A

**Abertura:** Aspecto da ética de dados que promove o acesso, uso e compartilhamento livre dos dados.

**Agenda**: Lista de compromissos agendados.

**Agregação de dados:** Processo de reunir dados de diferentes fontes e combiná-los em uma única coleção sintetizada.

**Agregação:** Processo de coletar ou reunir diferentes partes em um todo.

**Algoritmo:** Processo ou conjunto de regras seguidas para uma tarefa específica.

**Amostra:** Em Data Analytics, um segmento de uma população que representa a população como um todo.

**Amostragem aleatória:** Uma maneira de selecionar uma amostra de uma população para que cada tipo possível da amostra tenha a mesma chance de ser escolhido.

**Amostragem não enviesada:** Quando a amostra da população medida representa a população como um todo.

**Análise de dados:** A coleta, transformação e organização de dados para tirar conclusões, fazer previsões e orientar a tomada de decisão informada.

**Análise de lacunas:** Método para examinar e avaliar o estado atual de um processo, a fim de identificar oportunidades para melhoria no futuro.

**Análise:** Processo utilizado para entender os dados coletados.

**Analista de dados:** Alguém que coleta, transforma e organiza dados para tirar conclusões, fazer previsões e orientar a tomada de decisão informada.

**Aninhado:** Código que realiza uma função específica e está contido dentro do código que realiza uma função mais ampla.

**Anonimização dos dados:** Processo de proteger dados privados ou sensíveis das pessoas, eliminando informações que possam identificá-los.

**Anotação:** Texto que explica brevemente os dados ou ajuda a focalizar o público-alvo em um aspecto específico dos dados em uma visualização.

**Argumento (R):** Informações necessárias para executar uma função em R.

**Arquivo CSV (valores separados por vírgula):** Arquivo de texto delimitado que utiliza vírgulas para separar os valores.

**Arquivo de áudio:** Gravação de áudio digitalizada, geralmente em formato MP3, AAC ou outro formato comprimido.

**Arquivo de log:** Arquivo gerado pelo computador que registra eventos dos sistemas operacionais e outros programas de software.

**Arquivo de vídeo:** Coleção de imagens, arquivos de áudio e outros dados geralmente codificados em formato comprimido, como MP4, MV4, MOV, AVI, ou FLV.

**Árvore de decisão**: Ferramenta que ajuda os analistas a tomar decisões sobre pontos críticos de uma visualização.

**Atribuição de alias:** Nomeação temporária de uma tabela ou coluna em uma consulta, a fim de facilitar a leitura e gravação.

**Atributo**: Característica ou qualidade dos dados usada para rotular uma coluna em uma tabela.

**Atributos pré-atencionais:** Elementos de uma visualização de dados que um público-alvo reconhece automaticamente sem esforço consciente.

**AVERAGE**: Função de planilha que retorna a média dos valores de um intervalo selecionado.

**AVERAGEIF:** Função de planilha que retorna a média de todos os valores de células de determinado intervalo que atendem a uma condição especificada.

B

**Banco de dados normalizado:** Banco de dados em que são armazenados apenas dados relacionados em cada tabela.

**Banco de dados relacional:** Banco de dados que contém uma série de tabelas que podem ser conectadas para formar relacionamentos.

**Banco de dados:** Coleção de dados armazenados em um sistema de computador.

**Biblioteca:** Diretório contendo todos os pacotes instalados por um analista de dados.

**Big data:** Conjuntos de dados grandes e completos, que envolvem longos períodos e permitem aos analistas de dados tratarem de problemas de negócios de longo alcance.

**Bordas**: Linhas que podem ser adicionadas ao redor de duas ou mais células em uma planilha.

C

**C#:** Linguagem de programação orientada a objetos que é usada para criar jogos e aplicativos móveis na plataforma de desenvolvedor de código aberto .NET.

**C++:** Extensão da linguagem de programação C que também é usada para criar jogos de console, como aqueles para Xbox.

**Cabeçalho:** Primeira linha em uma planilha que identifica o tipo de dados em cada coluna.

**Cálculo:** Ramo da matemática que envolve o estudo das taxas de mudança e as mudanças entre valores que estão relacionados por uma função.

**Campo calculado:** Novo campo em uma tabela dinâmica que realiza determinados cálculos com base nos valores de outros campos.

**Campo:** Uma única informação de uma linha ou coluna de uma planilha; em uma tabela de dados, normalmente é uma coluna da tabela.

**Canal:** Aspecto visual ou variável que representa as características dos dados em uma visualização.

**CASE:** Instrução de SQL que retorna registros que atendem às condições ao incluir uma instrução if/then em uma consulta.

**CAST:** Função de SQL que converte dados de um tipo de dados para outro.

**Causa raiz:** Razão pela qual um problema ocorre

**Causalidade:** Quando uma ação leva diretamente a um resultado, como uma relação causa-efeito.

**Chave estrangeira:** Campo na tabela do banco de dados que é uma chave primária em outra tabela (Ver chave primária).

**Chave primária:** Identificador em um banco de dados que faz referência a uma coluna na qual cada valor é exclusivo (ver chave estrangeira).

**Ciclo de vida dos dados:** Sequência de etapas pelas quais os dados passam, e que incluem planejamento, coleta, gestão, análise, arquivamento e destruição.

**Ciência de dados:** Área de pesquisa que utiliza dados brutos para criar novas maneiras de modelar e entender o desconhecido.

**Classificação:** Processo de organizar os dados em uma ordem significativa para torná-los mais fáceis de entender, analisar e visualizar.

**Classificação:** Sistema para posicionar valores de um conjunto de dados dentro de uma escala de conquistas ou status.

**Classificar intervalo:** Função em um menu da planilha que classifica um intervalo especificado e preserva as células fora do intervalo.

**Classificar planilha:** Função em um menu da planilha que ordena todos os dados pela classificação de uma coluna classificada em específico e mantêm os dados juntos ao longo das linhas.

**Cluster:** Coleção de pontos de dados em uma visualização de dados com valores similares.

**COALESCE:** Função de SQL que retorna valores não nulos em uma lista.

**Codificação:** Processo de escrever instruções para um computador, na sintaxe de uma linguagem de programação específica.

**Código fonte aberto:** Código disponível gratuitamente, que pode ser modificado e compartilhado pelas pessoas que o utilizam.

**Combinação de dados:** Método do Tableau que combina dados a partir de múltiplas fontes de dados.

**Compatibilidade:** A forma como dois ou mais conjuntos de dados são capazes de trabalhar juntos.

**Composição de dados:** Processo de combinar partes individuais em uma visualização e mostrar essas partes juntas como um todo.

**Comprimento de campo:** Ferramenta para determinar quantos caracteres podem ser digitados em um campo da planilha.

**Comprimento:** Número de caracteres em uma string de texto.

**CONCAT:** Função de SQL que adiciona strings para criar strings de texto novas que podem ser usadas como chaves exclusivas.

**CONCATENATE:** Função de planilha que une duas ou mais strings de texto.

**Conjunto de dados:** Coleção de dados que pode ser manipulada ou analisada como uma unidade única.

**Consentimento:** Aspecto da ética de dados que presume o direito que um indivíduo tem de saber como e por que seus dados pessoais serão usados antes de concordar em fornecê-los.

**Consistência:** Grau em que os dados são repetíveis de diferentes pontos de entrada ou coleta.

**Consulta externa:** Instrução de SQL contendo uma subconsulta.

**Consulta interna:** Subconsulta de SQL que está dentro de outra instrução de SQL.

**Consulta:** Solicitação de dados ou informações de um banco de dados.

**Contexto:** Condição em que algo existe ou acontece.

**Controle de acesso:** Funcionalidades como proteção por senha, permissões do usuário e criptografia, usadas para proteger uma planilha.

**Convenções de nomenclatura:** Diretrizes consistentes que descrevem o conteúdo, a data de criação e a versão de um arquivo em seu nome.

**CONVERT:** Função de SQL que altera a unidade de medida de um valor nos dados.

**Cookie:** Arquivo pequeno armazenado em um computador, contendo informações sobre seus usuários.

**Correlação:** A medida do grau em que duas variáveis mudam em relação uma à outra.

**COUNT DISTINCT:** Função de SQL que retorna somente os valores distintos em um determinado intervalo.

**COUNT**: Função de planilha que conta o número de células dentro de um intervaloque atende a uma condição específica.

**COUNTA:** Função de planilha que conta o número total de valores dentro de um intervalo específico que atende aos critérios especificados.

**COUNTIF:** Função de planilha que retorna o número de células em um intervalo que corresponde a um valor específico.

**CRAN (Rede Abrangente de Arquivos R) (R):** Arquivo online com pacotes de R, código fonte, manuais e documentação.

**CREATE TABLE:** Cláusula de SQL que adiciona uma tabela temporária a um banco de dados que pode ser usado por várias pessoas.

**CSS (folhas de estilo em cascatas):** Linguagem de programação usada para o design de páginas da web que controla os elementos gráficos e a apresentação da página.

D

**Dados abertos:** Dados que estão disponíveis para o público

**Dados amplos:** Conjunto de dados em que cada assunto possui uma única linha com colunas múltiplas a fim de conter os valores de diversos atributos do assunto.

**Dados booleanos:** Tipo de dado com apenas dois valores possíveis, geralmente verdadeiro ou falso.

**Dados contínuos:** Dados que são medidos e que podem ter quase qualquer valor numérico.

**Dados de terceiros:** Dados fornecidos de fontes externas que não os coletaram diretamente.

**Dados desatualizados:** Quaisquer dados que tenham sido substituídos por informações mais recentes e mais precisas.

**Dados dinâmicos:** Dados que são atualizados automaticamente.

**Dados discretos:** Dados contados e que têm um número limitado de valores.

**Dados duplicados:** Qualquer registro que, inadvertidamente, compartilhe dados com outro registro.

**Dados estáticos:** Dados que não mudam depois de registrados

**Dados estruturados:** Dados organizados em determinado formato, como em linhas e colunas.

**Dados externos:** Dados que existem e são gerados fora de uma organização.

**Dados incompletos:** Dados onde estão faltando campos importantes.

**Dados inconsistentes:** Dados que utilizam formatos diferentes para representar a mesma coisa

**Dados incorretos/imprecisos:** Dados completos mas imprecisos.

**Dados internos:** Dados que residem nos próprios sistemas da empresa.

**Dados limpos:** Dados completos, corretos e relevantes para o problema a ser resolvido.

**Dados longos:** Conjunto de dados no qual cada linha é um ponto temporal por assunto, de modo que cada assunto possui dados em múltiplas linhas.

**Dados não estruturados:** Dados que não são organizados de modo a serem facilmente identificados.

**Dados nominais:** Tipo de dados qualitativos categorizados sem uma ordem definida.

**Dados ordinais:** Dados qualitativos com uma ordem ou escala definida.

**Dados primários:** Dados coletados por um indivíduo ou grupo usando seus próprios recursos

**Dados qualitativos:** Medida subjetiva e explicativa de uma qualidade ou característica.

**Dados quantitativos:** Medida específica e objetiva, como um número, quantidade ou intervalo.

**Dados secundários:** Dados coletados por um grupo diretamente de seu público-alvo, e vendidos posteriormente.

**Dados sujos:** Dados incompletos, incorretos, ou irrelevantes para o problema a ser resolvido.

**Dados tipo string:** Sequência de caracteres e pontuações que contêm informações textuais (também denominados dados tipo texto).

**Dados tipo texto:** Sequência de caracteres e pontuações que contêm informações textuais (também denominados dados tipo string).

**Dados:** Uma coleção de fatos.

**Data Analytics:** Ciência dos dados.

**Data frame:** Coleção de colunas contendo dados, semelhante a uma planilha ou tabela de SQL.

**DATEDIF:** Função de planilha que calcula o número de dias, meses ou anos entre duas datas.

**Delimitador:** Caractere que indica o início ou o fim de um item de dados.

**Design de dados:** O modo como a informação é organizada.

**Design Thinking:** Processo usado para resolver problemas complexos de uma forma centrada no usuário.

**Diagrama de Gantt:** Visualização de dados que mostra a duração de eventos ou atividades em uma linha do tempo.

**Diagrama em caixa:** Visualização de dados que mostra a distribuição dos valores ao longo de um eixo X.

**DISTINCT:** Palavra-chave que é adicionada a uma instrução SELECT SQL para extrair apenas entradas não duplicadas.

**Domínio do problema:** Área de análise que engloba todas as atividades afetando um problema ou afetadas por ele.

**dplyr (R):** Pacote de R no tidyverse que oferece um conjunto consistente de funções para completar tarefas comuns de manipulação de dados.

**DROP TABLE:** Cláusula de SQL que remove uma tabela temporária de um banco de dados.

E

**Ecossistema de dados:** Os vários elementos que interagem entre si para produzir, gerenciar, armazenar, organizar, analisar e compartilhar dados.

**Eixo X:** Linha horizontal de um gráfico, geralmente posicionada na base, que frequentemente é usada para representar escalas de tempo e categorias discretas.

**Eixo Y:** Linha vertical de um gráfico, geralmente posicionada à esquerda, que frequentemente é usada para representar frequências e outras variáveis numéricas.

**Elemento de dados:** Uma informação dentro de um conjunto de dados.

**Ênfase:** Princípio do design que consiste em arranjar os elementos visuais para focar a atenção do público-alvo nas informações importantes em uma visualização de dados.

**Engajamento:** Capturar e reter o interesse e a atenção de alguém durante uma apresentação de dados.

**Engenheiro de dados:** Profissional que transforma os dados em um formato útil para análise e lhe dá uma infraestrutura confiável.

**Equação:** Cálculo que envolve adição, subtração, multiplicação ou divisão (também chamada de expressão matemática).

**Equilíbrio:** Princípio do design que consiste em criar apelo estético e clareza em uma visualização de dados, distribuindo uniformemente os elementos visuais.

**Escopo de trabalho (SOW):** Delineamento acordado das tarefas a serem realizadas durante um projeto.

**Especialista em armazenamento de dados:** Profissional que desenvolve processos e procedimentos para efetivamente armazenar e organizar dados.

**Esquema:** Uma maneira de descrever de que forma algo, como os dados, é organizado.

**Estatística:** Estudo de como coletar, analisar, resumir, e apresentar dados.

**Estética (R):** Propriedade visual de um objeto em um gráfico.

**Estratégia de dados:** Gestão de pessoas, processos e ferramentas usadas na análise de dados

**Estrutura de dados:** Formato para organizar e armazenar os dados.

**Estrutura:** O contexto de uma apresentação precisa criar conexões lógicas que se liguem à tarefa de negócios e à métrica

**Estudo de caso:** Uma maneira comum para os empregadores avaliarem as habilidades de trabalho e obterem insight sobre como um candidato trata dos desafios comuns relacionados aos dados.

**Ética de dados:** Padrões bem fundamentados de certo e errado, que ditam como os dados são coletados, compartilhados e usados.

**Ética:** Padrões fundamentados de certo e errado, que prescrevem o que os humanos devem fazer, geralmente em termos de direitos, obrigações, benefícios sociais, imparcialidade ou virtudes específicas.

**Exibição de círculos:** Visualização de dados que mostra a força comparativa nos dados.

**Expressão matemática:** Cálculo que envolve adição, subtração, multiplicação ou divisão (também chamado de equação).

**Expressão regular (RegEx):** Regra que diz que os valores em uma tabela devem corresponder a um padrão prescrito.

F

**Facetas (R):** Série de funções que divide os dados em subconjuntos em uma matriz de painéis.

**Fator (R):** Objeto que armazena dados categóricos nos quais os valores dos dados são limitados e geralmente se baseiam em um grupo finito, como país ou ano.

**Filtragem:** Processo de exibir apenas os dados que atendem a um critério especificado, ocultando o restante.

**Filtro de painel:** Ferramenta para mostrar apenas os dados que satisfazem um critério específico enquanto esconde o resto

**Flutuante:** Número que contém um decimal.

**Fonte de dados bons:** Fonte de dados que é confiável, original, abrangente, atual e citada (ROCCC, na sigla em inglês).

**Fonte de dados inválidos:** Fonte de dados que não é confiável, original, abrangente, atual e citada (ROCCC, na sigla em inglês).

**Forma visual:** Aparência de uma visualização de dados que dá a ela estrutura e atrativo estético.

**Formatação condicional:** Ferramenta de planilha que muda a forma como as células aparecem quando os valores atendem a condições específicas

**Fórmula:** Conjunto de instruções usadas para realizar um cálculo usando os dados de uma planilha.

**Foto digital:** Uma imagem eletrônica ou computadorizada geralmente em formato BMP ou JPG.

**FROM:** Seção de uma consulta que indica de qual(is) tabela(s) para extrair os dados.

**Função (R):** Um corpo de código reutilizável para executar tarefas específicas em R.

**Função aninhada:** Função totalmente contida em outra função.

**Função matemática**: Função usada como parte de uma fórmula matemática.

**Função:** Comando predefinido que executa automaticamente um processo ou tarefa específica usando os dados de uma planilha.

**Fusão de dados:** Processo de combinar dois ou mais conjuntos de dados em um único conjunto de dados.

**Fusão:** Acordo que une duas organizações em uma única nova.

**FWF (arquivo com largura ajustada):** Texto com um formato específico, que permite salvar dados textuais de modo organizado.

G

**Geolocalização:** Localização geográfica de uma pessoa ou de um dispositivo, por meio da informação digital.

**Geom (R):** Objeto geométrico usado para representar dados.

**ggplot2 (R):** Pacote de R no tidyverse que cria uma variedade de visualizações de dados aplicando propriedades visuais diferentes às variáveis de dados em R.

**Governança dos dados:** Processo que garante a gestão formal dos ativos de dados de uma empresa.

**Gráfico de área:** Visualização de dados que usa pontos isolados dos dados para mudar uma variável conectada por meio de uma linha contínua com uma área preenchida por baixo.

**Gráfico de barras:** Visualização de dados que usa o tamanho para contrastar e comparar dois ou mais valores.

**Gráfico de bolhas em pacotes:** Visualização de dados que mostra dados em círculos agrupados.

**Gráfico de bolhas**: Visualização de dados que mostra pontos isolados dos dados como bolhas, comparando valores numéricos pelo seu tamanho relativo.

**Gráfico de combinação:** Visualização de dados que combina mais de um tipo de visualização.

**Gráfico de dispersão:** Visualização de dados que representa os relacionamentos entre diferentes variáveis com pontos isolados de dados sem uma linha que os conecta.

**Gráfico de distribuição:** Visualização de dados que mostra a frequência de vários resultados em uma amostra.

**Gráfico de linhas:** Visualização de dados que usa uma ou mais linhas para mostrar mudanças ou alterações nos dados ao longo do tempo.

**Gráfico de marcador:** Visualização de dados que mostra os dados como um gráfico de barras horizontais se movendo até um valor desejado.

**Gráfico de medidor:** Visualização de dados que mostra um único resultado dentro um intervalo progressivo de valores.

**Gráfico de pizza:** Visualização de dados que usa segmentos de um círculo para representar as proporções de cada categoria de dados comparada com o todo.

**Gráfico de rosca:** Visualização de dados em que os seguimentos de um anel representam os valores dos dados somados a um todo.

**Gráfico dinâmico:** Gráfico criado a partir dos campos de uma tabela dinâmica.

**Gráfico:** Representação gráfica dos dados de uma planilha.

**Gráficos de colunas:** Visualização de dados que usa pontos de dados individuais para uma variável dinâmica, representada em forma de colunas verticais.

**GROUP BY:** Cláusula de SQL que agrupa linhas que têm os mesmos valores de uma tabela em linhas de resumo.

H

**Habilidades analíticas:** Qualidades e características associadas ao uso de fatos para solucionar problemas.

**Habilidades comportamentais:** Características e comportamentos não técnicos relacionados à maneira como as pessoas trabalham.

**Habilidades transferíveis:** Habilidade e qualidades que podem ser transferidas de um trabalho a outro ou de um setor a outro.

**HAVING:** Cláusula de SQL que adiciona um filtro a uma consulta em vez da tabela subjacente que só pode ser usada com funções de agregação.

**head() (R):** Função de R que oferece uma pré-visualização dos nomes das colunas e das primeiras linhas de um conjunto de dados.

**Hipótese:** Uma teoria que se pode tentar provar ou refutar com dados.

**Histograma:** Visualização de dados que mostra com que frequência os valores de dados se enquadram em determinados intervalos.

**História:** Narrativa de uma apresentação de dados que a torna interessante e significativa.

**HTML5:** Linguagem de programação que fornece estrutura para páginas da web e conecta à plataformas de hospedagem.

I

**IDE (Ambiente de desenvolvimento integrado):** Aplicativo de software que reúne todas as ferramentas que um analista de dados pode querer usar em um único lugar.

**Imparcialidade:** Qualidade da análise de dados que não cria ou reforça um viés.

**INNER JOIN:** Função de SQL que retorna registros com valores correspondentes em ambas as tabelas.

**Instrução condicional:** Declaração de que caso haja uma determinada condição, um determinado evento deve ocorrer.

**Integridade dos dados:** A precisão, integridade, consistência e confiabilidade dos dados ao longo de seu ciclo de vida.

**Integridade:** Grau em que os dados contêm todos os componentes ou medidas desejados.

**Interoperabilidade dos dados:** Capacidade de integrar dados de múltiplas fontes, que é um fator importante que leva ao uso bem-sucedido de dados abertos entre empresas e governos.

**Intervalo de confiança:** Intervalo de valores que transmite a probabilidade de uma estimativa estatística refletir a população.

**Intervalo de dados:** Valores numéricos que se situam entre valores máximos e mínimos predefinidos.

**Intervalo:** Conjunto de duas ou mais células em uma planilha.

J

**Java:** Linguagem de programação amplamente usada para criar aplicativos da web que podem ser executados em vários ambientes.

**JOIN:** Função de SQL que é usada para combinar linhas de duas ou mais tabelas com base em uma coluna relacionada.

L

**LEFT JOIN:** Função de SQL que retornará todos os registros da tabela à esquerda e somente os registros correspondentes da tabela à direita.

**LEFT:** Função que retorna um determinado número de caracteres do lado esquerdo de uma string de texto.

**Legenda:** Ferramenta que identifica o significado de vários elementos em uma visualização de dados.

**LEN:** Função que retorna o comprimento de uma string de texto, contando o número de caracteres que ela contém.

**LIMIT:** Cláusula de SQL que especifica o número máximo de registros retornados em uma consulta.

**Linguagem de consulta estruturada:** Linguagem de programação computacional usada para se comunicar com um banco de dados.

**Linguagem de consulta:** Linguagem de programação computacional usada para se comunicar com um banco de dados.

**Linguagem de programação:** Sistema de palavras e símbolos usados para escrever instruções que os computadores seguem.

**Linha de suavização (R):** Linha na visualização de dados que utiliza a suavização para representar uma tendência.

**Lista:** Vetor cujos elementos podem ser de qualquer tipo.

**Localizar e substituir:** Ferramenta que localiza um determinado termo de pesquisa e o substitui por outro termo.

**Log de mudanças:** Arquivo que contém uma lista em ordem cronológica de todas as modificações feitas no projeto.

M

**Manipulação de dados:** Processo de mudança de dados para torná-los mais organizados e mais fáceis de ler.

**Mapa de calor:** Visualização de dados que usa contrastes de cor para comparar categorias em um conjunto de dados.

**Mapa de densidade:** Visualização de dados que representa concentrações, com cores representando o número ou frequência de pontos de dados em uma dada área em um mapa.

**Mapa de símbolos:** Visualização de dados que mostra uma marca sobre uma longitude e uma latitude determinadas.

**Mapa preenchido:** Visualização de dados que colore áreas em um mapa com base em medidas ou dimensões.

**Mapa:** Visualização de dados que organiza os dados geograficamente.

**Mapeamento (R):** Processo de equivalência de uma variável específica em um conjunto de dados com uma estética específica.

**Mapeamento dos dados:** Processo de correspondência de campos de uma fonte de dados para outra

**Marca:** Objeto visual em uma visualização de dados, tal como um ponto, linha ou forma.

**Margem de erro:** Quantidade máxima que se espera que os resultados da amostra sejam diferentes dos da população real.

**Margem de lucro:** Porcentagem que indica quantos centavos de lucro foram gerados para cada dólar de venda.

**MATCH:** Função de planilha usada para localizar a posição de determinado valor de pesquisa.

**Matriz unidimensional:** Coleção de valores em células da planilha

**Matriz:** Coleção bidimensional de elementos de dados com linhas e colunas

**MAX:** Função de planilha que calcula o maior valor numérico de um intervalo de células.

**MAXIFS:** Função de planilha que retorna o valor máximo de determinado intervalo que atende a uma condição especificada.

**Mentalidade técnica:** Capacidade de dividir itens em etapas ou partes menores e trabalhar com elas de maneira ordenada e lógica.

**Mentor:** Alguém que partilha seu conhecimento, suas habilidades e sua experiência para ajudar alguém a crescer profissionalmente e pessoalmente.

**Meta métrica:** Meta mensurável definida por uma empresa e é avaliada com o uso de métricas.

**Metadados administrativos:** Metadados que indicam a fonte técnica de um ativo digital.

**Metadados descritivos:** Metadados que descrevem uma parte dos dados e que podem ser usados para identificá-la posteriormente.

**Metadados estruturais:** Metadados que indicam como uma parte dos dados é organizada e se faz parte de uma ou mais de uma compilação de dados.

**Metadados:** Dados a respeito de dados.

**Método McCandless:** Método para apresentar visualizações de dados que passa de informações gerais para informações específicas

**Metodologia SMART:** Ferramenta para determinar a eficácia de uma pergunta, verificando se ela é específica, mensurável, orientada para a ação, relevante e limitada pelo tempo

**Métrica de negócios:** Um padrão de medição utilizado para resolver uma tarefa de negócios

**Métrica:** Tipo de dados quantificável e isolado que pode ser usado para medição.

**MID:** Função que retorna um segmento do meio de uma string de texto.

**Mídias sociais:** Websites e aplicativos por meio dos quais os usuários criam e compartilham conteúdo ou participam de redes sociais.

**MIN**: Função de planilha que retorna o menor valor numérico de um intervalo de células.

**MINIFS:** Função de planilha que retorna o valor mínimo de um determinado intervalo que atende a uma condição especificada.

**Modelo de dados:** Ferramenta para organizar os elementos dos dados e como eles se relacionam entre si.

**Modelo mental:** Linha de pensamento de um analista de dados e a forma como ele aborda um problema.

**Módulo:** Operador (%) que retorna o resto quando um número é dividido por outro.

**Moeda de troca:** O aspecto da ética de dados que presume os indivíduos devem estar cientes das transações financeiras resultantes do uso de seus dados pessoais e da escala dessas transações.

**Movimento:** Princípio do design que consiste em organizar os elementos visuais para guiar os olhos do público-alvo de uma parte da visualização de dados para outra.

**mutate() (R):** Função de R que faz alterações a uma estrutura de dados, separando e mesclando colunas ou criando novas variáveis.

N

**Narrativa:** (Ver: história)

**Networking:** Construção de relacionamentos por meio de encontros presenciais e online com outras pessoas.

**Nível de confiança:** A probabilidade de que o tamanho de uma amostra reflita com precisão a maior população.

**Notebook:** Ambiente interativo e editável de programação para criação de relatórios de dados e para exibição das habilidades dos dados.

**Nulo:** Indicação de que um valor não existe em um conjunto de dados.

**Nuvem:** Lugar para armazenar os dados online, em vez de no disco rígido de um computador.

O

**Obrigatório:** Valor de dados que não pode ser deixado em branco ou vazio.

**Observação:** Atributos que descrevem um dado contido em uma linha de uma tabela.

**Operador aritmético:** Operador usado para executar operações matemáticas básicas, como adição, subtração, multiplicação, e divisão.

**Operador de atribuição:** Operador usado para atribuir valores a variáveis e vetores.

**Operador lógico:** Operador que retorna um tipo de dado lógico.

**Operador relacional:** Operador usado para comparar valores (também chamado de comparador).

**Operador:** Símbolo que nomeia a operação ou o cálculo a ser realizado.

**Ordem das operações:** Uso de parênteses para agrupar os valores das planilhas para esclarecer a ordem em que as operações devem ser realizadas.

**ORDER BY:** Cláusula de SQL que classifica os resultados retornados em uma consulta.

**Organização Mundial da Saúde:** Organização cujo principal papel é dirigir e coordenar a saúde internacional dentro do sistema da Organização das Nações Unidas.

**OUTER JOIN:** Função de SQL que combina as funções RIGHT e LEFT JOIN para retornar todos os registros correspondentes em ambas as tabelas.

P

**Pacote (R):** Unidade de código R reproduzível.

**Padrão:** Princípio do design que consiste em usar elementos visuais similares para demonstrar tendências e relacionamentos em uma visualização de dados.

**Painel:** Ferramenta que monitora dados recebidos em tempo real.

**Paleta divergente de cores:** Tema colorido que mostra dois intervalos de valores dos dados usando duas matizes diferentes, com a intensidade da cor representando a magnitude dos valores

**Partes interessadas:** Pessoas que investem tempo e recursos em um projeto e estão interessadas em seu resultado.

**Patrocinador:** Profissional comprometido a dar suporte ao avanço da carreira de outro profissional

**Pensamento analítico:** Processo de identificar e definir um problema e, em seguida, resolvê-lo usando dados de maneira organizada e passo a passo.

**Pensamento estruturado:** Processo de reconhecer o problema ou situação atual, organizar as informações disponíveis, revelar lacunas e oportunidades e identificar as opções.

**Pergunta com tempo definido:** Pergunta que especifica um período de tempo a ser estudado.

**Pergunta específica:** Pergunta simples, significativa e focada em um único tópico ou em algumas ideias relacionadas entre si.

**Pergunta mensurável:** Pergunta cujas respostas podem ser quantificadas e avaliadas.

**Pergunta orientada para a ação:** Pergunta cujas respostas levam a uma alteração.

**Pergunta parcial:** Pergunta que faz suposições ou é difícil de se responder de forma honesta.

**Pergunta relevante:** Pergunta significativa para o problema a ser solucionado.

**Pergunta sugestiva:** Pergunta que direciona as pessoas para uma determinada resposta.

**PHP (pré-processador de hipertexto):** Linguagem de programação para desenvolvimento de aplicativos da web.

**Pipe (R):** Ferramenta de R para expressar uma sequência de múltiplas operações, representada por “%>%”.

**Pixel:** Nas imagens digitais, uma pequena área de iluminação em uma tela de exibição que, ao ser combinada com outras áreas adjacentes, forma uma imagem digital.

**Planilha:** Uma planilha digital.

**Poder estatístico:** Probabilidade de um teste de significância reconhecer um efeito que está presente.

**População:** Em Data Analytics, significa todos os valores de dados possíveis em um conjunto de dados.

**Portfólio:** Coleção de materiais que pode ser compartilhada com empregadores potenciais.

**Precisão:** Grau em que os dados estão de acordo com a entidade real que está sendo medida ou descrita.

**Preenchimento automático:** Boxe no canto inferior à direita na célula selecionada da planilha que pode ser arrastado para células próximas a fim de continuar uma instrução.

**Privacidade dos dados:** Preservar as informações de um titular de dados sempre que ocorrer uma transação com os dados.

**Processo de análise de dados:** As seis etapas de perguntar, preparar, processar, analisar, compartilhar e agir, cujo propósito é obter insights que orientem a tomada de decisão informada.

**Processo de validação de dados:** Processo de verificação e reavaliação da qualidade dos dados para que sejam completos, precisos, seguros e consistentes.

**Programação de computadores:** Processo de dar instruções para um computador para realizar uma ação ou conjunto de ações.

**Proporção:** Princípio do design que consiste em usar o tamanho relativo e o arranjo de elementos visuais para demonstrar informações em uma visualização de dados.

**Propriedade:** Aspecto da ética de dados que presume que os indivíduos possuem os dados brutos que fornecem e têm controle primário sobre o seu uso, processamento e compartilhamento.

**Python:** Linguagem de programação para fins gerais.

Q

**Quarteto de Anscombe:** Quatro conjuntos de dados que possuem sumários estatísticos quase idênticos, mas possuem pontos delimitadores diferentes.

R

**R:** Linguagem de programação usada para análise estatística, visualização, e outras análises de dados.

**readr (R):** Pacote de R no tidyverse usado para importar dados.

**Receita:** Valor total de renda gerada pela venda de produtos ou serviços.

**Redundância:** Quando a mesma parte dos dados é armazenada em dois ou mais locais.

**Referência absoluta:** Referência em uma função que é bloqueada, para que as linhas e colunas não sofram alterações se a função for copiada.

**Referência de célula:** Célula ou intervalo de células em uma planilha, que normalmente se usa em fórmulas e funções.

**Reformulação:** Processo de reformular um problema ou desafio e redirecioná-lo para uma resolução possível.

**Registro:** Coleção de dados relacionados em uma tabela de dados, geralmente sinônimo de linha.

**Regulamento Geral de Proteção de Dados (GDPR, na sigla em inglês):** Entidade que atua na criação de políticas na União Europeia, criada para ajudar a proteger as pessoas e seus dados.

**Relatividade:** Processo de considerar as observações em relação ou em proporção a algo mais.

**Relatório:** Coleção estática de dados entregue periodicamente às partes interessadas.

**Remover duplicatas:** Ferramenta de planilha que procura e elimina automaticamente as entradas duplicadas de uma planilha.

**Repetição:** Princípio do design que consiste em repetir elementos visuais para expressar significado em uma visualização de dados.

**Replicação de dados:** Processo de armazenamento de dados em múltiplos locais.

**Repositório de metadados:** Banco de dados criado para armazenar metadados.

**Restrições de dados:** Critérios que determinam se um dado é limpo e válido.

**Retorno sobre o investimento (ROI):** Fórmula que utiliza métricas de investimento e lucro para avaliar o sucesso de um investimento.

**RIGHT JOIN:** Função de SQL que retornará todos os registros da tabela à direita e somente os registros correspondentes da tabela à esquerda.

**RIGHT:** Função que retorna um determinado número de caracteres do lado direito de uma string de texto.

**Ritmo:** Princípio do design que consiste em criar movimento e fluxo em uma visualização de dados para engajar um público-alvo.

**Rótulo:** Texto em uma visualização que identifica um valor ou descreve uma escala.

**Rótulos e anotações (R):** Grupo de funções de R para customizar um gráfico.

**ROUND:** Função de SQL que retorna um número arredondado a determinado número de casas decimais

**Ruby:** Linguagem de programação orientada a objetos para desenvolvimento de aplicativos da web.

S

**Segurança de dados:** Proteger os dados contra acesso não autorizado ou corrupção por meio da adoção de medidas de segurança.

**SELECT INTO:** Cláusula de SQL que copia dados de uma tabela para uma tabela temporária sem adicionar a nova tabela no banco de dados.

**SELECT:** Seção de uma consulta que indica de qual(is) coluna(s) extrair os dados.

**Significância estatística:** Probabilidade de que os resultados da amostra não se devam a um acaso.

**Sintaxe:** Estrutura predeterminada de um idioma que inclui todas as palavras, símbolos e pontuação necessários, bem como sua correta colocação.

**Small data:** Pontos de dados pequenos e específicos, normalmente envolvendo um curto período de tempo, que são úteis para tomar decisões no dia a dia.

**SPLIT:** Função de planilha que divide o texto em torno de um caractere específico e coloca cada fragmento em uma célula nova e separada.

**Spotlightling:** Recurso que faz uma varredura pelos dados para identificar rapidamente os insights mais importantes.

**SQL:** (Ver Linguagem de consulta estruturada)

**Storytelling com dados:** Comunicar o significado de uma base de dados com visuais e uma narrativa que são personalizados para um público-alvo em particular.

**String de texto:** Grupo de caracteres dentro de uma célula, na maioria das vezes composto de letras.

**Suavização (R):** Processo para tornar as visualizações de dados em R mais claras e legíveis.

**Suavização GAM (modelo aditivo generalizado) (R):** Processo para suavizar gráficos com um grande número de pontos.

**Suavização Loess (R):** Processo usado para suavizar gráficos com menos de 1.000 pontos.

**Subconsulta:** Consulta SQL aninhada em uma consulta maior.

**Sublinhados**: Linhas usadas para sublinhar palavras e conectar caracteres de texto.

**SUBSTR:** Função de SQL que extrai uma substring de uma string variável.

**Substring:** Um subconjunto de uma string de texto.

**Subtítulo** Texto que apoia um título ao adicionar contexto e descrição.

**SUM:** Função de planilha que soma os valores de um intervalo de células selecionado.

**SUMIF:** Função de planilha que adiciona dados numéricos com base em uma condição.

**SUMPRODUCT:** Função que multiplica matrizes unidimensionais e retorna a soma desses produtos.

**Swift:** Linguagem de programação para macOS, iOS, watchOS e tvOS.

T

**Tabela de destaque:** Visualização de dados que usa cor e formatação condicional em uma tabela.

**Tabela dinâmica:** Ferramenta de resumo de dados utilizada para selecionar, reorganizar, agrupar, contar, calcular o total ou calcular a média dos dados.

**Tabela temporária:** Tabela de banco de dados que é criada e existe temporariamente em um servidor de banco de dados.

**Tableau:** Plataforma de inteligência e análise de negócios que ajuda as pessoas a visualizar, entender, e tomar decisões com dados.

**Tarefa de negócios:** Pergunta ou problema que a análise de dados resolve para uma empresa.

**Tarefa de resumo:** Tabela usada para resumir informações estatísticas sobre os dados.

**Taxa de resposta estimada**: Número médio de pessoas que normalmente completam uma pesquisa.

**Taxa de rotatividade:** Taxa que mensura a saída voluntária dos funcionários da empresa.

**Teste A/B:** Processo de testar duas variações da mesma página web para determinar qual página é mais bem sucedida em atrair tráfego de usuários e gerar receita.

**Teste de hipótese:** Processo para determinar se uma pesquisa ou experimento tem resultados significativos.

**Texto alternativo:** Texto que fornece uma alternativa ao conteúdo não textual, como imagens e vídeos

**Tibble (R):** Variação simplificada de data frames.

**tidy data (R):** Forma de padronizar a organização dos dados dentro de R.

**tidyr (R):** Pacote de R no tidyverse usado na limpeza de dados para criar dados ordenados.

**Tidyverse (R):** Sistema de pacotes de R com uma filosofia de design comum para manipulação, exploração e visualização de dados.

**Tipo de dado:** Atributo que descreve uma parte dos dados com base em seus valores, sua linguagem de programação ou as operações que pode realizar.

**Tipos de problema:** Os diversos problemas que os analistas de dados encontram, incluindo: categorizar as coisas, descobrir conexões, encontrar padrões, identificar temas, fazer previsões e detectar algo incomum.

**Título:** Texto no topo de uma visualização que comunica os dados que estão sendo apresentados.

**Tomada de decisão inspirada em dados:** Explorar diferentes fontes de dados para descobrir o que elas têm em comum.

**Tomada de decisões com base em dados:** Usar fatos para orientar a estratégia de negócios.

**Transferência de dados:** Processo de cópia de dados de um dispositivo de armazenamento para a memória de um computador ou de um computador para outro.

**Transparência de transações:** Aspecto da ética de dados que presume que todos os algoritmos e atividades de processamento de dados sejam explicáveis e entendidos pelo indivíduo que fornece os dados.

**TRIM:** Função que remove espaços à esquerda, à direita e repetidos nos dados

**TSV (arquivo de valores separados por tabulação):** Arquivo de texto que armazena uma tabela de dados separando colunas de dados por tabulação.

**Typecasting:** Conversão de dados de um tipo para outro.

U

**Único:** Valor que não pode ter um duplicado.

**Unidade:** Princípio do design que consiste em usar elementos visuais que se complementam mutuamente para criar apelo estético e clareza em uma visualização de dados.

**United States Census Bureau:** Agência do Departamento do Comércio dos EUA que atua como o principal provedor de dados de qualidade do país a respeito de seus habitantes e sua economia.

V

**Validação de campo cruzado:** Processo que garante o cumprimento de certas condições para múltiplos campos de dados

**Validação de dados:** Ferramenta para verificar a precisão e a qualidade dos dados.

**Validade:** Grau de conformidade dos dados com as restrições quando são inseridos, coletados ou criados.

**VALUE:** Função de planilha que converte uma string de texto que representa um número a um valor numérico.

**Variável (R):** Uma representação de um valor em R que pode ser armazenado para uso posterior

**Variedade:** Princípio do design que consiste em usar diferentes tipos de elementos visuais em uma visualização de dados para engajar um público-alvo.

**Verificação:** Processo para confirmar que uma tentativa de limpeza de dados foi executada com sucesso e que os dados derivados são precisos e confiáveis.

**Vetor (R):** Grupo de elementos de dados do mesmo tipo armazenados em uma sequência unidimensional em R.

**Viés de amostragem:** Representação excessiva ou deficiente de determinados membros de uma população como resultado de trabalhar com uma amostra que não representa a população como um todo.

**Viés de confirmação:** Tendência de buscar ou interpretar informações de forma a confirmar crenças existentes.

**Viés de dados:** Quando uma preferência a favor ou contra uma pessoa, um grupo de pessoas ou coisas sistematicamente distorce os resultados da análise de dados para uma determinada direção.

**Viés de interpretação:** Tendência a interpretar situações ambíguas de forma positiva ou negativa.

**Viés do experimentador:** A tendência de diferentes pessoas observarem as coisas de modo diferente (Ver viés do observador).

**Viés do observador:** Tendência de diferentes pessoas observarem as coisas de modo diferente (também denominada viés do experimentador).

**Viés:** Uma preferência consciente ou subconsciente que favorece ou desfavorece uma pessoa, um grupo de pessoas ou uma coisa.

**Vinheta (R):** Documentação de um pacote de R que descreve o problema que o pacote pode resolver, explica como suas funções podem ser usadas e lista quaisquer dependências com relação a outros pacotes.

**Visualização de dados:** Representação gráfica dos dados.

**Visualização estática:** Visualização de dados que não muda ao longo do tempo a menos que seja editada.

**Visualização:** (Ver visualização de dados).

**Visualizações dinâmicas:** Visualizações de dados que são interativas ou mudam ao longo do tempo.

**VLOOKUP:** Função de planilha que busca verticalmente um determinado valor em uma coluna para retornar uma informação correspondente.

W

**WHERE:** Seção de uma consulta que especifica critérios aos quais os dados requisitados devem atender.

**WITH:** Cláusula de SQL que cria uma tabela temporária que pode ser consultada múltiplas vezes.