

# Limpieza, Visualización y Análisis de Datos de Precios de Propiedades Inmobiliarias en Estados Unidos durante 2024



Universidad Americana (UAM)

**Marcos Ulises Sánchez**

musanchez@uamv.edu.ni

Inteligencia de Negocios

Facultad de Ingeniería y Arquitectura

7 de octubre de 2024

## Resumen

Este informe explora cómo las características físicas de las propiedades inmobiliarias en Estados Unidos influyen en su precio. Entre estas características están el tamaño del lote de la propiedad, el número de habitaciones y su ubicación geográfica. Los datos que se proporcionaron presentaron varios desafíos, como valores faltantes y casos poco realistas en algunos aspectos como el número de habitaciones o baños. Usamos modelos de regresión lineal para cuantificar la relación entre las características de las propiedades y sus precios, además de técnicas de visualización para encontrar tendencias y patrones. En los resultados obtenidos, encontramos que los atributos físicos considerados son significativos para explicar el precio. Sin embargo, el modelo también sugiere que el precio está determinado por otros factores de los que no tenemos información. Este análisis es relevante para las operaciones de compra y venta en el sector inmobiliario, siendo una guía para la valuación de propiedades y la planificación de estrategias de inversión.

**Palabras clave:** Regresión lineal, Precios inmobiliarios, Mínimos cuadrados ponderados (WLS), Visualización de datos, Outliers

## Introducción

El mercado de bienes inmuebles ha sido históricamente una de las piedras angulares de la economía global, y en el caso particular de Estados Unidos, su dinámica es tan compleja como diversa. Cada propiedad tiene características físicas únicas, como el número de habitaciones, el tamaño del terreno o atributos más complejos como su disposición geográfica o arquitectura, que hacen que el precio de viviendas que en papel serían iguales, tengan precios completamente distintos. Esta dificultad es más notoria cuando se consideran los mercados locales, que a menudo son tan dispares que los precios de las propiedades pueden variar drásticamente de un estado a otro, o incluso entre ciudades dentro de un mismo estado. El valor de una propiedad en California, por ejemplo, puede estar influenciado por una fuerte demanda y la limitada disponibilidad de terreno, mientras que en estados del medio oeste, como Ohio, los precios pueden estar mucho más estabilizados, respondiendo a dinámicas regionales distintas.

Este fenómeno presenta un reto significativo para los inversionistas, compradores y vendedores, quienes deben navegar por un mercado volátil y riesgoso. Tomar decisiones acertadas en este entorno requiere una comprensión profunda de los factores determinantes que impactan el valor de las propiedades, más allá de una simple evaluación superficial de sus características físicas. Las implicaciones de los factores geográficos, la influencia de las condiciones económicas locales y nacionales, y la interacción entre la oferta y la demanda son elementos clave que deben ser analizados para obtener una imagen clara del panorama inmobiliario.

El propósito de este análisis es mostrar la magnitud en la que características simples inciden en el valor de los inmuebles, con el objetivo de identificar patrones consistentes que permitan una valoración más precisa de las propiedades en diferentes contextos regionales. No se trata solo de entender cómo las características físicas de una propiedad afectan su precio, sino de explorar cómo las particularidades geográficas y las dinámicas del mercado local también juegan un papel crucial en la determinación del valor. Al identificar estos factores clave, el análisis no solo permitirá tomar decisiones más informadas en un mercado tan dinámico como el de bienes raíces, sino que además podría ofrecer una ventana hacia el estado general de la economía de un país. Es bien sabido que los precios de las propiedades suelen actuar como indicadores adelantados de la salud económica, volviéndose relevante no solo para actores directos en este mercado, sino también para economistas, urbanistas o planificadores de políticas públicas que podrían encontrar provecho en la estimación de precios de viviendas.

## Revisión de literatura

1. (Chowhaan et al., 2023). Por medio de algoritmos como Random Forest, Regresión Lineal, Regresión de Lasso y XGBoost, donde se utilizaron variables como baños, tamaño en pies cuadrados y la cantidad de balcones. En este estudio se obtuvo un modelo con un 90 % de precisión, siendo el más exacto. Para este estudio tomamos variables similares, siendo las más representativas el número de habitaciones, baños, área de la construcción y lote, incluyendo el estado como una variable categórica. La diferencia en el enfoque es que esta vez no se intenta hacer una predicción, más bien estamos calculando la sensibilidad de los precios cuando hay cambios en las variables exógenas o independientes, con una regresión log-lineal.
2. (Baldominos et al., 2018). En este trabajo se realizó un modelo predictivo para determinar oportunidades de inversión en bienes raíces. Para la predicción, al igual que en este análisis, se consideraron características añadidas a la ubicación, como la zona en el distrito de Salamanca, código postal, calle y número de piso en el que la propiedad se encuentra, diferencias debidas a la diferencia en el área geográfica abarcada.
3. El artículo de Zietz et al. (2008) intenta explicar cómo los precios de inmuebles son afectados por características propias o de ubicación, clasificando a las propiedades en varios rangos de precio a lo largo de la distribución de estos, implementando en su modelo factores clave como la accesibilidad a autopistas y servicios, que tienen un mayor impacto en viviendas en los umbrales de precio más bajo en el área de Orem/Provo, Utah.

## Recolección y preprocesamiento de datos

Tomamos los datos de Kaggle, de Sakib (2023), y los alojamos en GitHub. Los datos, según su documentación, fueron recopilados de fuentes públicas que proporcionan información sobre propiedades inmobiliarias en Estados Unidos. Cuando leímos el archivo usando la librería pandas de Python, identificamos las columnas principales, la cantidad de registros y los valores nulos.

En total hay 2, 221, 849 registros, con las siguientes columnas:

- brokered\_by: 0.02 % de los registros son nulos.

- status: No hay valores nulos.
- price: 0.07 % valores faltantes.
- bed: 21.62 % valores nulos.
- bath: 22.99 % de valores nulos.
- acre\_lot: 14.62 % registros nulos.
- street: 0.49 % registros nulos.
- city: 0.06 % registros nulos.
- state: Solo 8 valores nulos.
- zip\_code: Solamente el 0.01 % de los valores son nulos.
- house\_size: El 25 % de los valores son nulos.
- prev\_sold\_date: El 32.98 % no tienen información, es probable que las propiedades no fuesen vendidas antes.

Para nuestro análisis, identificamos como columnas importantes las siguientes: bed, bath, state, house\_size y acre\_lot. Consideramos estas como los factores determinantes en el precio de las viviendas.

## Manejo de valores faltantes

Los pasos que seguimos fueron:

1. Eliminamos los registros de las casas que no tienen precio.
2. Los valores nulos en las fechas previas de venta fueron reemplazados por "Not sold before".
3. Eliminamos los registros que tenían demasiados valores nulos (consideramos "demasiados" cuando faltaban la mitad o más de la mitad de los datos).
4. Para tratar los registros con faltante en la cantidad de baños, usamos la cantidad de baños promedio por número de habitaciones para reemplazar los baños nulos.

5. Los valores faltantes para la cantidad de habitaciones fueron reemplazados por la mediana.

```

Valores nulos en cantidad:
status      0
price       0
bed         0
bath        0
acre_lot    0
city        0
state       0
house_size  0
prev_sold_date 0
dtype: int64
Valores nulos en porcentaje:
status      0.00
price       0.00
bed         0.00
bath        0.00
acre_lot    0.00
city        0.00
state       0.00
house_size  0.00
prev_sold_date 0.00
dtype: float64

```

Figura 1: Datos después de tratamiento de valores faltantes.

## Tratamiento de outliers

Identificamos una cantidad significativa de valores atípicos que se alejan drásticamente del rango esperado. Los boxplots generados revelaron que había propiedades con hasta 800 baños o áreas que eran imposiblemente grandes o pequeñas. Este tipo de valores distorsionan la media de los datos y alteran la precisión de muchos modelos que dependen de la normalidad de la distribución.

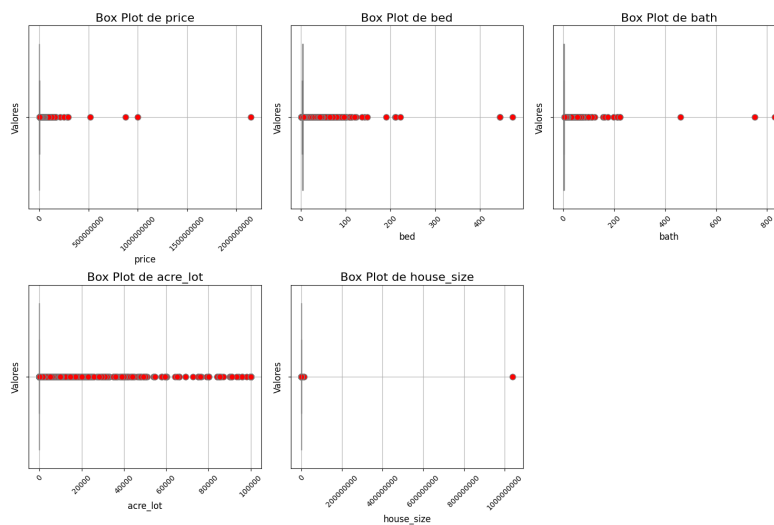


Figura 2: Distribución en gráficos de caja

Removimos la mayoría de outliers eliminando datos incoherentes, como casas sin baño o habitaciones, o precios iguales a cero. También eliminamos casas demasiado pequeñas. Aprovechamos para eliminar posibles incoherencias validando que el área del lote fuera mayor o igual al área que ocupaba la casa, convirtiendo el lote que estaba en acres a pies cuadrados.

Eliminamos outliers para las variables `house_size` y `acre_lot` usando rangos intercuartílicos, y conservamos registros de baños y habitaciones con un Z-score menor que 3, eliminando cantidades que se alejaban demasiado del rango que consideraríamos normal. También aplicamos el logaritmo a las variables `price` y `house_size`, ya que sus valores absolutos eran demasiado grandes. Aplicamos una winsorización a `price` para acercar los valores extremos a los percentiles 5 y 95.

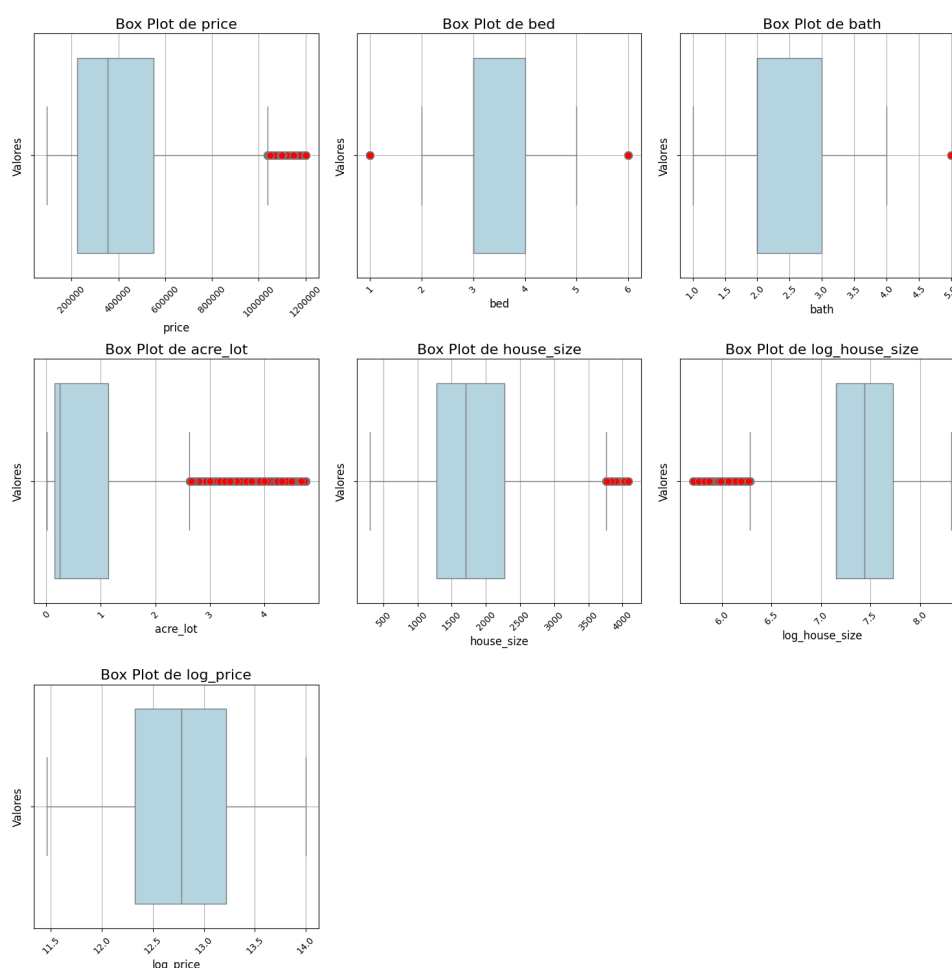


Figura 3: Distribución posterior a las transformaciones

En este análisis, hay outliers que decidimos conservar. A pesar de estar alejados de la media, permanecen en un rango razonable para bienes raíces, ya que podrían ser viviendas muy grandes o de lujo. Aunque sean menos comunes, forman parte de la realidad del mercado.

## Metodología

### Regresión lineal por métodos WLS y OLS

Para llevar a cabo el análisis de los factores que afectan el precio de las propiedades, usamos un modelo de regresión lineal ponderada (WLS) y uno de mínimos cuadrados ordinarios (OLS). Cuantificamos la relación entre el precio de las propiedades y sus características físicas, como habitaciones, baños y tamaño de lote, incluyendo al estado como variable categórica. Elegimos en particular una regresión ponderada para ajustar el modelo de manera que se tome en cuenta la importancia relativa de cada observación, esto debido a la naturaleza de los datos proporcionados. En el mercado inmobiliario, ciertas propiedades, como aquellas que pertenecen a áreas con alta densidad de población o propiedades con características de lujo, tienden a distorsionar el modelo, por lo que es crucial mitigar tanto como sea posible este impacto y evitar que estados caros o demasiado poblados dominen los resultados generales. Aunque usemos estos pesos, haremos una comparativa con el método OLS para determinar el modelo más robusto. Los supuestos son, acorde a Zach Robbitt (2020):

- **Relación lineal:** Se asume que existe una relación lineal entre las características de las propiedades, como el tamaño de propiedad, lote, o número de habitaciones, y la variable dependiente: el precio de venta.
- **Independencia de las observaciones:** Es decir, que las propiedades físicas de los inmuebles son independientes entre sí.
- **Homoscedasticidad:** Varianza constante para cualquier grupo de observaciones.
- **Normalidad:** Los residuos de la variable dependiente están normalmente distribuidos.

Sobre el cálculo de los pesos para la regresión WLS, estos fueron estimados usando el producto de los inversos de la varianza del logaritmo del precio y el número de observaciones por estado. El objetivo de esto fue dar un mayor peso a los estados con distribuciones de precios más estables, considerando también la distorsión que podrían causar en los resultados los estados con un gran número de observaciones. Pennsylvania State University (s.f.) aborda la WLS cuando se tienen problemas por heterocedasticidad, es decir, cuando la varianza no es constante para diversos subgrupos de datos.

## Gráficos para la visualización de los datos

Durante nuestro análisis, usamos varios análisis gráficos para entender la naturaleza de los datos, además de ser útiles para decidir qué transformaciones aplicar. Las visualizaciones son esenciales para comprender mejor las relaciones entre variables e identificar las razones de posibles problemas para cumplir con los supuestos de nuestro modelo. Los gráficos usados fueron:

- **Gráficos de densidad:** Usados principalmente para entender cómo estaba distribuida la variable dependiente.
- **Boxplots:** Útiles para identificar valores atípicos y proceder a su tratamiento.
- **Histogramas:** Permiten observar la frecuencia de los valores en diferentes rangos de las variables principales.
- **Matriz de correlación:** Ayuda a identificar problemas de multicolinealidad entre variables independientes.
- **Gráficos Q-Q (Quantile-Quantile):** Permiten evaluar, de manera más clara que un gráfico de densidad, si los datos siguen una distribución normal.
- **Gráfico de barras:** Útil para visualizar cómo se distribuyen las observaciones por variable categórica.

## Resultados

### Análisis de gráficos

Con el gráfico presentado en la Figura 4, al observar los precios medios de las viviendas por estado, podemos intuir algunas razones para estas distribuciones entre estados con precios de propiedades altos, moderados y bajos. Más adelante notaremos que hay estados en los que encontramos acumulación de propiedades en ciertos rangos de precio, afectando nuestra distribución.



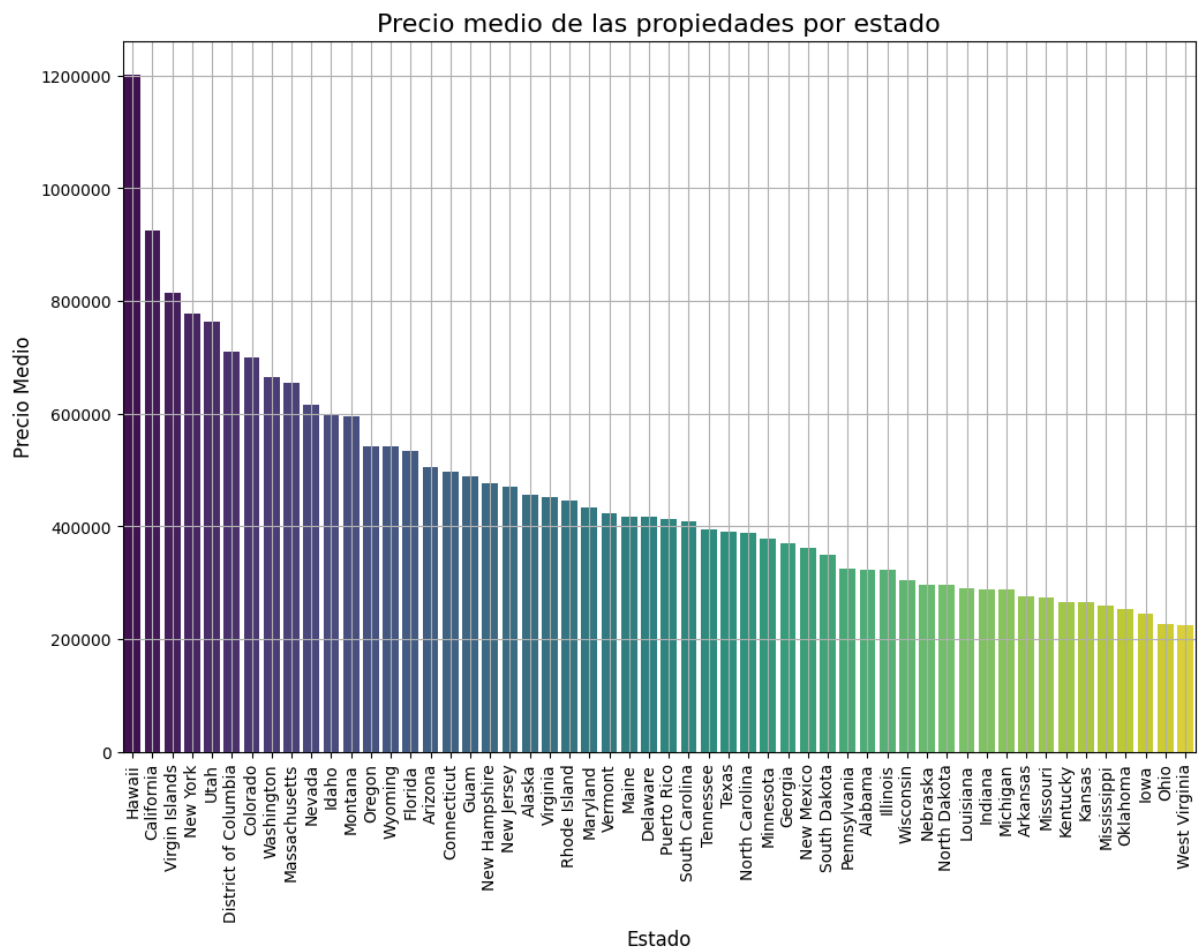


Figura 4: Precio medio de bienes inmuebles por estado

### 1. Estados con precios altos

- **Hawái:** Los precios son elevados debido a su condición de isla, con poca disponibilidad de terreno para construir, además de ser un destino turístico atractivo para residencias de lujo.
- **California:** Conocido por su alta demanda de vivienda, este estado alberga a Silicon Valley, lo que incrementa los precios debido a los altos ingresos de los trabajadores de la industria tecnológica.
- **Otros estados con precios elevados como Virginia y Nueva York:** Son centros económicos importantes y también relevantes para el turismo. Muchos cuentan con una oferta limitada de terreno para construcción, lo que restringe la capacidad de ofertar vivienda en áreas con alta demanda.

## 2. Estados con precios moderados

- Texas, Florida o Arizona: Sus precios no son tan elevados posiblemente por su gran extensión territorial y una demanda de vivienda menor en comparación a estados como California.
- La mayor disponibilidad de vivienda es esperable, y muchas personas podrían optar por migraciones internas en busca de hogares más asequibles.

## 3. Estados con precios bajos

Estados como Ohio, Iowa o Mississippi: Tienen economías menos desarrolladas y baja demanda inmobiliaria, lo que explica sus precios considerablemente más bajos.

En la Figura 5, se presenta la distribución de los precios. Podemos notar una acumulación de precios bajos y altos en las viviendas. Esto se debe a que los precios extremos están concentrados en un grupo específico de estados. Por ejemplo, los precios altos se encuentran principalmente en California, Florida y Nueva York, mientras que Iowa, Illinois y Florida presentan más acumulaciones de viviendas en rangos de precios bajos. Sorprendentemente, en Nueva York también se da una acumulación de viviendas en los rangos de precios más bajos.

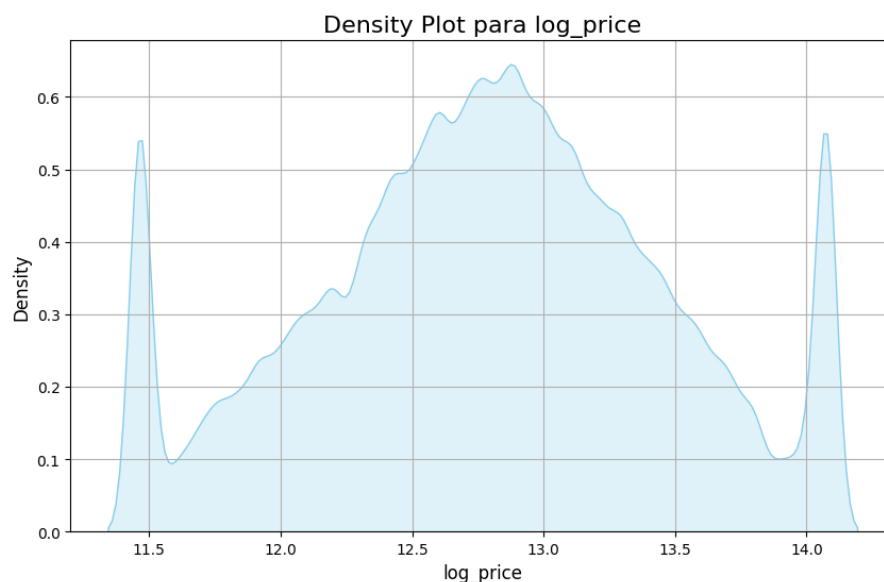


Figura 5: Gráfico de densidad del logaritmo del precio

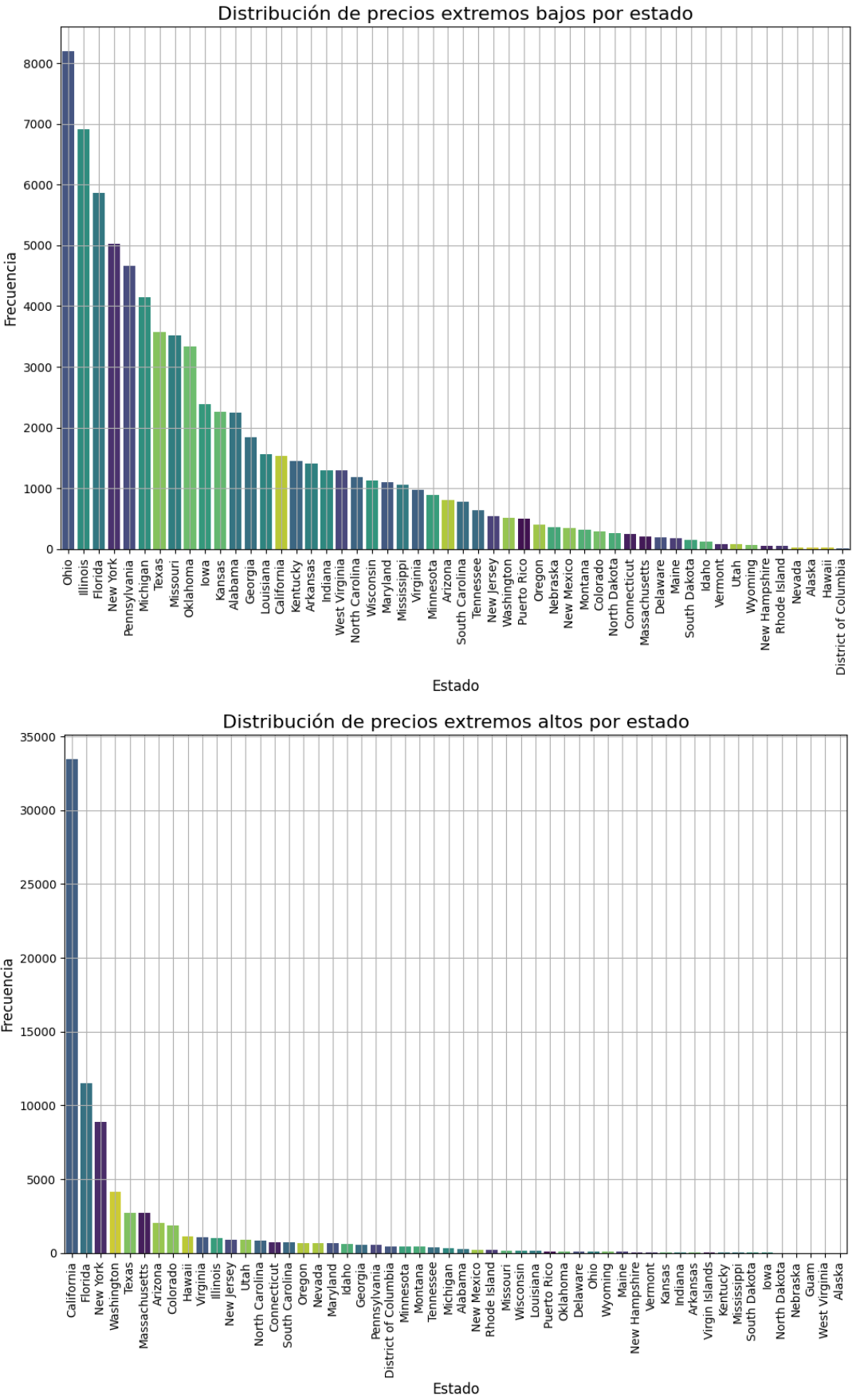


Figura 6: Frecuencia por estado de valores extremos de la distribución de precios

Para analizar los factores que influyen en el precio de las casas, usamos la matriz de correlación, que nos dará una visión global de las relaciones entre las variables tomadas. Entre las variables analizadas (precio, tamaño de casa, tamaño de lote, cantidad de habitaciones y cantidad de baños), el mayor coeficiente de correlación con el precio fue el del número de baños. El número de cuartos y baños de una casa están altamente relacionados, mientras que el tamaño del lote parece tener menos impacto en la determinación del valor de una propiedad. Sin embargo, el número de baños y el tamaño de la casa parecen ser determinantes más importantes.

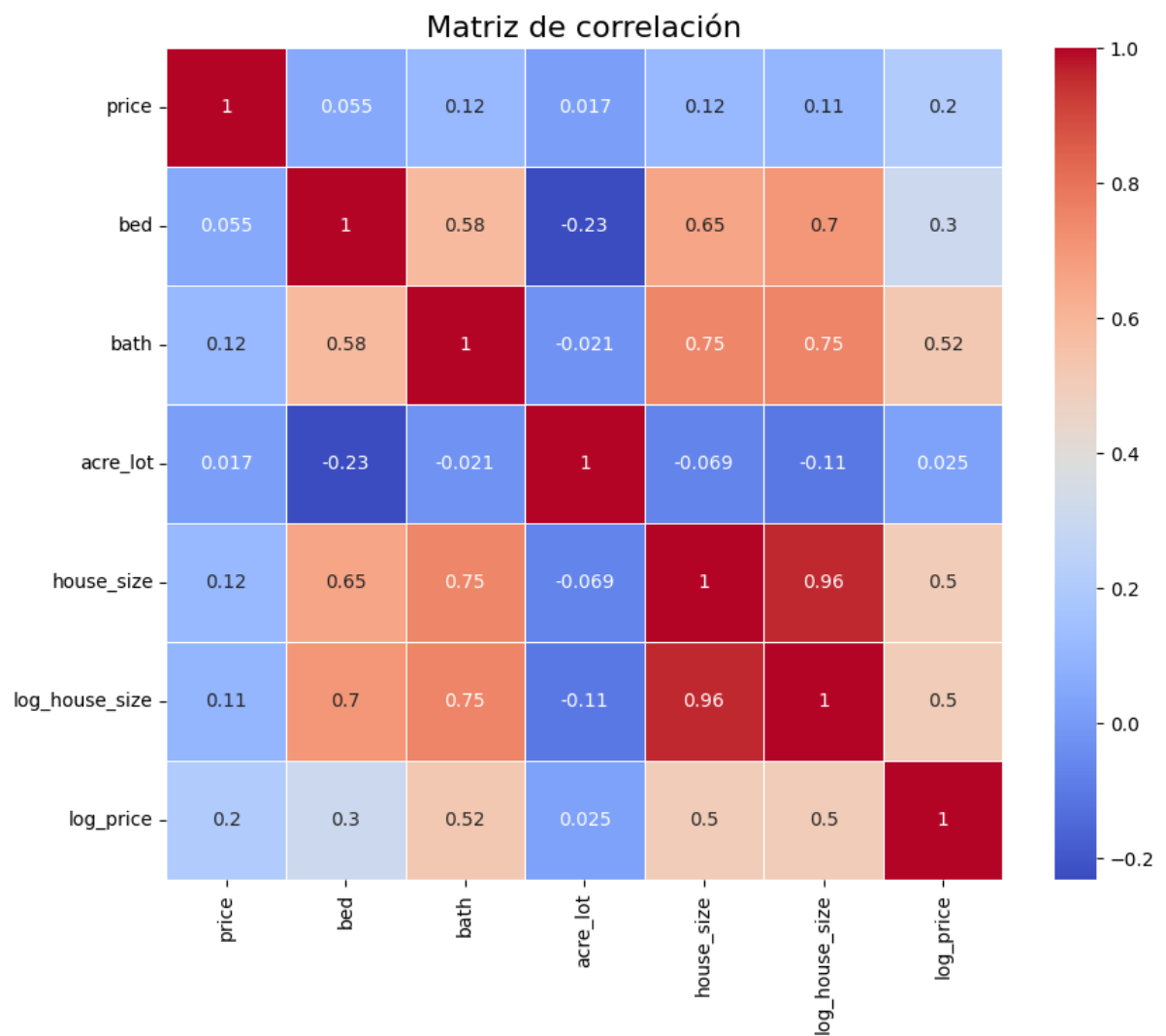


Figura 7: Matriz de correlación

## Modelos de regresión lineal

Usaremos como variable dependiente el logaritmo del **precio** de las viviendas. Esto es útil para analizar los cambios en el precio de las viviendas en términos porcentuales en lugar de valores absolutos. Usaremos como variables independientes el **estado**, categorizado, y el **tamaño de la casa, lote, número de cuartos y baños**.

Resumen WLS:						
WLS Regression Results						
=====						
Dep. Variable:	log_price	R-squared:	0.615			
Model:	WLS	Adj. R-squared:	0.615			
Method:	Least Squares	F-statistic:	4.148e+04			
Date:	do., 06 oct. 2024	Prob (F-statistic):	0.00			
Time:	05:36:04	Log-Likelihood:	-1.2769e+06			
No. Observations:	1480400	AIC:	2.554e+06			
Df Residuals:	1480342	BIC:	2.555e+06			
Df Model:	57					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
const	7.2758	0.009	810.141	0.000	7.258	7.293
bed	-0.0651	0.001	-128.933	0.000	-0.066	-0.064
bath	0.2056	0.001	362.427	0.000	0.205	0.207
acre_lot	0.0319	0.000	101.581	0.000	0.031	0.033
log_house_size	0.6458	0.001	487.533	0.000	0.643	0.648
state_Alaska	0.4204	0.004	112.928	0.000	0.413	0.428
state_Arizona	0.5569	0.004	156.814	0.000	0.550	0.564
state_Arkansas	-0.0449	0.004	-11.954	0.000	-0.052	-0.038
state_California	1.0534	0.004	290.845	0.000	1.046	1.061
state_Colorado	0.6372	0.004	177.926	0.000	0.630	0.644
state_Connecticut	0.3875	0.004	99.118	0.000	0.380	0.395
state_Delaware	0.2958	0.004	82.165	0.000	0.289	0.303

Figura 8: Modelo WLS incluyendo todas las variables

El modelo es bastante adecuado a simple vista. Los p-values de los parámetros estimados son casi todos significativos, con un  $R^2$  aceptable. Sin embargo, este análisis sugiere que un incremento en el número de habitaciones bajaría el precio al mantener constantes otras variables. Este resultado usualmente ocurre cuando hay problemas de multicolinealidad, como encontramos en la matriz de correlación en la Figura 6. El número de habitaciones está correlacionado con el tamaño de la casa y el número de baños. Para obtener un modelo más robusto, optamos por mantener solo una variable: el logaritmo del tamaño de la vivienda.

Dep. Variable:	log_price	R-squared:	0.573
Model:	WLS	Adj. R-squared:	0.573
Method:	Least Squares	F-statistic:	3.684e+04
Date:	do., 06 oct. 2024	Prob (F-statistic):	0.00
Time:	06:59:49	Log-Likelihood:	-1.3529e+06
No. Observations:	1480400	AIC:	2.706e+06
Df Residuals:	1480345	BIC:	2.707e+06
Df Model:	54		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	5.9604	0.007	865.329	0.000	5.947	5.974
log_house_size	0.8643	0.001	1041.995	0.000	0.863	0.866
state_Alaska	0.4343	0.004	110.903	0.000	0.427	0.442
state_Arizona	0.5298	0.004	142.293	0.000	0.522	0.537
state_Arkansas	-0.0804	0.004	-20.351	0.000	-0.088	-0.073
state_California	1.0608	0.004	279.006	0.000	1.053	1.068
state_Colorado	0.6552	0.004	174.022	0.000	0.648	0.663
state_Connecticut	0.4093	0.004	99.505	0.000	0.401	0.417
state_Delaware	0.2945	0.004	77.819	0.000	0.287	0.302
state_District of Columbia	1.1970	0.004	320.926	0.000	1.190	1.204
state_Florida	0.5040	0.004	123.317	0.000	0.496	0.512
state_Georgia	0.1482	0.004	39.116	0.000	0.141	0.156
state_Guam	0.6624	0.004	188.549	0.000	0.656	0.669

Figura 9: Regresión WLS con log\_house\_size y state como variables independientes

Realizando la regresión solo con el logaritmo del precio y el área de vivienda, usando el estado como variable categórica, obtenemos resultados más confiables. Al usar al estado como variable categórica, se tomó Alabama como estado de referencia. Algunas conclusiones interesantes incluyen:

- Una casa en Alabama, si estuviera en California y tuviera las mismas características, podría tener un incremento de precio del 189 % aproximadamente.
- Un incremento del 10 % en el área de una propiedad podría aumentar su precio en un 137 % aproximadamente.
- $R^2$ : El 57.4 % de las variaciones en el precio son explicadas por el área de la vivienda.
- Aunque a simple vista West Virginia presentaba el precio promedio más bajo, Ohio tiene el coeficiente menor (-0.2178), lo que implica un incremento de 19.6 % en comparación con una propiedad en Alabama.
- Hawái es el estado con las casas más caras. La misma casa en Ohio sería 363 % más cara en Hawái.

Cuando realizamos las mismas regresiones con el método de mínimos cuadrados ordinarios (OLS), los resultados son casi idénticos. Sin embargo, el valor de  $R^2$  es más alto con WLS, por lo tanto, conservamos este enfoque.

## Discusión de resultados

Los resultados obtenidos confirman que los atributos de una propiedad, como su tamaño, cantidad de cuartos o posición geográfica, influyen significativamente en el precio de bienes raíces. Estos hallazgos están alineados con estudios previos que consideran estas características como determinantes. Un hallazgo interesante fue que el tamaño del lote no impacta tanto en el precio de una propiedad como podría esperarse. Sin embargo, fue difícil llegar a conclusiones claras sobre otros factores físicos debido a la multicolinealidad de las variables independientes. Estos resultados tienen implicaciones importantes para los actores del mercado inmobiliario. Los inversionistas pueden usar esta información para identificar áreas con mayor potencial de apreciación de precios. Vendedores podrían determinar qué aspectos mejorar para aumentar el valor de una vivienda.

Un desafío importante fue la falta de normalidad en la distribución de los precios. Desde el inicio del análisis, los datos de precios mostraron dos picos marcados en los extremos de la distribución, lo que dificultó los análisis que asumen normalidad en los residuos, como la regresión lineal. Para mitigar este problema, aplicamos varias transformaciones, incluyendo la transformación logarítmica de los precios y la técnica de winsorización, que ajustó los valores extremos hacia los percentiles 5 y 95.

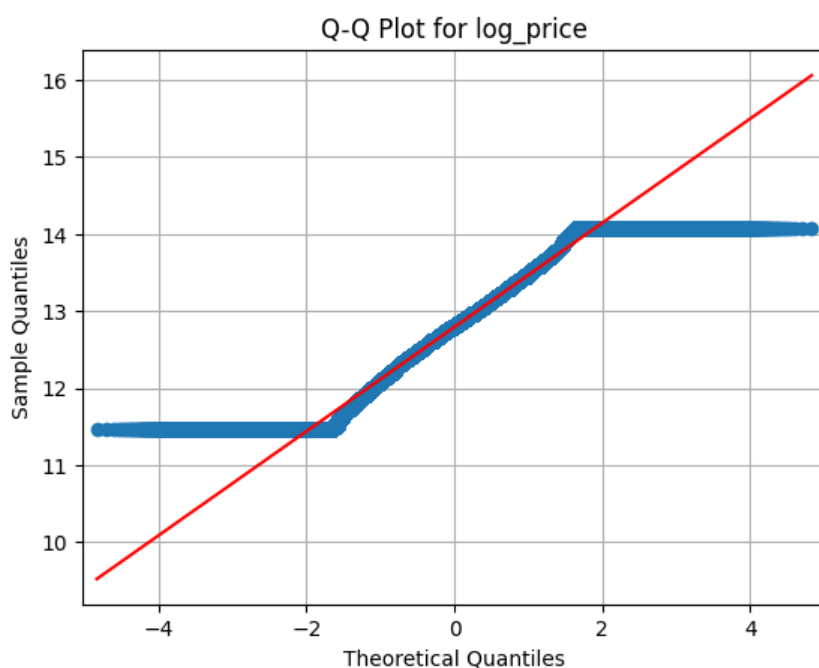


Figura 10: Gráfico Q-Q como prueba de normalidad

Si bien estas transformaciones ayudaron a suavizar la distribución y reducir el impacto de los valores atípicos, no lograron completamente normalizar los datos. El gráfico Q-Q mostró que, a pesar de las transformaciones, los extremos de la distribución todavía se desviaban significativamente de la normalidad, con una acumulación en los precios bajos y altos. Esta acumulación está concentrada en estados específicos, como Ohio e Illinois para los precios bajos, y California y Florida para los precios altos.

Este problema es relevante porque la regresión lineal, y en particular los modelos de mínimos cuadrados ordinarios (OLS), asumen que los residuos están normalmente distribuidos. Dado que este supuesto no se cumple completamente en este caso, las conclusiones deben ser interpretadas con cautela, especialmente en los rangos más extremos de los precios. La falta de normalidad podría estar afectando la precisión de los coeficientes estimados y, por lo tanto, las predicciones del modelo. Aunque la regresión ponderada (WLS) ayuda a mitigar en parte estos problemas, las conclusiones podrían estar influidas por la distribución sesgada de los datos. Si bien esta técnica tiene sentido para mitigar la influencia desproporcionada de estados con grandes fluctuaciones de precio o muchas observaciones, la ponderación podría estar subestimando o sobrestimando el impacto de ciertos estados con distribuciones más concentradas.

## Conclusión

El análisis realizado confirma que las propiedades físicas de una vivienda son determinantes en su precio en Estados Unidos, con un efecto distinto por estado. El objetivo principal de este estudio era analizar cómo las características físicas de las propiedades y la ubicación geográfica afectan el valor de las propiedades. Esto se cumplió al identificar patrones claros en los datos, que confirman la influencia significativa de los factores mencionados. Aunque no todas las suposiciones del modelo se cumplieron al 100 %, como la normalidad de los residuos, el estudio proporciona una base sólida para entender los principales determinantes del valor de las propiedades en un mercado inmobiliario tan diverso como el estadounidense.

Este análisis deja varias áreas que podrían beneficiarse de una mayor investigación:

1. Incorporación de más variables: Variables como la calidad percibida de la construcción, la edad de las propiedades, o el acceso a servicios, que están totalmente desligadas del tamaño de la propiedad, podrían mejorar el modelo planteado.



2. Análisis de submercados específicos: Podría ser interesante llevar a cabo un análisis más detallado de submercados específicos, como propiedades de lujo o en estados particulares.
3. Impacto de factores macroeconómicos: El estudio propuesto no toma en cuenta factores como las tasas de interés hipotecarias o el desempleo, lo que haría el análisis más rico e interesante.

## Referencias

- Baldominos, A., Blanco, Á., & Saez, Y. (2018). Real Estate Appraisal: A Machine Learning Approach. *Applied Sciences*, 8(2321), 1-14. <https://doi.org/https://doi.org/10.48550/arXiv.1809.04933>
- Chowhaan, J., Nitish, M., Akash, G., Nelli, S., & Shaik, S. (2023). Machine Learning Approach for House Price Prediction. *Asian Journal of Research in Computer Science*, 16, 54-61. <https://doi.org/10.9734/ajrcos/2023/v16i2339>
- Pennsylvania State University. (s.f.). Weighted Least Squares. <https://online.stat.psu.edu/stat501/lesson/13/13.1>
- Sakib, A. S. (2023). USA Real Estate Dataset. <https://www.kaggle.com/datasets/ahmedshahriarsakib/usa-real-estate-dataset/data>
- Zach Robbitt. (2020). Linear Regression Assumptions. <https://www.statology.org/linear-regression-assumptions/>
- Zietz, J., Zietz, E. N., & Sirmans, C. F. (2008). Determinants of House Prices: A Quantile Regression Approach. *Journal of Real Estate Finance and Economics*, 37(4), 317-333. <https://doi.org/10.1007/s11146-007-9053-7>