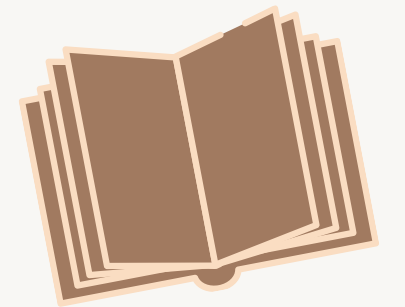


# VISUALISASI

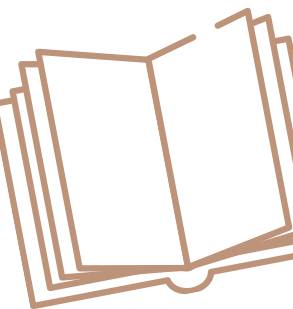


# IMPORT LIBRARY



```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

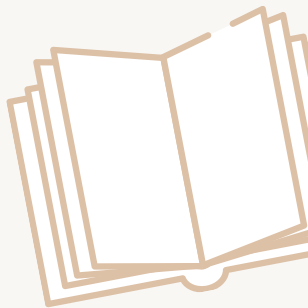
- Pandas: Untuk manipulasi dan analisis data berbentuk tabel.
- NumPy: Untuk komputasi numerik dan operasi pada array.
- Matplotlib: Untuk membuat grafik dan visualisasi data 2D.
- Seaborn: Untuk membuat visualisasi statistik yang lebih informatif dan estetik, seperti heatmap, box plot, dan pair plot.



# MEMANGGIL DATA

memanggil data menggunakan github

```
url="https://raw.githubusercontent.com/meilinnedafarera/Dataset/refs/heads/main/House%20Price%20Dataset.csv"  
data=pd.read_csv(url)  
data.head()
```



# VISUALISASI

Jenis Grafik	Digunakan untuk	Variabel	Sumbu X	Sumbu Y
Histogram	Distribusi frekuensi data numerik	1 (Numerik)	Nilai atau bin	Frekuensi
Scatter Plot	Korelasi/Hubungan dua variabel numerik	2 (Numerik)	Variabel 1	Variabel 2
Boxplot	Distribusi data, outlier, ringkasan statistik	1 (Numerik)	Kategori (opsional)	Nilai
Bar Chart	Membandingkan nilai antar kategori	1 (Kategorik + Numerik)	Kategori	Nilai
Pie Chart	Menunjukkan proporsi antar kategori	1 (Kategorik)	Tidak ada	Proporsi/Percentage

# VISUALISASI

## HISTOGRAM

- **Contoh Penggunaan:** untuk melihat distribusi data seperti distribusi nilai ujian, tinggi badan, atau pendapatan.
- **Contoh Visual:** Memperlihatkan berapa banyak nilai yang berada di dalam setiap kelompok (bin).

## SCATTER PLOT

- **Contoh Penggunaan:** untuk memvisualisasikan korelasi antara tinggi dan berat badan, atau antara suhu dan tekanan udara.
- **Contoh Visual:** Titik-titik tersebar di grafik, di mana setiap titik mewakili pasangan nilai dari dua variabel.

## BOXPLOT

- **Contoh Penggunaan:** untuk membandingkan distribusi data dalam beberapa grup, misalnya distribusi pendapatan di beberapa kota.
- **Contoh Visual:** Terdiri dari kotak yang menunjukkan kuartil bawah, median, dan kuartil atas, serta garis (whisker) yang menunjukkan rentang data dan outlier.

# VISUALISASI

## BAR-CHART

- **Contoh Penggunaan:** untuk membandingkan penjualan produk di berbagai daerah atau jumlah populasi di berbagai kota.
- **Contoh Visual:** Batang vertikal atau horizontal yang tinggi atau panjangnya sesuai dengan nilai atau frekuensi untuk setiap kategori.

## PIE-CHART

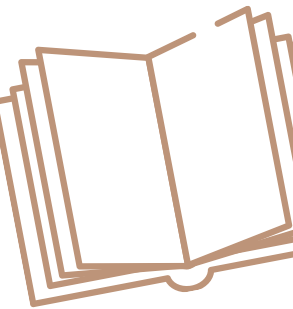
- **Contoh Penggunaan:** untuk menggambarkan pembagian pasar, alokasi anggaran, atau proporsi responden dalam survei.
- **Contoh Visual:** Lingkaran yang dibagi menjadi beberapa bagian, dengan setiap bagian mewakili kontribusi relatif dari suatu kategori.

# HISTOGRAM

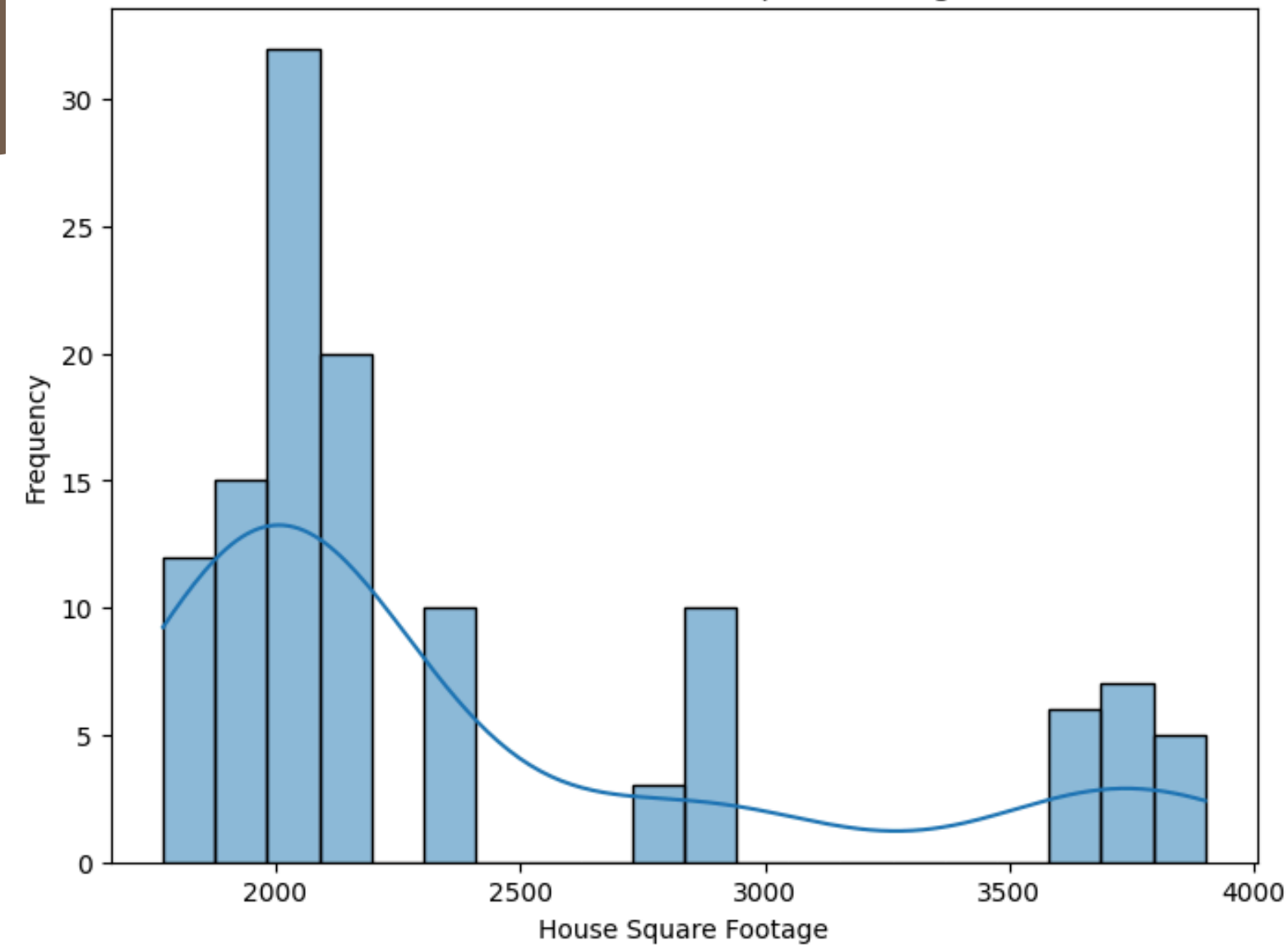


```
plt.figure(figsize=(8, 6))
sns.histplot(data['house_sqft'], bins=20, kde=True)
plt.title('Distribution of House Square Footage')
plt.xlabel('House Square Footage')
plt.ylabel('Frequency')
plt.show()
```

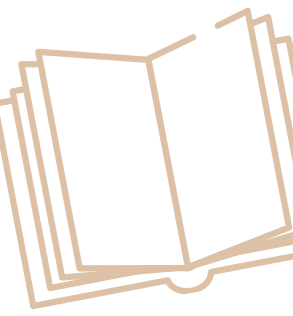
- `plt.figure(figsize=(8, 6))`: Membuat kanvas plot dengan ukuran 8x6 inci.
- `sns.histplot(data['house_sqft'], bins=20, kde=True)`: Membuat histogram dari data luas rumah (`house_sqft`) dengan 20 bin dan kurva distribusi (KDE).
- `plt.title()`: Menambahkan judul grafik.
- `plt.xlabel()`: Memberi label sumbu X (Luas Rumah).
- `plt.ylabel()`: Memberi label sumbu Y (Frekuensi)
- `plt.show()`: Menampilkan plot.



Distribution of House Square Footage



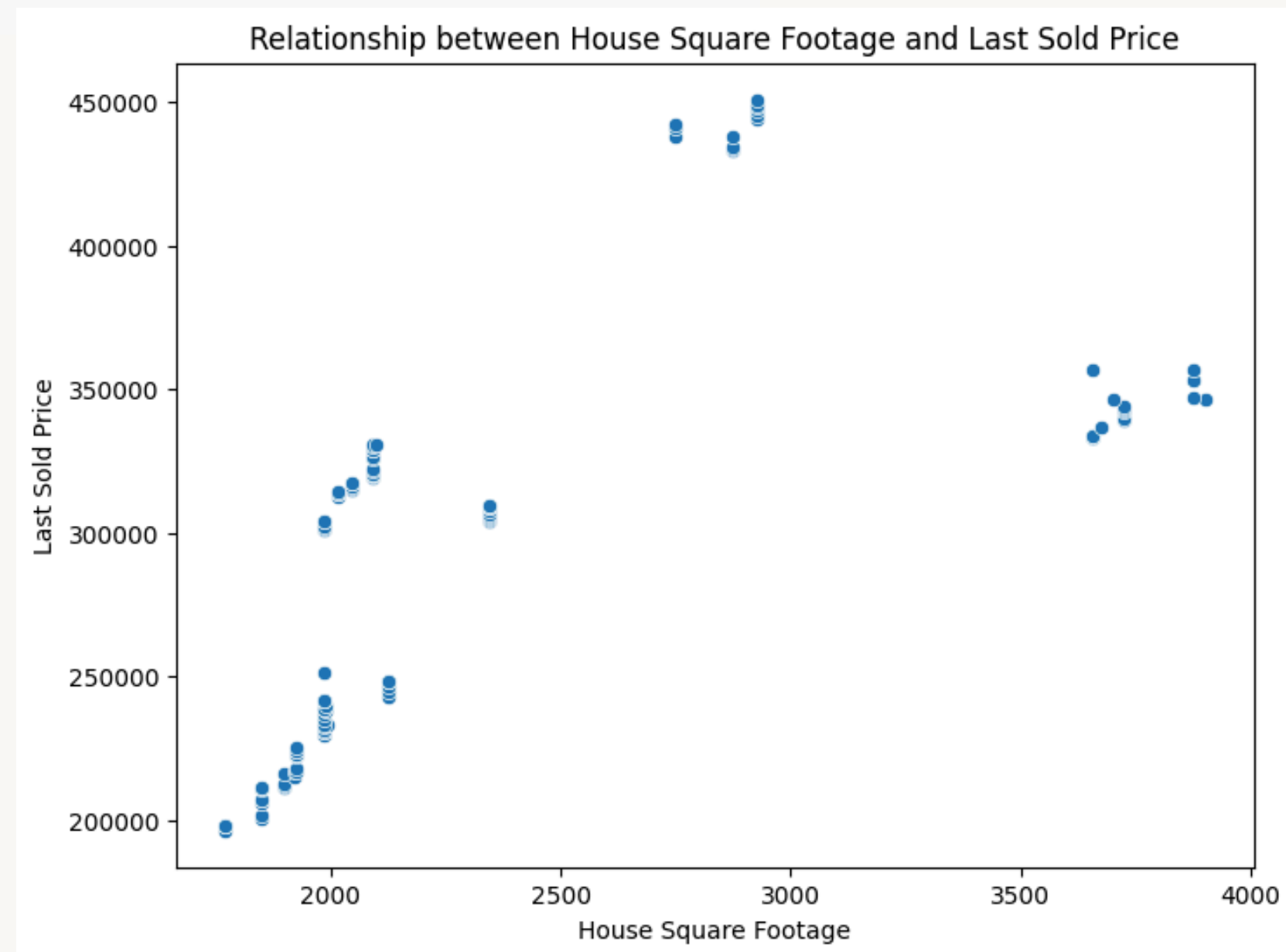
- Mayoritas rumah memiliki luas sekitar 2000 kaki persegi. Ini terlihat dari puncak tertinggi di histogram.
- Terdapat dua kelompok utama ukuran rumah: satu di sekitar 2000 kaki persegi dan yang lain di sekitar 3500 kaki persegi, meskipun kelompok kedua jauh lebih kecil.
- Distribusi tidak merata, dengan sedikit rumah di rentang luas di atas 3000 kaki persegi.
- Garis KDE menunjukkan pola distribusi dengan dua puncak, menandakan dua grup ukuran yang menonjol dalam data.
- Sebagian besar rumah berukuran lebih kecil (sekitar 2000 kaki persegi), sementara rumah yang lebih besar relatif jarang.
- Distribusi ini menunjukkan tren bahwa sebagian besar rumah dalam dataset memiliki ukuran yang cenderung standar di sekitar 2000 kaki persegi, dengan beberapa outlier yang lebih besar.



# SCATTER PLOT

## Menunjukkan Visualisasi Luas Rumah

```
plt.figure(figsize=(8, 6))
sns.scatterplot(x='house_sqft', y='last_sold_price', data=data)
plt.title('Relationship between House Square Footage and Last Sold Price')
plt.xlabel('House Square Footage')
plt.ylabel('Last Sold Price')
plt.show()
```



# SCATTER PLOT

Gambar tersebut adalah scatter plot yang menunjukkan hubungan antara luas rumah (house\_sqft) dan harga jual terakhir (last\_sold\_price). Berikut adalah penjelasan lebih rinci:

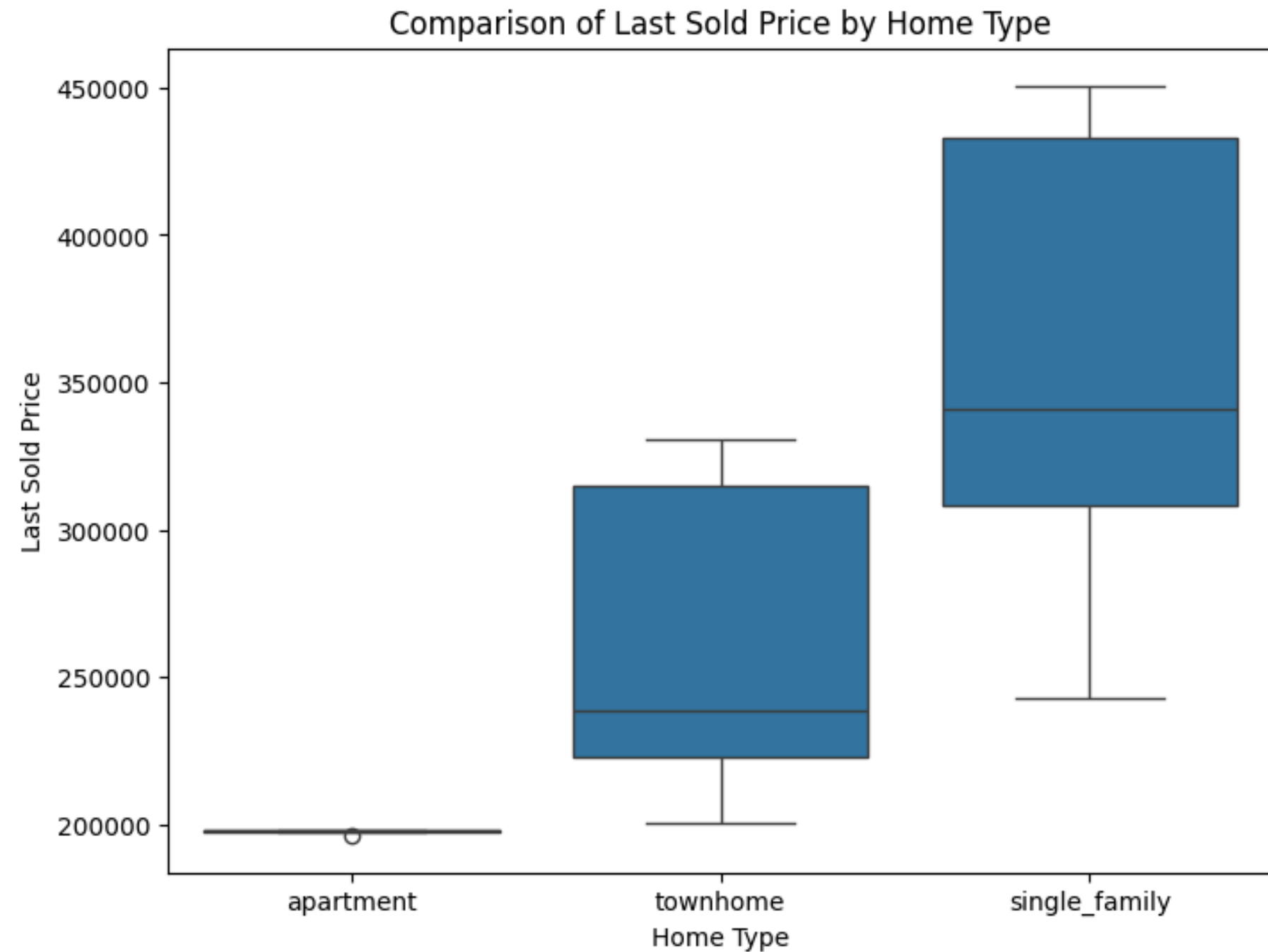
1. Sumbu X (House Square Footage): Menunjukkan ukuran luas rumah dalam kaki persegi, berkisar antara 2000 hingga 4000 kaki persegi.
2. Sumbu Y (Last Sold Price): Menunjukkan harga jual terakhir rumah, berkisar antara 200.000 hingga 450.000.
3. Titik Data: Setiap titik mewakili satu rumah. Sebagian besar titik berada di bagian bawah grafik, menunjukkan bahwa rumah dengan luas lebih kecil (sekitar 2000 kaki persegi) cenderung memiliki harga jual lebih rendah (sekitar 200.000 - 300.000).
4. Pola Umum: Ada kecenderungan positif; semakin besar luas rumah, harga jual terakhir juga cenderung meningkat. Namun, terlihat ada beberapa titik yang terpisah (outliers) di sisi kanan grafik (rumah dengan luas di atas 3000 kaki persegi yang dijual dengan harga di atas 350.000), menunjukkan bahwa rumah yang lebih besar memiliki variasi harga yang lebih tinggi.
5. Keteraturan dan Penyebaran: Titik-titik data menyebar secara relatif merata di sepanjang rentang ukuran, meskipun ada beberapa kumpulan di area tertentu (misalnya, banyak titik berada di kisaran 2000-2500 kaki persegi dengan harga sekitar 250.000-300.000).

Secara keseluruhan, scatter plot ini menunjukkan adanya hubungan positif antara luas rumah dan harga jual terakhir, meskipun dengan variasi yang cukup besar di harga untuk rumah yang lebih besar.

# BOXPLOT

Membandingkan distribusi harga jual terakhir untuk berbagai jenis rumah

```
plt.figure(figsize=(8, 6))
sns.boxplot(x='home_type', y='last_sold_price', data=data)
plt.title('Comparison of Last Sold Price by Home Type')
plt.xlabel('Home Type')
plt.ylabel('Last Sold Price')
plt.show()
```



# BOXPLOT

- Tipe Rumah:

Apartment: Median harga sekitar 200.000; harga jual relatif rendah dengan beberapa outliers.

Townhome: Median harga sekitar 300.000; variasi harga lebih besar.

Single Family: Median harga tertinggi sekitar 350.000; rentang harga lebih luas

- Variasi

Single Family memiliki rentang harga yang luas, sedangkan Apartment memiliki penyebaran sempit. Townhome menunjukkan variasi moderat/seimbang.

- Korelasi

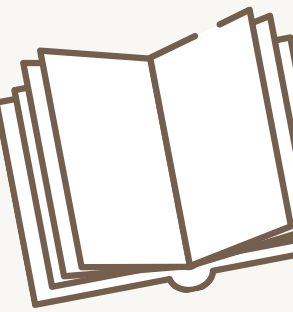
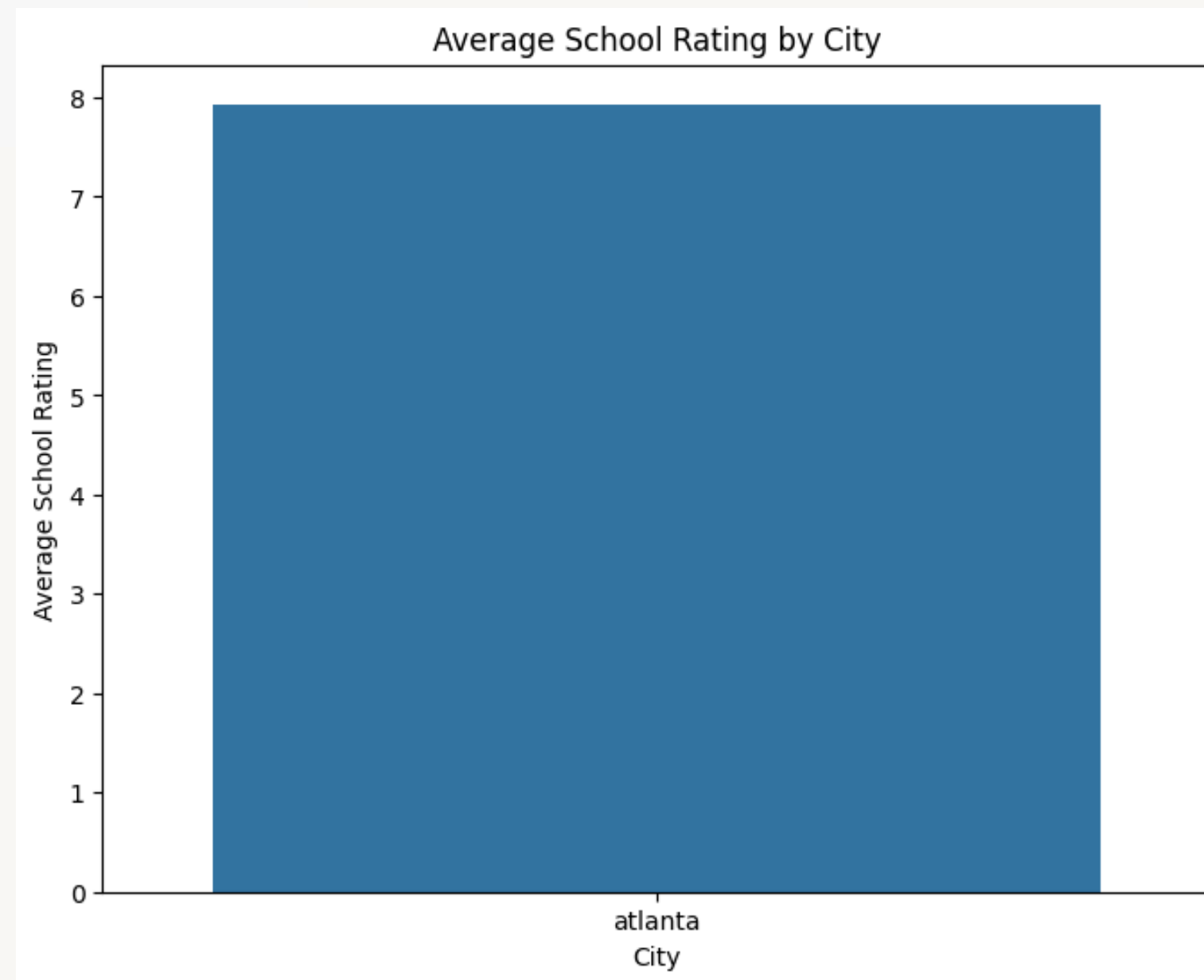
Ada hubungan positif antara tipe rumah dan harga jual: Single Family tertinggi, diikuti Townhome, dan Apartment terendah.

Kesimpulan : tipe rumah berpengaruh signifikan terhadap harga jual, dengan Single Family termahal dan Apartment termurah.

# BAR-CHART

Menampilkan rata-rata peringkat sekolah untuk setiap kota

```
plt.figure(figsize=(8, 6))
sns.barplot(x='city', y='school_rating_1to10', data=data, ci=None)
plt.title('Average School Rating by City')
plt.xlabel('City')
plt.ylabel('Average School Rating')
plt.show()
```

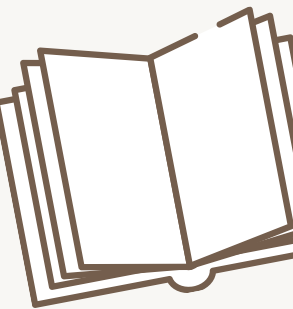


# BAR-CHART

- Kota yang dievaluasi: Hanya ada satu kota, yaitu Atlanta, yang ditampilkan dalam diagram.
- Nilai rata-rata sekolah: Rata-rata penilaian sekolah di kota Atlanta adalah sekitar 8, yang merupakan angka yang relatif tinggi dalam skala 1 sampai 10.

## Kesimpulan

sekolah-sekolah di Atlanta memiliki rata-rata penilaian yang cukup baik, yaitu sekitar 8. Namun, informasi ini terbatas hanya pada satu kota, sehingga tidak ada perbandingan dengan kota-kota lain.

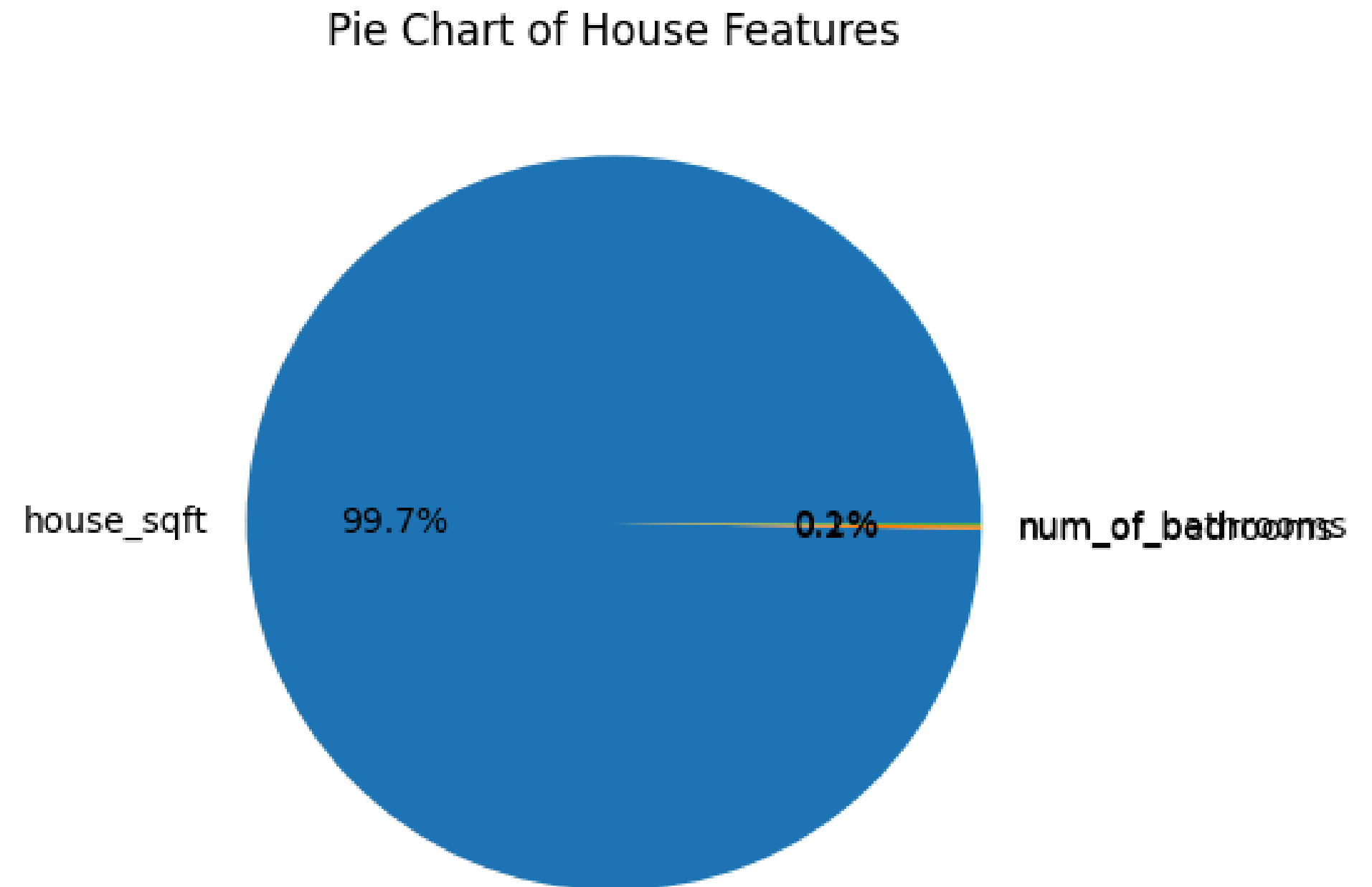


# PIE-CHART

## Membandingkan Fitur Rumah

```
# Data yang akan divisualisasikan
data = {
    'house_sqft': 1770,
    'num_of_bedrooms': 3,
    'num_of_bathrooms': 2
}

# Membuat pie chart
plt.pie(data.values(), labels=data.keys(), autopct='%1.1f%%')
plt.title('Pie Chart of House Features')
plt.show()
```



# PIE-CHART

- Fitur utama: Fitur `house_sqft` (luas rumah dalam satuan kaki persegi) mendominasi dengan persentase 99.7%. Ini menunjukkan bahwa luas rumah memiliki kontribusi yang sangat besar terhadap analisis fitur rumah.
- Fitur lain: Fitur `num_of_bathrooms` (jumlah kamar mandi) hanya berkontribusi sebesar 0.2%. Hal ini menunjukkan bahwa fitur ini memiliki dampak yang sangat kecil dalam analisis.

## Kesimpulan

Dalam konteks analisis ini, luas rumah (`house_sqft`) merupakan faktor paling signifikan, sedangkan jumlah kamar mandi hampir tidak berpengaruh. Mungkin ada ketidakseimbangan dalam data atau penekanan yang berlebihan pada satu fitur.