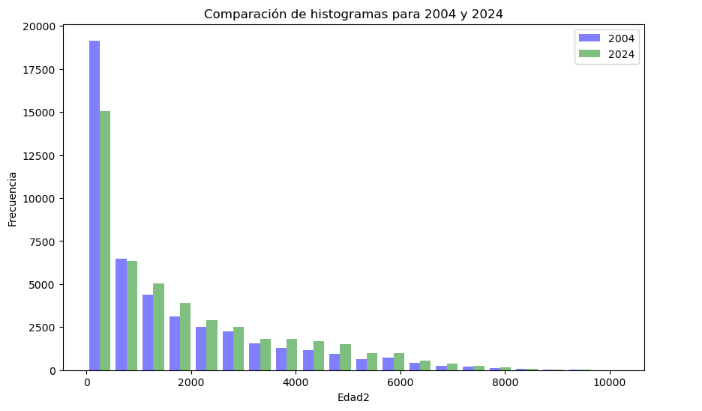
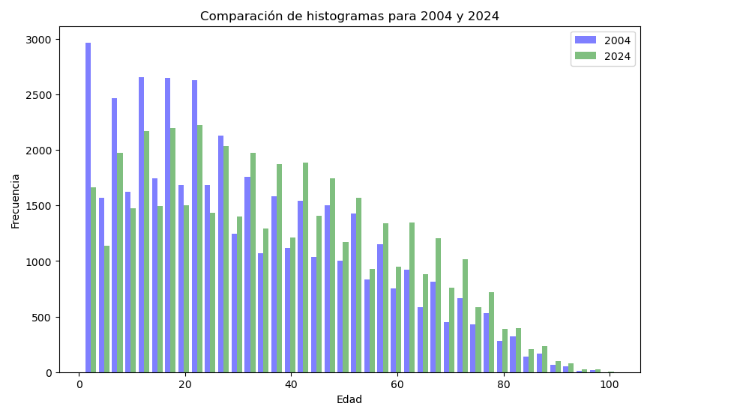
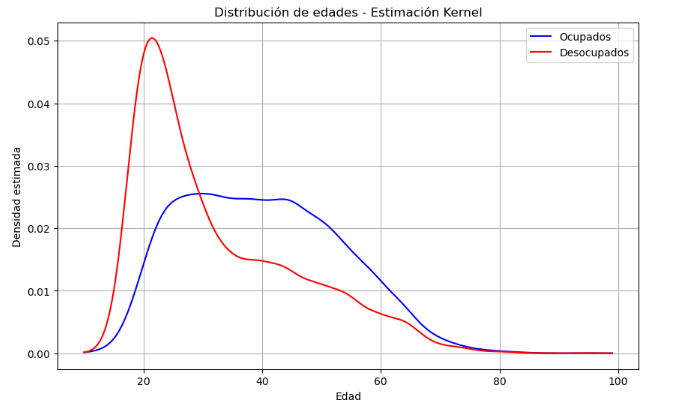
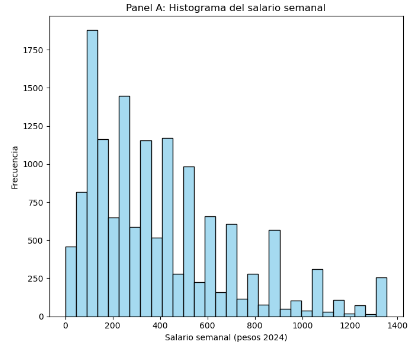
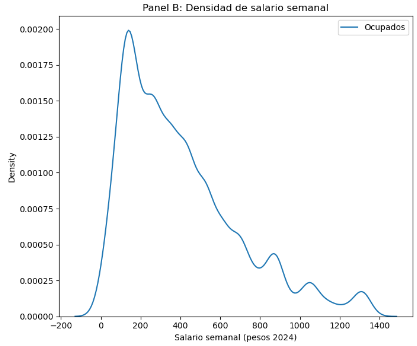
TP3 Big data and Machine Learning  
Franco Bustelo  
Juan Pablo Tomasello  
David Agudelo  
  
El gráfico muestra un histograma comparativo de la variable Edad2 (elevada al cuadrado y luego con raíz cuadrada, por lo que es equivalente a Edad2) para los años 2004 (en azul) y 2024 (en verde).

Se puede ver que tiene una distribución decreciente en ambos años. En ambos años, la mayor parte de los datos está concentrada en los valores bajos de edad2, lo que sugiere que la mayoría de las personas tienen valores pequeños de esta variable.  
Hay diferencias leves entre 2004 y 2024. En varios rangos medios de 1000 a 4000 se observa que hay ligeramente más casos en 2024 que en 2004, lo que podría indicar un pequeño envejecimiento de la población o cambios en la estructura etaria. Y por otro lado se puede ver también que la estructura poblacional no cambió drásticamente, aunque sí se aprecian algunos desplazamientos en ciertas edades.



El gráfico muestra que en ambos años la mayoría de la población se concentra en edades jóvenes. En 2024 hay una mayor proporción de personas en edades medias y mayores respecto a 2004, lo que sugiere un leve envejecimiento poblacional. Estas edades son más claras ya que se están usando las reales y no las que fueron elevadas al cuadrado

****  
La distribución muestra que los desocupados (línea roja) se concentran principalmente entre los 20 y 25 años, mientras que los ocupados (línea azul) están más distribuidos en edades mayores.

  
  
**Panel A**: Histograma del salario semanal   
Muestra que la mayoría de los ocupados gana entre 100 y 400 pesos semanales (ajustados a pesos de 2024), con una alta frecuencia en los tramos más bajos.  
  
  
**Panel B**: Densidad de salario semanal   
Refuerza la concentración en salarios bajos y revela una distribución asimétrica con larga cola derecha. Este histograma nos está indicando que hay pocos casos con ingresos altos.

|  |  |  |  |
| --- | --- | --- | --- |
|  | 2004 | 2024 | Total |
| Cant. de observaciones | 45289 | 46050 | 91339 |
| Can. De observaciones con nas en la variable “Estado” | 35 | 81 | 116 |
| Cantidad de ocupados | 17022 | 20325 | 37347 |
| Cantidad de desocupados | 2717 | 1362 | 4079 |
| Cantidad de variables limpias y homogeneizadas | 1362 | 1176 | 2538 |