A

**A/B Testing (Pengujian A/B):** Proses pengujian dua variasi laman web yang sama untuk menentukan laman mana yang lebih banyak digunakan pengguna dan menghasilkan pendapatan

***Absolute reference* (Referensi absolut):** Referensi dalam suatu *function* yang terkunci sedemikian hingga baris dan kolomnya tidak akan berubah jika *function* itu disalin

***Access control* (Kontrol akses):** Fitur seperti proteksi kata sandi, izin pengguna, dan enkripsi yang digunakan untuk melindungi aplikasi *spreadsheet*

***Accuracy* (Ketepatan):** Sejauh mana data sesuai dengan entitas aktual yang diukur atau digambarkan

**Action-oriented question (Pertanyaan yang mengarah pada tindakan):** Pertanyaan yang jawabannya mengarah pada perubahan

***Administrative Metadata* (Metadata administratif):** Metadata yang menunjukkan sumber teknis dari aset digital

***Aesthetic*(R) (Estetika (R)):** Properti visual dari sebuah objek dalam plot

**Agenda:** Daftar janji pertemuan yang telah dijadwalkan

**Aggregation (Agregasi):** Proses mengumpulkan atau menjadikan banyak potongan yang terpisah menjadi satu kesatuan

**Algoritma:** Proses atau seperangkat aturan yang harus diikuti untuk tugas tertentu

***Aliasing*:** Menamai tabel atau kolom dalam suatu kueri untuk sementara agar lebih mudah dibaca dan ditulis

***Alternative text* (Teks alternatif):** Teks yang menyediakan alternatif konten non-teks, seperti gambar dan video

**Analisis:** Proses yang digunakan untuk memahami data yang dikumpulkan

**Analytical skills (Keterampilan analitik):** Kualitas dan karakteristik yang terkait dengan fakta untuk memecahkan masalah

**Analytical thinking (Pemikiran analitis):** Proses untuk mengidentifikasi dan mendefinisikan permasalahan, kemudian menyelesaikannya dengan menggunakan data yang terorganisir, dilakukan dengan bertahap

***Annotation* (Anotasi):** Teks yang menjelaskan data secara singkat atau membantu audiens untuk fokus pada aspek tertentu dari visualisasi data.

***Anscombe’s quartet* (Kuartet Anscombe):** Empat *dataset* yang memiliki ringkasan statistik yang hampir mirip tetapi berisi nilai yang dipetakan berbeda

***Area chart* (Bagan area):** Visualisasi data yang menggunakan tiap *data point* untuk perubahan variabel yang dihubungkan sebuah garis kontinu dengan area yang terisi di bawahnya

**Argument (R):** Informasi yang dibutuhkan oleh suatu function di R untuk bekerja

***Arithmetic operator* (Operator aritmatika):** Operator yang digunakan untuk melakukan operasi matematika dasar seperti penambahan, pengurangan, perkalian, dan pembagian

***Array*:** Kumpulan nilai yang ada dalam sel-sel *spreadsheet*

***Assignment operator* (Operator penugasan/penetapan):** Operator yang digunakan untuk menetapkan nilai ke variabel dan vektor

**Atribut**: Karakteristik atau kualitas data yang digunakan untuk pelabelan kolom di dalam tabel

***Audio file* (File audio):** Audio digital biasanya disimpan dengan menggunakan format MP3, AAC, atau format terkompresi yang lainnya

**AVERAGE:** Fungsi di *spreadsheet* yang menampilkan rata-rata nilai dari rentang data yang dipilih

**AVERAGEIF:** Sebuah function di spreadsheet yang menampilkan nilai rata-rata dari semua sel pada suatu rentang yang memenuhi kondisi tertentu

B

**Bad data source (Sumber data yang buruk):** Sumber data yang tidak dapat diandalkan, tidak asli, tidak komprehensif, tidak terkini, dan tidak memiliki sumber yang jelas

***Balance* (Keseimbangan):** Prinsip desain yang menciptakan daya tarik estetika dan kejelasan pada visualisasi data dengan cara mendistribusikan elemen visual secara merata

**Bar graph (Bagan batang):** Visualisasi data yang menggunakan ukuran bentuk untuk mengontraskan dan membandingkan dua nilai atau lebih

**Bias:** Preferensi sadar atau bawah sadar yang mendukung atau berlawanan terhadap seseorang, sekelompok orang, atau hal tertentu

***Big Data*:** *Dataset* yang besar dan kompleks di mana biasanya mencakup jangka waktu panjang, yang memungkinkan analis data untuk mengatasi masalah bisnis dengan jangkauan yang luas

***Boolean data* (Data Boolean):** Data yang hanya memiliki dua kemungkinan nilai, biasanya benar atau salah

**Borders**: Garis yang dapat ditambahkan pada sekeliling dua sel atau lebih pada aplikasi spreadsheet

**Box plot (Bagan kotak garis):** Visualisasi data yang menampilkan distribusi nilai di sepanjang sumbu x

**Bubble chart (Bagan gelembung)**: Visualisasi data yang menampilkan titik data individual sebagai gelembung, membandingkan nilai numerik dengan ukuran relatifnya

***Bullet graph* (Bagan peluru/garis):** Visualisasi data yang menampilkan data sebagai bagan batang horizontal yang bergerak menuju nilai yang diinginkan

***Business metric* (Metrik bisnis):** Standar pengukuran yang digunakan untuk menyelesaikan tugas bisnis

***Business task* (Tugas bisnis):** Pertanyaan atau masalah yang diselesaikan melalui analisis data untuk suatu bisnis

C

**C#:** Bahasa pemrograman berorientasi objek yang digunakan untuk membuat *game* dan aplikasi seluler di *platform* pengembang *open source* .NET

**C++:** Perluasan bahasa pemrograman C yang digunakan untuk membuat game konsol, seperti untuk Xbox

***Calculated field* (Bidang terhitung):** Sebuah *field* baru di tabel pivot yang melakukan perhitungan tertentu berdasarkan nilai *field*-*field* lainnya

***Calculus* (Kalkulus):** Cabang ilmu matematika yang meliputi studi tentang tingkat perubahan dan perubahan nilai yang terkait dengan sebuah fungsi*.*

**Case study (Studi kasus):** Cara yang biasa dilakukan pemberi kerja untuk menilai keterampilan kerja dan mendapatkan wawasan tentang bagaimana seorang kandidat melakukan pendekatan terhadap permasalahan umum terkait data

**CASE:** Sebuah pernyataan SQL yang mengembalikan catatan yang memenuhi kondisi dengan memasukkan pernyataan jika/kemudian dalam kueri

**CAST:** Sebuah *function* SQL yang mengonversi data dari satu *datatype* (tipe data) ke yang lain

**Causation (Kausalitas):** Ketika suatu tindakan secara langsung mengarah ke hasil, contohnya seperti hubungan sebab-akibat

***Cell reference* (Referensi sel):** Sebuah sel atau kelompok sel dalam *worksheet* yang biasanya digunakan dalam *formula* dan *function*

**Changelog:** *File* yang berisi daftar modifikasi yang diurutkan secara kronologis yang dibuat untuk sebuah proyek

**Channel (Saluran):** Aspek visual atau variabel yang merepresentasikan karakteristik data pada suatu visualisasi

**Chart (Bagan):** Representasi grafis dari data di sebuah worksheet

***Circle view* (Tampilan melingkar):** Visualisasi data yang menunjukkan kekuatan komparatif dalam data

**Clean data (data bersih):** Data yang lengkap, benar, dan relevan dengan masalah yang dipecahkan

***Cloud*:** Tempat untuk menyimpan data secara *online*, bukan di *hard drive* komputer

***Cluster* (klaster):** Kumpulan *data point* di visualisasi data dengan nilai yang sama

**COALESCE:** Sebuah function SQL yang mengembalikan nilai-nilai bukan nol dalam daftar

***Coding* (Pengkodean):** Proses penulisan instruksi dalam *syntax* bahasa pemrograman tertentu ke sebuah komputer

***Column chart* (Bagan kolom):** Visualisasi data yang menggunakan tiap *data point* untuk perubahan variabel, yang direpresentasikan sebagai kolom vertikal

***Combo chart* (Bagan kombinasi):** Visualisasi data yang menggabungkan lebih dari satu jenis visualisasi

**Compatibility (Kompatibilitas):** Seberapa baik dua dataset atau lebih dapat digunakan secara bersamaan

***Completeness* (Kelengkapan):** Sejauh mana data berisikan semua komponen atau ukuran yang diinginkan

**Computer programming (pemrograman komputer):** Proses memberikan instruksi ke komputer untuk melakukan satu atau serangkaian tindakan

**CONCAT:** Suatu fungsi SQL yang menambahkan *string* secara bersamaan untuk membuat *string* teks baru yang dapat digunakan sebagai kunci unik

**CONCATENATE:** Sebuah function di spreadsheet yang menggabungkan dua atau lebih string teks

**Conditional formatting (Pemformatan bersyarat):** Alat pada spreadsheet yang mengubah penampakan sel pada saat ada nilai yang memenuhi kondisi tertentu

**Conditional statement (Pernyataan bersyarat):** Pernyataan di mana jika kondisi tertentu terpenuhi, maka suatu peristiwa harus terjadi

***Confidence interval* (Interval kepercayaan):** Rentang nilai yang menunjukkan seberapa besar estimasi statistik merefleksikan suatu populasi

**Confidence level (Tingkat kepercayaan):** Probabilitas bahwa ukuran sampel mencerminkan populasi yang lebih besar secara akurat

**Confirmation bias ( Bias konfirmasi):** Kecenderungan untuk mencari atau menafsirkan informasi dengan cara menegaskan keyakinan yang sudah ada sebelumnya

***Consent* (Persetujuan):** Aspek kode etik data yang menganggap bahwa tiap individu berhak untuk mengetahui bagaimana dan mengapa data pribadi mereka akan digunakan sebelum mereka menyetujui untuk memberikan data tersebut

**Consistency (Konsistensi):** Sejauh mana data dapat diulang dari tempat atau pengumpulan yang berbeda

***Context* (Konteks):** Kondisi pada sesuatu di mana hal tersebut dapat muncul atau terjadi.

***Continuous data* (Data Kontinu):** Data yang diukur dan dapat memiliki hampir semua nilai numerik

**CONVERT:** *Function* SQL yang mengubah unit pengukuran nilai dalam data

***Cookie*:** Sebuah *file* yang berukuran kecil yang tersimpan di komputer yang berisi informasi tentang penggunanya

***Correlation* (Korelasi):** Ukuran seberapa jauh perubahan pada dua variabel yang berhubungan satu sama lain

**COUNT DISTINCT:** Sebuah *function* SQL yang hanya menampilkan nilai yang berbeda di suatu rentang tertentu

**COUNT**: **Sebuah function di spreadsheet yang menghitung jumlah sel dalam rentang yang memenuhi kriteria tertentu**

**COUNTA:** **Sebuah function di spreadsheet yang menghitung jumlah total nilai dalam rentang tertentu**

**COUNTIF: Sebuah function di spreadsheet yang menghitung jumlah sel dalam rentang yang memenuhi kriteria tertentu**

**CRAN (Comprehensive R Archive Network) (R):** Arsip *online* yang berisi paket R, sumber kode, panduan, dan dokumentasi

**CREATE TABLE:** Klausa di SQL yang menambahkan tabel sementara ke database agar dapat digunakan oleh banyak orang

**Cross-field validation (Validasi lintas bidang):** Proses yang memastikan kondisi tertentu terpenuhi untuk beberapa field

**CSS (*Cascading Style Sheets*):** Bahasa pemrograman yang digunakan untuk mendesain laman web yang mengontrol elemen grafis dan presentasi halaman

**CSV file - comma-separated values (File CSV, ):** File teks terbatas yang menggunakan koma untuk memisahkan nilai

***Currency* (Nilai material/finansial):** Aspek kode etik data yang menganggap bahwa tiap individu perlu mengetahui transaksi keuangan yang dihasilkan dari penggunaan data pribadi mereka dan skala transaksi tersebut

D

**Dashboard (Dasbor):** Alat yang berfungsi untuk memonitor data yang masuk

***Dashboard filter* (Filter dasbor):** Alat untuk menampilkan hanya untuk data yang memenuhi kriteria tertentu saja, sementara sisanya disembunyikan

***Data aggregation* (Agregasi data):** Proses pengumpulan data dari berbagai sumber dan menggabungkannya menjadi bentuk yang ringkas

**Data analysis (Analisis data):** Proses pengumpulan, transformasi, dan penataan data untuk mengambil kesimpulan, membuat prediksi, dan memberi informasi untuk pengambilan keputusan yang tepat.

**Data analysis process (Proses analisis data):** Terdiri dari enam fase yaitu, bertanya, mempersiapkan, memproses, menganalisis, berbagi, dan bertindak yang tujuannya adalah untuk mendapatkan wawasan yang mendukung pengambilan keputusan

***Data analyst* (Analis data):** Seseorang yang mengumpulkan, mentransformasi, dan mengatur data untuk mengambil kesimpulan, membuat prediksi, dan mendorong pengambilan keputusan berdasarkan informasi yang tepat.

**Data analytics (Analitik data):** Ilmu tentang data

***Data anonymization* (Anonimisasi data):** Proses perlindungan data pribadi atau data sensitif dengan cara menghilangkan informasi pengenalnya

***Data bias* (Bias data):** Ketika preferensi yang cenderung mendukung atau menolak seseorang, sekelompok orang, atau hal lainnya secara sistematis mempengaruhi hasil analisis data, sehingga hasilnya condong ke arah tertentu

***Data blending* (Pencampuran data):** Metode Tableau yang menggabungkan data dari beberapa sumber

***Data composition* (Komposisi data):** Proses menggabungkan bagian-bagian individual menjadi visualisasi dan menampilkannya secara keseluruhan

**Data constraints (Batasan data):** Kriteria yang menentukan apakah kondisi data sudah bersih dan valid untuk digunakan

**Data design (Desain data):** Bagaimana informasi diatur

**Data ecosystem (Ekosistem data):** Berbagai elemen yang saling berinteraksi untuk menghasilkan, mengelola, menyimpan, mengatur, menganalisa, dan berbagi data

***Data element* (Elemen data):** Sekeping informasi pada *dataset*

**Data engineer (Insinyur data):** Seorang profesional yang mengubah data menjadi format yang berguna untuk analisis dan memberikannya infrastruktur yang bisa diandalkan

**Data ethics (Kode etik pada data):** Standar kebenaran atau kesalahan data yang berdasar pada bagaimana data dikumpulkan, dibagikan, dan digunakan

**Data frame (Bingkai data):** Kumpulan kolom yang berisi data, mirip dengan spreadsheet atau tabel SQL

**Data governance (Tata kelola data):** Sebuah proses untuk memastikan adanya manajemen formal terhadap aset data perusahaan

**Data integrity (Integritas data):** Keakuratan, kelengkapan, konsistensi, dan data yang terpercaya di sepanjang siklus hidup data tersebut

***Data interoperability* (Interoperabilitas data):** Kemampuan untuk mengintegrasikan data dari berbagai sumber dan menjadi faktor kunci yang menentukan keberhasilan penggunaan *open data* pada perusahaan dan pemerintah

***Data life cycle* (Siklus data):** Urutan tahapan pada data, yang meliputi perencanaan, menangkap, mengelola, menganalisis, mengarsipkan, dan menghancurkan data

***Data manipulation* (Manipulasi data):** Proses pengubahan data agar lebih teratur dan mudah dibaca

***Data mapping* (Pemetaan data):** Proses pencocokan *field* dari satu sumber data ke sumber data yang lain

**Data merging (Penggabungan data):** Proses menggabungkan dua dataset atau lebih ke dalam satu dataset

**Data model (Model data):** Alat untuk mengatur elemen data dan bagaimana mereka saling berhubungan

***Data privacy* (Privasi data):** Perlindungan informasi dan aktivitas subjek data setiap kali transaksi data terjadi

**Data range (Rentang data):** Nilai numerik yang terletak di antara nilai maksimum dan minimum yang telah ditentukan sebelumnya

**Data replication (Replikasi data):** Proses penyimpanan data di beberapa lokasi

**Data science (Ilmu data):** Bidang studi yang menggunakan data mentah untuk menciptakan pemodelan dan pemahaman yang baru pada hal yang belum diketahui

***Data security* (Keamanan data):** Perlindungan data dari akses yang tidak sah atau kerusakan dengan melakukan tahapan proses keamanan

**Data storytelling (Penceritaan data):** Mengkomunikasikan arti dataset dengan visual dan narasi yang disesuaikan untuk audiens

**Data strategy (Strategi data):** Mengelola manusia, proses, dan alat yang digunakan pada analisis data

***Data structure* (Struktur data):** Sebuah format untuk mengatur dan menyimpan data

**Data transfer (Transfer data):** Proses penyalinan data dari perangkat penyimpanan dari satu memori komputer ke memori komputer lain

***Data type* (Tipe data):** Atribut yang menggambarkan sekeping data berdasarkan nilai-nilainya, bahasa pemrogramannya, atau operasi yang dapat dilakukannya

***Data validation* (Validasi data):** Alat untuk memeriksa keakuratan dan kualitas pada data

**Data validation process (Proses validasi data):** Proses pengecekan kualitas data berulang kali sehingga data tersebut lengkap, akurat, aman dan konsisten

***Data visualization* (Visualisasi data):** Representasi grafis dari data

**Data warehousing specialist (Ahli pergudangan data):** Seorang profesional yang mengembangkan proses dan prosedur untuk menyimpan dan mengatur data secara efektif

**Data:** Sekumpulan fakta

**Database (Basis data):** Sekumpulan data yang disimpan di dalam sistem komputer

***Data-driven decision-making* (Pengambilan keputusan berdasarkan data):** Menggunakan fakta-fakta sebagai panduan merumuskan strategi bisnis

**Data-inspired decision-making (Pengambilan keputusan yang terinspirasi data):** Mengeksplorasi sumber data yang berbeda untuk mengetahui kesamaannya

**Dataset (Himpunan/Kumpulan data):** Kumpulan data yang dapat dimanipulasi atau dianalisis sebagai satuan unit

**DATEDIF:** Sebuah *function* dalam *spreadsheet* yang menghitung jumlah hari, bulan, atau tahun di antara dua tanggal

**Decision tree (Pohon keputusan)**: Alat yang membantu analis membuat keputusan mengenai fitur-fitur kritis dari suatu visualisasi

***Delimiter* (Pembatas):** Karakter yang menunjukkan awal atau akhir item data

***Density map* (Peta kepadatan):** Visualisasi data yang mewakili konsentrasi, dengan penggunaan warna yang mewakili jumlah atau frekuensi *data point* di area tertentu pada sebuah peta

**Descriptive metadata (Metadata deskriptif):** Metadata yang mendeskripsikan sebuah data dan dapat digunakan untuk mengidentifikasi data di waktu selanjutnya

***Design thinking* (Pemikiran desain):** Prosesyang digunakan untuk memecahkan masalah kompleks dengan cara yang mengutamakan pengguna

***Digital photo* (Foto digital):** Gambar elektronik atau gambar berbasis komputer, biasanya menggunakan format BMP atau JPG

***Dirty data* (Data kotor):** Data yang tidak lengkap, salah, atau tidak relevan dengan masalah yang harus dipecahkan

***Discrete data* (Data diskret):** Data yang dihitung dan memiliki jumlah nilai yang terbatas

**DISTINCT:** Sebuah kata kunci yang ditambahkan ke pernyataan *SELECT* pada SQL untuk mengambil entri non-duplikat saja

***Distribution graph* (Bagan distribusi):** Visualisasi data yang menampilkan frekuensi berbagai hasil sampel

***Diverging color palette*:** Tema warna yang menampilkan dua rentang nilai data dengan menggunakan dua rona warna yang berbeda, dengan intensitas warna yang mewakili besarnya nilai

***Donut chart* (Bagan donat):** Suatu visualisasi data di mana segmen sebuah cincin mewakili nilai data sesuai proporsinya dari keseluruhan

**dplyr (R):** Sebuah paket R dalam Tidyverse yang menawarkan serangkaian fungsi yang konsisten untuk menyelesaikan tugas manipulasi data yang umum

***DROP TABLE* (TABEL DROP):** Klausa SQL yang menghapus tabel sementara dari *database*

**Duplicate data (Data duplikat):** Setiap *record* yang secara tidak sengaja berbagi data dengan *record* yang lainnya

**Dynamic visualization (Visualisasi dinamis):** Visualisasi data yang interaktif atau berubah sepanjang waktu

E

***Emphasis* (Penekanan):** Prinsip desain yang mengatur elemen-elemen visual untuk memfokuskan perhatian audiens pada informasi penting dalam sebuah visualisasi data

***Engagement* (Pelibatan):** Menarik minat dan mempertahankan ketertarikan maupun perhatian seseorang selama suatu presentasi data

**Equation (Persamaan):** Penghitungan yang melibatkan penambahan, pengurangan, perkalian, atau pembagian (disebut juga sebagai ekspresi matematika)

**Estimated response rate (Estimasi tingkat respons)**: Jumlah rata-rata orang yang biasanya menyelesaikan survei

***Ethics* (Etika):** Standar mengenai kebenaran dan kesalahan yang manusia lakukan, biasanya terkait dengan kewajiban, hak, manfaat bagi masyarakat, keadilan, atau kebajikan tertentu.

**Experimenter bias (Bias pada eksperimen):** Kecenderungan di mana orang yang berbeda memiliki cara pengamatan yang berbeda juga (Silakan merujuk pada Bias pengamat)

***External data* (Data eksternal):** Data yang ada dan dihasilkan di luar organisasi

F

**Facets (R):** Serangkaian fungsi yang membagi data menjadi himpunan bagian dalam matriks panel

**Factor (Faktor) (R):** Suatu objek yang menyimpan data kategoris di mana nilai datanya terbatas dan biasanya didasarkan pada kelompok yang terbatas, seperti negara atau tahun

**Fairness (Kewajaran):** Kualitas pada analisis data dengan tidak menciptakan atau memperkuat bias pada data

**Field length (Panjang field):** Alat untuk menentukan berapa banyak karakter yang dapat dimasukkan ke dalam field yang ada di spreadsheet

***Field*:** Sebuah informasi dari baris atau kolom yang ada di *spreadsheet*; terletak di dalam tabel data, biasanya di kolom tabel

**Fill handle:** Kotak yang terletak di sudut kanan bawah sel spreadsheet yang dipilih yang dapat diseret melalui sel di sekitarnya untuk melanjutkan instruksi

***Filled map* (Peta yang terisi):** Visualisasi data yang mewarnai area pada peta berdasarkan pengukuran atau dimensi

**Filtering (Penyaringan):** Proses yang digunakan untuk menunjukkan data yang memenuhi kriteria tertentu dan menyembunyikan sisanya

**Find and replace (Mencari dan mengganti):** Suatu alat untuk menemukan istilah pencarian tertentu dan menggantikannya dengan sesuatu yang lain

***First-party-data* (Data pihak pertama):** Data yang dikumpulkan oleh individu atau kelompok menggunakan sumber daya mereka sendiri

**Float (Mengapung):** Suatu angka yang berisi desimal

**Foreign key:** Sebuah field di dalam tabel database yang merupakan primary key pada tabel lainnya (Silakan merujuk pada primary key)

**Formula:** Formula adalah sekumpulan instruksi yang digunakan untuk perhitungan dengan menggunakan data di spreadsheet.

***Framework* (Kerangka kerja):** Konteks presentasi yang harus membuat hubungan logis yang kembali ke tugas dan metrik bisnis.

**FROM:** Bagian kueri yang menunjukkan tabel mana yang akan digunakan untuk mengekstrak data

***Function* (Fungsi):** Perintah yang sudah diatur (*preset*), yang secara otomatis melakukan proses atau tugas tertentu dengan menggunakan data pada *spreadsheet*

**Function (R) {Fungsi (R)}:** Kumpulan kode yang dapat digunakan ulang untuk melakukan tugas-tugas tertentu dalam R

**FWF (fixed-width file):** File teks dengan format tertentu, yang memungkinkan penyimpanan data tekstual secara terorganisir

G

**GAM (generalized additive model) smoothing (R):** Sebuah proses untuk menghaluskan plot yang memiliki berbagai poin dalam jumlah besar

***Gantt chart* (Bagan Gantt):** Visualisasi data yang menampilkan durasi peristiwa atau aktivitas pada suatu garis waktu

**Gap analysis (Analisis kesenjangan):** Sebuah metode untuk memeriksa dan mengevaluasi proses yang sedang berlangsung untuk mengidentifikasi peluang untuk perbaikan di masa mendatang

***Gauge chart* (Bagan pengukur):** Visualisasi data yang menunjukkan satu hasil dalam rentang nilai progresif

**General Data Protection Regulation of the European Union - GDPR (Peraturan Perlindungan Data Uni Eropa):** Badan pembuat kebijakan di Uni Eropa yang didirikan untuk membantu melindungi masyarakat beserta data yang mereka miliki

**Geolocation (Geolokasi):** Lokasi geografis dari seseorang atau suatu perangkat yang diperoleh melalui informasi digital

**Geom (R):** Objek geometris yang digunakan untuk merepresentasikan data

**ggplot2 (R):** Sebuah paket R di Tidyverse yang membuat berbagai visualisasi data dengan menerapkan properti visual yang berbeda ke variabel data dalam R

**Good data source (Sumber data yang baik):** Sumber data yang dapat diandalkan, asli, komprehensif, terkini, dan dikutip (reliable, original, comprehensive, current, cited, ROCCC)

**GROUP BY:** Klausa SQL yang mengelompokkan baris yang memiliki nilai yang sama dari sebuah tabel ke dalam baris ringkasan

H

**HAVING:** Klausa SQL yang menambahkan filter ke kueri, selain tabel yang mendasarinya dan digunakan menggunakan fungsi agregat

**head() (R):** Sebuah function dalam R yang menampilkan pratinjau nama kolom dan beberapa baris pertama dari sebuah dataset

**Header (Tajuk):** Baris pertama pada spreadsheet yang memberi label untuk jenis data di setiap kolom

**Headline (Judul):** Teks pada bagian atas visualisasi yang mengkomunikasikan data yang disajikan

***Heat map* (Peta suhu):** Visualisasi data yang menggunakan kontras warna untuk membandingkan kategori dalam sebuah *dataset*

**Highlight table (tabel sorotan):** Visualisasi data yang menggunakan pemformatan dan warna bersyarat pada tabel

**Histogram:** Visualisasi data menunjukkan seberapa sering nilai data berada pada rentang tertentu.

**HTML5:** Bahasa pemrograman yang menyediakan struktur untuk laman web dan terhubung ke platform-platform hosting

***Hypothesis* (Hipotesis):** Sebuah teori yang mungkin coba dibuktikan atau disangkal dengan menggunakan data

**Hypothesis testing (Pengujian hipotesis):** Proses untuk menentukan apakah survei atau eksperimen yang dilakukan memiliki hasil yang berarti

I

**IDE (Integrated Development Environment):** Aplikasi perangkat lunak yang menyatukan semua alat yang mungkin ingin digunakan oleh seorang analis data di satu tempat

**Incomplete data (Data tidak lengkap):** Data yang kehilangan field penting

**Inconsistent data (Data yang tidak konsisten):** Data yang menggunakan format yang berbeda untuk menampilkan hal yang sama

**Incorrect/inaccurate data (Data yang tidak tepat/tidak akurat):** Data yang lengkap namun tidak akurat

**INNER JOIN:** Fungsi SQL yang menampilkan data yang mempunyai nilai sama di kedua tabel

**Inner query:** Suatu sub kueri SQL yang ada di dalam pernyataan SQL lain

**Internal data (Data internal):** Data yang ada di dalam sistem perusahaan itu sendiri

**Interpretation bias (Bias pada interpretasi):** Kecenderungan untuk menafsirkan situasi yang ambigu dengan cara yang positif atau negatif

J

**Java:** Bahasa pemrograman yang banyak digunakan untuk membuat aplikasi web skala perusahaan yang dapat berjalan di berbagai jenis client

**JOIN:** Function SQL yang digunakan untuk menggabungkan baris dari dua tabel atau lebih berdasarkan kolom terkait

L

**Label:** Teks dalam visualisasi yang mengidentifikasi suatu nilai atau menjelaskan suatu skala

**Labels and annotations (R) [Label dan anotasi (R)]:** Sekelompok function R yang digunakan untuk menyesuaikan plot

**Leading question (Pertanyaan yang terarah):** Pertanyaan yang mengarahkan peserta untuk memberikan respons tertentu

**LEFT JOIN:** *Function* SQL yang akan menampilkan semua data dari tabel kiri dan data dari tabel kanan yang cocok nilainya

**LEFT:** *Function* yang menghasilkan sejumlah karakter dari sisi kiri *string* teks

***Legend* (Legenda):** Fitur yang mengidentifikasi arti berbagai elemen dalam sebuah visualisasi data

**LEN*:*** *Function* yang menghasilkan panjang *string* teks dengan menghitung jumlah karakter yang dimiliki.

***Length* (Panjang)*:*** Jumlah karakter pada *string* teks

**Library (Pustaka):** Direktori yang berisi semua paket yang diinstal oleh analis data

**LIMIT:** Klausa SQL yang menentukan jumlah maksimum data yang dihasilkan dalam suatu kueri

***Line Graph* (Grafik garis):** Sebuah visualisasi data yang menggunakan satu atau beberapa baris untuk menampilkan pergeseran atau perubahan data dari waktu ke waktu

**List (Daftar):**  Vektor yang elemennya dapat berbentuk apa saja

***Live data* (Data langsung):** Data yang diperbarui secara otomatis

**Loess smoothing (R):** Sebuah proses yang digunakan untuk menghaluskan plot dengan nilai kurang dari 1.000 poin

***Log file* (Berkas log):** File yang dihasilkan komputer yang merekam kejadian dari sistem operasi dan program perangkat lunak lainnya

***Logical operator* (Operator logika):** Operator yang menampilkan tipe data yang logis

***Long data:*** *Dataset* yang di tiap barisnya memiliki satu titik waktu per subjeknya, sehingga setiap subjek memiliki data pada beberapa baris

M

**Mandatory (Wajib):** Nilai data yang tidak boleh dikosongkan

***Map* (Peta):** Visualisasi data yang mengatur data secara geografis

***Mapping* (Pemetaan) (R):** Proses mencocokkan variabel tertentu dalam kumpulan data dengan estetika tertentu

***Margin of error* (Margin kesalahan):** Jumlah maksimum hasil sampel yang di antisipasi akan berbeda dari populasi sebenarnya

**Mark (Tanda):** Objek visual dalam visualisasi data seperti sebuah titik, garis, atau bentuk

**MATCH:** Sebuah function di spreadsheet yang digunakan untuk menemukan posisi nilai pencarian tertentu

***Math expression* (Ekspresi matematika):** Penghitungan yang melibatkan penambahan, pengurangan, perkalian, atau pembagian (disebut juga dengan persamaan matematika)

**Math function (Fungsi matematika)**: Sebuah fungsi yang digunakan di sebagian formula matematika

***Matrix* (Matriks):** Kumpulan elemen data dua dimensi dengan baris dan kolom

**MAX:** Sebuah function di spreadsheet yang menghasilkannilai numerikterbesar dari suatu rentang sel

**MAXIFS:** Sebuah function pada spreadsheet yang menampilkan nilai maksimum dari rentang tertentu yang memenuhi kondisi tertentu

**McCandless *Method* (Metode McCandless):** Metode untuk menyajikan visualisasi data yang bergerak dari informasi umum ke informasi tertentu

**Measurable question (Pertanyaan terukur):** Sebuah pertanyaan yang jawabannya dapat diukur dan dinilai

***Mental model* (Model mental):** Proses pemikiran dan pendekatan terhadap suatu masalah oleh seorang analis data

***Mentor* (Pembimbing):** Seseorang yang berbagi pengetahuan, keterampilan, dan pengalaman untuk membantu orang lain berkembang menjadi seorang profesional maupun personal

***Merger* (Penggabungan):** Kesepakatan yang menyatukan dua organisasi menjadi satu organisasi baru

***Metadata repository* (Repositori metadata):** *Database* yang dibuat untuk menyimpan *metadata*

***Metadata*:** Data tentang data

**Metric (Metrik):** Jenis data tunggal yang dapat dihitung, yang digunakan untuk pengukuran

***Metric Goal* (Tujuan metrik):** Tujuan metrik adalah tujuan terukur yang ditetapkan perusahaan untuk dievaluasi menggunakan metrik.

**MID*:*** Sebuah fungsi yang mengembalikan segmen yang terletak di tengah *string* teks

**MIN:** Sebuah *function* di *spreadsheet* yang menghasilkan nilai numerik terkecil dari suatu rentang sel

**MINIFS:** Sebuah *function* pada *spreadsheet* yang menampilkan nilai minimum dari suatu rentang yang memenuhi kondisi tertentu

**Modulo:** Operator (%) yang mengembalikan sisa pembagian bilangan terhadap bilangan lainnya

***Movement* (Gerakan):** Prinsip desain yang mengatur elemen visual untuk memandu mata audiens dari satu bagian visualisasi data ke bagian lain

**mutate() (R):** Sebuah *function* pada R yang membuat perubahan pada kerangka data yang memisahkan dan menggabungkan kolom atau membuat variabel baru

N

***Naming conventions* (Konvensi penamaan):** Pedoman konsisten yang menjelaskan konten, tanggal pembuatan, dan versi *file* atas namanya

***Narrative* (Narasi):** (Silakan merujuk pada *story*)

***Nested* (Bersarang):** *Code* yang menjalankan *function* tertentu dan terkandung di dalam *code* yang menjalankan *function* yang lebih luas

***Nested function* (Fungsi bersarang):** Sebuah *function* yang sepenuhnya terkandung dalam *function* lain

**Networking (Berjejaring):** Membangun hubungan dengan bertemu orang-orang, baik secara langsung maupun online

***Nominal data* (Data nominal):** Jenis data kualitatif yang dikategorikan tanpa urutan tertentu

***Normalized database* (*Database* yang dinormalisasi):** Sebuah *database* di mana hanya data terkait yang disimpan di setiap tabel

***Notebook*:** Lingkungan pemrograman yang interaktif dan dapat diedit untuk membuat laporan data dan menampilkan keterampilan data

**Null (Kosong):** Indikasi adanya nilai yang tidak terdapat di dalam dataset

O

**Observasi:** Atribut yang menggambarkan sepotong data yang terkandung dalam suatu baris tabel

**Observer bias (Bias pengamat):** Kecenderungan di mana orang yang berbeda mengamati dengan cara yang berbeda (juga disebut bias eksperimenter)

**Open data (Data terbuka):** Data yang tersedia untuk umum

***Openness* (Keterbukaan):** Aspek etika data yang mengedepankan akses, penggunaan, dan pembagian data secara gratis

***Open-source* (Sumber terbuka):** *Code* yang tersedia secara bebas dan dapat dimodifikasi serta dibagikan oleh orang yang menggunakannya

**Operator:** Simbol yang menamai operasi atau penghitungan yang akan dilakukan

**ORDER BY:** Sebuah klausa SQL yang mengurutkan hasil dalam suatu kueri

***Order of operations* (Urutan operasi):** Menggunakan tanda kurung untuk mengelompokkan bersama nilai-nilai *spreadsheet* untuk memperjelas urutan operasi yang harus dilakukan

***Ordinal data* (Data ordinal):** Data kualitatif dengan urutan atau skala yang ditetapkan

**Outdated data (Data usang):** Setiap data yang telah digantikan oleh informasi terbaru dan lebih akurat

**OUTER JOIN:** *Function* SQL yang menggabungkan RIGHT JOINdan LEFT JOIN untuk menampilkan semua data yang cocok di kedua tabel

**Outer query:** Pernyataan SQL yang berisi sub-kueri

***Ownership* (Kepemilikan):** Aspek etika data yang menganggap bahwa setiap individu adalah pemilik data mentah yang mereka berikan dan memiliki kontrol utama atas penggunaan, pemrosesan, dan pembagiannya

P

***Package* (R) [Paket (R)]:** Satuan kode pada R yang dapat direproduksi

***Packed bubble chart (Bagan gelembung yang dikemas)*:** Visualisasi di mana data ditampilkan dalam lingkaran yang berkelompok.

***Pattern* (Pola):** Prinsip desain yang menggunakan elemen visual yang sama untuk menunjukkan tren dan hubungan dalam visualisasi data

**PHP (Hypertext Preprocessor):** Bahasa pemrograman untuk pengembangan aplikasi web

***Pie chart* (Bagan pai):** Visualisasi data yang menggunakan segmen lingkaran untuk mewakili proporsi setiap kategori data bila dibandingkan dengan keseluruhan data

**Pipe (R):** *Tool* dalam R yang digunakan untuk mengekspresikan urutan dari beberapa operasi, diwakili dengan <i>%&gt;%</i>

***Pivot chart* (Bagan pivot):** Bagan yang dibuat dari *field* dalam tabel pivot

**Pivot table (Tabel pivot):** Alat peringkas data yang digunakan untuk mengurutkan, mengatur ulang, mengelompokkan, menjumlahkan, atau menghitung rata-rata data

***Pixel* (Piksel):** Dalam pencitraan digital, area iluminasi kecil pada layar tampilan yang, bila dikombinasikan dengan area lain yang berdekatan, membentuk gambar digital

**Populasi:** Dalam analitik data, merupakan semua nilai data yang mungkin dalam *dataset*

**Portfolio:** Kumpulan materi yang dapat dibagikan dengan calon pemberi kerja

***Pre-attentive attributes*:** Aspek visualisasi data yang bisa dikenali audiens tanpa harus memikirkannya.

**Primary key:** Pengenal di database yang mereferensikan suatu kolom di mana tiap nilai bersifat unik(Silakan mengacu pada foreign key)

**Problem domain (Domain permasalahan):** Area analisis yang mencakup setiap aktivitas yang mempengaruhi atau dipengaruhi oleh masalah

***Problem types* (Jenis permasalahan):** Berbagai masalah yang dihadapi analis data, termasuk mengkategorikan berbagai hal, menemukan koneksi, menemukan pola, mengidentifikasi tema, membuat prediksi, dan mencari sesuatu yang tidak biasa

**Profit Margin (Margin keuntungan):** Persentase yang menunjukkan berapa persen keuntungan yang dihasilkan untuk setiap penjualan

**Programming language (Bahasa pemrograman):** Suatu sistem kata dan simbol yang digunakan untuk menulis instruksi yang diikuti oleh komputer

***Proportion* (Proporsi):** Prinsip desain menggunakan ukuran relatif dan pengaturan elemen visual untuk menunjukkan informasi dalam visualisasi data

**Python:** Bahasa pemrograman serbaguna

Q

**Qualitative data (Data kualitatif):** Ukuran subjektif yang memberikan penjelasan tentang suatu kualitas atau karakteristik

**Quantitative data (Data kuantitatif):** Ukuran spesifik dan objektif, seperti angka, kuantitas, atau rentang

***Query* (Kueri):** Permintaan untuk mengakses data atau informasi dari *database*.

**Query language (Bahasa kueri):** Sebuah bahasa pemrograman komputer yang digunakan untuk berkomunikasi dengan database

R

**R:** Bahasa pemrograman yang digunakan untuk analisis statistik, visualisasi, dan analisis data lainnya

**Random sampling (Sampling acak):** Cara memilih sampel dari suatu populasi sehingga setiap jenis sampel yang mungkin memiliki kesempatan yang sama untuk dipilih

***Range* (Rentang sel):** Kumpulan dua atau lebih sel dalam suatu *spreadsheet*

***Ranking* (Peringkat):** Sistem untuk memberi peringkat kepada nilai dari *dataset* dalam suatu skala pencapaian atau status.

**readr (R):** Sebuah paket dalam R di Tidyverse yang digunakan untuk mengimpor data

**Record (Catatan data):** Kumpulan data terkait dalam suatu tabel data, biasanya disebut juga dengan baris

***Redundancy* (Redundansi):** Ketika bagian yang sama dari data disimpan di dua atau lebih tempat berbeda

**Reframing (Pembingkaian ulang data):** Proses mengulangi pernyataan suatu masalah atau tantangan, kemudian mengarahkannya pada suatu resolusi yang potensial

**RegEx (Ekspresi reguler):** Aturan yang mengatakan nilai-nilai dalam tabel harus sesuai dengan pola yang ditentukan

***Relational database* (*Database* relasional):** *Database* yang berisi serangkaian tabel yang dapat dihubungkan sehingga terbentuk hubungan di antara tabel-tabel tersebut

***Relational operator* (Operator relasional):** Operator yang digunakan untuk membandingkan nilai, juga dikenal sebagai pembanding

**Relativity (Relativitas):** Proses yang mempertimbangkan pengamatan dalam kaitannya atau proporsi terhadap sesuatu yang lain

***Relevant question* (Pertanyaan yang relevan):** Sebuah pertanyaan yang memiliki arti penting terhadap masalah yang harus dipecahkan

**Remove duplicates (Hapus duplikat):** Sebuah fitur di spreadsheet yang secara otomatis mencari dan menghilangkan entri duplikat dari spreadsheet

***Repetition* (Repetisi):** Prinsip desain yang mengulang-ulang elemen visual untuk menunjukkan makna dalam sebuah visualisasi data

***Report* (Laporan):** Kumpulan data statis yang secara berkala diberikan kepada para pemangku kepentingan

***Return on Investment* (ROI, Laba atas investasi):** Formula yang menggunakan metrik dari perhitungan investasi dan keuntungan untuk mengevaluasi keberhasilan investasi

***Revenue* (Pendapatan):** Jumlah total pendapatan yang dihasilkan dari penjualan barang atau jasa

***Rhythm* (Ritme):** Prinsip desain dengan tujuan menciptakan gerakan dan aliran dalam visualisasi data untuk melibatkan audiens

**RIGHT JOIN:** *Function* SQL yang akan menampilkan semua data dari tabel kanan dan data dari tabel kiri yang cocok nilainya

**RIGHT:** Function yang mengembalikan sejumlah karakter yang terletak di sisi kanan string teks

**Root cause (Akar masalah):** Alasan mengapa masalah terjadi

**ROUND:** *Function* SQL yang menghasilkan angka yang dibulatkan ke sejumlah desimal tertentu.

**Ruby:** Bahasa pemrograman berorientasi objek untuk pengembangan aplikasi web

S

***Sample* (Sampel):** Dalam analitik data, segmen dari populasi yang mewakili seluruh populasi

***Sampling bias* (Bias pengambilan sampel):** Melebihkan atau mengurangi keterwakilan anggota populasi tertentu sebagai akibat dari pengambilan sampel yang tidak mewakili populasi secara keseluruhan

***Scatter plot* (Diagram pencar):** Visualisasi data yang mewakili hubungan antara beberapa variabel yang berbeda dengan titik data individual tanpa adanya garis penghubung

**Schema (Skema):** Cara menggambarkan bagaimana sesuatu, seperti data, diatur

**Scope of Work - SOW (Ruang Lingkup Kerja):** Garis besar tugas yang disepakati, yang akan dilakukan selama proyek berlangsung

**Second-party data (Data pihak kedua):** Data yang dikumpulkan oleh kelompok secara langsung dari audiensnya dan kemudian dijual

**SELECT INTO:** Klausa SQL yang menyalin data dari satu tabel menjadi tabel sementara tanpa menambahkan tabel baru ke database

**SELECT:** Bagian dari kueri yang menunjukkan dari mana asal kolom untuk mengekstrak data

**Small data:** Data dalam jumlah kecil dan spesifik, yang biasanya dikumpulkan dalam waktu singkat, yang berguna untuk membuat keputusan sehari-hari

**SMART Methodology (Metodologi SMART):** Alat untuk menilai efektivitas pertanyaan baik pertanyaan spesifik, terukur, berorientasi aksi, relevan, maupun terikat waktu

**Smoothing (R):** Sebuah proses yang digunakan untuk membuat visualisasi data dalam R menjadi lebih jelas dan lebih mudah dibaca

***Smoothing line* (Garis smoothing) (R):** Garis pada visualisasi data yang menggunakan pemulusan (smoothing) untuk menampilkan tren

***Social media* (Media sosial):** Situs web dan aplikasi di mana pengguna membuat dan berbagi konten atau berpartisipasi dalam jejaring sosial

**Soft skill (Keterampilan lunak):** Sifat dan perilaku nonteknis yang berhubungan dengan cara kerja orang

***Sort range* (Pengurutan rentang):** Sebuah *fungsi* dalam menu *spreadsheet* yang mengurutkan suatu rentang tertentu dan mengecualikan sel-sel di luar rentang tersebut

**Sort sheet (Pengurutan lembar):** Sebuah function dalam menu spreadsheet yang mengurutkan semua data berdasarkan peringkat kolom tertentu dan menyimpan data bersama di seluruh baris

**Sorting (Pengurutan):** Proses pengaturan data menjadi urutan yang memiliki makna untuk membuatnya lebih mudah dipahami, dianalisis, dan divisualisasikan

***Specific question* (Pertanyaan spesifik):** Sebuah pertanyaan yang sederhana, signifikan, dan berfokus pada satu topik atau beberapa ide yang terkait erat

***SPLIT:*** Sebuah *function* di *spreadsheet* yang membagi teks di sekitar karakter tertentu dan menempatkan setiap fragmen ke dalam sel baru yang terpisah.

**Sponsor:** Penasihat profesional yang berkomitmen untuk memajukan karier orang lain

**Spotlighting (Sorotan):** Memindai melalui data untuk mengidentifikasi wawasan yang paling penting dengan cepat

**Spreadsheet:** Lembar kerja digital

**SQL:** (Lihat Structured Query Language)

***Stakeholders* (Pemangku kepentingan):** Orang-orang yang menginvestasikan waktu dan sumber daya ke dalam sebuah proyek dan berkepentingan terhadap hasilnya

***Static data* (Data statis):** data yang tidak berubah setelah direkam,

***Static visualization* (Visualisasi statis):** Visualisasi data yang tidak berubah seiring waktu kecuali diedit

**Statistical power (Kekuatan statistik):** Probabilitas bahwa tes signifikansi akan mengenali adanya suatu pengaruh

**Statistical significance (Signifikansi statistik):** Probabilitas bahwa hasil sampel bukan karena kesempatan acak

***Statistics* (Statistik):** Studi tentang cara mengumpulkan, menganalisis, meringkas, dan menyajikan data

***Story* (Cerita):** Narasi saat presentasi data yang membuatnya bermakna dan menarik

**String data type (Tipe data string):** Urutan karakter dan tanda baca yang berisi informasi tekstual (juga disebut tipe data teks)

**Structural metadata (Metadata struktural):** Metadata yang menunjukkan bagaimana sepotong data diatur dan apakah itu merupakan bagian dari satu atau lebih kelompok data

***Structured data* (Data terstruktur):** Data yang diatur dalam format tertentu seperti baris dan kolom

**Structured Query Language (SQL):** Sebuah bahasa pemrograman komputer yang digunakan untuk berkomunikasi dengan database

***Structured thinking* (Pemikiran terstruktur):** Proses untuk mengenali masalah atau situasi yang sedang berlangsung, mengatur informasi yang tersedia, menemukan beberapa kesenjangan dan peluang, serta menentukan sejumlah opsi.

***Subquery*:** Kueri SQL yang disarangkan di dalam kueri yang lebih besar.

**SUBSTR:** Sebuah *function* SQL yang mengekstrak *substring* dari suatu variabel *string*

**Substring:** Sebuah subset dari string teks

***Subtitle* (Subjudul):** Teks yang mendukung judul dengan menambahkan konteks dan deskripsi

**SUM:** Suatu *function* di *spreadsheet* yang menjumlahkan nilai rentang sel yang dipilih

**SUMIF:** Suatu function di spreadsheet yang menambahkan data numerik berdasarkan satu kondisi

**Summary table (Tabel ringkasan):** Tabel yang digunakan untuk meringkas informasi statistik tentang data

**SUMPRODUCT:** Suatu *function* di *spreadsheet* yang menampilkan hasil perkalian antara baris-baris data yang bersesuaian pada beberapa *array* atau sekumpulan data.

**Swift:** Bahasa pemrograman untuk macOS, iOS, watchOS, dan tvOS

**Symbol map (Peta simbol):** Visualisasi data yang menampilkan tanda di atas garis bujur dan lintang tertentu

**Syntax (Sintaks/Sintaksis):** Struktur bahasa yang telah ditentukan yang mencakup semua kata, simbol, dan tanda baca yang diperlukan, serta penempatan tata bahasa yang tepat

T

**Tableau:** Platform intelijen dan analitik bisnis yang membantu penggunanya memvisualisasikan, memahami, dan membuat keputusan dengan data

***Technical mindset* (Pola pikir teknis):** Kemampuan untuk memecah sesuatu menjadi langkah atau bagian yang lebih kecil dan mengamatinya dengan cara yang teratur dan logis

***Temporary table* (Tabel sementara):** Tabel *database* yang dibuat dan ada untuk sementara di sebuah server *database*

**Text data type (tipe data teks):** Urutan karakter dan tanda baca yang berisi informasi tekstual (disebut juga dengan tipe data string)

***Text string* (*String* teks):** Sekelompok karakter yang ada di dalam sel, paling sering terdiri dari huruf-huruf

***Third-party data* (Data pihak ketiga):** Data dari sumber eksternal yang tidak mengumpulkannya secara langsung

**Tibble (R):** Variasi bingkai data (dataframe) yang disederhanakan

***Tidy data* (Data yang rapi) (R):** Cara standarisasi organisasi data dalam R

**tidyr (R):** Sebuah paket dalam R di Tidyverse yang digunakan untuk pembersihan data untuk membuat data yang rapi

**Tidyverse (R):** Sebuah sistem paket dalam R dengan filosofi desain umum untuk manipulasi data, eksplorasi, dan visualisasi

***Time-bound question* (Pertanyaan yang terikat waktu):** Sebuah pertanyaan yang menentukan jangka waktu yang akan dipelajari

***Transaction transparency* (Transparansi transaksi):** Aspek etika data yang menganggap bahwa semua kegiatan pengolahan data dan algoritma harus dijelaskan dan dipahami oleh individu yang memberikan data tersebut

***Transferable skills* (Keterampilan yang dapat ditransfer):** Keterampilan dan kualitas yang dapat mentransfer dari satu pekerjaan atau industri ke industri lainnya

**TRIM:** Function untuk menghilangkan spasi awal, spasi akhir, dan spasi berulang dalam data.

**TSV (Tab-separated values file):** *File* teks yang menyimpan tabel data dengan memisahkan kolom data dengan tab

***Turnover rate* (Tingkat pergantian karyawan):** Jumlah karyawan yang keluar dari perusahaan atas keinginan pribadi

**Typecasting:** Mengonversi data dari satu jenis ke tipe lainnya

U

***Unbiased sampling* (Sampling yang tidak bias):** Ketika sampel populasi yang diukur adalah perwakilan dari populasi secara keseluruhan

**Underscores (Garis bawah)**: Baris yang digunakan untuk menggarisbawahi kata-kata dan menghubungkan karakter teks

***Unfair question* (Pertanyaan yang tidak adil):** Sebuah pertanyaan yang menimbulkan asumsi atau sulit untuk dijawab dengan jujur

***Unique* (Unik):** Nilai yang tidak dapat memiliki duplikat

**United States Census Bureau (Biro Sensus Amerika Serikat):** Sebuah institusi di Departemen Perdagangan Amerika Serikat yang berfungsi sebagai penyedia data berkualitas terkemuka di negara tersebut tentang masyarakat dan ekonominya

***Unity* (Kesatuan):** Prinsip desain yang menggunakan elemen visual yang saling melengkapi untuk menciptakan daya tarik estetika dan kejelasan dalam visualisasi data

***Unstructured data* (Data tidak terstruktur):** Data yang tidak diorganisir dengan cara yang mudah diidentifikasi

V

**Validity (Validitas):** Tingkat data yang sesuai dengan kendala ketika dimasukkan, dikumpulkan, atau dibuat

**VALUE:** Sebuah *function* di *spreadsheet* yang mengubah *string* teks yang mewakili bilangan menjadi nilai numerik

**Variable (Variabel) (R):** Representasi suatu nilai dalam R yang dapat disimpan untuk digunakan nanti

**Variasi:** Prinsip desain menggunakan berbagai jenis elemen visual dalam visualisasi data untuk melibatkan audiens

***Vector* (R) [Vektor (R)]:** Sekelompok elemen data dari tipe yang sama disimpan dalam urutan satu dimensi dalam R

***Verification* (Verifikasi):** Sebuah proses yang menegaskan bahwa pembersihan data dijalankan dengan baik dan data yang dihasilkan akurat dan dapat diandalkan.

***Video File* (File video):** Kumpulan gambar, *file* audio, dan data lainnya yang biasanya dikodekan dalam format terkompresi seperti MP4, MV4, MOV, AVI, atau FLV

**Vignette (R):** Dokumentasi untuk paket dalam R yang menjelaskan masalah yang dirancang untuk dipecahkan oleh paket, menjelaskan bagaimana fungsinya dapat digunakan, dan mencantumkan semua dependensi pada paket lain

***Visual form* (Bentuk visual):** Tampilan dari visualisasi data yang memberikan struktur dan daya tarik estetika

**Visualisasi:** (Lihat *Data visualization*)

**VLOOKUP:** Suatu function di spreadsheet yang secara vertikal mencari nilai tertentu dalam kolom untuk menampilkan informasi yang sesuai

W

**WHERE:** Bagian kueri yang menentukan kriteria yang harus dipenuhi oleh data yang diminta

***Wide data* (Data lebar):** Sebuah *dataset* di mana setiap subjek data memiliki satu baris dengan beberapa kolom yang berisi nilai-nilai dari berbagai atribut subjek

**WITH:** Sebuah klausa dalam SQL untuk membuat tabel sementara yang dapat dikueri beberapa kali

**World Health Organization (WHO):** Sebuah organisasi yang peran utamanya adalah mengarahkan dan mengkoordinasikan kesehatan internasional dalam sistem Perserikatan Bangsa-Bangsa

X

**X-axis (Sumbu X):** Garis horizontal pada sebuah bagan yang biasanya ditempatkan di bagian bawah, dan sering digunakan untuk merepresentasikan skala waktu dan kategori diskrit.

Y

**Y-axis (Sumbu Y):** Garis vertikal pada sebuah bagan yang biasanya ditempatkan di sebelah kiri, yang sering digunakan untuk merepresentasikan frekuensi dan variabel numerik lainnya.