Data analysis on Family Income & Expenditure Analysis (Philippines)

Overview

This repository contains a data analysis project exploring the 'Filipino Family Income and Expenditure' dataset published by Francis Paul Flores (sourced from Kaggle). The goal is to extract meaningful insights, visualize trends, and derive actionable conclusions about how income and expenditure patterns vary across different socio-economic groups.

This set aims to explore and interpret patterns in household income, expenditures, and demographics. Use graphs and DataFrame operations to answer questions and uncover insights by answering the following questions:

1. **Urban vs Rural Differences**
    a. How do household incomes differ between urban and rural households? Compare NCR and Region VIII
2. **Impact of Education on Income**
    a. Do household heads with higher education levels earn more?
3. **Household Size and Spending**
    a. Do larger families spend more on food or education?
4. **Gender Differences in Household Economics**
    a. Is there a difference in income or spending patterns based on the sex of the household head?
5. **Regional Spending Patterns**
    a. Which regions spend the most on education, transportation, and food?
6. **Other Explorations**
    a. Which region has the highest Average Monthly Income per Family and Region?
    b. Examine the relationship of the number of family members and income, savings, expenditures

Insights will be summarized into bullet points.

Why this matters

- Real-world relevance: Household income \& spending are central to economics, public policy, and social programs.
- Transferable skills: The methods used (data cleaning, filtering, visualization, reporting) are directly applicable to business reporting, insights roles, and data science.
- Storytelling with data: Beyond numbers, this project shows how to communicate findings in a clear, effective way — crucial when working with non-technical stakeholders.

Tools & Technologies Used

- Python — core programming language
- pandas, numpy — data manipulation \& aggregation
- matplotlib, seaborn — for visualizations
- Google Colab Notebooks — for exploration and prototyping
- Version control via Git / GitHub

Urban vs Rural Differences

How do household incomes differ between urban and rural households? Compare NCR and Region VIII.

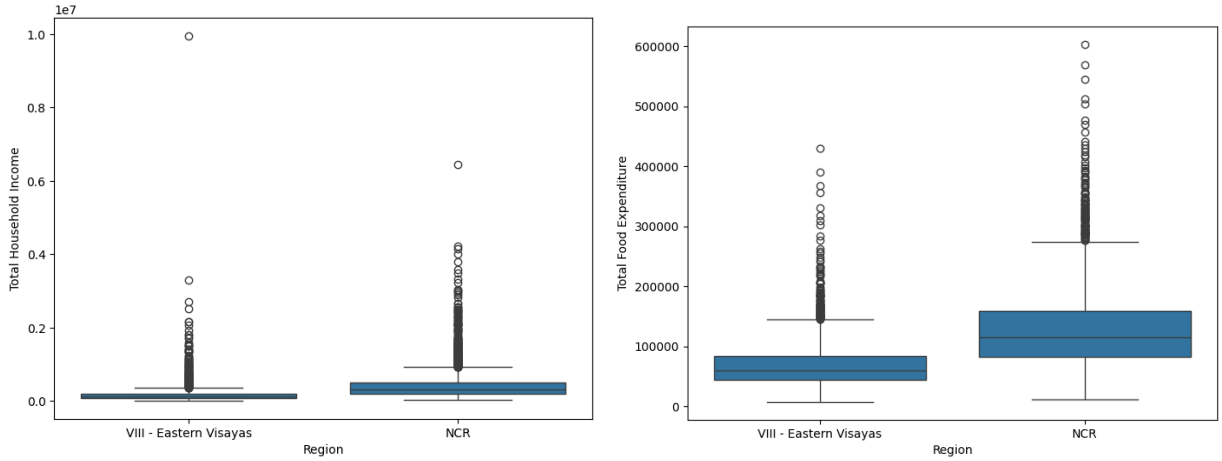


Figure 1: Boxplot comparing income Region VIII and NCR households based on income (with outliers or values over 500,000 PHP) & Figure 2: Boxplot comparing income Region VIII and NCR households based on income (with outliers or values less than 500,000 PHP)

Main Insights:

- NCR has a higher mean income than Region VIII.
- Most outliers above the mean average income exist for NCR households but one household in Region VIII earns the most among all surveyed households.

Impact of Education on Income

Do household heads with higher education levels earn more? (In the context of Preschool, Elementary, High School, University/ College, and Post Baccalaureate graduates)

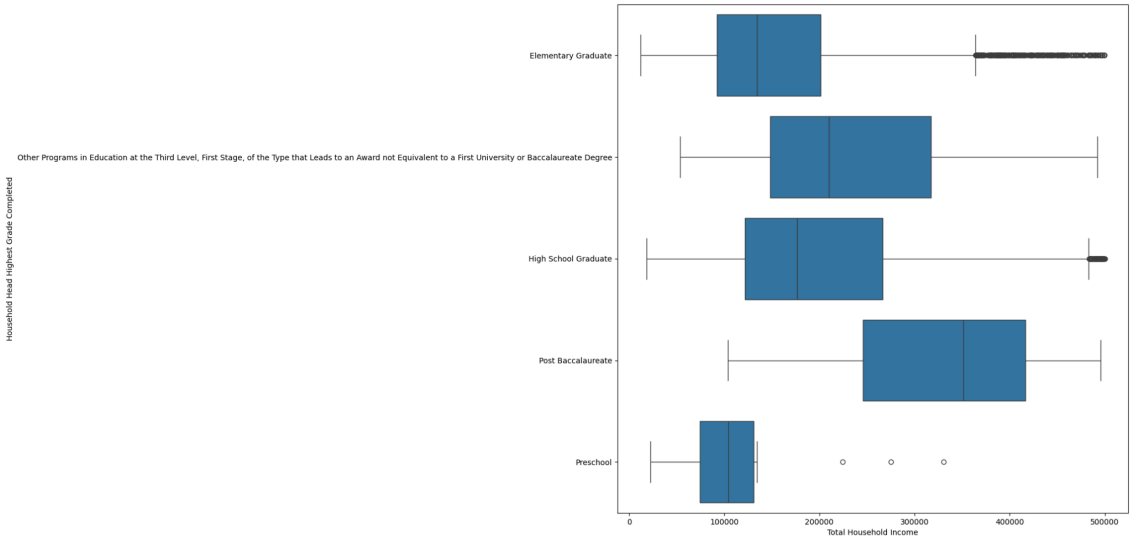


Figure 3: Boxplot comparing income of preschool, elementary, high school, university/ college, and post-baccalaureate graduates (with incomes above 500,000 PHP)

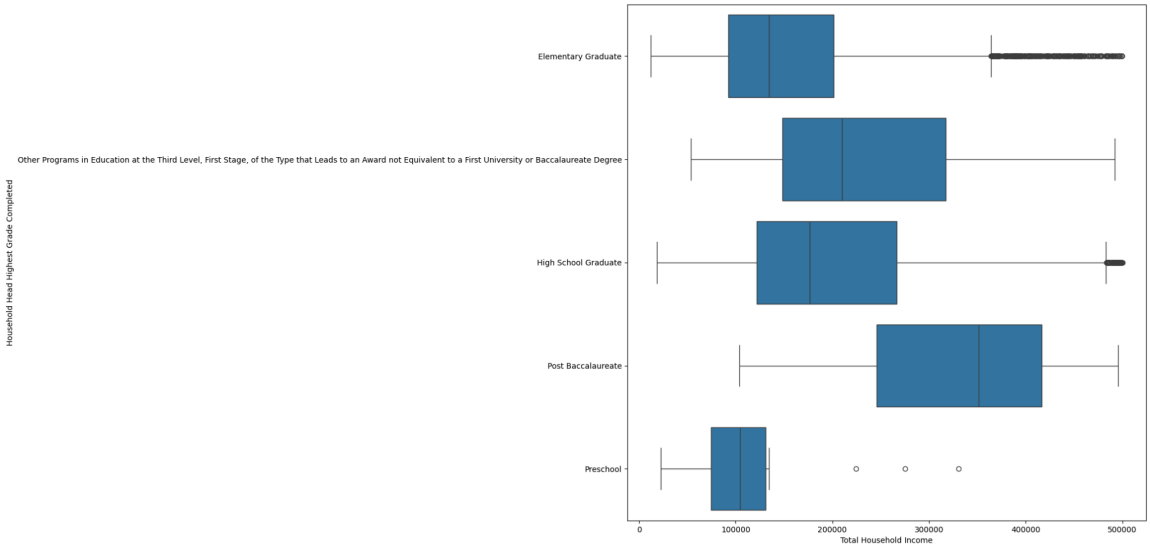


Figure 4: Boxplot comparing income of preschool, elementary, high school, university/ college, and post-baccalaureate graduates (with incomes below 500,000 PHP)

Main Insights:

- Post Baccalaureate graduates have a higher mean income
- A great amount of high school graduates and elementary school graduates are outliers, earning between 5-digit to 6-digit incomes
- Preschool graduates have the lowest mean income

Do household heads with higher education levels earn more? (In the context of STEM Programs)

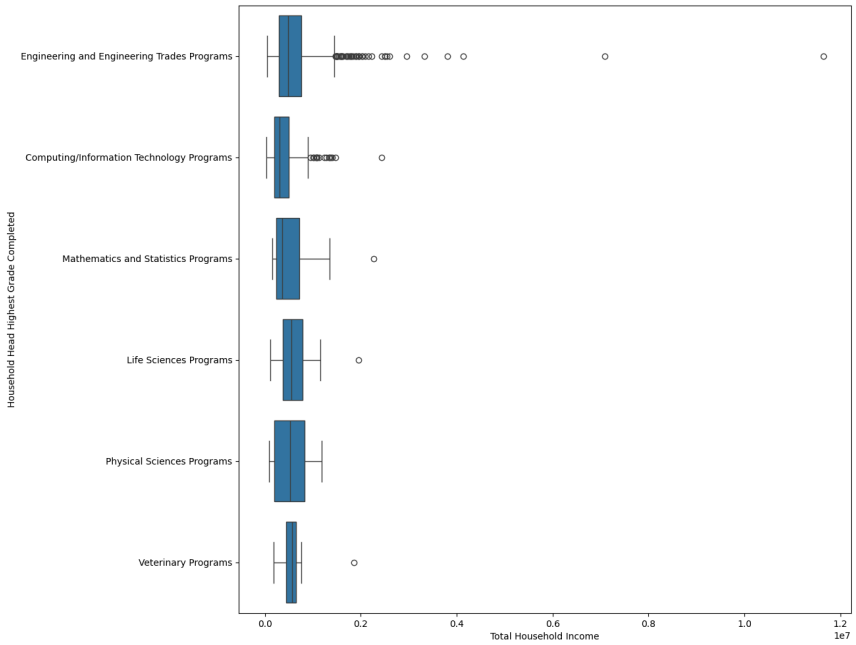


Figure 5: Boxplot comparing income of STEM program graduates (with incomes more than 500,000 PHP)

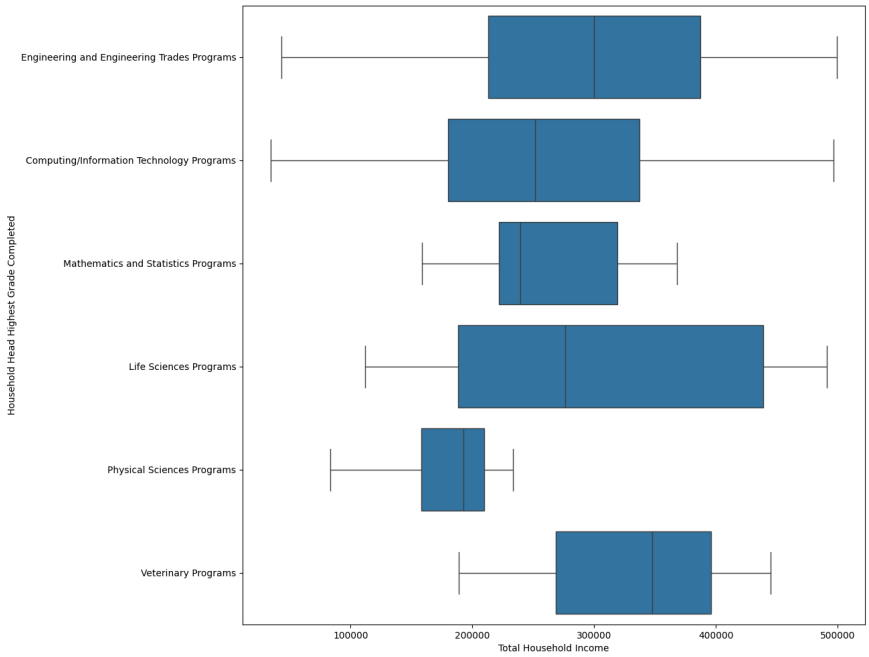


Figure 6: Boxplot comparing income of STEM program graduates (with incomes less than 500,000 PHP)

Main Insights:

- Veterinary Programs have the highest mean income, followed by Engineering and Engineering Trades Programs, and Life Sciences Programs
- Engineering and Engineering Trades Programs have the most outliers and have the capacity to earn the most

Household Size and Spending

Do larger families spend more on food or education?

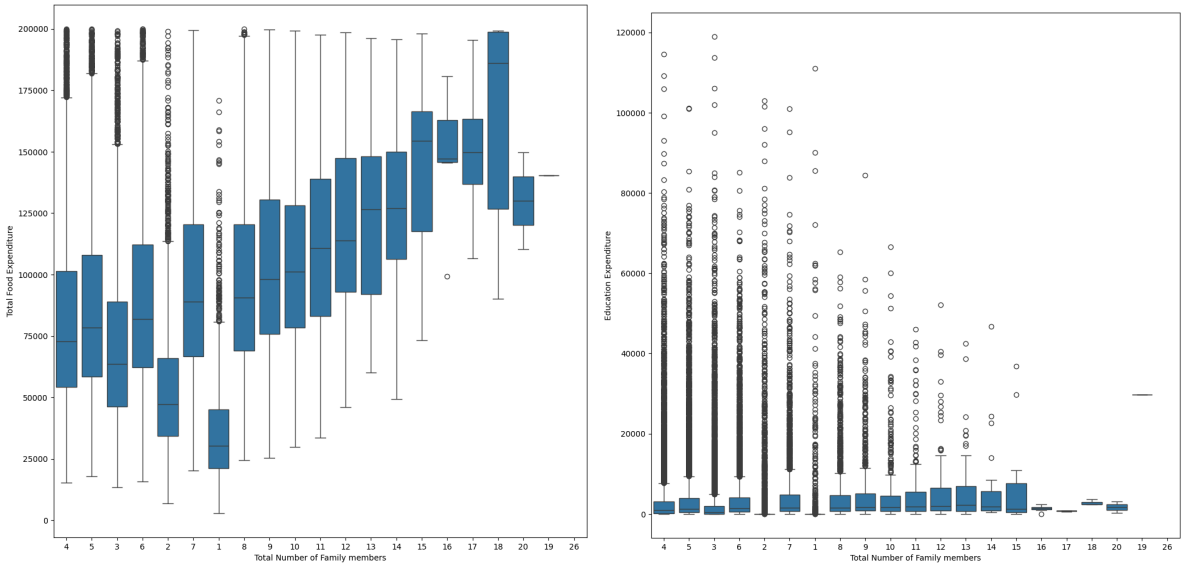


Figure 7: Boxplot comparing food expenditure per household size & Figure 8: Boxplot comparing education expenditure per household size

Main Insights:

- Bigger households spend more than smaller households on food
- Smaller households have more outliers than spend on education compared to bigger households

Gender Differences in Household Economics

Is there a difference in income or spending patterns based on the sex of the household head?

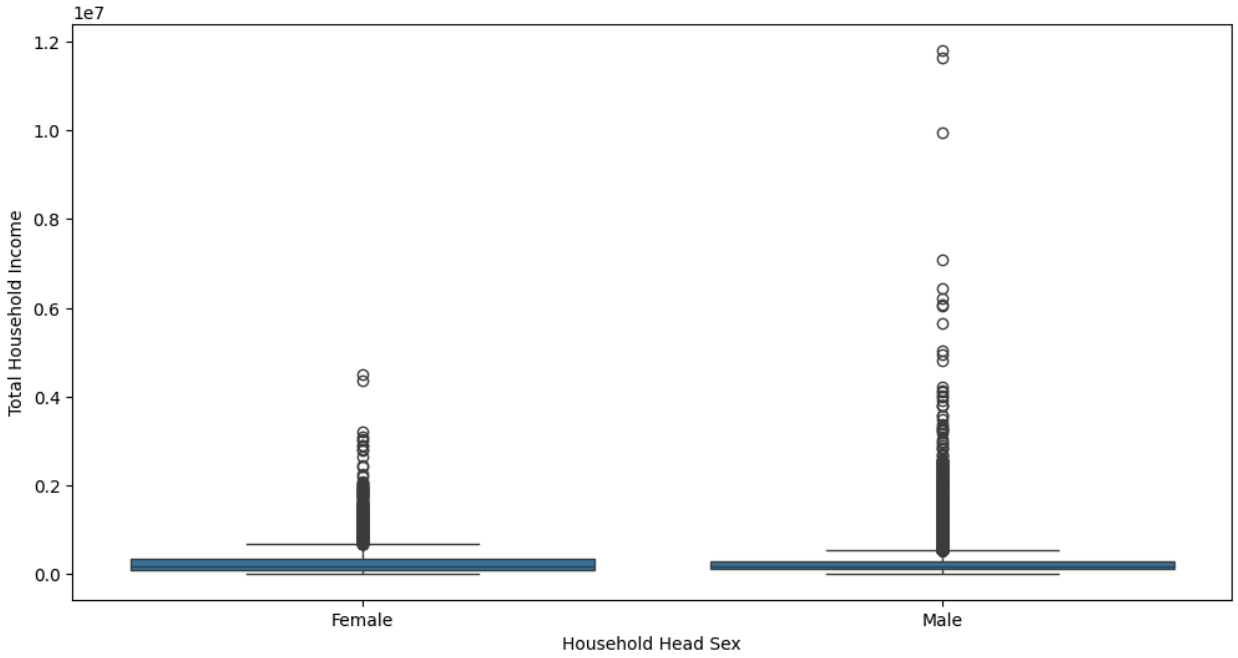


Figure 8: Boxplot comparing income per gender with income more than 500,000 PHP

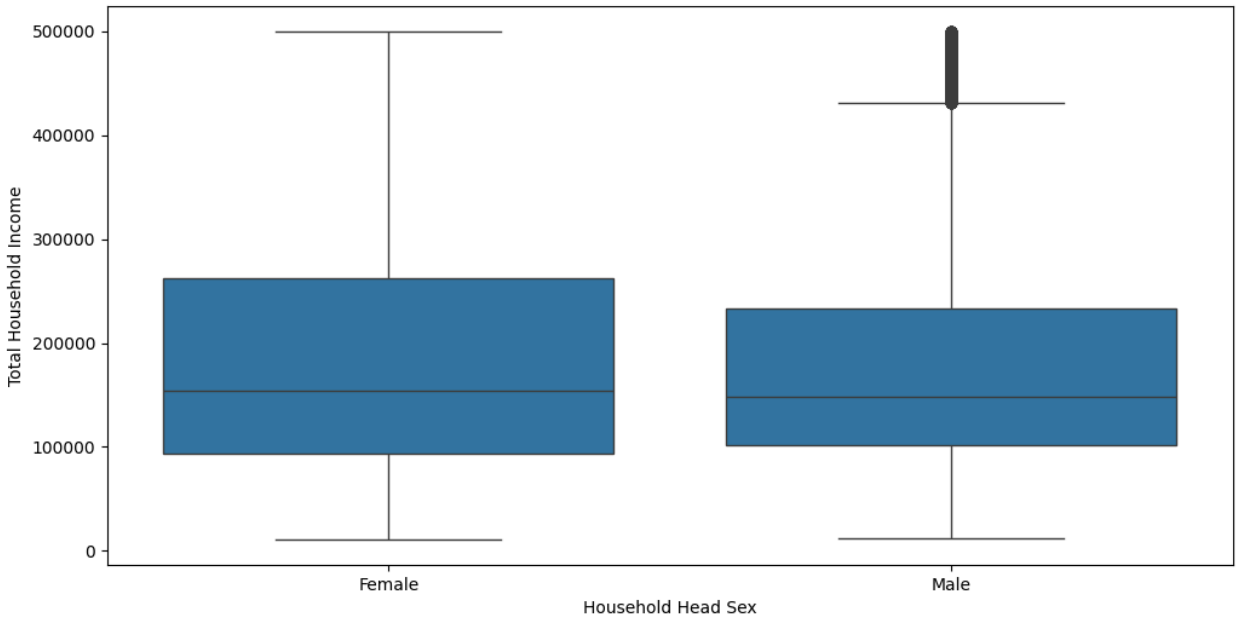


Figure 9: Boxplot comparing income per gender with income less than 500,000 PHP

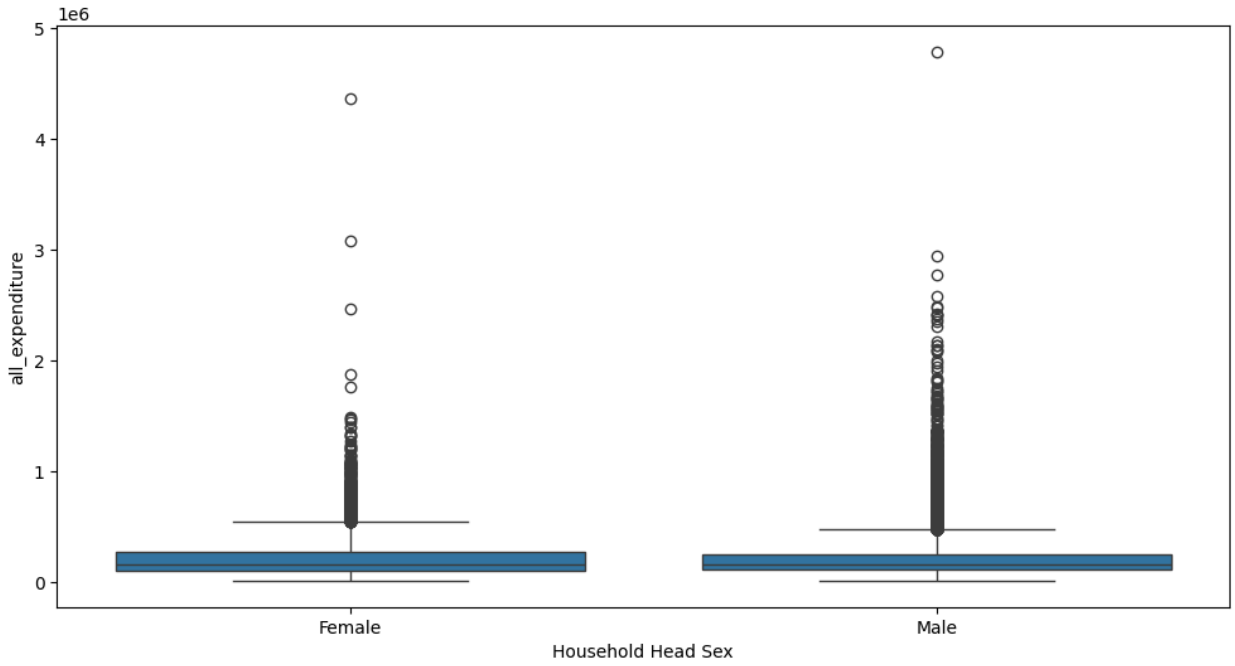


Figure 10: Boxplot comparing expenditures per gender with expenditures cost more than 300,000 PHP

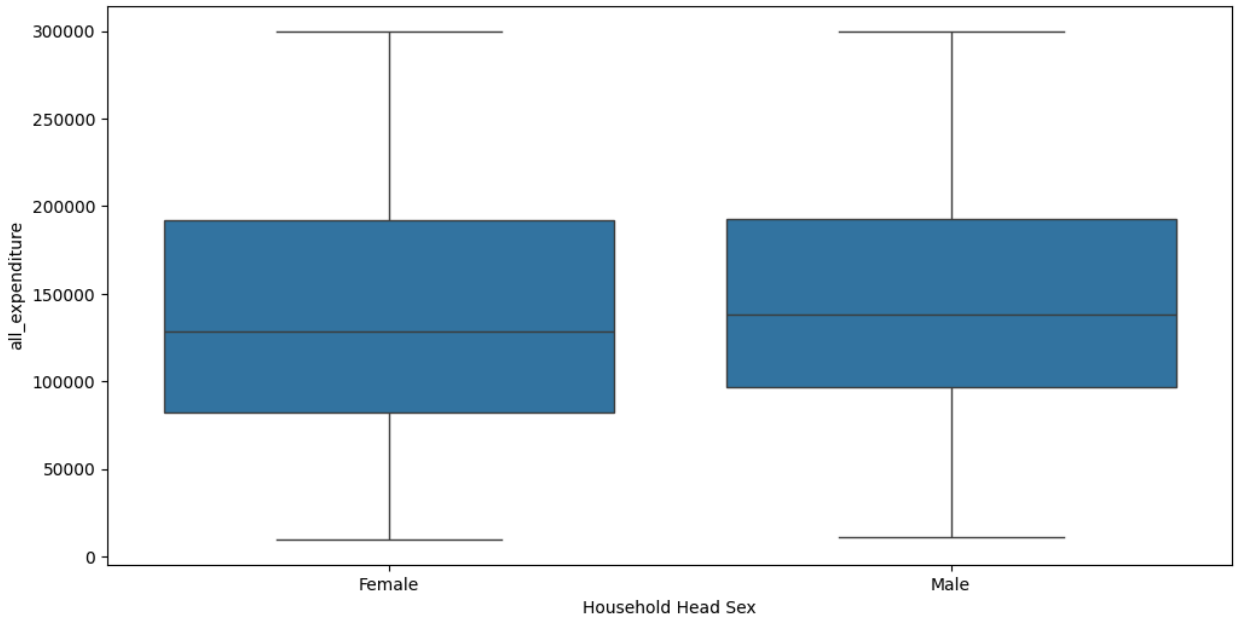


Figure 11: Boxplot comparing expenditures per gender with expenditures cost less than 300,000 PHP

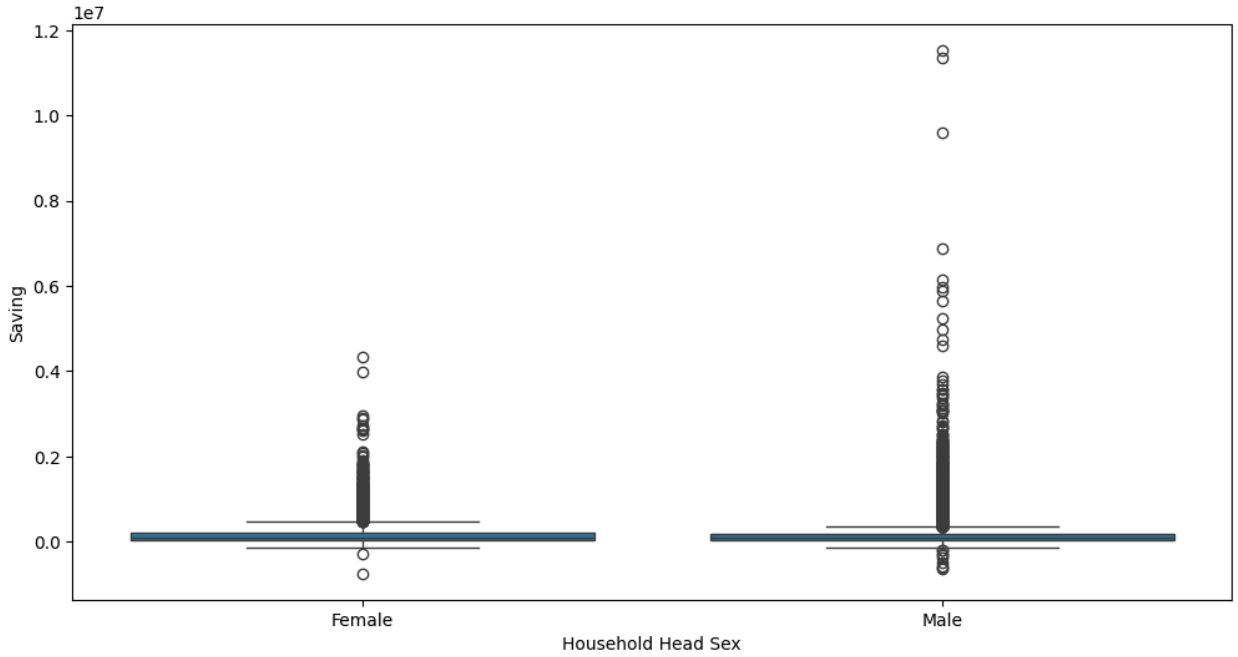


Figure 12: Boxplot comparing income per gender with savings more than 200,000 PHP

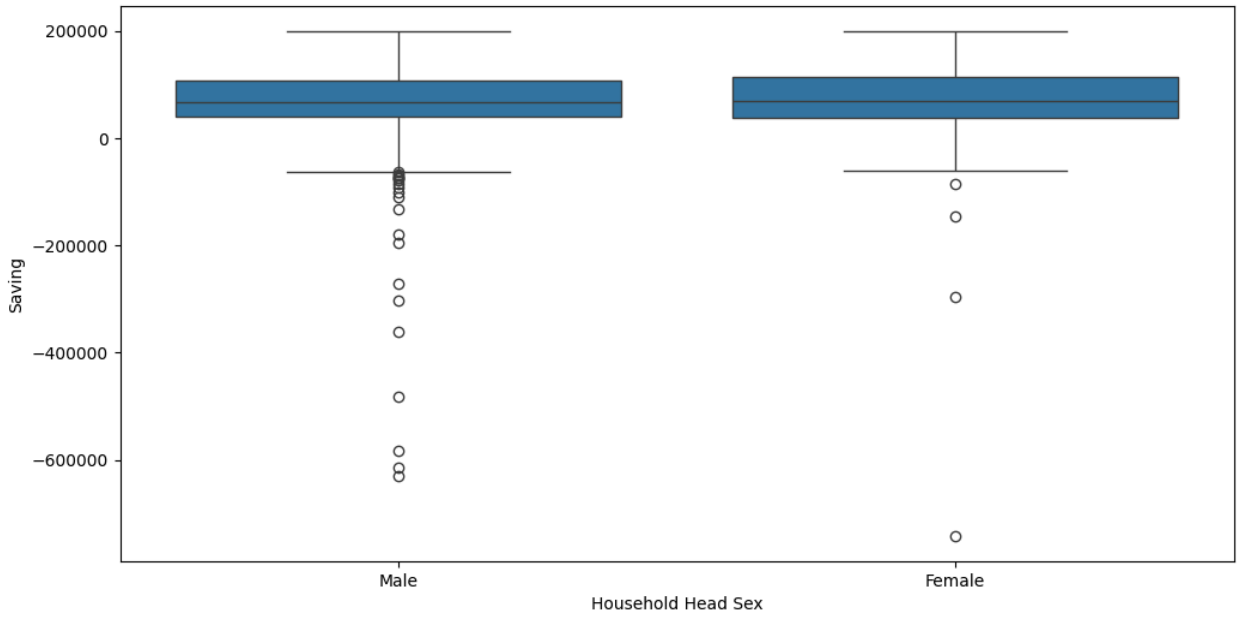


Figure 13: Boxplot comparing income per gender with savings less than 200,000 PHP

Main Insights:

- Women have a slightly higher mean income than men
- Men have more high outliers when it comes to income
- Men have a slightly higher mean expenditure than women
- Men have more high outliers when it comes to expenditures
- Women have a slightly higher mean savings than men
- Men have more negative outliers when it comes to savings than women

Regional Spending Patterns

Which regions spend the most on education, transportation, and food?

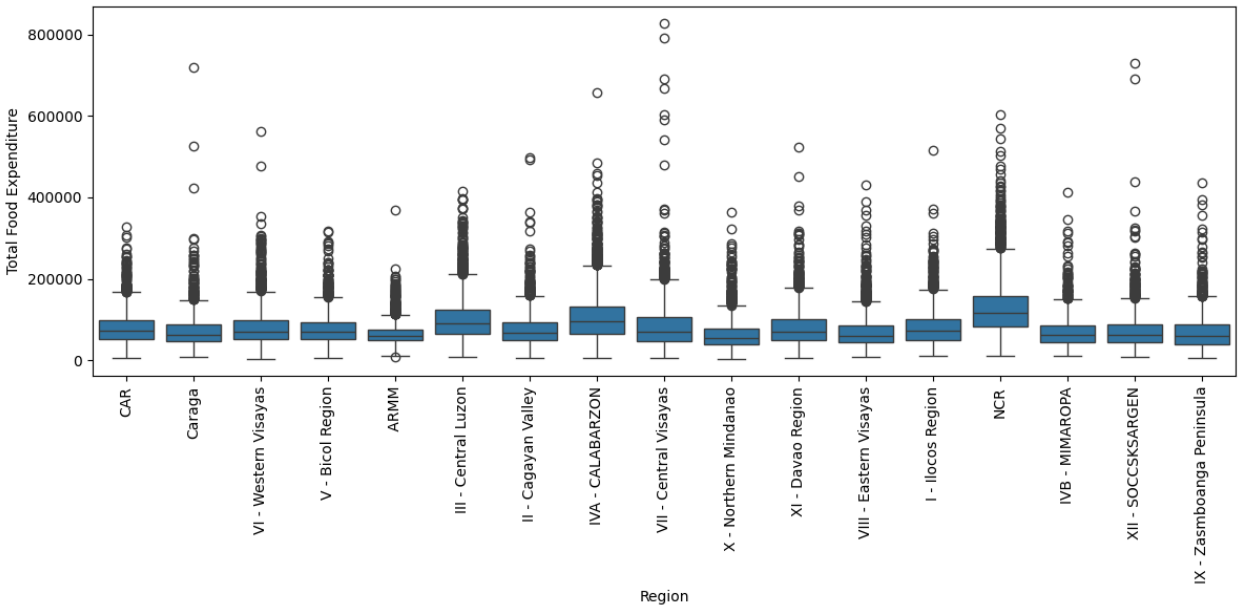


Figure 14: Boxplot comparing food expenditure per region

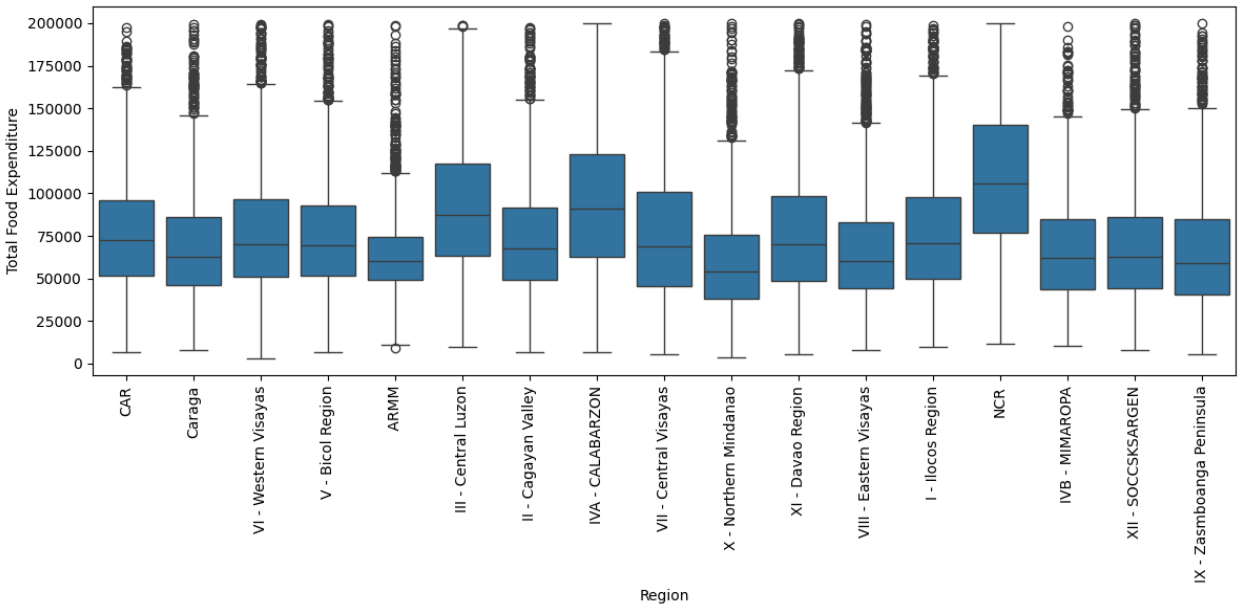


Figure 15: Boxplot comparing food expenditure per region (less than 200,000 PHP)

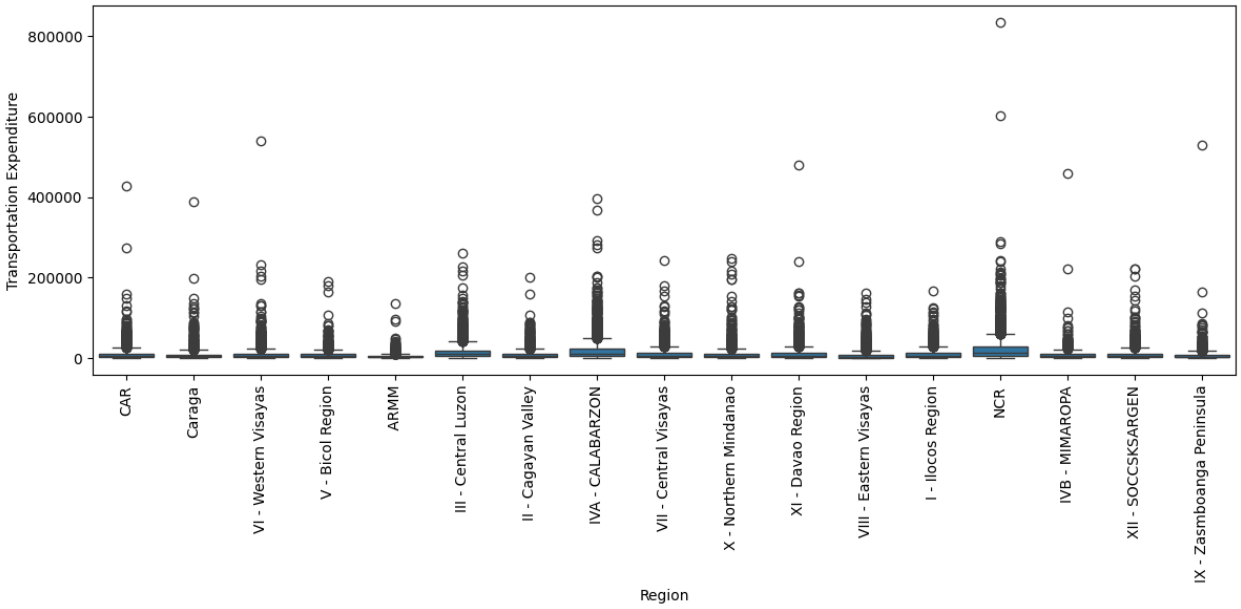


Figure 16: Boxplot comparing transportation expenditure per region

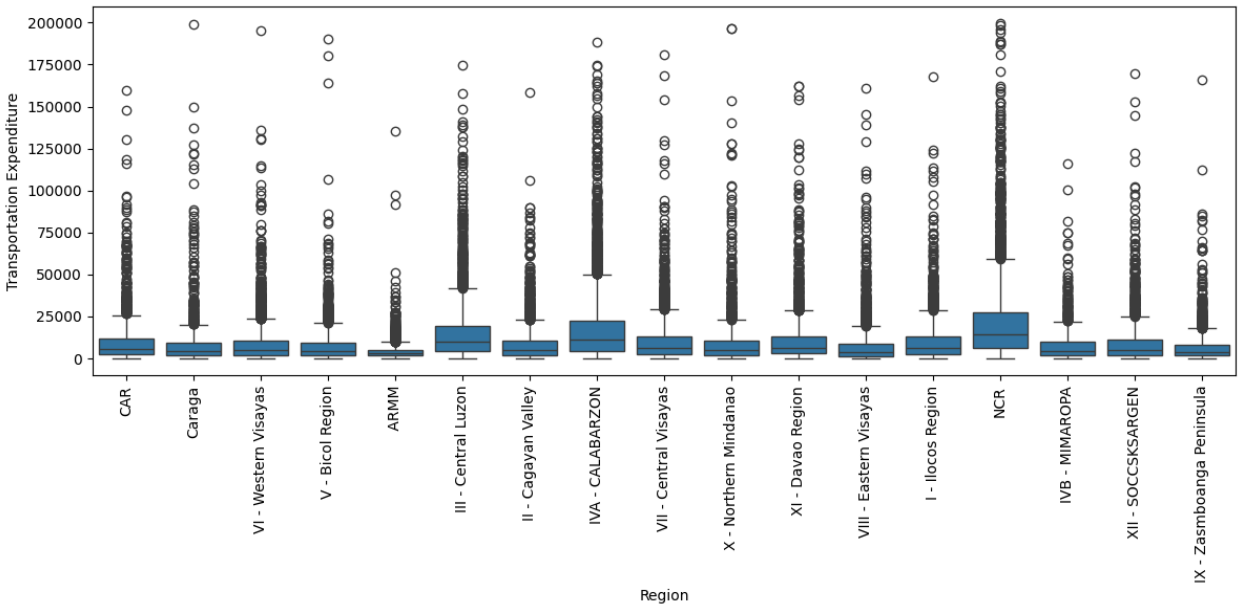


Figure 17: Boxplot comparing transportation expenditure per region (less than 200,000 PHP)

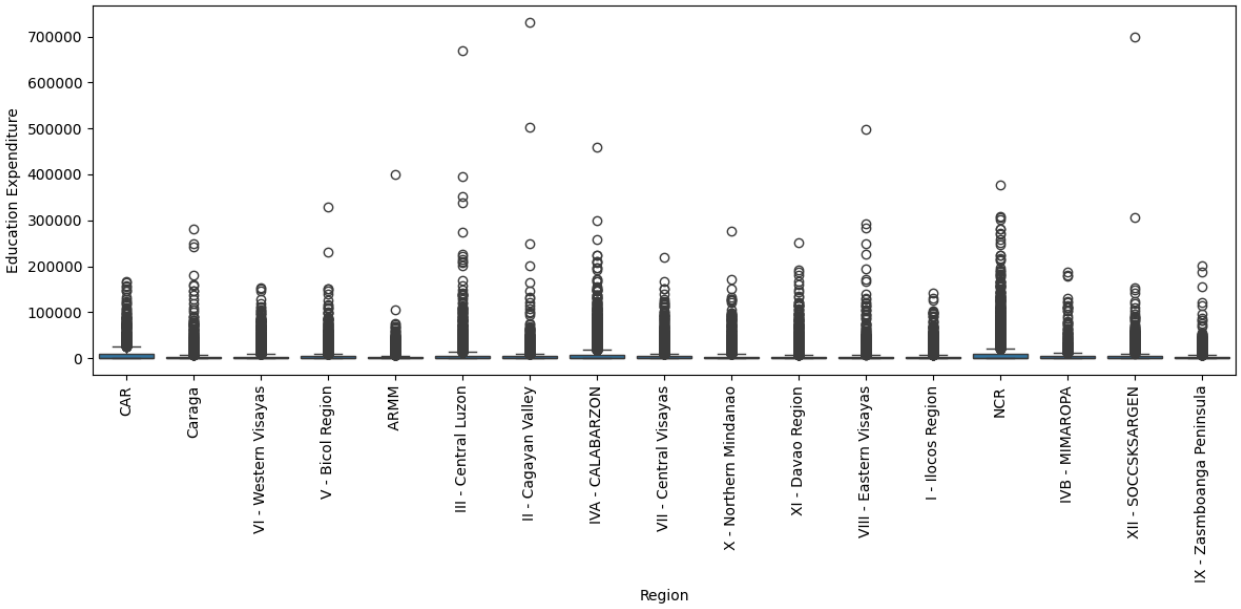


Figure 18: Boxplot comparing education expenditure per region

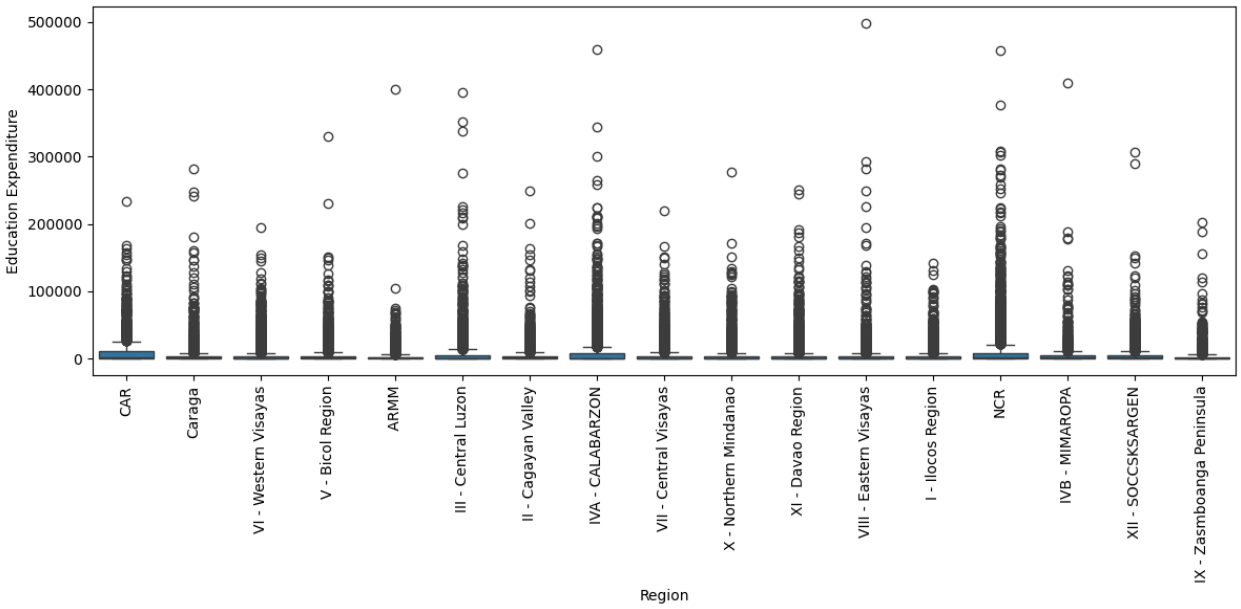


Figure 19: Boxplot comparing education expenditure per region (less than 500,000 PHP)

Main Insights:

- The top 3 regions that have the highest food expenditures are NCR, CALABARZON, and Central Luzon.
- Region VII has the highest amount of outliers when it comes to food expenditures
- The top 3 regions that have the highest transportation expenditures are NCR, CALABARZON, and Central Luzon.
- NCR has the highest amount of outliers when it comes to transportation expenditures
- The top 3 regions that have the highest transportation expenditures are NCR, CALABARZON, and CAR
- NCR has the highest amount of outliers when it comes to transportation expenditures
- The regions with the highest outliers for education expenditures are Eastern Visayas, CALABARZON, and NCR

Other Explorations

Which region has the highest Average Monthly Income per Family and Region?

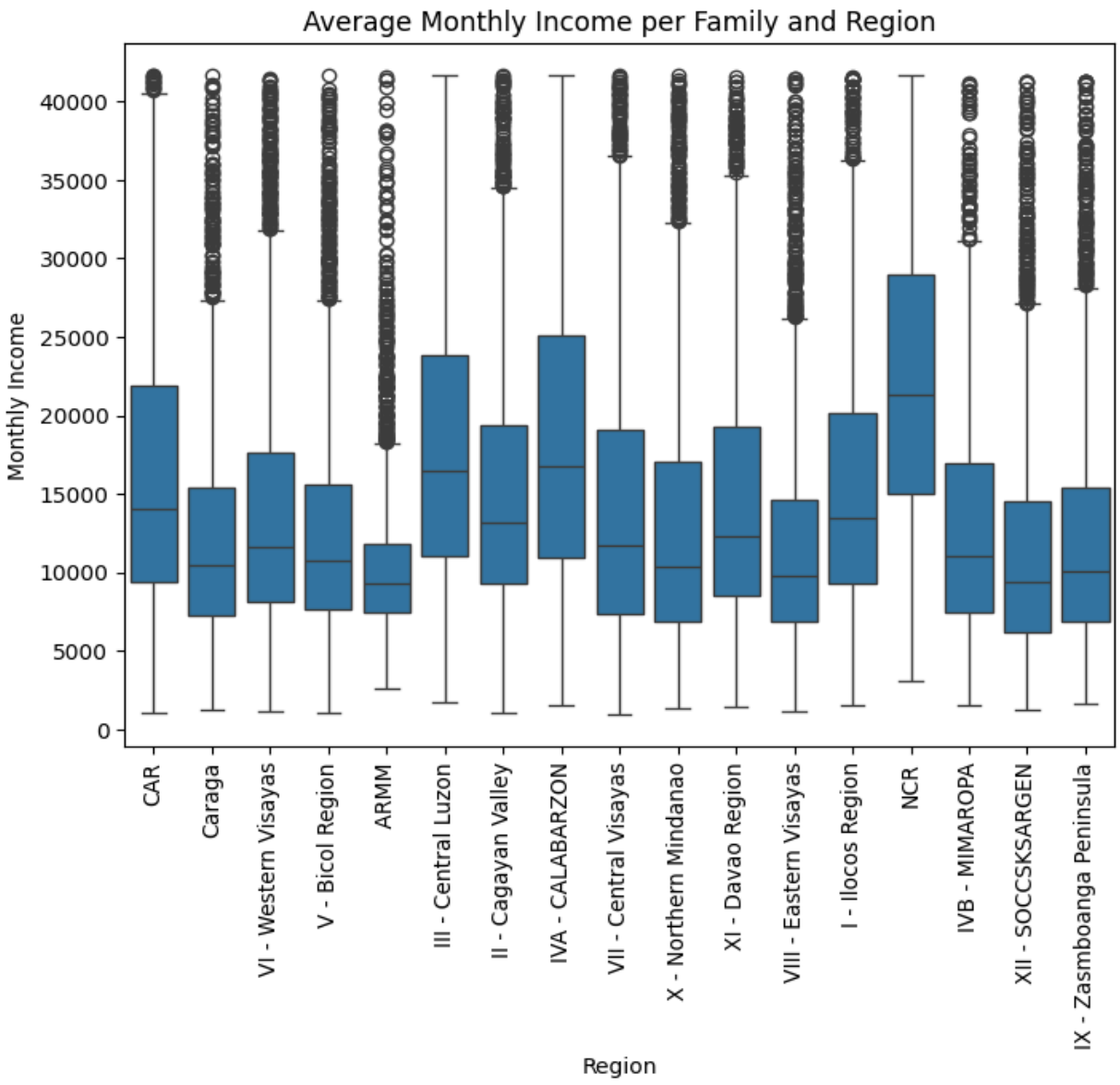
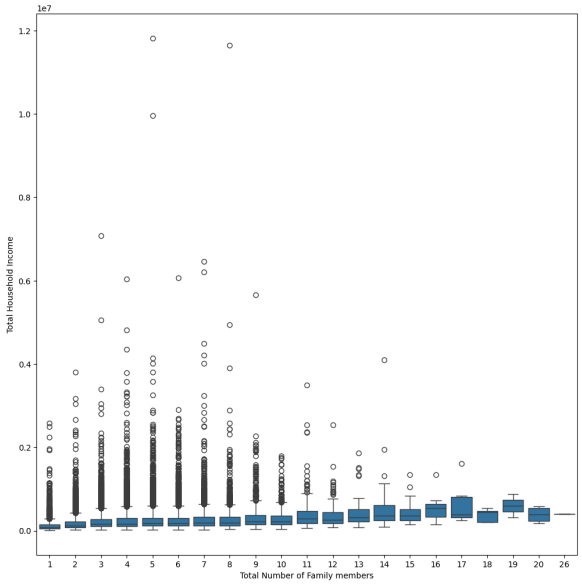


Figure 20: Boxplot comparing the average monthly income (less than 40,000 PHP)

Main Insights:

- The top 3 regions that have the mean income are NCR, CALABARZON, and Central Luzon. Outlets outside 40,000 PHP exist for these regions.

Examine the relationship of the number of family members and income, savings, expenditures



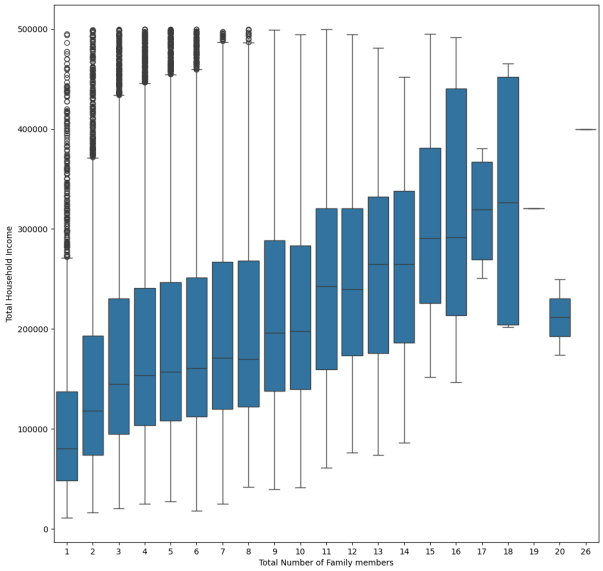
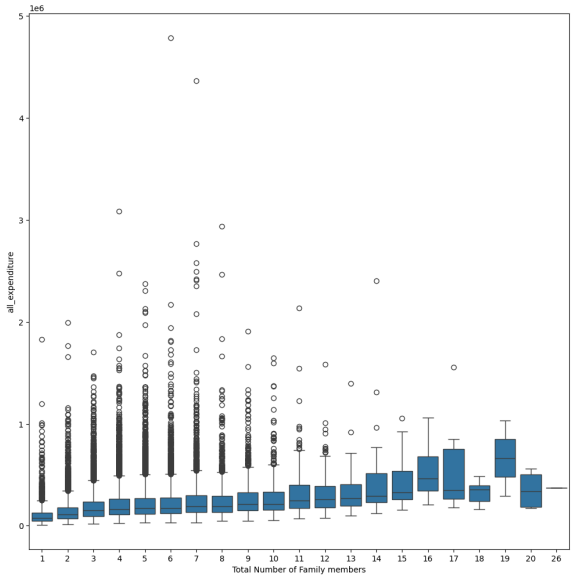


Figure 21: Boxplot comparing the income per total household size & Figure 22: Boxplot comparing the income per total household size (less than 500,000 PHP)



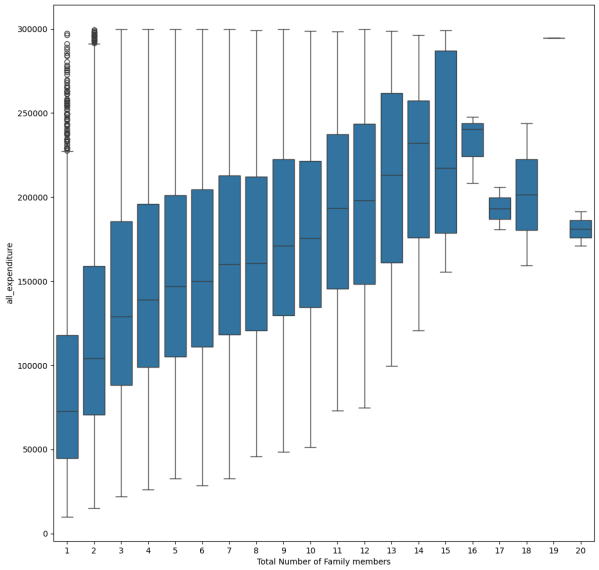


Figure 23: Boxplot comparing all expenditures per total household size & Figure 24: Boxplot comparing all expenditures per total household size (less than 300,000 PHP)

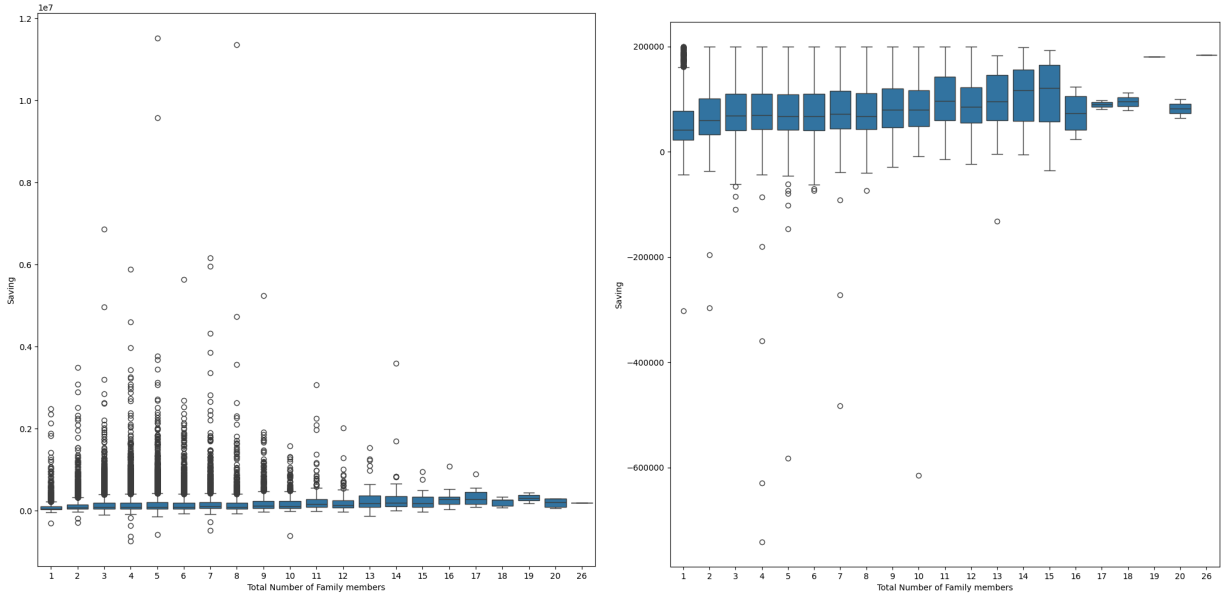


Figure 24: Boxplot comparing savings per total household size & Figure 24: Boxplot comparing savings per total household size (less than 200,000 PHP)

Main Insights:

- Households with 26 and 19 members have the highest mean income yet households with lower amount of members have higher outliers above the mean
- Households with 19 and 14 members have the highest mean expenditure compared to households with 26 members.
- Outliers for total expenditures exist for households with less than 2 members
- Households with 15 and 14 members have the highest mean savings with not negative outliers
- Households with 1 to 10 members, and 13 members report outliers who are not able to save
- High outliers for households with 1 member exist, meaning they are able to save up to 200,000 PHP