

# Weekly Status Submission Template

**Week Number: 4**

**Date: September 22 2025**

**Student Name: Abdullah Siddiqui**

## Original Plan for This Week: (5 points)

According to my project timeline and prior status report, my Week 4 plan was to:

- Move beyond descriptive statistics into exploring relationships between variables.
- Investigate correlations between predictors such as: education, income, exercise, smoking, alcohol use and poor mental health days.
- Create scatterplots and cross-tabulations to visualize relationships.
- Begin drafting initial regression model frameworks (baseline linear and logistic regressions).

## Tasks Accomplished This Week: (10 points)

This week I:

- Calculated pairwise correlations between numeric variables e.g., education, income, age group, poor mental health days.
- Generated a correlation heatmap and scatterplots e.g., income vs. poor mental health days; exercise vs. poor mental health days.
- Conducted cross-tabulations e.g., smoking vs. exercise; alcohol use vs. poor mental health days.
- Checked regression assumptions:
  - Assessed multicollinearity using VIF scores.
  - Examined linearity of relationships through scatterplots.
  - Began testing normality of residuals using Q–Q plots.
- Drafted baseline logistic regression (binary poor mental health days: frequent distress vs. not) and linear regression (continuous mental health days).

## Comparison of Planned vs Actual: (10 points)

Week	Planned	Actual
4	Correlation analysis, crosstabs, scatterplots, regression prep	Completed correlation matrix, scatterplots, heatmap, crosstabs, regression assumption checks, and drafted initial regression models

I was able to accomplish everything I had planned and slightly exceeded by already starting assumption checks and model drafts.

## Self-Rating of Progress: (5 points)

- ☒ Met the Planned Tasks
- ☐ Did Not Meet the Planned Tasks

(If you did not meet the planned tasks, please explain why and how you plan to get back on schedule)

I successfully met all planned tasks for this week and began building ahead toward Week 5.

## Plan for Next Week: (5 points)

For Week 5, I plan to:

- Implement gradient boosting models (XGBoost, LightGBM) as outlined in my project timeline.
- Compare their performance with the baseline regression models.
- Continue checking model assumptions and performance metrics (accuracy, RMSE, AUC depending on model type).
- Document the findings clearly in my GitHub repository with plots and Markdown summaries.

## Evidence of Tasks Completed: (15 points)

Dataset: Continued use of BRFSS2023\_subset\_clean.csv (subset of 2023 BRFSS).

Code: Updated R scripts (BRFSS2023\_correlation.R and BRFSS2023\_regression.R) for correlation analysis and regression prep.

Analysis Outputs:

- Correlation heatmap (correlation\_heatmap.png).
- Scatterplot: Income vs. Poor Mental Health Days (scatter\_income\_mentalhealth.png).
- Scatterplot: Exercise vs. Poor Mental Health Days (scatter\_exercise\_mentalhealth.png).
- Cross-tabulation tables (exercise × smoking, alcohol × mental health).
- Initial regression output summaries (regression\_baseline\_output.txt).

Repository Link:

<https://github.com/abdullahs1357/SeniorThesis.git>