

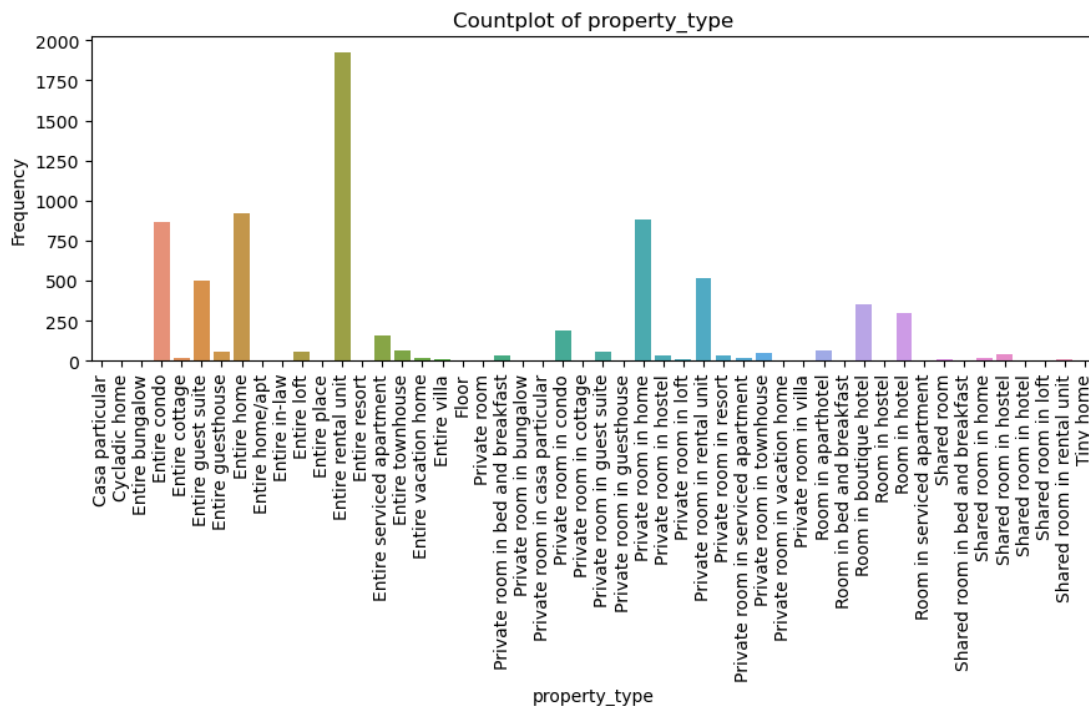
# Executive Summary

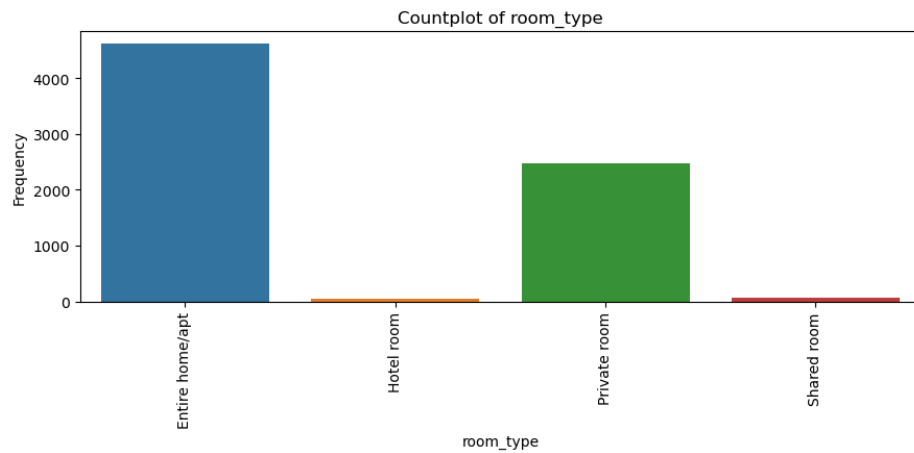
This report presents the findings and recommendations of a data science project focused on predicting Airbnb rental prices in San Francisco based on a dataset obtained from insideairbnb.com. The dataset spans from September 2022 to September 2023, containing 7208 rows and 75 columns. The predictive model, primarily utilizing Random Forest Regression, achieved a Mean Absolute Error (MAE) of approximately 0.33 on the test set.

## Data Exploration and Analysis

### Exploratory Data Analysis (EDA)

- Entire rental unit is the most common property type. 'Entire condo', 'Private Room in home', 'Entire home' are some other common property types.

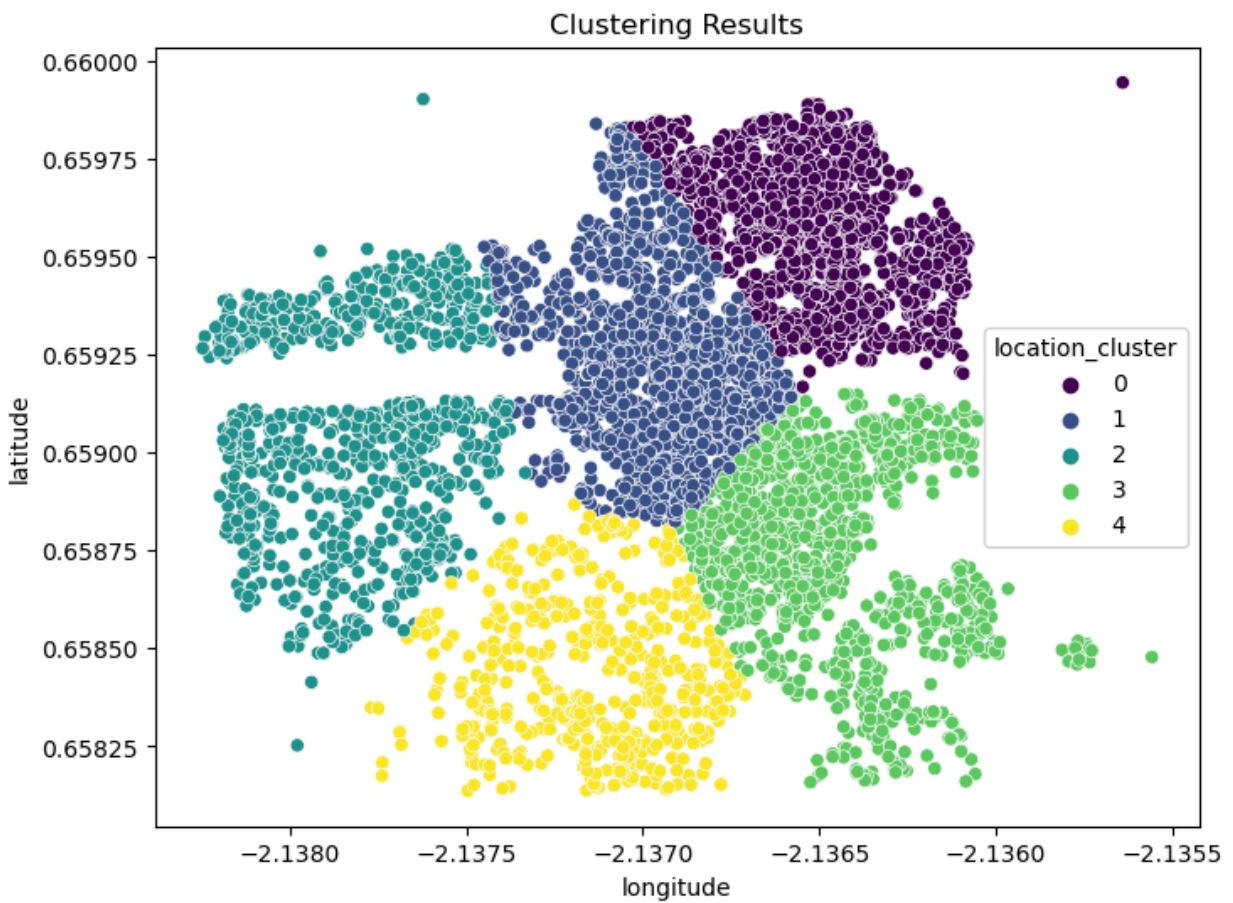




Most properties listed have either entire homes/apartments or private rooms available.

## Clustering based on Location

Clustering based on latitude and longitude was performed, creating clusters to replace the original coordinates.



# Model Selection

Given the dataset's outliers, Mean Absolute Error (MAE) was chosen as the evaluation metric, avoiding the amplification of extreme errors. Two models were considered:

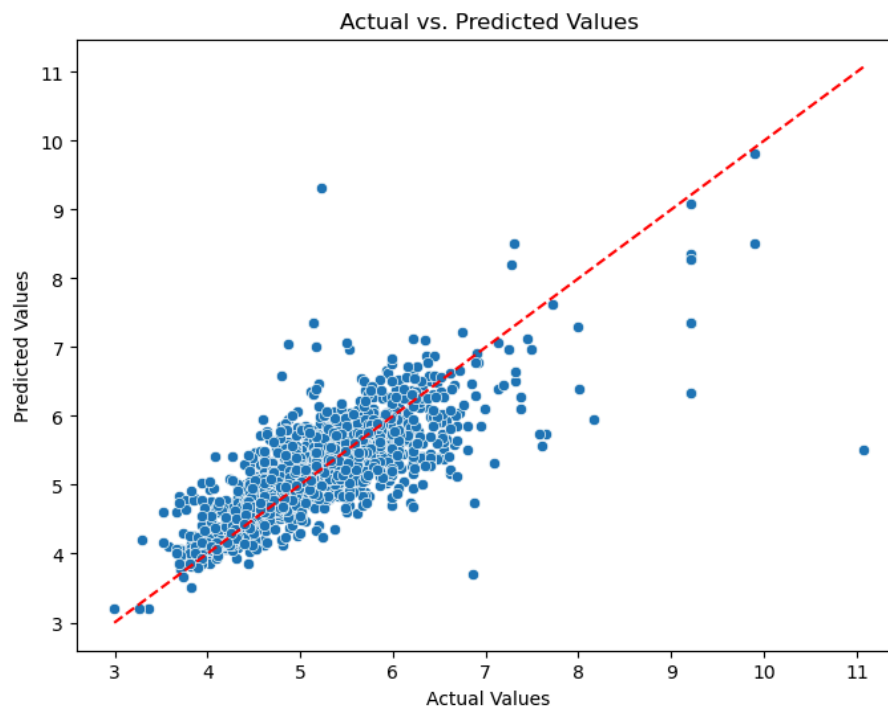
## 1. Linear Regression:

- Cross-Validation Mean MAE: 0.4275

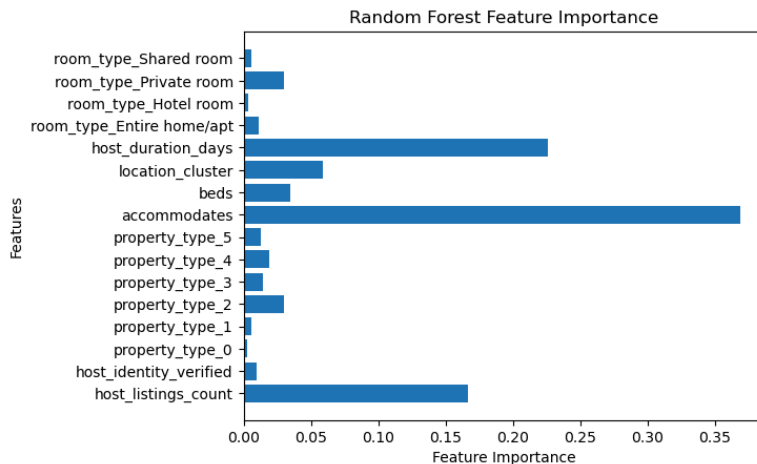
## 2. Random Forest Regression:

- Cross-Validation Mean MAE: 0.3303
- After hyperparameter tuning: 0.3261

The Random Forest Regressor demonstrated superior performance, making it the selected model for predicting Airbnb rental prices in San Francisco.



# Feature Importance



Key features influencing rental prices:

- 'Accommodates' (~37%)
- 'Host\_duration\_days' (~23%)
- 'Host\_listings\_count' (~17%)

'Location\_cluster' also held importance (~6%), highlighting the impact of specific neighborhood clusters on rental prices.

## Conclusion

The final Random Forest model achieved a reliable MAE of approximately 0.33 on the test set, establishing its effectiveness in estimating Airbnb rental prices in San Francisco. Notable insights include the significance of 'Accommodates,' emphasizing property capacity, and the importance of 'Host\_duration\_days' and 'Host\_listings\_count' in determining prices. Future work could explore enhancements like incorporating temporal trends or advanced modeling techniques for further refinement.

## Recommendations

### 1. Optimize Property Capacity:

- Property owners should consider optimizing the capacity of their listings ('Accommodates') to align with market demands and potentially enhance rental prices.

### 2. Pricing based on Host Experience & Portfolio:

- The duration of the property owner's tenure as an Airbnb host ('Host\_duration\_days') and the management of multiple listings ('Host\_listings\_count') can significantly impact the rental prices of the properties. Hosts can utilize this information to establish the rental prices for their properties.
- Property owners are encouraged to invest in building expertise and expanding their range of listings to leverage these influential factors in setting competitive and attractive rental prices for their properties.

### **3. Explore Neighborhood Clusters:**

- As 'Location\_cluster' influences prices, hosts and potential renters can explore specific neighborhood clusters that may offer unique pricing dynamics.

## **Future Work**

Future iterations of the project could explore:

- Incorporating temporal trends for more dynamic predictions.
- Experimenting with advanced modeling techniques to further enhance predictive capabilities.
- Continuous monitoring and updating of the model to adapt to evolving market conditions.