HIV PREDICTION MODEL

Targeting HIV risk reduction among Adolescent Girls and Young Women(AGYW) in Kenya

PROJECT OVERVIEW

38 M

Global HIV cases

Individuals living with HIV worldwide

62 %

Women in africa

New infections in Sub-Saharan Africa are women

78%

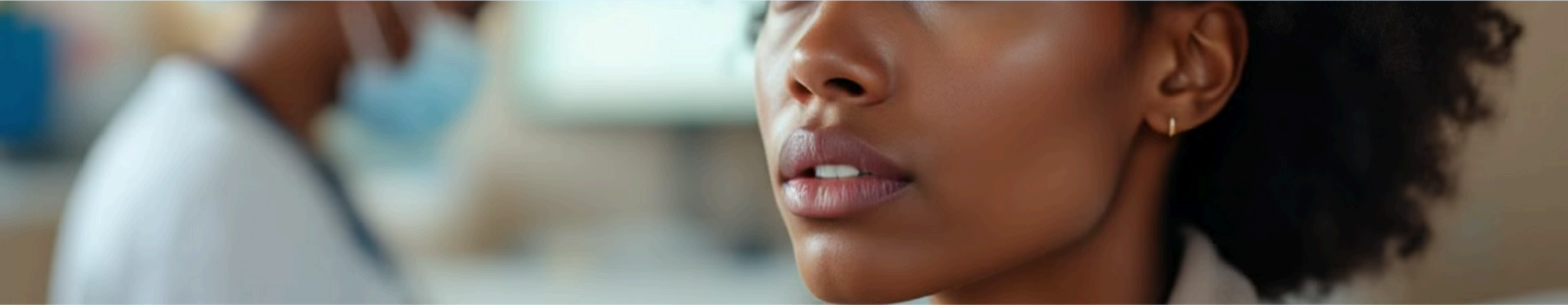
AGYW Impact

New HIV infection among AGYW

7,307

Kenya Youth Cases

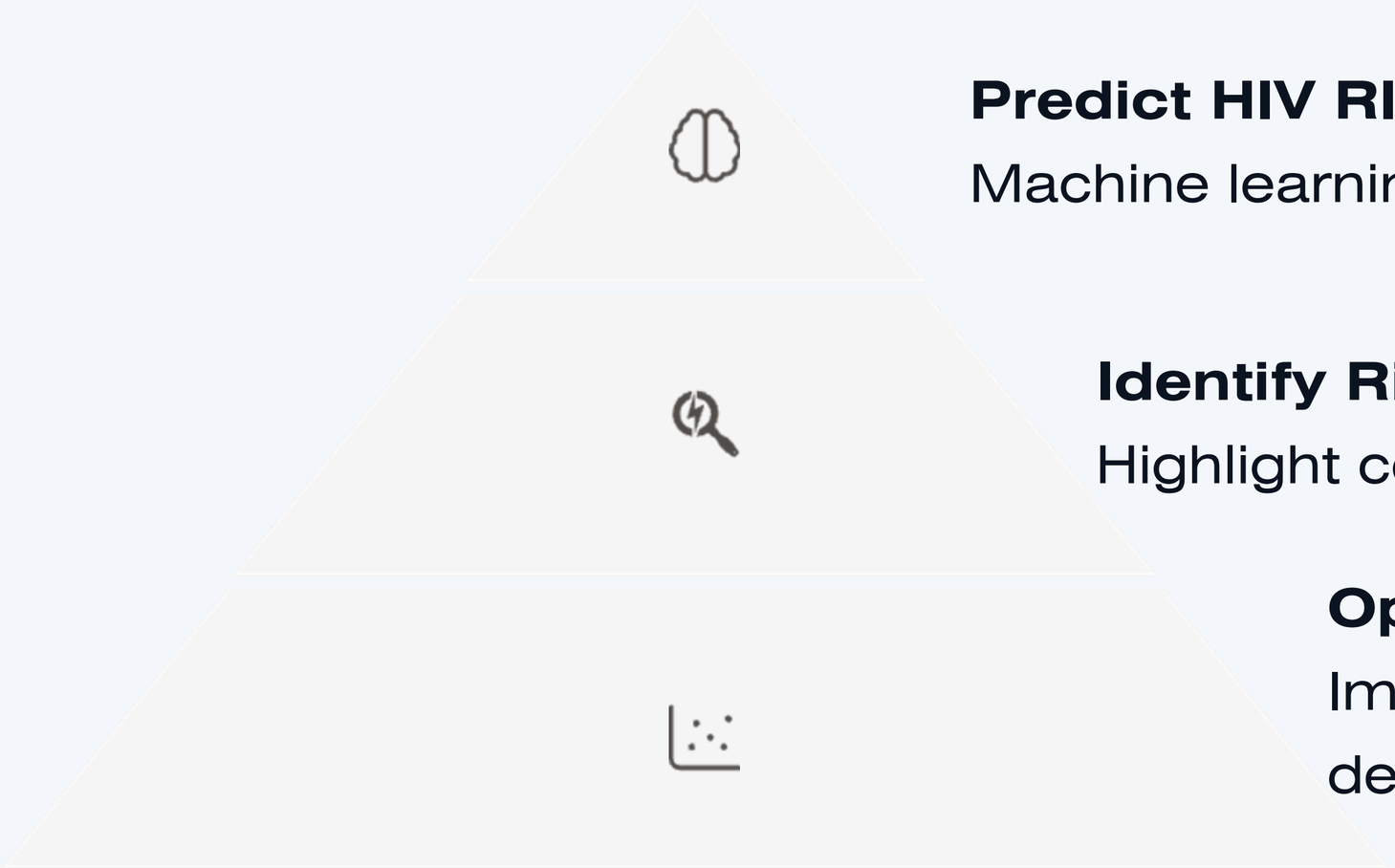
New infections ages 15-24 in 2022



PROBLEM STATEMENT

- Adolescent girls and young women (AGYW) face a **higher risk of HIV** due to biological, behavioral, and socioeconomic factors.
- New HIV infections among AGYW **persist despite significant investments.**
- **Urgent targeted interventions** are needed to tackle the underlying risk factors.

PROJECT GOAL



Predict HIV RISK among AGYW

Machine learning model based on multiple factors

Identify Risk Factors

Highlight contributors to vulnerability

Optimize Resources

Improve Program targeting and delivery

DATA UNDERSTANDING

- The project analysed data from 455,807 adolescent girls and young women (AGYW) in an HIV prevention program from 2015 to 2023.
- Data was collected during program intake and routine follow-ups.
- The dataset includes demographics, socioeconomic status, education and behaviour, HIV testing and status, HIV program interventions, and exit status.
- Privacy was maintained, no personally identifiable information (PII) was used, and the dataset was deidentified.

DATA PREPARATION

Completeness

Numerical columns imputed with median; categorical with mode and unknown

Uniformity

Feature engineering of dates, age groups

Feature selection was done using domain knowledge and significance

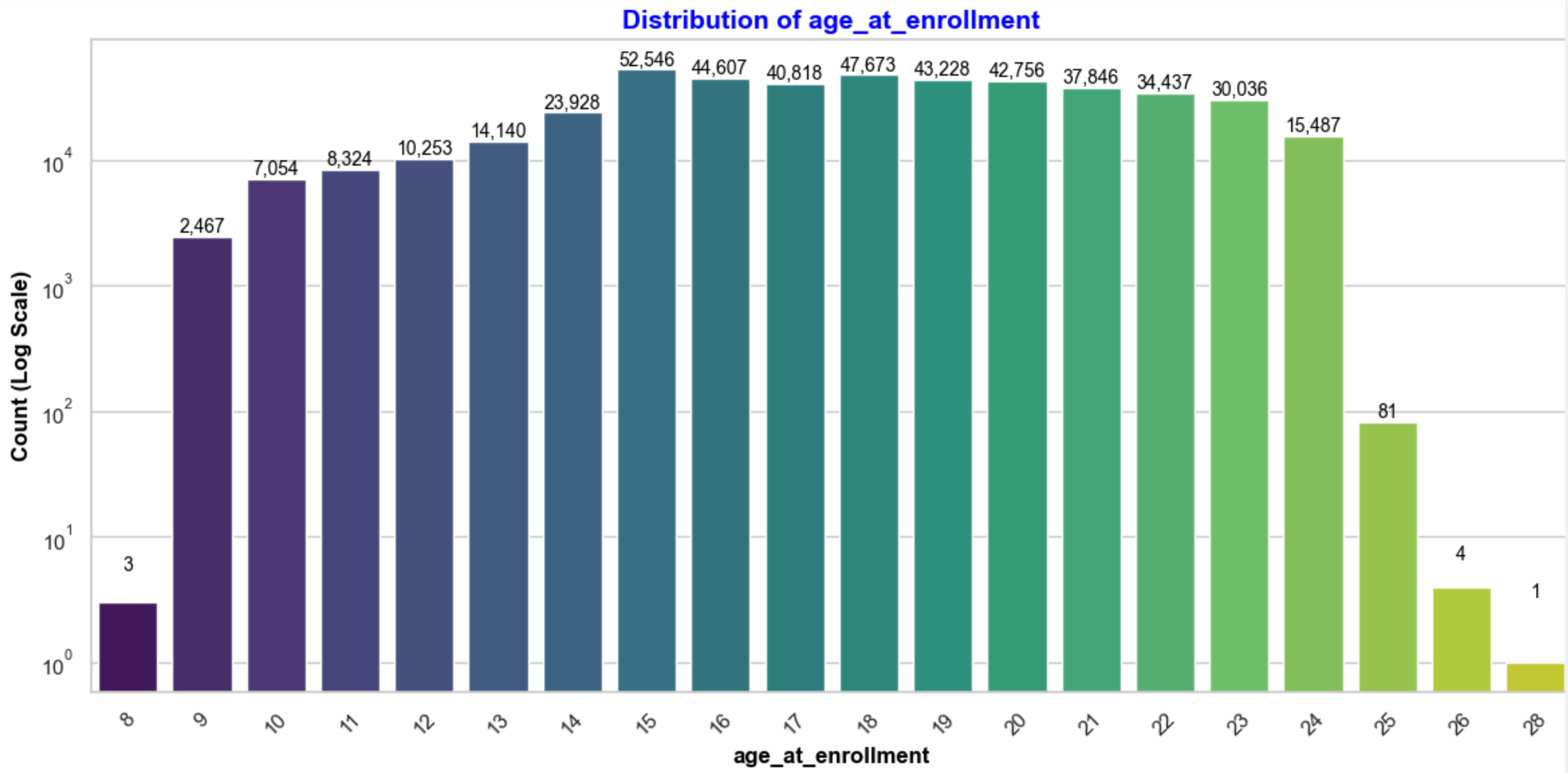
Accuracy

Outliers replaced with median

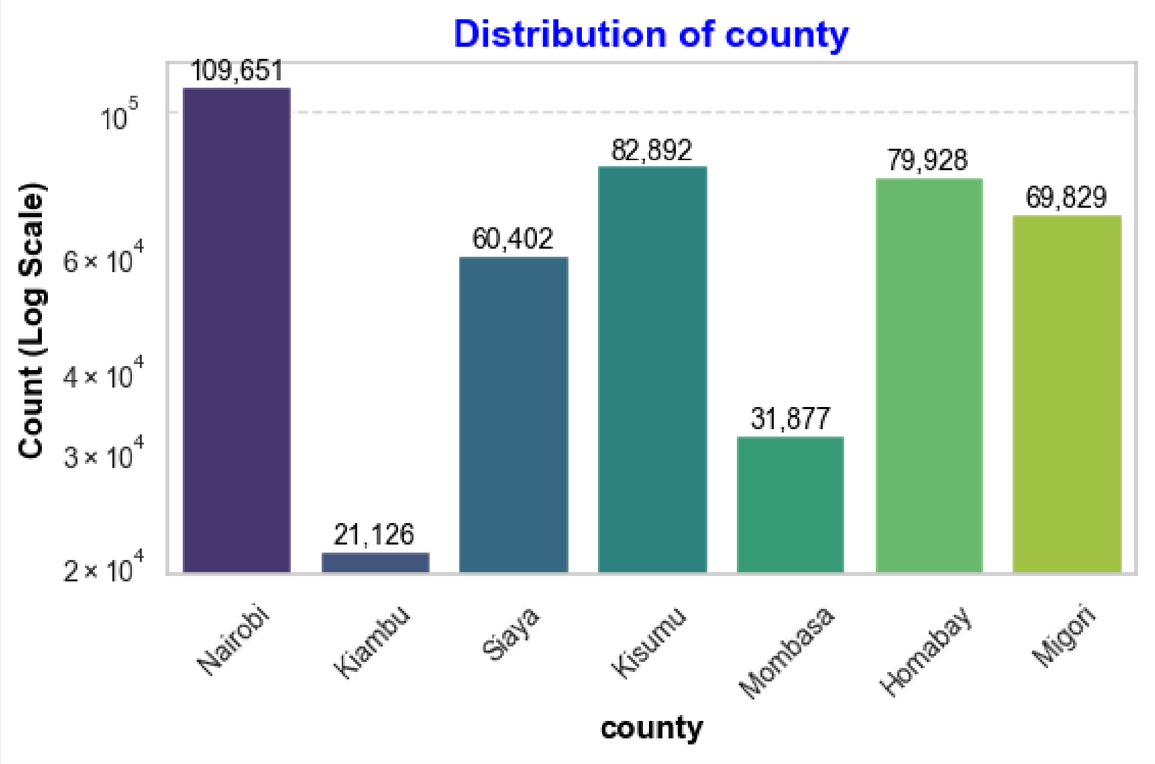
consistency

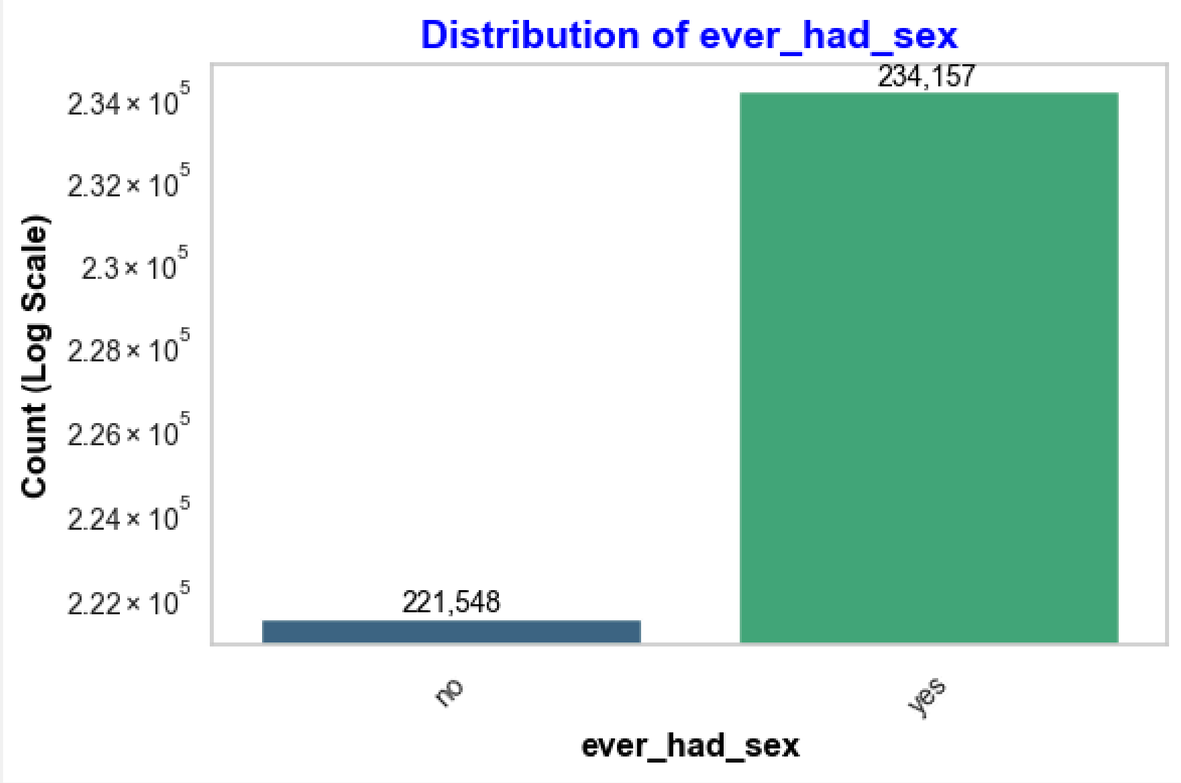
Dropped duplicates

EXPLORATORY DATA ANALYSIS



- Enrollment of adolescent girls and young women with HIV risk was highest between ages 14 and 23, with a peak at age 15 (52,546 individuals).

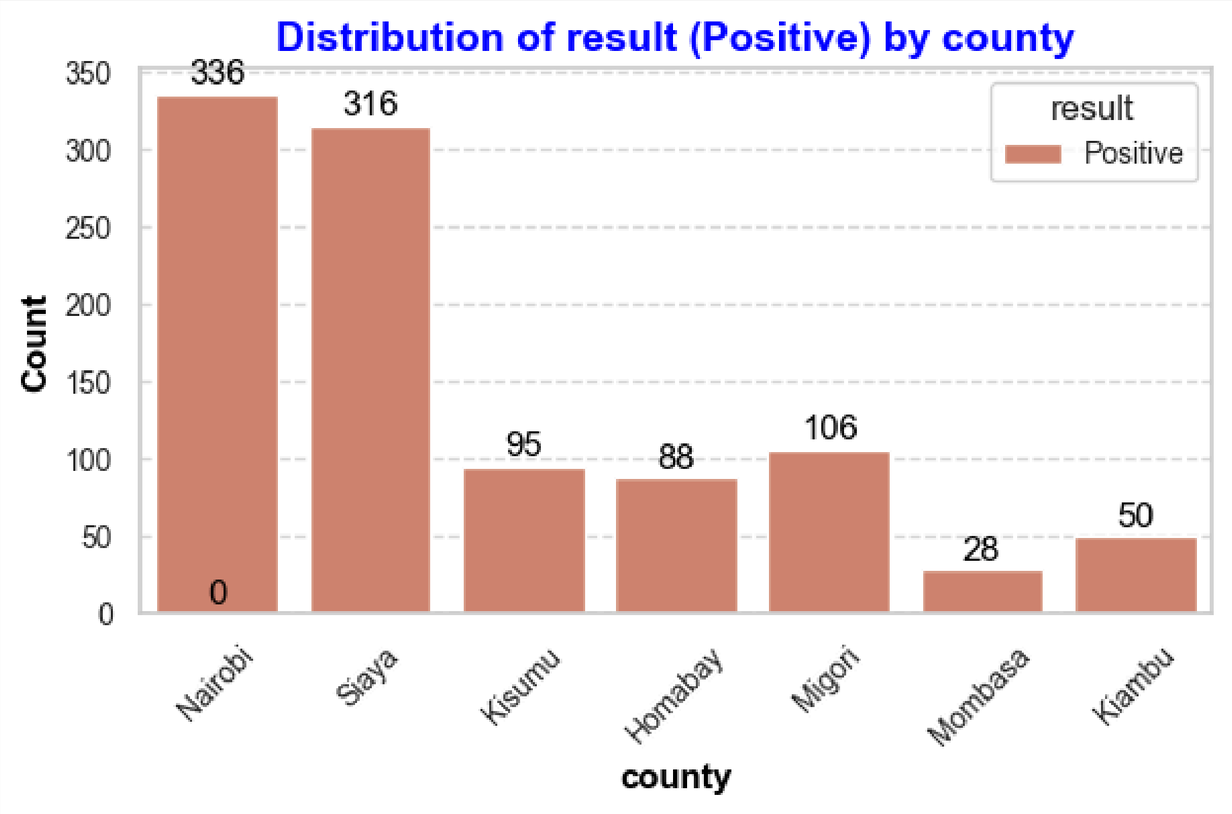


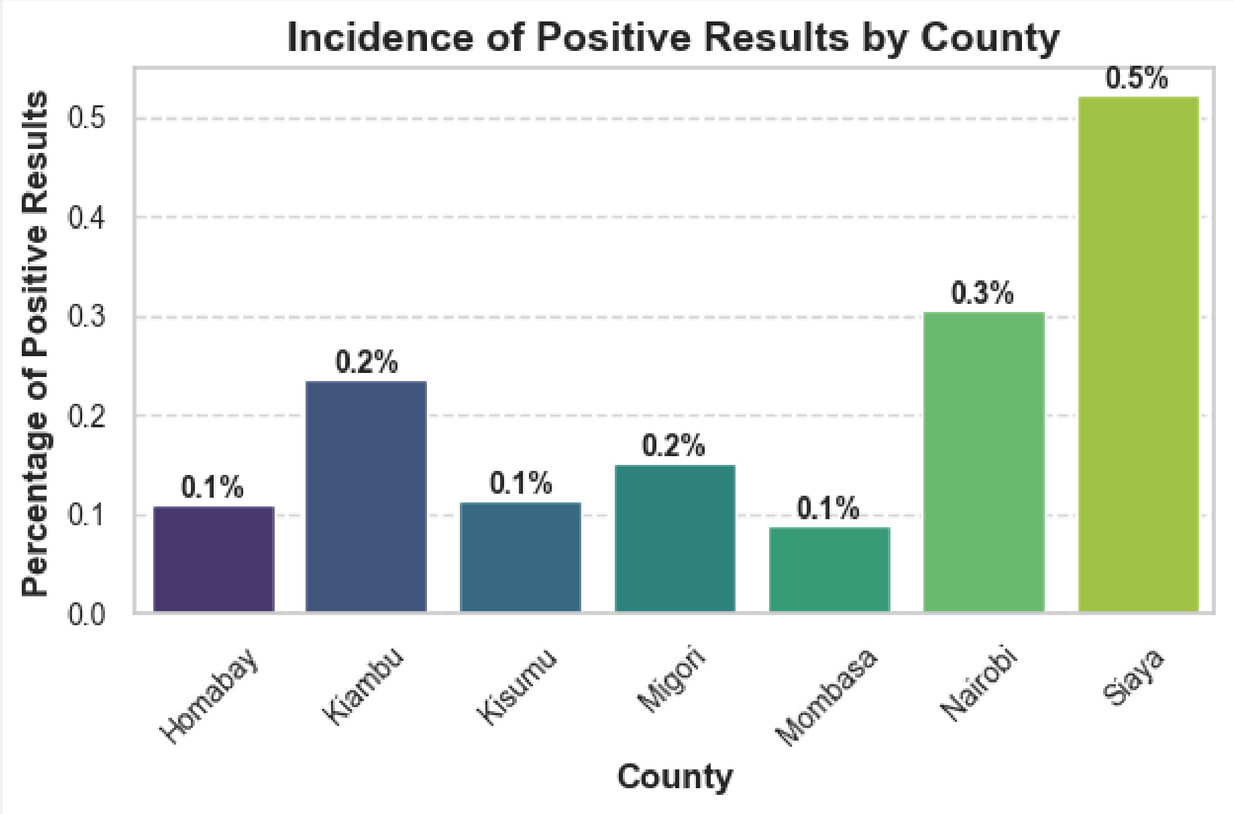


- Nairobi enrolled the highest number of AGYW in the program

Most of the AGYW enrolled in the program had ever had sex e

KEY INSIGHTS

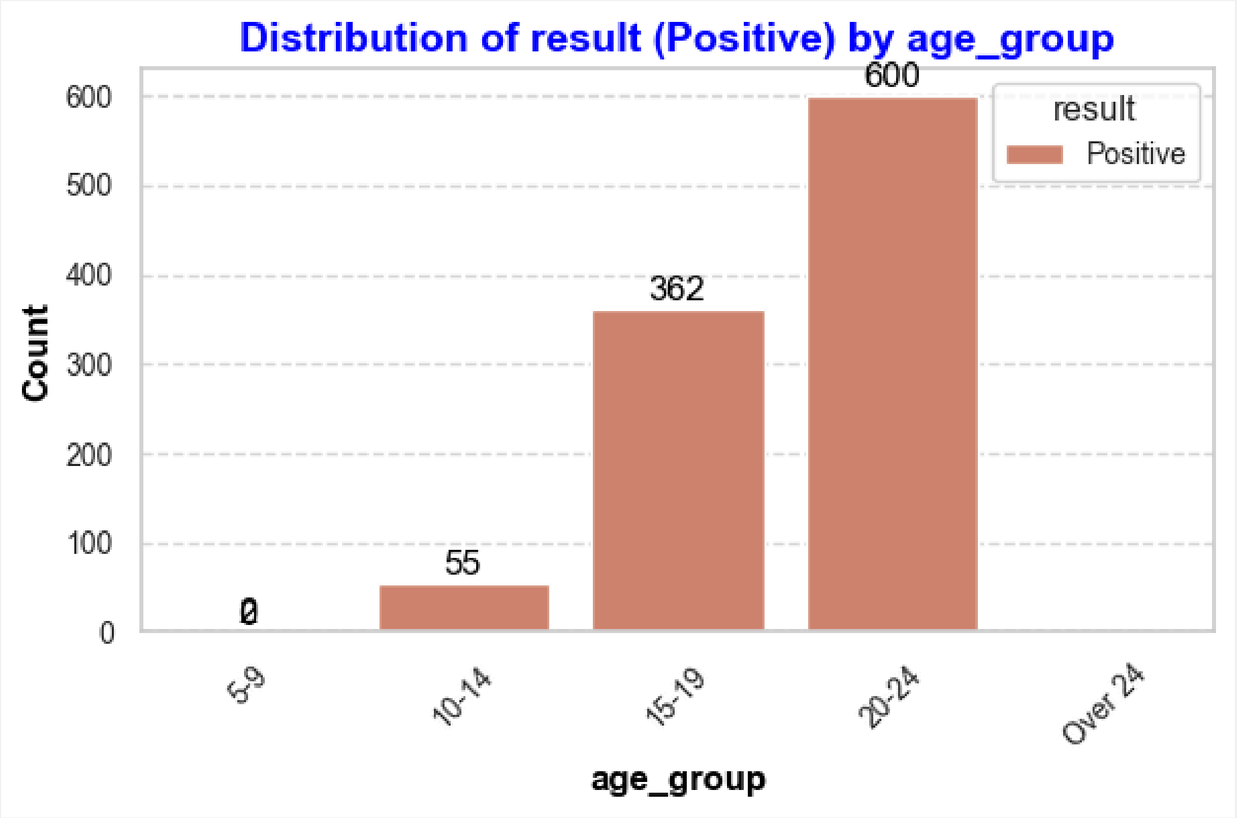


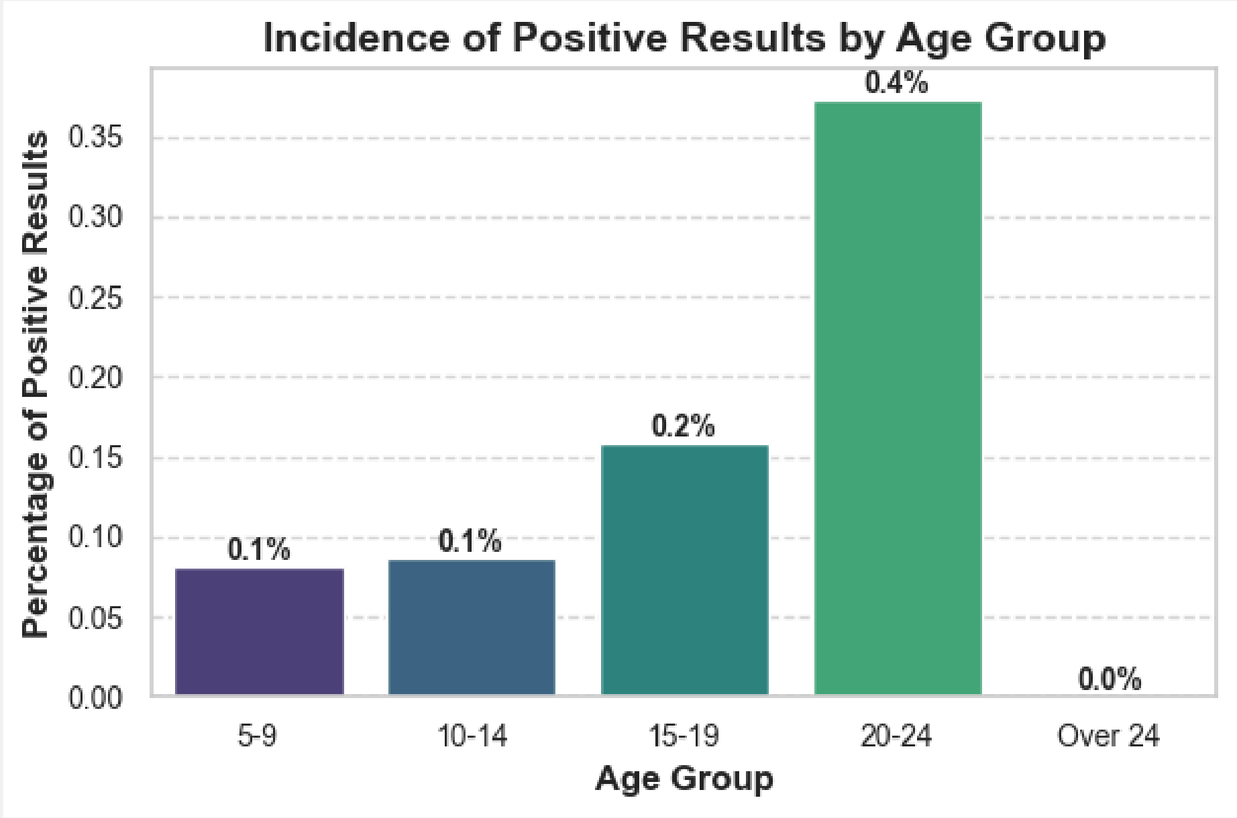


- Most of the girls who had a positive HIV infection were in Nairobi

- The highest incidence of HIV among the enrolled AGYW was in Siaya

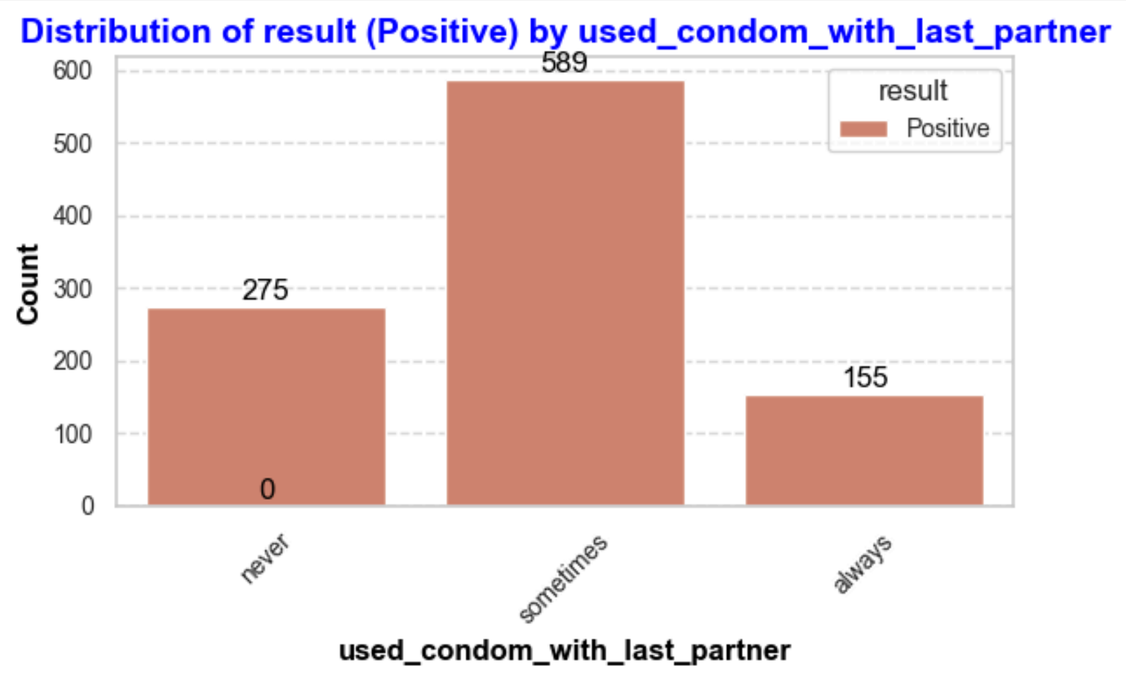
KEY INSIGHTS

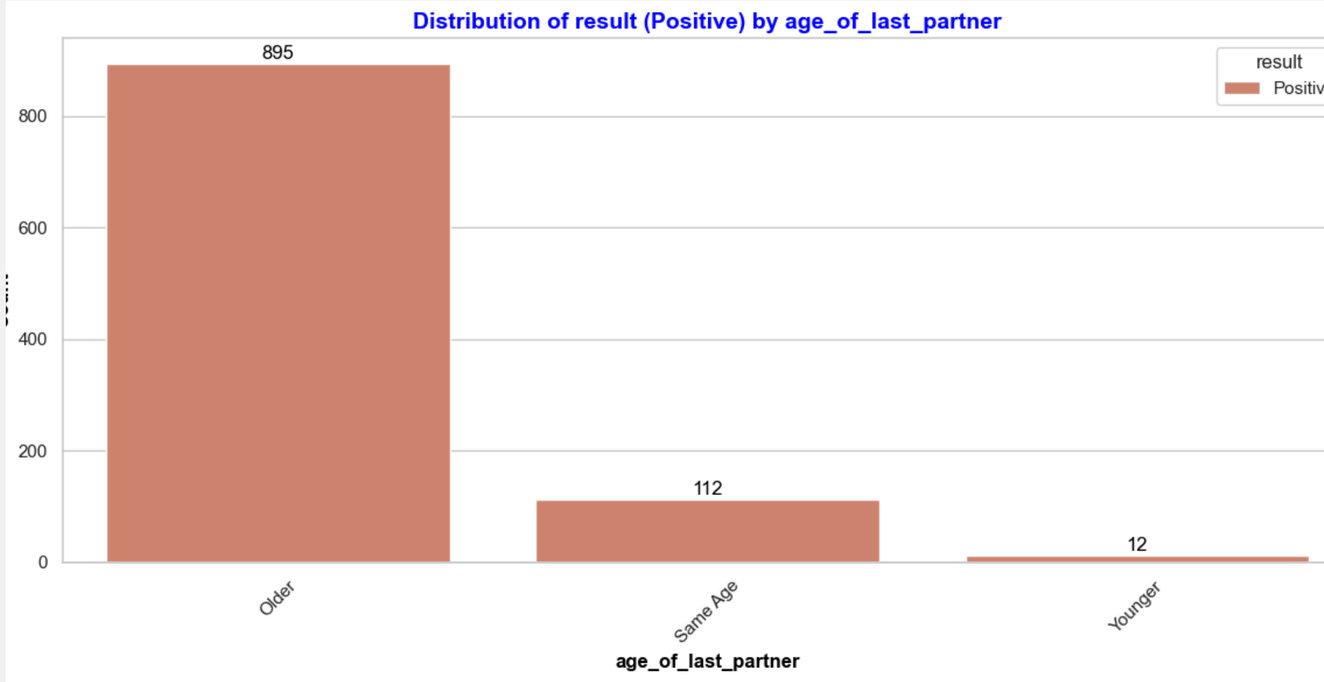




- Most of the young women who had a positive HIV infection were aged between 20-24 years old; they also had the highest incidence of HIV ($p < 0.05$)

KEY INSIGHTS

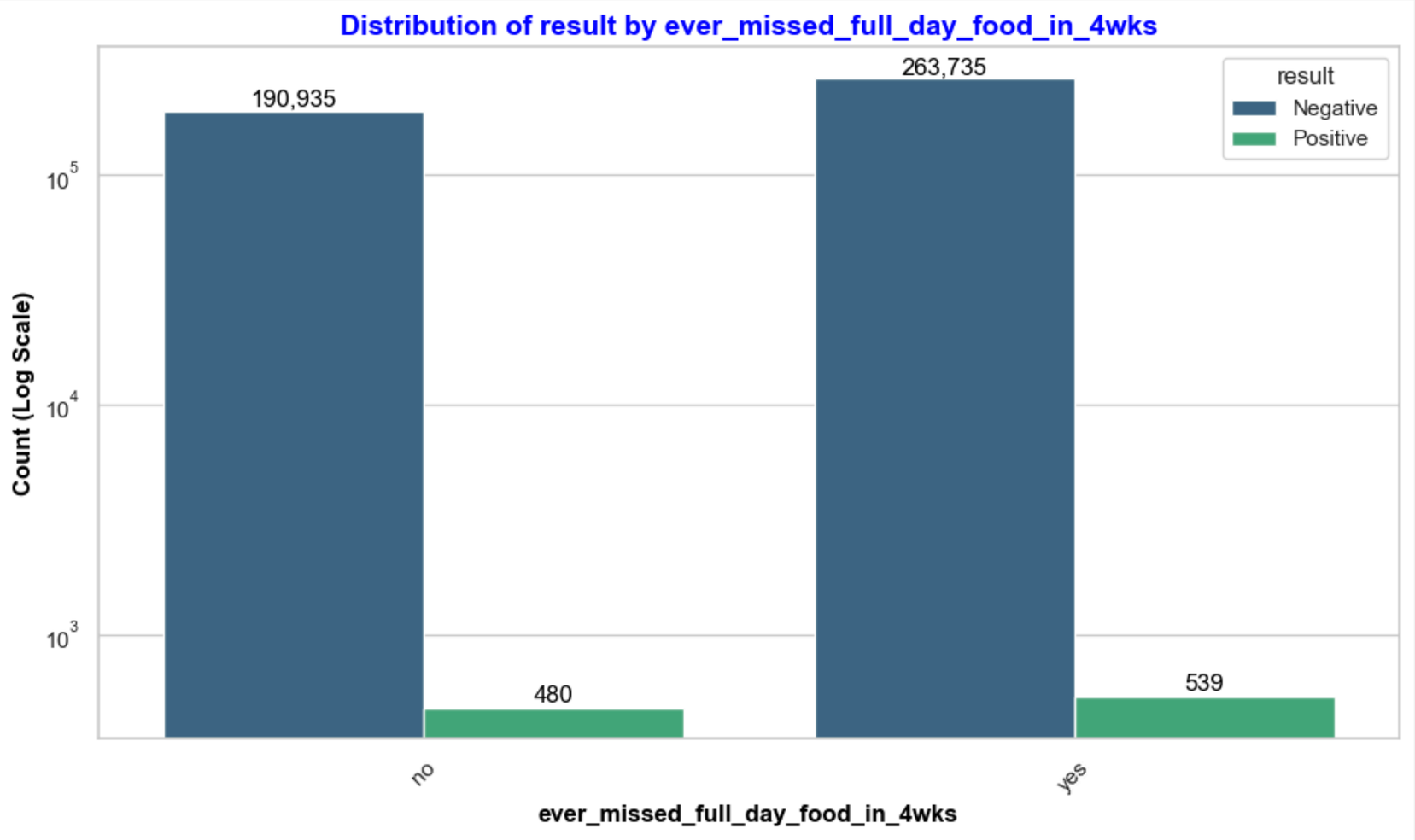




- Most of the young women who had a positive HIV infection did not use condom consistently ($p < 0.05$)

- Most of the young women who had a positive HIV infection had sex with partner who was older ($p < 0.05$)

KEY INSIGHTS



There is a statistical significance between food inadequacy and HIV test result ($p < 0.05$).

KEY RISK FACTORS

Education Level

Lower education correlates with higher risk

Food security

Insecurity increases vulnerability

Geographic Location

County-specific risk variations

Partner Age

Older partners linked to higher infection rates

MODEL EVALUATION

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	69.61%	68.88%	68.18%	68.53%
Random Forest Classifier	69.85%	62.23%	68.18%	68.78%
XGBoost	66.67%	66.64%	63.64%	64.95%

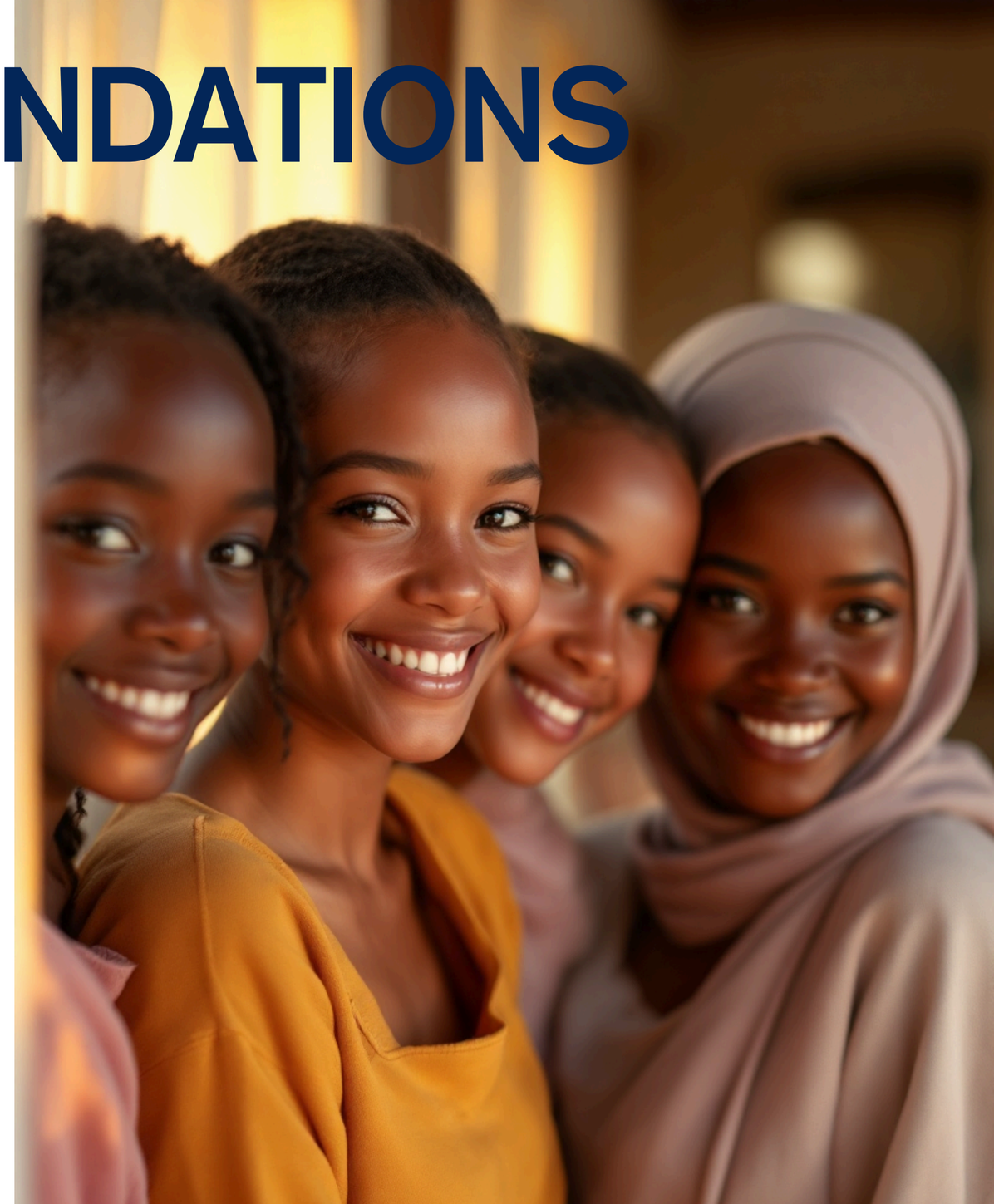
RECOMMENDATIONS

Enhance Age-Specific Interventions

Improve Education Access

Address Food Security

Implement County Specific Interventions



NEXT STEPS

- Perform model blending/hybrid
- Deploy in a controlled environment
- Refine and scale deployment

.