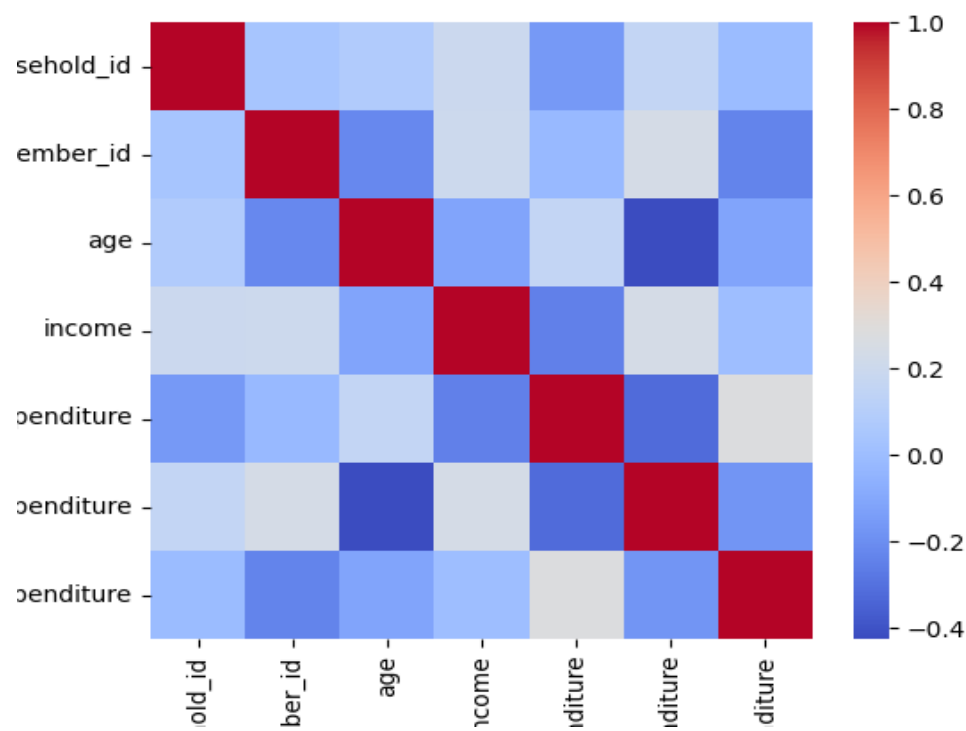


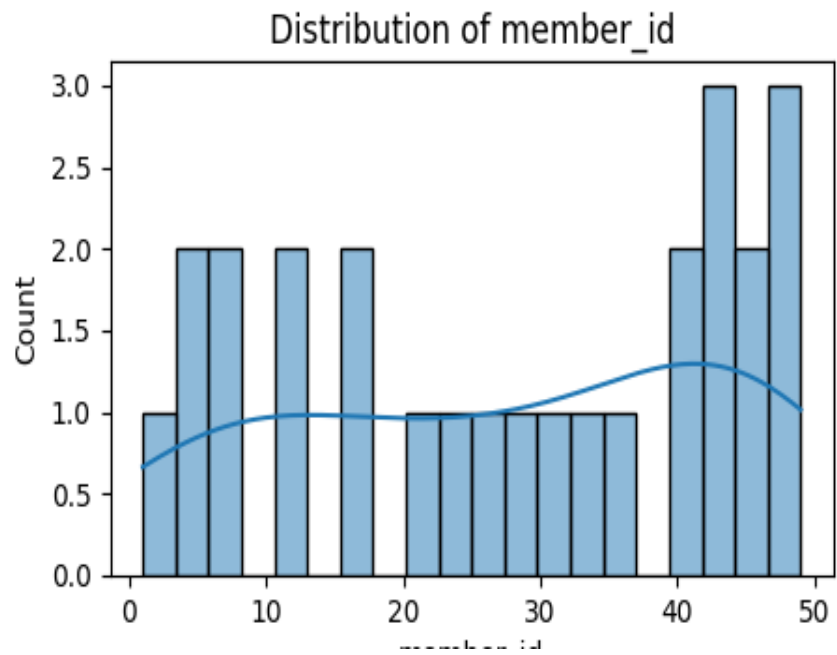
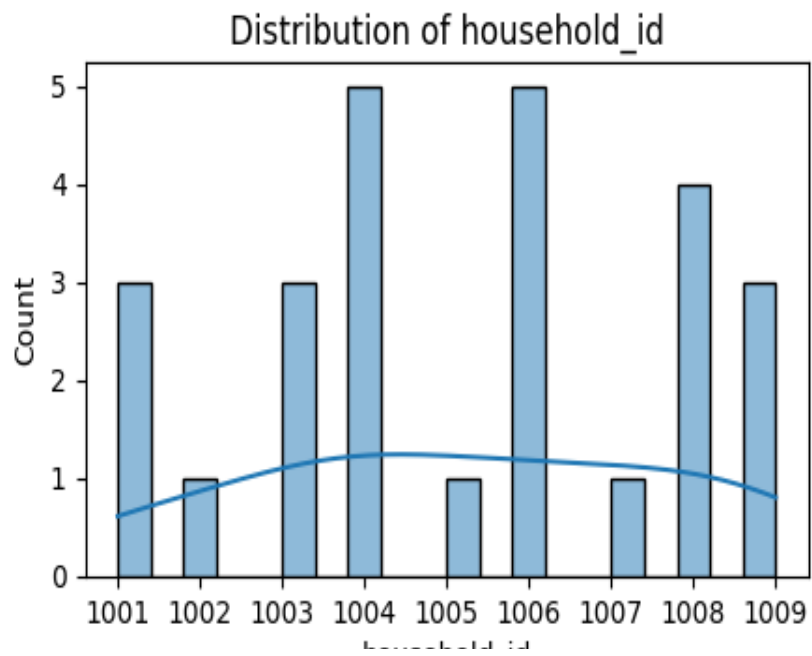
Data Preprocessing & EDA Report

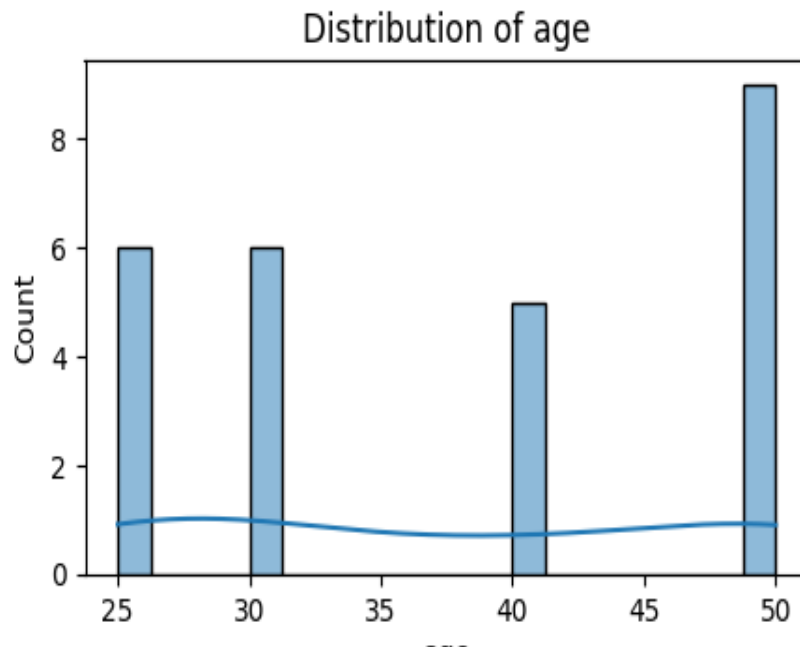
Generated on: 2025-08-20 18:17:35
Total Processing Steps: 10

Dataset Overview

Shape: 26 rows x 10 columns
Memory Usage: 5.40 KB







Processing Steps & Audit Trail

1. Uploaded CSV file 'sample_data.csv'
2. Schema mapped using 'household' template
3. Schema mapped using 'household' template
4. Data validation completed. Removed 24 invalid rows
5. Exploratory Data Analysis (EDA) report generated
6. Exploratory Data Analysis (EDA) report generated
7. Exploratory Data Analysis (EDA) report generated
8. Missing value imputation applied using median method
9. Outlier detection completed using zscore method
10. LLM-based report generated and converted to PDF

AI-Generated Summary

The dataset contains information on 26 households, with 10 variables describing each household member. The variables include identifying information (household_id, member_id), demographic data (age, gender, education, occupation), and economic data (income, expenditure, food_expenditure, housing_expenditure). The data likely represents a sample of households and their members, providing insights into household demographics and spending habits. Further analysis is needed to understand the distributions and relationships between variables.

Recommendations & Next Steps

- Review column distributions for anomalies
- Consider imputing missing values or dropping high-null columns

- Apply outlier detection methods on numeric features
- Perform feature engineering for categorical variables
- Document transformations for reproducibility