



Facultad de Ciencias Económicas y Empresariales

**IMPACTO DE UN SHOCK  
ENERGÉTICO EN EL  
COMPORTAMIENTO DEL  
CONSUMIDOR: ANÁLISIS MEDIANTE  
SIMULACIÓN Y MACHINE LEARNING  
APLICADO A LA COMUNIDAD DE  
MADRID**

Clave: 201707174

MADRID | Junio 2026

## **Resumen**

Este Trabajo de Fin de Grado analiza el comportamiento del consumidor ante un shock energético extremo, entendido como la eliminación del uso del gas, mediante una combinación de simulación y técnicas de Machine Learning.

El estudio se centra en la Comunidad de Madrid y parte de la construcción de un dataset sintético de 500 hogares, elaborado a partir de datos secundarios y variables económicas, estructurales y conductuales.

A partir de este conjunto de datos se incorporan variables como la renta anual, el gasto energético, el precio del shock, el coste de transición, la incertidumbre percibida, la exposición a la información y diversas interacciones como factores relevantes.

Sobre esta base se aplica un modelo de regresión logística que permita estimar la probabilidad de que un hogar decida sustituir el gas por otra fuente energética. Los resultados muestran que la decisión de cambio depende de forma significativa de la capacidad económica, la presión del shock, el acceso a información y las barreras asociadas al proceso de transición .

Los resultados en una primera aproximación permiten identificar patrones de comportamiento coherentes con la teoría económica y con la literatura sobre economía del comportamiento. En particular, se observa que la transición energética depende de múltiples factores. Ciertamente del precio, pero también de otras relaciones como la renta y el coste de adaptación, o de la vulnerabilidad del hogar y del papel de la incertidumbre.

En este sentido, el análisis ofrece una estimación útil sobre cómo podrían reaccionar los hogares ante escenarios de transición energética forzada y aporta una base metodológica para futuros estudios con datos reales.

## **Palabras clave**

Shock energético, comportamiento del consumidor, Machine Learning, regresión logística, simulación, Comunidad de Madrid, transición energética, incertidumbre

## ÍNDICE

CAPÍTULO I: INTRODUCCIÓN.....	5
<b>1.1 Planteamiento del problema</b> .....	5
<b>1.2 Justificación</b> .....	6
<b>1.3 Preguntas de investigación</b> .....	9
CAPÍTULO II: OBJETIVOS .....	10
<b>2.1 Objetivo general</b> .....	10
<b>2.2 Objetivos específicos</b> .....	10
2.2.1 <i>Evaluar el impacto del coste energético</i> <i>en la toma de decisiones del consumidor</i> .....	10
2.2.2. <i>Analizar el papel de la incertidumbre y la información</i> .....	10
2.2.3. <i>Analizar la sustitución energética en un contexto de transición forzada</i> .....	11
2.2.4. <i>Aplicar modelos predictivos para analizar el comportamiento del consumidor</i>	11
CAPÍTULO III: MARCO TEÓRICO .....	12
<b>3.1 Geopolítica energética y shocks de suministro</b> .....	12
<b>3.2 Teoría de shocks económicos</b> .....	14
<b>3.3 Economía del comportamiento</b> .....	15
<b>3.4 Economía experimental</b> .....	16
<b>3.5 Síntesis del marco teórico</b> .....	16
CAPÍTULO IV: METODOLOGÍA .....	17
<b>4.1 Datos secundarios</b> .....	17
4.1.1 <i>Consumo energético de los hogares</i> .....	17
4.1.2 <i>Precios de la energía (gas y electricidad)</i> .....	18
<b>4.2 Dataset sintético de hogares de la Comunidad de Madrid</b> .....	18
<b>4.3 Definición de variables y reglas de construcción del dataset sintético</b> .....	21
4.3.1 <i>Tipología de variables</i> .....	21
4.3.2 <i>Reglas de coherencia interna del dataset</i> .....	23
4.3.3 <i>Construcción de la variable objetivo</i> .....	24
4.3.4 <i>Justificación metodológica</i> .....	25

<b>4.4 Construcción técnica del dataset sintético</b> .....	26
4.4.1. <i>Generación de la estructura del dataset</i> .....	26
4.4.2. <i>Generación de variables mediante distribuciones estadísticas</i> .....	27
4.4.3. <i>Incorporación de relaciones estructurales entre variables</i> .....	29
4.4.4. <i>Simulación de escenarios de shock energético</i> .....	29
4.4.5. <i>Construcción de variables derivadas</i> .....	30
4.4.6. <i>Construcción de la variable objetivo</i> .....	31
4.4.7. <i>Validación y revisión del dataset</i> .....	31
4.4.8. <i>Consideraciones metodológicas</i> .....	34
<b>CAPÍTULO V: MACHINE LEARNING</b> .....	35
<b>5.1 Enfoque metodológico</b> .....	35
<b>5.2 Preparación de los datos</b> .....	37
<b>5.3 Modelo: Regresión logística</b> .....	39
<b>5.4 Evaluación del modelo</b> .....	41
<b>CAPÍTULO VI: RESULTADOS</b> .....	42
<b>6.1 Estadísticos descriptivos y patrones generales</b> .....	42
<b>6.2 Influencia del shock de precios</b> .....	45
6.2.1 <i>Efecto del precio shock</i> .....	45
6.2.2 <i>Relación entre precio y coste de transición</i> .....	47
6.2.3 <i>Impacto del precio sobre la probabilidad de cambio</i> .....	48
<b>6.3 Diferencias por perfil de hogar</b> .....	49
6.3.1 <i>Comparación entre renta alta y baja</i> .....	50
6.3.2 <i>Diferencias por vulnerabilidad</i> .....	51
6.3.3 <i>Diferencias por tipo de vivienda</i> .....	52
6.3.4 <i>Diferencias por sistema de calefacción</i> .....	53
6.3.5 <i>Diferencias por exposición a la información</i> .....	54
<b>6.4 Evaluación del modelo</b> .....	55
6.4.1 <i>Matriz de confusión</i> .....	56
6.4.2 <i>Métricas de rendimiento del modelo</i> .....	57
6.4.3 <i>Capacidad clasificatoria general</i> .....	59

<b>6.5 Interpretación de coeficientes</b> .....	60
6.5.1 <i>Signo de los coeficientes</i> .....	61
6.5.2 <i>Variables con mayor peso</i> .....	62
6.5.3 <i>Interacciones relevantes</i> .....	63
<b>6.6 Síntesis general de resultados</b> .....	64
<i>Tabla III. Síntesis de resultados principales</i> .....	65
CAPÍTULO VII: CONCLUSIONES .....	66
CAPÍTULO 8: LIMITACIONES DEL ESTUDIO.....	68
<b>8.1 Limitaciones metodológicas</b> .....	68
<b>8.2 Limitaciones del modelo predictivo</b> .....	69
<b>8.3 Limitaciones del contexto analizado</b> .....	70
<b>8.4 Limitaciones sistémicas</b> .....	70
CAPÍTULO IX: CONTRIBUCIÓN.....	72
<b>9.1 Aporte académico</b> .....	73
<b>9.2 Aporte innovador</b> .....	73
<b>9.3 Aporte prospectivo</b> .....	74
<b>9.4 Aporte metodológico</b> .....	75
BIBLIOGRAFÍA.....	76
<b>Literatura</b> .....	76
<b>Artículos</b> .....	78
<b>Webs</b> .....	80
Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado .....	81
ANEXOS.....	83
<b>Anexo 1. Definición de variables en Python</b> .....	83

# CAPÍTULO I: INTRODUCCIÓN

## 1.1 Planteamiento del problema

El presente trabajo tiene como objetivo analizar la posible reacción de los consumidores ante un shock regulatorio extremo<sup>1</sup> en el mercado energético. Concretamente, la eliminación fulminante e imprevista del uso del gas. El estudio se centrará en la Comunidad de Madrid, con el fin de acotar la investigación y trabajar en un entorno más homogéneo. Este planteamiento se enmarca en el contexto de creciente incertidumbre energética a nivel global, derivado de tensiones geopolíticas, la volatilidad de los mercados y la aceleración de las políticas de transición energética<sup>2</sup> impulsadas por organismos nacionales e internacionales.

En particular, se plantea un escenario simulado en el que, a partir del año 2027 se prohíba el uso y suministro de gas en todo el territorio nacional, afectando tanto al consumo doméstico como a la comercialización de productos y servicios asociados (calefacción, vehículos, electrodomésticos, etc.). El análisis se enfocará en los hogares de la Comunidad de Madrid. Lo cual permitirá trabajar con condiciones equiparables en cuanto a nivel de renta, tipo de vivienda y consumo energético.

En cuanto a la eliminación fulminante e imprevista del uso del gas se trataría de una medida de carácter abrupto, que rompería con la gradualidad habitual de las políticas energéticas; generaría un ambiente de alta incertidumbre<sup>3</sup> para los mercados y para los consumidores.

---

<sup>1</sup> **Shock regulatorio extremo:** cambio abrupto y de gran impacto en las normas o políticas de un sector, capaz de alterar significativamente el comportamiento de los agentes económicos. En el presente estudio, hace referencia a la eliminación total del gas como fuente energética, generando incertidumbre y necesidad de adaptación en los hogares.

<sup>2</sup> **Transición energética:** proceso de cambio desde sistemas energéticos basados en combustibles fósiles hacia fuentes de energía más sostenibles y renovables.

<sup>3</sup> **Incertidumbre:** situación en la que los agentes económicos no pueden prever con precisión la evolución futura de variables relevantes, como precios o disponibilidad de recursos.

El valor de este estudio podría radicar en comprender anticipadamente cómo los hogares tomarían decisiones socioeconómicas ante cambios inesperados en el mercado energético cuando implican: subidas de precios, restricciones en la oferta y situaciones de incertidumbre.

En este sentido, el trabajo combina elementos de la economía del comportamiento y la economía experimental, ya que no solo analiza decisiones basadas en el coste, sino también cómo influyen factores como la percepción del riesgo<sup>4</sup> y la información disponible<sup>5</sup>.

Como se trata de un escenario hipotético difícil de observar en la realidad, se empleará un enfoque basado en simulación y modelización que permitirá analizar cómo podrían reaccionar los consumidores de la Comunidad de Madrid ante este tipo de situación. De este modo, el estudio busca identificar tendencias en el comportamiento y aportar información útil para la toma de decisiones en contextos de cambio energético.

## 1.2 Justificación

La relevancia de este estudio se fundamenta en el protagonismo de la transición energética a nivel global y en la necesidad de comprender cómo reaccionan los consumidores ante cambios estructurales en el mercado energético. En los últimos años, el sistema energético ha experimentado transformaciones significativas impulsadas por la necesidad de reducir las

---

<sup>4</sup> La **percepción del riesgo** se refiere a la evaluación subjetiva que realizan los individuos sobre la probabilidad y las consecuencias de un evento incierto. A diferencia del riesgo objetivo, que puede medirse mediante datos cuantitativos, la percepción del riesgo está influida por factores psicológicos, experiencias previas, nivel de información y contexto en el que se toma la decisión.

En el ámbito económico, la percepción del riesgo desempeña un papel clave en la toma de decisiones, especialmente en situaciones de incertidumbre. Los individuos no reaccionan únicamente en función de los costes o beneficios esperados, sino también en función de cómo interpretan y sienten dicho riesgo, lo que puede generar comportamientos como la aversión a la pérdida o la resistencia al cambio, tal y como señalan Kahneman (2011) y Tversky y Kahneman (1979).

<sup>5</sup> El concepto de **Información Disponible** hace referencia al nivel de exposición y acceso que tienen los hogares a información relacionada con la transición energética, incluyendo el conocimiento sobre alternativas al gas, evolución de los precios energéticos, tecnologías disponibles, eficiencia energética o posibles medidas de apoyo e incentivos públicos. Esta variable busca representar cómo el acceso a la información puede influir en la percepción del riesgo, reducir la incertidumbre y condicionar la disposición de los consumidores a realizar cambios en su sistema energético.

emisiones de carbono o por factores externos como las tensiones geopolíticas, la volatilidad de los precios y las alteraciones en las cadenas de suministro.

La crisis energética europea posterior a 2020 puso de manifiesto la elevada dependencia de los combustibles fósiles importados y la vulnerabilidad del sistema energético ante eventos geopolíticos y interrupciones en el suministro energético. Este contexto ha dado lugar a un entorno caracterizado por una elevada incertidumbre energética, en el que las decisiones de los agentes económicos, especialmente los hogares, han adquirido una importancia capital (European Commission, 2023).

Varios organismos internacionales han señalado que los shocks energéticos derivados de conflictos geopolíticos impactan directamente en la economía global y afectan: a los precios, al crecimiento económico, a lo macroeconómico, al comportamiento de los consumidores (International Monetary Fund, 2026; Organisation for Economic Co-operation and Development, 2026). En el caso de los consumidores, deben adaptarse a cambios en los costes energéticos, la disponibilidad de recursos y las políticas regulatorias.

España, al tener mayor dependencia energética del exterior y ser dicho mercado tan inseguro, se encuentra en una situación y entorno de incertidumbre, en el que las tensiones geopolíticas y la volatilidad de los mercados energéticos afectan directamente al coste del gas y del petróleo, así como a la estabilidad del suministro (CaixaBank Research, 2026; BBVA Research, 2026). Asimismo, análisis sectoriales subrayan el papel clave del gas en la seguridad energética y su relevancia dentro del sistema energético actual (Industria Química, 2026).

En este contexto, la Comunidad de Madrid constituye un caso de estudio especialmente relevante debido a su carácter urbano, su elevada densidad poblacional y la alta dependencia del gas en los hogares. Particularmente en sistemas de calefacción. Estas características permiten analizar el comportamiento del consumidor en un entorno relativamente homogéneo donde las decisiones energéticas están condicionadas por factores económicos, limitaciones estructurales, el tipo de vivienda o la infraestructura disponible.

El análisis del comportamiento del consumidor ante un escenario de eliminación del gas resulta particularmente pertinente en un momento en el que las políticas energéticas están evolucionando hacia modelos más sostenibles, pero en los que persisten importantes desafíos relacionados con la adaptación de los hogares. Como señala Fressoz (2025), las transiciones energéticas no responden a procesos lineales de sustitución, sino a dinámicas complejas en las que distintas fuentes de energía coexisten durante largos periodos de tiempo. Esta visión refuerza la idea de que los cambios en el sistema energético generan incertidumbre y requieren procesos de adaptación gradual por parte de los consumidores.

Además, la literatura reciente destaca que los mercados energéticos actuales además de regirse por la lógica de oferta y demanda, están profundamente influenciados por factores geopolíticos y estratégicos, que contribuyen a una mayor volatilidad e incertidumbre (Kirsch, 2026; Geldard, 2026). En este sentido, el análisis del comportamiento del consumidor adquiere una dimensión clave: las decisiones individuales pueden amplificar o mitigar los efectos de estos shocks en el sistema económico.

Por otra parte, desde un punto de vista metodológico, este trabajo aporta valor al emplear un enfoque basado en simulación y técnicas de Machine Learning, lo que permite superar las limitaciones de los datos observacionales en escenarios futuros o hipotéticos.

La eliminación total del gas constituye un escenario difícilmente observable en la realidad actual, la construcción de un dataset sintético<sup>6</sup> y la aplicación de modelos predictivos permiten analizar de forma controlada cómo podrían reaccionar los consumidores ante distintos niveles de impacto. Este enfoque resulta coherente con la creciente utilización de herramientas analíticas avanzadas en el estudio de fenómenos económicos complejos, ya que permite identificar patrones de comportamiento, estimar probabilidades de decisión y analizar la influencia de distintas variables en contextos de incertidumbre. En este sentido, la

---

<sup>6</sup> *Dataset sintético: conjunto de datos generado artificialmente a partir de distribuciones estadísticas reales, utilizado para simular escenarios cuando no se dispone de datos observables.*

combinación de simulación y Machine Learning constituye una aproximación innovadora y adecuada para abordar el problema planteado.

Finalmente, el presente estudio tiene relevancia desde una perspectiva académica, pero también aporta valor práctico. Los resultados obtenidos pueden contribuir a una mejor comprensión de cómo los consumidores reaccionan ante cambios en el sistema energético, lo que resulta útil para el diseño de políticas públicas orientadas a la transición energética, así como para la planificación estratégica en contextos de crisis. En consecuencia, este trabajo se justifica por su capacidad para integrar contexto real, análisis teórico y herramientas metodológicas avanzadas; ofrece una aproximación sólida al estudio del comportamiento del consumidor en un entorno de cambio energético acelerado

### **1.3 Preguntas de investigación**

A partir del planteamiento anterior, se formulan las siguientes preguntas de investigación:

- 1) ¿Cómo reaccionan los consumidores ante la eliminación del gas en un contexto de incertidumbre energética?
- 2) ¿En qué medida el incremento del precio energético influye en la decisión de cambiar de fuente de energía?
- 3) ¿Qué variables económicas y conductuales presentan una mayor influencia sobre la decisión de cambio energético de los hogares?
- 4) ¿Cómo varía la capacidad de adaptación de los consumidores en función de variables como la renta, el gasto energético o el tipo de vivienda?
- 5) ¿En qué medida es posible predecir el comportamiento del consumidor mediante modelos de simulación y Machine Learning?

## **CAPÍTULO II: OBJETIVOS**

### **2.1 Objetivo general**

El objetivo principal es analizar el impacto de un shock regulatorio extremo en el mercado energético, concretamente la eliminación del gas, sobre el comportamiento del consumidor de la Comunidad de Madrid, y utilizando como metodología, técnicas analíticas como la simulación y el machine learning.

Este objetivo busca evaluar respuestas económicas y cómo influyen en la toma de decisiones de los hogares factores como la percepción del riesgo, la información disponible y la capacidad económica en la toma de decisiones de los hogares en un contexto de incertidumbre.

### **2.2 Objetivos específicos**

#### *2.2.1 Evaluar el impacto del coste energético en la toma de decisiones del consumidor*

Analizar cómo el incremento del coste energético influye en el comportamiento de los hogares, especialmente en términos de reducción del consumo, cambio de hábitos o adopción de alternativas energéticas. Se pretende estimar el grado de sensibilidad al precio y su variación en función de características socioeconómicas.

#### *2.2.2. Analizar el papel de la incertidumbre y la información*

Estudiar cómo la incertidumbre, así como la disponibilidad y calidad de la información, afectan a la toma de decisiones de los consumidores. Se considerarán escenarios con distintos niveles de información para evaluar su impacto en la percepción del riesgo y en la toma de decisiones.

### *2.2.3. Analizar la sustitución energética en un contexto de transición forzada*

Examinar en qué medida los hogares estarían dispuestos a sustituir el gas por otras fuentes energéticas, así como los factores que facilitan o dificultan este proceso, como la renta, el tipo de vivienda o los costes de adaptación.

### *2.2.4. Aplicar modelos predictivos para analizar el comportamiento del consumidor*

Utilizar técnicas de análisis cuantitativo, incluyendo modelos de regresión y Machine Learning, con el objetivo de estimar la probabilidad de determinadas decisiones tales como el cambio de fuente energética así como identificar las variables más influyentes.

## **CAPÍTULO III: MARCO TEÓRICO**

El presente trabajo se fundamenta en un conjunto de enfoques teóricos que permiten analizar el comportamiento de los agentes económicos en contextos de alta incertidumbre y cambio estructural. En particular, se integran aportaciones de la geopolítica energética, la teoría de shocks económicos, la economía del comportamiento y la economía experimental, con el objetivo de ofrecer una visión completa del problema planteado. Estos enfoques resultan especialmente relevantes para analizar cómo los consumidores reaccionan ante cambios bruscos en las condiciones del mercado energético, como el escenario simulado de eliminación fulminante y repentina del gas en la Comunidad de Madrid.

En este supuesto, interactúan en el contexto diferentes factores macroeconómicos, geopolíticos y conductuales de forma compleja. Estos factores influyen en la formación de los shocks y en la toma de decisiones de los hogares. A continuación, se describen las principales líneas teóricas que servirán de base para el desarrollo del análisis.

### **3.1 Geopolítica energética y shocks de suministro**

El análisis de los mercados energéticos contemporáneos requiere tomar en cuenta el papel central de la geopolítica. Factores estratégicos, relaciones internacionales y dinámicas de poder entre países productores y consumidores son elementos que inciden en el mercado de la energía. En particular el gas natural es un recurso altamente sensible, ya que su producción, transporte y distribución depende de infraestructuras críticas y de acuerdos internacionales que puede verse alterados por conflictos, tensiones diplomáticas o decisiones regulatorias.

Los shocks energéticos recientes están estrechamente vinculados a tensiones geopolíticas que afectan a la estabilidad del suministro y a la incertidumbre en la economía global. En particular, el gas natural presenta una elevada exposición a riesgos geopolíticos debido a su dependencia de redes de transporte específicas, como gasoductos y terminales de gas natural

licuado<sup>7</sup>, lo que limita la flexibilidad del sistema frente a interrupciones del suministro. Esta característica lo convierte en un elemento clave dentro de la seguridad energética de los países, especialmente en economías con alta dependencia de importaciones energéticas.

En el caso de España, esta vulnerabilidad se manifiesta en su elevada dependencia del exterior para el abastecimiento energético<sup>8</sup>, que implica una exposición directa a las dinámicas del mercado internacional (Eurostat, 2026; IDAE, 2026). Tal y como señalan distintos análisis sectoriales, las tensiones geopolíticas pueden traducirse en incrementos significativos en los precios energéticos, alteraciones en la disponibilidad de recursos y cambios en las políticas regulatorias (International Energy Agency, 2025; International Monetary Fund, 2026; BBVA Research, 2026; CaixaBank Research, 2026). Desde esta perspectiva, el escenario planteado en el presente estudio (la eliminación del uso del gas) puede interpretarse como una simulación extrema de un shock geopolítico o regulatorio, en el que se produce una ruptura abrupta en el sistema energético.

Este tipo de situaciones permite analizar cómo reaccionan los agentes económicos ante cambios drásticos en las condiciones del mercado. Asimismo, la literatura destaca que los efectos de los shocks geopolíticos no se limitan al ámbito macroeconómico, sino que se trasladan al comportamiento de los consumidores. El aumento de los precios, la incertidumbre sobre el suministro y la necesidad de adaptación generan cambios en las

---

<sup>7</sup> Las **terminales de gas natural licuado** (GNL) son infraestructuras que permiten recibir, almacenar, regasificar y distribuir gas natural transportado en estado líquido, facilitando su integración en la red energética tras su conversión nuevamente a estado gaseoso.

<sup>8</sup> La **dependencia energética exterior** se refiere a la necesidad de un país de importar una parte significativa de los recursos energéticos que consume, debido a una producción interna insuficiente para cubrir la demanda. En el caso de España, esta dependencia es especialmente relevante en relación con combustibles fósiles como el gas natural y el petróleo, cuya disponibilidad está condicionada por la capacidad de producción de otros países y por las infraestructuras de transporte internacional, como gasoductos y terminales de gas natural licuado. Esta situación implica que el sistema energético nacional se encuentra expuesto a factores externos que escapan al control directo del país, tales como la volatilidad de los precios en los mercados internacionales, las tensiones geopolíticas o posibles interrupciones en el suministro. Como consecuencia, la dependencia energética no solo tiene implicaciones económicas, sino también estratégicas, al afectar a la seguridad energética y a la estabilidad del sistema en contextos de incertidumbre.

decisiones individuales, que pueden variar en función de factores como la renta, la información disponible o la percepción del riesgo. La incorporación de la dimensión geopolítica en el análisis permite contextualizar el problema de investigación y entender el origen de los shocks energéticos que afectan al comportamiento del consumidor. De este modo, el estudio no solo analiza una situación hipotética, sino que se enmarca dentro de una realidad caracterizada por la creciente interdependencia entre energía, economía y geopolítica.

### **3.2 Teoría de shocks económicos**

La teoría de shocks económicos analiza el impacto de perturbaciones inesperadas en el funcionamiento de los mercados y en el comportamiento de los agentes económicos. Según Blanchard (2017), los shocks pueden afectar tanto a la oferta como a la demanda, generando cambios en los precios, el consumo y la inversión<sup>9</sup>. En el caso del mercado energético, los shocks suelen estar asociados a factores externos como conflictos geopolíticos, interrupciones en el suministro o cambios regulatorios. Hamilton (2009) demuestra que los shocks en los precios del petróleo tienen efectos significativos sobre la economía, afectando tanto al crecimiento como al comportamiento del consumidor.

En la actualidad, organismos internacionales como el Fondo Monetario Internacional (IMF, 2026) y la OCDE (OECD, 2026) han señalado que los shocks energéticos derivados de tensiones geopolíticas están generando un entorno de alta incertidumbre, caracterizado por la volatilidad de los precios y la inestabilidad de los mercados. En este contexto, la eliminación del gas puede interpretarse como un shock regulatorio extremo, cuyos efectos

---

<sup>9</sup> Los shocks económicos afectan simultáneamente a la oferta y a la demanda debido a que alteran las condiciones fundamentales del mercado. Por un lado, pueden impactar en la oferta al modificar los costes de producción, la disponibilidad de recursos o la capacidad productiva de las empresas, lo que puede traducirse en una reducción de la producción o en un encarecimiento de los bienes y servicios. Por otro lado, los shocks también influyen en la demanda, al afectar a los precios, la renta disponible de los consumidores y sus expectativas, generando cambios en los patrones de consumo. Como consecuencia, estas alteraciones simultáneas en oferta y demanda provocan variaciones en los precios, el nivel de consumo y las decisiones de inversión, especialmente en contextos de elevada incertidumbre.

incluyen aumentos en los precios energéticos, cambios en los patrones de consumo y una mayor incertidumbre en la toma de decisiones de los hogares.

### 3.3 Economía del comportamiento

Desde la perspectiva de la teoría de shocks económicos, resulta necesario analizar no solo el impacto agregado de estas perturbaciones, sino también cómo afectan a las decisiones individuales de los agentes económicos. En este sentido, la economía del comportamiento proporciona un marco teórico adecuado para comprender la respuesta de los consumidores ante contextos de incertidumbre energética. Las decisiones de los consumidores responden a criterios racionales, pero también están influenciadas por factores psicológicos y limitaciones cognitivas. Autores como Kahneman (2011) y Thaler (2015) señalan que los individuos utilizan atajos mentales (heurísticas<sup>10</sup>) y presentan sesgos sistemáticos que afectan a su toma de decisiones, especialmente en contextos de incertidumbre.

La teoría de la perspectiva desarrollada por Tversky y Kahneman (1979) demuestra que los individuos tienden a mostrar **aversión a la pérdida**, lo que implica que perciben las pérdidas con mayor intensidad que las ganancias equivalentes. Este comportamiento resulta especialmente relevante en el ámbito energético, donde incrementos en los precios o restricciones en el suministro pueden generar reacciones desproporcionadas en el consumo. Otros factores como la **percepción del riesgo**, la **información disponible** y la **incertidumbre** influyen en la toma de decisiones de los hogares, lo que implica que las respuestas ante un shock energético<sup>11</sup> pueden diferir significativamente entre individuos, incluso en condiciones similares.

---

<sup>10</sup> **Heurísticas:** reglas mentales simplificadas que facilitan la toma de decisiones en contextos de incertidumbre, reduciendo el esfuerzo cognitivo, pero pudiendo generar sesgos.

<sup>11</sup> **Shock energético:** perturbación repentina en el sistema energético que afecta a los precios, la oferta o la demanda de energía, generando efectos económicos significativos.

### 3.4 Economía experimental

La economía experimental permite estudiar la toma de decisiones en entornos controlados, donde es posible aislar variables y analizar su impacto de forma sistemática. Según Friedman y Sunder (1994), este enfoque facilita el análisis de situaciones en las que no existen datos observables o en las que se desea evaluar escenarios hipotéticos.

Para este trabajo, el escenario simulado de eliminación del gas, se alinea con este enfoque; ya que permite analizar cómo reaccionarían los consumidores ante un shock energético extremo. Tal y como señala Camerer (2003), los experimentos económicos permiten identificar patrones de comportamiento y comprender mejor la interacción entre factores económicos y conductuales.

Por eso, el enfoque de la economía experimental resulta especialmente adecuado para el presente estudio, porque la eliminación total del gas constituye un **escenario difícilmente observable** en la realidad actual, lo que justifica el uso de técnicas de simulación para aproximar el comportamiento de los consumidores.

### 3.5 Síntesis del marco teórico

En conjunto, los enfoques teóricos presentados permiten abordar el problema de investigación desde una perspectiva integral. La geopolítica energética explica el origen de los shocks, la teoría económica permite comprender su impacto en los mercados, la economía del comportamiento aporta una visión sobre la toma de decisiones de los consumidores y la economía experimental justifica el uso de simulaciones para el análisis de escenarios no observables.

Esta integración teórica proporciona una base sólida para el desarrollo metodológico del estudio, permitiendo conectar el contexto macroeconómico con el análisis microeconómico del comportamiento del consumidor y justificando la aplicación de técnicas de Machine Learning en el análisis empírico.

## CAPÍTULO IV: METODOLOGÍA

La metodología que se emplea en este estudio parte de un escenario simulado de shock energético extremo, en el que se plantea la eliminación del gas como fuente de suministro en los hogares.

Como se trata de una situación hipotética y no observable de forma directa en datos reales, se adopta un enfoque basado en simulación, construcción de un dataset sintético y aplicación de técnicas de Machine Learning.

El objetivo es aproximar cómo podrían reaccionar los hogares de la Comunidad de Madrid ante este cambio, prestando atención a su **capacidad de adaptación** y a los factores que influyen en la decisión de sustituir o no la fuente energética.

### 4.1 Datos secundarios

Se emplearán datos procedentes de fuentes oficiales con el objetivo de construir una base realista del contexto energético en la Comunidad de Madrid. En aquellos casos en los que no exista información desagregada a nivel regional, se utilizarán datos nacionales ajustados a las características del entorno urbano madrileño.

#### 4.1.1 Consumo energético de los hogares

Se utilizarán datos sobre consumo energético residencial para estimar el gasto energético y el grado de dependencia del gas en los hogares.

Para la construcción del dataset sintético, se asumió que aproximadamente el 72% de los hogares presentan dependencia del gas como sistema principal de calefacción. Este parámetro fue incorporado como supuesto de simulación para representar un escenario de alta exposición energética en la Comunidad de Madrid.

**Fuentes:**

- Instituto Nacional de Estadística (INE)
- IDAE (Instituto para la Diversificación y Ahorro de la Energía).

*4.1.2 Precios de la energía (gas y electricidad)*

Se emplearán datos de precios energéticos procedentes de mercados nacionales y europeos para construir distintos escenarios de incremento de costes derivados del shock regulatorio.

Dado que estos precios se determinan en mercados de ámbito superior al regional, se considerarán aplicables al contexto de la Comunidad de Madrid.

**Fuentes:**

- Red Eléctrica de España
- OMIE (Operador del Mercado Ibérico de Energía)
- Eurostat.

**4.2 Dataset sintético de hogares de la Comunidad de Madrid**

Dado que el estudio se basa en un escenario hipotético, se construirá un dataset sintético representativo de hogares de la Comunidad de Madrid, generado a partir de distribuciones estadísticas observadas en datos secundarios procedentes de fuentes oficiales como el Instituto Nacional de Estadística (INE), Eurostat y el IDAE.

El dataset estará compuesto por un total de **500 hogares simulados**, con el objetivo de recrear de forma aproximada la estructura real del consumo energético y las características socioeconómicas de los hogares madrileños. Este tamaño muestral permite disponer de un volumen suficiente de observaciones para realizar análisis cuantitativos y aplicar modelos de

Machine Learning, manteniendo al mismo tiempo la simplicidad y manejabilidad del conjunto de datos.

Para garantizar la coherencia con el contexto real, el *dataset* se diseñará respetando las proporciones observadas en la Comunidad de Madrid, especialmente en lo relativo a la alta dependencia del gas como sistema de calefacción, así como al predominio de viviendas tipo piso frente a viviendas unifamiliares. Asimismo, se incorporarán distribuciones realistas de renta, gasto energético, coste de transición, incertidumbre y exposición a la información, con el fin de reflejar la diversidad de situaciones económicas y energéticas existentes en la población.

El proceso de construcción del dataset consistirá en la generación de múltiples observaciones que representen distintos hogares, asignando valores a cada variable mediante distribuciones estadísticas plausibles<sup>12</sup>.

En concreto, se utilizarán distribuciones continuas, como la distribución normal para variables como la renta del hogar, el gasto energético y el coste de transición; y distribuciones discretas o proporcionales para variables categóricas como el tipo de vivienda y el sistema de calefacción. Este enfoque permite introducir heterogeneidad en el dataset, reflejando diferencias reales entre hogares.

Adicionalmente, se incorporarán relaciones coherentes entre variables, de forma que el dataset no solo reproduzca distribuciones individuales, sino también patrones estructurales realistas. Por ejemplo, niveles de renta más bajos estarán asociados a mayores niveles de vulnerabilidad económica, mientras que determinados tipos de vivienda podrán presentar mayores niveles de consumo energético y mayores costes de adaptación.

La incertidumbre percibida y la exposición a la información se incorporarán como factores que afectan al proceso de decisión del hogar ante el shock energético.

---

<sup>12</sup> Distribuciones de datos simuladas que intentan parecerse a cómo podrían comportarse los datos en la realidad.

En concreto, el dataset incluirá las siguientes variables principales: renta del hogar, gasto energético, precio del shock energético, tipo de vivienda, sistema de calefacción, vulnerabilidad económica, coste de transición, incertidumbre, exposición a la información y una variable objetivo binaria que representará la decisión de cambio de fuente energética. Estas variables permiten capturar tanto factores económicos como estructurales y conductuales que influyen en el comportamiento del consumidor.

Asimismo, se incorporará una **variable objetivo** que representará la decisión de cambio de fuente energética, definida como una variable binaria (0/1) a partir de reglas económicas que consideran el nivel de renta, el incremento del precio energético, el coste de transición y la dependencia inicial del gas. Esta variable permitirá analizar la probabilidad de que un hogar adopte alternativas energéticas en un contexto de shock regulatorio.

El dataset resultante permitirá simular el comportamiento de los consumidores bajo distintos escenarios de incremento de precios, restricciones en la oferta energética y niveles de incertidumbre. De este modo, constituye una herramienta fundamental para el desarrollo del análisis empírico, sirviendo como base para la aplicación posterior de modelos de análisis cuantitativo y técnicas de Machine Learning.

De tal manera que tenemos:

- **Variables socioeconómicas:** renta, gasto energético, vulnerabilidad.
- **Variables del escenario energético:** precio shock.
- **Variables estructurales:** tipo de vivienda, sistema de calefacción, dependencia gas.
- **Variables conductuales / barrera de adopción:** coste de transición, incertidumbre, exposición a información.
- **Variable objetivo:** cambio de energía (0/1).

### 4.3 Definición de variables y reglas de construcción del dataset sintético

Con el objetivo de analizar el comportamiento del consumidor ante un escenario de shock energético, se ha diseñado un dataset sintético representativo de hogares de la Comunidad de Madrid. Este dataset no solo reproduce distribuciones estadísticas realistas, sino que incorpora relaciones estructurales y conductuales coherentes entre variables, con el fin de aproximar de forma verosímil la toma de decisiones de los agentes económicos.

La construcción del dataset se fundamenta en la combinación de variables económicas, estructurales y conductuales, cuya interacción permite modelizar el proceso de decisión de los hogares ante la eliminación del gas como fuente energética.

#### 4.3.1 Tipología de variables

El conjunto de datos se estructura en cuatro grandes bloques:

##### *Variables socio-económicas*

Las variables socio-económicas capturan la capacidad financiera de los hogares y su exposición al gasto energético:

- **Renta anual del hogar (€):** variable continua que representa el nivel de ingresos del hogar. Se ha modelizado mediante distribuciones que reflejan la estructura real de renta en la Comunidad de Madrid, diferenciando entre hogares de renta baja, media y alta.
- **Gasto energético anual (€):** variable continua que recoge el gasto previo en energía. Esta variable se ha definido en función del tipo de vivienda y del sistema de calefacción, incorporando relaciones coherentes entre consumo energético y características estructurales del hogar.
- **Vulnerabilidad económica:** variable derivada que resume la fragilidad financiera del hogar frente a incrementos en el precio de la energía. Se construye a partir de la

relación entre renta y gasto energético, de modo que permite identificar hogares con mayor dificultad para absorber el impacto de un shock regulatorio.

### *Variables del escenario energético*

- **Precio shock (%):** variable que representa el incremento simulado del precio de la energía en el escenario analizado. Se define mediante distintos niveles de presión regulatoria o de mercado, con el fin de evaluar cómo afectan los cambios en el coste energético a la decisión de los hogares.

### *Variables estructurales*

Las variables estructurales describen las características físicas y tecnológicas del hogar:

- **Tipo de vivienda:** variable categórica que diferencia entre viviendas en piso y viviendas tipo casa. Esta clasificación se incorpora debido a que las características estructurales de la vivienda pueden afectar tanto al consumo energético como a la capacidad de adaptación ante un cambio de sistema energético.
- **Sistema de calefacción:** variable categórica que diferencia entre hogares con dependencia del gas y hogares con sistemas energéticos alternativos para calefacción. Esta variable resulta especialmente relevante dentro del estudio, ya que condiciona el grado de exposición de los hogares ante el escenario de shock energético planteado.
- **Dependencia del gas:** variable binaria derivada del sistema de calefacción, que indica si el hogar depende del gas como fuente principal de energía. Esta variable resulta clave para la simulación del escenario de eliminación del gas.

### *Variables conductuales o de barrera de adopción*

Las variables conductuales permiten incorporar factores no estrictamente económicos en el proceso de decisión:

- **Coste de transición energética (€):** variable continua que representa la inversión necesaria para sustituir el sistema energético actual por una alternativa. Este coste depende tanto del tipo de vivienda como del sistema de calefacción inicial.
- **Nivel de incertidumbre percibida (escala 1–10):** variable ordinal que refleja la percepción subjetiva del riesgo por parte del consumidor en un entorno de cambio regulatorio. Se asume que mayores niveles de incertidumbre reducen la propensión al cambio.
- **Exposición a la información (escala 1–10):** variable ordinal que mide el grado de acceso y comprensión de información relevante sobre el contexto energético y las alternativas disponibles. Se considera que una mayor exposición a información facilita la toma de decisiones.

#### *Variable objetivo*

- **Cambio de fuente energética (0/1):** variable binaria que representa la decisión del hogar de sustituir el gas por una fuente energética alternativa.

Esta variable constituye el elemento central del análisis y será utilizada como variable dependiente<sup>13</sup> en el modelo de regresión logística descrito en el apartado de Machine Learning.

#### *4.3.2 Reglas de coherencia interna del dataset*

Para garantizar la validez del dataset sintético, no solo se han replicado distribuciones individuales, sino que se han incorporado relaciones estructurales coherentes entre variables. Estas relaciones permiten aproximar patrones observados en la realidad y dotar al modelo de mayor consistencia económica.

---

<sup>13</sup> *Variable dependiente: variable cuyo valor se desea explicar o predecir.*

Entre las principales reglas de coherencia destacan:

- Los hogares con **mayor nivel de renta** presentan, en promedio, **menor vulnerabilidad económica** y mayor capacidad para asumir costes de transición energética.
- Las **viviendas unifamiliares** tienden a registrar un mayor gasto energético, debido a su mayor superficie y necesidades de climatización.
- La **dependencia del gas** se vincula directamente con el sistema de calefacción, siendo más frecuente en determinados tipos de vivienda.
- El **coste de transición energética** es significativamente mayor en hogares que utilizan gas, especialmente en viviendas unifamiliares, debido a la necesidad de adaptar infraestructuras existentes.
- La **vulnerabilidad económica** aumenta cuando el gasto energético representa una proporción elevada de la renta disponible del hogar.
- La **incertidumbre percibida** se relaciona inversamente con la exposición a información, de modo que hogares mejor informados presentan menores niveles de incertidumbre.

Estas relaciones permiten que el dataset no sea una simple agregación de valores aleatorios, sino una **representación estructurada y coherente** del comportamiento económico de los hogares.<sup>14</sup>

#### *4.3.3 Construcción de la variable objetivo*

La construcción de la variable objetivo se basa en supuestos coherentes con la literatura sobre comportamiento del consumidor, transición energética y toma de decisiones bajo incertidumbre. Diversos estudios señalan que factores económicos, estructurales y

---

<sup>14</sup> Ver Anexo I. Definición de variables en Python.

conductuales influyen en la disposición de los hogares a adoptar cambios energéticos o tecnologías alternativas (Wilson & Dowlatabadi, 2007; Frederiks et al., 2015).

En concreto, se asume que la probabilidad de cambio de fuente energética aumente cuando:

- El hogar presente **dependencia del gas**, lo que implica una mayor exposición al shock regulatorio.
- Se enfrente a un **incremento significativo del precio energético**.
- Dispone de un **nivel de renta suficiente** para asumir el coste de transición.
- Presente una **alta exposición a información** sobre alternativas energéticas.

Por el contrario, la probabilidad de cambio disminuye cuando:

- El **coste de transición energética es elevado**.
- El hogar presente **alta vulnerabilidad económica**.
- Existe un **alto nivel de incertidumbre percibida**.

A partir de estos criterios, se ha construido una función de decisión que combina los distintos factores mediante una puntuación agregada (score), permitiendo clasificar cada observación en función de su probabilidad de cambio.

#### *4.3.4 Justificación metodológica*

El uso de reglas de decisión en la construcción de la variable objetivo se alinea con el enfoque de simulación adoptado en el presente estudio. Este método permite reproducir comportamientos plausibles en ausencia de datos observados, especialmente en escenarios futuros o hipotéticos.

Además, la incorporación de variables conductuales responde a la evidencia aportada por la economía del comportamiento, según la cual las decisiones de los consumidores no dependen

únicamente de factores económicos, sino también de elementos como la percepción del riesgo, la información disponible y la incertidumbre.

En conjunto, el dataset construido constituye una base sólida para el análisis empírico, permitiendo aplicar modelos de Machine Learning y evaluar el impacto de distintos factores en la toma de decisiones de los hogares en un contexto de shock energético.

#### **4.4 Construcción técnica del dataset sintético**

La construcción del dataset sintético constituye una fase clave dentro del presente estudio, ya que permite trasladar el planteamiento teórico a una base de datos estructurada, coherente y apta para el análisis cuantitativo.

Dado que el escenario analizado (la eliminación total del gas en el contexto energético) es de carácter hipotético y no observable en la realidad actual, resulta necesario recurrir a técnicas de simulación que permitan aproximar el comportamiento de los consumidores bajo condiciones controladas.

El proceso de construcción del dataset se ha llevado a cabo mediante un enfoque sistemático que combina la generación de datos aleatorios con la aplicación de reglas económicas, estructurales y conductuales. Este procedimiento permite obtener un conjunto de datos plausible, en el que cada observación representa un hogar simulado de la Comunidad de Madrid, manteniendo coherencia con las características del contexto real descritas en apartados anteriores.

##### *4.4.1. Generación de la estructura del dataset*

El dataset se compone de un total de **500 observaciones**, donde cada fila representa un hogar. Esta dimensión muestral se considera suficiente para realizar análisis exploratorios y aplicar

modelos de clasificación, manteniendo al mismo tiempo la simplicidad necesaria para un trabajo de estas características.

Cada observación incluye un conjunto de variables previamente definidas, agrupadas en categorías económicas, estructurales y conductuales, así como una variable objetivo binaria que representa la decisión de cambio de fuente energética.

La generación de los datos se ha realizado mediante herramientas de programación orientadas al análisis de datos, lo que permite automatizar el proceso, garantizar la reproducibilidad del dataset y aplicar reglas condicionales de forma consistente en todas las observaciones.

#### *4.4.2. Generación de variables mediante distribuciones estadísticas*

Las variables del dataset se han generado utilizando distribuciones estadísticas adaptadas a la naturaleza de cada variable, con el objetivo de introducir variabilidad y heterogeneidad en los datos.

En el caso de variables continuas, como la renta anual o el gasto energético, se han empleado rangos definidos a partir de valores plausibles<sup>15</sup> para el contexto de la Comunidad de Madrid. Estas variables se han generado de forma que reflejen diferentes niveles socioeconómicos, permitiendo distinguir entre hogares con mayor o menor capacidad económica.

Por su parte, las variables categóricas, como el tipo de vivienda o el sistema de calefacción, se han generado mediante distribuciones probabilísticas que reproducen proporciones observadas en datos reales. En particular, se ha considerado una alta prevalencia del uso del gas como sistema de calefacción, en línea con la evidencia empírica disponible, lo que refuerza la relevancia del escenario de shock planteado.

---

<sup>15</sup> Valores posibles, razonables y coherentes con la realidad esperada

Este enfoque permite construir un dataset que, aun siendo sintético, mantiene una estructura coherente con la realidad y resulta adecuado para el análisis del comportamiento del consumidor.

Tabla I. Descripción de variables

<b>Variable</b>	<b>Descripción</b>
ID_hogar	Identificador único del hogar
Renta_anual	Ingresos anuales estimados del hogar
Gasto_energetico_anual	Gasto energético previo estimado
Tipo_vivienda	Piso o vivienda unifamiliar
Sistema_calefaccion	Gas, electricidad u otro sistema
Dependencia_gas	Variable binaria: 1 si depende del gas, 0 si no
Precio_shock	Incremento simulado del precio energético
Vulnerabilidad	Nivel de vulnerabilidad económica
Coste_transicion	Coste estimado de sustitución energética
Incertidumbre	Nivel de incertidumbre percibida
Exposicion_info	Nivel de exposición a información
Score_cambio	Puntuación agregada de probabilidad de cambio
Cambio_energia	Variable objetivo: 1 cambia, 0 no cambia

#### *4.4.3. Incorporación de relaciones estructurales entre variables*

Uno de los elementos fundamentales en la construcción del dataset ha sido la incorporación de relaciones coherentes entre variables, con el fin de evitar una generación completamente aleatoria de los datos.

En este sentido, se han definido reglas que reflejan patrones económicos plausibles. Por ejemplo, el gasto energético se ha vinculado al tipo de vivienda, de modo que las viviendas unifamiliares presentan, en promedio, un mayor consumo energético que los pisos. Del mismo modo, la dependencia del gas se deriva directamente del sistema de calefacción utilizado por el hogar.

Asimismo, se ha establecido una relación entre la renta y la vulnerabilidad económica, considerando que los hogares con menor nivel de ingresos o con un mayor peso del gasto energético sobre su renta presentan mayores niveles de vulnerabilidad. Este tipo de relaciones permite capturar dinámicas económicas relevantes y aporta mayor realismo al dataset.

Además, se han incorporado interacciones entre variables conductuales como la relación inversa entre la exposición a información y el nivel de incertidumbre percibida. Reflejando que un mayor acceso a información puede reducir la percepción de riesgo en contextos de cambio.

#### *4.4.4. Simulación de escenarios de shock energético*

Para representar el impacto del shock regulatorio, se ha introducido una variable que recoge distintos niveles de incremento del precio energético. En concreto, se han definido tres escenarios: moderado, alto y extremo, representados mediante incrementos del 20%, 40% y 60%, respectivamente.

Esta variable permite analizar cómo varía la respuesta de los consumidores ante diferentes niveles de presión económica, constituyendo un elemento central en la simulación del comportamiento.

La inclusión de múltiples escenarios aporta mayor profundidad al análisis, ya que permite comparar resultados y observar posibles cambios en los patrones de decisión en función de la intensidad del shock.

#### *4.4.5. Construcción de variables derivadas*

A partir de las variables básicas generadas, se han construido variables derivadas que permiten capturar relaciones más complejas entre los factores analizados.

En particular, se ha calculado un indicador de vulnerabilidad económica a partir de la relación entre el gasto energético y la renta del hogar. Este indicador permite identificar hogares con mayor exposición a situaciones de estrés económico ante incrementos en los precios energéticos.

Asimismo, se ha estimado el coste de transición energética en función del tipo de vivienda y del sistema de calefacción inicial. Se ha considerado que los hogares dependientes del gas requieren mayores inversiones para adaptarse a fuentes alternativas.

Estas variables derivadas enriquecen el dataset y permiten incorporar una mayor profundidad analítica en el modelo.

#### *4.4.6. Construcción de la variable objetivo*

Una vez definidas las variables relevantes del estudio, se procedió a la construcción técnica de la variable objetivo mediante un proceso de simulación probabilística. Para ello, se diseñó una función de puntuación agregada (score) capaz de sintetizar la influencia conjunta de las distintas variables económicas, estructurales y conductuales incluidas en el modelo.

Esta puntuación se calculó combinando variables como el precio del shock energético, el nivel de renta, el coste de transición, la incertidumbre percibida y la exposición a la información. Se asignaron distintos pesos relativos a cada una de ellas dentro de la función de decisión.

Posteriormente, dicha puntuación fue transformada en una probabilidad de cambio energético mediante reglas condicionadas por la dependencia del gas. En particular, los hogares con dependencia del gas presentan una probabilidad de transición significativamente superior a los hogares que utilizan sistemas energéticos alternativos.

Finalmente, la variable objetivo “Cambio\_energía” fue generada mediante una distribución binomial, permitiendo clasificar cada hogar en dos categorías:

- **Valor 1:** el hogar decide cambiar de fuente energética.
- **Valor 0:** el hogar mantiene su sistema energético actual.

Este procedimiento permite simular decisiones individuales de forma coherente con los supuestos establecidos en el estudio y generar un conjunto de datos adecuado para la aplicación posterior del modelo de regresión logística.

#### *4.4.7. Validación y revisión del dataset*

Una vez generado el dataset, se ha realizado una revisión básica de los datos con el objetivo de verificar su coherencia interna y detectar posibles inconsistencias.

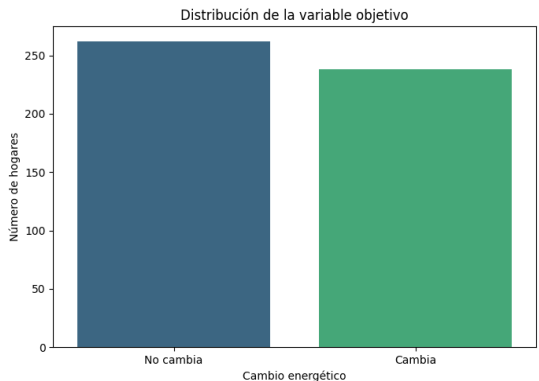
Esta revisión incluye el análisis de estadísticas descriptivas, la comprobación de rangos de valores y la evaluación de la distribución de la variable objetivo.

En particular, se ha prestado atención a que exista una proporción equilibrada entre hogares que cambian y que no cambian de fuente energética, evitando sesgos extremos que puedan afectar al rendimiento del modelo.

Asimismo, se ha verificado que las relaciones entre variables se mantengan dentro de los supuestos definidos, garantizando que el dataset refleja patrones plausibles desde el punto de vista económico.

La Figura I muestra la distribución de la variable objetivo “Cambio\_energia” dentro del dataset sintético. Puede observarse una distribución relativamente equilibrada entre hogares que realizan la transición energética y aquellos que mantienen su sistema energético original, evitando un desbalance extremo entre clases. Este aspecto resulta especialmente relevante para el entrenamiento del modelo predictivo, ya que contribuye a reducir posibles sesgos de clasificación y favorece una evaluación más estable del rendimiento del modelo.

Figura I. Distribución de la variable objetivo

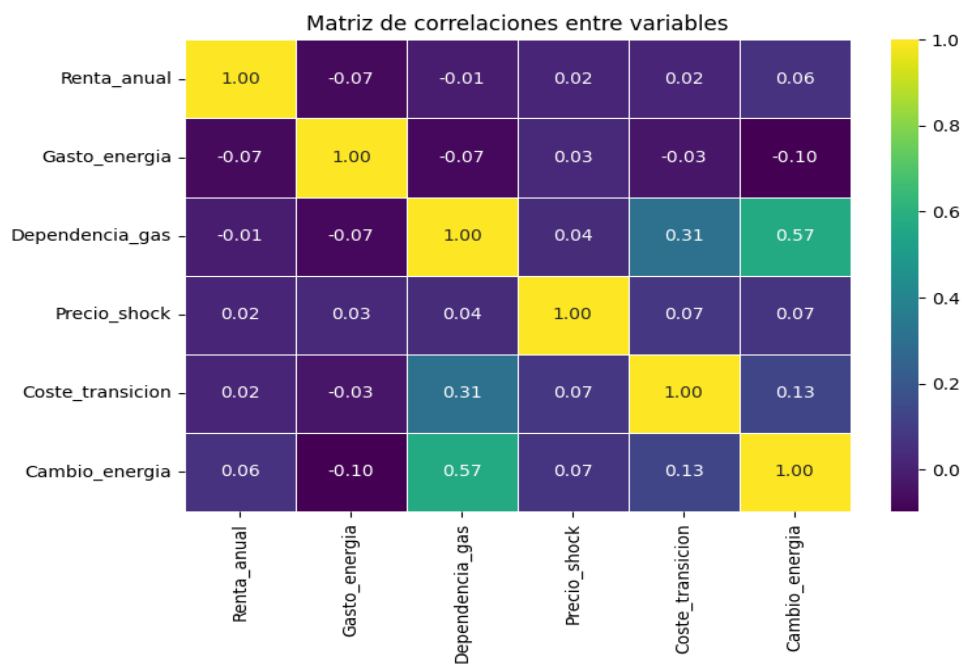


La Figura II presenta la matriz de correlaciones entre las principales variables numéricas incluidas en el estudio. Los resultados muestran que la variable “Cambio\_energia” presenta su relación más elevada con la variable “Dependencia\_gas”, indicando que los hogares con una mayor exposición al uso del gas muestran una mayor probabilidad de realizar una transición energética en el escenario simulado.

Por otro lado, variables como el “Coste\_transicion” presentan una relación positiva moderada con el cambio energético, mientras que el “Precio\_shock” muestra una correlación relativamente baja.

Esto sugiere que, dentro de la estructura del dataset sintético, la dependencia previa del gas actúa como el principal factor explicativo del comportamiento de los hogares. Asimismo, puede observarse una relación positiva entre “Dependencia\_gas” y “Coste\_transicion”, lo que resulta coherente desde el punto de vista económico, ya que los hogares más dependientes del gas requieren mayores inversiones para sustituir su sistema energético.

Figura II. Matriz de correlaciones entre variables



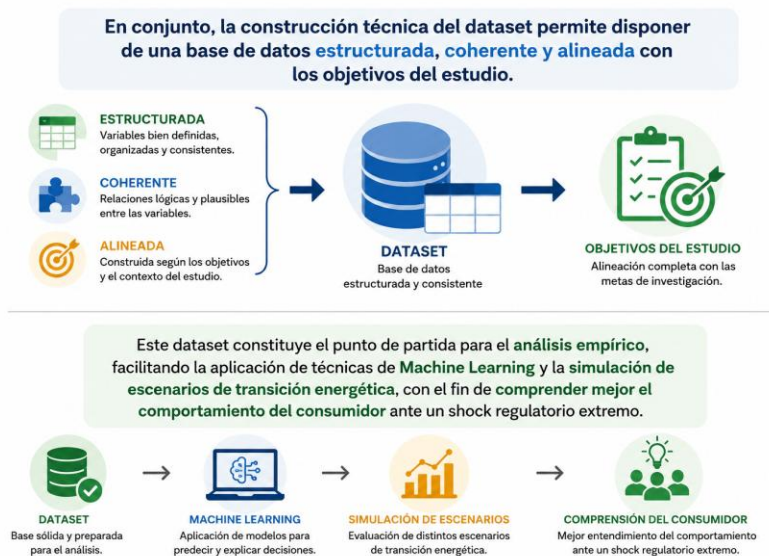
#### 4.4.8. Consideraciones metodológicas

El uso de técnicas de simulación en la construcción del dataset responde a la necesidad de analizar escenarios futuros o no observables en la realidad. Este enfoque permite explorar el comportamiento del consumidor en contextos de alta incertidumbre y evaluar el impacto potencial de políticas energéticas extremas.

Si bien los datos generados no representan observaciones reales, la incorporación de distribuciones estadísticas plausibles y de reglas de coherencia interna permite aproximar de forma razonable el comportamiento de los hogares.

En este sentido, el dataset sintético constituye una herramienta válida para el análisis exploratorio y la aplicación de modelos predictivos, siempre teniendo en cuenta las limitaciones inherentes a este tipo de aproximación, que serán abordadas en el apartado correspondiente. Ver Infografía I

#### Infografía I



## CAPÍTULO V: MACHINE LEARNING

### 5.1 Enfoque metodológico

Con el objetivo de complementar el análisis económico desarrollado en los apartados anteriores, el presente estudio incorpora un enfoque basado en técnicas de Machine Learning, concretamente mediante un modelo de clasificación supervisada. Este enfoque permite avanzar desde una perspectiva descriptiva hacia un análisis predictivo, facilitando la estimación del comportamiento esperado de los consumidores ante un escenario de shock energético.

El uso de técnicas de Machine Learning se justifica por la necesidad de analizar relaciones complejas entre múltiples variables que influyen en la toma de decisiones del consumidor. A diferencia de los enfoques puramente descriptivos, el Machine Learning permite identificar patrones en los datos y estimar probabilidades de comportamiento bajo distintos escenarios, lo que resulta especialmente útil en contextos de incertidumbre y en situaciones no observables directamente en la realidad, como el escenario planteado en este estudio.

En este sentido, el problema de investigación se plantea como un problema de **clasificación**, dado que la variable objetivo es de naturaleza binaria: el consumidor decide o no cambiar su fuente energética. Este tipo de decisiones discretas se modelizan de forma adecuada mediante técnicas de clasificación, que permiten asignar probabilidades a cada una de las posibles categorías y analizar qué factores influyen en dicha decisión.

Dentro de las distintas técnicas disponibles, se ha optado por utilizar un modelo de **regresión logística**. Esta elección responde a varias razones:

- En primer lugar, la regresión logística es especialmente adecuada para problemas de clasificación binaria, ya que permite estimar de forma directa la probabilidad de ocurrencia de un evento.

- En segundo lugar, se trata de un modelo altamente interpretable, lo que resulta fundamental en el contexto de este trabajo, donde no solo se pretende predecir el comportamiento del consumidor, sino también comprender el peso relativo de cada variable en la toma de decisiones.

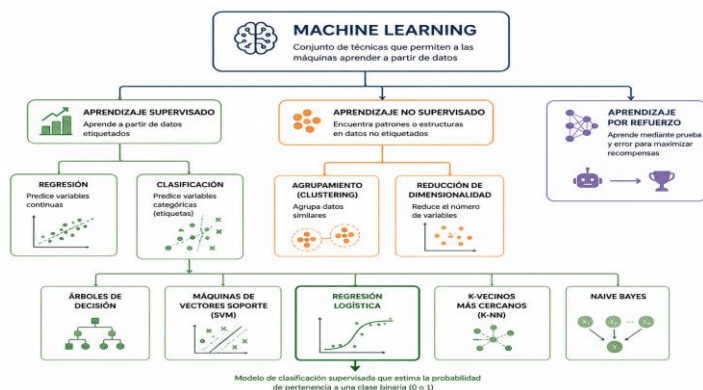
Asimismo, la regresión logística se alinea con el enfoque económico del estudio, ya que permite analizar la relación entre variables explicativas y la probabilidad de un determinado comportamiento, facilitando la interpretación de los resultados en términos de causalidad aproximada. Además, su simplicidad relativa la convierte en una herramienta adecuada para trabajar con datasets de tamaño reducido, como el utilizado en este trabajo, evitando problemas de sobreajuste que podrían surgir con modelos más complejos.

La Figura III muestra la clasificación de las principales técnicas de Machine Learning, destacando la regresión logística como un modelo de clasificación supervisada utilizado para predecir variables categóricas binarias.

Entre sus principales ventajas destacan:

- Permite estimar probabilidades de forma directa
- Es fácilmente interpretable
- Facilita el análisis del peso relativo de cada variable
- Es coherente con el enfoque económico del estudio

Figura III. Principales Técnicas de Machine Learning



En conjunto, la combinación de técnicas de Machine Learning y, en particular, el uso de la regresión logística, permite abordar el problema de investigación desde una perspectiva cuantitativa, interpretativa y coherente con los objetivos del estudio, proporcionando una base sólida para el análisis del comportamiento del consumidor en contextos de incertidumbre energética.

## 5.2 Preparación de los datos

Antes de la aplicación del modelo de clasificación (regresión logística), resulta necesario realizar una fase de preparación de los datos, con el objetivo de garantizar la coherencia del dataset y su adecuación para el análisis mediante técnicas de Machine Learning.

En primer lugar, se ha trabajado con un dataset sintético compuesto por 500 observaciones, donde cada fila representa un hogar de la Comunidad de Madrid. Este dataset incluye variables de distinta naturaleza, tanto numéricas como categóricas, lo que requiere un tratamiento específico para su correcta utilización en el modelo.

Las **variables numéricas**, como la renta anual, el gasto energético, el coste de transición o el nivel de incertidumbre, se han mantenido en su formato original, ya que la regresión

logística no requiere necesariamente procesos de normalización para su correcta aplicación en este contexto. No obstante, se ha verificado que estas variables se encuentren dentro de rangos y sin valores extremos que puedan distorsionar el modelo.

Por otro lado, las **variables categóricas**, como el tipo de vivienda o el sistema de calefacción, han sido transformadas en variables binarias mediante procesos de codificación, permitiendo su incorporación en el modelo de forma adecuada. En particular, variables como la dependencia del gas se han representado directamente en formato binario, facilitando su interpretación y análisis.

Asimismo, se ha prestado especial atención a la **coherencia interna del dataset**, verificando que las relaciones entre variables se mantengan dentro de los supuestos definidos en la fase de construcción. Por ejemplo, se ha comprobado que el coste de transición sea mayor en hogares dependientes del gas o que la vulnerabilidad económica esté relacionada con el nivel de renta y el gasto energético.

En relación con la variable objetivo, se ha analizado su distribución con el fin de evitar problemas de desbalanceo. En este sentido, se ha comprobado que existe una proporción razonable entre hogares que deciden cambiar de fuente energética y aquellos que no lo hacen, lo que permite aplicar el modelo sin necesidad de técnicas adicionales de balanceo.

Adicionalmente, el dataset se ha dividido en dos subconjuntos: un conjunto de entrenamiento y un conjunto de validación, siguiendo una proporción estándar (por ejemplo, **80% entrenamiento y 20% validación**). Esta división permite evaluar la capacidad predictiva del modelo y evitar problemas de sobreajuste, garantizando que los resultados obtenidos sean generalizables dentro del contexto del estudio.

En conjunto, esta fase de preparación de los datos permite asegurar que el dataset sea consistente, interpretable y adecuado para la aplicación del modelo de regresión logística, constituyendo un paso fundamental dentro del proceso de análisis.

### 5.3 Modelo: Regresión logística

Como técnica principal de análisis, se emplea un modelo de **regresión logística**<sup>16</sup>, ampliamente utilizado en problemas de clasificación binaria. Este modelo permite estimar la probabilidad de que ocurra un determinado evento (en este caso, el cambio de fuente energética) en función de un conjunto de variables explicativas.

La regresión logística se basa en la estimación de una función logística que transforma una combinación lineal de variables en una probabilidad comprendida entre 0 y 1:

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}}$$

Donde:

- $P(Y = 1)$  representa la probabilidad de que el hogar decida cambiar de fuente energética
- $X_1, X_2, \dots, X_n$  son las variables explicativas incluidas en el modelo
- $\beta_0, \beta_1, \dots, \beta_n$  son los coeficientes estimados

Desde un punto de vista interpretativo, los coeficientes del modelo permiten analizar la dirección y magnitud del efecto de cada variable sobre la probabilidad de cambio. Un coeficiente positivo indica que el aumento de la variable incrementa la probabilidad de cambio, mientras que un coeficiente negativo indica el efecto contrario.

En el contexto del presente estudio, se espera que variables como el incremento del precio energético o la exposición a la información tengan un efecto positivo sobre la probabilidad

---

<sup>16</sup> La **regresión logística** es un modelo estadístico utilizado para analizar la relación entre una variable dependiente categórica (generalmente binaria) y un conjunto de variables explicativas. A diferencia de la regresión lineal, este modelo estima la probabilidad de que ocurra un determinado evento mediante una función logística, que transforma una combinación lineal de variables en valores comprendidos entre 0 y 1, permitiendo interpretar los resultados en términos probabilísticos (Hosmer, Lemeshow & Sturdivant, 2013; James et al., 2021).

de cambio, mientras que factores como el coste de transición o la incertidumbre percibida ejerzan un efecto negativo.

El modelo se ha estimado utilizando el conjunto de entrenamiento previamente definido, y posteriormente se ha evaluado su capacidad predictiva mediante el conjunto de validación.

La elección de la regresión logística responde a su equilibrio entre simplicidad e interpretabilidad. A diferencia de modelos más complejos, como redes neuronales o métodos ensemble, la regresión logística permite identificar de forma clara el impacto de cada variable, lo que resulta especialmente relevante en un estudio cuyo objetivo no es únicamente predecir, sino también comprender el comportamiento del consumidor.

No obstante, es importante señalar que este modelo presenta ciertas limitaciones, como la asunción de relaciones lineales entre las variables explicativas y el logaritmo de las probabilidades. A pesar de ello, su adecuación al problema planteado y su coherencia con el enfoque económico del estudio lo convierten en una herramienta apropiada para el análisis desarrollado.

## 5.4 Evaluación del modelo

Métricas de evaluación: El rendimiento del modelo se evalúa mediante accuracy<sup>17</sup>, precision, recall<sup>18</sup>, F1-score<sup>19</sup> y ROC-AUC<sup>20</sup>. Además, se analiza la matriz de confusión<sup>21</sup> para observar el reparto entre aciertos y errores de clasificación.

---

<sup>17</sup> La **precisión (accuracy)** es una métrica utilizada para evaluar el rendimiento de un modelo de clasificación, que mide la proporción de predicciones correctas sobre el total de observaciones. Se calcula como el número de aciertos (tanto verdaderos positivos como verdaderos negativos) dividido entre el número total de casos. Esta métrica ofrece una visión general del desempeño del modelo, aunque puede resultar limitada en situaciones donde las clases están desbalanceadas.

<sup>18</sup> El **recall** es una métrica que mide la capacidad del modelo para identificar correctamente los casos positivos reales, es decir, el porcentaje de hogares que efectivamente cambian de sistema energético y que el modelo logra detectar correctamente.

<sup>19</sup> El F1-score representa una medida de equilibrio entre la precisión (precision) y el recall, siendo especialmente útil cuando existe desbalance entre clases o cuando se busca evaluar simultáneamente la capacidad de detección y la fiabilidad de las predicciones.

<sup>20</sup> El **ROC-AUC** (Receiver Operating Characteristic - Area Under the Curve) evalúa la capacidad general del modelo para diferenciar entre las distintas clases, indicando qué tan bien puede separar los hogares con probabilidad de cambio energético de aquellos que no presentan dicha probabilidad. Valores más cercanos a 1 indican un mejor rendimiento predictivo del modelo.

<sup>21</sup> La **matriz de confusión** es una herramienta utilizada para evaluar el rendimiento de un modelo de clasificación, permitiendo comparar las predicciones realizadas por el modelo con los valores reales observados. Esta matriz muestra de forma estructurada las clasificaciones correctas e incorrectas para cada una de las categorías analizadas, facilitando la identificación de aciertos y errores del modelo. Asimismo, permite calcular métricas de evaluación como accuracy, precision, recall y F1-score, las cuales ayudan a analizar la capacidad predictiva y clasificatoria del modelo desarrollado.

## **CAPÍTULO VI: RESULTADOS**

Los principales resultados obtenidos a partir del dataset sintético y de la aplicación del modelo de regresión logística muestran en primer lugar, los estadísticos descriptivos y las relaciones generales entre variables, con el objetivo de identificar patrones básicos de comportamiento en los hogares simulados.

Posteriormente, se examina el impacto del shock energético sobre la probabilidad de cambio de fuente energética, así como las diferencias existentes entre distintos perfiles de hogares.

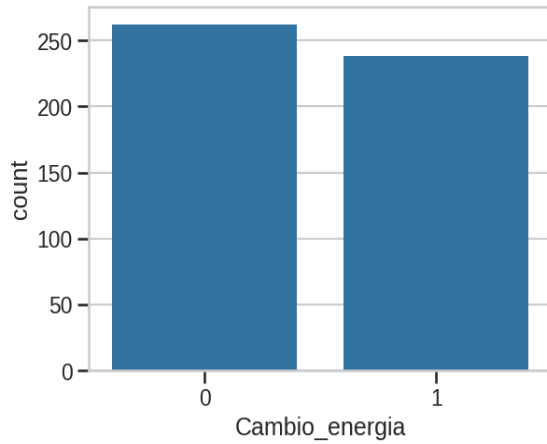
Finalmente, se evalúa el rendimiento del modelo predictivo mediante métricas de clasificación y se interpretan los coeficientes obtenidos, con el fin de identificar las variables con mayor influencia en la toma de decisiones de los consumidores en un contexto de incertidumbre energética.

### **6.1 Estadísticos descriptivos y patrones generales**

Con el objetivo de comprender la estructura general del dataset y los principales patrones presentes en los hogares simulados, se realizó un análisis descriptivo de las variables incluidas en el estudio. Este análisis permite identificar tendencias iniciales en el comportamiento de los consumidores ante el escenario de shock energético, así como explorar las relaciones existentes entre variables económicas, estructurales y conductuales.

En primer lugar, se analizó la distribución de la variable objetivo “Cambio\_energia”, la cual representa la decisión del hogar de sustituir o no su fuente energética en el contexto de eliminación del gas. Los resultados muestran que el dataset presenta una distribución relativamente equilibrada entre hogares que deciden realizar la transición energética y aquellos que mantienen su sistema actual, lo que resulta adecuado para la aplicación posterior del modelo de clasificación. La Figura IV muestra la distribución de la variable objetivo en el conjunto de datos simulado.

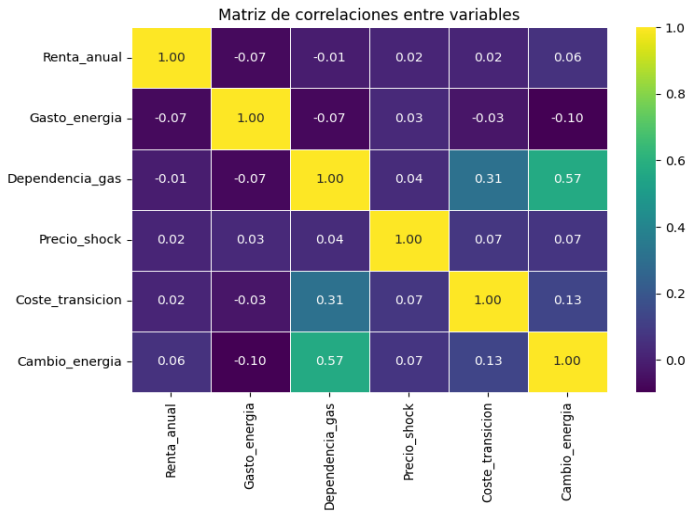
Figura IV. Distribución Variable Objetivo



A nivel general, puede observarse que la decisión de cambio energético no se distribuye de forma aleatoria, sino que responde a la combinación de distintos factores económicos y estructurales incorporados en el proceso de simulación. Esto sugiere que variables como el precio energético, la dependencia del gas o la capacidad económica del hogar desempeñan un papel relevante en la toma de decisiones.

Posteriormente, se analizaron las principales correlaciones entre variables, con el objetivo de identificar relaciones lineales significativas dentro del dataset. Para ello, se construyó una matriz de correlación entre las variables numéricas más relevantes del estudio. La Figura V presenta el mapa de correlaciones obtenido.

Figura V. Mapa de correlaciones



Los resultados de la matriz de correlaciones permiten identificar distintas relaciones entre las variables incluidas en el dataset. En términos generales, se observan asociaciones moderadas entre variables económicas y estructurales, aunque no todas las relaciones presentan una intensidad elevada.

En particular, la variable “Cambio\_energia” muestra relaciones más visibles con factores como la dependencia del gas y determinadas variables económicas; mientras que la relación directa con el precio del shock energético resulta más limitada de lo esperado. Esto sugiere que, dentro del escenario simulado, la decisión de cambio no depende exclusivamente del incremento del precio energético, sino también de la interacción entre múltiples factores, como la capacidad económica, el coste de transición o las barreras asociadas al proceso de adaptación.

Asimismo, se observa una relación negativa entre el coste de transición y la probabilidad de cambio energético, indicando que mayores costes de adaptación pueden actuar como una barrera relevante para la sustitución energética. Del mismo modo, las relaciones entre renta, vulnerabilidad económica y gasto energético muestran patrones coherentes con los supuestos planteados en la construcción del dataset.

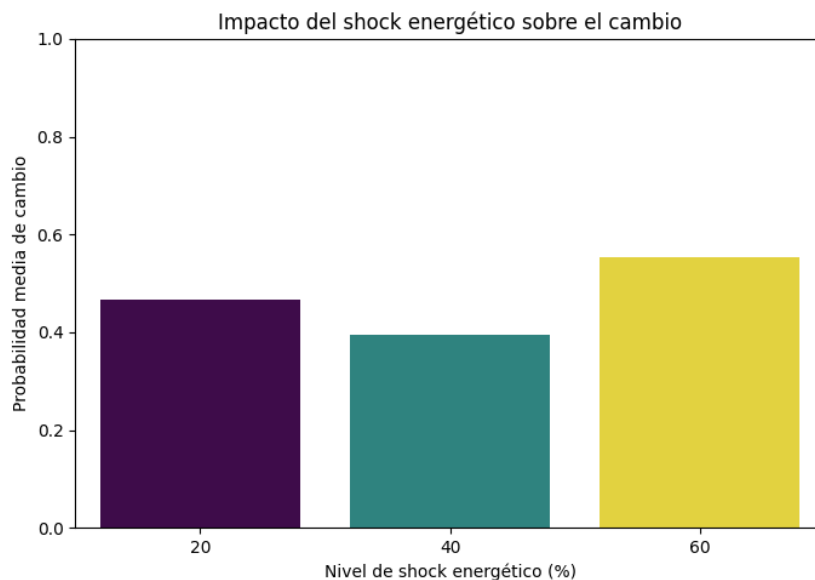
En conjunto, la matriz de correlaciones refleja que el comportamiento de los hogares responde a una combinación de factores económicos, estructurales y conductuales, lo que refuerza la complejidad del proceso de toma de decisiones en contextos de incertidumbre energética.

## 6.2 Influencia del shock de precios

Con el objetivo de analizar el efecto del shock energético sobre el comportamiento de los hogares, se evaluó la relación entre el incremento del precio energético, los costes de transición y la probabilidad de sustitución del sistema energético. Este análisis permite observar cómo los hogares reaccionan ante distintos niveles de presión económica y qué factores condicionan su capacidad de adaptación.

### 6.2.1 Efecto del precio shock

Figura VI. Impacto del shock energético sobre el cambio



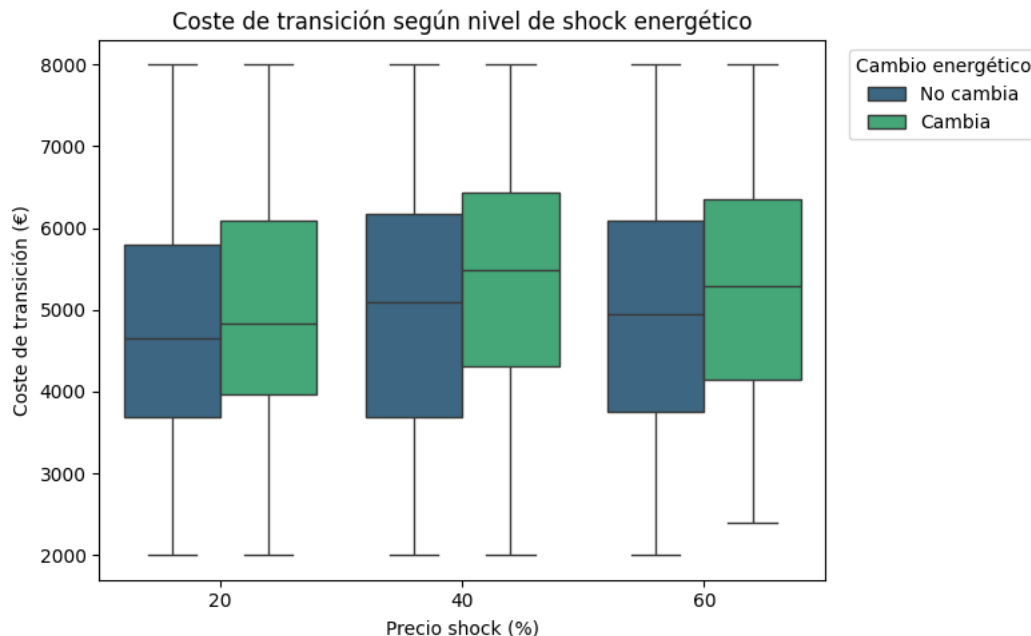
La Figura VI muestra la probabilidad media de cambio energético en función del nivel de shock simulado. En términos generales, se observa que el escenario de mayor intensidad correspondiente a un incremento del 60% en el precio energético y presenta la probabilidad media de cambio más elevada. Este resultado sugiere que un aumento significativo del precio energético puede actuar como un incentivo relevante para que los hogares consideren sustituir su fuente energética.

No obstante, la relación observada no es completamente lineal, ya que el escenario del 40% presenta una probabilidad media de cambio inferior a la observada en el escenario del 20%. Este comportamiento indica que el precio del shock, por sí solo, no explica de manera aislada la decisión de cambio energético. La respuesta de los hogares depende también de otros factores incorporados en el dataset, como la dependencia del gas, la renta disponible, el coste de transición, la incertidumbre percibida y la exposición a la información.

En consecuencia, los resultados sugieren que el shock de precios ejerce presión sobre la decisión de transición energética, aunque su impacto se encuentra condicionado por la capacidad económica y las barreras estructurales de cada hogar.

### 6.2.2 Relación entre precio y coste de transición

Figura VII. Coste de transición según nivel de shock energético



La Figura VII presenta la distribución del coste de transición según el nivel de shock energético y diferenciando entre hogares que realizan el cambio energético y aquellos que mantienen su sistema actual. En términos generales se observa que el coste de transición presenta una elevada dispersión en los tres escenarios analizados, lo que refleja la heterogeneidad económica existente entre los hogares simulados. Asimismo, los hogares que realizan la transición energética tienden a concentrarse en niveles de coste ligeramente superiores en comparación con aquellos que no cambian de sistema energético, especialmente en los escenarios de shock del 40% y 60%.

No obstante, las diferencias entre grupos no son completamente homogéneas, lo que sugiere que el coste de transición no actúa como único factor explicativo de la decisión de cambio. La probabilidad de sustitución energética parece depender de la interacción entre múltiples

variables, como la renta disponible, la vulnerabilidad económica, la dependencia del gas y el acceso a información sobre alternativas energéticas.

Por otro lado, el incremento del nivel de shock energético no elimina la dispersión observada en los costes de transición, indicando que incluso bajo escenarios de elevada presión económica continúan existiendo hogares con limitaciones relevantes para afrontar el proceso de adaptación energética.

En conjunto, los resultados muestran que la transición energética se encuentra condicionada tanto por incentivos económicos derivados del aumento de precios como por la capacidad estructural de los hogares para asumir los costes asociados al cambio de sistema energético.

### *6.2.3 Impacto del precio sobre la probabilidad de cambio*

Los resultados obtenidos permiten observar que el incremento del precio energético afecta de forma significativa la probabilidad de cambio de sistema energético, aunque dicho efecto varía entre hogares según sus características económicas y estructurales.

En los escenarios de mayor shock energético, el modelo tiende a asignar probabilidades más elevadas de transición energética, lo que indica que el aumento del coste asociado al consumo de gas incrementa el incentivo económico para considerar alternativas energéticas. Sin embargo, la variabilidad observada entre hogares demuestra que la respuesta al shock no es uniforme.

En este sentido, algunos hogares mantienen probabilidades reducidas de cambio incluso bajo escenarios de elevada presión económica, lo que puede explicarse por la presencia de barreras como altos costes de transición, limitaciones de renta o elevados niveles de incertidumbre. Por el contrario, otros hogares presentan una mayor capacidad de adaptación, aumentando significativamente su probabilidad de sustitución energética.

Estos resultados reflejan que la transición energética no depende exclusivamente del aumento de precios, sino de la interacción entre factores económicos, conductuales y estructurales.

Asimismo, evidencian que los efectos de un shock energético pueden generar respuestas heterogéneas dentro de la población, reforzando la importancia de considerar las diferencias socioeconómicas en el diseño de políticas públicas orientadas a facilitar procesos de transición energética.

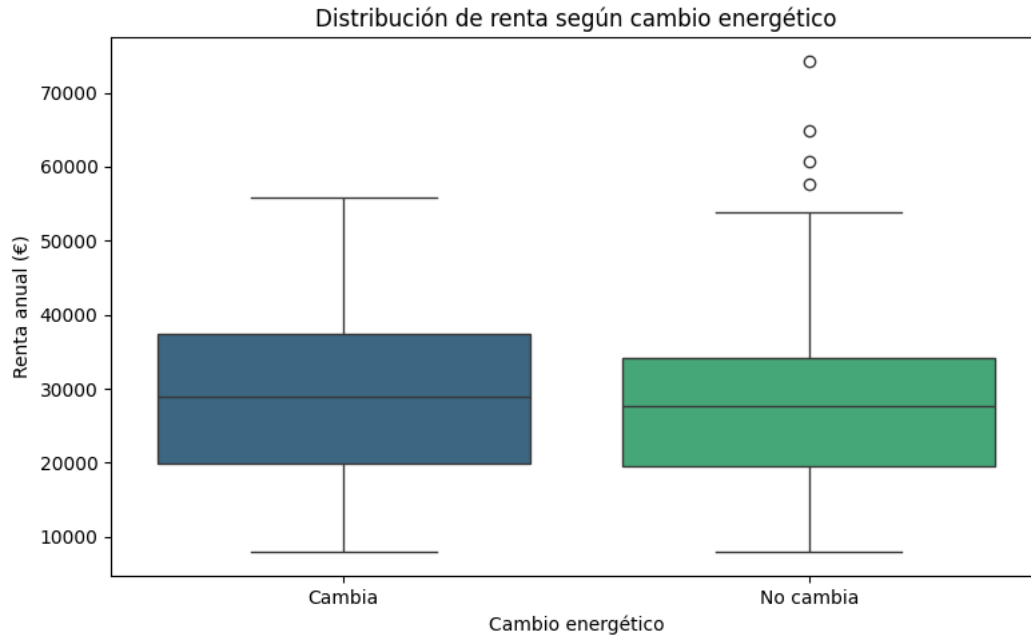
### **6.3 Diferencias por perfil de hogar**

Con el objetivo de profundizar en las diferencias existentes entre los hogares simulados, se analizó el comportamiento de distintos perfiles socioeconómicos y estructurales frente al escenario de shock energético. Este análisis permite identificar cómo variables como la renta, la vulnerabilidad económica, el tipo de vivienda, el sistema de calefacción y el acceso a información influyen sobre la probabilidad de transición energética.

Los resultados muestran que la respuesta de los hogares no es homogénea, sino que depende de múltiples factores económicos y estructurales que condicionan la capacidad de adaptación ante un contexto de incremento de precios energéticos.

### 6.3.1 Comparación entre renta alta y baja

Figura VIII. Distribución de renta según cambio energético



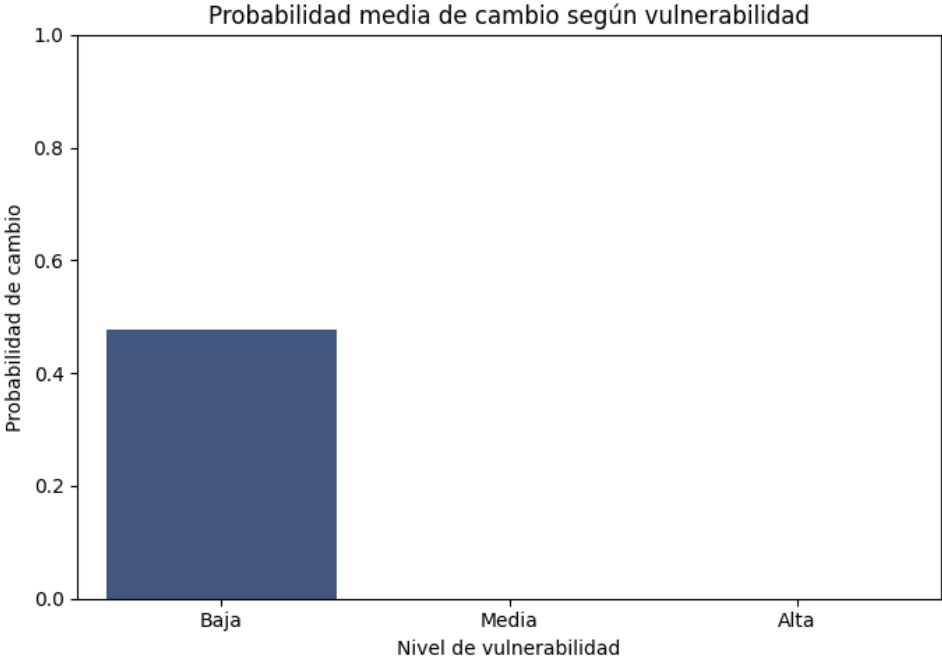
La Figura VIII presenta la distribución de la renta anual diferenciando entre hogares que realizan el cambio energético y aquellos que mantienen su sistema actual. En términos generales se observa que los hogares que realizan la transición energética tienden a concentrarse en niveles de renta ligeramente superiores en comparación con aquellos que no cambian de sistema energético. Este comportamiento sugiere que una mayor capacidad económica facilita la posibilidad de asumir los costes asociados a la adaptación energética.

Asimismo, los resultados reflejan una elevada dispersión en ambos grupos, indicando que la renta no actúa como único factor explicativo de la decisión de cambio. Existen hogares con rentas medias o incluso bajas que realizan la transición energética, así como hogares de renta elevada que mantienen su sistema energético original.

En conjunto, estos resultados sugieren que la capacidad económica influye sobre la probabilidad de adaptación energética, aunque dicha relación se encuentra condicionada también por otras variables estructurales y conductuales incluidas en el modelo.

6.3.2 Diferencias por vulnerabilidad

Figura IX. Probabilidad media de cambio según vulnerabilidad



La Figura IX muestra la probabilidad media de cambio energético según el nivel de vulnerabilidad económica de los hogares. Los resultados indican que la vulnerabilidad económica presenta una relación relevante con la probabilidad de transición energética. Los hogares con mayores niveles de vulnerabilidad muestran patrones de comportamiento diferenciados frente al shock energético, reflejando las dificultades económicas asociadas al proceso de adaptación.

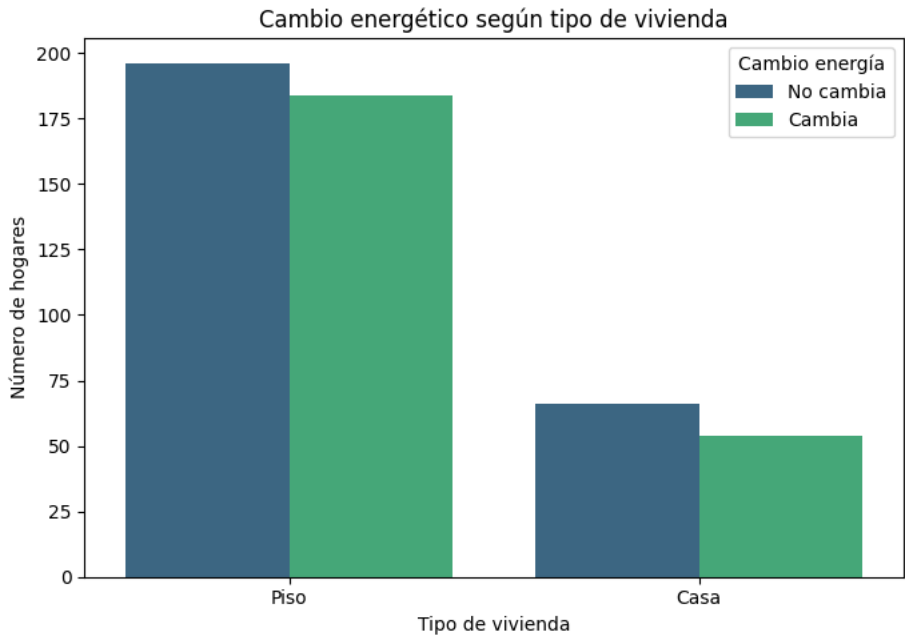
En este sentido, la vulnerabilidad puede actuar simultáneamente como un factor de presión y como una barrera para el cambio energético. Por un lado, los hogares vulnerables son más

sensibles al incremento de precios energéticos; sin embargo, también pueden enfrentar mayores limitaciones para asumir los costes de transición necesarios para sustituir su sistema energético.

Estos resultados ponen de manifiesto la existencia de desigualdades en la capacidad de adaptación energética, reforzando la importancia de considerar mecanismos de apoyo específicos para hogares económicamente vulnerables dentro de procesos de transición energética.

6.3.3 Diferencias por tipo de vivienda

Figura X. Cambio energético según tipo de vivienda



La Figura X presenta la distribución de hogares que realizan o no la transición energética según el tipo de vivienda. Los resultados muestran diferencias entre hogares que residen en pisos y aquellos que habitan en casas unifamiliares. Estas diferencias pueden estar

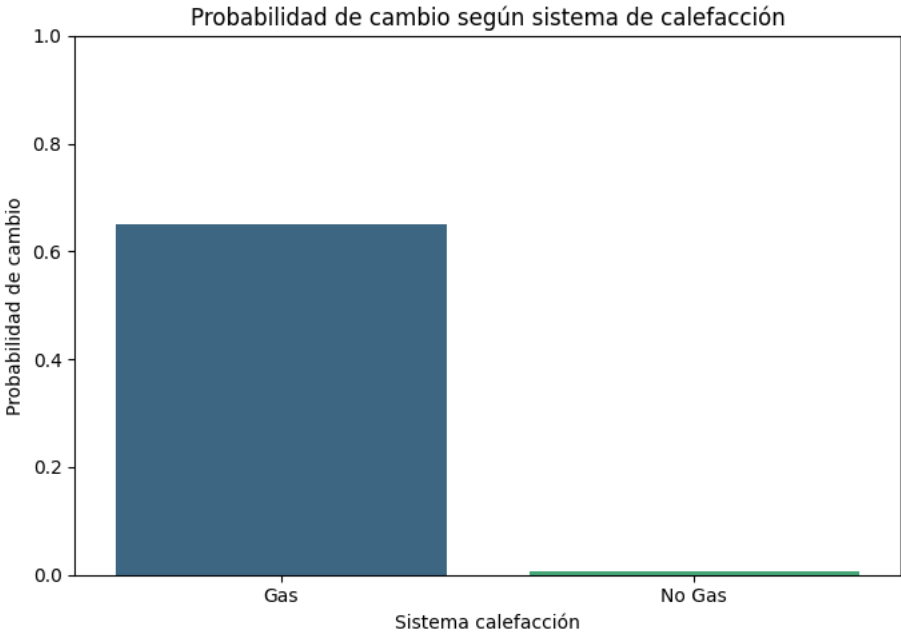
relacionadas con factores como el tamaño de la vivienda, las necesidades energéticas asociadas al sistema de calefacción y los costes de adaptación estructural.

En términos generales, las viviendas unifamiliares tienden a presentar mayores costes potenciales de transición energética debido a una mayor superficie energética y a sistemas de calefacción más complejos. Por el contrario, los pisos pueden presentar menores barreras estructurales para determinadas alternativas energéticas.

No obstante, las diferencias observadas no son absolutas, lo que indica que el tipo de vivienda constituye únicamente uno de los múltiples factores que influyen en el comportamiento energético de los hogares.

6.3.4 Diferencias por sistema de calefacción

Figura XI. Probabilidad de cambio según el sistema de calefacción



La Figura XI muestra la probabilidad de cambio energético según el sistema de calefacción utilizado por los hogares. Los resultados evidencian diferencias significativas entre hogares

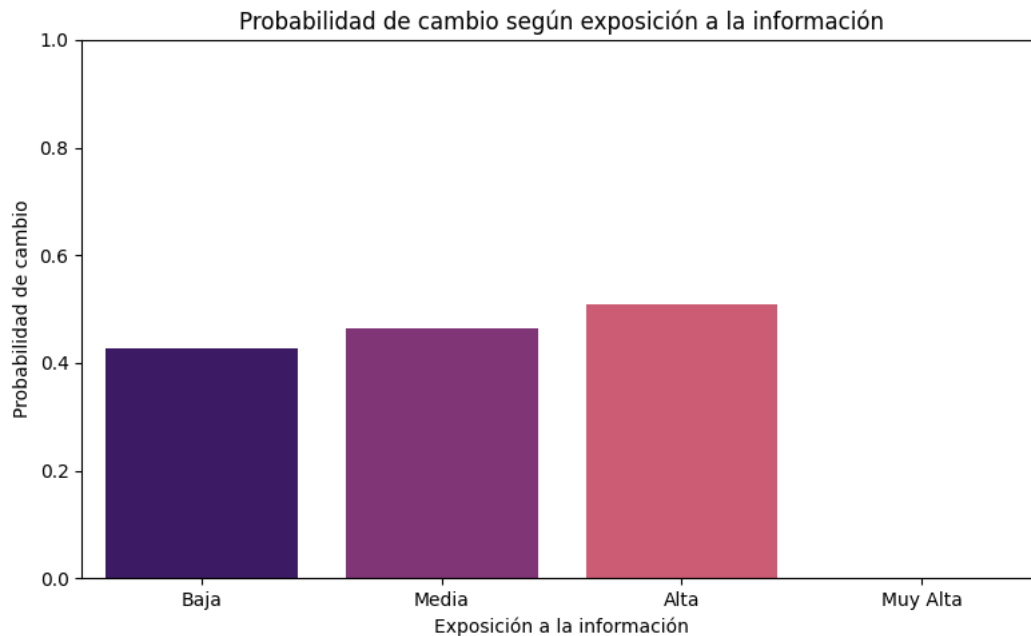
con dependencia del gas y hogares que utilizan otros sistemas energéticos. En términos generales, los hogares con calefacción basada en gas presentan una mayor probabilidad de transición energética, lo que resulta coherente con el escenario de shock planteado en el estudio.

Este comportamiento sugiere que la exposición directa al incremento del precio del gas incrementa el incentivo económico para considerar alternativas energéticas. Por el contrario, los hogares que ya utilizan sistemas no dependientes del gas muestran una menor necesidad de adaptación ante el shock energético simulado.

En consecuencia, la dependencia estructural del sistema energético inicial constituye uno de los factores más relevantes dentro del proceso de decisión analizado por el modelo.

### 6.3.5 Diferencias por exposición a la información

Figura XII. Probabilidad de cambio según exposición a la información



La Figura XII presenta la probabilidad media de cambio energético según el nivel de exposición a la información. Los resultados muestran que los hogares con mayores niveles de acceso a información presentan, en términos generales, una mayor probabilidad de transición energética. Este comportamiento sugiere que el conocimiento sobre alternativas energéticas, eficiencia y costes potenciales puede influir de manera positiva sobre la disposición al cambio.

Por el contrario, niveles reducidos de exposición a la información parecen estar asociados a una menor capacidad de adaptación, posiblemente debido a una mayor percepción de incertidumbre o desconocimiento sobre las opciones disponibles.

Estos resultados resultan coherentes con los planteamientos de la economía del comportamiento, según los cuales la percepción del riesgo, el acceso a información y la incertidumbre condicionan significativamente el proceso de toma de decisiones de los consumidores.

En conjunto, el análisis por perfiles de hogar refleja que la transición energética afecta de manera desigual a la población, dependiendo de factores económicos, estructurales y conductuales que condicionan la capacidad de adaptación frente a escenarios de shock energético.

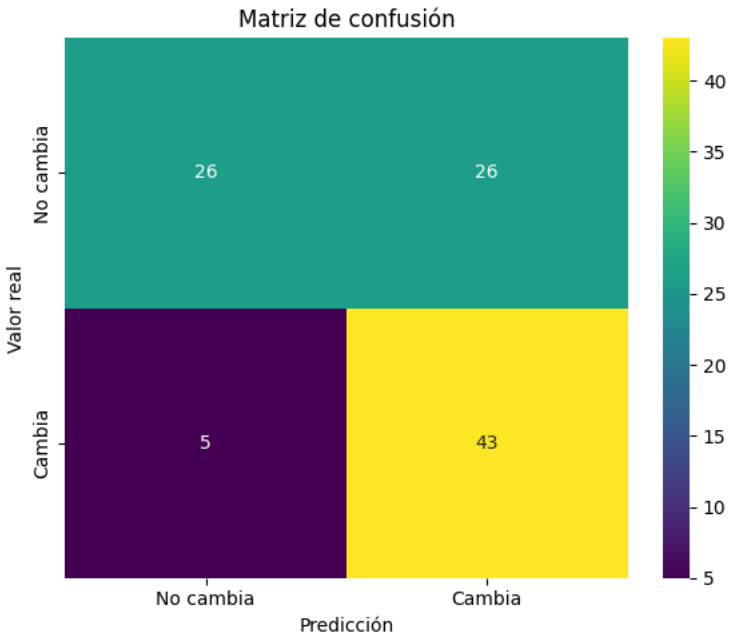
#### **6.4 Evaluación del modelo**

Con el objetivo de evaluar la capacidad predictiva del modelo de regresión logística, se analizaron distintas métricas de clasificación y herramientas de evaluación propias del aprendizaje supervisado. Para ello, el dataset fue dividido en un conjunto de entrenamiento y un conjunto de prueba mediante una proporción 80/20. De esta forma, el modelo fue entrenado utilizando el 80% de las observaciones y posteriormente evaluado sobre el 20% restante, equivalente a 100 hogares simulados.

Este procedimiento permite analizar la capacidad predictiva del modelo sobre datos no utilizados durante el entrenamiento, reduciendo el riesgo de sobreajuste y proporcionando una evaluación más realista del desempeño del modelo.

6.4.1 Matriz de confusión

Figura XIII. Matriz de Confusión



La Figura XIII presenta la matriz de confusión obtenida a partir de las predicciones realizadas sobre el conjunto de prueba. La matriz de confusión permite comparar las predicciones generadas por el modelo con los valores reales observados en el dataset de prueba. Dado que la evaluación se realiza sobre 100 hogares simulados, los valores representados en la matriz corresponden únicamente al subconjunto reservado para validación y no al total de 500 hogares generados en el estudio.

Los resultados muestran que el modelo logra identificar correctamente una gran proporción de hogares que realizan el cambio energético, obteniendo un número elevado de verdaderos positivos. En particular, el modelo clasifica correctamente 43 hogares que efectivamente realizan la transición energética, mientras que únicamente 5 hogares que cambian son clasificados incorrectamente como “No cambia”. Este comportamiento refleja una elevada sensibilidad del modelo para detectar la clase positiva.

No obstante, también se observa un número considerable de falsos positivos, ya que 26 hogares son clasificados como “Cambia” cuando realmente no realizan la transición energética. Esto indica que el modelo presenta cierta tendencia a sobreestimar la probabilidad de cambio energético en determinados perfiles de hogar.

Por otro lado, el modelo identifica correctamente 26 hogares que mantienen su sistema energético original, mostrando una capacidad moderada para reconocer la clase negativa.

En conjunto, los resultados sugieren que el modelo presenta una capacidad razonable para identificar patrones generales de transición energética, aunque mantiene limitaciones asociadas a la complejidad del comportamiento de los consumidores y a la naturaleza sintética del dataset utilizado.

#### 6.4.2 Métricas de rendimiento del modelo

La Tabla 2 resume las principales métricas utilizadas para evaluar el desempeño del modelo de regresión logística.

Tabla II. Principales Métricas

	<b>Metrica</b>	<b>Valor</b>
0	Accuracy	0,690
1	Precision	0,623

2	Recall	0,896
3	F1-score	0,735
4	ROC-AUC	0,735

El modelo obtuvo una *accuracy* de 0,690, indicando que aproximadamente el 69% de las observaciones del conjunto de prueba fueron clasificadas correctamente. Este resultado refleja una capacidad predictiva moderada, razonable dentro de un contexto de comportamiento humano influido por múltiples factores económicos y conductuales.

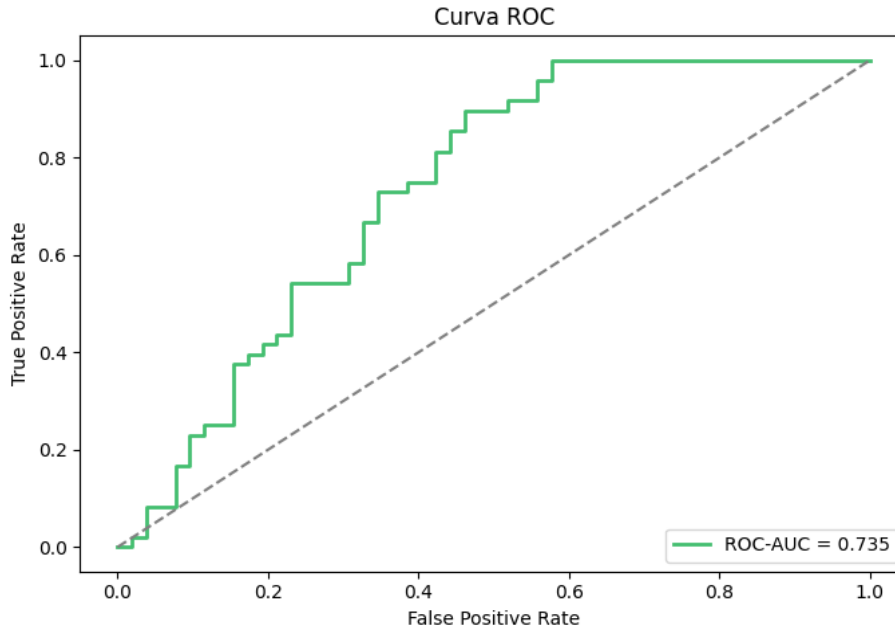
La métrica *recall* alcanzó un valor elevado de 0,896, mostrando que el modelo identifica correctamente una gran proporción de hogares que efectivamente realizan la transición energética. Este resultado constituye uno de los principales puntos fuertes del modelo, ya que demuestra una elevada capacidad para detectar hogares con probabilidad de cambio energético.

Por otro lado, la *precision* obtuvo un valor de 0,623, indicando que parte de las predicciones positivas realizadas por el modelo corresponden a hogares que finalmente no cambian de sistema energético. Este comportamiento resulta coherente con los falsos positivos observados en la matriz de confusión.

Asimismo, el modelo obtuvo un *F1-score* de 0,735, reflejando un equilibrio razonable entre precisión y sensibilidad. Finalmente, el valor *ROC-AUC* de 0,735 indica que el modelo presenta una capacidad discriminatoria adecuada para diferenciar entre hogares que cambian y hogares que no cambian de sistema energético.

### 6.4.3 Capacidad clasificatoria general

Figura XIV. Curva ROC



La Figura XIV presenta la curva ROC obtenida para el modelo de regresión logística. La curva ROC permite analizar la capacidad del modelo para distinguir entre ambas clases a diferentes umbrales de clasificación. En este caso, la curva se sitúa claramente por encima de la línea diagonal correspondiente a una clasificación aleatoria, indicando que el modelo posee capacidad predictiva superior al azar.

El valor ROC-AUC de 0,735 confirma que el modelo logra capturar patrones relevantes asociados al comportamiento energético de los hogares, aunque sin alcanzar niveles de clasificación extremadamente elevados. Este resultado es coherente con la complejidad inherente al fenómeno analizado, donde las decisiones de transición energética dependen simultáneamente de variables económicas, estructurales y conductuales.

En conjunto, los resultados obtenidos muestran que la regresión logística constituye una herramienta útil para analizar patrones generales de comportamiento energético dentro del

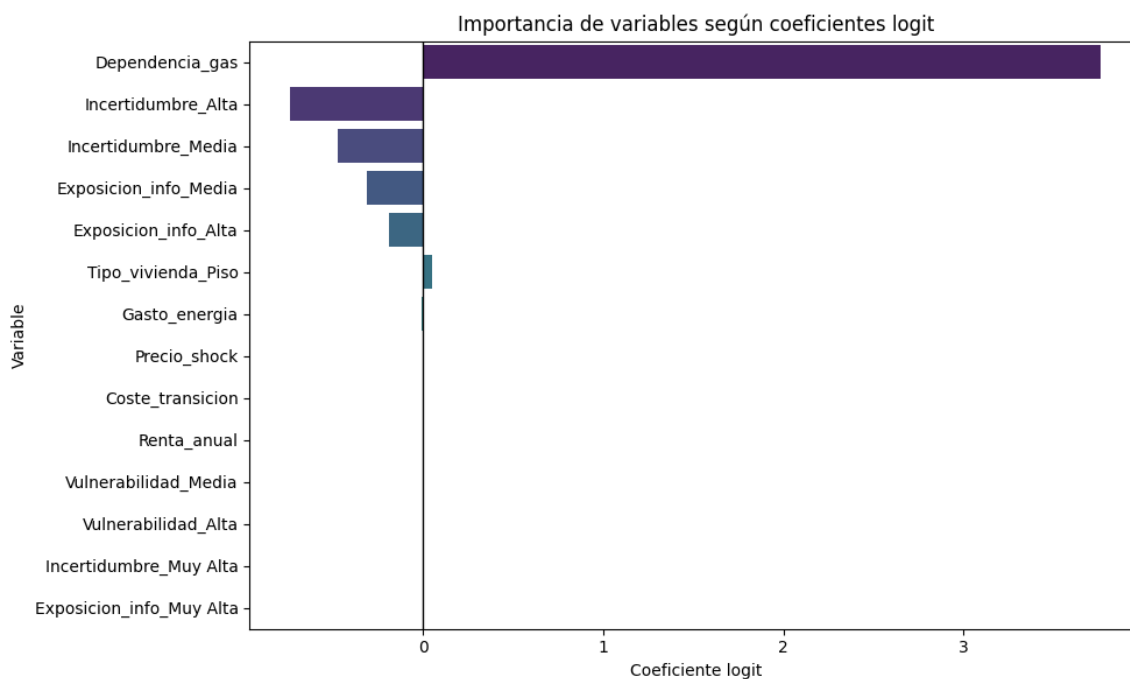
escenario simulado. Aunque el modelo no captura completamente toda la complejidad del proceso de toma de decisiones de los hogares, sí permite identificar relaciones relevantes entre variables asociadas a la transición energética y estimar probabilidades de cambio de forma razonablemente consistente.

### **6.5 Interpretación de coeficientes**

Con el objetivo de comprender la influencia de las distintas variables sobre la probabilidad de transición energética, se analizaron los coeficientes obtenidos por el modelo de regresión logística. En este tipo de modelos, los coeficientes permiten identificar tanto la dirección como la intensidad de la relación entre cada variable explicativa y la probabilidad de que un hogar realice el cambio energético.

En términos generales, un coeficiente positivo indica que el aumento o presencia de una variable incrementa la probabilidad de transición energética, mientras que un coeficiente negativo sugiere un efecto contrario, reduciendo la probabilidad de cambio.

Figura XV. Importancia de variables según coeficientes logit



La Figura XV presenta la importancia relativa de las variables incluidas en el modelo, representada mediante los coeficientes estimados por la regresión logística.

### 6.5.1 Signo de los coeficientes

Los resultados muestran que la variable con mayor influencia positiva sobre la probabilidad de transición energética es la dependencia del gas. El coeficiente asociado a esta variable presenta una magnitud significativamente superior al resto de variables del modelo, indicando que los hogares dependientes del gas poseen una probabilidad considerablemente mayor de realizar el cambio energético dentro del escenario simulado.

Este resultado refleja que la estructura energética inicial del hogar constituye un elemento determinante en el comportamiento observado ante el shock energético. En consecuencia, los hogares previamente dependientes del gas presentan una mayor exposición al escenario de

incremento de precios y, por tanto, una mayor predisposición a modificar su sistema energético.

Por otro lado, las variables relacionadas con la incertidumbre muestran coeficientes negativos relevantes, especialmente en las categorías de incertidumbre alta y media. Este comportamiento sugiere que niveles elevados de incertidumbre reducen la disposición de los hogares a realizar cambios energéticos, posiblemente debido al aumento de percepción de riesgo o a la dificultad para anticipar costes y beneficios futuros asociados a la transición.

Asimismo, las variables vinculadas a la exposición a la información presentan efectos positivos moderados sobre la probabilidad de cambio energético. Esto indica que un mayor acceso a información puede favorecer procesos de adaptación energética, facilitando la comprensión del contexto energético y reduciendo parcialmente la incertidumbre percibida.

En contraste, variables económicas tradicionales como la renta anual, el precio del shock energético o el coste de transición presentan coeficientes relativamente reducidos dentro del modelo. Aunque estas variables mantienen relación con el fenómeno estudiado, su capacidad explicativa resulta considerablemente menor en comparación con la fuerte influencia ejercida por la dependencia estructural del gas.

En conjunto, el signo de los coeficientes refleja relaciones coherentes con la lógica de construcción del dataset sintético y con los fundamentos teóricos desarrollados en el marco conceptual del estudio.

### *6.5.2 Variables con mayor peso*

El análisis de magnitud de los coeficientes permite identificar qué variables presentan una mayor influencia dentro del modelo predictivo.

Los resultados muestran que la dependencia del gas constituye, con diferencia, la variable más relevante dentro del modelo de regresión logística. La elevada magnitud de su

coeficiente indica que la estructura energética inicial del hogar representa el principal factor asociado a la probabilidad de transición energética en el escenario analizado.

Este comportamiento sugiere que los hogares previamente dependientes del gas presentan una sensibilidad significativamente mayor ante escenarios de shock energético, incrementando su predisposición a sustituir o modificar su sistema energético.

En segundo lugar, las variables relacionadas con la incertidumbre presentan una influencia negativa relevante sobre la clasificación final del modelo. Esto indica que niveles elevados de incertidumbre pueden actuar como barreras psicológicas o conductuales frente a la toma de decisiones vinculadas a procesos de transición energética.

Por otro lado, las variables asociadas a la exposición a la información muestran efectos positivos moderados, reflejando que el acceso a información puede facilitar la disposición al cambio energético y favorecer comportamientos más adaptativos frente al shock energético.

Finalmente, variables económicas como la renta anual, el coste de transición o el nivel de shock energético presentan una influencia comparativamente menor dentro del modelo. Este resultado puede estar relacionado tanto con la estructura metodológica utilizada para construir el dataset sintético como con la elevada capacidad explicativa de la variable de dependencia del gas.

En consecuencia, los resultados obtenidos sugieren que el comportamiento energético de los hogares no depende exclusivamente de variables económicas, sino también de factores estructurales y conductuales que condicionan la capacidad de adaptación ante escenarios de incertidumbre energética.

### *6.5.3 Interacciones relevantes*

Aunque la regresión logística utilizada en el presente estudio no incorpora explícitamente términos avanzados de interacción entre variables, los resultados permiten identificar relaciones indirectas relevantes entre distintos factores del modelo.

En particular, se observa que el impacto del shock energético depende parcialmente de la combinación entre dependencia del gas, incertidumbre y acceso a información. Los hogares con una elevada dependencia del gas muestran una mayor probabilidad de transición energética, aunque esta predisposición puede verse reducida cuando existen altos niveles de incertidumbre.

Del mismo modo, los resultados sugieren que la exposición a la información puede actuar como un factor moderador frente a la incertidumbre percibida, favoreciendo una mayor disposición al cambio energético en determinados perfiles de hogar.

Asimismo, se identifican diferencias entre perfiles de hogares en función de su estructura energética y características socioeconómicas, lo que indica que la transición energética responde a un fenómeno multifactorial condicionado simultáneamente por variables económicas, estructurales y conductuales.

En conjunto, la interpretación de los coeficientes confirma que el modelo logra capturar patrones coherentes con los fundamentos económicos y conductuales planteados en el estudio, permitiendo identificar variables relevantes asociadas al proceso de transición energética de los hogares dentro del escenario simulado.

## **6.6 Síntesis general de resultados**

Con el objetivo de integrar los principales hallazgos obtenidos a lo largo del análisis, la Tabla 16 resume las relaciones más relevantes identificadas entre las variables analizadas y la probabilidad de transición energética de los hogares.

*Tabla III. Síntesis de resultados principales*

<b>Factor analizado</b>	<b>Resultado observado</b>
Shock energético	Incrementa la presión hacia el cambio energético
Dependencia del gas	Principal factor asociado a la transición
Coste de transición	Actúa como barrera parcial al cambio
Renta del hogar	Mayor renta facilita adaptación
Vulnerabilidad económica	Reduce capacidad de transición
Exposición a información	Favorece el cambio energético
Incertidumbre percibida	Reduce la probabilidad de transición
Modelo predictivo	Capacidad clasificatoria moderada (ROC-AUC = 0,735)

## **CAPÍTULO VII: CONCLUSIONES**

El presente estudio ha analizado el comportamiento del consumidor ante un escenario de shock energético extremo: la eliminación súbita del gas como fuente energética principal en los hogares de la Comunidad de Madrid; y se ha realizado mediante un enfoque basado en simulación y técnicas de Machine Learning.

Los resultados obtenidos muestran que la transición energética constituye un fenómeno complejo y multifactorial, en el que las decisiones de los hogares dependen del incremento del precio energético y también de factores estructurales, económicos y conductuales.

El análisis evidencia que el incremento del precio energético actúa como un incentivo relevante para la sustitución de la fuente energética, especialmente en escenarios de shock elevado. Sin embargo, el efecto del precio no se comporta de manera completamente lineal, lo que indica que la reacción de los hogares depende simultáneamente de otras variables como la renta, el coste de transición, la vulnerabilidad económica o la dependencia inicial del gas.

Asimismo, los resultados muestran que la dependencia del gas constituye uno de los factores con mayor influencia sobre la probabilidad de transición energética. Los hogares altamente dependientes del gas presentan una mayor exposición al shock regulatorio y, en consecuencia, una mayor predisposición al cambio energético.

Desde el punto de vista conductual, el estudio confirma que variables como la incertidumbre percibida y la exposición a la información desempeñan un papel significativo en la toma de decisiones. En particular, mayores niveles de incertidumbre reducen la disposición al cambio, mientras que un mayor acceso a información favorece la capacidad de adaptación energética de los hogares.

En relación con las diferencias por perfil de hogar, se observa que la capacidad económica influye sobre la posibilidad de asumir los costes asociados a la transición energética. No obstante, la renta no actúa como único factor explicativo, ya que existen hogares de renta media o baja que realizan la transición energética y hogares de renta elevada que mantienen

su sistema energético original. Esto refuerza la idea de que las decisiones energéticas responden a la interacción entre múltiples factores económicos y conductuales.

Por otra parte, la aplicación del modelo de regresión logística permitió identificar patrones relevantes dentro del comportamiento energético de los hogares simulados. El modelo alcanzó un rendimiento razonable en términos de clasificación, obteniendo una capacidad predictiva superior al azar y mostrando una elevada sensibilidad para detectar hogares con probabilidad de cambio energético.

En este sentido, los resultados ponen de manifiesto el potencial de las técnicas de Machine Learning como herramienta de apoyo para el análisis de políticas energéticas y escenarios de transición. Aunque el modelo no pretende predecir con exactitud el comportamiento real de los consumidores, sí permite aproximar tendencias generales y analizar el impacto relativo de distintas variables en contextos de incertidumbre.

Finalmente, el estudio confirma que los procesos de transición energética pueden generar efectos desiguales entre hogares, dependiendo de su capacidad económica, estructura energética inicial y acceso a información. En consecuencia, las políticas públicas orientadas a acelerar la transición energética deberían considerar además de los incentivos económicos, los mecanismos de reducción de incertidumbre y apoyo específico a hogares vulnerables.

En conjunto, el trabajo aporta una aproximación metodológica útil para el análisis del comportamiento del consumidor en escenarios energéticos futuros, integrando simulación, economía del comportamiento y técnicas predictivas dentro de un mismo marco analítico.

## **CAPÍTULO 8: LIMITACIONES DEL ESTUDIO**

El estudio presenta una serie de limitaciones derivadas tanto del enfoque metodológico adoptado como de la naturaleza del problema analizado. Estas limitaciones deben ser consideradas a la hora de interpretar los resultados obtenidos y extraer conclusiones sobre el comportamiento de los consumidores ante un shock energético.

### **8.1 Limitaciones metodológicas**

Una de las principales limitaciones del estudio radica en el uso de un dataset sintético. Si bien este enfoque permite analizar escenarios hipotéticos no observables directamente en la realidad, como la eliminación total del gas, los datos generados corresponden a simulaciones construidas a partir de supuestos estadísticos y relaciones teóricas. En consecuencia, los resultados obtenidos deben interpretarse como aproximaciones teóricas y no como predicciones exactas del comportamiento de los consumidores.

Asimismo, la construcción del dataset implica la definición de reglas de generación y relaciones entre variables que, aunque fundamentadas en la literatura económica y en supuestos plausibles, simplifican la complejidad del comportamiento real. Factores como las interacciones sociales, las restricciones tecnológicas, las dinámicas del mercado inmobiliario o la influencia concreta de determinadas políticas públicas no se incorporan de forma explícita en el modelo.

Existe el riesgo de que ciertas relaciones detectadas por el modelo reflejen parcialmente las reglas utilizadas en la construcción del dataset sintético y no necesariamente patrones presentes en poblaciones reales. En este sentido, parte de la capacidad predictiva observada podría estar influida por las propias estructuras definidas durante la simulación de los datos.

Por otro lado, las variables conductuales incorporadas en el modelo, como la incertidumbre o la exposición a la información, se han construido mediante escalas simuladas, lo que introduce un cierto grado de subjetividad en su definición.

En contextos reales, estos factores pueden verse condicionados por elementos más complejos, como la confianza institucional, el contexto mediático, la percepción del riesgo o las experiencias previas de los individuos.

Por último, el modelo desarrollado presenta un carácter estático; analiza un único escenario temporal sin incorporar dinámicas evolutivas del comportamiento del consumidor. En la realidad, las decisiones relacionadas con la transición energética suelen producirse de forma gradual y adaptativa a lo largo del tiempo: en función de la evolución de los precios, las expectativas económicas o los cambios regulatorios. Los resultados permiten identificar asociaciones estadísticas entre variables y patrones de comportamiento dentro del escenario simulado, pero no permiten establecer relaciones causales directas entre los distintos factores analizados. En consecuencia, las relaciones observadas deben interpretarse como tendencias probabilísticas derivadas de la simulación y no como evidencia causal concluyente.

## **8.2 Limitaciones del modelo predictivo**

La regresión logística presenta limitaciones inherentes a su propia naturaleza. Aunque se trata de una herramienta adecuada para problemas de clasificación binaria y ofrece una elevada interpretabilidad, no permite capturar relaciones no lineales complejas ni interacciones avanzadas entre variables, lo que podría afectar a la precisión del análisis en contextos reales más dinámicos.

El modelo no ha sido validado mediante datos reales observados, por lo que su capacidad predictiva en escenarios reales no puede confirmarse empíricamente. En consecuencia, los resultados deben interpretarse principalmente como una simulación analítica orientada a explorar posibles patrones de comportamiento bajo determinadas condiciones hipotéticas.

Además, la distribución de la variable objetivo dentro del dataset sintético puede generar un cierto desbalance entre clases, favoreciendo una mayor capacidad del modelo para detectar hogares con probabilidad de cambio energético frente a hogares que mantienen su sistema

energético original. Esta situación puede influir en métricas como el recall o la precisión del modelo, afectando parcialmente el equilibrio de la capacidad clasificatoria.

### **8.3 Limitaciones del contexto analizado**

El estudio se centra exclusivamente en hogares de la Comunidad de Madrid, lo que limita la generalización de los resultados a otros contextos geográficos. Las características específicas de esta región, como su estructura urbana, nivel de renta o acceso a infraestructuras energéticas, pueden diferir significativamente de otras comunidades autónomas o países.

Además, el escenario analizado (la eliminación total del gas), constituye un supuesto extremo que no refleja necesariamente la evolución real de las políticas energéticas, las cuales suelen desarrollarse mediante procesos graduales de transición.

Si bien este planteamiento permite analizar reacciones ante situaciones límite, también implica que los resultados deben interpretarse dentro de este marco hipotético específico.

### **8.4 Limitaciones sistémicas**

Finalmente, el presente estudio no incorpora de forma explícita la respuesta de otros agentes económicos relevantes, como las empresas energéticas, los proveedores de servicios, los reguladores o las administraciones públicas, cuya actuación puede influir de manera decisiva en el comportamiento de los consumidores.

En un contexto real, la reacción de estos agentes ante un shock energético no es pasiva, sino que incluye la adaptación de estrategias comerciales, modificaciones de precios, inversiones en nuevas infraestructuras, desarrollo de tecnologías alternativas o implementación de políticas públicas orientadas a facilitar la transición energética, como subvenciones, incentivos fiscales o regulaciones específicas.

Asimismo, la interacción entre estos agentes y los consumidores genera dinámicas complejas que pueden amplificar o mitigar los efectos del shock. Por ejemplo, la existencia de ayudas públicas puede reducir el coste de transición energética y aumentar la probabilidad de

cambio, mientras que limitaciones en la infraestructura energética podrían generar cuellos de botella o incrementar la incertidumbre de los hogares.

En este sentido, el análisis desarrollado en el presente trabajo se centra en una visión parcial del sistema, limitada al comportamiento individual de los consumidores, sin incorporar de forma explícita las interacciones estratégicas entre los distintos agentes del mercado.

Esta simplificación, aunque permite aislar el proceso de toma de decisiones de los hogares y facilitar su modelización, implica que los resultados deben interpretarse con cautela.

A pesar de estas limitaciones, el enfoque adoptado permite ofrecer una aproximación válida al análisis del comportamiento del consumidor en contextos de incertidumbre energética, proporcionando una base útil para futuros estudios que incorporen datos reales, modelos dinámicos o técnicas predictivas más complejas.

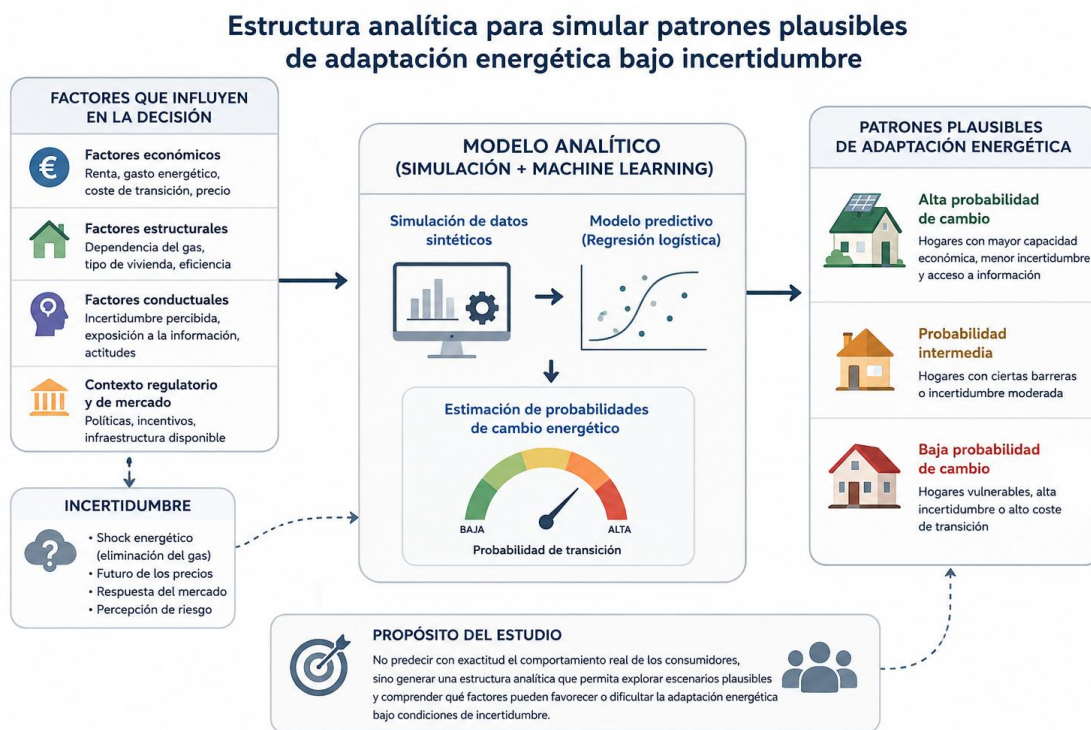
## CAPÍTULO IX: CONTRIBUCIÓN

El presente trabajo aporta una aproximación interdisciplinaria al análisis del comportamiento del consumidor en contextos de transición energética, integrando conceptos de economía, análisis de datos y Machine Learning dentro de un mismo marco metodológico.

La investigación combina herramientas cuantitativas y enfoques conductuales para explorar cómo distintos perfiles de hogares podrían reaccionar ante un escenario extremo de shock energético.

En este sentido, el estudio no pretende ofrecer predicciones exactas del comportamiento real de los consumidores, sino proporcionar una estructura analítica capaz de simular patrones plausibles de adaptación energética bajo condiciones de incertidumbre. Ver Infografía 2

Infografía II. Estructura para simular patrones plausibles



## 9.1 Aporte académico

Desde el punto de vista académico, el trabajo contribuye al estudio de la transición energética desde una perspectiva centrada en el comportamiento del consumidor, incorporando variables económicas, estructurales y conductuales dentro de un mismo modelo de análisis.

La investigación permite conectar distintos marcos teóricos, como la economía del comportamiento, la teoría de shocks económicos y el análisis predictivo, mostrando cómo las decisiones energéticas de los hogares pueden verse condicionadas simultáneamente por factores racionales y psicológicos.

Asimismo, el estudio aporta una aproximación aplicada al uso de técnicas de Machine Learning en el ámbito de la economía energética, contribuyendo a la comprensión de cómo los modelos predictivos pueden utilizarse para analizar patrones de comportamiento en escenarios complejos y altamente inciertos.

## 9.2 Aporte innovador

Uno de los principales elementos innovadores del trabajo radica en la combinación entre simulación de datos sintéticos y técnicas de aprendizaje automático para analizar escenarios energéticos hipotéticos que difícilmente podrían estudiarse mediante datos reales observados.

El desarrollo de un dataset sintético estructurado permitió construir un entorno controlado capaz de representar relaciones plausibles<sup>22</sup> entre variables económicas, energéticas y conductuales, facilitando el análisis de comportamientos potenciales ante un escenario extremo de transición energética.

Asimismo, la incorporación de variables conductuales simuladas, como la incertidumbre percibida o la exposición a la información, introduce una dimensión adicional al análisis

---

<sup>22</sup> Relaciones que tienen sentido lógico, económico y teórico, aunque no hayan sido observadas directamente en datos reales.

tradicional de la transición energética, normalmente centrado únicamente en variables económicas o tecnológicas.

Desde una perspectiva metodológica, el trabajo también resulta innovador al integrar simulación estadística, análisis económico y modelos predictivos dentro de un mismo proceso analítico, generando una aproximación híbrida entre investigación económica y ciencia de datos aplicada.

### **9.3 Aporte prospectivo**

El estudio aporta una visión prospectiva sobre posibles comportamientos futuros de los consumidores ante escenarios de transformación energética acelerada. Aunque el escenario analizado constituye una situación hipotética extrema, los resultados permiten identificar tendencias potenciales de adaptación energética y factores que podrían influir en futuras decisiones de los hogares.

En este sentido, el trabajo pone de manifiesto que la transición energética puede generar impactos desiguales entre distintos perfiles de consumidores, dependiendo de variables como la dependencia del gas, la vulnerabilidad económica, el acceso a información o la capacidad de asumir costes de transición.

Asimismo, los resultados sugieren que las políticas energéticas futuras podrían beneficiarse del uso de herramientas predictivas y modelos de simulación para anticipar posibles comportamientos de los consumidores y diseñar estrategias de transición más adaptativas y eficientes.

Del mismo modo, el estudio abre la posibilidad de desarrollar futuras investigaciones basadas en modelos dinámicos, simulaciones temporales o integración de datos reales procedentes de mercados energéticos, hogares o políticas públicas.

#### **9.4 Aporte metodológico**

Desde el punto de vista metodológico, el trabajo demuestra la utilidad de combinar datasets sintéticos con técnicas de Machine Learning como alternativa para explorar fenómenos complejos caracterizados por incertidumbre, ausencia de datos reales o imposibilidad de observación directa.

La construcción del dataset permitió establecer relaciones estructurales coherentes entre variables, incorporando tanto elementos económicos como factores conductuales dentro de una misma arquitectura de simulación.

Asimismo, la aplicación de un modelo de regresión logística facilitó la identificación de patrones relevantes asociados a la probabilidad de transición energética, ofreciendo una herramienta interpretable y adecuada para problemas de clasificación binaria.

El trabajo también contribuye metodológicamente al mostrar cómo herramientas de ciencia de datos pueden integrarse en investigaciones económicas y sociales sin perder interpretabilidad ni coherencia teórica.

En conjunto, la metodología desarrollada constituye una base útil para futuras investigaciones orientadas al análisis de comportamiento energético, evaluación de políticas públicas o simulación de escenarios económicos complejos mediante técnicas de análisis predictivo.

## BIBLIOGRAFÍA

### Literatura

- BBVA Research. (2026). *España: más resilientes al gas, pero no más competitivos*.
- Blanchard, O. (2017). *Macroeconomics* (7th ed.). Pearson.
- BP. (2024). *Statistical review of world energy*.
- CaixaBank Research. (2026). *España ante el nuevo shock energético*.
- Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.
- Fressoz, J.-B. (2025). *Sin transición. Una nueva historia de la energía*. Arpa Editores.
- Friedman, D., & Sunder, S. (1994). *Experimental methods: A primer for economists*. Cambridge University Press.
- Fundación Cajamar. (2025). *La estrategia para la transición energética*. Cajamar Caja Rural.
- Hamilton, J. D. (2009). *Causes and consequences of the oil shock of 2007–08*. Brookings Papers on Economic Activity.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.
- International Energy Agency. (2025). *World energy outlook*.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning* (2nd ed.). Springer.
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Smith, J., & Taylor, R. (2025). *Energy transition in the oil and gas industry*. Routledge.

Sovacool, B. K. (2016). *The political economy of energy*. Palgrave Macmillan.

Thaler, R. H. (2015). *Misbehaving: The making of behavioral economics*. W. W. Norton & Company.

Tversky, A., & Kahneman, D. (1979). *Prospect theory: An analysis of decision under risk*. *Econometrica*, 47(2), 263–291.

Varian, H. R. (2019). *Intermediate microeconomics: A modern approach* (9th ed.). Norton.

## Artículos

- BBVA Research. (2026, abril 13). *España: más resilientes al gas, pero no más competitivos*.  
<https://www.bbvaresearch.com/publicaciones/espana-mas-resilientes-al-gas-pero-no-mas-competitivos/>
- CaixaBank Research. (2026, abril). *España ante el nuevo shock energético: comparativa con Europa*.  
<https://www.caixabankresearch.com/es/economia-y-mercados/actividad-y-crecimiento/espana-ante-nuevo-shock-energetico-comparativa-europa>
- European Commission. (2023). *Energy prices and costs in Europe*. European Commission.  
[https://energy.ec.europa.eu/data-and-analysis/energy-prices-and-costs-europe\\_en](https://energy.ec.europa.eu/data-and-analysis/energy-prices-and-costs-europe_en)
- Geldard, R. (2026, March 24). *Energy shock shakes markets*. World Economic Forum.  
<https://www.weforum.org/stories/2026/03/energy-shock-shakes-markets/>
- International Monetary Fund. (2026, March 30). *How the war in the Middle East is affecting energy, trade, and finance*.  
<https://www.imf.org/en/blogs/articles/2026/03/30/how-the-war-in-the-middle-east-is-affecting-energy-trade-and-finance>
- Industria Química. (2026, abril 25). *El sector gasista destaca su papel clave en la seguridad energética y la competitividad industrial en España*.  
<https://www.industriaquimica.es/noticias/20260425/-sector-gasista-destaca-su-papel-clave-en-seguridad-energetica-competitividad-industrial-en-espana>

- Kirsch, J. (2026, April 12). *Energy market volatility reflects a deeper geopolitical repricing*. Forbes.  
<https://www.forbes.com/sites/investor-hub/2026/04/12/energy-market-volatility-reflects-a-deeper-geopolitical-repricing/>
- New Lines Institute. (2026, March 31). *2026 second quarter forecast report*.  
<https://newlinesinstitute.org/forecast/2026-second-quarter-forecast/>
- Organisation for Economic Co-operation and Development. (2026, March). *Global economic outlook remains robust but has weakened amid energy shock and geopolitical risks*.  
<https://www.oecd.org/en/about/news/press-releases/2026/03/global-economic-outlook-remains-robust-but-has-weakened-amid-energy-shock-and-geopolitical-risks.html>
- Torres, F. (2026, abril 6). *Energía en España: oportunidades en renovables, petróleo y gas para traders en 2026*. IG.  
<https://www.ig.com/es/ideas-de-trading-y-noticias/energia-en-espana-opportunidades-en-renovables-petroleo-y-gas-p-260406>

## Webs

- Eurostat. (2026). *Eurostat database*.  
<https://ec.europa.eu/eurostat>
- Instituto Nacional de Estadística. (2026). *Instituto Nacional de Estadística*.  
<https://www.ine.es>
- Instituto para la Diversificación y Ahorro de la Energía. (2026). *IDAE*.  
<https://www.idae.es>
- Red Eléctrica de España. (2026). *Red Eléctrica de España*.  
<https://www.ree.es>
- Operador del Mercado Ibérico de Energía. (2026). *OMIE*.  
<https://www.omie.es>

## Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

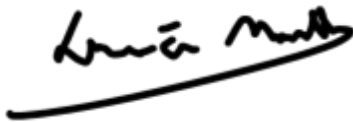
ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido 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, Lucía Catalina Murillo Gilmartín, estudiante del Doble Grado en Derecho y Business Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "Un shock energético en el comportamiento del consumidor: Análisis mediante simulación y Machine Learning aplicado a la Comunidad de Madrid", 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. Crítico: Para encontrar contra-argumentos a una tesis específica que pretendo defender.
3. Referencias: Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
4. Metodólogo: Para descubrir métodos aplicables a problemas específicos de investigación.
5. Interpretador de código: Para realizar análisis de datos preliminares.
6. Estudios multidisciplinares: Para comprender perspectivas de otras comunidades sobre temas de naturaleza multidisciplinar.
7. Constructor de plantillas: Para diseñar formatos específicos para secciones del trabajo.
8. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto.
9. Generador previo de diagramas de flujo y contenido: Para esbozar diagramas iniciales.
10. Sintetizador y divulgador de libros complicados: Para resumir y comprender literatura compleja.
11. Generador de datos sintéticos de prueba: Para la creación de conjuntos de datos ficticios.
12. Generador de problemas de ejemplo: Para ilustrar conceptos y técnicas.
13. Revisor: Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
14. Generador de encuestas: Para diseñar cuestionarios preliminares.
15. Traductor: Para traducir textos de un lenguaje a otro.

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: 17/06/2026

A handwritten signature in black ink, appearing to read "Lucía Muñoz", written over a horizontal line.A second handwritten signature in black ink, identical to the one above, also written over a horizontal line.

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

## ANEXOS

### Anexo 1. Definición de variables en Python

```
# Variables base

id_hogar = np.arange(1, n_hogares + 1)

renta = np.maximum(np.random.normal(28000, 12000, n_hogares),
8000).round(0).astype(int)

gasto_energia = np.maximum(np.random.normal(100, 30, n_hogares),
30).round(0).astype(int)

# Variables estructurales

tipo_vivienda_idx = np.random.choice([0, 1], n_hogares, p=[0.75, 0.25])

tipo_vivienda = np.where(tipo_vivienda_idx == 1, 'Casa', 'Piso')

sistema_idx = np.random.choice([0, 1], n_hogares, p=[0.736, 0.264])

sistema calefaccion = np.where(sistema_idx == 0, 'Gas', 'No Gas')

dependencia_gas = (sistema_idx == 0).astype(int)

# Variable del escenario energético

precio_shock = np.random.choice([20, 40, 60], n_hogares)

# Vulnerabilidad económica (variable derivada a partir de gasto/renta)

vulnerabilidad_score = (gasto_energia / renta) * 100 * 2
```

```

vulnerabilidad = pd.cut(
    vulnerabilidad_score,
    bins=[0, 25, 50, np.inf],
    labels=['Baja', 'Media', 'Alta']
)

# Coste de transición (ajustado por tipo de vivienda y dependencia del gas)
ajuste_casa = np.where(tipo_vivienda_idx == 1, 1.2, 1.0)
ajuste_gas = np.where(sistema_idx == 0, 1.3, 1.0)
coste_transicion = np.clip(
    np.random.normal(4000 * ajuste_casa * ajuste_gas, 1500, n_hogares),
    2000,
    8000
).round(0).astype(int)

# Variables conductuales simuladas
incertidumbre_num = np.clip(np.random.normal(5, 2, n_hogares), 1, 10)
incertidumbre = pd.cut(
    incertidumbre_num,
    bins=[0, 3.33, 6.66, 10, np.inf],

```

```
labels=['Baja', 'Media', 'Alta', 'Muy Alta']  
)  
  
informacion_num = np.clip(11 - incertidumbre_num + np.random.normal(0, 1.5, n_hogares),  
1, 10)
```