

# PREDIKSI HARGA HOTEL

JARINGAN HOTEL KYOZO

---



Edward Salim



Fathi Qushoyyi



Fakhri Robbani

Universitas Indonesia

# Latar Belakang

**10%**

**PDB secara  
global**

**\$4.548 milyar**

**Market industri perhotelan  
global**

**7%**

**Pertumbuhan tahunan  
2021 – 2025**



Bagaimana Jaringan Hotel Kyoza dapat mengembangkan hotel-hotel baru dengan **harga kamar yang kompetitif**?



# Tujuan dan Manfaat?

## Tujuan

- Mengembangkan model prediksi harga yang akurat.
- Meningkatkan pemahaman terhadap faktor-faktor yang mempengaruhi harga hotel.

## Manfaat

- Pengambilan keputusan yang lebih baik.
- Pengembangan hotel yang efisien.
- Keunggulan kompetitif.
- Peningkatan pendapatan dan keuntungan.







# Metode Analisis Data

PYTHON & GOOGLE COLAB



# Part 1



**Handling NaN dan  
Baris Duplikat**



**Ekstraksi Fasilitas  
dan Rating**

**Data Wrangling**



**Penambahan Fitur  
Baru yang Relevan**

- facilities\_count
- rating\_ratio
- rating\_review\_combined



**Encoding Kolom  
Kategorikal**

- Ordinal Encoding
- One-Hot Encoding

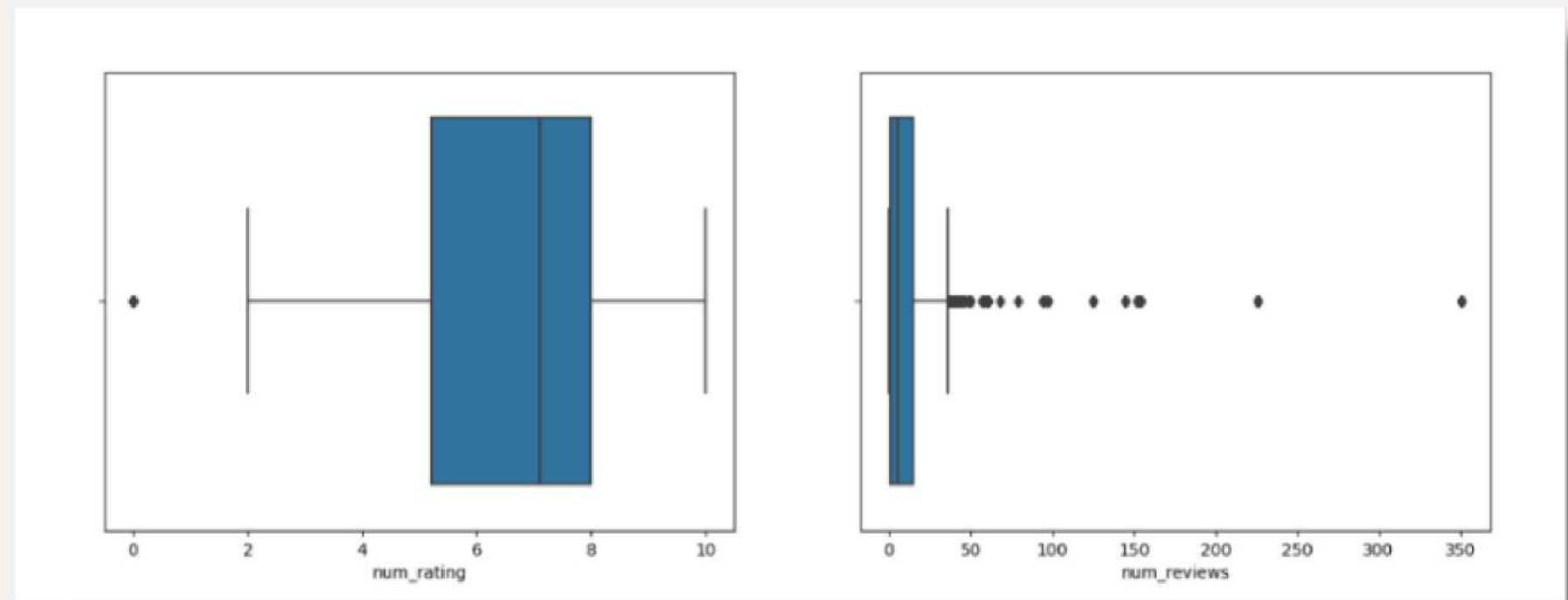
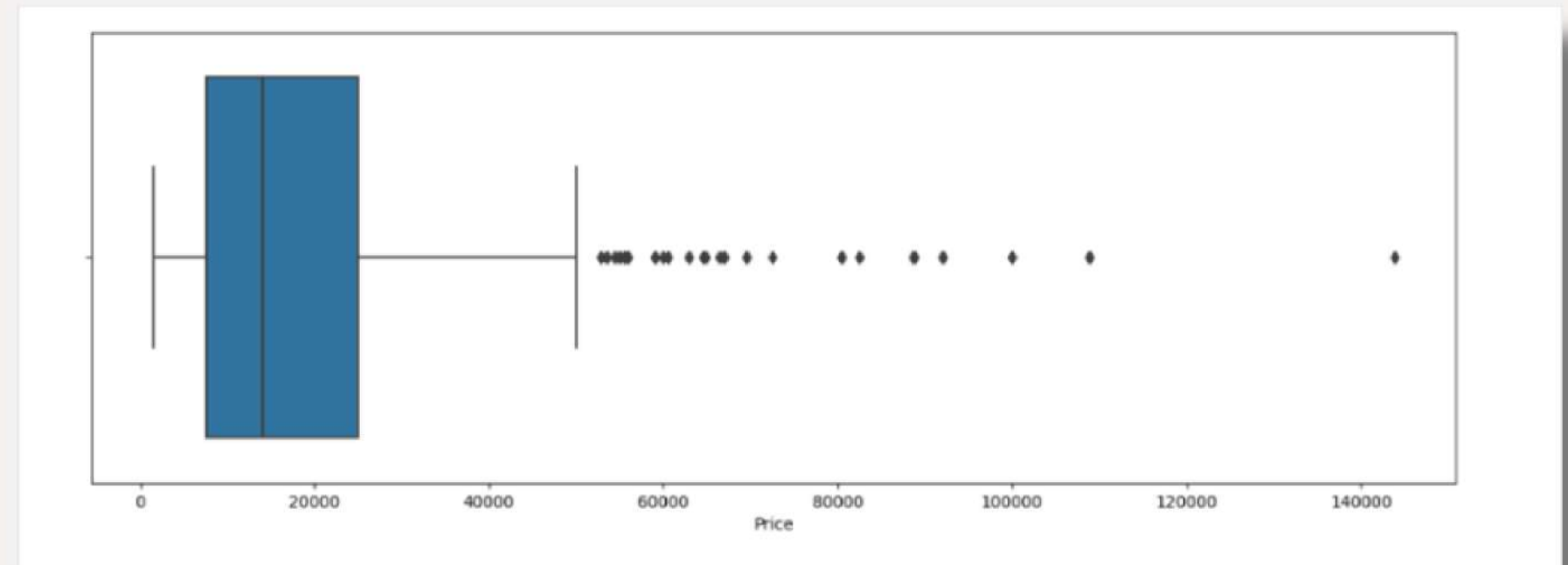
## Part 2:

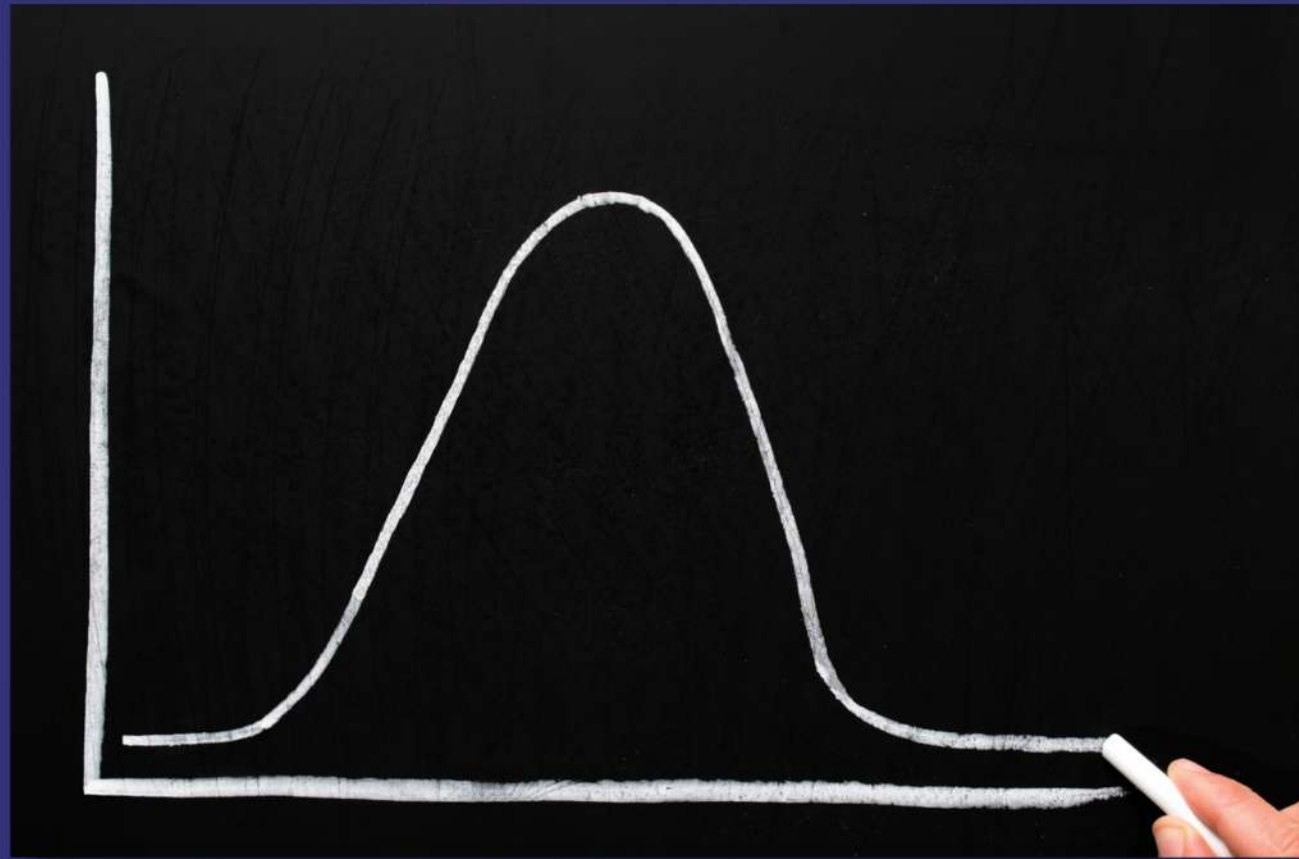
# PENANGANAN OUTLIERS DATA TRAIN

*Outliers* memiliki pengaruh signifikan pada hasil prediksi model regresi linear. *Outliers* dapat menyebabkan model menjadi tidak stabil / akurat.

Rentang Batas Outliers:

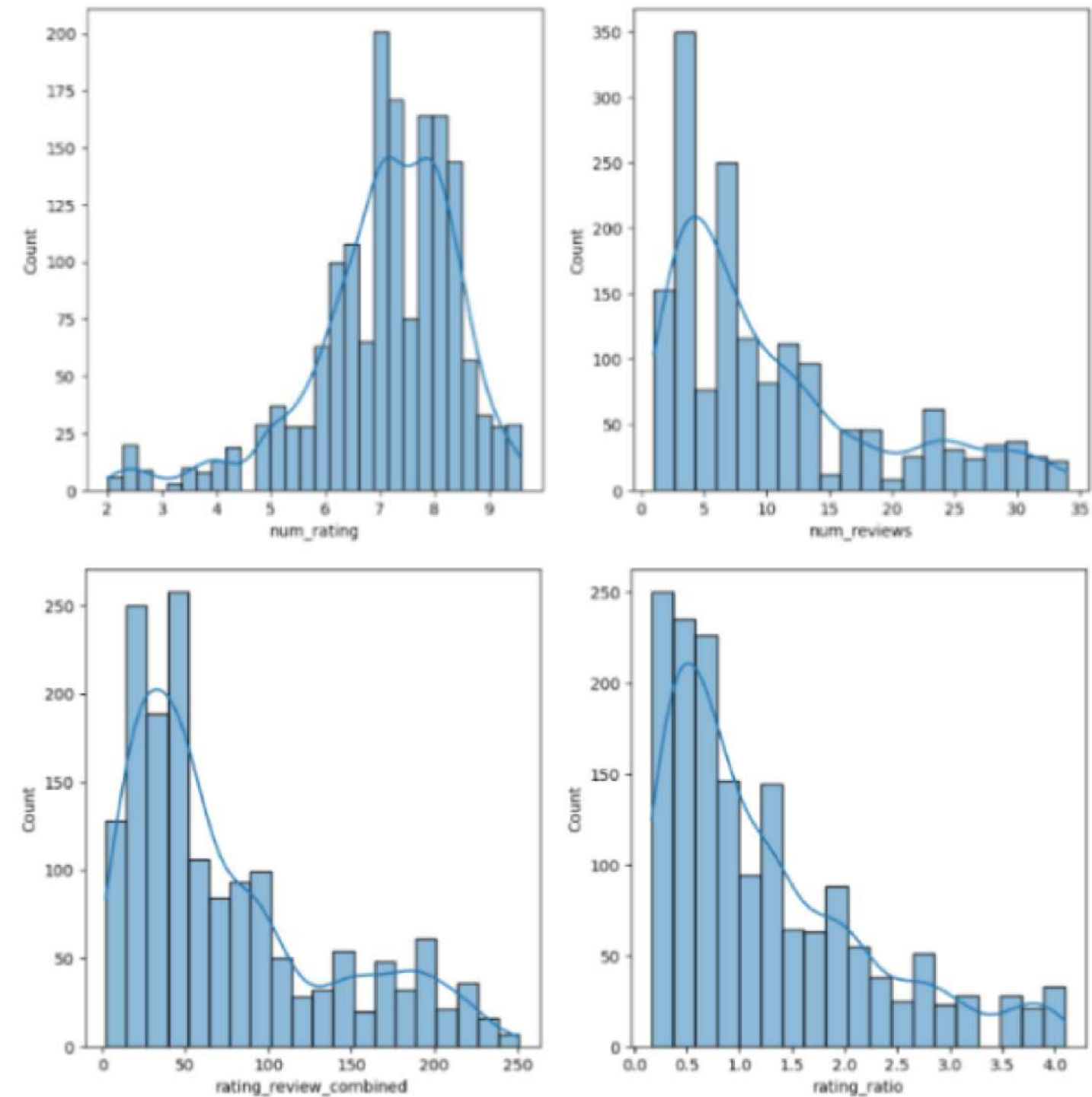
- **Lower bound:**  $Q1 - 1.5 \text{ IQR}$
- **Upper bound:**  $Q3 + 1.5 \text{ IQR}$
- **IQR:**  $Q3 - Q1$





# Part 3: Feature Scaling

dengan Normalization



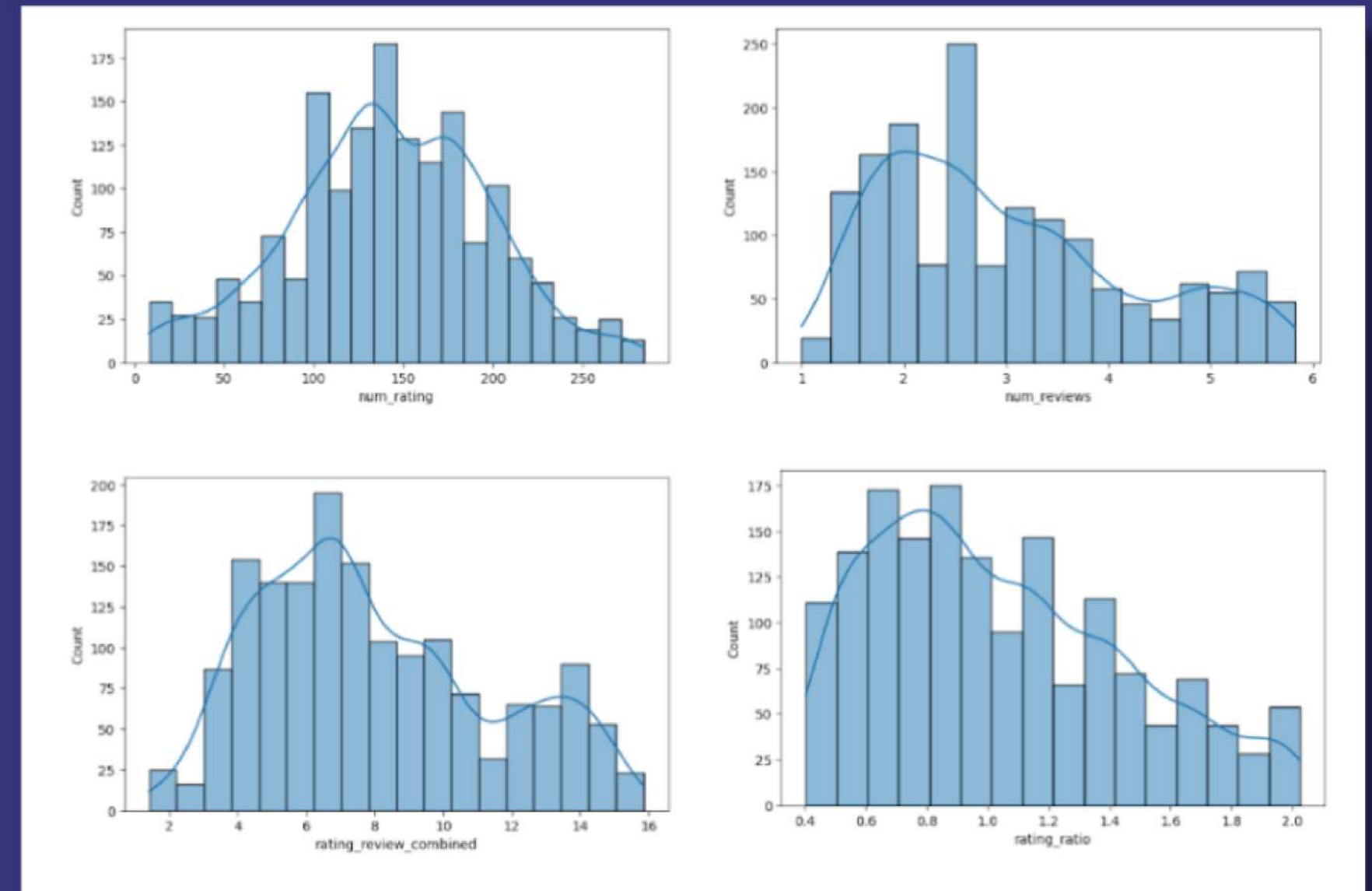
*Sebelum Normalization*





# Part 3: Feature Scaling

dengan Normalization



*Setelah Normalization*



Model	MAE Score
LinearRegression	6937.04
LassoCV	<b>7013.03</b>
RidgeCV	6943.48
BayesianRidge	6939.73
ElasticNet	7362.28
XGBRegressor	<b>825.07</b>

Menggunakan metode **KFold** dengan 10 split

Part 4

## Pemodelan dan Evaluasi Kinerja

**Fitur-fitur yang digunakan (X):**

- has\_swimmingpools
- has\_pool
- num\_rating
- num\_reviews
- rating\_review\_combined
- rating\_ratio
- facilities\_count



# XGBoost model terbaik?

\*untuk prediksi harga hotel\*



## Part 5

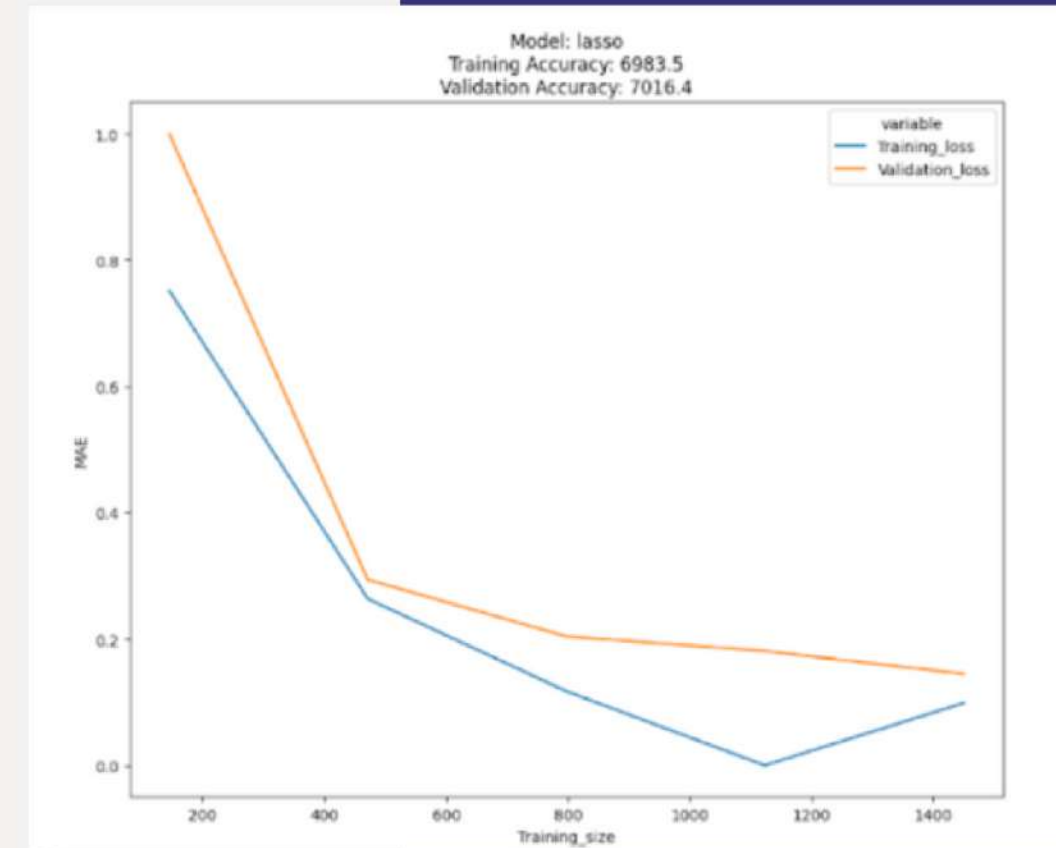
# Pengecekan Overfit dan Underfit

### Learning Curve yang Good

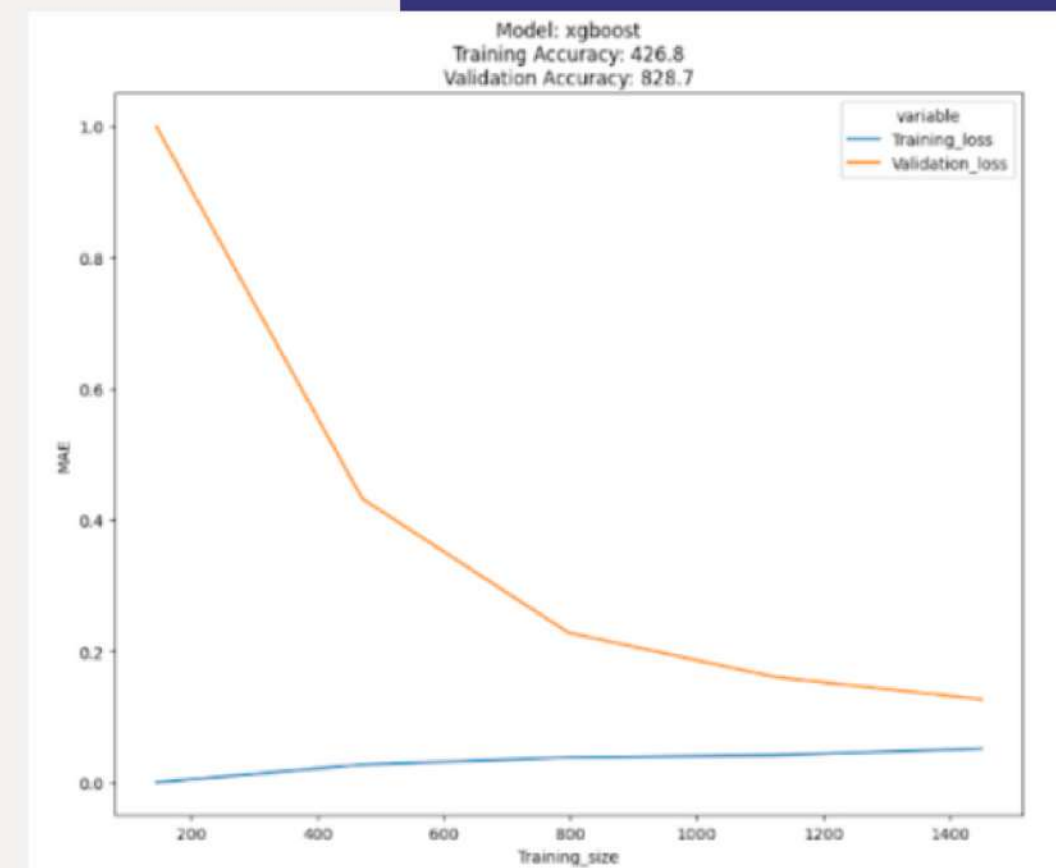
- Training dan validation loss saling mendekati
- Validation loss > training loss
- Awalnya penurunan bertahap
- Loss stabil pada titik tertentu

### Learning Curve yang Overfit

- Perbedaan signifikan training dan validation loss
- Validation loss terus menurun (tidak stabil)
- Training loss sangat rendah dan stabil



VS



LassoCV

XGBoost

**Note:** default parameter

**10990.70**

Private  
leaderboard

**11528.50**

Public  
leaderboard

Minim potensi  
***overfitting*** dan ***bias***  
pada model ini

\*Platform **Kaggle**\*



Streamlit

meefx-hotel-prediction-findit-main-kwvry3.streamlit.app/#enter-the-characteristics-of-the-hotel

## Hotel Price Predictor

Enter the characteristics of the hotel:

Number of Rating:

1.00 - +

Number of Reviews:

1.00 - +

Location:

Stokol

Provided facilities (Can more than one):

☐ Swimming Pools

☐ Pools



[ristek.link/PriceModel](https://ristek.link/PriceModel)

---

# Live Demo

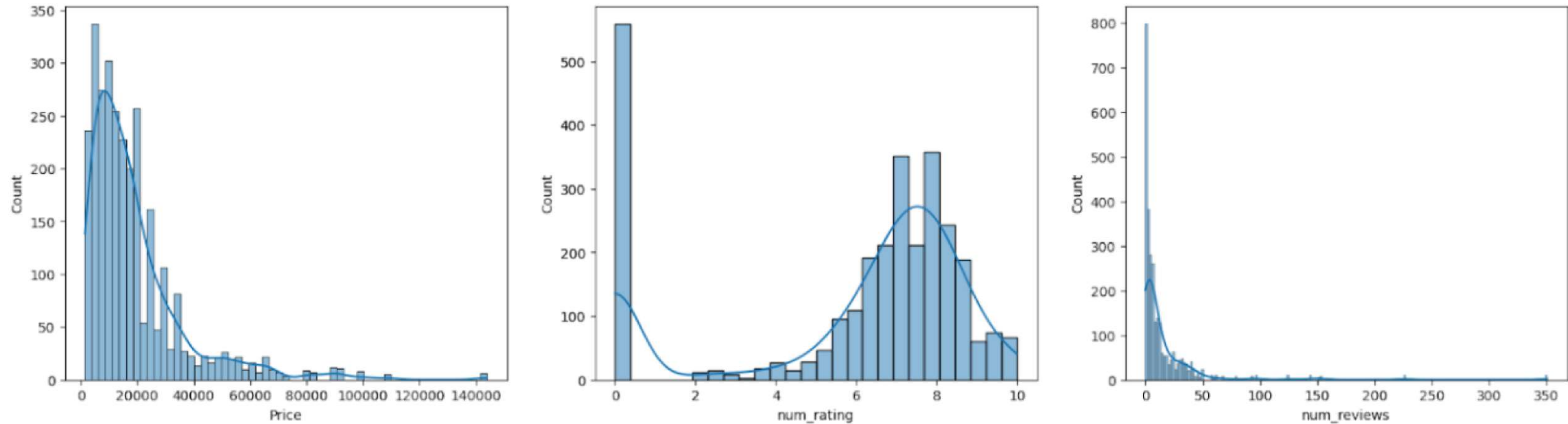


# Analisis

UNIVARIATE & BIVARIATE



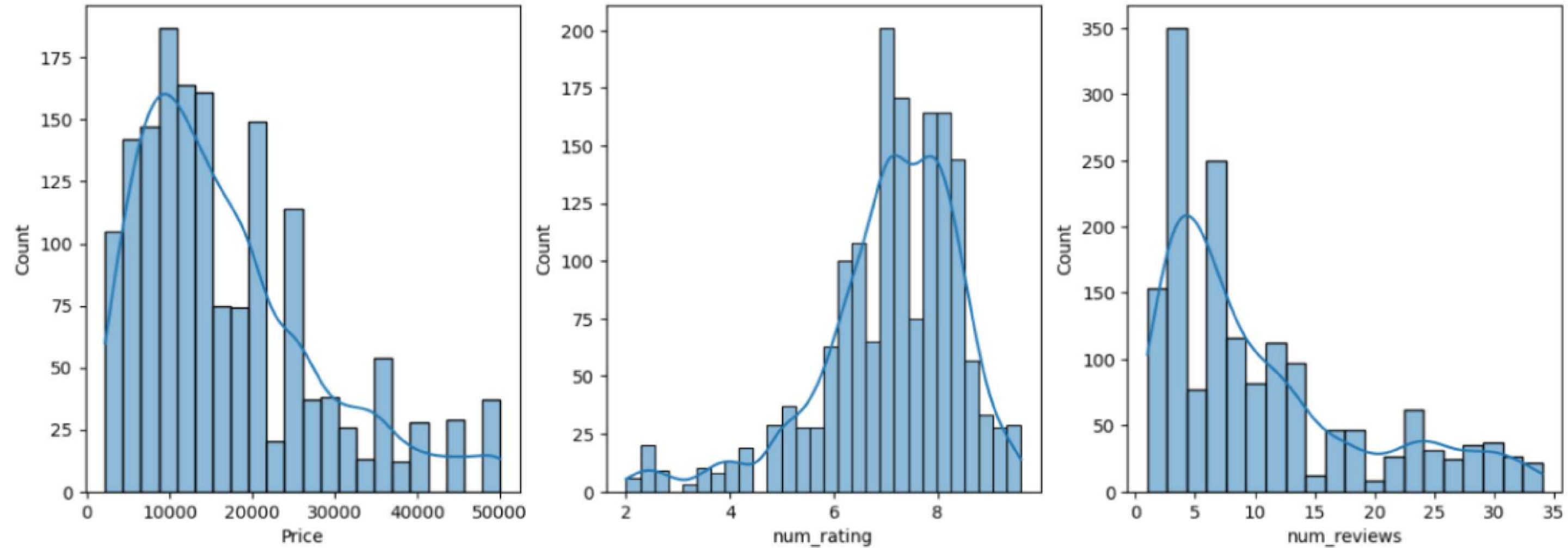




---

# Distribusi Numerikal

**Sebelum** drop outlier



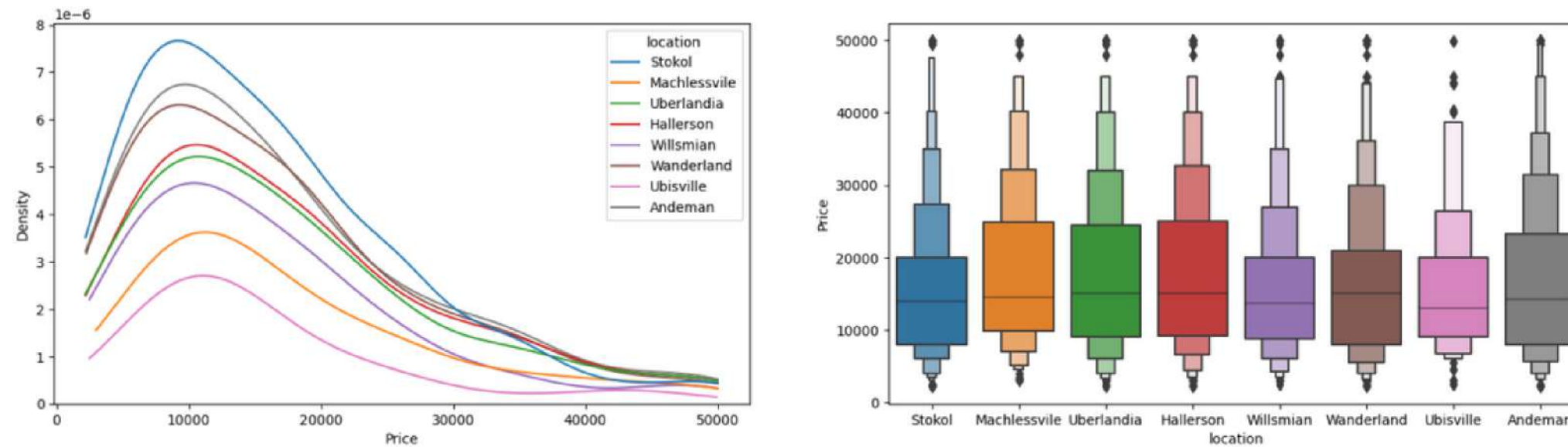
---

# Distribusi Numerikal

**Sesudah** drop outlier



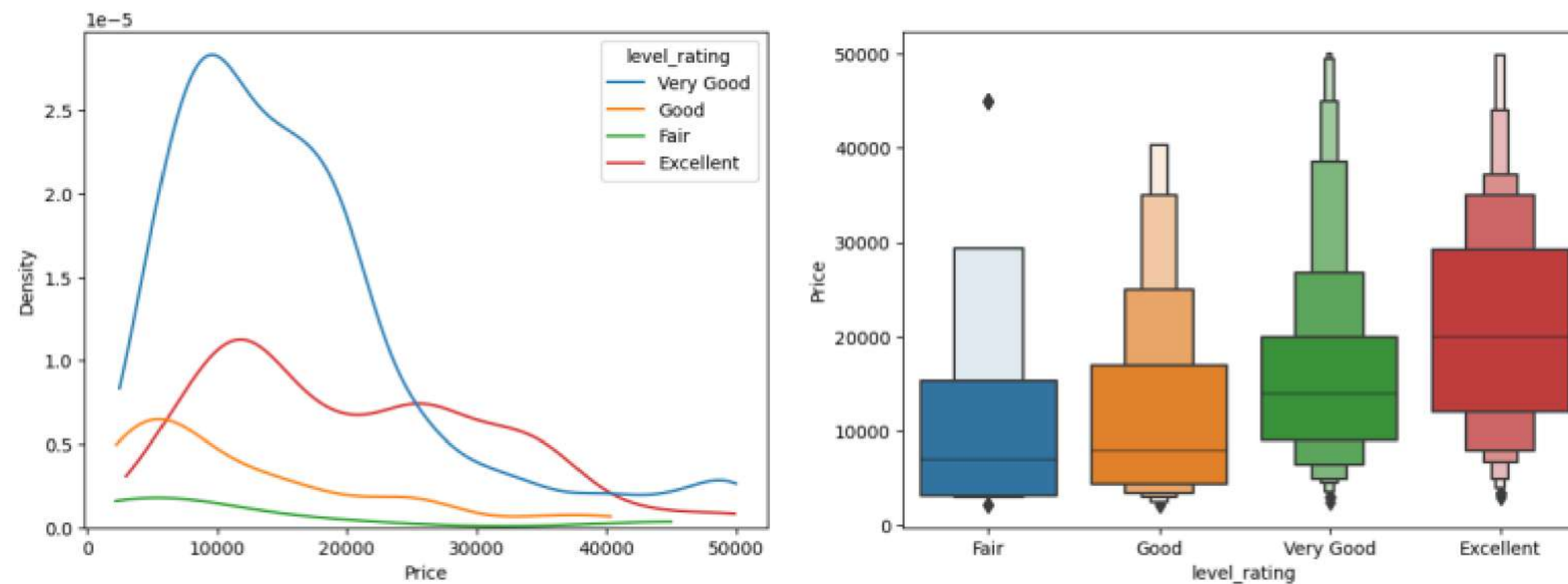
Hubungan Location dengan Price Setelah Remove Outliers



# Pengaruh Lokasi

terhadap **Harga Hotel**

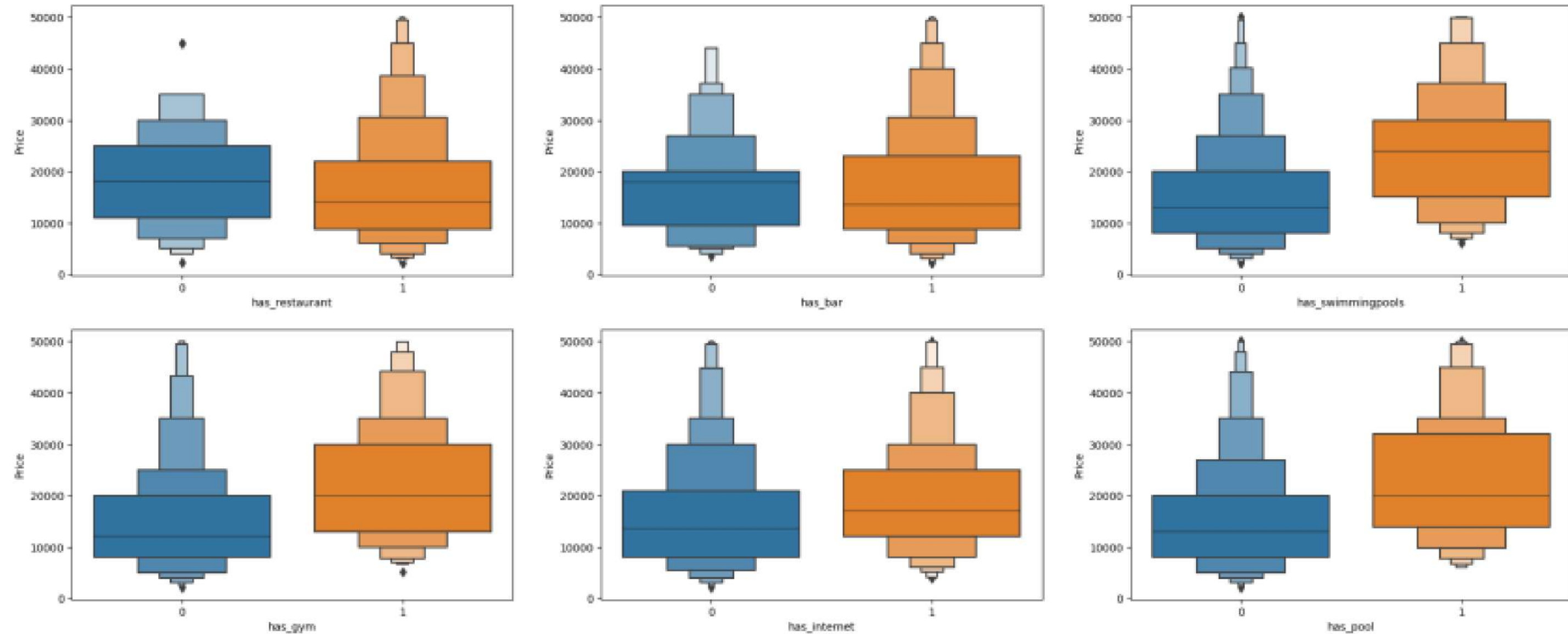
Hubungan Level Rating dengan Price Setelah Remove Outliers



# Pengaruh Level Rating

terhadap **Harga Hotel**

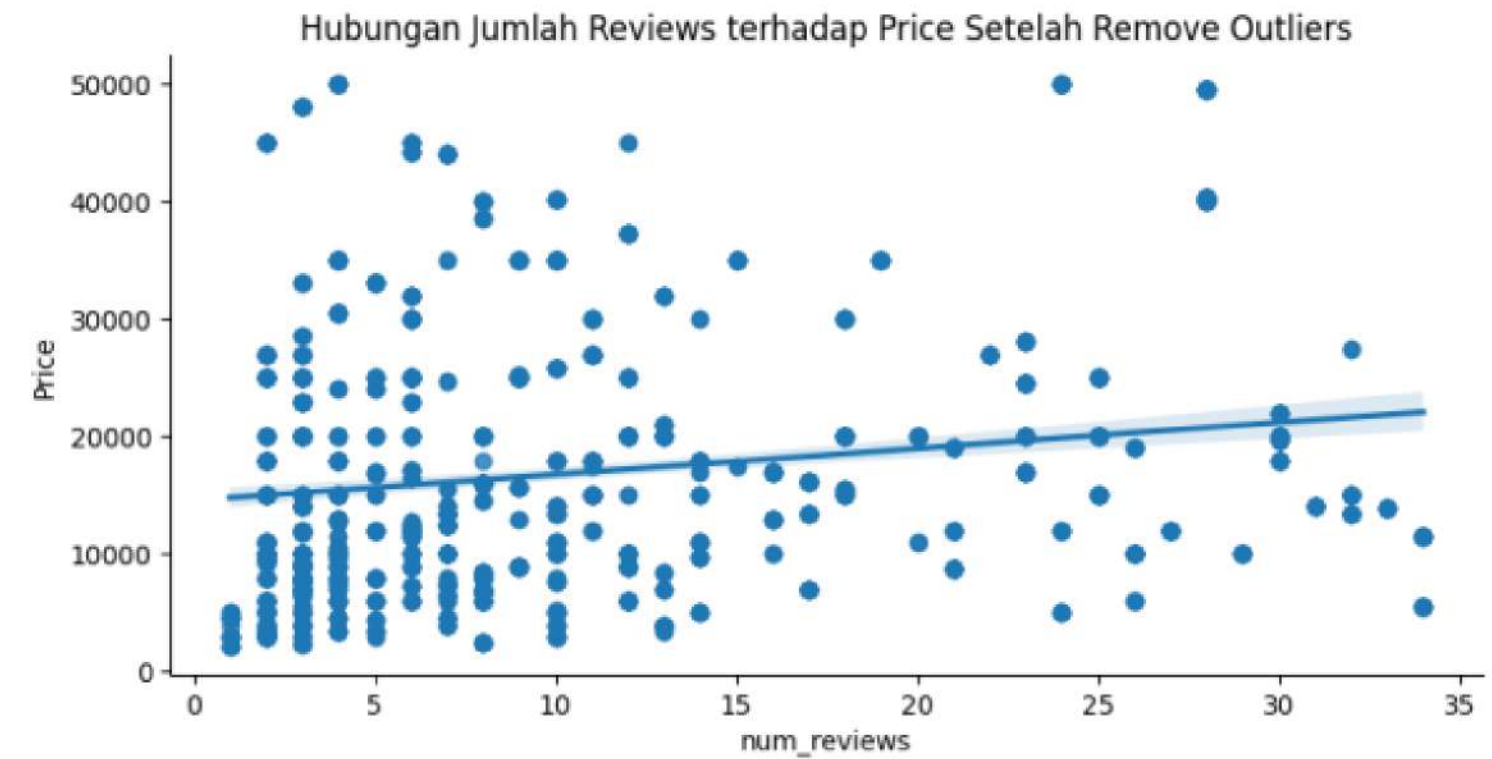
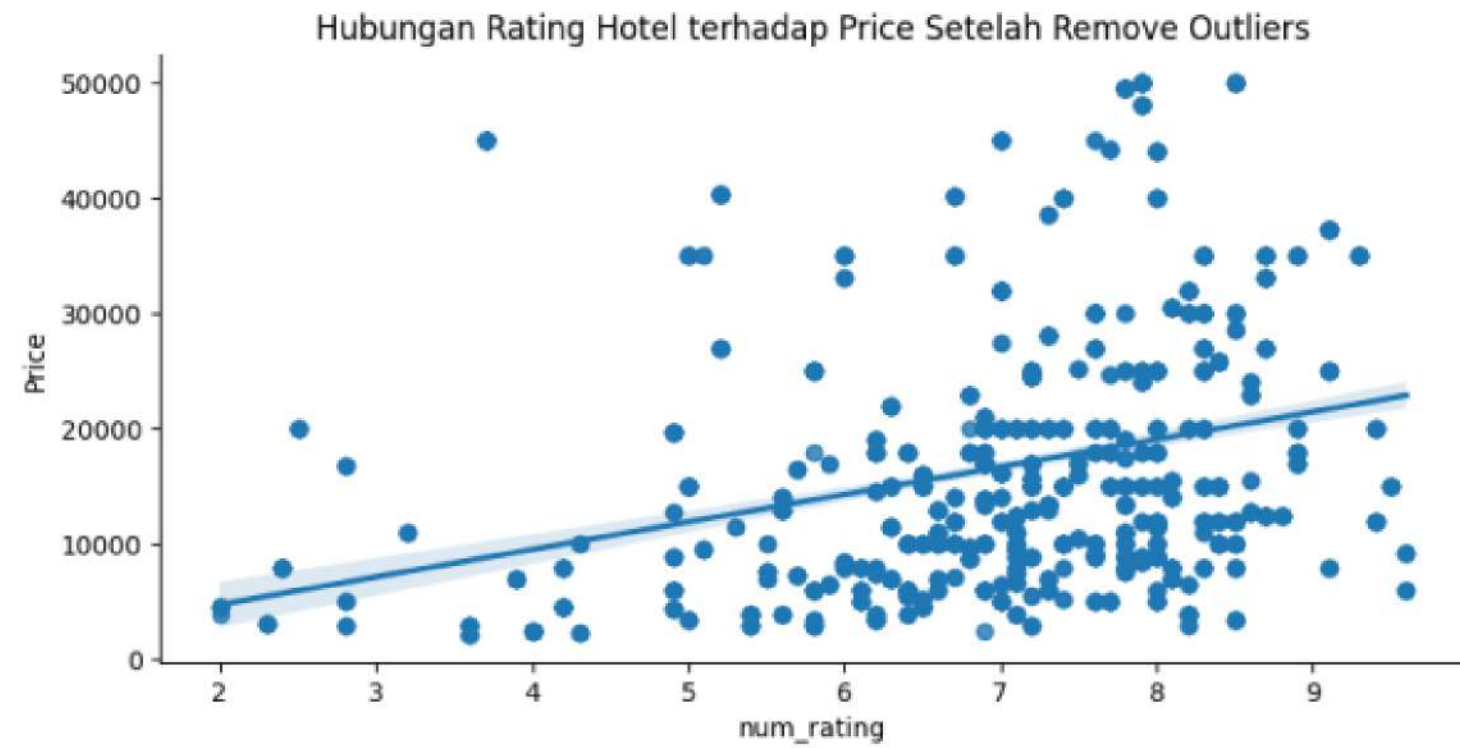
### Pengaruh Setiap Fasilitas terhadap Price Setelah Remove Outliers



# Pengaruh Fasilitas

terhadap **Harga Hotel**





---

# Pengaruh Rating dan Reviews

terhadap **Harga Hotel**

## Fitur yang **DIBUANG**



has\_bar

has\_internet

location

has\_restaurant

has\_gym

level\_rating

## Fitur yang **DIPAKAI**



num\_reviews

has\_pool

has\_swimmingpools

rating\_ratio

rating\_review\_combined

facilities\_count

num\_rating



# Kesimpulan

---



## 1. Model

**LassoCV** mampu menyesuaikan terhadap data *train*

+ Memiliki kemampuan **generalisasi** yang baik.

## 2. Analisis

*Outlier* menurunkan akurasi prediksi model.

"Beberapa" fasilitas, **jumlah review, rating**, dan **jumlah fasilitas** berpengaruh besar.

## 3. Penggunaan

Model dapat digunakan dalam penetapan harga kamar hotel baru.

- Memastikan **kompetitivitas** dan **kepuasan** pelanggan.



# Terima Kasih

\*QnA Session\*



---

# Daftar Pustaka

**McCain, A. (2023). 25 Hotel Industry Statistics [2023]: Hotel Rate Trends And Market Data. Zippia.**

**Kirenz, J. (2021). Lasso Regression with Python | Jan Kirenz. Jan Kirenz.**

**Frost, J. (2017). Choosing the Correct Type of Regression Analysis. Statistics By Jim**

**KSV Muralidhar (2021). Learning Curve to identify Overfitting and Underfitting in Machine Learning. Medium**

**Jordan Hollander (2022). 75+ Hospitality Statistics You Should Know (2023).**