

# EDA On Airbnb of NYC 2109

Harish Kollana

Data science trainee,

AlmaBetter, Bangalore.

## Abstract:

The Airbnb platforms offer you someone's home as a place to stay instead of a hotel and there are millions of listings on their platform. These millions of listings generate a lot of data- a data that can be analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behavior and performance on the platform, guiding marketing initiatives, implementation of innovative additional services, and much more. We were provided with one such already classified label in our data set containing 49,000 observations with 16 columns.

Our experiments can help understand what are the underlying trends in the data by Explorative data analysis.

**Keywords:** explorative data analysis, pricing.

## 1. Problem Statement

Data provided by a stay booking aggregator service Airbnb. Their customers can book their stay from the Airbnb platform listings as per their requirements from anywhere in

the cities they operate in. They, in turn, search for stays from various service providers and provide the best option to their clients across available options.

The main objective is to do an explorative data analysis to find some insights.

**id:** It is the Listing id and every listing on Airbnb has some listing id with some serial. it is not a null and unique value

**name:** It is the name of the listing in either term we can refer to as the property name.

**host\_id:** it is a number value of a host and a single host can contain multiple host ids as they list on multiple neighborhoods

**host\_name:** it is the name of the host who hosts the listing

**neighbourhood\_group:** it is a group of neighborhoods

**neighborhood:** these are neighborhoods in another way we can think of them as cities

**latitude:** these are the latitude coordinates of the location of the property.

**longitude:** these are the longitude coordinates of the location of the property.

**room\_type:** it contains different types of rooms available across listings

**price:** it is the price of the listing to stay

**minimum\_nights:** min no of days were spent on that listing or property

**number\_of\_reviews:** how many reviews do the listing receive from customers

**last\_review:** when did the customer leave a review for that listing the last date of the review

**reviews\_per\_month:** how many reviews received per month

**calculated\_host\_listings\_count:** no of listings submitted by the hosts. one host can submit multiple listings

**availability\_365:** how many days is a listing available through 365 days

## 2. Introduction

Our goal here is to find insights on these listings, which could help Airbnb take its business decision proactively.

## 3. Pricing:

The pricing of the listing or the property is based on the living cost of that neighborhood. The cost changes from neighborhood to neighborhood.

## 4. Room Type:

It is a customer's preference to choose rooms when he wants to stay in a particular location and he will decide which type of room he is comfortable with and can book that stay.

## 5. Minimum Nights Spent:

It makes a huge impact with pricing as the price increases when the no of nights the customers are spending. so the customers might want to stay for longer or less. it is completely based on them

## 6. Steps involved:

### Exploratory Data Analysis

After loading the dataset we performed this method by comparing our target variable that is Price with other independent variables. This process helped us figure out various aspects and relationships among the target and the independent variables. It gave us a better idea of which feature behaves in which manner compared to the target variable.

## Null values Treatment

Our dataset contains a large number of null values which might tend to disturb our analysis. hence we replace with the unknown for name, no\_name for host\_name, median values for price because the data is skewed and 0 for reviews per month.

## Outliners Treatment

we had removed outliers of price with the IQR method and found the distribution of pricing on independent variables without outliers as it will affect the distribution.

## Data Analysis:

### 1. Univariate Analysis:

In this one variable is compared with another to find the insights

### 2. Multivariate Analysis:

In this, more than one variable is used for comparison.

## 7. Conclusion:

That's it! We reached the end of our exercise.

1. Starting with loading the data so far we have done EDA, null values

treatment, outliers treatment, and data analysis.

2. Hillside Hotel is found to have listed more listings in entire NYC, followed by Home away from Home.
3. Host id 219517861 is found to have listed more listings in the entire NYC with 327, followed by 107434423.
4. Michael is found to have listed more listings in entire NYC with 417, followed by David.
5. Manhattan is found to have a high-listed neighborhood group with more listings in the entire NYC, followed by Brooklyn.
6. Williamsburg is found to have listed more listed neighborhoods in the entire NYC with 3920, followed by Bedford-Stuyvesant.
7. Entire home/Apt is found to have high listed room type with more listings in entire NYC, followed by Private room.
8. Manhattan has more listed properties with Entire home/apt around 27% of total listed properties followed by Brooklyn with around 19.6%.
9. Private rooms are more in Brooklyn as in 20.7% of the total listed properties followed by Manhattan with 16.3% of them. While 6.9% of private rooms are from Queens.

10. Manhattan is the most focused place for the host to do their business
11. Manhattan has the highest(the costliest) average price ranging to ~150dollars/day followed by Brooklyn
12. Queens and Staten Island have the same average price/day, while Bronx comes out to be the cheapest neighborhood group in NYC.
13. Brooklyn, Manhattan& Queens are the highest renting neighborhood groups
14. Brooklyn, Manhattan, Brooklyn & queens are the lowest renting neighborhood groups
15. Upper west side, Greenpoint & Astoria are the highest renting neighborhoods
16. There are 15 neighborhoods where the lowest cost is 10 dollars
17. Luxury 1 bedroom apt. -stunning Manhattan views, Furnished room in Astoria apartment & 1-BR Lincoln Center are the highest renting a property with 10,000 dollars
18. people mostly prefer living in an entire home/apt on an average of more than 8 nights followed by guests who stayed in a shared room where the average stay is 6-7 nights.
19. Manhattan is leading in all three room types on average price across NYC
20. Queens is top of the list followed by Brooklyn. Looks like all the top 3,4 & 5 costliest listings belongs to

- Manhattan itself(being the costliest place of living in NYC)
21. Among the top neighborhoods in each neighborhood group, the top 2 of them namely: Fort Wadsworth & Tribeca, origins from Staten Island & Manhattan.
  22. Looks like the top 2 costliest listings: Luxury 1 bedroom apt. -stunning Manhattan views & 1-BR Lincoln Center were from Brooklyn and Manhattan, hosted by Erin and Jelena respectively.
  23. Looks like a property/listing with Entire home/apt as room\_type wins the show at NYC followed by private rooms.
  24. Manhattan has the highest price for room types with Entire home/apt ranging to nearly 240 USD/night, followed by Private room with 110 USD/night. And it's obvious being the most expensive place to live in!
  25. Enjoy great views of the City in our Deluxe Room! with 58 reviews/month followed by Great Room in the heart of Times Square! with a total of 27 reviews/month. The top 2 are from Manhattan so we can decide the best-suited place for a comfortable stay but the prices are high to afford.
  26. Private rooms received the most no of reviews/month whereas Manhattan had the highest reviews received for Private rooms with more than 50 reviews/month, followed by Manhattan in the chase.

27. Manhattan & Queens got the most no of reviews for Entire home/apt room type.
28. Sonder(NYC) leads the most expensive neighborhood group in NYC with around 106929 properties listed followed by Blue ground and Kara.
29. Staten Island seems to be more available throughout the year to more than 300 days. On average, these listings are available for around 210 days every year followed by the Bronx where every listing is available for 150 on average every year.

#### References-

Analytics Vidhya

