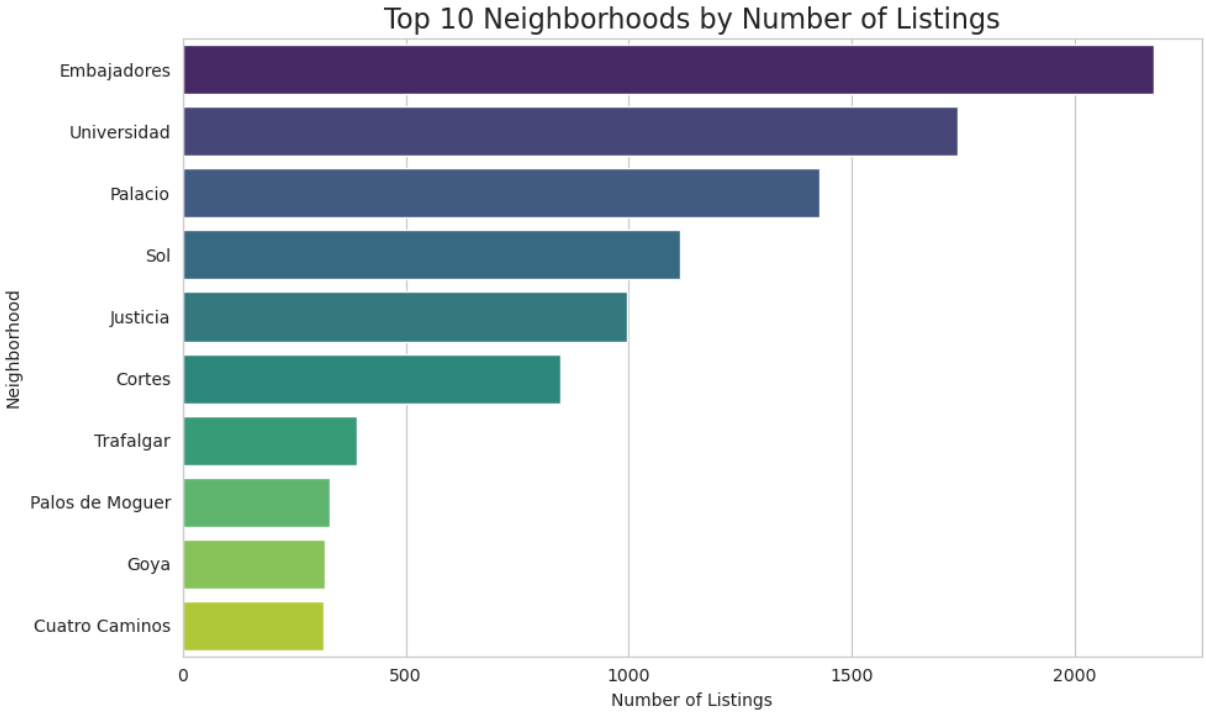
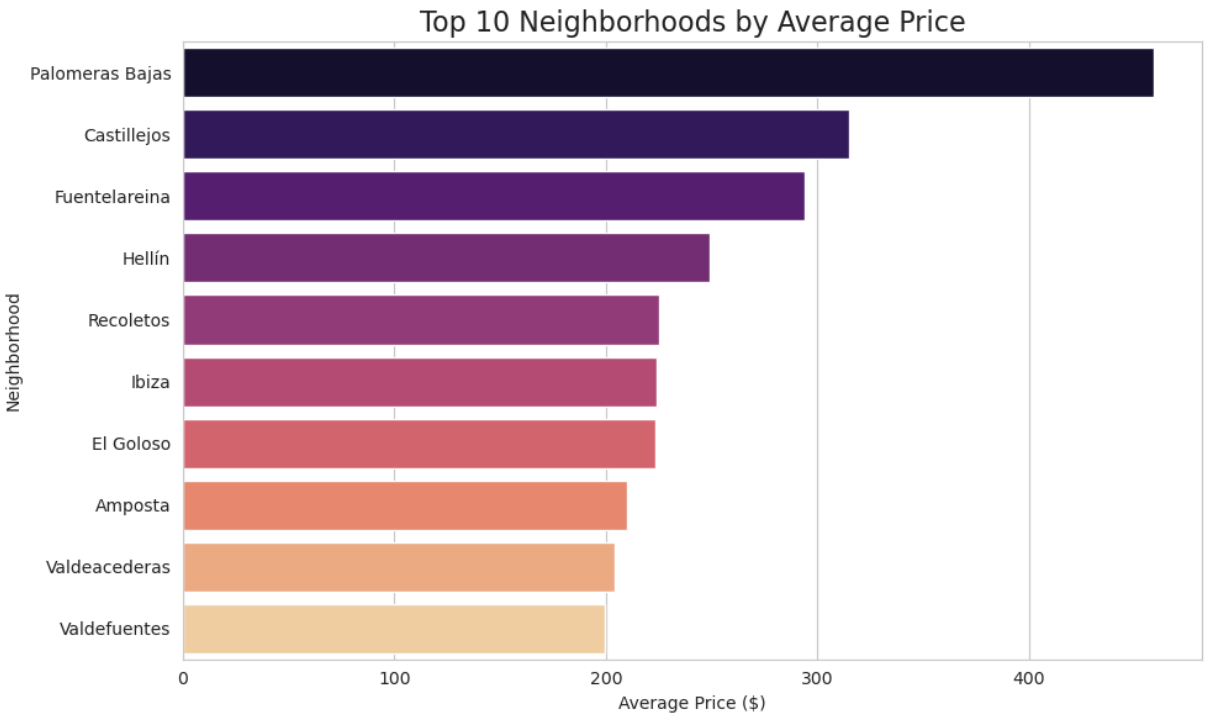
In [6]:
```
# 1. Import Libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Set style for plots
sns.set_style('whitegrid')

# 2. Load Dataset
try:
    df = pd.read_csv('listings.csv')
    print("Data loaded successfully!")
except FileNotFoundError:
    print("Error: The file 'listings.csv' was not found in the current directory.")

# 3. Quick Look at Data
print("\nDataset Shape:", df.shape)
print("\nColumn Names:", df.columns.tolist())
print("\nFirst 5 Rows:\n", df.head())

# 4. Data Cleaning

if 'price' in df.columns:
    df['price'] = df['price'].astype(str).replace({'\$':'', ',':''}, regex=True)
    df['price'] = pd.to_numeric(df['price'], errors='coerce')

df.dropna(subset=['price'], inplace=True)

df.fillna({'reviews_per_month': 0, 'last_review': 'Missing', 'neighbourhood': 'Unkn

# 5. Basic Analysis

# Top 10 neighborhoods by listing count
if 'neighbourhood' in df.columns:
    top_neighborhoods = df['neighbourhood'].value_counts().head(10)
    print("\nTop 10 Neighborhoods by Number of Listings:\n", top_neighborhoods)
else:
    print("\nColumn 'neighbourhood' not found in dataset.")

# Average price by neighborhood
if 'neighbourhood' in df.columns and 'price' in df.columns:
    avg_price_neighborhood = (
        df.groupby('neighbourhood')['price']
        .mean()
        .sort_values(ascending=False)
        .head(10)
    )
    print("\nTop 10 Neighborhoods by Average Price:\n", avg_price_neighborhood)
else:
    print("\nCannot calculate average price by neighborhood (columns missing).")

# 6. Visualizations

# Plot top neighborhoods by count
```

```
if 'neighbourhood' in df.columns:
    plt.figure(figsize=(10,6))
    sns.barplot(
        x=top_neighborhoods.values,
        y=top_neighborhoods.index,
        palette='viridis'
    )
    plt.title('Top 10 Neighborhoods by Number of Listings', fontsize=16)
    plt.xlabel('Number of Listings')
    plt.ylabel('Neighborhood')
    plt.tight_layout()
    plt.show()

# Plot average price by neighborhood
if 'neighbourhood' in df.columns and 'price' in df.columns:
    plt.figure(figsize=(10,6))
    sns.barplot(
        x=avg_price_neighborhood.values,
        y=avg_price_neighborhood.index,
        palette='magma'
    )
    plt.title('Top 10 Neighborhoods by Average Price', fontsize=16)
    plt.xlabel('Average Price ($)')
    plt.ylabel('Neighborhood')
    plt.tight_layout()
    plt.show()

# 7. Save Cleaned Data (Optional)
df.to_csv('airbnb_cleaned.csv', index=False)
print("\nCleaned data saved as 'airbnb_cleaned.csv'.")
```

```
Data loaded successfully!

Dataset Shape: (25288, 18)

Column Names: ['id', 'name', 'host_id', 'host_name', 'neighbourhood_group', 'neighbo
urhood', 'latitude', 'longitude', 'room_type', 'price', 'minimum_nights', 'number_of
_reviews', 'last_review', 'reviews_per_month', 'calculated_host_listings_count', 'av
ailability_365', 'number_of_reviews_ltm', 'license']

First 5 Rows:
         id                                              name  host_id host_name  \
0  21853.0                         Bright and airy room    83531     Abdel
1  30320.0                 Great Vacational Apartments   130907      Dana
2  30959.0              Beautiful loft in Madrid Center   132883    Angela
3  40916.0              Holiday Apartment Madrid Center   130907      Dana
4  62423.0  MAGIC ARTISTIC HOUSE IN THE CENTER OF MADRID   303845    Arturo

  neighbourhood_group neighbourhood  latitude  longitude        room_type  \
0              Latina      Cármenes  40.40381   -3.74130     Private room
1              Centro           Sol  40.41476   -3.70418  Entire home/apt
2              Centro   Embajadores  40.41259   -3.70105  Entire home/apt
3              Centro   Universidad  40.42247   -3.70577  Entire home/apt
4              Centro      Justicia  40.41884   -3.69655     Private room

   price  minimum_nights  number_of_reviews last_review  reviews_per_month  \
0   29.0               4                 33   7/15/2018               0.26
1    NaN               5                172   9/26/2022               0.96
2    NaN               3                  8   5/30/2017               0.07
3    NaN               5                 49  12/11/2021               0.28
4   77.0               1                227   2/20/2025               2.73

   calculated_host_listings_count  availability_365  number_of_reviews_ltm  \
0                               2               233                      0
1                               3                 0                      0
2                               1                 0                      0
3                               3                 0                      0
4                               3               298                     46

  license
0     NaN
1     NaN
2     NaN
3     NaN
4     NaN

Top 10 Neighborhoods by Number of Listings:
 neighbourhood
Embajadores        2177
Universidad        1737
Palacio            1427
Sol                1115
Justicia            996
Cortes              846
Trafalgar           391
Palos de Moguer     330
Goya                320
```

```
Cuatro Caminos      316
Name: count, dtype: int64

Top 10 Neighborhoods by Average Price:
 neighbourhood
Palomeras Bajas    458.888889
Castillejos        315.135965
Fuentelareina      294.000000
Hellín             249.142857
Recoletos          224.907692
Ibiza              223.797297
El Goloso          223.375000
Amposta            210.083333
Valdeacederas      204.241379
Valdefuentes       199.408333
Name: price, dtype: float64
```






```
Cleaned data saved as 'airbnb_cleaned.csv'.
```