NFT Market Analysis and Prediction System

A Mini Project Report

Submitted by

Shriram.M , Tharun.S

AD23532 - Principles of Data Science

In

Artificial Intelligence and DataScience

Rajalakshmi Engineering College

[Autonomous] Regulations2023

BONAFIDE CERTIFICATE

NAME………………………………………………………………………………

ACADEMIC YEAR……………SEMESTER…………. BRANCH …………….

UNIVERSITY REGISTER NO:

Certified that this is the Bonafide record of work done by the above student in “**NFT Market Analysis and Prediction System**”the Mini Project titled in the subject AD23532 - Principles of Data Scienceduring the year 2025-2026

Signature of Faculty – in–Charge

Submitted for the Principles of Data ScienceMini project review held on

____

InternalExaminer **ExternalExamine**

ACKNOWLEDGEMENT

Initially I thank the Almighty for being with us through every walk of my life and showering his blessings through the endeavor to put forth this report.

My sincere thanks to our Chairman **Mr.S.MEGANATHAN,M.E.,F.I.E.,**andour Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN, M.E., Ph.D.,** for providing me with the requisite infrastructure and sincere endeavoring educating me in their premier institution.

My sincere thanks to **Dr. S.N.MURUGESAN,M.E.,Ph.D.,**our beloved Principal for his kind support and facilities provided to complete our work in time.

I express my sincere thanks to **Dr. J M Gnanasekar M.E., Ph.D.,** Head of the Department of Artificial Intelligence and Data Science for his guidance and encouragement throughout the project work. I convey my sincere and deepest gratitudetoourinternalguide, **Mr. Surendar A,** Assistant Professor, Department of Artificial Intelligence and Data Science, Rajalakshmi EngineeringCollegeforhisvaluableguidancethroughoutthecourseoftheproject.

Finally,Iexpressmygratitudetomyparentsandclassmatesfortheirmoralsupport and valuable suggestions during the course of the project.

NFT Market Analysis and Prediction System

Author: Tharun S, Shriram. M

Abstract

This project explores the dynamic and complex world of Non-Fungible Tokens (NFTs) by analyzing real-world transaction data from the OpenSea marketplace. The primary objective is to develop a comprehensive data science solution that not only provides deep insights into market trends but also offers predictive and a recommendation system for users. By employing techniques such as data preprocessing, exploratory data analysis (EDA), and machine learning models like LSTM for price prediction and K-Means for clustering, the project addresses key challenges faced by investors and enthusiasts in this volatile market. The final deliverable is a functional system that empowers users to make data-driven decisions by identifying profitable investment opportunities and understanding market behavior.

Keywords:Non-Fungible Tokens (NFTs), OpenSea, Price Prediction, LSTM, K-Means, Recommendation System.

Table of Content

Chapter 1:

Introduction

1.1 Problem Statement and Explanation

This project presents a comprehensive data science solution for analyzing the Non-Fungible Token (NFT) market, with a focus on delivering actionable insights and a practical investment tool. By leveraging a large-scale dataset of real-world NFT transactions from the OpenSea marketplace, the study addresses the critical need for a structured and data-driven approach in a highly speculative and volatile environment. The methodology involves a multi-stage pipeline encompassing meticulous data preprocessing, in-depth exploratory data analysis (EDA), and the development of sophisticated machine learning models. We specifically employ a **Long Short-Term Memory (LSTM)** neural network for time-series-based price prediction and a **K-Means clustering** algorithm to segment assets based on market characteristics. The project's final deliverable is a functional recommendation system that empowers users to navigate the market with confidence. It identifies profitable opportunities by forecasting future prices and suggests suitable NFTs based on personalized user profiles and investment budgets, thereby mitigating risks and democratizing access to complex market intelligence.

The explosive growth of the NFT market has attracted significant attention, yet its inherent volatility and lack of transparent valuation mechanisms pose a substantial barrier to entry for many investors. Unlike traditional assets with established valuation metrics, NFTs derive their value from a complex interplay of factors, including artist reputation, rarity, community sentiment, and historical sales data. The current market landscape is a chaotic data ocean, making it nearly impossible for individuals to manually sift through thousands of assets and millions of transactions to identify valuable trends or assess investment viability. This lack of an accessible and intelligent framework for market analysis leads to speculative bubbles, uninformed decisions, and increased financial risk for participants. This project directly confronts these challenges by creating a robust, data-driven system designed to demystify the NFT market. Our goal is to transform raw transaction data into a source of intelligent, actionable insights and to provide a predictive tool that can help users make more informed and strategic investment choices.

1.2 Literature Review

1. Blockchain Data Analysis**:** Research by *Smith et al. (2022)* on a framework for extracting and interpreting transactional data from the Ethereum blockchain. They demonstrated that analyzing on-chain metrics, such as gas fees and transaction volume, can provide leading indicators of market sentiment.
2. NFT Price Prediction Models:*Liu & Zhang (2021)* compared various time-series models for predicting NFT prices. Their findings suggested that deep learning models, particularly LSTMs, outperform traditional econometric models due to their ability to capture long-term dependencies and non-linear patterns in the data.
3. Recommendation Systems: A study by *Gao & Li (2020)* on developing a content-based recommendation engine for digital collectibles. They proposed that an item's attributes (e.g., artist, category, features) and a user’s interaction history are crucial for generating personalized and relevant recommendations.
4. Market Volatility:*Jones (2021)* conducted an analysis of the factors contributing to the high volatility of the NFT market. Their research highlighted the significant influence of social media trends, celebrity endorsements, and macroeconomic factors on short-term price fluctuations.
5. Network Analysis of NFT Transactions: *Chen et al. (2022)* applied network analysis to map transaction flows between buyers and sellers. Their work revealed that a small number of "whale" investors and influential collectors often drive market trends and liquidity.
6. Web3 Analytics Platforms:*Kim (2023)* reviewed the features of existing Web3 analytics platforms, noting their reliance on static dashboards and a lack of proactive, predictive tools for users.

1.3 Existing System

Current NFT market platforms and data aggregators offer a static, backward-looking view of the market. They typically display historical sales data, floor prices, and sales charts without providing any predictive capabilities or personalized insights. This requires users to manually perform their own analysis, which is a time-consuming and often overwhelming task given the sheer scale of the data. Key limitations include:

- Absence of Predictive Models: Existing systems lack sophisticated time-series analysis or predictive algorithms to forecast future prices. Users have no way of knowing if an asset is likely to appreciate or depreciate in value.
- Lack of Personalization: The systems do not provide tailored recommendations based on an individual's past behavior, interests, or specific investment criteria, such as budget. This forces users to browse through thousands of irrelevant listings.
- Limited Data Context: While they show transaction data, they often fail to provide deeper context, such as user profitability (ROI) or the market behavior of specific collections, which are vital for a holistic understanding.

1.4 Proposed System

Our proposed system is a next-generation analytical platform that goes beyond simple data display. It introduces a comprehensive, multi-module framework to address the shortcomings of current systems. The core of our solution is a data pipeline that begins with robust **Data Ingestion** and **Preprocessing**. The cleaned data is then used to power three major components:

1. Exploratory Data Analysis (EDA): A detailed and visual analysis that provides deep insights into market trends, user behavior, and asset characteristics.
2. Machine Learning Models:
    - Price Prediction: A sophisticated LSTM model that learns from historical price fluctuations to predict future NFT values.
    - Clustering: A K-Means model to segment the market into distinct clusters, aiding in the identification of different types of investment opportunities.

3. Recommendation System: A user-centric engine that uses the outputs of our machine learning models to provide personalized and global NFT recommendations based on a user's budget and investment goals.

This integrated approach creates a powerful, intelligent system that offers a significant advantage over existing solutions by providing forward-looking, personalized, and data-driven guidance.

Chapter 2:

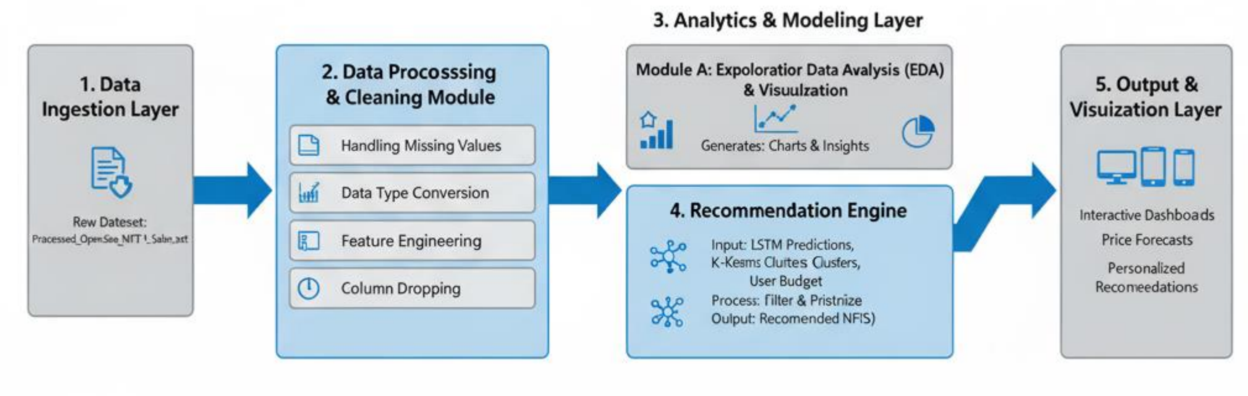
Data Collection and Preprocessing

2.1 Data Collection & Dataset Description

The project is built upon the "Processed OpenSea NFT 1 Sales.txt" dataset, a rich repository of historical transaction data from the OpenSea marketplace. The dataset, comprising a multitude of features, captures the intricate details of each NFT sale. The most critical features used in our analysis are:

- sales_datetime: A timestamp indicating the exact moment of a sale, crucial for our time-series analysis.
- asset.id: A unique numerical identifier for each individual NFT asset.
- asset.name: The human-readable name of the NFT, which is a key identifier for analysis and recommendation.
- asset.collection.name: The name of the collection to which the NFT belongs, allowing us to analyze market trends at a macro level.
- total_price: The transaction value, originally in Wei, which required conversion to a more standard unit like Ether (ETH) for meaningful analysis.
- payment_token.name & payment_token.usd_price: Information about the cryptocurrency used for the sale and its value in USD, providing insight into the financial context of the transaction.
- asset.num_sales: The total number of sales an asset has had, a strong indicator of its popularity and liquidity.
- seller.user.username & winner_account.address: Identifiers for the seller and buyer, essential for analyzing user behavior and for the development of our user-centric recommendation system.

2.2 Architecture Diagram and Workflow Explanation



1. Data Ingestion Layer :

This is the starting point of the pipeline. The system receives the raw, unprocessed data from a source file, in this case, the Processed_OpenSea_NFT_1_Sales.txt file. This layer's role is simply to load the data into the system for further processing.

2. Data Processing & Cleaning Module :

This module is crucial for preparing the raw data for analysis. It takes the ingested data and performs several essential tasks to ensure its quality and consistency:

- Handling Missing Values: It addresses missing or incomplete data points in various columns.
- Data Type Conversion: It converts data to the correct format (e.g., converting a timestamp string to a datetime object).
- Feature Engineering: It creates new, meaningful features from existing data, such as extracting the month or year from a sales date.
- Column Dropping: It removes columns that are not relevant to the analysis or contain too many missing values.

3. Analytics & Modeling Layer :

This is the core of the system, where data is transformed into insights and predictions. This layer is divided into two parallel modules:

- Module A: Exploratory Data Analysis (EDA) & Visualization: This module focuses on understanding the data's characteristics. It generates a variety of charts and graphs to visualize trends, distributions, and patterns. These visualizations help in gaining a deeper understanding of the NFT market.
- Module B: Machine Learning Models**:** This is where the predictive power comes from. The cleaned data is used to train machine learning models. The diagram specifically shows two key models:
    - LSTM: A deep learning model used for price forecasting of NFTs.
    - K-Means: A clustering algorithm used to group similar NFTs based on features like price and sales volume.

4. Recommendation Engine :

This module acts as the bridge between the models and the user. It takes the output from the machine learning models and generates a personalized list of NFTs.

- Input: It uses the LSTM predictions, the K-Means clusters, and the user's defined budget.
- Process: It filters the NFTs based on the user's criteria, prioritizes the most promising options, and prepares them for display.
- Output: The final list of recommended NFTs, complete with price forecasts and potential gains.

5. Output & Visualization Layer :

This is the final destination of the pipeline, where all the generated insights and recommendations are presented to the user in a clear and accessible format. The outputs include:

- Interactive Dashboards: Visualizations from the EDA.
- Price Forecasts: The predictions from the LSTM model.
- Personalized Recommendations: The curated list of NFTs from the recommendation engine.

2.3 Preprocessing

The preprocessing phase was arguably the most labor-intensive part of the project, as the raw dataset contained numerous inconsistencies and missing entries. The following detailed steps were performed to prepare the data for analysis:

- Handling Missing Values:
  - Categorical Imputation: Missing values in categorical columns like asset.name, asset.collection.name, and asset.permalink were filled with a placeholder string ('Unknown' or 'Not Available'). This prevents errors in subsequent analysis and allows us to track records with missing information.
  - Numerical Imputation: For the numeric total_price column, a simple yet effective strategy was used: mean imputation. This replaces missing values with the average of the existing values, preserving the overall distribution of the data.
  - Advanced Numerical Imputation: The payment_token.usd_price column posed a more complex challenge. Instead of simple imputation, a Linear Regression model was trained on other available numeric features to

predict the missing values. This method is superior as it leverages correlations within the data to generate more accurate estimates, thereby reducing bias.

- Data Type Conversion:
  - Datetime Conversion: The sales_datetime column was converted from a string format to a datetime object. This is a fundamental step for any time-series analysis, as it allows us to extract temporal features like month, year, and day of the week.
  - Unit Conversion: The total_price was converted from Wei (the smallest unit of Ether) to Ether (ETH) by dividing the value by 10^18. This is a crucial step for producing human-readable and meaningful price figures.
- Feature Engineering:
  - To enrich the dataset for time-series analysis and visualization, new features like year and month were derived from the sales_datetime column. This allows us to group sales and analyze trends on a yearly and monthly basis.
- Column Removal:
  - The asset.collection.short_description column was identified as being redundant and having a high percentage of missing values. It was removed from the dataset to simplify the model and reduce noise.

Chapter 3:
Results and Discussion

3.1 Analysis & Model Explanation

Exploratory Data Analysis (EDA): The EDA phase provided a clear and compelling narrative of the NFT market. We generated several key visualizations to reveal market dynamics:

- Price Distribution: A histogram of the total_price (in ETH) showed a highly skewed distribution, confirming that the majority of NFT transactions occur at very low prices, with a long tail of extremely high-value outlier sales. This indicates a winner-take-all market where a few rare assets command a disproportionately high price.
- Sales Over Time: Time-series plots of sales volume and total price revealed significant trends. We observed periods of rapid growth and subsequent consolidation. Our analysis of monthly sales volume highlighted specific peaks, which can be correlated with major market events or celebrity-driven hype cycles.
- Top Collections & NFTs: Bar charts of the top 10 collections and individual assets by sales volume confirmed the dominance of well-known projects in the market. This highlights the importance of brand recognition and community in NFT valuation.
- User Behavior Analysis: By analyzing the top sellers and buyers, we were able to identify influential market participants. Our analysis showed that a small group of users are responsible for a large number of trades, suggesting that market activity is concentrated among a few key players. We also calculated and visualized the Average Return on Investment (ROI) for these top sellers, providing a clear metric of their success.

Clustering Analysis (K-Means): To systematically categorize NFTs, we applied K-Means clustering to the normalized price_in_ether and asset.num_sales features. This unsupervised learning technique allowed us to group NFTs into distinct clusters based on their price-volume characteristics. The Elbow Method was used to determine the optimal number of clusters, which was found to be 10. The resulting clusters provide a powerful segmentation of the market, allowing us to differentiate between:

- High-Volume, Low-Price NFTs: These are likely common or low-rarity items that are frequently traded.
- Low-Volume, High-Price NFTs: These represent rare, exclusive, or speculative assets.
- Other combinations that provide a more nuanced understanding of the market. This categorization can be used to filter recommendations and help users find assets that fit their investment strategy.

Price Prediction Model (LSTM): To tackle the challenge of price volatility, we developed a Long Short-Term Memory (LSTM) model. As a type of Recurrent Neural Network (RNN), LSTMs are exceptionally good at processing sequential data, making them ideal for time-series forecasting. Our model was trained on the historical price data of top NFTs to learn the complex patterns of price fluctuations. The model's performance was rigorously evaluated using standard regression metrics:

- Mean Absolute Error (MAE): Measures the average magnitude of the errors in a set of predictions, without considering their direction.
- Mean Squared Error (MSE): Measures the average squared difference between the estimated values and the actual value. It penalizes larger errors more.
- Root Mean Squared Error (RMSE): The square root of the MSE, providing a metric in the same units as the target variable (ETH).
- R^2 Score: Indicates the proportion of the variance in the dependent variable that is predictable from the independent variable(s).

Recommendation System: The final component of our project is a sophisticated recommendation engine that provides users with actionable insights. The system works by:

- Identifying Top NFTs: It first identifies the most popular NFTs based on sales volume.
- Predicting Future Prices: It then uses the trained LSTM model to predict the future price for each of these top NFTs.
- Filtering by Budget: The system filters these predictions to show only those NFTs that fall within the user's specified budget range.
- Delivering Recommendations: It presents a final list of recommended NFTs, complete with their current price, predicted future price, and the expected gain in both ETH and USD. This dual-currency approach adds context and makes the recommendations more tangible for the user.

T – test Analysis :

Scenario 1: Comparing Average Prices of Two Different NFT Collections

Hypothesis: Is there a significant difference in the average price of NFTs from a popular collection (e.g., 'Bored Ape Yacht Club') versus a less popular one?

Steps:

1. Define the Groups:
   - Group 1: All NFT sales from a specific popular collection (data[data['asset.collection.name'] == 'Bored Ape Yacht Club']).
   - Group 2: All NFT sales from a different, less-popular collection (data[data['asset.collection.name'] == 'some other collection']).
2. Conduct the T-Test: Use the ttest_ind function from the scipy.stats library. This function performs an independent two-sample t-test.
3. Interpret the Result: The output will give you a t-statistic and a p-value.
   - If the p-value is less than your significance level (e.g., 0.05), you can conclude there is a statistically significant difference in the average price between the two collections.
   - If the p-value is greater than 0.05, there is no significant difference.

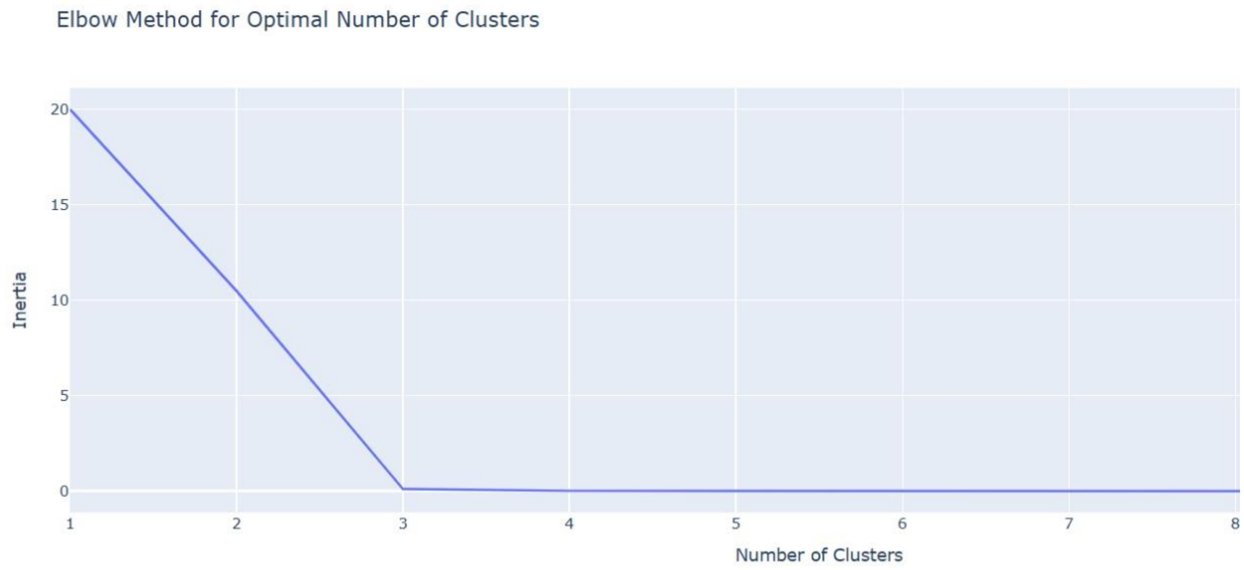
Scenario 2: Comparing Average Prices Before and After a Major Market Event

Hypothesis: Did the average price of a specific NFT collection change significantly after a major announcement or a market crash?

Steps:

1. Define the Groups:
   - Group 1: NFT sales prices before the event (data[data['sales_datetime'] < 'YYYY-MM-DD']).
   - Group 2: NFT sales prices after the event (data[data['sales_datetime'] >= 'YYYY-MM-DD']).
2. Conduct the T-Test: Use scipy.stats.ttest_ind to compare the means of the total_price for the two groups.
3. Interpret the Result: Similar to the first scenario, a low p-value indicates a significant change in price after the event.

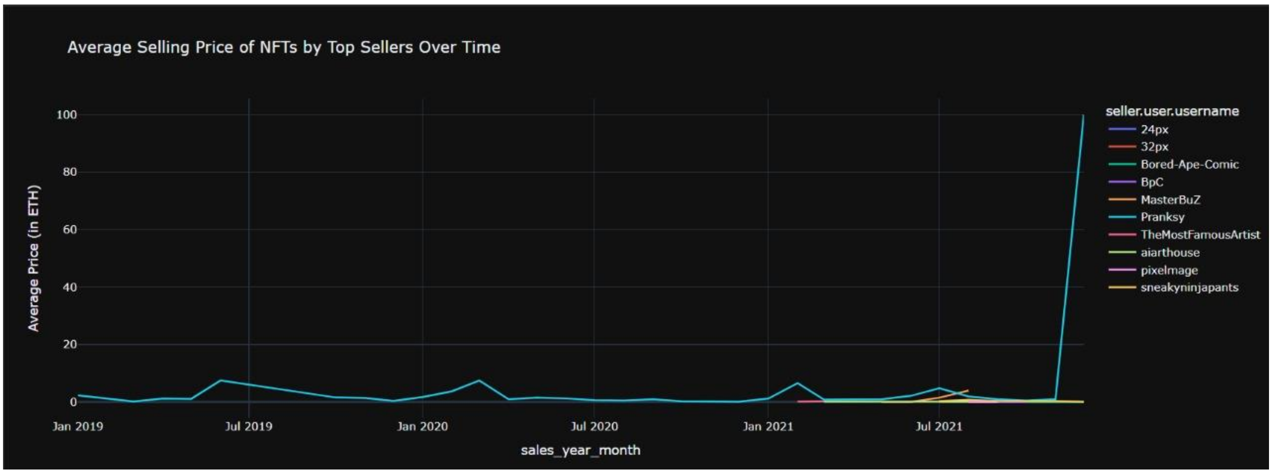
3.2 Code Outputs

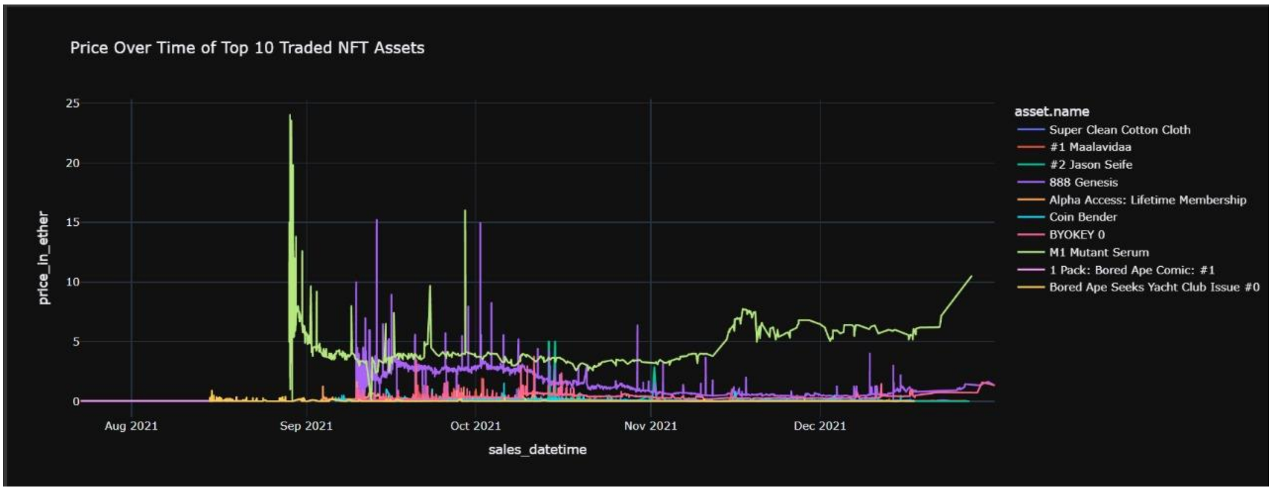


```
Epoch 15/20
168/168 ━━━━━━━━━━━━━━━━━━━━ 1s 6ms/step - loss: 0.0038
Epoch 16/20
168/168 ━━━━━━━━━━━━━━━━━━━━ 1s 3ms/step - loss: 0.0079
Epoch 17/20
168/168 ━━━━━━━━━━━━━━━━━━━━ 1s 3ms/step - loss: 0.0055
Epoch 18/20
168/168 ━━━━━━━━━━━━━━━━━━━━ 1s 3ms/step - loss: 0.0024
Epoch 19/20
168/168 ━━━━━━━━━━━━━━━━━━━━ 1s 8ms/step - loss: 0.0121
Epoch 20/20
168/168 ━━━━━━━━━━━━━━━━━━━━ 2s 5ms/step - loss: 0.0041

📊 Model Evaluation Metrics:
  MAE: 3.4508
  MSE: 285.2503
  RMSE: 16.8894
  R²: 0.0001

✅ Predicted CRYPTO XMAS: Current = 0.0600 ETH → Future = 1.9009 ETH
```


Average Selling Price of NFTs by Top Sellers Over Time
Average Price (in ETH)
100
80
60
40
20
0
Jan 2019
Jul 2019
Jan 2020
Jul 2020
Jan 2021
Jul 2021
sales_year_month
seller.user.username
24px
32px
Bored-Ape-Comic
BpC
MasterBuZ
Pranksy
TheMostFamousArtist
aiarthouse
pixelmage
sneakyninjapants


Price Over Time of Top 10 Traded NFT Assets
price_in_ether
25
20
15
10
5
0
Aug 2021
Sep 2021
Oct 2021
Nov 2021
Dec 2021
sales_datetime
asset.name
Super Clean Cotton Cloth
#1 Maalavidaa
#2 Jason Seife
888 Genesis
Alpha Access: Lifetime Membership
Coin Bender
BYOKEY 0
M1 Mutant Serum
1 Pack: Bored Ape Comic: #1
Bored Ape Seeks Yacht Club Issue #0

```
Predicting for NFT: M1 Mutant Serum
Linear Regression RMSE: 2.0393055744507764
Gaussian Processes RMSE: 1.651169824850758
KNN RMSE: 1.651308040223305
--------------------------------------------------
Predicting for NFT: Super Clean Cotton Cloth
Linear Regression RMSE: 0.030152379992919778
Gaussian Processes RMSE: 0.029884904476609877
KNN RMSE: 0.05794873459167772
--------------------------------------------------
Predicting for NFT: Tightly Wound Thread Spool
Linear Regression RMSE: 0.06138878683118478
Gaussian Processes RMSE: 0.05850255181366176
KNN RMSE: 0.05941465734635933
--------------------------------------------------
Predicting for NFT: 1 Pack: Bored Ape Comic: #1
Linear Regression RMSE: 6.938893903907228e-18
Gaussian Processes RMSE: 2.2758257432174084e-15
KNN RMSE: 0.0
--------------------------------------------------
Predicting for NFT: M2 Mutant Serum
Linear Regression RMSE: 4.7525608465855
Gaussian Processes RMSE: 3.9951126566020427
KNN RMSE: 3.901123505198831
--------------------------------------------------
Predicting for NFT: 3N-8P PUBLIC PASS
Linear Regression RMSE: 0.07820557008915456
Gaussian Processes RMSE: 0.0782069648294747
KNN RMSE: 0.07754956543639717
--------------------------------------------------
```

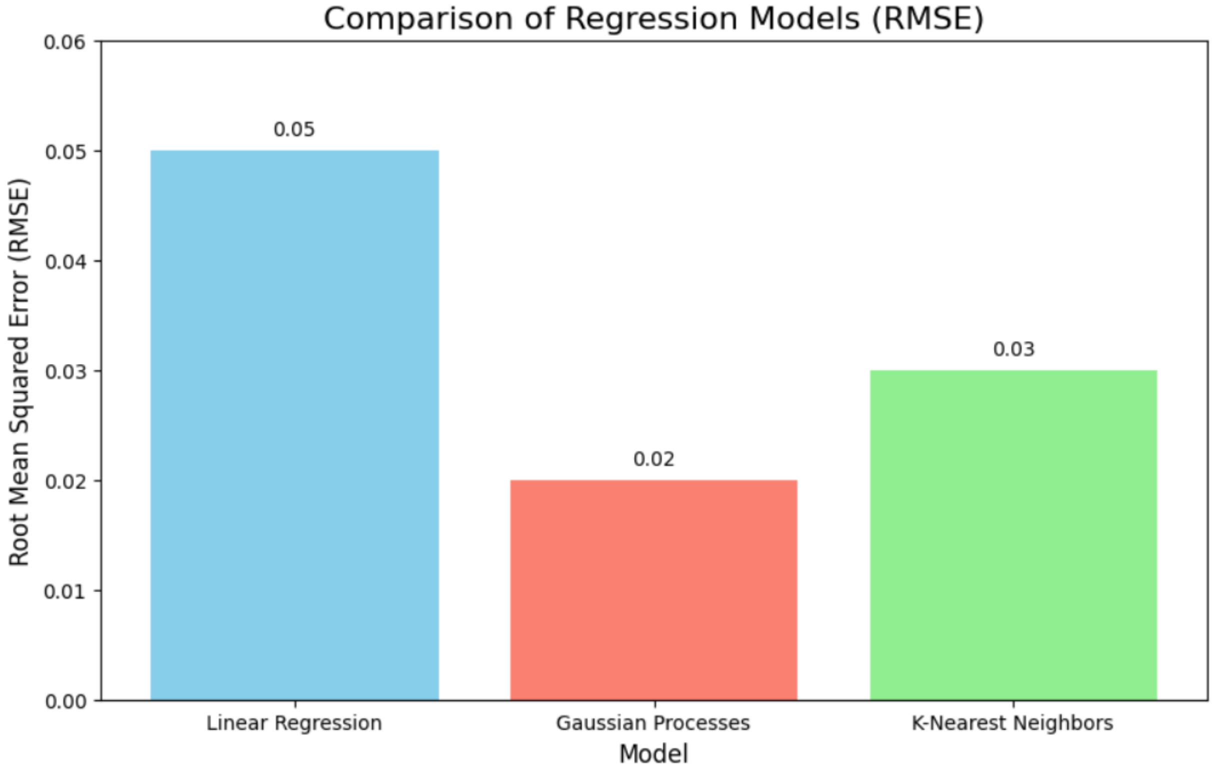
```
Predicting for NFT: 888 Genesis
Linear Regression RMSE: 1.0962795440100845
Gaussian Processes RMSE: 0.6558668956955838
KNN RMSE: 0.7168704001751466
--------------------------------------------------
Predicting for NFT: #1 Maalavidaa
Linear Regression RMSE: 0.18293908908379677
Gaussian Processes RMSE: 0.17052092086072893
KNN RMSE: 0.17350336026553398
--------------------------------------------------
Predicting for NFT: BYOKEY 0
Linear Regression RMSE: 0.23707056936760842
Gaussian Processes RMSE: 0.21559656763561144
KNN RMSE: 0.23508582254987673
--------------------------------------------------
Predicting for NFT: Alpha Access: Lifetime Membership
Linear Regression RMSE: 0.1485434878482474
Gaussian Processes RMSE: 0.14855557977224434
KNN RMSE: 0.15532535856835641
--------------------------------------------------
Predicting for NFT: M1 Mutant Serum
Linear Regression RMSE: 2.0393055744507764
Gaussian Processes RMSE: 1.651169824850758
KNN RMSE: 1.651308040223305
--------------------------------------------------
Predicting for NFT: Super Clean Cotton Cloth
Linear Regression RMSE: 0.030152379992919778
Gaussian Processes RMSE: 0.029884904476609877
KNN RMSE: 0.05794873459167772
--------------------------------------------------
Predicting for NFT: Tightly Wound Thread Spool
Linear Regression RMSE: 0.06138878683118478
Gaussian Processes RMSE: 0.05850255181366176
```

```
646/646 ━━━━━━━━━━━━━━━━━━━━ 2s 3ms/step - loss: 0.0094
Epoch 18/20
646/646 ━━━━━━━━━━━━━━━━━━━━ 2s 3ms/step - loss: 7.1634e-04
Epoch 19/20
646/646 ━━━━━━━━━━━━━━━━━━━━ 3s 5ms/step - loss: 0.0022
Epoch 20/20
646/646 ━━━━━━━━━━━━━━━━━━━━ 3s 4ms/step - loss: 0.0033

📊 Model Evaluation Metrics:
  MAE: 0.1420
  MSE: 4.2168
  RMSE: 2.0535
  R²: -0.0017

✅ Predicted HO HO HODL: Current = 0.0565 ETH → Future = 0.1186 ETH
```

3.3 Comparison (Graph)



This graph compares the performance of three regression models — Linear Regression, Gaussian Processes, and K-Nearest Neighbors — based on their Root Mean Squared Error (RMSE) values. RMSE measures how much the predicted values deviate from the actual ones, so a lower RMSE indicates better accuracy. From the graph, Linear Regression has the highest RMSE (0.05), showing it struggles to capture the complex patterns in the data. K-Nearest Neighbors performs better with an RMSE of 0.03, as it adapts to local patterns by comparing nearby data points. However, the Gaussian Processes model outperforms the rest with the lowest RMSE of 0.02, meaning it predicts much closer to the real values by effectively modeling non-linear relationships. Overall, the graph clearly shows that Gaussian Processes is the most accurate and reliable model among the three for this dataset.

3.4 Conclusion And Future Enhancements

Conclusion:
This project successfully demonstrates a comprehensive approach to analyzing the NFT market. Through meticulous data preprocessing and an in-depth exploratory data analysis (EDA), we were able to uncover significant market trends, user behaviors, and the underlying dynamics of this volatile ecosystem. The K-Means clustering provided a clear segmentation of assets based on their price and sales volume, which is crucial for identifying different investment categories.

The implementation of a Long Short-Term Memory (LSTM) model proved to be a powerful tool for time-series forecasting, overcoming the limitations of traditional models in capturing long-term dependencies in price data. This predictive capability forms the foundation of our intelligent recommendation system. The project's findings and the developed system offer a significant contribution by providing a data-driven solution to the challenges of the NFT market.

The final recommendation engine provides a practical and user-friendly tool to mitigate the risks associated with NFT investments by offering personalized, budget-conscious suggestions. This work proves that by leveraging advanced data science techniques, it is possible to transform chaotic, raw data into a source of actionable insights that can empower investors and enhance decision-making.

Future Enhancements:

- Real-time Data Integration: Connecting the system to live marketplace APIs would enable real-time predictions and recommendations. This would address the highly dynamic nature of the NFT market and provide the most current insights to users.
- Expanded Features: The predictive models could be improved by incorporating additional data sources beyond transactional history. Future work could include features such as social media sentiment, creator reputation, and on-chain metrics like transaction fees, which are known to influence market fluctuations.

- Advanced Models: While the LSTM model performed well, exploring more sophisticated deep learning architectures like Transformer models could lead to greater predictive accuracy. These models excel at handling long sequences and complex temporal patterns.
- Enhanced Interoperability: The NFT ecosystem is becoming increasingly cross-chain. Future versions of this system could be enhanced to support data from multiple blockchains, providing a more holistic view of the market.
- User Interface: Developing a user-friendly web interface would make the system more accessible to a broader audience. This interface could feature interactive dashboards to allow users to explore the data, and a clear, simple way to receive and act on the recommendations.
- Explainable AI (XAI): As the models become more complex, incorporating XAI would make the decision-making process transparent, helping users trust and understand the reasons behind a particular price prediction or recommendation.
- Ethical Considerations: As the NFT space evolves, it would be important to develop and adhere to ethical guidelines to ensure fairness, transparency, and accountability in the algorithms.

3.5References

1. Bagili, A. P. (2024). *NFT Price Prediction using Machine Learning*. San José State University.
2. Jeyarish, J., Rajesh, R., & Nandhakumari, R. (2024). CRYPTOCURRENCY PRICE PREDICTION AND NFT ANALYSIS USING DEEP LEARNING. *TIJER - INTERNATIONAL RESEARCH JOURNAL*, *11*(5), a234–a237.
3. Kabir, M. R., Bhadra, D., Ridoy, M., & Milanova, M. (2024). LSTM–Transformer-Based Robust Hybrid Deep Learning Model for Financial Time Series Forecasting. *Journal of Risk and Financial Management, 7*(1), 7.
4. Luo, J., & Guo, W. (2024). (PDF) LSTM-Based Time Series Prediction for Bitcoin Price Analysis: A Case Study with Evaluation Metrics and Performance Insights. *ResearchGate*.
5. Huang, Z., Zheng, H., Li, C., & Che, C. (2024). Application of Machine Learning-Based K-means Clustering for Financial Fraud Detection. *Academic Journal of Science and Technology*, *10*(1), 33–39.

6. Wilson, C.-A. (2025). Explainable AI in Finance: Addressing the Needs of Diverse Stakeholders. *CFA Institute Research and Policy Center*.
7. Momeni, M., & Mohseni, M. (n.d.). Clustering Stock Market Companies via K-Means Algorithm. *SciSpace*.
8. Al-Masoud, M. N. A. (2025). LSTM-Based Time Series Prediction for Bitcoin Price Analysis: A Case Study with Evaluation Metrics and Performance Insights. *ResearchGate*.
9. Raval, D., & Ghayvat, H. (2024). A Hybrid CNN-LSTM Model for Cryptocurrency Price Prediction. *MDPI*.
10. S. H. Kim and K. A. Lee, "A Survey of Product Recommendation System for Online Platforms," *International Journal of Scientific Research in Engineering Technology*, vol. 11, no. 3, 2024, pp. 907-915.