PREDICTING COSTA RICAN HOUSEHOLD POVERTY LEVEL

Belén Sánchez

October 2018

Avancemos, 2015

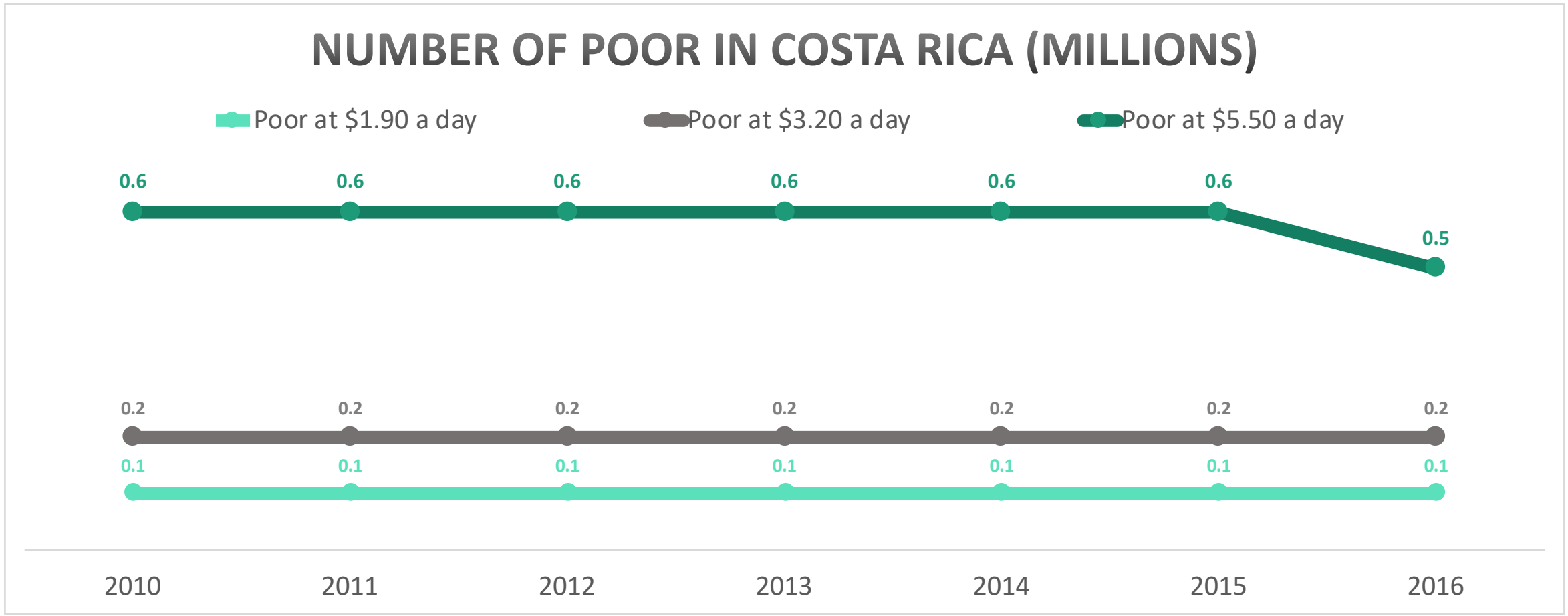
La Nación, 2016

Costa Rica

Capital city:	San José
Area:	51,100 km^2 (19.700 sq mi)
Population:	4,857,274
Density:	220/sq mi
Income level:	Upper middle Income
GDP per capita:	$12,144

Source: Wikipedia, 2018

Poverty in Costa Rica



Source: The World Bank, World Development Indicators, 2018

2016 – Poverty reduction due transfers

POVERTY

10,400

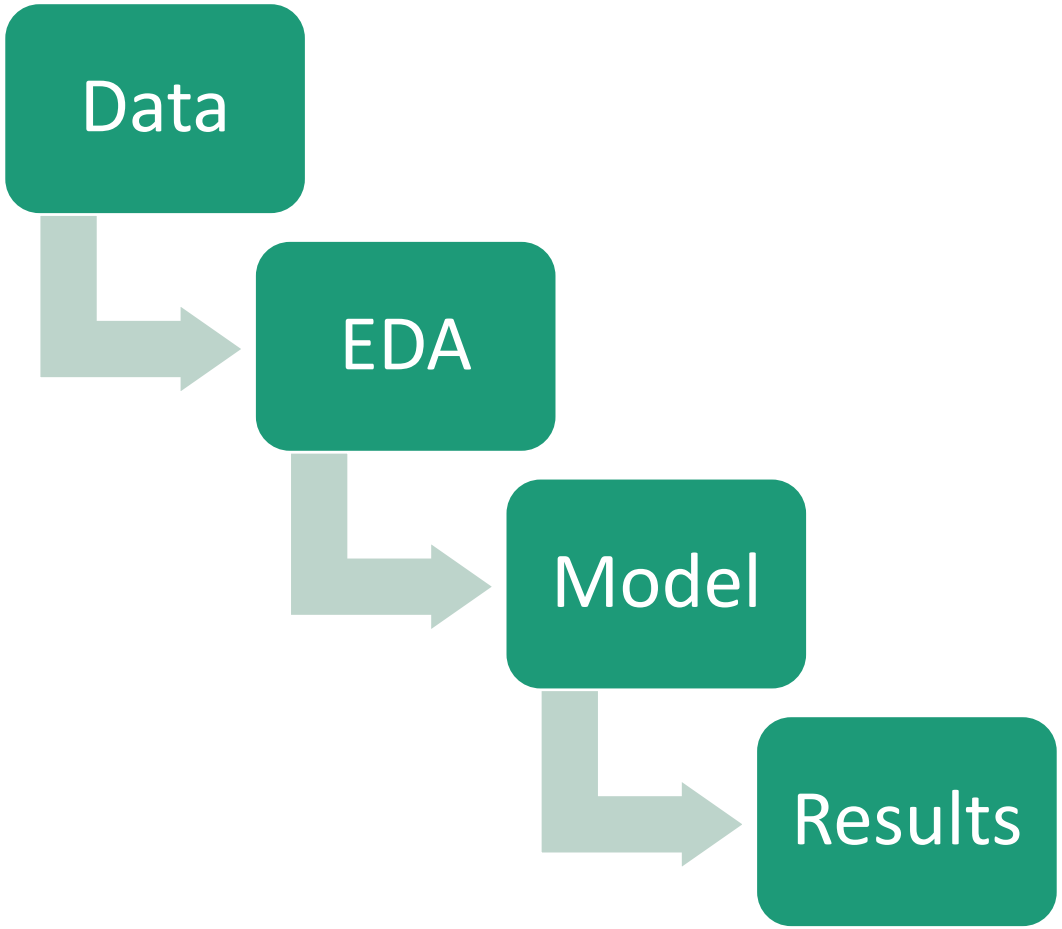
22,620

Source: Nacion, 2016 , Enaho, 2016

The project

Goal

To identify which households have the highest need for social welfare assistance in Costa Rica



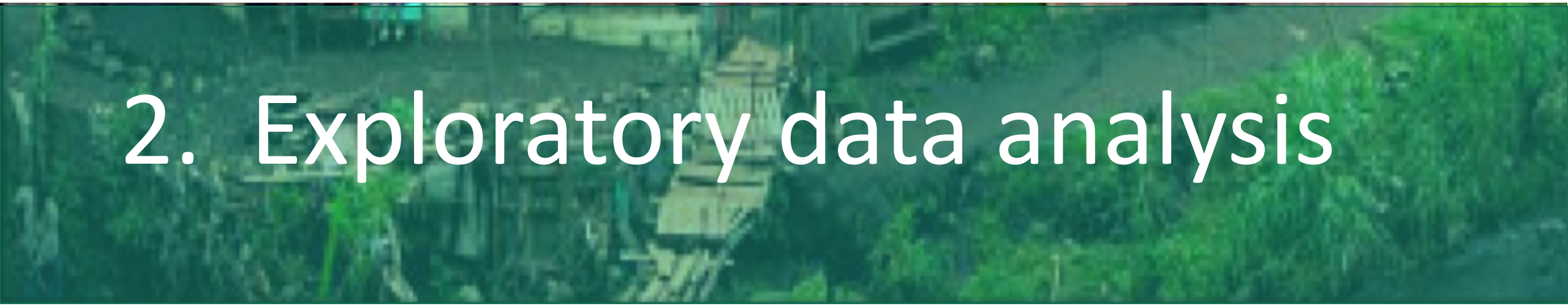
IDB Inter-American Development Bank

kaggle

1. Data

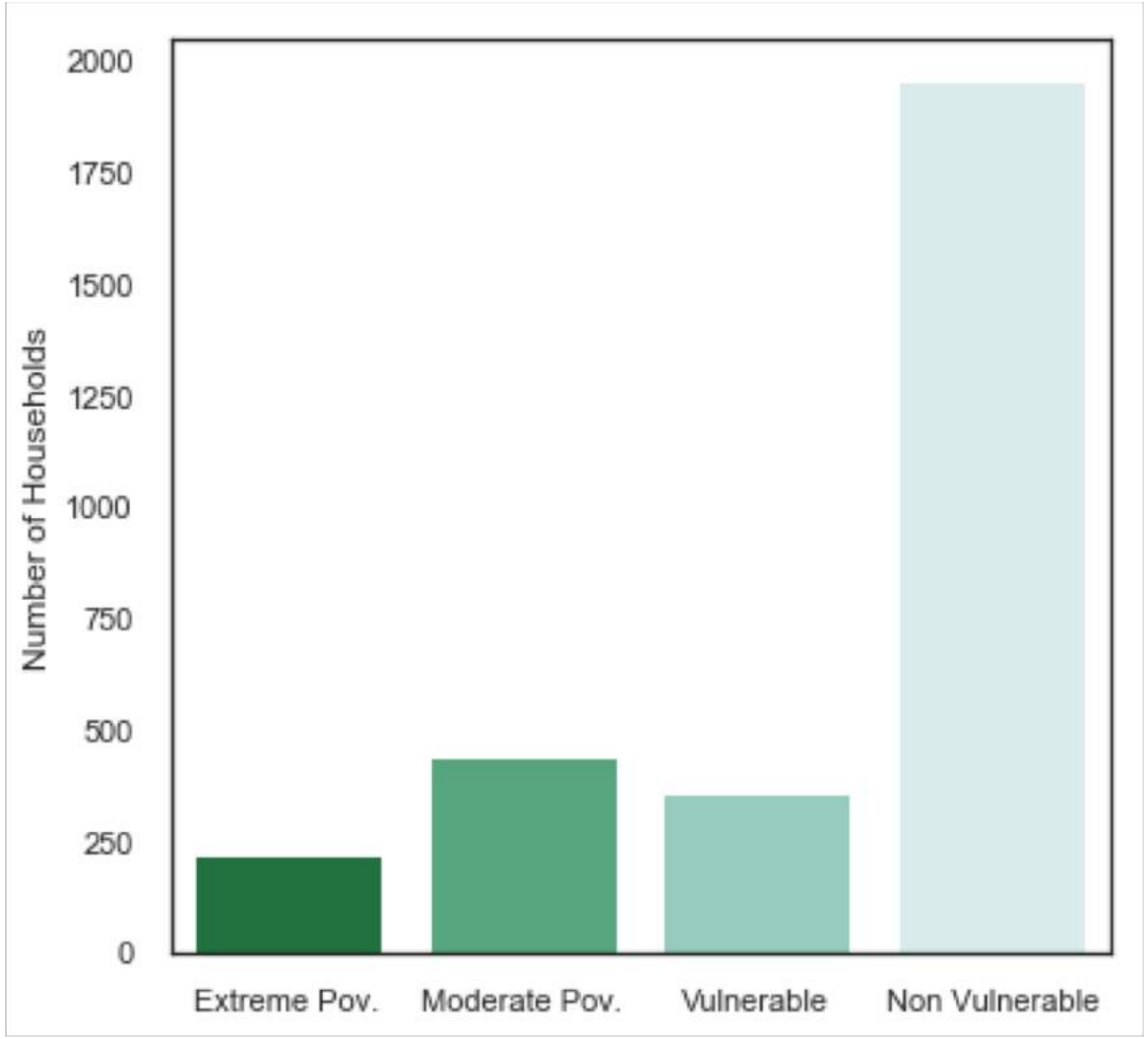
Data sets

TRAIN	TEST
9,557 individual observations	23,856 individual observations
2,973 households	7,352 households
143 features	142 features



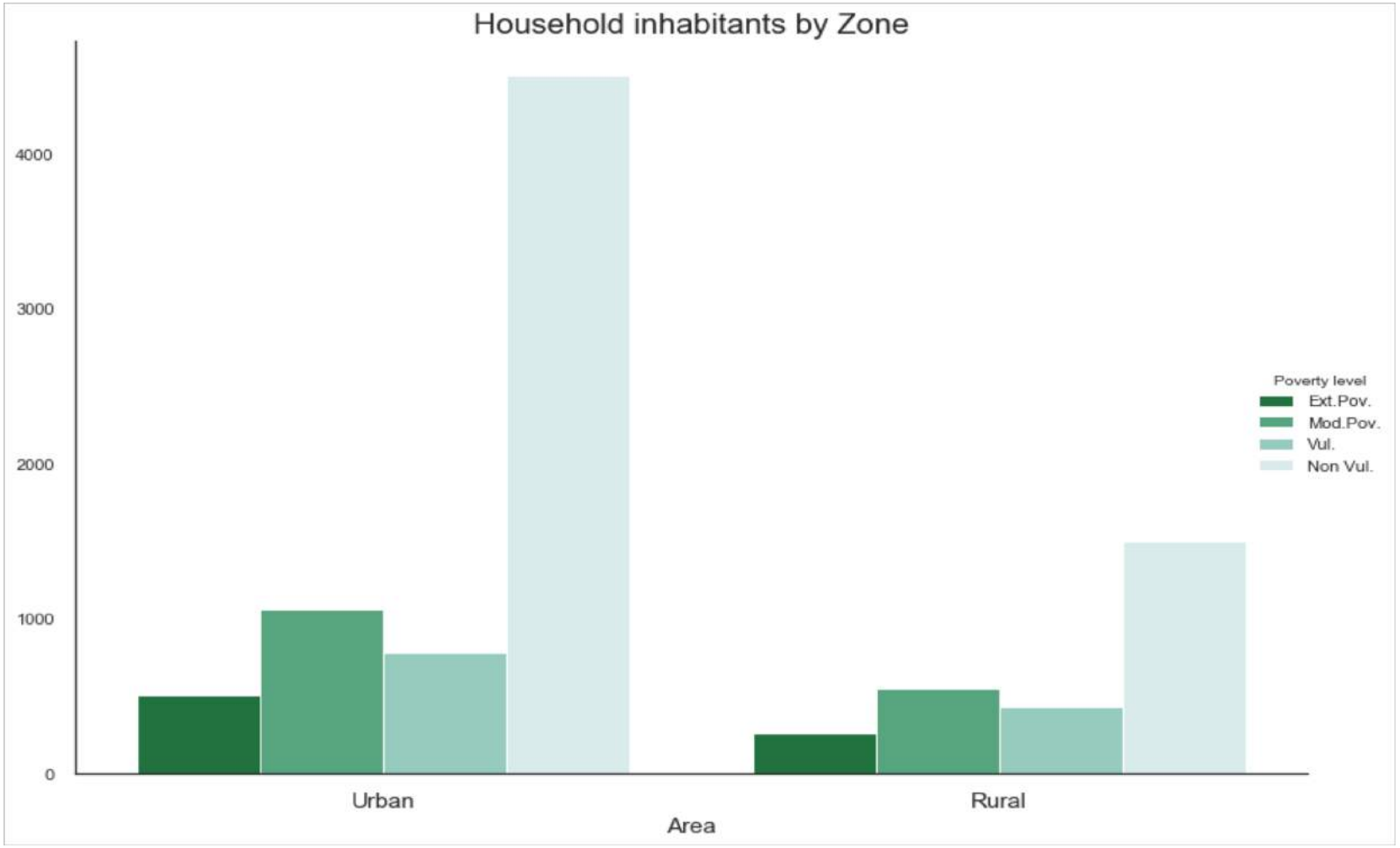
2. Exploratory data analysis

Poverty level of households

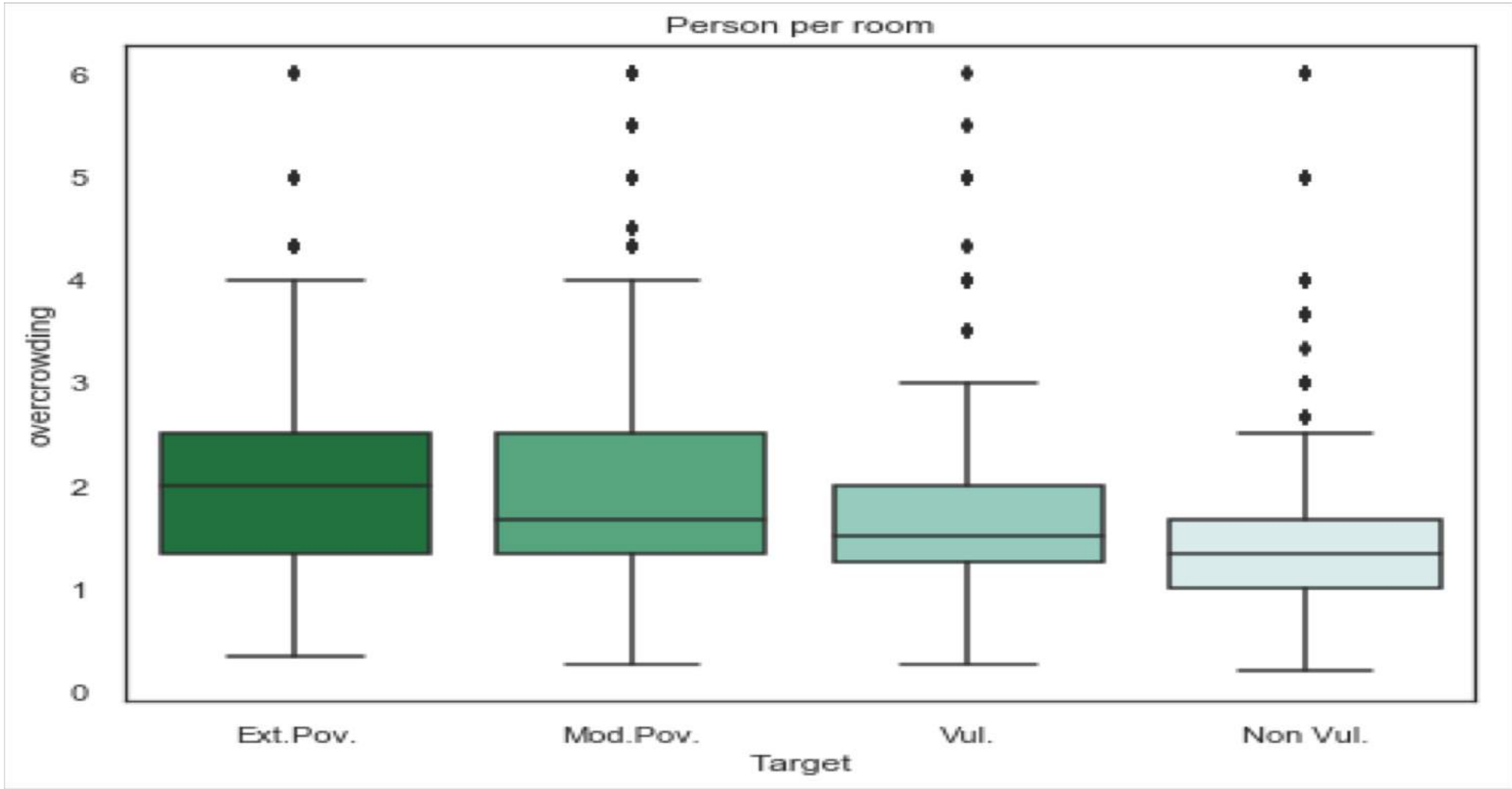


34% of households between extreme poverty and vulnerable.

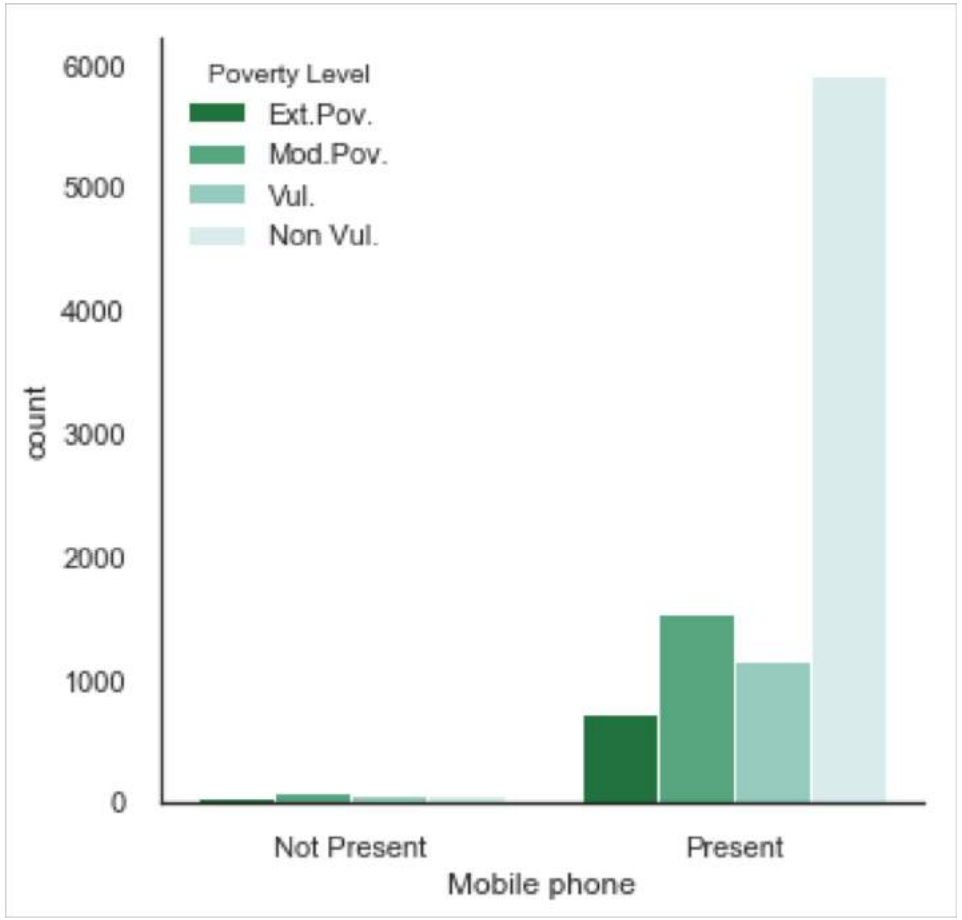
71% of households in urban areas

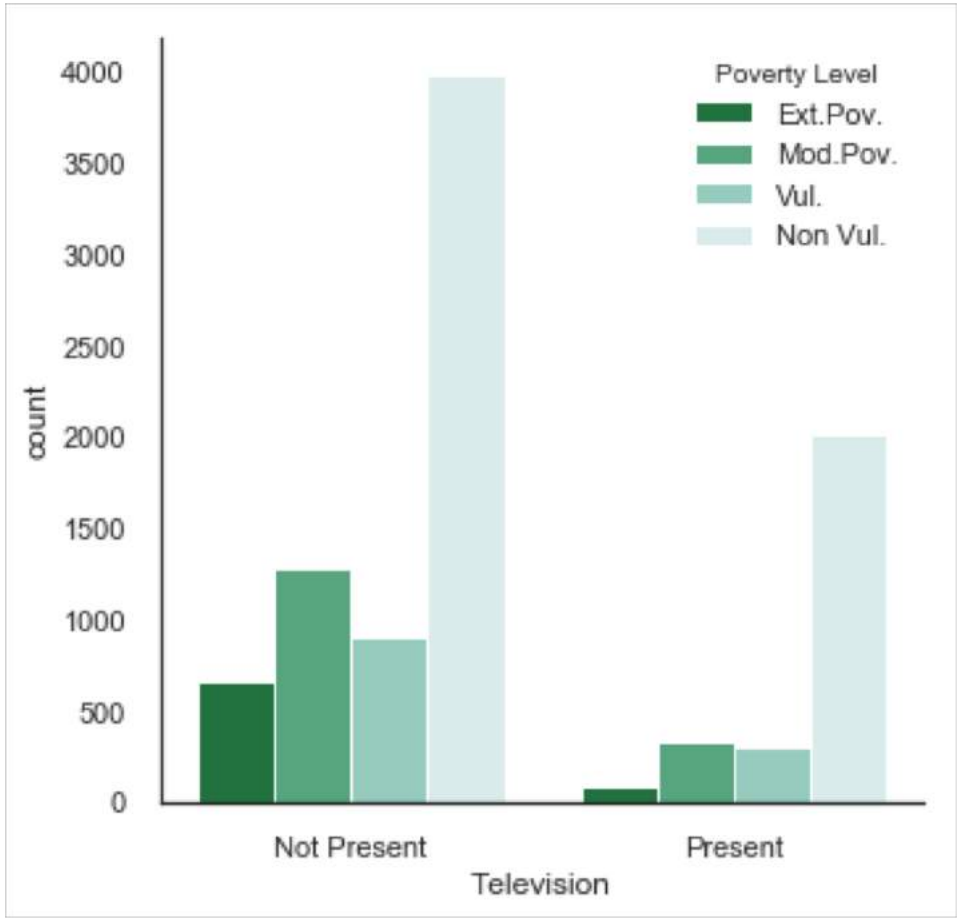


Overcrowding is higher in more vulnerable households.



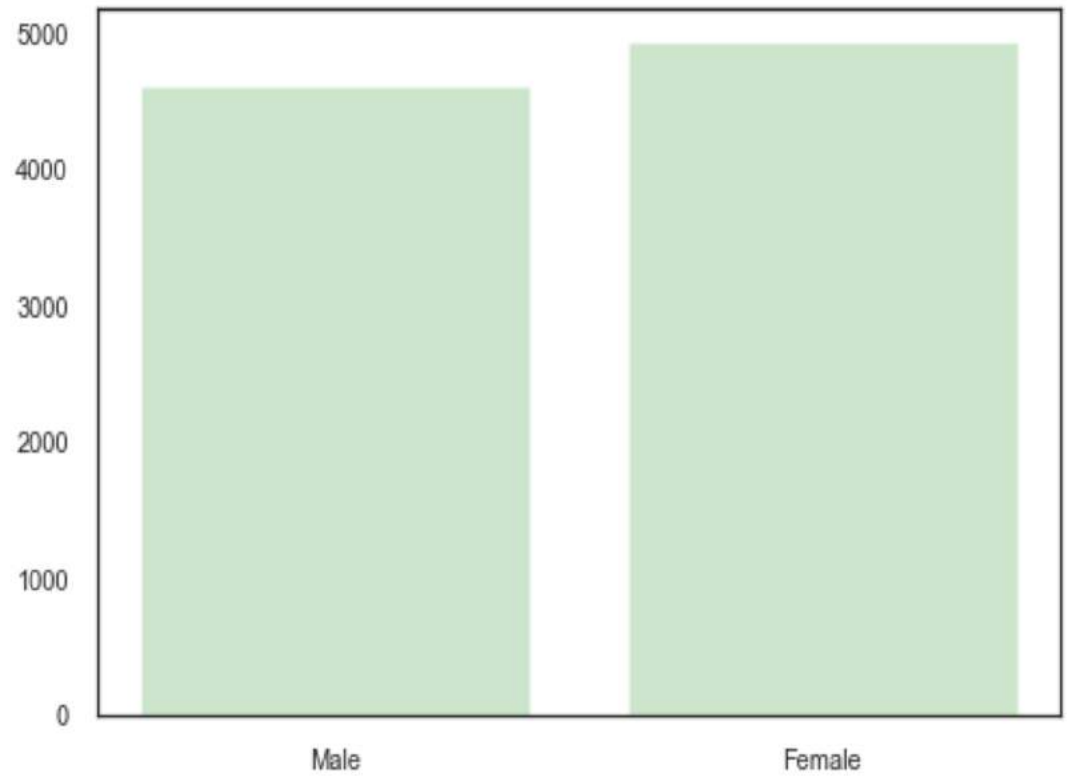
Assets


Poverty Level
Ext.Pov.
Mod.Pov.
Vul.
Non Vul.
count
6000
5000
4000
3000
2000
1000
0
Not Present
Present
Mobile phone

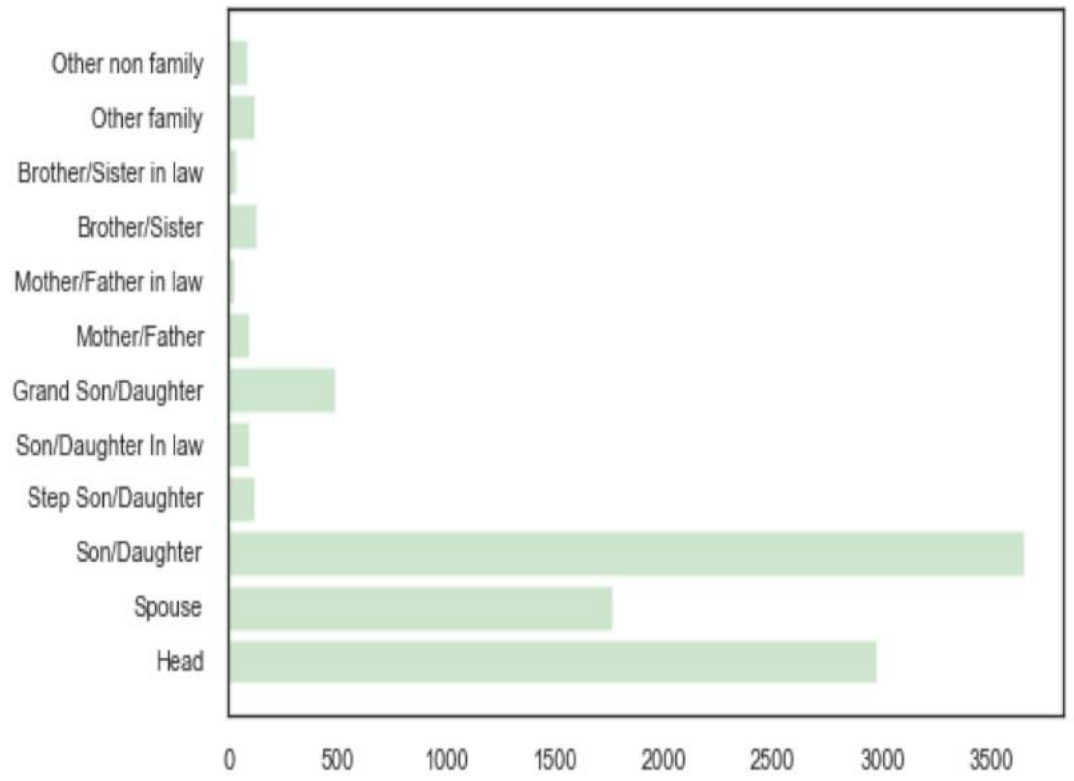

Poverty Level
Ext.Pov.
Mod.Pov.
Vul.
Non Vul.
count
4000
3500
3000
2500
2000
1500
1000
500
0
Not Present
Present
Television

Individual characteristics

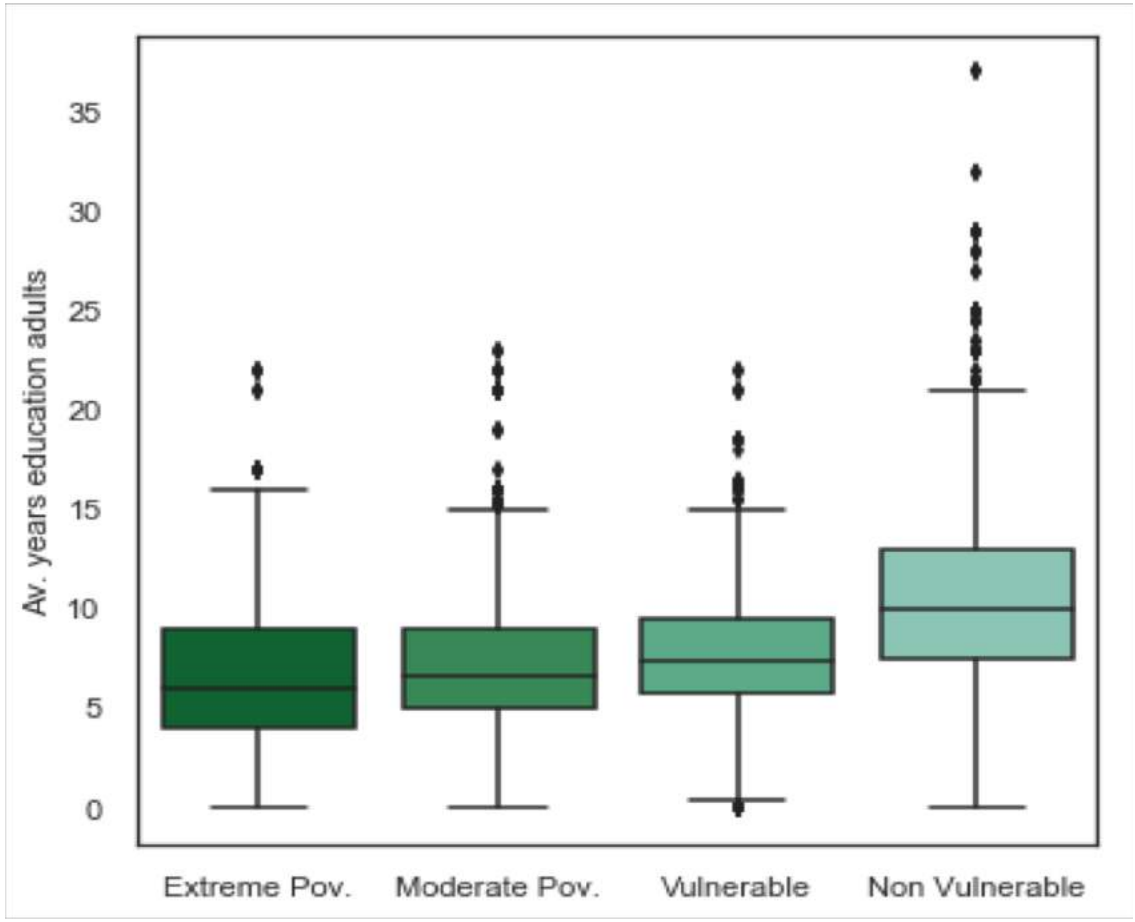
Gender distribution



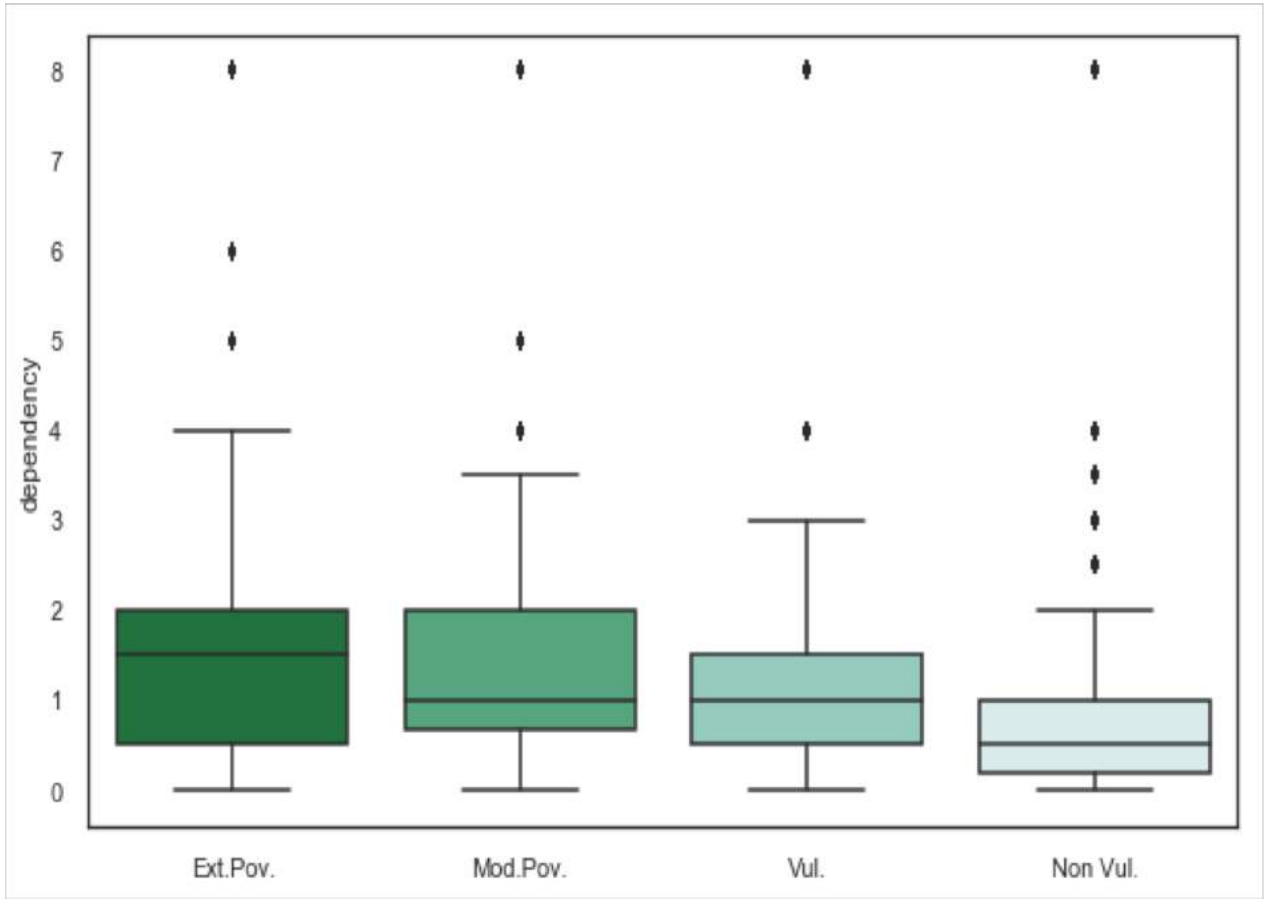
Family role



Education



Dependency



Feature engineering

- ✓Rent per capita
- ✓Male / female per household
- ✓ Children per household
- ✓Youth per household
- ✓Elderly per household
- ✓Ratio children per adult

3. Modeling

Model Comparison

MODEL	ACCURACY SCORE TRAIN	ACCURACY SCORE TEST
Random forest	1.0	0.93682
Extra tree classifier	1.0	0.85062
Adaboost	0.70294	0.68786
Gradient booster	0.81247	0.78410
Random forest with PCA	0.99413	0.75146

Best Model

Features:

New features

Dropped features with correlation > 0.90

Model:

Random Forest

Grid search:

- Max. depth: 50
- Min. samples split: 2
- N estimators: 60

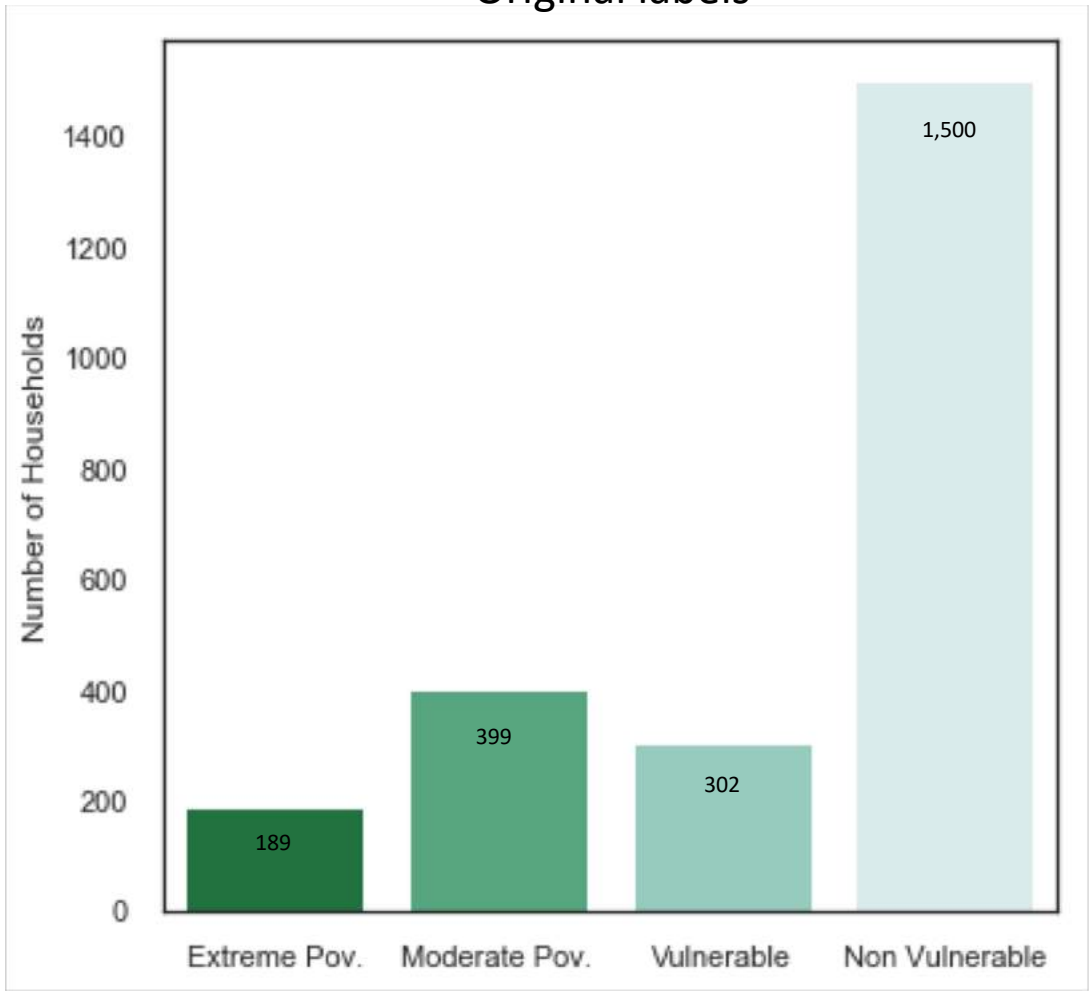
4. Results

Model scores

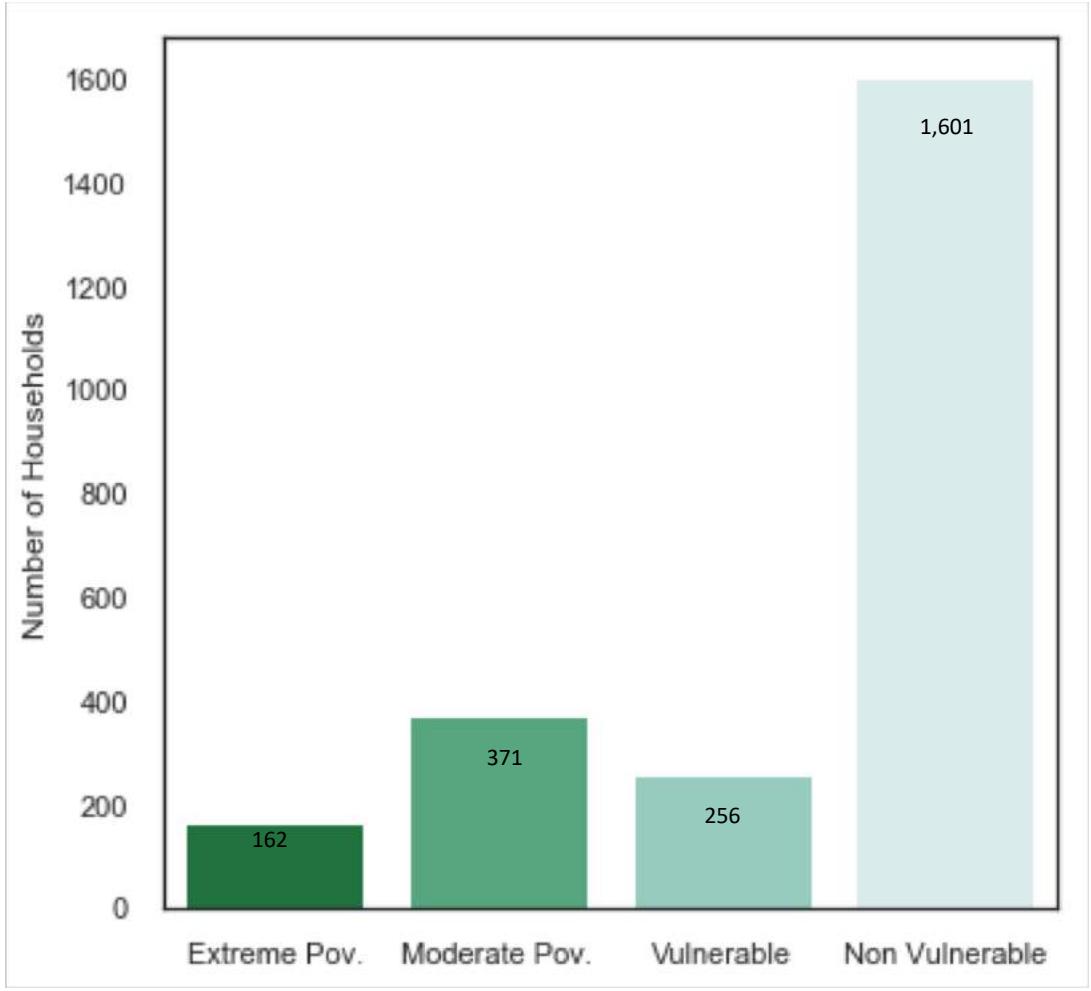
	PRECISION	RECALL	F1
Extreme poverty	0.96	0.82	0.88
Moderate poverty	0.93	0.86	0.89
Vulnerable	0.96	0.81	0.88
Non - vulnerable	0.93	1.00	0.96
Average	0.94	0.94	0.94

8% of values wrongly predicted

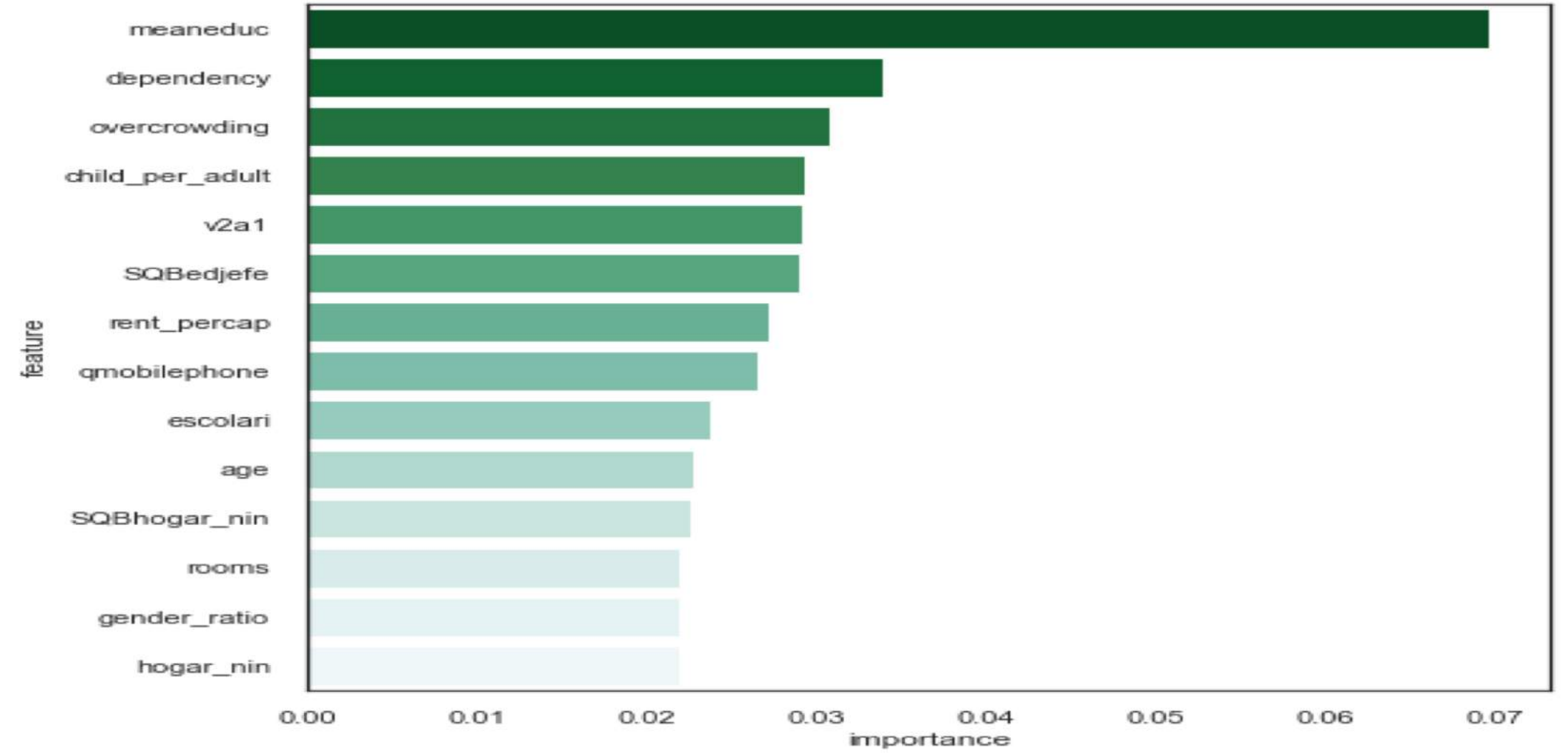
Original labels



Predicted labels



Feature Importance



5. Recommendations / Next Steps

Policy Recommendations

Focus support of monetary and non monetary welfare transfers with educational purposes.

Increase support to families with higher dependency levels.

Review housing policies and how are they addressing overcrowding problems.

Next steps

Train the model with more observations.

Engineer new features.