

Análisis exploratorio estadístico de la brecha salarial

Isabel Romo Bru

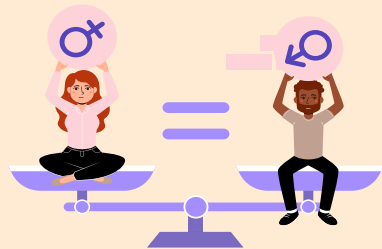


SOBRE LOS DATOS

Este dataSet contiene el salario promedio por hora en los empleados, por sexo, edad y ocupación, mediante el Indicador 8.5.1 (Ingresos promedio por hora de los empleados, por sexo, edad, ocupación). Los datos han sido recogidos en ILOSTAT (International Labour Organization), un portal estadístico que recoge los datos sociolaborales.

HIPÓTESIS:

- ¿Existe diferencia en ingresos entre hombres y mujeres?
- ¿Hay alguna profesión en la que las mujeres ganen más que los hombres?
- ¿Cuál es el sector laboral en el que la brecha de ingresos es mayor?
- ¿Cuál es el país con las mayores diferencias salariales?
- ¿Hay algún patrón en países respecto a la brecha salarial?
- ¿Ha evolucionado en los últimos años la brecha salarial?

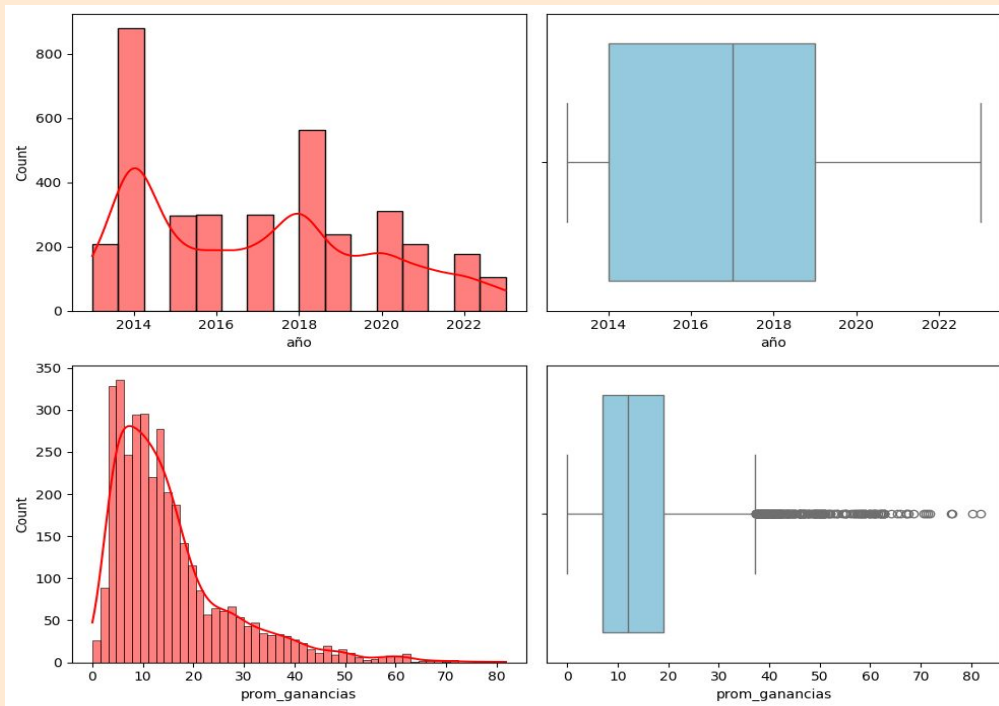
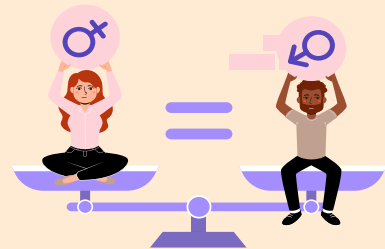


VARIABLES

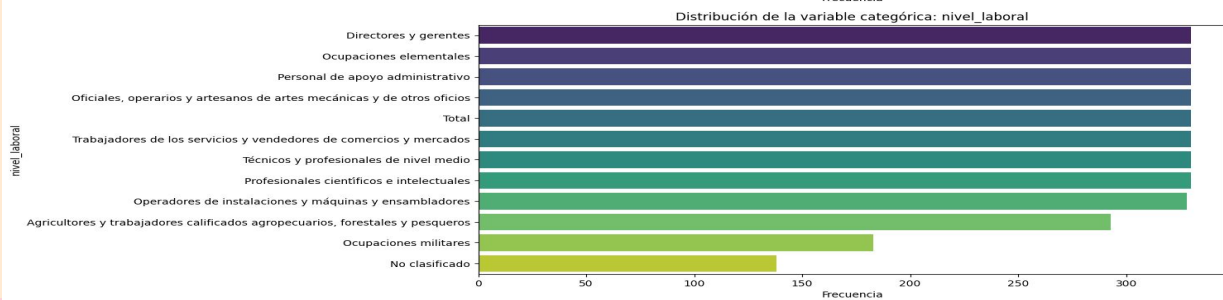
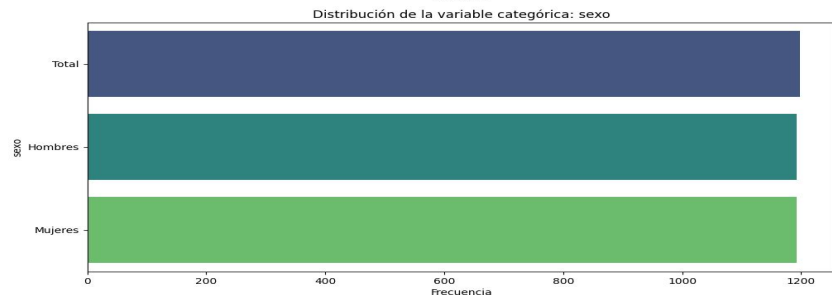
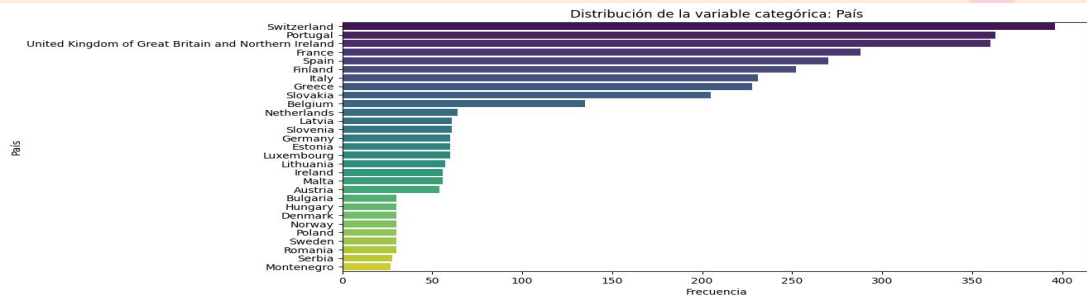
Variables tras los cambios

País	Países que están registrados con euros
Prom_ganancias	Promedio de ganancias por hora
sexo	Hombres y mujeres
Nivel_laboral	Registro de profesiones mediante ISCO-08
año	Desde 2013 a 2023
Moneda y note_source.label	Para comprobar las moneda esta medida igual y notas para saber detalles de los datos

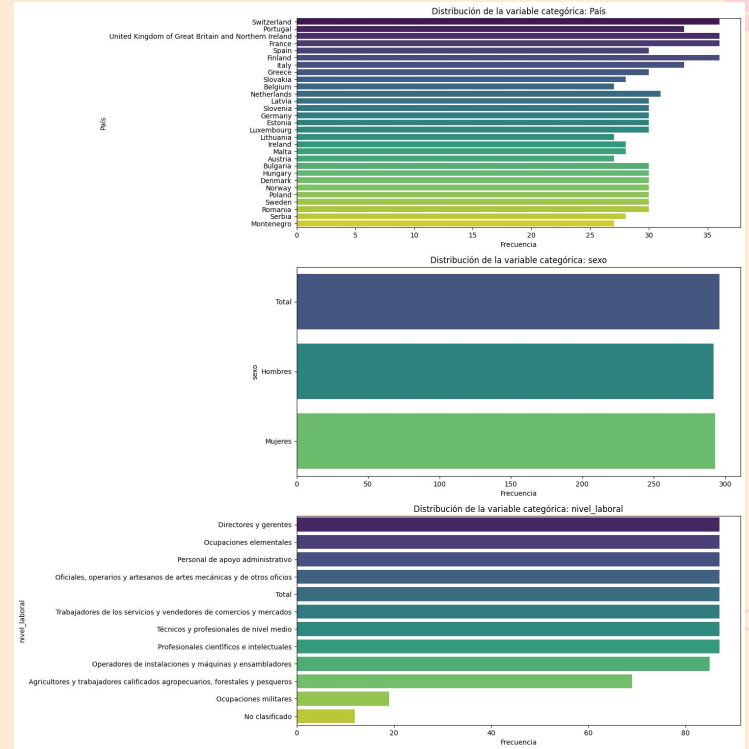
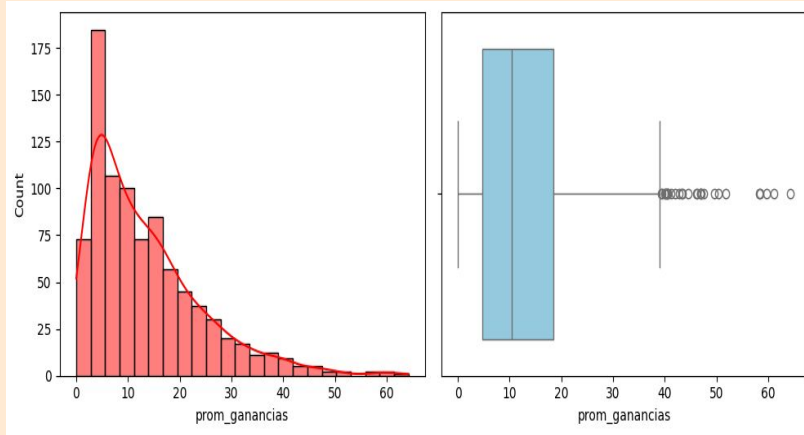
ANÁLISIS UNIVARIANTE



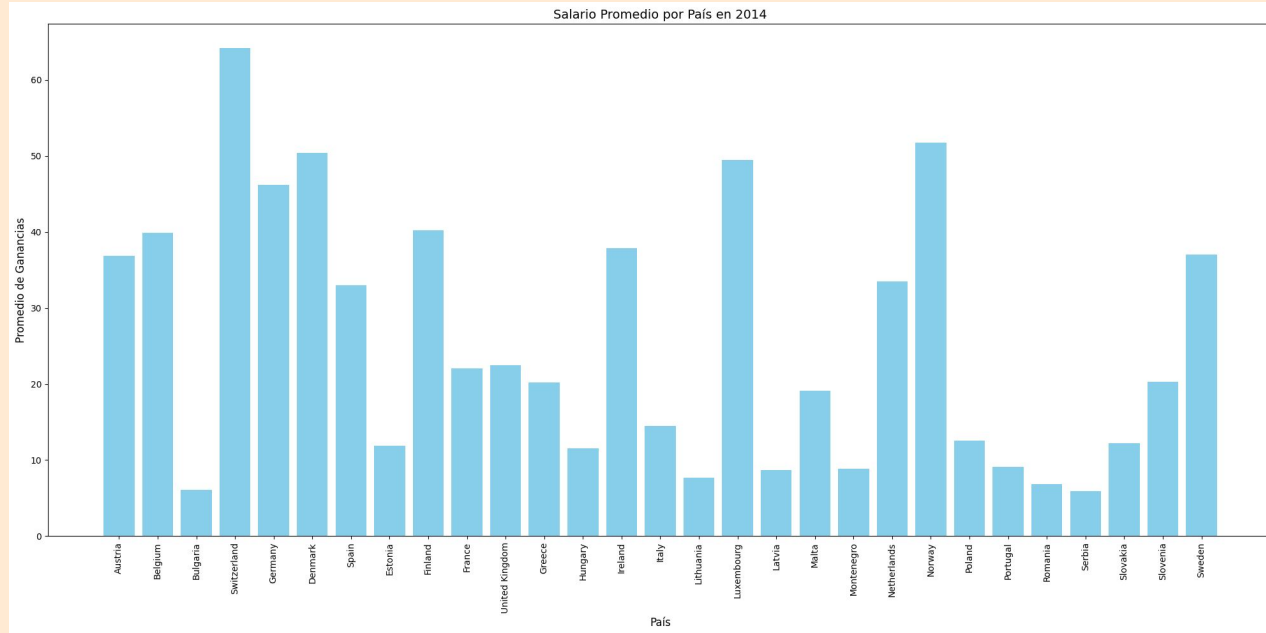
ANÁLISIS UNIVARIANTE



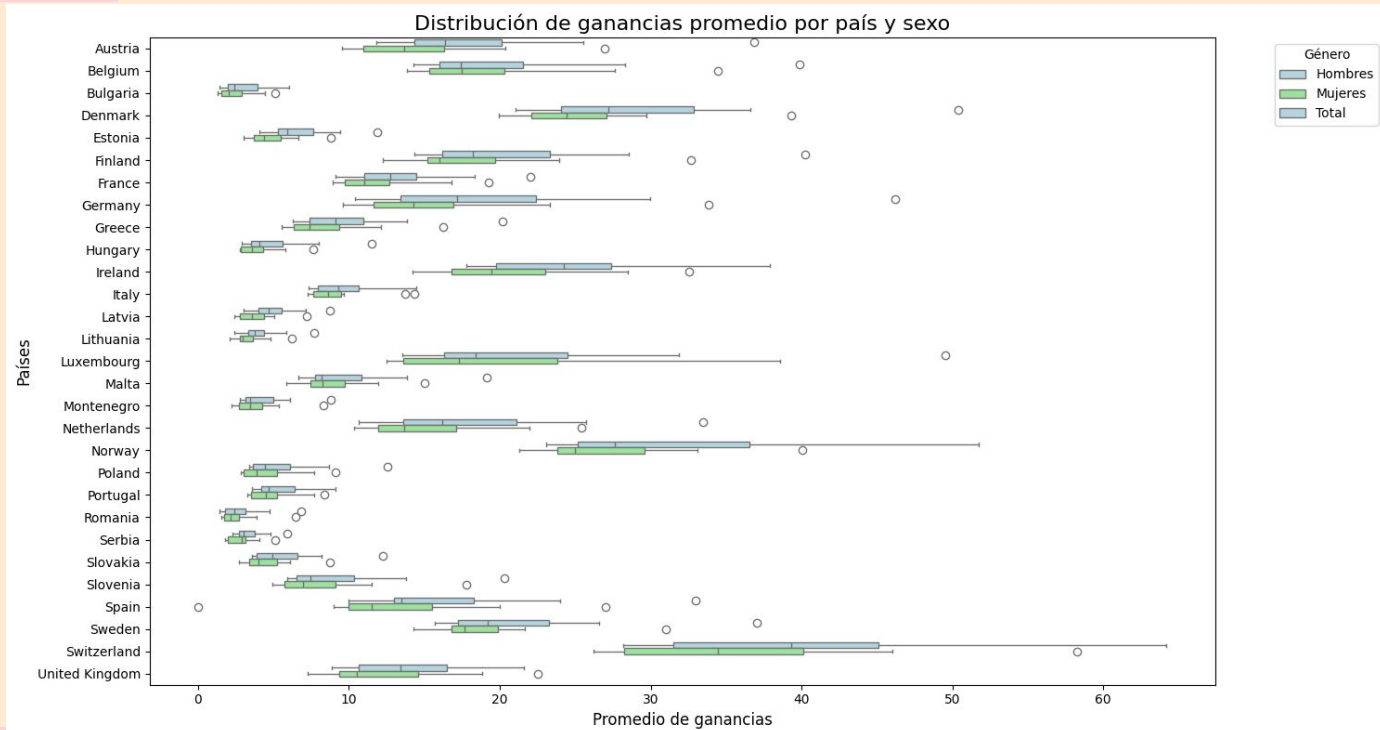
ANÁLISIS BIVARIANTE. 2014



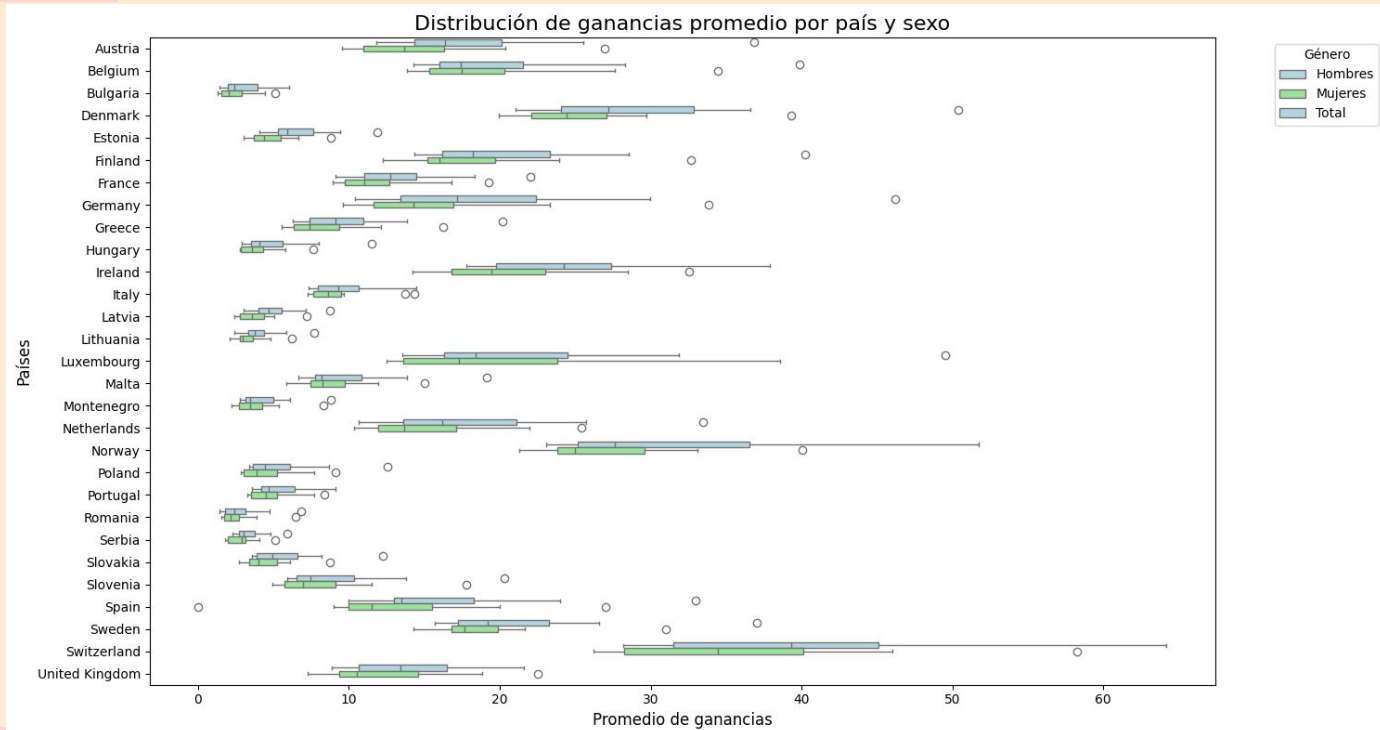
RELACIÓN SALARIO-PAÍS



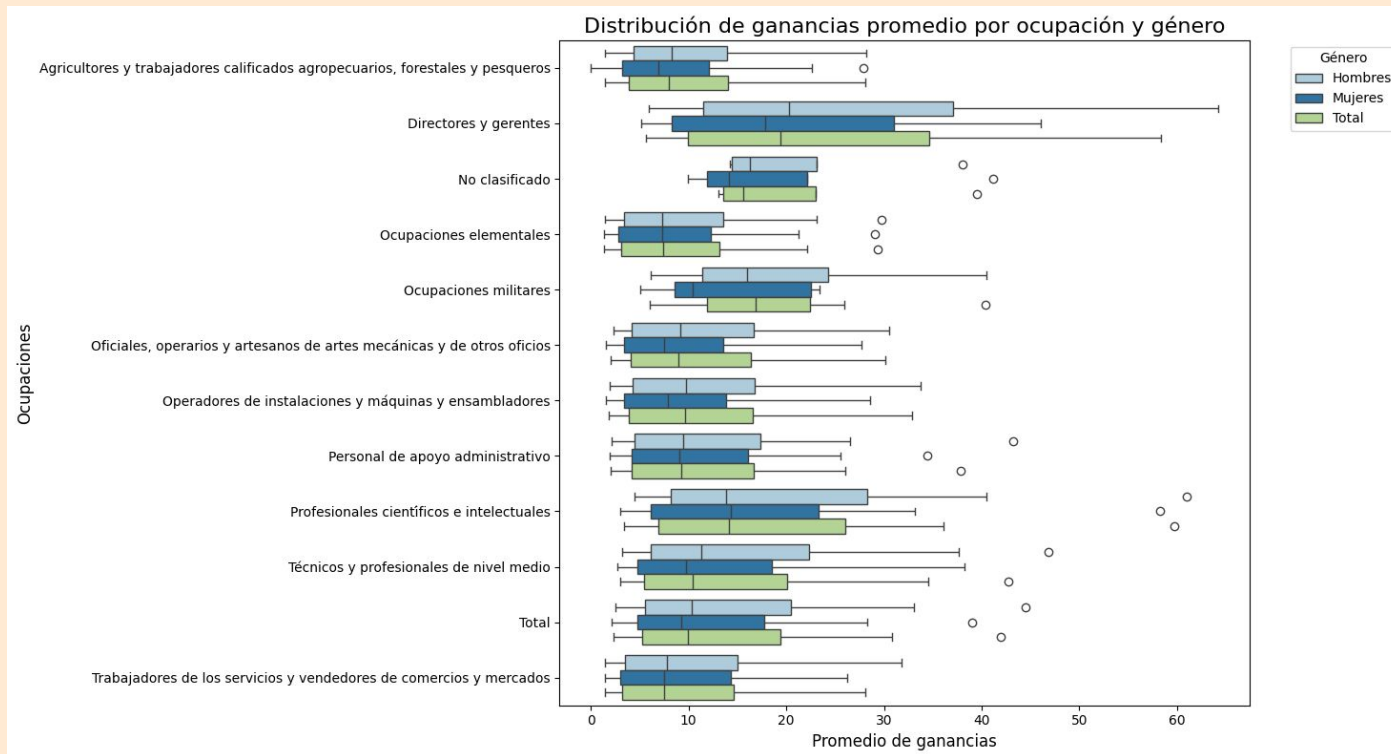
RELACIÓN GANANCIAS PROMEDIO POR PAÍS Y SEXO



RELACIÓN GANANCIAS PROMEDIO POR PAÍS Y SEXO



RELACIÓN GANANCIAS PROMEDIO POR PROFESIÓN Y SEXO



PRUEBA DE KRUSKAL-WALLIS

NIVEL LABORAL Y PROMEDIO

Estadístico H: 115.94809645599437
p-valor: 1.1811879385455551e-19
Hay diferencias significativas entre los grupos.

El resultado de estadístico H (115.94) sugiere que existe diferencias entre los grupos debido a que es un valor bastante grande.

Por otro lado el valor del P-valor es menor a 0,05 por lo que rechazamos la hipótesis nula y asumimos que existen diferencias significativas entre las diferentes ocupaciones

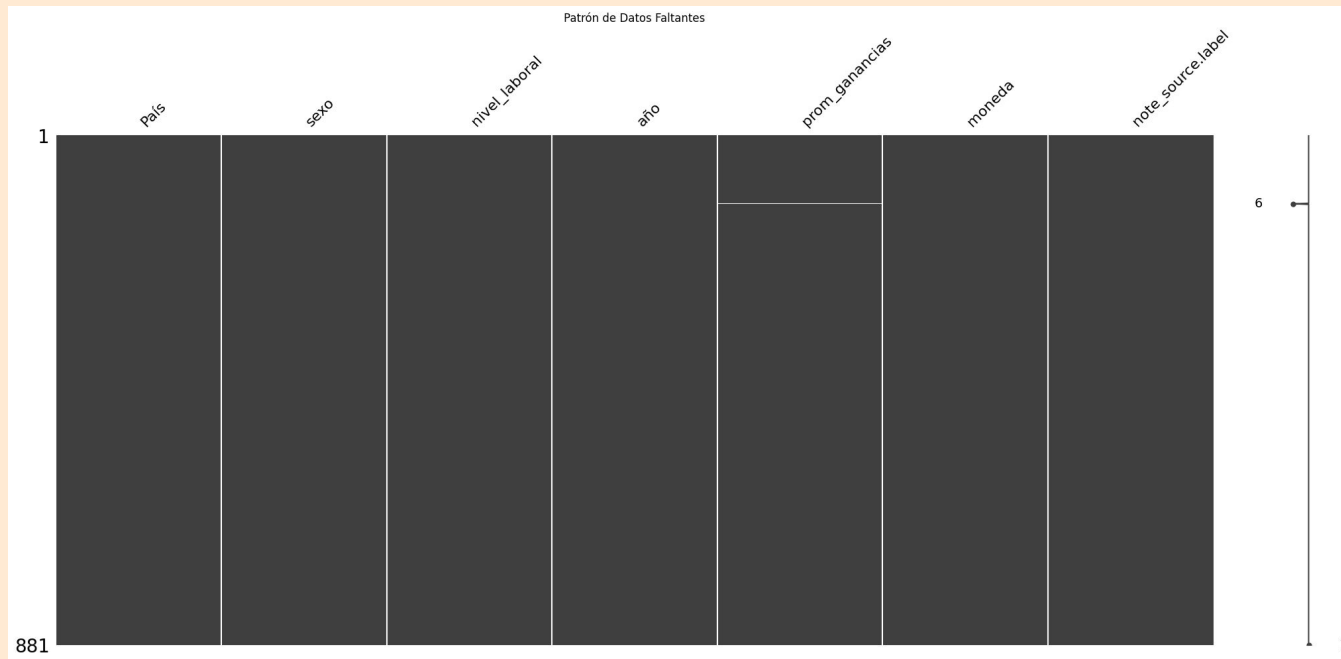
SEXO Y PROMEDIO

Estadístico H: 4.864112117283939
p-valor: 0.08785600964691005
No hay diferencias significativas entre los grupos.

El valor Estadístico (4,86) indica un valor muy bajo por lo que no se puede observar grandes diferencias.

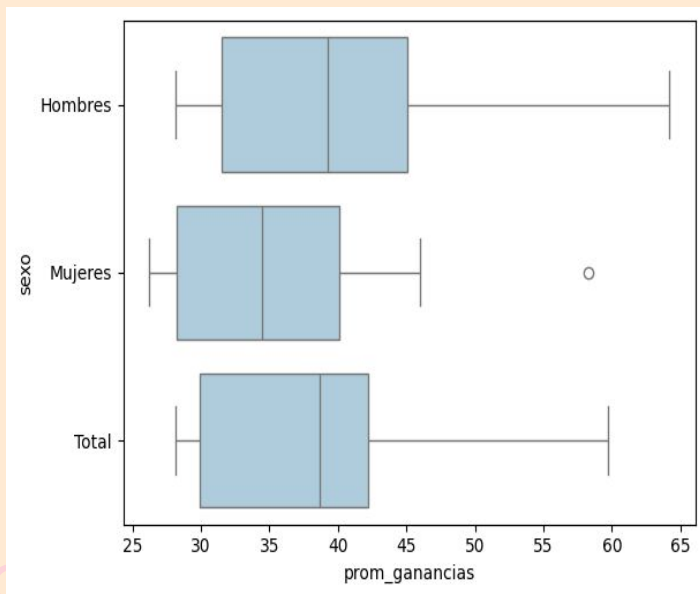
el Pvalor nos confirma que no existen diferencias estadísticamente hablando, debido a que 0,088 es mayor que 0,05 por lo que no rechazamos la hipótesis nula. Sin embargo, el valor es cercano a 0'05 por lo que podría sugerir una diferencia sutil.

TRATAMIENTO DE PERDIDOS



TRATAMIENTO DE PERDIDOS

El valor Na corresponde a las ocupaciones militares en mujeres en Suiza



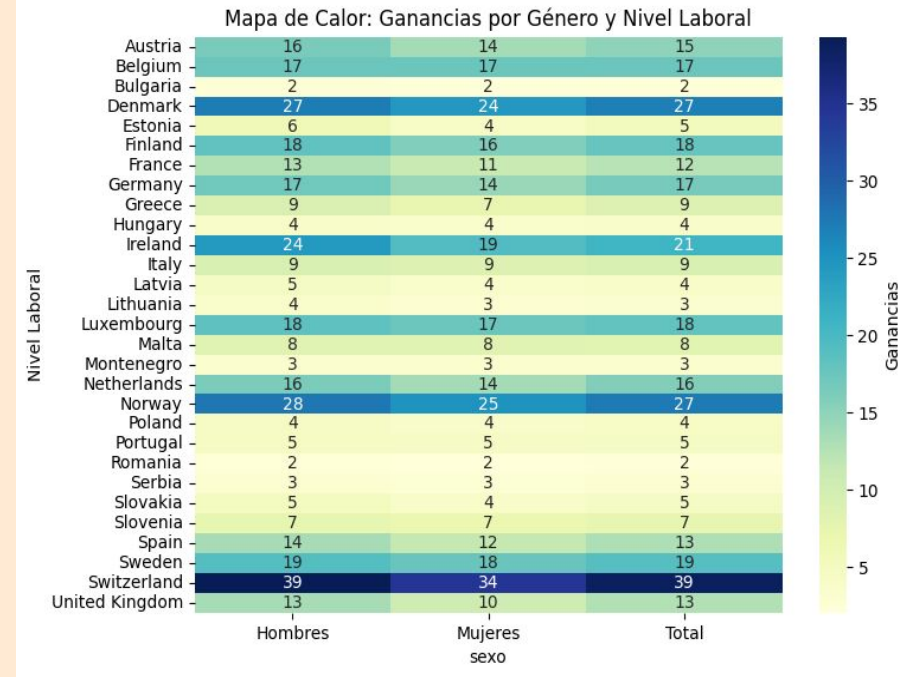
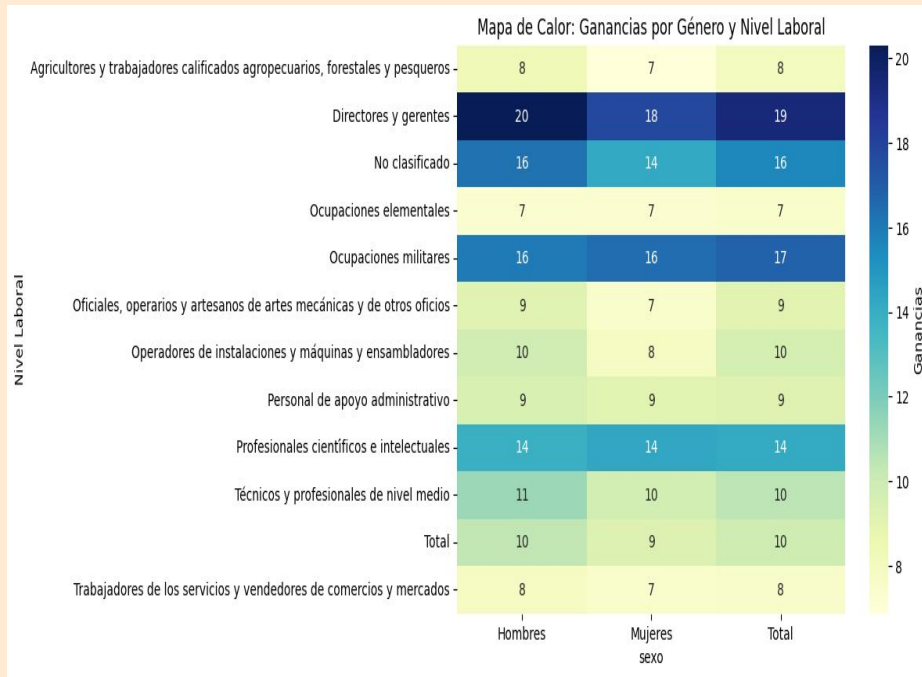
1. Este boxplot muestra si hay outliers para poder saber cual es la medida estadística más adecuada para rellenar el NA
2. Tratamiento MAR: al ser solo un dato faltante y haber tanta disparidad de promedio de ganancias según el país y el sexo he realizado la mediana, ya que presenta outliers en los datos de mujer
3. También hay presencia de valores de 0 que indican que no están incluidos la profesión de agricultores en España por lo que lo eliminaré ya que no significa que ganen 0

ANÁLISIS DESCRIPTIVO

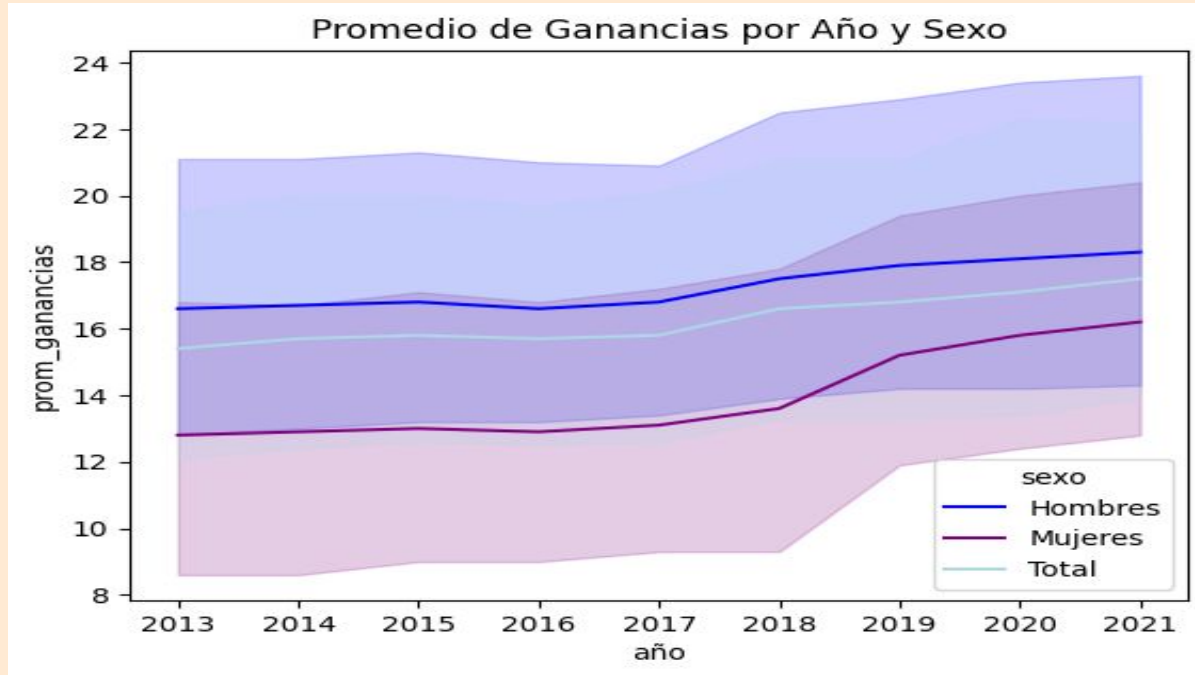
```
count      881.000000
mean       13.364915
std        10.743895
min         0.000000
25%         4.670000
50%        10.450000
75%        18.650000
max        64.190000
Name: prom_ganancias, dtype: float64
moda 13.0
curtosis 2.230668154164368
asimetria 1.4032421705281972
varianza 115.43127320297181
```

- Los registros en 2014 son de 881
- La media es de 13,36 euros por hora
- Se puede observar una desviación estándar alta de 10,74
- El mínimo es 0 y el máximo es 64,19
- Al menos el 25% gana 4,65 o menos
- Al menos el 50% gana 10,45 o menos
- Al menos el 75% gana 18,65 o menos, por lo que viendo el máximo se puede observar que hay muchos datos excepcionales

ANÁLISIS DESCRIPTIVO



SERIE TEMPORAL EN ESPAÑA



CONCLUSIONES

01 No se puede afirmar con los datos la brecha

02 Diferencia de medianas sutil

03 Directores y gerentes (otras variables)

Falta de datos representativos

04
05 Evidencia de las diferencias de ganancias entre profesiones