

1. Describe the dataset

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

<https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>

2. Why did you choose it?

I recently stayed at an AirBnB in PA and selecting the reservation was a lot of work. Looking through pictures, reviews, waiting for approval, a lot of work. Kaggle has a lot of really interesting datasets but once i saw one for NYC AirBnbs it intrigued me to see if I could find any cool patterns

3. What question are you hoping to answer by interrogating it?

I'd like to see what rooms are the most popular?, how important reviews are to the AirBnB?, Where in NY is the most popular place to stay, Where is the cheapest?

4. What type of file is it?

csv

5. How is the data structured?

Comma separated values 16 columns

6. How many cases(records) are there?

49080 total using a wc -l (subtract 1 for the first line columns)

7. How many features (dimensions) are there?

16

8. Are the features text or numbers?

Mix of Both

9. Quantitative or qualitative?

Mix of Both

10. Is there a common structure?

Yes

11. Is it a regular table of columns and rows?

Yes

12. Is there a header or footer portion that differs from the core data?

No, other than the 1st line of columns

13. Are there missing values?

Yes

14. What do you need to extract from it?

Price, location, room_type, reviews,

15. Is this a one off analysis or will you do it again with similar data?

I could see myself doing this analysis again on similar data

16. Will this data change in time?

Yes

17. What is the provenance of the data?

<http://insideairbnb.com/>

18. What is the reliability of the data?

Rated one of the best datasets for usability on Kaggle

19. Who paid for the data to get collected?

Insideairbnb creator Murray Cox

20. Does the organization who paid for the data collection have a vested interest?

"Murray Cox, an independent digital storyteller, community activist and technologist, conceived the project, compiled and analyzed the data and built the site."

21. What data do you need to extract?

Price, location, room_type, reviews,

22. How do you want to format the data storage for your use?

The data is already formatted okay for usage. I will need to do a little bit of wrangling using grep and some simple statistics like averages using bc