

Predicting SHIF Contributions



The Team



Cynthia Dalmas



Brian Ochieng



Nicole Bosibori



**Edwin
Mutendwa**



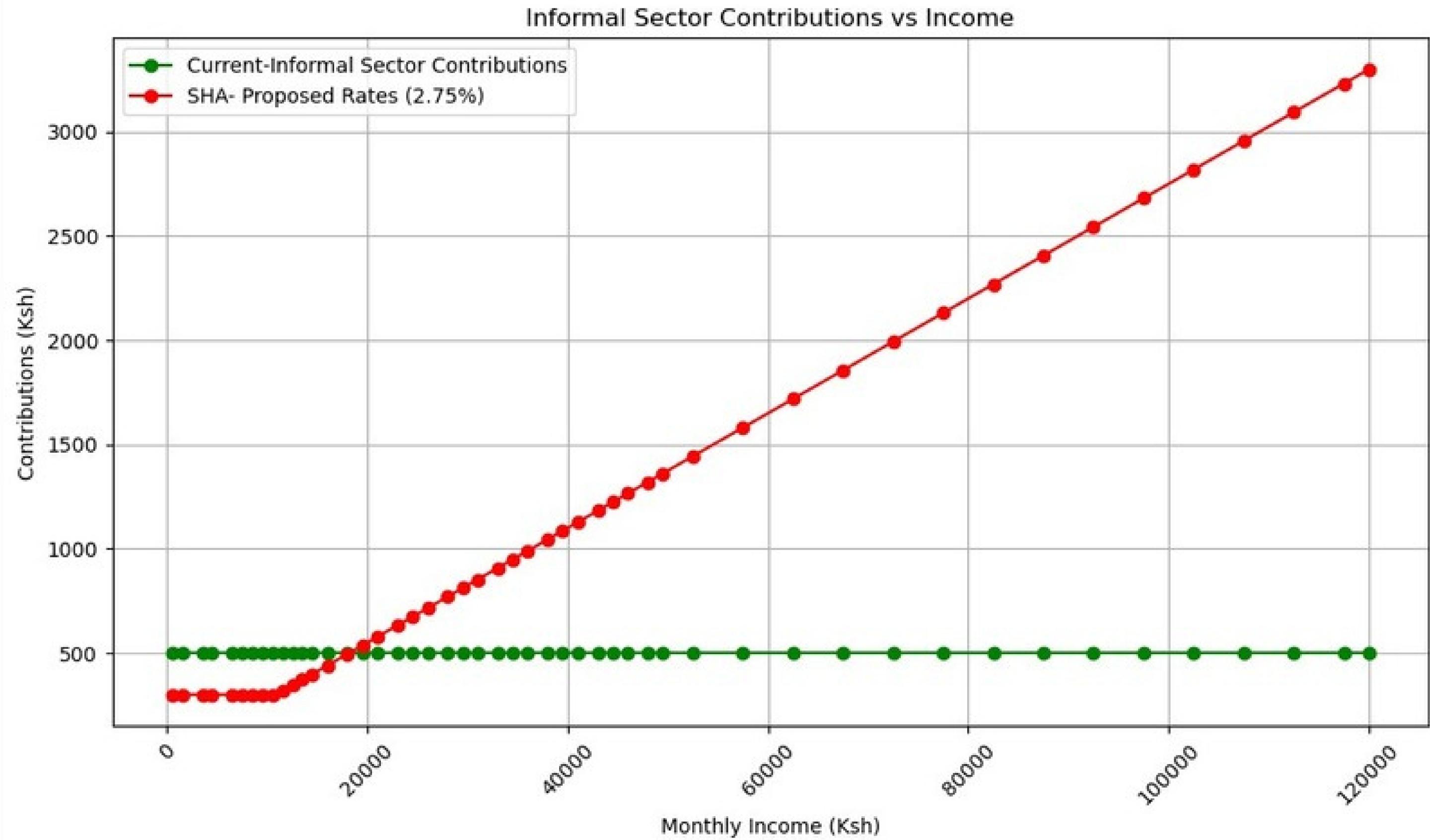
Nyakio Ruth

Introduction

The informal sector makes up 80% of the Kenyan population and it is difficult to determine their income and by extension the Social Health Insurance Fund (SHIF) contributions.

The Ministry of Health needs to come up with a fair methodology to determine the contributions before SHIF is assented in Oct 2024

NHIF vs SHIF



Business Understanding

To ensure a seamless transition from NHIF to SHIF, the Ministry of Health can use housing indicator data from the Kenya National Bureau of Statistics, to predict informal workers' income and ensure fair and accurate SHIF contributions



Problem Statement

Determining the SHIF contributions for the informal sector workers is a challenge due to the undocumented nature of their income.

Stakeholders

**Informal Sector
Workers**

**Social Health
Authority
(SHA)**

Objectives



Predict SHIF contributions for the informal sector



Understand the features and identify patterns

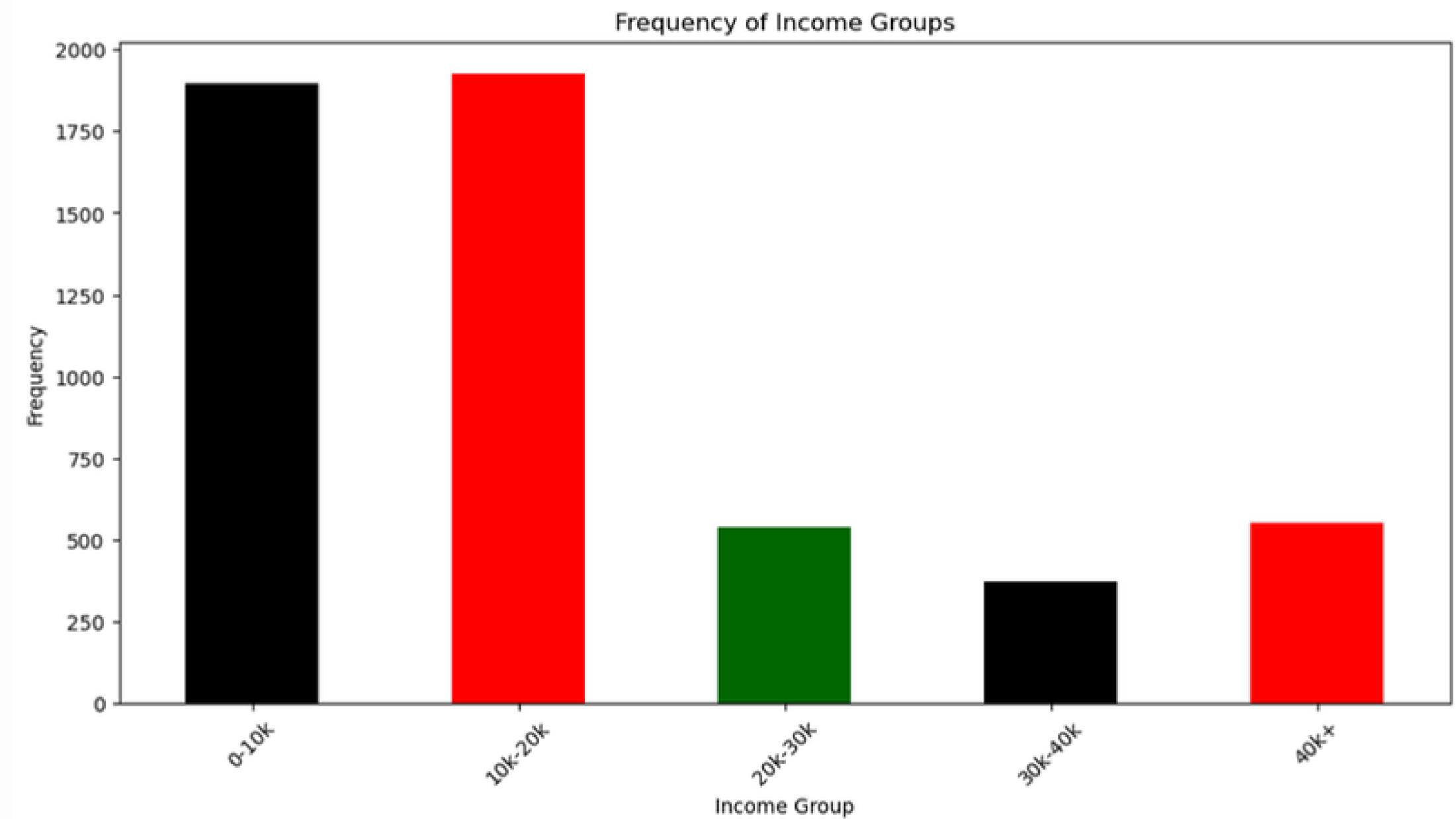


Generate Insights and Recommendations



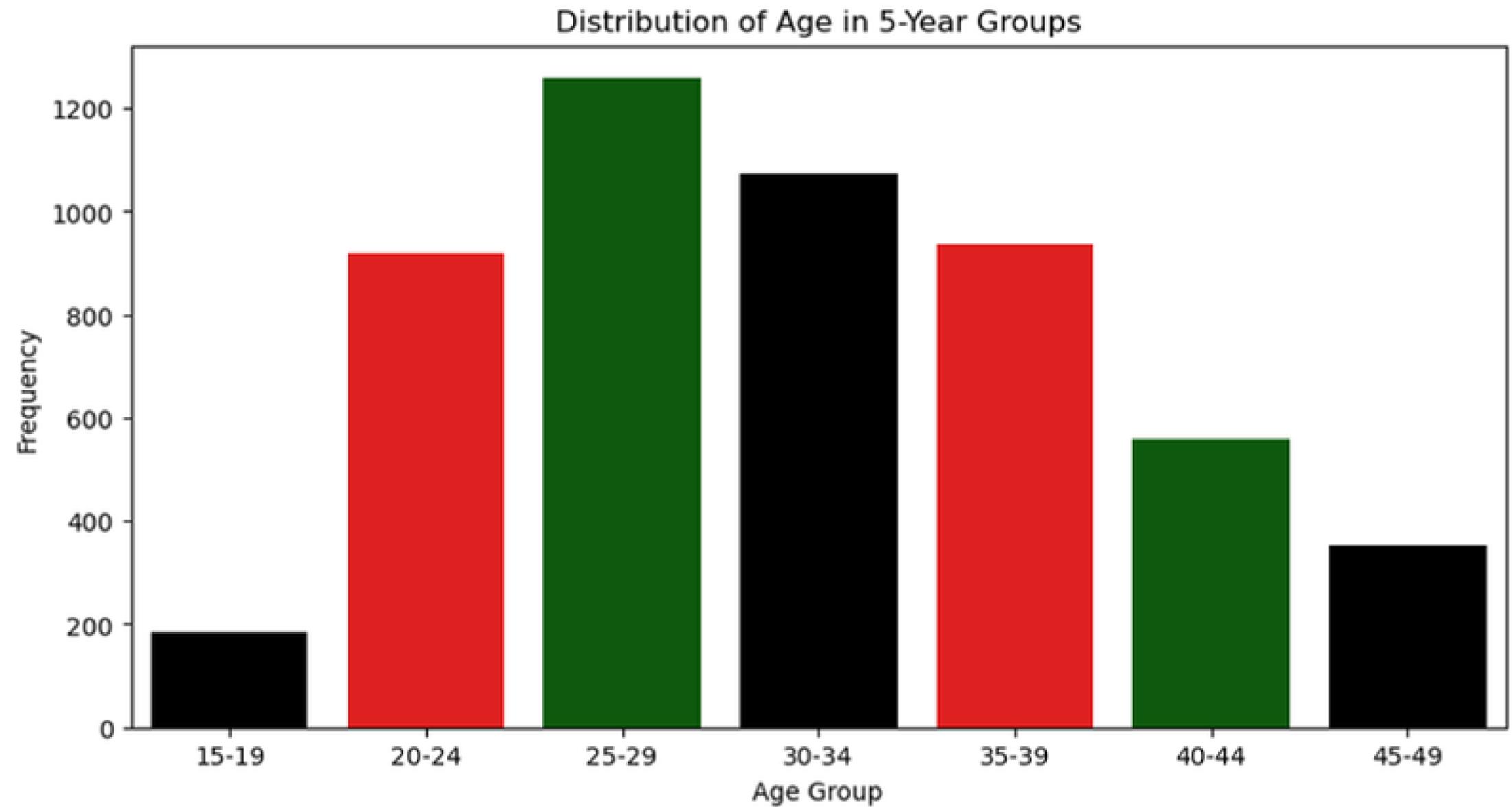
Frequency of income groups

Most individuals earn a low income of between 0-20k KSH



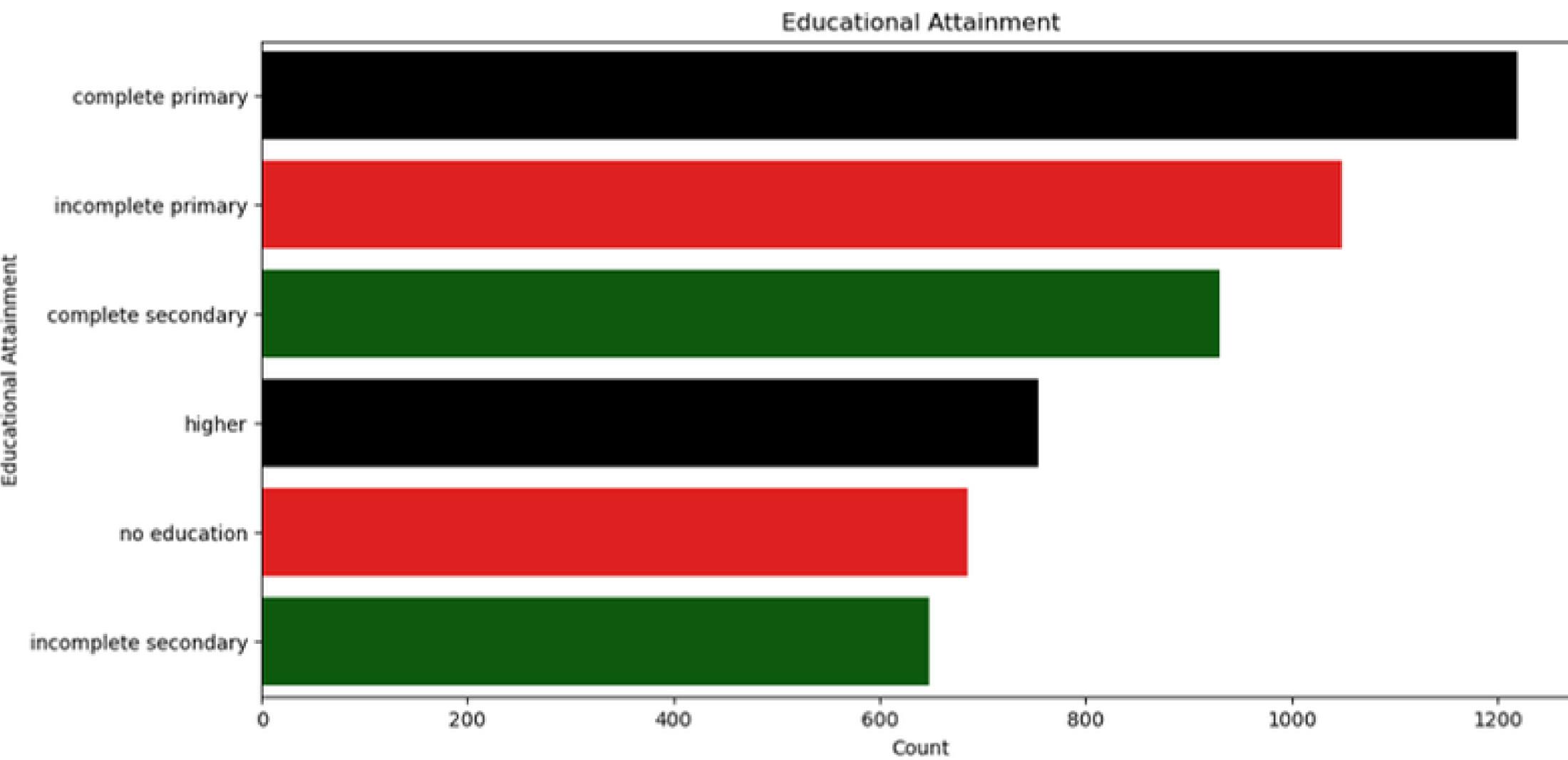
Distribution of Age

Youthful demographic
with the majority aged 25
to 29



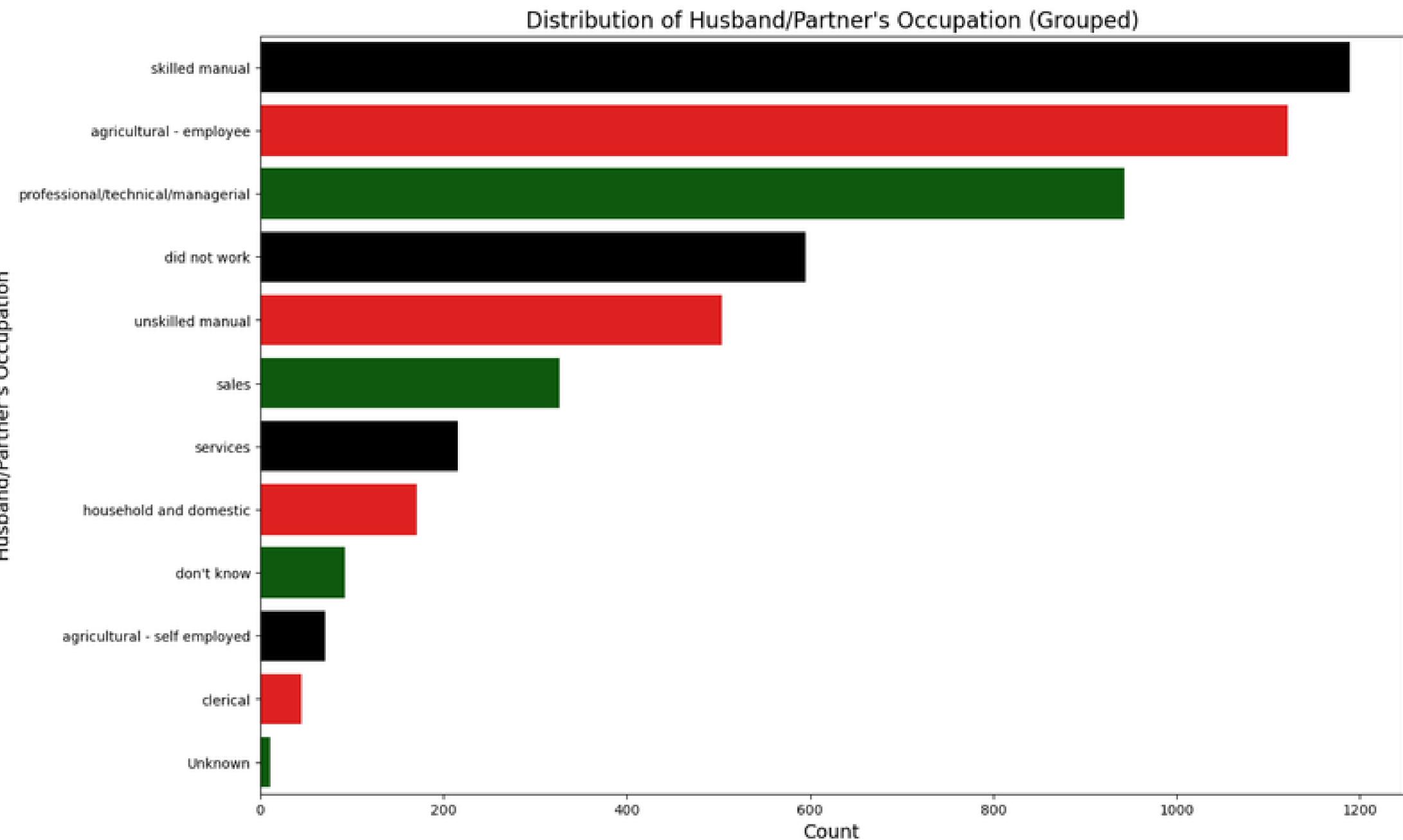
Distribution of Education Attainment

Most completed primary and secondary education. Drop-off rates are also high.



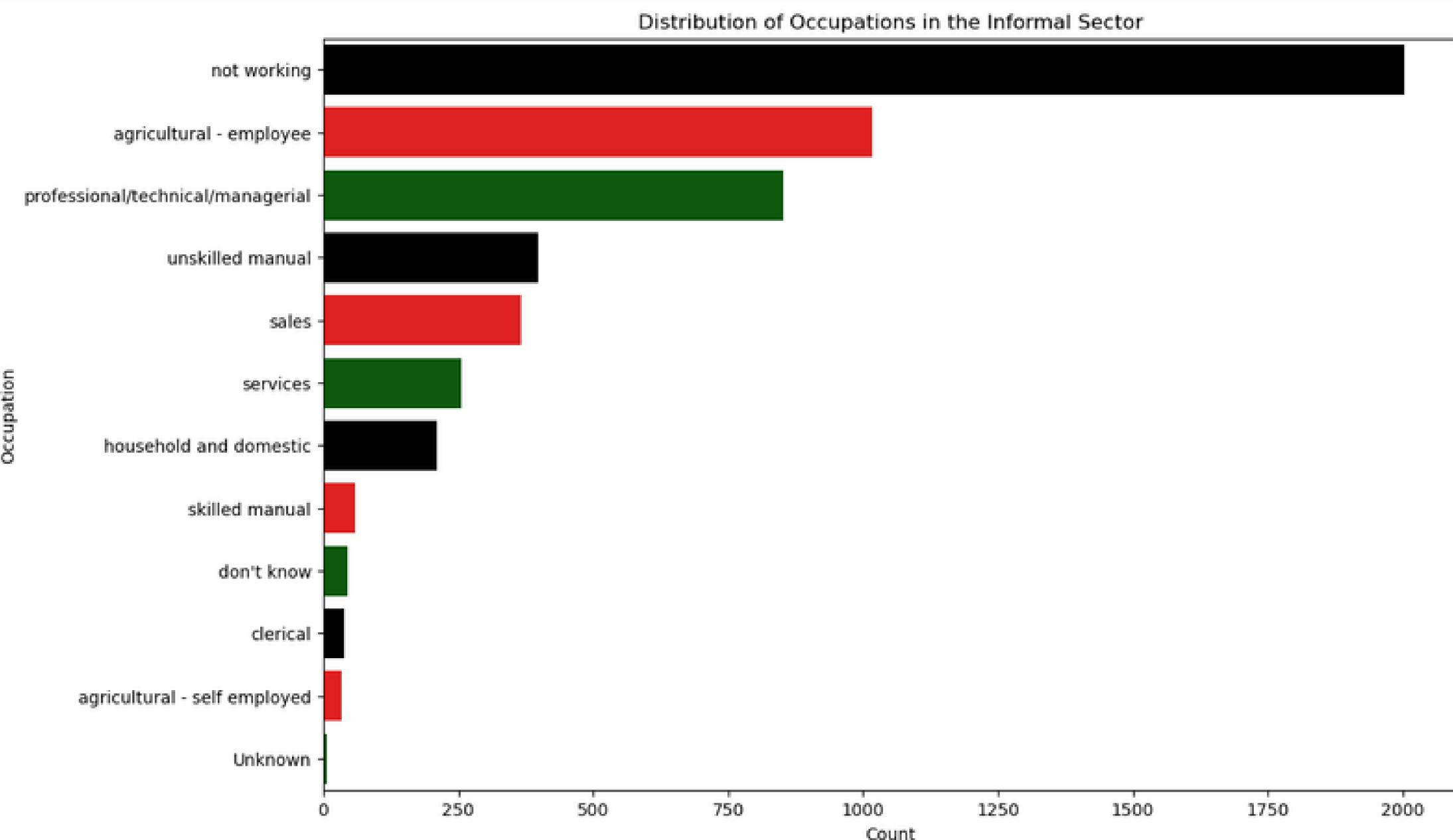
Distribution of Partners' Occupation

Most respondents' partners are skilled manual workers or work in the agricultural sector



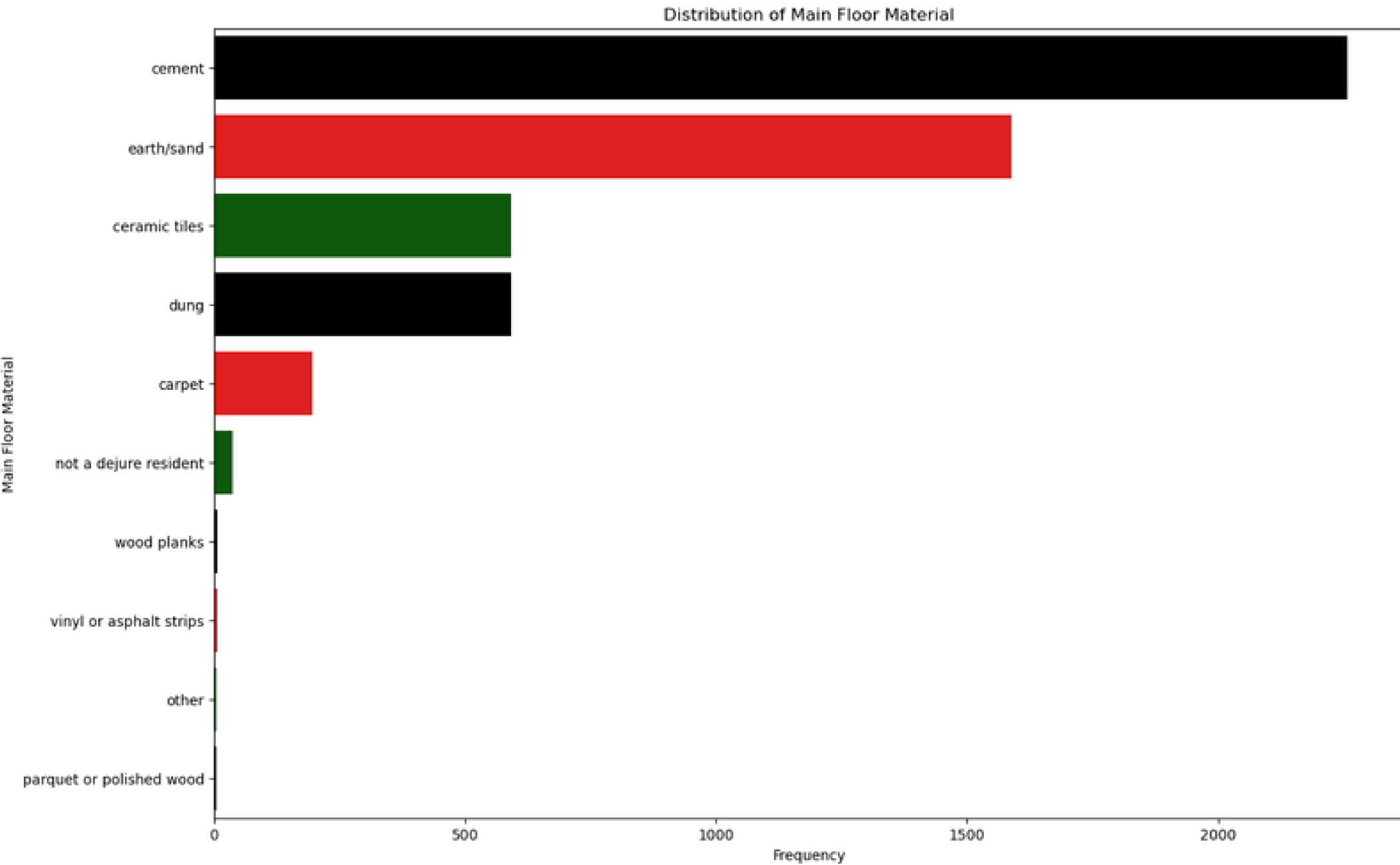
Distribution of Occupations

Most respondents are unemployed



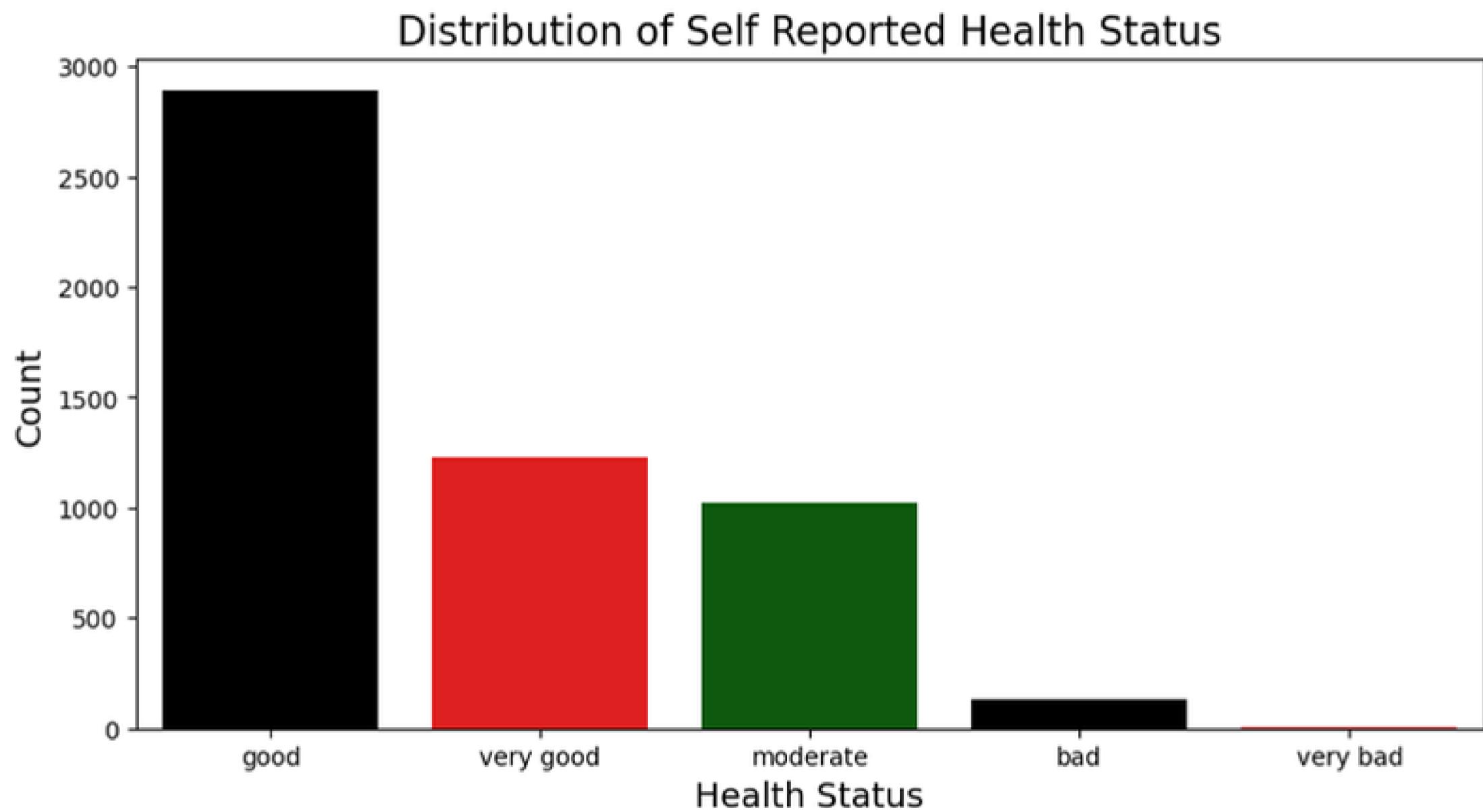
Distribution of Main Floor Material

Most individuals live in houses with cement or earth/sand as their main house floor material



Distribution of Health Status

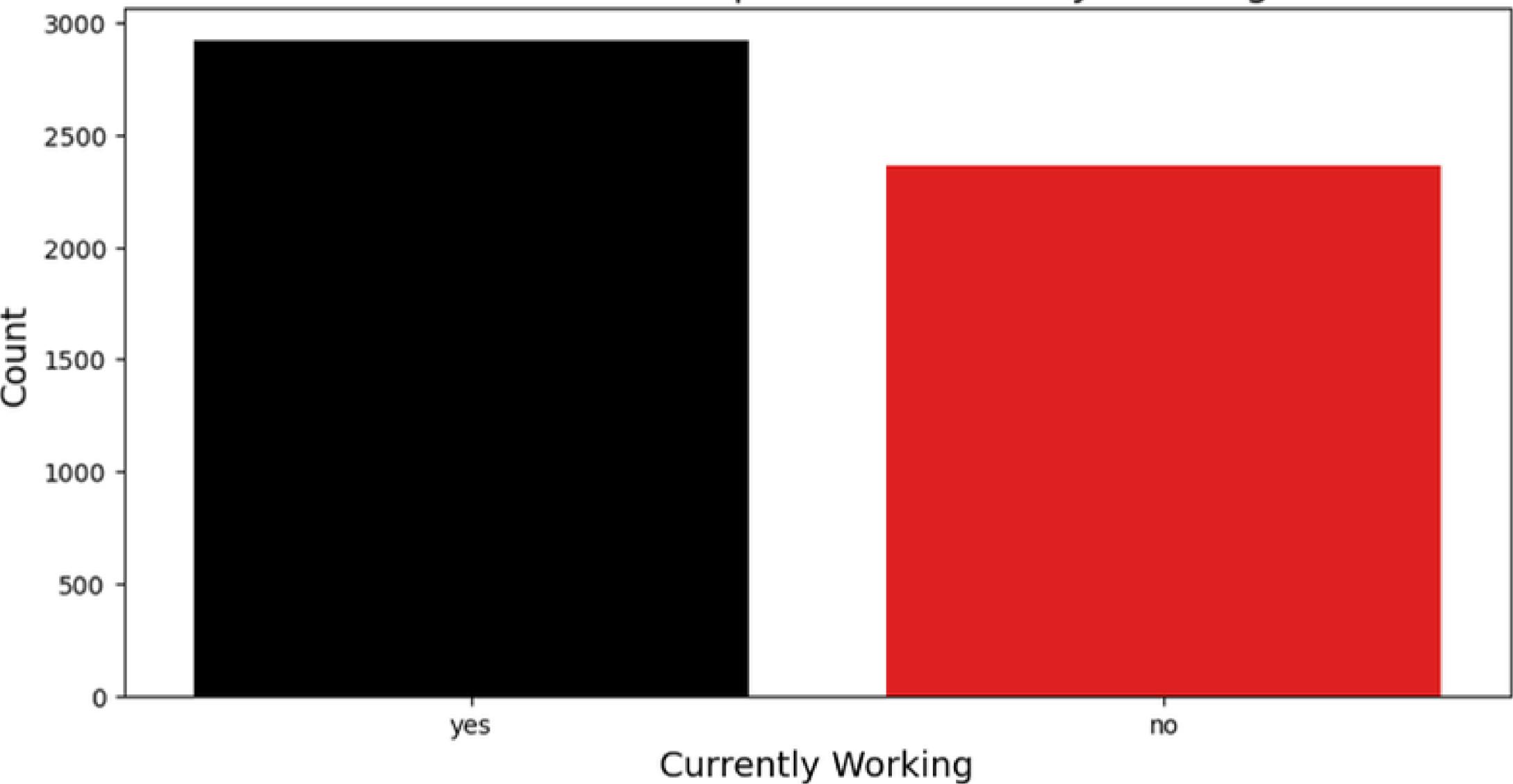
Most respondents have good health. Only a few are not healthy



Distribution of Working Status

Slightly more respondents are currently employed than not working

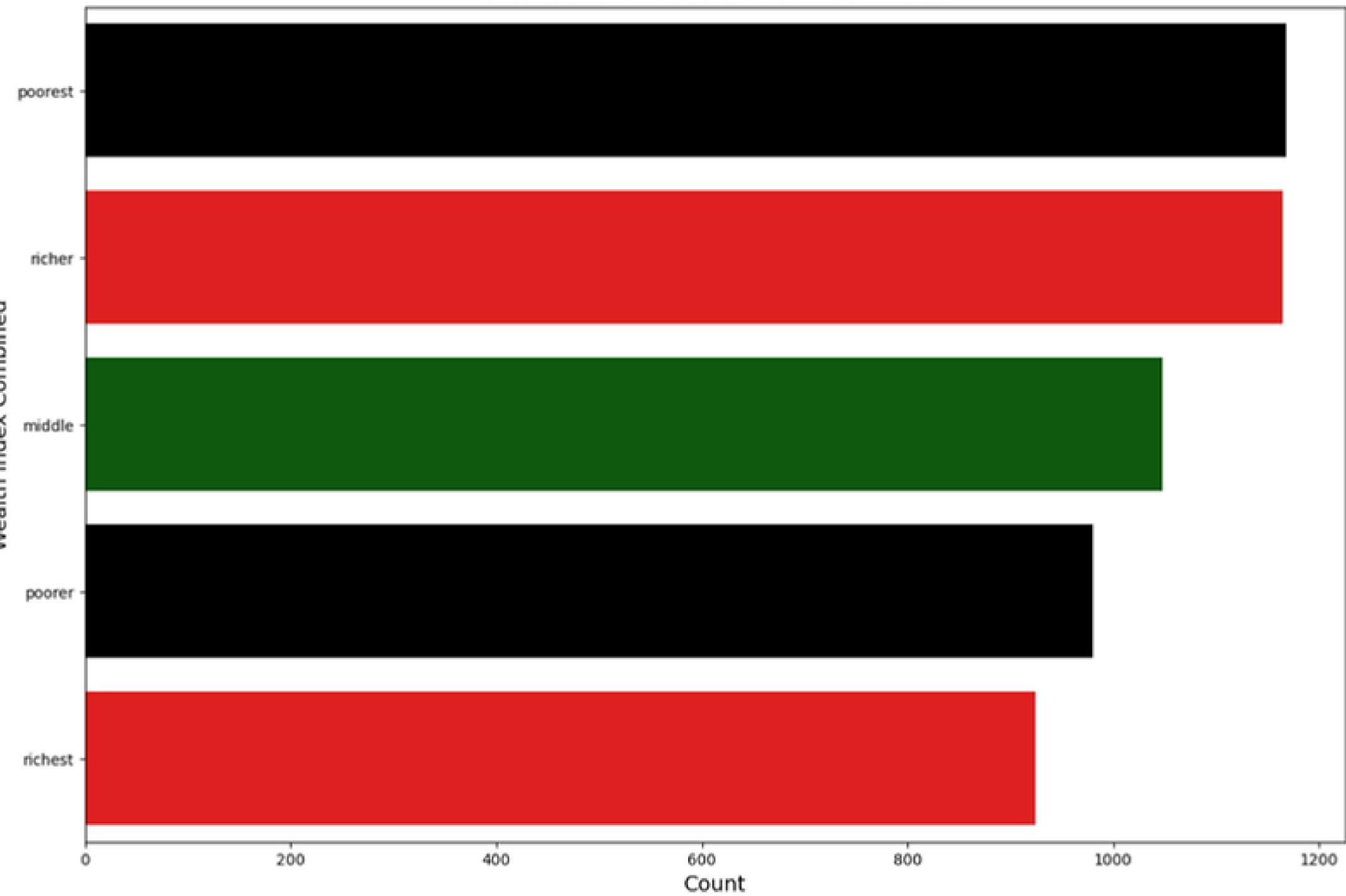
Distribution of Respondent Currently Working



Distribution of Wealth Index

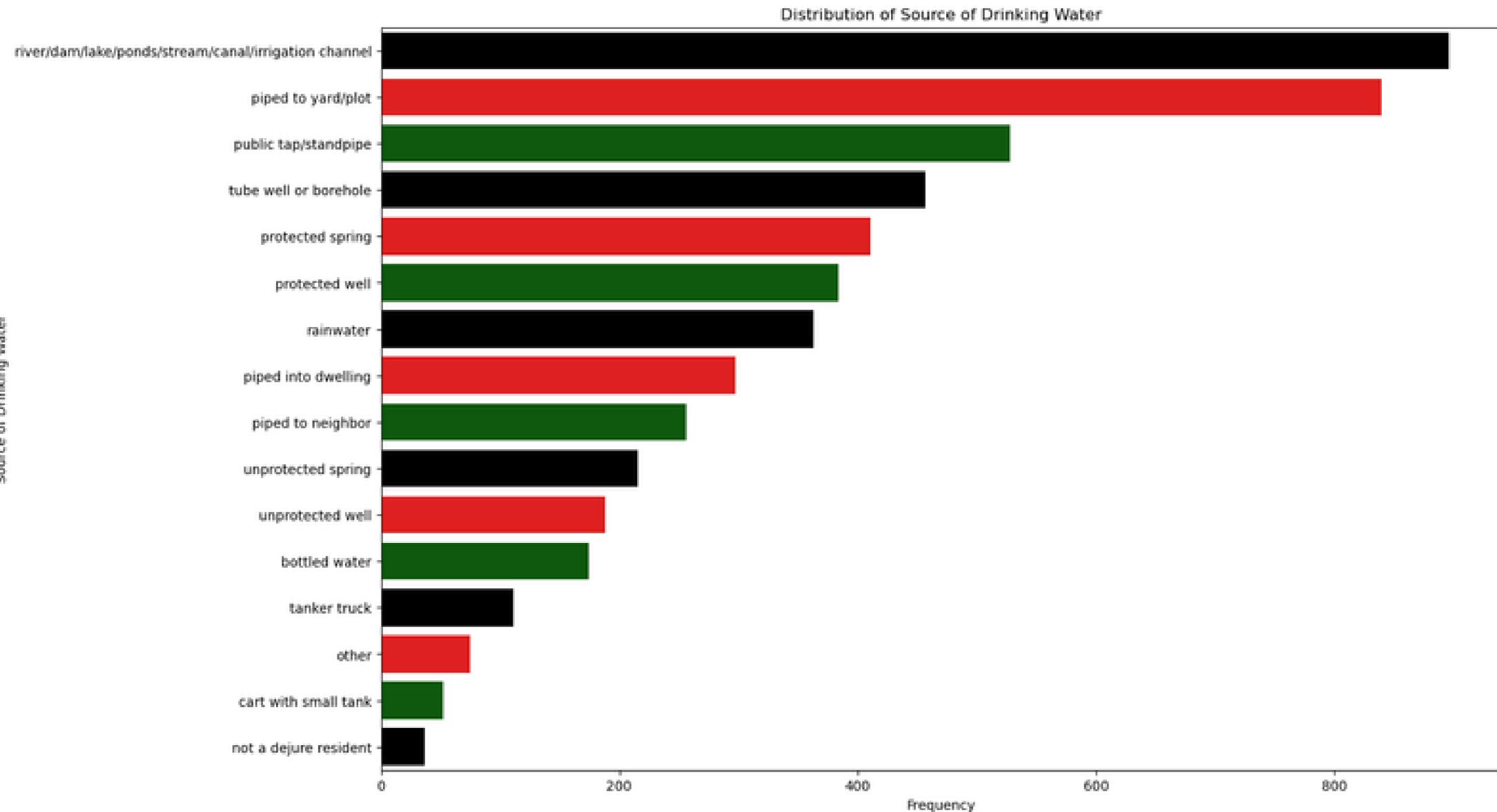
Wealth distribution is balanced, with a slight concentration in the "poorest" and "richer" categories

Distribution of Wealth Index Combined



Distribution of Sources of Drinking Water

Natural water bodies and piped water are the main source of drinking water



Income distribution by place of residence

Most rural residents earn less than 10k. Urban residents have a wider income distribution, peaking at 20k-30k KSH.



MODELS

Baseline Model: Logistic Regression

Accuracy: 95.52%

F1 Score: 0.95

Precision: 0.94

Recall: 0.96

- Precision, recall, and F1 scores were high across all classes, indicating good performance.
- Most predictions were correct, with few misclassifications across different classes.

Best Model: Random Forest Classifier

Accuracy: 99.38%

F1 Score: 0.99

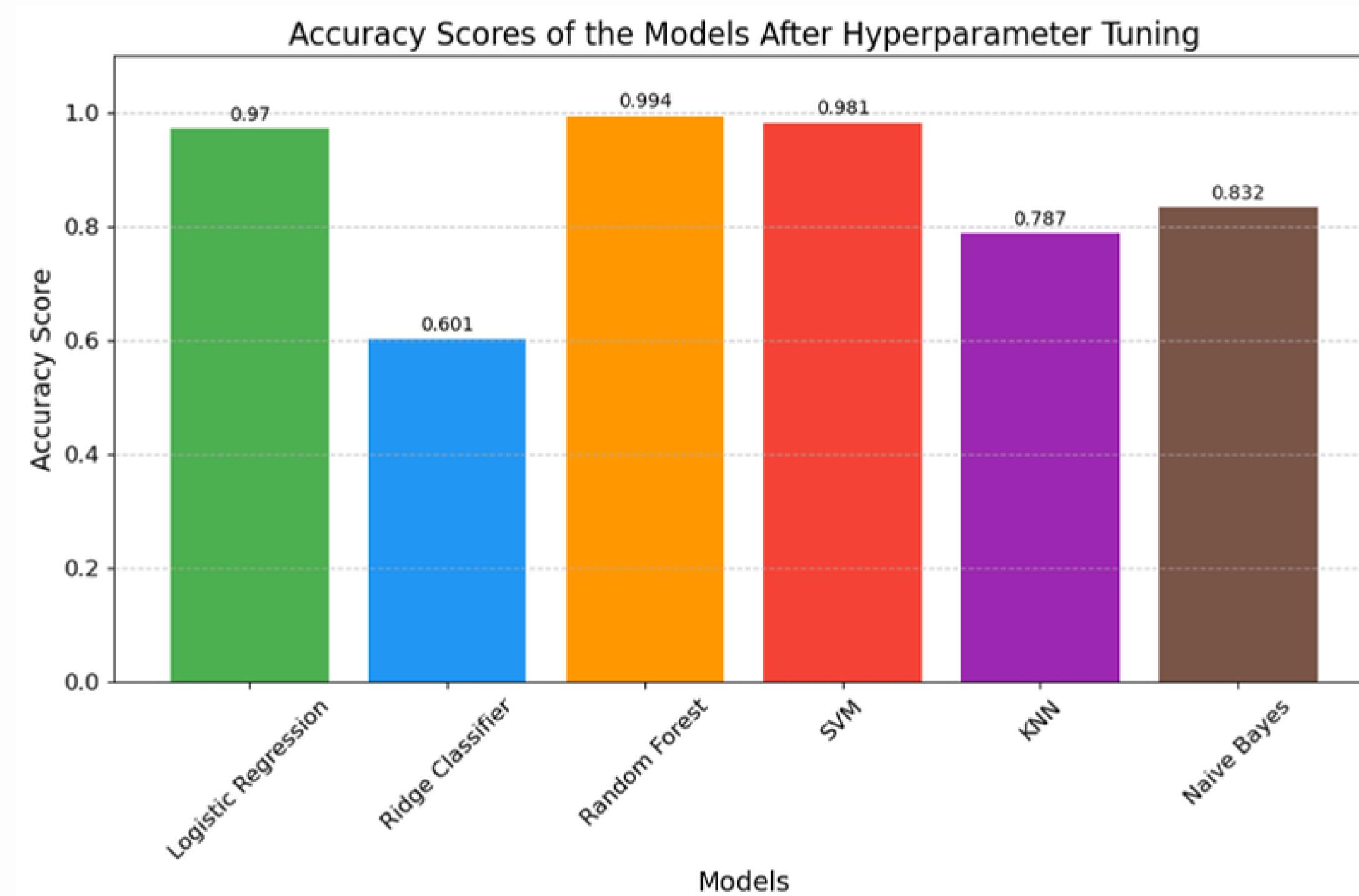
Precision: 0.98

Recall: 0.95

- The Random Forest Classifier performed exceptionally well, efficiently handling complex data and reducing overfitting through ensemble methods

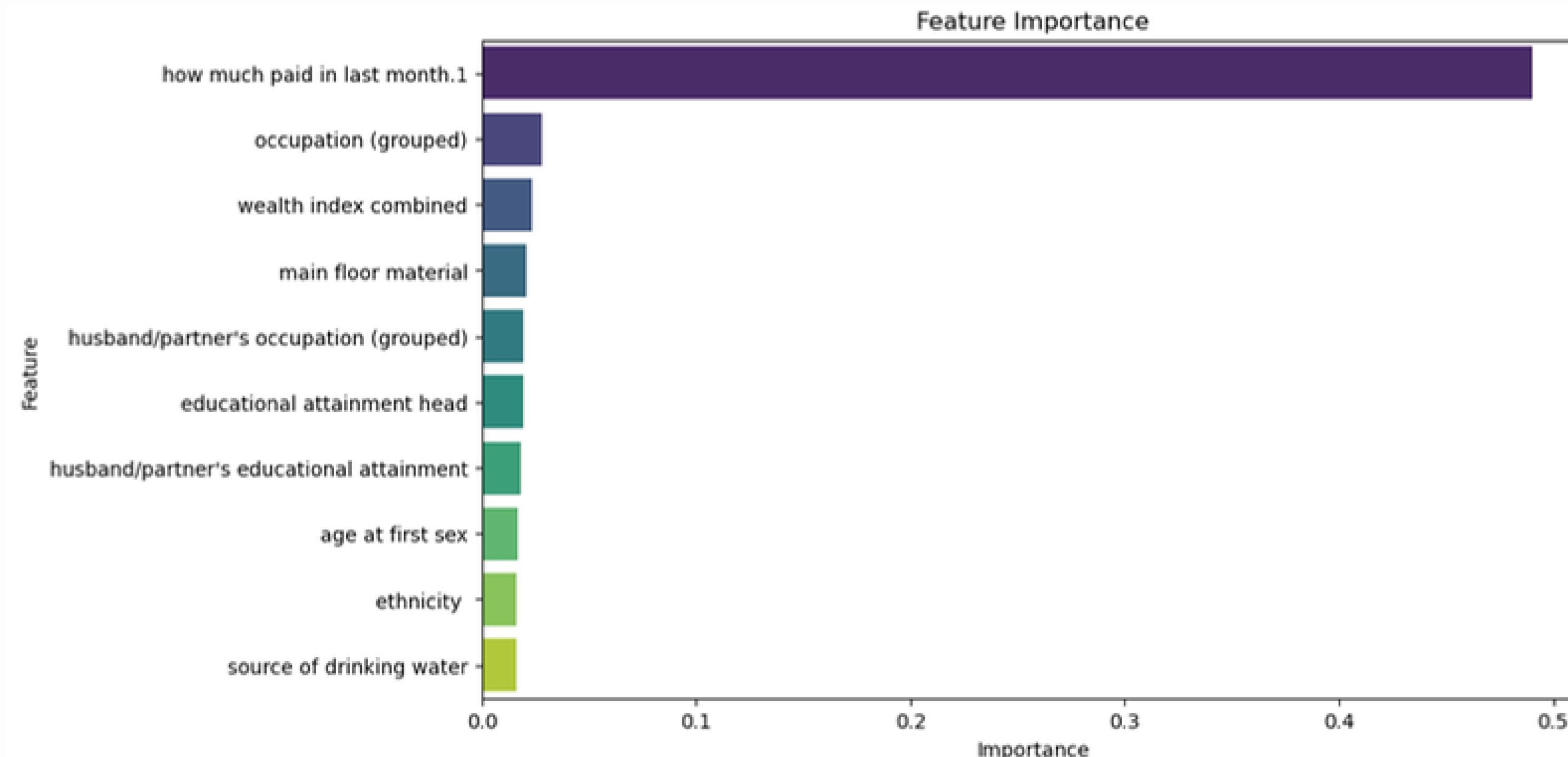
Models Accuracies

Our best model, the Random Forest Classifier had the highest accuracy of 99.38%, while the Ridge Classifier had the lowest at 60.12%, with other models showing varying performance levels and accuracy



Feature Importance

How much the respondents in the last month is the most important determinant on how much they will contribute to SHIF



Recommendations

- 01** **Targeted Educational Programs:** Implement programs to improve literacy and vocational skills among informal sector workers.
- 02** **Tailored Social Programs:** Develop social programs targeting vulnerable groups identified in the study.
- 03** **Policy Development:** Use data insights to inform policy decisions addressing the unique needs of informal sector workers.
- 04** **Enhanced Data Collection:** Improve data collection processes to capture comprehensive information on informal sector workers.



Next Steps

01

Model Refinement: Continue refining and optimizing the models for better performance

02

Comprehensive Evaluation: Evaluate models using additional metrics to ensure robustness.

03

User Engagement: Develop and deploy user-friendly tools for wider accessibility

04

Ongoing Data Collection: Establish continuous data collection mechanisms to keep the model updated

THANK YOU!

