



GRADO EN BUSINESS ANALYTICS

TRABAJO FIN DE GRADO

Tokenización de Inmuebles: Análisis de Datos y
Perspectivas Financieras

Autor: Alberto Martín Colino

Director: Gabriel Rodríguez Garnica

Madrid

Fecha:

Agradecimientos

A mí, como autor de este trabajo, me parecía conveniente que, puestos a hablar sobre *Real Estate* y mercado inmobiliario, empezásemos por lo más básico: la construcción. Y es que un proyecto así está hecho de muchísimas piedras, aunque a veces parezca que no son más que papeles y gráficos.

Este Proyecto de Fin de Grado, al igual que el entregado el curso anterior, representa las últimas piedras de un edificio que he venido construyendo durante los últimos cinco años, desde que comencé la carrera. Los humanos solemos celebrar la primera piedra, sacarnos fotos con ella y reservarle un lugar especial en la memoria. Sin embargo, hoy vengo a hacer lo contrario. Empezar es de todos, y son muchos los proyectos que se quedan a medias. Hoy celebro la última piedra, esa que, con esfuerzo y con muchísima ayuda, he sido capaz de colocar.

En primer lugar, gracias a Gabriel, mi tutor, por su implicación y por la asignatura que me impartió, que despertó en mí el interés por este tema.

En segundo lugar, a mis padres y a mi hermano, por su apoyo constante, tanto en lo económico como en lo emocional. No han dudado de mí ni en los momentos más difíciles, y eso lo cambia todo.

Por último, pero no menos importante, a todos y cada uno de mis compañeros de clase. Su alegría, su compromiso y su esfuerzo no solo me han empujado a querer estar a su altura, sino que han sido el mejor motivo para madrugar cada mañana y, día tras día, seguir adelante con esta aventura.

TOKENIZACIÓN DE INMUEBLES: ANÁLISIS DE DATOS Y PERSPECTIVAS FINANCIERAS

Autor: Martín Colino, Alberto.

Director: Rodríguez Garnica, Gabriel.

Entidad Colaboradora: ICAI – Universidad Pontificia Comillas

RESUMEN DEL PROYECTO

Este Trabajo Fin de Grado analiza comparativamente el mercado inmobiliario tradicional y el mercado inmobiliario tokenizado con el objetivo de estudiar sus diferencias estructurales en la formación de valor y rentabilidad. A partir de datasets representativos, se aplica un pipeline homogéneo de limpieza, análisis exploratorio, regresión y segmentación, prestando especial atención a la homogeneización de variables clave, como la transformación del rendimiento tokenizado en métricas comparables con el mercado tradicional.

En el mercado tradicional se examina la relación entre precio y fundamentos urbanos, mientras que en el mercado tokenizado se analiza el rendimiento anual en función de variables espaciales y financieras. El estudio se complementa con un análisis del marco regulatorio aplicable a la tokenización inmobiliaria, abordando su encaje dentro de la normativa financiera y los retos que plantea para la supervisión y protección del inversor

Palabras clave: *Tokenización inmobiliaria, Formación de valor, Rendimiento financiero, Mercado inmobiliario tradicional, Regulación financiera.*

REAL ESTATE TOKENIZATION: DATA ANALYSIS AND FINANCIAL PERSPECTIVES

Author: Martín Colino, Alberto.

Supervisor: Rodríguez Garnica, Gabriel.

Collaborating Entity: ICAI – Universidad Pontificia Comillas

ABSTRACT

This Thesis comparatively analyzes the traditional real estate market and the tokenized real estate market to study their structural differences in value creation and profitability. Using representative datasets, a homogeneous pipeline of data cleaning, exploratory analysis, regression, and segmentation is applied, paying particular attention to the homogenization of key variables, such as transforming tokenized returns into metrics comparable with the traditional market.

In the traditional market, the relationship between price and urban fundamentals is examined, while in the tokenized market, annual returns are analyzed based on spatial and financial variables. The study is complemented by an analysis of the regulatory framework applicable to real estate tokenization, addressing its integration within financial regulations and the challenges it poses for investor supervision and protection.

Keywords: Real estate tokenization, Value creation, Financial performance, Traditional real estate market, Financial regulation

Índice de la memoria

<i>Índice de la memoria</i>	<i>I</i>
<i>Índice de figuras</i>	<i>IV</i>
Capítulo 1. Introducción	6
Capítulo 2. Descripción de tecnologías	8
2.1 2.1. Procesamiento y estructuración de datos.....	8
2.2 2.2 Limpieza y transformación de datos.....	8
2.3 2.3 Modelización estadística y aprendizaje automático	9
2.4 Visualización de resultados.....	10
2.4 2.5 Organización técnica del pipeline	10
Capítulo 3. Estado de la Cuestión	11
Capítulo 4. Definición del trabajo	17
4.1 Justificación.....	17
4.1.1 Brecha tecnológica y de mercado	17
4.1.2 Relevancia científica y social	18
4.1.3 Objetivos del proyecto	19
4.2 4.2 Metodología.....	19
4.2.1 Análisis exploratorio de datos	19
4.2.2 Análisis de relaciones y capacidad explicativa	20
4.2.3 Análisis de segmentación (clustering).....	20
4.2.4 Síntesis e interpretación de resultados.....	21
Capítulo 5. Análisis exploratorio de datos	22
5.1 Datasets	22
5.1.1 Mercado Tradicional.....	23
5.1.2 Mercado Tokenizado	23
5.1.3 Consistencia y comparabilidad entre bases de datos.....	24

5.2	Estadística descriptiva	24
5.3	Variables principales analizadas.....	30
5.4	Distribuciones e interpretación preliminar	30
5.5	Análisis de relaciones entre variables.....	34
5.5.1	<i>Relaciones en el mercado inmobiliario tradicional</i>	34
5.5.2	<i>Relaciones en el mercado inmobiliario tokenizado</i>	37
5.5.3	<i>Interpretación conjunta</i>	41
5.6	Análisis por deciles	42
Capítulo 6. Comparación metodológica entre mercados		44
6.1	Comparación de estructura y dispersión.....	44
6.2	Análisis de correlaciones.....	45
6.3	Análisis explicativo	50
6.4	Análisis por segmentación.....	52
Capítulo 7. Discusión e implicaciones estructurales.....		59
Capítulo 8. Futuras líneas de investigación		62
Bibliografía 64		
	Bibliotecas de Python utilizadas	65
ANEXO I Declaración utilización IA Generativa.....		67
ANEXO II Implementación del pipeline analítico en Python		68
1.	Descripción general del proyecto.....	68
2.	Carga y preprocesado de los datos.....	68
3.	Construcción de variables derivadas.....	70
4.	Análisis exploratorio de datos (EDA).....	71
5.	Modelización estadística y regresión	73
6.	Segmentación mediante clustering (KMeans)	75
7.	Nota sobre reproducibilidad.....	77

Índice de figuras

Figura 1 Distribución del precio por metro cuadrado en el mercado tradicional	25
Figura 2 Distribución del precio por metro cuadrado en el dataset urbano (MRT)	26
Figura 3 Distribución del precio por metro cuadrado en el dataset tokenizado	28
Figura 4 Distribución del rendimiento en el mercado tokenizado.....	28
Figura 5 Boxplot distribución del precio por metro cuadrado en el Mercado tradicional...	31
Figura 6 Distribución del rendimiento en el mercado tokenizado.....	33
Figura 7 Relación entre tamaño (m ²) y precio – Mercado tradicional	35
Figura 8 Relación entre nº de habitaciones y precio Mercado tradicional	36
Figura 9 Relación entre precio por m ² y distancia al MRT	37
Figura 10 Relación entre inversión total y rendimiento-Mercado tokenizado	38
Figura 11 Relación entre liquidez (número de transferencias) y rendimiento – Mercado tokenizado.....	39
Figura 12 Relación entre liquidez (número de transferencias) y rendimiento – Mercado tokenizado.....	39
Figura 13 Relación entre distancia al centro y rendimiento – Mercado tokenizado	40
Figura 14 Relación entre accesibilidad (Walk Score) y rendimiento – Mercado tokenizado	41
Figura 15 Comparación de dispersión entre mercado tradicional y tokenizado.....	45
Figura 16 Matriz de correlaciones – Mercado tradicional.....	47
Figura 17 Matriz de correlaciones – Mercado tokenizado	49
Figura 18 Capacidad explicativa de los modelos	52
Figura 19 N°de clusters mercado tradicional.....	54
Figura 20 N°de clusters mercado tokenizado	55
Figura 21 Clusters mercado tradicional.....	56
Figura 22 Clusters mercado tokenizado	57

Índice de tablas

Tabla 1 Estadística descriptiva Mercado Tradicional.....	25
Tabla 2 Estadística descriptiva Mercado Tokenizado	29
Tabla 3 Estadística descriptiva del decil superior (10%) – Mercado inmobiliario tradicional	42
Tabla 4 Estadística descriptiva del decil superior (10%) – Mercado inmobiliario Tokenizado	43

Capítulo 1. INTRODUCCIÓN

El objetivo principal de este Trabajo Fin de Grado es analizar la tokenización de activos inmobiliarios como una alternativa emergente al modelo tradicional de inversión en bienes raíces, evaluando en qué medida esta tecnología puede transformar variables clave como la rentabilidad, la liquidez y el perfil de riesgo de este tipo de activos. Frente a la estructura tradicional del mercado inmobiliario —históricamente caracterizada por su iliquidez, sus elevados costes de transacción y sus altas barreras de entrada—, la tokenización parece prometer una solución a estos inconvenientes: mayor accesibilidad, fraccionamiento de la propiedad y una potencial mejora en la eficiencia del mercado. Este trabajo plantea un marco de análisis cuantitativo que combina técnicas estadísticas y herramientas de *Machine Learning* con el fin de estudiar empíricamente el comportamiento de los inmuebles tokenizados y compararlo con el de activos inmobiliarios tradicionales.

La finalidad de esta investigación es descomponer la complejidad inherente a la tokenización inmobiliaria y ofrecer una visión estructurada que permita comprender no solo sus ventajas potenciales, sino también sus limitaciones reales desde una perspectiva financiera, tecnológica y regulatoria. En un contexto marcado por la digitalización de los mercados financieros y la aparición de instrumentos basados en blockchain, resulta fácil quedarse en el discurso teórico sobre eficiencia y democratización. Sin embargo, este trabajo trata de bajar ese discurso al terreno de los datos: ¿realmente se comportan los activos tokenizados como promete la narrativa tecnológica? ¿O responden a dinámicas distintas a las del mercado inmobiliario tradicional? ¿Existe un cambio real?

El análisis combina un estudio exploratorio riguroso de distintos conjuntos de datos —tanto de activos inmobiliarios tokenizados como del mercado tradicional— con técnicas estadísticas y de aprendizaje automático orientadas a identificar patrones, segmentar activos y evaluar relaciones entre variables relevantes. A partir de un preprocesado cuidadoso de los datos, se aplican modelos de regresión, contrastes estadísticos y métodos de clustering con el objetivo de transformar información heterogénea en conclusiones interpretables. Este enfoque permite analizar si variables clásicas como la localización siguen desempeñando un papel determinante en el rendimiento de los activos tokenizados o si, por el contrario, la lógica de formación de valor se ve modificada por la arquitectura financiera del *token*.

Se ha puesto especial énfasis en el uso de técnicas de aprendizaje no supervisado como herramienta exploratoria, más que predictiva, dado el carácter todavía incipiente y novedoso del mercado tokenizado y la limitada disponibilidad de datos históricos. El objetivo no es “predecir el futuro”, sino comprender la estructura actual del mercado y sus patrones internos. En este sentido, la comparación sistemática entre mercado tradicional y *tokenizado*

constituye el eje central del trabajo, permitiendo evaluar hasta qué punto ambos comparten una lógica común o presentan diferencias estructurales significativas.

Todo el proceso analítico se ha diseñado con especial atención a la trazabilidad y reproducibilidad, documentando de manera explícita las decisiones metodológicas adoptadas y las limitaciones asociadas a la calidad y cantidad de los datos disponibles. La escasez de información pública y estandarizada sobre activos tokenizados no solo condiciona el análisis, sino que también refleja el estado todavía temprano de desarrollo del sector, algo que resulta especialmente relevante cuando se habla de transparencia y liquidez como promesas centrales de esta tecnología.

El objetivo último del trabajo no es evaluar una plataforma concreta ni emitir juicios categóricos sobre la viabilidad futura de la tokenización, sino construir un marco de análisis replicable que permita estudiar este fenómeno desde una perspectiva crítica y fundamentada. Este enfoque busca ser útil tanto para inversores interesados en nuevas formas de exposición inmobiliaria, como para reguladores que deben encajar estos instrumentos dentro del marco normativo existente, y para profesionales del sector financiero que tratan de entender si estamos ante una evolución incremental o ante una transformación más profunda del mercado.

En definitiva, y antes de embarcar mar adentro, se podría decir que el presente trabajo combina una aproximación técnica basada en el análisis de datos con una reflexión económica y regulatoria sobre un fenómeno en plena evolución. No se trata únicamente de aplicar modelos estadísticos, sino de entender qué está cambiando —y qué no— en la forma en que se estructura y percibe el valor inmobiliario cuando se introduce la tokenización. Esa mirada comparativa y crítica constituye la principal aportación del estudio y el punto de partida para futuras investigaciones en un ámbito que, con toda probabilidad, seguirá evolucionando en los próximos años.

Capítulo 2. DESCRIPCIÓN DE TECNOLOGÍAS

El desarrollo del proyecto se ha llevado a cabo en *Python* (versión 3.8), utilizando un entorno de trabajo gestionado mediante *Anaconda*, lo que ha permitido controlar de forma precisa las dependencias y versiones de las bibliotecas empleadas. La elección de *Python* responde a su consolidación como estándar en ciencia de datos y análisis cuantitativo, así como a la integración fluida entre bibliotecas especializadas en procesamiento de datos, modelización estadística y aprendizaje automático.

Las herramientas utilizadas pueden agruparse en cuatro grandes bloques funcionales: procesamiento y estructuración de datos, limpieza y transformación, modelización y aprendizaje automático, y visualización de resultados.

2.1. PROCESAMIENTO Y ESTRUCTURACIÓN DE DATOS

La manipulación de los datasets se ha realizado principalmente mediante la biblioteca *Pandas*, que ha constituido el eje central del tratamiento de datos. Esta herramienta ha permitido cargar archivos en formato *CSV* y *Excel*, organizar la información en estructuras tipo *DataFrame* y aplicar filtros condicionales necesarios para el análisis comparativo. Asimismo, *Pandas* ha facilitado la agrupación de observaciones, la segmentación por deciles y el cálculo de estadísticos descriptivos como medias, medianas, desviaciones estándar y percentiles.

Como complemento, se ha utilizado *NumPy* para operaciones numéricas más específicas, especialmente aquellas relacionadas con cálculos vectorizados, conversión y control de tipos de datos, generación de máscaras lógicas y preparación de matrices para su uso en modelos de regresión y algoritmos de *clustering*. La combinación de *Pandas* y *NumPy* ha permitido construir un entorno eficiente para el preprocesado y estructuración inicial de la información.

2.2 LIMPIEZA Y TRANSFORMACIÓN DE DATOS

La preparación de los datos se ha integrado dentro del pipeline analítico con el objetivo de garantizar la coherencia y comparabilidad entre los distintos conjuntos utilizados.

En primer lugar, se ha realizado una depuración básica, eliminando duplicados y filtrando observaciones incompletas en variables clave como el precio, asegurando la consistencia de la muestra.

Posteriormente, se han construido variables derivadas orientadas a capturar dimensiones económicas relevantes. En el mercado tradicional, destacan métricas como el precio por metro cuadrado o por habitación, mientras que en el dataset urbano se han incorporado indicadores de accesibilidad y entorno, como la proximidad al transporte o la densidad comercial.

Asimismo, se ha aplicado una segmentación por percentiles para identificar subconjuntos relevantes del mercado, como el decil superior de precios, facilitando el análisis comparativo entre segmentos

Finalmente, las variables han sido normalizadas en las fases necesarias, especialmente en los modelos de clustering, para evitar efectos derivados de diferencias de escala.

Todas estas transformaciones se han implementado de forma sistemática, garantizando la reproducibilidad del análisis.

2.3 MODELIZACIÓN ESTADÍSTICA Y APRENDIZAJE AUTOMÁTICO

La fase de modelización se ha desarrollado principalmente con la biblioteca *Scikit-learn* (*sklearn*), utilizada tanto para análisis supervisado como no supervisado, complementada con herramientas estadísticas de *SciPy* para contraste formal de hipótesis.

En el ámbito supervisado, se ha empleado el modelo de regresión lineal (*LinearRegression*) para estimar la relación entre las variables dependientes —precio por metro cuadrado en el mercado tradicional y *yield_real* en el tokenizado— y sus correspondientes variables explicativas. La evaluación del modelo se ha realizado mediante el coeficiente de determinación (R^2), validación cruzada con *cross_val_score* y procedimientos de bootstrap para analizar la estabilidad del poder explicativo. Asimismo, se incorporó un modelo no lineal mediante *RandomForestRegressor* con el objetivo de contrastar la posible existencia de relaciones no lineales y evaluar la capacidad de generalización fuera de muestra.

En el ámbito no supervisado, se ha aplicado el algoritmo *KMeans* para la segmentación de los datos, precedido por estandarización mediante *StandardScaler*. La calidad de los clusters se ha evaluado a través del *silhouette score*, y su relevancia estadística se ha contrastado mediante análisis de varianza (ANOVA). Adicionalmente, se utilizaron pruebas estadísticas como el test de Levene para comparar varianzas entre mercados y contrastes t para analizar diferencias entre deciles extremos.

En conjunto, estas herramientas configuran un marco técnico coherente y homogéneo que permite comparar de forma rigurosa la estructura interna del mercado tradicional y del mercado tokenizado, tanto desde una perspectiva predictiva como desde el análisis de segmentación y contraste estadístico.

2.4 VISUALIZACIÓN DE RESULTADOS

La visualización de resultados se ha realizado mediante las bibliotecas *Matplotlib* y *Seaborn*, que permiten generar representaciones gráficas de carácter estadístico. Entre los gráficos empleados se incluyen histogramas con estimación de densidad, boxplots por deciles, gráficos de dispersión para analizar relaciones bivariantes y mapas de calor de correlaciones.

Estas herramientas no solo facilitan la comunicación de los resultados, sino que también permiten inspeccionar visualmente la estructura de los datos y validar los patrones identificados mediante los modelos estadísticos.

2.5 ORGANIZACIÓN TÉCNICA DEL PIPELINE

El análisis se ha estructurado mediante un *pipeline* modular, compuesto por funciones diferenciadas para la carga de datos, el preprocesado, la construcción de variables, el análisis descriptivo, la segmentación por deciles, el estudio de correlaciones, la regresión y el clustering. Esta organización técnica permite ejecutar el flujo completo de análisis de forma automatizada y reproducible.

El presente apartado constituye una introducción a la arquitectura tecnológica empleada. El funcionamiento detallado del pipeline, así como la justificación metodológica de cada fase, se desarrollará con mayor profundidad en los capítulos 5 y 6, donde se describen el análisis exploratorio de datos y la modelización comparativa aplicada a ambos mercados.

Capítulo 3. ESTADO DE LA CUESTIÓN

El mercado inmobiliario ha sido históricamente considerado uno de los activos más estables y confiables dentro del sector financiero, gracias a su capacidad de generar ingresos recurrentes y servir como reserva de valor a largo plazo. Invertir en bienes raíces permite a los agentes económicos capturar rentas por arrendamientos y beneficiarse de la apreciación del capital, lo que explica su atractivo tanto para inversores individuales como institucionales. Sin embargo, la propia naturaleza de los inmuebles introduce ciertas características estructurales que condicionan el comportamiento del mercado y lo diferencian de otros activos: las propiedades inmobiliarias son físicamente ilíquidas, requieren un alto desembolso inicial y dependen de procesos de intermediación complejos y costosos, como la gestión de escrituras, tasaciones y la tramitación ante autoridades regulatorias y notariales (Geltner et al., 2014; Case & Shiller, 1989).

La liquidez limitada del mercado tradicional implica que la venta o transferencia de propiedades puede tardar semanas o incluso meses, lo que limita la capacidad de los inversores de reaccionar ante cambios inmediatos en las condiciones económicas o de diversificar rápidamente su portafolio. Además, los altos costes de transacción, incluidos impuestos, honorarios profesionales y comisiones, constituyen barreras de entrada significativas, restringiendo la participación de pequeños inversores y concentrando la propiedad en manos de agentes con mayor capacidad financiera (Fisher et al., 2007). Esta concentración genera, a su vez, una menor accesibilidad a oportunidades de inversión diversificadas, perpetuando la segmentación del mercado y reduciendo la flexibilidad de los agentes minoristas.

En paralelo, la evolución del sector inmobiliario ha estado marcada por la creciente digitalización y la aparición de nuevos instrumentos financieros. Durante las últimas dos décadas, el mercado ha visto la consolidación de los REITs (*Real Estate Investment Trusts*) y plataformas de crowdfunding inmobiliario, que permiten a los inversores acceder a activos fraccionados y a mercados más amplios sin necesidad de adquirir propiedades completas (Pagliari, 2017). Estas innovaciones han transformado parcialmente la naturaleza de la inversión inmobiliaria, ofreciendo mayor accesibilidad y liquidez relativa, aunque todavía condicionadas por procesos centralizados y limitaciones regulatorias. La literatura indica que, si bien estas plataformas reducen algunas fricciones, no eliminan por completo los problemas estructurales del sector, como la dependencia de intermediarios y la complejidad administrativa. La gestión de estos activos parece abrirse a una democratización, aunque todavía quedan muchos flecos por resolver. (Clayton & MacKinnon, 2003).

Asimismo, el perfil del inversor ha comenzado a transformarse. Los pequeños inversores muestran creciente interés en activos inmobiliarios como medio de diversificación, mientras que las instituciones financieras buscan instrumentos que combinen estabilidad de rentabilidad con oportunidades de liquidez. Esta evolución crea un contexto en el que las nuevas tecnologías financieras pueden ofrecer soluciones que amplíen la accesibilidad y la eficiencia del mercado (Eichholtz et al., 2010). En este sentido, la digitalización y el desarrollo de plataformas online representan un primer paso hacia un ecosistema inmobiliario más ágil y transparente, aunque su alcance sigue limitado por la falta de interoperabilidad, la estandarización de datos y la integración de procesos de cumplimiento regulatorio.

Un aspecto central del análisis del mercado tradicional es su papel como *benchmark* frente a alternativas emergentes. La valoración de propiedades sigue basada en métodos bien establecidos, como comparables de mercado, flujo de caja descontado o métodos de coste, y su desempeño puede evaluarse mediante índices de precios inmobiliarios o análisis de rentabilidad histórica (Ling & Archer, 2010). Estos métodos han demostrado ser robustos para estimaciones de largo plazo, pero presentan limitaciones cuando se busca analizar la dinámica de liquidez, acceso fraccionado o diversificación instantánea, aspectos que las tecnologías emergentes pretenden resolver.

En resumen, el mercado inmobiliario tradicional proporciona una base sólida de comprensión para evaluar los nuevos modelos de inversión, pero revela un conjunto de limitaciones que justifican la búsqueda de alternativas innovadoras. La iliquidez estructural, las barreras de entrada, los costes de transacción y la dependencia de intermediarios son problemas reconocidos en la literatura y constituyen un punto de partida indispensable para evaluar la relevancia de la tokenización y otras herramientas financieras emergentes. A partir de este diagnóstico, se puede establecer un marco comparativo que permita analizar de manera crítica el potencial y las limitaciones de los activos inmobiliarios tokenizados en términos de rentabilidad, liquidez y riesgo, conectando directamente con la propuesta y objetivos de este trabajo.

La tokenización de activos inmobiliarios representa una innovación disruptiva en la manera en que se estructura, distribuye y negocia la propiedad de bienes raíces. Este concepto se basa en la transformación de activos físicos, tradicionalmente ilíquidos y fragmentados, en unidades digitales negociables a través de tecnologías de registro distribuido, principalmente blockchain. A través de este proceso, cada propiedad puede dividirse en múltiples tokens que representan derechos económicos proporcionales, permitiendo la inversión fraccionada

y la creación de un mercado secundario potencialmente más líquido que el sistema tradicional (Catalini & Gans, 2016; Cong & He, 2019).

La tokenización no solo se limita a la representación digital de los activos, sino que integra un marco contractual automatizado mediante smart contracts, que ejecutan de forma automática las transacciones y la distribución de ingresos entre los poseedores de tokens. Este mecanismo reduce la necesidad de intermediarios, aumenta la transparencia y asegura trazabilidad en las operaciones, dos aspectos críticos para generar confianza en mercados emergentes y globales (Yermack, 2017). A diferencia de las plataformas tradicionales de crowdfunding o los REITs, los tokens permiten transacciones más rápidas y con una estandarización digital de los derechos sobre el activo, lo que facilita su negociación y seguimiento.

En términos de modelos de negocio, las plataformas de tokenización inmobiliaria suelen estructurar los activos mediante vehículos de propósito especial (SPV¹), que emiten tokens respaldados por propiedades concretas. Los derechos económicos de los inversores suelen estar vinculados a los flujos de caja generados por el inmueble, tales como rentas de arrendamiento o plusvalías por revalorización. Existen distintos enfoques según la plataforma: algunos tokens representan equity directo en la propiedad, mientras que otros funcionan bajo un esquema de revenue-sharing, ² donde los pagos dependen de los ingresos del activo subyacente. Esta flexibilidad permite adaptar la inversión a distintos perfiles de riesgo y preferencias de liquidez (Schär, 2021; Deloitte, 2023).

Entre las ventajas teóricas más citadas de la tokenización se encuentran la democratización del acceso a la inversión inmobiliaria, la potencial liquidez del mercado secundario, la reducción de costes de transacción y la mayor transparencia y trazabilidad de las operaciones. Al dividir los activos en unidades más pequeñas, se reduce la inversión mínimo de entrada, lo que permite a pequeños inversores participar en oportunidades que antes estaban reservadas a instituciones o grandes capitales. Además, el uso de registros distribuidos permite auditar y verificar transacciones sin depender de un intermediario centralizado, fortaleciendo la confianza y disminuyendo riesgos de fraude o error (Aste et al., 2017; Schär, 2021).

¹ Un **SPV (Special Purpose Vehicle o Vehículo de Propósito Especial)** es una entidad jurídica independiente creada específicamente para aislar los activos y riesgos de un proyecto o inversión.

² El modelo de **revenue-sharing** consiste en que los inversores reciben una proporción de los ingresos generados por el activo subyacente, como rentas o plusvalías, en lugar de ser propietarios directos del inmueble como en el modelo tradicional.

Sin embargo, la literatura también identifica una serie de limitaciones y riesgos que condicionan la aplicabilidad de estos modelos. En primer lugar, la liquidez real de los mercados secundarios sigue siendo limitada, dado que la adopción global aún es incipiente y las plataformas existentes manejan volúmenes relativamente bajos de transacción. En segundo lugar, persisten riesgos regulatorios y legales, ya que los tokens de seguridad deben cumplir con normativas financieras locales y, en muchos casos, internacionales, incluyendo requisitos de KYC/AML³, custodia de activos y registro de propiedad (Zetzsche et al., 2018; Gomber et al., 2017). Por otro lado, la dependencia de plataformas centralizadas para la emisión y negociación de tokens introduce un riesgo operativo que no se encuentra en mercados tradicionales más consolidados.

Un aspecto adicional relevante es la interoperabilidad tecnológica y la estandarización de tokens. La utilización de estándares como ERC-20, ERC-1400 o ERC-777 permite que los tokens sean compatibles con distintos ecosistemas de blockchain y faciliten su integración en plataformas secundarias de negociación. No obstante, la diversidad de estándares y la falta de uniformidad regulatoria crean incertidumbre sobre la transferibilidad y la validez legal de los derechos asociados (Cong & He, 2019; Catalini & Gans, 2016). Este vacío normativo y técnico refuerza la idea de que, aunque el potencial teórico es alto, la implementación práctica aún está en fase de experimentación.

En términos financieros, estudios recientes sugieren que los activos tokenizados pueden ofrecer rentabilidades comparables a las del mercado inmobiliario tradicional, aunque con perfiles de riesgo diferenciados, principalmente asociados a la volatilidad del ecosistema digital, riesgos operativos de las plataformas y la incertidumbre regulatoria (Allen et al., 2022; Fabian Schär, 2021).

Sin embargo, la evidencia empírica disponible sigue siendo limitada y, en muchos casos, basada en estudios de caso o análisis agregados. En este contexto, la aplicación de técnicas de análisis de datos y *Machine Learning* exploratorio —como se propone en este trabajo— permite avanzar más allá del enfoque descriptivo predominante en la literatura, facilitando la identificación sistemática de patrones en variables como la rentabilidad, la liquidez y la estructura de los activos tokenizados. Este enfoque proporciona una base cuantitativa más

³ **KYC (Know Your Customer)** y **AML (Anti-Money Laundering)** son procedimientos regulatorios que buscan prevenir el fraude y el lavado de dinero. KYC implica que la plataforma verifica la identidad de los inversores y su perfil financiero antes de permitirles participar en la inversión. AML se centra en monitorizar y reportar actividades sospechosas relacionadas con el flujo de fondos.

robusta para evaluar su comportamiento y comparar su desempeño con el mercado inmobiliario tradicional.

Finalmente, la literatura apunta que la tokenización de activos no es solo un fenómeno financiero, sino también un cambio estructural en la relación entre tecnología y mercado. Su relevancia radica en la posibilidad de redefinir la accesibilidad, liquidez y eficiencia de los mercados inmobiliarios, abriendo la puerta a nuevos modelos de inversión y a la creación de un ecosistema más transparente, pero sujeto a riesgos que deben gestionarse cuidadosamente. Esta visión teórico-práctica constituye la base conceptual sobre la cual se desarrollará el análisis empírico de este trabajo, permitiendo contrastar los beneficios prometidos de la tokenización con los desafíos reales observables en los datos disponibles.

Desde una perspectiva institucional, la literatura reciente coincide en que la tokenización inmobiliaria no se desarrolla en un vacío normativo, sino dentro del marco regulatorio financiero existente. A diferencia de otras innovaciones tecnológicas, su principal reto no radica únicamente en la viabilidad técnica, sino en su encaje jurídico dentro de los mercados de capitales (Zetsche et al., 2018; Gomber et al., 2017).

En el ámbito europeo, el desarrollo regulatorio ha avanzado significativamente con la aprobación del Reglamento MiCA, que establece un marco armonizado para los criptoactivos. No obstante, dicho reglamento excluye aquellos tokens que puedan ser considerados instrumentos financieros, los cuales quedan sujetos a la Directiva MiFID II. Esta distinción es especialmente relevante en el ámbito inmobiliario, donde la mayoría de los tokens representan derechos económicos asimilables a valores negociables (Zetsche et al., 2018).

En consecuencia, la emisión y comercialización de estos activos se encuentra sujeta a los requisitos tradicionales del mercado financiero, incluyendo obligaciones de transparencia, protección del inversor y supervisión por parte de autoridades como la Comisión Nacional del Mercado de Valores (CNMV, 2022). A su vez, iniciativas como el Régimen Piloto DLT reflejan un esfuerzo institucional por facilitar la integración de tecnologías blockchain en infraestructuras de mercado reguladas, especialmente en lo relativo a negociación y liquidación de instrumentos tokenizados (European Commission, 2022).

Desde una perspectiva comparada, organismos como la U.S. Securities and Exchange Commission han adoptado un enfoque basado en la naturaleza económica del activo, aplicando criterios como el *Howey Test* para determinar si un token constituye un valor. Este enfoque refuerza la idea de que la tokenización inmobiliaria, lejos de constituir un sistema paralelo, se integra dentro del perímetro regulatorio existente (SEC, 2019; Zetsche et al., 2018).

En este contexto, la literatura destaca que las diferencias entre el mercado inmobiliario tradicional y la tokenización no se encuentran en el activo subyacente, sino en la forma en que se estructuran la propiedad, la liquidez y los mecanismos de intermediación. La tokenización introduce modelos de fraccionamiento que reducen barreras de entrada, así como la posibilidad de mercados secundarios más ágiles, aunque condicionados por la regulación y la adopción real (Catalini & Gans, 2016; Schär, 2021). Asimismo, permite automatizar procesos mediante *smart contracts*, mejorando la eficiencia operativa, pero incorporando nuevos riesgos tecnológicos y regulatorios (Yermack, 2017; Aste et al., 2017).

En conjunto, el marco regulatorio actual configura un entorno en el que la innovación tecnológica coexiste con estructuras legales tradicionales, condicionando tanto el diseño de los productos tokenizados como su accesibilidad, liquidez y perfil de riesgo. Este contexto resulta fundamental para interpretar los resultados empíricos del presente trabajo, ya que limita y, al mismo tiempo, define el comportamiento observable de los activos analizados.

Capítulo 4. DEFINICIÓN DEL TRABAJO

4.1 JUSTIFICACIÓN

El mercado inmobiliario tradicional ha demostrado históricamente estabilidad y capacidad de generación de rentas a través del arrendamiento y la apreciación del capital. Sin embargo, su estructura presenta limitaciones bien documentadas: iliquidez estructural, elevados costes de transacción y barreras de entrada que dificultan la diversificación de pequeños inversores. La tokenización inmobiliaria surge como una alternativa tecnológica que promete fraccionamiento digital de la propiedad, automatización de flujos económicos mediante *smart contracts* y la posible creación de mercados secundarios más ágiles. No obstante, más allá del discurso tecnológico, la evidencia empírica sobre su comportamiento real en términos de rentabilidad, liquidez y riesgo sigue siendo limitada y fragmentaria.

El presente Trabajo Fin de Grado desarrolla un marco analítico comparativo basado en análisis de datos y técnicas de aprendizaje automático para evaluar si la tokenización reproduce la lógica tradicional de formación de valor o introduce una estructura económica diferenciada. A través de un pipeline homogéneo aplicado a datasets representativos de ambos mercados, se analizan métricas de dispersión, correlación, segmentación por deciles, regresión y clustering. Este enfoque permite no solo cuantificar diferencias en varianza y poder explicativo —mostrando una fuerte dependencia de fundamentos urbanos en el mercado tradicional frente a una comprensión estructural del rendimiento en el tokenizado— sino también identificar patrones internos que revelan una menor sensibilidad espacial del yield tokenizado y una mayor influencia de la arquitectura financiera del activo.

4.1.1 BRECHA TECNOLÓGICA Y DE MERCADO

Aunque existen estudios relevantes sobre tokenización y blockchain aplicados a activos financieros, la literatura específica sobre tokenización inmobiliaria presenta todavía importantes carencias estructurales. En primer lugar, la mayoría de los trabajos se centran en desarrollos conceptuales, marcos teóricos o análisis del potencial tecnológico de blockchain, sin incorporar evidencia empírica comparativa frente al mercado inmobiliario tradicional (Catalini & Gans, 2016; Schär, 2021). La discusión suele enfatizar ventajas teóricas como liquidez, democratización o eficiencia operativa, pero ofrece escasa validación cuantitativa sobre si dichas promesas se materializan en datos observables.

En segundo lugar, no se identifican estudios que integren de forma sistemática análisis exploratorio de datos, segmentación por deciles, modelización estadística y técnicas de clustering para evaluar conjuntamente rentabilidad, dispersión y sensibilidad a fundamentos urbanos. La literatura existente tiende a analizar la tokenización como fenómeno tecnológico o jurídico, pero raramente como estructura económica susceptible de contraste cuantitativo frente a benchmarks tradicionales.

Por último, la disponibilidad limitada de datos públicos sobre activos tokenizados dificulta la replicabilidad y el desarrollo de análisis comparativos robustos. Muchas plataformas operan con información fragmentaria o no estandarizada, lo que impide evaluar patrones estructurales de rentabilidad y liquidez con criterios homogéneos.

Así pues, este proyecto propone paliar todas estas carencias mediante un análisis cuantitativo que permita mapear el comportamiento de los activos tokenizados, detectar tendencias estructurales y compararlas con propiedades tradicionales, creando un marco interpretativo reproducible y transparente.

4.1.2 RELEVANCIA CIENTÍFICA Y SOCIAL

El trabajo presenta relevancia tanto metodológica como económica y regulatoria. En primer lugar, introduce un enfoque comparativo poco habitual en la literatura sobre tokenización inmobiliaria, combinando análisis exploratorio, contraste estadístico, regresión y segmentación para evaluar empíricamente la estructura de ambos mercados.

En segundo lugar, aporta evidencia cuantitativa sobre una cuestión central en el debate actual: si la tokenización realmente mejora accesibilidad y liquidez o si estas ventajas permanecen, por ahora, en el plano teórico. El análisis permite evaluar hasta qué punto el fraccionamiento digital modifica la formación de valor y el perfil de riesgo.

Asimismo, la metodología desarrollada es replicable y transferible a otros activos, plataformas o jurisdicciones, lo que amplía su utilidad más allá del caso concreto estudiado.

Por último, el trabajo conecta directamente con el marco regulatorio europeo y estadounidense, aportando elementos de análisis útiles para reguladores que deben supervisar estos instrumentos bajo normativas como MiCA y MiFID II.

4.1.3 OBJETIVOS DEL PROYECTO

- **Comparación estructural entre mercados:** analizar diferencias en dispersión, rentabilidad, liquidez y poder explicativo de variables urbanas entre inmuebles tradicionales y tokenizados.
- **Evaluación de formación de valor:** determinar si el rendimiento tokenizado responde a fundamentos espaciales similares a los del mercado tradicional o si presenta una lógica diferenciada.
- **Segmentación y detección de patrones:** aplicar técnicas de clustering para identificar estructuras internas y agrupaciones homogéneas dentro de cada mercado.
- **Análisis regulatorio contextualizado:** interpretar los resultados a la luz del marco normativo vigente y sus implicaciones para inversores y supervisores.
- **Síntesis operativa:** traducir los hallazgos en conclusiones claras sobre viabilidad, límites y potencial de la tokenización inmobiliaria.

4.2 METODOLOGÍA

El presente trabajo adopta un enfoque cuantitativo y exploratorio basado en la comparación entre el mercado inmobiliario tradicional y el mercado inmobiliario tokenizado. La metodología se estructura como un pipeline analítico que combina preprocesado de datos, análisis exploratorio, modelización estadística y técnicas de aprendizaje no supervisado, con el objetivo de identificar diferencias estructurales en la formación de valor entre ambos mercados.

En primer lugar, se lleva a cabo un proceso de preprocesado orientado a garantizar la coherencia y comparabilidad de los datos. Este proceso incluye la limpieza de valores nulos, la transformación de variables y la construcción de métricas derivadas, como el precio por metro cuadrado y el rendimiento anual (yield). Asimismo, se realiza una estandarización de variables cuando es necesario, especialmente en el contexto de técnicas de segmentación, con el fin de evitar sesgos derivados de escalas heterogéneas.

A continuación, el análisis se desarrolla en cuatro fases principales:

4.2.1 ANÁLISIS EXPLORATORIO DE DATOS

Se realiza una exploración descriptiva de ambos datasets con el objetivo de comprender la distribución, dispersión y comportamiento de las variables clave. Este análisis incluye el

estudio de histogramas, boxplots y estadísticas descriptivas, así como la identificación de outliers y asimetrías.

Asimismo, se analizan relaciones bivariantes entre variables relevantes —como precio y localización en el mercado tradicional, o rendimiento y variables urbanas en el mercado tokenizado— con el fin de detectar patrones preliminares y diferencias estructurales entre ambos entornos.

De forma complementaria, se emplea segmentación por deciles para analizar el comportamiento de los extremos de la distribución, lo que permite identificar posibles diferencias en los activos de mayor valor o rentabilidad.

4.2.2 ANÁLISIS DE RELACIONES Y CAPACIDAD EXPLICATIVA

Con el objetivo de evaluar en qué medida las variables observables explican el comportamiento de los activos, se aplican modelos de regresión lineal en ambos mercados.

En el mercado tradicional, el modelo permite analizar la relación entre el precio por metro cuadrado y variables estructurales y urbanas, mientras que en el mercado tokenizado se estudia la capacidad de estas variables para explicar el rendimiento (yield).

La evaluación de los modelos se realiza mediante el coeficiente de determinación (R^2), lo que permite comparar la capacidad explicativa en ambos entornos y analizar si el valor o la rentabilidad responden a fundamentos observables.

4.2.3 ANÁLISIS DE SEGMENTACIÓN (CLUSTERING)

Se aplican técnicas de aprendizaje no supervisado, principalmente el algoritmo K-Means, sobre variables previamente estandarizadas, con el objetivo de identificar estructuras internas en los datos.

La calidad de la segmentación se evalúa mediante el índice silhouette, que permite medir el grado de cohesión interna de los clusters y su separación respecto a otros grupos. Este análisis se utiliza para comparar la claridad estructural entre el mercado tradicional y el tokenizado, evaluando si existen agrupaciones homogéneas de activos en función de sus características.

4.2.4 SÍNTESIS E INTERPRETACIÓN DE RESULTADOS

Finalmente, se integran los resultados obtenidos en las distintas fases del análisis para construir una interpretación estructural de ambos mercados. Esta síntesis combina métricas de dispersión, correlaciones, capacidad explicativa de los modelos y resultados de segmentación, apoyándose en representaciones visuales como histogramas, diagramas de dispersión y mapas de calor.

El objetivo de esta fase no es únicamente describir los resultados, sino interpretar su significado económico, evaluando en qué medida la tokenización inmobiliaria reproduce o altera los mecanismos tradicionales de formación de valor.

Capítulo 5. ANÁLISIS EXPLORATORIO DE DATOS

El análisis exploratorio de datos (EDA) constituye una fase esencial en el desarrollo de este trabajo, ya que permite comprender la estructura interna de los *datasets*, detectar patrones preliminares y validar la coherencia de las variables antes de la aplicación de técnicas más avanzadas. En el contexto de este estudio, el EDA no se limita a una descripción estadística, sino que se orienta a identificar diferencias estructurales en la formación de valor entre el mercado inmobiliario tradicional y el mercado tokenizado.

Las variables analizadas se agrupan en cuatro bloques principales. En primer lugar, se consideran las variables financieras, como el precio del activo, el precio por metro cuadrado y la rentabilidad anual (*yield*), que permiten evaluar el retorno económico y establecer comparaciones entre ambos mercados. En segundo lugar, se analizan las variables de liquidez, aproximadas mediante el número de transacciones en el mercado secundario en el caso tokenizado, lo que permite evaluar la capacidad de negociación de los activos y su potencial de desinversión.

Asimismo, se incorporan variables estructurales del inmueble, como superficie, número de habitaciones o antigüedad, junto con variables urbanas como la distancia al centro urbano, la accesibilidad o el nivel de servicios disponibles (*Walk Score*, distancia a *MRT*). Estas variables son fundamentales en la literatura inmobiliaria tradicional y permiten analizar si los fundamentos del activo continúan siendo relevantes en el contexto tokenizado.

Finalmente, se consideran variables temporales y de historial de transacciones, que permiten capturar posibles dinámicas de evolución y volatilidad. El enfoque metodológico combina limpieza de datos, transformación de variables (como el cálculo de precio por metro cuadrado y *yield*), análisis descriptivo, visualización de distribuciones y segmentación por percentiles. Este proceso prepara la base para la posterior aplicación de técnicas de *Machine Learning* orientadas a la detección de patrones.

5.1 DATASETS

El análisis empírico desarrollado en este trabajo se apoya en tres conjuntos de datos diferenciados, agrupados en dos bloques conceptuales: mercado inmobiliario tradicional y mercado inmobiliario tokenizado.

5.1.1 MERCADO TRADICIONAL

En el caso del mercado tradicional, se han utilizado dos *datasets* públicos disponibles en Kaggle. El primero corresponde a un conjunto de transacciones inmobiliarias reales, que recoge información detallada sobre precios de venta (*Sale Price*), valoraciones estimadas, número de habitaciones y baños, superficie construida (*carpet_area*) y características estructurales del inmueble. Este conjunto de datos, con más de 6.000 observaciones, permite capturar una elevada heterogeneidad espacial y segmentación interna del mercado.

El segundo dataset incorpora variables urbanas explícitas, como la distancia a estaciones de transporte público (*distance_mrt*), el número de comercios cercanos, coordenadas geográficas y antigüedad del inmueble. Este conjunto permite analizar el precio desde una doble perspectiva: como resultado de atributos físicos y como función de variables espaciales observables, alineándose con la literatura clásica de valoración inmobiliaria.

Es importante destacar que los inmuebles de ambos datasets no coinciden necesariamente en localización ni en periodo temporal, por lo que el análisis no se basa en una comparación directa entre activos individuales, sino en una comparación agregada entre distribuciones de activos con características similares. Este enfoque permite evitar sesgos derivados de la falta de equivalencia directa y es consistente con estudios empíricos en mercados heterogéneos.

5.1.2 MERCADO TOKENIZADO

En el caso del mercado tokenizado, se utiliza el dataset *Real_Estate_Token_Data.xlsx*, procedente de Mendeley Data. Este conjunto incluye 173 observaciones correspondientes a activos tokenizados, principalmente emitidos a través de la plataforma *RealT*. A diferencia del mercado tradicional, este *dataset* presenta una elevada densidad financiera por observación, incorporando variables como rentas brutas y netas, precio del token, número total de tokens emitidos, inversión total, rendimiento esperado, costes operativos y métricas de liquidez (número de transferencias).

Asimismo, incluye variables urbanas comparables a las utilizadas en el mercado tradicional, como distancia al centro (*Distance_DTWN*), *Walk Score* o antigüedad, lo que permite mantener un marco metodológico homogéneo para la comparación entre ambos modelos. La disponibilidad de estas variables resulta especialmente relevante para analizar si los fundamentos urbanos continúan influyendo en el rendimiento de los activos tokenizados.

Cabe destacar que, debido a la novedad del fenómeno de la tokenización inmobiliaria y a la limitada disponibilidad de datos públicos, el uso de este tipo de datasets aún es reducido en la literatura académica, lo que refuerza el carácter exploratorio y la contribución empírica del presente trabajo.

5.1.3 CONSISTENCIA Y COMPARABILIDAD ENTRE BASES DE DATOS

A pesar de la diferente naturaleza de los datasets utilizados, existe un conjunto de variables comunes que permiten establecer comparaciones directas entre el mercado inmobiliario tradicional y el mercado tokenizado. En particular, variables como el precio total del activo, el precio por metro cuadrado (`price_m2`), la antigüedad del inmueble y determinados indicadores de localización urbana (como distancia al centro o accesibilidad medida mediante *Walk Score*) están presentes, con distinto grado de detalle, en ambas bases de datos. Esta coincidencia permite construir un marco de análisis homogéneo y facilita la comparación entre modelos, aunque es importante señalar que dichas variables pueden diferir en su forma de medición o escala. En consecuencia, el análisis se centra en comparaciones a nivel agregado y en la identificación de patrones estructurales, evitando interpretaciones directas a nivel individual entre activos.

5.2 *ESTADÍSTICA DESCRIPTIVA*

El análisis descriptivo de los datos permite identificar diferencias estructurales significativas entre el mercado inmobiliario tradicional y el mercado tokenizado, tanto en términos de distribución de variables como en los mecanismos subyacentes de formación de valor.

En primer lugar, el mercado inmobiliario tradicional, representado por el dataset de transacciones, muestra una elevada heterogeneidad en los precios. Tal y como se observa en la Figura 5.1, la distribución del precio por metro cuadrado presenta una fuerte asimetría positiva, con una alta concentración de observaciones en valores relativamente bajos y una

cola larga hacia valores extremos. La diferencia entre la media y la mediana, visible en la figura, confirma la presencia de outliers significativos y una distribución no normal.

Figura 1 Distribución del precio por metro cuadrado en el mercado tradicional

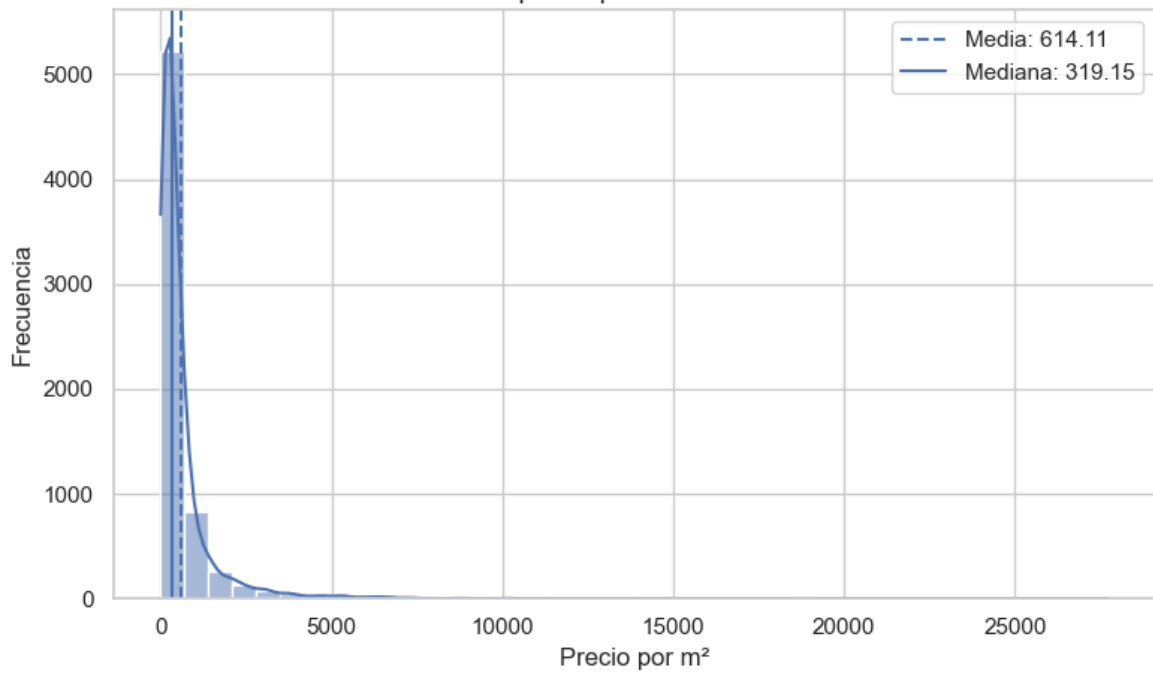


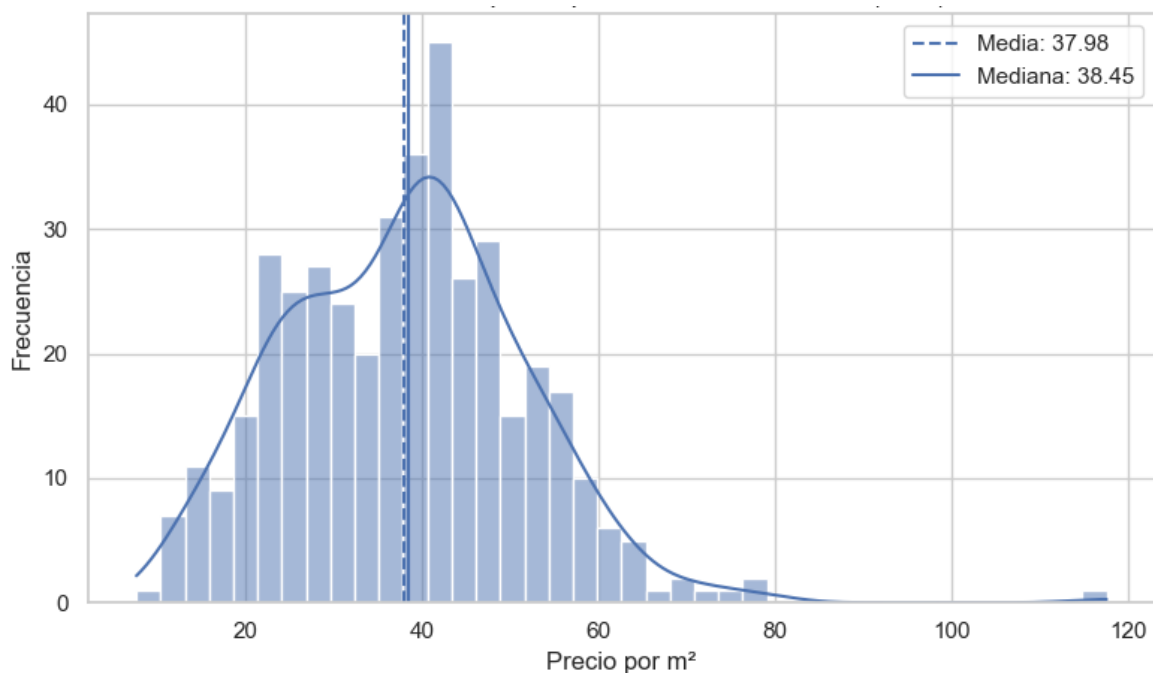
Tabla 1 Estadística descriptiva Mercado Tradicional

Variable	Media	Desv. Std	Mínimo	P25	Mediana	P75	Máximo
Precio de venta (€)	626,552	1,077,250	2,500	162,000	338,000	627,625	2,750,000
Precio por m ² (€)	614.11	1087.7	1.4	143.79	319.15	618.77	27,717.98
Número de habitaciones	3.33	0.88	3	3	3	3	8
Número de baños	2.33	1.21	1	1	2	3	8
Superficie (m ²)	1110.89	301.36	900	960	1022	1082	2989

Esta dispersión se refleja también en la estadística descriptiva (Tabla 1), donde la desviación estándar supera ampliamente el valor medio, lo que indica una elevada variabilidad entre activos. Este comportamiento es consistente con la existencia de múltiples segmentos dentro del mercado inmobiliario, incluyendo propiedades estándar y activos de alto valor, cuya presencia distorsiona la distribución global.

Por su parte, el dataset urbano (MRT) presenta un comportamiento sustancialmente distinto. Como se muestra en la Figura 5.2, la distribución del precio por metro cuadrado es más simétrica y concentrada, con menor dispersión y una menor presencia de valores extremos. Este resultado sugiere que, en este contexto, el precio está más estrechamente vinculado a variables estructurales y espaciales observables, como la accesibilidad o la proximidad a servicios.

Figura 2 Distribución del precio por metro cuadrado en el dataset urbano (MRT)



En conjunto, estos dos datasets del mercado tradicional evidencian que, aunque existe heterogeneidad, el precio inmobiliario sigue respondiendo a fundamentos del activo y del entorno, lo que permite su modelización mediante variables explicativas.

En contraste, el mercado inmobiliario tokenizado presenta una configuración significativamente diferente. A pesar de que el precio por metro cuadrado de los activos tokenizados se sitúa en rangos comparables al mercado tradicional, su distribución presenta características propias.

La Figura 3 muestra la distribución del precio por metro cuadrado en el mercado tokenizado. A diferencia del mercado inmobiliario tradicional, donde la dispersión es elevada y existen múltiples segmentos claramente diferenciados, la distribución en este caso presenta una mayor concentración en torno a valores medios, con una dispersión moderada. La proximidad entre la media (690,82 €/m²) y la mediana (624,55 €/m²) sugiere una menor asimetría, aunque se observan algunos valores extremos que generan una ligera cola hacia la derecha

Este resultado indica que, si bien el valor del activo subyacente mantiene cierta heterogeneidad, esta no se traduce de forma proporcional en el comportamiento del rendimiento.

Tal y como se observa en la Figura 4, la distribución del rendimiento anual (*yield_real*) presenta una fuerte concentración en torno al 11%, con una dispersión muy reducida en comparación con la observada en el mercado tradicional.

Figura 3 Distribución del precio por metro cuadrado en el dataset tokenizado

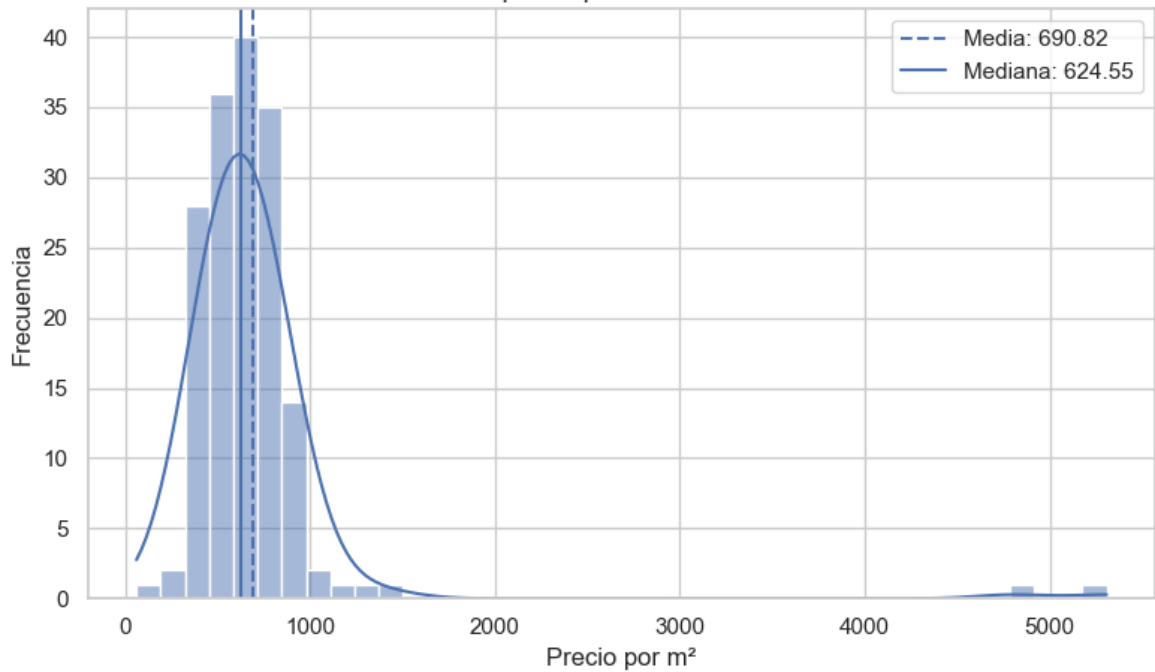
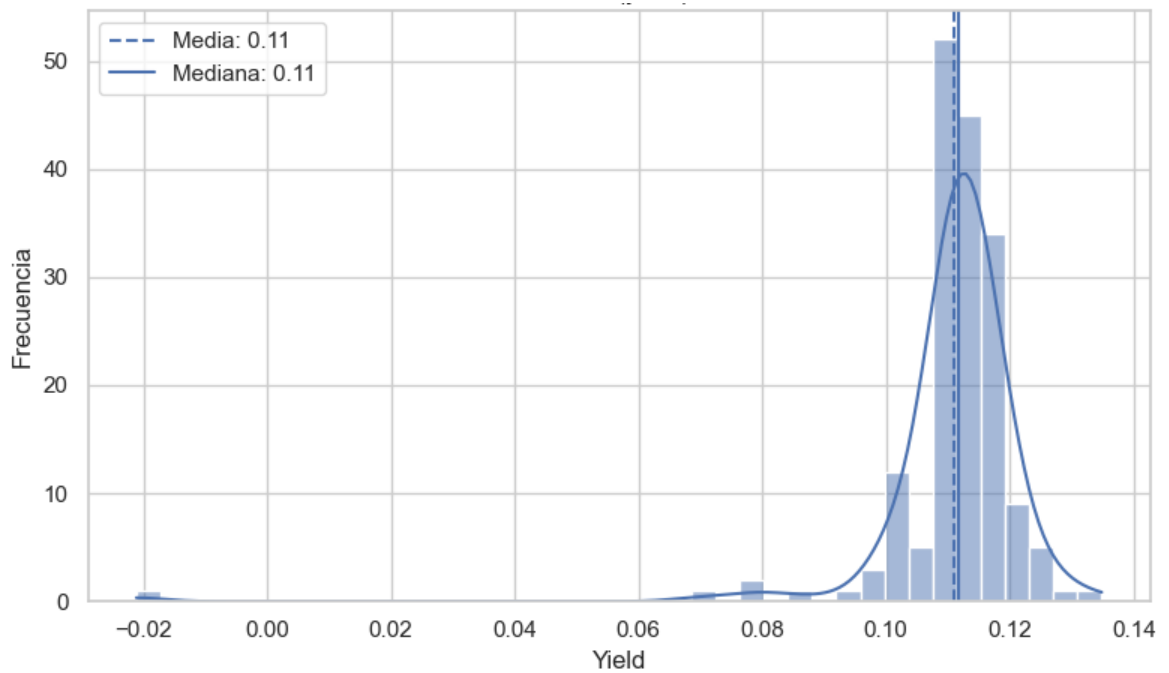


Figura 4 Distribución del rendimiento en el mercado tokenizado



Esta compresión de la rentabilidad se refleja también en la Tabla 2, donde la desviación estándar del *yield* es significativamente menor que la del precio en el mercado tradicional. En particular, el rango intercuartílico del rendimiento es extremadamente estrecho, lo que implica que la mayoría de los activos tokenizados ofrecen rentabilidades muy similares.

Tabla 2 Estadística descriptiva Mercado Tokenizado

Variable	Count	Mean	Std	Min	25%	Media n	75%	Max
price_total	173	168,040.93	205,537.26	48,080.00	60,576.00	66,599.00	144,450.00	985,910.00
price_m2	163	690.82	525.53	61.19	492.40	624.55	755.03	5,307.49
yield_real	173	0.11	0.01	-0.02	0.11	0.11	0.12	0.13
Transfers	173	83.51	203.36	0.00	1.00	15.00	49.00	1,972.00
Walk_Score	173	51.57	14.75	13.00	42.00	51.00	63.00	79.00

Este resultado constituye una diferencia estructural clave: mientras que en el mercado tradicional el valor emerge de la interacción entre características del activo y condiciones del entorno, en el mercado tokenizado se observa una disociación entre el valor del activo subyacente y su rendimiento. Aunque el precio por metro cuadrado presenta cierta heterogeneidad, el rendimiento permanece predefinido dentro de un rango estrecho, independientemente de factores como el tamaño del activo, su localización o su nivel de actividad en el mercado secundario.

En consecuencia, el análisis descriptivo preliminar apunta a que el proceso de formación de valor en la tokenización inmobiliaria podría estar más condicionado por la arquitectura financiera del producto y el diseño del token que por la heterogeneidad espacial del activo subyacente. Esta hipótesis será analizada en mayor profundidad en los siguientes apartados mediante el estudio de relaciones entre variables y técnicas de segmentación.

5.3 VARIABLES PRINCIPALES ANALIZADAS

En el mercado tradicional, la variable central del análisis es el precio por metro cuadrado, entendido como una medida del valor del activo ajustada por tamaño. Esta variable se analiza en función de características estructurales del inmueble y variables urbanas, lo que permite evaluar en qué medida el valor emerge de fundamentos observables.

En el mercado tokenizado, la variable principal es el *yield* real, definido como el ratio entre la renta neta anual y la inversión total. Esta métrica permite capturar la rentabilidad efectiva del activo desde la perspectiva del inversor.

No obstante, dado que el precio y el *yield* son magnitudes inversamente relacionadas, se ha procedido también a reconstruir el precio por metro cuadrado de los activos tokenizados, a partir del precio total (derivado del precio del token y el número total de tokens) y la superficie del inmueble. Este paso permite establecer una base de comparación más homogénea entre ambos mercados.

Adicionalmente, se incorporan variables como el número de transferencias (proxy de liquidez), el Walk Score y la distancia al centro, con el objetivo de analizar la relación entre rentabilidad, liquidez y fundamentos urbanos.

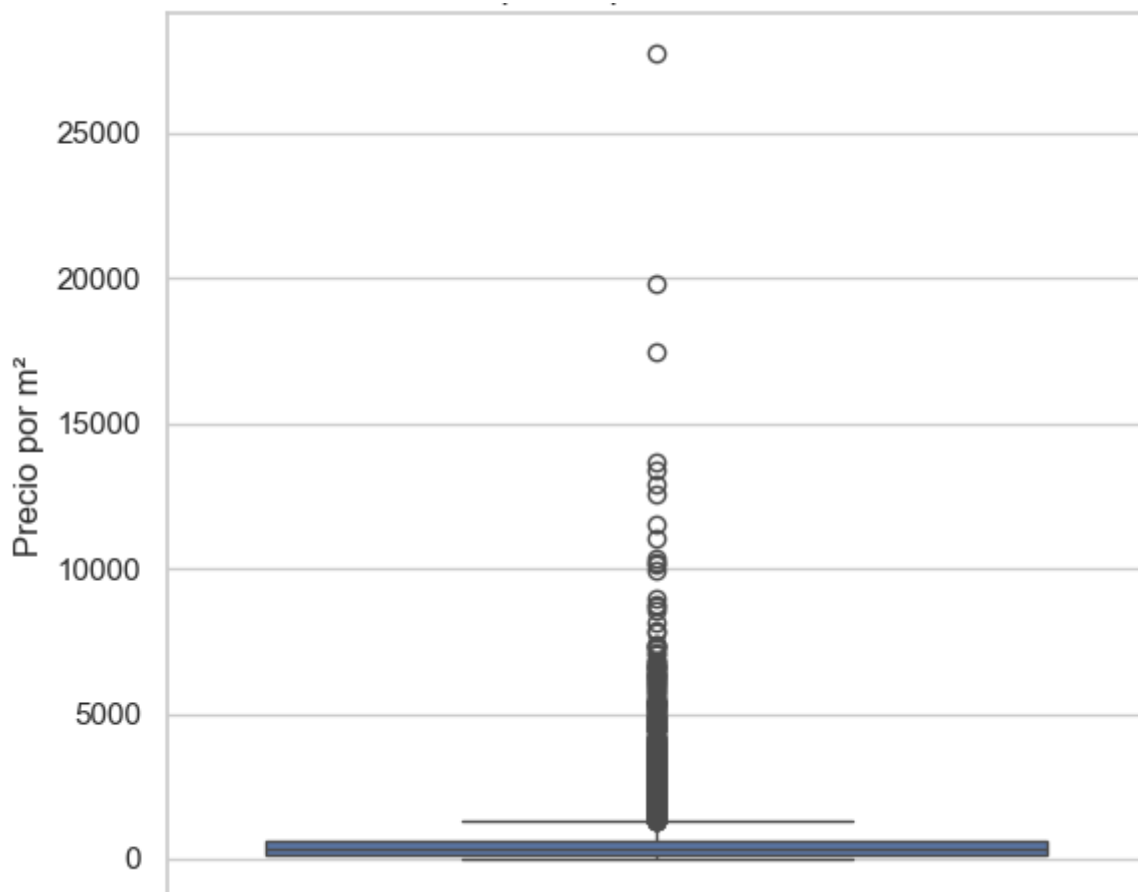
5.4 DISTRIBUCIONES E INTERPRETACIÓN PRELIMINAR

Una vez descritas las distribuciones en el apartado anterior, el análisis se centra en su interpretación a partir de medidas de dispersión y la identificación de valores atípicos, con el objetivo de comprender las diferencias estructurales entre ambos mercados.

En el mercado inmobiliario tradicional, la Figura 5 pone de manifiesto una dispersión extremadamente elevada en el precio por metro cuadrado. La mediana se sitúa en niveles

relativamente bajos dentro del rango total, mientras que la amplitud de los valores atípicos es muy significativa, alcanzando órdenes de magnitud muy superiores al grueso de las observaciones. La elevada concentración de outliers confirma que el mercado no solo es heterogéneo, sino que está fuertemente influido por una minoría de activos de alto valor.

Figura 5 Boxplot distribución del precio por metro cuadrado en el Mercado tradicional



Este comportamiento es consistente con una estructura segmentada, donde coexisten distintos submercados —desde vivienda estándar hasta propiedades premium—, lo que genera una distribución altamente desigual. Desde una perspectiva económica, esto refuerza la idea de que el precio en el mercado tradicional es el resultado de la interacción entre

múltiples factores, especialmente la localización, la calidad del activo y sus características específicas.

En contraste, la Figura 6 muestra que el mercado inmobiliario tokenizado presenta un comportamiento radicalmente distinto en términos de rentabilidad. El rendimiento (yield) se encuentra fuertemente concentrado en torno a un intervalo muy estrecho, con un rango intercuartílico reducido y una mediana prácticamente alineada con la media. Aunque existen algunos valores atípicos, estos son escasos y no alteran la estructura general de la distribución.

subyacente pueda variar en términos de valor, esta variación no se traduce en diferencias significativas en la rentabilidad ofrecida al inversor.

En consecuencia, estos resultados apuntan a que el mecanismo de formación de valor difiere sustancialmente entre ambos mercados. En el primero, el valor es endógeno y emerge de los fundamentos del activo y su entorno; en el segundo, el rendimiento parece venir determinado exógenamente por el diseño financiero del producto. Esta compresión del yield sugiere que la tokenización inmobiliaria introduce una capa de estandarización que transforma el activo en un instrumento más cercano a un producto financiero que a un activo inmobiliario tradicional.

Estas evidencias refuerzan la hipótesis de que la tokenización altera el proceso de formación de valor, desplazando el foco desde las características del inmueble hacia la estructuración del activo como vehículo de inversión. Este planteamiento será analizado en mayor profundidad en los siguientes apartados mediante el estudio de relaciones entre variables y técnicas de segmentación.

5.5 ANÁLISIS DE RELACIONES ENTRE VARIABLES

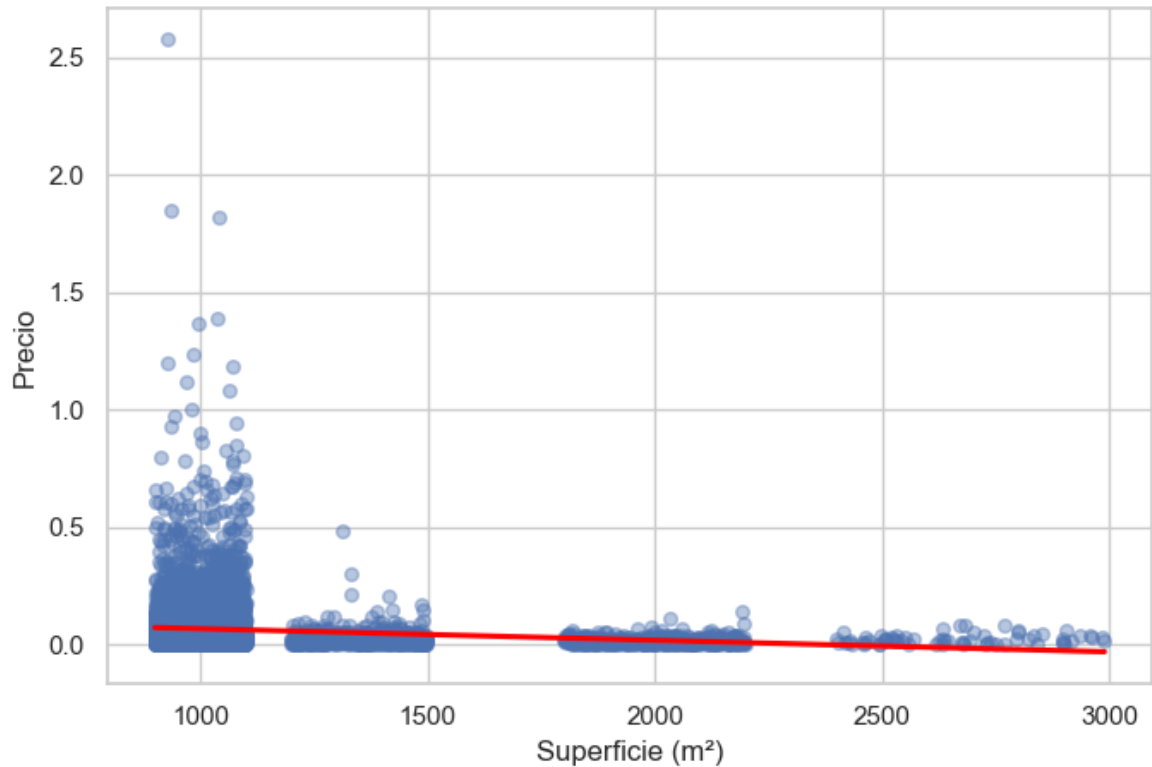
El análisis de relaciones entre variables permite profundizar en los mecanismos de formación de valor y comparar si el mercado tokenizado reproduce las dinámicas observadas en el mercado inmobiliario tradicional o, por el contrario, presenta un comportamiento diferenciado.

5.5.1 RELACIONES EN EL MERCADO INMOBILIARIO TRADICIONAL

En el mercado tradicional, las relaciones entre variables reflejan un comportamiento coherente con los fundamentos económicos del sector.

En primer lugar, la Figura 7 muestra la relación entre la superficie del inmueble y su precio. Se observa una elevada dispersión, pero con una ligera tendencia decreciente, lo que sugiere la presencia de economías de escala: los activos de mayor tamaño tienden a presentar un menor precio relativo o concentrarse en segmentos distintos del mercado. La dispersión también evidencia la influencia de otros factores relevantes, como la localización o la calidad del activo.

Figura 7 Relación entre tamaño (m²) y precio – Mercado tradicional

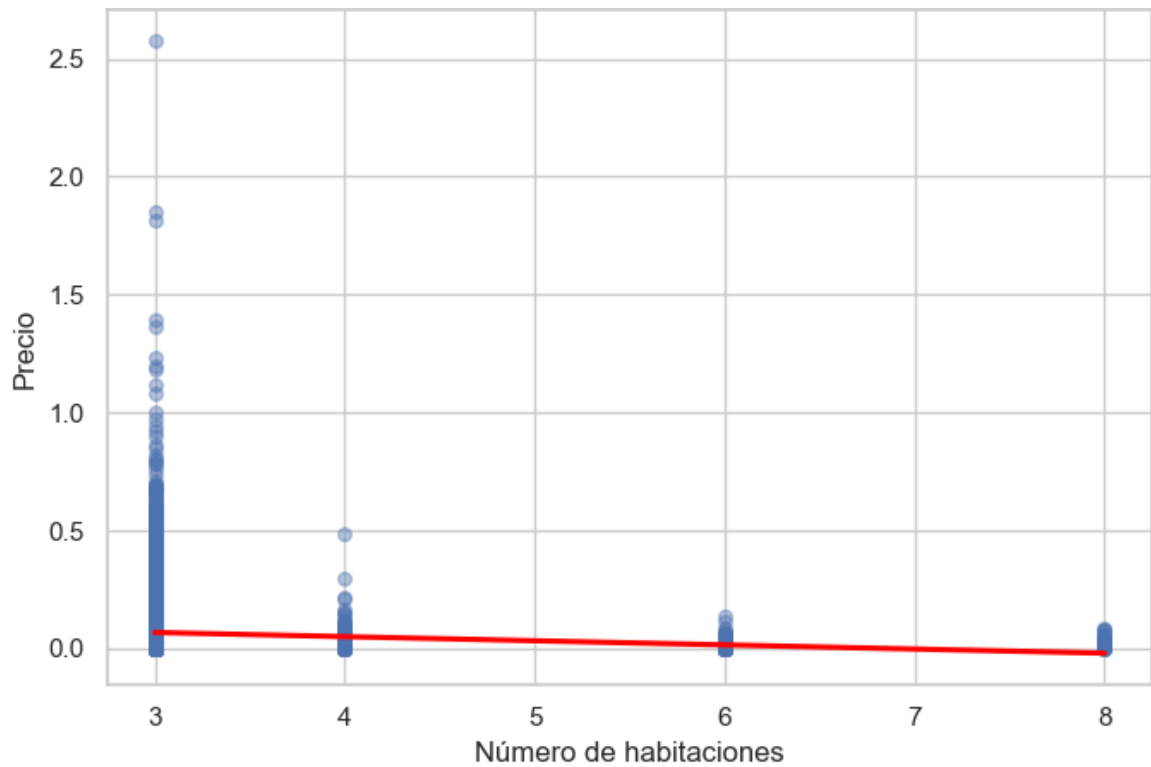


4

Por otro lado, la Figura 8 presenta la relación entre el número de habitaciones y el precio. La ausencia de una tendencia clara y la concentración de observaciones en ciertos valores discretos reflejan que esta variable, aunque relevante, no explica por sí sola la variabilidad del precio. Esto refuerza la idea de que el valor inmobiliario depende de la interacción de múltiples características.

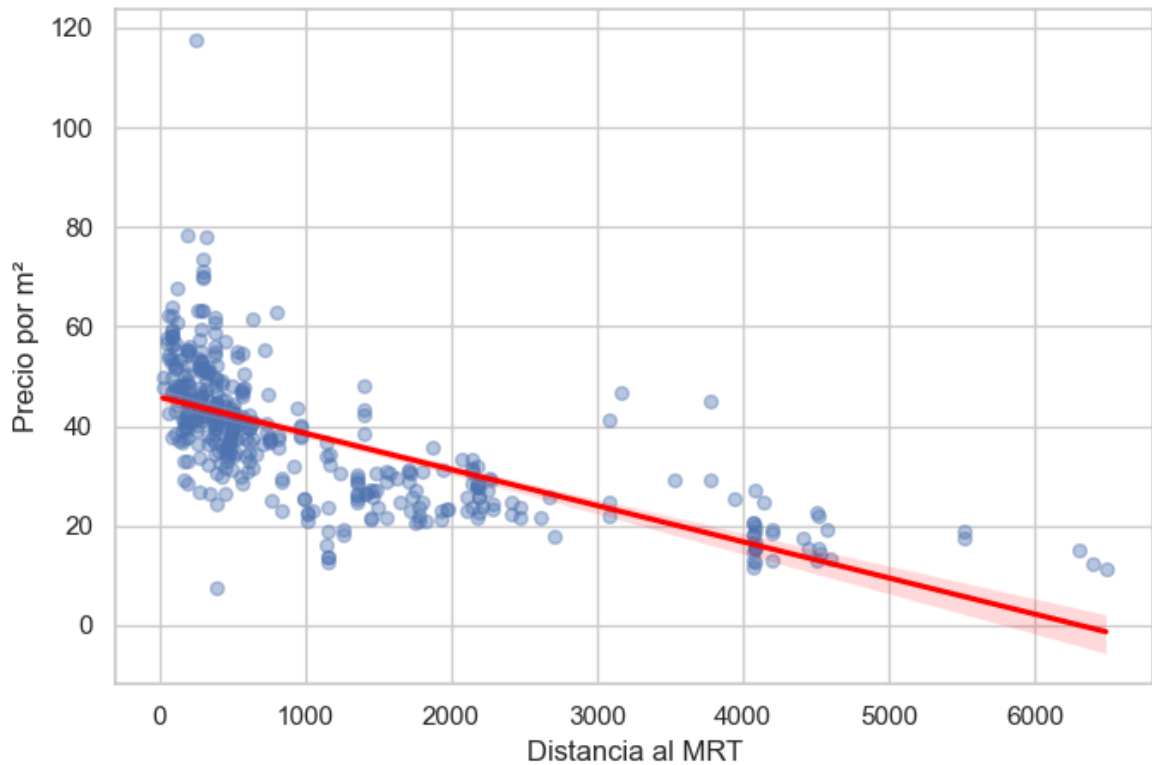
⁴ En el eje horizontal se representa la superficie del inmueble (m²), mientras que el eje vertical muestra el precio por unidad (normalizado), lo que permite analizar la relación entre tamaño y valor relativo del activo. Cada punto corresponde a una observación individual, y la línea roja representa una regresión lineal que resume la tendencia general de los datos, evidenciando una ligera relación decreciente entre superficie y precio relativo.

Figura 8 Relación entre n° de habitaciones y precio Mercado tradicional



Finalmente, la Figura 9 muestra una relación negativa clara entre el precio por metro cuadrado y la distancia al transporte público (*MRT*). Este resultado es consistente con la literatura inmobiliaria: los inmuebles mejor conectados presentan precios más elevados, evidenciando el papel fundamental de la accesibilidad en la formación de valor.

Figura 9 Relación entre precio por m² y distancia al MRT



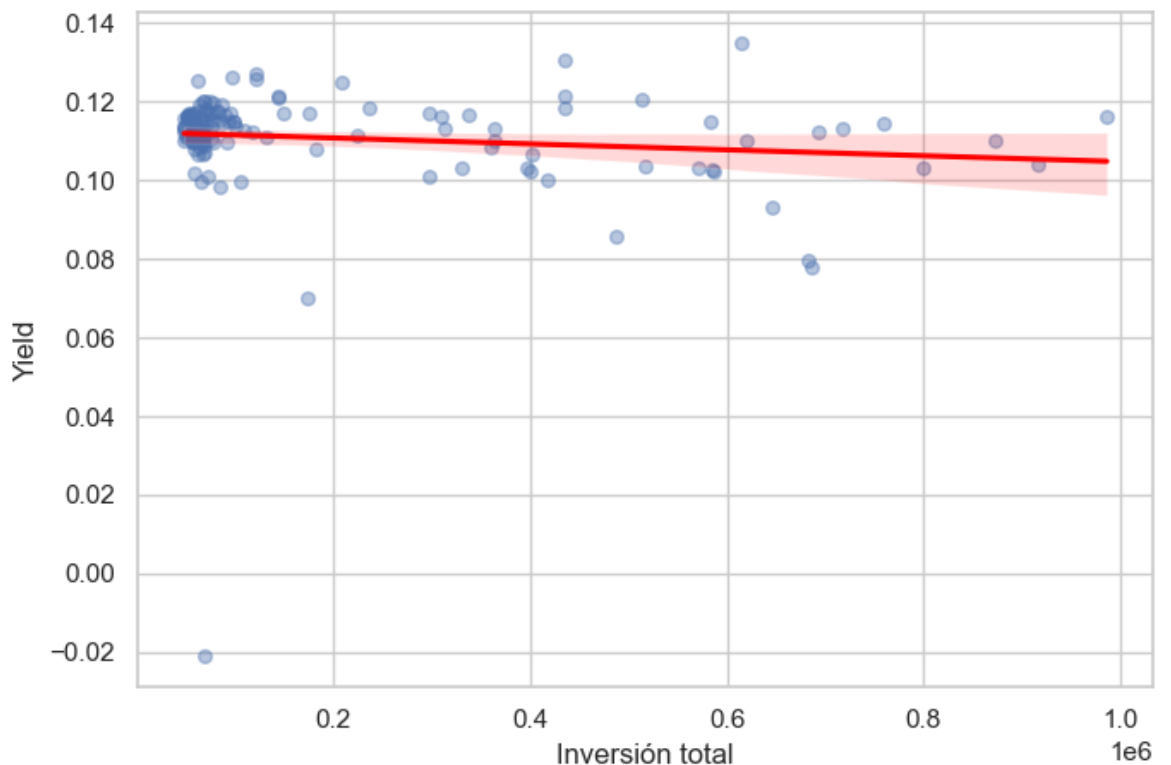
En conjunto, estas relaciones confirman que, en el mercado tradicional, el precio está determinado por factores estructurales del activo y del entorno, respondiendo a dinámicas de oferta y demanda.

5.5.2 RELACIONES EN EL MERCADO INMOBILIARIO TOKENIZADO

En contraste, el mercado tokenizado presenta un patrón radicalmente distinto, caracterizado por la ausencia de relaciones significativas entre el rendimiento y las variables analizadas.

La Figura 10 muestra la relación entre el rendimiento (yield) y la inversión total. La nube de puntos presenta una distribución prácticamente horizontal, con una pendiente cercana a cero, lo que indica que el tamaño del activo no influye en la rentabilidad ofrecida.

Figura 10 Relación entre inversión total y rendimiento-Mercado tokenizado



De manera similar, la Figura 11 analiza la relación entre el rendimiento y la liquidez, aproximada mediante el número de transferencias. A pesar de la elevada variabilidad en la liquidez, el rendimiento permanece concentrado en un rango muy estrecho, sin evidencia de una prima de liquidez.

Figura 11 Relación entre liquidez (número de transferencias) y rendimiento – Mercado tokenizado

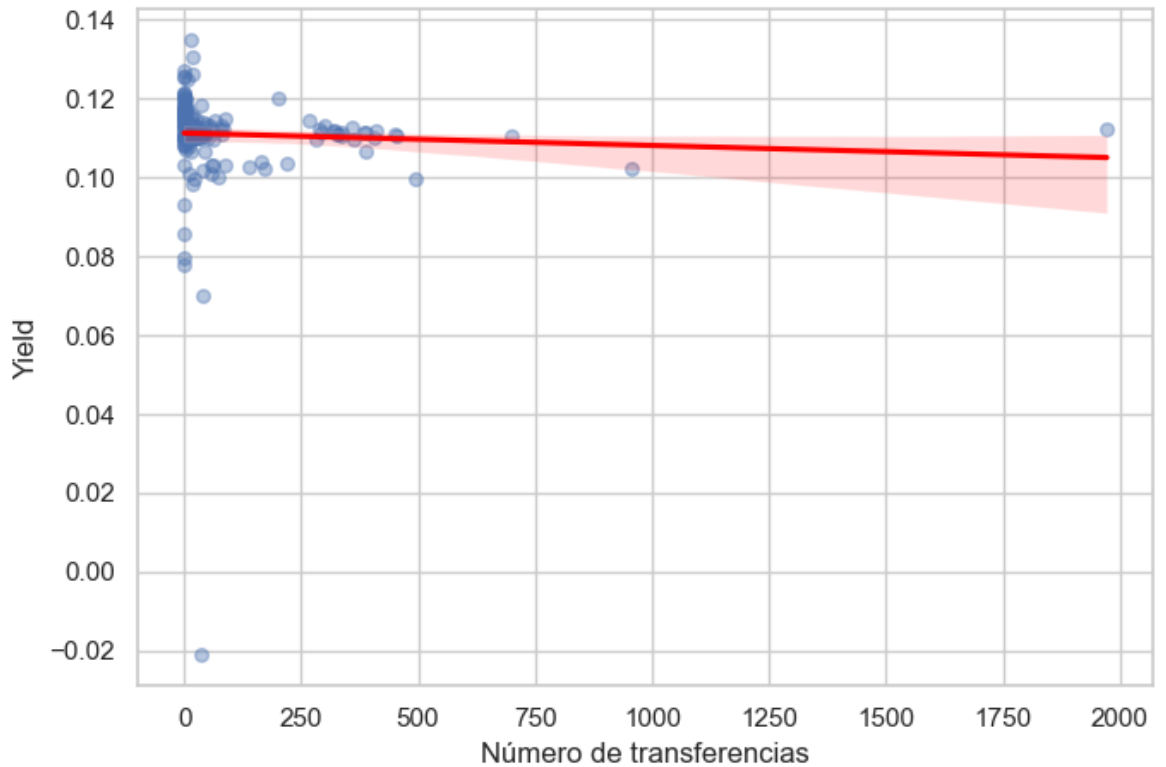
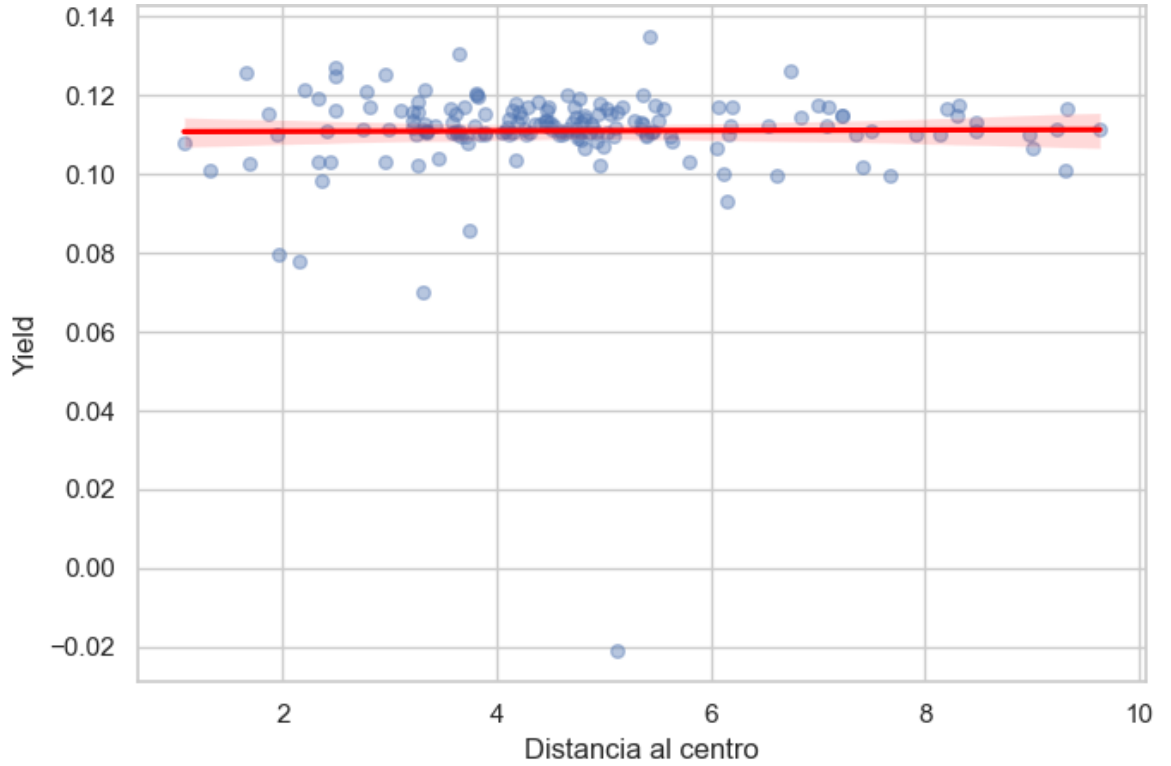


Figura 12 Relación entre liquidez (número de transferencias) y rendimiento – Mercado tokenizado

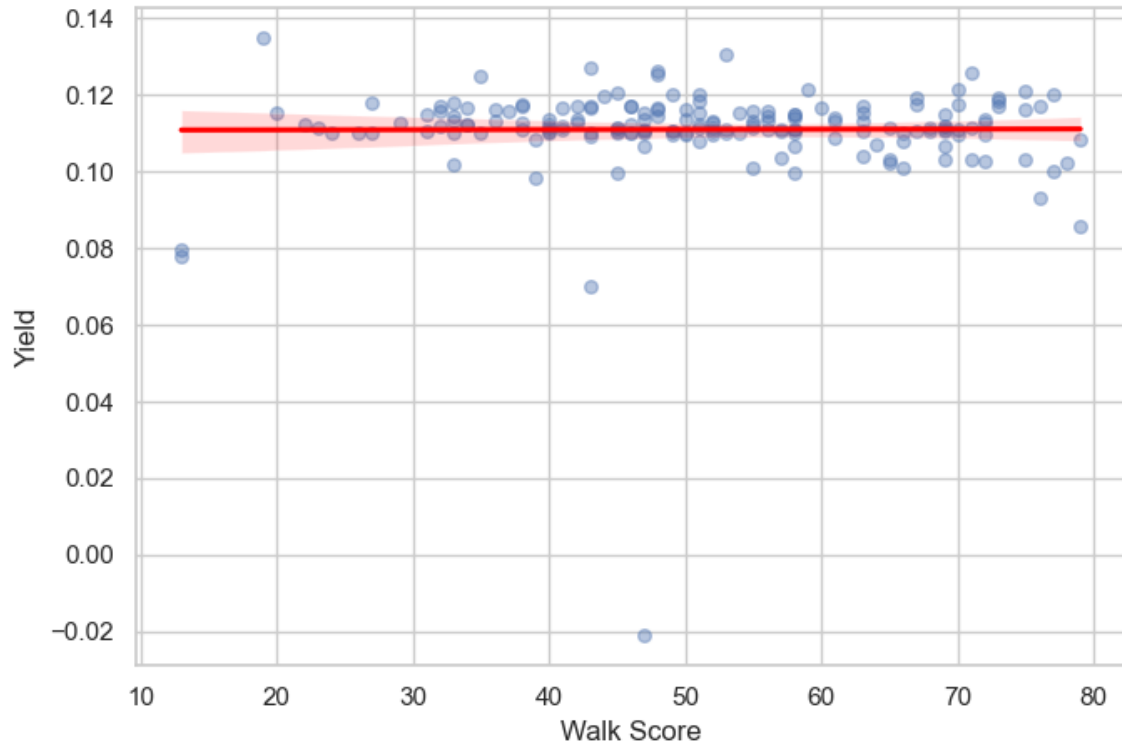
La Figura 12 presenta la relación entre el rendimiento y la distancia al centro urbano. La ausencia de pendiente en la línea de tendencia indica que la localización no tiene un impacto significativo en la rentabilidad, lo que contrasta con el comportamiento observado en el mercado tradicional.

Figura 13 Relación entre distancia al centro y rendimiento – Mercado tokenizado



Por su parte, la Figura 13 muestra la relación entre el rendimiento y la accesibilidad urbana, medida mediante el Walk Score. De nuevo, no se observa una relación definida, lo que refuerza la idea de que variables espaciales tradicionalmente relevantes pierden capacidad explicativa en este contexto.

Figura 14 Relación entre accesibilidad (Walk Score) y rendimiento – Mercado tokenizado



5.5.3 INTERPRETACIÓN CONJUNTA

El análisis comparado revela una diferencia estructural fundamental entre ambos mercados. Mientras que en el mercado tradicional las relaciones entre variables reflejan la influencia de factores económicos reales —como la localización, el tamaño o la accesibilidad—, en el mercado tokenizado el rendimiento permanece prácticamente invariable ante cambios en estas variables.

Este patrón sugiere que el yield en los activos tokenizados no emerge de la interacción entre oferta y demanda ni de las características del activo subyacente, sino que está determinado de forma exógena por la estructura financiera del producto.

5.6 ANÁLISIS POR DECILES

El análisis del 10% superior permite evaluar si las diferencias estructurales entre ambos mercados se mantienen en los activos más eficientes.

En el mercado inmobiliario tradicional, la Tabla.3 muestra un claro incremento en los niveles de precio, tanto en términos absolutos como por metro cuadrado, acompañado de una elevada dispersión. Este comportamiento confirma la existencia de un segmento premium diferenciado, donde el valor sigue dependiendo de factores como la localización y características del activo.

Tabla 3 Estadística descriptiva del decil superior (10%) – Mercado inmobiliario tradicional

Variable	Cou nt	Mean	Std	Min	25%	Media n	75%	Max
Sale Price	668	2,954,196.36	2,205,296.63	1,340,000	1,675,000	2,225,000	3,386,400	25,750,000
price_m2	668	2,946.38	2,243.85	639.27	1,661.09	2,225.61	3,322.84	27,717.98
num_rooms	668	3.02	0.16	3.00	3.00	3.00	3.00	6.00
num_bathrooms	668	2.06	0.82	1.00	1.00	2.00	3.00	4.00
carpet_area	668	1,009.11	85.99	900.00	953.00	1,001.50	1,058.25	2,190.00

Por el contrario, en el mercado tokenizado, la Tabla 4 refleja una fuerte homogeneidad en el rendimiento. El yield se mantiene concentrado en torno al 12–13%, con una variabilidad prácticamente nula, incluso entre los activos más rentables. Esta estabilidad se mantiene independientemente de variables como el tamaño del activo, la liquidez o la accesibilidad.

Tabla 4 Estadística descriptiva del decil superior (10%) – Mercado inmobiliario Tokenizado

Variabl e	Cou nt	Mean	Std	Min	25%	Median	75%	Max
price_tot al	18	209,484 .39	184,740 .23	61,896 .00	74,886 .50	121,787 .50	378,081 .25	614,992 .00
price_m 2	18	585.30	148.83	307.85	485.92	602.51	682.69	937.05
yield_re al	18	0.12	0.00	0.12	0.12	0.12	0.13	0.13
Transfer s	18	15.28	47.05	0.00	0.25	1.00	7.00	202.00
Walk_Sc ore	18	54.33	15.48	19.00	45.75	51.00	69.25	77.00

En conjunto, mientras que en el mercado tradicional el segmento superior amplifica la diferenciación entre activos, en el mercado tokenizado esta diferenciación desaparece en términos de rentabilidad. Este resultado refuerza la hipótesis de que el rendimiento en los activos tokenizados está condicionado principalmente por su diseño financiero, más que por los fundamentos del activo subyacente.

Capítulo 6. COMPARACIÓN METODOLÓGICA ENTRE MERCADOS

Los resultados obtenidos en el análisis exploratorio sugieren la existencia de diferencias estructurales en los mecanismos de formación de valor entre ambos mercados. Este capítulo tiene como objetivo validar estas evidencias mediante un análisis metodológico comparativo.

6.1 COMPARACIÓN DE ESTRUCTURA Y DISPERSIÓN

El análisis de dispersión permite evaluar la heterogeneidad en la formación de valor en ambos mercados. Como se observa en la Figura 14, existe una diferencia estructural significativa entre el mercado inmobiliario tradicional y el mercado tokenizado.

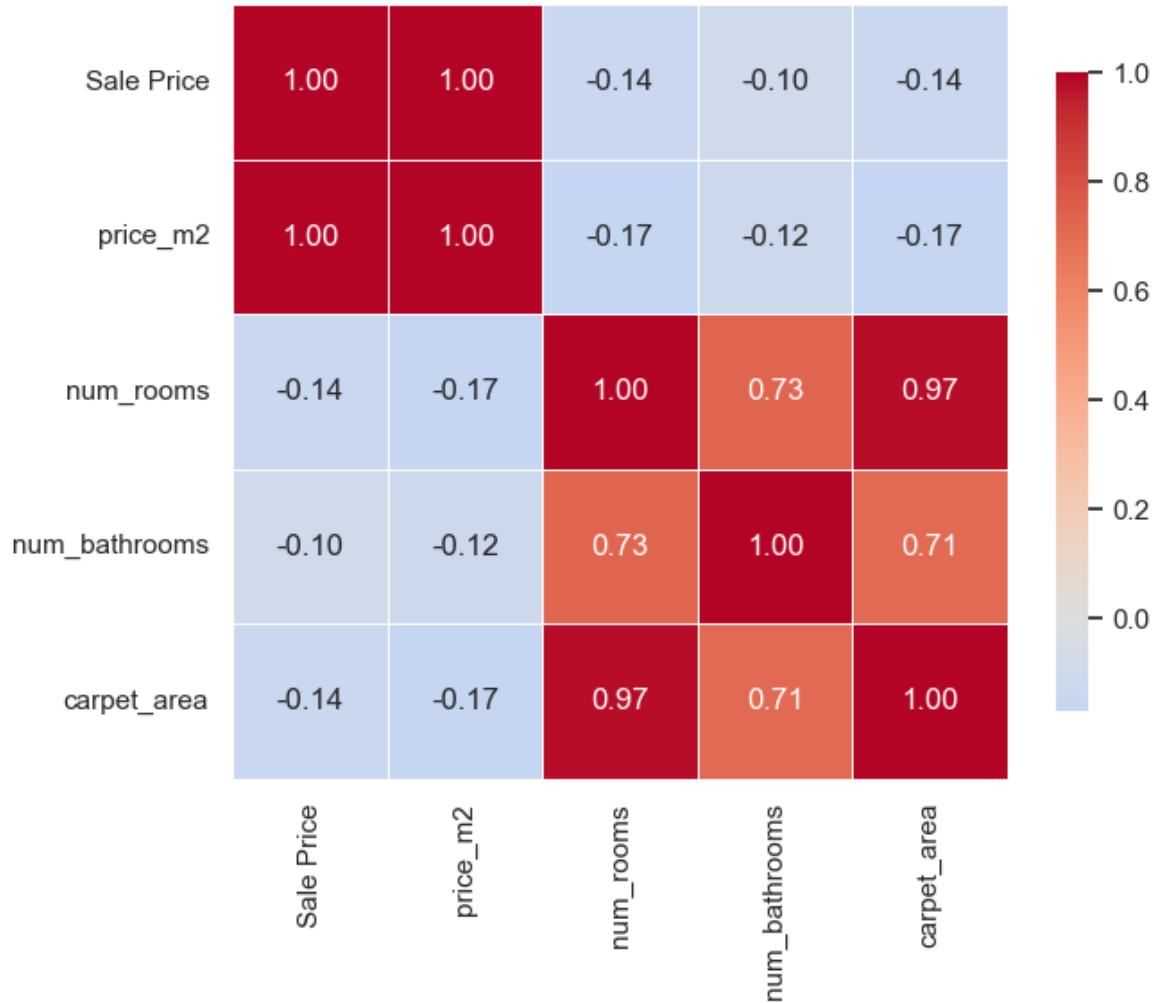
El precio por metro cuadrado en el mercado tradicional presenta una elevada dispersión, con una gran cantidad de valores extremos y una amplitud considerable en la distribución. Este comportamiento es consistente con un mercado donde el valor depende de múltiples factores —localización, características del activo y entorno— y donde existe una clara segmentación entre activos.

Por el contrario, el rendimiento (yield) en el mercado tokenizado aparece fuertemente comprimido, concentrándose en un rango muy estrecho y con una variabilidad mínima. Esta diferencia visual refuerza la idea de que ambos mercados operan bajo lógicas distintas: uno basado en la heterogeneidad y otro en la estandarización del retorno

número de habitaciones y la superficie (≈ 0.97), así como entre el número de baños y otras variables estructurales (≈ 0.71 – 0.73), lo que refleja una coherencia interna en las características físicas de los inmuebles.

Por el contrario, las correlaciones entre el precio (tanto absoluto como por metro cuadrado) y estas variables estructurales son relativamente reducidas. Este resultado sugiere que el valor del activo no depende de forma lineal de una única característica, sino de la interacción de múltiples factores, entre los que destacan variables espaciales y cualitativas no completamente capturadas en esta matriz. En este sentido, los resultados son consistentes con lo observado en la Sección 5.5, donde el análisis de dispersión evidenciaba relaciones no lineales y una elevada heterogeneidad en la formación de precios.

Figura 16 Matriz de correlaciones – Mercado tradicional

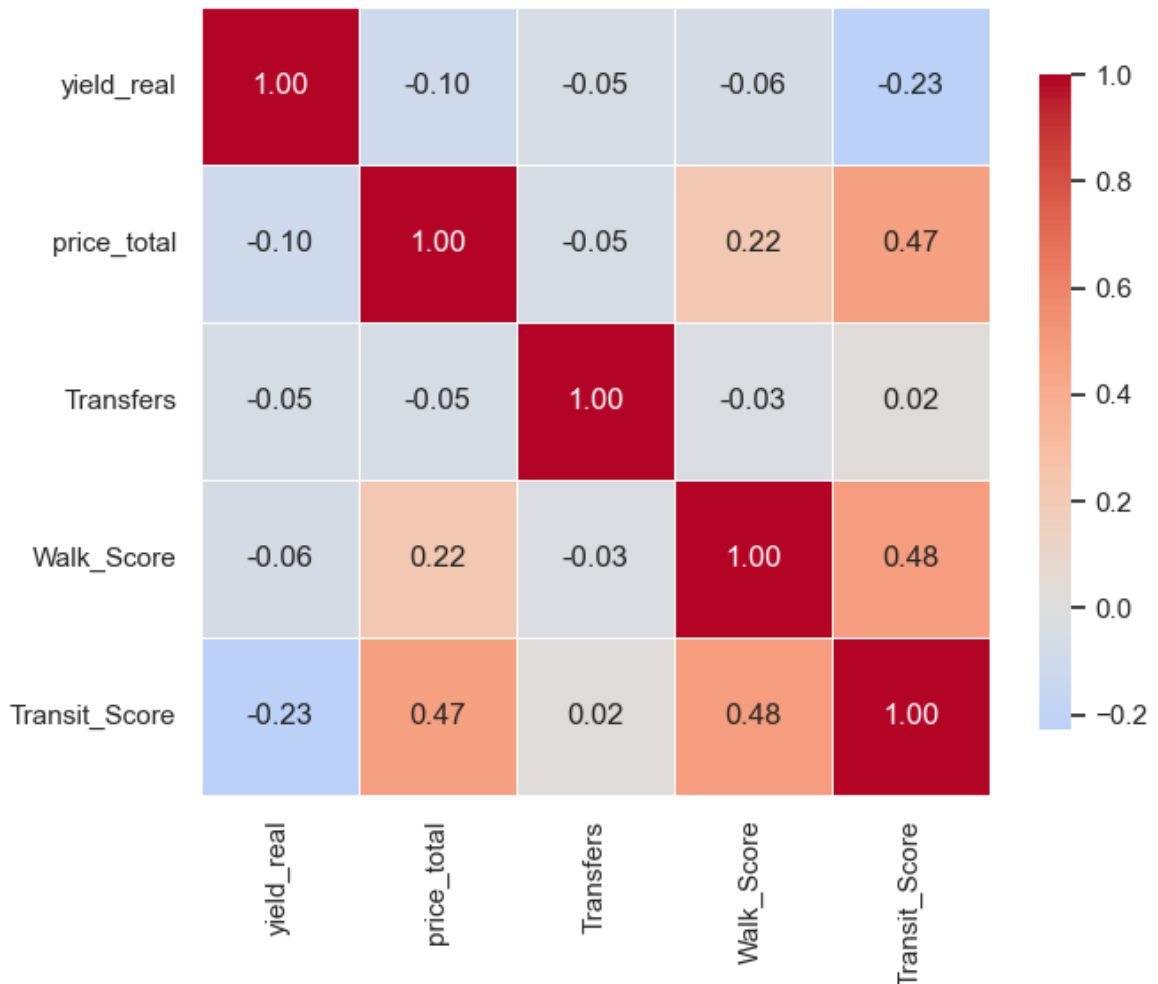


En contraste, la Figura 16 presenta la matriz de correlaciones del mercado inmobiliario tokenizado, donde el patrón observado es significativamente distinto. El rendimiento (yield_real) muestra correlaciones muy reducidas con el resto de variables: inversión total (≈ -0.10), liquidez (Transfers ≈ -0.05), accesibilidad (Walk Score ≈ -0.06) y conectividad (Transit Score ≈ -0.23).

Esta ausencia de relaciones significativas confirma los resultados obtenidos en la Sección 5.5, donde las representaciones gráficas mostraban nubes de puntos con pendientes prácticamente nulas. En conjunto, ambos enfoques convergen en la misma conclusión: la rentabilidad de los activos tokenizados es prácticamente independiente de sus características estructurales, espaciales y de mercado.

Por otro lado, sí se observan correlaciones moderadas entre variables relacionadas con el valor del activo subyacente, como la relación entre precio total y accesibilidad o conectividad ($\approx 0.22-0.48$). Esto sugiere que, aunque el valor del inmueble sigue respondiendo parcialmente a fundamentos tradicionales, dicha relación no se traslada al rendimiento ofrecido al inversor.

Figura 17 Matriz de correlaciones – Mercado tokenizado



En conjunto, la comparación entre ambas matrices permite identificar una diferencia estructural fundamental. Mientras que en el mercado tradicional las variables presentan relaciones coherentes que reflejan la interacción entre características físicas, localización y condiciones del entorno, en el mercado tokenizado el rendimiento aparece completamente desacoplado de dichas variables.

Este resultado refuerza la hipótesis central del trabajo: la tokenización inmobiliaria introduce un cambio en el mecanismo de formación de valor. En el mercado tradicional, el valor es endógeno y emerge de los fundamentos del activo; en el mercado tokenizado, la rentabilidad presenta un carácter exógeno, determinado principalmente por la arquitectura financiera del producto. En consecuencia, los activos tokenizados se comportan, desde la perspectiva del inversor, más como instrumentos financieros estructurados que como activos inmobiliarios tradicionales.

6.3 ANÁLISIS EXPLICATIVO

La Figura 6.3 compara la capacidad explicativa de los modelos lineales en ambos mercados mediante el coeficiente de determinación (R^2), permitiendo evaluar hasta qué punto las variables observables son capaces de explicar la variabilidad del precio o del rendimiento.

En el mercado tradicional, el modelo alcanza un R^2 aproximado de 0,54, lo que indica que más de la mitad de la variabilidad del precio puede explicarse a partir de variables estructurales como el tamaño, la localización o las características del inmueble. Este resultado es consistente con la teoría económica clásica, donde el valor de los activos inmobiliarios se fundamenta en atributos observables y medibles. En este sentido, el modelo no solo presenta un buen ajuste, sino que refleja una estructura de mercado coherente y relativamente predecible.

Por el contrario, en el mercado tokenizado, el modelo presenta un R^2 cercano a 0,09, evidenciando una capacidad explicativa muy limitada. Este resultado implica que las variables incluidas en el modelo apenas capturan la variabilidad del rendimiento, lo que sugiere que el proceso de formación de valor no está determinado por los fundamentos analizados. En términos prácticos, el rendimiento de estos activos parece depender en mayor medida de factores externos, de diseño financiero o de elementos no observables dentro del dataset.

Desde el punto de vista metodológico, el coeficiente de determinación (R^2) se ha obtenido a partir de modelos de regresión lineal ordinaria (OLS). Este indicador mide la proporción de

la varianza de la variable dependiente que es explicada por las variables independientes incluidas en el modelo. Formalmente, se define como:

Ecuación 1 Fórmula del coeficiente de determinación

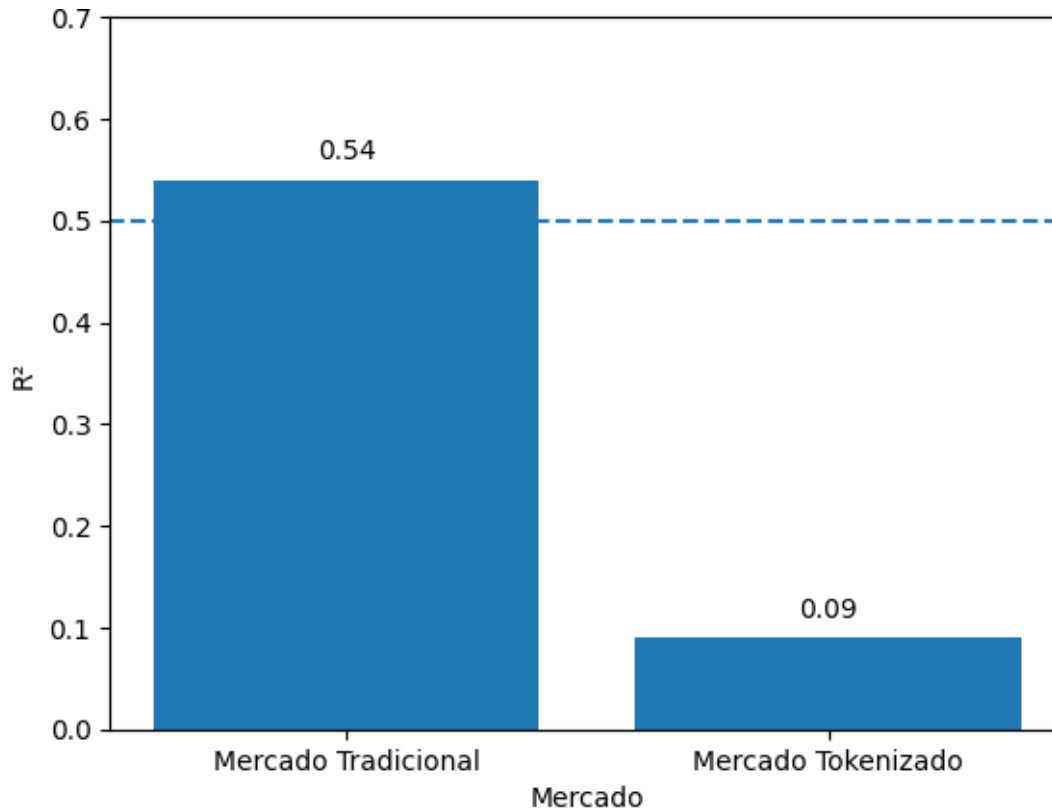
$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

donde SS_{res} representa la suma de los residuos al cuadrado (errores del modelo) y SS_{tot} la varianza total de la variable dependiente. Un valor de R^2 cercano a 1 indica un ajuste elevado, mientras que valores próximos a 0 reflejan una baja capacidad explicativa.

En este trabajo, el cálculo se ha realizado ajustando modelos lineales independientes para cada mercado, utilizando como variable dependiente el precio por metro cuadrado en el caso tradicional y el rendimiento (yield) en el mercado tokenizado. Las variables explicativas incluyen características estructurales y de contexto, previamente seleccionadas en el análisis exploratorio.

Tal y como se observa en la Figura 6.3, la diferencia en términos de R^2 es significativa, evidenciando la distinta capacidad explicativa de los modelos en cada mercado.

Figura 18 Capacidad explicativa de los modelos



6.4 ANÁLISIS POR SEGMENTACIÓN

El análisis de *clustering* permite identificar la existencia de estructuras internas en los datos, es decir, si los activos pueden agruparse en conjuntos homogéneos con características diferenciadas. A diferencia de los modelos supervisados, este enfoque no impone relaciones a priori entre variables, sino que permite observar si los patrones emergen de manera natural a partir de los datos.

La motivación principal de este análisis es complementar el enfoque explicativo anterior. Mientras que la regresión evalúa si el valor puede explicarse mediante relaciones entre variables, el *clustering* permite analizar si existen estructuras internas independientes del

modelo, es decir, si el mercado presenta segmentación natural en perfiles diferenciados de activos.

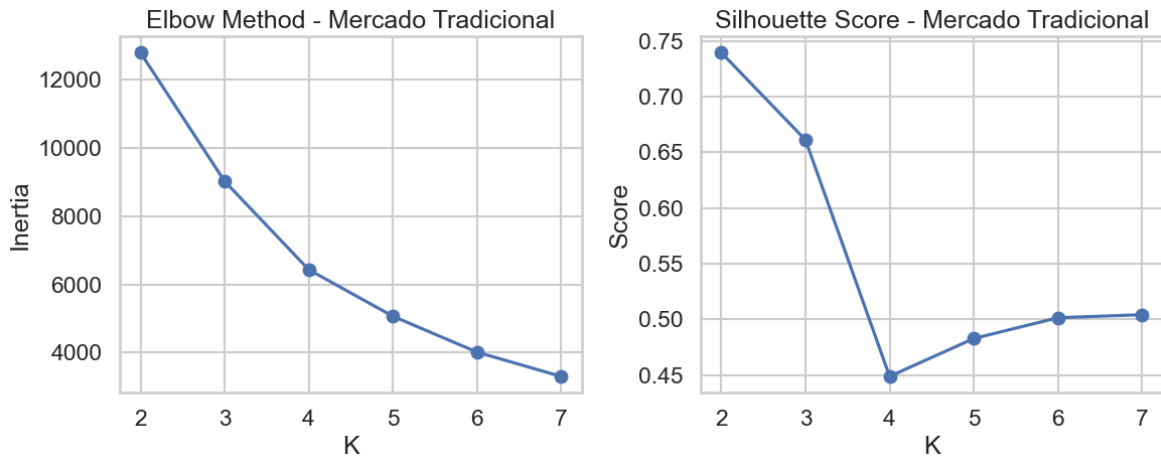
Desde el punto de vista metodológico, el análisis de segmentación se ha realizado exclusivamente mediante el algoritmo K-Means, priorizando la interpretabilidad y la comparabilidad entre ambos mercados. Esta elección responde a su capacidad para identificar particiones claras en espacios de variables continuas, así como a su facilidad de interpretación económica, permitiendo asociar los clusters a perfiles diferenciados de activos.

Las variables han sido previamente normalizadas mediante *StandardScaler*, con el objetivo de evitar sesgos derivados de diferencias de escala y garantizar una contribución equilibrada de cada dimensión en el proceso de agrupación.

Para evaluar la calidad de la segmentación se han utilizado dos enfoques complementarios: el método del codo (*elbow method*) y el índice *silhouette*. Mientras que el primero permite analizar la evolución de la inercia al aumentar el número de clusters, el segundo mide simultáneamente la cohesión interna y la separación entre grupos.

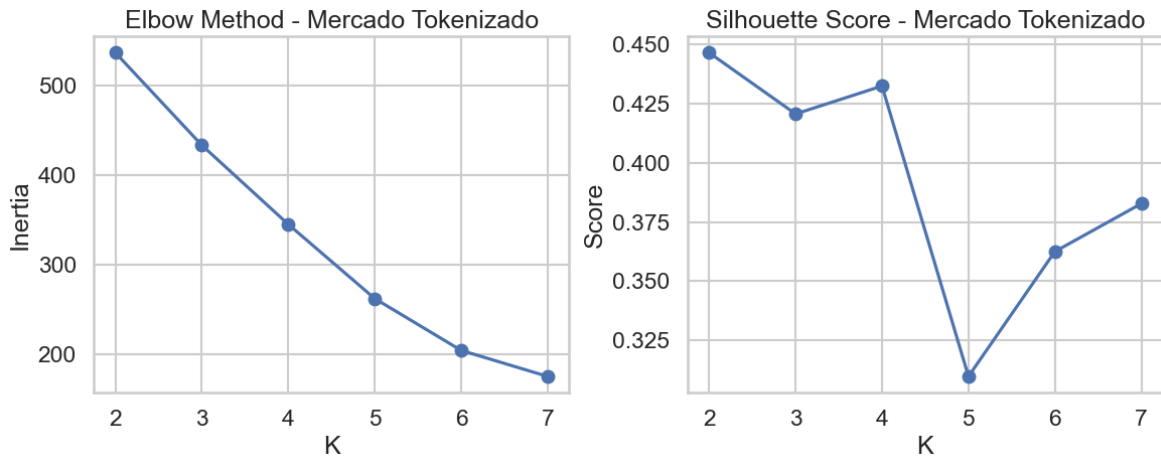
La Figura 18 muestra estos resultados para el mercado inmobiliario tradicional. En este caso, se observa una reducción significativa de la inercia a medida que aumenta el número de clusters, junto con un índice *silhouette* elevado para valores bajos de K , alcanzando valores cercanos a 0,74 para $K = 2$ y manteniéndose relativamente altos para $K = 3$. Este comportamiento sugiere la existencia de una estructura interna clara, donde los activos pueden agruparse en segmentos diferenciados con buena separación entre grupos.

Figura 19 N°de clusters mercado tradicional



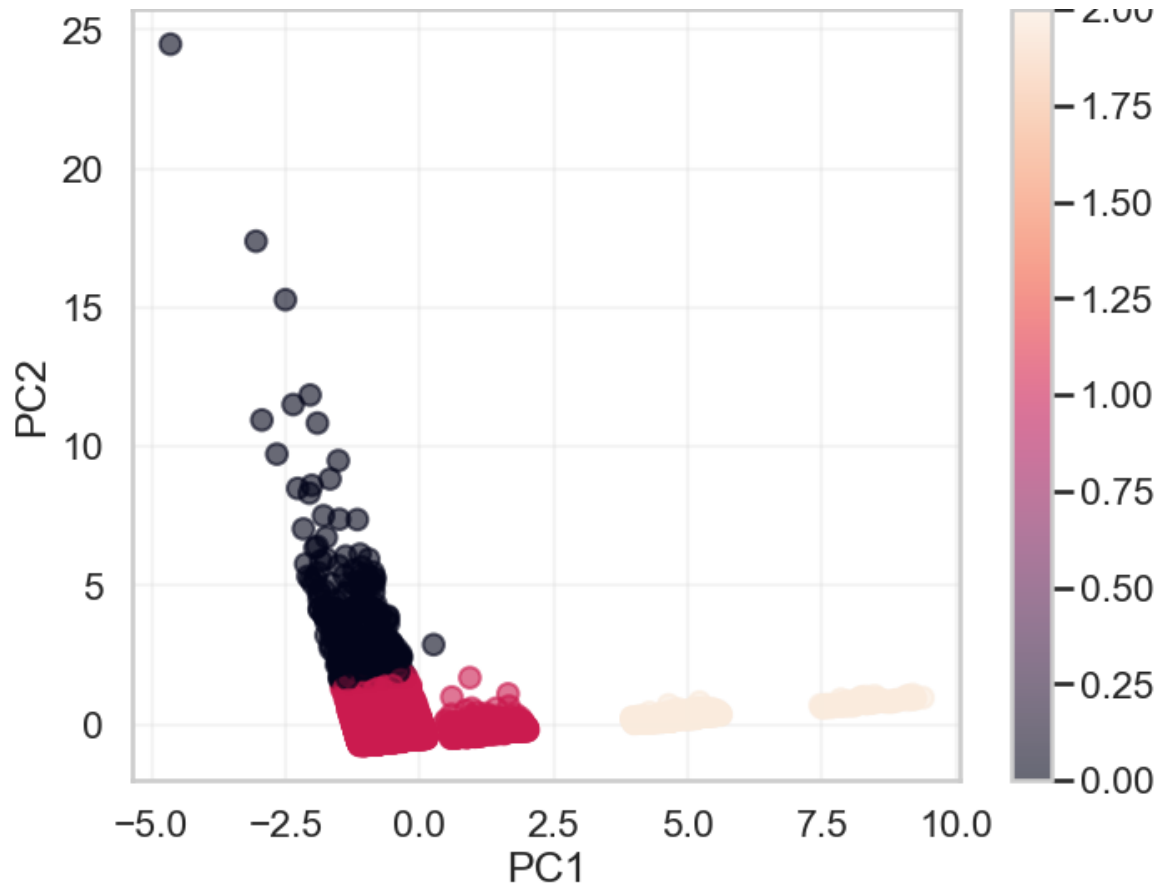
Por el contrario, la Figura 19 presenta los mismos análisis para el mercado tokenizado. En este caso, no se aprecia un punto de inflexión claramente definido en la curva del *elbow*, y los valores del índice *silhouette* son más bajos y relativamente estables (en torno a 0,42–0,45 para valores bajos de K). Este patrón indica una estructura más débil, donde la segmentación no genera grupos claramente diferenciados, sino particiones más simples y con menor capacidad de separación.

Figura 20 N°de clusters mercado tokenizado



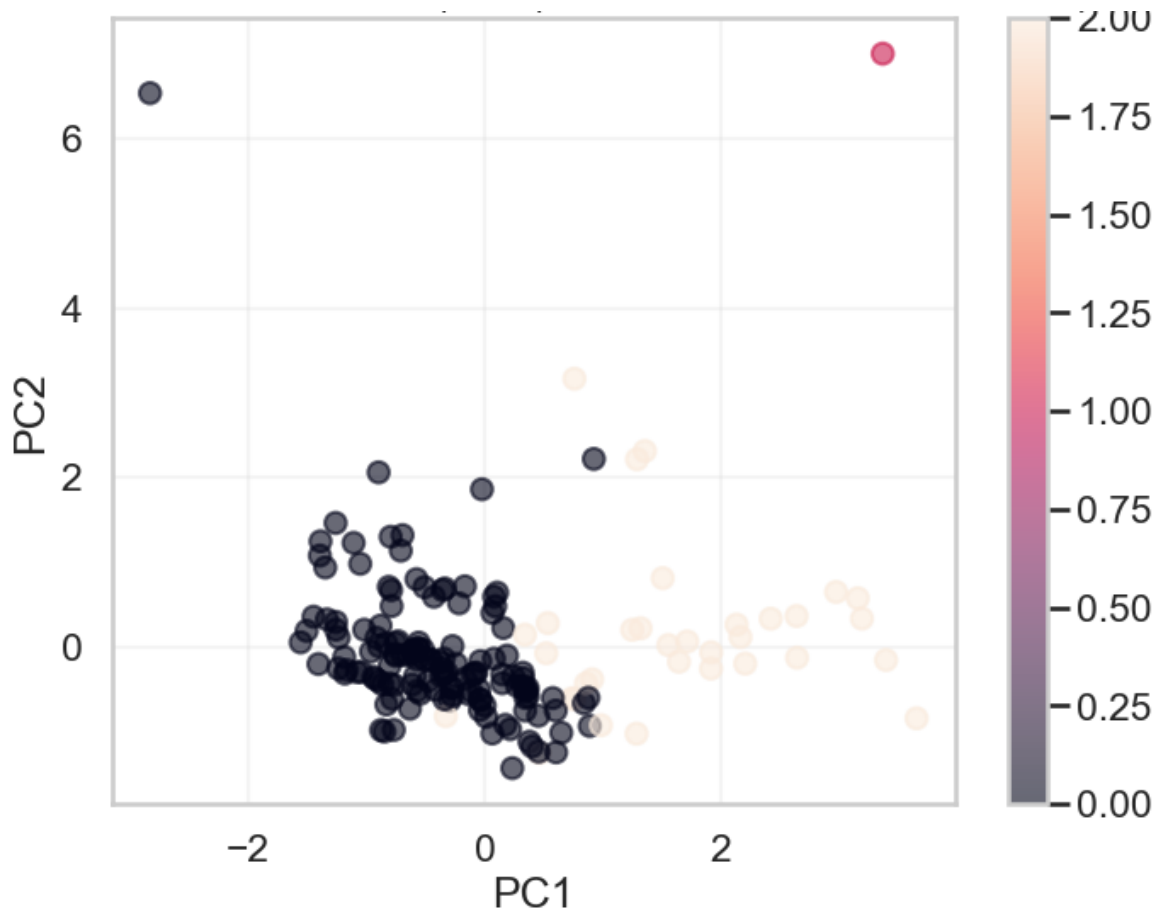
La evidencia visual se refuerza mediante la representación en componentes principales. La Figura 20 muestra el resultado del *clustering* en el mercado tradicional proyectado en dos dimensiones. En este caso, se observa una separación clara entre clusters, especialmente a lo largo de la primera componente principal, lo que indica la existencia de submercados diferenciados y una estructura interna consistente con la heterogeneidad del mercado.

Figura 21 Clusters mercado tradicional



En contraste, la Figura 21 presenta la misma representación para el mercado tokenizado. A diferencia del caso anterior, los clusters aparecen significativamente más solapados, sin fronteras claramente definidas entre grupos. Esta configuración visual sugiere una menor diferenciación entre activos y una estructura más homogénea.

Figura 22 Clusters mercado tokenizado



En conjunto, los resultados del análisis de segmentación apuntan a una diferencia estructural entre ambos mercados. Mientras que el mercado inmobiliario tradicional presenta una segmentación clara, con clusters bien definidos y separados, el mercado tokenizado muestra una estructura más simple, donde la agrupación de activos no permite identificar submercados diferenciados con la misma claridad.

Estos resultados son coherentes con la evidencia obtenida en apartados anteriores. La mayor dispersión, la dependencia de variables fundamentales y la mayor capacidad explicativa en el mercado tradicional se traducen en una estructura interna más rica y segmentada. Por el

contrario, la compresión del rendimiento y la menor relación con variables observables en el mercado tokenizado se reflejan en una segmentación más limitada y en un mayor grado de solapamiento entre activos.

En este sentido, la evidencia sugiere que la tokenización no reproduce plenamente la lógica de segmentación del mercado inmobiliario tradicional, sino que tiende hacia una estructura más homogénea, donde las diferencias entre activos tienen un menor impacto en su comportamiento económico.

Capítulo 7. DISCUSIÓN E IMPLICACIONES ESTRUCTURALES

Los resultados obtenidos a lo largo del análisis empírico permiten plantear una interpretación estructural sobre la naturaleza económica del mercado inmobiliario tokenizado y su relación con el mercado tradicional. Lejos de constituir una mera digitalización del activo inmobiliario, la evidencia sugiere que la tokenización introduce una lógica de funcionamiento distinta, tanto en la formación del valor como en la determinación de la rentabilidad.

En el mercado inmobiliario tradicional, los resultados muestran un comportamiento consistente con la teoría económica. El precio del activo presenta una elevada dispersión, una segmentación clara y una fuerte dependencia de variables observables, especialmente aquellas relacionadas con la localización y la accesibilidad. La capacidad explicativa de los modelos ($R^2 \approx 0,54$), junto con la coherencia observada en el análisis de correlaciones y segmentación, refuerza la idea de que el valor emerge de la interacción entre oferta y demanda en un entorno donde los fundamentos urbanos desempeñan un papel central.

Por el contrario, el mercado tokenizado presenta un patrón significativamente diferente. La evidencia empírica revela una fuerte compresión del rendimiento, concentrado en torno a valores cercanos al 11–12%, una ausencia sistemática de correlaciones relevantes con variables espaciales y una capacidad explicativa muy limitada de los modelos aplicados ($R^2 \approx 0,09$).

Asimismo, el análisis muestra que el rendimiento efectivo reproduce prácticamente de forma exacta el rendimiento esperado anunciado por la plataforma. Esta coincidencia no es trivial: indica que el rendimiento observado no resulta de la interacción entre oferta y demanda en el mercado, sino que está estrechamente vinculado a los parámetros definidos en el momento de estructuración del activo. En otras palabras, el *yield* no se ajusta dinámicamente en función de características como la localización, el tamaño o la liquidez, sino que permanece anclado a una rentabilidad previamente establecida.

Este comportamiento puede explicarse por la propia arquitectura financiera de los activos tokenizados. El rendimiento ofrecido al inversor se deriva directamente de los flujos de caja estimados (rentas netas) en relación con el precio de emisión del activo, ambos definidos ex ante por la plataforma. En consecuencia, la variabilidad del *yield* queda limitada, ya que depende más de decisiones de diseño —como la estimación de ingresos, los costes operativos o los márgenes incorporados— que de la heterogeneidad del activo subyacente.

En conjunto, estos resultados sugieren que el *yield* no se comporta como una variable plenamente emergente del mercado, sino como un parámetro altamente estandarizado, cuya estabilidad responde a la forma en que el producto financiero está estructurado.

Desde una perspectiva económica, este hallazgo implica una transformación en el mecanismo de formación de valor. Mientras que en el mercado tradicional el precio refleja información agregada sobre el activo y su entorno, en el mercado tokenizado la rentabilidad parece estar condicionada en mayor medida por la arquitectura financiera del producto. Elementos como la estructuración del flujo de rentas, los costes incorporados, las reservas iniciales o las condiciones de emisión desempeñan un papel central en la fijación del rendimiento ofrecido al inversor.

Esta diferencia introduce una separación conceptual relevante entre ambos mercados. En el entorno tradicional, el valor es fundamentalmente endógeno, resultado de dinámicas descentralizadas y competitivas. En el entorno tokenizado, en cambio, el rendimiento presenta un componente más exógeno, condicionado por decisiones de diseño y estructuración previas a la interacción con el mercado. Esta transición aproxima los activos tokenizados a la lógica de instrumentos financieros estructurados, donde el retorno está parcialmente definido *ex ante*, más que determinado por la evolución del mercado subyacente.

Las implicaciones de este fenómeno son relevantes para distintos agentes. Para los inversores, la estandarización del rendimiento puede traducirse en una mayor previsibilidad y facilidad de comparación entre activos. Sin embargo, esta misma homogeneidad puede limitar la capacidad de discriminar oportunidades de inversión basadas en la calidad intrínseca del inmueble o su potencial de revalorización. En este sentido, la toma de decisiones podría desplazarse desde el análisis del activo subyacente hacia la evaluación del diseño del producto financiero, alterando los criterios tradicionales de inversión inmobiliaria.

Para los reguladores, los resultados plantean la necesidad de analizar con mayor profundidad la naturaleza de la información ofrecida al inversor, especialmente en lo relativo a la formación del rendimiento y su relación con el activo subyacente. La aparente estabilidad del retorno puede ocultar riesgos asociados a la estructura del producto, lo que exige mecanismos adecuados de transparencia y supervisión. En este contexto, la tokenización no elimina la complejidad del mercado inmobiliario, sino que la reconfigura bajo nuevas estructuras contractuales y tecnológicas.

Desde el punto de vista académico, los hallazgos de este trabajo sugieren que la tokenización inmobiliaria no debe analizarse únicamente como una innovación tecnológica, sino como

una transformación institucional del mercado. La evidencia muestra que no se trata simplemente de mejorar la liquidez o el acceso, sino de alterar el propio proceso mediante el cual se genera y se percibe el valor del activo.

En conjunto, los resultados refuerzan la hipótesis central del trabajo: el mercado inmobiliario tokenizado opera bajo una lógica distinta a la del mercado tradicional. Mientras que este último se caracteriza por heterogeneidad, segmentación y dependencia de fundamentos, el primero muestra una estructura más homogénea, una relación más débil con variables observables y una mayor influencia del diseño financiero en la determinación del rendimiento.

En este sentido, la evidencia empírica sugiere que la tokenización no convierte el inmobiliario en un activo más líquido o accesible en su forma tradicional, sino que lo transforma en un producto financiero con características propias. El inmueble deja de ser el elemento central en la determinación del retorno, pasando a ocupar ese papel la estructura del instrumento que lo representa.

La tokenización no digitaliza el mercado inmobiliario: lo redefine como un producto financiero.

Capítulo 8. FUTURAS LÍNEAS DE INVESTIGACIÓN

El presente trabajo ha permitido avanzar en la comprensión empírica de la tokenización inmobiliaria desde una perspectiva comparativa. Sin embargo, la naturaleza emergente del fenómeno y las limitaciones en la disponibilidad de datos abren diversas líneas de investigación futuras que pueden ampliar y matizar los resultados obtenidos.

En primer lugar, resulta fundamental incorporar una dimensión temporal más robusta. El análisis actual se basa en datos transversales, lo que limita el estudio de la evolución del rendimiento (yield) a lo largo del tiempo. La construcción de series temporales permitiría evaluar la estabilidad del rendimiento y su comportamiento en distintos ciclos de mercado, determinando si la compresión observada es estructural o transitoria.

En segundo lugar, sería relevante ampliar el análisis a múltiples plataformas de tokenización. El uso de una única fuente puede introducir sesgos derivados del diseño financiero o la selección de activos. La comparación entre plataformas permitiría evaluar si los patrones identificados responden a una lógica general del mercado o a características específicas de cada emisor.

Asimismo, la incorporación de variables macroeconómicas constituye una extensión clave. Factores como los tipos de interés o la inflación han demostrado ser determinantes en el mercado inmobiliario tradicional. Analizar su impacto en los activos tokenizados permitiría evaluar si estos mantienen una dinámica independiente o si convergen hacia patrones tradicionales a medida que el mercado madura.

Por otro lado, futuras investigaciones podrían avanzar hacia enfoques causales. Frente al carácter descriptivo del presente trabajo, el uso de metodologías como variables instrumentales o técnicas de matching permitiría identificar relaciones más profundas en la formación de valor.

Adicionalmente, resulta relevante profundizar en la microestructura del mercado secundario. El análisis de variables como la liquidez efectiva, la frecuencia de transacciones o los diferenciales de precios permitiría evaluar si la liquidez prometida por la tokenización se materializa en la práctica.

Otra línea de interés se centra en el diseño financiero de los activos tokenizados. Un análisis más detallado de la estructura de costes, flujos y condiciones contractuales permitiría

comprender mejor cómo se determina el rendimiento y qué factores explican su homogeneidad.

Finalmente, desde una perspectiva regulatoria, será clave analizar el impacto de marcos como MiCA o el régimen piloto DLT en el desarrollo del mercado. La evolución normativa puede condicionar tanto la estructura de los activos como su accesibilidad y transparencia.

En conjunto, estas líneas reflejan que la tokenización inmobiliaria es un fenómeno aún en desarrollo, cuyo estudio requerirá datos más amplios y enfoques metodológicos más avanzados para comprender plenamente su funcionamiento y su papel en el sistema financiero.

BIBLIOGRAFÍA

La bibliografía empleada en este Trabajo Fin de Grado incluye referencias académicas, informes institucionales y normativa aplicable que fundamentan el análisis del mercado inmobiliario tradicional, la tokenización de activos y su encaje regulatorio. Estas fuentes han permitido construir un marco teórico sólido y contextualizar los resultados empíricos obtenidos.

Aste, T., Tasca, P., & Di Matteo, T. (2017). *Blockchain technologies: The foreseeable impact on society and industry*. *Computer*, 50(9), 18–28.

Case, K. E., & Shiller, R. J. (1989). *The efficiency of the market for single-family homes*. *American Economic Review*, 79(1), 125–137.

Catalini, C., & Gans, J. S. (2016). *Some simple economics of the blockchain*. MIT Sloan Research Paper No. 5191-16.

Comisión Nacional del Mercado de Valores (CNMV). (2020–2024). *Comunicaciones y criterios sobre criptoactivos y valores negociables*.

Cong, L. W., & He, Z. (2019). *Blockchain disruption and smart contracts*. *Review of Financial Studies*, 32(5), 1754–1797.

Deloitte. (2023). *Tokenization of real estate assets: Opportunities and challenges*.

Directiva 2014/65/UE del Parlamento Europeo y del Consejo. (2014). *Relativa a los mercados de instrumentos financieros (MiFID II)*.

Geltner, D., Miller, N. G., Clayton, J., & Eichholtz, P. (2014). *Commercial Real Estate Analysis and Investments* (3rd ed.). South-Western Cengage Learning.

Ling, D. C., & Archer, W. R. (2010). *Real Estate Principles: A Value Approach*. McGraw-Hill.

Pagliari, J. L. (2017). *Another look at real estate returns*. *Journal of Portfolio Management*, 43(6), 60–71.

Reglamento (UE) 2022/858 del Parlamento Europeo y del Consejo. (2022). *Relativo a un régimen piloto para infraestructuras de mercado basadas en tecnología de registro distribuido (DLT)*.

Reglamento (UE) 2023/1114 del Parlamento Europeo y del Consejo. (2023). *Relativo a los mercados de criptoactivos (MiCA)*.

Schär, F. (2021). *Decentralized finance: On blockchain- and smart contract-based financial markets*. Federal Reserve Bank of St. Louis Review, 103(2), 153–174.

U.S. Securities and Exchange Commission (SEC). (2019). *Framework for “Investment Contract” Analysis of Digital Assets*.

Yermack, D. (2017). *Corporate governance and blockchains*. Review of Finance, 21(1), 7–31.

Zetsche, D. A., Buckley, R. P., Arner, D. W., & Föhr, L. (2018). *The ICO gold rush: It’s a scam, it’s a bubble, it’s a super challenge for regulators*. Harvard International Law Journal, 60(2), 267–307.

.

BIBLIOTECAS DE PYTHON UTILIZADAS

El desarrollo empírico del trabajo se ha llevado a cabo íntegramente en Python, utilizando un conjunto de bibliotecas especializadas que han permitido implementar un pipeline analítico reproducible, desde la carga y preprocesamiento de los datos hasta la modelización y visualización de resultados.

La manipulación y estructuración de los datos se ha realizado mediante Pandas, que ha permitido gestionar datasets en formato CSV y Excel, tratar valores faltantes y construir variables derivadas clave como el precio por metro cuadrado o el rendimiento real anual (*yield_real*). Complementariamente, NumPy ha proporcionado soporte para operaciones numéricas vectorizadas y manipulación eficiente de arrays.

En la fase de modelización, se ha empleado Scikit-learn, utilizando modelos de regresión lineal y algoritmos basados en árboles como Random Forest para evaluar la capacidad explicativa de las variables. Asimismo, se han aplicado técnicas de aprendizaje no supervisado, como K-Means, para analizar la existencia de estructuras internas en los datos. La calidad de estas segmentaciones se ha evaluado mediante métricas como el *silhouette score*.

El análisis estadístico se ha apoyado en SciPy, permitiendo la realización de contrastes de hipótesis, incluyendo pruebas t, test de Levene y análisis ANOVA.

Finalmente, la visualización de resultados se ha desarrollado mediante Matplotlib y Seaborn, herramientas que han permitido representar distribuciones, correlaciones y agrupaciones, facilitando tanto la interpretación como la comunicación de los resultados.

En conjunto, este ecosistema de herramientas ha permitido desarrollar un análisis riguroso, coherente y replicable, adaptado a la naturaleza comparativa del presente trabajo.

ANEXO I DECLARACIÓN UTILIZACIÓN IA GENERATIVA

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, Alberto Martín Colino, estudiante de GITT+BA de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado “Tokenización de Inmuebles: Análisis de Datos y Perspectivas Financieras” 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:

8. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto.
9. Sintetizador y divulgador de libros complicados: Para resumir y comprender literatura compleja.
10. Revisor: Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
11. 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: 15 de abril de 2026

Firma: Alberto Martín Colino

ANEXO II IMPLEMENTACIÓN DEL PIPELINE

ANALÍTICO EN PYTHON

1. DESCRIPCIÓN GENERAL DEL PROYECTO

El presente anexo recoge una selección representativa del código Python desarrollado a lo largo del trabajo, organizado de forma coherente con las fases del pipeline analítico descrito en los capítulos 2, 5 y 6 de la memoria. Su propósito no es reproducir la totalidad del código empleado, sino ilustrar las decisiones técnicas fundamentales que sustentan el análisis comparativo entre el mercado inmobiliario tradicional y el mercado tokenizado.

El pipeline se estructura en cinco fases diferenciadas: (1) carga y preprocesado de datos, (2) construcción de variables derivadas, (3) análisis exploratorio y distribuciones, (4) modelización estadística y regresión, y (5) segmentación mediante clustering. Cada una de estas fases está implementada como un módulo funcional independiente, lo que garantiza la trazabilidad y reproducibilidad del análisis.

El entorno de trabajo es Python 3.8 gestionado con Anaconda. Las bibliotecas principales empleadas son pandas y numpy para el tratamiento de datos, scikit-learn para la modelización, scipy para los contrastes estadísticos y matplotlib/seaborn para la visualización.

2. CARGA Y PREPROCESADO DE LOS DATOS

El primer paso del pipeline consiste en la carga de los tres datasets utilizados en el estudio: el dataset de mercado inmobiliario tradicional, el dataset urbano (MRT) y el dataset de activos tokenizados. A continuación, se aplican operaciones de limpieza básica —eliminación de duplicados y valores nulos en variables clave— y se homogeneiza la nomenclatura de columnas para facilitar el análisis comparativo posterior.

Bloque 1 — Importación de librerías y carga de datasets

```
import pandas as pd

import numpy as np

from sklearn.preprocessing import StandardScaler

from sklearn.linear_model import LinearRegression

from sklearn.cluster import KMeans

from sklearn.model_selection import cross_val_score

from sklearn.metrics import silhouette_score

from sklearn.ensemble import RandomForestRegressor

from scipy import stats

import matplotlib.pyplot as plt

import seaborn as sns

# — Carga de los tres datasets —————

df_trad = pd.read_csv('data/mercado_tradicional.csv')

df_urban = pd.read_csv('data/dataset_mrt.csv')

df_token = pd.read_csv('data/mercado_tokenizado.csv')
```

Bloque 2 — Limpieza y estandarización de columnas

```
def limpiar_dataset(df, var_clave, nombre='dataset'):

    # Elimina duplicados y filas sin valor en la variable principal

    df = df.drop_duplicates()

    df = df.dropna(subset=var_clave)

    df.columns = df.columns.str.strip().str.lower().str.replace(' ', '_')

    print(f'[{nombre}] {len(df)} observaciones válidas tras limpieza.')

    return df
```

```
df_trad = limpiar_dataset(df_trad, ['price'], 'Tradicional')
df_urban = limpiar_dataset(df_urban, ['price_per_m2'], 'MRT-Urbano')
df_token = limpiar_dataset(df_token, ['annual_yield'], 'Tokenizado')
```

3. CONSTRUCCIÓN DE VARIABLES DERIVADAS

Una vez saneados los datos, se construyen las variables derivadas que sirven como métricas comparables entre ambos mercados. En el mercado tradicional, la variable central es el precio por metro cuadrado (`price_per_m2`), obtenida dividiendo el precio total entre la superficie. En el mercado tokenizado, se calcula el `yield_real` a partir del rendimiento anual bruto declarado, ajustado por la inversión mínima y el número de tokens en circulación. Esta homogeneización es imprescindible para que los modelos de regresión y los análisis de correlación sean comparables entre sí.

Bloque 3 — Variables derivadas: mercado tradicional

```
# Precio por metro cuadrado y por habitación

df_trad['price_per_m2'] = df_trad['price'] / df_trad['area_m2']

df_trad['price_per_room'] = df_trad['price'] /
df_trad['num_rooms'].replace(0, np.nan)

# Segmentación por deciles de precio

df_trad['decil_preco'] = pd.qcut(df_trad['price_per_m2'], q=10,
labels=False)

top_decil_trad = df_trad[df_trad['decil_preco'] == 9]
```

Bloque 4 — Variables derivadas: mercado tokenizado

```
# Yield real ajustado por inversión mínima

df_token['yield_real'] = (

    df_token['annual_yield'] / df_token['min_investment']

) * 100 # Expresado en porcentaje

# Variable de liquidez: número de transferencias normalizado

df_token['liquidez_norm'] = (

    df_token['num_transfers'] - df_token['num_transfers'].mean()

) / df_token['num_transfers'].std()
```

4. ANÁLISIS EXPLORATORIO DE DATOS (EDA)

El análisis exploratorio se realiza de forma paralela sobre ambos mercados, permitiendo identificar distribuciones, outliers y patrones preliminares. Se generan histogramas con estimación de densidad kernel (KDE), boxplots por deciles y matrices de correlación. Estas representaciones no solo sirven para caracterizar los datos, sino también para validar los supuestos que subyacen a los modelos estadísticos posteriores.

Bloque 5 — Distribuciones y estadísticos descriptivos

```
def resumen_estadistico(df, variables, nombre='mercado'):

    # Tabla de estadísticos: media, mediana, std, p10, p90

    stats_df = df[variables].agg(['mean', 'median', 'std',

                                  lambda x: x.quantile(0.1),

                                  lambda x: x.quantile(0.9)])

    stats_df.index = ['Media', 'Mediana', 'Desv. típ.', 'P10', 'P90']

    print(f'\n— Estadísticos {nombre} —')
```

```
print(stats_df.round(2).to_string())

return stats_df

resumen_estadistico(df_trad, ['price_per_m2', 'area_m2'], 'Tradicional')

resumen_estadistico(df_token, ['yield_real', 'min_investment'],
'Tokenizado')
```

Bloque 6 — Visualización: histograma con KDE y boxplot

```
# Histograma con KDE - mercado tradicional
fig, axes = plt.subplots(1, 2, figsize=(12, 4))

sns.histplot(df_trad['price_per_m2'], kde=True, ax=axes[0],
             color='steelblue', bins=40, edgecolor='white')
axes[0].set_title('Distribución precio/m² - Mercado tradicional')
axes[0].set_xlabel('Precio por m²')

# Boxplot por deciles - mercado tradicional
sns.boxplot(data=df_trad, x='decil_precio', y='price_per_m2',
            ax=axes[1], palette='Blues')
axes[1].set_title('Distribución por deciles - Mercado tradicional')

plt.tight_layout()

plt.savefig('output/fig_distribuciones.png', dpi=150, bbox_inches='tight')
```

5. MODELIZACIÓN ESTADÍSTICA Y REGRESIÓN

La fase de modelización aplica regresión lineal como modelo de referencia y Random Forest como contraste no lineal, sobre cada mercado de forma independiente. La capacidad explicativa se evalúa mediante el coeficiente de determinación R^2 en muestra y fuera de muestra (validación cruzada con $k=5$), complementado con bootstrapping para estimar la estabilidad del modelo. Asimismo, se realizan contrastes t entre los deciles extremos (D1 y D10) para evaluar si las diferencias de rendimiento entre segmentos son estadísticamente significativas.

Bloque 7 — Regresión lineal con validación cruzada

```
def regresion_con_validacion(df, features, target, nombre='modelo', k=5):  
  
    X = df[features].dropna()  
  
    y = df.loc[X.index, target]  
  
    modelo = LinearRegression()  
  
    # R2 en muestra completa  
  
    modelo.fit(X, y)  
  
    r2_train = modelo.score(X, y)  
  
    # R2 con validación cruzada (k-fold)  
  
    cv_scores = cross_val_score(modelo, X, y, cv=k, scoring='r2')  
  
    r2_cv = cv_scores.mean()  
  
    print(f'[{nombre}] R2_train={r2_train:.4f} | R2_CV={r2_cv:.4f} ±  
{cv_scores.std():.4f}')  
  
    return modelo, r2_train, r2_cv
```

```
# Mercado tradicional: predice precio/m²

regresion_con_validacion(

    df_trad,

    features=['area_m2', 'num_rooms', 'dist_mrt', 'walk_score'],

    target='price_per_m2', nombre='Tradicional'

)

# Mercado tokenizado: predice yield real

regresion_con_validacion(

    df_token,

    features=['min_investment', 'num_transfers', 'dist_centro',

'walk_score'],

    target='yield_real', nombre='Tokenizado'

)
```

Bloque 8 — Random Forest y contraste Levene entre mercados

```
# Random Forest - contraste de relaciones no lineales

rf = RandomForestRegressor(n_estimators=100, random_state=42)

X_rf = df_trad[['area_m2', 'num_rooms', 'dist_mrt', 'walk_score']].dropna()

y_rf = df_trad.loc[X_rf.index, 'price_per_m2']

rf.fit(X_rf, y_rf)

r2_rf = rf.score(X_rf, y_rf)

print(f'Random Forest R² tradicional: {r2_rf:.4f}')

# Test de Levene: compara varianzas entre mercados

levene_stat, levene_p = stats.levene(
```

```
df_trad['price_per_m2'].dropna(),  
  
df_token['yield_real'].dropna()  
  
)  
  
print(f'Levene: stat={levene_stat:.4f}, p-valor={levene_p:.4f}')  
  
if levene_p < 0.05:  
  
    print('-> Varianzas significativamente distintas entre mercados (p<0.05)')
```

6. SEGMENTACIÓN MEDIANTE CLUSTERING (KMEANS)

La fase de segmentación utiliza el algoritmo KMeans para identificar grupos estructurales dentro de cada mercado. Previamente, las variables se estandarizan con StandardScaler para evitar que diferencias de escala distorsionen la métrica de distancia. El número óptimo de clusters se determina combinando el método del codo (inerencia) con el silhouette score. La validez estadística de los clusters obtenidos se contrasta mediante ANOVA y una prueba de Tukey post-hoc cuando el resultado es significativo.

Bloque 9 — Selección del número óptimo de clusters

```
def seleccionar_k(X_scaled, k_max=10):  
  
    # Calcula inercia y silhouette para k = 2..k_max  
  
    inercias, silhouettes = [], []  
  
    ks = range(2, k_max + 1)  
  
    for k in ks:  
  
        km = KMeans(n_clusters=k, random_state=42, n_init=10)  
  
        labels = km.fit_predict(X_scaled)  
  
        inercias.append(km.inertia_)  
  
        silhouettes.append(silhouette_score(X_scaled, labels))  
  
    best_k = ks[np.argmax(silhouettes)]  
  
    print(f'k óptimo (silhouette): {best_k} - score={max(silhouettes):.4f}')
```

```
return best_k, inercias, silhouettes
```

Bloque 10 — KMeans final y validación ANOVA

```
# — Mercado tradicional —————
vars_cluster_trad = ['price_per_m2', 'area_m2', 'dist_mrt', 'walk_score']
X_trad = df_trad[vars_cluster_trad].dropna()
scaler = StandardScaler()
X_trad_sc = scaler.fit_transform(X_trad)
best_k_trad, _, _ = seleccionar_k(X_trad_sc)

km_trad = KMeans(n_clusters=best_k_trad, random_state=42, n_init=10)
df_trad.loc[X_trad.index, 'cluster'] = km_trad.fit_predict(X_trad_sc)

# ANOVA — ¿difieren los clusters en precio/m²?
grupos = [g['price_per_m2'].values
           for _, g in df_trad.groupby('cluster')]
f_stat, p_anova = stats.f_oneway(*grupos)
print(f'ANOVA tradicional: F={f_stat:.4f}, p={p_anova:.4f}')

# — Mercado tokenizado —————
vars_cluster_tok = ['yield_real', 'min_investment', 'dist_centro',
                   'liquidez_norm']
X_tok = df_token[vars_cluster_tok].dropna()
X_tok_sc = scaler.fit_transform(X_tok)
best_k_tok, _, _ = seleccionar_k(X_tok_sc)
```

```
km_tok = KMeans(n_clusters=best_k_tok, random_state=42, n_init=10)
df_token.loc[X_tok.index, 'cluster'] = km_tok.fit_predict(X_tok_sc)
```

7. NOTA SOBRE REPRODUCIBILIDAD

Todo el código presentado en este anexo ha sido diseñado con criterios de modularidad y reproducibilidad. Las semillas aleatorias están fijadas (`random_state=42`) en todos los modelos que las requieren, y las rutas de datos son relativas al directorio del proyecto, lo que permite ejecutar el pipeline completo desde cualquier entorno con las dependencias instaladas.

Los archivos de datos originales no se incluyen en este anexo por razones de extensión, pero están disponibles a través de las fuentes indicadas en el apartado 5.1 de la memoria. La ejecución del pipeline completo, desde la carga de datos hasta la obtención de las figuras del capítulo 5 y 6, requiere únicamente las bibliotecas listadas en el apartado "Bibliotecas de Python utilizadas" del índice.

Versión del entorno: Python 3.8 · pandas 1.5 · numpy 1.23 · scikit-learn 1.1 · scipy 1.9 · matplotlib 3.6 · seaborn 0.12