A

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

**Agenda**: Lista de compromissos agendados.

**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.

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

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

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

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

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

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.

**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:** 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.

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

**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, 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.

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

**Compatibilidade:** Como dois ou mais conjuntos de dados são capazes de trabalhar em conjunto.

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

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

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

**CONCATENATE:** Uma 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:** O 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:** O grau em que os dados são repetíveis de diferentes pontos de entrada ou coleta

**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.

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

**COUNT**: Função da planilha que conta o número de células em um intervalo que atendem a critérios especificados.

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

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

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

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

**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

**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.

**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.

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.

**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 (Ver Expressão matemática).

**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.

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

**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

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

**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).

**Formatação condicional:** Uma 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 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.

G

**Geolocalização:** 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.

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

**Integridade:** O 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:** Um 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.

L

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

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

**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

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

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

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

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

**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.

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

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

N

**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:** 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.

**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.

**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 significado reconhecer um efeito que está presente

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

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

**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.

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.

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

**Replicação de dados:** O processo de armazenamento de dados em múltiplos locais.

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

**Restrições de dados:** Os 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:** Função que retorna um determinado número de caracteres do lado direito de uma cadeia de texto.

S

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

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

**Significado estatístico:** Probabilidade de que os resultados da amostra não se devam a um acaso

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

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

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

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

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

**SUM:** Função que adiciona os valores de uma faixa selecionada de células

T

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

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

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

**Taxa de rotatividade:** Taxa que mensura a saída e entrada 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.

**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.

**Transferência de dados:** O 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.

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

U

**Único:** Um valor que não pode ter um duplicado

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

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

**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:** Representação gráfica dos dados.

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

**VLOOKUP:** Uma 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.