



International Women's Day Challenge

Helping the World

3 000 Points

19 days left

Skills you will learn

Prediction

108 joined28 active



About

South Africa is divided into over 4,000 wards. We have aggregated the target indicator and the other predictive variables from the 2011 census across all the households within each ward to create an aggregated value of each indicator per ward.

The target variable of interest is the percentage of households per ward that are both female-headed and earn an annual income that is below R19,600 (approximately \$2,300 USD in 2011). For context, the poverty line in South Africa is defined as R1,183 per month per person and the average individual salary in South Africa is R20,860 per month.

The objective of this challenge is to accurately model the target indicator using the predictive variables provided in the datasets. For the purposes of this competition, we have split the wards into a train and test set. You will train your model on the 2,822 wards in the train set and apply your model to the 1,013 wards in the test set.

Take note that the variables:

- "lat" and "lon" are the locations of the CENTER POINTS of the wards.
- "ADM4_PCODE" is the ward code used to find the ward polygon in the shapefile available [here](#).
- "NL" is the nightlights value for the center point and surrounding area. The values come from the Global Radiance-Calibrated Nighttime Lights database [here](#).

Files

DESCRIPTION	FILES
Train contains the target. This is the dataset that you will use to train your model.	<div>Train.csv</div> <div>2 MB</div> <div>↓</div>
Full list of variables and their explanations.	<div>variable_descriptions.csv</div> <div>4.3 KB</div> <div>↓</div>
This notebook will help you make your first submission to the leaderboard.	<div>Starter_Notebook.ipynb</div> <div>892.3 KB</div> <div>↓</div>
Test resembles Train.csv but without the target-related columns. This is the dataset on which you will apply your model to.	<div>Test.csv</div> <div>1.1 MB</div> <div>↓</div>
This shows the submission format for this competition, with the ‘ID’ column mirroring that of Test.csv and the ‘target’ column containing your predictions. The order of the rows does not matter, but the names of the ID must be correct.	<div>SampleSubmission.csv</div> <div>20.4 KB</div> <div>↓</div>



Join the largest network for
data scientists and AI builders

- About

Privacy Policy
- FAQs

Terms of Use
- Rules