

Reporte Big Data TP2 Grupo20.pdf

Punto 1 - Cómo Identificar personas desocupadas según el INDEC

Se identifica a las personas desocupadas como, las personas que por el momento no están trabajando, pero están en la búsqueda, es decir están dispuestos a trabajar.

Punto 2b - Variables

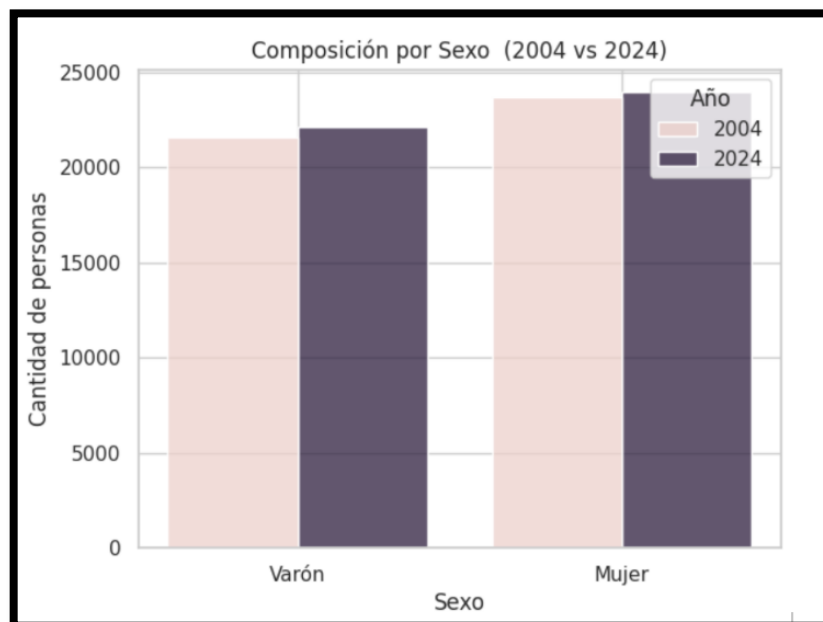
Las variables que tienen más valores faltantes son GDECCFR (Numero del conjunto de aglomerados de 500.000 y más habitantes) y PDECCFR (número de decil de ingreso per cápita familiar del conjunto de aglomerados de menos de 500.000 habitantes). El año es 2024.

Punto 2c – Proceso de limpieza

Se realizó una limpieza en las 15 variables de ambos años, lo que mostro que había más valores faltantes de los que habíamos observado en el paso anterior. Para corregir los valores sin sentido, buscamos en los diccionarios de EPH cuáles eran estos valores (9, 99, 999, 9999, -7, -8, 0), estos eran iguales en ambos años. Nos dio como resultado que el año que tenía mayor cantidad de valores faltantes era 2024, con 7647.

Punto 3 - Composición por sexo

En el gráfico, podemos observar que la cantidad de hombres y mujeres es similar. Podemos ver que en 2024 hay una mayor cantidad de personas registradas, esto se puede deber a crecimiento poblacional.

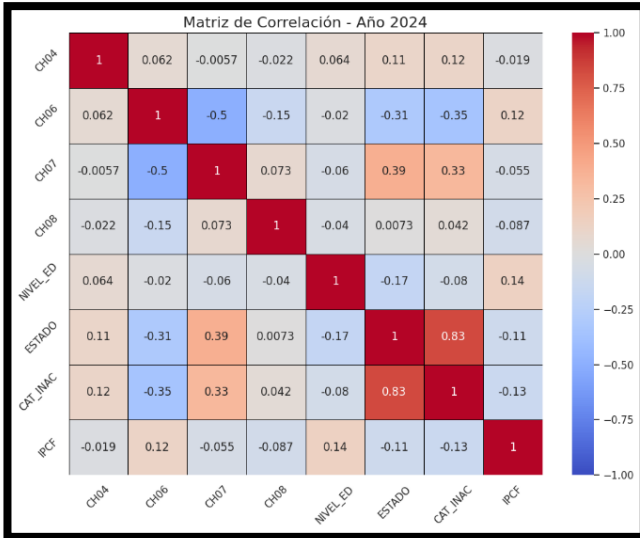
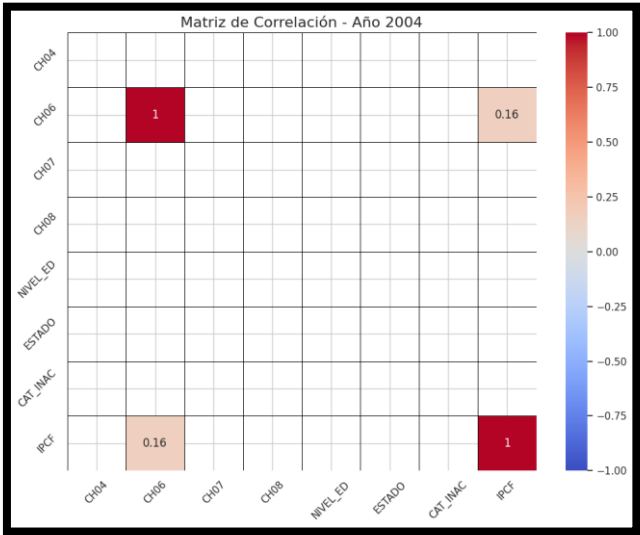


Punto 4 – Matriz de Correlación (2004 y 2024)

Matriz de correlaciones entre las variables seleccionadas: sexo (CH04), edad (CH06), estado civil (CH07), parentesco (CH08), nivel educativo (NIVEL_ED), condición de actividad (ESTADO), categoría de inactividad (CAT_INAC) e ingreso per cápita familiar (IPCF), para los años 2004 y 2024.

En ambas matrices observamos una correlación positiva moderada entre edad y categoría de inactividad, indicando que a mayor edad aumenta la probabilidad de inactividad. También encontramos una leve correlación negativa entre nivel educativo y condición de desocupación, lo que sugiere que, a mayor educación formal, menor es la probabilidad de desempleo.

Las correlaciones en general son débiles a moderadas, lo que implica que las variables analizadas no están fuertemente asociadas entre sí. La comparación entre 2004 y 2024 muestra patrones similares con pequeñas variaciones en las intensidades de las correlaciones.



Punto 5 – Desocupados e Inactivos

Para el año 2004, encontramos “X” desocupados e “Y” inactivos en la muestra, mientras que en 2024 encontramos “Z” desocupados y “W” inactivos. Analizando los ingresos per cápita familiares (IPCF) según estado laboral, los ocupados presentan los ingresos promedio más altos, seguidos por los inactivos y luego los desocupados.

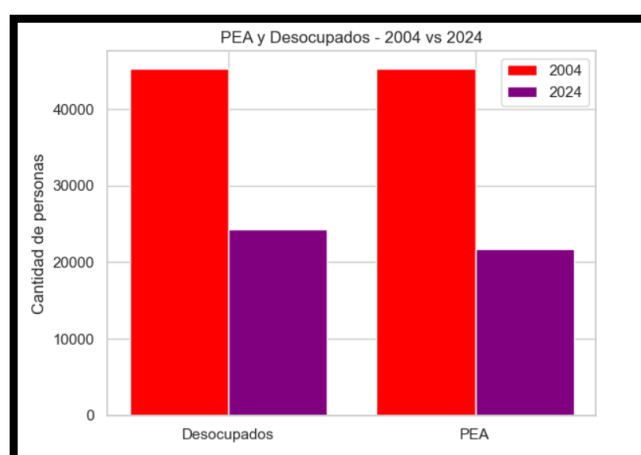
Esta tendencia se mantiene en ambos años, aunque los valores absolutos varían. Estos resultados reflejan las diferencias de acceso a recursos económicos entre los distintos estados de participación en el mercado laboral.

Punto 6 – Análisis de No Respuesta en la Condición de Actividad

Analizamos la variable ESTADO para identificar los casos de no respuesta en la condición de actividad. En el año 2004, encontramos X personas que no respondieron, mientras que en 2024 la cantidad fue de Y personas. Para continuar con el análisis, dividimos las bases en dos subconjuntos: "respondieron" (observaciones con respuesta válida en ESTADO) "no respondieron" (observaciones con ESTADO igual a o). La separación permite trabajar únicamente con datos consistentes, evitando sesgos por falta de respuesta en el análisis de actividad laboral.

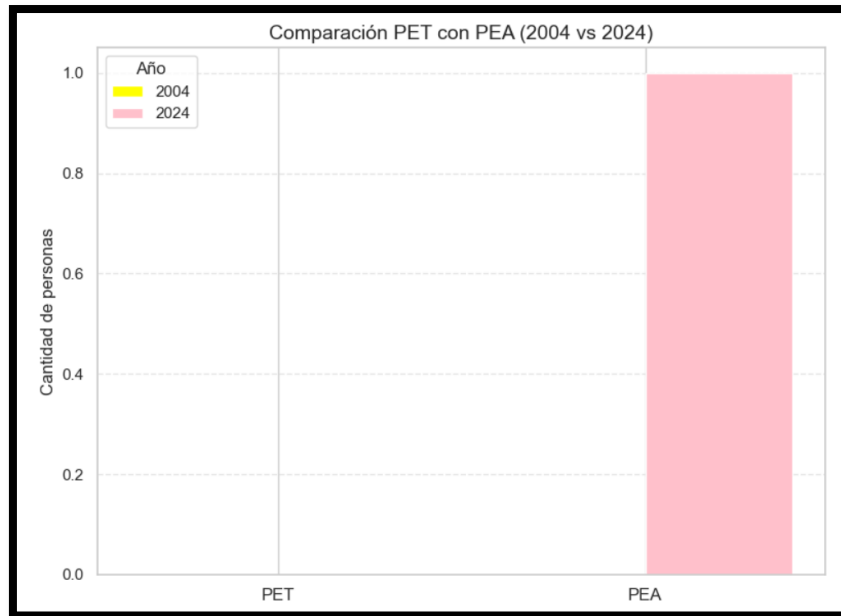
PUNTO 7 – Composición PEA para 2004 y 2024

En el gráfico se observa que tanto la cantidad de personas desocupadas como la población económicamente activa (PEA) disminuyeron en 2024 respecto a 2004.



Punto 8 – Comparación PET Y PEA

No hay datos con los cuáles trabajar para el caso de PET (Población en edad para trabajar). Ya sea por un error de código, definición o que simplemente la base de datos en esa categoría este vacía.



Punto 9a - ¿Hubo cambios de desocupados por nivel educativo?

Teniendo en cuenta que los registros en 2004 no tienen datos de personas desocupadas por nivel educativo. Lo tomaremos como si fuese un cero. En tal caso, para el 2024 hubo una distribución detallada según las subcategorías existentes. La desocupación afecta más a quienes tienen niveles de educación medios (en su mayoría incompleta).

Punto 9b - ¿Hubo cambios de desocupados por edad?

En 2004 no hay registros de personas desocupadas en ningún grupo etario (ya sea por error de código o falta de datos en la base). La desocupación en 2024 se concentra más en el grupo etario entre 20 y 30 años (jóvenes adultos), pero a medida que la edad aumenta, la desocupación disminuye. La gente adulta y los jubilados ya están dentro del mercado laboral, en su mayoría.