

Project 02

Housing Price

Outline :

- Overview
- Flowchart
- Visualisasi 1D
- Visualisasi 2D



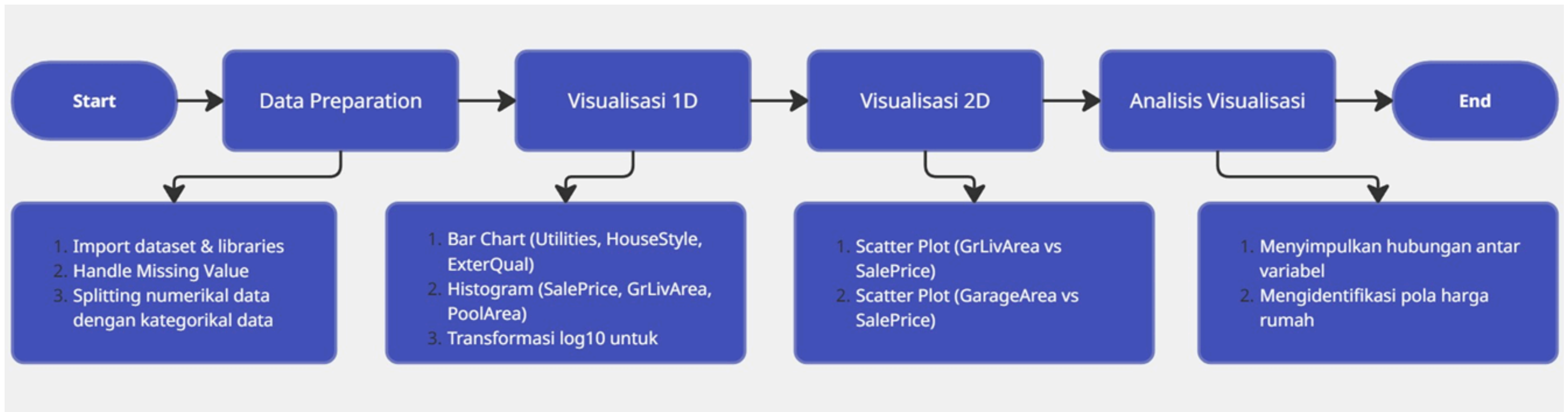
Overview

- Menentukan harga rumah juga bergantung pada berbagai faktor seperti luas bangunan, kualitas eksterior, dan fasilitas. Dataset Housing Price menyediakan informasi properti yang dijual, dan melalui Exploratory Data Analysis (EDA), dapat diidentifikasi pola serta faktor utama yang memengaruhi harga rumah.
- **Tujuan dari proyek ini adalah untuk:**
 - Membersihkan dan mempersiapkan data, termasuk menangani nilai yang hilang (NaN).
 - Menganalisis distribusi variabel numerikal dan kategorikal menggunakan visualisasi 1D.
 - Menganalisis hubungan antar-variabel menggunakan visualisasi 2D untuk melihat bagaimana faktor-faktor tertentu memengaruhi harga rumah.
 - Mendapatkan insight yang dapat membantu dalam pengambilan keputusan terkait valuasi properti.
- Hasil yang diharapkan adalah pemahaman yang lebih baik mengenai faktor-faktor utama yang menentukan harga rumah, serta bagaimana pola distribusi data dapat membantu dalam membuat prediksi atau rekomendasi harga properti.



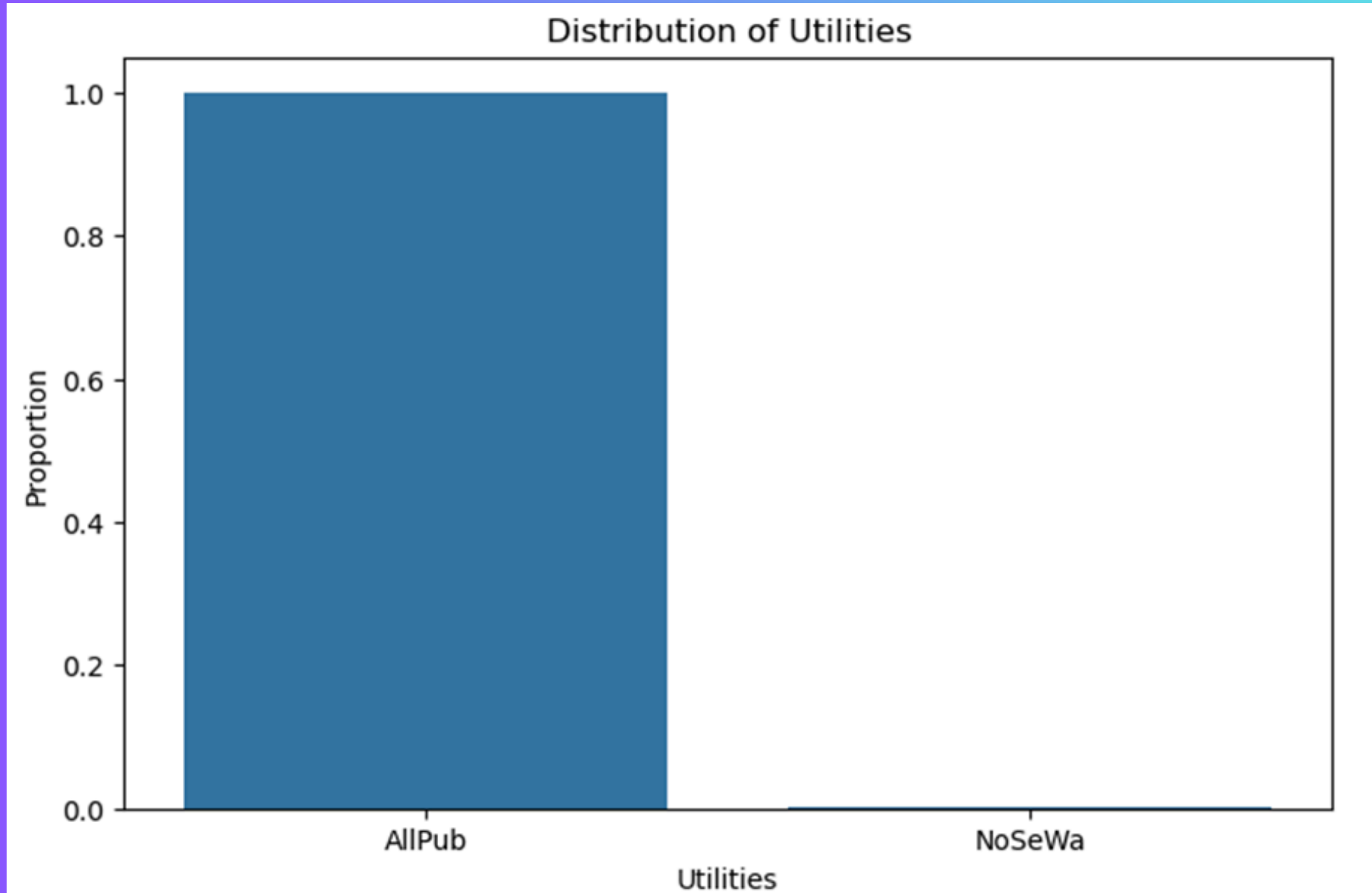
Flow Chart

Housing Price



Analisis 1D

Housing Price

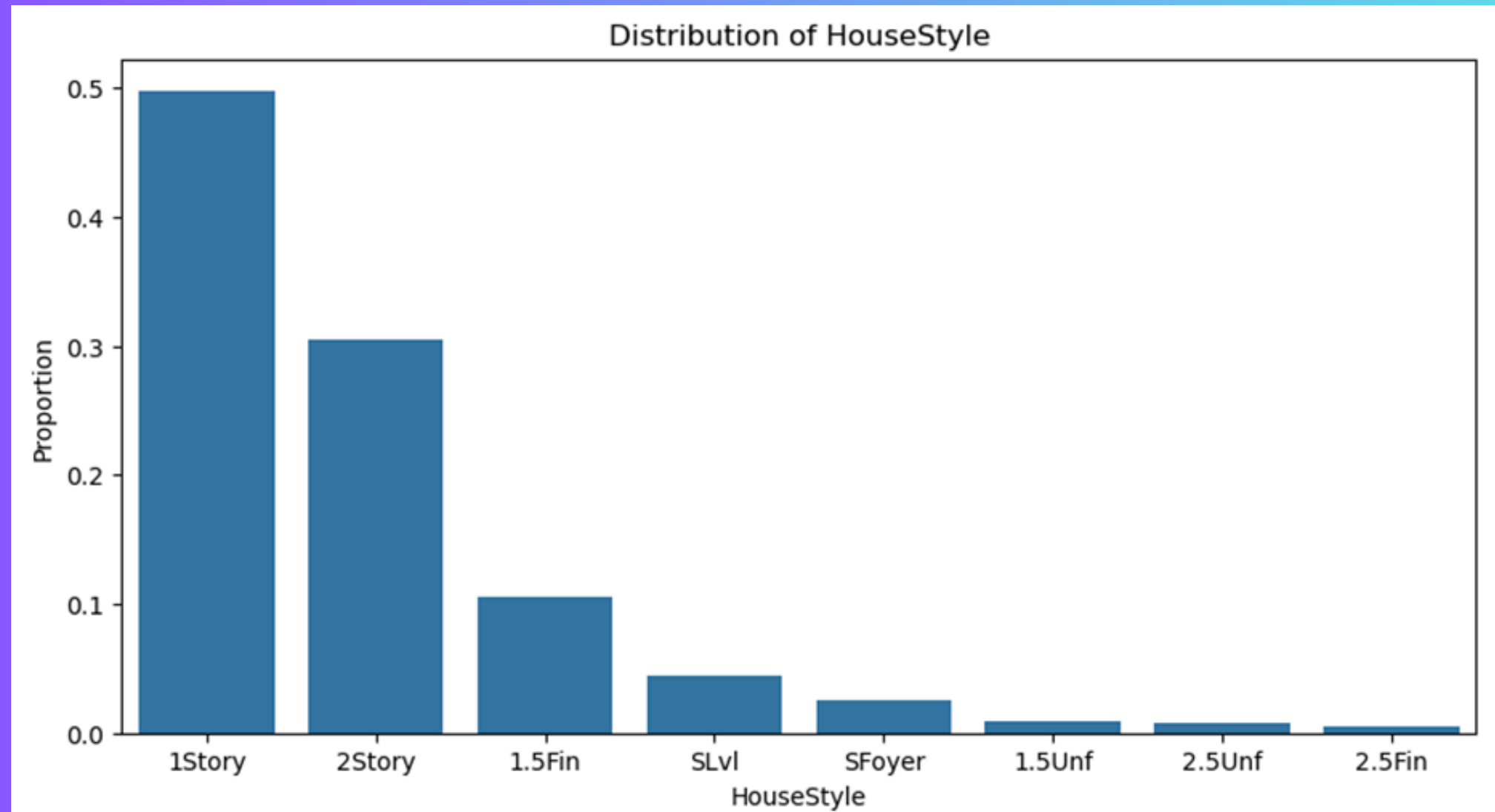


Dari bar chart Distribution of Utilities disamping dapat diketahui bahwa, hampir semua rumah memiliki fasilitas All public Utilities (AllPub), sementara rumah dengan fasilitas Electricity and Gas Only (NoSeWa) sangat jarang. Karena distribusi tidak seimbang, variabel ini kurang informatif dan bisa dipertimbangkan untuk dihapus atau direduksi menjadi variabel biner.



Analisis 1D

Housing Price



HouseStyle: Style properti (misal: 1 lantai, 2 lantai, dst)

1Story One story

1.5Fin One and one-half story: 2nd level finished

1.5Unf One and one-half story: 2nd level unfinished

2Story Two story

2.5Fin Two and one-half story: 2nd level finished

2.5Unf Two and one-half story: 2nd level unfinished

SFoyer Split Foyer

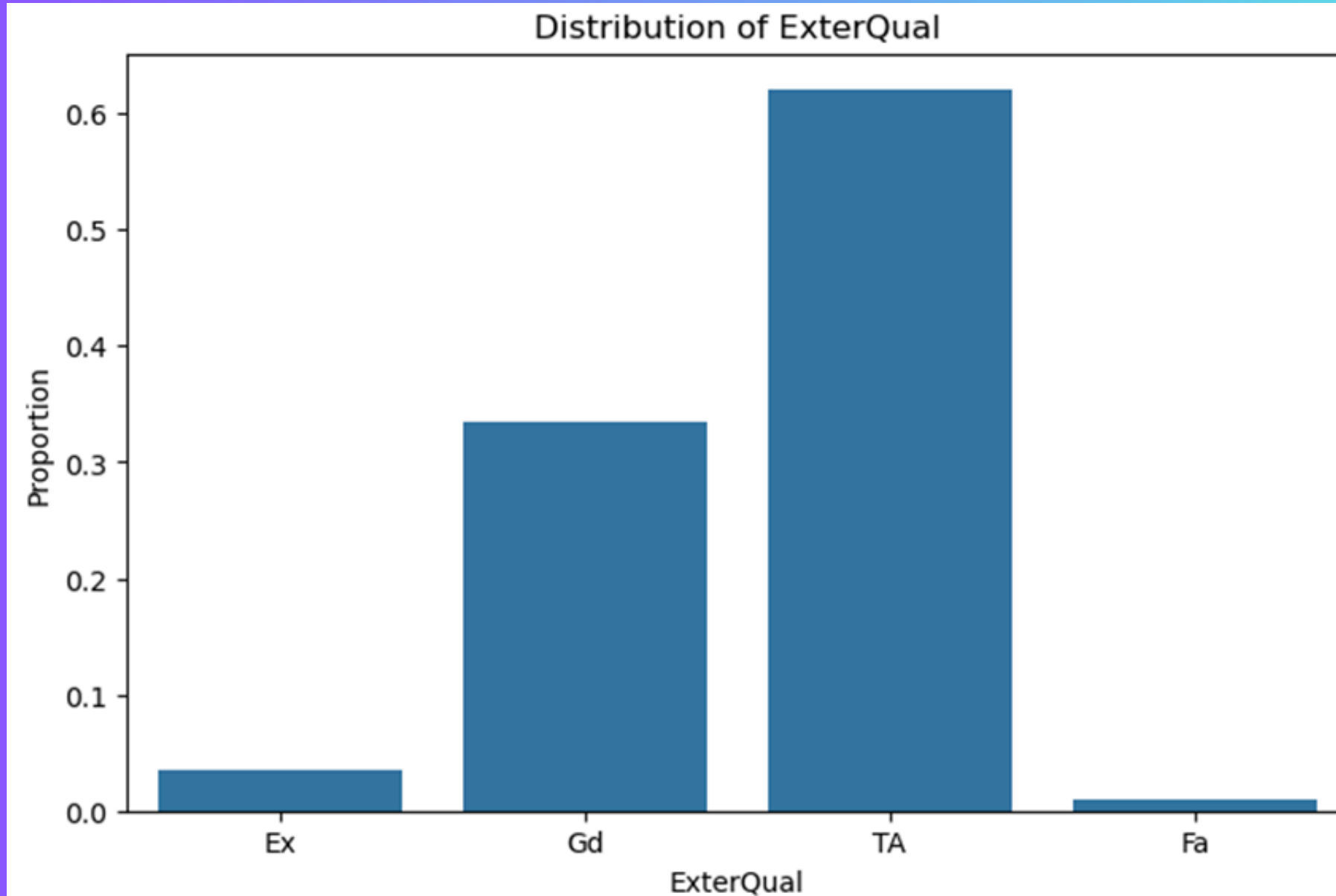
SLvl Split Level

Untuk bar chart Distribution of House Style diketahui bahwa, mayoritas rumah berjenis "1Story" dan "2Story", mencakup lebih dari 80% dataset. Gaya rumah lain seperti "1.5Unf", "2.5Fin", dan "2.5Unf" sangat jarang, menunjukkan bahwa rumah bertingkat sebagian atau split-level kurang umum diminati.



Analisis 1D

Housing Price



ExterQual: Kualitas material eksterior

Ex Excellent

Gd Good

TA Average/Typical

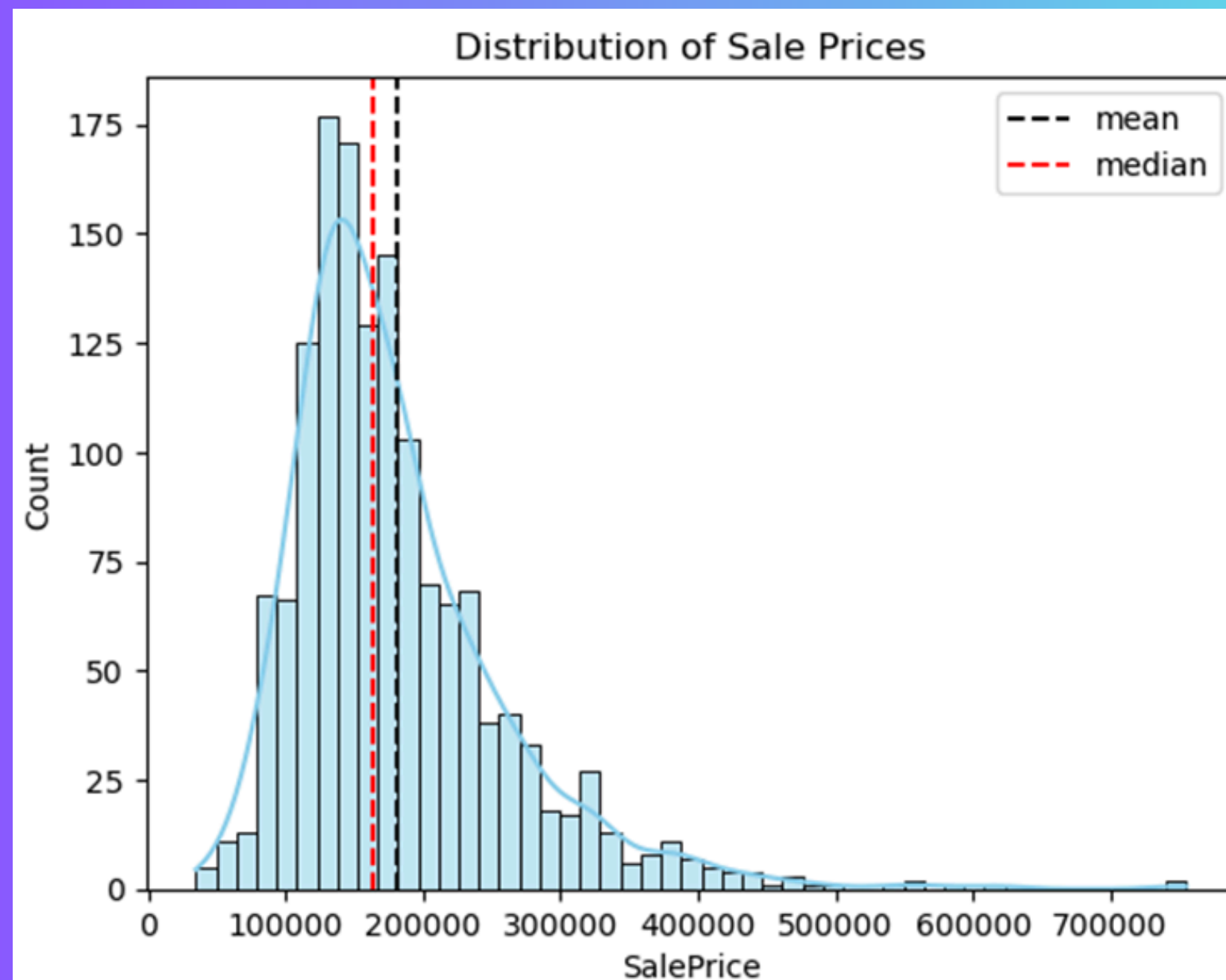
Fa Fair

Sedangkan pada bar chart Distribution of ExterQual, sebagian besar rumahnya memiliki kualitas eksterior "TA" (Typical/Average) dan "Gd" (Good), sementara "Ex" (Excellent) dan "Fa" (Fair) sangat jarang. Ini menunjukkan bahwa kebanyakan rumah memiliki material eksterior standar, dengan sedikit yang sangat bagus atau buruk.



Analisis 1D

Housing Price

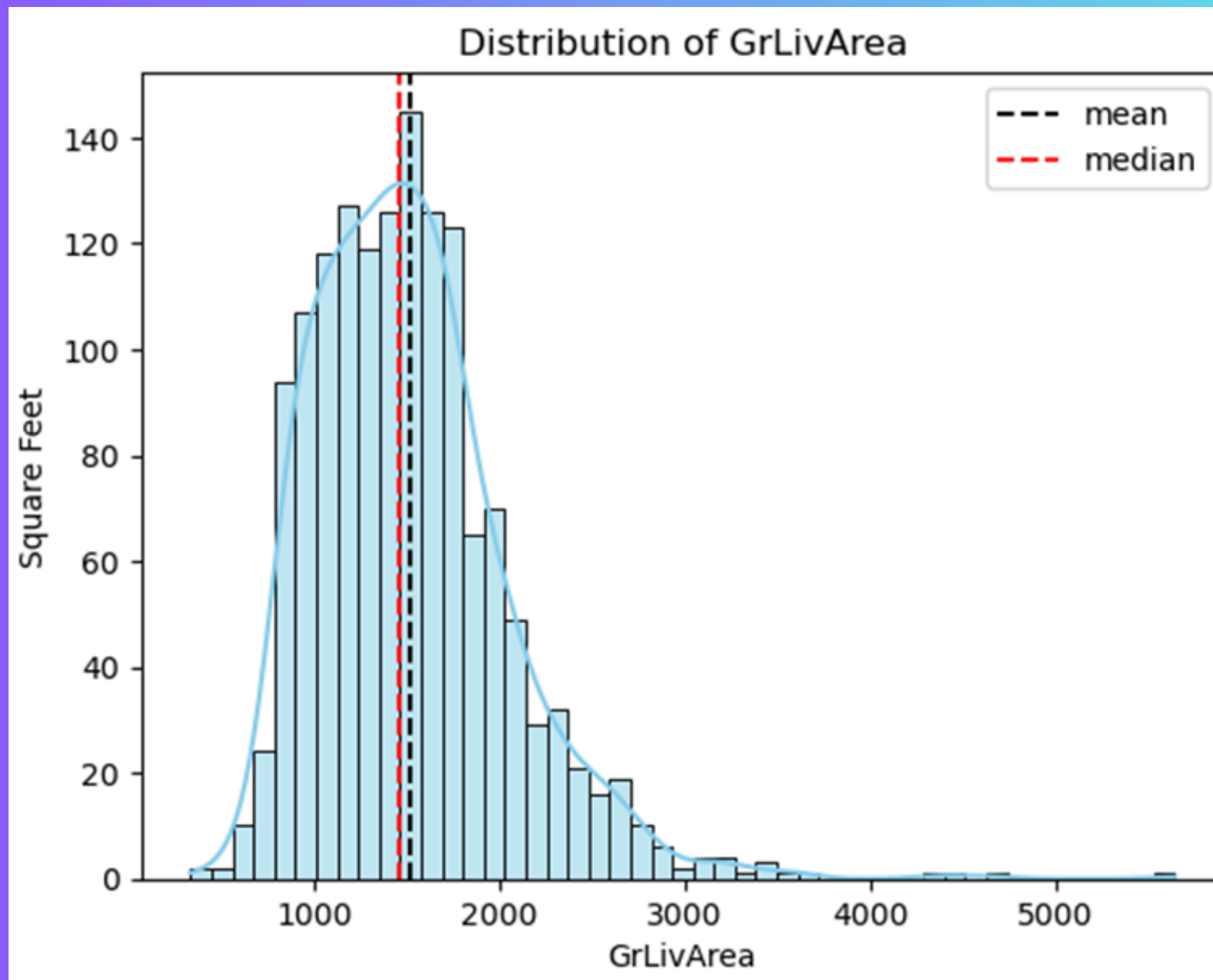


Dari histogram Distribution of Sale Prices diketahui bahwa, distribusi harga rumah memiliki skewness ke arah kanan (Skewed to the Right), dengan nilai mean (180,921) lebih besar dari median (163,000), menunjukkan adanya outlier pada harga tinggi. Sebagian besar rumah berada di kisaran 100,000–200,000, dengan standar deviasi 79,442, yang menandakan variasi harga yang besar. Karena median lebih representatif, analisis lebih lanjut dapat mempertimbangkan transformasi data untuk menangani skewness.



Analisis 1D

Housing Price

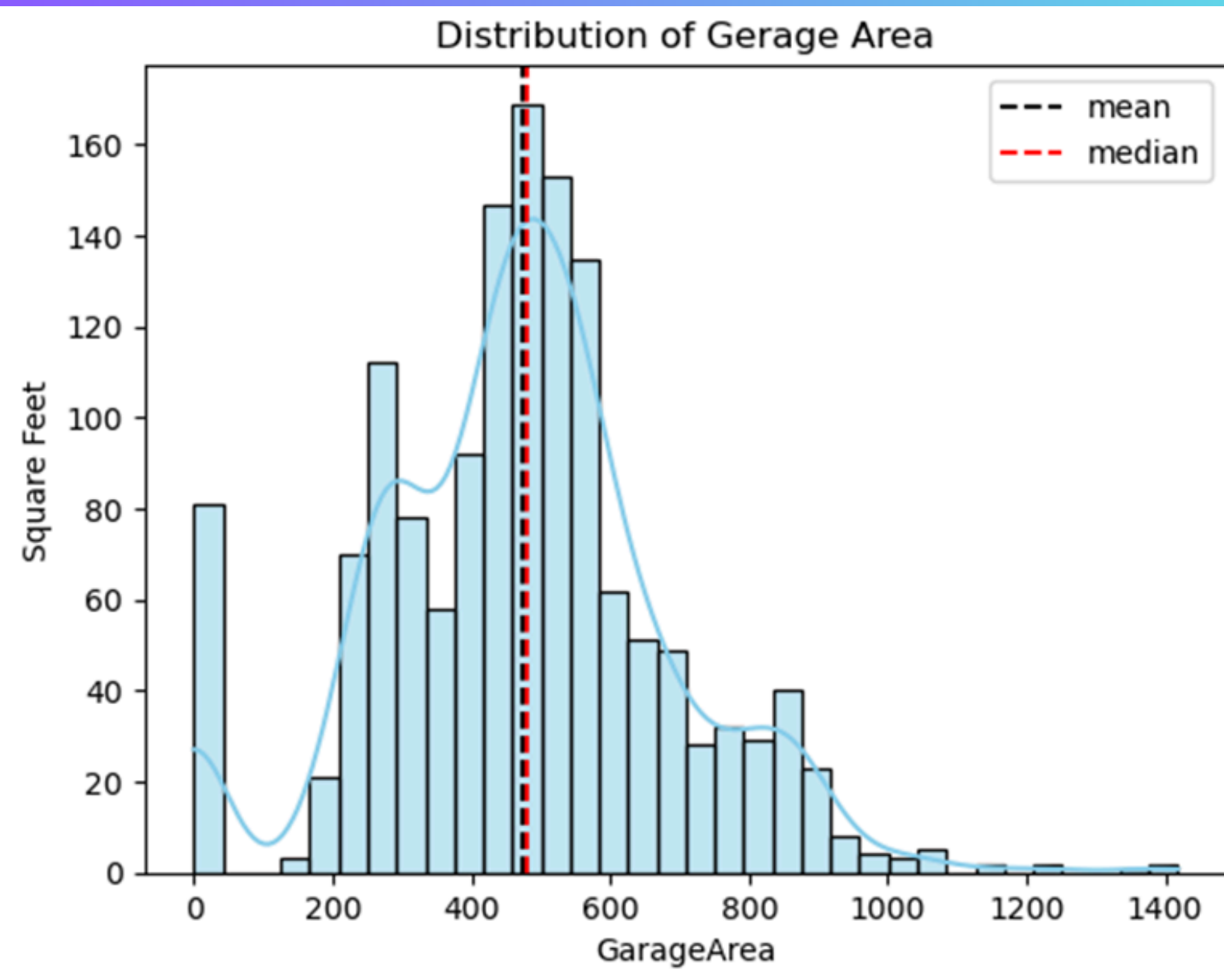


Histogram Distribution of GrLivArea menunjukkan distribusi luas area tinggal di atas tanah dalam satuan square feet. Distribusinya right-skewed, terlihat dari nilai mean yang lebih besar dari median. Hal ini menunjukkan bahwa ada beberapa rumah dengan luas yang jauh lebih besar dari mayoritas, yang kemungkinan menjadi outlier. Sebagian besar rumah memiliki luas sekitar 1.000 - 2.000 square feet, yang terlihat dari puncak distribusi.



Analisis 1D

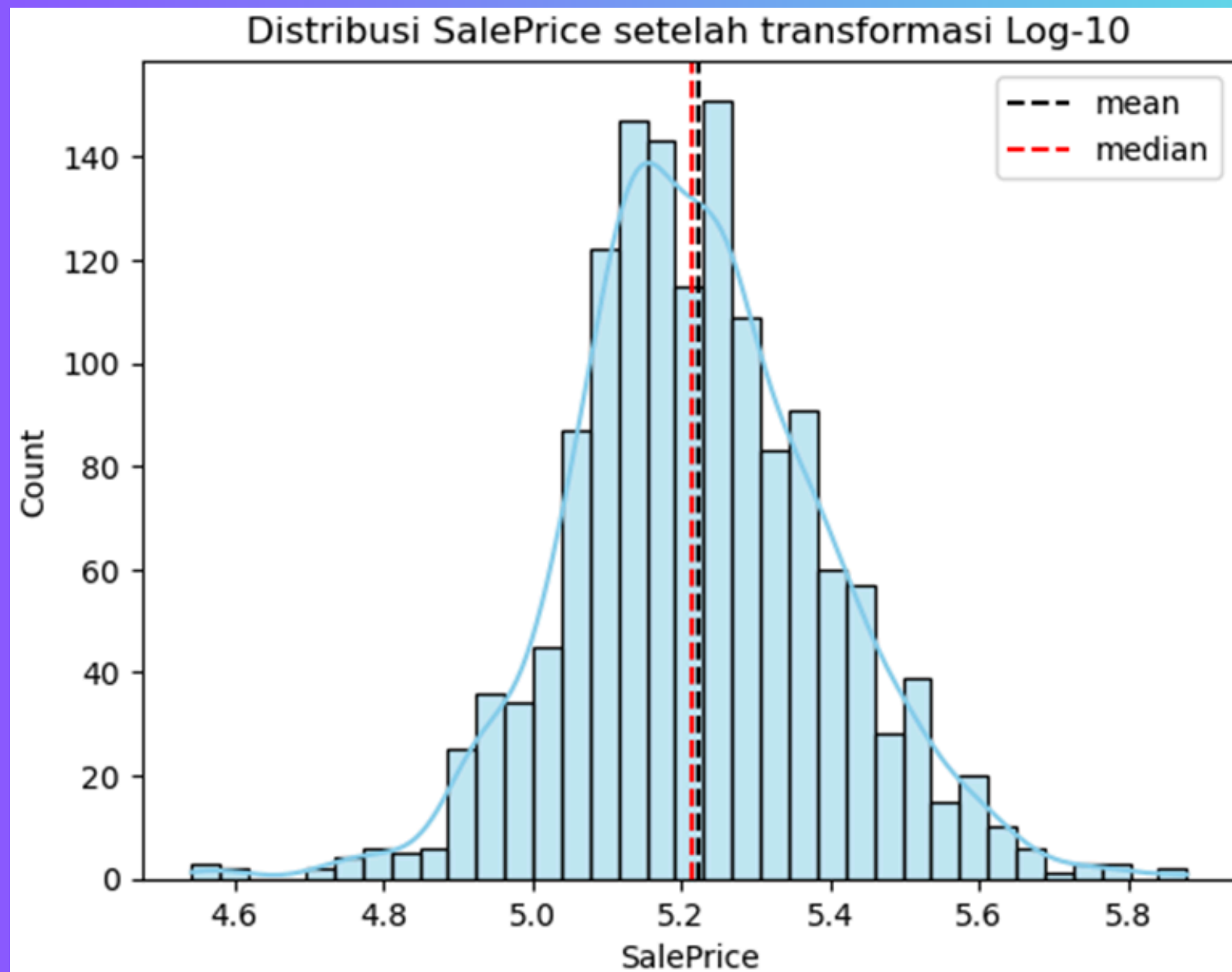
Housing Price



Histogram GarageArea menunjukkan distribusi luas garasi dalam square feet. Distribusinya skewed to the right, terlihat dari nilai mean yang lebih besar dari median walaupun nilainya hanya berbeda tipis tetapi memiliki long tail di kanan, ini menunjukkan adanya beberapa garasi berukuran sangat besar sebagai outlier. Sebagian besar rumah memiliki luas garasi sekitar 400 - 600 square feet, yang mencerminkan kapasitas umum untuk 1-2 mobil. Selain itu, terdapat 81 rumah yang tidak memiliki garasi dari total semua rumah yang ada di dalam dataset.



Analisis 1D

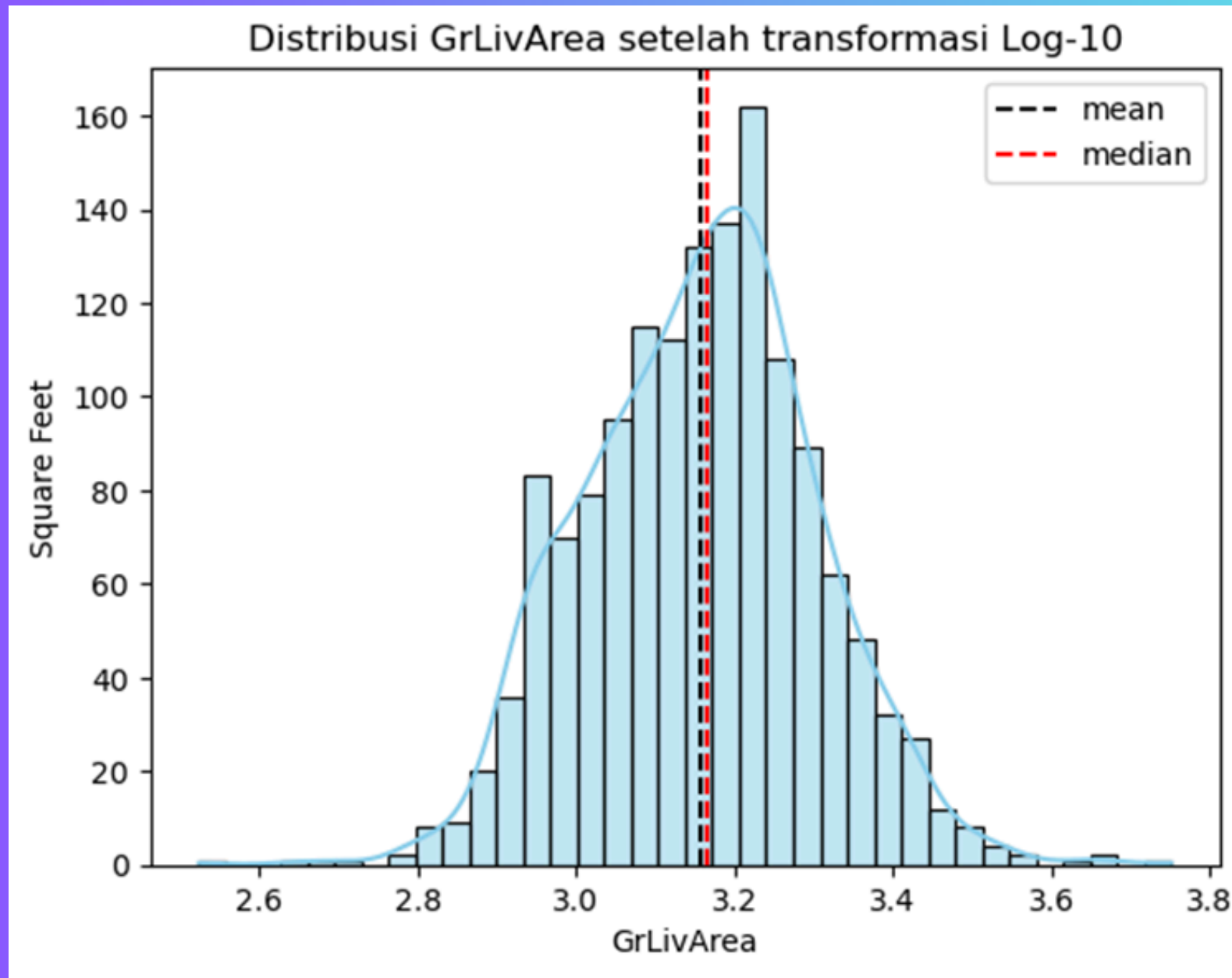


Setelah melakukan transformasi log10 pada SalePrice, nilai mean (5.22) dan median (5.21) hampir berimpit, menunjukkan distribusi yang lebih simetris dibandingkan sebelum transformasi. Standar deviasi yang lebih kecil (0.17) juga mengindikasikan bahwa variasi data lebih terkendali.



Analisis 1D

Housing Price



Sama seperti pada SalePrice, distribusi GrLivArea setelah transformasi log10 juga lebih simetris dengan perbedaan kecil antara nilai mean (3.16) dan median (3.17). Hal ini menunjukkan bahwa transformasi berhasil mengurangi skewness dan membuat data lebih normal.

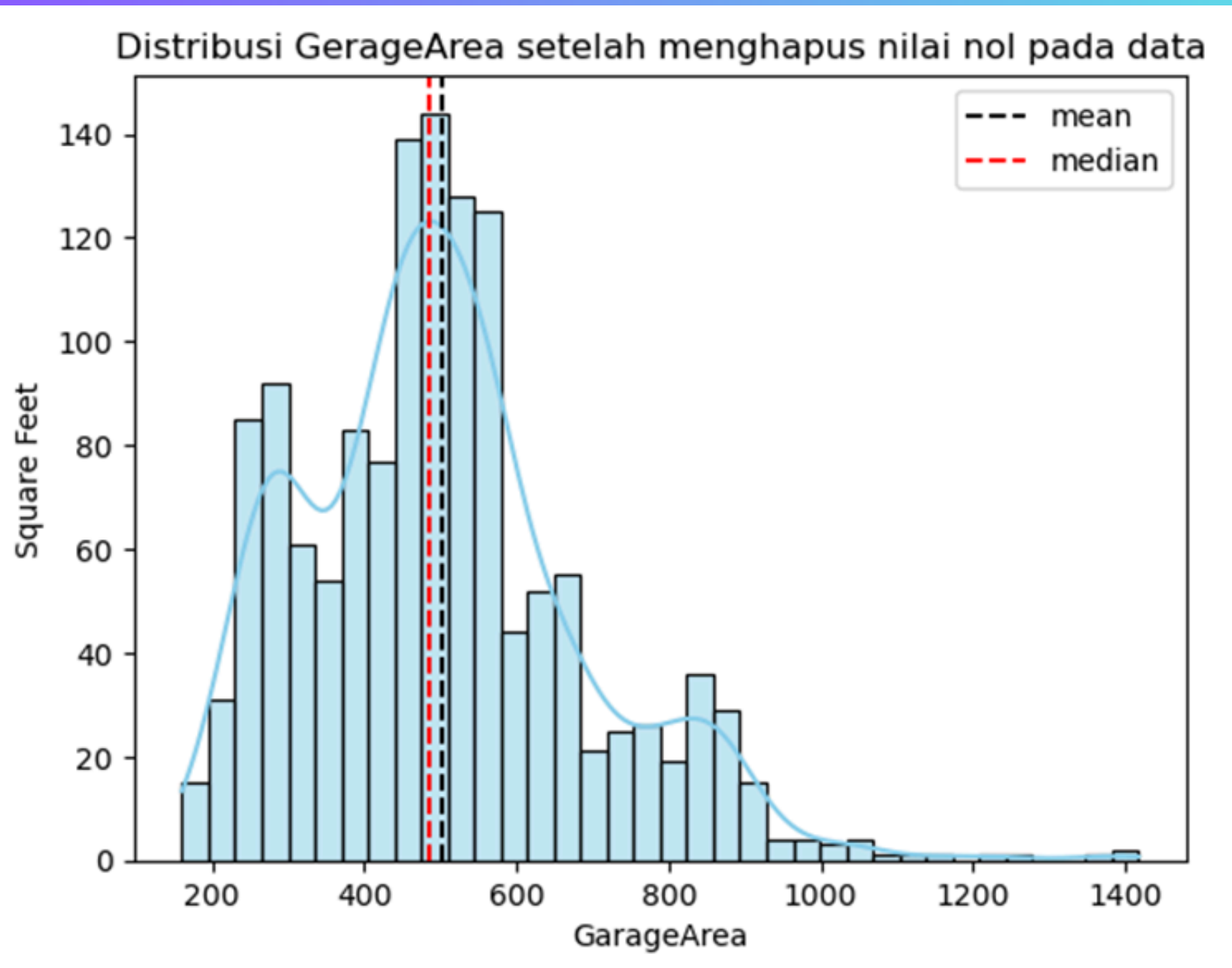


Analisis 1D

Housing Price

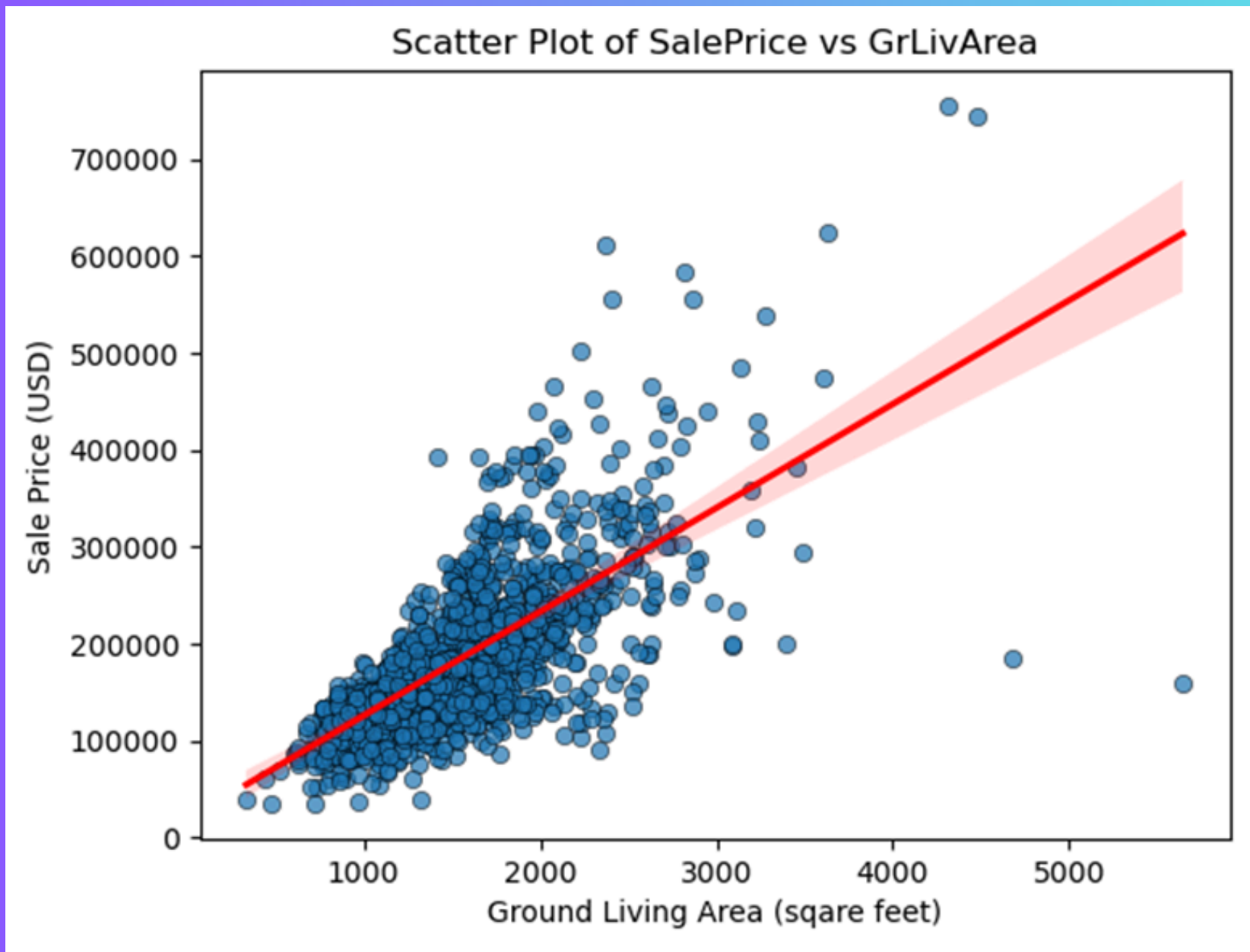
Untuk distribusi GarageArea, setelah menghapus nilai nol, distribusi tetap sedikit skewed ke kanan, dengan nilai mean (500.76) lebih besar dari median (484). Namun, dibandingkan sebelumnya, distribusi lebih terpusat dan memiliki variasi yang lebih terkendali (standar deviasi 185.69).

Secara keseluruhan, transformasi log10 efektif dalam menangani skewness pada SalePrice dan GrLivArea, sedangkan untuk GarageArea, langkah tambahan seperti transformasi atau penanganan outlier mungkin diperlukan agar distribusinya lebih normal.



Analisis 2D

Housing Price

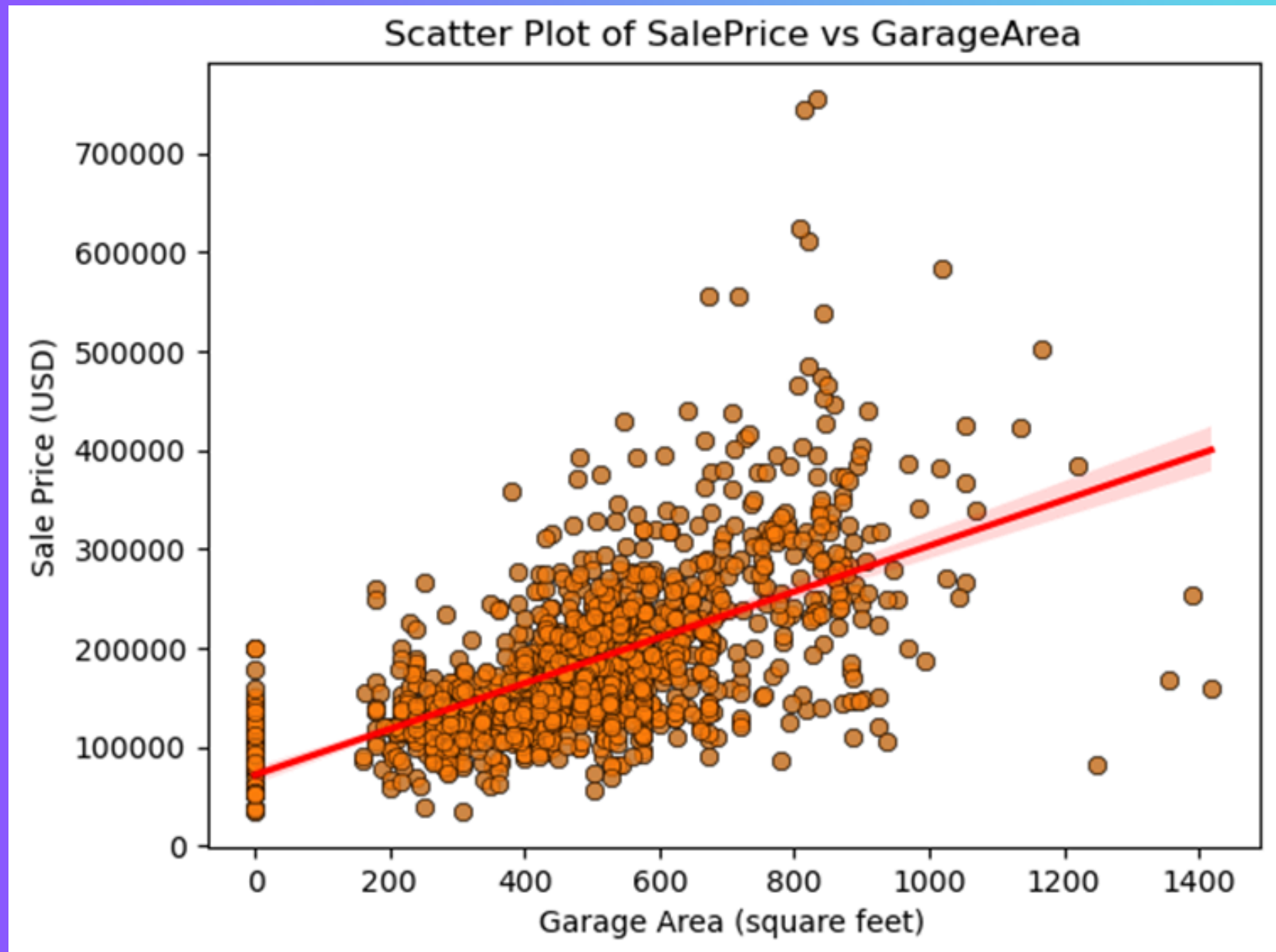


Scatter plot di samping menunjukkan hubungan positif antara Ground Living Area (GrLivArea) dan Sale Price (USD), di mana semakin luas area tempat tinggal, maka semakin tinggi harga jual rumah. Garis regresi merah menandakan tren linear, dengan band merah muda sebagai tingkat ketidakpastian. Beberapa outlier terlihat, seperti rumah berukuran besar dengan harga rendah atau sebaliknya, yang mungkin dipengaruhi oleh faktor lain seperti lokasi dan kualitas bangunan.



Analisis 2D

Housing Price



Sedangkan scatter plot ini menunjukkan hubungan positif antara Garage Area dan Sale Price (USD), di mana semakin luas garasi, maka semakin tinggi harga rumah. Garis regresi merah menunjukkan tren pada scatter plot dengan band ketidakpastian merah muda. Namun, korelasinya tampak lebih lemah dibandingkan dengan GrLivArea vs SalePrice, karena titik data lebih tersebar. Hal ini dapat dimengerti karena ruang utama rumah merupakan faktor utama dalam menentukan kenyamanan dan nilai properti, sedangkan garasi, meskipun

penting, bukanlah faktor utama dalam penentuan harga jual rumah.



Conclusion

- Analisis dataset Housing Price menunjukkan bahwa Ground Living Area (GrLivArea) memiliki korelasi positif yang kuat dengan harga rumah (SalePrice), menjadikannya faktor utama dalam menentukan nilai properti.
- Garage Area juga berpengaruh, meskipun korelasinya lebih lemah. Distribusi harga rumah cenderung right-skewed, dengan beberapa outlier berupa rumah berharga sangat tinggi atau rendah dibandingkan ukuran dan fitur yang dimilikinya.
- Faktor lain seperti kualitas eksterior (ExterQual) juga berkontribusi signifikan terhadap harga, sementara Utilities memiliki pengaruh yang lebih kecil.
- Transformasi log-10 membantu menormalkan distribusi harga untuk analisis yang lebih akurat.
- Secara keseluruhan, harga rumah dipengaruhi oleh kombinasi luas bangunan, fasilitas, dan faktor eksternal lainnya, sehingga pendekatan menyeluruh diperlukan dalam analisis properti.

