Airbnb NYC 2019

Airbnb, as in "Air Bed and Breakfast" is a service that lets property owners rent out their spaces to travellers looking for a place to stay. Travelers can rent a space for multiple people to share, a shared space with private rooms, or the entire property for themselves.

Overview

The data set I took it from Kaggle **Airbnb** NYC, NY 2019. In this project, the data set describes the listing activity and metrics. This data file includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions. We are going to find the answer to questions like:

- Where are the most properties in NYC?
- What is the average price for an **Airbnb** room in NYC?
- Who is the host with the most property owned in NYC?

The dataset contains:

- Id: Identifier unik untuk tiap tempat sewa
- Name : Nama tempat
- Host_id : Identifier penyedia kamar/tempat
- Host_name : Nama penyedia kamar/tempat
- Neighbourhood_group : Kelompok lingkungan dari tempat tinggal yang disediakan host (penyedia kamar/tempat), merupakan pengelompokan dari neighbourhood
- Neighbourhood : Nama dari lingkungan tempat tinggal yang disediakan host
- Latitude & longitude : Garis lintang dan garis bujur dari tempat tinggal yang disediakan
- Room_type : Tipe kamar yang disediakan
- Price : Harga sewa per malam dalam dolar Amerika
- Minimum_nights : Jumlah minimum malam yang disewa untuk setiap penyewaan
- Number_of_reviews : Jumlah ulasan oleh pelanggan
- Last_review : Tanggal review terakhir oleh pelanggan
- Reviews_per_month : Rata-rata ulasan jika dibagi per bulan
- Calculated_host_listings_count : Jumlah tempat milik host yang telah didaftarkan ke Airbnb
- Availability_365 : Beberapa hari dalam setahun, tempat tersedia untuk pemesanan

Tool

The tools that I used for this analysis are SQL and Tableau for the visualization.

Analyze

- Total of the Location

```sql
--Total of the location
SELECT COUNT(DISTINCT neighbourhood) AS total_location
FROM AB_NYC_2019$
WHERE neighbourhood IS NOT NULL
```

total_location
378

As we can see, there are 378 available property locations on Airbnb NYC.

- Average Price

```sql
--Average Price
SELECT ROUND(AVG(price), 2) as average_price
FROM AB_NYC_2019$
WHERE price IS NOT NULL AND price !< 0;
```

average_price
152,29

As we can see, the average price for one nights in NYC is 152.29 USD.

- Total of the Property

```sql
--Total of the Property
SELECT COUNT(DISTINCT name) AS total_place
FROM AB_NYC_2019$
```

total_place
47564

There are 47.564 properties listed on Airbnb on NYC.

- Total Host

```sql
--Total host
SELECT COUNT(DISTINCT host_id) AS total_host
FROM AB_NYC_2019$
```

total_host
37302

There are 37.302 host who listed there property on the Airbnb to got listed.

- Total Room by the Location

```sql
--Count Room by Location
]SELECT TOP 5 neighbourhood_group, COUNT(*) as total_room
FROM AB_NYC_2019$
GROUP BY neighbourhood_group
ORDER BY total_room DESC;
```

neighbourhood_group	total_room
Manhattan	21572
Brooklyn	20018
Queens	5591
Bronx	1079
Staten Island	369

As we can see Manhattan is the most property located the among the other borough in NYC.
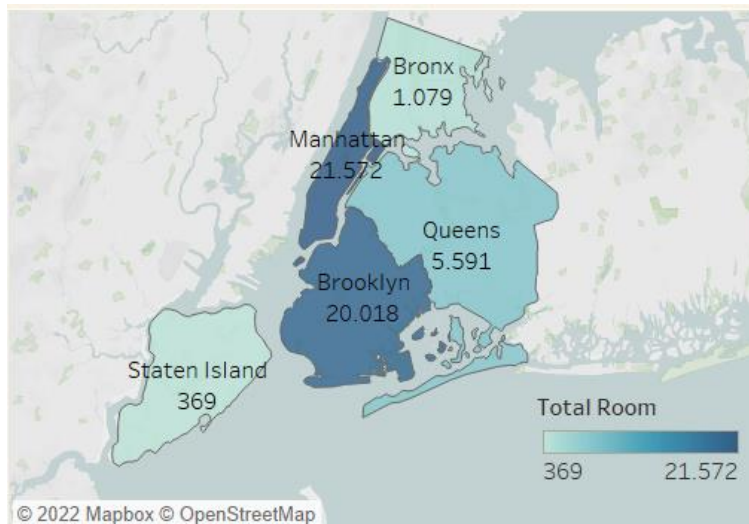

For the visualization I used map background of "Streets". The map is include five distinct values with the sum property by the location. Therefore I use colour for the categorization the location by the sum of the property.

- Total Room Type by Neighbourhood Group

```sql
-- Total Room Type by Neighbourhood Group
SELECT TOP 5 neighbourhood_group,
SUM(CASE WHEN room_type = 'Private room' THEN 1 ELSE 0 END) AS private_room,
SUM(CASE WHEN room_type = 'Entire home/apt' THEN 1 ELSE 0 END) AS entire_home_or_apt,
SUM(CASE WHEN room_type = 'Shared room' THEN 1 ELSE 0 END) AS shared_room,
COUNT(room_type) AS total_room
FROM AB_NYC_2019$
GROUP BY neighbourhood_group
ORDER BY total_room DESC;
```

neighbourhood_group	private_room	entire_home_or_apt	shared_room	total_room
Manhattan	7935	13160	477	21572
Brooklyn	10089	9517	412	20018
Queens	3327	2079	185	5591
Bronx	644	375	60	1079
Staten Island	186	174	9	369

In the Query above we already know the total property based on the location. Now I want to know the total the number of properties at a location based on the type of room provided. Manhattan is the location with the most Entire home or apt and Shared room types, while the most private room types are in Brooklyn.
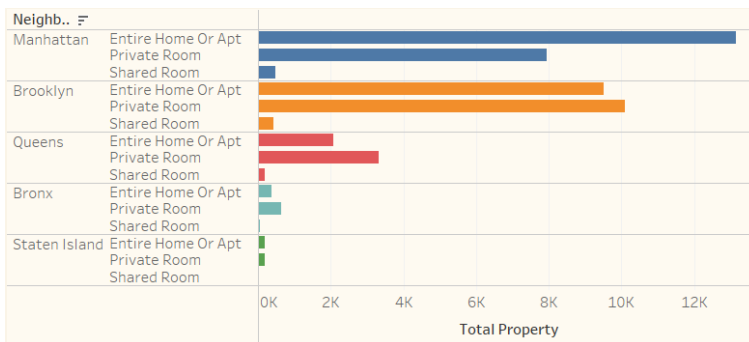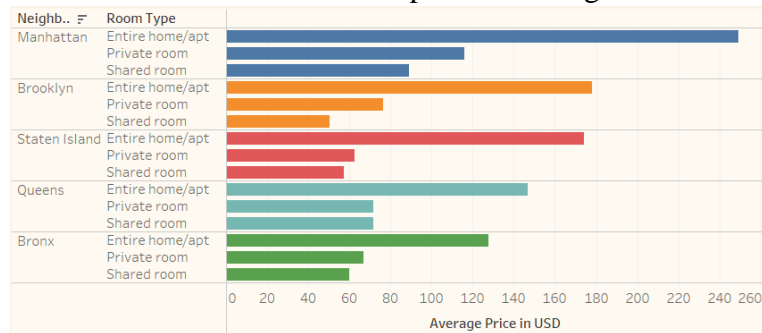

For the visualization I choose to use bar chart to make it more clear. Because there is a different location, I use colour to categorize based on the neighbourhood group

- Average Price by Room Type and Location

```sql
--Tipe kamar lokasi dan harga
SELECT neighbourhood_group, room_type, ROUND(AVG(price), 2) AS average_price
FROM AB_NYC_2019$
WHERE room_type IS NOT NULL AND room_type NOT LIKE '%[123456789%]%' AND price IS NOT NULL AND price !< 0
GROUP BY neighbourhood_group, room_type
ORDER BY average_price DESC
```

neighbourhood_group	room_type	average_price
Manhattan	Entire home/apt	249,38
Brooklyn	Entire home/apt	178,17
Staten Island	Entire home/apt	174,26
Queens	Entire home/apt	146,99
Bronx	Entire home/apt	127,79
Manhattan	Private room	116,05
Manhattan	Shared room	89,1
Brooklyn	Private room	76,5
Queens	Private room	71,86
Queens	Shared room	71,68
Bronx	Private room	66,89
Staten Island	Private room	62,49
Bronx	Shared room	59,8
Staten Island	Shared room	57,44
Brooklyn	Shared room	50,53

After we know it about the sum of the room type based on the location. Right now we dig a more deeper, with found out the average price of the room. And like the others, Manhattan is the location with the most expensive average room rates among others.

For the Visualization I choose to use bar chart to make it efficient. Because there is a different location, I use colour to categorize based on the neighbourhood group.

- Host with most Listing

```sql
--Top 5 host
SELECT TOP 5 host_id, host_name, COUNT(name) as Total
FROM AB_NYC_2019$
WHERE name IS NOT NULL AND host_name IS NOT NULL
GROUP BY host_id, host_name
ORDER BY Total DESC;
```

host_id	host_name	Total
219517861	Sonder (NYC)	327
107434423	Blueground	232
30283594	Kara	121
137358866	Kazuya	103
16098958	Jeremy & Laura	96

We can see Sonder(NYC) is the host with the most listing in NYC, Sonder(NYC) owned 327 room in NYC.
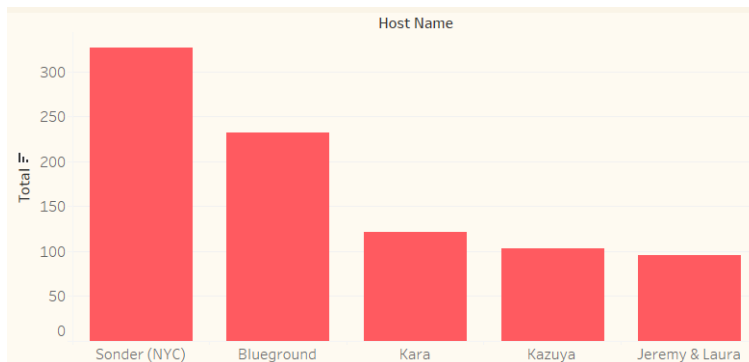
For the visualization I choose bar chart to visualize the data, because the data is the comparison from the sum of host with listing property in Airbnb NYC.

Dashboard airbnb NYC 2019
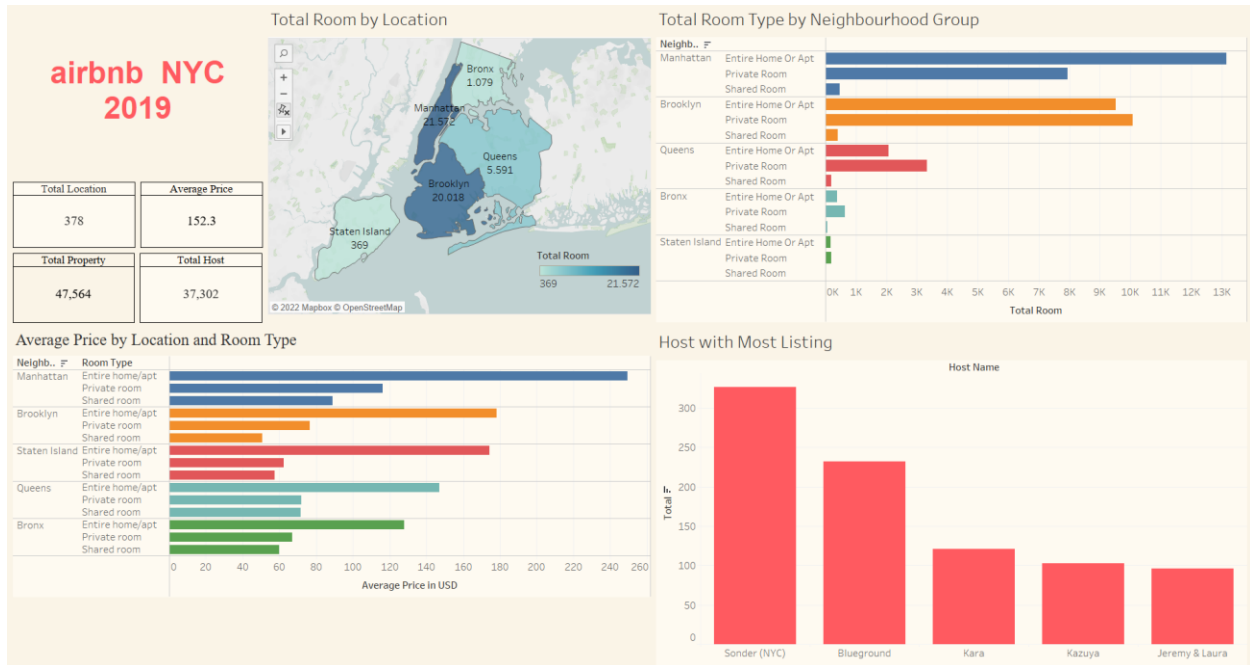
Full version of the dashboard click [here](#).