Task 6.1

Summary & Sources:
Airbnb is a US company that allows people seeking a stay and people who have a property to rent out to connect. Airbnb is designed for short-stay accommodations, ultimately replacing staying at a hotel. This dataset provides Airbnb's Market in New York City with a variety of metrics for over 40,000 properties.

Data Collection:
The data was downloaded from insideairbnb.com under the New York City section. The insideairbnb.com site empowers communities to understand the impact of Airbnb through data and information.

Data Content:
The data set contains information on over 40,000 in New York City with their hosts
 Information, geographical data, reviews, and property metrics.

Data Limitation & Ethics:
The data set is missing booking information for each Airbnb. The data may also include some human errors. There is some concern with the name is the host included in the data creating some PII issues.

Data Profile:

Column Name	Description	Qualitative / Quantitative	Nominal / Ordinal/Binary / Discrete/ Continuous
id	Unique ID number for each property	Qualitative	Nominal
name	Name of the Airbnb online	Qualitative	Nominal
host_id	Unique ID number for each host	Qualitative	Nominal
host_name	Name of the Airbnb host	Qualitative	Nominal
neighbourhood_group	Neighborhood Group	Qualitative	Nominal
neighbourhood	Neighborhood Name	Qualitative	Nominal
latitude	Coordinates of County	Quantitative	Continuous
longitude	Coordinates of County	Quantitative	Continuous
room_type	Type or room	Qualitative	Nominal
price	Nightly Rate to Book the Property	Quantitative	Discrete
minimum_nights	Minimum Nights required to book	Quantitative	Discrete
number_of_reviews	Number of Reviews	Quantitative	Discrete
last_review	Date of last Review	Quantitative	Discrete
reviews_per_month	Reviews per month	Quantitative	Discrete
calculated_host_listings_count	The number of listings the host has in New York City	Quantitative	Discrete
availability_365	Availability for the next 365 days (listing may not be available because it has been booked by a guest or blocked by the host)	Quantitative	Discrete
number_of_reviews_ltm	The number of reviews the listing has in the last 12 months	Quantitative	Discrete
license	Property's license	Qualitative	Nominal

Understand Data:

```
clean_nyc_airbnb.shape
```

(39160, 15)

	id	host_id	latitude	longitude	price	min_night_stay	no_reviews	no_monthly_reviews	host_listings_count	
count	3.916000e+04	3.916000e+04	39160.000000	39160.000000	39160.000000	39160.000000	39160.000000	28211.000000	39160.000000	3
mean	2.932430e+17	1.575643e+08	40.729372	-73.945897	205.995480	30.272165	25.584397	1.069862	39.493105	
std	3.943340e+17	1.693066e+08	0.056645	0.054958	466.662167	26.534306	57.488642	1.730834	117.040976	
min	2.595000e+03	1.678000e+03	40.500314	-74.251907	0.000000	1.000000	0.000000	0.010000	1.000000	
25%	1.991531e+07	1.654341e+07	40.688880	-73.982819	76.000000	30.000000	0.000000	0.110000	1.000000	
50%	4.574172e+07	7.831966e+07	40.726470	-73.953960	130.000000	30.000000	4.000000	0.410000	1.000000	
75%	7.354150e+17	2.783108e+08	40.762632	-73.926260	220.000000	30.000000	23.000000	1.505000	6.000000	
max	1.015028e+18	5.442621e+08	40.911380	-73.713650	30000.000000	1250.000000	1854.000000	73.720000	647.000000	

	365_avail	no_of_reviews_ltm
count	39160.000000	39160.000000
mean	153.915628	6.735674
std	144.305706	17.023977
min	0.000000	0.000000
25%	0.000000	0.000000
50%	130.000000	0.000000
75%	309.000000	5.000000
max	365.000000	900.000000

```
clean_nyc_airbnb['nbhd_grp'].value_counts()
```

```
nbhd_grp
Manhattan        17084
Brooklyn         14302
Queens            6036
Bronx             1366
Staten Island      372
Name: count, dtype: int64
```

```
clean_nyc_airbnb['nbhd'].value_counts()
```

```
nbhd
Bedford-Stuyvesant       2723
Williamsburg             2296
Midtown                  2077
Harlem                   1841
Bushwick                 1641
                         ...
Chelsea, Staten Island      1
Fort Wadsworth              1
New Dorp                    1
Neponsit                    1
Lighthouse Hill             1
Name: count, Length: 223, dtype: int64
```

```
clean_nyc_airbnb['room_type'].value_counts()
```

```
room_type
Entire home/apt    21236
Private room       17256
Shared room          537
Hotel room           131
Name: count, dtype: int64
```

Key Questions:

Which neighborhoods have the most Airbnb properties?
Is there a price difference in different neighborhoods?
Based on reviews, which properties were booked the most?
Are there any trends across New York?