

TRABAJO PRACTICO 2- _BIG DATA_ GRUPO 16

Grupo 16: Compuesto por: Máximiliano Petrini, Augusto Fasciani, Lautaro Solís.

Motivo del trabajo de investigación: El siguiente trabajo se realizó gracias a las encuestas realizadas, las EPH (Encuesta Permanente de Hogares), por el INDEC (Instituto Nacional de Estadísticas y Censos) en los años 2004 y 2024

Empezamos el trabajo filtrando y ordenado las bases de datos para posteriormente tratar sobre las distintas variables y sus cambios entre el 2004 y el 2024

Algunas de las variables que utilizamos fueron como: Cantidad de desocupados, Cantidad de ocupados, Cantidad de Inactivos, Sexo, etc.

Población ocupada:

Aquellos que trabajan, independientemente de la cantidad de horas trabajadas o el tipo de empleo.

Población desocupada:

Aquellos que no tienen trabajo pero que lo están buscando activamente. La búsqueda activa de empleo puede incluir acciones como enviar currículums, responder a anuncios de empleo o buscar información sobre oportunidades laborales.

Desarrollo del trabajo:

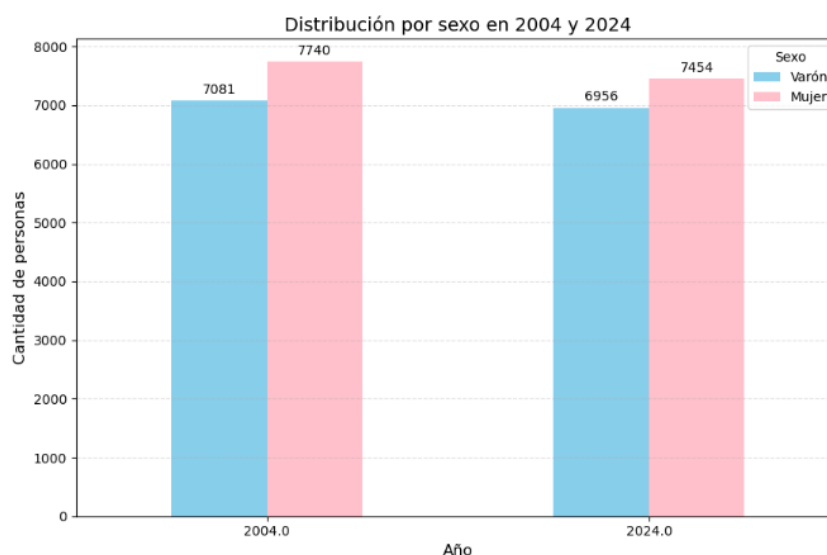
Cumpliendo las consignas preestablecidas, empezamos eligiendo una región y filtrando dicha base de datos. La región elegida fue la región Pampeana.

Continuamos nuestra investigación filtrando la base y eliminando ciertas apariciones de variables o casillas las cuales no aportaban información necesaria para el desarrollo del trabajo (limpieza de datos).

En esta etapa pudimos observar que en la base del año 2004 aparecen varios valores sin sentido, como nulos codificados con 'Ns./Nr.' o con ceros, que según la documentación de la EPH representan casos de no respuesta. En el año 2024, en cambio, se encontraron muchos valores directamente vacíos. Para poder trabajar mejor los datos y que los modelos no se vean afectados, esos valores se recodificaron como NaN. Esto ayuda a tener una estimación más limpia y precisa en los análisis que siguen.

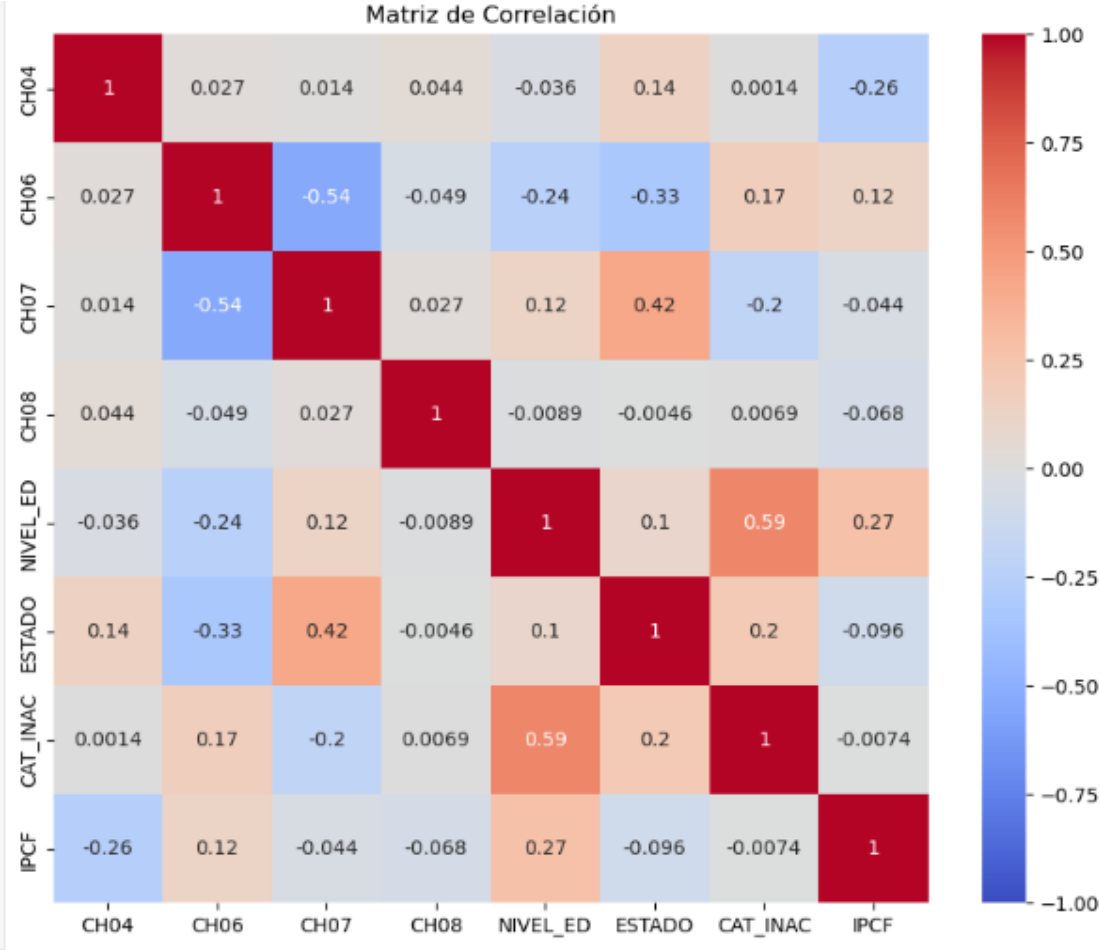
Una de las primeras variables a analizar fue el sexo. Aunque en ambos años la proporción de encuestados haya sido mayormente femenina, podemos observar un leve declive entre la proporción femenina encuestada en 2004 contra la de 2024. Esto genera que la muestra sea un tanto más representativa.

Cabe destacar que Aunque haya bajado la proporción de mujeres femeninas encuestadas, también bajo la de los hombres, esto nos podría llevar a concluir que hubo un aumento de las cantidades de personas que no hayan querido terminar de responder el cuestionario, o por ciertas otras variables que no son de completo interés para el desarrollo de dicho de trabajo (movimientos culturales del 2024 que manifiestan la expresión de múltiples géneros).



El siguiente gráfico, es una matriz de correlación. La matriz de correlación muestra que, en general, no hay relaciones fuertes entre las variables analizadas.

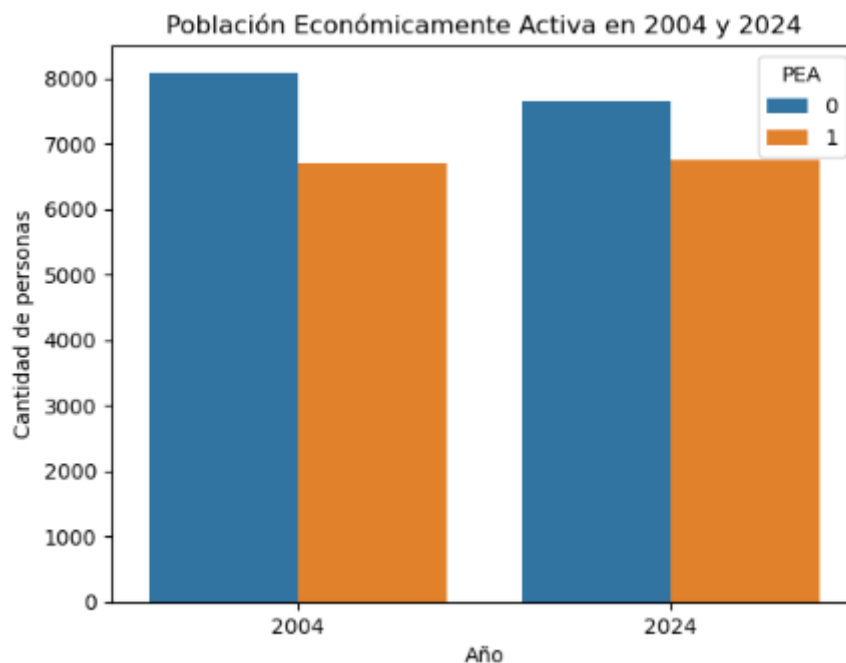
La mayoría de los valores están cerca de cero, lo que indica que las correlaciones son bajas. Esto sugiere que las variables no están fuertemente relacionadas entre sí.



Continuamos el desarrollo de nuestro trabajo filtrando a la base de datos según la cantidad de desocupados, cantidad de ocupados y la cantidad de personas inactivos de la muestra. Esto nos permitió posteriormente poder comparar la PEA.

La PEA, o Población Económicamente Activa, se refiere al conjunto de personas que tienen un trabajo o que están buscando activamente uno. En otras palabras, es la fuerza de trabajo de un país o región. La PEA incluye a la población ocupada (aquellos que trabajan) y a la población desocupada (aquellos que buscan trabajo activamente).

Según los resultados de las encuestas tomadas en 2004 y 2024 la cantidad de Desocupados & Ocupados [1] siempre fue menor al resto de la estadística (Inactivo, Menor de 10 años, No responde) [0], aunque la brecha se fue estabilizado para 2024.

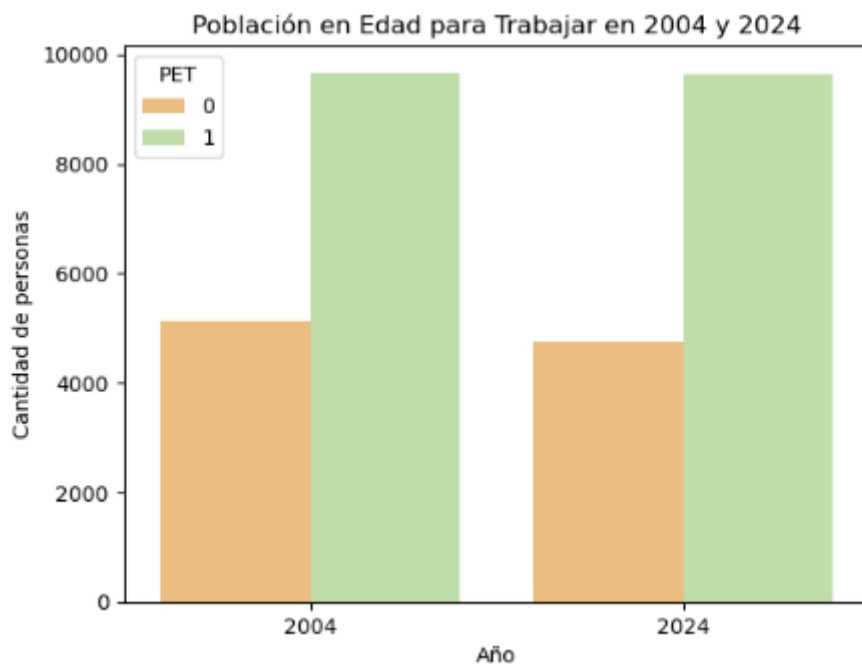


Esto nos podría llevar a asumir que se debe a la evolución de los problemas económicos a lo largo de la historia argentina lo cual provocaron el aumento del trabajo formal no registrado. Aunque dicha problemática haya disminuido (al observar la comparativa entre el año 2004 contra la del 2024), cabe destacar que es uno de los grandes problemas que enfrentan los institutos como el INDEC a la hora de tomar la EFH y derivar a ciertas conclusiones.

Población en edad de trabajar:

Podemos observar cómo durante ambos años hubo una gran mayoría población que estaban en edades para aptas para trabajar, llegando así casi a doblar en proporción a la población no apta para la actividad laboral debido a la condición de ser menores de edades. La población en edad para trabajar se mantuvo casi constante entre 2004 y 2024, mientras que la población no apta para trabajar se redujo en el 2024.

En comparación con el PEA, podemos notar que la población que participa del mercado laboral se mantuvo dentro de todo constante a través del tiempo.



Respetado las consignas establecidas, debimos crear una **variable categórica de años cumplidos** (ch06) agrupada de a 10 años. Como resultado se puede observar cambios en la proporción de desocupados entre 2004 y 2024. En general, la tasa de desocupación disminuyó en todos los grupos, esencialmente entre los jóvenes de 10 a 29 años. También hubo una reducción en los grupos de mayoría de edad, aunque en menor medida. (ver tabla en el código)

Por último, seleccionamos una **variable de interés**, está siendo la desocupación según el género. Esta variable muestra una disminución notable entre 2004 y 2024. En 2004, tanto mujeres como varones tenían una proporción de desocupación mayor, mientras que en 2024 esa proporción bajo. Esto indica que en general, la tasa de desocupación se redujo a la mitad en estos 20 años para ambos géneros, manteniéndose bastante parejo entre mujeres y varones,