

BSIT Weekly Budget and Expenses Survey Analysis

Renz Moquete, Adrian Sayson, Rashir John Castillano, Michael Simpron

2025-12-12

1. Introduction

In this survey, we aim to understand the weekly budget and spending habits of BSIT students. The purpose of this study is to answer key questions such as:

- How much budget do BSIT students usually have per week?
- How much budget do BSIT students usually have every year?
- Do student spend more on foods or transportation?
- What can we conclude at students top expense category?

Variables collected: - Gender

- YearLevel
- WeeklyBudget
- WeeklyExpenses
- TopExpenseCategory (Food, Transportation, boarding, etc.)

1. Libraries needed:

```
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
## 
##     filter, lag

## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union
library(readxl)
library(sentimentr)
```

2. Importing the excel survey

```
data <- read_excel("BSITBUDGET.xlsx")
```

3. Simplifying the column names:

```
data <- data %>%
  rename(
    Name = `1. Name (optional)`,
```

```

YearLevel = `2. Year Level`,
Sex = `3. Sex`,
Age = `4. Age`,
WeeklyBudget = `5. What is your average weekly budget or allowance?`,
FoodExpense = `6. How much do you usually spend on food per week? (School days only)`,
TransportExpense = `7. How much do you usually spend on transportation per week?`,
BudgetEnough = `8. Do you think your weekly budget is enough to support your needs as a BSIT student?`,
LivingSituation = `9. Living Situation`,
Rent = `10. If you live in a dorm/boarding house, how much is your monthly rent? (skip if not applicable)`,
Feedback = `11. What are your thoughts or feedback about this survey?`
)

```

4. Converting into Numeric

```

data <- data %>%
  mutate(
    # Food expenses
    FoodExpenseNum = case_when(
      FoodExpense == "P0 - P200" ~ 100,
      FoodExpense == "P201 - P500" ~ 350.5,
      FoodExpense == "P501 and above" ~ 750,
      TRUE ~ NA_real_
    ),
    # Transportation expenses
    TransportExpenseNum = case_when(
      TransportExpense == "P0 - P200" ~ 100,
      TransportExpense == "P201 - P500" ~ 350.5,
      TransportExpense == "P501 and above" ~ 750,
      TRUE ~ NA_real_
    ),
    # Weekly Budget Expense
    WeeklyBudgetNum = case_when(
      WeeklyBudget == "P0 - P500" ~ 250,
      WeeklyBudget == "P501 - P1,000" ~ 750,
      WeeklyBudget == "P1,001 and above" ~ 1250,
      TRUE ~ NA_real_
    )
  )

```

5. Visualization Graph

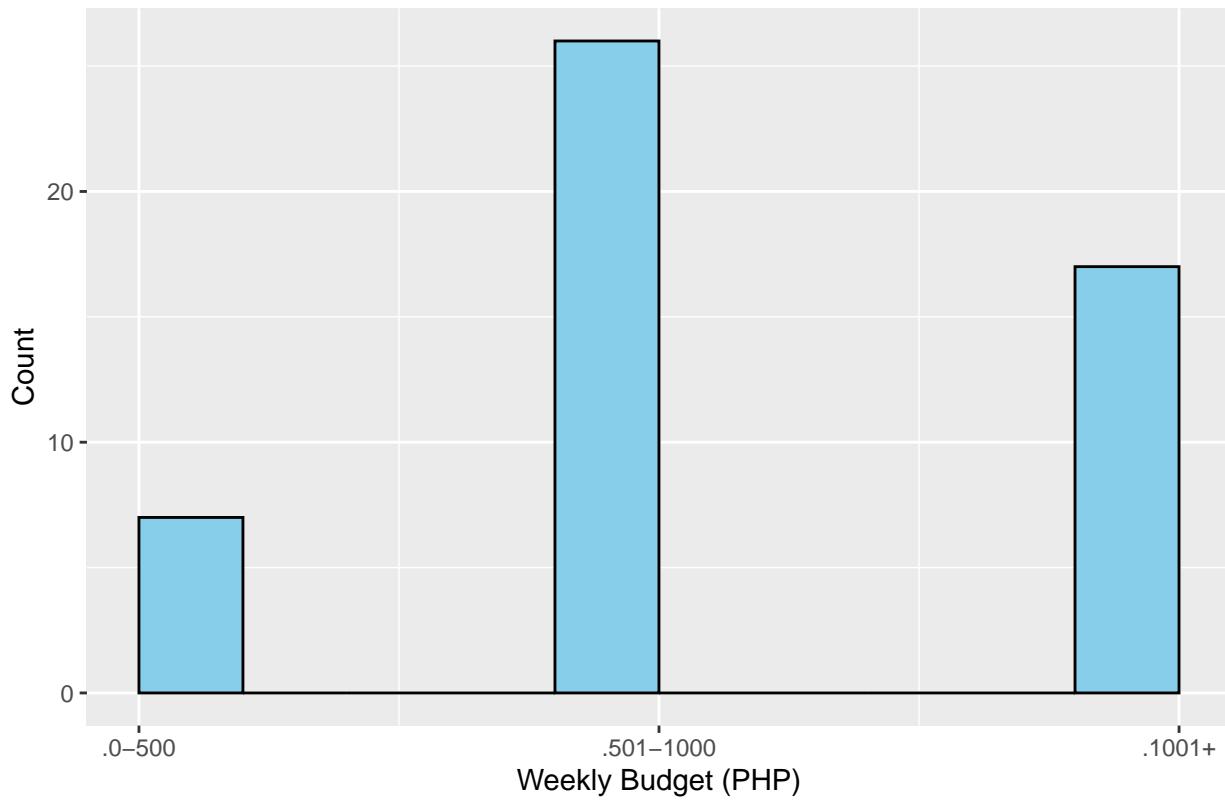
5.1 Plot 1 – Distribution of Weekly Budget

```

ggplot(data, aes(x = WeeklyBudgetNum)) +
  geom_histogram(binwidth = 100, fill = "skyblue", color = "black") +
  labs(
    title = "Distribution of Weekly Budget of BSIT Students",
    x = "Weekly Budget (PHP)",
    y = "Count"
  ) +
  scale_x_continuous(breaks = c(250, 750, 1250),
                     labels = c("P0-500", "P501-1000", "P1001+"))

```

Distribution of Weekly Budget of BSIT Students

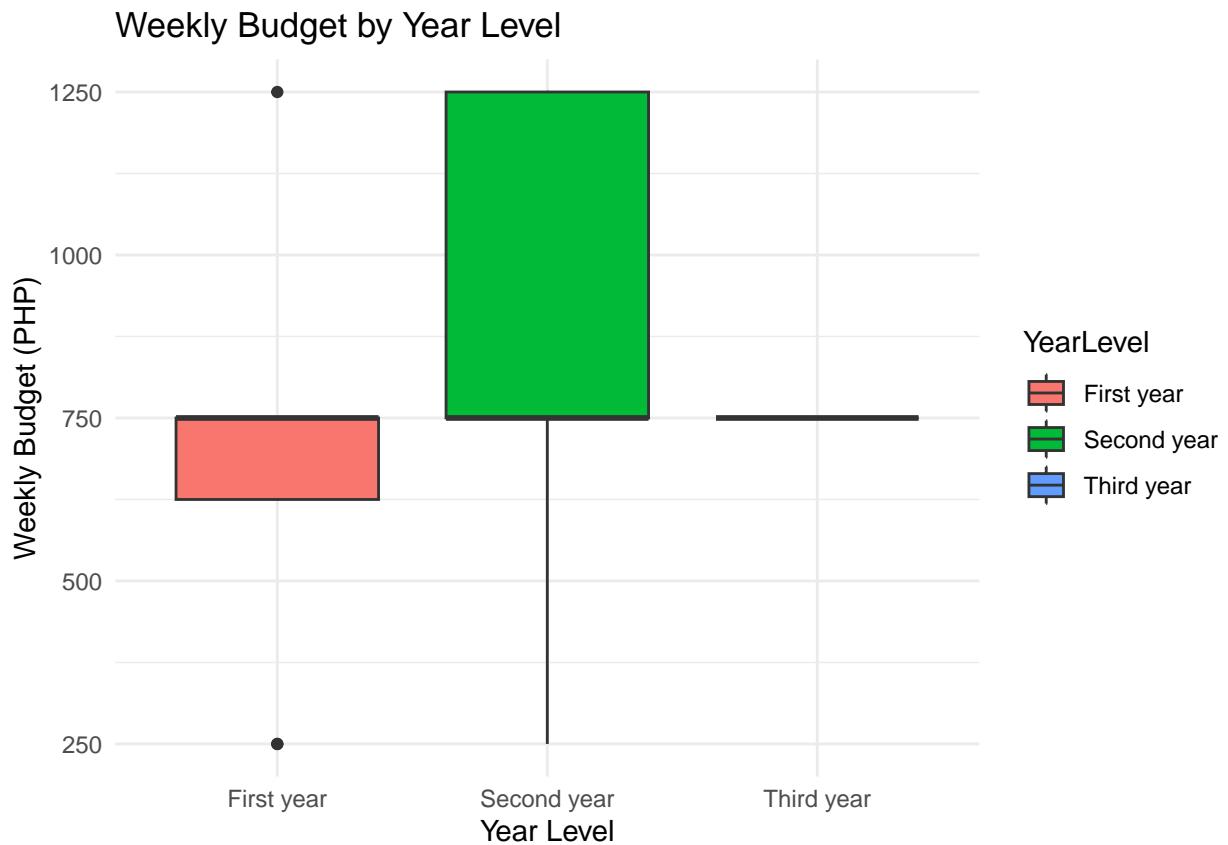


Interpretation

In the graph it shows that most students in BSIT have 501-1000 budget per week, there are also some BSIT students who have 1001+ weekly budget, and lastly there's only a little population of students who only have 0-500 weekly budget.

5.2 Plot 2 – Weekly Budget by Year Level

```
ggplot(data, aes(x = YearLevel, y = WeeklyBudgetNum, fill = YearLevel)) +  
  geom_boxplot() +  
  labs(  
    title = "Weekly Budget by Year Level",  
    x = "Year Level",  
    y = "Weekly Budget (PHP)"  
) +  
  theme_minimal()
```



Interpretation

This graph shows the weekly budget of students by year level. It shows that 1st year students have a budget between 600-750, second year students have a budget between 750-1250, and the third year has a budget of 750 per week. And the graph shows that the second year BSIT students have the highest weekly budget for this school year.

5.3 Plot 3 – Do student spend more on foods or transportation?

```

summary(data$FoodExpenseNum)

##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##    100.0   350.5  350.5   383.3  350.5   750.0

summary(data$TransportExpenseNum)

##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.    NA's
##    750    750    750    750    750    750     45

food_avg <- mean(data$FoodExpenseNum, na.rm = TRUE)
transport_avg <- mean(data$TransportExpenseNum, na.rm = TRUE)

cat("Average Food Expense: P", round(food_avg, 1), "\n")

## Average Food Expense: P 383.3
cat("Average Transport Expense: P", round(transport_avg, 1), "\n")

## Average Transport Expense: P 750

```

```

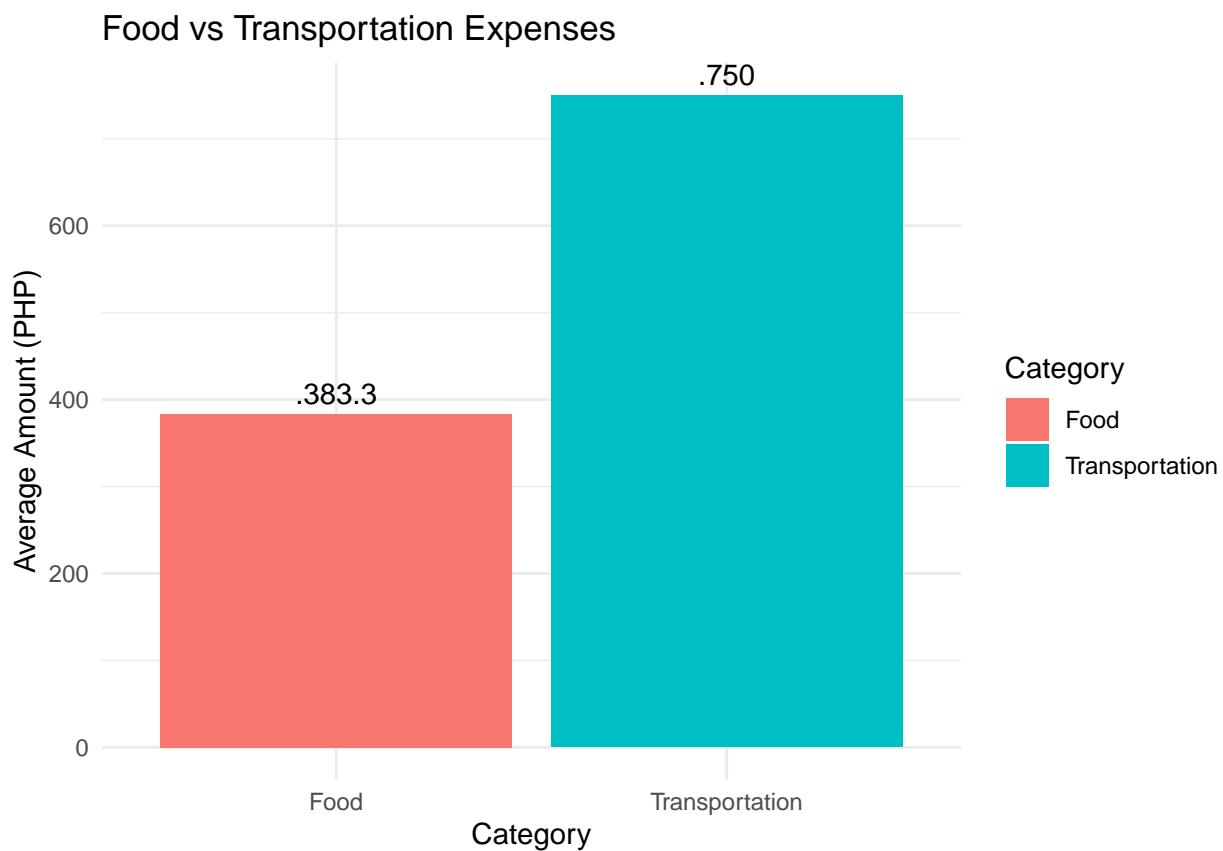
cat("Difference: P", round(food_avg - transport_avg, 1), "\n")

## Difference: P -366.7

plot_data <- data.frame(
  Category = c("Food", "Transportation"),
  Amount = c(food_avg, transport_avg)
)

ggplot(plot_data, aes(x = Category, y = Amount, fill = Category)) +
  geom_col() +
  geom_text(aes(label = paste0("P", round(Amount, 1)))), vjust = -0.5, size = 4) +
  labs(
    title = "Food vs Transportation Expenses",
    x = "Category",
    y = "Average Amount (PHP)"
  ) +
  theme_minimal()

```

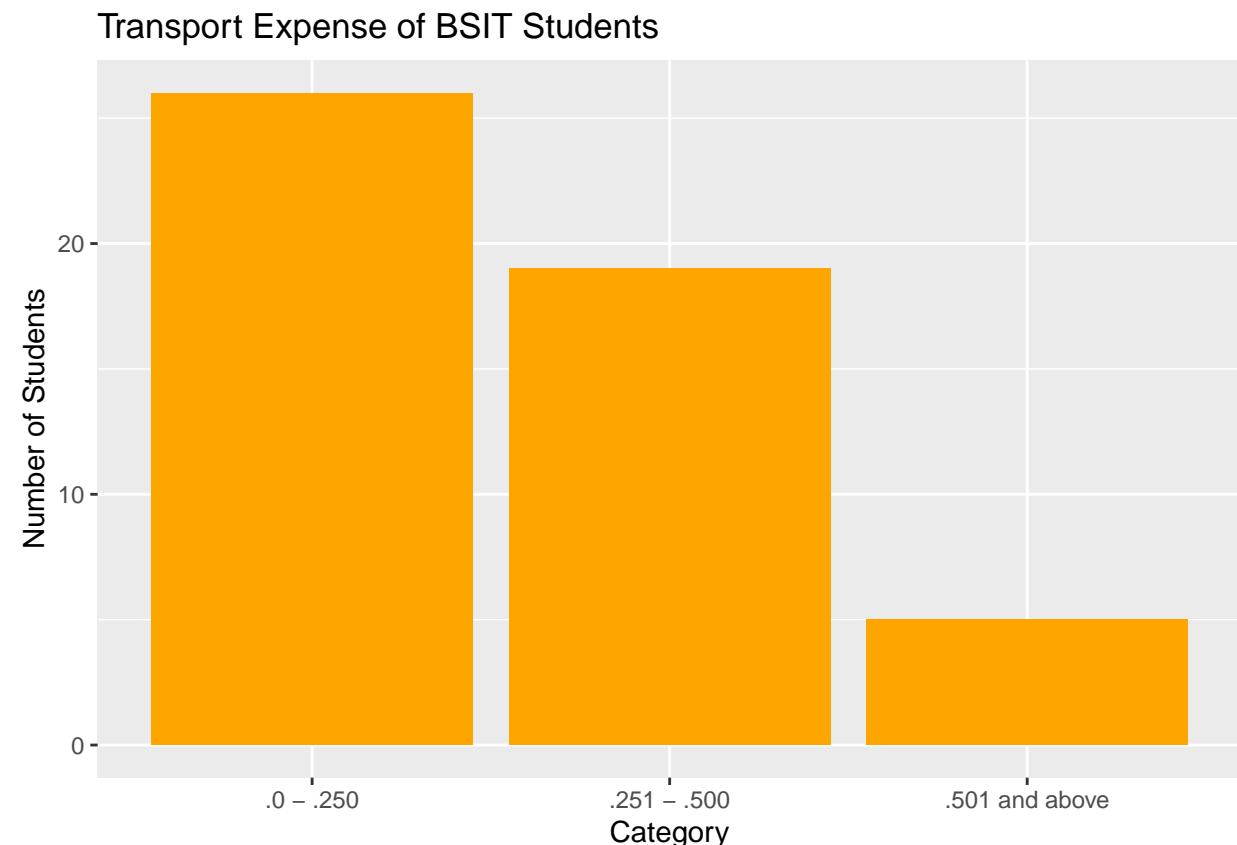


Interpretation

In the graph it shows that most students in BSIT spends more of their budget at transportation expense compared to foods, this made us conclude how costly it is to commute, and how it exceed the food expenses for most BSIT students, indicating commuting is a major financial challenge.

5.4 Plot 4 – Most Top Expense Category

```
ggplot(data, aes(x = TransportExpense)) +  
  geom_bar(fill="orange") +  
  labs(title = "Transport Expense of BSIT Students",  
       x = "Category",  
       y = "Number of Students")
```



Interpretation

The bar graph shows that students much prefer to save money on transportation or to have options like walking. Few goes past P500 and this goes to show that efficient public transportation is very important to BSIT students.

6. Sentiment analysis the feedbacks:

6.1 Cleaning the data

```
data <- data %>%  
  mutate(  
    Feedback_clean = tolower(Feedback), # Convert to lowercase  
    Feedback_clean = gsub("[^a-zA-Z0-9 ]", "", Feedback_clean) # Remove special chars  
)
```

6.2. Collecting the scores of sentiments

```
sentiment_scores <- sentiment_by(data$Feedback_clean)  
data$sentiment_score <- sentiment_scores$ave_sentiment
```

```

### 3. Categorize sentiment
data <- data %>%
  mutate(
    sentiment_category = case_when(
      sentiment_score > 0.1 ~ "Positive",
      sentiment_score < -0.1 ~ "Negative",
      TRUE ~ "Neutral"
    )
  )

sentiment_summary <- data %>%
  count(sentiment_category) %>%
  mutate(Percentage = n / sum(n) * 100)

print(sentiment_summary)

```

```

## # A tibble: 3 x 3
##   sentiment_category     n Percentage
##   <chr>           <int>     <dbl>
## 1 Negative          14        28
## 2 Neutral            14        28
## 3 Positive           22        44

```

6.3 Turning sentiments into Visualization

```

ggplot(sentiment_summary, aes(x = sentiment_category, y = n, fill = sentiment_category)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = paste0(n, "\n", round(Percentage, 1), "%))),
  vjust = 1.5, # Position inside bar (1.5 = near top inside)
  color = "white", # White text for contrast
  size = 6, # Slightly smaller
  fontface = "bold") +
  labs(
    title = "Sentiment Analysis of Student Feedback",
    subtitle = "Categorization of Survey Comments",
    x = "Sentiment Category",
    y = "Number of Comments"
  ) +
  scale_fill_manual(values = c("Positive" = "#2ecc71",
                              "Neutral" = "#f39c12",
                              "Negative" = "#e74c3c")) +
  theme_minimal() +
  theme(legend.position = "none")

```

Sentiment Analysis of Student Feedback

Categorization of Survey Comments



Interpretation

The sentiment analysis graph shows over 50% positive feedback, indicating the survey successfully engaged most students. However, the 28% negative responses highlight specific areas needing improvement in our future surveys.