

# Final Project

## Phân tích và đánh giá thị trường Airbnb tại New York

Nhóm 1:

- Nhan Như Ngọc
- Huỳnh Kim Ngân
- Văn Thụy Trâm Anh

# Mục lục

## 01

### Tổng quan

1. Giới thiệu dataset
2. Tiền xử lý dữ liệu

## 02

### Phân tích

Phân tích theo 5 bước  
(tổng quan, khu vực, giá,  
loại phòng, nhu cầu thực  
tế của khách hàng,...)

## 03

### Kết luận

Dựa trên kết quả phân  
tích đưa ra đề xuất

# 01

## Tổng quan

# Tổng quan

## Airbnb là gì?

Airbnb là một nền tảng trực tuyến kết nối giữa người cần thuê nhà, căn hộ, hoặc phòng trống với những người có không gian ở dư, muốn cho thuê. Được thành lập vào năm 2008, Airbnb đã trở thành một trong những công ty đầu tiên và lớn nhất trong lĩnh vực chia sẻ nơi ở. Người sử dụng có thể đăng ký tài khoản trên nền tảng này, tìm kiếm các lựa chọn ở trên khắp thế giới, đặt phòng và trả tiền trực tuyến. Đối với những người có không gian ở dư, họ có thể đăng thông tin về chỗ ở của mình lên Airbnb và kiếm thu nhập bằng cách cho thuê ngắn hạn.



# 1.1 Giới thiệu dataset

## **Bối cảnh**

Kể từ năm 2008, khách và chủ nhà đã sử dụng Airbnb để mở rộng khả năng du lịch và thể hiện cách trải nghiệm thế giới độc đáo, được cá nhân hóa hơn. Dataset này mô tả hoạt động niêm yết và số liệu ở NYC, NY cho năm 2019.

## **Nội dung**

Tập dữ liệu này bao gồm tất cả thông tin cần thiết để tìm hiểu thêm về máy chủ, tính khả dụng về mặt địa lý, các số liệu cần thiết để đưa ra dự đoán và đưa ra kết luận.

(ví dụ: địa điểm, giá cả, đánh giá, v.v.)



**Mô tả:** Bộ dataset có tất cả **16** cột, **48896** hàng, bao gồm các thông tin như:

**id:** Mã định danh duy nhất cho mỗi bản ghi trong cơ sở dữ liệu, giúp phân biệt giữa các mục khác nhau.

**name:** Tên của nơi lưu trú, có thể là căn hộ, nhà riêng, phòng trọ, vv.

**host\_id:** Mã định danh của người đăng bài cho thuê nơi lưu trú.

**host\_name:** Tên của người đăng bài cho thuê.

**neighbourhood\_group:** Khu vực hoặc quận lớn hơn mà nơi lưu trú thuộc về.

**neighbourhood:** Khu vực cụ thể mà nơi lưu trú đặt tại, thường là khu phố hoặc khu dân cư.

**latitude:** Vĩ độ của nơi lưu trú, thường được sử dụng để định vị trên bản đồ.

**longitude:** Kinh độ của nơi lưu trú, thường được sử dụng để định vị trên bản đồ

**room\_type:** Loại phòng được cung cấp (ví dụ: phòng riêng, phòng chia sẻ, căn hộ/nhà nguyên căn, vv.).

**price:** Giá thuê cho mỗi đêm hoặc cho mỗi đơn vị thời gian khác.

**minimum\_nights:** Số đêm tối thiểu mà người thuê đặt nơi lưu trú.

**number\_of\_reviews:** Tổng số lượt đánh giá mà nơi lưu trú đã nhận được từ khách hàng trước đó.

**last\_review:** Ngày đánh giá cuối cùng được thêm vào.

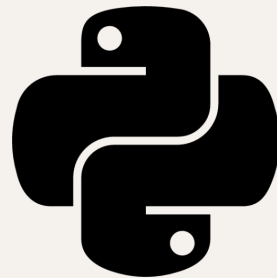
**reviews\_per\_month:** Số lượt đánh giá trung bình mỗi tháng.

**calculated\_host\_listings\_count:** Số lượng nơi lưu trú mà một người cho thuê đã đăng trên Airbnb.

**availability\_365:** Số ngày trong năm mà nơi lưu trú có sẵn để đặt phòng.

# Công cụ phân tích

Trong dự án này, nhóm đã áp dụng ngôn ngữ Python và một loạt các thư viện để lọc và làm sạch dữ liệu thô.



## Power BI

Dữ liệu sau khi làm sạch được đưa vào Power BI để tạo ra biểu đồ nhằm phân tích và đưa ra insight về các xu hướng cũng như nhu cầu về Airbnb ở NYC.

# Tiền xử lý dữ liệu

1/ Xóa các giá trị null, bị trùng lặp

```
▶ #Xóa các giá trị null, bị trùng lặp  
df=df.dropna()  
df = df.drop_duplicates( keep=False)  
print(df)
```

2/ Xác định các cột được sử dụng cho bài toán phân tích, loại bỏ bớt các cột không cần thiết như: `calculated_host_listings_count, id, last_review`

```
#Xóa các cột không cần thiết  
df = df.drop('calculated_host_listings_count', axis=1)  
df.drop(['id', 'last_review'], axis = 1, inplace = True)  
print(df)
```



# Tiền xử lý dữ liệu

## 3/ Xuất dữ liệu đã chuẩn hóa

```
#Xuất dữ liệu đã chuẩn hóa
df.to_csv('/content/AB_NYC_2019.csv',
          columns = ['name' , 'host_id' , 'host_name' , 'neighbourhood_group'
                    , 'neighbourhood' , 'latitude' , 'longitude' , 'room_type'
                    , 'price' , 'minimum_nights' , 'number_of_reviews'
                    , 'reviews_per_month' , 'availability_365' ],
          sep='\t', index = False)
```



02

Phân tích - Đánh giá

## 2. Nội dung phân tích - đánh giá





# 2.1

**Phân tích tổng quan dữ liệu**

## 1. Overview

47,458

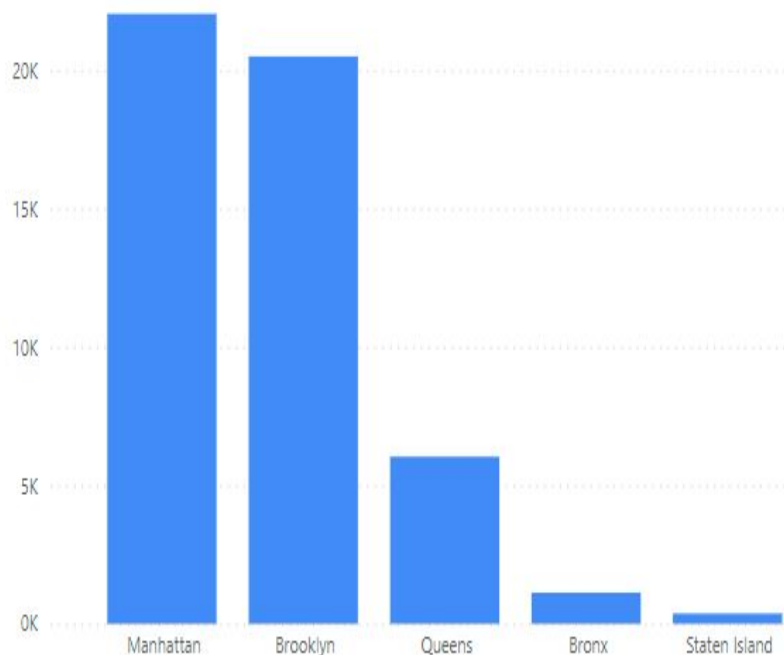
Tổng số nơi lưu trú  
tại New York

37,460

Chủ sở hữu

## 1. Overview: Số lượng listing dựa trên nhóm khu vực

Số lượng listing theo khu vực

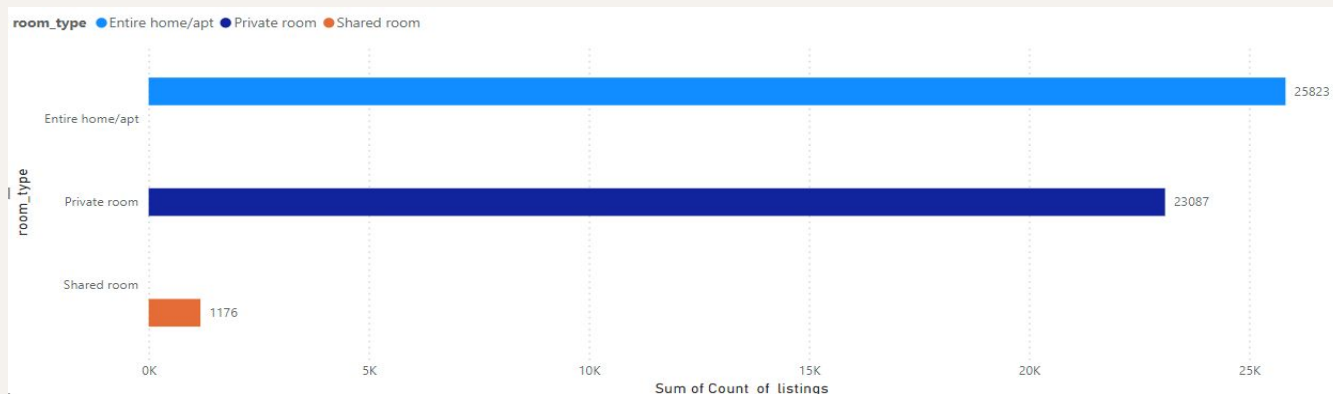


Tập trung chủ yếu tại 2 khu vực Manhattan và Brooklyn với tổng số ~43.000 listing, thấp nhất ở Staten Island với chỉ 374 listing được đăng tải trên Airbnb.

Manhattan và Brooklyn với mật độ dân cư đông đúc và là trọng tâm du lịch của New York với những điểm tham quan nổi tiếng, kiến trúc độc đáo. Ngoài ra còn là thủ phủ của văn hoá, tài chính và truyền thông, giải trí của thế giới => phần lớn listing nằm ở 2 quận này vì lượng nhu cầu lưu trú cho du lịch và công tác lớn.

Queens, the Bronx và Staten Island, mật độ dân cư thấp hơn cùng với vị trí cách xa trung tâm => ít tập trung đầu tư vào cho thuê lưu trú nên số lượng listing thấp hơn đáng kể.

# 1. Overview: Số lượng listing dựa trên room\_type



Entire home/apt chiếm phần lớn lượng listing, kế đến là private room và cuối cùng là shared room.

Các listing cho thuê Entire home/apt và private room chiếm phần lớn ở New York do nhiều lí do. Trước hết, thu nhập từ việc cho thuê toàn bộ căn hộ thường cao hơn và người thuê thường muốn không gian riêng tư. Nhu cầu này phản ánh mong muốn của du khách tìm kiếm không gian đầy đủ và tiện nghi, đặc biệt khi đi cùng gia đình hoặc nhóm bạn. Giá thuê toàn bộ căn hộ cũng có thể cạnh tranh hoặc rẻ hơn so với việc thuê phòng riêng. Mặc dù vậy, vẫn có một số du khách tìm kiếm sự giao tiếp với người địa phương hoặc muốn tiết kiệm chi phí, do đó, có nhiều listing cho thuê private room và shared room.

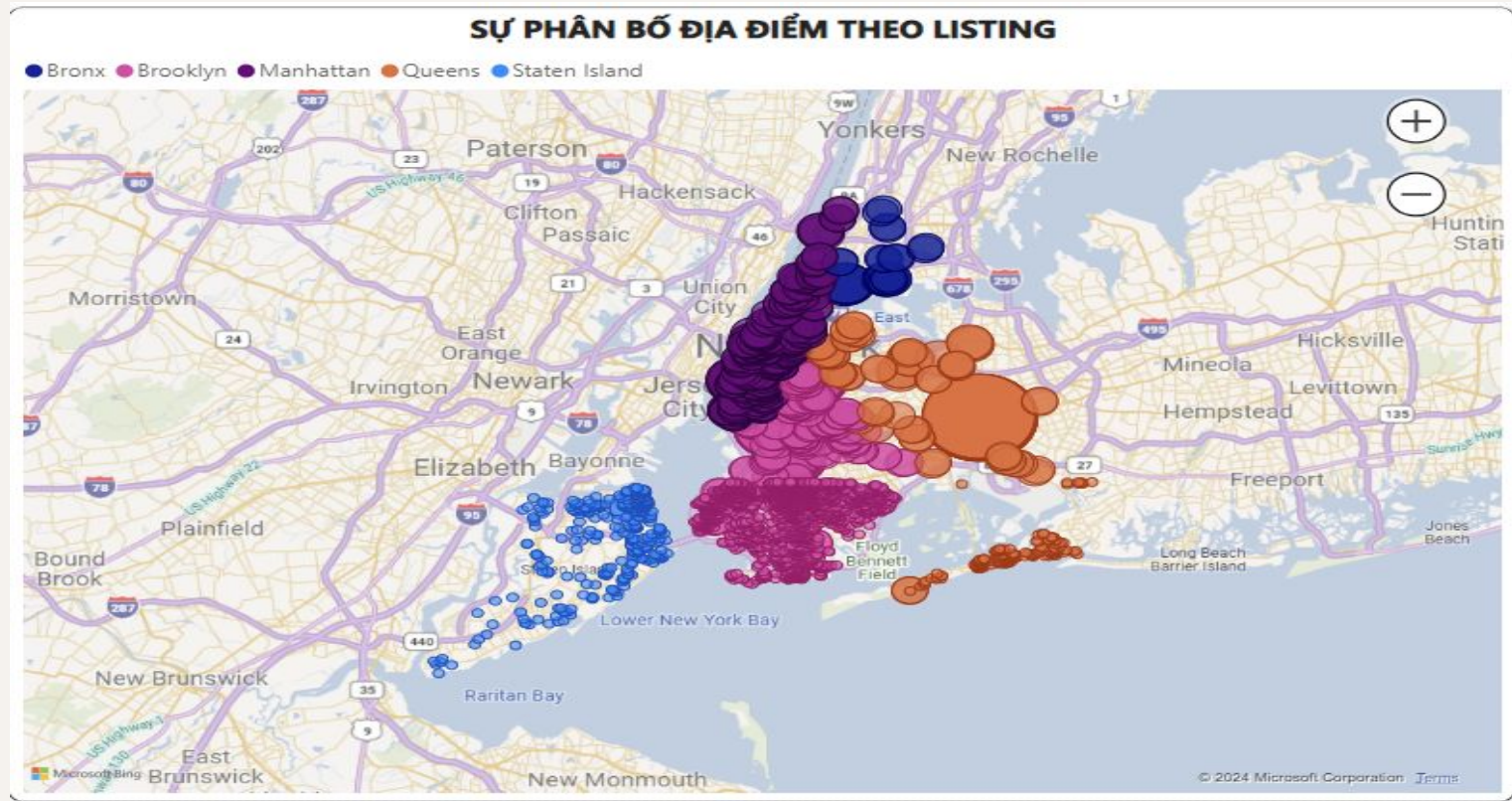
# 2.2

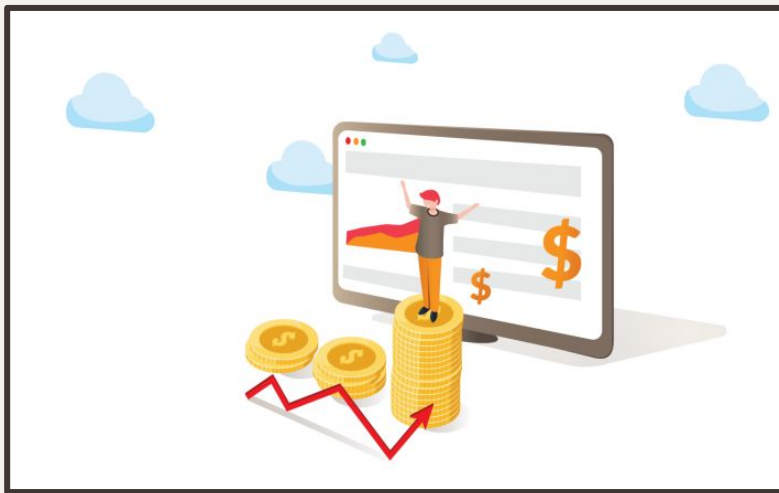
**Phân tích theo từng khu vực**





# Phân bố dựa trên **listing**

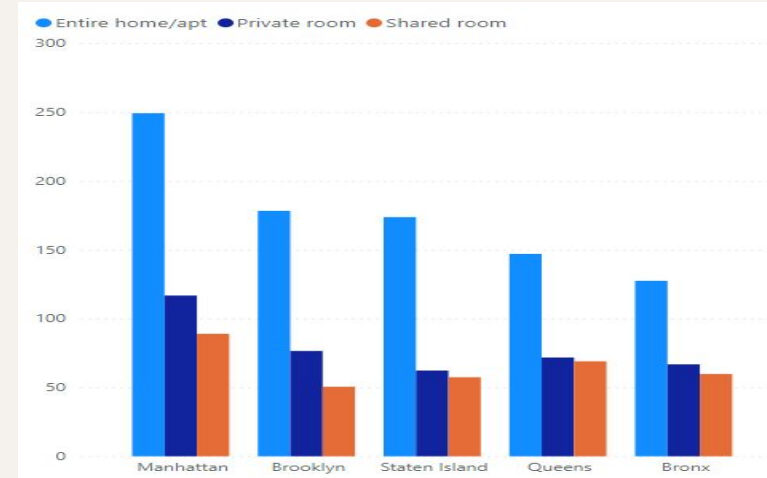
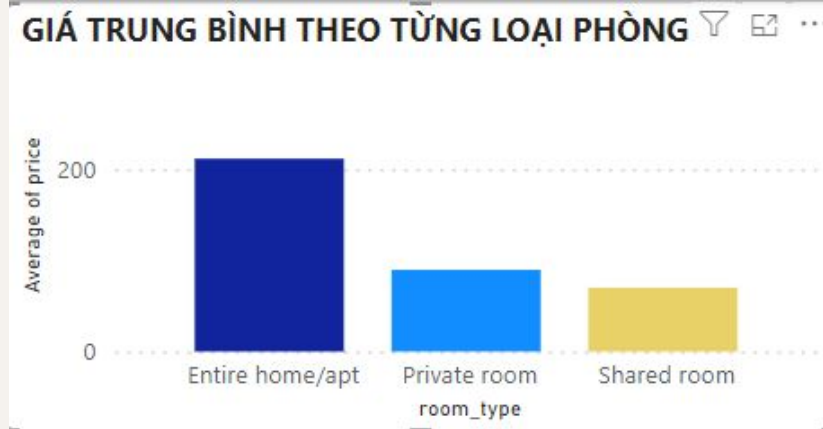




## 2.3

**Phân tích theo giá  
phòng**

# Giá trung bình theo từng loại phòng



**Entire home/apt** có mức giá trung bình cao nhất \$212, gần như gấp 3 so với 2 loại phòng còn lại ( và \$70). Lý do là khách thuê sẽ được sử dụng toàn bộ căn nhà với đầy đủ và riêng tư các tiện nghi như phòng ngủ, bếp, phòng tắm và ra vào tự do. Trong khi đó, với Private Room và là Shared Room, khách thuê sẽ phải dùng chung một số tiện nghi với chủ nhà hoặc các khách thuê khác hay thậm chí là chia sẻ giường tầng (Share Room). Sự chênh lệch mức giá này nhìn chung không có sự thay đổi nếu xét theo 5 khu vực.

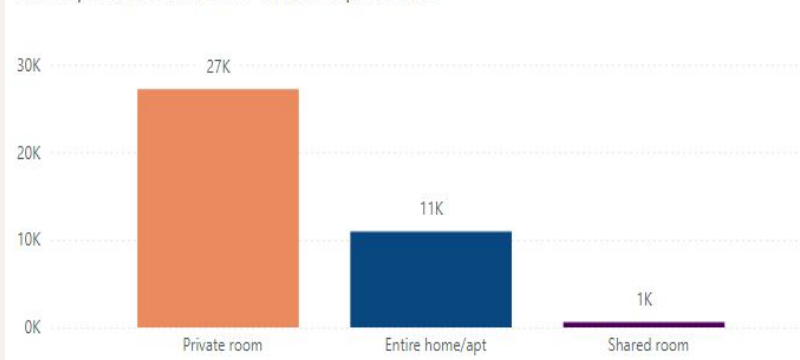
# 2.4

**Phân tích theo  
loại phòng**

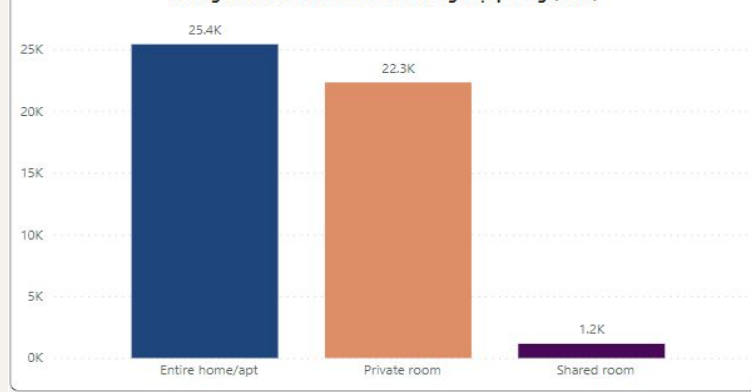


# Phân tích theo loại phòng

SỐ LƯỢT ĐÁNH GIÁ THEO TỪNG LOẠI PHÒNG



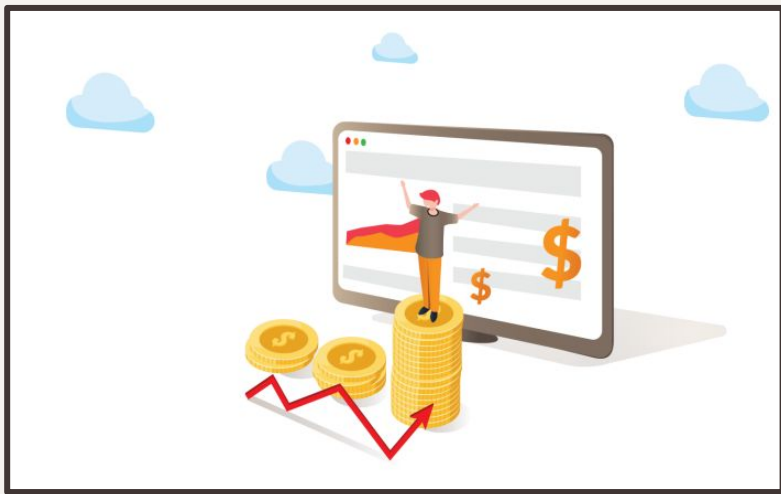
Thời gian lưu trú tối thiểu theo từng loại phòng (đêm)



Số lượt đánh giá nhiều cho loại phòng "private room" cho thấy đây là lựa chọn phổ biến của du khách. Điều này có thể xuất phát từ sự cân bằng giữa chi phí và tiện ích. Private room cung cấp sự riêng tư hơn so với shared room, trong khi vẫn tiết kiệm hơn so với entire home/apt. Private room thường phù hợp với các du khách cá nhân hoặc nhóm nhỏ, như những người đi công tác, sinh viên, hoặc du khách muốn trải nghiệm cuộc sống địa phương mà không cần toàn bộ căn hộ.

Tuy nhiên việc lưu trú lại qua đêm thì entire home/apt có phần nhỉnh hơn vì đối tượng là nhóm lớn, gia đình... Khách hàng có thể tìm kiếm sự ổn định và tiện lợi trong kế hoạch du lịch của họ, và do đó họ chọn lựa các căn hộ hoặc nhà riêng với thời gian lưu trú tối thiểu dài hơn để đảm bảo rằng họ có chỗ ở trong suốt thời gian du lịch của mình.

Qua hai bản này thì đối tượng khách hàng khác nhau sẽ có nhu cầu khác nhau, giúp cho host (chủ hộ) có cái nhìn tổng quan hơn để đầu tư loại hình lưu trú phù hợp hơn cho khu vực của mình dựa trên đối tượng khách hàng và nhu cầu của họ.

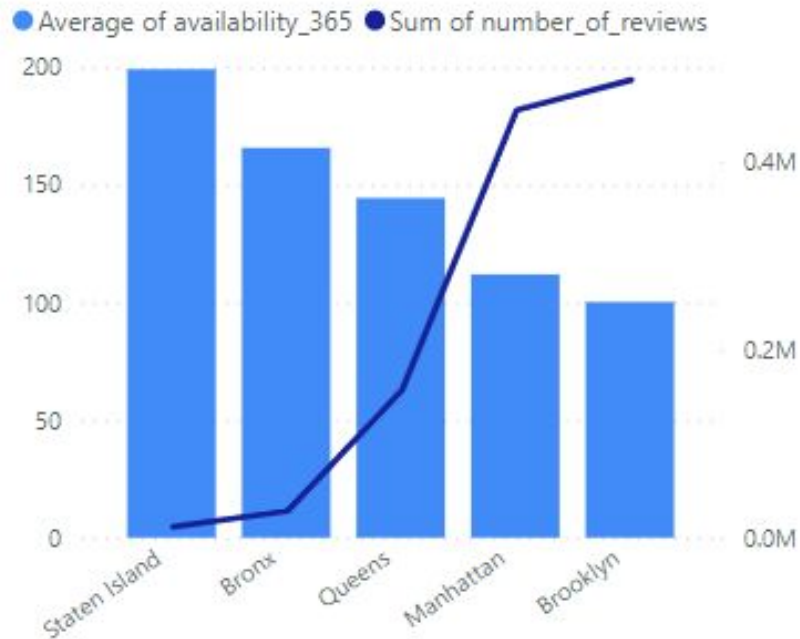


# 2.5

**Đánh giá nhu cầu  
khách hàng**

## Phân tích review & availability/365 ngày - Đánh giá nhu cầu

### Trung bình số phòng có thể đặt và tổng số lượt đánh giá theo từng khu vực



Ở Manhattan và Brooklyn có mức độ trống phòng thấp nhất, và cũng là 2 nơi có nhiều khách lưu trú nhất

=> Sự kết hợp giữa tỷ lệ trống phòng thấp và đánh giá cao từ khách hàng cho thấy tiềm năng đầu tư lớn tại Manhattan và Brooklyn. Nhà đầu tư nên cân nhắc việc đầu tư vào các dự án bất động sản hoặc các dịch vụ liên quan tại hai khu vực này để khai thác tối đa lợi ích kinh tế.

Tại ba khu vực còn lại (Staten Island, Bronx, Queens) có lượng trống phòng rất cao, nhưng lượt đánh giá lại thấp, cho thấy nhu cầu thực tế của khách hàng khá thấp ở ba khu vực này. Điều này cho thấy ba khu vực này không tiềm năng cho việc đầu tư cho thuê căn hộ vì những hạn chế này.



03

Tổng kết - Đề xuất



# Tổng kết

- Airbnb cho phép khách hàng dễ dàng tìm kiếm và đặt chỗ ở từ các căn hộ cao cấp tại các địa điểm ở New York theo loại phòng entire home/apt hoặc private room hay shared room
- Dựa trên phân bổ thì đầu tư nơi lưu trú theo các khu vực đông khách du lịch/công tác như Manhattan và Brooklyn. Tuy nhiên nếu nhu cầu đầu tư vào các khu vực khác thì có thể xem xét các vùng gần trung tâm du lịch.
- Việc chênh lệch về giá của các loại phòng không có khác biệt theo khu vực và theo vùng giúp chủ cho thuê đưa ra các mức giá phù hợp.
- Số lượng đánh giá và thời gian lưu trú phản ánh nhu cầu của người thuê giúp chủ thuê đưa ra các quyết định phù hợp về loại phòng mà mình muốn đầu tư.
- Việc Manhattan và Brooklyn có mức độ trống phòng thấp nhất và nhiều khách lưu trú nhất phản ánh lượng cung thấp và nhu cầu cao cho thấy đây là 2 khu vực nên đầu tư.

# Tài liệu

★ **Nguồn data :**

<https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data/data>

★ **Xử lý dữ liệu :**

[https://colab.research.google.com/drive/1vTqExmckanddSmZSvyKwj1QxcQBkxm\\_i?hl=vi#scrollTo=k9Y8gRJNRG98](https://colab.research.google.com/drive/1vTqExmckanddSmZSvyKwj1QxcQBkxm_i?hl=vi#scrollTo=k9Y8gRJNRG98)

★ **Power BI :**

<https://app.powerbi.com/groups/me/reports/3269397c-2c7a-4fdd-8269-f7f9e3c207be/ReportSection?experience=power-bi>

---

# Thanks

Cảm ơn bạn đã lắng nghe

---