# NEW YORK CITY AIRBNB

**Data Analysis and Presentation**


**Group 3**

Reynard Johns
Cortes Yip, Indira Hi Cheng
Gondaliya, Kenil Mansukhbhai
Salih, Aliah

George Brown College – Programming Fundation for Analytics

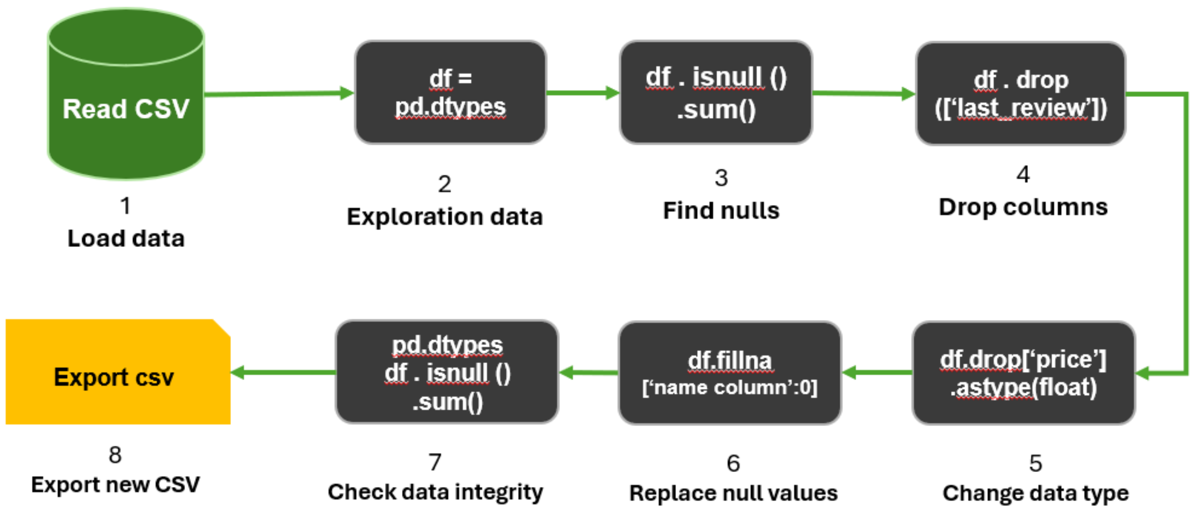
# AIRBNB REPORT

Table of Contents

## Cleaning data

# DATA CLEANING SECTION

### DATA CLEANING PROCESS




## Cleaning description process

### Data Loading and Initial Exploration

Importing necessary libraries and loading the Airbnb dataset:

```
[1]: #importing necessery libraries for future analysis of the dataset
     import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     import matplotlib.image as mpimg
     %matplotlib inline
     import seaborn as sns
```

```
[2]: #using pandas library and 'read_csv' function to read BlackFriday csv file as file already formated for us from Kaggle
     df=pd.read_csv("C:/Users/kenil/Downloads/AB_NYC_2019.csv/AB_NYC_2019.csv")
     #examing head of BlackFriday csv file
     df.head(3)
```

```
[3]: #checking amount of rows in given dataset to understand the size we are working with
     len(df)

[3]: 48895
```

```
[4]: #checking type of every column in the dataset
     df.dtypes

[4]: id                  int64
     name               object
     host_id             int64
```

| Variable | Data Types |
| --- | --- |
| id | int64 |
| name | object |
| host_id | int64 |
| host_name | object |
| neighbourhood_group | object |
| neighbourhood | object |
| latitude | float64 |
| longitude | float64 |
| room_type | object |
| price | int64 |
| minimum_nights | int64 |
| number_of_revie | int64 |
| last_review | object |
| reviews_per_month | float64 |
| calculated_host_listing_count | int64 |
| availability_365 | int64 |

These commands provide an overview of the dataset's structure, column names, data types, and basic statistics.

**Understanding, Wrangling and Cleaning Dataset:**

- **Finding the Null values from Dataset:**

```
]:
#looking to find out first what columns have null values
#using 'sum' function will show us how many nulls are found in each column in dataset
df.isnull().sum()
```

```
]:  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
```

From code we have found out the null values of name, host_name, Last_review and reviews_per_month.

- Drop unnecessary columns from the data set:

```
#dropping columns last_review
df.drop(['last_review'], axis=1, inplace=True)
df.head(3)
```

| | id | name | host_id | host_name | neighbourhood_group | neighbourhood | latitude | longitude | room_type | price | minimum_nights | number_of_reviews | revie |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2539 | Clean & quiet apt home by the park | 2787 | John | Brooklyn | Kensington | 40.64749 | -73.97237 | Private room | 149 | 1 | 9 | |
| 1 | 2595 | Skylit Midtown Castle | 2845 | Jennifer | Manhattan | Midtown | 40.75362 | -73.98377 | Entire home/apt | 225 | 1 | 45 | |
| 2 | 3647 | THE VILLAGE OF HARLEM....NEW YORK ! | 4632 | Elisabeth | Manhattan | Harlem | 40.80902 | -73.94190 | Private room | 150 | 3 | 0 | |

So, here we decided to remove last_review which mentions the date when review is placed so we this variable is insufficient and ineffective for future and also, we have a lot of null values on it so we decided to drop it.

- Changing the data type:

```
# changung data type of price from int to float
df['price'] = df['price'].astype(float)
```

Here, we are changing the data type of price from integer to float as we might have some values which are in float, so we decided to go with it and make it more useful for future predictive analysis and this variable is very crucial among all other variables.

- Replacing the null values:

```
34]: #replacing all NaN values in 'reviews_per_month' with 0
     df.fillna({'reviews_per_month':0}, inplace=True)
     df.reviews_per_month.isnull().sum()

34]: 0

36]: #replacing all Null values in 'host_name' with A
     df.fillna({'host_name':'A'}, inplace=True)
     #examing changes
     df.host_name.isnull().sum()

36]: 0

38]: #replacing all Null values in 'name' with z
     df.fillna({'name':'Z'}, inplace=True)
     df.host_name.isnull().sum()

38]: 0
```

- This operation replaces all NaN (Not a Number) values in the 'reviews_per_month' column with 0. The rationale behind this decision is likely based on the assumption that listings without any reviews per month can be considered to have zero reviews.
- This step replaces all null values in the 'host_name' column with the letter 'A'. The choice of 'A' as a placeholder is arbitrary and may serve as a flag for further analysis or processing.
- This operation replaces all null values in the 'name' column with the letter 'Z'. Again, 'Z' is used as an arbitrary placeholder for missing listing names.

## Cleaning data

- Checking data integrity and exporting the data:

```
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
reviews_per_month                 0
calculated_host_listings_count    0
availability_365                  0
dtype: int64
```

This indicates a complete dataset with no missing entries.

```
df.dtypes
```

```
id                                  int64
name                               object
host_id                             int64
host_name                          object
neighbourhood_group                object
neighbourhood                      object
latitude                          float64
longitude                         float64
room_type                          object
price                             float64
minimum_nights                      int64
number_of_reviews                   int64
reviews_per_month                 float64
calculated_host_listings_count      int64
availability_365                    int64
dtype: object
```

```
#df1.to_csv("Airbnb.csv")
df.to_csv ('C:/Users/kenil/Desktop/GBC/SEM 1/PROGRAMMING/final project/Airbnb.csv', index = None, header=True)
```

## Cleaning data

This analysis reveals a mix of data types:

- Integer (int64) for ID fields, counts, and night-related fields
- Float (float64) for geographical coordinates, price, and review frequency
- Object type for text-based fields like names and categories

The cleaned and verified dataset was exported to a CSV file. This operation saves the DataFrame to a CSV file named "Airbnb.csv" in the specified directory. The index=None parameter ensures that the DataFrame's index is not written to the CSV file, while header=True includes the column names as the first row of the CSV.

## SQL

# SQL SECTION

## Changing data set from CSV to Database

Import the dataset into a SQL database:

```
[46]: #Import the dataset into a SQL database.
      !pip install sqlite3

ERROR: Could not find a version that satisfies the requirement sqlite3 (from versions: none)
ERROR: No matching distribution found for sqlite3

[notice] A new release of pip is available: 24.1.1 -> 24.2
[notice] To update, run: python.exe -m pip install --upgrade pip

[47]: import sqlite3

[48]: connection = sqlite3.connect('C:/Users/kenil/Desktop/GBC/SEM 1/PROGRAMMING/final project/Airbnb.db')

[49]: df.to_sql('Airbnb',connection,if_exists ='replace')

[49]: 48895
```

By following these steps, the Airbnb dataset has been successfully imported into a SQL database using SQLite. This process involves installing the sqlite3 library, establishing a connection to the database, and using the to_sql method to transfer the DataFrame into a SQL table.

## Task 1: Average price of listings by neighborhood

## Result:

| | neighbourhood | Average_price |
|---|---|---|
| 1 | Allerton | 87.5952380952381 |
| 2 | Arden Heights | 67.25 |
| 3 | Arrochar | 115 |
| 4 | Arverne | 171.7792207792208 |
| 5 | Astoria | 117.18777777777778 |
| 6 | Bath Beach | 81.76470588235294 |
| 7 | Battery Park City | 367.5571428571429 |
| 8 | Bay Ridge | 144.43262411347519 |
| 9 | Bay Terrace | 142 |
| 10 | Bay Terrace, Staten I... | 102.5 |

221 0 affected 0.062 s

## Key Finding:

- The average price of Airbnb listings for each unique neighborhood in the dataset.
- Identification of neighborhoods with the highest and lowest average prices.
- Potential insights into price variations across different areas of the city.
- A basis for comparing affordability or pricing strategies between neighborhoods.

## SQL Querie – Task 1:

```
SELECT
   neighbourhood,
   AVG(price) AS Average_price
FROM
   Airbnb
GROUP BY
   neighbourhood;
```

**SQL**

# AIRBNB REPORT

## Task 2: Number of listings per neighborhood.

### Result:

| | neighbourhood | Number_of_listings |
|---|---|---|
| 1 | Allerton | 42 |
| 2 | Arden Heights | 4 |
| 3 | Arrochar | 21 |
| 4 | Arverne | 77 |
| 5 | Astoria | 900 |
| 6 | Bath Beach | 17 |
| 7 | Battery Park City | 70 |
| 8 | Bay Ridge | 141 |
| 9 | Bay Terrace | 6 |
| 10 | Bay Terrace, Staten I... | 2 |

221 0 affected 0.092 s

### Key Finding:

- The total count of Airbnb listings is calculated for each neighborhood.
- Neighborhoods with the highest number of listings are identified.
- Areas with fewer listings are pinpointed.
- The distribution of Airbnb properties across different neighborhoods is revealed.

### SQL Querie – Task 2:

```
SELECT
   neighbourhood,
   COUNT(*) AS Number_of_listings
FROM
   Airbnb
GROUP BY
   neighbourhood;
```

## Task 3: Availability of listings by neighborhood and room type.

### Result:

| | neighbourhood | room_type | Average_availability |
|---|---|---|---|
| 1 | Allerton | Entire home/apt | 193.9375 |
| 2 | Allerton | Private room | 145.03846153846155 |
| 3 | Arden Heights | Entire home/apt | 125.66666666666667 |
| 4 | Arden Heights | Private room | 0 |
| 5 | Arrochar | Entire home/apt | 213.9 |
| 6 | Arrochar | Private room | 293.90909090909093 |
| 7 | Arverne | Entire home/apt | 196.89583333333334 |
| 8 | Arverne | Private room | 180.25925925925927 |
| 9 | Arverne | Shared room | 95.5 |
| 10 | Astoria | Entire home/apt | 93.72965879265092 |

540 0 affected 0.115 s

### Key Finding:

- The average availability (in days) of listings is calculated for each combination of neighborhood and room type.
- Insights into which room types (e.g., entire home, private room, shared room) are available in specific neighborhoods are provided.
- Neighborhoods with the highest and lowest average availability for different room types can be identified.
- The data reveals patterns of availability, helping to understand seasonal or demand variations across different neighborhoods and room types.

**SQL Querie – Task 3:**

```
SELECT
   neighbourhood,
   room_type,
   AVG(availability_365) AS Average_availability
FROM
   Airbnb
GROUP BY
   neighbourhood,
   room_type;
```

**SQL**



## Task 4: Host information and the number of listings per host.

## Result:

| | host_id | host_name | Number_of_listings |
|---|---|---|---|
| 1 | 2438 | Tasos | 1 |
| 2 | 2571 | Teedo | 1 |
| 3 | 2787 | John | 6 |
| 4 | 2845 | Jennifer | 2 |
| 5 | 2868 | Letha M. | 1 |
| 6 | 2881 | Loli | 2 |
| 7 | 3151 | Eric | 1 |
| 8 | 3211 | Catherine | 1 |
| 9 | 3415 | Nataraj | 1 |
| 10 | 3563 | Arnie | 1 |

37457 0 affected 0.362 s

## Key Finding:

- The total number of listings managed by each host is calculated.
- Hosts with the highest number of listings are identified.
- The distribution of listings among different hosts is revealed, showing whether most hosts manage a few listings or many.
- Insights into the concentration of listings per host, which can indicate the presence of professional hosts versus casual hosts.

**SQL Querie – Task 4:**

```
SELECT
   host_id,
   host_name,
   COUNT(*) AS Number_of_listings
FROM
   Airbnb
GROUP BY
   host_id,
   host_name;
```

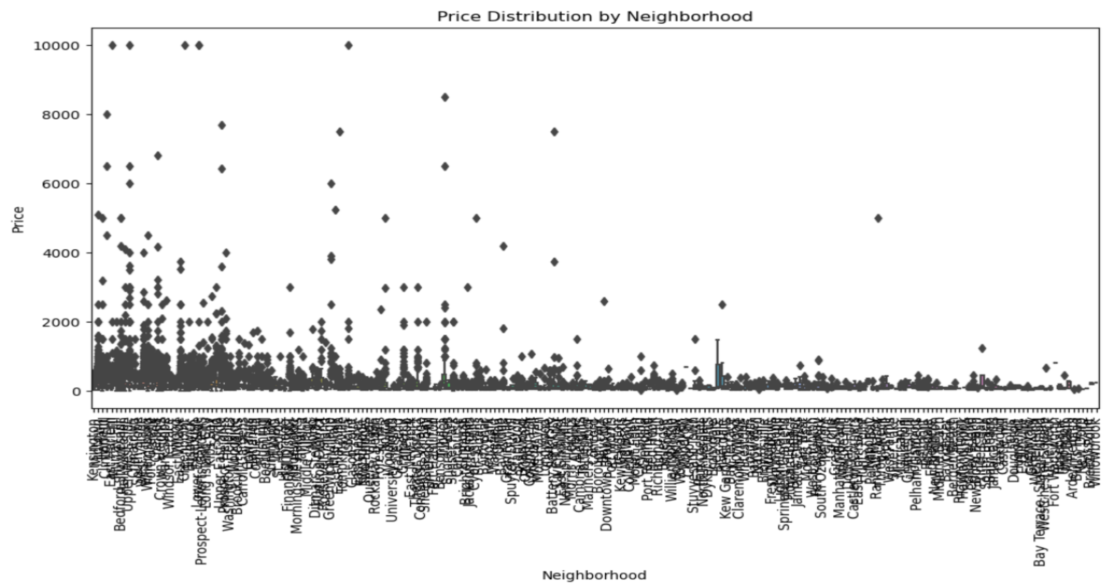
PYTHON

AIRBNB REPORT

# PYTHON SECTION

### Task 1: Visualization of price distribution by neighborhood using Matplotlib and Seaborn.

## Result:




## Key Finding:

1. **Price Range:** Prices range from approximately $0 to $10,000, with a majority of listings concentrated between $0 and $4,000.
2. **Price Clusters:** There are distinct price clusters, suggesting different tiers of accommodation options within the neighborhoods.
3. **Neighborhood Diversity:** A wide range of neighborhoods are represented, indicating a diverse selection of locations.
4. **Outliers:** Some neighborhoods have listings with significantly higher prices, potentially indicating luxury or unique properties.
5. **Price Variation:** There's considerable price variation within each neighborhood, suggesting factors beyond just location influence pricing.

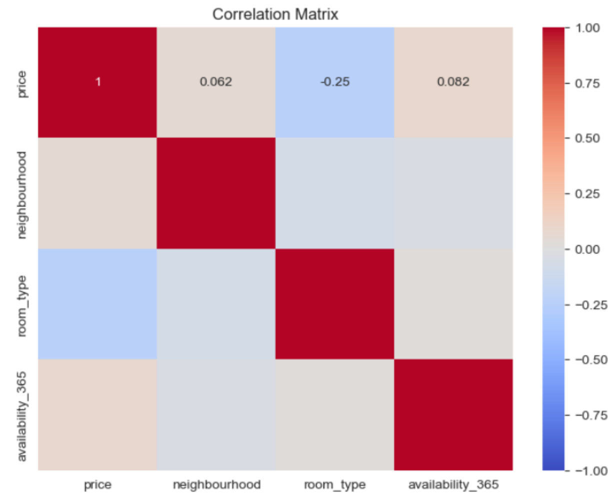
**Python code – Task 1:**

```
import seaborn as sns
import matplotlib.pyplot as plt

# Set the plot size
plt.figure(figsize=(12, 6))

# Create a boxplot for price distribution by neighborhood
sns.boxplot(x='neighbourhood', y='price', data=df)
plt.xticks(rotation=90)
plt.title('Price Distribution by Neighborhood')
plt.xlabel('Neighborhood')
plt.ylabel('Price')
plt.show()
```

# Task 2: Correlation analysis between price and factors such as neighborhood, room type, and availability.

## Result:

```
                     price  neighbourhood  room_type  availability_365
price             1.000000       0.062057  -0.249350          0.081829
neighbourhood     0.062057       1.000000  -0.071964         -0.034655
room_type        -0.249350      -0.071964   1.000000          0.022500
availability_365  0.081829      -0.034655   0.022500          1.000000
```




## Key Findings:

- Neighborhood characteristics have a minor impact on Airbnb prices.
- Room types are a more significant determinant of prices compared to neighborhoods and availability.
- Higher availability slightly correlates with higher prices, but the relationship is weak.
- Overall, the correlation analysis indicates that while neighborhood and availability have some influence, room type is a more substantial factor in determining Airbnb prices.

**Python Code – Task 2:**

```
# Encode categorical variables
df_encoded = df.copy()
df_encoded['neighbourhood'] = df_encoded['neighbourhood'].astype('category').cat.codes
df_encoded['room_type'] = df_encoded['room_type'].astype('category').cat.codes

# Calculate the correlation matrix
correlation_matrix = df_encoded[['price', 'neighbourhood', 'room_type', 'availability_365']].corr()

# Display the correlation matrix
print(correlation_matrix)

# Visualize the correlation matrix
plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', vmin=-1, vmax=1)
plt.title('Correlation Matrix')
plt.show()
```

# Task 3: Predictive analysis or modeling

## Result:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  price   R-squared:                       0.114
Model:                            OLS   Adj. R-squared:                  0.109
Method:                 Least Squares   F-statistic:                     27.95
Date:                Mon, 05 Aug 2024   Prob (F-statistic):               0.00
Time:                        11:09:58   Log-Likelihood:            -3.3444e+05
No. Observations:               48895   AIC:                         6.693e+05
Df Residuals:                   48671   BIC:                         6.713e+05
Df Model:                         223
Covariance Type:            nonrobust
```

## Key Finding:

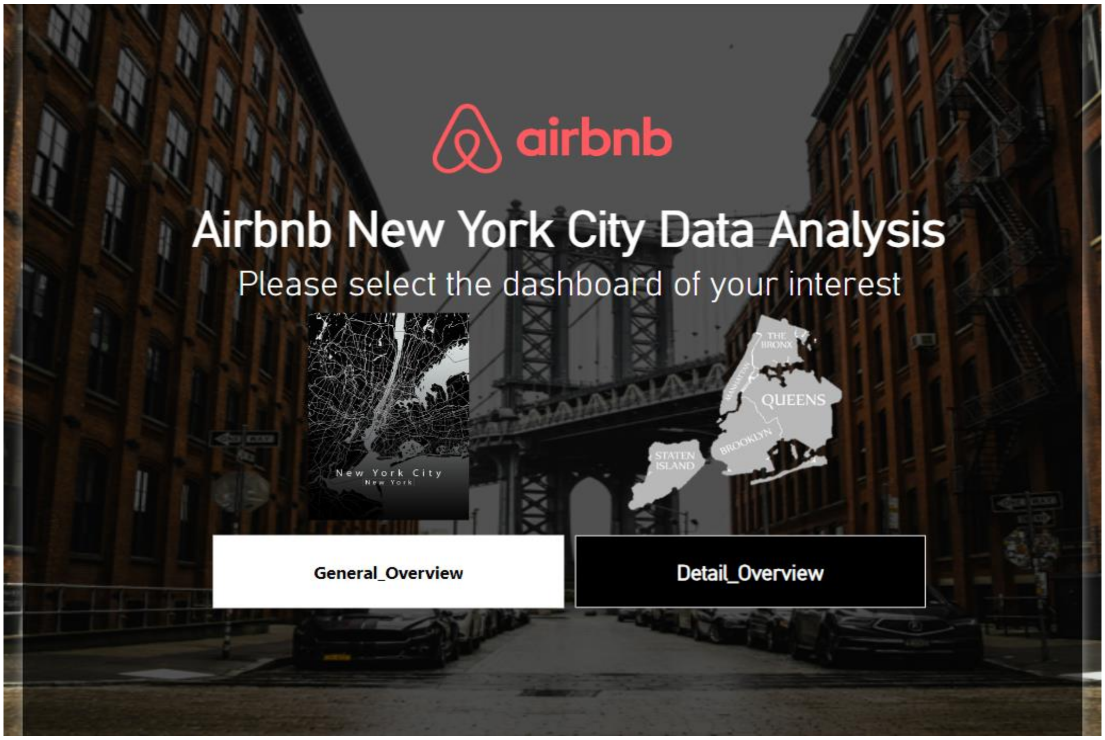
- **Neighborhood Impact on Prices**:
  Neighborhood characteristics significantly affect Airbnb prices. Certain neighborhoods command higher rates due to location, amenities, and demand. This helps stakeholders understand valuable areas.
- **Explained Variance**:
  The model explains approximately 11.4% of the variance in Airbnb prices (R-squared = 0.114), indicating that while other factors influence prices, the included variables provide meaningful insights.
- **Model Significance**:
  The model is statistically significant (F-statistic = 27.95, p-value = 0.00), confirming that the independent variables collectively impact Airbnb prices.
- **Log-Likelihood and Model Fit**:
  The log-likelihood value of -334440, along with AIC (669300) and BIC (671300), suggests the model fits the data but has room for improvement.
- **Practical Implications**:
  Insights from the analysis enable stakeholders to make informed decisions. Investors can target high-demand neighborhoods, sellers can price competitively, and urban planners can allocate resources effectively.

**Python Code – Task 3:**

```
import statsmodels.api as sm
import statsmodels.formula.api as smf

# Clean column names
df.columns = [mystring.replace(" ", "_").replace("(", "").replace(")", "") for mystring in df.columns]

# Correct OLS model specification
results = smf.ols(formula='price ~ neighbourhood + room_type + availability_365', data=df).fit()

print(results.summary())
```

## POWER BI SECTION

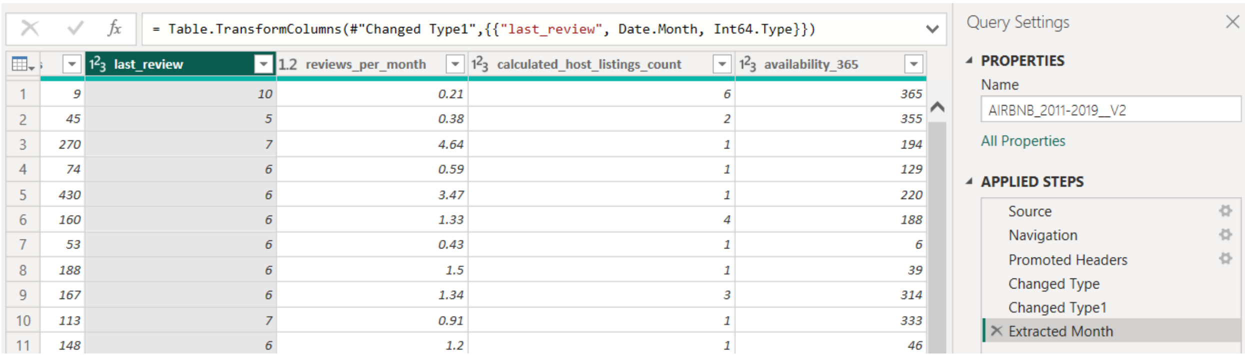
### Main page navigation




### Full Dashboard Airbnb New Your City Overview

## Data Manipulation: Review Month

# Result:




| | | last_review | reviews_per_month | calculated_host_listings_count | availability_365 |
|---|---|---|---|---|---|
| 1 | 9 | 10 | 0.21 | 6 | 365 |
| 2 | 45 | 5 | 0.38 | 2 | 355 |
| 3 | 270 | 7 | 4.64 | 1 | 194 |
| 4 | 74 | 6 | 0.59 | 1 | 129 |
| 5 | 430 | 6 | 3.47 | 1 | 220 |
| 6 | 160 | 6 | 1.33 | 4 | 188 |
| 7 | 53 | 6 | 0.43 | 1 | 6 |
| 8 | 188 | 6 | 1.5 | 1 | 39 |
| 9 | 167 | 6 | 1.34 | 3 | 314 |
| 10 | 113 | 7 | 0.91 | 1 | 333 |
| 11 | 148 | 6 | 1.2 | 1 | 46 |

With this first step in changing the type of the review column, in addition to only extracting the month of the review, more insights can be gained on the following:

1. The month with the most reviews
2. The months when guests visit/review properties New York City the most

In addition, to fully conceptualize trends on Power BI, missing values were removed – namely, prices less than USD 10. This was executed to capture valuable insights about Airbnb, where the platform's minimum charge for any listing in New York is USD.

# Key Findings

### Overview

The overview NYC's 2019 Airbnb listings has allowed for findings that cannot be understood in isolation from the context of New York City. To begin, there is a total of 5 neighbourhoods groups: Manhattan, Brooklyn, Queens, Bronx, and Staten Island. Within these groups, there are 217 neighbourhoods. The total number of hosts is approximately 18,000, with approximately 26,000 properties; meaning that many hosts manage more than one property.

### Hosts and Prices

The neighbourhood groups with the most hosts are Manhattan at the lead with 41% of the total hosts, followed by Brooklyn at 25%, Queens at 18%, Bronx at 9%, and Staten Island at 7%.

Furthermore, Power BI was used to visualize average prices in neighbourhood groups as well as neighbourhoods. This presented Manhattan as the highest average price per night, and Sea Gate in Brooklyn as the highest average price. The most reviewed neighbourhoods were also Manhattan and Brooklyn.

**Reviews**

For the date of the reviews, this was utilized as previously mentioned to seek an understanding of when NYC is visited the most; this can be the month with the most reviews and, when properties are most reviewed throughout the year. The line graph's visualization showed February, September, and December as the months with the highest reviews.

Reviews are also used as (numbers of reviews) to see how they correlate with prices through a plot graph. The scatter plot displays the relationship between the number of reviews (x-axis) and the prices (y-axis). The graph demonstrates an inverse relationship, where the general trend conveys that as the number of reviews increases, the price tends to decrease.

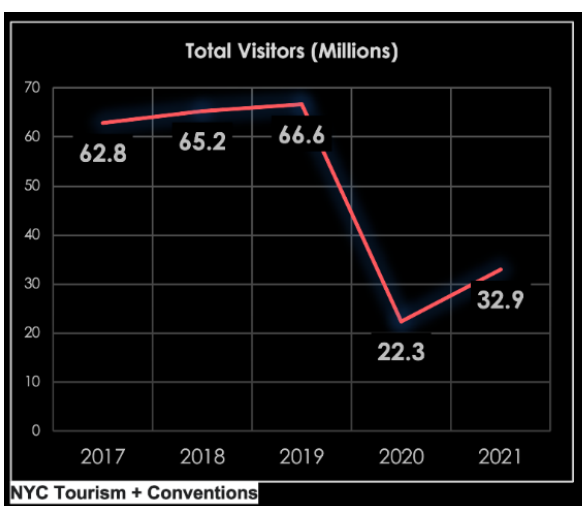
**Availability**

Finally, the availability of properties per year reveals that the most expensive neighbourhood groups, such as Manhattan and Brooklyn have the least availability of nights per year, while Staten Island, which has less properties and hosts, has the most availability throughout the year.

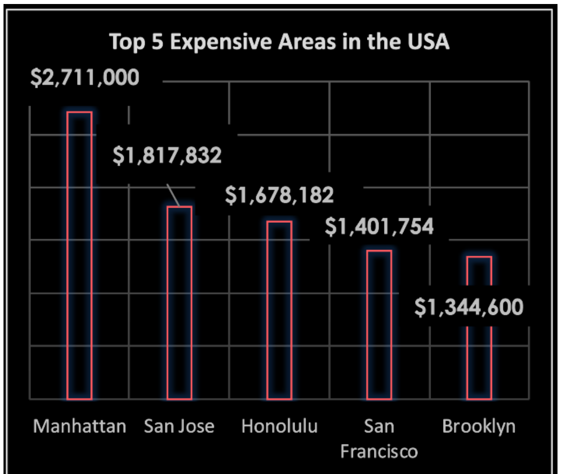
**Interpretation of Findings**

**NYC in 2019: The Manhattan and Brooklyn Phenomena**

The year 2019 was a benchmark year for NYC's tourism. It was recorded that 66+ Million visitors went to NYC. This number, like in many cities around the world, went down the year after, due to COVID-19, and tourism is slowly recovering. The graph below displays the number of visitors before and after 2019:



Moreover, according to the Case-Shiller U.S. National Home Price NSA Index, out of the most expensive real-estate areas/cities in the USA, Manhattan and Brooklyn are number one and five, respectively. This is in the same top 5 list against cities like San Jose, San Francisco, and Honolulu, which are different in their geography, offering more beaches and tropical climates. This is important to note when looking at NYC, which is far from a 'beach city' or setting, with mostly buildings and metropolitan atmospheres. However, this is due to NYC boroughs being central to the US's cultural and leisure scene – with most visitors seeking to experience diverse activities that reflect American big city lifestyle, as well as other factors, such as business trips.




Case-Shiller U.S. National Home Price NSA Index

Furthermore, the most expensive neighbourhood (average price) is Seagate, which is a gated and private community, and it is located in Coney Island – one of NYC's major attractions. Another neighbourhood noted for its high average price is Tribeca, where the Tribeca Film Festival takes place every year. This is followed by Soho and Noho, offering visitors unique cultural and historical atmospheres while visiting NYC.

To be more specific however, in 2019 major evets took place in NYC that can explain the high price in Manhattan, real-estate prices aside. In March 2019, the opening of the Vessel took place, and in April 2019, the opening of the Shed followed. This is part of the city's expansion of the Hudson Yards – a project that is still ongoing. These launches were one of the first, with the Hudson Yards being chief authentic 'NYC experience' areas for visitors. The Shed hosts hundreds of high-profile public events like concerts and art shows, while the Vessel is a structure that allows visitors to enter and view NYC from different vantage points. The Vessel, in 2019, had a range of 25,000 to 50,000 visitors per weekend – this undoubtedly makes Manhattan an attractive area to seek staying in while scanning NYC for several attractions, especially newly available ones.

Even more importantly, in June 2019, it was the 50th anniversary of the Stonewall Riots. This is a hisotircally significant event for the LGBTQ community in NYC, North America, and the world because it represents a milestone in the community's initiation of pride culture and celebrations. This event alone, in June, attracted 3 million visitors to NYC – and it took place in Manhattan as well.

## Reviews, Prices, and Listing Marketing: WordCloud Integration with Power BI

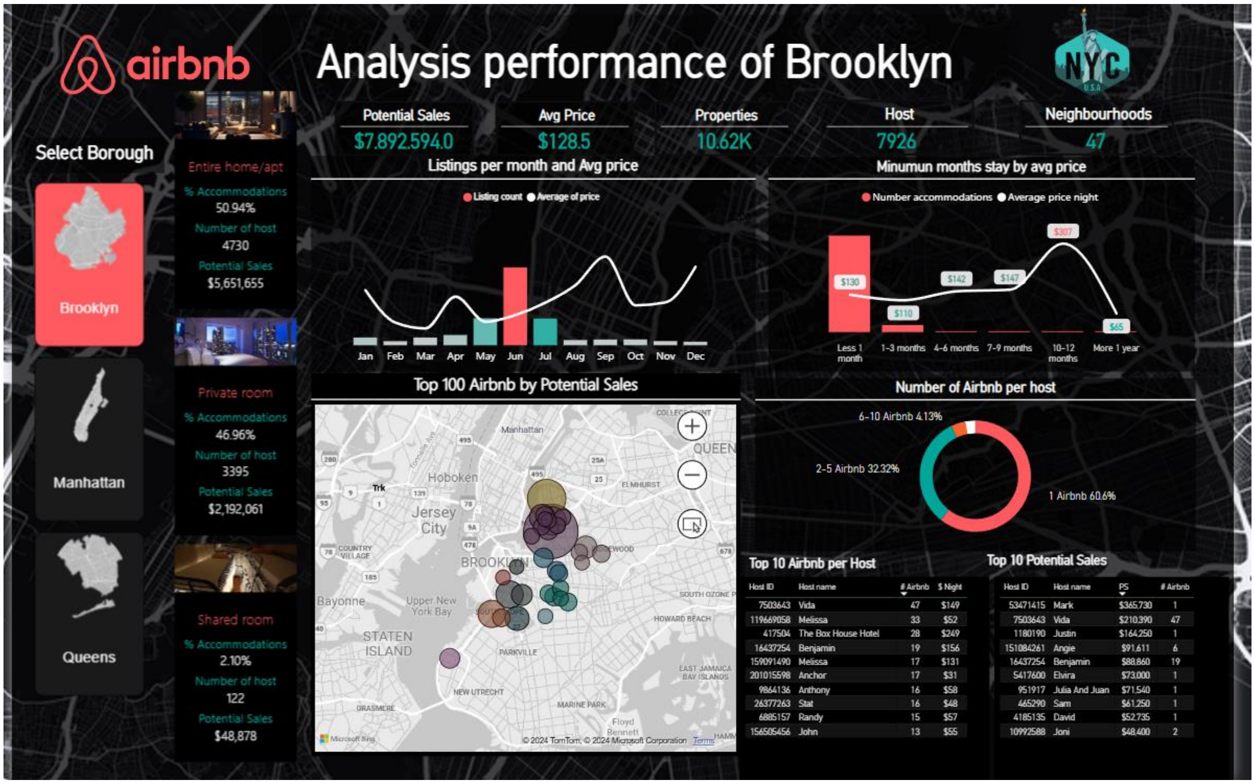
In platforms like Airbnb, two things can attract a guest: the description of the property and the photos. Since the data offers us the names of the properties, one useful tool would be to look at Word Clouds as a useful tool to integrate after seeing results in Power BI. Using Power BI, the properties with the most reviews were sorted in descending order and their titles were copied to produce a Word Cloud. The words that appeared the most were 'Private', 'Room', 'Bedroom', 'Cozy', 'JFK'. These words reflect the scatterplot's results, which illustrate that lower priced properties have more reviews. Lower prices in this dataset pertain to private rooms and bedrooms – usually for single guests, or groups of guests willing to share rooms, or business trips that do not require a bigger property for the visit. However, to understand this more, it can be hypothesized that these rooms receive more reviews because they are the most occupied units. In addition, they are also usually units that are more easily inspected for flaws or strengths, and thus, more reviews are generated.

Conclusively, however, it is worth noting that words like 'Cozy' and 'JFK' indicate a marketing strategy where the appeal for these smaller properties are usually the atmosphere of the unit – and its proximity to the airport, reflecting the hypothesis that these are units that appeal to singles, groups of friends, or businesspeople.

In another attempt to examine marketing strategies, Power BI was used to sort the most expensive units in descending orders. These properties are mostly in Manhattan and Brooklyn. The Word Cloud revealed that more affluent wording in marketing the unit, with recurring words such as 'Luxury', 'Loft', 'Townhouse', 'Manhattan', 'Central', 'Village', 'Beautiful', 'View', 'Park', and 'Soho' solidifying the exclusiveness of being in Manhattan, with types of properties mostly being Townhouses and Lofts, in places such as the Village and Soho, with proximity to Central Park, and features such as having Views, and being Central, with Luxury and Beautiful being major adjectives in these descriptions

## Full Dashboard Airbnb Neighborhood Group Overview

### Brooklyn View

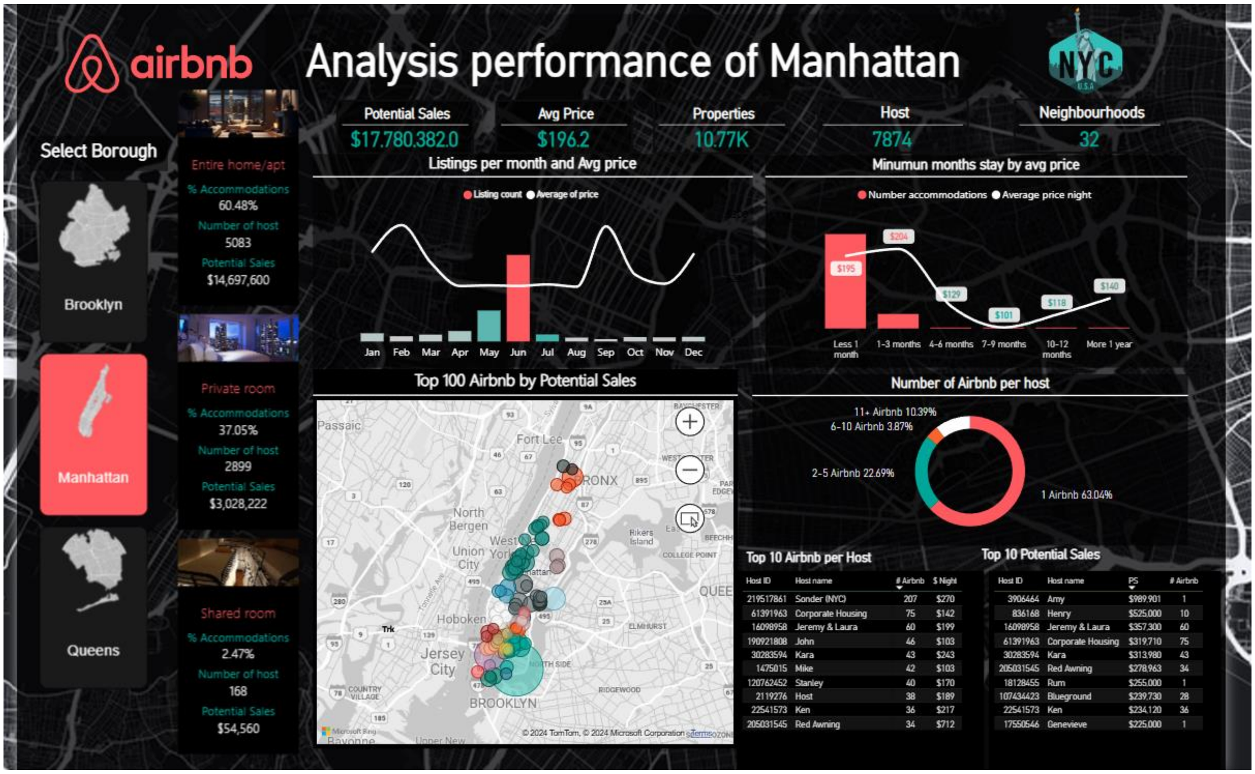



### Key Findings:

- Brooklyn has more hosts and almost the same number of accommodations compared with Manhattan, however, Brooklyn has 56% less sales potential than Manhattan because of the average price of night ($128.5 average price).
- The entire home/apartment and Private room have almost the same number of accommodations in Brooklyn, but with the entire home/apt Airbnb can get more potential sales due to the size of the properties-more expensive each night stay.
- Brooklyn offers the best value per night just until the guest stay more than 1 year stay ($65 average price per night).
- The average price per night increases considerably after the high season, which indicates hosts pursue more revenue per stay because of the low booking level.
- The host Vida leads the list of accommodations per host with 47 Airbnb, however, all the properties that the host owns do not generate more potential sales than the host Mark who only owns 1 Airbnb.

**POWER BI**

## Manhattan View



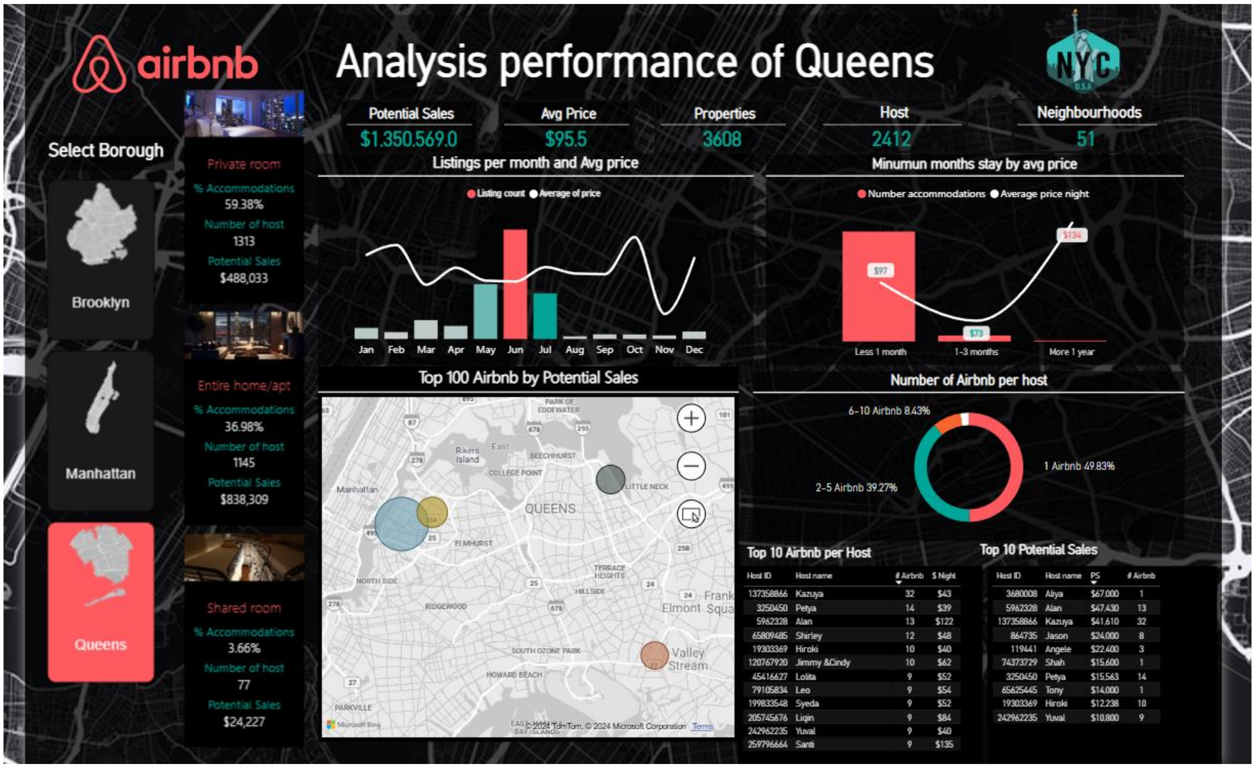

## Key Findings:

- Manhattan has more potential sales than Brooklyn due to this borough's higher average price per night.
- The months of May to Jul registered the highest number of listings, and combined with the average price per night indicates the host needs to reduce their prices because of the high tourist season.
- Manhattan offers more accommodations with a minimum requirement of 1-3 months night stay, but these accommodations don't offer a better value per night compared with accommodations of less than 1 month night stay.
- The entire home/apartment and Private room have almost the same number of accommodations in Brooklyn, but with the entire home/apt Airbnb can get more potential sales due to the size of the properties-more expensive each night stay.
- The number of hosts who own more than 11+ Airbnb is higher in Manhattan, where Sonder (NYC) leads the list of hosts with the most properties with 207 Airbnb.
- Manhattan has more Airbnb with potential sales in the top 100 compared with the other Boroughs, where the host Amy with the host ID 3906464 owns a private room with the most sales potential with $990 per complete stay.

# AIRBNB REPORT

**POWER BI**

## Queens View




## Key Findings:

- Queens offers the most affordable accommodation price per night ($95.5 average price), and also has more neighborhoods compared with the other boroughs.
- Although the entire home/apartment covers only 37% of the accommodation available in Queens, the sales potential of this type of room is higher than all the private rooms together.
- The average price during the year doesn't vary much during the year, only from October to February are the highest average prices.
- In June there were more listings where the highest number of listings per host was 721, however, in May the highest number of listings per host was 1,339 which indicates that some hosts are trying to position their Airbnb before the highest month of the year.
- Queens offers less variety of minimum required accommodation where there is no benefit in the price per night for the accommodation with a minimum stay of 1 year +
- The number of accommodations in the top 100 Airbnb by potential sales is the lowest compared to other boroughs.

# Conclusions

**All Boroughs**

- Manhattan's market is highly competitive with a mix of luxury and professionalized listings. Brooklyn offers a more balanced market, while Queens caters to budget-conscious travelers. Pricing strategies should align with these borough-specific trends.
- Manhattan is heavily saturated with professional hosts, while Brooklyn and Queens offer more opportunities for smaller or new hosts.
- High-end listings are most profitable in Manhattan, but Brooklyn and Queens offer steady, mid-range returns with lower competition.
- All boroughs experience peak bookings in summer (June-July), making dynamic pricing a key tactic.
- Queens and Brooklyn have strong demand for private rooms, while Manhattan favors entire homes/apartments.

**Brooklyn**

- Brooklyn's market appeals to a wide audience, from families seeking entire homes to solo travelers needing private rooms. It's an ideal entry market for new hosts, offering growth potential without the intense competition of Manhattan.

**Manhattan**

- Manhattan's market is dominated by professional hosts targeting high-end clientele. Success here requires offering premium experiences, targeting niche markets, or finding unique value propositions in a saturated environment.

**Manhattan**

- Queens is a budget-friendly market catering to students and long-term renters. It's an ideal location for hosts looking to enter the market with lower capital investment while focusing on affordability and essential amenities.

**Overall**

While premium neighborhoods command higher rates, location alone doesn't drive prices—factors like room type and amenities are more influential.
The most significant determinant of pricing is the type of accommodation offered (e.g., entire home, private room), outweighing location and availability.

Pricing varies widely within neighborhoods, creating distinct market tiers from budget options to luxury listings. This indicates that factors beyond location, such as property features and guest demand, are crucial.

Investors should focus on high-demand areas with room types that maximize returns, while sellers can use neighborhood data to set competitive, well-informed prices. Urban planners can leverage these insights for resource allocation and development.