**UNIVERSITA’ DEGLI STUDI DI NAPOLI**

**“PARTHENOPE”**



**SCUOLA INTERDIPARTIMENTALE**

**DI ECONOMIA E GIURISPRUDENZA**

**Dipartimento di**

**Studi Aziendali e Quantitativi**

**Corso di Laurea in**

**Statistica e Informatica per l’Azienda, la Finanza e le Assicurazioni**

*Tesi di Laurea in*

*Analisi e Previsione delle Serie Economiche e Finanziarie*

**Titolo tesi**

**MODELLI DI APPRENDIMENTO STATISTICO PER IL MERCATO NFT**

|  |  |
| --- | --- |
| TUTOR  Chiar.mo Prof.  Giovanni De Luca | CANDIDATO  Francesco Palmisano  MATR. 0265000075 |

ANNO ACCADEMICO 2022/2023

**Abstract**

Nell'ambito dell'analisi e previsione delle serie storiche economiche e finanziarie, questa ricerca esplora il mercato degli NFT attraverso l'applicazione di modelli di apprendimento statistico. Concentrandosi sulla collezione "Bored Ape Yacht Club", il lavoro adotta un approccio statistico avanzato, utilizzando analisi descrittive, di regressione e di serie storiche per comprendere le dinamiche sottostanti che influenzano il prezzo e la negoziazione degli asset digitali.

INDICE

[CAPITOLO 1 4](#_Toc153543469)

[INTRODUZIONE 4](#_Toc153543470)

[1.1 BLOCKCHAIN 4](#_Toc153543471)

[1.2 CRIPTOVALUTE 4](#_Toc153543472)

[1.3 NON-FUNGIBLE TOKEN (NFT) 5](#_Toc153543473)

[1.3.1 BORED APE YATCH CLUB (BAYC) 7](#_Toc153543474)

[CAPITOLO 2 8](#_Toc153543475)

[PREPARAZIONE DEI DATI 8](#_Toc153543476)

[2.1 PIANIFICAZIONE DEI DATI 8](#_Toc153543477)

[2.2 ESTRAZIONE DEI DATI 9](#_Toc153543478)

[2.3 PULIZIA DEI DATI 11](#_Toc153543479)

[CAPITOLO 3 13](#_Toc153543480)

[ANALISI DESCRITTIVA 13](#_Toc153543481)

[3.1 “DATE” 13](#_Toc153543482)

[3.2 “ID” 17](#_Toc153543483)

[3.3 “PRICE (K)” 18](#_Toc153543484)

[3.4 “RARITY” 20](#_Toc153543485)

[3.5 “CLOSE ETH ($)” 23](#_Toc153543486)

[3.6 “VOLUME ETH (B)” 25](#_Toc153543487)

[CAPITOLO 4 27](#_Toc153543488)

[ANALISI DI REGRESSIONE 27](#_Toc153543489)

[4.1 QUADRO TEORICO 27](#_Toc153543490)

[4.2 ANALISI 29](#_Toc153543491)

[4.2.1 ANALISI PER RARITA’ 31](#_Toc153543492)

[CAPITOLO 5 36](#_Toc153543493)

[ANALISI CON MODELLI PER SERIE STORICHE 36](#_Toc153543494)

[5.1 QUADRO TEORICO 36](#_Toc153543495)

[5.2 ANALISI CLASSICA 38](#_Toc153543496)

[5.3 ANALISI MODERNA 41](#_Toc153543497)

[5.3.1 MODELLO ARIMAX 46](#_Toc153543498)

[5.4 PREVISIONI 50](#_Toc153543499)

[CONCLUSIONI 53](#_Toc153543500)

[APPENDICE – Codice R 54](#_Toc153543501)

[BIBLIOGRAFIA 72](#_Toc153543502)

[RINGRAZIAMENTI 73](#_Toc153543503)

# CAPITOLO 1

# INTRODUZIONE

Nel corso degli ultimi anni, il mondo della finanza e della tecnologia ha vissuto una rivoluzione straordinaria, guidata da una serie di nuove e potenti innovazioni, quali la tecnologia blockchain, le criptovalute e i non-fungible token (NFT).

In questo lavoro di tesi, il focus è posto sul mercato degli NFT, con l'obiettivo di implementare procedure di apprendimento statistico finalizzate a comprendere l'andamento di un mercato tanto particolare quanto volubile, influenzabile e volatile. Questa analisi si concentra in modo particolare sulla collezione Bored Ape Yacht Club (BAYC), una delle più rilevanti e interessanti raccolte di token non fungibili attualmente disponibili.

Prima di esaminare nel dettaglio questo nuovo mercato, è essenziale gettare le basi per una piena comprensione di queste innovazioni che stanno ridefinendo il panorama finanziario e digitale su scala globale.

## 1.1 BLOCKCHAIN

Una blockchain è uno strumento gratuito per lo scambio di informazioni, che garantisce un migliore controllo dei dati degli utenti attraverso la crittografia.

Decentrata per natura, una blockchain si basa sul principio della rete peer-to-peer (P2P). La maggior parte delle reti non ha un server dedicato o un'unica autorità di controllo, ma funziona attraverso il consenso continuo della sua base di utenti.

Una blockchain conserva i dati applicando le informazioni a singoli file informatici, definiti “blocchi”, a loro volta concatenati tra loro. Una volta concatenati tra loro, i blocchi contengono informazioni immutabili perché ogni nuovo blocco porta con sé il DNA del blocco precedente ed è concatenato ad esso, rendendolo quindi virtualmente impossibile da alterare.

Lo scopo di una blockchain è quindi quello di garantire il flusso di informazioni da un mittente a un destinatario in modo decentralizzato.

## 1.2 CRIPTOVALUTE

La criptovaluta è una riserva di valore digitale che può essere scambiata e commercializzata come le normali valute globali.

La differenza tra queste ultime è semplice ma sostanziale: le normali valute globali, a differenza delle criptovalute, sono essenzialmente centralizzate e controllate dai governi e dalle aziende che gestiscono tali sistemi monetari.

I sistemi centralizzati hanno una storia di manipolazioni e interferenze e, dopo il crollo finanziario del 2007, molti individui desideravano una soluzione finanziaria più stabile ed equa.

Il Whitepaper di Satoshi Nakamotos su Bitcoin, pubblicato alla fine del 2008, era un documento innovativo che descriveva come la crittografia potesse creare una forma di denaro elettronico peer-to-peer senza fiducia e senza la necessità di coinvolgere alcun intermediario nelle transazioni.

Questa struttura decentralizzata offriva un sistema potenzialmente più equilibrato ed equo, con ogni utente o "nodo" connesso alla rete che svolgeva il proprio ruolo nel raggiungimento del consenso. Il consenso si ha quando ogni utente può partecipare alla convalida delle transazioni della blockchain scaricando un nodo che contiene l'intera blockchain dall'inizio.

Quando la maggioranza dei nodi è d'accordo nell'eseguire una transazione o nello scoprire un nuovo blocco, si parla di "consenso". In questo modo, diventa estremamente difficile manipolare la blockchain perché sarebbe necessario che più della metà dei nodi sia d'accordo nel riscrivere i blocchi.

Le criptovalute rappresentano quindi un importantissimo strumento decentralizzato, attraverso il quale gli utenti possono effettuare transazioni finanziarie in modo diretto, senza la necessità di intermediari, sfruttando la tecnologia blockchain come registro sicuro e distribuito.

## 1.3 NON-FUNGIBLE TOKEN (NFT)

I non-fungible token (NFT) sono asset crittografici presenti sulla blockchain che si distinguono l'uno dall'altro con codici di identificazione e metadati unici.

Un token non fungibile è tale (non fungibile) in quanto è un token che rappresenta un bene unico con caratteristiche che gli sono proprie: non può essere scambiato o sostituito da un altro token equivalente.

Le qualità uniche degli NFT che li differenziano dagli altri token sono:

* Raro: sebbene gli sviluppatori di token non fungibili possano creare tutti i token che vogliono, spesso mantengono limitata la fornitura di NFT in modo che possano essere scarsi per aumentarne il valore;
* Indivisibili: anche se non è il caso di tutti i token non fungibili, spesso sono indivisibili. Nella maggior parte dei casi non è possibile dividere a metà i token non fungibili e quindi è necessario acquistare l'intero token;
* Unicità: è la caratteristica più evidente dei token non fungibili. La loro unicità viene registrata in una scheda informativa permanente. Questa scheda è come un certificato di autenticazione.

Gli NFT possono essere utilizzati in diversi modi:

* Videogiochi: gli NFT possono essere utilizzate in svariati modi nei videogiochi. Nei giochi sparatutto in prima persona (FPS), possono assumere la forma di skin per armi o avatar. Nei giochi di ruolo (RPG), questi NFT possono assumere la forma di equipaggiamento, come armi o armature, e di alcune cavalcature (animali domestici). Per i giochi di carte (trading card games), gli NFT possono essere le carte stesse e anche la personalizzazione del loro terreno o del retro delle carte;
* Arte digitale: online, tutto è infinitamente duplicabile. Con gli NFT è diventato possibile quantificare il non quantificabile. L'arte digitale, che richiede molte ore di lavoro, può quindi essere valutata a livelli più alti rispetto all'arte fisica. Inoltre, è possibile verificare pubblicamente l'indirizzo di un portafoglio collegato al profilo di un artista. La firma dell'artista è quindi inclusa nella zecca NFT.
* Oggetti da collezione: gli oggetti da collezione digitali possono assumere molte forme, ma tutti esistono per lo stesso scopo: essere collezionati. Inizialmente, un progetto cosiddetto "da collezione" non ha altro scopo che quello di essere collezionato. Questo non significa che tutti gli oggetti da collezione rimarranno tali nel tempo. I creatori di prodotti da collezione spesso prevedono di utilizzarli in altri ambiti, come ad esempio nei videogiochi, o di utilizzarli come gettoni per entrare in un club o per premiare in altro modo gli utenti.

Tra le collezioni più rinomate e attivamente scambiate emergono Terraforms, Meebits, Bored Ape Yacht Club e CryptoPunks, ciascuna delle quali registra un volume di scambi che supera il milione di Ether (ETH), sottolineando così la loro rilevanza e popolarità all'interno del mercato NFT.

### 1.3.1 BORED APE YATCH CLUB (BAYC)

Bored Ape Yacht Club (BAYC) o Bored Ape è una collezione di 10.000 non-fungible token rilasciata sul mercato da Yuga Labs nel 2021.

La collezione NFT raffigura dei disegni di scimmie annoiate con dei tratti specifici, dal look ai colori, all’espressione e agli accessori. I disegni sono tutti unici e sono generati attraverso un algoritmo che combina diverse caratteristiche.

Il sempre maggior costo, insieme con l'esclusività dovuta al numero limitato, ha trasformato in pochissimi mesi le Bored Ape in una sorta di sorta di status symbol tanto da guadagnare l'attenzione di diversi personaggi famosi che hanno iniziato ad acquistarle. Tra le celebrità più famose ad aver acquistato una Bored Ape ci sono Jimmy Fallon, Neymar, Stephen Curry, Snoop Dogg, Gwyneth Paltrow, Shaquille O'Neal, Justin Bieber, Timbaland, Paris Hilton ed Eminem.

# CAPITOLO 2

# PREPARAZIONE DEI DATI

La preparazione dei dati svolge un ruolo cruciale nell'ambito dell'analisi del mercato degli NFT.

Disporre infatti di dati accurati e coerenti diventa fondamentale per comprendere appieno le dinamiche di questo settore in evoluzione.

Durante questo processo, vengono condotte operazioni quali la pianificazione, l’estrazione e la pulizia dei dati, garantendone l’affidabilità e la preparazione per le successive analisi.

## 2.1 PIANIFICAZIONE DEI DATI

In questo elaborato di tesi, per analizzare l’innovativo mercato degli NFT, si porrà l'attenzione sulla collezione Bored Ape Yacht Club (BAYC), che si configura come una delle raccolte di token non fungibili più significative e interessanti al momento.

Secondo i dati forniti da [CoinMarketCap](https://coinmarketcap.com/it/nft/collections/), infatti, BAYC occupa attualmente la 3° posizione nella classifica delle collezioni NFT più scambiate, basata sul calcolo del volume complessivo di vendite.

Le caratteristiche distintive e la sua notevole popolarità la rendono, quindi, oggetto di studio altamente promettente e, al contempo, rappresentativo per l'analisi del mercato degli NFT.

Al fine di ottenere una visione globale del fenomeno NFT, saranno utilizzate tutte le transazioni disponibili al momento della raccolta dei dati.

In particolare, saranno esaminate le transazioni relative al mercato OpenSea che, come riscontrato dai dati forniti da [Etherscan](https://etherscan.io/nft-trades?contractAddress=0xbc4ca0eda7647a8ab7c2061c2e118a18a936f13d), con 28.681 transazioni su un totale di 36.831 rappresenta il 77,87% del totale delle transazioni relative alla collezione.

Inoltre, verrà presa in considerazione la rarità di ciascun NFT, poiché si ritiene che essa possa influire sul suo prezzo.

Infine, dato che gli NFT della collezione sono acquistati e venduti in Ether (ETH), la criptovaluta nativa della blockchain Ethereum, verranno inclusi anche dati sul prezzo e sul volume giornaliero di ETH, insieme alle variabili ritardate di entrambe queste misurazioni.

Nel contesto delle analisi da effettuare, verrà implementata una regressione lineare multipla per identificare possibili relazioni tra il prezzo delle transazioni e le variabili relative a ETH.

È importante specificare che, sebbene l'approccio sia basato su una regressione lineare, si ipotizza che le distribuzioni del prezzo degli NFT e delle variabili legate a Ether potrebbero non essere strettamente normali. Tuttavia, l'analisi mira comunque a valutare le eventuali relazioni esistenti e a fornire informazioni rilevanti per una migliore comprensione dei fattori che influenzano il prezzo degli NFT, nonostante l’ipotesi di normalità potrebbe non essere completamente soddisfatta.

Successivamente, sarà valutato l’andamento della serie storica dei prezzi delle transazioni, attraverso le metodologie di analisi classica e moderna.

Sarà infine compiuto un tentativo di formulare previsioni basate sui risultati ottenuti.

Per l'esecuzione di queste analisi statistiche, si farà uso del software R, con la possibilità di utilizzare Excel qualora dovesse risultare necessario.

## 2.2 ESTRAZIONE DEI DATI

Le transazioni sono state analizzate a partire dal 1 maggio 2021 fino al 12 settembre 2023.

I dati relativi alle transazioni sono stati estratti da [Etherscan](https://etherscan.io/nft-trades?contractAddress=0xbc4ca0eda7647a8ab7c2061c2e118a18a936f13d), e comprendono le seguenti variabili:

* “DateTime (UTC)” rappresenta la data e l'orario delle transazioni, nel fuso orario Universal Time Coordinated;
* “Token ID” fornisce un identificativo unico assegnato a ciascun NFT all'interno della collezione Bored Ape Yacht Club;
* “Price” indica il prezzo di ogni transazione, espresso in dollari;
* “Market” identifica il mercato in cui è avvenuta la transazione.

Poiché Etherscan consente il recupero di dati per un massimo di 1000 transazioni per ciascun periodo, la raccolta dati è stata effettuata mensilmente durante il periodo preso in considerazione.

Nei mesi con più di 1000 transazioni, sono state effettuate più raccolte per lo stesso mese.

Per quanto riguarda i dati sulla rarità degli NFT, sono stati acquisiti da [Rarity Sniper.](https://raritysniper.com/bored-ape-yacht-club)

Poiché i dati desiderati non erano disponibili direttamente, è stata eseguita un'operazione di web scraping utilizzando il software Octoparse.

I dati raccolti includono le seguenti variabili:

* “NFT ID”, come per la variabile “Token ID” precedentemente descritta, fornisce un identificativo unico assegnato a ciascun NFT all'interno della collezione;
* “Rank” è un indicatore che varia da 1 a 10.000 e indica la rarità decrescente di ciascun NFT.

Un valore basso di Rank denota un NFT raro all'interno della collezione (più il valore di Rank è basso, più raro è il token) mentre un valore elevato indica una rarità inferiore;

* “Rarity” rappresenta la categoria di rarità assegnata a ciascun NFT in base al suo Rank.

La variabile è formata dalle seguenti categorie:

“MITICO” (Rank 1-99);

“LEGGENDARIO” (Rank 100-499);

“EPICO” (Rank 500-2499);

“RARO” (Rank 2500-4999);

“NON COMUNE” (Rank 5000-7499);

“COMUNE” (Rank 7500-10000).

Infine, i dati relativi al prezzo e al volume giornaliero di ETH sono stati raccolti quotidianamente per lo stesso periodo di analisi e sono stati ottenuti da [Yahoo Finance](https://it.finance.yahoo.com/quote/ETH-EUR/history?p=ETH-EUR).

Questi dati comprendono le seguenti variabili:

* “Date”, che a differenza della variabile “DateTime (UTC)”, rappresenta la data, ma non l’orario, in cui sono stati registrati i dati relativi al prezzo e al volume giornaliero di ETH;
* “Close” indica il prezzo di chiusura giornaliero di ETH;
* “Volume” rappresenta il volume giornaliero di scambi di ETH, ossia la quantità di Ether scambiata durante il giorno.

## 2.3 PULIZIA DEI DATI

I dati raccolti a livello mensile sono stati aggregati per anno e successivamente filtrati per identificare il mercato desiderato, ovvero il mercato OpenSea.

Successivamente, i dati sono stati trasferiti in un ambiente Excel per eseguire operazioni specifiche di pulizia dei dati.

In particolare, è stato condotto un processo di estrazione delle informazioni relative alla variabile "Price" delle transazioni, eliminando i dati espressi sia in dollari che in ETH e conservando esclusivamente i dati espressi in dollari.

Questa scelta è stata guidata dalla variabilità del valore di ETH nel tempo, che non offre una rappresentazione stabile dei prezzi delle transazioni.

Al contrario, l'utilizzo dei prezzi in dollari offre una misura più stabile per l'analisi. Successivamente, i dati sono stati importati in R.

Sono state quindi eliminate tutte le righe contenenti valori mancanti.

Al fine di migliorare la leggibilità e la scala dei dati, l'unità di misura della variabile “Price” delle transazioni è stata trasformata da dollari a migliaia di dollari.

Queste operazioni sono state effettuate per ciascun anno di riferimento e infine, i dati relativi a ciascun anno sono stati aggregati in un unico dataset finale.

Successivamente, sono stati rimossi i duplicati delle transazioni.

La variabile “DateTime (UTC)” delle transazioni è stata modificata per eliminare l'orario, ritenuto irrilevante ai fini dell'analisi, consentendo di mantenere solo la data.

Per quanto riguarda i dati relativi alla rarità, essi sono stati integrati nel dataset finale delle transazioni mediante una procedura di merge basata sulla corrispondenza tra le variabili "Token ID" e "NFT ID".

Successivamente, sono state rimosse dal dataset finale le nuove righe contenenti valori mancanti.

In relazione ai dati raccolti per Ether (ETH), è stata eseguita una trasformazione dell'unità di misura della variabile “Volume” da dollari a miliardi di dollari, migliorando così la leggibilità dei dati.

Inoltre, sono state create due nuove variabili per i ritardi di un periodo, sia per la variabile “Close” che per la variabile “Volume”.

Infine, il tutto è stato integrato all'interno del dataset finale attraverso una nuova procedura di merge, basata stavolta sulla corrispondenza tra le variabili “DateTime (UTC)” e “Date".

Il processo termina con un’ultima fase di pulizia, per rimuovere le ulteriori righe con valori mancanti, e col trattamento di alcuni valori anomali, opportunamente sostituiti con il valore medio delle tre transazioni antecedenti e successive, considerando la categoria di rarità.

Il dataset finale pulito conta quindi 24366 osservazioni suddivise in 9 variabili, di seguito elencate:

* “Date”;
* “ID”;
* “Price (k)”;
* “Rank”;
* “Rarity”;
* “Close ETH ($)”;
* “Close ETH\_L1 ($)”;
* “Volume ETH (B)”;
* “Volume ETH\_L1 (B)”.

Questo articolato processo di pulizia dei dati ha contribuito a garantire l’affidabilità, la qualità e la coerenza dei dati, preparandoli per l’imminente analisi statistica.

# CAPITOLO 3

# ANALISI DESCRITTIVA

L’analisi dei dati è senza dubbio la fase più importante e articolata del progetto di ricerca.

All’interno di una prima fase, saranno esplorati i dati mediante un approccio descrittivo, allo scopo di individuare tendenze e aspetti significativi. Grafici informativi e indicatori statistici saranno impiegati per fornire una visione chiara del panorama in questione.

Tali informazioni costituiranno una base solida per le successive analisi, all’interno delle quali saranno esplorate, quantificate e approfondite eventuali relazioni tra i dati delle diverse variabili. Inoltre, si porrà un'attenzione particolare all'analisi dei dati storici, con l'obiettivo di valutare le tendenze e i modelli che emergono da queste informazioni.

I risultati derivanti da queste analisi contribuiranno a una comprensione più approfondita delle dinamiche del mercato degli NFT, e saranno cruciali per la formulazione di previsioni accurate.

In questa prima fase, quindi, saranno valutate le caratteristiche del dataset pulito finale, variabile per variabile, attraverso l’utilizzo di molteplici grafici descrittivi.

**3.1 “DATE”**

Il grafico riportato in Figura 3.1 rappresenta la distribuzione delle transazioni, relative alla collezione, in funzione del tempo. L'asse delle ascisse ("Data") indica le date in cui sono avvenute le transazioni, mentre l'asse delle ordinate ("Frequenza") rappresenta il numero di transazioni registrate in ciascuna data.

Immagine che contiene testo, schermata, Carattere, diagramma

Descrizione generata automaticamente

Figura 3.1 - Distribuzione delle Transazioni in riferimento alla variabile Date

La Tabella 3.1 presenta un riepilogo delle statistiche chiave relative al grafico delle transazioni per data.

Immagine che contiene testo, schermata, Carattere, numero

Descrizione generata automaticamente

Tabella 3.1 – Statistiche Rilevanti della Distribuzione delle Transazioni in riferimento alla variabile Date

Emergono quindi diverse tendenze temporali significative:

* Il 2021 ha rappresentato circa l’80% del totale delle transazioni, con picchi rilevanti nei primi 3 mesi di vita della collezione. Difatti, nell’ambito delle transazioni totali, solamente nei primi tre giorni è stata superata la soglia del 10%, mentre nei i primi 3 mesi è stata superata la soglia del 50%.
* Il 2022 ha rappresentato circa il 16% del totale delle transazioni, indicando una forte diminuzione nell'attività rispetto all'anno precedente;
* Il 2023 ha rappresentato un numero ancora più limitato di transazioni, pari al 3.87% del totale delle transazioni. Nello specifico, negli ultimi 3 mesi di rilevazione, è stato osservato l’1.5% delle transazioni, pari a circa 1/3 di quelle osservate nel giorno di lancio.

L’analisi prosegue con l’illustrazione di un nuovo grafico (Figura 3.2), che mostra il prezzo medio delle transazioni relative alla collezione nel corso del tempo. Sull'asse delle ascisse ("Data") sono riportate le date delle transazioni, mentre sull'asse delle ordinate ("Prezzo (k)") è rappresentato il prezzo medio in migliaia di dollari.

Immagine che contiene testo, schermata, Carattere, diagramma

Descrizione generata automaticamente

Figura 3.2 – Prezzo Medio delle Transazioni in riferimento alla variabile Date

Dalla Figura 3.2 si evince chiaramente come la prima parte delle transazioni, seppur rappresenti un quantitativo molto elevato, risulti avere un prezzo medio molto basso. Ci sono poi, verso la fine del 2021, diversi picchi di prezzo, con conseguente aumento della volatilità. Il prezzo medio tende poi ad avere una tendenza al rialzo per i primi mesi del 2022, per poi passare ad una tendenza al ribasso per tutto il periodo rimanente.

Questa situazione è anche confermata dai boxplot in Figura 3.3.

Immagine che contiene testo, schermata, diagramma, Carattere

Descrizione generata automaticamente

Figura 3.3 – Boxplot dei Prezzi delle Transazioni (per anno)

I boxplot, infatti, mostrano chiaramente come nel 2021 i prezzi delle transazioni presentino una distribuzione con valori molto bassi all'interno della scatola, fortemente influenzati da numerosi valori anomali. Nel 2022, si osserva un significativo aumento della mediana e dell'ampiezza della scatola, con una minore presenza di valori anomali rispetto all'anno precedente. Nel 2023, nonostante la scatola dei prezzi sia comunque più alta rispetto a quella del 2021, i prezzi medi rimangono inferiori rispetto al 2022, confermando la tendenza al ribasso sopra descritta.

Immagine che contiene testo, Carattere, schermata, diagramma

Descrizione generata automaticamente

Figura 3.4 – Boxplot dei Prezzi delle Transazioni (per mese)

In linea con i picchi precedentemente osservati, da agosto 2021 fino a febbraio 2022 si osservano in Figura 3.4 prezzi delle transazioni fortemente variabili; nei primi mesi del 2021 e più in generale da fine 2022 in poi i prezzi osservati sono significativamente più stabili.

**3.2 “ID”**

Per questa variabile è riportato il grafico a barre (Figura 3.5). L’asse delle ascisse rappresenta l’ID univoco associato ad ogni NFT della collezione, mentre l'asse delle ordinate rappresenta il numero di transazioni effettuate per ciascun ID.

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

Figura 3.5 - Distribuzione delle Transazioni in riferimento alla variabile ID

Di seguito viene esplicitata la Tabella 3.2 in cui sono riportate alcune statistiche rilevanti.

Immagine che contiene testo, schermata, linea

Descrizione generata automaticamente

Tabella 3.2 – Statistiche Rilevanti della Distribuzione delle Transazioni in riferimento alla variabile ID

Emerge da questi dati un’informazione fondamentale: il 28,24% degli ID è caratterizzato da una transazione, il che significa che più di ¼ degli NFT sono stati acquistati per essere posseduti, e non scambiati.

Questo dato riflette una tendenza importante mostrando come, dai dati relativi alla collezione BAYC, un segmento significativo di partecipanti al mercato sia interessato a detenere questi token in modo simile a una collezione tradizionale, piuttosto che trattarli come asset finanziari.

**3.3 “PRICE (K)”**

L’istogramma in Figura 3.6 visualizza la distribuzione dei prezzi degli NFT della collezione, espressi in migliaia di dollari. Sull'asse delle ascisse, è espresso il valore del prezzo in migliaia, mentre sull'asse delle ordinate è rappresentata la frequenza con cui i prezzi ricadono in ciascun intervallo di valore.

Immagine che contiene testo, schermata, diagramma, linea

Descrizione generata automaticamente

Figura 3.6 – Distribuzione dei Prezzi degli NFT

Di seguito viene riportata la Tabella 3.3 in cui sono evidenziate alcune statistiche rilevanti.

Immagine che contiene testo, schermata, Carattere, linea

Descrizione generata automaticamente

Tabella 3.3 – Statistiche Rilevanti della Distribuzione dei Prezzi degli NFT

Sia la Figura 3.6 che la Tabella 3.3 forniscono risultati di chiara interpretazione: circa il 98% delle transazioni presenta un importo inferiore a $400.000, mentre più del 99% delle transazioni ha un importo inferiore a $500.000. E’ interessante notare come, nonostante questa predominanza, sia possibile osservare nel grafico valori che toccano quota $3.000.000.

La situazione appena descritta è ulteriormente confermata dagli ulteriori grafici in Figura 3.7 e Figura 3.8.

Immagine che contiene testo, schermata, Carattere, linea

Descrizione generata automaticamente

Figura 3.7 – Distribuzione Cumulativa dei Prezzi degli NFT

Immagine che contiene testo, schermata, diagramma, linea

Descrizione generata automaticamente

Figura 3.8 – Boxplot dei Prezzi degli NFT

Dalla Figura 3.7, infatti, è possibile osservare come in prossimità dei $500.00 la probabilità cumulativa sia praticamente pari a 1, mentre in Figura 3.8 si nota come la scatola dei prezzi sia fortemente compressa verso il basso, a causa dei numerosi valori anomali che influenzano la distribuzione verso valori più elevati.

**3.4 “RARITY”**

I prossimi due grafici mostrano come si distribuisce la rarità all'interno della collezione. Nel primo grafico a barre (Figura 3.9), è possibile osservare la frequenza delle diverse categorie di rarità. Nel secondo grafico a torta (Figura 3.10), invece, la percentuale di ciascuna categoria di rarità è rappresentata in modo più immediato.

Immagine che contiene testo, schermata, diagramma, Diagramma

Descrizione generata automaticamenteFigura 3.9 – Distribuzione della Rarità

Immagine che contiene testo, diagramma

Descrizione generata automaticamenteFigura 3.10 – Grafico a Torta della Rarità

Sono inoltre rappresentati in Figura 3.11 una serie di boxplot, ognuno dei quali rappresenta una diversa categoria di rarità e fornisce un'idea chiara delle tendenze dei prezzi all'interno di ciascuna categoria.

Immagine che contiene testo, diagramma, schermata, linea

Descrizione generata automaticamenteFigura 3.11 – Boxplot dei Prezzi degli NFT in riferimento alla variabile Rarity

Di seguito viene quindi esplicitata la Tabella 3.4 in cui sono riportate alcune statistiche rilevanti.

Immagine che contiene testo, schermata, Carattere, ricevuta

Descrizione generata automaticamente

Tabella 3.4 – Statistiche Rilevanti della Distribuzione della Rarità

Da questa serie di grafici si evince facilmente come, all'aumentare della rarità di un NFT, si osservi un aumento sia nella media dei prezzi che nella variabilità. Allo stesso tempo, al diminuire della categoria di rarità, si riscontra una diminuzione sia nelle medie che nella variabilità dei prezzi.

Questi risultati sottolineano e confermano, come era stato previsto nella pianificazione dei dati, il ruolo importante che la rarità gioca nel determinare i prezzi all'interno della collezione.

**3.5 “CLOSE ETH ($)”**

I seguenti grafici forniscono una visione completa del prezzo di Ether nel periodo di rilevazione. Nello specifico, il primo grafico (Figura 3.12) rappresenta l’andamento della serie storica dei prezzi, mentre il secondo (Figura 3.13) mostra le variazioni nel tempo delle differenze giornaliere nei prezzi.

Immagine che contiene testo, Carattere, Diagramma, diagramma

Descrizione generata automaticamente

Figura 3.12 – Serie Storica dei Prezzi Close di Ether

Immagine che contiene testo, Diagramma, schermata, linea

Descrizione generata automaticamente

Figura 3.13 – Differenze Giornaliere nei Prezzi Close di Ether

Di seguito viene quindi esplicitata una serie di tabelle (Tabella 3.5 e Tabella 3.6) in cui sono riportate alcune statistiche rilevanti.

Immagine che contiene testo, schermata, Carattere, linea

Descrizione generata automaticamente

Tabella 3.5 – Media e Deviazione Standard (per anno) della Serie Storica dei Prezzi Close di Ether

Immagine che contiene testo, schermata, ricevuta, Carattere

Descrizione generata automaticamente

Tabella 3.6 – Media e Deviazione Standard (per anno) delle Differenze Giornaliere nei Prezzi Close di Ether

Da Figura 3.12 e Figura 3.13 e dalle relative tabelle (Tabella 3.5 e Tabella 3.6), emerge chiaramente un andamento al ribasso nei prezzi di Ether nel corso del periodo considerato. Nel 2021, infatti, la media dei prezzi era di circa $2.865, ma questa cifra è notevolmente diminuita, fino ai $1.743,65 del 2023, con una diminuzione media giornaliera di prezzo pari a $0,05. Dettagliatamente, nel corso del 2022, si è verificata una notevole diminuzione giornaliera media di prezzo pari a $0,66, seguita da un aumento giornaliero medio di $0,44 per l’anno successivo.

La variabilità dei prezzi, misurata attraverso la deviazione standard, risulta essere significativamente elevata durante tutto il periodo considerato, con valori notevolmente fluttuanti, in particolare nei primi due anni di rilevazione. In media, la variabilità giornaliera dei prezzi si aggira intorno ai $20, confermando l'ampia variabilità che caratterizza il mercato delle criptovalute.

Tale situazione è finalmente confermata dai boxplot in Figura 3.14.

Immagine che contiene testo, schermata, diagramma, linea

Descrizione generata automaticamente

Figura 3.14 – Boxplot dei Prezzi Close di Ether (per anno)

**3.6 “VOLUME ETH (B)”**

Per l'analisi del volume di Ether, è stato creato il seguente grafico a barre (Figura 3.15) che rappresenta la variazione del volume medio nel corso del tempo, misurato in miliardi di dollari (B).

Immagine che contiene testo, schermata, Diagramma, Carattere

Descrizione generata automaticamente

Figura 3.15 – Volume Medio di Ether in riferimento alla variabile Date

Di seguito è esplicitata la Tabella 3.7 in cui sono riportate alcune statistiche rilevanti.

Immagine che contiene testo, schermata, Carattere, linea

Descrizione generata automaticamente

Tabella 3.7 – Media e Deviazione Standard (per anno) del Volume Medio di Ether in riferimento alla variabile Date

Sia la media che la variabilità del volume mostrano un calo significativo nel corso dei tre anni di rilevazione. Nello specifico, il volume medio è passato dai circa 28 miliardi del 2021 ai circa 7 miliardi e mezzo del 2023, indicando come l’interesse per lo scambio della criptovaluta sia notevolmente diminuito. La variabilità, misurata sempre attraverso la deviazione standard, allo stesso modo risulta essere in media, rispetto al 2021, quasi dimezzata nel 2022, e circa quattro volte minore nel 2023.

Tale analisi è finalmente confermata e verificata dai boxplot annuali in Figura 3.16.

Immagine che contiene testo, schermata, linea, numero

Descrizione generata automaticamente

Figura 3.16 – Boxplot del Volume di Ether (per anno)

# CAPITOLO 4

# ANALISI DI REGRESSIONE

Dopo un’attenta e approfondita analisi descrittiva, il focus si sposta sulla variabile di maggiore interesse, ossia il prezzo degli NFT. Nello specifico, si intende esaminare se e in che misura tale variabile sia influenzata dalle variabili relative al prezzo e al volume di Ether. L’analisi verrà poi ripetuta per ciascuna categoria di rarità, al fine di valutare l'influenza delle variabili esplicative di Ether per ciascuna categoria. A tal fine, viene eseguita un'analisi di regressione lineare multipla, benché sia importante ricordare che, in base all'ipotesi formulata nella progettazione dei dati, la normalità delle distribuzioni potrebbe non essere pienamente rispettata.

## 4.1 QUADRO TEORICO

**Regressione Lineare Multipla:** La regressione lineare multipla è una tecnica statistica utilizzata per modellare la relazione tra una variabile dipendente (o variabile di risposta) e due o più variabili indipendenti (o predittori). Questo metodo è estremamente utile nell'analisi dei dati e nella previsione poiché consente di valutare il contributo relativo delle variabili indipendenti nel predire la variabile dipendente. Il modello di regressione lineare multipla è definito come:

Y = β₀ + β₁X₁ + β₂X₂ + … + βpXp + ε

dove:

* Y è la variabile dipendente.
* X₁, X₂, …, Xp sono le variabili indipendenti.
* β₀ è il termine di intercetta.
* β₁, β₂, …, βp sono i coefficienti di regressione associati a ciascuna variabile indipendente X₁, X₂, …, Xp.
* ε rappresenta l'errore residuo.

L'obiettivo nella regressione lineare multipla è la stima dei coefficienti β₀, β₁, β₂, …, βp in modo che il modello si adatti ai dati osservati nel miglior modo possibile. L’obiettivo è raggiunto utilizzando il metodo dei minimi quadrati.

**Metodo dei Minimi Quadrati:** Il metodo dei minimi quadrati è una procedura per stimare i coefficienti del modello di regressione lineare multipla in modo da minimizzare la somma dei quadrati dei residui (SSE) definita come:

dove n è il numero di osservazioni. L'obiettivo è trovare i valori dei coefficienti β₀, β₁, β₂, …, βp che minimizzano SSE. Questo è solitamente fatto risolvendo un sistema di equazioni che deriva dalla derivazione parziale di SSE rispetto ai coefficienti e impostando le derivate a zero.

Nel contesto della regressione lineare, il modello può essere rappresentato come:

Y = Xβ + ε

dove:

* Y è il vettore delle osservazioni della variabile dipendente,
* X è la matrice dei predittori (variabili indipendenti),
* β è il vettore dei coefficienti da stimare,
* ε è il vettore degli errori.

L'obiettivo principale è trovare il vettore β che minimizza la somma dei quadrati degli errori, cioè:

minβ ​∣∣Y−Xβ∣∣2

La soluzione ottimale β∗ è data da:

β∗ = (XTX)−1 XTY

dove:

* XT rappresenta la trasposta di X,
* (XTX)−1 è l'inverso della matrice prodotto tra la trasposta di X e X.

Questo approccio fornisce una stima dei coefficienti che rappresenta il miglior adattamento lineare ai dati osservati.

**Ipotesi di Base:** Nell'ambito della regressione lineare multipla, sono fondamentali alcune ipotesi di base che devono essere verificate per garantire che le stime dei coefficienti siano affidabili. Queste ipotesi includono:

1. **Linearità:** La relazione tra le variabili indipendenti e la variabile dipendente è lineare nel modello. Questo può essere verificato esaminando i grafici di dispersione e cercando eventuali pattern non lineari.
2. **Indipendenza degli errori:** Gli errori residui (ε) non sono correlati tra loro. In altre parole, non esiste autocorrelazione tra gli errori.
3. **Normalità degli errori:** Gli errori residui seguono una distribuzione normale. Questo può essere verificato utilizzando test di normalità come il test di Shapiro-Wilk.
4. **Omoschedasticità:** La varianza degli errori è costante su tutti i livelli delle variabili indipendenti. Questo può essere verificato esaminando i residui rispetto alle previsioni.

Assicurarsi che queste ipotesi siano soddisfatte è essenziale per l'affidabilità del modello di regressione lineare multipla e delle stime dei coefficienti. Se una o più di queste ipotesi non sono verificate, potrebbe essere necessario apportare correzioni o considerare modelli alternativi.

## 4.2 ANALISI

In questa sezione viene effettuata una prima analisi di regressione, relativa al dataset finale nel suo insieme. Per implementare il modello di regressione, si stabilisce “Price (k)” come variabile di risposta, e si realizzano tanti modelli quante sono le diverse combinazioni tra le variabili esplicative riferite al prezzo e al volume di Ether, ossia “Close ETH ($)”, “Volume ETH (B)”, e le relative variabili ritardate di un giorno.

Nella Tabella 4.1 viene quindi esplicitato un confronto tra i modelli ottenuti, in termini di R2 corretto e p-value per il test di significatività globale del modello.

Immagine che contiene testo, Carattere, schermata, ricevuta

Descrizione generata automaticamente

Tabella 4.1 – Confronto tra modelli, in termini di R2 corretto e p-value per il test di significatività globale del modello

La Tabella 4.1 presenta i seguenti risultati:

* Il valore nullo del p-value per tutti i modelli indica che tutte le combinazioni di variabili considerate sono statisticamente significative per spiegare la variabile di risposta;
* Il modello “Close ETH\_L1 + Volume ETH” presenta il valore di R2 corretto più elevato (0,3340), indicando che il 33,4% della varianza del prezzo degli NFT può essere spiegata dal prezzo close di Ether nel giorno precedente e dal volume di Ether scambiato nella giornata di riferimento.

Il modello che ha per variabili esplicative “Close ETH\_L1 ($)” e “Volume ETH (B)” si configura quindi come il miglior modello tra i possibili, e sarà quindi oggetto di ulteriori analisi.

La Tabella 4.2 mostra e quantifica la relazione esistente tra le variabili esplicative del modello e la variabile di risposta.

Immagine che contiene testo, schermata, Carattere, linea

Descrizione generata automaticamente

Tabella 4.2 – Coefficienti di Regressione e p-value delle variabili esplicative del modello “Close ETH (L1) + Volume ETH”

La Tabella 4.2 può essere così commentata:

* Il valore nullo del p-value per entrambe le variabili indica che sia il prezzo close che il volume di Ether sono statisticamente significativi per spiegare il prezzo delle transazioni degli NFT;
* Il coefficiente associato alla variabile “Close ETH\_L1 ($)” di circa 0,072 indica che, mantenendo invariata la variabile “Volume ETH (B)”, ad un incremento unitario del prezzo close di Ether del giorno precedente è associato un aumento del prezzo degli NFT di circa 0,072 (k), ossia di circa $72;
* Il coefficiente associato alla variabile “Volume ETH ($)” di circa -4,912 indica che, mantenendo invariata la variabile “Close ETH\_L1 ($)”, ad un incremento unitario del volume di Ether durante la giornata di riferimento è associata una riduzione del prezzo degli NFT di circa 4,912 (k), ossia di circa $4912.

L'analisi fornisce quindi una chiara prospettiva sulle dinamiche che caratterizzano il mercato degli NFT in relazione alla criptovaluta della blockchain Ethereum. In un contesto in cui la crescita e la volatilità dei mercati digitali sono una costante, le osservazioni qui riportate sono rilevanti per comprendere come il prezzo degli NFT sia influenzato da fattori specifici legati a Ether.

Innanzitutto, l'alto livello di significatività statistica per entrambe le variabili, "Close ETH\_L1 ($)" e "Volume ETH ($)", sottolinea l'importanza di monitorare attentamente, anche durante l’arco della giornata stessa, il prezzo e il volume di Ether per predire le fluttuazioni nei prezzi degli NFT. Questo potrebbe rivelarsi particolarmente utile per gli investitori e gli operatori di mercato che cercano di prendere decisioni informate in un ambiente dinamico.

La relazione positiva tra il prezzo di Ether del giorno precedente e il prezzo degli NFT suggerisce che le variazioni nei mercati dove è possibile scambiare ETH possono avere un impatto significativo sull'andamento successivo degli NFT. Tuttavia, la relazione inversa tra il volume delle transazioni di ETH e il prezzo degli NFT aggiunge un livello di complessità, indicando che il comportamento del mercato degli NFT può essere influenzato da dinamiche di offerta e domanda più ampie.

### 4.2.1 ANALISI PER RARITA’

L’analisi precedentemente illustrata viene, in questa sezione, ulteriormente esplorata e arricchita, attraverso uno studio più dettagliato. Il dataset, infatti, non è più esaminato nella sua interezza, ma è suddiviso per categorie di rarità. Questa scelta è motivata, come evidenziato nel sottocapitolo 3.4, dalle notevoli disparità tra le diverse categorie. Pertanto, risulta ragionevole approfondire ulteriormente l'analisi al fine di catturare le differenze significative tra le categorie di rarità.

Vengono di seguito rappresentate due serie di scatterplot. Nella Figura 4.1, sono mostrati gli scatterplot distinti per ciascuna categoria di rarità, con il prezzo di Ether ritardato di un giorno sull'asse delle ascisse e il prezzo degli NFT sull'asse delle ordinate. Analogamente, nella Figura 4.2 sono presentati scatterplot distinti per ciascuna categoria di rarità, con il volume di Ether sull'asse delle ascisse e il prezzo degli NFT sull'asse delle ordinate.

Immagine che contiene testo, diagramma, mappa, schermata

Descrizione generata automaticamenteFigura 4.1 – Scatterplot tra Prezzi degli NFT e Prezzi di Ether del giorno precedente

Immagine che contiene testo, diagramma, schermata, Diagramma

Descrizione generata automaticamenteFigura 4.2 – Scatterplot tra Prezzi degli NFT e Volume di Ether

A partire da questi ultimi due grafici (Figura 4.1 e Figura 4.2), viene quindi implementato un nuovo modello di regressione per ciascuna categoria di rarità. Tale modello è caratterizzato dalle stesse variabili esplicative utilizzate per l’analisi del dataset nel suo complesso, ossia "Close ETH\_L1 ($)" e "Volume ETH (B)"; la variabile di risposta è ovviamente "Price (k)".

Nella Tabella 4.3 viene esplicitato un confronto tra i modelli ottenuti per ciascuna categoria di rarità, in termini di R2 corretto e p-value per il test di significatività globale del modello.

Immagine che contiene testo, schermata, Carattere, numero

Descrizione generata automaticamente

Tabella 4.3 – Confronto tra modelli, in termini di R2 corretto e p-value per il test di significatività globale del modello

La Tabella 4.3 presenta i seguenti risultati:

* Il valore tendente allo 0 del p-value per tutti i modelli indica che, per ognuna delle categorie di rarità, è possibile spiegare la variabile di risposta in maniera significativa, attraverso la combinazione di variabili esplicative considerata;
* I modelli riferiti alle categorie più rare (“MITICO” e “LEGGENDARIO”) presentano i valori di R2 corretto più deboli (rispettivamente 0,2344 e 0,2466), indicando che per queste due categorie di rarità è possibile spiegare una percentuale più bassa di varianza del prezzo degli NFT attraverso il prezzo close di Ether nel giorno precedente e il volume di Ether scambiato nella giornata di riferimento.
* I modelli riferiti alle categorie meno rare (“EPICO”, “RARO”, “NON COMUNE” e “COMUNE”), invece, presentano i valori di R2 corretto più elevati (rispettivamente 0,3918, 0,3857, 0,3810 e 0,3871), indicando che per queste quattro categorie di rarità è possibile spiegare una percentuale più alta di varianza del prezzo degli NFT attraverso il prezzo close di Ether nel giorno precedente e il volume di Ether scambiato nella giornata di riferimento.

L'analisi evidenzia chiaramente come, all'aumentare della rarità, risulti più complesso spiegare la variabile di risposta del modello attraverso le variabili esplicative. Al contrario, diminuendo la rarità, diventa più agevole spiegare la variabile di risposta del modello di regressione.

La Tabella 4.4, invece, mostra e quantifica, per ciascuna categoria di rarità, la relazione esistente tra le variabili esplicative del modello e la variabile di risposta.

Immagine che contiene testo, schermata, Carattere, Parallelo

Descrizione generata automaticamente

Tabella 4.4 - Coefficienti di Regressione e p-value delle variabili esplicative del modello, riferito a ciascuna categoria di rarità

La Tabella 4.4 esprime due statistiche particolarmente rilevanti:

* La categoria più rara (“MITICO”) presenta i coefficienti più elevati per entrambe le variabili esplicative. Nel dettaglio, il coefficiente associato alla variabile “Close ETH\_L1 ($)” di circa 0,263 indica che, mantenendo invariata la variabile “Volume ETH (B)”, ad un incremento unitario del prezzo close di Ether del giorno precedente è associato un aumento del prezzo degli NFT di circa 0,263 (k), ossia di circa $263. Il coefficiente associato alla variabile “Volume ETH ($)” di circa -16,690 indica invece che, mantenendo invariata la variabile “Close ETH\_L1 ($)”, ad un incremento unitario del volume di Ether durante la giornata di riferimento è associata una riduzione del prezzo degli NFT di circa 16,690 (k), ossia di circa $16690;
* La categoria meno rara (“COMUNE”) presenta i coefficienti più deboli per entrambe le variabili esplicative. Più nello specifico, il coefficiente associato alla variabile “Close ETH\_L1 ($)” di circa 0,064 indica che, mantenendo invariata la variabile “Volume ETH (B)”, ad un incremento unitario del prezzo close di Ether del giorno precedente è associato un aumento del prezzo degli NFT di circa 0,064 (k), ossia di circa $64. Il coefficiente associato alla variabile “Volume ETH ($)” di circa -4,517 indica invece che, mantenendo invariata la variabile “Close ETH\_L1 ($)”, ad un incremento unitario del volume di Ether durante la giornata di riferimento è associata una riduzione del prezzo degli NFT di circa 4,517 (k), ossia di circa $4517.

Emerge quindi chiaramente l’importante ruolo che la rarità gioca in questa analisi. Difatti è facile convincersi di come, all’aumentare della rarità all’interno della collezione BAYC, entrambi i coefficienti tendono a crescere, dimostrando come la relazione tra il prezzo degli NFT e le variabili relative a Ether sia soggetta a variazioni significative in base al livello di rarità.

# CAPITOLO 5

# ANALISI CON MODELLI PER SERIE STORICHE

Dopo aver analizzato le variabili che più influenzano il prezzo delle transazioni degli NFT, questa sezione si concentra sull'analisi delle serie storiche dei prezzi degli NFT. L'obiettivo è esaminare le tendenze nel tempo e formulare previsioni. Per raggiungere questo scopo, vengono impiegate tecniche legate all'analisi classica e moderna, incluso l'utilizzo di serie storiche specifiche di particolari regressori, i quali saranno successivamente identificati.

## 5.1 QUADRO TEORICO

Una serie storica è un insieme di dati ordinati nel tempo, offrendo una panoramica evolutiva di un fenomeno. Questa analisi è applicabile a contesti economici, fisici o demografici, con l'obiettivo di descrivere, spiegare o prevedere andamenti.

Le serie storiche si differenziano in base a diversi aspetti:

1. Spaziatura temporale: una serie storica può essere regolare, con intervalli costanti, o irregolare, con intervalli variabili.
2. Numero di fenomeni: una serie può concentrarsi su un singolo fenomeno (univariata) o coinvolgere più fenomeni (multivariata).
3. Natura del fenomeno (di stock o flusso): una serie storica può rappresentare una consistenza istantanea (stock) o una quantità riferita ad un intervallo di tempo (flusso).

Una rapida visualizzazione grafica di una serie storica avviene mediante un grafico cartesiano, con il tempo sull'asse delle ascisse e i valori della serie sull'asse delle ordinate. Questa rappresentazione offre un'istantanea delle caratteristiche principali della serie nel corso del tempo.

Due approcci principali all'analisi di serie storiche sono l'approccio classico e l'approccio moderno.

**Approccio Classico**

Nell'approccio classico all'analisi delle serie storiche, si assume che le osservazioni siano descritte da una funzione deterministica

Yt =f(t)+ϵt

dove:

Yt sono i valori osservati nel tempo t,

f(t) rappresenta la componente deterministica,

ϵt è la componente stocastica o di errore.

**Approccio Moderno: Modelli ARIMA**

Nell'approccio moderno, ci concentriamo sulle correlazioni tra i valori attuali e passati di una serie temporale. Utilizziamo i modelli AutoRegressivi Integrati a Media Mobile (ARIMA), i quali ci consentono di stimare l'andamento della serie. I modelli ARIMA sono definiti da tre componenti principali:

* AutoRegressiva (AR): Dipendenza lineare tra il valore attuale e i valori passati.
* Integrativa (I): Differenziazione della serie per renderla stazionaria.
* Media Mobile (MA): Dipendenza lineare tra il valore attuale e i residui passati.

**Fasi del Processo ARIMA**

Le fasi del processo ARIMA sono tre:

**Identificazione:** In questa fase, analizziamo i correlogrammi dell’autocorrelazione globale e parziale della serie. Questo ci aiuta a comprendere la correlazione tra le osservazioni e a formulare ipotesi sugli ordini delle componenti AR e MA del modello ARIMA.

**Stima:** Successivamente, procediamo con la stima dei coefficienti del modello ARIMA. Valutiamo la significatività di ciascun coefficiente, assicurandoci che le stime siano affidabili.

**Verifica:** Nella fase di verifica, analizziamo i correlogrammi dell’autocorrelazione globale e parziale dei residui del modello. L'obiettivo è assicurare che le correlazioni tra i residui rispettino determinati criteri, indicando la bontà del modello.

Questo approccio iterativo, sviluppato da Box e Jenkins, è fondamentale per ottenere un modello ARIMA robusto e affidabile. Ogni fase verifica e, se necessario, corregge la fase precedente, garantendo coerenza e accuratezza nell'analisi e nella previsione delle serie storiche.

**Previsioni**

Nel contesto delle serie storiche, la previsione è un obiettivo comune, distinto tra previsione puntuale e intervallo di previsione. La previsione puntuale al tempo T+1 basata sull'informazione fino al tempo T è data da:

yT (1)=E(yT+1 ∣IT )

dove IT rappresenta l'informazione fino al tempo T. La previsione puntuale minimizza l'errore quadratico medio di previsione (EQMP), una funzione di perdita simmetrica.

**L'Accuratezza delle Previsioni**

Per valutare l'accuratezza delle previsioni, si adotta una procedura ricorsiva, generando M previsioni per un periodo avanti (h = 1). La sequenza di previsioni parte dall'intervallo temporale da 1 a T−M, stima il modello e produce la previsione al periodo successivo, T−M+1, indicata con yT−M (1). Tale processo si ripete per gli intervalli successivi fino a T−1.

Successivamente, si confrontano le M previsioni da yT−M (1) a yT−1 (1) con i valori osservati da yT−M+1 a yT . Per valutare l'accuratezza delle previsioni, si considera l'errore di previsione, valutato tramite radice dell'errore quadratico medio (RMSE), errore assoluto medio (MAE) e errore percentuale assoluto medio (MAPE).

**Intervallo di Previsione**

Oltre alla previsione puntuale, si considera l'intervallo di previsione. Quest'ultimo è definito in corrispondenza di uno specifico livello di significatività α e si basa sull'errore standard della previsione (se(h)):

(yT (h)−zα/2 se(h),yT (h)+zα/2 se(h))

L'ampiezza dell'intervallo, determinata dal valore della variabile normale standard (zα/2), riflette il grado di incertezza nelle previsioni.

## 5.2 ANALISI CLASSICA

Per procedere con l’analisi classica e moderna della serie storica dei prezzi degli NFT, viene generato un nuovo set di dati a partire dal dataset finale. Inizialmente, è stata effettuata la selezione di un insieme specifico di variabili rilevanti, ossia: “Date”, “Price (k)”, “Close ETH ($)”, “Close ETH\_L1 ($)”, “Volume ETH (B)”, “Volume ETH\_L1 (B)”. Successivamente, i dati sono stati aggregati in base alla variabile “Date”, al fine di ottenere una visione più chiara delle tendenze nel tempo. Questo processo di aggregazione ha coinvolto il calcolo della media per ciascuna variabile numerica, fornendo così un set di dati aggregato che riflette le medie dei valori delle variabili considerate per ciascuna data. Questo processo si è reso necessario a causa della presenza di molteplici transazioni relative a singole date, rendendo difficile la costruzione diretta di una serie storica a frequenza giornaliera. L’aggregazione per data, attraverso la media, ha quindi contribuito a semplificare la struttura dei dati e a fornire una base solida per l'analisi statistica e la modellazione delle serie storiche.

Questa sezione, che ha per oggetto l’analisi attraverso metodologie classiche, mostra nella Figura 5.1 l’andamento della serie storica dei prezzi delle transazioni degli NFT. L'asse delle ascisse indica le date in cui sono avvenute le transazioni, mentre l'asse delle ordinate rappresenta la media dei prezzi per ciascuna data.

Immagine che contiene testo, Carattere, schermata, linea

Descrizione generata automaticamente

Figura 5.1 – Serie Storica dei Prezzi degli NFT

L’analisi classica viene implementata attraverso il metodo basato sul confronto tra R2 corretto: si considerano infatti modelli con trend polinomiali con ordini 1, 2 e 3 e si sceglie il trend polinomiale che presenta il valore di R2 corretto più elevato. Nella Tabella 5.1 è possibile visualizzare questo confronto.

Immagine che contiene testo, schermata, Carattere, linea

Descrizione generata automaticamente

Tabella 5.1 - Confronto tra modelli, in termini di R2 corretto e p-value per il test di significatività globale del modello

Il modello che presenta il valore di R2 corretto più elevato è il modello con trend polinomiale di ordine 3, che di seguito viene esplicitato nella Tabella 5.2.

Immagine che contiene testo, schermata, Carattere, linea

Descrizione generata automaticamente

Tabella 5.2 - Coefficienti di Regressione e p-value delle variabili esplicative del modello

Nella Figura 5.2 viene finalmente rappresentata la serie storica dei prezzi delle transazioni degli NFT, includendo il trend polinomiale di ordine 3.

Immagine che contiene testo, schermata, linea, Carattere

Descrizione generata automaticamente

Figura 5.2 – Serie Storica dei Prezzi degli NFT (in nero) e trend polinomiale di ordine 3 (in blu)

Dalla Figura 5.2 risulta evidente come il modello con trend cubico, caratterizzato da un R2 corretto di 0,6609, mostri solo una parziale adattabilità alla serie storica dei prezzi degli NFT di BAYC. Questo suggerisce che, nonostante la presenza di un trend polinomiale di ordine 3, il modello non riesce a spiegare completamente la variazione nei prezzi nel corso del tempo. Ciò potrebbe indicare l'influenza di altri fattori o la complessità della dinamica sottostante della serie storica dei prezzi NFT. Diventa quindi opportuno considerare approcci alternativi, come l’analisi moderna, oppure includere ulteriori variabili esogene nel modello per migliorare la sua capacità predittiva e comprensiva della dinamica dei prezzi NFT.

## 5.3 ANALISI MODERNA

Per procedere con l’analisi moderna della serie storica dei prezzi degli NFT, si fa riferimento allo stesso set di dati generato per l’analisi classica. Prima di applicare la procedura di Box-Jenkins, è importante eseguire una valutazione preliminare riguardo la stazionarietà della serie storica. Viene quindi visualizzato nella Figura 5.3 il correlogramma globale, all’interno del quale sono rappresentati i primi 50 ritardi.

Immagine che contiene testo, schermata, linea, Carattere

Descrizione generata automaticamente

Figura 5.3 – Correlogramma globale della serie storica ai primi 50 ritardi

L'analisi del correlogramma rivela un andamento che suggerisce la presenza di non stazionarietà nella serie storica. I primi 50 ritardi mostrano valori che decrescono lentamente da 0,949 a 0,561, indicando una forte autocorrelazione. Di conseguenza, si procede all'applicazione delle differenze prime per stabilizzare la serie. La Figura 5.4 illustra chiaramente l'andamento dell'autocorrelazione delle differenze prime.

Immagine che contiene testo, schermata, linea, diagramma

Descrizione generata automaticamente

Figura 5.4 – Correlogramma globale delle differenze prime ai primi 50 ritardi

L’andamento dell’ACF mostra un primo ritardo significativo pari a - 0,397, mentre i ritardi successivi rientrano all'interno della banda di confidenza, suggerendo che non ci sono correlazioni significative oltre il primo ritardo. Questo comportamento indica che la componente di tendenza è stata in gran parte catturata dalla differenziazione delle osservazioni, rendendo la serie stazionaria. Si può così applicare la procedura di Box-Jenkins.

Nella prima fase di identificazione, sono analizzate le funzioni di autocorrelazione globale e parziale, attraverso i correlogrammi rappresentati in Figura 5.5.

Immagine che contiene testo, schermata, diagramma, linea

Descrizione generata automaticamenteFigura 5.5 – Correlogramma globale e parziale delle differenze prime ai primi 50 ritardi

Si evince chiaramente, dalla Figura 5.5, come i coefficienti di autocorrelazione globale stimati si annullino dopo il primo ritardo, mentre i coefficienti di autocorrelazione parziale stimati tendano a zero all'aumentare del ritardo considerato: c'è un'evidenza a favore di un modello MA(1).

L’analisi procede quindi con la fase di stima, all’interno della quale sono stimati i parametri del modello ARIMA(0,1,1) mediante il metodo della massima verosimiglianza; successivamente, viene effettuato il test d’ipotesi sul coefficiente del modello. La Tabella 5.3, di seguito illustrata, descrive accuratamente questa fase di stima.

Immagine che contiene testo, schermata, Carattere, linea

Descrizione generata automaticamente

Tabella 5.3 - Coefficiente e p-value del parametro del modello ARIMA(0,1,1)

La fase di stima, con il relativo test di ipotesi sul coefficiente del modello, fornisce un’indicazione positiva circa la bontà della scelta effettuata nella fase di identificazione a favore di un modello ARIMA(0,1,1).

L’analisi termina con l’ultima fase di verifica in cui, attraverso i correlogrammi rappresentati in Figura 5.6, sono valutati i residui stimati dal modello.

Immagine che contiene testo, linea, schermata, diagramma

Descrizione generata automaticamente Figura 5.6 – Correlogramma globale e parziale dei residui ai primi 50 ritardi

Sulla base dei correlogrammi presentati in Figura 5.6, si nota come quasi tutti i coefficienti di autocorrelazione stimati rientrino nella regione di accettazione, ritenendo plausibile l'ipotesi sulla distribuzione dei residui εt ∼ WN.

Il modello scelto quindi, come si evince anche dalla Figura 5.7, risulta idoneo a rappresentare la serie storica dei prezzi delle transazioni degli NFT.

Immagine che contiene testo, schermata, Carattere, Diagramma

Descrizione generata automaticamenteFigura 5.7 – Serie Storica dei Prezzi degli NFT (in nero) e valori dal modello ARIMA(0,1,1) (in rosso)

L’analisi moderna, dunque, sembrerebbe aver prodotto risultati di gran lunga

migliori rispetto a quelli prodotti dall’analisi classica. Il confronto tra le due analisi è di seguito visualizzato in Figura 5.8, ed esplicitato nella Tabella 5.4 con l’accuratezza dei fit attraverso gli indici RMSE e MAE.

Immagine che contiene testo, schermata, Diagramma, linea

Descrizione generata automaticamente Figura 5.8 – Serie Storica dei Prezzi NFT (in nero), trend polinomiale di ordine 3 (in blu) e valori dal modello ARIMA(0,1,1) (in rosso)

Immagine che contiene testo, schermata, Carattere, linea

Descrizione generata automaticamente

Tabella 5.4 – Confronto tra modelli, in termini di RMSE e MAE

La Figura 5.8 mostra chiaramente come il modello ARIMA(0,1,1) si adatti molto meglio all’andamento della serie storica rispetto al trend polinomiale di grado 3. La Tabella 5.4 conferma l’analisi grafica, mostrando come il Root Mean Square Error e il Mean Absolute Error ottenuti dall’analisi moderna siano più che dimezzati rispetto a quelli ottenuti dall’analisi classica.

### 5.3.1 MODELLO ARIMAX

Questa sezione ha per oggetto di studio l’implementazione di un nuovo approccio basato sull’analisi moderna, in cui si ipotizza che la variabile di risposta dipenda non solo dai suoi valori passati ma anche da altri fattori esterni che possono influenzare la sua dinamica. Nello specifico l'analisi consiste nella realizzazione di tanti modelli ARIMAX quante sono le diverse combinazioni tra le variabili esplicative riferite al prezzo e al volume di Ether, ossia “Close ETH ($)”, “Volume ETH (B)”, e le relative variabili ritardate di un giorno. La combinazione migliore, in termine di RMSE e MAE, sarà utilizzata come regressore del modello ARIMAX(0,1,1) per valutare l’influenza della coppia di variabili sui prezzi delle transazioni NFT della collezione BAYC. Tale approccio mira a determinare se queste variabili esterne possono migliorare la capacità predittiva rispetto al precedente modello ARIMA(0,1,1), contribuendo così a una comprensione più approfondita dei fattori che incidono sui prezzi degli NFT.

Nella Tabella 5.5 viene quindi esplicitato un confronto tra i modelli ARIMAX(0,1,1) ottenuti, in termini di RMSE e MAE.

Immagine che contiene testo, schermata, ricevuta, Carattere

Descrizione generata automaticamente

Tabella 5.5 – Confronto tra modelli ARIMAX, in termini di RMSE e MAE

La Tabella 5.5 presenta i seguenti risultati:

* Diversamente da quanto constatato nell’analisi di regressione, la combinazione di regressori “Close ETH\_L1 + Volume ETH” risulta non essere più la migliore per l’analisi del fenomeno;
* La combinazione di regressori “Close ETH + Volume ETH” risulta stavolta essere la migliore, con un RMSE e un MAE pari rispettivamente a 26,5595 e 17,2203. Questo indica che le serie storiche del prezzo close e del volume di Ether della giornata di riferimento risultano essere le più adatte per l’analisi della serie storica dei prezzi NFT di BAYC.

Il modello ARIMAX(0,1,1) che ha per regressori “Close ETH ($)” e “Volume ETH (B)” si configura quindi come il miglior modello tra i possibili, e sarà quindi oggetto di ulteriori analisi attraverso la procedura di Box-Jenkins.

Identificato il modello ARIMAX, l’analisi procede quindi con la fase di stima, all’interno della quale sono stimati i parametri del modello mediante il metodo della massima verosimiglianza; successivamente, viene effettuato il test d’ipotesi sui coefficienti del modello. La Tabella 5.6, di seguito illustrata, descrive accuratamente questa fase di stima.

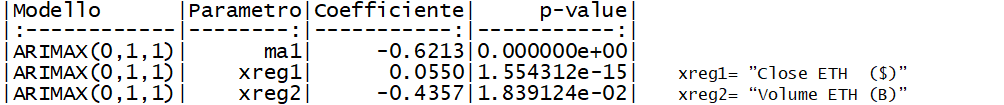


Tabella 5.6 - Coefficiente e p-value dei parametri del modello ARIMAX(0,1,1)

La fase di stima, con i relativi test di ipotesi sui coefficienti del modello, fornisce un’indicazione positiva circa la bontà della scelta effettuata nella fase di identificazione a favore di un modello ARIMA(0,1,1) con regressori “Close ETH ($)” e “Volume ETH (B)”.

L’analisi termina con l’ultima fase di verifica in cui, attraverso i correlogrammi rappresentati in Figura 5.9, sono valutati i residui stimati dal modello.

Immagine che contiene testo, schermata, linea, diagramma

Descrizione generata automaticamente Figura 5.9 – Correlogramma globale e parziale dei residui ai primi 50 ritardi

Sulla base dei correlogrammi presentati in Figura 5.9, si nota come tutti i coefficienti di autocorrelazione stimati rientrino nella regione di accettazione, ritenendo plausibile l'ipotesi sulla distribuzione dei residui εt ∼ WN.

Il modello scelto quindi, come si evince anche dalla Figura 5.10, risulta idoneo a rappresentare la serie storica dei prezzi delle transazioni degli NFT.

Immagine che contiene testo, schermata, Carattere, Diagramma

Descrizione generata automaticamenteFigura 5.10 – Serie Storica dei Prezzi degli NFT (in nero) e valori dal modello ARIMAX(0,1,1) con regressori “Close ETH ($)” e “Volume ETH (B)” (in verde)

L’analisi moderna attraverso i regressori sembrerebbe aver prodotto risultati simili rispetto a quelli prodotti dall’analisi moderna semplice. Il confronto dettagliato tra le due analisi è di seguito visualizzato in Figura 5.11, ed esplicitato nella Tabella 5.7 con l’accuratezza dei fit attraverso gli indici RMSE e MAE.

Immagine che contiene testo, schermata, Carattere, Diagramma

Descrizione generata automaticamente Figura 5.11 – Serie Storica dei Prezzi NFT, valori dal modello ARIMA(0,1,1) (in rosso) e valori dal modello ARIMAX(0,1,1) con regressori “Close ETH ($)” e “Volume ETH (B)” (in verde)

Immagine che contiene testo, schermata, Carattere, linea

Descrizione generata automaticamente

Tabella 5.7 – Confronto tra modelli, in termini di RMSE e MAE

La Figura 5.11 mostra come il modello ARIMAX(0,1,1) si adatti leggermente meglio all’andamento della serie storica rispetto al modello ARIMA(0,1,1). La Tabella 5.7 conferma l’analisi grafica, mostrando come il Root Mean Square Error e il Mean Absolute Error ottenuti dal modello modello ARIMAX siano, seppur di poco, leggermente migliori rispetto a quelli ottenuti dal modello ARIMA.

## 5.4 PREVISIONI

La fase di previsione elaborata in questa sezione si configura come un passo cruciale nell'analisi della serie storica dei prezzi NFT di Bored Ape Yacht Club. Difatti la metodologia adottata, basata sull'approccio di Box-Jenkins, permette di identificare un modello appropriato che può essere impiegato per proiettare futuri andamenti della serie temporale. In questo contesto, l'intervallo di analisi della serie storica di analisi viene ridotto al periodo che va dal 2 maggio 2021 al 2 maggio 2023; le successive osservazioni, sino al 12 settembre 2023, sono invece utilizzate per la valutazione delle previsioni.

Il modello auspicabile per queste proiezioni sarebbe il miglior modello individuato in precedenza, ovvero l'ARIMAX(0,1,1) che include come regressori le serie storiche dei prezzi e del volume di Ether relative al giorno di riferimento. Tuttavia, è importante sottolineare che, nella fase di previsione al tempo t+1, i valori dei regressori al tempo t+1 non sono noti. Considerando quindi questa limitazione, l’approccio scelto è quello di includere nel modello le due variabili con i ritardi. Tale decisione trova sostegno nei dati esposti nella Tabella 5.5, i quali rivelano un RMSE pari a 27,4629 e un MAE di 17,4183 per tali variabili: questi valori di errore sono da considerarsi accettabili, consolidando così la scelta adottata. Questa strategia, seppur non ottimale, si presenta come una soluzione pragmatica che può offrire un supporto utile nella fase di previsione.

Per massimizzare l'affidabilità delle previsioni è stato adottato un approccio ciclico. Il processo coinvolge la creazione di un modello utilizzando le prime 706 osservazioni per prevedere la 707esima, successivamente la stima dello stesso modello con 707 osservazioni per prevedere la 708esima, e così via fino a un modello basato sulle 822 osservazioni per prevedere la 823esima. Questo approccio consente di rendere ogni previsione il più affidabile possibile, poiché non dipende da altre previsioni. In questo modo, si riduce l'esposizione all'incertezza e si ottengono previsioni più stabili e informative.

Nella Figura 5.12, è possibile visualizzare graficamente il confronto tra le osservazioni realmente osservate nella serie storica dei prezzi NFT e le previsioni teoriche generate mediante l'impiego del semplice modello ARIMA(0,1,1) e del modello ARIMAX(0,1,1). Nella Tabella 5.8, invece, viene effettuato un confronto della bontà delle previsioni tramite gli indici del Root Mean Square Error e del Mean Absolut Error.

Immagine che contiene testo, schermata, diagramma, Diagramma

Descrizione generata automaticamenteFigura 5.12 – Serie Storica dei Prezzi degli NFT (in nero), previsione basata sul modello ARIMA(0,1,1) (in rosso) e previsione basata sul modello ARIMAX(0,1,1) (in verde)

Immagine che contiene testo, schermata, Carattere, linea

Descrizione generata automaticamente

Tabella 5.8 – Bontà delle previsioni del semplice modello ARIMA(0,1,1) e del modello ARIMAX(0,1,1), in termini di RMSE e MAE

Sulla base dei dati osservati e delle previsioni generate dai modelli è possibile osservare che le previsioni seguono in gran parte l'andamento generale della serie storica dei prezzi NFT di Bored Ape Yacht Club. Tuttavia, si notano alcune variazioni tra le previsioni e i valori effettivamente registrati. Ad esempio, nei primi periodi, le previsioni sembrano adattarsi bene alle fluttuazioni della serie storica, ma successivamente si verificano alcune discrepanze, specialmente in corrispondenza di picchi e minimi. Queste discrepanze, leggermente più accentuate nel modello ARIMAX rispetto al semplice ARIMA, potrebbero indicare la presenza di variazioni non catturate completamente dal modello o l'impatto di fattori esterni non inclusi nei regressori. Complessivamente però, come mostrano anche gli indici RMSE e MAE della Tabella 5.8, entrambi i modelli sembrano essere in grado di catturare, con una discreta precisione, le tendenze generali della serie storica.

# CONCLUSIONI

L'approfondita analisi condotta sul mercato dei non-fungible token, focalizzata sulla collezione Bored Ape Yacht Club, ha evidenziato la possibilità di analizzare questo mercato, pur riconoscendone la continua evoluzione, elevata volatilità e suscettibilità a influenze esterne. In particolare, è emerso che il prezzo degli NFT della collezione è strettamente correlato al prezzo e al volume di Ether nella giornata di scambio. Il prezzo in costante evoluzione di Ether, infatti, si è rivelato un fattore cruciale che influisce in tempo reale sui prezzi delle transazioni degli NFT.

L'analisi inoltre ha dimostrato come, pur non utilizzando i prezzi delle giornate di riferimento, bensì quelli dei giorni precedenti, sia possibile formulare previsioni soddisfacenti per i prezzi delle transazioni NFT. Tuttavia, è fondamentale tenere sempre in considerazione l'importante volatilità di questo mercato altamente influenzabile, il che rende sempre presente un forte grado di incertezza nelle previsioni a lungo termine.

Il lavoro svolto in questo elaborato costituisce quindi un punto di partenza per la conoscenza del mercato NFT. Tra le future direzioni di ricerca, potrebbe essere interessante esplorare l'applicazione di reti neurali, in particolare le reti neurali ricorrenti (RNN) e le reti neurali Long Short-Term Memory (LSTM), per migliorare l’analisi e la previsione dei prezzi del mercato NFT. Un approccio avanzato del genere sfrutterebbe la capacità delle reti neurali di apprendere modelli complessi dalle sequenze temporali, consentendo una migliore cattura delle dinamiche di questo mercato altamente volatile.

# APPENDICE – Codice R

##### CARICA PACCHETTI ####

library(readr)

library(dplyr)

library(openxlsx)

library(readxl)

library(ggplot2)

library(zoo)

library(lubridate)

library(knitr)

library(tseries)

library(forecast)

##### PREPARAZIONE DEI DATI ####

BAYC\_mag\_2021 <- read\_csv("BAYC/BAYC\_2021/BAYC\_mag\_2021.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_mag\_2021\_1 <- read\_csv("BAYC/BAYC\_2021/BAYC\_mag\_2021(1).csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_mag\_2021\_2 <- read\_csv("BAYC/BAYC\_2021/BAYC\_mag\_2021(2).csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_mag\_2021\_3 <- read\_csv("BAYC/BAYC\_2021/BAYC\_mag\_2021(3).csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_mag\_2021\_4 <- read\_csv("BAYC/BAYC\_2021/BAYC\_mag\_2021(4).csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_mag\_2021\_5 <- read\_csv("BAYC/BAYC\_2021/BAYC\_mag\_2021(5).csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_giu\_2021 <- read\_csv("BAYC/BAYC\_2021/BAYC\_giu\_2021.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_giu\_2021\_1 <- read\_csv("BAYC/BAYC\_2021/BAYC\_giu\_2021(1).csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_giu\_2021\_2 <- read\_csv("BAYC/BAYC\_2021/BAYC\_giu\_2021(2).csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_giu\_2021\_3 <- read\_csv("BAYC/BAYC\_2021/BAYC\_giu\_2021(3).csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_giu\_2021\_4 <- read\_csv("BAYC/BAYC\_2021/BAYC\_giu\_2021(4).csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_giu\_2021\_5 <- read\_csv("BAYC/BAYC\_2021/BAYC\_giu\_2021(5).csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_lug\_2021 <- read\_csv("BAYC/BAYC\_2021/BAYC\_lug\_2021.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_lug\_2021\_1 <- read\_csv("BAYC/BAYC\_2021/BAYC\_lug\_2021(1).csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_lug\_2021\_2 <- read\_csv("BAYC/BAYC\_2021/BAYC\_lug\_2021(2).csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_ago\_2021 <- read\_csv("BAYC/BAYC\_2021/BAYC\_ago\_2021.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_ago\_2021\_1 <- read\_csv("BAYC/BAYC\_2021/BAYC\_ago\_2021(1).csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_ago\_2021\_2 <- read\_csv("BAYC/BAYC\_2021/BAYC\_ago\_2021(2).csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_set\_2021 <- read\_csv("BAYC/BAYC\_2021/BAYC\_set\_2021.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_ott\_2021 <- read\_csv("BAYC/BAYC\_2021/BAYC\_ott\_2021.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_nov\_2021 <- read\_csv("BAYC/BAYC\_2021/BAYC\_nov\_2021.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_dic\_2021 <- read\_csv("BAYC/BAYC\_2021/BAYC\_dic\_2021.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_2021 <- bind\_rows(

BAYC\_mag\_2021,

BAYC\_mag\_2021\_1,

BAYC\_mag\_2021\_2,

BAYC\_mag\_2021\_3,

BAYC\_mag\_2021\_4,

BAYC\_mag\_2021\_5,

BAYC\_giu\_2021,

BAYC\_giu\_2021\_1,

BAYC\_giu\_2021\_2,

BAYC\_giu\_2021\_3,

BAYC\_giu\_2021\_4,

BAYC\_giu\_2021\_5,

BAYC\_lug\_2021,

BAYC\_lug\_2021\_1,

BAYC\_lug\_2021\_2,

BAYC\_ago\_2021,

BAYC\_ago\_2021\_1,

BAYC\_ago\_2021\_2,

BAYC\_set\_2021,

BAYC\_ott\_2021,

BAYC\_nov\_2021,

BAYC\_dic\_2021)

BAYC\_2021\_OpenSea <- BAYC\_2021 %>%

filter(Market == "OpenSea")

write.xlsx(BAYC\_2021\_OpenSea, "BAYC/BAYC\_2021/BAYC\_2021\_OpenSea.xlsx", rowNames = FALSE)

BAYC\_2021 <- read\_excel("BAYC/BAYC\_2021/BAYC\_2021.xlsx")

BAYC\_2021\_clean <- BAYC\_2021[complete.cases(BAYC\_2021$Price), ]

BAYC\_2021\_clean$Price <- as.numeric(gsub(",", "", BAYC\_2021\_clean$Price))

BAYC\_2021\_clean$Price <- BAYC\_2021\_clean$Price / 1000

BAYC\_2021\_clean$Price <- round(BAYC\_2021\_clean$Price, 3)

BAYC\_2021\_clean$Price <- sprintf("%.3f", BAYC\_2021\_clean$Price)

colnames(BAYC\_2021\_clean)[colnames(BAYC\_2021\_clean) == "Price"] <- "Price (k)"

BAYC\_2021\_clean <- BAYC\_2021\_clean[as.numeric(BAYC\_2021\_clean$`Price (k)`) != 0, ]

BAYC\_2021\_clean$`Price (k)` <- as.numeric(gsub(",", "", BAYC\_2021\_clean$`Price (k)`))

write.xlsx(BAYC\_2021\_clean, "BAYC/BAYC\_2021/BAYC\_2021\_clean.xlsx", rowNames = FALSE)

BAYC\_gen\_2022 <- read\_csv("BAYC/BAYC\_2022/BAYC\_gen\_2022.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_feb\_2022 <- read\_csv("BAYC/BAYC\_2022/BAYC\_feb\_2022.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_mar\_2022 <- read\_csv("BAYC/BAYC\_2022/BAYC\_mar\_2022.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_apr\_2022 <- read\_csv("BAYC/BAYC\_2022/BAYC\_apr\_2022.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_mag\_2022 <- read\_csv("BAYC/BAYC\_2022/BAYC\_mag\_2022.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_giu\_2022 <- read\_csv("BAYC/BAYC\_2022/BAYC\_giu\_2022.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_lug\_2022 <- read\_csv("BAYC/BAYC\_2022/BAYC\_lug\_2022.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_ago\_2022 <- read\_csv("BAYC/BAYC\_2022/BAYC\_ago\_2022.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_set\_2022 <- read\_csv("BAYC/BAYC\_2022/BAYC\_set\_2022.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_ott\_2022 <- read\_csv("BAYC/BAYC\_2022/BAYC\_ott\_2022.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_nov\_2022 <- read\_csv("BAYC/BAYC\_2022/BAYC\_nov\_2022.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_dic\_2022 <- read\_csv("BAYC/BAYC\_2022/BAYC\_dic\_2022.csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_dic\_2022\_1 <- read\_csv("BAYC/BAYC\_2022/BAYC\_dic\_2022(1).csv",

col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_2022 <- bind\_rows(

BAYC\_gen\_2022,

BAYC\_feb\_2022,

BAYC\_mar\_2022,

BAYC\_apr\_2022,

BAYC\_mag\_2022,

BAYC\_giu\_2022,

BAYC\_lug\_2022,

BAYC\_ago\_2022,

BAYC\_set\_2022,

BAYC\_ott\_2022,

BAYC\_nov\_2022,

BAYC\_dic\_2022,

BAYC\_dic\_2022\_1)

BAYC\_2022\_OpenSea <- BAYC\_2022 %>%

filter(Market == "OpenSea")

write.xlsx(BAYC\_2022\_OpenSea, "BAYC/BAYC\_2022/BAYC\_2022\_OpenSea.xlsx", rowNames = FALSE)

BAYC\_2022 <- read\_excel("BAYC/BAYC\_2022/BAYC\_2022.xlsx")

BAYC\_2022\_clean <- BAYC\_2022[complete.cases(BAYC\_2022$Price), ]

BAYC\_2022\_clean$Price <- as.numeric(gsub(",", "", BAYC\_2022\_clean$Price))

BAYC\_2022\_clean$Price <- BAYC\_2022\_clean$Price / 1000

BAYC\_2022\_clean$Price <- round(BAYC\_2022\_clean$Price, 3)

BAYC\_2022\_clean$Price <- sprintf("%.3f", BAYC\_2022\_clean$Price)

colnames(BAYC\_2022\_clean)[colnames(BAYC\_2022\_clean) == "Price"] <- "Price (k)"

BAYC\_2022\_clean <- BAYC\_2022\_clean[as.numeric(BAYC\_2022\_clean$`Price (k)`) != 0, ]

BAYC\_2022\_clean$`Price (k)` <- as.numeric(gsub(",", "", BAYC\_2022\_clean$`Price (k)`))

write.xlsx(BAYC\_2022\_clean, "BAYC/BAYC\_2022/BAYC\_2022\_clean.xlsx", rowNames = FALSE)

BAYC\_gen\_2023 <- read\_csv("BAYC/BAYC\_2023/BAYC\_gen\_2023.csv", col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_feb\_2023 <- read\_csv("BAYC/BAYC\_2023/BAYC\_feb\_2023.csv", col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_feb\_2023\_1 <- read\_csv("BAYC/BAYC\_2023/BAYC\_feb\_2023(1).csv", col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_mar\_2023 <- read\_csv("BAYC/BAYC\_2023/BAYC\_mar\_2023.csv", col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_apr\_2023 <- read\_csv("BAYC/BAYC\_2023/BAYC\_apr\_2023.csv", col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_apr\_2023\_1 <- read\_csv("BAYC/BAYC\_2023/BAYC\_apr\_2023(1).csv", col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_mag\_2023 <- read\_csv("BAYC/BAYC\_2023/BAYC\_mag\_2023.csv", col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_giu\_2023 <- read\_csv("BAYC/BAYC\_2023/BAYC\_giu\_2023.csv", col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_lug\_2023 <- read\_csv("BAYC/BAYC\_2023/BAYC\_lug\_2023.csv", col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_ago\_2023 <- read\_csv("BAYC/BAYC\_2023/BAYC\_ago\_2023.csv", col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_set\_2023 <- read\_csv("BAYC/BAYC\_2023/BAYC\_set\_2023.csv", col\_types = cols(`Txn Hash` = col\_skip(),

Blockno = col\_skip(), UnixTimestamp = col\_skip(),

Action = col\_skip(), Buyer = col\_skip(),

NFT = col\_skip(), Type = col\_skip(),

Quantity = col\_skip()))

BAYC\_2023 <- bind\_rows(

BAYC\_gen\_2023,

BAYC\_feb\_2023,

BAYC\_feb\_2023\_1,

BAYC\_mar\_2023,

BAYC\_apr\_2023,

BAYC\_apr\_2023\_1,

BAYC\_mag\_2023,

BAYC\_giu\_2023,

BAYC\_lug\_2023,

BAYC\_ago\_2023,

BAYC\_set\_2023)

BAYC\_2023\_OpenSea <- BAYC\_2023 %>%

filter(Market == "OpenSea")

write.xlsx(BAYC\_2023\_OpenSea, "BAYC/BAYC\_2023/BAYC\_2023\_OpenSea.xlsx", rowNames = FALSE)

BAYC\_2023 <- read\_excel("BAYC/BAYC\_2023/BAYC\_2023.xlsx")

BAYC\_2023\_clean <- BAYC\_2023[complete.cases(BAYC\_2023$Price), ]

BAYC\_2023\_clean$Price <- as.numeric(gsub(",", "", BAYC\_2023\_clean$Price))

BAYC\_2023\_clean$Price <- BAYC\_2023\_clean$Price / 1000

BAYC\_2023\_clean$Price <- round(BAYC\_2023\_clean$Price, 3)

BAYC\_2023\_clean$Price <- sprintf("%.3f", BAYC\_2023\_clean$Price)

colnames(BAYC\_2023\_clean)[colnames(BAYC\_2023\_clean) == "Price"] <- "Price (k)"

BAYC\_2023\_clean <- BAYC\_2023\_clean[as.numeric(BAYC\_2023\_clean$`Price (k)`) != 0, ]

BAYC\_2023\_clean$`Price (k)` <- as.numeric(gsub(",", "", BAYC\_2023\_clean$`Price (k)`))

write.xlsx(BAYC\_2023\_clean, "BAYC/BAYC\_2023/BAYC\_2023\_clean.xlsx", rowNames = FALSE)

BAYC <- bind\_rows(

BAYC\_2021\_clean,

BAYC\_2022\_clean,

BAYC\_2023\_clean)

BAYC=subset(BAYC, !duplicated(BAYC))

BAYC$`DateTime (UTC)` <- as.Date(BAYC$`DateTime (UTC)`)

colnames(BAYC)[colnames(BAYC) == "DateTime (UTC)"] <- "Date"

BAYC\_RarityRanking <- read\_excel("BAYC/BAYC\_RarityRanking.xlsx")

BAYC <- merge(BAYC, BAYC\_RarityRanking, by = "ID", all = TRUE)

BAYC <- na.omit(BAYC)

BAYC$ID <- as.factor(BAYC$ID)

BAYC$Rank <- as.factor(BAYC$Rank)

BAYC$Rarity <- as.factor(BAYC$Rarity)

ETH\_USD <- read\_csv("BAYC/ETH\_USD.csv", col\_types = cols(Open = col\_skip(),

High = col\_skip(), Low = col\_skip(),

`Adj Close` = col\_skip()))

ETH\_USD$Date=as.Date(ETH\_USD$Date)

ETH\_USD$Volume <- as.numeric(gsub(",", "", ETH\_USD$Volume))

ETH\_USD$Volume <- ETH\_USD$Volume / 1000000000

ETH\_USD$Volume <- round(ETH\_USD$Volume, 9)

ETH\_USD$Volume <- sprintf("%.9f", ETH\_USD$Volume)

ETH\_USD$Volume <- as.numeric(gsub(",", "", ETH\_USD$Volume))

ETH\_USD <- ETH\_USD %>% mutate(Close\_L1 = lag(Close, 1))

ETH\_USD <- ETH\_USD %>% mutate(Volume\_L1 = lag(Volume, 1))

ETH\_USD=ETH\_USD[, c("Date","Close","Close\_L1","Volume","Volume\_L1")]

BAYC <- merge(BAYC, ETH\_USD, by = "Date", all = TRUE)

View(BAYC)

colnames(BAYC)[colnames(BAYC) == "Close"] <- "Close ETH ($)"

colnames(BAYC)[colnames(BAYC) == "Close\_L1"] <- "Close ETH\_L1 ($)"

colnames(BAYC)[colnames(BAYC) == "Volume"] <- "Volume ETH (B)"

colnames(BAYC)[colnames(BAYC) == "Volume\_L1"] <- "Volume ETH\_L1 (B)"

BAYC\_MITICO=subset(BAYC, Rarity == "MITICO")

BAYC\_LEGGENDARIO=subset(BAYC, Rarity == "LEGGENDARIO")

BAYC\_EPICO=subset(BAYC, Rarity == "EPICO")

BAYC\_RARO=subset(BAYC, Rarity == "RARO")

BAYC\_NON\_COMUNE=subset(BAYC, Rarity == "NON COMUNE")

BAYC\_COMUNE=subset(BAYC, Rarity == "COMUNE")

which(BAYC\_NON\_COMUNE$`Price (k)`>800)

BAYC\_NON\_COMUNE[c(6012,6272),"Price (k)"]

mean(c(mean(BAYC\_NON\_COMUNE[(6012-3):(6012-1),"Price (k)"]),mean(BAYC\_NON\_COMUNE[(6012+1):(6012+3),"Price (k)"])))

mean(c(mean(BAYC\_NON\_COMUNE[(6272-3):(6272-1),"Price (k)"]),mean(BAYC\_NON\_COMUNE[(6272+1):(6272+3),"Price (k)"])))

which(BAYC$Rarity == "NON COMUNE" & BAYC$`Price (k)` > 800)

BAYC[22335,"Price (k)"]

BAYC[23269,"Price (k)"]

BAYC[22335,"Price (k)"]=mean(c(mean(BAYC\_NON\_COMUNE[(6012-3):(6012-1),"Price (k)"]),mean(BAYC\_NON\_COMUNE[(6012+1):(6012+3),"Price (k)"])))

BAYC[23269,"Price (k)"]=mean(c(mean(BAYC\_NON\_COMUNE[(6272-3):(6272-1),"Price (k)"]),mean(BAYC\_NON\_COMUNE[(6272+1):(6272+3),"Price (k)"])))

which(BAYC\_COMUNE$`Price (k)`>1000)

BAYC\_COMUNE[c(6073,6851),"Price (k)"]

mean(c(mean(BAYC\_COMUNE[(6073-3):(6073-1),"Price (k)"]),mean(BAYC\_COMUNE[(6073+1):(6073+3),"Price (k)"])))

mean(c(mean(BAYC\_COMUNE[(6851-3):(6851-1),"Price (k)"]),mean(BAYC\_COMUNE[(6851+1):(6851+3),"Price (k)"])))

which(BAYC$Rarity == "COMUNE" & BAYC$`Price (k)` > 1000)

BAYC[20338,"Price (k)"]

BAYC[22876,"Price (k)"]

BAYC[20338,"Price (k)"]=mean(c(mean(BAYC\_COMUNE[(6073-3):(6073-1),"Price (k)"]),mean(BAYC\_COMUNE[(6073+1):(6073+3),"Price (k)"])))

BAYC[22876,"Price (k)"]=mean(c(mean(BAYC\_COMUNE[(6851-3):(6851-1),"Price (k)"]),mean(BAYC\_COMUNE[(6851+1):(6851+3),"Price (k)"])))

which(BAYC\_LEGGENDARIO$`Price (k)`>1000)

mean(BAYC\_LEGGENDARIO[c(426,437,456,466,468,479,504,522,524,555,556,603,609),"Price (k)"])

which(BAYC\_LEGGENDARIO$`Price (k)`>1000 & BAYC\_LEGGENDARIO$`Close ETH ($)`<1504.92)

BAYC\_LEGGENDARIO[609,"Price (k)"]

mean(c(mean(BAYC\_LEGGENDARIO[(609-3):(609-1),"Price (k)"]),mean(BAYC\_LEGGENDARIO[(609+1):(609+3),"Price (k)"])))

which(BAYC$Rarity == "LEGGENDARIO" & BAYC$`Price (k)` > 1000 & BAYC$`Close ETH ($)`<1504.92)

BAYC[22989,"Price (k)"]

BAYC[22989,"Price (k)"]=mean(c(mean(BAYC\_LEGGENDARIO[(609-3):(609-1),"Price (k)"]),mean(BAYC\_LEGGENDARIO[(609+1):(609+3),"Price (k)"])))

##### ANALISI DESCRITTIVA ####

summary(BAYC)

summary(BAYC$Date)

barplot(table(BAYC$Date), main="Distribuzione delle Transazioni per Data", xlab="Data", ylab="Frequenza", col="lightblue")

round((sum(BAYC$Date == "2021-05-01")/length(BAYC$Date))\*100, 2)

round((sum(BAYC$Date == "2021-05-01")/length(BAYC$Date))\*100, 2)+round((sum(BAYC$Date == "2021-05-02")/length(BAYC$Date))\*100, 2)+round((sum(BAYC$Date == "2021-05-03")/length(BAYC$Date))\*100, 2)

round((sum(BAYC$Date >= "2021-05-01" & BAYC$Date <= "2021-08-01")/length(BAYC$Date))\*100, 2)

round((sum(format(BAYC$Date, "%Y") == "2021")/length(BAYC$Date))\*100, 2)

round((sum(format(BAYC$Date, "%Y") == "2023")/length(BAYC$Date))\*100, 2)

round((sum(BAYC$Date >= "2023-06-01" & BAYC$Date <= "2023-09-12")/length(BAYC$Date))\*100, 2)

plot(BAYC$Date, BAYC$`Price (k)`, type="l", main="Prezzo Medio delle Transazioni per Data",

xlab="Data", ylab="Prezzo (k)")

boxplot(BAYC$`Price (k)` ~ as.factor(format(BAYC$Date, "%Y")),

main="Boxplot dei Prezzi delle Transazioni (per Anno)", xlab="Anno", ylab="Prezzo (k)", col="lightblue")

boxplot(BAYC$`Price (k)` ~ as.factor(format(BAYC$Date, "%Y-%m")),

main="Boxplot dei Prezzi delle Transazioni (per Mese)", xlab="Mese", ylab="Prezzo (k)", col="lightblue")

summary(BAYC$ID)

barplot(table(BAYC$ID), main="Distribuzione delle Transazioni per ID", xlab="ID", ylab="Frequenza")

round(prop.table(table(table(BAYC$ID)))\*100, 2)

summary(BAYC$`Price (k)`)

hist(BAYC$`Price (k)`, breaks = 20, col = "lightblue", main = "Distribuzione dei Prezzi degli NFT", xlab = "Prezzo (k)", ylab = "Frequenza")

round((nrow(BAYC[BAYC$`Price (k)` <= 500, ])/nrow(BAYC))\*100, 2)

round((nrow(BAYC[BAYC$`Price (k)` <= 400, ])/nrow(BAYC))\*100, 2)

boxplot(BAYC$`Price (k)`, col = "lightblue", main = "Boxplot dei Prezzi degli NFT", ylab = "Prezzo (k)")

plot(ecdf(BAYC$`Price (k)`), col = "lightblue", main = "Distribuzione Cumulativa dei Prezzi degli NFT", xlab = "Prezzo (k)", ylab = "Probabilità Cumulativa")

summary(BAYC$Rarity)

round(prop.table(table(BAYC$Rarity)) \* 100, 2)

barplot(table(BAYC$Rarity), col = "lightblue", main = "Distribuzione della Rarità", xlab = "Rarità", ylab = "Frequenza")

pie(prop.table(table(BAYC$Rarity)) \* 100, labels = names(prop.table(table(BAYC$Rarity)) \* 100), main = "Grafico a Torta della Rarità")

boxplot(BAYC$`Price (k)` ~ BAYC$Rarity, col = "lightblue", main = "Boxplot dei Prezzi degli NFT (per Rarità)", xlab = "Rarità", ylab = "Prezzo (k)")

tapply(BAYC$`Price (k)`, BAYC$Rarity, mean)

tapply(BAYC$`Price (k)`, BAYC$Rarity, sd)

summary(BAYC$`Close ETH ($)`)

ggplot(BAYC, aes(x = Date, y = `Close ETH ($)`)) +

geom\_line() +

labs(title = "Serie Storica dei Prezzi Close di Ether", x = "Data", y = "Prezzo Close ($)") +

theme\_minimal()

mean(BAYC$`Close ETH ($)`[substr(BAYC$`Date`, 1, 4) == "2021"])

mean(BAYC$`Close ETH ($)`[substr(BAYC$`Date`, 1, 4) == "2022"])

mean(BAYC$`Close ETH ($)`[substr(BAYC$`Date`, 1, 4) == "2023"])

sd(BAYC$`Close ETH ($)`[substr(BAYC$`Date`, 1, 4) == "2021"])

sd(BAYC$`Close ETH ($)`[substr(BAYC$`Date`, 1, 4) == "2022"])

sd(BAYC$`Close ETH ($)`[substr(BAYC$`Date`, 1, 4) == "2023"])

sd(BAYC$`Close ETH ($)`)

ggplot(BAYC, aes(x = Date, y = c(0, diff(`Close ETH ($)`)))) +

geom\_line() +

labs(title = "Differenze Giornaliere nei Prezzi Close di Ether", x = "Data", y = "Differenza Giornaliera ($)") +

theme\_minimal()

mean(diff(BAYC$`Close ETH ($)`[substr(BAYC$`Date`, 1, 4) == "2021"]), na.rm = TRUE)

mean(diff(BAYC$`Close ETH ($)`[substr(BAYC$`Date`, 1, 4) == "2022"]), na.rm = TRUE)

mean(diff(BAYC$`Close ETH ($)`[substr(BAYC$`Date`, 1, 4) == "2023"]), na.rm = TRUE)

mean(diff(BAYC$`Close ETH ($)`), na.rm = TRUE)

sd(diff(BAYC$`Close ETH ($)`[substr(BAYC$`Date`, 1, 4) == "2021"]), na.rm = TRUE)

sd(diff(BAYC$`Close ETH ($)`[substr(BAYC$`Date`, 1, 4) == "2022"]), na.rm = TRUE)

sd(diff(BAYC$`Close ETH ($)`[substr(BAYC$`Date`, 1, 4) == "2023"]), na.rm = TRUE)

sd(diff(BAYC$`Close ETH ($)`, na.rm = TRUE))

ggplot(data = BAYC, aes(x = format(Date, "%Y"), y = `Close ETH ($)`)) +

geom\_boxplot() +

labs(title = "Boxplot dei Prezzi Close di Ether (per Anno)",

x = "Anno",

y = "Prezzo Close ($)")

summary(BAYC$`Volume ETH (B)`)

ggplot(na.omit(BAYC %>%

group\_by(Date) %>%

summarize(Average\_Volume = mean(`Volume ETH (B)`))), aes(x = Date, y = Average\_Volume)) +

geom\_bar(stat = "identity", fill = "blue") +

labs(

title = "Volume Medio di Ether per Data",

x = "Data",

y = "Volume Medio (B)"

) +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

ggplot(BAYC, aes(x = factor(substr(Date, 1, 4)), y = `Volume ETH (B)`)) +

geom\_boxplot(fill = "lightblue") +

labs(

title = "Boxplot del Volume di Ether (per Anno)",

x = "Anno",

y = "Volume (B)"

) +

theme\_minimal()

ggplot(na.omit(BAYC), aes(x = `Rarity`, y = `Volume ETH (B)`, fill = `Rarity`)) +

geom\_bar(stat = "identity") +

labs(

title = "Volume di BAYC per Livello di Rarità",

x = "Rarità",

y = "Volume (B)"

) +

theme\_minimal()

mean(BAYC$`Volume ETH (B)`[substr(BAYC$`Date`, 1, 4) == "2021"])

mean(BAYC$`Volume ETH (B)`[substr(BAYC$`Date`, 1, 4) == "2022"])

mean(BAYC$`Volume ETH (B)`[substr(BAYC$`Date`, 1, 4) == "2023"])

mean(BAYC$`Volume ETH (B)`)

sd(BAYC$`Volume ETH (B)`[substr(BAYC$`Date`, 1, 4) == "2021"])

sd(BAYC$`Volume ETH (B)`[substr(BAYC$`Date`, 1, 4) == "2022"])

sd(BAYC$`Volume ETH (B)`[substr(BAYC$`Date`, 1, 4) == "2023"])

sd(BAYC$`Volume ETH (B)`)

##### ANALISI ####

model\_BAYC=lm(`Price (k)`~ `Close ETH ($)`+`Volume ETH (B)`,data=BAYC)

model\_BAYC\_Close\_L1=lm(`Price (k)`~ `Close ETH\_L1 ($)`+`Volume ETH (B)`,data=BAYC)

model\_BAYC\_Volume\_L1=lm(`Price (k)`~ `Close ETH ($)`+`Volume ETH\_L1 (B)`,data=BAYC)

model\_BAYC\_L1=lm(`Price (k)`~ `Close ETH\_L1 ($)`+`Volume ETH\_L1 (B)`,data=BAYC)

knitr::kable(data.frame(Model = c("Close ETH + Volume ETH", "Close ETH (L1) + Volume ETH", "Close ETH + Volume ETH (L1)", "Close ETH (L1) + Volume ETH (L1)"),

`Adjusted R-squared` = c(summary(model\_BAYC)$adj.r.squared, summary(model\_BAYC\_Close\_L1)$adj.r.squared, summary(model\_BAYC\_Volume\_L1)$adj.r.squared, summary(model\_BAYC\_L1)$adj.r.squared),

`p-value` = c(((summary(model\_BAYC))$coefficients[2,4] + (summary(model\_BAYC))$coefficients[3,4]),

((summary(model\_BAYC\_Close\_L1))$coefficients[2,4] + (summary(model\_BAYC\_Close\_L1))$coefficients[3,4]),

((summary(model\_BAYC\_Volume\_L1))$coefficients[2,4] + (summary(model\_BAYC\_Volume\_L1))$coefficients[3,4]),

(summary(model\_BAYC\_L1))$coefficients[2,4] + (summary(model\_BAYC\_L1))$coefficients[3,4])),format = "markdown")

knitr::kable(data.frame(Variabile = c("Close ETH (L1)", "Volume ETH"),

`Coefficiente` = c(summary(model\_BAYC\_Close\_L1)$coefficients[2,1], summary(model\_BAYC\_Close\_L1)$coefficients[3,1]),

`p-value` = c((summary(model\_BAYC\_Close\_L1))$coefficients[2,4], (summary(model\_BAYC\_Close\_L1))$coefficients[3,4])), format = "markdown")

##### ANALISI PER RARITA' ####

ggplot(data = na.omit(BAYC), aes(x = `Close ETH\_L1 ($)`, y = `Price (k)`, color = Rarity)) +

geom\_point() +

labs(

title = "Scatterplot tra Prezzi degli NFT e Prezzi di Ether (L1)",

x = "Prezzi ETH\_L1 ($)",

y = "Prezzi NFT (k)",

color = "Rarità"

) +

facet\_wrap(~Rarity, scales = "free")

ggplot(data = na.omit(BAYC), aes(x = `Volume ETH (B)`, y = `Price (k)`, color = Rarity)) +

geom\_point() +

labs(

title = "Scatterplot tra Prezzi degli NFT e Volume di Ether",

x = "Volume ETH (B)",

y = "Prezzi NFT (k)",

color = "Rarità"

) +

facet\_wrap(~Rarity, scales = "free")

model\_MITICO <- lm(`Price (k)` ~ `Close ETH\_L1 ($)` + `Volume ETH (B)`, data = BAYC\_MITICO)

model\_LEGGENDARIO <- lm(`Price (k)` ~ `Close ETH\_L1 ($)` + `Volume ETH (B)`, data = BAYC\_LEGGENDARIO)

model\_EPICO <- lm(`Price (k)` ~ `Close ETH\_L1 ($)` + `Volume ETH (B)`, data = BAYC\_EPICO)

model\_RARO <- lm(`Price (k)` ~ `Close ETH\_L1 ($)` + `Volume ETH (B)`, data = BAYC\_RARO)

model\_NON\_COMUNE <- lm(`Price (k)` ~ `Close ETH\_L1 ($)` + `Volume ETH (B)`, data = BAYC\_NON\_COMUNE)

model\_COMUNE <- lm(`Price (k)` ~ `Close ETH\_L1 ($)` + `Volume ETH (B)`, data = BAYC\_COMUNE)

knitr::kable(data.frame(Model = c("MITICO", "LEGGENDARIO", "EPICO", "RARO", "NON COMUNE", "COMUNE"),

`Adjusted R-squared` = c(summary(model\_MITICO)$adj.r.squared,summary(model\_LEGGENDARIO)$adj.r.squared,summary(model\_EPICO)$adj.r.squared,summary(model\_RARO)$adj.r.squared,summary(model\_NON\_COMUNE)$adj.r.squared,summary(model\_COMUNE)$adj.r.squared),

`p-value` = c(summary(model\_MITICO)$coefficients[2, 4] + summary(model\_MITICO)$coefficients[3, 4],summary(model\_LEGGENDARIO)$coefficients[2, 4] + summary(model\_LEGGENDARIO)$coefficients[3, 4],summary(model\_EPICO)$coefficients[2, 4] + summary(model\_EPICO)$coefficients[3, 4],summary(model\_RARO)$coefficients[2, 4] + summary(model\_RARO)$coefficients[3, 4],summary(model\_NON\_COMUNE)$coefficients[2, 4] + summary(model\_NON\_COMUNE)$coefficients[3, 4],summary(model\_COMUNE)$coefficients[2, 4] + summary(model\_COMUNE)$coefficients[3, 4])),format = "markdown")

knitr::kable(data.frame(Rarità = c("MITICO", "MITICO"), Variabile = c("Close ETH\_L1", "Volume ETH"),

`Coefficiente` = c((summary(model\_MITICO))$coefficients[2,1], (summary(model\_MITICO))$coefficients[3,1]),

`p-value` = c(((summary(model\_MITICO))$coefficients[2,4] + (summary(model\_MITICO))$coefficients[3,4]),

((summary(model\_MITICO))$coefficients[2,4] + (summary(model\_MITICO))$coefficients[3,4]))), format = "markdown")

knitr::kable(data.frame(Rarità = c("LEGGENDARIO", "LEGGENDARIO"), Variabile = c("Close ETH\_L1", "Volume ETH"),

`Coefficiente` = c((summary(model\_LEGGENDARIO))$coefficients[2,1], (summary(model\_LEGGENDARIO))$coefficients[3,1]),

`p-value` = c(((summary(model\_LEGGENDARIO))$coefficients[2,4] + (summary(model\_LEGGENDARIO))$coefficients[3,4]),

((summary(model\_LEGGENDARIO))$coefficients[2,4] + (summary(model\_LEGGENDARIO))$coefficients[3,4]))), format = "markdown")

knitr::kable(data.frame(Rarità = c("EPICO", "EPICO"), Variabile = c("Close ETH\_L1", "Volume ETH"),

`Coefficiente` = c((summary(model\_EPICO))$coefficients[2,1], (summary(model\_EPICO))$coefficients[3,1]),

`p-value` = c(((summary(model\_EPICO))$coefficients[2,4] + (summary(model\_EPICO))$coefficients[3,4]),

((summary(model\_EPICO))$coefficients[2,4] + (summary(model\_EPICO))$coefficients[3,4]))), format = "markdown")

knitr::kable(data.frame(Rarità = c("RARO", "RARO"), Variabile = c("Close ETH\_L1", "Volume ETH"),

`Coefficiente` = c((summary(model\_RARO))$coefficients[2,1], (summary(model\_RARO))$coefficients[3,1]),

`p-value` = c(((summary(model\_RARO))$coefficients[2,4] + (summary(model\_RARO))$coefficients[3,4]),

((summary(model\_RARO))$coefficients[2,4] + (summary(model\_RARO))$coefficients[3,4]))), format = "markdown")

knitr::kable(data.frame(Rarità = c("NON COMUNE", "NON COMUNE"), Variabile = c("Close ETH\_L1", "Volume ETH"),

`Coefficiente` = c((summary(model\_NON\_COMUNE))$coefficients[2,1], (summary(model\_NON\_COMUNE))$coefficients[3,1]),

`p-value` = c(((summary(model\_NON\_COMUNE))$coefficients[2,4] + (summary(model\_NON\_COMUNE))$coefficients[3,4]),

((summary(model\_NON\_COMUNE))$coefficients[2,4] + (summary(model\_NON\_COMUNE))$coefficients[3,4]))), format = "markdown")

knitr::kable(data.frame(Rarità = c("COMUNE", "COMUNE"), Variabile = c("Close ETH\_L1", "Volume ETH"),

`Coefficiente` = c((summary(model\_COMUNE))$coefficients[2,1], (summary(model\_COMUNE))$coefficients[3,1]),

`p-value` = c(((summary(model\_COMUNE))$coefficients[2,4] + (summary(model\_COMUNE))$coefficients[3,4]),

((summary(model\_COMUNE))$coefficients[2,4] + (summary(model\_COMUNE))$coefficients[3,4]))), format = "markdown")

##### ANALISI CLASSICA ####

serie\_BAYC <- na.omit(BAYC)[, c("Date","Price (k)","Close ETH ($)","Close ETH\_L1 ($)","Volume ETH (B)","Volume ETH\_L1 (B)")]

serie\_BAYC <- serie\_BAYC %>%

group\_by(Date) %>%

summarise\_all(list(mean))

View(serie\_BAYC)

yt=ts(serie\_BAYC$`Price (k)`, frequency = 365, start = c(2021,5,2))

plot(yt, main = "Serie Storica dei Prezzi degli NFT", xlab="Data",ylab="Prezzo (k)")

modCla=tslm(yt ~ trend)

summary(modCla)

modCla2=tslm(yt ~ trend + I(trend^2))

summary(modCla2)

modCla3=tslm(yt ~ trend + I(trend^2) + I(trend^3))

summary(modCla3)

knitr::kable(data.frame(Modello = c("Trend", "I(Trend^2)", "I(Trend^3)"),

`Adjusted R-squared` = c(summary(modCla)$adj.r.squared, summary(modCla2)$adj.r.squared, summary(modCla3)$adj.r.squared),

`p-value` = c(((summary(modCla))$coefficients[2,4]),

((summary(modCla2))$coefficients[2,4]),

(summary(modCla3))$coefficients[2,4])),format = "markdown")

knitr::kable(data.frame(Variabile = c("Trend", "I(Trend^2)", "I(Trend^3)"),

`Coefficiente` = c(summary(modCla3)$coefficients[2, 1],

summary(modCla3)$coefficients[3, 1],

summary(modCla3)$coefficients[4, 1]),

`p-value` = c(summary(modCla3)$coefficients[2, 4],

summary(modCla3)$coefficients[3, 4],

summary(modCla3)$coefficients[4, 4])), format = "markdown")

plot(yt, main = "Serie Storica dei Prezzi degli NFT (Trend^3)", xlab="Data",ylab="Prezzo (k)")

lines(modCla3$fitted.values, col="blue")

##### ANALISI MODERNA ####

Acf(yt,50, main="Correlogramma globale di yt")

Dyt=diff(yt)

plot(Dyt)

Acf(Dyt,50, main="Correlogramma globale di Dyt")

par(mfrow=c(2,1))

Acf(Dyt)

Pacf(Dyt)

Acf(Dyt,50, main="Correlogramma globale di Dyt")

Pacf(Dyt,50, main="Correlogramma parziale di Dyt")

modARIMA=Arima(yt, c(0,1,1))

summary(modARIMA)

coefSTAT=modARIMA$coef/sqrt(diag(modARIMA$var.coef))

coefSTAT

T=length(yt)

r=length(modARIMA$coef)+1

pval=2\*(1-pt(abs(coefSTAT), T-r))

pval

resARIMA=modARIMA$residuals

Acf(resARIMA,50, main="Correlogramma globale resARIMA")

Pacf(resARIMA,50, main="Correlogramma parziale resARIMA")

par(mfrow=c(1,1))

plot(yt, main = "Serie Storica dei Prezzi degli NFT (ARIMA)", xlab="Data",ylab="Prezzo (k)")

lines(modARIMA$fitted, col="red")

summary(modARIMA)

sqrt(mean((yt-modCla3$fitted.values)^2))

mean(abs(yt-modCla3$fitted.values))

plot(yt, main="Serie Storica dei Prezzi degli NFT (Trend^3 + ARIMA)", xlab="Data",ylab="Prezzo (k)",lwd=2)

lines(modCla3$fitted.values, col="blue", lwd=2)

lines(modARIMA$fitted, col="red", lwd=2)

##### MODELLO ARIMAX ####

modARIMAX=Arima(yt, c(0,1,1), xreg= serie\_BAYC$`Close ETH ($)`)

summary(modARIMAX)

modARIMAX=Arima(yt, c(0,1,1), xreg= serie\_BAYC$`Close ETH\_L1 ($)`)

summary(modARIMAX)

modARIMAX=Arima(yt, c(0,1,1), xreg= serie\_BAYC$`Volume ETH (B)`)

summary(modARIMAX)

modARIMAX=Arima(yt, c(0,1,1), xreg= serie\_BAYC$`Volume ETH\_L1 (B)`)

summary(modARIMAX)

modARIMAX=Arima(yt, c(0,1,1), xreg= cbind(serie\_BAYC$`Close ETH ($)`,serie\_BAYC$`Volume ETH (B)`))

summary(modARIMAX)

modARIMAX=Arima(yt, c(0,1,1), xreg= cbind(serie\_BAYC$`Close ETH\_L1 ($)`,serie\_BAYC$`Volume ETH (B)`))

summary(modARIMAX)

modARIMAX=Arima(yt, c(0,1,1), xreg= cbind(serie\_BAYC$`Close ETH ($)`,serie\_BAYC$`Volume ETH\_L1 (B)`))

summary(modARIMAX)

modARIMAX=Arima(yt, c(0,1,1), xreg= cbind(serie\_BAYC$`Close ETH\_L1 ($)`,serie\_BAYC$`Volume ETH\_L1 (B)`))

summary(modARIMAX)

xreg=cbind(serie\_BAYC$`Close ETH ($)`,serie\_BAYC$`Volume ETH (B)`)

modARIMAX=Arima(yt, c(0,1,1), xreg=xreg)

summary(modARIMAX)

coefSTATX=modARIMAX$coef/sqrt(diag(modARIMAX$var.coef))

coefSTATX

T=length(yt)

r=length(modARIMAX$coef)+1

pval=2\*(1-pt(abs(coefSTATX), T-r))

pval

resARIMAX=modARIMAX$residuals

par(mfrow=c(2,1))

Acf(resARIMAX,50, main="Correlogramma globale resARIMAX")

Pacf(resARIMAX,50, main="Correlogramma parziale resARIMAX")

par(mfrow=c(1,1))

plot(yt, main = "Serie Storica dei Prezzi degli NFT (ARIMAX)", xlab="Data",ylab="Prezzo (k)")

lines(modARIMAX$fitted, col="green")

summary(modARIMA)

summary(modARIMAX)

plot(yt, main="Serie Storica dei Prezzi degli NFT (ARIMA + ARIMAX)", xlab="Data",ylab="Prezzo (k)",lwd=2)

lines(modARIMA$fitted, col="red", lwd=2)

lines(modARIMAX$fitted, col="green", lwd=2)

##### PREVISIONI ####

f\_modARIMA=matrix(,117,1)

for(i in 706:822) {

modARIMA=Arima(yt[1:i],c(0,1,1))

f\_modARIMA[i-705]=forecast(modARIMA,1)$mean}

f\_modARIMA

plot(yt[707:823],type="l",main="Previsione dei Prezzi degli NFT (ARIMA)",xlab="Previsione (n)",ylab="Prezzo NFT ($)",lwd=2)

lines(f\_modARIMA,type="l",col="red",lwd=2)

sqrt(mean((yt[707:823]-f\_modARIMA)^2))

mean(abs((yt[707:823]-f\_modARIMA)))

xreg= cbind(serie\_BAYC$`Close ETH\_L1 ($)`,serie\_BAYC$`Volume ETH\_L1 (B)`)

f\_modARIMAX=matrix(,117,1)

for(i in 706:822) {

modARIMAX=Arima(yt[1:i],c(0,1,1),xreg=xreg[1:i])

f\_modARIMAX[i-705]=forecast(modARIMAX,1,xreg=xreg[1+i])$mean}

f\_modARIMAX

plot(yt[707:823],type="l",main="Previsione dei Prezzi degli NFT (ARIMAX)",xlab="Previsione (n)",ylab="Prezzo NFT ($)",lwd=2)

lines(f\_modARIMAX,type="l",col="green",lwd=2)

sqrt(mean((yt[707:823]-f\_modARIMAX)^2))

mean(abs((yt[707:823]-f\_modARIMAX)))

plot(yt[707:823],type="l",main="Previsione dei Prezzi degli NFT (ARIMA + ARIMAX)",xlab="Previsione (n)",ylab="Prezzo NFT ($)",lwd=2)

lines(f\_modARIMA,type="l",col="red",lwd=2)

lines(f\_modARIMAX,type="l",col="green",lwd=2)

# BIBLIOGRAFIA

*Blockchain - Academy* *NonFungible.com*. Disponibile al sito: <https://nonfungible.com/academy/blockchain>

*NFT - Academy* *NonFungible.com*. Disponibile al sito: <https://nonfungible.com/academy/nft>

*Bored ape yacht club* (2023) *Wikipedia*. Disponibile al sito: <https://it.wikipedia.org/wiki/Bored_Ape_Yacht_Club>

*Cos’è il bored ape yacht club? Tutto quello che devi sapere sugli nft del momento!* *Young Platform*. Disponibile al sito: <https://youngplatform.com/blog/news/cos-e-bored-ape-yacht-club-guida-nft/>

*Raccolte NFT elencate per volume di vendita* *CoinMarketCap*. Disponibile al sito: <https://coinmarketcap.com/it/nft/collections/>

*NFT trades | Etherscan*. Disponibile al sito: <https://etherscan.io/nft-trades?contractAddress=0xbc4ca0eda7647a8ab7c2061c2e118a18a936f13d>

*Bored ape yacht club NFT rarity ranking* *Rarity Sniper*. Disponibile al sito: <https://raritysniper.com/bored-ape-yacht-club>

*Storico dei Prezzi e Dati Storici per Ethereum EUR (ETH-EUR)* *Yahoo! Finance*. Disponibile al sito: <https://it.finance.yahoo.com/quote/ETH-EUR/history?p=ETH-EUR&guccounter=1>

James, G. (2021) *An introduction to statistical learning: With applications in R*. New York: Springer

De Luca, G. (2023) *Analisi e previsione delle serie storiche economiche e finanziarie*