Final Project Report

Airbnb and the Effect of Demographic Factors on Airbnb Hosts

GROUP 3:

Alexa Goff

Jason Sue

Alexandra Hoyos

INFO 246 - 11 Information Visualization

San José State University

**Introduction**

It’s hard to believe that less than a decade ago nobody had ever heard of Airbnb. Founded in late 2007, Airbnb got its start when Joe Gebbia and Brian Chesky—a pair of entrepreneurs— rented three air mattresses in their San Francisco loft to “make a few bucks.” (Carson, 2016). And from those modest beginnings, Airbnb has grown in less than 10 years into a $31,000,000 dollar business that spans 191 countries (Bort, 2018; Smith, 2018).

With 150 million users, Airbnb has definitively reshaped the way the world travels, but it should not be forgotten that Airbnb’s popularity is due in large part to property owners who have been quick to embrace the Airbnb’s platform as a source of supplemental income (Smith, 2018). The types of properties listed vary widely. Whether a traveler prefers the privacy of renting an entire house or the affordability of a shared room, there is something to match every preference. Towards that end, Group 3 of INFO 246-11 has decided to examine which factors influence nightly listing price on the Airbnb website and to try and quantify the earnings potential of would-be hosts. This three-person team examined how elements such as neighborhood real estate value, crime rate, transportation, and accessibility to public parks affected the Airbnb listing price with a particular focus in using information visualization to analyze and illustrate the nature of the relationships between these variables. In order to accomplish this undertaking, the group narrowed the scope of their analysis to the city of Portland, Oregon. As a regional center of business and commerce and an up-and-coming travel destination; the Portland Metropolitan Area—or PDX as it is commonly referred to—was an ideal candidate to analyze which factors influence Airbnb listing price. Furthermore, in a recently published article by Airbnb; Portland, Oregon was highlighted as the number two most hospitable city in the United States (2018).

Group 3 of class INFO 246-11’s motivation to focus their analysis on the Airbnb data of Portland, Oregon was partially due to their shared interest in travel, and their familiarity with Airbnb. The impetus for specifically choosing Airbnb as the subject of their investigation was the group’s passion for traveling abroad and the personal experience of the group’s membership with using the Airbnb website to reduce the cost of their past travels. Another factor was the personal familiarity that Alexandra Hoyos had of people who list their properties on Airbnb as a major source of their personal income. When narrowing down the scale of Group 3’s investigation, the selection of the city of Portland, Oregon was due to the personal familiarity of Alexa Goff and Alexandra Hoyos as residents of the Portland Metropolitan Area (PDX), as well as the availability of current data on Inside Airbnb’s Portland listings and the availability of recent demographic data for the PDX area.

To accomplish the group’s goal in identifying which factors are most influential, Group 3 of class INFO 246-11 cross-referenced Inside Airbnb’s recently published data set concerning Airbnb listings within the Portland Metropolitan Area with Portland neighborhood data made available through *Monthly Portland* *Magazine’s* website (DeNies, 2016; Murray, 2018). With these results, Group 3 hoped that the relationships that they discover will have validity in other cities and countries around the globe so that in the near term they may reduce the cost of accommodations when they travel and perhaps in the future list their own properties as an Airbnb host.

**Data Selection**

With accommodations in 65,000 cities, Airbnb is a massive enterprise and an examination of Airbnb listings in its entirety is beyond the resources of this group’s three-person team (Smith, 2018). For this analysis, Group 3 narrowed the domain to a single city—the city of Portland, Oregon. Data were taken from two sources. The first source that Group 3 of INFO 246-11 drew from was from the third party website Inside Airbnb. This website’s Portland, Oregon data set was published on February 8, 2018 and provided a comprehensive directory on Airbnb listings located in the Portland Metropolitan Area (Murray, 2018). Not only did it include information on the location of each listing—through fields for neighborhood, latitude, and longitude—and their advertised price per night, Inside Airbnb’s PDX database also included information on whether a particular listing was instantly bookable and the minimum number of nights required to rent a particular accommodation. Many of the fields inside Airbnb’s PDX database were dedicated to the ratings and reviews of said Airbnb listings. Examples of these fields that described Airbnb listings included ratings such as the cleanliness of an accommodation, the total number of reviews, and the frequency of said reviews. Group 3 took this information and cross-referenced it with the Portland demographic neighborhood data made available through *Monthly Portland* *Magazine*’s website (DeNies, 2016).

*Monthly Portland*’s PDX data consisted of 2015 demographic data including information on real estate value, lifestyle factors, and crime rates with the exception of its criminal statistics where 2014 data were used. Of particular interest to Group 3 were lifestyle including factors relating to commute and transportation and the accessibility of amenities such as public parks. By combining the *Monthly Portland*’s demographic data set of PDX neighborhoods with the data from Airbnb listings provided by the Inside Airbnb website Group 3 identified three factors influence which influenced the nightly listing price of Airbnb Portland listings and the income potential of Airbnb hosts. These factors were real estate value, downtown commute, and accessibility of public parks.

**Design Process**

Group 3 of INFO 246-11 relied on Tableau to generate visualizations that illustrated the relationships between fields of Inside Airbnb’s data set and the Portland Metropolitan Area neighborhood data set from *Portland Monthly Magazine*. This was accomplished by joining the two data sets within a Tableau Workbook. The three members of Group 3 were largely unfamiliar with Tableau. Working with Tableau was a process of trial and error with the emphasis on error; but after a prolonged period of experimentation, Group 3 was able to illustrate the strength of relationships between the quantitative demographic factors and the Airbnb nightly listing price in a series of scatter plots. The scatter plots were made possible by aggregating the listings in the Portland Airbnb listings (or database records) by neighborhood, calculating their mean listing price and then plotting them with their corresponding value for the demographic factor under consideration from the *Monthly Portland* PDX data set. For the series of scatter plots used in this project to illustrate the relationships between demographic data and the Airbnb listing price, Group 3 elected to maintain a consistent design—consistent with Edward Tufte’s Integrity Principles—and minimized chart junk—consistent with Tufte’s Design Principles—to allow the data to stand on its own. To minimize interference, Group 3 controlled the perceptual properties utilized by the marks by setting the mark’s shape as solid circles—or dots—and calibrating the hue of said marks to match the hue of the Airbnb logo while simultaneously increasing the contrast between them and the design space. In addition, the axis, tick marks, and labels were darkened to increase visibility. (See Appendix A for a scatter plot example).

Since the study was spatial in nature, choropleth maps were created to accompany the scatter plots and to provide the reader with a more accurate sense of statistical trends across space. The maps were created using GIS data for Portland neighborhood boundaries made available by the City of Portland Open Data (2018). These boundaries were imported into a Tableau workbook and joined with Airbnb and Portland demographic data. Individual map dashboards were created for each factor of the study, where a corresponding measure was placed on the Marks card to display the statistical measure by magnitudes of filled colored regions. Visual details like the light gray basemap were left intact, but the color palette was updated to Tableau’s “Temperature Diverging” bi-polar color progression of green-yellow-red in order to provide contrast and reveal the extreme ends of the data to the reader (Robinson et al., 1995). This color palette made the extreme ends of the data more apparent, but the user must rely on the corresponding legends placed near the maps in order to interpret the values of the colors.

Since the study involved analyzing multiple factors, Group 3 decided to organize the design space of the graphic into a series of graphs that isolated individual demographic inquiries. This minimized the amount of chart junk in a single frame and allowed the reader to interpret a subject with minimal distractions. Group 3 utilized the visual narrative form of an interactive slideshow using default settings in a Tableau Story, in order to create a linear series of graphs that allow for reader-interaction mid-narrative (Segel & Heer, 2010). The graphic was uploaded to Tableau Public so that users could interact with the data through tooltips and highlighting. Exploration of the graphic is guided by a highlighted progress bar with captions describing the subjects of the graphs in the series. Visual signs like arrows on the progress bar give the viewer additional clues to click forward and move throughout the graph series. Though the overall narrative structure reflects an author-driven approach, it is balanced with reader-driven control because the graph series does not use static messaging throughout to give an interpretation, but it instead presents several graphs that users may explore and use to form their own hypotheses (Segel & Heer, 2010).

**Data Analysis**

Overall, Group 3 of INFO 246-11 identified three relationships that influenced Airbnb’s nightly listing price, and therefore, the earnings potential of Airbnb’s host. The strongest influence was real estate value, which has a moderately strong relationship with Airbnb listing price. Group 3 also discovered a minor relation between park accessibility and Airbnb listing price. In addition, the commute to downtown was identified as a minor determinant of the nightly Airbnb listing price. Group 3 also tested whether crime rate and instant booking ability factored into Airbnb listing prices, but neither factor generated a statistically significant relationship.

The data set compiled by *Portland Monthly Magazine* included three metrics to measure the real estate value of each neighborhood in Portland. These three metrics were the mean neighborhood real estate value, the median neighborhood real estate value, and the neighborhood’s cost per square foot. By plotting each of these metrics in a scatter plot against the Airbnb listing price, the group used regression analysis to determine which of these measures is the most appropriate.

When plotting the mean neighborhood real estate value in relation to Airbnb listing price, the scatter plot returned a moderate correlation of r2 = .417364. The median neighborhood real estate value when plotted against the Airbnb listing price yielded a coefficient of determination of r2 = .37315. The relationship between the neighborhood’s cost per square foot was even weaker. Linear regression analysis between the Airbnb listing price and the neighborhood’s cost per square foot generated a coefficient of determination of r2 = .242059. Given these values. it can be reasonably concluded that the mean neighborhood real estate value—by having the strongest correlation—is the most appropriate measure for comparison with the Airbnb listing prince. A coefficient of determination of r2 = .417364indicated that the mean neighborhood real estate value was predictive of 41.7% of the average variation of Airbnb listing price.

 This moderate correlation implies that there were other factors that affect listing price. Group 3  was curious on whether the listing type—entire homes or apartment, private rooms, or shared room—would improve the correlation. Leveraging the power of Tableau, the group was able to partition the original scatter plot into three component scatter plots. When plotting the correlation between mean neighborhood real estate value and the Airbnb listing price of entire homes or apartments, the component scatter plot yielded a coefficient of determination of r2 = .380948. When plotting the correlation between mean real estate value and the Airbnb listing price of private rooms, the coefficient of determination was r2 = .117352. When plotting the correlation between mean real estate value and the Airbnb listing price of shared rooms, the scatter plot yielded a coefficient of determination of r2 = .0833628. At first glance, this data seemed to indicate that the mean real estate value had a moderate correlation with the Airbnb listing price of homes and apartments, and a minor correlation with both the Airbnb listing price of shared rooms, and the Airbnb listing price of private rooms, but a visual inspection of all three component scatter plots showed that the scatter plot that illustrated the relationship between the mean neighborhood real estate value and the Airbnb listing price was affected by two outliers (see Appendix A).

The most prominent of these outliers was the data point from the Eliot neighborhood. Excluding this data point, did not substantially affect the correlation between mean neighborhood real estate value and the Airbnb listing price of entire homes or apartment of the correlation between mean neighborhood real estate values and Airbnb listing price. As expected, eliminating the Eliot neighborhood outlier increase the correlation between mean neighborhood real estate value and the Airbnb per night listing price of shared rooms from r2 = .0833628 to r2 = .243057 (see Appendix A). In layman’s terms, this means that variation in mean neighborhood real estate value could account for 24.4% of the variation in the Airbnb nightly listing price of shared rooms instead of the much smaller 8.3% when excluding the Eliot neighborhood as an outlier.

To summarize Group 3’s analysis of real estate value, it was found that mean neighborhood real estate value was moderately predictive of the Airbnb nightly listing price with variation in mean real estate value being predictive of 41.7% of variation or Airbnb nightly listing price. When segregating this data by accommodation type—whole homes or apartments, private rooms, or shared rooms—and excluding data from the outlier Eliot neighborhood variation in mean real estate value could account for 38.1% of variation in the listing price of whole homes or apartments on the Airbnb website, 11.7% of variation in the listing price of private rooms, and 24.4% of the variation in the listing price of shared rooms. In short, real estate value was a moderate determinate of Airbnb listing price; but when analyzing the data and segregating by accommodation type, the relationship between real estate value and Airbnb listing prices was weakened, but maintained a positive predictive relationship.

There are two metrics that measure the accessibility of parks for each Portland neighborhood. The first measured the acreage of parks as a percentage of the total neighborhood’s footprint. The second measured the total acreage of park land within a neighborhood. When plotting park acreage as a percentage of the total neighborhood’s footprint versus Airbnb listing price in a scatter plot, the linear regression analysis returned a coefficient of determination of r2 = .0906087 for the correlation. Similarly, a scatter plot of public park acreage versus nightly accommodation listing price generated a correlation of r2 = .28604. From these values, one would assume that with its stronger correlation that public park acreage was the better metric. Unfortunately, a visual inspection of the scatter plots revealed that the linear regression analysis of public park acreage versus Airbnb listing price was highly dependant upon outlier data point representing the Forest Park neighborhood; therefore park acreage as a percentage of the total neighborhood’s footprint was the better metric. In short, accessibility of public parks is a minor determinant of Airbnb nightly listing price where variation in the percentage of a neighborhood’s acreage utilized as a public park space is predictive of 9.1% of the variation in Airbnb listing price.

Since many of the tourist attractions such as Voodoo Doughnut and the bookstore Powell's City of Books was found in downtown Portland, and much of Portland’s business was connected to its downtown district, Group 3 tested whether the vehicular commute to Portland’s downtown measured in minutes influenced Airbnb listing price. When plotting both the vehicular commute to Portland’s downtown and the nightly Airbnb listing price in a scatter plot, the resulting trendline had a negative relationship and a correlation coefficient of r2 = .0902248. This meant that variation in distance to downtown as measured by vehicle commute was responsible for 9% of the variation in Airbnb listing price and that vehicular commute is a mild determinate of Airbnb per night listing price.

Other factors that Group 3 were interested in was whether if crime rate or if designating an Airbnb listing as instantly bookable affected the Airbnb listing price. When plotting the crime rate vs. Airbnb listing price in a scatter plot, linear regression analysis generated a coefficient of determination measuring the effect of at r2 = 0.009242. Such a small coefficient indicated that crime rate has a negligible effect on Airbnb listing price. When testing to see if making an Airbnb listing as instantly bookable had a discernible difference in the listing price the mean listing price for instantly bookable Airbnb listings in Portland, Oregon was $120.48. This was in comparison to a mean $117.26 for Airbnb listings in Portland, Oregon without instant bookings. To put this in perspective, the difference of 2.7% was hardly significant. Another measure Group 3 used was overall review scores as a measure of satisfaction. Again, the difference between the two groups of Airbnb listing—accommodations that were instantly bookable and accommodations that were not instantly bookable—was not significant. On a hundred-point scale, Portland Airbnb listings that were instantly bookable had a mean rating of 96.6016; and the mean rating for Portland Airbnb listings that were not instantly bookable was 96.9585.

Overall, Group 3 of INFO 246-11 identified three relationships that influenced Airbnb’s nightly listing price; and therefore, the earnings potential of Airbnb’s host. The strongest influence was real estate value which has a moderately strong relationship with Airbnb listing price where more expensive properties corresponding with higher listing prices. There was a minor relationship between parks and Airbnb listing prices with a greater acreage of parks corresponding with higher Airbnb listing prices. In addition, the commute to downtown was discovered to have a slightly negative relationship with Airbnb listing prices. Group 3 also tested crime rate and whether instant bookings factored into Airbnb listing prices, but neither factor generated a statistically significant relationship.

**Insights and Implications**

The results of this study provided only limited strategic implications for a host seeking to use Airbnb as a source of revenue. Out of the five factors that were studied, only real estate value had any significant relationship with Airbnb listing price. Thus, a host may generally be able to charge more per night for a rental in neighborhood with higher property values. However, the data used for this study shows there is a potential limit to such a strategy because the neighborhood with the highest average real estate price in Portland, Dunthorpe, which has an average home price of about $1.1 million (DeNies & Stanton, 2016), did not contain any Airbnb listings. Some reasons for this might include neighborhood association restrictions, the suburban location, or that residents of the neighborhood do not wish to seek extra income on this platform.

One insight found in the course of the study was the slightly positive relationship of a listing’s proximity to downtown and price. What surprised some members of Group 3 was the weakness in the correlation between downtown commute and Airbnb price given that many of Portland’s tourist attractions are located in a near their downtown district. This weakness of the relationship may be partially attributed to Portland’s great public transit where Airbnb listings with fewer social problems and easy access to public transport may be drawing guests away from staying in the downtown neighborhood.

Airbnb rentals in the Forest Park neighborhood stood out as an outlier in average listing price data, where hosts asked for an average of $346 per night versus the city-wide average of $120. The Forest Park neighborhood sits against Portland’s largest urban park of the same name, which is composed of an extensive 70-mile network of hiking trails in dense second-growth forest. Forest Park is highly recommended on local travel websites like TravelPortland.com as an ideal location for Pacific Northwest urban tourism and outdoor recreation (Pullen, 2017). Access to these features and its picturesque landscape likely influenced the neighborhood’s desirability as a more expensive Airbnb destination. Airbnb hosts who have the flexibility to select a location may want to invest in locations in Forest Park or nearby, where guests have access to Portland’s largest and most desirable park.

**Conclusions**

One of the challenges Group 3 faced was the difficulty in utilizing host ratings and reviews as a larger part of the analysis. The majority of these ratings were based on a scale of 1–10, where most of the ratings were 9’s or 10’s and very few 8’s. This is due to the unfortunate tendency among Airbnb travelers to give a host 10’s across the board even when their experience would indicate a lower rating would be more appropriate. Such a behavior wards off retaliation from hosts who might “warn” other Airbnb’s hosts that this honest, if disgruntled traveler was a less than ideal guest. This would, in turn, make it more difficult for said traveler to book future accommodations through Airbnb. Group 3 worked around this difficulty by utilizing Airbnb nightly listing price as the primary metric when comparing demographic factors and using the general rating which has a scale of 1–100 as a measure of guest satisfaction when warranted.

Out of the five factors that Group 3 of INFO 246-11 for their relationships with average Airbnb listing price. we found that only a few of them actually had a statistically significant relationship. Only real estate value had a moderately strong relationship with the average listing price where higher the property values in the neighborhood were correlated with more expensive rates for Airbnb listings. Two of the factors that had showed only a weak relationship to average listing price were distance to downtown and access to parks; so that Airbnb accommodations that were either close to downtown or near large public parks were more likely to charge a higher price than their peers. Finally, two of the factors—crime and instant bookings—showed no statistical correlation to average listing price. As to the original question on discovering what factors influence the income potential of would be hosts, Group 3’s analysis reveal that would be hosts have very strategic options available short of investing in a highly desirable property.

References

Airbnb (2018, May 4). Airbnb unveils top 10 most hospitable cities in the U.S.. Retrieved from https://press.atairbnb.com/airbnb-unveils-top-10-most-hospitable-cities-in-the-us/

Bort, J. (2018, February 6). Airbnb made $93 million in profit on $2.6 billion in revenue, but an internal clash sent the CFO out the door. Retrieved from http://www.businessinsider.com/airbnb-profit-revenue-2018-2

Carson, B. (2016, February 15). How 3 guys turned renting an air mattress in their apartment into a $25 billion company. Retrieved from http://www.businessinsider.com/how-airbnb-was-founded-a-visual-history-2016-2

City of Portland, Oregon Open Data. (2018, January 5). Neighborhoods (regions). Retrieved from https://gis-pdx.opendata.arcgis.com/datasets/neighborhoods-regions

DeNies, R., & Stanton, S., (2016, April 1). Portland neighborhoods by the numbers 2016: The city. *Portland* *Monthly.* Retrieved from: https://www.pdxmonthly.com/articles/2016/4/1/real-estate-2016-the-city

Jasper. (2015, September 18). The Airbnb founder story: From selling cereals to a $25b company. Retrieved from https://getpaidforyourpad.com/blog/the-airbnb-founder-story/

Marum, A. (2018, January 8). Downtown Portland’s paradox: Crime, public perceptions threaten growth, retailers say. Retrieved from http://www.oregonlive.com/business/index.ssf/2018/01/downtown\_portlands\_paradox.html

Murray, C. (2018, February 8). Get the data. Retrieved from http://insideairbnb.com/get-the-data.html

Pullen, J.P. (2017, January 3). Forest Park: Escape the city without leaving Portland in this massive urban wilderness that’s a haven for animals and adventurers alike. Retrieved from https://www.travelportland.com/article/forest-park

Robinson, A.H., Morrison, J.L., Muehrke, P.C., Kimmerling, A.J. & Guptill, S.C. (1995). *Elements of Cartography.* New York: Wiley.

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 1139-1148. Retrieved from http://vis.stanford.edu/files/2010-Narrative-InfoVis.pdf

Smith, C. (2018, May 8). 100 Airbnb statistics and facts (March 2018): By the numbers. https://expandedramblings.com/index.php/airbnb-statistics/

Appendix A

Scatter plots used to determine the strength of the correlation between real estate value and Airbnb Listing Price by accomodation type.



*Figure 1*. The Real Estate Value vs Airbnb Listing Price by Listing Type. Two outliers have an outsized effect of which the most egregious is the Eliot neighborhood on the correlation of shared rooms component scatter plot r2 = .0833628..



*Figure 2*. The Real Estate Value vs Airbnb Listing Price by Listing Type Excluding the Eliot Neighborhood. Correlation in Eliot neighborhood increases to r2 = .243057, while the correlation of the other component scatter plots remain unaffected. Figure 1 and Figure 2 do not share the same scale on their vertical axis.