

Airbnb Data Analysis Project

Objective:

To perform exploratory analysis on Airbnb Listings data to uncover pricing trends, listing types , and neighbourhood patterns for actionable business insights.

Dataset Source:

[Kaggle - Airbnb Listings](#)

file used : 'Listings.csv'

Tools and Libraries Used:

python(Pandas ,Numpy)

Data visualization: Seaborn ,matplotlib

jupyter notebook

Key Analysis Performed:

--Data Cleaning (nulls,outliers , formatting)

--Exploratory Data Analysis(EDA)

--Visualization: Room types , Price trends , Reviews , Availability

--Business Insights and recommendation

Key Insights:

- Entire homes are most preferred but costliest
- Some neighbourhoods are overcrowded , others underutilized
- Availability doesn't always correlate with reviews (underperforming listings)
- Pricing distribution is heavily skewed , showing need for dynamic pricing.

```
In [15]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_csv("listings.csv")
df.head()
```

Out[15]:

	id	listing_url	scrape_id	last_scraped	source	name	description	neighborhood_overview
0	2539	https://www.airbnb.com/rooms/2539	20250617032754	2025-06-17	city scrape	Superfast Wi-Fi. Clean & quiet home by the park	Bright, serene room in a renovated apartment h...	Close to Prospect and Historic Ditmas
1	2595	https://www.airbnb.com/rooms/2595	20250617032754	2025-06-17	city scrape	Skylit Midtown Manhattan Spacious Studio	Beautiful, spacious skylit studio in the heart...	Centrally located in heart of Manhattan
2	5136	https://www.airbnb.com/rooms/5136	20250617032754	2025-06-17	city scrape	Spacious Family Friendly Duplex w/ Patio + Yard	We welcome you to stay in our lovely 2 br, 130...	
3	6848	https://www.airbnb.com/rooms/6848	20250617032754	2025-06-17	city scrape	Only 2 stops to Manhattan studio	Comfortable studio apartment with super comfor...	
4	6872	https://www.airbnb.com/rooms/6872	20250617032754	2025-06-17	previous scrape	Uptown Sanctuary w/ Private Bath (Month to Month)	This charming distancing-friendly month-to-mon...	This sweet Ha... sanctuary is a 10 minu

5 rows × 79 columns



In [17]:

```
print("Shape of dataset:", df.shape)
df.info()
df.isnull().sum().sort_values(ascending = False)
```

Shape of dataset: (36322, 79)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36322 entries, 0 to 36321
Data columns (total 79 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   id               36322 non-null   int64  
 1   listing_url      36322 non-null   object  
 2   scrape_id        36322 non-null   int64  
 3   last_scraped     36322 non-null   object  
 4   source            36322 non-null   object  
 5   name              36320 non-null   object  
 6   description       35374 non-null   object  
 7   neighborhood_overview 19084 non-null   object  
 8   picture_url      36322 non-null   object  
 9   host_id           36322 non-null   int64  
 10  host_url          36322 non-null   object  
 11  host_name         36307 non-null   object  
 12  host_since        36307 non-null   object  
 13  host_location     28913 non-null   object  
 14  host_about        20956 non-null   object  
 15  host_response_time 21550 non-null   object  
 16  host_response_rate 21550 non-null   object  
 17  host_acceptance_rate 21720 non-null   object  
 18  host_is_superhost 35830 non-null   object  
 19  host_thumbnail_url 36307 non-null   object  
 20  host_picture_url  36307 non-null   object  
 21  host_neighbourhood 29034 non-null   object  
 22  host_listings_count 36307 non-null   float64 
 23  host_total_listings_count 36307 non-null   float64 
 24  host_verifications 36307 non-null   object  
 25  host_has_profile_pic 36307 non-null   object  
 26  host_identity_verified 36307 non-null   object  
 27  neighbourhood      19085 non-null   object  
 28  neighbourhood_cleansed 36322 non-null   object  
 29  neighbourhood_group_cleansed 36322 non-null   object  
 30  latitude           36322 non-null   float64 
 31  longitude          36322 non-null   float64 
 32  property_type      36322 non-null   object  
 33  room_type          36322 non-null   object  
 34  accommodates       36322 non-null   int64  
 35  bathrooms          21682 non-null   float64
```

36	bathrooms_text	36286	non-null	object
37	bedrooms	30274	non-null	float64
38	beds	21620	non-null	float64
39	amenities	36322	non-null	object
40	price	21459	non-null	object
41	minimum_nights	36322	non-null	int64
42	maximum_nights	36322	non-null	int64
43	minimum_minimum_nights	36322	non-null	int64
44	maximum_minimum_nights	36322	non-null	int64
45	minimum_maximum_nights	36322	non-null	int64
46	maximum_maximum_nights	36322	non-null	int64
47	minimum_nights_avg_ntm	36322	non-null	float64
48	maximum_nights_avg_ntm	36322	non-null	float64
49	calendar_updated	0	non-null	float64
50	has_availability	30705	non-null	object
51	availability_30	36322	non-null	int64
52	availability_60	36322	non-null	int64
53	availability_90	36322	non-null	int64
54	availability_365	36322	non-null	int64
55	calendar_last_scraped	36322	non-null	object
56	number_of_reviews	36322	non-null	int64
57	number_of_reviews_ltm	36322	non-null	int64
58	number_of_reviews_l30d	36322	non-null	int64
59	availability_eoy	36322	non-null	int64
60	number_of_reviews_ly	36322	non-null	int64
61	estimated_occupancy_l365d	36322	non-null	int64
62	estimated_revenue_l365d	21459	non-null	float64
63	first_review	25171	non-null	object
64	last_review	25171	non-null	object
65	review_scores_rating	25171	non-null	float64
66	review_scores_accuracy	25161	non-null	float64
67	review_scores_cleanliness	25170	non-null	float64
68	review_scores_checkin	25157	non-null	float64
69	review_scores_communication	25165	non-null	float64
70	review_scores_location	25154	non-null	float64
71	review_scores_value	25155	non-null	float64
72	license	5329	non-null	object
73	instant_bookable	36322	non-null	object
74	calculated_host_listings_count	36322	non-null	int64
75	calculated_host_listings_count_entire_homes	36322	non-null	int64
76	calculated_host_listings_count_private_rooms	36322	non-null	int64
77	calculated_host_listings_count_shared_rooms	36322	non-null	int64

```
78 reviews_per_month          25171 non-null float64
dtypes: float64(19), int64(24), object(36)
memory usage: 21.9+ MB

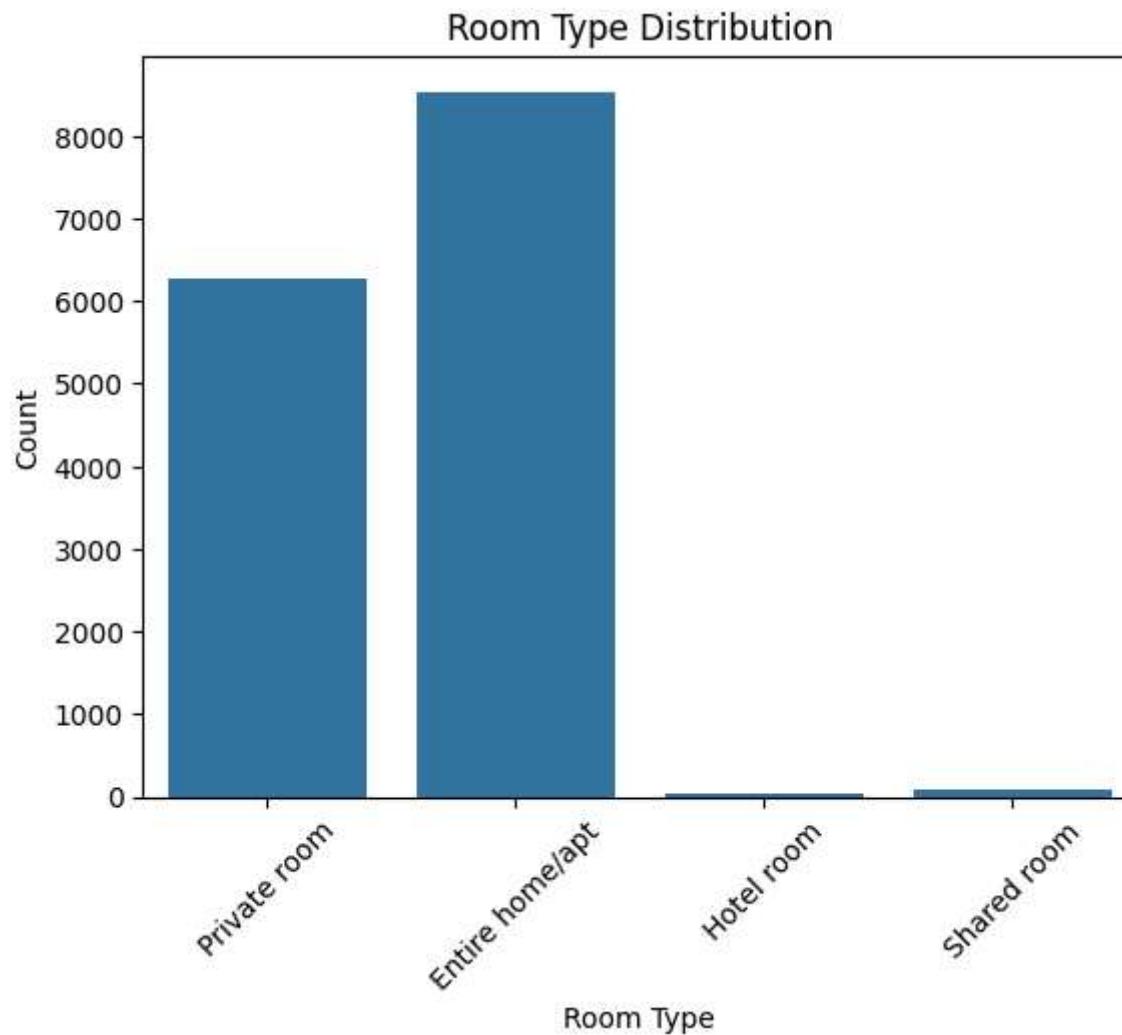
Out[17]: calendar_updated      36322
license                      30993
neighborhood_overview        17238
neighbourhood                 17237
host_about                     15366
...
instant_bookable                  0
calculated_host_listings_count      0
calculated_host_listings_count_entire_homes 0
calculated_host_listings_count_private_rooms 0
calculated_host_listings_count_shared_rooms 0
Length: 79, dtype: int64
```

```
In [18]: df=df[['id','name','host_id','neighbourhood','room_type','price','minimum_nights','number_of_reviews',
           'last_review','reviews_per_month','calculated_host_listings_count','availability_365']]
```

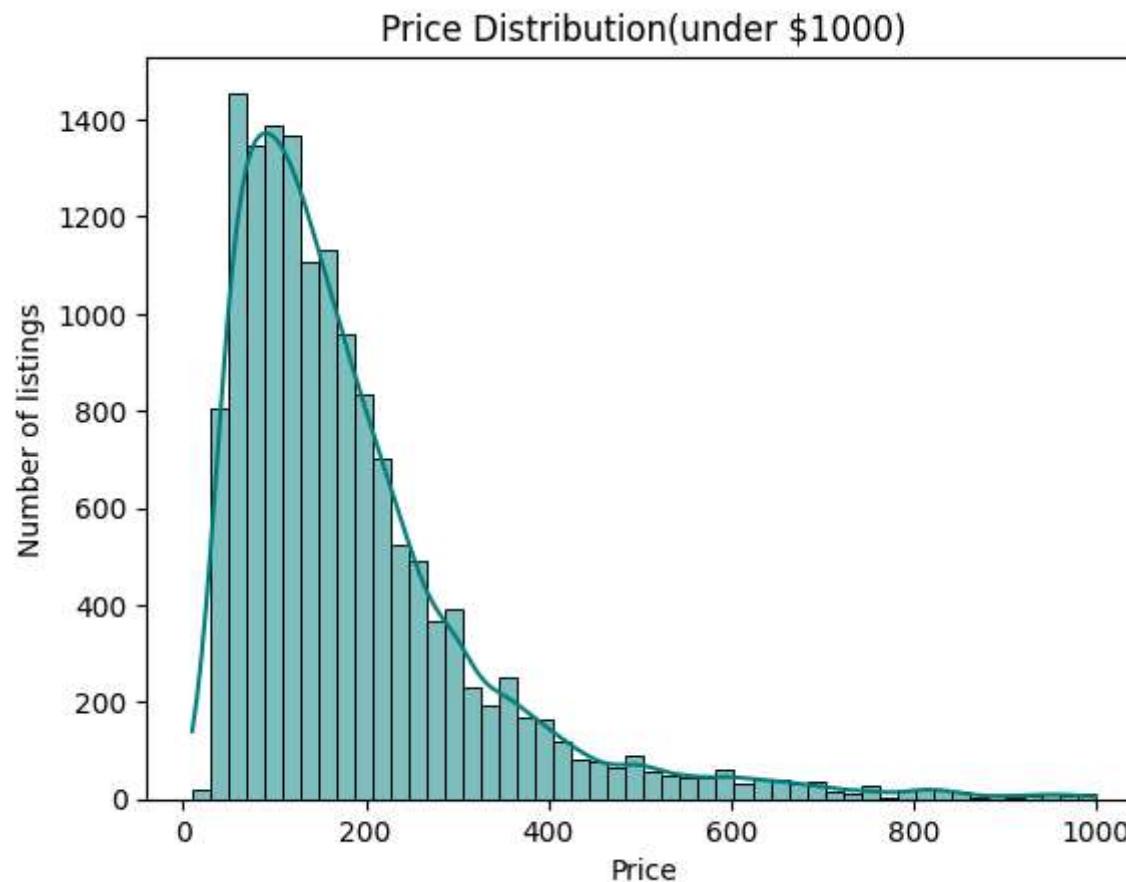
```
In [19]: df['reviews_per_month']=df['reviews_per_month'].fillna(0)
df.dropna(subset=['last_review'] , inplace = True)
```

```
In [20]: df['price'] = df['price'].replace(r'[\$,]', '', regex = True).astype(float)
df=df[df['price'] > 0]
df=df[df['price']<1000]
```

```
In [22]: sns.countplot(data=df,x='room_type')
plt.title('Room Type Distribution')
plt.xlabel('Room Type')
plt.ylabel('Count')
plt.xticks(rotation = 45)
plt.show()
```

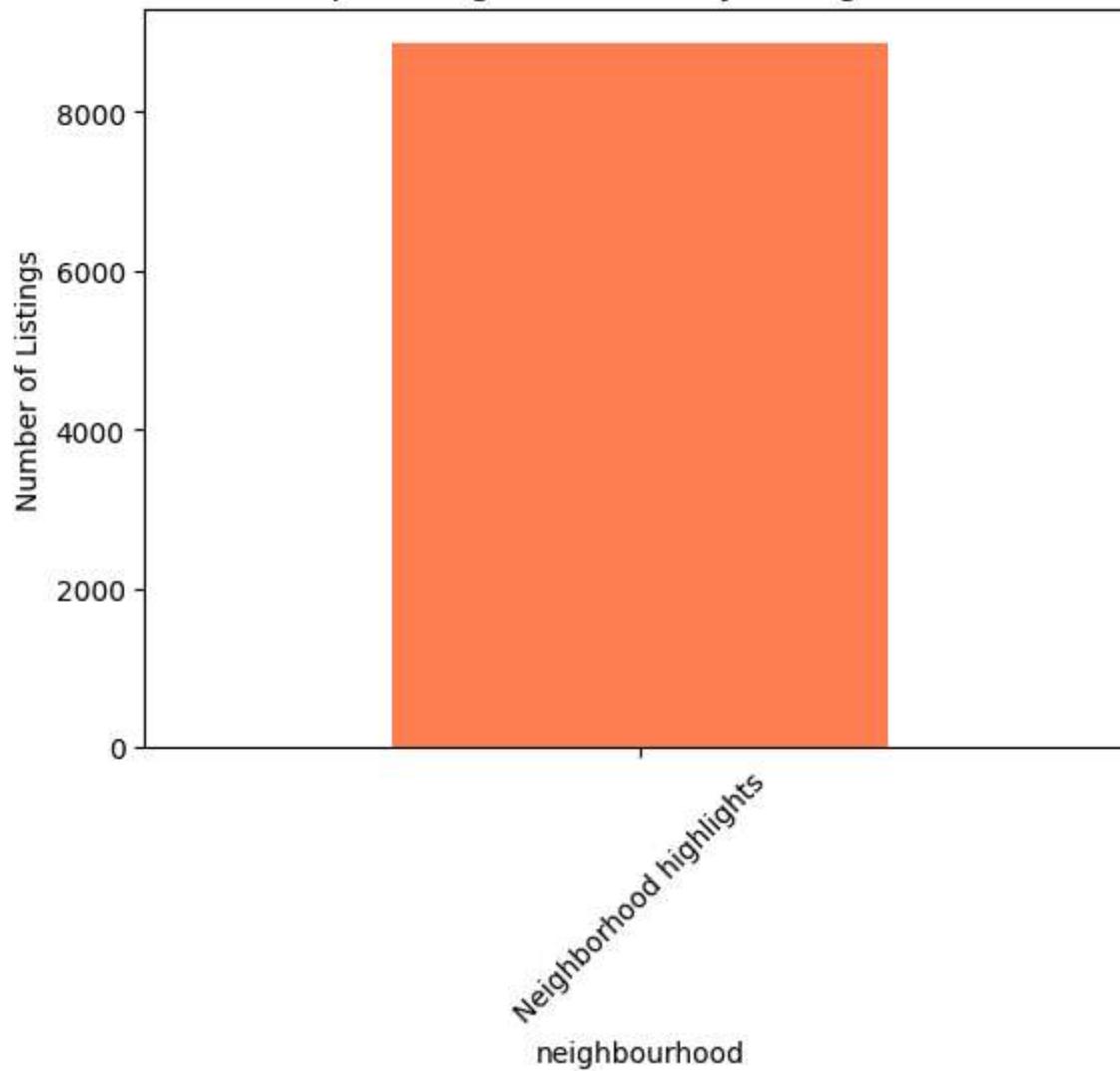


```
In [23]: sns.histplot(df['price'], bins=50, kde=True, color='teal')
plt.title('Price Distribution (under $1000)')
plt.xlabel('Price')
plt.ylabel('Number of listings')
plt.show()
```

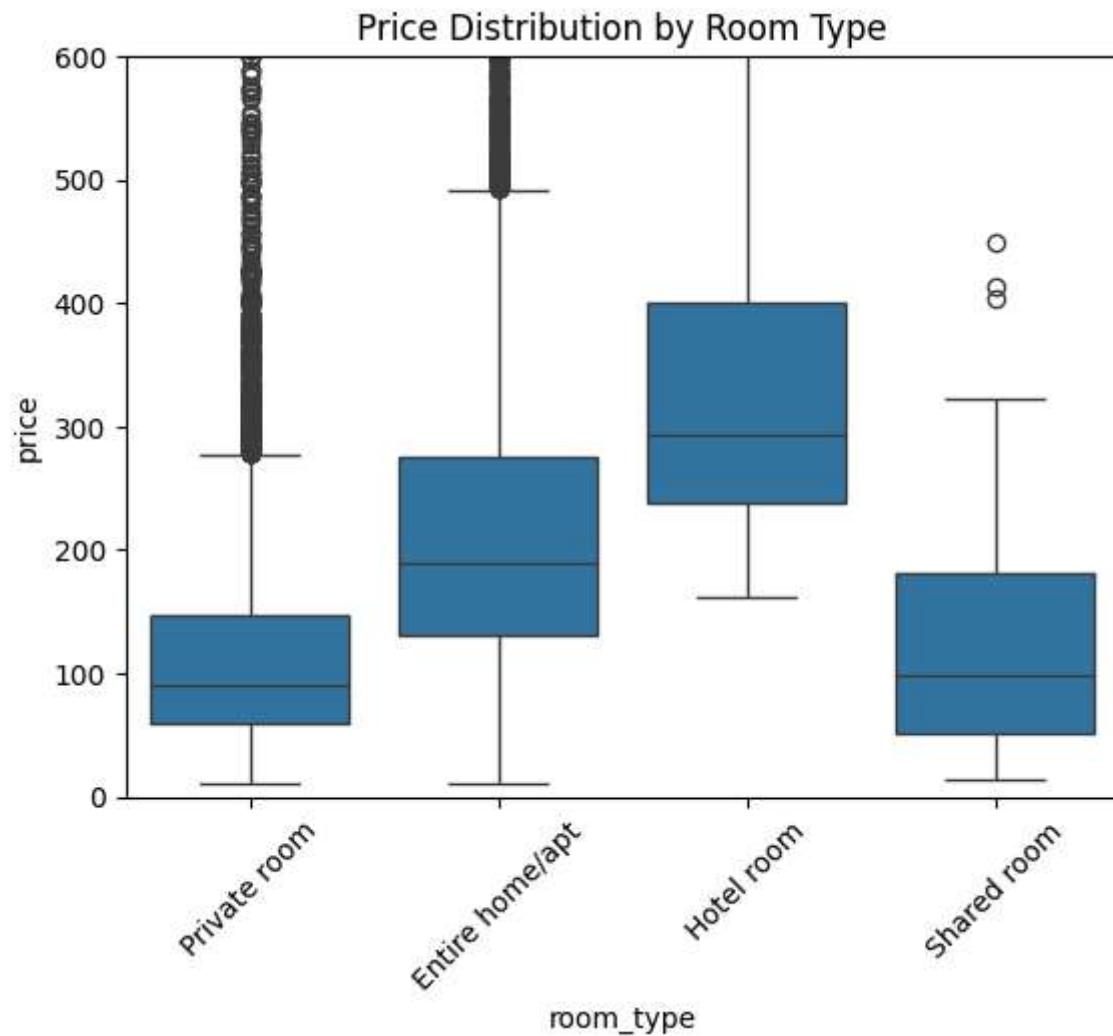


```
In [24]: top_neigh = df['neighbourhood'].value_counts().head(10)
top_neigh.plot(kind='bar', color ='coral')
plt.title('Top 10 Neighbourhoods by Listings Count')
plt.ylabel('Number of Listings')
plt.xticks(rotation = 45)
plt.show()
```

Top 10 Neighbourhoods by Listings Count

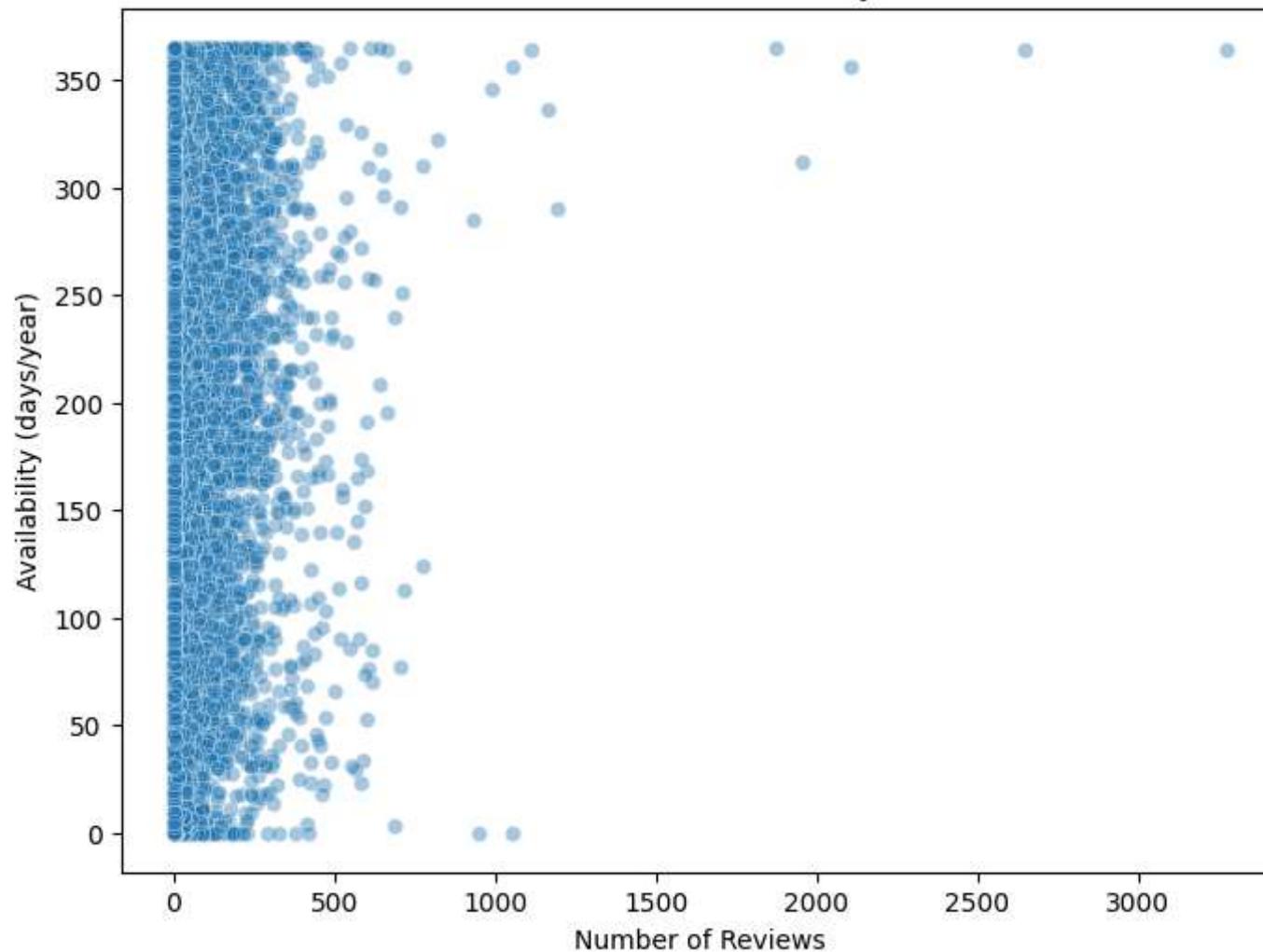


```
In [26]: sns.boxplot(data=df,x='room_type',y='price')
plt.title('Price Distribution by Room Type')
plt.xticks(rotation=45)
plt.ylim(0,600)
plt.show()
```



```
In [30]: plt.figure(figsize=(8,6))
sns.scatterplot(data=df,x='number_of_reviews',y='availability_365',alpha =0.4)
plt.title('Reviews vs Availability')
plt.xlabel('Number of Reviews')
plt.ylabel('Availability (days/year)')
plt.show()
```

Reviews vs Availability



Business Insights From EDA ---

- 1 . Most Common Room Type:** The majority of listings are of type "*Entire home/apt*", indicating users completee privacy. This room type also tends to be priced higher.
- 2 . Affordable Neighbourhood:** Certain neighbourhoods(like XYZ -add from your graph) have a high number of listings with relatively lower average price.These areas are good for budget travelers.

3 . High Availability with Low Reviews: Some listings show high availability (close to 365 days), but very low number of reviews-- may indicate new or underperforming listings.

4 . Price Distribution Skewed: Prices are heavily skewed --most listings fall under \$300 , but a few luxury listings raise the average. This suggests Airbnb should use **median pricing** for better analysis.

5 . Top 10 Neighbourhood Dominate Supply: A small number of neighbourhood have the highest number of listings, which may indicate **over-saturation** in certain areas, and **growth opportunity in underserved zones**

Summary:

This analysis explored Airbnb Listings data to uncover pricing patterns, room type distribution, neighbourhood saturation, and user engagement. Key trends in availability and review activity were also identified to help optimize platform performance.

Business Recommendation

1. **Dynamic Pricing Strategy:** Airbnb should recommend hosts to adjust pricing based on room type and neighbourhood median to stay competitive.
2. **Promote Private/Shared Rooms in Expensive Areas:** To serve budget travelers in premium areas like XYZ , encourage hosting of affordable shared or private rooms.
3. **Highlight Underused Listings:** Listings with high availability but few reviews should be flagged for promotion or support.
4. **Neighbourhood Expansion Opportunity:** Less crowded neighbourhoods with decent demand are ideal for expansion or marketing push.
5. **Data Driven Host Guidance:** Use these patterns to guide new hosts with optimal pricing and availability suggestions.

In []: