

Project Information

The data source is from inside Airbnb; they are activists who gather information from the Airbnb websites. The main objective for them *“Provide data that quantifies the impact of short-term rentals on housing and residential communities; and also provides a platform to support advocacy for policies to protect our cities from the impacts of short-term rentals.”*

I have chosen this data set because it would be interesting to know more regarding what causes rent prices to hike and whether there are exciting factors that I can see.

Data profile

All the changes that have been made on the data set to provide cleaner data.

Column	Reason for change	resolution
Name	Mix data type	From object to string
Host_name	Mix data type	From object to string
Last_review	Mix data type	Object to datetime64
Reviews_per_month	Null values	Change all null values to 0
Last_review	Null Values	Change all null values with the mean of this column
Price	0 as the price	Change all the 0 price with the mean of price column

Descriptive Analysis

	latitude	longitude	price	minimum_nights	number_of_reviews	reviews_per_month	calculated_host_listings_count	availability_365
count	48895.000000	48895.000000	48895.000000	48895.000000	48895.000000	48895.000000	48895.000000	48895.000000
mean	40.728949	-73.952170	152.755045	7.029962	23.274466	1.090910	7.143982	112.781327
std	0.054530	0.046157	240.143242	20.510550	44.550582	1.597283	32.952519	131.622289
min	40.499790	-74.244420	10.000000	1.000000	0.000000	0.000000	1.000000	0.000000
25%	40.690100	-73.983070	69.000000	1.000000	1.000000	0.040000	1.000000	0.000000
50%	40.723070	-73.955680	106.000000	3.000000	5.000000	0.370000	1.000000	45.000000
75%	40.763115	-73.936275	175.000000	5.000000	24.000000	1.580000	2.000000	227.000000
max	40.913060	-73.712990	10000.000000	1250.000000	629.000000	58.500000	327.000000	365.000000

The descriptive analysis provides information of all the variables where it would provide us a certain information like popular location listing, price, and minimum nights. The descriptive analysis also helps us provide information on where there may be any dirty columns.

Limitations and Ethics

1. Since the data that was provided by the Airbnb website, that could be a sample bias where Airbnb does not release all New York listings and only a certain listing.
2. The data was gathered during that particular time, which means many listing can be deleted and added, resulting in inaccurate data.
3. The address of the actual listings have been anonymized by Airbnb, which is good in terms of privacy, but this could cause an issue in the analysis since the data provided is not accurate.

There is no ethics issue that I can see; Airbnb has worked to the best of its ability to protect its customer's data.

Questions:

1. Which location would be the most popular? Is there a reason for it?
2. How many houses and apartments are being rented out frequently to tourists and not to long-term residents?
3. How many houses and apartments are being rented out for people with long-term stays?
4. How much are hosts making from renting to tourists compared to long-term rentals?
5. What is the most popular price that provides the best revenue?
6. Does the number of reviews have an impact on the revenue?
7. Which room type is the most popular, and which room type generates revenue?
8. Is there any specific host that stands out in generating revenue, and why?