

Big Data y Machine Learning (UBA) - 2025

Trabajo Práctico 2: Un primer encuentro con la EPH

Grupo: 10

Integrantes: Bautista Benetti, Luca D'adderio y Tongkun Weng

Según la información disponible en la página del INDEC, las personas desocupadas son aquellas que no tienen trabajo, están buscando activamente trabajo, y están disponibles para comenzar a trabajar.

	CODUSU	ANO4	TRIMESTRE	NRO_HOGAR	COMPONENTE	H15	REGION	MAS_500	A
0	TQRMNOPUYHMOKPCDEGOIH00800329	2024.0	1	1.0	1	1	40	N	
1	TQRMNOPUYHMOKPCDEGOIH00800329	2024.0	1	1.0	2	1	40	N	
2	TQRMNOPUYHMOKPCDEGOIH00800329	2024.0	1	1.0	3	1	40	N	
3	TQRMNOPUYHMOKPCDEGOIH00800329	2024.0	1	1.0	4	1	40	N	
4	TQRMNOTSQHJKLMCDEGOIH00795856	2024.0	1	1.0	1	1	40	N	
...
19087	288280	2004.0	1er. Trimestre	1.0	3.0	0.0	40	S	
19088	288280	2004.0	1er. Trimestre	1.0	4.0	0.0	40	S	
19089	288280	2004.0	1er. Trimestre	1.0	5.0	0.0	40	S	
19090	288337	2004.0	1er. Trimestre	1.0	1.0	Sí	40	S	
19091	288337	2004.0	1er. Trimestre	1.0	2.0	Sí	40	S	
19092	rows x 181 columns								

Valores faltantes en df_2024:

```
ANO4          0
REGION        0
PONDERA       0
CH03          0
CH04          0
CH08          0
CH06          0
CH07          0
NIVEL_ED      0
ESTADO        0
CAT_OCUP      5249
CAT_INAC      0
PP04G         5738
PP07C         6777
PP08D1        6577
dtype: int64
```

Valores faltantes en df_2004:

AN04	0
REGION	0
PONDERA	0
CH03	0
CH04	0
CH08	725
CH06	0
CH07	0
NIVEL_ED	8256
ESTADO	8
CAT_OCUP	5624
CAT_INAC	3942
PP04G	9393
PP07C	9393
PP08D1	7464

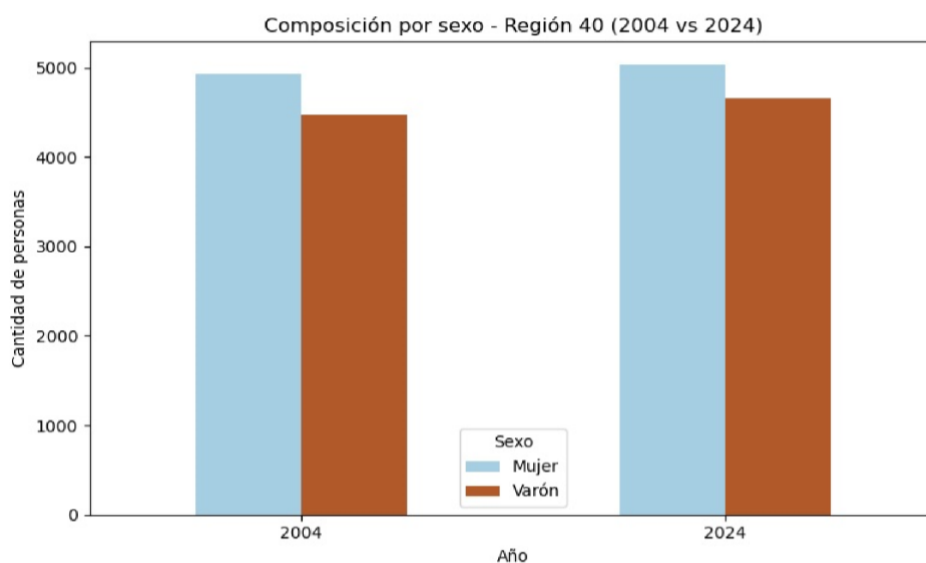
dtype: int64

Los variables CH08, NIVEL_ED, ESTADO, CAT_OCUP, CAT_INAC, PP04G, PP07C, PP08D1 del 2004 y los variables CAT_OCUP, PP04G, PP07C, PP08D1 del 2024 son los que tienen valores faltantes, y los variables PP07C y PP04G del 2004 son los que tiene más valores faltantes.

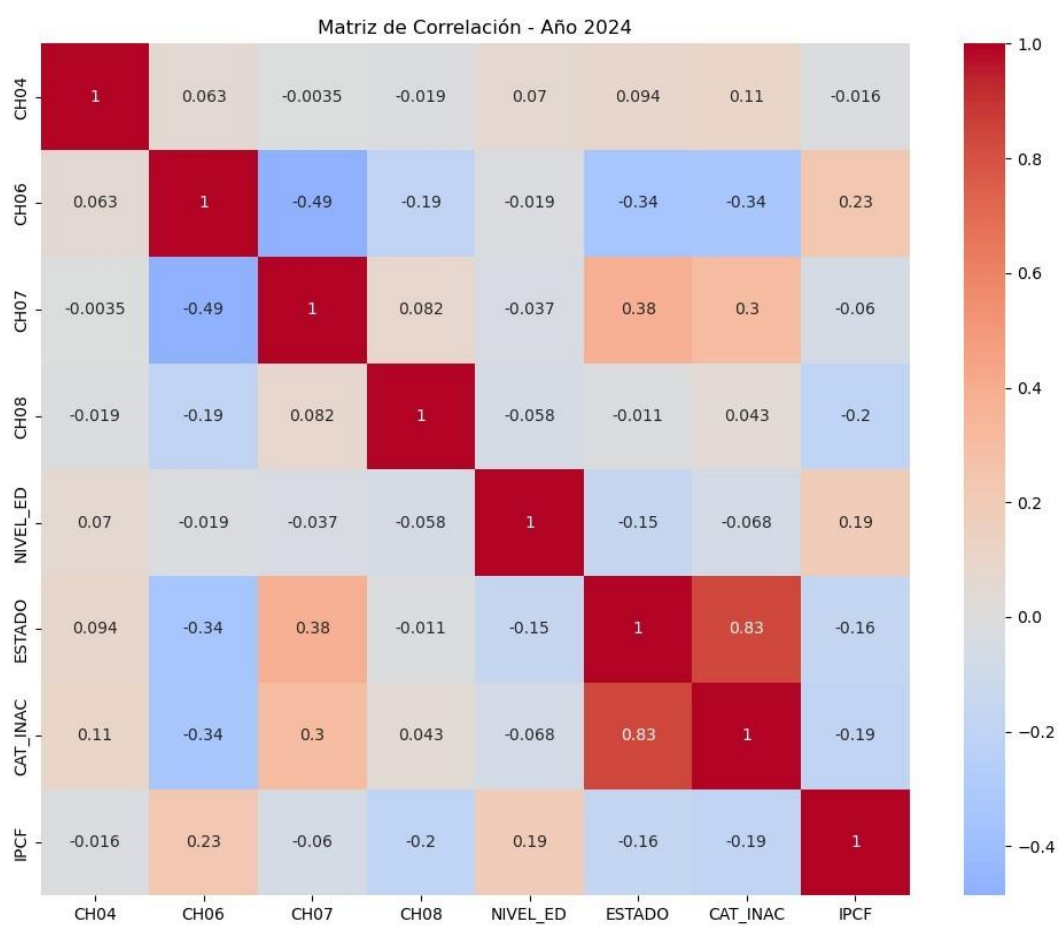
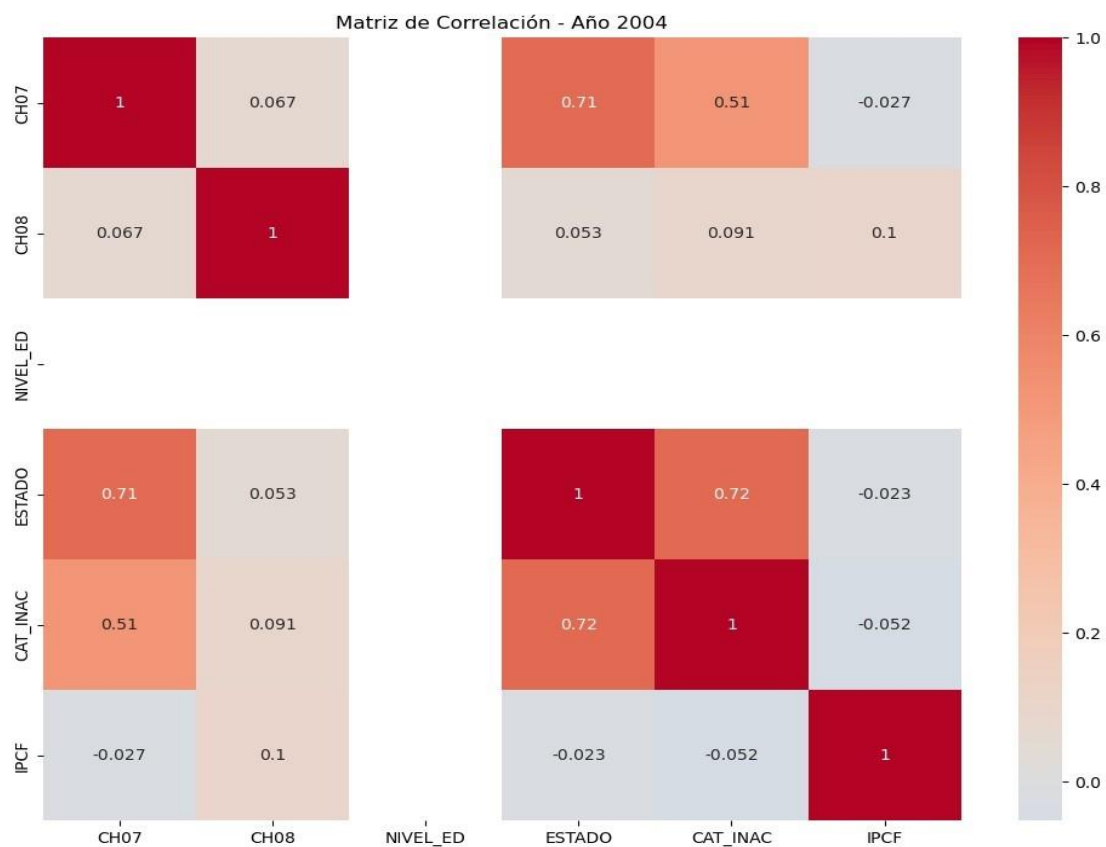
Filas eliminadas por datos sin sentido (2024): 9699

Filas eliminadas por datos sin sentido (2004): 9393

Este proceso de limpieza consiste en filtrar todos los datos que tengan valores razonables eliminando todos los valores no razonables y los valores faltantes.



La cantidad de mujeres en el 2004 son superiores a las de los hombres y en el 2024 también.



En la matriz de correlación del año 2004, en CH07 y ESTADO existe una fuerte correlación positiva. Esto indica que el valor de CH07 está fuertemente asociado al ESTADO. A medida que cambia CH07, también cambia ESTADO en la misma dirección; en ESTADO y CAT_INAC también se observa una fuerte correlación positiva; en CH07 y CAT_INAC es una correlación moderada. Indica cierta relación entre CH07 y CAT_INAC, aunque no tan fuerte como con ESTADO; en CH08 y el resto de variables prácticamente no muestra correlación significativa (todos los valores cerca de 0), lo cual sugiere que no guarda una relación directa con las otras variables en ese año; en IPCF, su relación con todas las demás variables es muy baja. Esto podría implicar que en 2004 el IPCF no estaba directamente vinculado con las otras variables.

En la matriz de Correlación del año 2024, en CH07 y CAT_INAC, la relación se ha fortalecido considerablemente con respecto a 2004, lo que indica que la variable CH07 ahora tiene una asociación mucho más clara con CAT_INAC; en ESTADO y CAT_INAC, también más fuerte que en 2004; en CH06 y CH07, es una correlación negativa importante. A medida que aumenta CH06, disminuye CH07. Esto podría indicar que son variables contrapuestas o que miden fenómenos inversos; en CH06 y los variables ESTADO y CAT_INAC, son correlaciones negativas. Reforzando la idea de que CH06 representa un aspecto opuesto al de estas variables; en NIVEL_ED y ESTADO, es una relación positiva débil pero mayor que en 2004. Puede reflejar una mayor influencia de la educación sobre la situación laboral; en IPCF, aunque sigue mostrando correlaciones bajas, hay un leve incremento respecto a 2004, especialmente con CH08 y NIVEL_ED, lo que sugiere que ahora el IPCF está más relacionado con CH08 y NIVEL_ED.

Cantidad de personas por estado:

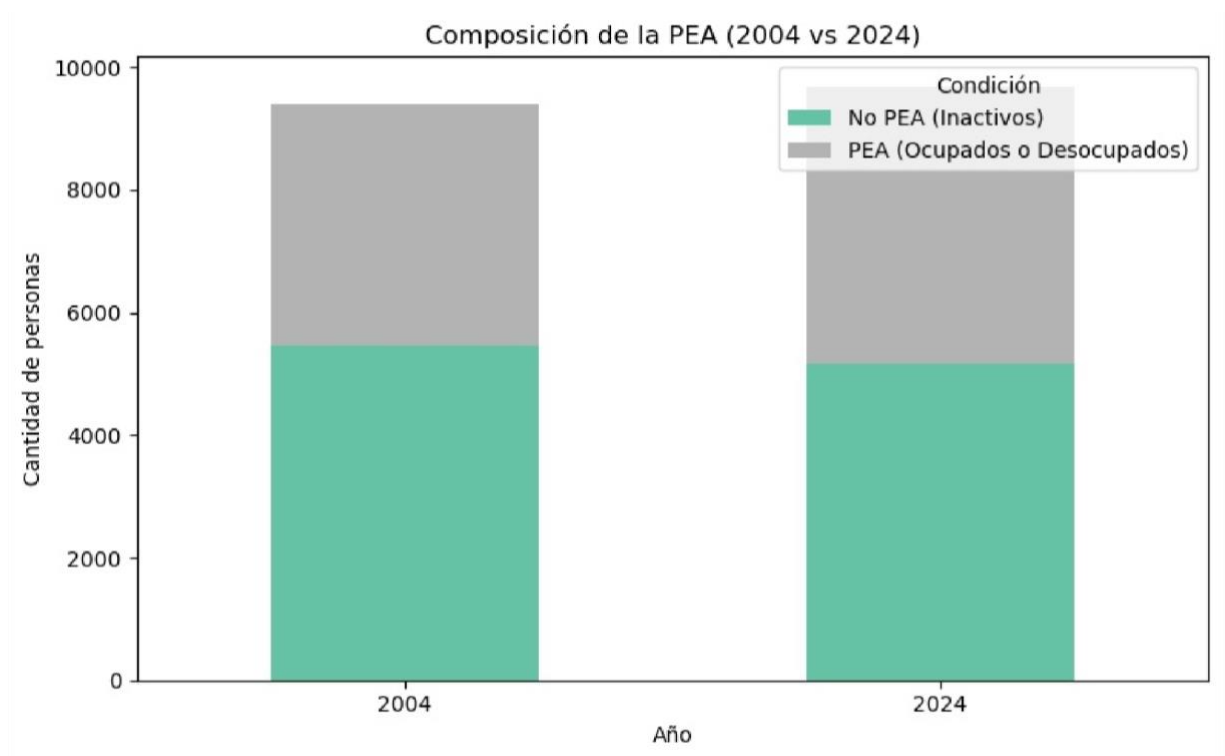
```
ESTADO
Ocupado      7600
Inactivo     7596
Desocupado    853
Name: count, dtype: int64
```

Media de IPCF por estado:

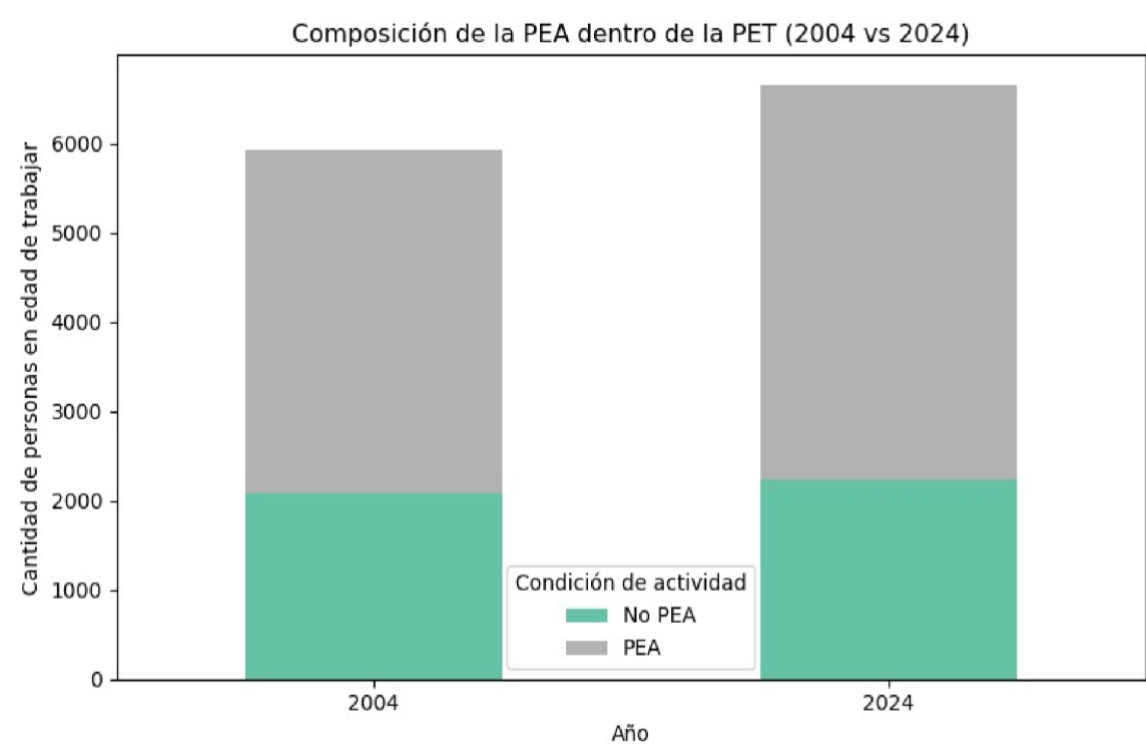
```
ESTADO
Desocupado    33607.05
Inactivo      75589.48
Ocupado       102329.39
Name: IPCF, dtype: float64
```

Hay 853 de desocupados en la muestra, hay 7596 inactivos en la muestra y hay 7600 ocupado. También la media de ingreso per cápita familiar según estado es por desocupado es 33607.05, de inactivo de 75589.48 y ocupado 102329.39.

Cantidad de personas que no respondieron su condición de actividad: 9



En el 2004 la cantidad de inactivos es mayor que el 2024 y su cantidad de PEA también es mayor que el 2024.



La cantidad de personas en edad de trabajar en el 2004 tiene menores inactivos que el 2024, también tiene menores ocupados y desocupados que el 2024. Sin embargo, la cantidad de personas en edad de trabajar en el 2024 es mayor que el 2004.

Cantidad de personas desocupadas por año:

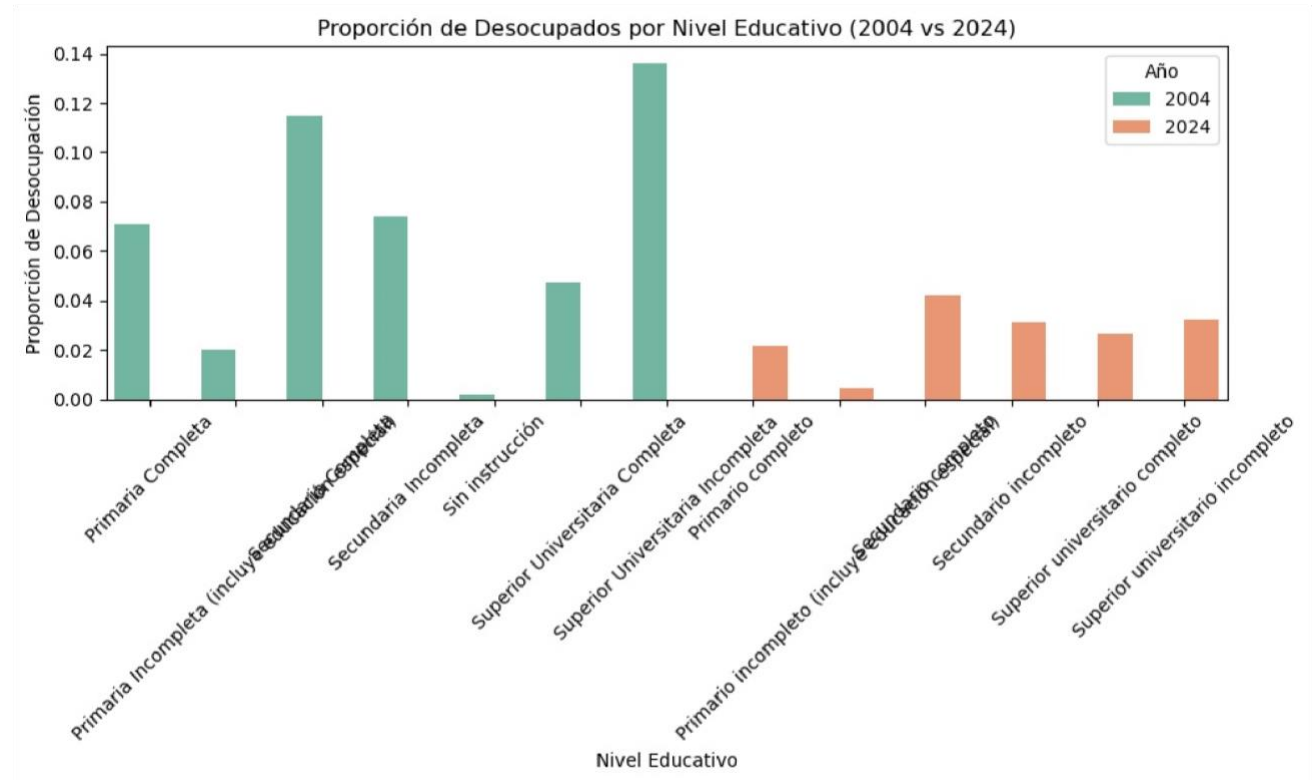
AÑO

2004 597

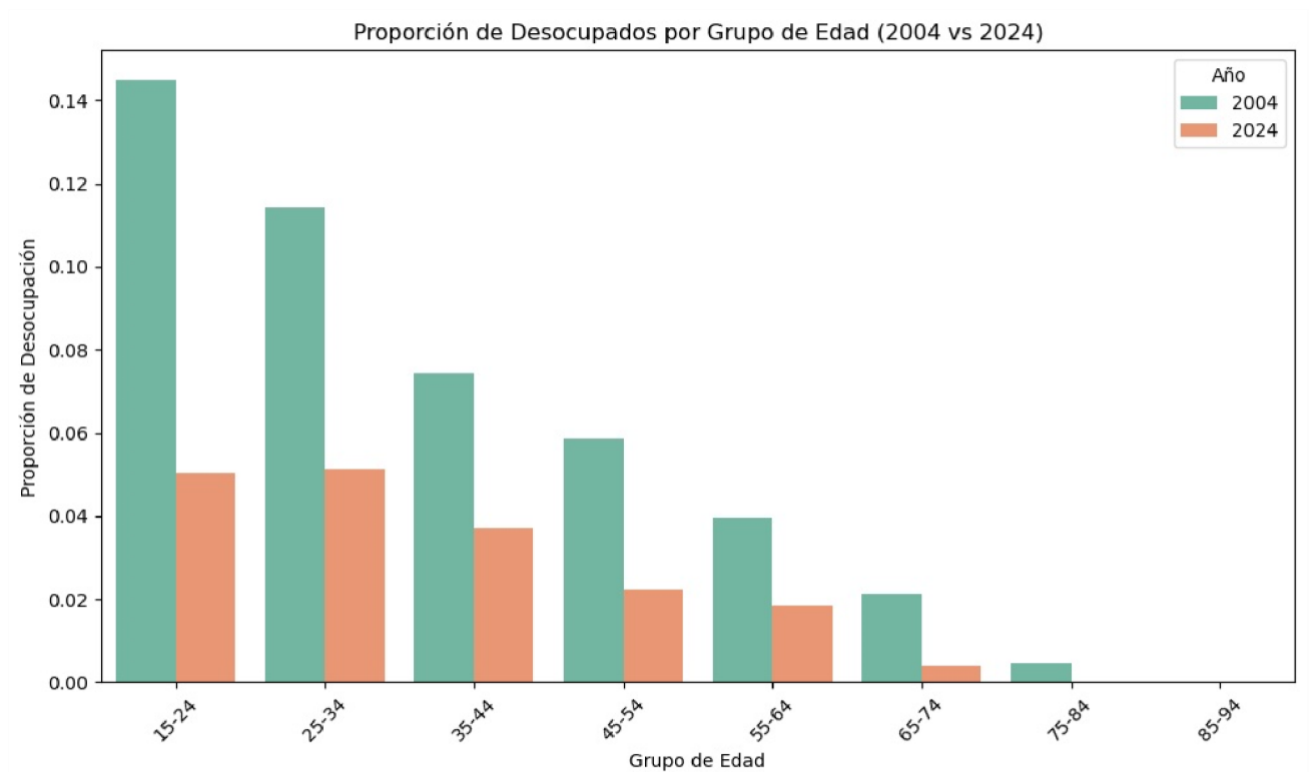
2024 256

Name: desocupado, dtype: int64

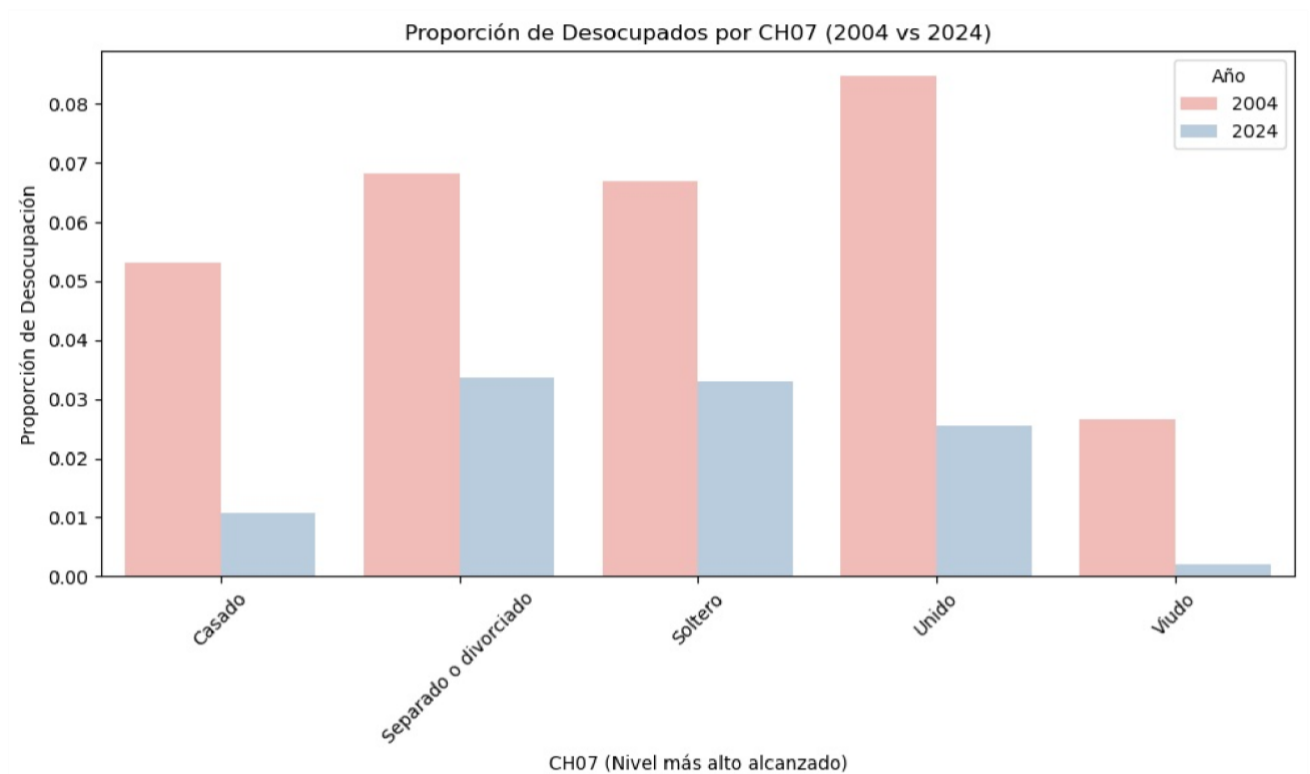
En el 2004 la cantidad de personas desocupadas es 597 y en el 2024 se disminuye a ser 256.



Sí, hubieron cambios de desocupados por nivel educativo.



Sí, hubo cambios de desocupados por edad.



Sí, hubo cambios de desocupados por edad.