

Female-headed households in South Africa



Alexandra, Christine, Jana

Keyindicator for aid distribution

Stakeholder: **NGO**, has limited amount of money for aid

Goal: identify wards, where aid is needed most

How: **KEYINDICATOR**

Prediction of % of female-headed households per ward

with income < 19.600 R per year



<https://thecircle.ngo/tag/south-africa/>



Income South Africa:

poverty line: 14.196 R per year

average income: 20.860 R per month

The Data



Womxn in Big Data South Africa: Female-Headed Households in South Africa

source <https://zindi.africa/competitions/womxn-in-big-data-south-africa-female-headed-households-in-south-africa/data>

- data aggregated from census 2011
- 2822 observations, 64 variables
- variables: housing forms, school attendance, access to water etc
- no missing values
- data as percentage per ward

target: female headed households with annual income < 19.600 R per year

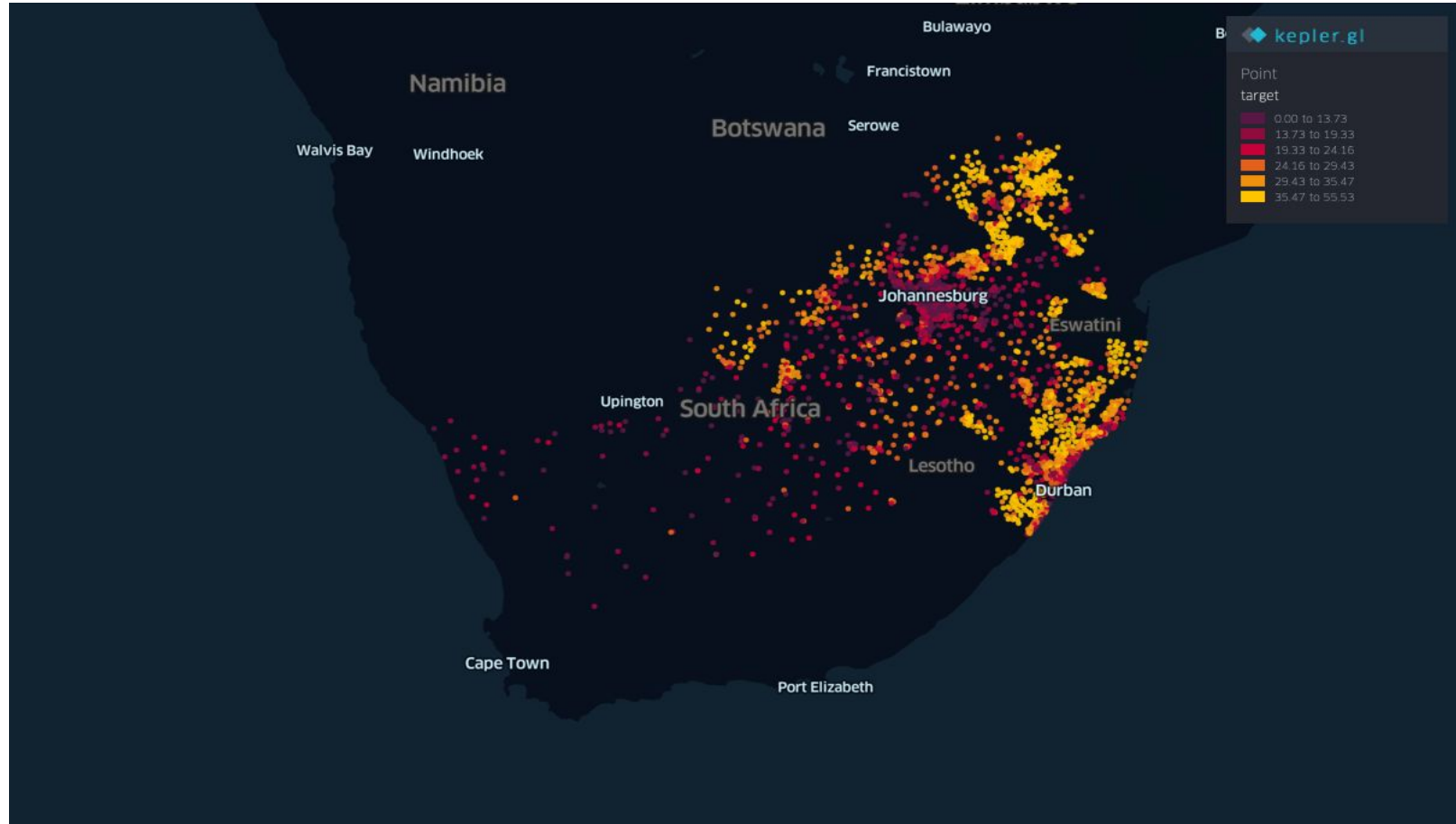
Wards in South Africa

source:
[https://en.wikipedia.org/wiki/Ward_\(South_Africa\)#/media/File:South_Africa_electoral_wards_2016_blank.svg](https://en.wikipedia.org/wiki/Ward_(South_Africa)#/media/File:South_Africa_electoral_wards_2016_blank.svg), created by Htonl

4,392 wards in South Africa

ward: district of a city or town

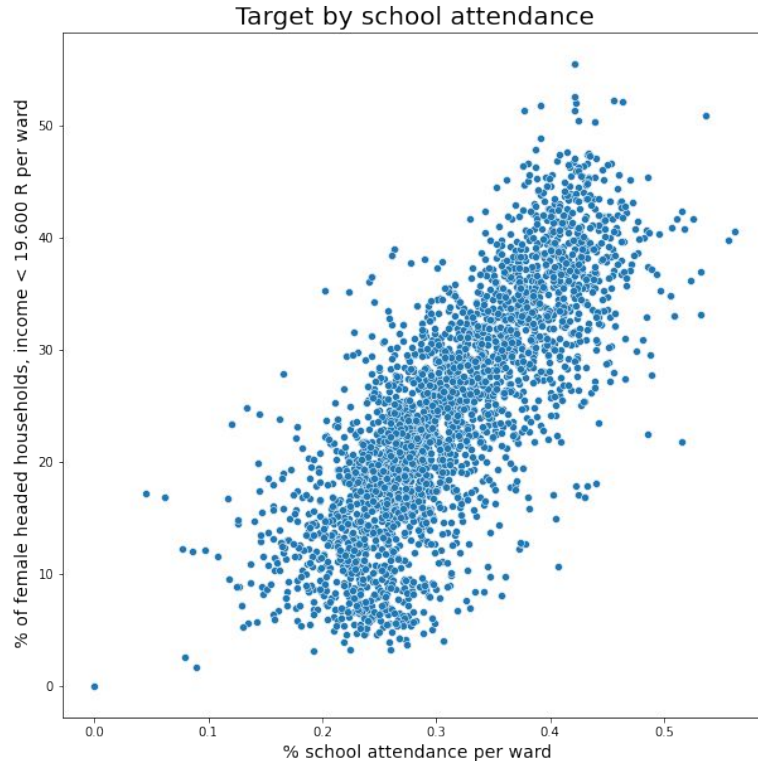
Distribution of target in South Africa



Hypotheses

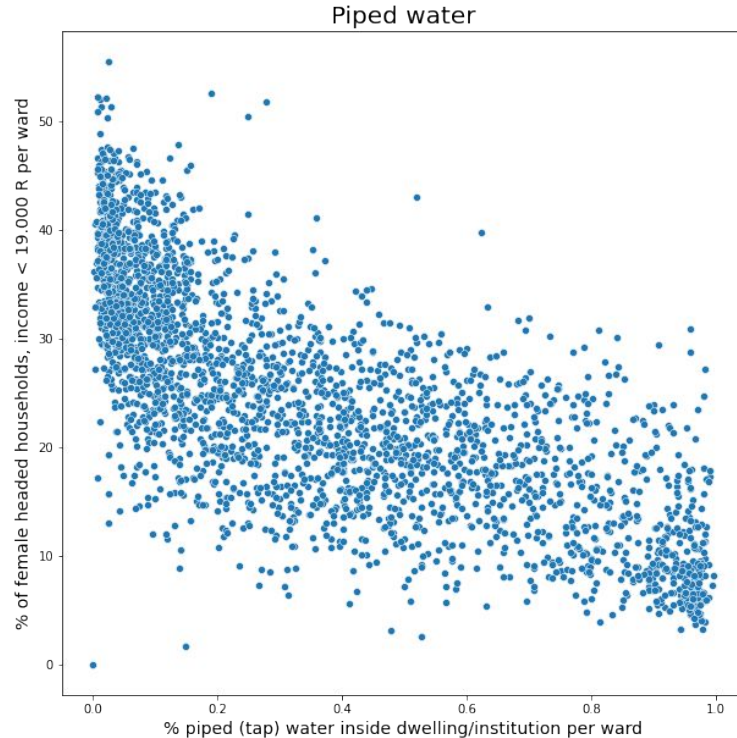
1. target is high in wards with low school attendance
2. target is high in regions with little water access
3. target higher in rural regions

1. School attendance



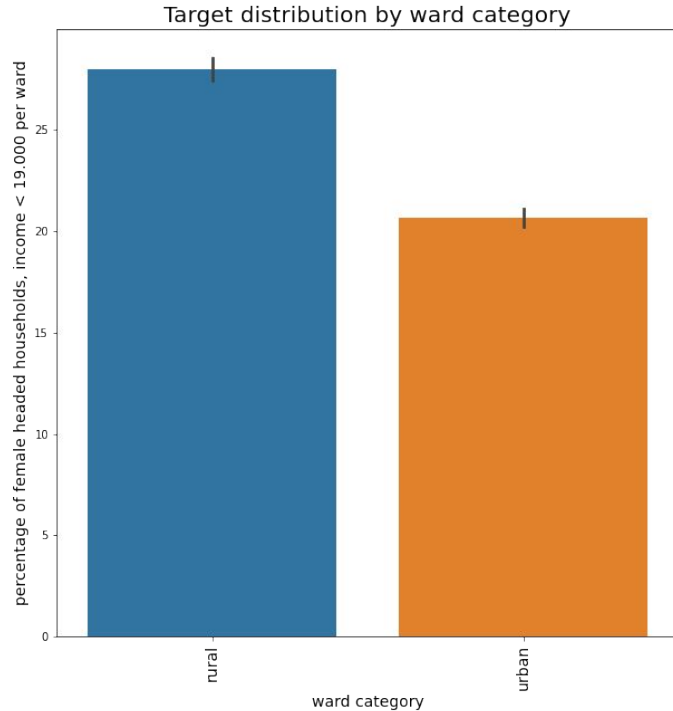
- even wards with highest school attendance have a maximum of ~ 50%
→ school attendance overall low
- target positively correlated with % of school attendance
→ unexpected

2. Access to water



- target high in regions with low % of piped water in dwelling

3. Ward category (rural vs. urban)



- Target higher in rural wards

Baseline Model

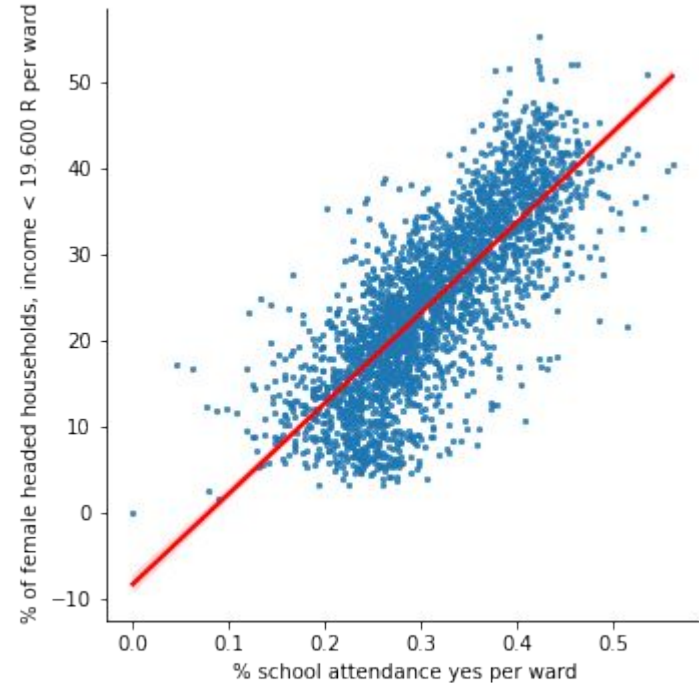
Hypothesis:

Target is high in wards with low school attendance.

model: target ~ school attendance

Evaluation of the model:

- Score(RMSE): 6.90
- Errors too high



Final Model

XGBoost:

selected features:

- school attendance
- car ownership
- access to water
- land ownership
- population group (black african)
- housing type (tent, caravan)

Evaluation of the models by RSME:

Baseline: 6.90

XGBoost: 3.92

Recommendations for the stakeholder

- collect data from important features only
- invest in water access for wards
- target group is present in rural and urban wards → not focus only on rural

Limitations

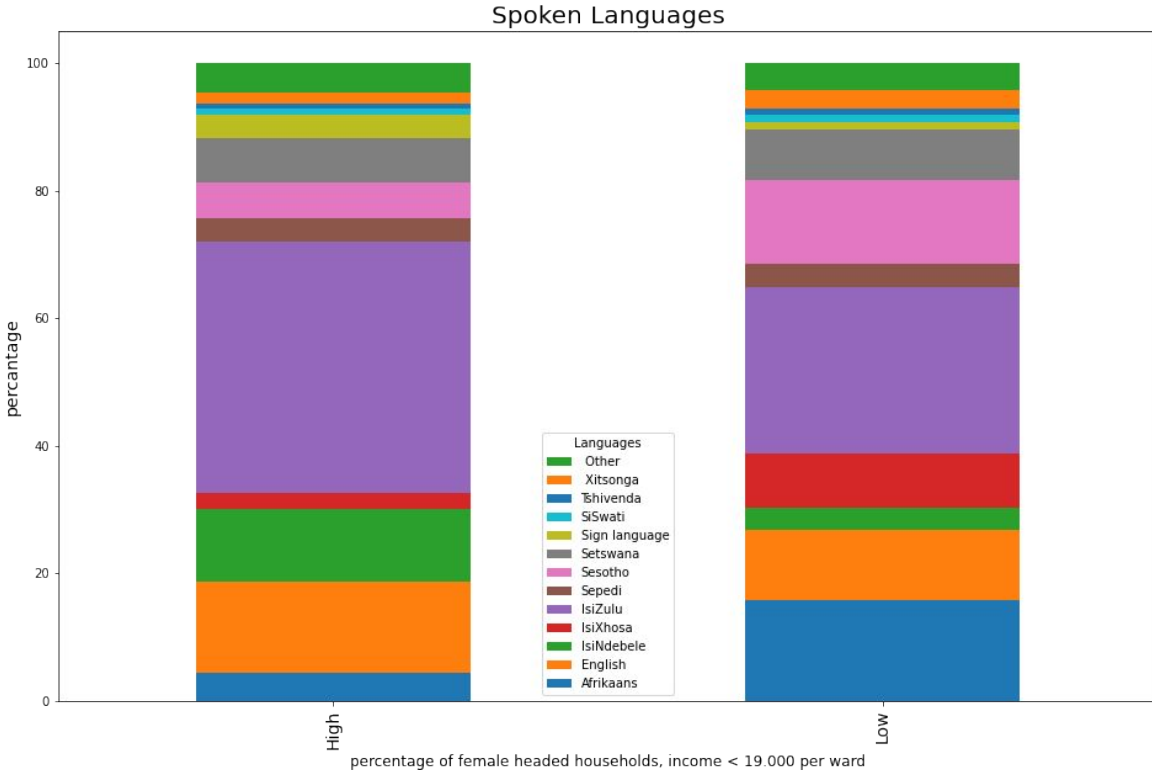
- no insights into needs of the target group

Future work

- look for trends in subpopulations

Thank you for your attention!

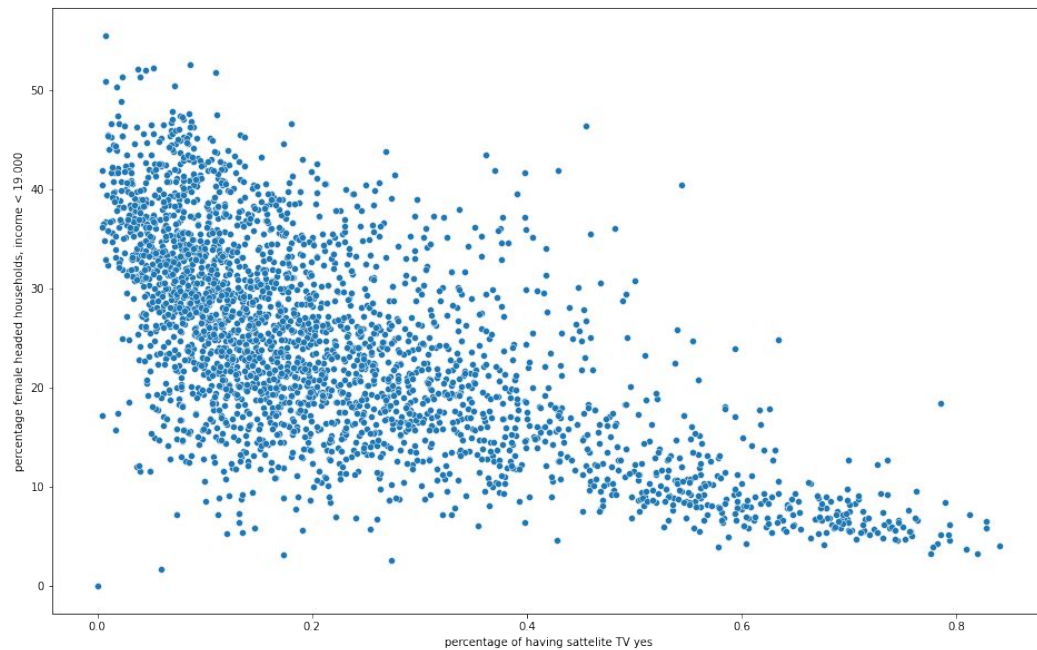
Spoken languages associated with target



highest difference:

- Afrikaans
- IsiZulu
- English

Satellite TV



Milestone 1

Value of product: Monitor the percentage of female headed households between census years. Since female-headed households are prone to poverty, the stakeholder (NGO) uses it as a key indicator to estimate the amount of relief supplies in certain regions of South Africa.

Prediction: Percentage of female headed households with income under 19,600 R per year per ward (in South Africa)

Evaluation Metrics: R^2 and RMSE

Baseline Model: Linear Regression,

Hypothesis: “wards with low school attendance have a high percentage of female-headed households with income under 19,600 R per year”

model: target ~ school attendance

Score: R^2 : 0.533, RMSE: 6.90