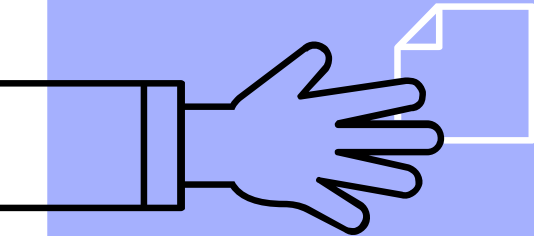
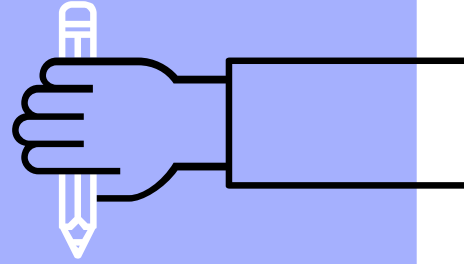


Will a person
have a bank
account or not ?

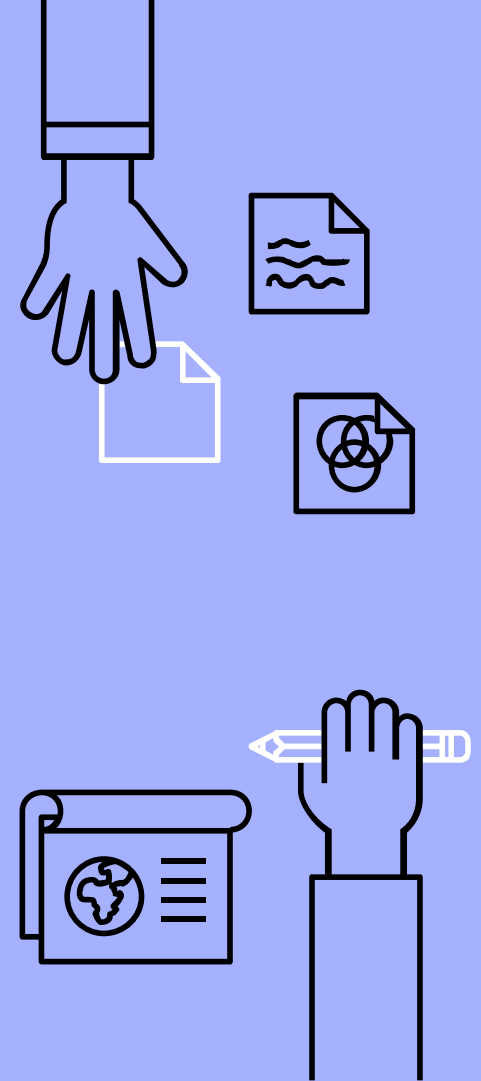


Aymen Soussi



Dataset Source :

- ▶ **Data Source:** Zindi challenge Financial Inclusion in Africa
- ▶ **Location:** <http://zindi.africa/competitions/financial-inclusion-in-africa/data>
- ▶ **Filename:** Train_v2.csv



Data set

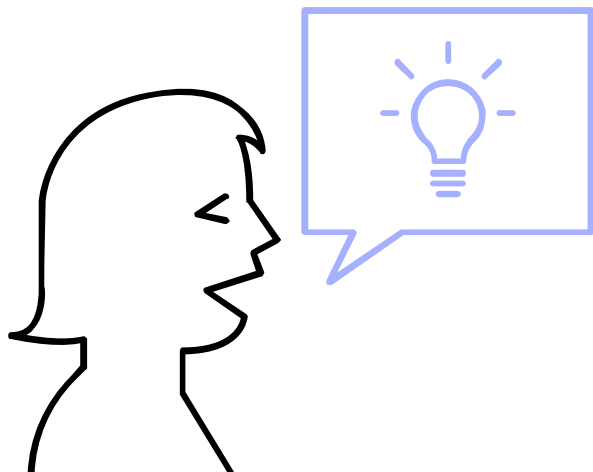
23524 rows and 13 features:

9 categorical:

Gender, Education_level and
Bank_account (target)...

3 numerical :

Year , Age...



Data Quality Assessment

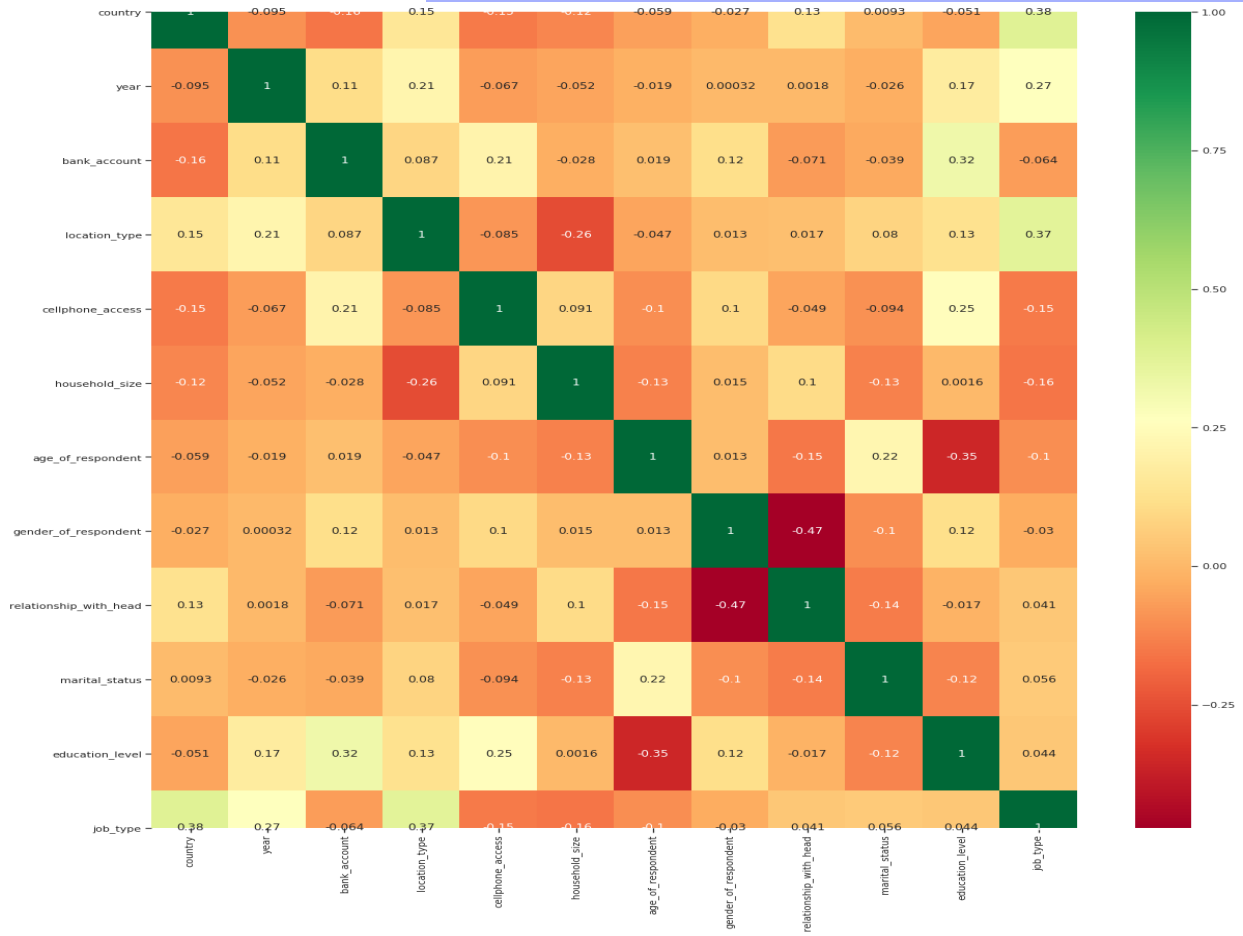
```
[ ] data.isnull().sum()
```

```
country          0
year             0
uniqueid         0
bank_account     0
location_type    0
cellphone_access 0
household_size   0
age_of_respondent 0
gender_of_respondent 0
relationship_with_head 0
marital_status   0
education_level   0
job_type         0
dtype: int64
```

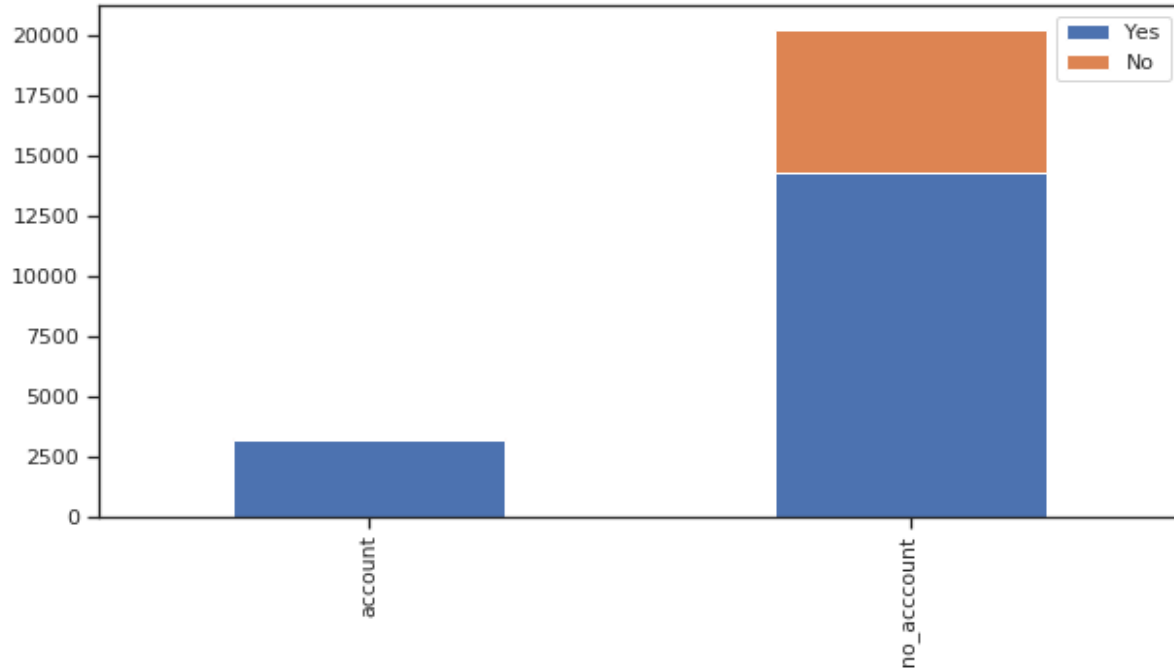
This show that the data is already cleaned

Data Exploration

Correlation Matrix



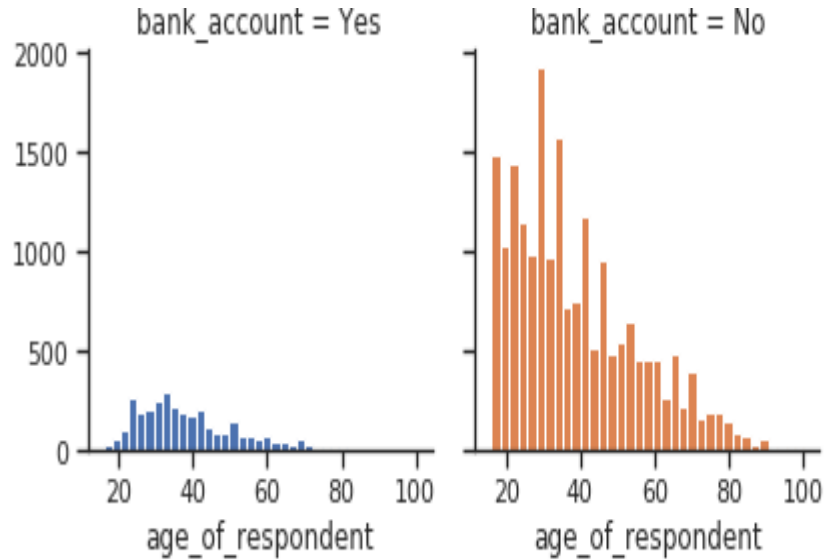
Data Exploration



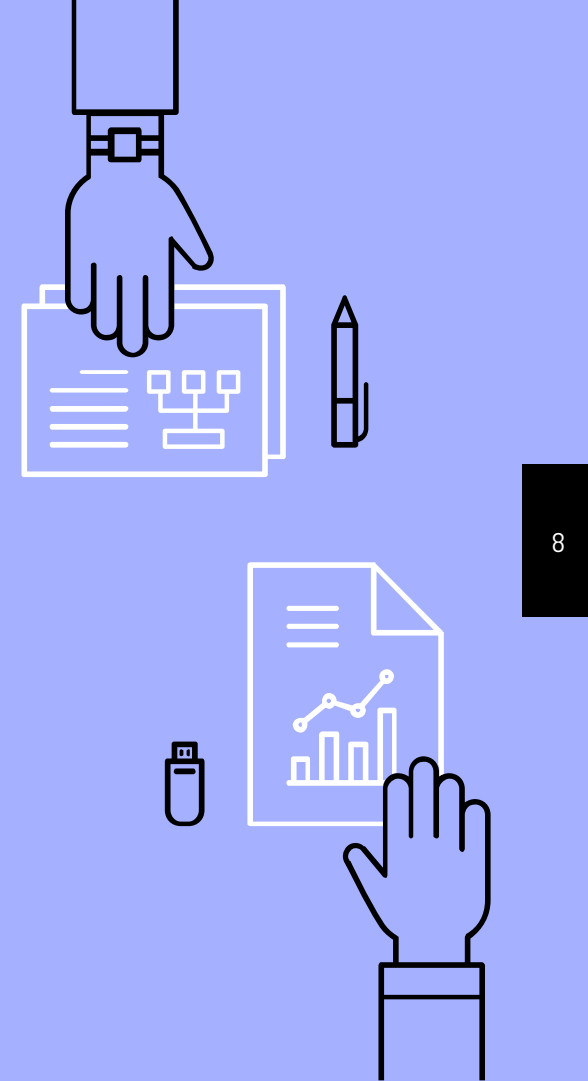
Cellphone access per bank account



Data Exploration



Age Distribution



Features engineering

Label encoded categorical features .

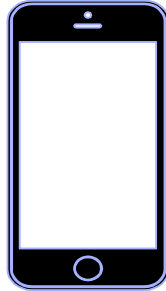
```
le = LabelEncoder()
data['gender_of_respondent'] = le.fit_transform(data['gender_of_respondent'])
data['bank_account'] = le.fit_transform(data['bank_account'])
data['cellphone_access'] = le.fit_transform(data['cellphone_access'])
data['location_type'] = le.fit_transform(data['location_type'])
data['relationship_with_head'] = le.fit_transform(data['relationship_with_head'])
data['gender_of_respondent'] = le.fit_transform(data['gender_of_respondent'])
data['marital_status'] = le.fit_transform(data['marital_status'])
data['education_level'] = le.fit_transform(data['education_level'])
data['job_type'] = le.fit_transform(data['job_type'])
data['country'] = le.fit_transform(data['country'])
```

Model Selection

```
↳ The accuracy of random forest was : 0.8625672995182772  
The accuracy of xgboost was : 0.880136015868518  
The accuracy of catboost was : 0.8821195806177388  
The accuracy of Light gbm was : 0.8841031453669594  
The accuracy of stack model was : 0.8855199773306885  
The accuracy of the deep learning model was : 0.8659676962312269
```

Conclusion

- ▶ The best accuracy was 88.55% (stacking)
- ▶ Deploy the model to build a web app and a mobile app



Thank you very much

