



UBA FCE
Universidad de Buenos Aires
Facultad de Ciencias Económicas

Taller de Programación – Trabajo Práctico N.º 2 (2025)

Tema: Métodos No Supervisados y Visualización

Grupo 4

Ángel Enrique Zapata Barros

David Andrés Robalino Chica

Federico Walter Kiswa

Docente: María Noelia Romero

Repositorio GitHub del Grupo:

https://github.com/Analisis1983/Grupo4_UBA_2025

1 Introducción

El presente trabajo aplica métodos no supervisados de análisis estadístico —principalmente Análisis de Componentes Principales (PCA) y Clustering— sobre una base combinada de la Encuesta Permanente de Hogares (EPH) correspondiente a los años 2005 y 2025, con el objetivo de identificar patrones socioeconómicos en la región patagónica. A partir de variables demográficas, educativas y de ingresos, se busca reconocer agrupamientos y relaciones latentes que permitan caracterizar distintos perfiles de hogares sin la necesidad de una variable dependiente predefinida. Este abordaje integra los contenidos teóricos y prácticos del Taller de Programación, articulando la programación en Python con la interpretación empírica de resultados, en coherencia con los fundamentos de la estadística aplicada y el aprendizaje automático. De esta manera, el trabajo combina rigor metodológico e intencionalidad analítica, contribuyendo a la comprensión de la estructura social y económica patagónica en un contexto temporal comparativo.

2 Base de datos, preparación y metodología

El trabajo utiliza información de la Encuesta Permanente de Hogares (EPH) de los años 2005 y 2025, combinada en una base única para la región patagónica. La integración de ambas bases permitió comparar cambios en variables demográficas, educativas y económicas, con especial énfasis en el ingreso total familiar (ITF), la edad y el nivel educativo de los individuos. El proceso de preparación incluyó la limpieza de datos, la eliminación de registros incompletos, la normalización de categorías y la creación de una variable numérica de educación para facilitar el análisis cuantitativo. Una vez depurada la base, las variables fueron estandarizadas mediante escalamiento y posteriormente analizadas con técnicas de aprendizaje no supervisadas. En particular, se aplicó el Análisis de Componentes Principales (PCA) para reducir la dimensionalidad y los métodos de K-Means y Clustering Jerárquico para identificar grupos con características socioeconómicas similares. La metodología adoptada combina programación en Python y análisis estadístico.

Tabla 1

Estadísticas descriptivas comparativas entre 2005 y 2025

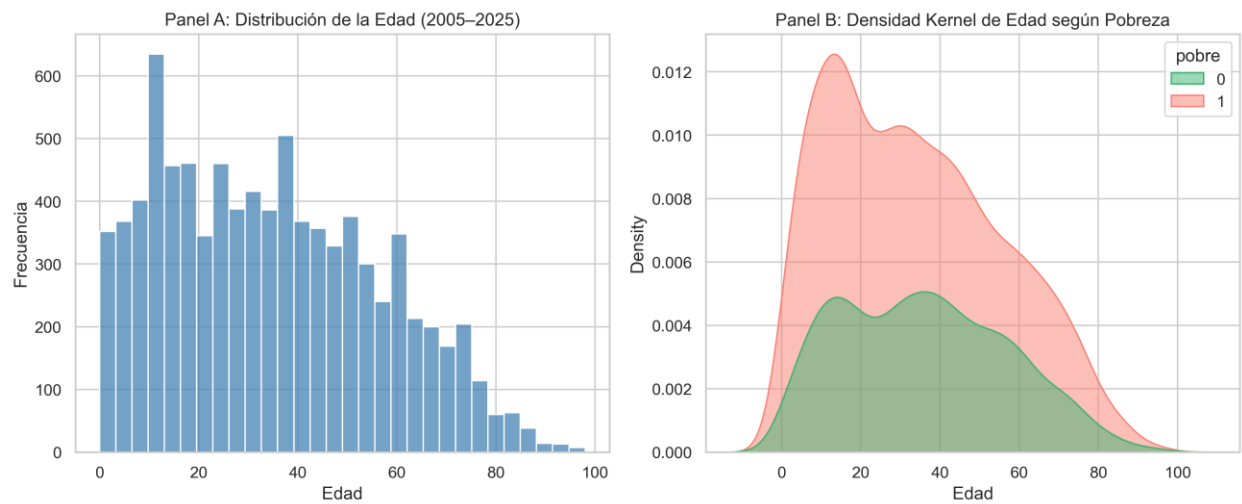
Año	Ingreso familiar total (media)	Ingreso familiar total (mediana)	Edad promedio (años)
2005	21 201,32	16 800,00	29,69
2025	1 550 391,22	1 250 000,00	37,63

Nota. Fuente: elaboración propia con base en la Encuesta Permanente de Hogares (EPH), procesamiento en Python (Grupo 4, 2025).

El ingreso familiar total registra un incremento nominal sustancial entre 2005 y 2025, observable tanto en los medios como en la mediana, reflejando la evolución de los valores monetarios en pesos argentinos. Este aumento responde a factores macroeconómicos como la inflación y la expansión del ingreso nominal, sin implicar necesariamente mejoras reales en el poder adquisitivo. La edad promedio aumenta levemente, lo que sugiere una población económicamente activa de mayor edad y un cambio demográfico gradual.

Figura 1

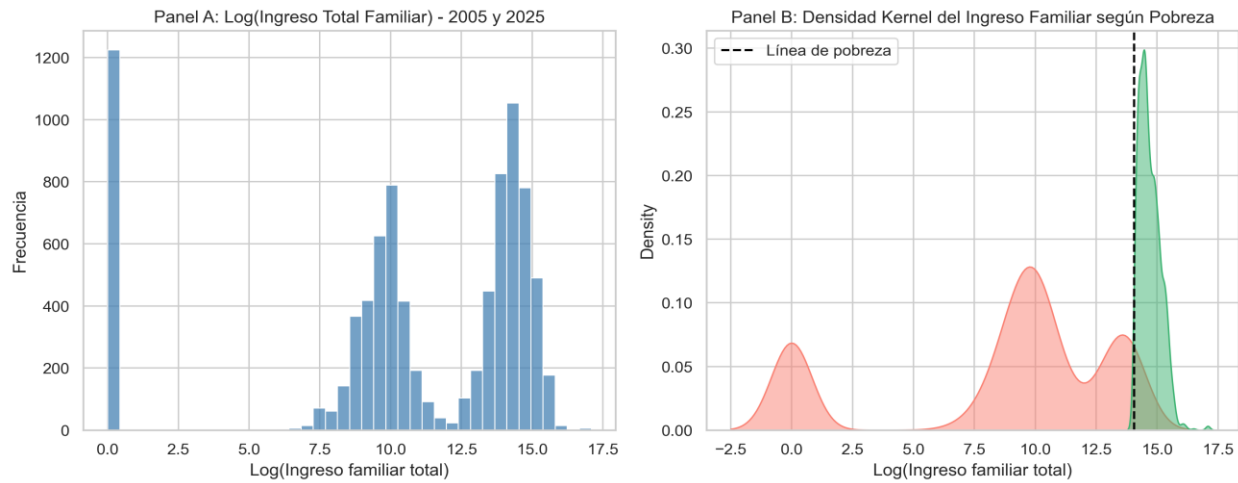
Distribución de la edad y densidad según condición de pobreza (2005–2025)



La distribución de la edad presenta una concentración mayor en los grupos jóvenes y adultos hasta los 40 años, con una disminución progresiva en edades avanzadas. La densidad por condición de pobreza muestra que las personas en situación de pobreza tienden a centrarse en edades menores, mientras que los no pobres presentan una distribución más amplia. Esta diferencia sugiere contrastes demográficos vinculados a la inserción laboral y a la estructura etaria de los hogares.

Figura 2

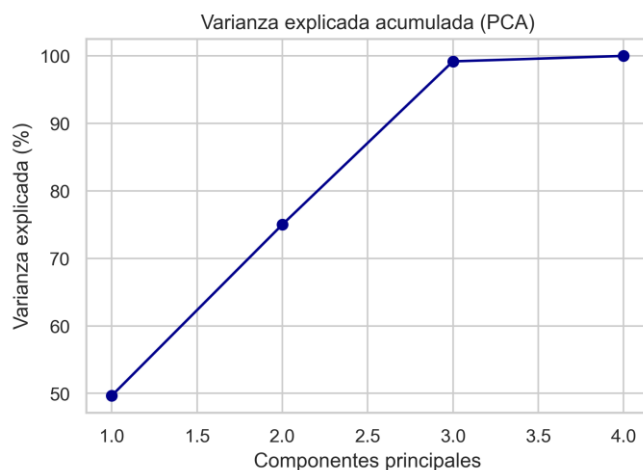
Distribución del ingreso familiar total y densidad según condición de pobreza (2005-2025)



En el Panel A, se observa que la mayoría de los hogares tiene ingresos bajos y que solo una parte pequeña alcanza niveles altos, lo que indica una fuerte desigualdad. En el Panel B, las líneas de densidad muestran que los hogares pobres (en rojo) se concentran por debajo de la línea de pobreza, mientras que los no pobres (en verde) se ubican claramente por encima. En conjunto, el gráfico refleja una diferencia marcada entre ambos grupos y una estructura de ingresos muy desigual.

Figura 3

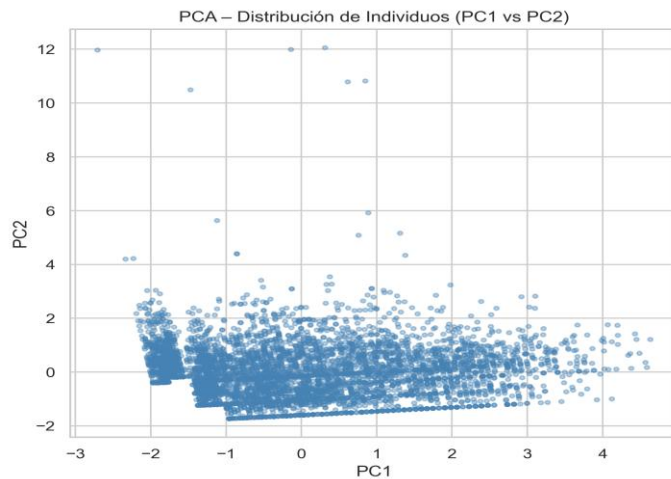
Variación explicada acumulada



La varianza explicada acumulada indica que los tres primeros componentes concentran casi la totalidad de la variabilidad del conjunto de datos, lo que sugiere una estructura bien representada con pocas dimensiones. En particular, los dos primeros componentes ya capturan una proporción sustancial de la información, por lo que el modelo puede simplificarse sin pérdida significativa de contenido analítico.

Figura 4

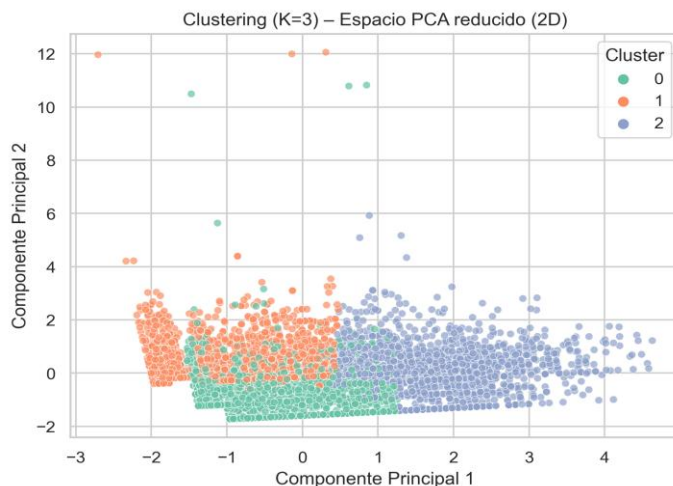
Distribución de los individuos según los dos primeros componentes principales (PC1 y PC2)



La mayor dispersión sobre el primer componente sugiere que las diferencias socioeconómicas entre los individuos de la región patagónica se explican principalmente por una dimensión asociada a variables como el ingreso y la educación. La concentración central indica similitud en gran parte de la población, posiblemente vinculada a estructuras laborales y niveles de ingreso intermedios. La presencia de puntos más alejados refleja heterogeneidad interna, típica de economías regionales con contrastes entre áreas urbanas y rurales o entre sectores productivos de distinta intensidad económica.

Figura 5

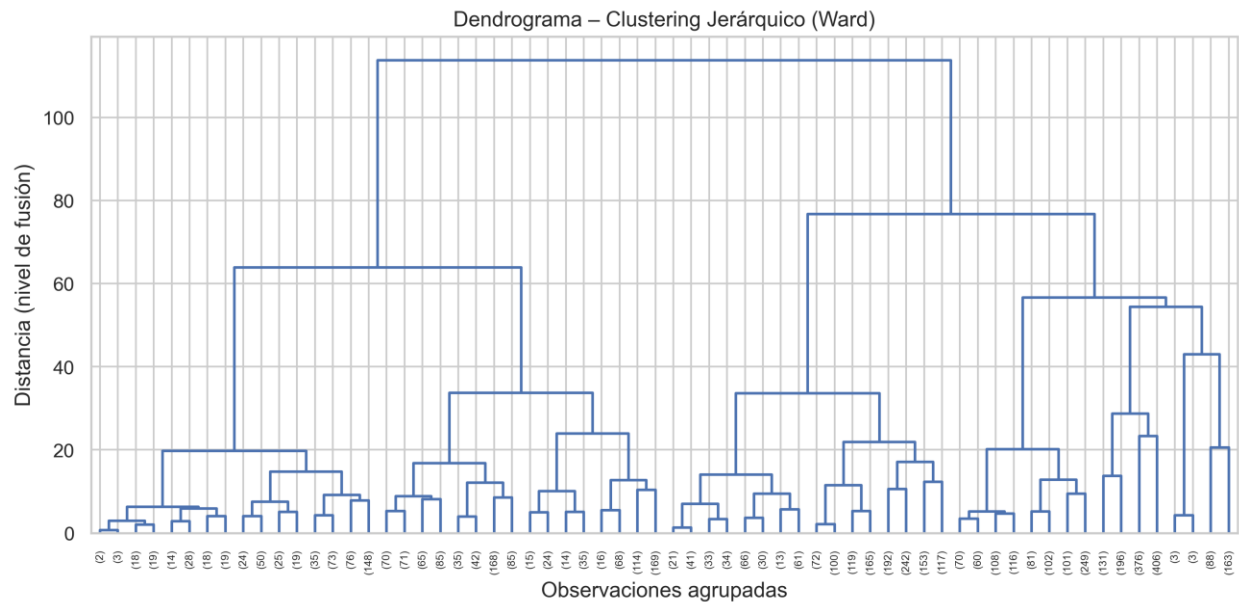
Agrupación de individuos ($K = 3$) en el espacio PCA reducido (2D)



El agrupamiento distingue tres perfiles socioeconómicos: uno de bajos ingresos y educación, otro intermedio y un tercero con mejores condiciones. Esta segmentación resume la desigualdad estructural de la región y la coexistencia de grupos con niveles claramente diferenciados de bienestar.

Figura 6

Dendrograma del agrupamiento jerárquico (Ward)



El dendrograma sugiere una estructura regional heterogénea, con al menos tres grupos socioeconómicos diferenciados por nivel educativo, edad e ingreso. Las distancias de fusión más amplias reflejan brechas marcadas entre segmentos de la población, posiblemente asociadas a desigualdades en el acceso a educación y oportunidades laborales. Esta configuración evidencia que, dentro del espacio patagónico analizado, coexisten perfiles con características socioeconómicas contrastantes que explican parte de la dispersión observada en los indicadores de bienestar.