



Facultad de Ciencias Económicas y Empresariales

# **UN CAMINO HACIA LA SOSTENIBILIDAD: LA HUELLA MEDIOAMBIENTAL EN LA INDUSTRIA ALIMENTARIA**

EL PAPEL DE LA ANALÍTICA DE DATOS

Autor: Diego Díez Borque

Director: Raúl González Fabre

MADRID | Junio 2024

*“La emergencia climática es una carrera que estamos perdiendo,  
pero es una carrera que podemos ganar. La crisis climática  
está causada por nosotros y las soluciones deben venir de nosotros.  
Tenemos las herramientas: la tecnología está de nuestro lado.”*

ANTONIO GUTERRES  
SECRETARIO GENERAL DE LA ONU

## **Resumen**

Desde el cultivo hasta el consumo, toda la cadena de valor de la industria alimentaria impone una gran presión sobre los recursos naturales de nuestro planeta, contribuyendo a graves problemas medioambientales como el cambio climático, la pérdida de biodiversidad, la contaminación del agua y la degradación del suelo. Abordar estos impactos resulta crucial, no solo por motivos ambientales, sino también por razones de justicia social y económica. La seguridad alimentaria, la salud pública y el bienestar económico dependen de ecosistemas saludables y de prácticas agrícolas sostenibles. Por lo tanto, la industria alimentaria enfrenta el reto de producir alimentos de manera eficiente y accesible, mientras disminuye su huella de carbono. Superar este desafío requiere herramientas innovadoras como la analítica de datos y tecnologías avanzadas, tales como la inteligencia artificial. Estas tecnologías permiten optimizar procesos, reducir desperdicios, mejorar la eficiencia energética y desarrollar prácticas agrícolas sostenibles, facilitando la toma de decisiones basada en datos para avanzar hacia sistemas alimentarios más resilientes y sostenibles.

**Palabras clave** – Industria Alimentaria, Huella Medioambiental, Emisiones de Gases de Efecto Invernadero, Machine Learning, Análisis Descriptivo.

## **Abstract**

From farming to consumption, the entire food industry value chain puts great pressure on our planet's natural resources, contributing to serious environmental problems such as climate change, biodiversity loss, water pollution and soil degradation. Addressing these impacts is crucial, not only for environmental reasons, but also for reasons of social and economic justice. Food security, public health and economic well-being depend on healthy ecosystems and sustainable agricultural practices. Therefore, the food industry faces the challenge of producing food efficiently and affordably, while decreasing its carbon footprint. Overcoming this challenge requires innovative tools such as data analytics and advanced technologies such as artificial intelligence. These technologies allow optimizing processes, reducing waste, improving energy efficiency and developing sustainable agricultural practices, facilitating data-driven decision-making to move towards more resilient and sustainable food systems.

**Key Words** – Food Industry, Environmental Footprint, Greenhouse Gas Emissions, Machine Learning, Descriptive Analysis.

## ÍNDICE DE CONTENIDOS

☞ Resumen / Abstract .....	3
☞ Palabras clave / Key words .....	3
☞ Tabla de Figuras e Ilustraciones .....	5
1. INTRODUCCIÓN .....	6
1.1. Contextualización del problema .....	6
1.2. Objetivos .....	7
1.3. Metodología .....	8
2. HUELLA MEDIOAMBIENTAL .....	9
2.1. El efecto de la industria sobre el planeta. Ejemplificación cuantitativa .....	9
2.2. Indicadores existentes .....	10
2.3. Situación actual. Riesgos que valorar como sociedad .....	11
3. ESTRATEGIAS ANALÍTICAS APLICABLES .....	14
3.1. Inteligencia Artificial y Machine Learning .....	14
3.2. Aplicaciones de la IA y el ML .....	15
4. CASO DE ESTUDIO: LA INDUSTRIA ALIMENTARIA .....	17
4.1. Introducción. Huella ecológica de lo que comemos .....	17
4.2. La responsabilidad social corporativa en la industria alimentaria .....	19
4.3. La sostenibilidad como oportunidad estratégica .....	20
5. APLICACIÓN PRÁCTICA DE TÉCNICAS DE MACHINE LEARNING AL CASO DE ESTUDIO .....	23
5.1. Exploración de la base de datos y <i>Data Cleaning</i> .....	23
5.2. Análisis descriptivo e interpretación de resultados .....	27
5.3. Análisis de Correlación .....	30
5.4. Análisis de Componentes Principales (PCA) .....	32
5.5. Análisis de Regresión Lineal Simple .....	34
6. CONCLUSIONES .....	37
7. BIBLIOGRAFÍA .....	40
ANEXO I: CÓDIGO DE PYTHON .....	43

## TABLA DE FIGURAS E ILUSTRACIONES

FIGURA 1: MUESTRA DEL CONJUNTO DE DATOS .....	24
FIGURA 2: CLASIFICACIÓN DE LOS ALIMENTOS POR CATEGORÍAS .....	26
FIGURA 3: EMISIONES TOTALES POR PRODUCTO, FILTRADOS POR CATEGORÍA .....	27
FIGURA 4: EMISIONES TOTALES DEL TOP 10 DE ALIMENTOS .....	28
FIGURA 5: EMISIONES TOTALES POR KG DE CATEGORÍA A LO LARGO DE LA CADENA DE SUMINISTRO .....	29
FIGURA 6: MATRIZ DE CORRELACIÓN DE LAS EMISIONES GEI .....	31
FIGURA 7: PCA DE LAS EMISIONES GEI .....	33
FIGURA 8: COEFICIENTES DE CARGA EN LOS COMPONENTES PRINCIPALES .....	34
FIGURA 9: CÓDIGO Y OUTPUT DEL MODELO DE REGRESIÓN LINEAL .....	35

## **1. INTRODUCCIÓN**

### **1.1. CONTEXTUALIZACIÓN DEL PROBLEMA**

La industria alimentaria se posiciona en el corazón del desarrollo sostenible a nivel global, desempeñando un papel central en la nutrición humana y la economía, pero siendo también fuente significativa de presión sobre el medioambiente. Su cadena de valor, desde la producción agrícola hasta el procesamiento, distribución y consumo de alimentos, incide significativamente en los recursos naturales, contribuyendo a desafíos ambientales de gran envergadura como el cambio climático, la deforestación, la pérdida de biodiversidad, la contaminación del agua y la degradación del suelo. La creciente demanda de alimentos, impulsada por el aumento de la población mundial y el cambio en los patrones de consumo, agrava aún más estos impactos, poniendo en riesgo la capacidad del planeta para sostener cualquier vida futura.

La urgencia de abordar estos impactos no es solo una cuestión de conservación ambiental, sino también de equidad social y económica. La seguridad alimentaria, la salud humana y el bienestar económico dependen intrínsecamente de ecosistemas saludables y de la sostenibilidad de las prácticas agrícolas y alimentarias. Por lo tanto, la industria alimentaria se encuentra en una encrucijada crítica, donde debe reconciliar la necesidad de producir alimentos de manera eficiente y accesible con el imperativo de reducir su huella medioambiental y transitar hacia prácticas más sostenibles.

En este contexto, la huella medioambiental de la industria alimentaria emerge como un indicador clave para evaluar y entender la magnitud de su impacto en el ecosistema. Esta huella engloba diversas dimensiones, incluyendo el uso del agua, la emisión de gases de efecto invernadero, el uso del suelo y la energía consumida a lo largo de toda la cadena de suministro alimentario. La identificación y cuantificación de estas secuelas es fundamental para diseñar estrategias que permitan una producción y consumo de alimentos más sostenibles.

Sin embargo, enfrentar este desafío requiere de herramientas y enfoques innovadores capaces de manejar la complejidad y la escala de la problemática. Aquí es donde la

analítica de datos juega un papel transformador, ofreciendo soluciones avanzadas para recopilar, procesar y analizar grandes volúmenes de datos relacionados con la producción y consumo de alimentos. Mediante el uso de tecnologías y enfoques analíticos precisos como la Inteligencia Artificial – en adelante, IA –, que incluye el Machine Learning – en adelante, ML – y uso del Big Data, es posible obtener *insights* precisos sobre cómo optimizar los procesos, reducir el desperdicio, mejorar la eficiencia energética y desarrollar prácticas agrícolas más sostenibles.

La adopción de la analítica de datos en la industria alimentaria no solo facilita la toma de decisiones basada en evidencia para minimizar la huella medioambiental, sino que también abre caminos hacia la innovación en productos y servicios alimentarios que son esenciales para la transición hacia sistemas alimentarios más resilientes y sustentables. Por lo tanto, comprender su papel en este contexto es crucial para enfrentar los desafíos actuales y futuros de la industria alimentaria, marcando el camino hacia la sostenibilidad. Este enfoque fundamenta el presente ensayo y guiará el análisis y la discusión de todo su contenido.

## 1.2. OBJETIVOS

El objetivo de este trabajo es explorar profundamente el rol que la analítica de datos puede desempeñar en la promoción de una industria alimentaria más sostenible desde el punto de vista ecológico. La relevancia de esta investigación radica en su potencial para identificar, a través del análisis de datos, los aspectos que deben ser prioritariamente abordados para mitigar los impactos negativos de la industria alimentaria en el medio ambiente. Para lograr este fin, se propone una investigación que no solo se enfoque en el papel transformador de la analítica de datos, sino que también aborde una serie de preguntas que ayudarán a contextualizar y profundizar en el análisis:

- ¿Cuál es el efecto de la industria sobre el planeta? ¿En qué punto nos encontramos?
- ¿Es posible la aplicación de herramientas de IA para limitar el efecto destructor de las acciones industriales sobre el planeta? ¿Es esto viable con la industria alimentaria?

### 1.3. METODOLOGÍA

Para abordar este proyecto con la amplitud y profundidad que requiere, se adoptará un enfoque multidisciplinario que integre el análisis de literatura científica y técnica actualizada en el ámbito de la sostenibilidad industrial, con especial énfasis en la industria alimentaria. Este análisis permitirá comprender las tendencias actuales, los desafíos clave y las mejores prácticas en el sector, sentando las bases de un modelo de operación ecológicamente eficiente para empresas alimentarias. La revisión de la literatura abarcará estudios de caso, investigaciones empíricas, teorías de gestión sostenible y tecnologías emergentes aplicadas a la sostenibilidad.

La segunda parte se centrará en el análisis de una relevante base de datos que refleja una muestra de la huella de carbono producida por la industria alimentaria. Esta base incluirá información sobre las emisiones de gases de efecto invernadero producidas por una serie de alimentos a lo largo de toda su cadena de producción. La selección del *dataset* se guiará por criterios de relevancia, fiabilidad y actualidad, asegurando que los datos sean representativos del impacto ambiental global e histórico del sector.

Para extraer *insights* significativos de estos datos, se aplicarán técnicas de ML para el análisis descriptivo. El objetivo será desarrollar un enfoque que permita identificar patrones, entender relaciones y evaluar el impacto de diferentes prácticas de gestión en la huella de carbono y otros medidores ambientales críticos. Este análisis descriptivo servirá como herramienta fundamental para comprender la situación actual y orientar futuras investigaciones y acciones estratégicas hacia la minimización de los impactos negativos y la optimización de la eficiencia ecológica. Además, se realizará una demostración de visualización de datos para presentar de manera clara y comprensible los hallazgos del estudio. Las visualizaciones ayudarán a ilustrar las conexiones entre diferentes variables, el impacto de diversas prácticas industriales en la sostenibilidad y el potencial de mejoras específicas. En definitiva, se pretende mostrar cómo la integración de la IA en las operaciones industriales puede proporcionar un conocimiento profundo de la situación actual y señalar áreas clave para la mejora continua en el camino hacia los objetivos de sostenibilidad, identificando tanto oportunidades como desafíos asociados con su implementación.



## 2. HUELLA MEDIOAMBIENTAL

### 2.1. EL EFECTO DE LA INDUSTRIA SOBRE EL PLANETA. EJEMPLIFICACIÓN CUANTITATIVA

Resulta evidente que la actividad industrial, al igual que la mayoría de las actividades humanas, afecta al entorno en el que se desarrolla. Algunas veces, este impacto es inmediato y de gran magnitud, como lo demuestran desastres como el de Chernóbil, los cuales han sido catalizadores de cambios significativos en la legislación industrial a nivel mundial. En otras situaciones, el efecto es más restringido geográficamente y no produce eventos dramáticos, pero se mantiene a lo largo del tiempo debido a la naturaleza continua de la actividad que lo origina. Ejemplos de esto son las acumulaciones de residuos mineros, el deterioro gradual de ríos, o la polución del aire en zonas con alta concentración industrial, tal como sucedió en el pasado en Bilbao.

Está demostrado que la calidad del ambiente constituye un requisito indispensable para la salud humana y el desarrollo sostenible. El plan de acción de las Naciones Unidas para el desarrollo en el siglo XXI, denominada “Agenda 21”, acordada en la Cumbre de la Tierra de 1992 efectuada en Río de Janeiro, reitera que *“los seres humanos constituyen el centro de las preocupaciones para un desarrollo sostenible. Ellos tienen derecho a una vida saludable y productiva, en armonía con la naturaleza”*. Es responsabilidad de la ONU y de distintos organismos internacionales brindar información sobre procesos de producción menos contaminantes a través del Programa de Producción Limpia del PNUMA (Suárez Tamayo & Molina, 2014).

El impacto de la industria se suele producir y medir en las siguientes cuatro áreas: aire, agua, energía y residuos. El último informe del Ministerio de Industria y Turismo nos muestra, por un lado, que en 2022 las emisiones de gases de efecto invernadero en España ascendieron a 293.778 kilotoneladas equivalentes de CO<sub>2</sub>, reflejando un ligero incremento respecto al año anterior, si bien se observa una tendencia general de reducción en las emisiones totales, desde las 383.276 kilotoneladas producidas en el año 2000. Como grandes contribuciones hacia el desarrollo sostenible, destacan las industrias del sector energético y las de combustibles sólidos (Ministerio de Industria y Turismo, 2024).

De forma paralela, mientras que la polución en cuerpos de agua tanto continentales como marinos parece estar bajo control, según los análisis realizados en ríos y costas de España, el desafío más significativo radica en la limitada disponibilidad de agua apta para el consumo y uso industrial. A pesar de que la agricultura (77%) y el uso residencial (18%) constituyen los mayores consumidores de agua, el sector industrial ha incrementado su demanda paralelamente al crecimiento económico. Aunque este consumo se ha estabilizado gracias a las iniciativas implementadas por la industria, dichas medidas resultan insuficientes teniendo en cuenta la presión de la alta demanda urbana y la reducción de las lluvias (INE, 2020).

El crecimiento de la actividad industrial va acompañado de un aumento tanto de consumo energético como de la consecuente generación de residuos. A pesar de la mejor calidad del consumo energético industrial y la transición hacia la electricidad, el gas natural y las energías renovables en detrimento del carbón y el petróleo, este aún se halla muy por encima de la media europea, y con el ritmo actual tardaríamos al menos 40 años en alcanzarla. En lo que respecta a la producción de residuos peligrosos, en el año 2021 alcanzó los 3,7 millones de toneladas. El 70,46% de ellos acabó siendo reciclado, el 21,63% fue vertido y el 7,91% restante fue incinerado (INE, 2023).

## 2.2. INDICADORES EXISTENTES

Medir la huella medioambiental de la industria sobre el planeta implica considerar varios indicadores que reflejan el alcance y la variedad del impacto que las actividades industriales tienen en el medio ambiente. Entre todos ellos, la huella de carbono es quizás el indicador más conocido, que mide la cantidad total de gases de efecto invernadero emitidos directa o indirectamente por una actividad industrial, expresada en equivalente de dióxido de carbono (CO<sub>2</sub>eq).

La huella hídrica evalúa el volumen total de agua dulce utilizado para producir bienes y servicios industriales, incluyendo el agua consumida y contaminada durante todo el proceso de producción. Por su parte, existe un medidor sobre la huella ecológica, utilizando como objeto de estudio la demanda humana que se impone sobre la biosfera,

incluyendo la tierra y el agua necesarias para absorber los residuos generados por la industria y para regenerar los recursos que consume.

La huella energética se refiere a la cantidad de energía consumida en los procesos industriales, incluyendo tanto las fuentes renovables como no renovables; mientras que la huella de materiales contabiliza el peso total de los recursos materiales extraídos para apoyar la actividad industrial. En esa misma línea, la huella toxicológica mide el impacto de las sustancias liberadas por las actividades industriales que pueden ser tóxicas para los ecosistemas y la salud humana.

Finalmente encontramos la huella de residuos, que calcula la cantidad y el impacto de los residuos sólidos y líquidos generados por la industria, y la huella de biodiversidad, que analiza cómo las operaciones industriales afectan la variedad y abundancia de especies en un área dada, incluyendo la pérdida de hábitat y la fragmentación ecológica.

### 2.3. SITUACIÓN ACTUAL. RIESGOS QUE VALORAR COMO SOCIEDAD

Desde la década de 1970 se empezó a dar importancia a la dimensión ambiental como elemento crucial para un desarrollo sostenible, comenzando con la Conferencia de las Naciones Unidas sobre el Medio Humano, celebrada en Estocolmo en 1972, la cual marcó el reconocimiento internacional de la escalada de problemas ambientales a nivel global. La Evaluación de Impacto Ambiental (EIA) nació como una estrategia para proteger el medio ambiente, permitiendo una mejor toma de decisiones en políticas, planes, programas y proyectos, integrando consideraciones ambientales en los proyectos de inversión desde su concepción. Institucionalizada por primera vez en la legislación estadounidense con la Ley Nacional sobre Política Medioambiental (NEPA) en 1970 y la creación del Consejo de Calidad Ambiental (CEQ), la EIA se posicionó como una medida preventiva para mitigar impactos nocivos de la actividad humana en la naturaleza y las personas. Con el tiempo, la noción de desarrollo sostenible evolucionó para incluir al medio ambiente como un pilar fundamental de la economía, promoviendo un progreso que fuera técnicamente factible, económicamente viable y socialmente aceptable, enfocado en la conservación de recursos naturales. Hoy en día, la EIA es un proceso

analítico esencial que proyecta los impactos tanto negativos como positivos de las actividades propuestas, permitiendo la selección de alternativas y la implementación de controles para prevenir o mitigar efectos adversos y fomentar aquellos beneficiosos (Maza, 2007).

Hablar de la Tierra como “un planeta en riesgo” encierra una novedad que tiene poco que ver con los abundantes –por no decir constantes- desastres que han afectado a unos seres humanos u otros a lo largo de la Historia. Hablar hoy de planeta en riesgo es hablar de una situación de auténtica *emergencia planetaria* (Bybee, 1991), en el sentido de que los riesgos no están acotados localmente, sino que se extienden a todo el planeta, afectando a todos los ecosistemas y amenazando, si no se adoptan urgentemente las medidas que se precisan, con una nueva gran extinción de la que la humanidad sería principal causante y, en última instancia, víctima (Lewin, 1997). La expresión emergencia planetaria pretende así llamar la atención sobre la gravedad del conjunto de riesgos y serios problemas de alcance global y estrechamente relacionados a los que la humanidad ha de hacer hoy frente y que podemos resumir de la siguiente manera (Vilches & Gil, 2008):

- Una contaminación pluriforme y omnipresente que afecta a tierras, ríos y océanos, deteriorando la calidad del aire y acelerando el cambio climático, lo que representa una amenaza para la habitabilidad de la Tierra.
- La degradación y agotamiento de una amplia gama de recursos, que va desde la energía hasta los caladeros de pesca, pasando por los bosques, las fuentes de agua dulce y los suelos agrícolas, y que conduce a un incremento de la desertificación y a una disminución en la diversidad biológica.
- Una expansión urbana rápida y caótica que intensifica la contaminación originada por medios como el transporte y la calefacción, y contribuye al agotamiento de recursos mediante la destrucción de áreas agrícolas y un mayor consumo de energía, debido al incremento en los tiempos de traslado.
- La destrucción extendida de ecosistemas, incluyendo bosques, praderas, glaciares, zonas polares, humedales y arrecifes de coral, que se debe a factores como el aumento del efecto invernadero, la sobreexplotación, incendios y la

urbanización desmedida. Esta degradación se manifiesta en una mayor frecuencia e intensidad de eventos climáticos extremos, como sequías, huracanes, inundaciones y deslizamientos, así como en una pérdida de biodiversidad y un avance de la desertificación, afectando desproporcionadamente a millones de personas en situación de pobreza extrema.

- Desigualdades críticas que se evidencian cuando un quinto de la población mundial consume de manera desmedida, frente a cientos de millones que padecen de hambre y viven en condiciones extremadamente difíciles. Todo esto sucede mientras la población global crece más allá de lo que el planeta puede sostener.
- Conflictos variados que se extienden desde guerras destructivas, frecuentemente ligadas al deseo de dominar recursos naturales, hasta las operaciones de mafias y corporaciones multinacionales que anteponen sus propios intereses. Esto incluye actos de terrorismo, genocidios étnicos y la erosión de la diversidad cultural, elementos todos ellos que constituyen un valioso patrimonio mundial en constante riesgo.

### 3. ESTRATEGIAS ANALÍTICAS APLICABLES

En este punto del estudio, nos centraremos en explorar algunas de las estrategias analíticas disponibles para medir la huella medioambiental. Esta fase es fundamental para entender cómo podemos cuantificar y evaluar el impacto de diversas actividades sobre el medio ambiente, abarcando desde la industria hasta las acciones cotidianas individuales. Se trata de profundizar en una serie de metodologías propias del *Business Analytics* que permiten identificar, analizar y gestionar eficazmente la huella que dejamos en el planeta, estableciendo un marco para la implementación de prácticas sostenibles y la toma de decisiones informadas hacia una menor repercusión ambiental.

#### 3.1. INTELIGENCIA ARTIFICIAL Y MACHINE LEARNING

La IA se ha establecido como una rama fundamental de las ciencias de la computación, enfocada en el desarrollo de sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, la resolución de problemas y la percepción. Esta disciplina busca no solo replicar sino también superar la capacidad intelectual humana a través de máquinas y programas altamente eficientes que minimicen la probabilidad de error. Sus aplicaciones son vastas y diversas, incluyendo robótica, videojuegos, y automatización de tareas, contribuyendo a eficiencias operativas en múltiples sectores. Dentro de la IA, el Machine Learning (ML) o Aprendizaje Automático otorga a las máquinas la capacidad de aprender y mejorar a partir de la experiencia, sin necesidad de ser programadas de manera específica para cada tarea. El ML se enfoca en la creación de programas que pueden acceder a datos y utilizarlos para aprender “por su propia cuenta”. Este proceso de aprendizaje se asemeja en gran medida al aprendizaje humano, que sucede a través de la observación y el procesamiento de información, aunque salvando las distancias, pues el ML necesita de un conjunto de datos mucho mayor para extraer un “conocimiento”.

Este proceso comienza con la recopilación de observaciones y datos, a partir de los cuales buscan patrones para hacer inferencias. El objetivo principal del ML es que los sistemas puedan aprender de manera autónoma, sin la necesidad de intervención humana, para posteriormente adaptar sus acciones con el conocimiento adquirido (Escobar, 2022).

Existen hoy en día tres modelos principales de ML:

- **Supervisado:** Entrena máquinas con datos etiquetados para mejorar su precisión, como diferenciar conejos de otras imágenes con etiquetas humanas.
- **No Supervisado:** Descubre patrones en datos no etiquetados, identificando, por ejemplo, diferentes tipos de clientes en compras online.
- **Reforzado:** Utiliza prueba y error y un sistema de recompensas para enseñar a las máquinas a tomar decisiones, aplicado en juegos y conducción autónoma.

Estos modelos se aplican en diversos campos de la IA, como el Procesamiento de Lenguaje Natural (NLP), que permite a las máquinas interpretar y generar lenguaje humano; las Redes Neuronales, algoritmos inspirados en el cerebro humano para procesar información; y el *Deep Learning*, redes neuronales de múltiples capas que requieren significativa capacidad de cómputo, pero que proporcionan un análisis más profundo.

### 3.2. APLICACIONES DE LA IA Y EL ML

Ahora bien, en relación al modo en que estas herramientas pueden contribuir a la consecución de un futuro más medioambientalmente sostenible, cabe nombrar las siguientes aplicaciones:

- **Optimización del uso de recursos:** podemos analizar mediante algoritmos patrones de consumo de recursos naturales (agua, energía, etc.) para optimizar su uso y reducir el desperdicio. Esto incluye la identificación de eficiencias en sistemas de producción y distribución de energía, así como en la gestión de recursos hídricos.
- **Monitoreo ambiental:** los modelos de *deep learning* y redes neuronales permiten el análisis de grandes volúmenes de datos provenientes de satélites y sensores terrestres para monitorear la deforestación, la contaminación del aire y del agua, y otros indicadores ambientales críticos. Esto facilita la detección temprana de problemas ambientales y la toma de decisiones informadas para su mitigación.

- **Agricultura sostenible:** el ML puede ayudar en el desarrollo de sistemas de agricultura de precisión que optimicen el uso de agua y fertilizantes, minimizando el impacto en el suelo y aumentando incluso los rendimientos de los cultivos, reduciendo así la necesidad de deforestación y contribuyendo a la seguridad alimentaria.
- **Gestión de residuos:** se pueden desarrollar sistemas inteligentes para la gestión y reciclaje de residuos, identificando los materiales reciclables y optimizando los procesos de selección y tratamiento de desechos. Asimismo, puede ayudar a implementar sistemas de economía circular, donde los productos al final de su vida útil se reutilicen, reciclen o remanufacturen.
- **Predicción de fenómenos meteorológicos extremos:** ahora podemos analizar datos históricos y actuales para predecir con una mayor precisión, permitiendo una mejor preparación y mitigación de sus efectos sobre las comunidades y los ecosistemas.
- **Conservación de la biodiversidad:** servimos de modelos de ML para identificar patrones y contribuir en la conservación de especies en peligro y en la gestión de áreas protegidas.
- **Eficiencia energética en edificaciones:** el uso de algoritmos de aprendizaje supervisado puede optimizar el consumo de energía en edificios, ajustando automáticamente la iluminación, la calefacción y el aire acondicionado, lo que contribuye a la reducción de la huella de carbono.
- **Optimización en la logística:** la IA puede optimizar las rutas de diversos medios de transporte para reducir sus emisiones y el consumo de combustible.
- **Desarrollo de nuevos materiales:** la IA puede acelerar el descubrimiento y el diseño de nuevos materiales y procesos químicos que sean más respetuosos con el medio ambiente, como plásticos biodegradables o procesos industriales que emitan menos CO<sub>2</sub>.

Esta integración de IA y ML no solo busca mejorar la productividad y la eficiencia, sino también facilitar la vida humana, asumiendo responsabilidades y tareas, lo que a su vez plantea desafíos y oportunidades para el futuro desarrollo tecnológico y aplicaciones innovadoras (Alvarado, 2015).



## **4. CASO DE ESTUDIO: LA INDUSTRIA ALIMENTARIA**

### **4.1. INTRODUCCIÓN. HUELLA ECOLÓGICA DE LO QUE COMEMOS**

La huella ecológica se ha convertido en un asunto de gran importancia en la discusión sobre la sostenibilidad y el impacto ambiental de nuestras decisiones diarias, especialmente en lo que respecta a lo que comemos. El sistema agroalimentario global, con su compleja red de producción, procesamiento, distribución y consumo de alimentos, juega un papel crucial en la configuración de esta huella, afectando de manera significativa a los recursos naturales y a nuestros ecosistemas.

Y es que, de todos los flujos que componen el metabolismo de las ciudades (energía, materiales, agua, biomasa, residuos, etc.), los relacionados con la alimentación (biomasa) tienen una gran importancia tanto cuantitativa como cualitativa. No sólo porque, en términos de Análisis de Ciclo de Vida (ACV), vienen representando anualmente entre el 20 y el 30% del impacto ambiental relacionado con todo el consumo público y privado (Tukker & Jansen, 2006), sino porque su gestión y organización da lugar a un importante despilfarro en términos de residuos alimentarios no aprovechados (Stuart, 2011) (FAO, 2011).

Investigaciones sobre el impacto de diferentes modelos alimentarios en el medio ambiente han demostrado que las dietas ricas en carne requieren aproximadamente tres veces más territorio cultivado que las dietas vegetarianas, subrayando la necesidad de repensar nuestros patrones de consumo alimentario. Cohen (1995) y Goodland (1997) destacan las implicaciones de estas dietas en términos de requerimientos de tierra, energía y recursos, sugiriendo que una transición hacia dietas basadas en plantas es crucial para la sostenibilidad ambiental. Explicaban también que los requisitos para una correcta dieta vegetariana eran de aproximadamente 2.500 kcal/día, pero que si quisiéramos obtener el 30% de esa energía a través del ganado esto supondría elevar, directa e indirectamente, nuestra ingesta hasta las 9.250 kcal/día, lo que, por ejemplificar, supone 3,7 veces más calorías de las que se obtuvieron como media en la década de 1990 con la cosecha mundial (Cohen, 1995). Por otro lado, las desigualdades en el impacto ambiental exceden a aquellas ocasionadas por el mero consumo de energía interna. Esto es algo que se expresa

en la alta proporción representada por la “huella animal” en el total de los países ricos, llegando a valores cercanos al 50-60%, lo cual casi duplica el promedio global del 35% y triplica la correspondiente huella de los habitantes de África y Asia. Este tipo de huella se refiere al impacto ambiental generado por la producción y consumo de productos de origen animal, incluyendo la cría de ganado, el uso de recursos como agua y tierra destinados a esa cría, y las emisiones de gases de efecto invernadero asociadas. En el caso de España, por ejemplo, desde el inicio del siglo XXI, esa proporción de la “huella alimentaria animal” se sitúa cercana al 50% (Carpintero, 2006).

Recientes estudios han evaluado los impactos de distintas dietas en términos de uso de recursos, emisiones de gases de efecto invernadero y requerimientos territoriales, concluyendo en la importancia de transitar hacia dietas con menor consumo de carne y mayor consumo de productos vegetales por motivos ya no solo ecológicos, sino también de salud humana y animal. Además, en muchos países, especialmente los más desarrollados, se observa una marcada disparidad entre los requerimientos energéticos estables de la población, que se han mantenido constantes desde tiempos prehistóricos, y el incremento significativo del gasto alimenticio. Este contraste no solo refleja una divergencia entre las necesidades fisiológicas y el gasto económico, sino también entre dichas necesidades y el impacto ambiental de su satisfacción. Los cambios en la dieta, lejos de aumentar el consumo físico de alimentos, ejercen una presión creciente sobre las áreas productivas, afectando tanto al país de origen como a otros. Esta situación invita a una reflexión sobre el impacto sociocultural del consumo alimentario y cómo este se ve influenciado por el estilo de vida urbano y la estructura social y laboral contemporáneas (Gerbens-Leenes, Nonhebel & Ivens, 2002).

En última instancia, el fomento de la agricultura urbana y periurbana aparece como una estrategia prometedora para reducir la dependencia de las ciudades de recursos externos, apoyando la soberanía alimentaria y disminuyendo la huella ecológica de lo que comemos. Iniciativas como el “Pacto de Milán sobre Política Alimentaria Urbana” han insistido en los beneficios de incrementar la producción local de alimentos, tanto para el medio ambiente como para la economía y la salud de las comunidades locales (Fernández & Morán, 2015).

## 4.2. LA RESPONSABILIDAD SOCIAL CORPORATIVA EN LA INDUSTRIA ALIMENTARIA

La Responsabilidad Social Corporativa (RSC) en la industria alimentaria de España desempeña un papel crucial, reflejando su profunda conexión con la salud de los consumidores, la sostenibilidad y el desarrollo económico. La industria alimentaria, que representa el 22,8% del sector industrial total y el 18,9% del valor añadido, se compone principalmente de PYMES, marcando una significativa contribución al tejido económico del país (MAPA, 2021). A pesar de los desafíos presentados por la crisis económica de 2008 y otras crisis globales (como pudo suponer la reciente pandemia del COVID-19), este sector ha demostrado resiliencia, adaptándose a las fluctuaciones de demanda y manteniendo una posición estable en términos de empleo y facturación (Claver-Cortés et al., 2018).

El análisis DAFO de la industria alimentaria española revela varios aspectos cruciales. Las fortalezas incluyen su capacidad de generación de empleo y el significativo peso en la economía nacional. Sin embargo, enfrenta debilidades como la falta de confianza de los consumidores respecto a la implementación de RSC y la escasez de recursos económicos para afrontar cambios necesarios (Escolán & Mejuto, 2018). Las oportunidades emergen en la modernización de técnicas productivas y la aparición de nuevos nichos de mercado, mientras que las amenazas se ven relacionadas en mayor medida con el cambio climático.

La RSC en la industria alimentaria abarca la necesidad de atender y responder de manera incluso más responsable que otros sectores a las necesidades de la sociedad, considerando la directa repercusión de la producción alimentaria en la salud de los consumidores (OIT, 2007). Casos de crisis alimentarias, como el brote de listeria en 2019 y la enfermedad de las "vacas locas", subrayan la importancia de mantener altos estándares de seguridad y responsabilidad (Ballester et al., 2015). Para asegurar una correcta higiene alimentaria, es esencial que los alimentos conserven sus propiedades organolépticas (aroma, textura, sabor, color) y su inocuidad alimentaria, además de garantizar la higiene personal de quienes los manipulan.

Por otro lado, la creciente preocupación por la obesidad y la salud pública insta a la industria a asumir un rol más activo en la promoción de un consumo responsable. La contradicción entre comercializar productos poco saludables y al mismo tiempo promover una dieta equilibrada destaca la complejidad de la RSC en este sector. Además, el interés del consumidor por el origen y el impacto ambiental de los alimentos demanda a las empresas una transparencia y responsabilidad adicionales, situando a la industria alimentaria española a la vanguardia en términos de RSC (OCU, 2020). Sin embargo, el desafío del *greenwashing*, donde las empresas adoptan prácticas aparentemente sostenibles solo para mejorar su imagen, representa un obstáculo significativo para la autenticidad de las iniciativas de RSC (Pearse, 2014) (Follos, 2014).

Para que la RSC sea efectiva y auténtica, es crucial que las empresas vayan más allá del cumplimiento normativo y contribuyan de manera voluntaria a la sostenibilidad y al bienestar social. Esto incluye iniciativas como la reducción del desperdicio alimentario, la implementación de prácticas agrícolas sostenibles, y el apoyo a comunidades locales. En este sentido, la RSC no solo se trata de cumplir con las leyes, sino de adoptar prácticas que generen un impacto positivo y duradero en la sociedad y el medio ambiente.

En resumen, la RSC en la industria alimentaria de España es fundamental para el desarrollo sostenible y el bienestar social. Las empresas enfrentan el reto de integrar prácticas responsables que vayan más allá del cumplimiento normativo, buscando generar un impacto positivo en la sociedad y el medio ambiente. La implementación efectiva de la RSC puede no solo mitigar los riesgos y desafíos actuales, sino también abrir nuevas oportunidades de negocio, mejorando la competitividad y fortaleciendo la reputación corporativa en un mercado cada vez más consciente de los valores éticos y sostenibles.

#### 4.3. LA SOSTENIBILIDAD COMO OPORTUNIDAD ESTRATÉGICA

La sostenibilidad ha surgido como una oportunidad estratégica clave para las empresas, especialmente en la industria alimentaria, donde los desafíos son diversos y complejos. A medida que el cambio climático se acelera y la población mundial crece, la

industria alimentaria se encuentra bajo una presión sin precedentes para adaptarse y transformarse.

En 2016, el Banco Mundial estimó que esta industria es responsable de un tercio de las emisiones de gases de efecto invernadero del planeta, el 80% de la pérdida de biodiversidad y la deforestación y el 70% del consumo de agua dulce. Asimismo, destacó la urgente necesidad de prácticas sostenibles (World Bank Group, 2022). La vulnerabilidad de la agricultura al cambio climático y la inseguridad alimentaria en ciertas regiones del mundo pone de relieve la importancia de la sostenibilidad. Organizaciones como la FAO han identificado desafíos clave para la industria, incluido el aumento de la productividad agrícola sostenible, la garantía de una base de recursos naturales sostenible y la prevención de plagas y enfermedades transfronterizas. Estos esfuerzos son consistentes con los Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030 y resaltan la importancia de un enfoque integrado para abordar estos temas (FAO, 2017).

Empresas del sector de la alimentación como Nestlé y Danone han demostrado que integrar la sostenibilidad en la estrategia empresarial puede crear valor tanto para las empresas como para la sociedad. Nestlé, por ejemplo, se ha comprometido a lograr cero emisiones netas para 2050, lo que incluirá una revisión exhaustiva de sus operaciones, productos y cadena de suministro. Este enfoque no solo considera el impacto ambiental de una empresa, sino que también fomenta la innovación y la colaboración con proveedores para promover prácticas agrícolas sostenibles (Nestlé, 2021). Mientras tanto, Danone, a través de una empresa conjunta con Grameen Bank, lanzó Grameen Danone para luchar contra la desnutrición en Bangladesh y desarrollar productos nutritivos accesibles a los segmentos más pobres de la población. Esta iniciativa se centra en los aspectos sociales de la sostenibilidad y en la erradicación de la pobreza y del hambre, objetivos que están directamente relacionados con los ODS.

Por poner otro ejemplo, nos encontramos con casos de empresas como Ecoalf, que, a pesar de pertenecer al sector textil, puede servir como modelo para cualquier compañía de cualquier otro sector. Ecoalf basa su modelo de negocio en la utilización de recursos

reciclados para la fabricación de sus productos, demostrando cómo la integración de prácticas sostenibles puede también ser un motor de innovación, de creación de valor y de implantación de tendencias de moda. Este enfoque no solo responde a una necesidad ambiental, sino que también se alinea con las crecientes demandas de los consumidores por productos responsables y éticos (Ecoalf, s.f.).

En definitiva, todas estas prácticas no sólo mejoran la sostenibilidad ambiental y social, sino que también brindan oportunidades estratégicas para las empresas. La sostenibilidad puede diferenciar productos en el mercado, abrir nuevas áreas de negocio y crear ventajas competitivas. Sin embargo, persisten desafíos, particularmente en torno a los costos y la necesidad de un cambio sistémico en la industria, de tal manera que una transición hacia prácticas más sostenibles requiere la colaboración entre empresas, gobiernos y sociedad civil. La adopción de prácticas sostenibles no sólo es importante para abordar los desafíos ambientales y sociales actuales, sino que también proporciona beneficios económicos a largo plazo al fomentar la innovación, aumentar la eficiencia y fortalecer la reputación de una empresa. Para lograr una industria alimentaria verdaderamente sostenible a medio y largo plazo, la cooperación entre varios actores de la cadena de valor y el compromiso con los ODS resultan esenciales.

## 5. APLICACIÓN PRÁCTICA DE TÉCNICAS DE MACHINE LEARNING AL CASO DE ESTUDIO

### 5.1. EXPLORACIÓN DE LA BASE DE DATOS Y *DATA CLEANING*

La metodología de la investigación comienza con la recopilación de datos, para lo cual vamos a servirnos de *Our World in Data*, una reconocida y confiable fuente de bases de datos que ofrece información rigurosamente recopilada y verificada sobre diversos aspectos del desarrollo humano y ambiental. En concreto, el *dataset* que extraemos contiene información detallada sobre el uso de recursos y emisiones de gases de efecto invernadero (GEI) asociadas a la producción de distintos alimentos (Ritchie, Rosado & Roser, 2022). Se compone de 43 registros, que incluyen los productos alimenticios más comunes y que han sido analizados (tipos de cereales, carnes, frutas, aceites, derivados lácteos, pescados, entre otros), y ocho variables, siete de las cuales se refieren a las distintas etapas de la cadena de producción del alimento y una última que engloba el total de las emisiones. Para cada alimento se muestra la huella de carbono asociada a su producción, tomando como medida los kilogramos de CO<sub>2</sub>eq producidos en cada etapa por cada kilogramo de alimento.

Para la exploración de la base de datos se utilizará el lenguaje de programación Python. En primer lugar, importamos el CSV y las librerías *math*, *matplotlib*, *matplotlib.pyplot*, *numpy*, *pandas* y *seaborn*, que nos ayudarán a una interpretación inicial más sencilla y accesible por vía de distintas visualizaciones gráficas. Como primera aproximación, elaboramos un análisis previo mostrando la información general del *dataframe*, las estadísticas descriptivas (recuento, media, mediana, desviación estándar, etc.), el número de filas y columnas y pedimos que nos saque cualquier valor nulo (ver Anexo, páginas 1-3). En este caso, nuestra base de datos no muestra ningún NaN, por lo que no será necesario aplicar técnicas de imputación de valores desconocidos.

En la Figura 1 aparece una muestra de nuestro *dataset*, incluyendo únicamente los diez primeros alimentos.

	Food	Land use change (kgCO <sub>2</sub> eq per kg)	Farm (kgCO <sub>2</sub> eq per kg)	Animal feed (kgCO <sub>2</sub> eq per kg)	Processing (kgCO <sub>2</sub> eq per kg)	Transport (kgCO <sub>2</sub> eq per kg)	Retail (kgCO <sub>2</sub> eq per kg)	Packaging (kgCO <sub>2</sub> eq per kg)	Total emissions (kgCO <sub>2</sub> eq per kg)
0	Apples	-0.028951	0.225816	0.000000	0.003820	0.095804	0.016578	0.044526	0.357593
1	Bananas	-0.025534	0.269618	0.000000	0.060146	0.295830	0.020981	0.065670	0.686712
2	Barley	0.008676	0.176377	0.000000	0.127703	0.035351	0.263587	0.496786	1.108480
3	Beef (beef herd)	23.237536	56.228063	2.680979	1.811083	0.494125	0.233538	0.352084	85.037407
4	Beef (dairy herd)	1.266022	21.915529	3.503673	1.547664	0.592410	0.254032	0.374549	29.453881
5	Beet Sugar	0.000052	0.535721	0.000000	0.241885	0.628544	0.039322	0.090289	1.535814
6	Berries & Grapes	0.024601	0.717848	0.000000	0.000000	0.236392	0.017057	0.211092	1.206990
7	Brassicas	0.002242	0.277745	0.000000	0.000000	0.094641	0.016844	0.045303	0.436776
8	Cane Sugar	1.263041	0.491270	0.000000	0.037381	0.794510	0.036723	0.084277	2.707202
9	Cassava	0.589436	0.219994	0.000000	0.000000	0.093597	0.039322	0.044822	0.987171

Figura 1: Muestra del conjunto de datos

Las variables sobre las que vamos a sostener nuestro estudio son las siguientes:

- Food: listado de alimentos analizados, compuesto por un total de 43. El tipo de dato es *object*, que en este caso viene exclusivamente contenido por cadenas de texto.
- Land use change (kgCO<sub>2</sub>eq per kg): emisiones de GEI producidas en la etapa de cambio de uso del suelo por cada kg de alimento, medidas en kg de dióxido de carbono equivalente (CO<sub>2</sub>eq).
- Farm (kgCO<sub>2</sub>eq per kg): emisiones de GEI producidas en la etapa de cultivación por cada kg de alimento, medidas en kg CO<sub>2</sub>eq.
- Animal feed (kgCO<sub>2</sub>eq per kg): emisiones de GEI producidas en la etapa de alimentación animal por cada kg de alimento, medidas en kg CO<sub>2</sub>eq.
- Processing (kgCO<sub>2</sub>eq per kg): emisiones de GEI producidas en la etapa de procesamiento por cada kg de alimento, medidas en kg CO<sub>2</sub>eq.
- Transport (kgCO<sub>2</sub>eq per kg): emisiones de GEI producidas en la etapa de transportación por cada kg de alimento, medidas en kg CO<sub>2</sub>eq.
- Retail (kgCO<sub>2</sub>eq per kg): emisiones de GEI producidas en la etapa de comercialización por cada kg de alimento, medidas en kg CO<sub>2</sub>eq.
- Packaging (kgCO<sub>2</sub>eq per kg): emisiones de GEI producidas en la etapa de empaquetado por cada kg de alimento, medidas en kg CO<sub>2</sub>eq.



- Total emissions (kgCO<sub>2</sub>eq per kg): suma de todas las emisiones de GEI producidas por cada kg de alimento a lo largo de toda la cadena de producción, medidas en kg CO<sub>2</sub>eq.

Como podemos observar, todos los datos de nuestra base de estudio son de tipo *float64*, lo que significa que almacenan valores numéricos de coma flotante, y se miden por CO<sub>2</sub>eq, que es la unidad de medida empleada para calcular la huella de carbono y que engloba la emisión de todos los gases considerados de efecto invernadero: dióxido de carbono (CO<sub>2</sub>), metano (CH<sub>4</sub>), óxido de nitrógeno (N<sub>2</sub>O), hidrofluorocarburos (HFC), perfluorocarburos (PFC) y hexafluoruro de azufre (SF<sub>6</sub>).

A continuación, como parte del proceso de limpieza del CSV o *Data Cleaning*, tratamos también de simplificar al máximo la información (acortando títulos, eliminando espacios entre palabras, etc.) y establecemos la columna del tipo de alimento ("*Food*") como índice, lo que nos facilitará la búsqueda y el acceso a los datos basados en los nombres de los productos.

En total, nuestro *dataset* contiene información de 43 tipos de productos alimenticios. Para seguir facilitando el posterior procesamiento de los datos, vamos a realizar un trabajo de *clustering* para agrupar los productos por distintas categorías, de tal manera que también consigamos reducir el ruido de cara a visualizar estos datos. La clasificación consta de ocho grupos de alimentos, estos son: cereales, frutos secos, verduras, frutas, azúcares y derivados, aceites, lácteos y derivados, productos de origen animal, y una categoría residual donde incluimos productos más difícilmente clasificables en nuestra base de datos, como el café y el chocolate, entre otros. La siguiente figura muestra el listado global de los alimentos estudiados y la correspondiente categoría asignada.

	<b>Food</b>	<b>Category</b>
0	Apples	Fruta
1	Bananas	Fruta
2	Barley	Cereal
3	Beef (beef herd)	Animal
4	Beef (dairy herd)	Animal
5	Beet Sugar	Azúcar
6	Berries & Grapes	Fruta
7	Brassicas	Verdura
8	Cane Sugar	Azúcar
9	Cassava	Verdura
10	Cheese	Lácteos
11	Citrus Fruit	Fruta
12	Coffee	Otros
13	Dark Chocolate	Otros
14	Eggs	Animal
15	Fish (farmed)	Animal
16	Groundnuts	Frutos_Secos
17	Lamb & Mutton	Animal
18	Maize	Cereal
19	Milk	Lácteos
20	Nuts	Frutos_Secos
21	Oatmeal	Cereal
22	Olive Oil	Aceite
23	Onions & Leeks	Verdura
24	Other Fruit	Fruta
25	Other Pulses	Verdura
26	Other Vegetables	Verdura
27	Palm Oil	Aceite
28	Peas	Verdura
29	Pig Meat	Animal
30	Potatoes	Verdura
31	Poultry Meat	Animal
32	Rapeseed Oil	Aceite
33	Rice	Cereal
34	Root Vegetables	Verdura
35	Shrimps (farmed)	Animal
36	Soy milk	Lácteos
37	Soybean Oil	Aceite
38	Sunflower Oil	Aceite
39	Tofu	Otros
40	Tomatoes	Verdura
41	Wheat & Rye	Cereal
42	Wine	Otros

Figura 2: Clasificación de los alimentos por categorías

## 5.2. ANÁLISIS DESCRIPTIVO E INTERPRETACIÓN DE RESULTADOS

La primera visualización que elaboramos es un *scatter plot* o gráfico de dispersión (Figura 3) que nos muestra la relación de emisiones totales que conlleva la producción de cada alimento, es decir, a lo largo de toda su cadena de producción y de suministro, categorizándolo por los respectivos grupos de alimentos.

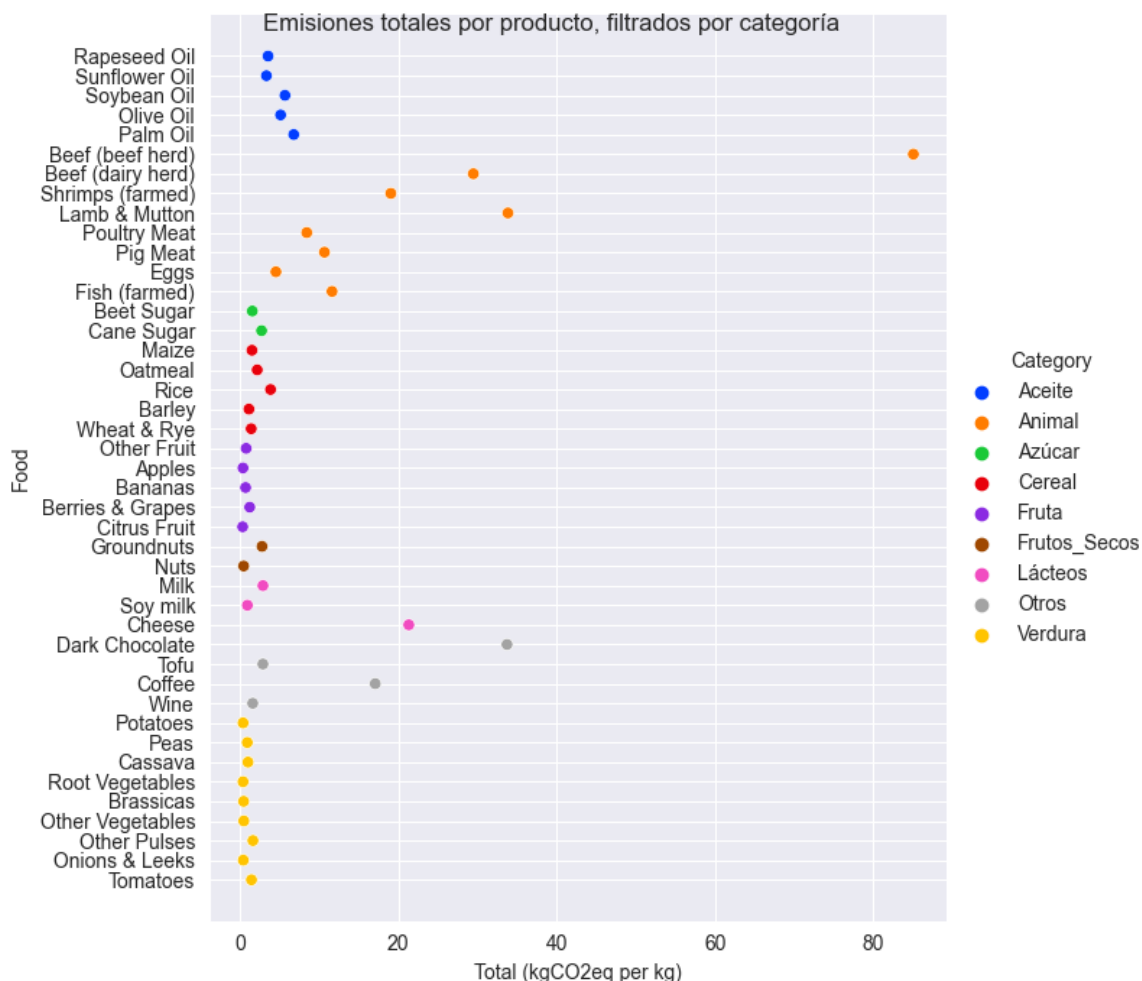
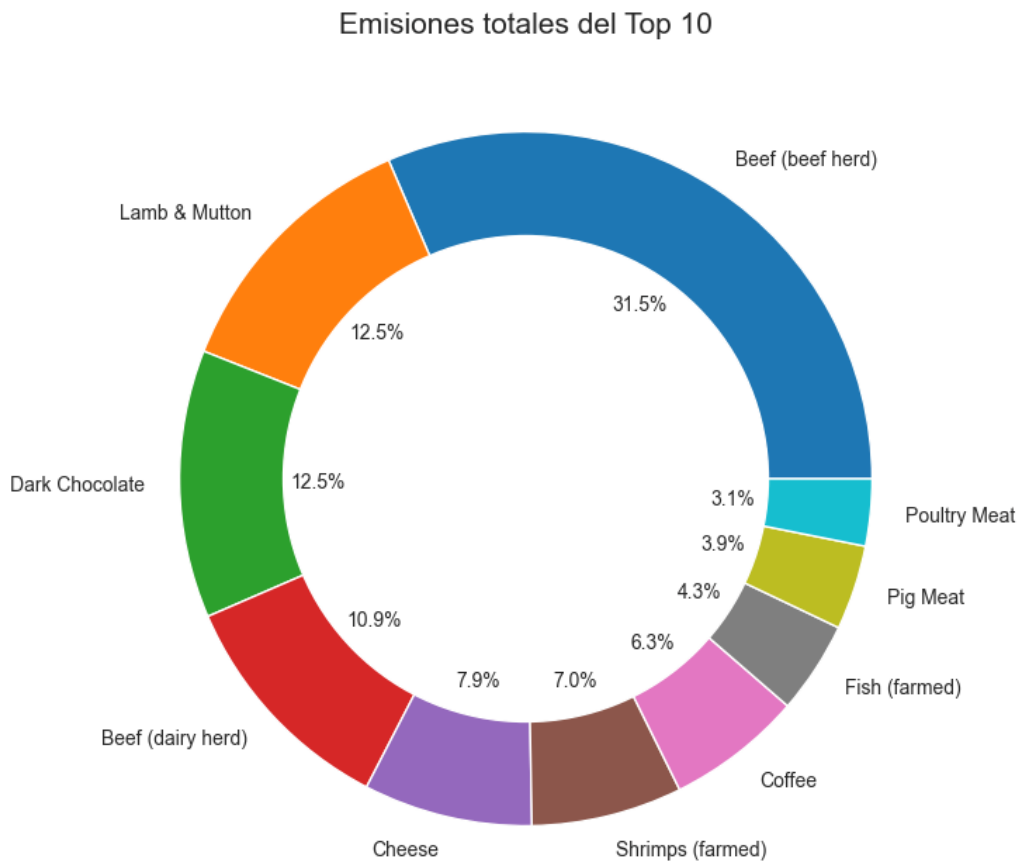


Figura 3: Emisiones totales por producto, filtrados por categoría

Como podemos observar, los productos de origen animal son comprensiblemente más erráticos, mientras que las frutas, verduras y, de hecho, la mayoría de las demás categorías, son relativamente estables en lo que a volatilidad se refiere. Es también inevitable percatarse de la presencia de determinados *outliers*, como es el caso de la carne de ganado vacuno (“Beef herd”), cuya mayor contribución a la huella de carbono se encuentra, sorprendentemente, en las fases de cultivación y cambio de uso del suelo. Por otro lado, la categoría que hemos denominado “Otros” y que incluye el café, el vino, el

tofu y el chocolate negro, también presenta un mayor nivel de variabilidad debido a que la mayoría de ellos son productos procesados y no crudos y, por tanto, se espera que presenten mayores fluctuaciones en las mediciones. En síntesis, parece que la agrupación por categorías nos ha ayudado a obtener información muy necesaria en esta primera tarea.

A continuación (Figura 4), se presentan las emisiones totales del Top 10 de productos más “nocivos” para el medio ambiente de entre todos los analizados. Como podemos observar, el impacto de este pequeño grupo representa algo más del 80% sobre la suma global de emisiones. Siendo el global de 334,94 kg de CO<sub>2</sub>eq, el conjunto de estos 10 alcanza ya los 270,08 kg de CO<sub>2</sub>eq.



Suma global de emisiones: 334.94 kgCO<sub>2</sub>eq per kg  
 Suma de emisiones de los 10 principales productos: 270.08 kgCO<sub>2</sub>eq per kg  
 Porcentaje del Top 10 sobre el global: 80.6%

*Figura 4: Emisiones totales del Top 10 de alimentos*

La Figura 5 representa las emisiones totales de CO<sub>2</sub>eq agrupadas directamente por cada kg producido de cada categoría y divididas según las diversas etapas de la cadena de suministro. Cada barra horizontal representa un grupo de alimentos, y los diferentes colores dentro de las barras indican las emisiones correspondientes a cada etapa de la cadena.

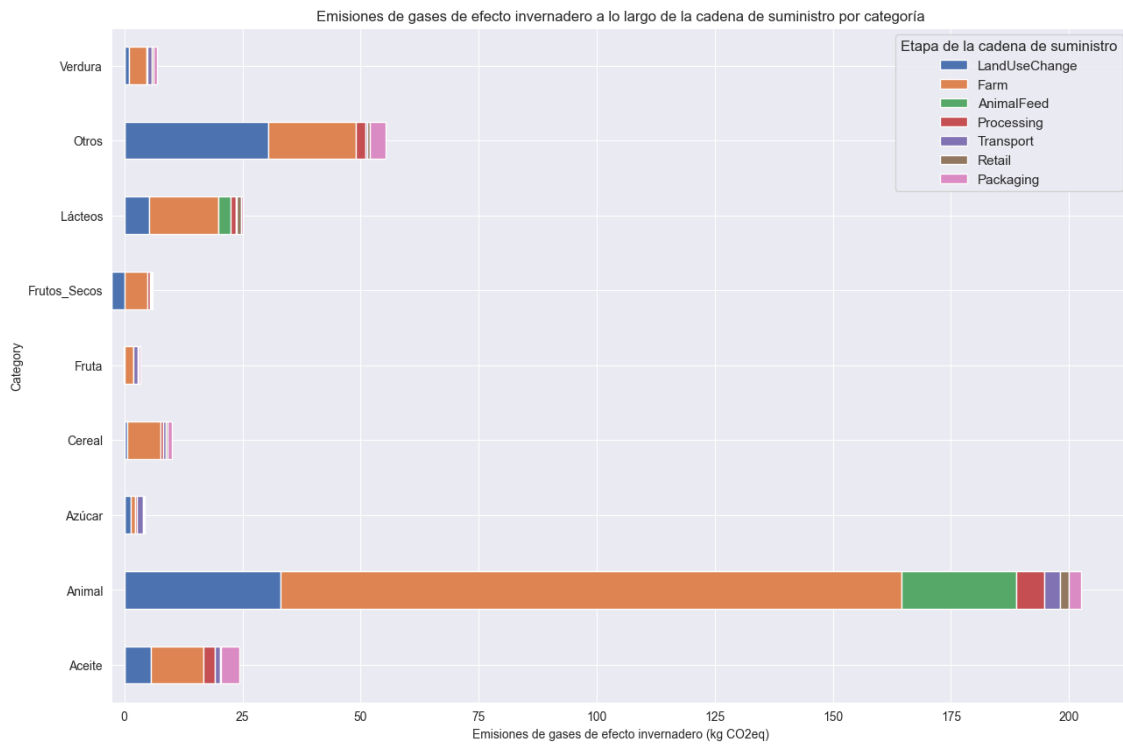


Figura 5: Emisiones totales por kg de categoría a lo largo de la cadena de suministro

Como principales observaciones, podemos verificar de nuevo que la categoría animal destaca significativamente por sus altas emisiones GEI por cada kg de alimento producido en comparación con el resto. La mayor parte de estas emisiones proviene de la producción agrícola que lleva asociado cualquier alimento de origen animal (*Farm*) y el cambio de uso de la tierra (*LandUseChange*). La producción de lácteos, aceites y otros (café, vino, chocolate...) también presentan la mayor parte de sus emisiones en esas etapas. En contrapartida, vemos que los frutos secos, la fruta, la verdura, el azúcar y el cereal presentan menores emisiones en general, con una distribución relativamente equilibrada a lo largo de las diferentes etapas de la cadena de suministro, si bien la fase de cultivo o producción agrícola obtiene un peso importante en todas ellas.

El análisis de este gráfico sugiere que las estrategias de mitigación de GEI en la industria alimentaria deben centrarse particularmente en las líneas de producción de alimentos de origen animal. La implementación de prácticas agrícolas sostenibles, el fomento de la ganadería extensiva y la pesca artesanal y la optimización de procesos de transporte podrían contribuir significativamente a la reducción de estas emisiones totales. No en vano, como ya se ha expuesto con anterioridad en este trabajo, promover dietas basadas en plantas y la reducción del consumo de productos cárnicos y pescados y sus derivados podría tener también un impacto muy considerable en la reducción de la huella de carbono del sector alimentario. Sirva así este último gráfico como herramienta visual para identificar las áreas críticas donde se deben enfocar los esfuerzos de sostenibilidad y políticas ambientales.

### 5.3. ANÁLISIS DE CORRELACIÓN

El análisis de correlación es una herramienta fundamental para entender las relaciones lineales entre diferentes variables dentro de un conjunto de datos. Esta técnica nos permite identificar y cuantificar la fuerza y dirección de las relaciones entre dos o más variables, lo cual es esencial tanto para la interpretación como la predicción en diversos contextos.

En este caso, hemos analizado las correlaciones entre las etapas de la cadena de suministro de alimentos y las emisiones totales de gases de efecto invernadero. La matriz de correlación generada nos proporciona una visión clara de cómo cada etapa de la cadena de suministro se relaciona con las demás y con las emisiones totales, entendiendo cómo cada una de ellas puede contribuir de manera diferente a la huella de carbono total.

Para su interpretación general debemos conocer que cualquier valor cercano a 1 indica una fuerte correlación positiva, es decir, a medida que una variable aumenta, la otra también tiende a aumentar. Los valores cercanos a 0 indican poca o ninguna correlación lineal entre las variables, y los que se aproximan a -1 indican una fuerte correlación negativa, es decir, que a medida que una variable aumenta, la otra tiende a disminuir.

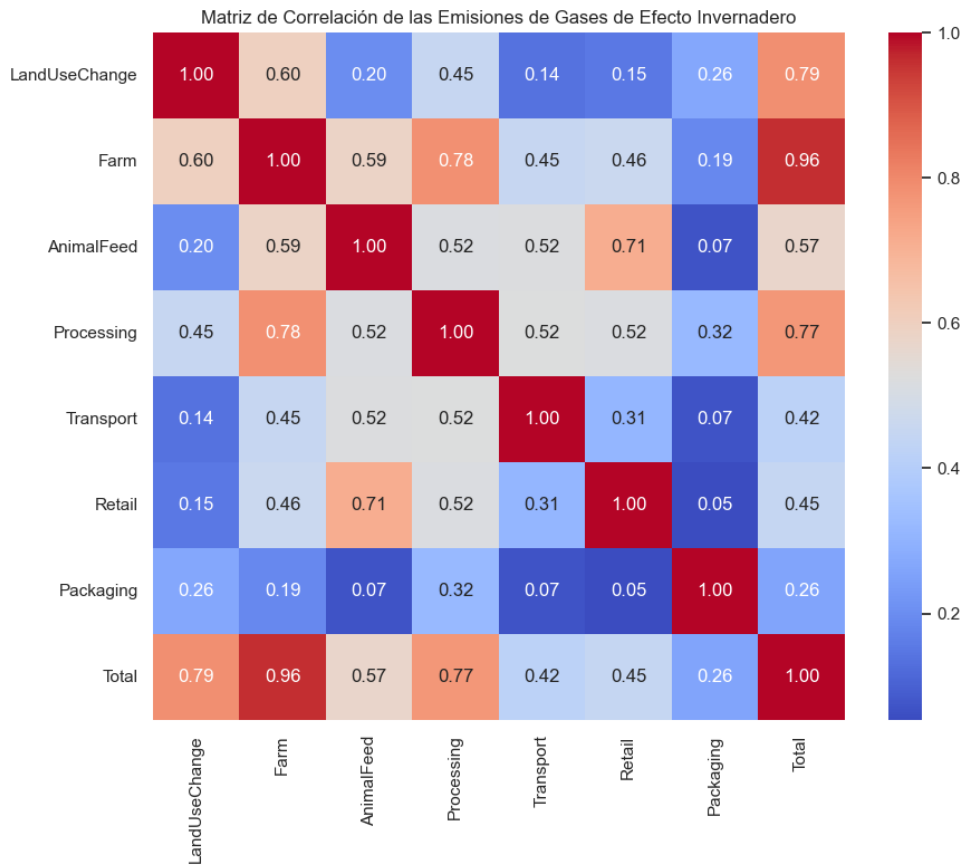


Figura 6: Matriz de Correlación de las emisiones GEI

La Figura 6 muestra que “Farm” tiene una correlación extremadamente alta con “Total”, alcanzando 0,96. Esto destaca la importancia de la etapa de cultivo como el principal contribuyente a las emisiones totales. Además, los cambios en el uso de la tierra y la etapa de procesamiento muestran también una fuerte correlación de 0,79 y 0,77 respectivamente con respecto a las emisiones totales.

En cuanto a correlaciones moderadas, “AnimalFeed” presenta una correlación de 0,57 con “Total”, lo que sugiere una interdependencia meramente considerable entre estas variables. “Processing”, por ejemplo, mantiene igualmente correlaciones moderadas con varias otras variables, como “Transport” (0,52) y “Retail” (0,52), indicando una conexión directa entre el procesamiento de determinados productos y las posteriores etapas de transporte y comercialización.

Finalmente, las correlaciones débiles se observan en variables como “Transport”, con un valor de 0,14 con respecto a “LandUseChange”, y “Packaging”, con una correlación de 0,26 con “Total”. Estos datos sugieren que, aunque estas etapas son importantes, su impacto en el total de emisiones es generalmente menos significativo en comparación con otras.

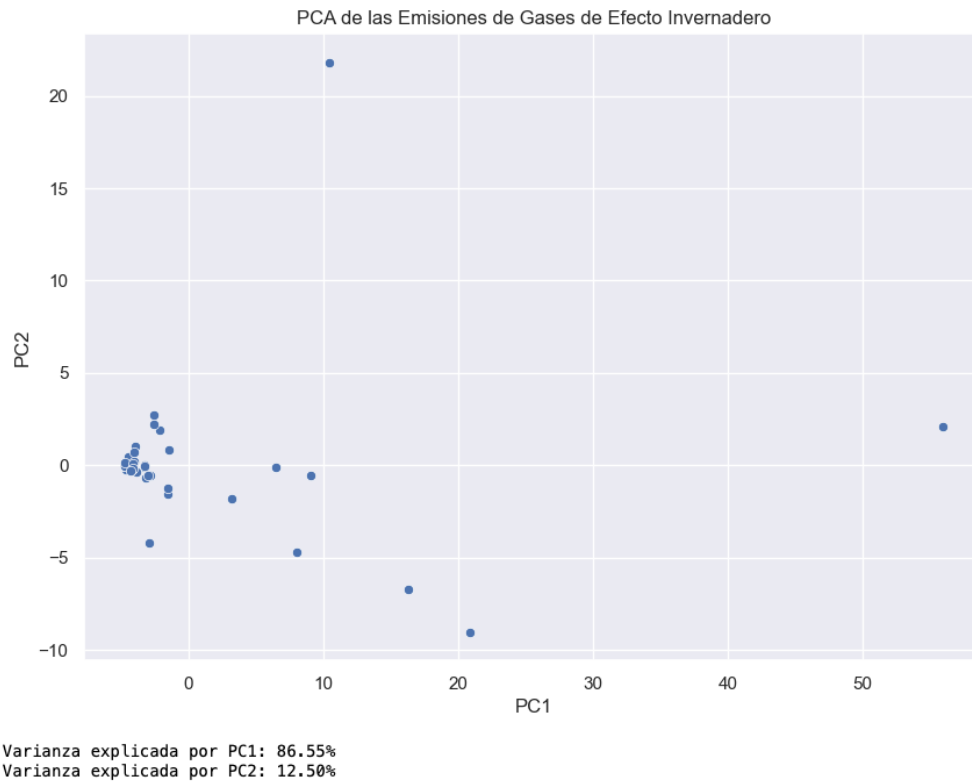
En conclusión, este análisis de correlación proporciona una visión clara de las relaciones entre las diferentes etapas de la cadena de suministro y sus contribuciones a las emisiones totales de gases de efecto invernadero. La etapa de cultivo muestra ser el mayor contribuyente a las emisiones totales, seguido por el cambio en los usos del suelo y el procesamiento, lo que implica que cualquier estrategia de reducción de emisiones deberá enfocarse principalmente en estas etapas para lograr un impacto significativo. Las etapas de transporte y comercio, aunque menos influyentes en las emisiones totales, siguen siendo importantes áreas para la mejora continua. Este análisis subraya la importancia de una evaluación holística de todas las etapas de la cadena de suministro para desarrollar políticas y estrategias efectivas en la reducción de emisiones GEI.

#### 5.4. ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

El análisis de Componentes Principales o PCA (por sus siglas en inglés) es una técnica de ML que permite reducir la dimensionalidad y eliminar redundancias de los datos y encontrar combinaciones lineales de las variables que expliquen la mayor parte de la varianza. Esto nos puede ayudar a identificar patrones y relaciones entre las variables que no sean inmediatamente obvios. PCA transforma los datos originales en un conjunto de nuevas variables, llamadas componentes principales, que son combinaciones lineales de las variables originales.

Para nuestro *dataset*, vamos a aplicar PCA reduciendo a dos componentes principales y crearemos un *dataframe* con ellos. La Figura 7 recoge el resultado obtenido.





*Figura 7: PCA de las emisiones GEI*

Con el gráfico de PCA y la varianza explicada por los componentes principales podemos sacar varias conclusiones importantes, como que el primer componente principal (PC1) explica el 86,55% de la varianza en los datos, mientras que PC2 explica el 12,50%. Juntas, las dos componentes principales explican el 99,05% de la varianza total en los datos, lo que implica que la mayoría de la información contenida en las variables originales puede ser representada en solo dos dimensiones con una pérdida mínima de información.

En cuanto al gráfico, observamos que la mayoría de los puntos se agrupan cerca del origen en el espacio de componentes principales (PC1 y PC2), pero hay algunos puntos que se desvían significativamente del resto, lo que sugiere la presencia de *outliers* o alimentos con características muy distintas y que presentan emisiones significativamente más altas o bajas en alguna etapa específica de la cadena de suministro. La agrupación cerca del origen sugiere que muchos alimentos tienen perfiles de emisión de gases de efecto invernadero similares.

Ahora bien, dado que PC1 explica la mayor parte de la varianza, cualquier análisis adicional debería centrarse en entender qué variables contribuyen más a PC1. Esto lo vamos a llevar a cabo mediante un examen de los coeficientes de carga o *loadings* de las variables originales en el componente principal. Aunque PC2 no explica tanta varianza como PC1, aún es importante y puede revelar variaciones secundarias en los datos que no son capturados por PC1.

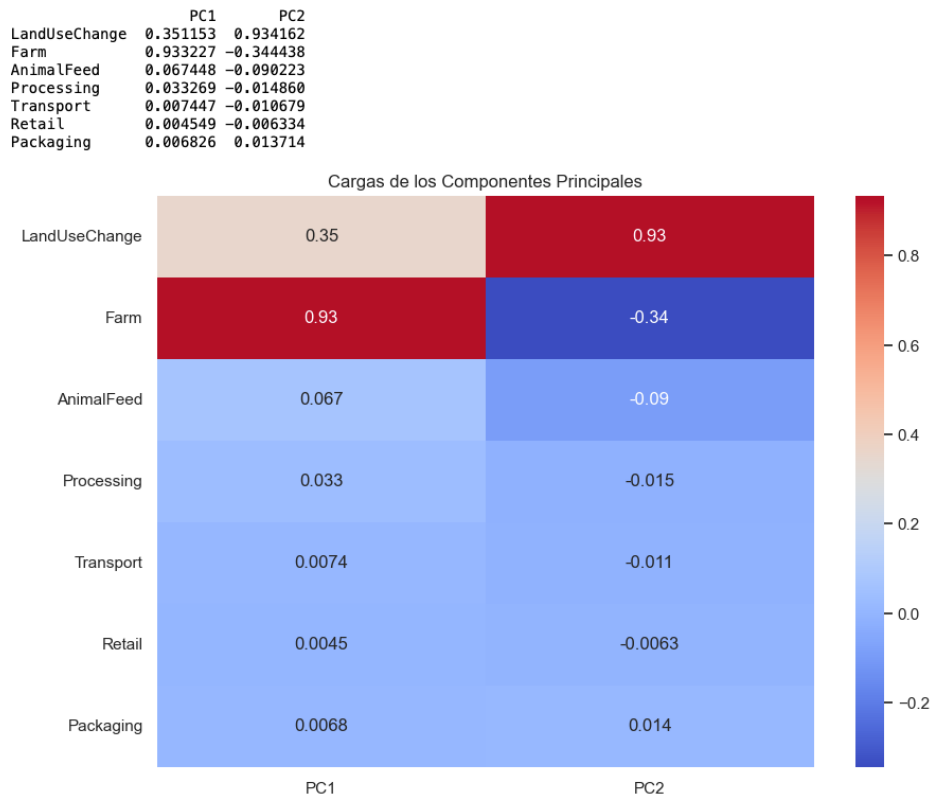


Figura 8: Coeficientes de carga en los Componentes Principales

La Figura 8 revela la influencia de cada variable original en los dos componentes principales (PC1 y PC2). Como podemos observar, la variabilidad en las emisiones de gases de efecto invernadero capturada por PC1 está mayor y fuertemente influenciada por las emisiones en la etapa de “Farm”, seguida por las del cambio en los usos del suelo. Las demás variables tienen cargas muy pequeñas en PC1, lo que indica que su contribución a la variabilidad capturada por este componente es mínima. Con respecto a PC2, “LandUseChange” es ahora la mayor contribuyente (0,93), mientras que, por ejemplo, “Farm” tiene una carga negativa significativa (-0,34), indicando una relación inversa con PC2 en comparación con su relación en PC1.

Así, confirmamos que cualquier estrategia para reducir las emisiones de gases de efecto invernadero deberá centrarse en las etapas de cultivo y cambios en el uso de la tierra, ya que son estas las que dominan la variabilidad en las emisiones. Las políticas y medidas de manejo ambiental que apunten a estas dos etapas probablemente serán las más efectivas para reducir las emisiones totales.

## 5.5. ANÁLISIS DE REGRESIÓN LINEAL SIMPLE

En último lugar, vamos a realizar un análisis de regresión lineal simple. Dado que la variable “Total” es simplemente la suma de las otras variables, no tiene sentido llevar a cabo un modelo de predicción del total de emisiones a partir de sus componentes mediante técnicas de regresión complejas como Random Forest, ya que la relación es puramente aditiva. Sin embargo, sí que podemos usar un análisis de regresión lineal para demostrar y validar la relación aditiva entre las variables independientes y “Total”. Este análisis nos permitirá corroborar que nuestra base de datos es correcta y que sigue una estructura lógica, confirmando así la integridad y coherencia de los datos utilizados.

```
# Análisis de Regresión Lineal Simple
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# Definimos las características (X) y la variable objetivo (y)
X = df[cadena_suministro]
y = df['Total']

# Creamos y entrenamos el modelo de regresión lineal
model = LinearRegression()
model.fit(X, y)

# Hacemos las predicciones
y_pred = model.predict(X)

# Evaluamos el modelo
mse = mean_squared_error(y, y_pred)
r2 = r2_score(y, y_pred)

print(f'Mean Squared Error: {mse}')
print(f'R^2 Score: {r2}')

# Coeficientes del modelo
coefficients = pd.DataFrame(model.coef_, X.columns, columns=['Coefficient'])
coefficients.loc['Intercept'] = model.intercept_
print(coefficients)
```

Mean Squared Error: 3.6718078335085023e-29

R^2 Score: 1.0

	Coefficient
LandUseChange	1.000000e+00
Farm	1.000000e+00
AnimalFeed	1.000000e+00
Processing	1.000000e+00
Transport	1.000000e+00
Retail	1.000000e+00
Packaging	1.000000e+00
Intercept	-1.776357e-15

Figura 9: Código y Output del Modelo de Regresión Lineal

Los resultados obtenidos muestran que las predicciones del modelo son prácticamente perfectas, mediante un MSE extremadamente pequeño. Además, el  $R^2$  Score demuestra que el modelo explica el 100% de la variabilidad en las emisiones de GEI, lo que indica un excelente desempeño predictivo. Ambos resultados resultan consistentes con la relación matemática exacta entre “Total” y sus componentes.

El Error Cuadrático Medio o MSE es una medida utilizada para evaluar la calidad de un modelo de regresión. Representa la media de los cuadrados de los errores, es decir, las diferencias entre los valores predichos por el modelo y los valores reales. Un MSE tan bajo como el nuestro indica que los valores predichos son prácticamente iguales a los valores reales, lo que significa que el modelo tiene un buen desempeño. Mientras tanto, el Coeficiente de Determinación o  $R^2$  es una medida estadística que mide cómo de bien se ajustan los valores predichos del modelo a los valores reales. Si el resultado es igual a 1, el 100% de la varianza en las emisiones de GEI puede explicarse mediante las características incluidas en el modelo.

Por otro lado, todos los coeficientes (“LandUseChange”, “Farm”, “AnimalFeed”, “Processing”, “Transport”, “Retail”, “Packaging”) son exactamente 1, lo que indica que cada una de estas variables contribuye lineal y directamente a “Total”. Por su parte, el valor obtenido para el Intercepto es extremadamente cercano a cero, lo que confirma que no hay un término constante adicional en la relación entre el total de las emisiones y las variables independientes. Es decir, si todas las variables independientes son iguales a cero, el “Total” también deberá ser cero.

En conclusión, como ya habíamos introducido, este modelo de regresión lineal simple es más una validación matemática que una predicción estadística, ya que la relación entre las variables es conocida y exacta. Nos confirma así que la estructura de los datos de nuestro *dataset* es correcta y que el total es calculado como la suma del resto de variables sin errores.

## 6. CONCLUSIONES

En el transcurso de este estudio, hemos navegado por los intrincados caminos de la huella medioambiental de la industria alimentaria. Desde el inicio se ha tratado de contextualizar un problema que se erige como uno de los más urgentes de nuestro tiempo: la creciente presión que la producción de alimentos ejerce sobre los recursos naturales y su contribución al cambio climático, la pérdida de la biodiversidad y la degradación del suelo. A través de un objetivo y una metodología bien delineados, hemos explorado cómo la analítica de datos, la Inteligencia Artificial y, en concreto, el Machine Learning, pueden convertirse en poderosos aliados en la búsqueda de un equilibrio sostenible.

El examen de estrategias analíticas aplicables a la medición de la huella medioambiental revela cómo la analítica de datos no solo tiene el potencial de optimizar procesos y reducir desperdicios, sino también de transformar por completo nuestra aproximación a la sostenibilidad. Además, en nuestro análisis de la industria alimentaria, se ha destacado la importancia de la responsabilidad social corporativa, no solo como una obligación ética y legal, sino como una oportunidad estratégica para innovar y liderar en un mercado cada vez más consciente de la importancia del asunto.

La literatura científica revisada destaca la gran presión que esta industria ejerce sobre los recursos naturales. Estudios previos han demostrado que la alimentación representa aproximadamente entre el 20% y el 30% del impacto ambiental relacionado con todo el consumo público y privado. Además, esta industria es responsable de un tercio de las emisiones de gases de efecto invernadero del planeta, el 80% de la pérdida de biodiversidad y la deforestación y el 70% del consumo de agua dulce. También es bien conocido que las dietas basadas en productos de origen animal tienen un impacto significativamente mayor en comparación con las que están basadas en productos vegetales, requiriendo las primeras aproximadamente tres veces más territorio cultivado que las dietas vegetarianas.

En aras de corroborar y expandir tales hallazgos científicos, se ha llevado a cabo un análisis empírico mediante la implementación de métodos estadísticos y técnicas de ML,

cuyas conclusiones nos han revelado una coherencia con lo estudiado inicialmente. Mediante el análisis de datos de emisiones de gases de efecto invernadero para diferentes alimentos, se han podido identificar patrones claros y se ha logrado cuantificar el impacto relativo que cada etapa de la cadena de suministro supone sobre el medioambiente. Los modelos para el análisis descriptivo y el análisis de componentes principales (PCA) han demostrado que la etapa de cultivo y el cambio de uso del suelo son los principales responsables de las emisiones totales de GEI en la producción de un alimento. El análisis de correlación ha podido reforzar también estos resultados, identificando las etapas que resultan más críticas y la relación que existe entre todas ellas. En última instancia, la utilidad de estas técnicas para evaluar y prever los impactos ambientales en la industria alimentaria se ha logrado validar mediante un análisis de regresión lineal simple, que ha servido igualmente para confirmar la estructura correcta del *dataset* empleado y la exactitud de los cálculos. Los resultados del modelo de regresión mostraron una precisión casi perfecta.

En síntesis, consideramos que la integración de los hallazgos de la literatura con los resultados empíricos ha contribuido a resaltar una vez más la importancia de focalizar los esfuerzos de mitigación de la huella de carbono en las etapas de producción que generan un mayor impacto. La aplicación de técnicas de ML no solo nos permite una mejor comprensión de los impactos ambientales, sino que también proporciona una base sólida para la toma de decisiones informadas que nos permita avanzar hacia un sistema alimentario más resiliente y ecológicamente responsable. Así, defendemos desde aquí que cualquier política dirigida a la sostenibilidad en la industria alimentaria debe incluir la adopción de prácticas agrícolas más sostenibles, como la rotación de cultivos y el uso eficiente del agua, además de la promoción de dietas con menor contenido de productos de origen animal. Las empresas del sector deben también tomar el papel que les corresponde en la implementación de estas políticas sostenibles, utilizando estos *insights* para optimizar sus procesos de producción, distribución y comercialización. Esto incluye la mejora en la eficiencia energética, la reducción del desperdicio de alimentos y la adopción de tecnologías limpias.

## DECLARACIÓN DE USO DE HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL GENERATIVA EN TRABAJOS FIN DE GRADO

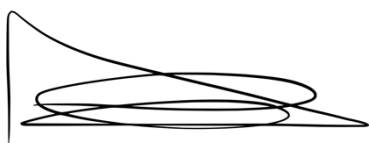
**ADVERTENCIA:** Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Diego Díez Borque, estudiante de Derecho y Business Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado “Un camino hacia la sostenibilidad: La huella medioambiental en la industria alimentaria. El papel de la analítica de datos”, declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Brainstorming de ideas de investigación:** Utilizado para idear y esbozar posibles áreas de investigación.
2. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
3. **Interpretador de código:** Para realizar análisis de datos preliminares.
4. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
5. **Generador previo de diagramas de flujo y contenido:** Para esbozar diagramas iniciales.
6. **Generador de problemas de ejemplo:** Para ilustrar conceptos y técnicas.
7. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 18 de junio de 2024.



Firma: \_\_\_\_\_

## 7. BIBLIOGRAFÍA

- Alvarado Rojas, M. (2015). Una mirada a la inteligencia artificial. *Revista de Ingeniería, Matemáticas y Ciencias de la Información*, 2(3). Recuperado de: <http://ojs.urepublicana.edu.co/index.php/ingenieria/article/view/234>
- Ballester, R. et al. (2015). *Política, salud y enfermedad en España*. Universidad Miguel Hernández de Elche.
- Bybee, R. (1991). Planet Earth in crisis: how should science educators respond? *The American Biology Teacher*, 53(3), 146-153.
- Carpintero, O. (2006). La huella ecológica de la agricultura y la alimentación en España, 1955-2000. *Áreas. Revista Internacional de Ciencias Sociales*, 25, 31-45.
- Claver-Cortés, E. et al. (2018). *Industria alimentaria: fortalezas y debilidades*. Egítania Scientia.
- Cohen, J. (1995). *How Many People Can the Earth Support?* WW Norton.
- Ecoalf. (s. f.). *Propósito Ecoalf*. Ecoalf. <https://ecoalf.com/pages/proposito>
- Emisiones de gases efecto invernadero por actividad*. (2024, febrero). Ministerio de Industria y Turismo. Recuperado de: [https://www.mintur.gob.es/es-es/IndicadoresyEstadisticas/BoletinEstadistico/Energía%20y%20emisiones/IV\\_14.pdf](https://www.mintur.gob.es/es-es/IndicadoresyEstadisticas/BoletinEstadistico/Energía%20y%20emisiones/IV_14.pdf)
- Escobar, R. S. (2022). *Machine Learning: Qué es, funcionamiento y aplicaciones*. Openwebinars.net Recuperado de: <https://openwebinars.net/blog/machine-learning-que-es-funcionamiento-y-aplicaciones/>
- Escolán, E., & Mejuto, S. (2018). La industria alimentaria, polo de competitividad, dinamismo y... ¿empleo? *Equipos y Talento*, 12-15.
- Escrig Zaragoza, D. (2008). El impacto ambiental de las actividades industriales: el cambio necesario. En Rodríguez Jiménez, J. J. (dir.), *Hacia un uso sostenible de los recursos naturales* (pp. 53-61). Universidad Internacional de Andalucía..
- FAO. (2011). *Global food losses and food waste – Extent, causes and prevention*. Recuperado de: <https://www.fao.org/4/mb060e/mb060e00.htm>
- FAO. (2017). *El futuro de la alimentación y la agricultura. Tendencias y desafíos*. Recuperado de: <https://www.fao.org/agrifood-economics/publications/detail/es/c/1475527/>



- Fernández Casadevante «Kois», J. L., & Morán, N. (2015). *Raíces en el asfalto. Pasado, presente y futuro de la agricultura urbana*. Libros en Acción.
- Follos, F. (2014). El Greenwashing o cómo engañar con el medio ambiente. *La Calidad Ambiental*. Recuperado de: <http://ferfollos.blogspot.com/2014/04/el-greenwashing-o-como-enganar-con-el.html>
- Gerbens-Leenes, P. W., Nonhebel, S., & Ivens, W. P. M. F. (2002). A method to determine land requirements relating to food consumption patterns. *Agriculture, Ecosystems and Environment*, 90(1), 47-58.
- INE (2020). Estadística sobre el suministro y saneamiento del agua. Últimos datos. Recuperado de: [https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica\\_C&cid=1254736176834&menu=ultiDatos&idp=1254735976602](https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176834&menu=ultiDatos&idp=1254735976602)
- INE (2023). Estadísticas sobre generación de residuos. Últimos datos. Recuperado de: [https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=estadistica\\_C&cid=1254736176841&menu=ultiDatos&idp=1254735976612](https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=estadistica_C&cid=1254736176841&menu=ultiDatos&idp=1254735976612)
- Informe anual de la industria alimentaria española, período 2019 – 2020*. (2021). MAPA. Recuperado de: [https://www.mapa.gob.es/es/alimentacion/temas/industria-agroalimentaria/\\_20210114informeanualindustria2019-2020ok\\_tcm30-542507.pdf](https://www.mapa.gob.es/es/alimentacion/temas/industria-agroalimentaria/_20210114informeanualindustria2019-2020ok_tcm30-542507.pdf)
- Lewin, R. (1997). *La sexta extinción*. Tusquets Editores.
- De La Maza Asquet, C. (2007). Evaluación de impactos ambientales. En *Manejo y conservación de recursos naturales* (pp. 579-609). Editorial Universitaria. Universidad de Chile.
- Nestlé. (2021). *Climate action*. Nestlé Global. Recuperado de: <https://www.nestle.com/sustainability/climate-change>
- OCU. (2020). *Hábitos alimentarios durante el confinamiento*. Organización de Consumidores y Usuarios. Recuperado de: <https://www.ocu.org/consumo-familia/derechos-consumidor/noticias/encuesta-habitos-coronavirus>
- OIT. (2007). *La incidencia de las cadenas mundiales de alimentación en el empleo en el sector de alimentación y bebidas*.
- Pearse, G. (2014). *The Greenwash Effect: Corporate Deception, Celebrity Environmentalists, and What Big Business Isn't Telling You about Their Green Products and Brands*. Skyhorse.

- Ritchie, H., Rosado, P., & Roser, M. (2022). *Environmental Impacts of Food Production*. Our World In data. Recuperado de: <https://ourworldindata.org/environmental-impacts-of-food>
- Stuart, T. (2011). *Despilfarro. El escándalo global de la comida*. Alianza Editorial.
- Suárez Tamayo, S., & Molina Esquivel, E. (2014). El desarrollo industrial y su impacto en el medio ambiente. *Revista cubana de higiene y epidemiología*, 52(3), 357-363.
- Tukker, A., & Jansen, B. (2006). Environmental impacts of products: A detailed review of studies. *Journal of Industrial Ecology*, 10(3), 159-182.
- Vilches, A., & Gil, D. (2008). La construcción de un futuro sostenible en un planeta en riesgo. *Alambique*, 55(1), 9-19.
- World Bank Group. (2022). *El Banco Mundial anuncia acciones previstas para responder a la crisis mundial de alimentos*. World Bank. Recuperado de: <https://www.bancomundial.org/es/news/press-release/2022/05/18/world-bank-announces-planned-actions-for-global-food-crisis-response>

# ANEXO I

In [1]:

```
import math as math
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib as mpl
```

In [2]:

```
df = pd.read_excel("/Users/diego/Downloads/Food_Greenhouse_gas_emissions.xlsx")
```

In [3]:

```
# Mostrar las primeras filas del archivo
print("Primeras filas del archivo:")
print(df.head())

# Mostrar la información general del dataframe
print("\nInformación general del dataframe:")
print(df.info())

# Mostrar estadísticas descriptivas del dataframe
print("\nEstadísticas descriptivas del dataframe:")
print(df.describe())

# Mostrar la lista de columnas
print("\nColumnas en el dataframe:")
print(df.columns)

# Mostrar el número de filas y columnas
print(f"\nNúmero de filas y columnas: {df.shape}")

# Mostrar cualquier valor nulo en el dataframe
print("\nValores nulos en el dataframe:")
print(df.isnull().sum())
```

Primeras filas del archivo:

	Food	Land use change (kgCO <sub>2</sub> eq per kg)	Farm (kgCO <sub>2</sub> eq per kg)
0	Apples	-0.028951	0.2258
1	Bananas	-0.025534	0.2696
2	Barley	0.008676	0.1763
3	Beef (beef herd)	23.237536	56.2280
4	Beef (dairy herd)	1.266022	21.9155
	Animal feed (kgCO <sub>2</sub> eq per kg)	Processing (kgCO <sub>2</sub> eq per kg)	
0	0.000000	0.003820	

1	0.000000	0.060146
2	0.000000	0.127703
3	2.680979	1.811083
4	3.503673	1.547664

	Transport (kgCO <sub>2</sub> eq per kg)	Retail (kgCO <sub>2</sub> eq per kg)	\
0	0.095804	0.016578	
1	0.295830	0.020981	
2	0.035351	0.263587	
3	0.494125	0.233538	
4	0.592410	0.254032	

	Packaging (kgCO <sub>2</sub> eq per kg)	Total emissions (kgCO <sub>2</sub> eq per kg)
0	0.044526	0.357593
1	0.065670	0.686712
2	0.496786	1.108480
3	0.352084	85.037407
4	0.374549	29.453881

Información general del dataframe:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 43 entries, 0 to 42

Data columns (total 9 columns):

#	Column	Non-Null Count	Dtype
0	Food	43 non-null	object
1	Land use change (kgCO <sub>2</sub> eq per kg)	43 non-null	float64
2	Farm (kgCO <sub>2</sub> eq per kg)	43 non-null	float64
3	Animal feed (kgCO <sub>2</sub> eq per kg)	43 non-null	float64
4	Processing (kgCO <sub>2</sub> eq per kg)	43 non-null	float64
5	Transport (kgCO <sub>2</sub> eq per kg)	43 non-null	float64
6	Retail (kgCO <sub>2</sub> eq per kg)	43 non-null	float64
7	Packaging (kgCO <sub>2</sub> eq per kg)	43 non-null	float64
8	Total emissions (kgCO <sub>2</sub> eq per kg)	43 non-null	float64

dtypes: float64(8), object(1)

memory usage: 3.1+ KB

None

Estadísticas descriptivas del dataframe:

	Land use change (kgCO <sub>2</sub> eq per kg)	Farm (kgCO <sub>2</sub> eq per kg)	\
count	43.000000	43.000000	
mean	1.731747	4.523337	
std	5.263927	9.933745	
min	-3.257812	0.092768	
25%	0.001556	0.339656	
50%	0.179972	0.927822	
75%	1.076473	2.925752	
max	25.814833	56.228063	

	Animal feed (kgCO <sub>2</sub> eq per kg)	Processing (kgCO <sub>2</sub> eq per kg)	\
count	43.000000	43.000000	
mean	0.625230	0.295458	
std	1.272308	0.453325	
min	0.000000	0.000000	
25%	0.000000	0.001910	
50%	0.000000	0.078319	
75%	0.000000	0.372331	

```
max                4.298476                1.811083
```

```

      Transport (kgCO2eq per kg)  Retail (kgCO2eq per kg)  \
count                43.000000                43.000000
mean                 0.219390                0.098952
std                  0.184951                0.108912
min                  0.035351                0.015135
25%                  0.094924                0.038195
50%                  0.134528                0.039322
75%                  0.262454                0.161757
max                  0.794510                0.352361

      Packaging (kgCO2eq per kg)  Total emissions (kgCO2eq per kg)
count                43.000000                43.000000
mean                 0.295123                7.789238
std                  0.337939                14.973011
min                  0.041748                0.316055
25%                  0.074762                0.905071
50%                  0.137537                2.153031
75%                  0.403418                6.215340
max                  1.687496                85.037407

```

Columnas en el dataframe:

```
Index(['Food', 'Land use change (kgCO2eq per kg)', 'Farm (kgCO2eq per kg)',
      'Animal feed (kgCO2eq per kg)', 'Processing (kgCO2eq per kg)',
      'Transport (kgCO2eq per kg)', 'Retail (kgCO2eq per kg)',
      'Packaging (kgCO2eq per kg)', 'Total emissions (kgCO2eq per kg)'],
      dtype='object')
```

Número de filas y columnas: (43, 9)

Valores nulos en el dataframe:

```
Food                0
Land use change (kgCO2eq per kg)  0
Farm (kgCO2eq per kg)            0
Animal feed (kgCO2eq per kg)      0
Processing (kgCO2eq per kg)        0
Transport (kgCO2eq per kg)        0
Retail (kgCO2eq per kg)           0
Packaging (kgCO2eq per kg)        0
Total emissions (kgCO2eq per kg)  0
dtype: int64
```

```
In [4]: df.head(10)
```

Out [4]:

	Food	Land use change (kgCO <sub>2</sub> eq per kg)	Farm (kgCO <sub>2</sub> eq per kg)	Animal feed (kgCO <sub>2</sub> eq per kg)	Processing (kgCO <sub>2</sub> eq per kg)	Transport (kgCO <sub>2</sub> eq per kg)	Retail (kgCO <sub>2</sub> eq per kg)	Packaging (kgCO <sub>2</sub> eq per kg)
0	Apples	-0.028951	0.225816	0.000000	0.003820	0.095804	0.016578	0.044526
1	Bananas	-0.025534	0.269618	0.000000	0.060146	0.295830	0.020981	0.065670
2	Barley	0.008676	0.176377	0.000000	0.127703	0.035351	0.263587	0.496786
3	Beef (beef herd)	23.237536	56.228063	2.680979	1.811083	0.494125	0.233538	0.352084
4	Beef (dairy herd)	1.266022	21.915529	3.503673	1.547664	0.592410	0.254032	0.374549
5	Beet Sugar	0.000052	0.535721	0.000000	0.241885	0.628544	0.039322	0.090289
6	Berries & Grapes	0.024601	0.717848	0.000000	0.000000	0.236392	0.017057	0.211092
7	Brassicas	0.002242	0.277745	0.000000	0.000000	0.094641	0.016844	0.045303
8	Cane Sugar	1.263041	0.491270	0.000000	0.037381	0.794510	0.036723	0.084277
9	Cassava	0.589436	0.219994	0.000000	0.000000	0.093597	0.039322	0.044822

In [5]:

```
print(df.columns)
```

```
Index(['Food', 'Land use change (kgCO2eq per kg)', 'Farm (kgCO2eq per kg)', 'Animal feed (kgCO2eq per kg)', 'Processing (kgCO2eq per kg)', 'Transport (kgCO2eq per kg)', 'Retail (kgCO2eq per kg)', 'Packaging (kgCO2eq per kg)', 'Total emissions (kgCO2eq per kg)'], dtype='object')
```

In [6]:

```
# Acortamos los nombres de las variables
df.rename(columns = {'Land use change (kgCO2eq per kg)': 'LandUseChange',
                    'Farm (kgCO2eq per kg)': 'Farm',
                    'Animal feed (kgCO2eq per kg)': 'AnimalFeed',
                    'Processing (kgCO2eq per kg)': 'Processing',
                    'Transport (kgCO2eq per kg)': 'Transport',
                    'Retail (kgCO2eq per kg)': 'Retail',
                    'Packaging (kgCO2eq per kg)': 'Packaging',
                    'Total emissions (kgCO2eq per kg)': 'Total'},
          inplace = True)
print(df.columns)
```

```
Index(['Food', 'LandUseChange', 'Farm', 'AnimalFeed', 'Processing', 'Transport', 'Retail', 'Packaging', 'Total'], dtype='object')
```

In [7]:

```

# CLUSTERING - Creamos categorías para agrupar productos
df["Category"] = df["Food"]

Cereal = ["Wheat & Rye", "Maize", "Oatmeal", "Barley", "Rice"]
Frutos_Secos = ['Nuts', 'Groundnuts']
Verdura = ["Potatoes", "Cassava", 'Other Pulses', "Peas", 'Tomatoes', 'Onions']
Fruta = ['Citrus Fruit', 'Bananas', 'Apples', 'Berries & Grapes', 'Other Fru
Azúcar = ['Cane Sugar', 'Beet Sugar',]
Aceite = ['Soybean Oil', 'Palm Oil', 'Sunflower Oil', 'Rapeseed Oil', 'Oliv
Lácteos = ["Soy milk", 'Milk', 'Cheese']
Animal = ['Beef (beef herd)', 'Beef (dairy herd)', 'Lamb & Mutton', 'Pig Mea
Otros = ["Tofu", "Coffee", "Dark Chocolate", "Wine"]

# Sustituimos todos los productos en la nueva columna por su respectiva cat
for i in df["Category"]:
    if i in Cereal:
        df["Category"].replace([i], "Cereal", inplace=True)
    elif i in Frutos_Secos:
        df["Category"].replace([i], "Frutos_Secos", inplace=True)
    elif i in Verdura:
        df["Category"].replace([i], "Verdura", inplace=True)
    elif i in Fruta:
        df["Category"].replace([i], "Fruta", inplace=True)
    elif i in Azúcar:
        df["Category"].replace([i], "Azúcar", inplace=True)
    elif i in Aceite:
        df["Category"].replace([i], "Aceite", inplace=True)
    elif i in Lácteos:
        df["Category"].replace([i], "Lácteos", inplace=True)
    elif i in Animal:
        df["Category"].replace([i], "Animal", inplace=True)
    elif i in Otros:
        df["Category"].replace([i], "Otros", inplace=True)

display(df[['Food', 'Category']])

```

	Food	Category
0	Apples	Fruta
1	Bananas	Fruta
2	Barley	Cereal
3	Beef (beef herd)	Animal
4	Beef (dairy herd)	Animal
5	Beet Sugar	Azúcar
6	Berries & Grapes	Fruta
7	Brassicas	Verdura
8	Cane Sugar	Azúcar
9	Cassava	Verdura
10	Cheese	Lácteos

11	Citrus Fruit	Fruta
12	Coffee	Otros
13	Dark Chocolate	Otros
14	Eggs	Animal
15	Fish (farmed)	Animal
16	Groundnuts	Frutos_Secos
17	Lamb & Mutton	Animal
18	Maize	Cereal
19	Milk	Lácteos
20	Nuts	Frutos_Secos
21	Oatmeal	Cereal
22	Olive Oil	Aceite
23	Onions & Leeks	Verdura
24	Other Fruit	Fruta
25	Other Pulses	Verdura
26	Other Vegetables	Verdura
27	Palm Oil	Aceite
28	Peas	Verdura
29	Pig Meat	Animal
30	Potatoes	Verdura
31	Poultry Meat	Animal
32	Rapeseed Oil	Aceite
33	Rice	Cereal
34	Root Vegetables	Verdura
35	Shrimps (farmed)	Animal
36	Soy milk	Lácteos
37	Soybean Oil	Aceite
38	Sunflower Oil	Aceite
39	Tofu	Otros
40	Tomatoes	Verdura
41	Wheat & Rye	Cereal
42	Wine	Otros



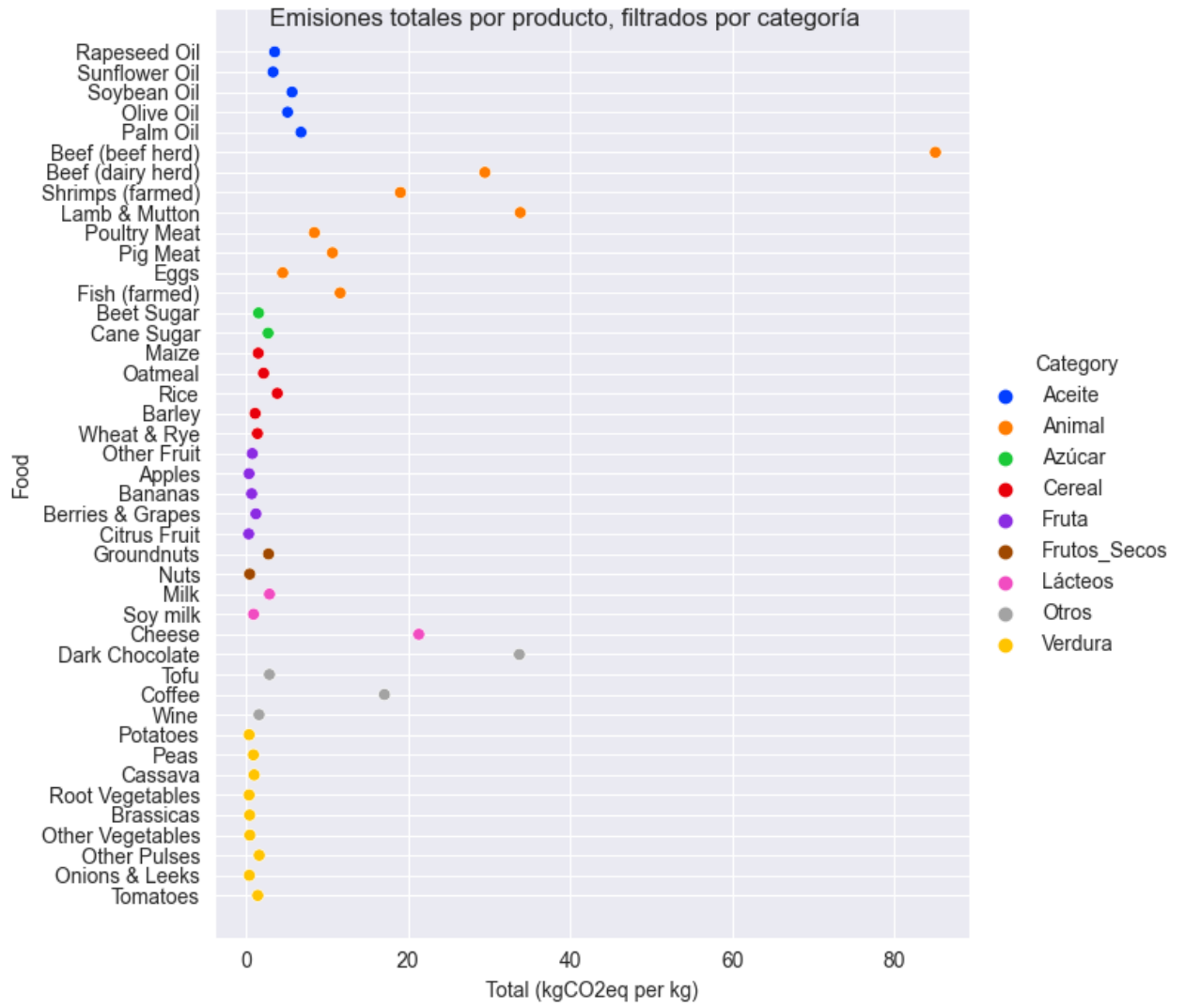
```
In [8]: # Establecemos la columna 'Food' como índice del DataFrame
df.set_index('Food', inplace=True)
df.head(10)
```

```
Out[8]:
```

	LandUseChange	Farm	AnimalFeed	Processing	Transport	Retail	Packa
<b>Food</b>							
<b>Apples</b>	-0.028951	0.225816	0.000000	0.003820	0.095804	0.016578	0.04
<b>Bananas</b>	-0.025534	0.269618	0.000000	0.060146	0.295830	0.020981	0.06
<b>Barley</b>	0.008676	0.176377	0.000000	0.127703	0.035351	0.263587	0.49
<b>Beef (beef herd)</b>	23.237536	56.228063	2.680979	1.811083	0.494125	0.233538	0.35
<b>Beef (dairy herd)</b>	1.266022	21.915529	3.503673	1.547664	0.592410	0.254032	0.37
<b>Beet Sugar</b>	0.000052	0.535721	0.000000	0.241885	0.628544	0.039322	0.09
<b>Berries &amp; Grapes</b>	0.024601	0.717848	0.000000	0.000000	0.236392	0.017057	0.2
<b>Brassicas</b>	0.002242	0.277745	0.000000	0.000000	0.094641	0.016844	0.04
<b>Cane Sugar</b>	1.263041	0.491270	0.000000	0.037381	0.794510	0.036723	0.08
<b>Cassava</b>	0.589436	0.219994	0.000000	0.000000	0.093597	0.039322	0.04

```
In [11]: # Scatter plot de emisiones totales por producto y categoría
sns.set_style("darkgrid")
rel_plot = sns.relplot(data=df.sort_values(by='Category'), x='Total', y='Food',
                        hue='Category', palette="bright",
                        height=7, aspect=1)
rel_plot.fig.suptitle("Emisiones totales por producto, filtrados por categoría")
rel_plot.set_axis_labels("Total (kgCO2eq per kg)", "Food")
```

Out[11]: <seaborn.axisgrid.FacetGrid at 0x7f904e2170d0>



In [12]:

```
# Gráfica de pastel de las emisiones totales de los 10 principales productos
df_top10totalemm = (df.groupby('Food')['Total']
                    .sum()
                    .nlargest(10))

# Calculamos el porcentaje de las emisiones totales
diff_top10 = round((df_top10totalemm.sum()/df['Total'].sum()*100, 1)

labels = df_top10totalemm.index
num_labels = len(labels)

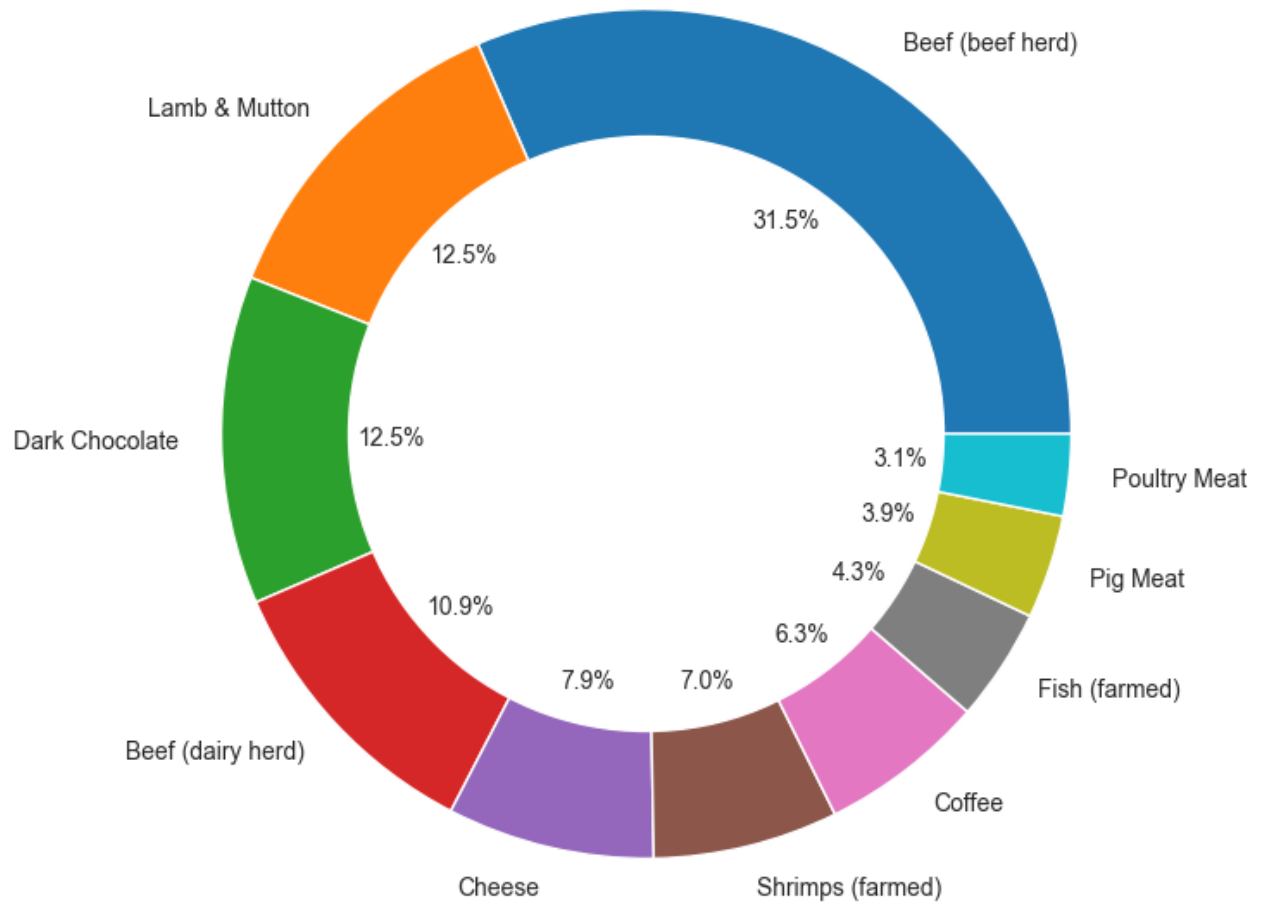
fig, ax = plt.subplots(figsize=(13, 8), facecolor='#FFFFFF')
wedges, texts, autotexts = ax.pie(df_top10totalemm, labels=labels, autopct=

donut = plt.Circle((0, 0), 0.7, color='white')
ax.add_artist(donut)

ax.set_title('Emisiones totales del Top 10', fontsize=15)
plt.show()

print(f"Suma global de emisiones: {df['Total'].sum().round(2)} kgCO2eq per
      f"Suma de emisiones de los 10 principales productos: {df_top10totalemm
      f"Porcentaje del Top 10 sobre el global: {diff_top10}%")
```

## Emisiones totales del Top 10



Suma global de emisiones: 334.94 kgCO<sub>2</sub>eq per kg

Suma de emisiones de los 10 principales productos: 270.08 kgCO<sub>2</sub>eq per kg

Porcentaje del Top 10 sobre el global: 80.6%

In [13]:

```

# Emisiones de gases de efecto invernadero a lo largo de toda la cadena de
temp_df = df.sort_values(by='Total', ascending=True)

# Definimos las columnas que representan las etapas de la cadena de suministro
cadena_suministro = ['LandUseChange', 'Farm', 'AnimalFeed', 'Processing', '
Transport', 'Retail', 'Packaging']

temp_df = df[cadena_suministro + ['Category']]
category_df = temp_df.groupby('Category').sum()

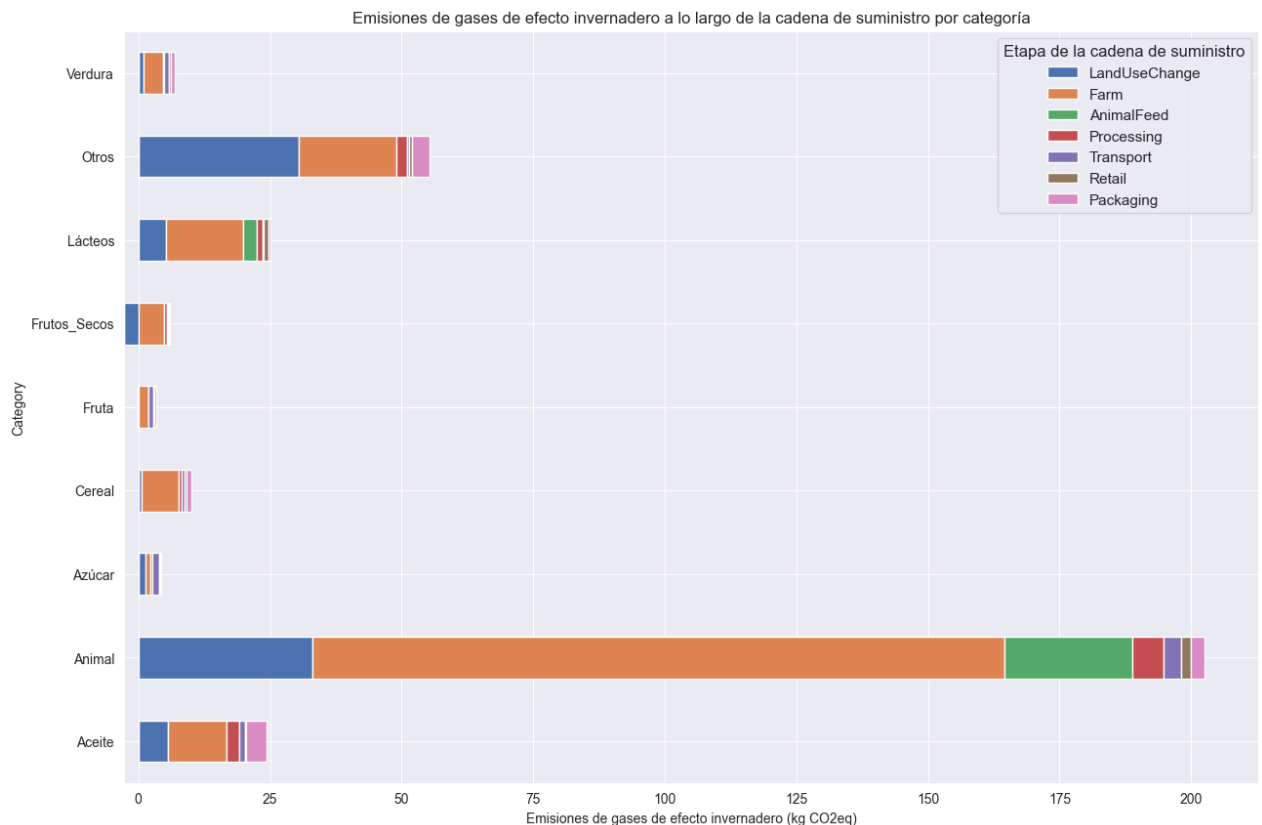
fig, ax = plt.subplots(figsize=(15, 10))

sns.set()
category_df.plot(kind='barh', stacked=True, ax=ax)

plt.xlabel("Emisiones de gases de efecto invernadero (kg CO2eq)")
plt.title("Emisiones de gases de efecto invernadero a lo largo de la cadena de suministro")

plt.legend(title='Etapa de la cadena de suministro')
plt.show()

```



In [14]:

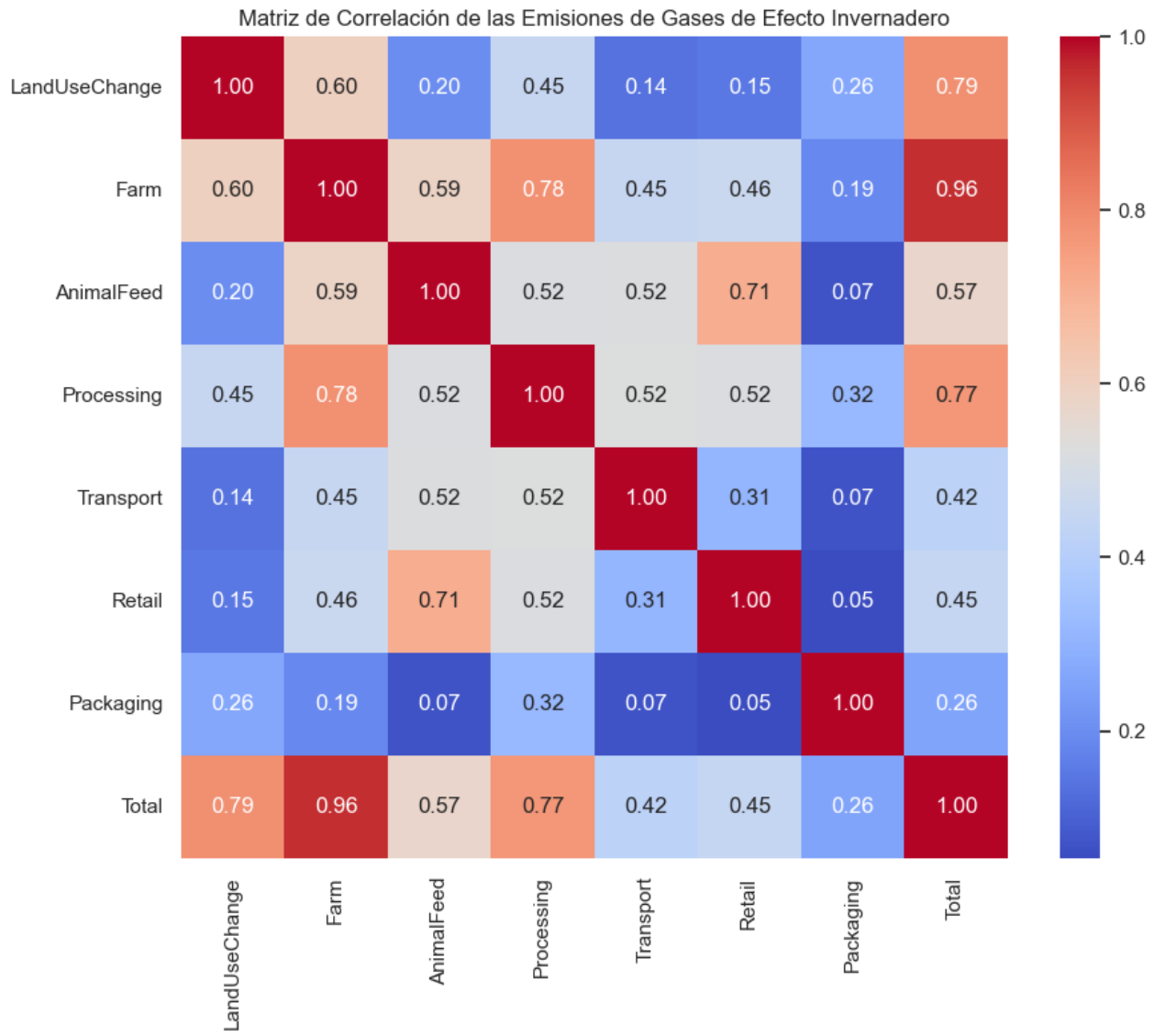
```

# Análisis de Correlación

corr_matrix = df.corr() # Calculamos la matriz de correlación

# Visualizamos
plt.figure(figsize=(10, 8))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Matriz de Correlación de las Emisiones de Gases de Efecto Invernadero')
plt.show()

```



In [15]:

```

# Análisis de Componentes Principales (PCA)
from sklearn.decomposition import PCA

# Seleccionamos las columnas numéricas para el análisis
data = df[cadena_suministro]

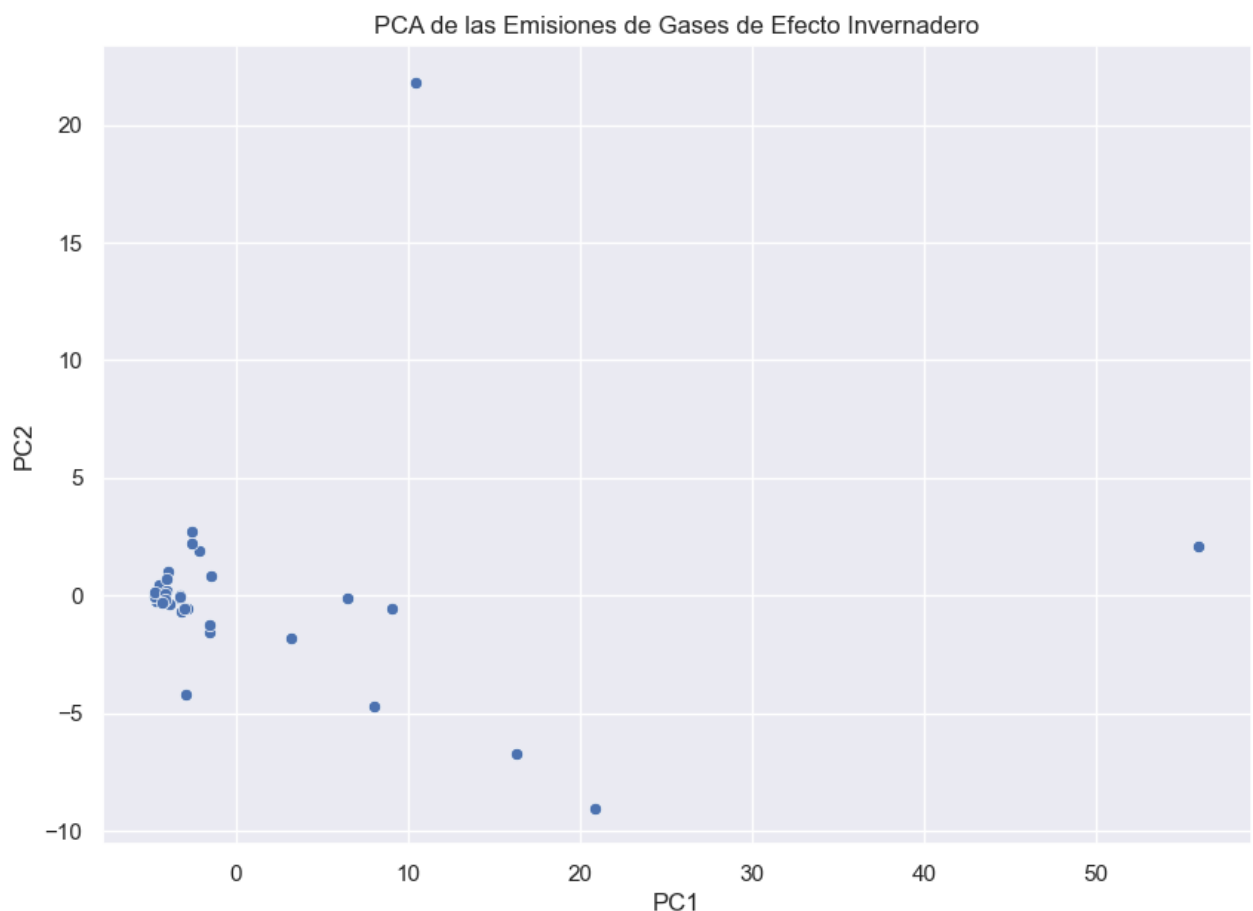
# Aplicamos PCA reduciendo a 2 componentes principales y creamos un dataframe
pca = PCA(n_components=2)
componentes_principales = pca.fit_transform(data)
pca_df = pd.DataFrame(data=componentes_principales, columns=['PC1', 'PC2'])

# Añadimos la categoría al DataFrame PCA para la visualización
pca_df['Category'] = df['Category']

# Visualizamos los componentes principales
plt.figure(figsize=(10, 7))
sns.scatterplot(x='PC1', y='PC2', hue='Category', data=pca_df, palette='bright')
plt.title('PCA de las Emisiones de Gases de Efecto Invernadero')
plt.show()

# Explicamos la varianza
varianza_explicada = pca.explained_variance_ratio_
print(f"Varianza explicada por PC1: {varianza_explicada[0] * 100:.2f}%")
print(f"Varianza explicada por PC2: {varianza_explicada[1] * 100:.2f}%")

```



Varianza explicada por PC1: 86.55%

Varianza explicada por PC2: 12.50%

In [16]:

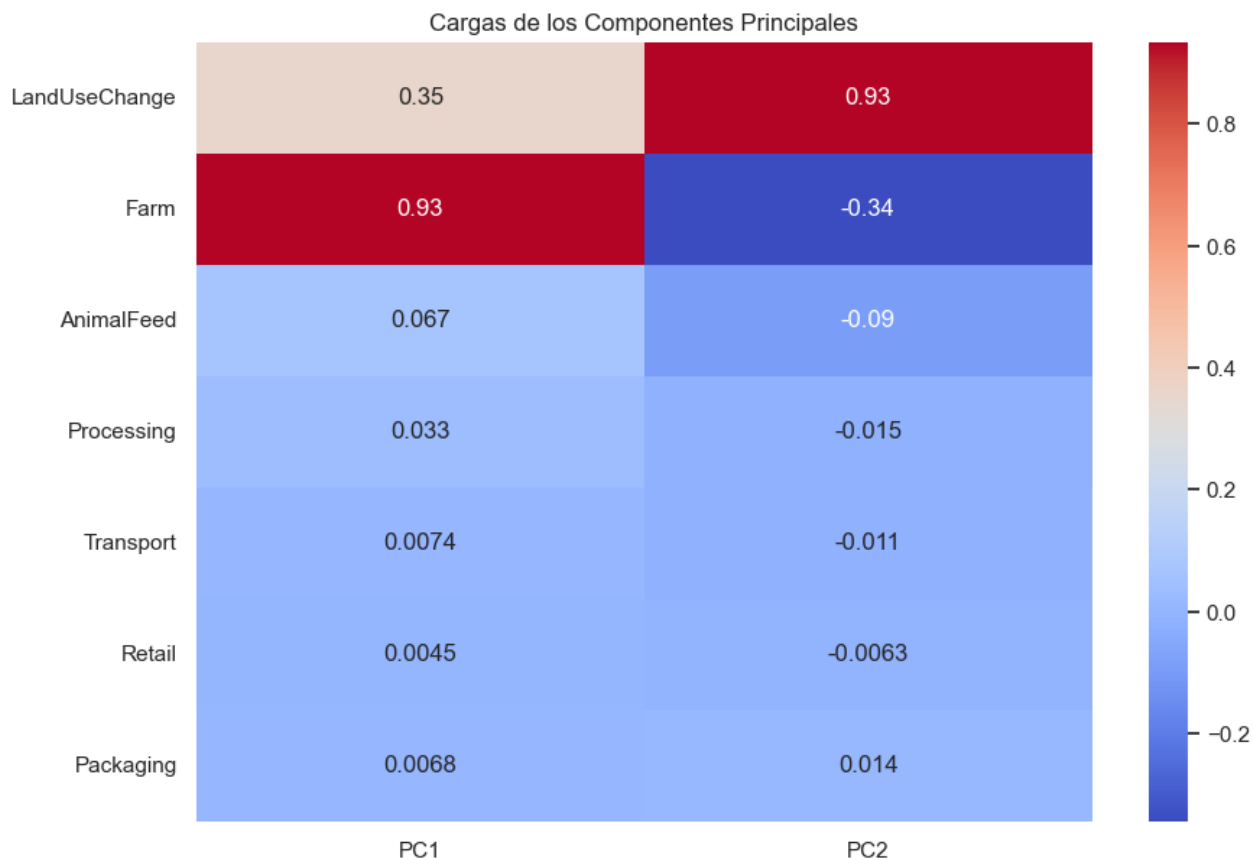
```

# Extraemos las cargas de los componentes principales
loadings = pd.DataFrame(pca.components_.T, columns=['PC1', 'PC2'], index=ca
print(loadings)

# Visualizamos las cargas
plt.figure(figsize=(10, 7))
sns.heatmap(loadings, annot=True, cmap='coolwarm')
plt.title('Cargas de los Componentes Principales')
plt.show()

```

	PC1	PC2
LandUseChange	0.351153	0.934162
Farm	0.933227	-0.344438
AnimalFeed	0.067448	-0.090223
Processing	0.033269	-0.014860
Transport	0.007447	-0.010679
Retail	0.004549	-0.006334
Packaging	0.006826	0.013714





In [17]:

```
# Análisis de Regresión Lineal Simple
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# Definimos las características (X) y la variable objetivo (y)
X = df[cadena_suministro]
y = df['Total']

# Creamos y entrenamos el modelo de regresión lineal
model = LinearRegression()
model.fit(X, y)

# Hacemos las predicciones
y_pred = model.predict(X)

# Evaluamos el modelo
mse = mean_squared_error(y, y_pred)
r2 = r2_score(y, y_pred)

print(f'Mean Squared Error: {mse}')
print(f'R^2 Score: {r2}')

# Coeficientes del modelo
coefficients = pd.DataFrame(model.coef_, X.columns, columns=['Coefficient'])
coefficients.loc['Intercept'] = model.intercept_
print(coefficients)
```

Mean Squared Error: 3.6718078335085023e-29

R^2 Score: 1.0

	Coefficient
LandUseChange	1.000000e+00
Farm	1.000000e+00
AnimalFeed	1.000000e+00
Processing	1.000000e+00
Transport	1.000000e+00
Retail	1.000000e+00
Packaging	1.000000e+00
Intercept	-1.776357e-15

In [ ]: