

WAF Data Challenge

An Analysis of Airbnb Listings Across Hawaii

Darren Chen

Fall 2022

Exploratory Data Analysis (EDA)

Interesting findings:

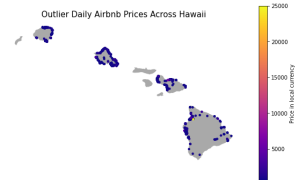
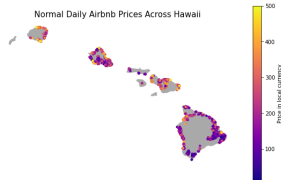
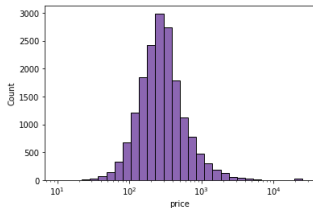
- Listings with daily prices above \$20,000
- Large majority of ratings between 4 and 5; very few less than 3
- Very few hosts without profile pictures; ~ 50 in a sample of 17,500
- Locations predominantly spread along coast

Data preprocessing:

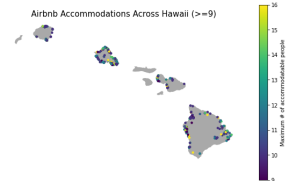
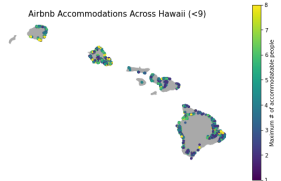
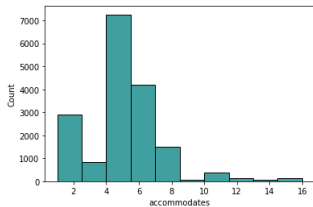
- Removed entries with missing values
- Removed entries with significant outliers
- Converted every entry to a numerical value (i.e. 85% \rightarrow 0.85)
- Created location coordinates from latitude and longitude for plotting purposes

EDA Cont.

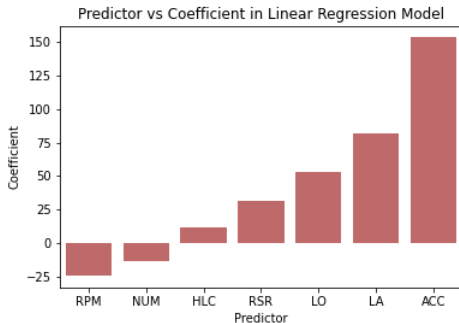
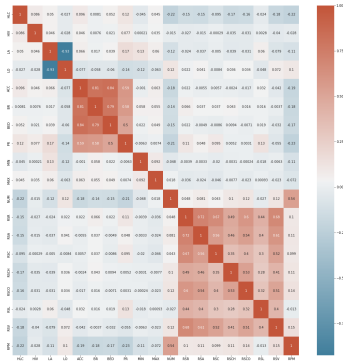
Prices:



Accommodations:



Regression

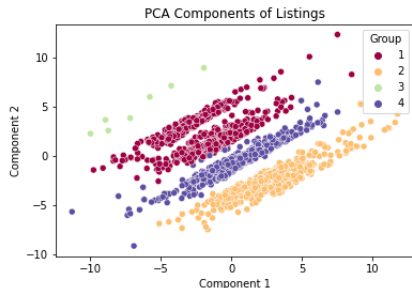


Key takeaway:

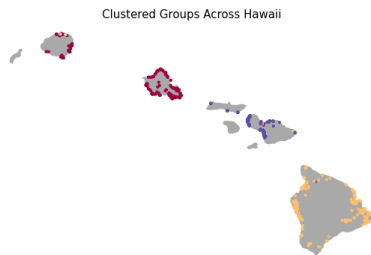
Accommodations, latitude, longitude, and review scores are the most predictive factors for price.

Clustering

Grouping similar listings:



Visualizing the groups:



Key takeaway:

The location of a listing is the predominant factor that groups different listings together.

Summary

Important findings:

- The price of a listing is strongly correlated with the number of people it can accommodate
- The locations of listings not only predict price but also group listings together
- Data science is fun!

Tools used:

- Python, Numpy, Pandas for data analysis
- Matplotlib, Seaborn, Geoplot for data visualization
- Scikit-learn for models (LinearRegression, PCA, SpectralClustering)