**5 Jenis Visualisasi Data Populer di Python dengan Matplotlib**

Salah satu bagian penting dalam *data science* adalah visualisasi data atau *data visualization*. Saat kita melakukan analisis untuk menemukan *insight* dari suatu data, kita akan sangat membutuhkan visualisasi agar data yang dimiliki lebih mudah dipahami. Apalagi jika dataset yang sedang dianalisis berukuran sangat besar, tentunya visualisasi data akan sangat membantu.

[Matplotlib](https://matplotlib.org/3.3.3/index.html) adalah salah satu *library* yang populer yang dapat digunakan dengan mudah untuk *data visualization*. Pada artikel ini saya akan membahas 5 jenis visualisasi data di Python menggunakan Matplotlib yang sering digunakan untuk menampilkan berbagai informasi dari suatu dataset.

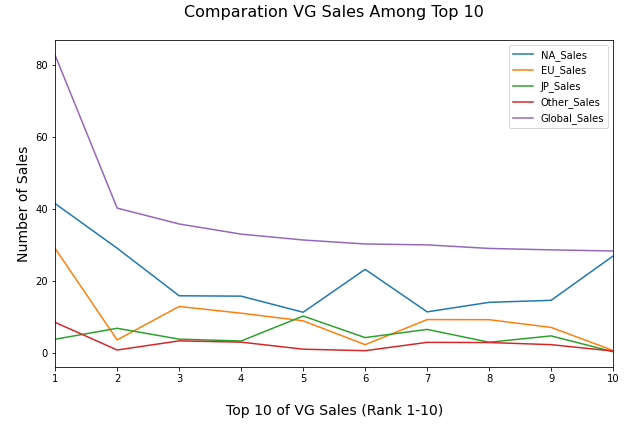
**Line Plot**

Line plot atau line chart adalah jenis plot yang menampilkan informasi berupa rangkaian titik data yang terhubung dengan segmen garis lurus. Line plot merupakan jenis plot dasar dalam visualisasi data. Plot jenis ini dapat digunakan pada dataset yang memiliki nilai kontinu untuk melihat pergerakan data.

Contoh kode Python untuk membuat Line Plot menggunakan dataset Video Game Sales

df\_join = df\_vg[['Rank','NA\_Sales', 'EU\_Sales', 'JP\_Sales', 'Other\_Sales', 'Global\_Sales']].head(10)  
df\_join = df\_join.set\_index('Rank')  
  
# create line plot  
df\_join.plot(kind='line', figsize=(10, 6))  
  
plt.title('Comparation VG Sales Among Top 10\n', size=16)  
plt.ylabel('\nNumber of Sales', size=14)  
plt.xlabel('\nTop 10 of VG Sales (Rank 1-10)', size=14)  
plt.show()

**Output:**



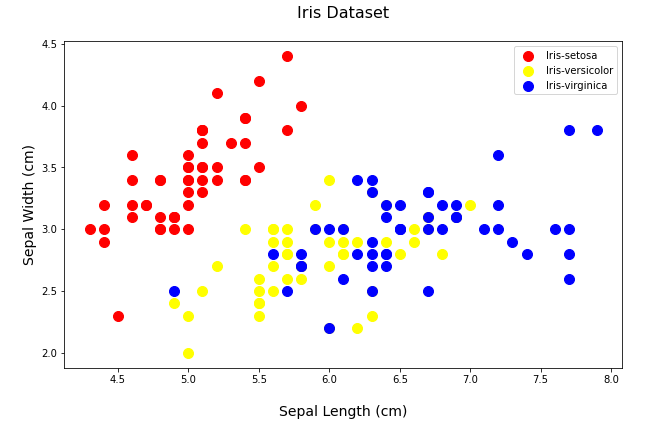
**Scatter Plot**

Scatter plot adalah jenis visualisasi data untuk menunjukkan hubungan antara dua variabel. Dengan scatter plot, Anda dapat melihat distribusi data dan bagaimana *trend*-nya. Anda juga dapat menggunakan scatter plot untuk melihat hubungan dari beberapa kelompok data dengan memberikan warna yang berbeda untuk membedakan tiap kelompok data.

Contoh kode Python untuk membuat scatter plot menggunakan dataset Iris

fig, ax = plt.subplots(figsize=(10, 6))  
  
df\_group = df.groupby('class')  
colors = {'Iris-setosa':'red', 'Iris-versicolor':'yellow', 'Iris-virginica':'blue'}  
  
# create scatter plot  
for name, group in df\_group:  
    group.plot(ax=ax, kind='scatter', x='sepal\_length', y='sepal\_width', label=name, color=colors[name], s=100)  
  
plt.title('Iris Dataset\n', size=16)  
plt.xlabel('\nSepal Length (cm)', size=14)  
plt.ylabel('\nSepal Width (cm)', size=14)  
plt.show()

**Output:**



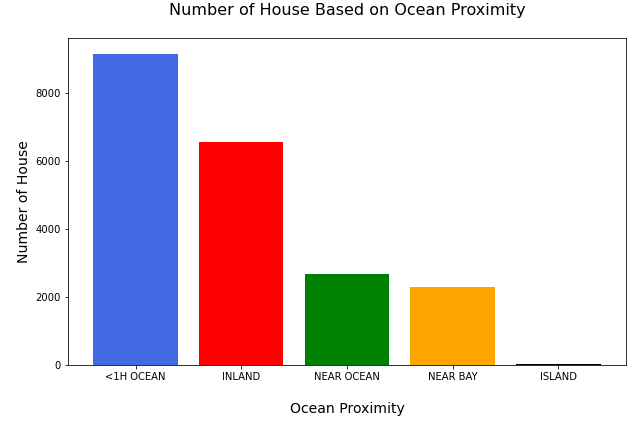
**Bar Plot**

Bar plot atau bar chart adalah jenis plot yang direpresentasikan dengan bar, dimana panjang bar adalah representasi dari ukuran sebuah fitur atau variabel. Bar plot biasanya merepresentasikan variabel numerik dan kategori yang dikelompokkan dalam interval.

Contoh kode Python untuk membuat bar plot menggunakan dataset California Housing Prices

count = df\_pri['ocean\_proximity'].value\_counts()  
plt.figure(figsize=(10, 6))  
  
# create bar plot  
plt.bar(['<1H OCEAN', 'INLAND', 'NEAR OCEAN', 'NEAR BAY', 'ISLAND'], count, color=['royalblue','red','green', 'orange', 'black'])  
  
plt.xlabel('\nOcean Proximity', size=14)  
plt.ylabel('\nNumber of House', size=14)  
plt.title('Number of House Based on Ocean Proximity\n', size=16)  
plt.show()

**Output:**



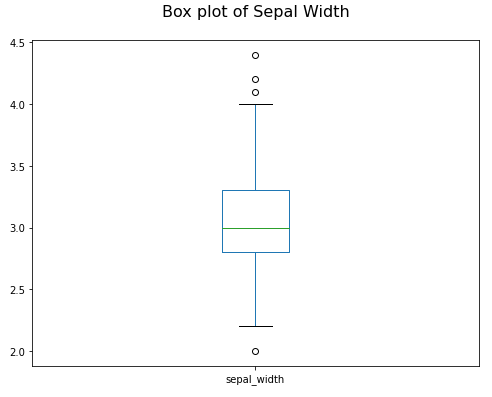
**Box Plot**

Box plot adalah jenis visualisasi data yang secara statistik merepresentasikan distribusi data melalui lima dimensi utama, yaitu nilai minimum, kuartil 1, kuartil 2 (median), kuartil 3, dan nilai maksimum. Biasanya box plot digunakan untuk memeriksa keberadaan *outlier* dalam dataset.

Contoh kode Python untuk membuat box plot menggunakan dataset Iris

# create box plot  
df['sepal\_width'].plot(kind='box', figsize=(8, 6))  
  
plt.title('Box plot of Sepal Width\n', size=16)  
plt.show()

**Output:**



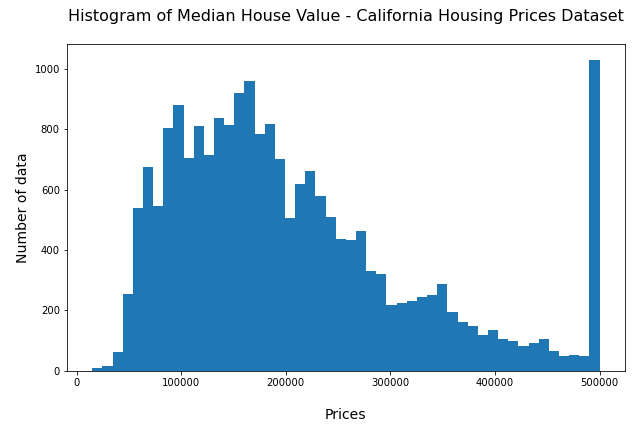
**Histogram**

Histogram adalah jenis visualisasi data untuk merepresentasikan distribusi frekuensi dari dataset numerik. Sumbu X merepresentasikan bin dari *data point*, sedangkan sumbu Y merepresentasikan frekuensi atau banyaknya *data point* pada setiap bin. Ukuran bin dapat disesuaikan agar distribusinya terlihat bagus. Namun, *default*-nya Python akan membagi menjadi 10 bin.

Contoh kode Python untuk membuat histogram menggunakan dataset California Housing Prices

# create histogram  
df\_pri['median\_house\_value'].plot(kind='hist', figsize=(10, 6), bins=50)  
  
plt.title('Histogram of Median House Value - California Housing Prices Dataset\n', size=16)  
plt.ylabel('\nNumber of data', size=14)  
plt.xlabel('\nPrices', size=14)  
plt.show()

**Output:**



Contoh-contoh di atas adalah visualisasi data sederhana yang dapat dibuat dengan Matplotlib di Python. Anda juga dapat membuat plot-plot yang lebih kompleks dengan mengatur parameternya serta memanfaatkan fungsi-fungsi lain yang dapat diterapkan.