IMPORTING LIBRARIES

In [ ]:
```
import pandas as py
import numpy as np
df=py.read_csv(r"C:\Users\evang\Downloads\AB_NYC_2019.csv")
df
```

In [30]:
```
df.shape
```

Out[30]:
```
(48895, 14)
```

CHECKING IF THERE IS NULL VALUES

In [4]:
```
df.isnull().sum()
```

Out[4]:
```
id                                    0
name                                 16
host_id                               0
host_name                            21
neighbourhood_group                   0
neighbourhood                         0
latitude                              0
longitude                             0
room_type                             0
price                                 0
minimum_nights                        0
number_of_reviews                     0
last_review                       10052
reviews_per_month                 10052
calculated_host_listings_count        0
availability_365                      0
dtype: int64
```

GETTING INFO ABOUT DATAFRAME

In [8]:
```
df.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 [11]:
```
df1.shape
```

Out[11]:
```
(48895, 16)
```

DROPPING COLUMNS(last_review),(reviews_per_month) WHICH HAS MORE THAN 50% OF NULL VALUES

In [6]:
```
df.drop(['last_review'],axis=1,inplace=True)
```

In [7]:
```
df.drop(['reviews_per_month'],axis=1,inplace=True)
```

TO CHECK WHETHER ANY NULL VALUES ARE PRESENT

In [8]:
```
df.isnull().sum()
```

Out[8]:
```
id                                0
name                             16
host_id                           0
host_name                        21
neighbourhood_group               0
neighbourhood                     0
latitude                          0
longitude                         0
room_type                         0
price                             0
minimum_nights                    0
number_of_reviews                 0
calculated_host_listings_count    0
availability_365                  0
dtype: int64
```

REPLACING NULL VALUES IN NAME USING NULL VALUES

In [13]:
```
df['name']=df['name'].bfill(axis=0)
```

In [ ]:
```
TO CHECK WHETHER NULL VALUES ARE THERE OR NOT
```

In [15]:
```
df['name'].isnull().sum()
```

Out[15]:
```
0
```

In [26]:
```
df['host_name']=df['host_name'].bfill(axis=0)
```

In [27]:
```
df['host_name'].isnull().sum()
```

Out[27]:
```
0
```

In [ ]:
```
CHECKING WHETHER NULL PRESENT IN ANY OF THE COLUMNS
```

In [28]:
```
df.isnull().sum()
```

Out[28]:
```
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
calculated_host_listings_count    0
availability_365                  0
dtype: int64
```