****

**Análisis y Predicción del**

**Mercado Inmobiliario en España**

**para la Optimización de la venta y compra**

**Germán Bosch Estévez**

**Máster de Ciencia de Datos de Kschool**

**23 de julio de 2024**

**Título:** Análisis y Predicción del Mercado Inmobiliario en España para la Optimización de la venta y compra

**Elaborado por** Germán Bosch Estévez

**Repositorio:** [git](https://github.com/GermanBo1/TFM-Data-Science-GBosch), [video](https://drive.google.com/file/d/1V-vSdDHIkyQeKPB_80HXE39QZs2swSNP/view?usp=drive_link)

**ABSTRACT**

The project focuses on developing a predictive model for real estate prices using advanced machine learning techniques, specifically Long Short-Term Memory (LSTM) neural networks. In the current context of the Spanish real estate market, accurately anticipating price trends has become a crucial factor for maintaining a competitive advantage. Fluctuations in real estate prices deeply impact investment strategies, buying and selling decisions, and long-term planning for sector companies.

The proposed model aims to overcome limitations of traditional methods, which often rely on static and less dynamic analyses. By utilizing historical data and key variables such as macroeconomic, demographic, and policy conditions, LSTM neural networks can capture complex and nonlinear patterns in time-series data, thereby enabling more precise forecasting of future price variations across different segments of the real estate market.

The project follows a structured approach that includes several crucial phases: data collection and preprocessing, LSTM model construction and training, performance evaluation using metrics like MAE, RMSE, and R², and finally implementation in an operational environment. A deployment strategy has been designed to update predictions quarterly based on new data.

During model evaluation, tests were conducted on different regional groups with varying appraisal prices, from high to low. Results demonstrated that the LSTM model was capable of generating accurate and useful predictions, providing companies with tools to make informed strategic decisions.

Despite its achievements, the project faces significant challenges such as limited data availability and infrequent updates, which can affect the model's predictive ability in dynamic market environments. This underscores the importance of implementing a continuous monitoring plan to adjust and enhance the model in response to market fluctuations and changes in economic variables.

In conclusion, the developed model represents a step forward in the ability to anticipate and manage risks in the real estate sector. It offers a valuable tool for companies and investors looking to optimize investment strategies and maximize market opportunities, despite inherent challenges related to data availability and update frequency.

Índice

[1. Introducción 7](#_Toc172661214)

[2. Objetivos 8](#_Toc172661215)

[3. Descripción de los datos utilizados en el proyecto 9](#_Toc172661216)

[3.1. Descripción de las variables 11](#_Toc172661217)

[Tasación media de vivienda libre (tasacion\_media) 11](#_Toc172661218)

[Precio medio del metro cuadrado de suelo urbano (pre\_m2) 12](#_Toc172661219)

[Superficie de las transacciones de suelo (total) (sup\_tran) 13](#_Toc172661220)

[Número total de transacciones de suelo (num\_tran) 14](#_Toc172661221)

[3.2. Exploración de los datos 15](#_Toc172661222)

[Tasación media de vivienda libre (tasacion\_media) 15](#_Toc172661223)

[Precio medio del metro cuadrado de suelo urbano (pre\_m2) 17](#_Toc172661224)

[Número total de transacciones de suelo (num\_tran) 18](#_Toc172661225)

[Superficie total de transacciones de suelo (sup\_tran) 19](#_Toc172661226)

[3.3. Relación entre variables 21](#_Toc172661227)

[Análisis general 21](#_Toc172661228)

[Correlación entre variables 22](#_Toc172661229)

[4. Plataforma tecnológica: arquitectura de referencia para el modelado 24](#_Toc172661230)

[4.1. Procesamiento de datos 25](#_Toc172661231)

[Descarga y preparación de datos 25](#_Toc172661232)

[Tecnologías y herramientas utilizadas 25](#_Toc172661233)

[4.2. Desarrollo de modelos predictivos 26](#_Toc172661234)

[Enfoques de modelado 26](#_Toc172661235)

[4.3. Evaluación de modelos 27](#_Toc172661236)

[4.4. Despliegue de modelos 27](#_Toc172661237)

[4.5. Previsiones futuras 28](#_Toc172661238)

[5. Modelización 28](#_Toc172661239)

[5.1. Modelo analítico básico 28](#_Toc172661240)

[Descripción del modelo 28](#_Toc172661241)

[Evaluación del modelo 29](#_Toc172661242)

[Resultados obtenidos 30](#_Toc172661243)

[5.2. Modelo de Machine Learning con Random Forest 31](#_Toc172661244)

[Descripción del modelo 31](#_Toc172661245)

[Evaluación del modelo 33](#_Toc172661246)

[Resultados obtenidos 33](#_Toc172661247)

[5.3. Modelo de red neuronal con LSTM 38](#_Toc172661248)

[Descripción del modelo 38](#_Toc172661249)

[Evaluación del modelo 42](#_Toc172661250)

[Resultados obtenidos 43](#_Toc172661251)

[5.4. Conclusión sobre los modelos 46](#_Toc172661252)

[6. Resultados 47](#_Toc172661253)

[6.1. Cumplimiento de los objetivos empresariales 47](#_Toc172661254)

[6.2. Limitaciones y consideraciones futuras 48](#_Toc172661255)

[7. Despliegue en la arquitectura tecnológica de explotación 49](#_Toc172661256)

[7.1. Estrategia de despliegue del modelo 49](#_Toc172661257)

[7.2. Comparación con otras opciones de despliegue 50](#_Toc172661258)

[8. Puesta en valor 51](#_Toc172661259)

[9. Conclusión 52](#_Toc172661260)

[Contribución 53](#_Toc172661261)

# Introducción

En el actual contexto del mercado inmobiliario, cada vez más tensionado en ciertas regiones, anticipar con precisión las tendencias de precios se ha vuelto crucial para mantener una ventaja competitiva. Las fluctuaciones en los precios de los bienes inmobiliarios tienen un impacto significativo en las estrategias de inversión, las decisiones de compra y venta, y la planificación a largo plazo de las empresas del sector. Este proyecto tiene como objetivo desarrollar un modelo predictivo de precios inmobiliarios para proporcionar predicciones precisas y oportunas que faciliten la toma de decisiones estratégicas.

Los modelos predictivos permiten prever movimientos del mercado con mayor precisión que los métodos tradicionales, gracias a su capacidad para manejar secuencias de datos temporales y capturar patrones complejos en los datos históricos. Es por ello que los modelos de machine learning, y las redes neuronales en particular, son ideales para predecir tendencias de precios inmobiliarios.

Para llevar a cabo este proyecto, se seguirán varias fases clave, que incluyen la recopilación y preprocesamiento de datos, la construcción y entrenamiento de diversos modelos, la evaluación y resultados de los mismos, y la implementación de una estrategia de despliegue. El modelo será probado en diferentes grupos de regiones para evaluar su capacidad de generalización y su robustez en diversos contextos del mercado.

# Objetivos

El objetivo principal de este proyecto es desarrollar un modelo predictivo robusto y preciso para los precios de tasación del mercado inmobiliario en España, utilizando para ello una variedad de datos históricos y macroeconómicos relacionados con el sector inmobiliario. La meta es que este modelo permita a las empresas del sector anticipar tendencias y tomar decisiones informadas para optimizar sus estrategias de negocio, reducir riesgos y mejorar la rentabilidad.

Además del objetivo principal, el proyecto persigue varios objetivos específicos. Primero, se pretende recopilar y analizar datos públicos relevantes del mercado inmobiliario. Segundo, se buscará identificar y evaluar las variables más influyentes en la evolución de los precios inmobiliarios y la correlación entre estas. Y tercero, se desarrollará una herramienta analítica útil para las empresas, que facilite la planificación estratégica y la gestión de inversiones

# Descripción de los datos utilizados en el proyecto

Los datos empleados en el presente proyecto se obtienen del sitio web del Ministerio de Transportes, Movilidad y Agenda Urbana. Los enlaces a los datos disponibles son los siguientes:

* Datos de transacciones de suelo y superficie
* Datos de precios del suelo urbano

A partir de estos enlaces, se descargaron y procesaron los archivos correspondientes a los siguientes conjuntos de datos:

* **Número total de transacciones de suelo**: Archivo denominado "Número total de transacciones de suelo.XLS", representado por la variable **num\_tran**. Este archivo contiene información sobre el valor tasado medio de vivienda libre, con la unidad expresada en euros por metro cuadrado (€/m²).
* **Precio medio del metro cuadrado de suelo urbano**: Archivo denominado "Precio medio del metro cuadrado de suelo urbano.XLS", representado por la variable **pre\_m2**. Este archivo proporciona datos sobre el precio medio del metro cuadrado de suelo urbano desglosado por comunidades autónomas y provincias, con la unidad en euros por metro cuadrado (€/m²).
* **Superficie de las transacciones de suelo**: Archivo denominado "Superficie de las transacciones de suelo. Total.XLS", representado por la variable **sup\_tran**. Este archivo recoge datos sobre la superficie total de las transacciones de suelo (m²).
* **Tasación media de vivienda libre**: Archivo denominado "valor\_tasado\_vivienda\_libre.xls", representado por la variable **tasacion\_media**. Este archivo presenta los datos históricos sobre la tasación media de la vivienda libre, con la unidad en euros por metro cuadrado (€/m²).

Dado que estos datos provienen de fuentes oficiales del gobierno, se considera que poseen un alto grado de fiabilidad, por lo que no se realiza un proceso de limpieza de datos. Sin embargo, uno de los inconvenientes encontrados es que las tablas están presentadas de manera muy manual, lo cual requiere un proceso significativo de ingeniería de variables para su posterior análisis.

Los datos disponibles están muy agrupados, ya que se encuentran distribuidos a nivel de comunidades autónomas y por trimestres. Además, como el registro de datos abarca desde 2004 hasta 2023, contamos con una tabla de 80 entradas por región.

No se ha logrado obtener datos públicos con una mayor granularidad debido a que otras posibles fuentes de datos, como portales inmobiliarios, presentan fuertes restricciones para la recolección de información.

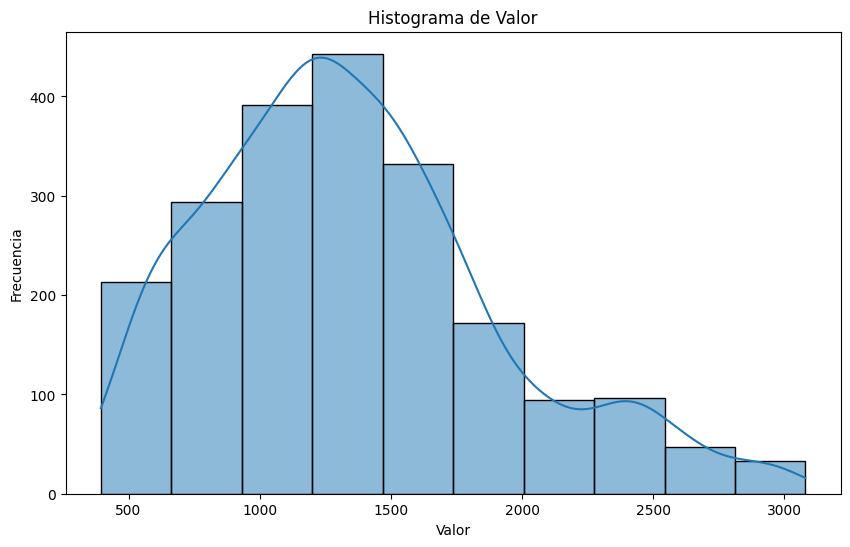
Entre las variables disponibles, **tasacion\_media** es la más relevante, ya que constituye la variable de interés principal del proyecto y el objetivo de su predicción. Adicionalmente, **tasacion\_media** cuenta con el registro histórico más extenso, comenzando en el año 1995, mientras que las demás variables tienen un inicio de registro más reciente.

Otras variables macroeconómicas podrían valorarse en estudios más profundos, tales como la renta media comunidad o el desempleo por comunidad autónoma.

A continuación, se presenta una descripción detallada de las variables empleadas en el análisis:

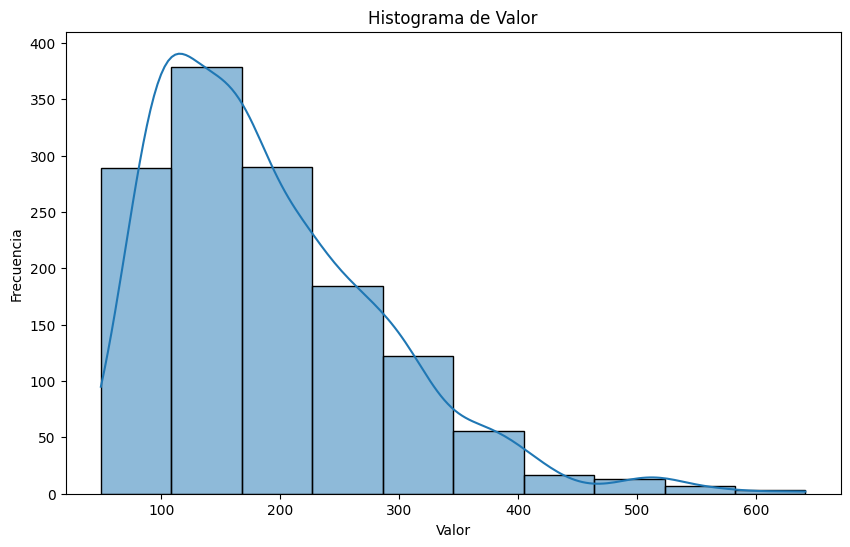
## Descripción de las variables

### Tasación media de vivienda libre (tasacion\_media)



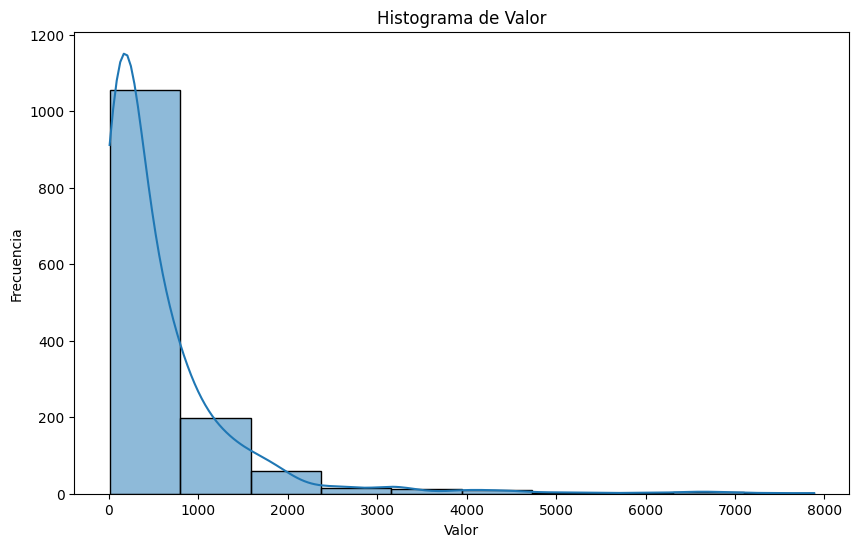
| **Estadística** | **Valor** |
| --- | --- |
| Media | 1354.04 € |
| Mediana | 1285.30 € |
| Desviación Típica | 553.98 € |
| Cuartil 25% | 953.20 € |
| Cuartil 50% (Mediana) | 1285.30 € |
| Cuartil 75% | 1654.95 € |
| Rango Intercuartílico | 701.75 € |

### Precio medio del metro cuadrado de suelo urbano (pre\_m2)



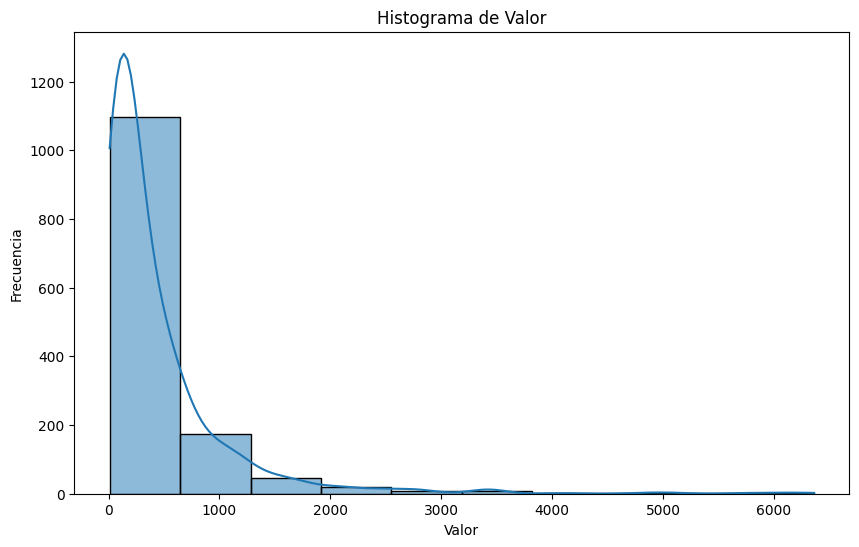
| **Estadística** | **Valor** |
| --- | --- |
| Media | 190.34 € |
| Mediana | 170.55 € |
| Desviación Típica | 96.28 € |
| Cuartil 25% | 114.52 € |
| Cuartil 50% (Mediana) | 170.55 € |
| Cuartil 75% | 244.48 € |
| Rango Intercuartílico | 129.96 € |
|  |  |

### Superficie de las transacciones de suelo (total) (sup\_tran)



| **Estadística** | **Valor** |
| --- | --- |
| Media | 592.84 m² |
| Mediana | 288.58 m² |
| Desviación Típica | 867.77 m² |
| Cuartil 25% | 116.37 m² |
| Cuartil 50% (Mediana) | 288.58 m² |
| Cuartil 75% | 713.62 m² |
| Rango Intercuartílico | * 1. ² |

### Número total de transacciones de suelo (num\_tran)



| **Estadística** | **Valor** |
| --- | --- |
| Media | 436.78 |
| Mediana | 204.00 |
| Desviación Típica | 647.46 |
| Cuartil 25% | 94.00 |
| Cuartil 50% (Mediana) | 204.00 |
| Cuartil 75% | 525.25 |
| Rango Intercuartílico | 431.25 |

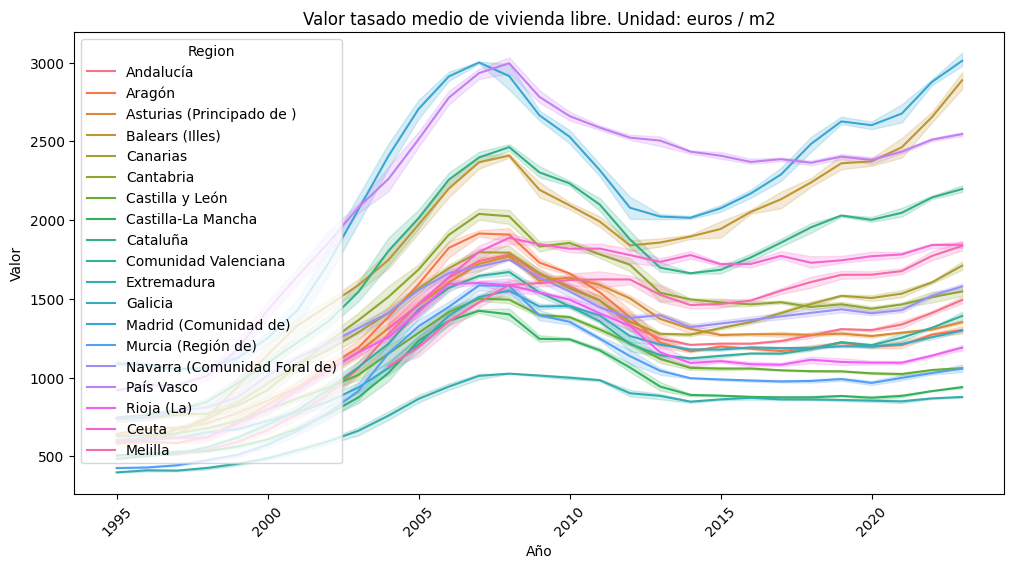
Estas estadísticas descriptivas permiten una comprensión preliminar de la naturaleza de los datos y orientan el análisis subsiguiente en el proyecto.

## Exploración de los datos

A continuación, se presenta un análisis detallado de las distintas variables, incluyendo observaciones sobre las tendencias históricas y actuales en las distintas comunidades autónomas.

### Tasación media de vivienda libre (tasacion\_media)

El análisis de la **tasación media de vivienda libre** para el año 2023 revela que las comunidades autónomas con las tasaciones más elevadas son Madrid, Islas Baleares, País Vasco y Cataluña. A continuación, se detallan las tasaciones medias para estas comunidades en 2023:



* **Comunidad de Madrid**: 3,011.25 €/m²
* **Islas Baleares**: 2,887.58 €/m²
* **País Vasco**: 2,545.98 €/m²
* **Cataluña**: 2,197.38 €/m²

El análisis histórico muestra que el precio de tasación más alto registrado se produjo en el cuarto trimestre de 2023 en la Comunidad de Madrid, alcanzando un valor de 3,011.25 €/m².

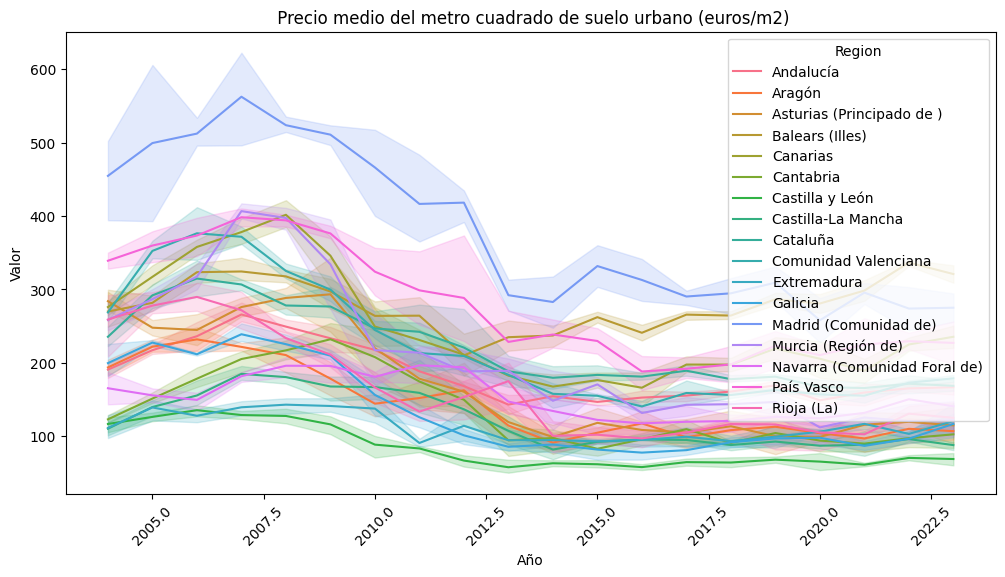
La variable **tasación media de vivienda libre es la que se predice en los posteriores modelos realizados, ya que nos permite extrapolar de una manera sencilla los datos al estado del mercado inmobiliario.**

Examinando las tendencias a lo largo del tiempo, se observa que, en los años previos a la crisis económica de 2008, los precios de tasación experimentaron un incremento constante durante el auge del mercado inmobiliario. Este período de crecimiento sostenido se interrumpió con la crisis económica global que comenzó en 2008, durante la cual los precios de tasación de viviendas disminuyeron hasta mediados de la década de 2010. A partir de 2014, los precios comenzaron a recuperarse, y hasta el año 2023, los datos indican que las comunidades con los precios más altos son, en orden descendente, la Comunidad de Madrid, Islas Baleares, País Vasco y Cataluña.

Este patrón histórico de fluctuación en los precios de tasación evidencia una fase de crecimiento acelerado seguida por una recesión durante la crisis económica, y una posterior fase de recuperación que continúa en la actualidad.

### Precio medio del metro cuadrado de suelo urbano (pre\_m2)

Para el año 2023, los precios más altos del metro cuadrado de suelo urbano se registran en **Islas Baleares, Madrid y Canarias.** Los valores respectivos para estas comunidades son:



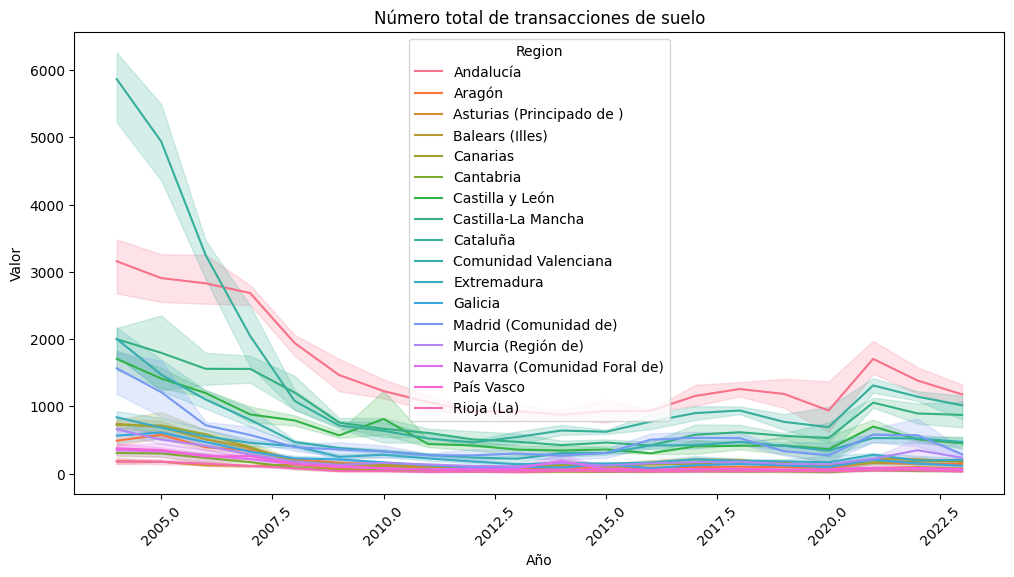
* **Islas Baleares**: 320.93 €/m²
* **Madrid**: 275.09 €/m²
* **Canarias**: 235.52 €/m²

El **"Precio Medio del Metro Cuadrado de Suelo Urbano"** refleja el costo promedio por metro cuadrado de terreno urbano en una región durante un período específico. Este indicador es crucial para evaluar el valor del suelo en diferentes ubicaciones geográficas.

A lo largo de los años, se observa que el **precio medio del metro cuadrado de suelo urbano** muestra una tendencia similar a la vista para la tasacion\_media, pero con un aumento más moderado después de la crisis inmobiliaria. Esto indica que, aunque hay una tendencia al alza en los precios del suelo urbano, esta tendencia es menos marcada en comparación con el aumento de los precios de las viviendas.

### Número total de transacciones de suelo (num\_tran)

Para el año 2023, se observa que **Andalucía** presenta el mayor número de transacciones de suelo, seguida de **Cataluña** y **Castilla-La Mancha**. Los valores respectivos para estas comunidades son:



* **Andalucía**: 1,177.75 transacciones
* **Cataluña**: 1,015.25 transacciones
* **Castilla-La Mancha**: 872.50 transacciones

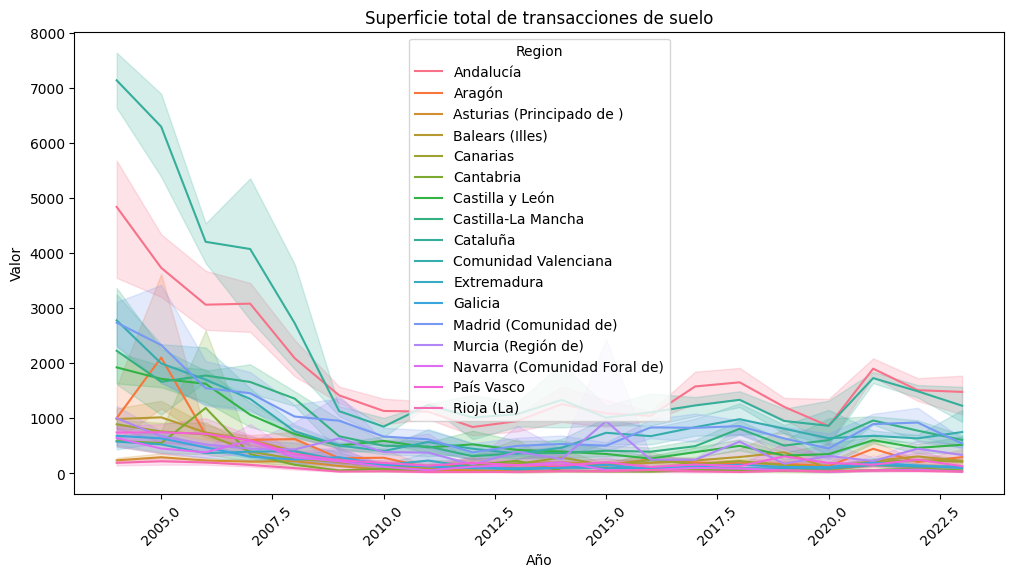
Sin embargo, estos valores son significativamente más bajos en comparación con los registros de los años 2004 y 2005, cuando se observaron los mayores volúmenes de transacciones.

Este indicador es fundamental para el análisis del mercado inmobiliario y el desarrollo urbano, ya que refleja la actividad económica asociada a la adquisición y el uso del suelo. Los datos históricos muestran que los años previos al auge del mercado inmobiliario, como los años 2004 y 2005, registraron los niveles más altos de transacciones de suelo. Posteriormente, se produjo una disminución en el número de transacciones, lo cual coincide con el período de crisis económica que comenzó en 2008.

A pesar de una demanda constante y un aumento en los precios de los bienes raíces, la oferta de terrenos disponibles en el mercado puede estar limitada por diversas razones como restricciones en el desarrollo urbano, regulaciones ambientales, limitaciones geográficas, o la escasez de terrenos adecuados para nuevos desarrollos.

### Superficie total de transacciones de suelo (sup\_tran)

En el año 2023, las comunidades autónomas con mayores superficies totales de transacciones de suelo son **Andalucía**, **Cataluña**, y **Comunidad Valenciana**. Los valores respectivos para estas comunidades son:



* **Andalucía**: 1,480.15 m²
* **Cataluña**: 1,229.14 m²
* **Comunidad Valenciana**: 750.76 m²

Al igual que el número total de transacciones, los valores de la superficie total de transacciones en 2023 son relativamente bajos en comparación con los picos alcanzados en los años 2004 y 2005.

El indicador de la **"Superficie Total de Transacciones de Suelo"** es similar al de **"Número Total de Transacciones de Suelo"** en el sentido de que ambos reflejan la actividad de compra y venta de terrenos en un área geográfica específica durante un período de tiempo determinado. A pesar de la demanda continua y el aumento de los precios, es posible que la oferta de terrenos disponibles esté restringida por factores como regulaciones de desarrollo urbano, restricciones ambientales, limitaciones geográficas, o la falta de terrenos adecuados para nuevos proyectos.

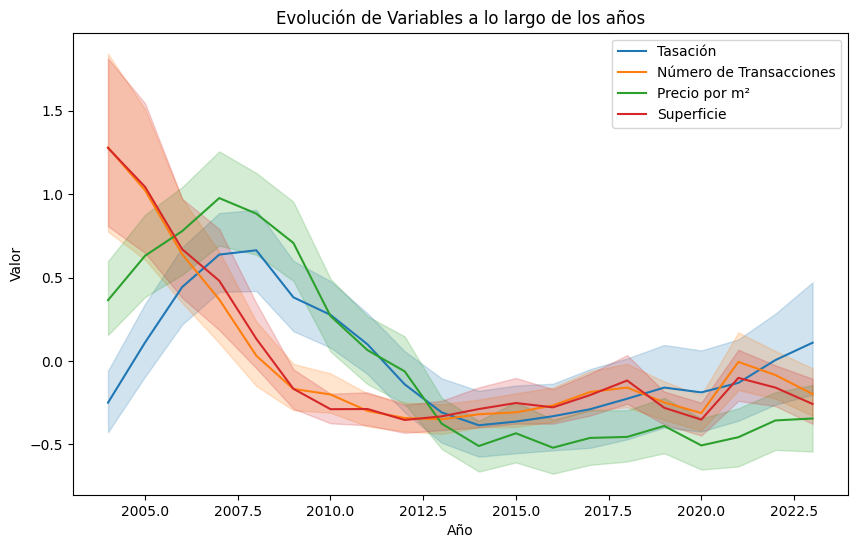
## Relación entre variables

La exploración de las relaciones entre las variables Número Total de Transacciones (num\_tran), Superficie Total de Transacciones de Suelo (sup\_tran), Tasación Media de Vivienda Libre (tasacion\_media) y Precio Medio del Metro Cuadrado de Suelo Urbano (pre\_m2) revela patrones importantes en el comportamiento del mercado inmobiliario.

### Análisis general

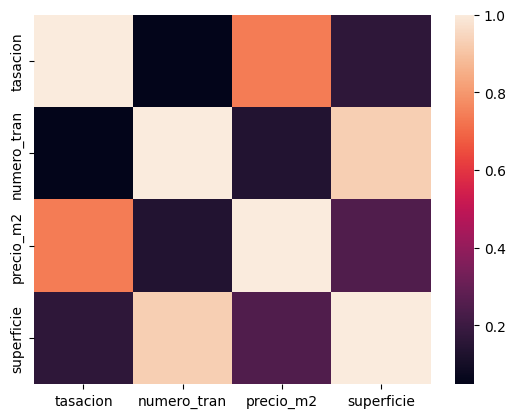
Al analizar un gráfico conjunto de las cuatro variables mencionadas, se observa que las variables **Número Total de Transacciones** y **Superficie Total de Transacciones de Suelo** experimentaron una disminución abrupta antes de que descendieran los valores de **Tasación Media** y **Precio por Metro Cuadrado de Suelo Urbano**. Esto sugiere que las primeras dos variables pueden ser indicadores adelantados de cambios en el mercado inmobiliario.

En los años más recientes, aunque los valores de **Tasación Media** y **Precio por Metro Cuadrado de Suelo Urbano** han aumentado, los valores de **Número Total de Transacciones** y **Superficie Total de Transacciones de Suelo** muestran una tendencia negativa. Esto indica una posible desconexión entre la demanda y la oferta en el mercado, y sugiere que, a pesar de la apreciación de los precios, la actividad de transacciones de suelo ha disminuido. Esto podría ser una de las causas de la subida progresiva de los precios del mercado inmobiliario.



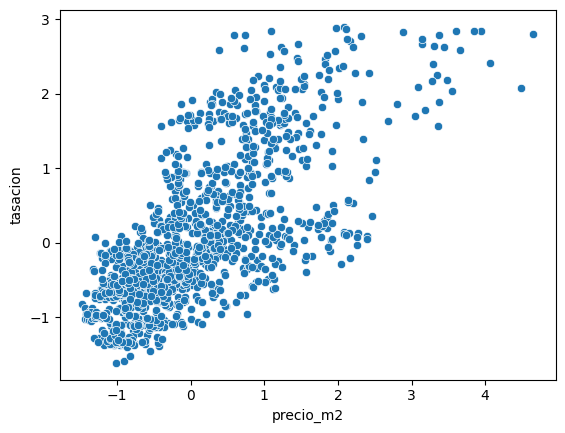
### Correlación entre variables

Para profundizar en las relaciones entre las variables, se utilizó un heatmap de correlación que muestra una relación significativa entre las distintas variables.



**Se da una correlación entre las** variables **Número Total de Transacciones y Superficie Total de Transacciones de Suelo como ya se apreciaba en la gráfica del apartado anterior.** La correlación se verificó mediante un test de correlación de Pearson, obteniéndose un coeficiente de correlación de 0.9280 y un p-valor de 0.0 (siendo por tanto significativa).

**Se da también una correlación entre Tasación Media** y **Precio por Metro Cuadrado de Suelo Urbano**. La correlación se verificó mediante un test de correlación de Pearson, obteniéndose un coeficiente de correlación de 0.7369 y un p-valor de 1.4576e-227 (siendo por tanto significativa). Esta correlación también se observa en una gráfica de nube de puntos, donde se destaca la relación positiva entre estas dos variables. La existencia de esta relación puede ser beneficiosa para la predicción de la variable **tasación\_media**.



# Plataforma tecnológica: arquitectura de referencia para el modelado

En el desarrollo de modelos analíticos y predictivos para el proyecto, se han seguido una serie de procedimientos tecnológicos y metodológicos para el tratamiento de datos, la creación de modelos predictivos, y la evaluación de los resultados. Esta sección describe en detalle la arquitectura tecnológica empleada en el proyecto, abarcando desde la descarga y procesamiento de datos hasta el desarrollo de modelos y su evaluación. Dado el carácter estático y de baja frecuencia de actualización de los datos, no se ha considerado necesario diseñar una arquitectura de despliegue compleja para el entorno de producción.

## Procesamiento de datos

### ****Descarga y preparación de datos****

Los datos utilizados en este proyecto se obtienen a partir de fuentes oficiales, específicamente del Ministerio de Transportes, Movilidad y Agenda Urbana. Dado que los datos se actualizan trimestralmente de manera manual, no se han implementado estrategias de extracción, transformación y carga (ETL) automatizadas. En su lugar, el proceso de descarga se realiza manualmente y los datos son tratados utilizando herramientas específicas para su preparación y análisis.

### ****Tecnologías y herramientas utilizadas****

* **Lenguaje de programación:** Python 3.12.0
* **Librerías de tratamiento de datos:**
  + import pandas as pd
  + import matplotlib.pyplot as plt
  + import seaborn as sns
  + from sklearn.preprocessing import StandardScaler
  + from scipy.stats import pearsonr

La librería Pandas se ha utilizado para la carga, limpieza y transformación de datos desde archivos Excel. Las transformaciones incluyen la combinación de datos en un único DataFrame y la exportación del DataFrame tratado a un archivo Excel final (final.xlsx). Dado que la cantidad de datos es relativamente pequeña y las actualizaciones son poco frecuentes, no se ha considerado necesario implementar una solución de almacenamiento en bases de datos SQL u otros sistemas de gestión de datos más avanzados.

## Desarrollo de modelos predictivos

### ****Enfoques de modelado****

Para predecir la variable objetivo, la tasación, se han explorado tres enfoques diferentes de modelado. Estos enfoques incluyen métodos analíticos básicos, modelos de Random Forest, y redes neuronales. A continuación, se detallan las herramientas y librerías utilizadas en cada enfoque:

**Modelos analíticos básicos y Machine Learning:**

* **import numpy as np**
* **import pandas as pd**
* **from sklearn.ensemble import RandomForestRegressor**
* **from sklearn.model\_selection import train\_test\_split**
* **from sklearn.metrics import mean\_squared\_error, mean\_absolute\_error, r2\_score**
* **import matplotlib.pyplot as plt**
* **from sklearn.preprocessing import MinMaxScaler**

Para los modelos analíticos y el Random Forest, se han utilizado herramientas de Python para realizar la segmentación de datos, el entrenamiento de modelos, y la evaluación de su rendimiento. RandomForestRegressor ha sido empleado para realizar predicciones basadas en múltiples variables, mientras que train\_test\_split, mean\_squared\_error, mean\_absolute\_error, y r2\_score han servido para validar y evaluar los modelos.

**Redes Neuronales:**

* **from tensorflow.keras.models import Sequential**
* **from tensorflow.keras.layers import LSTM, Dense**
* **from tensorflow.keras.callbacks import EarlyStopping**
* **from tensorflow.keras import Input**

En el desarrollo de redes neuronales, se ha utilizado Keras con TensorFlow como backend. Se ha creado una red neuronal con capas LSTM para predecir el valor de tasación un trimestre antes. Las técnicas de regularización y optimización, como EarlyStopping, han sido implementadas para evitar el sobreajuste y mejorar la generalización del modelo.

## Evaluación de modelos

La evaluación de los modelos se ha realizado utilizando métricas estándar para medir el rendimiento de los modelos predictivos, tales como el error cuadrático medio (MSE), el error absoluto medio (MAE), y el coeficiente de determinación (R²). Estas métricas permiten determinar la precisión de las predicciones y la calidad general de los modelos.

## Despliegue de modelos

Dado que el proyecto se basa en la creación de predicciones para variables que se actualizan trimestralmente, no se ha desarrollado una arquitectura de referencia específica para el despliegue en producción. La naturaleza estática de las actualizaciones de datos hace que no sea rentable ni necesario construir una infraestructura de despliegue compleja. En su lugar, se ha optado por una estrategia basada en informes a demanda, elaborados por el departamento de analítica a solicitud de los distintos departamentos responsables de las decisiones estratégicas.

## Previsiones futuras

La arquitectura tecnológica diseñada para este proyecto ha sido efectiva para la descarga, procesamiento, modelado y evaluación de datos. La solución empleada considera el bajo volumen de datos y la frecuencia de actualización, eligiendo herramientas y enfoques que son adecuados para los requisitos actuales del proyecto. Para futuros desarrollos, se podrían considerar otras fuentes de datos que complementen mejor el modelo, la búsqueda de datos más detallados e incluso la automatización de la captura de los datos deseados.

# Modelización

En esta sección se detallarán las técnicas de modelización empleadas para predecir la tasación de propiedades inmobiliarias en España, el proceso de evaluación de los modelos y los resultados analíticos obtenidos. El objetivo principal ha sido construir modelos predictivos que permitan estimar la variable de interés, **tasacion\_media**, basándose en datos históricos y diversos enfoques metodológicos. Se han utilizado tanto métodos analíticos básicos como técnicas avanzadas de machine learning para alcanzar este objetivo.

## Modelo analítico básico

### ****Descripción del modelo****

El primer enfoque consistió en un modelo analítico simple basado en una regresión lineal para predecir la tasación futura. El código utilizado para este modelo es el siguiente:

*def* calculo\_prevision(*df*, *año\_corte*):

    pendientes\_por\_region = {}

    for region, data\_region in *df*[*df*["Año"] > *año\_corte*].groupby("Region"):

        pendiente, \_ = np.polyfit(data\_region["Año"], data\_region["Valor"], 1)

        media\_2023 = data\_region[data\_region["Año"] == 2023]["Valor"].mean()

        # Calculo de la media esperada para el año 2024

        media\_esperada\_2024 = media\_2023 + pendiente

        # Calculo de la media esperada para el año 2025

        media\_esperada\_2025 = media\_esperada\_2024 + pendiente

        pendientes\_por\_region[region] = {"Pendiente": pendiente, "Media\_2023": media\_2023,

                                         "Media\_esperada\_2024": media\_esperada\_2024, "Media\_esperada\_2025": media\_esperada\_2025}

    pendientes\_df = pd.DataFrame.from\_dict(pendientes\_por\_region, *orient*="index")

    pendientes\_df = pendientes\_df.sort\_values(*by*="Pendiente", *ascending*=False)

    return pendientes\_df

### ****Evaluación del modelo****

Este modelo ofrece una estimación básica de las tasaciones futuras usando una extrapolación lineal simple basada en la tendencia histórica. Sin embargo, este enfoque tiene limitaciones significativas:

* **Simplicidad:** No considera la complejidad de las relaciones entre variables ni los patrones no lineales.
* **Sin Validación Formal:** No se realizó una validación formal del modelo ya que se basa en técnicas analíticas más que en métodos de machine learning.

### ****Resultados obtenidos****

El modelo proporciona estimaciones directas de tasación para los años 2024 y 2025. Sin embargo, la naturaleza simple de este modelo limita su capacidad para capturar patrones complejos en los datos históricos.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Comunidad Autónoma** | **Pendiente** | **Media 2023** | **Media esperada 2024** | **Media esperada 2025** |
| Madrid (Comunidad de) | 112,67375 | 3011,25 | 3123,92375 | 3236,5975 |
| Balears (Illes) | 106,252917 | 2887,575 | 2993,827917 | 3100,080833 |
| Cataluña | 60,330417 | 2197,375 | 2257,705417 | 2318,035833 |
| Canarias | 43,818333 | 1710,95 | 1754,768333 | 1798,586667 |
| Melilla | 43,69375 | 1836,525 | 1880,21875 | 1923,9125 |
| Navarra (Comunidad Foral de) | 40,07 | 1579 | 1619,07 | 1659,14 |
| Andalucía | 32,281667 | 1493,075 | 1525,356667 | 1557,638333 |
| Comunidad Valenciana | 28,87625 | 1391,825 | 1420,70125 | 1449,5775 |
| País Vasco | 18,075 | 2545,975 | 2564,05 | 2582,125 |
| Ceuta | 15,193333 | 1843,35 | 1858,543333 | 1873,736667 |
| Aragón | 12,91125 | 1303,35 | 1316,26125 | 1329,1725 |
| Galicia | 12,24875 | 1298,775 | 1311,02375 | 1323,2725 |
| Rioja (La) | 8,531667 | 1190,25 | 1198,781667 | 1207,313333 |
| Murcia (Región de) | 7,52375 | 1057,4 | 1064,92375 | 1072,4475 |
| Asturias (Principado de) | 7,155 | 1352 | 1359,155 | 1366,31 |
| Cantabria | 6,231667 | 1547,1 | 1553,331667 | 1559,563333 |
| Castilla-La Mancha | 5,699583 | 939,875 | 945,574583 | 951,274167 |
| Extremadura | 0,280833 | 877,35 | 877,630833 | 877,911667 |
| Castilla y León | -1,32375 | 1059,975 | 1058,65125 | 1057,3275 |

## Modelo de Machine Learning con Random Forest

### ****Descripción del modelo****

El segundo enfoque utiliza un modelo de Random Forest por comunidad autónoma para predecir la tasación de propiedades. El código para este modelo es el siguiente:

import numpy as np

import pandas as pd

from sklearn.ensemble import RandomForestRegressor

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import mean\_squared\_error, mean\_absolute\_error, r2\_score

import matplotlib.pyplot as plt

df\_final = pd.read\_excel("final.xlsx")

df\_final["Trimestre"] = df\_final["Trimestre"].apply(*lambda* *x*: int(*x*[0]))

predicciones\_por\_region = {}

# Definir las variables predictoras

variables\_predictoras = ["Año", "Trimestre", "numero\_tran", "precio\_m2", "superficie"]

# Iterar sobre cada región

for region, datos\_region in df\_final.groupby("Region"):

    if all(var in datos\_region.columns for var in variables\_predictoras):

        # Crear las variables predictoras (X) y la variable objetivo (y)

        X = datos\_region[variables\_predictoras].values

        y = datos\_region["tasacion"].values

        X\_entrenamiento, X\_prueba, y\_entrenamiento, y\_prueba = train\_test\_split(X, y, *test\_size*=0.2, *random\_state*=42)

        modelo\_regresion = RandomForestRegressor(*n\_estimators*=100, *random\_state*=42)

        modelo\_regresion.fit(X\_entrenamiento, y\_entrenamiento)

        y\_pred = modelo\_regresion.predict(X\_prueba)

        error\_cuadratico\_medio = mean\_squared\_error(y\_prueba, y\_pred)

        mae = mean\_absolute\_error(y\_prueba, y\_pred)

        rmse = np.sqrt(error\_cuadratico\_medio)

        r2 = r2\_score(y\_prueba, y\_pred)

        mean\_tasacion = np.mean(y\_prueba)

        mae\_relacion\_media = (mae / mean\_tasacion) \* 100

        rmse\_relacion\_media = (rmse / mean\_tasacion) \* 100

        print(*f*"Región: {region} -->  MAE: {mae*:.2f*}, RMSE: {rmse*:.2f*}, R²: {r2*:.2f*}, MAE en relación con la media: {mae\_relacion\_media*:.2f*}%, RMSE en relación con la media: {rmse\_relacion\_media*:.2f*}%")

        valores\_recientes = datos\_region[variables\_predictoras[2:]].iloc[-1].values.reshape(1, -1)  # Tomar las últimas variables predictoras (sin Año y Trimestre)

        predicciones\_2023 = []

        for trimestre in range(1, 5):

            X\_prediccion = np.hstack((np.array([[2023, trimestre]]), valores\_recientes))

            prediccion\_trimestre = modelo\_regresion.predict(X\_prediccion)

            predicciones\_2023.append(prediccion\_trimestre[0])

        predicciones\_por\_region[region] = predicciones\_2023

for region, datos\_region in df\_final.groupby("Region"):

    plt.figure(*figsize*=(10, 6))

    plt.plot(datos\_region["Año"] + (datos\_region["Trimestre"] - 1)/4, datos\_region["tasacion"], *label*="Histórico")

    trimestres\_2023 = [2023 + (t - 1)/4 for t in range(1, 5)]

    plt.scatter(trimestres\_2023, predicciones\_por\_region[region], *color*='red', *label*="Predicciones 2023")

    plt.title(*f*"Predicciones para la región {region}")

    plt.xlabel("Año")

    plt.ylabel("Tasación")

    plt.legend()

    plt.show()

### Evaluación del modelo

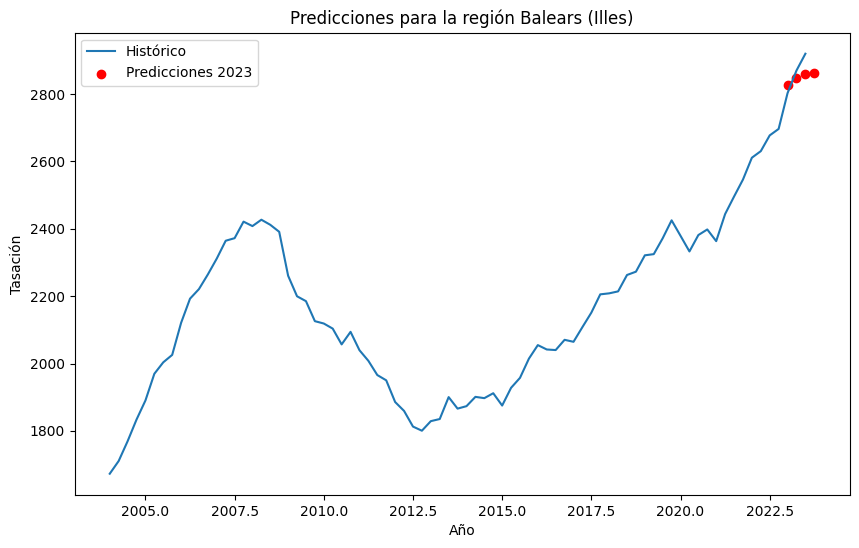
El modelo de Random Forest muestra un desempeño notable en términos de error absoluto medio (MAE), raíz del error cuadrático medio (RMSE) y coeficiente de determinación (R²). Los resultados indican que:

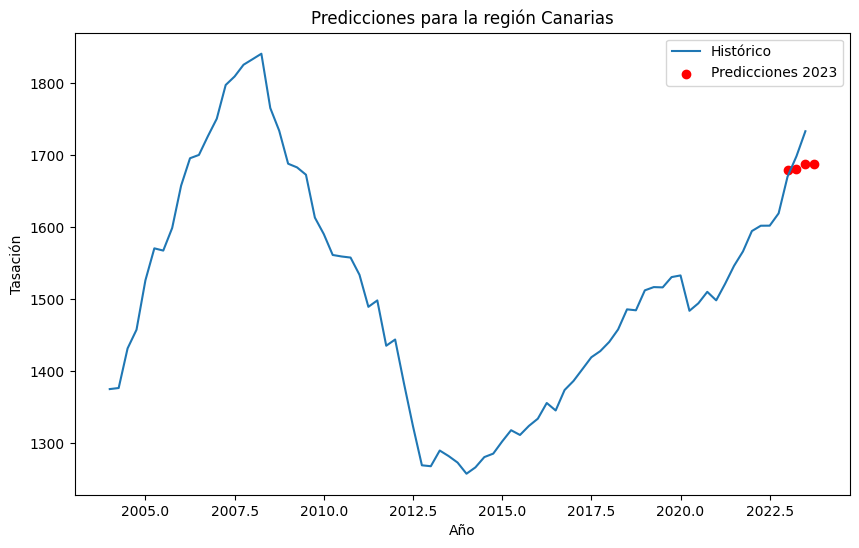
* **Rendimiento General:** La mayoría de las regiones muestran un R² alto (por encima de 0.70), indicando que el modelo explica una proporción significativa de la variabilidad en las tasaciones.
* **Errores Relativos:** El MAE y el RMSE en relación con la media de tasación son relativamente bajos en la mayoría de las regiones, con un R² promedio de 0.79.

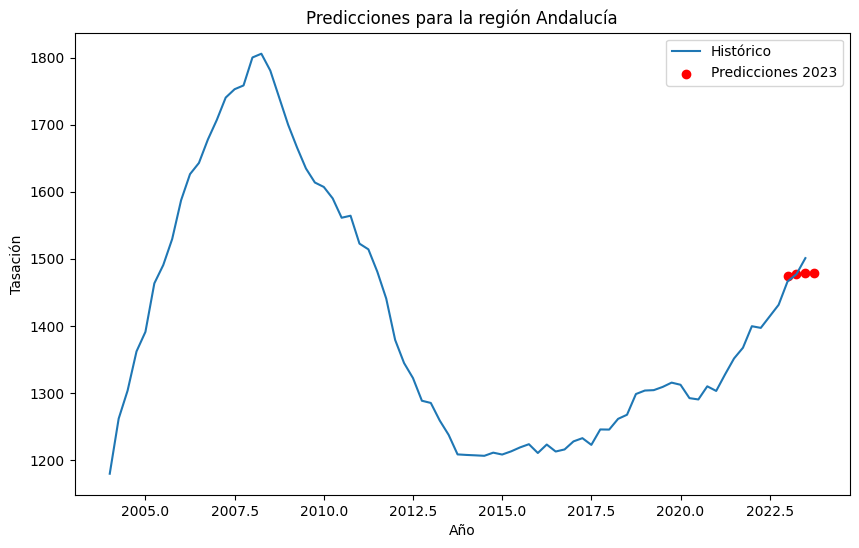
### Resultados obtenidos

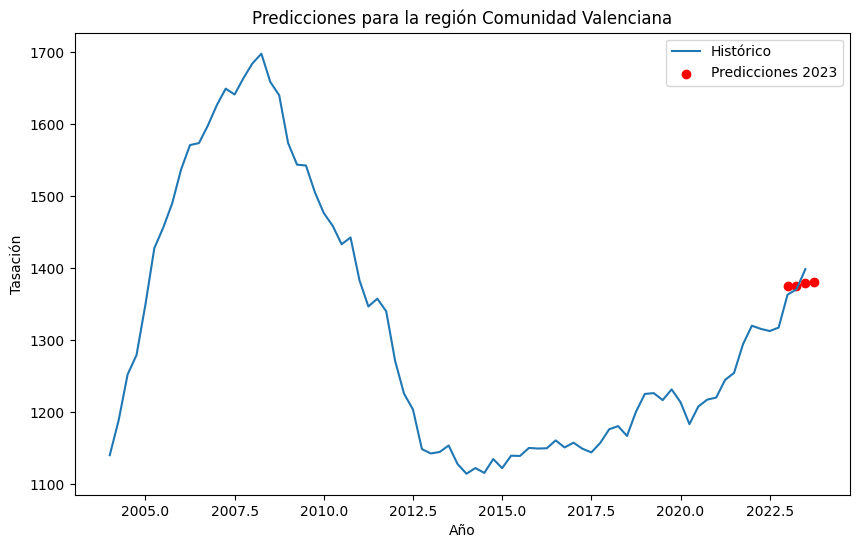
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Región** | **MAE** | **RMSE** | **R²** | **MAE en relación con la media (%)** | **RMSE en relación con la media (%)** |
| Andalucía | 67.34 | 84.60 | 0.78 | 4.72% | 5.93% |
| Aragón | 61.49 | 87.06 | 0.88 | 4.19% | 5.94% |
| Asturias | 44.67 | 68.79 | 0.82 | 3.06% | 4.70% |
| Balears (Illes) | 55.53 | 73.30 | 0.91 | 2.65% | 3.50% |
| Canarias | 52.56 | 81.27 | 0.70 | 3.48% | 5.39% |
| Cantabria | 75.63 | 93.81 | 0.77 | 4.49% | 5.56% |
| Castilla y León | 31.80 | 57.18 | 0.88 | 2.60% | 4.67% |
| Castilla-La Mancha | 29.74 | 41.73 | 0.95 | 2.71% | 3.80% |
| Cataluña | 79.10 | 102.84 | 0.80 | 3.90% | 5.08% |
| Comunidad Valenciana | 46.25 | 60.73 | 0.88 | 3.47% | 4.55% |
| Extremadura | 41.33 | 51.61 | 0.59 | 4.54% | 5.67% |
| Galicia | 52.13 | 94.07 | 0.56 | 4.03% | 7.27% |
| Madrid | 101.86 | 138.82 | 0.80 | 4.12% | 5.62% |
| Murcia | 53.64 | 74.98 | 0.86 | 4.48% | 6.26% |
| Navarra | 41.09 | 52.81 | 0.85 | 2.70% | 3.47% |
| País Vasco | 59.04 | 87.04 | 0.83 | 2.31% | 3.40% |
| Rioja | 31.93 | 36.55 | 0.96 | 2.39% | 2.74% |

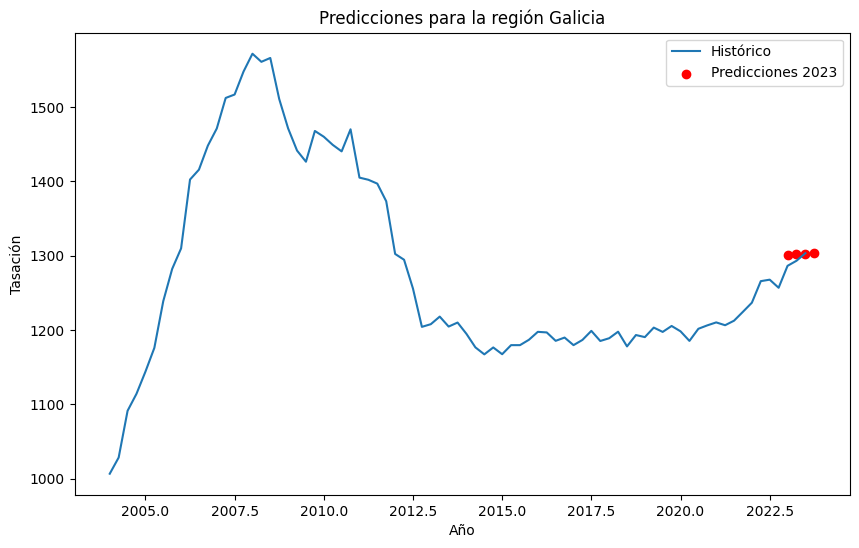
A diferencia del **modelo analítico**, que ofrece una estimación basada en una tendencia lineal simple, el **Random Forest**, siendo una técnica avanzada de machine learning, es capaz de capturar una mayor complejidad y variabilidad presentes en los datos históricos. Sin embargo, a pesar de obtener resultados aceptables en la evaluación, la comparación entre los resultados predichos y los reales muestran que, en las comunidades con fuertes tendencias en los datos, las predicciones tienden a ser más erróneas. Esto se aprecia en las siguientes figuras.

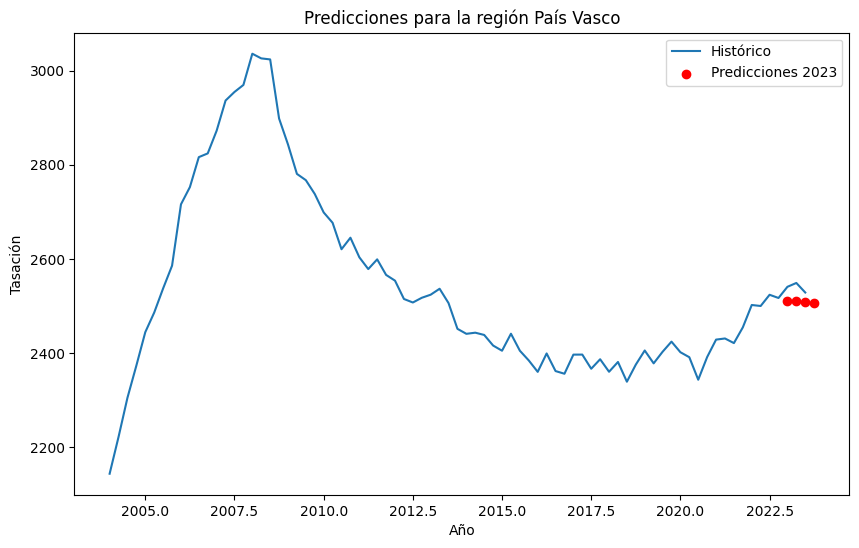


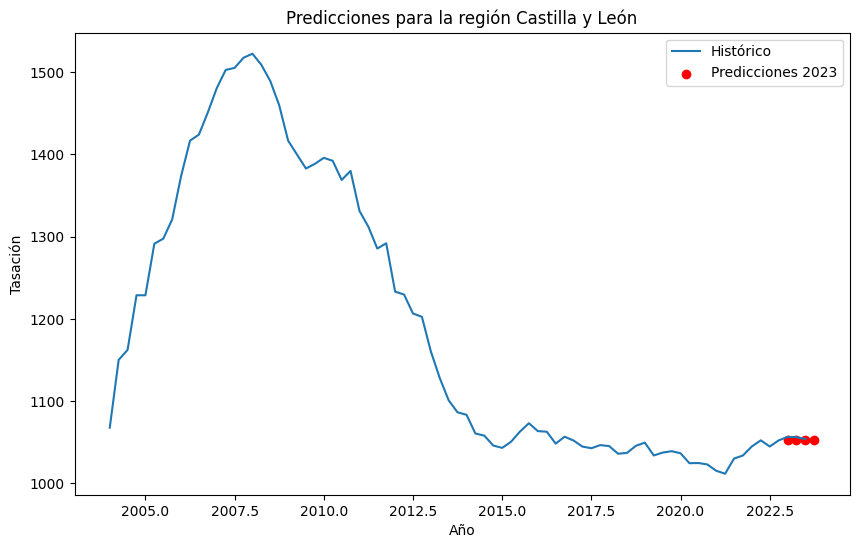












Además, este modelo de Random Forest presenta una limitación significativa en el contexto del negocio: no realiza predicciones futuras, sino que estima el valor actual de las variables en cuestión, lo que lo hace inútil, ya que los valores actuales ya son conocidos.

Por lo tanto, dado que el objetivo es realizar **predicciones a futuro en base a datos históricos**, se plantea el desarrollo de un **modelo de red neuronal**, modelos que son capaces de aprender patrones complejos y generalizar mejor para escenarios futuros.

## Modelo de red neuronal con LSTM

### ****Descripción del modelo****

El modelo de red neuronal se basa en una red neuronal recurrente LSTM (Long Short-Term Memory) para capturar patrones secuenciales en los datos históricos. Se desarrollaron tres modelos de red neuronal LSTM, cada uno entrenado en un grupo de regiones según el rango de precios de tasación: regiones con tasación alta, media y baja. El siguiente código se utilizó para construir y evaluar estos modelos:

import pandas as pd

import numpy as np

from sklearn.preprocessing import MinMaxScaler

from sklearn.metrics import mean\_squared\_error, mean\_absolute\_error, r2\_score

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import LSTM, Dense

from tensorflow.keras.callbacks import EarlyStopping

from tensorflow.keras import Input

import matplotlib.pyplot as plt

*def* red\_neuronal(*df*, *window\_size*):

    # Variables a utilizar

    features = ['tasacion', 'numero\_tran', 'precio\_m2', 'superficie']

    target = 'tasacion'

    # Normalización de los datos

    scaler = MinMaxScaler()

*df*.loc[:, features] = scaler.fit\_transform(*df*[features])

    # Función para crear ventanas deslizantes

*def* create\_dataset(*df*, *features*, *target*, *window\_size*):

        X, y = [], []

        for i in range(len(*df*) - *window\_size*):

            X.append(*df*[*features*].iloc[i:i + *window\_size*].values)

            y.append(*df*[*target*].iloc[i + *window\_size*])

        return np.array(X), np.array(y)

    # Creación del dataset

    X, y = create\_dataset(*df*, features, target, *window\_size*)

    # División de los datos en entrenamiento y prueba

    X\_train, X\_test = X[:-5], X[-5:]

    y\_train, y\_test = y[:-5], y[-5:]

    # Definición del modelo LSTM

    model = Sequential()

    model.add(Input(*shape*=(*window\_size*, len(features))))

    model.add(LSTM(50, *activation*='relu'))

    model.add(Dense(1))

    model.compile(*optimizer*='adam', *loss*='mse')

    # Early stopping

    early\_stopping = EarlyStopping(*monitor*='val\_loss', *patience*=10, *restore\_best\_weights*=True)

    # Entrenamiento del modelo

    history = model.fit(X\_train, y\_train, *validation\_data*=(X\_test, y\_test), *epochs*=100, *callbacks*=[early\_stopping], *verbose*=0)

    # Evaluación del modelo

    train\_loss = model.evaluate(X\_train, y\_train)

    val\_loss = model.evaluate(X\_test, y\_test)

    print(*f*'Training Loss: {train\_loss}')

    print(*f*'Validation Loss: {val\_loss}')

    # Predicciones

    y\_pred = model.predict(X\_test)

    temp\_test = pd.DataFrame(np.c\_[y\_test, np.zeros((len(y\_test), len(features)-1))], *columns*=features)

    temp\_pred = pd.DataFrame(np.c\_[y\_pred, np.zeros((len(y\_pred), len(features)-1))], *columns*=features)

    y\_test\_descaled = scaler.inverse\_transform(temp\_test)[:, 0]

    y\_pred\_descaled = scaler.inverse\_transform(temp\_pred)[:, 0]

    # Métricas de evaluación

    mse = mean\_squared\_error(y\_test\_descaled, y\_pred\_descaled)

    mae = mean\_absolute\_error(y\_test\_descaled, y\_pred\_descaled)

    r2 = r2\_score(y\_test\_descaled, y\_pred\_descaled)

    rmse = np.sqrt(mse)

    # Calcular media y mediana de los datos reales

    mean\_tasacion = np.mean(y\_test\_descaled)

    median\_tasacion = np.median(y\_test\_descaled)

    print(*f*'Media de tasacion: {mean\_tasacion}')

    print(*f*'Mediana de tasacion: {median\_tasacion}')

    print(*f*'MAE: {mae}')

    print(*f*'RMSE: {rmse}')

    print(*f*'R²: {r2}')

    print(*f*'MAE en relación con la media: {mae / mean\_tasacion \* 100*:.2f*}%')

    print(*f*'RMSE en relación con la media: {rmse / mean\_tasacion \* 100*:.2f*}%')

    # Gráfica de puntos: Valores reales vs Predicciones

    # Mostrar las predicciones junto con los datos históricos

    plt.figure(*figsize*=(14, 7))

    plt.scatter(range(len(*df*) - len(y\_test), len(*df*)), y\_test\_descaled, *color*='black', *label*="Valores reales")

    plt.scatter(range(len(*df*) - len(y\_test), len(*df*)), y\_pred\_descaled, *color*='red', *label*="Predicciones")

    plt.xlabel('Índice de Tiempo')

    plt.ylabel('Tasación')

    plt.title('Valores Reales vs Predicciones')

    plt.legend()

    plt.show()

    # Distribución de los Errores

    errors = y\_test\_descaled - y\_pred\_descaled

    plt.figure(*figsize*=(14, 7))

    plt.hist(errors, *bins*=20)

    plt.xlabel('Error')

    plt.ylabel('Frecuencia')

    plt.title('Distribución de los Errores')

    plt.show()

    # Curvas de Aprendizaje

    plt.figure(*figsize*=(14, 7))

    plt.plot(history.history['loss'], *label*='Pérdida de Entrenamiento')

    plt.plot(history.history['val\_loss'], *label*='Pérdida de Validación')

    plt.xlabel('Épocas')

    plt.ylabel('Pérdida')

    plt.title('Curvas de Aprendizaje')

    plt.legend()

    plt.show()

    # Devolver el modelo y el scaler

    return model, scaler

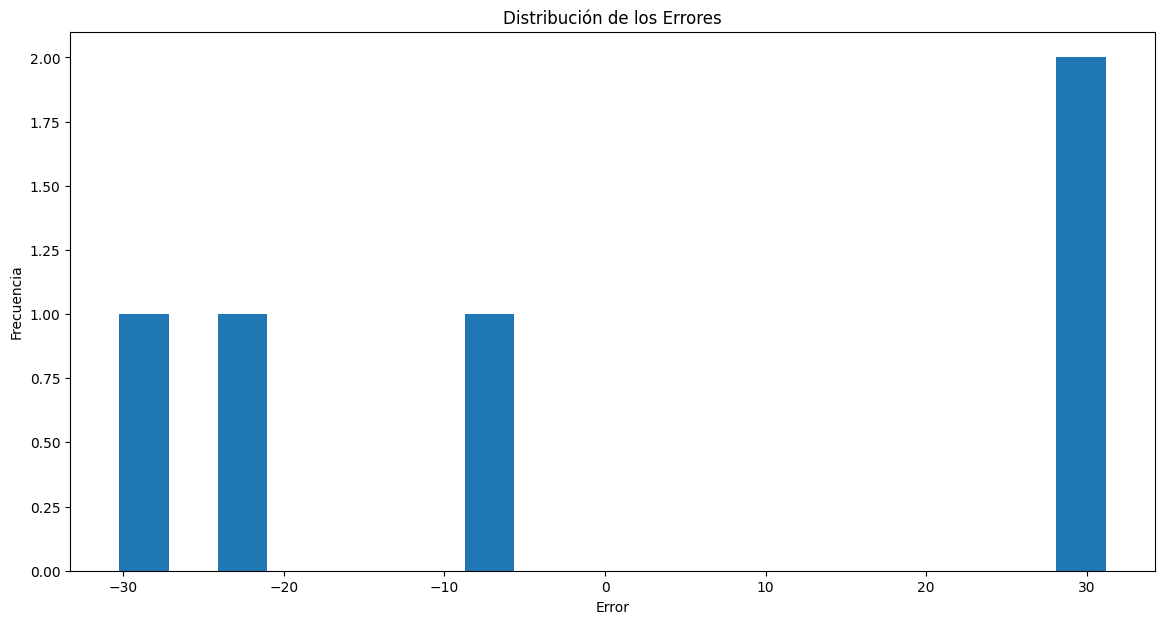
### Evaluación del modelo

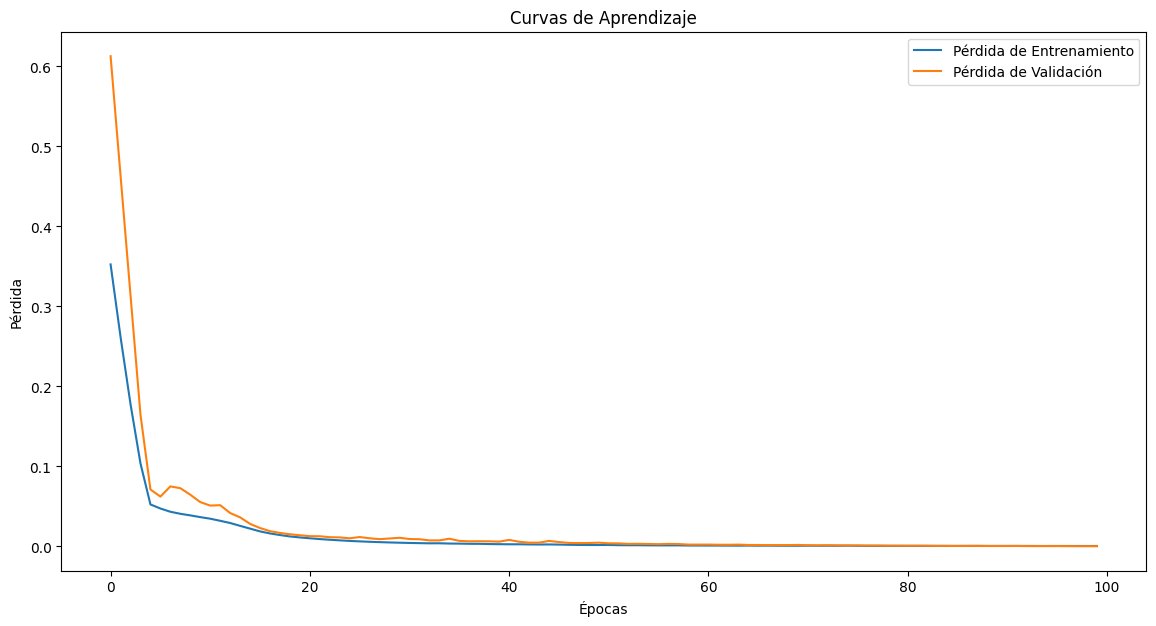
* **Rendimiento general:** El modelo de red neuronal muestra un **excelente desempeño** en términos de las métricas de evaluación. Los **errores absolutos promedio** (MAE) y la **raíz del error cuadrático medio** (RMSE) son bajos en todos los grupos, y el **coeficiente de determinación** (R²) es alto, indicando una fuerte capacidad del modelo para capturar la variabilidad en los datos de tasación.
* **Errores relativos:** El **MAE** y el **RMSE** en relación con la media de tasación son bastante bajos, con porcentajes que varían entre 1.14% y 2.42% para el MAE y entre 1.31% y 2.83% para el RMSE, lo que sugiere que el modelo hace predicciones muy precisas en comparación con los valores reales de tasación.

### Resultados obtenidos

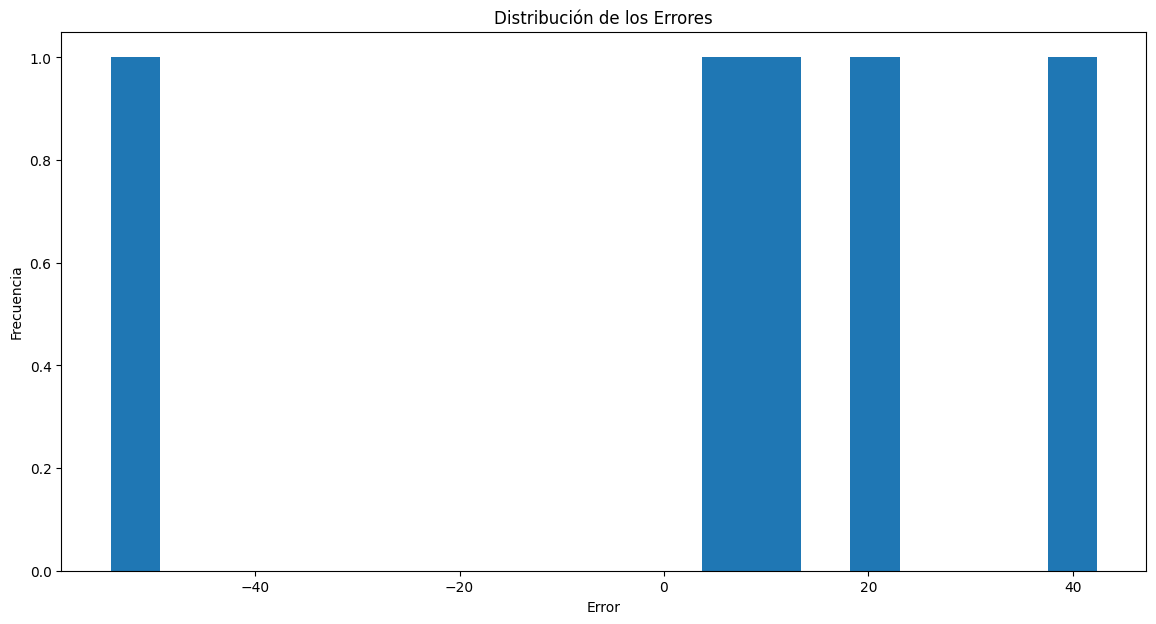
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Grupo de Regiones** | **MAE** | **RMSE** | **R²** | **MAE en Relación con la Media** | **RMSE en Relación con la Media** |
| Regiones con Tasación Alta | 30.20 | 34.65 | 986 | 1.14% | 1.31% |
| Regiones con Tasación Media | 26.91 | 39.27 | 913 | 1.76% | 2.57% |
| Regiones con Tasación Baja | 26.12 | 30.48 | 964 | 2.42% | 2.83% |

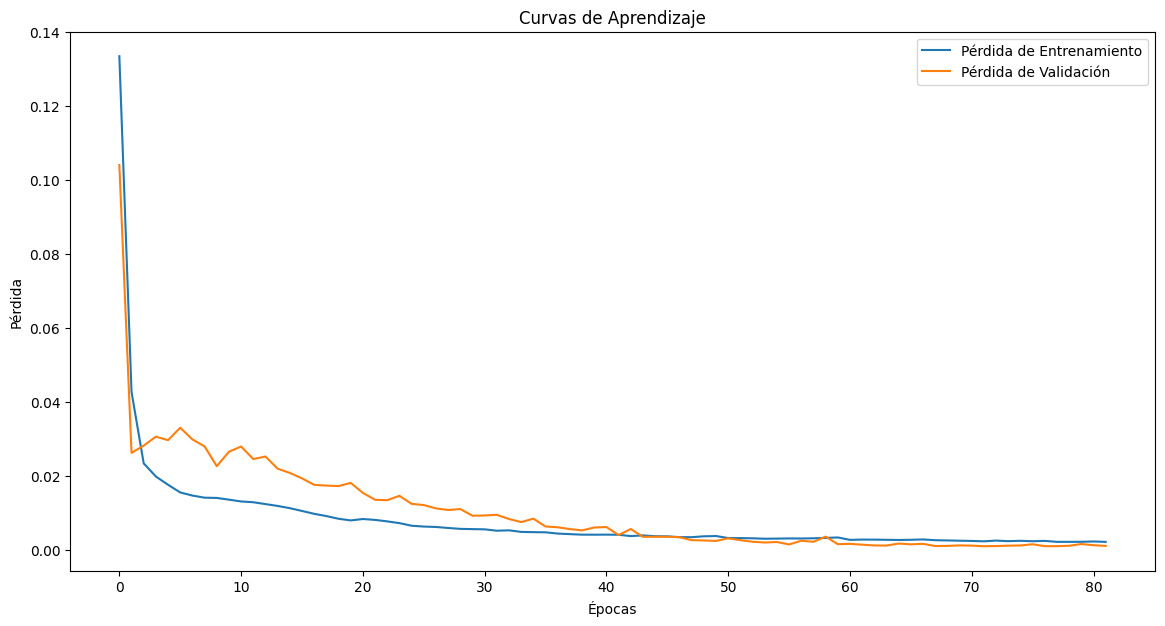
Respecto a las gráficas obtenidas, vemos que en el caso de las **regiones con Tasación Alta la distribución de los errores** es más o menos uniforme, y las **curvas de aprendizaje** muestran que tanto la **pérdida de entrenamiento** como la de **validación** disminuyen eficazmente, alcanzando niveles bajos en ambas métricas.



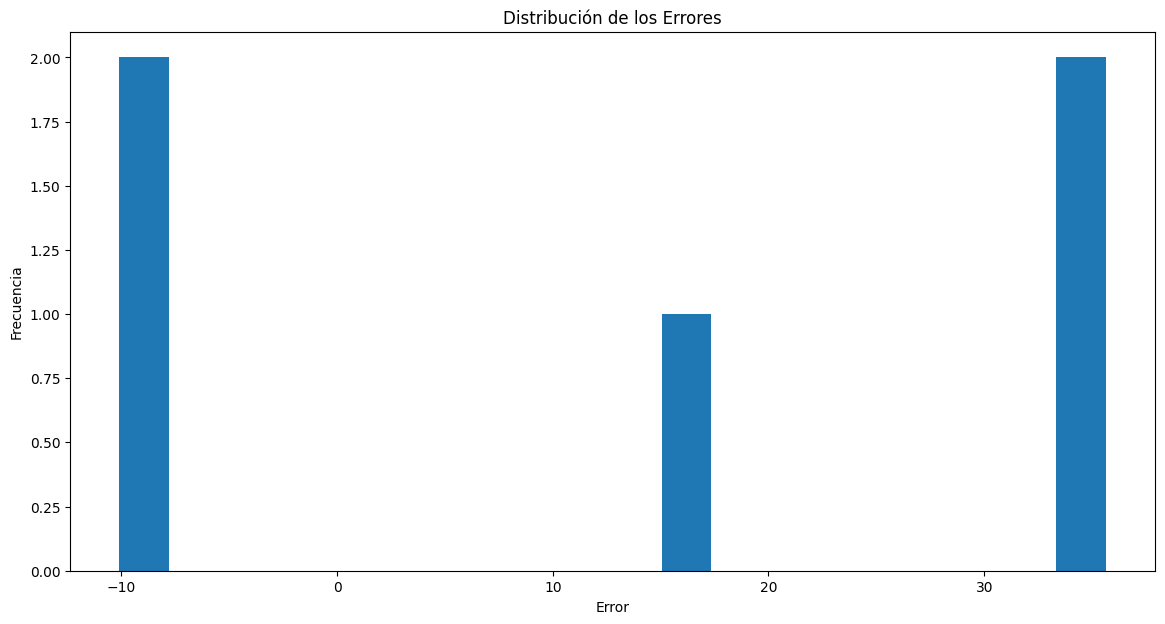


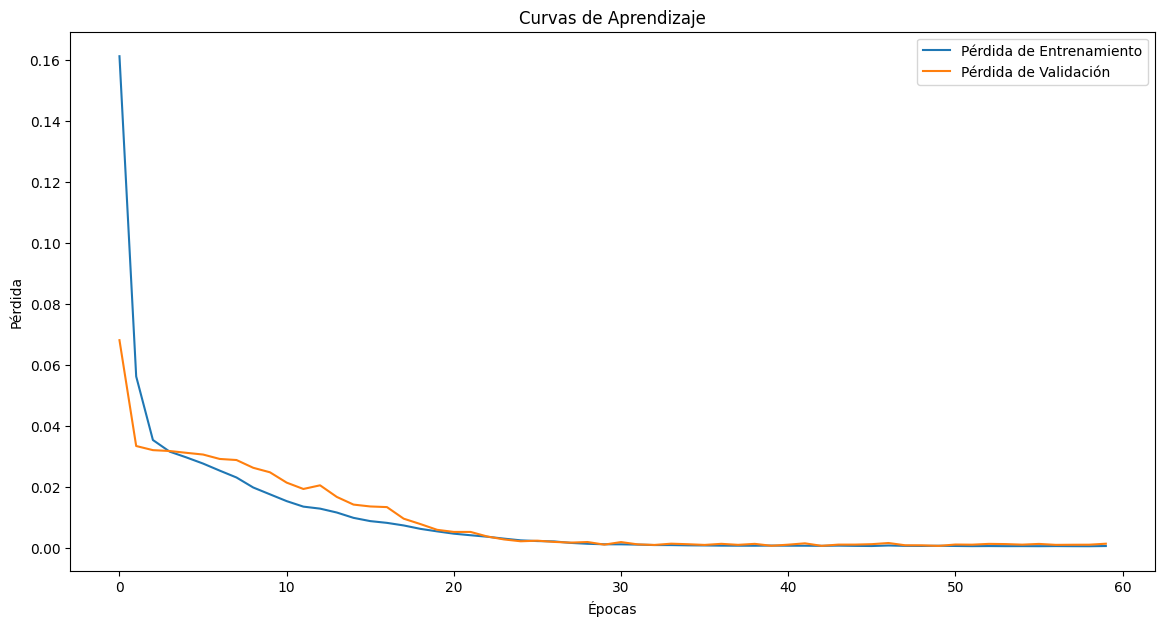
Respecto a las **regiones con Tasación Media, los errores tienden a ser más negativos** en comparación con el grupo de tasación alta. Las **curvas de aprendizaje** demuestran una **convergencia efectiva** del modelo, aunque con **mayor variabilidad en los errores**.





Por último, en el caso de las r**egiones con Tasación Baja, l**as **gráficas** muestran patrones similares a los observados en las regiones de tasación alta, con una **distribución de errores** relativamente uniforme y **curvas de aprendizaje** que reflejan una **buena convergencia** del modelo.





## Conclusión sobre los modelos

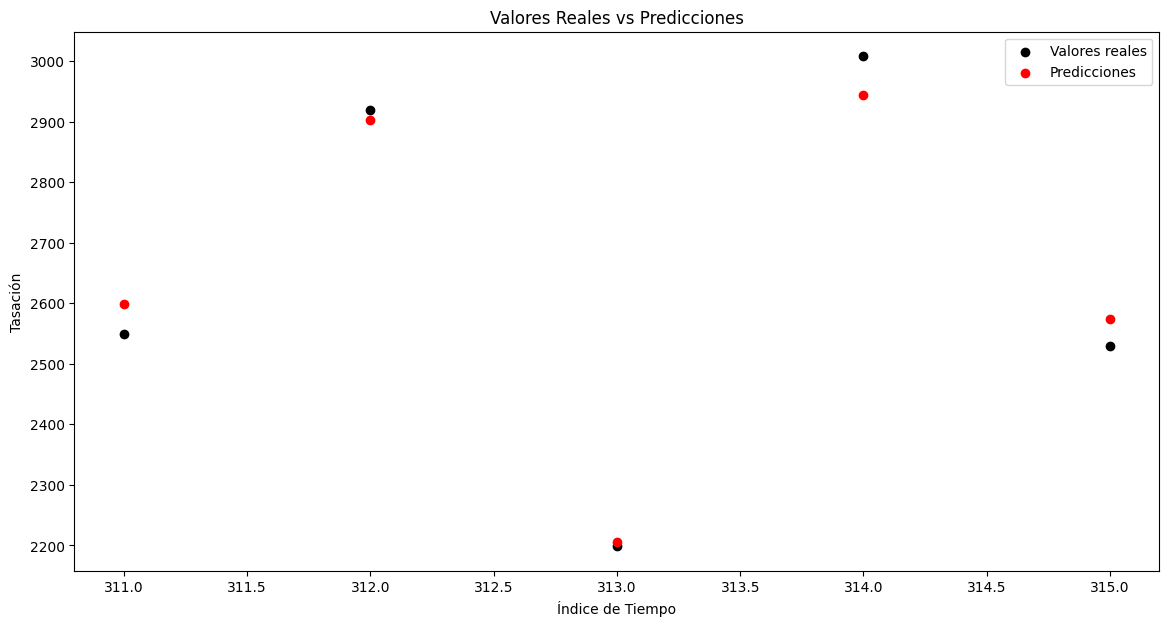
Después de la evaluación de los tres modelos desarrollados —el modelo analítico básico, el Random Forest y la red neuronal LSTM— se concluye que**, la red neuronal LSTM es la mejor opción para predecir las tasaciones futuras de manera fiable y efectiva**.

Los resultados obtenidos evidencian que el LSTM no solo ofrece **precisión en las predicciones,** sino que también proporciona una **solución práctica** para anticipar las tasaciones en escenarios futuros, a diferencia del modelo analítico básico y del Random Forest, que presentan limitaciones en estos aspectos.

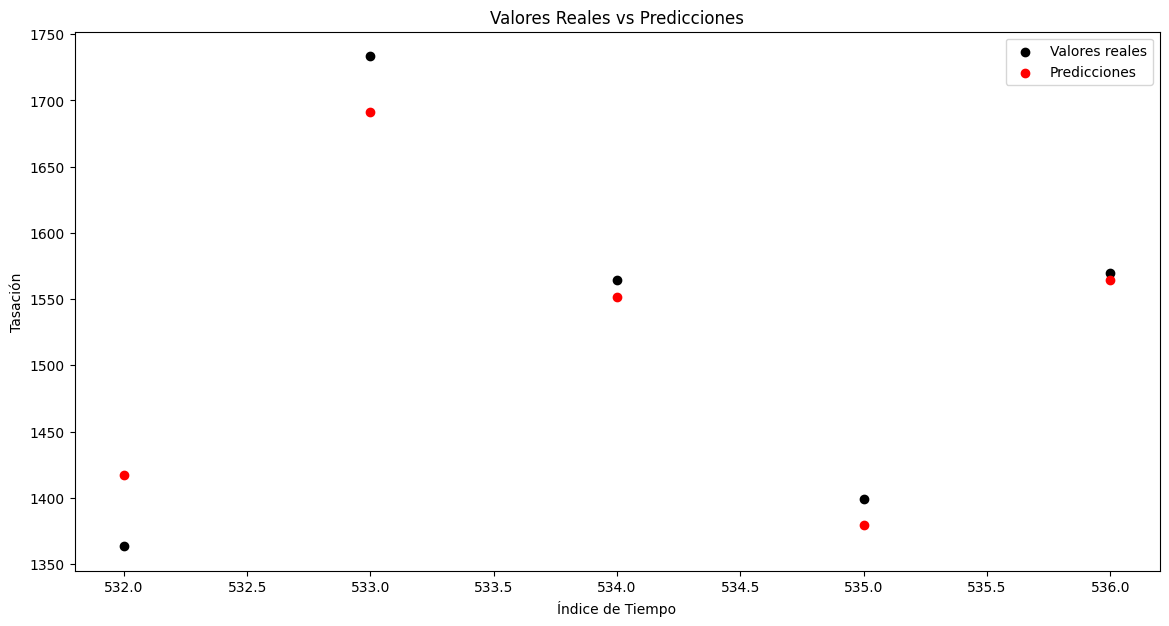
# Resultados

## Cumplimiento de los objetivos empresariales

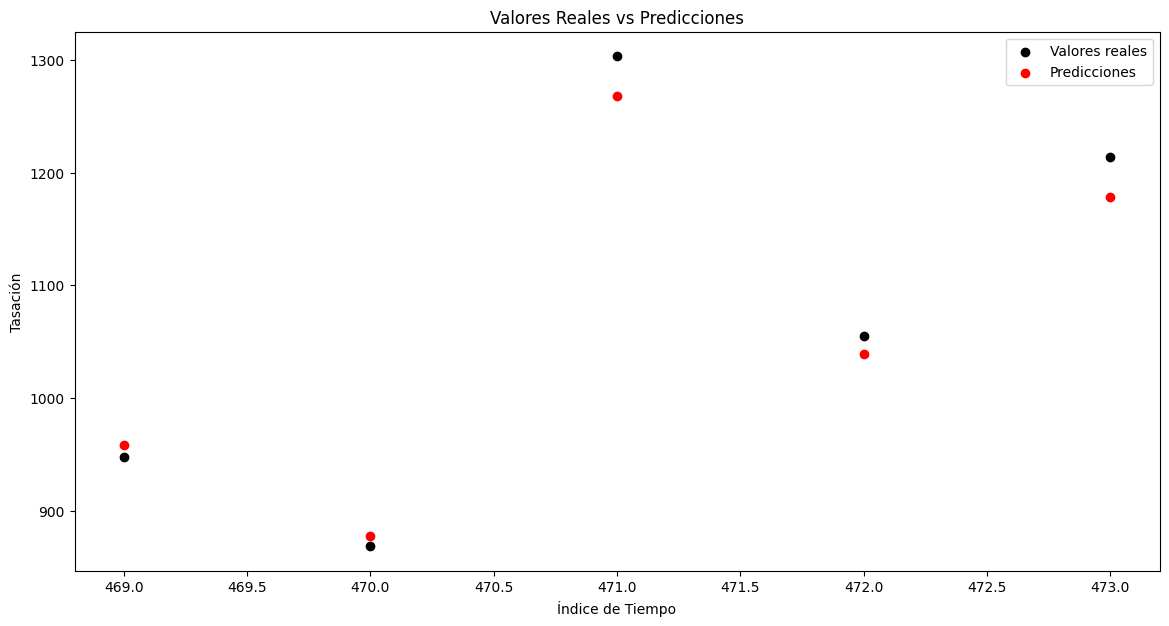
El modelo de red neuronal LSTM demuestra ser una herramienta útil gracias a su capacidad de predecir valores futuros, cumpliendo así con los objetivos empresariales definidos inicialmente. La empresa busca obtener una ventaja competitiva anticipando las tendencias de precios en el mercado inmobiliario. Este objetivo se logra de manera efectiva gracias a las capacidades predictivas del modelo LSTM, como se ha descrito en el apartado anterior. Las siguientes figuras muestran la comparación entre las predicciones del modelo y los datos reales, evidenciando su precisión.



Precios altos



Precios medios



Precios bajos

## Limitaciones y consideraciones futuras

Una de las principales limitaciones es la ingesta de datos futuros. Actualmente, los nuevos datos deben capturarse manualmente a partir de fuentes públicas, lo cual puede provocar retrasos significativos debido a la demora en la publicación de estos datos. Este retraso puede afectar la capacidad de la empresa para prever con suficiente antelación las fluctuaciones del mercado, comprometiendo así la ventaja competitiva que el modelo ofrece. Como se ha mencionado en apartados anteriores, se podría considerar la opción de utilizar otras fuentes de datos para abordar esta problemática.

Otra limitación significativa es la cantidad de datos disponibles. La escasez de datos ha obligado a agrupar las regiones en tres categorías para entrenar el modelo de red neuronal, lo que puede haber limitado la capacidad del modelo para capturar detalles más específicos del mercado inmobiliario. La falta de datos suficientes puede restringir la capacidad del modelo para generalizar y predecir con precisión en contextos más detallados o en mercados más segmentados. De manera similar a la limitación anterior, este problema podría resolverse mediante el uso de fuentes de datos alternativas.

# Despliegue en la arquitectura tecnológica de explotación

## Estrategia de despliegue del modelo

Para el despliegue del modelo de red neuronal LSTM en un entorno de explotación, se ha diseñado una solución a demanda para responder de manera efectiva a los objetivos del proyecto y a las necesidades operativas de la empresa. La estrategia de despliegue considera las siguientes fases y componentes clave:

La estrategia se basa en la predicción de los valores de tasación para cada comunidad, mediante los modelos LSTM entrenados previamente, cada vez que se dé una actualización de los datos del mercado inmobiliario. Este proceso implica las siguientes tareas:

* **Obtención de datos:** Captura de los datos inmobiliarios actualizados desde fuentes públicas y guardado en formato csv en local.
* **Preprocesamiento:** Aplicación de las mismas técnicas de preprocesamiento utilizadas durante el entrenamiento del modelo, como la normalización de datos.
* **Generación de predicciones:** Ejecución del modelo LSTM para prever las tendencias de precios futuros del mercado inmobiliario.
* **Generación de informes:** Creación de informes detallados con las predicciones realizadas, que incluirán análisis de los resultados, comparación con datos históricos y recomendaciones estratégicas para el negocio.

El **departamento de análisis de datos** será responsable de llevar a cabo estas tareas, asegurando que el informe se elabore de manera profesional y oportuna para facilitar la toma de decisiones.

## Comparación con otras opciones de despliegue

Se ha considerado la opción de empaquetar el modelo en un **Docker container,** pero se ha desestimado por la suma de c**omplejidad que conlleva.** Usar Docker para el despliegue requiere que el usuario tenga al menos **nociones básicas de Docker** para ejecutar el contenedor, lo que podría no ser ideal para todos los usuarios en el entorno empresarial. En su lugar, la solución a demanda se presenta como una opción más **eficiente** y **accesible** para los usuarios finales del modelo.

# Puesta en valor

En la Puesta de Valor se aborda la estrategia de implantación del modelo de predicción de precios inmobiliarios en los procesos funcionales de la empresa, así como el plan de monitoreo para asegurar la efectividad continua del modelo y su alineación con los objetivos empresariales, garantizando así que las predicciones se mantengan precisas y útiles para la toma de decisiones estratégicas.

La estrategia de implantación tiene como objetivo garantizar la puesta en marcha efectiva del modelo de predicción de precios inmobiliarios en los procesos funcionales de la empresa. La estrategia se basa en una solución a demanda, que se activa con la actualización trimestral de los datos del mercado inmobiliario. Este enfoque permite a la empresa obtener informes precisos y actualizados, facilitando decisiones estratégicas basadas en las tendencias futuras del mercado. Los pasos necesarios incluyen la captura y preprocesamiento de nuevos datos, la ejecución del modelo LSTM para generar predicciones, y la creación de informes detallados. Estos informes serán utilizados por los equipos de negocio para tomar decisiones anticipadas y maximizar las oportunidades del mercado.

Para asegurar la efectividad continua del modelo de predicción y su alineación con los objetivos empresariales, se implementa un plan de monitoreo. Este plan se basa en la supervisión regular del rendimiento del modelo mediante métricas clave como el MAE, RMSE y R², las cuales permiten medir el rendimiento del modelo y compararlo con los datos reales del mercado. Cualquier desviación significativa debe conllevar un ajuste del modelo para mantener su precisión.

# Conclusión

La implementación del modelo de red neuronal LSTM ha demostrado ser eficaz para anticipar las tendencias de precios en el mercado inmobiliario, proporcionando a la empresa una herramienta valiosa para la toma de decisiones estratégicas.

A pesar de ello, también se ha identificado ciertos desafíos que deben ser abordados en el futuro. Uno de los principales retos es la baja cantidad de datos disponibles y la frecuencia de su actualización. La ingesta manual de datos y la dependencia de fuentes públicas pueden generar retrasos y limitar la capacidad del modelo para prever con suficiente antelación las fluctuaciones del mercado. Además, la necesidad de agrupar datos de diferentes regiones para entrenar el modelo refleja una limitación en la granularidad de los datos disponibles.

# Contribución

Autor único: Germán Bosch Estévez