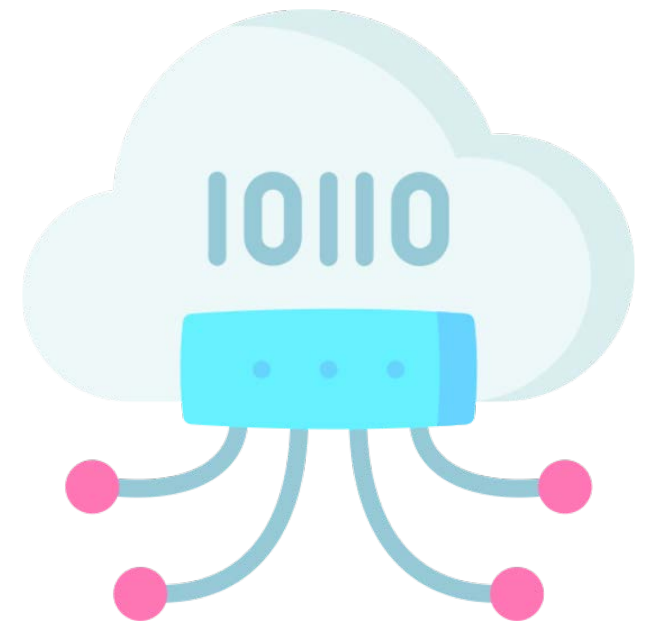


O que são Dados?

Dados são **fatos brutos**, ou seja, que ainda não foram processados, interpretados ou organizados, que **representam informações coletadas e registradas sobre algo**. Eles podem ser:

- Números, alfabetos, símbolos, sons ou uma combinação de todos esses.
- Quantitativos ou qualitativos.



O que são Dados?

Dados estão sempre vinculados:

A um “evento” e a uma “característica” (que é o que observamos do evento – e podemos observar por diversas lógicas diferentes).

Exemplo 1: no problema “assar bolos” podemos ter a informação de que um bolo (evento) é de chocolate (característica 1), pesa 100 gramas (característica 2) e demorou 30 minutos para ser assado (característica 3).

Exemplo 2: no problema “vender carros”, podemos analisar cada mês de trabalho (evento) pela quantidade de carros vendidos (característica 1), valor vendido (característica 2) e até pelo preço médio das vendas (característica 3).

Por estas razões dizemos que dados são sempre dependentes de um contexto.

Trabalhando com Dados

Dados

Informação

- Variáveis
- Atributos



Adulto ▾	Peso (kg) ▾	Altura (m) ▾
1	70	1.75
2	65	1.68
3	80	1.80
4	75	1.70
5	72	1.73

- Entrada individual de dados

Trabalhando com Dados

Trabalhar com dados nada mais é do que entender como gerenciar as informações ao nosso redor para a melhor tomada de decisão possível.

Algumas atividades precisam ser feitas:

- Garantir a confiabilidade das informações;
- Cruzar informações de diferentes fontes e sobre diferentes características;
- Contextualizar as informações;
- Usar técnicas para visualizar e entender o que está acontecendo frente a uma “enxurrada de informações”.



Trabalhando com Dados

Dados podem representar uma população ou uma amostra (parte dessa população).

Podem ser obtidos por meio de estudos:

- Observacionais;
- Amostrais;
- Experimentais.

Definição

População é o conjunto de elementos para os quais se deseja estudar determinada(s) característica(s).

Amostra é um subconjunto da população.



Trabalhando com Dados

A partir da estruturação que acabamos de ver, é possível fazer a **Análise de Dados Amostrais**.

Ela possibilita que se faça inferência sobre a distribuição de probabilidades das variáveis de interesse, definidas sobre a população da qual a amostra foi colhida.

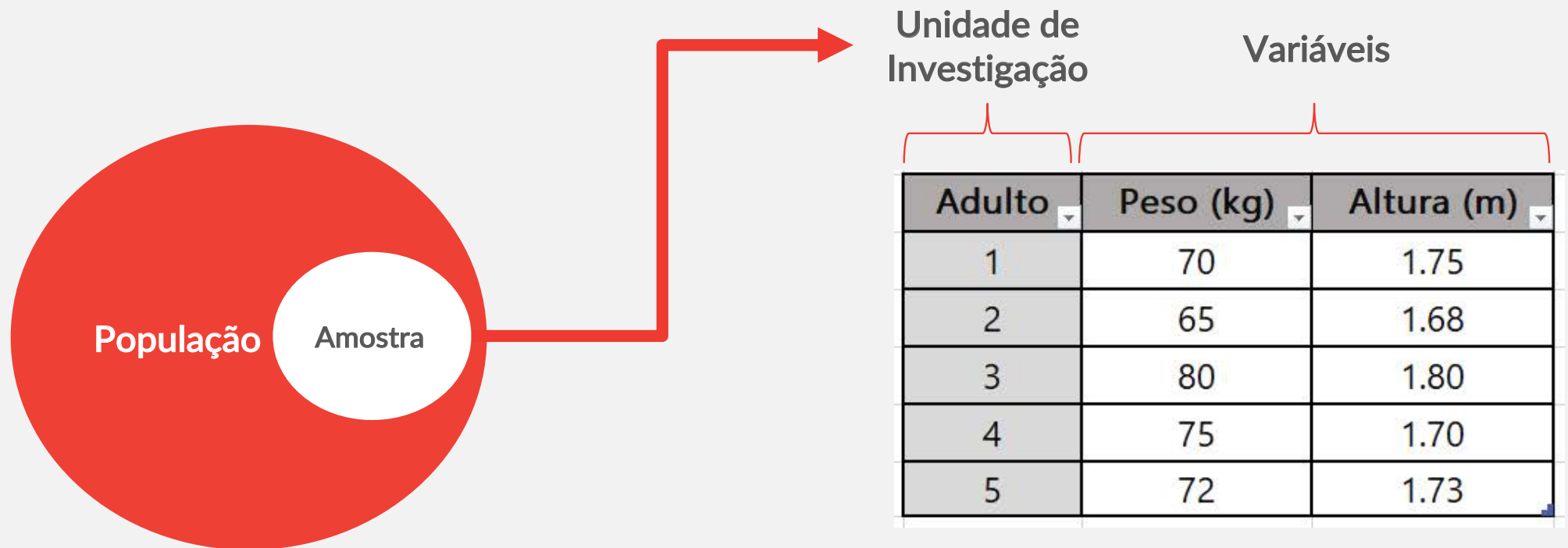
Nesse contexto, a **Estatística** é uma ferramenta importante para organizá-los, resumi-los, analisá-los e utilizá-los para tomada de decisões.

Inferência Estatística se refere ao processo de se tirar conclusões sobre uma população com base em uma amostra dela.



Trabalhando com Dados

Mais comumente, os dados envolvem valores de **várias variáveis** obtidos da observação de **unidades de investigação** que constituem uma amostra de uma população.



Dados x Informação

Para que os dados tenham propósito e gerem conhecimento, eles precisam ser processados, organizados e analisados. Só aí, a partir desse processamento, gerenciamento e estruturação, é que os dados podem ser considerados **informações**.

Pode-se dizer, portanto, que a informação é uma coleção de dados.



Quando usar Dados?

Os dados são usados em uma ampla variedade de contextos para tomar decisões, realizar análises, entender padrões e tendências, etc.

Conforme discutido em "*Introduction to Statistical Learning*" por Gareth James et al., os dados formam a **base essencial para análises estatísticas, modelagem preditiva e inferência estatística.**

A seguir, estão algumas situações comuns em que os dados são utilizados:



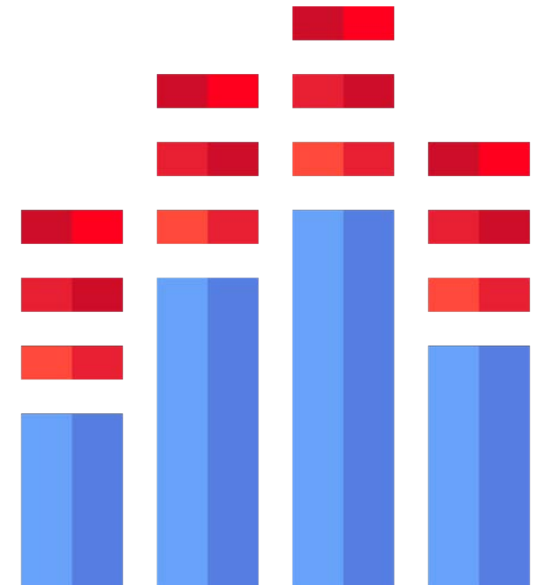
Quando usar Dados?

Tomada de Decisão Empresarial:

- Avaliar o desempenho;
- Identificar oportunidades de crescimento;
- Tomar decisões estratégicas.

Análise de Mercado:

- Entender as preferências dos clientes;
- Identificar tendências e prever demandas futuras.



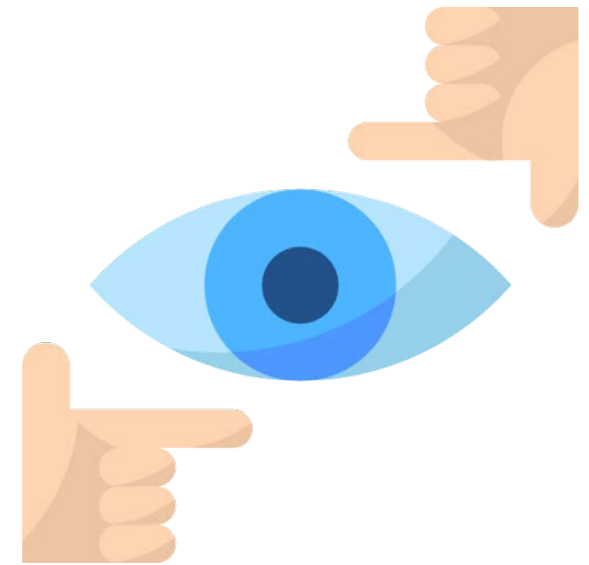
Quando usar Dados?

Pesquisa Científica:

- Realizar experimentos;
- Validar hipóteses;
- Descobrir novos conhecimentos em diversas áreas.

Análise Financeira:

- Avaliar riscos;
- Identificar oportunidades de investimento;
- Monitorar o desempenho de ativos financeiros.



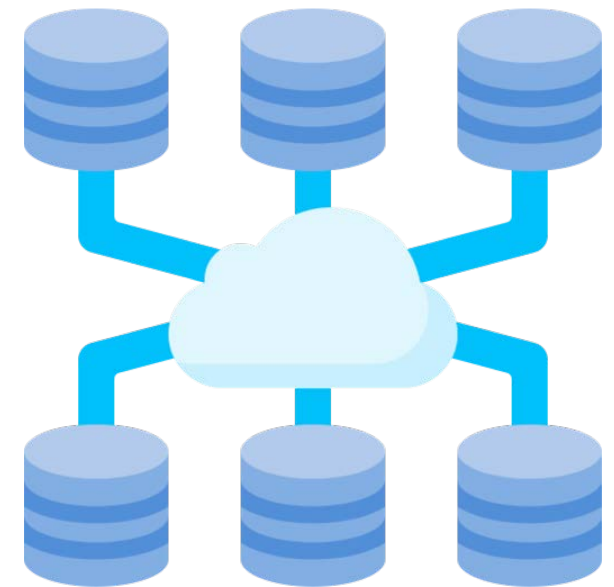
Como Dados Influenciam nossa Vida

Ao compreendermos onde os dados são empregados, podemos refletir sobre como esses dados influenciam profundamente em nossa vida diária.

De que forma?

- Moldando nossas interações sociais;
- Nossas escolhas de consumo;
- E até mesmo nossa percepção do mundo ao nosso redor.

A seguir, veremos alguns exemplos.



Como Dados Influenciam nossa Vida

Tomada de Decisão Pessoal

- Escolher um restaurante com base nas avaliações online;
- Decidir qual rota seguir com base no tráfego em tempo real.

Consumo

- Personalizar experiências de compra;
- Recomendar produtos com base em nossas preferências;
- Ajustar preços com base em demanda e oferta.

Educação e Carreira

- Desempenho acadêmico;
- Tendências do mercado de trabalho e oportunidades de carreira;
- Tomar decisões educacionais e profissionais informadas.

Outros: Saúde e Bem-Estar; Mobilidade Urbana; Viagens...

Como Dados Influenciam o Ambiente Corporativo

Tomada de Decisão

- Embasar suas decisões estratégicas;
- Incluir desde a identificação de novas oportunidades de mercado até a alocação de recursos de forma mais eficiente.

Otimização de Processos

- Identificar áreas de ineficiência em processos operacionais;
- Implementar melhorias para aumentar a eficiência e reduzir custos.

Gestão de Riscos

- Identificar e mitigar riscos, seja no contexto financeiro, de segurança cibernética ou de conformidade regulatória.

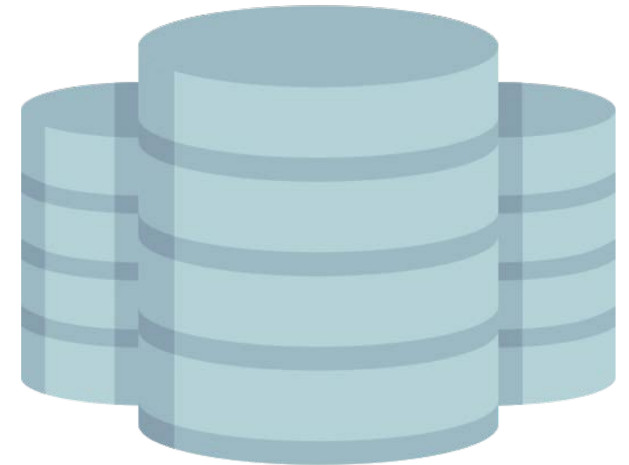
Inovação e Desenvolvimento de Produto

- Ao analisar feedback do cliente e dados de mercado, as empresas podem identificar oportunidades para inovação e desenvolvimento de novos produtos ou serviços que atendam às necessidades do mercado.

Próximo tópico...

Ao refletir sobre o impacto dos dados em nossa vida cotidiana, torna-se evidente a necessidade de compreender profundamente os diferentes tipos de dados, como e quando utilizá-los.

No próximo tópico, exploraremos os tipos de dados na ciência de dados, examinando sua natureza, características e aplicações.



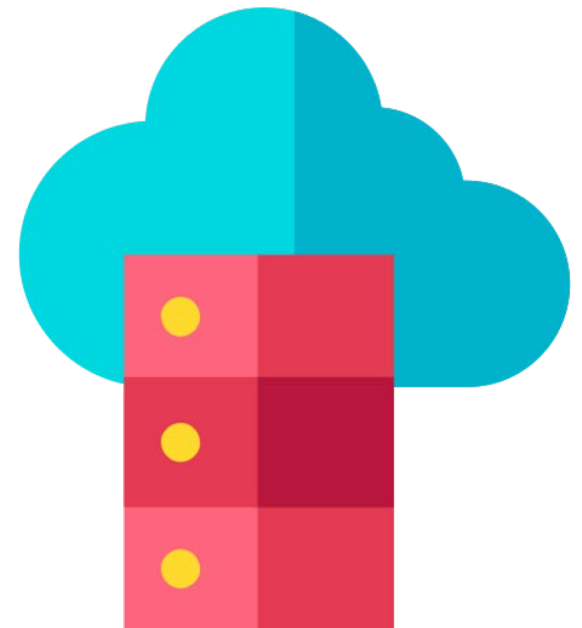
Tipos de Dados

Por que é importante conhecer os Tipos de Dados?

Compreender a natureza de cada tipo de dado é essencial para selecionar as técnicas adequadas de **análise, modelagem e visualização**, bem como para **interpretar corretamente** os resultados obtidos.

Ao compreender profundamente os diferentes tipos de dados, os profissionais de ciência de dados podem:

- Conduzir análises mais precisas;
- Desenvolver modelos mais robustos;
- Tomar decisões mais informadas;
- Dentre outros.



Quais são os Tipos de Dados?

A seguir, estão alguns dos tipos de dados que são fundamentais na ciência de dados e que são amplamente utilizados em uma variedade de aplicações:

- Dados Numéricos;
- Dados Categóricos;
- Dados Textuais.

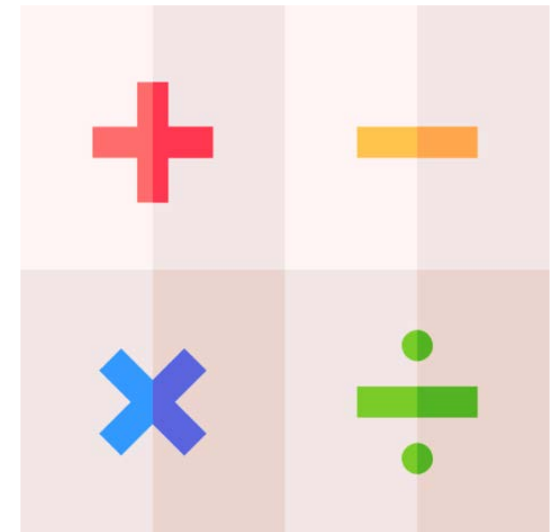


Dados Numéricos

Conceito: São dados representados por números e são utilizados para medir quantidades ou realizar cálculos matemáticos.

Importância: Esses dados são fundamentais na análise estatística e na modelagem matemática, permitindo análises quantitativas e a aplicação de algoritmos de *machine learning*.

Aplicação: Exemplos incluem dados de vendas, temperatura, idade, altura, entre outros.



Dados Numéricos

Exemplo: Dados de Vendas Mensais de uma loja. Cada entrada representa o valor das vendas em dólares para um determinado mês.

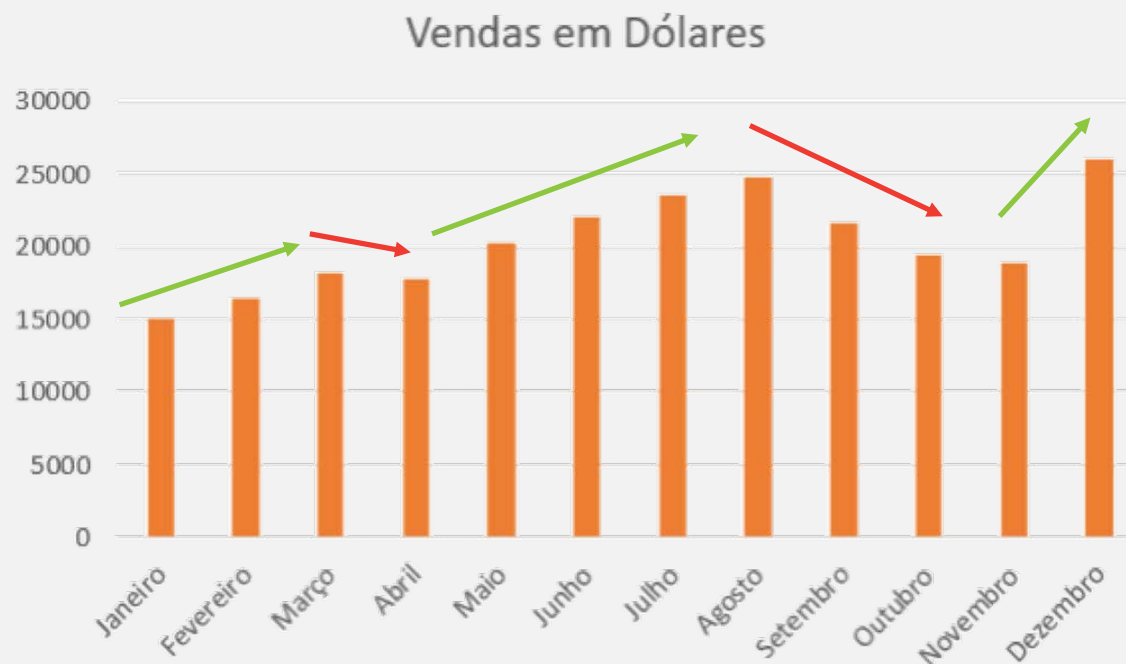
Vantagens dos Dados Numéricos nesse exemplo:

- Comparação de Desempenho;
- Análise de Sazonalidade;
- Previsão de Vendas Futuras;
- Identificação de Tendências.
 - A seguir, um exemplo de gráfico para Identificação de Tendências:

Mês	Vendas em Dólares
Janeiro	15000
Fevereiro	16500
Março	18200
Abril	17800
Maio	20300
Junho	22000
Julho	23500
Agosto	24800
Setembro	21700
Outubro	19400
Novembro	18900
Dezembro	26000

Dados Numéricos

Mês	Vendas em Dólares
Janeiro	15000
Fevereiro	16500
Março	18200
Abril	17800
Maio	20300
Junho	22000
Julho	23500
Agosto	24800
Setembro	21700
Outubro	19400
Novembro	18900
Dezembro	26000



Dados Categóricos

Conceito: São dados que representam categorias ou grupos e não possuem uma ordem natural.

Importância: Permitem a segmentação e classificação de dados, sendo essenciais em análises de mercado, segmentação de clientes e classificação de textos.

Aplicação: Exemplos incluem gênero, estado civil, tipo de produto, entre outros.



Dados Categóricos

Exemplo: Estado civil de clientes de uma empresa.

As categorias podem incluir solteiro, casado, divorciado e viúvo.

Cliente	Estado Civil
Cliente A	Solteiro
Cliente B	Casado
Cliente C	Divorciado
Cliente D	Viúvo
Cliente E	Solteiro

Os dados sobre o estado civil dos clientes podem ser úteis para a empresa de várias maneiras:

- Segmentação de Mercado;
- Personalização de Produtos e Serviços;
- Planejamento de Eventos e Promoções;
- Análise de Demanda;
- Fidelização dos Clientes.

Dados Textuais

Conceito: São dados compostos por texto e são comumente encontrados em documentos, redes sociais, e-mails, entre outros.

Importância: Permitem análises de sentimentos, processamento de linguagem natural e extração de informações úteis a partir de grandes volumes de texto.

Aplicação: Exemplos incluem *tweets*, artigos de jornal, *e-mails* corporativos, entre outros.



Dados Textuais

Exemplo: Comentários de clientes em uma plataforma de *e-commerce*. Cada entrada é um texto que expressa a opinião ou experiência do cliente com um produto ou serviço.

Vantagens dos Dados Textuais nesse exemplo:

- Análise de Sentimento;
- *Feedback* do Cliente;
- Personalização da Experiência do Cliente;
- Identificação de Tendências e Padrões;
- Melhoria Contínua.

Cliente	Comentário
Cliente A	"Adorei o produto! Ótima qualidade e entrega super rápida."
Cliente B	"A entrega demorou mais do que o esperado, mas o produto é excelente."
Cliente C	"Produto de baixa qualidade. Não correspondeu às minhas expectativas."
Cliente D	"Recomendo! Produto exatamente como descrito e entrega dentro do prazo."
Cliente E	"O produto chegou danificado. Entrei em contato com o suporte e fui muito bem atendido."

Dados Estruturados x Dados Desestruturados

