

# Análisis de Factores que Influyen en el Precio por Metro Cuadrado de Viviendas

Moises León

July 3, 2023

- 1 Introducción
- 2 Factores que influyen en el precio por metro cuadrado
- 3 EDA
- 4 Otros métodos implementados
- 5 Recursive Feature Elimination (RFE)
- 6 Conclusiones

- El precio por metro cuadrado de una vivienda es un factor clave en el mercado inmobiliario.
- Diversos factores influyen en el valor por metro cuadrado de una propiedad.
- En esta presentación, analizaremos algunos de los factores más importantes que afectan el precio por metro cuadrado de viviendas.
- Abordaremos el uso de diferentes técnicas para el análisis de bases de datos.

# Factores que influyen en el precio por metro cuadrado

- **Ubicación:** La ubicación de una vivienda es uno de los factores más influyentes en su precio por metro cuadrado. Las propiedades ubicadas en áreas urbanas o zonas con alta demanda suelen tener precios más elevados.



Figure: Ubicacion

# Factores que influyen en el precio por metro cuadrado

- **Tamaño y distribución:** El tamaño total de la propiedad y la distribución de los espacios también impactan en el precio por metro cuadrado. Las viviendas con una distribución bien diseñada y mayor número de habitaciones suelen tener un mayor valor por metro cuadrado.



Figure: Distribución

# Factores que influyen en el precio por metro cuadrado

- **Estado de conservación:** El estado de mantenimiento y conservación de la propiedad influye en su precio. Las viviendas en buen estado y con renovaciones recientes tienden a tener un precio por metro cuadrado más alto que aquellas que requieren reparaciones o remodelaciones.



Figure: Conservación

# Factores que influyen en el precio por metro cuadrado

- **Amenidades y servicios:** La disponibilidad de comodidades y servicios dentro o cerca de la vivienda puede influir en su precio. Por ejemplo, la presencia de una piscina, gimnasio, seguridad las 24 horas o estacionamiento subterráneo puede aumentar el valor de la propiedad.



#2586717

Figure: Amenidades

# Factores que influyen en el precio por metro cuadrado

- **Nivel de demanda:** La oferta y demanda del mercado inmobiliario en una determinada zona también afecta el precio por metro cuadrado. Si hay una alta demanda de viviendas en una ubicación específica y una oferta limitada, los precios tienden a subir.



Figure: Oferta-Demanda



# Factores que influyen en el precio por metro cuadrado

- **Características adicionales:** Otros factores que pueden influir en el precio por metro cuadrado incluyen la orientación de la propiedad, la calidad de los materiales utilizados en la construcción, la edad del edificio, la accesibilidad para personas con discapacidades y la existencia de áreas verdes o espacios al aire libre.



Figure: Materiales de construcción

# Factores que influyen en el precio por metro cuadrado

- **Condiciones económicas:** Las condiciones económicas generales del país o región también pueden tener un impacto en el mercado inmobiliario. Durante períodos de crecimiento económico, los precios pueden aumentar, mientras que en épocas de recesión, los precios pueden estabilizarse o incluso disminuir.



Figure: Condiciones económicas

# Análisis Exploratorio de Datos (EDA)

- El análisis exploratorio de datos es una etapa inicial y crucial en cualquier proyecto de análisis de datos.
- En este proyecto, se llevó a cabo un análisis exploratorio para comprender mejor el conjunto de datos de precios de viviendas.
- Algunas de las tareas realizadas en el EDA incluyen:
  - Carga y revisión del conjunto de datos para identificar su estructura y características.
  - Identificación y manejo de valores faltantes o datos atípicos.
  - Visualización de distribuciones y relaciones entre variables numéricas.
  - Análisis de correlación para identificar relaciones entre el precio por metro cuadrado y otras variables.

# Análisis Exploratorio de Datos (EDA)

Método/Técnica	Descripción
Resumen estadístico	Obtener medidas descriptivas básicas como la media, mediana, desviación estándar, mínimo y máximo de las variables numéricas.
Histogramas	Representación gráfica de la distribución de una variable numérica. Permite identificar la forma de la distribución y la presencia de valores atípicos.
Diagramas de dispersión	Gráfico que muestra la relación entre dos variables numéricas. Permite identificar si existe una correlación lineal entre ellas.

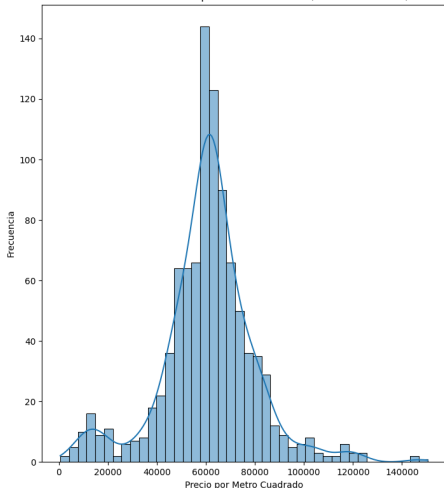
# Métodos para el Análisis Exploratorio de Datos

Método/Técnica	Descripción
Gráficos de cajas y bigotes (boxplots)	Representación visual de la distribución y los valores atípicos de una variable numérica o varias categorizadas por una variable categórica.
Matriz de correlación	Tabla o mapa de calor que muestra la correlación entre pares de variables numéricas. Valores cercanos a 1 indican una correlación positiva, valores cercanos a -1 indican una correlación negativa, y valores cercanos a 0 indican una correlación débil.

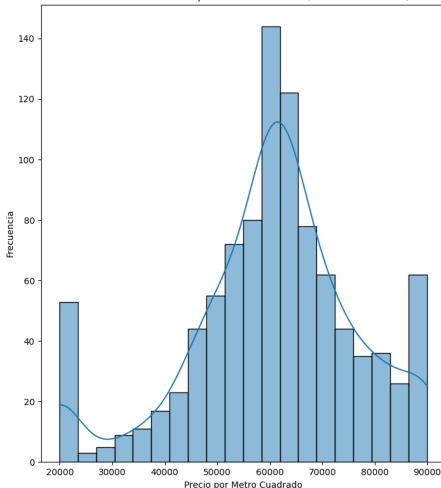
Método/Técnica	Descripción
Correlación de Pearson	Mide la relación lineal entre dos variables numéricas. Toma valores entre -1 y 1. Una correlación de 1 indica una correlación positiva perfecta, -1 indica una correlación negativa perfecta, y 0 indica que no hay correlación.
Correlación de Spearman	Calcula la correlación entre dos variables manteniendo el orden de los datos, no asumiendo una relación lineal. Es adecuado para datos no lineales o con valores atípicos.
Correlación de Kendall	Similar a la correlación de Spearman, pero es más adecuada para datos categóricos o de rangos.

# Resultados EDA- Histogramas

Distribución del Precio por Metro Cuadrado (Sin Winsorization)

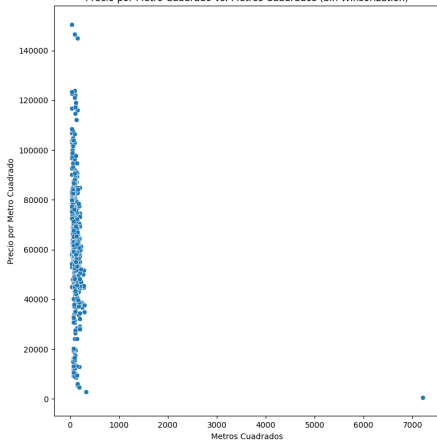


Distribución del Precio por Metro Cuadrado (Con Winsorization)

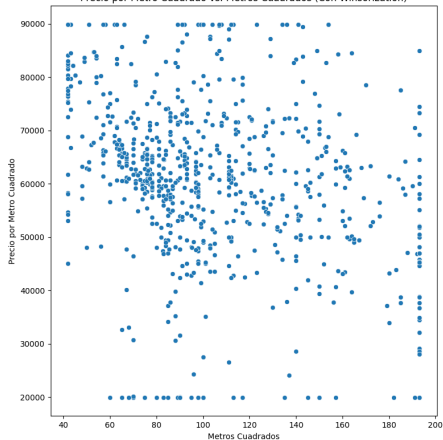


# Resultados EDA- Diagrama de dispersión

Precio por Metro Cuadrado vs. Metros Cuadrados (Sin Winsorization)

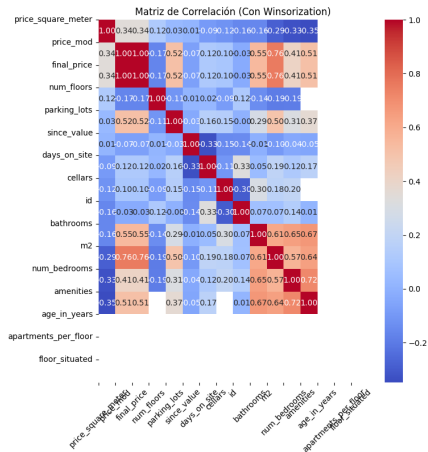
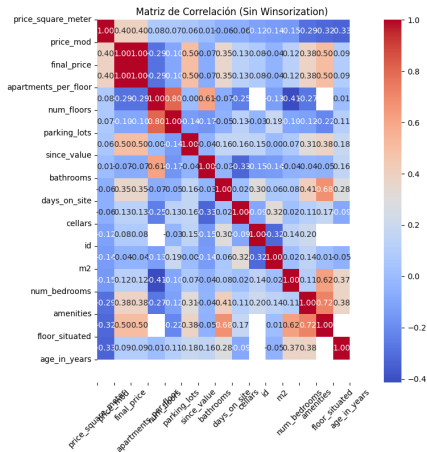


Precio por Metro Cuadrado vs. Metros Cuadrados (Con Winsorization)





# Resultados EDA- Mapa de calor



# Métodos de ML para encontrar correlaciones

Método	Tipo de Método	Uso	Ventajas	Limitaciones
Regresión Lineal	Supervisado	Relación lineal	Fácil interpretación	Solo captura relaciones lineales
Árboles de Decisión	Supervisado	Relaciones no lineales	Fácilmente interpretables	Puede sobreajustar datos
Random Forest	Supervisado	Clasificación y regresión	Alta precisión y robustez	Difícil de interpretar
SVM	Supervisado	Clasificación y regresión	Eficaz en datos de alta dimensionalidad	Puede requerir ajuste de hiperparámetros
Redes Neuronales	Supervisado	Clasificación y regresión	Capacidad para aprender relaciones complejas	Requiere grandes conjuntos de datos y tiempo
PCA	No supervisado	Reducción de dimensionalidad	Resumen de información y correlaciones	Pérdida de interpretación
Selección de Características	Supervisado	Identificar características relevantes	Mejora la eficiencia y precisión de modelos	Depende de algoritmos y técnicas específicas

**Table:** Métodos de Machine Learning para encontrar correlaciones

# PCA (Principal Component Analysis) - Análisis de Componentes Principales.

Es importante recordar que la interpretación de los resultados de PCA puede ser más compleja que en otros modelos, ya que los componentes principales son combinaciones lineales de las características originales. Por lo tanto, es necesario realizar un análisis detallado y contextualizar los resultados para obtener conclusiones significativas sobre los factores que influyen en el precio por metro cuadrado en las propiedades.

# Resultados PCA

```
mleon@Leon-Legion:~/Escritorio/Retos/Oblatorio$ python3 Reto_PCA.py
Características más influyentes en el precio por metro cuadrado:
price_mod          5.041469e-01
final_price        5.041469e-01
parkingLots        3.594495e-01
num_bedrooms       3.514549e-01
amenities           3.077422e-01
bathrooms           2.979233e-01
days_on_site       1.366633e-01
m2                  1.045920e-01
price_square_meter  1.025155e-01
since_value         7.964706e-02
lat                 5.961183e-02
lon                 4.492221e-02
apartments_per_floor 2.869455e-02
id                  2.829212e-02
cellars             2.640025e-02
num_floors           2.583234e-02
floor_situated       2.132238e-02
age_in_years         2.168404e-19
Name: 0, dtype: float64
```

```
mleon@Leon-Legion:~/Escritorio/Retos/Oblatorio$ python3 Reto_PCA.py
Características más influyentes en el precio por metro cuadrado:
price_mod          0.504147
final_price        0.504147
parkingLots        0.359449
num_bedrooms       0.351455
amenities           0.307742
bathrooms           0.297923
days_on_site       0.136663
m2                  0.104592
price_square_meter  0.102515
since_value         0.079647
lat                 0.059612
lon                 0.044922
apartments_per_floor 0.028695
id                  0.028292
cellars             0.026400
num_floors           0.025832
floor_situated       0.021322
age_in_years         0.000000
Name: 0, dtype: float64
```

Figure: Resultados prueba con PCA.

Para aplicar selección de características en lugar de PCA, podemos utilizar un método de selección de características univariado como la prueba F o la correlación de Pearson para determinar qué características tienen una mayor relación con la variable objetivo (precio por metro cuadrado en este caso).

# Resultados

```
mleon@Leon-Legion:~/Escritorio/Retos/Oblatorio$ python3 Reto_Caracteristicas.py
Características más influyentes en el precio por metro cuadrado (por correlación):
price_square_meter      1.000000
final_price             0.395871
price_mod               0.395871
num_bedrooms            0.292771
amenities               0.172156
m2                      0.152549
id                      0.138673
lat                     0.110886
lon                     0.075853
days_on_site           0.061991
parking_lots            0.057810
floor_situated          0.057399
bathrooms              0.055017
num_floors              0.021049
cellars                 0.019618
apartments_per_floor    0.006656
since_value             0.006627
age_in_years            NaN
Name: price_square_meter, dtype: float64
```

Figure: Resultados selección de características

En este código, primero imputamos los valores faltantes usando la media de cada columna. Luego, calculamos la matriz de correlación entre todas las características y la variable objetivo (precio por metro cuadrado). Finalmente, ordenamos las características según su correlación absoluta con el precio por metro cuadrado para determinar cuáles tienen una mayor influencia en la variable objetivo.

Es importante tener en cuenta que la selección de características univariada como esta no tiene en cuenta las relaciones complejas entre las características y puede pasar por alto interacciones importantes entre ellas. Si deseas una selección de características más sofisticada que tome en cuenta interacciones y relaciones no lineales, puedes explorar métodos de selección de características basados en modelos o en técnicas de aprendizaje automático más avanzadas.



Utilizando el método Recursive Feature Elimination (RFE) con un modelo de regresión lineal. Este método utiliza la eliminación recursiva de características para seleccionar las características más importantes de manera iterativa.

# Recursive Feature Elimination (RFE)

- En este ejemplo, primero cargamos el archivo CSV y realizamos el mismo preprocesamiento que antes, imputando los valores faltantes con la media de cada columna.

# Recursive Feature Elimination (RFE)

- Luego, separamos las características ( $X$ ) de la variable objetivo ( $y$ ). A continuación, creamos un modelo de regresión lineal y un selector de características RFE. Especificamos que queremos seleccionar las 5 características más importantes ( $n\_features\_to\_select = 5$ ), pero este valor puede ajustarse según tus necesidades.

# Recursive Feature Elimination (RFE)

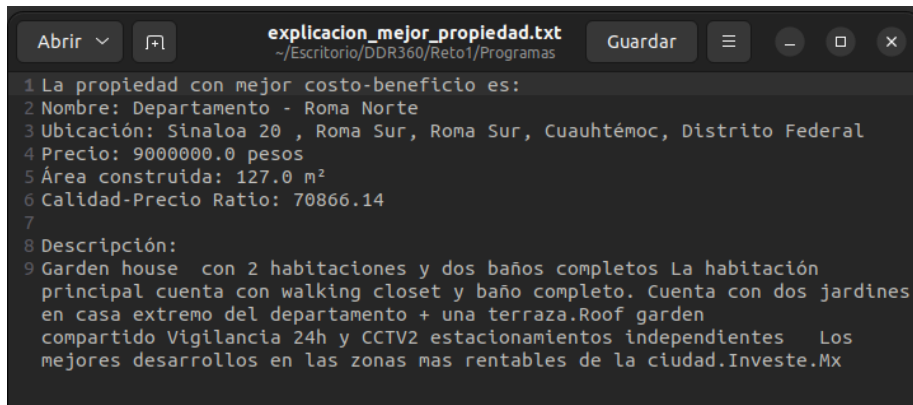
- Después de ajustar el selector RFE con los datos, obtenemos las características seleccionadas (*columnas\_seleccionadas*) y las imprimimos. Estas serán las características más importantes según el método RFE y el modelo de regresión lineal utilizado.

# Recursive Feature Elimination (RFE)

- Recuerda que RFE es solo uno de los muchos métodos de selección de características disponibles en scikit-learn, y dependiendo del problema, otros métodos podrían ser más adecuados. También puedes experimentar con diferentes modelos en lugar de la regresión lineal para obtener diferentes resultados de selección de características.

En resumen, la mejor propiedad se determina utilizando RFE para seleccionar la característica más relevante, que en este caso es el precio por metro cuadrado (`price_square_meter`). Esta característica está relacionada con la relación calidad-precio y, por lo tanto, nos permite identificar la mejor propiedad en términos de valor por metro cuadrado.

# Resultados



A screenshot of a text editor window with a dark theme. The title bar shows the filename 'explicacion\_mejor\_propiedad.txt' and the path '~/Escritorio/DDR360/Reto1/Programas'. The window has buttons for 'Abrir', 'Guardar', and window controls. The text content is as follows:

```
1 La propiedad con mejor costo-beneficio es:  
2 Nombre: Departamento - Roma Norte  
3 Ubicación: Sinaloa 20 , Roma Sur, Roma Sur, Cuauhtémoc, Distrito Federal  
4 Precio: 90000000.0 pesos  
5 Área construida: 127.0 m2  
6 Calidad-Precio Ratio: 70866.14  
7  
8 Descripción:  
9 Garden house con 2 habitaciones y dos baños completos La habitación  
principal cuenta con walking closet y baño completo. Cuenta con dos jardines  
en casa extremo del departamento + una terraza.Roof garden  
compartido Vigilancia 24h y CCTV2 estacionamientos independientes Los  
mejores desarrollos en las zonas mas rentables de la ciudad.Investe.Mx
```

Figure: Resultado análisis RFE

# Mapa de propiedades

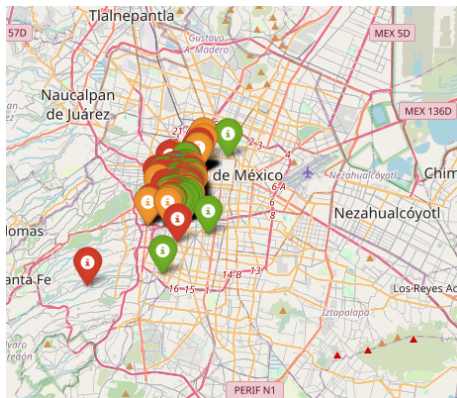
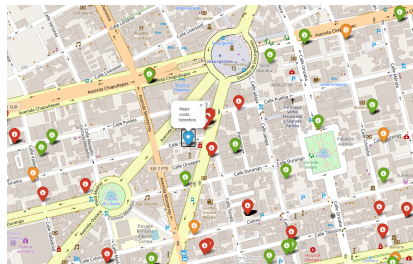
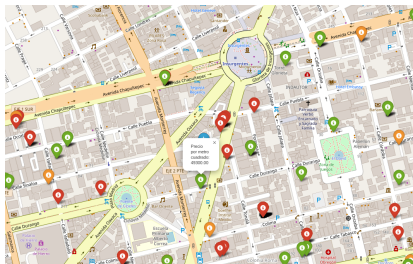
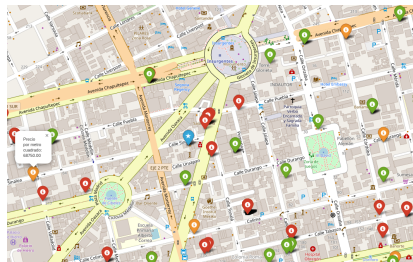
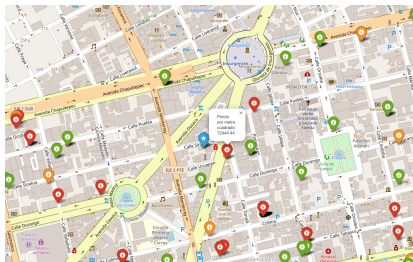


Figure: Mapa de propiedades



# Mapa de propiedades



# Conclusión

- El precio por metro cuadrado de una vivienda es resultado de múltiples factores que interactúan entre sí.
- Al evaluar una propiedad, es fundamental considerar todos estos factores para tomar decisiones informadas en el mercado inmobiliario.
- Cada región y mercado tiene sus particularidades, por lo que es esencial contar con el asesoramiento de expertos en bienes raíces.

¿Preguntas?

Gracias por la oportunidad

Para calcular la mejor propiedad en términos de costo-beneficio, el código utiliza el algoritmo de "Recursive Feature Elimination" (RFE) junto con un modelo de regresión lineal. A continuación, se explico el proceso detallado paso a paso:

- **Normalización de datos:** Primero, se normalizan las características relevantes (m2 y final\_price) utilizando la clase MinMaxScaler de sklearn.preprocessing. La normalización escala los datos a un rango entre 0 y 1, lo que ayuda a que las características tengan la misma escala para el algoritmo de RFE.
- **Selección de características:** A continuación, se crea un objeto LinearRegression que servirá como el modelo de regresión lineal para la selección de características. Luego, se inicializa el objeto RFE con este modelo y se establece n\_features\_to\_select=1, lo que significa que RFE seleccionará la mejor característica.

- **Ejecución de RFE:** Se ejecuta RFE en los datos normalizados. RFE funciona eliminando recursivamente las características menos importantes y ajustando el modelo de regresión lineal en cada iteración hasta que se seleccione el número de características especificado (`n_features_to_select=1`).
- **Obtención del índice de la mejor característica:** Una vez finalizada la ejecución de RFE, se obtiene el ranking de características (ranking) y se busca la posición donde se encuentra el valor 1 en el ranking. Esto corresponderá al índice de la mejor característica seleccionada por RFE.
- **Obtención de la mejor propiedad:** Finalmente, se utiliza el índice obtenido para obtener la fila correspondiente en los datos originales, lo que dará como resultado la información de la mejor propiedad en términos de costo-beneficio.