



Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado



Informe TP 1 – Taller de Programación

PRIMER ENCUENTRO CON LA ENCUESTA PERMANENTE DE HOGARES (EPH)

Maestría: Economía Aplicada

Materia: Taller de Programación

Profesor: Romero, María Noelia

Alumnos: Godino, Franco; Sulca, Ronny; y Alvarez, Martín

Año 2025

Enlace al repositorio: <https://github.com/alvarezmartin2411-debug/Big-Data-UBA--Grupo-2/tree/main>

INTRODUCCIÓN

El presente trabajo práctico tiene como objetivo introducirnos en el manejo y análisis de bases de datos socioeconómicas utilizando herramientas de programación aplicadas a la Encuesta Permanente de Hogares (EPH), elaborada por el Instituto Nacional de Estadística y Censos (INDEC). A lo largo del informe se realiza un proceso completo que abarca desde la descarga y limpieza de microdatos correspondientes al primer trimestre de los años 2005 y 2025, hasta la exploración descriptiva de variables sociodemográficas y laborales relevantes, con el fin de comprender su evolución en el tiempo. Asimismo, se aborda la medición de la pobreza a nivel de hogares mediante la incorporación del concepto de adulto equivalente y el cálculo del ingreso necesario en relación con la Canasta Básica Total, lo que permite identificar la proporción de población en situación de pobreza en ambos períodos.

INFORME

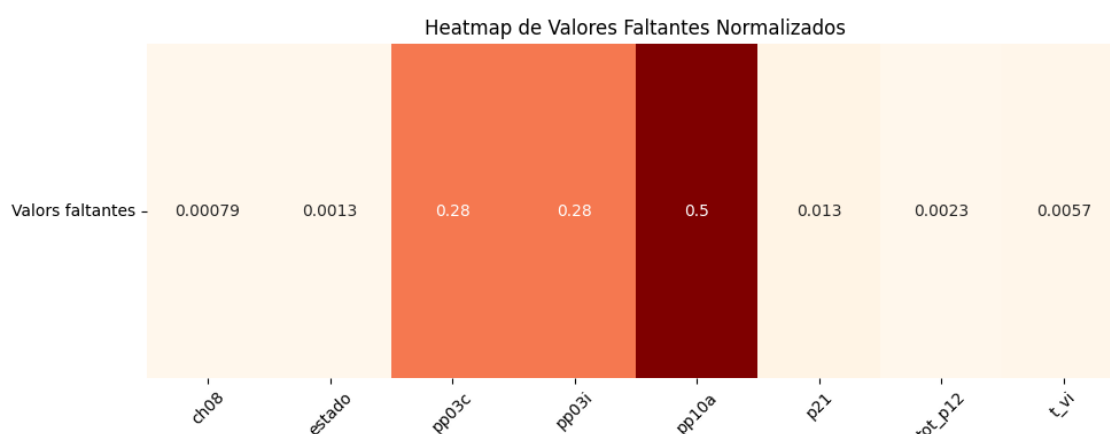
PARTE I: FAMILIARIZÁNDONOS CON LA BASE EPH Y LIMPIEZA

CONSIGNA 1

Según la metodología oficial utilizada por el INDEC, una persona o un hogar se considera pobre cuando su ingreso total familiar (ITF) no alcanza el valor de la Canasta Básica Total (CBT) correspondiente a su composición y tamaño. La CBT representa el monto mínimo necesario para cubrir tanto la canasta básica alimentaria, que asegura los requerimientos nutricionales esenciales, como un conjunto de bienes y servicios no alimentarios indispensables, como transporte, vestimenta, educación y salud. Para realizar esta comparación, se ajusta la canasta según la cantidad de integrantes del hogar mediante el concepto de adulto equivalente, que pondera las necesidades de consumo según edad y sexo. Si el ingreso disponible del hogar es inferior al ingreso necesario determinado por la CBT ajustada por adulto equivalente, todas las personas que integran ese hogar son clasificadas como pobres.

CONSIGNA 2

Se descargaron y procesaron los microdatos correspondientes al primer trimestre de los años 2005 y 2025 de la Encuesta Permanente de Hogares (EPH), disponibles en la página oficial del INDEC. A partir de estos archivos, se seleccionó la región Noroeste del país para trabajar, filtrando la base de datos según la variable REGION y eliminando los registros que no pertenecían a dicha zona. Posteriormente, ambas bases fueron unificadas en un único archivo para facilitar el análisis comparativo entre los dos períodos. Asimismo, se seleccionaron 15 variables de interés, entre las cuales se incluyen las obligatorias (CH04, CH06, CH07, CH08, NIVEL_ED, ESTADO, CAT_INAC, IPCF), con el objetivo de realizar un análisis descriptivo y evaluar su calidad y completitud. A continuación, se elaboró un heatmap de valores faltantes que permitió visualizar de forma gráfica la proporción de datos ausentes en cada variable y año.



Los resultados muestran que la cantidad de valores faltantes varía entre las variables y los períodos analizados. En general, se observa que el año 2025 presenta un mayor número de valores ausentes en comparación con 2005, reflejando un desafío creciente en la calidad de los datos reportados en la EPH. Entre las variables con mayor proporción de

datos faltantes se destaca IPCF (Ingreso Per Cápita Familiar), lo cual resulta consistente con la literatura que señala un aumento en la no respuesta a las preguntas vinculadas a ingresos. Este análisis inicial permitió identificar las variables que requieren un tratamiento especial en el proceso de limpieza, en el cual se eliminaron valores sin sentido (como ingresos negativos).

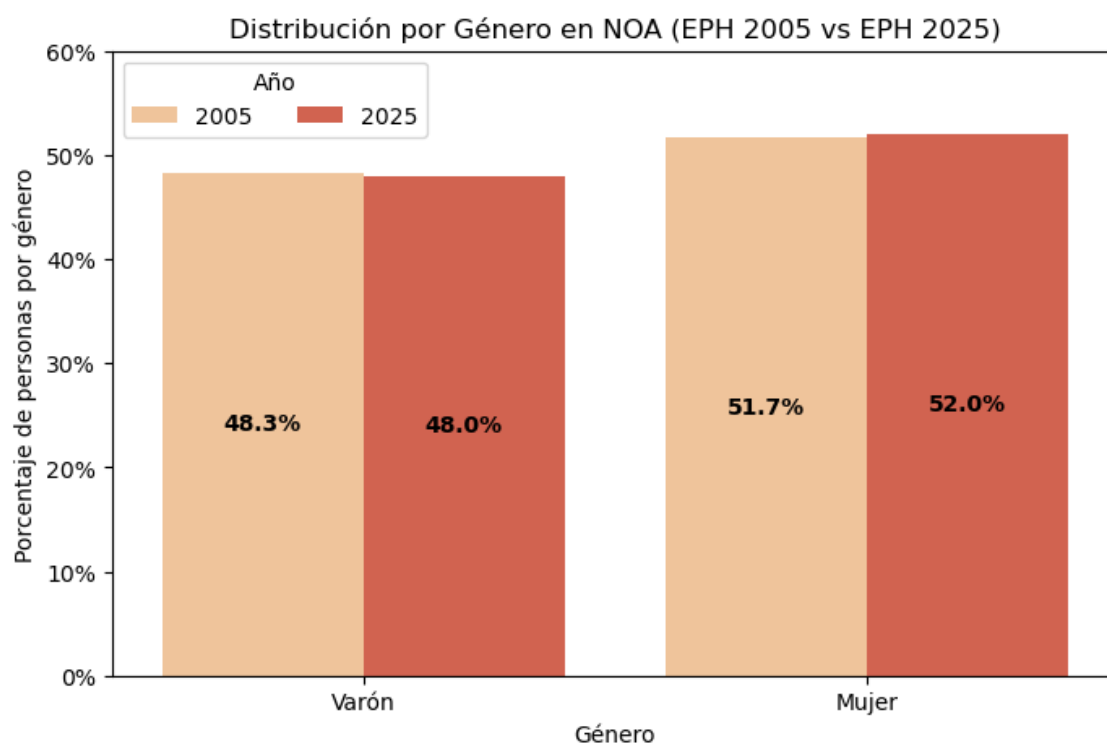
Reporte de unión de base Hogares e Individuos de la EPH.

Tipo de unión	Nro. de filas	Nro. de Columnas	Total de NAs
unión por intersección	19090	24	20734
unir por izquierda	19090	24	20734
unir por derecha	19090	24	20734
unión conjunta	19090	24	20734

PARTE II: PRIMER ANÁLISIS EXPLORATORIO

CONSIGNA 3

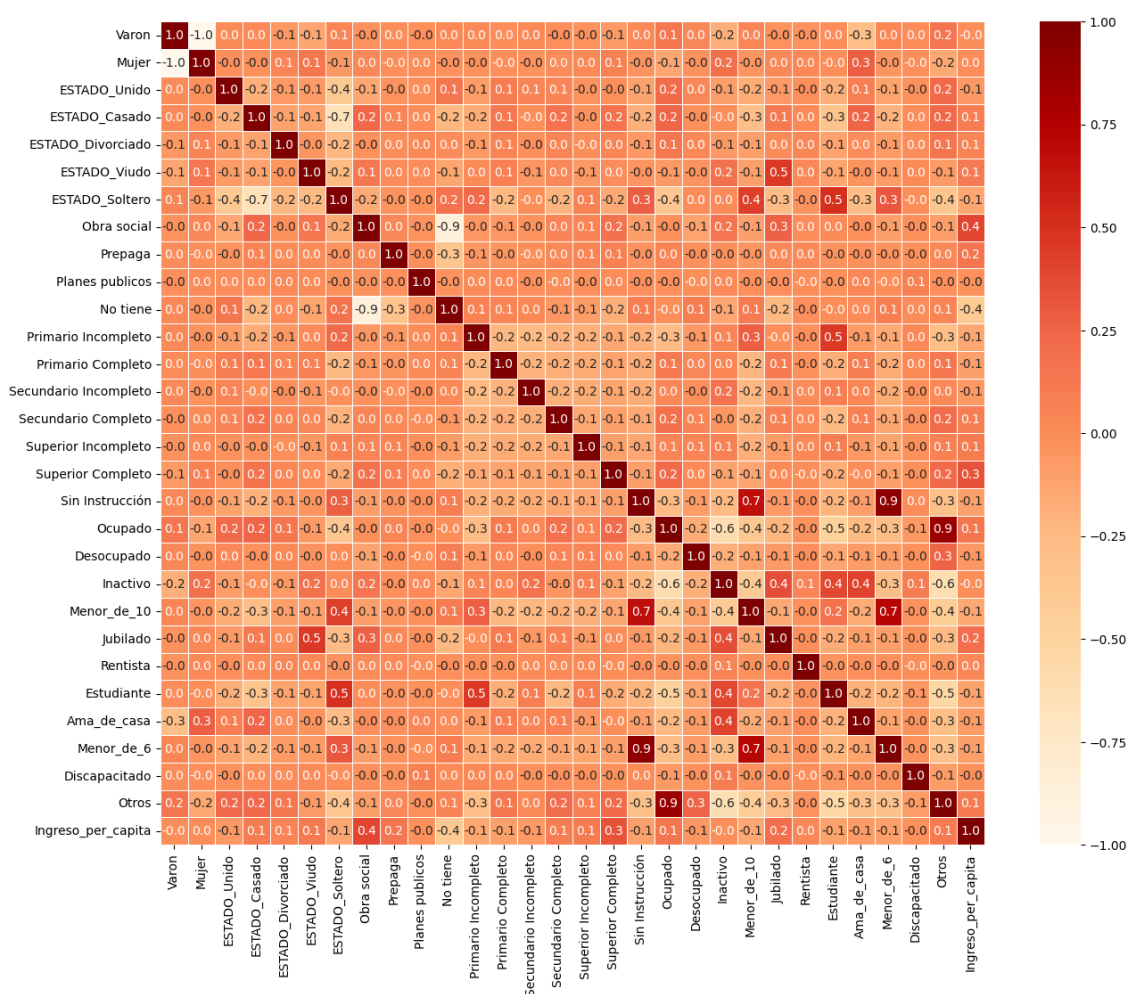
El análisis de la composición por sexo para los años 2005 y 2025 en la región seleccionada muestra que la distribución entre varones y mujeres se mantiene relativamente estable a lo largo del tiempo, sin cambios estructurales significativos en la participación de cada grupo. En ambos períodos, la población femenina representa una proporción levemente mayor que la masculina.

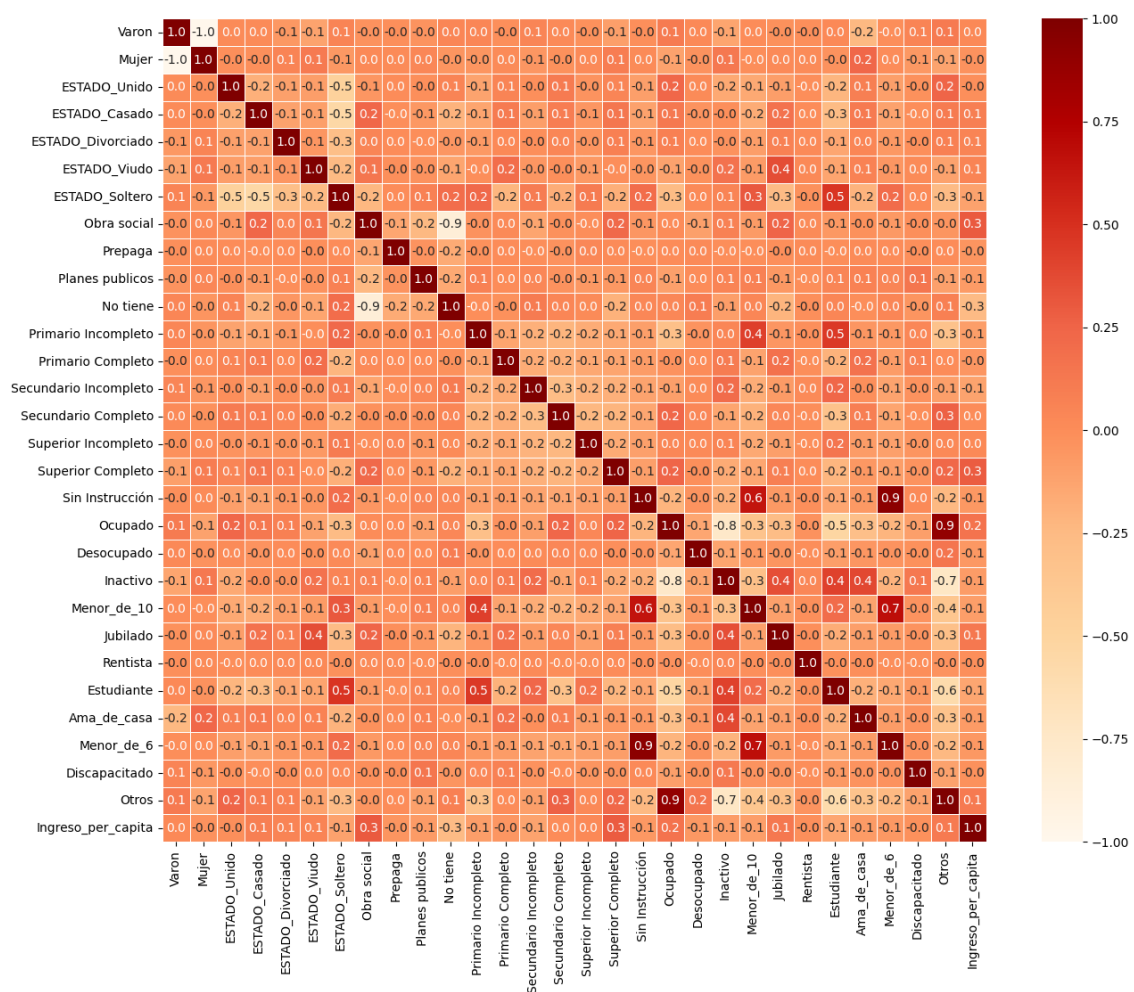


CONSIGNA 4

La matriz de correlaciones permite observar las relaciones existentes entre las principales variables sociodemográficas y laborales de la población en los años 2005 y 2025. En ambos períodos, se destaca una fuerte asociación positiva entre el nivel educativo (NIVEL_ED) y el ingreso per cápita familiar (IPCF), lo cual confirma la relevancia de este variable como determinante de los ingresos y las condiciones socioeconómicas de los hogares. Además, se observa una correlación significativa entre la condición de actividad (ESTADO) y la categoría de inactividad (CAT_INAC). Las variables de edad (CH06) y nivel educativo también muestran una relación positiva moderada, lo que implica que a mayor edad suele alcanzarse un mayor nivel educativo, probablemente vinculado a trayectorias laborales más largas o a la finalización de estudios en etapas posteriores de la vida.

Matriz de correlaciones de las variables seleccionadas de EPH. Año 2005.



Matriz de correlaciones de las variables seleccionadas de EPH. Año 2025.

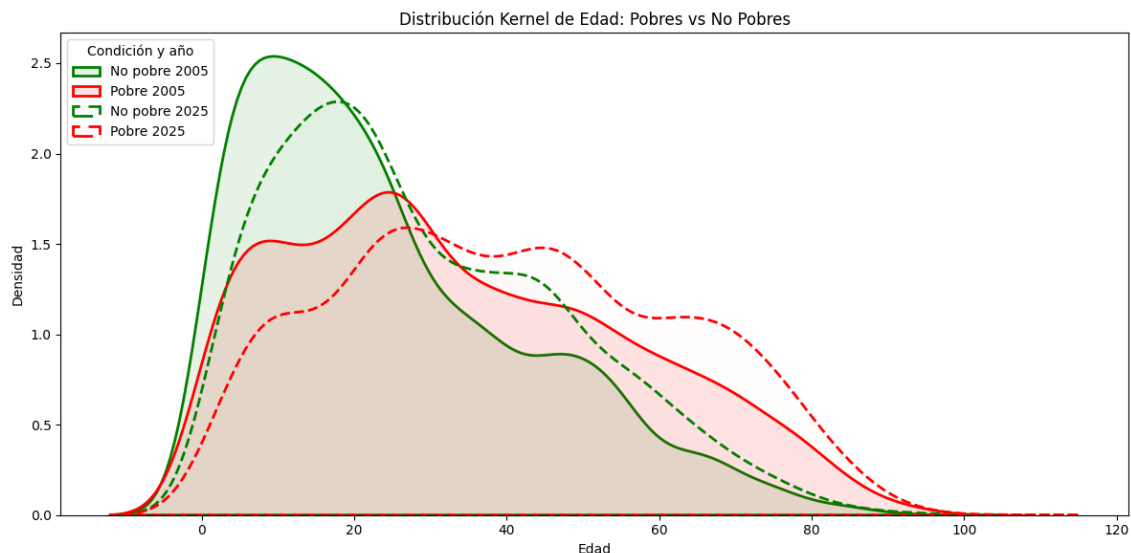
Al comparar los dos años analizados, se aprecia que las correlaciones tienden a fortalecerse en 2025 respecto de 2005, particularmente en el vínculo entre educación e ingresos, lo que podría estar asociado a cambios estructurales en el mercado laboral y a la creciente importancia de las credenciales educativas en la determinación de las oportunidades económicas. Por otro lado, la relación entre las variables sociodemográficas básicas (como sexo y edad) y las variables económicas es más débil, lo que indica que por sí solas no explican en gran medida las diferencias en el ingreso o en la situación laboral.

PARTE III: CONOCIENDO A LOS POBRES Y NO POBRES**CONSIGNA 8**

Para el primer trimestre de 2005, se identificaron 4.072 personas en situación de pobreza sobre un total de 9.278 observaciones, lo que representa un 43,9% de la muestra. Veinte años después, en el primer trimestre de 2025, la cantidad absoluta de personas pobres disminuye levemente a 3.888, sobre un total también menor de 8.695 observaciones, aunque el porcentaje de pobreza se eleva a 44,7%, reflejando un aumento relativo en la incidencia de la pobreza en comparación con el período anterior.

CONSIGNA 9

El análisis de la variable “pobre” en relación con la edad muestra que la pobreza afecta de manera desproporcionada a los sectores más jóvenes de la población. Tal como se observa en el gráfico, los individuos menores de 30 años concentran un porcentaje significativamente mayor de pobreza en comparación con los grupos etarios más avanzados. Este patrón se mantiene tanto en 2005 como en 2025.



Por otro lado, el gráfico que relaciona la pobreza con el nivel educativo revela una correlación negativa clara: la pobreza es considerablemente más elevada entre las personas con bajo nivel de instrucción, especialmente aquellas que no completaron la educación secundaria. En cambio, a medida que aumenta el nivel educativo, la incidencia de la pobreza disminuye de forma notable.

