Análisis Exploratorio de los Datos de Propiedades del portal Fazwaz

1. Descripción General de los Datos

El dataset consiste en 60 registros con información de propiedades listadas en Quito, Ecuador. Los datos incluyen detalles como el precio de la propiedad, el precio por m², el área en m², el número de habitaciones y baños, además de descripciones textuales y enlaces a las páginas de las propiedades.

La estructura de las columnas es la siguiente:

- **Precio**: El precio total de la propiedad (USD).
- **Título**: Título de la propiedad, que generalmente incluye información sobre la ubicación y tipo de propiedad.
- **Ubicación**: Ciudad y ubicación específica de la propiedad (e.g., Quito, Cumbayá).
- **Precio por m²**: Precio por metro cuadrado (USD/m²), que es un indicador clave del valor relativo de la propiedad.
- Área: Tamaño de la propiedad en metros cuadrados.
- Habitaciones: Número de habitaciones disponibles en la propiedad.
- Baños: Número de baños en la propiedad.
- **Descripción Características**: Detalles adicionales sobre las características de la propiedad (e.g., piscina, balcón).
- **Descripción Título**: Descripción corta o resumen de la propiedad.
- Última Actualización: Fecha de la última actualización de la lista de la propiedad.
- Tipo de Propiedad: Indica si es un apartamento, casa, o algún otro tipo.
- Enlace de Propiedad: URL del listado original.
- Extras: Características adicionales, como vistas, acceso a áreas recreativas, y otras amenidades.

2. Análisis de Tipos de Datos

Los tipos de datos se dividen en tres categorías principales:

1. Numéricos:

o Precio: int64

o Precio por m²: int64

Área: float64

Habitaciones: int64

o Baños: float64

2. Categóricos (Texto):

 Título, Ubicación, Descripción Características, Descripción Título, Tipo de Propiedad, Enlace de Propiedad, Extras

3. Tiempo:

 Última Actualización: object (se puede transformar en un formato de fecha si se requiere un análisis temporal más profundo). La presencia de columnas numéricas correctamente formateadas es crucial para un análisis cuantitativo y la aplicación de técnicas estadísticas y de modelado.

3. Análisis Exploratorio de las Variables Numéricas

1. Distribución de Precios:

El precio de una propiedad es uno de los factores más importantes en la compra y venta de inmuebles. Un análisis inicial de la distribución del precio permite comprender cómo se distribuyen los precios de las propiedades en la muestra.

2. Precio por Metro Cuadrado:

El precio por m² es un indicador útil para comparar el valor relativo de las propiedades, independientemente de su tamaño. En muchos casos, las propiedades más pequeñas tienden a tener un precio por m² más alto debido a su ubicación o características específicas.

3. Relación entre Precio y Área:

La relación entre el área de la propiedad y su precio total ayuda a identificar si existe una relación proporcional entre el tamaño de la propiedad y su precio. En muchos mercados inmobiliarios, se espera que exista una relación positiva entre estas dos variables, es decir, a mayor área, mayor precio.

4. Evaluación Inicial de Datos Faltantes

En la evaluación del conjunto de datos, se ha identificado algunas columnas con valores faltantes:

- Extras: Solo dos registros contienen información sobre extras, lo que implica que esta columna tiene muchos valores nulos (NaN). Esto podría deberse a que la mayoría de las propiedades no incluyen detalles adicionales en la fuente de datos original, o bien, que estos detalles no se consideraron en algunos listados.
- Última Actualización: Aunque todos los registros tienen un valor, algunos valores no son exactos o están mal estructurados, por lo que esta columna requiere revisión si queremos realizar un análisis temporal.

Dado que los valores faltantes en estas dos columnas no son críticos para el análisis exploratorio inicial de precios y características físicas de las propiedades, se pueden dejar como están o analizar en profundidad en futuras iteraciones.

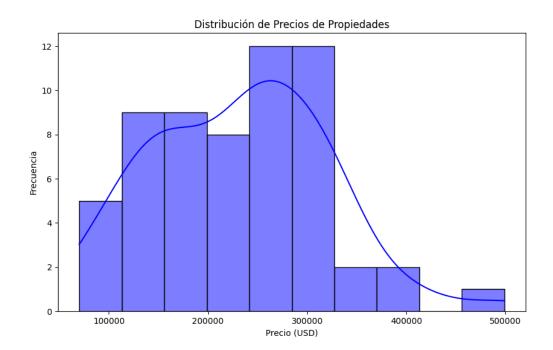
Análisis Visual de las Variables Clave

A continuación, se generarán gráficos que permitirán explorar la distribución de precios, la relación entre variables, y la existencia de valores atípicos.

1. Distribución de Precios

```
plt.figure(figsize=(10, 6))
sns.histplot(df_fazwaz['Precio'], bins=10, kde=True, color='blue')
plt.title('Distribución de Precios de Propiedades')
plt.xlabel('Precio (USD)')
plt.ylabel('Frecuencia')
```

plt.savefig('images/FazWaz_precio_dist.png')
plt.show()



Descripción: Este gráfico muestra la distribución de precios entre las propiedades listadas. Un análisis de esta gráfica nos permitirá ver si hay una concentración de precios en un rango específico o si la distribución es más uniforme.

Observación General: La gráfica de distribución muestra cómo se agrupan los precios totales de las propiedades.

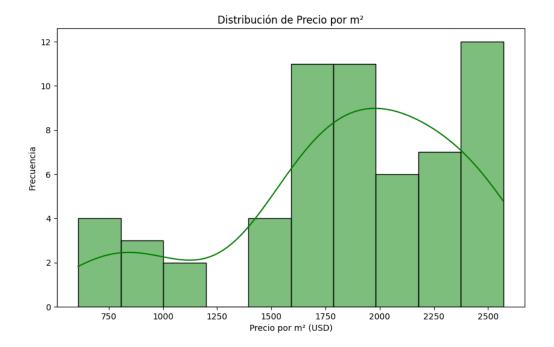
Conclusiones:

- La mayoría de las propiedades se encuentra en el rango de \$150,000 a \$300,000.
 Este rango de precios parece ser el más atractivo para los compradores en el mercado de Fazwaz, reflejando una tendencia hacia propiedades dirigidas a una clase media-alta.
- Propiedades por encima de \$400,000 son escasas, lo que sugiere que este segmento de lujo representa una pequeña proporción del total.
- Propiedades por debajo de \$150,000 también son menos comunes, lo que podría reflejar una falta de oferta en el mercado más económico.

Recomendación: Realizar un análisis de demanda para ver si existe una oportunidad de expansión en los segmentos económicos o de lujo, y estudiar si hay factores que podrían hacer aumentar la oferta de propiedades en estos segmentos.

2. Distribución de Precio por m²

```
plt.figure(figsize=(10, 6))
sns.histplot(df_fazwaz['Precio por m²'], bins=10, kde=True,
color='green')
plt.title('Distribución de Precio por m²')
plt.xlabel('Precio por m² (USD)')
plt.ylabel('Frecuencia')
plt.savefig('images/Fazwaz/Fazwaz_exp_precio_m2_dist.png')
plt.show()
```



Descripción: El precio por m² es un indicador clave del valor del terreno o la propiedad en relación con su tamaño. Este análisis puede mostrar si hay grandes variaciones entre propiedades similares en diferentes ubicaciones.

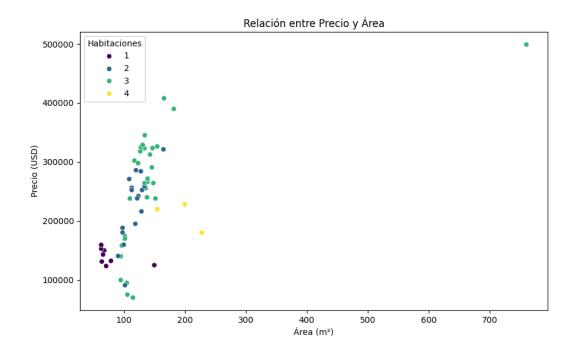
Observación General: El gráfico de barras muestra la distribución de precios por metro cuadrado (USD/m²).

- Los precios por m² varían considerablemente, desde aproximadamente \$750/m² hasta más de \$2,500/m².
- La mayoría de las propiedades tienen un precio por m² que oscila entre los \$1,500 y \$2,500, lo que indica una alta concentración en este rango de precios. Esto puede ser indicativo de áreas que presentan una demanda constante y en crecimiento, especialmente en zonas metropolitanas.
- El segmento con precios más bajos (por debajo de \$1,250/m²) es menos común, posiblemente vinculado a propiedades más antiguas o ubicadas en zonas con menor desarrollo.

Recomendación: Se debe investigar si las propiedades con precios altos por metro cuadrado están relacionadas con factores como vistas exclusivas, zonas de alto valor como Cumbayá, o nuevas construcciones.

3. Relación entre Precio y Área

```
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Área', y='Precio', data=df_fazwaz,
hue='Habitaciones', palette='viridis')
plt.title('Relación entre Precio y Área')
plt.xlabel('Área (m²)')
plt.ylabel('Precio (USD)')
plt.savefig('images/Fazwaz/Fazwaz_exp_precio_vs_area.png')
plt.show()
```



Descripción: Este gráfico nos permite explorar cómo el área de la propiedad impacta el precio total. Es de esperar que las propiedades más grandes tengan un precio mayor, aunque las propiedades con ciertas características (como ubicación o extras) podrían estar fuera de esta tendencia general.

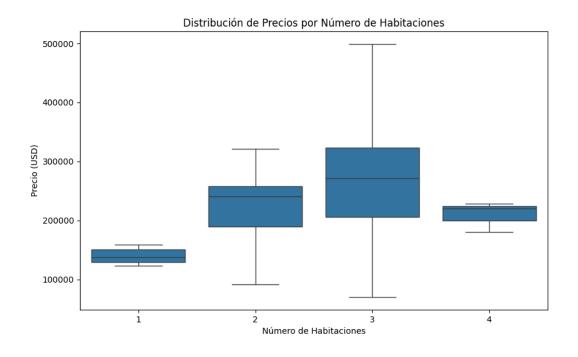
Observación General: La gráfica de dispersión muestra una correlación entre el área de la propiedad y el precio.

- Las propiedades de mayor tamaño, en su mayoría, tienen precios más elevados.
 Esto sigue la lógica esperada de que una mayor área construida requiere una mayor inversión.
- Sin embargo, se observan algunas propiedades grandes con precios relativamente bajos, lo que podría indicar una desvalorización por ubicación o antigüedad de la propiedad.
- Las propiedades con 3 habitaciones, aunque tienden a tener mayor área, muestran una mayor dispersión en cuanto a precio. Esto sugiere que, además del tamaño, otras variables (ubicación, extras) impactan fuertemente en el precio final.

Recomendación: Sería útil realizar un análisis de regresión para cuantificar la relación entre el área y el precio, y entender qué otras variables están contribuyendo significativamente a la variabilidad en los precios.

4. Boxplot de Precios por Número de Habitaciones

```
plt.figure(figsize=(10, 6))
sns.boxplot(x='Habitaciones', y='Precio', data=df_fazwaz)
plt.title('Distribución de Precios por Número de Habitaciones')
plt.xlabel('Número de Habitaciones')
plt.ylabel('Precio (USD)')
plt.savefig('images/Fazwaz/Fazwaz_exp_precio_habitaciones_boxplot.pn
g')
plt.show()
```



Descripción: Este gráfico de caja y bigotes nos permite visualizar la dispersión de los precios dentro de cada categoría de número de habitaciones, identificando posibles outliers que podrían requerir análisis más detallados.

Observación General: En el gráfico de caja (boxplot) se muestra la distribución de precios según el número de habitaciones.

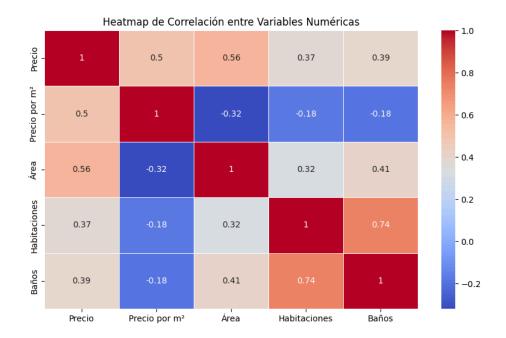
Conclusiones:

- 1 Habitación: Las propiedades con 1 habitación tienden a estar concentradas en un rango de precios relativamente estrecho, entre \$100,000 y \$150,000. La baja dispersión de los datos sugiere que estas propiedades son homogéneas en cuanto a características, lo que podría reflejar la falta de amenidades adicionales o ubicaciones más estándar.
- 2 Habitaciones: Este grupo presenta una mayor dispersión de precios, con valores que oscilan entre \$100,000 y \$300,000. Esta variabilidad sugiere la presencia de propiedades tanto en áreas más accesibles como en ubicaciones más premium o con extras como vistas y amenidades.
- **3 Habitaciones:** Se observa la mayor dispersión en este grupo, con precios que van desde \$200,000 hasta cerca de \$500,000. Las propiedades de 3 habitaciones parecen representar una oferta más amplia y diversa, posiblemente vinculada a distintos segmentos de mercado, desde clase media hasta alta.
- 4 Habitaciones: Aunque el número de datos es menor, las propiedades con 4 habitaciones tienen precios más estables y se concentran alrededor de los \$200,000. Esto podría indicar que este tipo de propiedades son menos comunes y podrían estar ubicadas en áreas más homogéneas en términos de precios.

Recomendación: Se recomienda realizar un análisis más detallado de los factores que explican la dispersión de precios en las propiedades con más habitaciones, como la ubicación exacta, la edad de la propiedad, la presencia de extras (vistas, gimnasio, piscina) y la demanda en el mercado.

5. Heatmap de Correlación entre Variables Numéricas

```
plt.figure(figsize=(10, 6))
correlacion = df_fazwaz[['Precio', 'Precio por m²', 'Área',
   'Habitaciones', 'Baños']].corr()
sns.heatmap(correlacion, annot=True, cmap='coolwarm',
linewidths=0.5)
plt.title('Heatmap de Correlación entre Variables Numéricas')
plt.savefig('images/Fazwaz/Fazwaz_exp_heatmap_correlacion.png')
plt.show()
```



Descripción: Este heatmap nos muestra la relación entre las diferentes variables numéricas en nuestro dataset. Un valor de correlación cercano a 1 indica una correlación fuerte positiva, mientras que un valor cercano a -1 indica una relación fuerte negativa. Este análisis es útil para identificar qué variables están más relacionadas entre sí y cómo influyen en los precios.

Observación General: El heatmap muestra la correlación entre las variables numéricas en el conjunto de datos.

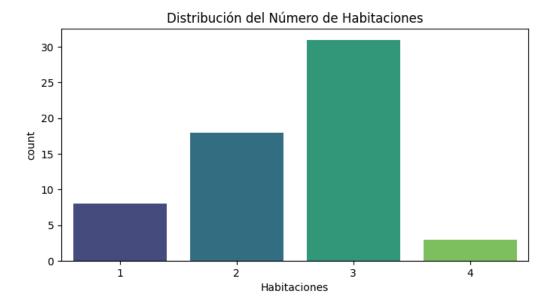
Conclusiones:

- La correlación más fuerte se da entre el precio total y el área (0.56), lo cual era esperable. Sin embargo, la correlación del precio con el número de habitaciones y baños es moderada (0.37 y 0.39 respectivamente), lo que indica que otros factores, como la ubicación y las amenidades, tienen un impacto significativo en el precio.
- La relación entre el precio por metro cuadrado y el área es negativa (-0.32), lo cual sugiere que las propiedades más grandes tienden a tener un precio por m² menor.

Recomendación: El análisis indica que el precio no solo depende del área o las habitaciones, por lo que sería útil realizar un análisis más profundo utilizando modelos de regresión que incluyen variables adicionales como ubicación y extras para predecir el precio.

6. Distribución de Habitaciones y Baños

```
plt.figure(figsize=(8, 4))
sns.countplot(x='Habitaciones', data=df_fazwaz, palette='viridis')
plt.title('Distribución del Número de Habitaciones')
plt.savefig('images/Fazwaz/Fazwaz_exp_distribucion_habitaciones.png')
plt.show()
```



Descripción: Estas gráficas nos permiten visualizar cómo se distribuyen las propiedades en términos de número de habitaciones y baños. Es útil para identificar si hay predominancia de propiedades con cierto número de habitaciones o baños, lo cual puede reflejar el tipo de oferta disponible en el mercado.

Observación General: La gráfica de barras muestra cuántas habitaciones predominan en las propiedades.

Conclusiones:

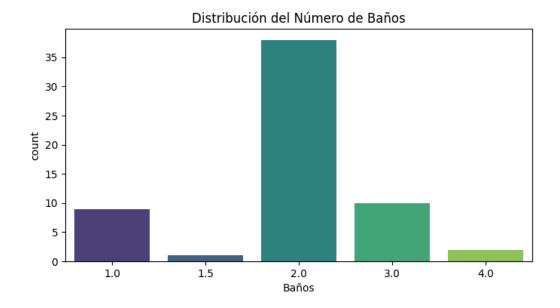
- Las propiedades con 3 habitaciones son las más comunes, lo que probablemente refleje la demanda de familias de tamaño mediano que buscan espacio adicional para oficinas en casa o habitaciones para hijos.
- Las propiedades con **2 habitaciones** son las segundas más comunes, lo que puede estar orientado a parejas o familias pequeñas.
- Las propiedades de 1 habitación y 4 habitaciones son considerablemente menos comunes. Las de 1 habitación pueden estar enfocadas en solteros o parejas sin hijos, mientras que las de 4 habitaciones parecen ser opciones más exclusivas, probablemente destinadas a familias grandes o compradores de propiedades de lujo.

Recomendación:

Explorar más a fondo las características de las propiedades de **4 habitaciones** para entender si están principalmente asociadas con zonas de lujo o si otros factores influyen en

su menor disponibilidad. Por otro lado, también sería útil analizar el impacto del número de habitaciones en el precio total, ya que las propiedades de **3 habitaciones** parecen dominar el mercado.

```
# Distribución de Baños
plt.figure(figsize=(8, 4))
sns.countplot(x='Baños', data=df_fazwaz, palette='viridis')
plt.title('Distribución del Número de Baños')
plt.savefig('images/Fazwaz/Fazwaz_exp_distribucion_baños.png')
plt.show()
```



Observación General: La gráfica de barras muestra la cantidad de baños que predominan en las propiedades.

Conclusiones:

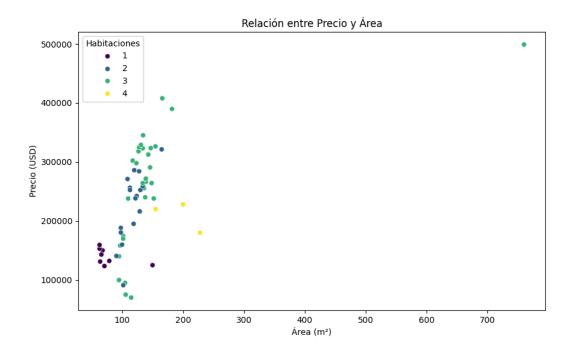
- La mayoría de las propiedades tienen 2 baños, lo que parece ser un estándar en las ofertas actuales, posiblemente reflejando la demanda de viviendas para familias pequeñas o medianas.
- Propiedades con 3 baños también son relativamente comunes, lo que sugiere que existen opciones orientadas a familias más grandes.
- **Propiedades con 1 baño** son menos comunes, lo que podría indicar que las viviendas más pequeñas, como estudios o apartamentos de una habitación, tienen una menor presencia en el mercado.
- Las propiedades con **1.5 o 4 baños** son muy escasas, lo que sugiere que estas configuraciones no son comunes o están orientadas a un mercado muy específico.

Recomendación:

Sería interesante investigar si las propiedades con más de 2 baños ofrecen mayores beneficios en cuanto a precio o si simplemente están dirigidas a un segmento de mercado diferente. Además, explorar si las propiedades con 4 baños están enfocadas en compradores de lujo o si hay factores específicos que explican su menor presencia.

7. Relación entre Precio y Área

```
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Área', y='Precio', data=df_fazwaz,
hue='Habitaciones', palette='viridis')
plt.title('Relación entre Precio y Área')
plt.xlabel('Área (m²)')
plt.ylabel('Precio (USD)')
plt.savefig('images/Fazwaz/Fazwaz_exp_precio_vs_area.png')
plt.show()
```



Descripción: Este gráfico nos permite explorar cómo el área de la propiedad impacta el precio total. Es de esperar que las propiedades más grandes tengan un precio mayor, aunque las propiedades con ciertas características (como ubicación o extras) podrían estar fuera de esta tendencia general.

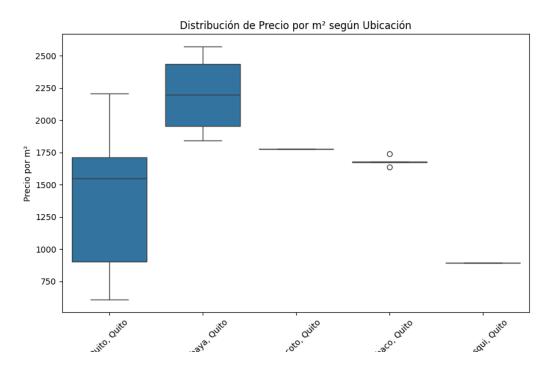
Observación General: La gráfica de dispersión muestra una correlación entre el área de la propiedad y el precio.

- Las propiedades de mayor tamaño, en su mayoría, tienen precios más elevados.
 Esto sigue la lógica esperada de que una mayor área construida requiere una mayor inversión.
- Sin embargo, se observan algunas propiedades grandes con precios relativamente bajos, lo que podría indicar una desvalorización por ubicación o antigüedad de la propiedad.
- Las propiedades con 3 habitaciones, aunque tienden a tener mayor área, muestran una mayor dispersión en cuanto a precio. Esto sugiere que, además del tamaño, otras variables (ubicación, extras) impactan fuertemente en el precio final.

Recomendación: Sería útil realizar un análisis de regresión para cuantificar la relación entre el área y el precio, y entender qué otras variables están contribuyendo significativamente a la variabilidad en los precios.

8. Distribución de Precio por Metro Cuadrado según Ubicación

```
plt.figure(figsize=(10, 6))
sns.boxplot(x='Ubicación', y='Precio por m²', data=df_fazwaz)
plt.xticks(rotation=45)
plt.title('Distribución de Precio por m² según Ubicación')
plt.savefig('images/Fazwaz/Fazwaz_exp_precio_m2_vs_ubicacion.png')
plt.show()
```



Descripción: Este boxplot nos muestra cómo varía el precio por metro cuadrado según la ubicación de las propiedades. Esto es esencial para identificar cuáles son las áreas más caras y cómo la ubicación impacta directamente en el precio.

Observación General: La gráfica de caja muestra cómo varía el precio por metro cuadrado (m²) en diferentes zonas de Quito y sus alrededores.

Conclusiones:

- Cumbaya, Quito tiene los precios por m² más altos, con valores entre \$2,000 y \$2,500, lo que refleja su estatus como una de las zonas más exclusivas de la ciudad.
- La Carolina, Quito presenta una mayor dispersión de precios por m², desde alrededor de \$1,000 hasta casi \$2,500, lo que indica una variabilidad considerable dentro de esta área debido a la calidad de las propiedades o su proximidad a lugares de interés
- Nayón, Quito se sitúa en un rango medio, alrededor de los \$1,500 por m², con una baja dispersión.
- Ponciano Alto, Quito y La Mariscal, Quito presentan los precios más bajos, sugiriendo que son zonas más accesibles o menos exclusivas.

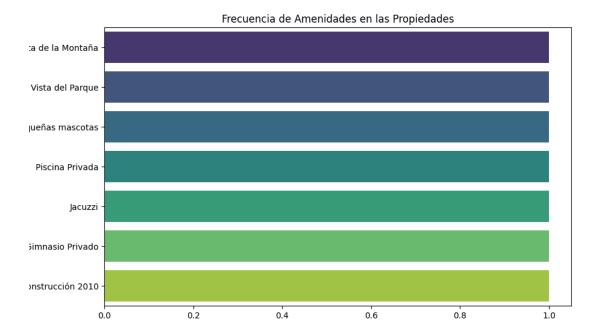
Recomendación:

Realizar un análisis adicional sobre las características específicas de las propiedades en **Cumbaya** y **La Carolina** que justifiquen estos precios elevados. Además, explorar por qué zonas como **Ponciano Alto** y **La Mariscal** tienen precios significativamente más bajos.

9. Distribución de Amenidades (Extras)

```
df_extras = df_fazwaz['Extras'].dropna().str.split(', ')
all_extras = [extra for sublist in df_extras for extra in sublist]
extras_freq = pd.Series(all_extras).value_counts()

plt.figure(figsize=(10, 6))
sns.barplot(x=extras_freq.values, y=extras_freq.index,
palette='viridis')
plt.title('Frecuencia de Amenidades en las Propiedades')
plt.savefig('images/Fazwaz/Fazwaz_exp_frecuencia_extras.png')
plt.show()
```



Descripción: Este gráfico muestra la frecuencia con la que aparecen diferentes amenidades (extras) en las propiedades. Nos proporciona información útil sobre qué características adicionales son más comunes y, en consecuencia, qué buscan los compradores o inquilinos.

Observación General: Este gráfico muestra las amenidades más frecuentes en las propiedades analizadas.

Conclusiones:

- Las amenidades más comunes incluyen vistas de la montaña, vistas del parque y la aceptación de pequeñas mascotas, lo que sugiere una preferencia por propiedades que ofrecen una conexión con la naturaleza o servicios amigables con las mascotas.
- Amenidades como piscina privada, gimnasio y jacuzzi son menos frecuentes, pero siguen siendo un atractivo importante para el segmento medio-alto.
- Al tratarse de un análisis exploratorio y la pequeña cantidad de datos y de que la columna solamente presentaba 2 filas sin null (12 Extras 2 non-null object) la gráfica no es muy representativa del total de amenidades del conjunto de datos.

Recomendación: Las propiedades con estas amenidades tienden a ser más atractivas para compradores de segmento medio-alto. Se recomienda que los desarrolladores de propiedades de lujo ofrezcan más de estas amenidades para competir efectivamente en el mercado.

10. Segmentación de Mercado por Precio

bins = [0, 150000, 300000, 500000]

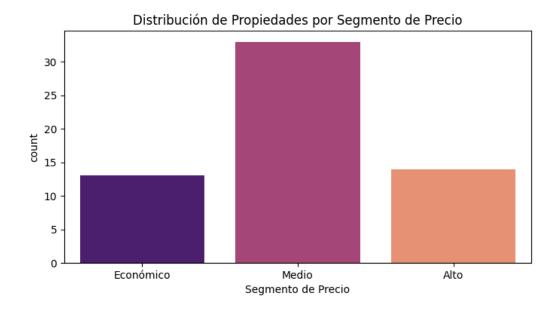
```
labels = ['Económico', 'Medio', 'Alto']

df_fazwaz['Segmento de Precio'] = pd.cut(df_fazwaz['Precio'],

bins=bins, labels=labels)

plt.figure(figsize=(8, 4))
sns.countplot(x='Segmento de Precio', data=df_fazwaz,

palette='magma')
plt.title('Distribución de Propiedades por Segmento de Precio')
plt.savefig('images/Fazwaz/Fazwaz_exp_segmento_precio.png')
plt.show()
```



Descripción: Este gráfico nos permite observar cómo se distribuyen las propiedades en función de su rango de precios. Segmentar el mercado es esencial para identificar patrones dentro de cada categoría, lo que ayuda a enfocar los análisis en cada tipo de cliente o grupo de compradores potenciales.

Observación General: El gráfico de barras muestra la cantidad de propiedades en cada segmento de precio: Económico, Medio y Alto.

- El segmento medio domina el mercado, lo que refleja una demanda robusta para propiedades de rango medio, probablemente dirigidas a familias y compradores de clase media.
- El segmento alto muestra una menor cantidad de propiedades, lo que puede indicar que el mercado de lujo está menos desarrollado o es más exclusivo.

 El segmento económico, aunque menor en comparación, sigue representando una porción significativa del mercado, posiblemente vinculado a propiedades más antiguas o en zonas periféricas.

Recomendación: Se debe explorar si la demanda en el segmento de lujo está siendo satisfecha adecuadamente y si hay espacio para el desarrollo de nuevas propiedades en el segmento económico.

Conclusión General del Análisis Exploratorio:

El análisis exploratorio de las propiedades en Quito revela varias tendencias y características clave en el mercado inmobiliario:

- 1. Relación entre Precio y Área: Las propiedades más grandes tienden a ser más caras, lo cual es esperado. Sin embargo, también se observan algunas propiedades con grandes áreas pero precios relativamente bajos, lo que indica que factores como la ubicación, antigüedad y características adicionales pueden influir significativamente en el valor de las propiedades.
- 2. Distribución de Precios por Número de Habitaciones: Las propiedades con más habitaciones muestran una mayor variabilidad en los precios. Las de 3 habitaciones, en particular, abarcan un amplio rango de precios, sugiriendo que, además del tamaño, factores como la ubicación y las amenidades influyen considerablemente. Las propiedades de 1 y 2 habitaciones tienden a estar en rangos de precio más definidos.
- 3. **Precio por m² según Ubicación**: Cumbayá destaca como la zona más cara en términos de precio por metro cuadrado, seguida de La Carolina. Ambas son zonas consideradas de alta demanda y exclusivas. Zonas como Ponciano Alto y La Mariscal presentan precios más bajos por metro cuadrado, lo que indica un mercado más accesible o menos demandado en esas áreas.
- 4. Número de Habitaciones y Baños: La mayoría de las propiedades en el análisis tienen 3 habitaciones y 2 baños, lo que refleja una preferencia por propiedades orientadas a familias de tamaño medio. Las propiedades con más de 3 habitaciones son más escasas y probablemente estén enfocadas en compradores de lujo o grandes familias, mientras que las de 1 habitación parecen estar dirigidas a solteros o parejas sin hijos.
- 5. Amenidades y Extras: Características como vistas panorámicas (Vista de la Montaña, Vista del Parque) y amenidades como piscinas privadas, gimnasios y jacuzzis aparecen en muchas propiedades, especialmente en las de precio más alto. Estas amenidades son clave para justificar las diferencias de precio en propiedades similares en términos de área y ubicación.
- 6. Correlaciones entre Variables: El análisis de correlación muestra que el precio tiene una relación moderada con el área de la propiedad y el precio por m². Además, el número de baños y habitaciones también tiene un impacto en el precio, aunque en menor medida. Estos resultados sugieren que el precio de una propiedad no depende exclusivamente de una sola variable, sino de una combinación de varias características.