



HomeFinder: Encuentra tu nuevo hogar

Autores: Rubén Martín Blázquez, David Monroy González
Madrid, 2023

1. Beneficiarios del proyecto

- *Valor y Utilidad:*

Este proyecto está diseñado para convertirse en una herramienta útil para aquellas personas que estén buscando comprar una vivienda, de segunda mano, en Madrid.

La utilidad principal del mismo radica en proporcionar una evaluación detallada y objetiva del valor de las propiedades, analizando variables como el número de habitaciones, baños, superficie y otros aspectos clave, permitiendo a los compradores identificar ofertas que se adecuen perfectamente a sus preferencias y necesidades.

- *Resolución de problemas:*

El proyecto aborda los desafíos que enfrenta cada persona que desea adquirir una vivienda al intentar determinar el valor real de una vivienda de segunda mano en el mercado. Ofrece una solución que facilita la toma de decisiones al ser capaces de identificar ofertas justas y ajustadas a las expectativas de cada uno.

- *Audiencia objetivo:*

Esta herramienta está dirigida a aquellas personas que desean tener una visión clara y objetiva del valor de mercado de las propiedades en las que puedan estar interesados. Proporciona una base sólida para tomar decisiones financieras inteligentes de cara a la compra de una vivienda de segunda mano.

- *Objetivo del proyecto:*

Desarrollar una herramienta precisa que permita a los compradores evaluar el valor real de las viviendas en base a sus características clave. Al utilizar análisis detallados y datos históricos, este proyecto busca ofrecer una herramienta de confianza permitiendo una toma de decisiones fundamentadas en la compra de propiedades de segunda mano.

Pero, ¿por qué crear dicha herramienta?

El precio de las viviendas en Madrid ha ido subiendo cada año más y más, haciendo muy difícil, y en ocasiones casi imposible, poder llegar siquiera a plantearse la compra de una propiedad propia. La creación de esta herramienta surge de esta misma necesidad, buscando ofrecer una solución que genere confianza y seguridad a todos sus usuarios, ya que hay una falta de herramientas precisas y de fácil acceso en el mercado inmobiliario actual.

2. Ejecución del proyecto:

- Proceso de desarrollo:

Para llevar a cabo este proyecto lo primero que se hizo fue investigar la página de ‘Idealista’ pero dado que tenía una API propia de pago, se descartó la idea de realizar un scraping de dicha web, por lo que seguimos investigando más webs, llegando así hasta la página de ‘Yaencontre’, la cual permitía llevar a cabo un scraping de toda la información que había publicada en el apartado de viviendas de segunda mano en Madrid.

Tras realizar el scrapping de la web, se consiguieron un total de 9.995 anuncios, de los cuales, se observó como había más de 2.000 anuncios duplicados y alguno que no tenía ningún dato, por lo que se eliminaron dichos anuncios para quedarnos con un dataset sin datos duplicados.

Una vez realizado todo ese proceso, comenzamos a limpiar el dataset de forma que fuéramos capaces de analizar todo el dataset en conjunto de manera más simple, siendo así más fácil de entender qué es lo que necesitábamos de cara a realizar el EDA (Análisis Exploratorio de Datos).

Durante el EDA pudimos observar la distribución de las variables, observando como algunas de ellas, como el número de habitaciones, baños o superficie, tenían una importancia mayor que el resto en relación al precio.

Por último, realizamos una imputación de los valores faltantes para llevar a cabo el último paso restante, entrenar nuestro modelo para comprobar si nuestra hipótesis podría llevarse a cabo.

- Metodologías:

Durante el proceso de realización y desarrollo del proyecto se han llevado a cabo varias de las lecciones aprendidas a lo largo del curso.

- 1. Comenzamos usando Web Scraping para sacar toda la información de los 9.995 anuncios publicados en la página ‘Yaencontre’.
- 2. Una vez logramos sacar toda la información y tuvimos montado nuestro Dataframe, procedimos con una limpieza previa al EDA usando la librería de Pandas.
- 3. Llegados a este punto, comenzamos con el EDA, en el que combinamos la librería de Pandas con otras como Seaborn o Matplotlib.
- 4. Por último, una vez finalizado el EDA, comenzamos con los modelos de entrenamiento, en el que usamos 3 tipos diferentes: Regresión Lineal, XG Boost y Random Forest.

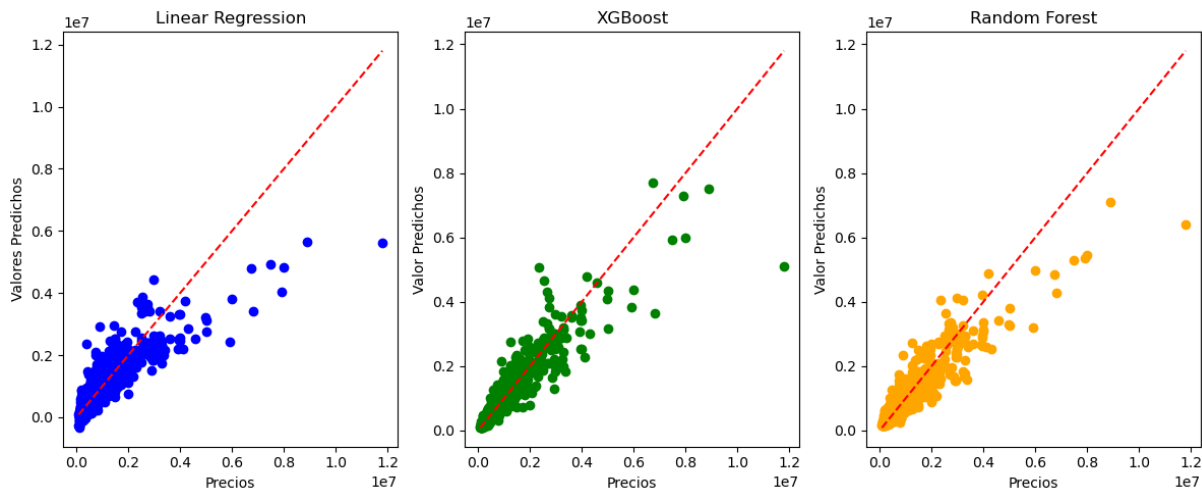
3. Impacto y resultado:

Una vez llevamos a cabo los modelos, nos dimos cuenta de que el margen de error obtenido al usar todos los anuncios era aún muy alto, por lo que decidimos reducir nuestros datos en función del precio, eliminando los outliers y dejando como límite de precio 950.000€ como precio máximo de las viviendas, consiguiendo así unos resultados más precisos y con menos error.

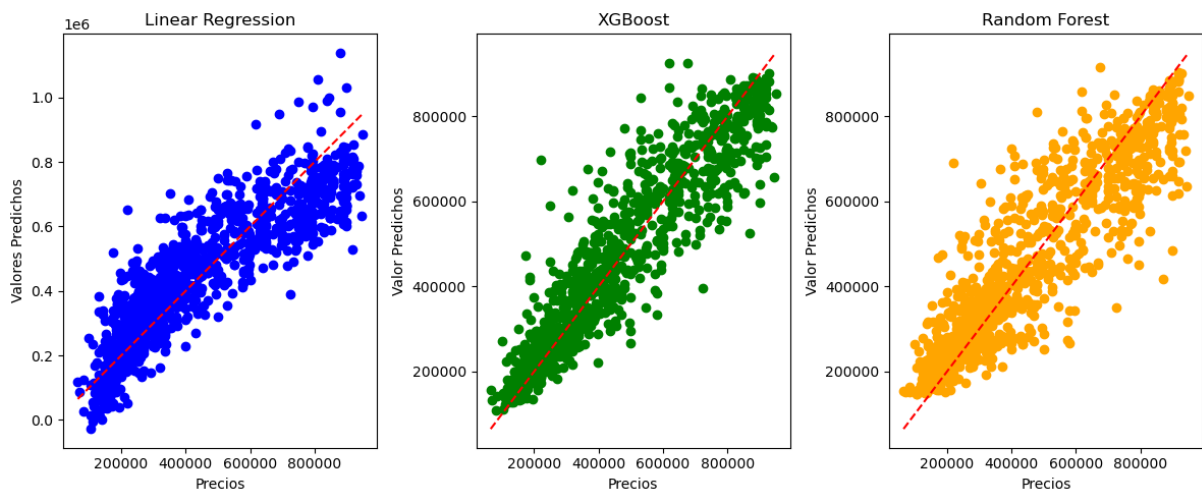
- Resultados de los modelos:

Modelo	Modelos con Outliers			Modelos sin Outliers		
	Error (€)	Score	Mape	Error	Score	Mape
Regresión Lineal	443.622,24	0,761	39%	11.861,56	0,789	24%
XGBoost	343.563,63	0,857	19%	85.828,93	0,876	16%
Random Forest	356.518,08	0,846	26%	103.423,69	0,82	22%

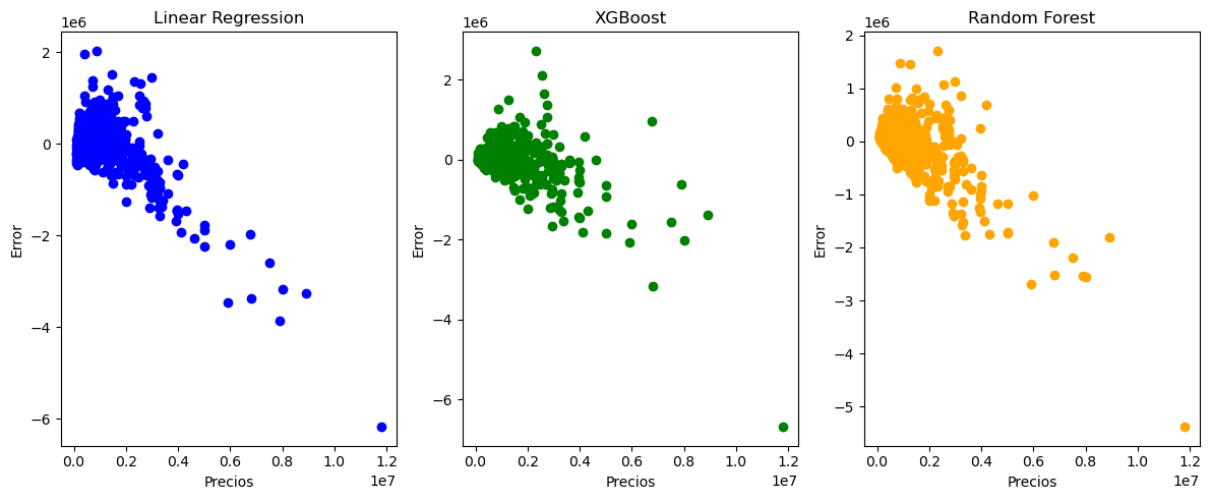
- Resultados predichos vs reales con Outliers:



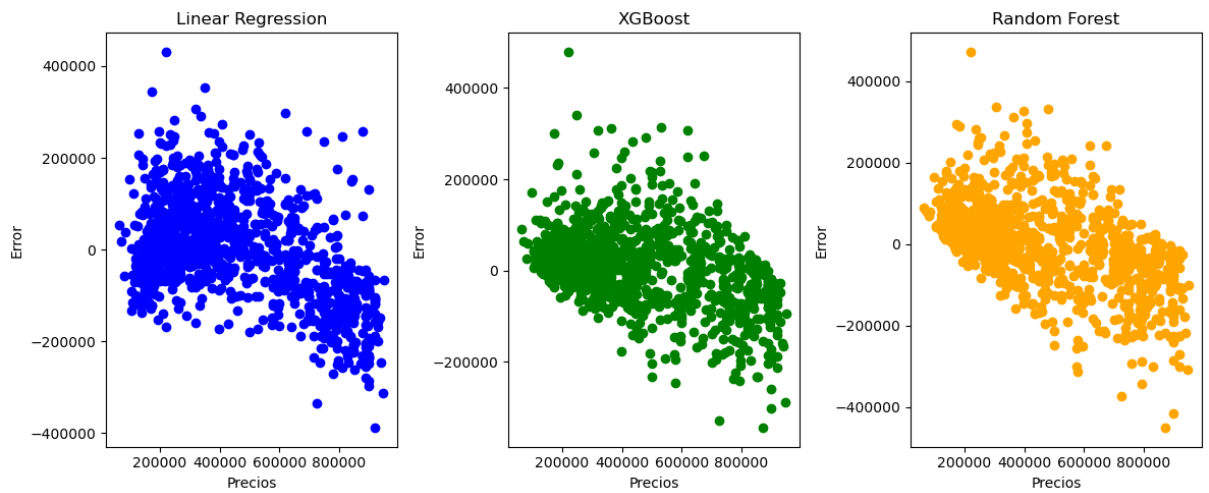
- Resultados predichos vs reales sin Outliers:



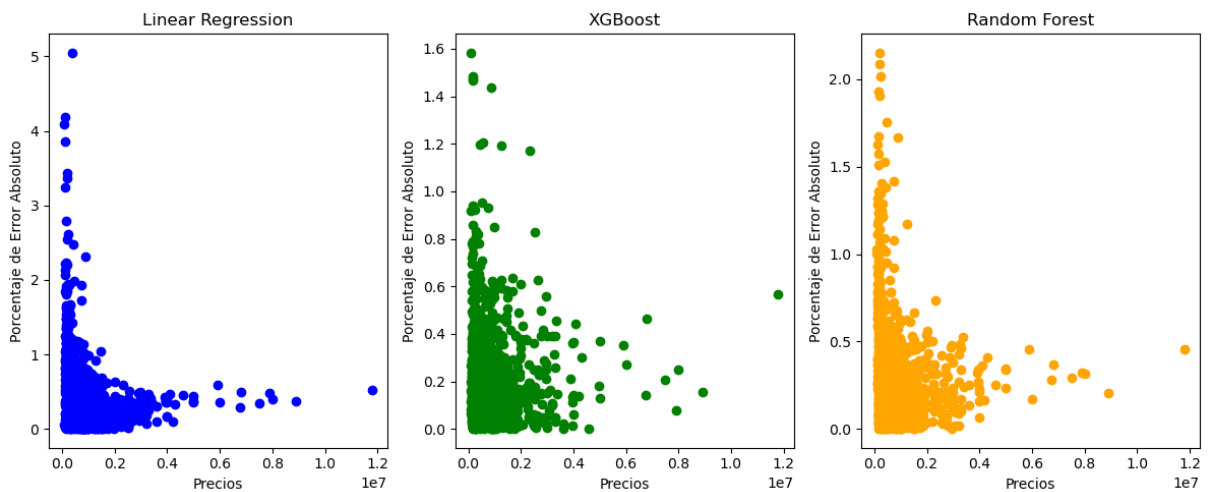
- **Error con Outliers:**



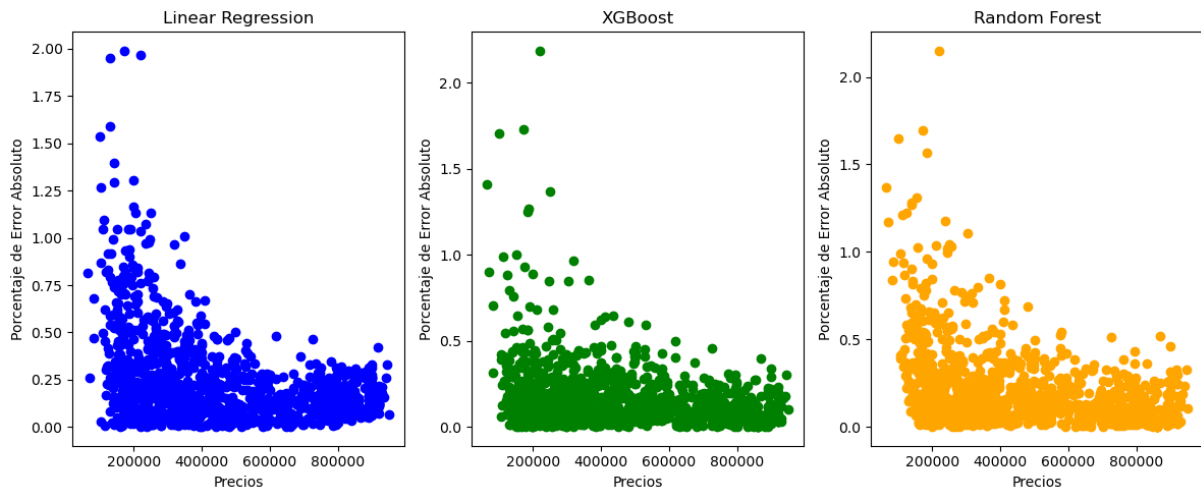
- **Error sin Outliers:**



- **Mape con Outliers:**



- ***Mape sin Outliers:***



4. Líneas de continuidad:

De continuar con este proyecto el camino a seguir sería conseguir una mayor cantidad de datos ya que consideramos que, con apenas 7.400 anuncios, no se llega a alcanzar una cantidad de datos tan grande como para lograr un modelo más preciso.

Sin embargo, tras finalizar el proyecto, somos optimistas con el futuro de esta herramienta, ya que pensamos que si con una cantidad de datos tan reducida como era la obtenida al realizar el scraping de 'Yaencontre', hemos logrado unas métricas que, a priori, consideramos muy buenas, si se consiguiera mayor cantidad de datos, se podría realizar un análisis más profundo de cara al posterior entrenamiento de los modelos para poder producir esta herramienta.

5. Conclusiones

En relación al proceso de desarrollo del proyecto si bien costó encontrar con una idea interesante y atractiva que se pudiera enfocar a un aspecto más empresarial, el tema de la venta de viviendas fue perfecto para poder probar una hipótesis en la que nos planteamos si sería posible crear una herramienta que permitiera al usuario buscar las mejores ofertas que se adaptaran a sus necesidades y expectativas. Además nos ha permitido poner a pruebas varias de las lecciones aprendidas durante el curso, como el Web Scraping, el uso de librerías como Pandas, Numpy, Seaborn o Matplotlib, probar entrenamiento de diferentes modelos y usar el Procesamiento del Lenguaje Natural.