

PS1 - Equipo 03

Catalina Leal Rojas, Lucas Daniel Carrillo Aguirre, Lucas Eduardo Veras
Costa

17 de octubre de 2025

El link del repositorio es: https://github.com/mbastol06/PS1_Equipo3

1. Introducción

2. Datos

Los datos empleados en este análisis provienen del DANE y de la misión para el “Empalme de las Series de Empleo, Pobreza y Desigualdad - MESE”. En el repositorio *Kaggle* se encuentran cuatro conjuntos de datos divididos en muestras de entrenamiento y de prueba, tanto a nivel de hogar como a nivel individual. En ambos casos, es posible vincular las bases de hogar y persona mediante la variable *id*. Como la herramienta de *GitHub* en su versión más básica no nos permite trabajar simultáneamente con los 4 archivos - debido a que su tamaño ultrapasa los 45MB - cada persona del grupo descargó los datos en su computadora local.

Varias de las variables contenidas en el conjunto de prueba conservan los mismos códigos utilizados en la encuesta original del DANE. Por esta razón, se realizó una consulta al diccionario de la encuesta y se modificaron los nombres de las variables con el fin de hacerlos más interpretables, facilitando así el proceso de análisis.

Es importante señalar que las variables incluidas en el conjunto de datos de entrenamiento no coinciden completamente con las del conjunto de prueba. Dado que únicamente pueden emplearse como predictores las variables presentes en el conjunto de prueba, se procedió a eliminar del conjunto de entrenamiento aquellas variables que no se encuentran disponibles en el conjunto de prueba.

En el conjunto de datos individuales se optó por no usar las variables relacionados a ingreso una vez que las mismas contienen una gran cantidad de datos no disponibles. Por ejemplo, para la variable que indicar el ingreso neto mensual no hay datos disponible para alrededor de 80 % para los individuos de la muestra. Sin embargo, nos parece relevante incluir si el jefe del hogar está ocupado o no pues es intuitivo pensar que en hogares pobres hay más chance que el jefe de hogar esté desempleado.

3. Modelos y Resultados
4. Conclusiones
5. Anexos