

Airbnb Data Visualization

Application Development

Group Members:

1. Syed Umer Farooq (27262)
2. Faiq-uz-Zaman (27255)
3. Hamza Mujahid (25292)

Instructor: Usman Ali

Business Model

The Airbnb business model is an example of the sharing economy, so much so that the company itself considers it a *community built on sharing*. The sharing economy industry allows people to earn from assets that are idle or underused. In this case, accommodation spaces.

Airbnb's revenue comes from two main sources:

- **Commission from Hosts:** This fee varies from 3 to 5% of the amount charged for the rent. The host can list their properties for free and, with each booking, Airbnb deducts its commission. For hosts offering experiences (events and activities), the commission is 20%
- **Fees from Guests:** Airbnb charges 0% to 20% of the booking amount as transaction fees. This charge is also made with each reservation and varies according to the courses involved. To end this up, it's worth saying that, considering Uber for short-term rentals, Airbnb is currently a multimillion-dollar company and, while there are several competitors on the market, none are even close to taking its lead.

Customer Segments

Two customer segments define the Airbnb business model. They are:

- **Hosts:** The people with spaces available for rent want to make money from that. In the app, they can include their properties under certain conditions, such as available periods, check-in and out times, and other rules. They can also accept or reject the requested reservations and evaluate travelers' details later
- **Guests:** The people who are looking for a place to stay. They can search by location, property type, price, and other filters the app offers. Book and pay through Airbnb

Proposed analytics on the Dataset

The data for this analysis is from Kaggle New York City Airbnb Open Data. The dataset describes the listing activity and metrics in NYC, NY for 2019. It includes information such as the location of the listing properties, the neighborhood of the properties, room type, price, minimum nights required, customer reviews, and availability of the listing.

The purpose of this analysis is to perform exploratory data analysis as well as data visualization to understand how different factors influence the demand for listing properties on Airbnb. Univariate analysis is conducted on the dataset to understand the properties of the variables and how they behave

independently. Understanding the variables gives a better picture to explore and analyze the dataset and conduct multivariate analysis to better understand how different variables interact and whether there is some correlation among the variables.

The following questions will be answered in the course of this analysis:

- Where most of the properties are listed and where are the busiest areas?
- What type of rooms are most popular?
- How do different areas/neighborhoods affect the listing property price and demands?
- What are the most important factors when customers choose an Airbnb property?
 - Price
 - Location
 - Room Type

About Dataset

Context

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. This dataset describes the listing activity and metrics in NYC, NY for 2019. This dataset falls into the category of clustering.

Content

This data file includes all the needed information to find out more about hosts, geographical availability, and necessary metrics to make predictions and draw conclusions.

The data is available in the form of the following:

Room type

Airbnb hosts can list entire homes/apartments, private, shared rooms, and more recently hotel rooms.

Depending on the room type and activity, a residential Airbnb listing could be more like a hotel, disruptive for neighbors, taking away housing, and illegal.

Host name

Some Airbnb hosts have multiple listings.

A host may list separate rooms in the same apartment, or multiple apartments or homes available in their entirety.

Hosts with multiple listings are more likely to be running a business, are unlikely to be living on the property, and violate most short-term rental laws designed to protect residential housing

Balance Data or Imbalanced Data:

This dataset seems to be balanced as per our knowledge. The consequences of an imbalanced dataset are very severe mean instances of one of the two classes being higher than the other, in another way, the number of observations is not the same for all the classes in a classification dataset.

Thus, the problem with training the model with an imbalanced dataset is that the model will be biased towards the majority class only. This causes a problem when we are interested in the prediction of the minority class.

Data Table

Attributes	Data Type	Missing Values			
id	int64	0			
name	object	16			
host_id	int64	0			
host_name	object	21			
neighbourhood_group	object	0			
neighbourhood	object	0			
latitude	float64	0			
longitude	float64	0			
room_type	object	0			
price	int64	0			
minimum_nights	int64	0			
number_of_reviews	int64	0			
last_reviews	object	10052			
reviews_per_months	float64	10052			
calculated_host_listings_count	int64	0			
availability_365	int64	0			
dtype: object					

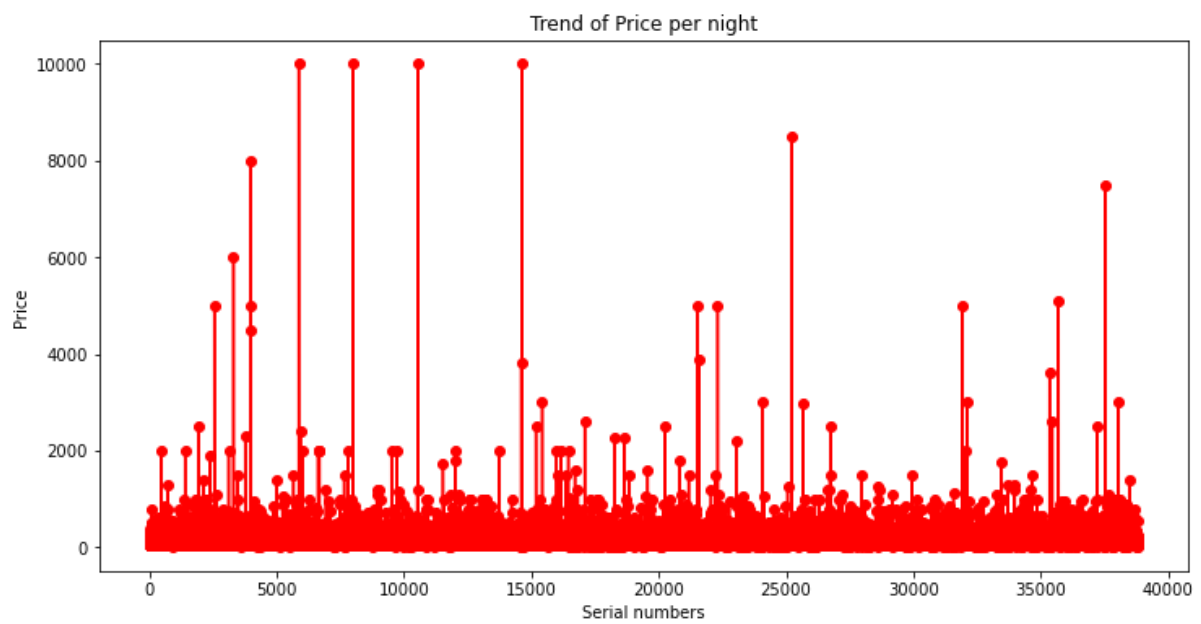
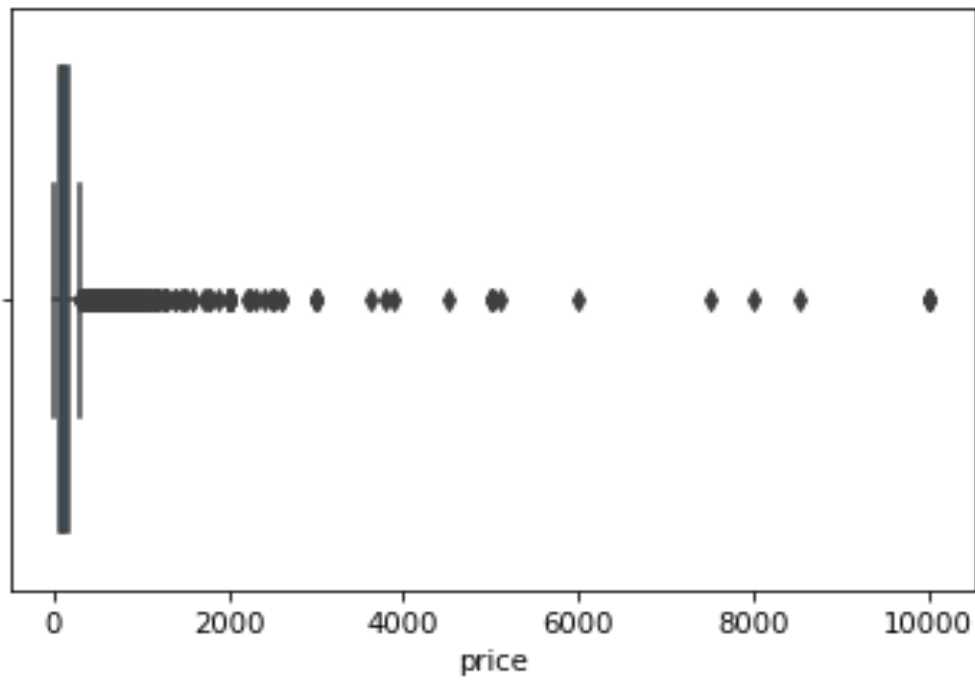
Feature Importance

For our data exploration purposes, it also would be nice to have a couple of additional features, such as positive and negative numeric (0-5 stars) reviews or 0-5 star average reviews for each listing; the addition of these features would help to determine the best-reviewed hosts for NYC along with 'number_of_review' column that is provided. Overall, we discovered a very good number of interesting features and tried out best to explain each step of the process. This data analytics is very much on a higher level on the Airbnb Data/Machine Learning team for better business decisions according to our domain knowledge.

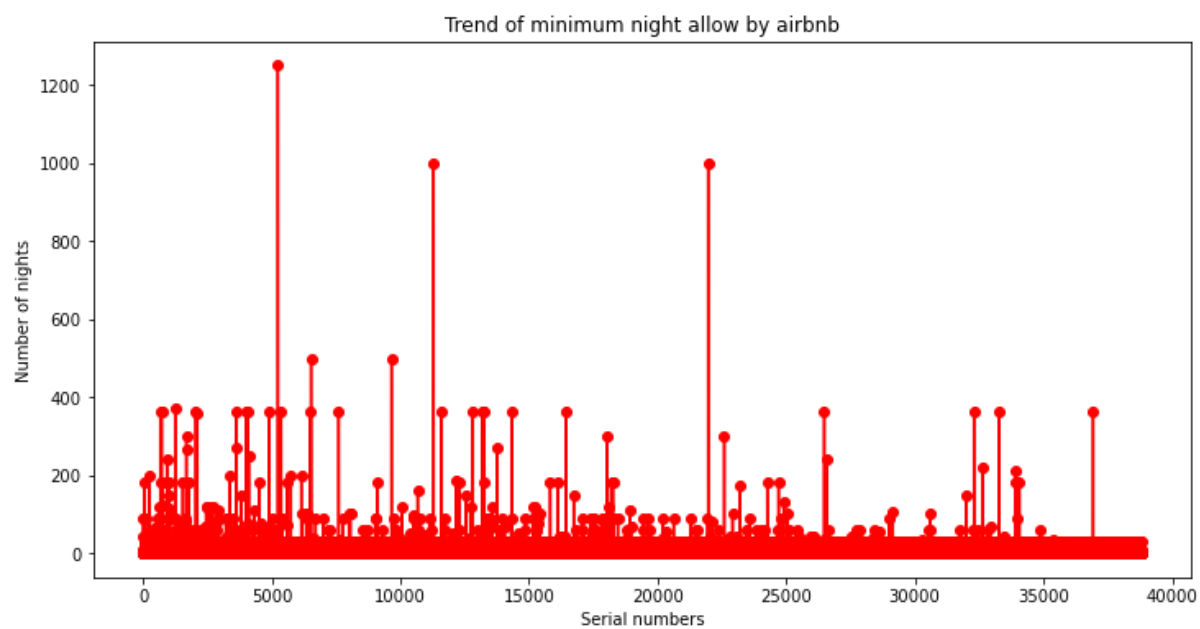
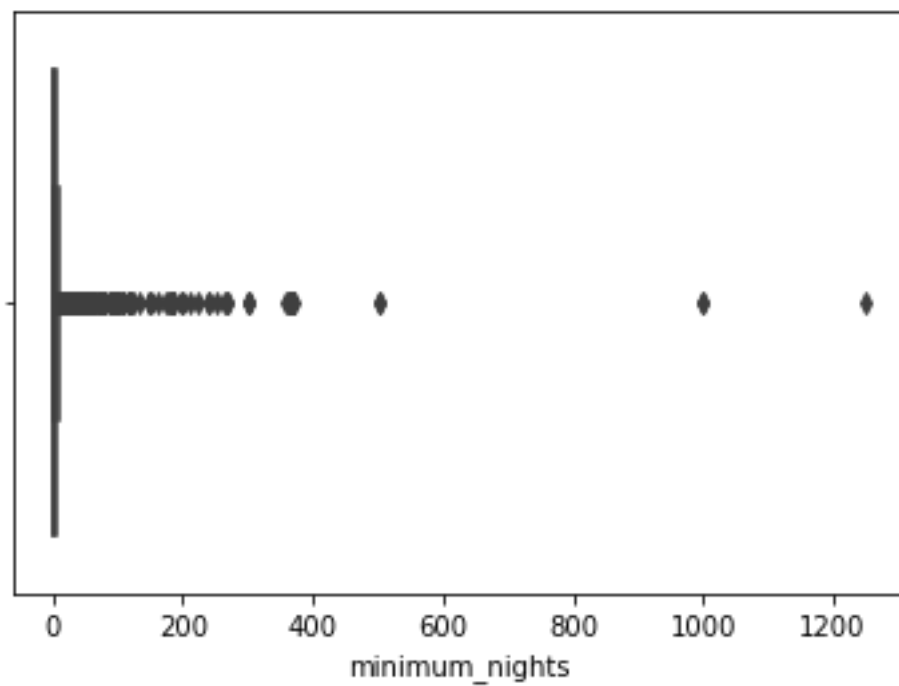
Graphs & Boxplots

Univariate Analysis

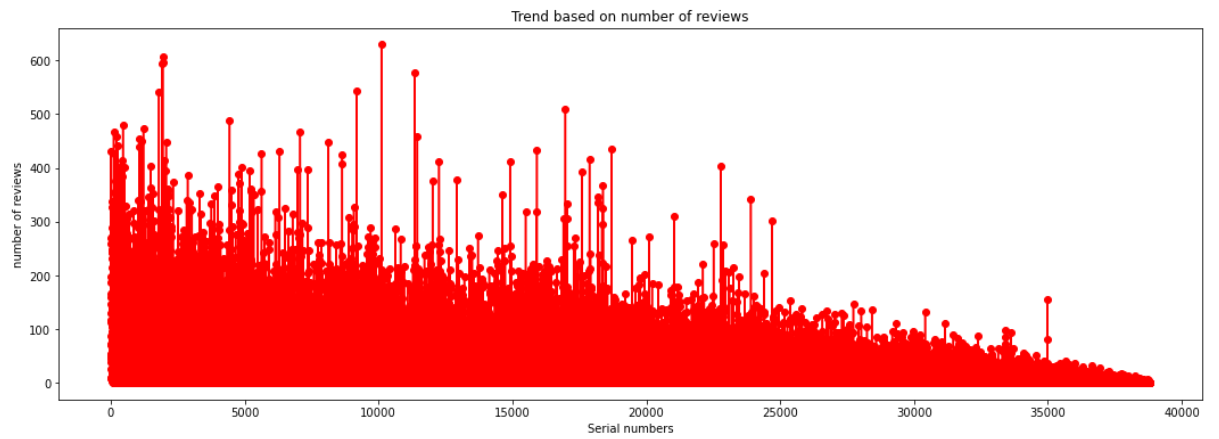
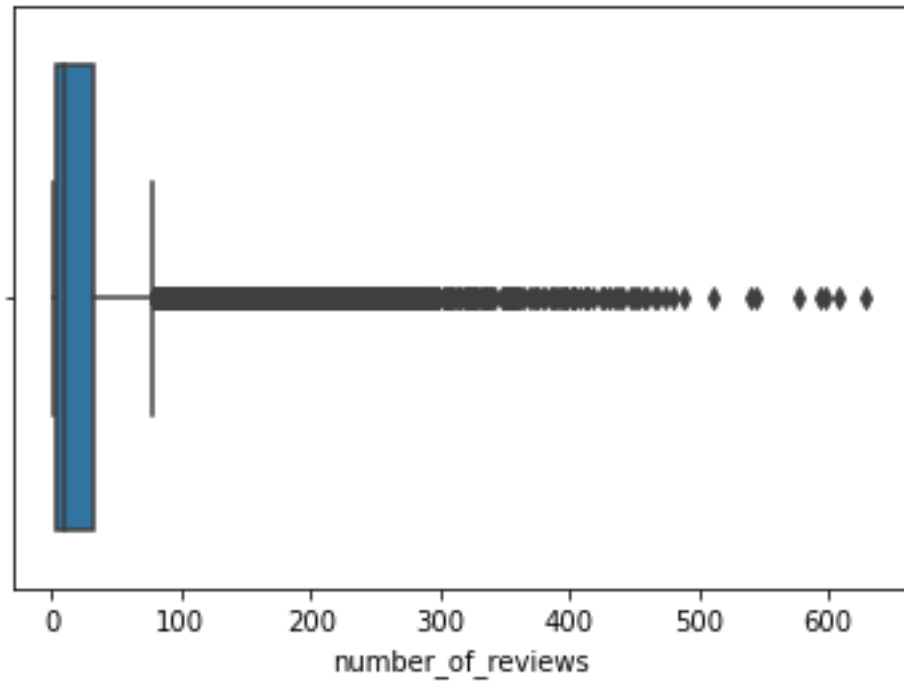
1. Price:



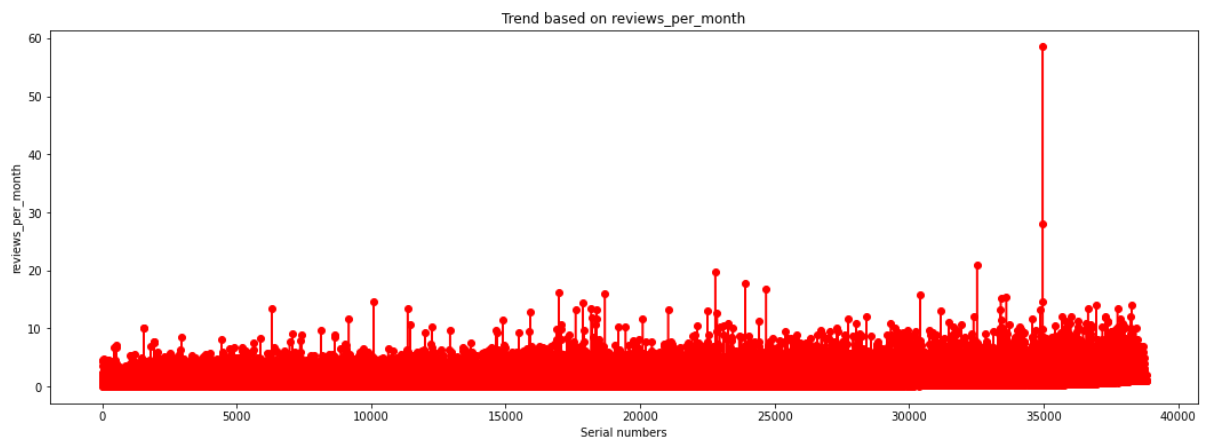
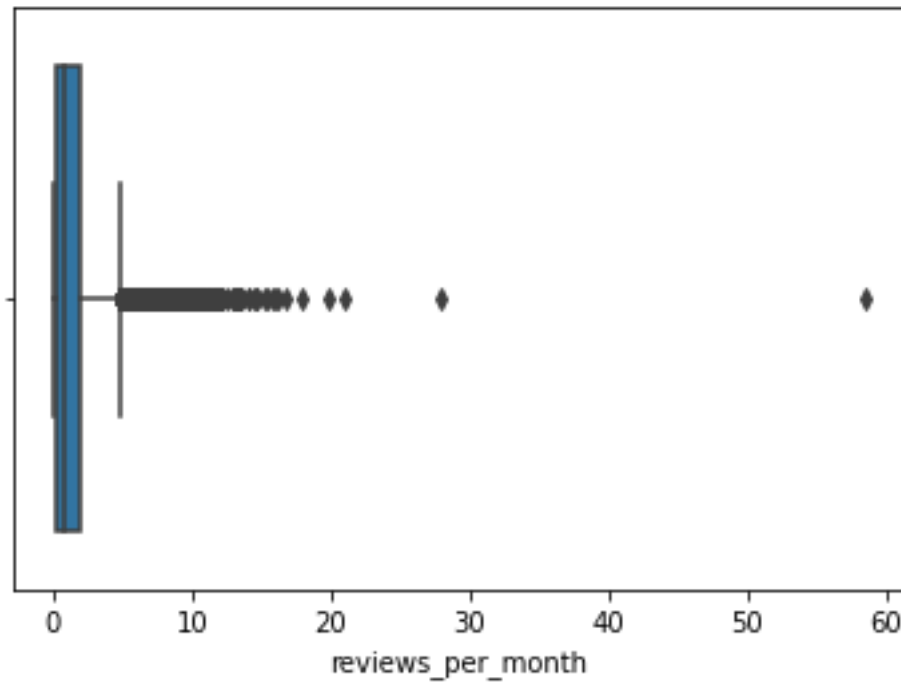
2. Minimum Nights



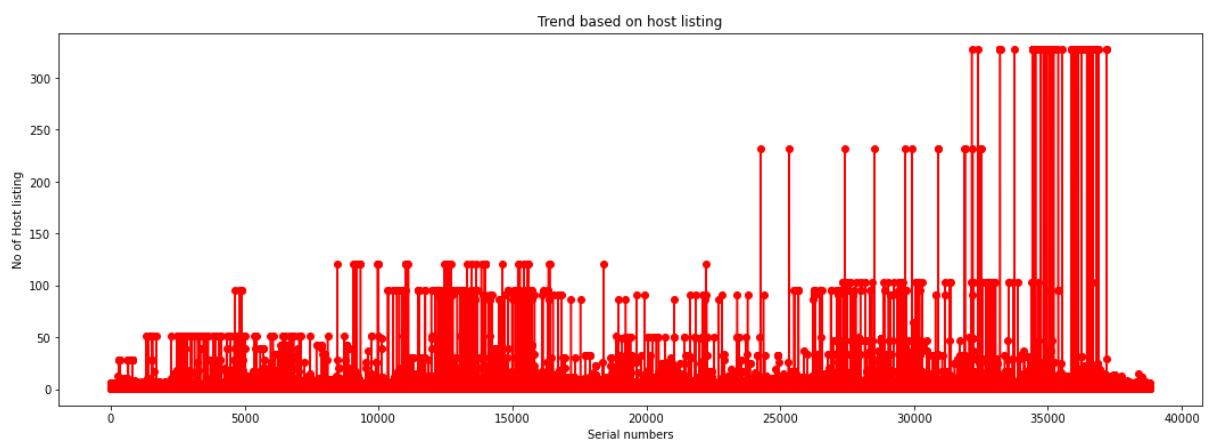
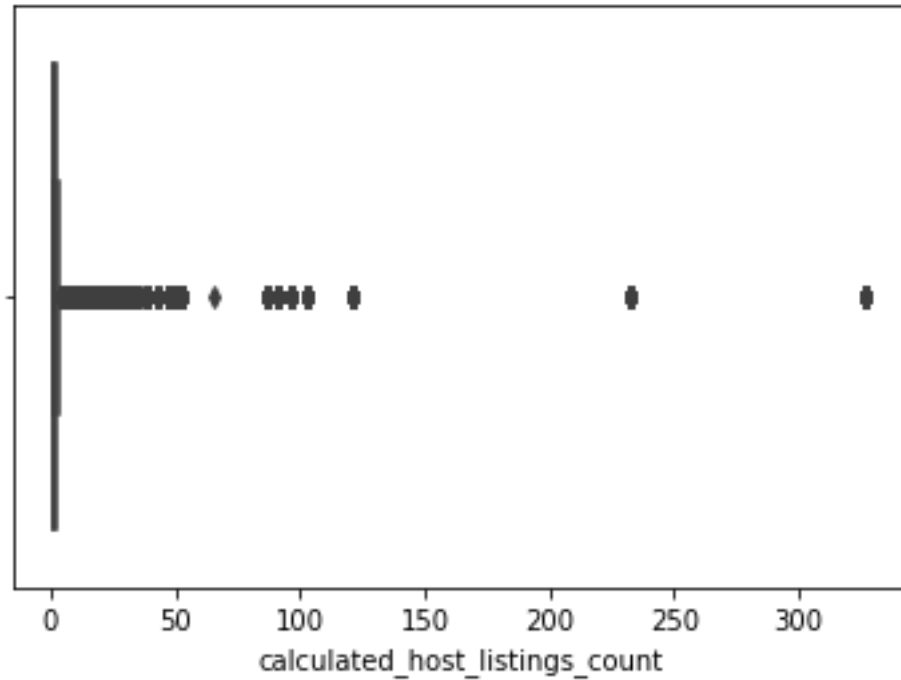
3. Reviews



4. Reviews Per Month

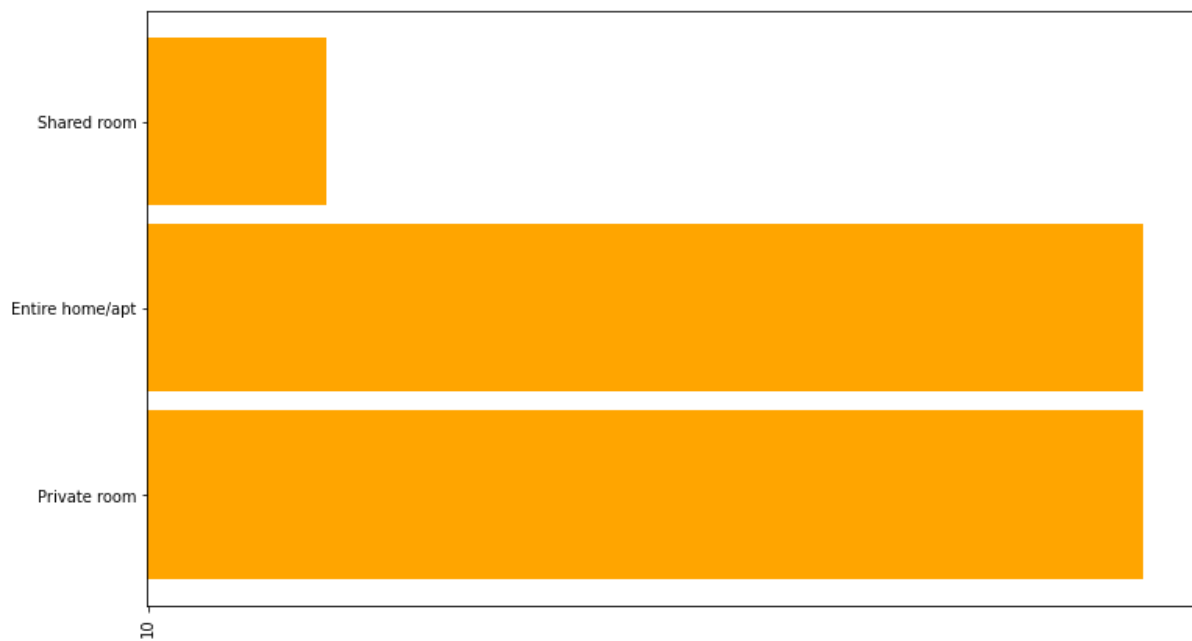


5. Hosts Listing



Bivariate Analysis

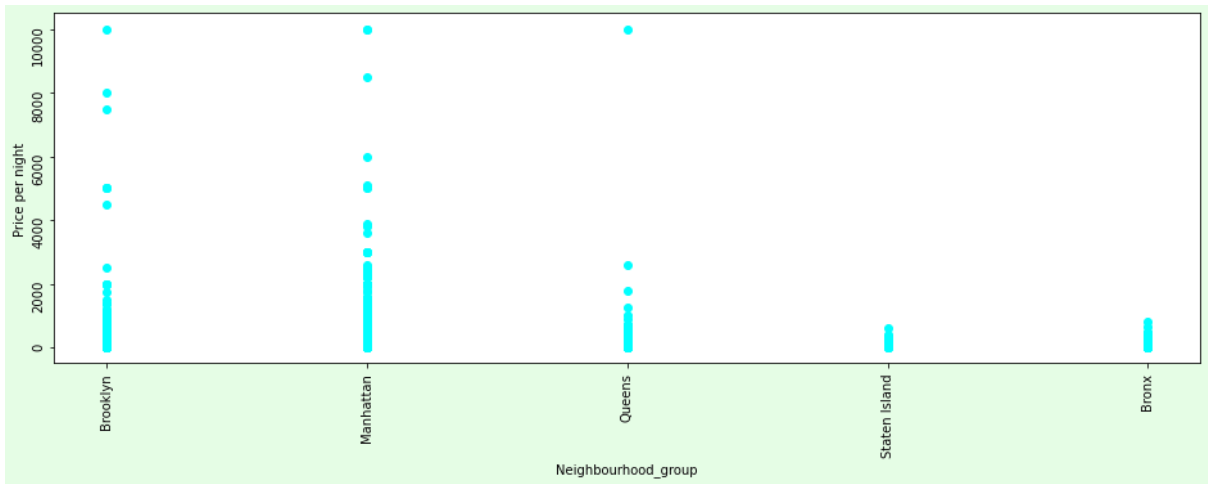
1. Room Type & Price:



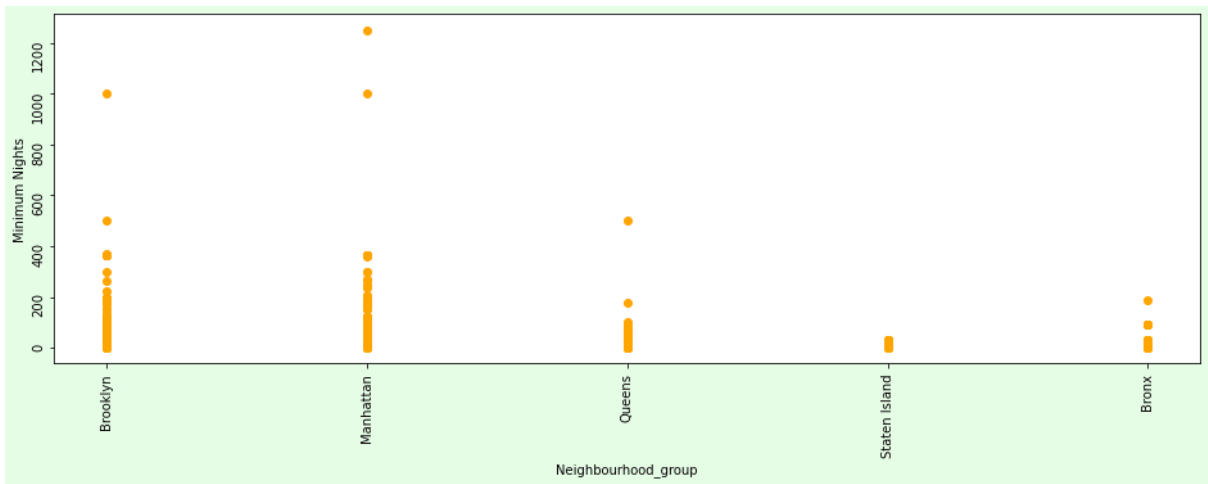
2. Room Type & Minimum Nights:



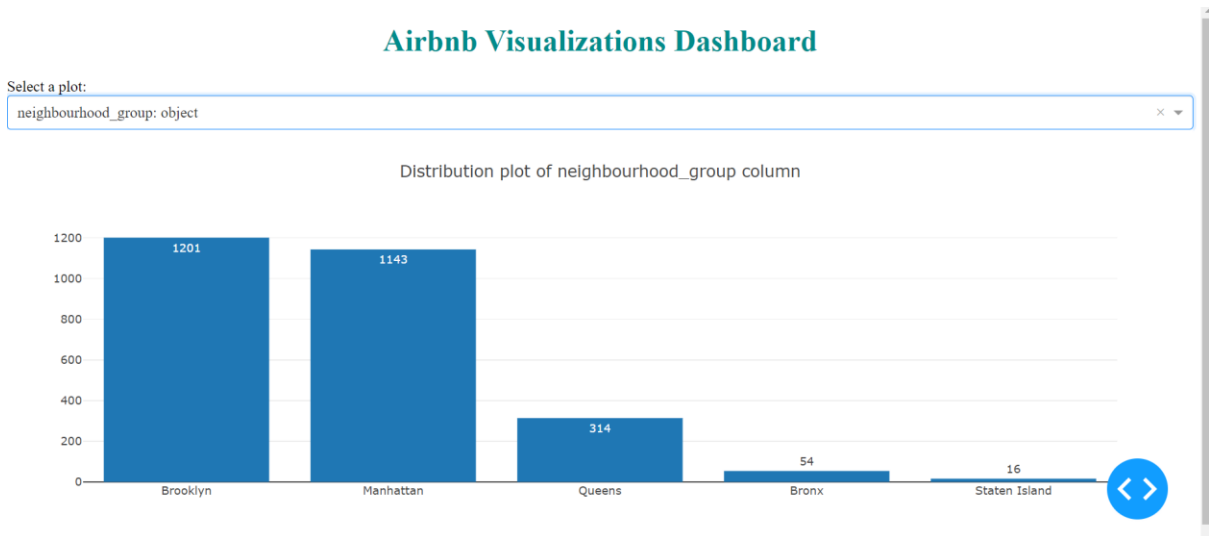
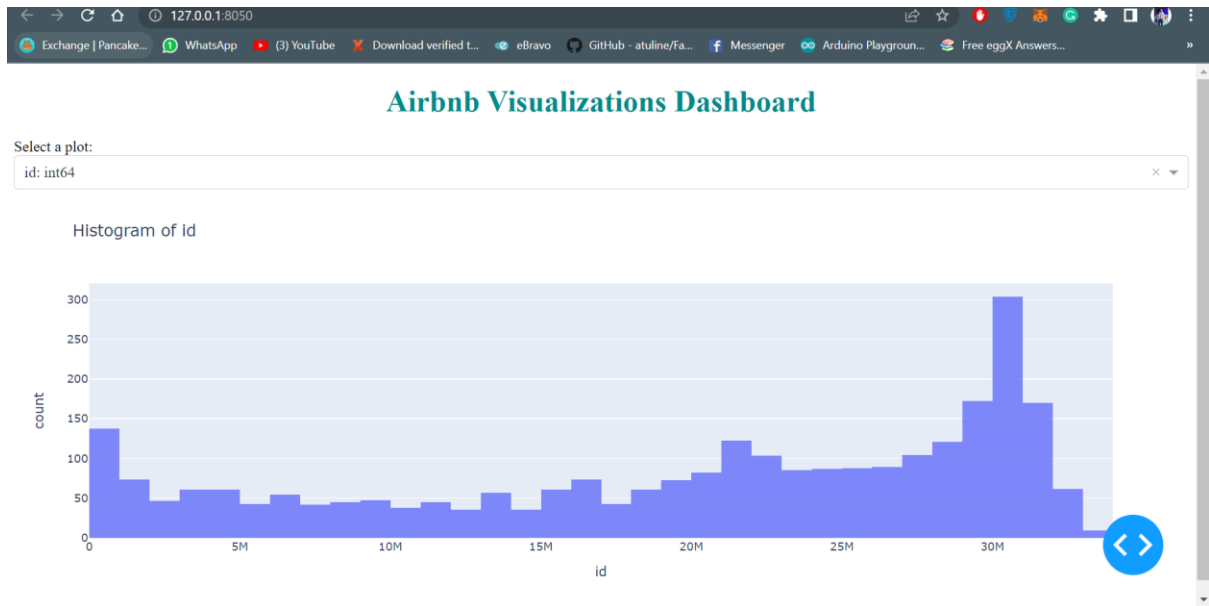
3. Neighbourhood Group & Price



4. Neighbourhood Group & Minimum Nights:



Application Dashboard:

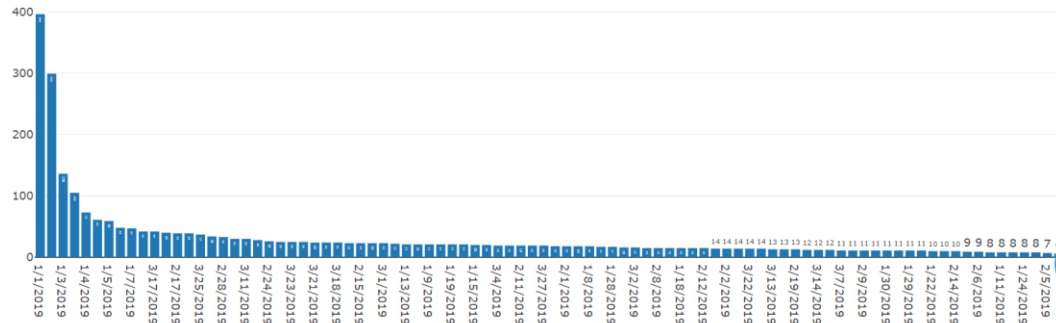


Airbnb Visualizations Dashboard

Select a plot:

last_review: object

Distribution plot of last_review column



Airbnb Visualizations Dashboard

Select a plot:

room_type: object
name: object
host_id: int64
host_name: object
neighbourhood_group: object
neighbourhood: object
room_type: object



Airbnb Visualizations Dashboard

Room Prices and Reviews

