

**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**

**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) siguiendo las indicaciones del diccionario de variables del INDEC, dejando la base lista para los análisis exploratorios posteriores.

Gráfico, Gráfico de rectángulos

El contenido generado por IA puede ser incorrecto.

**Tabla 1. Reporte de union 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 |

Son todos iguales básicamente por el hecho de que se deben corresponder uno a uno. Es decir, para cada persona hay un hogar asociado y viceversa, por lo tanto cada intento de unión arroja el mismo resultado.

**PARTE II**

**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, lo cual es consistente con la estructura demográfica nacional y con la tendencia observada en otras regiones del país. Sin embargo, se observa un ligero aumento en la participación relativa de las mujeres en 2025 respecto de 2005, lo que puede vincularse tanto a cambios demográficos naturales (como la mayor esperanza de vida femenina) como a transformaciones en la dinámica del mercado laboral y los patrones de participación económica. Este resultado constituye un primer acercamiento descriptivo al perfil sociodemográfico de la población analizada y servirá como base para interpretar con mayor profundidad las diferencias en otras dimensiones, como el nivel educativo, la condición de actividad y la situación de pobreza.

Gráfico, Gráfico en cascada

El contenido generado por IA puede ser incorrecto.