

Estudio Descriptivo del Mercado Inmobiliario en Ames, Iowa

Data Science II – Aaron Ruiz Díaz



Agenda del Estudio

1. Abstract: motivación y audiencia
2. Metadata del dataset
3. Preguntas de análisis
4. Dsistribución de precios
5. Relación entre tamaño y precio
6. Variables correlacionales
7. Barrios con precios mas altos
8. Comportamiento de variables
9. Outliers
10. Insights



Motivación y audiencia

En este proyecto realicé un análisis exploratorio del dataset Ames Housing, que contiene información de casas vendidas en la localidad de Ames, Iowa dentro de los Estados Unidos de América.

Audiencia

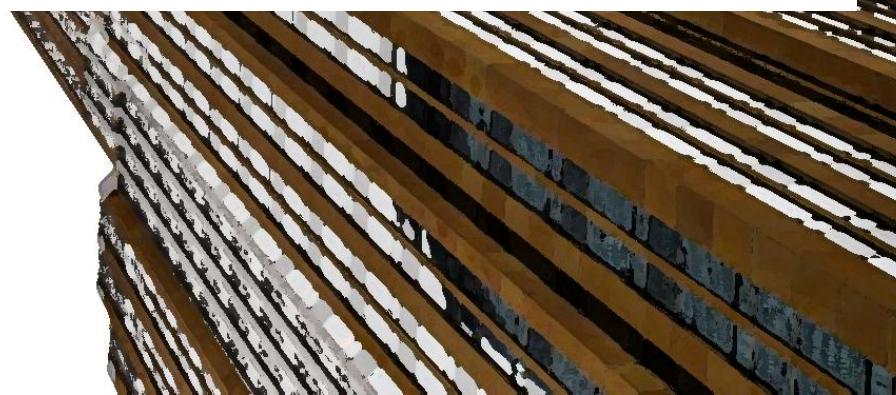
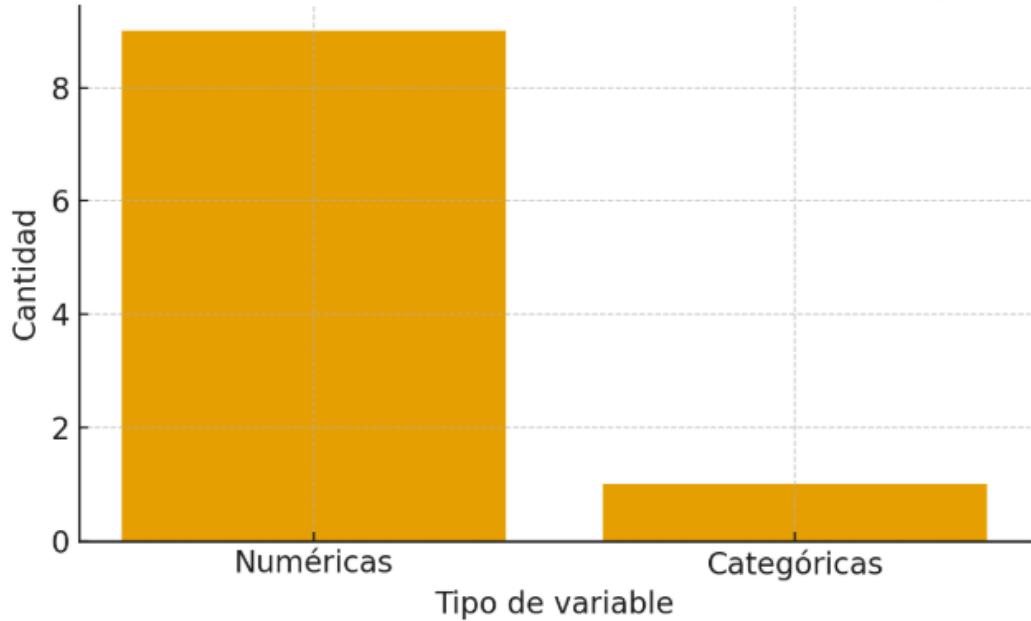
- personas interesadas en estudiar precios de viviendas
- agentes inmobiliarios
- estudiantes que trabajan con datos por primera vez
- cualquier persona que quiera entender qué variables pueden influir en el precio de una casa

Limitaciones

No se incluyen variables económicas del tipo inflación o tasas de interés.



Distribución de tipos de variables en Ames Housing (Reducido)



Metadata

- Año del registro: 2006–2010
- Unidad de análisis: propiedades residenciales de la ciudad de Ames, I
- Fuente original: Ames Assessor's Office
- Formato del dataset: CSV
- Variables incluidas en la versión reducida:
 - Información estructural (ej.: superficie, número de habitaciones)
 - Calidad y estado de la vivienda
 - Variables de ubicación
 - Precio de venta (*SalePrice*)



Preguntas

- 1- ¿Cómo se distribuyen los precios de las casas?
- 2- ¿Qué relación hay entre el tamaño de la casa y el precio?
- 3- ¿Cuáles son las variables numéricas más correlacionadas con el precio de venta?
- 4- ¿Qué barrios (Neighborhood) tienden a tener precios más altos?
- 5- ¿Cómo se comportan las principales variables en general? (superficie, lotes, calidad)
- 6- ¿Hay valores atípicos (outliers) importantes en los precios?

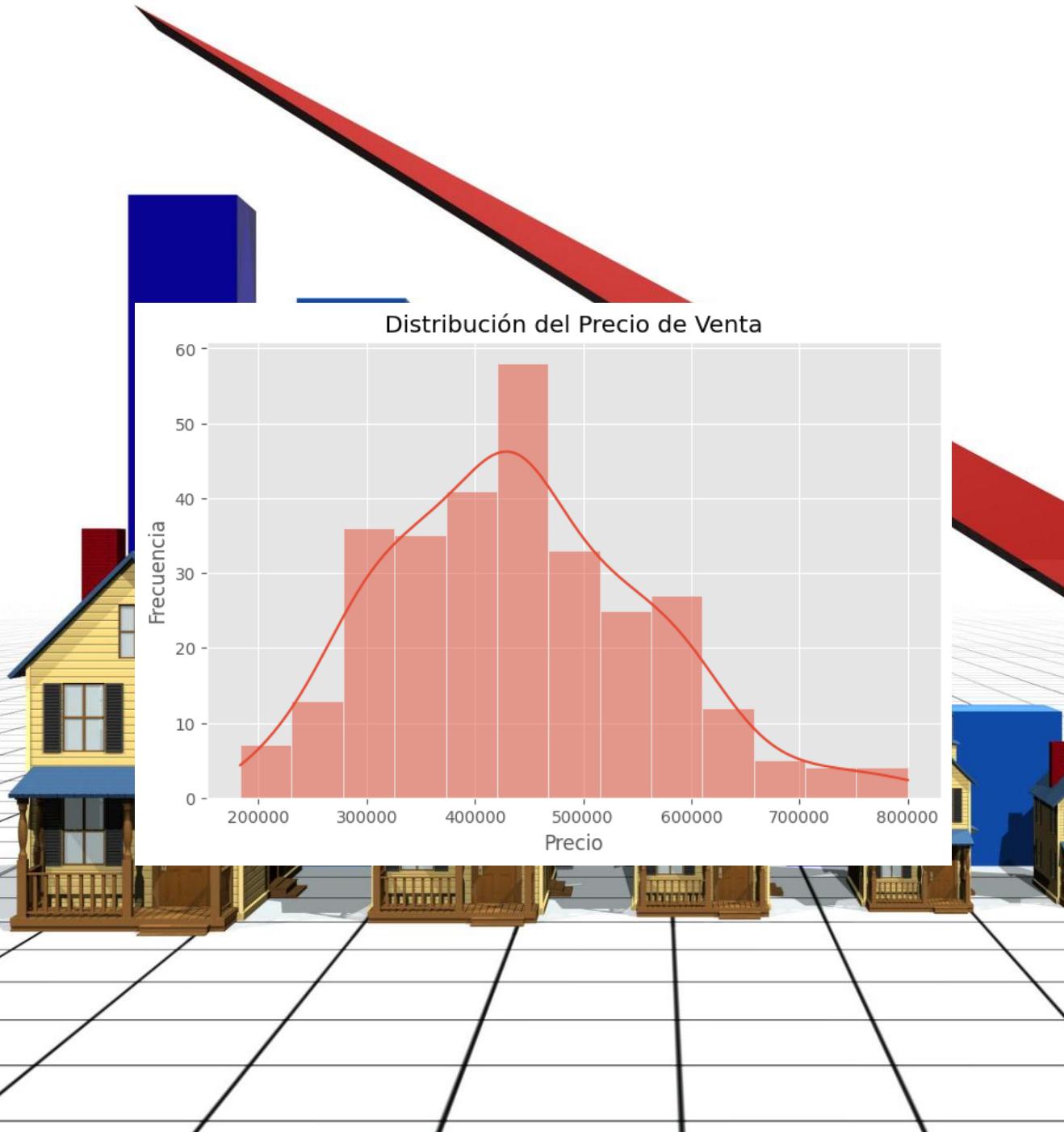
Distribución de precios

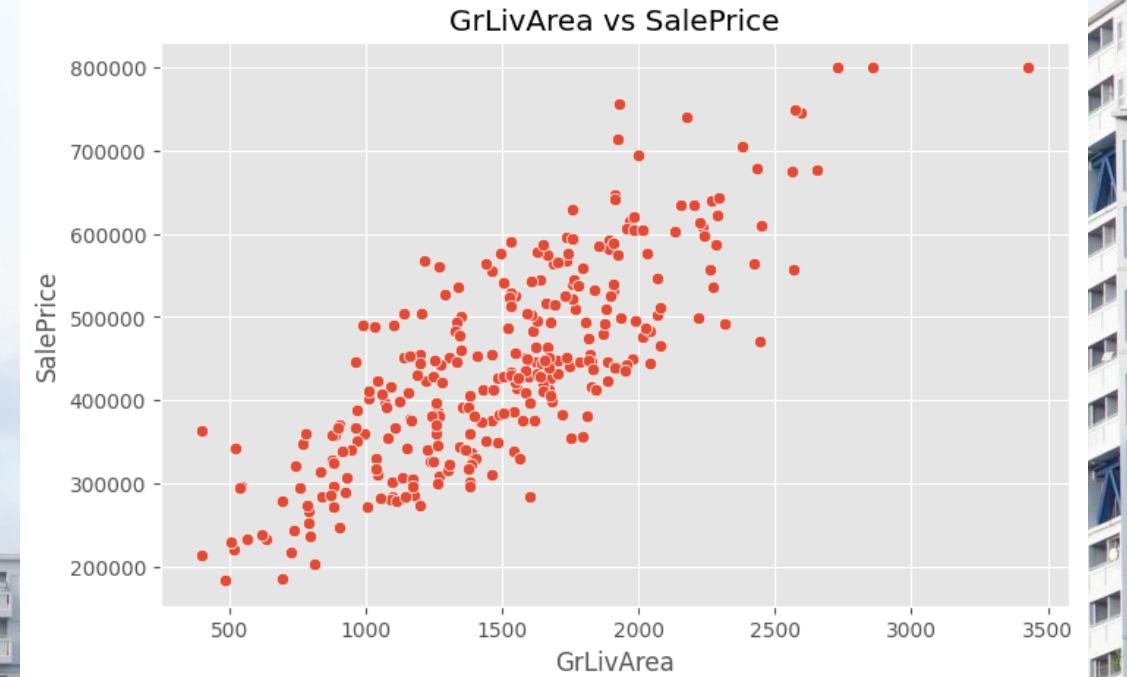
Los precios de venta muestran una distribución entre los **300.000 y 500.000 USD**, donde se encuentra la mayor parte del mercado analizado.

El pico de frecuencia está alrededor de los **400.000–450.000 USD**, lo que indica que este rango representa el valor “típico” de las casas en el dataset.

A medida que los precios aumentan por encima de los **550.000 USD**, la frecuencia disminuye. Esto sugiere que las propiedades muy costosas son menos comunes y podrían representar casas de mayor calidad o ubicadas en zonas específicas.

En resumen, el mercado se comporta como uno moderadamente concentrado, con pocas propiedades extremas hacia valores muy altos.





Relación entre tamaño y precio

El gráfico muestra una relación claramente positiva entre la superficie habitable y el precio de venta: a medida que la casa es más grande, su valor tiende a aumentar.

Pero también se observa que para superficies parecidas puede haber diferencias en el precio, lo cual indica que no solo importa el tamaño, sino también factores como la calidad de construcción, el barrio o características internas.

Aun así, la tendencia general es que las casas de mas grandes suelen alcanzar precios más altos, lo que confirma el rol fundamental del espacio habitable como determinante del valor de mercado.



Variables mas correlacionadas con el precio de venta

Del análisis de correlación surge que las variables más asociadas al precio de venta son:

GrLivArea (Área habitable)

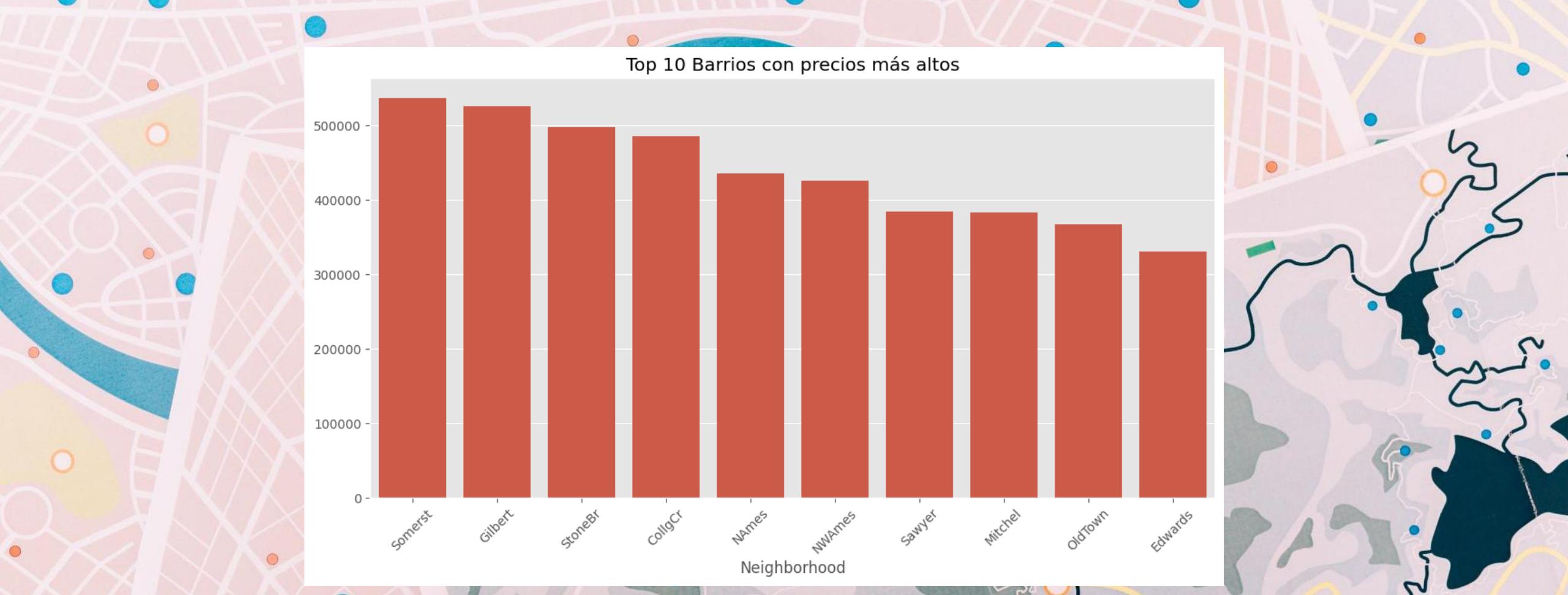
OverallQual (Calidad general de la vivienda)

TotalBsmtSF (Superficie total del sótano)

GarageArea (Área del garage)

Estas variables muestran correlaciones fuertes y positivas, lo que indica que afectan directamente al valor final de la propiedad. En particular, la calidad general aparece como uno de los factores más determinantes, incluso más que el tamaño en algunos rangos.

Esto confirma que el precio depende tanto del espacio disponible como del nivel de construcción y terminaciones.



Barrios con precios mas altos

El análisis por barrios revela que algunas zonas presentan valores consistentemente más altos. Los barrios con precios superiores se destacan por casas más grandes y de mejor calidad, mientras que los de menor precio muestran mayor variabilidad y propiedades más modestas.

La diferencia entre barrios ilustra cómo la ubicación sigue siendo uno de los factores clave en el valor inmobiliario, y cómo el entorno puede elevar o limitar el precio incluso cuando la estructura de la casa es similar.

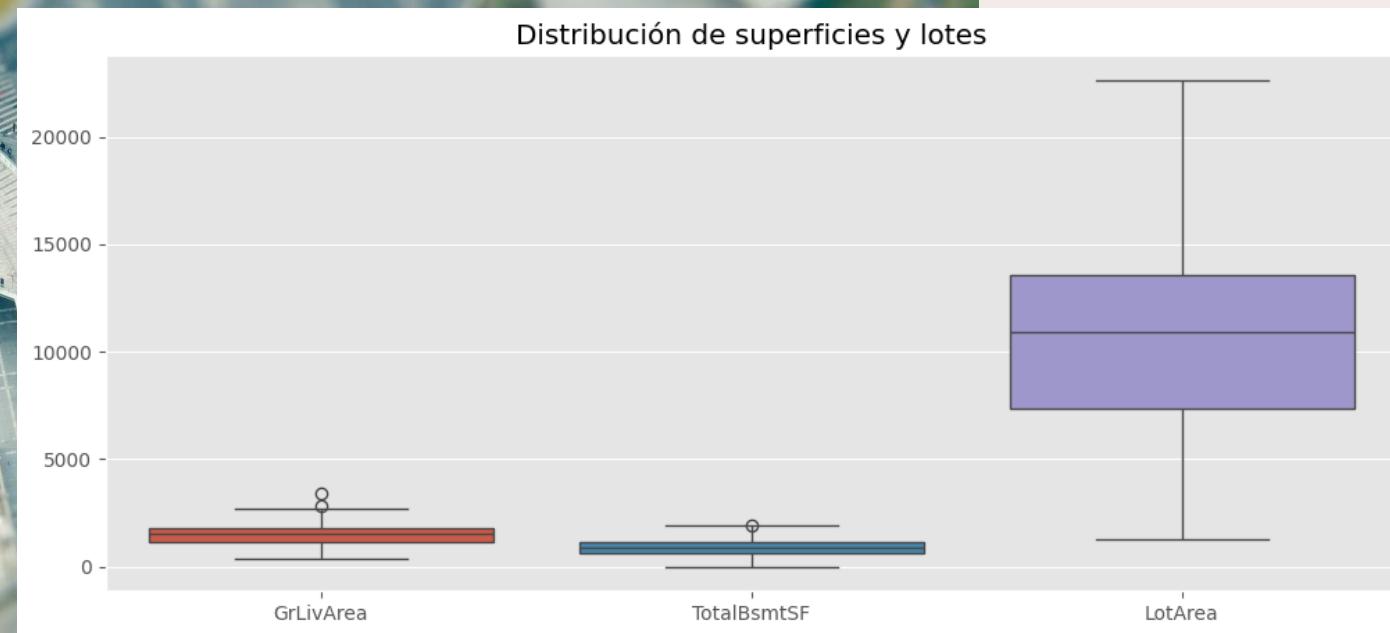
Comportamiento de las variables principales

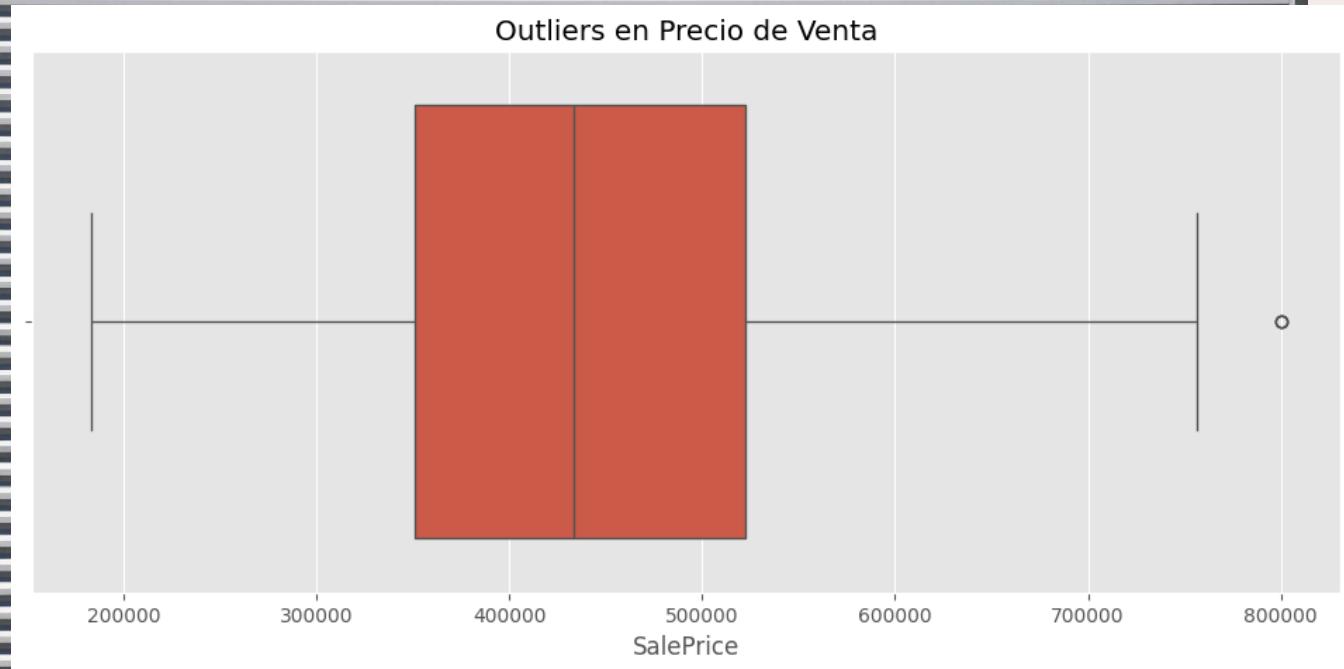
Las principales variables muestran distribuciones relativamente equilibradas, aunque con ciertas tendencias:

La superficie habitable y el lote presentan una mayoría de valores en rangos medios, con algunos grupos de casas más grandes que elevan el rango superior.

La calidad general (OverallQual) tiende a concentrarse en niveles intermedios, pero con un grupo claro de propiedades de alta calidad que influencian notablemente el precio.

En conjunto, estas variables muestran un mercado compuesto mayoritariamente por viviendas estándar, complementado por un segmento más premium que se destaca por mejores terminaciones y mayor superficie.





Hay valores atípicos importantes en los precios?

Sí, el análisis revela algunos valores atípicos hacia precios muy altos.

Estos outliers representan propiedades excepcionales dentro del mercado: son más grandes, de alta calidad o ubicadas en barrios de mayor valor.

Si bien son pocos, pueden influir significativamente en métricas estadísticas como la media o las correlaciones si no se los considera adecuadamente.

Estos casos permiten identificar el segmento premium del mercado, distinto del rango típico donde se concentra la mayoría de las viviendas.

Insights

Insights generales

La variable GrLivArea (superficie habitable) tiene una correlación fuerte y positiva con el precio (SalePrice).

Los barrios NAmes y CollgCr muestran los precios medianos más altos.

Algunas casas presentan precios atípicamente altos (outliers), lo cual podría influir en análisis futuros.

El precio por m² (Price_per_sqft) permite comparar casas de distintos tamaños de forma más justa.

El dataset tiene algunas columnas con muchos valores faltantes, que fueron eliminadas para evitar sesgos.

Pros

El análisis permite identificar rápidamente barrios más caros y variables que afectan al precio.

Las transformaciones de ingeniería de atributos facilitan la comparación entre casas de distintos tamaños.

La visualización con mapas de calor y scatterplots ayuda a entender correlaciones y relaciones clave.

Contras

No se realizó imputación de valores faltantes, solo eliminación de columnas. Esto podría hacer que se pierda información.

Algunos insights podrían cambiar si se usan datos más completos o actuales.

El análisis es descriptivo, no predictivo; no permite estimar precios de nuevas casas