TALLER 1

BIG DATA Y MACHINE LEARNING PARA ECONOMÍA APLICADA

MECA

# INTRODUCCIÓN

# DATOS

La base de datos utilizada proviene de la Gran Encuesta Integrada de Hogares-GEIH de 2018, es realizada por el DANE en la que se recogen datos sobre las condiciones de empleo, las condiciones de la ocupación que tienen, temas de seguridad social, o si están en busca de un empleo. Además, se definen algunas variables para caracterizar a los hogares como temas de edad, sexo o nivel educativo.

Personas empleadas mayores de 18 años

## Descripción de los datos

Los datos se tomaron de la página presentada para el taller[[1]](#footnote-1). Los datos solo cuentan con información para Bogotá D.C. y se encuentran distribuidos en 10 sub grupos que se se unen para tener la base final de trabajo.

## Proceso de adquisición de los datos

## Limpieza de los datos

## Análisis descriptivo de los datos

# PERFIL EDAD-SALARIO

Regresión

Interpretación de los coeficientes

Ajuste del modelo

Gráfico del perfil estimado

# BRECHA SALARIAL DE GÉNERO

# PREDICCIÓN DE LAS GANANCIAS POR MODELO

Comparar los modelos vistos en clase

Repositorio GirHub:

1. <https://ignaciomsarmiento.github.io/GEIH2018_sample/> [↑](#footnote-ref-1)