



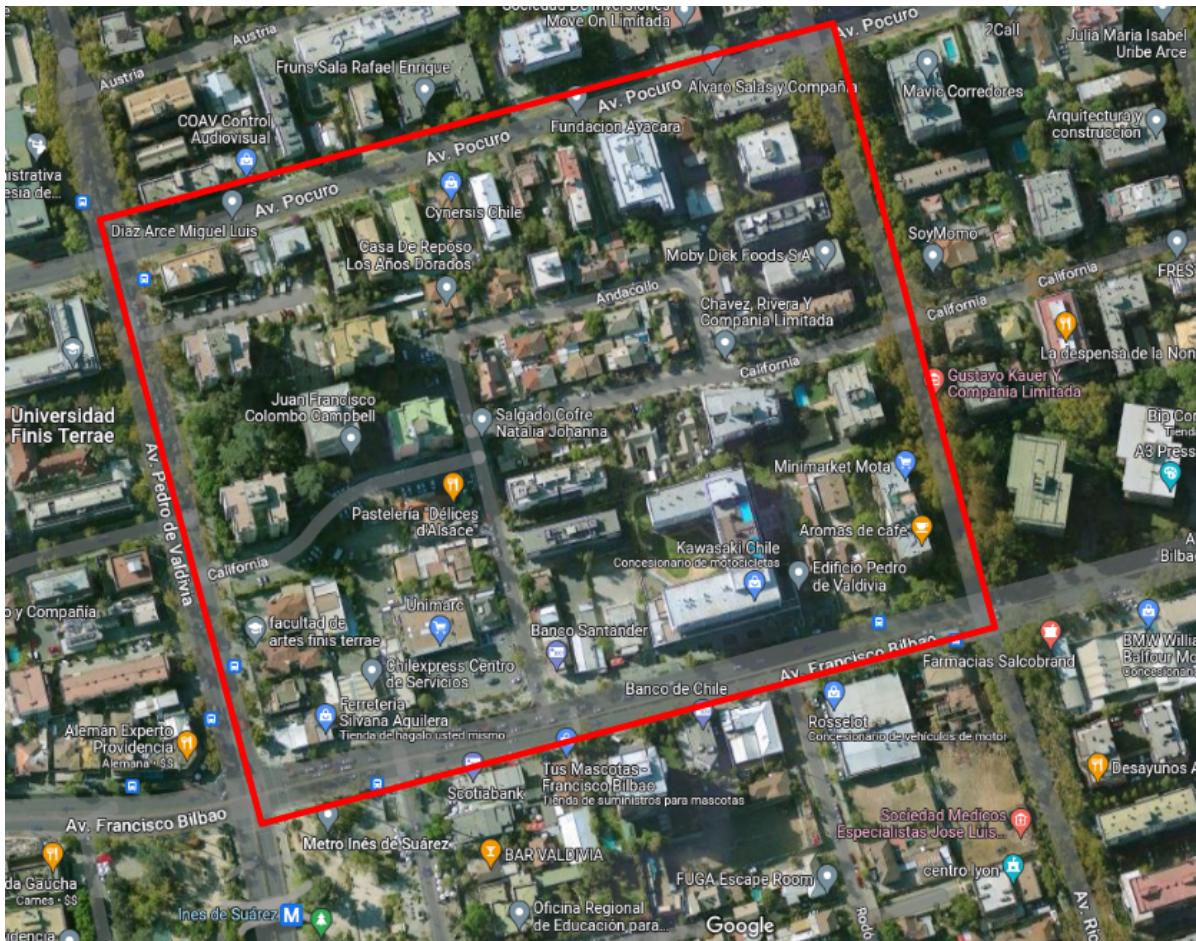
Estratificación socioeconómica del marco muestral de viviendas a través del uso de imágenes satelitales

Experiencia durante el programa de aceleración de ciencia de datos

Agosto 2023

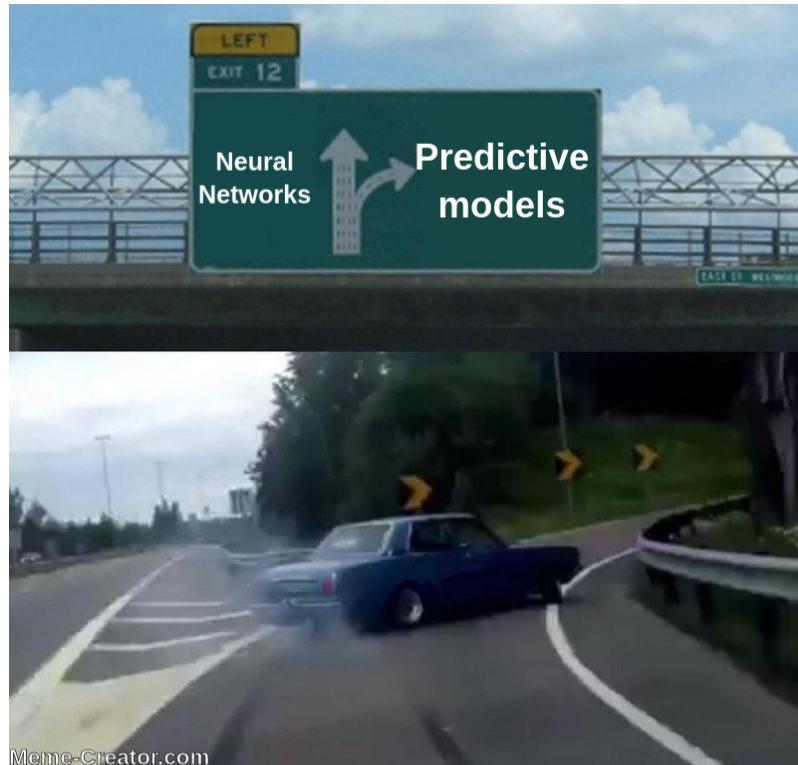
Objetivo inicial

Estratificar algunas porciones del marco muestral de viviendas en la región metropolitana, identificando UPM estratos alto, medio y bajo, por medio de imágenes satelitales y visión computacional



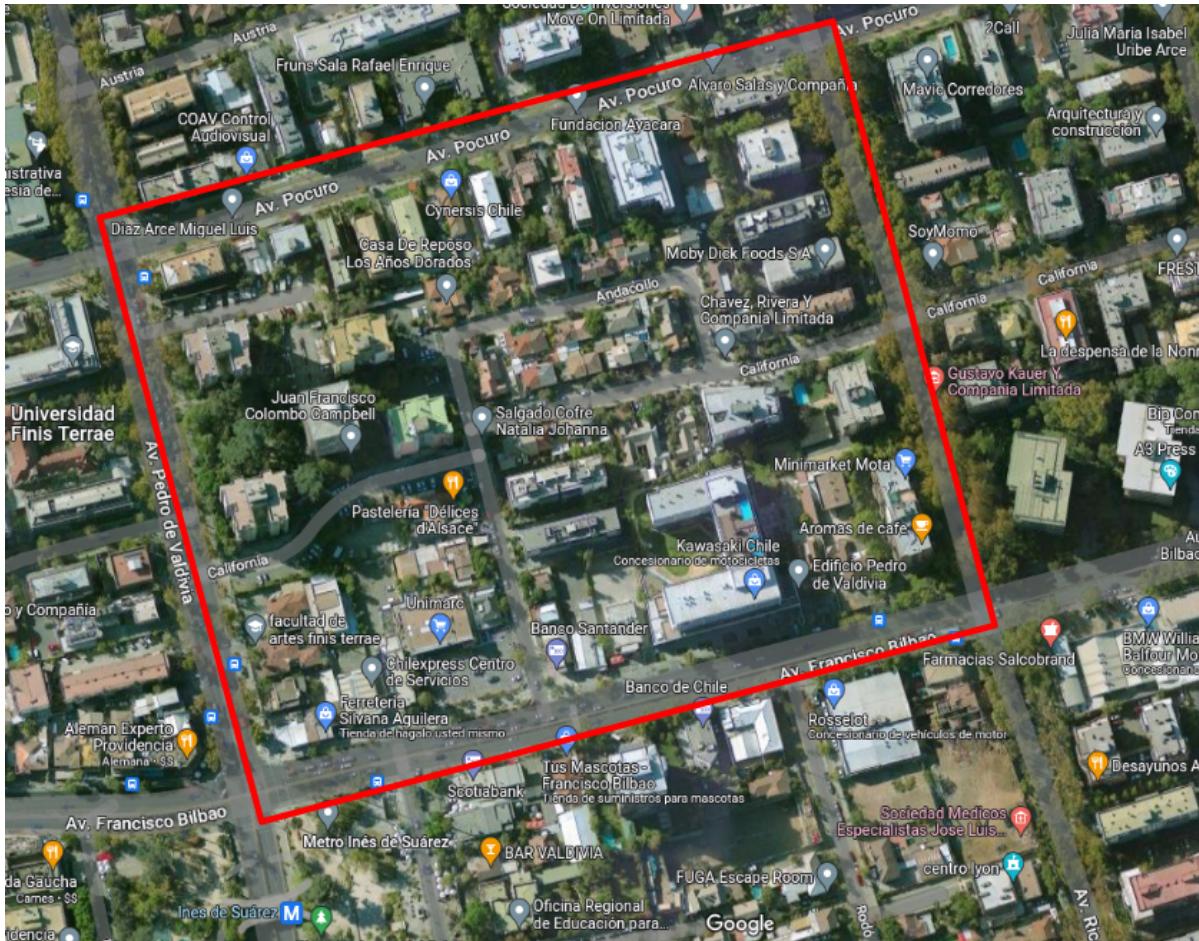
Objetivo alcanzable en el programa

En el camino las cosas cambiaron un poco...



Estratificar algunas porciones del marco muestral de viviendas en la región metropolitana, identificando el ingreso como una **variable continua**, por medio de imágenes satelitales y otras fuentes

Complejidad de la imagen



Dataset y etiquetas

Tenemos ~1450 UPMs (e imágenes) con datos de ingreso

Ingreso per cápita de 3 encuestas:

- EPF
- Piloto CASEN
- ESI



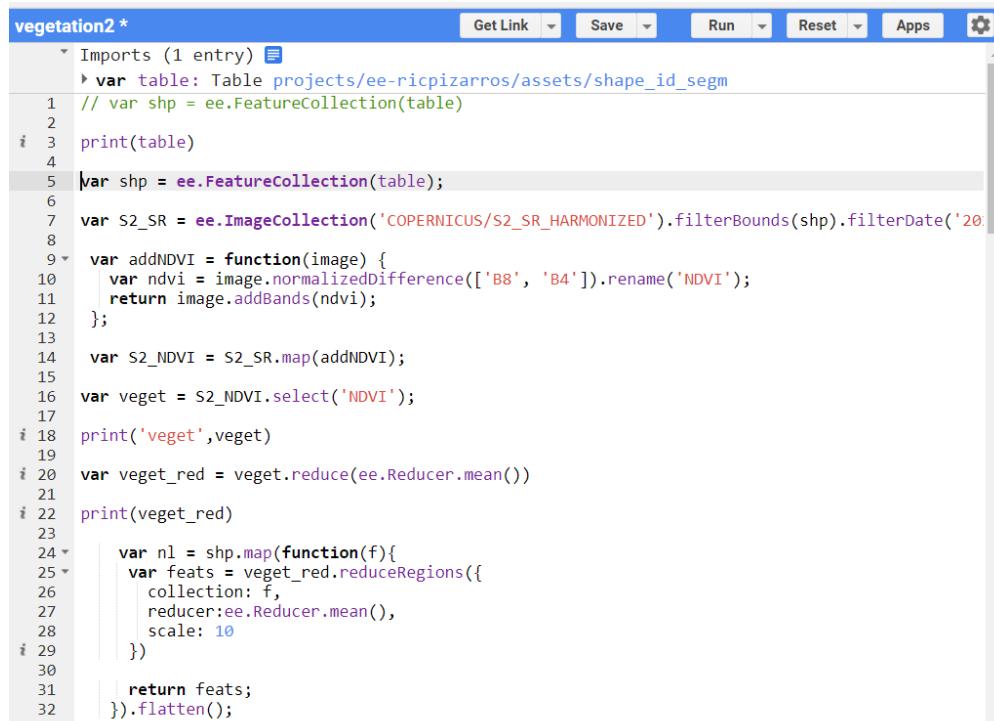
Extracción de características

google earth engine

índices

Bandas de Sentinel-2

- bandas
- operaciones sobre las bandas (mínimo, máximo, mediana, sd)
- índices (construcción, vegetación, etc.)



The screenshot shows the Google Earth Engine code editor interface. The title bar says "vegetation2 *". The toolbar includes "Get Link", "Save", "Run", "Reset", "Apps", and a settings gear icon. The code area contains the following JavaScript code:

```
vegetation2 *
  ▾ Imports (1 entry) ⓘ
    ▾ var table: Table projects/ee-ricpizarros/assets/shape_id_segm
      1 // var shp = ee.FeatureCollection(table)
      2
      i 3 print(table)
      4
      5 var shp = ee.FeatureCollection(table);
      6
      7 var S2_SR = ee.ImageCollection('COPERNICUS/S2_SR_HARMONIZED').filterBounds(shp).filterDate('2018-01-01/2018-02-28')
      8
      9 var addNDVI = function(image) {
        10   var ndvi = image.normalizedDifference(['B8', 'B4']).rename('NDVI');
        11   return image.addBands(ndvi);
        12 };
      13
      14 var S2_NDVI = S2_SR.map(addNDVI);
      15
      16 var veget = S2_NDVI.select('NDVI');
      17
      i 18 print('veget',veget)
      19
      i 20 var veget_red = veget.reduce(ee.Reducer.mean())
      21
      i 22 print(veget_red)
      23
      24 var nl = shp.map(function(f){
      25   var feats = veget_red.reduceRegions({
      26     collection: f,
      27     reducer:ee.Reducer.mean(),
      28     scale: 10
      29   })
      30
      31   return feats;
      32 }).flatten();
```

Más características

árboles

piscinas

luces nocturnas

características



Entrenamiento

Partición 80/20

xgboost

Optimizamos algunos hiperparámetros

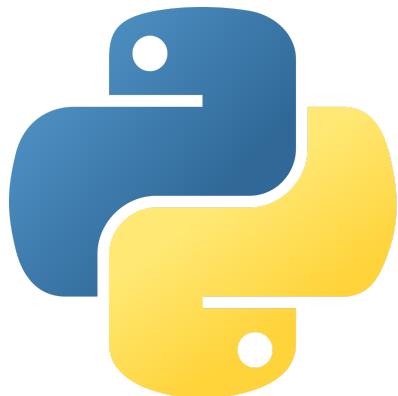
- rounds
- profundidad
- eta

Flujo completo

```
1 #!/bin/bash
2
3 # Trabajo con dataset de colegios
4 Rscript scripts/colegios.R
5
6 # Generar la tabla que contiene las etiquetas de las UPM
7 python3 scripts/create_psu_label.py
8
9 # Crear medidas de distancia
10 Rscript scripts/distance_upm.R
11
12 # Crear el data set con toda la información
13 Rscript scripts/feature_matrix.R
14
15 # Separar set de entrenamiento y testeo
16 python3 scripts/split_train_test.py
17
18 # Buscar hiperparámetros para xgboost
19 python3 scripts/tune_hyperparameter.py
20
21 #Generar inputs para el aplicativo
22 python3 scripts/train_many_models.py
23
```



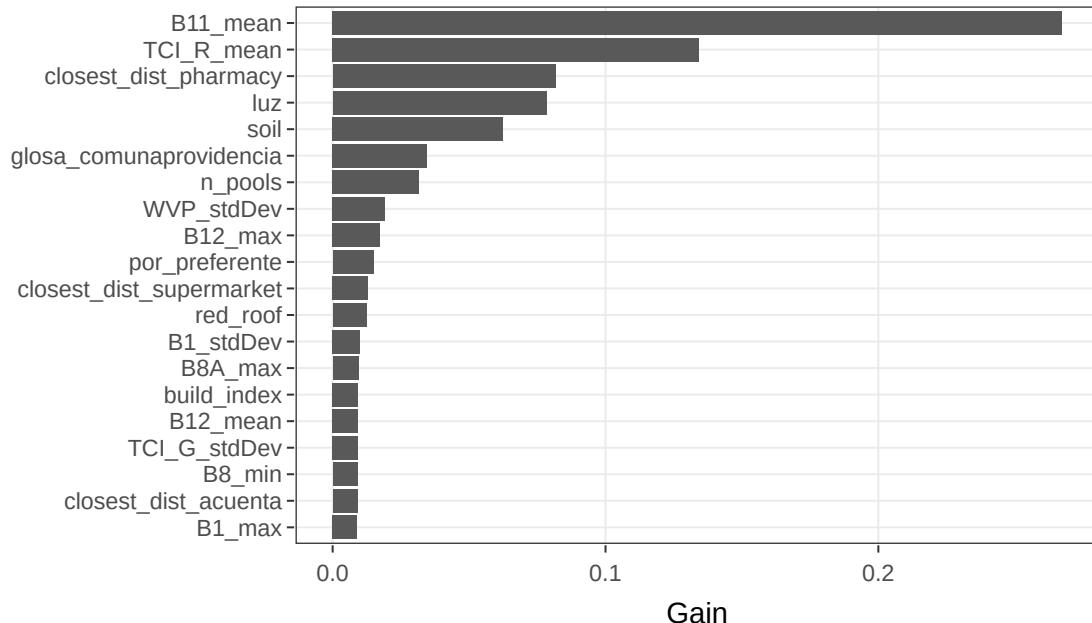
Download PNG PSD SVG EPS Formats In Single Zip
www.VectorSeek.com



Google Earth Engine

Resultados

Most important variables



Resultados

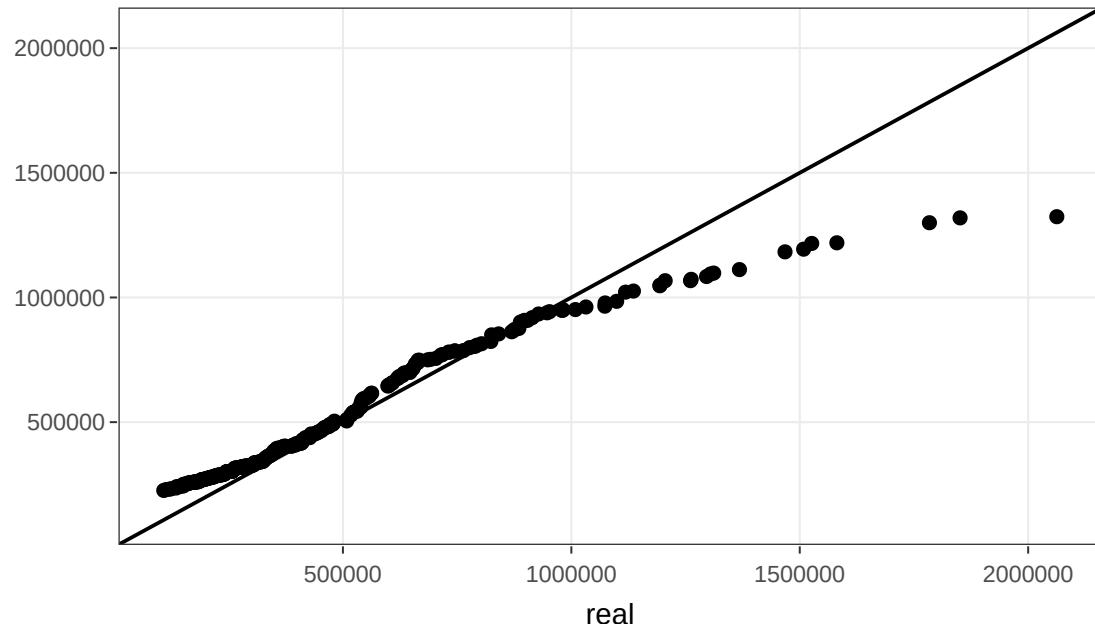
qqplot

line-plot

density-plot

scatter-plot

QQ plot. Predicted Versus real



Más allá del programa

Detección de techos (SAM)



Más datos

- CASEN oficial
- EPF

Veamos un prototipo



Estratificación socioeconómica del marco muestral de viviendas a través del uso de imágenes satelitales

Experiencia durante el programa de aceleración de ciencia de datos

Agosto 2023

