Explository Data Analysis on New York Airbnb listing

Importing relevant Module

```
In [2]: import pandas as pd
        import numpy as np
        import os
        import seaborn as sn
        from matplotlib import pyplot as plt
```

```
In [3]: ny_airbnb=pd.read_csv("C:/Users/Sanayak/Desktop/AB_NYC_2019.csv")
```

```
In [4]: ny_airbnb.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48895 entries, 0 to 48894
Data columns (total 16 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   id                              48895 non-null  int64
 1   name                            48879 non-null  object
 2   host_id                         48895 non-null  int64
 3   host_name                       48874 non-null  object
 4   neighbourhood_group             48895 non-null  object
 5   neighbourhood                   48895 non-null  object
 6   latitude                        48895 non-null  float64
 7   longitude                       48895 non-null  float64
 8   room_type                       48895 non-null  object
 9   price                           48895 non-null  int64
 10  minimum_nights                  48895 non-null  int64
 11  number_of_reviews               48895 non-null  int64
 12  last_review                     38843 non-null  object
 13  reviews_per_month               38843 non-null  float64
 14  calculated_host_listings_count  48895 non-null  int64
 15  availability_365                48895 non-null  int64
dtypes: float64(3), int64(7), object(6)
memory usage: 6.0+ MB
```

```
In [5]: ny_airbnb.describe()
```

Out[5]:

	id	host_id	latitude	longitude	price	minimum_ni
count	4.889500e+04	4.889500e+04	48895.000000	48895.000000	48895.000000	48895.00(
mean	1.901714e+07	6.762001e+07	40.728949	-73.952170	152.720687	7.02!
std	1.098311e+07	7.861097e+07	0.054530	0.046157	240.154170	20.51(
min	2.539000e+03	2.438000e+03	40.499790	-74.244420	0.000000	1.00(
25%	9.471945e+06	7.822033e+06	40.690100	-73.983070	69.000000	1.00(
50%	1.967728e+07	3.079382e+07	40.723070	-73.955680	106.000000	3.00(
75%	2.915218e+07	1.074344e+08	40.763115	-73.936275	175.000000	5.00(
max	3.648724e+07	2.743213e+08	40.913060	-73.712990	10000.000000	1250.00(

In [6]: `ny_airbnb.head()`

Out[6]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latit
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75
2	3647	THE VILLAGE OF HARLEM….NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79

In [7]: `ny_airbnb.nunique()`

```
Out[7]:  id                              48895
         name                            47905
         host_id                         37457
         host_name                       11452
         neighbourhood_group                 5
         neighbourhood                     221
         latitude                        19048
         longitude                       14718
         room_type                           3
         price                             674
         minimum_nights                    109
         number_of_reviews                 394
         last_review                      1764
         reviews_per_month                 937
         calculated_host_listings_count     47
         availability_365                  366
         dtype: int64
```

In [9]: `ny_airbnb.shape`

Out[9]: `(48895, 16)`

In [16]: `ny_airbnb.count()`

```
Out[16]:  id                              48895
          name                            48879
          host_name                       48874
          neighbourhood_group             48895
          neighbourhood                   48895
          latitude                        48895
          longitude                       48895
          room_type                       48895
          price                           48895
          minimum_nights                  48895
          number_of_reviews               48895
          last_review                     38843
          reviews_per_month               38843
          calculated_host_listings_count  48895
          availability_365                48895
          dtype: int64
```

In [18]: `ny_airbnb.drop(["id","latitude","longitude",], axis=1, inplace=True)`

In [49]: `ny_airbnb.head()`

Out[49]:

	name	host_name	neighbourhood_group	neighbourhood	room_type	price	mini
0	Clean & quiet apt home by the park	John	Brooklyn	Kensington	Private room	149	
1	Skylit Midtown Castle	Jennifer	Manhattan	Midtown	Entire home/apt	225	
3	Cozy Entire Floor of Brownstone	LisaRoxanne	Brooklyn	Clinton Hill	Entire home/apt	89	
4	Entire Apt: Spacious Studio/Loft by central park	Laura	Manhattan	East Harlem	Entire home/apt	80	
5	Large Cozy 1 BR Apartment In Midtown East	Chris	Manhattan	Murray Hill	Entire home/apt	200	

From the above New York Airbnb listing Data set, we have rows with null values and i think the best approach in cleansing the dataset is by droping the said rows.
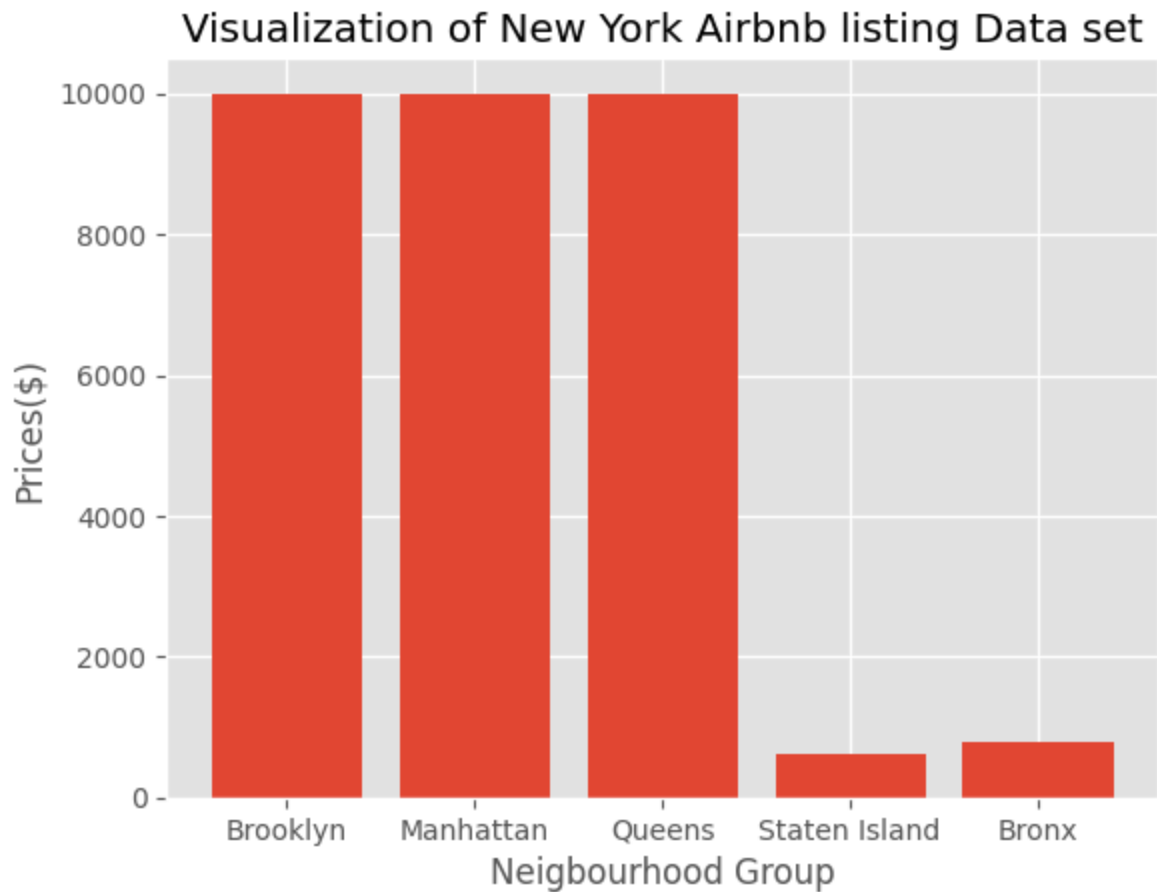
In [23]:
```python
ny_airbnb.dropna(axis=0,how="any",inplace=True)
```

In [26]:
```python
ny_airbnb.info()
```
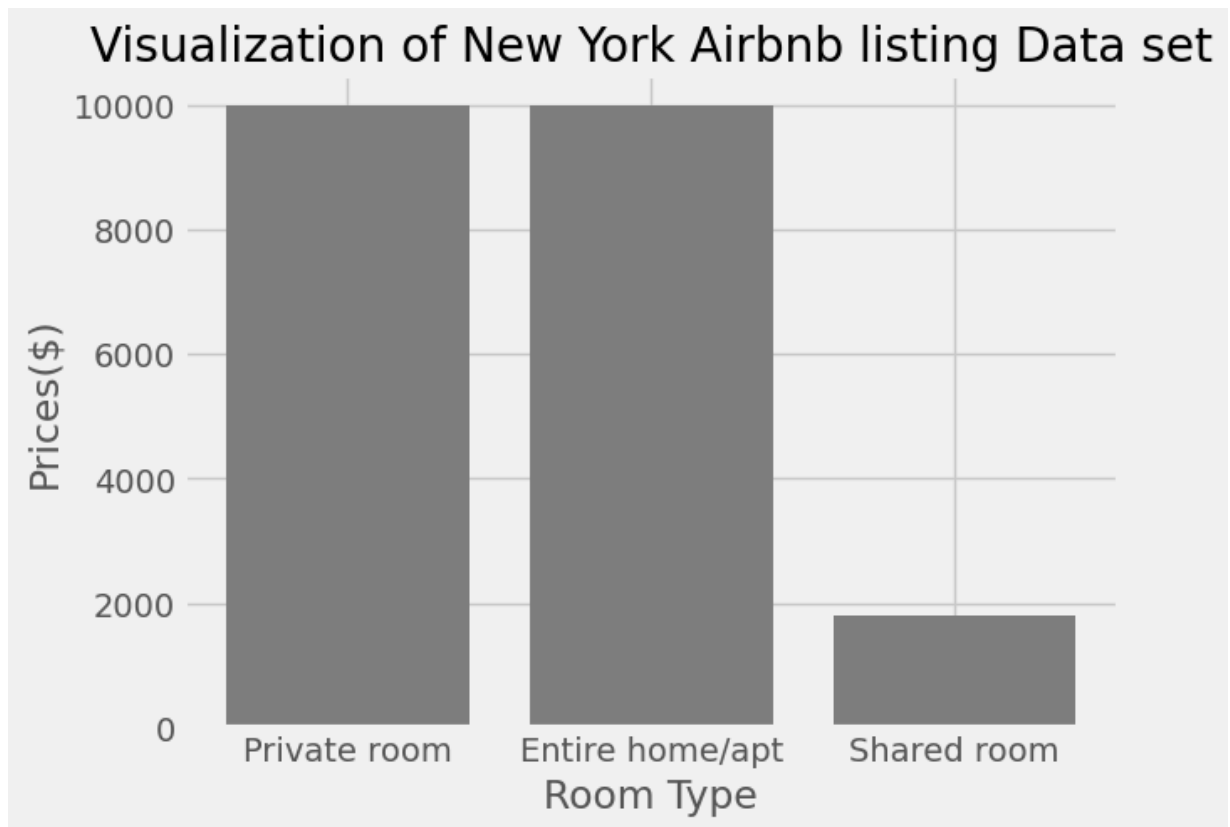```
<class 'pandas.core.frame.DataFrame'>
Index: 38821 entries, 0 to 48852
Data columns (total 12 columns):
 #   Column   Non-Null Count Dtype
--- ------   -------------- -----
 0   name   38821 non-null object
 1   host_name   38821 non-null object
 2   neighbourhood_group   38821 non-null object
 3   neighbourhood   38821 non-null object
 4   room_type   38821 non-null object
 5   price   38821 non-null int64
 6   minimum_nights   38821 non-null int64
 7   number_of_reviews   38821 non-null int64
 8   last_review   38821 non-null object
 9   reviews_per_month   38821 non-null float64
 10 calculated_host_listings_count 38821 non-null int64
 11 availability_365   38821 non-null int64
dtypes: float64(1), int64(5), object(6)
memory usage: 3.9+ MB
```
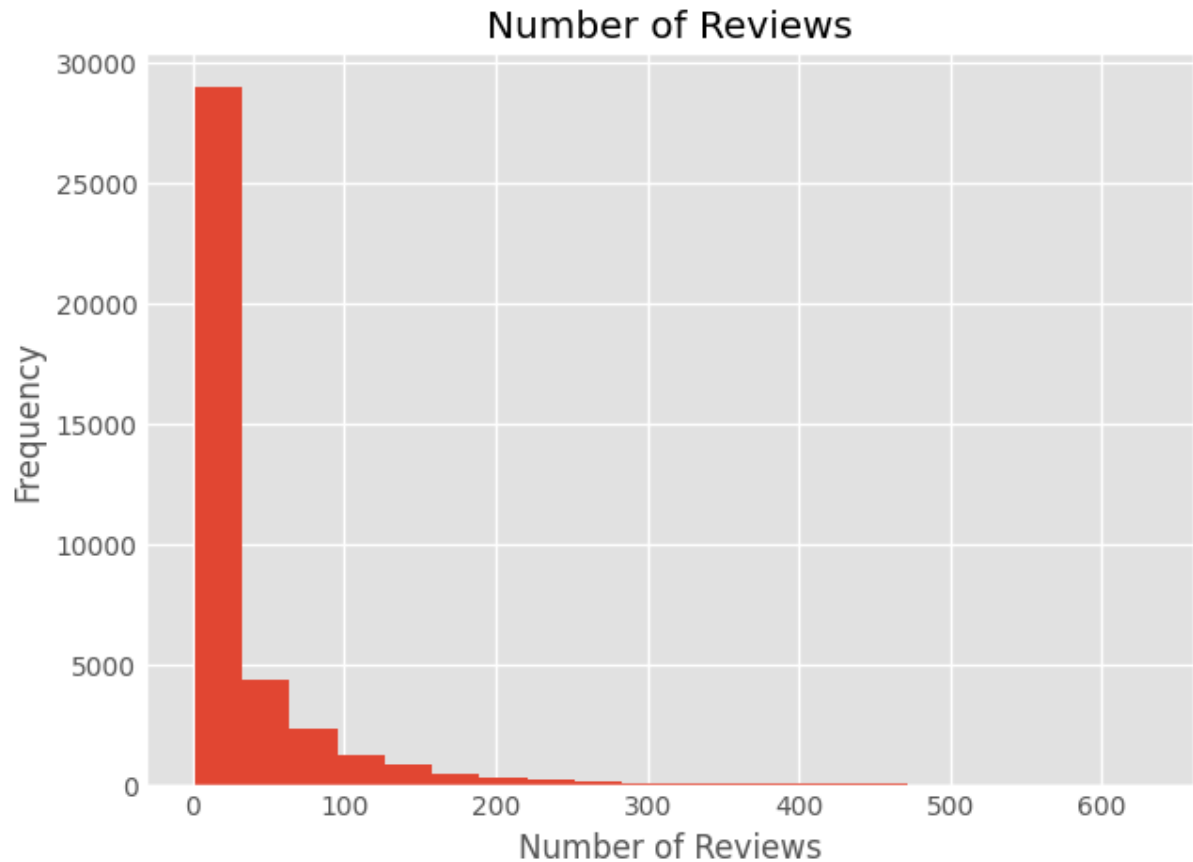
Visualization of New York Airbnb listing Dataset

In [32]:
```python
plt.bar(ny_airbnb.neighbourhood_group,ny_airbnb.price)
plt.xlabel("Neigbourhood Group")
plt.ylabel("Prices($)")
plt.title("Visualization of New York Airbnb listing Data set")
plt.style.use("ggplot")
plt.show()
```


Visualization of New York Airbnb listing Data set

In [34]:
```python
plt.bar(ny_airbnb.room_type,ny_airbnb.price,color="grey")
plt.xlabel("Room Type")
plt.ylabel("Prices($)")
plt.title("Visualization of New York Airbnb listing Data set")
plt.style.use("fivethirtyeight")
plt.show()
```

```
In [42]: plt.hist(ny_airbnb.number_of_reviews,bins=20)
         plt.xlabel("Number of Reviews")
         plt.ylabel("Frequency")
         plt.title("Number of Reviews")
         plt.style.use("ggplot")
         plt.show()
```

Number of Reviews

Correlation between Price and other variables

```
In [50]: print("Correlation Between price,reviews_per_month and number_of_reviews,minimum_ni
         print(ny_airbnb[["price","reviews_per_month","number_of_reviews","minimum_nights"]]
```

```
Correlation Between price,reviews_per_month and number_of_reviews,minimum_nights
                     price  reviews_per_month  number_of_reviews  \
price              1.000000          -0.030623          -0.035924
reviews_per_month -0.030623           1.000000           0.549699
number_of_reviews -0.035924           0.549699           1.000000
minimum_nights     0.025501          -0.121712          -0.069366

                   minimum_nights
price                    0.025501
reviews_per_month       -0.121712
number_of_reviews       -0.069366
minimum_nights           1.000000
```

Prepared by Victor Iniobong Ititnah Data Scientist and Analyst | Economist | Mathematics Tutor

```
In [ ]:
```