

Trabajo Practico 2 ‘Big Data and Machine Learning’

Integrantes: Paulo Gonzalez, Andrea Duette, Uriel Masserdotti.

Parte 1: Familiarizándonos con la E.P.H. y la limpieza de datos

I

Resumido de manera breve, el INDEC considera a las personas ‘desocupadas’ como aquellas personas mayores a los 10 años que, estando disponibles para trabajar (en cierto tiempo de referencia), no tienen trabajo y que lo hayan buscado activamente en los últimos treinta días.

Definido en el registro de diseño y estructura, el estado del trabajador se define por la variable ‘ESTADO’, donde el valor que indica el estado de desocupación es el valor ‘2’.

II

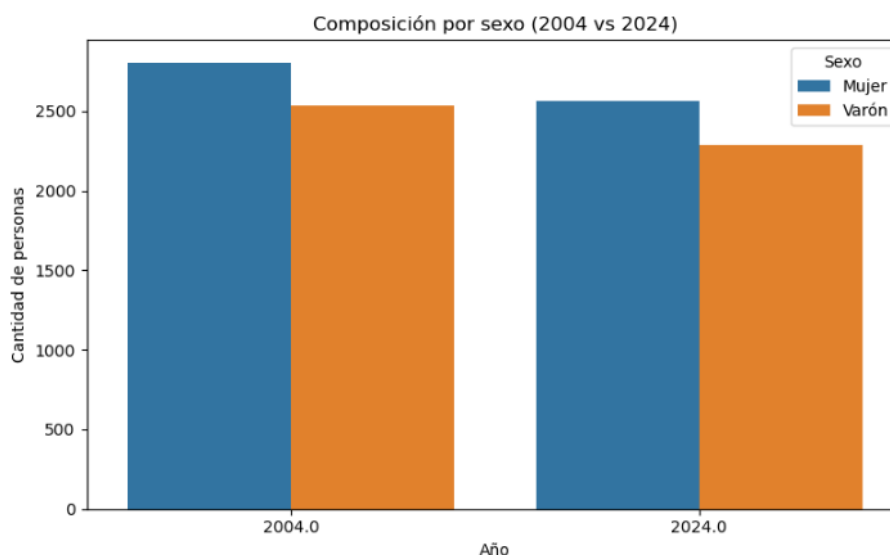
La región utilizada para el trabajo fue el Noreste Argentino (NEA), en los años 2004 y 2024, que fueron unificados en una única base de datos. Las demás variables de interés seleccionadas fueron: AGLOMERADO; PONDERA; CH04; CH06; NIVEL_ED; ESTADO; CAT_OCUP; CAT_INAC; PP04A; P21; RDECINDR; P47T; IPCF.

Una vez unificada, la variable PP04A fue la única que presentó valores faltantes (2.739 valores faltantes en total en el año 2024) y, en las variables seleccionadas, no se encontraron valores atípicos o sin sentido que eliminar.

Parte 2: Primer Análisis Exploratorio

III

Antes que nada, es de destacar que: en 2024 se entrevistaron a menos personas en general que en 2004 y, en ambos años, hubo una mayor cantidad de mujeres entrevistadas que de hombres:

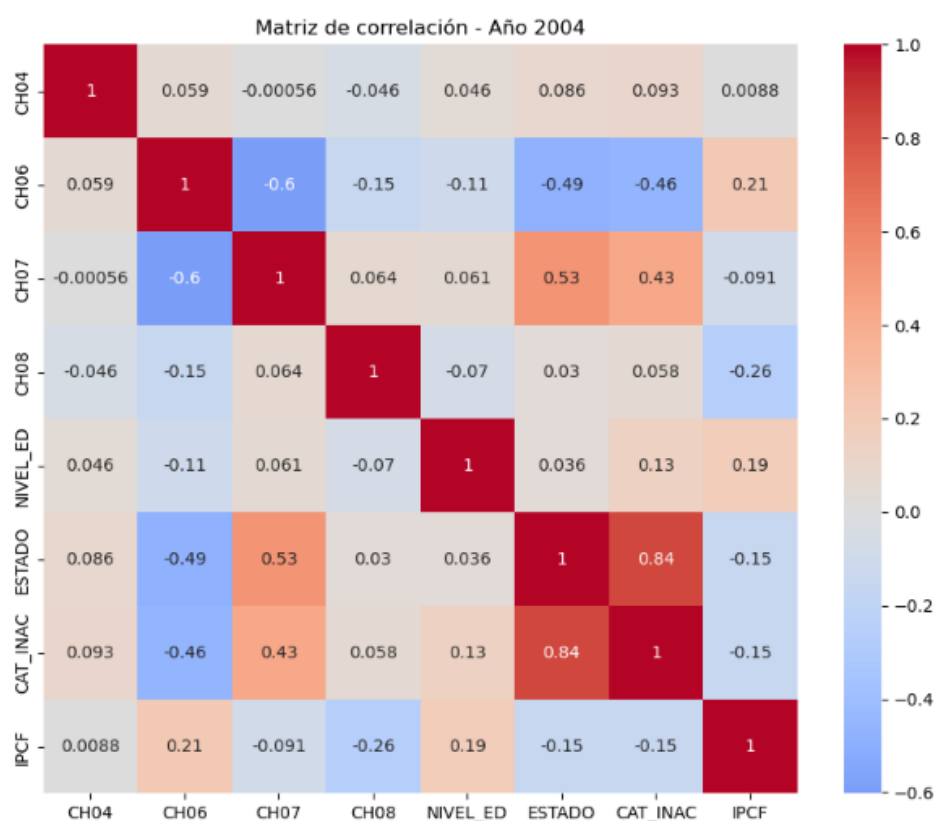


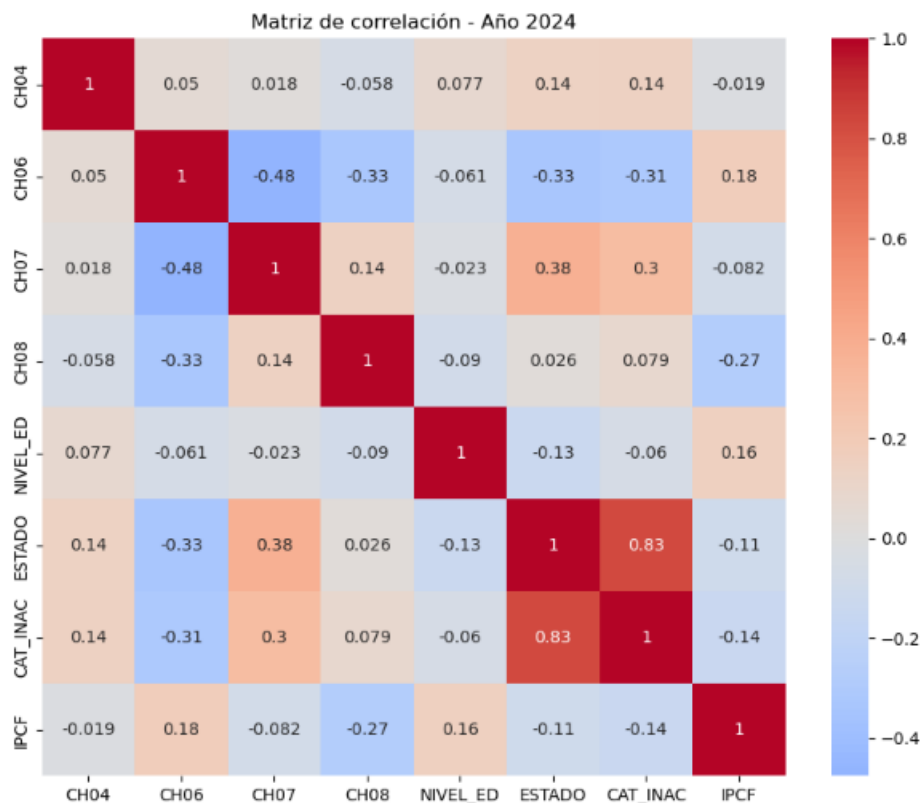
IV

Sobre la correlación entre variables para ambos años; se nota, en primer lugar, una fuerte relación entre el estado de ocupación ('ESTADO') y la edad ('CH06'). En este escenario se ve una fuerte relación negativa (-0,49), sugiriendo una mayor probabilidad de encontrar jóvenes activos ocupados o buscando trabajo y, por el otro lado, personas mayores no pudiendo encontrar puestos de trabajo o dejando de buscarlos. Si se compara la misma relación para el 2024 (-0,33), puede verse un aumento positivo de la relación que puede deberse a factores como periodos de vulnerabilidad económica que fuerzan a aquellos sectores de mayor edad a seguir dentro del mercado laboral trabajando o buscando trabajo de forma activa o, por otro lado, un mercado laboral que dificulta el acceso a nuevas generaciones a puestos de trabajo (ambos escenarios siendo no excluyentes uno del otro y sin considerar los avances en salud que alargan la esperanza y la calidad de vida).

Otro punto destacable puede ser la relación entre el nivel de educación ('NIVEL_ED') y el ingreso per cápita familiar ('IPCF') (0.19), señalando un mayor nivel de ingreso a mejores niveles de educación alcanzados.

En general, para 2024, las relaciones entre las variables son más débiles que en 2004.





Parte 3: Conociendo a los ocupados y desocupados

Como la región que elegimos para trabajar es el Noreste Argentino, los resultados presentados son de dicha región:

V

Algunas consideraciones sobre los ingresos promedios: primero, puede llegar a tener sentido que los inactivos posean ingresos, incluso más altos que los desocupados, aunque parezca contra intuitivo. La categoría de ‘inactivo’ incluye a personas que no tienen trabajo y que no lo buscan activamente, como jubilados, amas de casa, estudiantes, personas con discapacidades, etc. Estas personas, aunque no trabajen, perciben prestaciones económicas del Estado que cuentan como ingresos y, por lo general, viven con otras que sí generan ingresos de una ocupación laboral, lo que empuja el ingreso promedio familiar hacia arriba.

CANTIDAD DE PERSONAS POR ESTADO Y AÑO			
Año	Ocupado	Desocupado	Inactivo
2004	1.789	215	2.303
2024	2.111	127	1.981
MEDIA DE IPCF POR ESTADO Y AÑO (en pesos)			
Año	Ocupado	Desocupado	Inactivo
2004	244	148	196
2024	144.161	85.540	119.954

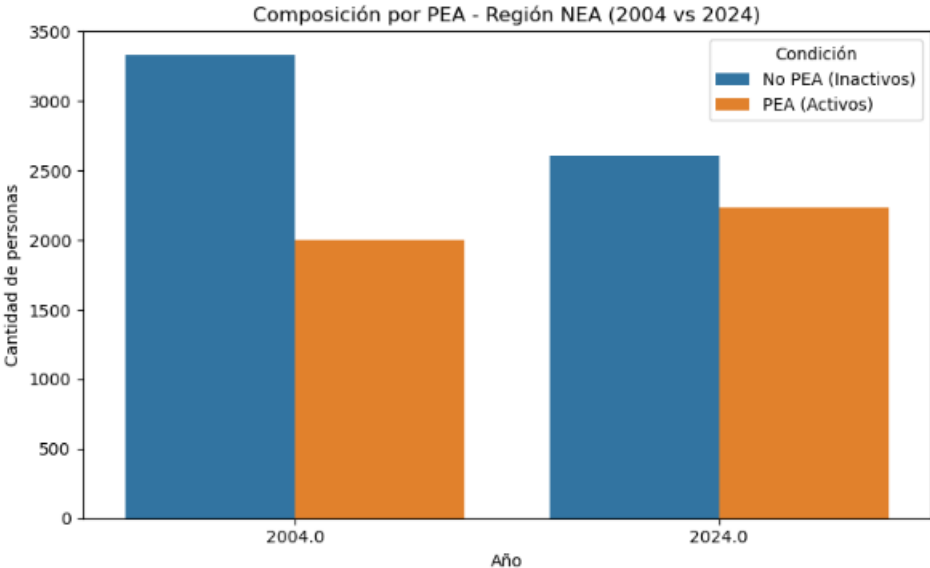
Y, además, aunque los ingresos de los ocupados puedan parecer bajos, hay que tener en cuenta que el ingreso familiar se promedia con todos los miembros del hogar (independientemente de su condición laboral y también contando a los niños).

VI

En la muestra seleccionada, no respondieron once personas, un valor que no posee mucha significancia para el análisis al ser tan bajo.

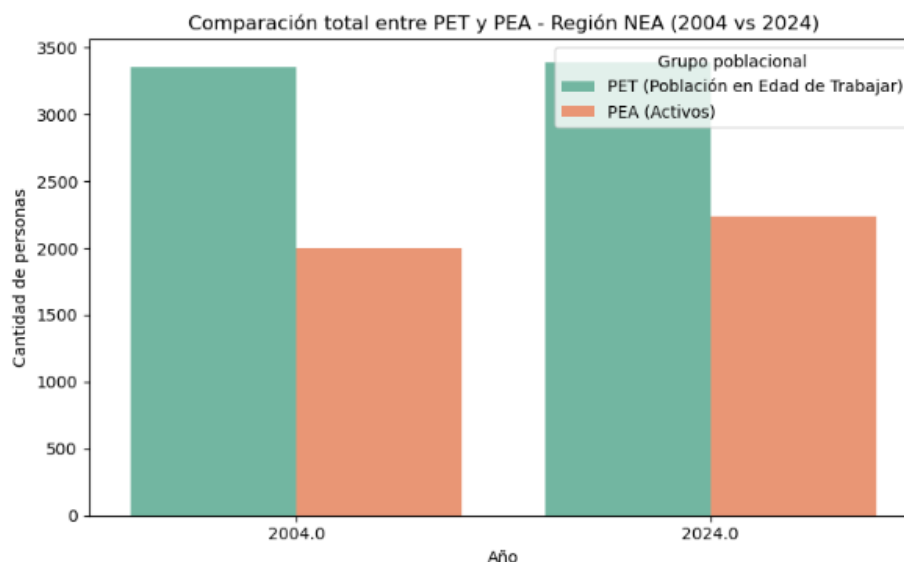
VII

El grafico muestra cómo la población económicamente activa aumentó desde el 2004 en comparación con el 2024. Más en profundidad (ver cuadro del punto V), este aumento se dio con fuerza en la población ocupada, mientras que hubo un descenso en la población desocupada. Este descenso de la población desocupada se observa como un traspaso de población desocupada a población ocupada, ya que hay que tener en cuenta que la cantidad de gente inactiva también cayó (en resumen, hay más capacidad para acceder a puestos de trabajo en 2024 que en 2004).



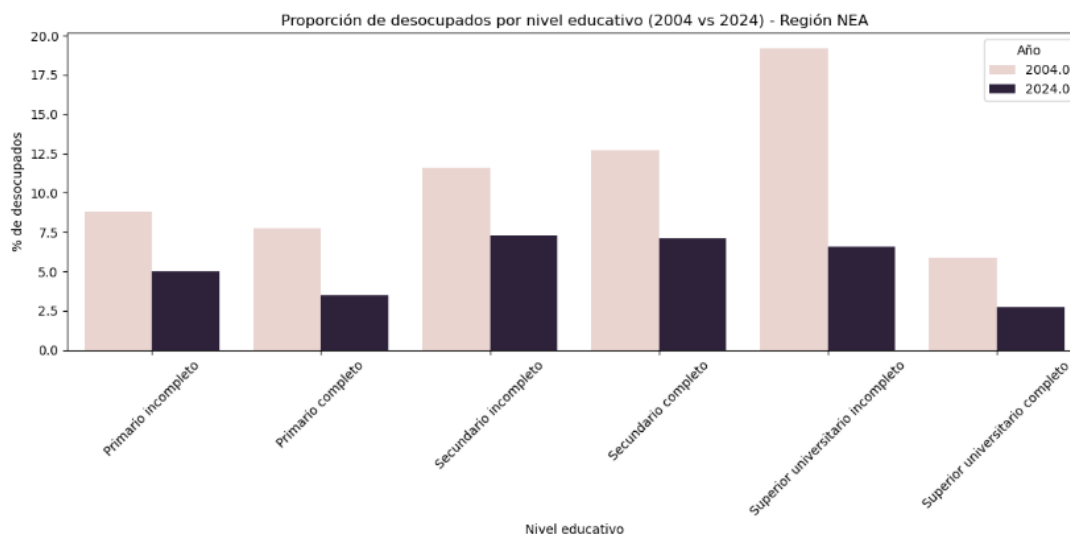
VIII

Dentro de la población en edad de trabajar, los resultados son similares, pero, viendo la alta cantidad de gente de la p.e.t. en relación a la población económicamente activa, es un dato desalentador, ya que muestra que hay mucha cantidad de personas en edad de trabajar que no lo están haciendo.

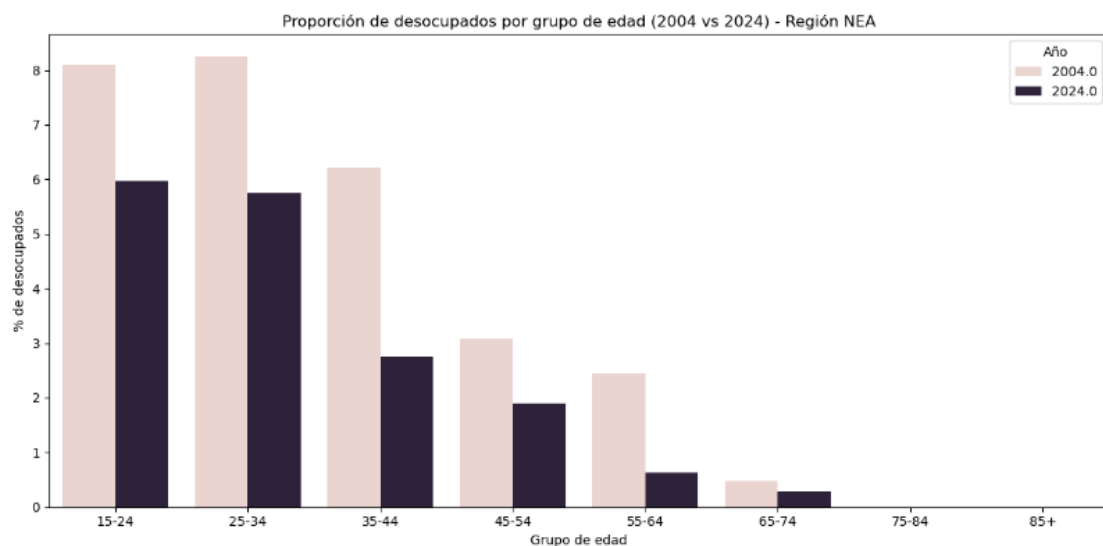


IX

Ya se vio en el punto V que la cantidad total de desocupados del 2004 al 2024 descendió (de 215 a 127). Si se divide a estas personas por su nivel educativo se obtiene que, en 2004, hubo una altísima tasa de desocupación para aquellas personas que se encontraban cursando una carrera universitaria. Para el 2024 no solo que se reducen todas las tasas de desocupación para todos los niveles, sino que también se invierte la tendencia de desocupación, protegiendo más a aquellas personas que tienen sus estudios completos, sea el nivel que sea.



En cuanto a la división de la muestra en grupos de edades, se ve una desocupación más alta en jóvenes y adultos en edad de terminar sus estudios universitarios, quizá por dificultad de acceder al mercado laboral y por razones explicadas anteriormente. También se ve una fuerte reducción a partir de los 50 años aproximadamente debido a personas mayores que empiezan a abandonar el mercado laboral (personas que se jubilan) o que están en el último puesto de trabajo antes de jubilarse.



Luego, habiendo elegido dividir el análisis por aglomerados urbanos del Noreste Argentino, se ve una alta reducción de la desocupación en Posadas, un relativo estancamiento en Gran Resistencia y reducciones constantes en Corrientes y Formosa.

