

Nama : Mauricio Bethoven Tigauw  
Nim : 1103204099  
Kelas : TK-44-G4

### **Exploratory Data Analysis (EDA)**

Exploratory Data Analysis (EDA) adalah proses untuk memahami dataset dengan melakukan berbagai analisis dan visualisasi data. EDA membantu kita untuk mengidentifikasi pola, hubungan, dan anomali dalam dataset. Berikut adalah beberapa langkah yang dapat dilakukan dalam EDA untuk dataset house pricing:

1. Memahami struktur dataset:
  - Melihat dimensi dataset (jumlah baris dan kolom).
  - Menampilkan beberapa baris pertama dataset untuk melihat struktur data.
  - Memeriksa tipe data dari setiap kolom (numerik, kategorikal, dll.).
  - Mengidentifikasi kolom target (misalnya harga rumah).
2. Menangani missing values:
  - Mengidentifikasi kolom dengan missing values.
  - Memutuskan strategi penanganan missing values (menghapus baris/kolom, mengisi dengan nilai rata-rata/median/modus, dll.).
  - Mengimplementasikan strategi yang dipilih.
3. Melakukan analisis statistik:
  - Menghitung statistik deskriptif untuk variabel numerik (rata-rata, median, standar deviasi, dll.).
  - Melihat distribusi variabel numerik dengan histogram atau box plot.
  - Menghitung frekuensi dan proporsi untuk variabel kategorikal.
4. Mengeksplorasi hubungan antar variabel:
  - Menganalisis korelasi antar variabel numerik dengan menggunakan matriks korelasi atau scatter plot.
  - Melihat hubungan antara variabel numerik dan target dengan scatter plot atau line plot.
  - Menganalisis perbedaan harga rumah berdasarkan variabel kategorikal dengan menggunakan box plot atau bar plot.
5. Melakukan visualisasi data:
  - Membuat visualisasi yang sesuai untuk memvisualisasikan data, seperti histogram, box plot, scatter plot, bar plot, dll.
  - Membuat visualisasi yang berguna untuk memahami hubungan antar variabel, seperti heat map korelasi, pair plot, dll.
6. Melakukan analisis outlier:
  - Mengidentifikasi adanya outlier dalam data numerik dengan menggunakan metode seperti IQR atau Z-score.
  - Memutuskan apakah outlier perlu dihapus atau diatasi.
7. Melakukan analisis time series (jika ada):
  - Jika dataset memiliki variabel waktu, melakukan analisis time series seperti plot tren, plot musimanitas, dll.
8. Membuat kesimpulan:
  - Menarik kesimpulan dari hasil analisis dan visualisasi data.
  - Mengidentifikasi temuan menarik atau pola dalam dataset.
  - Menyusun rekomendasi atau langkah selanjutnya berdasarkan temuan tersebut.

EDA untuk dataset house pricing dapat memberikan wawasan yang berharga tentang faktor-faktor yang mempengaruhi harga rumah dan membantu dalam pengambilan keputusan yang lebih baik dalam bisnis properti.