

# Evolución salarial en España y comunidades (2008-2020)

Un estudio de caso para el certificado Google de analista de datos

Por Diego Reguera

16-11-2022

## 1. Introducción

Soy Diego Reguera tras terminar mi última etapa universitaria en un MA de Economía Internacional y Desarrollo, y seguido de un periodo de prácticas y voluntariados para organismos públicos e internacionales, me decidí centrar el mundo de la ciencia y análisis de datos. Estuve haciendo formación y prácticas primero para un puesto de DBA e ingeniero de *Datastage* durante 3 meses. Para posteriormente centrarme en formarme en esta posición de analista de datos. Este documento, es un trabajo que se encuadra en el proyecto final para el “certificado *Google* de análisis de datos” que busca poner en práctica los conocimientos y herramientas adquiridas durante dicho curso.

Pretendo hacer una breve muestra de la evolución de los salarios en mi país España atendiendo a diversos elementos de género y diferencias locales. Respecto al por qué de mi elección en esta pregunta actualmente hay un interés social por situar las dinámicas del mercado de trabajo en lo que se refiere al género. Se delimita desde el año 2008 por acotarlo temporalmente desde la última crisis española hasta la actualidad, que en este caso se cuentan con datos hasta el 2020. Se resaltarán datos a nivel nacional y de 3 comunidades aleatorias con una renta superior a la media española como son C. de Madrid, Cataluña y C. de Valencia. Los datos provienen de fuentes oficiales del INE extraídos a fecha de la realización del estudio.

## 2. Limpieza de datos

Dado la fiabilidad de la fuente (INE) no es necesario un proceso de limpieza de datos profundo por lo que con Excel es suficiente para realizar tal limpieza de datos. En caso de necesitar un proceso más largo de limpieza y documentación de dichos cambios se recomienda la herramienta de *RMarkdown*. Consultar el documento de limpieza de datos para ver nombres de ficheros, y el “log” manual de cambios en las tablas de Excel en este [enlace](#).

No se observan grandes incongruencias salvo por los salarios negativos de mujeres en el sector de la construcción que serán eliminados del análisis. Se procede a acotar los datos de interés para nuestro estudio. Se filtran columnas y pasan de 2808 filas a 624 tras descartar las comunidades autónomas que se examinarán y se conservan “Total Nacional”, “Cataluña”, “Madrid, Comunidad de” y “Comunitat Valenciana”. Se copia en un nuevo archivo cvs los datos filtrados excluyendo al resto. Se ha eliminado el sector de la construcción por contar con números negativos.

Respecto a los datos de divorcios el INE ya proveía de las tasas de variaciones en las fechas indicadas por lo que no tuve que procesarlos. Posteriormente se procedió al cálculo de las tasas de variación de la situación de empleo y sueldo a nivel nacional de cara a vislumbrar posibles relaciones entre ambas variables. Se elimina el año 2021 y 2006-2008 de la tabla de divorcios ya que no contamos con datos a dicha fecha en el fichero copia de divorcios.

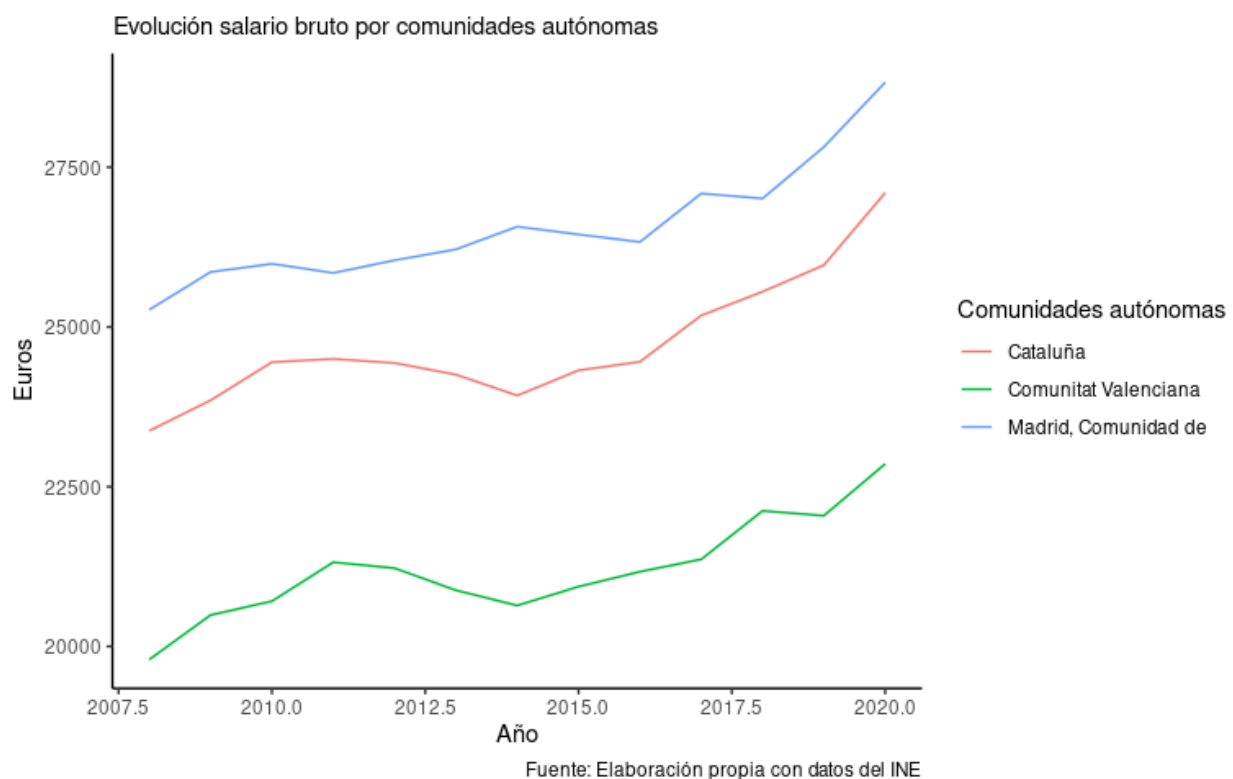
Se calculan las tasas de variación de la tabla de salarios en Excel. Como los datos están vertical, hay que tener cuidado de no mezclar categorías al calcular dichas variaciones por lo que se eliminan los valores del primer año en este caso el 2008, cuya celda se aprovechará para calcular la media de lo que ha variado el salario por categoría en ese sector. Posteriormente se guarda, filtra y eliminan valores por fecha a 2008 para poder visualizar las variaciones sin los datos de media en R. A su vez se seleccionan los valores de 2008 por medio de filtrado y eliminación de los valores “Ambos sexos” de la copia, que no nos interesan de cara a mostrar las diferencias.

### 3. Análisis y presentación de datos

En este [documento](#) en R de mi repositorio en [Github](#), se pueden seguir las líneas de código y todos los “plots” que he creado para filtrar las filas y que se visualicen aquellas que nos interesen. Por ejemplo, para la gráfica 1 se han creado un “plot” con varios “dataframe” seleccionando mediante la función “filter” hasta crear “df\_w\_comunidades3” donde se incluyen las Comunidades autónomas, el sector de actividad y el sexo que nos interesa en esta gráfica en concreto. La metodología del resto de “plots” se han creado de manera muy similar, tal y como viene detallado en el documento *Markdown*.

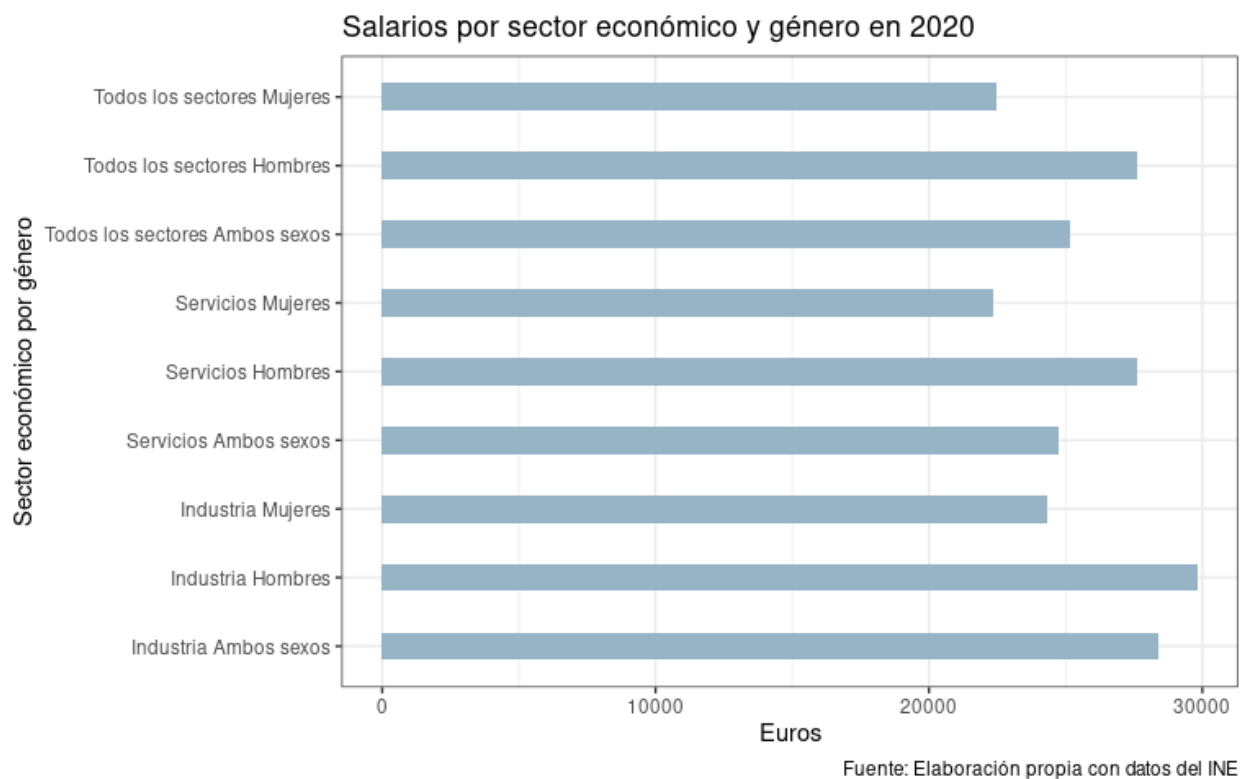
Una vez que los datos están limpios se procede a la observación más detenida de datos. En el gráfico 1 se observa la “evolución salario bruto por comunidades autónomas” para hacernos una idea general de diferencias de renta dentro del territorio español en términos absolutos.

Gráfica 1



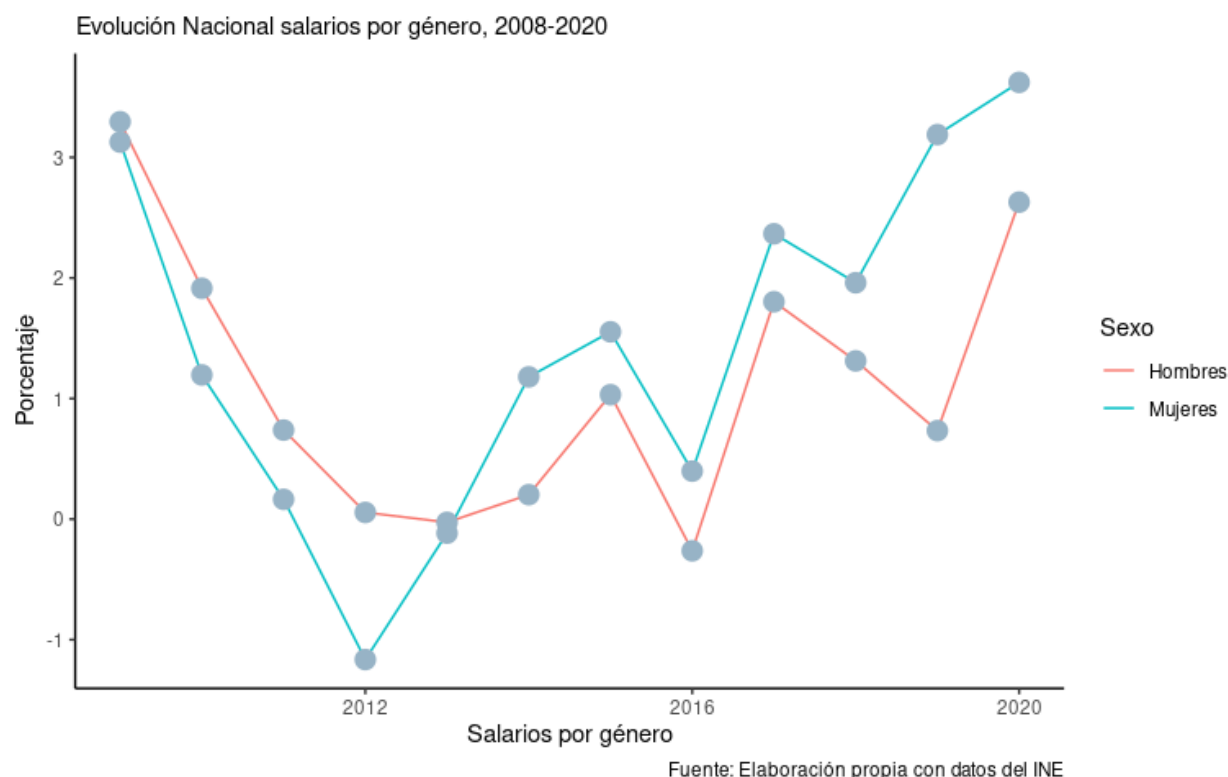
Por otro lado, en la se han seleccionado los datos de salarios por industrias por género para el año 2020 a nivel nacional. Se puede observar que sigue habiendo cierta disparidad a lo largo de todas las industrias en términos absolutos, donde el hombre gana más que la mujer siendo más pronunciada la diferencia en el sector industrial, vuelvo a recordar que en este análisis se ha excluido el sector de la construcción, el cual históricamente es y ha sido sector mayoritariamente masculino.

Grafica 2



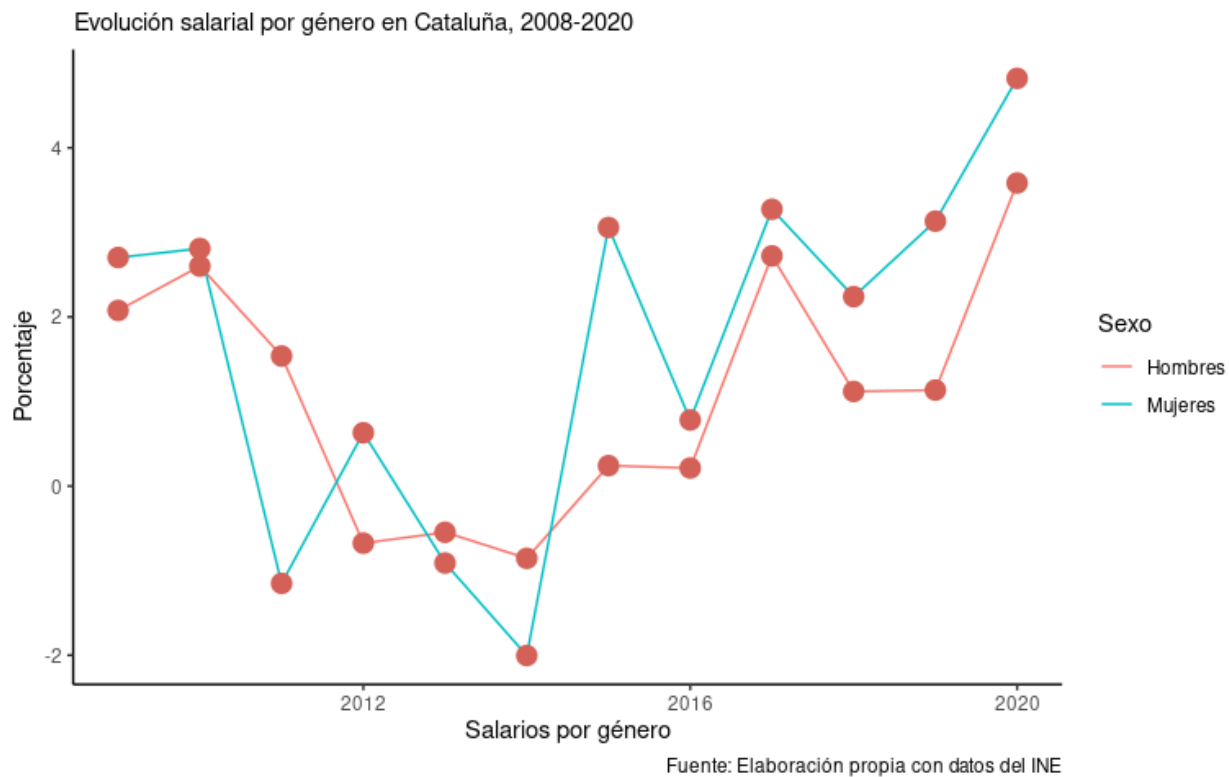
Respecto a la evolución porcentual expuesta en la gráfica 3 a nivel nacional se puede ver que en general hasta 2012 hubo una caída generalizada de los salarios como consecuencia de la crisis del 2008, sin embargo, el empleo femenino ha tenido una mejor recuperación y crecimiento salarial en España lo que ayudaría a reducir la existente brecha salarial ya expuesta en el gráfico 2.

Gráfica 3

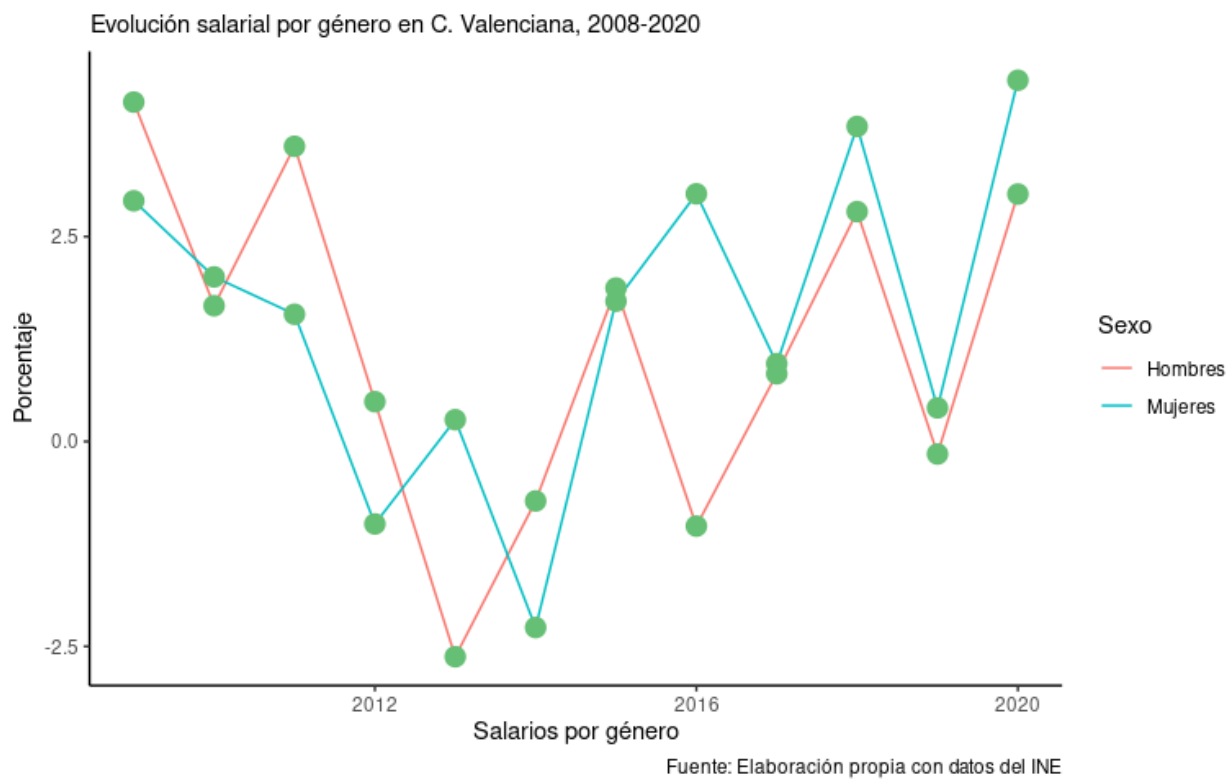


Si lo observamos de manera desagregada por comunidades autónomas (gráficas 4 a 6), observamos que la tónica descrita en la gráfica 3 se repite. En el caso de la Comunidad de Madrid de manera menos pronunciada. Otro aspecto por resaltar son los crecimientos salariales en 2020 superiores al 4% por encima de la media nacional que ese ligeramente superior al 3. Lo que hace denotar procesos de centralización económicos en la capital y de zonas industriales catalanes vitales en España. A su vez, se producen caídas salariales durante el pico de la crisis de manera menos pronunciada en estas regiones respecto, en este caso, a Valencia.

Gráfica 4



Gráfica 5

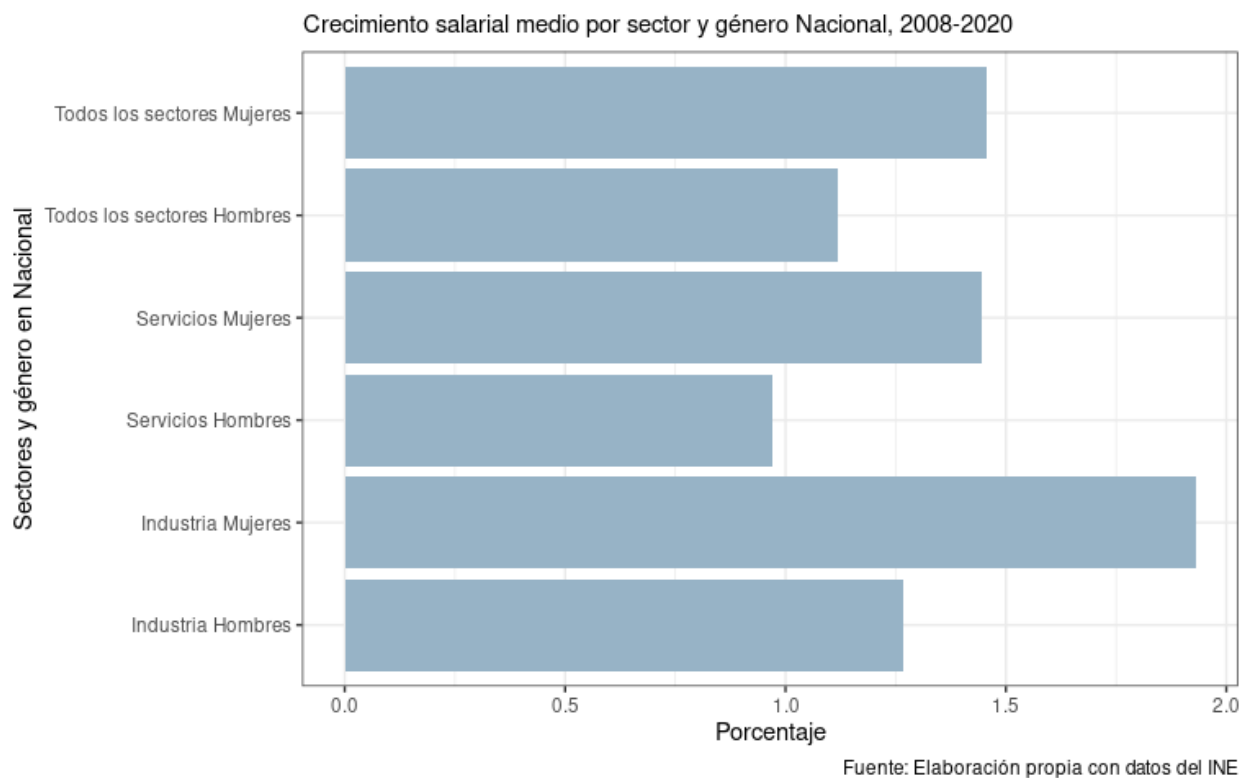


Gráfica 6



A nivel Nacional y sectorial volvemos a ver un mayor crecimiento medio salarial del empleo femenino a lo largo de los sectores económicos y del país en su conjunto, durante el periodo 2008-2020. Siendo especialmente pronunciado en el sector industrial. Los hombres crecen por debajo del 1% de media en el sector servicios, un sector con tendencias recientes a la precarización en algunos trabajos.

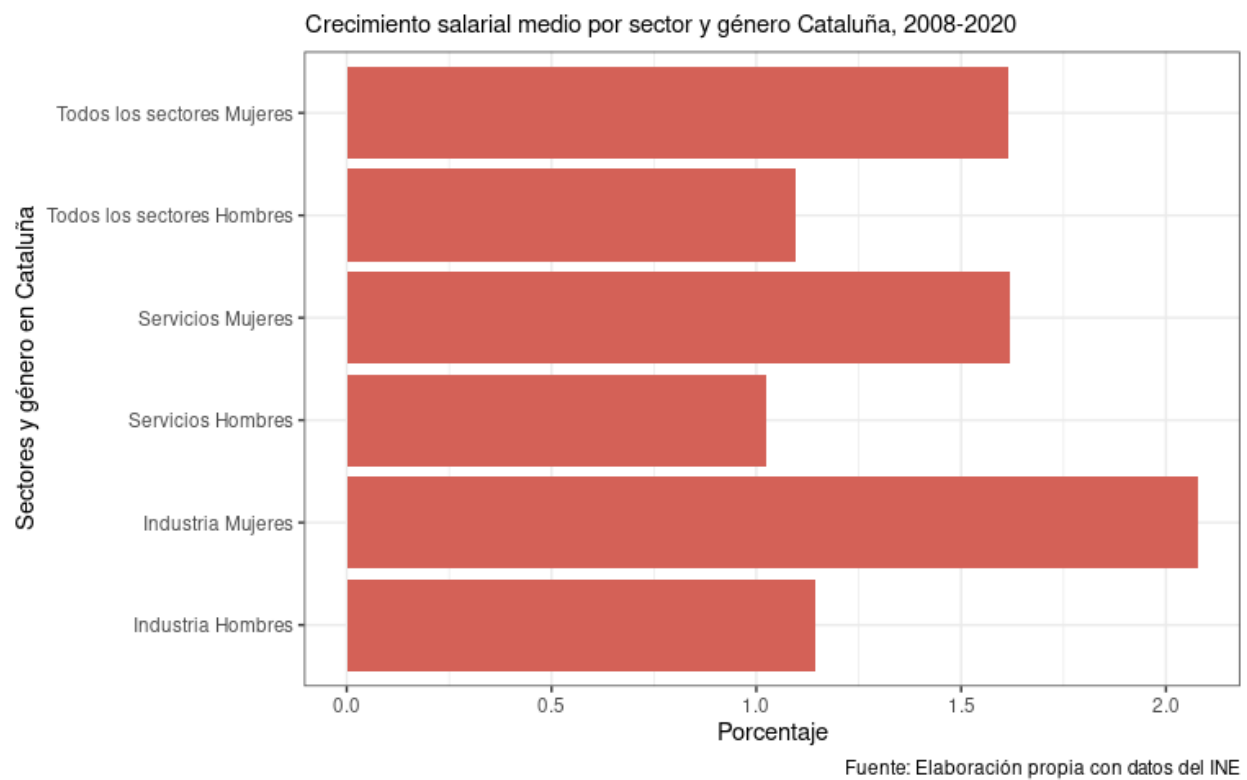
Gráfica 7

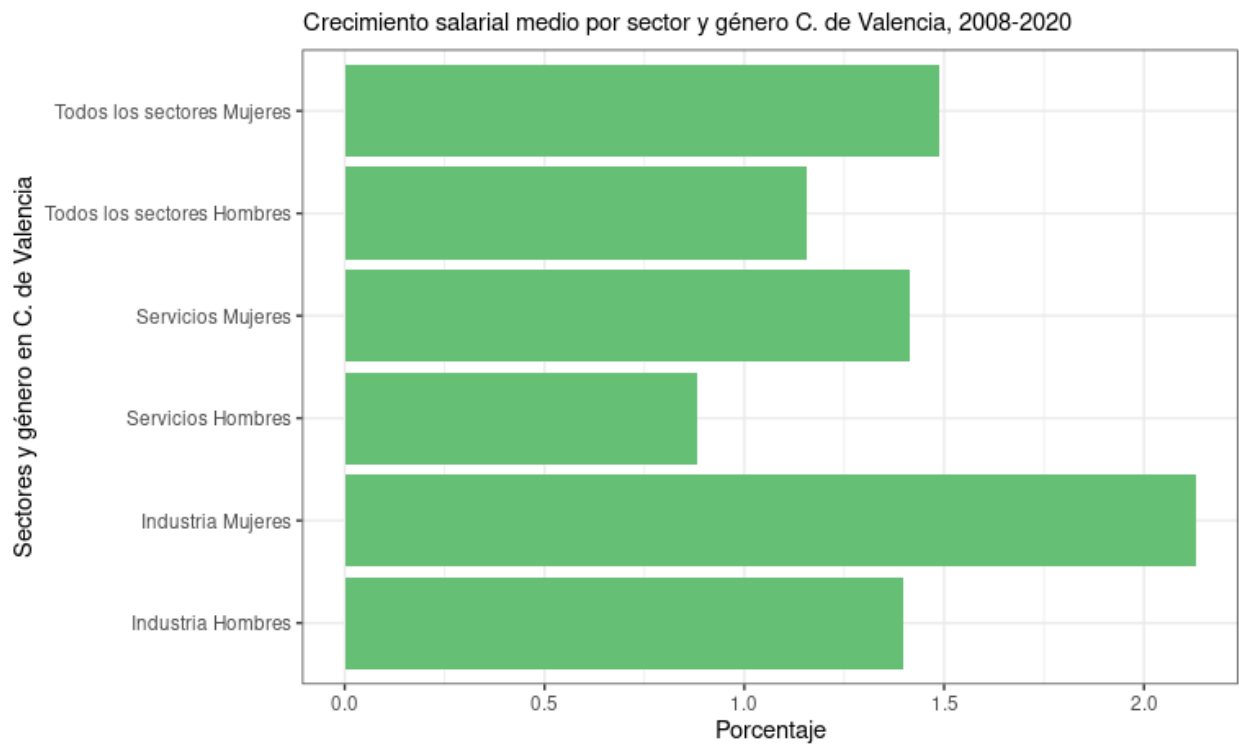


Mientras que desagregado regionalmente (gráficas 8-10) se observa tendencias similares al nacional con algunas diferencias. El crecimiento salarial de hombres en Madrid en todos los sectores no llega al 1% de media, mientras que el salario del empleo femenino supera el 1.5% de media en las 3 regiones. También resalta que la región de las 3 que mayor crecimiento medio de salario femenino se da en Cataluña.

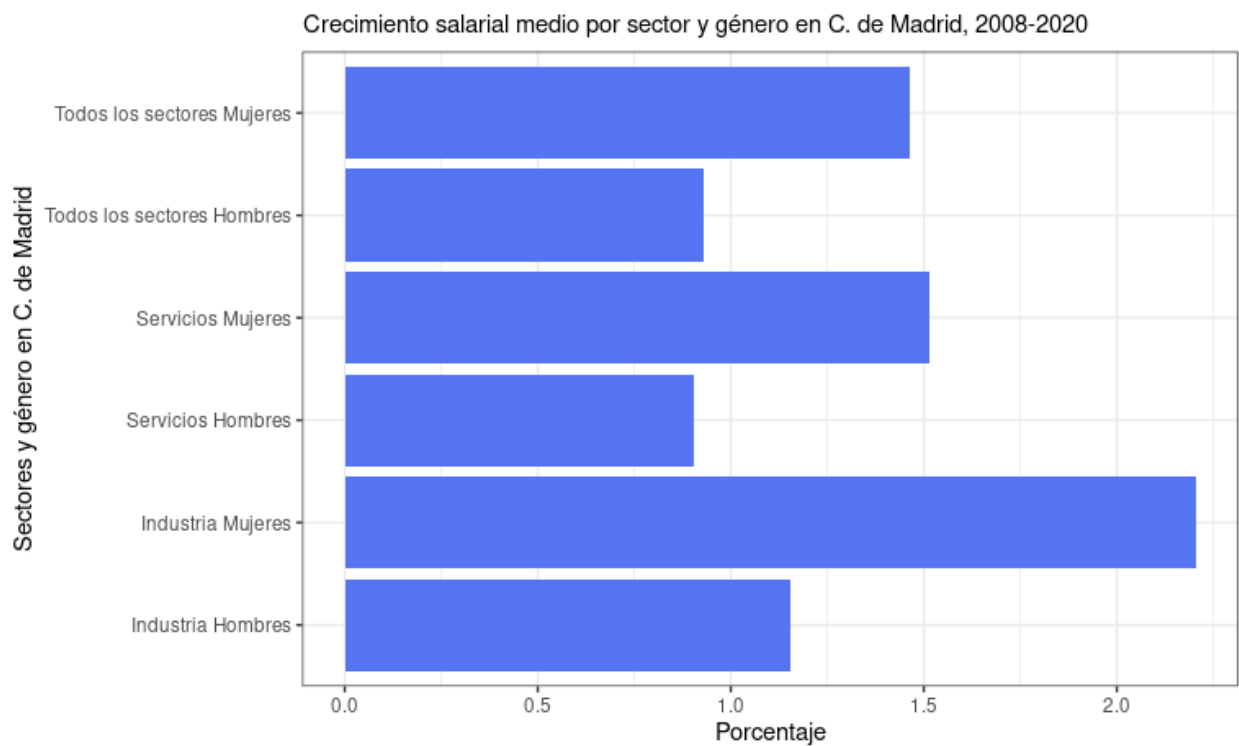


Graficas 8-10





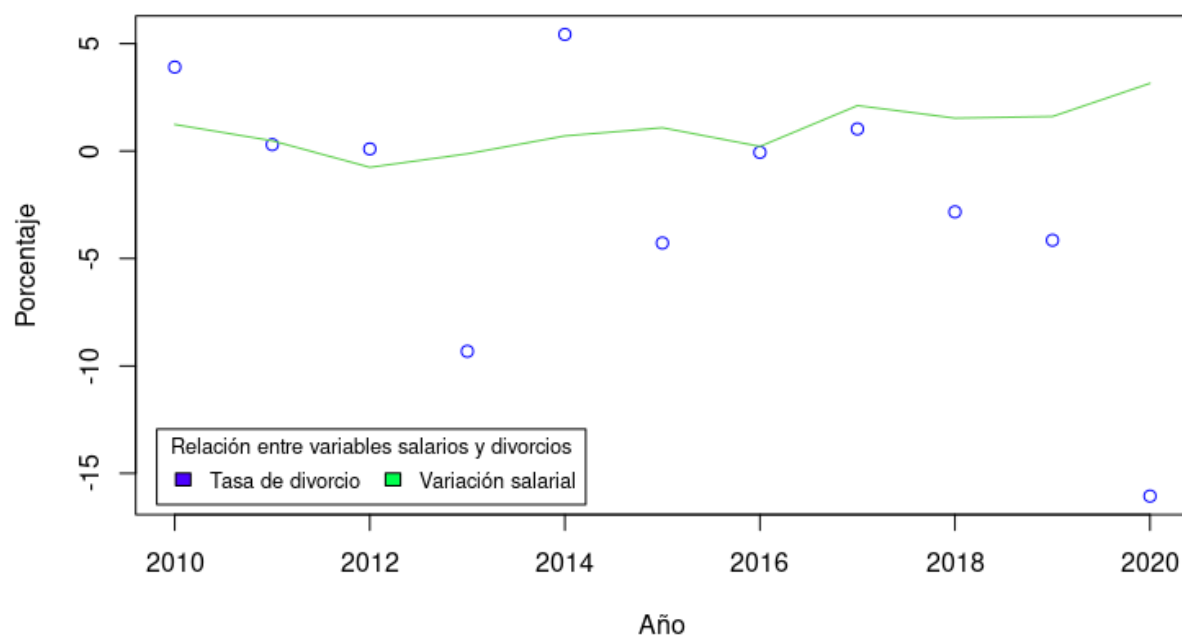
Fuente: Elaboración propia con datos del INE



Fuente: Elaboración propia con datos del INE

Finalmente, de cara a ver buscar correlaciones en comportamiento salarial y tasas de divorcio, la gráfica 11 no apunta algunas ideas, pero no de forma rotunda. Hay cierta correlación entre vistas al crecimiento y divorcios, donde tras los primeros años de la crisis hasta el 2012 los divorcios se redujeron hasta un 10% en los años de recuperación. Posteriormente en los últimos años donde se vuelven a dar mejores salariales se produce otro pico de bajada de salarios de más del 5%. Aunque es significativo este gráfico tampoco se pueden arrojar afinaciones rotundas a este aspecto, ya que influyen más aspectos no contemplados en esta investigación.

Gráfica 11



Fuente: elaboración propia con datos del INE

#### 4. Conclusiones

- La primera idea que se puede resaltar de esta investigación es que tras el pico de la crisis del 2008 en España se empezaron a recuperar los salarios a partir del 2012.
- En España hay una “gap” salarial entre hombres y mujeres en los principales sectores económicos.
- Sin embargo, los salarios de los hombres han tenido una peor recuperación lo que significa que el salario del empleo femenino ha crecido más anualmente. Lo que puede llevar a una futura convergencia o equidad de salarios por un mismo empleo.
- Ha habido en cierta medida una mayor caída de divorcios en los años de crecimiento de salaria.

## **5. Fuentes:**

INE: <https://www.ine.es/index.htm>  
<https://n9.cl/bfpn>