Predicting Airbnb Rental Prices: A Data Science Project Overview

Agenda

- Project Overview

- Data Exploration and Analysis

- Model Selection

- Key Feature Importance

- Conclusion

- Recommendations

- Optimize Property Capacity

- Future Work

Project Overview

> This data science project focuses on predicting Airbnb rental prices in San Francisco.

> Utilizes a dataset from insideairbnb.com, spanning from September 2022 to September 2023, containing 7208 rows and 75 columns.

> Primary model used is Random Forest Regression, achieving a Mean Absolute Error (MAE) of approximately 0.33 on the test set.

> The report includes in-depth Data Exploration and Analysis, Model Selection process, and key findings such as Feature Importance and clusters based on location.
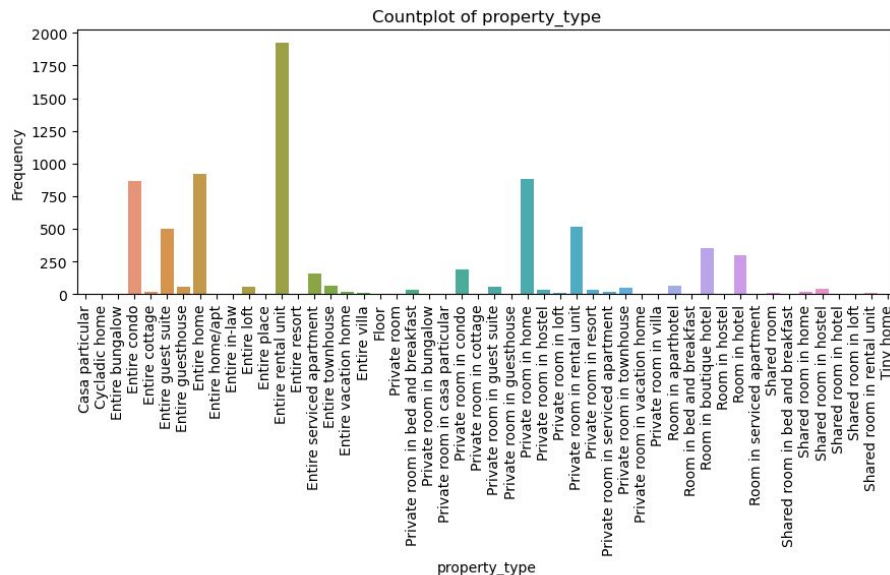
Data Exploration and Analysis

The majority of properties listed on Airbnb offer either entire homes/apartments or private rooms.


Countplot of room_type
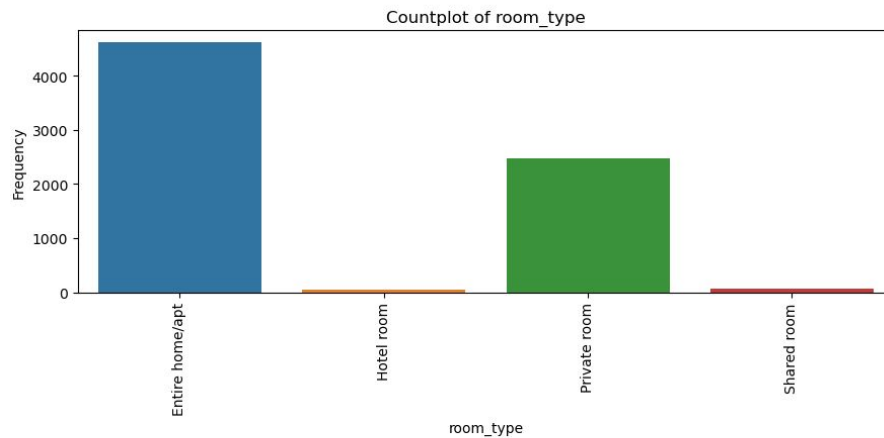
Data Exploration and Analysis

The most common property type in the Airbnb listings is 'Entire rental unit'. Other common types include 'Entire condo', 'Private Room in home', 'Entire home'.

Countplot of property_type

Clustering based on Location

The clustering process revealed that properties in the same cluster tend to have similar rental price dynamics, suggesting a spatial component to the pricing trends.
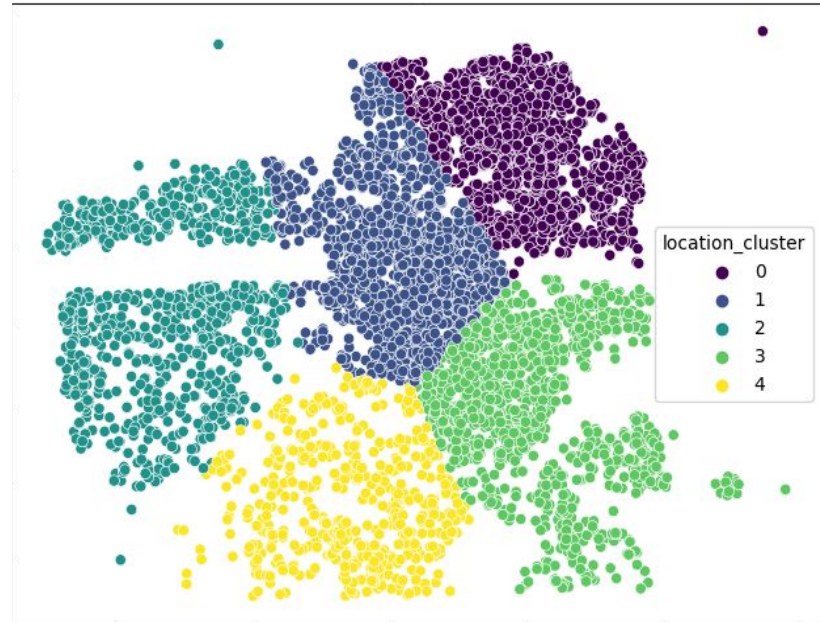
Model Selection

Linear Regression
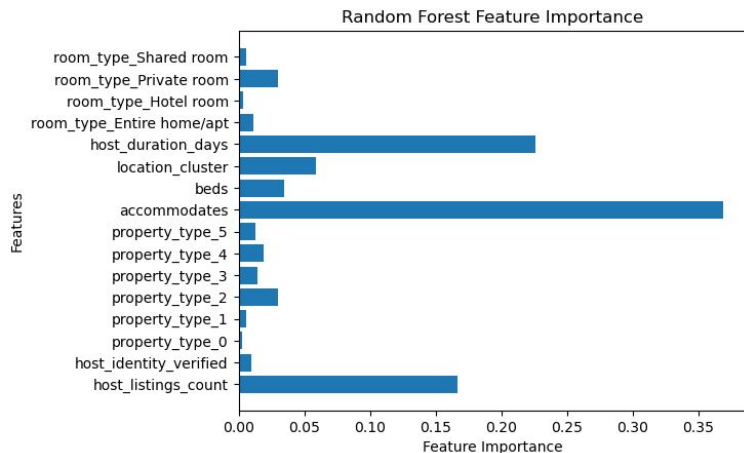
Cross-Validation Mean MAE - **0.4275**

Random Forest Regression

Cross-Validation Mean MAE - **0.3303**

After hyperparameter tuning - **0.3261**

Key Feature Importance

Key features influencing Rental Prices

Insights from Feature Importance

- The feature **'Accommodates'** accounts for approximately **37%** of the model's predictive capability, making it the most influential factor on rental prices.

- **'Host_duration_days'** and **'Host_listings_count'** , the host-related features, also play a significant role, collectively influencing around **40%** of the rental price.

- **'Location_cluster'**, the geographical feature, accounts for nearly **6%** of the price variation, indicating specific impact on certain neighborhood clusters.

- The model's other 40% is influenced by features not specified here, potentially capturing unmodeled neighborhood dynamics or other unique property attributes.

Conclusion

1

The final Random Forest model achieved a reliable MAE of approximately **0.33** on the test set.

2

The model's effectiveness in estimating Airbnb rental prices in San Francisco highlights the significance of 'Accommodates' as a key feature, emphasizing property capacity.

3

Key features like 'Host_duration_days' and 'Host_listings_count' also proved crucial, showcasing the impact of property owner's experience and portfolio on rental prices.

4

The insights provided by the model create opportunities for property owners to optimize their listings and for hosts to establish competitive and attractive rental prices.

Recommendations

Optimize Property Capacity

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

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.

Explore Neighborhood Clusters

'Location_cluster' influences prices, hosts and potential renters can explore specific neighborhood clusters that may offer unique pricing dynamics.

Future Work

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.