A

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

**Agenda:** Lista de compromissos agendados.

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

**Análise de dados:** Ciência dos dados

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

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

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

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

**Arquivo CSV (comma-separated values):** 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 vídeo:** Coleção de imagens, arquivos de áudio e outros dados geralmente codificados em formato comprimido, como MP4, MV4, MOV, AVI, ou FLV.

**Atributo**: Uma característica ou qualidade dos dados usados para rotular uma coluna em uma tabela

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

**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

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

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

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

**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, 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.

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

**Consentimento:** O aspecto da ética dos 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.

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

**Contexto:** A condição na qual algo existe ou acontece

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

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

**COUNT**: Uma função de planilha que conta o número de células em uma faixa atende a um critério específico

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

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

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

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

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

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

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

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

**Dados tipo texto:** Uma sequência de caracteres e pontuação que contém informações textuais (também chamada de string data type)

**Dados:** Uma coleção de fatos

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

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

E

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

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

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

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

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

**Ética em 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.

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

F

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

**Fonte de dados bons:** Uma fonte de dados confiável, original, abrangente, atual e citada (ROCCC) (Consulte a Fonte de dados ruim)

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

**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 onde vêm os dados selecionados.

**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.

G

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

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

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

H

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

I

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

**Interoperabilidade dos dados:** Capacidade de integrar dados de múltiplas fontes, um fator importante no uso bem-sucedido de dados abertos entre empresas e governos.

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

L

**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.

M

**MAX:** Função da planilha que calcula o maiorvalor numérico de um intervalo de células.

**MÉDIA:** Função da planilha que fornece a média dos valores de um intervalo selecionado

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

**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.

**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:** Um tipo de dados quantificável e isolado que pode ser usado para medição.

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

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

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

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

N

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

**Nuvem:** Um lugar para manter os dados on-line, em vez de um disco rígido de computador

O

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

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

**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.

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

P

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

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

**Pensamento analítico:** O processo de identificação e definição de um problema, depois resolvê-lo usando os dados de forma 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 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:** Uma pergunta que direciona as pessoas para uma determinada resposta.

**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

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

**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 pergunta, preparação, processamento, análise, compartilhamento e ação, cujo propósito é obter insights que orientem a tomada de decisão informada.

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

R

**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 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.

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

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

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

S

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

**SELECT:** Seção de uma consulta que indica o subconjunto de um conjunto de dados.

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

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

T

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

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

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

**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.

**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

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

U

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

V

**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.

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

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

W

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