In [1]:
```
pip install pandas
```

```
Defaulting to user installation because normal site-packages is not writeableNote: you may need to restart the kernel to use updated packages.

Requirement already satisfied: pandas in c:\users\jesus\appdata\roaming\python\python310\site-packages (2.3.0)
Requirement already satisfied: pytz>=2020.1 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from pandas) (2025.2)
Requirement already satisfied: numpy>=1.22.4 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from pandas) (1.24.3)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: tzdata>=2022.7 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from pandas) (2025.2)
Requirement already satisfied: six>=1.5 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
WARNING: You are using pip version 21.2.3; however, version 25.1.1 is available.
You should consider upgrading via the 'C:\Program Files\Python310\python.exe -m pip install --upgrade pip' command.
```

In [2]:
```
pip install numpy
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: numpy in c:\users\jesus\appdata\roaming\python\python310\site-packages (1.24.3)
Note: you may need to restart the kernel to use updated packages.
WARNING: You are using pip version 21.2.3; however, version 25.1.1 is available.
You should consider upgrading via the 'C:\Program Files\Python310\python.exe -m pip install --upgrade pip' command.
```

In [3]:
```
pip install matplotlib
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: matplotlib in c:\users\jesus\appdata\roaming\python\python310\site-packages (3.10.3)
Requirement already satisfied: packaging>=20.0 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib) (25.0)
Requirement already satisfied: numpy>=1.23 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib) (1.24.3)
Requirement already satisfied: cycler>=0.10 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib) (0.12.1)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib) (1.3.2)
Requirement already satisfied: pillow>=8 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib) (11.3.0)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib) (1.4.8)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib) (3.2.3)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib) (2.9.0.post0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib) (4.58.5)
Requirement already satisfied: six>=1.5 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from python-dateutil>=2.7->matplotlib) (1.17.0)
Note: you may need to restart the kernel to use updated packages.
WARNING: You are using pip version 21.2.3; however, version 25.1.1 is available.
You should consider upgrading via the 'C:\Program Files\Python310\python.exe -m pip install --upgrade pip' command.
```

In [4]:
```
pip install seaborn
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: seaborn in c:\users\jesus\appdata\roaming\python\python310\site-packages (0.13.2)
Requirement already satisfied: pandas>=1.2 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from seaborn) (2.3.0)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from seaborn) (3.10.3)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from seaborn) (1.24.3)
Requirement already satisfied: pillow>=8 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (11.3.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.58.5)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.2.3)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.3.2)
Requirement already satisfied: packaging>=20.0 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (25.0)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.8)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.9.0.post0)
Requirement already satisfied: cycler>=0.10 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)
Requirement already satisfied: tzdata>=2022.7 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from pandas>=1.2->seaborn) (2025.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from pandas>=1.2->seaborn) (2025.2)
Requirement already satisfied: six>=1.5 in c:\users\jesus\appdata\roaming\python\python310\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.17.0)
Note: you may need to restart the kernel to use updated packages.
WARNING: You are using pip version 21.2.3; however, version 25.1.1 is available.
You should consider upgrading via the 'C:\Program Files\Python310\python.exe -m pip install --upgrade pip' command.
```

In [5]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

In [6]:
```
df= pd.read_csv("airbnb.csv")
```

In [7]:
```
df
```

Out[7]:

	id	host_name	neighbourhood_group	latitude	longitude	room_type	price	number_of_reviews	reviews_per_month	availability_365	last_review
0	2539	John	Brooklyn	40.64749	-73.97237	Private room	149	9	0.21	365	2018-10-19
1	2595	Jennifer	Manhattan	40.75362	-73.98377	Entire home/apt	225	45	0.38	355	2019-05-21
2	3647	Elisabeth	Manhattan	40.80902	-73.94190	Private room	150	0	0.00	365	NaN
3	3831	LisaRoxanne	Brooklyn	40.68514	-73.95976	Entire home/apt	89	270	4.64	194	2019-05-07
4	5022	Laura	Manhattan	40.79851	-73.94399	Entire home/apt	80	9	0.10	0	2018-11-19
...	...	...	...	...	...	...	...	...	...	...	...
48869	36484665	Sabrina	Brooklyn	40.67853	-73.94995	Private room	70	0	0.00	9	NaN
48870	36485057	Marisol	Brooklyn	40.70184	-73.93317	Private room	40	0	0.00	36	NaN
48871	36485431	Ilgar & Aysel	Manhattan	40.81475	-73.94867	Entire home/apt	115	0	0.00	27	NaN
48872	36485609	Taz	Manhattan	40.75751	-73.99112	Shared room	55	0	0.00	2	NaN
48873	36487245	Christophe	Manhattan	40.76404	-73.98933	Private room	90	0	0.00	23	NaN

48874 rows × 11 columns

In [8]:
```
df.shape # no. of rows and columns
```

Out[8]: (48874, 11)

In [9]:
```
df.info() #data types and values
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48874 entries, 0 to 48873
Data columns (total 11 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   id                   48874 non-null  int64  
 1   host_name            48874 non-null  object 
 2   neighbourhood_group  48874 non-null  object 
 3   latitude             48874 non-null  float64
 4   longitude            48874 non-null  float64
 5   room_type            48874 non-null  object 
 6   price                48874 non-null  int64  
 7   number_of_reviews    48874 non-null  int64  
 8   reviews_per_month    48874 non-null  float64
 9   availability_365     48874 non-null  int64  
 10  last_review          38827 non-null  object 
dtypes: float64(3), int64(4), object(4)
memory usage: 4.1+ MB
```

In [10]:
```
df.describe() #summary statistics
```

Out[10]:

	id	latitude	longitude	price	number_of_reviews	reviews_per_month	availability_365
count	4.887400e+04	48874.000000	48874.000000	48874.000000	48874.000000	48874.000000	48874.000000
mean	1.901988e+07	40.728946	-73.952172	152.738634	23.266358	1.090785	112.793755
std	1.098318e+07	0.054529	0.046156	240.199728	44.544330	1.597119	131.619934
min	2.539000e+03	40.499790	-74.244420	0.000000	0.000000	0.000000	0.000000
25%	9.474068e+06	40.690100	-73.983078	69.000000	1.000000	0.040000	0.000000
50%	1.967936e+07	40.723065	-73.955680	106.000000	5.000000	0.370000	45.000000
75%	2.915342e+07	40.763110	-73.936273	175.000000	24.000000	1.580000	227.000000
max	3.648724e+07	40.913060	-73.712990	10000.000000	629.000000	58.500000	365.000000

In [11]:
```
df.columns #names of the columns
```

Out[11]:
```
Index(['id', 'host_name', 'neighbourhood_group', 'latitude', 'longitude',
       'room_type', 'price', 'number_of_reviews', 'reviews_per_month',
       'availability_365', 'last_review'],
      dtype='object')
```

In [12]:
```
print("Missing values before cleaning:\n")
df.isnull().sum() #checking the missing values
```

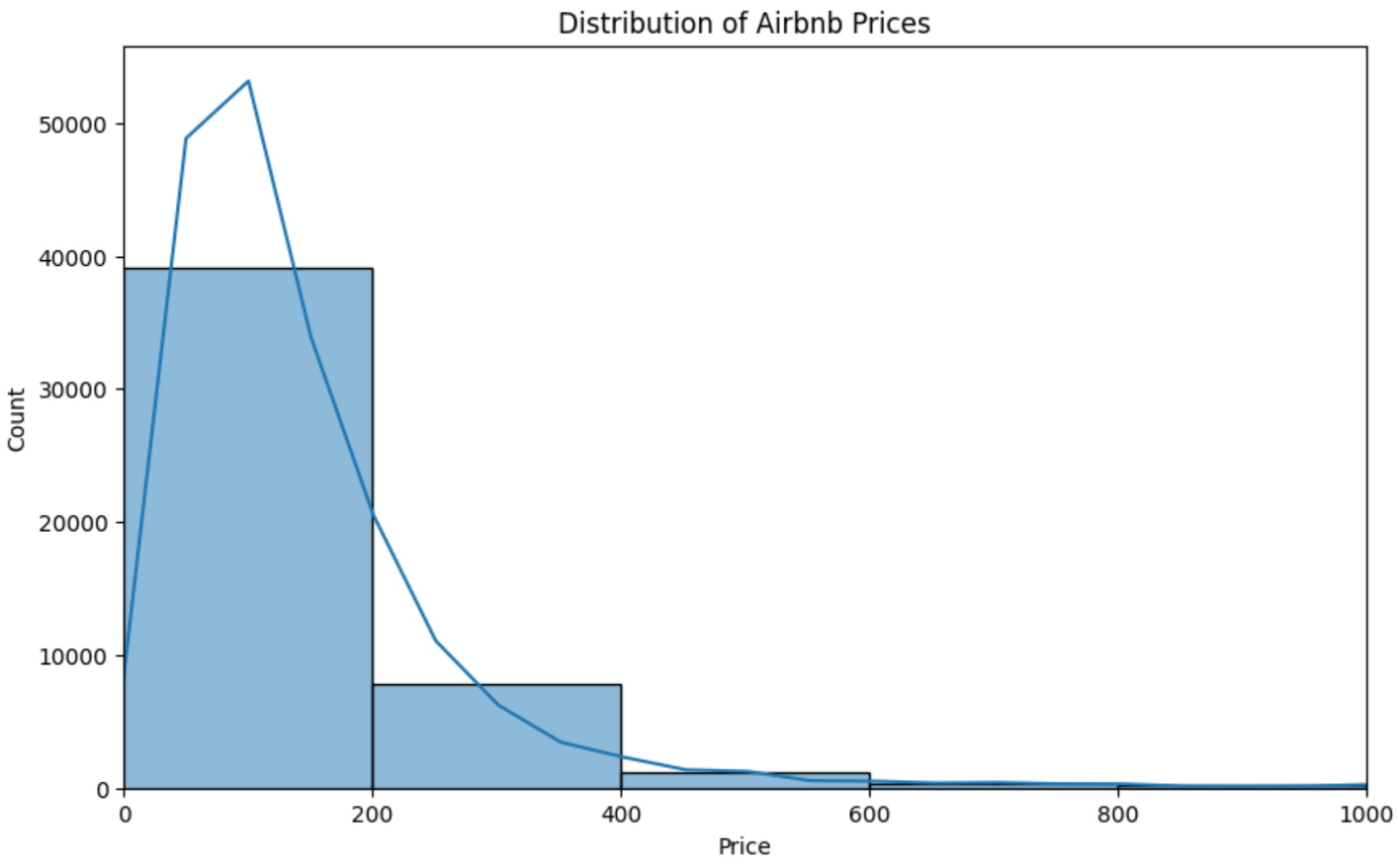
```
Missing values before cleaning:

id                            0
host_name                     0
neighbourhood_group           0
latitude                      0
longitude                     0
room_type                     0
price                         0
number_of_reviews             0
reviews_per_month             0
availability_365              0
last_review               10047
dtype: int64
```

In [13]:
```
df.dropna(subset=['host_name'], inplace=True) #Drop rows where 'name' or 'host_name' is missing
df['reviews_per_month'].fillna(0, inplace=True) #Fill missing 'reviews_per_month' with 0
print("\n Missing values after cleaning:\n")
print(df.isnull().sum())
```

```
 Missing values after cleaning:

id                            0
host_name                     0
neighbourhood_group           0
latitude                      0
longitude                     0
room_type                     0
price                         0
number_of_reviews             0
reviews_per_month             0
availability_365              0
last_review               10047
dtype: int64
```

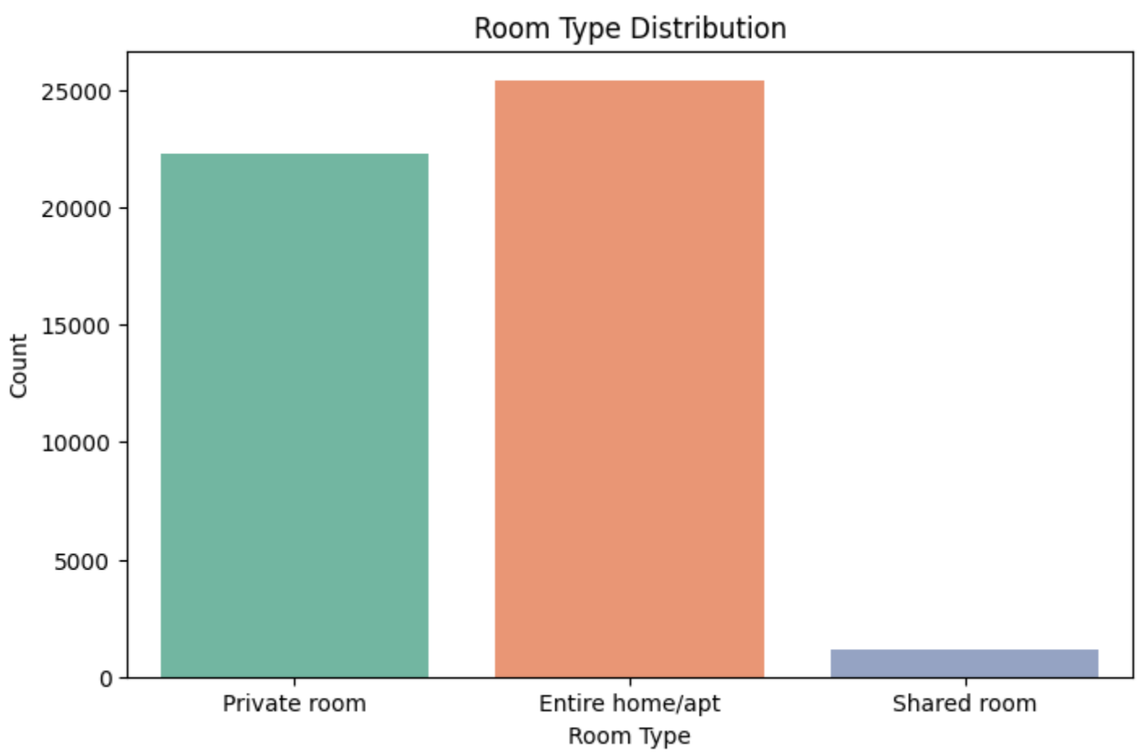
```
C:\Users\JESUS\AppData\Local\Temp\ipykernel_6924\3216078995.py:2: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.


  df['reviews_per_month'].fillna(0, inplace=True) #Fill missing 'reviews_per_month' with 0
```

In [14]:
```
#Distribution of Listing Prices
plt.figure(figsize=(10, 6))
sns.histplot(df['price'], bins=50, kde=True)
plt.title('Distribution of Airbnb Prices')
plt.xlabel('Price')
plt.ylabel('Count')
plt.xlim(0, 1000)  # Limit x-axis to ignore extreme outliers
plt.show()
```



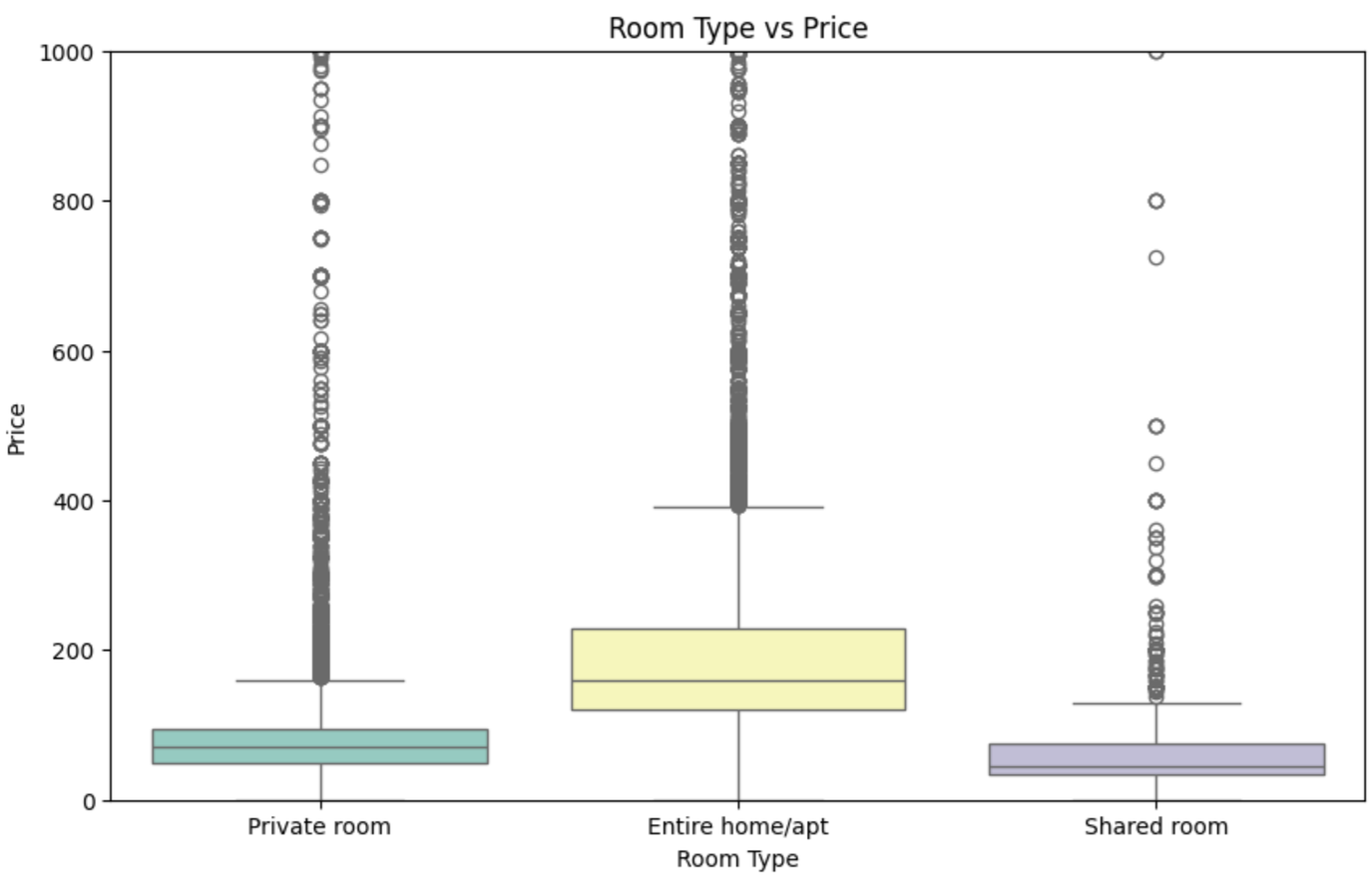
In [15]:
```
#  Frequency of each Room Type
plt.figure(figsize=(8, 5))
sns.countplot(x='room_type', data=df, palette='Set2')
plt.title('Room Type Distribution')
plt.xlabel('Room Type')
plt.ylabel('Count')
plt.show()
```

```
C:\Users\JESUS\AppData\Local\Temp\ipykernel_6924\2749544932.py:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.countplot(x='room_type', data=df, palette='Set2')
```



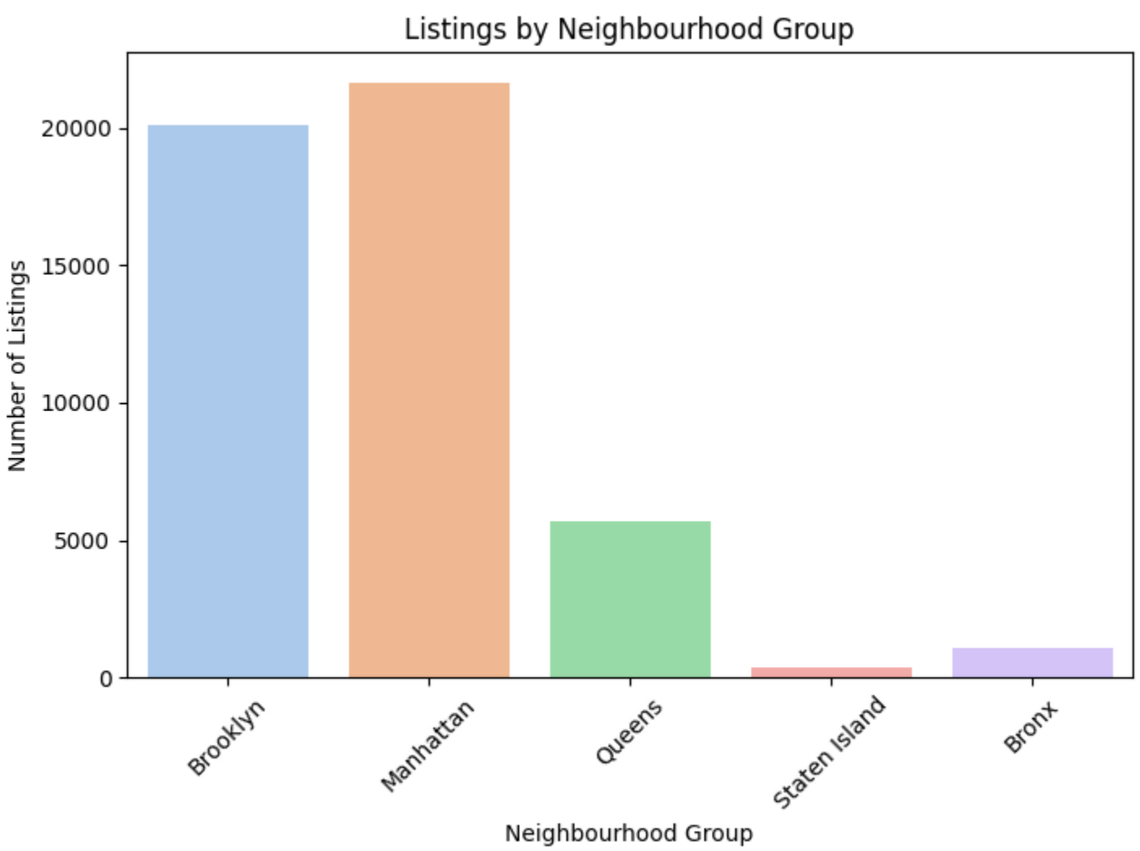
In [16]:
```
#Compare price across room types
plt.figure(figsize=(10, 6))
sns.boxplot(x='room_type', y='price', data=df, palette='Set3')
plt.title('Room Type vs Price')
plt.xlabel('Room Type')
plt.ylabel('Price')
plt.ylim(0, 1000)  # Limit y-axis to reduce outlier noise
plt.show()
```

```
C:\Users\JESUS\AppData\Local\Temp\ipykernel_6924\3549648928.py:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.boxplot(x='room_type', y='price', data=df, palette='Set3')
```

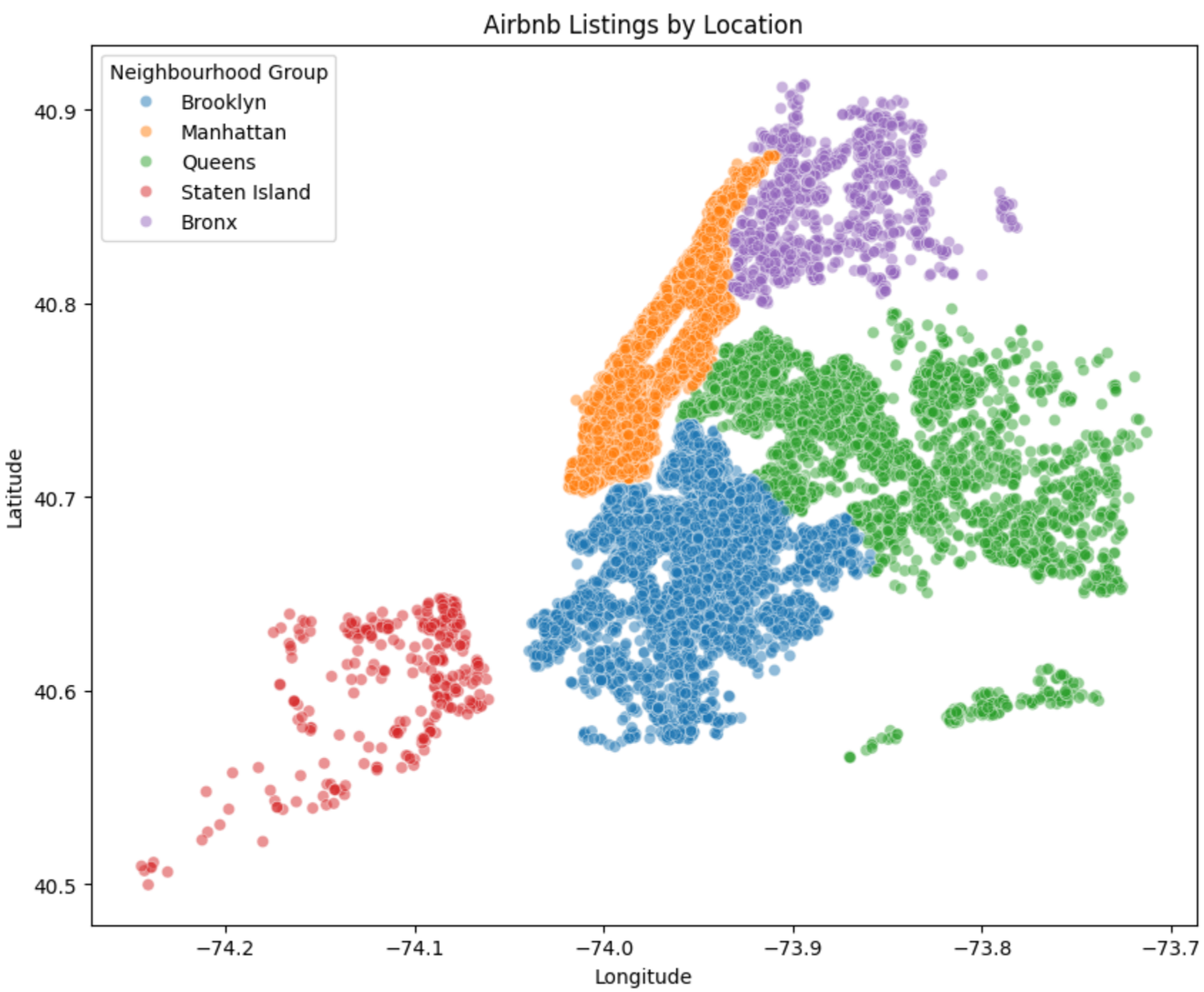


In [17]:
```
# Count of listings per neighbourhood group
plt.figure(figsize=(8, 5))
sns.countplot(x='neighbourhood_group', data=df, palette='pastel')
plt.title('Listings by Neighbourhood Group')
plt.xlabel('Neighbourhood Group')
plt.ylabel('Number of Listings')
plt.xticks(rotation=45)
plt.show()
```

```
C:\Users\JESUS\AppData\Local\Temp\ipykernel_6924\2025867288.py:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.countplot(x='neighbourhood_group', data=df, palette='pastel')
```

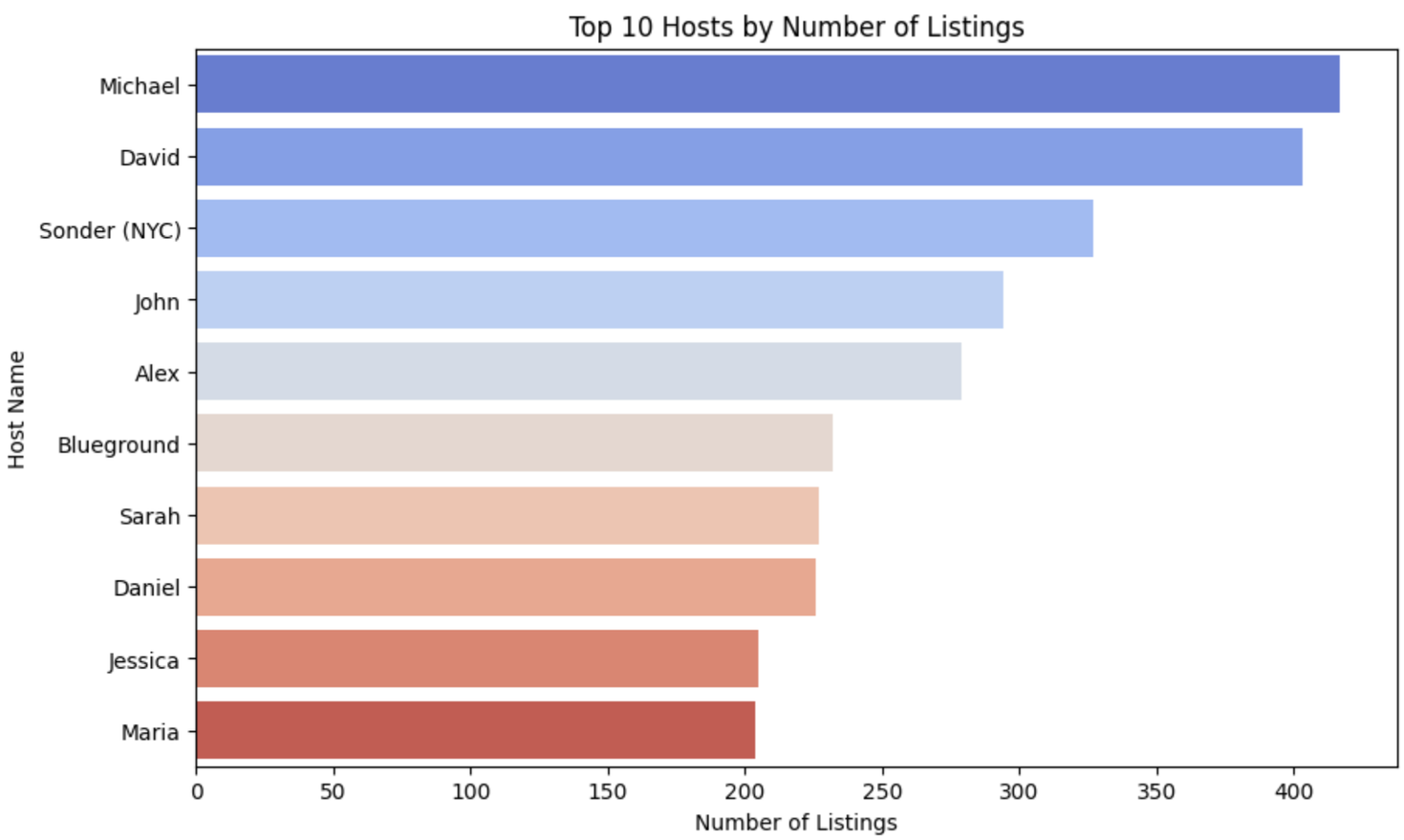


In [18]:
```
# Scatter plot of listings by location
plt.figure(figsize=(10, 8))
sns.scatterplot(data=df, x='longitude', y='latitude', hue='neighbourhood_group', alpha=0.5)
plt.title('Airbnb Listings by Location')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.legend(title='Neighbourhood Group')
plt.show()
# This gives a geographical idea of where listings are located in the city
```



In [19]:
```
# Hosts with the most listings
top_hosts = df['host_name'].value_counts().head(10)

plt.figure(figsize=(10, 6))
sns.barplot(x=top_hosts.values, y=top_hosts.index, palette='coolwarm')
plt.title('Top 10 Hosts by Number of Listings')
plt.xlabel('Number of Listings')
plt.ylabel('Host Name')
plt.show()
```

```
C:\Users\JESUS\AppData\Local\Temp\ipykernel_6924\2770422307.py:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

  sns.barplot(x=top_hosts.values, y=top_hosts.index, palette='coolwarm')
```



In [20]:
```python
# Availability scatter plot
plt.figure(figsize=(10, 6))
sns.scatterplot(x='availability_365', y='price', data=df, alpha=0.4)
plt.title('Availability vs Price')
plt.xlabel('Availability (Days per Year)')
plt.ylabel('Price')
plt.xlim(0, 370)
plt.ylim(0, 1000)
plt.show()
```

🧠 Key Insights from Airbnb EDA

- 🏠 **Most listings are for entire homes/apartments**, followed by private rooms, making them the most common room types.
- 💰 **Prices mostly fall below ₹1000**, but a few high-priced outliers indicate the presence of luxury properties.
- 🕵️‍♂️ **Top hosts manage multiple listings**, with some having over 100 properties, showing a trend toward commercial hosting.
- 📍 **Listings are denser in popular neighbourhoods** like Manhattan and Brooklyn, suggesting high demand in these areas.
- 🔁 **No strong correlation between availability and price** — pricing appears independent of how often the property is available.
- ✍️ **Many listings have no reviews**, but those with reviews average **1–2 reviews per month**, indicating moderate user engagement.