

# Análisis del mercado laboral en España

Catret Ruber, Pablo<sup>1</sup>, Palazón Caballero, José Miguel<sup>1,\*</sup>, Rosique Martínez, Marcos<sup>1</sup>

<sup>1</sup> Universitat de València - Escola Tècnica Superior d'Enginyeria (ETSE) Avinguda de l'Universitat, 46100 Burjassot, Valencia;

\* Correspondence: [jomipaca@alumni.uv.es](mailto:jomipaca@alumni.uv.es).

**Simple Summary:** A Simple summary goes here.

**Abstract:** A single paragraph of about 200 words maximum. For research articles, abstracts should give a pertinent overview of the work. We strongly encourage authors to use the following style of structured abstracts, but without headings: 1) Background: Place the question addressed in a broad context and highlight the purpose of the study; 2) Methods: Describe briefly the main methods or treatments applied; 3) Results: Summarize the article's main findings; and 4) Conclusion: Indicate the main conclusions or interpretations. The abstract should be an objective representation of the article, it must not contain results which are not presented and substantiated in the main text and should not exaggerate the main conclusions.

**Keywords:** keyword 1; keyword 2; keyword 3 (list three to ten pertinent keywords specific to the article, yet reasonably common within the subject discipline.).

## 1. Introducción

El mercado laboral en España es un sistema complejo y dinámico que refleja las interacciones entre diversos factores económicos, sociales y demográficos. Este trabajo de análisis exploratorio de datos combina dos pilares fundamentales para entender las dinámicas laborales: la Encuesta de Población Activa (EPA) y la Clasificación Nacional de Actividades Económicas (CNAE-2009). A través de estas fuentes, se busca ofrecer una visión integral de las tasas de empleo, actividad y ocupación, desglosadas por género, grupo de edad y comunidad autónoma, así como por ramas de actividad económica.

La EPA, realizada trimestralmente por el Instituto Nacional de Estadística (INE), proporciona información detallada sobre la participación laboral de la población, permitiendo analizar las diferencias en el acceso al empleo según el género, la edad y el territorio. En esta parte del estudio, se examinan las tasas del mercado laboral en las distintas comunidades autónomas y cómo estas se ven influenciadas por eventos como la crisis financiera de 2008 o la pandemia de COVID-19. Además, se busca identificar desigualdades estructurales y dinámicas regionales que puedan servir como base para estudios más específicos o el diseño de políticas públicas.

Por otro lado, la CNAE-2009 aporta un marco estándar para clasificar las actividades económicas en las que se desempeñan los trabajadores, permitiendo analizar cómo se distribuye la fuerza laboral entre diferentes sectores. Este enfoque complementario permite explorar no solo el "dónde" y "quién" trabaja, sino también el "en qué" trabaja la población, revelando patrones de especialización sectorial y el impacto de la transformación económica a lo largo de los años.

La combinación de ambas perspectivas —la distribución demográfica y regional desde la EPA y la estructura sectorial desde la CNAE— permite abordar preguntas clave sobre el mercado laboral, tales como:

- ¿Cómo varía la participación laboral entre comunidades autónomas?
- ¿Cómo varía la participación laboral entre grupos de edad?

**Citation:** Catret Ruber, P.; Palazón Caballero, J.M.; Rosique Martínez, M. Análisis del mercado laboral en España. *Journal Not Specified* **2023**, *1*, 0. <https://doi.org/>

Received:

Revised:

Accepted:

Published:

**Copyright:** © 2025 by the authors. Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

- ¿Qué sectores han experimentado los mayores cambios tras eventos históricos como la crisis de 2008 o la pandemia de COVID-19?

Este análisis exploratorio busca no solo describir el estado actual del mercado laboral en España, sino también identificar tendencias, desigualdades y oportunidades de mejora. El objetivo final es proporcionar una base sólida para futuros estudios y contribuir al diseño de estrategias efectivas en el ámbito económico, social y político.

2. Importación y tratamiento de datos

En primer lugar, se han descargado los datos directamente desde la base de datos abierta INE Base. Estos datos se presentan en un formato de csv, separado por el carácter ‘;’, y con el uso de marca de decimales española ‘,’.

2.1. Encuesta de Población Activa (EPA)

Como se ha mencionado anteriormente en la introducción, la EPA es realizada trimestralmente por el Instituto Nacional de Estadística, por lo que se han extraído cinco datasets de su base de datos: población total, activa, inactiva, ocupada y parada.

Cabe destacar que estos datasets están desglosados por comunidad y ciudad autónoma, sexo y grupo de edad, con las tres columnas expresadas en “miles de personas” como unidad. El periodo de los datos figura desde el primer trimestre de 2002 hasta el tercer trimestre de 2024.

Al tratarse de un análisis del mercado laboral eliminaremos al grupo de población menor a 16 años pues no tienen edad suficiente para trabajar. Notamos que en el dataset Población hay una columna que solo contiene la cadena “Total Nacional”, por lo que la eliminamos. Además, en la columna “Comunidades y Ciudades Autónomas” hay valores faltantes que coinciden con las filas del propio Total Nacional, de modo que los completamos con dicha cadena.

Notamos que para el conjunto de datos de población total en edad de trabajar y población inactiva hay un intervalo de edad más que para el resto de datasets: se divide l”55 y más años” en los grupos “De 55 a 64 años” y “65 y más años”. Dado que buscamos obtener un dataset único y compacto uniremos ambos grupos de edad como en el resto de datasets. Una vez normalizada la estructura de los datasets podemos unificar todas las tablas en una sola.

Otra circunstancia a corregir que los datos de la columna “Periodo” son cadenas; para solucionarlo, definimos y empleamos la función “SacarFechas” para transformarlos a tipo Date. Una vez corregido, comprobaremos si existen datos faltantes en nuestro dataset.

##	Sexo	Comunidades y Ciudades Autónomas
##	0	0
##	Edad	Periodo
##	0	0
##	Población en edad de trabajar	Activos
##	0	43
##	Inactivos	Ocupados
##	0	277
##	Parados	
##	272	

Existe únicamente un porcentaje muy bajo de valores faltantes entre las columnas “Activos”, “Ocupados” y “Parados”. Veamos un ejemplo de estos casos.

##	\$‘Comunidades y Ciudades Autónomas‘	
##		
##	02 Aragón	03 Asturias, Principado de
##	1	12
##	04 Balears, Illes	05 Canarias

##	1	5	89			
##	06 Cantabria	11 Extremadura	90			
##	18	6	91			
##	15 Navarra, Comunidad Foral de	16 País Vasco	92			
##	12	3	93			
##	17 Rioja, La	18 Ceuta	94			
##	19	182	95			
##	19 Melilla		96			
##	247		97			
##			98			
##	\$Edad		99			
##			100			
##	55 y más años	De 16 a 19 años	De 20 a 24 años	De 25 a 34 años	De 35 a 44 años	101 años
##	101	345	18	1		102 13
##	De 45 a 54 años	Total				103
##	27	1				104

Los datos se concentran en Comunidades y Ciudades Autónomas con proporcionalmente poca población y en rangos de edades donde es poco habitual estar parado (ya que para que se cumpla dicha condición se debe estar buscando activamente trabajo). De hecho, el único valor que podría resultar extraño es que exista un valor faltante para un Total de edades, pero al comprobar la localización y fecha notamos que es razonable.

##	#	A tibble: 1 x 3	110			
##	Sexo	‘Comunidades y Ciudades Autónomas’	Periodo	111		
##	<chr>	<chr>	<date>	112		
##	1	Hombres	19	Melilla	2002-06-01	113

Dicho patrón parece reflejar que muy poca población de esas características se encontraba parada; por tanto, sustituiremos los valores faltantes del dataset por ceros.

2.2. Clasificación Nacional de Actividades Económicas (CNAE-2009)

Este dataset divide los datos según rama de actividad, sexo y fecha en el periodo comprendido entre el primer trimestre de 2008 y el tercer trimestre de 2024, usando las mismas unidades que en el resto de datasets. No obstante, en este caso podemos encontrar los datos segmentado en dos subconjuntos principales:

- Porcentajes (“Total\_abs”): Representan la proporción de ocupación de cada rama de actividad con respecto al total del sexo correspondiente.
- Valores absolutos (“Total\_porc”): Proporcionan el número total de empleados en cada rama.

Esta separación permite analizar tanto las tendencias globales (valores absolutos) como la estructura relativa de los sectores (porcentajes).

Nos encontramos de nuevo con el problema del formato de las fechas, por lo que volveremos a aplicar la función SacarFechas. Además, para diferenciar entre cada tipo de dato (valor absoluto y porcentaje) vamos a extraerlos en dos datasets para posteriormente realizar un join, esencialmente como si se realizara un pivote. Una vez hemos unificado la tabla es recomendable realizar un análisis de datos faltantes.

##	Rama.de.actividad.CNAE.2009	Sexo	132
##	0	0	133
##	Periodo	Total_abs	134
##	0	313	135
##	Total_porc		136
##	313		137

Observamos que para un pequeño subconjunto de filas no existen cifras totales. Veamos si siguen algún patrón.

```
## [1] "05 Extracción de antracita, hulla y lignito"
## [2] "06 Extracción de crudo de petróleo y gas natural"
## [3] "07 Extracción de minerales metálicos"
## [4] "09 Actividades de apoyo a las industrias extractivas"
```

Notamos que los valores faltantes se encuentran en ramas de actividades poco comunes relacionadas con la industria pesada, por lo que dichos NA reflejarán que muy poca población se dedica a ello, probablemente menos del mínimo registrable; en consecuencia sustituiremos dichos valores por ceros.

Por último, añadiremos una columna que calcula la variación entre un periodo y el siguiente, lo que permitirá identificar momentos de cambio significativo en el mercado laboral, detectando tendencias positivas o negativas a lo largo del tiempo. Esto también permitirá ver de manera rápida la estacionalidad de la ocupación, sobre todo en ciertas ramas de actividad. Es importante mencionar que realizar las diferencias entre un periodo y el siguiente genera NAs para el primer periodo de cada rama, por lo que sustituiremos estos valores faltantes por ceros.

### 3. Representación y análisis de datos

Una vez preparados, los datos se visualizan a través de gráficos y tasas que permiten interpretar patrones, tendencias y distribuciones de manera más clara y directa.

#### 3.1. Encuesta de Población Activa (EPA)

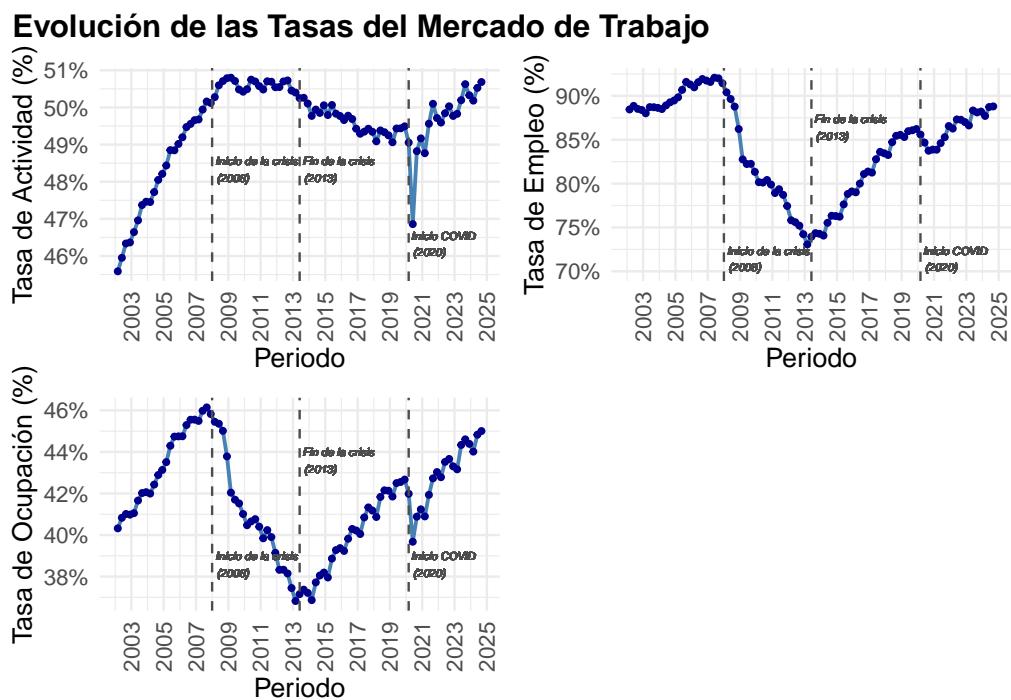
Las tasas del mercado de trabajo, como la de actividad, empleo y ocupación, son fundamentales para este análisis porque permiten comparar de forma clara y estandarizada las dinámicas laborales entre comunidades autónomas y grupos poblacionales. Estas métricas facilitan el estudio de tendencias temporales y desigualdades estructurales, como las brechas de género o regionales, que no serían evidentes al observar valores absolutos. Además, su uso es clave para identificar áreas críticas y evaluar el impacto de eventos económicos, como la crisis de 2008 o la pandemia de COVID-19, en la participación laboral de la población.

Por ello, las siguientes tasas quedan incorporadas dentro del dataset anteriormente mencionado:

- Tasa de Actividad (TA):  $\frac{\text{Activos}}{\text{Población}_{>16}} * 100$
- Tasa de Inactividad (TI):  $\frac{\text{Inactivos}}{\text{Población}_{>16}} * 100$
- Tasa de Ocupación (TO):  $\frac{\text{Ocupados}}{\text{Población}_{>16}} * 100$
- Tasa de Empleo (TE):  $\frac{\text{Ocupados}}{\text{Activos}} * 100$
- Tasa de Desempleo (TD):  $\frac{\text{Parados}}{\text{Activos}} * 100$

Notamos que las tasas de inactividad y desempleo no proporcionan información nueva ya que se pueden obtener mediante la de actividad y empleo respectivamente:  $TI = 1 - TA$  y  $TD = 1 - TE$ .

La figura 1 muestra la evolución de tres indicadores clave del mercado laboral en España: la tasa de actividad, la tasa de empleo y la tasa de ocupación, desde 2002 hasta prácticamente la actualidad, reflejando periodos de estabilidad y cambios abruptos relacionados con eventos económicos y sociales. Entre 2002 y 2008, todas las tasas experimentan un crecimiento sostenido, impulsado por la bonanza económica previa a la crisis financiera global, con un auge en sectores como la construcción y los servicios, y una incorporación activa de la población al mercado laboral.



**Figure 1.** Evolución de la tasa de actividad en España

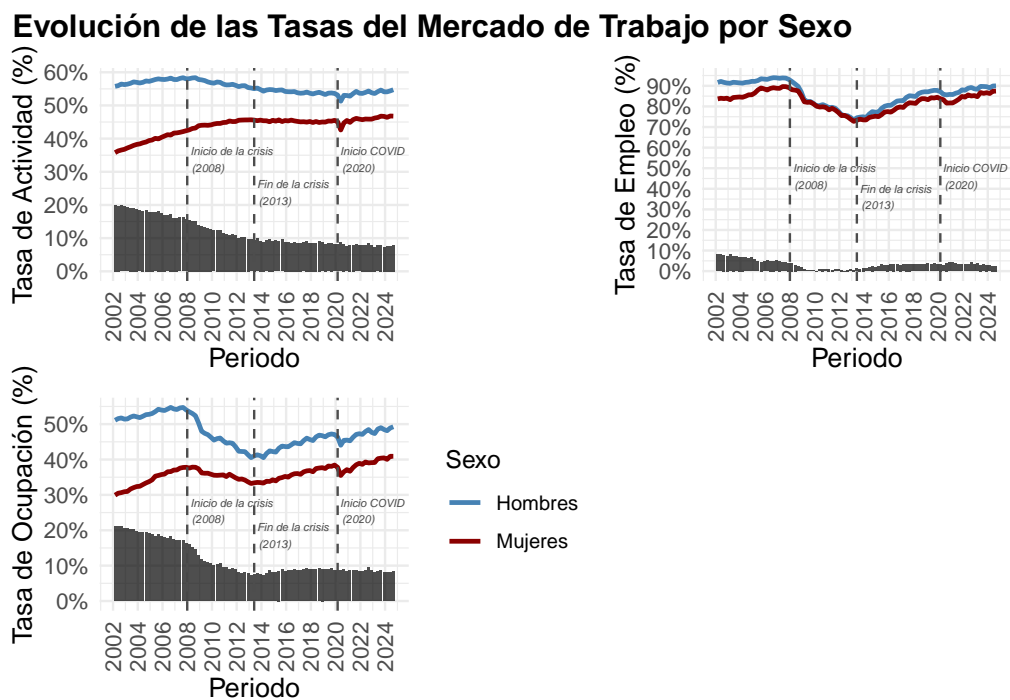
El inicio de la crisis en 2008 genera un comportamiento diferenciado: mientras que la tasa de actividad se mantiene relativamente estable hasta 2013, las tasas de empleo y ocupación muestran un descenso significativo, reflejando el impacto directo de la destrucción de empleo y la pérdida de capacidad de absorción del mercado laboral. Esto indica que, aunque muchas personas siguieron buscando trabajo activamente (evitando una caída en la actividad), el deterioro de las condiciones laborales afectó severamente las oportunidades de empleo. La tasa de empleo pasó de más del 90% a por debajo del 75% en 2013, o que es lo mismo, un desempleo de más del 25% en la población.

A partir de 2013, con los primeros signos de recuperación económica, la tasa de actividad decrece levemente, posiblemente asociada a una menor presión en el mercado laboral y una reducción del abandono escolar, mientras que las tasas de empleo y ocupación comienzan una recuperación progresiva. Este periodo refleja una estabilización del mercado laboral con la recuperación de ciertos sectores clave.

El impacto de la pandemia de COVID-19 en 2020 es evidente en todas las tasas, con caídas abruptas, especialmente en la tasa de ocupación, debido al cierre temporal de actividades económicas y restricciones de movilidad. Este impacto fue más pronunciado que el de la crisis financiera. Desde 2021, se observa una recuperación acelerada en las tres tasas, marcada por la reincorporación de trabajadores al mercado laboral y el dinamismo de sectores como los servicios digitales, el comercio y el turismo.

La figura muestra la evolución de las tasas de actividad, empleo y ocupación desglosadas por género, evidenciando patrones de desigualdad, convergencia y algunas diferencias clave entre estos indicadores. La tasa de actividad masculina, consistentemente más alta que la femenina, experimenta un descenso gradual entre 2008 y 2020, probablemente vinculado a los cambios estructurales tras la crisis financiera. En contraste, la tasa de actividad femenina ha crecido de forma sostenida desde 2002, impulsada por cambios sociales y legislativos, aunque este crecimiento se ralentiza a partir de 2008. Este proceso ha reducido la brecha de género de casi un 20% en 2002 a cerca de un 8% en años recientes, como reflejan las barras grises que representan esta diferencia.

En la tasa de empleo, sin embargo, las diferencias entre hombres y mujeres son mucho más reducidas, lo que sugiere que la principal disparidad radica en la población activa.



**Figure 2.** Evolución de la tasa de actividad en España

Esto se explica porque la tasa de empleo mide únicamente a las personas empleadas dentro de la población activa, mientras que la tasa de ocupación considera a toda la población en edad de trabajar. Por ello, las diferencias en la tasa de ocupación entre hombres y mujeres son más marcadas, reflejando no solo las desigualdades en el empleo, sino también las disparidades en la incorporación al mercado laboral.

Durante la pandemia de COVID-19 en 2020, todas las tasas muestran un descenso significativo, seguido de una recuperación que resalta la resiliencia del mercado laboral. Este análisis pone en evidencia tanto los avances hacia la igualdad de género en términos laborales como las áreas donde persisten desigualdades estructurales, especialmente en la población activa y su impacto en las tasas de ocupación.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.