

TP 2 de Big Data sobre la EPH

1)

El INDEC encuentra a las personas desocupadas mediante un censo donde se le pregunta a la gente sobre su situación laboral. Para el INDEC una persona esta en estado de desocupación si no realiza ninguna actividad laboral por lo menos una semana anterior a la encuesta y que también este en condiciones de trabajar.

Este censo que controla la desocupación entre muchos otros datos se llama Encuesta Permanente de Hogares EPH. La EPH tiene códigos para definir estas variables, por ejemplo para saber si esta desocupado, con trabajo, inactivo o si es menor de 10 años se usa el código ESTADO.

El código CAT_INAC sirve para identificar aquellos que están inactivos como los jubilados, estudiantes y discapacitados.

Otro código interesante son los diferentes que hay para describir como estuvo buscando esa persona trabajo, por ejemplo PP02C1 si estuvo haciendo entrevistas, PP02C3 si se presento en algún lugar etc.

2)

Para este segundo punto lo que buscamos hacer primero es subir las bases de datos del año 2004 y 2024 a nuestra notebook, luego fusionamos estas base de datos para trabajar mejor.

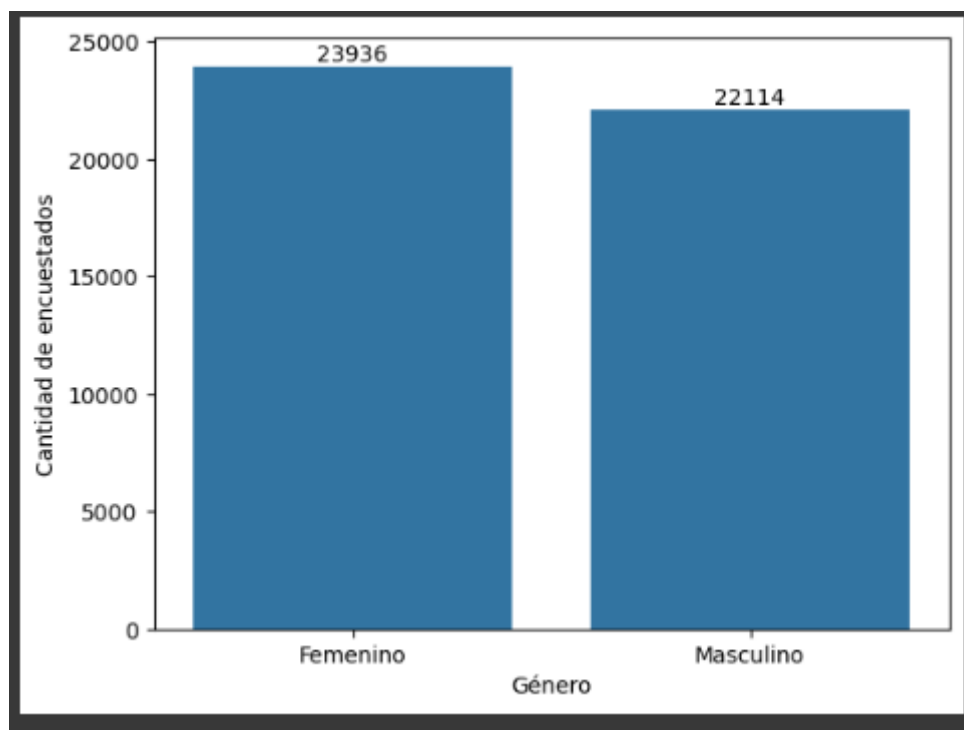
Luego lo que hicimos como indica la consigna fue elegir 15 variables para profundizar.

Después de todo esto llevamos un proceso de limpieza de datos, en particular sacando los nulos y datos no correspondientes.

```
ultimo_df = pd.concat([df_filtrado,dfs],axis=0)
ultimo_df = ultimo_df.replace("NaN", 0)
ultimo_df = ultimo_df.replace([np.nan, -np.inf], 0)
ultimo_df = ultimo_df.apply(pd.to_numeric, errors='coerce').fillna(0)
ultimo_df[ultimo_df <= 0] = 0
ultimo_df
```

3)

Para este ejercicio utilizamos la variables “CH03” que en la EPH define el sexo de los encuestados para hacer un grafico de barras comparativo.

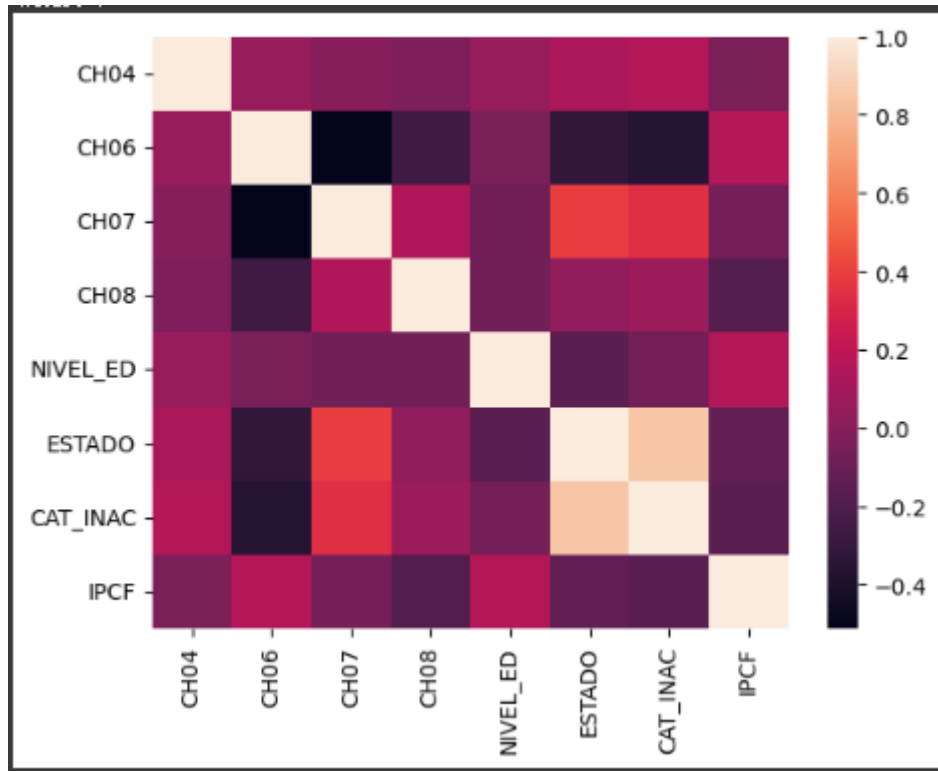


La observación que podemos ver en este grafico es que son mas las encuestadas mujeres que los encuestados hombres en esta región.

4)

En esta parte hicimos un cuadro de correlación usando justamente una matriz de correlación que contiene estas variables: CH04, CH06, CH07, CH08, NIVEL_ED, ESTADO, CAT_INAC, IPCF.

Esto es lo que devolvió:



Se puede ver relaciones interesantes entre el sexo de los encuestados, su nivel educativo y su estado laboral.

5)

Para este punto usamos la variable “ESTADO” de la EPH para calcular los desempleados y después usamos funciones estadísticas para sacar la media del ingreso IPCF.

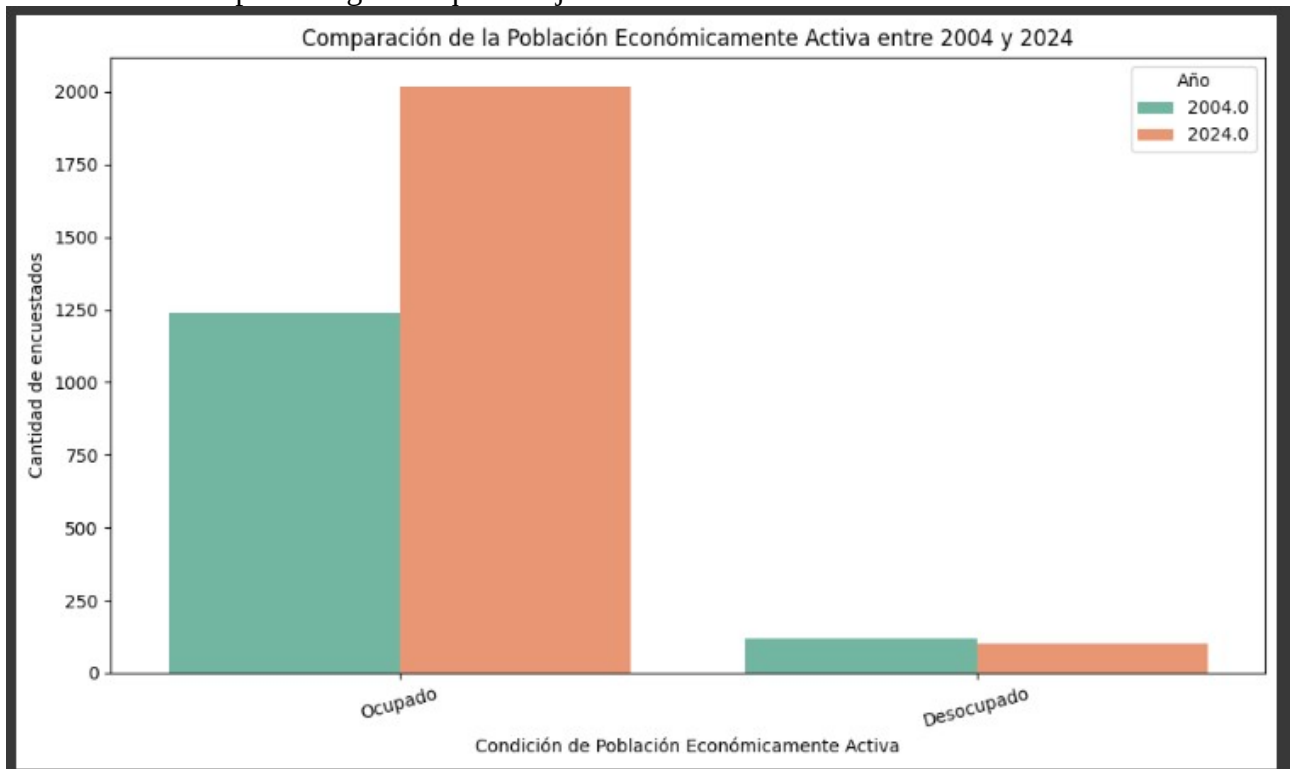
ESTADO	
Ocupado	2025
Desocupado	1821
Inactivo	578
Menor de 10 años	100
Entrevista individual no realizada	6

6)

Acá creamos las dos bases de datos filtrado por si respondieron a las encuestas o no, para eso utilizamos la variable “ESTADO” de la EPH.

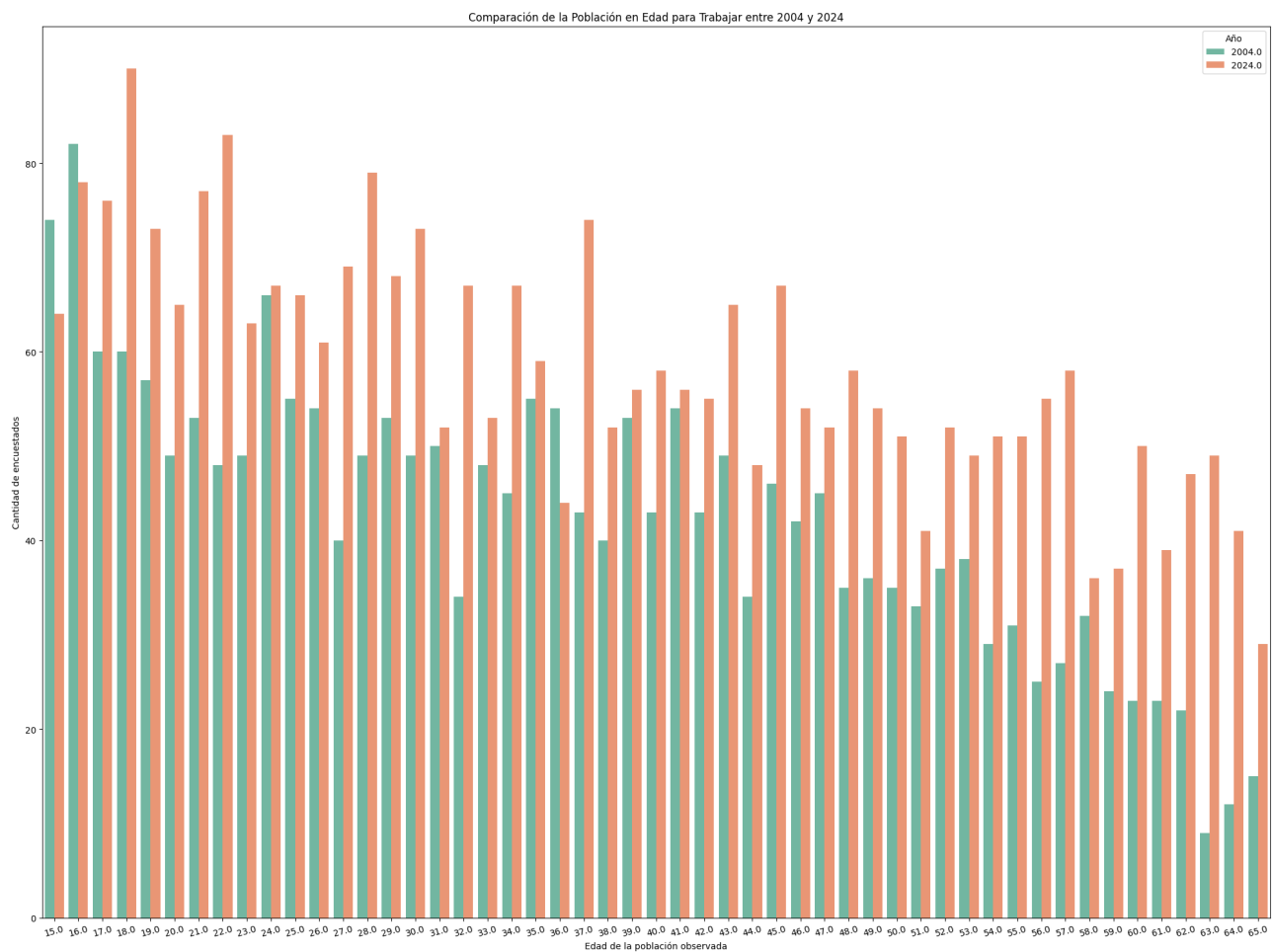
7)

De nuevo usamos “ESTADO” para hacer una columna PEA donde figuren los desempleados, luego hacemos una comparativa grafica que refleje los datos tomados:



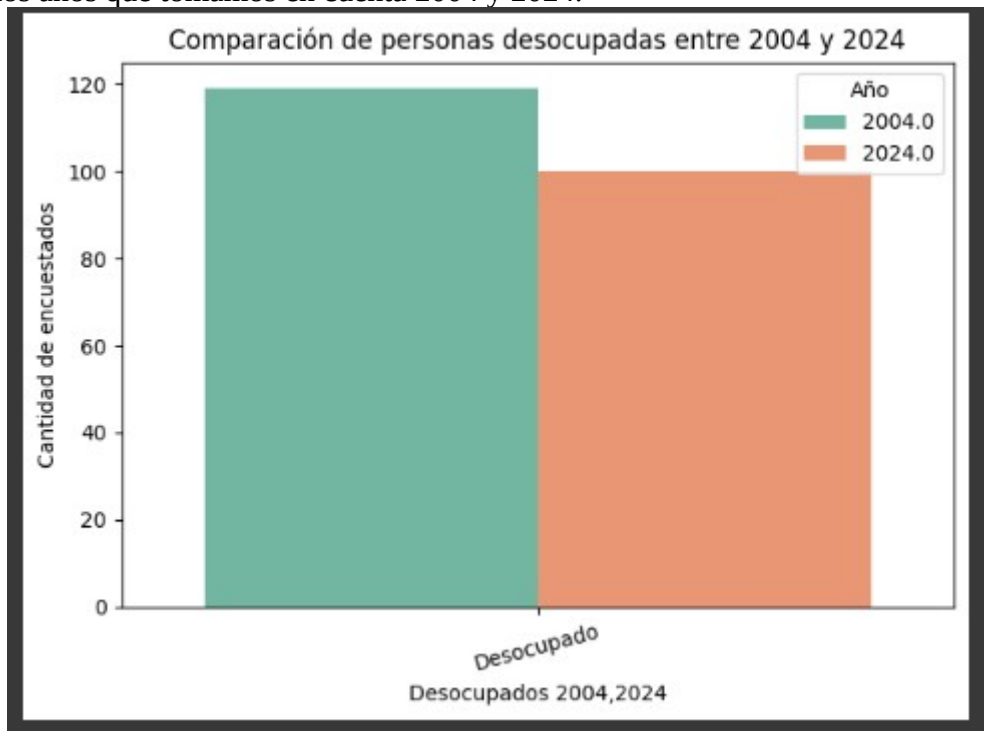
8)

Seguimos la metodología del anterior punto añadimos una columna PET y usamos en esta ocasión, la variable CH06 que maneja las edades de los encuestados para graficar una comparativa de edades entre los desempleados por año.



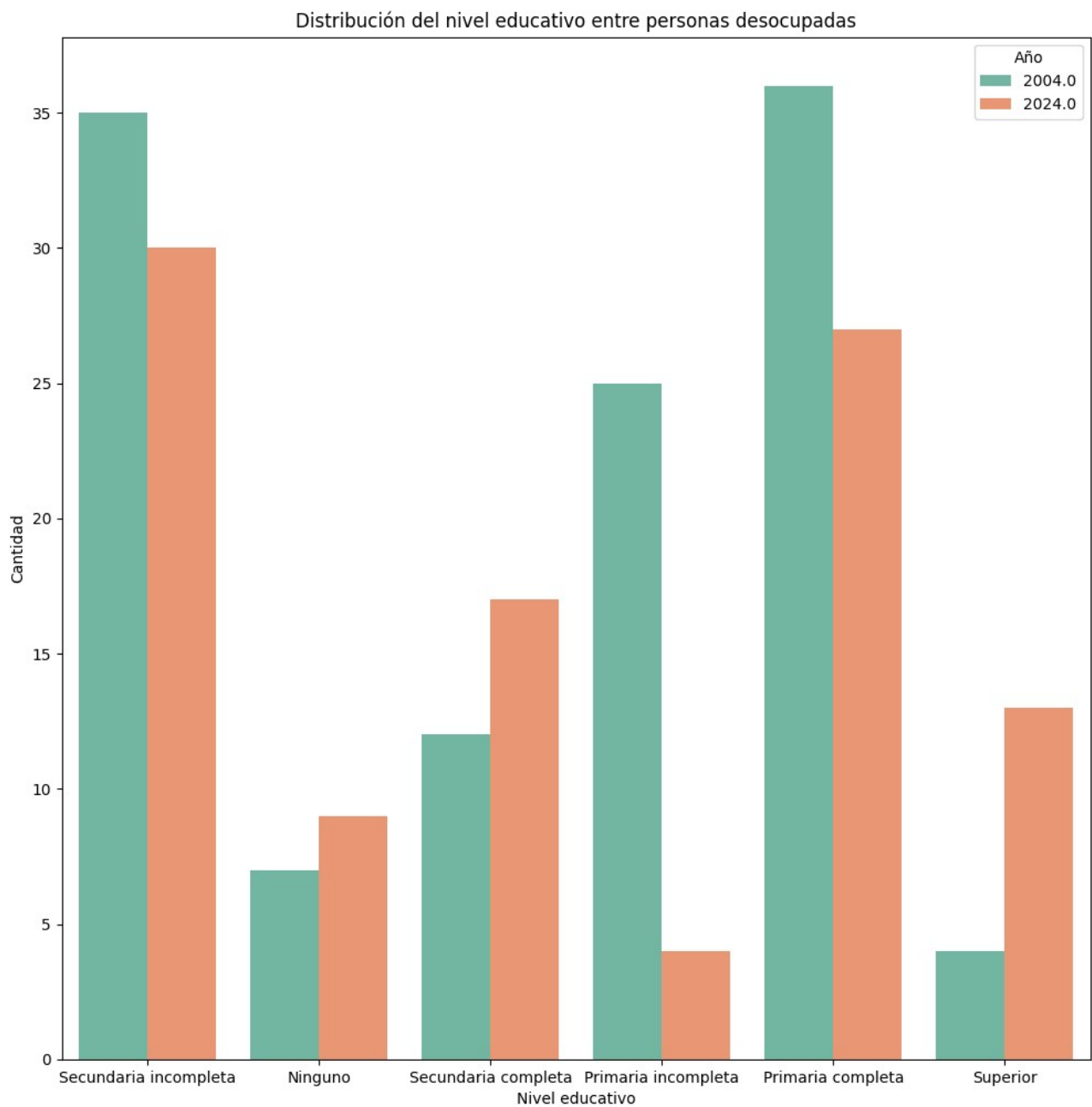
9)

Para este ultimo punto utilizamos las variables ya creadas para definir la cantidad de desocupados según los dos años que tomamos en cuenta 2004 y 2024.



A)

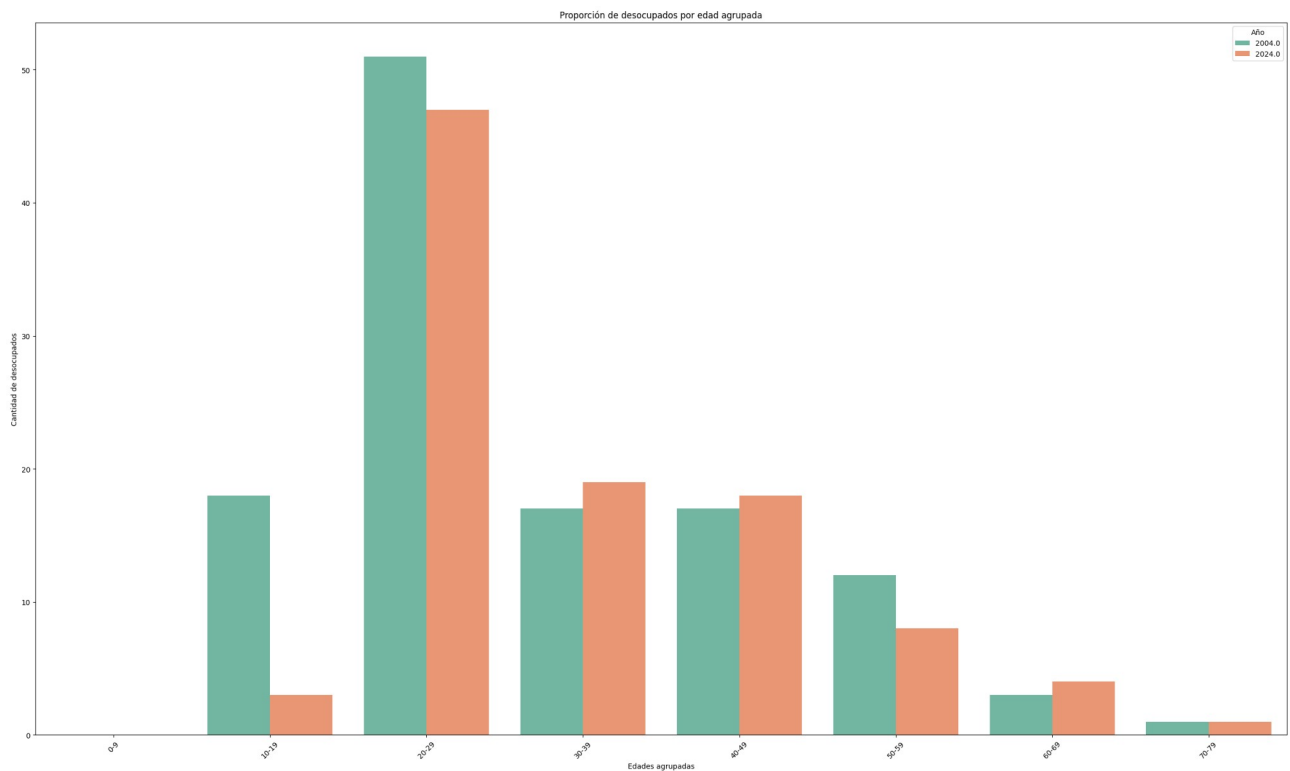
Acá se pide una comparativa de nivel de educación de los desempleados.



Lo que se puede ver que en los últimos años es que incluso aquellos que tienen superior completo tienen problemas de desempleo, incluso mayor que en el 2004.

B)

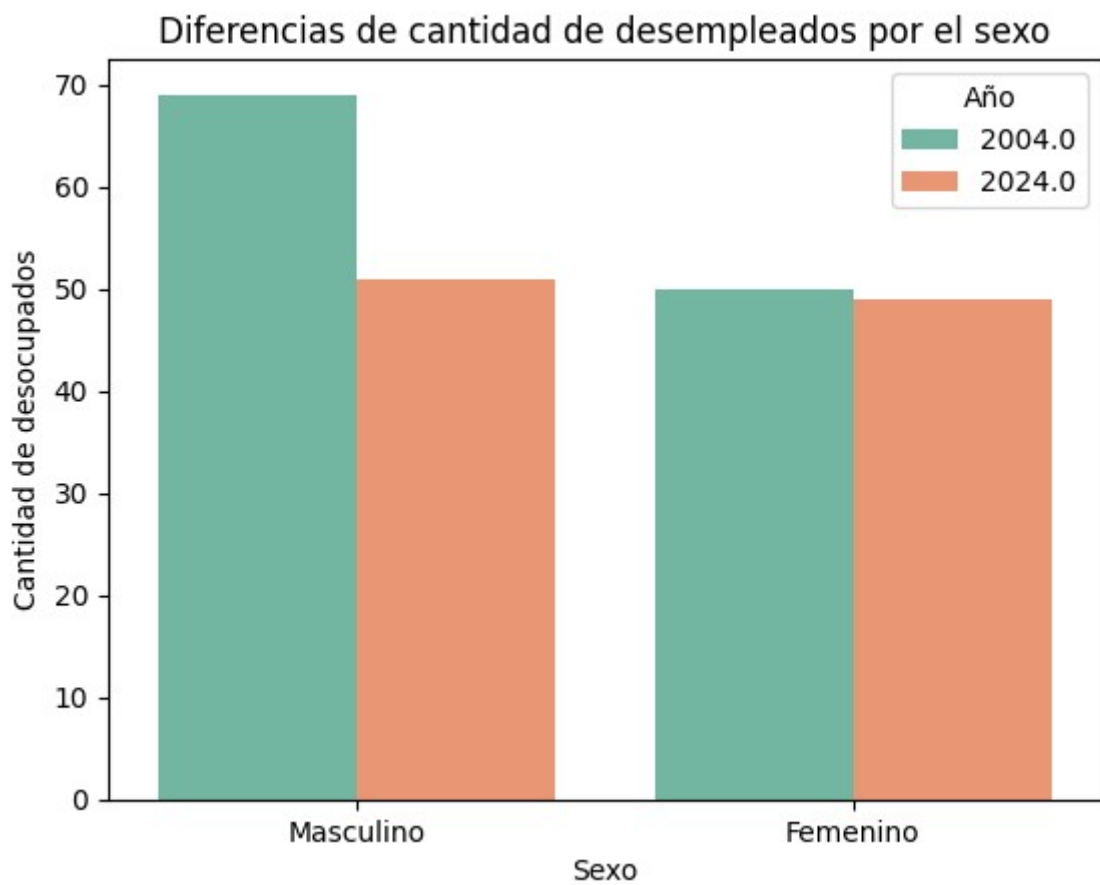
En este inciso se pide que con la variable CH06 se grafica una agrupación de edades de los desempleados entre 2004 y 2024.



Se muestra un gran nivel de desempleo entre los 20-29, esto según van creciendo las personas encuestadas va cayendo.

C)

La variable que elegimos es el sexo de los desocupados.



Lo que se puede ver es que el desempleo masculino se disminuyo en este periodo de 20 años.