

Data Visualization

Presented by



Table of Contents

A.	Statistical, Scientific, Financial and Spatial Charts	03
B.	Gestalt Principles of Visual Perception	14
C.	Machine Learning Visualization (ML Insight & Storytelling)	20
D.	Mengukur kualitas Visualisasi	28

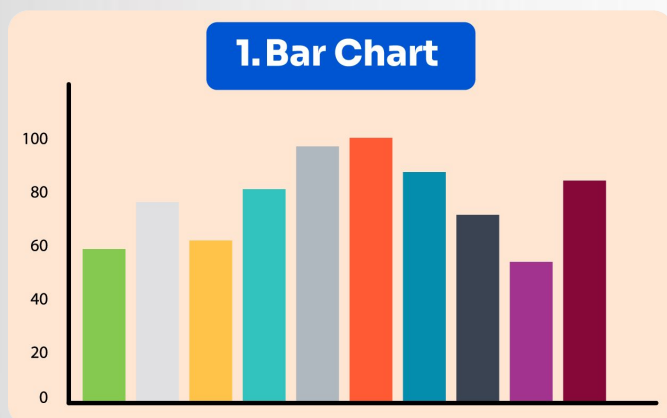
Statistical, Scientific, Financial and Spatial Charts

Di dalam dunia yang semakin terhubung dan canggih ini, data telah menjadi elemen kunci dalam pengambilan keputusan yang cerdas. Namun, seringkali data itu sendiri tidak dapat memberikan pemahaman yang cukup. Inilah mengapa kita membutuhkan visualisasi data.

Dalam pembelajaran ini, kita akan menjelajahi berbagai jenis grafik yang membantu kita memvisualisasikan data secara efektif. Kita akan mempelajari grafik statistik, ilmiah, keuangan, dan spasial. Setiap jenis grafik memiliki kegunaan dan kekuatan sendiri dalam membantu kita menganalisis dan menginterpretasikan data.

Statistical Chart

Pertama, mari kita melangkah ke dunia grafik statistik. Grafik ini memungkinkan kita untuk melihat pola dan tren dalam data.



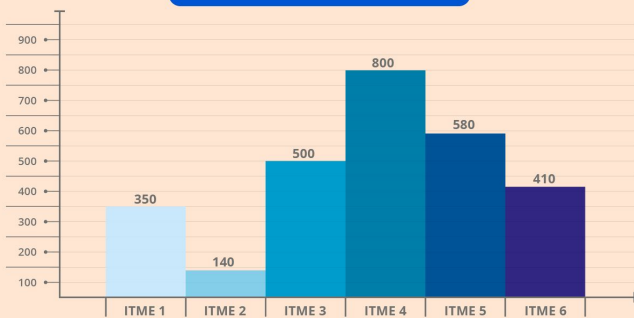
1. Bar Chart

Bar chart adalah jenis grafik yang menggunakan batang vertikal atau horizontal untuk membandingkan kategori atau variabel. Setiap batang mewakili nilai atau proporsi dari setiap kategori. Bar chart berguna untuk memvisualisasikan perbandingan antara beberapa kategori atau variabel.

Ia membantu mengidentifikasi perbedaan yang signifikan antara data dan melihat pola atau tren yang mungkin tersembunyi di dalamnya. Misalnya, kita dapat menggunakan bar chart untuk membandingkan penjualan produk pada beberapa bulan dalam setahun. Setiap batang akan mewakili penjualan produk pada bulan tersebut, dan tinggi batang akan menunjukkan jumlah penjualan.

Statistical, Scientific, Financial and Spatial Charts

2. Histogram

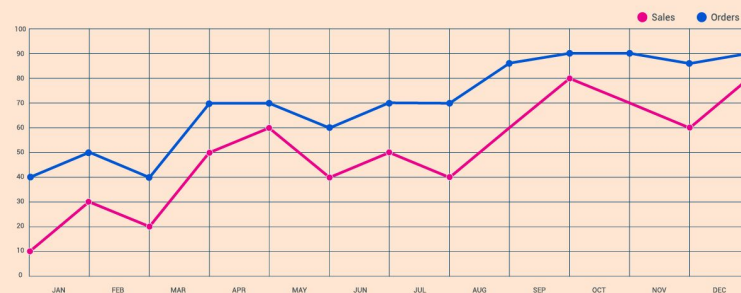


2. Histogram

Histogram adalah jenis grafik yang digunakan untuk menunjukkan distribusi frekuensi dari suatu variabel numerik. Rentang nilai dibagi menjadi interval atau kelas, dan sumbu horizontal mewakili kelas tersebut sedangkan sumbu vertikal mewakili frekuensi atau jumlah data di setiap kelas. Histogram membantu kita memahami sebaran data dan pola distribusi dari variabel numerik. Ia mem-

perlihatkan apakah data cenderung terkumpul dalam kelas tertentu atau terdistribusi secara merata di seluruh rentang nilai. Misalnya, histogram dapat digunakan untuk memvisualisasikan distribusi tinggi badan dari sekelompok orang. Interval tinggi badan akan dibagi, dan histogram akan menunjukkan jumlah orang di setiap interval tinggi badan.

3. Line Chart



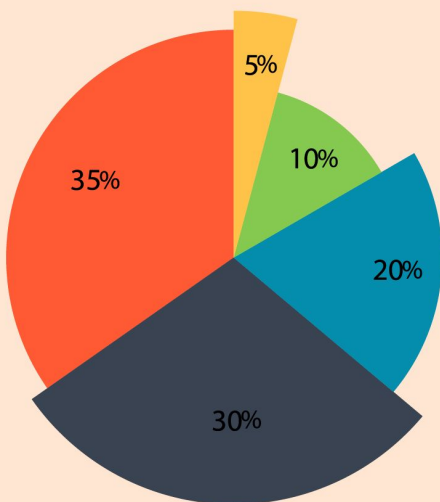
3. Line Chart

Line chart adalah jenis grafik yang menggunakan garis untuk menggambarkan perubahan data seiring waktu. Pada sumbu horizontal, biasanya terdapat skala waktu atau variabel independen, sedangkan sumbu vertikal menunjukkan nilai data.

Statistical, Scientific, Financial and Spatial Charts

Line chart membantu melacak dan memvisualisasikan tren atau perubahan data seiring waktu. Ia berguna untuk mengidentifikasi pola, fluktuasi, atau hubungan antara dua variabel yang berubah dalam periode waktu tertentu. Misalnya, line chart dapat digunakan untuk memperlihatkan perubahan harga saham suatu perusahaan selama satu tahun. Garis akan menghubungkan titik data harian dan memberikan gambaran visual tentang pergerakan harga saham seiring waktu.

4. Pie Chart



4. Pie Chart

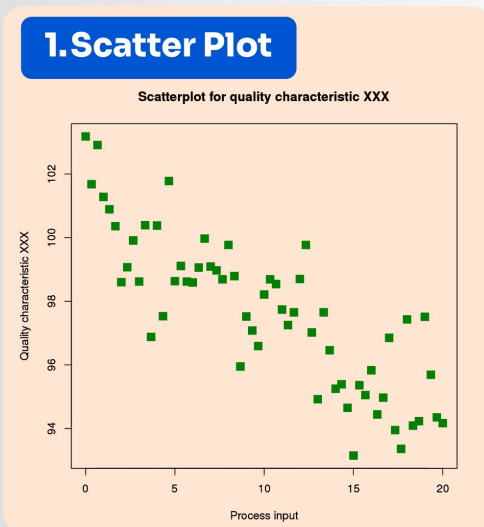
Pie chart adalah jenis grafik yang menggunakan sektor melingkar untuk menunjukkan proporsi relatif dari setiap kategori. Setiap sektor mewakili persentase atau bagian dari keseluruhan data. Pie chart berguna untuk memvisualisasikan komposisi atau proporsi relatif dari kategori yang berbeda. Ia membantu dengan cepat memahami persentase masing-masing kategori dalam satu kesatuan. Misalnya, pie chart dapat digunakan untuk menunjukkan persentase penjualan dari beberapa region dalam suatu perusahaan. Setiap sektor pie chart akan mewakili persentase penjualan suatu region.

Statistical, Scientific, Financial and Spatial Charts

Scientific Chart

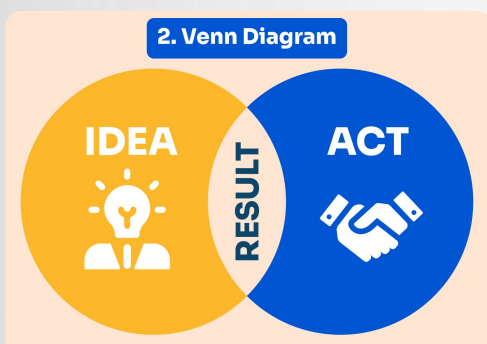
Selanjutnya, ada grafik ilmiah yang memungkinkan kita untuk memvisualisasikan data eksperimen dan penelitian.

1. Scatter Plot



Scatter plot adalah jenis grafik yang menggunakan titik-titik untuk menunjukkan hubungan antara dua variabel. Setiap titik pada grafik mewakili satu observasi dan ditempatkan pada koordinat yang sesuai dengan nilai dari dua variabel tersebut. Scatter plot digunakan untuk memvisualisasikan pola atau hubungan antara dua variabel. Ia membantu kita memahami korelasi, pola sebaran data, dan mengidentifikasi anomali atau outlier dalam data. Misalnya, scatter plot dapat digunakan untuk memperlihatkan hubungan antara usia dan tinggi badan dalam suatu kelompok populasi. Setiap titik akan mewakili satu individu, dengan sumbu horizontal mewakili usia dan sumbu vertikal mewakili tinggi badan.

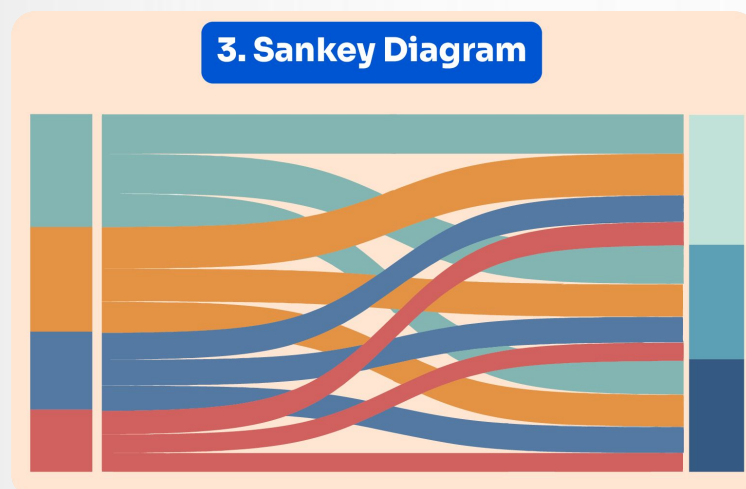
2. Venn Diagram



Venn diagram adalah jenis diagram yang digunakan untuk memvisualisasikan hubungan dan kesamaan antara beberapa himpunan data. Setiap himpunan direpresentasikan oleh lingkaran atau elips yang tumpang tindih, dan area tumpang tindih menunjukkan elemen yang sama di antara himpunan-himpunan tersebut.

Statistical, Scientific, Financial and Spatial Charts

Venn diagram membantu kita memahami persamaan dan perbedaan antara himpunan data yang berbeda. Ia berguna untuk menganalisis relasi logika, interseksi, atau relasi kompleks di antara kategori atau elemen yang berbeda. Misalnya, Venn diagram dapat digunakan untuk memvisualisasikan hubungan antara tiga kelompok orang berdasarkan minat mereka. Setiap lingkaran akan mewakili kelompok orang dengan elemen tumpang tindih menunjukkan minat yang sama di antara kelompok-kelompok tersebut.



3. Sankey Diagram

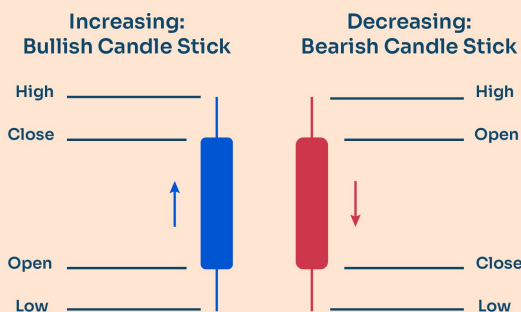
Sankey diagram adalah jenis diagram yang digunakan untuk memvisualisasikan aliran atau keterkaitan antara elemen yang berbeda. Diagram ini menggunakan garis yang berbentuk lebar yang mewakili jumlah atau proporsi aliran, dan lebar garis menunjukkan besarnya aliran antara elemen. Sankey diagram membantu kita memahami aliran atau keterkaitan kompleks antara elemen yang berbeda. Ia berguna untuk menggambarkan pola aliran data, proses, energi, atau sumber daya dalam suatu sistem. Misalnya, Sankey diagram dapat digunakan untuk memvisualisasikan aliran energi dalam suatu proses produksi. Garis yang lebar akan mewakili jumlah energi yang mengalir di antara berbagai komponen dalam sistem tersebut.

Statistical, Scientific, Financial and Spatial Charts

Financial Chart

Grafik keuangan yang digunakan untuk menganalisis data terkait pasar keuangan dan investasi.

1. Candlestick Chart

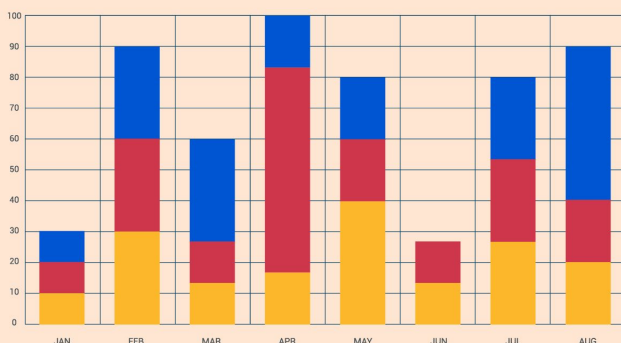


1. Candlestick Chart

Candlestick chart adalah jenis grafik yang digunakan dalam analisis pasar keuangan untuk memvisualisasikan pergerakan harga saham atau aset keuangan dalam periode waktu tertentu. Setiap "lilin" pada grafik mencerminkan rentang harga pembukaan, penutupan, tertinggi, dan terendah.

Candlestick chart membantu memvisualisasikan perubahan harga saham atau aset keuangan dalam jangka waktu tertentu. Ia memberikan informasi tentang sentimen pasar, tekanan pembeli dan penjual, serta pola dan tren harga. Misalnya, candlestick chart dapat digunakan untuk memvisualisasikan pergerakan harga saham perusahaan dalam satu minggu. Setiap lilin akan menunjukkan harga pembukaan, penutupan, tertinggi, dan terendah untuk setiap sesi perdagangan.

2. Stacked Bar Chart



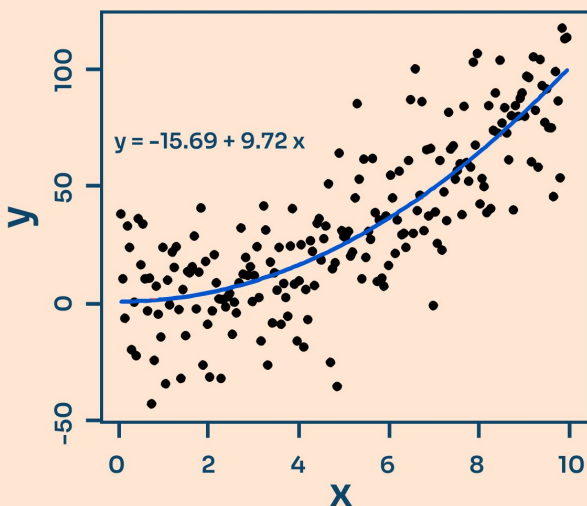
2. Stacked Bar Chart

Stacked bar chart adalah jenis grafik yang menggunakan batang bertumpuk untuk memvisualisasikan komposisi atau pembagian proporsi antara beberapa kategori atau sub-kategori.

Statistical, Scientific, Financial and Spatial Charts

Setiap batang mewakili jumlah total dari variabel yang ditumpuk di dalamnya. Stacked bar chart membantu memvisualisasikan kontribusi setiap kategori terhadap keseluruhan. Ia berguna untuk membandingkan proporsi antara kategori-kategori tersebut dan melihat perubahan komposisi dari waktu ke waktu. Misalnya, stacked bar chart dapat digunakan untuk memvisualisasikan komposisi penjualan produk berdasarkan kategori seperti makanan, minuman, dan aksesoris. Setiap batang akan menunjukkan total penjualan dan tumpukan warna di dalamnya akan mewakili kontribusi setiap kategori.

3. Line & Scatter Plot



3. Line & Scatter Plot

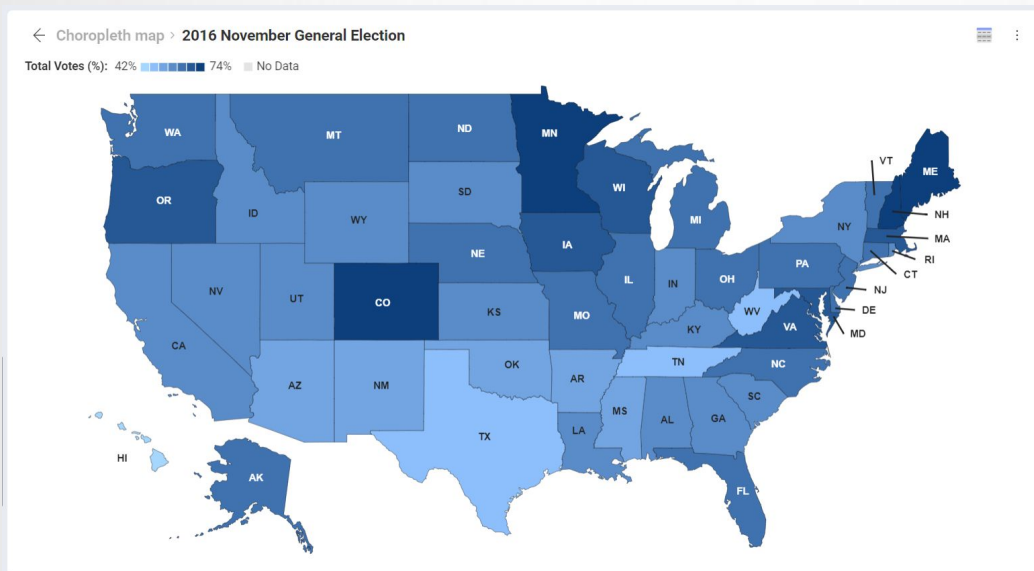
Line and scatter plot adalah kombinasi dari grafik garis dan grafik scatter yang menggabungkan titik-titik data dengan garis yang menghubungkannya. Pada grafik ini, titik-titik data mewakili observasi individu, sedangkan garis menghubungkan titik-titik tersebut untuk memperlihatkan tren atau pola yang mungkin ada. Line and scatter plot membantu memvisualisasikan hubungan antara dua variabel dan melacak perubahan data seiring waktu.

Ia berguna untuk memahami pola, korelasi, dan fluktuasi data. Misalnya, line and scatter plot dapat digunakan untuk memperlihatkan hubungan antara suhu dan penjualan es krim dalam satu tahun. Titik-titik data akan mewakili observasi penjualan dan garis yang menghubungkannya akan menunjukkan tren penjualan seiring perubahan suhu.

Statistical, Scientific, Financial and Spatial Charts

Spatial Chart

Grafik spasial yang memvisualisasikan data di peta atau ruang geografis.



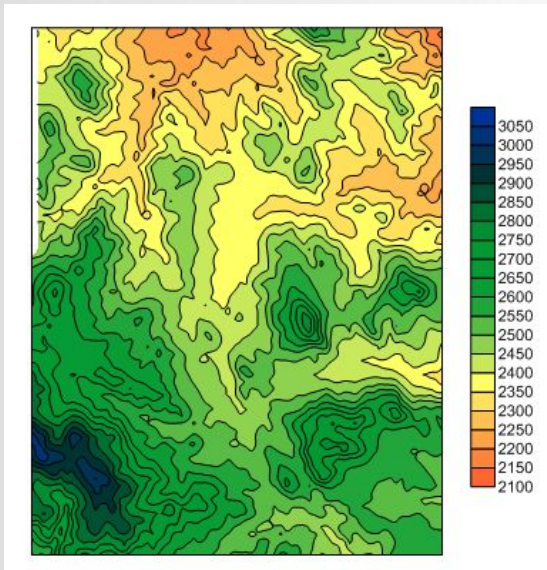
1. Choropleth Map

Choropleth map adalah jenis peta yang menggunakan warna atau pola yang berbeda pada wilayah geografis untuk menunjukkan data statistik atau properti yang berbeda. Setiap wilayah dalam peta mewakili kategori atau variabel tertentu, dan warna atau pola yang diterapkan pada wilayah tersebut mencerminkan nilai atau proporsi dari variabel tersebut. Choropleth map membantu memvisualisasikan pola spasial dari data statistik pada tingkat wilayah. Ia memungkinkan pemahaman yang cepat tentang perbedaan regional dan variabilitas data pada skala geografis. Misalnya, choropleth map dapat digunakan untuk memvisualisasikan tingkat pengangguran di berbagai negara bagian dalam suatu negara. Setiap negara bagian akan mewakili tingkat pengangguran dengan menggunakan warna yang berbeda, di mana warna lebih terang menunjukkan tingkat pengangguran yang lebih tinggi.

Statistical, Scientific, Financial and Spatial Charts

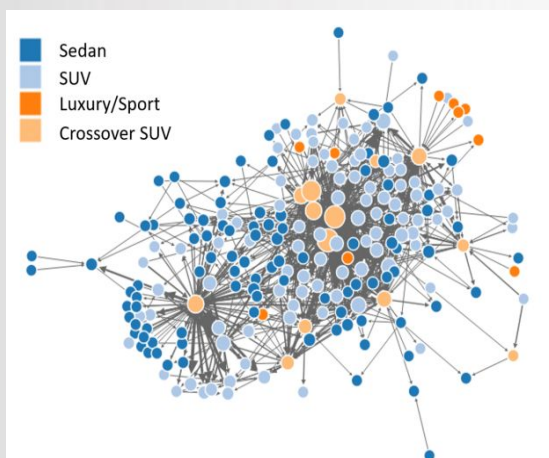
2. Contour Map

Contour map adalah jenis peta yang digunakan untuk memvisualisasikan perbedaan ketinggian atau permukaan bumi dalam bentuk garis kontur. Garis kontur menghubungkan titik-titik dengan ketinggian yang sama di permukaan bumi, membentuk pola garis yang menunjukkan perubahan ketinggian yang halus atau tajam.



Contour map membantu memvisualisasikan bentuk topografi dan perbedaan ketinggian suatu daerah. Ia berguna untuk memahami struktur lahan, mengidentifikasi cekungan atau puncak, serta merencanakan rute perjalanan di daerah yang bergunung atau berbukit. Misalnya, contour map dapat digunakan untuk memvisualisasikan topografi pegunungan dengan garis kontur yang menggambarkan perubahan ketinggian di wilayah tersebut.

3. Network Graph



Network graph, juga dikenal sebagai grafik jaringan, adalah jenis grafik yang digunakan untuk memvisualisasikan hubungan kompleks antara entitas atau elemen yang berbeda. Grafik ini terdiri dari simpul (node) yang mewakili entitas dan tepi (edge) yang menghubungkan simpul-simpul tersebut untuk menunjukkan relasi antara entitas tersebut. Network graph membantu memvisualisasikan interaksi dan hubungan antara entitas dalam suatu sistem atau jaringan.

Statistical, Scientific, Financial and Spatial Charts

Ia berguna untuk memahami struktur jaringan, mengidentifikasi simpul pusat atau penghubung yang penting, serta menganalisis aliran informasi atau interaksi antara entitas. Misalnya, network graph dapat digunakan untuk memvisualisasikan interaksi sosial dalam media sosial, di mana setiap akun pengguna akan menjadi simpul dan tepi akan menghubungkan pengguna yang saling berinteraksi.

Metode Pemilihan Grafik

Setiap kali Anda berencana untuk membuat visualisasi data, ada beberapa praktik terbaik yang dapat Anda gunakan agar visualisasi tersebut lebih mudah dipahami dan efektif. Berikut adalah panduan yang lebih detail:

- Jika Anda memiliki data kategorikal, pertimbangkan penggunaan Grafik batang jika terdapat lebih dari 5 kategori. Namun, jika jumlah kategori terbatas, Anda dapat menggunakan Grafik lingkaran untuk memvisualisasikan proporsi setiap kategori.
- Jika Anda memiliki data nominal, terdapat beberapa pilihan. Jika data Anda terdiri dari observasi terpisah, Anda dapat menggunakan bagan batang atau histogram. Namun, jika data Anda bersifat kontinu, lebih tepat menggunakan bagan garis atau bagan area.
- Jika tujuan Anda adalah untuk menunjukkan hubungan antara nilai dalam suatu set data, Anda dapat mempertimbangkan penggunaan Grafik sebar, bagan gelembung, atau bagan garis. Grafik sebar membantu memvisualisasikan korelasi antara dua variabel, sedangkan bagan gelembung memberikan informasi tambahan dengan ukuran titik yang merepresentasikan variabel ketiga.

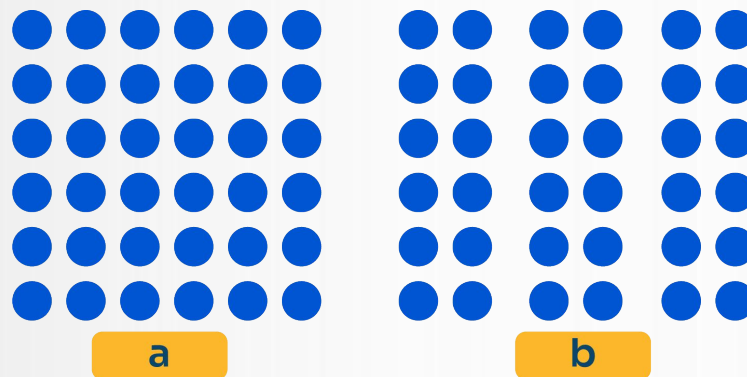
Statistical, Scientific, Financial and Spatial Charts

- Jika Anda ingin membandingkan nilai, Grafik lingkaran dapat digunakan untuk memberikan perbandingan relatif antara kategori atau sub-kategori. Namun, jika Anda memerlukan perbandingan yang lebih tepat, Grafik batang lebih disarankan.
- Jika fokus Anda adalah pada perbandingan volume antara kategori atau sub-kategori, Anda dapat mempertimbangkan penggunaan bagan area atau bagan gelembung. Bagan area memvisualisasikan volume dengan menggunakan area di bawah garis, sedangkan bagan gelembung menggunakan ukuran titik untuk menggambarkan besarnya nilai.
- Jika Anda ingin memperlihatkan tren dan pola dalam data Anda, grafik yang direkomendasikan adalah bagan garis, bagan batang, atau bagan sebar. Bagan garis membantu memvisualisasikan perubahan nilai seiring waktu, sedangkan bagan batang dan bagan sebar memperlihatkan pola sebaran dan variasi data.

Dengan memilih grafik yang sesuai sesuai dengan jenis data yang Anda miliki dan tujuan visualisasi Anda, Anda dapat mengkomunikasikan informasi dengan lebih jelas dan efektif kepada audiens Anda.

Gestalt Principles of Visual Perception

Prinsip Gestalt adalah prinsip/hukum persepsi manusia yang menggambarkan bagaimana manusia mengelompokkan unsur-unsur yang serupa, mengenali pola dan menyederhanakan gambar kompleks ketika kita melihat objek. Desainer menggunakan prinsip ini untuk mengatur konten di situs web dan antarmuka lain sehingga menyenangkan secara estetika dan mudah dipahami. Prinsip-prinsip ini membantu kita memahami bagaimana kita mempersepsi dan menginterpretasikan dunia visual di sekitar kita. Dalam materi ini, kita akan menjelaskan tujuh prinsip Gestalt yang utama beserta contoh-contohnya.



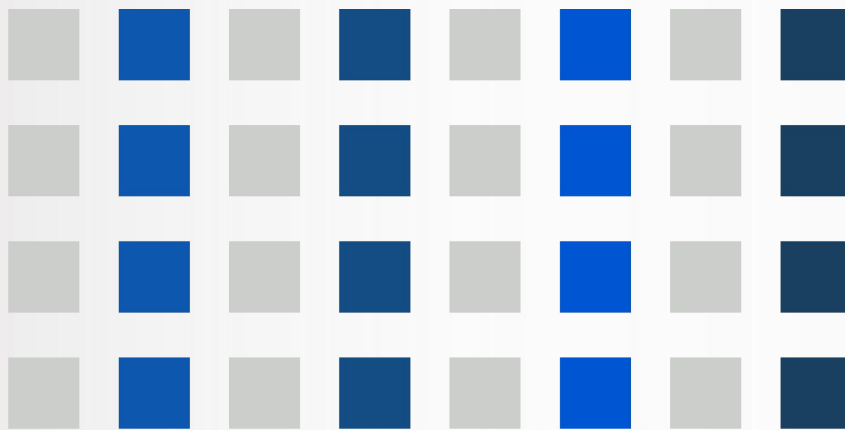
1. Prinsip Kedekatan (Principle of Proximity)

Prinsip kedekatan menyatakan bahwa kita cenderung mengelompokkan objek yang berdekatan spasial menjadi satu kelompok atau kesatuan. Ketika objek-objek berada dalam jarak yang relatif dekat satu sama lain, kita cenderung mempersepsikan mereka sebagai satu kelompok yang terkait. Contoh-contoh prinsip kedekatan termasuk:

- Mengelompokkan titik-titik yang berdekatan menjadi satu pola atau garis.
Contoh: Jika kita melihat sekelompok titik yang tersebar secara acak, tetapi beberapa di antaranya berdekatan satu sama lain, kita cenderung mengelompokkan titik-titik tersebut menjadi garis atau pola yang berkelanjutan.
- Mengelompokkan kata-kata yang berdekatan dalam satu kalimat atau paragraf.
Contoh: Ketika membaca sebuah teks, kita secara otomatis mengelompokkan kata-kata yang berdekatan menjadi satu unit makna yang utuh.

Gestalt Principles of Visual Perception

2. Prinsip Kesamaan (Principle of Similarity)

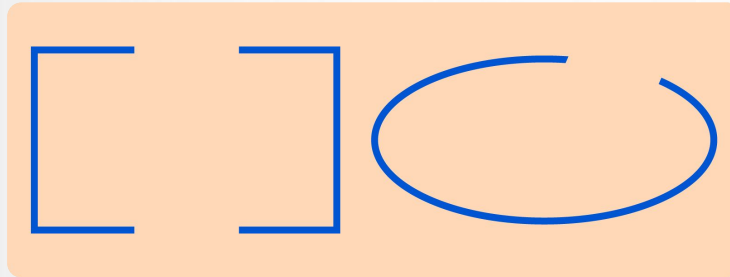


Prinsip kesamaan menyatakan bahwa kita cenderung mengelompokkan objek yang memiliki kesamaan dalam bentuk, ukuran, warna, atau tekstur menjadi satu kesatuan. Ketika objek-objek memiliki atribut yang mirip, kita cenderung mempersepsikan mereka sebagai bagian dari satu kelompok atau kategori yang sama. Contoh-contoh prinsip kesamaan termasuk:

- Mengelompokkan benda-benda dengan warna yang sama menjadi satu kelompok. Contoh: Jika ada sekelompok bola dengan warna yang berbeda, tetapi beberapa di antaranya memiliki warna yang sama, kita cenderung mengelompokkan bola-bola tersebut berdasarkan warna mereka.
- Mengelompokkan benda-benda dengan bentuk yang sama menjadi satu kelompok. Contoh: Jika ada sekelompok benda geometris, seperti segitiga dan lingkaran, kita cenderung mengelompokkan segitiga-segitiga dan lingkaran-lingkaran menjadi dua kelompok yang berbeda berdasarkan bentuk mereka.

Gestalt Principles of Visual Perception

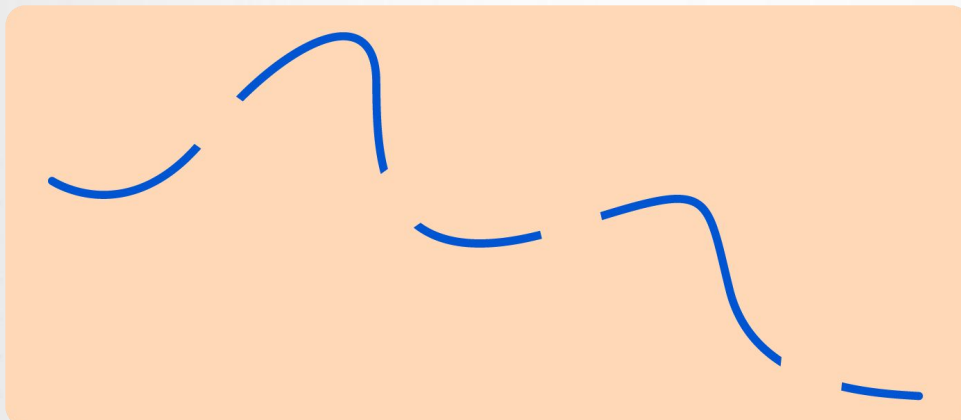
3. Prinsip Kesimpangan (Principle of Closure)



Prinsip kesimpangan menyatakan bahwa kita cenderung melihat bentuk yang tidak lengkap atau terputus sebagai bentuk yang utuh dan lengkap. Kita memiliki kecenderungan untuk melengkapi celah atau bagian yang hilang dalam suatu pola untuk membentuk suatu gambaran yang lebih terorganisir. Contoh-contoh prinsip kesimpangan termasuk:

- Melihat lingkaran yang terputus sebagai lingkaran utuh. Contoh: Jika ada sebuah lingkaran yang terputus di beberapa titik, kita cenderung melihatnya sebagai lingkaran lengkap meskipun ada celah atau bagian yang hilang.
- Melihat huruf "C" yang terbuka sebagai huruf "O" yang lengkap. Contoh: Saat melihat huruf "C" yang hanya memiliki setengah lingkaran, kita secara otomatis melengkapi bagian yang hilang dan melihatnya sebagai huruf "O" yang utuh.

4. Prinsip Kesenambungan (Principle of Continuity)

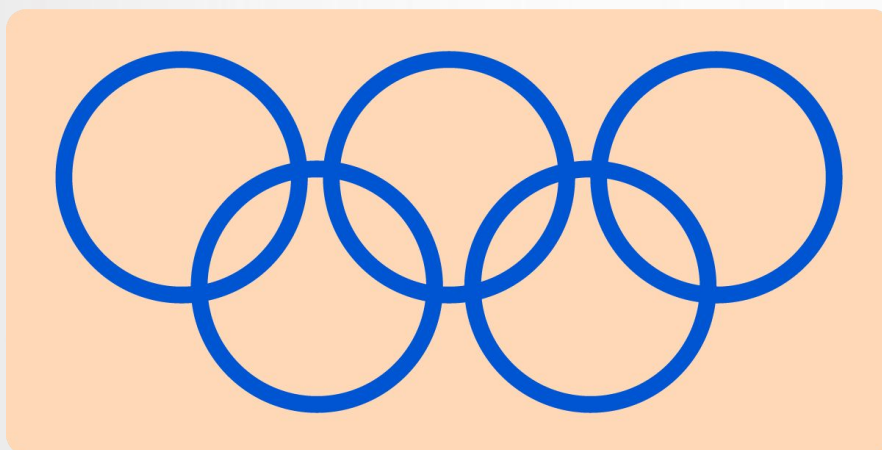


Gestalt Principles of Visual Perception

Prinsip kesinambungan mengatakan bahwa kita cenderung melihat garis yang lancar dan berkelanjutan daripada garis yang terputus-putus. Pikiran kita secara alami mengisi celah atau jeda dalam suatu pola, sehingga menciptakan kelanjutan yang tidak nyata. Contoh-contoh prinsip kesinambungan termasuk:

- Melihat garis yang lancar dan berkelanjutan meskipun ada objek yang memotongnya. Contoh: Ketika kita melihat garis yang terpotong oleh sebuah objek, kita masih melihat garis tersebut sebagai satu garis yang berkelanjutan, meskipun ada bagian yang tersembunyi oleh objek tersebut.
- Melihat garis yang dibengkokkan atau melengkung sebagai garis yang tidak terputus. Contoh: Jika ada sebuah garis yang mengalami lengkungan atau melingkar, kita cenderung melihatnya sebagai garis yang terus berkelanjutan tanpa ada putus-putus di dalamnya.

5. Prinsip Simetri (Principle of Symmetry)

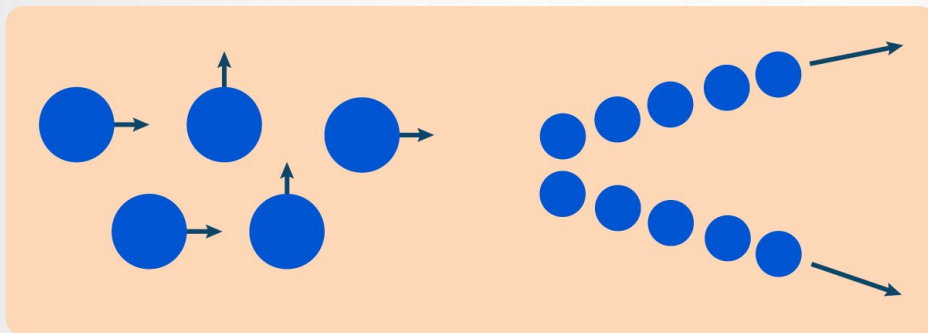


Prinsip simetri menyatakan bahwa kita cenderung mempersepsi objek sebagai simetris atau seimbang berdasarkan sumbu sentral. Kita cenderung melihat kesamaan dalam bentuk, ukuran, atau posisi objek yang berada di sebelah kanan dan sebelah kiri sumbu sentral. Contoh-contoh prinsip simetri termasuk:

Gestalt Principles of Visual Perception

- Melihat wajah manusia sebagai simetris. Contoh: Ketika melihat wajah manusia, kita cenderung melihat kesamaan dalam bentuk mata, hidung, dan bibir yang berada di kedua sisi sumbu sentral wajah.
- Melihat objek seperti bunga dengan kelopak simetris. Contoh: Saat melihat bunga dengan kelopak yang simetris, kita cenderung melihat kesamaan bentuk dan pola pada setiap sisi bunga.

6. Prinsip Nasib Bersama (Principle of Common Fate)

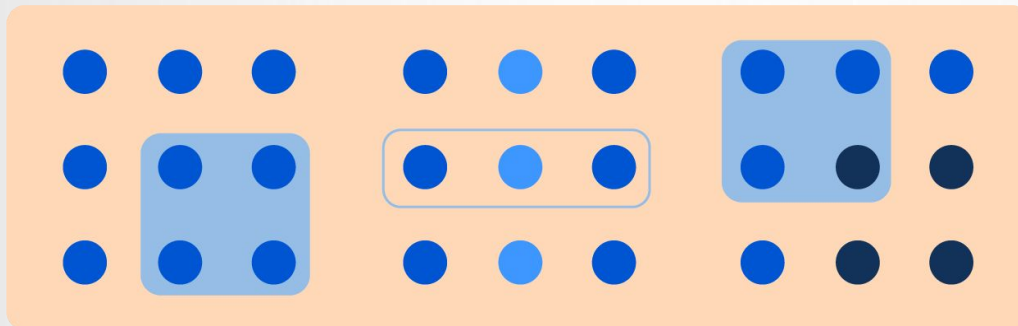


Prinsip nasib bersama menyatakan bahwa kita cenderung mengelompokkan objek yang memiliki arah atau gerakan yang serupa menjadi satu kelompok. Kita mempersepsi objek-objek tersebut sebagai bagian dari satu entitas yang sama karena mereka bergerak atau berubah dengan cara yang serupa. Contoh-contoh prinsip nasib bersama termasuk:

- Melihat sekelompok burung yang terbang dalam formasi yang seragam. Contoh: Ketika melihat sekelompok burung yang terbang bersama dan mengubah formasi secara seragam, kita cenderung melihatnya sebagai satu kelompok burung yang bergerak bersama-sama.
- Melihat sekelompok orang yang berjalan dalam arah yang sama. Contoh: Saat melihat sekelompok orang yang berjalan dalam arah yang sama di jalan, kita cenderung mengelompokkan mereka sebagai satu kelompok yang memiliki nasib bersama.

Gestalt Principles of Visual Perception

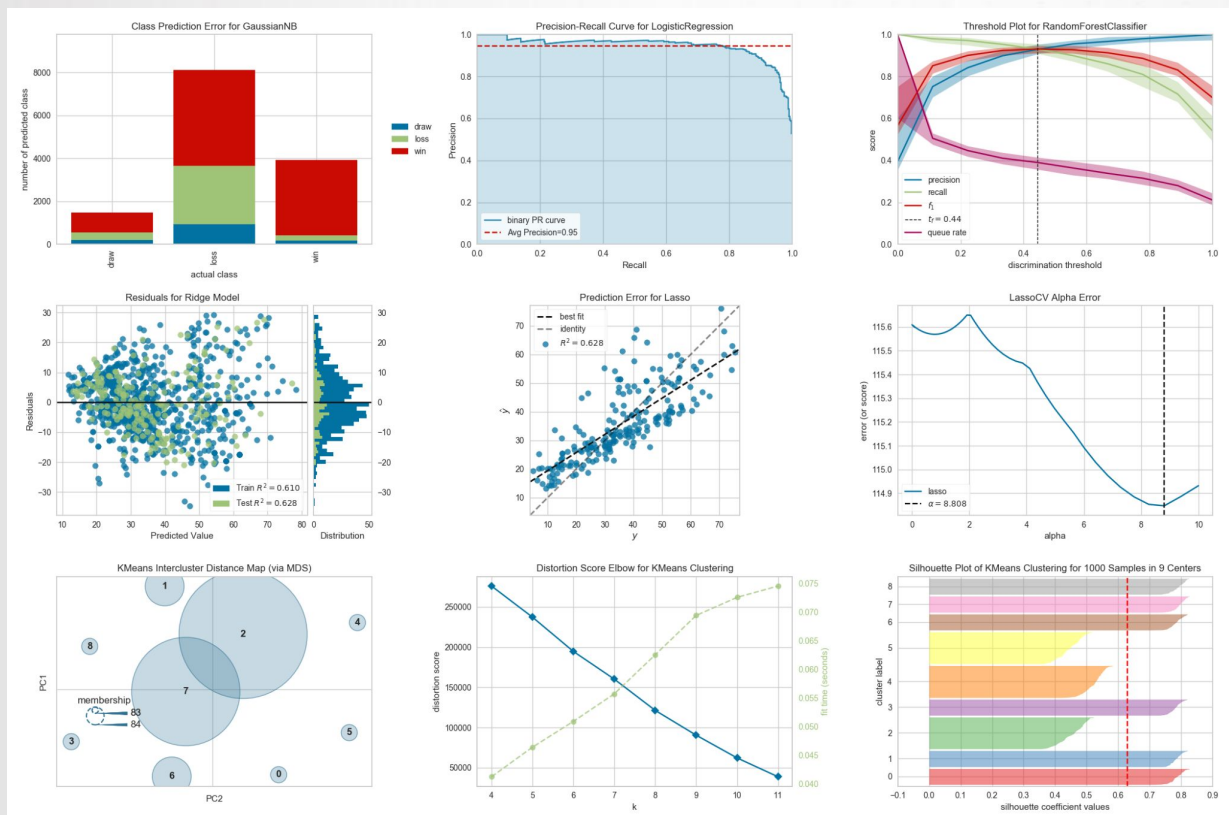
7. Prinsip Wilayah Bersama (Common Region)



Prinsip Wilayah Bersama menyatakan bahwa kita cenderung mengelompokkan objek-objek yang berada dalam wilayah yang sama menjadi satu kelompok. Kita mempersepsi objek-objek tersebut sebagai entitas yang terpisah dari objek-objek di luar wilayah tersebut. Prinsip ini membantu kita mengenali batasan visual dan mengelompokkan objek-objek berdasarkan wilayah yang mereka bagi. Contoh-contoh prinsip Wilayah Bersama antara lain:

- Mengelompokkan objek dalam suatu bingkai atau kotak. Contoh: Jika ada beberapa objek, seperti gambar, yang diletakkan dalam bingkai atau kotak yang sama, kita cenderung mempersepsi objek-objek tersebut sebagai kelompok yang terpisah dari objek-objek di luar bingkai tersebut.
- Mengelompokkan teks yang ditempatkan dalam kolom atau area yang terpisah. Contoh: Saat membaca artikel yang memiliki kolom teks yang terpisah, kita cenderung mempersepsi setiap kolom sebagai satu wilayah yang berbeda dan memproses teks di dalam kolom tersebut secara terpisah dari teks di kolom lainnya.
- Mengelompokkan objek-objek dalam diagram atau grafik yang berbeda. Contoh: Ketika melihat sebuah diagram yang terdiri dari beberapa area atau bagian yang berbeda, kita cenderung mengelompokkan objek-objek atau data yang berada di dalam setiap bagian tersebut menjadi satu kelompok yang terpisah.

Machine Learning Visualization (ML Insight & Storytelling)



Dalam Machine Learning, Exploratory Data Analysis (EDA) adalah tahap penting dalam pemrosesan data yang bertujuan untuk memahami dan menganalisis data secara visual sebelum membangun model. Visualisasi data memainkan peran krusial dalam EDA karena membantu kita memahami pola, tren, keterkaitan, dan anomali yang terdapat dalam dataset. Dalam materi ini, kita akan menjelaskan pentingnya visualisasi dalam EDA dan beberapa teknik visualisasi yang umum digunakan.

1. Pentingnya Visualisasi dalam EDA

Visualisasi dalam EDA memiliki beberapa manfaat penting, antara lain:

- Memahami struktur data: Melalui visualisasi, kita dapat melihat distribusi, tipe data, dan atribut-atribut yang ada dalam dataset. Hal ini membantu kita memahami karakteristik data dan memutuskan langkah-langkah selanjutnya.

Machine Learning Visualization (ML Insight & Storytelling)

- Mengidentifikasi pola dan tren: Dengan menggambarkan data dalam bentuk grafik atau plot, kita dapat mengenali pola, tren, dan keterkaitan antara variabel-variabel dalam dataset. Hal ini membantu kita membuat hipotesis dan memperoleh wawasan yang lebih dalam tentang data.
- Mendeteksi anomali dan outliers: Visualisasi membantu kita mengidentifikasi nilai-nilai yang tidak biasa atau jauh dari pola umum dalam dataset. Dengan menyoroti outliers, kita dapat melakukan analisis lebih lanjut untuk memahami penyebabnya atau memutuskan tindakan yang tepat.

Exploratory Data Analysis (EDA) adalah tahap awal dalam proses Machine Learning yang bertujuan untuk memahami data yang akan digunakan untuk membangun model. Visualisasi memainkan peran krusial dalam EDA karena membantu kita menjelajahi, menganalisis, dan memahami data secara visual. Dalam materi ini, kita akan menjelaskan langkah-langkah penting dalam EDA dan bagaimana visualisasi dapat digunakan untuk memperoleh wawasan tentang data yang akan digunakan.

1. Memahami Struktur Data

Pertama-tama, kita perlu memahami struktur data yang ada dalam dataset. Beberapa langkah yang dapat dilakukan meliputi:

- Memeriksa dimensi dataset (jumlah baris dan kolom) untuk memahami seberapa besar dataset yang kita miliki.
- Menampilkan beberapa baris pertama dari dataset untuk mendapatkan gambaran awal tentang atribut-atribut yang ada.
- Melihat tipe data dari setiap kolom untuk memastikan pemahaman yang benar tentang variabel-variabel yang ada.

Machine Learning Visualization (ML Insight & Storytelling)

2. Mengidentifikasi dan Membersihkan Missing Values

Missing values atau nilai yang hilang adalah nilai yang tidak ada dalam dataset. Untuk memahami sejauh mana dataset memiliki missing values, kita dapat melakukan hal berikut:

- Menghitung jumlah missing values dalam setiap kolom dan memvisualisasikannya dalam bentuk grafik seperti bar plot atau heatmap.
- Menentukan apakah missing values bersifat acak atau terdapat pola tertentu di antara variabel-variabel dalam dataset.
- Mengambil keputusan tentang bagaimana menangani missing values, seperti menghapus baris atau mengisi nilai yang hilang dengan metode seperti mean atau median.

3. Analisis Statistik Deskriptif

Statistik deskriptif memberikan ringkasan tentang atribut-atribut dalam dataset. Beberapa teknik visualisasi yang dapat digunakan meliputi:

- Histogram untuk melihat sebaran nilai pada variabel numerik.
- Box plot untuk memahami distribusi, kuartil, rentang, dan outliers dalam variabel.
- Bar chart atau pie chart untuk memvisualisasikan sebaran kategori pada variabel kualitatif.

4. Menganalisis Korelasi

Korelasi menggambarkan hubungan antara variabel-variabel dalam dataset. Beberapa visualisasi yang berguna untuk analisis korelasi meliputi:

- Scatter plot untuk melihat hubungan antara dua variabel numerik.
- Heatmap untuk memvisualisasikan matriks korelasi antara variabel-variabel dalam bentuk warna.

Machine Learning Visualization (ML Insight & Storytelling)

5. Mengeksplorasi Pola dan Tren

Visualisasi dapat membantu kita menemukan pola dan tren dalam data. Beberapa teknik yang berguna meliputi:

- Line chart untuk memvisualisasikan perubahan variabel numerik seiring waktu.
- Time series plot untuk melihat pola seiring waktu pada data deret waktu.
- Area chart atau stacked bar chart untuk membandingkan kontribusi relatif dari variabel-variabel pada setiap periode waktu.

Data Storytelling and Actionable Insight

Data storytelling adalah seni mengkomunikasikan informasi yang terkandung dalam data melalui visualisasi yang kuat dan narasi yang efektif. Tujuan dari data storytelling adalah untuk menginspirasi tindakan dan pengambilan keputusan yang didasarkan pada wawasan yang diperoleh dari data. Dalam materi ini, kita akan menjelaskan pentingnya data storytelling dan bagaimana visualisasi dalam Machine Learning dapat digunakan untuk menghasilkan wawasan yang dapat diimplementasikan (actionable insights).

1. Pentingnya Data Storytelling

Data storytelling memiliki beberapa manfaat penting, antara lain:

- Menggabungkan fakta dan narasi: Data storytelling menggabungkan fakta-fakta yang didapatkan dari analisis data dengan narasi yang menarik. Ini membantu audiens memahami konteks dan relevansi informasi yang disajikan.
- Membantu pengambilan keputusan: Melalui data storytelling, kita dapat mengilustrasikan temuan-temuan penting dari analisis data yang dapat digunakan sebagai dasar pengambilan keputusan yang lebih baik.

Machine Learning Visualization (ML Insight & Storytelling)

- **Memotivasi tindakan:** Dengan menggabungkan wawasan dari visualisasi data dengan narasi yang menginspirasi, data storytelling mendorong orang untuk mengambil tindakan yang relevan berdasarkan informasi yang disajikan.

2. Menghasilkan Actionable Insights

Mengidentifikasi insight yang bermanfaat dari analisis data adalah langkah penting dalam menghasilkan actionable insights. Berikut adalah beberapa langkah yang dapat membantu dalam menghasilkan actionable insights melalui analisis data:

- **Membatasi fokus:** Tentukan tujuan dan pertanyaan bisnis yang ingin dijawab melalui analisis data. Ini membantu dalam memfokuskan upaya pada temuan yang paling relevan dan actionable.
- **Gunakan pemahaman domain:** Memiliki pemahaman yang baik tentang domain bisnis atau industri yang sedang dianalisis membantu dalam mengidentifikasi insight yang memiliki dampak nyata dan dapat diimplementasikan.
- **Pertimbangkan konteks:** Memahami konteks yang melingkupi data sangat penting untuk menghasilkan insight yang relevan. Pertimbangkan faktor eksternal, tren industri, atau perubahan lingkungan yang dapat mempengaruhi interpretasi dan tindakan yang diperlukan.
- **Menganalisis pola dan tren:** Identifikasi pola atau tren yang signifikan dalam data yang dapat memberikan wawasan berharga. Misalnya, apakah ada tren kenaikan penjualan pada saat-saat tertentu atau pola musiman yang dapat dimanfaatkan untuk pengambilan keputusan.

Machine Learning Visualization (ML Insight & Storytelling)

- **Identifikasi penyebab:** Melalui analisis data yang mendalam, upayakan untuk mengidentifikasi penyebab atau faktor yang berkontribusi terhadap temuan yang dihasilkan. Ini membantu dalam memahami akar permasalahan dan menyusun tindakan yang tepat.
- **Lakukan segmentasi:** Jika memungkinkan, lakukan segmentasi data untuk mengidentifikasi kelompok atau sub-kelompok yang memiliki karakteristik atau perilaku yang serupa. Ini dapat membantu dalam memahami preferensi, kebutuhan, atau kecenderungan yang berbeda antara segmen tersebut.
- **Gunakan metrik yang relevan:** Gunakan metrik yang sesuai dengan tujuan bisnis dan analisis yang sedang dilakukan. Misalnya, jika fokusnya adalah pada retensi pelanggan, gunakan metrik seperti churn rate atau customer lifetime value.
- **Validasi hasil:** Validasi hasil analisis data dan temuan yang dihasilkan dengan menguji hipotesis atau membandingkannya dengan informasi atau pengetahuan lain yang ada. Ini membantu memastikan kebenaran dan keandalan dari actionable insights yang dihasilkan.
- **Sampaikan temuan dengan jelas:** Saat menyajikan actionable insights, pastikan informasi yang disampaikan mudah dipahami dan jelas. Gunakan visualisasi data yang efektif, narasi yang kuat, dan komunikasikan temuan secara singkat dan langsung.
- **Jelaskan implikasi dan tindakan yang dapat diambil:** Sertakan implikasi dari insight yang dihasilkan dan rekomendasi tindakan yang dapat diambil berdasarkan temuan tersebut. Pastikan rekomendasi tersebut praktis, relevan, dan dapat diimplementasikan.
- Dengan mengikuti langkah-langkah ini, kita dapat menghasilkan actionable insights yang dapat digunakan untuk mendukung pengambilan keputusan yang lebih baik dan mencapai tujuan bisnis yang diinginkan.

Machine Learning Visualization (ML Insight & Storytelling)

4. Visualisasi Actionable Insights

Setelah mengidentifikasi temuan kunci, langkah selanjutnya adalah mengkomunikasikan actionable insights melalui visualisasi data yang efektif. Beberapa tips yang dapat membantu dalam proses ini adalah:

- Gunakan visualisasi yang relevan: Pilih jenis visualisasi yang dapat memperkuat pesan dan insight yang ingin disampaikan. Misalnya, jika insight yang ingin disampaikan adalah perbandingan antara beberapa kategori, bar chart atau stacked bar chart dapat digunakan.
- Beri penekanan pada perbedaan yang signifikan: Gunakan elemen visual seperti warna atau ukuran untuk mempertegas perbedaan yang signifikan antara data. Ini membantu dalam menyoroti insight yang paling relevan.
- Gunakan anotasi dan teks yang jelas: Sertakan anotasi dan teks untuk memberikan konteks dan penjelasan tambahan pada visualisasi. Ini membantu audiens memahami temuan secara lebih mendalam dan menghubungkannya dengan tindakan yang dapat diambil.
- Gunakan multiple visualizations: Jika diperlukan, gunakan beberapa visualisasi yang saling melengkapi untuk menggambarkan berbagai aspek insight yang ingin disampaikan. Misalnya, kombinasikan line chart dengan bar chart untuk menunjukkan tren dan perbandingan kategori secara bersamaan.
- Sederhanakan dan fokuskan pesan: Pastikan visualisasi yang dibuat tidak terlalu rumit atau terlalu banyak informasi yang membingungkan. Sederhanakan pesan yang ingin disampaikan sehingga audiens dapat dengan mudah memahaminya.

Machine Learning Visualization (ML Insight & Storytelling)

5. Memaparkan Insight dan Mengajukan Tindakan

Langkah terakhir dalam data storytelling adalah memaparkan insight yang ditemukan dan mengajukan tindakan yang dapat diambil berdasarkan temuan tersebut. Beberapa tips yang dapat membantu dalam proses ini adalah:

- Gunakan narasi yang kuat: Buatlah narasi yang mengaitkan temuan dengan tujuan bisnis atau permasalahan yang ada. Jelaskan mengapa temuan tersebut penting dan apa implikasinya bagi pengambilan keputusan.
- Sertakan rekomendasi tindakan: Berikan rekomendasi yang jelas mengenai tindakan yang dapat diambil berdasarkan insight yang ditemukan. Pastikan rekomendasi tersebut praktis dan dapat diimplementasikan.
- Berikan konteks yang relevan: Sertakan konteks yang relevan seperti informasi tentang target pasar, persaingan, atau tren industri yang dapat membantu audiens memahami implikasi insight secara lebih baik.
- Ajukan pertanyaan tambahan: Dorong audiens untuk berpikir lebih jauh dengan mengajukan pertanyaan tambahan yang dapat mengarah pada penelitian atau analisis lebih lanjut. Ini mendorong diskusi dan pemikiran kritis yang lebih dalam.

Mengukur kualitas Visualisasi

Dalam dunia yang semakin terhubung secara digital, visualisasi data telah menjadi alat yang penting dalam memahami dan menyampaikan informasi. Namun, dalam membaca visualisasi data, penting untuk mempertahankan sikap kritis dan berhati-hati. Meskipun visualisasi sering kali menarik perhatian dengan tampilannya yang menarik, kesalahan interpretasi tetap dapat terjadi. Dalam materi ini, kita akan menjelaskan karakteristik visualisasi data yang ideal dan bagaimana mengadopsi sikap kritis dalam memahami informasi yang disampaikan.

1. Akurat

Ketika membaca visualisasi data, penting untuk memastikan bahwa visualisasi tersebut merepresentasikan data dan tren dengan akurat. Beberapa langkah yang dapat diambil untuk memastikan keakuratan visualisasi adalah:

- Periksa sumber data: Pastikan visualisasi didasarkan pada sumber data yang andal dan valid. Periksa integritas data dan pastikan bahwa data yang digunakan relevan dengan konteks yang sedang dibahas.
- Verifikasi metode visualisasi: Periksa metode dan algoritma yang digunakan dalam pembuatan visualisasi. Pastikan bahwa metode ini tepat untuk jenis data yang digunakan dan sesuai dengan tujuan analisis.
- Periksa skala dan representasi grafis: Perhatikan skala sumbu dan representasi grafis yang digunakan dalam visualisasi. Pastikan bahwa mereka tidak menyesatkan atau memperlihatkan informasi yang salah.

2. Jelas

Visualisasi data harus mudah dimengerti oleh pembaca. Beberapa prinsip yang perlu dipertimbangkan untuk memastikan kejelasan visualisasi adalah:

- Pilih visualisasi yang sesuai: Pilih jenis visualisasi yang paling cocok untuk data yang sedang dianalisis dan tujuan yang ingin dicapai. Pertimbangkan apakah grafik garis, diagram batang, scatter plot, atau visualisasi lainnya yang paling efektif untuk menyampaikan pesan.

Mengukur kualitas Visualisasi

- Gunakan label yang jelas: Pastikan setiap elemen dalam visualisasi memiliki label yang jelas dan dapat dipahami. Ini termasuk sumbu, judul grafik, dan penjelasan untuk setiap elemen yang mungkin membingungkan bagi pembaca.
- Sederhana dan terorganisir: Hindari kelebihan informasi dalam visualisasi. Sederhanakan visualisasi dengan hanya menyertakan elemen yang paling relevan dan penting untuk menjelaskan pesan yang ingin disampaikan.

3. Memberdayakan

Visualisasi data harus memberdayakan pembaca untuk mengambil tindakan atau membuat keputusan yang tepat. Beberapa langkah yang dapat diambil untuk mencapai tujuan ini adalah:

- Jelaskan insight dan implikasi: Sertakan narasi yang jelas yang menjelaskan insight yang ditemukan dari visualisasi. Jelaskan implikasi dari insight tersebut dan bagaimana hal itu dapat mempengaruhi pengambilan keputusan atau tindakan yang dapat diambil.
- Sajikan data komparatif: Gunakan visualisasi data untuk membandingkan kategori, tren, atau metrik yang relevan. Ini membantu pembaca dalam memahami perbandingan antara data dan mengidentifikasi perbedaan yang signifikan. Misalnya, grafik batang dapat digunakan untuk membandingkan penjualan produk antar tahun atau antar wilayah.
- Tawarkan interaktivitas: Jika memungkinkan, gunakan elemen interaktif dalam visualisasi data. Ini memungkinkan pembaca untuk menjelajahi data lebih lanjut, memfilter atau menyortir data sesuai kebutuhan mereka, dan mendapatkan wawasan yang lebih mendalam.

Mengukur kualitas Visualisasi

- Sertakan konteks: Visualisasi data yang efektif juga menyertakan konteks yang relevan. Berikan informasi tambahan tentang faktor eksternal, tren pasar, atau kebijakan yang dapat mempengaruhi data yang disajikan. Hal ini membantu pembaca memahami visualisasi dalam konteks yang lebih luas.

4. Singkat

Visualisasi data yang efektif harus menyampaikan pesan secara singkat. Beberapa prinsip untuk membuat visualisasi yang singkat adalah:

- Sederhana dan ringkas: Kurangi kelebihan informasi yang tidak relevan atau membingungkan. Fokuskan pada informasi inti yang ingin disampaikan dan sederhanakan visualisasi sesuai kebutuhan.
- Gunakan elemen visual yang tepat: Gunakan elemen visual seperti grafik, ikon, atau warna untuk menyampaikan informasi secara ringkas. Misalnya, penggunaan warna yang berbeda untuk membedakan kategori atau grafik garis yang menunjukkan tren dengan jelas.
- Gunakan headline yang menonjol: Sertakan headline yang singkat dan menarik untuk menyoroti pesan utama dalam visualisasi. Headline yang jelas dan menonjol membantu pembaca dalam memahami informasi penting dengan cepat.
- Hilangkan kelebihan desain: Hindari penggunaan elemen desain yang berlebihan atau terlalu rumit. Visualisasi yang terlalu "ramai" dapat mengaburkan pesan yang ingin disampaikan.

EXERCISE

Berikut adalah langkah - langkah untuk mengerjakan *exercise*:

1. Buka *link* Google Colab menggunakan Google Chrome.
2. Klik 'File'.
3. Klik 'Save a Copy in Drive'.
4. Kerjakan *exercise* dengan mengisi coding pada baris yang tertera dan menjawab Insight yang ditanyakan.
5. Untuk baris coding yang sudah terisi, anda disarankan untuk langsung melakukan run baris tersebut tanpa mengubah isinya.
6. Perhatikan instruksi lain yang tertera pada Google Colab.

Google Colab dapat diakses pada:

[Exercise Modul 2 DS](#)

THANK YOU