

AirYourBnB

Anthony Rezzonico, Bruce Darcy, Daniel Rhoades

Dataset:

We will be working with an Airbnb public data set that is hosted from Inside Airbnb. The data set that we were able to attain is approximately ~1.9 GBs in size and is in a CSV format. Inside Airbnb collected this data for non-commercial use and was built using all of the public fields that Airbnb hosts. Some of the attributes in this set include:

Country, City, type of property, the number of people the listing can accommodate for, they types of amenities that are provided at the listing, the unit price, along with plenty of other fields.

Problem Characterization:

Different countries have different standards of living as a result of how wealth is distributed. With this distribution, different regions can view certain amenities as essential for living conditions. By analyzing the data of the listed amenities and grouping the trends at the country and city level, what can be derived about the standard of living for those regions? The goal of this problem is to see what are the common amenities and what do those amenities say about the socio economic status of the particular region. While amenities may provide a good baseline for some of these traits, other attributes in the data set may be used to derive the standard of living for the region.

Currently Published Work:

There are multiple existing studies that reference this dataset provided by Inside Airbnb. After reviewing existing studies, most of the publications cover problems in a very broad analysis and fail to derive further meaning from the data.

Some of the questions answered in currently published work include: What can one say about the positive and negative tourists reviews about their hosts? What are the main factors that influence Airbnb daily renting price? How well can one predict the Airbnb daily renting price based on the data from different cities? These questions covered three cities: Seattle, Boston and Copenhagen (Lakubovskyi, 2018). Another analysis of the dataset covered New York City, asking questions such as: How popular has airbnb become? Which area is the best? Which area is the most expensive? (Gupta, 2019).

There are multiple other studies using this dataset asking similar questions on small subsets of the data. Other references that utilize the Inside Airbnb that were not

mentioned in this section answer very similar questions. Our question provides a more of an in depth analysis and has not been answered yet.

Proposed Analytic Tasks:

In order to assess the socio economic status of the communities, deriving the price estimate for the commonly found amenities is necessary for further data analysis and comparison. By grouping listings together based upon the location of the listing and the type of housing (apartment, house, etc.), find the average amenities available along with the price for one person to rent out the space. Grouping the entries based upon the housing type is important, because the style of housing can contribute a good amount to how much the posted listing is going up for. By factoring out the pricing for the types of housing, a price can be estimated for the average amenity bundle. With our analysis we will make the following inferences. If the pricing for the average amenity bundle is high, then we can say that the amenities are high in demand but are not vital to living at the location. If the pricing for the average amenity bundle is relatively low, then we can say that these amenities are either expected, necessary for daily functions, or are low in value. It is with these low priced amenity bundles where the socio economic status of the community can be derived.

Effectiveness of the Solution:

To judge the effectiveness of our outcome we will compare it to multiple existing indexes. One such index is Numbeo, a crowd sourced data collection and analysis platform for cost of living. More specifically Numbeo offers a rent index on the country level for the same timeframe as our data. If our solution correlates well with Numbeo's we will have a strong indicator that airbnb listings and their amenities correlate well with each country's cost of living and more specifically the rent index for that country. We will be able to judge the effectiveness of our solution based upon how well our data fits Numbeo's global rent index (Numbeo, 2019).

We will also compare our results regarding standard of living to the Human Development Index created and estimated by the United Nations. If we do a decent job, our results should generally correlate with those indices. The Human Development Index includes a lot of soft measurements such as freedom of expression therefore it will not be a perfect comparison but will be a useful index to see how our solution holds up.

References

- Cox, M. (2017, March 1). The Face Of Airbnb, New York City - Inside Airbnb. Adding data to the debate. Retrieved April 7, 2019, from <http://insideairbnb.com/face-of-airbnb-nyc/>
- Gupta, S. (2019, January 04). Airbnb Rental Listings Dataset Mining. Retrieved April 7, 2019, from <https://towardsdatascience.com/airbnb-rental-listings-dataset-mining-f972ed08ddece>
- Lakubovskiy, D. (2018, October 01). Digging into Airbnb data: Reviews sentiments, superhosts, and prices prediction (part1). Retrieved April 7, 2019, from <https://towardsdatascience.com/digging-into-airbnb-data-reviews-sentiments-superhosts-and-prices-prediction-part1-6c80ccb26c6a>
- Numbeo. (n.d.). Cost of Living. Retrieved April 9, 2019, from https://www.numbeo.com/cost-of-living/rankings_by_country.jsp
- Wallace, N. (2018, February 20). Where Do Airbnb Hosts Make the Most Money? Retrieved April 7, 2019, from <https://smartasset.com/mortgage/where-do-airbnb-hosts-make-the-most-money>