File descriptions

* {train|test}.csv - the training set
  + This is the main table, broken into two files for Train (with a Target column) and Test (without the Target column).
  + One row represents one person in our data sample.
  + Multiple people can be part of a single household. Only predictions for heads of household are scored.
* sample\_submission.csv - a sample submission file in the correct format
  + This file contains all test IDs and a default value.
  + Note that ONLY the heads of household are used in scoring. All household members are included in test + the sample submission, but only heads of households are scored.

Core Data fields

* Id - a unique identifier for each row.
* Target - the target is an ordinal variable indicating groups of income levels.   
  1 = extreme poverty   
  2 = moderate poverty   
  3 = vulnerable households   
  4 = non vulnerable households
* idhogar - this is a unique identifier for each household. This can be used to create household-wide features, etc. All rows in a given household will have a matching value for this identifier.
* parentesco1 - indicates if this person is the head of the household.
* This data contains 142 total columns.

All Data fields

Variable name, Variable description

v2a1, Monthly rent payment

hacdor, =1 Overcrowding by bedrooms

rooms, number of all rooms in the house

hacapo, =1 Overcrowding by rooms

v14a, =1 has toilet in the household

refrig, =1 if the household has refrigerator

v18q, owns a tablet

v18q1, number of tablets household owns

r4h1, Males younger than 12 years of age

r4h2, Males 12 years of age and older

r4h3, Total males in the household

r4m1, Females younger than 12 years of age

r4m2, Females 12 years of age and older

r4m3, Total females in the household

r4t1, persons younger than 12 years of age

r4t2, persons 12 years of age and older

r4t3, Total persons in the household

tamhog, size of the household

tamviv, number of persons living in the household

escolari, years of schooling

rez\_esc, Years behind in school

hhsize, household size

paredblolad, =1 if predominant material on the outside wall is block or brick

paredzocalo, "=1 if predominant material on the outside wall is socket (wood, zinc or absbesto"

paredpreb, =1 if predominant material on the outside wall is prefabricated or cement

pareddes, =1 if predominant material on the outside wall is waste material

paredmad, =1 if predominant material on the outside wall is wood

paredzinc, =1 if predominant material on the outside wall is zink

paredfibras, =1 if predominant material on the outside wall is natural fibers

paredother, =1 if predominant material on the outside wall is other

pisomoscer, "=1 if predominant material on the floor is mosaic, ceramic, terrazo"

pisocemento, =1 if predominant material on the floor is cement

pisoother, =1 if predominant material on the floor is other

pisonatur, =1 if predominant material on the floor is natural material

pisonotiene, =1 if no floor at the household

pisomadera, =1 if predominant material on the floor is wood

techozinc, =1 if predominant material on the roof is metal foil or zink

techoentrepiso, "=1 if predominant material on the roof is fiber cement, mezzanine "

techocane, =1 if predominant material on the roof is natural fibers

techootro, =1 if predominant material on the roof is other

cielorazo, =1 if the house has ceiling

abastaguadentro, =1 if water provision inside the dwelling

abastaguafuera, =1 if water provision outside the dwelling

abastaguano, =1 if no water provision

public, "=1 electricity from CNFL, ICE, ESPH/JASEC"

planpri, =1 electricity from private plant

noelec, =1 no electricity in the dwelling

coopele, =1 electricity from cooperative

sanitario1, =1 no toilet in the dwelling

sanitario2, =1 toilet connected to sewer or cesspool

sanitario3, =1 toilet connected to septic tank

sanitario5, =1 toilet connected to black hole or letrine

sanitario6, =1 toilet connected to other system

energcocinar1, =1 no main source of energy used for cooking (no kitchen)

energcocinar2, =1 main source of energy used for cooking electricity

energcocinar3, =1 main source of energy used for cooking gas

energcocinar4, =1 main source of energy used for cooking wood charcoal

elimbasu1, =1 if rubbish disposal mainly by tanker truck

elimbasu2, =1 if rubbish disposal mainly by botan hollow or buried

elimbasu3, =1 if rubbish disposal mainly by burning

elimbasu4, =1 if rubbish disposal mainly by throwing in an unoccupied space

elimbasu5, "=1 if rubbish disposal mainly by throwing in river, creek or sea"

elimbasu6, =1 if rubbish disposal mainly other

epared1, =1 if walls are bad

epared2, =1 if walls are regular

epared3, =1 if walls are good

etecho1, =1 if roof are bad

etecho2, =1 if roof are regular

etecho3, =1 if roof are good

eviv1, =1 if floor are bad

eviv2, =1 if floor are regular

eviv3, =1 if floor are good

dis, =1 if disable person

male, =1 if male

female, =1 if female

estadocivil1, =1 if less than 10 years old

estadocivil2, =1 if free or coupled uunion

estadocivil3, =1 if married

estadocivil4, =1 if divorced

estadocivil5, =1 if separated

estadocivil6, =1 if widow/er

estadocivil7, =1 if single

parentesco1, =1 if household head

parentesco2, =1 if spouse/partner

parentesco3, =1 if son/doughter

parentesco4, =1 if stepson/doughter

parentesco5, =1 if son/doughter in law

parentesco6, =1 if grandson/doughter

parentesco7, =1 if mother/father

parentesco8, =1 if father/mother in law

parentesco9, =1 if brother/sister

parentesco10, =1 if brother/sister in law

parentesco11, =1 if other family member

parentesco12, =1 if other non family member

idhogar, Household level identifier

hogar\_nin, Number of children 0 to 19 in household

hogar\_adul, Number of adults in household

hogar\_mayor, # of individuals 65+ in the household

hogar\_total, # of total individuals in the household

dependency, Dependency rate, calculated = (number of members of the household younger than 19 or older than 64)/(number of member of household between 19 and 64)

edjefe, years of education of male head of household, based on the interaction of escolari (years of education), head of household and gender, yes=1 and no=0

edjefa, years of education of female head of household, based on the interaction of escolari (years of education), head of household and gender, yes=1 and no=0

meaneduc,average years of education for adults (18+)

instlevel1, =1 no level of education

instlevel2, =1 incomplete primary

instlevel3, =1 complete primary

instlevel4, =1 incomplete academic secondary level

instlevel5, =1 complete academic secondary level

instlevel6, =1 incomplete technical secondary level

instlevel7, =1 complete technical secondary level

instlevel8, =1 undergraduate and higher education

instlevel9, =1 postgraduate higher education

bedrooms, number of bedrooms

overcrowding, # persons per room

tipovivi1, =1 own and fully paid house

tipovivi2, "=1 own, paying in installments"

tipovivi3, =1 rented

tipovivi4, =1 precarious

tipovivi5, "=1 other(assigned, borrowed)"

computer, =1 if the household has notebook or desktop computer

television, =1 if the household has TV

mobilephone, =1 if mobile phone

qmobilephone, # of mobile phones

lugar1, =1 region Central

lugar2, =1 region Chorotega

lugar3, =1 region PacÃƒÂ­fico central

lugar4, =1 region Brunca

lugar5, =1 region Huetar AtlÃƒÂ¡ntica

lugar6, =1 region Huetar Norte

area1, =1 zona urbana

area2, =2 zona rural

age, Age in years

SQBescolari, escolari squared

SQBage, age squared

SQBhogar\_total, hogar\_total squared

SQBedjefe, edjefe squared

SQBhogar\_nin, hogar\_nin squared

SQBovercrowding, overcrowding squared

SQBdependency, dependency squared

SQBmeaned, square of the mean years of education of adults (>=18) in the household

agesq, Age squared