

Krtin Juneja and Elizabeth Veilleux

Professor Carter

MA 346

Due: June 29, 2021

Final Project Report – Predicting Airbnb Prices in Los Angeles

In our final project we looked at the Los Angeles Airbnb data that provided details about the Airbnb rental itself and the individual who is hosting the listing. This data was chosen since we are interested in traveling, but also need to keep in mind staying in a college student budget (which can be limited). So, this got us thinking about what plays a role in predicting Airbnb prices.

The predictors we utilized were the number of bedrooms, beds, and accommodates along with the dummy variables that specified if the rental was an entire property, a private room, a hotel room, or a shared room. We chose another dataset to utilize which provided the rent prices of each LA neighborhood between the years 2010-2016 (we only used the most recent data from 2016). These values were also used as a predictor since it can give some insight on what areas are more expensive to live in than others.

Before we performed any exploration, we had to clean the data by creating functions to eliminate the price formatting so that we could convert it to a quantitative variable type (float). Additionally, the dataset had several unused columns as it originally had 31,543 rows with 74 columns which we dropped to 12 columns to reduce the size of the dataset. This would also speed up the execution of the code and be more efficient. After cleaning the data, we created visualizations like a pair plot, scatterplot, and heatmap using the seaborn and matplotlib packages to enhance the readability and exploration of the data as well as learn about the correlation with between the variables as our goal was to create a model to predict the price of Airbnbs.

Next, a multiple linear regression model was created on these variables and run, utilizing the packages called sklearn and statsmodels, and then we used machine learning to see if that would be a better method of prediction. The three different models were not as promising as we wanted since they lacked precision in predicting the Airbnb prices. This led us to believe that maybe other predictors would be more useful, but we did not have access to that in our timeframe. Moreover, intangible features like the design of the house and the view from it could also play a role in determining the price which can be difficult to measure. So, for the future, we could look into other factors that are involved in Airbnb prices which could be simply the appearance or maybe a variable like ocean proximity since California is on the West Coast.

Additionally, we ran some hypothesis tests to determine whether some LA neighborhoods were pricier than others (in a statistically significant way). To perform this, we used the group by function to create a data frame of all the neighborhoods with their mean price. When we utilized all the LA neighborhoods we had, we could not run a hypothesis test comparing the maximum and minimum mean Airbnb price since the minimum mean price neighborhood of Cudahy comprised of a single value. We resorted to utilizing the top 10 neighborhoods to live in LA according to the HomeSnacks website. We were able to run the hypothesis test on the minimum (West Los Angeles) and maximum (Bel-Air) for this set and concluded that the mean price of the two neighborhoods were significantly different. Therefore, we were able to reach the previously defined goal that some LA neighborhoods tend to be significantly pricier than others (at least when looking at those 10 neighborhoods).

In addition to our Deepnote Jupyter notebook, an exploratory dashboard was created on Heroku which involved data exploration opportunities for non-technical users that were looking for an Airbnb in LA. Our dashboard has a home page complimented by 2 additional pages that

allows one to find an Airbnb based on certain specifications like price, neighborhood and room type. Once an individual finds an Airbnb they are interested in, they can find the ID for that Airbnb and get description about that specific Airbnb rental and host like the minimum cost, amenities, and whether the host is a super host. We decided to do this as some of these specifications play an important role when selecting which Airbnb to stay in.

Overall, this project was interesting to partake in since we were exposed to new technology like Github, Streamlit, Heroku and Python packages. Additionally, it showed us that we won't always get the best results and will need to fix and edit in the future.