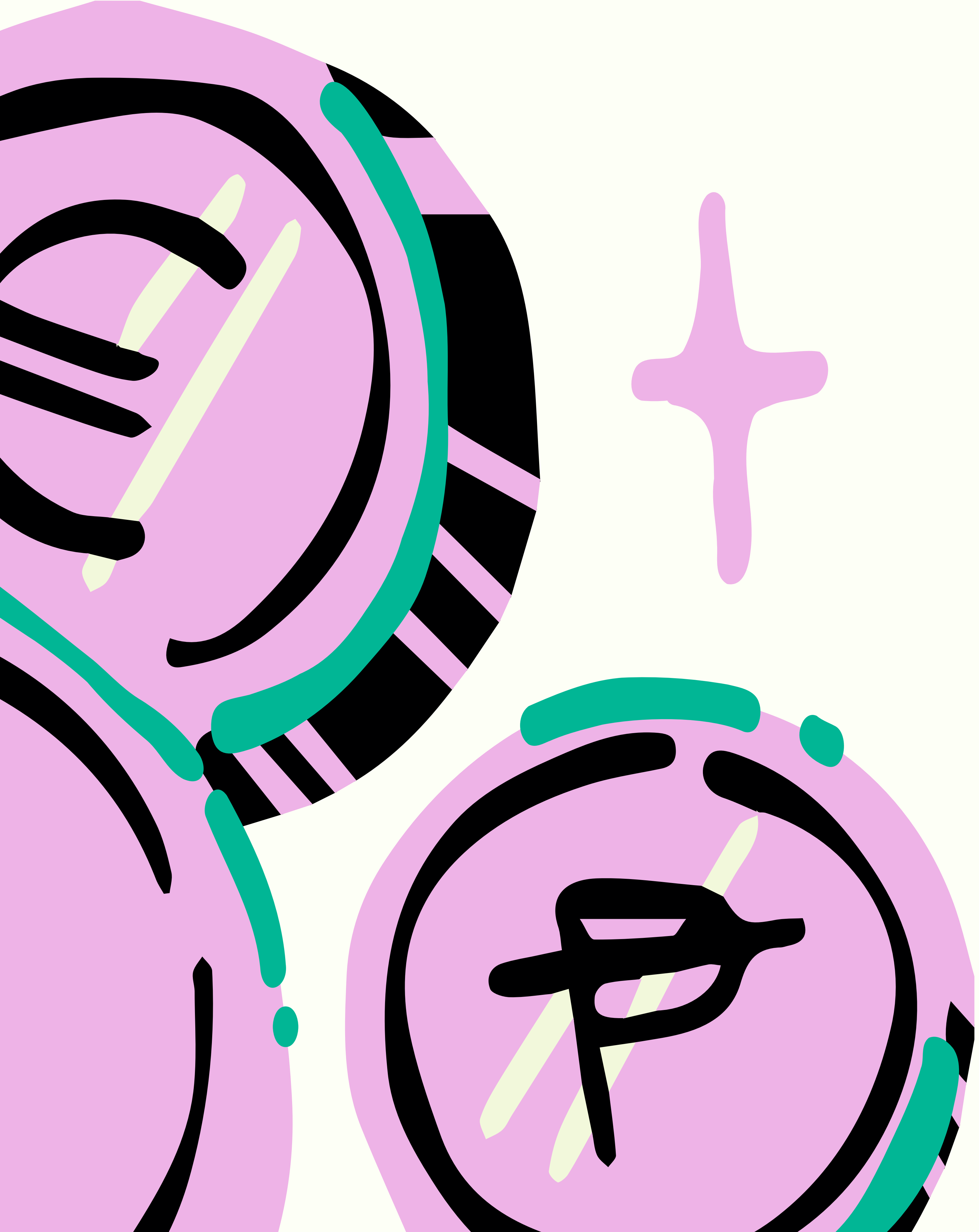
PERSONAL BUDGETING ANALYSIS BASED ON SURVEY

DONE BY
YERKIMAN LAZER



INTRODUCTION

This project analyzes personal financial habits and budgeting behavior based on survey data collected from 80 respondents. The study aims to explore how individuals allocate their expenses, save for short- and long-term goals, and manage their overall budget. The project combines data preprocessing, exploratory data analysis (EDA), and analytical methods to provide actionable insights for financial planning and personal finance education.

ACTUALITY & RELEVANCE

Effective personal financial management is essential in today’s society, especially due to rising living costs and economic uncertainty. Many individuals struggle to track spending, plan budgets, and save for both short- and long-term goals, often leading to financial stress and inefficiency.

This project aims to analyze personal budgeting habits and financial decision-making using survey data. By identifying patterns in expenses and savings, the study provides insights that can help individuals make informed financial decisions.

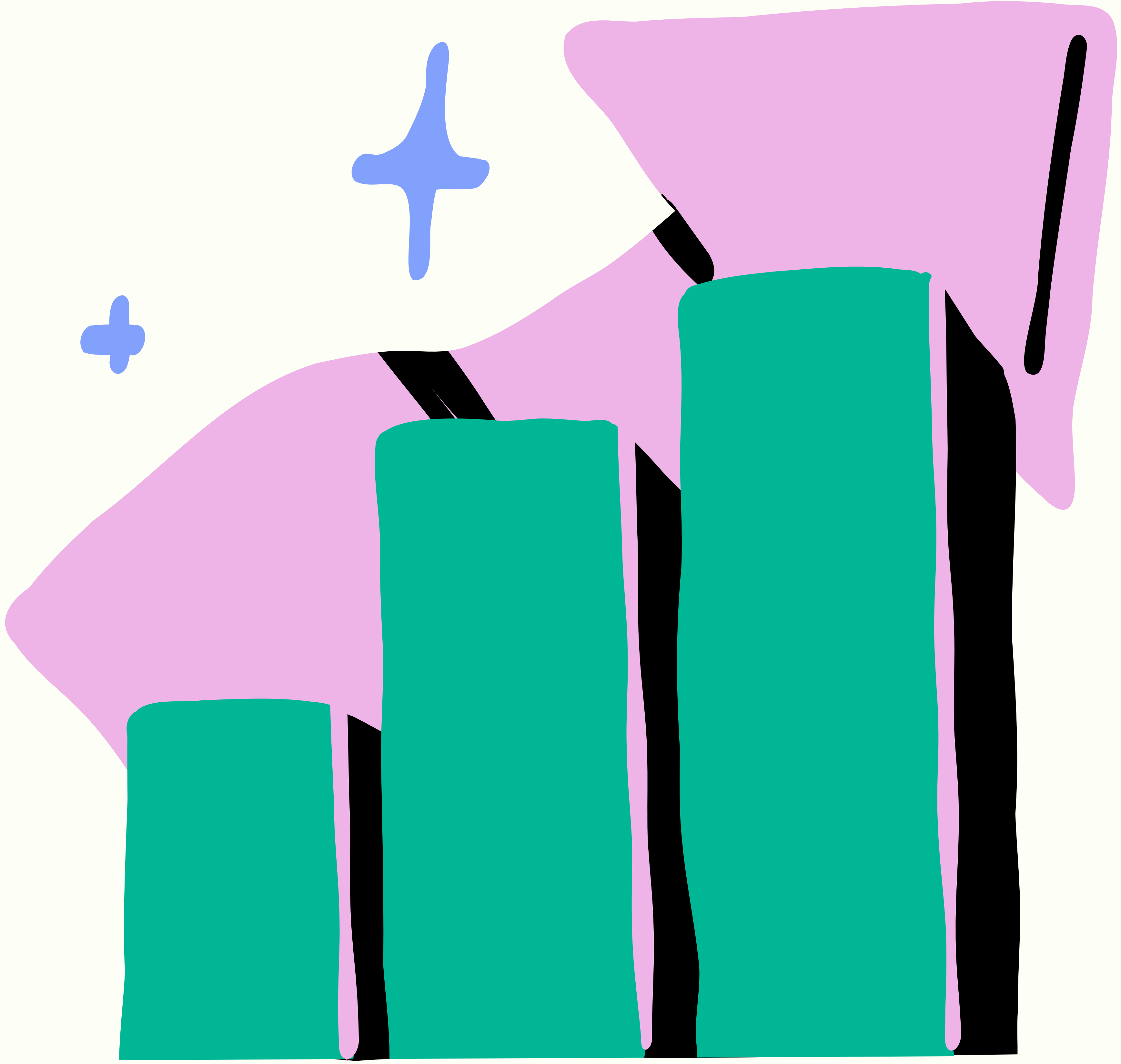
NOVELTY & ORIGINALITY

This project stands out due to its reliance on primary survey data, collected specifically to capture participants' personal budgeting habits. Unlike many existing studies that focus on large-scale transactional datasets or generalized financial literacy assessments, our dataset incorporates subjective information, including individual spending preferences, saving goals, and financial discipline. This allows for a more nuanced understanding of real-world financial behavior.

RELATED WORKS

Several studies have explored personal financial behavior and budgeting using survey or analytical approaches.

- how financial knowledge impacts spending and saving habits among university students.
- personal financial planning and expenditure control using survey responses from a diverse population.



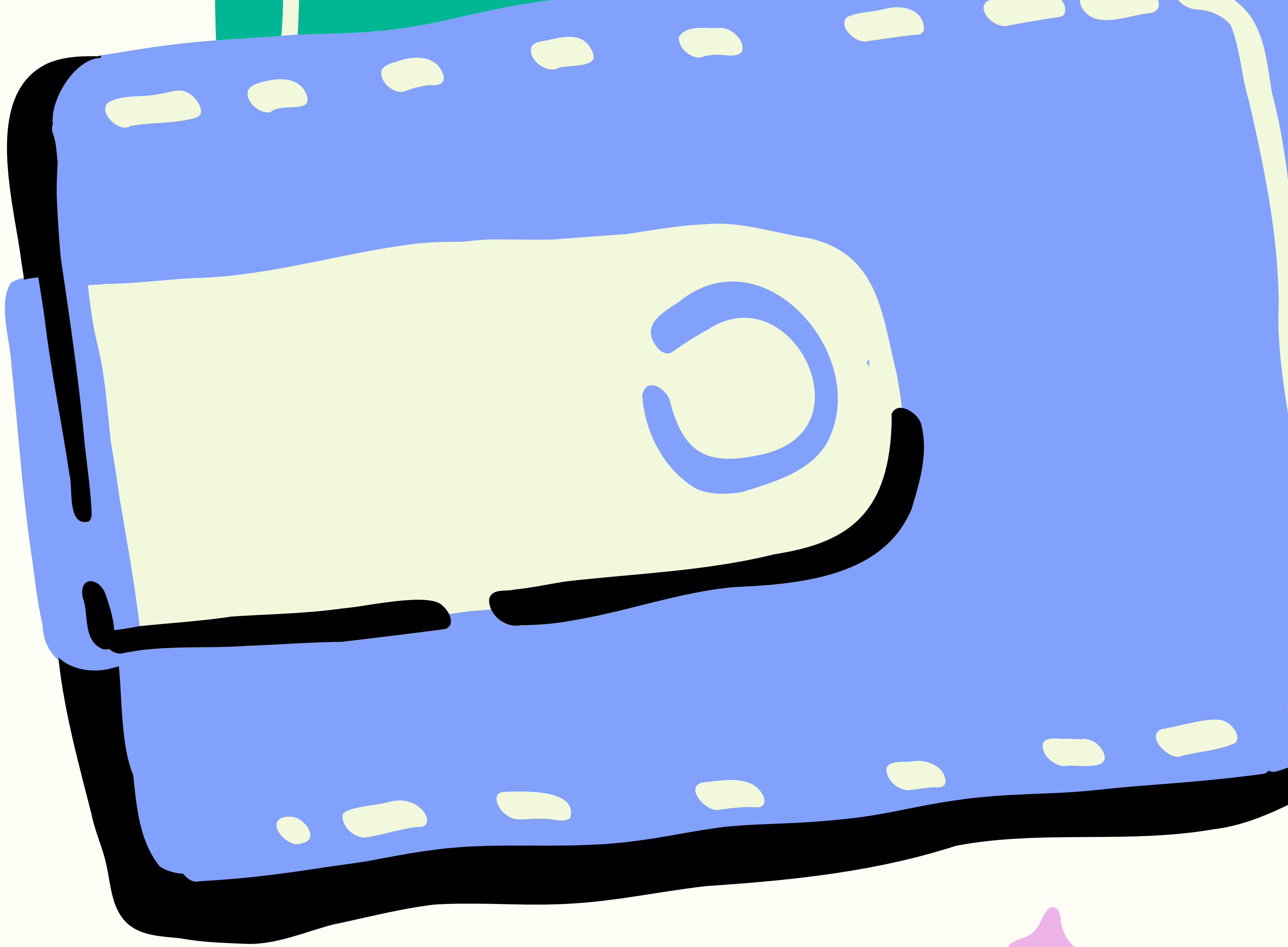
DATA SOURCE

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 80 entries, 0 to 79
Data columns (total 21 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   age                        80 non-null     int64
 1   gender                     80 non-null     object
 2   monthly_income_range       80 non-null     object
 3   total_monthly_spending     80 non-null     int64
 4   food_spending              80 non-null     int64
 5   transport_spending         80 non-null     int64
 6   entertainment_spending     78 non-null     float64
 7   shopping_spending          78 non-null     float64
 8   rent_spending              80 non-null     int64
 9   other_spending             77 non-null     float64
 10  expense_tracking           79 non-null     object
 11  budget_adherence           80 non-null     object
 12  impulse_purchase_frequency 80 non-null     int64
 13  saves_monthly              80 non-null     object
 14  amount_saved_last_month    80 non-null     int64
 15  monthly_savings_goal       60 non-null     object
 16  saving_difficulty_score    80 non-null     int64
 17  main_saving_obstacle       75 non-null     object
 18  income_stability           76 non-null     object
 19  impulse_purchase_trigger   73 non-null     object
 20  reason_for_not_saving      61 non-null     object
dtypes: float64(3), int64(8), object(10)
memory usage: 13.3+ KB
```

The dataset used in this project was collected through an online survey focused on personal budgeting and financial behavior. The final dataset consists of 80 responses and contains 21 variables, including demographic information, income ranges, detailed spending categories, saving behavior, and subjective behavioral indicators such as impulse purchase frequency and saving difficulty scores.

The data includes both numerical variables (e.g., total spending, category-level expenses, savings amounts) and categorical variables (e.g., gender, income range, budgeting habits, reasons for not saving). Since the data is self-reported, it reflects real-world personal financial perceptions and behaviors.

DATA PREPROCESSING AND FEATURE ENGINEERING



SEVERAL PREPROCESSING STEPS ENSURED DATA QUALITY AND CONSISTENCY.

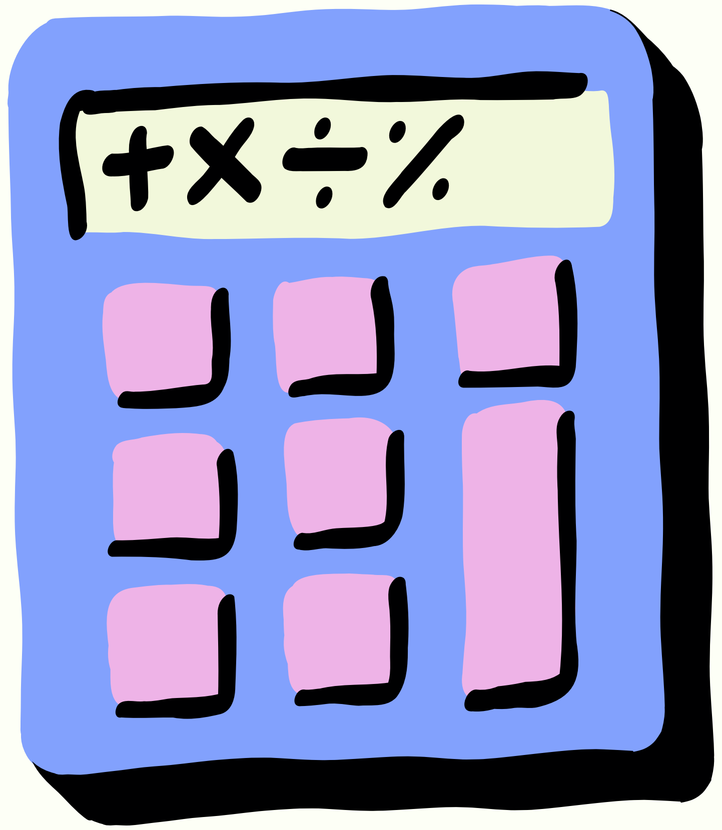
MISSING VALUES WERE IMPUTED USING THE MEDIAN

NUMERICAL SPENDING VARIABLES WERE CONVERTED TO NUMERIC FORMAT

WHILE CATEGORICAL MISSING VALUES WERE FILLED WITH AN "UNKNOWN" CATEGORY

ANALYTICAL AND CLUSTERING METHODS

K-MEANS CLUSTERING ALGORITHM

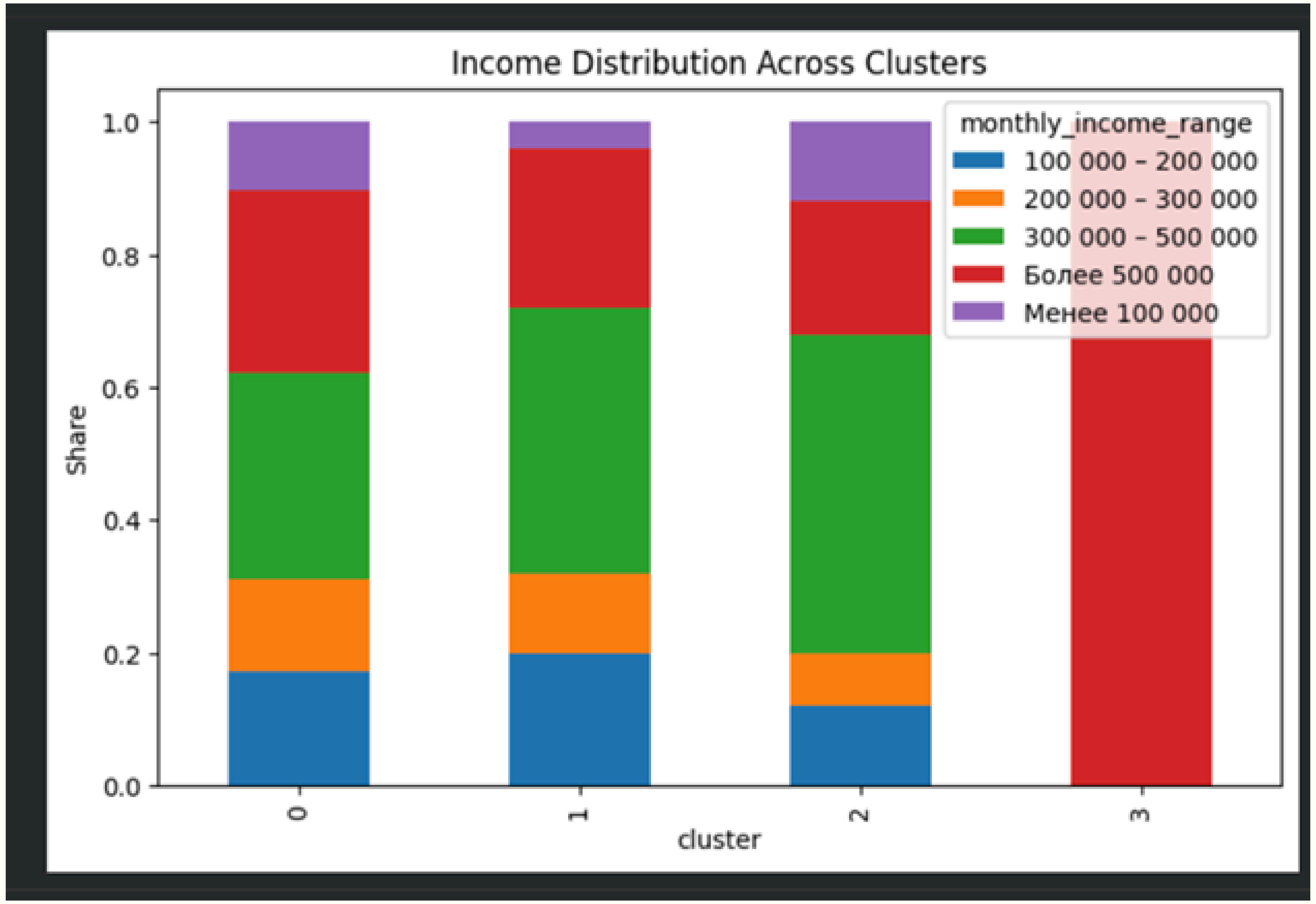


SILHOUETTE ANALYSIS

RESULTS OF CLUSTERING

The probability of saving money differs substantially across the identified clusters. In some clusters, saving behavior is nearly universal, whereas in others the majority of users do not save at all. This pronounced contrast indicates that saving behavior is not primarily determined by income level, but rather by the behavioral profile of the user, including spending discipline, impulsivity, and financial habits. As a result, the presence or absence of savings reflects underlying behavioral patterns rather than purely economic capacity.

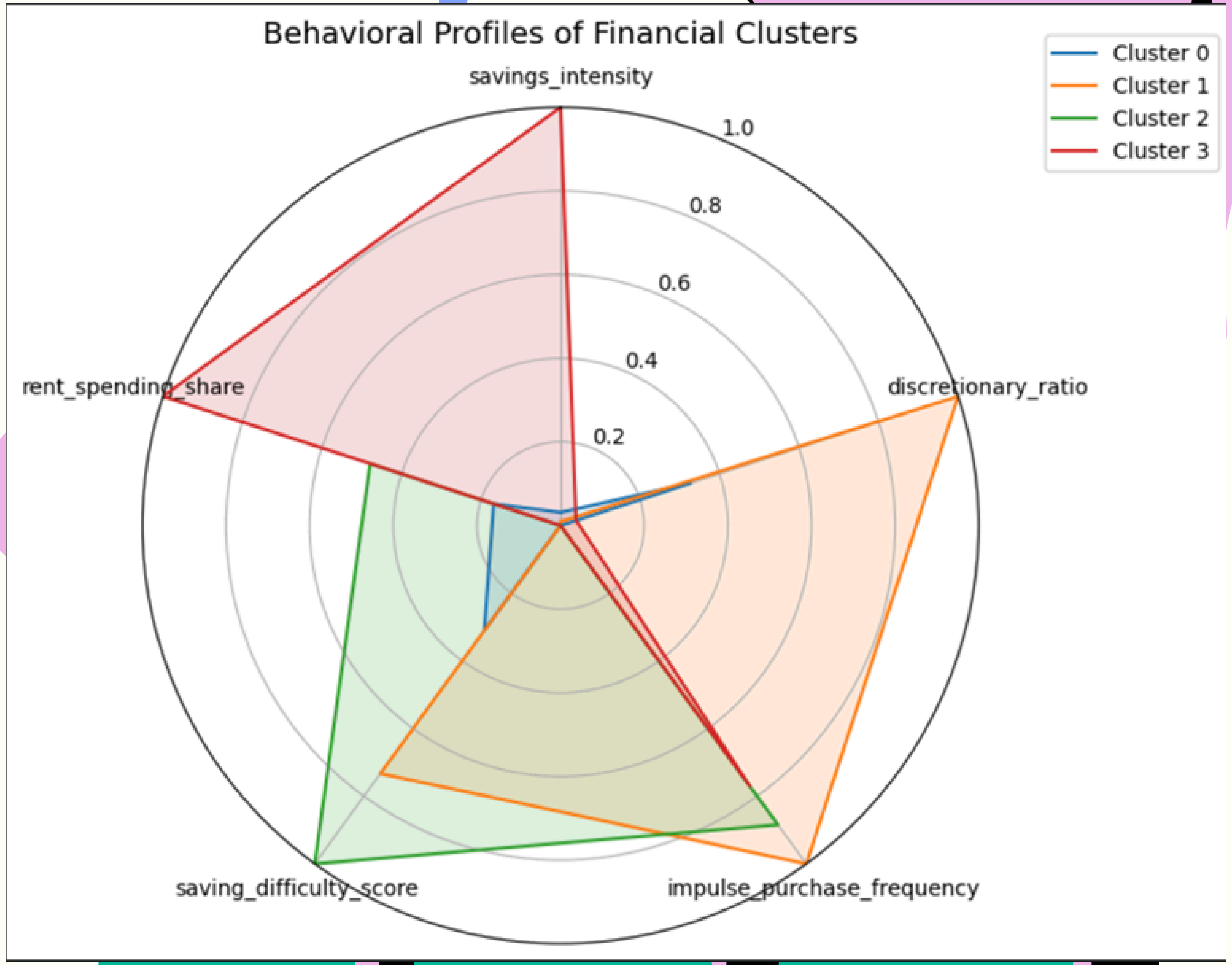
RESULTS OF CLUSTERING



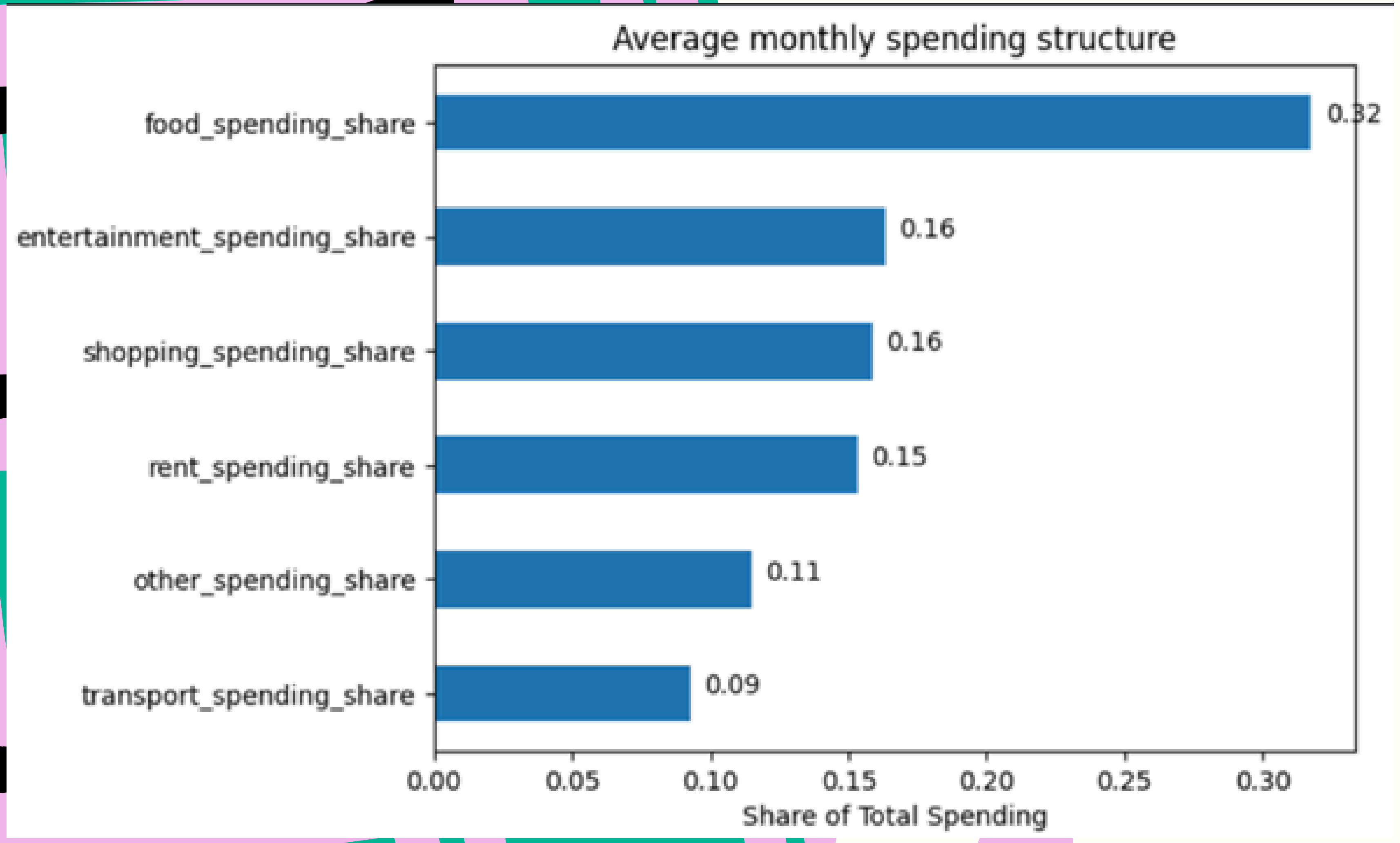
Despite clear behavioral differences between clusters, income levels are not directly aligned with cluster membership. Each cluster contains users from multiple income ranges, and no cluster corresponds exclusively to a specific income group. This finding suggests that segmentation based solely on income is insufficient for understanding financial behavior, as individuals with similar income levels may exhibit significantly different budgeting and saving patterns. Behavioral characteristics therefore provide a more informative basis for financial segmentation.

RESULTS OF CLUSTERING

Radar chart shows that each cluster has a distinct behavioral profile defined by a unique combination of financial indicators. Differences arise from discretionary spending, impulse purchases, perceived saving difficulty, savings intensity, and structural cost burden, highlighting the heterogeneity of budgeting strategies.
Cluster 0: balanced, moderate spending and saving difficulty, financially adaptive.
Cluster 1: high discretionary spending and impulsivity, moderate saving, impulsive consumers.
Cluster 2: low savings intensity, high saving difficulty, moderate impulsivity, struggle to save.
Cluster 3: high rent expenses, strong saving behavior, minimal impulsivity, disciplined despite constraints.
These clusters show that financial outcomes result from complex behavioral and structural interactions, not income alone.

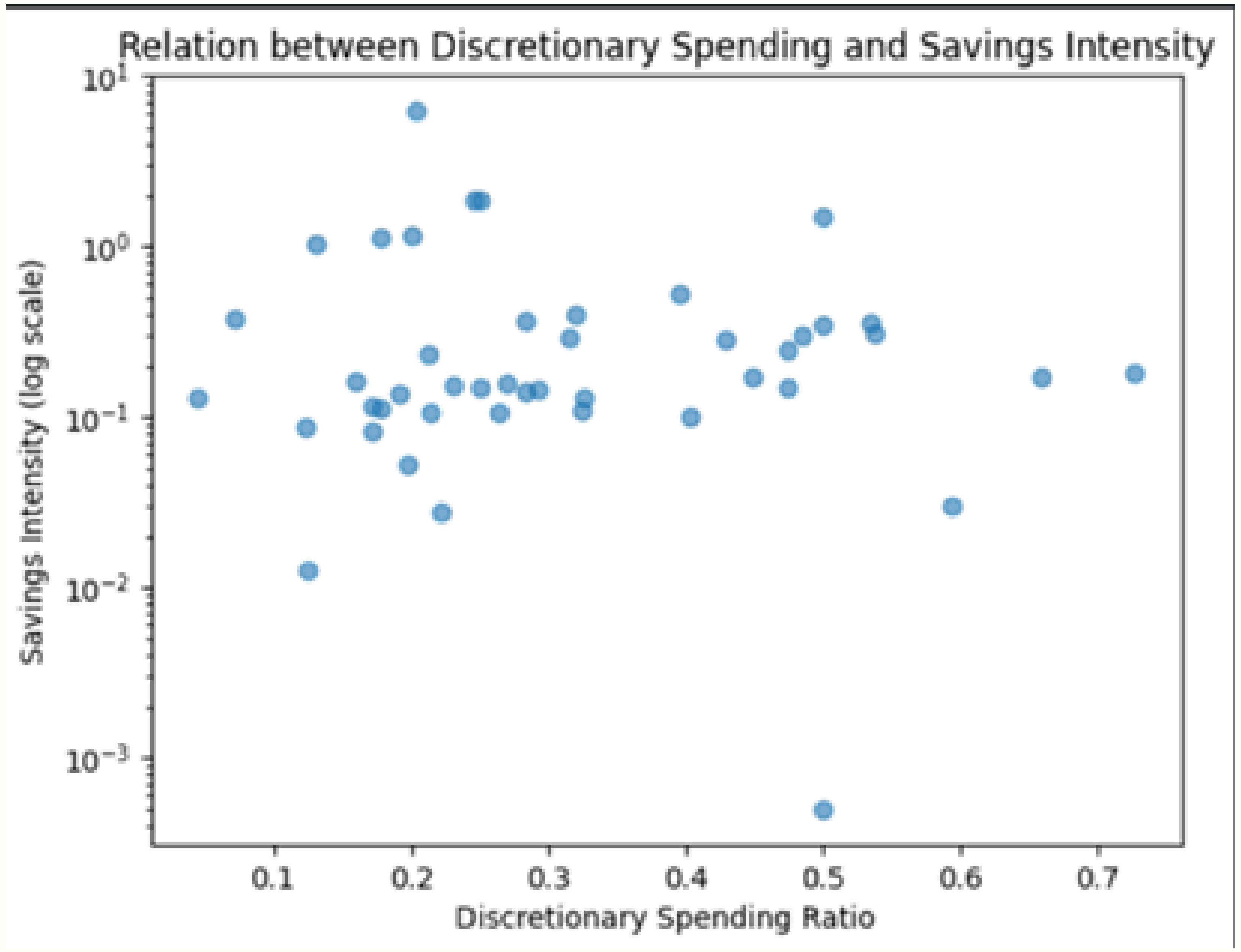


RESULTS OF VIZUALIZATIONS



The visualization of average monthly spending structure shows that food expenses constitute the largest share of total monthly spending, followed by entertainment and shopping categories. This distribution indicates that discretionary expenses account for a substantial proportion of household budgets. Such a structure suggests that non-essential spending categories represent potential leverage points for savings-oriented interventions, as even small behavioral adjustments in these areas could lead to meaningful improvements in savings outcomes.

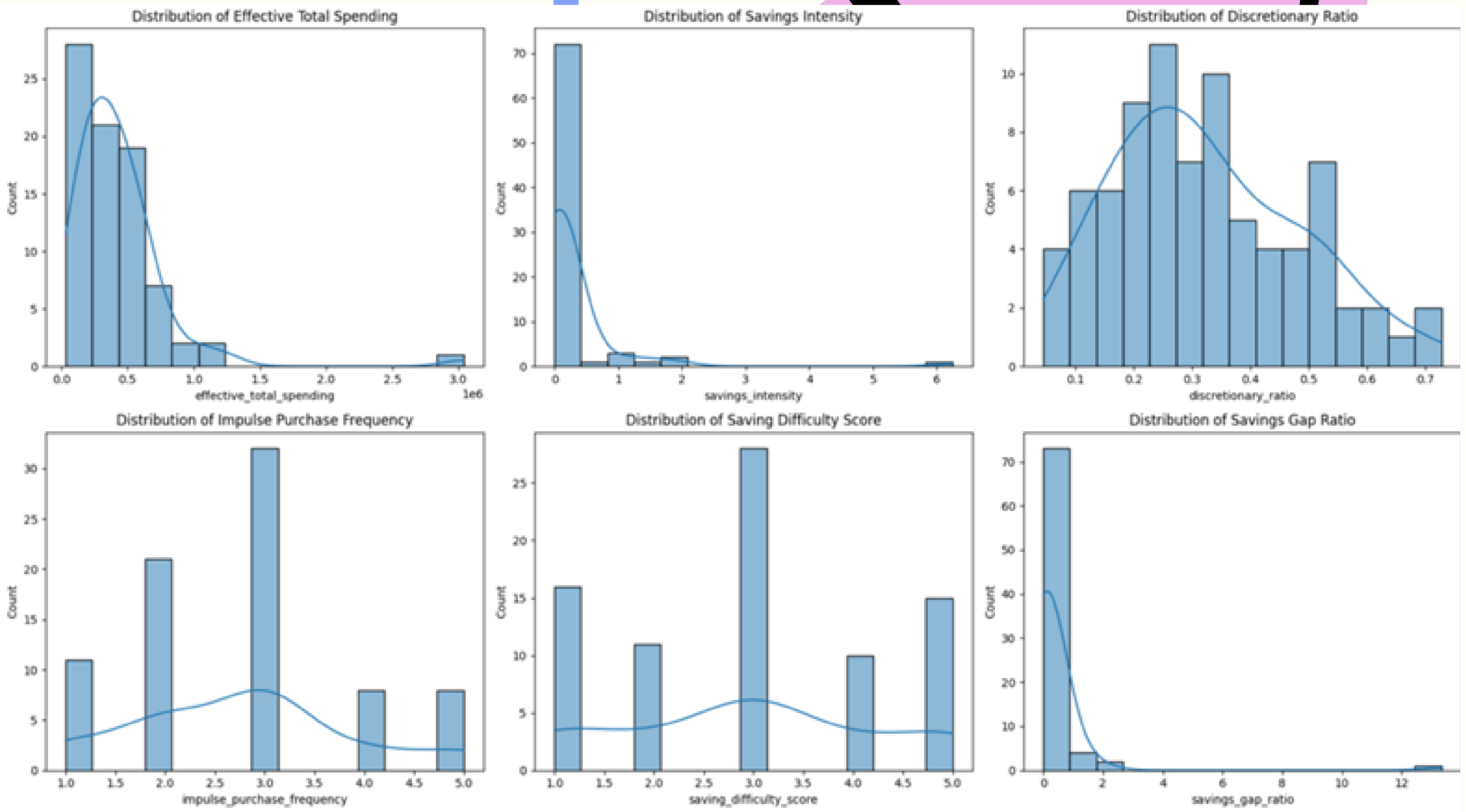
RESULTS OF VIZUALIZATIONS



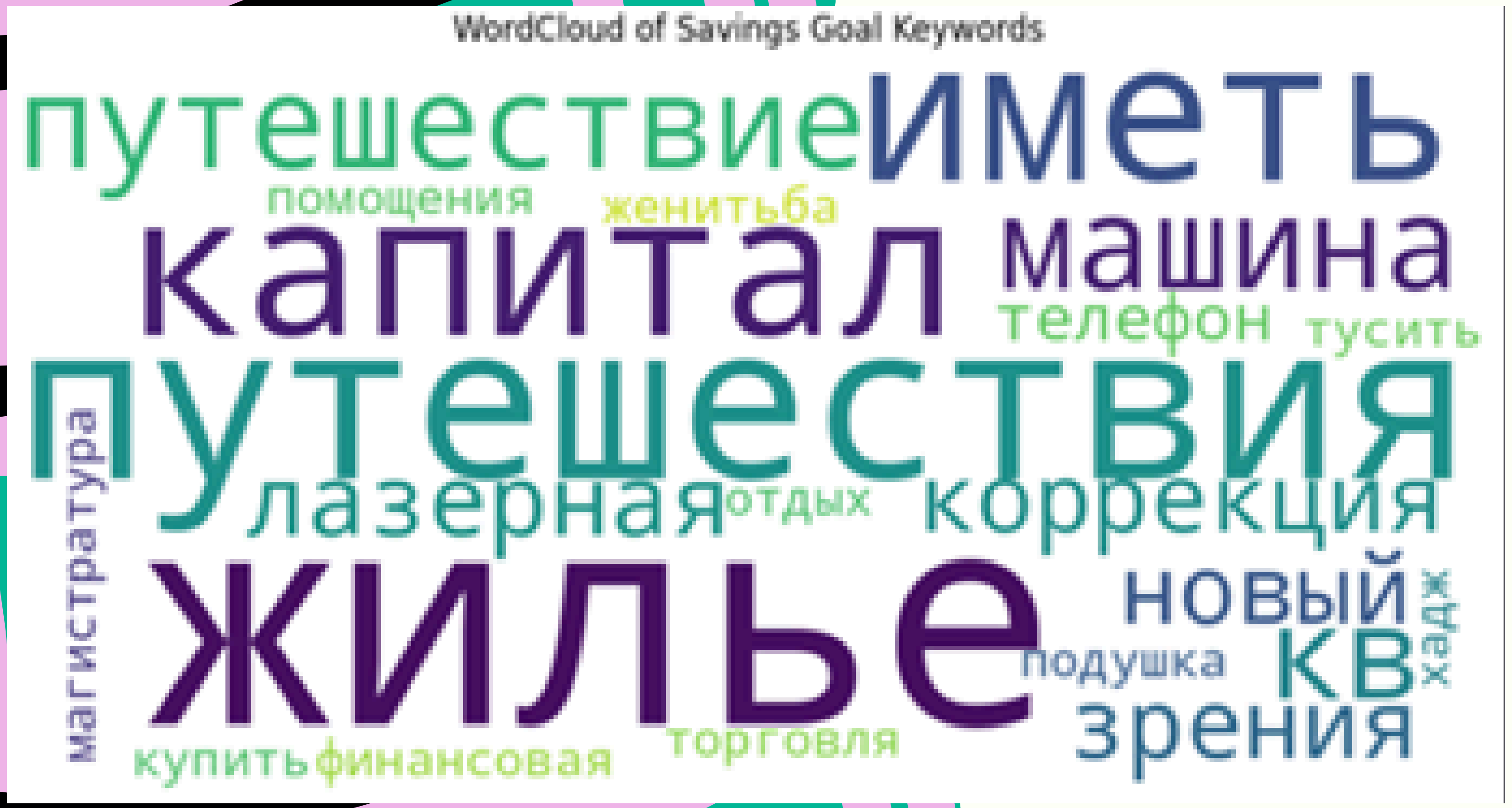
The scatter plot examining the relationship between discretionary spending ratio and savings intensity reveals a weak inverse association. Individuals with higher proportions of discretionary expenses tend to exhibit lower savings efficiency. However, the presence of several outliers indicates significant heterogeneity in financial behavior, suggesting that discretionary spending alone does not fully determine saving outcomes and must be interpreted alongside behavioral discipline and structural constraint.

RESULTS OF VIZUALIZATIONS

Histogram shows that effective total spending is strongly right-skewed, with most users reporting moderate expenses and a few exhibiting very high spending. Savings intensity is also skewed, mostly low to moderate, with some extreme values reflecting strong discipline or anomalies. The discretionary spending ratio is fairly uniform (0.2–0.4), highlighting differences in non-essential spending. Impulse purchase frequency peaks at moderate values, showing variability in impulsive behavior. Saving difficulty score is near-symmetric, centered on moderate difficulty. The savings gap ratio is highly right-skewed, with most users having small gaps but a minority showing very large gaps, useful for detecting misaligned saving strategies.



RESULTS OF VIZUALIZATIONS



The word cloud generated from open-ended survey responses highlights the most frequently mentioned savings goals, which are primarily related to housing, travel, vehicle purchases, capital accumulation, education, and health. These findings suggest that users' financial objectives are strongly life-oriented and practical rather than abstract monetary targets. This reinforces the real-world relevance of the dataset and enhances the practical value of the project, as the identified goals align closely with common long-term financial planning priorities.

CONCLUSION

Key Findings:

- Financial behavior and outcomes are highly heterogeneous and cannot be fully explained by income alone.
- Behavioral factors—discretionary spending, impulsivity, perceived saving difficulty, and expense stability—strongly influence savings outcomes.
- Individuals with high impulsivity or high discretionary spending often struggle to save, regardless of income.
- Clustering analyses reveal distinct behavioral segments, each with unique spending and saving patterns.
- Visualizations highlight the interactions between behaviors, showing how patterns of spending and saving affect financial outcomes.

THANK YOU