



Analisis Data dengan Statistika dan R

HOUSE PRICES PREDICTION

Presented by Kelompok 01

ANGGOTA



6182001001

Jenson Mark Lowell



6182101040

Samuel Edward Winoto



6182101054

Steffi Widjaya



Overview

1. Introduction
2. Attribute
3. Penyesuaian Data
4. Eksplorasi Data
5. Analisis Data
6. Model Prediksi
7. Hasil Analisis
8. Kesimpulan
9. Insight
10. Actionable Insight



INTRODUCTION

Mintalah pembeli rumah untuk mendeskripsikan rumah impian mereka.

Kemungkinan besar jawaban mereka bukan, ketinggian langit-langit, ruang bawah tanah, atau seberapa dekat dengan kereta api (transportasi umum).

Namun, mereka cenderung mempertimbangkan harga daripada faktor lainnya.

Maka dari itu melalui dataset Hotel Prices, ingin diprediksi harga setiap rumah berdasarkan fitur-fitur yang dipilih & membuat model prediksi harga rumah.



ATTRIBUTE

- **SalePrice:** Harga jual rumah.
- **OverallQual:** Kualitas keseluruhan material dan penyelesaian rumah.
- **GrLivArea:** Luas area tinggal di atas tanah.
- **GarageCars:** Kapasitas mobil di garasi.
- **TotalBsmtSF:** Total luas basement.
- **X1stFlrSF:** Luas lantai pertama.
- **FullBath:** Jumlah kamar mandi.
- **YearBuilt:** Tahun dibangun.
- **OverallCond:** Kondisi keseluruhan rumah.
- **Utilities:** Tipe utilitas yang tersedia.
- **Neighborhood:** Lokasi lingkungan.
- **ExterQual:** Kualitas eksterior.
- **ExterCond:** Kondisi eksterior.
- **Street:** Jenis jalan akses ke properti.
- **LotShape:** Bentuk lot properti.

PENYESUAIAN DATA

Mengubah Tipe Data: Mengonversi variabel kategorikal menjadi tipe faktor

```
categorical_vars <- c("MSSubClass", "MSZoning", "Street", "Alley", "LotShape", "LandContour", "Utilities", "LotConfig", "LandSlope",  
"Neighborhood", "Condition1", "Condition2", "BldgType", "HouseStyle", "RoofStyle", "RoofMatl", "Exterior1st",  
"Exterior2nd", "MasVnrType", "ExterQual", "ExterCond", "Foundation", "BsmtQual", "BsmtCond", "BsmtExposure",  
"BsmtFinType1", "BsmtFinType2", "Heating", "HeatingQC", "CentralAir", "Electrical", "KitchenQual", "Functional",  
"FireplaceQu", "GarageType", "GarageFinish", "GarageQual", "GarageCond", "PavedDrive", "PoolQC", "Fence",  
"MiscFeature", "SaleType", "SaleCondition")  
  
house_data[categorical_vars] <- lapply(house_data[categorical_vars], as.factor)
```

Menangani Nilai Hilang:

- Numerik: Mengisi dengan median
- Faktor: Menambahkan level eksplisit untuk nilai hilang

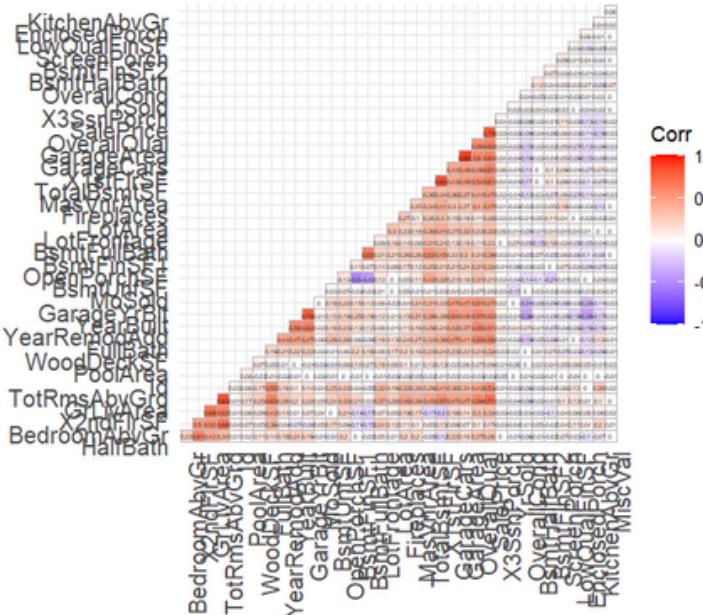
```
house_data <- house_data %>%  
  mutate(across(where(is.numeric), ~ ifelse(is.na(.), median(., na.rm = TRUE), .))) %>%  
  mutate(across(where(is.factor), ~ fct_explicit_na(.)))
```



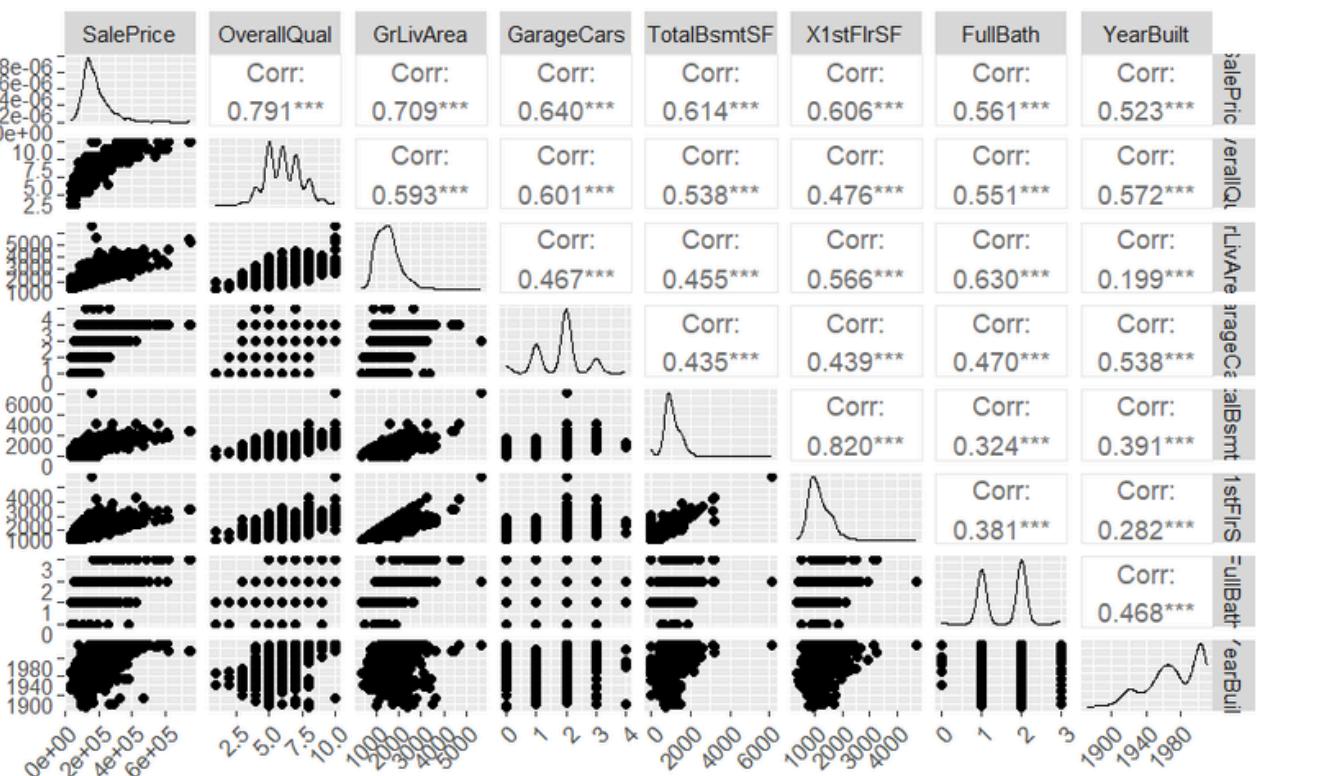
EKSPLORASI DATA

Korelasi Antar Variabel

- Analisis Korelasi: Mengidentifikasi hubungan antara variabel numerik.



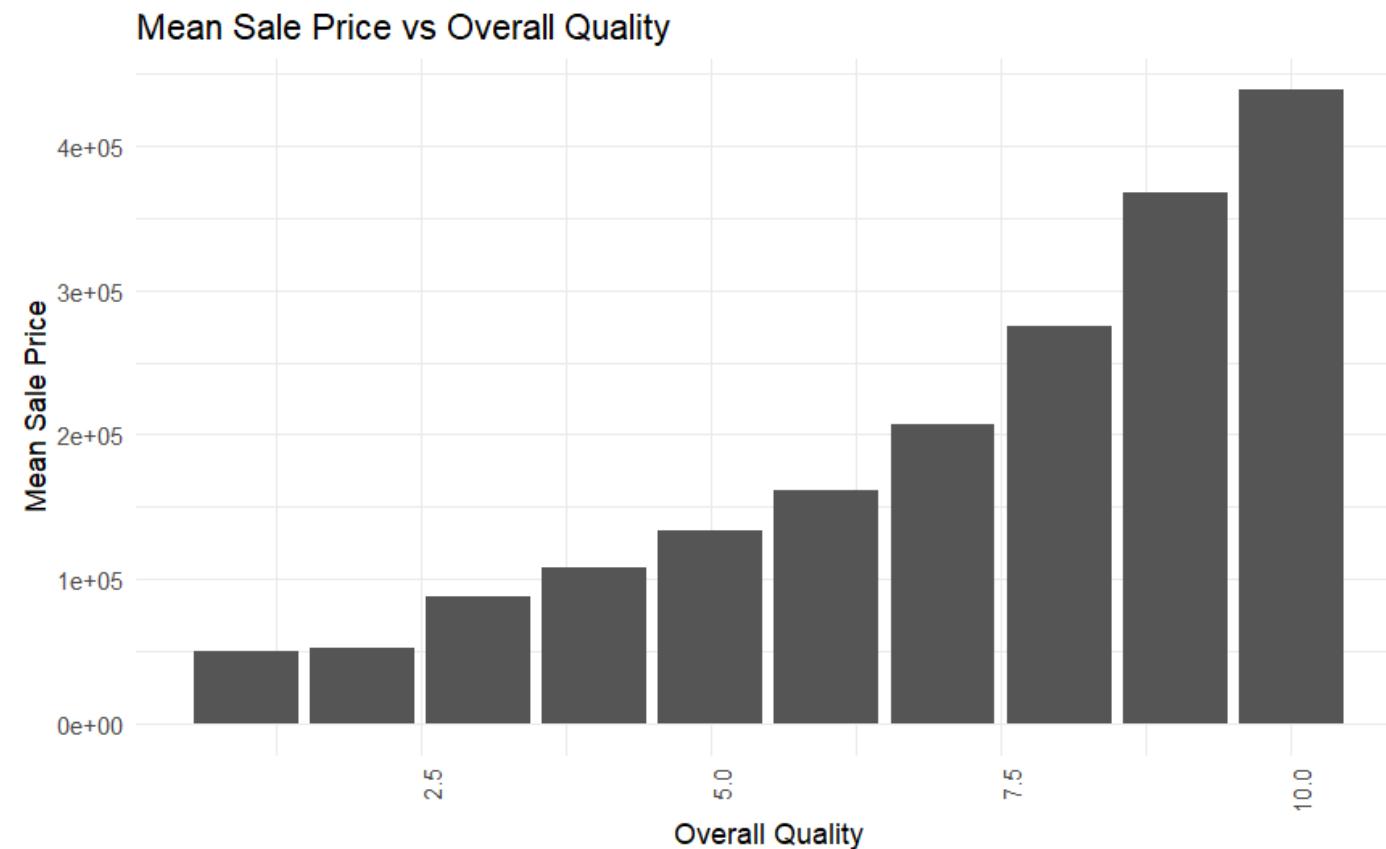
- Pair Plot: Memvisualisasikan hubungan antar variabel terpilih.



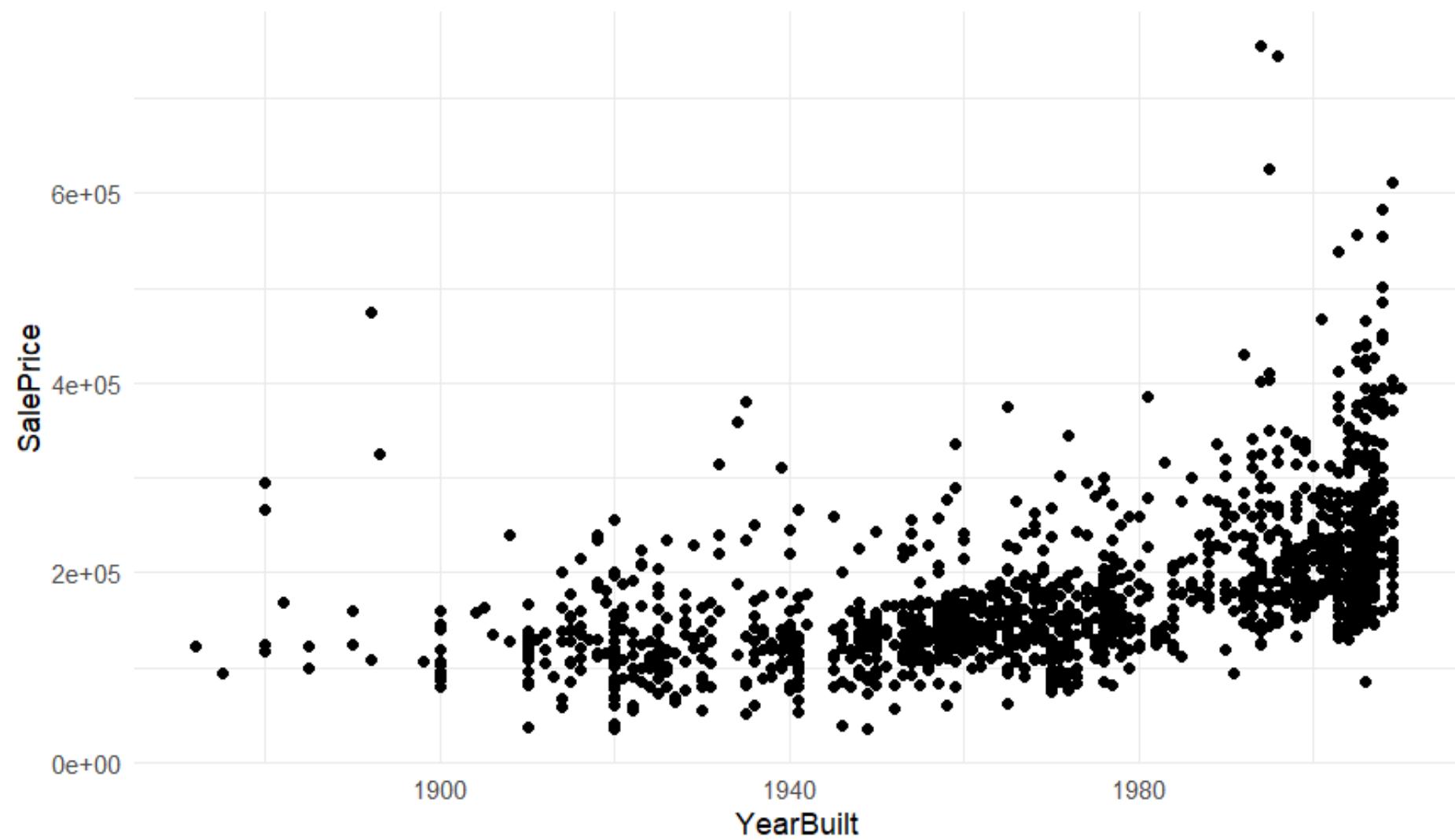
ANALISIS DATA

Kualitas Keseluruhan dan Harga Jual:

Kualitas keseluruhan memiliki pengaruh besar terhadap harga jual.



YearBuilt vs Overall Quality



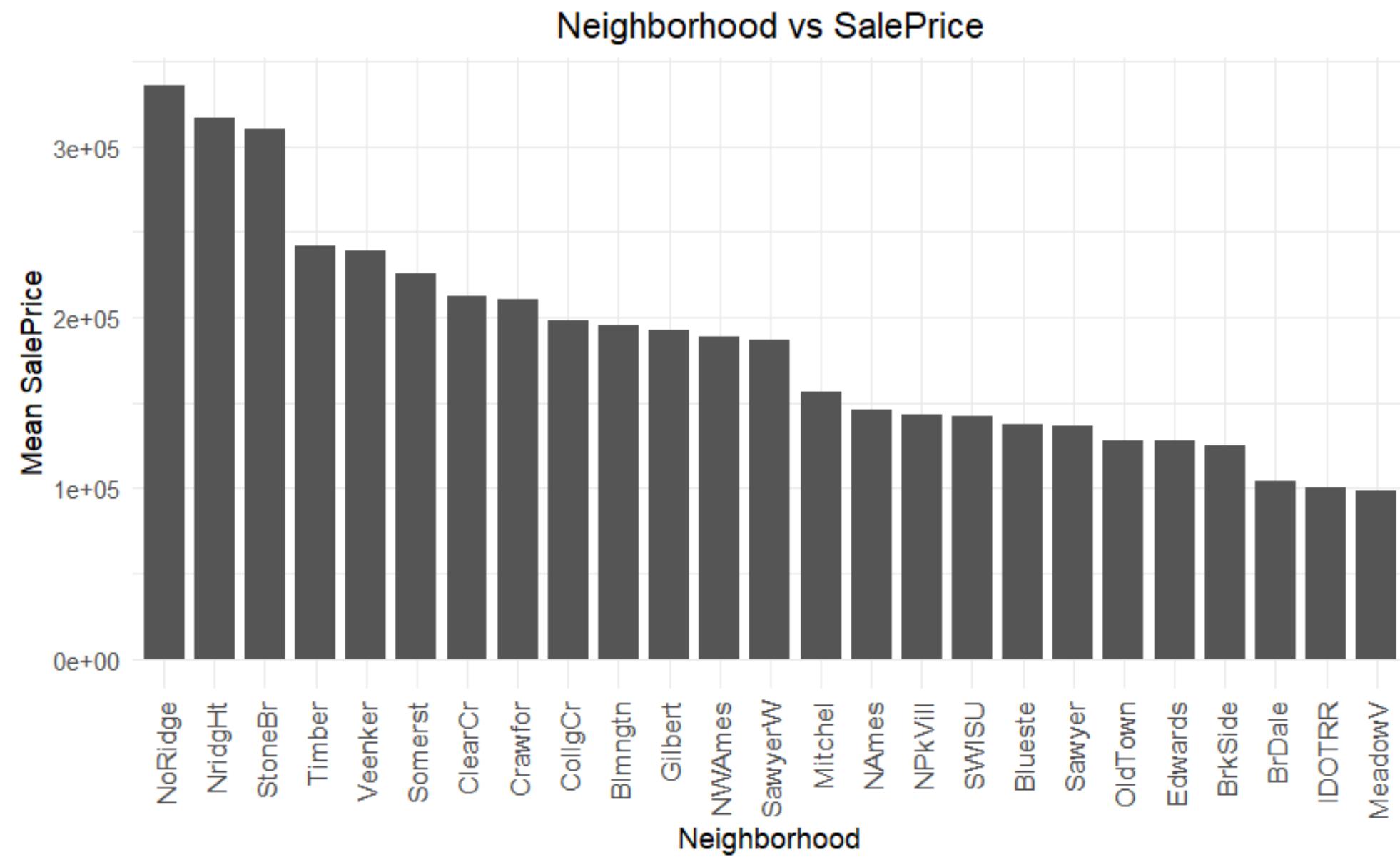
Tahun Dibangun dan Harga Jual:

Rumah yang lebih baru cenderung lebih mahal.

ANALISIS DATA (2)

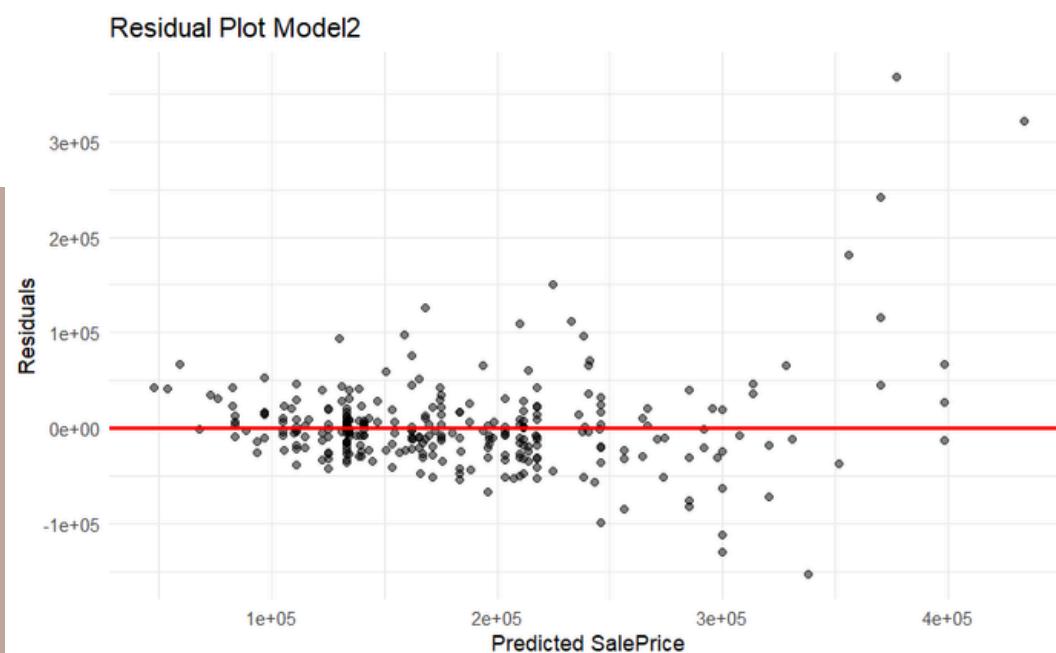
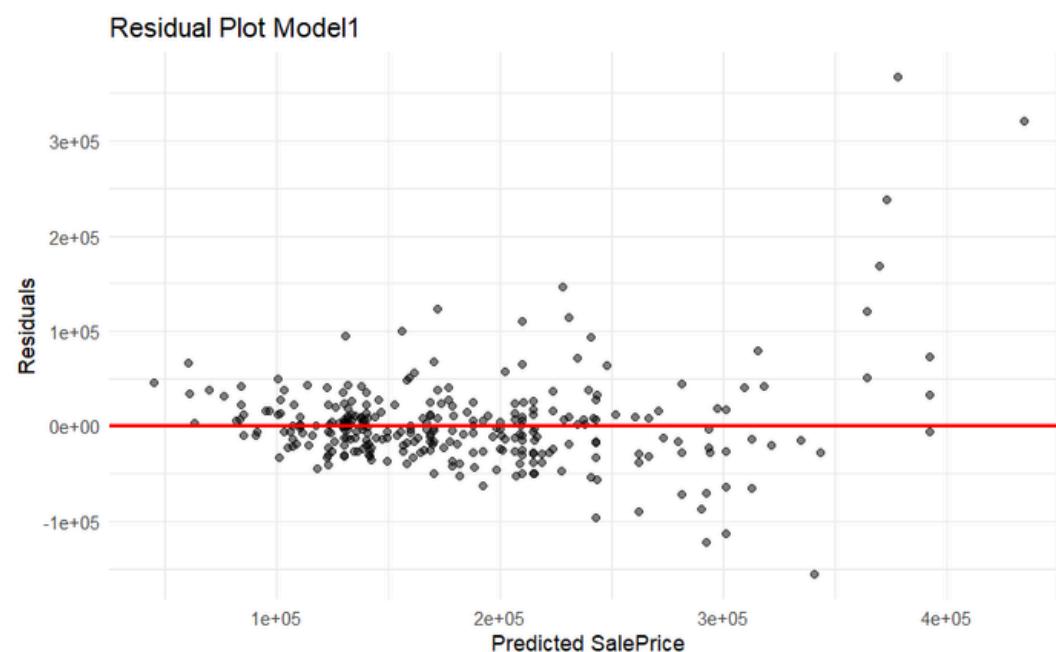
Lingkungan dan Harga Jual:

Lokasi lingkungan menunjukkan variasi harga jual yang signifikan.

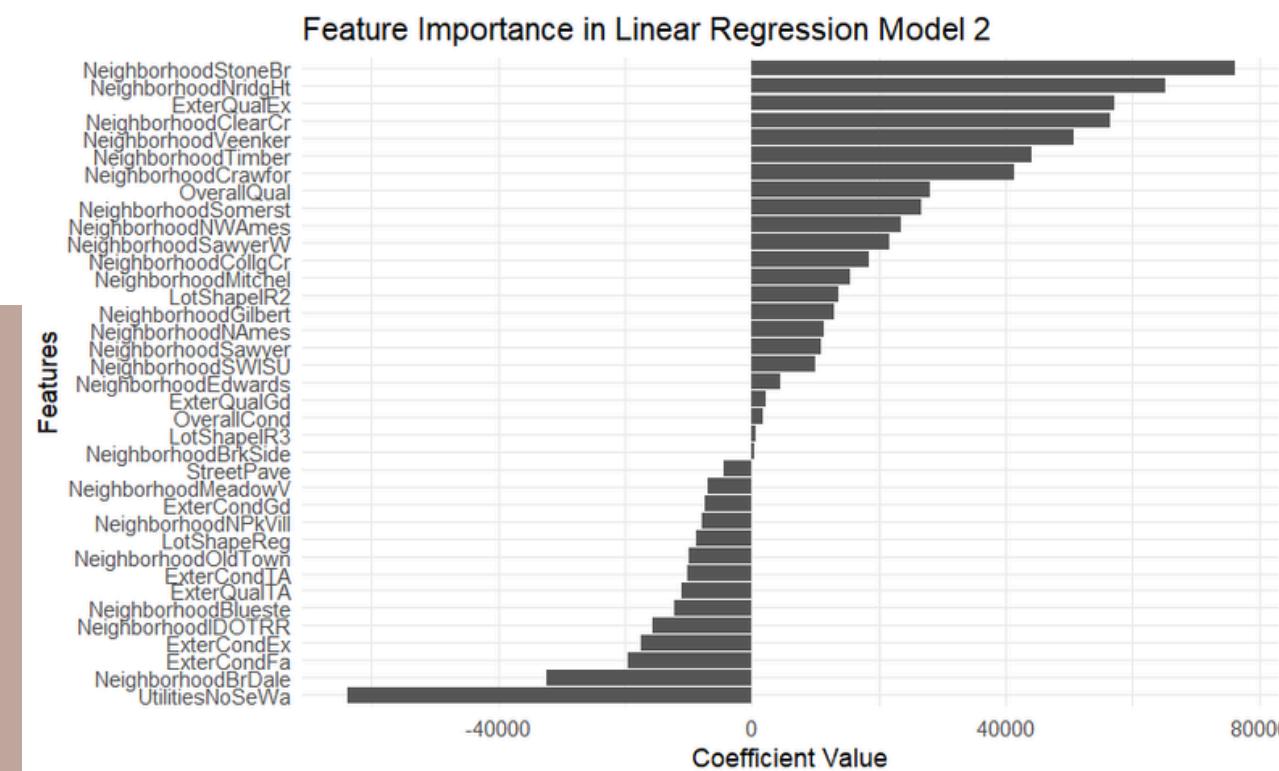
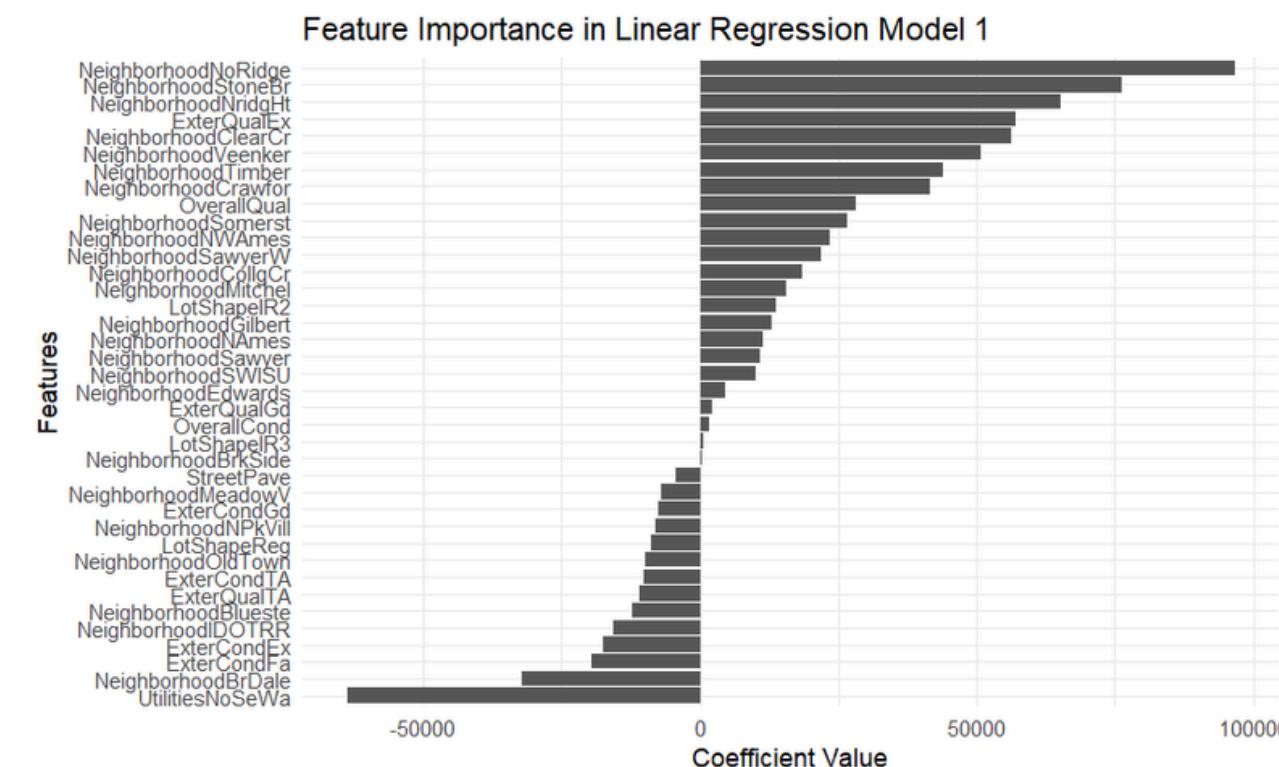


MODEL PREDIKSI

Model 1 dan Model 2 menunjukkan kemampuan prediksi yang baik, dengan distribusi residual yang relatif seimbang di sekitar garis nol.



Feature Importance: Variabel Neighborhood, OverallQual, dan OverallQual, ExterQualEx menunjukkan pengaruh yang signifikan terhadap harga jual rumah dalam kedua model.



MODEL PREDIKSI (2)



Pemilihan Model Terbaik:
membandingkan RMSE (Root Mean
Squared Error) dari kedua model.

Model dengan RMSE lebih rendah
dipilih sebagai model terbaik.

RMSE Model 1 = 49583.86

RMSE Model 2 = 49753.79

HASIL ANALISIS

- Model 1 direkomendasikan sebagai model terbaik untuk memprediksi harga jual rumah dalam kasus ini.
- Mengeliminasi variable tidak secara signifikan meningkatkan performa model.
- Harga jual aktual dan yang diprediksi oleh model menunjukkan tren yang baik hal ini dapat terlihat dari sebagian besar titik data berada dekat dengan garis kesesuaian.
- Distribusi residual relatif seimbang di sekitar garis nol, menunjukkan bahwa tidak ada pola yang jelas dalam residuals yang mengindikasikan masalah dengan model.



KESIMPULAN

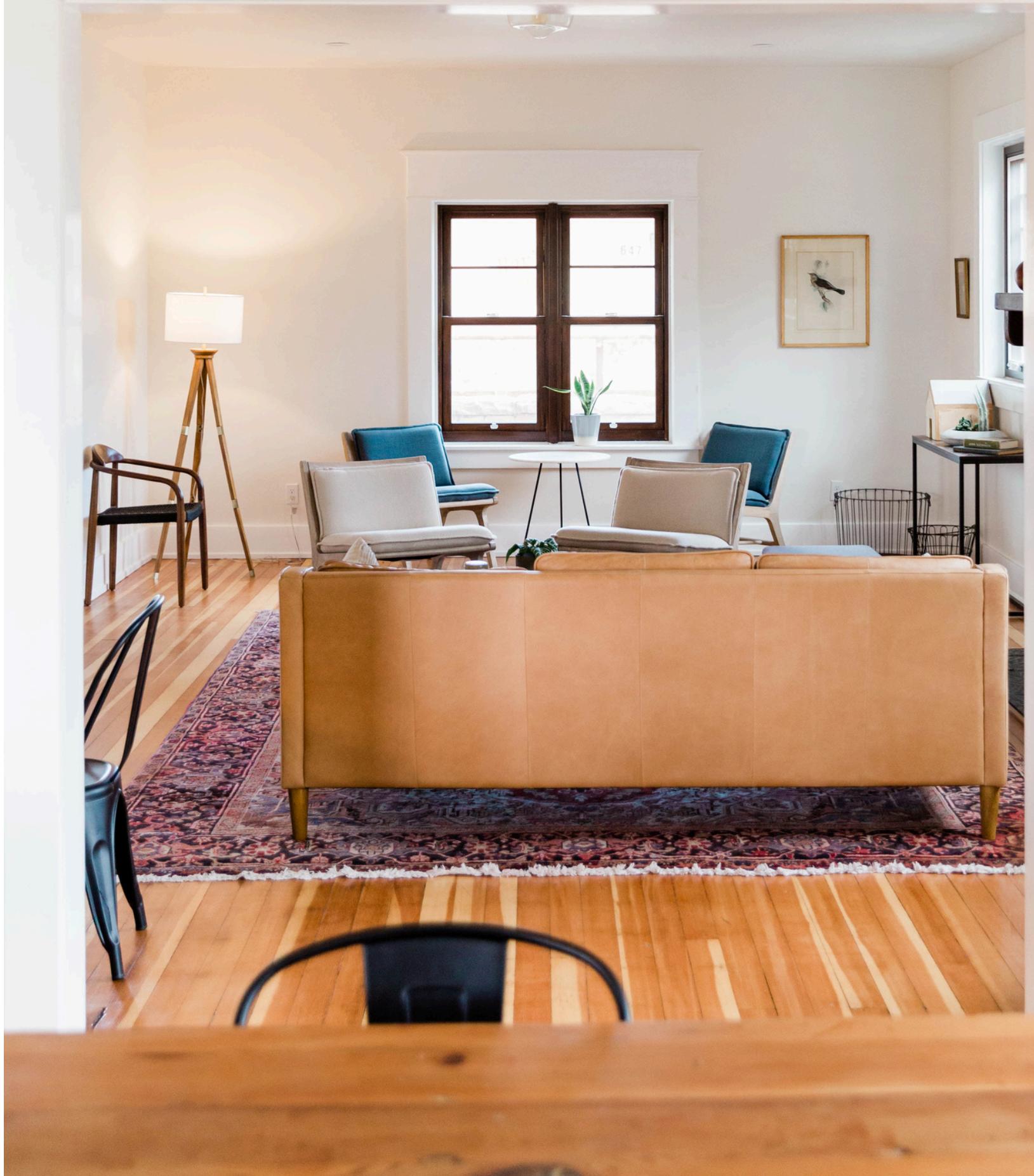
Faktor-faktor utama yang mempengaruhi harga rumah:



- **Kualitas Keseluruhan** (OverallQual): kualitas material dan hasil akhir rumah secara keseluruhan.
- **Kualitas Eksterior** (ExterQual): kualitas bahan dan hasil akhir eksterior rumah.
- **Lingkungan** (Neighborhood): Lokasi rumah sangat berpengaruh terhadap nilai jual. Lingkungan yang lebih baik cenderung memiliki harga jual yang lebih tinggi.
- **Tahun Dibangun** (YearBuilt): Rumah yang dibangun lebih baru cenderung memiliki harga jual yang lebih tinggi karena kondisi yang lebih baik dan fitur yang lebih modern.

INSIGHT

- **Renovasi untuk Meningkatkan Kualitas:**
Fokus pada peningkatan kualitas keseluruhan dan eksterior.
- **Investasi di Lokasi Strategis:**
Memilih lokasi dengan permintaan tinggi.
- **Rumah baru/Renovasi:**
Pembangunan atau renovasi rumah untuk tampilan dan fungsionalitas yang lebih modern.



ACTIONABLE INSIGHT

- **Peningkatan kualitas:**

Renovasi untuk meningkatkan kualitas rumah.

- **Pemilihan lokasi:**

Investasi di lingkungan yang strategis dan populer.

- **Rumah baru atau renovasi:**

Menarik pembeli dengan rumah baru atau yang telah direnovasi.



THANK YOU

Do You Have Any Question?

