A

**Teste A/B:** O processo de testar duas variações da mesma página da web para determinar qual página tem mais sucesso em atrair tráfego de usuários e gerar receita

**Referência absoluta:**Uma referência dentro de uma função que está bloqueada para que linhas e colunas não sejam alteradas se a função for copiada

**Controle de acesso:** Recursos como proteção por senha, permissões de usuário e criptografia usados ​​para proteger uma planilha

**Precisão:**O grau em que os dados estão em conformidade com a entidade real que está sendo medida ou descrita

**Pergunta orientada para a ação:**Uma pergunta cujas respostas levam à mudança

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

**Estética (R):**Uma propriedade visual de um objeto em um gráfico

**Agenda**: Uma lista de compromissos agendados

**Agregação:**O processo de coletar ou reunir muitas peças separadas em um todo

**Algoritmo:**Um processo ou conjunto de regras seguido para uma tarefa específica

**Aliasing:**Nomear temporariamente uma tabela ou coluna em uma consulta para facilitar a leitura e a gravação

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

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

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

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

**Quarteto de Anscombe:** Quatro conjuntos de dados que possuem estatísticas resumidas quase idênticas, mas contêm valores plotados diferentes

**Gráfico de área**:Uma visualização de dados que usa pontos de dados individuais para uma variável variável conectada por uma linha contínua com uma área preenchida abaixo

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

**Operador aritmético:** Um operador usado para realizar operações matemáticas básicas, como adição, subtração, multiplicação e divisão

**Variedade:**Uma coleção de valores em células da planilha

**Operador de atribuição:**Um operador usado para atribuir valores a variáveis ​​e vetores

**Atributo**: Uma característica ou qualidade de dados usada para rotular uma coluna em uma tabela

**Arquivo de áudio:** Armazenamento de áudio digitalizado geralmente em MP3, AAC ou outro formato compactado

**MÉDIA**: uma função de planilha que retorna uma média dos valores de um intervalo selecionado

**MÉDIASE:** Uma função de planilha que retorna a média de todos os valores de células de um determinado intervalo que atendem a uma condição especificada

B

**Fonte de dados incorreta:**Uma fonte de dados que não é confiável, original, abrangente, atual e citada (ROCCC)

**Equilíbrio:**O princípio de design de criar apelo estético e clareza em uma visualização de dados, distribuindo uniformemente os elementos visuais

**Gráfico de barras:**Uma visualização de dados que usa tamanho para contrastar e comparar dois ou mais valores

**Viés:**Uma preferência consciente ou subconsciente a favor ou contra uma pessoa, grupo de pessoas ou coisa

**Grandes dados:** Conjuntos de dados grandes e complexos, normalmente envolvendo longos períodos de tempo, que permitem aos analistas de dados resolver problemas de negócios de longo alcance

**Dados booleanos:**Um tipo de dados com apenas dois valores possíveis, geralmente verdadeiro ou falso

**Fronteiras**: linhas que podem ser adicionadas ao redor de duas ou mais células em uma planilha

**Gráfico de caixa:** Uma visualização de dados que exibe a distribuição de valores ao longo de um eixo x

**Gráfico de bolhas**: uma visualização de dados que exibe pontos de dados individuais como bolhas, comparando valores numéricos por seu tamanho relativo

**Gráfico com marcadores:** Uma visualização de dados que exibe dados como um gráfico de barras horizontais movendo-se em direção a um valor desejado

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

**Tarefa comercial:**A questão ou problema que a análise de dados resolve para uma empresa

C

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

**C++:**Uma extensão da linguagem de programação C usada para criar jogos de console, como os do Xbox

**Campo calculado:** Um novo campo dentro de uma tabela dinâmica que realiza determinados cálculos com base nos valores de outros campos

**Cálculo:** Um ramo da matemática que envolve o estudo das taxas de variação e das variações entre valores relacionados por uma função.

**CASO:**Uma instrução SQL que retorna registros que atendem às condições, incluindo uma instrução if/then em uma consulta

**Estudo de caso:** Uma maneira comum para os empregadores avaliarem as habilidades profissionais e obterem insights sobre como um candidato aborda desafios comuns relacionados a dados

**ELENCO:** Uma função SQL que converte dados de um tipo de dados para outro

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

**Referência de célula:**Uma célula ou intervalo de células em uma planilha normalmente usada em fórmulas e funções

**Registro de alterações:** Um arquivo contendo uma lista ordenada cronologicamente das modificações feitas em um projeto

**Canal:** Um aspecto visual ou variável que representa características dos dados em uma visualização

**Gráfico:**Uma representação gráfica de dados de uma planilha

**Visualização circular:** Uma visualização de dados que mostra a força comparativa dos dados

**Dados limpos:** Dados completos, corretos e relevantes para o problema que está sendo resolvido

**Nuvem:**Um lugar para manter dados online, em vez de um disco rígido de computador

**Conjunto:** Uma coleção de pontos de dados em uma visualização de dados com valores semelhantes

**COALESCE:**Uma função SQL que retorna valores não nulos em uma lista

**Pedaço de código:**Um trecho de código adicionado em um arquivo R Markdown que é usado para processar, visualizar ou analisar dados

**Codificação:** O processo de escrever instruções em um computador na sintaxe de uma linguagem de programação específica

**Gráfico de colunas**: uma visualização de dados que usa pontos de dados individuais para uma variável variável, representada como colunas verticais

**Gráfico combinado:** Uma visualização de dados que combina mais de um tipo de visualização

**Compatibilidade:** Quão bem dois ou mais conjuntos de dados são capazes de trabalhar juntos

**Completude:** O grau em que os dados contêm todos os componentes ou medidas desejadas

**Programação de computadores:** O processo de dar instruções a um computador para executar uma ação ou conjunto de ações

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

**CONCATENAR:** Uma função de planilha que une duas ou mais sequências de texto

**Formatação condicional:** Uma ferramenta de planilha que altera a forma como as células aparecem quando os valores atendem a condições específicas

**Afirmação condicional:**Uma declaração de que se uma determinada condição for válida, então um determinado evento deve ocorrer

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

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

**Viés de confirmação:**A tendência de procurar ou interpretar informações de uma forma que confirme crenças pré-existentes

**Consentimento:** O aspecto da ética dos dados que pressupõe o direito de um indivíduo saber como e por que seus dados pessoais serão usados ​​antes de concordar em fornecê-los

**Consistência:** O grau em que os dados são repetíveis a partir de diferentes pontos de entrada ou coleta

**Contexto:** A condição em que algo existe ou acontece

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

**CONVERTER:**Uma função SQL que altera a unidade de medida de um valor nos dados

**Biscoito:**Um pequeno arquivo armazenado em um computador que contém informações sobre seus usuários

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

**CONTAR**: uma função de planilha que conta o número de células dentro de um intervaloo atender a uma condição especificada

**CONTAGEM:** Uma função de planilha que conta o número total de valores dentro de um intervalo especificado que atende a critérios especificados

**CONT.SE:**Uma função de planilha que retorna o número de células dentro de um intervalo que corresponde a um valor especificado

**CONTAGEM DISTINTA:**Uma função SQL que retorna apenas valores distintos em um intervalo especificado

**CRAN (rede abrangente de arquivos R) (R):** Um arquivo online com pacotes R, código-fonte, manuais e documentação

**CRIAR A TABELA:** Uma cláusula SQL que adiciona uma tabela temporária a um banco de dados que pode ser usada por várias pessoas

**Validação entre campos:** Um processo que garante que certas condições para vários campos de dados sejam satisfeitas

**CSS (folhas de estilo em cascata):** Uma linguagem de programação usada para design de páginas da web que controla elementos gráficos e apresentação de páginas

**Arquivo CSV (valores separados por vírgula):**Um arquivo de texto delimitado que usa vírgula para separar valores

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

D

**Painel:**Uma ferramenta que monitora dados recebidos em tempo real

**Dados:**Uma coleção de fatos

**Agregação de dados:**O processo de coletar dados de múltiplas fontes e combiná-los em uma coleção única e resumida

**Análise de dados:**A coleta, transformação e organização de dados para tirar conclusões, fazer previsões e impulsionar a tomada de decisões informadas

**Processo de análise de dados:**As seis fases de perguntar, preparar, processar, analisar, compartilhar e agir cujo objetivo é obter insights que conduzam à tomada de decisões informadas

**Analista de informações:**Alguém que coleta, transforma e organiza dados para tirar conclusões, fazer previsões e orientar a tomada de decisões informadas

**Análise de dados:**A ciência dos dados

**Anonimização de dados:**O processo de proteção dos dados privados ou confidenciais das pessoas, eliminando informações de identificação

**Dados tendenciosos:** Quando uma preferência a favor ou contra uma pessoa, grupo de pessoas ou coisa distorce sistematicamente os resultados da análise de dados em uma determinada direção

**Combinação de dados:** Um método do Tableau que combina dados de várias fontes de dados

**Composição dos dados:** O processo de combinar as partes individuais em uma visualização e exibi-las juntas como um todo

**Restrições de dados:** Os critérios que determinam se uma parte dos dados é limpa e válida

**Projeto de dados:**Como as informações são organizadas

**Tomada de decisão baseada em dados:**Usando fatos para orientar a estratégia de negócios

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

**Elemento de dados:**Ainformação em um conjunto de dados

**Engenheiro de dados:**Um profissional que transforma dados em um formato útil para análise e lhes dá uma infraestrutura confiável

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

**Quadro de dados:** Uma coleção de colunas contendo dados, semelhante a uma planilha ou tabela SQL

**Gestão de dados:**Um processo para garantir a gestão formal dos ativos de dados de uma empresa

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

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

**Interoperabilidade de dados:**A capacidade de integrar dados de múltiplas fontes e um fator-chave que leva ao uso bem-sucedido de dados abertos entre empresas e governos

**Ciclo de vida dos dados:**A sequência de estágios que os dados passam, que incluem planejar, capturar, gerenciar, analisar, arquivar e destruir

**Manipulação de dados:** O processo de alteração de dados para torná-los mais organizados e fáceis de ler

**Mapeamento de dados:**O processo de correspondência de campos de uma fonte de dados para outra

**Mesclagem de dados:** O processo de combinar dois ou mais conjuntos de dados em um único conjunto de dados

**Modelo de dados:**Uma ferramenta para organizar elementos de dados e como eles se relacionam entre si

**Dados privados:**Preservar as informações do titular dos dados sempre que ocorrer uma transação de dados

**Intervalo de dados:** Valores numéricos que ficam entre valores máximos e mínimos predefinidos

**Replicação de dados:** O processo de armazenamento de dados em vários locais

**Ciência de dados:** Um campo de estudo que utiliza dados brutos para criarnovas formas de modelar e compreender o desconhecido

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

**Contação de histórias de dados:** Comunicar o significado de um conjunto de dados com recursos visuais e uma narrativa personalizada para um público

**Estratégia de dados:**A gestão das pessoas, processos e ferramentas utilizadas na análise de dados

**Estrutura de dados:**Um formato para organizar e armazenar dados

**Transferência de dados:** O processo de copiar dados de um dispositivo de armazenamento para a memória do computador ou de um computador para outro

**Tipo de dados:**Um atributo que descreve um dado com base em seus valores, sua linguagem de programação ou as operações que ele pode realizar

**Data de validade:** Uma ferramenta para verificar a precisão e a qualidade dos dados

**Processo de validação de dados:** O processo de verificação e nova verificação da qualidade dos dados para que sejam completos, precisos, seguros e consistentes

**Visualização de dados:**A representação gráfica dos dados

**Especialista em armazenamento de dados:** Um profissional que desenvolve processos e procedimentos para armazenar e organizar dados de forma eficaz

**Base de dados:**Uma coleção de dados armazenados em um sistema de computador

**Conjunto de dados:** Uma coleção de dados que podem ser manipulados ou analisados ​​como uma unidade

**DATA SE:** Uma função de planilha que calcula o número de dias, meses ou anos entre duas datas

**Árvore de decisão**: uma ferramenta que ajuda os analistas a tomar decisões sobre recursos críticos de uma visualização

**Delimitador:** Um caractere que indica o início ou o fim de um item de dados

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

**Metadados descritivos:**Metadados que descrevem um dado e podem ser usados ​​para identificá-lo posteriormente

**Pensamento de design:** Um processo usado para resolver problemas complexos de maneira centrada no usuário

**Foto digital:**Uma imagem eletrônica ou baseada em computador, geralmente em formato BMP ou JPG.

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

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

**DISTINTO:** Uma palavra-chave adicionada a uma instrução SQL SELECT para recuperar apenas entradas não duplicadas

**Gráfico de distribuição:** Uma visualização de dados que exibe a frequência de vários resultados em uma amostra

**Paleta de cores divergentes:**Um tema de cores que exibe dois intervalos de valores de dados usando dois matizes diferentes, com intensidade de cor representando a magnitude dos valores

**Gráfico de rosca:**Uma visualização de dados em que segmentos de um anel representam valores de dados que somam um todo

**dplyr (R):**Um pacote R no Tidyverse que oferece um conjunto consistente de funções para completar tarefas comuns de manipulação de dados

**TABELA DE QUEDA:** Uma cláusula SQL que remove uma tabela temporária de um banco de dados

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

**Visualizações dinâmicas:**Visualizações de dados interativas ou que mudam com o tempo

E

**Passo do elevador:** Uma breve declaração descrevendo uma ideia ou conceito

**Ênfase:**O princípio de design de organizar elementos visuais para focar a atenção do público em informações importantes em uma visualização de dados

**Noivado:**Capturar e manter o interesse e a atenção de alguém durante uma apresentação de dados

**Equação:** Um cálculo que envolve adição, subtração, multiplicação ou divisão (também chamado de expressão matemática)

**Taxa de resposta estimada**: o número médio de pessoas que normalmente completam uma pesquisa

**Ética:**Padrões bem fundamentados de certo e errado que prescrevem o que os humanos devem fazer, geralmente em termos de direitos, obrigações, benefícios para a sociedade, justiça ou virtudes específicas

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

F

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

**Fator (R):**Um objeto que armazena dados categóricos onde os valores dos dados são limitados e geralmente baseados em um grupo finito, como país ou ano

**Justiça:** Uma qualidade de análise de dados que não cria ou reforça preconceitos

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

**Comprimento do campo:** Uma ferramenta para determinar quantos caracteres podem ser inseridos em um campo de planilha

**Alça de preenchimento:** Uma caixa no canto inferior direito de uma célula selecionada da planilha que pode ser arrastada pelas células vizinhas para continuar uma instrução

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

**Filtragem:** O processo de mostrar apenas os dados que atendem a um critério especificado enquanto oculta o restante

**Encontre e substitua:**Uma ferramenta que encontra um termo de pesquisa específico e o substitui por outro

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

**Flutuador:** Um número que contém um decimal

**Chave estrangeira:** Um campo dentro de uma tabela de banco de dados que é uma chave primária em outra tabela (consulte a chave primária)

**Fórmula:**Um conjunto de instruções usadas para realizar um cálculo usando os dados de uma planilha

**Estrutura:** O contexto que uma apresentação precisa para criar conexões lógicas que se relacionem com a tarefa e as métricas de negócios

**DE:** A seção de uma consulta que indica de quais tabelas extrair os dados

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

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

**FWF (arquivo de largura fixa):** Um arquivo de texto com formato específico, que permite salvar dados textuais de forma organizada

G

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

**Gráfico de Gantt:** Uma visualização de dados que exibe a duração de eventos ou atividades em uma linha do tempo

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

**Gráfico de medidor**: uma visualização de dados que mostra um único resultado dentro de um intervalo progressivo de valores

**Regulamento Geral de Proteção de Dados da União Europeia (GDPR):**Órgão de elaboração de políticas na União Europeia criado para ajudar a proteger as pessoas e os seus dados

**Geolocalização:**Ta localização geográfica de uma pessoa ou dispositivo por meio de informações digitais

**Geom (R):** O objeto geométrico usado para representar dados

**ggplot2 (R):** Um pacote R no Tidyverse que cria uma variedade de visualizações de dados aplicando diferentes propriedades visuais às variáveis ​​de dados em R

**Boa fonte de dados:**Uma fonte de dados confiável, original, abrangente, atual e citada (ROCCC)

**Agrupar por:** Uma cláusula SQL que agrupa linhas que possuem os mesmos valores de uma tabela em linhas de resumo

H

**TENDO:**Uma cláusula SQL que adiciona um filtro a uma consulta em vez da tabela subjacente que só pode ser usada com funções agregadas

**cabeça() (R):**Uma função R que retorna uma visualização dos nomes das colunas e das primeiras linhas de um conjunto de dados

**Cabeçalho:** A primeira linha em uma planilha que rotula o tipo de dados em cada coluna

**Título:** Texto na parte superior de uma visualização que comunica os dados apresentados

**Mapa de calor:**Uma visualização de dados que usa contraste de cores para comparar categorias em um conjunto de dados

**Tabela de destaque:** Uma visualização de dados que usa formatação condicional e cores em uma tabela

**Histograma:** Uma visualização de dados que mostra com que frequência os valores dos dados se enquadram em determinados intervalos

**HTML (linguagem de marcação de hipertexto):**O conjunto de símbolos ou códigos de marcação usados ​​para criar uma página da web

**HTML5:** Uma linguagem de programação que fornece estrutura para páginas web e se conecta a plataformas de hospedagem

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

**Testando hipóteses:**Um processo para determinar se uma pesquisa ou experimento tem resultados significativos

EU

**IDE (Ambiente de Desenvolvimento Integrado):** Um aplicativo de software que reúne todas as ferramentas que um analista de dados pode querer usar em um único lugar

**Dados incompletos:**Dados que está faltando campos importantes

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

**Dados incorretos/imprecisos:**Dados isso está completo, mas impreciso

**Código embutido:** Código que pode ser inserido diretamente no texto de um arquivo R Markdown

**JUNÇÃO INTERNA :**Uma função SQL que retorna registros com valores correspondentes em ambas as tabelas

**Consulta interna:** Uma subconsulta SQL que está dentro de outra instrução SQL

**Dados internos:** Dados que residem nos próprios sistemas de uma empresa

**Viés de interpretação:** A tendência de interpretar situações ambíguas de forma positiva ou negativa

J.

**Java:**Uma linguagem de programação amplamente utilizada para criar aplicativos web corporativos que podem ser executados em vários clientes

**JUNTAR:**Uma função SQL usada para combinar linhas de duas ou mais tabelas com base em uma coluna relacionada

**Caderno Jupyter:** Um aplicativo da web de código aberto usado para criar e compartilhar documentos que contêm código ativo, equações, visualizações e texto narrativo

K

eu

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

**Etiquetas e anotações (R):** Um grupo de funções R usadas para personalizar um gráfico

**Pergunta principal:**Uma pergunta que direciona as pessoas para uma determinada resposta

**ESQUERDA:** Uma função que retorna um determinado número de caracteres do lado esquerdo de uma string de texto

**ASSOCIAÇÃO À ESQUERDA:**Uma função SQL que retornará todos os registros da tabela esquerda e apenas os registros correspondentes da tabela direita

**Lenda:**Uma ferramenta que identifica o significado de vários elementos em uma visualização de dados

**APENAS:** Uma função que retorna o comprimento de uma string de texto contando o número de caracteres que ela contém

**Comprimento:** O número de caracteres em uma string de texto

**Biblioteca:**Um diretório contendo todos os pacotes instalados de um analista de dados

**LIMITE:** Uma cláusula SQL que especifica o número máximo de registros retornados em uma consulta

**Gráfico de linha:** Uma visualização de dados que usa uma ou mais linhas para exibir mudanças ou mudanças nos dados ao longo do tempo

**Lista:** Um vetor cujos elementos podem ser de qualquer tipo

**Dados ao vivo:**Dados que são atualizados automaticamente

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

**Arquivo de log:**Um arquivo gerado por computador que registra eventos de sistemas operacionais e outros programas de software

**Operador lógico:** Um operador que retorna um tipo de dados lógico

**Dados longos:** Um conjunto de dados em que cada linha representa um ponto no tempo por assunto, de modo que cada assunto tenha dados em várias linhas

M

**Obrigatório:**Um valor de dados que não pode ser deixado em branco ou vazio

**Mapa:** Uma visualização de dados que organiza os dados geograficamente

**Mapeamento (R):** O processo de combinar uma variável específica em um conjunto de dados com uma estética específica

**Margem de erro**: O valor máximo que se espera que os resultados da amostra sejam diferentes daqueles da população real

**Remarcação (R):**Uma sintaxe para formatar arquivos de texto simples

**Marca:** Um objeto visual em uma visualização de dados, como um ponto, linha ou forma

**CORRESPONDER:**Uma função de planilha usada para localizar a posição de um valor de pesquisa específico

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

**Função matemática**: Uma função usada como parte de uma fórmula matemática

**Matriz:**Uma coleção bidimensional de elementos de dados com linhas e colunas

**MÁXIMO:**Uma função que retornao maior númerovalor de uma série de células

**MAXIFOS:** Uma função de planilha que retornao valor máximo de um determinado intervalo que atende a uma condição especificada

**Método McCandless:** Um método para apresentar visualizações de dados que passam de informações gerais para informações específicas

**Pergunta mensurável:**Uma pergunta cujas respostas podem ser quantificadas e avaliadas

**Modelo mental:**O processo de pensamento e a abordagem de um analista de dados para um problema

**Mentor:**Alguém que compartilha conhecimentos, habilidades e experiência para ajudar outra pessoa a crescer tanto profissionalmente quanto pessoalmente

**Fusão:** Um acordo que une duas organizações em uma única e nova

**Metadados:**Dados sobre dados

**Repositório de metadados:**Um banco de dados criado para armazenar metadados

**Métrica:** Um tipo único e quantificável de dados usado para medição

**Meta métrica:**Uma meta mensurável definida por uma empresa e avaliada por meio de métricas

**MEIO:** Uma função que retorna um segmento do meio de uma string de texto

**MÍNIMO**: uma função de planilha que retornao menor numéricovalor de uma série de células

**MINIFS:** Uma função de planilha que retornao valor mínimo de um determinado intervalo que atende a uma condição especificada

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

**Movimento:** O princípio de design de organizar elementos visuais para guiar os olhos do público de uma parte da visualização de dados para outra

**mutação() (R):**Uma função R que faz alterações em um dataframe separando e mesclando colunas ou criando novas variáveis

N

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

**Narrativa:** (Consulte a história)

**Aninhado:**Código que executa uma função específica e está contido em código que executa uma função mais ampla

**Função aninhada:**Uma função que está completamente contida em outra função

**Rede:**Construindo relacionamentos conhecendo pessoas pessoalmente e online

**Dados nominais:** Um tipo de dados qualitativos categorizados sem uma ordem definida

**Banco de dados normalizado:** Um banco de dados no qual apenas dados relacionados são armazenados em cada tabela

**Caderno:** Um ambiente de programação interativo e editável para criar relatórios de dados e demonstrar habilidades em dados

**Nulo:** Uma indicação de que um valor não existe em um conjunto de dados

O

**Observação:**Os atributos que descrevem um dado contido em uma linha de uma tabela

**Viés do observador:**A tendência de pessoas diferentes observarem as coisas de maneira diferente (também chamado de viés do experimentador)

**Dados abertos:**Dados que estão disponíveis ao público

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

**Abertura:**O aspecto da ética de dados que promove o livre acesso, uso e compartilhamento de dados

**Operador:**Um símbolo que nomeia a operação ou cálculo a ser executado

**ORDENAR POR:**Uma cláusula SQL que classifica os resultados retornados em uma consulta

**Ordem de operações:** Usar parênteses para agrupar valores da planilha a fim de esclarecer a ordem em que as operações devem ser realizadas

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

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

**JUNÇÃO EXTERIOR:**Uma função SQL que combina RIGHT e LEFT JOIN para retornar todos os registros correspondentes em ambas as tabelas

**Consulta externa:** Uma instrução SQL contendo uma subconsulta

**Propriedade**: O aspecto da ética de dados que presume que os indivíduos são proprietários dos dados brutos que fornecem e têm controle primário sobre seu uso, processamento e compartilhamento

P

**Pacote (R):** Uma unidade de código R reproduzível

**Gráfico de bolhas compactadas:** Uma visualização de dados que exibe dados em círculos agrupados

**Padrão:** O princípio de design de usar elementos visuais semelhantes para demonstrar tendências e relacionamentos em uma visualização de dados

**PHP (pré-processador de hipertexto):** Uma linguagem de programação para desenvolvimento de aplicações web

**Gráfico de pizza:** Uma visualização de dados que usa segmentos de um círculo para representar as proporções de cada categoria de dados em comparação com o todo

**Tubo (R):** Uma ferramenta em R para expressar uma sequência de múltiplas operações, representada por “%>%”

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

**Tabela dinâmica:**Uma ferramenta de resumo de dados usada para classificar, reorganizar, agrupar, contar, totalizar ou média de dados

**Pixel:** Na imagem digital, uma pequena área de iluminação em uma tela que, quando combinada com outras áreas adjacentes, forma uma imagem digital

**População:** Na análise de dados, todos os valores de dados possíveis em um conjunto de dados

**Portfólio:** Uma coleção de materiais que podem ser compartilhados com potenciais empregadores

**Atributos pré-atentivos:** Os elementos de uma visualização de dados que um público reconhece automaticamente, sem esforço consciente

**Chave primária:** Um identificador em um banco de dados que faz referência a uma coluna na qual cada valor é único (consulte a chave estrangeira)

**Domínio do problema:** A área de análise que abrange todas as atividades que afetam ou são afetadas por um problema

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

**Margem de lucro:** Uma porcentagem que indica quantos centavos de lucro foram gerados para cada dólar vendido

**Linguagem de programação:**Um sistema de palavras e símbolos usado para escrever instruções que os computadores seguem

**Proporção:** O princípio de design de usar o tamanho relativo e a disposição dos elementos visuais para demonstrar informações em uma visualização de dados

**Pitão:** Uma linguagem de programação de uso geral

P

**Dados qualitativos:**Uma medida subjetiva e explicativa de uma qualidade ou característica

**Dados quantitativos:**Uma medida específica e objetiva, como um número, quantidade ou intervalo

**Consulta:**Uma solicitação de dados ou informações de um banco de dados

**Linguagem de consulta:**Uma linguagem de programação de computador usada para se comunicar com um banco de dados

R

**R:** Uma linguagem de programação usada para análise estatística, visualização e outras análises de dados

**Remarcação R:** Um formato de arquivo para criar documentos dinâmicos com R

**Caderno R:** Um documento para executar o código e exibir os gráficos e tabelas que visualizam o código

**Amostragem aleatória:** Uma maneira de selecionar uma amostra de uma população de modo que todos os tipos possíveis de amostra tenham chances iguais de serem escolhidos

**Faixa:**Uma coleção de duas ou mais células em uma planilha

**Classificação:** Um sistema para posicionar valores de um conjunto de dados dentro de uma escala de realização ou status

**ler (R):** Um pacote R no Tidyverse usado para importar dados

**Registro:** Uma coleção de dados relacionados em uma tabela de dados, geralmente sinônimo de linha

**Redundância:** Quando o mesmo dado é armazenado em dois ou mais locais

**Reenquadramento:** O processo de reafirmar um problema ou desafio e, em seguida, redirecioná-lo para uma resolução potencial

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

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

**Operador relacional:**Um operador usado para comparar valores, também conhecido como comparador

**Relatividade:** O processo de considerar observações em relação ou proporção a outra coisa

**Pergunta relevante:**Uma questão que tem significado para o problema a ser resolvido

**Remover duplicatas:** Uma ferramenta de planilha que procura e elimina automaticamente entradas duplicadas de uma planilha

**Repetição:**O princípio de design de repetir elementos visuais para demonstrar o significado em uma visualização de dados

**Relatório:**Uma coleção estática de dados fornecida periodicamente às partes interessadas

**Retorno sobre o investimento (ROI):**Uma fórmula que utiliza as métricas de investimento e lucro para avaliar o sucesso de um investimento

**Receita:** O valor total da receita gerada pela venda de bens ou serviços

**Ritmo:**O princípio de design de criar movimento e fluxo em uma visualização de dados para envolver o público

**CERTO:**Uma função que retorna um determinado número de caracteres do lado direito de uma string de texto

**JUNTE-SE À DIREITA:**Uma função SQL que retornará todos os registros da tabela direita e apenas os registros correspondentes da esquerda

**Causa raiz:** A razão pela qual ocorre um problema

**REDONDO:**Uma função SQL que retorna um número arredondado para um determinado número de casas decimais.

**Rubi:** Uma linguagem de programação orientada a objetos para desenvolvimento de aplicações web

S

**Amostra:**Na análise de dados, um segmento de uma população que representa toda a população

**Viés de amostragem:**Representar sobre ou sub-representar certos membros de uma população como resultado de trabalhar com uma amostra que não é representativa da população como um todo

**Gráfico de dispersão:**Uma visualização de dados que representa relacionamentos entre diferentes variáveis ​​com pontos de dados individuais sem uma linha de conexão

**Esquema:**Uma maneira de descrever como algo, como dados, é organizado

**Escopo do trabalho (SOW):**Um esboço acordado das tarefas a serem executadas durante um projeto

**Dados de terceiros:**Dados coletados por um grupo diretamente de seu público e depois vendidos

**SELECIONE:** A seção de uma consulta que indica de quais colunas extrair os dados

**SELECIONE EM:**Uma cláusula SQL que copia dados de uma tabela em uma tabela temporária sem adicionar a nova tabela ao banco de dados

**Brilhante (R):** Um pacote R usado para construir aplicativos web interativos com código R

**Dados pequenos:** Pontos de dados pequenos e específicos, normalmente envolvendo um curto período de tempo, que são úteis para a tomada de decisões do dia a dia

**Metodologia SMART:**Uma ferramenta para determinar a eficácia de uma pergunta com base no fato de ela ser específica, mensurável, orientada para a ação, relevante e com prazo determinado

**Suavização (R):**Um processo usado para tornar as visualizações de dados em R mais claras e legíveis

**Linha de suavização (R):** Uma linha em uma visualização de dados que usa suavização para representar uma tendência

**Mídia social:**Sites e aplicativos por meio dos quais os usuários criam e compartilham conteúdo ou participam de redes sociais

**Habilidades interpessoais:** Traços e comportamentos não técnicos relacionados à forma como as pessoas trabalham

**Intervalo de classificação:** Uma função de menu de planilha que classifica um intervalo especificado e preserva as células fora do intervalo

**Classificar planilha:**Uma função de menu de planilha que classifica todos os dados pela classificação de uma coluna classificada específica e mantém os dados juntos nas linhas

**Ordenação:**O processo de organizar os dados em uma ordem significativa para facilitar a compreensão, análise e visualização

**Pergunta específica:**Uma pergunta simples, significativa e focada em um único tópico ou em algumas ideias intimamente relacionadas

**DIVIDIR:**Uma função de planilha que divide o texto em torno de um caractere especificado e coloca cada fragmento em uma célula nova e separada

**Patrocinador:**Um defensor profissional comprometido em promover a carreira de outra pessoa

**Destaque:** Varredura de dados para identificar rapidamente os insights mais importantes

**Planilha:** Uma planilha digital

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

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

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

**Visualização estática:** Uma visualização de dados que não muda com o tempo, a menos que seja editada

**Poder estatístico:** A probabilidade de um teste de significância reconhecer um efeito que está presente

**Significado estatístico:**A probabilidade de que os resultados da amostra não sejam devidos ao acaso

**Estatisticas:** O estudo de como coletar, analisar, resumir e apresentar dados

**História:**A narrativa de uma apresentação de dados que a torna significativa e interessante

**Tipo de dados de sequência:** Uma sequência de caracteres e pontuação que contém informações textuais (também chamada de tipo de dados de texto)

**Metadados estruturais:**Metadados que indicam como um dado está organizado e se faz parte de uma ou mais de uma coleta de dados

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

**Linguagem de consulta estruturada:**Uma linguagem de programação de computador usada para se comunicar com um banco de dados

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

**Subconsulta:**Uma consulta SQL aninhada dentro de uma consulta maior

**SUBSTR:** Uma função SQL que extrai uma substring de uma variável de string

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

**Legenda:** Texto que suporta um título adicionando contexto e descrição

**SOMA:** Uma função de planilha que adiciona os valores de um intervalo selecionado de células

**SOMASE:**Uma função de planilha que adiciona dados numéricos com base em uma condição

**Tabela de resumo:**Uma tabela usada para resumir informações estatísticas sobre dados

**SUMPRODUTO:** Uma função que multiplica arrays e retorna a soma desses produtos

**Rápido:** Uma linguagem de programação para macOS, iOS, watchOS e tvOS

**Mapa de símbolos:** Uma visualização de dados queexibe uma marca sobre uma determinada longitude e latitude

**Sintaxe:** A estrutura predeterminada de um idioma que inclui todas as palavras, símbolos e pontuação necessários, bem como seu posicionamento adequado

T

**Pintura:** Uma plataforma de business intelligence e análise que ajuda as pessoas a visualizar, compreender e tomar decisões com dados

**Mentalidade técnica:**A capacidade de dividir as coisas em etapas ou pedaços menores e trabalhar com elas de maneira ordenada e lógica

**Mesa temporária:** Uma tabela de banco de dados criada e existente temporariamente em um servidor de banco de dados

**Tipo de dados de texto:**Uma sequência de caracteres e pontuação que contém informações textuais (também chamada de tipo de dados string)

**Sequência de texto:** Um grupo de caracteres dentro de uma célula, geralmente composto por letras

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

**Tibble (R):** Uma variação simplificada de quadros de dados

**Dados organizados (R):**Uma forma de padronizar a organização dos dados dentro do R

**não (R):** Um pacote R no Tidyverse usado para limpeza de dados para torná-los organizados

**Organizadoverso (R):** Um sistema de pacotes em R com uma filosofia de design comum para manipulação, exploração e visualização de dados

**Pergunta com limite de tempo:**Uma pergunta que especifica um prazo a ser estudado

**Transparência nas transações:** O aspecto da ética dos dados que pressupõe que todas as atividades e algoritmos de processamento de dados devem ser explicáveis ​​e compreendidos pelo indivíduo que fornece os dados

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

**APARAR:**Uma função que remove espaços iniciais, finais e repetidos nos dados

**TSV (arquivo de valores separados por tabulação):**Um arquivo de texto que armazena uma tabela de dados separando colunas de dados com guias

**Taxa de rotatividade:** A taxa na qual os funcionários deixam voluntariamente uma empresa

**Tipografia:** Convertendo dados de um tipo para outro

EM

**Amostragem imparcial:**Quando a amostra da população que está sendo medida é representativa da população como um todo

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

**Pergunta injusta:**Uma pergunta que faz suposições ou é difícil de responder honestamente

**Exclusivo:**Um valor que não pode ser duplicado

**Escritório do Censo dos Estados Unidos:** Uma agência do Departamento de Comércio dos EUA que atua como principal fornecedora de dados de qualidade sobre seu povo e sua economia.

**Unidade:** O princípio de design de usar elementos visuais que se complementam para criar apelo estético e clareza em uma visualização de dados

**Dados não estruturados:**Dados que não estão organizados de maneira facilmente identificável

EM

**Validade:**O grau em que os dados estão em conformidade com as restrições quando são inseridos, coletados ou criados

**VALOR:**Uma função de planilha que converte uma sequência de texto que representa um número em um valor numérico

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

**Variedade:** O princípio de design de usar diferentes tipos de elementos visuais em uma visualização de dados para envolver o público

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

**Verificação:** Um processo para confirmar que um esforço de limpeza de dados foi bem executado e que os dados resultantes são precisos e confiáveis

**Arquivo de vídeo:** Uma coleção de imagens, arquivos de áudio e outros dados geralmente codificados em um formato compactado, como MP4, MV4, MOV, AVI ou FLV

**Vinheta (R):** Documentação para um pacote R que descreve o problema que o pacote foi projetado para resolver, explica como suas funções podem ser usadas e lista quaisquer dependências de outros pacotes

**Forma visual:**A aparência de uma visualização de dados que lhe confere estrutura e apelo estético

**Visualização:**(Consulte Visualização de dados)

**PROCV:** Uma função de planilha que pesquisa verticalmente um determinado valor em uma coluna para retornar uma informação correspondente

EM

**ONDE:**A seção de uma consultaque especifica os critérios que os dados solicitados devem atender

**Dados amplos:** Um conjunto de dados em que cada titular de dados tem uma única linha com múltiplas colunas para armazenar os valores de vários atributos do assunto

**COM:** Uma cláusula SQL que cria uma tabela temporária que pode ser consultada diversas vezes

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

X

**Eixo X:**A linha horizontal de um gráfico geralmente colocada na parte inferior, que é frequentemente usada para representar escalas de tempo e categorias discretas

E

**Eixo Y:**A linha vertical de um gráfico geralmente colocada à esquerda, que é frequentemente usada para representar frequências e outras variáveis ​​numéricas

**YAML:**Uma linguagem que traduz dados para melhorar a legibilidade

COM