Daniel Park

DATA 606

**Delivery 2 Summary**

After looking through similar projects, I noticed that there are many similarities. Just like me, each author used insideairbnb.com for the source of their dataset. Insideairbnb.com is an anti-Airbnb lobby group that scrapes Airbnb data from multiple cities around the world. The dataset that is given by insideairbnb consists of 84 columns with thousands of rows. The dataset for Washington D.C. has 9,153 rows. That is a lot of data to process. To make things easier, the authors transformed and cleaned the data by removing most of the columns and keeping the features they feel were important. Some features that all the authors kept are accommodates, bedrooms, bathrooms, beds, price, minimum\_nights, maximum\_nights, and number\_of\_reviews. These were proven to be important features that determined prices throughout the projects. From what I’ve been seeing, the preliminary exploratory data analysis is very similar to each other. Some authors created graphs based on the data they initially found, like number of beds, accommodates, etc. One author (4) went ahead and created visualizations based on zip codes which I think is very interesting and clever. Another author (3) started with very basic machine learning models and applied them to only one feature (number of bedrooms vs price). These machine learning models involved KNN (K-nearest neighbor) and using RMSE (root mean square error) to evaluate the model. I plan on implementing simple machine learning models myself and putting in visuals to allow the audience to better understand what I am trying to accomplish for this project. More details will be present in part three of the project.

References

1. <https://towardsdatascience.com/predicting-airbnb-prices-with-machine-learning-and-location-data-5c1e033d0a5a>
2. <https://towardsdatascience.com/predicting-airbnb-prices-with-machine-learning-and-deep-learning-f46d44afb8a6>
3. <https://www.dataquest.io/blog/machine-learning-tutorial/>
4. <https://airbnb-pricing-prediction.herokuapp.com/index.html>