Financial Incluion in Africa

Agenda

- Our team
- Client case
- Hypotheses
- Target Customers profile
- Our model

Our Team

We are a Financial analytics solutions provider.

We offers tailored data analysis and machine learning solutions for financial providers.

Data scientists:

Alex Volosha

Hiroi Isomura

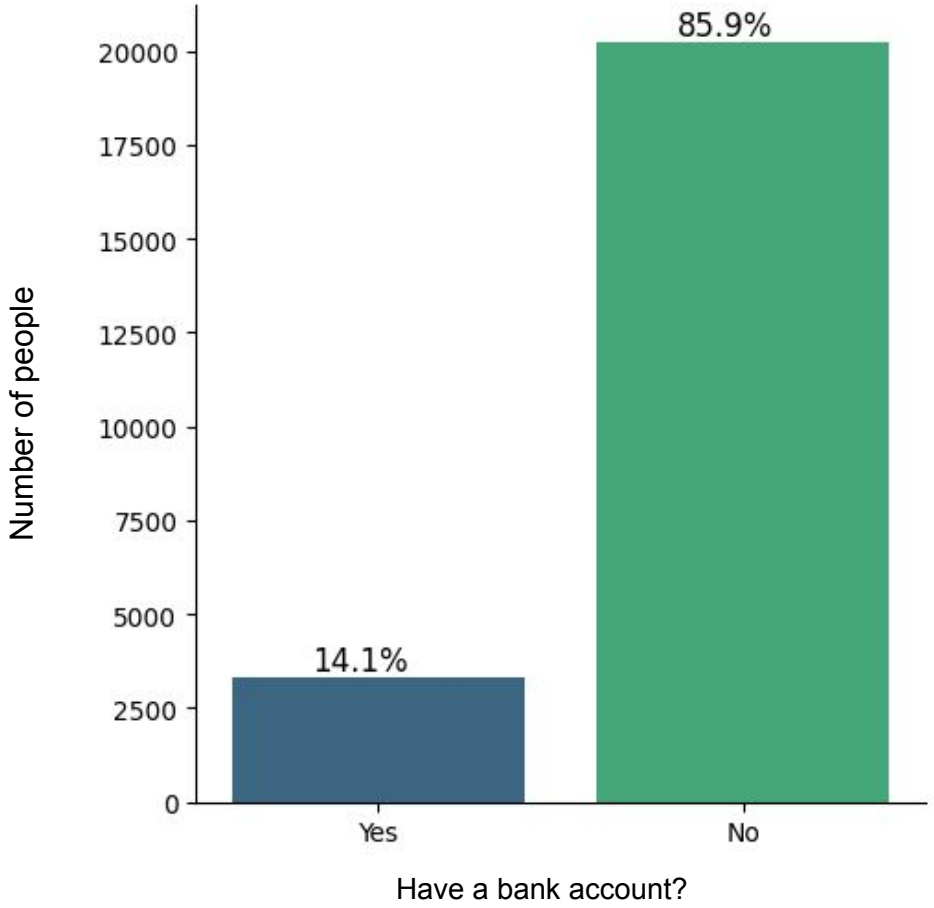
Ron Silva

Our Client Case

Who in Africa is most likely to have a bank account?

- Only around 14% of adults have a bank account in Kenya, Rwanda, Tanzania and Uganda

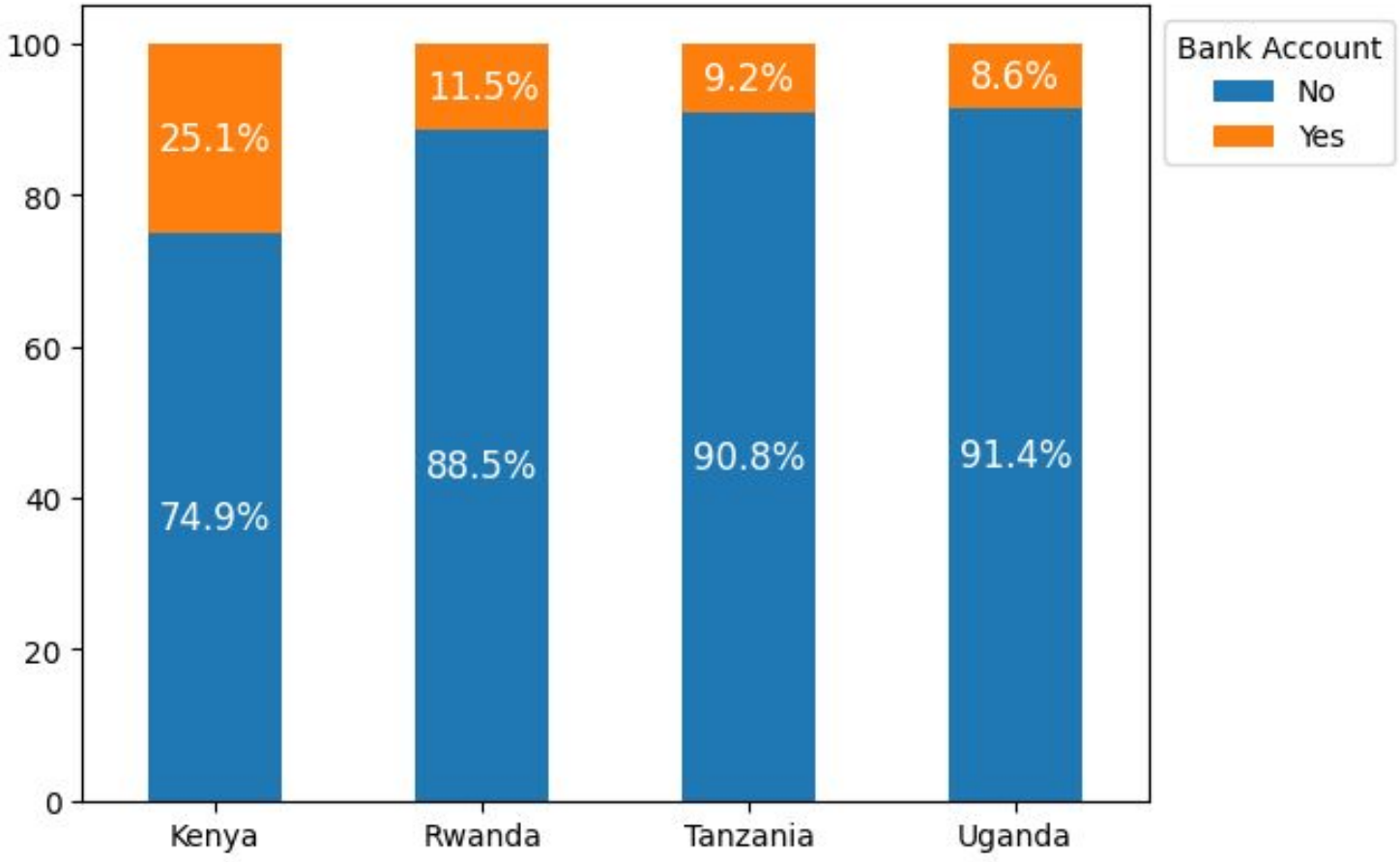
- Our goal is to develop a model that can predict which individuals are most likely to have or use a bank account.



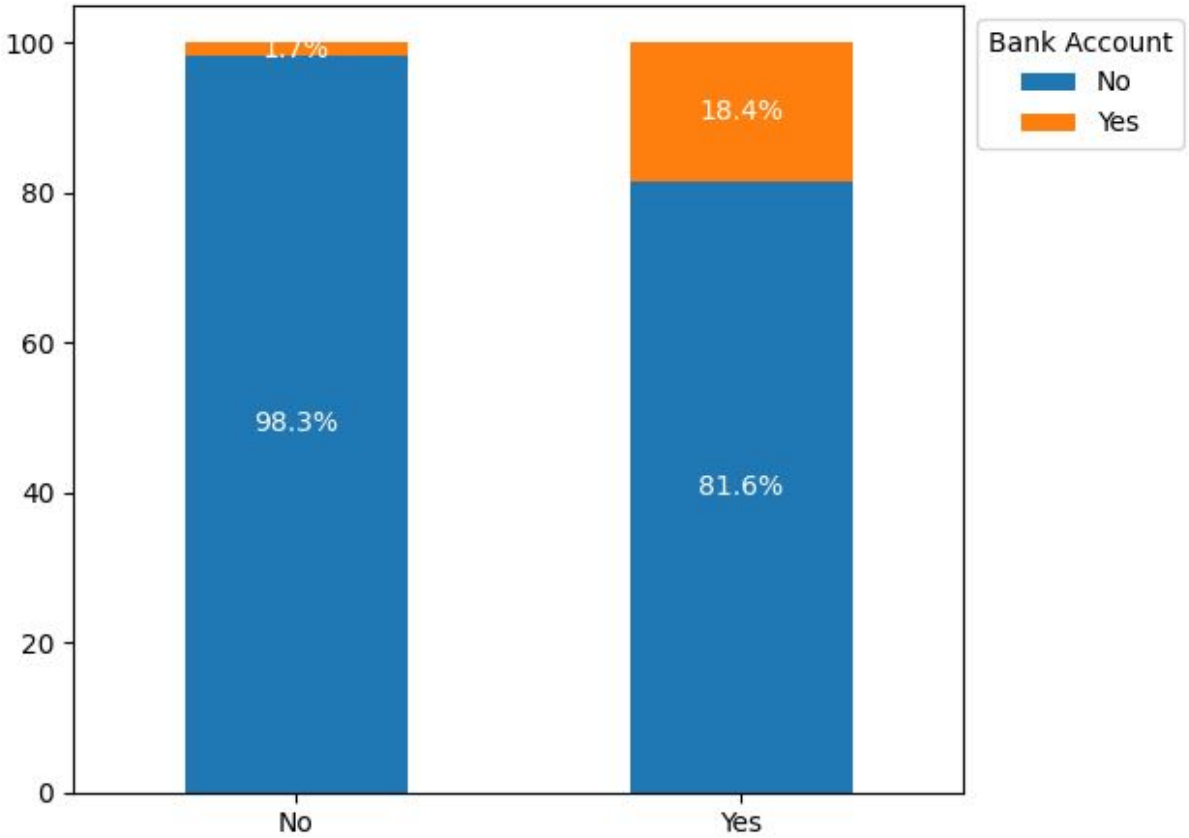
Hypotheses for bank holder’s profile

1. People from specific country has more access to bank account.
2. Most people that don't have bank account are from rural area.
3. People with cell phones have no bank account.
4. People with high education have more access to bank account.
5. People who are employed have more access to bank account.

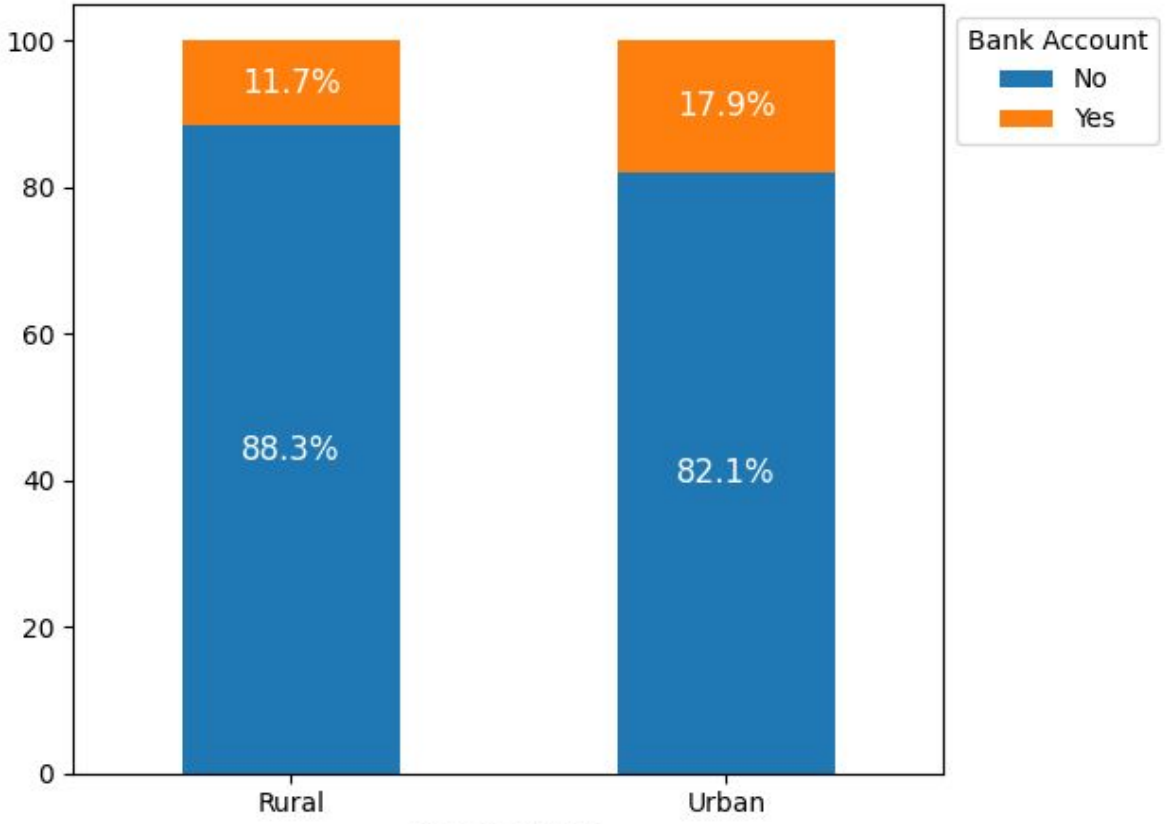
Bank Account Holder Rate per Countries



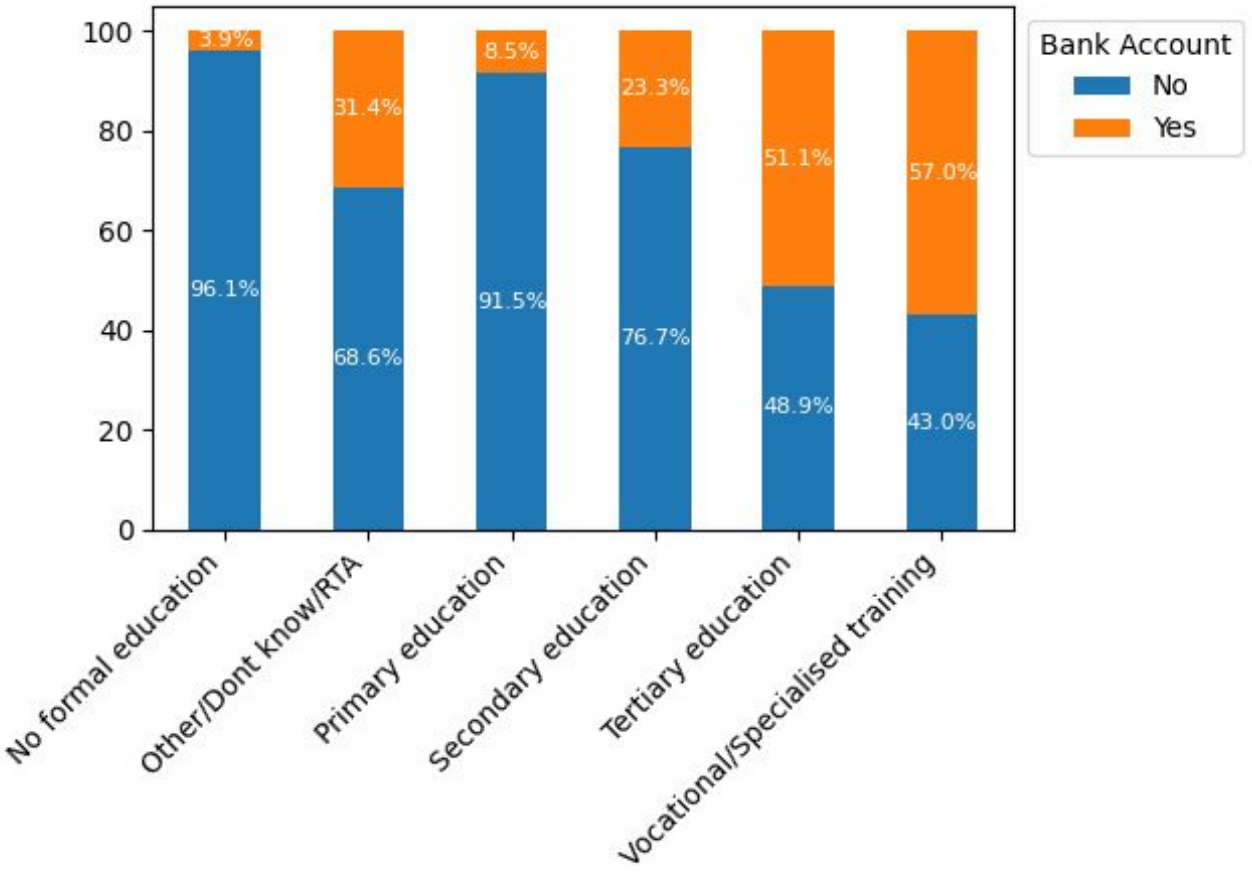
Bank Account Holder Rate per Cellphone Possession



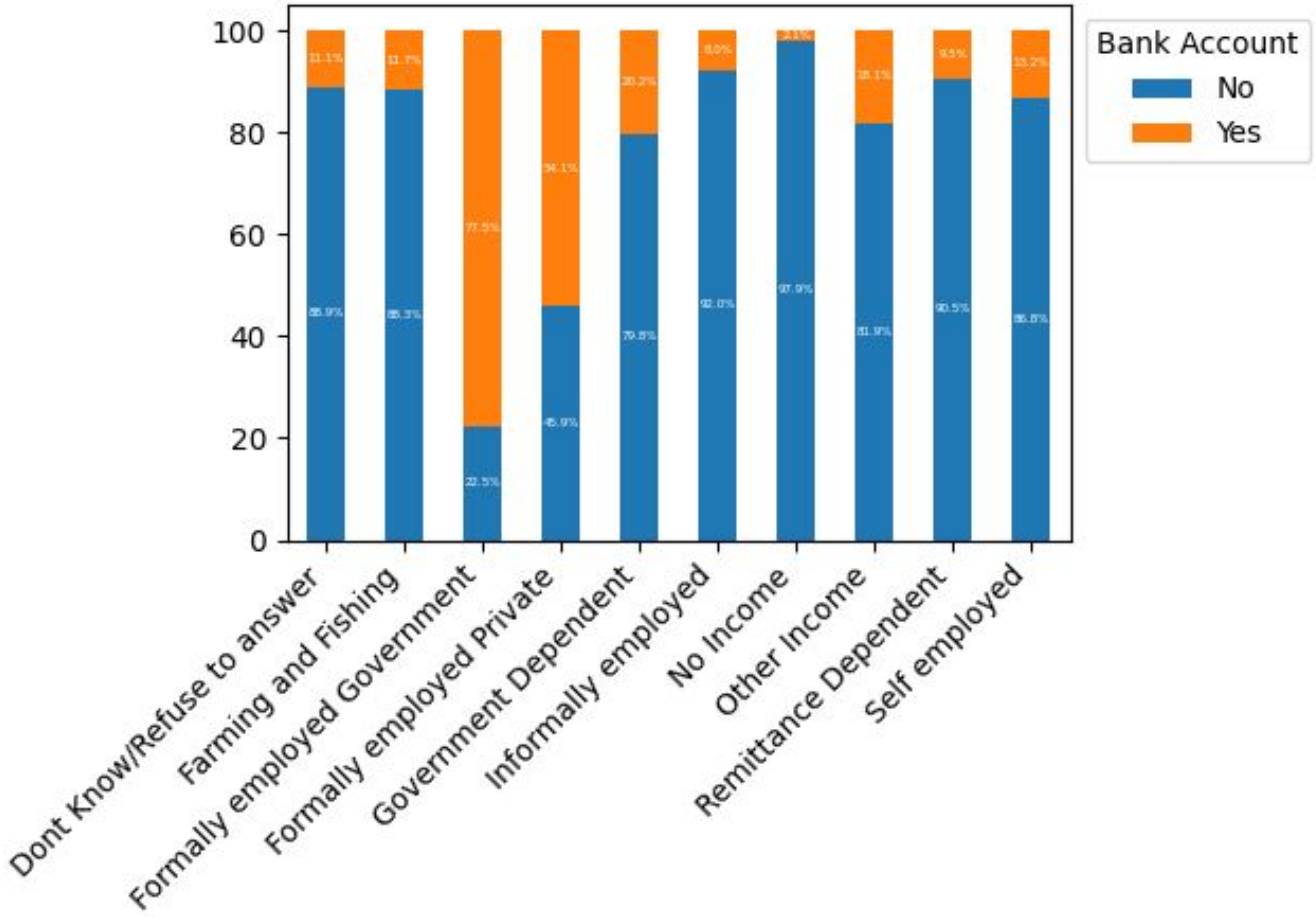
Bank Account Holder Rate per Location Type



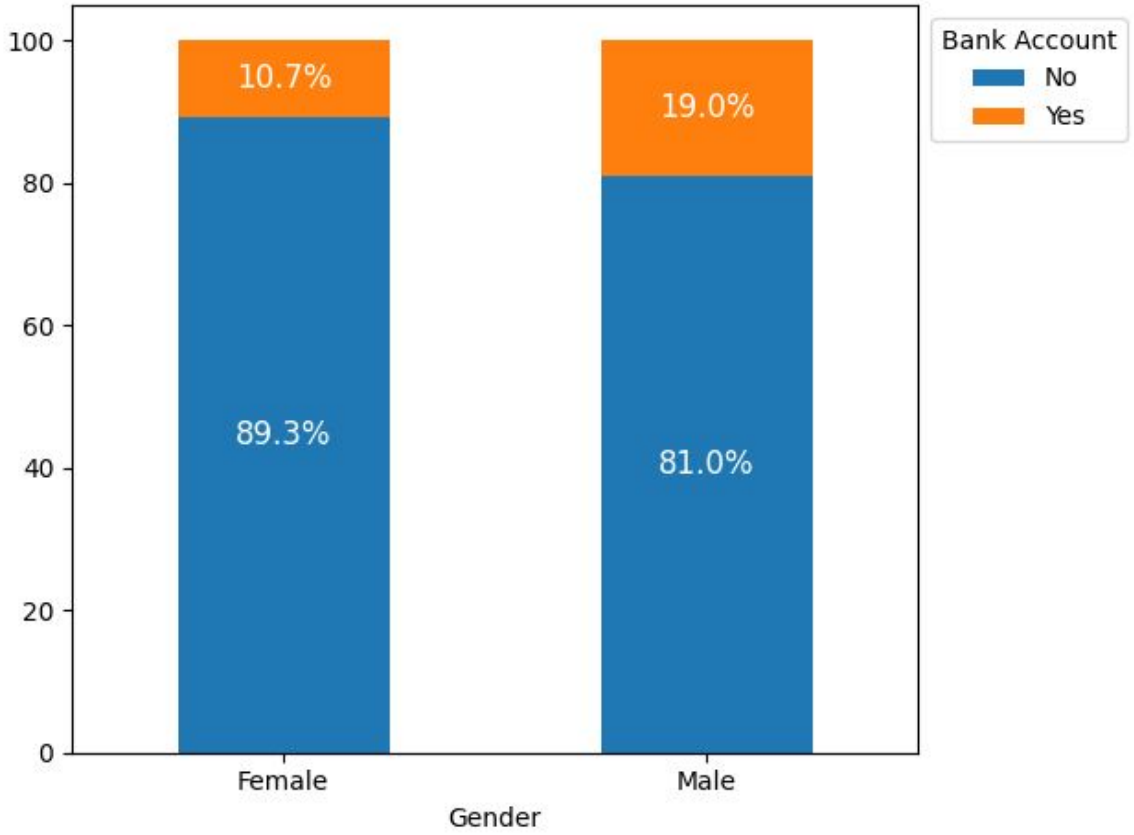
Bank Account Holder Rate per Education type



Bank Account Holder Rate per Job type



Bank Account Holder Rate per Gender

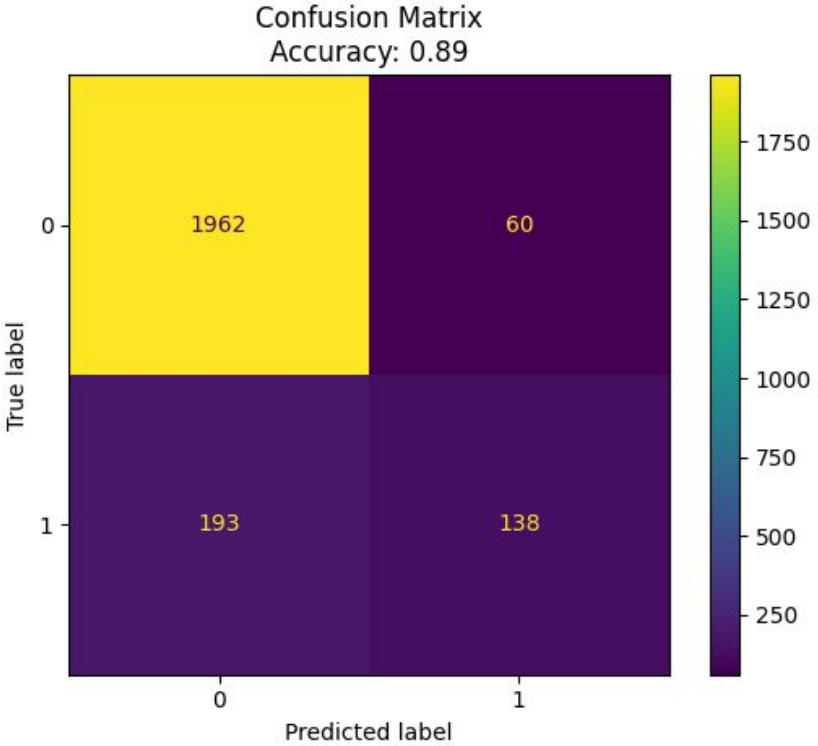


Customer portrait:

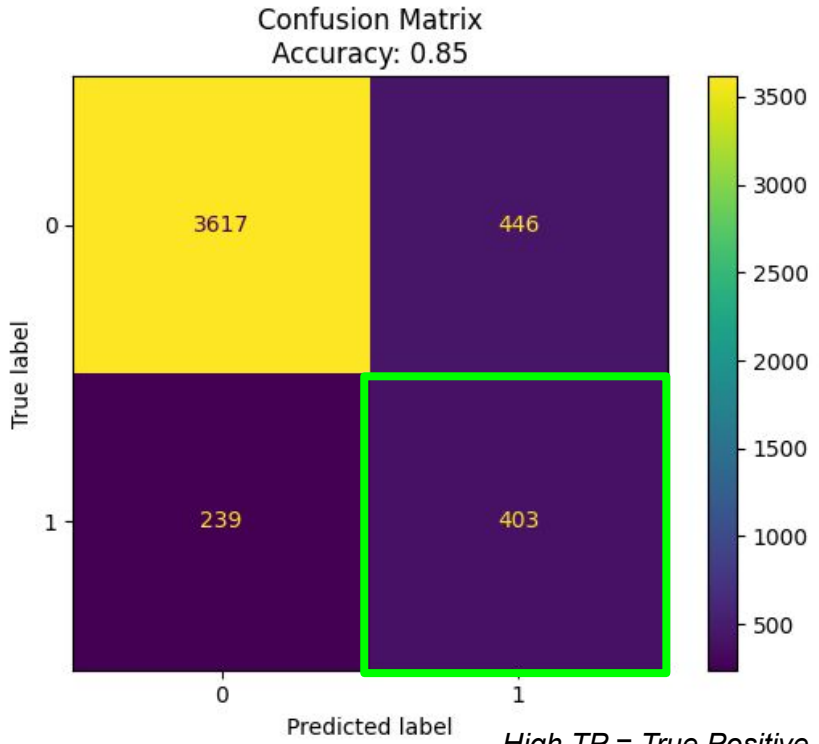
- A man, between 30 and 40 years old
- Having a cell phone
- From an urban area
- With a technical or graduation degree
- Formally employed

is more likely to have a bank account.

- Baseline model



- Final model 1



High TP = True Positive. Most importantly to capture potential clients.

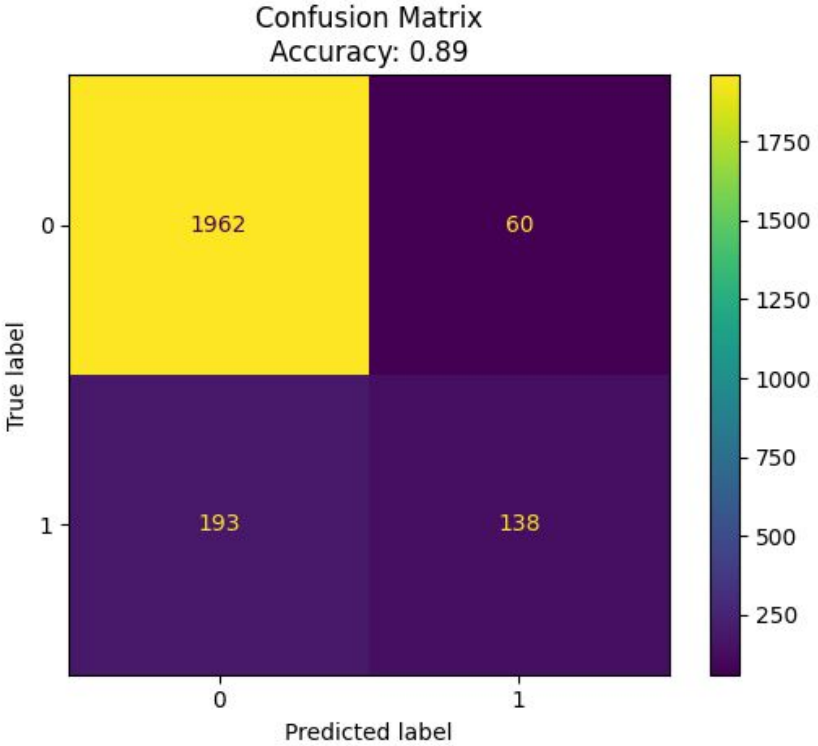
- Basis model

	precision	recall	f1-score	support
0	0.91	0.97	0.94	2022
1	0.70	0.42	0.52	331
accuracy			0.89	2353
macro avg	0.80	0.69	0.73	2353
weighted avg	0.88	0.89	0.88	2353

- Final model

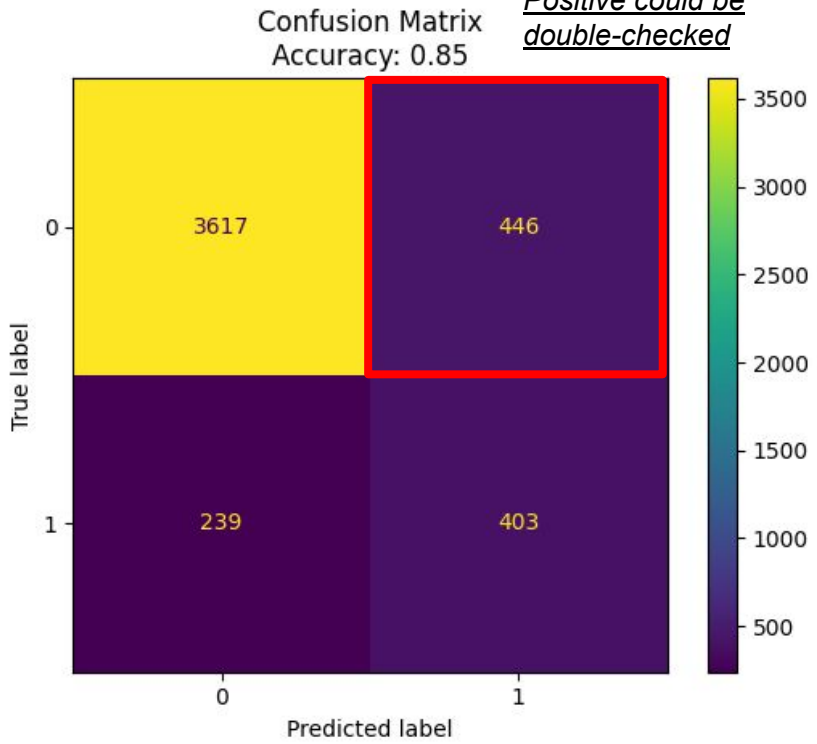
	precision	recall	f1-score	support
0	0.94	0.89	0.91	4063
1	0.47	0.63	0.54	642
accuracy			0.85	4705
macro avg	0.71	0.76	0.73	4705
weighted avg	0.87	0.85	0.86	4705

- Baseline model



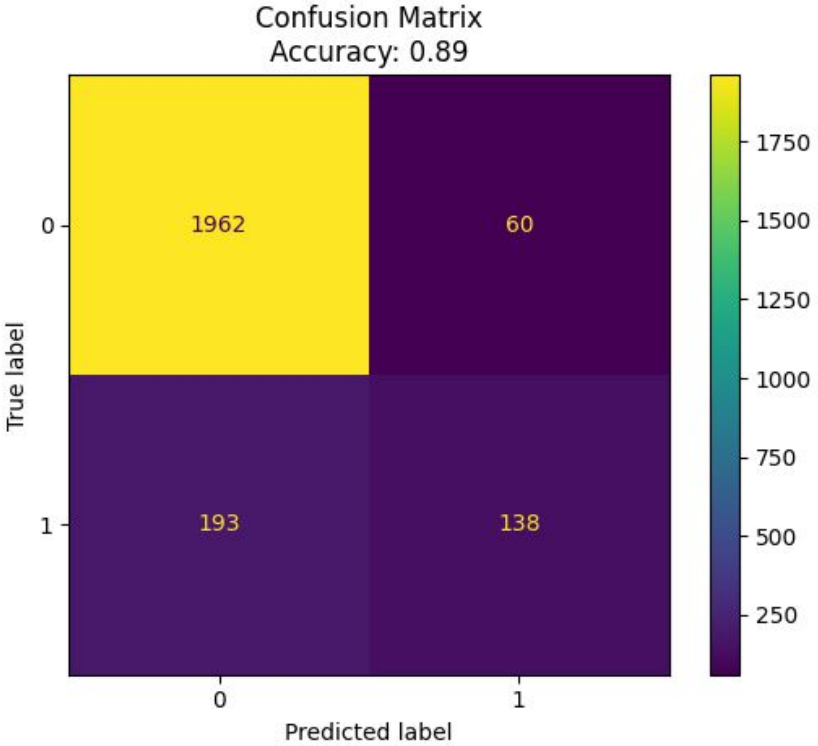
- Final model 1

Higher FP = False Positive could be double-checked



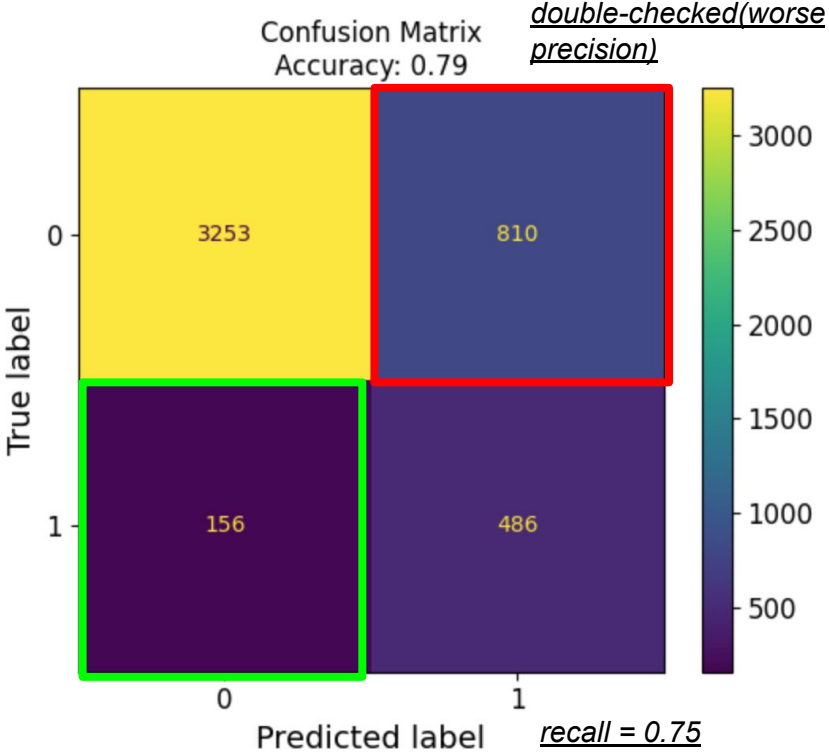
Observations scale between models is disproportioned due to oversampling technique applied to minority class in models

- Basis model



- Model 2

Higher FP = could be double-checked(worse precision)



recall = 0.75

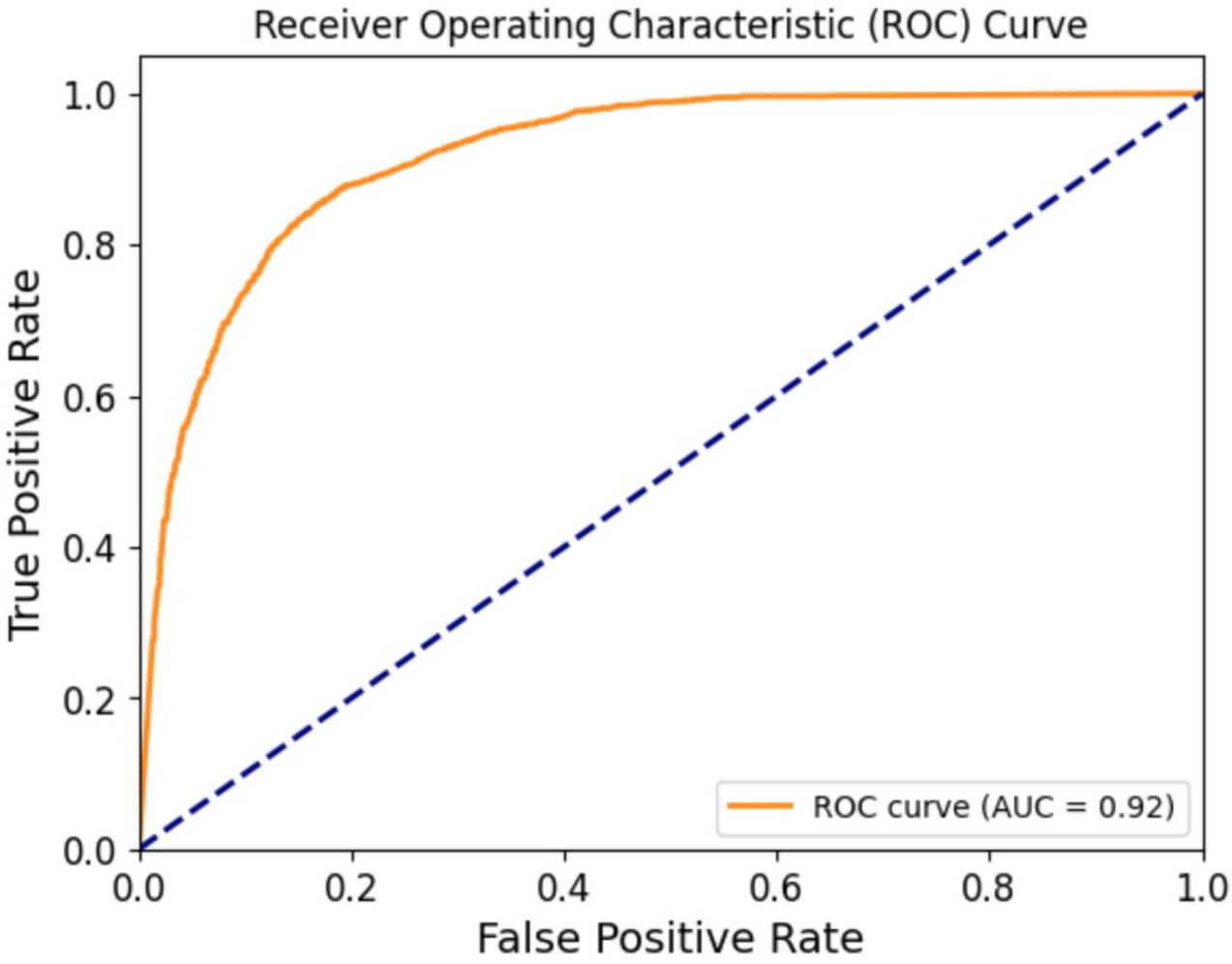
Observations scale between models is disproportioned due to oversampling technique applied to minority class in models

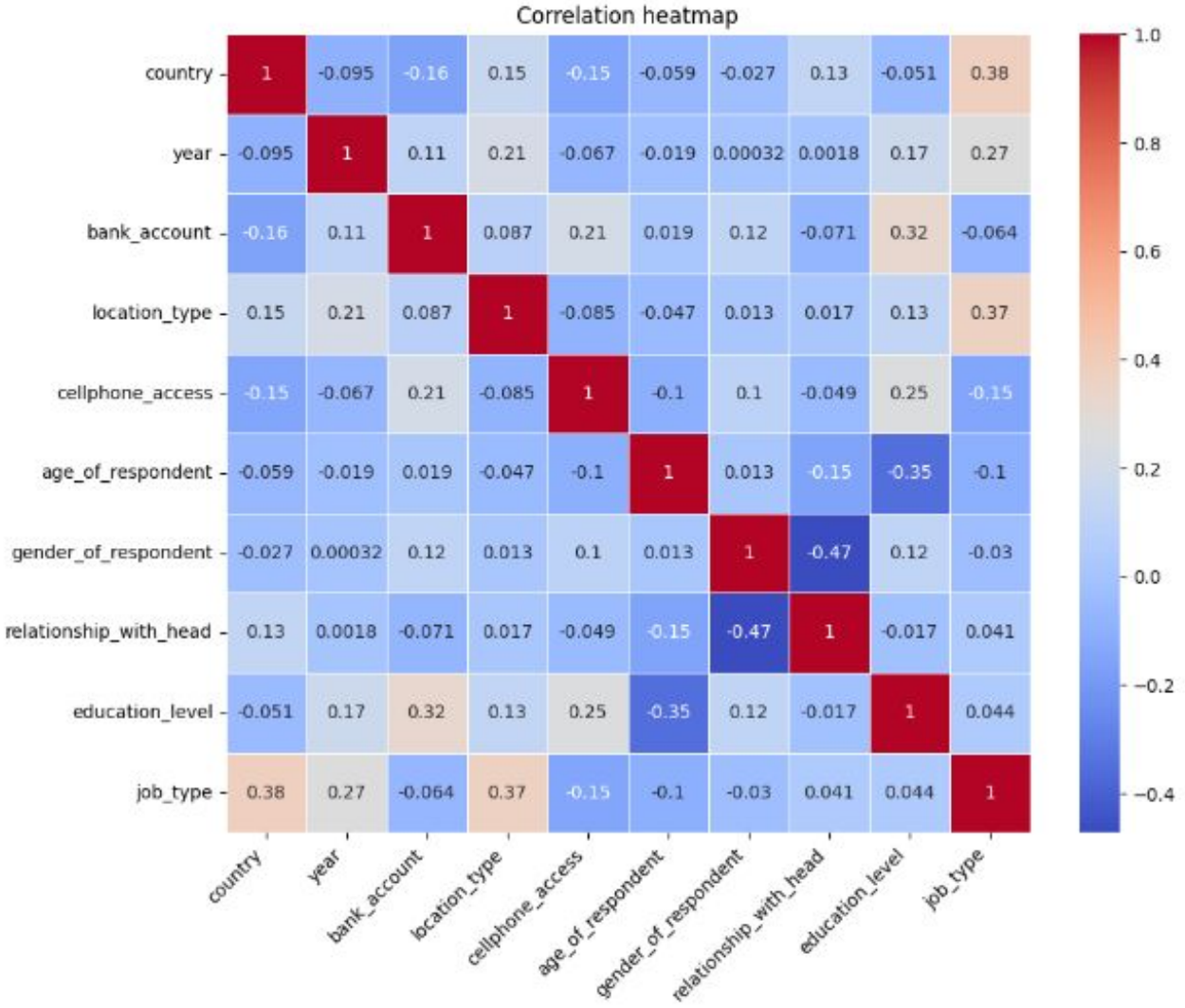
Next steps

We see, that we could capture more potential customers (true positives) at the expense of increased number of false positives as well.

If the cost of screening process is significantly lower than value from onboarded customer, we will tune the model to capture as many potential customers as possible.

Thanks!


Receiver Operating Characteristic (ROC) Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (AUC = 0.92)


Correlation heatmap
country 1 -0.095 -0.16 0.15 -0.15 -0.059 -0.027 0.13 -0.051 0.38
year -0.095 1 0.11 0.21 -0.067 -0.019 0.00032 0.0018 0.17 0.27
bank_account -0.16 0.11 1 0.087 0.21 0.019 0.12 -0.071 0.32 -0.064
location_type 0.15 0.21 0.087 1 -0.085 -0.047 0.013 0.017 0.13 0.37
cellphone_access -0.15 -0.067 0.21 -0.085 1 -0.1 0.1 -0.049 0.25 -0.15
age_of_respondent -0.059 -0.019 0.019 -0.047 -0.1 1 0.013 -0.15 -0.35 -0.1
gender_of_respondent -0.027 0.00032 0.12 0.013 0.1 0.013 1 -0.47 0.12 -0.03
relationship_with_head 0.13 0.0018 -0.071 0.017 -0.049 -0.15 -0.47 1 -0.017 0.041
education_level -0.051 0.17 0.32 0.13 0.25 -0.35 0.12 -0.017 1 0.044
job_type 0.38 0.27 -0.064 0.37 -0.15 -0.1 -0.03 0.041 0.044 1
country
year
bank_account
location_type
cellphone_access
age_of_respondent
gender_of_respondent
relationship_with_head
education_level
job_type
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4