

VISUALISASI DATA

Arif Laksito, M. Kom www.ariflaksito.net



VISUALISASI DATA

Visualisasi data merupakan salah satu hal penting untuk mempermudah memahami data

Dalam menganalisa data, sering kali menggunakan visualisasi data untuk memahami karakteristik maupun trend di dalam data yang dimiliki.

Dalam Python kita bisa dengan mudah melakukan visualisasi data dengan library Matplotlib atau library Seaborn

Seaborn adalah library untuk visualisasi data yang dibangun di atas matplotlib. Seaborn memiliki lebih banyak fungsi untuk visualisasi data dan lebih mudah digunakan



MATPLOTLIB

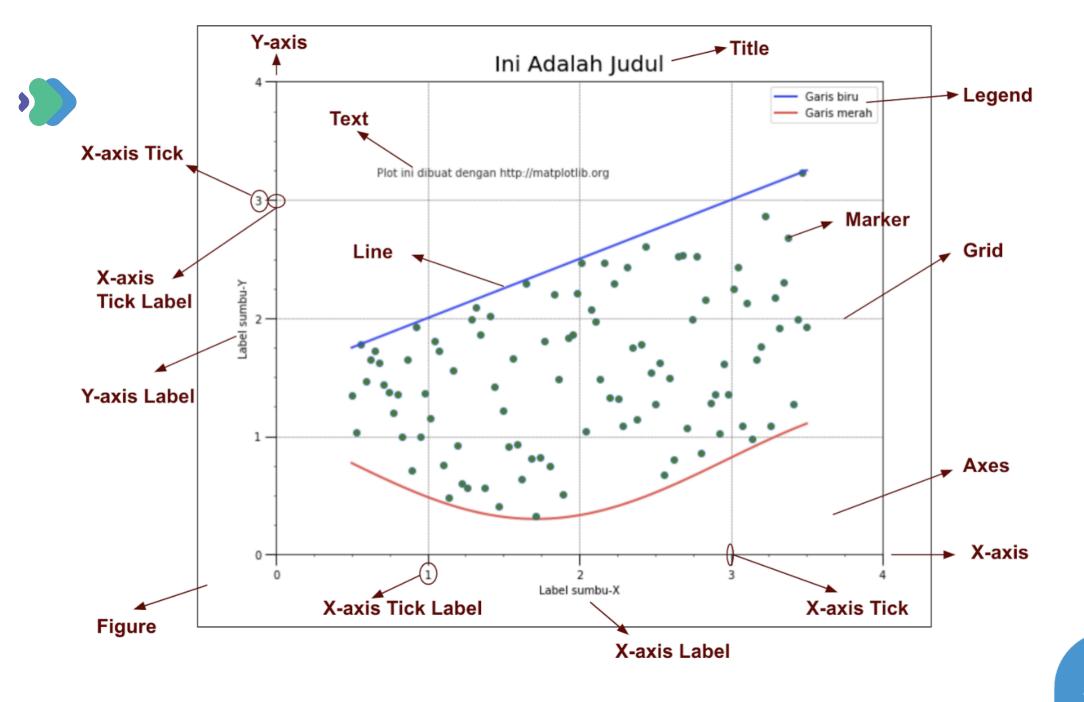
matplotlib.pyplot adalah kumpulan fungsi yang membuat beberapa perubahan pada gambar: membuat gambar, membuat area plot dalam gambar, menambah label di plot dan lainnya

Biasanya untuk mempermudah secara umum matplotlib.pyplot disingkat menjadi plt: import matplotlib.pyplot as plt

Biasanya

Gunakan fungsi plot() untuk melakukan plotting terhadap nilai X dan Y. Fungsi ylabel() untuk mencetak label di sumbu Y dan fungsi xlabel() untuk mencetak label sumbu X









Ada beberapa istilah dan komponen di sebuah plot:

Figure adalah keseluruhan visualisasi yang kita plot dalam 1 kali menjalankan code.

Sedangkan satu plot (yang dibuat saat memanggil .plot() atau .scatter()) disebut Axes. Sebuah Figure bisa terdiri dari beberapa Axes.

Setiap Axes biasanya memiliki sumbu-X (X-axis) dan sumbu-Y (Y-axis). Masing-masing sumbu memiliki komponen sebagai berikut:

Axis Label

Tick

Tick Label

Untuk isi grafiknya sendiri, bisa berupa line (untuk line plot), atau marker (untuk scatter plot), bisa juga bentuk lain seperti bar (untuk bar plot / histogram).

Aksesoris lain yang bisa ditambahkan, di antaranya Grid untuk mempermudah melihat tick yang sejajar, dan Text untuk memberikan informasi tambahan berbentuk teks di grafik.





```
plt.figure(figsize=(15, 5))
dataset.groupby(['order_month'])['gmv'].sum().plot()
plt.title('Monthly GMV Year 2019', loc='center', pad=40, fontsize=20, color='blue')
plt.xlabel('Order Month', fontsize=15)
plt.ylabel('Total Amount', fontsize=15)
plt.show()
```

Monthly GMV Year 2019







Seaborn merupakan API yang berdasar dari Matplotlib, memiliki high-level function untuk tipe plot yang umum digunakan untuk statistik.

Seaborn juga dapat diintegrasikan dengan Dataframes dari Pandas.

Ada beberapa jenis plot yang dapat dilakukan dengan seaborn seperti



SEABORN DATASET

Seaborn juga mempunyai repository dataset yang dapat digunakan untuk eksperimen

Untuk load dataset gunakan fungsi load_dataset()

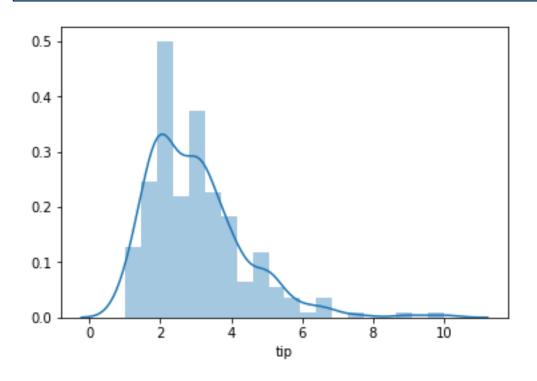
Ada banyak dataset yang dapat kita gunakan seperti titanic.csv, tips.csv, iris.csv, flight.csv, planets.csv dan lain sebagainya

Untuk mengetahui dataset apa saja yang disediakan seaborn gunakan fungsi **get_dataset_names()**

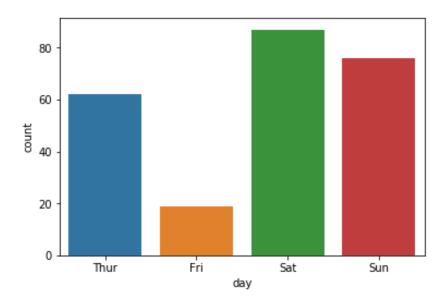




tips = sns.load_dataset('tips')
sns.distplot(tips['tip'], bins=20, kde=True)



tips = sns.load_dataset('tips')
sns.countplot(x='day', data=tips)







DQLab, Data visualization part 1, https://academy.dqlab.id/main/package/practice/164

Ngodingdata.com, Visualisasi Matplotlib: Mengenal Library Matplotlib, https://ngodingdata.com/pengenalan-library-matplotlib/

Ngodingdata.com, Tutorial Seaborn: Pengenalan Library Seaborn, https://ngodingdata.com/pengenalan-library-seaborn/



TERIMAKASIH

about us:

www.tlab.co.id