



Salford & Co.

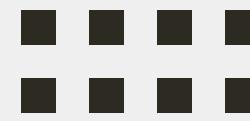
# 2025 INDONESIA HOUSING PRICE PREDICTION

Prepared By:

**Raditya Erlang Arkananta**



Let's Connect!  
[Linkedin Profile](#)



Transforming Data Into  
Actionable Insight

# Table of Contents

---

**01** Introduction

**02** Previous Projects

**03** Main Project

**04** Data & Business  
Understanding

**05** Data Pre-Processing

**06** EDA & Feature  
Engineering

**07** Machine Learning  
Modeling & Evaluation

**08** Insight &  
Recommendation



# About Me

**Raditya Erlang Arkananta**

Data Science Enthusiast

## Experience

- PT Hekikai Indonesia  
QA/QC Japanese Translator Staff
- PT Indonesia Indikator  
HR Officer

**Sep 2024 - Apr 2025**

**Jan 2024 - Sep 2024**

## Education

- Technology Institute of Sepuluh Nopember  
Industrial Engineering

**Aug 2019 - Sep 2023**

- Dibimbing  
Data Science

**Nov 2024 - Jun 2025**

# Previous Projects

---



## Pakistan E-Commerce Customer RFM Analysis

- Identified Seasonal Pattern to Optimize Marketing Effort during peak Seasons.
- Segmented Customers into Groups to focus Marketing Efforts tailored to each Segment.
- Identified the Need to Improve Payment Methods due to high error and Cancellations.
- Discovered Potential Market to push a Segment with Highest Customer Count.

## Customer Sentiment in Ticketing System

- Analyzed customer satisfaction using key metrics such as CSAT, CES, and NPS
- Gathered insights on Customer sentiment and feedback comments by Analyzing comments
- Observed that customer service area need improvement
- Identified the need to make more people recommend the product

## Advertising Impact A/B Testing Analysis

- Analyzed the test results of two groups, each exposed to different ads, and identified the one with significantly better performance
- Calculated the duration and sample size for the A/B test to be held reliable.

# Main Project

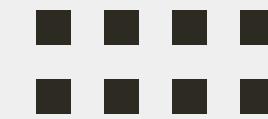
---

## Background

This project was initiated in response to the rising house prices, after COVID, property values have surged at a rate far exceeding wage growth, making homeownership increasingly out of reach for a large segment of the population.

The traditional milestones of financial independence are now being delayed or abandoned entirely due to the financial strain.

The project hopes to provide potential solutions, insight and guidance for those navigating the complexities of today's housing landscape.

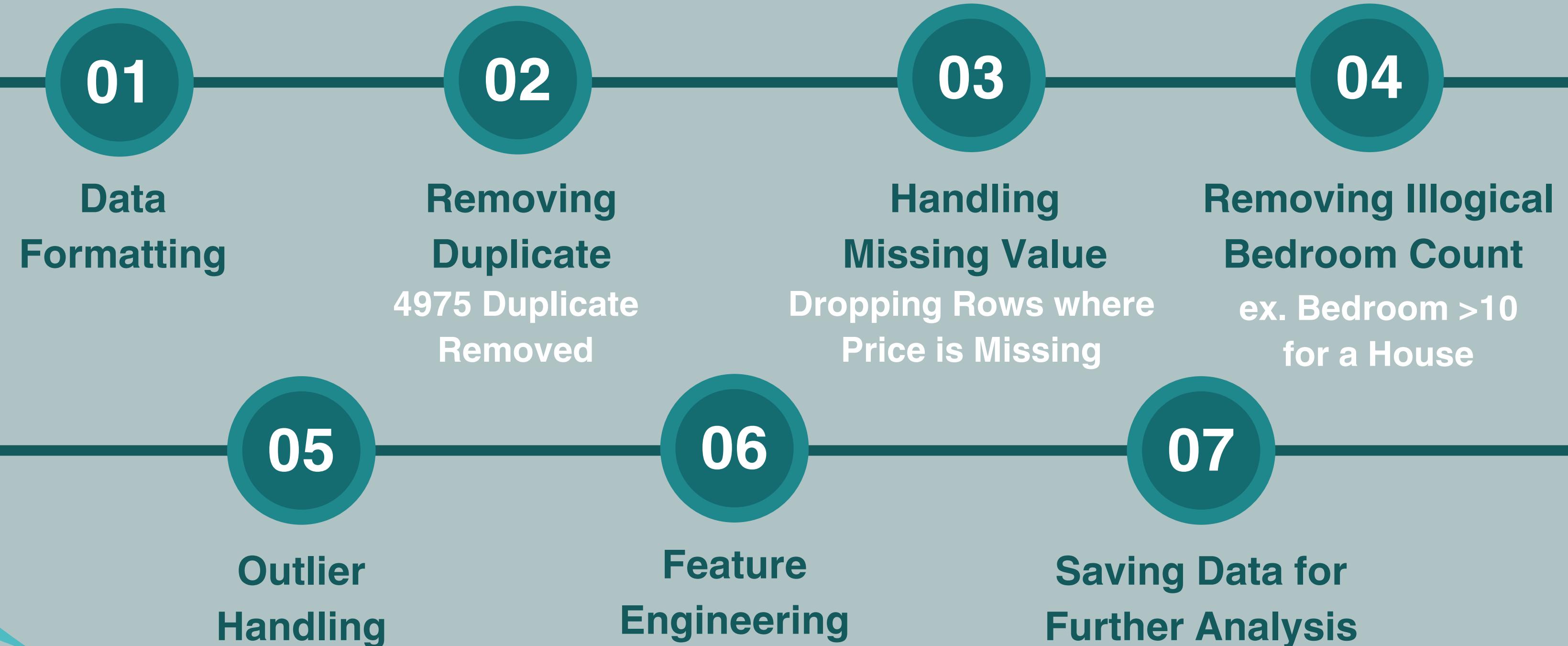


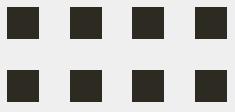
## Goals

The goal of this project is to identify and analyze the key factors that influence housing prices in Indonesia—such as property features, location, and market conditions—and to develop a machine learning model that can accurately predict house prices based on these factors.

# Data Scraping & Pre-Processing

....





# Data Understanding

## About the Data

This dataset, scraped from rumah123.com, contains current property listings for sale in Indonesia, including housing, land, apartments, and other types. It provides valuable insights into real estate prices across various regions.

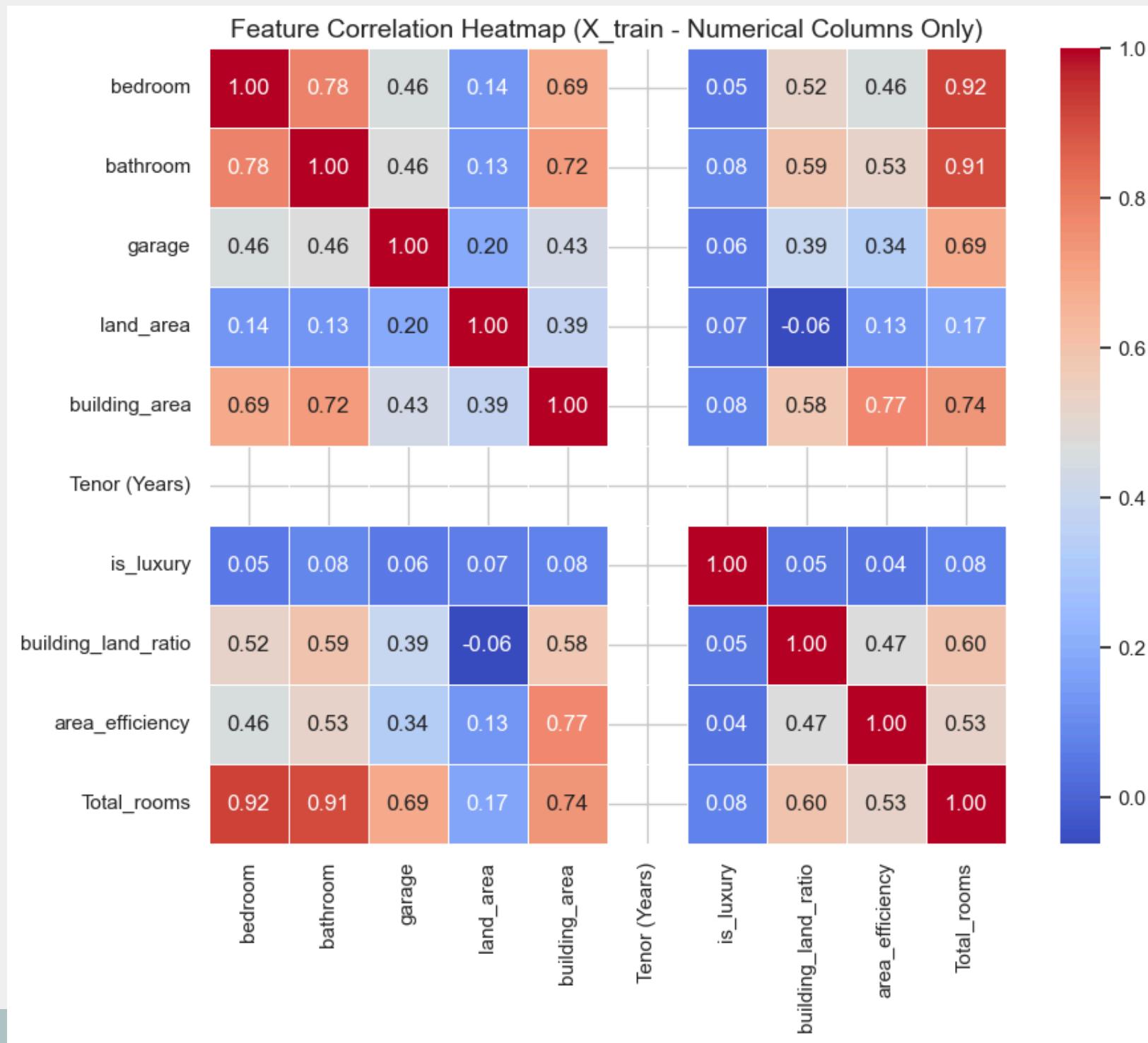
The Dataset Contains the following:

**50000 Rows of Data**

**9 Columns**

**Current available  
Property**

# Multicollinearity Study & Feature Engineering



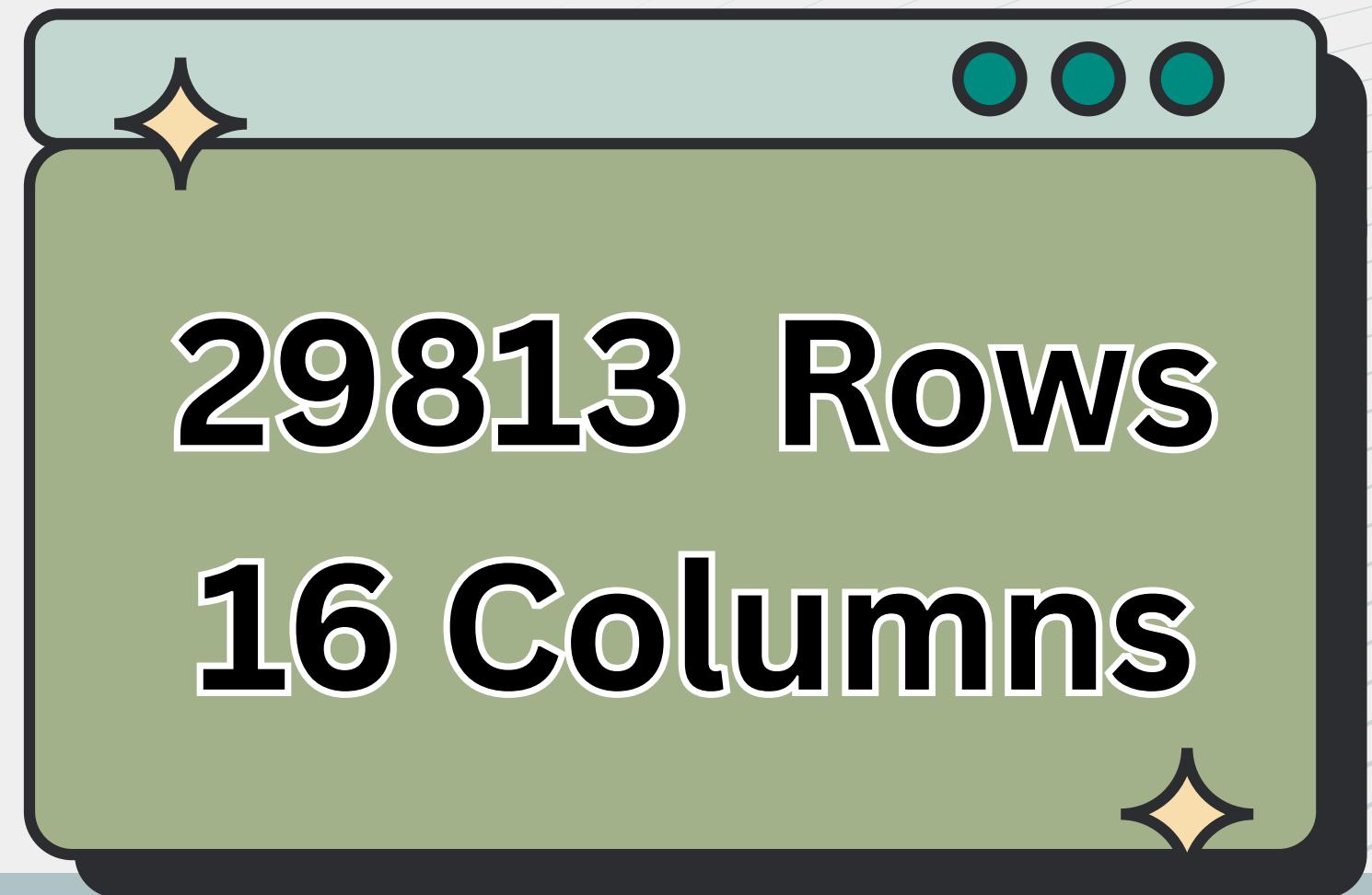
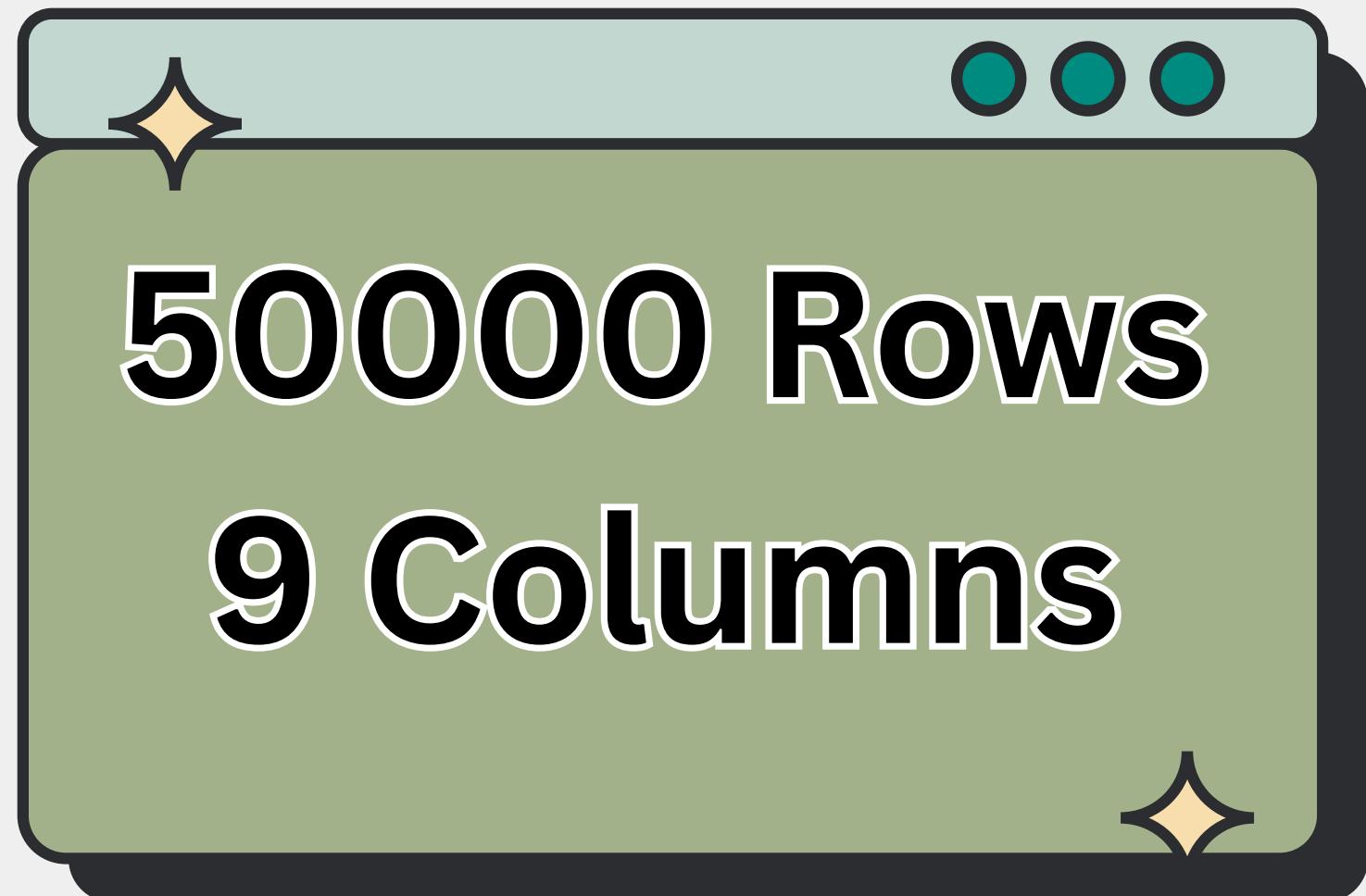
As seen in the heatmap, due to high multicollinearity between Bedroom, Bathroom, and Total Room, we will drop the Total Room column to reduce redundancy.

Additionally, we will perform feature engineering to create new variables:

- Area Efficiency (Building Area per Total Room)
- Is\_Luxury (derived from keywords in the listing name)
- Area Category
- Building-to-Land Ratio.



# Data Pre-Processing Result

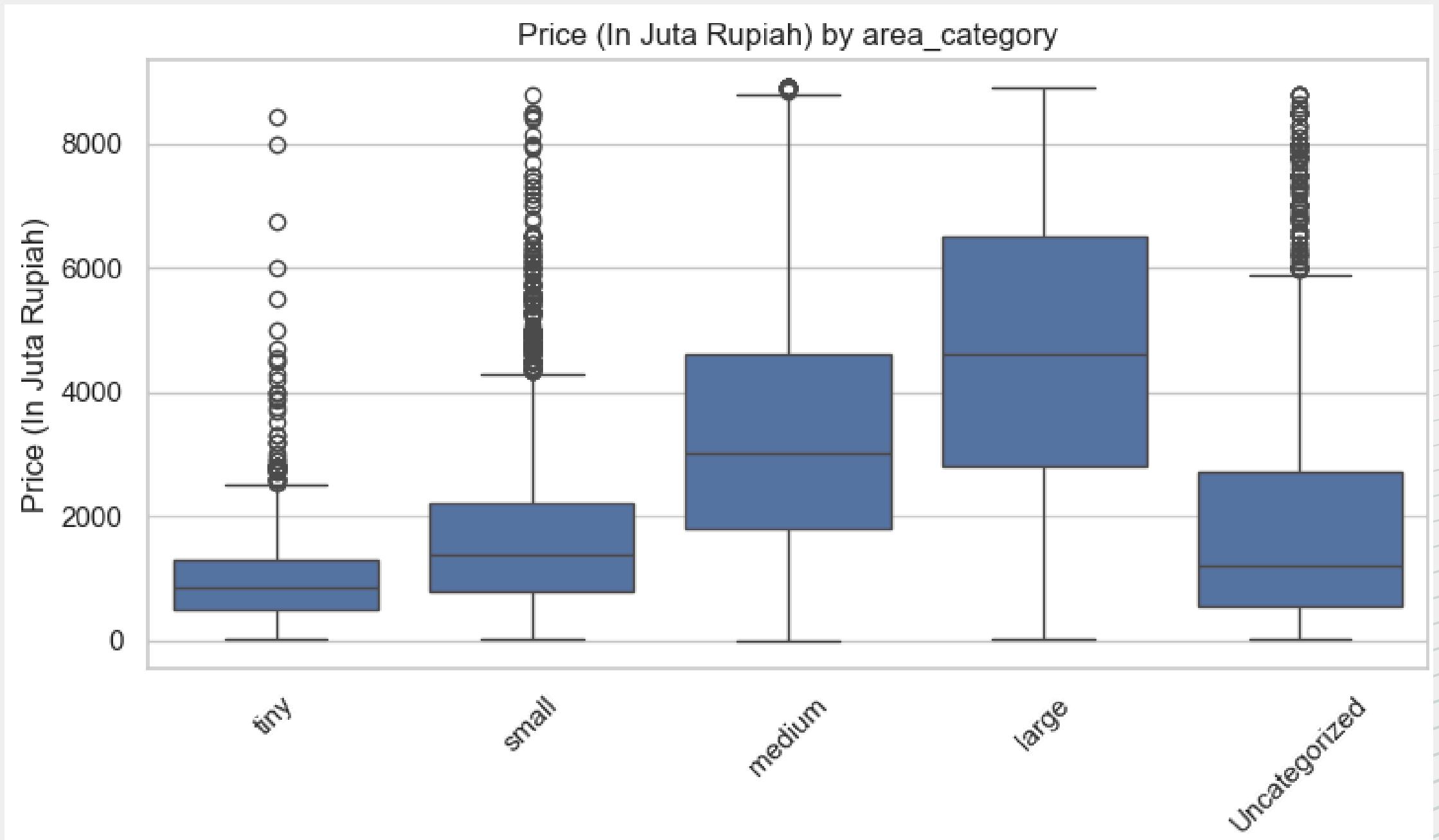


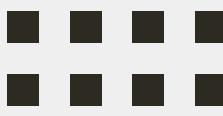
# **EDA - Exploratory Data Analysis**



# EDA - Size Effect to Price

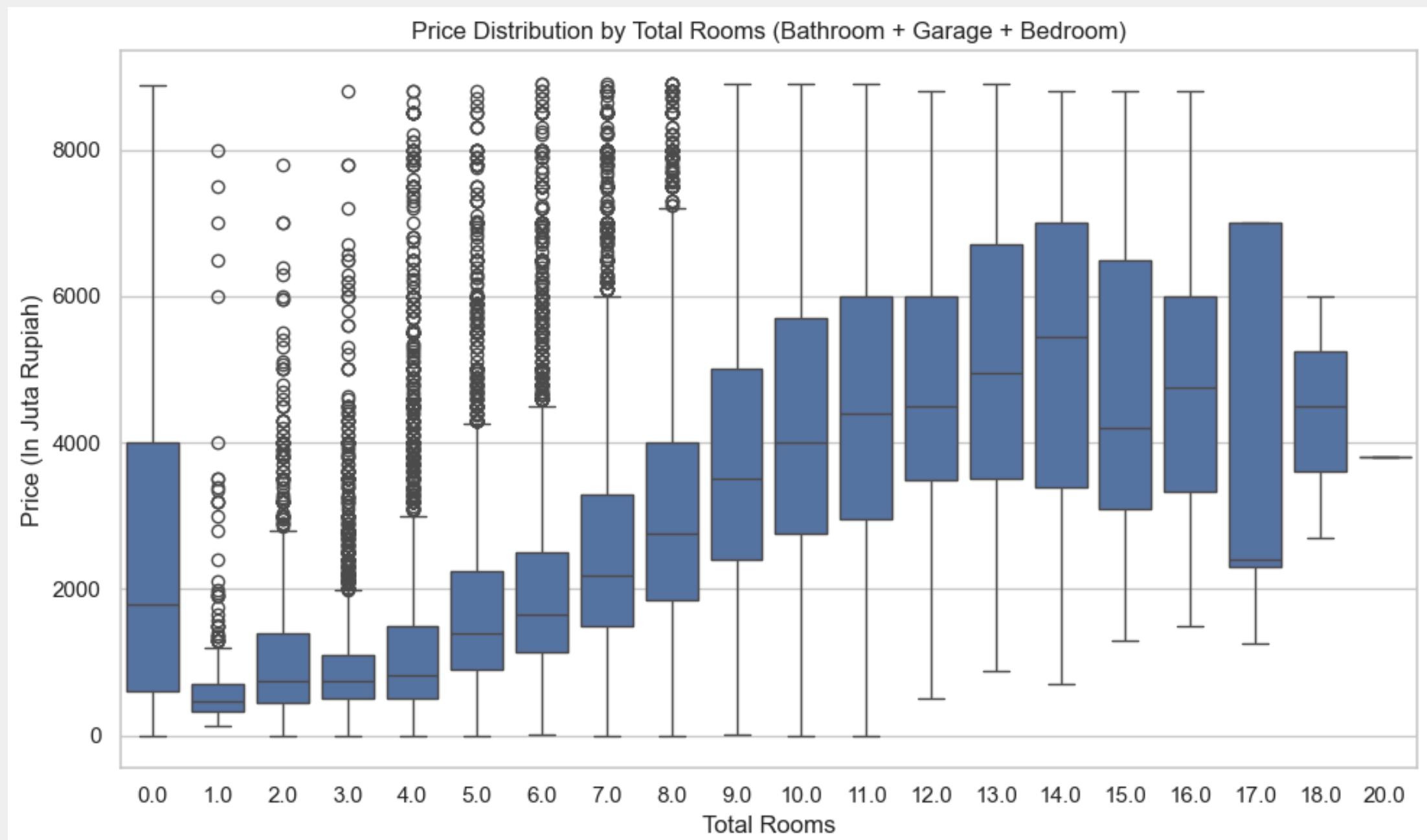
The boxplot shows a clear positive relationship between land size and price, there is some outliers, but larger properties tend to have significantly higher prices compared to smaller ones.





# EDA - Utility Count to Price

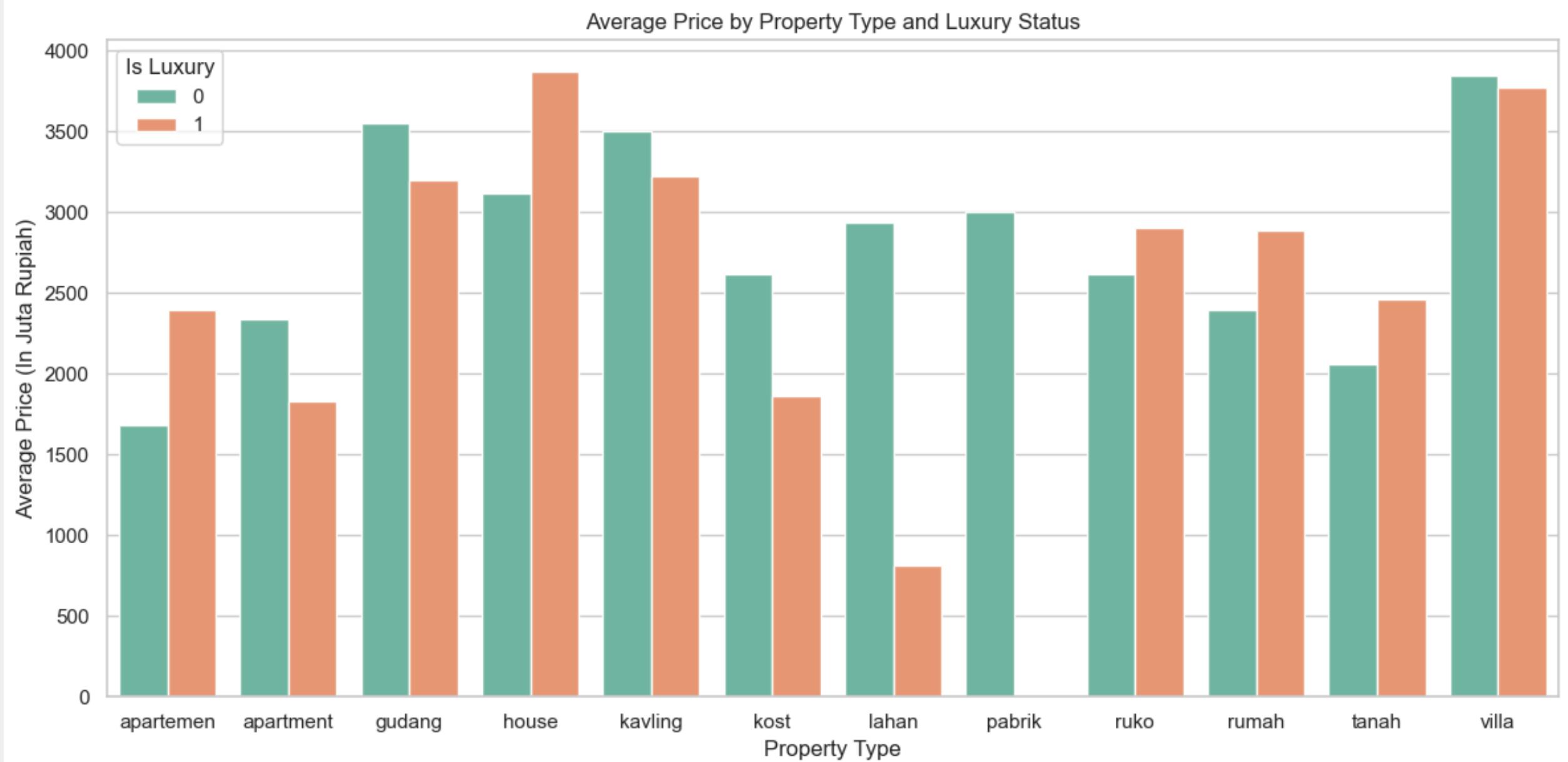
There is a clear positive correlation between the total number of rooms and the average price up to 14 rooms; beyond that point, the relationship becomes less pronounced, its still higher than previous totals, but no longer in a straight line.

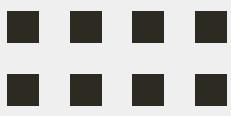




# EDA - Luxury to Price

There doesn't appear to be a clear indication that the "Luxurious" tag significantly affects property prices, which might be due to the way the website formats listings—since the label isn't an official designation but rather inferred from the seller's chosen title.

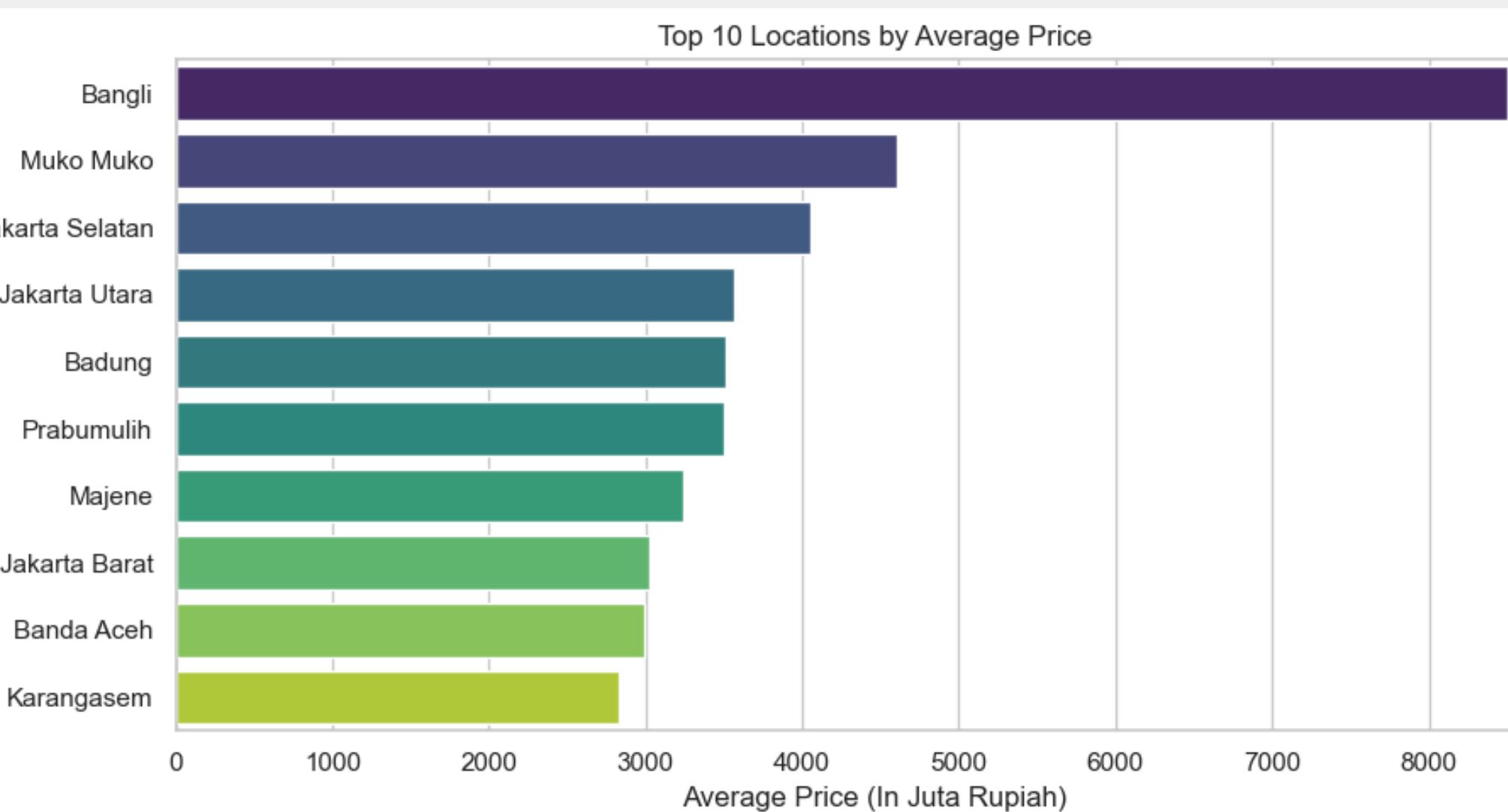




# EDA - Priciest Location

Bangli stands out significantly with an average property price nearing 8,500 juta Rupiah, far exceeding all other locations, suggesting a possible outlier or a small number of high-value listings.

This is followed by Muko Muko, Jakarta Selatan, and Jakarta Utara, with average prices around 3,500–4,000 juta Rupiah. The mix of both urban centers (ex. Jakarta Barat) and more remote or tourist-heavy areas (ex. Karangasem and Badung) indicates that factors beyond just urbanization—such as tourism or listing volume—may be influencing average prices.

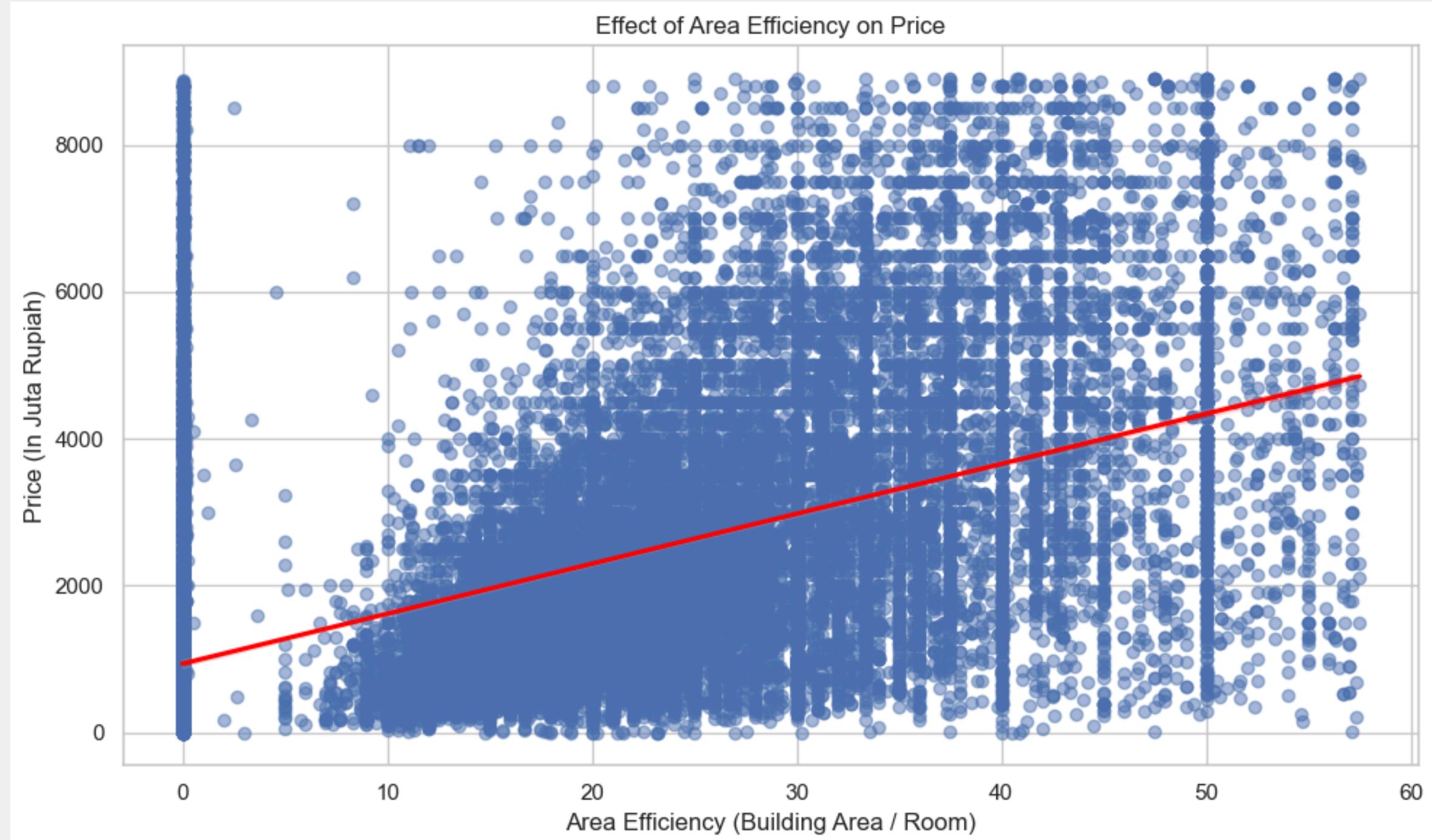


# EDA - Size Utilization to Price

There is a correlation between Area Efficiency and Price, likely because a higher number of total rooms relative to the building size indicates better space utilization, which can make a property more functional and appealing—especially for larger households, increasing its market value.

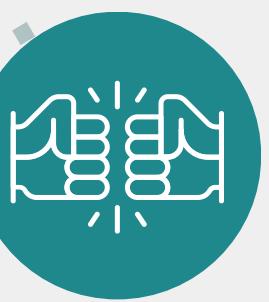
**Pearson Correlation Score = 0.44**

This Score means there is Moderate Linear Correlation between the two, however the other Features are still contributing considerably as well.



# Machine Learning Modeling

# Tested Models



## Ridge Regression

A linear regression model with L2 regularization that penalizes large coefficients to reduce overfitting.

## Random Forest Regression

An ensemble of decision trees to build multiple trees and average their predictions, improving accuracy and reducing overfitting.

## XGBoost Regressor

Gradient boosting algorithm that builds trees sequentially, optimizing for performance and speed using regularization

## GradientBoost Regressor

Builds an ensemble of trees incrementally, each trying to correct the errors of the previous one by minimizing a loss function

## Light GBM

A fast, efficient gradient boosting framework that uses leaf-wise tree growth and histogram-based splitting

# Result & Evaluation



MAE indicates the average errors in predictions, with lower values being better. On the other hand, the  $R^2$  Score reflects how well the model explains the variance in the target variable, with values closer to 1 indicating better performance.

While the Random Forest model achieves the lowest MAE, its performance drops significantly on the test set, suggesting potential overfitting — the model learns the training data very well but struggles to generalize to unseen data.

XGBoost, although not having the lowest MAE, achieves the highest  $R^2$  score on the test set (0.7472), indicating it generalizes better while still maintaining a good balance between bias and variance.

Therefore, we choose the XGBoost model as our final model due to its superior ability to explain the variance in the data while maintaining relatively low prediction error on the test set.

Model	Metric	Training Performance	Test Performance
<b>Ridge Regression</b>	MAE	786.86	806.47
	$R^2$	0.6673	0.6592
<b>Random Forest</b>	MAE	308.11	624.45
	$R^2$	0.9318	0.7439
<b>XGBoost</b>	MAE	452.91	644.13
	$R^2$	0.8877	0.7472
<b>Gradient Boost</b>	MAE	473.56	652.88
	$R^2$	0.8694	0.7447
<b>LightGBM</b>	MAE	494.79	645.67
	$R^2$	0.8531	0.7447

# Evaluation

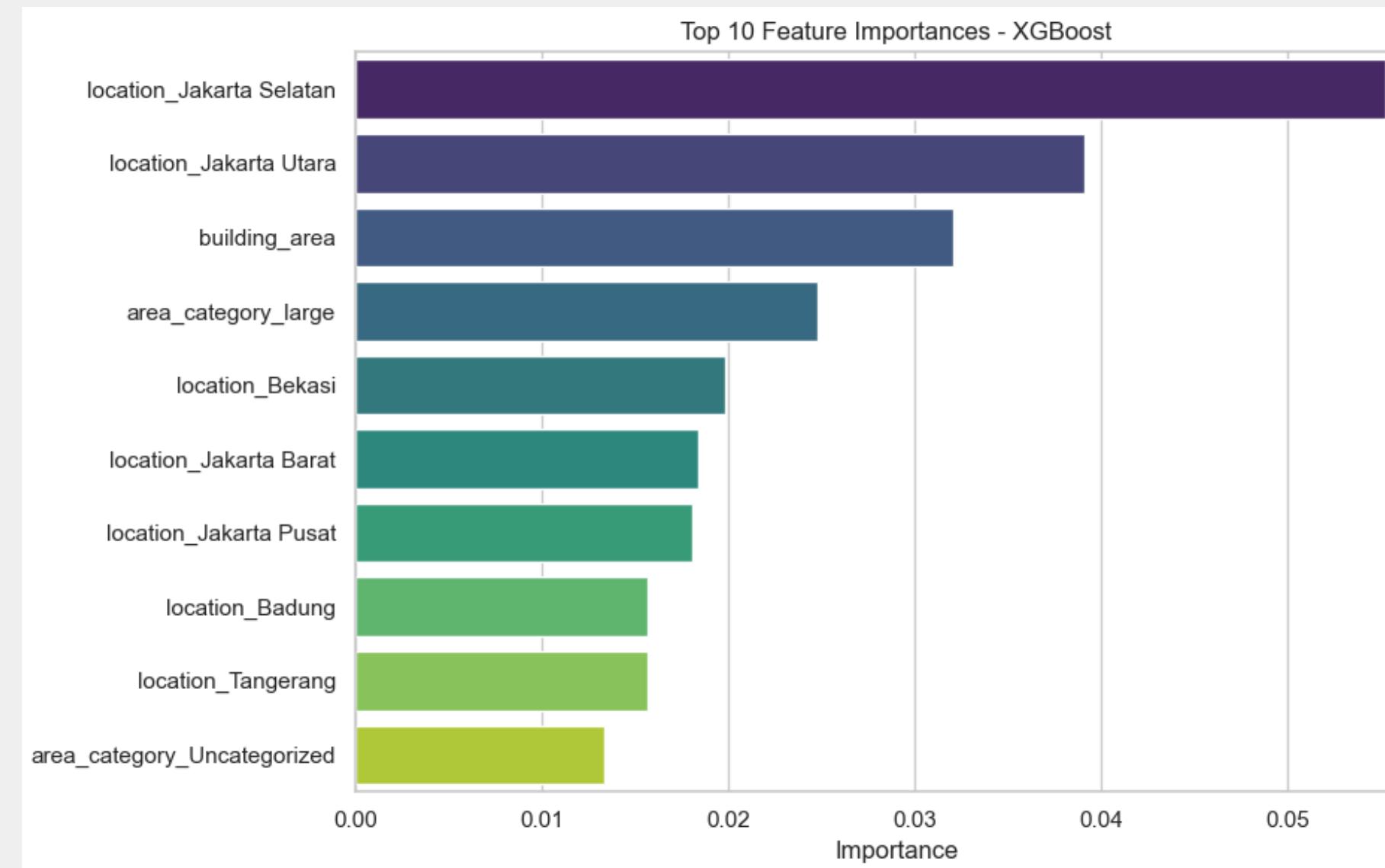


The feature importance plot from the XGBoost model shows that Jakarta Selatan and Jakarta Utara are the top two most influential features in predicting housing prices. This highlights the significance of location, for their prime residential appeal and dynamic urban development.

These regions likely drive up property values due to better infrastructure, amenities, and overall desirability, making location a critical factor in the model.

The next two important features are building\_area and area\_category\_large, both of which relate to the size of the property.

This indicates that larger buildings and properties classified as having a large area also play a substantial role in determining price. Together, these insights show that both where a property is located and how big it is are the dominant factors influencing housing price predictions in this dataset.

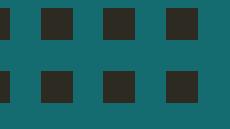


# Application

The application interface shows various input fields for property features:

- Number of Bedrooms: 0
- Tenor (Years): 1
- Number of Bathrooms: 0
- Is it a luxury property?: No
- Number of Garages: 0
- Building-to-Land Ratio: 0.00
- Land Area (m<sup>2</sup>): 0.00
- Area Efficiency: 0.00
- Building Area (m<sup>2</sup>): 0.00
- Total Rooms: 0
- Location: Aceh Besar
- Property Type: tanah
- Area Category: tiny

Predicted Price: 2,247 Million Rupiah (Rp 2,247 Miliar)



Using the Model we Trained, I have deployed a Streamlit App to Predict Indonesia Housing Prices.

**Try it  
Yourself!**

# Insight & Recommendation



- 01 Location-based features have emerged as the most influential factors in predicting housing prices. In particular, South Jakarta (5.5%) leads the importance ranking, followed by North Jakarta (3.9%) then Building and Land Size Follows
- 02 Bedrooms, bathrooms, and garage spaces do influence property prices, but their effect diminishes after a certain point and is less impactful compared to location.
- 03 Space efficiency (how well the available area is used) tends to positively affect property pricing, with more efficient designs often Priced higher
- 04 Location that are Top Urban Center such as Jakarta and Top Tourist Spot such as Badung tend to be the most Important Feature.
- 05 Even for homes with similar specifications (size, room count, features), property prices can vary widely across cities, emphasizing the importance of geographic context.

# Recommendation



01

For Real Estate Agents, it's important to improve listings and user experience because location has become the most influential factor in property pricing, they can add things like school and transportsations nearby.

02

For Investors and Stakeholders Invest in properties located in high-demand areas—such as South Jakarta, North Jakarta, and Bekasi, Monitor urban expansion and infrastructure projects to identify emerging hotspots, focusing on improvements in accessibility and livability

03

For Customers and Home Buyers When evaluating properties, prioritize location over size, as where a property is situated often impacts value more than its dimensions Use region-based price comparison tools to identify area-specific price trends and determine if you're paying a premium or getting a deal



# Thank You For Attending

**Let's Connect!**



[wa.me/6281218900315](https://wa.me/6281218900315)



[Github](#)



[LinkedIn](#)

