



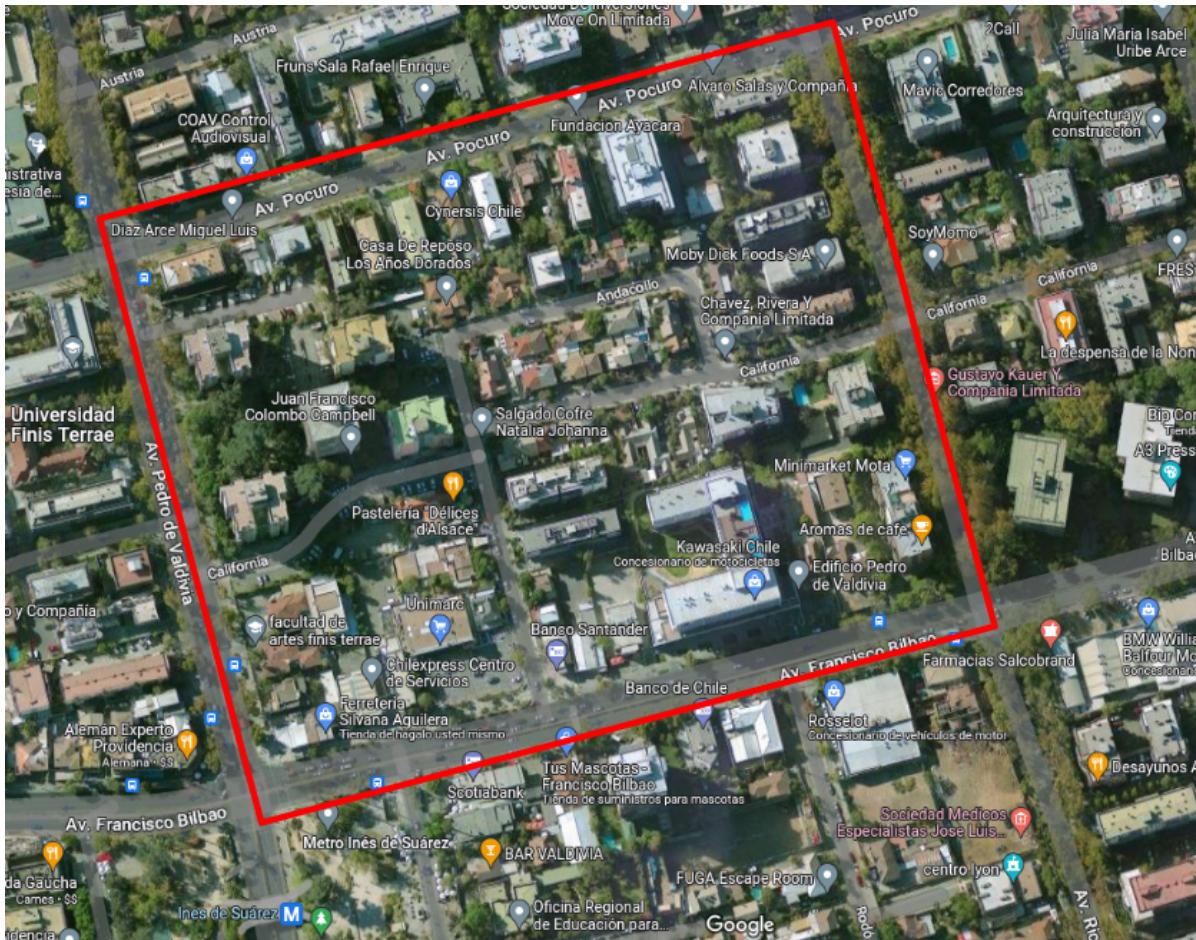
Data Science Accelerator Program

Socioeconomic stratification of the sampling frame of dwellings through the
use of satellite images

June 8, 2023

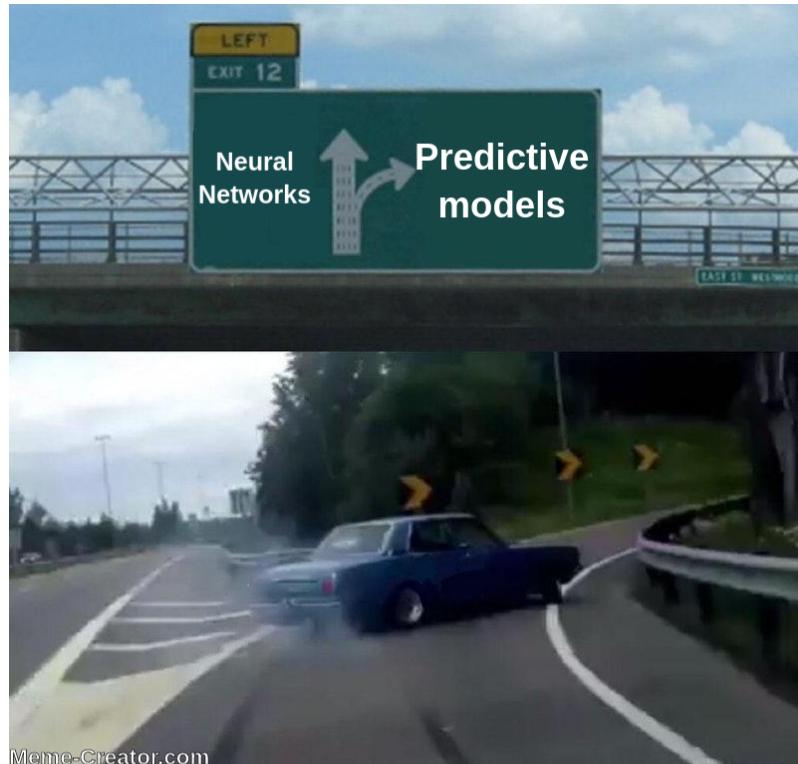
Objective 1

Stratify selected portions of the dwelling sampling frame in the **Metropolitan Region**, identifying PSU of **high, medium and low** strata, using satellite images and **visual recognition**.



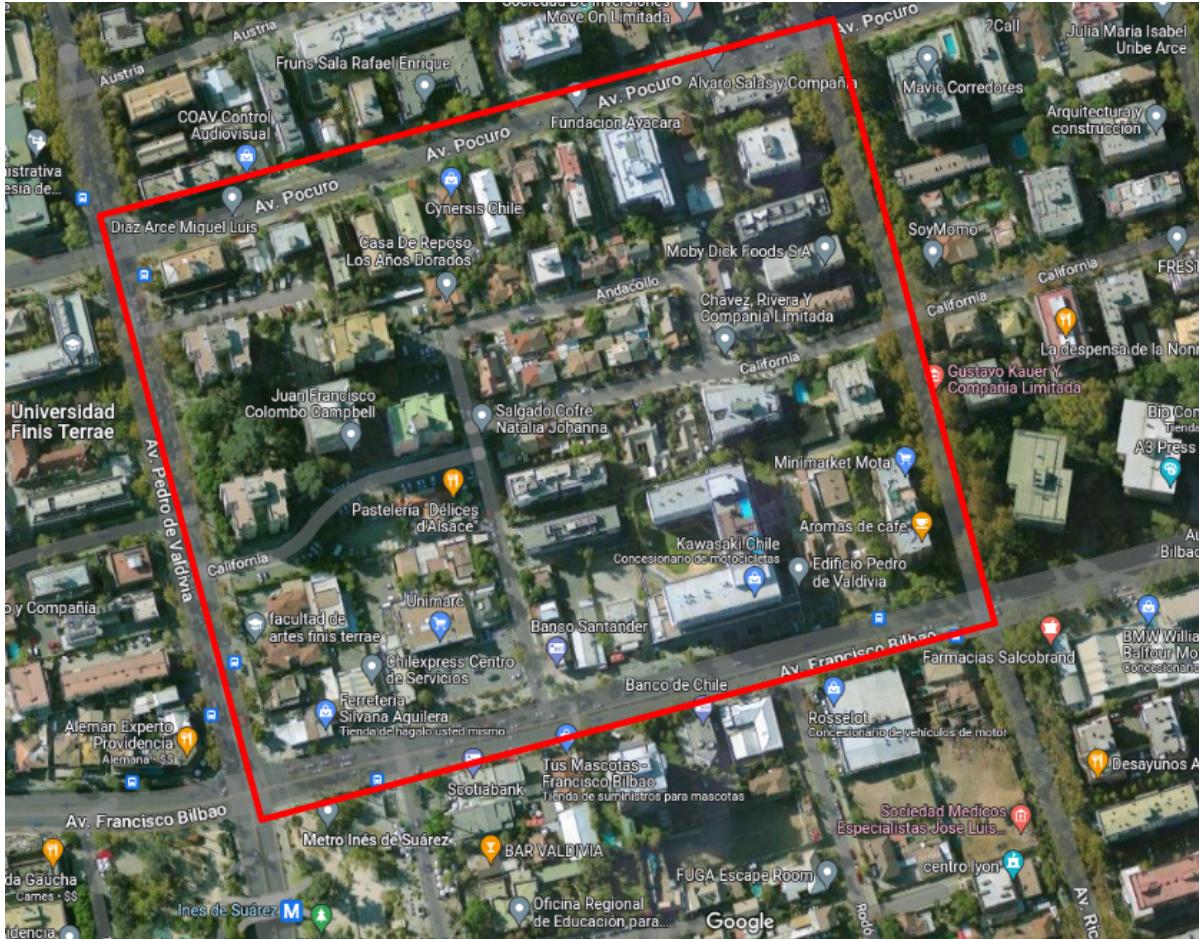
Objective 2

But along the way we've changed a bit...



Stratify selected portions of the dwelling sampling frame in the **Metropolitan Region**, identifying PSU income as a **continuous variable** , using satellite images, **visual recognition** and **and other data sources**.

Example of PSU



Dataset and labels

We have ~1450 labeled images (PSU)

The income data come from 3 surveys:

- Household budget survey (EPF)
- Socioeconomic characterization survey (CASEN)
- Supplementary Income Survey (ESI)



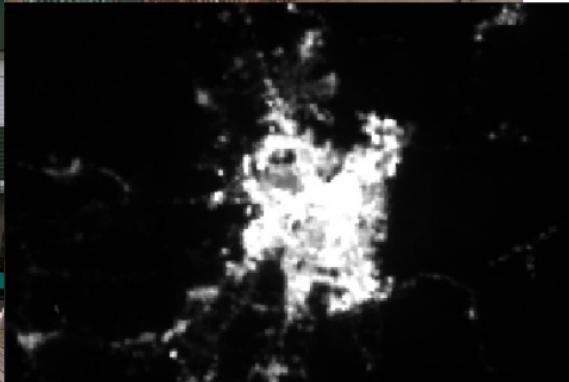
Features

Sentinel-2 bands

- indices (build index, vegetation index, etc.)
- bands
- operations over bands (minimum, maximum, median, sd)

indices + bands + operations = 148 features

Features



Training

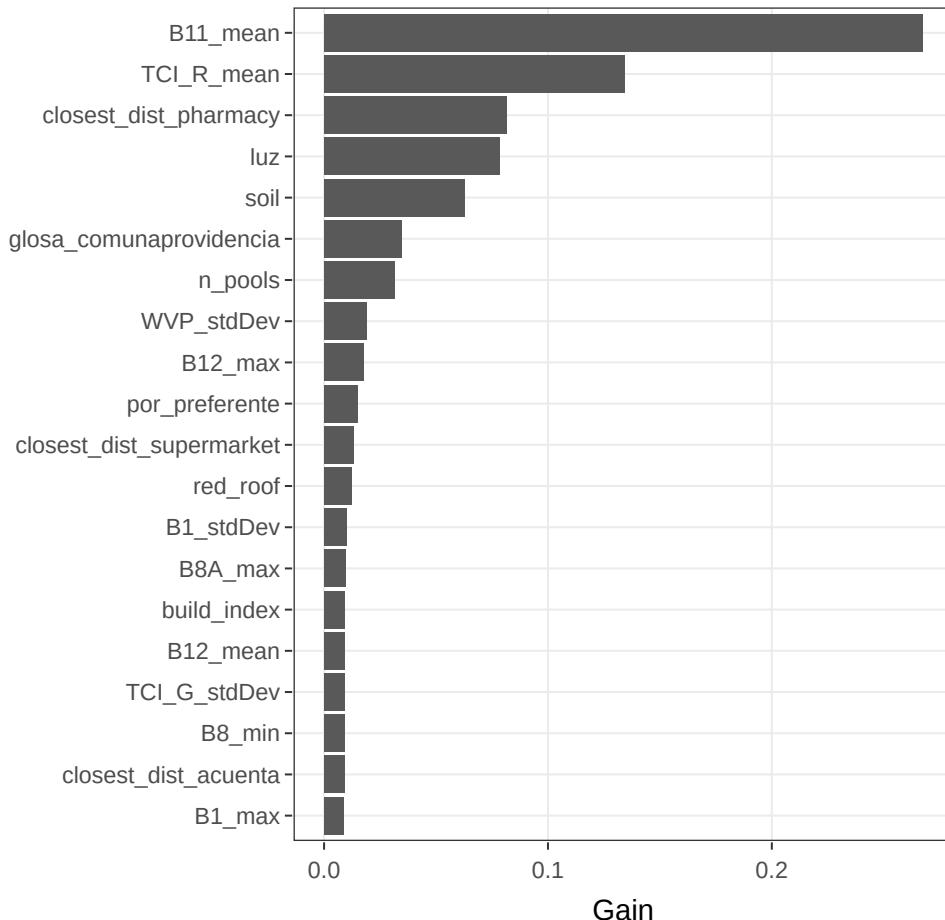
80/20 partition

xgboost

hyper parameters optimization

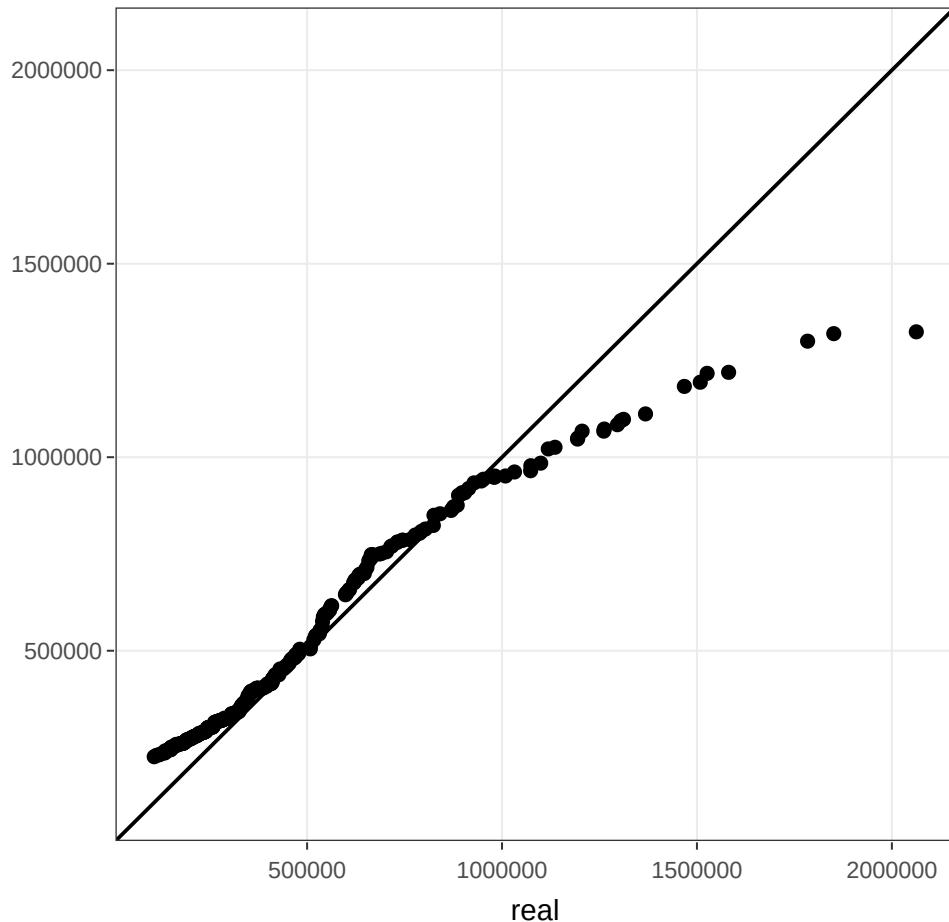
Results

Most Important Variables



Results

QQ plot. Predicted versus Real



Beyond the program

Roofs detection (SAM)



We will receive new data

- More data
- More consistent data

Shinyapp

Let's take a look at the app



Presentation DS Accelerator Program

Socioeconomic stratification of the sampling frame of dwellings through the
use of satellite images

June 8, 2023

