



Factores determinantes en el precio de las propiedades EE. UU.

Con enfoque hacia Inversionistas y
Agentes Inmobiliarios

DAVID MAYA LONDOÑO 2023



Problemática

- El mercado inmobiliario de Estados Unidos es altamente complejo y requiere considerar múltiples factores al momento de tasar una propiedad.
- Actualmente no existe una fórmula matemática ni un método analítico que permita obtener un precio de vivienda de una manera que maximice el costo de oportunidad.
- La especulación juega un papel crucial y, por ende, puede llevar a afectar negativamente una inversión inmobiliaria

Preguntas y Objetivos

¿Es posible obtener un modelo de machine learning que permita responder estas preguntas?



¿Afecta el área del lote el precio de una vivienda construida para venta?



¿Cuáles son los factores más determinantes que afectan el precio de compra de una vivienda?



¿Es posible obtener un precio de una propiedad acorde al histórico de mercado y su crecimiento orgánico?

Conjunto de datos

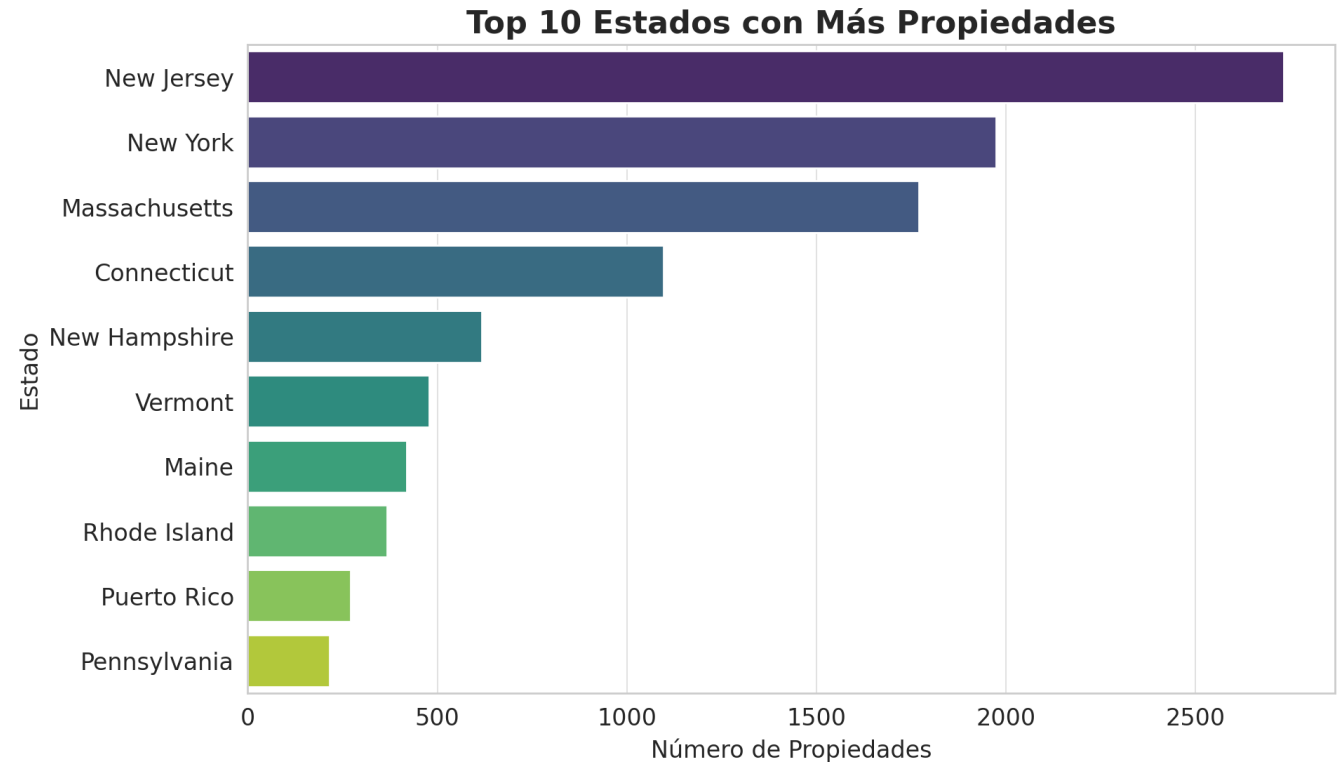
Conjunto de datos con los precios de las viviendas de los últimos 20 años

- **Variables numéricas:**

- HABITACIONES: Número de habitaciones en la propiedad.
- WC: Número de baños.
- AREA_CASA: Área de la casa en pies cuadrados o alguna otra unidad de medida (no especificada en los datos).
- AREA_LOTE: Área de la propiedad en yardas cuadradas (no especificada)
- PRECIO: Precio de la propiedad.

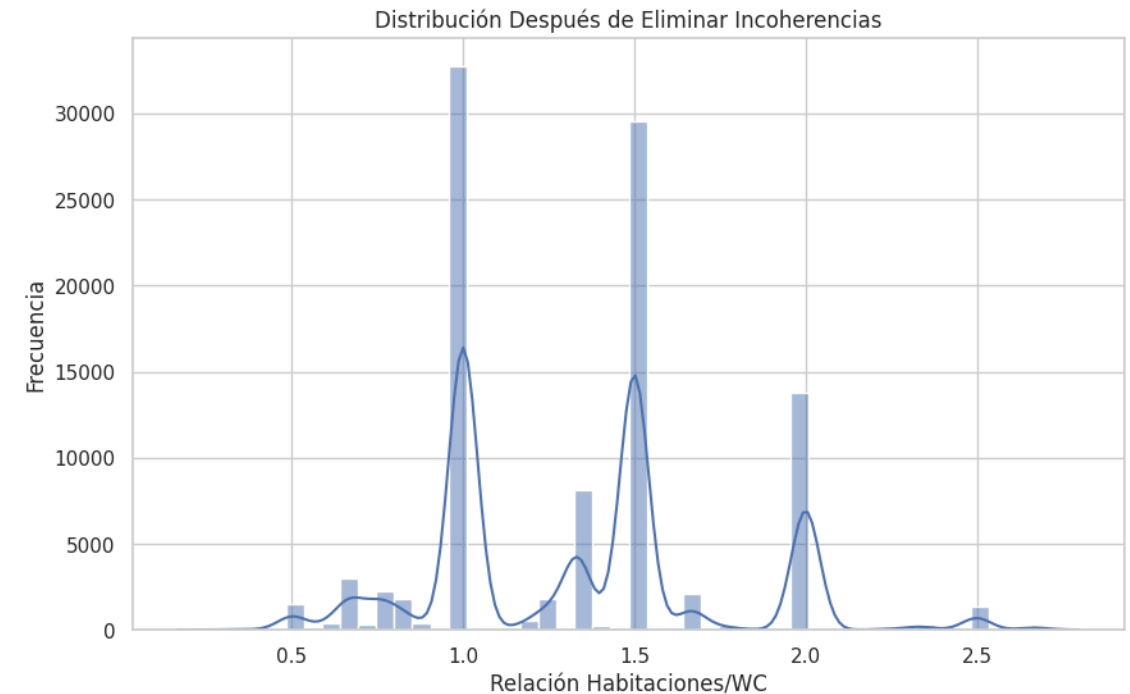
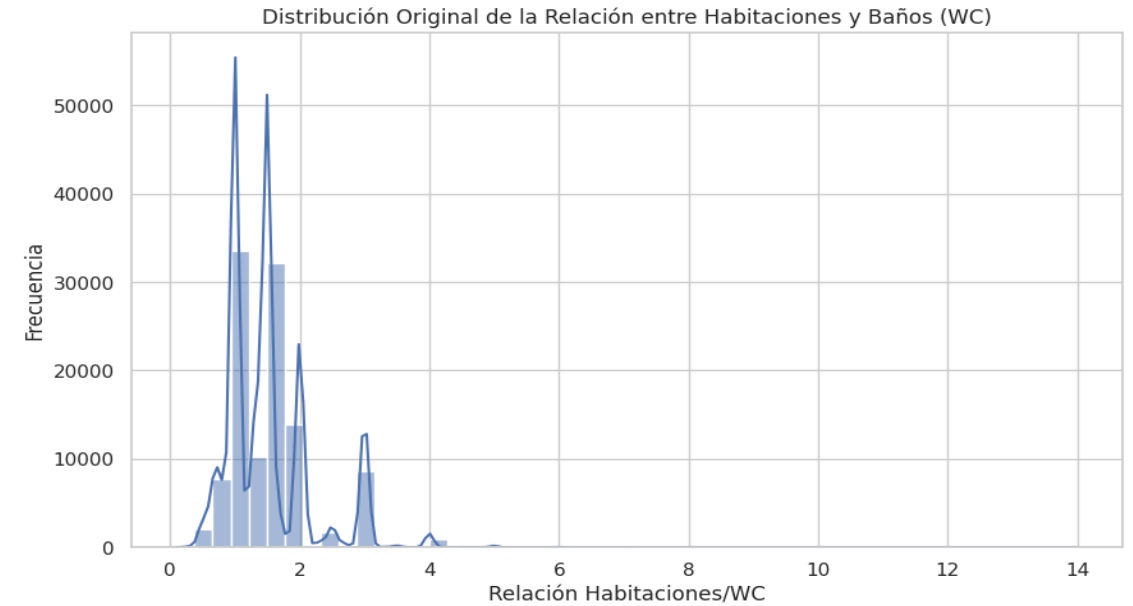
- **Variables Categóricas:**

- CIUDAD: Ciudad en la que se encuentra la propiedad.
- ESTADO: Estado en el que se encuentra la propiedad.
- ZIP_CODE: Código postal.



Análisis de Outliers

- Habitaciones: Outliers en propiedades con más de 10 habitaciones, posiblemente representando casas grandes o errores de datos.
- Baños: Outliers en propiedades con 7 o más baños, indicando posibles propiedades de lujo o comerciales.
- Área de la Casa: Varias propiedades con áreas excepcionales, reflejando mansiones, propiedades de lujo o inconsistencias en los datos.
- Precio: Outliers en precios altos, representando propiedades de lujo, ubicaciones privilegiadas o factores económicos y de mercado.
- Relaciones incoherentes entre número de habitaciones y baños



Exploración inicial de los datos

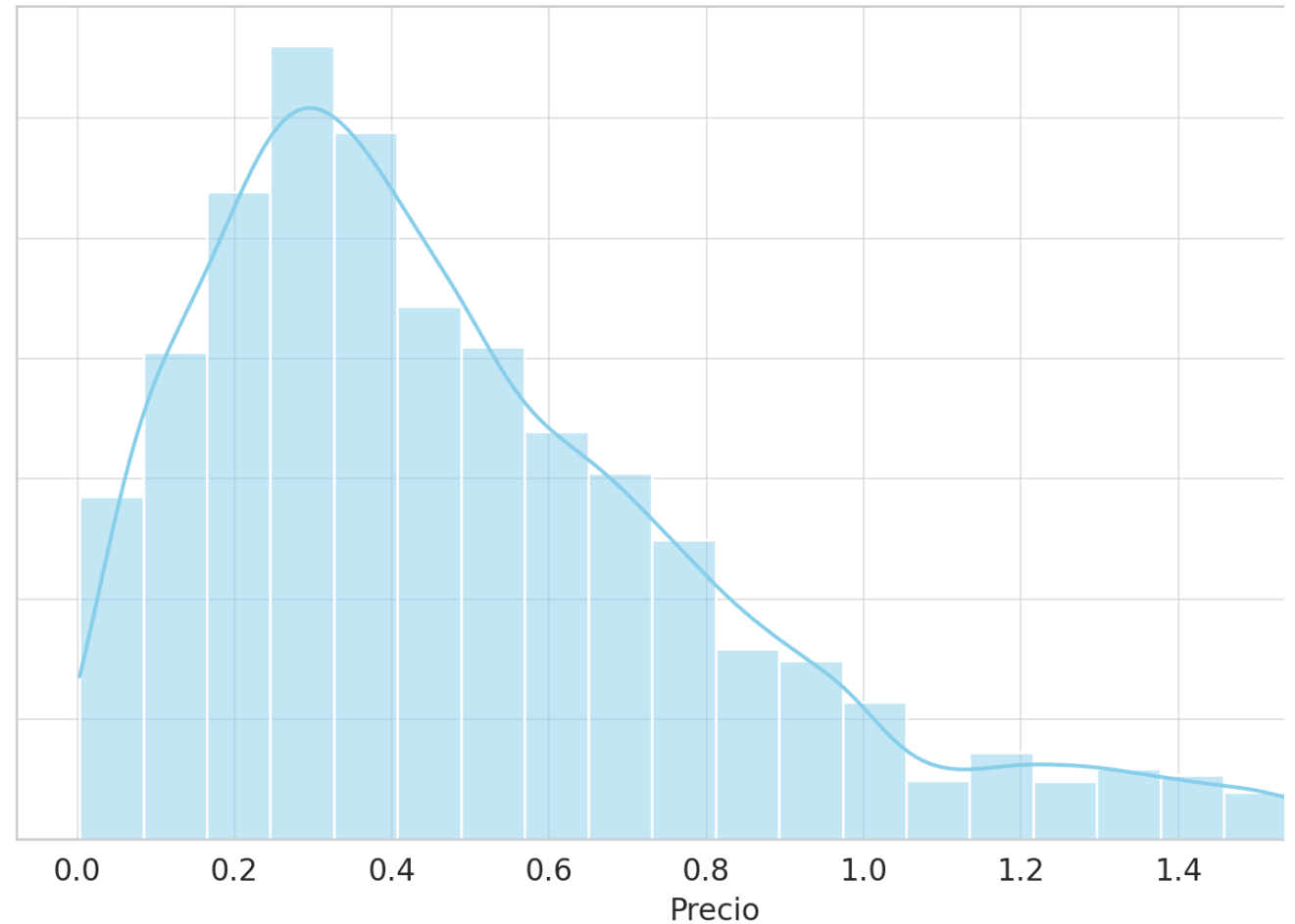
Diversidad Geográfica: Las propiedades están distribuidas en 1518 ciudades diferentes en 12 estados

Precios: Desde tan solo \$3,000 hasta \$1,620,000, lo que sugiere que el conjunto de datos incluye propiedades de varios segmentos del mercado.

Tamaño y Comodidades: La mayoría de las propiedades tienen entre 2 y 4 habitaciones y 1 y 3 baños,

Oportunidades de Segmentación: La variedad de ciudades, estados, y características de las casas como el número de habitaciones y baños, ofrece oportunidades para segmentar el mercado y dirigir estrategias de marketing o ventas específicas.

Distribución del Precio de las Propiedades



Factores determinantes

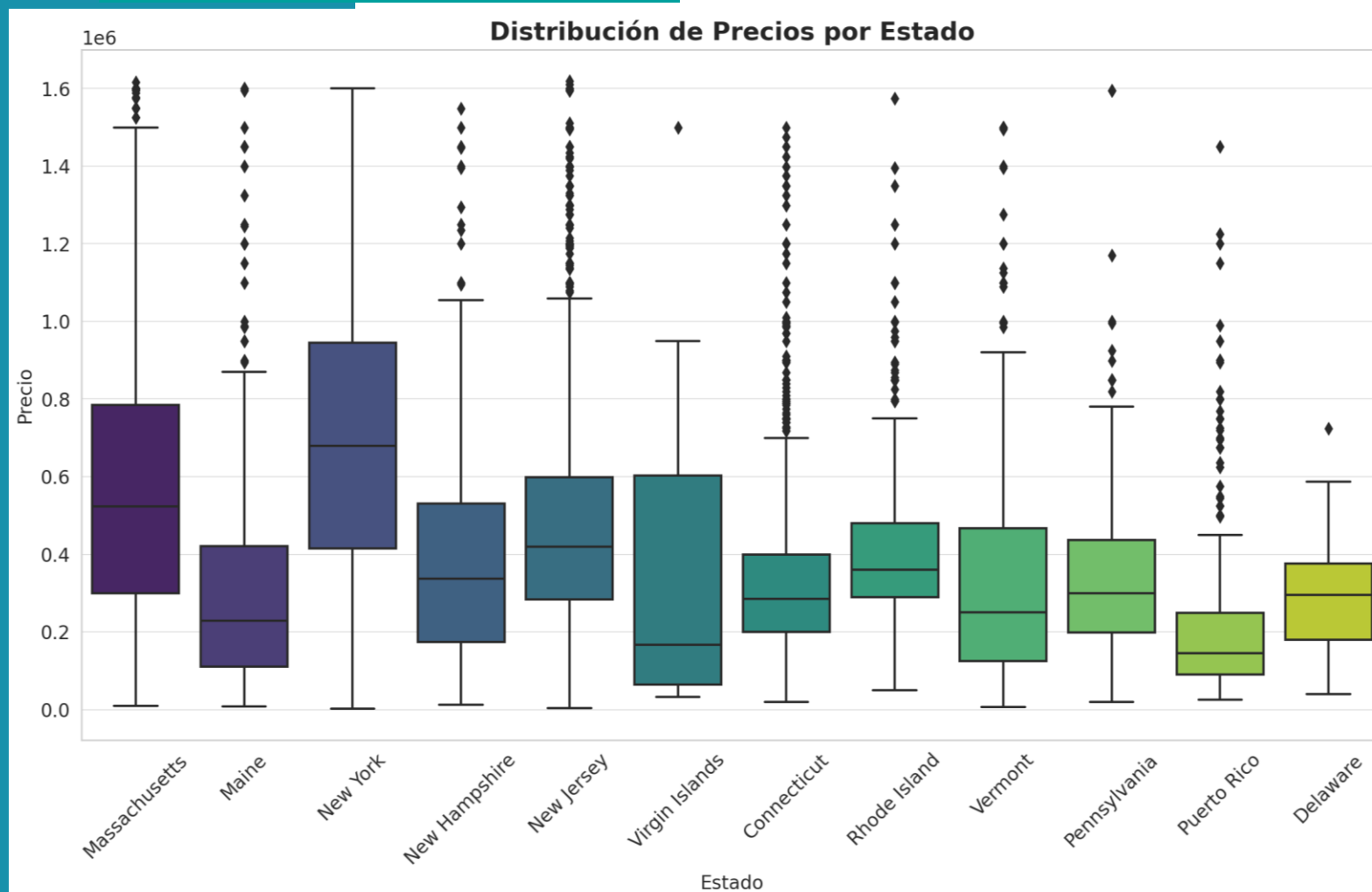
Estado vs Precio:

Variabilidad entre Estados: Precios medios varían ampliamente entre estados, algunos mucho más altos o bajos.

Diferencias Significativas: Prueba ANOVA confirma diferencias estadísticas en precios medios entre estados.

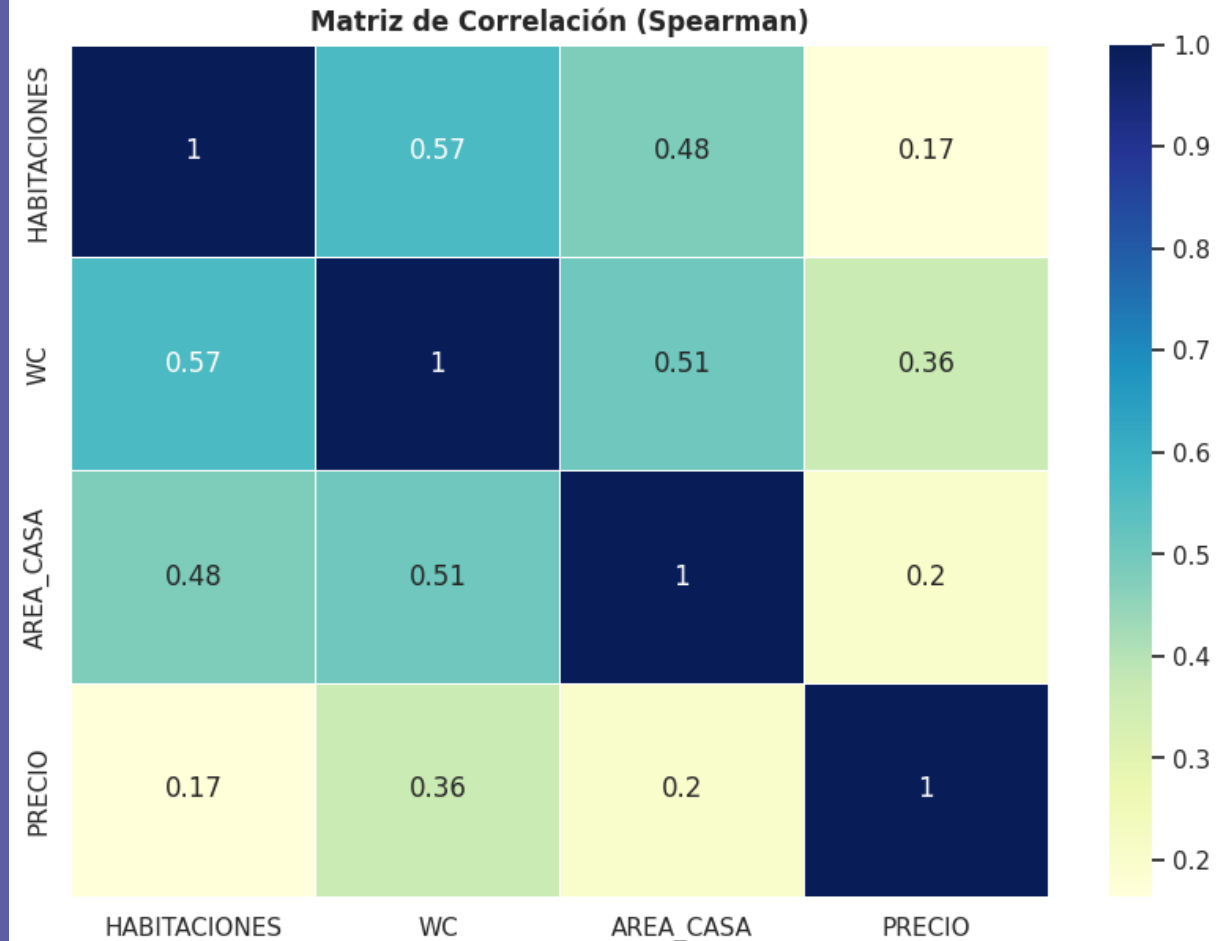
Importancia del Estado: Ubicación impacta el precio. Adaptar estrategias de precio, marketing e inversión por estados.

Guía Estratégica: Información detallada sobre cómo el estado influye en precios, orientando decisiones estratégicas y operativas.



Factores Determinantes

- Las correlaciones de Spearman revelan relaciones no lineales entre las variables, especialmente para el área de la casa en relación con el precio.
- La prueba de hipótesis nula valida que la correlación de Spearman para el área de la casa y el precio no es simplemente una relación lineal .
- La ubicación (número de habitaciones y baños) y el tamaño (área de la casa) juegan roles en la determinación del precio, pero las relaciones no son estrictamente lineales.



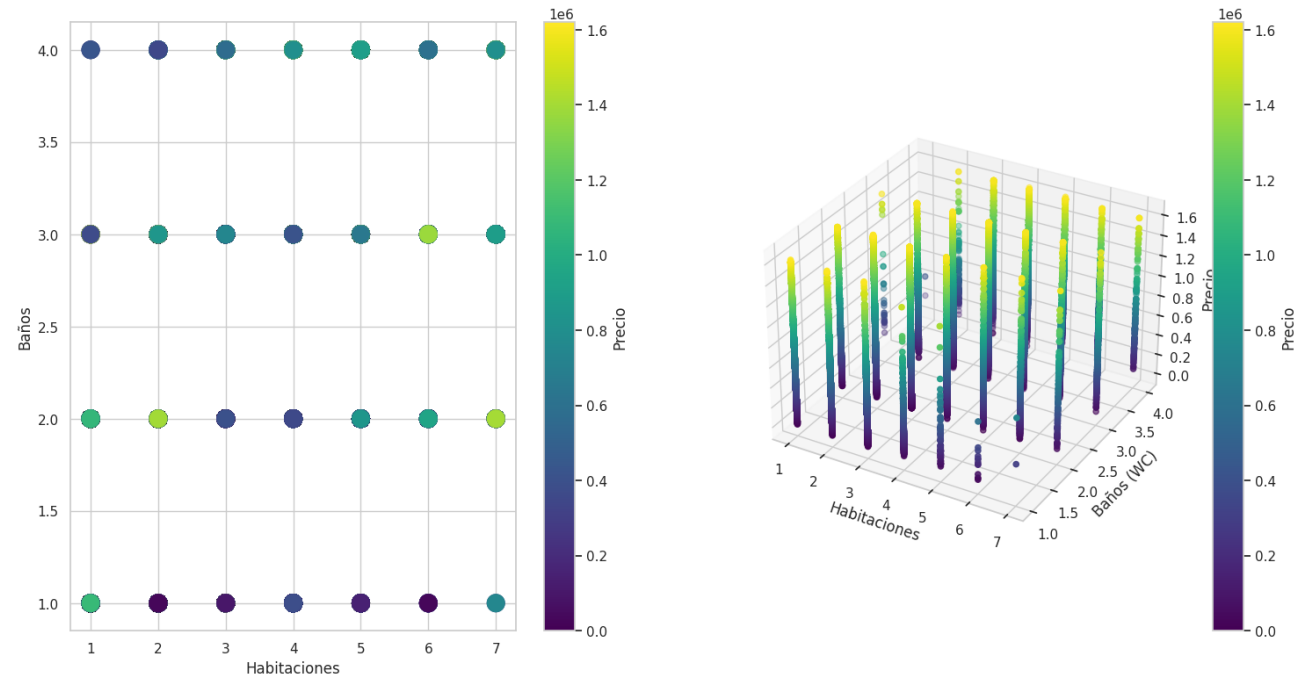
Factores Determinantes

La relación entre el Precio, las Habitaciones y los Baños es una dimensión vital en la evaluación de propiedades inmobiliarias.

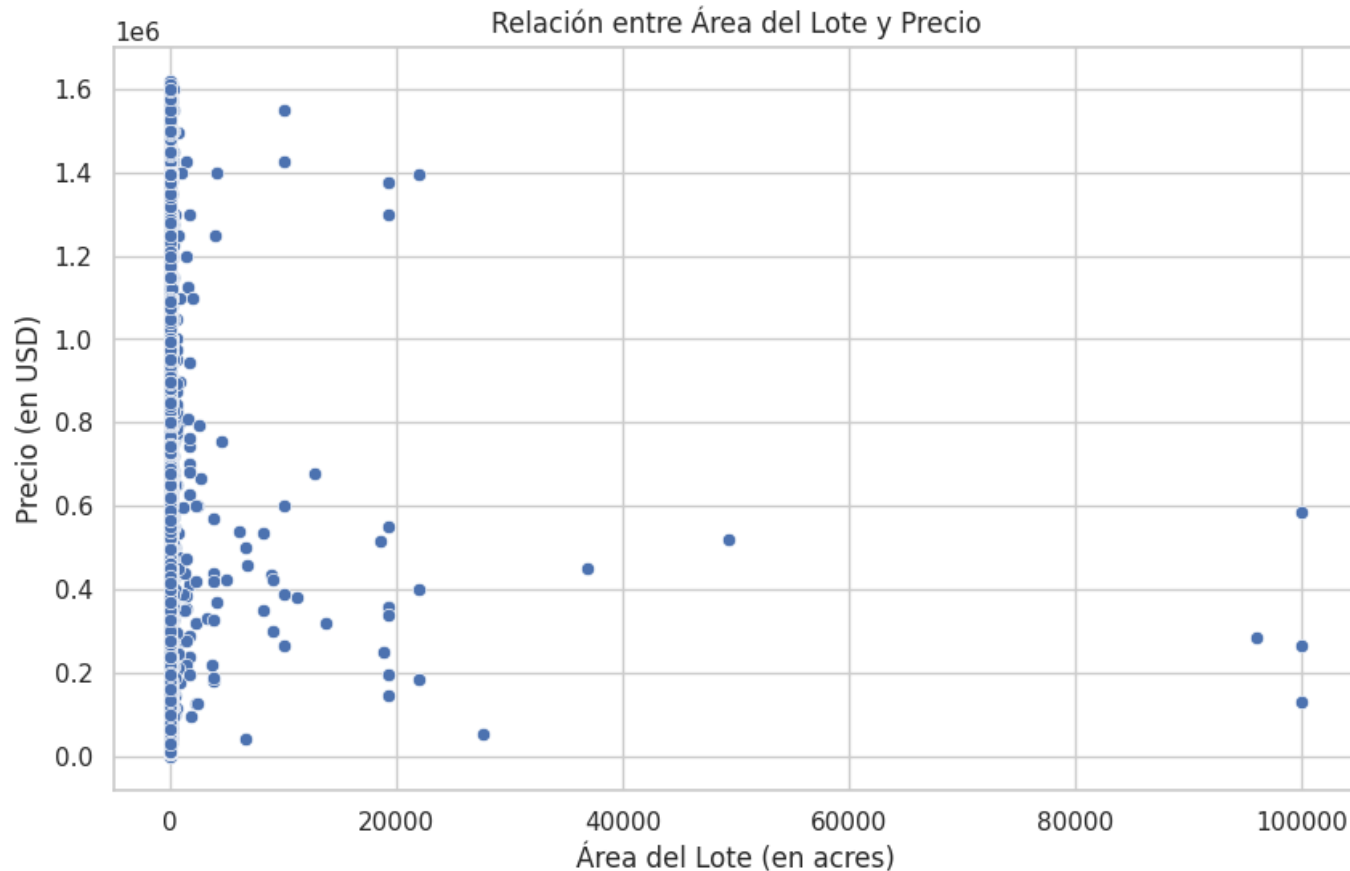
Los gráficos destacan que, si bien hay una correlación positiva general, la relación es multifacética y puede estar influenciada por otros factores y características.

Estos insights pueden ser útiles para agentes inmobiliarios, inversionistas y compradores al evaluar el valor de una propiedad y desarrollar estrategias de precios y marketing ajustadas al mercado actual.

Relación entre el Precio, las Habitaciones y los Baños



Área del Lote vs Precio



- El coeficiente de correlación de Pearson es aproximadamente -0.002, lo cual indica una correlación muy débil y cercana a cero entre el área del lote y el precio.
- En otras palabras, *no hay evidencia de que el área del lote tenga un efecto significativo en el precio*. Además, el valor R² es 0.000, lo que indica que el modelo no explica prácticamente ninguna de la variación en el precio.

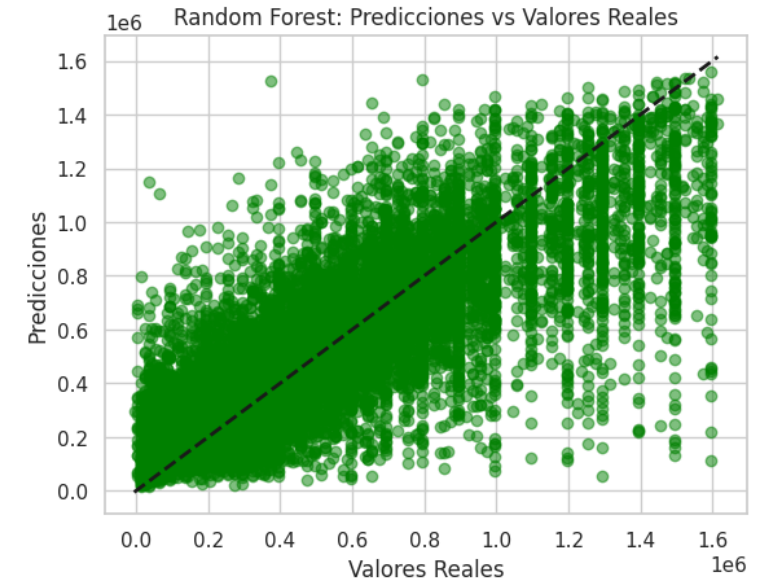
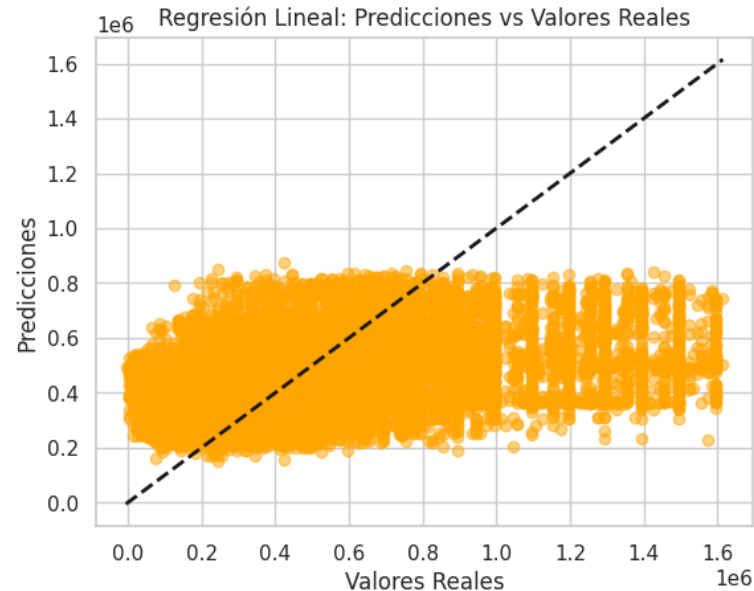
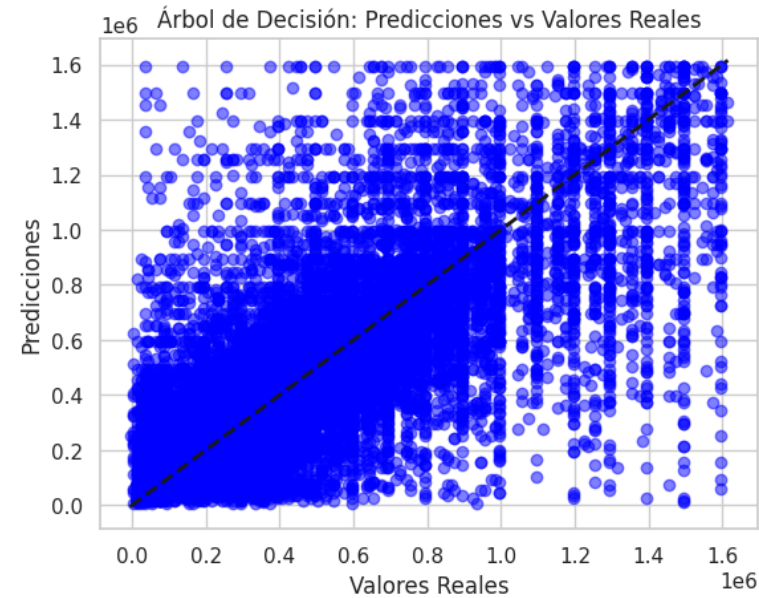
Modelos de ML para predicción del Precio

- Se evaluaron 3 modelos: Árbol de Decisiones, Random Forest y Regresión Lineal.

Random Forest: modelo más efectivo.
Relaciones no lineales en los datos.
Potencial para más ajustes y pruebas.

Resultados

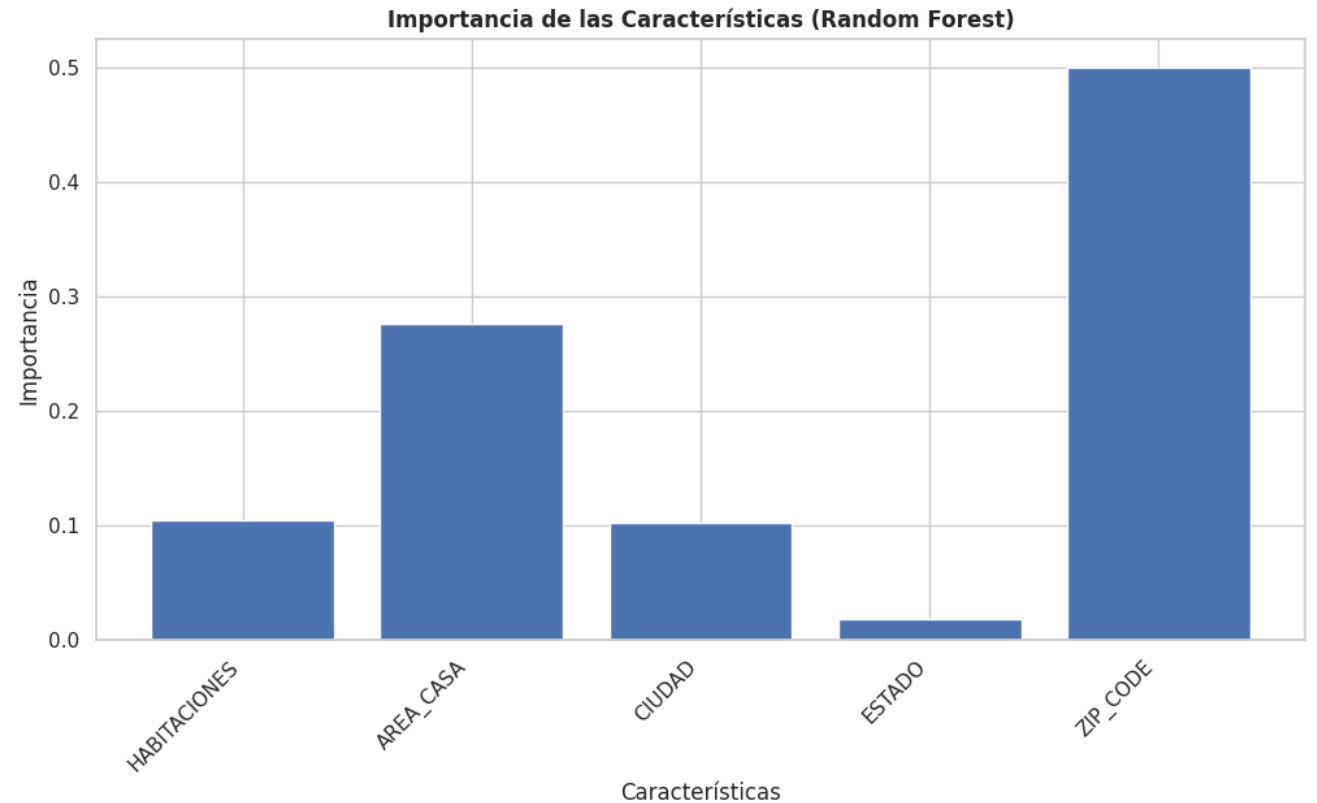
- Mejor Modelo: Random Forest.
- MSE: 4.30×10^{10} 4.30×10^{10}
- R2: 0.6143
- Árbol de Decisión:
- MSE: 7.19×10^{10} 7.19×10^{10}
- R2: 0.3554
- Regresión Lineal:
- MSE: 9.55×10^{10} 9.55×10^{10}
- R2: 0.1432



Modelos de ML para predicción del Precio

- Se intentaron optimizar los modelos con:
 - Tratamiento de valores atípicos
 - Cross validation
 - Categorización de variables Precio y Área Casa en intervalos

No se obtuvieron cambios significativos



Conclusiones

- Mejor Enfoque: Los modelos basados en árboles (especialmente Random Forest) parecen ser más apropiados para estos datos, reflejando la complejidad y las relaciones no lineales.
- Potencial de Mejora: Hay margen para explorar más características, realizar ajustes adicionales y probar otros modelos para posiblemente mejorar aún más el rendimiento.
- Implicaciones Prácticas: El análisis puede informar decisiones sobre la valoración de propiedades y la identificación de factores clave en la determinación de los precios.
- Factores determinantes: El número de habitaciones, los baños, el código postal y el área de la vivienda son los factores más determinantes para incidir en el precio de una vivienda.
- Área del Lote: El área del lote no afecta el precio de la vivienda construida. Esto es coherente ya que generalmente en el precio de la vivienda construida ya se encuentra implícito los gastos por el valor del lote.
- Las correlaciones de las variables que afectan el precio son débiles, esto se debe a que no es un modelo lineal bivariado. La combinación de diversos factores es la que influye drásticamente en el precio. Como se observó con la relación HABITACIONES/WC vs PRECIO. Ya que estas de manera individual no son fuertes, pero en conjunto si presentan mayor correlación.