New York Airbnb EDA Project with Python

In [3]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
```

In [5]:

```
data=pd.read_csv('airbnb_dataset.csv',encoding_errors='ignore')
data.head()
```

Out[5]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude
0	1.312228e+06	Rental unit in Brooklyn · ★5.0 · 1 bedroom	7130382	Walter	Brooklyn	Clinton Hill	40.683710	-73.964610
1	4.527754e+07	Rental unit in New York · ★4.67 · 2 bedrooms ·...	51501835	Jeniffer	Manhattan	Hell's Kitchen	40.766610	-73.988100
2	9.710000e+17	Rental unit in New York · ★4.17 · 1 bedroom · ...	528871354	Joshua	Manhattan	Chelsea	40.750764	-73.994605
3	3.857863e+06	Rental unit in New York · ★4.64 · 1 bedroom · ...	19902271	John And Catherine	Manhattan	Washington Heights	40.835600	-73.942500
4	4.089661e+07	Condo in New York · ★4.91 · Studio · 1 bed · 1...	61391963	Stay With Vibe	Manhattan	Murray Hill	40.751120	-73.978600

5 rows × 22 columns

In [196…

```
data.shape
```

Out[196…

```
(20770, 22)
```

In [198…

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20770 entries, 0 to 20769
Data columns (total 22 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   id                              20770 non-null  float64
 1   name                            20770 non-null  object 
 2   host_id                         20770 non-null  int64  
 3   host_name                       20770 non-null  object 
 4   neighbourhood_group             20770 non-null  object 
 5   neighbourhood                   20763 non-null  object 
 6   latitude                        20763 non-null  float64
 7   longitude                       20763 non-null  float64
 8   room_type                       20763 non-null  object 
 9   price                           20736 non-null  float64
 10  minimum_nights                  20763 non-null  float64
 11  number_of_reviews               20763 non-null  float64
 12  last_review                     20763 non-null  object 
 13  reviews_per_month               20763 non-null  float64
 14  calculated_host_listings_count  20763 non-null  float64
 15  availability_365                20763 non-null  float64
 16  number_of_reviews_ltm           20763 non-null  float64
 17  license                         20770 non-null  object 
 18  rating                          20770 non-null  object 
 19  bedrooms                        20770 non-null  object 
 20  beds                            20770 non-null  int64  
 21  baths                           20770 non-null  object 
dtypes: float64(10), int64(2), object(10)
memory usage: 3.5+ MB
```

In [200…

```
data.describe()
```

Out[200…

	id	host_id	latitude	longitude	price	minimum_nights	number_of_reviev
count	2.077000e+04	2.077000e+04	20763.000000	20763.000000	20736.000000	20763.000000	20763.0000
mean	3.033858e+17	1.749049e+08	40.726821	-73.939179	187.714940	28.558493	42.6106
std	3.901221e+17	1.725657e+08	0.060293	0.061403	1023.245124	33.532697	73.5234
min	2.595000e+03	1.678000e+03	40.500314	-74.249840	10.000000	1.000000	1.0000
25%	2.707260e+07	2.041184e+07	40.684159	-73.980755	80.000000	30.000000	4.0000
50%	4.992852e+07	1.086990e+08	40.722890	-73.949597	125.000000	30.000000	14.0000
75%	7.220000e+17	3.143997e+08	40.763106	-73.917475	199.000000	30.000000	49.0000
max	1.050000e+18	5.504035e+08	40.911147	-73.713650	100000.000000	1250.000000	1865.0000

Data Cleaning

In [7]:

```
data.isnull().sum()
```

Out[7]:

```
id                                 0
name                               0
host_id                            0
host_name                          0
neighbourhood_group                0
neighbourhood                      7
latitude                           7
longitude                          7
room_type                          7
price                             34
minimum_nights                     7
number_of_reviews                  7
last_review                        7
reviews_per_month                  7
calculated_host_listings_count     7
availability_365                   7
number_of_reviews_ltm              7
license                            0
rating                             0
bedrooms                           0
beds                               0
baths                              0
dtype: int64
```

In [9]:

```
data.dropna(inplace=True)
```

In [11]:

```
data.isnull().sum()
```

Out[11]:

```
id                                0
name                              0
host_id                           0
host_name                         0
neighbourhood_group               0
neighbourhood                     0
latitude                          0
longitude                         0
room_type                         0
price                             0
minimum_nights                    0
number_of_reviews                 0
last_review                       0
reviews_per_month                 0
calculated_host_listings_count    0
availability_365                  0
number_of_reviews_ltm             0
license                           0
rating                            0
bedrooms                          0
beds                              0
baths                             0
dtype: int64
```

In [13]:

```
#Duplicate Rows
data.duplicated().sum()
```

Out[13]:

```
12
```

In [15]:

```
data[data.duplicated()]
```

Out[15]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longit
6	4.527754e+07	Rental unit in New York · ★4.67 · 2 bedrooms ·...	51501835	Jeniffer	Manhattan	Hell's Kitchen	40.766610	-73.988
7	9.710000e+17	Rental unit in New York · ★4.17 · 1 bedroom · ...	528871354	Joshua	Manhattan	Chelsea	40.750764	-73.994
8	3.857863e+06	Rental unit in New York · ★4.64 · 1 bedroom · ...	19902271	John And Catherine	Manhattan	Washington Heights	40.835600	-73.942
9	4.089661e+07	Condo in New York · ★4.91 · Studio · 1 bed · 1...	61391963	Stay With Vibe	Manhattan	Murray Hill	40.751120	-73.978
10	4.958498e+07	Rental unit in New York · ★5.0 · 1 bedroom · 1...	51501835	Jeniffer	Manhattan	Hell's Kitchen	40.759950	-73.992
20736	7.990000e+17	Rental unit in New York · 2 bedrooms · 2 beds ...	224733902	CozySuites Copake	Manhattan	Upper East Side	40.768970	-73.957
20737	5.930000e+17	Rental unit in New York · ★4.79 · 2 bedrooms ·...	23219783	Rob	Manhattan	West Village	40.730220	-74.002
20738	9.230000e+17	Loft in New York · ★4.33 · 1 bedroom · 2 beds ...	520265731	Rodrigo	Manhattan	Greenwich Village	40.728390	-73.999
20739	1.336161e+07	Rental unit in New York · ★4.89 · 2 bedrooms ·...	8961407	Jamie	Manhattan	Harlem	40.805700	-73.946
20740	5.119566e+07	Rental unit in New York · Studio · 1 bed · 1 bath	51501835	Jeniffer	Manhattan	Chinatown	40.718360	-73.995
20741	2.523473e+07	Rental unit in New York · ★4.41 · 1 bedroom · ...	1497427	Mara	Manhattan	Upper East Side	40.774030	-73.950

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longit
20742	3.339399e+06	Rental unit in New York · ★4.73 · 1 bedroom · ...	2119276	Urban Furnished	Manhattan	West Village	40.732030	-74.006

12 rows × 22 columns

In [17]:

```
#changibg data types
data.drop_duplicates(inplace=True)
data.duplicated().sum()
```

Out[17]:

```
0
```

In [19]:

```
data.dtypes
```

Out[19]:

```
id                                float64
name                               object
host_id                             int64
host_name                          object
neighbourhood_group                object
neighbourhood                      object
latitude                          float64
longitude                         float64
room_type                          object
price                             float64
minimum_nights                    float64
number_of_reviews                 float64
last_review                        object
reviews_per_month                 float64
calculated_host_listings_count    float64
availability_365                  float64
number_of_reviews_ltm             float64
license                            object
rating                             object
bedrooms                           object
beds                                int64
baths                              object
dtype: object
```

In [21]:

```
data['id']=data['id'].astype(object)
```

In [23]:

```
data.dtypes
```

Out[23]:

```
id                                 object
name                               object
host_id                             int64
host_name                          object
neighbourhood_group                object
neighbourhood                      object
latitude                          float64
longitude                         float64
room_type                          object
price                             float64
minimum_nights                    float64
number_of_reviews                 float64
last_review                        object
reviews_per_month                 float64
calculated_host_listings_count    float64
availability_365                  float64
number_of_reviews_ltm             float64
license                            object
rating                             object
bedrooms                           object
beds                                int64
baths                              object
dtype: object
```

In [212…

```
data['host_id']=data['host_id'].astype(object)
```

In [214…

```
data.dtypes
```

Out[214…

```
id                                float64
name                               object
host_id                            object
host_name                          object
neighbourhood_group                object
neighbourhood                      object
latitude                          float64
longitude                         float64
room_type                          object
price                             float64
minimum_nights                    float64
number_of_reviews                 float64
last_review                        object
reviews_per_month                 float64
calculated_host_listings_count    float64
availability_365                  float64
number_of_reviews_ltm             float64
license                            object
rating                             object
bedrooms                           object
beds                                int64
baths                              object
dtype: object
```

Exploratory Data Analyis

Univariate Analysis

In [26]:

```
#price Distributiom

data['price']
```

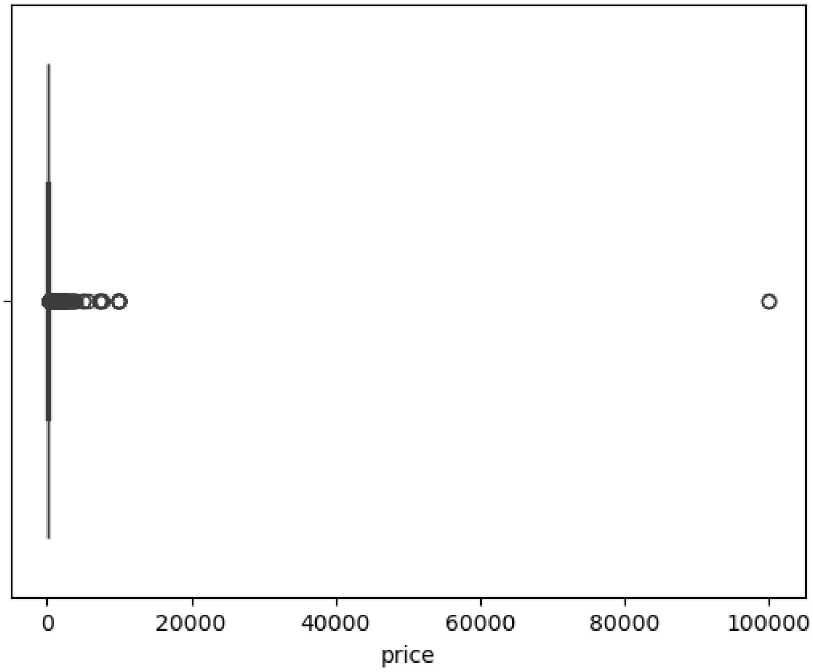
Out[26]:

```
0          55.0
1         144.0
2         187.0
3         120.0
4          85.0
          ...  
20765      45.0
20766     105.0
20767     299.0
20768     115.0
20769     102.0
Name: price, Length: 20724, dtype: float64
```

Boxplot

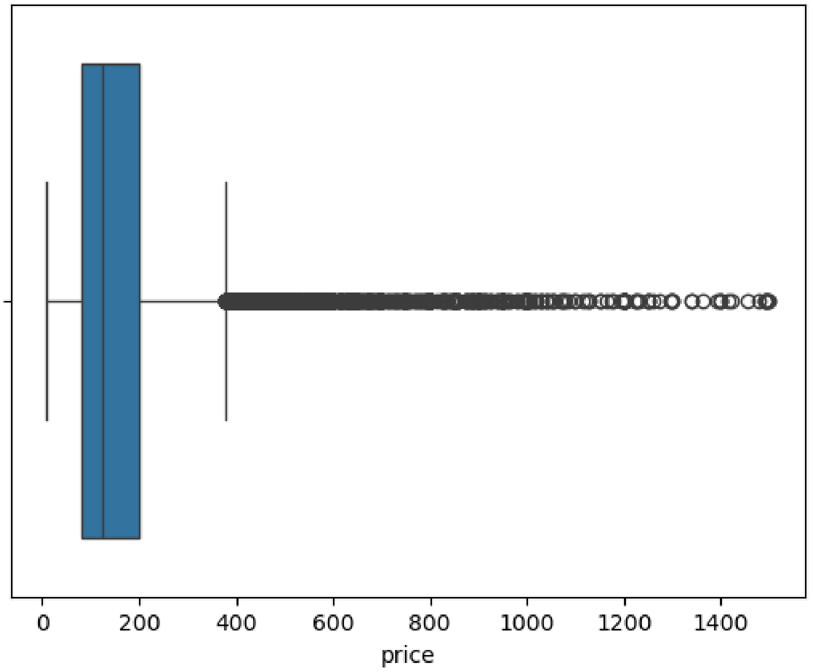
In [28]:

```
#Identifying outliers in price
sns.boxplot(data=data,x='price')
plt.show()
```



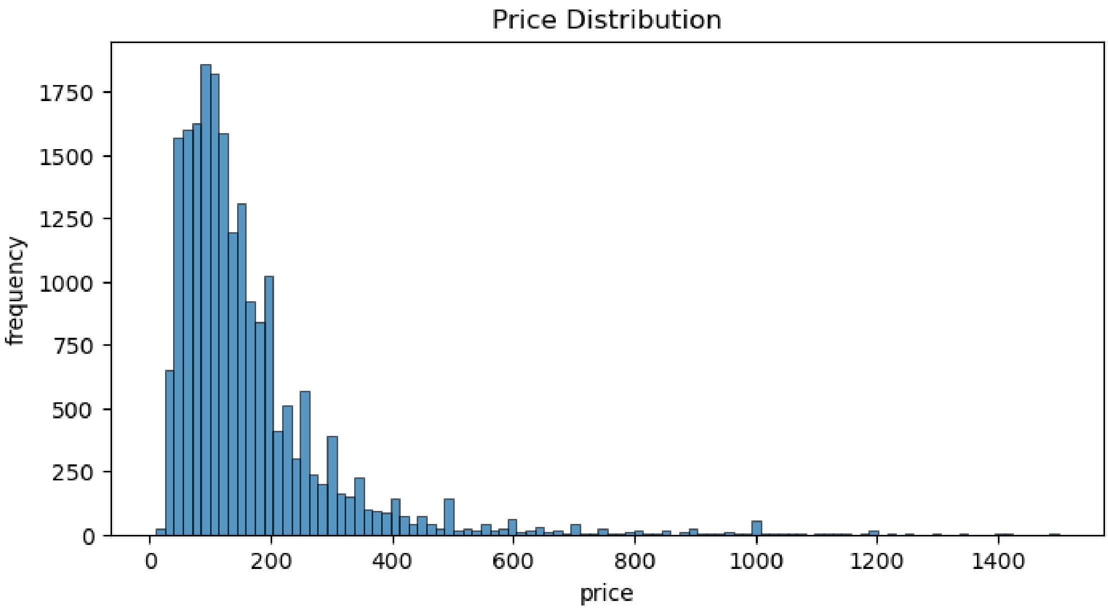
```
In [30]: df=data[data['price']<1500]
```

```
In [32]: sns.boxplot(data=df,x='price')
         plt.show()
```



Price Distribution

```
In [34]: plt.figure(figsize=(8,4))
         sns.histplot(data=df ,x='price',bins=100)
         plt.ylabel('frequency')
         plt.title('Price Distribution')
         plt.show()
```



A significant proportion of Airbnb listings are priced between $0and200$, indicating a prevalence of relatively affordable options.

Price per Bed by Neighbourhood Group

```
In [36]: df.groupby(by='neighbourhood_group')['price'].mean()
```

```
Out[36]: neighbourhood_group
         Bronx            107.990506
         Brooklyn         155.138317
         Manhattan        204.146014
         Queens           121.681939
         Staten Island    118.780069
         Name: price, dtype: float64
```

```
In [38]: #creating new column
         df['price per bed']=df['price']/df['beds']
         df.head()
```

```
C:\Users\KRIPESH\AppData\Local\Temp\ipykernel_13340\2784808981.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.ht
ml#returning-a-view-versus-a-copy
  df['price per bed']=df['price']/df['beds']
```

Out[38]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	lo
0	1312228.0	Rental unit in Brooklyn · ★5.0 · 1 bedroom	7130382	Walter	Brooklyn	Clinton Hill	40.683710	-73
1	45277537.0	Rental unit in New York · ★4.67 · 2 bedrooms ·...	51501835	Jeniffer	Manhattan	Hell's Kitchen	40.766610	-73
2	971000000000000000.0	Rental unit in New York · ★4.17 · 1 bedroom · ...	528871354	Joshua	Manhattan	Chelsea	40.750764	-73
3	3857863.0	Rental unit in New York · ★4.64 · 1 bedroom · ...	19902271	John And Catherine	Manhattan	Washington Heights	40.835600	-73
4	40896611.0	Condo in New York · ★4.91 · Studio · 1 bed · 1...	61391963	Stay With Vibe	Manhattan	Murray Hill	40.751120	-73

5 rows × 23 columns

In [169…

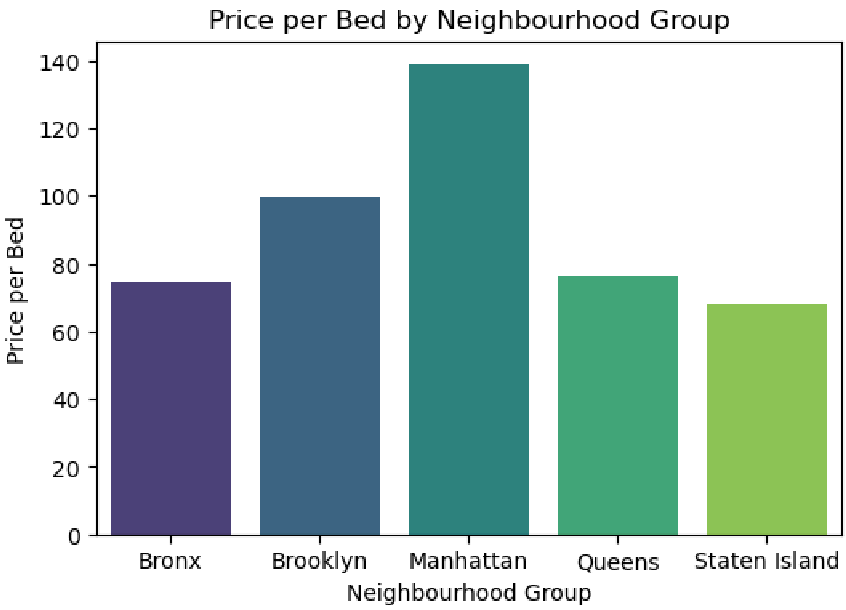
```
df.groupby(by='neighbourhood_group')['price per bed'].mean()
```

Out[169…

```
neighbourhood_group
Bronx             74.713639
Brooklyn          99.788493
Manhattan        138.708057
Queens            76.336210
Staten Island     67.728101
Name: price per bed, dtype: float64
```

In [52]:

```
bed_price=df.groupby(by='neighbourhood_group')['price per bed'].mean()
plt.figure(figsize=(6,4))
plt.xlabel('Neighbourhood Group')
plt.ylabel('Price per Bed')
plt.title('Price per Bed by Neighbourhood Group')
sns.barplot(x='neighbourhood_group',y='price per bed',data=bed_price.reset_index(),hue='neighbourhood_gr
plt.show()
```



The average price per bed in Manhattan is significantly higher compared to other neighborhood groups, exceeding $140. This indicates that Manhattan is the most expensive area for Airbnb accommodations, likely due to its prime location, tourist attractions, and high demand.

Price per Bed by Room Type

In [60]:

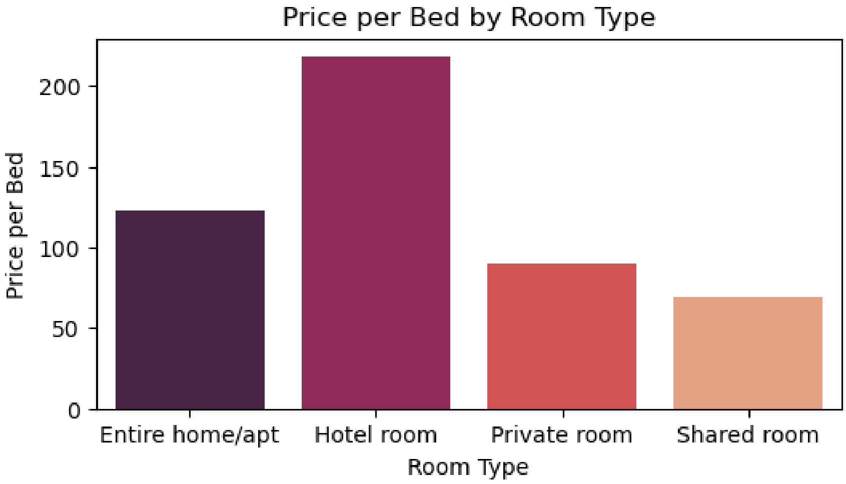
```
df.groupby(by='room_type')['price per bed'].mean()
```

Out[60]:

```
room_type
Entire home/apt    123.272485
Hotel room         218.330275
Private room        90.149760
Shared room         69.019928
Name: price per bed, dtype: float64
```

In [71]:

```
room=df.groupby(by='room_type')['price per bed'].mean()
plt.figure(figsize=(6,3))
plt.xlabel('Room Type')
plt.ylabel('Price per Bed')
plt.title('Price per Bed by Room Type')
sns.barplot(x='room_type',y='price per bed',data=room.reset_index(),hue='room_type',palette='rocket' )
plt.show()
```

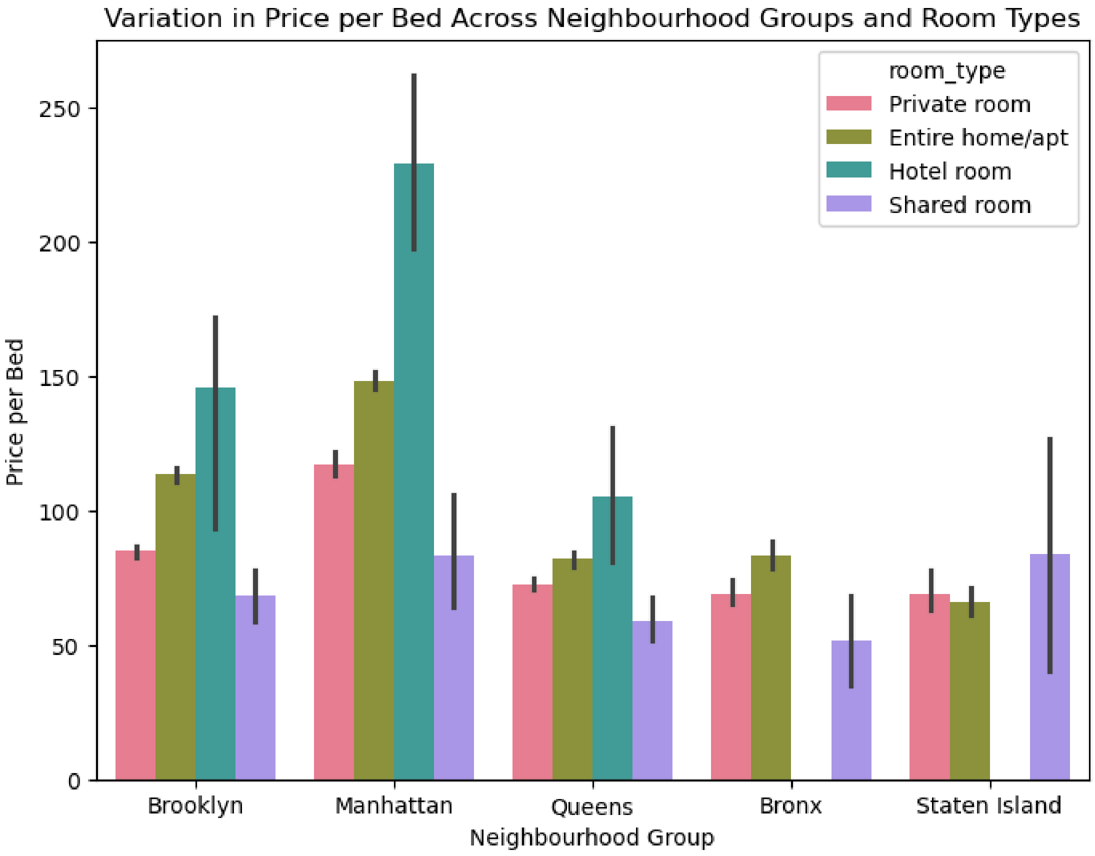


Hotel rooms command the highest price per bed, exceeding 200 units, making them the most expensive accommodation type on Airbnb. Conversely, shared rooms offer the lowest price per bed, at just over 50 units, representing the most budget-friendly option.

Variation in Price per Bed Across Neighbourhood Groups and Room Types

In [125…

```
plt.figure(figsize=(8,6))
plt.xlabel('Neighbourhood Group')
plt.ylabel('Price per Bed')
plt.title('Variation in Price per Bed Across Neighbourhood Groups and Room Types')
sns.barplot(x='neighbourhood_group',y='price per bed',data=df,hue='room_type',palette='husl' )
plt.show()
```

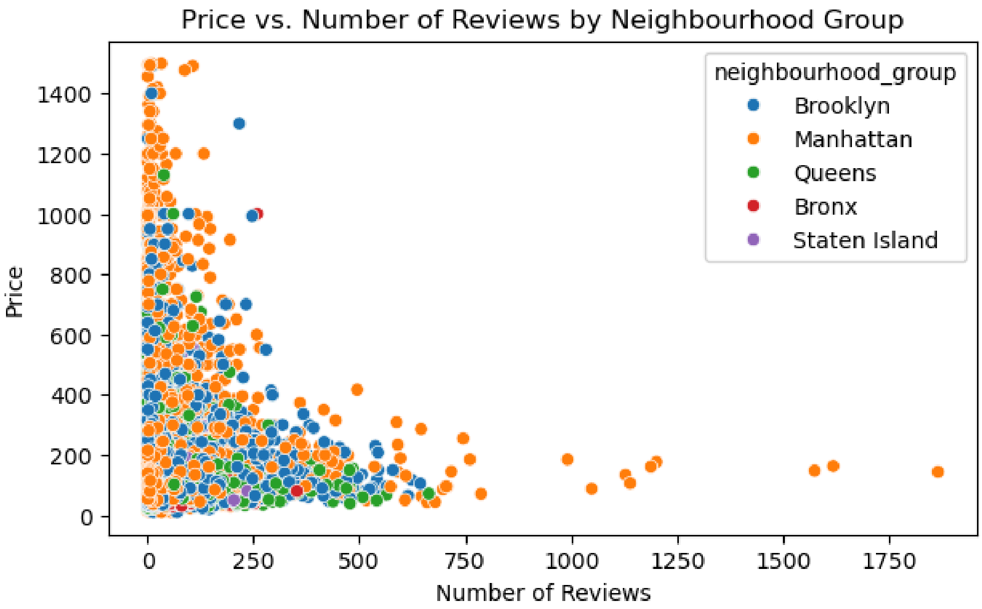


Neighbourhood group and room type significantly impact the price per bed on Airbnb ***Manhattan is generally more expensive than other neighbourhoods for all room types*** ***Shared rooms offer the most affordable option, while hotel rooms are the most expensive***

Impact of Number of Reviews on Price Across Different Neighbourhood Groups

In [142…

```
plt.figure(figsize=(7,4))
plt.xlabel('Number of Reviews')
plt.ylabel('Price')
plt.title('Price vs. Number of Reviews by Neighbourhood Group')
sns.scatterplot(data=df,x='number_of_reviews',y='price',hue='neighbourhood_group')
plt.show()
```

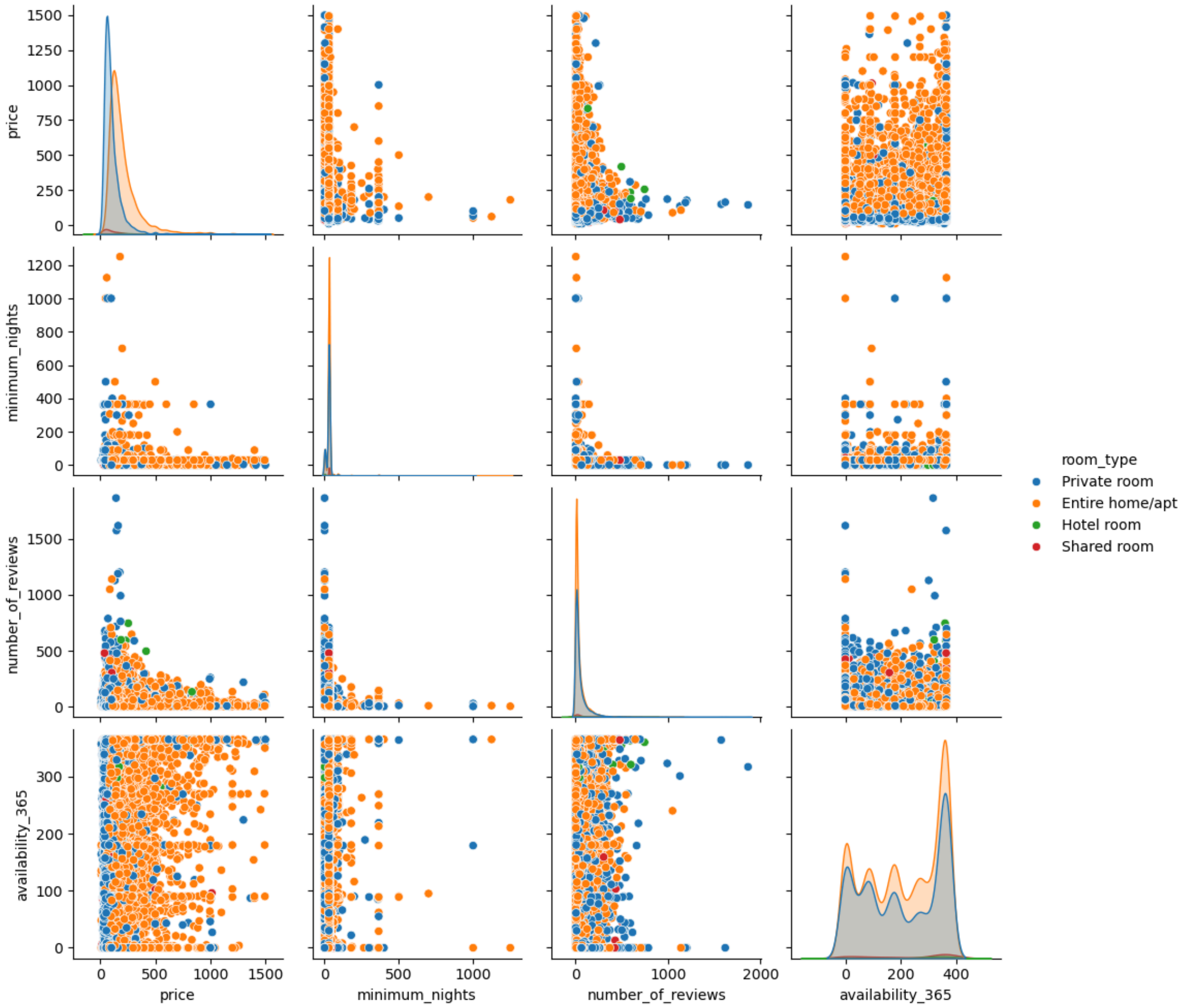


Most listings have prices concentrated in the lower range (0-400 units). Listings with higher prices (above 400 units) are less frequent and tend to have fewer reviews. *Across all neighbourhood groups, the number of reviews decreases as the price increases. This suggests that more affordable listings tend to receive more reviews, possibly due to higher occupancy rates.*

Exploring Relationships Between Listing Characteristics by Room Type

In [153…

```
sns.pairplot(data=df,vars=['price','minimum_nights','number_of_reviews','availability_365'],hue='room_ty
plt.show()
```

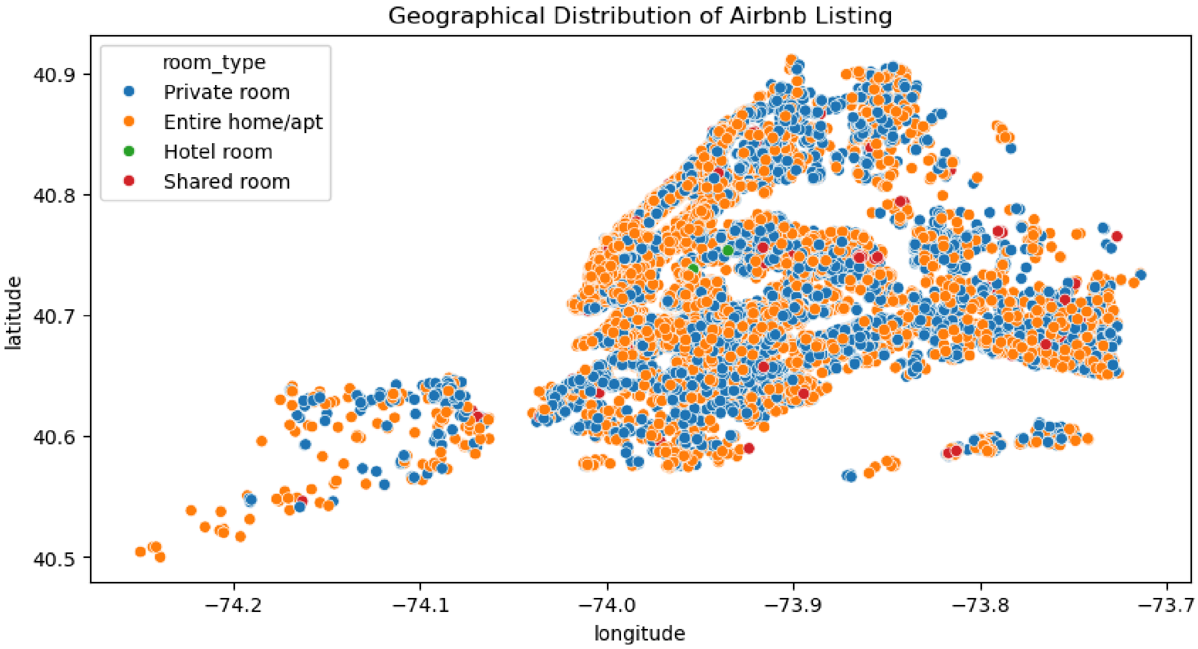


In [ ]:

Geographical Distribution of Airbnb Listing

In [169…

```
plt.figure(figsize=(10,5))
sns.scatterplot(data=df,x='longitude',y='latitude',hue='room_type')
plt.title('Geographical Distribution of Airbnb Listing')
plt.show()
```



Concentration of Listings:The majority of Airbnb listings are concentrated in specific areas, particularly in Manhattan and Brooklyn. These boroughs show a high density of listings, indicating their popularity among hosts and guests.

Room Type Distribution:Entire homes/apartments (orange dots) and private rooms (blue dots) are the most common types of listings. Shared rooms (red dots) and hotel rooms (green dots) are less frequent. Entire homes/apartments and private rooms are densely packed in certain regions, showing their prominence in the Airbnb market.

Correlation Matrix of Airbnb Listing Features

In [178…

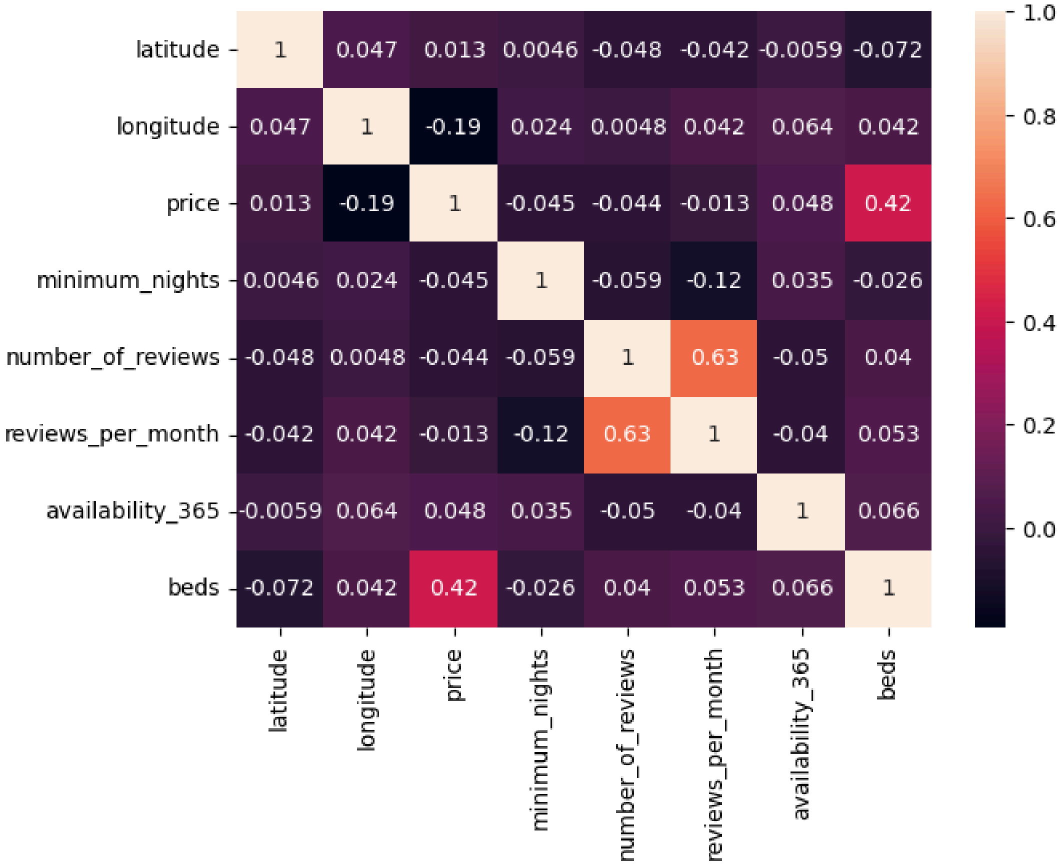
```
corr=df[['latitude','longitude','price','minimum_nights','number_of_reviews','reviews_per_month','availa
corr
```

Out[178…

	latitude	longitude	price	minimum_nights	number_of_reviews	reviews_per_month	avai
latitude	1.000000	0.047369	0.012686	0.004590	-0.047953	-0.041673	
longitude	0.047369	1.000000	-0.193728	0.023890	0.004820	0.041720	
price	0.012686	-0.193728	1.000000	-0.044635	-0.043533	-0.012775	
minimum_nights	0.004590	0.023890	-0.044635	1.000000	-0.059049	-0.122509	
number_of_reviews	-0.047953	0.004820	-0.043533	-0.059049	1.000000	0.631005	
reviews_per_month	-0.041673	0.041720	-0.012775	-0.122509	0.631005	1.000000	
availability_365	-0.005941	0.063523	0.048036	0.035466	-0.049656	-0.040116	
beds	-0.071753	0.041832	0.415278	-0.025852	0.040071	0.053496	

In [190…

```
plt.figure(figsize=(7,5))
sns.heatmap(data=corr,annot=True)
plt.show()
```



	latitude	longitude	price	minimum_nights	number_of_reviews	reviews_per_month	availability_365	beds
latitude	1	0.047	0.013	0.0046	-0.048	-0.042	-0.0059	-0.072
longitude	0.047	1	-0.19	0.024	0.0048	0.042	0.064	0.042
price	0.013	-0.19	1	-0.045	-0.044	-0.013	0.048	0.42
minimum_nights	0.0046	0.024	-0.045	1	-0.059	-0.12	0.035	-0.026
number_of_reviews	-0.048	0.0048	-0.044	-0.059	1	0.63	-0.05	0.04
reviews_per_month	-0.042	0.042	-0.013	-0.12	0.63	1	-0.04	0.053
availability_365	-0.0059	0.064	0.048	0.035	-0.05	-0.04	1	0.066
beds	-0.072	0.042	0.42	-0.026	0.04	0.053	0.066	1

Price and Beds: There's a moderate positive correlation (0.42) between price and the number of beds. This indicates that listings with more beds tend to have higher prices

There's a moderate negative correlation (-0.044) between price and number of reviews. This suggests that as price increases, the number of reviews tends to decrease.

```
In [ ]:
```