Exploratory Analysis Report – Oxford Behavioural Finance Task

1. Introduction

This report presents the main findings from the exploratory analysis conducted on a dataset that integrates psychological measures and financial decisions. The objective is to identify preliminary patterns that can guide future modeling and inference stages, particularly regarding the relationship between individual profiles and investment strategies.

2. General Overview of the Dataset

The dataset contains a total of 786 records, corresponding to investment decisions made by 297 unique individuals. Each row represents an asset in which a person invested and includes information about:

- Psychological profile (five normalized dimensions: *confidence*, *risk_tolerance*, *composure*, *impulsivity*, *impact_desire*).
- Type of asset and invested amount.
- Investment currency.
- Date of the operation.

No missing values or duplicate records were detected. Additionally, the identification columns were correctly typed as text strings to avoid misinterpretation as numerical variables.

However, interpreting some of the psychological variables requires caution, as the assessment method is not specified. In particular:

- **Confidence** could refer to self-confidence (self-efficacy, self-esteem) or confidence in the selected asset. Given the dataset's context, we assume it refers to the former.

- **Impact_desire** could represent the desire for the investment to have an impact (a psychological intention) or the magnitude of the desired impact (a quantifiable goal). Assuming it belongs to the personality block, we interpret it as a measure of desire or intrinsic motivation.

Variable	Count	Mean	Std Dev	Min	25%	Median	75%	Max
confidence	786	0.5030	0.1026	0.176	0.432	0.505	0.565	0.885
risk_tolerance	786	0.5010	0.0776	0.299	0.449	0.500	0.551	0.745
composure	786	0.5042	0.0720	0.311	0.455	0.502	0.547	0.700
impulsivity	786	0.5009	0.2944	0.005	0.228	0.507	0.724	0.997
impact_desire	786	0.4891	0.2819	0.006	0.237	0.497	0.718	0.999
asset_value	786	6498.97	12972.08	100.52	193.45	261.22	410.74	47919.16

Table 1. Descriptive statistics of the numerical variables.

3. Univariate Analysis

3.1. Psychological Variables

The five psychological dimensions are normalized between 0 and 1. Different dispersion patterns were observed:

- **Confidence**, **risk_tolerance**, and **composure** showed centered distributions with low variability, suggesting homogeneous profiles within the sample.

- In contrast, **impulsivity** and **impact_desire** displayed greater dispersion, with nearly flat distributions. This may reflect high interindividual heterogeneity, or weaknesses in the scales' discriminative capacity.

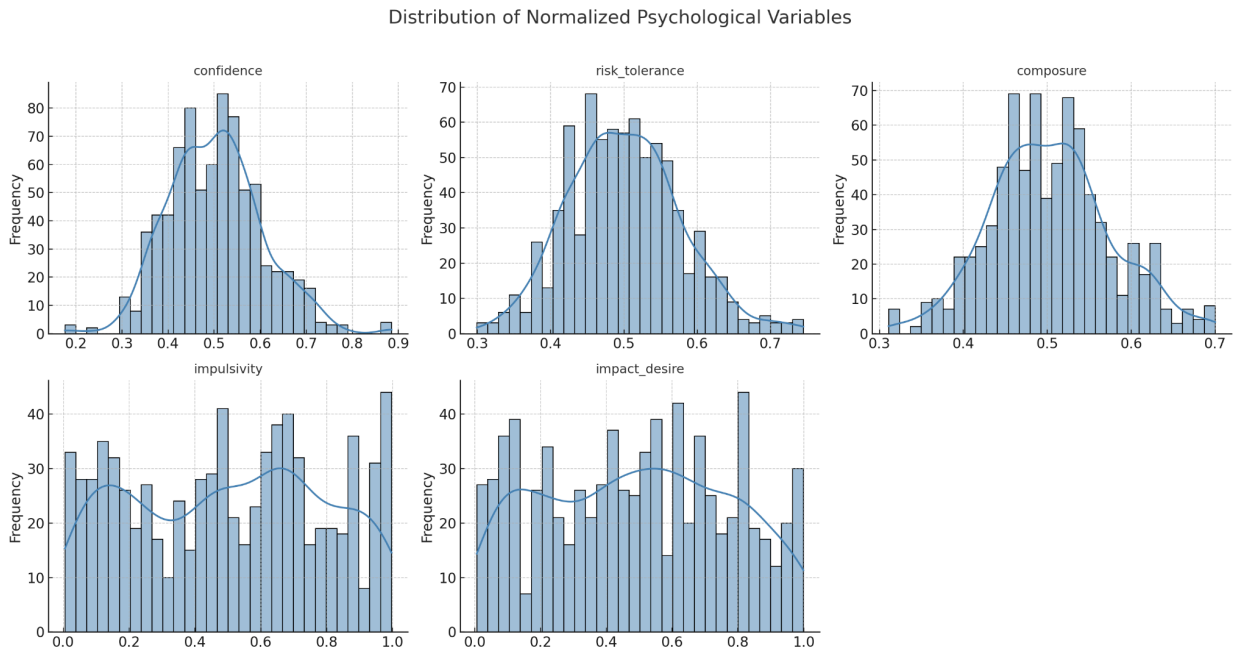

Fig. 1. Histograms of psychological variables.

3.2. Asset Value

The asset value variable showed strong positive skewness: although some values exceed 47.000 units, the median is only 261. This disparity may indicate that most assets have low amounts, while a few extreme cases heavily influence the mean, but it's important to note that asset value is conditioned by the investment currency, so interpreting these values without normalization can be misleading. This issue will be addressed in later sections.

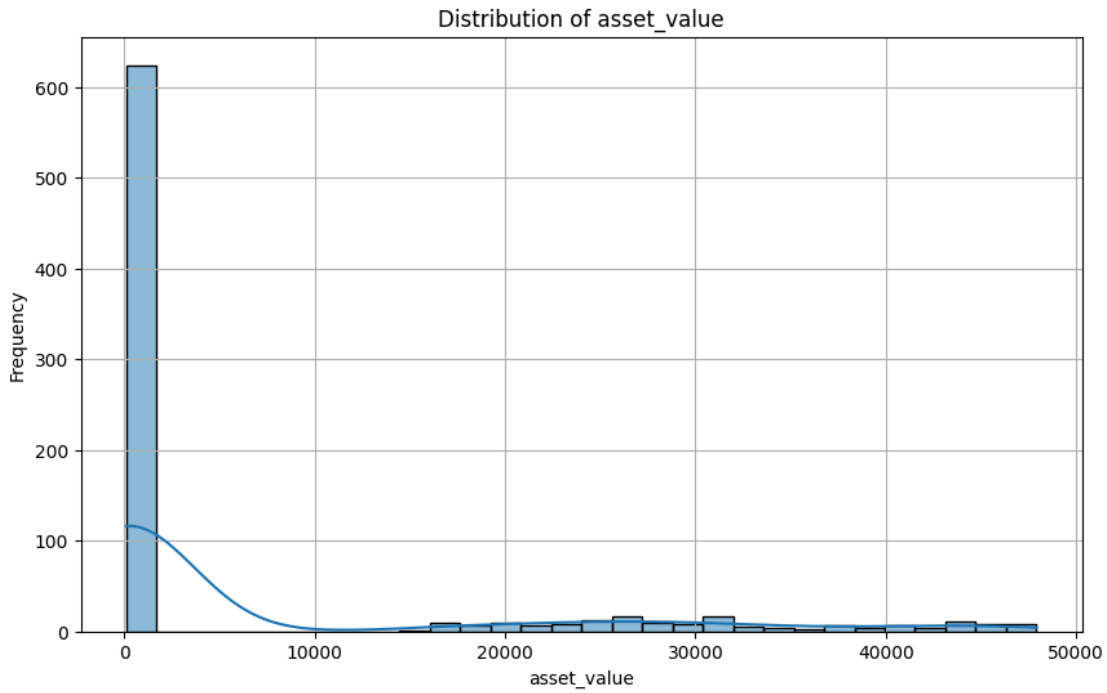

Fig. 2. Distribution of asset value

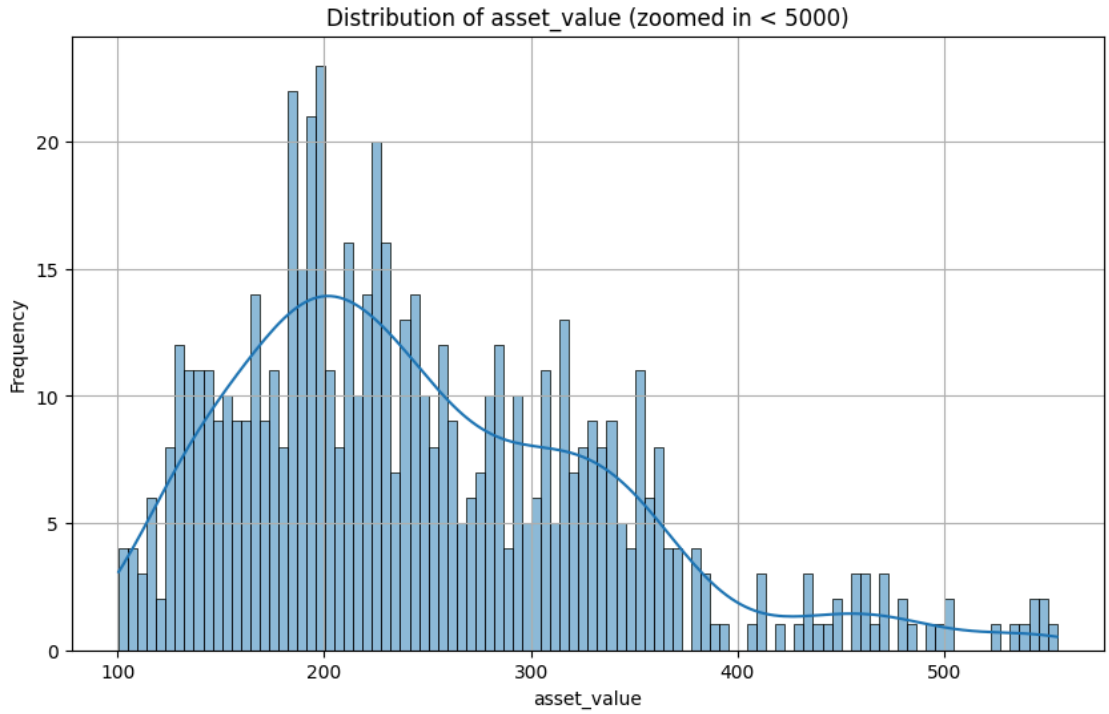

Fig. 3. Zoom on `asset value` for values below 5000 units.

3.3. Categorical Variables

Regarding asset types, **Crypto** was the most frequent, followed by **Equities** and **Commodities**. As for investment currency, the **Australian dollar (AUD)** was the most used, although the overall distribution across currencies was fairly balanced.

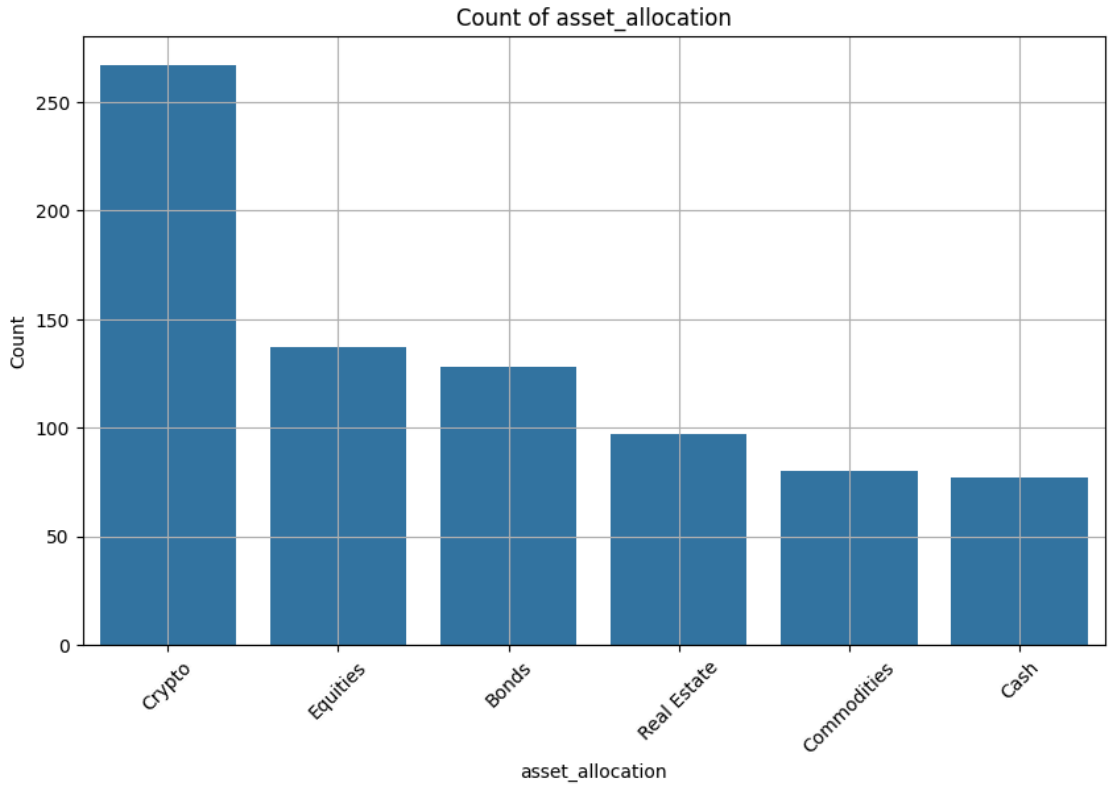

Fig. 4. Frequency of asset allocation

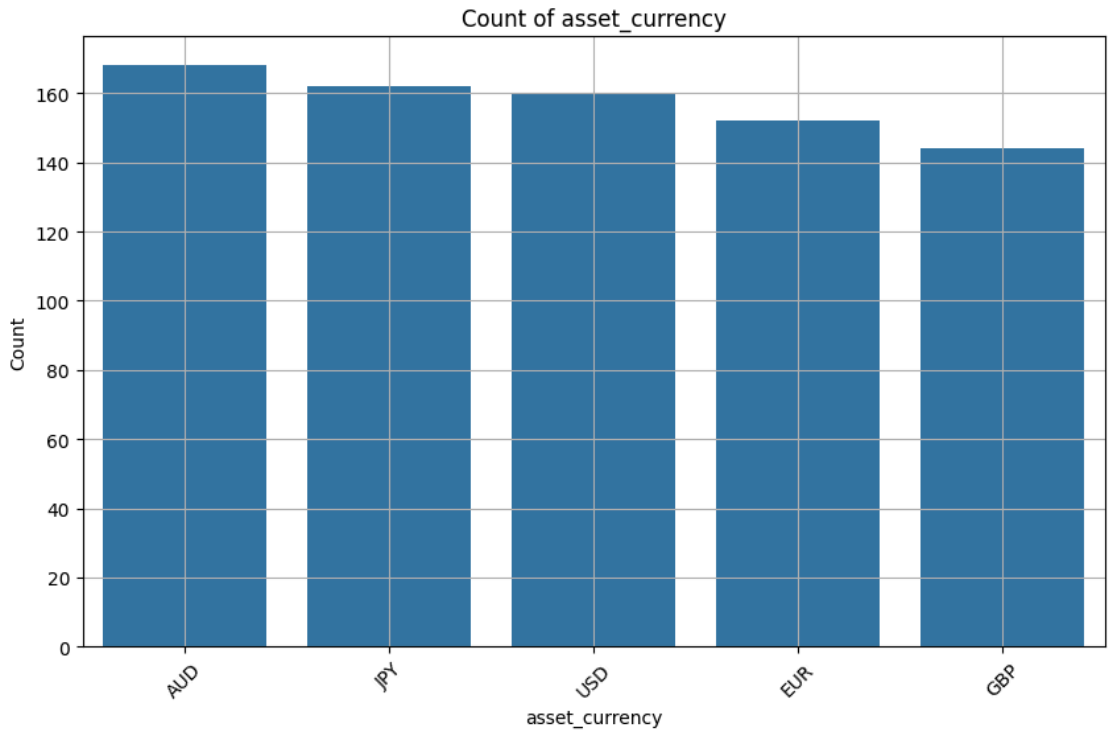

Fig. 5. Frequency of asset currency.

4. Bivariate Analysis

4.1. Correlations Between Psychological Variables

Three notable correlations stand out:

- **Confidence and Risk Tolerance (r = 0.93):** This extremely high correlation can be interpreted in two ways. On one hand, greater confidence in one's abilities may lead to a lower perception of risk, thereby increasing tolerance. On the other, both variables may be capturing the same latent construct, indicating measurement redundancy.

- **Confidence and Composure (r = 0.55):** Self-confidence may be supported by effective emotional regulation in uncertain contexts. Emotionally stable individuals tend to maintain their self-efficacy judgments even in volatile or high-pressure situations.

- **Risk Tolerance and Composure (r = 0.51):** The ability to remain calm during periods of instability may facilitate greater risk tolerance, helping avoid reactive or erratic responses to potential losses.

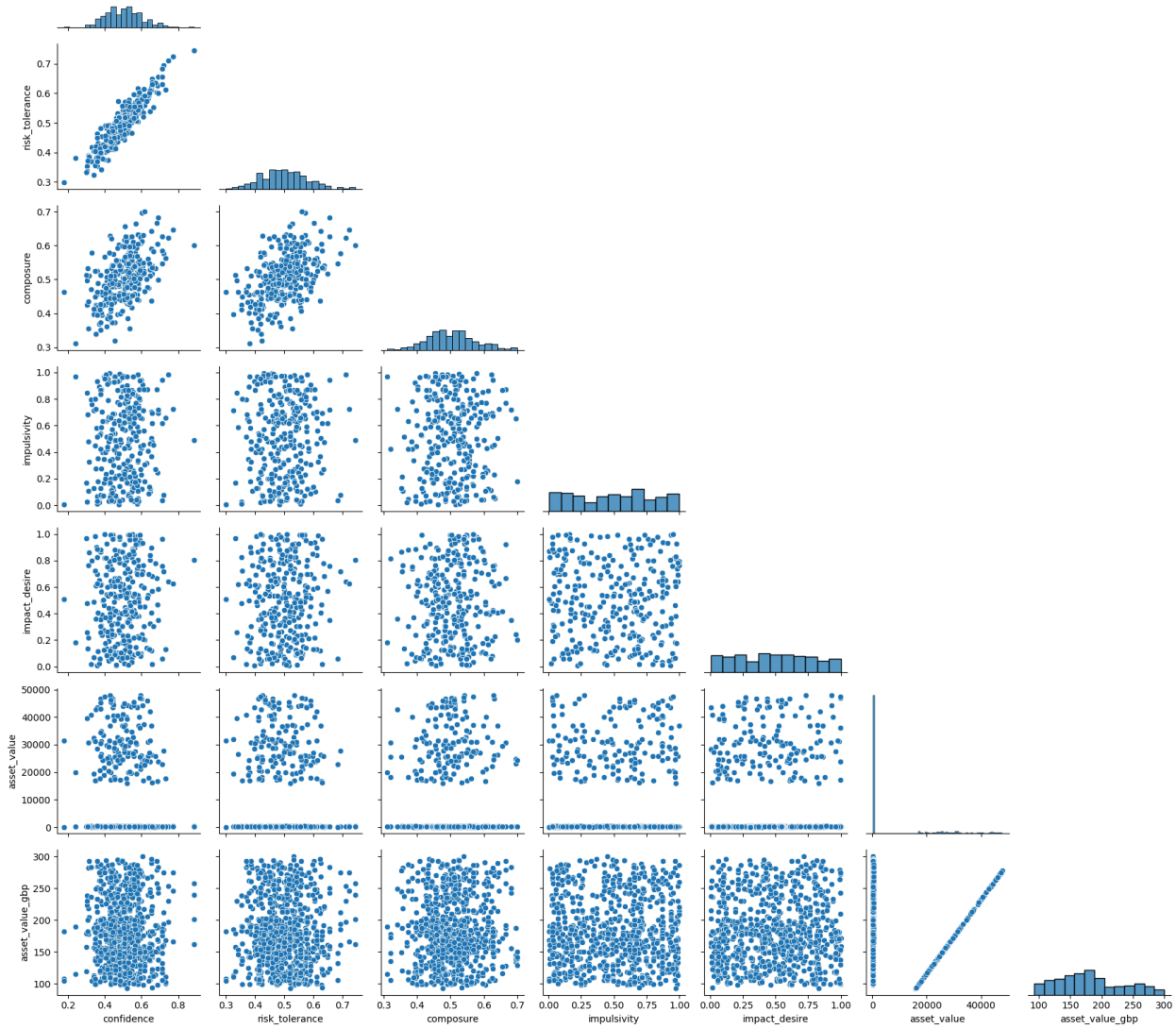

Fig. 6. Pairplot of numerical variables (asset value normalized included).

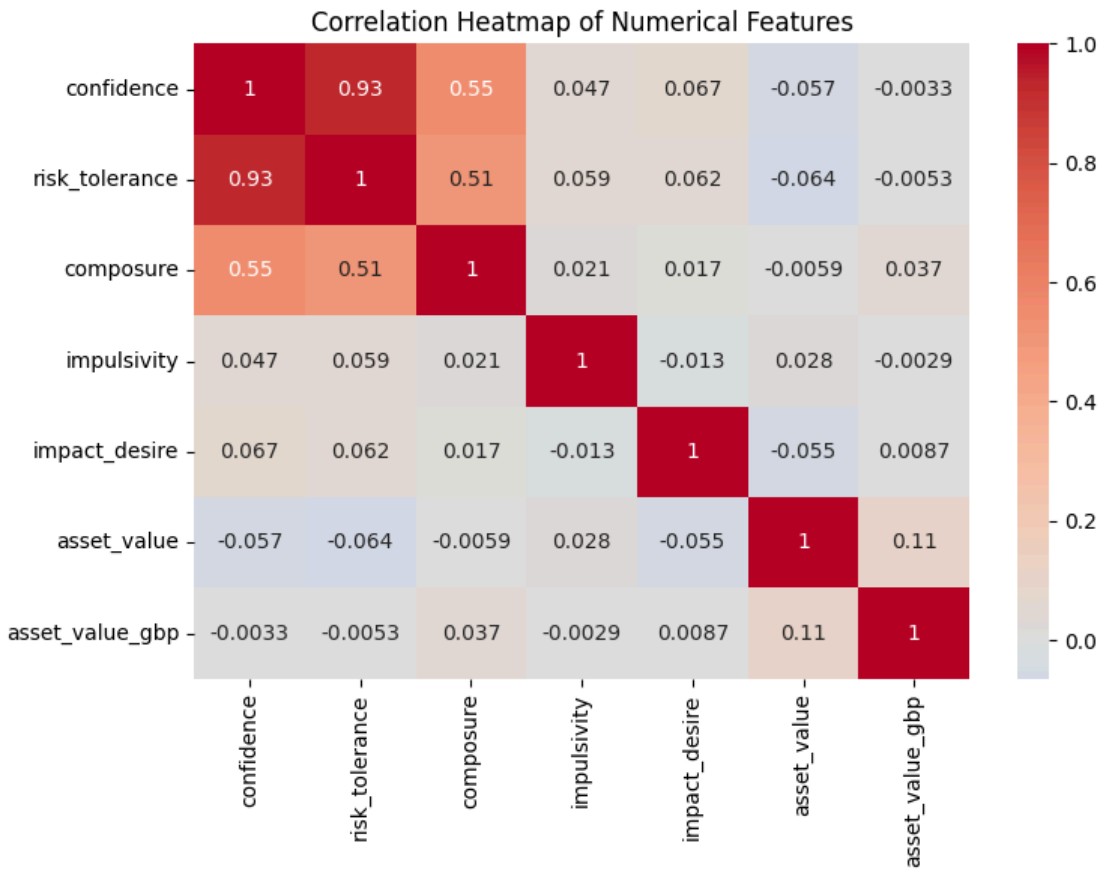

Fig. 7. Heatmap of correlations between numerical variables.

Emerging Hypothesis: These correlations suggest that the emotional component (*composure*) is moderately linked to perceived control (*confidence*) and risk attitude. In contrast, variables like *impulsivity* and *impact_desire* show no significant correlations, possibly due to their high variability. This could reflect real differences between individuals or a poor psychometric discrimination by the instruments used.

4.2. Psychological Variables vs. Asset Value

No significant correlations were found between asset value and any psychological variables. No clear associations were found when normalizing asset value to GBP either. This suggests that this amount is not directly associated with individuals' psychological profiles—at least not at the descriptive level.

The absence of relationships may be due to mediating variables not included in the dataset, such as personal income, prior investment experience, or financial goals.

4.3. Relationships Between Categorical Variables

No clear associations were found between asset type and the currency used. The proportions of currencies within each asset category were relatively stable, suggesting independence between the two variables.

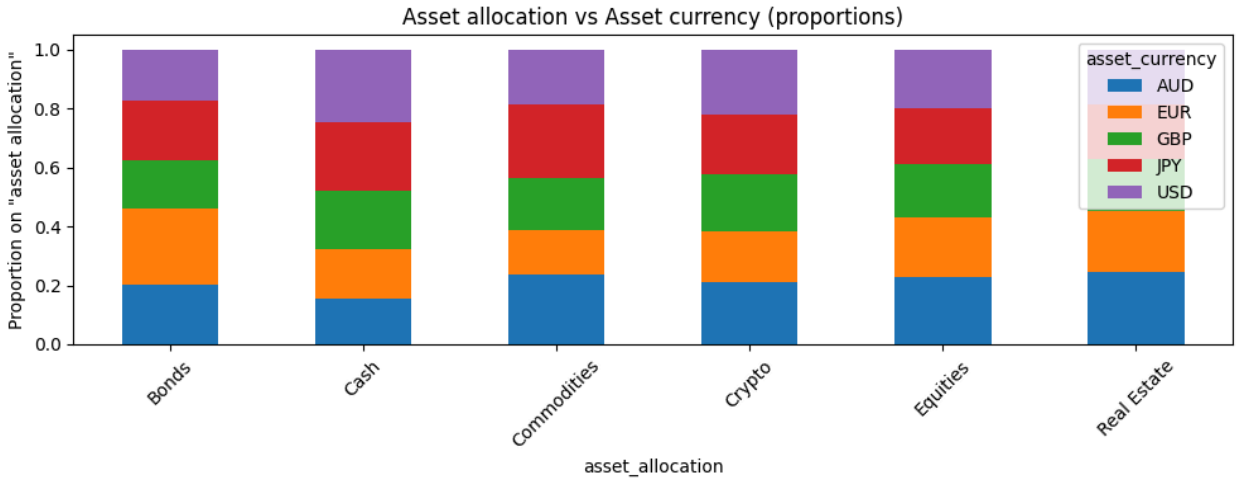

Fig. 9. Stacked bar chart of asset allocation and currencies

5. Special Case: Highest Invested Value

The individual with the highest total asset value in standardized currency (GBP) was identified as the subject 134. The sum was £542.86, and the *risk_tolerance* score was 0.555.

6. Conclusion

This exploratory analysis revealed preliminary insights into the behavioral and financial patterns captured in the dataset. The psychological dimensions displayed varying degrees of dispersion, with traits like confidence and risk tolerance showing tight clustering, while impulsivity and impact desire exhibited greater heterogeneity.

At the bivariate level, strong interrelations between confidence, risk tolerance, and composure suggest that emotionally grounded self-assurance may underlie investment risk attitudes. However, these psychological variables showed no significant association with the monetary value of investments, even after normalization. This points to the probable influence of omitted contextual factors—such as income, experience, or investment objectives—that modulate financial behavior beyond the assessed psychological disposition.

7. Next Steps

To build upon the exploratory findings using the current dataset, I propose the following directions:

1. **Cluster Analysis of Investor Profiles**: Use unsupervised learning techniques to segment individuals based on their psychological traits. This could reveal behavioral subgroups with distinct investment styles, offering a more nuanced understanding of decision-making profiles.

2. **Temporal Trend Analysis**: Analyze the investment behavior over time by leveraging the operation dates. This can help identify potential shifts in asset preferences or psychological patterns across different time periods within the dataset.

3. **Intra-individual Consistency Checks**: Examine whether individual subjects exhibit stable decision-making behaviors across multiple investments.

4. **Non-linear and Interaction Models**: Implement models capable of capturing complex relationships between variables (such as decision trees or random forests) to explore whether combinations of psychological traits better predict investment behaviors than linear associations alone.

These steps aim to refine our understanding of the behavioral signatures within the dataset and offer stronger foundations for future predictive or inferential modeling.