**Caros Gonzalez**

**Jessica Diniz**

### 

### 

### 

### 

### 

### PRECIO DEL ALQUILER

### EN BARCELONA

**Data Science**

**IronHack**

**27.07.2024**

### I**ntroducción**

En este proyecto, nos proponemos predecir el precio del alquiler en Barcelona en función de varias características determinadas. Utilizamos un conjunto de datos sobre precios de alquiler en Barcelona, que abarca el período desde 2014 hasta 2022. Este conjunto de datos ha sido elaborado y proporcionado por el Ajuntament de Barcelona (Ayuntamiento de Barcelona) y está disponible en Kaggle.

Mediante el uso de varios modelos de Machine Learning, intentaremos predecir el precio del alquiler en Barcelona a lo largo de los años, buscando identificar tendencias y factores clave que afectan el mercado de alquiler en la ciudad.

### 

### Acerca del Conjunto de Datos

Este conjunto de datos incluye información sobre los precios de alquiler en Barcelona, España. Los datos fueron recopilados durante el período de 2014 a 2022, divididos en trimestres. Los precios están organizados por barrios y distritos, lo que permite un análisis detallado de las variaciones geográficas dentro de la ciudad.

Además, este conjunto de datos incluye tanto los precios mensuales como los precios por metro cuadrado, facilitando así la comparación entre diferentes inmuebles y áreas. Esto nos permitirá obtener una visión más precisa del mercado de alquiler en Barcelona y analizar las tendencias a lo largo del tiempo.

Los datos se cargan desde un archivo CSV que contiene 4622 filas y 5 variables de diferentes tipos. Estas variables incluyen el año, la región y los precios.

### **Análisis del Negocio**

Para comprender mejor la repercusión de cada variable sobre el negocio y la variable objetivo, es crucial analizar los rangos, valores medios y otras características del conjunto de datos. Nuestro conjunto de datos consta de seis variables sin elementos faltantes ni nulos, divididas en tres variables cuantitativas y dos variables cualitativas.

La variable cualitativa dicotómica: “Average Rent”estaba dividida en dos categorías y ambas con 2311 filas o registros:

* **Average Rent (euro/month)**: El alquiler promedio en euros por mes.
* **Average Rent per Surface (euro/m²)**: El alquiler promedio por metro cuadrado en euros.

Para facilitar el análisis y la comparación, hemos modificado el conjunto de datos y organizado una nueva columna que representa el precio por metro cuadrado (precio/m²). Esto nos permitirá tener una visión más clara de cómo varían los precios del alquiler según la superficie del inmueble, facilitando la comparación entre diferentes áreas y periodos de tiempo.

Con esta modificación, tenemos las siguientes variables:

#### Variables Cuantitativas

* **Año** (Year): Representa el año en el que se registraron los datos. Rango de datos: 2014 a 2022
* **Trimestre** (Trimester): Indica el trimestre del año en el que se recopilaron los datos. Rango de datos: 1 a 4
* **Precio por mes** (Price/month): El precio del alquiler por mes en euros. Rango de datos: 187 a 2034
* **Precio por metro cuadrado** (Price/meter): El precio del alquiler por metro cuadrado en euros. Rango de datos: 4 a 18

#### Variables Cualitativas

* **Distrito** (District): Variable politómica que indica el distrito de Barcelona al que pertenece el inmueble. Clasificación: 10 distritos
* **Barrio** (Neighbourhood): Variable politómica que especifica el barrio dentro del distrito. Clasificación: 77 Barrios

### **Análisis Exploratorio**

#### Variable Objetivo**:** Precio por Metro Cuadrado (Price/m2)

##### **Tipo de Dato:** La variable objetivo es el precio por m2, que es un dato numérico continuo. Esto significa que el precio por metro cuadrado puede tomar cualquier valor dentro de un rango determinado y se mide en euros por metro cuadrado.

##### **Estadística Descriptiva:** Para la variable objetivo fue calculado las siguientes estadísticas descriptivas por año: Media, máximo, mínimo, mediana y desviación estándar.

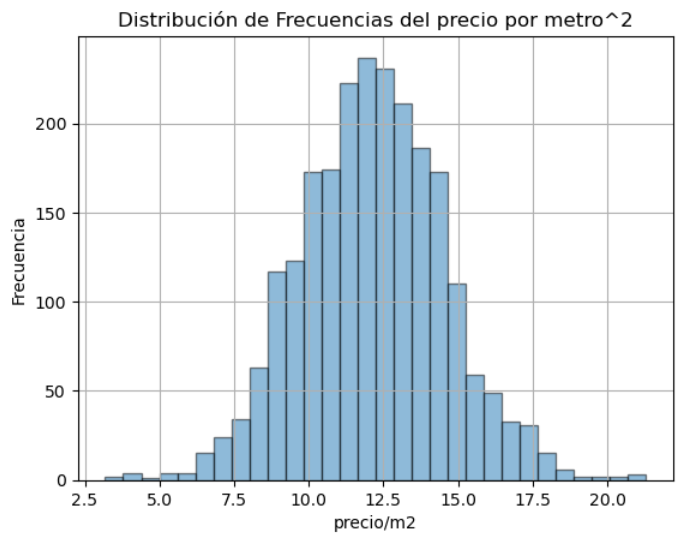
**Media**: El precio medio de alquiler ha aumentado consistentemente de 9.73 en2014 a 13.35 en 2022.

**Máximos y Mínimos**: El precio máximo de alquiler ha aumentado de 16.14 en 2014 a 21.30 en 2022. El precio mínimo no muestra una tendencia tan clara, con fluctuaciones a lo largo de los años.Esto puede se dar por varias razones como, por ejemplo, la nueva ley de alquileres o la influencia del Covid19.

**Mediana de los Precios**: La mediana de los precios sigue una tendencia similar a la media, aumentando de 9.66 en 2014 a 13.40 en 2022.

**Desviación Estándar (std)**: La desviación estándar ha fluctuado, indicando variabilidad en la dispersión de los precios de alquiler. En 2014, la desviación estándar era de 1.76, y alcanzó un pico en 2017 con 2.47, antes de estabilizarse alrededor de 1.70 a 2.27 en los últimos años.

* + **Distribución de frecuencia:**



* Forma: Aproximadamente simétrica con ligera desviación a la derecha.
* Centro: Pico alrededor de 12.5 euros/m², indicando la mayoría de los alquileres.
* Rango: De 2.5 a 20+ euros/m², mostrando amplia variabilidad.
* Dispersión: Baja dispersión con mayor concentración entre 10 y 15 euros/m².
* Frecuencia: Mayor frecuencia cerca de 12.5 euros/m², con algunos valores extremos.

**Prueba de Kolmogorov-Smirnov**: Estadístico=0.020, p=0.323 Distribución probablemente normal

**Variables independientes:** Year, District, Neighbourhood

##### **Tipo de Dato:**

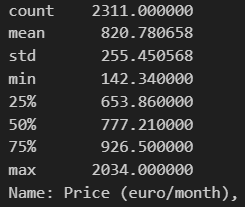
##### **Year**: Numérico (nm)

##### **District**: Categórico (i) (para clasificación)

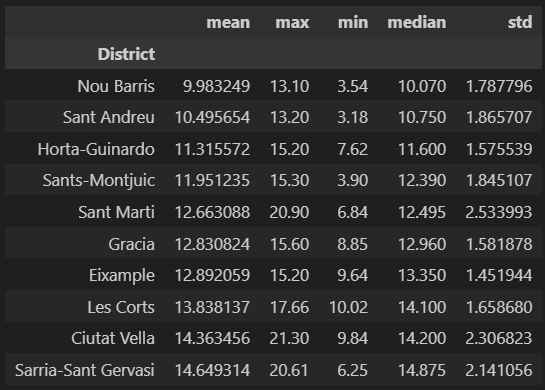
##### **Neighbourhood**: Categórico (i) (para clasificación)

##### **Estadística descriptiva:**

Por precio por mes:



Por precio por m²:



**Medias y Medianas:** Los distritos con precios más altos son Ciutat Vella y Sarrià-Sant Gervasi, con medias de 14.36 €/m² y 14.65 €/m² respectivamente. Nou Barris tiene el precio medio más bajo, con 9.98 €/m².

**Desviación Estándar:** La mayor variabilidad en los precios se observa en Sant Martí (2.53), Ciutat Vella (2.31) y Sarrià-Sant Gervasi (2.14), lo que indica que en estos distritos hay una amplia gama de precios por metro cuadrado. Los distritos con menor variabilidad son Eixample (1.45) y Horta-Guinardó (1.58).

**Rango de Precios (Máximos y Mínimos):** El rango más amplio lo tiene Ciutat Vella, con un mínimo de 9.84 €/m² y un máximo de 21.30 €/m². Sarrià-Sant Gervasi también muestra un rango amplio, con un mínimo de 6.25 €/m² y un máximo de 20.61 €/m².

### Calidad de los datos

#### Presencia de outliers:

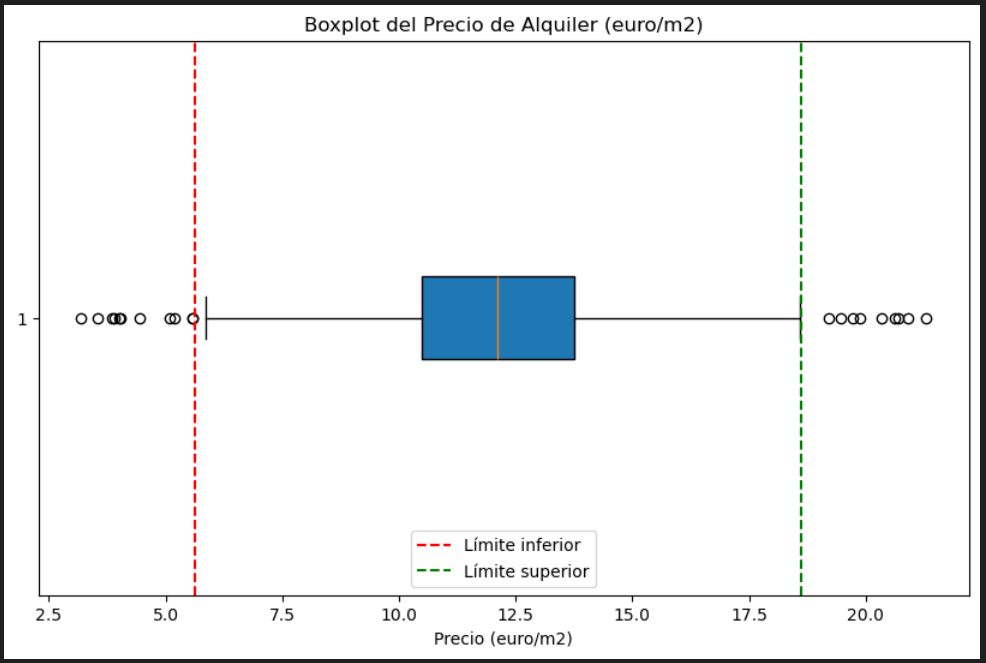
El procesamiento de outliers se realizó dos veces debido a la necesidad de ajustar el modelo.

#### Primer Intento:

#### Identificación de Outliers

Utilizamos el cálculo del rango intercuartílico (IQR) para identificar los valores atípicos.

Encontramos la presencia de 20 outliers en nuestros datos, representados mediante diagramas de cajas y bigotes, como se muestra a continuación.



##### 

##### **Corrección o imputación de outliers:** Inicialmente reemplazamos los outliers con la mediana y obtuvimos este resultado:

#### 

**Después del Tratamiento de Outliers**:

* **Distribución**: La distribución después del tratamiento es más compacta, con precios que oscilan entre 6 y 18 euros/m².
* **Frecuencia Alta**: El pico de frecuencia se mantiene alrededor de 12 euros/m², pero la distribución ahora es más ajustada y con colas menos pronunciadas.
* **Reducción de Outliers**: La eliminación de outliers ha resultado en una distribución más normalizada y menos sesgada por valores extremos.

#### Conclusiones del procesamiento de outliers:

* **Reducción de Variabilidad**: Al sustituir los outliers por la media, se ha conseguido una distribución de precios más uniforme, eliminando valores extremos que podrían influir negativamente en análisis posteriores.
* **Normalización de la Distribución**: La distribución de precios se ha vuelto más simétrica y ajustada alrededor del valor medio, lo que facilita un análisis más preciso y menos sesgado.

En resumen, el tratamiento de outliers ha mejorado la calidad de los datos, ofreciendo una distribución más representativa del mercado de alquiler de pisos en Barcelona, lo que permitirá realizar análisis y predicciones más fiables.

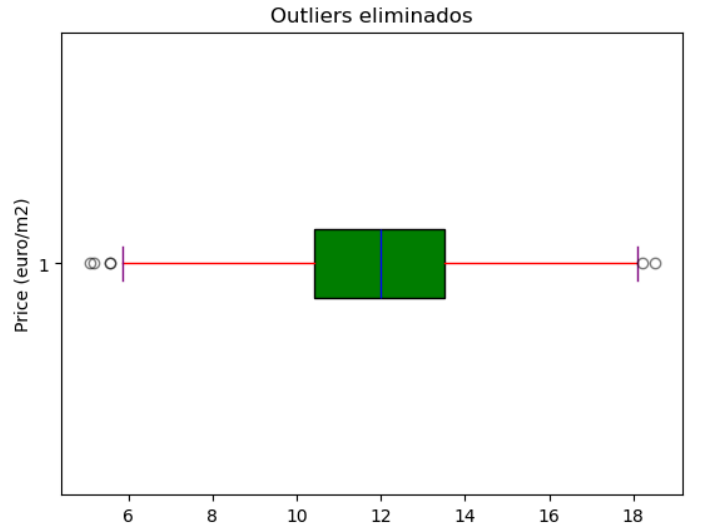
#### Segundo Intento: Corrección de Outliers con K-means

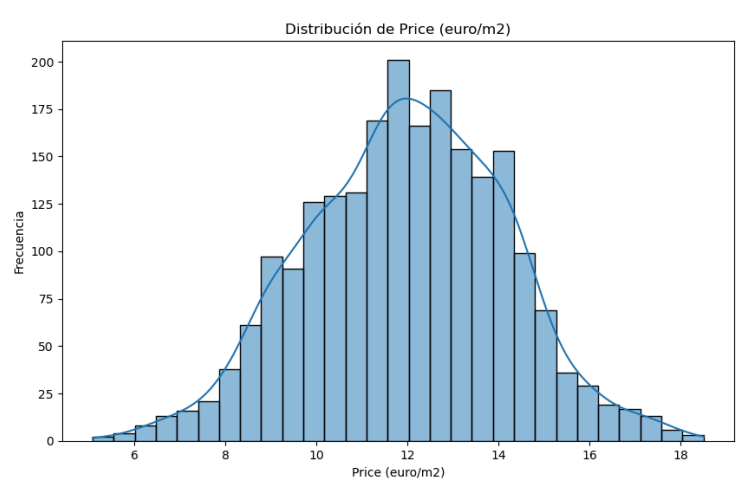
Debido a que el modelo inicial no funcionó adecuadamente, realizamos una segunda corrección de outliers utilizando el algoritmo K-means.

Calculamos el tamaño de los clusters y confirmamos que el número adecuado de clusters es 4.

* **Resultados de la Detección**:
  + **Número de outliers**: 116
  + **Número total de registros**: 2311
  + **Porcentaje de outliers**: 5.02%

Los gráficos a continuación muestran la distribución de precios después de la eliminación de outliers:

* El diagrama de cajas y bigotes indica una distribución más compacta y sin valores extremos.
* El histograma muestra una distribución normalizada de los precios (euro/m²), evidenciando que la eliminación de outliers ha mejorado la calidad de los datos para análisis posteriores.

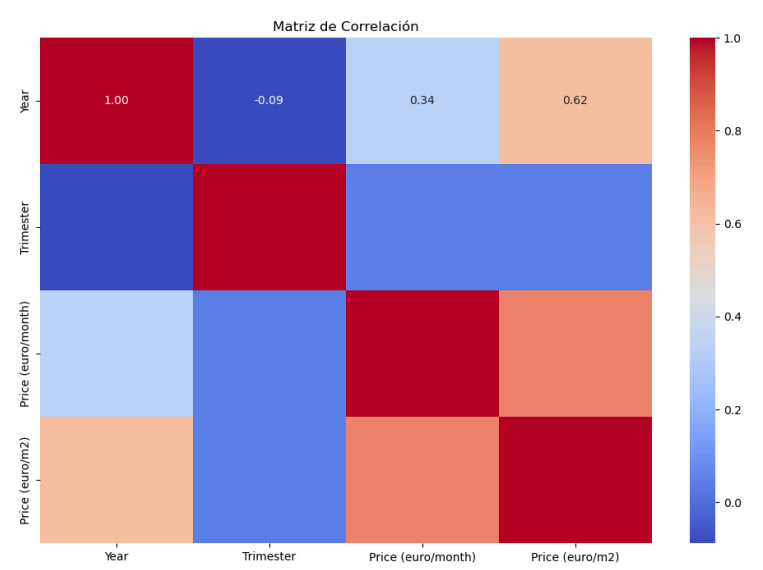


**Prueba de Kolmogorov-Smirnov Estadístico=** 0.019, p= 0.374

Distribución probablemente normal

Al evaluar el modelo tras la eliminación de outliers utilizando K-means, se confirma mediante la prueba de Kolmogórov-Smirnov que la distribución de los datos sigue una distribución normal.

### Correlación de las variables:



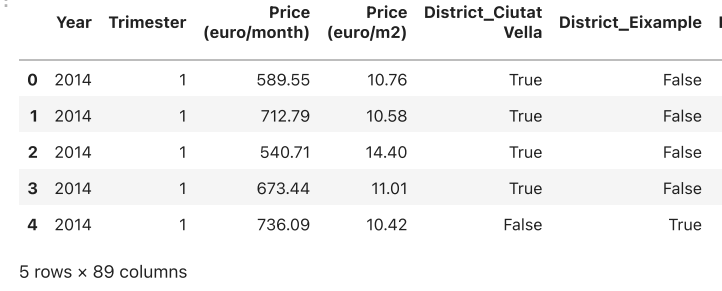
### 

**Price (euro/month) y Price (euro/m2)**: La muy fuerte correlación positiva (0.87) entre los precios por mes y los precios por metro cuadrado indica que ambas medidas de precio se mueven en la misma dirección y están fuertemente relacionadas.

**Year y Price (euro/m2)**: La fuerte correlación positiva (0.62) sugiere que a medida que pasan los años, los precios por metro cuadrado tienden a aumentar. Esto es consistente con la tendencia general de aumento en los precios del mercado inmobiliario.

### Binarización de Categorías:

Para preparar los datos para el aprendizaje automático, fue necesario realizar la binarización de las variables categóricas. Utilizamos la técnica de One-Hot Encoding, que nos permitió transformar las variables categóricas "District" y "Neighbourhood" en un formato numérico. Esto facilita el procesamiento por los algoritmos, mejorando el análisis y el modelado predictivo.



### División Train Test:

Dividimos el conjunto de datos en conjuntos de entrenamiento y prueba con una **proporción de 80-20**. Esto nos permitió entrenar el modelo con el 80% de los datos y evaluarlo con el 20% restante para verificar su rendimiento en datos no vistos.

### Transformación:

Para preparar los datos para Machine Learn y mejorar el rendimiento y la precisión de los modelos, utilizamos la técnica de **MinMaxScaler**. Esta técnica transforma las características para que todas estén dentro de un rango específico, generalmente entre 0 y 1. Al hacerlo, se asegura que ninguna característica domine sobre las otras debido a diferencias en escala, proporcionando un campo de juego equilibrado para los algoritmos de machine learning.

### Modelo, Entrenamiento y Evaluación

Para evaluar el rendimiento del modelo, seguimos los siguientes pasos:

**Primer Intento (Median\_Model):**  Usamos el Modelo de regresión lineal simple

#### Evaluación del Modelo:

#### **‘MAE (Mean Absolute Error)**: 0.6076

#### **MSE (Mean Squaed Error)**: 0.7825

#### **RMSE (Root Mean** r**Squared Error)**: 0.8846

#### **R² (Coeficiente de Determinación)**: 0.8475

**R²**: Este coeficiente de determinación indica qué tan bien los datos observados se explican por las predicciones del modelo. El valor de 0.8475 significa que el modelo explica el 84.75% de la variabilidad de los datos, lo que refleja un buen desempeño del modelo.

#### **Segundo intento (Modelo\_Cluster):** Usamos el Modelo de Ridge

#### **AE (Ridge):** 0.4916088822009097,

#### **MSE (Ridge):** 0.4607188546634674,

#### **RMSE (Ridge):** 0.6787627381224366,

#### **R² (Ridge):** 0.9117413927163492

**R²**: El coeficiente de 0.9117, indica que el modelo ahora explica el 91.17% de la variabilidad de los datos, en comparación con el 84.75% anterior.

#### El ajuste del modelo utilizando K-means para la gestión de outliers ha mejorado considerablemente su desempeño. El valor de R² aumentó de 0.84 a 0.91, lo que refleja una mejora sustancial en la capacidad del modelo para explicar la variabilidad de los datos. Las métricas de error también muestran una reducción significativa, lo que indica una mayor precisión y fiabilidad en las predicciones del modelo.

#### Tercer intento con Feature\_engineering: Usamos el Modelo de Ridge

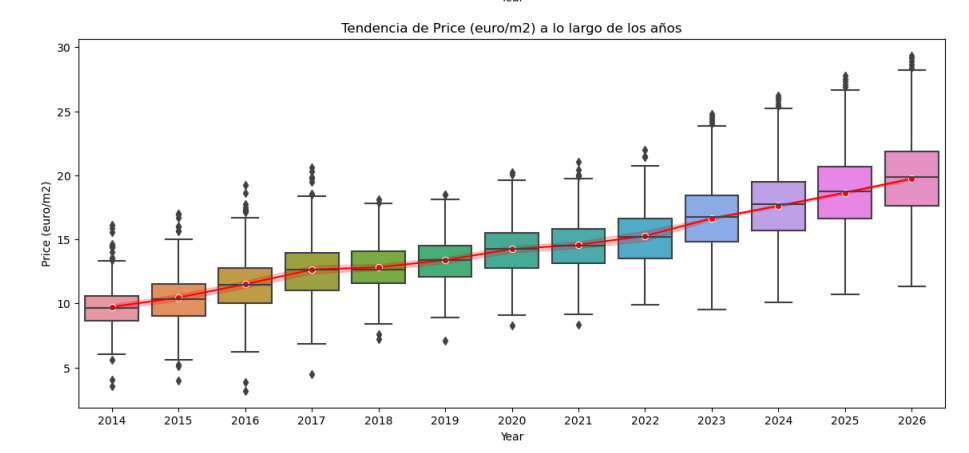
En los modelos de predicción anteriores, se observó que las proyecciones a futuro se "aplanaron" debido a la caída significativa en los precios durante el período de la pandemia de COVID-19, específicamente entre 2020 y 2022. Esta caída en los precios afectó la capacidad del modelo para hacer predicciones precisas más allá de ese período, ya que los patrones de precios durante la pandemia fueron inusuales y no reflejaron las tendencias históricas.

* **Reconstrucción del Dataset con Feature Engineering:** Se llevó a cabo un proceso de ingeniería de características (feature engineering) para ajustar el dataset a las condiciones específicas del período 2020-2022. Esto implicó la creación de nuevas variables o la modificación de las existentes para capturar mejor los efectos de la pandemia en los precios.Se actualizaron y ajustaron los datos para reflejar con mayor precisión los cambios en el mercado durante este período atípico.
* **Aplicación del Modelo de Ridge Regression:** Utilizamos Ridge Regression, una técnica de regresión que incorpora una penalización a los coeficientes del modelo para evitar el sobreajuste.
* **Price (euro/m2):** 
  + **MAE:** 0.4594534772448738,
  + **MSE:** 0.36858901136417416,
  + **RMSE:** 0.6071153196586083,
  + **R²:** 0.9440063761832949
  + **Validación cruzada R²:** 0.9257207163008297, 0.00921761638 814432
* **Price (euro/month):**
  + **MAE:** 48.995374177329715,
  + **MSE:** 4430.829879621476,
  + **RMSE:** 66.56447911327389,
  + **R²:** 0.9002916859612925
  + **Validación cruzada R²**: 0.8939557056686407, 0.01112776 1580973701

**Modelo seleccionado:** Los modelos anteriores no pudieron capturar adecuadamente las variaciones de precios durante la pandemia, lo que resultó en predicciones planas a futuro. Para mejorar la precisión, se reconstruyó el dataset mediante feature engineering y se utilizó Ridge Regression para ajustar el modelo a las condiciones excepcionales del período 2020-2022. Esto ayudará a obtener proyecciones más precisas y ajustadas a las realidades del mercado actual.

### Predicción:

### 



### Conclusiones e Inferencias sobre el Mercado de Alquiler en Barcelona

**1. Tendencias Generales del Mercado**

Aumento Sostenido de los Precios: Desde 2014 hasta 2026, se observa una tendencia ascendente constante en los precios de alquiler por metro cuadrado y por mes. Esto refleja un crecimiento sostenido en el mercado inmobiliario de Barcelona.

Impacto de la Pandemia: Aunque hubo una ligera desaceleración en 2020 y 2021, los precios volvieron a aumentar, indicando una rápida recuperación del mercado tras la pandemia.

**2. Diferencias por Distritos**

Disparidades entre Distritos: Distritos como Ciutat Vella y Sarrià-Sant Gervasi presentan precios significativamente más altos en comparación con distritos como Nou Barris y Sant Andreu. Esto puede deberse a factores como la localización, la oferta de servicios, y la demanda turística.

Influencia del Turismo: Áreas con alta afluencia turística, como Ciutat Vella, tienden a tener precios más elevados, posiblemente debido a la presencia de pisos turísticos que influyen en el mercado de alquiler residencial.

**3. Posibles Factores de Influencia**

Licencias Turísticas: La alta concentración de pisos turísticos puede estar elevando los precios de alquiler en ciertos distritos, reduciendo la oferta de viviendas disponibles para alquiler residencial.

Seguridad Jurídica y Ocupación: La percepción de seguridad jurídica y la incidencia de ocupaciones ilegales pueden influir en la disposición de los propietarios a alquilar sus propiedades, afectando la oferta y, por ende, los precios.

Tasas de Interés Hipotecarias: Un aumento en las tasas de interés hipotecarias podría desincentivar la compra de viviendas, aumentando la demanda de alquiler y, consecuentemente, los precios.

**4. Predicciones Futuras**

Proyección de Aumento: Las predicciones para los años 2023 a 2026 sugieren que los precios de alquiler continuarán aumentando, con una media estimada de alrededor de 19 euros/m2 para 2026.

Inversión en Nuevas Áreas: Los distritos con precios más bajos pueden atraer inversiones futuras, lo que podría equilibrar la disparidad en los precios entre diferentes áreas de la ciudad.

Recomendaciones

Monitoreo Continuo del Mercado: Es crucial seguir monitorizando los datos del mercado de alquiler para identificar cambios en las tendencias y posibles factores disruptivos.

Políticas de Regulación: Implementar políticas que regulen los pisos turísticos y promuevan la oferta de alquiler residencial puede ayudar a estabilizar los precios en distritos con alta demanda turística.

Incentivos para la Vivienda Asequible: Crear incentivos para la construcción y oferta de viviendas asequibles en distritos con precios elevados puede equilibrar la demanda y oferta en el mercado de alquiler.

Análisis de Factores Externos: Continuar analizando cómo factores externos, como las tasas de interés y la seguridad jurídica, afectan el mercado de alquiler puede proporcionar insights valiosos para futuras estrategias.