

**Maestría en Economía Aplicada**

**Materia:**

**Taller de Programación**

**Informe – Trabajo Práctico # 1**

**Primer encuentro con la EPH**

**Grupo # 5:**

**Cammisi, Andrés**

**Porco, Matías**

**Pineda, David**

**Octubre, 2025**

**PARTE I. FAMILIRIANZANDONOS CON LA BASE EPH Y LIMPIEZA**

La primera parte de este Trabajo Práctico consistió en revisar y preparar las bases de microdatos de la Encuesta Permanente de Hogares (EPH) del Instituto Nacional de Estadística y Censos (INDEC) de Argentina correspondientes al primer trimestre de 2005 y 2025.

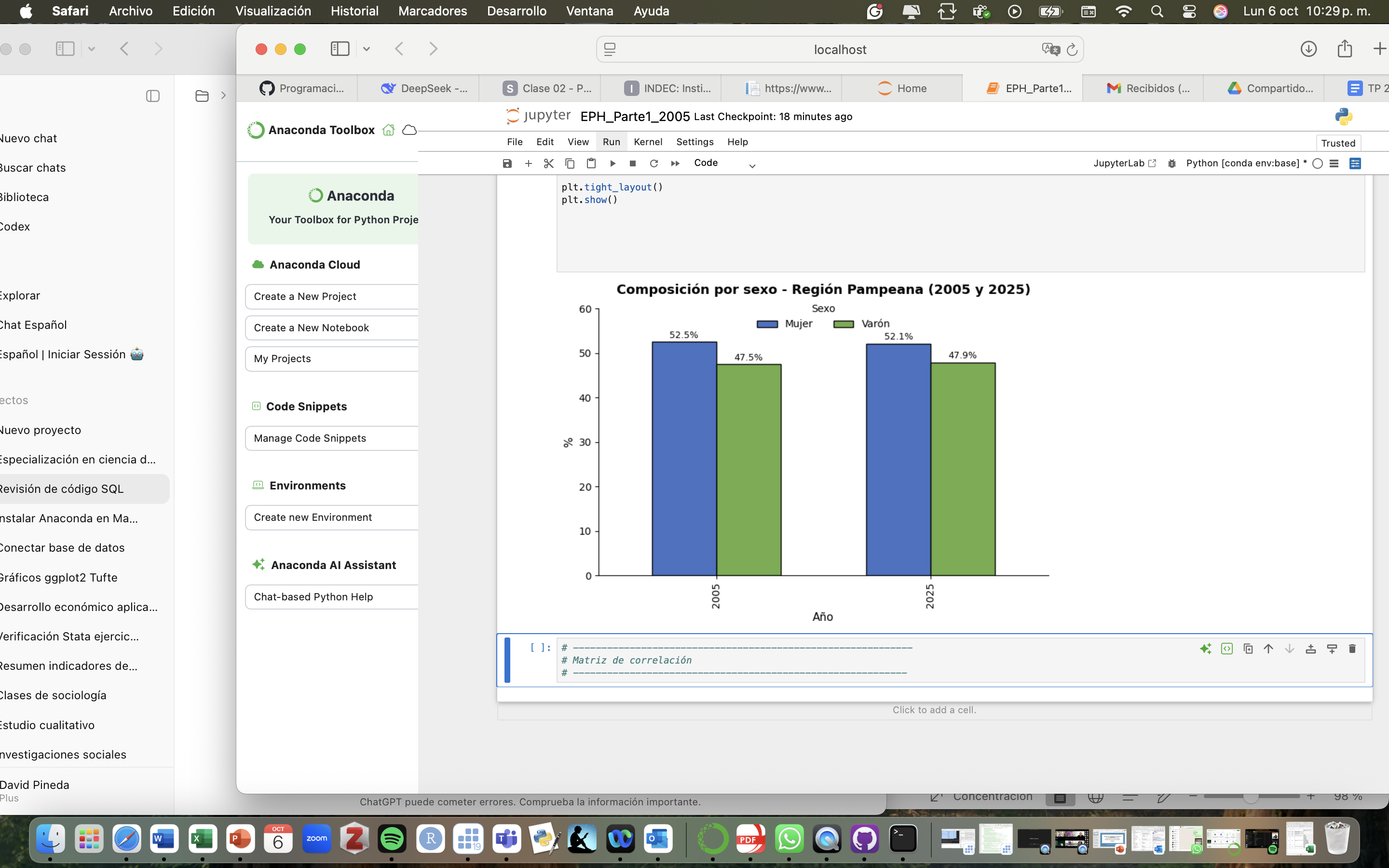
Las bases fueron limitadas a la Región Pampeana (43) y contemplan 15 variables de interés comunes en ambos años: 1) codusu, 2) número de hogar (nro\_hogar), 3) sexo (ch04), 4) años cumplidos (ch06), 5) estado civil (ch07), 6) cobertura médica (ch08), 7) nivel educativo (nivel\_ed), 8) condición de actividad (estado), 9) categoría de inactividad (cat\_inac), 10) ingreso per cápita familiar (ipcf), 11) categoría ocupacional (cat\_ocup), 12) cantidad de ocupaciones (pp03d), 13) total de horas que trabajó en la semana en la ocupación principal (pp3e\_tot), 14) total de horas que trabajó en la semana en otras ocupaciones (pp3f\_\_tot), 15) tipo de empresa o negocio (pp04a), 16) tamaño del establecimiento (pp04c), 17) tiempo buscando trabajo (pp10a), y finalmente, los factores de expansión pondera y pondih.

Se procedió a un diagnóstico inicial de valores faltantes por medio de mapas de calor. En 2005, las variables total de horas que trabajó en la semana en la ocupación principal y total de horas que trabajó en la semana en otras ocupaciones presentaron la mayor cantidad de valores faltantes. Por su parte, en el 2025 las variables cantidad de ocupaciones, total de horas que trabajó en la semana en la ocupación principal, total de horas que trabajó en la semana en otras ocupaciones, tipo de empresa o negocio, tamaño del establecimiento y tiempo buscando trabajo, presentaron altos porcentajes de valores faltantes, lo cual podría explicarse porque únicamente contienen información de la población ocupada o asalariada.

La limpieza de las bases de datos se orientó por los criterios establecidos en los diseños de registro y estructura para las bases de microdatos de la EPH. Los valores negativos de la variable edad (ch06) se convirtiero a cero (0). En la variable horas trabajadas a la semana se convirtieron a cero (0) los valores negativos y aquellos superiores a 120 se codificaron como NaN. Para las variables horas trabajadas en la ocupación principal y en otras ocupaciones se recodificaron los valores 99 y 999 como NaN. Los valores negativos del ingreso per cápita familiar se establecieron en NaN. Este procedimiento se realizó en ambas bases. Finalmente, se consolidó una base (eph\_panel\_05\_25.csv), la que contenía 28,454 observaciones: en el 2005 14,651 y en 2025 13,803.

**PARTE II. PRIMER ANÁLISIS EXPLORATORIO**

Para el análisis exploratorio se realizó una comparación de la composición por sexo en la región Pampeana:

****

En ambos años las mujeres representan una ligera mayoría: en 2005 eran el 52.5% y en 2025 el 52.1%. La variación en la composición por sexo es mínima, por tanto, no se observan cambios demográficos significativos en las dos décadas analizadas.

Matriz de correlación

PARTE III. CONOCIENDO A LOS POBRES Y NO POBRES