



Article

# Performing Wash Trading on NFTs: Is the Game Worth the Candle?

Gianluca Bonifazi <sup>1</sup>, Francesco Cauteruccio <sup>1</sup>, Enrico Corradini <sup>1</sup>, Michele Marchetti <sup>1</sup>, Daniele Montella <sup>2</sup>, Simone Scarponi <sup>2</sup>, Domenico Ursino <sup>1,\*</sup> and Luca Virgili <sup>1</sup>

<sup>1</sup> Department of Information Engineering (DII), Polytechnic University of Marche, 60121 Ancona, Italy

<sup>2</sup> Inconsulting, 40033 Casalecchio di Reno, Italy

\* Correspondence: d.ursino@univpm.it

**Abstract:** Wash trading is considered a highly inopportune and illegal behavior in regulated markets. Instead, it is practiced in unregulated markets, such as cryptocurrency or NFT (Non-Fungible Tokens) markets. Regarding the latter, in the past many researchers have been interested in this phenomenon from an “ex-ante” perspective, aiming to identify and classify wash trading activities before or at the exact time they happen. In this paper, we want to investigate the phenomenon of wash trading in the NFT market from a completely different perspective, namely “ex-post”. Our ultimate goal is to analyze wash trading activities in the past to understand whether the game is worth the candle, i.e., whether these illicit activities actually lead to a significant profit for their perpetrators. To the best of our knowledge, this is the first paper in the literature that attempts to answer this question in a “structured” way. The efforts to answer this question have enabled us to make some additional contributions to the literature in this research area. They are: (i) a framework to support future “ex-post” analyses of the NFT wash trading phenomenon; (ii) a new dataset on wash trading transactions involving NFTs that can support further future investigations of this phenomenon; (iii) a set of insights of the NFT wash trading phenomenon extracted at the end of an experimental campaign.

**Keywords:** wash trading; NFTs; correlation; causality; distance; blockchain



**Citation:** Bonifazi, G.; Cauteruccio, F.; Corradini, E.; Marchetti M.; Montella D.; Scarponi S.; Ursino D.; Virgili L. Performing Wash Trading on NFTs: Is the Game Worth the Candle? *Big Data Cogn. Comput.* **2023**, *7*, 38. <https://doi.org/10.3390/bdcc7010038>

Academic Editor: Min Chen

Received: 9 January 2023

Revised: 13 February 2023

Accepted: 17 February 2023

Published: 21 February 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

A Non-Fungible Token (NFT, for short) is a digital content that represents real-world objects, such as artwork, music, games and collections of any kind. In a very short time, prosperous markets have developed around NFTs. Their evolution and dynamics follow the classic rules of any market [1,2], flanked by several peculiarities. Similar to any technology, NFTs present both opportunities and risks [3–5]. As for the latter, a very significant one concerns market manipulation, which aims to artificially increase the selling price of an asset. This practice is called wash trading and consists of a series of sell and buy transactions on the same asset performed by a single trader so as to generate a false interest in it, thereby artificially increasing its value. In regulated markets, wash trading is judged a highly improper and illegal behavior [6,7]. However, the cryptocurrency and NFT markets are not regulated; therefore, wash trading activities have been carried out within them several times in the past [8].

In the literature, several investigations concerning NFTs in general, and the wash trading phenomenon in particular, have been proposed. As for the former, studies are very varied in that they focus on different issues. Among them, we highlight: (i) the analysis and characterization of NFT ecosystems [9–11], (ii) the study of the role of social media (e.g., Twitter), in the context of NFTs [12], and (iii) the investigation of geopolitical risks and market factors in this scenario [13]. The wash trading phenomenon was analyzed from a variety of perspectives. However, all of them are “ex-ante”, that is, they aimed to identify and classify wash trading activities before or at the exact time they occur [8,14–16].

In contrast, there are no “ex-post” analyses on wash trading, i.e., investigations aiming to determine how much wash trading activities have already achieved success in increasing (at least apparently) interest in the NFT involved, leading to the subsequent increase of its value. Such a kind of analysis could have a variety of interesting applications. The first of them concerns the ability to understand whether an NFT having a current selling price much higher than its floor one (i.e., the lowest price at which it was sold in the past) has achieved such a success because of its characteristics or because it is the subject of speculation.

A second, even more important, application is the following. The practice of wash trading is illicit, improper, and potentially harmful (at least to the reputation of those who perform it, if they are found out). For this potential very high cost, what are the benefits? In other words: is the game worth the candle? A study by the company Chainalysis, whose results were published in a report in February 2022 (<https://blog.chainalysis.com/reports/2022-crypto-crime-report-preview-nft-wash-trading-money-laundering/> (accessed on 25 November 2022)), gives us a clue that this question is not all that far-fetched. In fact, this study points out that most wash trading tasks on NFTs did not report high profits in the past. Only in a few cases was there a significant profit. This report certainly is an important clue to answering the previous question. However, it is not the result of a “structured” study, taking into account the intrinsic meaning of the various features that characterize an NFT. Being able to conduct such a kind of study would provide insight into whether there are correlations, causal relationships, or other forms of relationships among the features characterizing an NFT that, if taken into account, could lead a wash trading activity on it to significantly increase its value. Conversely, if there are no such relationships, a wash trading activity on an NFT will not be able to impact its value significantly, so the game will not be worth the candle. Analyzing the wash trading phenomenon in a “structured” way and “ex-post” to determine whether it is worth the candle is the main objective of this paper.

An NFT (In this paper, when we talk about an NFT, we mean a collection of NFTs, all of the same type; therefore, in the following, we will use the terms “NFT” and “NFT collection” interchangeably.) is defined by several features, such as sales volume, price, owner number, etc. The values of these features can change over time; as a consequence, our analysis will have to take the time component into account. In addition, as mentioned above, the various features can be related to each other by several relationships, e.g., correlation, causality, distance, etc. In the following, we will use the term “facets” to refer to such relationships, because they are able to capture different dimensions of the same phenomenon. These facets, considered for all or part of the features of an NFT, allow us to perform a “structured” analysis of the wash trading phenomenon. Nevertheless, this analysis, and those derived from it that we present in this paper, will not be possible if we do not have a congruent model available for representing NFTs. Therefore, our paper is also concerned with defining such a model.

The first motivation behind our paper is the desire to define a structured framework that allows for an “ex-post” analysis of the NFT wash trading phenomenon, and to apply it to verify if the game is worth the candle. The second motivation is to define a new data model, capable of providing a rich and comprehensive representation of NFT transactions, as well as to support analyses of the phenomenon of interest from multiple facets. The third motivation is to build a dataset that can support our framework in the analyses of the NFT wash trading phenomenon and other investigations that we and/or other researchers plan to conduct in the future. The fourth and final motivation is to enrich the knowledge of the wash trading phenomenon already derived in the literature with new insights gained thanks to our framework and dataset.

Specifically, the main contributions of this paper are:

- A framework to support the investigation of NFT-related phenomena. Specifically, this framework consists of: (i) a model to represent NFTs, their features and the evolution of the latter over time; (ii) a set of facets, which represent relationships among features.

The framework is general and allows for an easy expansion of both the model and the set of facets in the future.

- A new dataset regarding the NFT scenario. Specifically, this dataset records wash trading transactions made on some NFTs in the past. It holds substantial importance since, to the best of our knowledge, there are no other datasets that include all the information needed to analyze the wash trading phenomenon “ex-post”. We had to derive such a dataset for our analysis and we made it freely available to all those researchers who want to perform similar analyses and develop novel approaches for an “ex-post” investigation of the NFT wash trading phenomenon.
- The illustration of the results of an experimental campaign in which we applied the proposed framework to the dataset we constructed in order to answer the question representing the core of this paper, and several other ones related to it.

The outline of this paper is as follows: in Section 2, we review the related literature. In Section 3, we present our framework. In Section 4, we describe the dataset we constructed. In Section 5, we illustrate the experiments we conducted. In Section 6, we present a discussion regarding our framework and its ability to achieve the goals for which it was designed. Finally, in Section 7, we draw our conclusion and outline some possible future developments of our research efforts in this field.

## 2. Related Work

In recent years, thanks to the development of blockchain technology, the use of NFTs to represent and identify the ownership of a unique asset has become popular. The versatility characterizing NFTs has facilitated their diffusion and use in many fields, leading researchers to address related challenges and needs [17–19]. In particular, the NFT market is young and has no shared rules, which could lead to some problems [20,21]. One of the most important of these problems is market manipulation performed by malicious users. This problem also afflicts traditional markets, for example the stock one. However, in these contexts, there are already established techniques to detect such practices [22].

One market manipulation mechanism also widespread in the NFT context is wash trading [23,24]. This illicit practice consists of the repetition of buy and sell transactions performed by the same user. The goal is to fake a growth in demand for an asset, thereby attracting other investors and buyers to it by creating the illusion of a large business. Wash trading in traditional markets is illegal. However, this is not the case with NFTs, since this market is new and has little regulation because it has not yet been decided to which asset class NFTs belong. The ability to manipulate the digital asset market is increased by the fact that a user can make, buy or sell transactions simply by creating a wallet, without having to provide additional identifying information.

Wash trading has been extensively investigated in the literature for traditional markets [22]. Recently, the analysis of illegal practices in cryptocurrency markets has also attracted much interest [25–28]. In contrast, to the best of our knowledge, only a few studies analyze wash trading in the NFT market [8,14–16]. All of these techniques aim to identify and flag suspicious transactions that are thought to have been made for carrying out wash trading activities. Therefore, they aim to perform an “ex-ante” examination [14]. Unlike all these approaches, ours aim is to perform an “ex-post” analysis. In fact, it aims to study transactions that have already been identified as wash trading ones in order to verify if the game is worth the candle.

Among the “ex-ante” approaches, there is the one described in [29], which aims to identify anomalous transactions based on the probability of making a profit. Specifically, the approach considers as anomalous those transactions that made a profit whose probability was below a predetermined threshold. In [14], the authors introduce three new strategies to detect potential wash trades in the NFT market. The first, called “Closed loop token trades”, checks whether the pair of wallets making the transactions are involved in trading the same token. The second, called “Closed loop value trades”, checks whether the same value is transacted across multiple tokens. The third, called “high transaction volumes”, checks

for transactions handling large volumes of cryptocurrency between recently generated accounts. It is particularly suited for detecting automated wash traders.

Other papers aimed to understand what generates fame and price around NFTs and whether there is a mechanism that can control them [30–35]. For example, the authors of [32] analyze the relationship between NFT sales and NFT active wallets, focusing on two cryptocurrencies, namely Bitcoin (BTC) and Ether (ETH). To reach their objective, they perform a Granger causality test, which aims to verify whether the change in a feature produces a change in another one. Using this approach, they find that the BTC price Granger-causes the NFT sales, while the ETH price Granger-causes the number of NFT active wallets. Thus, they show that large cryptocurrency markets affect small NFT markets, while the opposite effect does not happen. The approach described in [32] and ours share the use of the Granger causality test to investigate the relationships between different features in the context of NFTs. However, the purpose of the approach of [32] is the identification of the relationships existing between BTC and ETH on the one hand, and NFTs on the other hand. Therefore, [32] does not deal with wash trading, which is the core of the analysis performed by our approach. In [31], the authors propose an approach whose goal is very similar to the approach of [32], in that they also want to understand whether the price of NFTs is correlated with that of cryptocurrencies. Their result indicates that NFTs represent a class of assets weakly correlated with cryptocurrencies.

In [5], the authors propose a study to assess whether or not the wash trading phenomenon has a significant presence in the NFT market. From this point of view, the objective of [5] and that of our paper are certainly related, although quite different from each other. Indeed, it could be the case that there is a massive presence of the phenomenon and yet it turns out in the end that the game is not worth the candle, or vice versa.

Finally, there are papers that analyze the effects of wash trading within the NFT market. Our paper is related to them in that it analyzes what changes in the market are brought about by wash trading and whether this practice benefits people performing it.

The authors of [15] show that much of wash trading is unprofitable because of high transaction costs. However, in those cases where such activities are successful, they manage to create a high profit. The authors of [8] conduct a series of analyses to understand to what extent wash trading occurs in NFT markets and to what extent this practice distorts prices. They first create transaction graphs to identify suspicious transactions, and then use these graphs to understand how wash trades alter market performance. Their analysis reveals that an NFT for sale has on average an increased price after a flagged wash trade involving it. However, a regression on panel data, designed to measure the price impact, showed no significant change for the majority of NFTs. While the approach of [8] and our own share one of their objectives, their ways of proceeding are completely different. In fact, the approach of [8] focuses on suspicious transactions, while our approach aims to determine whether a set of relationships exists between the features characterizing wash trading activities.

At the end of this discussion, it seems interesting to us to point out that there are research topics on NFTs that are close to wash trading and, at the same time, orthogonal to it. For example, in [4], the authors analyze the emerging Flash Loan service. This allows users to apply for non-collateral loans within the decentralized finance ecosystem. It, too, could in principle be used for both legitimate and illegitimate purposes. For example, wash traders can use it to manipulate the market without a large amount of capital as long as they can afford the potential loss and the gas fee to make transactions.

### 3. A Framework to Support the Investigation of NFT-Related Phenomena

In this section, we propose a framework to model NFTs and investigate phenomena involving them. Specifically, in Section 3.1, we illustrate a model for representing a set of NFTs and its evolution over time. In Section 3.2, we propose a set of operators (called facets) that allow us to evaluate the mutual behavior of a set of NFTs involved in one or more phenomena we want to analyze. We recall that the phenomenon under investigation in

this paper is wash trading. As a consequence, the facets that we propose in Section 3.2 are well-suited for studying this phenomenon. However, our framework is easily extendable so that one or more of these facets can be removed and other facets can be easily added in case we want to investigate different NFT-related phenomena in the future.

### 3.1. A Model for NFTs

In this section, we illustrate our model for representing a set of NFT collections. An NFT collection is a set of NFTs of the same type created by the same author. In particular, let  $\mathcal{S} = \{C_1, C_2, \dots, C_l\}$  be a set of NFT collections of interest. We assume that all the NFT collections of  $\mathcal{S}$  are characterized by the same set  $\mathcal{F} = \{f_1, f_2, \dots, f_m\}$  of features. Let  $C_i$  be a generic NFT collection of  $\mathcal{S}$ . It can be represented as:

$$C_i = \langle Name_i, \mathcal{F} \rangle$$

In this case,  $Name_i$  is the name of  $C_i$ ; this name is unique. The set  $\mathcal{F}$  consists of numerical features whose values may vary over time. To model such variation, we borrow the concepts of time series analysis [36] and describe  $\mathcal{F}$  as a multivariate time series, that is, a set of contiguous observations, one for each feature of interest. For instance, some features of interest for an NFT collection might be the sales price, the number of sellers and the number of owners.

In our model, the evolution of features is crucial; therefore, it is essential to model time. For this purpose, given an overall interval  $T$  of interest, we can think of modeling it as an ordered sequence of  $z$  time slices,  $T = T_1, \dots, T_z$ . For example,  $T$  could be a certain month, say October 2022, and could be represented by a succession of 31 time slices, one for each day. It is necessary for our time representation to allow the indexing of the sequence of time slices so that we can select a particular interval of contiguous time slices of  $T$  (e.g., the second decade of October 2022). To this end, our model uses the notation  $T[x..y]$ ,  $1 \leq x \leq y \leq z$ , to denote the interval of contiguous time slices in  $T$  that begins at  $T_x$  and ends at  $T_y$ . If  $x = y$ , then it means that we want to select a single time slice; in this case, we will use the abbreviated notation  $T_x$  or  $T[x]$  to represent  $T[x..x]$ . If  $x = 1$  and  $y = z$ , then it means we are considering the overall interval of interest; in this case, we will use the abbreviated notation  $T$ , instead of  $T[1..z]$ , to denote this interval.

The previous notation about intervals and time slices can be extended to model the features of an NFT collection. Specifically, given a generic NFT collection  $C_i = \langle Name_i, \mathcal{F} \rangle$  of  $\mathcal{S}$ , we denote by  $\mathcal{F}[x..y]$  the trend of the values of  $\mathcal{F}$  assumed by  $C_i$  in the time interval  $T[x..y]$ . In turn,  $f_j[x..y]$ ,  $1 \leq j \leq m$ , denotes the trend of the value of  $f_j$  assumed by  $C_i$  in the time interval  $T[x..y]$ . Clearly,  $\mathcal{F}[x..y] = \{f_1[x..y], f_2[x..y], \dots, f_m[x..y]\}$ . We will use the notation  $\mathcal{F}[x]$ , or  $\mathcal{F}_x$ , (resp.,  $f_j[x]$ , or  $f_{j,x}$ ) to represent the values of  $\mathcal{F}$  (resp., the value of  $f_j$ ) assumed by  $C_i$  during the time slice  $T_x$ , i.e.,  $\mathcal{F}[x] = \mathcal{F}[x..x]$  (resp.,  $f_j[x] = f_j[x..x]$ ). Finally, we will use the abbreviated notation  $\mathcal{F}$  (resp.,  $f_j$ ) to denote the trend of the values of  $\mathcal{F}$  (resp., the value of  $f_j$ ) assumed by  $C_i$  during the overall time interval  $T$ , i.e.,  $\mathcal{F} = \mathcal{F}[1..z]$  (resp.,  $f_j = f_j[1..z]$ ).

As a consequence of the previous modeling,  $\mathcal{F}$  (resp.,  $f_j$ ) represents a multivariate (resp., univariate) time series.

### 3.2. The Selected Facets

In this section, we examine the facets used by our framework to investigate the wash trading phenomenon for NFTs. They are correlations, causal relationships, and distances between features. They are capable of capturing and evaluating different aspects of the phenomenon under consideration. However, we recall that our research framework is easily extendable so that some of these facets could be removed and other ones could be easily added in the future should it be deemed appropriate.



### 3.2.1. Correlation

The first facet considered in our research framework is the correlation occurring between different features related to NFT transactions (e.g., sales, floor price, wash sales, active wallets). Correlation analysis is carried out through two techniques: the first determines the correlation degree of a pair of features by computing the Pearson coefficient between the corresponding time series [37]. More specifically, suppose  $f_j$  and  $f_k$  are the two features of interest. Furthermore, suppose that the reference time interval for the time series is  $T[1..z]$ . Let  $f_{j_x}$  (resp.,  $f_{k_x}$ ) be the value of  $f_j$  (resp.  $f_k$ ) in the time slice  $T_x$ ,  $1 \leq x \leq z$ . Let  $\bar{f}_j$  (resp.,  $\bar{f}_k$ ) be the mean of the values of  $f_j$  (resp.,  $f_k$ ) in the time interval  $T$ . The Pearson coefficient is defined as:

$$\rho_{jk} = \frac{\sum_{x=1}^z ((f_{j_x} - \bar{f}_j) \cdot (f_{k_x} - \bar{f}_k))}{\sqrt{\sum_{x=1}^z (f_{j_x} - \bar{f}_j)^2} \cdot \sqrt{\sum_{x=1}^z (f_{k_x} - \bar{f}_k)^2}} \quad (1)$$

The computation of the correlation using the Pearson coefficient is straightforward and can already provide interesting insights. However, it does not take into account the cross-correlation between the time series involved, i.e., the possibility that two features are correlated, but one of them lags behind the other. The second correlation coefficient we use takes this fact into account. In order to define it, we must first standardize the time series associated with features. Standardization is achieved by subtracting the corresponding mean from each value and dividing the result obtained by the corresponding standard deviation. At this point, it is possible to calculate the cross-correlation between two features by considering the  $x$ th element of one and the  $(x + h)$ th element of the other. The parameter  $h$  represents the lag that we want to consider. In principle,  $h$  could range from 0 to  $z - x$ . In practice, low values of  $h$  are chosen; for example, we considered values of  $h$  less than or equal to 2 in our experiments.

By analyzing the cross-correlation between two features, it is possible to understand whether one of them can be used to predict the values of the other. Given two features, after computing the cross-correlation for different values of  $h$ , we select the maximum value as the overall cross-correlation coefficient between the two features. This policy allows us to capture the maximum similarity existing between the two features.

Correlation represents a facet of our approach, i.e., a point of view through which to study the phenomenon of interest. Specifically, once we standardize the time series associated with features, it is able to show us the similarity degree of each pair of them regardless of the lag, whether or not that similarity is due to a cause-and-effect phenomenon.

### 3.2.2. Causality

The second facet considered in our research framework takes into account the causal relationship that may exist between two features. Unlike correlation, here we are really looking for a cause-and-effect relationship between a pair of features such that the values of one of them depend on the values of the other. To compute a causal index between a pair of features  $f_j$  and  $f_k$ , we first consider the test proposed by Granger [38]. This test is based on the idea of comparing the ability to predict the values of  $f_k$  using all the information in the universe, denoted by  $\mathcal{U}$ , with the ability to predict the same values using all the information in  $\mathcal{U}$  except the values of  $f_j$ . We will adopt the notation  $\mathcal{U}|f_j$  to denote the latter. If discarding  $f_j$  reduces the predictive power on  $f_k$ , then it means that  $f_j$  has unique information about  $f_k$ . In this case, we say that  $f_j$  Granger-causes  $f_k$ . In order to apply the Granger test, it is necessary to ensure that the series are stationary. To verify it, one can adopt the Augmented Dickey–Fuller test [39]. If one or both series are not stationary, it is necessary to make them stationary through a differentiation process.

To carry out the Granger causality test, it is necessary to compare the values returned by a first linear model, called restricted, obtained by means of an autoregressive model (Vector AutoRegression—VAR) [40] of the time series of  $f_k$ , with the values obtained by a second linear model, called unrestricted, which considers both the time series  $f_j$  and  $f_k$  [41].

Equation (2) represents the restricted model, while Equation (3) denotes the unrestricted one.

$$f_{k_x} = \gamma_0 + \sum_{h=1}^p (\gamma_h \cdot f_{k_{x-h}}) + e_x \quad (2)$$

$$f_{k_x} = \alpha_0 + \sum_{h=1}^p (\alpha_h \cdot f_{k_{x-h}}) + e_x + \sum_{h=1}^p (\beta_h \cdot f_{j_{x-h}}) + u_x \quad (3)$$

In these equations:

- $f_{k_x}$  denotes the value of the feature  $f_k$  at time  $T_x$ ;
- $\gamma_h$ ,  $\alpha_h$  and  $\beta_h$  indicate linear coefficients;
- $p$  represents the maximum number of lagged observations included in the model;
- $e_x$  and  $u_x$  denote time-dependent noise terms;
- $f_{k_{x-h}}$  and  $f_{j_{x-h}}$  indicate the values of the features  $f_k$  and  $f_j$  at time  $T_{x-h}$ .

Equation (2) assumes that  $f_k$  depends linearly only on its past values and the noise term  $e_x$ . Equation (3) considers  $f_k$  to be linearly dependent on the past values of  $f_j$  and  $f_k$  and two time-dependent noise terms  $e_x$  and  $u_x$ .

Similarly to what we have seen for the refined computation of the correlation of two features in Section 3.2.1, the computation of causality values is performed considering a certain number  $p$  of lagged observations. In the experimental tests, we set  $p$  less than or equal to 3.

Being a statistical test, it is important to define the null hypothesis. In our case, it is as follows:  $H_0$ : “ $f_j$  does not Granger-cause  $f_k$ ”. This is equivalent to saying that  $\beta_h = 0$ ,  $1 \leq h \leq p$ , in Equation (3). The  $p$ -value resulting from the test determines the acceptance or the rejection of the null hypothesis.

Equations (2) and (3), used previously in the Granger test, express a linear model and, therefore, can be used if the features  $f_j$  and  $f_k$  exhibit linear trends. If this is not the case, the Granger test can still be applied, but the linear models represented by Equations (2) and (3) must be replaced with nonlinear ones. Neural networks, which can handle nonlinearity, can be employed for this purpose. For instance, two neural networks suitable for this purpose are the Multi-Layer Perceptron (MLP) and the Long-Short Term Memory (LSTM) models [42,43]. These nonlinear models are used as restricted and unrestricted models in the Granger test. As in the previous case, the resulting predictions of the two models are compared in order to derive the possible existence of the causality relationship between the input features.

Causality is an important facet. In fact, it represents a stronger relationship than correlation. In other words, if  $f_j$  causes  $f_k$ , we can say not only that  $f_k$  is correlated with  $f_j$ , but also that the values of  $f_k$  depend on the values of  $f_j$  or, also, that the knowledge of the values of  $f_j$  allows the prediction of the values of  $f_k$ . It is worth recalling that two features  $f_j$  and  $f_k$  might be correlated without one depending on the other. In this sense, correlation still remains a facet to be considered, regardless of causality.

### 3.2.3. Distance

The third and final facet considered by our research framework takes into account the distance between features. Since each feature has a time series associated with it, classical metrics, such as Euclidean, Manhattan or Minkowski distances, cannot be used to compute the distance between features because they do not take the temporal nature of data into account. In other words, they are unable to understand whether the two time series align but with a certain lag from each other. So, ad hoc distance metrics are needed to handle such a scenario. Among them, a well-known one is the Dynamic Time Warping (DTW) distance [44].

The DTW distance receives two time series  $f_j$  and  $f_k$  as input. These must be standardized. If they are not, it is first necessary to proceed with a standardization task by which the corresponding mean is subtracted from each element of the series and the result

obtained is divided by the corresponding standard deviation. As a first task, the DTW distance identifies the optimal alignment between the two series. This alignment is defined through the optimal alignment path  $\pi$ . This path can be represented as a succession of pairs  $(f_{j_u}, f_{k_v})$  such that  $f_{j_u}$  is the element of  $f_j$  that best aligns to  $f_{k_v}$ , and vice versa (see [44] for all details).

Once the optimal alignment path  $\pi$  has been identified, the DTW distance computes, for each pair  $(f_{j_u}, f_{k_v}) \in \pi$ , the difference  $(f_{j_u} - f_{k_v})$ , which represents the distance between the two values. Then, it calculates the square of that distance to eliminate the sign dependence. Finally, it sums the squares of the differences of each pair of  $\pi$  and computes the square root of the value thus obtained [45]. This whole process is represented in Equation (4).

$$DTW(f_j, f_k) = \sqrt{\sum_{(f_{j_u}, f_{k_v}) \in \pi} (f_{j_u} - f_{k_v})^2} \quad (4)$$

Distance is another important critical facet of our research framework. It can be computed for any pair  $(f_j, f_k)$  of features, regardless of the existence of a correlation or causal relationship between them. From this point of view, distance is a “universal” facet because it can be employed in any circumstance. However, it is clear that, at equal distances, if two features  $f_j$  and  $f_k$  are also related by correlation or even causality, the link between them is much stronger, and this should be taken into account. The adoption of the correlation and causality facets introduced above is for just that purpose. This reasoning reinforces the idea that a multi-facet approach to the analysis of the wash trading phenomenon can provide more accurate results than an approach that considers only one form of relationship between features.

#### 4. Dataset

In order to investigate the phenomenon of NFT wash trading, we had to build a dataset from scratch, capable of supporting our investigation. This was necessary because there were no ready datasets in the literature that could be used for our analysis. In fact, the few datasets available referred only to the most popular collections of NFTs or only to a small number of specific features or, again, they covered too short time periods. More relevantly, most of these datasets were designed to search for suspicious wash trading transactions and did not provide information on the volume of money traded, which is important for the objectives of our analysis.

We built our dataset by extracting data from Cryptoslam (<http://criptoslam.io> (accessed on 24 November 2022)). This is an aggregator of NFT collectibles data derived from the Ethereum, WAX, and FLOW blockchains. According to the Cryptoslam website policy, we created a Python script that downloads all NFT sales data recorded by the Cryptoslam explorer. We downloaded data on all NFT sales carried out from the entry of each NFT collection into the market to the last day of our data extraction task, i.e., 7 February 2023. Proceeding in this way, we were able to extract many NFT collections. Afterward, we selected the top 2000 of them, indexed on the basis of their USD sales volumes. In this way, we constructed a ranking from which to draw the NFT collections to be considered for further analysis. In addition, we downloaded daily ETH/USD exchange rate data from Yahoo Finance in order to standardize the money amounts within the dataset, because some of them were expressed in USD and others in ETH. For each NFT collection, Cryptoslam explorer (<https://cryptoslam.gitbook.io/docs/> (accessed on 25 November 2022)) identifies and reports wash trading sales and transactions, which are stored in our dataset. The complete dataset has 672,098 rows, with a mean of 336 rows for each NFT collection. A sample of 112,512 rows of the dataset is available for further experiments at the following GitHub address: <https://github.com/Bored3DS/NFT-Scrap> (accessed on 24 November 2022).

After constructing the dataset, we performed ETL operations on it to facilitate the next data analysis activities. The first step in this operation concerns the identification of the



features of NFT collections useful for studying the phenomenon of interest. At the end of this step, we identified the following variables:

- *Name*: it represents the name of the NFT collection to which all the next features refer;
- *Date*: it indicates the day to which all the next features refer;
- *Sales*: it denotes the sales volume expressed in USD;
- *ETH*: it represents the ETH/USD exchange rate;
- *Floor Price*: it indicates the minimum price of an NFT in the collection;
- *Active Wallets*: it denotes the number of active wallets;
- *Sales Txns*: it represents the number of sales transactions;
- *Total Owners*: it indicates the total number of owners;
- *Wash Sales*: it indicates the volume of wash sales expressed in USD;
- *Wash Txns*: it denotes the number of transactions made for performing wash sale activities.

All these variables, except for the first (representing the name of the NFT collection) and the second (denoting the time slices of interest), form the set  $\mathcal{F}$  of features of our interest.

The second step of the ETL activity we carried out on the dataset regarded the homogenization of time slices. In fact, different features were surveyed at different cadences. In the end, we decided to use the daily cadence as the reference one. A snapshot of the dataset obtained after all these tasks is shown in Figure 1.

Name	Date	Sales	ETH	Floor Price	Active Wallets	Sales Txns	Total Owners	Wash Sales	Wash Txns
3landers	2022-03-15	450336.43	2620.149658	1.8000	204	78	2610	13122.92	3
3landers	2022-03-16	349155.43	2772.055664	1.6800	201	71	2629	0.00	0
3landers	2022-03-17	244921.73	2814.854492	1.5000	145	48	2646	4572.37	1
3landers	2022-03-18	494468.93	2945.343018	1.7000	187	95	2669	0.00	0
3landers	2022-03-19	779696.11	2946.257080	2.0500	237	116	2695	0.00	0
3landers	2022-03-20	585574.21	2860.459229	1.9790	245	87	2724	16090.52	1
3landers	2022-03-21	392488.08	2897.976562	1.8000	224	66	2755	13761.08	2
3landers	2022-03-22	555448.44	2973.131104	1.9390	221	92	2780	0.00	0
adam-bomb-squad	2021-09-14	241354.99	3429.169678	0.2200	236	136	3008	2135.69	3
adam-bomb-squad	2021-09-15	268383.54	3615.282715	0.2100	259	151	3033	1484.46	2
adam-bomb-squad	2021-09-16	305900.51	3571.294922	0.2099	263	157	3057	819.91	1
adam-bomb-squad	2021-09-17	227620.82	3398.538818	0.2190	263	136	3078	0.00	0
adam-bomb-squad	2021-09-18	259199.37	3432.018311	0.2200	248	156	3098	781.37	1

Figure 1. A snapshot of the reference dataset after the ETL tasks.

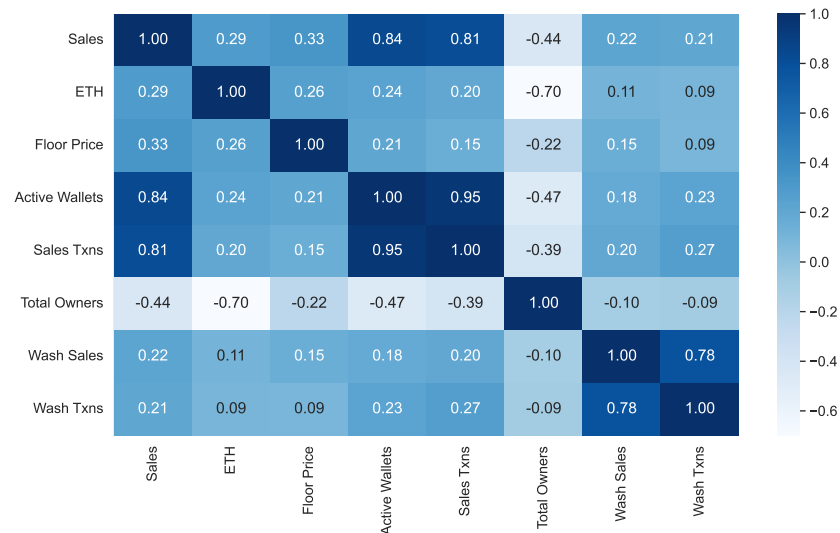
## 5. Experiments

In this section, we apply the model and facets described in Section 3 on the dataset introduced in Section 4. Our ultimate goal is to address the challenge underlying this paper, which is to understand whether, in the case of NFT wash trading, the game is worth the candle. In the Introduction, we said we wanted to answer this question by means of a “structured” study, which takes into consideration the intrinsic meaning of the various features that characterize an NFT, and especially the relationships between them. These relationships can be expressed through the three facets we introduced in Section 3.2, namely: correlation, causality and distance. Therefore, we divide the experiments in this section into three groups, one for each facet.

### 5.1. Experiments on Correlation

The first group of experiments focused on the first facet we introduced in Section 3.2, namely correlation. Regarding this facet, we saw that there are two different forms of correlation to consider, namely the Pearson and cross-correlation coefficients.

We started our analysis by computing, for each NFT collection, the Pearson coefficient relative to each pair of features of interest. Then, we averaged the corresponding values on all available NFT collections. The results obtained are shown in Figure 2.



**Figure 2.** Values of the Pearson coefficient averaged on all available NFT collections.

From examining this figure, we can see that the Pearson coefficient generally has medium-low values except for the following pairs of features:

- (Sales, Active Wallets): in this case, the coefficient has a value of 0.84. This correlation indicates that the sales volume of a certain collection tends to increase as the number of active wallets referring to it increases.
- (Sales, Sales Txns): in this case, the coefficient has a value of 0.81. This correlation indicates that an increase in the number of sales transactions leads to an increase in the corresponding volumes.
- (Sales Txns, Active Wallets): in this case, the coefficient has a value of 0.95. This correlation tells us that increasing the number of active wallets also leads to an increase in the number of sales transactions. The two trends are very strongly correlated.
- (ETH, Total Owners): in this case, the coefficient has a value of  $-0.70$ . This correlation tells us that a decrease in the ETH/USD exchange rate leads to an increase in the number of NFT owners.
- (Wash Sales, Wash Txns): in this case, the coefficient has a value of 0.78. This correlation tells us that an increase in wash trading transactions leads to an increase in the corresponding volume of wash sales.

All the correlation pairs highlighted above have an immediate intuitive explanation except for the pair (ETH, Total Owners). In this case, an explanation can be identified by means of the following reasoning: A decrease in the ETH cost leads to fewer USDs needed to purchase an NFT from the collection. So, more people can afford to buy those NFTs, which leads to an increase in the number of total owners.

In order to perform a deeper analysis of the correlation of features, we investigated their cross-correlation. Specifically, as mentioned in Section 3.2.1, we considered three cases (It is worth pointing out that, when  $h = 0$ , the cross-correlation coincides with the Pearson coefficient.) corresponding to a number  $h$  of lags equal to 0, 1 and 2. For each of these cases, we computed the value of the cross-correlation for each pair of features and for each NFT collection. Finally, we averaged the values thus obtained over all the NFT collections. The results are shown in Figure 3 (for  $h = 1$ ) and Figure 4 (for  $h = 2$ ). The results for  $h = 0$  are those reported in Figure 2, since they coincide with the ones of the Pearson coefficient computation.

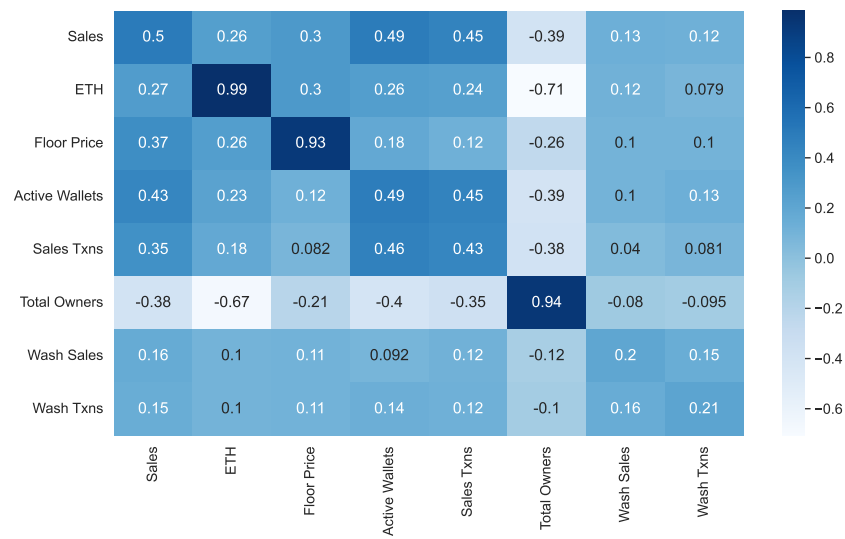


Figure 3. Cross-correlation values averaged on all available NFT collections ( $h = 1$ ).

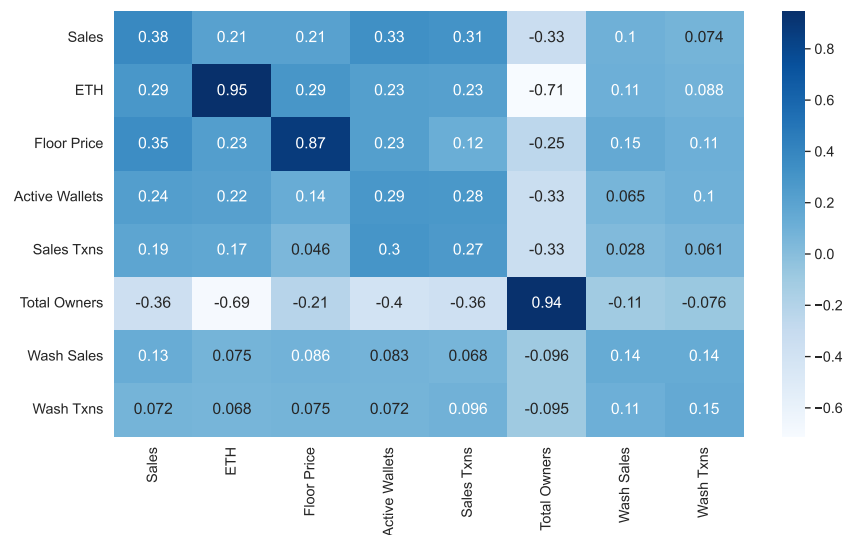


Figure 4. Cross-correlation values averaged on all available NFT collections ( $h = 2$ ).

From the analysis of Figures 2–4, we can observe that, as the value of  $h$  increases, the values of cross-correlation for the pairs of features generally decrease, and sometimes by a great deal. Only for some pairs of features do they remain essentially constant. In contrast, we never observe a significant increase in the values of cross-correlation for any pair. This is already an interesting result because it indicates the absence of a predictive capability of one feature with respect to any other.

Let us now focus on the features most involved in wash trading, namely Wash Sales and Wash Txns. The values of the correlations between these features and all the others in our dataset are shown in the last two rows of Figures 2–4. From the analysis of these figures, we can extract the following insights:

- The values of the correlation between Wash Sales and Wash Txns are very high for  $h = 0$ . This was quite expected because an increase in the number of wash trading transactions generally corresponds to an increase in the amount of money involved. However, observe that the strength of this correlation decreases rapidly when passing from  $h = 0$  to  $h = 1$  or  $h = 2$ . This implies that an increase in the amount of money

currently involved in wash trading has no impact on future wash trading transactions, and vice versa.

- The correlations of Wash Sales with the other features are positive, except for Total Owners, with which there is a slight negative correlation. This can be explained by considering that, in doing wash trading, a person sells their NFTs after buying them to artificially inflate their price.
- While the correlation values between Wash Sales and Wash Txns are very high, those between each of these features and the other ones in the model are low or very low. This suggests that wash trading has a low or very low influence on the main economic and financial variables of an NFT.

The examination of these three figures gives us an initial clue about the main question underlying our paper. However, we think it is appropriate to continue with the other facets before drawing stable conclusions about it. In addition, we have derived two other insights about the wash trading phenomenon that, albeit collateral, are nonetheless of interest in deepening our comprehension of this phenomenon.

### 5.2. Experiments on Causality

The second group of experiments focused on the second facet introduced in Section 3.2, namely causality. In Section 3.2.1, we have said that, in order to test whether the values of a feature  $f_j$  allowed us to predict the values of a feature  $f_k$ , we could use the Granger causality test. We used such a test considering the features Wash Sales and Sales to verify whether Wash Sales Granger-cause Sales. Indeed, if this occurs, we would have a clue that, in wash trading, the game is worth the candle. If not, we would have an additional clue, alongside that already obtained in the previous section, that the game is not worth the candle. Recall that the following null hypothesis must be considered to perform this verification:  $H_0$ : "Wash Sales does not Granger-cause Sales". Then, for each NFT collection, we perform the test and calculate the  $p$ -value. If the latter is less than 0.05, the null hypothesis can be rejected.

Recall also that the Granger causality test for verifying whether  $f_j$  Granger-causes  $f_k$  is based on the idea of comparing the ability to predict the values of  $f_k$  using all the information in the universe ( $\mathcal{U}$ ) with the ability to predict the same values using all the information in the universe except  $f_j$  ( $\mathcal{U}|f_j$ ). For the test to be meaningful, we need to have many values of  $f_j$ . In fact, if this did not happen, there would be a risk that the prediction of  $f_k$  using  $\mathcal{U}$  and the one obtained using  $\mathcal{U}|f_j$  are identical not because of the presence of a causal relationship between  $f_j$  and  $f_k$ , but because  $\mathcal{U}$  and  $\mathcal{U}|f_j$  tend to coincide. For this reason, we had to remove from our analysis all NFT collections that had many null values of Wash Sales. In particular, we kept only those NFT collections that had at least 10% of their rows with not null values of Wash Sales. At the end of this activity, we maintained 1386 of the 2000 original NFT collections.

Once the NFT collections that could be involved in the Granger causality test were identified, we checked for how many of them the  $p$ -value associated with the null hypothesis expressed above was less than 0.05. Specifically, we performed this computation both using a linear autoregressive model (VAR) and adopting the two deep learning models mentioned in Section 3.2.2, namely MLP and LSTM. The results obtained are reported in Table 1. In addition to considering the three different models mentioned above, we considered different values of  $p$  (representing the maximum number of lagged observations). In particular, we took into consideration three values of  $p$ , namely 1, 2 and 3.

**Table 1.** Percentage of the NFT collections for which the null hypothesis  $H_0$  can be rejected.

<i>Model</i>	<i>p</i> = 1	<i>p</i> = 2	<i>p</i> = 3
VAR	4.25%	4.03%	4.01%
MLP	1.04%	1.04%	0.52%
LSTM	2.13%	1.54%	1.53%

From the analysis of Table 1, we can deduce the following insights:

- The null hypothesis is almost always confirmed when the linear model is used within the Granger causality test, and this trend is even more evident when the MLP and LSTM models are employed.
- As the value of  $p$  increases, there is no significant change in the percentage of NFT collections for which the null hypothesis  $H_0$  is confirmed in the Granger causality test.

As for the 614 NFT collections that we had to discard because there were not enough Wash Sales values, from a formal point of view we cannot draw any conclusions regarding the existence of a causal relationship between Wash Sales and Sales. However, we can conjecture that the high number of null values for Wash Sales in these NFT collections is already an indicator of the little influence that Wash Sales can exert on Sales as well as on all the other features. Therefore, we can think that, for all or most of these NFT collections, Wash Sales do not Granger-cause Sales.

At the end of this series of tests, we can conclude that there is generally no causal relationship between Wash Sales and Sales. This represents a confirmation of what we had already found through the correlation analysis, namely that, in the case of wash trading, the game is not worth the candle. However, before drawing any final conclusion, we think it appropriate to examine the third facet described in Section 3.2, namely distance.

### 5.3. Experiments on Distance

The third group of experiments focuses on the third facet introduced in Section 3.2, namely distance. In Section 3.2.3, we have specified that we use Dynamic Time Warping (DTW) to evaluate this facet quantitatively. In particular, we computed the value of DTW for each pair of features and for each NFT collection of the dataset. Then, we averaged the DTW values thus obtained over the NFT collections. The final results are shown in Figure 5.

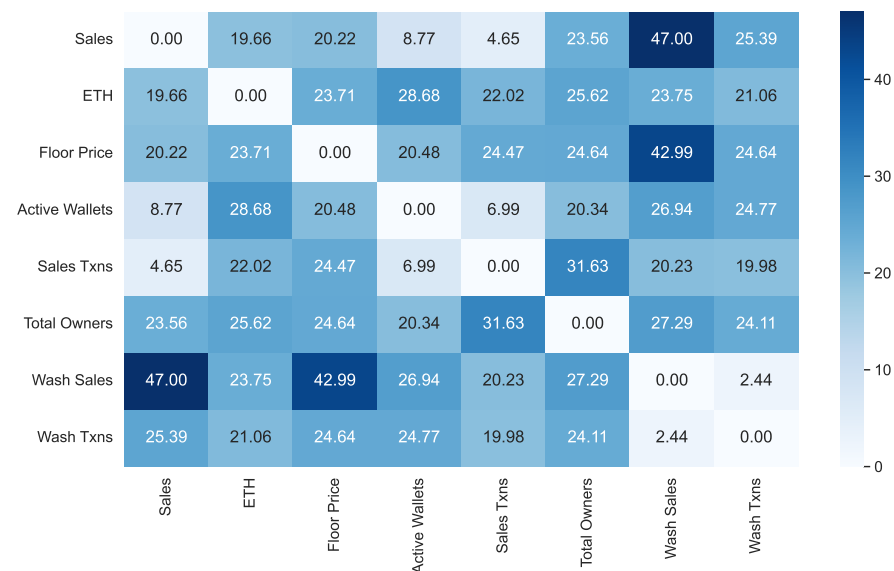


Figure 5. Distance values averaged on all available NFT collections.

Some critical insights can be obtained from the analysis of this figure. In particular:

- The pair of features (Wash Sales, Wash Txns) has a very low value of DTW. This is justified by the fact that both of them relate directly to wash trading activities. In fact, this result is fully congruent with what we have reported in Section 5.1, where we have seen that there is a strong correlation between these two features.



- The pair of features (Sales, Sales Txns) has a very low value of DTW. The same explanation as seen for the previous pair applies to it.
- The pairs of features (Sales Txns, Active Wallets) and (Sales, Active Wallets) have low values of DTW, although they are higher than those of the previous two pairs. The same explanation as for the previous pairs applies to them. However, note that, while in Section 5.1, the correlation of the pair (Sales Txns, Active Wallets) is much greater than that of the pair (Sales, Active Wallets), and in the case of distance there is not a great difference between the two pairs.
- The pairs of features (Wash Sales, Sales) and (Wash Sales, Floor Price) have high DTW values. These values are much higher than the ones characterizing the previous pairs. In particular, the pair (Wash Sales, Floor Price) has a much higher distance than the previous pairs, but this distance is slightly lower than that of the other pairs. In contrast, the distance relative to the pair (Wash Sales, Sales) is the highest one among the distances of all the pairs of features.

The insight concerning the pair (Wash Sales, Floor Price) mentioned above is extremely interesting because it may be a clue that wash trading activities may have an influence, albeit limited, on Floor Price. This information, although it is not the central one we are interested in for this paper, represents a collateral knowledge we extracted that is certainly useful and worthy of future investigation.

Finally, the insight concerning the pair (Wash Sales, Sales) mentioned above is very interesting. We premise that, in contrast to the other two facets, for the DTW parameter there is no predefined maximum value. Therefore, the considerations that can be drawn concern comparisons of the values of this parameter for the various pairs of features. All that said, the fact that the value of DTW for the pair (Wash Sales, Sales) is the highest value of DTW among all possible pairs of features is still a clue that, for Wash Trading, the game is not worth the candle. If we had only this clue, it would be rash to draw a conclusion, given the lack of the maximum value for DTW. However, in this case, this clue is perfectly in line with the ones obtained from the previous two facets. In light of this, it seems reasonable for us to say that, in the case of wash trading, the game is not worth the candle. Therefore, we have found an unambiguous and, in some ways, surprising answer to the question that is the central focus of this paper.

#### 5.4. Further Verification of Knowledge Detected by Our Framework

The experiments in the previous section showed how our framework was able to extract important knowledge about the NFT wash trading phenomenon, and thus to achieve the goal for which it was designed. In particular, it derived that the game is not worth the candle. In this section, we describe a further experiment designed to test whether the knowledge detected by our framework has a foundation or not. For this purpose, we examine three main NFT projects, namely Azuki, Bored Ape Yacht Club (BAYC) and Milady. Our objective is to determine whether in these three cases the wash trading activities carried out had a significant impact on sales. To answer this question, we need to analyze the relationships between the features Wash Sales and Sales in the three NFT projects considering the three different facets managed by our framework.

We begin our verification by analyzing the possible cross-correlations between the features Wash Sales and Sales related to the three projects for the first three lags. The plots are shown in Figures 6–8. From the analysis of this figure, we can observe that Wash Sales and Sales show similar patterns (in the sense that, for example, both show peaks or troughs) but are not directly correlated (in that, for example, the peaks occur at different times, and the same happens for the troughs). This reveals that there is no strong correlation between Wash Sales and Sales, and this is a first confirmation of our framework's ability to extract correct knowledge about the NFT wash trading phenomenon.

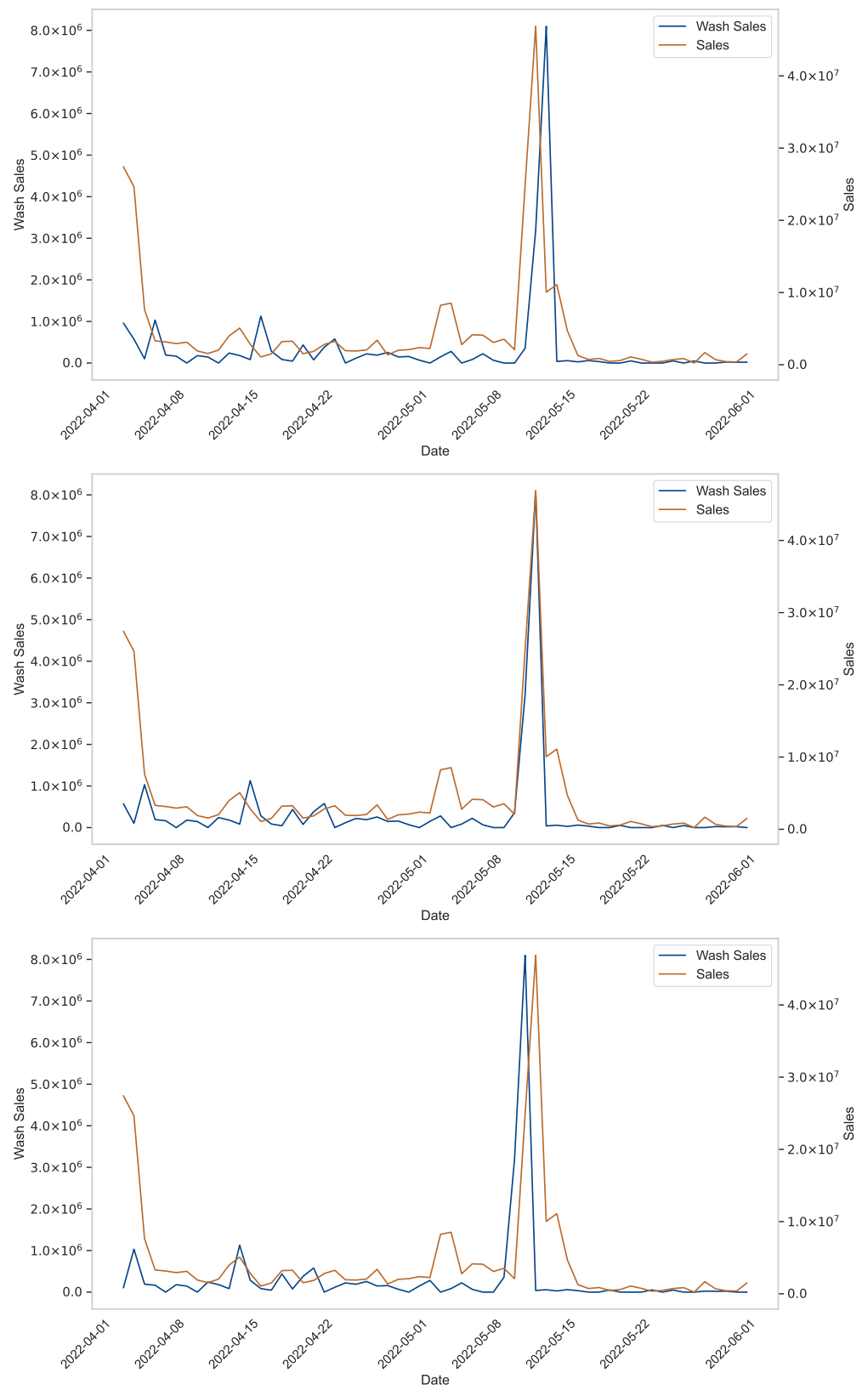


Figure 6. Cross-correlation of Wash Sales and Sales for Azuki when  $h = 0$ ,  $h = 1$ , and  $h = 2$ .

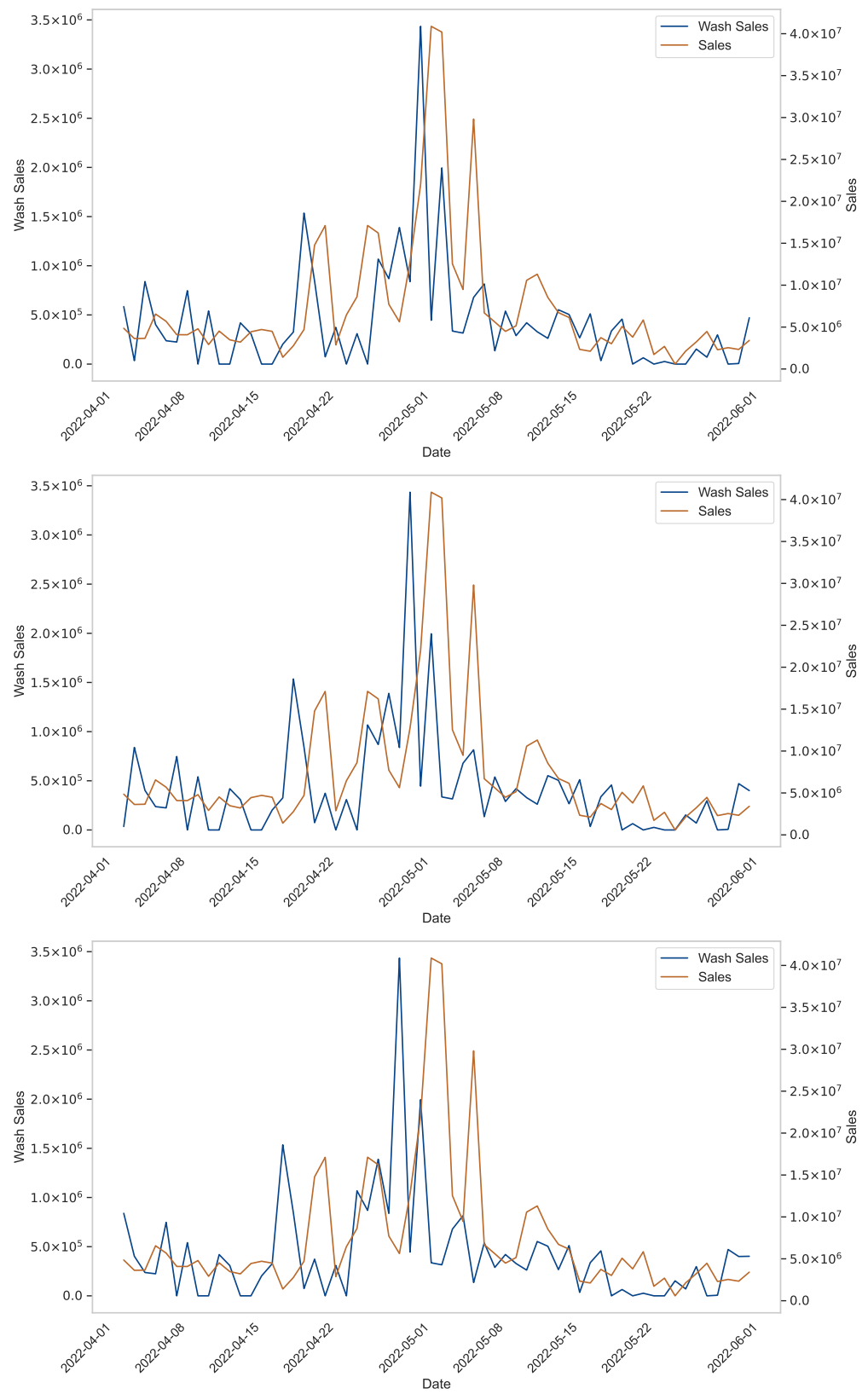
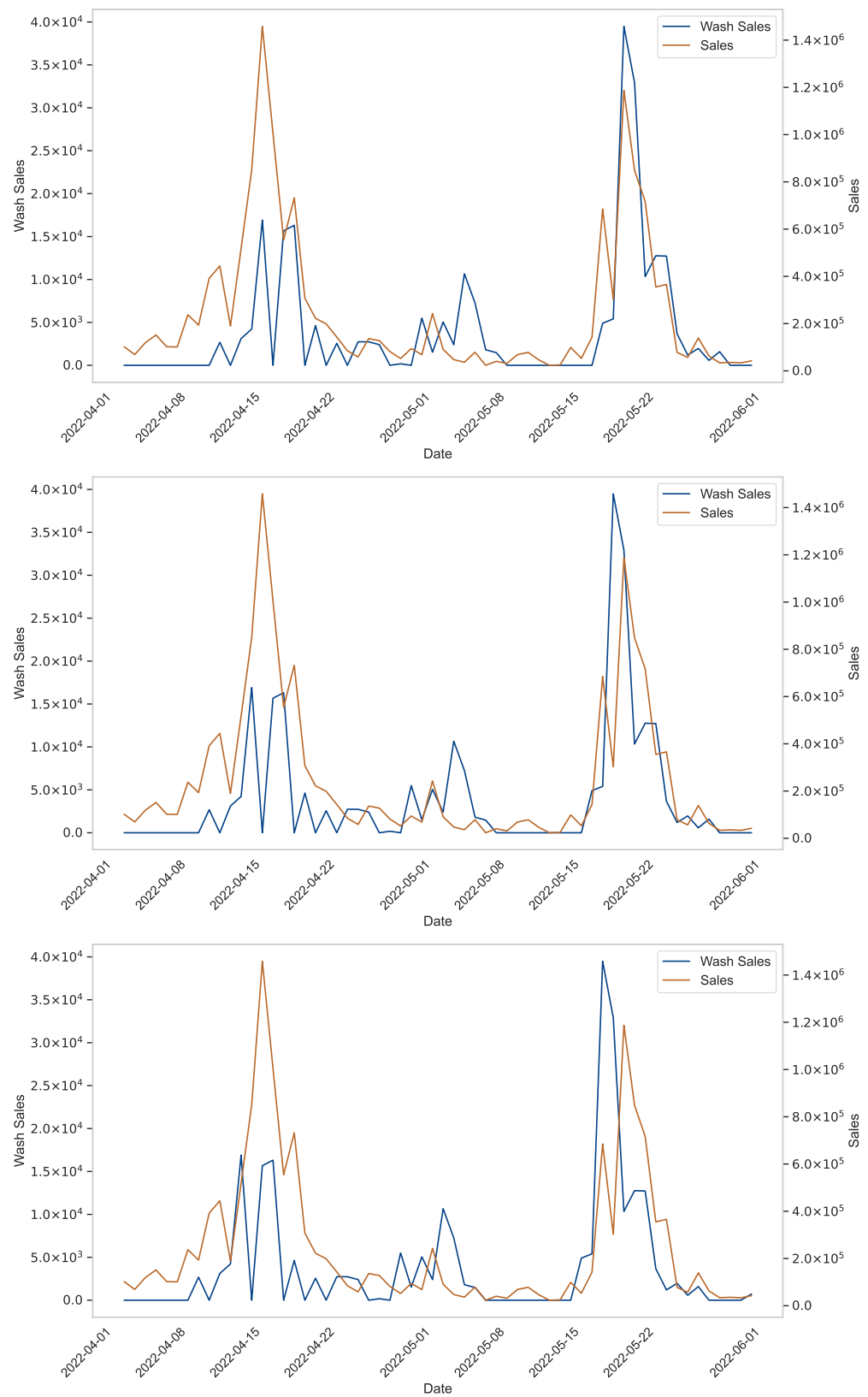


Figure 7. Cross-correlation of Wash Sales and Sales for BAYC when  $h = 0$ ,  $h = 1$ , and  $h = 2$ .



**Figure 8.** Cross-correlation of Wash Sales and Sales for Milady when  $h = 0$ ,  $h = 1$ , and  $h = 2$ .

The next step in our verification involved the causality facet. To this end, we performed the Granger tests to determine whether a causality relationship exists between the features Wash Sales and Sales on the three NFT projects. The corresponding  $p$ -values are presented

in Table 2. From the analysis of this table, we can observe that the  $p$ -values are always greater than 0.05. As a consequence, we cannot reject the null hypothesis of the Granger tests. As mentioned in Section 3.2.2, this implies that wash trading activities have no effect on the sales of an NFT, since the knowledge of the current values of Wash Sales does not allow us to say anything on the future values of Sales.

**Table 2.**  $p$ -values returned by the Granger tests applied on the features Wash Sales and Sales for Azuki, BAYC and Milady.

<i>Model</i>	$p = 1$	$p = 2$	$p = 3$
VAR	0.62	0.18	0.48
MLP	0.39	0.55	0.35
LSTM	0.27	0.8	0.41

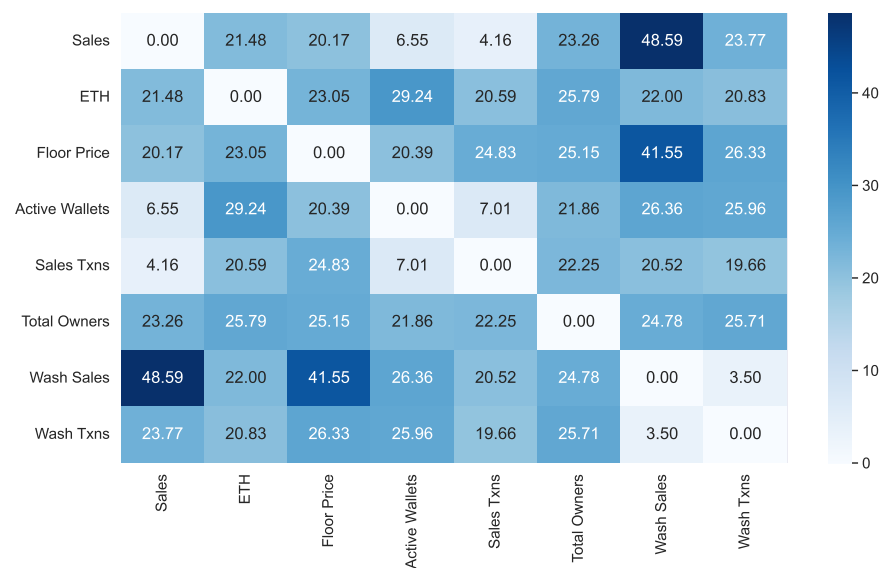
<i>Model</i>	$p = 1$	$p = 2$	$p = 3$
VAR	0.64	0.86	0.16
MLP	0.64	0.5	0.24
LSTM	0.78	0.45	0.73

<i>Model</i>	$p = 1$	$p = 2$	$p = 3$
VAR	0.11	0.23	0.73
MLP	0.23	0.23	0.75
LSTM	0.38	0.93	0.46

The results derived from examining this second facet also confirm the ability of our framework to extract correct knowledge regarding the NFT wash trading phenomenon.

As a final verification, we computed the DTW distance between all possible pairs of features of the three NFT projects. The corresponding heatmaps are reported in Figure 9. From the analysis of this figure, we can see that the pair of features (Wash Sales, Sales) has associated the largest DTW values among all possible pairs. This implies that Wash Sales and Sales are very distant features, and therefore that it is not possible to predict the values of one of them knowing the values of the other. This conclusion is a further confirmation that our framework was able to extract correct knowledge regarding the NFT wash trading phenomenon.



**Figure 9.** Cont.





Figure 9. DTW distance values between all the pairs of features for Azuki, BAYC, and Milady.

As a consequence, all facets examined for all NFT projects considered in this test agree in confirming the correctness of the conclusions derived from our framework about the (negligible) role of wash trading in stimulating sales.

### 6. Discussion

The experiments presented in Section 5 allowed us to conclude that our framework is capable of supporting the analysis of the NFT scenario and, in particular, wash trading activities that occurred in it. As proof of this, we obtained interesting results for all the facets considered.

As for the correlation facet, we found that the correlations and the cross-correlations between the features related to wash trading and the others were very low, suggesting that wash trading activities had a limited impact on the financial features of an NFT.

This inference was confirmed by the causality facet. In fact, the null hypothesis of the Granger test related to the pair (Wash Sales, Sales) was almost never rejected for either linear and nonlinear models. This fact implies that there is generally no causal relationship between Wash Sales and Sales.

The latest confirmation to our framework's conclusions came from the distance facet. In fact, the DTW value for the pair (Wash Sales, Sales) was the highest among all pairs of features. This means that the corresponding features were very distant, implying that it is very unlikely that there is an impact of wash trading activities on NFT values.

It is exactly the joint and unambiguous result of the three facets that allowed us to conclude that wash trading is not worth the candle. Therefore, the framework we proposed has proven its validity to perform the task for which it was designed.

The presence of facets provides our framework with the potential to derive (through some facets), and next confirm (through the other ones), insights into the NFT wash trading phenomenon. The correlation facet implies the computation of the Pearson coefficient and cross-correlation coefficients between two features. They give us an idea of the possible influence of one feature over another and, in our case study, of the impact of wash trading on NFT sales. The correlation facet is able to determine the similarity degree of each pair of features, but cannot show the presence or absence of a cause-and-effect phenomenon. For this reason, we have included the causality facet in our framework. In this case, thanks to the Granger tests, we are able to verify whether one feature has a predictive power over the other. The last facet, i.e., the distance one, has a confirmatory role. In fact, if the distance between two features are low and they are related by a correlation or causality relationship, we can conclude that their connection is much stronger.

An additional merit of our framework is its extensibility, so that at any time we can add other facets capable of capturing new features that we deemed appropriate to consider.

## 7. Conclusions

In this paper, we aimed to verify whether, in the case of wash trading in the NFT market, the game was worth the candle. To achieve our goal, we had to define a framework for analyzing wash trading transactions in the context of the NFT market from an "ex-post" perspective, rather than from an "ex-ante" one, as it had been done previously in the literature. To the best of our knowledge, this is the first paper to address the wash trading phenomenon in the NFT market from this perspective and in a "structured" way. In fact, in the past, there have been blogs and magazines that have dealt with doing "ex-post" statistical investigations of the wash trading phenomenon. However, they proposed only descriptive and diagnostic analyses, and not a framework, complete and "structured", based on a well-defined data model, capable of supporting this kind of analysis. Our framework, in addition to a model tailored specifically for this type of analysis, introduces several "facets" (i.e., perspectives) through which the phenomenon can be analyzed. The formalization of "facets" is also a new feature of our framework, compared to past analyses. At the end of our investigation, we were able to show that wash trading in the NFT market is not worth the candle.

Although the main objective of this paper was to answer this question, it provides some other contributions, namely: (i) a framework to support future "ex-post" investigations on the NFT wash trading phenomenon; (ii) a new dataset, freely available, which stores data on wash trading transactions involving NFTs in the past and may support further analyses on the NFT wash trading phenomenon from an "ex-post" perspective; and (iii) the results of an experimental campaign that not only answers the core question behind this paper, but also provides several additional insights into the NFT wash trading phenomenon.

This paper does not represent an ending point, but rather a starting point, in our investigations on the NFT wash trading phenomenon. In fact, we can imagine several future developments in this line of research. For example, we might consider using Machine Learning algorithms to predict wash trading and sales volumes and to extract knowledge patterns characterizing the features involved in the wash trading phenomenon. In addition, it would be interesting to investigate the correlation between other NFT features, such as rarity and asset type, and the corresponding wash trading activity. Indeed, some types of NFTs might receive more attention from malicious users than others. To this end, a categorization of them based on their features could help predict whether and to what

extent a new NFT might be subject to speculation. Finally, we plan to study what the factors that drive NFT prices are, with a focus on the community that supports an NFT project. In particular, we want to investigate the social media ecosystem around an NFT project and characterize the behavior of users participating in it, as well as the characteristics of contents produced by them.

In addition to the tasks we intend to pursue, we want to outline some potential future recommendations that may assist any researcher who wishes to extend this research. First, it would be interesting to investigate the role of NFT marketplaces in facilitating or preventing wash trading activities. Researchers could enhance the data we collected by integrating it with information about marketplaces and reasoning about the role of the latter in promoting fairness and preventing malicious behavior. Another interesting recommendation is to analyze the motivations behind wash trading. Indeed, our study provides an “ex-post” view of the phenomenon, but says little about the motivations of those people who have engaged in wash trading. Researchers interested in this topic could hypothesize certain motivations and test the truth of their hypotheses by analyzing discussions in social networks among those who engage in NFT transactions. More technical recommendations for researchers interested in this context are to study the impact of NFT ownership concentration and to investigate the impact of NFT trading bots. The first topic is a critically important aspect of market behavior, particularly when much of the ownership is in the hands of a few large holders. The second topic would concern the possibly automatic identification of trading bots. This would be a challenging issue that could allow researchers to study and understand the role of bots in promoting or preventing illicit behaviors.

**Author Contributions:** Conceptualization, G.B. and L.V.; methodology, F.C. and D.U.; software, D.M. and S.S.; validation, E.C. and M.M.; formal analysis, G.B. and F.C.; investigation, E.C. and D.U.; resources, M.M. and L.V.; data curation, G.B. and E.C.; writing—original draft preparation, F.C. and L.V.; writing—review and editing, M.M. and D.U.; visualization, D.M. and S.S.; supervision, G.B., E.C. and M.M.; project administration, F.C., D.U. and L.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** We used data accessible through [Pushshift.io](https://pushshift.io).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Mahdy, A.; Lotfy, K.; El-Bary, A. Use of optimal control in studying the dynamical behaviors of fractional financial awareness models. *Soft Comput.* **2022**, *26*, 3401–3409. [[CrossRef](#)]
2. Mahdy, A. A numerical method for solving the nonlinear equations of Emden-Fowler models. *J. Ocean. Eng. Sci.* **2022**, *in press*.
3. Wilson, K.; Karg, A.; Ghaderi, H. Prospecting non-fungible tokens in the digital economy: Stakeholders and ecosystem, risk and opportunity. *Bus. Horizons* **2022**, *65*, 657–670. [[CrossRef](#)]
4. Wang, D.; Wu, S.; Lin, Z.; Wu, L.; Yuan, X.; Zhou, Y.; Wang, H.; Ren, K. Towards a first step to understand flash loan and its applications in defi ecosystem. In Proceedings of the Ninth International Workshop on Security in Blockchain and Cloud Computing (SBC '21), Taipei, Taiwan, 7–11 June 2021; pp. 23–28.
5. Ante, L. Non-fungible token (NFT) markets on the Ethereum blockchain: Temporal development, cointegration and interrelations. *Econ. Innov. New Technol.* **2022**, 3904683 [[CrossRef](#)]
6. Fox, M.; Glosten, L.; Rauterberg, G. Stock market manipulation and its regulation. *Yale J. Regul.* **2018**, *35*, 67.
7. Shamsi, M.A.; Smith, D.; Gleason, K. Space transition and the vulnerabilities of the NFT market to financial crime. *J. Financ. Crime* **2023**. [[CrossRef](#)]
8. Wachter, V.V.; Jensen, J.R.; Regner, F.; Ross, O. NFT Wash Trading: Quantifying suspicious behaviour in NFT markets. *arXiv* **2022**, arXiv:2202.03866.
9. Nadini, M.; Alessandretti, L.; Giacinto, F.D.; Martino, M.; Aiello, L.M.; Baronchelli, A. Mapping the NFT revolution: Market trends, trade networks, and visual features. *Sci. Rep.* **2021**, *11*, 20902. [[CrossRef](#)]
10. Vasani, K.; Janosov, M.; Barabási, A.L. Quantifying NFT-driven networks in crypto art. *Sci. Rep.* **2022**, *12*, 2769.

11. White, B.; Mahanti, A.; Passi, K. Characterizing the OpenSea NFT marketplace. In Proceedings of the ACM Web Conference 2022 (TheWebConf'22), Lyon, France, 25–29 April 2022; pp. 488–496.
12. Casale-Brunet, S.; Zichichi, M.; Hutchinson, L.; Mattavelli, M.; Ferretti, S. The impact of NFT profile pictures within social network communities. In Proceedings of the ACM Conference on Information Technology for Social Good (GOODIT '22), Limassol, Cyprus, 7–9 September 2022; pp. 283–291.
13. Urom, C.; Ndubuisi, G.; Guesmi, K. Dynamic dependence and predictability between volume and return of Non-Fungible Tokens (NFTs): The roles of market factors and geopolitical risks. *Financ. Res. Lett.* **2022**, *50*, 103188. [[CrossRef](#)]
14. Serneels, S. Detecting wash trading for nonfungible tokens. *Financ. Res. Lett.* **2022**, 103374. [[CrossRef](#)]
15. Weber, D. *Exploring Markets: Non-Fungible Token*; Technical report; IU Discussion Papers-Business & Management: Erfurt, Germany, 2022.
16. Tahmasbi, N.; Fuchsberger, A. Non-fungible Tokens-Exploring Suspicious Washtrader Communities in NFT Networks. In Proceedings of the International Conference on Information System (ICIS'22), Zhuhai, China, 26–28 June 2022.
17. Wang, Q.; Li, R.; Wang, Q.; Chen, S. Non-fungible token (NFT): Overview, evaluation, opportunities and challenges. *arXiv* **2021**, arXiv:2105.07447.
18. Bao, H.; Roubaud, D. Non-Fungible Token: A Systematic Review and Research Agenda. *J. Risk Financ. Manag.* **2022**, *15*, 215. [[CrossRef](#)]
19. He, D.; Liu, Z.; Yang, Q.; Ma, L. The Development of Digital Collection Platform under Responsible Innovation Framework: A Study on China's Non-Fungible Token (NFT) Industry. *J. Open Innov. Technol. Mark. Complex.* **2022**, *8*, 203.
20. Das, D.; Bose, P.; Ruaro, N.; Kruegel, C.; Vigna, G. Understanding security Issues in the NFT Ecosystem. *arXiv* **2021**, arXiv:2111.08893.
21. Bhujel, S.; Rahulamathavan, Y. A Survey: Security, Transparency, and Scalability Issues of NFT's and Its Marketplaces. *Sensors* **2022**, *22*, 8833.
22. Khodabandehlou, S.; Golpayegani, S.A.H. Market manipulation detection: A systematic literature review. *Expert Syst. Appl.* **2022**, *210*, 118330. [[CrossRef](#)]
23. Imisiker, S.; Tas, B.K.O. Wash trades as a stock market manipulation tool. *J. Behav. Exp. Financ.* **2018**, *20*, 92–98. [[CrossRef](#)]
24. Mukhopadhyay, M.; Ghosh, K. Market Microstructure of Non Fungible Tokens. *arXiv* **2021**, arXiv:2112.03172.
25. Cong, L.; Li, X.; Tang, K.; Yang, Y. Detecting and Quantifying Crypto Wash Trading. In Proceedings of the 3rd International Conference on Blockchain Economics, Security and Protocols (Tokenomics '21), New York University, New York, NY, USA, 18–19 November 2021.
26. Pennec, G.L.; Fiedler, I.; Ante, L. Wash trading at cryptocurrency exchanges. *Financ. Res. Lett.* **2021**, *43*, 101982. [[CrossRef](#)]
27. Aloosh, A.; Li, J. Direct evidence of bitcoin wash trading. *Soc. Sci. Res. Netw.* **2019**, 3362153.
28. Chen, J.; Lin, D.; Wu, J. Do cryptocurrency exchanges fake trading volumes? An empirical analysis of wash trading based on data mining. *Phys. A Stat. Mech. Its Appl.* **2022**, *586*, 126405. [[CrossRef](#)]
29. Pelechrinis, K.; Liu, X.; Krishnamurthy, P.; Babay, A. Spotting Anomalous Trades in NFT Markets: The Case of NBA Topshot. *arXiv* **2022**, arXiv:2202.04013.
30. Dowling, M. Fertile LAND: Pricing non-fungible tokens. *Financ. Res. Lett.* **2022**, *44*, 102096. [[CrossRef](#)]
31. Dowling, M. Is non-fungible token pricing driven by cryptocurrencies? *Financ. Res. Lett.* **2022**, *44*, 102097. [[CrossRef](#)]
32. Ante, L. The non-fungible token (NFT) market and its relationship with Bitcoin and Ethereum. *FinTech* **2022**, *1*, 216–224. [[CrossRef](#)]
33. Pinto-Gutiérrez, C.; Gaitán, S.; Jaramillo, D.; Velasquez, S. The NFT Hype: What Draws Attention to Non-Fungible Tokens? *Mathematics* **2022**, *10*, 335. [[CrossRef](#)]
34. Kireyev, P.; Lin, R. *Infinite But Rare: Valuation and Pricing in Marketplaces for Blockchain-Based Nonfungible Tokens*; INSEAD Working Paper: Fontainebleau, France, 2021.
35. Kapoor, A.; Guhathakurta, D.; Mathur, M.; Yadav, R.; Gupta, M.; Kumaraguru, P. Tweetboost: Influence of social media on NFT valuation. In Proceedings of the Web Conference (TheWebConf'22), Lyon, France, 25–29 April 2022; pp. 621–629.
36. Reinsel, G. *Elements of Multivariate TIME Series Analysis*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2003.
37. Pearson, K. Note on Regression and Inheritance in the Case of Two Parents. *Proc. R. Soc. Lond.* **1895**, *58*, 240–242.
38. Granger, C. Investigating causal relations by econometric models and cross-spectral methods. *Econom. J. Econom. Soc.* **1969**, *37*, 424–438. [[CrossRef](#)]
39. Dickey, D.; Fuller, W. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* **1979**, *74*, 427–431.
40. Holden, K. Vector auto regression modeling and forecasting. *J. Forecast.* **1995**, *14*, 159–166. [[CrossRef](#)]
41. Giannini, C.; Mosconi, R. Predictions from unrestricted and restricted VAR models. *Giornale degli Economisti e Annali di Economia* **1987**, *46*, 291–316.
42. Tank, A.; Covert, I.; Foti, N.; Shojaie, A.; Fox, E. Neural granger causality for nonlinear time series. *stat* **2018**, *1050*, 16.
43. Rosoł, M.; Młyńczak, M.; Cybulski, G. Granger causality test with nonlinear neural-network-based methods: Python package and simulation study. *Comput. Methods Programs Biomed.* **2022**, *216*, 106669. [[CrossRef](#)] [[PubMed](#)]

44. Müller, M. Dynamic time warping. *Inf. Retr. Music. Motion* **2007**, 69–84.
45. Giorgino, T. Computing and visualizing dynamic time warping alignments in R: The dtw package. *J. Stat. Softw.* **2009**, 31, 1–24. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.