

Estudio estadístico de brecha Salarial

1.PRIMERA EXPLORACIÓN

Este dataSet contiene el salario promedio por hora en los empleados, además de variables como sexo, edad y ocupación. Los datos han sido recogidos en ILOSTAT (International Labour Organization), un portal estadístico que recoge los datos sociolaborales. Los salarios promedio están medidos mediante el indicador 8.5.1 (Ingresos promedio por hora de los empleados, por sexo, edad, ocupación) y se puede observar que son recogidos de diferentes fuentes estadísticas.

1. Al principio del notebook hace una presentación de la problemática y los objetivos de los estudios. A continuación, se realiza una primera visión del dataset donde se encuentra información detallada del dataframe y los nulos y se clasifica los tipos de datos y se hace un estudio de la cardinalidad.

Observando las características hay que cambiar el tipo de datos para poder manejarlos y como cambio adicional cambiar el nombre de variables para que sea más fácil identificarlas.

Se realiza un cambio a tipo categórica y datetime para la variable año.

2. Cambio de nombre de las variables para que sea más intuitiva en Castellano y escritas todas en 'snake_case'.

3. Limpieza de variables: hay varias variables que no sirven para nada debido a que no tienen importancia o gran número de nulos

(obs_satus.label, note_indicator.label, indicator.label) otras variables como note_source.label la conservaré por si da alguna información adicional del dataset.

4. Identificación de valores duplicados o errores evidentes: se realiza este estudio para observar si existen algunos problemas de imputación o valores que han sido escritos duplicados, en este caso no existen duplicados exactos, aunque si parciales, le veo sentido debido a que hay datos diferentes de los mismos trabajos y sexo.

5. Ver valores perdidos: Se puede observar 56 valores perdidos en la columna obs_Value que se tiene que ver el porqué, aquí se puede observar los países donde falta los datos, en otros faltan años. Por lo que se tendrá que decidir los años donde se obtenga más datos.

Variable Ocupación: Se ha hecho un filtrado para seleccionar que tipo de Indicador de trabajadores se iba a utilizar para el análisis. Estaba compuesto por dos índices ISCO-08 y ISCO-88. Después de documentarme en la plataforma de datos estadísticos de donde provienen los datos, he decidido elegir ISCO-08 debido a que es la más reciente. Por otro lado, se cambiará las etiquetas para que sea más sencillo a la hora de graficar

- También se realiza un cambio en los nombres para que sea más claros en los gráficos, los nombres son la traducción que proporciona la documentación de ISCO-08

Variable País: Como hay diferentes monedas y es imposible comparar por el diferente valor que tiene cada moneda voy a realizar un filtrado con los países que tienen euros.

Se ha elegido los datos de los países que tienen euro para poder hacer el análisis en una misma moneda, por otro lado, debido a que cada país tiene los datos años distintos se hará un análisis macro del año 2014 o 2018 que tiene la mayoría de países y la serie temporal de España. Finalmente se guarda los datos procesados.

2.EDA (INFORME TÉCNICO)

Se importa los datos y las librerías. Realizo el cambio de variables debido que no se cambian correctamente los tipos de datos.

1.Análisis Univariante:

Estudio descriptivo de las variables numéricas acompañada de histogramas y boxplots. Tras el estudio de los años decido realizarlo del año 2014 para todos los países debido a que es el que proporciona más datos dentro de los años que nos proporciona.

También se observa la cantidad Outliers en el promedio de ganancias, por lo que existen grandes diferencias entre países y diferentes trabajos.

Por otro lado, para acompañar el gráfico se realiza la prueba de normalidad para verificar estadísticamente que no sigue una distribución normal.

En segundo lugar, se hace un estudio de las categóricas con gráficos de barras y estudio de frecuencias. Me parecía importante saber la proporción de hombres y mujeres que había y las profesiones para intentar no cometer sesgos a la hora del análisis.

Por otro lado, volví hacer el análisis de cardinalidad con los datos transformados y con las variables definitiva, debido a la limpieza y selección de datos.

Análisis Bivariante: En primer lugar, hago un filtrado del año 2014 y lo guardo en otra variable con los datos únicamente de este año. Veo las características generales de los datos seleccionados y la cantidad.

Hago unos gráficos de las variables numéricas (histograma y boxplot para ver la distribución en el año 2014). Lo mismo hago para las categóricas para ver las proporciones de los diferentes tipos de trabajos y de mujeres por si existía alguna diferencia tenerlo en cuenta a la hora del análisis.

- Relación Salario País: Realizo un gráfico de barras para observar primero las diferencias generales de los países y sobre todo cuales son aquellos que tienen salarios más altos y bajos.
- Relación ganancias promedio por país y sexo:
 - Boxplot: Con este gráfico quería principalmente la diferencia de medianas entre hombre y mujer y las diferencias de los países, mediante el análisis se puede observar patrones en los tipos de países, sin embargo, tengo en cuenta que la diferencia de niveles de trabajo afecta en dicha diferencia por lo que analizaré más a fondo.

- Barplot: ver de una manera gráfica las diferencias más claras sin conocimientos estadísticos.
- También he realizado un mapa para ver de forma más visual las diferencias entre los países de salario.
- Relación Salario- Profesión Género:
 - Primero observo la proporción de mujeres y hombres por profesión y observo que no hay grandes diferencias entre ellos, si hubiese grandes diferencias podría sesgar los resultados entre los grupos, es decir, si solo hay 3 mujeres profesionales en comparación a 50 hombres, no sería representativa la diferencia.
 - Boxplot: Para observar las medianas de las profesiones o existencia de outliers, sin embargo, teniendo en cuenta las diferencias entre países sería interesante realizarla de un país en concreto y ver la diferencia dentro de los patrones reconocidos en los países (más ricos- más brecha)
- Comparar medianas: Prueba de Kruskal-Wallis:
- He elegido esta prueba debido a que no sigue una distribución normal por lo que se necesita una medida no paramétrica para comparar la mediana de dos grupos, en este caso, nivel laboral y sexo. Se agrupa los datos por nivel laboral y sexo mediante un groupby y se realiza la prueba de kruskal.

Por último, como análisis bivariado he querido hacer uno más específico de uno de los países que más diferencia se observaba y menos diferencias, sin embargo, hay un sesgo representativo ya que solo había un representante por categoría, por lo que no se puede llegar a conclusiones representativas por falta de datos.

3. TRATAMIENTO DE PERDIDOS

Primero observo mediante `isnull()` la cantidad de perdidos y donde, en este caso en el promedio de ganancias. Posteriormente lo guardo en una variable para poder observar exactamente cuál es en el dataframe, es el caso de las mujeres en Suiza en las ocupaciones militares. He realizado un gráfico también para mostrar donde se encuentra el dato NA.

Por otro lado, he realizado una caja de bigotes para observar si había Outliers en las mujeres para poder realizar la mejor imputación, además los datos no siguen una distribución normal, por lo que se aplicará la imputación MAR mediante la mediana, debido a la disparidad de los datos y los outliers. Sin embargo, al ser un caso que se ven diferencias de medianas por sexo y país he realizado la imputación del NA por la mediana de las mujeres suizas.

Este Na era solo en el año 2014, sin embargo, en el dataframe original investigue el resto de los perdidos para poder observar algún patrón y se puede observar que la

mayoría es en ocupaciones militares en Suiza, indague y descubrí que al ser un país con servicio militar obligatorio en los hombres, el número de mujeres militares se reduce, otros países como Gran Bretaña también falta datos en ocupaciones militares en mujeres, no he realizado ningún tratamiento ya que no era necesario para la investigación. Más adelante, para realizar la serie temporal de España se eliminará la profesión de agricultores debido a que como indica el source.label los datos excluyen esta profesión por lo que no se puede tener información de este sector y no significa que ganen 0 euros, por lo que pueden sesgar los resultados.

4.ANÁLISIS DESCRIPTIVO

He realizado un análisis descriptivo del año 2014 para poder observar las estadísticas más concretamente.

Después he creado unas visualizaciones con mapas de calor donde se puede observar las diferencias de medianas de los diferentes sectores laborales y por otro lado de países.

Para ver de manera visual las medianas y países y sectores que más ganan.

Por otro lado, he realizado una serie temporal de España, he elegido España porque tenía una gran cantidad de años y así poder conocer una situación que estamos todos familiarizados, y como explique en el punto anterior elimine la categoría agricultura.

ANÁLISIS DE GRÁFICOS. Y ESTADÍSTICAS.

1.Numericas

año

count 3582.000000

mean 2016.983529

std 2.822894

min 2013.000000

25% 2014.000000

50% 2017.000000

75% 2019.000000

max 2023.000000

Name: año, dtype: float64

moda 2014

curtosis -0.9742304455179438

asimetria 0.37949803769577495

varianza 7.968731694090547

=====

prom_ganancias

count 3563.000000

mean 15.401392

std 11.931298

min 0.000000
25% 7.000000
50% 12.050000
75% 19.125000
max 81.770000
Name: prom_ganancias, dtype: float64
moda 13.0
curtosis 3.5217319023695755
asimetria 1.7277104858483137
varianza 142.3558781850805

Análisis de los resultados de las estadísticas descriptivas

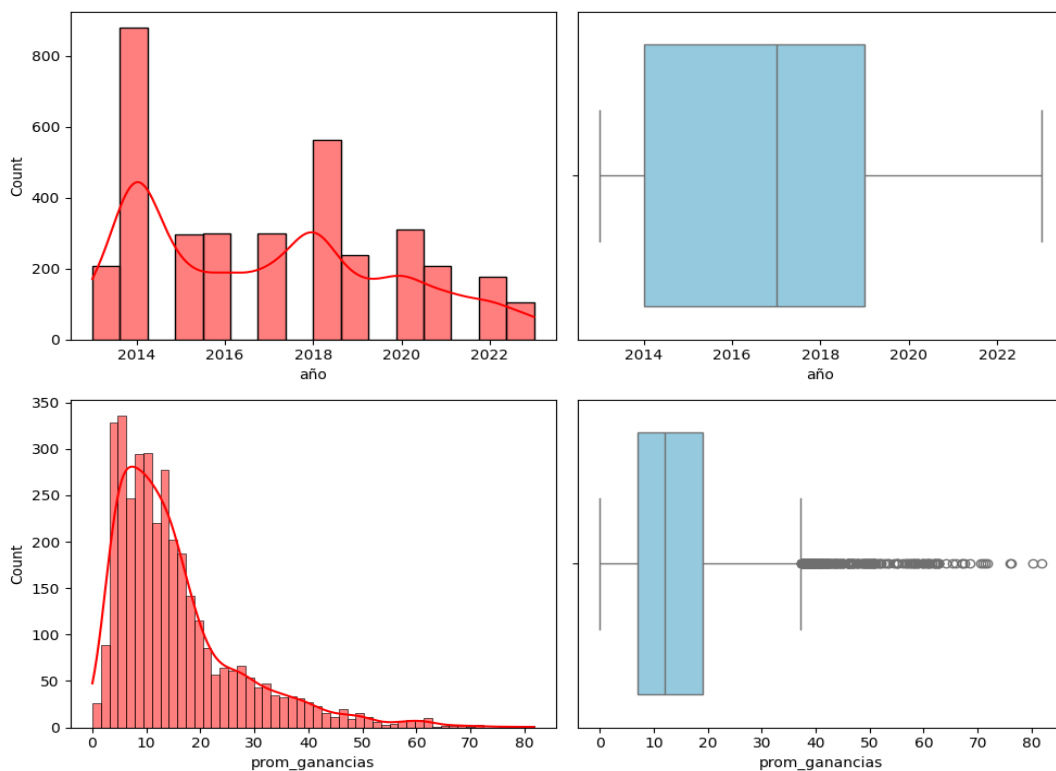
1. Años:

- Hay un registro de 3582, es decir, es la cantidad de años proporcionados en el dataset.
- La media es de 2016,98, esto quiere decir que el año promedio este alrededor del año 2017
- La desviación típica, que nos muestra la dispersión de los datos, en este caso, 2,82 años, por lo que sugiere una dispersión de alrededor de 3 años entre los dato.
- Los años se encuentran reunidos entre el año 2014 y 2019, más adelante con el gráfico de boxplot se verá con más claridad.
- El año mínimo es el 2013 y el máximo el 2023
- El año que más se repite es el 2014
- La curtosis con un -0,97, nos indica que la distribución de los años es plana y no muestra grandes picos. Debido a que es menor a 3 es platicúrtica
- Asimetría: La asimetría de 1.72 significa que la mayoría de los registros están en años más antiguos, pero hay algunos años recientes (como 2023) con menos datos que son mucho más altos y están alejados del promedio, creando un sesgo hacia la derecha.
- La varianza: mide cuanto de dispersos estas los datos respecto a la media, en este caso es de 7,97, lo que nos indica que los datos están bastante agrupados alrededor de la media.

2. Promedio de ganancias:

- Hay un registro de 3563, es decir, es la cantidad de datos de las ganancias proporcionados en el dataset
- La media es de 15.40 euros por hora.
- La desviación estándar es de 11.93, por lo que nos muestra una dispersión alta entre los datos.
- El mínimo es de 0 y el máximo de 81,77

- La distribución de las ganancias nos muestra que menos del 25% de la población gana menos de 7 euros por hora, la mitad de las personas ganan menos de 12,05 y el 75% gana menos de 19,03
- La curtosis con 3,52 nos indica una distribución leptocurtica ya que es mayor que 3, o que quiere decir que tiene colas más pesadas y un pico más alto que una distribución normal, lo que indica que hay valores extremos.
- La Varianza con 142,36, muestra una gran dispersión entre los datos y que hay valores muy extremos



Análisis de los gráficos.

1. Años:

- Histograma: El histograma muestra la distribución de los años, se puede observar la moda en 2014 y casi podríamos decir que es bimodal debido al pico del 2018. Se puede observar que la mayoría de los datos se encuentra en los años más antiguos, viendo un descenso en los años como 2022 o 2023, una hipótesis de porque falten puede ser debido a la falta de datos después de la pandemia.
- Boxplot: Los años van del 2013 al 2023. La mediana, que es el centro del boxplot es el año 2017. El 25 % de los datos está en 2014 y el 75 % están por debajo del año 2019. No hay outliers y se muestra datos bastantes simétricos alrededor de la mediana.

2. Promedio de ganancias:

- Histograma:

1. La mayoría de los individuos ganan entre 0 y 20, donde se encuentra la mayor concentración de la curva
2. A medida que aumenta el promedio del salario disminuye la cantidad de individuos, creando una cola larga hacia la derecha.
3. La línea nos indica que el promedio de ganancias tiene una asimetría positiva, también se puede ver mediante la media y la moda, ya que la media es mayor que la mediana.
 - Boxplot: Lo que más destaca de la distribución de las ganancias es la gran cantidad de outliers, por lo que hay individuos que gana una gran cantidad de dinero en comparación con la mayoría. La mayoría de las ganancias se encuentra entre 7 y 19 y la mediana, como se podía observar en el análisis descriptivo.

Resumen

- El análisis para observar los países se realizará en el año 2014 debido a que es el año que más datos se tiene
- Hay una gran cantidad de Outliers, esto significa que hay una gran disparidad en el promedio de ganancias entre los individuos.

Prueba de normalidad

1. Promedio de ganancias

Estadístico de prueba: 1186.0341605460699

P-valor: 2.8572902750320396e-258

Como el P- valor está muy por debajo de 0,05, se rechaza la hipótesis nula y muestra que los datos no muestran una distribución normal, aunque ya se podía confirmar mediante los datos de curtosis y gráficos.

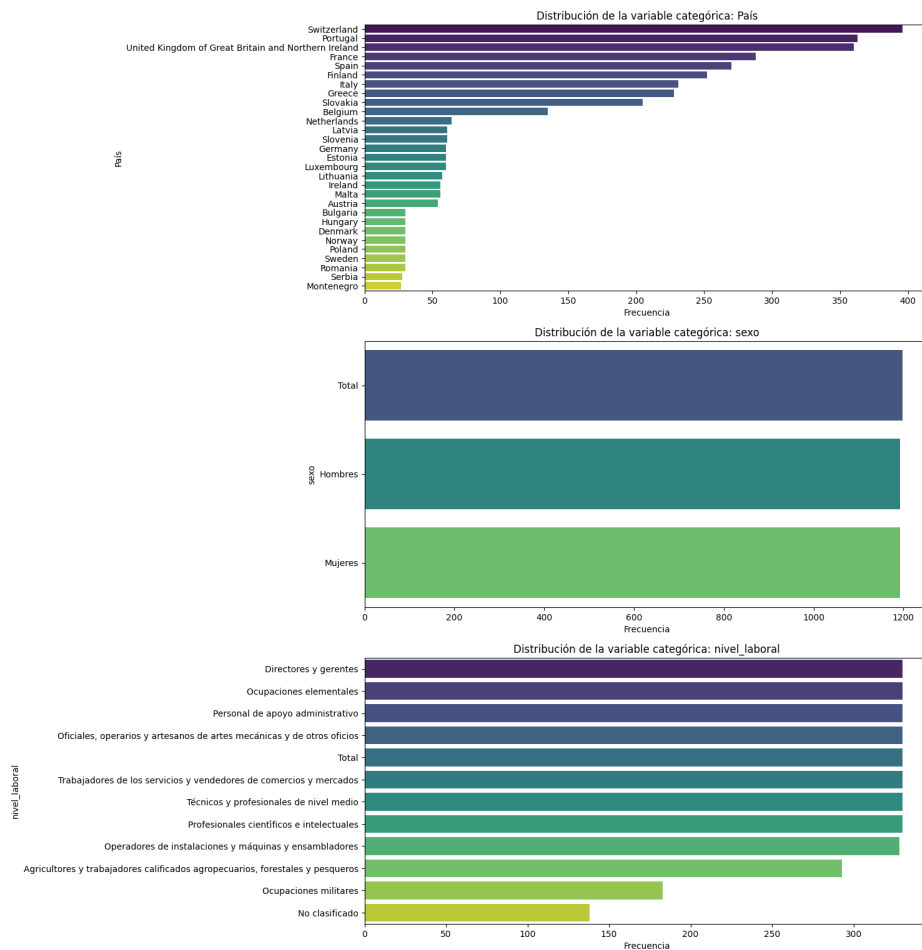
2. Año

Estadístico de prueba: 786.0081652179827

P-valor: 2.0916815105262303e-171

Como el P- valor está muy por debajo de 0,05, se rechaza la hipótesis nula y muestra que los datos no muestran una distribución normal, aunque ya se podía confirmar mediante los datos de curtosis y gráficos.

CATEGÓRICAS

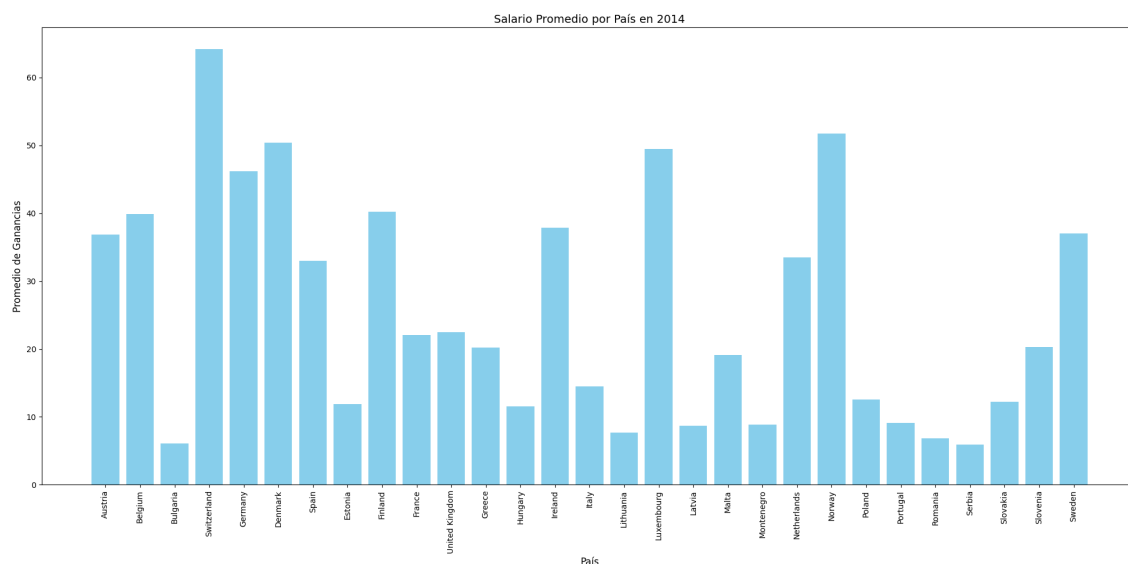


Con estos gráficos se puede observar las frecuencias de las variables categóricas que tenemos en el dataset. Algunas nos proporcionan más información que otras, por ejemplo, la de moneda a la hora de la investigación no nos aporta gran información.

- El primer gráfico nos muestra cuales son los países que hay más datos, como es el caso de Switzerland, Portugal o Gran Bretaña.
- El segundo gráfico muestra la proporción de hombres y mujeres que hay en el dataset, esto si nos da una información relevante debido a que es importante al comparar que las proporciones de hombres y mujeres sean las mismas.
- El tercer gráfico nos muestra la distribución de las diferentes profesiones y como están repartidas, la mayoría tienen la misma proporción menos Agricultores y trabajadores calificados agropecuarios, forestales y pesqueros, ocupaciones no militares y trabajos no clasificados.

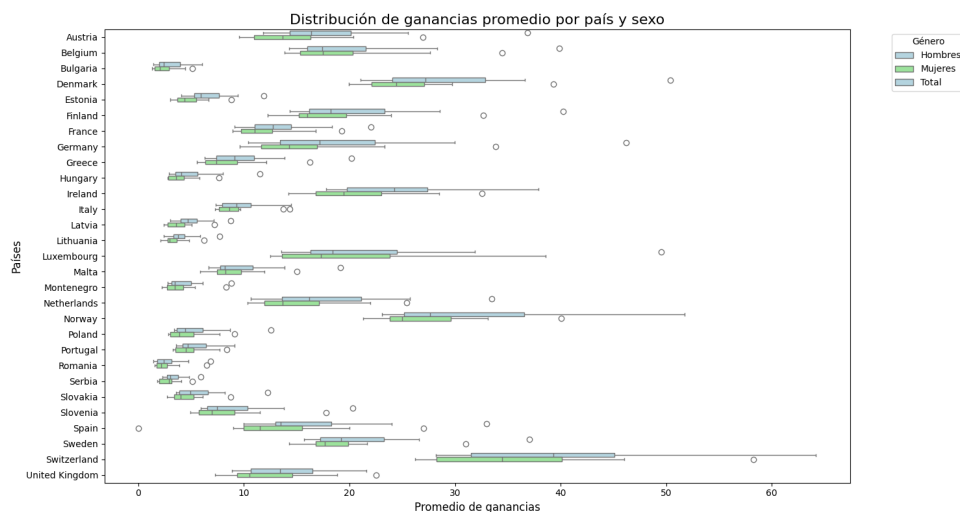
ANALISIS 2014

Relación Salario País



En este gráfico se muestra la relación del promedio de ganancias por hora en el año 2014 por país, donde se puede observar la diversidad de diferencias salariales sin tener en cuenta el género. Dentro de Europa existe grandes disparidades entre los países, por ejemplo, el país que más dinero gana por hora es Suiza cerca de 40 euros la hora o Noruega con 30 euros la hora, sin embargo, hay otros países como Rumania, Serbia que no llegan los 5 euros por hora.

Relación ganancias promedio por país y sexo



En este gráfico de boxplot nos da bastante información sobre la diferencia entre hombres y mujeres en los países de la Unión Europea en el año 2014, sobre todo fijándonos en las medianas. Haciendo una vista general, se puede observar que todos los países muestran una desigualdad entre hombres y mujeres, ya que los bigotes son mayores y las medianas en todos los países son mayores en hombres que en mujeres.

1. Países con salarios moderadamente altos

- Austria: La mediana en los hombres coincide con el 75% del salario promedio de las mujeres, esto quiere decir que al menos el 75% de las mujeres en Austria ganan lo mismo que al menos el 50% de los hombres, por lo cual la mayoría de los hombres ganan mucho más dinero que las mujeres. Además, hay presencia de Outliers en ambos casos, es decir, casos excepcionales que ganan más, sin embargo, el valor de la mujer está bastante por debajo de outliers del hombre.
- Bélgica: Las medianas de los hombres y las mujeres es prácticamente igual, por lo que podemos decir que este país no presenta una brecha salarial muy marcada. Si que muestra diferencias significativas en los Outliers y que los extremos son un poco más bajos en las mujeres.
- Dinamarca: Ocurre similar que, en Austria, el percentil 25% de los hombres coincide con la mediana de las mujeres, esto quiere decir que las mujeres están por debajo que las mujeres.
- Alemania: La extensión del boxplot en hombres está movido hacia la derecha, es decir, además coincide la mediana de los hombres con el percentil 75% de las mujeres. Cabe destacar que existe un outliers en las mujeres, por encima de lo que gana los hombres.

2. Países con Salarios más Altos:

- Suiza: Es el país con los salarios más altos y presenta aparentemente una brecha salarial muy marcada. La mediana de los hombres es bastante cercana al percentil 75% de las mujeres, por lo que el 75% de las mujeres ganan menos que el 50 % de los hombres. Los extremos del boxplot en los hombres muestran que hay gran disparidad entre los salarios promedios, el de las mujeres es más corto.
- Noruega: Ocurre algo muy similar a Suiza, aunque las diferencias son menos pronunciadas que en Suiza
- Luxemburgo: Las diferencias entre las medianas no es tan grande como en otros casos, además en el caso de las mujeres existen outliers muy altos, alrededor de 50, sin embargo, el primer cuartil, la mediana y el tercer cuartil son menos que en los hombres, a pesar de este dato excepcional.
- Irlanda: Muestra una gran diferencia entre las medianas en los diferentes sexos, coincidiendo la mediana de las mujeres con el primer cuartil de los hombres, mostrando el desplazamiento a la derecha, y, en consecuencia, los salarios más altos de los hombres. Las mujeres muestran un Outliers, sin embargo, coincide con el extremo de los hombres.
- Finlandia y Suecia: Los países nórdicos muestran datos similares, ya que la mediana de los hombres es bastante superior que el de las mujeres, coincidiendo con el primer cuartil, sin embargo, los salarios en Suecia se encuentran más concentrados que los de Finlandia.

3. Países con salarios bajos:

- Portugal, Polonia, Malta, Bulgaria, Lituania, Montenegro, Rumanía, Serbia, Eslovaquia y Eslovenia: Son países que tienen unos salarios más bajos, sin embargo, las medianas de entre hombres y mujeres son muy similares. También cabe destacar que los salarios promedio están muy concentrados, es decir, el rango intercuartílico es muy pequeño, por lo que los salarios son similares. Aun así, se puede observar que los hombres ganan más y la presencia de outliers.
- España, Francia, Gran Bretaña, Grecia, Italia, Estonia, Letonia: Estos países la mayoría tienen unos salarios promedio un poco mayores, sin embargo, se muestra más diferencia entre las medianas de género. La extensión del boxplot es mayor por lo que existe una mayor variabilidad entre los salarios. La mediana de las mujeres se acerca al tercer cuartil de los hombres

RESUMEN:

Las mayores disparidades salariales entre géneros se observan en:

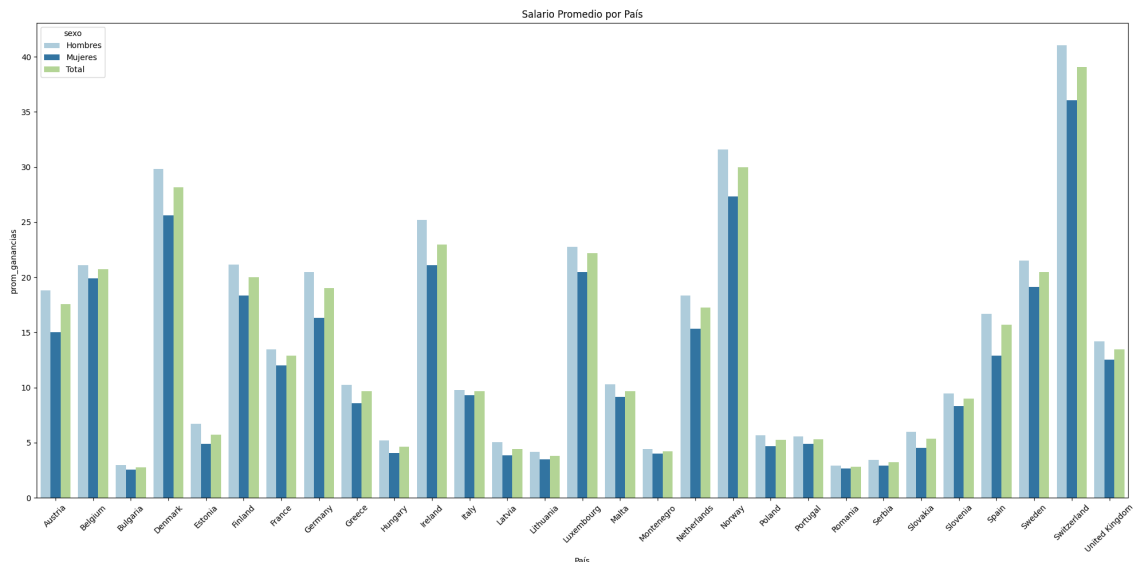
1. Suiza
2. Noruega
3. Luxemburgo
4. Irlanda
5. Países Bajos

Las menores brechas salariales (aunque con salarios más bajos) aparecen en:

1. Bulgaria
2. Rumania
3. Montenegro
4. Estonia
5. Lituania

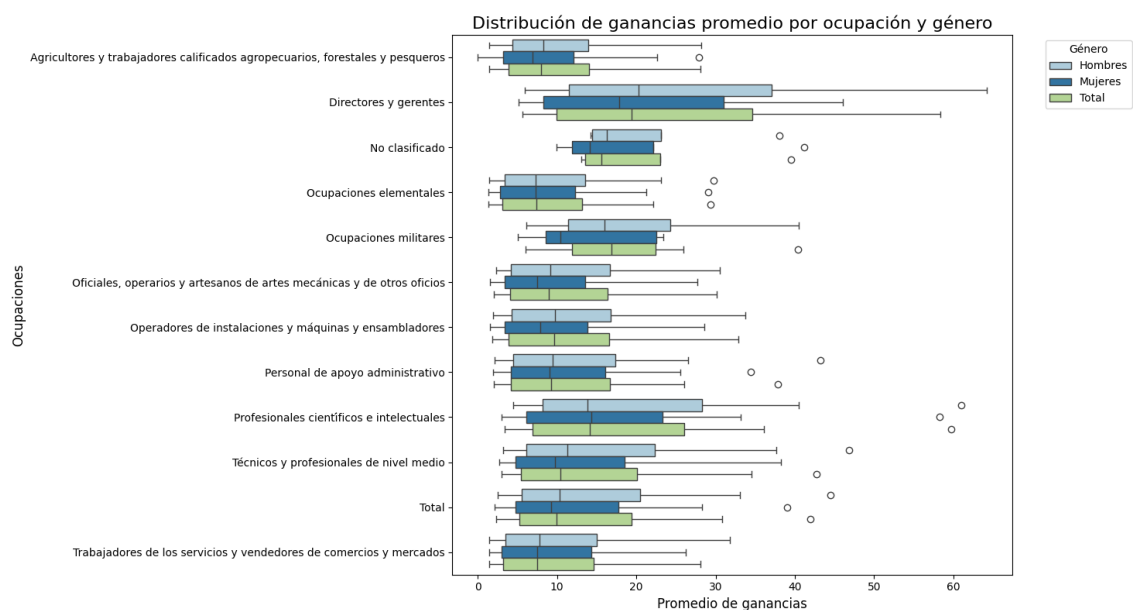
Los países que más se gana existe una mayor brecha salarial aparentemente, sin embargo, los países con salarios más bajos son más igualitarios, los países con salarios intermedios si son más igualitarios, pero con una tendencia a la brecha salarial.

Relación Salario-País-Género



Este gráfico de barras en la que nos muestra el promedio de ganancias por países podemos observar que en el caso de las mujeres se gana menos dinero por hora, en este primer gráfico, se puede ver que los países con salarios más bajos son más igualitarios o el caso de Bélgica. Suiza siendo el país que más gana y parece que más brecha de género que existe, sin embargo, Rumanía parece el más igualitario y con salarios más bajos.

Relación Salario Profesión Género



Este gráfico boxplot por grupos nos muestra la distribución de las ganancias promedio clasificado por profesión, en comparación con los países, las diferencias son menores.

- Directores y gerentes: Los hombres tienen una mediana más alta que las mujeres, indicando una disparidad en los salarios centrales. Por otro lado, los hombres tienen mayor IQR, lo que significa que las ganancias son más

variables, sin embargo, las mujeres presentan un rango más comprimido, reflejando menor variabilidad.

- Profesionales científicos e intelectuales: La mediana de las mujeres está un poco por encima de los hombres, por lo que la brecha salarial entre profesionales y científicos es menor que en directores y gerentes. Cabe destacar que los hombres tienen un rango intercuartílico más amplio, menor que las mujeres, además se encuentra desplazada a la derecha. Tiene algunos outliers superiores (particularmente en hombres).
- Técnicos y profesionales de nivel medio: La mediana de los hombres es mayor que la de las mujeres, aunque las diferencias no son muy destacables, tienen unos rangos muy parecidos por lo que la variabilidad en las ganancias es similar. Existen algunos outliers en hombres y mujeres, pero no tan pronunciados como en los profesionales científicos e intelectuales.
- Personal de apoyo administrativo: Las ganancias centrales de hombres y mujeres son similares, reflejando una menor disparidad de género. Ambos géneros tienen rangos comprimidos, indicando menos variabilidad en las ganancias. Existen algunos outliers en ambos géneros, pero no muy extremos.
- Trabajadores de los servicios y vendedores de comercios y mercados: Los hombres tienen una mediana más alta, pero la diferencia con las mujeres es pequeña. Hombres y mujeres tienen IQR similares, reflejando rangos de ganancias similares, sin embargo, la disparidad de los hombres es mayor.
- Oficiales, Operarios y artesanos de artes mecánicas y de otros oficios: La mediana de los hombres es considerablemente mayor que la de las mujeres. Además, el IQR de los hombres es amplio, lo que sugiere variabilidad en los salarios en comparación con el de las mujeres, ya que su rango es mucho más corto.
- Operadores de instalaciones y máquinas y ensambladores: Los hombres tienen una mediana más alta que las mujeres, ambos géneros tienen IQR similares, pero el rango de los hombres es un poco más amplio.
- Ocupaciones elementales: Ambos sexos tienen una mediana muy similar, por otro lado, tienen rangos comprimidos, lo que existe poca variabilidad de ganancias, sin embargo, el de los hombres está ligeramente desplazado a las derechas. Se puede observar que existen algunos Outliers.
- Agricultores y trabajadores calificados agropecuarios, forestales y pesqueros: Los hombres tienen una mediana más alta que las mujeres, reflejando una notable disparidad, además existe una mayor dispersión en los hombres, indicando más variabilidad en los salarios. El rango de las mujeres es más

comprimido, por lo que la variabilidad de salarios es menor, por otro lado, presenta outliers que se asemeja al extremo de boxplot de hombres.

- Ocupaciones militares: La mediana de los hombres es mucho mayor a la de las mujeres, por otro lado, las mujeres muestran un rango más comprimido que el de los hombres, por lo que la variabilidad de ganancias en los hombres es mucho mayor que en el de las mujeres, lo que podría indicar que mayoritariamente solo llegan hombres a los altos rangos militares.
- No Clasificado: Los hombres tienen una mediana más alta que las mujeres, sin embargo, ambos sexos tienen un rango muy comprimido, coincidiendo su tercer cuartil con el extremo del boxplot. El boxplot de los hombres se encuentra ligeramente desplazado a la derecha.

Conclusión estadística general:

Mayoritariamente se puede observar una diferencia en los salarios de los hombres y las mujeres, donde el salario de los hombres es ligeramente superior, no existe unas evidencias tan claras como en la comparación entre países. Si que existe diferencias más evidentes en profesiones como directores, esto se puede deber que por problemas de conciliación la mujer tiene menos acceso a estos puestos. También existe una diferencia más evidente en ocupaciones militares, puede ser influido por razones históricas y culturales o los agricultores. El resto de las profesiones existe una brecha salarial menor.

Comparar medianas. Prueba de Kruskal-Wallis

Nivel Laboral:

Estadístico H: 115.94809645599437

p-valor: 1.181187938545551e-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:

Estadístico H: 5.187845270254686

p-valor: 0.07472634025606531

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.

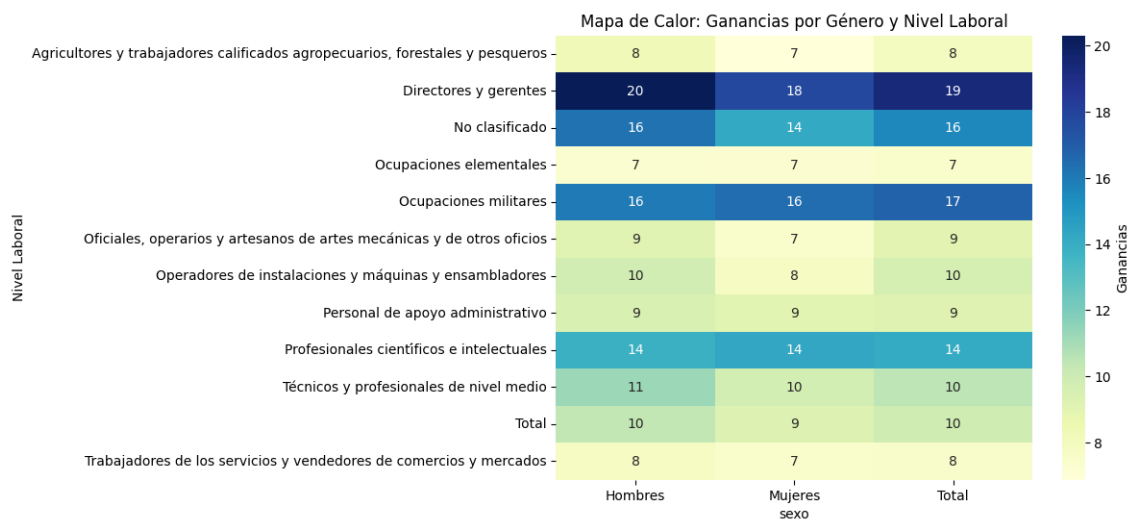
el Valor nos confirma que no existen diferencias estadísticamente hablado, debido a que 0,088 es mayor que 0,05 por lo que no rechazamos la hipótesis nula. Sin embargo, el valor es cercano a 0'05 por lo que podría sugerir una diferencia sutil.

```

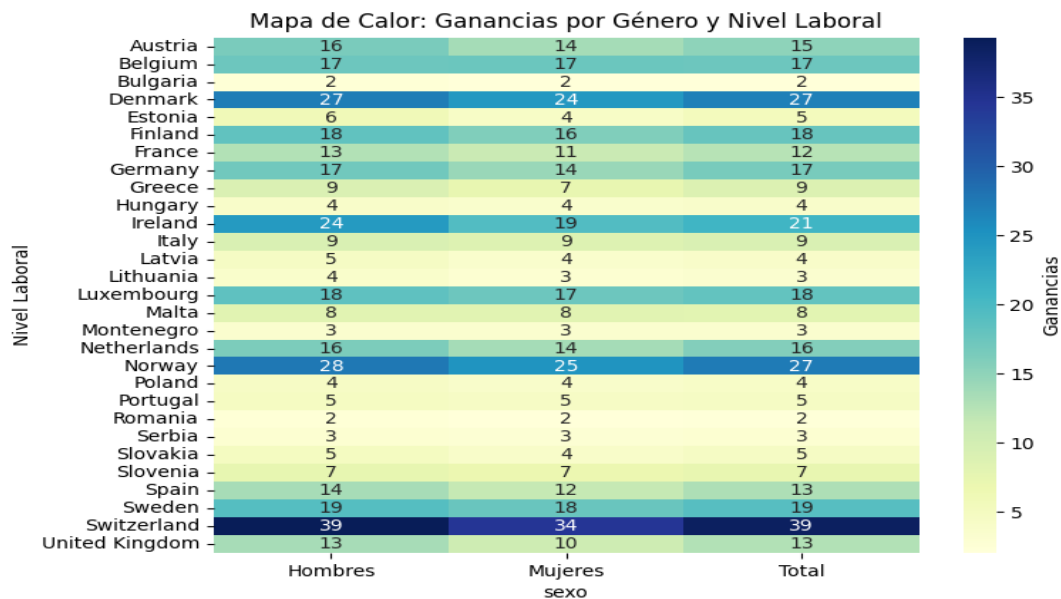
prom_ganancias
count    881.000000
mean      13.364915
std       10.743895
min        0.000000
25%        4.670000
50%       10.450000
75%       18.650000
max       64.190000
moda 13.0
curtosis  2.230668154164368
asimetría  1.4032421705281972
varianza  115.43127320297181

```

- 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

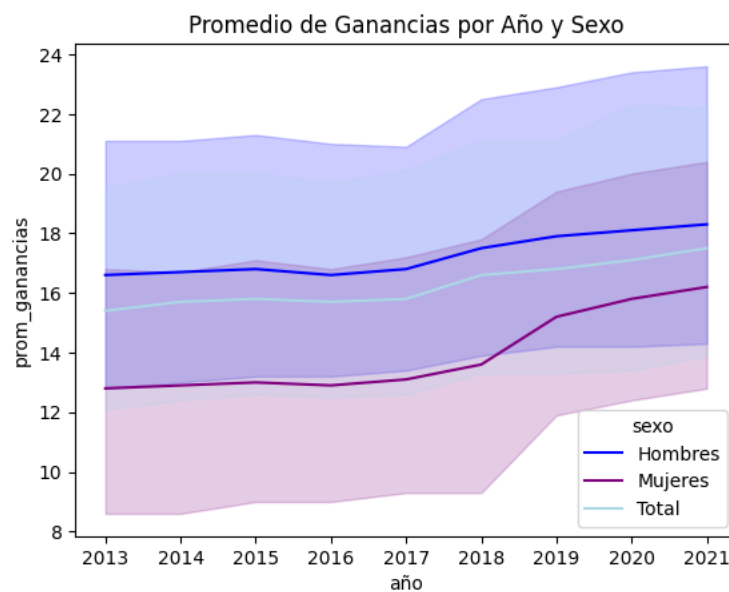


En este gráfico donde muestra las medianas por cada profesión en 2014, se puede observar las diferencias son muy sutiles, con poca variabilidad, en muchas son iguales, por lo que puede complementar la información de la prueba de Kruskal. La diferencia más evidente es en directores y gerentes.



Este gráfico de calor muestra las diferencias de países donde también complementa la información de las pruebas de kruskal y los boxplot, ya que se puede observar los patrones identificados de ganancias y diferencias.

Serie Temporal España



La serie temporal de España muestra como ha sido la evolución de los salarios en España por sexo, en este se puede observar que el de los hombres es más alto, sin embargo, a partir del año 2018 se observa un aumento considerable en las ganancias de las mujeres, esto se puede deber debido a la implementación de reformas legislativas en el ámbito laboral para la igualdad, también promovidas por la unión europea

CONCLUSIONES

- No se puede afirmar la brecha salarial, por varias razones, ya que las pruebas no lo muestran claramente, por otro lado, hay falta de representación en algunos casos por lo que se puede sesgar. Además, para un análisis más acertado sería interesante tener otras variables como tener hijos o no, estado civil o nivel de estudios.

- Si que existe una diferencia en medianas en todos los casos y que existe más diferencia entre países que en las profesiones, cabe destacar la diferencia evidente de promedio de ganancias entre profesiones, también mostrada estadísticamente.
- Aunque solo se ve diferencias muy sutiles y no se puede afirmar la brecha salarial, cabe destacar en la serie temporal la mejora de los salarios de las mujeres coincidiendo con el año 2018, año muy representativo en la lucha feminista,