



Understanding financial inclusion in East Africa

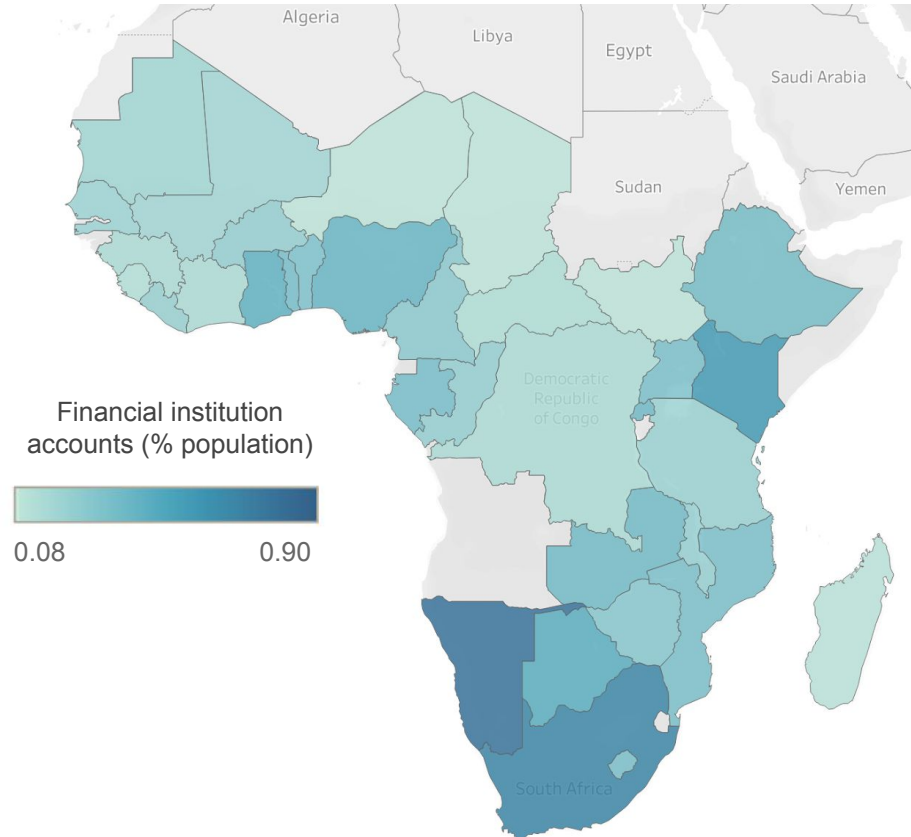
Predicting the unbanked population in
Kenya, Tanzania, Uganda and
Rwanda

Tawney Kirkland
10 February 2021

Agenda

- Introduction
- Methods
- Data
- Findings
- Insights

Across Sub Saharan Africa, formal financial inclusion is low



Given this reality, the project aimed to address the following objective

Predict who is **unbanked** in order to identify the most important features linked to financial inclusion

Both quantitative and qualitative methods were deployed as part of this investigation

Quantitative methods

- Exploratory data analysis of [FinScope National Surveys](#) and the [World Bank Global Findex database](#)
- Predictive analysis on FinScope National Surveys using logistic regression ¹

Qualitative methods

- Review of literature on financial inclusion in East Africa
- Consultation with a financial inclusion expert in Kenya

1. Logistic regression with oversampling and Naive Bayes was also tested but the simple logistic regression model was prioritized given performance / the focus on interpretability of results. See appendix for results and tools used for the analysis

FinScope National Surveys data provided a range of features used to predict bank account holding

Base features

- Age of respondent
- Gender
- Household size
- Cellphone (y / n)
- Location type (urban / rural)
- Position in household
- Marital status
- Level of education
- Primary source of income

Target:
Bank account
(no / yes)

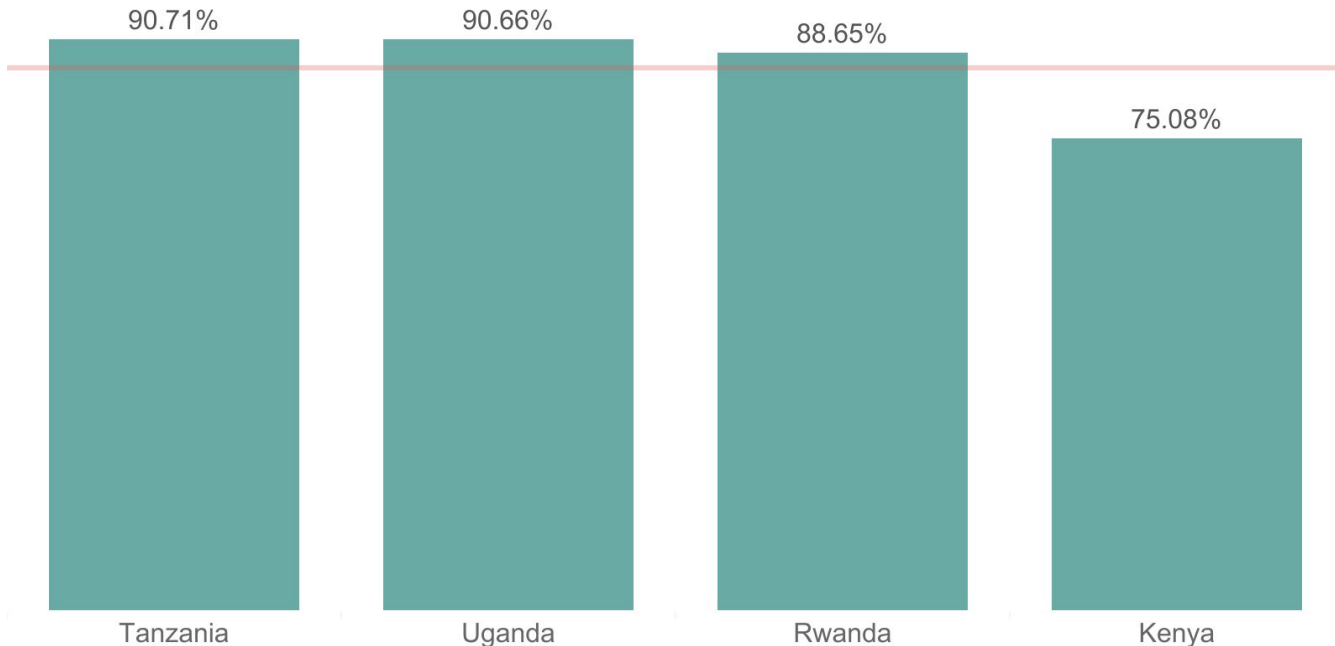


Findings from the EDA are consistent with broader literature on the state of financial inclusion in SSA

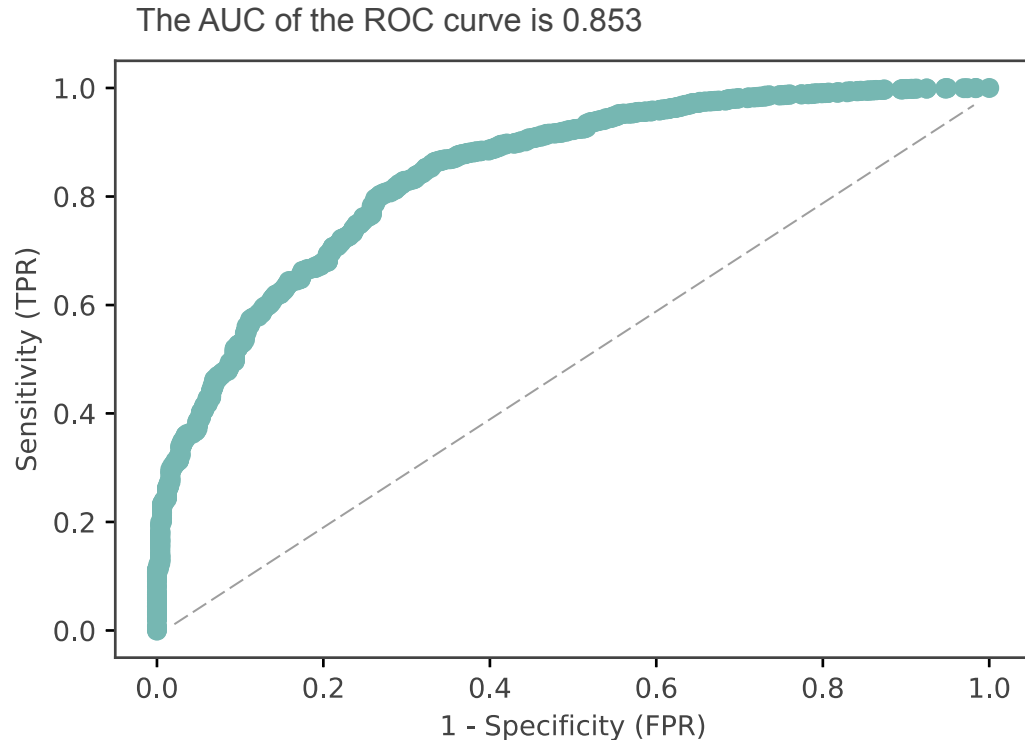
86%

of respondents
are unbanked

Tanzania and Uganda have the highest share of unbanked population in the sample



A logistic regression model was used to identify the unbanked population, with strong results



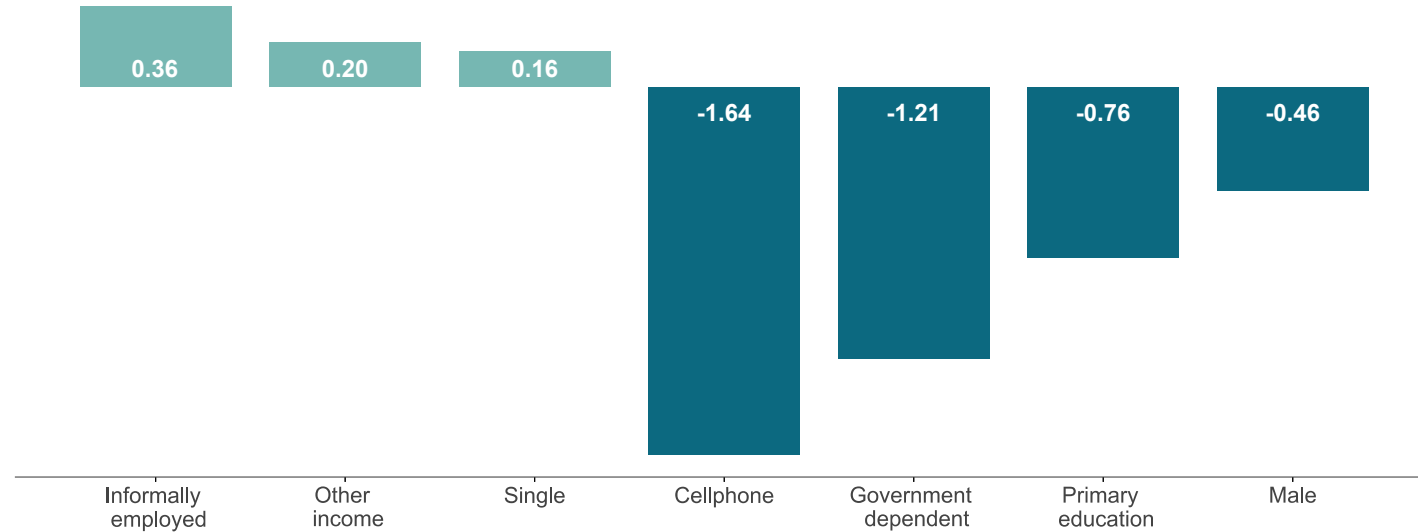
Results

- Recall: 0.97
- Precision: 0.90
- F1: 0.94
- Accuracy: 0.88

Highly confident, false predictions linked to lower levels of education and primary source of income¹

This model identified some of the strongest features in predicting the unbanked

Positive regression coefficients increase the odds of being unbanked, while negative coefficients decrease the odds of being unbanked



The results provide important insights on opportunities to advance financial inclusion

Unbanked

Informally employed single women have higher odds of being unbanked - target through DSDs and WEE

Banked

- Individuals who have completed primary school / government dependents have higher odds of having a bank account but may benefit from financial education
- Individuals with a cell phone are more likely to be banked - consider extending financial products through mobile channels



Appendix

- Tools
- Algorithm
- Dashboard

Tools

Data processing

- Python
- Pandas
- Numpy

Modeling

- Sklearn
- statsmodels

Visualization

- Matplotlib
- Seaborn
- Tableau



Coefficients for the selected logistic regression algorithm

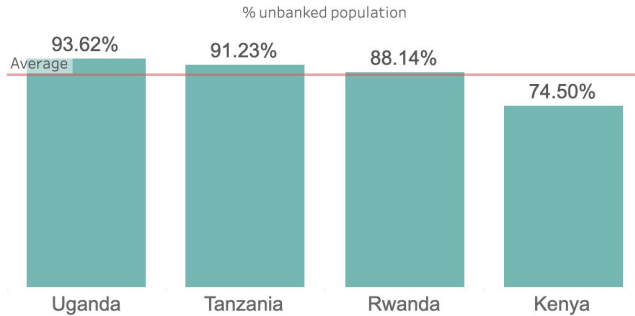
-0.013	age_of_respondent
-0.273	is_urban
-1.636	cellphone
-0.463	is_male
0.336	country_Rwanda
1.275	country_Tanzania
1.207	country_Uganda
-0.766	relationship_with_head_Head of Household
-0.551	relationship_with_head_Spouse
0.157	marital_status_Single/Never Married
-0.035	marital_status_Widowed
-0.764	education_level_Primary education
-1.765	education_level_Secondary education
-3.234	education_level_Tertiary education
-2.776	education_level_Vocational/Specialised training
-1.614	job_type_Formally employed Government
-1.730	job_type_Formally employed Private
-1.212	job_type_Government Dependent
0.360	job_type_Informally employed
0.189	job_type_No Income
0.202	job_type_Other Income
-0.672	job_type_Remittance Dependent
-0.462	job_type_Self employed
5.608	const

Selected algorithm versus tested algorithms

		Model	Validation accuracy	Test accuracy	Precision	Recall	F1
Selected		Logistic regression	0.87	0.88	0.90	0.97	0.93
		Logistic regression with random oversampling	0.79	0.79	0.95	0.80	0.87
		Naives Bayes	0.88	0.87	0.91	0.97	0.93

Dashboard

Uganda has the highest unbanked population as a percent of total respondents



Self employed, Informally employed and farming and fishing were most frequently reported as the main source of income



The majority of the highly confident, false predictions (~56%) were individuals with lower levels of education, whose primary income was from farming and fishing, self employed or informally employed

