

# TRABAJO PRÁCTICO N° 2

## Histogramas, Kernels & Métodos no supervisados usando la EPH

Ricardo Javier Gutiérrez Vistín - Martín Gabriel Cargnel

Link al repositorio: <https://github.com/mcargnel/Big-Data-UBA-Grupo-001/tree/main/TP1>

### 1 Creación de variables, histogramas, kernels y resumen de la base de datos final

Tabla 1: Tabla con la cantidad y proporción de pobres por año

	Año	Pobre	Cantidad	Proporción
0	2005.0	0	6878	0.7340
2	2005.0	1	2493	0.2660
1	2025.0	0	2950	0.6889
3	2025.0	1	1332	0.3111

### 2 Métodos No Supervisados

Parte I - Creación de variables, histogramas, kernels y resumen de la base de datos final

La distribución de edades muestra una concentración importante en los grupos jóvenes, principalmente entre los 10 y 30 años, que representan alrededor de un tercio de la población analizada, mientras que la frecuencia disminuye progresivamente a medida que aumenta la edad. En términos de pobreza, los datos indican que la incidencia es mayor en los grupos etarios más jóvenes con más del 40% de personas pobres entre 0 y 30 años y tiende a reducirse en los grupos de mayor edad, donde las tasas descienden por debajo del 15%. Esto sugiere que la vulnerabilidad económica afecta con mayor intensidad a la población joven, posiblemente debido a su menor estabilidad laboral y menor acumulación de capital humano.

La distribución del ingreso total familiar se concentra fuertemente en los niveles bajos, ya que casi el 98% de los hogares tiene ingresos inferiores a 9,3 millones de pesos de 2025, mostrando

una clara asimetría hacia la derecha. En cuanto a la condición de pobreza, se observa que los hogares pobres se concentran principalmente en los tramos de menor ingreso, mientras que en los niveles altos prácticamente no existen, reflejando una marcada desigualdad en la distribución del ingreso.

### **3 Métodos no supervisados**

#### **3.1 PCA**

#### **3.2 Cluster**