
Rent Holiday Houses

(Topik 3)

Kelompok Regu Buaya



Anggota Kelompok



01 **Galan Gahara Triguna**

2006597121



02 **M. Ihsan Al Farisi**

2006596693



03 **Yosabad Torando S.**

2006597046

Table of Contents

01

Project Background

02

Overview & EDA

03

Data Preprocessing

04

Modelling

Project Background

01 Problem Statements

Semakin banyak variasi penyewaan rumah liburan membuat wisatawan sulit untuk menentukan rumah liburan yang akan mereka sewa

02 Goals

Mengetahui tipe ruangan sesuai dengan fasilitas yang diberikan, mengetahui harga rumah sesuai dengan fasilitas yang diberikan, dan mengetahui segmentasi review

03 Objective

Melakukan analisis menggunakan metode klasifikasi, regresi, dan klaster

Overview and EDA

Dataset

- Terdiri dari 5834 baris data dan 22 fitur
- Memiliki 14 fitur numerik dan 8 fitur kategori
- Memiliki lebih dari 30% data yang kosong
- Memiliki 158 data duplikat

Dataset

Numerical Features

Feature	Description
accommodates	Jumlah maksimal tamu dalam fasilitas
bathrooms	Jumlah kamar mandi
bedrooms	Jumlah kamar tidur
beds	Jumlah kasur
calculated_host_listings_count	Jumlah post host masuk trending list
guests_included	Jumlah tamu diluar accommodates
host_listings_count	Jumlah penyewa menekan post milik host
maximum_nights	Maksimal waktu meningap (dalam satuan malam)
number_of_reviews	Jumlah review
review_scores_checkin	Skor review berdasarkan proses checkin
review_scores_communication	Skor review berdasarkan komunikasi
review_scores_location	Skor review berdasarkan lokasi
review_scores_value	Skor review berdasarkan kesesuaian properti dengan harga
review_scores_rating	Skor review keseluruhan

Dataset

Categorical Features

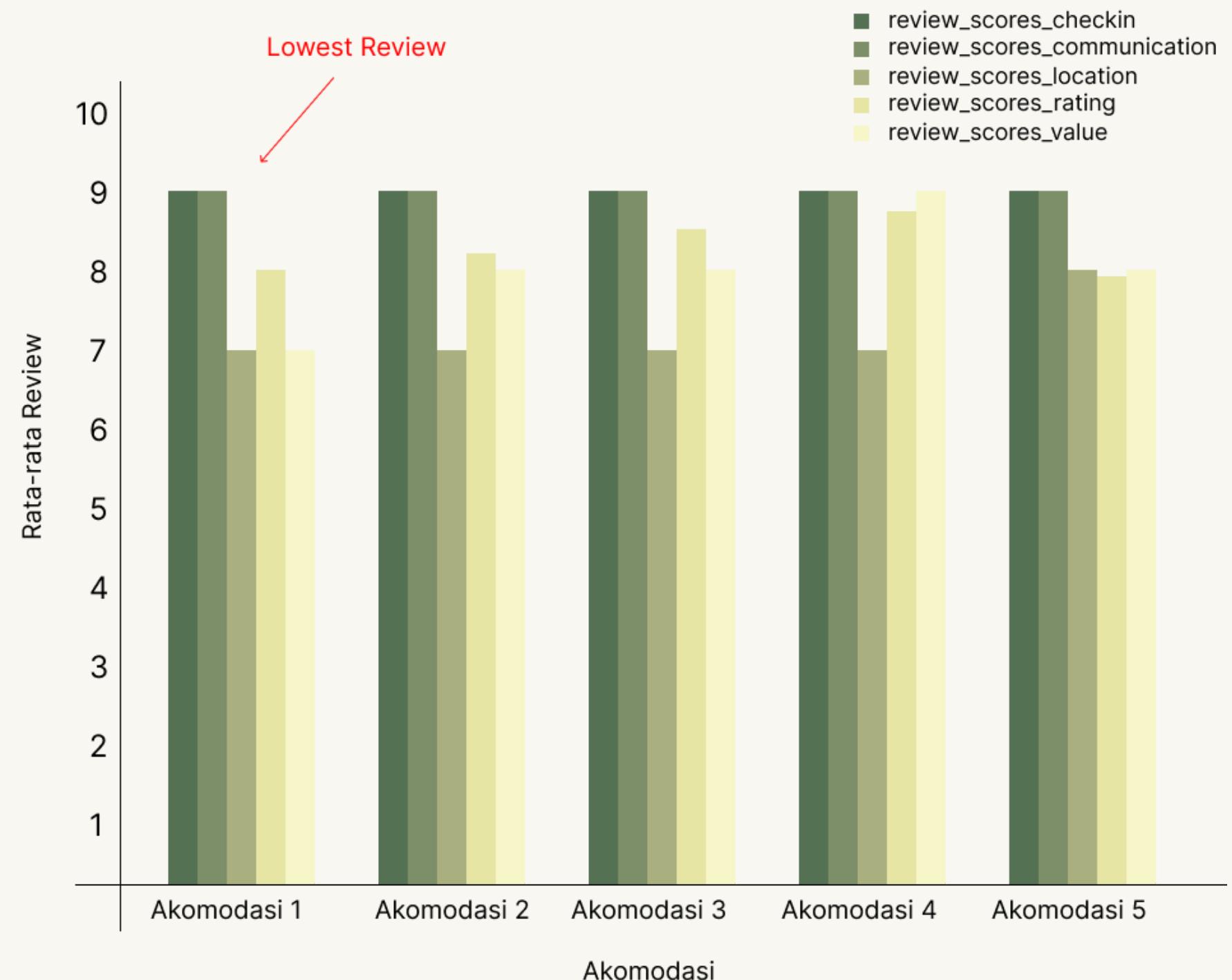
Feature	Description
bed_type	Tipe kasur
cancellation_policy	Ketentuan pembatalan
has_availability	Ketersediaan rumah
host_is_superhost	host yang memiliki gelar superhost
instant_bookable	Menyediakan booking instan
property_type	Tipe properti
room_type	Tipe ruangan
price	Harga

Insight (1a)

Akomodasi atau layanan yang diberikan oleh sebuah rumah dengan rata-rata semua review paling rendah adalah Akomodasi 1, yaitu:

- Dapat menerima penyewa sebanyak **3** orang
- Memiliki **sebuah** kamar mandi, kamar tidur, dan kasur bertipe **real bed**
- Dapat menerima tamu lain sebanyak **2** orang
- Penyewa dapat menginap hingga **1125** hari
- Kebijakan pembatalan adalah **moderate**
- **Rumah** adalah propertinya dengan **entire home/apt** sebagai tipe ruangannya
- Pemilik rumah bukanlah **Superhost**
- Tidak melayani **instant booking**

Karakteristik Akomodasi dengan Rata-Rata Review Terendah

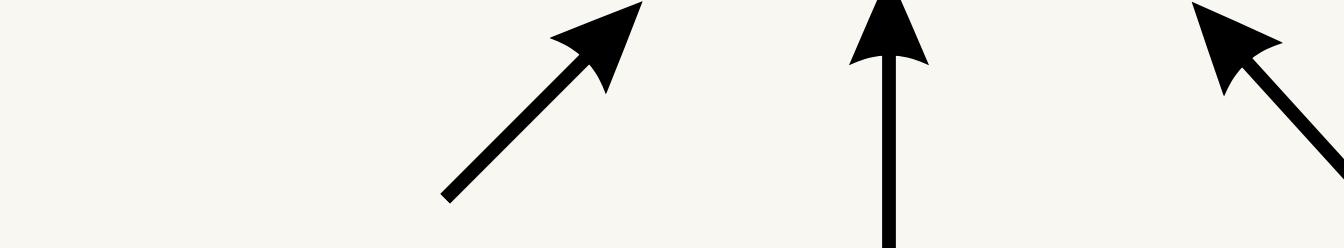


Insight (1b)

Fasilitas yang diberikan oleh sebuah rumah dengan harga termahal adalah

- Dapat menerima penyewa sebanyak **8** orang
- Memiliki **2** kamar mandi, **3** kamar tidur, dan **3** kasur bertipe **real bed**
- Dapat menerima tamu lain sebanyak **6** orang
- Penyewa dapat menginap hingga **14** hari
- Kebijakan pembatalan adalah **strict**
- **Rumah** adalah propertinya dengan **entire home/apt** sebagai tipe ruangannya
- Pemilik bukanlah **Superhost**
- Tidak melayani **instant booking**

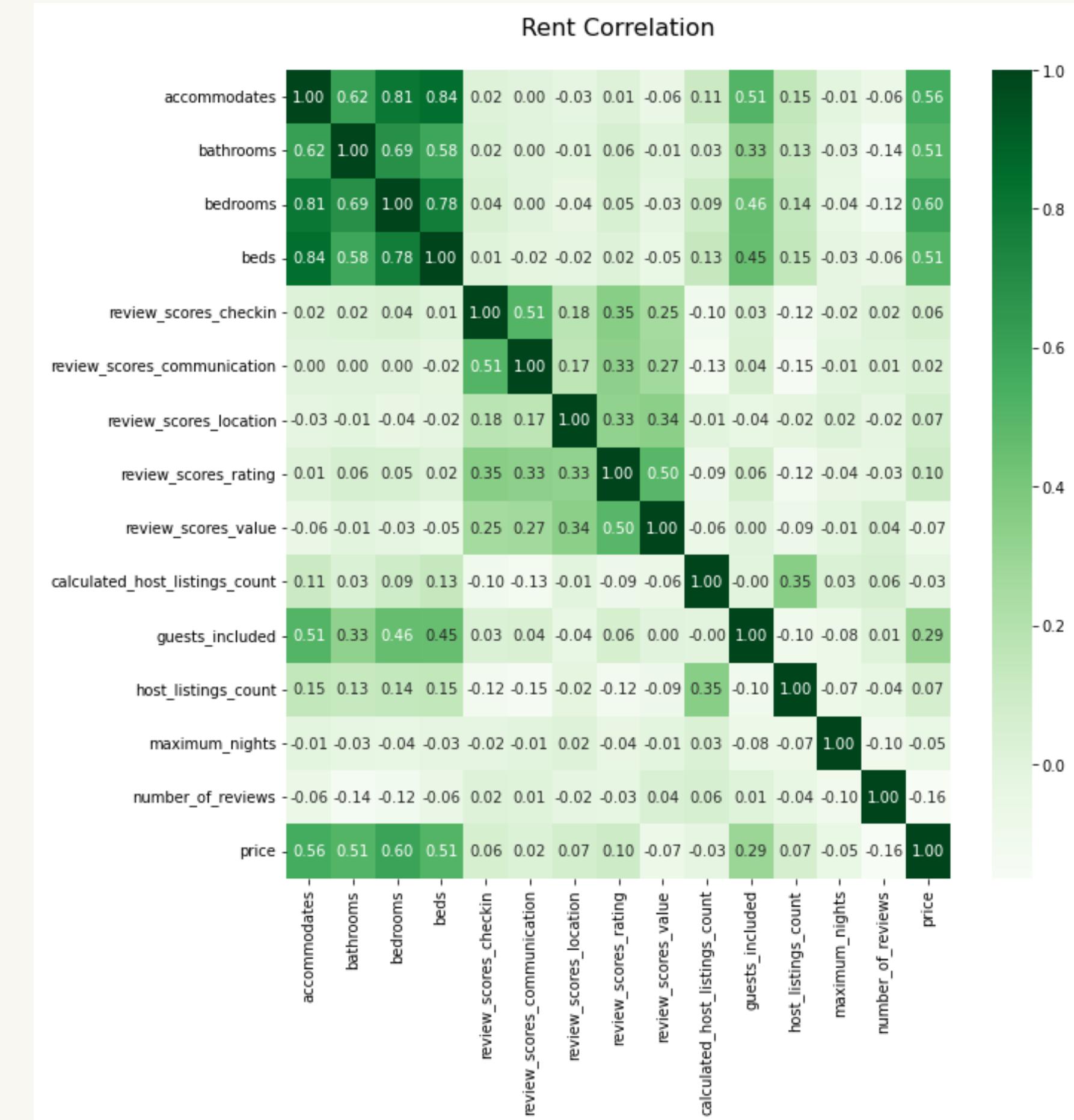
\$2500



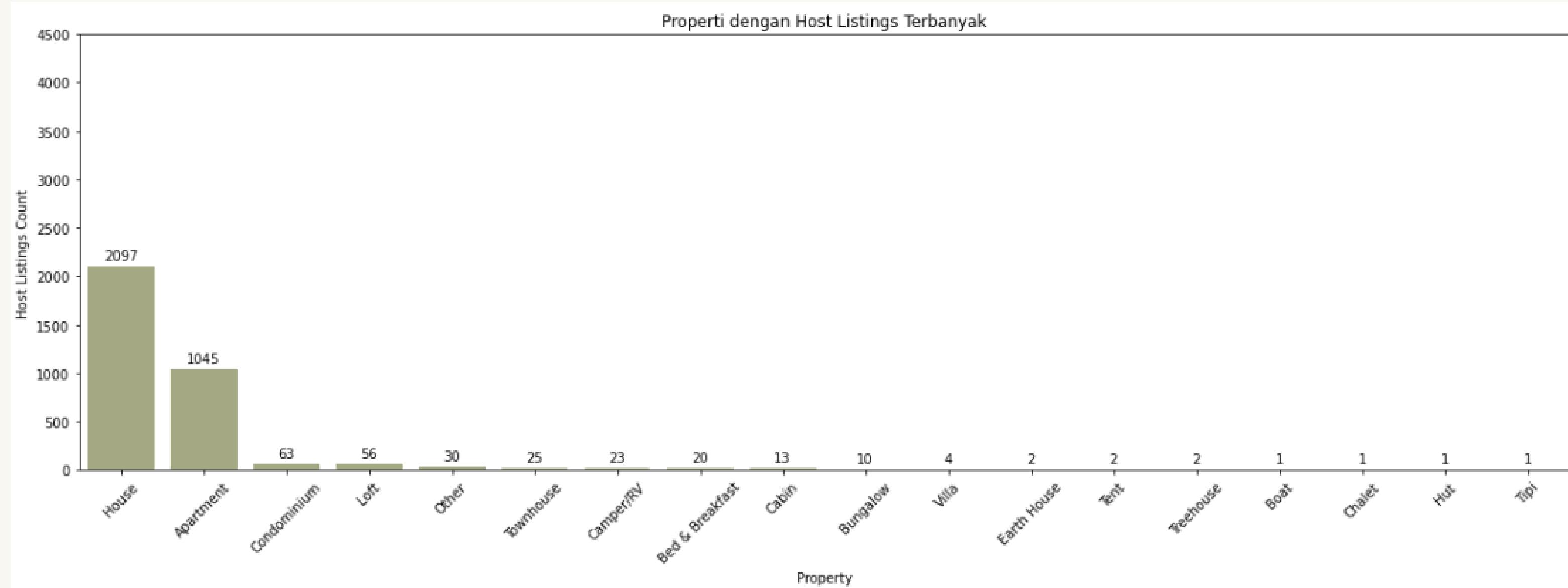
Insight (1c)

Terdapat beberapa fitur yang memiliki cukup korelasi dengan target (price), yaitu *accomodates* dengan nilai **0.56**, *bathrooms* dengan nilai **0.51**, *bedrooms* dengan nilai **0.60**, dan *beds* dengan nilai **0.51**.

Fitur *review* tidak memiliki korelasi dengan target (price) karena nilai korelasinya sangat kecil.



Insight (1d)

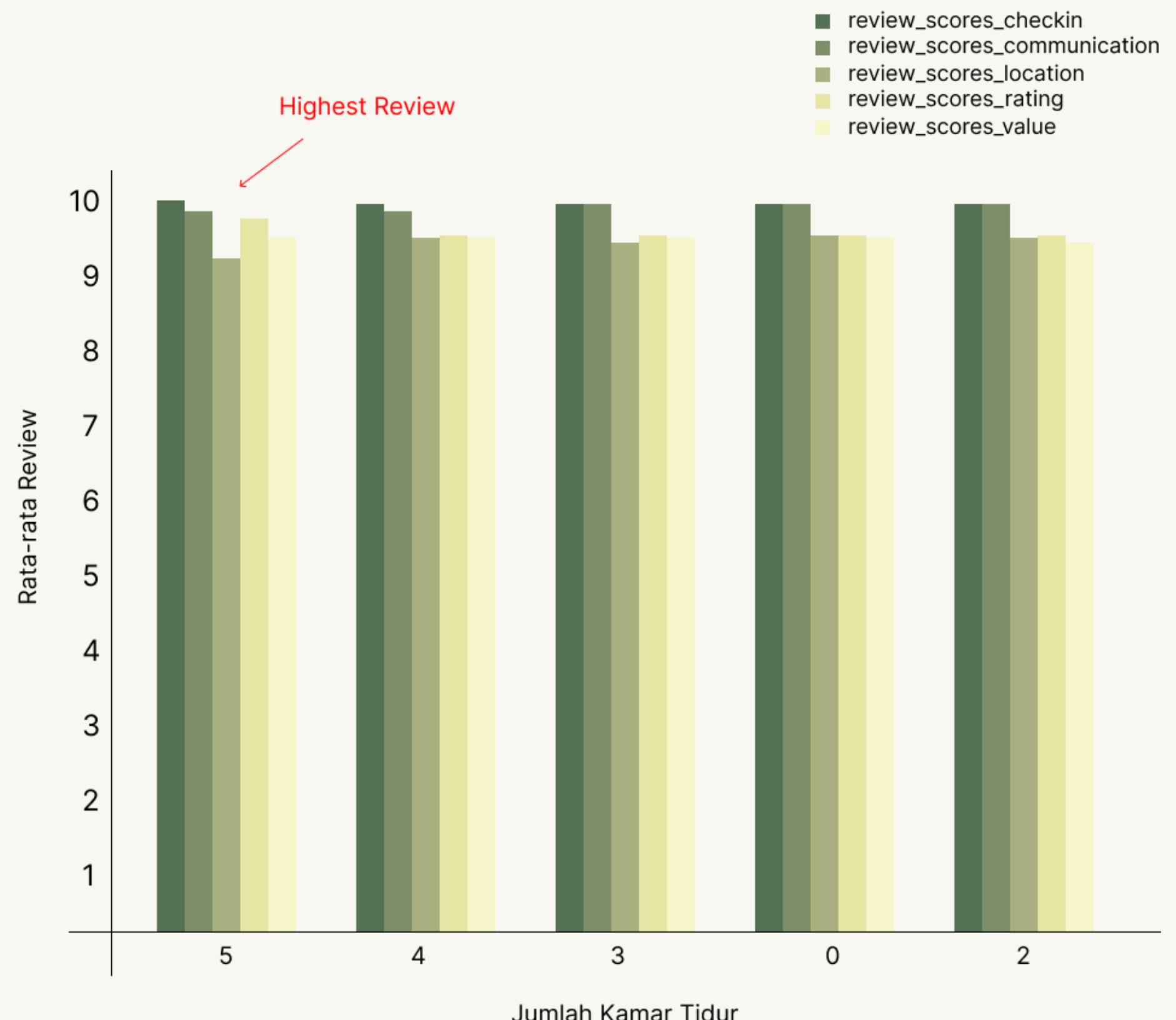


3 tipe properti dengan *host listings* terbanyak adalah (1) [House](#) dengan jumlah [2097](#), (2) [Apartment](#) dengan jumlah [1045](#), dan [Condominium](#) dengan jumlah [63](#)

Insight tambahan

Banyaknya kamar tidur yang memiliki review tertinggi pada suatu rumah adalah **5** kamar

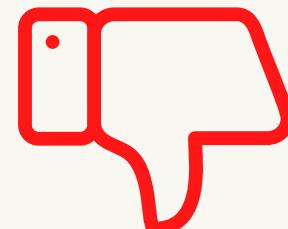
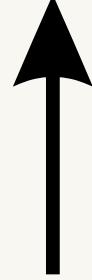
Jumlah Kamar Tidur dengan Rata-Rata Review Tertinggi



Insight tambahan

Harga rumah yang memiliki skor review berdasarkan kesesuaian properti terhadap harga terendah adalah \$890 dengan skor 7.0

\$890



Data Preprocessing

Data Cleaning

- Menghapus data duplikat
- Mengisi baris ber-*missing value* dengan median dan modus untuk fitur tertentu
- Menghapus baris ber-*missing value* untuk fitur 'review_xxx'
- Mentransformasi skewed data dengan *log distribution*
- Mengatasi *outlier* dengan *z-score* untuk data normal
- Menghapus fitur 'has_availability' karena hanya memiliki satu value

Data Transformation (Classification)

- Melakukan *label encoding* pada target
- Melakukan *one hot encoding* pada fitur lainnya
- Melakukan *feature selection* menggunakan Chi2
- Melakukan split data train dan test dengan rasio 80:20

Data Transformation (Regression)

- Melakukan *one hot encoding* pada fitur selain target
- Melakukan feature selection menggunakan Heatmap Correlation
- Melakukan split data train dan test dengan rasio 80:20



Data Transformation (Clustering)

- Menambah fitur baru berdasarkan hasil kalkulasi dari beberapa fitur
- Melakukan standarisasi menggunakan StandardScaler

A tall, modern residential building with a light-colored facade and many windows. The building has a grid-like pattern of vertical and horizontal lines, creating a sense of depth and perspective. The windows are evenly spaced and reflect the surrounding environment.

Modelling

Classification

Model	Accuracy (train)	Accuracy (test)
KNN Classifier	0.439	0.494
KNN Classifier (Tuning Params)	0.567	0.637
Decision Tree Classifier	0.497	0.558
Random Forest Classifier	0.486	0.574
Random Forest Classifier (Tuning Params)	0.511	0.532

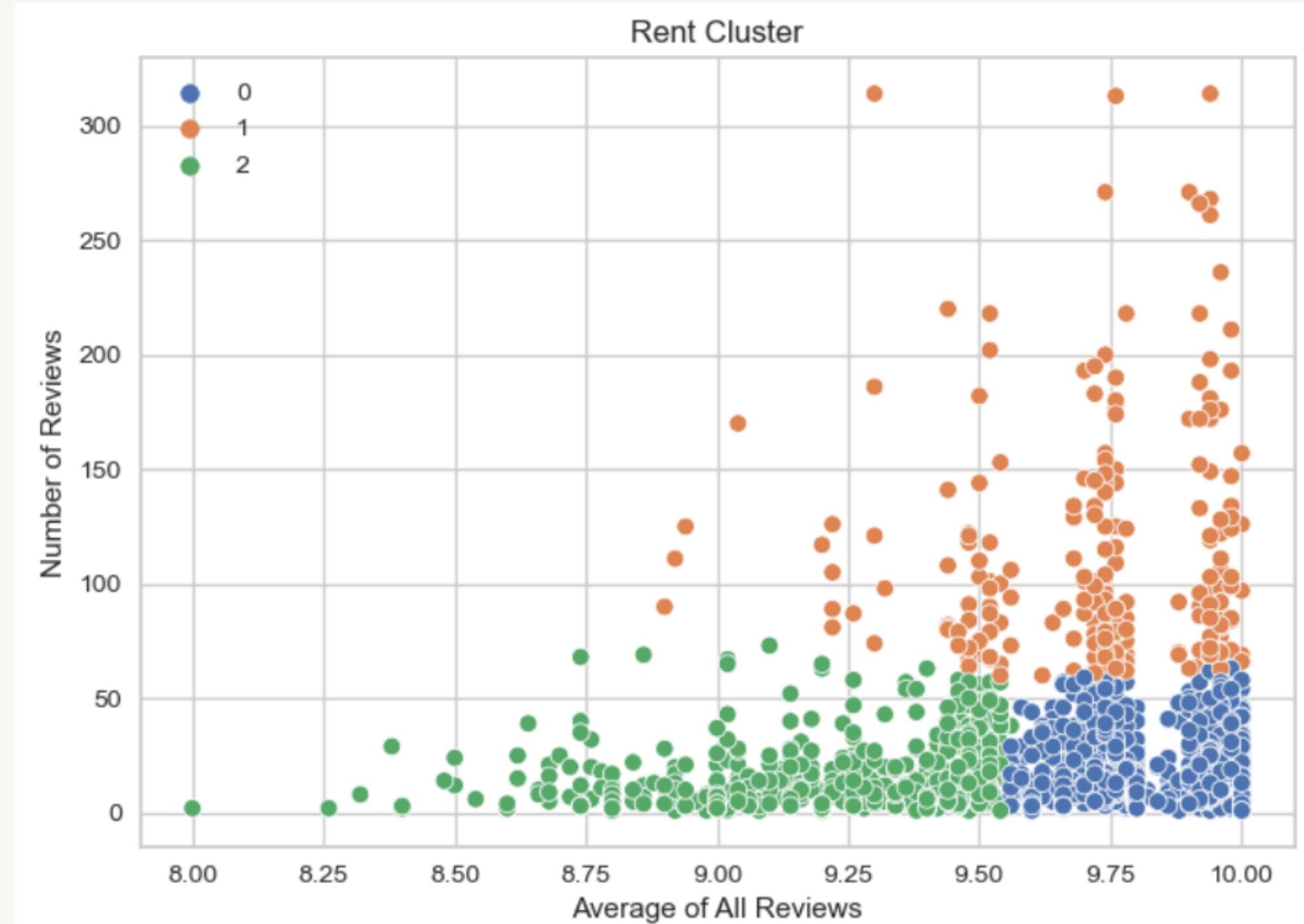
Regression

Model	R2 (train)	R2 (test)
Linear Regression	0.463	0.449
Ridge Regression	0.463	0.449
Decision Tree Regressor	0.492	0.549
Random Forest Regressor	0.505	0.547
Random Forest Regressor (Tuning Params)	0.509	0.544
AdaBoost Regressor	0.502	0.495
GradientBoosting Regressor	0.519	0.536
XGBoost Regressor	0.520	0.536

Clustering

Berdasarkan hasil cluster di samping, dapat diambil beberapa informasi yaitu:

- Klaster biru merupakan klaster dengan rata-rata review sangat baik namun masih sedikit orang yang memberikan review. Melalui hal tersebut dapat dilakukan pengiklanan untuk mempromosikan beberapa rumah sehingga banyak orang yang tahu dan jumlah review menjadi meningkat.
- Klaster hijau merupakan klaster dengan rata-rata rating cukup baik namun tidak banyak orang yang tahu. Melalui hal tersebut dapat dilakukan pengiklanan untuk mempromosikan rumah yang akan disewa dan perlu meningkatkan pelayanan.
- Klaster orange merupakan klaster akomodasi dengan rata-rata rating sangat baik dan banyak orang yang tahu sehingga hanya perlu melakukan peningkatan pelayanan agar semakin banyak orang yang ingin menyewa.



**TERIMA
KASIH**

Regu Buaya

