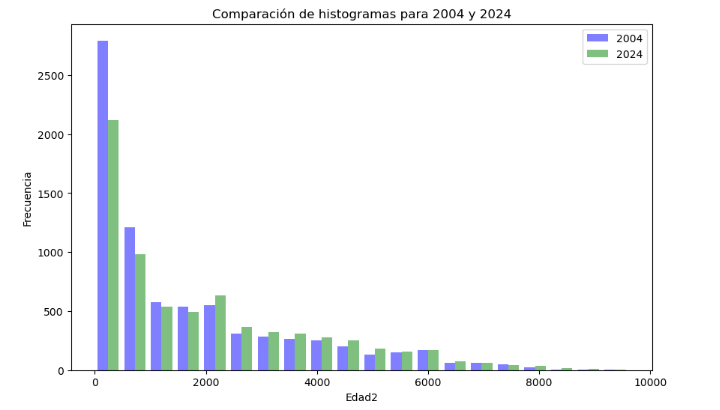
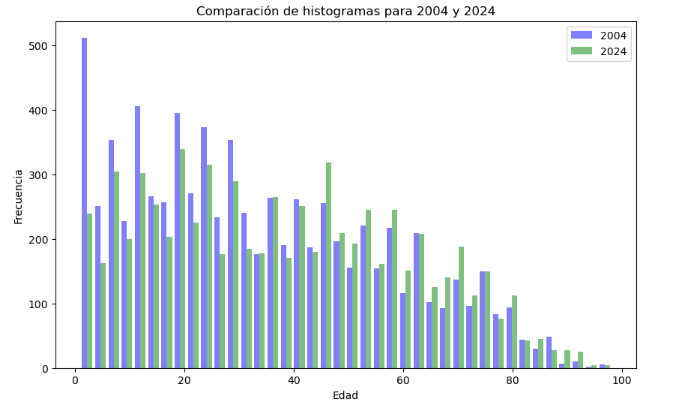
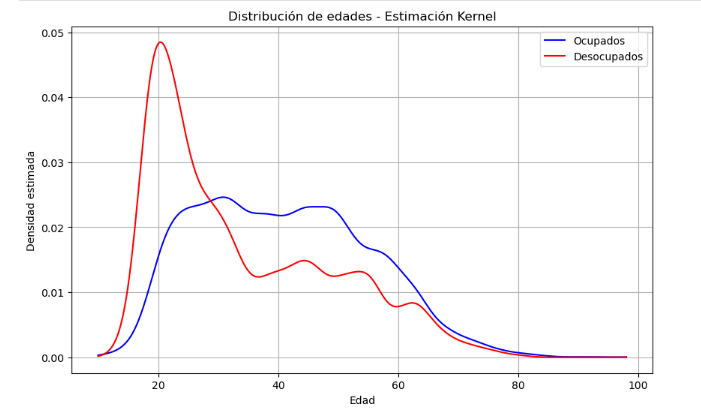
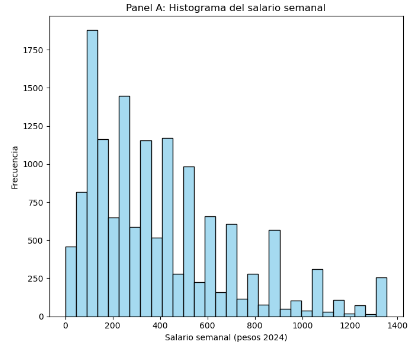
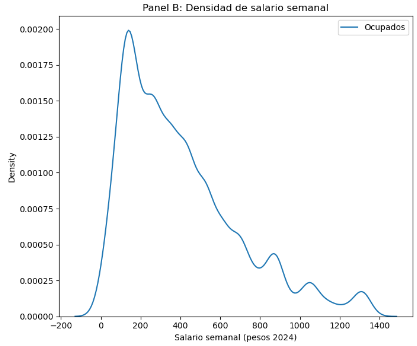
TP3 Big data and Machine Learning  
Franco Bustelo  
Juan Pablo Tomasello  
David Agudelo.   
El gráfico muestra un histograma comparativo de la variable Edad2 (elevada al cuadrado y luego con raíz cuadrada, por lo que es equivalente a Edad2) para los años 2004 (en azul) y 2024 (en verde).

El gráfico compara la distribución de la **edad al cuadrado** entre 2004 y 2024. En ambos años, la mayor frecuencia se da en valores bajos, lo que indica predominancia de personas jóvenes. Sin embargo, en 2024 hay un leve aumento en los valores más altos, lo que sugiere un envejecimiento relativo de la población en ese período. 

El gráfico muestra la distribución de la **edad** entre 2004 y 2024. En 2004 predominan los grupos más jóvenes, especialmente menores de 30 años. En 2024 se observa una mayor frecuencia relativa en edades medias y altas (de 40 años aproximadamente), lo que indica un **envejecimiento de la población** en el período analizado.

****La mayor densidad de desocupados se concentra en edades **jóvenes de alrededor de 20 años,** mientras que los ocupados presentan una distribución **que abarca más edades**, con mayor densidad entre los **22 y 60 años aproximadamente**. Esto podría tratarse de un indicador de que el desempleo afecta principalmente a personas jóvenes.   
  
  
**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 | 7647 | 7051 | 14698 |
| Can. De observaciones con nas en la variable “Estado” | 10 | 41 | 51 |
| Cantidad de ocupados | 3079 | 3224 | 6303 |
| Cantidad de desocupados | 528 | 311 | 839 |
| Cantidad de variables limpias y homogeneizadas | 311 | 232 | 543 |