

AIRBNB, U.S. DATA - 2020

11/14/2022

KAYLA MARTINEZ



Executive Summary

Airbnb is an online marketplace that connects people who want to rent out their homes to people who are looking for accommodations in specific locals. They make a stress-free environment by allowing homeowners to earn additional income from their property while allowing travelers to stay in cheaper and homier locations than hotels. In this analysis we will explore the geographical trends and look at a price prediction model that can allow hosts to have a better understanding of what income can be made from their property.

Data Source

[U.S. Airbnb Open Data](#)

Limitations and Ethics

- Difficult to predict year-to-year trends because the data set is from 2020
- I have removed personal identifying information (PII), like 'host_name'

Data Cleaning and Data Consistency Checks

- Removed PII columns
- Adjusting data types
- Checking for duplicates: None found
- Removing outliers that didn't fit the column

Data Profile

- Number of rows and columns in the cleaned dataset: 22,579 x 16

Column Details:

<i>Column</i>	<i>Column Description</i>	<i>Data Type</i>	<i>Time Variant</i>
<i>Id</i>	ID for each listing	qualitative, ordinal	No
<i>Name</i>	Name of listing on Airbnb	quantitative, discrete	No
<i>Host_id</i>	ID for each host	qualitative, ordinal	No
<i>Neighbourhood_group</i>	District or neighbourhood	quantitative, discrete	No
<i>Neighbourhood</i>	ID for each neighbourhood	qualitative, ordinal	No
<i>Latitude</i>	Geographic coordinates of the Airbnb	quantitative, continuous	No
<i>Longitude</i>	Geographic coordinates of the Airbnb	quantitative, continuous	No
<i>Room_type</i>	Defines the style of lodging for rent	quantitative, discrete	No
<i>Price</i>	Defines the price for each listing	qualitative, ordinal	no
<i>Minimum_nights</i>	Minimum nights required for booking	qualitative, ordinal	Yes
<i>Number_of_reviews</i>	Number of reviews on Airbnb	qualitative, ordinal	Yes
<i>Last_review</i>	When the last review for the listing was submitted	qualitative, ordinal	Yes
<i>Reviews_per_month</i>	Averages the number of reviews received each month	qualitative, ordinal	Yes
<i>Calculated_host_listing_count</i>	Number of listings a host has	qualitative, ordinal	No
<i>Availability_365</i>	Availability through the year	qualitative, ordinal	Yes
<i>City</i>	Name of City for listing	quantitative, discrete	No

Questions to explore:

- Can we predict the price of a house in different regions?
- Can we estimate the popularity of a listing based on given features?
- What can we learn about different regions from the data?