








# Project Progress Assessment




## Step 1: Data Loading and Understanding - 100% Complete

-  Data successfully loaded
-  Structure examined (columns, data types)
-  Basic data cleaning and preprocessing performed (fixed\_female\_farmers\_data.xlsx)
-  Data dictionary/summary created (evidenced by the codebook)




## Step 2: Exploratory Data Analysis - 100% Complete

-  Descriptive statistics generated for both numerical and categorical variables
-  Distribution analysis performed with histograms, boxplots, and pie charts
-  Basic visualizations created (multiple charts for each variable)




## Step 3: Key Relationships Analysis - 100% Complete

-  Correlation analysis completed for numerical variables (correlation\_matrix.csv)
-  Cross-tabulations created for categorical variables (numerous crosstab files in subdirectories)
-  Relationship visualizations generated (heatmaps, grouped bar charts, etc.)




## Step 4: Health Risk Factors Analysis - 70% Complete

-  Basic analysis of factors associated with health issues
-  Protective equipment usage patterns thoroughly analyzed
-  More rigorous statistical tests for significant relationships needed
  - Missing formal hypothesis testing and significance values for key relationships
  - Could benefit from regression analysis or other statistical modeling

## Step 5: Multivariate Analysis - 0% Complete

-  Principal Component Analysis (PCA) not yet conducted
-  Multiple Correspondence Analysis not performed
-  Still need to identify deeper hidden relationships using multivariate techniques

## Step 6: Final Reporting - 20% Complete

-  Initial findings compiled but not yet synthesized
-  Executive summary not yet created
-  Basic visualizations produced, but final presentation-quality visuals still needed

## Overall Progress: Approximately 65% Complete

### Strengths of Current Analysis:

- Very thorough exploratory analysis of both numerical and categorical variables
- Excellent documentation of data properties and distributions
- Strong foundation for understanding the dataset characteristics

### Key Next Steps:

1. **Complete Health Risk Analysis:**

- Add formal statistical tests (chi-square, t-tests, ANOVA)
- Create risk ratio calculations for protective behaviors and health outcomes

## **2. Perform Multivariate Analysis:**

- Conduct PCA on numerical variables to identify key factors
- Perform Multiple Correspondence Analysis on categorical variables
- Look for latent variable patterns across the dataset

## **3. Create Final Reports:**

- Synthesize findings into a cohesive narrative
- Generate executive summary highlighting key insights
- Produce professional-quality visualizations for presentation

The analysis so far provides a strong foundation, but the multivariate analysis will be crucial for uncovering the deeper patterns and relationships that aren't immediately apparent from univariate and bivariate analyses. This is where you're likely to find the most valuable and actionable insights.