Detecting Fraud

March 07, 2025
Data Visualization
Prof. Jodi Hill

Ryan Dsouza,
Priyanka Girish Smart,
Junaid Shaik,
Namratha Peddamalla,
Shradha Shedge

Introduction

In the rapidly evolving metaverse scenario, blockchain technology plays a crucial role in enabling secure, transparent, and decentralized transactions. However, with greater monetary activity online comes the danger of fraudulent trades, scams, and illicit activities. In counteracting these dangers, the knowledge of transaction trends and the ability to find anomalies are essential in spotting and preventing fraudulent operations in the metaverse.

We have a data set of 78,600 blockchain financial transactions of the Open Metaverse, which is a virtual world that exists on blockchain technology. The data set has been created to reflect a comprehensive and realistic data set, recording various types of transactions, user behaviors, and risk profiles in a global network of metaverse actors. It is a handy tool for constructing and testing models for fraud detection, anomaly detection, and predictive analytics in virtual worlds.

The dataset contains a rich collection of transaction records, each providing valuable insight into financial activity within the Open Metaverse. Timestamp and Hour of Day track when transactions occur and Sending Address and Receiving Address track the blockchain addresses involved. Amount captures the transaction value, and Transaction Type classifies each transaction (e.g., transfer, sale, scam). The Location Region, Sending Address, Receiving Address, and IP Prefix variables were not so effective for this analysis because the data was uniformly sampled by regions such that significant patterns could not be recognized based on location.

User behavior is modeled by Login Frequency, which measures the frequency at which a user logs in, and Session Duration, which measures the duration at which users stay active within the metaverse. Purchase Pattern categorizes user expenditure behavior. Age Group categorizes users into new, established, or veteran, based on activity history, and Risk Score assigns a risk value to every transaction based on several variables. Finally, the Anomaly variable classifies transactions by their risk status (e.g., high_risk, moderate_risk, low_risk).

These variables form a solid foundation for identifying and analyzing potential fraud and anomalous behavior within the blockchain environment which will be explained in detail in the exploratory data analysis section.

Exploratory Analysis

Our exploratory analysis focused on understanding the patterns and relationships within the Metaverse financial transaction dataset, with particular emphasis on identifying risk factors and behavioral indicators of fraudulent activities. Through iterative visualization and analysis, we discovered several key insights that shaped our investigation.

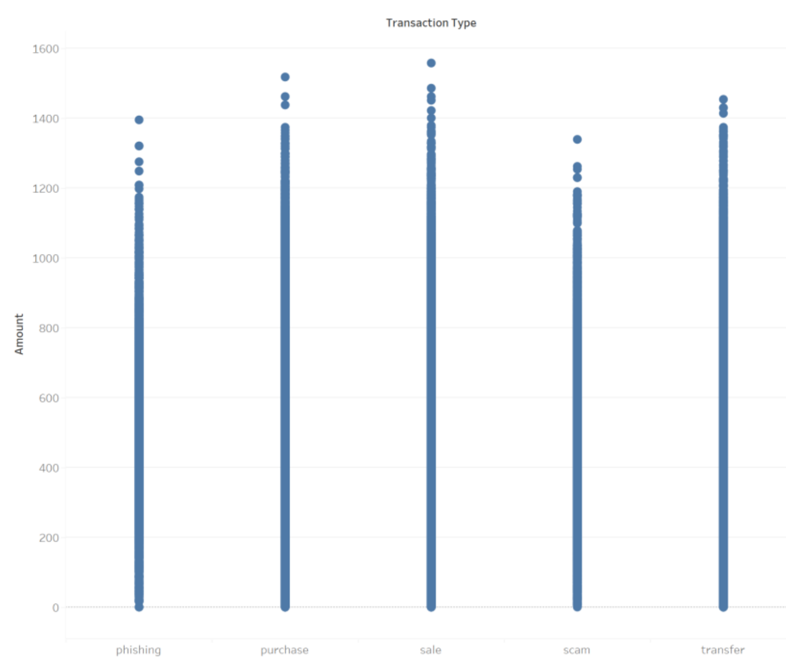
Initial Approach to Data Exploration

Our investigation into the Metaverse financial transaction dataset began with a fundamental challenge: identifying patterns of fraudulent activity within a complex digital ecosystem. With 78,600 transaction records at our disposal, we needed to uncover the hidden relationships that might distinguish legitimate activities from potential scams.

Like detectives following leads, we approached this exploration methodically, starting with the broadest possible view before narrowing our focus. Our initial steps involved examining the raw distributions of individual variables – transaction amounts, risk scores, and frequency of different transaction types – using histograms and basic bar plots. We analyzed box plots to understand the spread and central tendencies of key numerical variables across different categories.

These preliminary visualizations helped us understand the overall landscape but didn't immediately reveal the patterns we were seeking.

As we progressed, we began to notice subtle connections between user behaviors and transaction characteristics that warranted deeper investigation. The relationship between session duration and transaction type showed promise, as did patterns in purchase behavior across different user age groups.



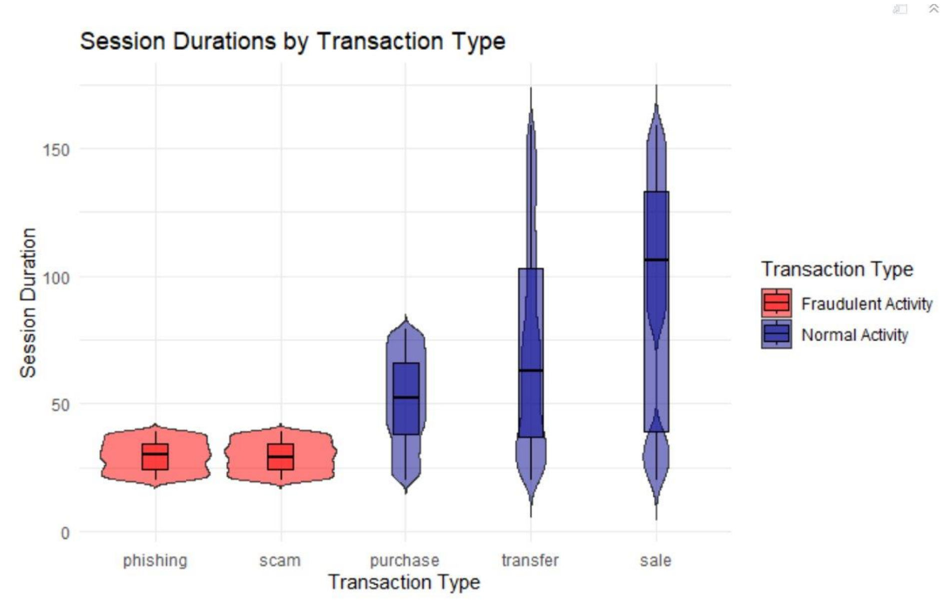
Our exploratory process evolved iteratively, with each visualization informing the next. We refined our approach based on emerging insights, gradually shifting from general distribution analysis to more focused examination of specific variable relationships. This process was guided by several key questions:

- How do behavioral patterns differ between legitimate and fraudulent transactions?
- Which user demographics show the highest vulnerability to potential scams?
- What combinations of transaction characteristics are most strongly associated with risk?
- Can we identify clear indicators that might serve as early warning signs of fraudulent activity?

As we'll demonstrate through our visualizations, this exploratory journey led us to discover several critical patterns that not only answered these questions but also provided a foundation for developing more sophisticated fraud detection approaches in the Metaverse financial ecosystem.

Visualizations

I. **Violin Plot – Session Durations by Transaction Type**



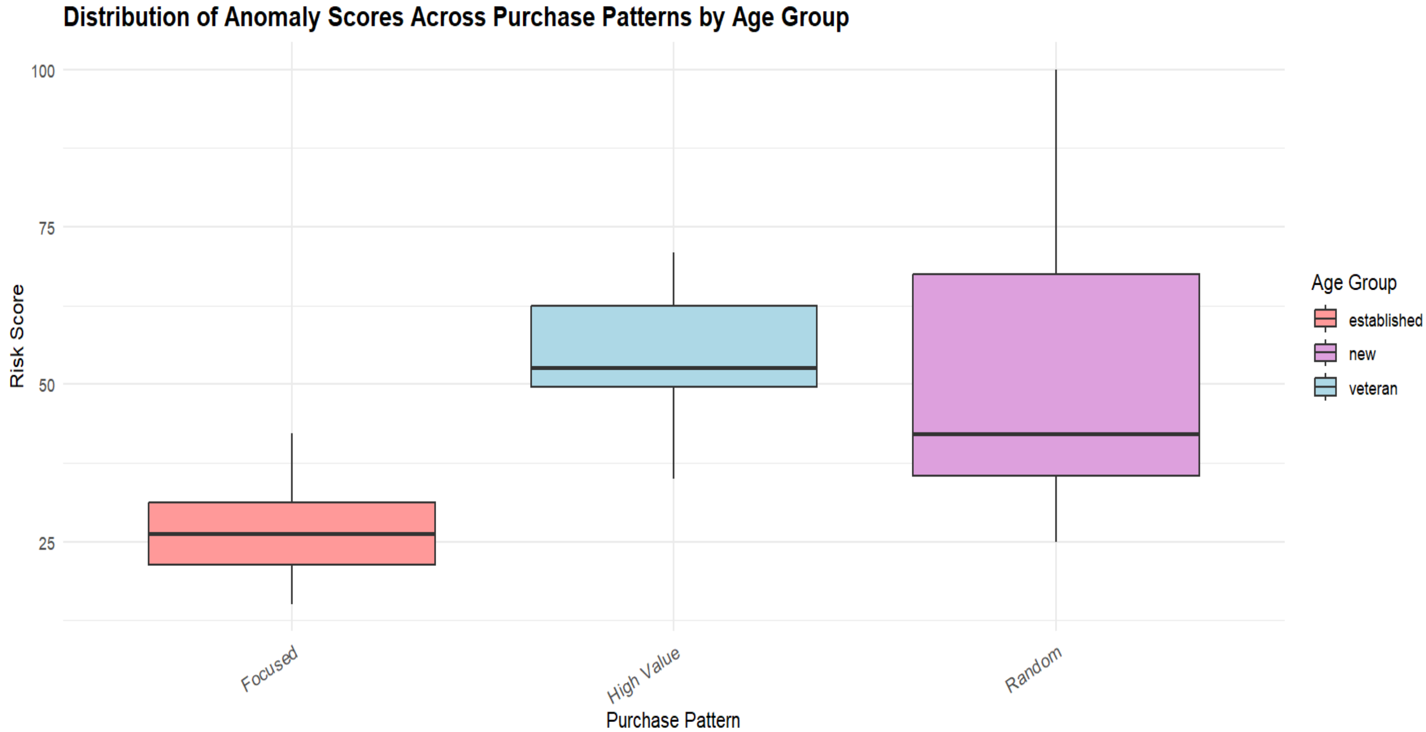
Originally, I created a rather simple box plot to analyze the different types of transactions with respect to session durations. However, with five types of transactions in my dataset, I felt that shifting the presentation toward a violin plot with an overlay box plot would best depict the distribution of session durations by transaction type. The violin plot presented a clearer overview of the way in which distributions are shaped, their density, and the occurrence of outliers. In that regard, exploration into earlier identified behavioral patterns associated with various transaction types became quite fruitful. To boost the visualization further, I changed the ordering of the transaction categories on the plot. By default, the categories were laid out in alphabetical order, but this presentation inhibited certain meaningful patterns from being perceived. The ordering was changed, with scam and phishing transactions placed relatively side by side so that watchers could comparatively clear what transpires with these higher-risk behaviors. The reordering, in a way, highlighted the similarities between the distributions of session durations, adding greater support to their association as possibly related to suspicious activity patterns.

Beyond the rearrangement of the categories, the next touch involved enhancing the color scheme of the plot to accentuate key insights. Each category was initially assigned a different color, but this was perhaps an unsatisfactory result. Instead, I decided to group the phishing and scam categories in red, highlighting a dangerous risk profile, while granting all the other categories a blue hue, indicative of less risky or more typical transactions. A legend defining the red and blue colors was included to further support the visual distinction, allowing viewers to quickly understand what the data is telling us.

In addition, I've modeled the axes and given them titles to be more readable. L made changes to the y-axis to reflect that its values are session durations in minutes--and adding a little bit of context above the plot on the data used. These changes helped form a fuller view of how people typically behave and of some of the fraud risks.

By refining the plot through these steps — using a violin plot now, reordering the categories, applying a meaningful color scheme, and labeling the different aspects — I really improved the clarity and impact of the plot. This new design helped reveal important behavioral trends tied to suspicious transactions and hence provided great insight into our team's analysis.

II. Box Plot - Distribution of anomaly score across Purchase Pattern by Age Group



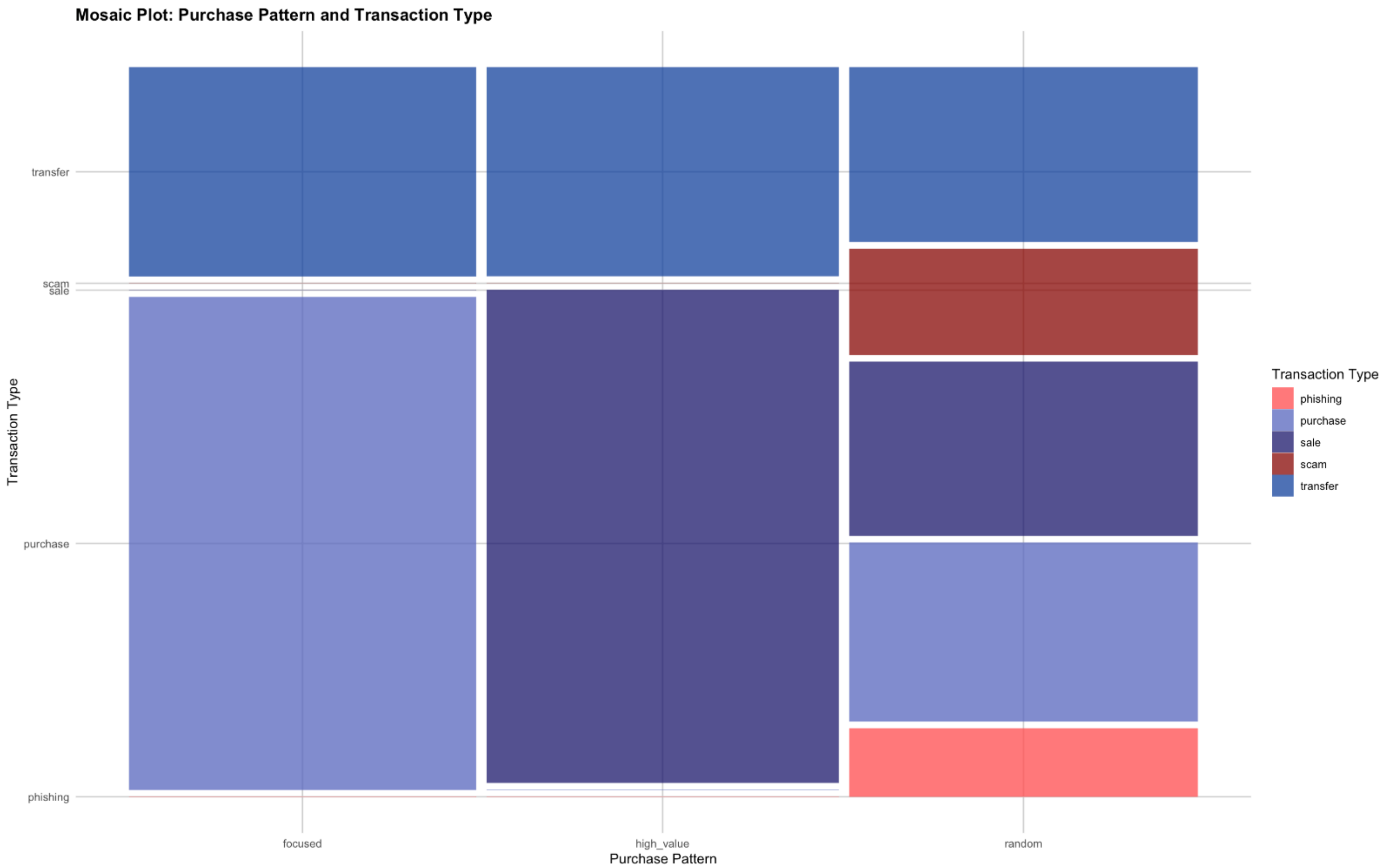
To analyze the data more appropriately, I experimented with a few visualization techniques to get a better insight into the data distribution. Techniques like Histogram to see the frequency of the amount to analyze the amount spent by the users. Scatter plots to see what amount is spent more in which risk score. Bar plot to see the transaction type spread to get which type of transaction the users are more likely to associate with.

These graphs did not provide that much information or any patterns to investigate further. While exploring the variables in the dataset, the purchase pattern intrigued as it had different categories like high value, random and focus. It is interesting to uncover how users of different age groups like new, established and veteran engage with the purchases they make and to detect if there is any purchase pattern having high risk. The relationship between purchase pattern and age group is one to one as there is only data for each purchase pattern to age group. For example, the new age group has data in random purchase pattern and not in any other purchases.

The use of Box plot is very helpful as it helps in multiple comparisons and provides an easy glance to get a quick overview of which age group is associated with which purchase pattern. Visualization provides a clear distribution of which age group has more experience and in which purchase pattern. Cleaning the dataset was very important and to see if it had any missing value in any columns, which helped more any having accurate data. To see any patterns, a pivot table plotted to see how the data is distributed in the dataset, by which it was found that there is only one value for each variable in each category.

The graph shows the distribution of risk score across by age group. It is observed that the new age group is at high risk getting targeted in random purchase. Veterans are more entitled to high value purchase pattern and for risk score it is moderate. The established age group are in low risk and in focused purchase pattern.

III. Mosaic Plot – Purchase Pattern and Transaction Type



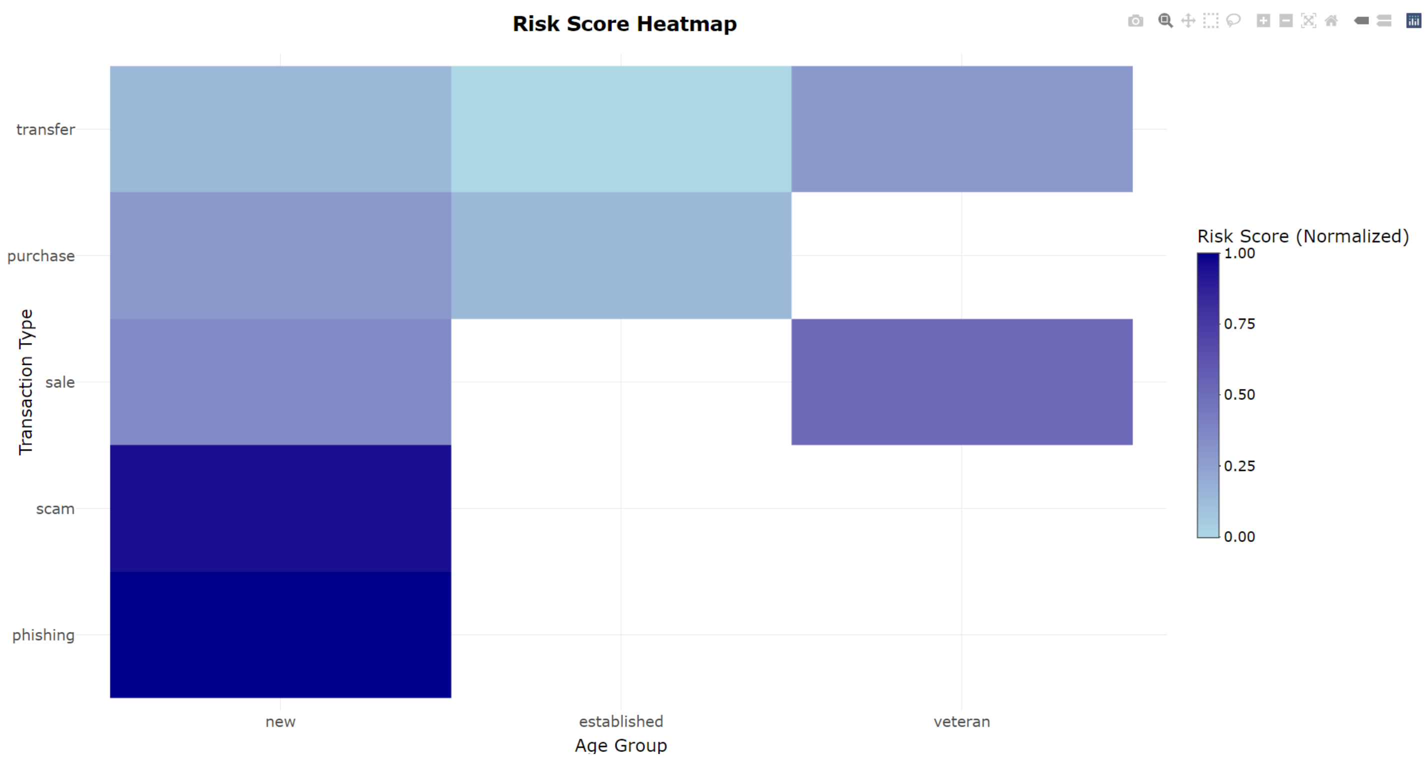
To analyze the relationship between Purchase Pattern and Transaction Type, I used a mosaic plot, which visually represents the connection between two categorical variables: Purchase Pattern (focused, high value, random) and Transaction Type (transfer, scam, purchase, phishing). The x-axis maps to the Purchase Pattern categories, while the y-axis represents the Transaction Type categories. The size of each rectangular tile indicates the frequency or proportion of observations in each combination, and the color scheme distinguishes between transaction types—phishing in red, purchase and sale in blue, and scam in dark red.

Initially, I considered using bar charts and heatmaps to display the relationship between these variables. However, I found these methods lacked the ability to effectively convey proportional relationships. That's when I decided to use the mosaic plot, which allowed me to not only show the frequency of each combination but also highlight the conditional probability distribution between the two categorical variables. The color scheme was adjusted to emphasize fraudulent transactions—phishing and scam in red tones, while legitimate transactions were represented in blue tones, making it easy to intuitively distinguish between them. I also carefully adjusted the tile sizes to reflect the underlying data's proportional relationships, ensuring an accurate representation. To enhance readability, I added clear labels and a legend for easy interpretation.

The resulting visualization uncovered some critical patterns in the data. One striking observation is that the random purchase pattern category is strongly linked to fraudulent transaction types, particularly phishing and scam, which appear prominently as red tiles in the rightmost column. In contrast, the focused and high-value purchase patterns are more likely to be associated with legitimate transaction types, such as purchase, sale, and transfer, evident from the dominance of blue tiles in the left and middle columns. This pattern suggests that random purchasing behavior is a significant risk factor for fraud in the Open Metaverse blockchain environment. The insights from this mosaic plot align with our risk score analysis, which showed that random purchase patterns were associated with the highest risk scores across age groups.

In conclusion, this mosaic plot serves as a powerful tool for identifying high-risk behaviors in blockchain transactions, providing valuable insights that can guide targeted fraud prevention strategies.

IV. Heat Map – Risk score analysis



It visualizes the **normalized risk score** across different **transaction types** and **age groups** in the metaverse transaction dataset and provides an intuitive way to identify high-risk transactions across different user demographics.

Traditional heatmaps are great for quickly spotting trends, but they can sometimes be static and lack depth. To enhance usability and interactivity, I decided to build the heatmap using ggplot2 combined with Plotly in R. This allowed users to hover over different tiles and instantly see transaction details, making the visualization more engaging and informative. This heatmap highlights how different age groups (new, established, veteran) engage in various transactions (transfer, purchase, sale, scam, phishing, etc.) and how each category correlates with risk levels. By normalizing the risk scores between 0 and 1, I ensured consistency across the dataset, making it easier to compare risk levels between different transactions.

Several enhancements were made to improve clarity and interpretation. The addition of dynamic tooltips eliminates the need for manual estimation, providing precise risk values at a glance. White grid lines were included to enhance readability, preventing color blending. New users were found to be at the highest risk, especially in phishing and scam transactions, while established users engaged in more legitimate activities like purchases and transfers. Interestingly, sales transactions showed mixed risk levels, suggesting that some may be fraudulent while others are legitimate.

This visualization offers key insights for fraud detection teams. By highlighting high-risk behaviors and vulnerable user groups, it provides a data-driven approach to identifying fraud patterns. The model can be further enhanced with real-time data integration, enabling live anomaly detection and more proactive fraud prevention.

Moving forward, this visualization could be further improved by incorporating real-time transaction data, enabling live fraud detection and anomaly detection. Additionally, adding filtering capabilities would allow users to customize their analysis based on specific risk thresholds, making it even more effective for fraud prevention teams. Ultimately, the interactive heatmap bridges the gap between complex data and actionable insights, transforming raw transaction logs into a powerful tool for understanding fraud risks in the evolving digital economy.

Analysis & Discussion

The data set provided here is of great interest to financial transactions within the Open Metaverse and shows significant characteristics for fraud detection, anomaly detection, and predictive analytics. Based on an analysis of the data, we are able to discern how transactional attributes, user behavior, and risk profiles contribute to fraud.

Our visualization provided additional clarity to these patterns. The first visualization, a violin plot, contrasted Transaction Type with Session Duration. This plot demonstrated that fraudulent transactions had significantly longer session lifespans, suggesting that scammers would keep fraudulent behavior active for extended periods. The second graph, a Risk Score vs. Purchase Pattern plot, depicted how users who followed erratic spending patterns tended to have riskier transactions. The third graph, a mosaic plot showing the relationship between Purchase Pattern and Transaction Type, revealed that random purchase patterns were strongly associated with fraudulent transaction types, particularly phishing and scam. This pattern highlights that random purchasing behavior is a key indicator of potentially fraudulent activity. Lastly, our final visualization presented Average Risk Score per Transaction Type and reiterated that those transactions categorized as scams and sales consistently had higher risk scores, indicating these transaction types as the biggest scam tell-tales.

The combined observations from these visualizations and general data insights reveal the value of integrating transaction data with behavioral intelligence for improved fraud detection. Analysis of transaction types, value, and user interaction metrics can significantly aid in identifying suspicious behavior. The presence of high-risk transactions during off-peak hours and correlation of specific types of transactions with anomalies point towards areas best suited for model improvement.

To improve fraud detection capability, the following is recommended. First, more advanced feature engineering could involve generating new features such as rolling transaction averages, cumulative user spending habits, or flagged address histories, which would improve anomaly detection. Dynamic thresholds that change risk scores based on time of day, transaction frequency, and user behavior would also improve predictive accuracy. Also, the unification of behavior analytics models with minimal transaction data would create a stronger real-time suspicious behavior detection. Finally, possessing the original dataset with complete variables without trimmed or normalized distributions would contain more meaningful information. By analyzing real distributions of single variables, more useful patterns and correlations can be realized, resulting in improved overall performance of fraud detection models.

Overall, the dataset presents a robust foundation for developing effective fraud detection models. By leveraging key variables like transaction timing, amount, and user behavior metrics, more accurate and proactive security systems can be implemented to safeguard financial activities in the Open Metaverse.

Appendix

Code

Exploratory

Box plots to check different distributions of variables

```
for (var in numeric_vars) {
  plot <- ggplot(transactions, aes(x = transaction_type, y = get(var), fill = transaction_type)) +
    geom_violin(trim = FALSE, alpha = 0.7) +  # Violin plot with full density
    geom_boxplot(width = 0.1, outlier.colour = "black", outlier.shape = 16, outlier.size = 2) +  # Add boxplot inside violin
    scale_fill_manual(values = c(custom_colors, setNames(rep(default_color,
                      length(levels(transactions$transaction_type)) - length(custom_colors)),
                      setdiff(levels(transactions$transaction_type), names(custom_colors))))) +
    labs(title = paste("Session Duration by Transaction Type"), x = "Transaction Type", y = var) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))

  print(plot)
}
```

Explanatory

Violin Plot - Session Durations by Transaction Type

```
```{r}
# Load necessary libraries
library(ggplot2)
library(dplyr)



# Convert Transaction Type to factor
# Reorder transaction types
df$transaction_type <- factor(df$transaction_type,
                              levels = c("phishing", "scam", "purchase", "transfer",
"sale"))

# Assign colors for fraudulent and normal activities
df$color <- ifelse(df$transaction_type %in% c("phishing", "scam"),
                   "Fraudulent Activity",
                   "Normal Activity")

# Create the plot
ggplot(df, aes(x = transaction_type, y = session_duration, fill = color)) +
  geom_violin(trim = FALSE, alpha = 0.5) +
  geom_boxplot(width = 0.2, color = "black", outlier.size = 2, alpha = 0.5) +
  scale_fill_manual(values = c("Fraudulent Activity" = "red",
                               "Normal Activity" = "darkblue"),
                    name = "Transaction Type",
                    labels = c("Fraudulent Activity", "Normal Activity")) +
  labs(title = "Session Durations by Transaction Type",
       x = "Transaction Type",
       y = "Session Duration") +
  theme_minimal()

```
```

Boxplot for Risk Score with Purchase pattern and age group-

Code:

library(ggplot2)
color_palette <- c("established" = "#FF9999", # Light Red "new" = "#DDA0DD", # Light Violet "veteran" = "#ADD8E6") # Light Blue

Create the Boxplot-----

ggplot(df, aes(x = purchase_pattern, y = risk_score, fill = age_group)) + geom_boxplot(outlier.shape = NA) + # Removes outliers scale_fill_manual(values = color_palette) + # Apply custom colors labs(title = "Distribution of Anomaly Scores Across Purchase Patterns by Age Group", x = "Purchase Pattern", y = "Risk Score", fill = "Age Group") + theme_minimal() + theme(axis.text.x = element_text(angle = 30, hjust = 1, size = 10), plot.title = element_text(size = 14, face = "bold"))

Mosaic Plot – Purchase Pattern and Transaction Type

Code:

```
# Load necessary libraries
library(ggplot2)
library(ggmosaic)

# Read the dataset
data <- read.csv("/Users/namratha/Downloads/metaverse_transactions_dataset.csv")

# Convert categorical variables to factors
data$purchase_pattern <- as.factor(data$purchase_pattern)
data$transaction_type <- as.factor(data$transaction_type)

# Create custom color palette with more variation in lightness
# Keeping red tones for fraudulent activities and blue tones for normal activities
transaction_colors <- c(
  "phishing" = "#FF5252",  # Brighter red for phishing
  "scam" = "#8B0000",      # Very dark red for scam
  "purchase" = "#5C6BC0",  # Medium blue for purchase
  "sale" = "#191970",      # Dark navy blue for sale
  "transfer" = "#0D47A1"   # Very dark blue for transfer
)

# Create the improved mosaic plot with varied lightness color scheme
ggplot(data) +
  geom_mosaic(aes(x = product(purchase_pattern), fill = transaction_type)) +
  # Use custom color palette with varied lightness
  scale_fill_manual(values = transaction_colors) +
  theme_minimal() +
  labs(title = "Mosaic Plot: Purchase Pattern and Transaction Type",
       x = "Purchase Pattern",
       y = "Transaction Type",
       fill = "Transaction Type") +
  theme(
    axis.text.x = element_text(angle = 0, hjust = 0.5),
    plot.title = element_text(face = "bold"),
    legend.position = "right",
    panel.grid.major = element_line(color = "gray80")
  )
```

Heat Map – Risk Score by Transaction Type and Age Group

Code:

```
# Load necessary libraries
  library(ggplot2)
```

```
> library(dplyr)
> library(tidyr)
> library(readr)
> library(plotly)  # Load plotly for interactivity
> # Load the dataset
> df <- read_csv("D:\\Depaul courses\\Data Visualization\\Final
project\\metaverse_transactions_dataset.csv", show_col_types = FALSE)
> # Define the correct order for transaction type based on risk score
> transaction_order <- df %>%
+  group_by(transaction_type) %>%
+  summarise(avg_risk = mean(risk_score, na.rm = TRUE)) %>%
+  arrange(desc(avg_risk)) %>%
+  pull(transaction_type)
> # Define the correct order for age groups
> age_group_order <- c("new", "established", "veteran")
> # Create aggregated data for heatmap
> heatmap_data <- df %>%
+  group_by(transaction_type, age_group) %>%
+  summarise(avg_risk = mean(risk_score, na.rm = TRUE)) %>%
+  ungroup()

> # Normalize risk scores
> min_risk <- min(heatmap_data$avg_risk, na.rm = TRUE)
> max_risk <- max(heatmap_data$avg_risk, na.rm = TRUE)
> heatmap_data$normalized_risk <- (heatmap_data$avg_risk - min_risk) / (max_risk - min_risk)
> # Convert to factor with correct ordering
> heatmap_data$transaction_type <- factor(heatmap_data$transaction_type, levels =
transaction_order)
> heatmap_data$age_group <- factor(heatmap_data$age_group, levels = age_group_order)
> # Create the heatmap with ggplot2 (base for plotly)
> p <- ggplot(heatmap_data, aes(x = age_group, y = transaction_type, fill = normalized_risk, text =
paste(
+  "Transaction Type:", transaction_type, "<br>",
+  "Age Group:", age_group, "<br>",
+  "Normalized Risk Score:", round(normalized_risk, 2)
+ ))) +
+  geom_tile(color = "white") +  # Add white grid lines for clarity
+  scale_fill_gradient(low = "lightblue", high = "darkblue", name = "Risk Score (Normalized)") +
+  labs(title = " Risk Score Heatmap",
+      x = "Age Group",
+      y = "Transaction Type") +
+  theme_minimal() +
+  theme(
+    text = element_text(size = 14),
+    axis.text.x = element_text(angle = 0, hjust = 0.5, size = 12),
+    axis.text.y = element_text(size = 12),
+    plot.title = element_text(hjust = 0.5, face = "bold", size = 16)
+  )
> # Convert ggplot2 heatmap into interactive plotly chart
> interactive_heatmap <- ggplotly(p, tooltip = "text")
> # Save the interactive plot as an HTML file
> htmlwidgets::saveWidget(interactive_heatmap, "interactive_heatmap.html")
> # Display the interactive heatmap in RStudio viewer
> interactive_heatmap
```

Individual Reports

I. Ryan Dsouza

Role & Contributions

My role within our project was to manage submission of key milestones and keep our group well organized and on track in the process. I maintained and created our group documents and shared files, which ran efficiently. I also periodically checked in with team members to double check that work was proceeding on their visualizations to avoid duplicating efforts. This sophisticated coordination enabled us to maintain our project schedule and made sure that each member's work complemented the overall analysis.

On the data side, I conducted exploratory data analysis to gain an insight into the dataset and identify meaningful trends. My primary contribution in this area involved the session duration data to investigate user behavior and identify patterns regarding phishing or scam transactions. I built a violin plot visualization highlighting the distribution of session duration across different transaction types, focusing particularly on distinguishing user behavior for high-risk transactions such as phishing and scams. The visualization facilitated easy formation of an insight into how session duration might be correlated with suspicious activity to inform anomaly detection in blockchain transactions.

In addition to my technical contribution, I also assisted in our presentation facilitation. Since two of our team members were not present, I offered to present two of our visualizations, ensuring our team's results were well articulated. I also assisted in formatting the overall narrative of our presentation, ensuring our narrative was brief, logical, and aligned with our data-based conclusions.

Reflection

Throughout this project, I discovered a few extremely significant things about data visualization techniques and how they assist in conveying critical information. Firstly, I discovered the importance of applying the correct type of visualization to the data—the violin plot came in very handy when illustrating the distribution of session length while also depicting behavioral trends. I also observed how exploratory data analysis was essential in uncovering unexpected patterns that helped inform our team's decision on what sections of the dataset to investigate in more detail.

Team coordination also required good communication and careful planning. A combination of technical activity and team coordination improved my leadership and time management skills. Coming in to provide two visuals and help format the presentation script taught the lesson of being adaptable and team-oriented in defining success. Overall, the project provided a better understanding of how data visualization can make storytelling easier in data analysis, especially when dealing with complicated sets of data such as blockchain transactions within the Open Metaverse.

II. Priyanka Girish Smart

Role & Contributions

My contribution focused on data selection, exploration, analysis, visualization, collaboration, presentation ensuring well-disciplined and strategic outcome.

Data Exploration and analysis:

While researching various datasets, which included searching, evaluating, analyzing, and finalizing the dataset for our project. After exploring many datasets, I came across the Metaverse dataset, a little structured and was a little unique as it contained records of how users interact in the data world.
Identifying the dataset, conducted a detailed study to analyze the structure, variables, to conclude solid insights.
Performed data cleaning and preprocessing to have consistent and accurate data for team analysis.

Uncovering trends and patterns, experimenting with many various visualization techniques which included histograms, bar plot, scatter plot, spine plot, mosaic plot, box plot and violin plot.

Team collaboration:
Communicated with team members to know their ideas, understand their perspectives and opinions for the project.
Organized regular check-in meetings, including everyone's availability to have a smooth discussion and conversations.
Assisted team members in any issues, challenges or have any doubts and tried to provide a solution to best resolve the queries.
To make sure tasks are divided correctly, not overworking, making sure all the team members get equal work and contribution. Helped in diving the work to ensure every member's contribution.
Communicated with Professor to discuss, verify that the project was on the right track, understand the requirements of the projects, seeking feedback to improve as we go.
Presentation:

Worked upon developing the story of our findings and ensuring it is in narrative format, making sure our findings are logically accurate. Participated and worked on presenting the findings, title, overview, introduction and significance of the analysis. Helped in organizaing the contents of the presentation to be as accurate, clean and to the point so that it is easy to interpret.

Reflection:

Working on this project was an amazing experience as I got to learn data cleaning, exploring various datasets, data understanding, different techniques to understand the data, how to proceed with data exploration. Learned different data visualization techniques and determined which technique is useful to present the findings. Even though each technique is very important it depends on the context of the data and what analysis we are working on. It helped me to develop my visualization skills to not only focus on technicality of the graph but also work on making it clear and easy to read by non-technical audience. Working on the project helped us to collaborate with other amazing people to understand their views, ideas and share how we can proceed with the project findings. It is very important to work in a group to develop communication and collaboration skills. Organizing and taking initiative to track the progress of the group was also very interesting to do.

III. Junaid Shaik

Role & Contributions

My primary contribution to this project was designing an interactive heatmap that visualizes risk scores across transaction types and age groups. I focused on enhancing the interactivity of the visualization, allowing users to hover over the heatmap to view detailed risk score information dynamically. To achieve this, I implemented plotly integration in R, which transformed the static heatmap into an engaging and user-friendly exploration tool. Additionally, I carefully structured the ordering of transaction types, ensuring that high-risk behaviors such as phishing and scams appeared at the top, making fraudulent activities more noticeable at a glance. The color scheme was also fine-tuned, with light blue representing lower risk and dark blue highlighting high-risk transactions, improving clarity and interpretability.

Beyond the visualization, I conducted extensive exploratory data analysis (EDA) to understand the dataset's structure and identify key risk patterns. I examined transaction distributions, risk score trends, and user behavior patterns to determine the best way to present these insights visually. Through this analysis, I discovered that new users were particularly vulnerable to phishing and scam transactions, while established users tended to engage in safer activities like purchases and transfers. This insight guided the design and refinement of the heatmap, ensuring that it effectively communicated these trends.

In addition to my technical contributions, I played a role in collaborating with the team to ensure smooth workflow and division of tasks. I regularly checked in with teammates, provided input on their visualizations, and incorporated their feedback to enhance my own work. I also assisted in structuring our final report, making sure that our analysis and visualizations were well-organized and clearly communicated.

Reflection

This project significantly enhanced my data visualization skills, particularly in making visualizations interactive and more insightful. I learned how small refinements, such as reordering categories or choosing the right color scale, can drastically improve the clarity of a visualization. More importantly, this experience reinforced the importance of interactivity in data exploration—allowing users to engage with the data rather than just viewing static charts made a significant difference. From a teamwork perspective, I gained valuable experience in collaborative data storytelling, ensuring that our findings were not just technically sound but also effectively communicated to a wider audience. This project underscored how visualizations are not just about aesthetics but about making complex data understandable and actionable.

IV. Namratha Peddamalla

Role & Contributions

In our fraud detection project, I focused on visualization development and team coordination. I created various visualizations including bar plots, box plots, violin plot, density plots, Sankey diagram, 2D binning, scatterplots, and mosaic plots to explore patterns in our blockchain data.

I stayed connected with my teammates throughout the project, joining all meetings and helping divide tasks. I also reached out to our professor to confirm we were on track and get feedback for improvements.
I helped analyze the dataset during our exploratory phase, identifying patterns that helped guide our understanding of what would work best for our objectives. My primary contribution was the mosaic plot showing relationships between purchase patterns and transaction types, which the team selected for our final report. This visualization revealed how random purchase patterns strongly correlate with fraudulent activities like phishing and scams.

I also took part in presenting our findings, helping explain our work and creating a coherent narrative for our audience.

Reflection

This project taught me a lot about effective data visualization. I discovered that exploring multiple visualization approaches leads to stronger insights. Testing different visualization types helped me understand the strengths of each for different analytical questions.

Team communication proved incredibly valuable. Regular discussions with teammates exposed me to different perspectives, while checking in with our professor helped us stay focused and make improvements. Seeing which visualizations made it to our final report (including my mosaic plot) taught me what makes certain visualizations more compelling and effective.

The presentation experience improved my ability to explain technical findings clearly. I learned to build information gradually and highlight key insights in a way that resonates with the audience.

Overall, this project showed me that effective data visualization requires both technical skills and design sensibility, along with the ability to communicate findings as part of a larger analytical story. These lessons will be valuable as I continue developing my data analysis and visualization skills.

V. Shradha Shedge

Role & Contributions

My role encompassed data selection, exploration, analysis, visualization, collaboration, and presentation, ensuring a strategic and well-structured outcome.

In this project, I contributed by selecting relevant datasets, exploring trends, identifying patterns, and experimenting with various visualization techniques, including histograms, bar plots, scatter plots, spine plots, mosaic plots, box plots, and violin plots to uncover fraud risks in the Metaverse.

My role also involved designing clear and insightful visualizations that effectively communicated key findings, ensuring stakeholders could easily interpret fraud patterns.

Through collaboration, I helped refine the presentation of results, making complex data more accessible and actionable.

Team Collaboration

This project was a collaborative effort, where team members played a vital role in ensuring a comprehensive and accurate analysis.

Through active communication and shared expertise, we effectively divided tasks, refined methodologies, and enhanced the clarity of visualizations. Regular feedback sessions helped in iterating on the presentation, ensuring the findings were impactful and easy to interpret.

By working together, we streamlined the process, improved accuracy, and ensured the final report effectively addressed the risks and solutions related to fraud in the Metaverse.