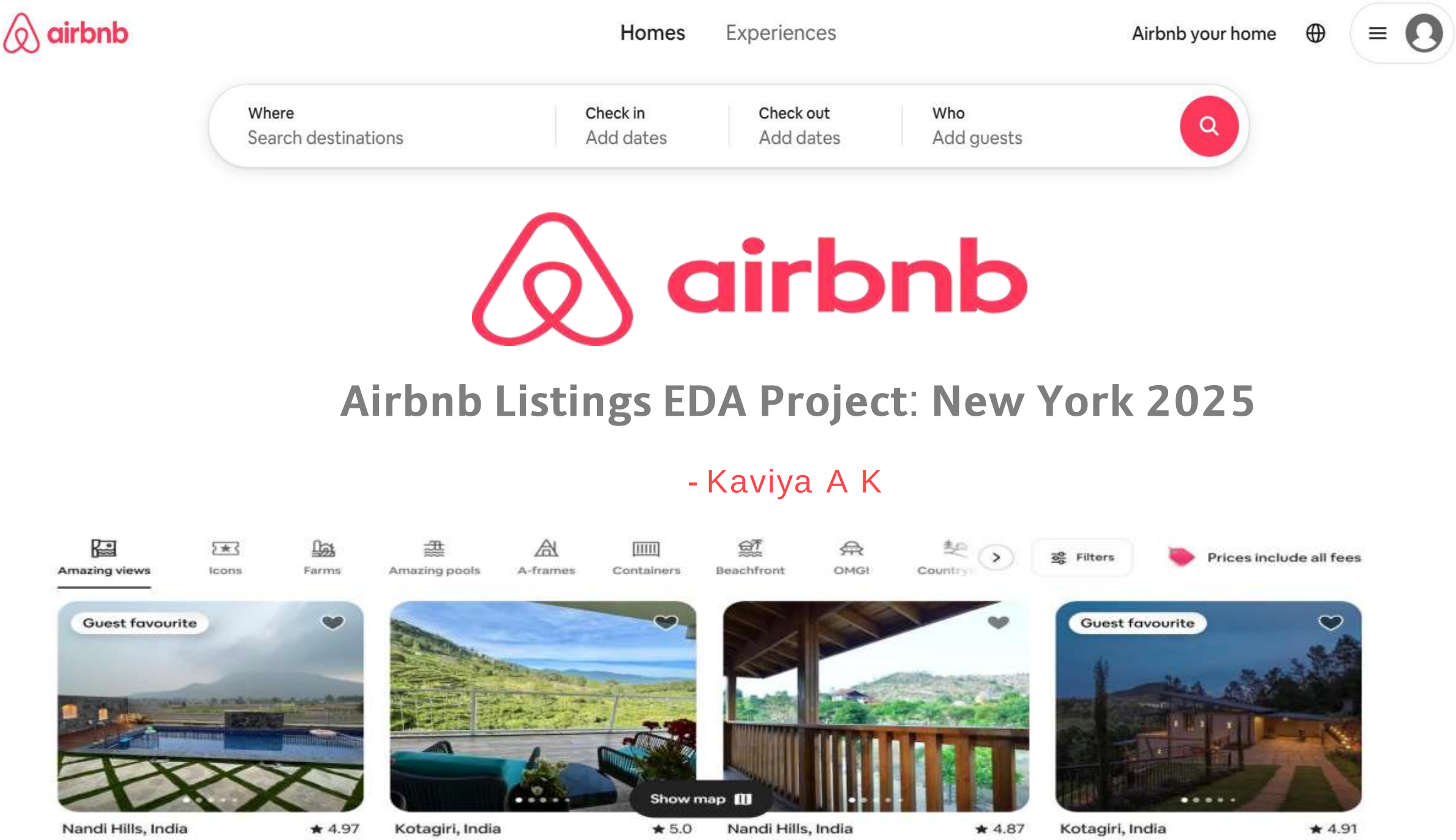
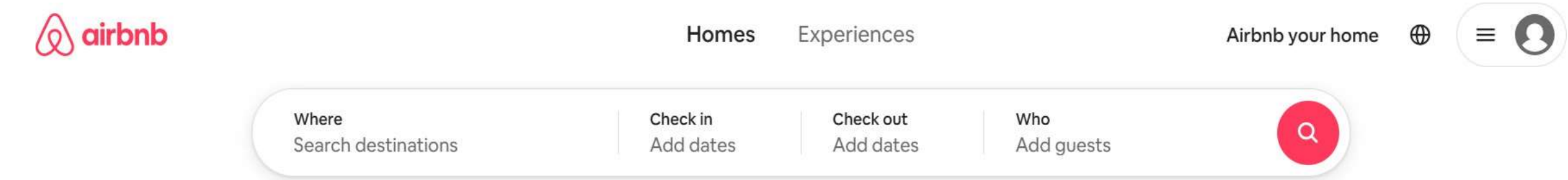
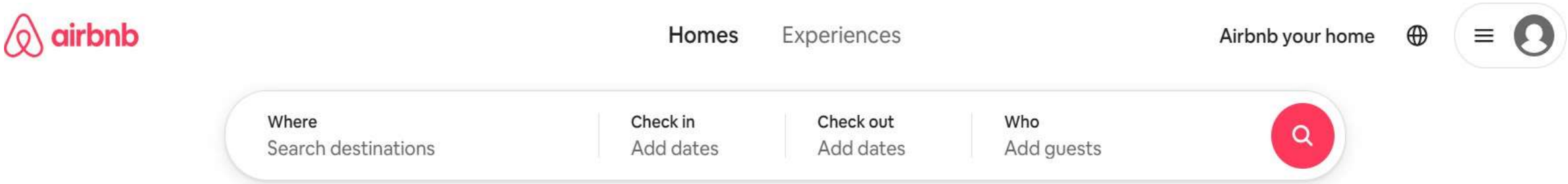

airbnb
Homes
Experiences
Airbnb your home
Where
Search destinations
Check in
Add dates
Check out
Add dates
Who
Add guests
airbnb
Airbnb Listings EDA Project: New York 2025
- Kaviya A K
Amazing views
Icons
Farms
Amazing pools
A-frames
Containers
Beachfront
OMG!
Filters
Prices include all fees
Guest favourite
Guest favourite
Show map
Nandi Hills, India
★ 4.97
Kotagiri, India
★ 5.0
Nandi Hills, India
★ 4.87
Kotagiri, India
★ 4.91



Project Overview

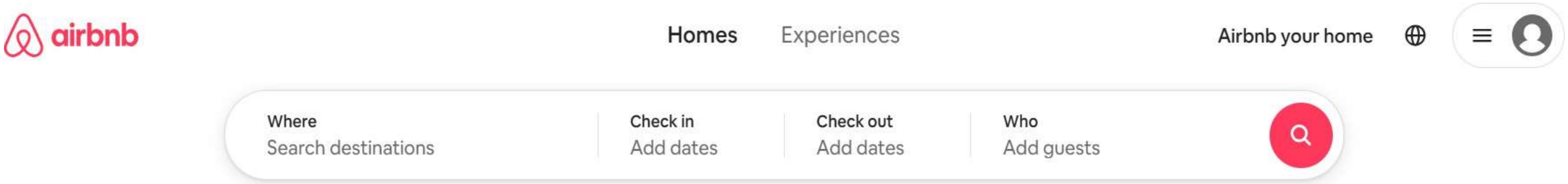
This project performs Exploratory Data Analysis (EDA) on New York Airbnb data to uncover trends and patterns in rental listings. Using libraries like Pandas, Numpy, Matplotlib, Seaborn for cleaning, visualization and analysis.



Objective

The goal of this project is to:

- Analyze room types, prices, and availability across different neighborhoods.
- Understand host behavior and listing patterns.
- Detect potential outliers in prices.
- Provide recommendations for guests and hosts based on insights.



Dataset Overview

- Data source – Kaggle
- Number of rows and columns – 38821 X 19
- Key variables / Features - Latitude, Longitude, Price, Minimum nights, Number of reviews, Reviews per month, Availability 365, Beds, Neighbourhood group

airbnb

Import libraries

```
[51]
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Loading dataset

```
2]
df = pd.read_csv(r'D:\Airbnb EDA\my_env2\Airbnb_dataset.csv')
df.head()
```

airbnb

Data exploration

```
df.head()
```

```
df.tail()
```

```
df.shape
```

```
df.info()
```

```
df.describe()
```

Data cleaning

```
df.isnull().sum()

#dropping the missing value rows
df.dropna(inplace=True)

df.isnull().sum()
```

```
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
bedrooms                          0
beds                              0
baths                             0
dtype: int64
```

```
# Handling duplicate datas
df.duplicated().sum()

#dropping duplicate records
df.drop_duplicates(inplace=True)
df.duplicated().sum()
```

```
np.int64(0)
```

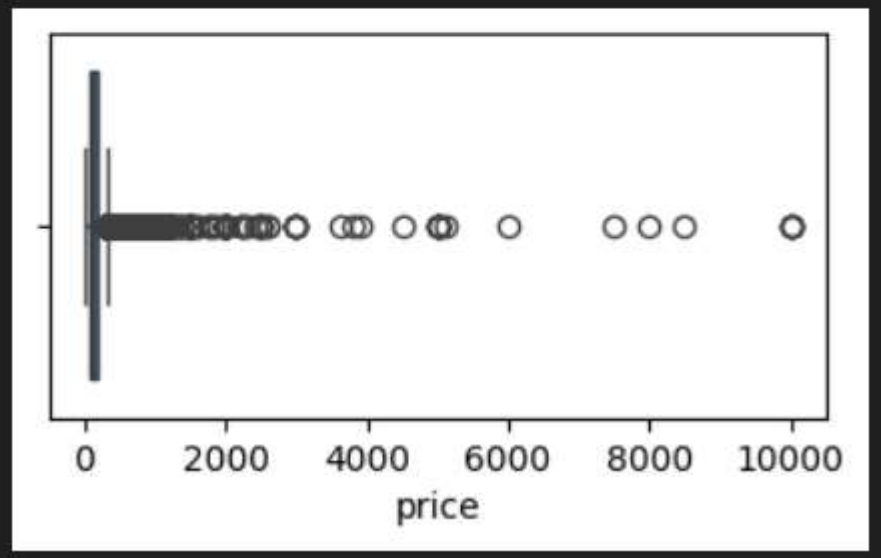
```
# Type casting
df.dtypes

df['id']=df['id'].astype(object)
df.dtypes
df['host_id']=df['host_id'].astype(object)
df.dtypes
```

EDA – Univariate Analysis

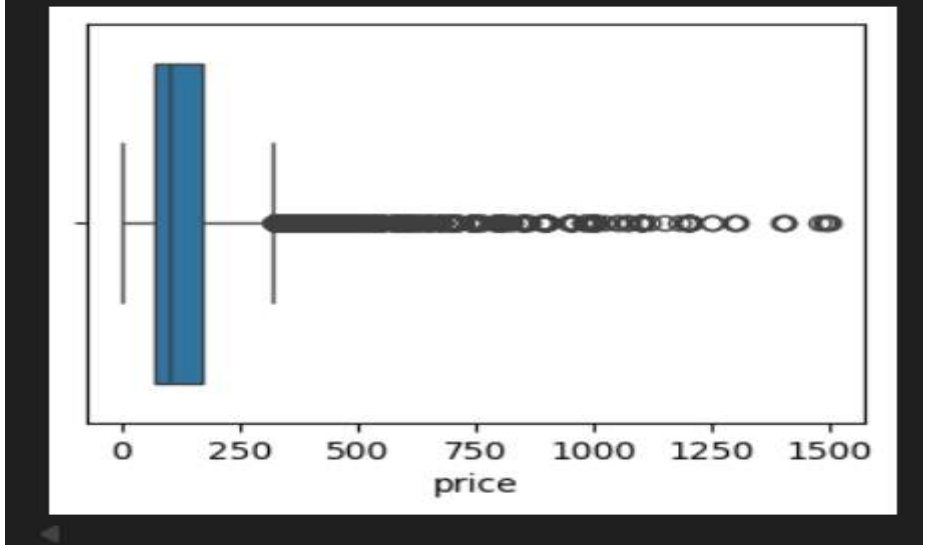
Checking Outliers for price column

```
plt.figure(figsize=(4,2))
sns.boxplot(data=df,x='price')
plt.show()
```




```
data=df[df['price']<1500]
```

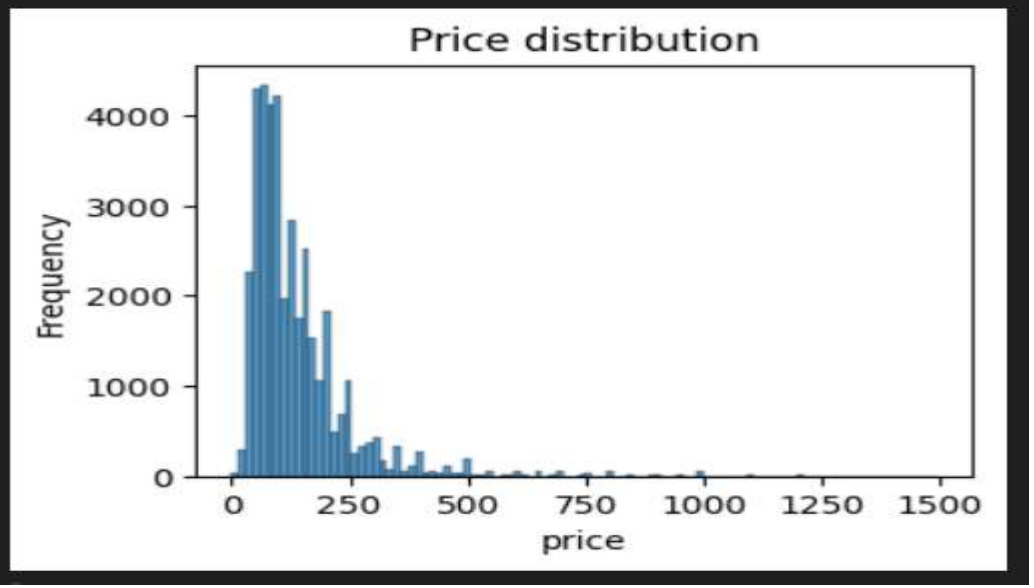
```
plt.figure(figsize=(4,3))
sns.boxplot(data=data,x='price')
plt.show()
```



EDA – Univariate Analysis

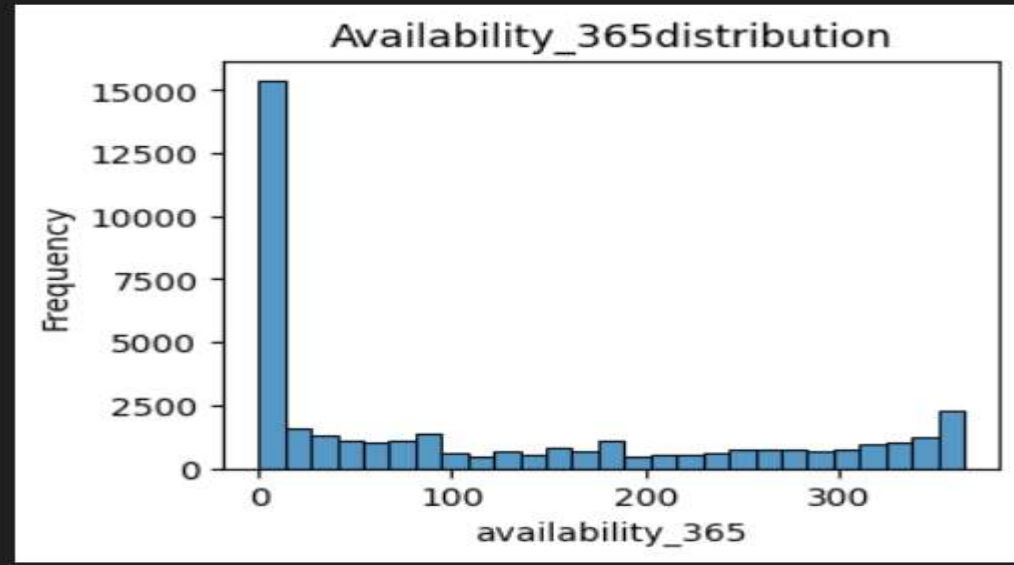
Data distribution of price column

```
# Data distribution of each columns
plt.figure(figsize=(4,3))
sns.histplot(data=data,x='price',bins=100)
plt.title('Price distribution')
plt.ylabel('Frequency')
plt.show()
```



Data distribution of Availability 365 column

```
#Data distribution of availability_365
plt.figure(figsize=(4,3))
sns.histplot(data=df,x='availability_365')
plt.title('Availability_365distribution')
plt.ylabel('Frequency')
plt.show()
```



EDA – Univariate Analysis

Data summary of neighbourhood group with price

```
data.groupby('neighbourhood_group')['price'].mean()
```

```
neighbourhood_group
Bronx             79.558857
Brooklyn         118.074090
Manhattan        171.894394
Queens            92.694900
Staten Island     89.964968
Name: price, dtype: float64
```

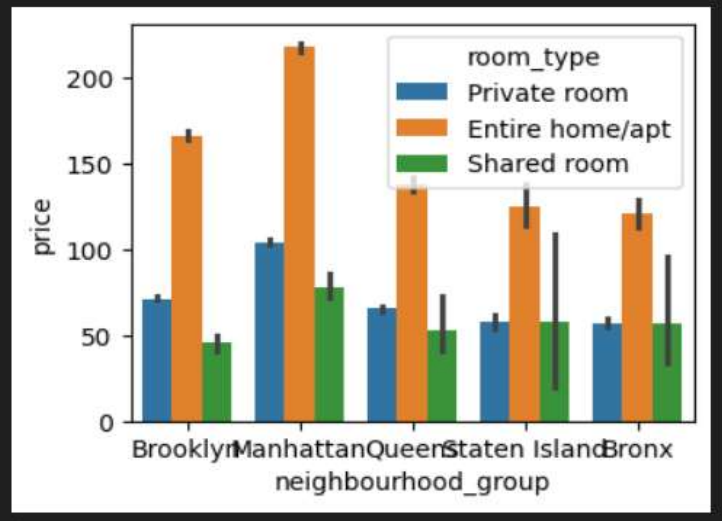
Feature engineering

```
# Adding "Price Per bed" column
data['price per bed']= data['price']/data['beds']
df.head()
```

EDA – Bivariate Analysis

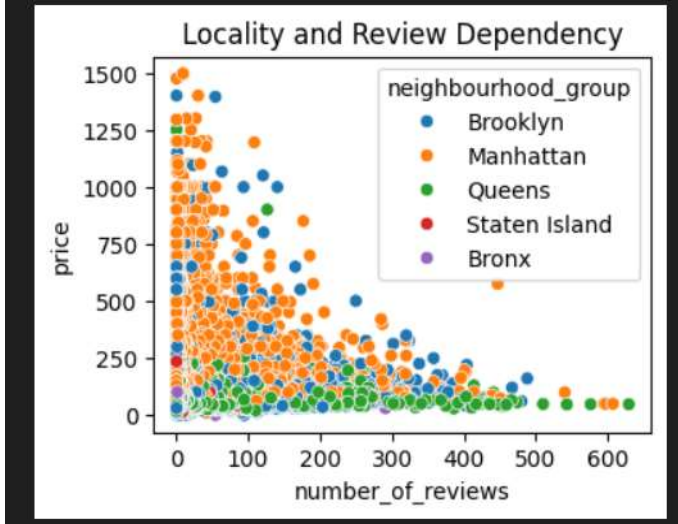
Price dependency on room type

```
#Price dependency on room type
plt.figure(figsize=(4,3))
sns.barplot(data=data,x='neighbourhood_group',y='price',hue='room_type')
plt.show()
```



Price relationship with number of reviews

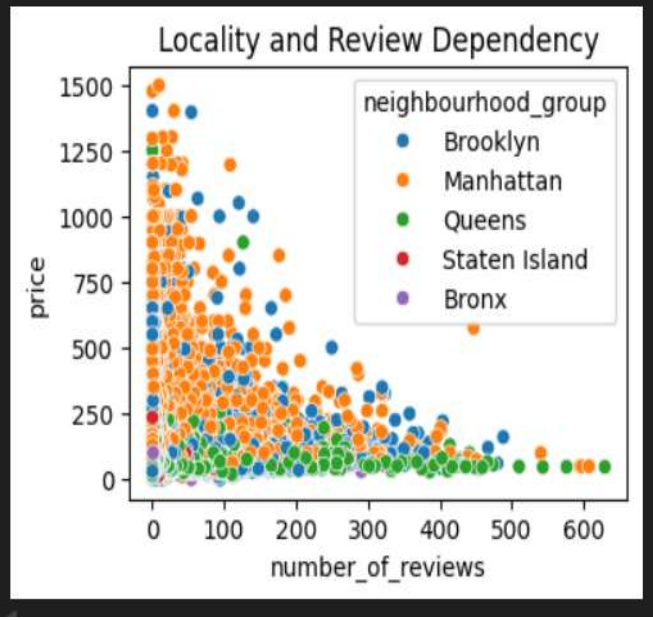
```
#Price relationship with no of Reviews
plt.figure(figsize=(4,3))
sns.scatterplot(data=data,x='number_of_reviews',y='price',hue='neighbourhood_group')
plt.title('Locality and Review Dependency')
plt.show()
```



EDA – Bivariate Analysis

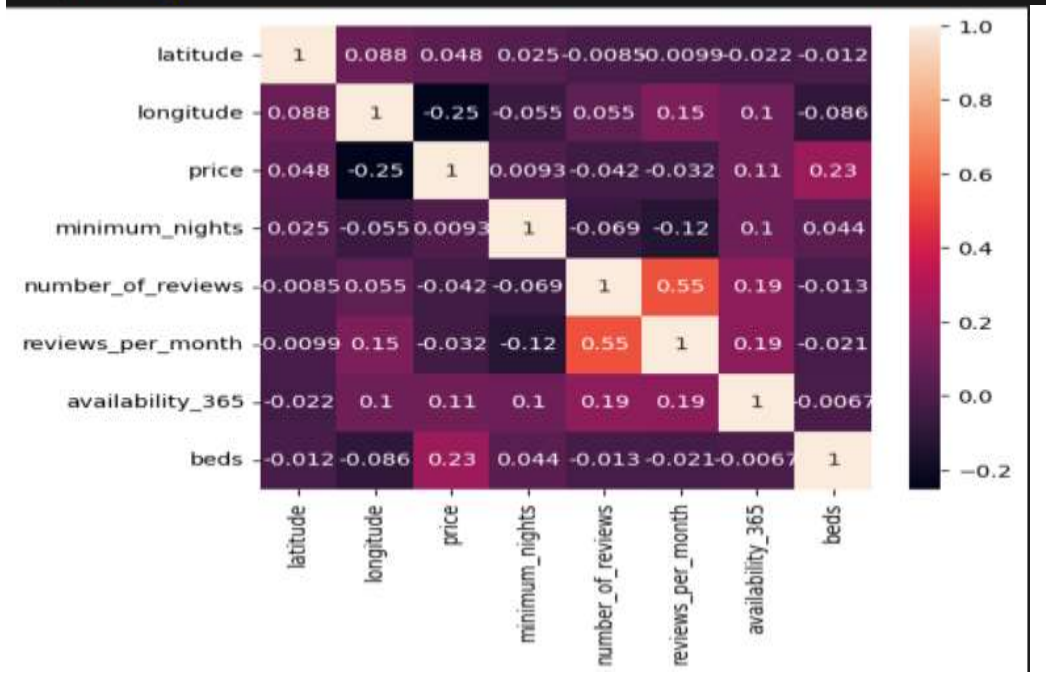
Geographical distribution of Airbnb listing

```
#Price relationship with no of Reviews
plt.figure(figsize=(4,3))
sns.scatterplot(data=data,x='number_of_reviews',y='price',hue='neighbourhood_group')
plt.title('Locality and Review Dependency')
plt.show()
```



Correlation between numerical columns

```
#heatmap : To see correlation between numerical columns
plt.figure(figsize=(6,5))
corr=data[['latitude','longitude','price','minimum_nights','number_of_reviews',
           'reviews_per_month','availability_365','beds']].corr()
corr
sns.heatmap(data=corr,annot=True)
plt.show()
```



Key Insights

1. **Price Trends**:
    - **Manhattan** has the most expensive listings, followed by Brooklyn.
    - **Entire homes/apartments** cost significantly more than private or shared rooms.

2. **Room Type Distribution**:
    - **Entire homes/apartments** are the most common, but **private rooms** offer budget-friendly options.

3. **Outliers in Price**:
    - Few listings priced at **$10,000+** were detected, indicating the need to filter such extreme values.

4. **Availability Patterns**:
    - Listings with **high availability** tend to have lower prices and more reviews, likely due to better guest experience.

5. **Host Behavior**:
    - Some hosts manage **multiple listings**, indicating a trend toward professional hosting.

Recommendations

- **For Guests**:
  - Look for listings with high availability and good reviews for a better experience.
  - **Private rooms** in Brooklyn offer affordable stays compared to Manhattan.
- **For Hosts**:
  - Improve **availability** and **review response rates** to attract more bookings.
  - Manage pricing effectively to compete within the borough's market.

Conclusion

This project offers valuable insights into the New York Airbnb market, helping both guests and hosts make informed decisions. By using **EDA techniques**, I have identified key trends and developed actionable recommendations. Future improvements can involve advanced analytics and predictive modeling to further enhance the findings.

Thank you