to Jesus

Contents

Chapter 1

Introduction

Decentralized Finance (DeFi), a financial system based on blockchain technology, has recently gained popularity worldwide, leading to a rocketing adoption by existent financial players and technology enthusiasts. The interest in the crypto economy has exponentially increased, especially with the rise of Turing complete protocols such as Ethereum, leading to a total value locked (TVL) above 180bn[1] registered at the end of 2021 across different networks and markets. In only five years, from 2018 to 2022, the total number of unique DeFi users has raised from below one thousand addresses to nearly five millions[2] as of today, in September 2022, making this technology one of the most promising of the current decade.

Consequently, when a new technology makes its rise, there is an old system that is threatened. According to a research report from Deloitte [1], DeFi has the potential to disrupt and transform traditional finance, to which we will refer as Centralized Finance (CeFi) or Traditional Finance (TradFi), forcing financial institutions to redefine their business strategies, embracing the new rising economy of digital assets. However, according to a recent survey conducted by Deloitte [2] in June of 2022, DeFi is perceived more as an opportunity, rather than a threat, for existent financial players seeking to maximize their profits, grow and innovate their product line.

Given the fast adoption of distributed ledger technologies, the increasing popularity of

[1] `https://defillama.com`
[2] `https://dune.com/rchen8/defi-users-over-time`

cryptocurrencies and dApps (decentralized applications), altogether with the enormous and increasing amount of money involved, traditional financial institutions have started to lend an ear and invest in blockchain-related solutions, as reported in the Forbes Blockchain 50 collection [3], that studies the involvement of more than 100 big companies, with a minimum valuation of $1 billion, in blockchain-related solution. In particular, according to this report, well-known banks such as BBVA, Citigroup, ING and Credit Suisse have already invested in several blockchain ventures or launched pilot projects. In addition, according to a recent survey of PwC [4], among six hundred executives, 84% confirmed that blockchain-related projects had been founded within their companies, suggesting the fear of missing out on potential profits and market share is palpable. Interestingly, according to a research report from Blockdata [5], institutional DeFi could grow into a market worth 1 trillion dollars by 2025 if the largest financial institutions would invest 1% of their AUM (asset under management).

One of the core aspects of finance, both centralized or decentralized, refers to the practices of lending and borrowing assets. In its essence, the financial world comprises agents that need finances and agents that have a surplus instead. Through the power of innovative distributed ledger technologies (DLTs) and smart contracts development, a well-known DeFi primitive that will be later extensively illustrated, lending and borrowing have never been this simple compared to the ordinary system in TradFi. For this reason, one of the most notable and successful DeFi applications are protocols for loanable funds (PLFs), which enable every user to supply liquidity at attractive annual percentages yields (APY), namely interest rates over a year, or to obtain a loan in a fast, anonymized but transparent manner, revolutionizing the financial world. [6]

Given the public availability of the blockchain, where every single action is recorded and accessible for inspection, it can be observed that the number of borrowing and lending transactions has skyrocketed. Focusing just on the Ethereum blockchain, the total number of borrowings performed in the main PLFs (Aave, MakerDAO and Compound) increased from 44 thousand in 2019 to reach a peak after just a year of 507 thousand, an increase of almost 1200%[3]. In crypto trading, borrowings are mainly used

[3]https://dune.com/queries/1012049

to augment buying power by engaging in margin trading, also commonly referred to as leverage, which uses borrowed assets to perform trades. [7] However, leverage, which is also a financial product commonly adopted in centralized finance, is inherently risky in the digital assets world due to the well-known volatility of the assets involved. In this regard, crypto margin trading leads to the diffusion of over-collateralized loans, namely, as the word suggests, defined as borrowing positions where the monetary value of the deposited asset, the collateral, is larger than the value of the borrowed one.

Commonly in leverage trading, the borrowing position is deemed at risk of liquidation whenever the underlying collateral asset decreases in value under a certain threshold. Liquidations, when a third user redeems the debt of the risky position and collects the collateral, are performed differently depending on whether the position was opened in a centralized or decentralized context. Specifically, in CeFi, liquidations, which will be deeply analyzed in section 5.1, are mainly negotiations where courts, directors, shareholders, individuals, accountants, borrowers or liquidators can be possibly included, depending on the nature of the borrower: if a company or a physical person. On the other hand, in DeFi, the deemed risky positions get cleared by automated, transparent and publicly available smart contracts, a coded set of rules according to predetermined conditions unique to each PLF, independently of the nature of the users. More accurately, by directly pulling out data from the Ethereum blockchain and inspecting the main PLFs, namely Compound, AAve and MakerDAO, it can be noticed that liquidations have repaid risky borrowing positions aggregating for more than 1 billion dollars just in 2021[4]. This demonstrates the fundamental role liquidators play in the DeFi leverage trading world, acting in the background and providing stability and equilibrium to the system by clearing unhealthy borrowing positions.

In parallel, while the blockchain-powered world is emerging, the data-powered world is consolidating with widespread adoption and innovation of machine learning techniques. Data science is not anymore a myth, but it is a well-known and established reality according to an article in Harvard Business Review [8], described as a continuously evolving field, projected to grow until 2029. With this new awareness by companies

[4] `https://dune.com/queries/1110658`

in judging the power of data, an increasing amount of information gets collected and analyzed daily, leading to more accurate, faster and useful predictions. Furthermore, given that the blockchain is, by construction, constantly producing publicly available new information, mostly misunderstood but more accessible every day to get, the union between these two realms is immediate.

This thesis aims to show how liquidations in decentralized finance are affected by the choice of different protocols for loanable funds (PLF) by firstly providing a qualitative assessment of the state of these protocols by directly querying data from their smart contracts. This qualitative analysis is augmented with a quantitative one, through an inferential approach, by studying the impact of the choice of a borrowing platform with a similar liquidation mechanism on the probability of being liquidated. Additionally, it provides a predictive approach for estimating the likelihood through which an address with unique characteristics gets liquidated. Before entering the details of the results obtained on actual data, an introduction to the basic concept of blockchain, smart contracts and decentralized finance is provided to the reader. Additionally, a comprehensive illustration of the chosen PLFs available on the Ethereum blockchain is given with related structural differences, backed-up by updated dashboards directly querying the underlying blockchain itself, with primary emphasis on the liquidation phenomena.

The fundamental concepts of distributed ledger technology and smart contract are described in chapter 2. Closely related to it, decentralized finance and its main primitives are introduced in chapter 3. In chapter 4, a comprehensive illustration of borrowing and lending practices in DeFi, along with the description and comparison between the leading protocols for loanable funds(PLFs) on the Ethereum blockchain and related differences, is given to the reader. Then, liquidations are introduced, and the different ways of performing it per protocol, altogether with summarized data and an introduction to flash loans, are described in chapter 5. Consequently, chapter 6 introduces how data have been collected and reshaped into a relational dataset.

Moreover, it illustrates the inference model for understanding the effect of the choice of different PLFs on the probability of being liquidated. Finally, it provides a collection of prediction models for forecasting the probability for users with open borrowing positions to get liquidated. Lastly, chapter 7 discusses the future and actual steps related to the development of liquidation practices performed on different blockchains rather than Ethereum, discussing limitations and further extensions of the study delivered by this paper. Finally, the last chapter concludes this thesis by summarizing the main results and achievements obtained.

Chapter 2

Theoretical Background

The following section introduces the reader to the basic concepts of the blockchain realm, smart contract ideas and statistical learning necessary to understand the rest of the paper.

2.1 Bitcoin: History

On October 31st, 2008, Satoshi Nakamoto published a short whitepaper [9] referring to a technology called Bitcoin on a cryptographic mailing list: this was the beginning of the decentralized revolution. In particular, this paper signifies the starting point of a radical change against all third-party guarantors in numerous markets, such as the financial and insurance ones. Bitcoin, and in general, most of the blockchains, are essentially a database, which differs from the traditional ones. It is governed by a set of rules called consensus protocol that governs how data is added to this ledger in a decentralized manner. Unlike traditional databases, the blockchain data structure is defined as append-only [10], which means that new data can only be appended to the end of the chain, but the already present data are immutable. It needs to be immediately stated that the blockchain is immutable according to a probabilistic guarantee which depends on the consensus protocol governing the blockchain, which will be later explained in this chapter.

In order to comprehend how the Bitcoin blockchain works, its design peculiarities and what enormous innovations it has brought to our world, it is crucial to understand why Bitcoin was created in the first place and, in general, the reasons and needs behind the shift toward a decentralized digital economy. Focusing on the financial realm, the first appearance of digital money was in 1946 when John Biggins, a banker in Brooklyn, released the prototype of a credit card, the so-called “Charg-it”. Until then, ownership of financial assets has always been based on physical assets (either commodity money, minted coins or fiat money) and transfer of ownership through an intermediary such as a bank, acting as a trustworthy third party, or through physical transfer between the two agents. With the development of digital money, ownership of assets became merely an entry into a digital database, which is legally entrusted by centralized institutions such as national governments and financial intermediaries. Following the Bitcoin paper, in the digital world, the data itself should guarantee the existence and ownership or transfer of ownership of an asset instead of relying entirely on financial institutions assisting as trusted third parties and encouraging the transition from a trust-based model to a cryptographic based one, enabling secure digital payments without the need of an intermediary.

Bitcoin represents the natural evolution of money. Its decentralized nature eliminates any single point of failure, as the financial intermediary is still an institution that could potentially fail or misbehave. Bitcoin was the milestone, the genesis, of a more significant project with a broader scope than the financial one: distributed ledger technologies (DLT) with numerous applications across markets and industries. The umbrella term DLT usually generically refers to the append-only data structures (known as blockchain), particularly the protocols that enable the secure functioning of it [10]. The following sections generally discuss the technical elements that constitute a blockchain, with later more emphasis on Ethereum.

2.2 Blockchain Systems

As previously introduced, a blockchain, such as Bitcoin, Litecoin, Ethereum, Dash and many others, is a decentralized append-only data structure. The term blockchain refers to the fact that transactions are first assembled in blocks and then added to the end of a time-ordered and immutable chain of blocks, in a decentralized manner, therefore without a centralized entity and according to a predetermined set of rules unique to each consensus protocol. The main ingredients for a blockchain to function and be secure, decentralized, immutable and public are listed below, and an introductive explanation is given to the reader.

2.2.1 Peer-to-Peer (PLP) Network

The blockchain achieves decentralization of data and eliminates any single point of failure thanks to the public peer-to-peer architecture that maintains it, which partitions jobs and workloads between different digital nodes [11]. In particular, there is no single server connected to many clients; in this network, every node is simultaneously a server for the other nodes and a client for the other servers. Every node needs to maintain an updated version (a copy) of the current state of the system so that if one of the servers is shut down, the whole system remains intact, making the network and the public ledger (the blockchain) safer in terms of data availability compared to a centralized option. Moreover, the majority of blockchains systems not only achieve decentralization of data through a P2P network but also the decentralization of the source code, given that its software and the code behind their technologies are open source, so it can be audited or improved by its users (not only the nodes' owners), making the whole system more transparent and prone to innovate. [12]

2.2.2 Hash Functions

One of the underlying features that make the blockchain function is represented by using cryptographic hash functions. Hash functions are mathematical transformations that map input of any size into the output of fixed size and are efficiently computable. Moreover, hash functions are one-way, meaning it is statistically impracticable to find the inverse of such a function where the average case behaviour can be nearly optimal (minimal collision). [13] Hash functions are used in the blockchain as hash pointers between blocks and serve as a way to verify that the information inside each block has not changed, therefore ensuring integrity. In particular, a blockchain is a variation of a linked list data structure where each block has a hash pointer to the previous block. Consequently, hash functions represent a fundamental part of the digital signature scheme described in the next paragraph.

2.2.3 Digital Signature Scheme

The decentralization introduced by distributed ledger technologies does not stop only in public availability but also in the decentralization of asset ownership. In the blockchain, as there is no authorized central third party, decentralization also refers to asset ownership, particularly to who is authorized to change the data on the blockchain. Blockchains such as Bitcoin and Ethereum are based on a digital signature scheme, that is, a pair of public and private keys, which enables everybody to independently check if a user has the authority to propose a change or not. A digital signature is an analogue of a handwritten signature that is unforgeable, which means it is computationally infeasible to replicate. It is tied to a particular document, therefore, cannot be reused or copied. Specifically, whenever an agent wants to sign or enable a particular action, for example, a transfer, the private key will be used to sign it. The transaction is broadcasted to the network, whereas the public key is publicly available. The nodes that are listening to the peer-to-peer network can verify the ownership of this transaction by using the agent's disclosed public key: this is an example of how asymmetric encryption works, a foundational pillar of distributed ledger technologies. [14]

2.2.4 Consensus Protocol: Proof of Work

As it was discussed before, the way a block gets added to the chain in DLT in a decentralized manner is established through a distributed consensus protocol. It has been previously explained that a blockchain is based on a peer-to-peer network composed of different nodes that keep a copy of the blockchain and are responsible for adding new blocks. These nodes are "listening" in the open network (called mempool) for transactions proposed by other users. Then after grouping these transactions into a candidate block, the node proposes it to the other nodes, thus requesting validation of these transactions. The nodes must agree on which transactions were broadcasted and the order in which they happened. When the nodes agree on the next block by confirming it, that is, adding it to their ledger copy, the winning block will be added to the public chain after a specific confirmation and time.

More clearly, all the nodes have reached a consensus on the current blockchain and are all listening for different transactions happening in the network. This means that each node could propose a different block that needs to be validated. Achieving consensus between nodes on the state of the blockchain and especially on which is the next block to be included is a hard task: because nodes may crash, be malicious, or simply due to the latency in the network. This is usually an impossibility result in distributed computing, specifically the byzantine generals' problem. [15]

According to the official Bitcoin paper, Satoshi Nakamoto addresses this problem by proposing a new form of consensus among nodes named proof of work (PoW). Proof of work is a rather simple concept very much criticized for its environmental consequences. It creates competition between nodes in solving a cryptographic or hash puzzle by consuming computing (CPU) power. [16] The first node that solves the puzzle is enabled to propose the next block to the network. Consequently, all the other nodes validate the transaction inside and include it in their current copy of the chain, validating the transactions. The proof of work consensus protocol is based on economic incentives that will be described in the next section, and it is statistically safe, as described by the founder of Bitcoin, given that it requires the majority of the nodes that retain more than 50% of CPU power to behave in an honest way [9]. Considering the Bitcoin blockchain, the average time between two consecutive blocks to be added to the chain is around 10 minutes, whereas in the Ethereum blockchain is between 10 to 19 seconds, making the block time of Ethereum faster than the one of Bitcoin. [17]

2.2.5 Mining and Financial Incentives

In the blockchain jargon, nodes are often called miners, solving a hash puzzle and mining the next block of the public chain. Miners do not engage in such activities without being compensated for their job. Miners get rewarded when they successfully include their candidate blocks in the public chain. Moreover, miners' fees encourage the decentralization of incentives to inclusion, in particular given the economic aspect at stake, making the inclusion of transactions in candidate blocks a real auction, thus not based on miners' preferences. Each user that wants its transaction to be included in the system will add an arbitrary sum that will incentivize a miner to include this transaction in its block. Miner's fees can be augmented by a block prize, such as in Bitcoin or depend on the congestion of the blockchain, that is, the amount and willingness to pay of the users to have their transaction included in the chain, such as in Ethereum. According to the Ethereum whitepaper, the miner's fee is usually referred to as gas and varies according to the number of users submitting transaction requests. [17]

2.2.6 Ethereum

The Bitcoin debut presented many novelties, revolutionizing the digital economy with the introduction of decentralized assets ownership and transfer, the native Bitcoin (BTC) token, characterized by uncensorable access, as anybody could create a public/private key pair and anonymously transfer value. Moreover, as described by the block's prizes scheme, its transparent monetary policy made this cryptocurrency the most famous still in 2022, with the current highest market cap as of September 2022[1]. However, Bitcoin's starting idea soon became overrated as this system was slow. As explained before, it takes ten minutes to validate a transaction, and most importantly is not Turing complete as explained in the Ethereum whitepaper. The scripting language of Bitcoin does not allow loops. Moreover, the Bitcoin blockchain transactions lack a state which means that either the existing amount is spent entirely or not, making it impossible to create multistage contract logic. Consequently, in 2014, building on the original

[1] https://www.coingecko.com/en/coins/bitcoin

idea of Bitcoin, the Ethereum blockchain was released with some essential differences from Bitcoin. In particular, Ethereum was built with a Turing-complete programming language that would allow anybody to build applications and smart contracts. The Ethereum blockchain, with its native asset, the ether ETH, has an account-based model compared to the transaction one of Bitcoin. In particular, state information in Ethereum is stored in accounts that can either be externally owned, controlled by a user with its private key, or smart contracts governed by code. An account in Ethereum is usually identified through its address, a public key hash. Moreover, each account publicly displays its current ether balance, storage and code presence if it is a smart contract. [18] The next section will be dedicated to introducing smart contracts in Ethereum and its most famous use cases. This will be later linked to the explanation of Decentralized Finance (DeFi), described in chapter 3.

2.3 Smart Contracts

This section describes smart contracts, which are building blocks of publicly available blockchains such as Ethereum or Avalanche[2]. These contracts are controlled by code, not by a private key or external agent. Once a smart contract is deployed, it cannot be changed, and its code is immutable. The following sections describe Ethereum smart contracts in more in-depth and provide critical smart contracts applications fundamental for understanding the core concept of this paper.

2.3.1 Ethereum Smart Contracts

As previously introduced, a smart contract is controlled by code, not by an external user. Each smart contract account has an address and publicly stores the code describing the functions invoked by EOA accounts or even by other smart contract accounts. Moreover, a smart contract, like any EOA, can receive funds and transactions that are method invocations. Usually, smart contracts in Ethereum are written in high-level languages such as Solidity or Vyper, with the intent to enable everybody to understand what

[2] https://www.avax.network

the contract does, rendering its function transparent and trustworthy. These smart contracts are deployed on the Ethereum Virtual Machine (EVM), a stack-based VM that executes the smart contract's bytecode compiled to the EVM. [18] As introduced before, each transaction that changes the state of the Ethereum blockchain consumes a miner's fee, called gas, measured in ETH. Gas is a vital element that, according to the Ethereum whitepaper, enables smart contracts to terminate their operation whenever the gas limit is reached, allowing loopings and potentially infinite programs to terminate. [19]

2.3.2 Tokens

A typical application of smart contacts is tokens, which are used to represent crypto assets such as ETH or AVAX, synthetic assets and derivatives, or even governance tokens that give the right to vote, for example, in decentralized autonomous organizations (DAO). In general, the underlying contract code of tokens follows a standard interface to facilitate the interaction with protocols. In particular, the most adopted tokens can be classified into fungible and non-fungible (most known through the buzzword NFT), even if other token typologies are present. Fungible tokens in Ethereum are usually coded according to the ERC-20 interface, whereas standard NFTs follow the ERC-721. The difference between these two can be easily explained through an example: fungible tokens are like EURO coins. There is no difference between the two EURO. They can be easily exchanged. Instead, non-fungible tokens are unique; a typical example could be a stadium ticket or a work of art, unique assets that cannot be reproduced or counterfeit. [20]

2.3.3 Oracles

Smart contracts often need to access information not present on-chain, only externally available data. These systems that provide this connection to the external environment are smart contracts called oracles. An example could be an oracle for a financial derivative smart contract that needs to retrieve the value of its underlying stock or a prediction market smart contract that needs to know the result of a specific event to determine the success or failure of a bet. Oracles[3] are fundamental in decentralized finance, as it will be described in chapter 3, as they are commonly used to retrieve the real-time value of assets.

2.3.4 Stablecoins

Another example of a smart contract is a stablecoin, which, as the name suggests, are stable cryptocurrencies, which are stable in terms of value compared to volatile ones, such as ETH or BTC. Stablecoins are pegged to a currency, in general, the USD, and they aim to solve the volatility issue that prevents the use of cryptocurrencies for everyday payment purposes. Stablecoins can be of two different types: custodial and non-custodial. Custodial stablecoins, such as USDT or USDC, are asset-backed off-chain and rely on a trusted third (the custodial) party to hold the collateral. On the other hand, non-custodial stablecoins, such as DAI, are asset-backed on-chain which means that other cryptocurrencies back them. The DAI stablecoin is at the core of the MakerDAO protocol for loanable funds(PLF), which will be thoroughly analyzed in chapter 4 and chapter 5. Another type of stablecoin is algorithmic such as Terra (Luna), now rebranded as Terra 2.0, which relies on a combination of smart contracts to maintain the price equilibrium. However, trust in such stablecoins has dramatically dropped after the failure of Luna in May 2022. [21]

2.4 Statistical Learning Background

Switching from the blockchain realm to the data science one, statistical learning refers to the disciplinary field that studies the mathematical relationship between a set of inputs and an output (or a set of outputs). In particular, suppose, in the simplest case, that we observe a quantitative response Y and a set of regressors (or predictors) d, $X_1, \ldots, X_d$, then let us assume that there is a relation between the response and its

[3]Compound: Open price feed (2022). `https://compound.finance/prices`

predictors in the form of:

$$Y = f(X_1, ..., X_p) + e \tag{2.1}$$

Where f is some fixed but unknown function of the predictors, and e is the error term, assumed to be independent of $X_1, \ldots, X_d$ and centred around zero.

In this context, statistical learning focuses on estimating f to either understand or replicate the relationship between the set of predictors and the response Y. In turn, while studying the same mathematical object f, inference and forecasting have two different aims. On the one hand, forecasting is often seen as a black-box problem. In particular, it does not prioritize the interpretation of every single predictor. However, it only tries to provide a closer estimate $\hat{f}$ of f such that the predictive power is maximized, monitoring the out-of-sample error term. In prediction, it is not essential to know the exact form of f as model interpretability is not the scope of the analysis.

On the other hand, knowing the exact form of f is crucial in inference. This realm does not focus on predicting Y given X but on understanding the relationship between every single predictor and the response, mainly how Y changes as a function of these regressors. [22]

In chapter 6, as explained in the introduction, inference will have the scope of measuring the role of a protocol in the context of liquidations by estimating the marginal effect on the probability of a user being liquidated. On the other hand, in parallel with inference, a predictive approach is proposed. In particular, we are devising a statistical technique that predicts the probability of being liquidated for every single address that has opened a borrowing position.

Chapter 3

Decentralized Finance

This chapter introduces the main concepts related to decentralized finance (DeFi), providing the necessary information to understand the rest of the paper. Starting with a brief comparison with the centralized finance model (CeFi) by discussing the advantages and benefits of DeFi, this chapter focuses on listing the major DeFi protocols. It gives an intuition on the magnitude of credit protocols compared to other protocols' typologies. The rest of the chapter introduces flash loans, a novel DeFi-specific lending mechanism widely used in liquidations, further analyzed in chapter 5.

3.1 Blockchain and DeFi

As introduced in chapter 2, a new ecosystem started to emerge with the development of Ethereum and the adoption of smart contracts. In particular, users were given the freedom to develop on top of blockchains whatever decentralized application they wanted by just submitting a transaction, in the case of Ethereum, and deploying an autonomous smart contract account. Thus, Decentralized Finance (DeFi) was beginning to take shape, with multiple blockchain companies that started to emerge and populate an environment with dApps (decentralized applications), to the point of achieving a total value locked exceeding 100 Billion in 2021[1].

Decentralized finance presents a non-custodial financial model, which means that

[1] `https://www.defipulse.com`

owners of private keys have all the rights and control over their funds at any moment. On the contrary, Centralized Finance presents a custodial model, meaning that users have partial control over their assets in exchange for security.

Due to the permissionless nature of the blockchain and DeFi world, anyone can interact with a smart contract independently of his personal information. This means that anybody is allowed to make use of decentralized protocols without the possibility of censorship or discrimination. On the other hand, centralized finance is very much based on personal interaction, where transparency is often the critical component through strict KYC practices.

3.2 DeFi vs CeFi

DeFi and CeFi as briefly explained before, are two faces of the same medal. At its essence, as DeFi follows a non-custodial model, it eliminates the need for a third trusted party when performing financial operations while making finance more accessible, transparent, uncensorable and less costly. Apart from replicating existing financial services already available in CeFi, DeFi introduces new financial products unique in its realms, such as flash loans and highly-leveraged financial assets and ad-hoc blockchain solutions. The lower intermediary transaction costs, greater transparency, easier accessibility and fast execution, altogether with extensive documentation and control over the financial products, make DeFi solutions more appealing, compared to the same services offered in CeFi, following a custodial model, characterized by silent and unwanted intermediary fees altogether with slow performance. [23]

Moreover, the financial opportunities that DeFi presents compared to CeFi are more appealing. While the APY (annual percentage yield) of USD in a centralized bank is around 0.01%, in DeFi at the time of writing, the APYs offered are beyond 3% on multiple crypto assets[2]. However, the risks connected to the decentralized nature of DeFi, the high volatility of crypto assets and the crash of essential players in DeFi, such as the previously stated failure of the stablecoin TerraUSD, keeps limiting the adoption

[2] `https://defirate.com/lend`

of DeFi solutions compared to CeFi.

3.3 DeFi Protocols

At their essence, DeFi protocols are a set of smart contracts that, by introducing innovative solutions, mirror classical financial services in a decentralized setting. According to a report from the ECB [24], DeFi protocols are categorized into seven classes by the type of operation they supply. These categories include assets, auxiliary, credit, insurance, payments, staking and trading. As an example, asset protocols include derivatives, synthetics, and options. In contrast, insurance protocols offer financial protection against unpredictable events in the decentralized environment, such as smart contracts bugs or hacker attacks. As this paper is focused on studying the liquidation phenomena on credit protocols, or named loanable funds markets for on-chain assets (PLFs), it is fundamental to understand the TVL of these protocols compared to the other DeFi protocols previously listed.

From Table 3.1, derived from Defillama[3], it can be noticed that the credit protocol category is second only to Dexes for combined TVL, accounting for 14.64 billion at the time of writing. However, only 181 credit protocols exist compared to the 577 in Dexes.

Table 3.1: TVL by category.

Category	Protocols	Combined TVL
Dexes	577	$23.9b
Lending	181	$14.64b
Bridge	30	$10.83b

3.4 Flash Loans

Flash loans are a novel and DeFi-specific financial lending instrument. Flash loans are fundamental components of liquidations that are usually even provided by the

[3] `https://defillama.com/categories`

same lending platform (PLF) where the liquidation is happening, such as Aave v2 and MakerDAO, which will be later discussed in detail in the last section of chapter 5.

Before jumping into the technicalities of flash loans, it is essential to remind the reader that blockchain transactions are atomic. [17] A state transaction that can include a collection of individual transactions is either executed entirely, which means all the individual transactions succeed together or fail jointly. Thus none of these transactions is considered valid.

Flash loans exploit the atomicity property of blockchain transactions by initiating and repaying a loan inside a single block. In particular, a user can set up a system; usually, a bot that borrows from a platform performs a set of actions with the borrowed asset and repay the loan, plus interest, in the same block. The beauty of flash loans is that if the loan cannot be repaid, then this state transaction is reverted, which means that this set of actions never happened, and the user only needs to pay the necessary fees to start a flash loan, plus gas fees. Flash loans are a potent tool widely used to exploit arbitrage opportunities and to perform liquidations. However, DeFi suffers from using flash loans to perform malicious attacks against DeFi protocols. They give users access to billions of USD without providing any collateral and by just paying small transaction fees, accounting for more than 100 million dollars in 2020. [25], [23]

Chapter 4

Borrowing and Lending in DeFi

This chapter extensively discusses one of the aforementioned applications in DeFi, focusing on how borrowing and lending are administered in a decentralized context on the constantly developing protocol for loanable funds (PLFs). It also introduces the necessary base concepts to understand how liquidations work in decentralized finance on the Ethereum blockchain, in-depth explained in the next chapter. An introduction to the DeFi lending markets by illustrating a comparison to traditional CeFi lending practices is provided in <section 4.1. The jargon used when describing PLFs and liquidations is described in section 4.2. In section 4.3, the selected most famous PLFs functioning on the Ethereum blockchain are briefly described and explained from a historical, technical and economic viewpoint. Finally, section 4.4 aims at showing the comparison between these protocols according to qualitative and quantitative metrics, gathering data directly from the Ethereum Blockchain.

4.1 From centralized to decentralized loans

In a financial context, centralized or decentralized, credit is defined as a financial contract between a borrower and a lender, where the former receives an amount of money from the latter, which has to be repaid at a later stage with the addition of predetermined interests. Historically, lending or borrowing has always represented essential financial services provided by public and private institutions, such as banks, credit unions, online lenders and even governmental entities. Interestingly, whenever an agent uses a credit card, he performs a borrowing operation, or analogously he has been given credit toward another agent, a sourcing one.

In a centralized context, as explained throughout the paper, particularly in chapter 3, given the presence of a middleman, defined as a third trusted party, such as a bank or a generic financial institution, borrowings and lending's success likelihood is affected depending on the estimated default risk, and on specific borrower's characteristics. In particular, in CeFi, the lender carries the uncertainty from the possibility that a borrower may default on its debt, failing to pay its loan either on time or not paying it entirely. The lender commonly alleviates this uncertainty by attentively scrutinizing and statistically forecasting the borrower's creditworthiness. The potential debtor must provide evidence and data on its credit score, credit report and payment history, as well as financial documents such as earning income and owned assets, in case of a personal loan, that could facilitate the loan's approval. In most cases, a borrower is also required to provide collateral, an asset with tangible financial worthiness, to the financial institution to be eligible for the desired loan. As it can be easily deduced, the success and speed of this financial operation are heavily influenced by the borrower's unique characteristics (also depending if it is an individual, a company or an institution) and internal lender's policies, making the lending and borrowing activity a slow, long and tedious process, especially when some required documents do not align with the lender's indications. This statement is supported by empirical data, with usual market conditions. According to the leading mortgage software company Ellie Mae, in the US, the average number of days (in 2019) for an agent to get a mortgage is around 47 days, considering all the different typologies of loans. [26] In addition, regarding personal loans instead, according to a study published by Forbes [27], the average time for such an operation to be finalized uniquely depends on the lender, where both approval and funding of the loan may vary between one and seven days, respectively.

On the contrary, in DeFi, the lack of a centralized figure, the absence of interference by any legal or governmental institution, as there is still a regulatory gap at the time of writing [24], and the anonymization of the users' information immensely facilitate the lending process. In this way, by avoiding discrimination based on personal or financial characteristics, such as the previously described measures to assess creditworthiness, and by mitigating the default risk through the introduction of over-collateralized loans, everybody has the opportunity to become a liquidity provider (lender) or a borrower in just a few clicks. Over-collateralized loans, as the word suggests, are techniques where the total monetary value of the provided collateral requested is higher than the total monetary value of the borrowed asset, put the collateral deposited is worth more than the loan obtained. Overcollateralization is not a novelty in the traditional financial world; it is a credit enhancement technique that aims at limiting the risk faced by the lender. It is commonly found in asset-backed securities such as mortgages or collateralized loan obligations (CLO).

However, considering the legal aspects, lending and borrowing in DeFi are not protected by any regulations or governmental institutions compared to CeFi. In centralized finance, for example, in Europe, deposit guarantee schemes (DGS) reimburse a confined amount if the bank goes bankrupt and act as a safeguard for the client. In DeFi, given the non-custodial model and absence of a central authority, effective supervision and regulation is still an opaque matter, a worrying gap that governments and institutions must, and probably will soon address. [24]

4.2 Background terminology of PLFs

The background terminology[1] in the context of PLFs is illustrated in this section to facilitate the understanding of the reader and to enable the subjective comparison of the different protocols.

Locked Funds: defined as the aggregate amount of tokens that a protocol's pool holds. Computed as the difference between the funds supplied with the ones borrowed.

Supplier: An agent that deposits funds into the protocol's pool.

[1] `https://compound.finance/docs/comptroller`

Borrower: An agent borrows funds from the protocol's pool. Given that, in most cases, borrowing implies the deposit of collateral, a borrower is a supplier first.

Liquidator: A third user that buys a borrower's open position whose health factor is below a certain threshold collects the collateral.

Reserve: Portion of borrowers paid interest from the protocol set aside as cash to mitigate default risks further or utilize the governance.

Reserve Factor: the percentage of borrowers paid interest that gets transferred in a pool of reserves.

Collateral : a crypto asset that must be deposited to guarantee a borrowing position.

Governance Token: usually an ERC-20 token that enables users to vote, propose changes (as the nature of the protocols is decentralized), actively participate and contribute to the development of the selected PLF.

Supply/ Mint: The act of supplying funds to a lending pool that can be borrowed by other agents or used to open collateralized debt positions.

Borrow: Getting funds from a lending pool.

Collateralization Ratio (Collateral-to-borrow ratio): the ratio between the collateral value and the loan value. Determines how much collateral needs to be provided to borrow the desired asset. If larger than 1, these are overcollateralized loans.

Collateral Factor: this is defined as the inverse of collateralization ratio. It determines the fraction of a deposit that can be used as collateral.

Close Factor: related to liquidations, the percentage for a position deemed liquidatable that can be repaid in a single transaction.

Liquidation Threshold: threshold for the value of the deposited collateral to flag a position to be undercollateralized and, therefore, could be liquidated.

Health Factor: proxy of how safe a borrower's asset is against a liquidation scenario. If the health factor is smaller than one, the opened position is subject to liquidation. The equation for the health factor is defined as follows:

$$HF = \frac{\sum ValueofCollateral_i \times LT_i}{\sum ValueOfDebt_i} \tag{4.1}$$

Liquidation Fee: fee that works as an incentive for agents to become liquidators to stabilize the protocol lending pool and clean risky borrowing positions.

Exchange Rate: conversion rate between the deposited token value and its derivative within the protocol depends on the total supply, borrows and reserves of that particular asset in the PLF lending pool.

4.3 Protocol for Loanable Funds: PLF

This section describes PLFs functions, focusing on the three selected credit protocols, MakerDAO, Aave and Compound, and their ecosystems on the Ethereum blockchain. Given the decentralized nature and the lack of a central authority, the strategic decisions of each platform are administered through a governance token. The governance token is at the basis of decentralized autonomous organizations (DAOs) [17], where every user can propose changes in the protocol to the community, and governance tokens' owners can vote on that proposal. Obviously, the higher the governance token an address possesses, the more voting power it retains.

In the credit protocols framework, each protocol's community votes on a set of tokens that can be borrowed or supplied as collateral in the protocol. Moreover, for every of these selected tokens, the community votes on the collateralization ratio depending on how volatile and risky the asset is. The collateralization ratio, also defined as the collateral-to-borrow ratio, represents the percentage of collateral that must be provided to borrow another asset. In DeFi, it is typically greater than 100%, as these loans are overcollateralized. For example, if an agent wants to borrow 1000 USD worth of USDT (Tether) while using BTC as collateral, at a ratio of 1.3 or, in percentage terms, 130%, then the agent should lock 1300 USD worth of BTC in order to get the 1000 USD and secure the position.

Lending and borrowing mostly take place in lending pools. Lending pools are the central smart contracts of each protocol. The lending pool is a central account regulated by code, where users can deposit and borrow funds. By depositing funds, users become liquidity providers and earn interest based on the asset's annual percentage yield (APY). In addition, lending pools also coordinate the role of other vital participants, price oracles and liquidators, which is fundamental in maintaining the pool's stability.

On the other hand, apart from a few cases, such as credit delegation or flash loans, borrowers can open a borrowing position only by first acting as a lender and depositing the collateral. Therefore, users become borrowers after supplying the collateral to the lending pool and according to the corresponding collateralization ratio between the deposited and desired asset. They will have to repay the amount taken plus predetermined, calculated interest. Given the volatility of the deposited collateral, lending pools, and more generally PLFs, compute each open position's health factor, defined previously in Equation 4.1, through the interaction with price oracles, that, as introduced previously, retrieves the current price of the cryptocurrency deposited. Most PLFs force their debtor to keep their health factor greater than one to maintain the over-collateralization state of borrowing positions and reduce potential default risks. In default cases, the PLFs encourage third parties, liquidators, to exploit these arbitrage opportunities by repaying the borrower's debt and collecting the still higher-in-value collateral, thus making a profit and benefiting the whole protocol by reducing overall exposure to risk.

The three largest platforms, PLFs that operate on the Ethereum blockchain are Compound, MakerDAO and AAve, representing more than 15b of TVL at the time of writing[2].

[2] `https://defillama.com/protocols/lending/Ethereum`

4.3.1 Compound

According to its whitepaper, Compound, which is entirely working on the Ethereum blockchain, is a protocol that establishes money markets. In particular, these are pools of crypto assets whose interest rates depend on the supply and demand for the asset in the lending pool. Being decentralized, the interaction with the protocol does not need a third-party interaction or negotiation. Thus, every supplier can deposit its assets in its specific money market and earn an interest rate that depends on the current pool's balance. On the other hand, after providing the required collateral, borrowers can borrow from the pool and start paying a predetermined interest rate.

The main technical sets of smart contracts of Compound v2 are the cTokens and the Comptroller, together with liquidationCall, the smart contract related to liquidations that will be later explained in chapter 5. The cTokens, which follow the usual ERC-20 specifications, are the main crypto asset used in the Compound protocol. In particular, whenever a user supplies one of the available tokens to its corresponding lending pool, an equal amount of cTokens is returned to the user. For example, if a user supplies ETH worth 100 USD, it will receive cETH worth 100 USD. Additionally, cTokens are special smart contracts as they can be used as collateral and directly earn interest; therefore, users accumulate yield by simply holding them in an externally owned account (EOA).

The other relevant set of smart contracts of Compound v2 is the Comptroller, which is the risk management layer of the protocol. This layer has important functions, such as determining how much collateral a user must maintain to keep its position healthy by computing its account liquidity. The latter represents the borrowing power of a user in USD terms, and whenever the account liquidity is negative, the Comptroller flags an account to be liquidatable. [28]

4.3.2 Aave

Similarly to Compound, AAve is one of the significant lending protocols available on the Ethereum blockchain. The first version of AAve, v1, was deployed on the Ethereum blockchain in 2019. Then, a second version was deployed just a year later, introducing exciting protocol features such as credit delegation, the ability to choose between a stable and a variable interest rate and flash loans, compared to the present competition on Ethereum, such as in Compound, MakerDAO and Dydx, still available on the Mainnet of Ethereum in 2019. Likewise, Compound v2, AAve v2 presents a tokenized interface through the aTokens, which mirrors the same function of cTokens.

In November 2021, AAve deployed a third version, Aave v3, that signalled the beginning of a new epoch for the lending platform. In particular, Aave v3, by still maintaining the core concepts of the Aave protocol such as aTokens, flash loans, stable rate borrowing and credit delegation, was now enabling the flow of assets over different networks such as Avalanche and Fantom and introducing specific designs for Layer 2.0 blockchains such as Polygon and Arbitrum. At the time of writing, more than 10 billion USD of liquidity is locked in Aave across seven networks and over 13 markets. [29]

4.3.3 MakerDAO

Finally, the Maker protocol is the first lending protocol deployed on Ethereum. It allows users to mint DAI, the protocol's stablecoin, by leveraging collateral assets approved by the community. Like the other lending platforms, the Maker protocol is DAOs, a decentralized autonomous organization; in the case of MakerDAO, the governance token is the MKR. As we introduced in chapter 3, DAI is an example of an asset-backed on-chain stablecoin. DAI is easy to generate; it requires depositing collateral assets into the smart contracts called Market Vaults within the Maker protocol. Every DAI in circulation is directly backed by excess collateral; ergo, it is over-collateralized. Generating DAI entails creating an obligation to repay the debt along with a stability fee to withdraw the collateral leverage and lock it into a vault. However, even though generating DAI is easy, many users prefer to obtain DAI by exchanging them in CEX or DEX instead of creating vaults. MakerDAO represents one of the most successful DeFi projects ever built, with a current TVL of $8.5 billion, almost four times more than Compound and 50% more than Aave[3]. Finally, as Aave, MakerDAO provides the

[3] `https://defillama.com`

opportunity to generate flash loans through the flash mint module.

4.4 Comparison between PLFs

This section provides a comparative overview between the selected PLFs, focusing on data directly queried from the Ethereum blockchain. Moreover, it compares the different interest rate models offered by these platforms altogether with the services and benefits provided, presenting to the reader sufficient material to assess each protocol objectively.

Before entering into the core of the comparison among these protocols, it needs to be specified that even if MakerDAO is classified as a lending platform, therefore as a PLF, users cannot borrow anything else other than DAI. Moreover, DAI can be minted by opening a maker vault, but it can be acquired from other sources by trading on CEX or DEX or transferring from other addresses as a payment source. When comparing MakerDAO, the focus will be on the protocol dedicated to lending and borrowing, not on the much bigger ecosystem around DAI. This difference between MakerDAO and Compound and Aave is also one of the reasons MakerDAO will not be integrated with the empirical analysis presented in chapter 6.

In order to compare these three different protocols, a set of metrics has been chosen, and in particular, the rest of this section will focus on studying the following aspects:

1. Usage and Monetary Outflows and Inflows (TVL, number of unique borrowers and suppliers, average amount borrowed and supplied, conversion rate)

2. Supplying and Borrowing: token compositions

3. Profile of suppliers and borrowers

4. Interest rate model: focus on ETH, DAI and BTC

Moreover, a final comparison, including other qualitative metrics such as the availability of the platform in different markets, the flash loans module, the size of the community and information related to governance, will be given at the end of this chapter.

4.4.1 Usage and Monetary Outflows and Inflows

As previously stated, the three selected protocols are the lending platforms with the highest TVL on the Ethereum blockchain. In order to compare these three protocols, we started studying the data coming from the blockchain after December 2020, when Aave v2 was released. [29]

Regarding the total number of unique suppliers, it can be observed from Table 4.1 that this data is higher in Compound v2 with almost 250k unique addresses, Aave v2 is second with 89k, and finally MakerDAO, according to the official MCD Vaults Tracker[4], states that the total number of vaults ever opened is 30k, but now active are only around 2.5k. By rearranging data from official sources from the MakerDAO community, specifically DeFi Explore[5], the total number of Maker Vaults opened from December 2020 are 12822, with currently only 4,492 active vaults. On the other hand, interestingly, the number of unique borrowing addresses is higher in Aave v2 compared to those in Compound. In contrast, in MakerDAO, counting the number of suppliers is not immediate, as DAI can be acquired from multiple sources.

Table 4.1: Unique Participants and Conversion Rate: Aave and Compound.

	Aave	**Compound**
Total Unique Suppliers	89,259	250,934
Total Unique Borrowers	44,242	16,878
Conversion Rate	49.4%	6.4%

The numerical analysis of the number of suppliers and borrowers reflects the difference in nature between these three platforms: on the one hand, Aave v2 and Compound v2 are credit protocols that are more focused on providing lending and borrowing services to a variety of end users, whereas MakerDAO's lending and borrowing functions are targeting a niche of addresses, as there are only around 4k of opened vaults at this moment but an enormous amount of USD collateral in these, also as the main objective

[4] `http://mcdstate.info`

[5] `https://defiexplore.com/stats`

of its ecosystem is in the stablecoin. For this reason, let the reader focus for the rest of the paper on two similar platforms, Compound and Aave, while making, when possible, a comparison with MakerDAO.

The difference in unique suppliers and borrowers needs to be assessed with the total amount of USD deposited and borrowed and the average amount per deposit and borrow. As shown in Table 4.2, the average USD per borrow in Compound v2 amounted to 715,976$ against 362,535$ in Aave v2. However, looking at the total number of borrowings and borrowers, Aave V2 attracts more unique addresses but less capital than Compound. On the other hand, analyzing the deposits more closely in Table 4.3, it can be seen that the average USD amount supplied in Aave v2 amounts to 668,841$ compared to the lower 550,240$ in Compound v2. In this case, it can be seen that Compound attracts more suppliers, more than three times those in Aave (250k vs 89k), which brings less capital per capita in the ecosystem.

Table 4.2: Borrowing Statistics: Aave and Compound.

	Aave	**Compound**
Total Borrows	181,429	117,652
Total Amount Borrowed	65,774,459,799 $	84,235,349,065$
Average $ per Borrow	362,535.00$	715,970.00$

Table 4.3: Supplying Statistics: Aave and Compound.

	Aave	**Compound**
Total Deposits	344,413	489,925
Total Amount Deposited	230,357,731,137$	269,576,816,561 $
Average $ per Deposit	668,841.00$	550,240.00$

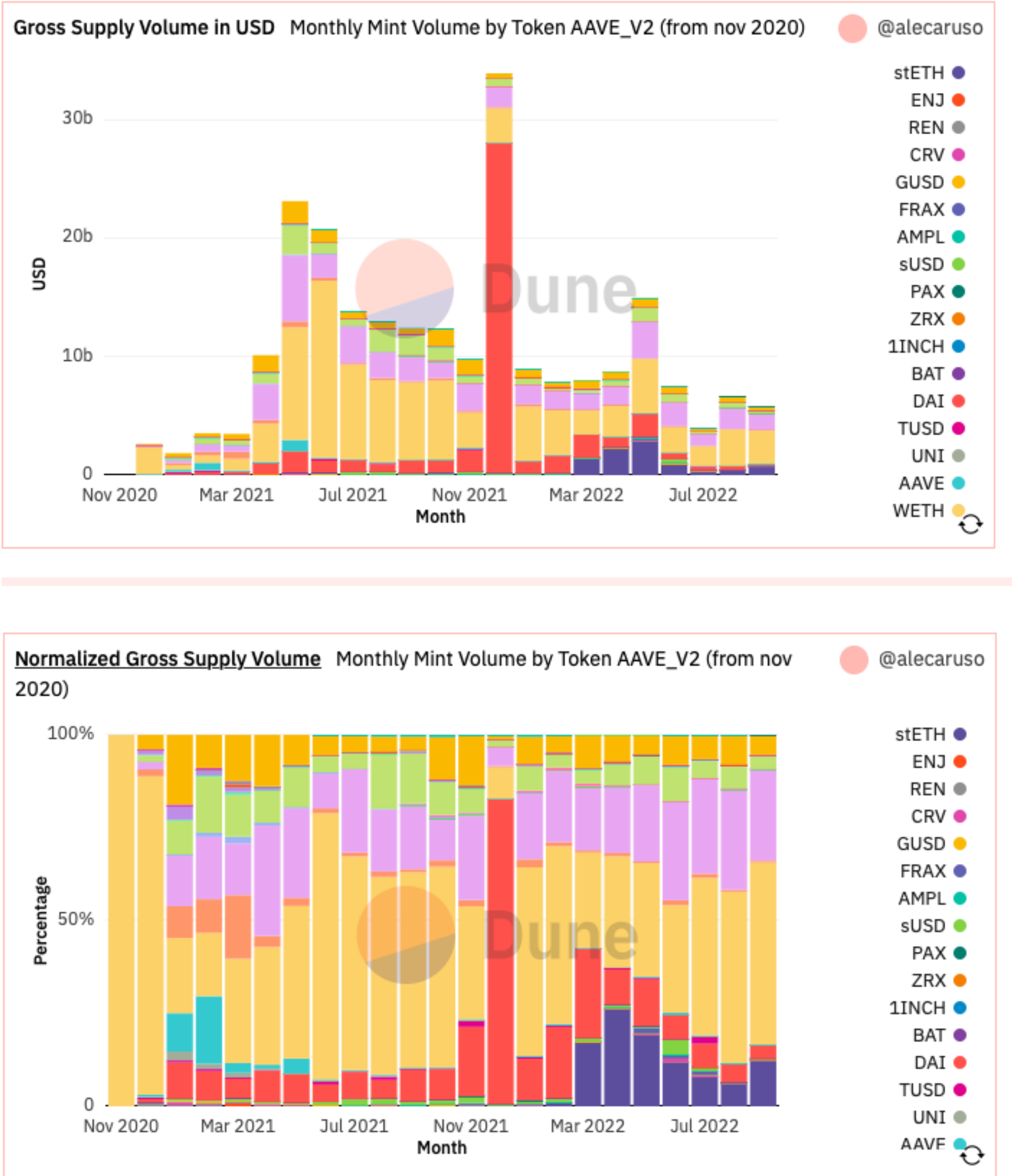
By just looking at the usage, monetary outflows and inflows of these two platforms, it can be seen that Aave v2 seems more balanced in terms of borrowings and lending operations. Compound's users mainly perform supply operations; however, out of these 250k unique suppliers, only 16k become borrowers (only 6.4%). On the other hand, Aave v2 seems to offer more incentives to its suppliers to convert into borrowers; among the 89k supplying assets, 44k also converted, therefore almost 50%, quite an astonishing difference. Finally, analyzing the conversion rate in MakerDAO is superfluous as all the users that supply the protocol are immediately minting DAI and becoming borrowers by construction.

4.4.2 Supplying and Borrowing: Token Composition

The amounts of USD supplied and borrowed in each protocol can be seen in Table 4.3 and Table 4.2. However, the following graphs show how these amounts are divided according to the token composition. Notice that in MakerDAO, the outflow, the borrowing composition, is 100% DAI by the protocol's design; as explained before, an address can only borrow DAI in this context.

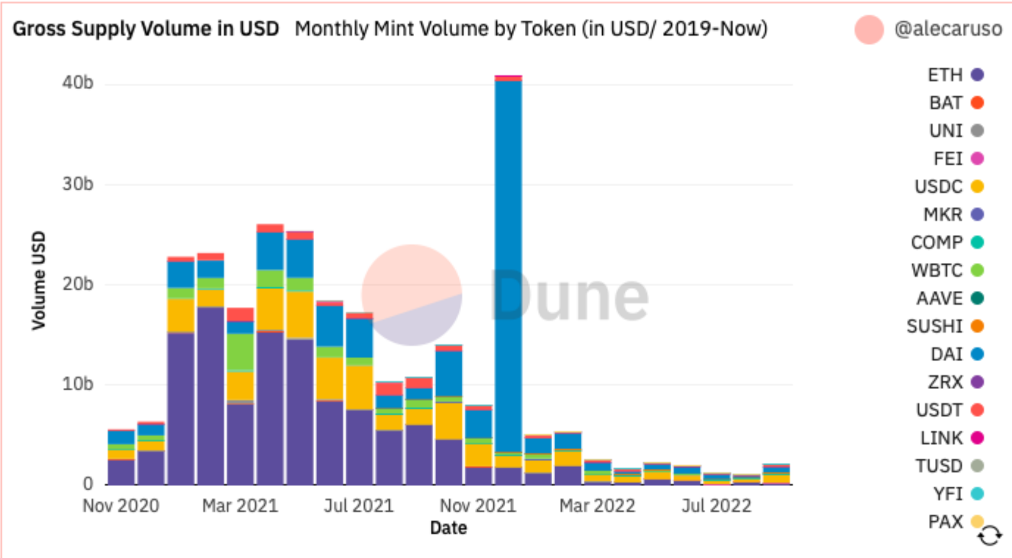
On the supply side, the composition of the supplied amount of Compound V2 and AAve v2 is shown in the next set of graphs. In AAve v2, it can be seen from Figure 4.1 that the supply volume has followed a decreasing trend starting May 2021. In particular, in terms of token composition, users have mainly supplied ETH (in the form of sETH and WETH) and BTC (as WBTC). Stablecoins such as DAI, USDC, and USDT have also been supplied, especially in December 2021, a memorable event where 82% of the total supply represents DAI.

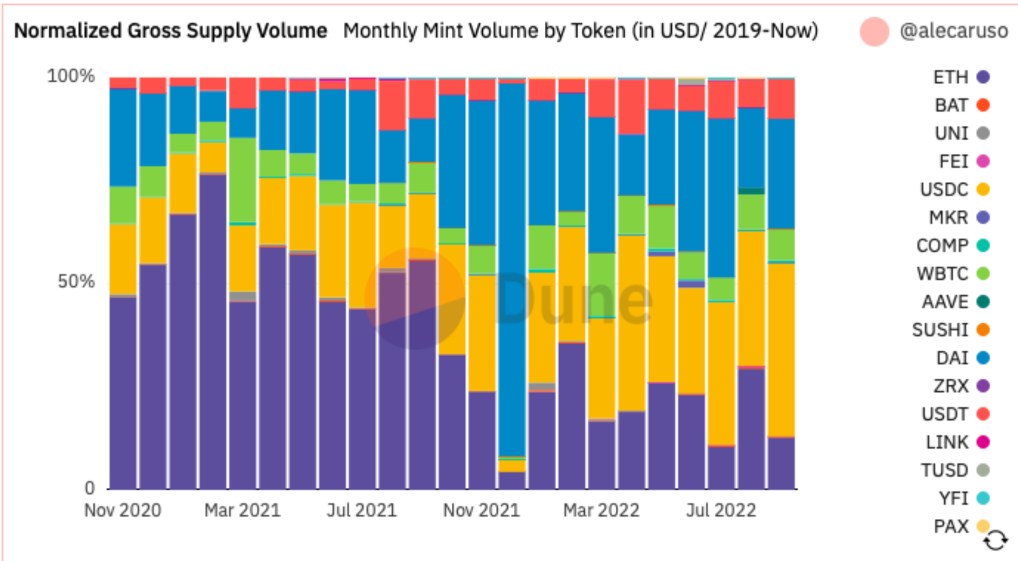
Figure 4.1: Supply Volume Aave



On the other hand, analyzing Compound v2 supply composition, shown in Figure 4.2, it can also be noticed how the USD deposited volume has shrunk over time. Additionally, stablecoins represent most of the tokens supplied to the protocol, particularly USDC, USDT and DAI, with the more volatile tokens such as ETH, BTC and LINK that constitute a minority group. On the supply side, one clear difference between these two platforms is that users prefer to supply more stablecoins in Compound v2 compared to Aave v2. This can be translated as a signal coming from the users' behaviour which may be interpreted as finding Compound riskier than AAve.

Figure 4.2: Supply Volume Compound





Finally, regarding MakerDAO, given the fact that users do not just supply from the platform to earn interest, but they are forced to open a borrowing position, the current situation of active Maker Vaults is provided in one of the MakerDAO official statistics websites[6], where maker vault information can be extracted. In particular, the collateral where most active vaults are is represented by ETH-A, WBTC-A, ETH-C and USDC-A. Class A-B-C, in a nutshell, refers to scenarios to generate DAI under

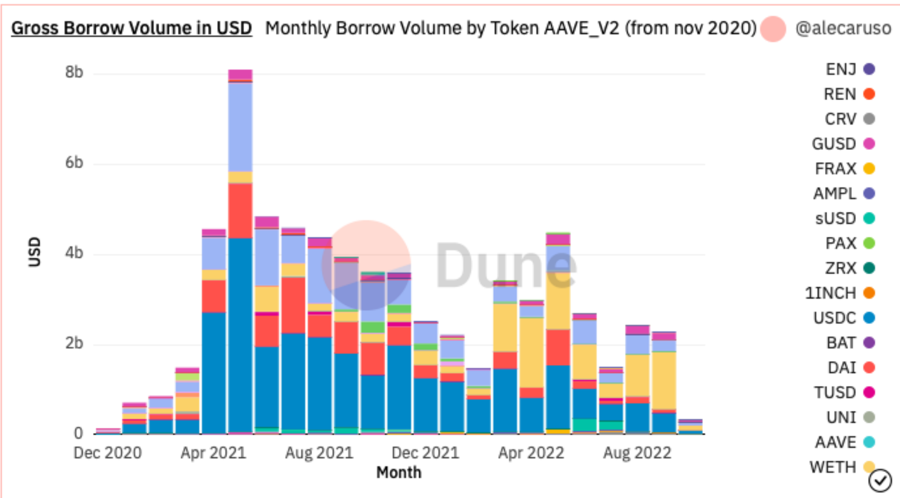
[6] http://mcdstate.info

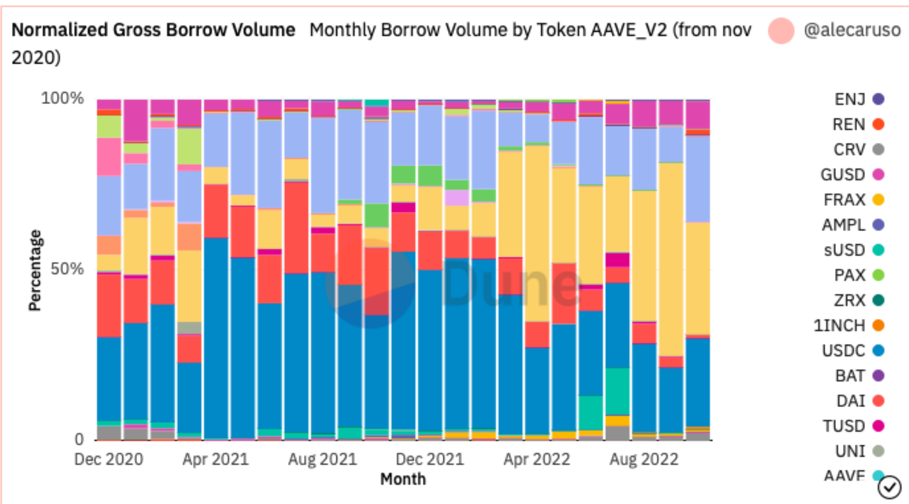
different sets of risk parameters[7] (stability fees, liquidation ratios, liquidation penalty), where A is less risky, and C is riskier.

Regarding the amount of USD locked value in these active vaults, it can be inspected that the collateral with the highest locked value is PSM-USDC-A (the Peg Stability Module against USDC), followed by ETH-C and GUNIV3DAIUSDC2-A (the Uniswap vault) and finally ETH-A.

On the other hand, focusing on the borrowing behaviour, according to Figure 4.3, in Aave v2, users keep borrowing similar amounts over time. In particular, the typology of borrowed tokens seems to have shifted from most stablecoin borrowings in 2021 to more volatile tokens such as WETH and WBTC in 2022.

Figure 4.3: Borrow Volume Aave

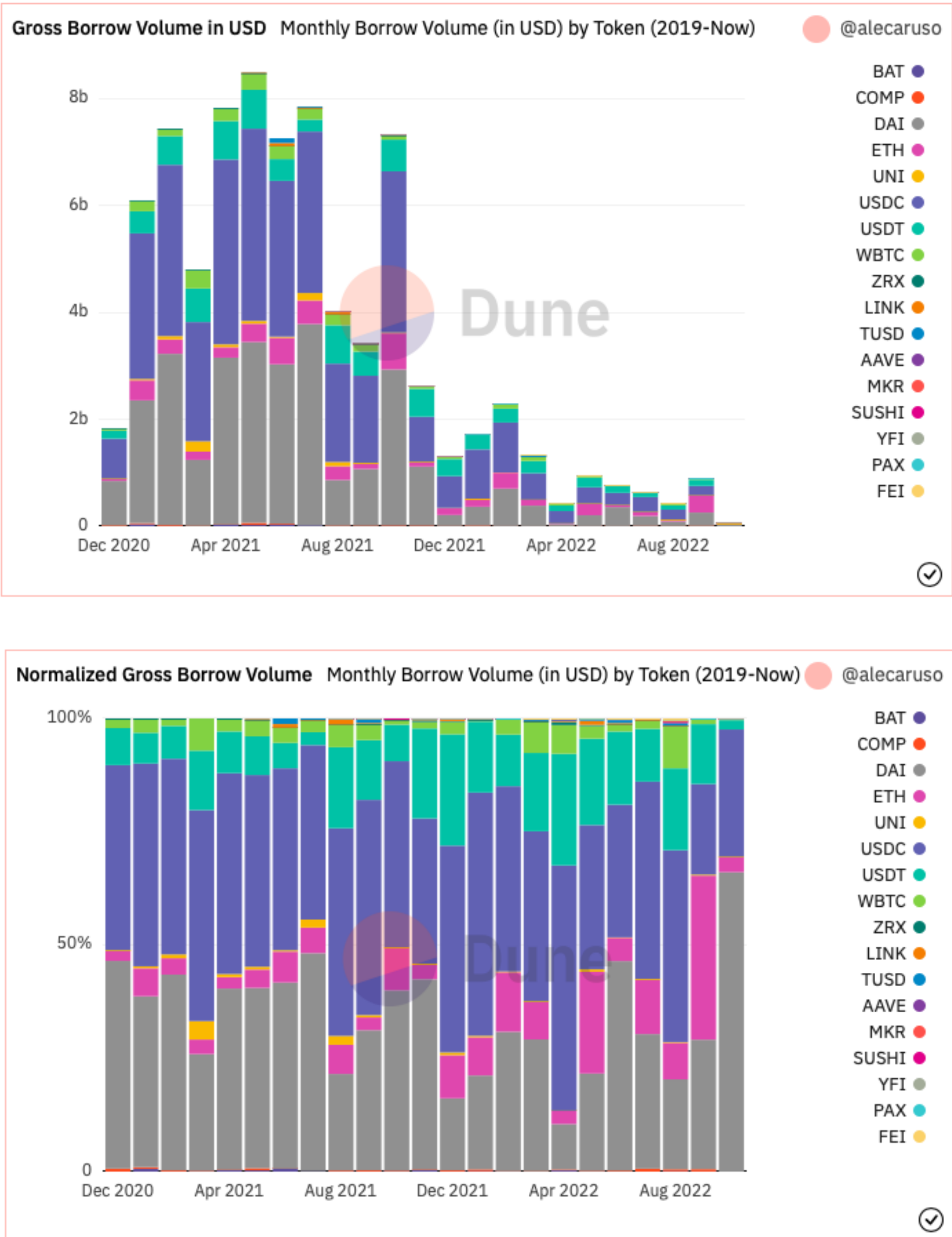




[7] https://blog.makerdao.com/a-guide-to-dai-stats/

On the contrary, in Compound v2, as shown in Figure 4.4, the number of borrowings by month dramatically decreased from 2021 to 2022, and the token composition suggests that the majority of borrowings involved stablecoins such as USDC, DAI and USDT.

Figure 4.4: Borrow Volume Compound

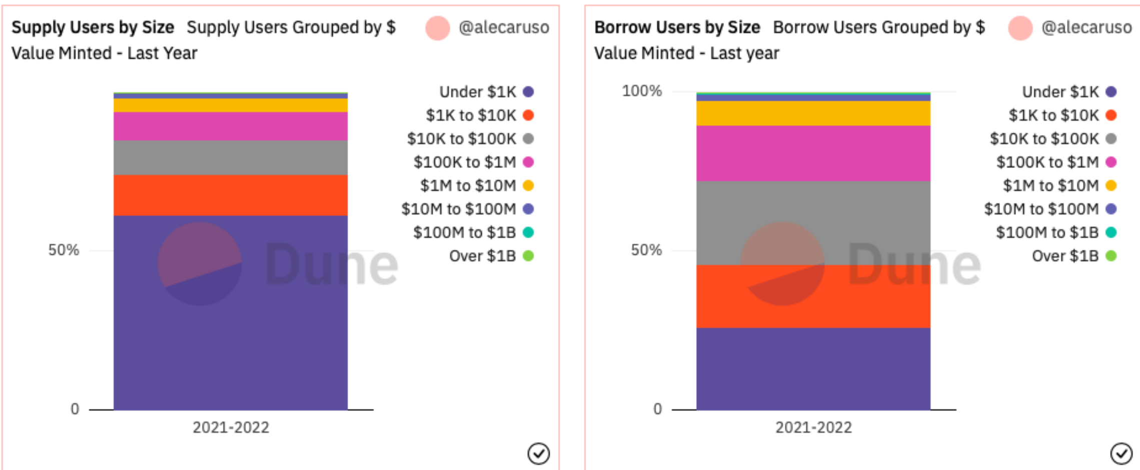


Finally, this last comparison between Compound v2 and Aave v2 gives another exciting signal in the user's behaviour: the former seems to be a riskier platform, as addresses mainly borrow stablecoins, compared to the latter, where actual leverage (margin trading) operations are performed.

4.4.3 Profile of suppliers and borrowers

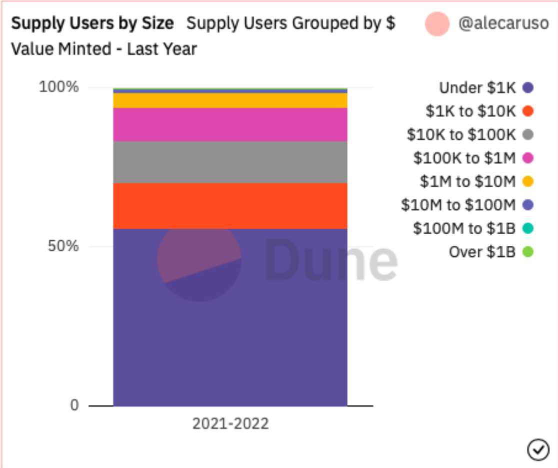
In addition to the two previous analyses, this section studies the end users of these protocols and divides them according to the amounts supplied and borrowed. Another time, the main focus is on Compound and Aave, leaving MakerDAO on the side due to the nature of the protocol. According to Figure 4.5, in Compound, most suppliers (more than 50%) have supplied tokens valued at less than 1,000$. On the other hand, most borrowers (more than 70%) have opened individual positions where less than 100,000$ were involved.

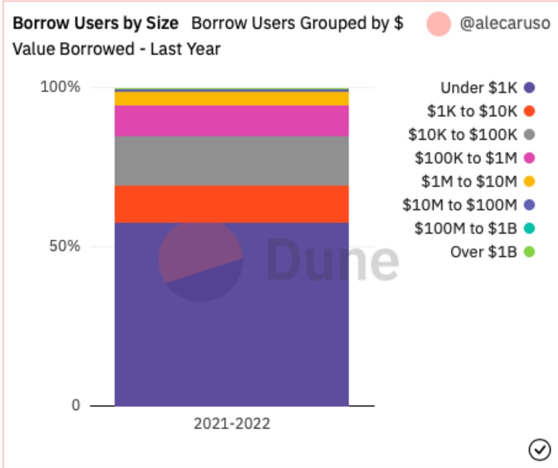
Figure 4.5: Profile of Compound Users



Contrarily, looking at Aave suppliers and borrowers profiles in Figure 4.6, it looks like suppliers and borrowers follow similar patterns. More than 80% of the suppliers have deposited relatively small amounts, below 100,00$, which is, however, reflected in the borrowing operations.

Figure 4.6: Profile of Aave Users





The analysis of these two graphs strengthens the previous claim regarding the conversion rate of these two platforms. On the one hand, in Compound, it is clear that the majority of the borrowing positions have been opened by the wealthier suppliers (above 100k), as over-collateralization needs to be taken into account. On the other hand, the profiles of the users in Aave v2, as the two bar charts in Figure 4.6, look alike, suggesting a more healthy and balanced platform where small players are also involved in the borrowing ecosystem.

4.4.4 Interest rates: focus on ETH, DAI and BTC

Looking at the interest rates and collateral factor between these three different platforms, in Table 4.4, Table 4.5 and Table 4.6, it can be seen that Aave offers higher APY on the most borrowed/supplied tokens, and also the possibility to choose between a stable and variable borrowing interest rate. Moreover, Aave enables more assets to be used as collateral, making it a more interesting and profitable platform.

Table 4.4: Lending APY

	ETH	**BTC**	**DAI**
Compound	0.26%	0.03%	1.23%
Aave	5.48%	0.14%	1.29%
Maker	-	-	0.01%

Table 4.5: Borrow APY and Collateral Factor

	ETH	BTC	DAI
Compound	3.72% (82%)	2.86% (70%)	2.9% (83%)
Aave (Variable-Stable)	3% - 7% (82%)	2.68%-11.9% (70%)	1.46%-4.93% (77%)
Maker[8]	-	-	-

Table 4.6: Distinct Collateralized Assets

	Number of Assets Used as Collateral
Compound	9
Aave	24
Maker	18

4.4.5 Final Comparison

Finally, Table 4.7 shows the previously described metrics with other relevant design aspects, services offered, and qualitative details to give the reader a summary when objectively assessing the selected PLFs.

[8]When it comes to ETH-backed loans, Maker has a higher stability fee (9%) for the higher LTV Ratio product and a lower stability fee (5.5%) for the lower LTV Ratio product. Maker stability fees are the borrowing interest rate and continuously accrue on the debt over time.

Table 4.7: Comparison of Selected PLFs

	Aave	**Compound**	**Maker**
TVL[9]	\$4.29B	\$2.15B	\$7.46B
Borrowers/Suppliers	49.4%	6.4%	-
Flash Loans	Yes	No	Yes
Community (Twitter)	~500,000[10]	~236,000[11]	~230,000[12]
Availability in Different Markets	Yes	No	Yes
Credit Delegation	Yes	No	No
Governance Token Holders	~121,000	~197,000	~87,000
Level of Centralization	85.57%	90.32%	82.23%
Variable vs Stable Interest Rates	Yes	No	No

According to Table 4.7, Aave looks like the most attractive lending platform among the three taking into account all the selected metrics. In particular, it is interesting to see how these protocols, even though they claim to be decentralized, the governance tokens distribution of AAVE, MKR and COMP, are retained by few addresses that influence the choices of the whole community given the higher voting power. This is shown in the level of centralization item, which measures how much of the total supply of the governance token is owned by the top 100 addresses.

[9] https://www.defipulse.com
[10] https://twitter.com/AaveAave
[11] https://twitter.com/compoundfinance
[12] https://twitter.com/MakerDAO

Chapter 5

Liquidations

This chapter examines in depth the concept of liquidations in DeFi lending protocols. In particular, a general overview of this mechanism is given at the beginning by firstly introducing liquidations in the traditional finance context, describing the main focus points related to this phenomenon. Consequently, liquidations in DeFi are explained as what they represent: blockchain extractable value (BEV) opportunities. Then, for every selected PLF, the current liquidation function smart contract is introduced, and statistics regarding these liquidations are directly derived from the Ethereum blockchain through visualizations. Lastly, a focus on the usage of flash loans in the liquidation ecosystem is provided to the reader, specifically looking at the Aave protocol.

5.1 Liquidations in TradFi

In traditional finance, whenever a company or natural person cannot pay the creditors in case of a loan, or if, in a trading context, to maintain the value of the collateralized asset, a liquidation can happen. However, this phenomenon is not immediate, as it has been introduced throughout this paper, which is one of the main differences between TradFi and DeFi. In traditional finance, the liquidation process highly depends on the situations the company or individual faces and the type of assets involved. Moreover, being publicly liquidated has dramatic effects on both companies and individuals. On the one hand, companies' reputation is deeply affected, and on the other hand, individual credit rating deteriorates, making it difficult for a person even to open another borrowing position. As described by the New Zealand Insolvency and Trustee Service governmental institution[1], the liquidation life-cycle that affects companies is a long and tedious process where different actors are involved such as courts, directors, shareholders, liquidators and creditors.

In particular, this phenomenon starts when a company cannot pay its debts (or part of it). Then the director, the shareholders, the courts and creditors start the liquidation process. Consequently, a liquidator, usually an insolvency expert, is appointed when a company is liquidated. This agent notifies the Companies Office and advertises the appointment. Then, a creditors meeting is held to confirm the liquidator and start the liquidation administration. This involves investigations of possible solutions, such as the closure of business and making payments to creditors (as dividends), while updating through reports the involved creditors on the state of the liquidation. Finally, once the liquidation is completed, a final report is sent to the creditors, and the Companies Office is informed. DeFi has radically transformed how liquidations and liquidators are perceived and how this phenomenon affects the borrowing and lending landscape. As readily understood, liquidations in DeFi are a completely different object of study compared to the TradFi counterpart. Given the decentralized nature of the environment they happen in, none of the previously mentioned third parties is involved, like courts, shareholders or directors. Moreover, individual credit ratings and business reputation are not affected due to the anonymization of the borrowers.

5.2 Blockchain Extractable Value (BEV) and Liquidations

The blockchain is a complex system where numerous agents are involved. However, as a new and emerging technology, it presents several profitable opportunities that can be exploited by studying the game's rules. In particular, the publicly available

[1] `https://www.insolvency.govt.nz/business-debt/the-liquidation-process/`

and easily accessible information about every transaction makes the blockchain a land of transparent dishonesty, where automated bots are constantly hunting for victims from whom value can be extracted. In particular, these bots, basically smart contracts, screen the blockchain public ledger looking for valuable transactions where blockchain extractable value (BEV) is present. According to Gervais et Al. [30], during 32 months from December 2018 to August 2021, the BEV amounted to 540M USD in profit, coming from three different sources: sandwich attacks, arbitrage opportunities and liquidations. We refer the reader to the cited paper for a more detailed background on the first two source types. Regarding liquidations, from April 2019 to April 2021, in the significant lending platforms, namely Aave (v1 and v2), MakerDAO, Compound and DyDx, there have been a total of 28138 successful liquidations, amounting to 807.46M USD of collateral sold, demonstrating the immense magnitude of such an unpopular phenomenon.

Nevertheless, what are liquidations? Furthermore, why are they important? As previously explained throughout the paper, liquidation entails selling the collateral of the borrower whose position is risky. Whenever the collateral price drops below a certain threshold, the health factor, Equation 4.1, declines below 1, then a borrowing account becomes liquidatable. A liquidator can be any third external user, and the platforms themselves encourage the users to become liquidators through tutorials[2], as it helps in decreasing the overall risk of the system against unhealthy positions, favouring the entire lending and borrowing protocol.

There are mainly two different types of liquidations, according to Qin et al.[31]:

1. Fixed Spread Liquidations: popular in Aave and Compound, are completed in one blockchain transaction. The liquidator repays the debt of a borrowing position and, in return, acquires the collateral at a discounted price. In fixed spread liquidation, users can adopt two different extracting strategies: either front-running other competing liquidators, which means being faster in finding liquidatable positions than other users, or back-running, that is anticipating the price oracle updates

[2] `https://docs.aave.com/developers/v/1.0/tutorials/liquidations`

and computing the health factor before the competition, and even the protocol itself. The majority of fixed-spread liquidations are of the front-running type.

2. Auction Liquidations: mainly present in MakerDAO. This protocol was firstly characterized by English auctions presented in the Liquidation 1.2 smart contract. Recently, the MakerDAO liquidation mechanism has been updated to the Liquidation 2.0 version, upgrading the auction mechanism to the Dutch type, defined by instant settlements. In this context, a liquidator starts an auction that lasts a predetermined amount of time (6 hours in Liquidation 1.2), and multiple users can compete by bidding on the collateral price. The one with the lowest bid wins the opportunity to liquidate that position. English auctions, or open ascending auctions, are characterized by bidders that outbid each other increasingly in an interactive manner. On the other hand, Dutch auctions are the opposite of English ones, as the auction starts with a high asking price that decreases until termination.

The following sections describe the liquidation smart contracts adopted by each lending protocol, providing summary statistics regarding the total number of users participating, the token composition of liquidated positions and the dollar amounts involved.

5.3 Aave v2 - liquidationCall

Aave v2 adopts a fixed spread liquidation mechanism encoded in the liquidationCall smart contract, publicly available for inspection in the official Github repository. Whenever the health factor of an account, considering the total amount of open credit positions, drops below 1, the account is deemed liquidatable. Therefore, any user can interact with the liquidationCall function in the Lending pool smart contract[3] repository of the protocol, by paying part of the debt owed and obtaining a discounted amount of collateral. In particular, it must be noted that, in Aave v2, a liquidator cannot

[3]`https://github.com/aave/aave-protocol/blob/master/contracts/lendingpool/LendingPool.sol`

repay the entire position but can only repay a portion of it, defined by the close factor described in section 4.2. The current close factor of Aave v2 is 50%, which means that in a single transaction, a liquidator can only repay half of the debt. In Aave v3, which is present in other blockchains rather than Ethereum, liquidationCall allows 100% of the debt to be liquidated in a single transaction as long as the health factor is smaller than the close health factor threshold. Otherwise, the close factor remains still 50%. As this paper focuses on the Ethereum blockchain, Aave v2 liquidation function data will be analyzed, as Aave v3 is not deployed on this blockchain. [29]

As it can be seen from Figure 5.1 and Figure 5.2, the number of liquidations in Aave v2, starting from December 2020, the date when the second version was released, were 25,773 performed by only 241 liquidators repaying collateral for 1,008,422,539$. This is the first signal of how enormous the amount of money involved in liquidations is and how small this niche market is compared to it. Regarding the number of liquidations, it can be seen that liquidations surpassed 300M in May 2021, mainly due to the decline in price on May 19 of ETH of approximately 41% from 3400$ to 2000$. As can be noticed from the pie chart, the majority of these liquidations include WETH (55.9%), WBTC (18.7%) and LINK (10.1%), demonstrating how sensible liquidations are to ETH price fluctuations.

Figure 5.1: Liquidation in Aave

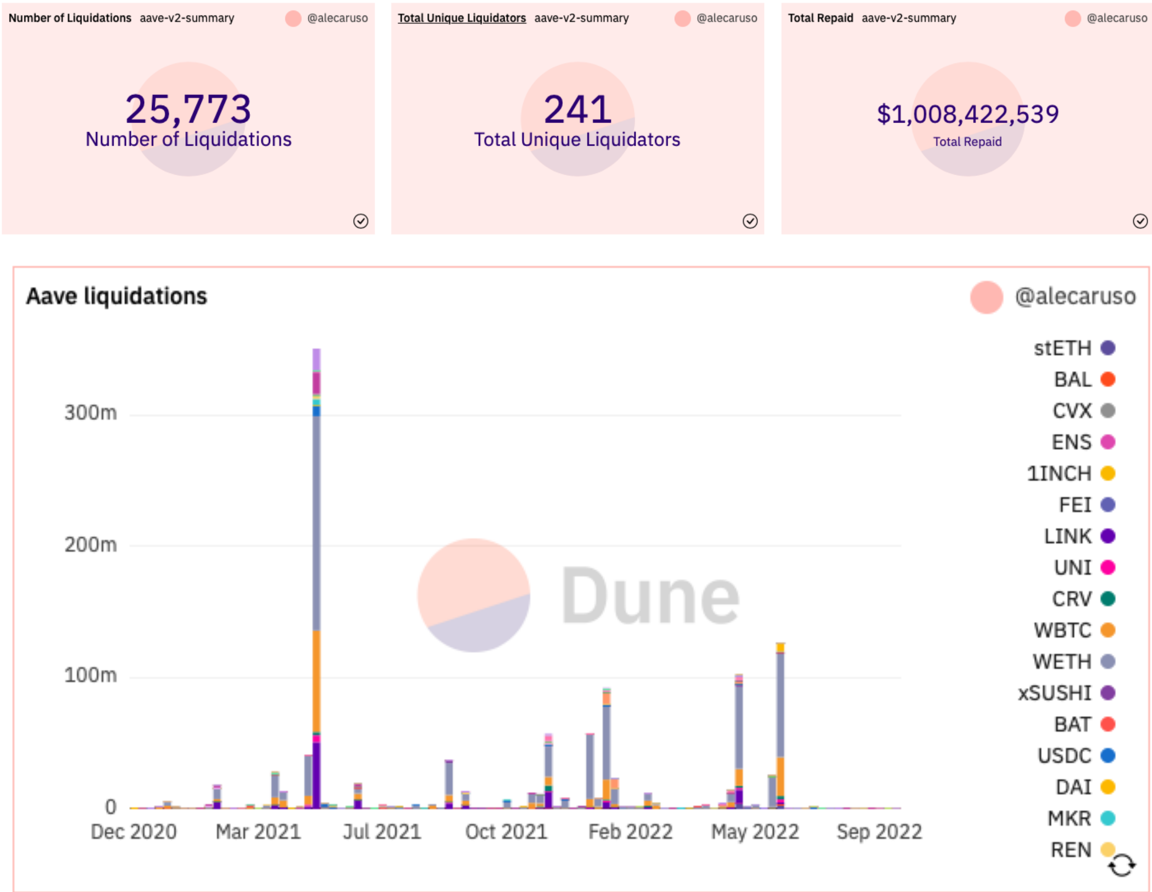
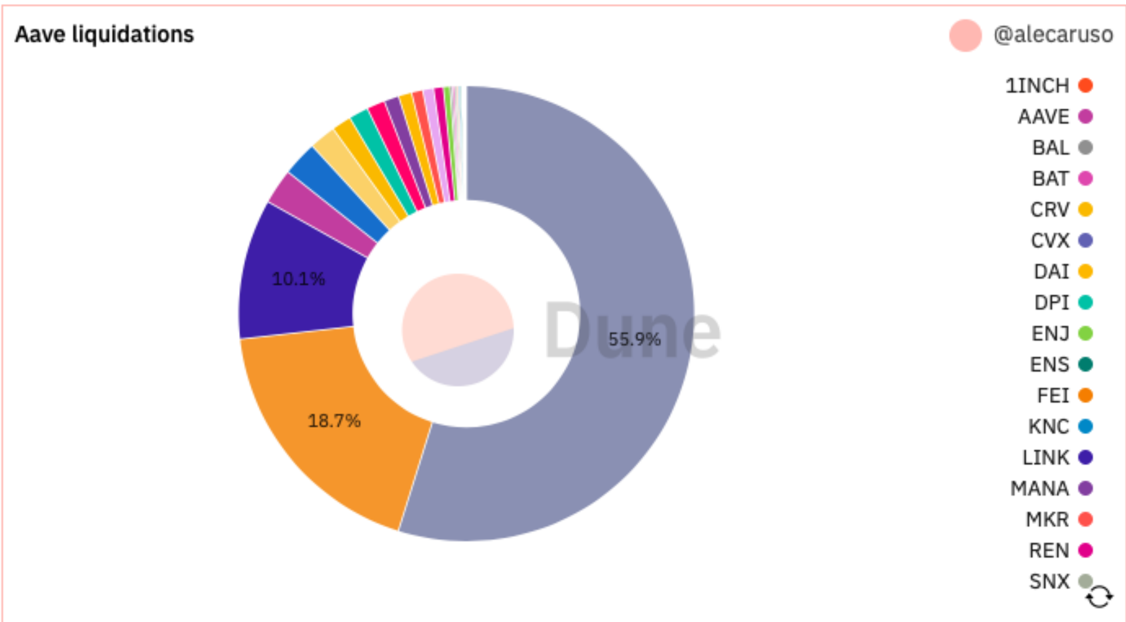


Figure 5.2: Liquidation in Aave: Pie Chart



5.4 Compound v2 - LiquidateBorrow

Similarly to AAve v2, Compound v2 liquidations are also of the fixed spread type. The designated liquidation smart contract is called LiquidateBorrow and has similar characteristics as liquidationCall. Comparatively, in this case, an entire account is deemed liquidatable whenever the account liquidity, computed by the Comptroller smart contract, is negative. Consequently, a liquidator can repay either part or all of an outstanding loan on behalf of the borrower but only up to a certain percentage (the close factor) of the total positions. After repaying the position, the liquidator will receive the discounted amount of that collateral, whereas, in the Compound ecosystem, this discount is defined as the liquidation incentive. The close factor is not fixed as in Aave v2, but it ranges between 0-100% depending on the borrowed asset. [28]

Additionally, Figure 5.3 and Figure 5.4 shows the data regarding liquidations in Compound v2, directly querying from the liquidateBorrow smart contract deployed in the Ethereum blockchain. By aggregating all liquidations weekly, it can be seen that similar to what has been reported in Aave v2, the week ranging from May 17 to May 24 2021, reports the highest total liquidated collateral, amounting to more than 200 Million only in Compound v2.

Figure 5.3: Liquidation in Compound

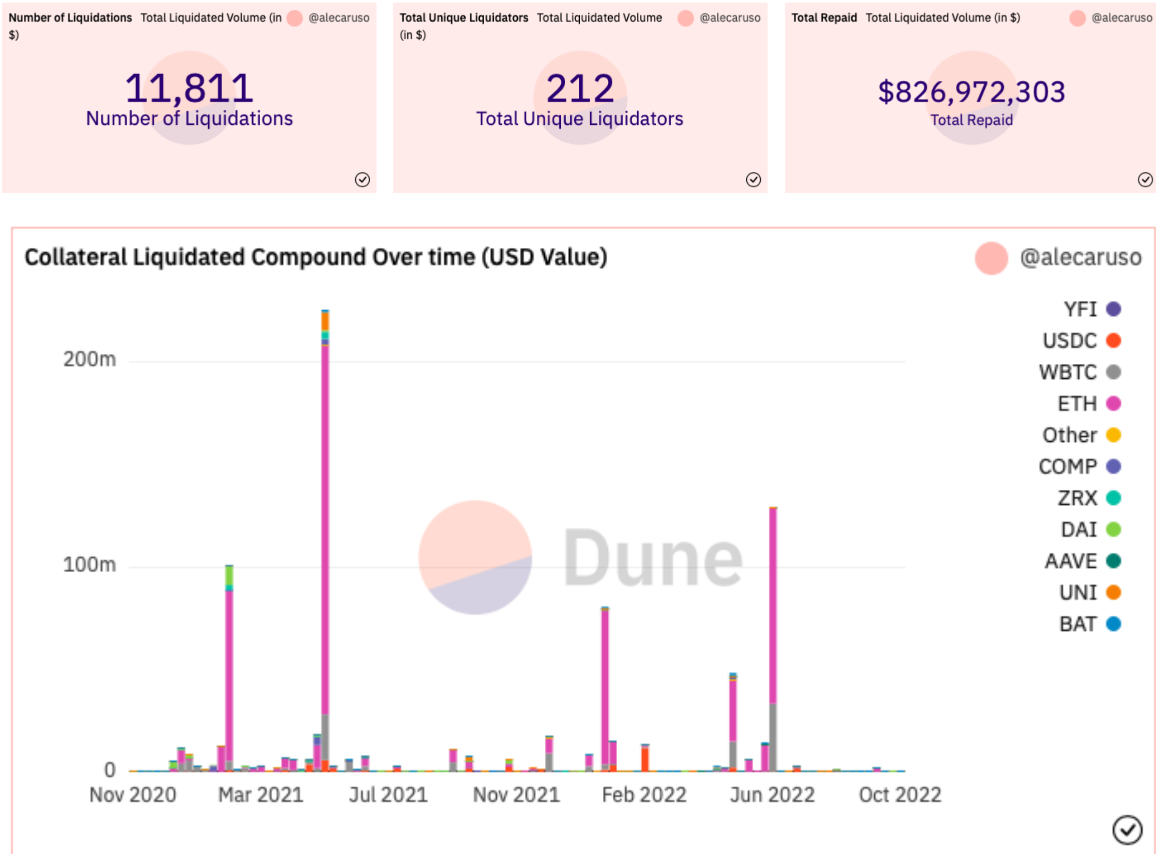
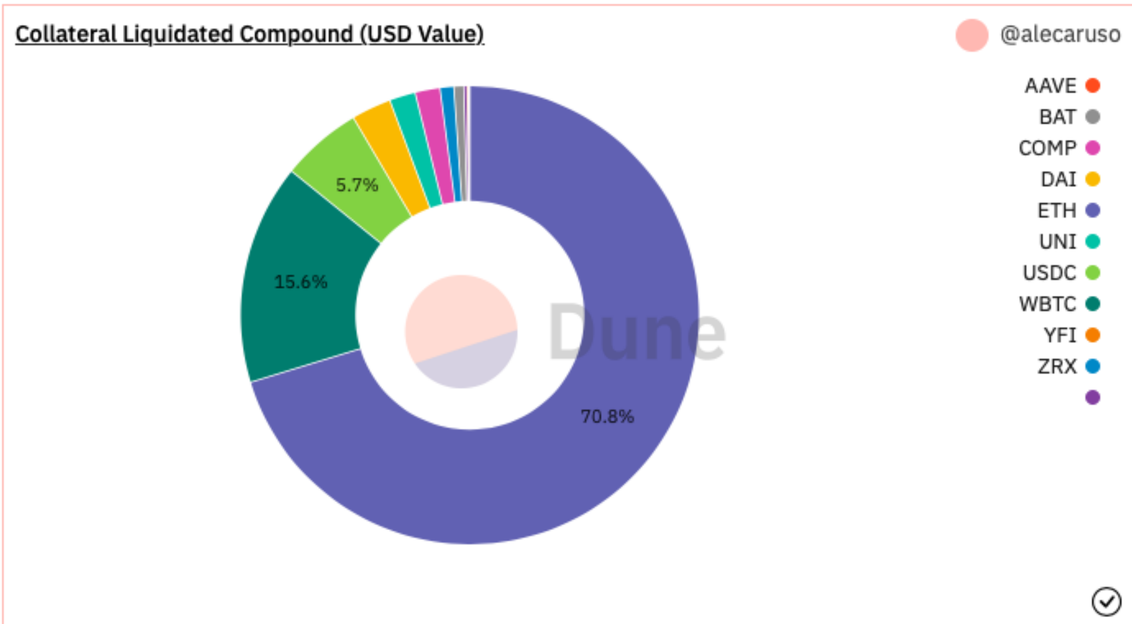


Figure 5.4: Liquidation in Compound: Pie Chart



Similarly to Aave v2, in Compound v2, only 212 unique addresses correspond to liquidators, and the total number of liquidations from December 2020 was 11,811 for a total repaid amount of $826,972,303. By focusing on the pie chart, it can be seen that the total liquidated collateral was composed again of a strong presence of the ETH asset, amounting to 70.8%, followed by BTC and USDC, respectively 15.6% and 5.7% of the total.

5.5 MakerDAO: Liquidation 1.2 Liquidation 2.0

As presented before, the MakerDAO lending platform adopts a different liquidation system than Aave and Compound, where an auction-based one is adopted instead of a fixed spread mechanism. In particular, two main modules have been famous in the liquidation process in the MakerDAO ecosystem, namely Liquidation 1.2 and Liquidation 2.0 smart contracts. The former adopted an English auction through a two-phase process called tend-dent auction, and the latter a Dutch type, characterized by instant settlement. [31] In April 2021, the Liquidation 1.2 was deprecated[4] in favor of the Dutch auction module.

Liquidation 2.0[5] introduces an important feature compared to the previous Liquidation 1.2, which is the flash lending of collateral. This eliminates any capital requirements for bidders, apart from gas fees, that can use any external flash loan smart contract involving DAI in the DeFi system, conforming to a particular interface, to interact with the Liquidation 2.0 auction contract. Compared to its previous version, the liquidation contract (called *Dog* and not *Cat* anymore), allows partial purchases of the collateral. The liquidation starts with the asset's current price delivered by the Oracle Security Module (OSM) times *buf*, which is a multiplicative factor that increases the starting price of an auction. Then *calc*, an overtime price calculator function, usually a linear decreasing one, is applied to the price. *Tail*(*tau*) instead refers to the maximum amount of time an auction can last, altogether with *cusp*, which fixes the maximum percentage price drop of the collateral. Whenever the auction's time and price exceeds *tail* and

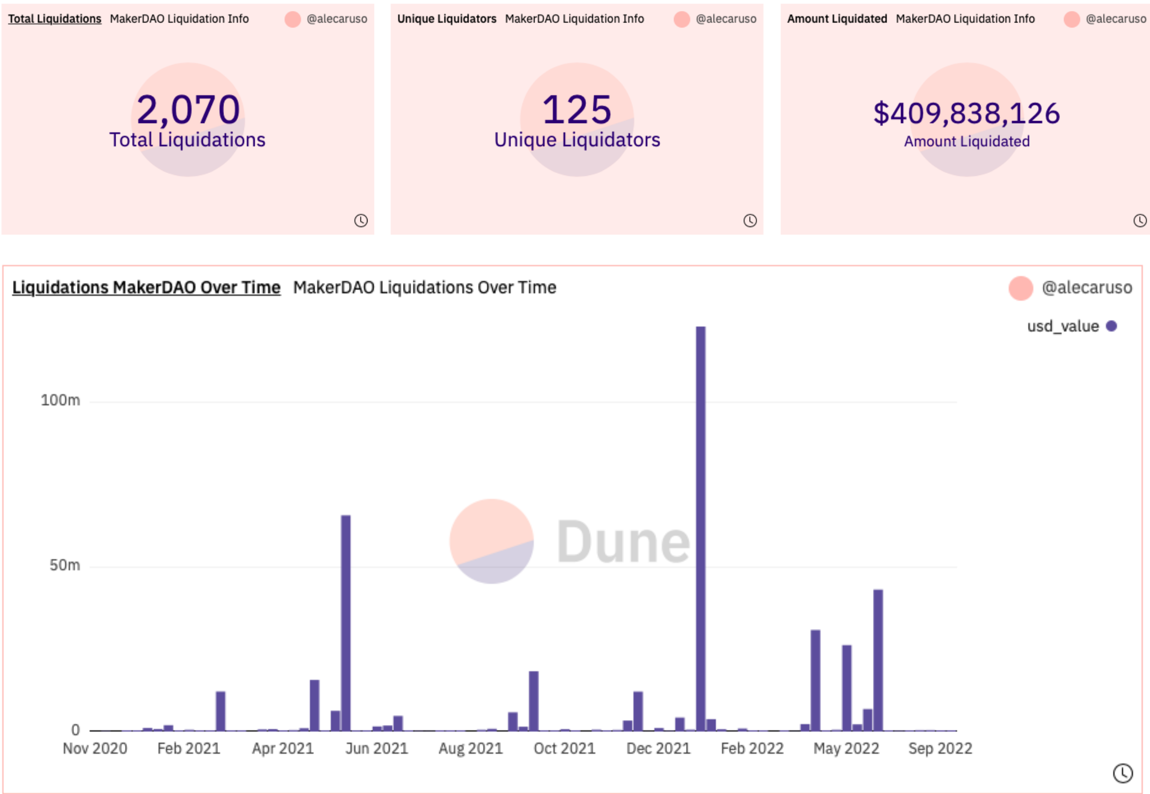
[4] `https://mips.makerdao.com/mips/details/MIP45`
[5] `https://docs.makerdao.com/smart-contract-modules/dog-and-clipper-detailed-documentation`

cusp respectively, the auction must be reset. *Tab* is the target DAI to raise from the auction, corresponding to the debt, liquidity fees and liquidation penalty. Liquidators bid through the *Dog* smart contract function called *take*.

Let us explain how Liquidation 2.0 works through an example. Imagine Alice and Bob are two liquidators in a ETH-A scenario vault: where the reported OSM price is 300 DAI, the *buf* is 15%, *tau* of 15,000 seconds, *tab* 80,000 DAI and the price follows a linearly decreasing *calc* function. Firstly, Bob calls the function *take*, offering 60,000 DAI for 280 DAI per ETH, cashing out 214.28 ETH collateral. Then the price of the collateral (ETH) keeps falling, and Alice bids the remaining amount of 20,000 DAI at 240 DAI per ETH, getting 83.3 ETH collateral. If in this auction the time elapsed has exceeded *tau* or if the price has dropped below *cusp* x *top*, then the auction gets reverted. [87]

MakerDAO liquidation 2.0, compared to Compound and Aave liquidation smart contracts, presents higher complexity and intersection between different smart contract functions. For this reason, less information can be extracted from the blockchain related to liquidated collateral token composition. In particular by directly querying the Liquidation 1.2 *Cat* contract and Liquidation 2.0 *Dog* contract from the Ethereum Blockchain, the summary information can be found in Figure 5.5. It can be noticed that the total number of unique liquidators is lower compared to Aave and Compound. In MakerDAO, there have been only 125 unique liquidators for a total of 2070 liquidations, still considering from December 2020. The total amount liquidated, however, corresponds to 409,838,126 $, with the highest liquidations amount in January 2022 surpassing 100M dollars.

Figure 5.5: Liquidation in MakerDAO

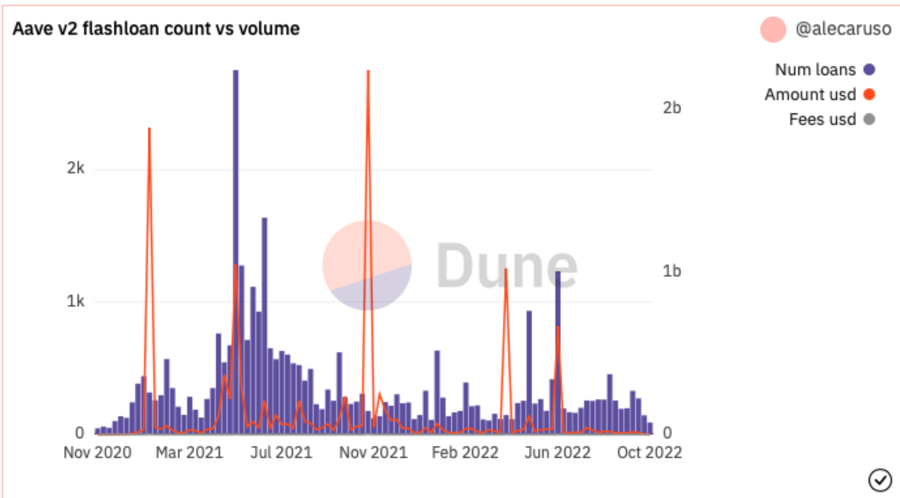


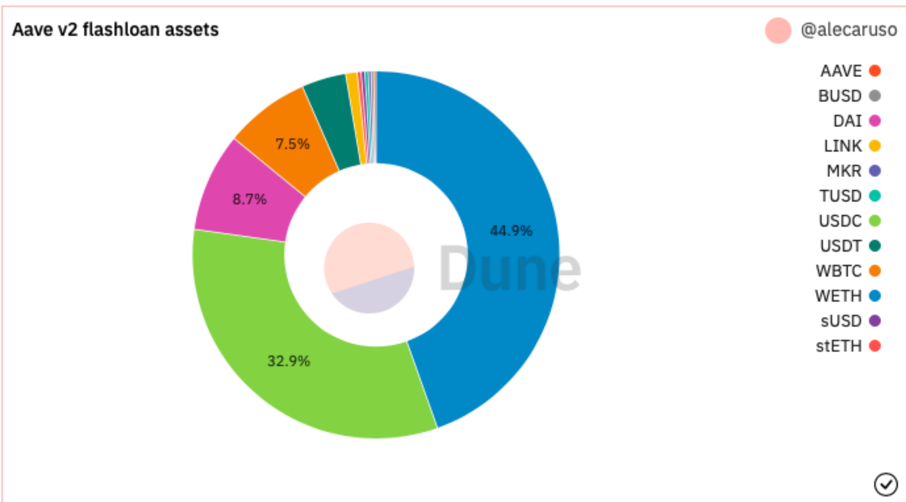
5.6 Flash Loans in Liquidations

One of the main actors behind liquidations is flash loans, which, as introduced in chapter 3, are very revolutionary financial instruments that enable every user to get immediate access to enormous amounts of money without depositing any collateral. As fixed spread liquidation happens in one transaction, flash loans are ideal to be used in this context by liquidators. As explained in Quin et al. [31], a typical process in flash loan liquidations follows four steps. Inside a single atomic transaction, the liquidator initiates a flash loan in currency A to repay the liquidatable position. Once repaid the debt, the liquidator collects the collateral plus the liquidation incentive in the collateral's currency B. On an exchange, for example, Uniswap, the liquidator swaps the received collateral in B for currency A, which is used to repay the flash loan plus the accumulated interest. If the liquidator cannot repay the flash loan, all these sub-transactions are reverted. On the other hand, if the liquidator can repay the flash loan, then what remains is its profit.

As we have introduced through the paper, Aave v2 and, recently, MakerDAO enable the creation of flash loans through the *flashloan* smart contract and the Flash Mint module, respectively. Focusing on Aave v2, by querying the respective smart contract, Figure 5.6 shows the number of flash loans that have been opened over time, with a peak in May 2021, reasonable event given the number of liquidations happening in that month, as previously mentioned, altogether with the USD value borrowed and fees paid. Moreover, according to the pie chart of the different tokens used in flash loans, it can be noticed that the majority of tokens used corresponds to stablecoins, namely USDC and DAI, respectively with 40.9% and 10.3%, and ETH with 39.8%.

Figure 5.6: Flash Loans in Aave v2





Chapter 6

Data, Inference and Prediction

This chapter proposes a statistical analysis of the liquidation phenomena, looking at it from inference and predictive standpoint. Firstly, by studying the effect on the probability of being liquidated from the borrower's choice of a protocol. Then, a predictive approach is adopted where machine learning algorithms are trained on historical data to forecast the probability of being liquidated for a given borrower and propose a new strategy for bots, to prioritize borrowers more likely to get liquidated. In section 6.1, the data extracted from the blockchain are described, along with further cleaning, manipulation into a relational dataset and feature engineering. In section 6.2, the inference problem is tackled by interpreting the direction of the linear effect coming from the choice of a platform on the probability of being liquidated. Additionally, section 6.3 is dedicated to the forecasting problem: training different machine learning algorithms to predict the likelihood of being liquidated for any given borrower, measuring out-of-sample metrics. Finally, section 6.4 summarizes and interprets the results obtained by providing additional explanations regarding limitations and further improvements of the selected procedures.

6.1 Problem Definition and Dataset Creation

The creation of the relational dataset starts by extracting from the blockchain all the borrowings, collateral changes and repayments transactions from the two fixed spread auction selected protocols, Aave v2 and Compound v2. It is crucial to notice how MakerDAO has been left outside of this analysis, given the structural differences in how the protocol has been built presented in chapter 4, with a diverse liquidation system based on auctions explained in chapter 5.

Given the different release date times of the latest version of the selected protocols, with Aave v2 being released in late 2020, it has been chosen to extract and aggregate all the borrowers' addresses whose first borrow is greater than the first of January 2021. More specifically, those addresses that have borrowed from either Compound v2 or Aave v2, that, as Table 6.1 shows, amounts to 40691 unique borrowing addresses, where 72% are unique borrowers of Aave, 23% are unique borrowers of Compound and 5% are both borrowers of Aave and Compound.

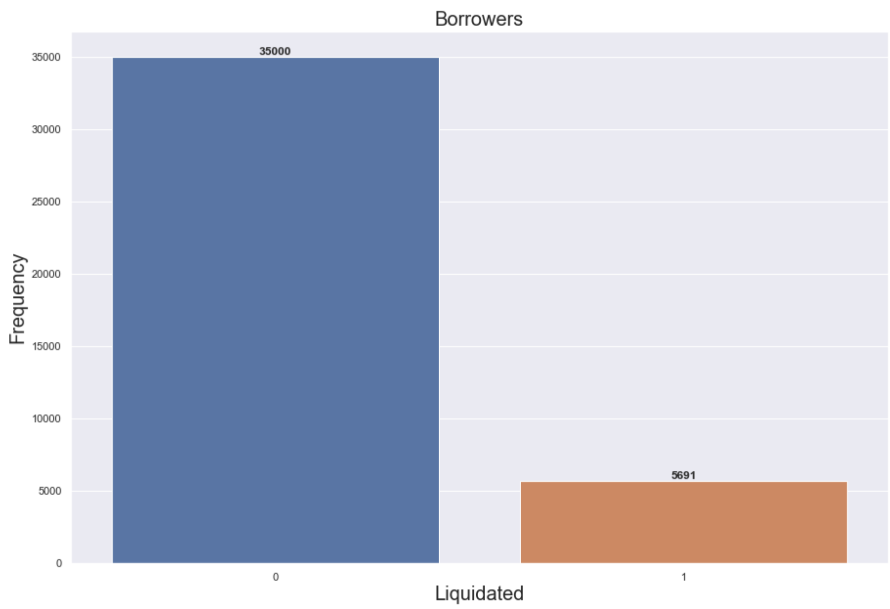
Table 6.1: Dataset Statistics Overview

	Borrowers	**Percentage**
Aave	29059	72%
Compound	8801	23%
Both	2381	5%
Total	40691	100%

Both the inference and the prediction tasks aim at looking at users that have never been liquidated (the unliquidated borrowers) and users that have only once, grouping their actions and investing strategies before the first liquidation event. This approach has been taken to try to mitigate any structural difference between the typologies of users, trying to capture similar users' characteristics instead. Therefore, given the possibility for a user to be liquidated multiple times, in the data aggregation, it has been considered to aggregate the liquidated users' information according to the history before the first liquidation event.

To sum up, to try to capture similar users, similar time periods and similar investing strategies, three elements have been chosen in aggregating the data. First, similar protocols: Aave and Compound on the Ethereum Blockchain, with the same typology of liquidation mechanism, assuming that the typology of users that engage in these two platforms are similar. Secondly, similar period: users that have borrowed for the first time from 2021 until August 2022, aiming to control for the same market conditions (drops in ETH prices, bull or bear runs). Thirdly, similar investing strategies: first-time borrowers that have never been liquidated altogether with first-time borrowers' actions before the first liquidation, trying to predict and make inferences on what brings a user to be liquidated in the first place. Finally, the ultimate dataset counts 40691 borrowers, of which 5691 have been liquidated, as Figure 6.1 shows.

Figure 6.1: Dataset Composition



6.2 Feature Engineering

Different features have been created coming from the borrowing behaviour of the users, the changes in collateral, supplying and withdrawing assets, and the debt repayment transactions. The computed features are summarized in Table 6.2, and a brief description is given when necessary.

Table 6.2: Extracted Features

Features	**Description**
Percentage Stablecoin Borrowed	% stablecoin over total borrowed
Distinct Asset Borrowed	-
Percentage Changes ETH	% supply/reedem of ETH assets over total
Distinct Asset used as Collateral	-
Balance Positive	Sum of supply/redeem of collateral action 1 if Positive 0 if Negative
Percentage of Stablecoin Repaid	% stablecoin over total assets repaid
Days between last borrow and supply	-
Project Compound	Borrowing started in Compound
Project Aave	Borrowing started in Aave

6.3 Inference

Firstly, the problem is observed from an inferential perspective where the probability for an address to be liquidated is modelled as a linear function of the regressors mentioned above. The aim of this analysis is mainly to understand the role of the borrowing protocol's choice and if it affects the target variable positively or negatively. Multivariate logistic regression is therefore applied, and the results are summarized in Table 6.3. As can be seen, all the coefficients are statistically significant, choosing a significance level alpha of 5%. It is also interesting to see that the sign of the coefficient of borrowing by using the protocol Compound is positive, meaning that if everything stays constant, the probability (measured in odds ratio) for an address of being liquidated increases, even though by a small amount. On the other hand, by looking at the coefficient of the Aave project and noticing its negative sign, still assuming everything else is held constant, the probability for an address to be liquidated decreases. In addition, inspecting the coefficients of the control variable, such as the balance, it can be seen that a positive balance decreased the probability of being liquidated, whereas mainly borrowing stablecoins (which can be interpreted as using volatile addresses as collateral) increases it, still considering everything else held constant. Moreover, by looking at the adjusted R-squared, it can be noticed that the (linearly) explained variability is only 13%, suggesting that the relationship between the target variable and chosen regressors is not linear.

Table 6.3: Results Logistic Regression

Dep. Variable:	y	**R-squared:**	0.134
Model:	OLS	**Adj. R-squared:**	0.134
Method:	Least Squares	**F-statistic:**	698.2
Date:	Sat, 08 Oct 2022	**Prob (F-statistic):**	0.00
Time:	17:26:07	**Log-Likelihood:**	-11728.
No. Observations:	40691	**AIC:**	2.348e+04
Df Residuals:	40681	**BIC:**	2.356e+04
Df Model:	9		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**
const	0.1836	0.009	20.166	0.000
perc_stablecoin_borrowed	0.2546	0.005	50.682	0.000
distinct_asset_borrowed_total	0.0290	0.002	16.087	0.000
project_Aave	-0.0983	0.008	-12.721	0.000
project_Compound	0.0327	0.007	4.586	0.000
perc_changes_eth	-0.1867	0.004	-45.579	0.000
distinct_asset_used_as_collateral	-0.0069	0.001	-5.815	0.000
balance_positive	-0.0472	0.004	-12.422	0.000
n_days_last_borrow_and_supply	-0.0005	1.66e-05	-27.394	0.000
perc_stablecoin_repayed	-0.2514	0.004	-61.244	0.000

Omnibus:	12108.874	**Durbin-Watson:**	1.869
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	27567.747
Skew:	1.740	**Prob(JB):**	0.00
Kurtosis:	5.039	**Cond. No.**	866.

The following section proposes a predictive approach by using well-known machine learning strategies in modelling the probability of being liquidated.

6.4 Forecasting

Three different models have been selected as predictive strategies, and the results are summarized in Table 6.4. In particular, the selected models are: as a benchmark, a logistic regression with L2 penalty term, and two ensemble models: a random forest and gradient boosting. It needs to be specified that for all these models, a cross-validation approach has been adopted to tune the hyperparameters by monitoring the out-of-sample error, which can be found in the corresponding notebook[1]. Given the unbalanced data, the metric reported to assess the performance of each classifier will not be the accuracy but the f1 score, which combines the precision and recall by taking their harmonic mean. Moreover, considering this binary classification problem, the threshold is optimized.

Table 6.4: Models Performance

	Macro Avg	**F1 Score**	**Threshold**
Logistic Regression (L2 loss)	0.67	0.46	0.54
Random Forest	0.78	0.70	0.57
Gradient Boosting	0.85	0.74	0.37

As it can be seen from results shown above in Table 6.4, the gradient boosting approach substantially improves the quality of predictions compared to the linear case,

[1] `https://github.com/alessandrocaruso/defi_liquidations_thesis`

bringing the f1 score to 0.74 on unseen data and delivering a macro average of precision and recall of 0.85, the optimized threshold is 0.37. The complete analysis with the confusion matrix and optimization of the described models can be found in the notebook on the dedicated GitHub page.

6.5 Results and Interpretation

To sum up, this chapter proposed statistical methods to study the liquidation phenomena, looking at it firstly from an inferential perspective, aiming at understanding the sign of the choice of the two studied protocols, and secondly proposing an algorithmic approach for predicting the probability of being liquidated for a user, based on features related to transactional data.

Regarding the choice of protocols, the empirical results confirm the qualitative assessment proposed in chapter 4 and chapter 5 when analyzing the two different protocols, suggesting that Aave v2 is, on average, a better choice compared to Compound v2 as a borrowing platform, given the statistical validity due the signs and magnitudes of the regression coefficients.

On the other hand, by proposing a predictive approach, promising results are achieved. A user could combine statistical and programming knowledge to create a liquidation bot that tends to focus on those sets of addresses that are more likely to get liquidated by their platform, introducing a new strategy compared to previously described front-running and back-running. This bot will not more expensively grid-search all the open positions looking for one address to suddenly become liquidatable, but give attention to different users depending on their assigned score, aiming at beating competing bots.

Chapter 7

Future Applications and limitations

The borrowing and lending DeFi landscape do not stop in the three analyzed protocols on the Ethereum Blockchain. Ethereum is one of the most important platforms that allow the creation of dApps (Decentralized applications), but it is not the only one. The Ethereum Mainnet blockchain is costly, slow, and has limits, as the base layer (Layer 1.0) can only process around 15 transactions per second (TPS), and it is not enough to support the massive growth of DeFi. This is why Layer 2.0[1] blockchain solutions have been built that aim at extending the Ethereum blockchain (base layer) by making it less costly and taking care of the scaling part, dramatically increasing the number of TPS, through systems such as rollups, solving the well-known congestion problem of the PoW Ethereum blockchain. These solutions, such as Polygon, Arbitrum One and Optimism, have gained popularity and have encouraged dApps developers to build their solutions on top of these systems directly, sometimes even entirely bypassing the Ethereum mainnet. Popular PLFs such as Dydx or Aave v3 are by design only working on top of these Layer 2.0 solutions, exploiting the higher number of TPS and taking advantage of the lower gas fees by proposing more exciting and innovative investment solutions. The future of liquidations needs to be assessed by considering the bigger picture and not just the Ethereum blockchain: understanding if borrowing has changed in these alternative solutions and, consequently, assessing if the liquidation phenomena have been affected. Moreover, given the fast development of DeFi and DLTs solutions,

[1]https://ethereum.org/en/layer-2/

this paper could not include in the analysis the impact of the release of Compound V3 (happening in August 2022, through a Twitter[2] announcement) and the effect of "The Merge" on liquidations, an iconic event[3] for the whole blockchain ecosystem, happening on the 15th of September where Ethereum upgraded from a proof of work (PoW) consensus mechanism to proof of stake (PoF).

Additionally, related to the data aggregation, more features could have been created, primarily related to the behaviour of a user outside the lending platforms, trying to map the public balance available in its address, achieving a comprehensive understanding of the likelihood for a borrower to maintain its health factor greater than one. Furthermore, lastly, the statistical procedures adopted are working on aggregated data without looking at the time dimension and the sequence of the transactions. In this context, recurrent neural networks (RNN) could be used for time series classification in the context of liquidations.

[2] https://twitter.com/compoundfinance/status/1562969434360549378?s=20&t=nHUScYjBHjH1QhsM3ZNR_Q

[3] https://ethereum.org/en/upgrades/merge/

Chapter 8

Conclusion

The rise of decentralized finance has opened numerous profitable opportunities that only a few agents fully exploit. Lending and borrowing have never been so accessible compared to traditional finance. Every day, new exciting and innovative applications are developed, enlarging the gap between those inside and outside the decentralized world. This paper presents one approach that tries to achieve convergence between two worlds running in parallel but never mentioned in the same context. A more consolidated world, composed of statisticians, data scientists and analytical enthusiasts aiming to discover hidden patterns and relationships in data, developing the next-generation algorithms that exploit the enormous amount of information collected daily. On the other hand, the world of computer scientists, financial rebels, and decentralized believers aims at rethinking any source of an organization whose purely service is ensuring trust between two unknown agents: such as banks, insurance companies and financial institutions. Remarkably, this paper shows the centralization in a hidden space of a claimed decentralized world, where few agents exploit tremendous arbitrage opportunities and make enormous profits. Moreover, after comparing the main available protocols for loanable funds (PLFs) on the Ethereum blockchain and their liquidation systems, by taking an inferential approach, it recommends what platform a first-time user should choose based on qualitative and quantitative assessments. Lastly, by taking a predictive approach, it aims at forecasting the likelihood for a borrowing user to be liquidated, proposing a strategy based on statistical evidence for the potential creation of a liquidation bot in screening the available, open and still healthy positions to beat the competition and increase the probability to obtain the liquidation prize.

List of Tables

List of Figures

Bibliography

[1] Deloitte. *DeFi deciphered: Navigating disruption within financial services*. March 2022. URL: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/risk/us-financial-advisory-defi-march-2022.pdf.

[2] Arjun Gupta Marc Christian Bender. *DeFi: An opportunity or an existential threat to banks?* June 2022. URL: https://www2.deloitte.com/au/en/blog/financial-advisory-financial-services-blog/2022/defi-opportunity-existential-threat-banks.html.

[3] Michael Del Castillo. *Blockchain 50: Billion Dollar Babies*. April 2019. URL: https://www.forbes.com/sites/michaeldelcastillo/2019/04/16/blockchain-50-billion-dollar-babies/?sh=3d7822d057cc.

[4] PwC. *Time for trust The trillion-dollar reasons to rethink blockchain*. October 2020. URL: https://www.pwc.com/blockchainsurvey.

[5] Jonathan Knegtel. *Could a Trillion Dollars be Waiting to Flood into DeFi From Institutions Over the Next 5 Years?* September 2021. URL: https://www.blockdata.tech/blog/general/could-a-trillion-dollars-be-waiting-to-flood-into-defi-from-institutions-over-the-next-5-years.

[6] Lewis Gudgeon et al. *DeFi Protocols for Loanable Funds: Interest Rates, Liquidity and Market Efficiency*. 2020. DOI: 10.48550/ARXIV.2006.13922. URL: https://arxiv.org/abs/2006.13922.

[7] Binance Academy. *What Is Leverage in Crypto Trading?* October 2022. URL: https://academy.binance.com/en/articles/what-is-leverage-in-crypto-trading.

[8] Thomas H. Davenport and DJ Patil. "Is Data Scientist Still the Sexiest Job of the 21st Century?" In: *Harvard Business Review* (July 2022). URL: `https://hbr.org/2022/07/is-data-scientist-still-the-sexiest-job-of-the-21st-century`.

[9] Satoshi Nakamoto. "Bitcoin: A Peer-to-Peer Electronic Cash System". In: *Cryptography Mailing list at https://metzdowd.com* (Mar. 2009).

[10] Michael Kuperberg. "Towards Enabling Deletion in Append-Only Blockchains to Support Data Growth Management and GDPR Compliance". In: *IEEE International Conference on Blockchain, Blockchain 2020, Rhodes, Greece, November 2-6, 2020*. IEEE, 2020, pp. 393–400. DOI: `10.1109/Blockchain50366.2020.00057`. URL: `https://doi.org/10.1109/Blockchain50366.2020.00057`.

[11] Quang Tung Thai et al. "Design and implementation of NDN-based Ethereum blockchain". In: *Journal of Network and Computer Applications* 200 (2022), p. 103329. ISSN: 1084-8045. DOI: `https://doi.org/10.1016/j.jnca.2021.103329`. URL: `https://www.sciencedirect.com/science/article/pii/S1084804521003143`.

[12] Matt Shealy. *How open source underpins blockchain technology*. October 2020. URL: `https://opensource.com/article/20/10/open-source-blockchain`.

[13] Donald E. Knuth. *The Art of Computer Programming, Volume 3: (2nd Ed.) Sorting and Searching*. USA: Addison Wesley Longman Publishing Co., Inc., 1998. ISBN: 0201896850.

[14] Abdul Khan, Sana Basharat, and Muhammad Riaz. *Analysis of asymmetric cryptography in information security based on computational study to ensure confidentiality during information exchange*. Oct. 2018. DOI: `10.13140/RG.2.2.30495.61602`.

[15] Leslie Lamport, Robert E. Shostak, and Marshall C. Pease. "The Byzantine Generals Problem." In: *ACM Trans. Program. Lang. Syst.* 4.3 (1982), pp. 382–401.

URL: http://dblp.uni-trier.de/db/journals/toplas/toplas4.html#LamportSP82.

[16] Daniel Perez, Jiahua Xu, and Benjamin Livshits. "Revisiting Transactional Statistics of High-scalability Blockchains". In: *Proceedings of the ACM Internet Measurement Conference*. ACM, Oct. 2020. DOI: 10.1145/3419394.3423628. URL: https://arxiv.org/pdf/2003.02693.pdf.

[17] Vitalik Buterin. *Ethereum: A next-generation smart contract and decentralized application platform*. Last Modified: 2022-10-07. 2014. URL: https://ethereum.org/en/whitepaper/.

[18] Gavin Wood. *Ethereum: A secure decentralised generalised transaction ledger EIP-150 REVISION (759dccd - 2017-08-07)*. Accessed: 2022-10-07. 2017. URL: https://ethereum.github.io/yellowpaper/paper.pdf.

[19] Elvira Albert et al. *GASOL: Gas Analysis and Optimization for Ethereum Smart Contracts*. 2019. DOI: 10.48550/ARXIV.1912.11929. URL: https://arxiv.org/abs/1912.11929.

[20] Sam M. Werner et al. *SoK: Decentralized Finance (DeFi)*. 2021. DOI: 10.48550/ARXIV.2101.08778. URL: https://arxiv.org/abs/2101.08778.

[21] Antonio Briola et al. "Anatomy of a Stablecoin's failure: The Terra-Luna case". In: *Finance Research Letters* (Sept. 2022), p. 103358. DOI: 10.1016/j.frl.2022.103358. URL: https://doi.org/10.1016%2Fj.frl.2022.103358.

[22] Gareth James et al. *An Introduction to Statistical Learning: with Applications in R*. Springer, 2013, pp. 17–21. URL: https://faculty.marshall.usc.edu/gareth-james/ISL/.

[23] Kaihua Qin et al. *CeFi vs. DeFi – Comparing Centralized to Decentralized Finance*. 2021. DOI: 10.48550/ARXIV.2106.08157. URL: https://arxiv.org/abs/2106.08157.

[24] Alexandra Born et al. "Decentralised finance – a new unregulated non-bank system?" In: *ECB Macroprudential Bulletin* (July 2022). URL: https://www.ecb.europa.eu/pub/financial-stability/macroprudential-bulletin/focus/2022/html/ecb.mpbu202207_focus1.en.html.

[25] Kaihua Qin et al. *Attacking the DeFi Ecosystem with Flash Loans for Fun and Profit*. 2020. DOI: 10.48550/ARXIV.2003.03810. URL: https://arxiv.org/abs/2003.03810.

[26] Kenny Zhu. "Mortgage Guide: What's the Average Time to Close on a House?" In: *valuepenguin.com* (Sept. 2022). URL: https://www.valuepenguin.com/mortgages/average-time-to-close-on-a-house.

[27] Zina Kumok and Jordan Tarver. "How Long Does It Take To Get A Personal Loan?" In: *Forbes* (Oct. 2021). URL: https://www.forbes.com/advisor/personal-loans/how-long-does-it-take-to-get-a-personal-loan/.

[28] Robert Leshner and Geoffrey Hayes. *Compound: The Money Market Protocol*. Version 1.0: 2019-02-01. 2019. URL: https://compound.finance/documents/Compound.Whitepaper.pdf.

[29] Aave Team. *Aave Protocol Whitepaper V2.0*. 2020. URL: https://github.com/aave/protocol-v2/blob/master/aave-v2-whitepaper.pdf.

[30] Kaihua Qin, Liyi Zhou, and Arthur Gervais. *Quantifying Blockchain Extractable Value: How dark is the forest?* 2021. DOI: 10.48550/ARXIV.2101.05511. URL: https://arxiv.org/abs/2101.05511.

[31] Kaihua Qin et al. "An empirical study of DeFi liquidations". In: *Proceedings of the 21st ACM Internet Measurement Conference*. ACM, Nov. 2021. DOI: 10.1145/3487552.3487811. URL: https://doi.org/10.1145%2F3487552.3487811.