



DATA METHODOLOGY OF AIRBNB NYC

BY: DEEPAK,URVASHI,NRK PAVAN

STEP 1: STORY BOARDING

- ❖ Verified the data in AB_NYC_2019 and made a note of the key fields.
- ❖ created a mind map of the presentation's key slides.

STEP 2: DATA WRANGLING

- ❖ Determined the unique values and assessed the category and numerical variables using the Python Jupyter Notebook.
- ❖ A check was made for duplicate, missing, and null values. and accordingly changed the Nan values.
- ❖ Based on the computation below, it was discovered that Price is significantly skewed and has many outliers.

```
# Skewness, Kurtosis
print("Skewness: %f" % df_airbnb['price'].skew())
print("Kurtosis: %f" % df_airbnb['price'].kurt())

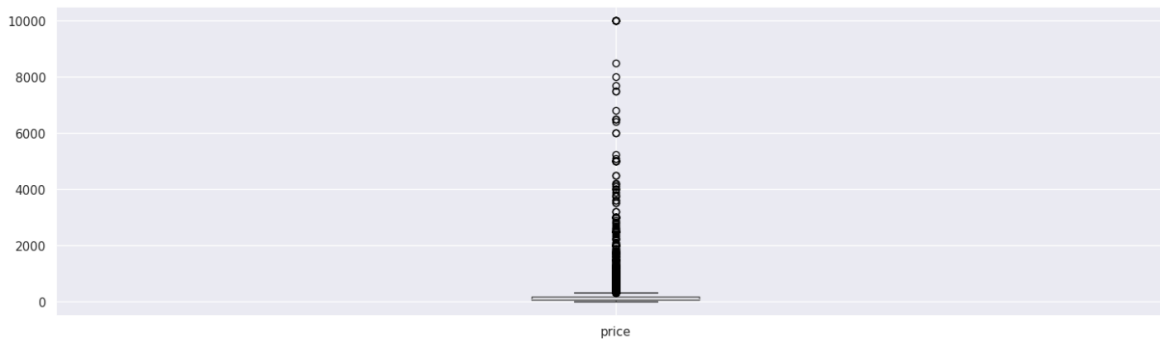
Skewness: 19.118939
Kurtosis: 585.672879
```

Observation:

- Looks the skewness and kurtosis came out very large. Since the skewness has value greater than 1 it is highly skewed.
- Also kurtosis looks high as well which indicates presence of good amount of outliers.

and box plot of Price with many outliers.

```
df_airbnb.boxplot(column=['price'])
plt.show()
```



- ❖ Using the IQR technique, the outlier treatment was carried out as follows.

```
# Removing outliers using IQR approach
```

```
Q1 = df_AB.quantile(0.25)
```

```
Q3 = df_AB.quantile(0.75)
```

```
IQR = Q3 - Q1
```

```
print(IQR)
```

```
id                1.968023e+07
host_id           9.961239e+07
latitude          7.301500e-02
longitude         4.679500e-02
price             1.060000e+02
minimum_nights    4.000000e+00
number_of_reviews 2.300000e+01
reviews_per_month 1.000000e+00
calculated_host_listings_count 1.000000e+00
availability_365  2.270000e+02
dtype: float64
```

```
# Function for removing outliers!
```

```
def outlier_treatment(datacolumn):
```

```
    sorted(datacolumn)
```

```
    Q1,Q3 = np.percentile(datacolumn , [25,75])
```

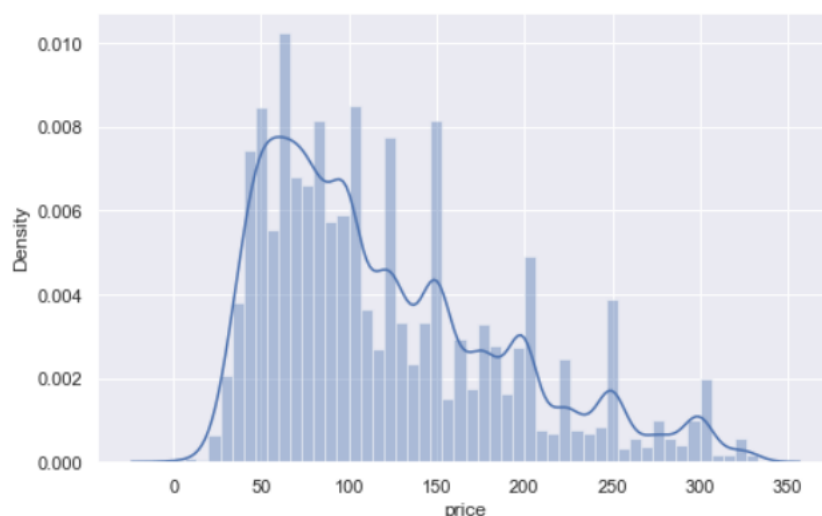
```
    IQR = Q3 - Q1
```

```
    lower_range = Q1 - (1.5 * IQR)
```

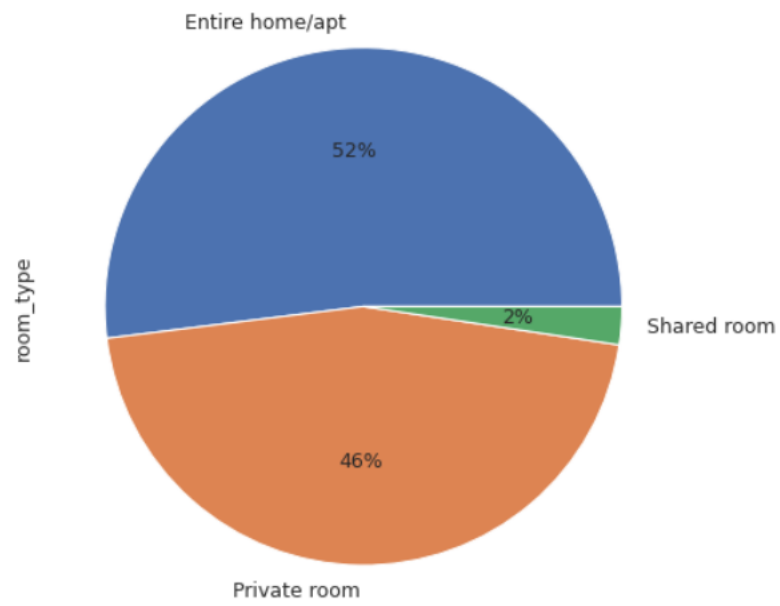
```
    upper_range = Q3 + (1.5 * IQR)
```

```
    return lower_range,upper_range
```

❖ The price dist. The plot of those results is displayed below.

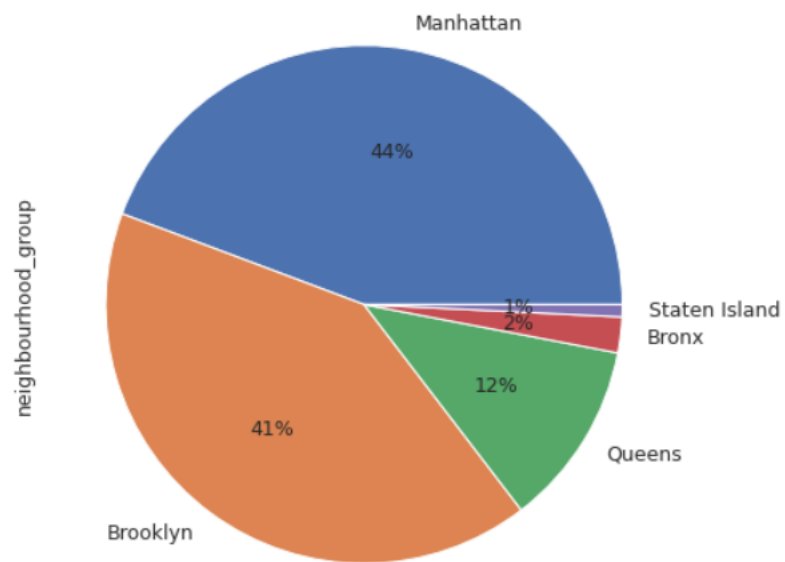


- ❖ carried out a Univariate analysis on various fields to examine the distribution. We discovered the following insights:
 - The pie graph shows that most available room types are complete homes/apartments, while shared rooms are the least common.



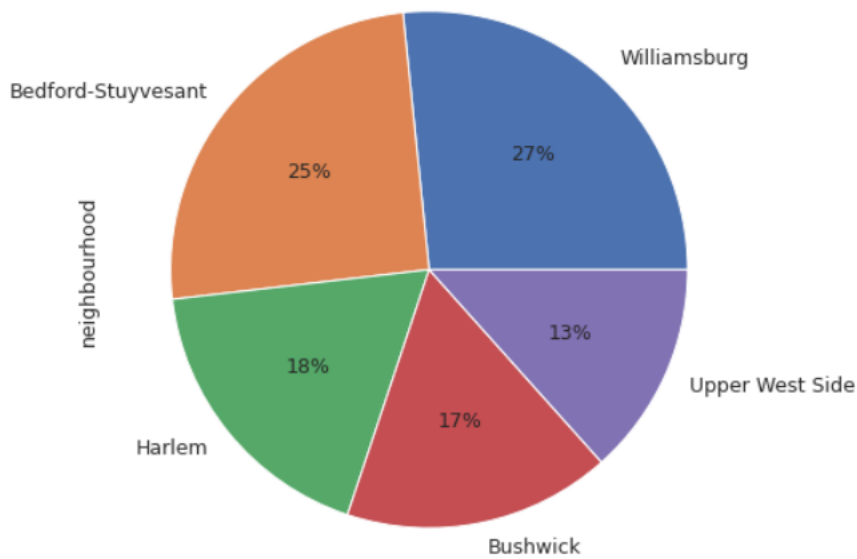
- In all of NYC, the neighbourhood group "Manhattan" has the most listings—nearly 44%!

`Out[41]: <Axes: ylabel='neighbourhood_group'>`

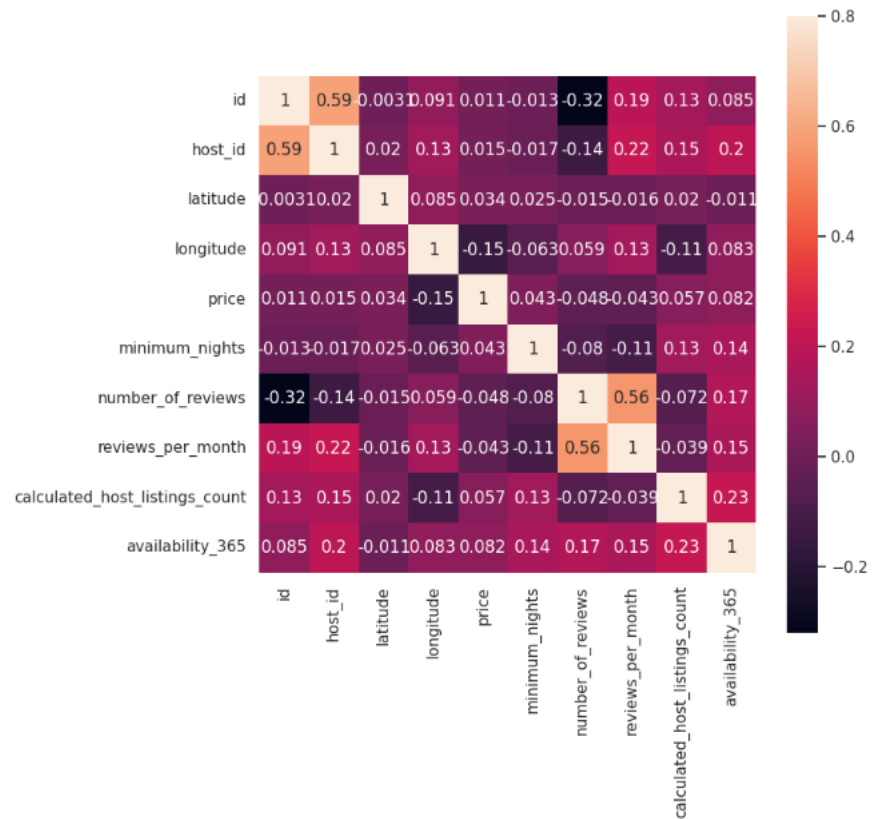


- based on listings in NYC, I looked up the top 10 neighborhoods and displayed the distribution of the top five.

```
<Axes: ylabel='neighbourhood'>
```



- ❖ Correlation analysis revealed a relationship between host_id, reviews_per_month, and availability_365. Additionally, there is a clear relationship between the minimum nights and the number of postings and

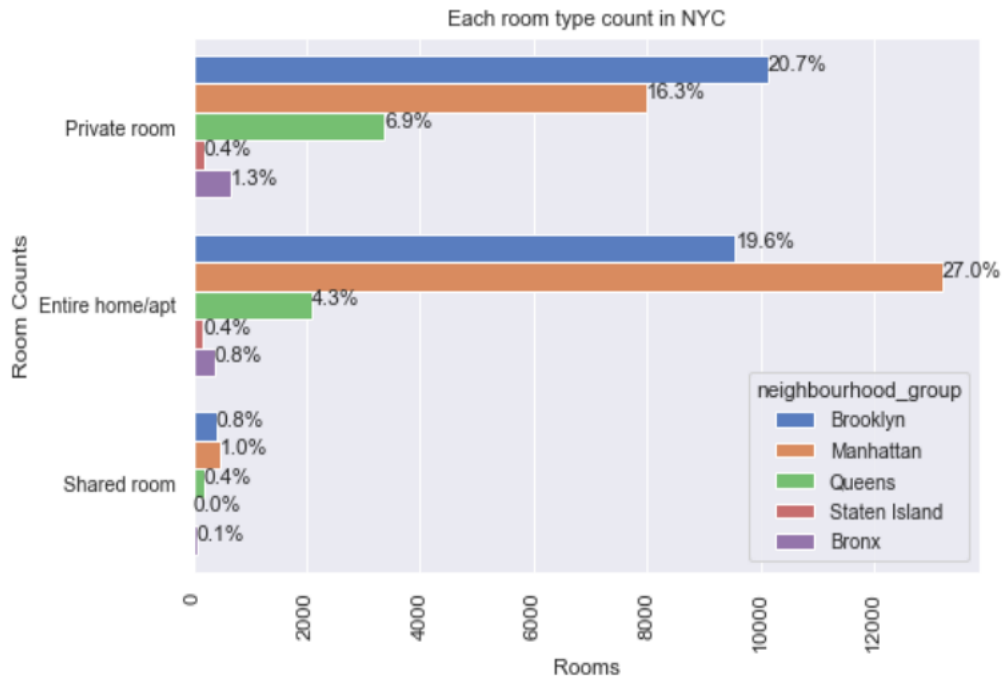


availability_365.

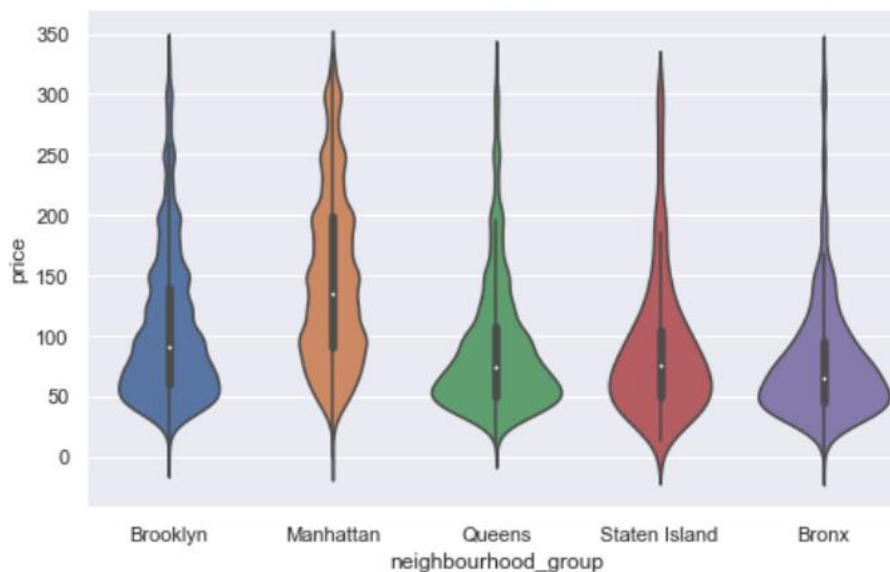
STEP 3: DATA ANALYSIS

- ❖ Brooklyn comes in second with about 19.6% of all listed properties, followed by Manhattan with about 27% of homes/apartments.

- We may infer that Manhattan, which has the biggest number of listings in all of NYC, has more complete home/apt room types, whereas Brooklyn, Queens, and the Bronx have more private room types. There are hardly any, if any, shared rooms in Staten Island and the Bronx, hence very few of the total listed have shared rooms available on Airbnb.

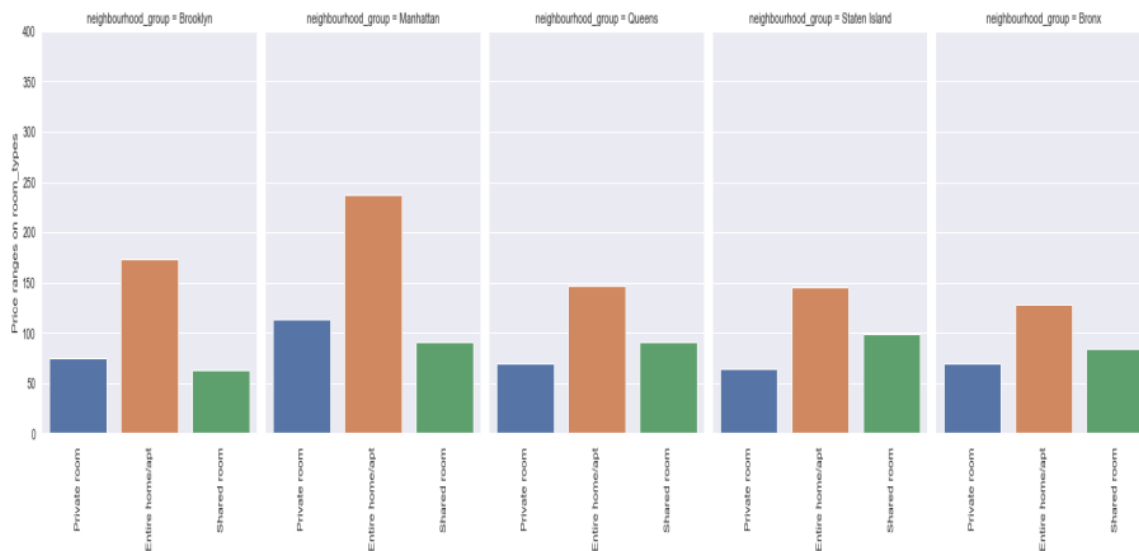


- ❖ We can deduce the following from a violin plot between the neighbour group and price:
 - The most expensive city to reside in is Manhattan, with an average listing price of more than 140 USD. Brooklyn is next, with an average listing price of over 80 USD.
 - Queens and Staten Island are priced similarly.



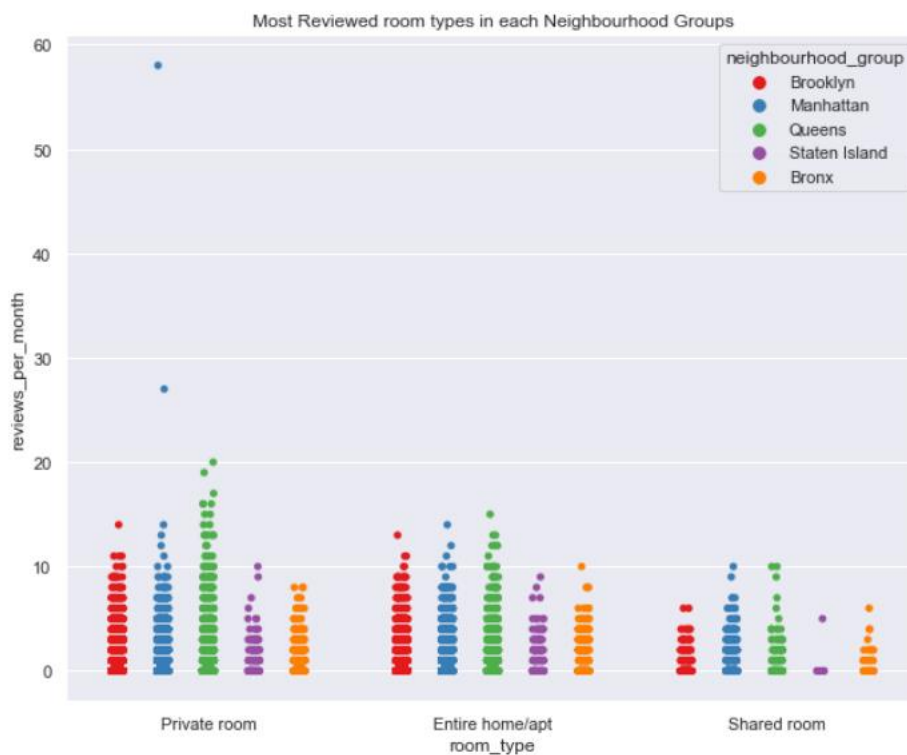
- ❖ According to the price distribution by room type and neighbourhood, Manhattan has the most expensive rooms, with Private rooms costing 110 USD/night and Entire homes/apartments costing around 240 USD/night. the least expensive are:
 - Entire home/Apt: Bronx
 - Private room: Staten Island

- Shared room: Brooklyn

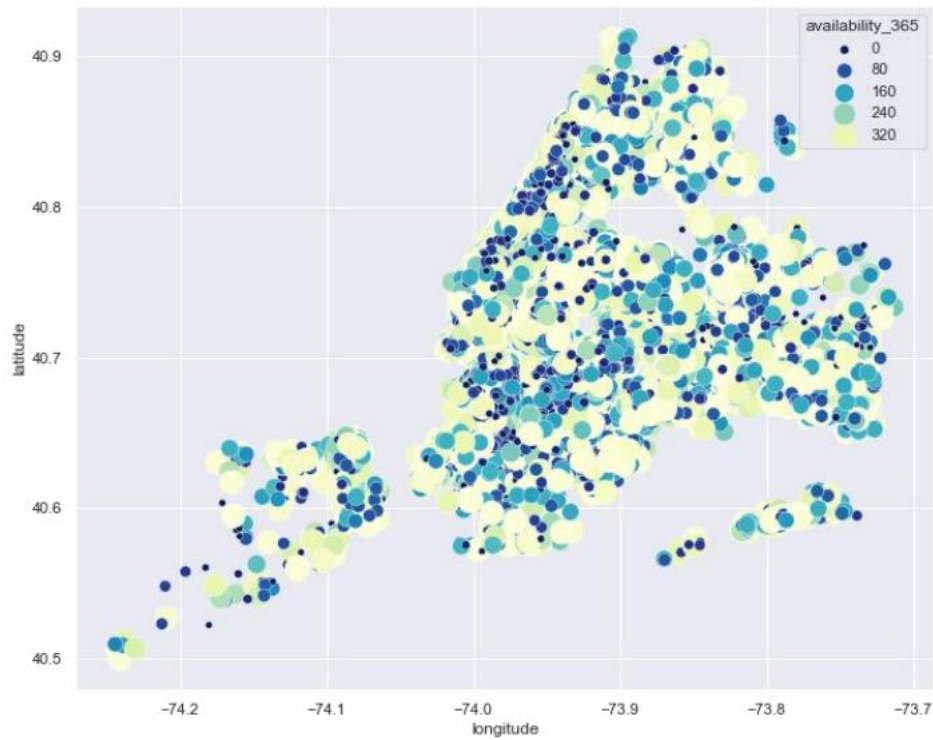


- Price reduction of entire apartments in Manhattan by at least 15% and an increase the in price of private rooms in Queens by at least 10% is advised.

❖ Private rooms were the accommodations with the highest number of reviews per month, with Manhattan receiving more than 50 reviews per month, followed closely by Queens. The most reviews for the entire home/apartment room type were found in Manhattan & Queens.



- Except for Manhattan, shared rooms are unpopular.
- ❖ Created a scatter plot to determine whether the listings are available year-round. Listings are typically accessible year-round in the Bronx and Staten Island, which may be the case given that they are not significantly more expensive than other boroughs like Manhattan, Brooklyn,



queens.

STEP 4: PRESENTATION

- ❖ Created a presentation using the best techniques.
- ❖ Added suggestions for the concerned departments.