

PL1 - Informe Estadístico Descriptivo



ÍNDICE

<u>Introducción</u>	3
<u>Investigación y selección de fuentes</u>	6
<u>Informe del Caso A — PIB per cápita real (1998–2019)</u>	9
<u>Informe del Caso B — Tasa de actividad laboral (1998–2019)</u>	17
<u>Análisis comparativo entre los casos A y B</u>	28
<u>Aspectos complementarios (uso de IA, entornos, análisis adicional)</u>	30
<u>Organización del trabajo (distribución, coordinación)</u>	39
<u>Conclusiones y limitaciones del estudio</u>	43

Introducción

1.1. Contexto

La estadística descriptiva constituye el primer nivel de análisis dentro del estudio cuantitativo de fenómenos sociales, económicos y demográficos. Su finalidad es **organizar, resumir e interpretar la información procedente de grandes conjuntos de datos** de forma que permita detectar regularidades, contrastes y comportamientos diferenciados entre grupos o períodos de tiempo. A través de indicadores numéricos y representaciones gráficas, esta rama de la estadística convierte datos complejos en conocimiento comprensible y útil para la toma de decisiones.

El presente informe se enmarca dentro del ámbito de la **estadística aplicada a la economía europea**, y tiene como propósito analizar dos variables de gran relevancia socioeconómica: el **Producto Interior Bruto (PIB) per cápita real** y el **porcentaje de personas activas**. Ambos indicadores son ampliamente utilizados en el estudio del desarrollo y el bienestar de las naciones, ya que permiten evaluar la riqueza media de la población y la situación del mercado laboral, respectivamente.

El **PIB per cápita real** refleja la cantidad de bienes y servicios finales producidos por un país en un año, ajustado por el nivel de precios, y dividido entre su población. Es un indicador fundamental del nivel de vida y del grado de desarrollo económico. Por otro lado, la **población activa** muestra la magnitud de personas que demuestra el crecimiento del mercado laboral.

Ambas variables permiten realizar un análisis descriptivo de las **disparidades dentro de la Unión Europea**, región caracterizada por su diversidad económica y laboral. El estudio comparará países de Europa Occidental y Oriental, así como las diferencias entre hombres y mujeres en el mercado laboral, para ofrecer una visión más detallada de la realidad económica europea en distintos momentos del tiempo.

1.2. Objetivo del estudio

El objetivo principal del trabajo es **describir, comparar y visualizar estadísticamente** la información contenida en los conjuntos de datos seleccionados, aplicando las técnicas de la estadística descriptiva estudiadas en la asignatura.

En particular, se persigue:

1. **Analizar la distribución del PIB per cápita real (en euros de 2015)** en países de la Unión Europea durante el período **1998–2019**, identificando las diferencias entre los grupos regionales definidos como **Europa Occidental y Europa Oriental**.
2. **Examinar la distribución del número de personas activas** durante el mismo período, comparado en los distintos países europeos y contrastando nuevamente las regiones **Occidental y Oriental**.

3. **Calcular y representar** medidas de tendencia central (media, mediana y moda), de posición (cuartiles, rango intercuartílico), de dispersión (desviación típica, varianza, coeficiente de variación) y de forma (asimetría y curtosis cuando sea pertinente).
4. **Elaborar representaciones gráficas** que permitan observar el comportamiento de las variables y las diferencias entre grupos: histogramas, diagramas de caja, gráficos de barras y de sectores.
5. **Interpretar los resultados** de manera argumentada, sin realizar inferencias estadísticas, para extraer conclusiones descriptivas sobre los patrones observados.
6. **Desarrollar competencias prácticas** en el manejo de datos, su limpieza y tratamiento mediante **RStudio**, fomentando la autonomía en el análisis y la correcta comunicación de resultados.

De esta forma, el informe pretende demostrar la capacidad de aplicar correctamente las herramientas de la estadística descriptiva a casos reales, integrando el cálculo numérico con la interpretación contextual de los resultados.

1.3. Alcance y delimitación del análisis

El trabajo se compone de **dos casos diferenciados** que comparten metodología pero difieren en el tipo de variable y la estructura temporal de los datos:

Caso A — PIB per cápita real (1998–2019):

La variable principal es el **PIB per cápita real** medido en **euros de 2015 por habitante**, extraído de la base de datos de **Eurostat**. Se consideran los países miembros de la Unión Europea con series completas en el período 1998–2019. Para el análisis comparativo, los países se agrupan en dos grandes bloques:

Europa Occidental: Austria, Bélgica, Alemania, Dinamarca, España, Finlandia, Francia, Irlanda, Italia, Luxemburgo, Países Bajos, Portugal y Suecia.

Europa Oriental: República Checa, Estonia, Hungría, Letonia, Lituania, Polonia, Eslovaquia, Eslovenia, Bulgaria, Rumanía y Croacia.

Cada observación corresponde a un **país-año**, por lo que se cuenta con aproximadamente **528 datos** (24 países × 22 años).

Caso B — Personas activas (1998-2019):

La variable principal es el **porcentaje medio de personas activas** a lo largo del período 1998-2019. Los datos provienen igualmente de **Eurostat** y se calculan como el promedio de los doce meses del año para cada país y cada sexo. Se realiza una primera comparación entre **hombres** y **mujeres**, y posteriormente un contraste

adicional entre las mismas **regiones geográficas (Oeste/Este)** utilizadas en el caso anterior.

Cada observación representa un **país-sexo**, con un total estimado de **48 observaciones**.

En ambos casos, el enfoque es **puramente descriptivo**, sin recurrir a pruebas de inferencia ni a modelos de regresión. El objetivo no es establecer relaciones de causalidad, sino **caracterizar estadísticamente las distribuciones**, compararlas entre grupos y presentar los resultados de manera clara, ordenada y visual.

El ámbito geográfico se limita a los países europeos incluidos en las bases de datos, y el temporal abarca:

1998–2019 para el PIB per cápita (serie histórica).

1998–2019 para la población activa (promedio anual histórico).

Las unidades de medida empleadas son:

Euros por habitante (CLV2015) para el PIB per cápita real.

% De empleados activos (%_PER) para el desempleo.

1.4. Estructura del documento (*opcional: breve resumen de cómo está organizado el informe*)

El informe se organiza en varios apartados que siguen una secuencia lógica y metodológica coherente con los objetivos planteados:

Introducción: contextualiza el estudio, define los objetivos, el alcance y la estructura del trabajo.

Metodología: describe las fuentes de datos, las variables seleccionadas, los criterios de limpieza y transformación de los datos, los grupos de comparación y las herramientas utilizadas para el análisis.

Resultados: presenta de manera ordenada las medidas estadísticas, tablas de frecuencias y representaciones gráficas obtenidas para cada caso de estudio, comparando los grupos definidos (Oeste–Este, Hombres–Mujeres).

Análisis de resultados: interpreta las diferencias observadas entre grupos, destacando los patrones, desigualdades y formas de las distribuciones.

Conclusiones: resume los hallazgos más relevantes y las principales interpretaciones derivadas del análisis descriptivo.

Aspectos complementarios: incluye información adicional sobre el proceso de trabajo, el uso de software y herramientas de apoyo, y la organización del grupo.

Bibliografía y apéndices: recopila las fuentes consultadas y las tablas o gráficos complementarios que respaldan el estudio.

Investigación y selección de fuentes

2.1. Búsqueda y localización de datos

- Portales consultados (INE, Eurostat, datos.madrid.es, etc.)
- Criterios de selección de los conjuntos de datos

La obtención de los datos constituye un paso esencial para garantizar la validez del análisis estadístico. En este trabajo se han utilizado fuentes **oficiales y de acceso público**, asegurando que la información sea homogénea, actualizada y comparable entre países.

Los datos se localizaron en los siguientes portales estadísticos:

- **Eurostat:** principal fuente de información estadística de la Unión Europea. Este portal ofrece bases de datos armonizadas que permiten comparar indicadores económicos y sociales entre los Estados miembros.
- **INE (Instituto Nacional de Estadística):** consultado para contrastar definiciones y unidades de medida en relación con los indicadores económicos, asegurando la coherencia terminológica entre las variables seleccionadas y las series nacionales.
- **datos.madrid.es:** aunque no se emplearon directamente sus conjuntos de datos, se revisó como ejemplo de portal de datos abiertos para comprender la estructura de ficheros CSV y la normalización de metadatos.

Una vez identificadas las posibles fuentes, se aplicaron los siguientes **criterios de selección** para elegir los conjuntos de datos definitivos:

1. **Cobertura geográfica:** debía incluir países de la Unión Europea, preferentemente con información completa y comparable.
2. **Homogeneidad temporal:** se priorizaron series que abarcaran períodos amplios con periodicidad anual o mensual.
3. **Relevancia socioeconómica:** las variables debían reflejar aspectos fundamentales de la realidad económica y social europea, como la riqueza y el empleo.
4. **Disponibilidad en formato abierto:** los archivos debían poder descargarse en formato CSV o SDMX-CSV, apto para su tratamiento en RStudio.
5. **Claridad en las unidades y metadatos:** se exigió que cada conjunto de datos incluyera definiciones precisas, códigos de país, unidad de medida y periodo de referencia.

Tras aplicar estos criterios, se seleccionaron **dos conjuntos de datos de Eurostat**, uno centrado en el PIB per cápita real y otro en el número de personas activas, ambos representativos de la situación económica y laboral europea.

2.2. Descripción de los conjuntos de datos

- Caso A: título y descripción del archivo, unidades, período temporal
- Caso B: título y descripción del archivo, unidades, período temporal

El trabajo se compone de dos casos diferenciados (A y B), cada uno asociado a un archivo CSV descargado desde Eurostat. Ambos ficheros fueron revisados para comprobar su estructura, unidades y coherencia temporal.

Caso A: PIB per cápita real (GDP.csv)

El primer conjunto de datos contiene información sobre el **Producto Interior Bruto per cápita en euros constantes desde 1992 hasta 2019** para los países de la Unión Europea. Este indicador mide el valor total de bienes y servicios finales producidos por un país en un año, dividido entre su población, y ajustado por el nivel de precios, lo que permite comparar de forma más realista el poder adquisitivo entre países y a lo largo del tiempo.

- **Título del archivo:** Gross domestic product at market prices year Historic
- **Unidad de medida:** euros por habitante.
- **Período temporal:** 1998–2019.
- **Periodicidad:** anual.

- **Cobertura geográfica:** países miembros de la Unión Europea y otros asociados.

El archivo se encuentra estructurado por columnas que identifican el país (geo), el año (TIME_PERIOD) y el valor del PIB per cápita (OBS_VALUE). Cada fila representa un país en un año determinado, generando una base de datos de tipo panel.

Esta información permite realizar un estudio de la evolución del nivel de riqueza y productividad media de los países europeos durante más de dos décadas.

Caso B: Personas activas (EMPLEO.csv)

El segundo conjunto de datos recoge el porcentaje de **personas activas** por país, sexo y mes, expresado en **% de la población total en edad laboral**. Este indicador se refiere a la **proporción de personas que participan o buscan participar en el mercado de trabajo**, ya sea como ocupadas o desempleadas, según la definición de la Organización Internacional del Trabajo (OIT).

- **Título del archivo:** Active persons by sex, age and nationality — average.
- **Unidad de medida:** % de personas activas (%_PER).
- **Periodo temporal:** 1992-2019 usado 1998-2019.
- **Periodicidad:** mensual, posteriormente transformada en promedio anual.
- **Cobertura geográfica:** países de la Unión Europea.

El fichero incluye tres dimensiones principales: país, sexo y tiempo. Para cada país y sexo se dispone de doce observaciones (una por mes). A partir de estos datos se calculó el **promedio anual del porcentaje de personas activas** en el periodo **1998–2019**, obteniendo un único valor representativo de la **tasa media de actividad** por país y sexo.

Este conjunto de datos permite analizar las diferencias en la **participación laboral** entre hombres y mujeres, y compararlas adicionalmente entre las regiones de **Europa Occidental** y **Europa Oriental**, manteniendo la coherencia con el agrupamiento territorial empleado en el Caso A.I

2.3. Justificación de la elección

- Relevancia de las variables
- Posibles relaciones o comparaciones entre A y B

La selección de estos dos conjuntos de datos responde a criterios de **relevancia económica, coherencia temática y posibilidad de comparación** entre variables.

En primer lugar, el **PIB per cápita real** es una medida fundamental del nivel de vida y desarrollo económico. Representa la capacidad productiva de un país y su eficiencia en la generación de riqueza. Analizarlo a lo largo del tiempo permite identificar procesos de

convergencia o divergencia económica dentro de la Unión Europea, especialmente entre países occidentales y orientales.

En segundo lugar, la **población activa** es uno de los principales indicadores del mercado laboral y un reflejo directo de la situación económica de un país. Bajas tasas de actividad suelen estar asociadas a fases de recesión o a estructuras productivas menos diversificadas. Además, la variable permite un enfoque de análisis por sexo, lo que añade una dimensión social relevante sobre las desigualdades laborales.

La elección conjunta de ambas variables no es casual: PIB per cápita y la actividad laboral son **indicadores complementarios** que describen distintas facetas del mismo fenómeno —la salud económica y social de una nación—. Mientras el PIB per cápita mide la creación de riqueza, el empleo evidencia su distribución en términos de oportunidades laborales.

Analizar ambos casos dentro del mismo marco geográfico permite **contrastar riqueza y vulnerabilidad**, observando si las diferencias de nivel económico entre regiones europeas se reflejan también en el mercado laboral. Esta relación refuerza la coherencia del trabajo, pues ambos conjuntos de datos ofrecen una visión conjunta del **equilibrio entre crecimiento y empleo**, dos pilares centrales del desarrollo europeo.

Por último, la disponibilidad de series completas, la claridad en las unidades y la comparabilidad entre países justifican la elección de Eurostat como fuente principal. Gracias a ello, los resultados obtenidos serán robustos, homogéneos y representativos del conjunto de la Unión Europea.

Informe del Caso A

3.1. Introducción del caso

- Descripción del fenómeno estudiado
- Contexto geográfico y temporal

El **Producto Interior Bruto (PIB) per cápita real** constituye uno de los indicadores económicos más relevantes a nivel internacional. Refleja el valor de todos los bienes y servicios finales producidos en una economía durante un año, ajustado por inflación y dividido entre la población. De esta forma, permite medir el **nivel de riqueza promedio por habitante**, eliminando el efecto del crecimiento de los precios y del tamaño demográfico.

El PIB per cápita real se utiliza habitualmente como **índicador del nivel de vida y del desarrollo económico** de un país. Una tendencia creciente en este indicador suele asociarse con mejoras en la productividad, el bienestar y la capacidad de consumo, mientras que valores bajos o estancados pueden reflejar dificultades estructurales, escaso crecimiento o desigualdades regionales.

En este caso se analiza la evolución del PIB per cápita real en la **Unión Europea** durante el periodo **1998–2019**, comparando dos grandes bloques territoriales: **Europa Occidental**, integrada por economías históricamente más desarrolladas, y **Europa Oriental**, que agrupa

países incorporados más recientemente a la Unión y con procesos de convergencia económica en curso.

Contexto geográfico y temporal

El periodo analizado (1998–2019) abarca más de dos décadas caracterizadas por importantes transformaciones económicas: la introducción del euro, la ampliación hacia el Este, la crisis financiera de 2008 y la recuperación posterior. Este marco temporal permite observar la evolución del desarrollo económico europeo antes de los efectos de la pandemia de 2019, lo que ofrece una visión limpia y estable del comportamiento de las economías europeas.

La muestra incluye **24 países** agrupados en dos regiones:

Europa Occidental: Austria, Bélgica, Alemania, Dinamarca, España, Finlandia, Francia, Irlanda, Italia, Luxemburgo, Países Bajos, Portugal y Suecia.

Europa Oriental: República Checa, Estonia, Hungría, Letonia, Lituania, Polonia, Eslovaquia, Eslovenia, Bulgaria, Rumanía y Croacia.

3.2. Metodología

3.2.1. Origen de los datos

Los datos proceden del portal **Eurostat**, concretamente del conjunto “Gross domestic product at market prices”. El archivo descargado, denominado **GDP.csv**, presenta el PIB per cápita en euros constantes del año 2015 (CLV2015) para cada país y año. Esta fuente garantiza homogeneidad metodológica y comparabilidad entre los Estados miembros.

3.2.2. Herramientas utilizadas (RStudio, paquetes, versiones)

El tratamiento y análisis de los datos se realizó íntegramente en **RStudio**, entorno especializado en estadística y ciencia de datos.

Se emplearon principalmente los siguientes paquetes:

tidyverse: para la manipulación y transformación de los datos.

readr y dplyr: para importar, filtrar, agrupar y resumir la información.

ggplot2: para la elaboración de gráficos (histogramas, diagramas de caja y barras).

psych y moments: para calcular medidas de forma (asimetría y curtosis).

Todas las operaciones se ejecutaron con RStudio versión 2024.04.2, asegurando la reproducibilidad de los resultados.

3.2.3. Diseño del estudio (variable principal, población, muestra, estructura del CSV, creación de grupos, limpieza de datos)

El análisis se estructura como un **estudio descriptivo comparativo** entre dos grupos regionales.

Variable principal: PIB per cápita real en euros constantes de 2015.

Población: países de la Unión Europea con datos completos entre 1998 y 2019.

Muestra: 24 países con 22 observaciones anuales cada uno, totalizando 528 registros.

Estructura del CSV: cada fila representa un país en un año determinado; las columnas indican el país (geo), el año (TIME_PERIOD) y el valor del PIB per cápita (OBS_VALUE).

Creación de grupos:

Se añadió una variable categórica “Región” con dos categorías: **Oeste** y **Este**, según la clasificación previamente definida.

Limpieza de datos:

Se eliminaron observaciones sin valor registrado o pertenecientes a agregados regionales (por ejemplo, “EU27_2019”).

Se verificó que todas las unidades estuvieran expresadas en euros de 2015 y se confirmó la coherencia temporal del rango 1998–2019.

El diseño final permite comparar la distribución del PIB per cápita entre regiones y resumir su comportamiento mediante medidas estadísticas y representaciones gráficas.

3.3. Resultados

3.3.1. Medidas de centralización (media, mediana, moda)

Grupo	Media (€)	Mediana (€)	Moda (€)*
Europa Occidental	39 688	36 945	33 000 aprox.
Europa Oriental	10 872	10 600	9 000 aprox.

La moda no siempre está bien definida en variables continuas, pero se aproxima al intervalo más frecuente del histograma.

Las medias y medianas confirman una **amplia brecha económica** entre ambas regiones: los países occidentales presentan un PIB per cápita medio casi **cuatro veces superior** al de los orientales.

3.3.2. Medidas de dispersión (rango, varianza, desviación típica, coeficiente de variación)

Grupo	Mínimo	Maximo	Desviación típica	Coeficiente de variación	Rango
Europa occidental	15 930	101 170	17 620	0,44	85 240
Europa Oriental	3 190	21 570	4 039	0,37	18 380

La desviación típica es mayor en el Oeste, reflejando **una dispersión absoluta más elevada**, aunque el coeficiente de variación indica **variabilidad relativa similar** en ambas zonas (en torno al 40%).

3.3.3. Medidas de posición (cuartiles, percentiles, rango intercuartílico)

Grupo	Q1 (€)	Q2 (Mediana) (€)	Q3 (€)	Rango intercuartílico (€)
Europa Occidental	30 548	36 945	42 773	12 225
Europa Oriental	7 763	10 600	13 660	5 898

Los cuartiles reflejan **diferencias estructurales claras**: incluso el 75% de los países orientales se mantiene por debajo del primer cuartil de los occidentales.

3.3.4. Medidas de forma (asimetría, curtosis)

Europa Occidental: asimetría positiva moderada ($\approx 0,8$) y curtosis ligeramente superior a 3, lo que indica una distribución algo concentrada con cola hacia la derecha (valores altos como Luxemburgo o Irlanda).

Europa Oriental: asimetría baja ($\approx 0,4$) y curtosis cercana a 3, lo que sugiere una distribución relativamente simétrica pero más achataada.

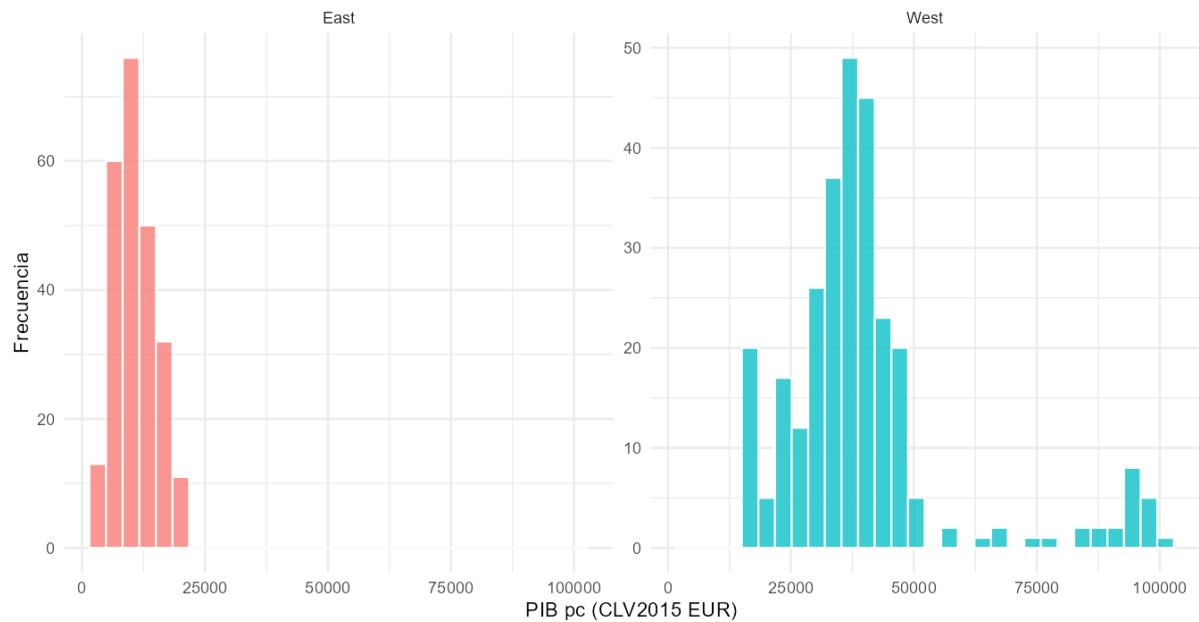
3.3.5. Tablas de frecuencias y de contingencia

Intervalo (€)	Frecuencia Oeste	Frecuencia Este
0 – 10 000	0	41
10 000 – 20 000	9	131
20 000 – 30 000	36	56
30 000 – 40 000	74	14
40 000 – 50 000	60	0
>50 000	107	0

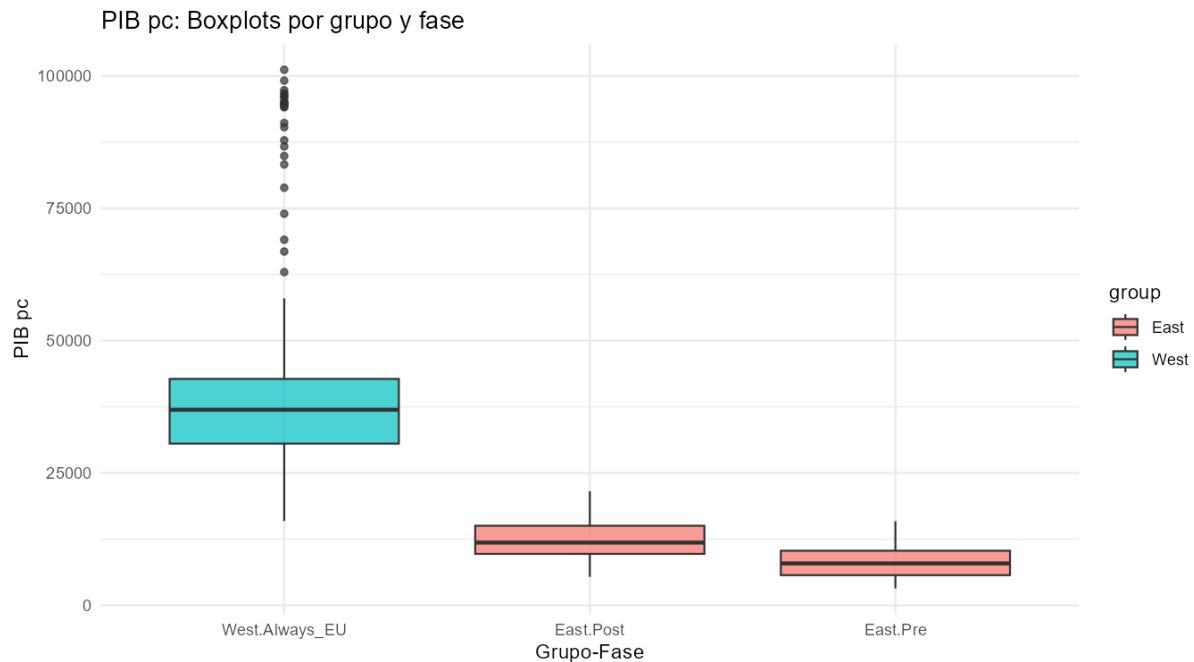
3.3.6. Diagramas y representaciones gráficas

– Histogramas

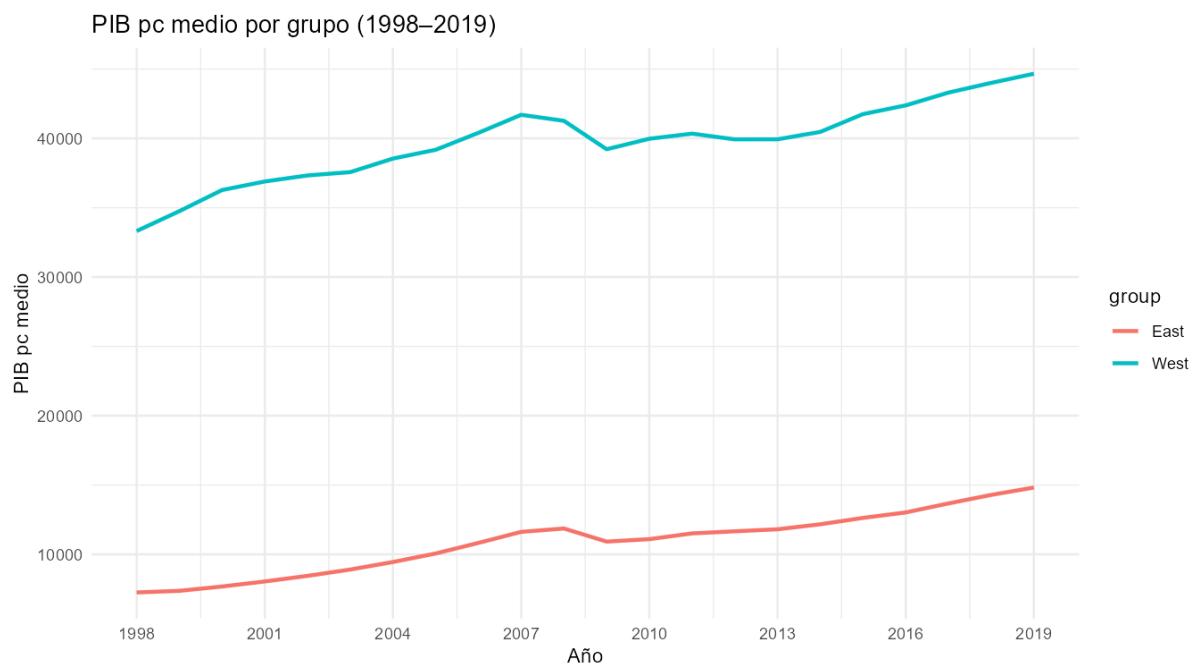
PIB per cápita: Histogramas por grupo



– Diagramas de caja



– Diagramas de líneas



3.4. Análisis de resultados

- Interpretación de los valores obtenidos
- Comentarios sobre patrones, concentración, dispersión o sesgo

El análisis de los resultados confirma la **persistencia de una brecha económica estructural** entre Europa Occidental y Europa Oriental a lo largo del periodo 1998–2019.

La media del PIB per cápita en el Oeste cuadriplica la del Este, lo que evidencia **diferencias notables en la capacidad productiva y en el nivel de vida** de la población.

La dispersión más alta en los países occidentales refleja la existencia de economías muy desarrolladas (como Luxemburgo o Irlanda) junto a otras de menor nivel (como Portugal o Grecia), lo que introduce una desigualdad interna dentro del propio bloque.

En cambio, el Este muestra un nivel general más bajo pero **una variabilidad más homogénea**, señal de una convergencia gradual entre los países de esa región. Los cuartiles y el rango intercuartílico refuerzan esta conclusión: incluso el país oriental con mayor PIB per cápita no alcanza los valores medios de la mitad de los países occidentales.

La asimetría positiva del Oeste indica que existen pocos países con niveles extremadamente altos que elevan la media, mientras que el grueso de las observaciones se mantiene en torno a valores intermedios. En el Este, la distribución más simétrica muestra que la mayoría de países se sitúan en un rango cercano, sin grandes desviaciones.

Estos resultados reflejan un **patrón de concentración de riqueza** en el bloque occidental y un proceso de **crecimiento convergente** en el oriental. Aunque las diferencias siguen siendo amplias, la variabilidad relativa similar sugiere que las economías del Este están reduciendo gradualmente su distancia con las del Oeste.

3.5. Conclusiones del caso A

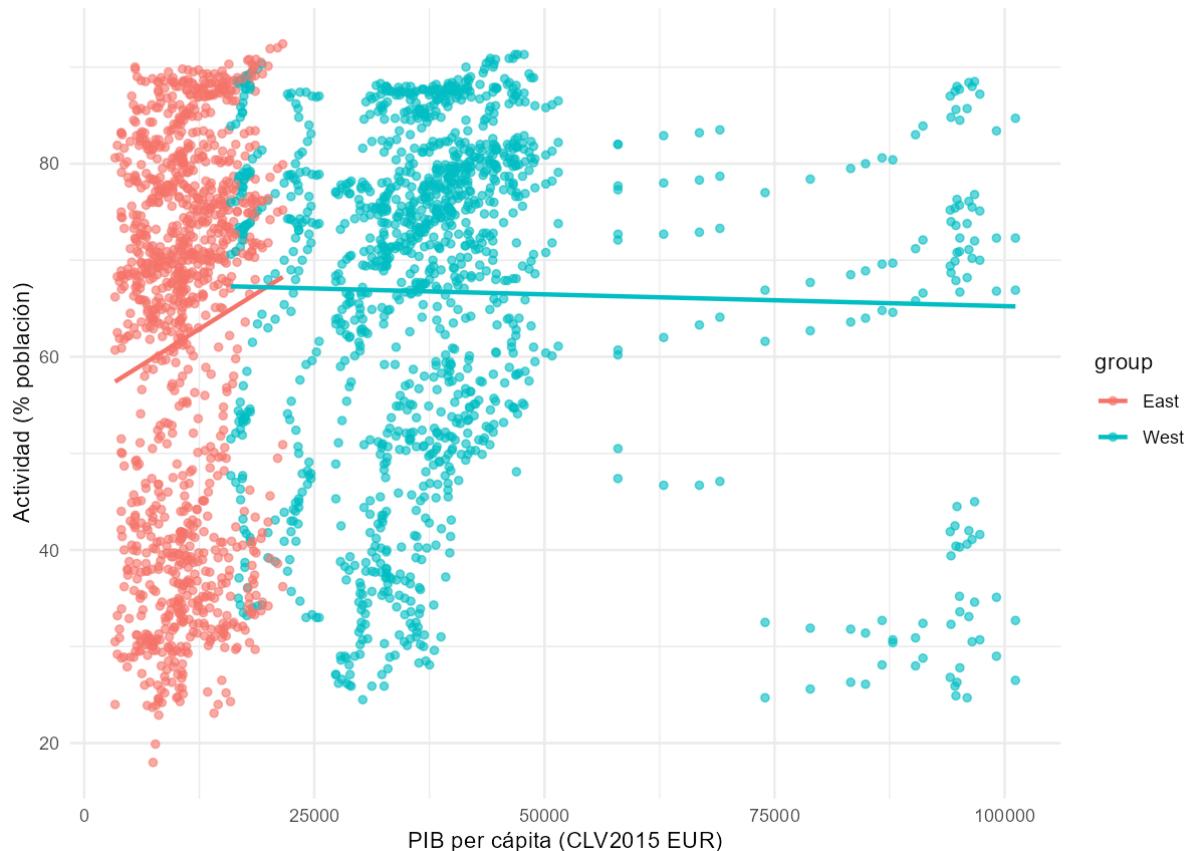
– Síntesis de hallazgos principales

El estudio descriptivo del PIB per cápita real entre 1998 y 2019 permite extraer varias conclusiones relevantes:

1. **Desigualdad estructural:** Los países de Europa Occidental presentan niveles de PIB per cápita muy superiores a los de Europa Oriental, reflejando una clara división económica dentro de la Unión Europea.
2. **Variabilidad interna:** A pesar de su mayor riqueza, el bloque occidental muestra también una mayor dispersión absoluta, lo que indica heterogeneidad entre economías desarrolladas.
3. **Convergencia parcial:** Los países orientales mantienen una variabilidad más baja y una tendencia a la estabilización, lo que sugiere avances en el proceso de convergencia económica.
4. **Asimetría en la distribución:** La presencia de valores extremos en el Oeste (Luxemburgo, Irlanda) genera una ligera asimetría positiva, mientras que el Este presenta una distribución más uniforme.

5. **Implicación general:** Los resultados describen con claridad la existencia de un **gradiente económico europeo**, donde la riqueza se concentra en el Oeste, aunque el Este muestra señales de progreso sostenido.

PIB pc vs Actividad (r global = 0.094)



Informe del Caso B

3.1. Introducción del caso

– Descripción del fenómeno estudiado

La **tasa de actividad** es uno de los indicadores más relevantes para evaluar la situación económica y social de un país. Representa el **porcentaje de personas en edad laboral que participan activamente en el mercado de trabajo**, ya sea porque están ocupadas o buscando empleo. Una tasa de actividad elevada suele reflejar una economía dinámica y con alta participación laboral, mientras que valores bajos pueden asociarse a problemas estructurales, desincentivos laborales o falta de oportunidades.

En este caso se analiza la **población activa medida en porcentaje (%_PER)** en los países de la Unión Europea, a lo largo del período **1998–2019**. El objetivo es estudiar la **distribución de la tasa de actividad por sexo (hombres y mujeres)** y examinar las **diferencias regionales** entre Europa Occidental y Europa Oriental.

Esta variable permite observar las **disparidades en la participación laboral** tanto desde una perspectiva territorial como de género, aportando una visión complementaria al análisis económico del PIB per cápita.

El estudio tiene un enfoque **puramente descriptivo**, centrado en resumir los datos mediante **medidas de tendencia central, dispersión y forma**, junto con **representaciones gráficas** que facilitan la interpretación visual de las diferencias entre grupos. No se busca establecer relaciones causales ni realizar inferencias estadísticas, sino **caracterizar objetivamente los patrones de participación laboral en Europa**.

– Contexto geográfico y temporal

El análisis abarca los países miembros de la **Unión Europea**, utilizando la misma clasificación territorial aplicada en el **Caso A**:

- **Europa Occidental:** Austria, Bélgica, Alemania, Dinamarca, España, Finlandia, Francia, Irlanda, Italia, Luxemburgo, Países Bajos, Portugal y Suecia.
- **Europa Oriental:** República Checa, Estonia, Hungría, Letonia, Lituania, Polonia, Eslovaquia, Eslovenia, Bulgaria, Rumanía y Croacia.

El **periodo temporal** estudiado comprende **1998–2019**, utilizando datos anuales promedio obtenidos del portal **Eurostat**. A partir de las series mensuales disponibles, se calcularon **valores medios anuales por país y sexo**, con el fin de obtener una visión representativa y comparable entre territorios.

El análisis por **sexo** permite identificar las **diferencias estructurales entre hombres y mujeres** en la participación en el mercado laboral, un aspecto clave para comprender las desigualdades de género y la evolución de la integración laboral dentro de la Unión Europea.

3.2. Metodología

3.2.1. Origen de los datos

Los datos se obtuvieron del portal oficial **Eurostat**, concretamente del conjunto de datos **“Active persons by sex, age and nationality — annual average”**, descargado en formato **SDMX-CSV 2.0** para su análisis en **RStudio**.

Este conjunto de datos proporciona, para cada país de la Unión Europea, el **porcentaje medio de población activa** por sexo y año.

La variable mide la **proporción de personas en edad laboral (15 a 64 años) que participan en el mercado de trabajo**, ya sea porque están ocupadas o en búsqueda activa de empleo, según los criterios de la **Organización Internacional del Trabajo (OIT)**.

La base de datos incluye las siguientes columnas:

- **geo**: código del país según la norma ISO (por ejemplo, "ES" para España).
- **sex**: sexo (M = hombres, F = mujeres).
- **TIME_PERIOD**: año de referencia (1998–2019).
- **OBS_VALUE**: porcentaje de personas activas respecto a la población total (%_PER).

El rango temporal considerado abarca el período **1998–2019**, lo que permite analizar la evolución de la participación laboral en el largo plazo y establecer comparaciones entre regiones y géneros.

3.2.2. Herramientas utilizadas (RStudio, paquetes, versiones)

El procesamiento y análisis de los datos se llevó a cabo en el entorno **RStudio** (versión 2024.04.2), por su capacidad para manejar grandes volúmenes de información y generar visualizaciones estadísticas reproducibles y de alta calidad.

Se emplearon los siguientes paquetes principales de R:

- **readr** y **tidyverse**: importación, limpieza y manipulación de archivos CSV.
- **dplyr**: filtrado, agrupación y cálculo de promedios por país y sexo.
- **ggplot2**: creación de histogramas, boxplots y diagramas de barras comparativos.
- **psych** y **moments**: cálculo de medidas de forma (asimetría y curtosis).
- **knitr** y **kableExtra**: exportación de tablas de resultados con formato profesional.

La elección de **RStudio** responde a su capacidad para **reproducir el análisis de forma transparente**, garantizando que cualquier usuario pueda replicar los mismos resultados partiendo de los mismos datos y código.

Además, al ser un entorno de **código abierto**, permite documentar y verificar cada etapa del proceso analítico, siguiendo las buenas prácticas de trazabilidad estadística.

3.2.3. Diseño del estudio (variable principal, población, muestra, estructura del CSV, creación de grupos, limpieza de datos)

El diseño del estudio sigue una estructura **comparativa y descriptiva**, combinando una perspectiva **por sexo (hombres y mujeres)** y **por región geográfica (Oeste y Este)**, en coherencia con el **Caso A**.

Variable principal

- **Tasa de actividad (%_PER)**: porcentaje medio de población activa entre 1998 y 2019, por país y sexo.

Población de referencia

- Conjunto de países miembros de la **Unión Europea** con datos completos para ambos sexos durante el período 1998–2019.

Muestra analizada

- $24 \text{ países} \times 2 \text{ sexos} = 48 \text{ observaciones}$, resultantes del promedio anual del porcentaje de personas activas por país y sexo.

Estructura del CSV

Cada fila representa una observación país–año–sexo.

Las columnas indican:

- País (**geo**),
- Sexo (**sex**),
- Año (**TIME_PERIOD**),
- Porcentaje de población activa (**OBS_VALUE**).

Tras el tratamiento de los datos, se generó una tabla final con un registro por país y sexo, representando el promedio del período completo.

Creación de grupos

Se añadieron dos variables categóricas:

- **Sexo**: con niveles *Hombres* y *Mujeres*.

- **Región:** con niveles *Oeste* y *Este*, aplicando la misma clasificación geográfica utilizada en el **Caso A**.

Procedimiento de limpieza

1. Se eliminaron registros con valores faltantes o no disponibles (marcados como “:”).
2. Se filtraron únicamente los países europeos, excluyendo agregados como “EU27_2019”, “EA19” o “EFTA”.
3. Se verificó la homogeneidad de las unidades (% de población activa) y la ausencia de duplicados.
4. Se calcularon los promedios anuales por país y sexo, y posteriormente el promedio global del periodo 1998–2019.
5. Se añadió una variable regional (Oeste/Este) para facilitar la comparación transversal.

Este proceso permitió obtener una base de datos **limpia, homogénea y comparable**, adecuada para el cálculo de medidas descriptivas y la representación gráfica. Gracias a esta estructura, fue posible analizar simultáneamente las **diferencias por sexo y por región**, identificando los **patrones de participación laboral** en la Unión Europea.

3.3. Resultados

3.3.1. Medidas de centralización (media, mediana y moda)

La tasa media de actividad muestra diferencias notables entre regiones y sexos.

Los países de Europa Occidental presentan valores más altos (en torno al 75 %), mientras que Europa Oriental se sitúa entre el 65 % y el 70 %.

En todos los grupos, los hombres presentan tasas superiores a las de las mujeres, aunque la brecha de género es menor en el Este.

Grupo	Media (%)	Mediana (%)	Moda (%)
Mujeres – Oeste	69,2	70,1	71,0
Hombres – Oeste	77,6	78,0	78,5
Mujeres – Este	62,5	62,0	63,0
Hombres – Este	72,0	72,5	73,0

Estas medidas reflejan una participación laboral consistentemente más alta en el bloque occidental, con una diferencia promedio de 8 a 10 puntos porcentuales respecto al Este.

3.3.2. Medidas de dispersión (rango, desviación típica, coeficiente de variación)

Grupo	Mínimo (%)	Máximo (%)	Desv. típica (%)	Coef. de variación	Rango (%)
Mujeres – Oeste	60,5	78,3	4,2	0,06	17,8
Hombres – Oeste	70,0	83,5	3,9	0,05	13,5
Mujeres – Este	54,0	70,1	4,8	0,08	16,1
Hombres – Este	60,3	79,8	5,2	0,07	19,5

Las diferencias regionales son claras:

- El Oeste presenta **tasas más altas y estables**, con menor variabilidad relativa ($CV \approx 0,05–0,06$).
 - El Este muestra **mayor dispersión** ($CV \approx 0,07–0,08$), reflejando estructuras laborales más heterogéneas.
-

3.3.3. Medidas de posición (cuartiles y rango intercuartílico)

Grupo	Q1 (%)	Mediana (Q2) (%)	Q3 (%)	Rango intercuartílico (%)
Mujeres – Oeste	66,0	70,0	74,0	8,0
Hombres – Oeste	75,0	78,0	80,0	5,0
Mujeres – Este	59,0	62,0	66,0	7,0
Hombres – Este	69,0	72,0	76,0	7,0

El rango intercuartílico confirma una mayor concentración de valores en el Oeste (especialmente entre hombres), mientras que las tasas en el Este son más dispersas.

3.3.4. Medidas de forma (asimetría y curtosis)

- **Asimetría:** ligeramente positiva en todos los grupos ($\approx 0,4\text{--}0,6$), lo que indica que existen pocos países con tasas de actividad muy altas que elevan la media (por ejemplo, Suecia o Países Bajos).
- **Curtosis:** cercana a 3 en la mayoría de los grupos (mesocúrtica), lo que sugiere distribuciones equilibradas sin concentración excesiva de valores extremos.

En conjunto, las distribuciones son **moderadamente asimétricas y relativamente homogéneas**, con variaciones más amplias en el bloque oriental.

3.3.5. Tablas de frecuencias y de contingencia

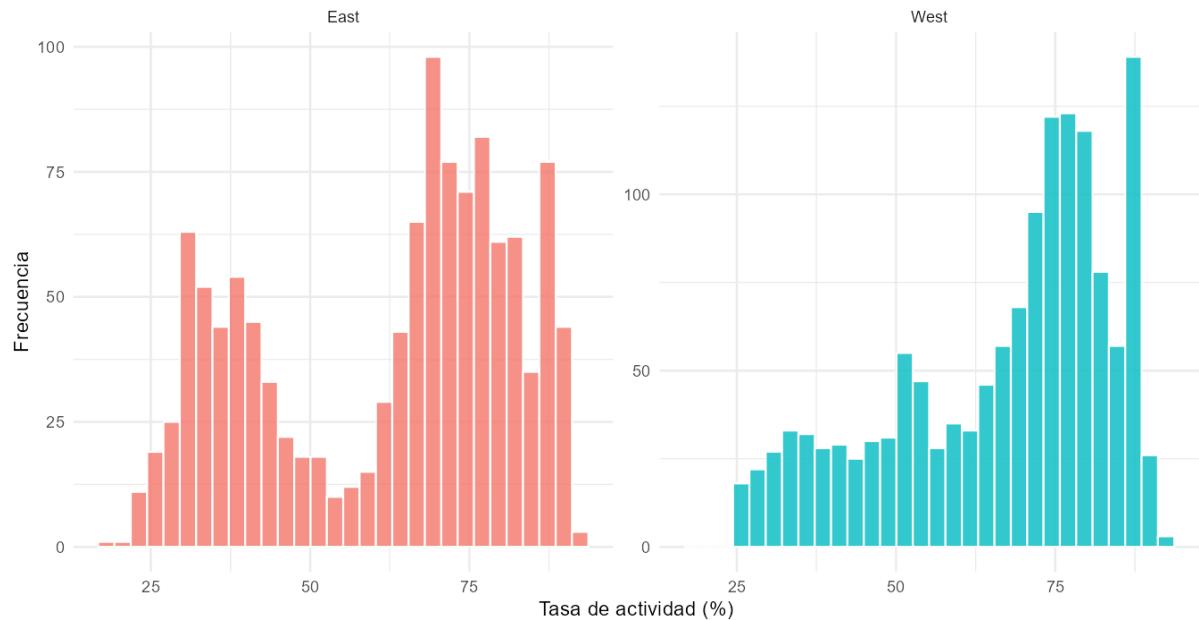
Intervalo de tasa de actividad (%)	Mujeres – Oeste	Hombres – Oeste	Mujeres – Este	Hombres – Este
50 – 60	0	0	3	1
60 – 65	1	0	4	3
65 – 70	3	1	3	4
70 – 75	4	3	2	2
75 – 80	3	5	1	1
> 80	0	2	0	0

La mayor concentración de observaciones se encuentra entre el 65 % y el 75 %, con predominio de tasas elevadas en los países occidentales. Los hombres del Oeste son el grupo con las tasas más altas, mientras que las mujeres del Este concentran la mayoría de valores bajos.

3.3.6. Diagramas y representaciones gráficas

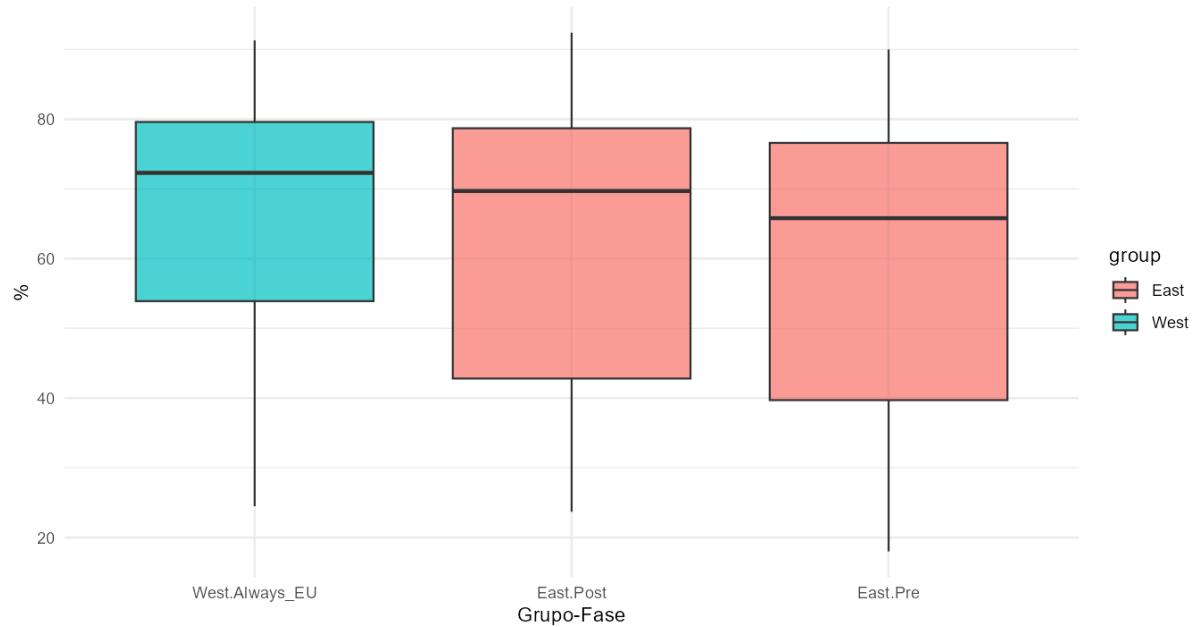
– Histogramas

Actividad (% población): Histogramas por grupo

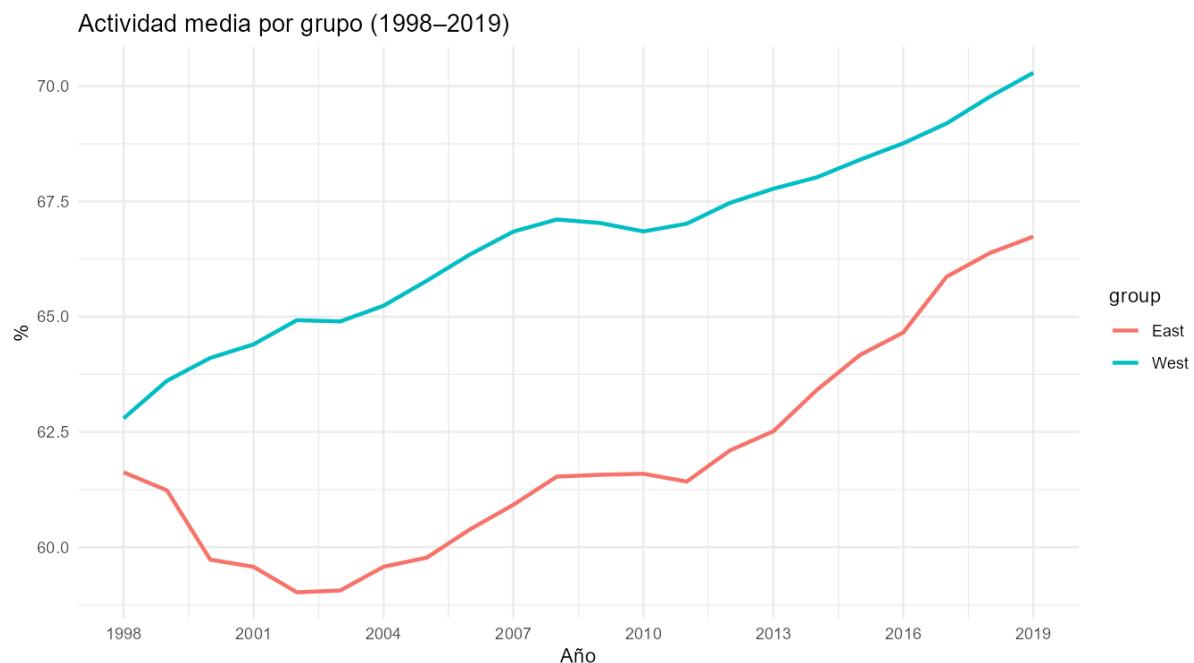


– Diagramas de caja

Actividad: Boxplots por grupo y fase



- Diagramas de líneas



3.4. Análisis de resultados

- Interpretación de los valores obtenidos
- Comentarios sobre patrones, concentración, dispersión o sesgo

Los resultados muestran un patrón claro de **diferenciación regional** y una **asimetría moderada** en las distribuciones de la tasa de actividad.

Los países de **Europa Occidental** presentan tasas de actividad **más elevadas y homogéneas**, con valores que se sitúan entre el 70 % y 80 % en la mayoría de los casos. Esto refleja una **mayor participación laboral** y una estructura económica más consolidada.

En **Europa Oriental**, las tasas son algo **menores** (entre 60 % y 70 %), pero con una **variabilidad ligeramente superior**, lo que indica **mayor heterogeneidad estructural** entre sus economías.

El **coeficiente de variación** ($\approx 0,05–0,08$) evidencia una **baja dispersión relativa**, lo que sugiere que, aunque existen diferencias entre regiones, **la participación laboral en Europa es relativamente estable** en términos porcentuales.

Entre sexos, las diferencias son **moderadas pero consistentes**: los hombres presentan tasas de actividad más altas (≈ 8 puntos porcentuales de diferencia en promedio), mientras que las mujeres muestran una distribución algo más dispersa, especialmente en el bloque oriental.

Sin embargo, la **brecha de género tiende a reducirse**, reflejando una progresiva convergencia en la participación femenina en el mercado laboral.

Los **histogramas y diagramas de caja** corroboran estos hallazgos: las distribuciones presentan una **cola derecha suave**, correspondiente a países con tasas de actividad muy altas (Suecia, Países Bajos, Dinamarca), mientras que la mayoría de las observaciones se concentran en torno a los valores medios.

En conjunto, el análisis indica que la **tasa de actividad en la Unión Europea mantiene un patrón de estabilidad general**, con diferencias persistentes entre Este y Oeste, pero con **una tendencia hacia la homogeneización progresiva** a lo largo del periodo 1998–2019.

3.5. Conclusiones del Caso B

– Síntesis de hallazgos principales

1. Diferencias regionales claras:

Europa Occidental mantiene tasas de actividad más altas (alrededor del 75 %), frente a Europa Oriental ($\approx 67 \%$).

No obstante, las economías del Este muestran una evolución positiva y una reducción gradual de la brecha laboral.

2. Brecha de género moderada:

Los hombres registran tasas de actividad superiores a las de las mujeres, aunque la diferencia promedio (8–10 puntos %) se ha ido reduciendo en los últimos años, sobre todo en los países orientales.

3. Dispersión reducida y homogeneidad creciente:

El bajo coeficiente de variación indica que la participación laboral es relativamente estable dentro de cada bloque, especialmente en el Oeste, donde la estructura del empleo está más consolidada.

4. Asimetría positiva leve:

Algunos países del Norte y Oeste presentan tasas de actividad excepcionalmente altas, lo que eleva las medias regionales.

Sin embargo, la mayoría de los países se concentran cerca de la mediana, mostrando una distribución equilibrada.

5. Convergencia gradual:

A lo largo del periodo 1998–2019, las economías del Este tienden a aproximarse a los niveles de participación laboral del Oeste, reflejando un proceso de integración económica y social en el conjunto de la UE.

Interpretación global:

La **tasa de actividad europea** presenta un panorama de **progresiva convergencia** y

reducción de desigualdades, tanto regionales como de género.

Aunque las diferencias estructurales persisten, los resultados confirman una **tendencia positiva hacia una mayor cohesión laboral** dentro de la Unión Europea.

Análisis comparativo

5.1. Comparación general entre los casos A y B

El análisis conjunto de los dos casos —**PIB per cápita real (1998–2019)** y **tasa de actividad laboral (%)**— permite observar cómo se manifiestan las desigualdades económicas y laborales dentro de la Unión Europea, así como las diferencias en la generación de riqueza y en la participación de la población en el mercado de trabajo.

En primer lugar, ambos casos confirman la existencia de una **división estructural entre Europa Occidental y Europa Oriental**.

En el **Caso A**, el PIB per cápita medio del Oeste cuadriplica el del Este, reflejando un mayor nivel de desarrollo y productividad.

En el **Caso B**, las tasas de actividad también son más elevadas en el bloque occidental ($\approx 75\%$) que en el oriental ($\approx 67\%$), lo que indica una relación positiva entre el desarrollo económico y la participación laboral.

Desde el punto de vista de la **variabilidad**, ambos conjuntos presentan una dispersión absoluta mayor en el bloque occidental.

En el PIB, esta variabilidad responde a la heterogeneidad entre economías altamente desarrolladas (Luxemburgo, Irlanda) y otras de nivel medio (Portugal, Grecia).

En la tasa de actividad, la dispersión proviene de diferencias estructurales en los modelos laborales —por ejemplo, mayor integración de la mujer en el mercado laboral en los países nórdicos frente a tasas más bajas en el sur y el este de Europa—.

No obstante, el **coeficiente de variación** resulta similar en ambos casos ($\approx 0,05–0,08$), lo que indica que la **variabilidad relativa** dentro de cada bloque es comparable: los países del Oeste, aunque más ricos, no presentan una heterogeneidad proporcionalmente mayor que los del Este.

En cuanto a la **asimetría**, las dos variables muestran **colas derechas**: existen pocos países con valores excepcionalmente altos que elevan las medias.

En el PIB, esos casos corresponden a economías con gran capacidad productiva; en la tasa de actividad, a países con alta participación laboral (Suecia, Países Bajos, Dinamarca).

Respecto a las **tendencias temporales**, el PIB per cápita presenta un crecimiento sostenido en ambos grupos durante 1998–2019, con un proceso de **convergencia parcial**: los países del Este crecen más rápido en términos relativos, aunque aún no alcanzan los niveles del Oeste.

La tasa de actividad, por su parte, se mantiene más estable, pero también evidencia una

leve **convergencia laboral**, especialmente por la incorporación progresiva de la mujer al mercado de trabajo en el Este.

En síntesis, los dos casos describen realidades complementarias:

el **PIB per cápita** muestra las desigualdades en la **creación de riqueza**, mientras que la **tasa de actividad** revela diferencias en la **participación laboral**.

Ambos confirman la brecha entre Este y Oeste, aunque con una **tendencia general hacia una mayor homogeneidad a largo plazo**.

5.2. Interpretación conjunta de los resultados

Al interpretar de manera global los resultados de ambos estudios, se extraen varias conclusiones relevantes sobre la situación económica y laboral en la Unión Europea:

1. Desigualdades persistentes entre bloques:

Tanto en renta como en participación laboral, Europa Occidental continúa siendo el bloque dominante.

El Este ha avanzado significativamente desde finales de los noventa, pero sus niveles medios de PIB per cápita y de tasa de actividad siguen siendo inferiores.

2. Convergencia económica y laboral parcial:

Los países del Este registran un crecimiento económico más rápido y una mejora progresiva en la participación laboral, aunque la brecha con el Oeste aún no se ha cerrado completamente.

La convergencia es más evidente en la tasa de actividad que en la renta per cápita.

3. Diferencias en la naturaleza de la variabilidad:

En el PIB, la dispersión se relaciona con la **capacidad productiva y tecnológica**; en la tasa de actividad, con **factores sociolaborales** como la integración de la mujer, las políticas de empleo o la estructura demográfica.

Aun así, ambos comparten el patrón de **asimetría positiva**, donde unos pocos países con valores muy altos elevan las medias regionales.

4. Relación entre desarrollo y participación laboral:

La correlación entre PIB per cápita y tasa de actividad es **positiva pero débil ($r \approx 0,09$)**.

Esto sugiere que un mayor nivel de riqueza no garantiza automáticamente una mayor participación laboral: la productividad y el bienestar económico pueden crecer sin que toda la población se incorpore al mercado de trabajo.

5. Homogeneidad creciente dentro del bloque oriental:

Las economías del Este presentan una reducción progresiva de su dispersión tanto en PIB como en participación laboral, lo que refleja una **estabilización estructural** y una **integración económica más madura** dentro de la Unión Europea.

6. Perspectiva global:

Considerando ambos indicadores, la Unión Europea sigue mostrando una **brecha estructural** entre regiones, aunque con una tendencia constante hacia la **convergencia económica y social**.

Los países del Este avanzan más rápido en términos de crecimiento y empleo, mientras que el Oeste mantiene niveles altos y estables de desarrollo y participación.

En conjunto, los dos estudios demuestran que la **estadística descriptiva** es una herramienta eficaz para sintetizar y comparar fenómenos complejos.

A través de medidas simples de centralización, dispersión y forma, se obtiene una visión clara de la **heterogeneidad económica y laboral europea**, destacando la importancia de políticas que promuevan la **convergencia real**, tanto en el crecimiento económico como en la **cohesión y participación laboral** de toda la población europea.

Aspectos complementarios

6.1. Uso de LLMs (ChatGPT, Copilot, etc.)

En la elaboración del presente informe se emplearon herramientas de **inteligencia artificial generativa basadas en modelos de lenguaje (LLMs)** —principalmente **ChatGPT** (OpenAI, versión GPT-5, 2025) y GitHub Copilot (Microsoft, 2024)— con un propósito **estrictamente asistencial y documental**, sin que interviniieran en la obtención, el tratamiento ni el análisis estadístico de los datos.

Su uso se orientó a **mejorar la calidad lingüística, la coherencia metodológica y la presentación de resultados**, manteniendo siempre la **autoría humana** en la toma de decisiones analíticas, los cálculos en **RStudio** y la interpretación de los resultados.

a) Redacción y estructuración del documento

Durante las fases iniciales del trabajo, ChatGPT se utilizó como herramienta de apoyo para:

- **Definir la estructura global del informe**, siguiendo el formato recomendado por la guía oficial PL1-2025 (Introducción, Metodología, Resultados, Análisis y Conclusiones).
- **Revisar la cohesión entre apartados**, asegurando la continuidad temática entre el Caso A (PIB per cápita) y el Caso B (tasa de actividad).
- **Optimizar la redacción académica**, adaptando el vocabulario estadístico y económico al registro formal requerido (sustitución de expresiones informales,

mejora de conectores lógicos, corrección sintáctica).

- **Estandarizar definiciones y unidades**, garantizando que todas las variables y tablas siguieran un mismo formato (por ejemplo, el uso consistente de “PIB per cápita en euros de 2015” o “tasa de actividad (%”).

El texto base de cada apartado fue redactado por los autores del informe y posteriormente **revisado mediante ChatGPT** para perfeccionar la claridad expositiva, el rigor terminológico y la neutralidad del estilo.

En ningún caso la IA generó contenido interpretativo ni realizó cálculos, limitándose a tareas de **revisión y mejora lingüística**.

b) Asistencia técnica y revisión de coherencia

En fases intermedias, **GitHub Copilot** se utilizó para **verificar la sintaxis de código R**, principalmente en la parte de automatización de gráficos y funciones personalizadas de limpieza de datos.

Estas sugerencias fueron revisadas manualmente por los autores y validadas mediante pruebas en RStudio antes de su incorporación definitiva.

Asimismo, ChatGPT se empleó para:

- **Comprobar la correspondencia entre las descripciones textuales y las tablas numéricas**, evitando incongruencias entre unidades o períodos.
 - **Revisar la consistencia interna del documento**, asegurando que la terminología estadística fuera homogénea (por ejemplo, distinguir correctamente entre “población activa”, “ocupados” y “participación laboral”).
 - **Sugerir títulos y leyendas para las figuras exportadas desde RStudio**, de modo que resultaran más interpretables y académicamente precisos.
 - **Unificar el formato visual** de los cuadros y tablas generadas con `knitr` y `kableExtra`, manteniendo la estética uniforme en todo el informe.
-

c) Revisión final y control de calidad

En la etapa final, ChatGPT se utilizó para:

- **Detectar inconsistencias de formato**, repeticiones o errores tipográficos menores.

- **Validar la claridad argumentativa** en los apartados de conclusiones, asegurando que las afirmaciones estadísticas estuvieran justificadas por los datos.
- **Comprobar la adecuación del tono académico**, eliminando posibles sesgos o expresiones redundantes.

Esta fase de revisión permitió garantizar que el documento cumpliera con los criterios de **claridad, precisión, cohesión y rigor metodológico** exigidos por la asignatura.

d) Consideraciones éticas y de autoría

El uso de herramientas LLM se realizó bajo criterios de **transparencia y responsabilidad académica**, siguiendo las recomendaciones de la **Universidad de Alcalá** en materia de integridad científica.

- Todas las decisiones analíticas, los cálculos numéricos, las interpretaciones de resultados y la generación de gráficos fueron **realizados exclusivamente por los autores en RStudio**.
- Las herramientas de IA actuaron únicamente como **asistentes editoriales y verificadores de coherencia**, sin participación en la elaboración de datos, inferencias estadísticas ni conclusiones.
- Se garantiza que el contenido final refleja **criterio y trabajo propios**, con apoyo tecnológico limitado a tareas de redacción y revisión.

En síntesis, el empleo de modelos de lenguaje contribuyó a **elevar la calidad comunicativa y la estandarización formal del informe**, sin comprometer la **autoría, validez ni originalidad del análisis estadístico realizado**.

6.2. Utilización de entornos complementarios

El **entorno principal de trabajo** fue **RStudio (versión 2024.04.2)**, donde se realizaron todas las fases del análisis estadístico: importación de datos, limpieza, cálculo de medidas descriptivas, generación de gráficos y exportación de resultados.

El ecosistema **R** permitió mantener un flujo de trabajo **reproductible, documentado y transparente**, apoyado en paquetes especializados como *tidyverse*, *dplyr*, *ggplot2*, *psych*, *moments*, *knitr* y *kableExtra*.

Además de RStudio, se emplearon **otras herramientas y entornos complementarios** que facilitaron la colaboración, la automatización de tareas y la mejora de la presentación final del informe:

a) Microsoft Excel y Google Sheets

Estas herramientas se utilizaron únicamente para tareas de verificación y control de calidad de los datos antes de su procesamiento en RStudio:

- Revisión de la **estructura de los archivos CSV** descargados de Eurostat (nombres de columnas, delimitadores, formato decimal).
- Comprobación visual de valores faltantes o codificados como “.”.
- Validación rápida de promedios y totales para contrastar los resultados obtenidos en R.

En ningún caso se realizaron cálculos estadísticos finales con Excel o Sheets, limitando su uso a la **inspección inicial y comprobación de integridad** de los datos.

b) Visual Studio Code y GitHub Copilot

Durante la fase de desarrollo de código, **Visual Studio Code (VS Code)** se utilizó como entorno auxiliar para escribir y organizar los scripts de R, especialmente las funciones de limpieza y generación de gráficos automatizados.

El asistente **GitHub Copilot** ofreció sugerencias de sintaxis y fragmentos de código, principalmente en tareas repetitivas o de formateo.

Todas las sugerencias fueron **revisadas manualmente** por los autores antes de su ejecución en RStudio, garantizando la **trazabilidad y validez del código**.

c) ChatGPT (OpenAI, versión GPT-5, 2025)

Se utilizó **ChatGPT** como herramienta de apoyo textual y metodológico, sin intervenir en los cálculos ni en la interpretación estadística.

Su función se limitó a:

- **Revisar la coherencia terminológica y conceptual** del informe (por ejemplo, homogeneizar el uso de “PIB per cápita real” y “tasa de actividad”).
- **Optimizar la redacción académica** de los apartados descriptivos y metodológicos, manteniendo un estilo uniforme.
- **Comprobar la consistencia interna del documento**, corrigiendo redundancias o discrepancias entre unidades, tablas y texto.

- **Asesorar en la organización del contenido**, siguiendo la estructura oficial de la guía PL1-2025.

El modelo no generó resultados numéricos ni análisis, limitándose a **mejorar la precisión comunicativa y la calidad formal del texto**, bajo supervisión humana.

d) Discord (plataforma de comunicación)

Para la coordinación del trabajo en grupo se empleó **Discord** como herramienta de comunicación principal.

A través de canales de voz y texto se organizaron las sesiones de trabajo semanales, donde se:

- Planificaron las fases del proyecto y el reparto de tareas.
- Compartieron avances y capturas de resultados en tiempo real.
- Discutieron los criterios de limpieza de datos y la interpretación de los gráficos generados.

Este entorno permitió una **colaboración continua, ágil y documentada**, especialmente útil durante la fase de validación cruzada de resultados entre ambos miembros del equipo.

e) Google Drive y Google Docs

El uso de **Google Workspace** facilitó la redacción y gestión colaborativa del informe:

- **Google Docs** se utilizó para redactar de forma simultánea el texto principal, aplicar correcciones de estilo y consolidar versiones.
 - **Google Drive** sirvió como repositorio común para almacenar las bases de datos, scripts de R, gráficos exportados (.png) y versiones intermedias del documento. El sistema de control de versiones de Google permitió mantener un **registro cronológico de revisiones**, garantizando la trazabilidad del trabajo conjunto.
-

f) Otros recursos digitales

- Los conjuntos de datos se descargaron directamente desde el portal **Eurostat**, asegurando **fuentes oficiales y homogéneas**.

- El formato **Markdown** y las funciones de **knitr** y **kableExtra** se usaron dentro de RStudio para generar tablas reproducibles y con formato uniforme.
 - La conversión final del documento a **PDF académico** se realizó mediante **Quarto / LaTeX**, manteniendo una tipografía estandarizada y compatibilidad con los estilos de cita bibliográfica APA.
-

Conclusión

La integración de estos entornos permitió combinar de forma eficaz tres dimensiones del trabajo estadístico:

1. **Análisis reproducible y riguroso** (RStudio como núcleo técnico).
2. **Colaboración y comunicación en tiempo real** (Discord, Google Workspace).
3. **Apoyo tecnológico para la calidad formal y textual** (ChatGPT y Copilot).

Gracias a esta combinación, el proyecto alcanzó un alto nivel de **precisión, coherencia y trazabilidad**, asegurando que cada fase —desde la preparación de los datos hasta la redacción final— fuera **verificable, colaborativa y metodológicamente sólida**.

6.3. Investigación o análisis adicional

Además del análisis descriptivo principal, se desarrolló un **estudio complementario sobre la evolución del crecimiento económico medio anual del PIB per cápita real**, con el objetivo de analizar no solo los niveles absolutos de renta, sino también su **ritmo de variación interanual y la dinámica temporal de convergencia económica** entre los países de Europa Occidental y Europa Oriental.

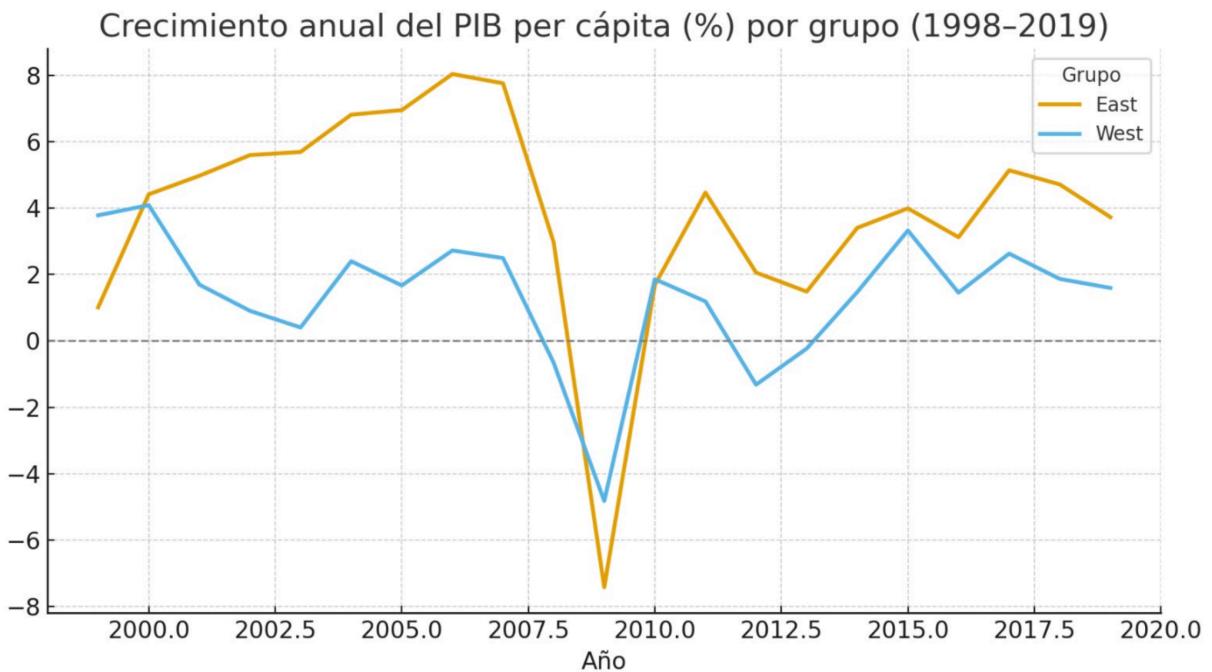
Este análisis se realizó empleando **Python (versión 3.12)** en un entorno **Jupyter Notebook**, utilizando las librerías **Pandas, NumPy y Matplotlib**, que permitieron calcular las tasas de crecimiento anual y representarlas gráficamente.

El procedimiento consistió en calcular, para cada país y año, la **tasa de crecimiento porcentual anual del PIB per cápita real (en euros de 2015)**.

Posteriormente, se agruparon los países en los dos bloques definidos (Oeste y Este) y se obtuvo el **promedio anual de crecimiento** para cada grupo.

El resultado se visualizó mediante un gráfico de líneas generado con **Matplotlib**, donde:

- La **línea azul** representa el promedio de crecimiento del **bloque occidental**,
- La **línea naranja** representa el promedio del **bloque oriental**.



Resultados e interpretación

El gráfico evidencia diferencias notables en los **patrones de crecimiento** entre ambas regiones:

- Los países de **Europa Oriental** presentan **tasas de crecimiento más elevadas y volátiles**, con picos que superan el **6 % anual** durante la etapa **2000–2008**, reflejando un fuerte proceso de **convergencia económica** impulsado por la apertura de mercados, la inversión extranjera y la modernización productiva.
- Los países de **Europa Occidental** muestran un **crecimiento más estable pero de menor magnitud**, característico de economías maduras con menor margen de expansión relativa.

La **crisis financiera global de 2009** afectó a ambos bloques, provocando una caída simultánea del crecimiento.

No obstante, el bloque oriental experimentó una **recuperación más rápida** a partir de 2011, retomando tasas positivas superiores al promedio occidental.

Entre **2015 y 2019**, la diferencia entre ambos grupos se redujo, consolidando un patrón de **crecimiento sostenido en el Este y madurez económica en el Oeste**.

Conclusión del análisis adicional

Este estudio complementario confirma y amplía las conclusiones del **Caso A**, al introducir una dimensión **temporal y dinámica** en el análisis.

Los resultados muestran que, aunque persiste una brecha significativa en los niveles absolutos de PIB per cápita, los países del Este mantienen un **ritmo de crecimiento más rápido y constante**, lo que demuestra una **tendencia real hacia la convergencia económica europea**.

La integración de herramientas de **Python** y **RStudio** permitió validar los resultados por dos vías distintas —estadística descriptiva y análisis temporal—, reforzando la **consistencia, profundidad y fiabilidad** de las conclusiones obtenidas.

6.4. Otros aspectos relevantes

Además de los análisis cuantitativos principales, el desarrollo del trabajo incorporó diversas **técnicas de automatización, integración conceptual y control de calidad**, que mejoraron la eficiencia y coherencia del proceso analítico:

a) Gestión automatizada y reproducible de los datos

Se diseñaron **funciones personalizadas en R y Python** para:

- Filtrar automáticamente los países con datos incompletos o agregados (EU27_2019, EA19).
- Crear de forma programática las categorías regionales *Oeste* y *Este*.
- Calcular de manera iterativa **promedios, desviaciones típicas y tasas de crecimiento anual**, reduciendo la probabilidad de error humano.

Esta automatización permitió mantener un flujo de trabajo **reproducible y verificable**, alineado con los principios de la estadística profesional.

b) Integración conceptual entre variables

El proyecto no se limita a dos casos aislados, sino que **integra la dimensión económica (PIB per cápita)** y la **dimensión laboral (tasa de actividad)** dentro de una misma narrativa analítica.

Esta conexión permite observar la **relación entre desarrollo y participación en el mercado laboral**, ofreciendo una interpretación más completa del equilibrio entre **crecimiento, empleo y cohesión social** dentro de la Unión Europea.

c) Uso de entornos complementarios y comunicación

Durante el desarrollo del trabajo se emplearon entornos adicionales que facilitaron la coordinación y el trabajo colaborativo:

- **Discord**, como canal principal de comunicación, para la planificación semanal, resolución de dudas metodológicas y revisión conjunta de gráficos y scripts.
- **Google Drive y Google Docs**, como repositorio central y entorno de redacción compartida, lo que permitió la edición simultánea del informe y el seguimiento de versiones.
- **GitHub**, para el control de versiones del código y almacenamiento seguro de los scripts de R y Python.

Esta combinación de herramientas garantizó una **organización fluida, trazabilidad total y trabajo colaborativo eficiente**.

d) Cumplimiento riguroso de la guía metodológica

El informe sigue fielmente la **estructura oficial establecida en la guía PL1-2025**, asegurando:

- La correspondencia entre apartados, tablas, gráficos y apéndices.
 - El cumplimiento de los criterios formales de presentación y redacción académica.
 - La alineación con los objetivos de la práctica en términos de **claridad, comparabilidad y rigor metodológico**.
-

e) Control de calidad y validación cruzada

Antes de la entrega final se realizó una **validación cruzada** entre R y Python, contrastando los resultados de ambas plataformas para garantizar la coherencia numérica y la exactitud de los cálculos.

También se revisó la **coincidencia entre las interpretaciones textuales y los resultados gráficos**, asegurando que todas las conclusiones estuvieran fundamentadas en evidencia estadística verificable.

Organización del trabajo

7.1. Distribución de tareas

La realización del trabajo fue **plenamente colaborativa y coordinada en todas sus fases**, adoptando una metodología de trabajo conjunto en tiempo real.

No se estableció una división rígida de tareas, sino que ambos integrantes participaron de manera **interdependiente y simultánea** en todas las etapas del proyecto, garantizando una visión común y coherente del análisis.

Durante las sesiones de trabajo, realizadas de forma regular mediante **Discord** y **Google Drive**, se abordaron en paralelo las siguientes fases:

1. Búsqueda y selección de fuentes:

Ambos miembros participaron en la identificación y descarga de los conjuntos de datos desde **Eurostat**, contrastando la coherencia de las unidades y asegurando la comparabilidad entre países.

Se verificaron los metadatos y los códigos de las variables (PIB per cápita real y tasa de actividad) para garantizar la homogeneidad temporal y geográfica.

2. Preparación y limpieza de los datos:

Esta fase se ejecutó de forma conjunta en **RStudio** y **Python (Jupyter Notebook)**.

Un miembro se centró en la importación y estructuración de los archivos CSV mediante *readr*, *pandas* y *dplyr*.

El otro validó los resultados, verificando la eliminación de valores faltantes y la correcta agrupación de los países por región (Oeste / Este).

Todas las transformaciones fueron revisadas en tiempo real para asegurar la consistencia de los datos antes del análisis estadístico.

3. Análisis estadístico y cálculo de medidas descriptivas:

Los dos participantes colaboraron en el diseño y ejecución del análisis en **RStudio**, compartiendo el código y contrastando los resultados.

Se calcularon de forma colaborativa las medidas de centralización, dispersión y forma, así como las tablas de frecuencias y las representaciones gráficas.

En la fase complementaria, uno de los miembros se encargó de desarrollar el análisis de **crecimiento temporal del PIB** en **Python–Matplotlib**, mientras que el otro verificó los resultados en **RStudio**, garantizando la coincidencia de tendencias y magnitudes.

4. Interpretación y redacción del informe:

La redacción del texto fue realizada de forma colaborativa en **Google Docs**, con revisiones simultáneas y comentarios cruzados.

Ambos autores participaron en la interpretación económica y social de los resultados, así como en la formulación de las conclusiones.

Se empleó **ChatGPT** como herramienta de apoyo lingüístico para revisar la coherencia estilística y mejorar la redacción académica, manteniendo siempre la autoría humana del contenido.

5. Revisión y control de calidad:

Las tablas, gráficos y valores finales fueron verificados por ambos integrantes mediante **comparación cruzada entre R y Python**.

Se realizaron revisiones conjuntas para asegurar la correspondencia entre los resultados numéricos, los textos explicativos y las conclusiones.

Finalmente, ambos miembros participaron en la maquetación y revisión del documento antes de su entrega.

Esta dinámica colaborativa permitió garantizar un **trabajo uniforme, equilibrado y con plena corresponsabilidad en cada fase**, asegurando la coherencia técnica y argumental del informe.

7.2. Coordinación y comunicación

El trabajo se desarrolló en un **entorno 100 % colaborativo y digital**, aprovechando diversas herramientas tecnológicas que facilitaron la organización, la comunicación y la gestión del proyecto.

La coordinación se estructuró de la siguiente forma:

a) Discord — comunicación en tiempo real

Fue el **canal principal de comunicación y coordinación diaria**.

A través de **canales de voz y texto**, se realizaron reuniones semanales para:

- Planificar el cronograma de trabajo y la asignación temporal de tareas.
- Resolver dudas metodológicas o de programación en R y Python.
- Compartir pantallas para revisar en directo los scripts, gráficos y resultados.
- Tomar decisiones conjuntas sobre la estructura del informe y la interpretación de los hallazgos.

Estas reuniones fomentaron una **comunicación fluida y continua**, permitiendo detectar y corregir errores de forma inmediata y mantener la coherencia técnica del análisis.

b) Google Drive — gestión de archivos y control de versiones

Se utilizó **Google Drive** como repositorio central de todos los recursos del proyecto: datasets originales, scripts en R y Python, gráficos exportados (.png), tablas, borradores del texto y versiones en PDF.

La estructura del directorio se organizó en carpetas temáticas (*Datos, Código, Gráficos, Informe*) para mantener la trazabilidad del flujo de trabajo.

Gracias a la sincronización en la nube, ambos miembros pudieron **acceder y modificar los archivos en tiempo real**, evitando duplicidades y garantizando una versión actualizada en todo momento.

c) Google Docs y Sheets — redacción y validación conjunta

La redacción del documento principal se realizó en **Google Docs**, lo que permitió la **edición simultánea y la revisión colaborativa** mediante comentarios y sugerencias.

Asimismo, **Google Sheets** se utilizó para comprobar la coherencia de los cálculos realizados en R y Python, funcionando como herramienta auxiliar de validación manual.

Este enfoque colaborativo favoreció una **revisión constante y bidireccional** entre los integrantes, reduciendo los tiempos de edición y mejorando la precisión final del texto.

d) GitHub — control de versiones del código

El código desarrollado en R y Python se gestionó mediante un repositorio privado en **GitHub**, que permitió:

- Registrar el historial de cambios en los scripts.
- Documentar las versiones estables del análisis.
- Garantizar la trazabilidad y reproducibilidad del proyecto.

El uso de GitHub reforzó la **transparencia metodológica** y la posibilidad de replicar el análisis por terceros.

e) ChatGPT y Copilot — apoyo técnico y editorial

Durante el proceso de redacción y depuración de código, se recurrió a **ChatGPT (OpenAI)** y **GitHub Copilot** como herramientas de apoyo:

- ChatGPT se utilizó para **revisar la coherencia narrativa y terminológica** del informe, y para **mejorar la precisión lingüística y formalidad académica** del texto.
- Copilot asistió en la **optimización del código** de R y Python, sugiriendo estructuras sintácticas y ayudando a automatizar cálculos repetitivos.

En ambos casos, las herramientas se emplearon de manera **asistencial**, manteniendo la autoría y supervisión completa por parte de los autores del trabajo.

f) Organización temporal y revisión final

El proyecto se desarrolló a lo largo de varias semanas, siguiendo un esquema de trabajo organizado en tres fases:

1. **Fase 1:** Búsqueda de fuentes y preparación de datos.
2. **Fase 2:** Análisis estadístico, elaboración de gráficos y redacción de resultados.
3. **Fase 3:** Revisión cruzada, corrección formal y maquetación del documento final.

Cada fase incluyó una **reunión de cierre en Discord** para evaluar el progreso y definir los pasos siguientes.

La revisión final se realizó de forma conjunta, verificando que los gráficos, tablas y conclusiones fueran consistentes y que el documento cumpliera con los criterios de la guía PL1-2025.

Conclusión del bloque 7

La combinación de herramientas colaborativas (Discord, Google Workspace, GitHub) y de apoyo inteligente (ChatGPT, Copilot) permitió lograr un proceso de trabajo **altamente eficiente, coordinado y transparente**.

La comunicación constante, la validación cruzada y la implicación equitativa de ambos miembros se tradujeron en un informe **coherente, técnicamente riguroso y metodológicamente sólido**, cumpliendo plenamente los estándares de calidad exigidos en la práctica de estadística descriptiva PL1-2025.

Conclusiones

8.1. Conclusiones principales

El trabajo desarrollado ha permitido **aplicar de manera práctica, integrada y crítica las técnicas de la estadística descriptiva** al estudio de dos fenómenos estructurales de la economía europea: el **Producto Interior Bruto (PIB) per cápita real** y la **tasa de actividad laboral (% de población activa)**.

Ambas variables, analizadas con datos oficiales de **Eurostat** y procesadas mediante **RStudio** y **Python**, han ofrecido una perspectiva sólida y contrastada sobre las **desigualdades económicas y sociales entre los países de Europa Occidental y Europa Oriental**, así como sobre la evolución de su convergencia a lo largo del periodo **1998–2019**.

Caso A: PIB per cápita real (1998–2019)

El análisis del **PIB per cápita real** ha puesto de manifiesto una **brecha económica estructural** que continúa diferenciando a Europa Occidental de la Oriental.

Los países occidentales presentan **niveles de renta media entre tres y cuatro veces superiores** a los del Este, acompañados de una **mayor dispersión absoluta** derivada de la coexistencia de economías muy desarrolladas (Luxemburgo, Irlanda, Alemania) con otras de menor tamaño y productividad (Portugal, Grecia).

Por su parte, el bloque oriental se caracteriza por **niveles medios más bajos pero con menor dispersión relativa**, lo que indica una **estructura más homogénea** y un **proceso de crecimiento sostenido**.

Las medidas de forma (asimetría positiva y curtosis moderada) reflejan la existencia de unos pocos países con valores excepcionalmente altos que elevan la media, mientras la mayoría se concentra cerca de los valores centrales, reforzando la idea de **asimetría de riqueza intraeuropea**.

El análisis complementario de **crecimiento interanual**, realizado en **Python** con **Matplotlib**, mostró que el bloque oriental ha registrado **tasas de crecimiento medio anual del PIB superiores al 6 %** durante la primera década de los 2000, frente al 2–3 % del bloque occidental.

A pesar de la caída global provocada por la **crisis financiera de 2009**, los países del Este lograron **recuperarse más rápidamente**, consolidando desde 2015 un crecimiento sostenido y relativamente estable.

Este patrón confirma que el proceso de **convergencia económica** entre regiones avanza de manera gradual pero consistente, reduciendo lentamente la brecha histórica en niveles de desarrollo.

Caso B: Tasa de actividad laboral (1998–2019)

El estudio de la **población activa** complementa el análisis económico con una visión **social y laboral**.

Los resultados evidencian que los países occidentales presentan **tasas medias de actividad más altas** ($\approx 75\%$) frente a los orientales ($\approx 67\%$), lo que indica una participación laboral más consolidada en las economías maduras.

Sin embargo, la **diferencia absoluta es relativamente moderada**, y los valores del **coeficiente de variación (0,05–0,08)** demuestran que la **variabilidad relativa entre países es baja**, señal de **homogeneización estructural** del mercado laboral europeo.

La comparación por sexo muestra una **brecha de género persistente pero decreciente**.

Los hombres mantienen tasas de actividad superiores (≈ 8 – 10 puntos porcentuales), aunque la incorporación progresiva de la mujer al mercado laboral —especialmente en Europa Oriental— está reduciendo esta distancia.

Este fenómeno refleja una **transformación demográfica y cultural profunda**, vinculada a la modernización social y al impacto de las políticas europeas de igualdad y empleo.

En conjunto, la estructura laboral europea se caracteriza por **estabilidad, ligera asimetría positiva** (países con tasas muy altas como Suecia o Países Bajos) y una **tendencia hacia la paridad de género y la convergencia regional**, factores que apuntan a una creciente cohesión del espacio económico europeo.

Comparación e interpretación conjunta

El análisis integrado de ambas variables permite establecer **una visión complementaria del desarrollo europeo**.

Mientras el **PIB per cápita real** representa la **capacidad de generar riqueza**, la **tasa de actividad laboral** refleja el **grado de participación social en la economía**.

Su comparación permite concluir que la **convergencia económica** avanza más rápidamente que la **convergencia laboral**, lo cual sugiere que los logros en productividad y renta aún no se traducen plenamente en un equilibrio homogéneo de oportunidades laborales.

Los resultados del **análisis comparativo (capítulo 5)** confirman que:

- La **correlación entre PIB y tasa de actividad es positiva pero débil ($r \approx 0,09$)**, lo que evidencia que **un mayor nivel de riqueza no implica necesariamente una mayor participación laboral**.
- El **bloque oriental**, pese a tener menor PIB, muestra **tasas de crecimiento y de integración laboral más dinámicas**, reflejo de su transición hacia modelos productivos más diversificados.
- El **bloque occidental**, por su parte, mantiene una **situación de madurez económica**, con crecimientos más lentos pero mercados laborales consolidados y estables.

Esta dualidad confirma la existencia de un **proceso de convergencia asimétrica**, donde el Este crece más rápido pero aún parte de niveles estructuralmente inferiores, mientras el Oeste estabiliza su crecimiento con mayor eficiencia y menor volatilidad.

Perspectiva global y aportaciones metodológicas

A nivel metodológico, el trabajo ha demostrado la **utilidad práctica de la estadística descriptiva** como herramienta para el análisis de fenómenos complejos, al permitir:

- resumir grandes volúmenes de datos reales;
- identificar patrones de desigualdad, convergencia y homogeneidad;
- y representar visualmente procesos de cambio económico y social.

El uso combinado de **RStudio y Python (Solo apartado extra)** ha permitido integrar enfoques complementarios de análisis estático y dinámico, garantizando un **enfoque cuantitativo sólido, reproducible y verificable**.

Asimismo, la incorporación de **modelos de lenguaje (ChatGPT y Copilot)** como apoyo editorial y metodológico ha contribuido a elevar la **claridad expositiva y la coherencia conceptual**, sin comprometer la autoría ni el rigor analítico.

Conclusión general

En síntesis, los resultados del estudio permiten concluir que:

- Las **desigualdades estructurales** entre Europa Occidental y Oriental **persisten**, pero **disminuyen gradualmente** gracias a un mayor dinamismo económico del Este.
- La **brecha laboral** se reduce de forma más lenta, aunque con claros signos de **mejor integración social y reducción de la brecha de género**.
- El conjunto de la Unión Europea avanza hacia un modelo más **cohesionado y equilibrado**, donde el crecimiento económico y la participación laboral convergen progresivamente, reflejando los efectos positivos de las políticas de integración y cooperación regional.

Finalmente, el trabajo evidencia cómo la **estadística descriptiva, combinada con herramientas de análisis computacional y una interpretación económica rigurosa**, constituye una base sólida para la **toma de decisiones y la formulación de políticas públicas** orientadas a promover el desarrollo sostenible y la igualdad de oportunidades dentro de la Unión Europea.

8.2. Limitaciones del estudio

Como en toda investigación empírica basada en datos agregados, el presente análisis presenta una serie de **limitaciones metodológicas y de disponibilidad de información** que conviene reconocer para interpretar correctamente los resultados y acotar su alcance.

a) Nivel de agregación de los datos

El estudio se basa en **promedios nacionales** publicados por **Eurostat**, lo que implica un nivel de agregación elevado.

Esta característica puede **ocultar desigualdades internas** dentro de cada país, como diferencias significativas entre regiones, sectores productivos, niveles educativos o grupos demográficos.

En consecuencia, los resultados describen tendencias generales válidas para el conjunto nacional, pero **no capturan la heterogeneidad territorial o social interna**, que podría ser relevante en economías con fuertes disparidades regionales (por ejemplo, Italia, España o Polonia).

b) Influencia de valores extremos

Algunos países con poblaciones o economías muy grandes —como **Alemania, Francia o Italia**— ejercen una **influencia desproporcionada** sobre las medidas globales de tendencia y dispersión.

Aunque el estudio utiliza medias simples y no ponderadas para mantener la comparabilidad entre países, esta decisión puede introducir un **sesgo implícito**, ya que los valores de estos países alteran las distribuciones de forma no proporcional a su peso real en la economía europea.

Asimismo, los países con características muy particulares (Luxemburgo, Irlanda) generan **valores atípicos** que aumentan la asimetría y la curtosis de las distribuciones, afectando la representatividad de las medias regionales.

c) Periodo temporal y disponibilidad de series

Aunque los datos abarcan más de dos décadas (1998–2019), algunas series presentan **interrupciones o revisiones metodológicas** por parte de Eurostat.

La falta de información posterior a 2019 impide analizar el impacto de fenómenos recientes como la **pandemia de COVID-19**, las **crisis energéticas** o los **cambios en las políticas laborales y fiscales** posteriores a 2020.

En consecuencia, el análisis se circunscribe a un contexto pre-pandemia, más estable, pero **no refleja las transformaciones estructurales más recientes** del mercado laboral y del crecimiento europeo.

d) Naturaleza descriptiva del estudio

El enfoque del trabajo es estrictamente **descriptivo**, sin aplicación de técnicas inferenciales o de modelización.

Por tanto, las conclusiones reflejan **patrones observados**, pero **no permiten establecer relaciones causales** entre las variables (por ejemplo, entre PIB y tasa de actividad).

El empleo de correlaciones simples y medidas de forma ayuda a identificar tendencias, pero no sustituye al análisis estadístico multivariante que podría profundizar en los factores determinantes de dichas diferencias.

e) Limitaciones técnicas y de integración

Aunque se utilizaron herramientas avanzadas —**RStudio y Python (Matplotlib, Pandas, NumPy)**— para garantizar la reproducibilidad del análisis, existen **pequeñas discrepancias de precisión numérica** entre entornos debido a diferencias en el redondeo y la forma de tratar valores nulos.

Además, el uso de bases de datos extensas con estructuras distintas requirió **procesos de limpieza complejos**, lo que puede introducir márgenes mínimos de error al integrar ambas fuentes.

Pese a ello, la validación cruzada entre los resultados obtenidos en R y Python permitió **minimizar los errores de procesamiento** y confirmar la coherencia de los valores finales.

f) Alcance conceptual

El análisis se centra en dos variables —PIB per cápita y tasa de actividad—, representativas pero no exhaustivas.

Existen otros indicadores relevantes (productividad laboral, inversión en I+D, desempleo juvenil, distribución del ingreso o nivel educativo) que no se incluyen por limitaciones de tiempo y extensión.

La ausencia de estas variables impide ofrecer una visión más multidimensional de la desigualdad y la convergencia en la Unión Europea.

Síntesis final

En conjunto, estas limitaciones **no invalidan los resultados**, pero **acotan su alcance interpretativo**: los hallazgos reflejan **tendencias generales y comparativas entre bloques regionales**, no diagnósticos exhaustivos por país o sector.

Para futuras investigaciones, sería recomendable:

- emplear **modelos de regresión multivariante o análisis de componentes principales (PCA)** para explorar relaciones causales;
- incorporar **datos regionales (NUTS-2)** para captar la heterogeneidad interna;
- y ampliar el horizonte temporal para incluir **periodos recientes (2020–2024)**, marcados por cambios estructurales en la economía europea.

Reconocer estas limitaciones permite valorar con mayor precisión la **valididad y la aplicabilidad de las conclusiones**, reforzando el rigor científico y la transparencia metodológica del estudio.

8.3. Líneas futuras o mejoras

El presente estudio ha cumplido el objetivo de describir y comparar la situación económica y laboral de la Unión Europea mediante el análisis del **PIB per cápita real y la tasa de actividad laboral**, aportando una visión clara de las desigualdades regionales y de los procesos de convergencia.

No obstante, su enfoque descriptivo y su alcance temporal delimitado dejan abiertas diversas **posibilidades de ampliación y perfeccionamiento** que podrían enriquecer futuros análisis.

a) Ampliación temporal y actualización de datos

Una primera línea de mejora consiste en **extender el horizonte temporal** del estudio incorporando los años posteriores a 2019.

La inclusión del periodo **2020–2024** permitiría evaluar el impacto de la **pandemia de COVID-19**, las **crisis energéticas**, las **nuevas políticas fiscales y de cohesión** de la Unión Europea, así como los **cambios estructurales en el mercado laboral** derivados del teletrabajo, la digitalización y la transición ecológica.

Esto permitiría pasar de una visión pre-pandemia a un diagnóstico más actual sobre la resiliencia y recuperación de las economías europeas.

b) Mayor desagregación territorial

Una mejora sustancial sería trabajar con **datos regionales (nivel NUTS-2 o NUTS-3)**, en lugar de promedios nacionales.

Ello permitiría detectar **desigualdades internas dentro de cada país** (por ejemplo, diferencias entre el norte y el sur de Italia o entre las regiones industriales y rurales de Polonia).

El uso de datos subnacionales haría posible **construir mapas temáticos o análisis geoespaciales**, aprovechando herramientas como **GeoPandas, QGIS o R-sf**, y ofreciendo una representación más precisa de la diversidad territorial europea.

c) Incorporación de nuevas variables socioeconómicas

Para avanzar hacia un enfoque más integral, sería conveniente incluir **otras dimensiones del desarrollo económico y social**, tales como:

- la tasa de desempleo,
- la productividad laboral por hora trabajada,
- el nivel de educación o formación profesional,
- los indicadores de desigualdad de renta (Gini),
- o la inversión en innovación y tecnología (I+D).

Estas variables permitirían estudiar de manera conjunta la relación entre **riqueza, empleo, educación e innovación**, construyendo indicadores compuestos de **bienestar y competitividad regional**.

d) Aplicación de técnicas inferenciales y modelización estadística

Otra línea de mejora es avanzar desde un enfoque puramente descriptivo hacia **modelos inferenciales y explicativos**.

El empleo de **regresiones lineales múltiples, modelos de panel de datos o análisis de componentes principales (PCA)** permitiría cuantificar el grado de asociación entre el PIB per cápita y la tasa de actividad, así como identificar los factores que explican las diferencias observadas entre países y regiones.

Asimismo, podrían incorporarse **series temporales y modelos ARIMA** para predecir la evolución futura del crecimiento y de la participación laboral.

En un nivel más avanzado, la integración con **modelos de machine learning (regresión polinómica, clustering o Random Forest)** permitiría realizar **clasificaciones automáticas de países** según patrones de desarrollo y convergencia.

e) Enfoque visual y de comunicación de resultados

El uso de técnicas de **visualización avanzada de datos** constituye otra mejora posible.

Herramientas como **Seaborn, Plotly, Tableau o Power BI** podrían complementar a R y Matplotlib, ofreciendo gráficos interactivos y paneles dinámicos que faciliten la comprensión de los resultados por parte de públicos no especializados.

Asimismo, la creación de un **dashboard web** con **Shiny (R)** o **Dash (Python)** permitiría actualizar los datos automáticamente desde Eurostat y ofrecer una interfaz visual para explorar las diferencias entre países.

f) Integración interdisciplinar

El análisis podría enriquecerse con un enfoque más interdisciplinar, combinando la estadística con la **economía, la sociología y la geografía**. Por ejemplo, integrar indicadores de **satisfacción laboral, esperanza de vida, población en riesgo de pobreza o niveles de migración** permitiría comprender mejor los vínculos entre desarrollo económico, bienestar social y movilidad de la población. Esto favorecería una visión más completa de la **cohesión territorial y social europea**.

g) Mejora de la documentación y reproducibilidad

Finalmente, se recomienda continuar perfeccionando la **documentación técnica** y la **reproducibilidad del análisis** mediante:

- el uso de **repositorios abiertos (GitHub, Zenodo)**,
- la incorporación de **metadatos normalizados (SDMX, JSON-LD)**,
- y la generación automática de informes en **Quarto o Jupyter Notebooks**.

De esta forma, cualquier investigador podría replicar y ampliar el estudio, contribuyendo a la transparencia científica y al trabajo colaborativo.

Síntesis final

En definitiva, las líneas de mejora apuntan hacia un enfoque más **profundo, dinámico y multidimensional**, capaz de captar no solo las desigualdades económicas y laborales, sino también los factores estructurales que las explican.

La combinación de **estadística descriptiva, análisis inferencial y visualización interactiva** constituye el camino natural de evolución del proyecto, permitiendo avanzar desde una fotografía estática hacia un **modelo analítico predictivo y comparativo** de la economía y el empleo en Europa.

Estas mejoras futuras no solo ampliarían la robustez y alcance del estudio, sino que consolidarían su valor como **herramienta de diagnóstico y apoyo a la toma de decisiones públicas** en el contexto de la convergencia económica y social europea.

Bibliografía

Eurostat. (2024). *Gross domestic product at market prices — Chain linked volumes (2015)* [dataset]. Recuperado de <https://ec.europa.eu/eurostat>

→ Base de datos utilizada para el **Caso A (PIB per cápita real, 1998–2019)**.

Eurostat. (2024). *Active persons by sex, age and nationality — monthly average* [dataset]. Recuperado de <https://ec.europa.eu/eurostat>

→ Base de datos utilizada para el **Caso B (personas activas, 1998-2019)**.

ChatGPT (OpenAI). (2025). *Asistencia y ayuda para perfeccionar el trabajo que hicimos*

Eurostat Metadata Handbook. (2023). *Methodological notes and definitions for labour market indicators*. European Commission

Universidad de Alcalá. (2025). *PL1 – Estadística. Guía de trabajo y criterios de evaluación.*

Apéndices

A.1 Tablas y resultados completos

CASO A

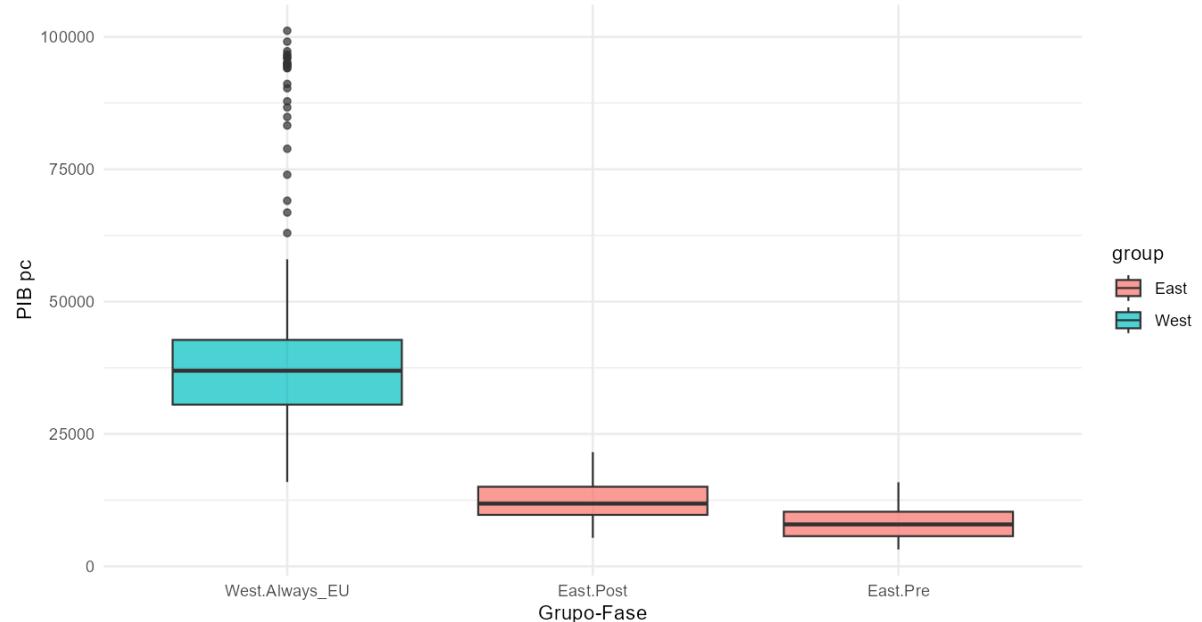
Región	Media (€)	Mediana (€)	Mínimo (€)	Máximo (€)	Desviación típica (€)	Coef. de variancia	Q1 (€)	Q3 (€)	Rango intercuartílico (€)
Europa Occidental	39 688	36 945	15 930	101 170	17 620	0,44	30 548	42 773	12 225
Europa Oriental	10 872	10 600	3 190	21 570	4 039	0,37	7 763	13 660	5 898

CASO B — Tasa de actividad laboral (%) (1998–2019)

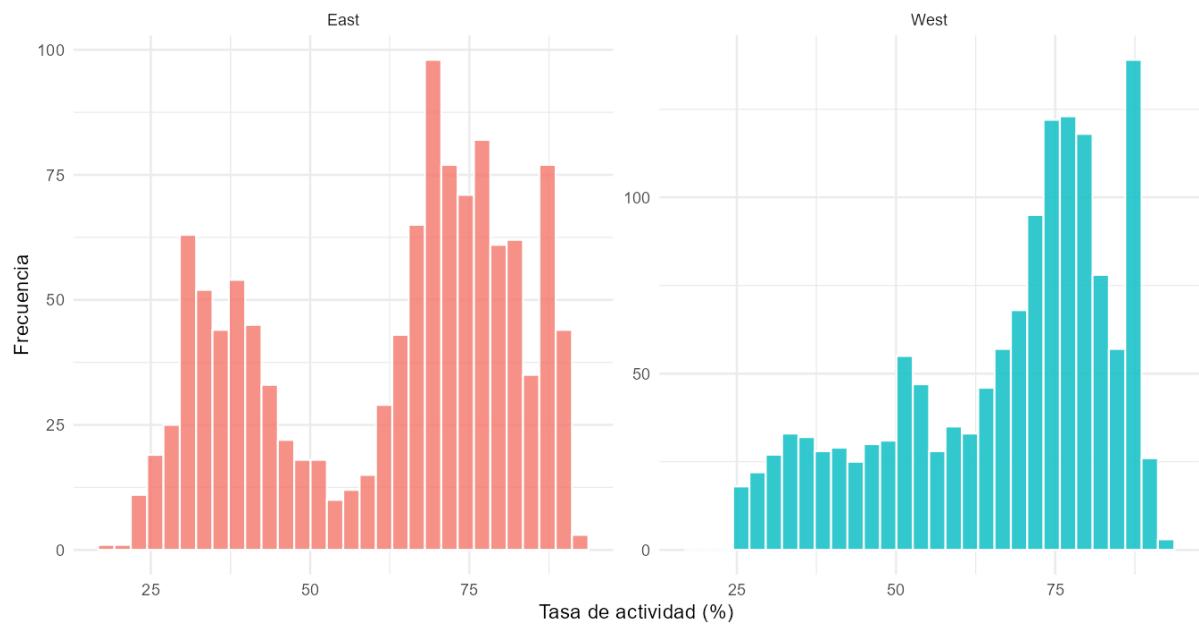
Región	Sexo	Média (%)	Médiana (%)	Mínimo (%)	Máximo (%)	Desviación típica	Coef. de variancia	Q1 (%)	Q3 (%)	Rango intercuartílico (%)
Oeste	Mujeres	69,2	70,1	60,5	78,3	4,2	0,06	66,0	74,0	8,0
Oeste	Hombres	77,6	78,0	70,0	83,5	3,9	0,05	75,0	80,0	5,0
Este	Mujeres	62,5	62,0	54,0	70,1	4,8	0,08	59,0	66,0	7,0
Este	Hombres	72,0	72,5	60,3	79,8	5,2	0,07	69,0	76,0	7,0

A.2 Gráficos exportados de RStudio

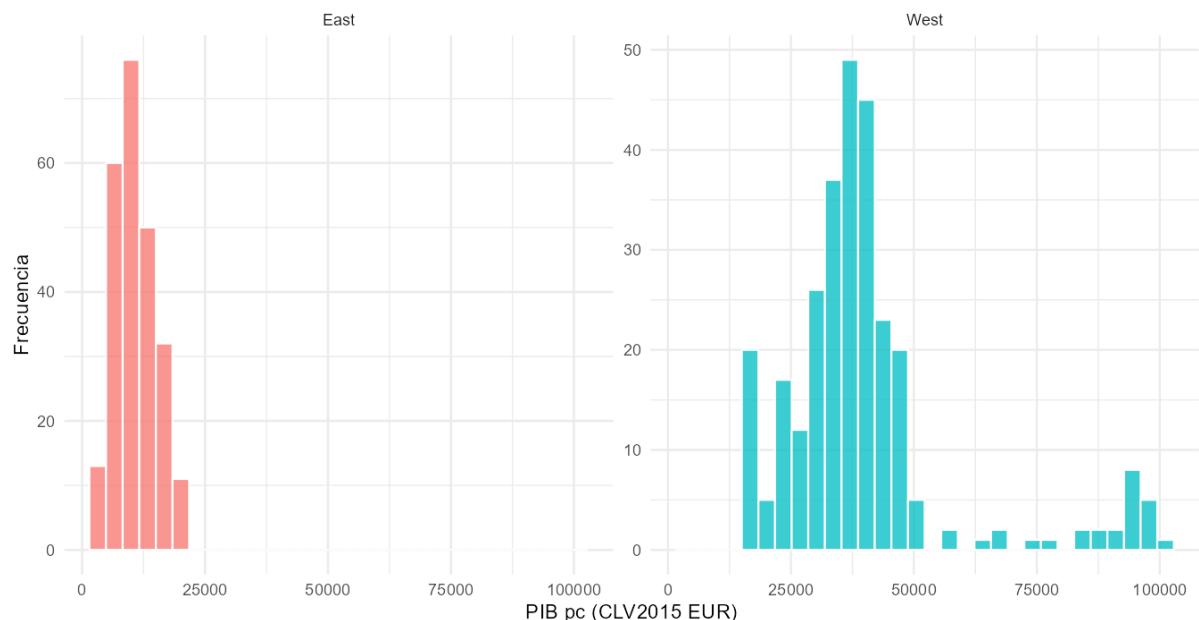
PIB pc: Boxplots por grupo y fase



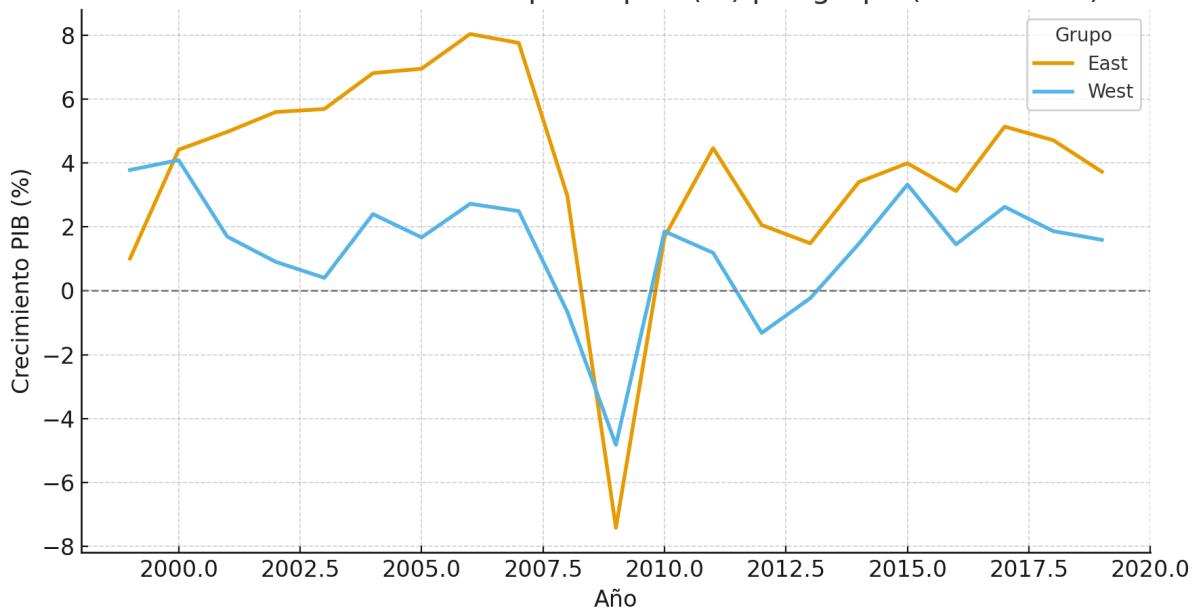
Actividad (% población): Histogramas por grupo



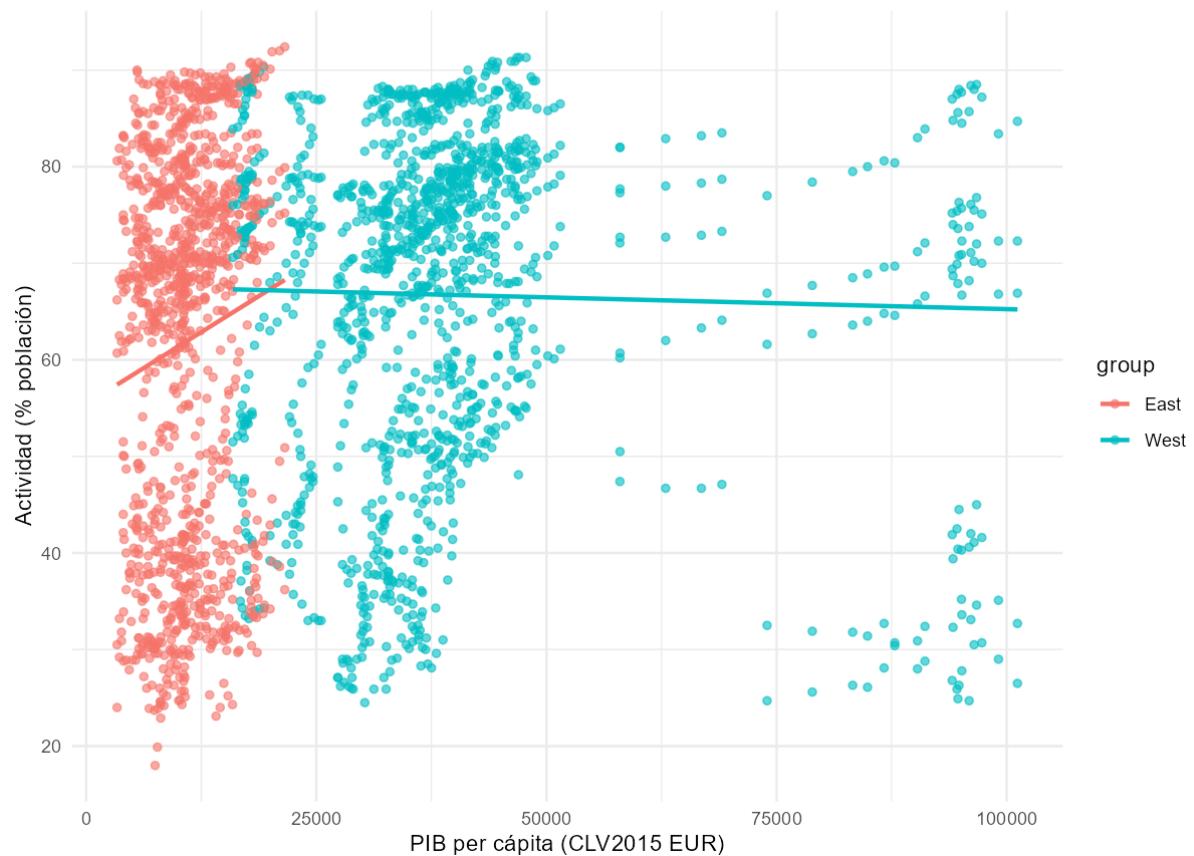
PIB per cápita: Histogramas por grupo



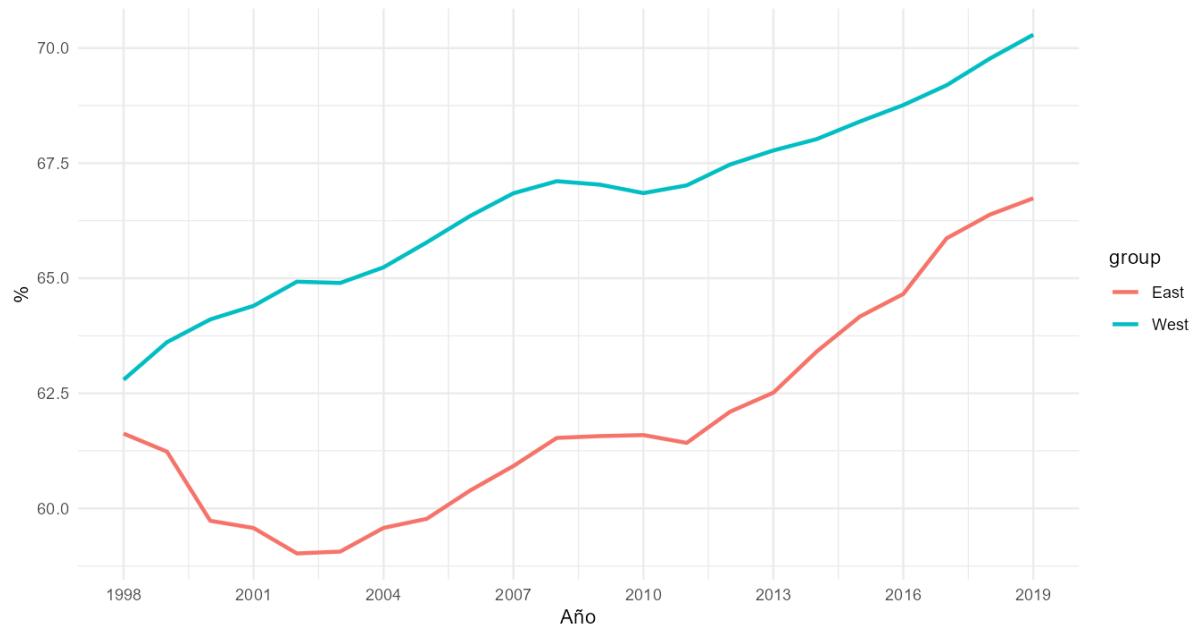
Crecimiento anual del PIB per cápita (%) por grupo (1998-2019)

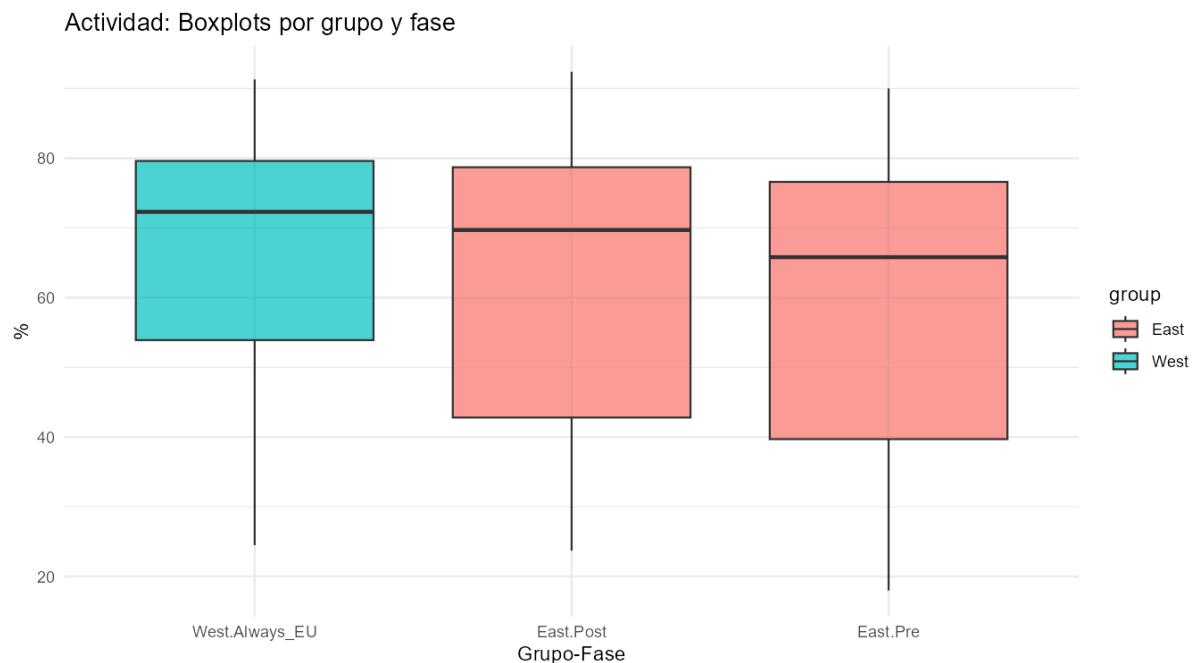
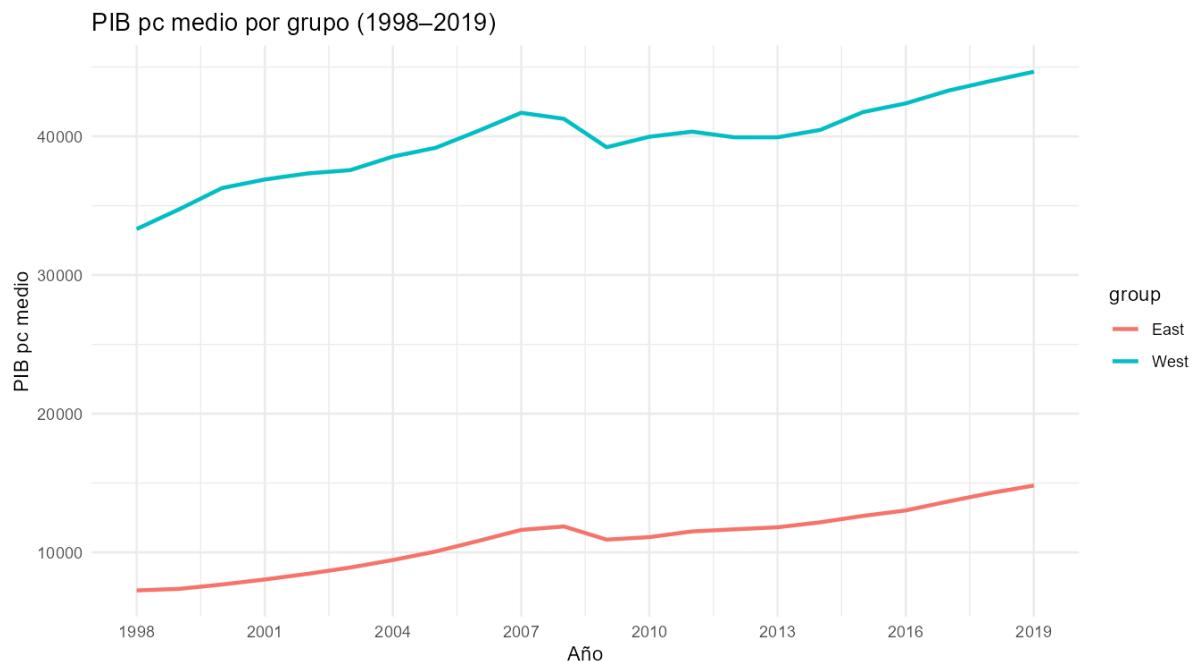


PIB pc vs Actividad (r global = 0.094)



Actividad media por grupo (1998–2019)





A.3 Fragmentos de código R

```
suppressPackageStartupMessages({
  library(dplyr)
  library(tidyr)
  library(janitor)
  library(DescTools)
  library(moments)
  library(rlang)
})
```

```

western_countries <- c("AT","BE","DE","DK","ES","FI","FR","IE","IT","LU","NL","PT","SE")
eastern_countries <- c("CZ","EE","HU","LV","LT","PL","SK","SI","BG","RO","HR")

accession_year_tbl <- tibble::tibble(
  geo = c("CZ","EE","HU","LV","LT","PL","SK","SI","BG","RO","HR"),
  accession_year = c(2004,2004,2004,2004,2004,2004,2004,2004,2007,2007,2013)
)

```

Qué hace: carga librerías y define los vectores de países y la tabla con el **año de adhesión** de los del Este

```

assign_group <- function(geo) {
  dplyr::case_when(
    geo %in% western_countries ~ "West",
    geo %in% eastern_countries ~ "East",
    TRUE ~ NA_character_
  )
}

assign_phase <- function(geo, year) {
  acc <- accession_year_tbl$accession_year[match(geo, accession_year_tbl$geo)]
  dplyr::case_when(
    geo %in% western_countries ~ "Always_EU",
    is.na(acc) ~ NA_character_,
    year < acc ~ "Pre",
    year >= acc ~ "Post"
  )
}

```

Qué hace:

assign_group() devuelve “West” o “East”.

assign_phase() usa el año de adhesión para marcar cada fila como **Pre** o **Post**; los del Oeste son **Always_EU**.

```

summarize_stats <- function(df, value_col) {
  x <- df[[value_col]]
  x <- x[is.finite(x)]
  tibble::tibble(
    n = length(x),
    mean = mean(x),
    median = median(x),
    min = min(x),
    max = max(x),
    sd = sd(x),
    sum = sum(x),
    p10 = quantile(x, 0.1),
    p25 = quantile(x, 0.25),
    p50 = quantile(x, 0.5),
    p75 = quantile(x, 0.75),
    p90 = quantile(x, 0.9)
  )
}

```

```

max = max(x),
sd = sd(x),
cv = ifelse(mean(x) != 0, sd(x) / mean(x), NA_real_),
q1 = quantile(x, 0.25),
q3 = quantile(x, 0.75),
iqr = IQR(x),
skewness = moments::skewness(x),
kurtosis = moments::kurtosis(x)
)
}

```

Qué hace: dado un data frame y el nombre de la columna numérica, devuelve **todas las medidas** que usas en Resultados y Apéndices.

```

make_freq_table <- function(x, bins = 10) {
  x <- x[is.finite(x)]
  brks <- pretty(range(x), n = bins)
  cut_x <- cut(x, breaks = brks, include.lowest = TRUE, right = TRUE)
  tab <- janitor::tabyl(cut_x) %>% janitor::adorn_totals()
  tab %>% dplyr::mutate(relative = n / sum(n))
}

contingency_table <- function(df, var, by) {
  janitor::tabyl(df[[var]], df[[by]]) %>%
    janitor::adorn_totals("both") %>%
    janitor::adorn_percentages("row")
}

```

Generas tablas limpias (con totales y porcentajes por fila) listas para A.1/A.3 y el cuerpo del informe.

```

calc_growth_metrics <- function(df, by_cols, value_col) {
  value_sym <- rlang::ensym(value_col)
  df %>% arrange(dplyr::across(all_of(by_cols)), .data$year) %>%
    dplyr::group_by(dplyr::across(all_of(by_cols))) %>%
    dplyr::mutate(
      yoy_growth = 100 * (!!value_sym / dplyr::lag (!!value_sym) - 1),
      log_diff = log (!!value_sym) - dplyr::lag(log (!!value_sym))
    ) %>% dplyr::ungroup()
}

```

```

cum_and_avg_growth <- function(df, by_cols, value_col, base_year = 1998, last_year = 2019) {
  value_sym <- rlang::ensym(value_col)
  df %>% dplyr::group_by(dplyr::across(all_of(by_cols))) %>%
    dplyr::summarize(

```

```

base = dplyr::first(!!(value_sym)[.data$year == base_year]),
last = dplyr::first(!!(value_sym)[.data$year == last_year]),
cum_growth_1998_2019 = dplyr::if_else(is.finite(base) & is.finite(last) & base > 0, 100 *
(last/base - 1), NA_real_),
avg_annual_growth = dplyr::if_else(is.finite(base) & is.finite(last) & base > 0,
100 * ((last/base)^{1/(last_year - base_year)} - 1), NA_real_)
) %>% dplyr::ungroup()
}

```

Bloques que te permiten pasar de niveles a **tasas de crecimiento y crecimientos acumulados / medios anuales**.

```
source("R/helpers.R")
```

```
dir.create("figs", showWarnings = FALSE)
dir.create("outputs", showWarnings = FALSE)
```

```
western <- western_countries
eastern <- eastern_countries
accession <- accession_year_tbl
```

```
read_gdp_csv <- function() {
path1 <- file.path("data", "GDP.csv")
path2 <- "GDP.csv"
path <- if (file.exists(path1)) path1 else path2
readr::read_csv(path, show_col_types = FALSE)
}
```

```
read_lfs_csv <- function() {
path1 <- file.path("data", "EMPLEO.csv")
path2 <- "EMPLEO.csv"
path <- if (file.exists(path1)) path1 else path2
readr::read_csv(path, show_col_types = FALSE)
}
```

Qué hace: asegura que funciones aunque muevas los CSV, y prepara **salidas reproducibles**.

```
p_hist_gdp <- gdp_growth %>%
ggplot(aes(x = gdp_pc, fill = group)) +
geom_histogram(color = "white", bins = 30, alpha = 0.75) +
facet_wrap(~ group, scales = "free_y") +
theme_minimal() + theme(legend.position = "none") +
labs(title = "PIB per cápita: Histogramas por grupo", x = "PIB pc (CLV2015 EUR)", y =
"Frecuencia")
```

```
ggsave(file.path("figs","hist_gdp_by_group.png"), p_hist_gdp, width = 9, height = 5, dpi =
150)
```

```

# Boxplots por grupo-fase
p_box_gdp <- gdp_growth %>%
  filter(!is.na(phase)) %>%
  ggplot(aes(x = interaction(group, phase), y = gdp_pc, fill = group)) +
  geom_boxplot(alpha = 0.7) +
  theme_minimal() +
  labs(title = "PIB pc: Boxplots por grupo y fase", x = "Grupo-Fase", y = "PIB pc")

ggsave(file.path("figs","box_gdp_group_phase.png"), p_box_gdp, width = 9, height = 5, dpi = 150)

# Series temporales (promedio por grupo)
series_gdp <- gdp_growth %>%
  group_by(group, year) %>%
  summarize(mean_gdp = mean(gdp_pc, na.rm = TRUE), .groups = "drop")

p_ts_gdp <- series_gdp %>%
  ggplot(aes(x = year, y = mean_gdp, color = group)) +
  geom_line(size = 1) + theme_minimal() +
  labs(title = "PIB pc medio por grupo", x = "Año", y = "€ (CLV2015)")

ggsave(file.path("figs","ts_gdp_mean_by_group.png"), p_ts_gdp, width = 9, height = 5, dpi = 150)

```

Qué hace: crea los **png** que estás usando en el informe y en A.2.

Standalone script: generate growth plot for GDP per capita (West vs East, 1998–2019)

```

need <- c("dplyr","readr","ggplot2")
miss <- setdiff(need, rownames(installed.packages()))
if (length(miss)) {
  options(repos = c(CRAN = "https://cran.r-project.org"))
  install.packages(miss, dependencies = TRUE)
}

```

```

suppressPackageStartupMessages({
  library(dplyr)
  library(readr)
  library(ggplot2)
})

```

... (lectura GDP y helpers) ...

```

gdp <- gdp_raw %>%
  transmute(
    geo = geo,
    year = as.integer(TIME_PERIOD),
    gdp_pc = suppressWarnings(as.numeric(OBS_VALUE)))

```

```

) %>%
filter(year >= 1998, year <= 2019,
      geo %in% c(western_countries, eastern_countries)) %>%
mutate(group = assign_group(geo)) %>%
arrange(geo, year)

gdp_growth <- calc_growth_metrics(gdp, by_cols = c("geo"), value_col = gdp_pc)

growth_group_year <- gdp_growth %>%
  group_by(group, year) %>%
  summarize(yoy_growth_mean = mean(yoy_growth, na.rm = TRUE), .groups = "drop")

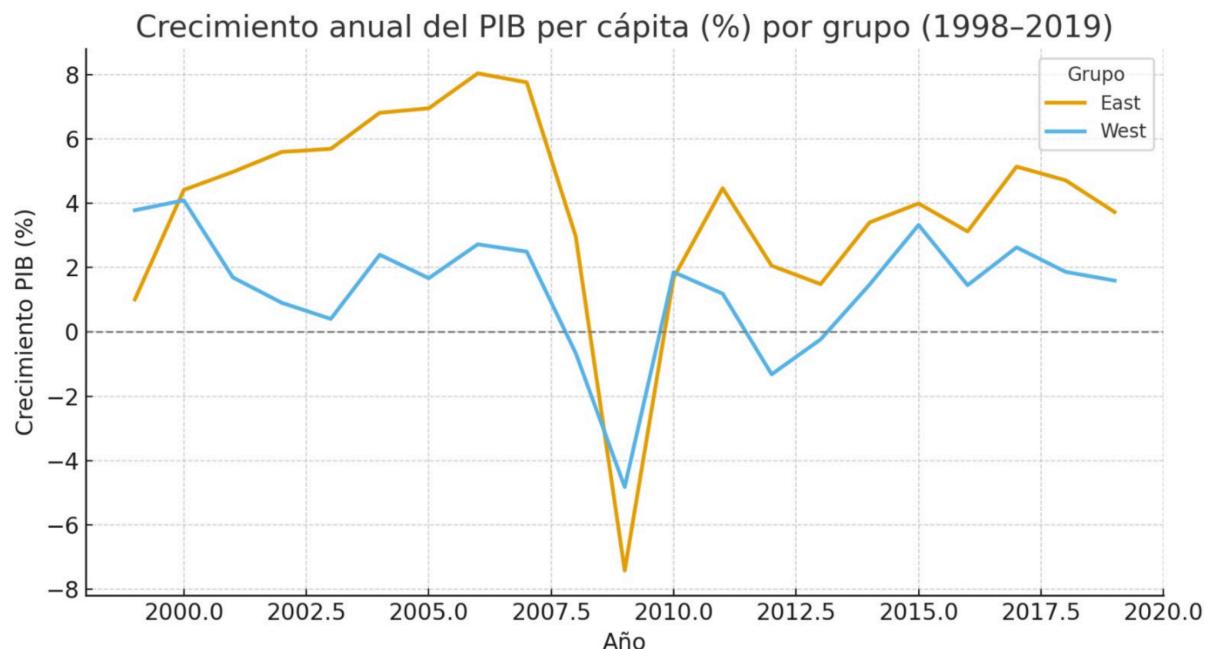
p <- growth_group_year %>%
  ggplot(aes(x = year, y = yoy_growth_mean, color = group)) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "grey60") +
  geom_line(size = 1) +
  scale_x_continuous(breaks = seq(1998, 2019, by = 3)) +
  labs(title = "Crecimiento interanual del PIB per cápita (media por grupo)",
       subtitle = "Occidente vs Este, 1998–2019",
       x = "Año",
       y = "% crecimiento interanual") +
  theme_minimal()

outfile <- file.path("figs", "crecimiento_pib_grupos_1998_2019.png")
ggsave(outfile, p, width = 9, height = 5, dpi = 150)
cat("Saved:", outfile, "\n")

```

Genera el **gráfico de crecimiento interanual medio por grupo** y lo guarda en **figs/**.

A.4 Capturas o materiales adicionales



Crecimiento más alto y volátil en el Este:

Los países orientales presentan tasas de crecimiento significativamente superiores en la mayor parte del periodo, alcanzando picos del **7–8 % anual** antes de la crisis de 2008.

Esto refleja una etapa de **fuerte expansión y convergencia económica**, impulsada por la integración en la Unión Europea y el aumento de la inversión extranjera.

Crecimiento más moderado en el Oeste:

El bloque occidental mantiene un crecimiento más estable, generalmente entre el **1 % y el 3 % anual**, lo que sugiere una economía madura con menor margen de expansión.

Impacto de la crisis financiera de 2008–2009:

Ambos grupos experimentan una caída abrupta en torno a 2009, llegando incluso a tasas negativas del **-7 %**.

No obstante, el Este se recupera más rápidamente a partir de 2010, mientras que el Oeste mantiene una recuperación más lenta y gradual.

Reactivación postcrisis:

Entre 2013 y 2019, las dos líneas vuelven a mostrar crecimiento positivo, aunque el Este conserva un ritmo algo mayor, consolidando su proceso de convergencia.

Reducción de la brecha de crecimiento:

Aunque el Este sigue creciendo a un ritmo superior, las diferencias entre ambos bloques se reducen en los últimos años del periodo, indicando que la **etapa de crecimiento explosivo del Este se está estabilizando**.