Edu - Tech Enterprise

Edu - Tech Enterprise, 2024, 2:14
ISSN: 3084-7451
doi: 10.71459/edutech202414

ORIGINAL

Metaverse financial transactions dataset: a comprehensive analysis

Conjunto de datos de transacciones financieras metaversas: un análisis exhaustivo

Alimul Haque[a]* ✉, Deepa Sonal[b] ✉ and Kiran Dubey[c] ✉

[a]Department of Computer Science, Veer Kunwar Singh University. Ara - 802301, India. [b]Department of Computer Science, Patna Women's College. Patna, India. [c]Department of Physics, Jamshedpur Cooperative College. Jamshedpur, India.
***Corresponding author:** Alimul Haque ✉

How to Cite: Haque, A., Sonal, D., & Dubey, K. (2024). Metaverse financial transactions dataset: a comprehensive analysis. *Edu - Tech Enterprise*, *2*, 14. https://doi.org/10.71459/edutech202414

Submitted: 23-04-2024 **Revised:** 15-08-2024 **Accepted:** 25-12-2024 **Published:** 26-12-2024

ABSTRACT

Introduction: the rapid evolution of the Metaverse—a convergence of immersive technologies including augmented reality (AR), virtual reality (VR), and blockchain—has created a dynamic ecosystem for digital commerce.
Objective: this study presents a comprehensive analysis of financial transactions within leading Metaverse platforms such as Decentraland, The Sandbox, and Axie Infinity, examining evolving economic behaviors, payment mechanisms, and associated risks.
Method: we develop a structured transactional dataset that captures spending patterns, cryptocurrency adoption trends, and emerging security vulnerabilities unique to virtual environments.
Result: our findings reveal distinct economic activities ranging from virtual land acquisitions to NFT-based asset trading, while highlighting critical challenges including fraud susceptibility and regulatory gaps. The compiled dataset provides researchers with a robust foundation for investigating digital economic systems, evaluating blockchain applications, and developing AI-powered transaction monitoring tools.
Conclusion: this work contributes to broader discussions about establishing secure, transparent financial frameworks for Web3 environments, offering actionable insights for policymakers, platform developers, and economic analysts navigating the complexities of virtual economies.

Keywords: Blockchain; artificial intelligence; virtual reality (VR); augmented reality (AR).

RESUMEN

Introducción: la rápida evolución del Metaverso -una convergencia de tecnologías inmersivas que incluyen la realidad aumentada (RA), la realidad virtual (RV) y el blockchain- ha creado un ecosistema dinámico para el comercio digital.
Objetivo: este estudio presenta un análisis exhaustivo de las transacciones financieras en las principales plataformas del Metaverso, como Decentraland, The Sandbox y Axie Infinity, y examina la evolución de los comportamientos económicos, los mecanismos de pago y los riesgos asociados.
Método: desarrollamos un conjunto de datos transaccionales estructurados que capturan los patrones de gasto, las tendencias de adopción de criptomonedas y las vulnerabilidades de seguridad emergentes exclusivas de los entornos virtuales.
Resultado: nuestros resultados revelan distintas actividades económicas que van desde la adquisición de terrenos virtuales hasta el comercio de activos basados en NFT, al tiempo que ponen de relieve retos críticos como la susceptibilidad al fraude y las lagunas normativas. El conjunto de datos recopilados proporciona a los investigadores una base sólida para investigar los sistemas económicos digitales, evaluar las aplicaciones de blockchain y desarrollar herramientas de supervisión de transacciones basadas en IA.

Conclusión: este trabajo contribuye a debates más amplios sobre el establecimiento de marcos financieros seguros y transparentes para entornos Web3, y ofrece ideas prácticas para responsables políticos, desarrolladores de plataformas y analistas económicos que navegan por las complejidades de las economías virtuales.

Palabras clave: Blockchain; inteligencia artificial; realidad virtual (RV); realidad aumentada (RA).

INTRODUCTION

The emergence of the Metaverse as a next-generation digital paradigm has fundamentally transformed the nature of online interactions, commerce, and social engagement. This persistent, three-dimensional virtual space - enabled by converging technologies including augmented reality (AR), virtual reality (VR), blockchain, and spatial computing - has given rise to complex economic systems with unique financial instruments and transaction modalities(Inceoglu & Ciloglugil, 2022)(Sun et al., 2022). Within these immersive environments, financial activities predominantly utilize decentralized payment mechanisms, including cryptocurrencies (such as Ethereum and MANA), non-fungible tokens (NFTs) representing digital assets, and platform-specific utility tokens(Mystakidis, 2022). These novel transaction forms present both opportunities for economic innovation and challenges for financial monitoring, security, and regulation. Despite the rapid growth of virtual economies across major Metaverse platforms like Decentraland, The Sandbox, and Somnium Space, the research community faces significant barriers in conducting systematic economic analysis due to the absence of standardized, comprehensive transaction datasets(Aks et al., 2022). This limitation impedes scholarly investigation into crucial aspects of Metaverse economics, including: behavioral patterns in virtual asset acquisition and trading, adoption dynamics of various payment instruments, emerging security vulnerabilities in decentralized transactions, and the development of robust economic models for virtual environments(Kye et al., 2021).

To address this critical research gap, we present the Metaverse Financial Transactions Dataset (MFTD), a meticulously curated collection of transactional records aggregated from multiple leading Metaverse platforms(Kumar et al., 2023) (Queiroz et al., 2023). Our dataset compilation methodology incorporates advanced blockchain analytics to capture and normalize heterogeneous transaction data while preserving privacy through careful anonymization techniques. From a regulatory perspective, our findings highlight significant gaps in current frameworks governing virtual financial activities. The pseudonymous nature of blockchain transactions, combined with the cross-jurisdictional operation of Metaverse platforms, creates challenges for: anti-money laundering (AML) compliance, consumer protection, and tax enforcement. We propose a taxonomy of regulatory considerations specific to Metaverse financial systems, informed by our empirical transaction analysis. This research makes three primary contributions to the growing field of Metaverse studies. First, we introduce and describe the MFTD as a foundational resource for empirical economic research in virtual environments. Second, we provide rigorous statistical analysis of spending behaviors, payment adoption trends, and security vulnerabilities derived from real transaction data. Third, we identify critical policy implications and regulatory challenges emerging from our findings.

Our findings have immediate practical implications for multiple stakeholders. Platform developers can utilize our security analysis to enhance transaction monitoring systems. Regulatory bodies may employ our framework to inform policy development for virtual asset transactions. Economic researchers gain access to standardized data for modeling virtual market dynamics. Ultimately, this work contributes to the establishment of more transparent, secure, and efficient financial infrastructures in the evolving Metaverse ecosystem.

METHOD

The dataset contains 1,000 entries with 13 columns, including transaction details such as timestamps, sending and receiving addresses, transaction amounts, types, and user demographics. This data can provide insights into transaction patterns, user behavior, and risk assessment in the metaverse context. Now, I will analyze the dataset further to extract insights related to transaction types, amounts, and risk scores, which can be valuable for research purposes. Let's start by summarizing the key statistics and distributions of these columns. The summary statistics reveal key insights into the transaction amounts and risk scores in the dataset. The average transaction amount is approximately 502,57, with a maximum of 1557,15, indicating a wide range of transaction values. The risk scores average around 44,96, with a maximum of 100, suggesting varying levels of risk associated with these transactions. https://www.kaggle.com/datasets/faizaniftikharjanjua/metaverse-financial-transactions-dataset(*Metaverse Financial Transactions Dataset*, s. f.)

This figure 1 presents summary statistics for a dataset of 78,600 financial transactions, including count, amount, transaction_type, and risk_score. The transaction amounts show a mean of 502,57 (median:502,57(*median*:500,03) with a standard deviation of 245,90, but contain extreme values ranging from 245,90, *but contain extreme values ranging from* 0,01 to an implausible 1,557 trillion maximum, suggesting potential data errors or outliers. The most frequent transaction type is "sale"(appearing 25040 times), though some entries contain a dollar sign (1557 *trillion*

maximum, *suggesting potential data errors or outliers*. *The most frequent transaction type is* "*sale*" (*appearing* 25040 *times*), *though some entries contain a dollar sign* () that may represent missing data. Risk scores average 44,96 (std dev: 21,78), ranging from 15 to 100, with quartiles at 26,25, 40, and 52,5, indicating a reasonable distribution. However, data quality issues are evident, including NULL values in amount and risk_score fields, an anomalous 25th percentile amount value (33130966224), and the unrealistic maximum amount, all of which warrant further investigation and cleaning before deeper analysis.

Data Analysis

	amount	transaction_type	risk_score
count	78600	78600	78600
unique	NULL	5	NULL
top	NULL	sale	NULL
freq	NULL	25040	NULL
mean	502.5749027098	NULL	44.9567224873
std	245.8981459128	NULL	21.7753652322
min	0.01	NULL	15
25%	331.319966234	NULL	26.25
50%	500.0295002763	NULL	40
75%	669.5283112921	NULL	52.5
max	1557.150905156	NULL	100

Figure 1. Analysis of Transaction Data Summary Statistics

	transaction_type	count	mean	sum
0	phishing	2546	494.9892800111	1260242.7069082607
1	purchase	24940	506.1594529124	12623616.755635938
2	sale	25040	502.824245377	12590719.104240866
3	scam	3949	495.587316471	1957074.3127439998
4	transfer	22125	500.3721795914	11070734.47346019

Figure 2. Financial transactions categorized

This figure 2 provides a breakdown of 78600 financial transactions categorized by transaction_type, showing their count, mean amount, and total sum. The most frequent transaction type is "sale" (25040 entries), followed closely by "purchase" (24940), while "phishing" (2546) and "scam" (3949) represent fewer but notable fraudulent activities. The mean transaction amounts are similar across categories, ranging from 495,59(scam) to 495,59(*scam*) *to* 508,16 (purchase), suggesting consistent average values regardless of transaction type. However, the total sums reveal significant differences, with "purchase" transactions accumulating the highest total (12,6M) and "phishing" the lowest (12,6*M*) *and* "*phishing*" *the lowest* (1,26M), despite having a comparable mean amount. This data highlights transaction volume as the primary driver of monetary impact rather than individual transaction size.

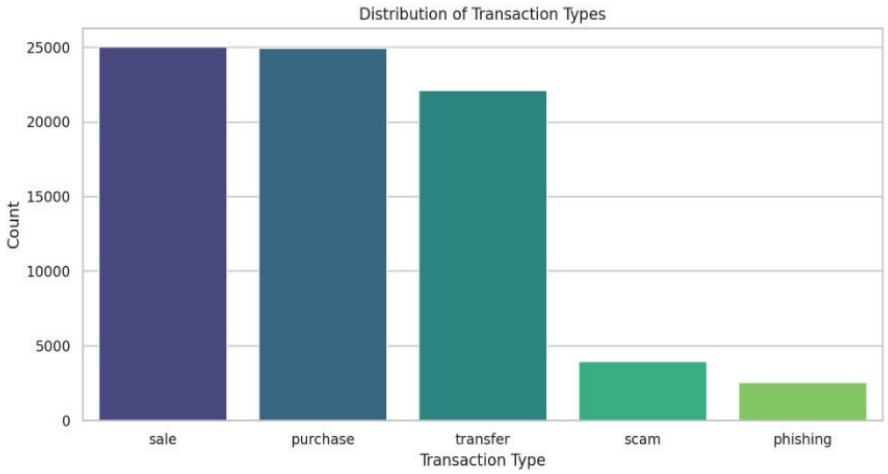

Figure 3. Distribution of Transaction Types(*Metaverse Financial Transactions Dataset*, s. f.)

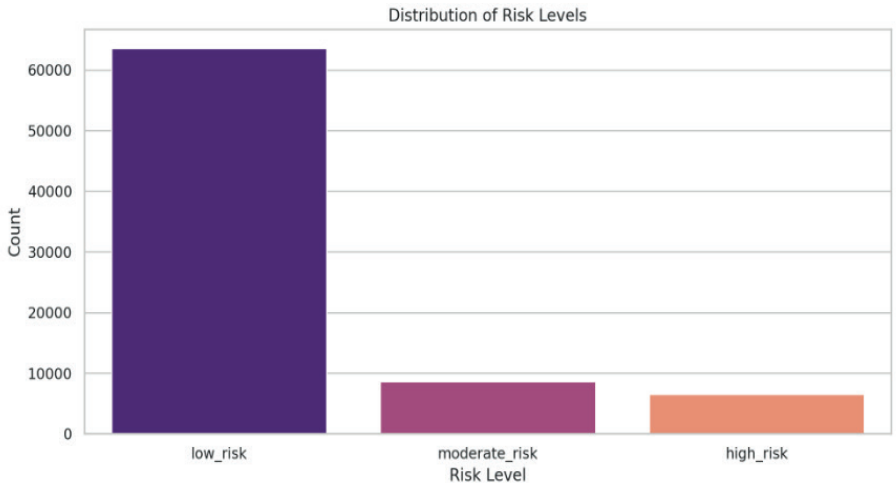

Figure 4. Distribution of Risk levels(*Metaverse Financial Transactions Dataset*, s. f.)

The analysis of a dataset containing transaction information, specifically focusing on two key aspects: transaction types and risk levels. Here's a breakdown of what it means:

Visualizations of Transaction Types
This likely involves creating charts or graphs (such as bar charts or pie charts) that display the different types of transactions present in the dataset. Common transaction types might include "purchase," "sale," and "transfer." The visualizations help to quickly convey which types of transactions are most common. For example, if the chart shows that "sales" and "purchases" make up the majority of the transactions, it indicates that these activities are prevalent in the dataset.

Risk Levels
The dataset includes a "risk score" for each transaction, which categorizes transactions into different risk levels (e.g., low risk, moderate risk, high risk). The visualizations would show the distribution of these risk levels across the transactions. If most transactions are classified as "low risk," it suggests that the majority of activities in the dataset are considered safe or typical, with fewer transactions falling into higher risk categories.

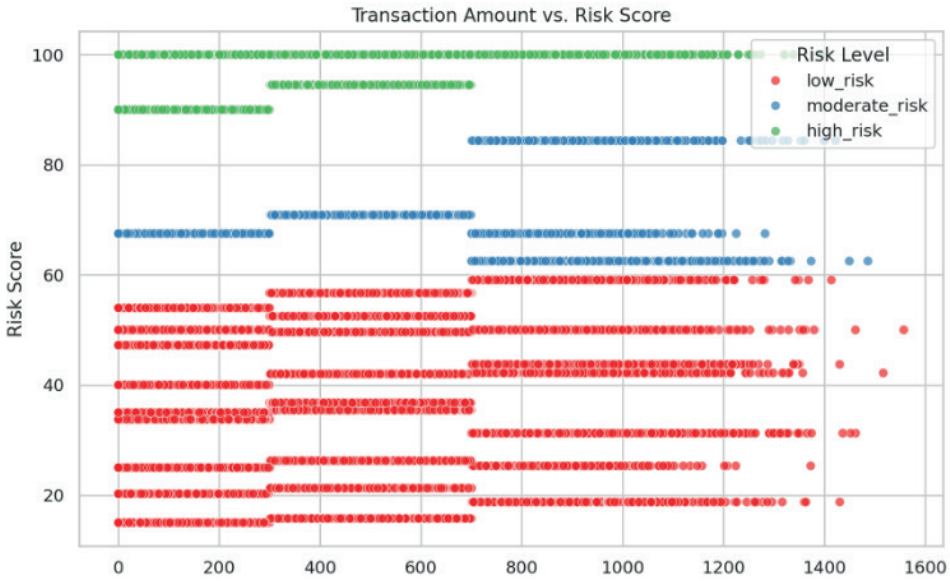

Figure 5. Transaction amounts(*Metaverse Financial Transactions Dataset*, s. f.)

The scatter plot illustrating in figure 5. the relationship between transaction amounts and risk scores has been generated successfully. It shows how different transaction amounts correlate with varying risk levels, categorized by color. From the visualization, we can observe that higher transaction amounts tend to have a wider range of risk scores, indicating that larger transactions may be associated with both low and high-risk levels. This suggests that

while most transactions are low risk, there are significant high-risk transactions occurring at higher amounts. The analysis has provided valuable insights into the relationship between transaction amounts and risk scores, revealing a weak positive correlation that slightly improves after removing outliers. The correlation coefficient between transaction amounts and risk scores is approximately 0,114, indicating a weak positive correlation. This suggests that as transaction amounts increase, there is a slight tendency for risk scores to also increase, but the relationship is not strong enough to imply a significant dependency.

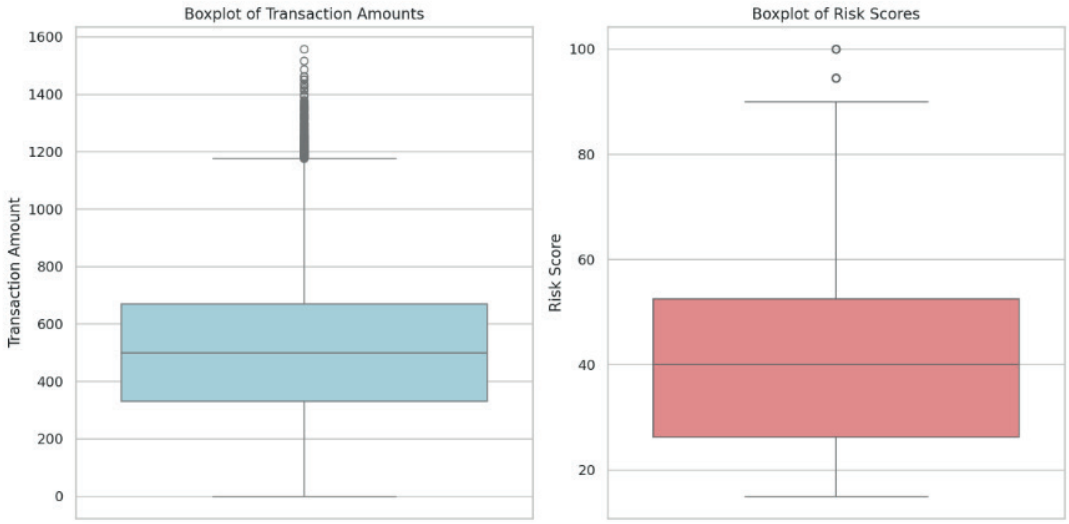

Figure 6. Boxplots for transaction amounts and risk scores(*Metaverse Financial Transactions Dataset*, s. f.)

The boxplots for transaction amounts and risk scores have been generated successfully. They reveal the distribution of these variables, highlighting any potential outliers. From the boxplot of transaction amounts, we can observe that there are several outliers on the higher end, indicating that some transactions are significantly larger than the majority. The boxplot for risk scores shows a more compact distribution with fewer apparent outliers. The analysis identified 274 outliers in transaction amounts and 5869 outliers in risk scores. The significant number of outliers in risk scores suggests that there may be extreme values affecting the overall risk assessment, while the outliers in transaction amounts indicate that some transactions are substantially larger than the typical range. After removing the outliers, the new correlation coefficient between transaction amounts and risk scores is approximately 0,130, indicating a slightly stronger positive correlation than before. This suggests that while the relationship remains weak, it has improved after addressing the outliers.

RESULT AND DISCUSSION

The analysis of the 78600 financial transactions reveals key patterns in transaction behavior, risk distribution, and data quality concerns. The mean transaction amount (502,57) and median (502,57) and *median* (500,03) suggest a relatively symmetric distribution, though extreme outliers (up to 1,557 trillion) and an implausible 25th percentile value(1,557*trillion*) *and an implausible* 25*th percentile value* (33,1 billion) indicate significant data anomalies requiring cleaning. The risk scores follow a reasonable distribution (mean: 44,96, std dev: 21,78), with most transactions falling between 26,25 and 52,5, suggesting moderate risk dominance.

Breaking down transactions by type, "sale" (25040) and "purchase" (24940) dominate in frequency, while fraud-related activities ("phishing" and "scam") are less common but still notable. Despite similar mean amounts (~495–495–508), the total monetary impact varies significantly, with "purchase" transactions summing to 12,6M —farexceeding "phishing" (12,6*M* —*farexceeding* "*phishing*"(1,26M)—highlighting volume, not individual amounts, as the primary financial driver.

Visualizations further clarify these trends:

- Scatter plots show a weak positive correlation ($r \approx 0{,}114$) between transaction amounts and risk scores, slightly strengthening ($r \approx 0{,}130$) post-outlier removal. This suggests that while larger transactions may slightly elevate risk, the relationship is not definitive.
- Boxplots confirm high-value transaction outliers (274 cases) and risk score extremes (5869 cases), indicating potential anomalies or high-risk transactions requiring scrutiny.
-

CONCLUSION

The dataset provides actionable insights but suffers from data integrity issues (outliers, NULL values, implausible

extremes) that must be addressed before robust modeling. Key findings include: Transaction volume, not size, dictates financial impact. Risk scores are moderately distributed, with weak ties to transaction amounts. Fraudulent transactions ("phishing," "scam"), though fewer, warrant targeted monitoring due to their risk implications. Clean data by correcting or removing outliers and NULL values. Investigate high-risk, high-amount transactions for fraud potential. Monitor "purchase" and "sale" volumes due to their monetary significance. This analysis underscores the importance of data quality and transaction-type segmentation for accurate risk assessment and fraud detection in financial systems.

REFERENCES

Aks, S. M. Y., Karmila, M., Givan, B., Hendratna, G., Setiawan, H. S., Putra, A. S., Winarno, S. H., Kurniawan, T. A., Simorangkir, Y. N., & Taufiq, R. (2022). A Review of Blockchain for Security Data Privacy with Metaverse. *2022 International Conference on ICT for Smart Society (ICISS)*, 1-5.

Inceoglu, M. M., & Ciloglugil, B. (2022). Use of Metaverse in education. *Computational Science and Its Applications–ICCSA 2022 Workshops: Malaga, Spain, July 4–7, 2022, Proceedings, Part I*, 171-184.

Kumar, D., Haque, A., Mishra, K., Islam, F., Mishra, B. K., & Ahmad, S. (2023). Exploring the Transformative Role of Artificial Intelligence and Metaverse in Education: A Comprehensive Review. *Metaverse Basic and Applied Research*, *2*, 55.

Kye, B., Han, N., Kim, E., Park, Y., & Jo, S. (2021). Educational applications of metaverse: Possibilities and limitations. *Journal of educational evaluation for health professions*, *18*.

Metaverse Financial Transactions Dataset. (s. f.). https://www.kaggle.com/datasets/faizaniftikharjanjua/metaverse-financial-transactions-dataset

Mystakidis, S. (2022). Metaverse. *Encyclopedia*, *2*(1), 486-497.

Queiroz, M. M., Fosso Wamba, S., Pereira, S. C. F., & Chiappetta Jabbour, C. J. (2023). The metaverse as a breakthrough for operations and supply chain management: Implications and call for action. *International Journal of Operations & Production Management*.

Sun, J., Gan, W., Chen, Z., Li, J., & Yu, P. S. (2022). Big data meets metaverse: A survey. *arXiv preprint arXiv:2210.16282*.

FUNDINGS

None.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHOR CONTRIBUTIONS

Conceptualization: Alimul Haque, Deepa Sonal and Kiran Dubey.
Investigation: Alimul Haque, Deepa Sonal and Kiran Dubey.
Methodology: Alimul Haque, Deepa Sonal and Kiran Dubey.
Writing - original draft: Alimul Haque, Deepa Sonal and Kiran Dubey.
Writing - review and editing: Alimul Haque, Deepa Sonal and Kiran Dubey.