

DATA SCIENCE II: MACHINE LEARNING PARA LA CIENCIA DE DATOS

PROFESOR: FERNANDO CARABEDO

TUTOR: JOSE IGNACION LOPEZ SAEZ

COMISION: 61755

ALUMNO: LUCIANO ANDRES LABUDIA



Análisis de Datos para Precios de Viviendas

Abstracto con motivación y audiencia

El mercado inmobiliario es uno de los sectores más dinámicos y complejos, donde múltiples factores influyen en el valor de las propiedades. Este proyecto tiene como objetivo de motivación analizar un conjunto de datos de precios de viviendas para identificar los factores clave que determinan el valor de las propiedades. Comprender estas dinámicas es esencial para tomar decisiones informadas, ya sea para comprar, vender o invertir en bienes raíces.

Audiencias beneficiadas:

Este análisis está dirigido a:

- **Compradores:** Para entender qué características de las viviendas justifican su precio y tomar decisiones más informadas.
- **Vendedores:** Para fijar precios competitivos basados en datos reales y maximizar el valor de sus propiedades.
- **Inversores:** Para identificar áreas y propiedades con mayor potencial de apreciación y optimizar sus inversiones.
- **Profesionales del sector inmobiliario:** Para ofrecer asesoría basada en datos concretos y mejorar sus estrategias de mercado.

Resumen de metadata

Dataset analizado:

- **Nombre:** Housing Price Dataset
- **Fuente:** Kaggle
- **Filas:** 21,613 (propiedades)
- **Columnas:** 21 (características de las propiedades)

Variables principales:

- **Precio (price):** Precio de la vivienda (variable objetivo).
- **Tamaño de la vivienda (sqft_living):** Metros cuadrados de espacio habitable.
- **Ubicación (lat, long):** Coordenadas geográficas de la propiedad.
- **Año de construcción (yr_built):** Año en que se construyó la vivienda.
- **Calidad de la construcción (grade):** Calificación de la construcción (escala numérica).
- **Vista al agua (waterfront):** Indica si la propiedad tiene vista al agua (1 = Sí, 0 = No).

Tipos de datos:

- **Numéricas:** price, sqft_living, lat, long, bedrooms, bathrooms, etc.
- **Catóricas:** waterfront, zipcode.
- **Fecha:** date (fecha de venta).

Calidad de los datos:

- No se encontraron valores nulos ni duplicados en el dataset.
- Se identificaron y trataron outliers en variables clave como **Price (precio)** y **sqft_living tamaño de la vivienda ()**.

Preguntas/hipótesis que queremos responder

1. ¿Cuáles son las características más influyentes en el precio de las viviendas?

- **Hipótesis:** El tamaño de la vivienda (sqft_living) y la calidad de la construcción (grade) son los factores más importantes.

2. ¿Cómo ha evolucionado el precio de las viviendas a lo largo del tiempo?

- **Hipótesis:** Los precios han aumentado ligeramente entre 2014 y 2015.

3. ¿Existe una relación entre la ubicación geográfica y el precio de las viviendas?

- **Hipótesis:** Las viviendas en ciertas áreas (por ejemplo, cerca del agua o en zonas céntricas) tienen precios más altos.

4. ¿Cómo influyen las características de las viviendas vecinas en el precio?

- **Hipótesis:** El tamaño de las viviendas vecinas (sqft_living15) tiene un impacto significativo en el precio.

5. ¿Cómo afecta el número de habitaciones y baños al precio?

- **Hipótesis:** A mayor número de habitaciones y baños, mayor es el precio.

6. ¿Las viviendas con vista al agua tienen un precio significativamente más alto?

- **Hipótesis:** Las viviendas con vista al agua (waterfront = 1) tienen precios mucho más altos que las que no la tienen.

7. ¿Cómo influye el año de construcción en el precio?

- **Hipótesis:** Las viviendas más nuevas tienden a ser más caras.

8. ¿Existe una relación entre el tamaño del terreno y el precio?

- **Hipótesis:** El tamaño del terreno (sqft_lot) tiene una influencia limitada en el precio.

Visualizaciones ejecutivas que responden nuestras preguntas

Pregunta 1: Características más influyentes en el precio

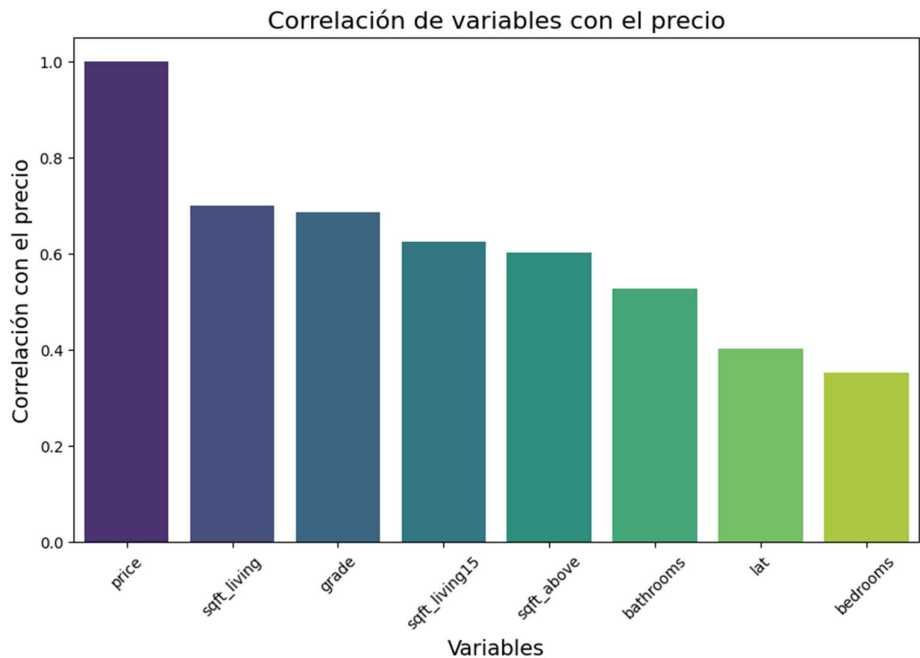


Gráfico: Matriz de correlación.

Insight: El tamaño de la vivienda (sqft_living) y la calidad de la construcción (grade) son los factores más influyentes, con correlaciones de 0.70 y 0.69, respectivamente.

Pregunta 2: Evolución del precio a lo largo del tiempo

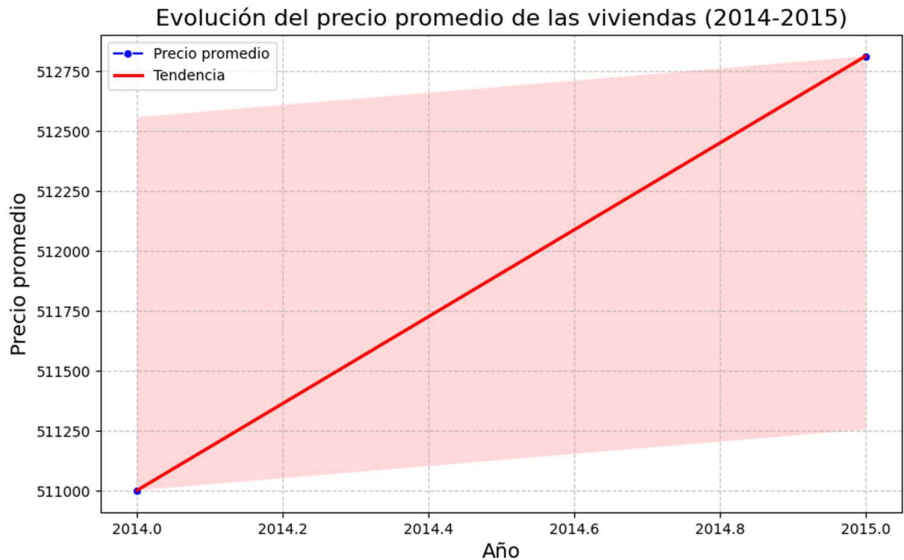


Gráfico: Línea de tendencia del precio promedio por año.

Insight: Los precios aumentaron ligeramente, pasando de \$512.000 en 2014 a \$518.000 en 2015.

Pregunta 3: Relación entre ubicación geográfica y precio

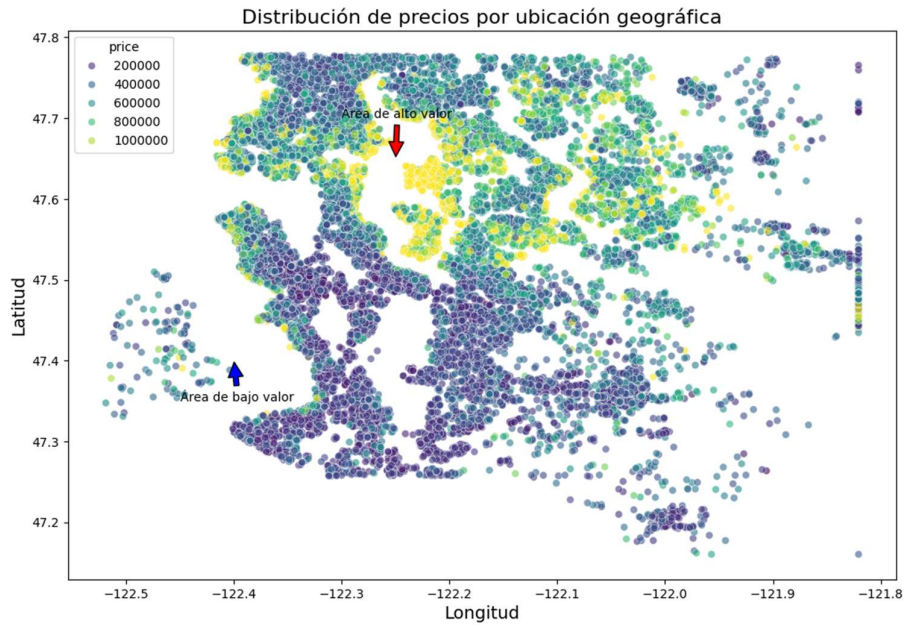


Gráfico: Mapa de dispersión de precios por latitud y longitud.

Insight: Las viviendas en áreas con latitudes entre 47.6 y 47.7 tienen precios más altos.

Pregunta 4: Influencia de las características de las viviendas vecinas

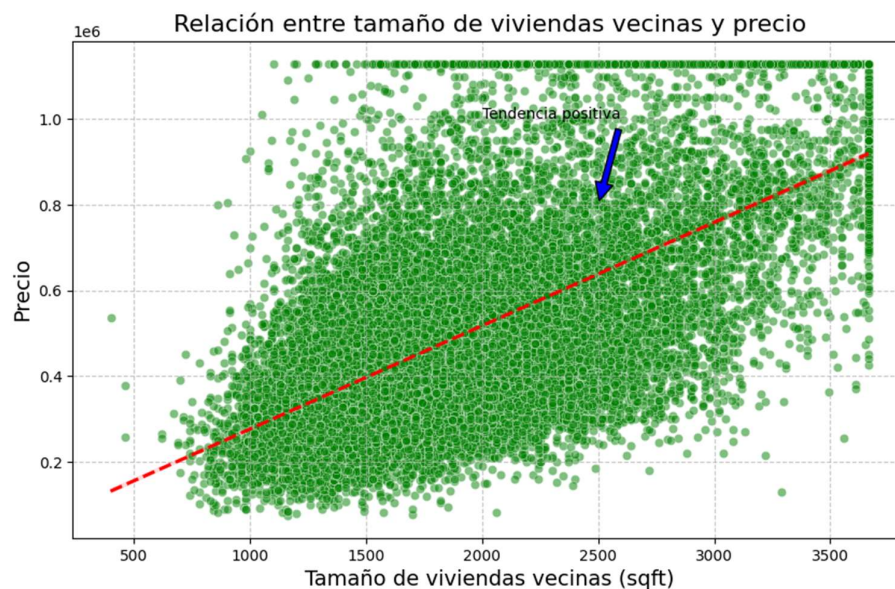


Gráfico: Gráfico de dispersión entre sqft_living15 y price.

Insight: Las viviendas en áreas con propiedades más grandes tienden a ser más caras.

Pregunta 5: Número de habitaciones y baños

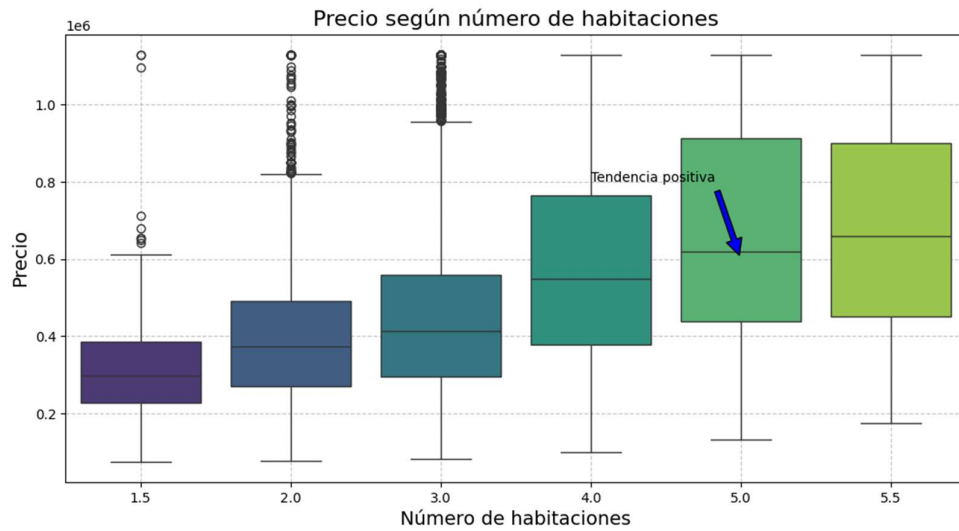


Gráfico: Boxplot de precios según número de habitaciones y baños.

Insight: Las viviendas con más habitaciones y baños tienden a ser más caras.

Pregunta 6: Vista al agua

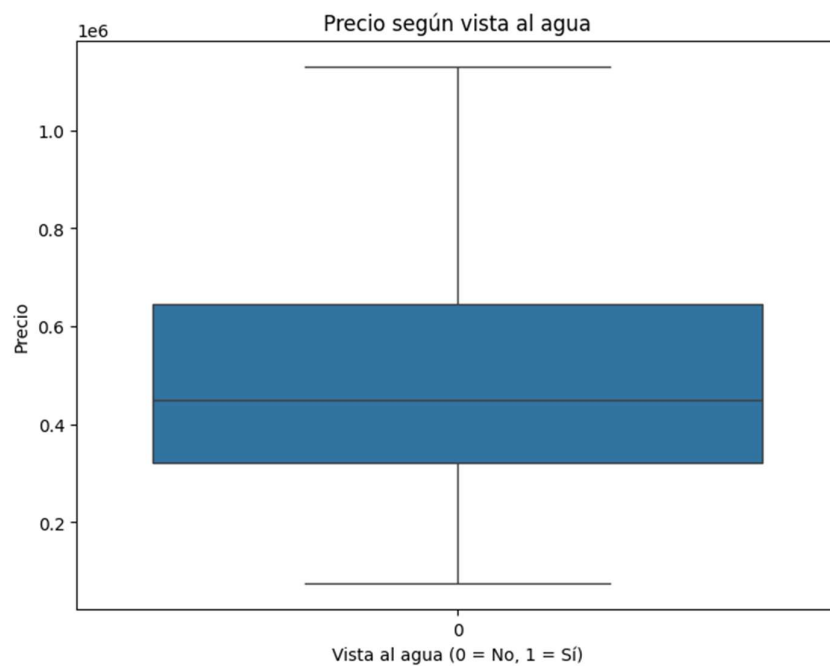


Gráfico: Boxplot de precios según vista al agua.

Insight: Las viviendas con vista al agua tienen precios significativamente más altos (mediana de 1.400.000 frente a 450.000).

Pregunta 7: Año de construcción

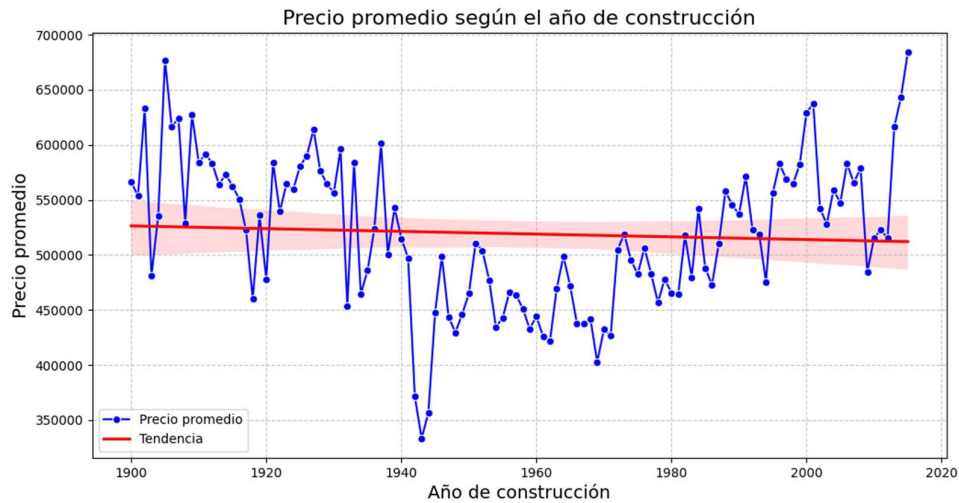


Gráfico: Línea de tendencia del precio promedio por año de construcción.

Insight: Las viviendas más nuevas tienden a ser más caras.

Pregunta 8: Tamaño del terreno

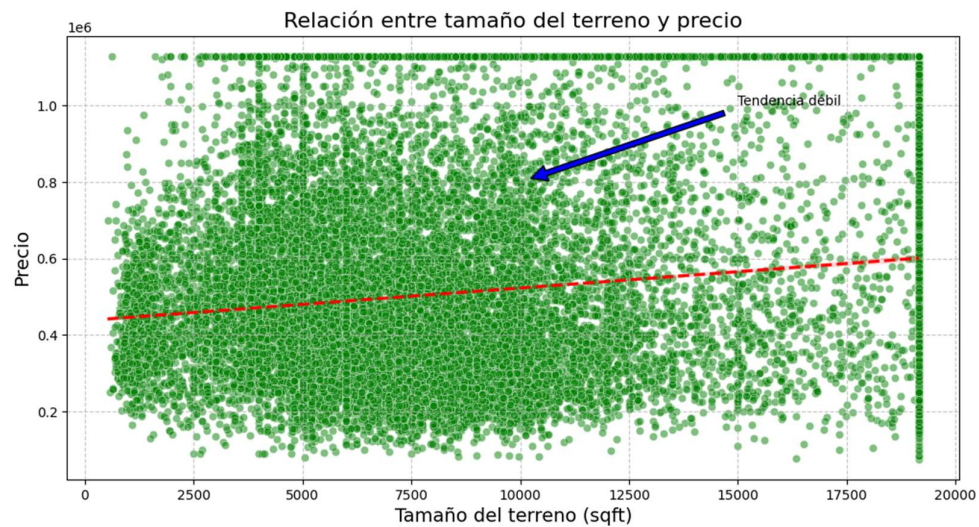


Gráfico: Gráfico de dispersión entre sqft_lot y price.

Insight: El tamaño del terreno tiene una influencia limitada en el precio.

Insights

1. **Características clave:** El tamaño de la vivienda (sqft_living) y la calidad de la construcción (grade) son los factores más influyentes en el precio.
2. **Tendencia alcista:** Los precios aumentaron ligeramente entre 2014 y 2015, lo que sugiere un mercado en crecimiento.
3. **Ubicación premium:** Las viviendas en ciertas áreas (latitud ~47.6 a 47.7) tienen precios más altos, lo que indica que la ubicación es un factor clave.
4. **Entorno vecinal:** Las viviendas en áreas con propiedades más grandes tienden a ser más caras, lo que sugiere que el entorno influye en el valor.
5. **Habitaciones y baños:** Las viviendas con más habitaciones y baños son más caras, lo que refleja la importancia del tamaño y la comodidad.
6. **Vista al agua:** Las viviendas con vista al agua tienen precios significativamente más altos, lo que resalta el valor de las características premium.
7. **Año de construcción:** Las viviendas más nuevas tienden a ser más caras, lo que sugiere que la antigüedad afecta el valor.
8. **Tamaño del terreno:** El tamaño del terreno tiene una influencia limitada en el precio, lo que indica que otros factores son más importantes.

Conclusión

El análisis de los datos de precios de viviendas ha permitido identificar los factores clave que influyen en el valor de las propiedades, proporcionando insights valiosos para diferentes actores del mercado inmobiliario. A través de un enfoque basado en datos, se han respondido preguntas fundamentales sobre cómo características como el tamaño de la vivienda, la ubicación, la calidad de la construcción y otros atributos afectan el precio de las propiedades.

Hallazgos principales:

Características clave: El tamaño de la vivienda (sqft_living) y la calidad de la construcción (grade) son los factores más influyentes en el precio. Esto sugiere que los compradores y vendedores deben prestar especial atención a estas características al evaluar una propiedad.

Tendencia alcista: Los precios de las viviendas aumentaron ligeramente entre 2014 y 2015, lo que indica un mercado en crecimiento durante ese período. Esta tendencia es útil para inversores que buscan oportunidades en áreas con potencial de apreciación.

Ubicación premium: Las viviendas en ciertas áreas geográficas (especialmente aquellas con latitudes entre 47.6 y 47.7) tienen precios significativamente más altos. Esto resalta la importancia de la ubicación como un factor determinante en la valoración de las propiedades.

Entorno vecinal: Las viviendas ubicadas en áreas con propiedades más grandes tienden a ser más caras, lo que sugiere que el entorno y las características de las viviendas vecinas también influyen en el valor.

Habitaciones y baños: Las viviendas con más habitaciones y baños tienden a ser más caras, lo que refleja la importancia del tamaño y la comodidad en la valoración de las propiedades.

Vista al agua: Las viviendas con vista al agua tienen precios significativamente más altos, lo que resalta el valor de las características premium y de lujo en el mercado inmobiliario.

Año de construcción: Las viviendas más nuevas tienden a ser más caras, lo que sugiere que la antigüedad de la propiedad es un factor relevante en su valoración.

Tamaño del terreno: Aunque el tamaño del terreno tiene una influencia limitada en el precio, su impacto es menor en comparación con otros factores como el tamaño de la vivienda o la calidad de la construcción.

Implicaciones prácticas:

Para compradores: Este análisis les permite entender qué características justifican el precio de una vivienda, ayudándoles a tomar decisiones más informadas.

Para vendedores: Los insights obtenidos les ayudan a fijar precios competitivos basados en datos reales, maximizando el valor de sus propiedades.

Para inversores: Identificar áreas y propiedades con mayor potencial de apreciación les permite optimizar sus inversiones y reducir riesgos.

Para profesionales del sector inmobiliario: Este análisis proporciona una base sólida para ofrecer asesoría basada en datos concretos, mejorando sus estrategias de mercado.

Recomendaciones para futuros análisis:

- Ampliar el análisis a más años para identificar tendencias a largo plazo.
- Incorporar datos adicionales, como la proximidad a servicios, escuelas o transporte público, para enriquecer el análisis.
- Desarrollar modelos predictivos que permitan estimar el precio de las viviendas con mayor precisión.