Big Data y Machine Learning

Segundo Trabajo Práctico

# Parte I.

1. Viendo los indicadores socioeconómicos del mercado de trabajo de la EPH del [cuarto cuatrimestre de 2024](https://www.indec.gob.ar/uploads/informesdeprensa/mercado_trabajo_eph_4trim24083C6B9E41.pdf), en la página 3, muestra un cuadro donde incluye a los desocupados en la población económicamente activa describiéndolos como personas que, aunque estén desempleadas, activamente están buscando empleo.

Se destacó que para los hombres la tasa de desocupación es del 6.1% y para las mujeres del 6.9%, que las regiones de mayor tasa fueron Gran Buenos Aires y la Pampeana, que en aglomerados de menos de 500.000 habitantes había menos desocupación que aquellos que sobrepasaban la cifra y por último que el 71.2% de éstos, tenían el secundario completo y el otro 28.8% llegaban hasta un nivel universitario.

1. B. Elegimos el aglomerado 17 de Neuquén – Plottier, eligiendo las siguientes variables: sexo, edad, condición de actividad, nivel educativo, ingreso per cápita familiar, posibilidad de descuento jubilatorio por trabajo, aporte por si mismo a algún sistema jubilatorio, monto de ingreso por jubilación, subsidios, alquiler, cuota alimentaria y/u ocupación principal, categoría de inactividad, estado civil y tipo de cobertura médica.

Al buscar valores faltantes, sorpresivamente y como se puede ver en el siguiente cuadro, las únicas dos variables con valores faltantes fueron las ingreso por alquiler y el de cuota alimentaria, con 1142 y 1134 variables faltantes respectivamente en el primer trimestre de 2024.

Tabla

El contenido generado por IA puede ser incorrecto.

c. Para el proceso de limpieza decidimos separar las variables en numéricas y categóricas, poniendo límites entre los cuales se encontrarían los valores lógicos con tal de descartar cualquier valor fuera de este rango, por ejemplo, pusimos edad de 0 a 130 años, descartando todo valor negativo o mayor a 130. Así tomando un criterio distinto para cada una llevamos a cabo el proceso.

# Parte II.

1. Hicimos los gráficos para ver el porcentaje de hombres/mujeres en 2004 y 2024 dando los siguientes resultados:

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

Viendo que la distribución no sufrió cambios notables.

1. Al hacer la matriz, pudimos ver que en 2004, vemos relaciones fuertes en edad y soltería (-0.76) lo que puede referirse a que se distribuyen opuestamente los grupos etarios frente a los puestos de trabajo, el tipo de empleo con la edad (0.59) y la tasa de inactividad en estudiantes y edad (-0.59), lo que quiere decir que a mayor edad menor inactividad y viceversa.

Algunas relaciones moderadas/débiles son, la tasa de inactividad de amas de casa y edad (0.27) y las personas viudas y edad (0.31).

A destacar, en 2024 podemos ver una gran estabilidad, sin nuevas correlaciones fuertes o cambios notables en ninguna variable.

Gráfico

El contenido generado por IA puede ser incorrecto.

Gráfico

El contenido generado por IA puede ser incorrecto.