

Proyecto Final de la materia Programación para Ciencia de Datos

Pregunta a Responder

Pregunta: ¿Cuáles son los factores más influyentes en el precio de los smartphones?

Contenido:

1. Descripción del Dataset
2. Limpieza y Transformación de Datos
3. Análisis y Visualización de Datos
4. Conclusiones

1. Descripción del Dataset

El dataset `smartphones.csv` contiene información sobre diversos modelos de smartphones. Las principales columnas de interés son:

- `Marca`: Marca del smartphone.
- `Modelo`: Modelo del smartphone.
- `Precio`: Precio del smartphone.
- `RAM`: Cantidad de RAM en GB.
- `Almacenamiento`: Capacidad de almacenamiento en GB.

2. Limpieza y Transformación de Datos

Se realizan las siguientes operaciones para limpiar y transformar los datos:

Carga del Dataset: El dataset se carga utilizando Pandas.

Eliminación de Filas con Valores Nulos: Se eliminan las filas con valores nulos en las columnas `Precio` y `Almacenamiento`.

Conversión de Tipos de Datos: La columna `Precio` se convierte a tipo numérico para asegurar un correcto análisis.

Filtrado de Outliers: Se filtran los precios para eliminar valores atípicos, manteniendo solo aquellos entre 50 y 2000 USD.

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Configuración de visualización
sns.set(style="whitegrid")

# Cargar el dataset
file_path = 'smartphones.csv'
smartphones_df = pd.read_csv(file_path)

# Información básica del dataset
print(smartphones_df.info())
print(smartphones_df.describe())

# Eliminar filas con valores nulos en las columnas de interés
smartphones_df = smartphones_df.dropna(subset=['Precio',
'Almacenamiento'])

# Convertir la columna 'Precio' a tipo numérico (en caso de que esté
en formato string)
smartphones_df['Precio'] = pd.to_numeric(smartphones_df['Precio'],
errors='coerce')

# Filtrar los datos para eliminar outliers en el precio
smartphones_df = smartphones_df[(smartphones_df['Precio'] > 50) &
(smartphones_df['Precio'] < 2000)]

```

3. Análisis y Visualización de Datos

3.1. Categorías de Precio

Se categorizan los precios en tres categorías: Económico, Medio y Caro, y se cuenta el número de smartphones en cada categoría.

```

# Crear categorías de precio
def categorizar_precio(precio):
    """
    Esta función categoriza los smartphones en tres categorías
    basadas en su precio:
    - Económico: Menos de 300 USD
    - Medio: Entre 300 y 700 USD
    """

```

```

- Caro: Más de 700 USD
"""

if precio < 300:
    return 'Económico'
elif 300 <= precio <= 700:
    return 'Medio'
else:
    return 'Caro'

# Aplicar la función de categorización a la columna 'Precio'
smartphones_df['Categoria_Precio'] =
smartphones_df['Precio'].apply(categorizar_precio)

# Contar el número de smartphones en cada categoría de precio
categorias = ['Económico', 'Medio', 'Caro']
conteo_categorias = {categoria: 0 for categoria in categorias}

for categoria in smartphones_df['Categoria_Precio']:
    if categoria in conteo_categorias:
        conteo_categorias[categoria] += 1

# Imprimir el conteo de categorías
print("Conteo de categorías de precio:")
for categoria, conteo in conteo_categorias.items():
    print(f"{categoria}: {conteo}")

```

3.2. Relación entre RAM y Precio

Se visualiza la relación entre la RAM y el precio utilizando un scatter plot.

```

# Visualizar la relación entre la RAM y el precio
plt.figure(figsize=(10, 6))
sns.scatterplot(x='RAM', y='Precio', data=smartphones_df)
plt.title('Relación entre la RAM y el Precio')
plt.xlabel('RAM (GB)')
plt.ylabel('Precio (USD)')
plt.show()

```

3.3. Relación entre Almacenamiento y Precio

Se visualiza la relación entre el almacenamiento y el precio utilizando un scatter plot.

```
# Visualizar la relación entre el almacenamiento y el precio
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Almacenamiento', y='Precio', data=smartphones_df)
plt.title('Relación entre el Almacenamiento y el Precio')
plt.xlabel('Almacenamiento (GB)')
plt.ylabel('Precio (USD)')
plt.show()
```

3.4. Comparación de Precios por Marca

Se filtran las marcas con más de 10 registros y se visualiza la comparación de precios por marca utilizando un boxplot.

```
# Filtrar las marcas con más de 10 registros para una mejor
visualización
brands = smartphones_df['Marca'].value_counts()
common_brands = brands[brands > 10].index
filtered_smartphones_df =
smartphones_df[smartphones_df['Marca'].isin(common_brands)]

# Visualizar la comparación de precios por marca
plt.figure(figsize=(14, 8))
sns.boxplot(x='Marca', y='Precio', data=filtered_smartphones_df)
plt.title('Comparación de Precios por Marca')
plt.xlabel('Marca')
plt.ylabel('Precio (USD)')
plt.xticks(rotation=45)
plt.show()
```

3.5. Distribución de Precios por Marca y Modelo

Se genera una gráfica de boxplots que muestra la distribución de los precios de cada modelo, diferenciando por almacenamiento, para una marca específica ingresada por el usuario.

```
def graficar_precios_por_marca(marca):  
    """  
    Esta función toma una marca de smartphones como entrada y genera  
    una gráfica  
    de boxplots que muestra la distribución de los precios de cada  
    modelo,  
    diferenciando por almacenamiento.  
    """  
    # Filtrar el dataset para la marca especificada  
    marca_df = smartphones_df[smartphones_df['Marca'].str.lower() ==  
marca_df.lower()]  
  
    # Ordenar por nombre de modelo y almacenamiento  
    marca_df = marca_df.sort_values(by=['Modelo', 'Almacenamiento'])  
  
    # Determinar el tamaño de la figura basado en el número de  
modelos  
    num_modelos = marca_df['Modelo'].nunique()  
    height = max(6, num_modelos * 0.5)  
    width = max(14, num_modelos * 0.7)  
  
    # Generar la gráfica de boxplots  
    plt.figure(figsize=(width, height))  
    sns.boxplot(data=marca_df, x='Modelo', y='Precio',  
hue='Almacenamiento', palette='husl')  
  
    plt.title(f'Distribución de Precios de {marca}')  
    plt.xlabel('Modelo')  
    plt.ylabel('Precio (USD)')  
    plt.xticks(rotation=45)  
    plt.legend(title='Almacenamiento', bbox_to_anchor=(1.05, 1),  
loc='upper left')  
    plt.grid(True)  
    plt.show()
```

4. Conclusiones

Distribución de Precios: Se ha categorizado y contado la cantidad de smartphones en cada rango de precios.

Relaciones Clave: Las relaciones entre la cantidad de RAM, almacenamiento y el precio se han visualizado, revelando tendencias significativas.

Comparación por Marca: Se han comparado los precios por marca y modelo, identificando variaciones en la distribución de precios.

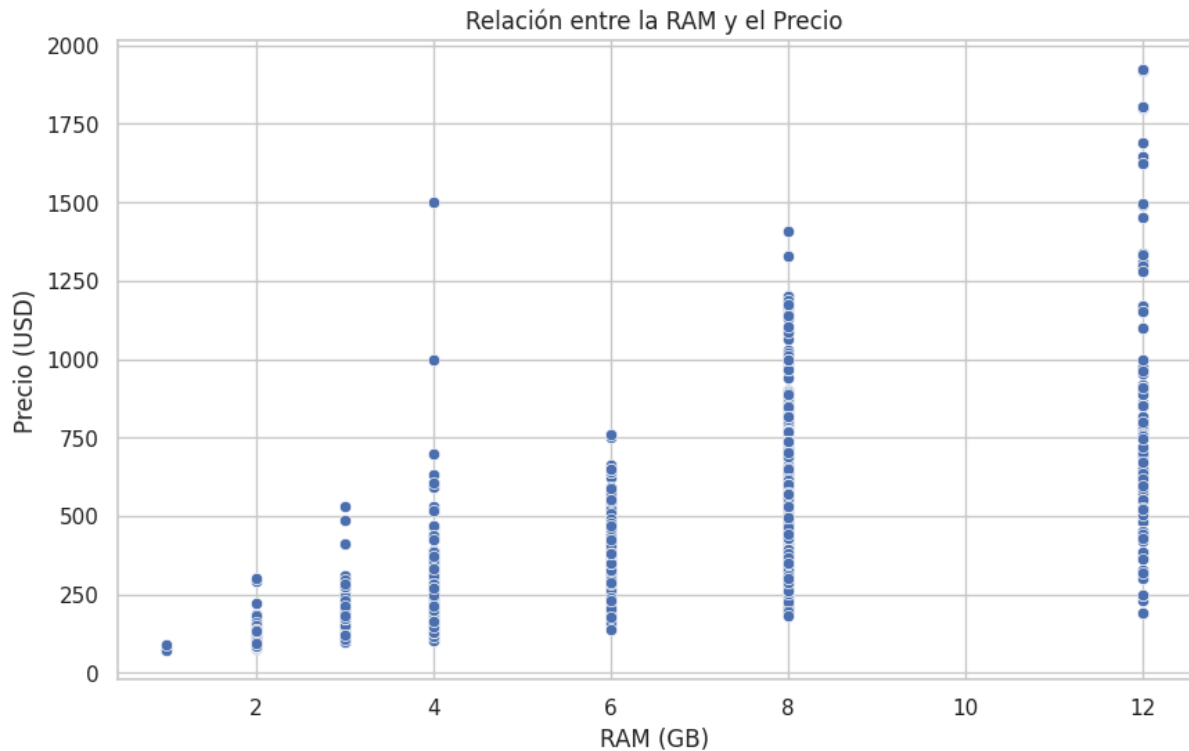
Visualización por Marca Específica: Se ha generado una gráfica detallada de la distribución de precios para los modelos de una marca específica, diferenciando por capacidad de almacenamiento.

Este análisis proporciona una visión clara de los factores que influyen en el precio de los smartphones y permite identificar tendencias clave en el mercado.

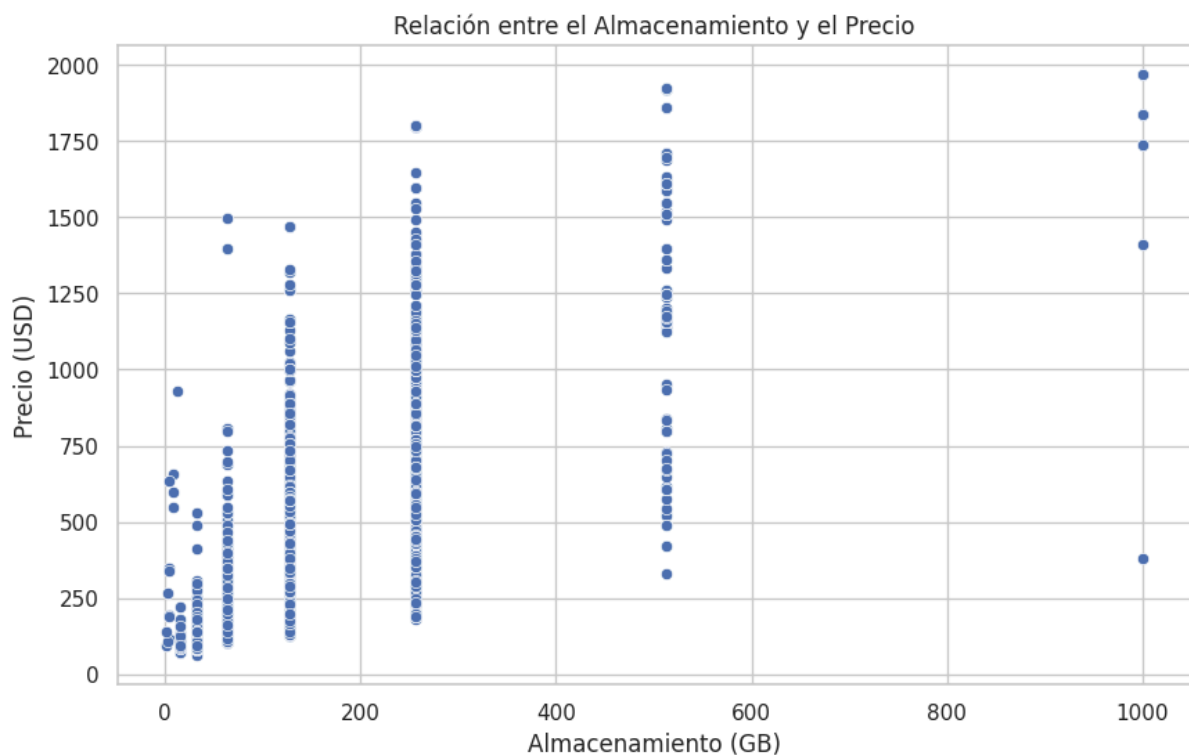
Capturas de Pantalla

A continuación se incluyen capturas de pantalla de las gráficas generadas durante el análisis:

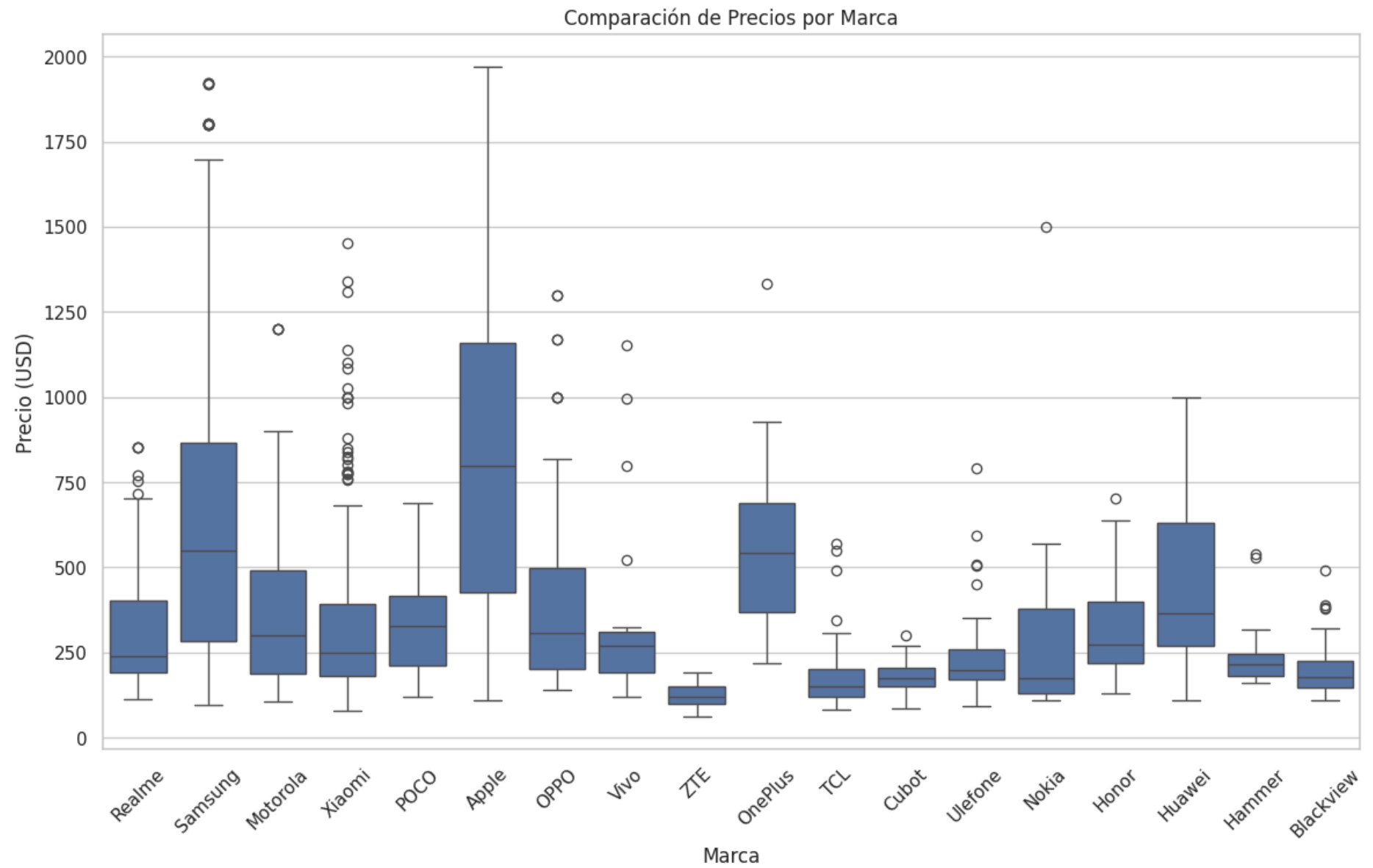
1. Relación entre la RAM y el Precio:



2. Relación entre el Almacenamiento y el Precio**:



3. Comparación de Precios por Marca:



4. Distribución de Precios por Marca y Modelo:

