



Institut Informatika & Bisnis
DARMAJAYA
Kepercayaan Allah dan Manusia



**Kampus
Merdeka**
INDONESIA JAYA

**MERDEKA
BELAJAR**

Statistika untuk Sains Data

SSD23401

PERTEMUAN 1

Egi Safitri, S.Mat., M.Si

`egisafitri@ darmajaya.ac.id`

September 28th, 2023



Perkenalan

Kontrak Kuliah

Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor

Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



- **Kehadiran** : 70% dari 16 kali pertemuan.
- **Keterlambatan** : Max. 20 Menit
- **Aktif dalam pembelajaran**
- **Pakaian** : Rapih, Sopan, sesuai dengan aturan Kampus.
 - 1 Laki-laki : Kemeja / Kaos BERKERAH , Bersepatu
 - 2 Wanita : Kemeja/Blouse, Sopan, Bersepatu.
- **Penilaian** : UTS, UAS, Tugas dan Keaktifan di dalam kelas



Perkenalan

Kontrak Kuliah

Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data

Analisis Data di Beberapa Sektor

Siklus Analisis Data

Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data

Memilih Visualisasi Data yang Tepat



Daftar Isi

- 1 Perkenalan**
 - Kontrak Kuliah
 - Daftar Isi
- 2 Pengantar Analisis Data Eksploratif**
 - Pengenalan Analisis Data
 - Analisis Data di Beberapa Sektor
 - Siklus Analisis Data
 - Langkah-Langkah Analisis Data
- 3 Data Preprocessing**
 - Pengantar Data Preprocessing
- 4 Statistika Deskriptif**
 - Ukuran Pemusatan Data
- 5 Visualisasi Data**
 - Pentingnya Visualisasi Data
 - Memilih Visualisasi Data yang Tepat



Perkenalan

- Kontrak Kuliah
- Daftar Isi

Pengantar Analisis Data Eksploratif

- Pengenalan Analisis Data
- Analisis Data di Beberapa Sektor
- Siklus Analisis Data
- Langkah-Langkah Analisis Data

Data Preprocessing

- Pengantar Data Preprocessing

Statistika Deskriptif

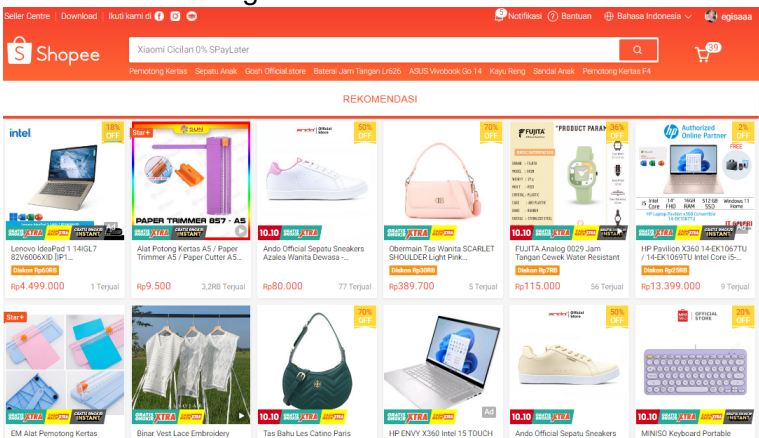
- Ukuran Pemusatan Data

Visualisasi Data

- Pentingnya Visualisasi Data
- Memilih Visualisasi Data yang Tepat



Sebelum membahas terkait analisis data, mari simak gambar di bawah ini.





Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



Analisis data adalah proses menyelidiki, membersihkan, mengubah, dan memodelkan data dengan tujuan menemukan informasi yang berguna, menginformasikan kesimpulan, mendukung pengambilan keputusan, dan masih banyak lagi.

Mengapa data begitu penting?



Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

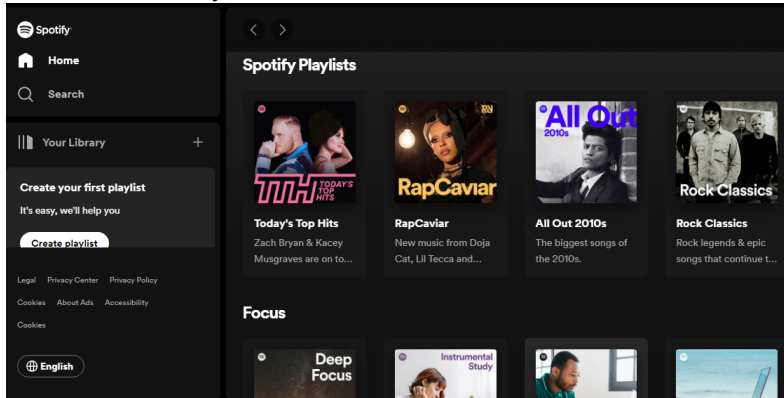
Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



Berikut merupakan peran analisis data di beberapa sektor. Contohnya di sektor **media dan hiburan**.





Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa
Sektor
Siklus Analisis Data
Langkah-Langkah Analisis
Data

Data Preprocessing

Pengantar Data
Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi
Data
Memilih Visualisasi Data
yang Tepat



Analisis data di bidang **retail dan wholesale**.





Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

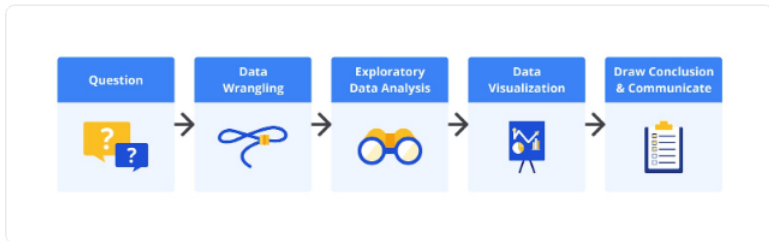
Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



Setelah memahami semua pembahasan dalam materi Berkenalan dengan Analisis Data, pada bagian ini, kita akan fokus pada tahapan analisis data yang umum dijumpai di industri. Siklus analisis data umumnya seperti yang tertera di bawah ini.





Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



Mendefinisikan Pertanyaan

Pada awal proyek analisis data, kita perlu mendefinisikan berbagai pertanyaan analisis terkait data ataupun permasalahan yang ingin diselesaikan. Pertanyaan tersebut akan membantu kita dalam memahami masalah bisnis yang dihadapi serta tujuan yang ingin dicapai. Bukan hanya itu, dengan adanya pertanyaan tersebut, kita akan merasa terbantu karena bisa lebih fokus pada bagian data yang relevan dengan tujuan.



Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



Data Wrangling

Setelah memiliki pertanyaan analisis yang baik, tahap selanjutnya adalah data wrangling. Tahap ini dimulai dengan mengumpulkan data yang dibutuhkan untuk menjawab pertanyaan analisis yang telah dibuat sebelumnya. Kemudian Anda perlu menilai kualitas dan struktur dari data tersebut. Hal ini dilakukan untuk mengidentifikasi masalah yang terdapat dalam data serta membuat strategi pembersihan data yang sesuai.



Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



Exploratory Data Analysis (EDA)

Tahap ini merupakan tahap eksplorasi terhadap data yang telah dibersihkan untuk memperoleh *meaningful insights* dan menjawab pertanyaan analisis. EDA memungkinkan *analyst* memahami isi data yang digunakan, mulai dari distribusi, frekuensi, korelasi, dan lainnya. Dalam praktiknya, memiliki rasa *curiosity* merupakan hal yang sangat penting. Pemahaman dalam konteks data juga perlu diperhatikan karena hal tersebut dapat menjawab masalah-masalah dasar.



Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



Data Visualization

Hasil yang diperoleh dari EDA perlu divisualisasikan melalui tahap data visualization. Pada tahap ini, kita akan menerapkan berbagai teknik dan jenis visualisasi data yang tepat untuk mengomunikasikan temuan atau insight dari hasil analisis secara efektif.



Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



Data Visualization

Hasil yang diperoleh dari EDA perlu divisualisasikan melalui tahap data visualization. Pada tahap ini, kita akan menerapkan berbagai teknik dan jenis visualisasi data yang tepat untuk mengomunikasikan temuan atau insight dari hasil analisis secara efektif.

Draw conclusion & communicate

Pada akhir proyek analisis data, kita perlu membuat conclusion atau kesimpulan dari hasil analisis. Kesimpulan yang dibuat haruslah menjawab semua pertanyaan dan tujuan yang telah didefinisikan di awal.



Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



Menurut TechTarget

Data preprocessing adalah segala bentuk pemrosesan terhadap data mentah agar siap diproses dalam tahapan analisis data berikutnya.

Dalam data preprocessing (pra-pemrosesan data), sejumlah langkah penting dilakukan untuk mempersiapkan data mentah sebelum dianalisis atau dimodelkan.



Perkenalan

Kontrak Kuliah

Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data

Analisis Data di Beberapa Sektor

Siklus Analisis Data

Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data

Memilih Visualisasi Data yang Tepat



Berikut adalah beberapa aspek utama yang dipelajari dalam data preprocessing :

- Pembersihan Data (Data Cleaning).
- Pengisian Data yang Hilang (Data Imputation)
- Transformasi Data (Data Transformation)
- Penghapusan Atribut yang Tidak Diperlukan (Feature Selection)
- Ekstraksi Fitur (Feature Extraction)
- Penanganan Data Kategorikal (Categorical Data Handling)



Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



Berikut adalah beberapa aspek utama yang dipelajari dalam data preprocessing :

- Penghapusan Data yang Duplikat (Duplicate Data Removal)
- Penanganan Outlier (Outlier Handling)
- Sampling Data (Data Sampling)
- Pengurangan Noise (Noise Reduction)
- Scaling Data
- Penanganan Data Waktu (Time Series Data Handling)



Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



Apa Itu Statistika Deskriptif?

Statistika deskriptif menggambarkan, menunjukkan, sekaligus meringkas fitur dasar dari kumpulan data yang ditemukan dalam studi tertentu agar lebih mudah dipahami. Kemudian, fitur-fitur tersebut disajikan dalam sebuah ringkasan yang menggambarkan sampel data dan pengukurannya. Ini akan membantu analisis dalam memahami data dengan lebih baik.



Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



Ukuran Pemusatan Data

Central tendency (ukuran pemusatan data), biasanya kita lebih mengenal *central tendency* dengan Mean, Median, dan Modus.

Means

Jika Anda memiliki sampel sejumlah n nilai, x_i , maka rata-rata, μ , adalah hasil penjumlahan semua nilai tersebut dibagi dengan jumlah nilai tersebut; dengan kata lain, μ adalah jumlah dari semua nilai tersebut dibagi dengan n .

dapat dituliskan ke dalam persamaan berikut :

$$\mu = \frac{1}{N} \sum_i^N x_i \quad (4.1)$$



Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



Median merupakan datum yang terletak di tengah dari sebuah data. Syarat perhitungan median adalah diurutkan dari nilai terkecil hingga terbesar. Perhitungan median ini akan berbeda jika data berjumlah ganjil atau genap. Perhatikan bentuk berikut :

$$Q_2 = \begin{cases} X_{\frac{n+1}{2}}, & \text{jika } n \text{ genap} \\ \frac{1}{2} \left\{ X_{\frac{n}{2}} + X_{\frac{n+1}{2}} \right\}, & \text{jika } n \text{ ganjil} \end{cases} \quad (4.2)$$



Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



Varian

adalah salah satu ukuran statistik yang mengukur sejauh mana data dalam sebuah himpunan tersebar atau menyebar dari nilai rata-ratanya. Variansi menggambarkan tingkat ketidakhomogenan atau variasi dalam data. Semakin besar nilai variansi, semakin besar variasi atau penyebaran data tersebut, sedangkan nilai variansi yang lebih kecil menunjukkan bahwa data cenderung lebih mendekati nilai rata-ratanya.



Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



Varian dari suatu populasi dinotasikan dengan :

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (4.3)$$

Varian dari suatu sampel dinotasikan dengan :

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (4.4)$$

Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa
Sektor
Siklus Analisis Data
Langkah-Langkah Analisis
Data

Data Preprocessing

Pengantar Data
Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi
Data
Memilih Visualisasi Data
yang Tepat

Standar Deviasi

Standar deviasi adalah ukuran statistik yang mengukur sejauh mana data dalam suatu himpunan tersebar atau menyebar dari nilai rata-ratanya. Ini mengukur variabilitas atau deviasi data dari nilai tengah atau rata-rata. Standar deviasi digunakan untuk memahami sebaran data dan sejauh mana data berbeda dari nilai rata-ratanya.



Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data
Memilih Visualisasi Data yang Tepat



Standar Deviasi

Standar Deviasi Populasi (*Population Standard Deviation*):

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}} \quad (4.5)$$

Standar Deviasi Sampel (*Sample Standard Deviation*):

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (4.6)$$



Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data

Memilih Visualisasi Data yang Tepat



Pentingnya Visualisasi Data

Pentingnya Visualisasi Data dalam 4 Poin Utama:

- 1 **Menggambarkan Informasi:** Visualisasi data mengubah angka-angka dan data kompleks menjadi gambar yang lebih mudah dipahami, membantu kita melihat pola, tren, dan anomali dengan lebih jelas.



Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data

Memilih Visualisasi Data yang Tepat



Pentingnya Visualisasi Data

Pentingnya Visualisasi Data dalam 4 Poin Utama:

- 1** Menggambarkan Informasi: Visualisasi data mengubah angka-angka dan data kompleks menjadi gambar yang lebih mudah dipahami, membantu kita melihat pola, tren, dan anomali dengan lebih jelas.
- 2** Meningkatkan Pengambilan Keputusan: Memungkinkan pengambilan keputusan yang lebih baik dengan memberikan pandangan yang lebih lengkap tentang situasi atau masalah yang dihadapi.



Perkenalan

Kontrak Kuliah

Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data

Analisis Data di Beberapa Sektor

Siklus Analisis Data

Langkah-Langkah Analisis Data

Data

Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data

Memilih Visualisasi Data yang Tepat



Pentingnya Visualisasi Data

Pentingnya Visualisasi Data dalam 4 Poin Utama:

- 1** Menggambarkan Informasi: Visualisasi data mengubah angka-angka dan data kompleks menjadi gambar yang lebih mudah dipahami, membantu kita melihat pola, tren, dan anomali dengan lebih jelas.
- 2** Meningkatkan Pengambilan Keputusan: Memungkinkan pengambilan keputusan yang lebih baik dengan memberikan pandangan yang lebih lengkap tentang situasi atau masalah yang dihadapi.
- 3** Membantu Komunikasi: Mempermudah komunikasi informasi dan temuan kepada orang lain dengan cara yang jelas dan efektif.



Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

Visualisasi Data

Pentingnya Visualisasi Data

Memilih Visualisasi Data yang Tepat



Pentingnya Visualisasi Data

Pentingnya Visualisasi Data dalam 4 Poin Utama:

- 1** Menggambarkan Informasi: Visualisasi data mengubah angka-angka dan data kompleks menjadi gambar yang lebih mudah dipahami, membantu kita melihat pola, tren, dan anomali dengan lebih jelas.
- 2** Meningkatkan Pengambilan Keputusan: Memungkinkan pengambilan keputusan yang lebih baik dengan memberikan pandangan yang lebih lengkap tentang situasi atau masalah yang dihadapi.
- 3** Membantu Komunikasi: Mempermudah komunikasi informasi dan temuan kepada orang lain dengan cara yang jelas dan efektif.
- 4** Memfasilitasi Penjelajahan Data : Visualisasi data interaktif memungkinkan pengguna untuk berinteraksi dengan data, memperdalam pemahaman, dan menggali lebih dalam ke dalam informasi.



Memilih Visualisasi Data yang Tepat

Perkenalan

Kontrak Kuliah
Daftar Isi

Pengantar Analisis Data Eksploratif

Pengenalan Analisis Data
Analisis Data di Beberapa Sektor
Siklus Analisis Data
Langkah-Langkah Analisis Data

Data Preprocessing

Pengantar Data Preprocessing

Statistika Deskriptif

Ukuran Pemusatan Data

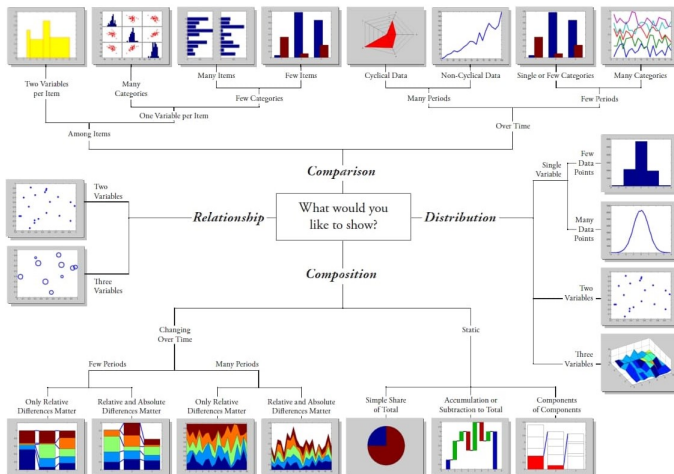
Visualisasi Data

Pentingnya Visualisasi Data

Memilih Visualisasi Data yang Tepat



Chart Suggestions—A Thought-Starter



Modified with permission -Doug Hull
blogs.mathworks.com/videos
hull@mathworks.com 2009
www.ExtremePresentat
© 2009 A. Abela — axcablagga

Thank You.

Egi Safitri, S.Mat., M.Si

