Assignment: Final Term Project Paper

Name: Muley, Tushar

Date: 21 November, 2020

My Term Project was performing EDA (Explore Data Analysis) on Airbnb data I obtained from an independent site called *Inside Airbnb.* Below where the questions I thought I would be able to address.

My questions of the data:

* Do seasoned hosts make more money compared to newbie hosts?
* Do the number of review impact price?
* Which types of rentals (homes, apartments, private rooms) get most bookings?
* What elements have an impact on price?

After spending some time with the data, I learned a few of the questions I will not be able to answer. The original questions are below after the references.

The outcome of my EDA. The first question, do seasoned hosts make more money compared to newbie hosts? The short answer is maybe. While the null hypotheses cannot be ignored having a 0.231 p-value and R-Squared of zero. It does reduce the impact of the original hypothesis. There is small relationship I would not call it significant enough to impact ones ability to make money. There are assumptions that can be made. A host running a business of renting properties for example. Some appeared to be hotels or even apartment complexes. Or the host has gained enough experience to figure out the seasonality or combination of location, price and amenities to pull in renters. I would need additional data or spend more time transforming the data to answer some of the more in-depth questions.

Do the number of reviews impact price? Based on the scatter plot it appears there was good amount of impact. The impact was negative by -0.0605. I would be guessing that certain types of reviews would impact price. This is an assumption that with a little more detail of the reviews like average ratings it would be easier to figure out if overall rating was driving price.

What elements have an impact on price? This was an interesting question. I was not able to answer this question fully. I did get an answer that years of experience, location and accommodations (number of people) have an impact but they are small. The over p-value was 2.40e-33 and R-squared was 0.06. All small and not very significant. The high experience numbers might be attributed to other factors not related to this analysis.

Property type was not very consistent data to use. It appeared host are allowed to define the type of property leading to many options. Making it hard to do a comparison to each other and using price as relationship indicator. There are common types like private room in apartment.

The information from *Inside Airbnb* is one sided. You know the host and what kind of rentals property they are renting. What is missing is the other side of that coin. You do not know who is renting the property or what their needs are for renting the property. A simple example, the individual or individuals doing the renting could be a couple or a family or a single person. You also do not know the motivation for requesting the rental. Is it business travel or family vacation or something else? Other information that would be helpful would be accurate availability to figure out occupancy. Since the availability is updated by the host it is inconsistent for of measure.

I believe I struggle with trying to understand the data. While there is a lot analysis on Airbnb data all of it is based on categorical information. Some have done a lot of transformation to get the data to conform. For example, trying to find a definition of the terms was difficult. The data that can be gathered is not the same for all markets and regions. Some of the terms are different and some terms are not available in all markets. The data is free it is somewhat consistent within regions.

In conclusion I believe I can’t rule out the null hypothesis and variables that play a part in the analysis. While host make money, I was unable to prove how much or which elements attribute to a successful host. With more data variables I believe I could have answered this question.

GitHub URL: <https://github.com/chemistrusaz/GitHub/tree/master/DSC530_MMetzger_Fall2020/Week_12_Final_Project>

**References**

1. Downey A (2014), “Think Stats, 2nd Edition”, Published by O'Reilly Media, Inc., <https://learning.oreilly.com/library/view/think-stats-2nd/9781491907344/ch13.html>
2. Cox, M (October 2020), *Inside Airbnb Adding Data to the Debate*, From Inside Airbnb, <http://insideairbnb.com/about.html>

Original Questions:

* Does a host make money renting their home, apartment, private room?
* Do seasoned hosts make more money compared to newbie hosts?
* How much does location impact rental rates?
* Do most hosts live where their list property is located?
* Which types of rentals (homes, apartments, private rooms) get most bookings?
* What is the impact of COVID-19 on Airbnb host’s listings, prices? (this question might be difficult and I am leaving it as optional)