

# Initial Report

## Spring 2020 CSCI 3390 Network Science

Group Members: Lizzy Hanley, Kieran Roth, Catriona Sullivan, Kevin Li

Wednesday, February 19th, 2020

<b>Title</b>	<b>1</b>
<b>Abstract</b>	<b>1</b>
Data Sources	2
<b>References</b>	<b>2</b>
Summaries	2
The Effect of Home-Sharing on House Prices and Rents: Evidence from Airbnb	2
The economic costs and benefits of Airbnb	2
Predicting Airbnb prices with machine learning and location data	3
Understanding Airbnb's impact through census data	3
Neural Networks and Deep Learning	4
An End-To-End Data Science Project That Will Boost Your Portfolio	4
Key factors affecting the price of airbnb listings: A geographically weighted approach	4
Citations	5
<b>Motivation &amp; Problem Statement</b>	<b>5</b>
<b>Potential Challenges</b>	<b>6</b>
<b>Timeline</b>	<b>6</b>
Midterm Report (~March 20)	6
Final Presentation (~April 25)	7
Ultimate Goal	7

## Title

Do Airbnb prices predict wealth distribution in a city?

## Abstract

We will be investigating if the distribution of Airbnb prices in a city accurately represents the wealth distribution of that city. In order to do this, we plan to visualize on a map the wealth distribution of a city, and then add a layer to the map that visualizes Airbnb prices across that city. Our goal is to do this for

three cities-- New York, New York, San Francisco, California, and Austin, Texas-- and provide measurements of how similar the Airbnb price layer is to the wealth distribution layer of each map.

## Data Sources

Airbnb data: <http://insideairbnb.com/amsterdam/>

Wealth data: <https://datausa.io/profile/geo/austin-tx/#economy>

## Additional Data Sources

Median household income tiers:

<https://www.6sqft.com/interactive-map-reveals-the-income-gap-that-divides-nycs-richest-and-poorest/>

2019 USA Median Household Income:

<https://www.arcgis.com/home/item.html?id=a95db032417f40579fb68ff98ca14847>

## References

### Summaries

#### *Getting Started: Tutorial on gmaps*

This is the documentation for gmaps: a plugin for Jupyter. The “Getting started” page has an overview of basic concepts and tools for creating heat maps. While this is not a research article, this will be a crucial reference for our project because we plan to create a heat map of wealth distribution in selected cities, and thus having this tutorial allows us to create such a map. Utilizing the Google Maps API will be incredibly important as it will allow us to create beautiful data visualization maps!

#### *The Effect of Home-Sharing on House Prices and Rents: Evidence from Airbnb*

This is a research article exploring the effects of home-sharing on house prices and rents. The researchers used data from the Airbnb website, Zillow, and the Census Bureau’s American Community Survey. Specifically they looked at zipcodes and house and rental prices. They found that a 1% increase in Airbnb listings leads to a 0.018% increase in rents and 0.026% increase in house prices. Our project will be more of a visual representation and will not encompass the entire U.S., but it is beneficial to look at prior research when analyzing our own results.

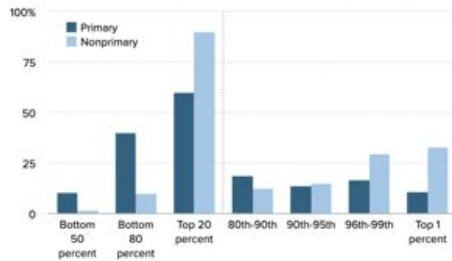
#### **The Economic Costs and Benefits of Airbnb**

This study aims to understand the economic costs and benefits of Airbnb. The author classifies Airbnb as an “internet-based service firm”, which have been known to cause controversy in discussions as many see them as attempts by “capitalists to profit by flouting regulations.” Specifically, this study aimed to assess the economic costs and benefits of Airbnb in order to see how to regulate the rising force of IBSF’s.

In order to conduct this study, they using housing wealth data from the Federal Reserve Board Survey of Consumer Finances.

### Housing wealth—particularly wealth from owning a nonprimary residence—is skewed

Share of total primary and nonprimary household housing wealth in the U.S. economy held by each wealth class, 2016



Using this figure, one can see that the potential benefits of Airbnb to property owners are extremely focused on those who are already deemed “wealthy.” The researchers also used data to show how white households disproportionately benefit from the housing wealth. The study goes further into detail with other potential benefits and costs of increased Airbnb presence. Overall, as stated before, the results showed that Airbnb benefits those who are white and deemed already “wealthy”, thus the costs defeat benefits that only apply to a small group.

This study could be greatly beneficial for us as we could analyze the overall wealth of a city over a specific time vs. the presence of Airbnb in that time period. It would be interesting to see if there are any connections with overall household income/wealth and increased Airbnb presence.

### *Predicting Airbnb prices with machine learning and location data*

In this article, the author explores the relationship between Airbnb prices and its determinants using July 21, 2019 Airbnb data for the city of Edinburgh, Scotland. Carillo’s project uses several listing features to try and predict price, and includes a predictor based on space (the property’s proximity to certain venues). His data source is an open-source tool called Inside Airbnb and he hypothesizes that access to restaurants, shops, cafes, bars and pubs are important for Airbnb users. He also conducts geospatial analysis using the Foursquare API, OpenStreetMap (OSM), Pandas, Pandana, and GeoDataFrame by looking into locational features as possible predictors. Carillo concludes that it seems price is affected by previous renters positive rating of location instead of the actual accessibility of the property with respect to main points of interest. From this article, our group can consider how seasonality affects Airbnb prices and also follow the methods for conducting geospatial analysis for mapping the distribution of prices. Our analysis differs as we will not be investigating how to predict Airbnb prices.

### *Understanding Airbnb’s impact through census data*

This study aims to understand and see if Airbnb does contribute to the “displacement of longtime, poorer residents” in cities. Specifically, in this study they wanted to see the effects on a much more local level – neighborhoods or even blocks.

As we were planning to use, they used Inside Airbnb for all data on the location and number of Airbnbs, however they filtered the data to see neighborhood specific details. Firstly, the researchers took the data and put it in block groups to show the concentration of listings per square mile. Then, they compared this to the number of available housing units. Comparing the Airbnb market to the rental market did not reveal

very surprising results, as the more the expensive areas yielded higher rental rates. However, after reading through this study we did not seem to see their answer to the first question presented in the introduction.

The visualizations that were created for this study will be incredibly useful for us because we try and use their wealth data for NYC and apply and compare it to the Census data we have been able to find. Using their map can help us see over time the effects Airbnb have had on the housing prices for example, or the wealth of a city overall.

### *Neural Networks and Deep Learning*

This chapter explains what a neural network is by writing a computer program to implement a neural network that learns to recognize handwritten digits. The author explains that a neural network uses a set of training examples to automatically infer rules for recognizing handwritten digits. Furthermore, by increasing the number of training examples, the network can learn more about handwriting, and so improve its accuracy. This chapter is a great source for our group to learn about neural networks, as this is a model several of the other references that we have included mention. The information in this chapter will help us better understand the rest of our references, however our project differs as we will not be implementing a neural network on handwritten digits.

### *An End-To-End Data Science Project That Will Boost Your Portfolio*

In this article, Sannazzaro uses an Amsterdam Airbnb dataset for predicting the price of a flat given a list of predictors. He creates a heatmap using gmaps to better understand where flats are located and how the price is affected by the location. Additionally, he implements feature engineering by creating a list of point of interests (POI) and builds a variety of models for predicting price. From this analysis, Sannazzo found that locations in the city centre are more expensive, while the outskirts are cheaper. He also concludes that the best model was the Gradient Boosting Tree (LightGBM). Sannazzaro's work will be an excellent resource for our group for using pandas\_profiling to explore our data and using gmaps for creating a heatmap. Our project differs in that we will not be building a model for predicting price.

### *Key factors affecting the price of airbnb listings: A geographically weighted approach*

This is a research article exploring factors that affect the prices of Airbnb listings. The data used included 794 samples of Airbnb listings of business units in Metro Nashville. The results indicated that Airbnb listing prices are more sensitive to the distance from the convention center than in other areas. We are only looking at how wealth distribution affects Airbnb prices, but it will be interesting to see if our visual reflects this research.

### *Where Do Airbnb Hosts Make the Most Money?*

This is an article that discusses some key findings on the profits of Airbnb renters and how some people find "Airbnb roommates" to help pay for the cost of renting an apartment. They used data from BeyondService, a service that optimizes price listing for Airbnb hosts and rental owners. While some Airbnb hosts rent out part of their own homes, others rent out a room from apartments that they are renting themselves. Our visual representation of Airbnb price listings and that area's wealth could maybe

reveal some more insight on how economic situations affect Airbnb price listening and how that would contribute to the state or city's wealth distribution.

### *Cities Tell Airbnb to Make Room for Affordable Housing*

This is a short article that brings up the issue on how Airbnb are pushing long term housing options (homes, apartments) out of the market with their short term rental housing. The article also discusses some of Airbnb's impact on hotels as well. This article not only discusses the problems between Airbnb, the city, housing, and its economy, but also how Airbnb has different regulations that makes it more accessible in comparison to a contract with a landlord. Using ATTOM Data Solutions, a company that has a database for tracking property, as well as other studies, the article makes some claims on the affordability of using Airbnb. We could use it as a reference and a motivation for our project to build a visual that can give us insight on whether the price listings on Airbnb are actually affordable for that area.

### Citations

Barron, K., Kung, E., & Proserpio, D. (2020). The effect of home-sharing on house prices and rents: Evidence from airbnb. *Ssrn*, Retrieved from Barron, Kyle and Kung, Edward and Proserpio, Davide, The Effect of Home-Sharing on House Prices and Rents: Evidence from Airbnb (January 22, 2020). Available at SSRN: <https://ssrn.com/abstract=3006832> or <http://dx.doi.org/10.2139/ssrn.3006832>

Bivens, J. (2019, January 30). The economic costs and benefits of Airbnb: No reason for local policymakers to let Airbnb bypass tax or regulatory obligations. Retrieved February 18, 2020, from <https://www.epi.org/publication/the-economic-costs-and-benefits-of-airbnb-no-reason-for-local-policymakers-to-let-airbnb-bypass-tax-or-regulatory-obligations/>

Bugnion, P. (2016). Getting started. Retrieved from <https://jupyter-gmaps.readthedocs.io/en/latest/tutorial.html>

Carrillo, G. (2020, January 29). Predicting Airbnb prices with machine learning and location data. Retrieved February 19, 2020, from <https://towardsdatascience.com/predicting-airbnb-prices-with-machine-learning-and-location-data-5c1e033d0a5a>

Carto. (2016, March 15). Understanding Airbnb's impact through Census data. Retrieved from <https://carto.com/blog/airbnb-impact/>

Nielsen, & A., M. (1970, January 1). Neural Networks and Deep Learning. Retrieved February 19, 2020, from <http://neuralnetworksanddeeplearning.com/chap1.html>

Povich, E. S. (2018, October 18). Cities Tell Airbnb to Make Room for Affordable Housing. Retrieved February 19, 2020, from <https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2018/10/18/cities-tell-airbnb-to-make-room-for-affordable-housing>

Sannazzaro, R. (2020, January 9). An End-To-End Data Science Project That Will Boost Your Portfolio. Retrieved February 19, 2020, from

<https://towardsdatascience.com/an-end-to-end-data-science-project-that-will-boost-your-portfolio-c53cfe16f0e3>

Wallace, N. (2018, February 20). Where Do Airbnb Hosts Make the Most Money? Retrieved from

<https://smartasset.com/mortgage/where-do-airbnb-hosts-make-the-most-money>

Zhang, Z., Chen, R., Han, L., & Yang, L. (2017). Key factors affecting the price of airbnb listings: A geographically weighted approach. *Sustainability*, 9, 1635. doi:10.3390/su9091635

## Motivation & Problem Statement

Our goal is to test the hypothesis that the distribution of Airbnb prices in a city accurately represents the wealth distribution in that city. For our investigation, we will define ‘accurate’ at 70% similarity or higher. We will be using median household incomes from a census dataset to measure wealth. Additionally, we will be using Airbnb listing data from Inside Airbnb.

We are motivated to investigate this hypothesis because some of our group members have used Airbnb before, and have been curious about how a list’s price reflects the actual population of folks who live in that area. In our experiences, we have used Airbnb when traveling to a new city. The decision to rent an Airbnb rather than a hotel was motivated by the belief that an Airbnb is a personalized experience, and thus an authentic way to explore a new city.

The results of our analysis will create value for renters on the Airbnb platform who are looking for authentic ways to explore new cities. Our results will inform renters if Airbnbs are as genuine of an experience as Airbnb intends to platform to be. Additionally, as noted in our analysis of the sources, it seems that Airbnb may be the cause of disproportionate wealth distribution - leaning towards white and wealthy people. A great way to publicize our work would be to incorporate it into the Inside Airbnb tool. Because Inside Airbnb is an independent, non-commercial set of tools and data that allows you to explore how Airbnb is really being used in cities around the world, this sort of analysis would enrich their platform. Our widget could be incorporated into the “see the data visually” page on InsideAirbnb’s website!

## Potential Challenges

One potential problem that we may encounter is if there is not enough economic diversity in the data for the cities we have selected. In order to be mindful of this concern, we have selected relatively big cities that should have wealth differences between areas, but it is still important to consider if our data is representative of most large US cities. Additionally, we will keep track of the data we are using and make sure there is a diversity in the median wealth among households in our three cities.

Another challenge is learning how to use new tools to visualize our data and create a web-app. We are considering using a python tool called gmaps to create maps, and we have intentionally included projects that have used this tool in our references.

Additionally, we have a bit of trouble with the Census data; this dataset for a specific city will be quite large, and we may not know exactly what needs to be filtered out. We do have some examples in the studies we researched that filtered the Census data in order for a more user-friendly set, however their methods are not entirely clear. Hopefully we will be able to pull-out the significant data for our project, and not be looking for a needle in a haystack of numbers.

Some members of the team have prior experience with web development. One team member took the Visualizations course at BC where she learned some basic techniques to create graphs using D3, Vega-Lite, and Tableau. We have access to the textbooks/learning material from her course which we plan to refer to for learning these technologies.

## Timeline

### Midterm Report (~March 20)

By the midterm, we hope to have finished the Airbnb price mapping and most of the wealth distribution mapping. We also hope to have developed metrics that will be used for comparing airbnb price distribution to wealth distribution.

#### *Tsengs Comments*

Completeness: 45/50

Feasibility: 26/30

Presentation: 17/20

Topic not that related, but interesting. The work would be harder than you think, especially the web app part. I'd recommend you to try some web app stuff by midterm, and if necessary, you can downscale the project or pivot your project.

Idea is nice, but lacks technical depth. For example, some descriptions are a bit arbitrary and out of nowhere. Why this line: "we will define 'accurate' at 70% similarity or higher. We will be using median household incomes from a census dataset to measure wealth"? How did you decide this criteria?

<https://scholar.harvard.edu/files/resseger/files/glaeserressegeretobiojrs.pdf>

This paper seeks to measure inequality in cities and uses household income as one of the metrics for measuring wealth.

Why did you mention neural network? Work cited should be closely related to your project.

## Final Presentation (~April 25)

Developing the web app will be the goal for the second half since most of us will need to learn how to use the technologies.

## Ultimate Goal

The ultimate product will be a web app that will display a heat map of the wealth distribution of a city and with the price listing data points of each Airbnb onto it. Our web app will allow a user to select which city they want to look at. Each data point will be colored based on its price/value. For example: Red (Most expensive) -> Yellow (Mean/Median) -> Green (Least expensive). The heat map would also follow this color scale but indicate the areas with the highest median income all the way to areas with the lowest median income. If the color of the datapoint is similar to the color of its surrounding area on the heatmap, then it would indicate a trend between the pricing of an Airbnb and the wealth of that area.

If we are able to complete everything above, we can use our extra time to add in comparison metrics of between cities. ie. rankings of which cities fair better or worse in our analysis and perhaps explore why this happened. Additionally, we could possibly use various years of Census data against various years of Airbnb data and fully examine how cities' wealth may have changed with the growth of Airbnb.