

hack(io)

PROYECTO FINAL

Modelo predictivo de la rentabilidad de la compra de una vivienda para alquiler utilizando Machine Learning.



Las Casas de David

Alumno: David Franco

Máster en Data Science e Inteligencia Artificial
Madrid, Febrero 2025

Resumen Ejecutivo

Este proyecto desarrolla una herramienta basada en Machine Learning que analiza la rentabilidad de la compra de viviendas para alquiler en Zaragoza. Se centra en ofrecer estimaciones sobre ingresos, costes y métricas relevantes, además de identificar las nuevas propiedades que salen al mercado. Así, los pequeños y medianos inversores pueden evaluar con mayor precisión su inversión y el estado de la vivienda, disminuyendo la incertidumbre y mejorando la toma de decisiones.

En primer lugar, se lleva a cabo un proceso ETL, en el que se recopilan datos de Idealista sobre precios de compra y alquiler, así como datos geoespaciales de Geoapify para asignar distritos. A continuación, se aplican procesos de limpieza y transformación de datos, asegurando su calidad y consistencia para el modelado predictivo.

Uno de los aspectos más destacados, es el uso de modelos de visión artificial para evaluar cocinas y baños, factores determinantes en la rentabilidad. Mediante YOLO, se identifican imágenes de estos espacios, mientras que BLIP genera descripciones, validando la identificación realizada. Posteriormente, Sonnet puntúa el estado de las estancias y estima su tamaño, proporcionando un análisis automatizado del estado de conservación.

Posteriormente, se entrena un modelo de regresión para estimar el precio de alquiler de cada vivienda. La evaluación de varios algoritmos concluyó que, el Random Forest alcanzó los mejores resultados, con un R^2 de 0.89 y un RMSE de 76 euros. Paralelamente, se calculan métricas financieras como el ROI y el ROCE, ofreciendo una visión global de la rentabilidad.

Los resultados obtenidos nos indican que, cumple su cometido de identificar automáticamente propiedades con alto potencial. Además, reduce el tiempo y esfuerzo en la evaluación manual. Al ofrecer predicciones precisas y análisis de imágenes, reduce la incertidumbre en la toma de decisiones, factor relevante inversores con menos liquidez.

Aunque la herramienta ha sido probada en Zaragoza, su diseño permite escalarlo a otras grandes ciudades, que presentan una mayor complejidad. En futuras versiones, se propone la integración de variables macroeconómicas, información sobre el barrio y datos sobre la financiación ofrecida en el mercado.

En conclusión, este proyecto ofrece una solución innovadora que combina datos, inteligencia artificial y análisis financiero en una sola herramienta. Su implementación permitirá a los usuarios optimizar sus estrategias de compra, reducir riesgos y maximizar la rentabilidad de sus inversiones en el sector inmobiliario español.

Índice

Resumen Ejecutivo	1
1. Descripción del caso de negocio	3
2. Pipeline ETL	4
1. Extracción:	4
2. Transformación:.....	4
3. Preprocesamiento:	4
4. Carga:.....	4
3. Resultados del modelo	5
4. Impacto de Negocio	6
5. Conclusiones y Próximos Pasos	9
Acceso al repositorio de GitHub y la Aplicación	10

1. Descripción del caso de negocio

- La inversión inmobiliaria ha sido, históricamente, una de las estrategias más seguras y rentables para aumentar el patrimonio, gracias a la estabilidad que ofrece este sector. En España, el mercado inmobiliario ha crecido notablemente en las grandes ciudades en la era post-COVID, impulsado por el aumento de la demanda de viviendas y el auge del alquiler de corta duración en plataformas como Airbnb. Estos factores han consolidado la inversión en vivienda como una opción atractiva para inversores que buscan rentabilidad sostenida y más segura.



Las Casas de David

- Este proyecto, orientado a pequeños y medianos inversores, tiene como objetivo analizar la rentabilidad de adquirir una vivienda destinada al alquiler en Zaragoza. La ciudad destaca como un mercado estratégico por su crecimiento económico, impulsado por su posición como nodo logístico clave y la reciente instalación de data centers, que generan empleo y aumentan la demanda de vivienda. Este desarrollo, sumado al acceso a precios de compra hasta un 40% más bajos que en grandes capitales, convierte a Zaragoza en una oportunidad atractiva para inversiones iniciales más asequibles.

- En los últimos años, el precio medio de la vivienda en Zaragoza ha crecido un 10%, mientras que la demanda de alquiler ha aumentado un 15% en la última década. Estos datos reflejan un mercado en expansión que ofrece tanto estabilidad como oportunidades de crecimiento. Además, la ciudad cuenta con una población diversificada de estudiantes, trabajadores y turistas, lo que fortalece su atractivo tanto para el alquiler tradicional como para el de corta duración. Elegir Zaragoza como punto de partida, facilita la implementación y validación de los modelos predictivos, que puedan ser escalables a otras ciudades más heterogéneas como Madrid, Barcelona o Valencia.

- Para alcanzar el objetivo del proyecto, se emplearán herramientas de ciencia de datos e inteligencia artificial, utilizando datos obtenidos de plataformas inmobiliarias. Asimismo, se desarrollará un modelo que evalúe las condiciones de las cocinas y baños, elementos clave para determinar la rentabilidad de una vivienda. Este enfoque integral proporcionará a los inversores información precisa y fundamentada para tomar decisiones estratégicas en el mercado inmobiliario español.

2. Pipeline ETL

El proceso ETL consta de cuatro fases principales:

1. Extracción:

- Se extraen datos de Idealista sobre precios de venta y alquiler, incluyendo características clave como tamaño, número de habitaciones y ubicación.
- Se extraen datos de Geoapify para asignar el distrito a cada vivienda y analizar la distribución geográfica de la inversión.

2. Transformación:

- Eliminación de duplicados y datos inconsistentes.
- Asignación de distritos a viviendas mediante el cruce de datos geoespaciales.
- Conversión de coordenadas geográficas a un formato estandarizado (CRS).
- Filtrado de viviendas que no forman parte del mercado objetivo del proyecto (por ejemplo, las que requieren una reforma integral).
- Eliminación de columnas por no aportar valor al análisis o contener información errónea o incompleta.
- Conversión de variables categóricas a formatos numéricos mediante *target encoding*.
- Normalización y escalado de variables para mejorar el rendimiento del modelo.

3. Preprocesamiento:

- Detección y eliminación de outliers mediante el método IFO.
- Sustitución de valores nulos en variables numéricas con la mediana y en variables categóricas con "ND" (No Disponible).

4. Carga:

- Almacenamiento de los datos transformados en MongoDB para consultas rápidas y análisis espacial.

Con la ETL finalizada, se procede a crear el modelo y presentar los datos a través de una aplicación de Streamlit.

3. Resultados del modelo

Se evaluaron múltiples modelos de regresión para predecir el precio del alquiler de viviendas en venta.

Los modelos probados incluyen:

- **Decision Tree:** Mostró un R^2 de ~ 0.49 , indicando una baja capacidad predictiva debido al sobreajuste.
- **Gradient Boosting:** Alcanzó un R^2 de ~ 0.68 , mejorando la precisión, pero sin superar otros modelos.
- **XGBoost:** Obtuvo un R^2 de ~ 0.74 , con un buen equilibrio entre precisión y velocidad de ejecución.
- **Random Forest:** Fue el modelo con mejor rendimiento, alcanzando un R^2 de 0.89 y un RMSE de 75.97 euros.

Las características más influyentes en el precio del alquiler son:

- *Tamaño de la vivienda:* Explicó el 35% de la variabilidad en el precio.
- *Distrito:* Aportó un 15% de la variabilidad.
- *Número de baños y la planta:* Representaron una influencia del 12% y 11%, respectivamente.

Para mejorar el análisis, se emplearon modelos de visión para identificar cocinas y baños en las viviendas en venta.

- **YOLO (Ultralytics):** Modelo de detección de objetos que identificó cocinas y baños en el 72% de las imágenes analizadas.
- **BLIP (Salesforce):** Generó descripciones textuales de las imágenes, con un 88% de coincidencia con YOLO en cocinas y un 94% en baños.
- **Anthropic Sonnet 3.5:** Evaluó el estado de conservación y estimó el tamaño de cada estancia identificada.

En ninguno de los casos se llevó a cabo *fine-tuning*.

4. Impacto de Negocio

El mercado inmobiliario en España ha experimentado cambios significativos en la última década, impulsados por diversos factores económicos, políticos y tecnológicos. Desde la crisis de 2008 hasta la recuperación progresiva en los años siguientes, y más recientemente, el impacto de la pandemia de COVID-19, la dinámica del sector ha cambiado considerablemente.

En la actualidad, el sector inmobiliario sigue siendo una de las principales opciones de inversión debido a:

- **Alta demanda de alquiler:** La dificultad de acceso a la vivienda por parte de los jóvenes y la escasez de viviendas asequibles han generado un incremento en la demanda de alquiler.
- **Aumento del interés de inversores extranjeros:** Ciudades como Madrid, Barcelona, Valencia y Málaga han atraído capital extranjero, incrementando la competitividad del sector.
- **Regulación y cambios legislativos:** La Ley de Vivienda ha introducido límites al precio del alquiler en algunas regiones, lo que ha generado incertidumbre en el mercado.
- **Impacto del turismo y alquileres a corto plazo:** Plataformas como Airbnb han transformado el mercado inmobiliario, generando una competencia directa con el alquiler tradicional.

El desarrollo de un modelo predictivo basado en Machine Learning y análisis de imagen supone una oportunidad para los pequeños y medianos inversores, que muchas veces carecen de herramientas avanzadas para tomar decisiones fundamentadas. En este sentido, se ofrecen soluciones innovadoras a problemas comunes en la inversión inmobiliaria:

1. Identificación de Oportunidades de Inversión con Mayor Rentabilidad

Uno de los principales desafíos para los inversores inmobiliarios es identificar propiedades con el mayor potencial de rentabilidad. Tradicionalmente, los inversores han dependido de análisis subjetivos o métodos rudimentarios para evaluar una propiedad. Con la implementación de un modelo basado en datos:

- Se eliminan sesgos subjetivos en la toma de decisiones.
- Se identifican patrones en los precios de compra y alquiler que permiten detectar oportunidades con mejor rendimiento.
- Se optimiza el proceso de selección de propiedades, reduciendo el tiempo y esfuerzo en la búsqueda manual.

2. Automatización del Análisis de Propiedades con Machine Learning

El uso de modelos de regresión y visión por computadora introduce un nivel de automatización en la evaluación de propiedades que no se había explotado en el sector de inversión inmobiliaria a pequeña escala. Algunas ventajas clave incluyen:

- **Precisión en la predicción del precio del alquiler:** El modelo Random Forest utilizado en este proyecto ha demostrado una precisión del 89% en la estimación del alquiler, permitiendo a los inversores calcular de manera más fiable sus ingresos futuros.
- **Análisis del estado de las viviendas:** La identificación de cocinas y baños mediante modelos de visión artificial como YOLO y BLIP permite detectar propiedades con alto potencial o aquellas que requieren reformas, lo que impacta directamente en la rentabilidad de la inversión.
- **Automatización del cálculo de métricas financieras:** La inclusión de indicadores como ROI, ROCE y COCR permite que los inversores evalúen fácilmente los riesgos y beneficios de cada propiedad sin necesidad de cálculos manuales.

3. Reducción de la Incertidumbre en la Evaluación de Viviendas

Uno de los mayores desafíos en la inversión inmobiliaria es la incertidumbre. Los inversores deben considerar múltiples factores, desde el estado físico de la propiedad hasta la estabilidad del mercado. Este modelo reduce significativamente esa incertidumbre al proporcionar:

- *Evaluaciones objetivas de propiedades:* Los modelos de visión artificial identifican cocinas y baños en mal estado, permitiendo estimar costos de renovación antes de la compra.
- *Predicciones más precisas de ingresos por alquiler:* Con un error medio de 75 euros en la estimación del alquiler, los inversores pueden calcular con confianza los ingresos mensuales esperados.
- *Análisis de mercado en tiempo real:* Al integrar datos de Idealista, el sistema permite comparar las tendencias del mercado y anticipar fluctuaciones en los precios.

Desde una perspectiva financiera, el modelo también tiene implicaciones relevantes en términos de:

- **Planificación fiscal:** Conocer la rentabilidad proyectada de una inversión permite calcular de antemano los impuestos y costes asociados, como el IBI o el IRPF.
- **Optimización de préstamos hipotecarios:** Al predecir con mayor precisión los ingresos por alquiler, se pueden negociar mejores condiciones con las entidades bancarias, aumentando el apalancamiento financiero.
- **Impacto en el cash flow:** La automatización del análisis de ingresos y costos ayuda a prever el flujo de caja mensual y evaluar si una inversión es sostenible en el tiempo.

Aunque es posible obtener métricas financieras de forma manual, tenerlas calculadas automáticamente y representadas de una forma visual amigable, es el objetivo del proyecto.

Por otro lado, si bien el modelo ha sido probado en Zaragoza, su estructura permite una fácil escalabilidad a otras ciudades de España. La expansión a mercados más dinámicos como Madrid, Barcelona y Valencia, que presentan una mayor variabilidad en los precios y demanda de alquiler, haría que el modelo fuera aún más valioso.

Este proyecto representa una oportunidad para acercar la inversión inmobiliaria a los pequeños y medianos inversores. Mediante el uso de inteligencia artificial y modelos de regresión, el proceso de toma de decisiones pasa de ser una estimación subjetiva a un modelo basado en datos cuantificables y objetivos. El impacto se refleja entonces en:

1. Reducción del riesgo en la inversión inmobiliaria.
2. Aumento en la precisión de las estimaciones de rentabilidad.
3. Democratización del acceso a herramientas de análisis de mercado.

En un contexto de un mercado inmobiliario dinámico, contar con herramientas basadas en ciencia de datos y modelos predictivos puede ser un factor diferenciador clave para cualquier inversor que busque maximizar su rentabilidad y minimizar los riesgos.

5. Conclusiones y Próximos Pasos

- El estudio confirma la viabilidad de utilizar modelos predictivos y visión por ordenador para evaluar inversiones inmobiliarias. Las herramientas desarrolladas permiten realizar un análisis detallados y automatizados, reduciendo el tiempo y el esfuerzo requerido por inversores objetivo. Como próximos pasos, se propone:

1. Mejorar la precisión de la evaluación del estado de las viviendas a través de, por ejemplo, *fine-tuning* de los modelos de reconocimiento de imágenes. Hacer que el análisis de la estancia pueda realizarse a partir de múltiples imágenes, y no una como se hace actualmente.
2. Explorar la posibilidad de predecir el coste de las reformas en cocinas y baños a partir de las puntuaciones obtenidas y la estimación de los metros cuadrados. En ese sentido, identificar las viviendas con plano y utilizar OCR para extraer el tamaño preciso de la cocina y los baños.
3. Extraer el gasto de comunidad de las viviendas que lo mencionen en la descripción.
4. Explorar otras modalidades de alquiler como el vacacional y por habitaciones, utilizando datos de plataformas como AirBnB.
5. Valorar integrar nuevas fuentes de datos de viviendas, como Redpiso.
6. Ampliar la cobertura geográfica de la aplicación a otras ciudades.
7. Utilizar un sistema de grillas como H3 para mejorar la predicción del alquiler.
8. Explorar el impacto de variables macroeconómicas en la rentabilidad, y añadirlas como elementos adicionales de decisión. Además, la integración de datos sobre tasas de interés, inflación, crecimiento del PIB y otros, puede ayudar a identificar tendencias en el mercado inmobiliario.
9. Añadir información adicional sobre el entorno de la vivienda, con nuevas fuentes de información. Por ejemplo, las características del barrio, seguridad, servicios, ocio y transportes, que puedan servir como elementos adicionales de decisión cuando el inversor no está familiarizado con la ciudad.
10. Incluir en la aplicación información sobre financiación, obtenida de distintas entidades financieras, como un elemento más de comparación para el usuario.
11. Crear un sistema de notificaciones que avise al usuario vía correo electrónico cuando surja una vivienda con las características deseadas.
12. Desarrollar una funcionalidad para que el usuario reciba un informe semanal con las oportunidades más interesantes de inversión.
13. Añadir a la aplicación de Streamlit una página con gráficas a modo de resumen de los filtros del usuario y la situación del mercado, por ejemplo, por barrio o distrito.
14. Desplegar el *backend* en una plataforma como Azure o AWS, de modo que, la aplicación pueda obtener datos del mercado en tiempo real.

Acceso al repositorio de GitHub y la Aplicación

Se puede acceder al código del proyecto y directamente a la aplicación a través de los siguientes enlaces:

- **Repositorio:** <https://github.com/davfranco1/Proyecto-Rentabilidad-Viviendas/>
- **Aplicación:** <https://lascasasdedavid.streamlit.app/>