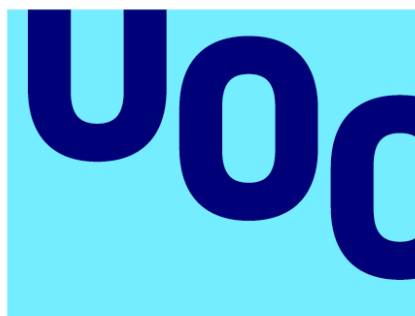


## PEC 4 – Memoria

### ***Desigualdad de precios en la cadena alimentaria: Análisis del IPOD en los productos con mayor diferencia Origen-Destino***



Universitat  
Oberta  
de Catalunya

Estudiante: **Alberto Rodríguez Álvarez**

Profesor/a responsable: **Susana Acedo Nadal**

Profesor colaborador: **Xavier Florit Medina**

Asignatura: **22.536 - Trabajo final de grado**  
Estudios: **Grado de Ciencia de Datos Aplicada**  
Fecha entrega: **03/06/2025**



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial- SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

## FICHA DEL TRABAJO FINAL

<b>Título del trabajo:</b>	Desigualdad de precios en la cadena alimentaria: Análisis del IPOD en los productos con mayor diferencia Origen-Destino
<b>Nombre del autor:</b>	Alberto Rodríguez Álvarez
<b>Nombre del director/a:</b>	Xavier Florit Medina
<b>Nombre del PRA:</b>	Susana Acedo Nadal
<b>Fecha de entrega (mm/aaaa):</b>	03/06/2025
<b>Titulación o programa:</b>	Grado de Ciencia de Datos Aplicada
<b>Área del Trabajo Final:</b>	Análisis de datos y tendencias
<b>Idioma del trabajo:</b>	Castellano
<b>Palabras clave</b>	alimentos precio desigualdad
<b>Resumen del Trabajo</b>	
<p>Este Trabajo de Fin de Grado estudia las disparidades de precios en el sector agroalimentario español, centrado en los productos con mayor diferencia entre precios en origen y destino (IPOD). Se analizan patrones de variación y factores influyentes mediante modelos de series temporales y machine learning, aplicando la metodología CRISP-DM. Se espera identificar causas clave y proponer estrategias para mejorar la transparencia y equidad en la cadena de valor, aportando información útil para políticas públicas y decisiones empresariales.</p>	
<b>Abstract</b>	
<p>This Final Degree Project examines price disparities in the Spanish agri-food sector, focusing on products with the highest origin-destination price gap (IPOD). Using time series models and machine learning under the CRISP-DM methodology, it analyzes price variation patterns and influencing factors. The study aims to identify key causes and propose strategies to enhance transparency and fairness in the supply chain, providing valuable insights for public policies and business decisions.</p> <p>Si necesitas más ajustes, dime y lo afinamos.</p>	

# Índice

## Tabla de contenido

PEC 4 – Memoria.....	1
Índice .....	1
Lista de figuras.....	2
1. Introducción.....	3
1.1. Contexto y justificación del Trabajo .....	3
1.2. Objetivos del Trabajo.....	4
1.3. Impacto en sostenibilidad, ético-social y de diversidad .....	5
1.4. Enfoque y método seguido .....	5
1.5. Planificación del Trabajo.....	7
1.6. Breve resumen de productos obtenidos.....	9
2. Materiales y métodos .....	10
2.1 Fuentes de datos.....	10
2.2 Preparación y fusión de datos (pipeline ETL) .....	11
2.3 Variables de análisis.....	13
3. Modelado y análisis.....	14
3.1 Verificación inicial .....	14
3.2 Análisis temporal diferencia Origen-Destino .....	16
3.3 Modelos de predicción.....	19
3.4 Análisis de la evolución temporal diferencia Origen-Destino.....	21
3.5 Evolución precio Origen-Destino por producto.....	23
4. Conclusiones y trabajos futuros.....	24
5. Glosario.....	26
4. Bibliografía .....	28

# Lista de figuras

## Índice de figuras y tablas

Ilustración 1 Esquema del ciclo CRISP-DM[8] .....	6
Ilustración 2 Fuentes de datos (elaboración propia).....	10
Ilustración 3 Diferencias precios Origen.....	14
Ilustración 4 Diferencias precios Origen Scatterplot.....	14
Ilustración 5 Diferencias precio Destino .....	15
Ilustración 6 Diferencias precio Destino Scatterplot .....	15
Ilustración 7 Diferencias Origen - Mercado - Destino .....	16
Ilustración 8 Precio Patata O-M-D.....	18
Ilustración 9 Resultados ARIMA .....	20
Ilustración 10 Resultados Auto ARIMA .....	20
Ilustración 11 Proyecciones ARIMA .....	21
Ilustración 12 Evolucion temporal Dif O-D eventos .....	22
Ilustración 13 Precio O-D por producto .....	23

# 1. Introducción

## 1.1. Contexto y justificación del Trabajo

El sector agroalimentario español constituye uno de los pilares esenciales para la **economía nacional y para la seguridad alimentaria** y la estabilidad territorial del país: vertebrando amplias zonas rurales, genera empleo directo e indirecto y refuerza el arraigo poblacional allí donde existen pocas actividades económicas alternativas. La cadena de valor agroalimentaria involucra a multitud de actores —productores primarios, cooperativas, mayoristas, logística, distribución minorista y consumidores— cuyas relaciones, con frecuencia asimétricas, influyen decisivamente en la formación de precios y en la redistribución del valor generado. El Índice de Precios en Origen y Destino (IPOD)[1], que en 2024 promedió 4,25, evidencia esa desigualdad: **el consumidor llegó a pagar más de cuatro veces lo que percibió el agricultor**, y en determinados productos el diferencial superó el 400 %, es decir, el precio final quintuplicó al de origen. Este margen comercial desproporcionado impacta negativamente en la rentabilidad de las explotaciones, fomenta el abandono del campo y la despoblación rural, y a la vez encarece la cesta básica de los hogares con menor renta, comprometiendo la **equidad en el acceso a los alimentos**.

La tensión se agrava por factores globales que han introducido mayor volatilidad e incertidumbre: **inflación generalizada, encarecimiento de insumos y energía, crisis climática y eventos extremos**. Los costes de producción se han disparado tal y como reflejan los informes de costes del Ministerio de Agricultura [2] y los registros del ‘Observatori de Preus’ de la Generalitat de Catalunya[3], [4] mientras que los márgenes de los agricultores han permanecido estancados o incluso se han reducido. El sistema actual de fijación de precios resulta, por tanto, insuficiente tanto desde el punto de vista de la **justicia económica** como de la eficiencia distributiva y genera consecuencias sociales (precariedad laboral, pérdida de tejido productivo) y éticas (desequilibrios de poder, falta de transparencia).

En este contexto, el presente Trabajo Final de Grado tiene como propósito analizar en profundidad las disparidades de precios a lo largo de la cadena agroalimentaria, identificar las causas (estructurales, coyunturales o asociadas a dinámicas opacas de mercado) y proponer soluciones basadas en evidencia objetiva. Para ello se construirá un conjunto de datos abiertos, públicos y verificables (MAPA, COAG, Generalitat) sobre el que se aplicarán modelos de series temporales y algoritmos de *machine learning* capaces de detectar patrones, segmentar productos y generar pronósticos. La visualización mediante librerías especializadas de Python facilitará la identificación de tendencias, anomalías y puntos críticos. El objetivo final es **identificar con precisión los puntos críticos y los factores que agravan las desigualdades de precios a lo largo de la cadena de valor agroalimentaria**, generando evidencia rigurosa que sirva de base para la formulación de políticas públicas, la adopción de prácticas empresariales responsables y decisiones de consumo mejor informadas contribuyendo así a la sostenibilidad económica, social y ambiental del sistema alimentario español.

## 1.2. Objetivos del Trabajo

El objetivo principal de este proyecto es analizar de forma rigurosa y sistemática las **disparidades de precios existentes en el sector agroalimentario**, con especial énfasis en las diferencias entre el precio en origen —lo que perciben los productores— y el precio en destino —lo que paga el consumidor final—. A través del estudio del Índice de **Precios en Origen y Destino** (IPOD), se pretende identificar patrones de comportamiento, explorar variaciones significativas a lo largo del tiempo y entre productos, y evaluar las causas estructurales y coyunturales que originan dichas desigualdades. Este análisis busca no solo describir la situación actual, sino también ofrecer una base sólida para la formulación de propuestas que fomenten una mayor equidad en la cadena de valor agroalimentaria, contribuyendo así al diseño de políticas públicas más justas y a la mejora de la eficiencia económica del sistema.

Los objetivos concretos planteados han sido:

- Detectar los productos con mayores brechas de precios origen-destino a través del análisis de los índices IPOD acumulados en los últimos años. Esta selección permitirá centrar el estudio en los casos más representativos y relevantes desde el punto de vista económico y social.
- Estudiar la evolución temporal de los precios agroalimentarios, identificando tendencias, estacionalidades y cambios estructurales que permitan explicar la dinámica de los márgenes comerciales. Este análisis incluirá tanto series cronológicas como comparativas entre distintas campañas agrícolas o periodos económicos.
- Incorporar el análisis de factores externos que influyen en la formación de precios, tales como políticas agrarias, inflación, condiciones climáticas, costes logísticos o presión de los intermediarios. Se buscará evaluar el grado de correlación o causalidad entre estos factores y las oscilaciones detectadas en el IPOD.
- Aplicar técnicas de modelado y visualización que permitan representar de forma clara y comprensible los hallazgos obtenidos. Se utilizarán modelos de series temporales y algoritmos de machine learning para predecir el comportamiento futuro de los márgenes y detectar posibles anomalías o situaciones de ineficiencia en el sistema.
- Proponer estrategias y recomendaciones orientadas a mejorar la transparencia en la cadena de suministro agroalimentaria. Estas propuestas estarán fundamentadas en los resultados del análisis y tendrán como objetivo reducir las desigualdades en la fijación de precios, asegurando un reparto más justo del valor generado entre los distintos eslabones de la cadena.

### 1.3. Impacto en sostenibilidad, ético-social y de diversidad

La desigualdad de precios repercute directamente en la **sostenibilidad económica** de las explotaciones familiares, limita su capacidad de innovar y debilita la resiliencia del tejido productivo rural. En el plano **social**, encarece la cesta básica, acelera la despoblación del campo y perpetúa brechas estructurales —por ejemplo, de género y de condiciones laborales precarias—. Desde la perspectiva **ambiental**, unos márgenes comerciales que no cubren los costes reales desincentivan la adopción de prácticas regenerativas y la inversión en adaptación climática, comprometiendo la salud y continuidad de los ecosistemas agrarios.

Estas repercusiones se inscriben en los **Objetivos de Desarrollo Sostenible (ODS)**[5] de la Agenda 2030, especialmente los relativos a hambre cero y agricultura sostenible, reducción de las desigualdades y producción y consumo responsables, que sirven aquí como marco de referencia para valorar la magnitud del problema.

Comprender estas dimensiones éticas, sociales y ambientales resulta fundamental para dimensionar adecuadamente la problemática y sienta la base sobre la que se definen los objetivos de este estudio, orientados a fortalecer una cadena de valor más justa y resiliente.

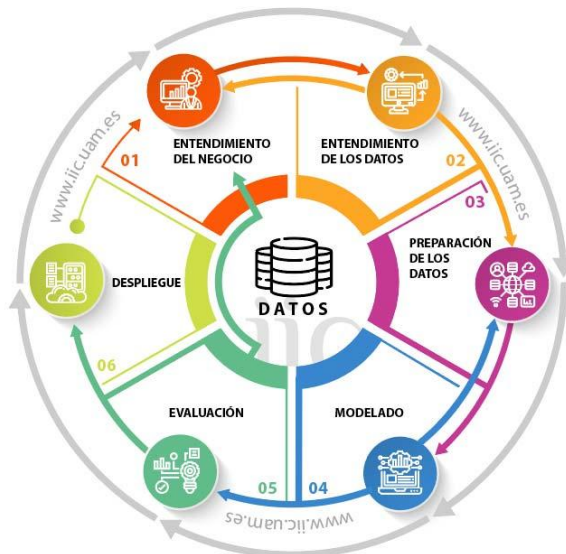
### 1.4. Enfoque y método seguido

Para comprender las causas de la brecha de precios entre el campo y el consumidor no basta con obtener los datos estáticos del IPOD; se necesita un **proceso iterativo** que acerque el análisis de datos a las preguntas de negocio y garantice que cada hallazgo vuelva a alimentar la toma de decisiones. Así se puede transformar los datos en información para luego generar conocimiento del ecosistema al que pertenecen y representa. Por ello se adoptó un **enfoque empírico-cuantitativo basado en datos abiertos**, articulado mediante el estándar **CRISP-DM**[6] (Cross-Industry Standard Process for Data Mining), cuyo ciclo se representa en la Ilustración 1.

1. **Comprensión del negocio** Se depuraron las preguntas clave: magnitud y tendencia de la brecha, productos más afectados y factores que la amplifican.
2. **Comprensión de los datos** Se inventariaron las fuentes públicas (Ministerio de Agricultura, Observatori de Preus Generalitat, COAG) y se evaluó su cobertura temporal y fiabilidad.
3. **Preparación de los datos** Se construyó un flujo ETL reproducible (extracción, limpieza, normalización y *fuzzy-matching* de nomenclaturas) que alinea series heterogéneas.
4. **Modelado** Se definieron las familias de técnicas (series temporales y *machine learning*) cuya implementación detallada se describe en el capítulo 2.
5. **Evaluación** Se contrastó la capacidad explicativa preliminar con criterios de negocio y métricas de error, retroalimentando ajustes en los pasos anteriores.
6. **Despliegue** Se documentaron notebooks, scripts y dashboards interactivos para que los resultados pudieran ser consumidos por perfiles técnicos y no técnicos.



Durante todo el ciclo se aplicaron principios transversales de reproducibilidad y ética: control de versiones, uso exclusivo de datos públicos y cumplimiento del RGPD[7]. Este marco metodológico aseguró que cada hallazgo fuera trazable y que el conocimiento generado resultara verificable y útil para orientar decisiones sobre la cadena agroalimentaria.



*Ilustración 1 Esquema del ciclo CRISP-DM[8]*

## 1.5. Planificación del Trabajo

La ejecución del proyecto se estructuró con las buenas prácticas del **PMBOK**[9] (Project Management Body of Knowledge), adaptándolas al plazo disponible de doce semanas y a la naturaleza iterativa del análisis de datos descrito en la sección 1.4. El resultado fue un plan realista, con controles regulares y márgenes de seguridad para absorber imprevistos sin comprometer la calidad.

Tabla 1 Cronograma – Elaboración propia

Duración		Hitos
<b>Fase 1 Inicio</b>	Semanas 1 a 3	<ul style="list-style-type: none"> <li>Definición de objetivos generales y específicos.</li> <li>Establecimiento del alcance del proyecto.</li> <li>Selección de productos agroalimentarios clave a analizar.</li> <li>Identificación y recopilación inicial de datos.</li> <li>Verificación preliminar de calidad y consistencia de datos.</li> </ul>
<b>Fase 2 Planificación</b>	Semanas 4 y 5	<ul style="list-style-type: none"> <li>Elaboración del plan detallado de actividades, técnicas y herramientas a utilizar.</li> <li>Creación del cronograma definitivo, ajustando prioridades según la complejidad de cada tarea.</li> <li>Selección y prueba inicial de las técnicas analíticas más adecuadas (ARIMA, regresiones, etc.).</li> </ul>
<b>Fase 3 Ejecución y Análisis</b>	Semanas 6 a 9	<ul style="list-style-type: none"> <li>Limpieza y preprocesamiento de datos</li> <li>Implementación de modelos de series temporales para analizar patrones históricos.</li> <li>Aplicación de algoritmos de machine learning para predicción de tendencias futuras.</li> <li>Enfoque iterativo <b>CRISP-DM</b>[6] (se seguirá esta metodología)</li> <li>Generación de resultados preliminares y primeras visualizaciones (gráficos, mapas de calor).</li> </ul>
<b>Fase 4 Seguimiento y Validación</b>	Semanas 10 y 11	<ul style="list-style-type: none"> <li>Revisión exhaustiva de resultados preliminares, asegurando su consistencia y validez.</li> <li>Ajuste y optimización de los modelos analíticos según las necesidades identificadas.</li> <li>Creación de la estructura preliminar del informe final.</li> </ul>
<b>Fase 5 Redacción y Presentación Final</b>	Semana 12	<ul style="list-style-type: none"> <li>Redacción del informe final, integrando resultados, visualizaciones y recomendaciones.</li> <li>Revisión general del documento (estructura, ortografía, coherencia).</li> <li>Preparación de la presentación final del trabajo.</li> </ul>

Para la **gestión de riesgos y calidad**, se identificaron cuatro riesgos críticos a priori:

- **Disponibilidad de datos**

Dado que los datos constituyen la base de todo el análisis, se realizó previamente un estudio preliminar sobre su accesibilidad. En un primer momento, tanto su disponibilidad como su calidad parecían estar garantizadas por la fiabilidad de las fuentes consultadas. No obstante, se contemplaron posibles soluciones como la interpolación lineal o el uso de fuentes alternativas en caso de necesidad.

Tras el análisis detallado, se constató que, si bien la disponibilidad de los datos fue adecuada, su calidad resultó ser inferior a lo esperado. Las fuentes carecían de metadatos y referencias bibliográficas claras, y los conjuntos presentaban una estructura poco consistente, con numerosos valores perdidos, erróneos o inválidos. Esto añadió complejidad a la generación de un dataset válido y a la posterior implementación de procesos ETL (Extract, Transform, Load), que resultaron ser más laboriosos de lo previsto.

- **Complejidad de los modelos**

El enfoque inicial se basaba en el análisis de series temporales, utilizando modelos estadísticos como ARIMA, que no requieren recursos computacionales elevados. Sin embargo, el volumen de datos y la diversidad de productos a analizar incrementaron significativamente la carga computacional.

Como medida preventiva, se reservó un periodo de tiempo adicional en la fase de ejecución y análisis, lo que permitió afrontar este riesgo sin comprometer la calidad de los resultados. Asimismo, se contempló la posibilidad de presentar una solución parcial o representativa en caso de que no se pudiera cubrir todo el conjunto de productos.

- **Retrasos en validaciones**

Se planificaron posibles retrasos derivados de validaciones intermedias o contratiempos imprevistos. Por ejemplo, el *apagón* inesperado alteró temporalmente el calendario de trabajo. No obstante, la planificación inicial ya contemplaba un margen de tiempo de reserva, y la flexibilidad ofrecida por la Universidad permitió reprogramar tareas clave sin afectar la entrega final. Este enfoque proactivo permitió mitigar el impacto de los retrasos y mantener el cronograma general del proyecto.

- **Errores de última hora en la memoria**

El riesgo de cometer errores en la versión final de la memoria se afrontó con una planificación detallada de las últimas etapas. No obstante, como ocurre en muchos proyectos, la presión de los plazos generó cierta tensión durante el cierre. Aunque se corrigieron la mayoría de las inconsistencias, esta experiencia resalta la importancia de anticipar aún más tiempo para la revisión en trabajos futuros.

La calidad se controló mediante **revisiones periódicas del código**, pruebas unitarias en las funciones ETL y verificación interna de la estructura de los datos. La consistencia de los datos es positiva dado sus enlaces permanentes y la no modificación de la estructura de cada uno de ellos.

La comunicación y coordinación con los tutores durante la realización del Trabajo ha sido constante mediante correos electrónicos o videoconferencia cuando ha sido necesario. Con este plan se mantuvo un equilibrio entre la flexibilidad necesaria en un proyecto de análisis de datos y la disciplina de la gestión clásica, asegurando que cada fase dispusiera de los recursos y el tiempo requeridos para cumplir los objetivos del TFG.

## 1.6. Breve resumen de productos obtenidos

Al cierre del proyecto se generó un conjunto de documentos que consolidan todo el trabajo técnico y facilitan su reutilización:

- Dataset integrado `df_final.csv`  
Conjunto de datos formado por aproximadamente 25 000 registros en un periodo de tiempo desde 2008 hasta 2024 con los precios de origen, mercado y destino procedentes de Ministerio de Agricultura, Mercabarna y COAG, así como el indicador *dif OD revisada* y la marca temporal normalizada; se exportó directamente desde el notebook para su distribución o futuros usos futuros.
- Notebook reproducible que documenta los flujos ETL completos (extracción, limpieza, *fuzzy-matching* y uniones externas), los análisis exploratorios y la aplicación secuencial de modelos. Es la base para auditorías o extensiones futuras.
- Conjunto de visualizaciones:  
Comparativas de precios origen-destino-mercado de algunos productos representativos. Series de diferencias porcentuales IPOD-Ministerio tanto en origen como en destino.
- Modelos predictivos ARIMA/SARIMA para los nueve productos con datos completos, con gráficos de ajuste in-sample y scripts que automatizan la selección de parámetros mediante *auto-arima*.
- Listados de anomalías (productos y fechas con saltos de precio atípicos) y funciones de comparación gráfica que facilitan su inspección manual

## 2. Materiales y métodos

En este apartado se detallan los procesos realizados para obtener los datos y la aplicación de técnicas estadísticas y modelos de análisis de datos y así generar los resultados que nos permitirán obtener las conclusiones a los planteamientos iniciales.

Los procesos que generan este trabajo se han creado íntegramente en un **Jupyter Notebook**, de modo que cualquier lector pueda rehacer cada paso con solo ejecutar las celdas en el orden indicado. El uso de Jupyter facilita la **reproducibilidad** (código y comentario juntos), la portabilidad entre sistemas operativos y la inserción directa de gráficos exploratorios.

### 2.1 Fuentes de datos

Los datos necesarios para realizar este trabajo se obtienen de diferentes bases de datos públicas. Gracias a la implantación de las políticas de transparencia en la información, se puede acceder a gran cantidad de datos públicos de manera abierta y gratuita. En este caso los organismos que publican la información son:



Ilustración 2 Fuentes de datos (elaboración propia)

Estas bases de datos proporcionan información histórica sobre precios en origen y destino, así como indicadores relevantes sobre las condiciones del mercado.

Los datos disponibles son Opensource (gestionados por administraciones públicas) y no disponen de herramienta para poder acceder a ellos de manera automática (API). Dada esta situación, el primer paso del ETL; la extracción de la información ha sido más complicada. Requiere más tiempo y una estructura adecuada para su posterior transformación y carga.

## 2.2 Preparación y fusión de datos (pipeline ETL)

El conjunto de datos sobre el que se construyeron los análisis se generó mediante un proceso ETL reproducible programado íntegramente en un Jupyter Notebook de Python. El núcleo del proceso se desarrolló en Python 3.11 con las librerías *pandas*, *numpy*, *requests* + *BeautifulSoup* (descarga y *scraping* de páginas sin API), *rapidfuzz* (alineamiento difuso de nombres) y *matplotlib* / *seaborn*.

La etapa de extracción ha sido particularmente exigente, ya que ninguno de los organismos consultados publica ficheros aptos para análisis. El IPOD de COAG se difunde mensualmente en boletines **PDF**, mientras que el Ministerio de Agricultura y Mercabarna ofrecen hojas **Excel** elaboradas como resúmenes informativos con celdas combinadas, logotipos e indicadores de formato que impiden la carga directa en *pandas*.

Para salvar esta heterogeneidad se desarrollaron **funciones específicas para cada fuente** dentro del notebook:

- Boletines PDF — se procesaron con *pdfplumber*, que permitió leer página a página y localizar las tablas mediante coordenadas relativas; los encabezados se normalizaron y los valores se volcaron en listas intermedias antes de convertirse en *DataFrame*.
- Plantillas Excel — se cargaron con *openpyxl* (*load\_workbook*) para acceder a la hoja que contiene los precios, ignorando celdas de formato y fórmulas. Se recorrieron filas y columnas con un mapeo explícito de coordenadas, reconstruyendo un esquema tabular estándar (producto, año, mes, precio)

Una vez extraídos, los datos pasaron a *pandas* para las transformaciones siguientes. Se unificaron codificaciones y se corrigieron erratas (como, por ejemplo: “MELONE”: “MELÓN”), se normalizaron unidades y se armonizaron los nombres de producto mediante **RapidFuzz** (umbral de similitud del 75 %). Esto permite solventar una parte importante de errores tipográficos.

Con los vocabularios ya armonizados, los tres orígenes de información se integraron en dos **union** sucesivas de tipo **outer** para maximizar la retención de registros:

1. IPOD  $\otimes$  Mercabarna.  
*IPOD* aporta el precio en origen y en destino, mientras que *Mercabarna* describe la cotización mayorista (mercado intermedio). La unión se realizó sobre la clave (PRODUCTO, YEAR, MES). Al ser *outer*, si un producto-mes aparecía sólo en una de las fuentes quedaba igualmente representado y el campo ausente se marcaba como NaN. Esto fue esencial para no perder meses en los que, por ejemplo, Mercabarna interrumpió la publicación de un producto pero COAG sí reportó su IPOD .

## 2. (IPOD ⊗ Mercabarna) ⊗ Ministerio Agricultura.

El dataframe resultante se volvió a unir, también con *outer join*, con la tabla del Ministerio de Agricultura, que ofrece precios de origen y destino calculados con una metodología distinta. Mantener este segundo par de observaciones permitió, más tarde, comparar y validar la coherencia entre fuentes. Para resolver duplicidades se aplicó la siguiente regla:

- Si ambos organismos reportaban el mismo producto-mes, se conservó la media simple de los dos valores.
- Si únicamente uno aportaba dato, se heredó sin modificación y se añadió un flag (`fuentes_minist = 0/1`) que indica su procedencia.

La elección de uniones *outer* (en lugar de *inner join*) respondió a un criterio de cobertura temporal: hubiera sido preferible perder menos del 6 % de filas antes que descartar meses completos por un hueco puntual. Tras la fusión, el dataframe conservó **todas las observaciones disponibles entre 2008 y 2024**, con etiquetas de procedencia que más tarde facilitaron los análisis de consistencia y la imputación selectiva de valores faltantes.

Una vez integradas las tres fuentes, se añadió al dataframe combinado un indicador nuevo que resume la brecha de precios y una marca temporal homogénea: **dif\_OD\_revisada y dif%\_revisada**. Se definió como el cociente entre el precio medio de venta al consumidor (DESTINO) y el precio medio recibido por el agricultor (ORIGEN) para cada producto (*i*) y mes (*t*) y su versión porcentual:

*Ecuación 1 dif O-D*

$$\text{dif OD revisada}_{i,t} = \frac{\text{DESTINO}_{i,t}}{\text{ORIGEN}_{i,t}}$$

*Ecuación 2 dif % O-D*

$$\text{dif \% revisada} = (\text{dif OD revisada}_{i,t} - 1) * 100$$

Los registros con ORIGEN = 0 se descartaron (< 0.2 % *del total*) para evitar divisiones indeterminadas, y cualquier valor ausente se conservó como *NaN*, garantizando la trazabilidad de puntos con información incompleta.

Estas dos métricas cumplían un doble propósito. Primero, permitieron verificar la coherencia de los valores publicados por COAG mediante la comparación con las series del Ministerio; segundo, proporcionaron una variable estandarizada—en forma de razón y en forma porcentual—idónea para los **análisis estadísticos posteriores** (pruebas de estacionariedad, modelado SARIMA, contraste de eventos), asegurando que todos los cálculos se basaran en un indicador homogéneo y directamente interpretable.

Finalmente, el conjunto depurado se exportó como **df\_final.csv**.



## 2.3 Variables de análisis

El fichero *df\_final.csv*, que constituye el punto de partida de todos los análisis posteriores, agrupa algo más de veinticinco mil observaciones comprendidas entre enero de 2008 y diciembre de 2024. Cada registro identifica, sin ambigüedades, un producto y un mes concretos; para ello se conserva la descripción normalizada del artículo junto con los campos de año, mes y una fecha ISO que simplifica cualquier operación de series temporales.

Sobre esa plantilla temporal se proyectan tres valores fundamentales: el precio que recibe el agricultor en origen, la cotización mayorista de Mercabarna y el importe final que paga el consumidor. Todas las cifras se expresan en euros por kilogramo y se han armonizado (si coinciden dos se guarda la media y, etiqueta que señala su procedencia), de modo que las comparaciones sean consistentes a lo largo de toda la serie.

A partir de esos precios básicos se incorporan dos métricas explicadas en el apartado anterior (la razón destino-origen, que cuantifica la brecha absoluta, y su versión porcentual, que facilita la interpretación inmediata de las diferencias). Por último, cada fila recibe una etiqueta de contexto: normal, COVID o sequía. Esto permitirá comprobar, más adelante, si los episodios de choque alteraron de manera significativa la relación de precios.

Esta combinación de identificadores precisos, precios homogéneos e indicadores derivados ofrece una base suficientemente rica para describir la evolución de la brecha, alimentar los modelos de predicción y, llegado el caso, contrastar la influencia de factores externos sin introducir redundancias ni cálculos superfluos que ya se explicaron en el epígrafe anterior.



## 3. Modelado y análisis

### 3.1 Verificación inicial

Como primer punto del análisis se procede a validar la **calidad de los datos**. Este proceso se realiza mediante el cálculo de la diferencia porcentual del valor del precio del Ministerio respecto al valor IPOD, que es el valor de referencia. Esto se calcula por un lado para el valor (precio) de Origen:

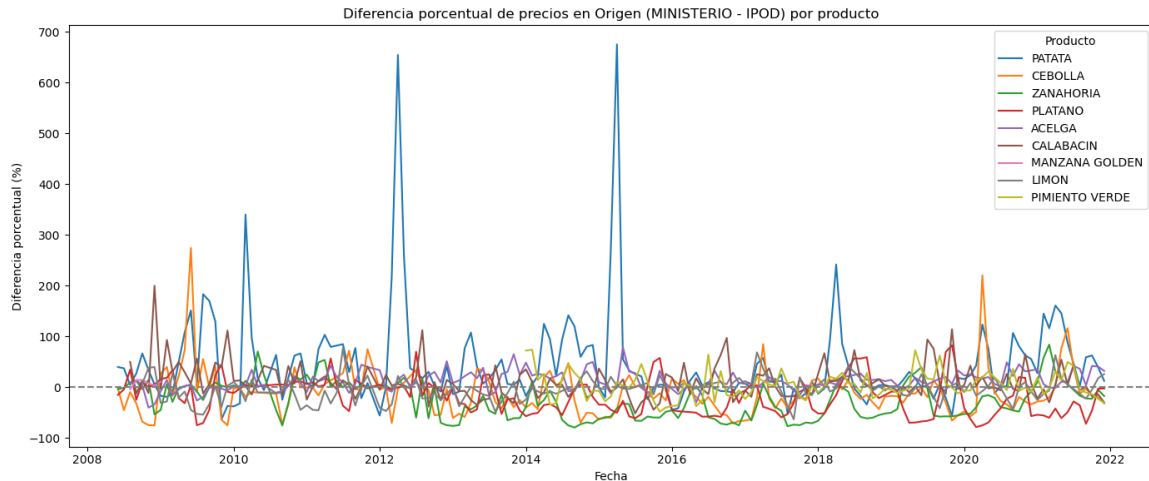


Ilustración 3 Diferencias precios Origen

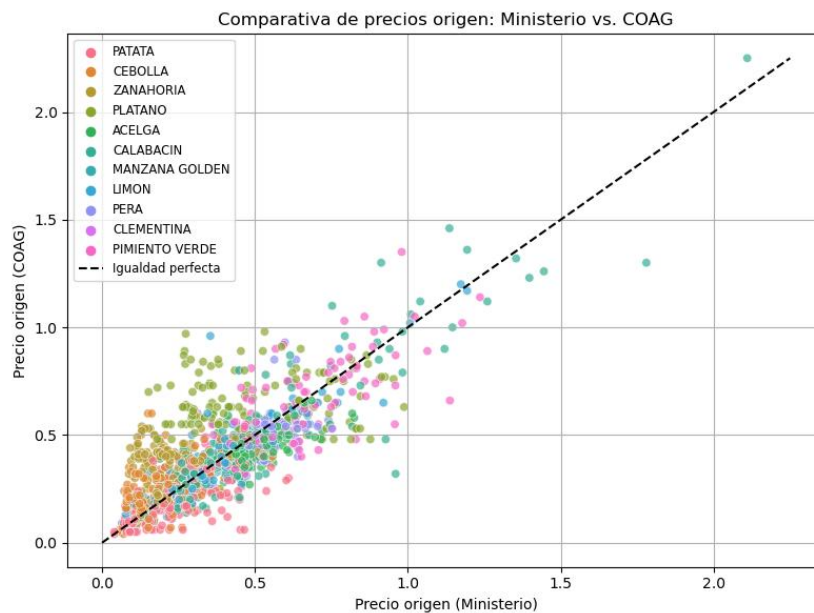
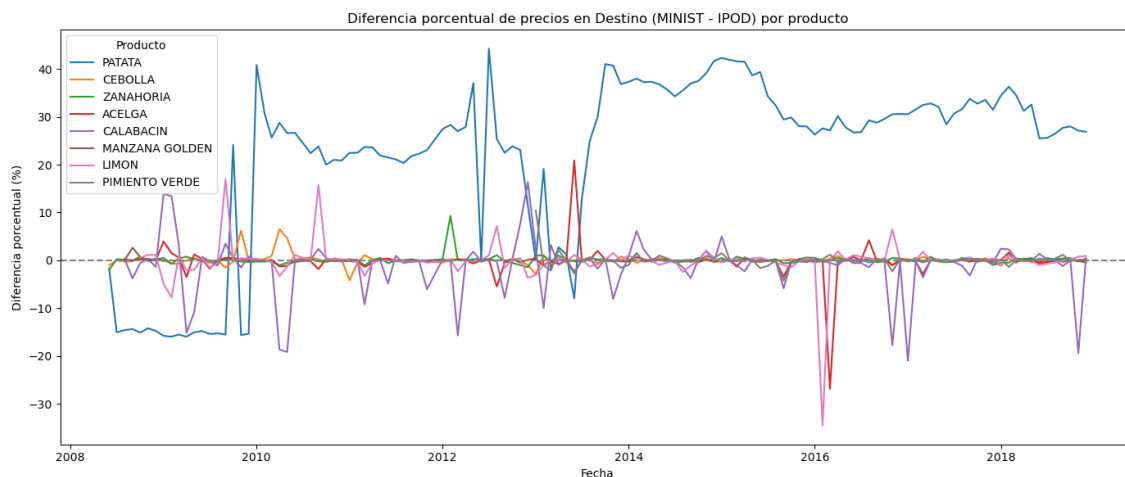
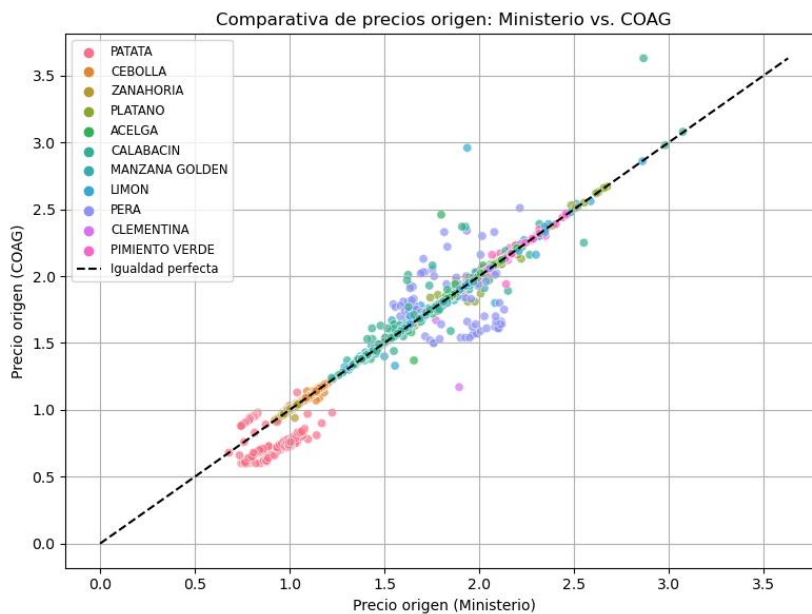


Ilustración 4 Diferencias precios Origen Scatterplot

Y posteriormente con el valor de Destino:



*Ilustración 5 Diferencias precio Destino*



*Ilustración 6 Diferencias precio Destino Scatterplot*

Estas gráficas confirman que, en términos generales, las dos series describen la misma realidad. En los precios de origen las curvas de la mayoría de los productos oscilan muy cerca del eje cero, de manera que las discrepancias permanecen reducidas durante la mayor parte del periodo 2008-2024. Las excepciones se concentran en episodios muy concretos, visibles como picos pronunciados en la patata (y, en menor medida, en la cebolla y la alcachofa) durante los inviernos de 2010, 2013 y 2016. Esos sobresaltos se explican por la conjunción de valores inusualmente bajos en el IPOD y fuertes tensiones estacionales del mercado, de modo que una pequeña diferencia absoluta se tradujo en un porcentaje desproporcionado.

La validación con el precio de destino refuerza esta lectura. Salvo la patata, cuyos desvíos indican un caso especial, este producto requiere un trato extraordinario. El sesgo, por tanto, no es estructural ni sistemático, sino circunscrito a un reducido número de productos de elevada volatilidad o metodologías de captura específicas. Este hecho respalda la decisión metodológica de trabajar con ambas fuentes fusionadas y de promediar sus valores siempre que coincidan, manteniendo a la vez un indicador de procedencia que permita, en análisis posteriores, evaluar la sensibilidad de los resultados a la presencia de los datos ministeriales.

En síntesis, la prueba preliminar avala la calidad global del conjunto y señala únicamente unos pocos meses y productos que exigirán precaución adicional: bien aplicando filtros de suavizado, bien excluyéndolos de modelos especialmente sensibles a los valores extremos. De este modo, la base numérica sobre la que se construyen los modelos puede considerarse, en general, homogénea y fiable.

## 3.2 Análisis temporal diferencia Origen-Destino



Ilustración 7 Diferencias Origen - Mercado - Destino

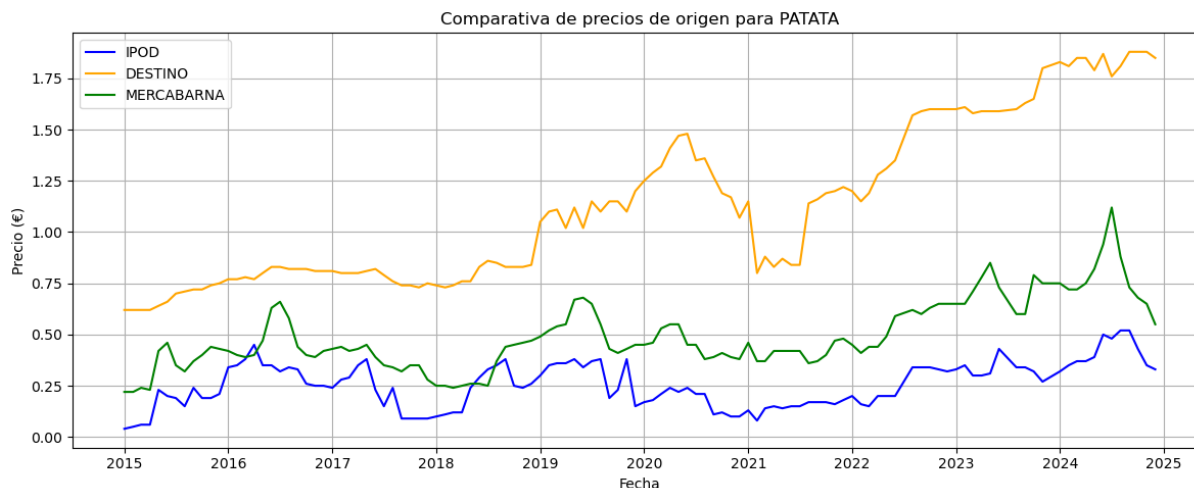
Las seis series comparan, para cada producto, la evolución mensual del precio que recibe el agricultor (línea azul, IPOD-origen), el precio mayorista en Mercabarna (verde) y el que paga el consumidor (naranja). Todas están expresadas en euros por kilogramo y cubren de 2015 a 2025.

En el conjunto de los gráficos se aprecia, ante todo, la **jerarquía esperada**: el precio al consumidor se mantiene de forma sistemática por encima del mayorista y éste, a su vez, supera al de origen. No obstante, la **distancia entre niveles y la forma de los movimientos** difiere según la naturaleza del cultivo.

*En cebolla, zanahoria y acelga* predomina un **patrón de inflación gradual**. El precio al consumidor arranca en torno al euro y medio y sobrepasa los dos euros en 2024, mientras que el mayorista sube con cierto retardo y el origen permanece casi plano durante buena parte del periodo. Ese divorcio progresivo sugiere un **aumento del margen comercial que no se explica por costes coyunturales** sino, más bien, por una renegociación estructural de la cadena de valor: el productor apenas captura la tendencia alcista que sí trasladan los escalones intermedios.

*En calabacín, limón y pimiento verde* la lectura cambia: la línea mayorista (y en menor medida la de origen) presenta **picos muy pronunciados** que responden a la estacionalidad y a la sensibilidad climática de estos cultivos. La presencia de esas crestas también en el precio al consumidor indica que, en episodios de oferta corta, la traslación es rápida: el sistema reacciona al alza y el diferencial entre niveles se mantiene relativamente constante. Con todo, se observa que a partir de 2021 los máximos del calabacín y del pimiento se sitúan un peldaño por encima de los años previos, lo que apunta a un **nuevo régimen de precios altos** desencadenado, en parte, por la sequía de 2022-23.

Por último, destaca la **patata** que exhibe la brecha más pronunciada del conjunto. Entre 2015 y 2025 el precio al consumidor pasa de  $\approx 0,65$  €/kg a rozar los 2 €, mientras el agricultor apenas supera los 0,50 € al final del periodo. El mayorista marca picos estacionales que se trasladan casi de inmediato al destino, pero apenas llegan al origen, de modo que el margen se amplía ciclo tras ciclo. El patrón confirma una transmisión de precios asimétrica y un poder de fijación claramente volcado hacia los eslabones intermedios de la cadena.



Il·lustració 8 Precio Patata O-M-D

En conjunto, las figuras muestran dos conclusiones provisionales:

1. en los productos menos estacionales (cebolla, zanahoria, acelga) la brecha destino-origen se ha ido incrementando de manera paulatina;
2. en los cultivos muy estacionales (calabacín, limón, pimiento) el margen se ensancha o se estrecha al compás de la oferta, pero sin llegar a romper la proporcionalidad entre niveles, salvo en los episodios extremos de 2022-24.

Las gráficas muestran que el precio al consumidor se distancia de forma persistente (creciente) del que percibe el agricultor, lo cual sugiere un reparto **desigual** del valor y, por extensión, una posible concentración de poder en los tramos intermedios.

Tabla 2 Diferencia OD máxima

	PRODUCTO	dif OD revisada
173	PIMIENTO CALIFORNIA VERDE	14.867000
156	PATATA (Temprana)	10.429000
135	MELON PIEL DE SAPO	9.653000
58	COL	8.850750
172	PIMIENTO CALIFORNIA ROJO	8.567000
115	LIMON FINO	7.852000
56	CLEMENTINA	7.603827
141	NARANJA TIPO NAVEL	7.455484
9	ACEITUNA EN VERDE	7.341875
114	LIMON	6.969925

La tabla recoge los casos extremos en los que se ha multiplicado el precio. Encabeza la lista el pimiento California verde: un coeficiente de **14,9** implica que, en el mes medido, el importe final multiplicó casi por quince lo que cobró el agricultor. Estos datos revelan una pérdida de poder de negociación en el primer eslabón de la cadena. Para estos artículos concretos el margen comercial absorbido por intermediarios y distribución resulta excepcionalmente alto y, por tanto, merecen un análisis pormenorizado en las pruebas de transmisión de precios y en la discusión sobre posibles intervenciones de política agraria o de competencia.

### 3.3 Modelos de predicción

El ajuste preliminar de un modelo ARIMA[10] se concibió como un punto de partida común para los nueve cultivos incluidos en la muestra (representativa). Por un lado, todas las series comparten la misma periodicidad y un margen temporal suficiente para estimar la dinámica propia de cada producto; por otro, el modelo autorregresivo integrado constituye la referencia más sencilla cuando se desconoce de antemano la posible presencia de raíces unitarias, tendencias o estructuras de medias móviles[11].

Para garantizar la comparabilidad, se fijó el mismo orden (1, 1, 1). Con ello se obligó a todas las series a someterse a una única diferencia de primer grado (necesaria para estabilizar la media en la mayoría de los cultivos) y se dejó a la interacción entre un término autorregresivo y otro de media móvil la tarea de capturar la inercia inmediata. Se optó conscientemente por esta **parametrización mínima** antes de explorar variantes con componentes estacionales o exógenas, de modo que cualquier mejora posterior pudiera atribuirse sin ambigüedad a la nueva especificación.

Las gráficas comparan los valores observados (línea azul) con la trayectoria que reproduce el ARIMA (línea roja discontinua) dentro de la muestra. El resultado confirma que, aun con una estructura tan sobria, el modelo consigue seguir la pauta de fondo en todos los productos: las tendencias ascendentes de la acelga y del calabacín, la caída prolongada de la cebolla tras su pico de 2013 o las oscilaciones de la zanahoria quedan reflejadas con escaso rezago.

Cuando aparecen discrepancias significativas, como los picos extremos del limón o los sobresaltos anuales del pimiento verde, el error se concentra en los valores máximos, lo que sugiere que la varianza condicional crece durante los episodios de oferta escasa y que, llegado el caso, un término estacional o un modelo con heterocedasticidad cambiaría la foto.

La aplicación de este modelo cumple, por tanto, tres objetivos: demuestra que los datos resultantes del proceso ETL son modelizables sin transformaciones adicionales; ofrece una primera medida de ajuste que servirá de referencia frente a variantes más complejas (SARIMA, ARIMAX, modelos no lineales); y ayuda a identificar, cultivo a cultivo, los meses en los que la brecha registra saltos exagerados, facilitando así el análisis de anomalías y la posterior evaluación de la transmisión de precios.



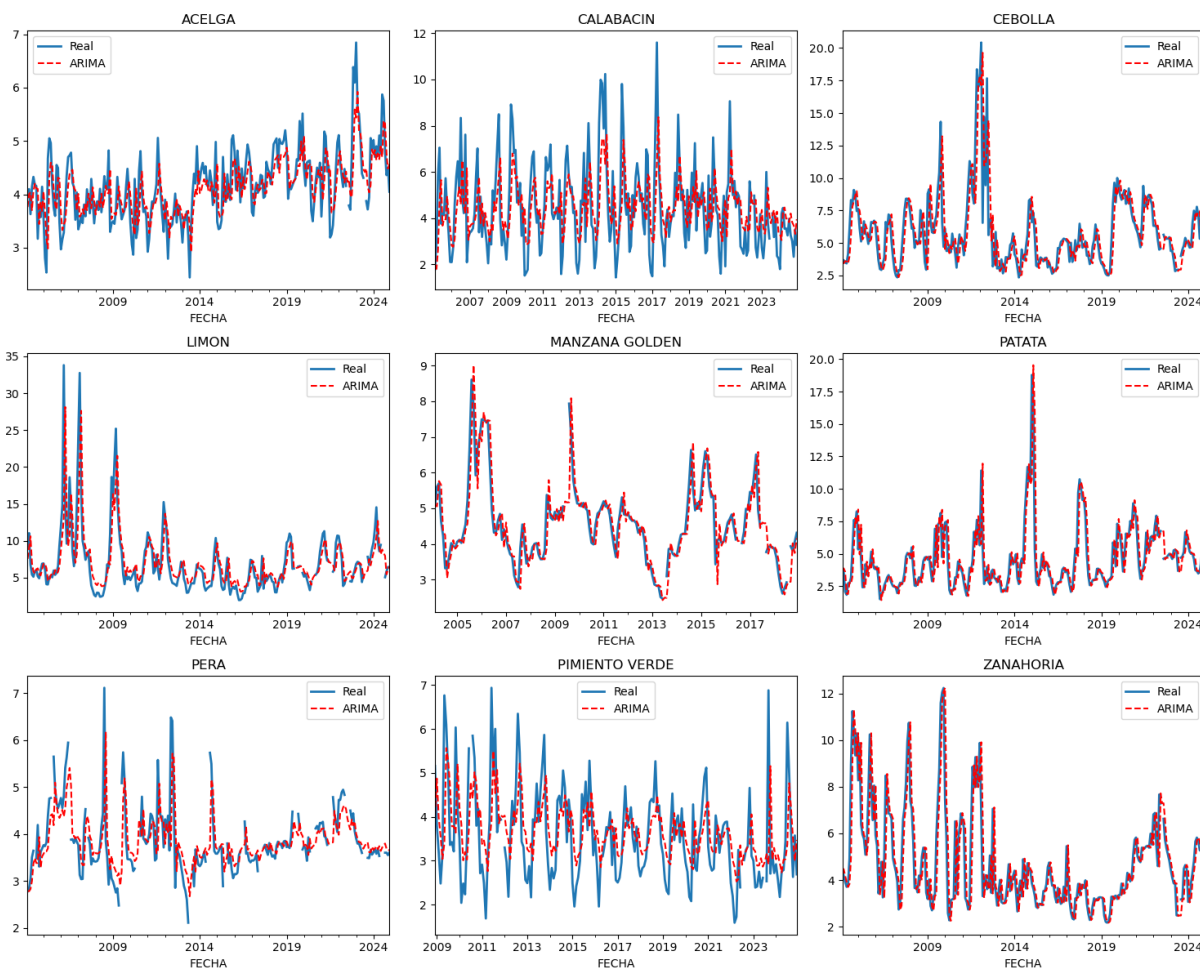


Ilustración 9 Resultados ARIMA

A la luz de estos resultados de escasa utilidad, se aplica la estimación automática seleccionada por *auto\_arima* (con estacionalidad mensual ( $m = 12$ ) y parámetros elegidos según el AIC) reproduce con notable fidelidad la trayectoria de la patata a lo largo de toda la serie. El trazado rojo, aunque suaviza el gran repunte de 2015 y los máximos de 2018-2020, acompaña la tendencia ascendente y el patrón de picos anuales que caracteriza al cultivo; el modelo, por tanto, capta la estacionalidad y la inercia sin necesidad de definir manualmente los órdenes autorregresivos y de medias móviles.

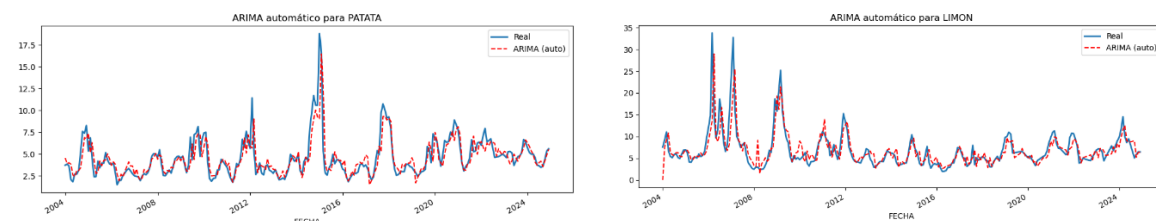


Ilustración 10 Resultados Auto ARIMA

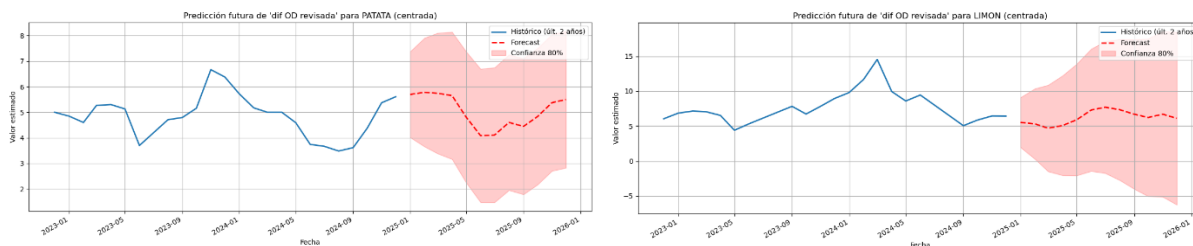
Una vez ajustado el modelo `auto_arima`, se procedió a la generación de predicciones a 12 meses vista. La predicción se realizó sobre la base del mejor modelo identificado según los criterios AIC y BIC, permitiendo tanto la detección de estacionalidad como el ajuste de los parámetros autoregresivos, de diferencia e integración ( $p$ ,  $d$ ,  $q$ ).

En el caso concreto de la **patata y el limón**, el modelo capturó correctamente las fluctuaciones observadas en los últimos dos años y generó una previsión para los próximos doce meses, acompañada de un intervalo de confianza del 80%. El resultado se visualizó gráficamente, mostrando:

- La evolución histórica reciente (últimos 24 meses).
- La proyección de valores futuros (línea discontinua).
- La banda de incertidumbre, representando el margen de error de la estimación.

Este procedimiento permitió no solo anticipar posibles comportamientos de la variable en el futuro, sino también evaluar el nivel de incertidumbre asociado a cada predicción. En los casos en los que el modelo original generó valores inválidos, se incorporó un mecanismo de reintento con parámetros más conservadores (sin estacionalidad), garantizando así la robustez del análisis.

Las figuras reflejan una previsión relativamente estable, aunque con una dispersión creciente en los meses centrales del horizonte temporal, lo cual sugiere cautela a la hora de interpretar los valores puntuales de predicción.



Il·lustració 11 Proyecciones ARIMA

### 3.4 Análisis de la evolución temporal diferencia Origen-Destino

El siguiente gráfico representa la evolución mensual de la diferencia revisada entre precios de origen y destino (O-D) para cinco productos agrícolas seleccionados: **calabacín, cebolla, limón, patata y pimiento verde**, en el periodo comprendido entre 2010 y 2025.

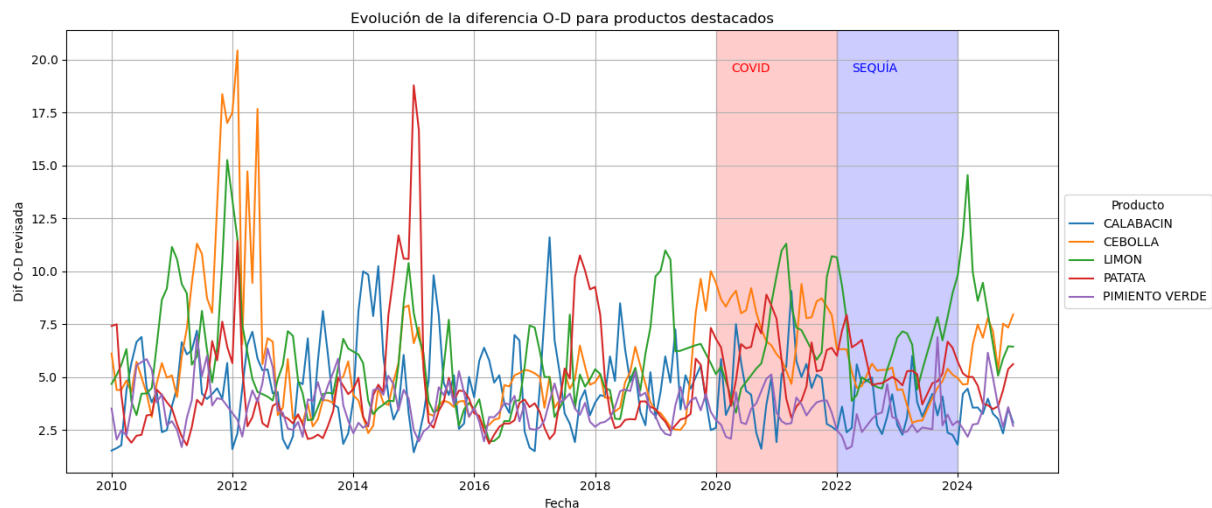
Cada línea de color refleja la variabilidad de esta diferencia para un producto concreto, permitiendo observar comportamientos diferenciados a lo largo del tiempo. Algunos productos presentan mayor inestabilidad o picos puntuales, mientras que otros mantienen una evolución más contenida, lo cual puede reflejar diferencias en estructura de mercado, estacionalidad o vulnerabilidad a perturbaciones externas.



En la parte superior del gráfico se han destacado dos franjas temporales que corresponden a eventos con alto impacto en los precios agroalimentarios:

- **COVID-19 (2020–2021):** marcó alteraciones severas en transporte, demanda y distribución.
- **SEQUÍA (2022–2023):** afectó directamente a la producción y disponibilidad de ciertos cultivos.

Este tipo de representación permite identificar de forma visual correlaciones temporales entre eventos externos y aumentos en las desigualdades de precios a lo largo de la cadena alimentaria, sirviendo como base para análisis más detallados o propuestas de intervención.



*Ilustración 12 Evolucion temporal Dif O-D eventos*

Siempre se pueden añadir más elementos que quizás hacen fluctuar los precios, como sequías o lluvias torrenciales locales, el precio del combustible o conflictos armados. El resumen general en este apartado, como se observa en las gráficas, es que la variación de los precios (excepto en momentos puntuales asociados a crisis concretas) tiende a mantenerse dentro de un rango relativamente estable.

Salvo episodios como la pandemia de COVID-19 o la reciente sequía prolongada, que generaron distorsiones significativas en algunos productos, el comportamiento general revela una cierta regularidad estacional y una resiliencia relativa del sistema de precios.

En un mercado globalizado, la disponibilidad casi permanente de productos (incluso fuera de temporada) debido a la importación, podría hacer pensar que los precios deberían presentarse más estables a lo largo del año. Sin embargo, esta estabilidad no siempre se refleja en la diferencia entre los precios de origen y destino.

Esto se debe a varios factores: los productos importados no siempre compiten en igualdad de condiciones (por calidad, costes logísticos o acuerdos comerciales), y muchas veces los márgenes se ajustan en la distribución y no en el precio pagado al agricultor local. Además, la percepción de escasez o de “fuera de temporada” en determinados productos sigue influyendo en el comportamiento de los precios minoristas, independientemente de la disponibilidad real.

En este sentido, el efecto de la globalización sobre los precios resulta parcial, amortiguando las subidas extremas, pero sin corregir las desigualdades estructurales en la cadena alimentaria.

### 3.5 Evolución precio Origen-Destino por producto

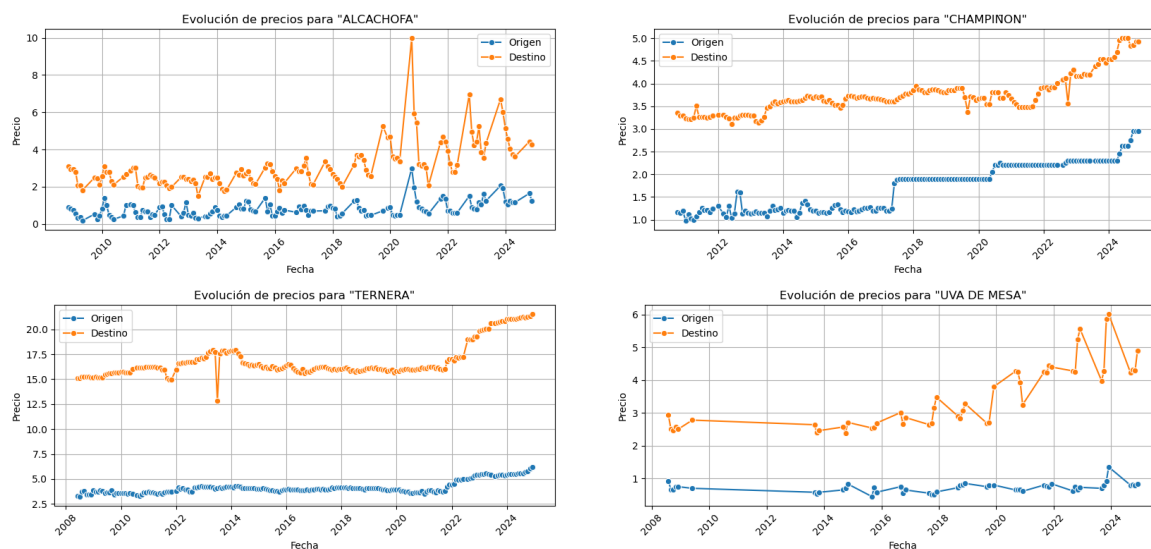


Ilustración 13 Precio O-D por producto

La figura muestra la evolución comparada de los precios en origen y en destino para cuatro productos agroalimentarios: alcachofa, champiñón, ternera y uva de mesa. En todos los casos, se observa una constante estructural: el precio de destino es **sistemáticamente superior** al precio de origen, con una brecha que, lejos de reducirse con el tiempo, tiende a mantenerse o incluso ampliarse en algunos casos.

En general, los precios en origen presentan una evolución mucho más estable y contenida, lo que sugiere que los agricultores reciben ingresos relativamente constantes, sin beneficiarse de manera proporcional de las subidas de precio que sí experimenta el consumidor final. Esta asimetría en la transmisión de precios es especialmente visible en productos como la carne de ternera y el champiñón, donde el precio de destino sigue una tendencia claramente ascendente, mientras que el precio en origen apenas varía durante largos periodos.

Además, en algunos productos como la alcachofa o la uva de mesa, se observa una alta volatilidad en los precios de destino, con picos puntuales y oscilaciones abruptas, frente a un origen mucho más lineal. Esta disparidad sugiere que las variaciones de precios en el mercado no siempre responden a los costes de producción, sino que están influenciadas por factores como la intermediación, la estacionalidad, el posicionamiento comercial o incluso dinámicas especulativas en la distribución.

En conjunto, la imagen ilustra de forma clara cómo, pese a operar en un entorno globalizado y tecnológicamente avanzado, las cadenas de valor agroalimentarias siguen presentando importantes desequilibrios. Estos datos refuerzan la necesidad de políticas que fomenten una mayor transparencia en la formación de precios y un reparto más equitativo del valor entre los distintos eslabones de la cadena, especialmente en defensa de los productores primarios.

## 4. Conclusiones y trabajos futuros

El presente Trabajo de Fin de Grado ha permitido analizar de manera rigurosa y empírica la desigualdad de precios existente en la cadena agroalimentaria española, centrando el estudio en los productos con mayores diferencias entre el precio en origen y el precio en destino. A partir de datos abiertos provenientes de distintas fuentes institucionales (COAG, MAPA, Generalitat), se construyó un conjunto integrado y reproducible sobre el que se aplicaron técnicas de análisis temporal, modelado predictivo y visualización.

Los resultados obtenidos confirman la existencia de una **brecha estructural persistente** entre el precio recibido por el productor y el precio final pagado por el consumidor. Esta brecha no solo se mantiene en el tiempo, sino que tiende a incrementarse en productos poco estacionales como la cebolla, la zanahoria o la acelga, donde los márgenes se ensanchan gradualmente sin que los agricultores se beneficien de la subida en destino. En cultivos más sensibles a factores climáticos o de mercado —como el calabacín, el pimiento verde o el limón— la brecha se comporta de forma más elástica, pero igualmente revela una **asimetría en la transmisión de precios**, especialmente en episodios extremos como la sequía de 2022–2023.

El análisis conjunto de los precios en origen, mercado y destino ha permitido **evidenciar patrones diferenciados según la naturaleza del producto**, así como identificar casos especialmente llamativos como el del pimiento California verde, cuya diferencia llegó a multiplicar por quince el precio en origen. Esta situación refleja una **pérdida de poder de negociación** del primer eslabón de la cadena y la concentración del valor añadido en tramos intermedios, lo cual justifica la necesidad de intervención tanto desde el ámbito público como desde prácticas empresariales más justas.

En el plano metodológico, la aplicación de modelos ARIMA y auto\_arima ha demostrado que las series resultantes del proceso ETL son suficientemente estables y modelizables. Aunque los modelos simples logran capturar las tendencias principales y replicar con fidelidad los patrones estacionales de cultivos como la patata, las limitaciones aparecen al tratar de modelar picos extremos o eventos abruptos. La predicción a 12 meses vista proporciona un horizonte útil para la anticipación de desigualdades futuras, aunque los **intervalos de confianza amplios** en ciertos productos invitan a interpretar los resultados con cautela. Desde un punto de vista práctico, este trabajo ofrece una herramienta objetiva para la toma de decisiones en materia de política agraria, transparencia comercial y defensa del consumidor. La información derivada de los análisis —anomalías detectadas, productos críticos y periodos de mayor tensión— puede orientar estrategias de intervención o regulación de márgenes comerciales.

No obstante, deben señalarse algunas **limitaciones** que condicionan la interpretación de los resultados. La calidad de los datos, aunque suficiente para el análisis, presenta deficiencias estructurales: ausencia de metadatos, falta de estandarización y pérdida parcial de registros históricos. Asimismo, no se han incorporado variables contextuales de carácter económico (PIB, inflación general), productivo (volumen comercializado, superficie cultivada) o climático (precipitaciones, temperatura), que sin duda enriquecerían el modelado. Por razones de alcance y viabilidad, tampoco se ha aplicado inferencia causal ni se ha explorado la elasticidad de precios ante cambios externos.

Como líneas de trabajo futuro se plantea la posibilidad de:

- **Ampliar el conjunto de variables** con indicadores económicos y climáticos que permitan modelar dinámicas de precios más complejas;
- **Aplicar modelos multivariantes** o causales (ARIMAX, VAR, Prophet o modelos estructurales) que exploren la relación entre factores exógenos y variaciones de precios;
- **Analizar con mayor granularidad territorial** la evolución de precios en comunidades autónomas o regiones específicas;
- **Integrar modelos de clasificación o clustering** para tipificar productos según su nivel de desigualdad estructural.

En definitiva, este trabajo contribuye a generar evidencia empírica que visibiliza una problemática compleja y estructural en la cadena agroalimentaria. Su valor reside tanto en la construcción metodológica como en la aplicabilidad de los resultados a distintos niveles de toma de decisión, desde el diseño de políticas públicas hasta la sensibilización del consumidor.

## 5. Glosario

### **Agroalimentario**

Conjunto de actividades económicas relacionadas con la producción, transformación, distribución y comercialización de productos agrícolas y alimentarios.

### **ARIMA (AutoRegressive Integrated Moving Average)**

Modelo estadístico de series temporales que combina componentes autorregresivos, de diferenciación y de media móvil. Se emplea para capturar patrones y realizar predicciones en variables que presentan autocorrelación temporal.

### **Auto\_ARIMA**

Versión automatizada del modelo ARIMA que selecciona de forma óptima los parámetros del modelo (p, d, q) y su posible estacionalidad mediante criterios de información como AIC o BIC.

### **Brecha Origen-Destino (O-D)**

Diferencia entre el precio que recibe el productor agrícola (origen) y el que paga el consumidor final (destino). Se utiliza como indicador de desigualdad en la cadena de valor.

### **COAG (Coordinadora de Organizaciones de Agricultores y Ganaderos)**

Organización profesional agraria que publica, entre otros indicadores, el Índice de Precios en Origen y Destino (IPOD), utilizado como referencia en este trabajo.

### **CRISP-DM (Cross-Industry Standard Process for Data Mining)**

Metodología estándar para la realización de proyectos de análisis de datos, estructurada en seis fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue.

### **Dataset**

Conjunto estructurado de datos utilizado como base para análisis y modelado. En este trabajo se hace referencia a df\_final.csv, resultado del proceso ETL aplicado sobre fuentes oficiales.

### **ETL (Extract, Transform, Load)**

Proceso técnico que comprende la extracción de datos desde distintas fuentes, su transformación para normalizar y corregir errores, y su carga en un entorno preparado para el análisis.

### **Fuzzy Matching**

Técnica utilizada para emparejar cadenas de texto que no son exactamente iguales pero son similares (por ejemplo, nombres de productos con errores tipográficos), mediante medidas de similitud textual.

### **IPOD (Índice de Precios en Origen y Destino)**

Indicador que mide la relación entre el precio pagado por el consumidor y el recibido por el agricultor, calculado y publicado por COAG mensualmente. Se usa como herramienta para evaluar la equidad de la cadena agroalimentaria.

### **Mercabarna**

Mercado mayorista de Barcelona que publica cotizaciones de productos agroalimentarios. Se utiliza como fuente intermedia (nivel mercado) entre productor y consumidor.

### **Modelo predictivo**

Herramienta analítica que permite estimar valores futuros a partir de patrones pasados en los datos. En este trabajo se aplican modelos ARIMA a series temporales de precios.

### **Outer Join**

Tipo de unión entre tablas de datos que conserva todos los registros de ambas tablas, independientemente de que coincidan en la clave. Se utiliza en este trabajo para no perder observaciones relevantes.

### **PMBOK (Project Management Body of Knowledge)**

Conjunto de buenas prácticas en gestión de proyectos definido por el Project Management Institute (PMI). Se utiliza como referencia para planificar y estructurar el trabajo por fases.

### **RapidFuzz**

Librería de Python utilizada para implementar fuzzy matching, útil en la normalización de nombres de productos con distintas variantes textuales.

### **SARIMA**

Extensión del modelo ARIMA que incorpora componentes estacionales explícitos. Se utiliza cuando los datos presentan ciclos recurrentes, como ocurre en muchos productos agrícolas.

### **Transparencia alimentaria**

Principio que defiende que toda la información sobre la formación de precios, márgenes, costes y orígenes de los productos debe estar accesible y ser comprensible para todos los agentes de la cadena, incluidos los consumidores.



## 4. Bibliografía

- [1] COAG, “IPOD Índice de precios en origen y destino de los alimentos.” Accessed: Dec. 15, 2024. [Online]. Available: <https://coag.org/post/ipod-indice-de-precios-en-origen-y-destino-de-los-alimentos-122677>
- [2] P. y A. Ministerio de Agricultura, “Observatorio de le Cadena Alimentaria.” Accessed: Dec. 10, 2024. [Online]. Available: <https://www.mapa.gob.es/es/estadistica/temas/estadisticas-alimentacion/observatorio-precios/>
- [3] R. P. i A. Departament d’Agricultura, “Dades. Departament d’Agricultura, Ramaderia, Pesca i Alimentació,” 30.08.2024. Accessed: Dec. 10, 2024. [Online]. Available: <https://agricultura.gencat.cat/ca/departament/estadistiques/observatori-agroalimentari-preus/preus-origen/preus-percebuts-pages/dades/>
- [4] R. P. i A. Departament d’Agricultura, “Consum alimentari. Departament d’Agricultura, Ramaderia, Pesca i Alimentació,” 05.12.2024. Accessed: Dec. 10, 2024. [Online]. Available: <https://agricultura.gencat.cat/ca/departament/estadistiques/alimentacio/consum-alimentari/>
- [5] “La Asamblea General adopta la Agenda 2030 para el Desarrollo Sostenible - Desarrollo Sostenible.” Accessed: Jun. 03, 2025. [Online]. Available: [https://www.un.org/sustainabledevelopment/es/2015/09/la-asamblea-general-adopta-la-agenda-2030-para-el-desarrollo-sostenible/?utm\\_source=chatgpt.com](https://www.un.org/sustainabledevelopment/es/2015/09/la-asamblea-general-adopta-la-agenda-2030-para-el-desarrollo-sostenible/?utm_source=chatgpt.com)
- [6] J. S. Saltz, “CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps,” in *2021 IEEE International Conference on Big Data (Big Data)*, IEEE, Dec. 2021, pp. 2337–2344. doi: 10.1109/BigData52589.2021.9671634.
- [7] “DIRECTIVA (UE) 2019/ 790 DEL PARLAMENTO EUROPEO Y DEL CONSEJO - de 17 de abril de 2019 - sobre los derechos de autor y derechos afines en el mercado único digital y por la que se modifican las Directivas 96/ 9/ CE y 2001/ 29/ CE.”
- [8] Pablo Haya, “La metodología CRISP-DM en ciencia de datos - IIC,” <https://www.iic.uam.es/innovacion/metodologia-crisp-dm-ciencia-de-datos/>.
- [9] K. H. Rose, “A Guide to the Project Management Body of Knowledge (PMBOK Guide) Fifth Edition,” *Project Management Journal*, vol. 44, no. 3, pp. e1–e1, Jun. 2013, doi: 10.1002/pmj.21345.
- [10] “Modelo autorregresivo integrado de media móvil - Wikipedia, la enciclopedia libre.” Accessed: Jun. 03, 2025. [Online]. Available: [https://es.wikipedia.org/wiki/Modelo\\_autorregresivo\\_integrado\\_de\\_media\\_m%C3%B3vil](https://es.wikipedia.org/wiki/Modelo_autorregresivo_integrado_de_media_m%C3%B3vil)
- [11] “ARIMA Time Series Models,” in *Statistics of Financial Markets*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 203–225. doi: 10.1007/978-3-540-76272-0\_12.