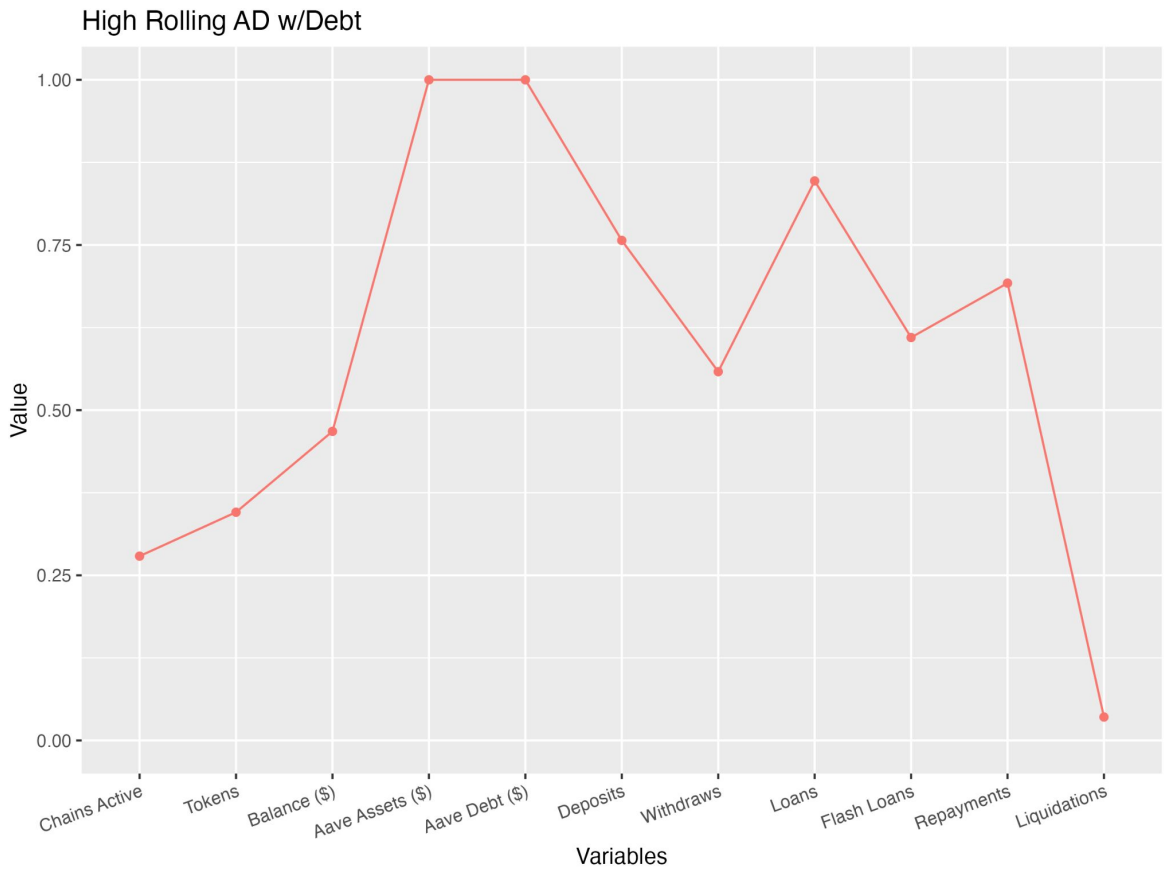
Middle-Class Low Checking - High Savings (9,477 | 8.25%) - This cluster of users keep a low wallet balance (checking), but a higher deposit balance (savings). Interestingly, neither of the middle class profiles use debt.

/// Aave Wallet Profiles



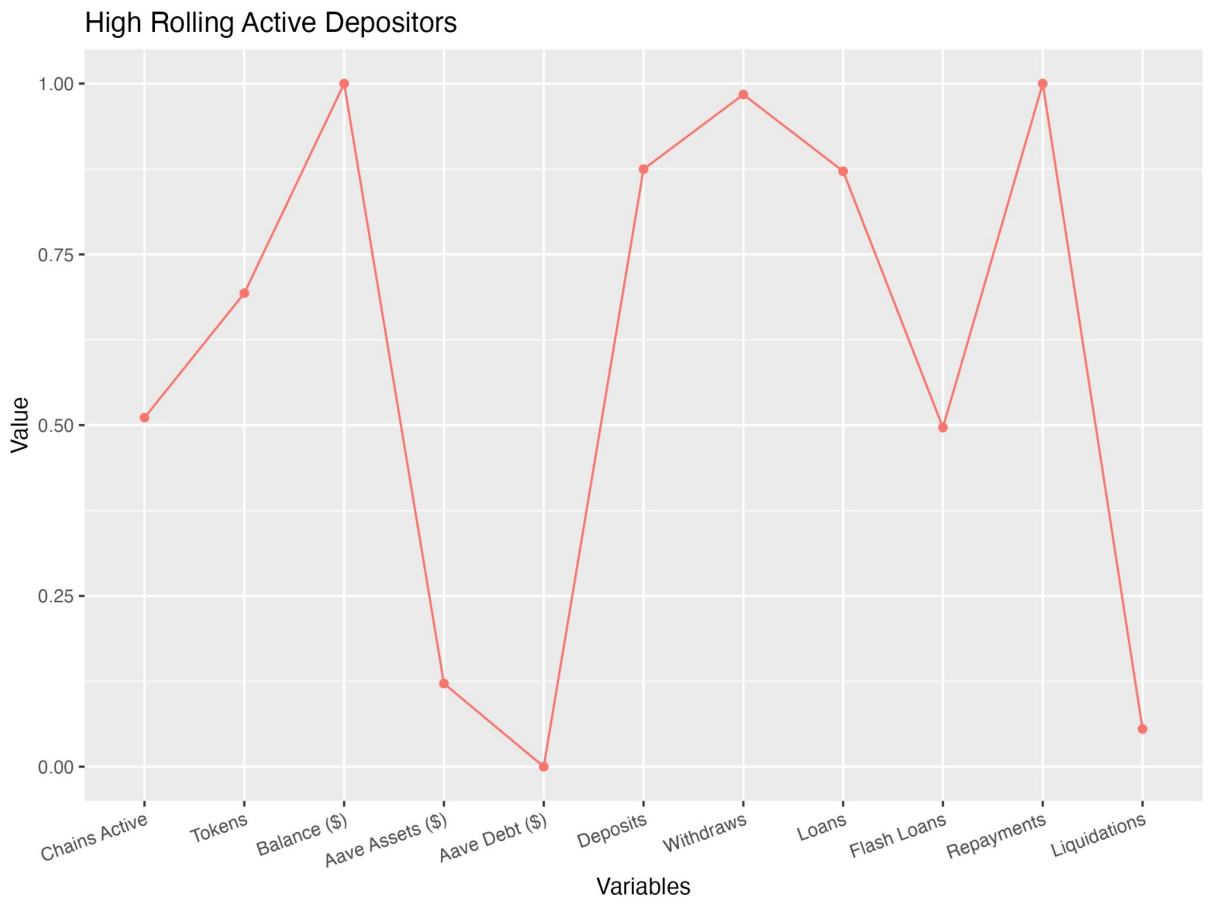
ETH Testers (9,321 | 8.03%) - Testers are those who use the platform once or twice then leave. They usually have one or two deposits that they subsequently withdraw. They usually don't borrow and have only tried the platform on ETH, as indicated by their low "Num_Chains_Active" value.

/// Aave Wallet Profiles



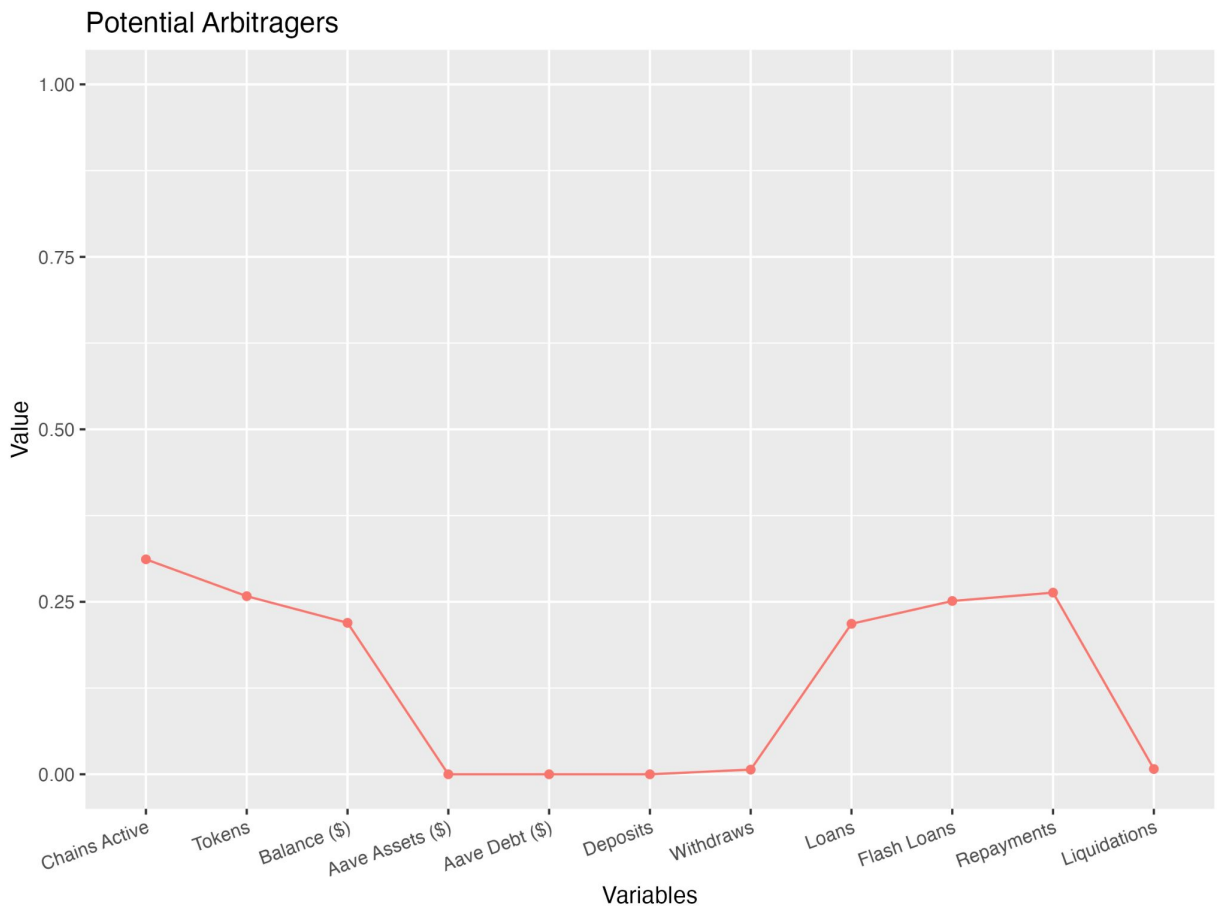
High Rollers AD w/Debt (7,016 | 6.11%) - These High rolling accounts possess a six figure Aave deposit value, on average, and while still maintaining a meaningful wallet balance. These accounts regularly make deposits, but fewer withdraws and choose to keep some outstanding debt that they often repay. Interestingly enough, those who regularly keep debt in this cluster are liquidated less frequently than the other high rolling profile that chooses to not use debt.

/// Aave Wallet Profiles



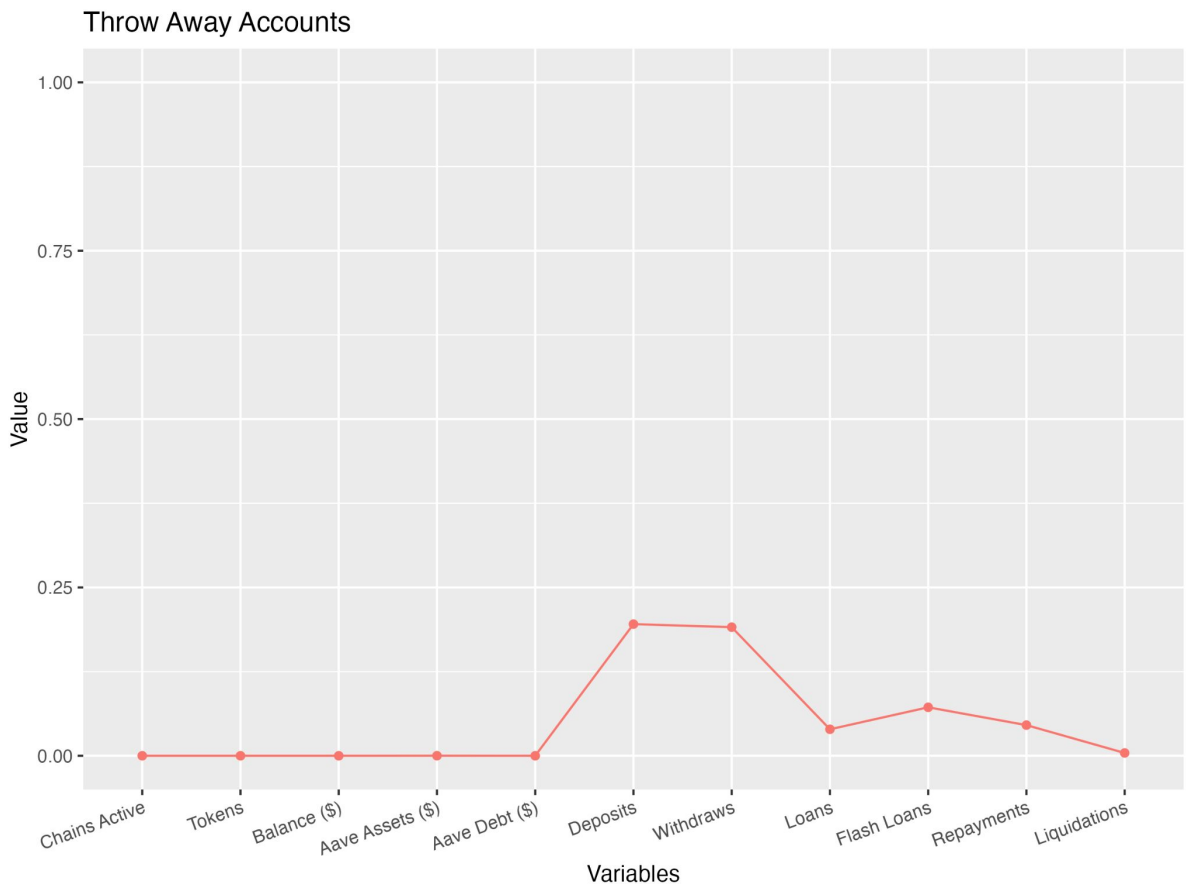
High Rollers AD (6,553 | 5.7%) - These High rolling accounts possess a six figure wallet balance, on average, and have some assets deposited in Aave. These accounts regularly make deposits and withdraws.

/// Aave Wallet Profiles



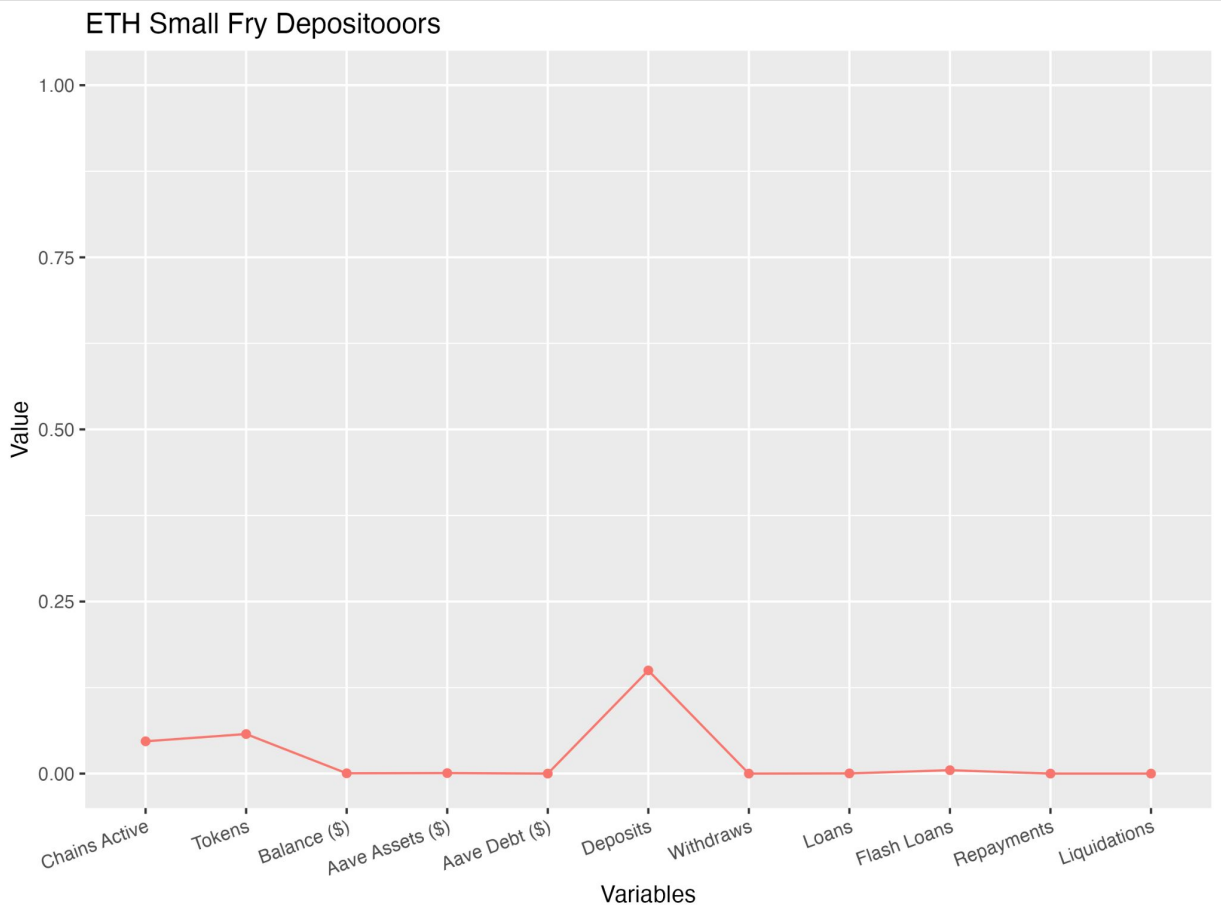
Potential Arbitragers (3,824 | 3.33%) - These savvy accounts are rarely liquidated but almost exclusively use flashloans and borrow without ever depositing. Despite that, they repay on time, then withdraw to keep their assets left on the platform very low.

/// Aave Wallet Profiles



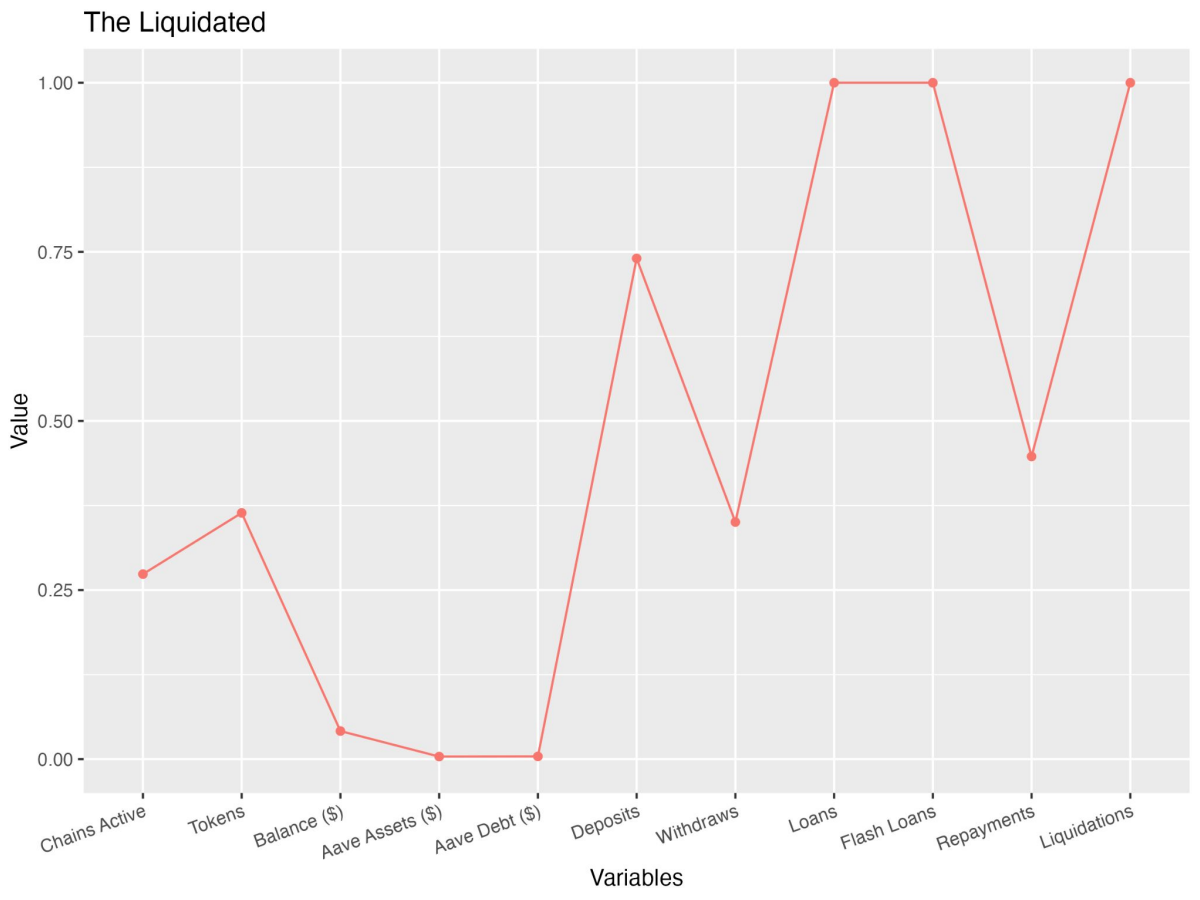
Throw Away Accounts (3,191 | 2.78%) - Similar to the "tester" profile, these are accounts that were used to interact with the contract once, and then were drained or discarded. Despite having a positive asset balance on average, most accounts in this category have $0 on the platform.

/// Aave Wallet Profiles



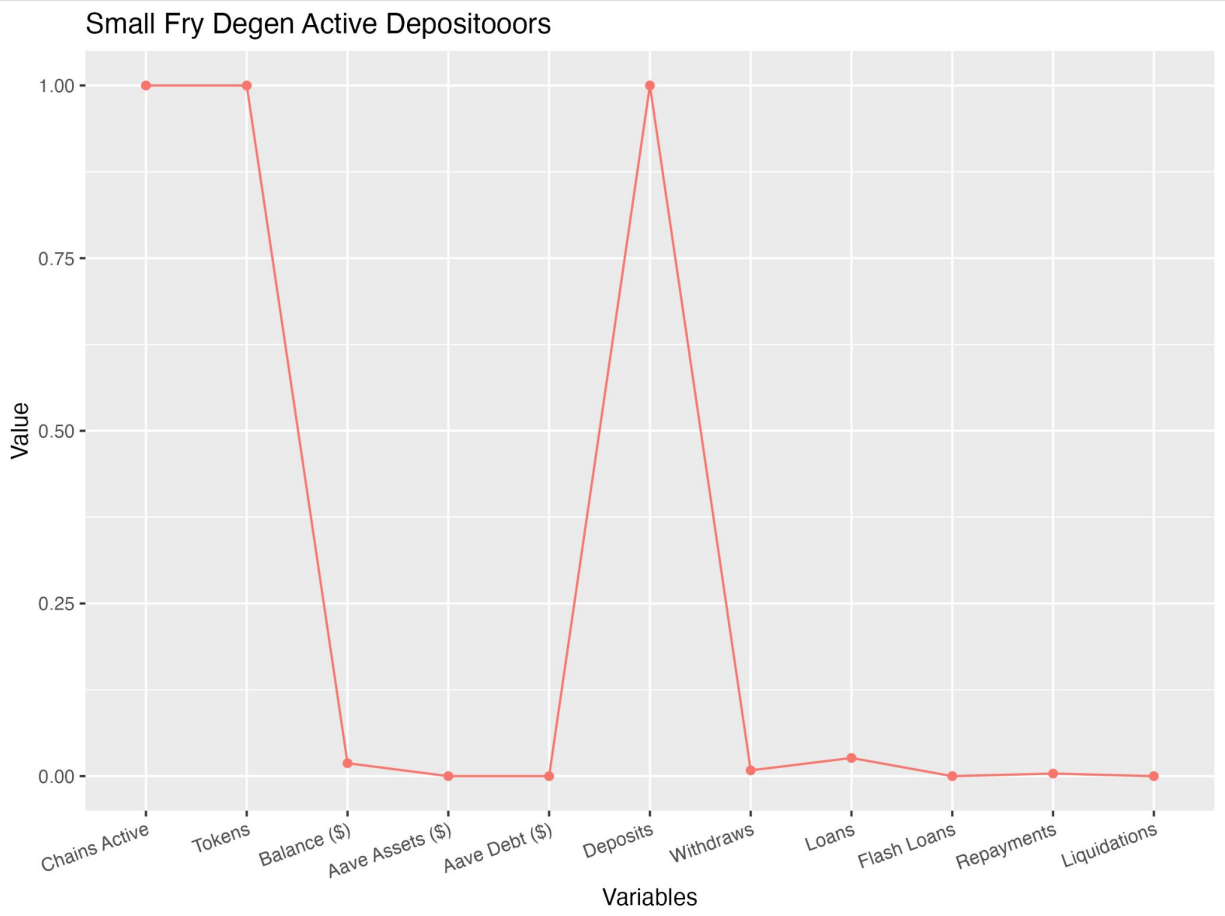
ETH Small Fry Depositooors (2,773 | 2.41%) - These small fries are trying their best to survive in Ethereum's high gas fee environment. They are active exclusively on Ethereum, hold only Ethereum, deposit once and just let it sit. The value of their deposits is often than their wallet balance, which suggests they are just trying to save and hope for the best.

/// Aave Wallet Profiles



The Liquidated (2,324 | 2.02%) - These are the active accounts that Supply, Lend, Borrow, use Flash loans, and generally do everything on the platform. Unfortunately they pay the cost for their experimentation because they are liquidated (or do the liquidating) at a higher rate than any other profile.

/// Aave Wallet Profiles



Small Fry Degen Active Depositooors (1,202 | 1.05%) - This group of small fry accounts are active on the most number of chains and hold a large number of tokens in addition to making the largest number of deposit calls to the Aave ethereum contracts. They rarely withdraw, but also don't have large deposits. Airdrop farmers?

/// Aave Wallet Profiles Supporting EDA

cluster	LensProfile	GitcoinPassport	TrustaLabSuspects	AtLeast1SnapshotVote	MeanSnapshotVotes	Mean(trustalabs_score)	Mean(agedays)
Binance Exchange	0.00%	0.00%	0.00%	0.00%	•	•	1414.4704256
Contract	0.00%	0.00%	0.42%	0.13%	5.2777777778	80.626968966	827.33285843
ETH Small Frys Depositooors	0.00%	0.00%	0.58%	0.04%	1	69.6875	731.84101404
ETH Testers	0.01%	0.00%	0.77%	0.03%	1	64.594360302	882.89772745
High Rollers AD	12.60%	6.61%	7.00%	7.22%	25.344608879	73.809582506	1036.7058826
High Rollers AD w/Debt	3.63%	1.23%	3.42%	2.01%	12.156028369	70.815874249	762.3040087
Maker Vault Owner	0.00%	0.00%	0.00%	0.00%	•	•	1281.449056
Middle-Class HC-LS	6.03%	3.01%	2.51%	3.81%	20.481958763	72.14735525	946.21552585
Middle-Class LC-HS	3.96%	4.69%	2.15%	2.81%	25.387218045	75.829781153	833.11375394
Multi-Chain Testers	4.65%	2.64%	2.98%	1.07%	20.953488372	72.899043721	1042.26986
MultiSig	0.00%	0.00%	0.26%	1.03%	5.75	79.5	557.34036914
Potential Arbitragers	2.85%	1.52%	3.43%	0.52%	16.95	78.591147548	799.98477062
Small Fry Degen Active Depositoors	55.82%	63.73%	19.80%	39.35%	37.179704017	75.537416926	816.97390059
Small Fry Degen Debt Users	58.75%	44.26%	23.72%	34.37%	34.417004049	81.114985828	683.94482931
The Good Guys	8.26%	5.44%	5.01%	3.21%	24.695798319	73.845126646	1000.1433297
The Liquidated	2.97%	1.03%	3.06%	1.08%	6.36	69.71455432	932.70157078
Throw Away Accounts	0.03%	0.00%	0.72%	0.00%	•	69.89828947	876.02434053

Figure: To further understand the differences in the groups, we created a table of proportions, counts and values comparing the profiles against a set of exogenous variables collected that were not included in the segmentation analysis. The cells within this table correspond to a numerical value for a specific variable or its proportion within that group. For example, 55.8% of "Small Fry DADs" had a Lens profile.

/// Aave Wallet Profiles Supporting EDA

cluster	LensProfile	GitcoinPassport	TrustaLabSuspects	AtLeast1SnapshotVote	MeanSnapshotVotes	Mean(trustalabs_score)	Mean(agedays)
Binance Exchange	0.00%	0.00%	0.00%	0.00%	•	•	1414.4704256
Contract	0.00%	0.00%	0.42%	0.13%	5.2777777778	80.626968966	827.33285843
ETH Small Frys Depositooors	0.00%	0.00%	0.58%	0.04%	1	69.6875	731.84101404
ETH Testers	0.01%	0.00%	0.77%	0.03%	1	64.594360302	882.89772745
High Rollers AD	12.60%	6.61%	7.00%	7.22%	25.344608879	73.809582506	1036.7058826
High Rollers AD w/Debt	3.63%	1.23%	3.42%	2.01%	12.156028369	70.815874249	762.3040087
Maker Vault Owner	0.00%	0.00%	0.00%	0.00%	•	•	1281.449056
Middle-Class HC-LS	6.03%	3.01%	2.51%	3.81%	20.481958763	72.14735525	946.21552585
Middle-Class LC-HS	3.96%	4.69%	2.15%	2.81%	25.387218045	75.829781153	833.11375394
Multi-Chain Testers	4.65%	2.64%	2.98%	1.07%	20.953488372	72.899043721	1042.26986
MultiSig	0.00%	0.00%	0.26%	1.03%	5.75	79.5	557.34036914
Potential Arbitragers	2.85%	1.52%	3.43%	0.52%	16.95	78.591147548	799.98477062
Small Fry Degen Active Depositoors	55.82%	63.73%	19.80%	39.35%	37.179704017	75.537416926	816.97390059
Small Fry Degen Debt Users	58.75%	44.26%	23.72%	34.37%	34.417004049	81.114985828	683.94482931
The Good Guys	8.26%	5.44%	5.01%	3.21%	24.695798319	73.845126646	1000.1433297
The Liquidated	2.97%	1.03%	3.06%	1.08%	6.36	69.71455432	932.70157078
Throw Away Accounts	0.03%	0.00%	0.72%	0.00%	•	69.89828947	876.02434053

These represent how active the different profiles are across various platforms, how committed to Web3 governance and also how long they've been active on the Ethereum blockchain. Lens is a social media platform, currently invite only, that allows users to participate in a Web3 designed Twitter like experience. At the time of this writing, a substantial percentage of both "Small Fry" profiles had Lens profiles with the third highest group being the "High Rollers AD" with only 12.6% of their members having a profile. We have a similar result for the proof of personhood protocol, Gitcoin Passport where the same set of profiles have the most interaction with the platform. Curiously, these same set of profiles have the highest proportion of their members flagged by the Trustalabs anti-sybil detection platform, though the numbers are much lower as a percentage of the population. Here, at the highest, 24% of the "Small Fry Degen Debt Users" have been flagged as having been involved in sybil-like behavior. This shouldn't be too surprising because their profiles are similar to those of airdrop farmers who spread small amounts of activity across many different protocols and chains, where some of which do so in an automated manner. Activity within the governance platform, Snapshot, tells a similar story where the same three profiles are again the most active. It is only when we look at the last variable, the age of the wallet, do we see a break in this pattern. The "age" of a wallet in this context has been taken to be a measurement of time from its first transaction until this writing. Outliers aside, the "High Rollers AD" are the oldest group with the average age of a wallet in this group being approximately 2 years and 10 months old. The next oldest group is only a month younger and consists of the "Good Guys", the group who regularly borrow and pay back their loans. An interesting fact happens to be that the youngest profile actually are the hyper-active "Small Fry Degen Debt Users", who are active across multiple chains, try out all the functionality of the platform all while doing so with limited resources. They are on average have only been around since the 2021, being approximately 1 year and 10 months old.

/// Extensions

What can we do now!?

- Think about targeted marketing and curated product development.
- Create a customized website experience based on the wallet profile.
- Extend the analysis beyond just the Ethereum mainnet chain.
- Perform more advanced EDA based on the profiles (Time to Default analysis or Customer Journey Path analysis).
- Analyze the interconnection between Web3 platforms from the Aave perspective (network plot of shared protocols) to identify potential partnerships.
- Use an autoencoder to improve the results (potentially).

<The End>

Closing out!

<Appendix>

/// Resources

Resources

1. https://www.aaveql.org/
2. https://docs.aave.com/developers/v/1.0/developing-on-aave/the-protocol/lendingpool
3. https://docs.aave.com/developers/v/2.0/the-core-protocol/lendingpool
4. https://docs.aave.com/developers/deployed-contracts/v3-mainnet/ethereum-mainnet
5. www.aaveql.com
6. Contracts of interest on Ethereum:
    a. https://etherscan.io/address/0x398eC7346DcD622eDc5ae82352F02bE94C62d119
    b. https://etherscan.io/address/0x7d2768de32b0b80b7a3454c06bdac94a69ddc7a9
    c. https://etherscan.io/address/0x87870Bca3F3fD6335C3F4ce8392D69350B4fA4E2
7. https://stats.stackexchange.com/questions/263539/clustering-on-the-output-of-t-sne

/// Appendix


Value
Binance Exchange
Contract
ETH Small Frys Depositoors
ETH Testers
High Rollers AD
High Rollers AD w/Debt
Maker Vault Owner
Middle-Class HC-LS
Middle-Class LC-HS
Multi-Chain Testers
MultiSig
Potential Arbitragers
Small Fry Degen Active Depositoors
Small Fry Degen Debt Users
The Good Guys
The Liquidated
Throw Away Accounts