

**Maestría en Economía Aplicada**

**Materia:**

**Taller de Programación**

**Docente:**

**Noelia Romero**

**Informe – Trabajo Práctico # 2**

**Histogramas, Kernels & Métodos No Supervisados usando la EPH**

**Grupo # 5:**

**Cammisi, Andrés**

**Porco, Matías**

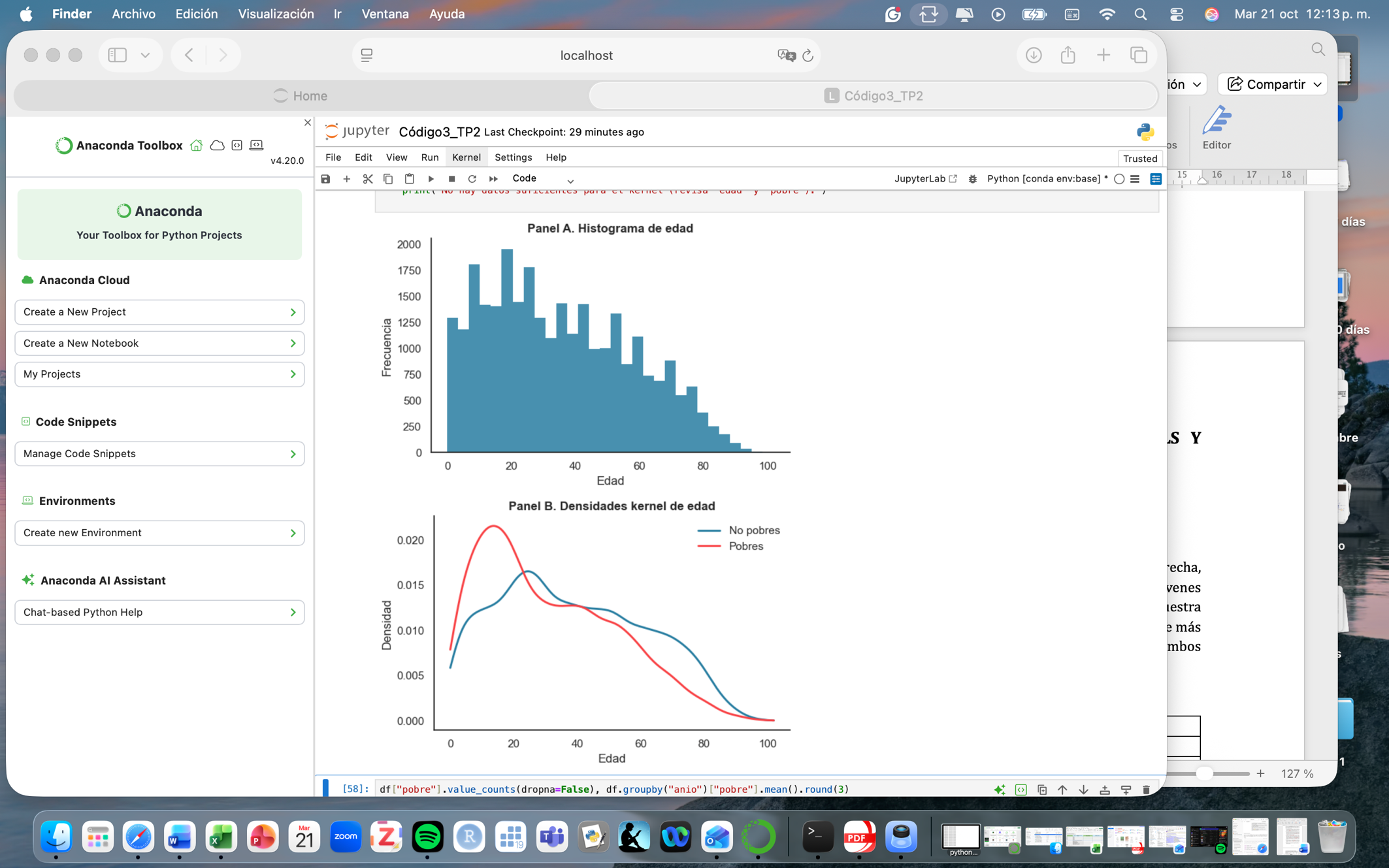
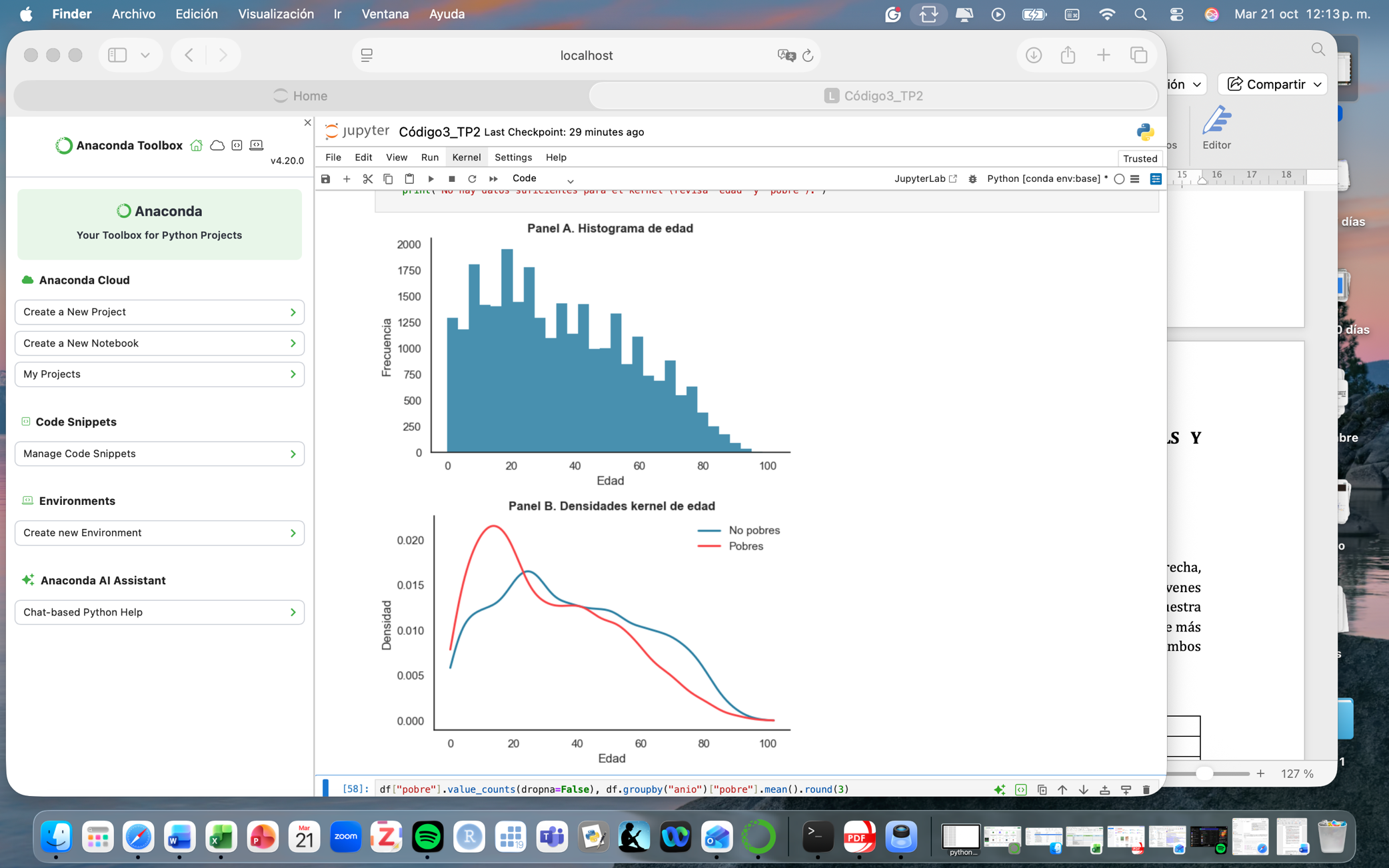
**Pineda, David**

**Link GitHub:** [**https://github.com/DavidPT0902/Grupo\_5\_UBA\_2025.git**](https://github.com/DavidPT0902/Grupo_5_UBA_2025.git)

**Octubre, 2025**

**PARTE I. CREACIÓN DE VARIABLES, HISTOGRAMAS, KERNELS Y RESUMEN DE LA BASE DE DATOS FINAL**

**Ejercicio 1**

****

El panel A muestra que la distribución de la edad es asimétrica a la derecha, confirmando una mayor concentración de personas en edades jóvenes y adultos jóvenes y una cola que tiende a reducirse desde los 60 años. Por su parte, el Panel B muestra que la curva de pobres se mueve hacia edades más jóvenes y la de no pobres tiene más peso relativo entre los 30 y 60 años. En edades altas la densidad se reduce para ambos grupos.

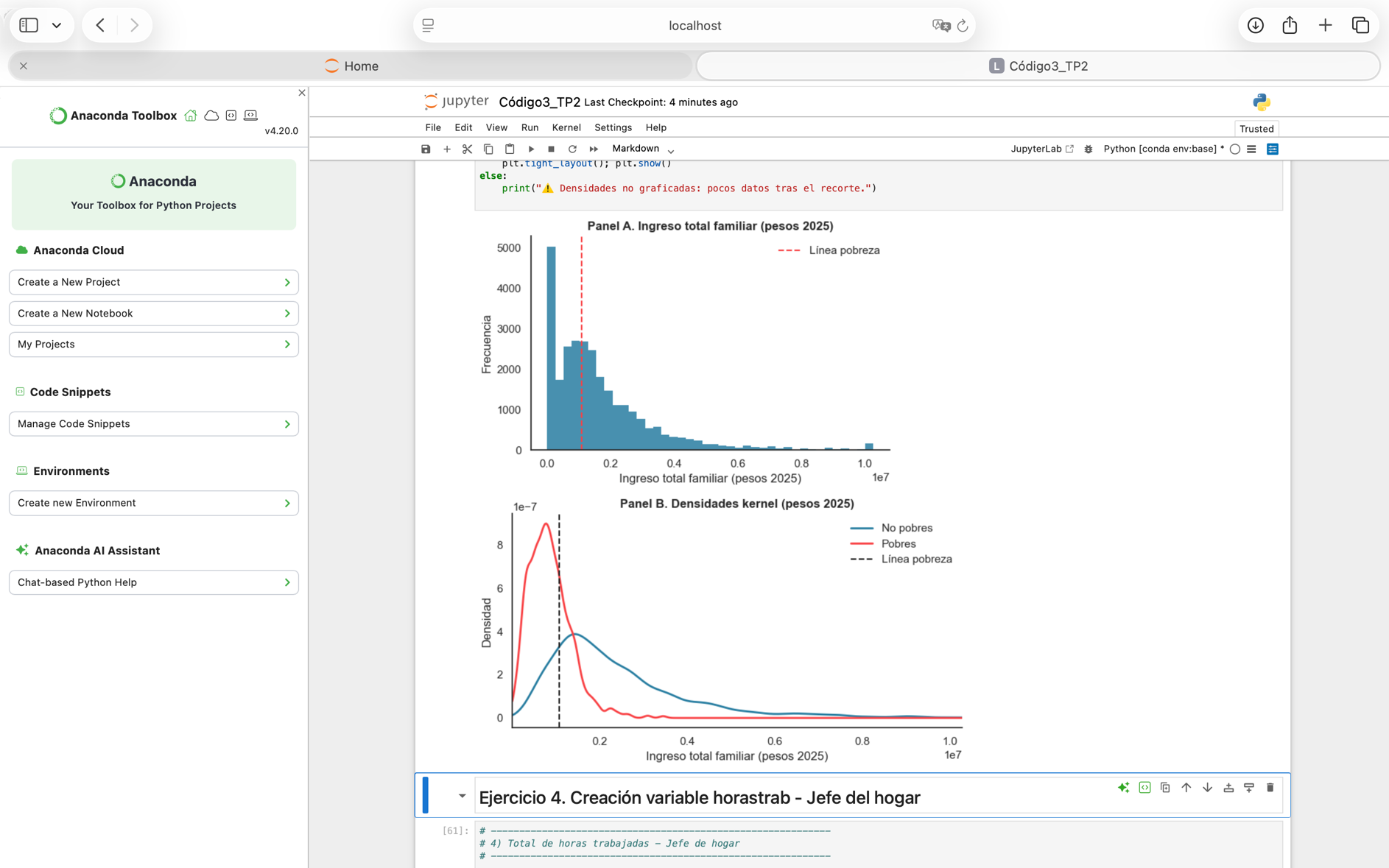
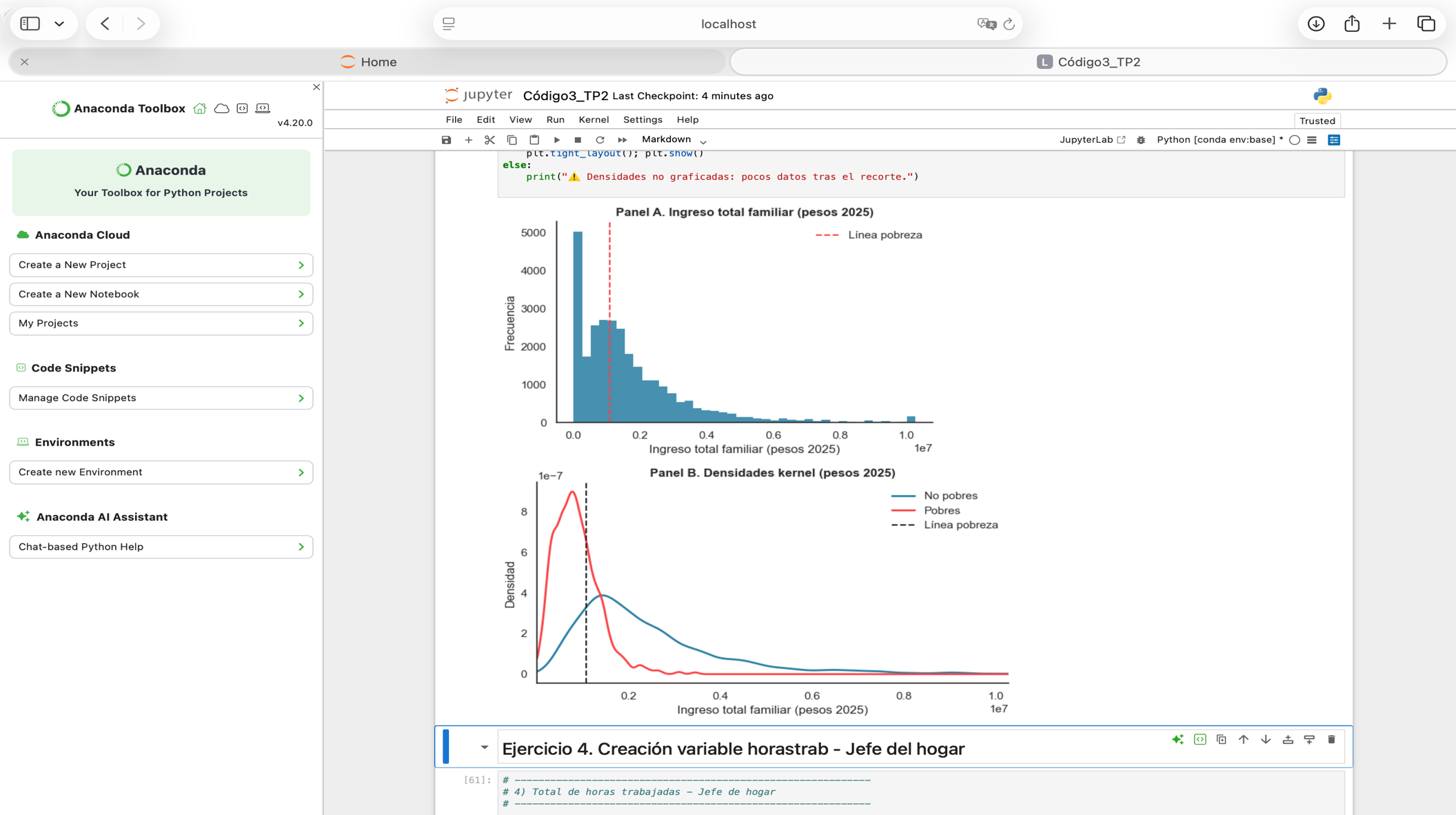
**Ejercicio 2**

**Tabla 1. Región Pampeana: Cantidad de años de educación (2005 y 2025)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Año** | **Mean** | **Std** | **Min** | **P50** | **Max** |
| **2005** | 8.72 | 4.7 | 0.0 | 8.0 | 20.0 |
| **2025** | 9.98 | 4.67 | 0.0 | 12.0 | 20.0 |

Para el 2005, la región Pampeana presentó un promedio de 8.72 años de educación y un ascenso a 9.98 años en 2025, lo que reflejaría mejoras educativas en esta región. En ambos años la desviación estándar es significativa, lo que indica que coexisten personas con reducida o nula educación formal y otras con formación a nivel superior.

**Ejercicio 3**



El histograma del ingreso familiar (Panel A) evidencia una asimetría positiva: una alta concentración de ingresos bajos y una reducción de valores altos. La mayoría de los valores se aproximan a la mediana de ingresos. Por su parte, en el Panel B se muestra que alto porcentaje de pobres se concentran cerca del eje X y la brecha entre pobres y no pobres es amplia, lo que sugiere una estructura de ingresos muy desigual.

**Ejercicio 4**

**Tabla 1. Región Pampeana: estadística descriptiva de horas trabajadas**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Año** | **Mean** | **Std** | **Min** | **P50** | **Max** |
| **2005** | 26.89 | 27.11 | 0.0 | 24.0 | 126.0 |
| **2025** | 21.91 | 23.30 | 0.0 | 18.0 | 140.0 |

Fuente: elaboración propia a partir de EPH (2005) & EPH (2025).

Para el 2005 la horas trabajadas promedio de los jefes de hogar era aproximadamente 27 horas y para el 2025 de 22 horas. Estas cifras pueden estar asociadas con una reducción del empleo formal y, a su vez, un incremento de la inactividad o subocupación laboral.

**Ejercicio 5**

**Tabla 2. Región Pampeana: Resumen de la base unificada**

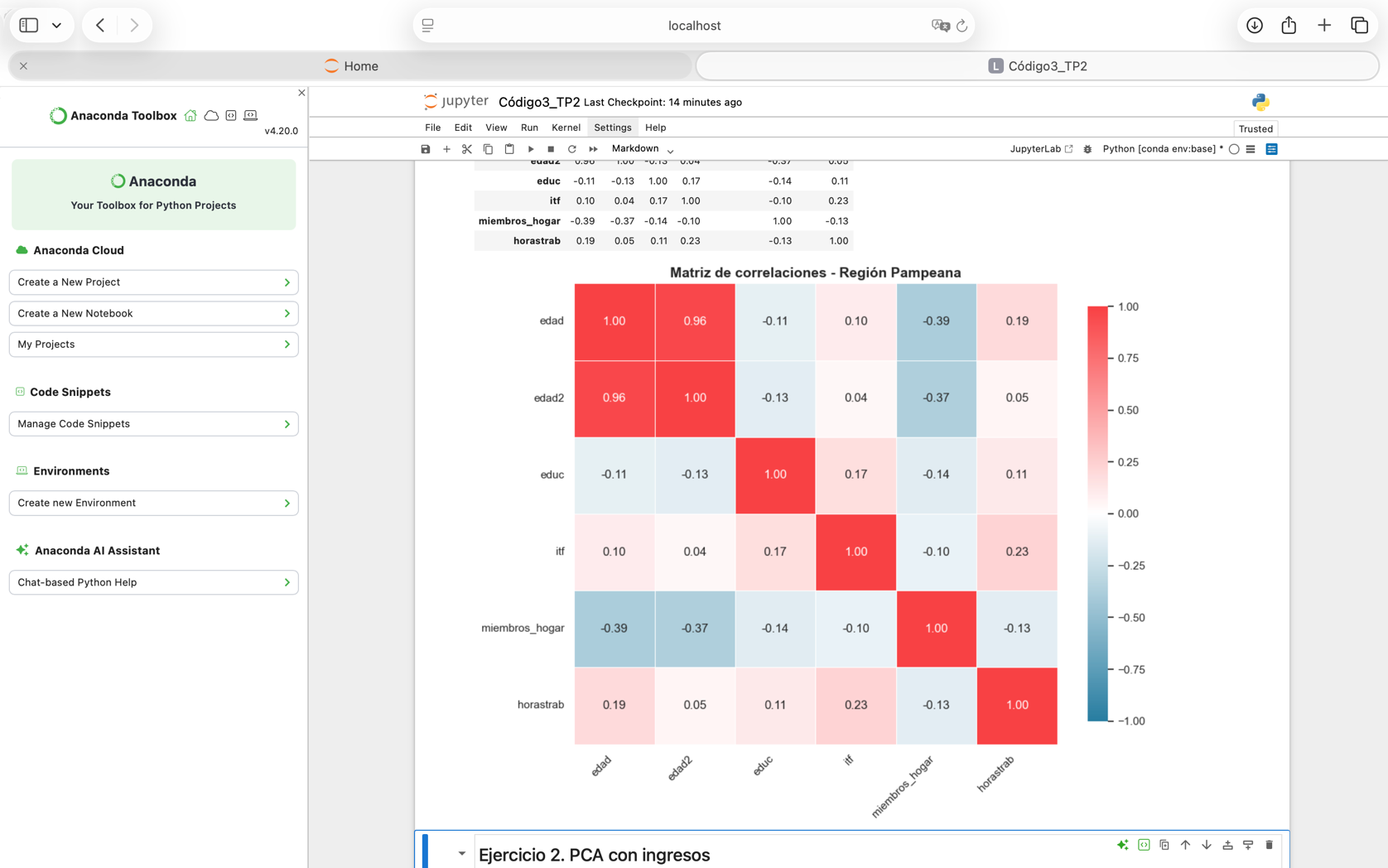
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Año** | **Cantidad de observaciones** | **Observaciones con NA en “pobre”** | **Cantidad de pobres** | **Cantidad de “no pobres”** | **Variables limpias y homogeneizadas** |
| **2005** | 14,651 | 0 | 855 | 13,796 | 35 |
| **2025** | 13,803 | 0 | 5,045 | 8,758 | 35 |
| **Total** | 28,454 | 0 | 5,900 | 22,554 | 35 |

Fuente: elaboración propia a partir de EPH (2005) & EPH (2025).

La base final de la región Pampeana contiene 28,454 observaciones: 14,651 para 2005 y 13,803 para 2025. La variable “pobre” no presentó observaciones faltantes. Los datos muestran un incrementos importante de personas pobres, de 855 a 5,045 en los dos años de análisis, lo que indica un deterioro en las condiciones de vida de la población, tomando como referencia el ingreso familiar equivalente. En total se trabajaron 35 variables en cada año, las cuales fueron recodificadas hasta alcanzar su homogenización.

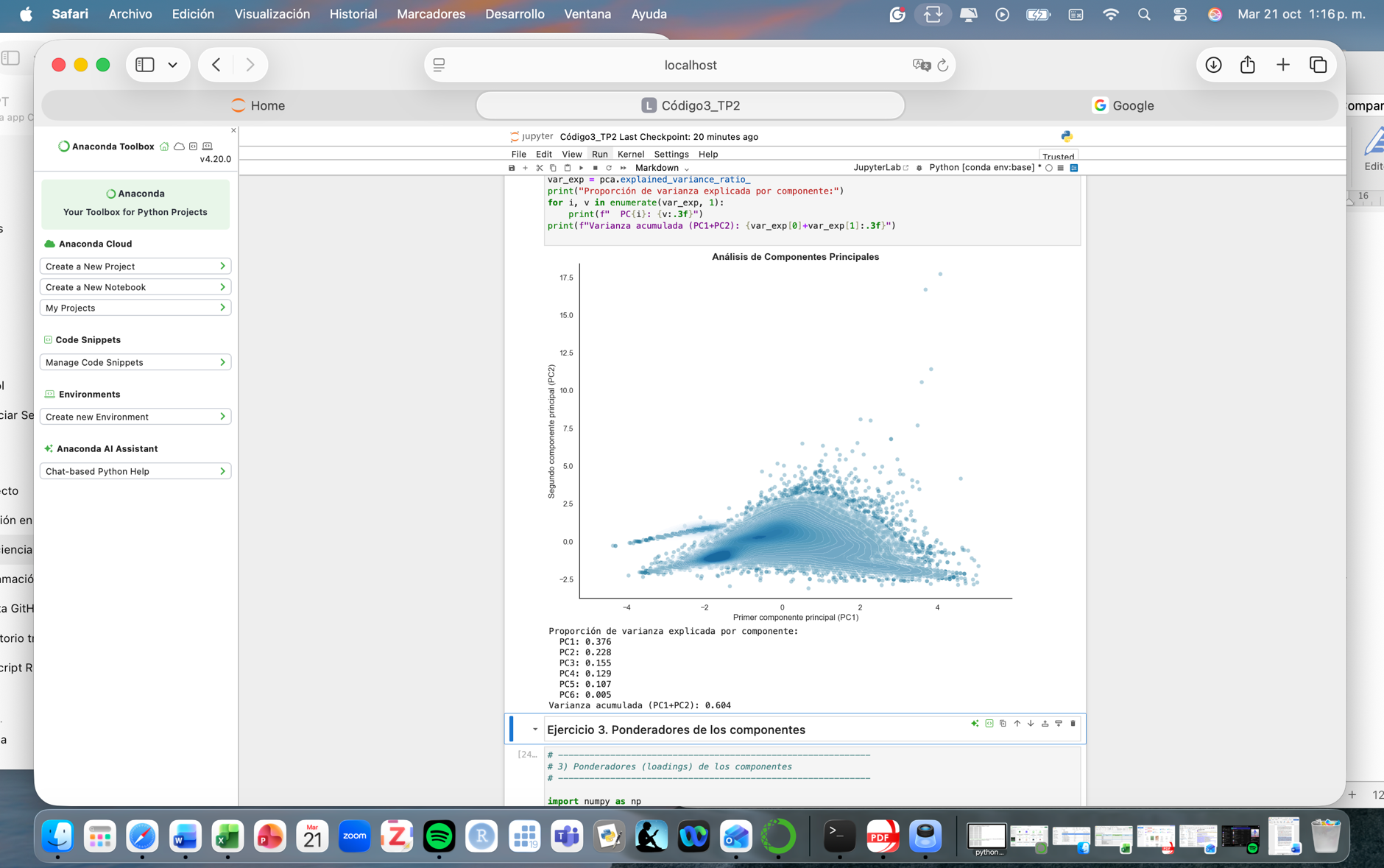
**PARTE II. MÉTODOS NO SUPERVISADOS**

**Ejercicio 1**

****

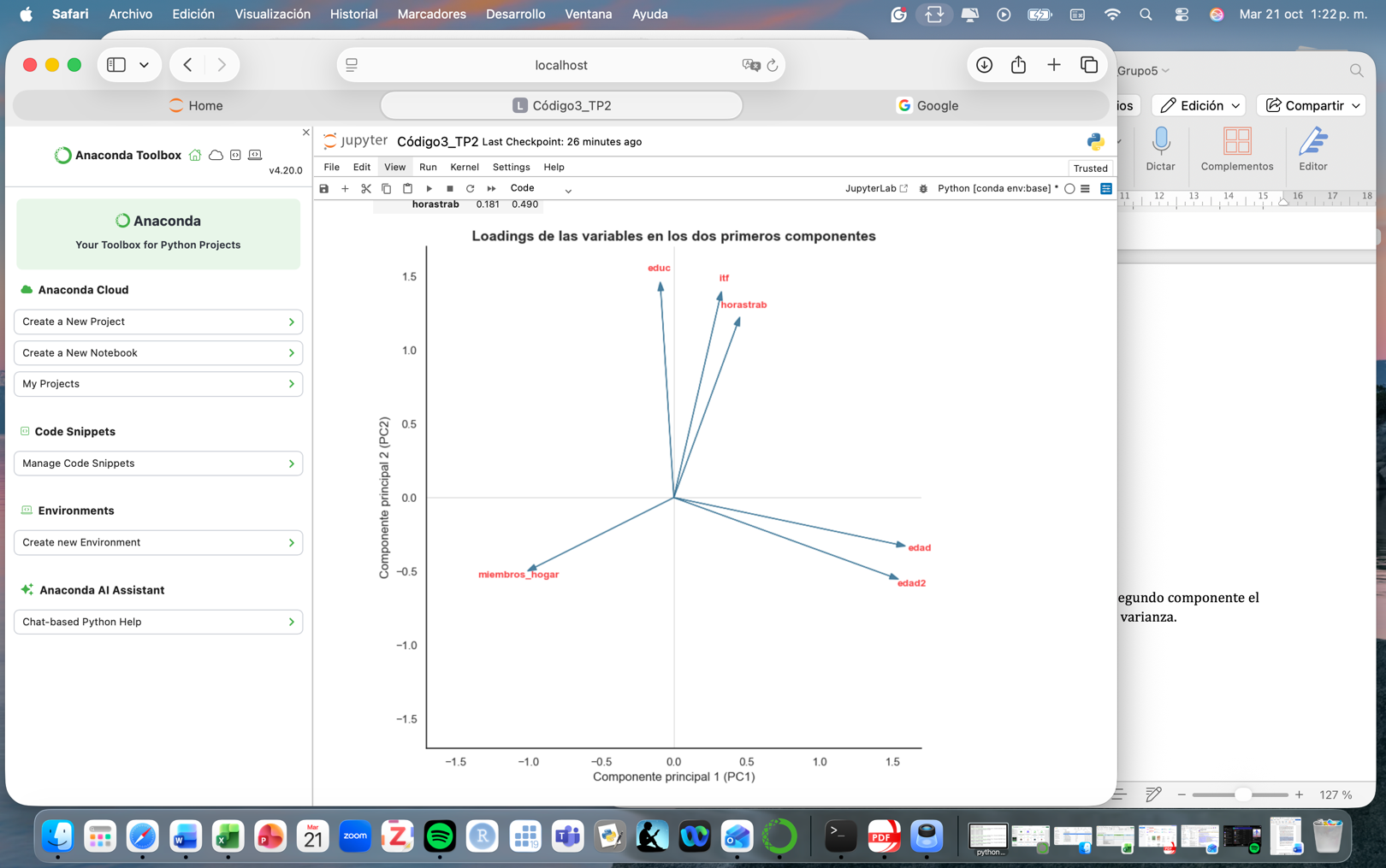
Las correlaciones son bajas o moderadas. Se observa una relación positiva entre horas trabajadas e ingreso familiar y una relación negativa entre miembros del hogar y edad.

**Ejercicio 2**

****

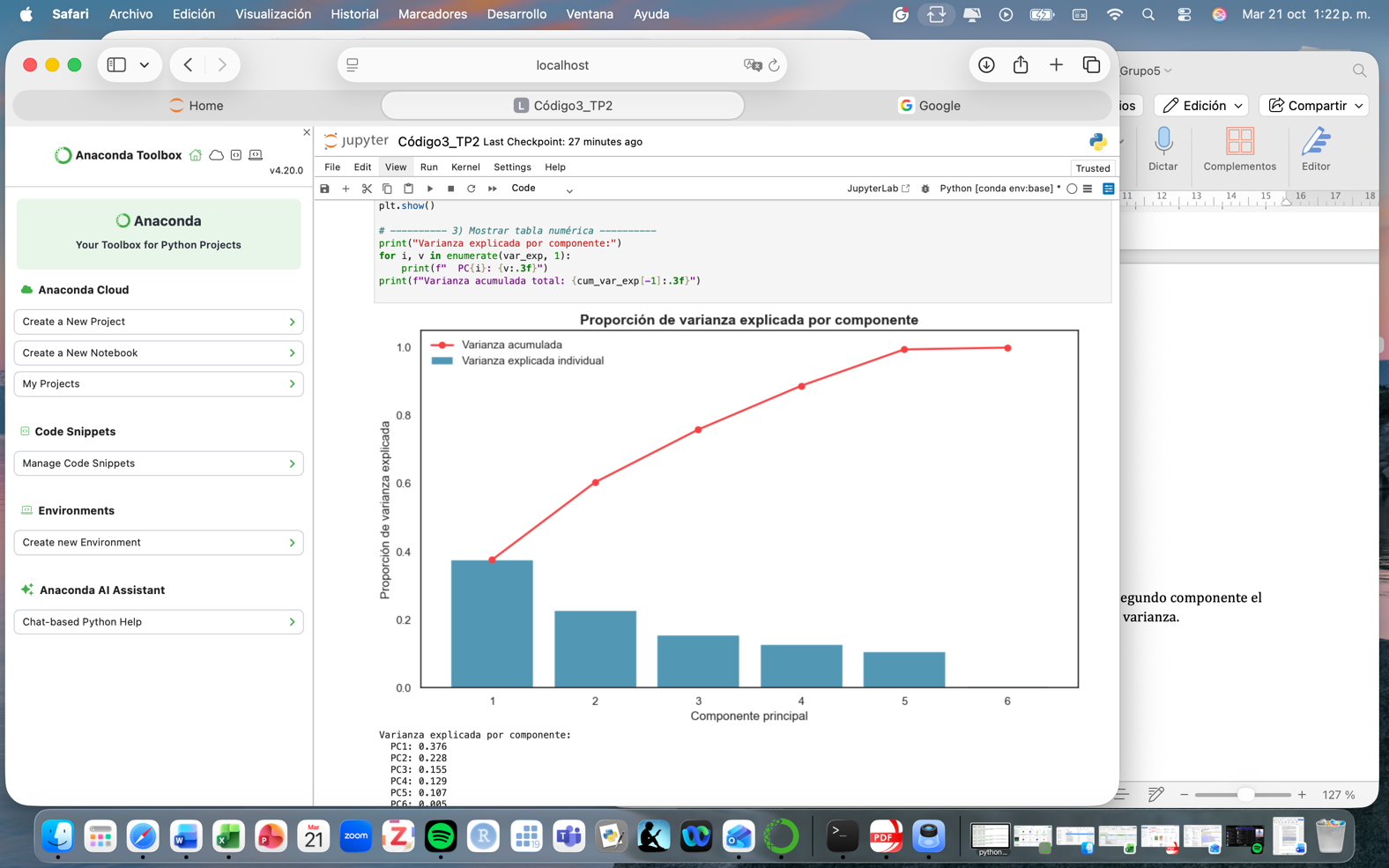
El primer componente explicar el 37.6% de la varianza y el segundo componente el 22.8%, en total, ambos componentes explican el 60.4% de la varianza.

**Ejercicio 3**



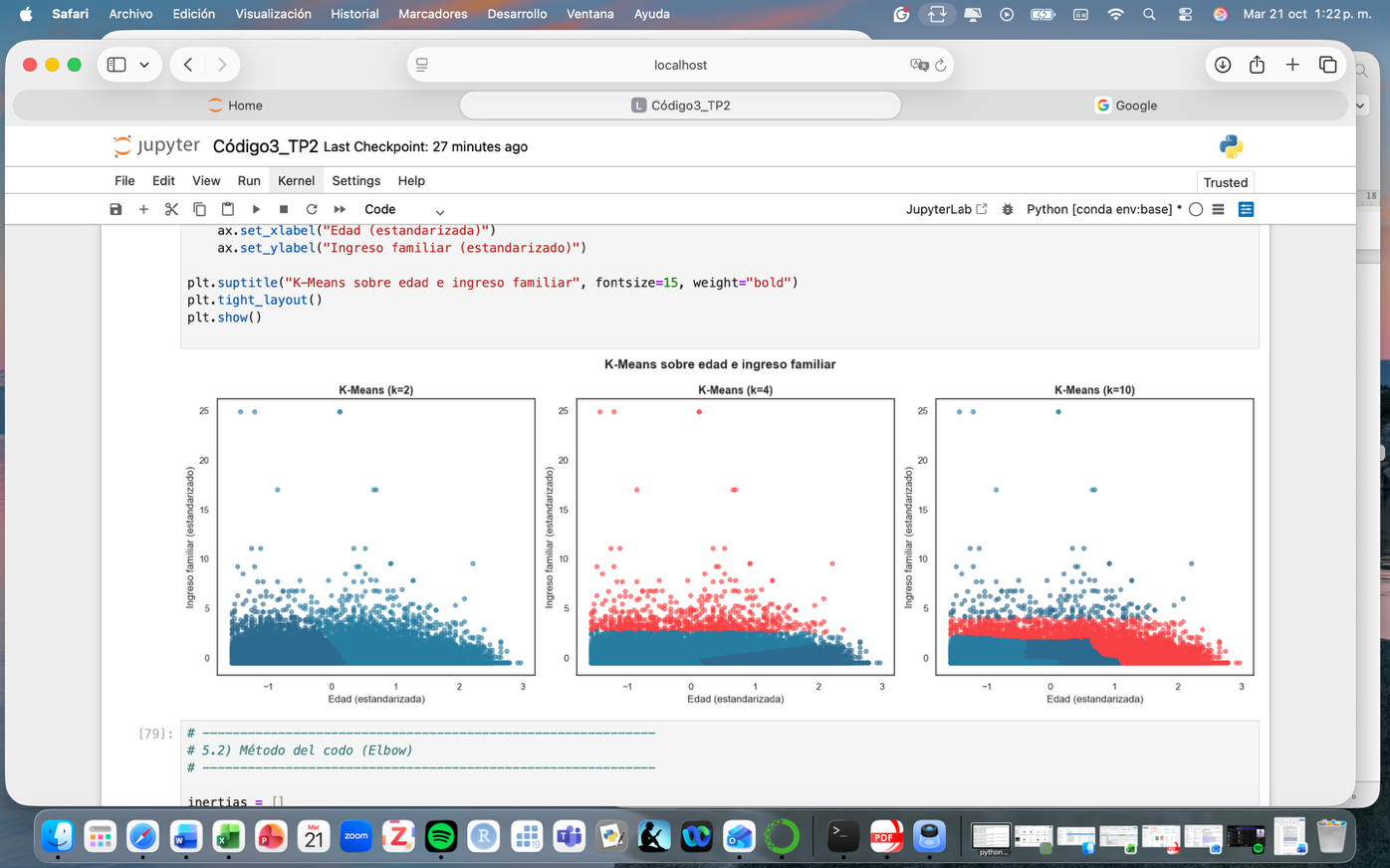
Las variables ingreso total familiar y horas de trabajo están relacionadas con individuos por encima de la línea de la pobreza. En cambio, la cantidad de miembros del hogar se relaciona negativamente con la línea de la pobreza (mayores miembros del hogar = mayor probabildad de ser pobre).

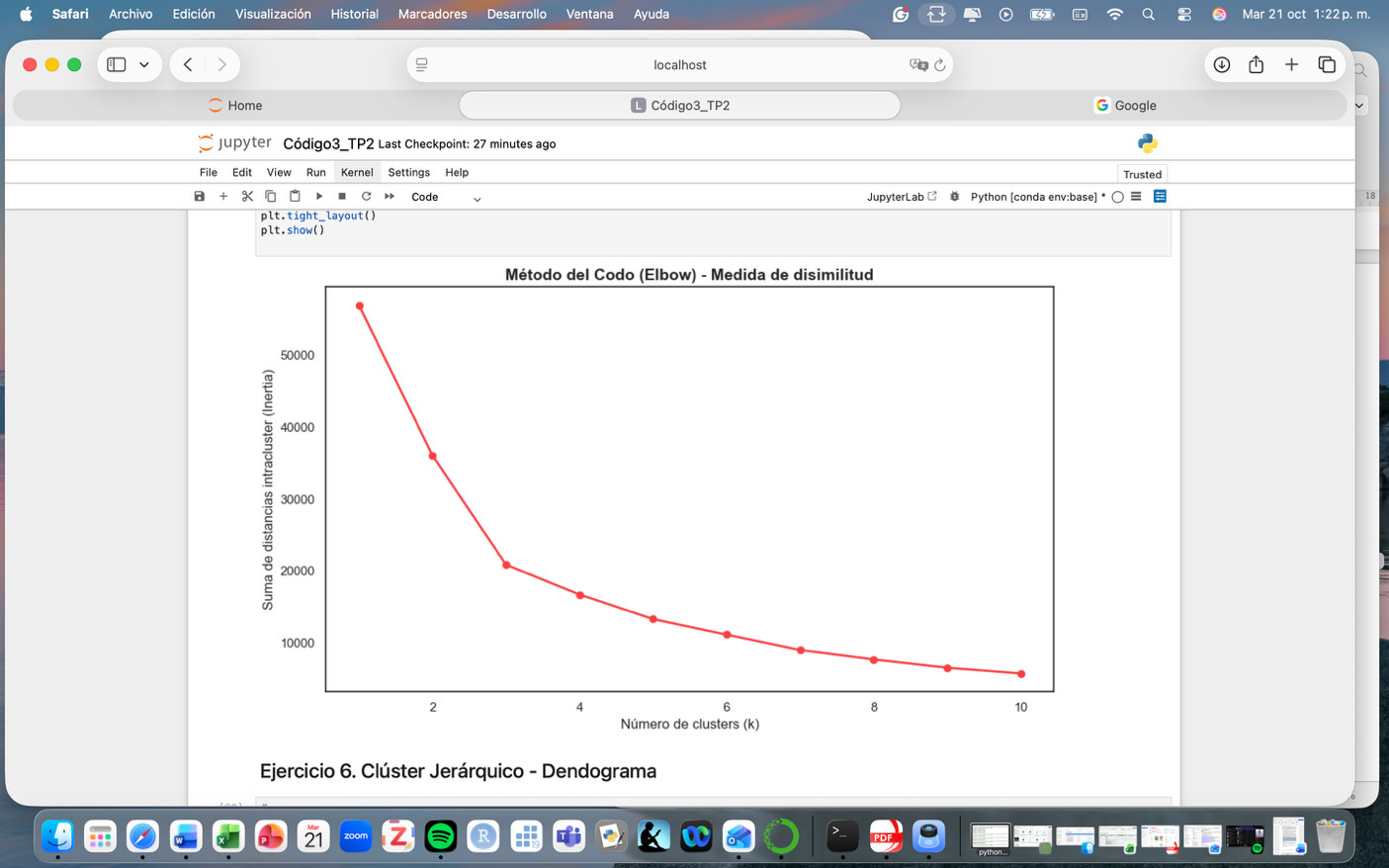
**Ejercicio 4**



El primer componente explicar el 37.6% de la varianza y el segundo componente el 22.8%, en total, ambos componentes explican el 60.4% de la varianza. A partir del tercer componente se observa un aporte mínimo en la varianza explicada.

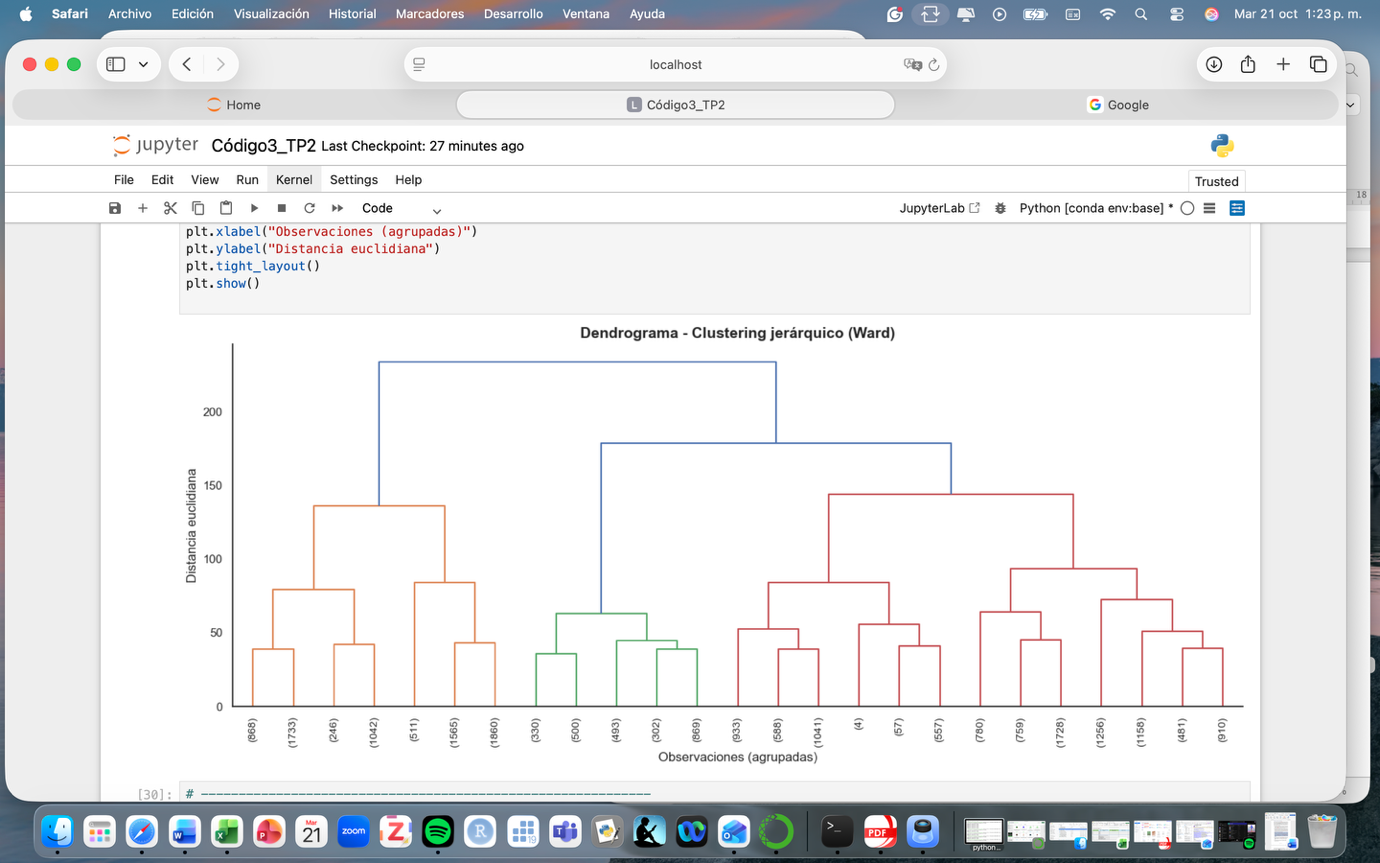
**Ejercicio 5**





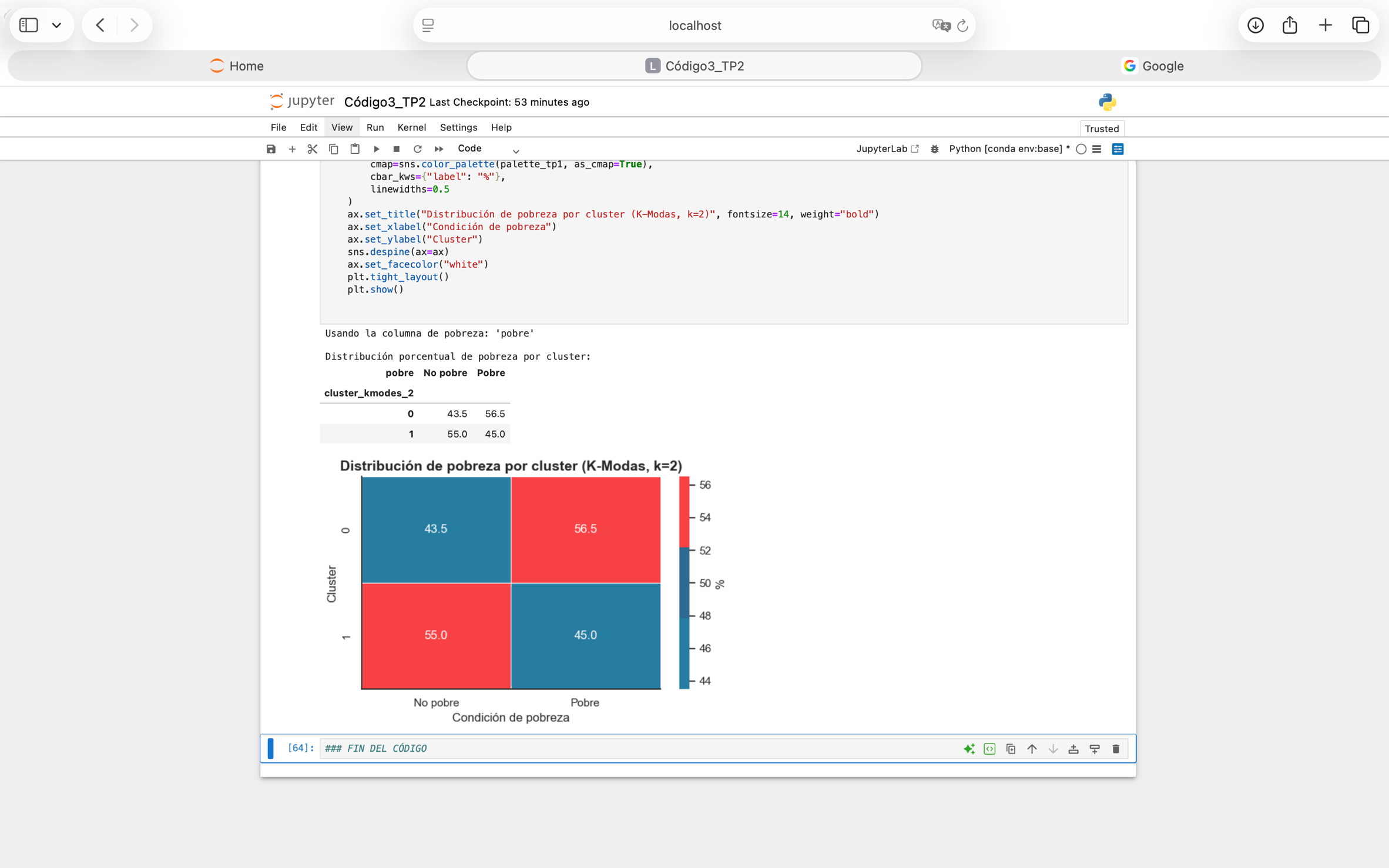
El método del codo indica que la mayor parte de la variabilidad se captura en un máximo de 3 cluster (k=3), a partir de ese punto, el aporte en la explicación de la varianza tiende a ser marginal.

**Ejercicio 6**



El dendograma permite diferenciar entre tres y cuatro conglomerados principales conformados por individuos con perfiles socioeconómicos homogéneas dentro de cada grupo.

**Ejercicio 7**

****

Con k=2, los cluster presentan proporciones internas de 56.5% y 45.0% de pobreza, muy cercanas a la distribución total de la muestra (51.5% pobres y 48.5% de no pobres). Esto indicaría que el modelo no logra discriminar con claridad a la población por debajo y por encima de la línea de pobreza.