

# 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); [rosique2@alumni.uv.es](mailto:rosique2@alumni.uv.es)

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

**Abstract:** Este trabajo aborda el análisis exploratorio del mercado laboral en España como parte de la asignatura de Análisis Exploratorio de Datos. Utiliza datos de la Encuesta de Población Activa (EPA) y la Clasificación Nacional de Actividades Económicas (CNAE-2009), abarcando el período entre 2002 y 2024. En el estudio se realizan técnicas de ciencia de datos para limpiar y analizar información sobre tasas de empleo, actividad y ocupación, desglosadas por género, grupos de edad, regiones y sectores económicos. Se emplearon herramientas de tratamiento y transformación de datos para unificar conjuntos, corregir valores faltantes y estandarizar formatos, incluyendo la normalización de fechas y la imputación de datos ausentes. Además, se integraron métricas como tasas de actividad, empleo y desempleo, calculadas mediante fórmulas específicas para identificar patrones y tendencias en el tiempo. Los métodos incluyen visualizaciones gráficas para representar dinámicas temporales y regionales, cálculos de variación entre periodos para detectar cambios significativos y técnicas de clusterización para analizar de manera segmentada la evolución del mercado laboral en distintos grupos. El análisis revela el impacto de eventos como la crisis de 2008 y la pandemia de COVID-19 en el mercado laboral, destacando desigualdades estructurales y patrones sectoriales.

**Keywords:** Mercado laboral; Tasa; Clustering; Pandemia COVID-19; Análisis exploratorio de datos

## 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.

**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/>).

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?
- ¿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 “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 unificarlos.

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.

Table 1. Frecuencias de datos faltantes en el conjunto de datos EPA

Variable	Valores faltantes
Sexo	0
Comunidades y Ciudades Autónomas	0
Edad	0
Periodo	0
Población en edad de trabajar	0
Activos	43
Inactivos	0
Ocupados	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.

**Table 2.** Frecuencias de datos faltantes por Comunidad Autónoma en el conjunto de datos EPA

Comunidad	Valores faltantes
02 Aragón	1
03 Asturias, Principado de	12
04 Balears, Illes	1
05 Canarias	5
06 Cantabria	18
11 Extremadura	6
15 Navarra, Comunidad Foral de	12
16 País Vasco	3
17 Rioja, La	19
18 Ceuta	182
19 Melilla	247

**Table 3.** Frecuencias de datos faltantes por edad en el conjunto de datos EPA

Grupo de edad	Valores faltantes
55 y más años	101
De 16 a 19 años	345
De 20 a 24 años	18
De 25 a 34 años	1
De 35 a 44 años	13
De 45 a 54 años	27
Total	1

Los datos se concentran en Comunidades y Ciudades Autónomas con proporcionalmente poca población y en rangos de edades donde es poco habitual estar ocupado (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.

**Table 4.** Valor faltante en el total de edades en el conjunto de datos EPA

Sexo	Comunidades y Ciudades Autónomas	Periodo
Hombres	19 Melilla	2002-06-01

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.

**Table 5.** Frecuencias de datos faltantes en el conjunto de datos CNAE

Variable	Valores faltantes
Rama.de.actividad.CNAE.2009	0
Sexo	0
Periodo	0
Total_abs	313
Total_porc	313

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

**Table 6.** Primeras cuatro actividades con valores faltantes en `Total_abs`

Rama de actividad
05 Extracción de antracita, hulla y lignito
06 Extracción de crudo de petróleo y gas natural
07 Extracción de minerales metálicos
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**

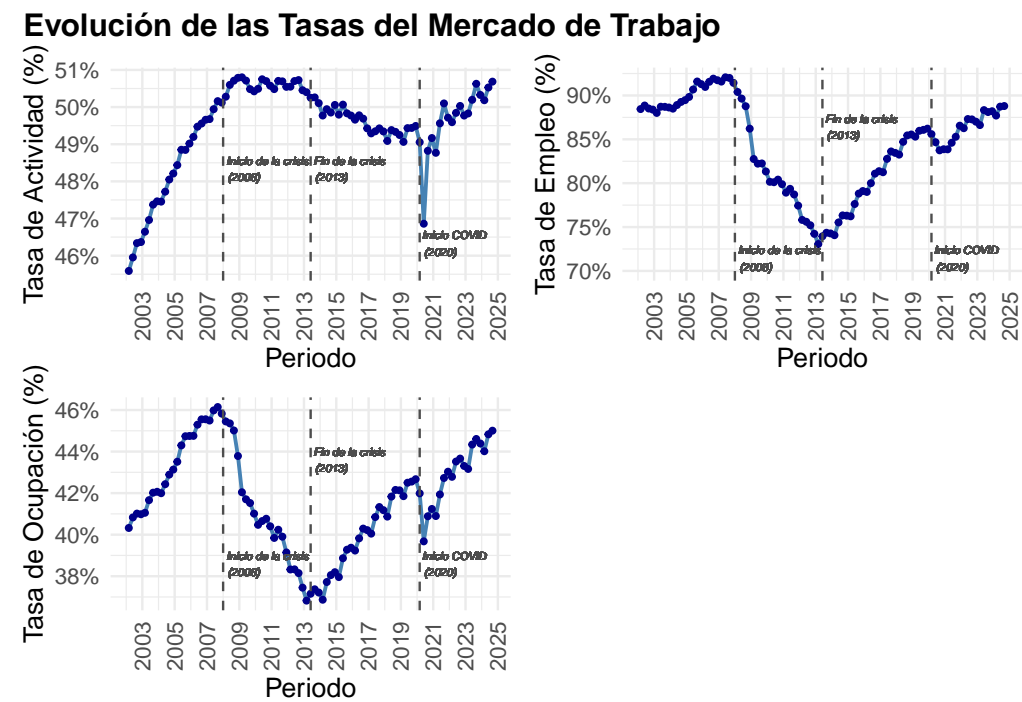
Una vez preparados, los datos se visualizan a través de gráficos y tasas que permiten interpretar patrones, tendencias y distribuciones de manera más clara y directa.

**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) =  $\frac{\text{Activos}}{\text{Población} \geq 16} \cdot 100$ 128
- Tasa de Inactividad (TI) =  $\frac{\text{Inactivos}}{\text{Población} \geq 16} \cdot 100 = 1 - TA$ 129
- Tasa de Ocupación (TO) =  $\frac{\text{Ocupados}}{\text{Población} \geq 16} \cdot 100$ 130
- Tasa de Empleo (TE) =  $\frac{\text{Ocupados}}{\text{Activos}} \cdot 100$ 131
- Tasa de Desempleo (TD) =  $\frac{\text{Parados}}{\text{Activos}} \cdot 100 = 1 - TE$ 132

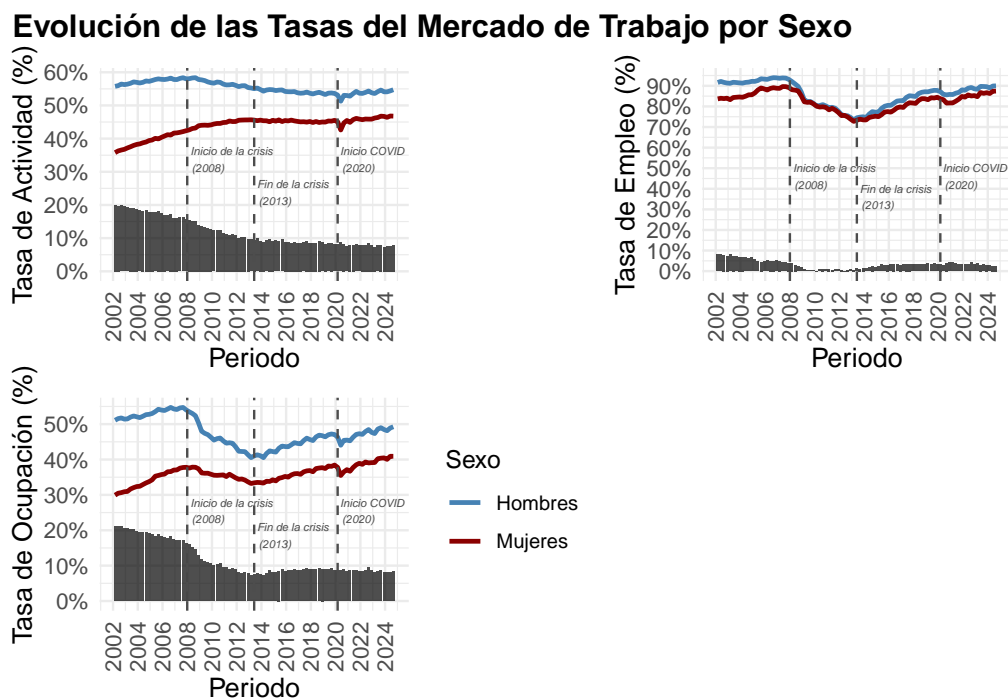


**Figure 1.** Evolución del mercado de trabajo en España

La Figura 1 muestra la evolución de la tasa de actividad, empleo y ocupación en España desde 2002 hasta la actualidad, reflejando cambios asociados a eventos económicos y sociales. Entre 2002 y 2008, todas las tasas crecieron sostenidamente gracias a la bonanza económica, mientras que la crisis de 2008 provocó una fuerte caída en las tasas de empleo y ocupación, aunque la tasa de actividad se mantuvo estable hasta 2013. En ese periodo, el empleo cayó a niveles del 75%, evidenciando un desempleo superior al 25%. Desde 2013, con signos de recuperación económica, las tasas de empleo y ocupación comenzaron a mejorar progresivamente, mientras que la tasa de actividad decreció ligeramente debido a factores como la reducción del abandono escolar. La pandemia de COVID-19 en 2020 generó caídas abruptas en todas las tasas, especialmente en la ocupación, aunque a partir de 2021 se observa una recuperación marcada por la reincorporación laboral y el dinamismo de sectores clave como el turismo y los servicios digitales.

En cambio, la Figura 2 muestra la evolución de las tasas de actividad, empleo y ocupación por género, evidenciando desigualdades y avances hacia la igualdad. La tasa de actividad masculina, aunque más alta, disminuyó entre 2008 y 2020, mientras que la femenina creció sostenidamente desde 2002, reduciendo la brecha de género del 20% al 8%. En la tasa de empleo, las diferencias son menores, ya que mide solo a la población activa, mientras que en la tasa de ocupación las disparidades son más evidentes al incluir a toda la población en edad de trabajar.

La pandemia de 2020 generó descensos significativos en todas las tasas, seguidos de una recuperación que refleja la resiliencia del mercado laboral. Este análisis resalta los



**Figure 2.** Evolución del mercado de trabajo por género en España

avances hacia la igualdad laboral, aunque persisten desigualdades estructurales, especialmente en la población activa y su impacto en la ocupación.

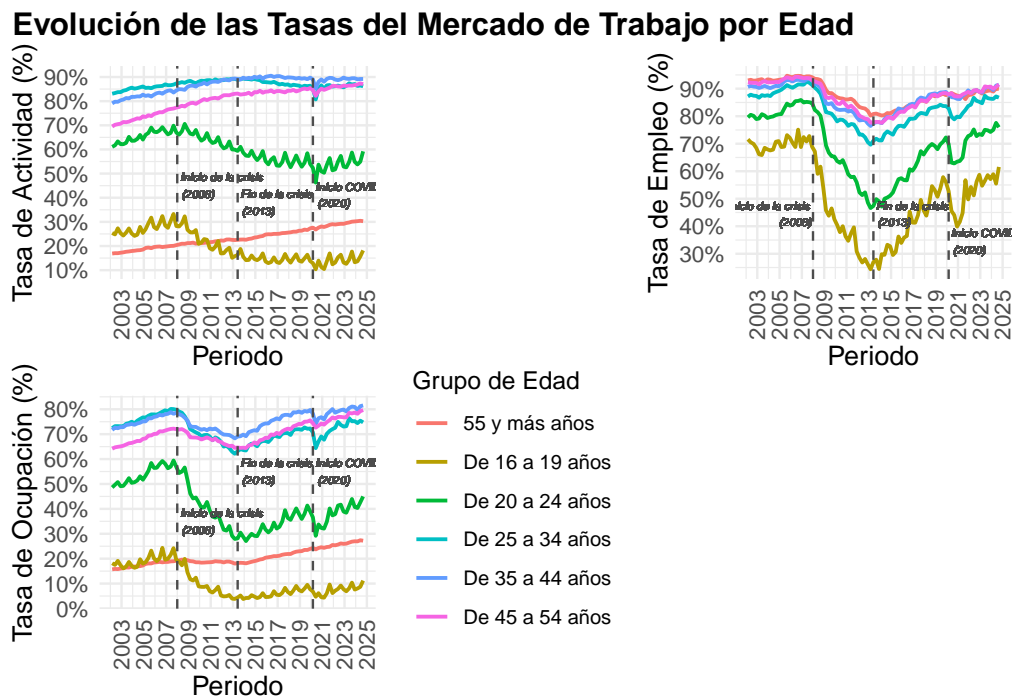
Por último, referenciando a la Figura 3 que desglosa la población por grupos de edad, se puede observar cómo los grupos de edad han experimentado dinámicas distintas en el mercado laboral. Los mayores de 35 a 54 años mantienen tasas de actividad superiores al 80 %, reflejando una integración estable. En contraste, los jóvenes de 16 a 24 años presentan tasas mucho más bajas, con un desempleo superior al 70 % en 2013, influido por la crisis en sectores como la construcción y el aumento de la escolarización.

Los grupos jóvenes también exhiben estacionalidad, especialmente en verano, cuando se incorporan al mercado laboral en sectores como la hostelería. Por otro lado, el grupo de 55 y más años presenta tasas de actividad y ocupación bajas debido a la jubilación, aunque esto no afecta de igual manera a la tasa de empleo, que excluye a los inactivos como los jubilados.

3.1.1. Clusterización por Comunidades y Ciudades Autónomas

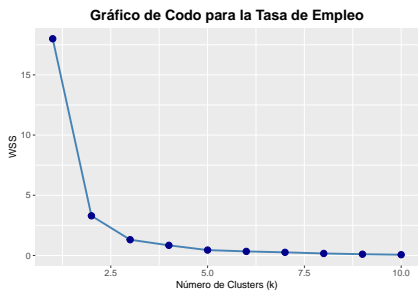
En este apartado del proyecto se va a proceder a realizar un clustering de la tasa de empleo para las distintas Comunidades y Ciudades Autónomas de España en los siguientes años: 2007, 2013 y 2023. El principal motivo por el cual escogemos esta tasa se debe a que se mide como el porcentaje de personas ocupadas respecto a la población activa, es decir, las que se encuentran activamente en el mercado laboral.

Como método de clusterización se utilizará el k-means, que organiza los datos en grupos basados en su similitud. Este algoritmo comienza eligiendo un número de clusters (k) y seleccionando centroides iniciales al azar. Cada punto se asigna al clúster cuyo centroide está más cercano, y luego los centroides se recalculan como el promedio de los puntos en cada grupo. Es un proceso que se repite hasta que los centroides dejan de cambiar o se alcanza un límite de iteraciones. Al finalizar, las comunidades quedan agrupadas en k clusters representando cada uno un conjunto de puntos con características similares, en este caso CCAA.



**Figure 3.** Evolución del mercado de trabajo por grupo de edad en España

Por otro lado, para la selección del hiperparámetro  $k$  se opta por el método del codo. Este consiste en graficar la suma de cuadrados dentro del clúster (WSS) frente a distintos valores de  $k$ . El punto donde la disminución de WSS se ralentiza significativamente, formando un “codo”, indica el número adecuado de clusters.



**Figure 4.** Gráfico de codo global para la tasa de empleo

Observando el gráfico de la Figura 4, se puede considerar que el número de clusters a partir del cual la WSS se ralentiza es en  $k=3$ , pues es en donde se forma el “codo”. Posteriormente, y como se ha mencionado al principio del apartado, vamos a hacer un clustering con el objetivo de agrupar las CCAA según su tasa de empleo en distintos periodos. Posteriormente los colores de los clusters se representan en un mapa de España dividido por dichas regiones para visualizar de forma clara y efectiva los patrones espaciales. La intención de esta medida es identificar patrones regionales y analizar las diferencias en las dinámicas del mercado laboral a nivel territorial.

La Figura 5 muestra el análisis de clustering de la tasa de empleo en España en 2007. Todas las comunidades de la península a excepción de Extremadura y Andalucía se agrupan en un clustering único debido a la similitud en su tasa, con valores superiores al 90% representando mercados laborales sólidos. Ceuta y Melilla, con tasas por debajo del 80%, pertenecen al clúster rojo, reflejando mayores desafíos laborales. El clúster verde, con

Análisis de Clustering y Mapa para Tasa de Empleo (2007)

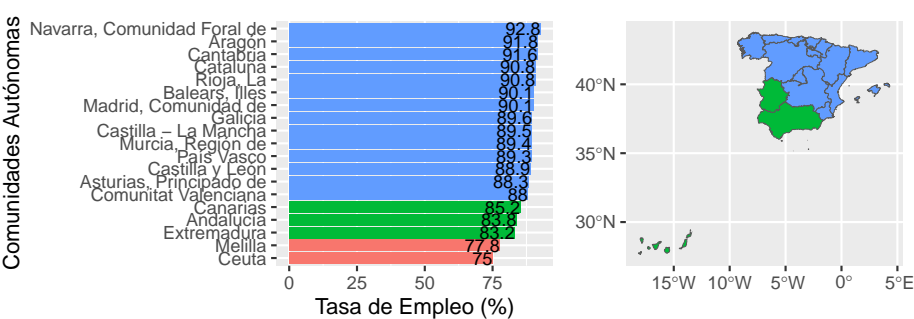


Figure 5. Clusterización regional con 3 centroides para la tasa de empleo en España en 2007

tasas entre el 80% y el 85%, incluye las regiones de Extremadura, Andalucía y Canarias, que muestran estabilidad moderada. En conjunto, la figura evidencia claras disparidades regionales en el empleo respecto al sur del país antes de la crisis económica.

Análisis de Clustering y Mapa para Tasa de Empleo (2013)

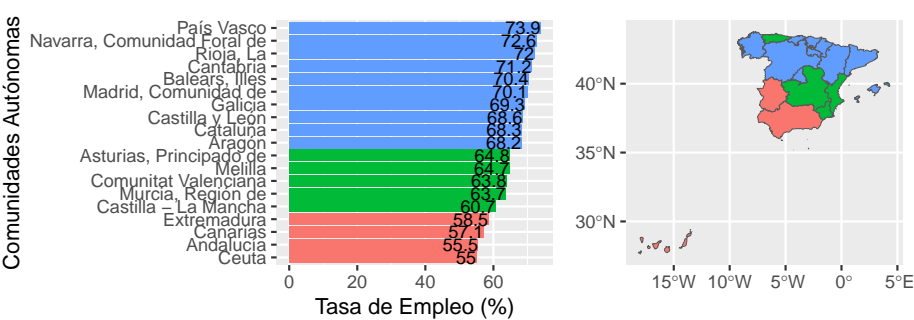
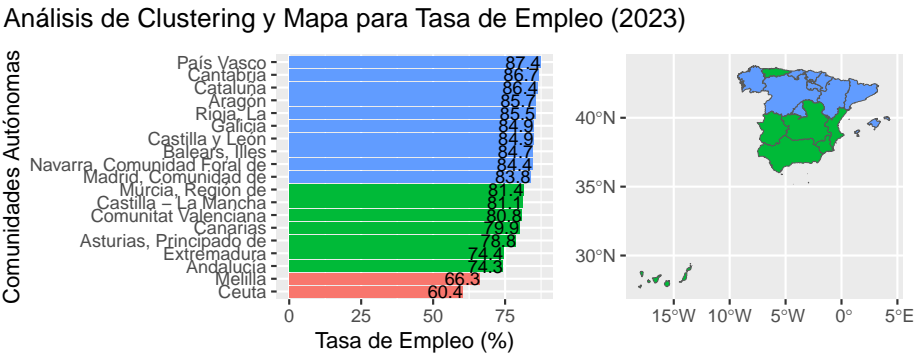


Figure 6. Clusterización regional con 3 centroides para la tasa de empleo en España en 2013

En cambio, la Figura 6 muestra el análisis de clustering de la tasa de empleo en España en 2013, el peor momento de la crisis económica. En comparación con 2007, las tasas de empleo han disminuido significativamente en todas las comunidades autónomas, reflejando el impacto generalizado de la recesión. Mientras que en 2007 gran parte de la península estaba agrupada en un único clúster, en 2013 se observa una mayor dispersión. El clúster azul mantiene las tasas más altas, superiores al 70%, mientras que el clúster rojo refleja las



más bajas, por debajo del 60%. Esto evidencia las desigualdades regionales acentuadas por la crisis.



**Figure 7.** Clusterización regional con 3 centroides para la tasa de empleo en España en 2023

En la actualidad, la Figura 7 muestra una recuperación general en las tasas de empleo respecto a 2013, aunque persisten desigualdades regionales. El clúster azul, que representa la mitad norte de España (a excepción de Asturias), destaca con tasas superiores al 83%, reflejando un mercado laboral sólido en esta área. El clúster verde, que representa la mitad sur, presenta tasas intermedias entre el 74% y el 82%, reflejando una mayor estabilidad en el empleo, aunque todavía sin alcanzar la uniformidad observada antes de la crisis. Por el contrario, el clúster rojo, que incluye las ciudades de Ceuta y Melilla, presenta tasas inferiores al 70%, evidenciando desafíos laborales persistentes.

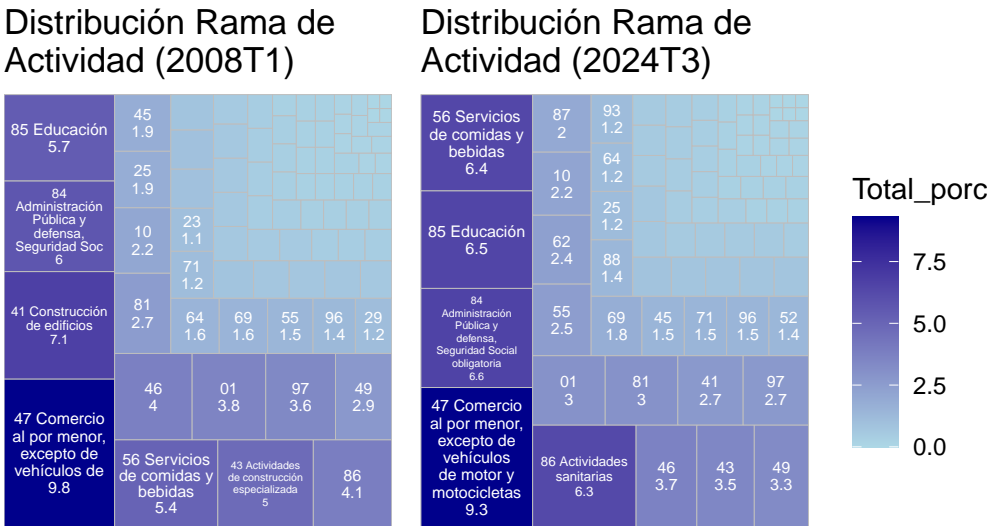
En conclusión, se evidencia una mejora general, pero con brechas regionales significativas.

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

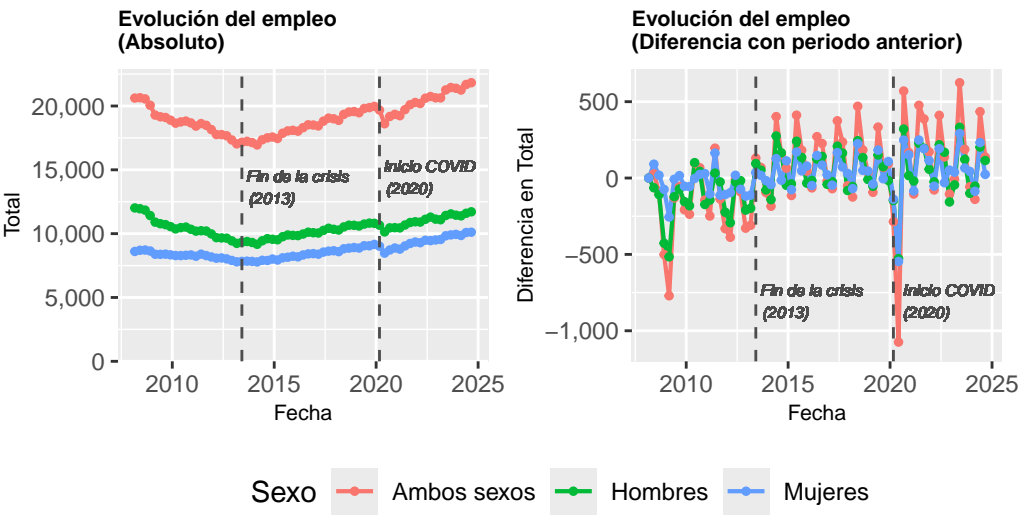
En primer lugar se analiza la distribución de peso de las ramas de actividad en el total de la población empleada. Además, se busca comparar la evolución y el cambio sufrido por la estructura laboral española desde el primer trimestre de 2008 y el tercer trimestre de 2024. Para ello se generan treemaps para los periodos inicial (2008T1) y final (2024T3) del análisis. Estos gráficos muestran cómo se distribuye el empleo entre las diferentes ramas económicas según su peso relativo.

En la Figura 8 podemos observar cómo algunas ramas han ganado o perdido relevancia en términos de empleo total a lo largo del tiempo. Por ejemplo, sectores relacionados con la educación y la administración pública han aumentado su proporción, mientras que ramas relacionadas con la construcción y la industria han experimentado una marcada disminución.

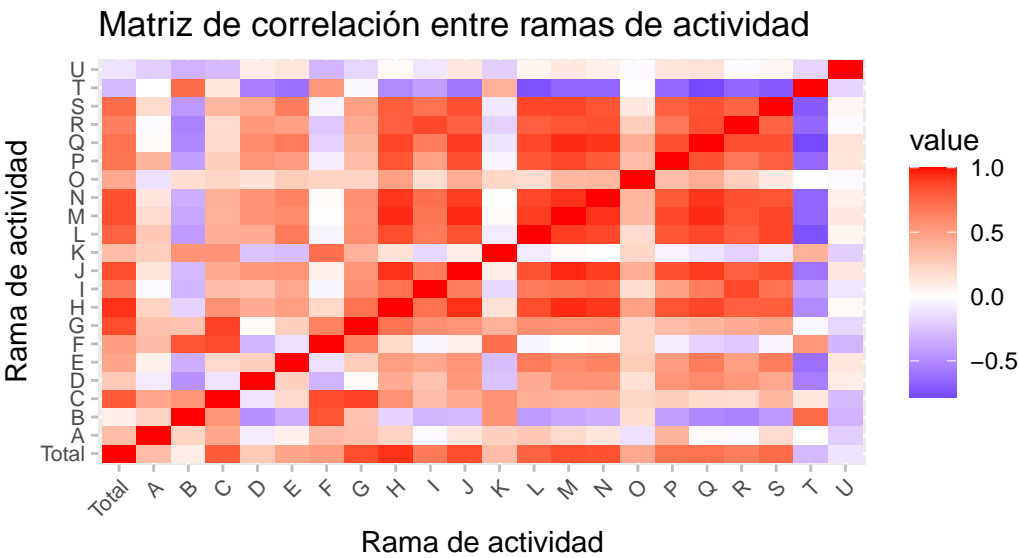
Además de realizar un análisis comparativo de cómo se distribuye por ramas de actividad la población trabajadora, se realiza un análisis visual de la evolución del empleo en España de manera absoluta agrupando por sexo, en conjunto con un análisis de la evolución de la diferencia trimestral en el empleo. Para ello se han empleado gráficos de líneas. Estos gráficos incluyen referencias visuales a eventos históricos relevantes, como el fin de la crisis económica de 2013 y el inicio de la pandemia en 2020.



**Figure 8.** Distribución de las ramas de actividad en el primer trimestre de 2008 y el tercer trimestre de 2024.



**Figure 9.** Evolución del empleo en España y de las diferencias ientre dos periodos consecutivos.



**Figure 10.** Matriz de correlación entre agrupaciones generales de ramas de actividad.

En la Figura 9 destaca cómo el empleo masculino y femenino responden de manera diferente a los shocks económicos. Por ejemplo, los hombres experimentaron caídas más pronunciadas durante la crisis de 2008, mientras que las mujeres mostraron una recuperación más gradual. Además, se puede apreciar que la diferencia entre sexos se redujo en gran medida durante la crisis de 2008, desde la cual la población masculina ha sido incapaz de recuperar las cifras precrisis, mientras que las mujeres han superado marcadamente dichos valores.

**3.2.1. Estudio de correlaciones entre Ramas de Actividad**

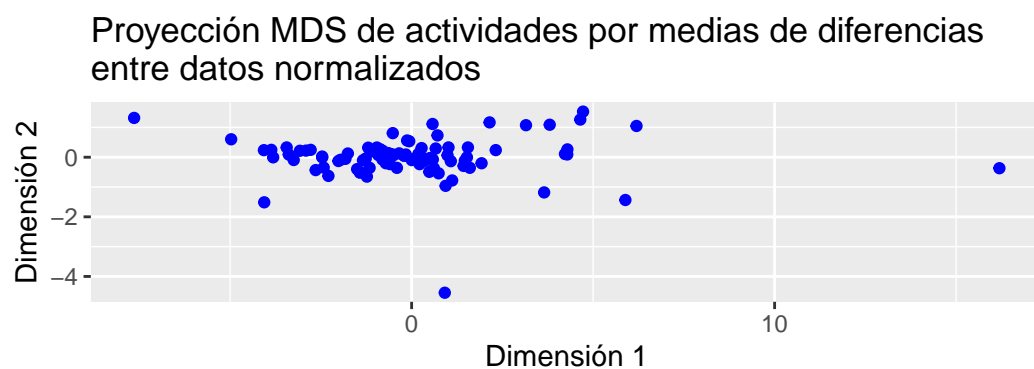
A continuación, se realiza un estudio de las correlaciones lineales de Pearson entre distintas ramas de actividad (en ambos sexos en conjunto). Los datos de CNAE vienen dispuestos en dos órdenes de agrupación, unos más generales, caracterizados por empezar por una letra en su nombre, y otros más concretos, empezando por un número. Realizaremos el estudio sobre los datos agrupados de forma más general, usando solo la letra como etiqueta; el nombre completo de la rama se puede consultar en el Apéndice.

En la matriz de correlación de la Figura 10 se contempla que la mayoría de las ramas parecen estar correlacionadas positivamente, salvo aquellos puestos relacionados con el personal doméstico e industrias extractivas con una correlación negativa con la mayoría de las ramas, así como aquellas relacionadas con la agricultura, ganadería, construcción, actividades financieras entre otras que no presentan una marcada correlación con el resto.

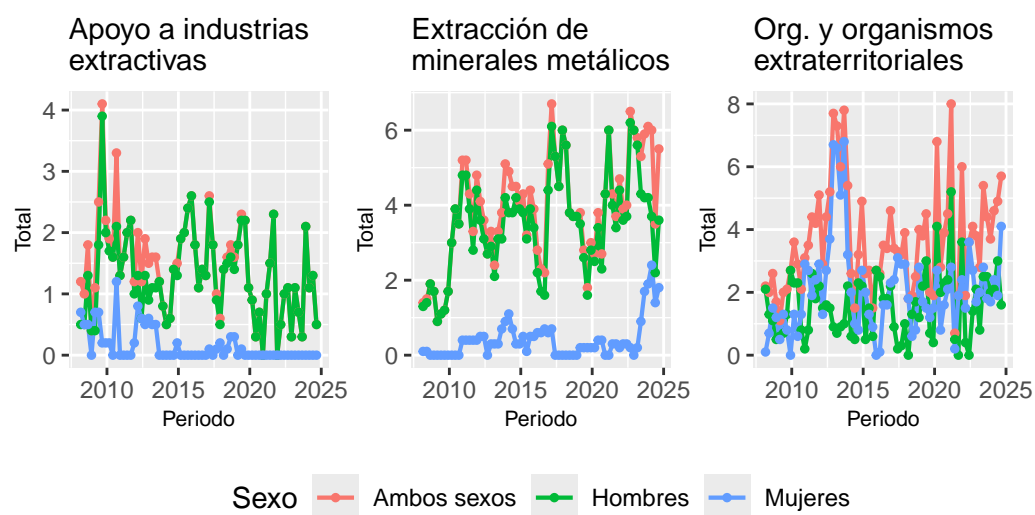
**3.2.2. Tendencias relativas a partir de datos normalizados**

En este punto, se busca estudiar las tendencias relativas, eliminando la influencia de las diferencias iniciales entre ramas. Para ello, primero se normalizan los datos dividiendo por el número inicial de trabajadores en cada rama, y a continuación se obtiene una métrica de distancia entre dichos datos normalizados, obteniéndose como la media de las diferencias absolutas de las cantidades normalizadas. A partir de estas distancias, se proyectan en dos dimensiones utilizando MDS, lo que permite visualizar las relaciones entre ramas en función de sus patrones de empleo a lo largo del tiempo. La proyección MDS es escogida frente a otros métodos debido a su enfoque en conservar las distancias establecidas.

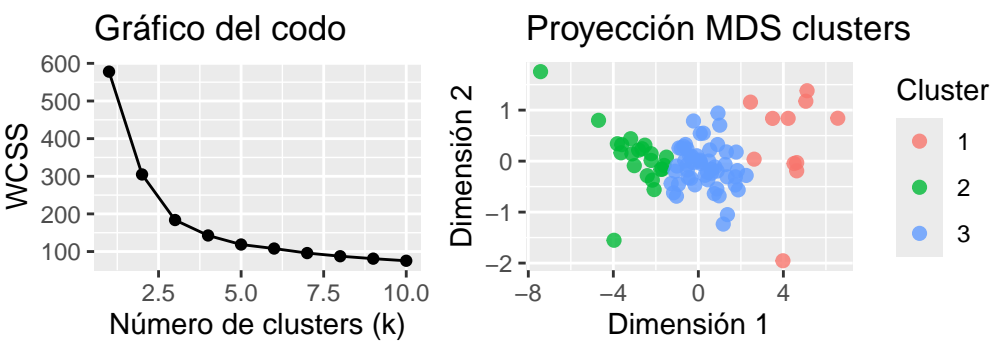
Se forman agrupaciones naturales de ramas con patrones similares en la Figura 11. Además, se observan claramente algunos outliers, que puede ser de gran interés analizarlos individualmente.



**Figure 11.** Proyección MDS de las distancias entre la evolución de ramas de actividad normalizadas.



**Figure 12.** Evolución del empleo en posibles outliers de ramas de actividad.



**Figure 13.** Gráfico del método del codo y representación mediante una proyección MDS de los clusters de k-means, con el parámetro  $k$  escogido.

Ramas como “Extracción de minerales metálicos”, “Actividades de apoyo a las industrias extractivas” o “Actividades de organizaciones y organismos extraterritoriales” se destacan en la Figura 12 por su comportamiento atípico. Estas ramas tienen una dinámica de empleo inusual debido a su naturaleza específica, tamaño reducido o dependencia de factores externos como la demanda global. Observamos claramente que son datos con mucho ruido, y por lo tanto es normal que nos apareciesen como outliers.

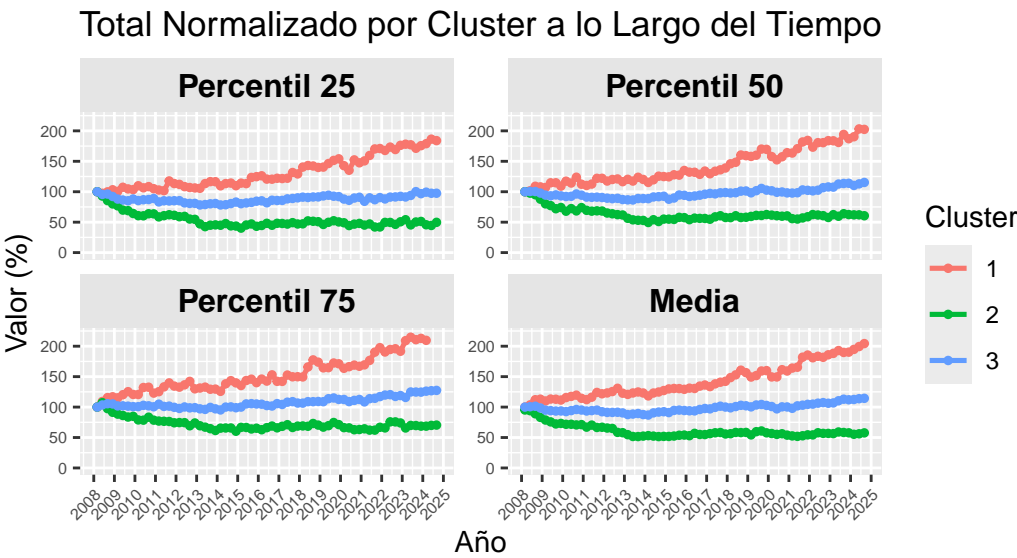
3.2.3. Clusterización por Ramas de Actividad

Tras observar la proyección tratamos de segmentar los datos en agrupaciones naturales, es por ello que tratamos de hacer un clustering que nos permita agrupar cuantitativamente aquellas ramas de actividad que han tenido un comportamiento similar entre ellas (con los datos normalizados). Para ello se opta en primer lugar por el algoritmo más típico en clustering (k-means), y se analiza sus resultados para ver si son satisfactorios. Además, como método para la selección del hiperparámetro  $k$ , se opta por el método del codo. Sin embargo, este resultado se analizará para saber si es adecuado, si sería necesario alterar ligeramente el número de clusters obtenidos o si directamente deberíamos escoger otro método.

El método del codo ilustrado en la Figura 13 parece indicarnos que la mejor selección es  $k = 3$  y observamos que los resultados parecen satisfactorios. Por lo tanto, debido a la simplicidad del modelo nos quedamos con esta elección. Pasamos a estudiar por separado el comportamiento de cada clúster y trataremos de averiguar si tienen una interpretación natural clara.

Analizando los cuantiles y medias de cada clúster presentes en la Figura 14, se observa que se agrupan según rendimiento, cosa ya esperada proviniendo de la métrica definida. Se observa que la mayoría de las ramas de actividad se acumulan en los clústeres con la evolución negativa o mediocre, en los cuales ha habido un decrecimiento (clúster 2) o estancamiento (clúster 3) de los trabajadores. Por otra parte el clúster 1 parece presentar ramas con resultados positivos, aunque con menor número de ramas que los anteriores.

Podemos observar que aquellos sectores relacionados con la descontaminación o la programación y la informática han avanzado mucho en los últimos años, mientras que los sectores más relacionados con la construcción han caído fuertemente. Sin embargo, la conclusión general es que la evolución del empleo en España durante los últimos años ha sido marcadamente negativa para la mayoría de las ramas de actividad, salvándose de ella ejemplos muy concretos.



**Figure 14.** Evolución de los percentiles 25, 50, 75 y media de la evolución normalizada agrupados por clúster.

**Appendix D**

Letra	Descripción
A	Agricultura, ganadería, silvicultura y pesca
B	Industrias extractivas
C	Industria manufacturera
D	Suministro de energía eléctrica, gas, vapor y aire acondicionado
E	Suministro de agua, actividades de saneamiento, gestión de residuos y descont.
F	Construcción
G	Comercio al por mayor y al por menor y reparación de vehículos de motor
H	Transporte y almacenamiento
I	Hostelería
J	Información y comunicaciones
K	Actividades financieras y de seguros
L	Actividades inmobiliarias
M	Actividades profesionales, científicas y técnicas
N	Actividades administrativas y servicios auxiliares
O	Administración Pública y defensa, Seguridad Social obligatoria
P	Educación
Q	Actividades sanitarias y de servicios sociales
R	Actividades artísticas, recreativas y de entretenimiento
S	Otros servicios
T	Actividades de los hogares como empleadores de personal doméstico
U	Actividades de organizaciones y organismos extraterritoriales

**Table A7.** Clasificación de Ramas de Actividad según letra

**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.