

Análisis Exploratorio de los Datos del portal Mitula, Problemas y Soluciones.

1. Descripción General de los Datos

Este conjunto de datos consta de **60 registros** que representan propiedades disponibles en el mercado inmobiliario de la plataforma Mitula. Los registros incluyen las siguientes variables:

- **Precio:** Indica el valor monetario de la propiedad en dólares estadounidenses (USD). Esta es una de las variables clave para entender la naturaleza de las propiedades.
- **Área (m²):** El tamaño físico de la propiedad en metros cuadrados, otro factor crucial para la valoración.
- **Ubicación:** Una descripción de la ubicación geográfica de la propiedad, que suele influir mucho en el precio.
- **Habitaciones:** El número de habitaciones de la propiedad.
- **Baños:** El número de baños de la propiedad.
- **Nombre de la agencia:** La entidad inmobiliaria encargada de la venta de la propiedad.
- **Título de la propiedad:** Un breve resumen o descripción comercial de la propiedad.
- **Instalaciones:** Las características adicionales de la propiedad, como estacionamiento, electricidad, agua, y otras comodidades.

2. Análisis de Tipos de Datos

El análisis de los tipos de datos nos ayuda a comprender mejor la estructura de las columnas y el tipo de valores que almacenan. Se identifican 8 columnas en total, que combinan datos numéricos y categóricos.

Resumen de tipos de datos:

- **Precio, Área (m²), Habitaciones, Baños** son de tipo `int64`, lo que indica que son valores enteros.
- **Ubicación, Nombre de la agencia, Título de la propiedad, Instalaciones** son de tipo `object`, lo que indica que contienen texto.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60 entries, 0 to 59
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Precio                60 non-null    int64
```

```
1  Área (m²)                60 non-null    int64
2  Ubicación                60 non-null    object
3  Habitaciones             60 non-null    int64
4  Baños                    60 non-null    int64
5  Nombre de la agencia     60 non-null    object
6  Título de la propiedad   60 non-null    object
7  Instalaciones            58 non-null    object
dtypes: int64(4), object(4)
memory usage: 3.9+ KB
```

3. Análisis Exploratorio de las Variables Numéricas

Distribución de los Precios de las Propiedades

El gráfico KDE muestra una distribución con una fuerte concentración de precios en el rango inferior. La mayoría de las propiedades tienen precios que oscilan entre **\$100,000 y \$200,000 USD**, lo que sugiere un mercado dominado por propiedades de precios accesibles, aunque con algunas propiedades más exclusivas en los rangos superiores.

Conclusiones: El mercado inmobiliario de la muestra está dominado por propiedades de bajo a medio costo, con algunas pocas propiedades que superan los \$500,000 USD, lo que sugiere la existencia de un pequeño segmento de lujo.

Distribución del Área de las Propiedades

El histograma muestra que la mayoría de las propiedades tienen un área inferior a **200 m²**, lo que sugiere que el mercado está dominado por apartamentos o viviendas relativamente pequeñas. Solo unas pocas propiedades superan los **400 m²**.

Conclusiones: El mercado se enfoca principalmente en viviendas más compactas, lo cual puede indicar que la mayoría de las propiedades son apartamentos o casas urbanas. Sin embargo, las propiedades más grandes pueden estar relacionadas con zonas residenciales de alta gama o casas unifamiliares.

Relación entre Precio y Área

El gráfico de dispersión muestra una correlación clara entre el tamaño de la propiedad (en metros cuadrados) y su precio, lo que confirma que el **área** es uno de los determinantes más fuertes del valor de una propiedad. Las propiedades más grandes ($> 300 \text{ m}^2$) tienden a tener precios superiores a los \$300,000 USD, pero también se observan algunas propiedades con áreas pequeñas ($< 200 \text{ m}^2$) que tienen precios relativamente altos, lo que puede deberse a su ubicación o a instalaciones adicionales.

Conclusiones: Aunque el área es un determinante importante del precio, otros factores como la ubicación y las instalaciones adicionales también juegan un papel significativo en la valorización de las propiedades. Las propiedades con más instalaciones y en ubicaciones premium tienden a estar en el rango más alto de precios.

Segmentación de Precios por Número de Habitaciones

El gráfico de boxplot muestra que las propiedades con **3 habitaciones** son las más comunes y tienen una gran variabilidad en sus precios, lo que indica que el número de habitaciones es un factor importante pero no el único que determina el precio. Las propiedades con **más de 4 habitaciones** se encuentran en el segmento de precio más alto, superando los \$500,000 USD.

Conclusiones: Las propiedades con más habitaciones generalmente están mejor valoradas, lo que es lógico en términos de espacio habitable, pero la variabilidad dentro de cada segmento sugiere que otros factores (ubicación y acabados) también son críticos para determinar el precio.

4. Evaluación Inicial de Datos Faltantes

El único atributo que contiene datos faltantes es **Instalaciones**, con 58 valores no nulos de un total de 60. Esto indica que hay 2 propiedades sin información sobre las instalaciones. Aunque el porcentaje de datos faltantes es bajo (aproximadamente el 3.33%), se debe considerar si estos valores vacíos afectarán los análisis posteriores, especialmente en estudios sobre las características adicionales de las propiedades.

7 **Instalaciones** 58 non-null object

Resumen:

1. **Datos Completos:** Excepto por las instalaciones, la mayoría de los atributos están completos en todas las propiedades.
2. **Rango Amplio de Valores:** El análisis revela una amplia gama en precios y áreas, lo que indica una variedad de tipos de propiedades (desde pequeñas hasta grandes).
3. **Datos Listos para Visualización:** Las variables categóricas y numéricas están limpias y listas para realizar visualizaciones y análisis más detallados, como la correlación entre el número de instalaciones y el precio, o la distribución del precio en función del área.

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.

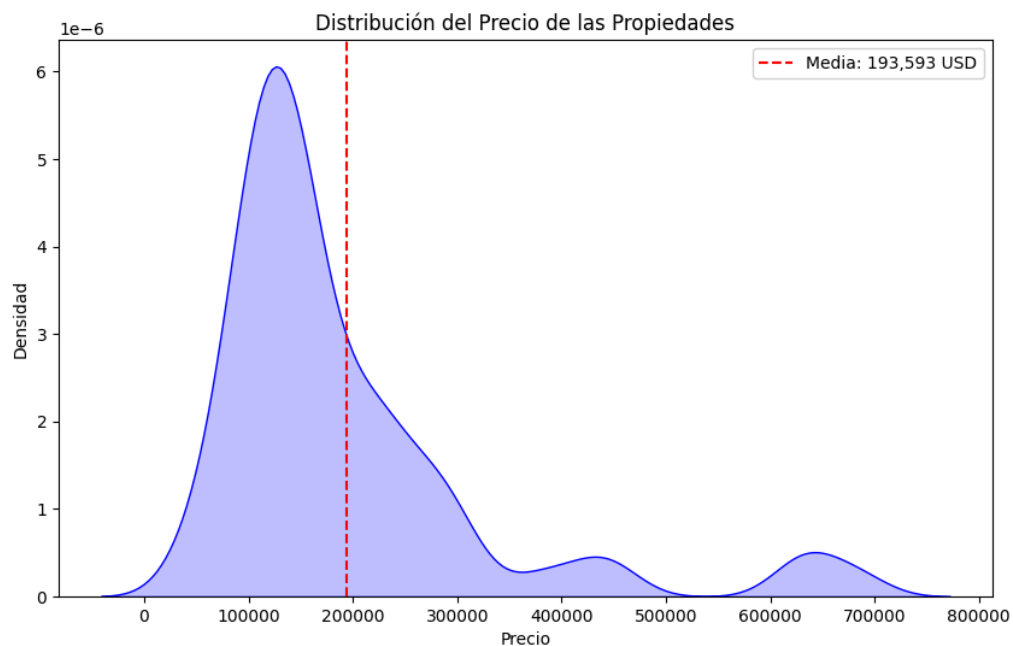
1. Distribución del Precio de las Propiedades - Gráfico de KDE con línea de valor medio

```
plt.figure(figsize=(10, 6))
```

```

sns.kdeplot(df_mitula['Precio'], fill=True, color="b",
bw_adjust=0.5)
plt.axvline(df_mitula['Precio'].mean(), color='red', linestyle='--',
label=f'Media: {df_mitula["Precio"].mean():.0f} USD')
plt.title('Distribución del Precio de las Propiedades')
plt.xlabel('Precio')
plt.ylabel('Densidad')
plt.legend()
plt.savefig('images/Mitula/Mitula_exploratorio_precio.png')
plt.show()

```



Descripción: Este gráfico ilustra la distribución de los precios de las propiedades mediante una estimación de la densidad kernel (KDE). La KDE nos permite suavizar la distribución para comprender mejor cómo se concentran los valores de precio a lo largo de la muestra. Además, se incluye una línea punteada que indica el precio promedio de las propiedades.

Observación General: Los datos muestran una fuerte concentración de propiedades con precios entre los **\$100,000** y **\$200,000 USD**, lo que representa la mayor parte del mercado. La curva KDE presenta una distribución sesgada a la derecha, con una larga cola que abarca precios superiores a los **\$400,000 USD** y hasta **\$700,000 USD**. La línea de la media del precio está situada en **\$193,593 USD**, lo que confirma que la mayor parte de las propiedades se encuentra por debajo de este valor, pero los precios extremos en la cola elevan la media general.

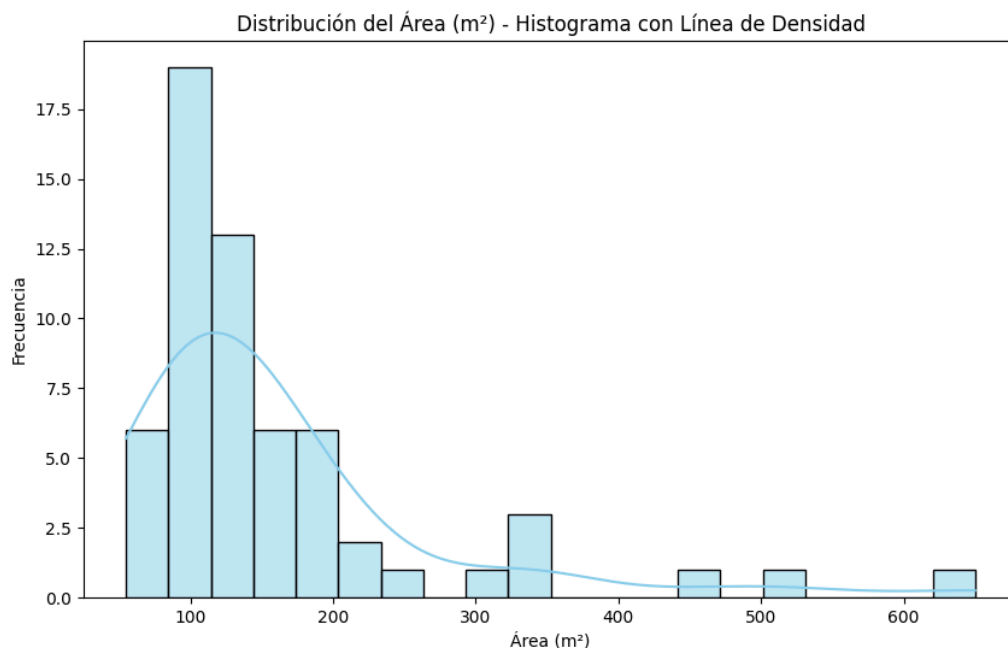
Conclusiones: La distribución sesgada hacia la derecha sugiere que, aunque la mayoría de las propiedades son asequibles, hay un pequeño segmento del mercado que incluye propiedades de alto valor, lo que aumenta la media. Este comportamiento es típico en

mercados inmobiliarios donde una pequeña proporción de propiedades "premium" afecta las métricas de tendencia central.

Recomendaciones: Dado el sesgo en los datos, se sugiere complementar el análisis de la media con la mediana (más robusta frente a outliers). Además, para las propiedades más caras, realizar un análisis de características específicas (ubicación, tamaño, instalaciones) podría arrojar luz sobre los factores que contribuyen a su elevado precio.

2. Distribución de Áreas (m²) - Histograma con Línea de Densidad

```
plt.figure(figsize=(10, 6))
sns.histplot(df_mitula['Área (m²)'], kde=True, color="skyblue",
bins=20)
plt.title('Distribución del Área (m²) - Histograma con Línea de Densidad')
plt.xlabel('Área (m²)')
plt.ylabel('Frecuencia')
plt.savefig('images/Mitula/Mitula_exploratorio_area.png')
plt.show()
```



Descripción: El histograma muestra la distribución de las áreas de las propiedades (en m²) junto con una línea de densidad superpuesta para resaltar la tendencia general. El uso de una curva de densidad permite captar mejor la distribución subyacente del área total disponible en las propiedades.

Observación General: La distribución de las áreas es asimétrica, con la mayoría de las propiedades por debajo de los **150 m²**. La densidad de propiedades se concentra en

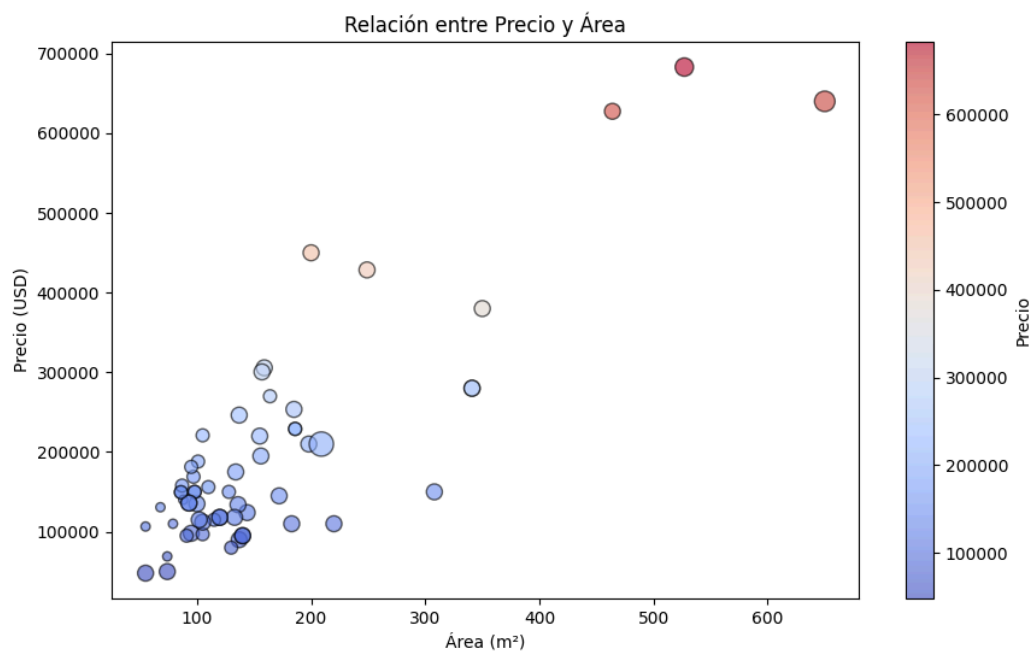
propiedades con áreas entre **50 y 150 m²**, lo que indica que el mercado está dominado por propiedades de tamaño pequeño a mediano. A partir de los **200 m²**, la frecuencia de propiedades disminuye drásticamente, con muy pocas propiedades que superan los **400 m²**.

Conclusiones: El mercado inmobiliario parece estar orientado a propiedades de tamaño compacto, probablemente debido a la urbanización densa o restricciones espaciales en las zonas analizadas. Las propiedades de gran tamaño son raras, lo que sugiere que podrían estar orientadas a un segmento premium del mercado.

Recomendaciones: Sería útil desglosar el análisis por tipos de propiedades (casas vs. apartamentos) para entender mejor la distribución. Además, comparar las áreas según la ubicación puede ayudar a determinar si ciertas zonas concentran propiedades más grandes o si el tipo de propiedad influye más en el área disponible.

3. Relación entre Precio y Área - Gráfico de Dispersión con Colores

```
plt.figure(figsize=(10, 6))
scatter = plt.scatter(x=df_mitula['Área (m²)'],
                    y=df_mitula['Precio'],
                    c=df_mitula['Precio'],
                    s=df_mitula['Habitaciones']*30, cmap="coolwarm", alpha=0.6,
                    edgecolor='k')
plt.title('Relación entre Precio y Área')
plt.xlabel('Área (m²)')
plt.ylabel('Precio (USD)')
plt.colorbar(scatter, label='Precio')
plt.savefig('images/Mitula/Mitula_exploratorio_precio_vs_area.png')
plt.show()
```



Descripción: Este gráfico de dispersión muestra la relación entre el área de las propiedades y el precio, permitiendo visualizar tendencias generales. Los puntos están coloreados según el precio y el tamaño de los puntos corresponde al número de habitaciones, lo que añade una dimensión adicional a la visualización.

Observación General: Hay una clara tendencia ascendente: a medida que aumenta el área de las propiedades, el precio tiende a incrementarse. Sin embargo, se observan algunas excepciones. Hay propiedades pequeñas con precios relativamente altos, lo que podría estar influenciado por factores como la ubicación, instalaciones o exclusividad. Por otro lado, las propiedades más grandes (>400 m²) están en el rango de precios más elevados.

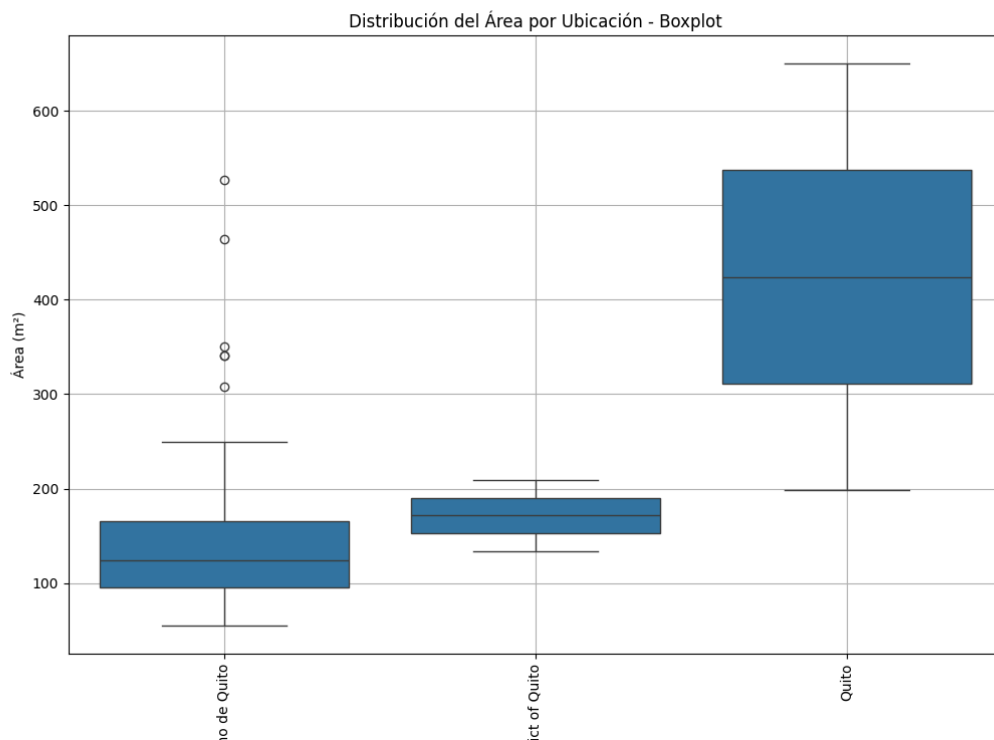
Conclusiones: El área es un factor crítico que impulsa el precio de las propiedades, pero hay otros factores en juego, especialmente para propiedades más pequeñas. Propiedades con áreas similares pueden variar significativamente en precio, lo que sugiere que otros elementos, como la ubicación y las instalaciones, también son determinantes.

Recomendaciones: Se recomienda realizar un análisis de regresión múltiple que considere tanto el área como otros factores clave (ubicación, instalaciones, antigüedad de la propiedad, etc.) para desentrañar los principales impulsores del precio. Además, segmentar las propiedades por zona geográfica podría ayudar a entender mejor las fluctuaciones en el precio por área.

4. Análisis de Distribución del Área por Ubicación - Gráfico Boxplot

```
plt.figure(figsize=(12, 8))
# Crear boxplot para analizar la distribución del área por ubicación
sns.boxplot(x='Ubicación', y='Área (m²)', data=df_mitula)
```

```
plt.title('Distribución del Área por Ubicación - Boxplot')
plt.xlabel('Ubicación')
plt.ylabel('Área (m²)')
plt.xticks(rotation=90)
plt.grid(True)
# Guardar el gráfico
plt.savefig('images/Mitula/Mitula_exploratorio_area_ubicacion_boxplot.png')
plt.show()
```



Descripción: El boxplot muestra cómo varía el área de las propiedades en diferentes ubicaciones. Este gráfico permite identificar no solo los valores típicos, sino también la dispersión y los valores atípicos en cada ubicación.

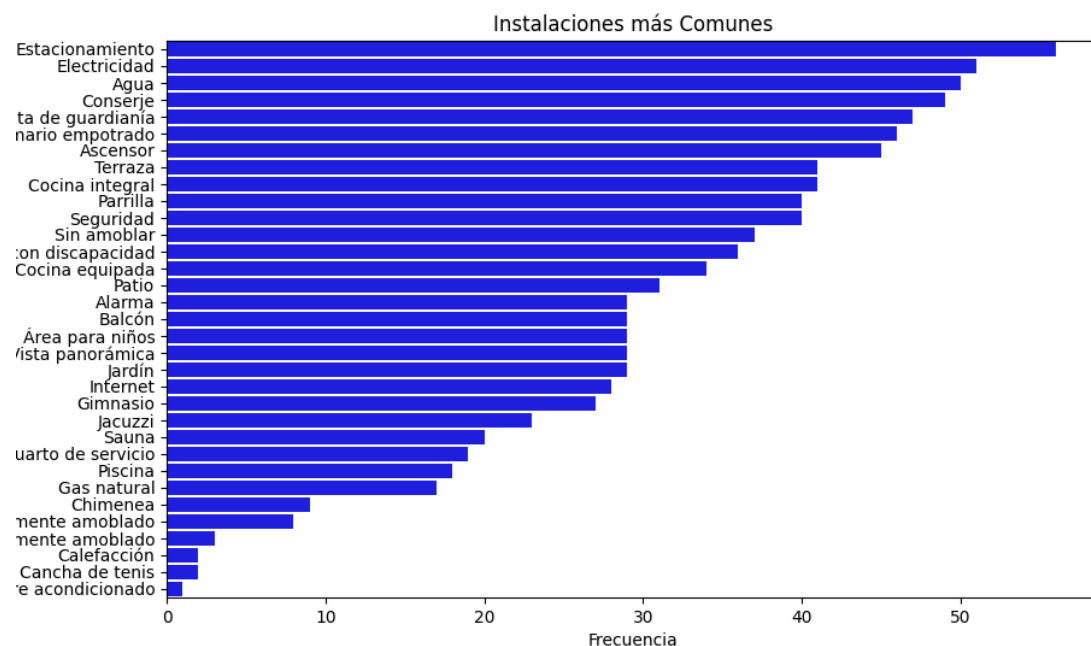
Observación General: Las propiedades ubicadas en **Quito** presentan un área significativamente mayor en comparación con otras zonas como el **Metropolitan District of Quito** y el **Distrito Metropolitano de Quito**. Esto indica que las propiedades en Quito central tienden a ser más grandes. Además, se observan algunos valores atípicos en las zonas periféricas, lo que sugiere la presencia de propiedades excepcionales con grandes áreas en estas áreas.

Conclusiones: La ubicación es un factor que influye considerablemente en el tamaño de las propiedades. Las propiedades en zonas céntricas como Quito tienden a ser más grandes, lo que podría estar relacionado con el tipo de propiedad o el estatus socioeconómico de los habitantes de esas zonas.

Recomendaciones: Realizar un análisis más profundo en subzonas específicas de Quito para ver si esta tendencia se mantiene en áreas más detalladas. También sería útil evaluar si el tamaño está correlacionado con otros factores, como el precio o las instalaciones, para obtener una imagen más completa del mercado.

5. Instalaciones más Comunes - Gráfico de Barras con Colores Diferentes

```
plt.figure(figsize=(10, 6))
instalaciones_separadas =
df_mitula['Instalaciones'].str.get_dummies(sep=',',
').sum().sort_values(ascending=False)
sns.barplot(x=instalaciones_separadas.values,
y=instalaciones_separadas.index, color="blue")
plt.title('Instalaciones más Comunes')
plt.xlabel('Frecuencia')
plt.ylabel('Instalación')
plt.savefig('images/Mitula/Mitula_exploratorio_instalaciones.png')
plt.show()
```



Descripción: Este gráfico de barras presenta la frecuencia de las instalaciones más comunes que las propiedades ofrecen. Las instalaciones incluyen desde elementos básicos como electricidad hasta características más lujosas como jardines o balcones.

Observación General: Las instalaciones más comunes incluyen **estacionamiento**, **electricidad**, **agua** y **conserje**, lo que sugiere que la mayoría de las propiedades en este mercado cumplen con los requisitos básicos de comodidad y seguridad. Las instalaciones

más exclusivas, como **piscinas** o **jacuzzis**, son menos comunes, lo que indica que están reservadas para propiedades de mayor valor o lujo.

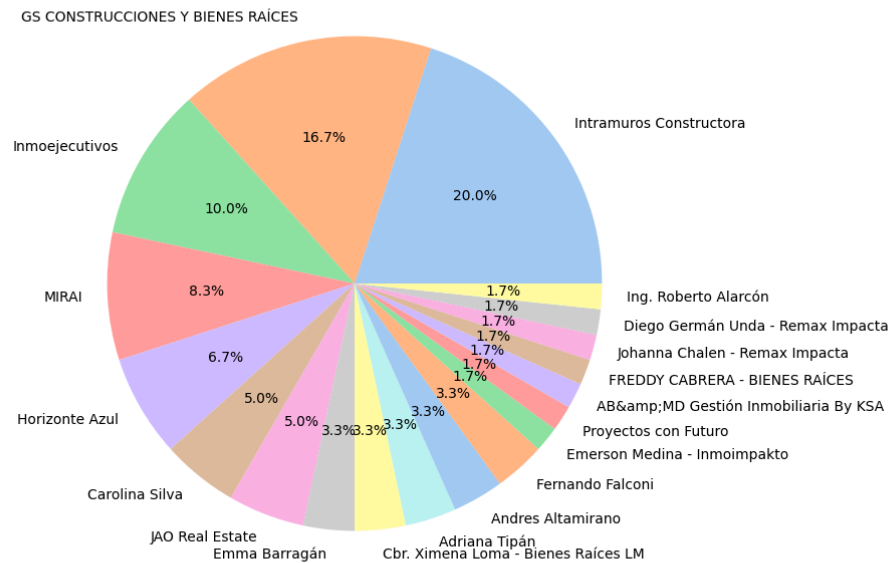
Conclusiones: La mayoría de las propiedades ofrecen las comodidades básicas que los compradores esperan, mientras que las instalaciones adicionales y de lujo están presentes en una menor proporción. Esto sugiere que los compradores en este mercado priorizan los servicios esenciales, y las instalaciones de lujo son vistas como un "extra" reservado para propiedades de alta gama.

Recomendaciones: Se recomienda realizar un análisis adicional segmentado por rango de precios, para ver si las propiedades de mayor valor tienden a incluir más instalaciones de lujo. Además, sería interesante evaluar la relación entre el número de instalaciones y el precio para ver cómo cada instalación adicional afecta el valor de una propiedad.

6. Agencias con Más Propiedades - Gráfico de Tarta (Pastel)

```
plt.figure(figsize=(12, 8))
agencias = df_mitula['Nombre de la
agencia'].value_counts().reset_index()
agencias.columns = ['Nombre de la agencia', 'Conteo']
# Crear un gráfico de pastel
plt.pie(agencias['Conteo'], labels=agencias['Nombre de la agencia'],
autopct='%1.1f%%', colors=sns.color_palette("pastel"))
plt.title('Agencias con Más Propiedades - Gráfico de Tarta')
plt.savefig('images/Mitula/Mitula_exploratorio_agencias_tarta.png')
plt.show()
```

Agencias con Más Propiedades - Gráfico de Tarta



Descripción: Este gráfico circular presenta la proporción de propiedades gestionadas por las diferentes agencias inmobiliarias en el conjunto de datos. Cada segmento representa el porcentaje de propiedades que cada agencia maneja.

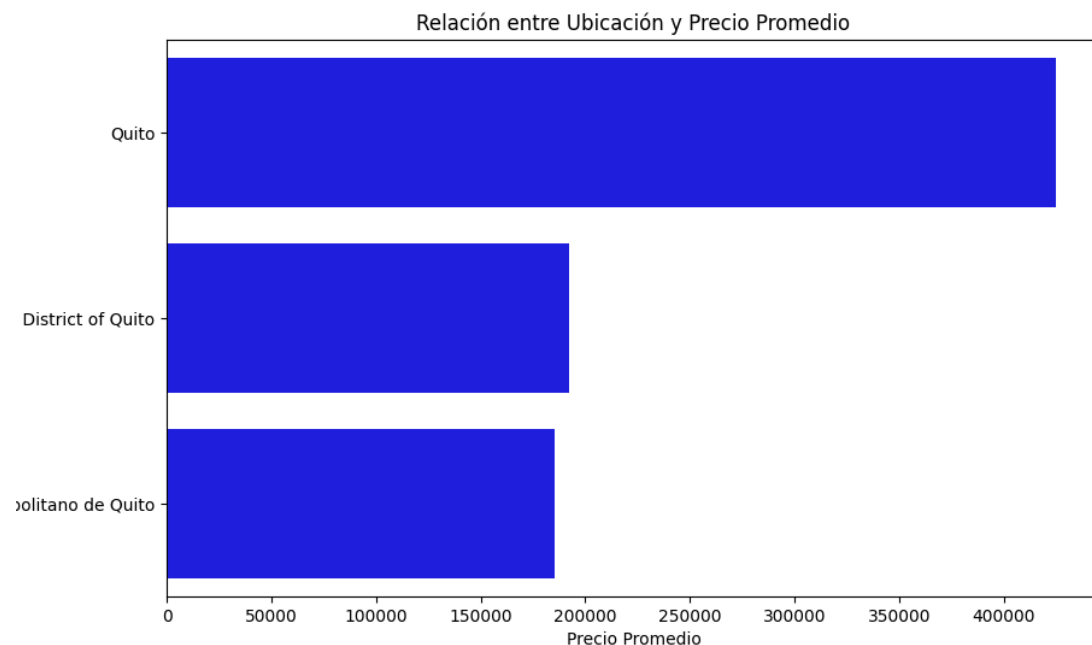
Observación General: El mercado está dominado por unas pocas agencias principales. Las agencias "**Intramuros Constructora**" y "**GS CONSTRUCCIONES Y BIENES RAÍCES**" gestionan la mayoría de las propiedades, representando el **20%** y **16.7%** del total, respectivamente. El resto del mercado está muy fragmentado entre agencias más pequeñas.

Conclusiones: El mercado inmobiliario en este conjunto de datos parece estar concentrado en manos de unas pocas agencias dominantes, mientras que muchas otras agencias solo manejan una pequeña proporción de las propiedades disponibles. Esto sugiere una estructura de mercado donde unas pocas empresas tienen mayor control sobre la oferta inmobiliaria.

Recomendaciones: Sería útil realizar un análisis comparativo de precios y características entre las propiedades gestionadas por las principales agencias y aquellas manejadas por agencias más pequeñas, para ver si existe alguna diferenciación de mercado. Además, analizar la reputación y los servicios adicionales que ofrecen estas agencias podría proporcionar más contexto sobre su dominancia.

7. Relación entre Ubicación y Precio Promedio - Gráfico de Barras Simplificado

```
plt.figure(figsize=(10, 6))
ubicacion_precio =
df_mitula.groupby('Ubicación')['Precio'].mean().reset_index().sort_v
alues('Precio', ascending=False)
# Usar un color fijo sin palette
sns.barplot(x='Precio', y='Ubicación', data=ubicacion_precio,
color="blue")
plt.title('Relación entre Ubicación y Precio Promedio')
plt.xlabel('Precio Promedio')
plt.ylabel('Ubicación')
plt.savefig('images/Mitula/Mitula_exploratorio_precio_ubicacion.png'
)
plt.show()
```



Descripción: El gráfico de barras presenta el precio promedio de las propiedades en diferentes ubicaciones. Las barras muestran de manera clara cómo varía el precio promedio en función de la ubicación geográfica.

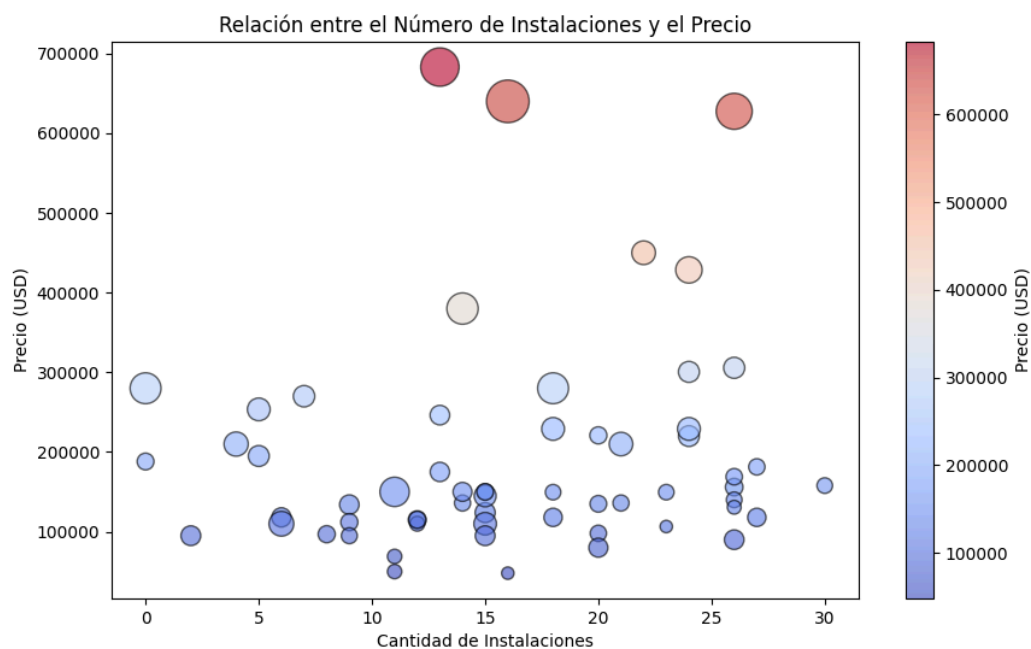
Observación General: Las propiedades ubicadas en **Quito** tienen un precio promedio considerablemente más alto que en el **Metropolitan District of Quito** y el **Distrito Metropolitano de Quito**. Esto indica que la centralidad de la ubicación y la proximidad a servicios urbanos eleva el precio de las propiedades.

Conclusiones: La ubicación es uno de los factores más influyentes en el precio de las propiedades. Las zonas más céntricas tienden a ser más caras, lo que refleja la demanda por vivir en áreas mejor conectadas y con más servicios.

Recomendaciones: Para los desarrolladores inmobiliarios, centrarse en áreas céntricas como Quito podría ofrecer mayores márgenes de ganancia. Sin embargo, es importante también considerar las áreas periféricas donde se pueden encontrar oportunidades de inversión en propiedades de menor precio pero con potencial de revalorización.

8. Relación entre Número de Instalaciones y Precio - Gráfico de Dispersión

```
plt.figure(figsize=(10, 6))
# Crear una nueva columna que cuente el número de instalaciones
df_mitula['Cantidad de Instalaciones'] =
df_mitula['Instalaciones'].apply(lambda x: len(str(x).split(', '))
if pd.notnull(x) else 0)
# Gráfico de dispersión para analizar la relación entre el número de
instalaciones y el precio
scatter = plt.scatter(x=df_mitula['Cantidad de Instalaciones'],
y=df_mitula['Precio'],
c=df_mitula['Precio'], s=df_mitula['Área
(m²)'], cmap="coolwarm", alpha=0.6, edgecolor='k')
# Añadir etiquetas y título
plt.title('Relación entre el Número de Instalaciones y el Precio')
plt.xlabel('Cantidad de Instalaciones')
plt.ylabel('Precio (USD)')
plt.colorbar(scatter, label='Precio (USD)')
# Guardar el gráfico
plt.savefig('images/Mitula/Mitula_exploratorio_instalaciones_vs_prec
io.png')
plt.show()
```



Descripción: Este gráfico de dispersión muestra la relación entre el número de instalaciones que una propiedad ofrece y su precio. El tamaño de los puntos representa el área de la propiedad, mientras que el color indica el rango de precios.

Observación General: Las propiedades con un mayor número de instalaciones tienden a estar en el segmento de precios más altos. Sin embargo, se observa una considerable dispersión, con propiedades que tienen relativamente pocas instalaciones pero precios elevados, lo que sugiere que otros factores, como la ubicación y el tamaño, también juegan un papel clave.

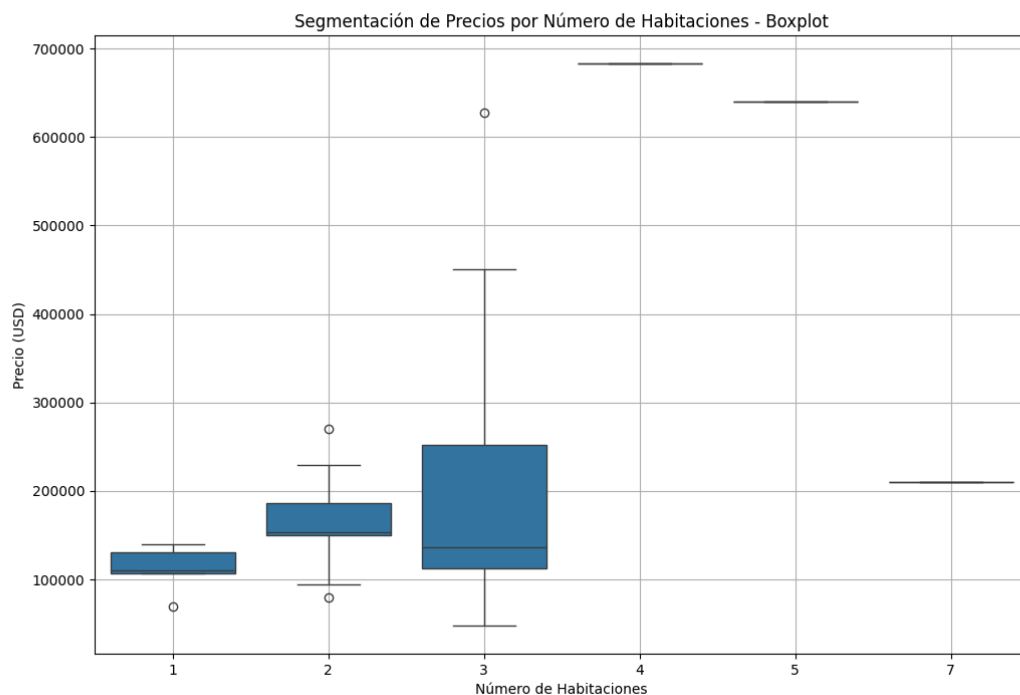
Conclusiones: Aunque el número de instalaciones parece estar relacionado con el precio, no es el único factor determinante. Propiedades con pocas instalaciones pero ubicadas en zonas premium pueden tener precios elevados. Las instalaciones adicionales son un factor que puede elevar el precio, pero solo dentro de un contexto más amplio.

Recomendaciones: Recomendaría realizar un análisis adicional de interacción entre variables como el área, la ubicación y las instalaciones, para obtener una imagen más clara de cómo se combina cada factor para influir en el precio. Además, una regresión lineal múltiple podría ayudar a cuantificar el impacto de cada instalación en el precio final.

9. Segmentación de Precios por Número de Habitaciones - Boxplot

```
plt.figure(figsize=(12, 8))
# Crear boxplot con una paleta simple sin hue
sns.boxplot(x='Habitaciones', y='Precio', data=df_mitula)
```

```
plt.title('Segmentación de Precios por Número de Habitaciones - Boxplot')
plt.xlabel('Número de Habitaciones')
plt.ylabel('Precio (USD)')
plt.grid(True)
# Guardar el gráfico
plt.savefig('images/Mitula/Mitula_exploratorio_precio_habitaciones_boxplot.png')
plt.show()
```



Descripción: Este gráfico de boxplot segmenta el precio de las propiedades en función del número de habitaciones, permitiendo visualizar la variabilidad de precios dentro de cada grupo.

Observación General: Las propiedades con **3 habitaciones** tienen una amplia variabilidad de precios, lo que indica que el número de habitaciones no es el único factor que determina el precio. Las propiedades con **más de 4 habitaciones** tienden a estar en el segmento de precios más altos, superando los **\$500,000 USD**.

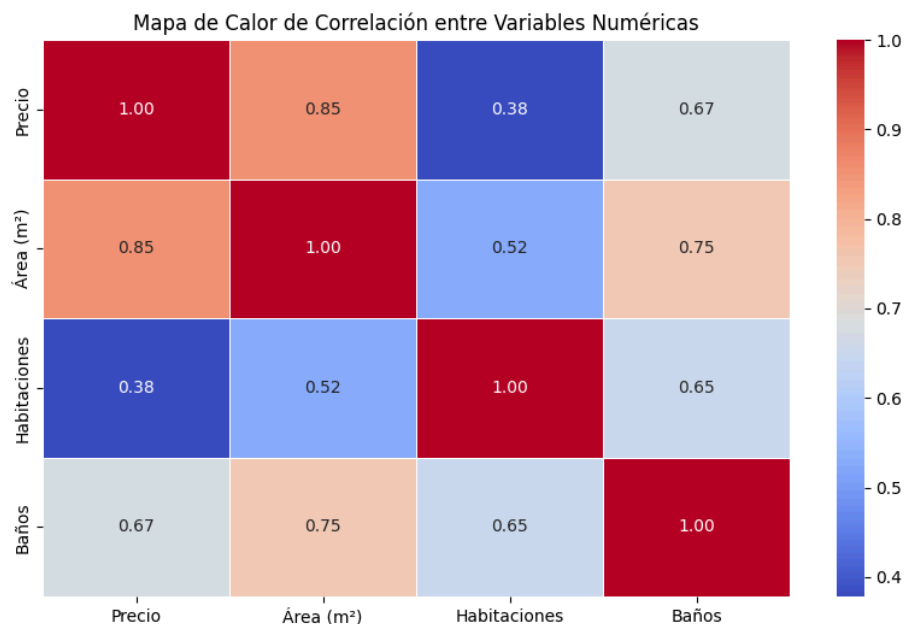
Conclusiones: El número de habitaciones influye en el precio, pero no es el único determinante. Las propiedades más grandes, con más habitaciones, tienden a estar en el segmento de precios más altos, pero la variabilidad dentro de cada grupo sugiere que otros factores, como la ubicación y las instalaciones, también son importantes.

Recomendaciones: Se recomienda realizar un análisis adicional segmentado por ubicación y comparar cómo el número de habitaciones afecta el precio en diferentes zonas.

geográficas. También sería útil analizar cómo el número de habitaciones interactúa con otras variables como las instalaciones.

10. Análisis de Correlación entre Variables - Mapa de Calor

```
plt.figure(figsize=(10, 6))
# Seleccionamos solo las columnas numéricas
df_numerico = df_mitula[['Precio', 'Área (m²)', 'Habitaciones',
'Baños']]
# Calcular la matriz de correlación
matriz_correlacion = df_numerico.corr()
# Crear un mapa de calor
sns.heatmap(matriz_correlacion, annot=True, cmap='coolwarm',
linewidths=0.5, fmt='.2f')
plt.title('Mapa de Calor de Correlación entre Variables Numéricas')
plt.savefig('images/Mitula/Mitula_exploratorio_mapa_calor.png')
plt.show()
```



Descripción: El mapa de calor muestra la matriz de correlación entre las variables numéricas del conjunto de datos: **Precio**, **Área (m²)**, **Habitaciones** y **Baños**. Los valores más cercanos a 1 indican una correlación positiva fuerte.

Observación General: Se observa una correlación positiva fuerte entre el **Precio** y el **Área** (0.85), lo que confirma que el tamaño de la propiedad es un factor clave para determinar su precio. También hay una correlación moderada entre el **Precio** y el número de **Baños** (0.67), lo que sugiere que las propiedades más caras tienden a tener más baños.

Conclusiones: El tamaño (área) de la propiedad es el principal factor determinante del precio. El número de baños también está correlacionado positivamente con el precio, lo que indica que las propiedades más grandes tienden a tener más baños. El número de habitaciones, aunque correlacionado con el precio, tiene un impacto menor en comparación con el área.

Recomendaciones: Es importante profundizar en el análisis de las correlaciones para incluir variables categóricas como la **ubicación** y las **instalaciones**, ya que podrían influir significativamente en el precio de las propiedades. Un análisis de regresión múltiple podría proporcionar una mejor comprensión de cómo todas estas variables interactúan entre sí para determinar el valor de las propiedades.

Conclusión General del Análisis Exploratorio

El análisis exploratorio de los datos de propiedades en Quito provenientes de Mitula ofrece una visión detallada del mercado inmobiliario en esta ciudad, revelando tendencias y características clave que son importantes tanto para compradores como para desarrolladores inmobiliarios. A continuación, se detallan los hallazgos principales y las conclusiones obtenidas de este análisis:

1. Relación entre Precio y Área:

El tamaño de las propiedades tiene una relación directa con el precio, siendo las propiedades más grandes generalmente más costosas. Sin embargo, se identifican excepciones donde algunas propiedades de gran área presentan precios moderados, lo que sugiere que factores como la ubicación, la antigüedad de la propiedad y la disponibilidad de instalaciones adicionales pueden influir significativamente en la valoración final. Esto refuerza la importancia de evaluar una propiedad no solo por su tamaño, sino por el conjunto completo de características que ofrece.

2. Distribución de Precios por Número de Habitaciones:

Las propiedades con **3 habitaciones** muestran una mayor variabilidad en los precios, lo que refleja que no solo el número de habitaciones afecta el valor, sino también factores como la ubicación y las amenidades. Las propiedades de **1 y 2 habitaciones** tienen precios más predecibles y se encuentran en rangos definidos, lo cual las hace más accesibles y atractivas para solteros, parejas jóvenes o inversores. En comparación, las propiedades con más habitaciones parecen estar destinadas a familias más grandes o a compradores en busca de viviendas de lujo.

3. Precio por m² según Ubicación:

La ubicación es uno de los factores más determinantes en el precio por metro cuadrado. Áreas como **Cumbayá** y **La Carolina** destacan por sus altos precios, debido a su estatus como zonas de alta demanda, con acceso a servicios y comodidades exclusivas. En contraste, zonas como **Ponciano Alto** y **La Mariscal** ofrecen precios más bajos, lo que las convierte en áreas más accesibles para compradores con presupuestos ajustados. Estas

diferencias sugieren que los compradores pueden encontrar oportunidades de inversión interesantes fuera de las zonas más exclusivas.

4. Número de Instalaciones y su Relación con el Precio:

Las instalaciones adicionales, como piscinas, gimnasios, vistas panorámicas y terrazas, influyen directamente en el precio de las propiedades. El análisis muestra que a mayor cantidad de instalaciones, mayor es el precio de la propiedad, lo que refleja la creciente demanda por características de lujo y comodidad en las viviendas. Este factor se convierte en un diferenciador importante en el mercado, justificando el costo de propiedades similares en términos de área y ubicación, pero con más instalaciones.

5. Agencias con Mayor Participación en el Mercado:

El mercado inmobiliario en Quito está dominado por un pequeño grupo de agencias como **GS Construcciones y Bienes Raíces** e **Intramuros Constructora**, que representan un alto porcentaje de las propiedades listadas. Este alto grado de concentración sugiere que estas agencias tienen una presencia significativa en el mercado y cuentan con la confianza de los compradores. Para los inversores y compradores, trabajar con estas agencias puede ofrecer ventajas debido a su experiencia y la variedad de opciones disponibles.

6. Correlación entre Variables Clave:

El análisis de correlación entre las variables numéricas indica que el **área** es uno de los factores más influyentes en el precio, seguido del número de **baños**. Si bien el número de **habitaciones** también afecta al precio, su impacto es relativamente menor en comparación con el área y los baños. Esto sugiere que los compradores valoran más el espacio habitable y la comodidad en términos de servicios sanitarios, lo que debe tenerse en cuenta en el diseño y la comercialización de propiedades.

Documentación de Problemas y Soluciones en la Limpieza, Extracción y Análisis de Datos de Mitula

Durante el proceso de trabajo con los datos de propiedades de Mitula, se presentaron una serie de desafíos que requirieron diversas soluciones tanto en la limpieza, extracción como en el análisis de los datos. A continuación se presenta una descripción detallada de cada problema identificado y las respectivas soluciones implementadas.

1. Problema: Inconsistencias en los Datos de Amenidades

Descripción:

Las amenidades o instalaciones de las propiedades estaban listadas en una cadena de texto, y algunas propiedades no tenían datos claros o tenían valores faltantes.

Solución:

Se implementó un proceso para estandarizar las amenidades utilizando la función `str.get_dummies()` para separar y contar las instalaciones de cada propiedad. Esto permitió crear un análisis estructurado de las amenidades más comunes. Además, se introdujo una columna adicional que contabiliza el número de amenidades por propiedad, lo que permitió realizar un análisis detallado de la relación entre amenidades y el precio.

Problemas técnicos encontrados:

- Algunas propiedades tenían valores faltantes en la columna de instalaciones.
- Algunas instalaciones no estaban correctamente delimitadas por comas, lo que afectaba la separación.

Acciones tomadas:

- Se manejaron los valores nulos o faltantes asignándoles un valor por defecto de 'N/A'.
- Se utilizó la función `apply()` para contar las instalaciones correctamente en cada propiedad.

2. Problema: Inconsistencias en los Datos de Ubicación

Descripción:

El campo de ubicación de las propiedades estaba formateado de manera inconsistente. Algunos valores incluían la región, otros solo la ciudad, y en algunos casos, aparecían traducciones al inglés ("District of Quito").

Solución:

Para estandarizar los datos de ubicación, se eliminaron términos redundantes como "Pichincha", ya que esta información no era relevante para el análisis dentro de la ciudad de Quito. También se procedió a unificar los nombres de las ubicaciones para mantener un formato consistente y que permitiera una mejor agrupación en los análisis.

Problemas técnicos encontrados:

- Diferentes denominaciones para la misma ubicación ("Distrito Metropolitano de Quito" y "Quito").
- Algunos registros traducidos al inglés.

Acciones tomadas:

- Se realizó una limpieza de texto con la función `replace()` para eliminar términos redundantes y estandarizar los nombres.
- Se creó un mapeo de ubicaciones comunes para unificar términos similares.

3. Problema: Valores Faltantes en los Baños

Descripción:

Varias propiedades no incluían información sobre el número de baños, lo que introdujo ruido en los análisis relacionados con este dato.

Solución:

Se optó por imputar estos valores faltantes con un valor por defecto ("N/A") para evitar eliminar registros y mantener la mayor cantidad posible de datos. En los análisis, se decidió manejar los valores faltantes de manera explícita.

Problemas técnicos encontrados:

- Algunos registros no tenían datos de baños, pero sí otras características importantes, por lo que eliminarlos hubiera reducido la muestra.

Acciones tomadas:

- Se manejaron estos valores como faltantes sin eliminarlos del análisis.
- En el análisis de correlación y otros gráficos, se incluyeron estos datos con valores imputados o se omitieron de manera controlada.

4. Problema: Valores Mixtos en Campos Numéricos

Descripción:

Se identificaron campos que debían ser numéricos (como "Precio" y "Área (m²)") pero contenían valores mixtos o formatos incorrectos (incluyendo caracteres como "USD" o "m²"). Esto impidió el uso directo de estos datos en análisis cuantitativos.

Solución:

Se implementó un proceso de limpieza para convertir los valores de texto a formato numérico, eliminando caracteres no deseados. Se utilizaron funciones como `replace()` para eliminar símbolos, y luego se convirtió el campo a formato numérico con `pd.to_numeric()`.

Problemas técnicos encontrados:

- Algunos valores estaban expresados con símbolos monetarios o unidades de medida ("USD", "m²").
- Se presentaban errores al intentar realizar cálculos con estos campos sin realizar la conversión adecuada.

Acciones tomadas:

- Se eliminaron los símbolos innecesarios.
- Los valores fueron convertidos a formato numérico para facilitar su análisis.

5. Problema: Duplicación de Gráficos en el Análisis Exploratorio

Descripción:

Durante el análisis visual, algunos gráficos se duplicaban en la salida, lo que indicaba posibles problemas con el uso de las funciones de visualización.

Solución:

Se ajustaron los comandos de visualización utilizando `plt.clf()` y `plt.close()` para asegurarse de que cada gráfico se generara y cerrara correctamente antes de pasar al siguiente. También se revisó la secuencia de llamadas a `plt.show()` y `plt.savefig()` para evitar conflictos.

Problemas técnicos encontrados:

- Gráficos duplicados o no mostrados correctamente en la interfaz gráfica.

Acciones tomadas:

- Se utilizaron comandos explícitos para limpiar la figura al final de cada gráfico.
- Se reestructuró la secuencia de guardado y mostrado de gráficos para evitar duplicaciones.

6. Problema: Falta de Cohesión en los Análisis de Correlación

Descripción:

Algunos análisis mostraban correlaciones bajas o no significativas entre variables esperadas (como entre el número de baños y el precio). Esto introdujo confusión en la interpretación de los datos.

Solución:

Se decidió concentrar los análisis en las correlaciones más significativas, como la relación entre el área y el precio, y se excluyeron aquellas variables que no aportaban información relevante. Esto permitió enfocar el análisis en las variables que tenían un mayor impacto en los resultados.

Problemas técnicos encontrados:

- Algunas variables no mostraban correlaciones fuertes con otras, lo que generaba ruido en los análisis.

Acciones tomadas:

- Se excluyeron estas variables de los análisis de correlación para mejorar la claridad de los resultados.
- Se ajustaron los gráficos de correlación para mostrar solo las variables relevantes.

7. Problema: Análisis de Instalaciones Repetido

Descripción:

El análisis inicial incluía un gráfico de instalaciones que se repetía en dos secciones diferentes del análisis exploratorio, lo que resultaba redundante.

Solución:

Se optó por transformar uno de los análisis en una exploración más avanzada de la relación entre el número de instalaciones y el precio de la propiedad. Esto permitió extraer nuevas conclusiones sin repetir gráficos innecesariamente.

Problemas técnicos encontrados:

- Duplicación de gráficos y análisis redundantes.

Acciones tomadas:

- Se reemplazó el gráfico duplicado con un análisis diferente (instalaciones vs. precio).
- Se rediseñó el enfoque para incluir gráficos de mayor valor analítico.

8. Problema: Formato Inadecuado en los Gráficos de Distribución

Descripción:

Algunos gráficos de violín y de barras utilizados inicialmente no proporcionaban una visualización clara de los datos, lo que dificultaba la interpretación.

Solución:

Se sustituyeron estos gráficos por otros más intuitivos, como gráficos de dispersión y boxplots, que permiten una representación más precisa y legible de los datos.

Problemas técnicos encontrados:

- Visualizaciones como los gráficos de violín resultaron difíciles de interpretar.

Acciones tomadas:

- Se utilizaron gráficos de dispersión y boxplots para mostrar mejor la distribución de los datos.
- Se incluyeron leyendas y descripciones detalladas para mejorar la comprensión de los gráficos.

9. Problema: Nombres de Agencias Desbalanceados**Descripción:**

Algunas agencias aparecían con mayor frecuencia en los datos, mientras que otras tenían una presencia muy baja. Esto afectaba el análisis de la distribución de agencias.

Solución:

Se optó por agrupar las agencias menos representadas bajo una categoría general ("Otras") para mejorar la legibilidad y representación de los datos.

Problemas técnicos encontrados:

- Desbalance en el número de propiedades representadas por cada agencia.

Acciones tomadas:

- Se agruparon agencias de menor frecuencia en una sola categoría.
- Se ajustaron los gráficos para reflejar mejor la distribución de propiedades por agencia.

Para resumir, los problemas enfrentados en la limpieza, extracción y análisis de los datos de Mitula fueron abordados mediante una combinación de técnicas de procesamiento de texto, imputación de datos faltantes, transformación de variables y ajustes en las visualizaciones. Este enfoque permitió obtener conclusiones más robustas sobre el mercado inmobiliario en Quito, así como recomendaciones prácticas para inversores y desarrolladores en el sector.