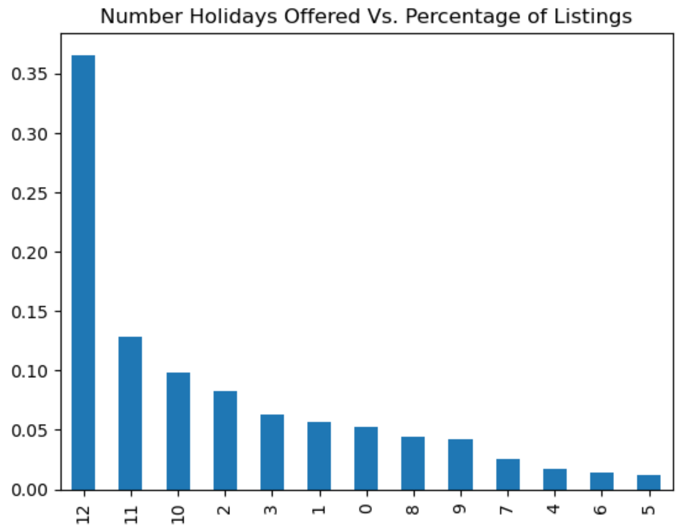
Can You Vacation On A Holiday using AirBNB?

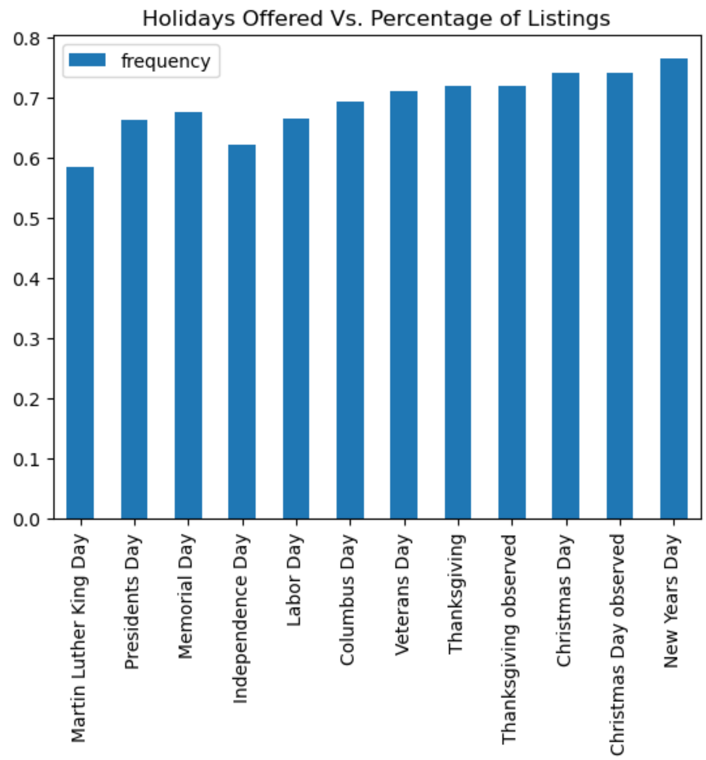
This paper details a basic analysis of listings on AirBnB and their availability during U.S. federal holidays. U.S. federal holidays were picked due to their likelihood to be reflected in the data of a U.S. city and they are days that people are likely to require accommodation. Findings for the analysis indicate that linear regression is only slightly better at determining if a listing will be available – at 78% – than defaulting to a listing as available – 69%. The findings also indicate that hosts throughout 2016 were more likely to make listings available as the year progressed, even when taking into account listings that did not appear to be available throughout the year. Additional findings show that AirBnB listings are generally available for all holidays, indicating hosts purchased properties to act as investment vehicles. The number of listings that are only available for a few number of holidays indicate that only a few hosts are renting our their main residence periodically.

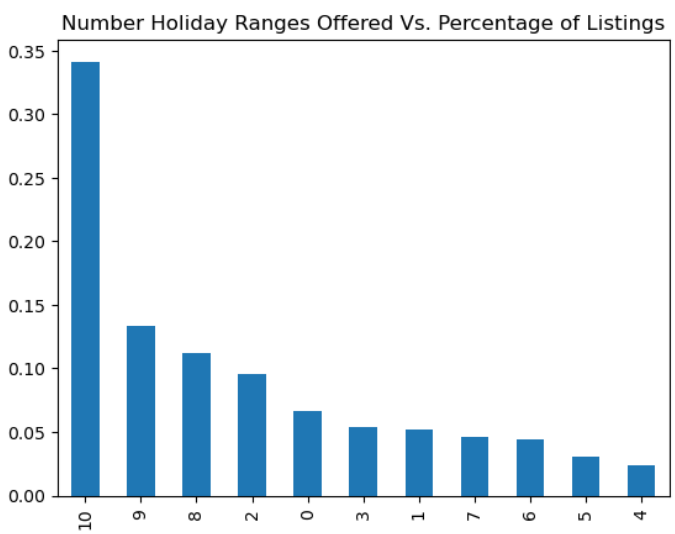
Section I: Listing Availability On Holidays

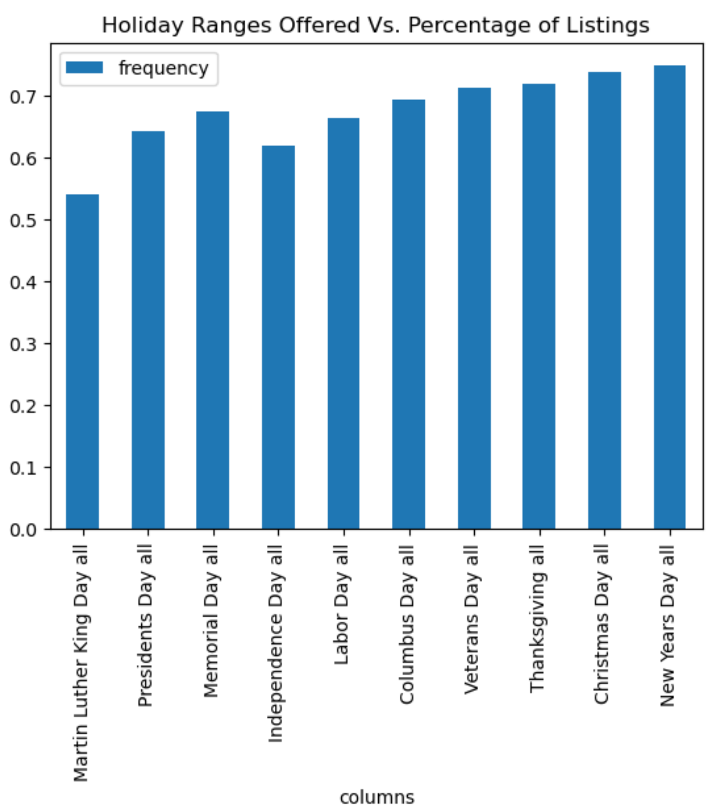
Data available for Seattle in 2016 indicates that most listings made all holidays available, with the next two largest groups making all but one or two holidays available. Listings making no holidays available represented only about five percent of listings. Looking at the availability percentage per holiday across the listings, it appears that hosts increased their availability as the year progressed. Viewing the same data for date ranges that include the holiday and corresponding weekend appear to show the same trends. This data shows that hosts are making their listings available year-round, an indication that the properties may have been purchased as an investment vehicle. The few properties that are only available on some holidays indicate that these properties could possibly be either either: 1) the host’s main residence that is being made available while the host is out of town; or 2) the host is using the listing themselves that holiday and blocked off the listing for others.

Holiday Availability on AirBnB in Seattle



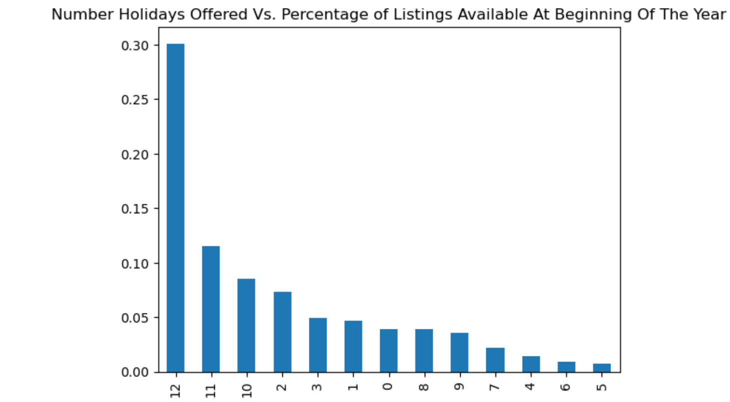


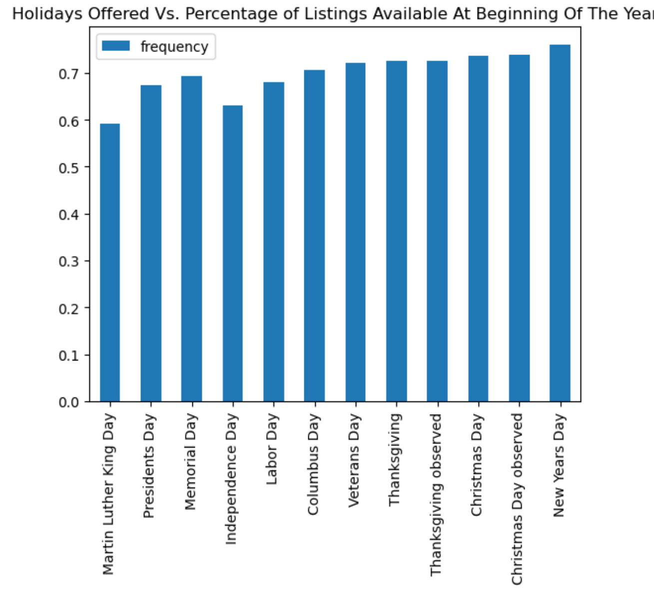


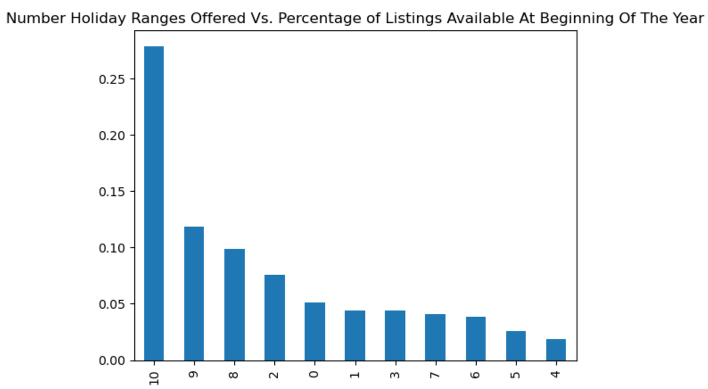


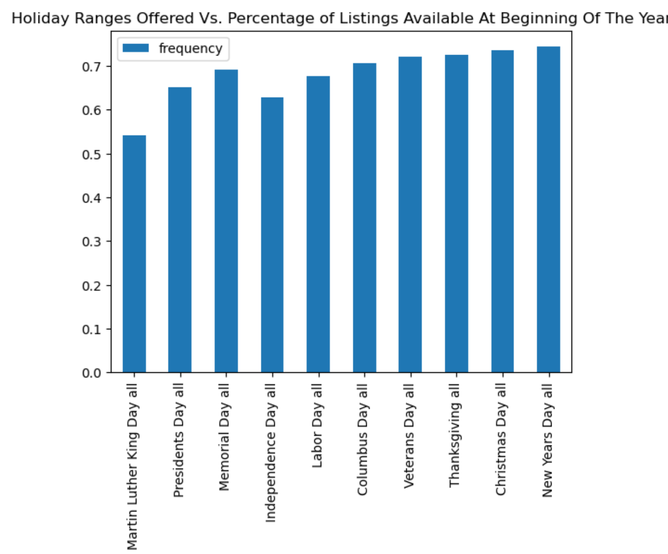
Section II: Listing Availability After Removing Non-Yearlong Listings

Removing listings that did not have a review dated before 2016 – indicating the listing was more likely to be available throughout 2016 – showed the same trends. A majority of listings were available for all or almost all holidays, indicating the listings were potentially purchased by hosts to act as an investment vehicle.





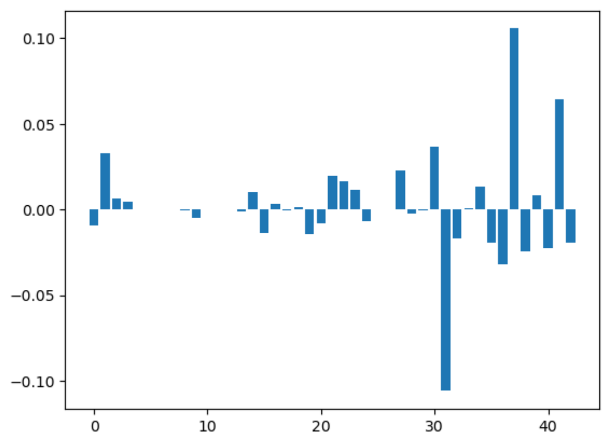




Section III: Predicting A Listing's Availability On Holidays

A listing's various features – such as cost, square footage, number of properties made available by the host – may indicate if a property will be available on a holiday. By encoding the data using value ranges and one-hot encoding enables these features to be analyzed through various statistical methods, such as linear regression. Using every feature available returns a prediction rate of only 56 percent. This is below the success rate of defaulting to assuming every listing will be available on every holiday. However, removing features based on the percentage of False values increases the prediction rate, likely by filtering out noise generated by infrequently referenced columns.

By running a linear analysis of all columns minus those filled by 99%, 98%, 97%, etc. of False values identifies that removing columns that contain 6% or more of False values generates the highest prediction rate – 78%. This is likely due to rarely used features generating no value and potentially generating noise that interfered with the linear regression. Another note, when performing the same reduction in features against the entire data set and then again against listings that were available the entire year, the columns to remove were inconsistent, ranging from removing columns that were filled with 6% to 96% False, respectively. This large ranges indicates that there is no readily available percentage to use when removing columns in this manner.



The graph to the left shows the coefficients of the features input into the model. Reviewing the coefficient weights applied to each feature identifies two columns with extremes – compared to other coefficients – that have opposite values from each other. These represent if a listing makes a washer and dryer available. Given most listings include both a washer and dryer and are therefore highly correlated, removing tone of these columns would make sense, as it is still represented by the other highly correlated feature. Removing the washer feature generated the same result while requiring less computation.

Removing other features did not improve the model. While some coefficients appeared to have zero affect, as indicated in the above chart, removing them decreased the results by as much as 30 percent.

Conclusion

The results of this analysis were not extraordinary. The analysis performed only nine percent better than defaulting to a listing as available. However, it would generate fewer false positives than defaulting to available. Provided this model could carry on to the following year, this analysis would more accurately predict availability during upcoming holiday seasons. Such an analysis could be used to set rates for hosts and help travelers identify availability. Continued analysis of this data, both from different date ranges and regions, would help improve this model and verify such predictions could be made.