

# *Como utilizar a análise de dados para obter informações açãoáveis para a tomada de decisão*

Princípios e dicas práticas de análise de dados



Superintendência de  
Inovação da Gestão  
Pública

Subsecretaria de  
Inovação da Gestão e  
dos Serviços Públicos

SEAD  
Secretaria de Estado de  
Administração



## **GOVERNADOR**

Ronaldo Ramos Caiado

## **SECRETÁRIO DA ADMINISTRAÇÃO**

Alan Farias Tavares

## **SUBSECRETÁRIO DE INOVAÇÃO DA GESTÃO E DOS SERVIÇOS PÚBLICOS**

Rômulo Mendonça Bailão

## **SUPERINTENDENTE DE INOVAÇÃO DA GESTÃO PÚBLICA**

Daniel Soares Santana

## **GERENTE DE INTELIGÊNCIA DE NEGÓCIOS PÚBLICOS**

Camila do Nascimento Freitas

## **EQUIPE TÉCNICA E EDIÇÃO**

Carlos Alexandre Alencar de Assis

Giovanna Marques Grassini

Jessé Daniel de Souza

Rafaela Moraes Rosa

## **REVISÃO**

Mirielly dos Santos Dourado

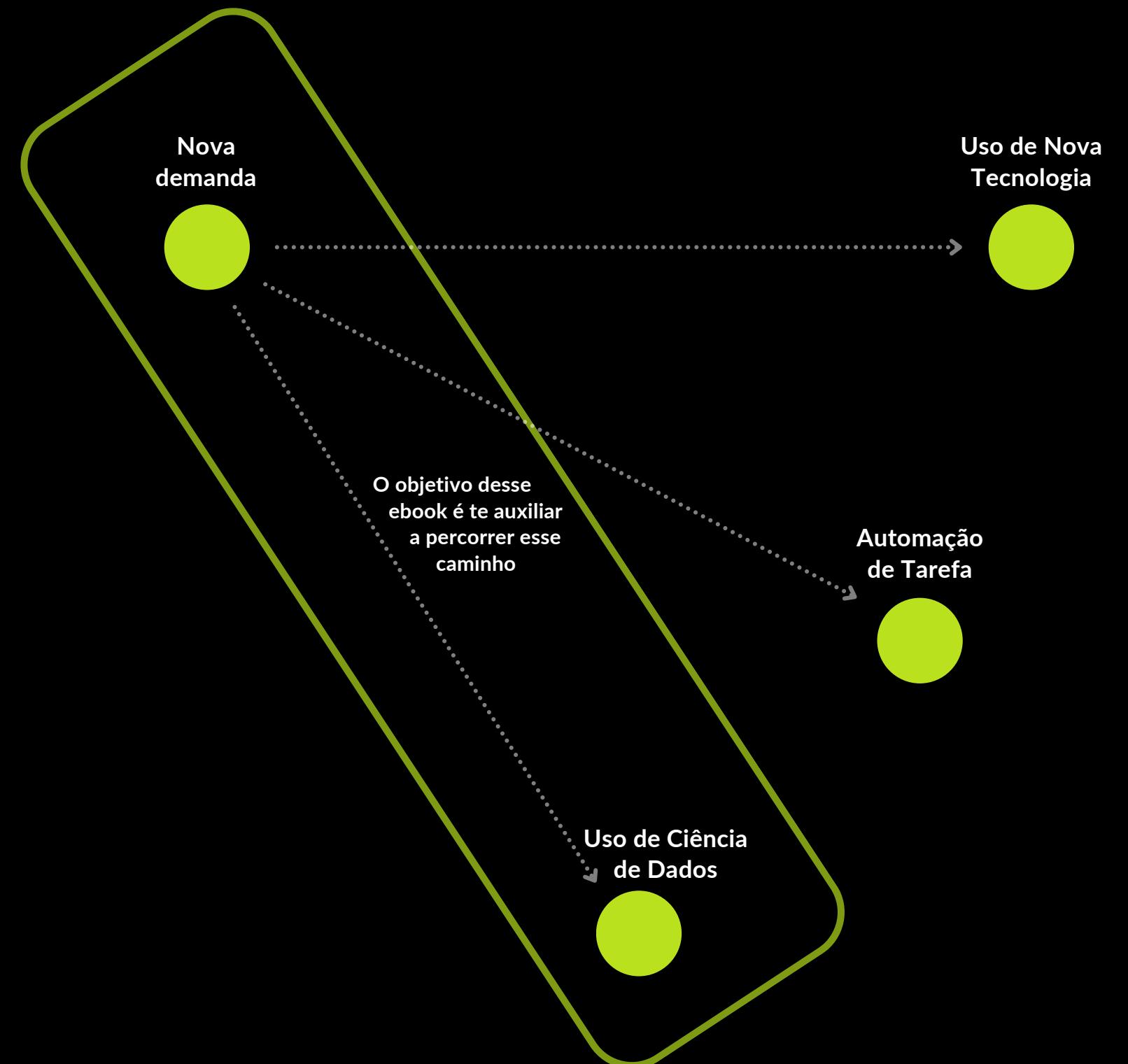
# Conheça a GEINP

Somos a **Gerência de Inteligência de Negócios Públicos (GEINP)**, uma equipe multidisciplinar e colaborativa dedicada a buscar soluções eficazes e inovadoras para a administração pública nas áreas de novas tecnologias, automatização de tarefas e ciência de dados.

Nosso propósito é garantir que a gestão pública esteja sempre à frente, antecipando problemas e identificando oportunidades, para oferecer soluções eficazes e transparentes que atendam às necessidades da população de maneira exemplar.

Em nossa Jornada, atuamos em **três frentes**:

- Prospecção e adoção de novas tecnologias;
- Automação de tarefas manuais e repetitivas; e
- Aplicação de Ciência de Dados.



# Índice

1	<i>Introdução</i>	5
2	<i>Principais Métodos de Análise de Dados</i>	8
3	<i>Processos de Análise de Dados</i>	11
4	<i>Ferramentas de Análise de Dados</i>	14
5	<i>Análise de Dados na Gestão Pública</i>	17
6	<i>Conclusão</i>	19



# 1 Introdução

A análise de dados se tornou uma ferramenta essencial para organizações públicas que desejam tomar **decisões informadas e estratégicas**. Com a quantidade de dados disponíveis atualmente, é crucial **saber como coletar, analisar e interpretar** essas informações de maneira eficiente para transformar dados brutos em insights valiosos e acionáveis.

A análise de dados ajuda a descobrir insights valiosos de conjuntos de dados complexos



## Dica



***“Acima de tudo, mostre os dados.”***

Edward Rolf Tufte

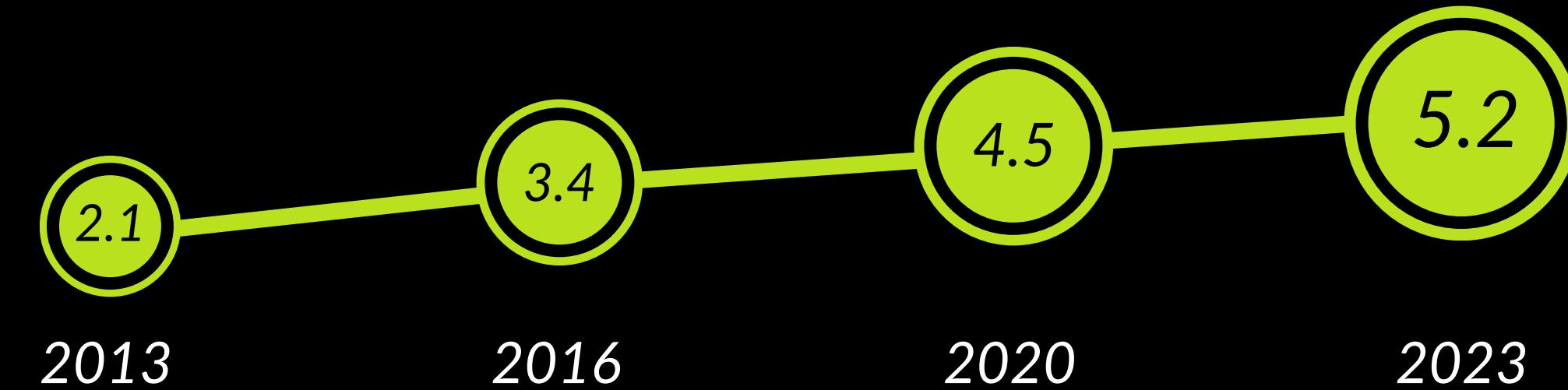
**Dica**

*No mundo digital atual,  
a análise de dados é  
ainda mais essencial,  
especialmente no  
cenário de Big Data*



## Crescimento Populacional Global da Internet

(Em bilhões)



Por meio da análise de dados, podemos entender **padrões** e **tendências**, identificar correlações e compreender **como diferentes variáveis se relacionam**.



Disponível em Domo - Data Never Sleeps

*Você sabia? <<<<<<<<<<<<<<*

*Os dados nunca dormem*

*A quantidade total de **dados criados, capturados, copiados e consumidos** globalmente em 2023 foi de **120 zettabytes**, um número que deve crescer para **181 zettabytes** em 2025*

## 2 Principais Métodos de Análise de Dados

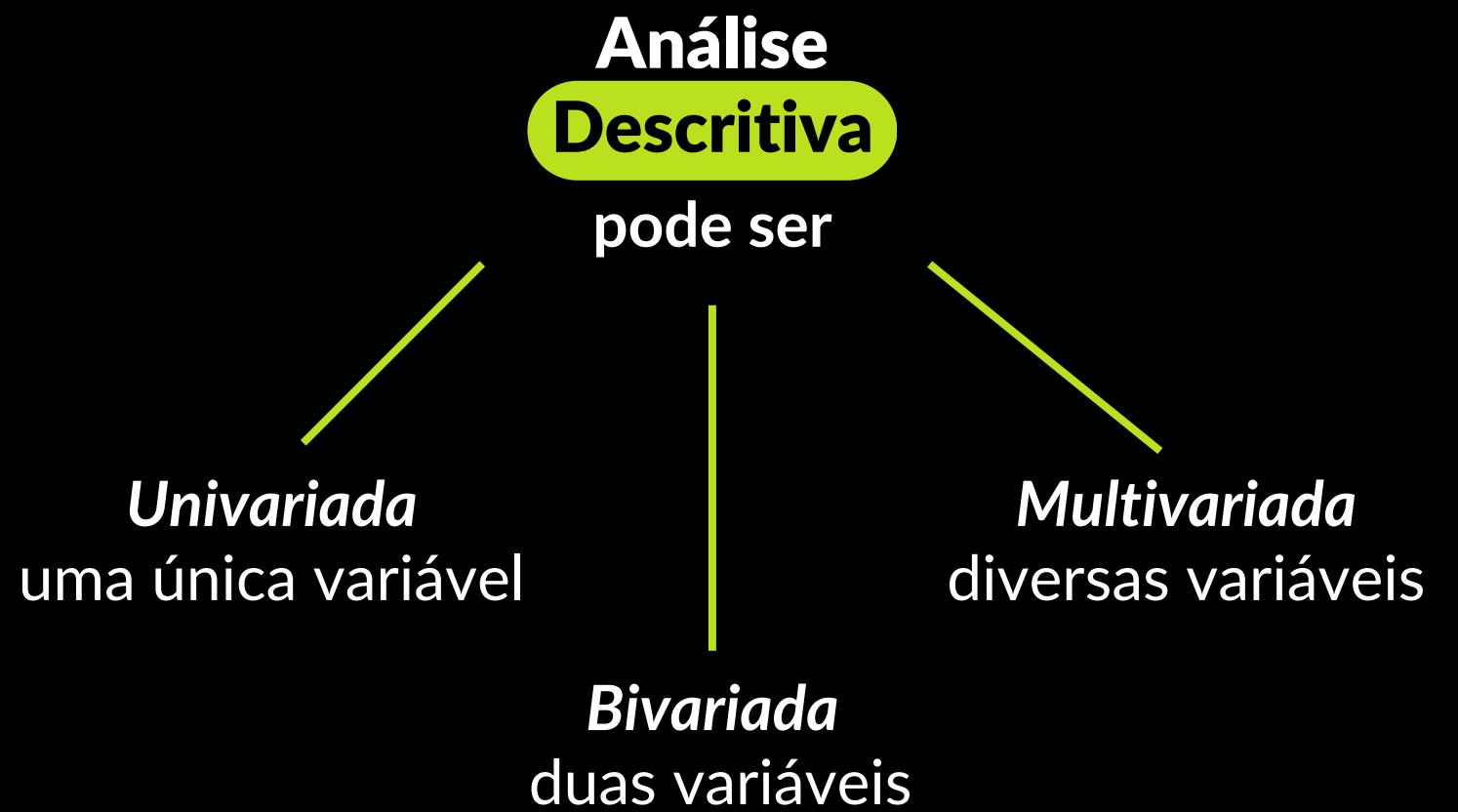


» **Análise Descritiva:** Resume dados históricos para entender o que aconteceu no passado.

**Aplicação:** em relatórios de despesas públicas mensais para identificar tendências de gasto.

» **Análise Diagnóstica:** Investiga os motivos pelos quais algo aconteceu.

**Aplicação:** em análise da causa raiz de um aumento nas taxas de criminalidade em uma área específica.



**Univariada:** investiga uma variável ou coluna no banco de dados  
**Bivariada:** investiga a relação entre duas variáveis  
**Multivariada:** investiga a relação entre diversas variáveis ou colunas do banco

» **Análise Preditiva:** Usa modelos estatísticos e algoritmos para prever futuros resultados.

**Aplicação:** na previsão de demanda por serviços com base em dados demográficos e históricos.

» **Análise Prescritiva:** Sugere ações a serem tomadas com base em análises descritivas e preditivas.

**Aplicação:** na recomendações de rotas de transporte para reduzir o congestionamento com base em padrões de tráfego.

**Dica**

*Defina seu objetivo e entenda muito bem o seu problema antes de começar a usar a análise preditiva*



# Processos de Análise de Dados

## *Knowledge Discovery in Databases - KDD*

O KDD, sigla em inglês para descoberta de conhecimento em bases de dados, é um **processo completo, interativo e iterativo** que envolve a descoberta de conhecimento a partir de grandes volumes de dados. Ele inclui várias etapas, **desde a preparação dos dados até a interpretação dos resultados**, passando pela mineração de dados.

A escolha de métodos e ferramentas depende das necessidades específicas do projeto, nível de experiência da equipe e complexidade da análise a ser realizada.



### *Dica*



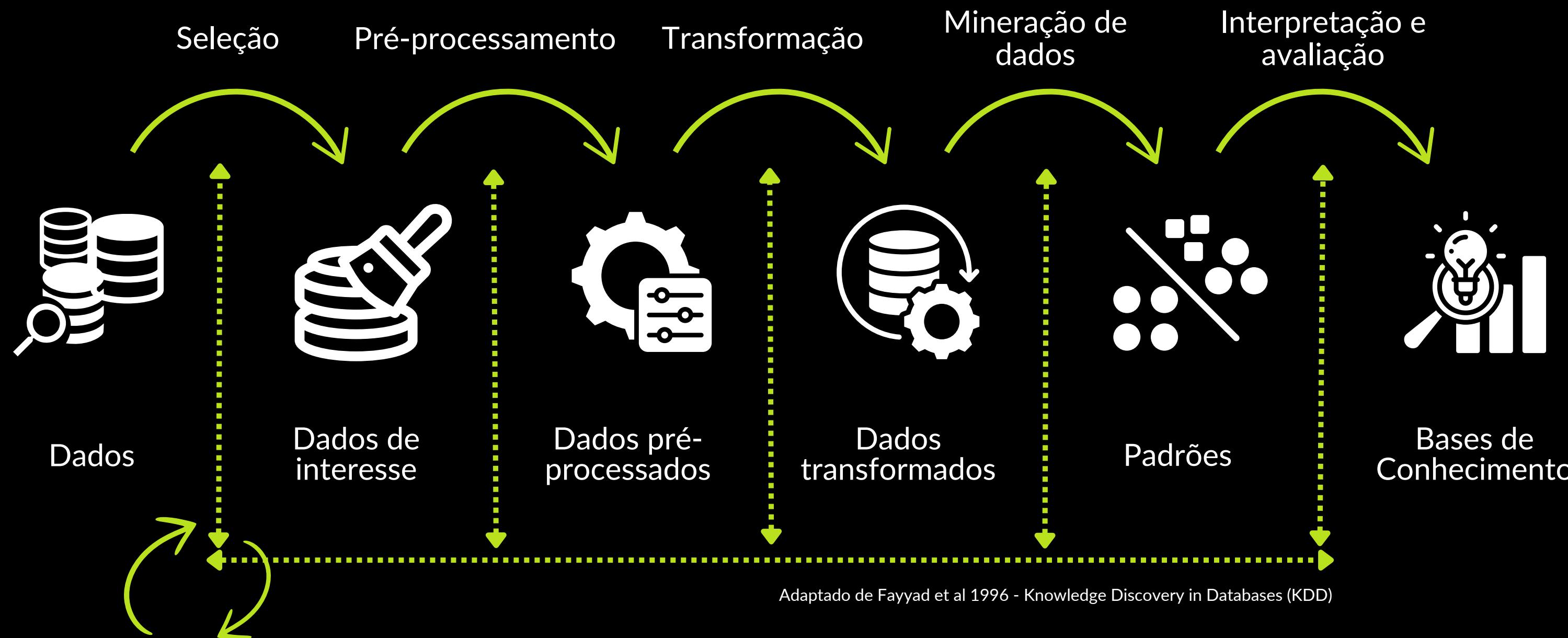
***Você gastará 60% do tempo da análises limpando e organizando dados. Este cuidado garantirá a qualidade dos dados***

## Etapas do KDD

- »» **Seleção de Dados:** Identificação dos dados relevantes para a análise.
- »» **Pré-processamento:** Limpeza e integração dos dados para garantir a qualidade.
- »» **Transformação:** Conversão dos dados em formatos apropriados para a mineração.
- »» **Mineração de Dados:** Aplicação de técnicas e algoritmos para extrair padrões e insights.
- »» **Interpretação/Avaliação:** Análise dos padrões encontrados para extrair conhecimento útil.



Você passará 60% do tempo  
da análises nesta etapa



Você irá **repetir esse processo muitas vezes**. Lembre-se é um processo interativo e iterativo!

# Ferramentas de Análise de Dados



Destaque para  
o Orange



**Microsoft Excel:** O Excel é amplamente utilizado para análise de dados. Permite manipular, visualizar e calcular dados em planilhas. Além disso, possui recursos como tabelas dinâmicas e gráficos.

**Tableau:** Ferramenta de visualização de dados que permite criar painéis interativos e gráficos a partir de várias fontes de dados.

**Power BI:** Plataforma de análise de negócios que permite criar relatórios interativos e painéis de controle com base em dados de várias fontes.

**Python:** Linguagem de programação versátil e popular para análise de dados. As bibliotecas Pandas e NumPy oferecem funcionalidades avançadas para manipulação e análise de dados.

**R:** Outra linguagem de programação amplamente usada em análise de dados. O pacote tidyverse inclui ferramentas como **ggplot** (para visualização) e **dplyr** (para manipulação de dados).

**SQL:** Linguagem essencial para consultas e manipulação de bancos de dados relacionais. Utilizada para **extrair**, **modificar** e **gerenciar dados armazenados** em bancos de dados.

**Google Data Studio:** Ferramenta gratuita de visualização de dados que permite **criar relatórios e painéis personalizados e interativos**. Se integra facilmente com outras ferramentas do Google, como Google Analytics e Google Sheets, além de várias fontes de dados externas.

**Orange:** Ferramenta de código aberto para visualização e análise de dados. Oferece uma **interface visual** para **criar fluxos de trabalho de análise de dados**, tornando acessível a aplicação de técnicas de mineração de dados e **aprendizado de máquina**, mesmo para usuários sem experiência em programação.

## 5

## Análise de Dados na Gestão Pública

Estabeleça claramente o que deseja alcançar com a análise de dados.

**Dedique tempo** na **coleta dos dados**, garanta a coleta de dados relevantes e de **alta qualidade**.

A análise de dados é fundamental no setor público, áreas como Saúde Pública, Segurança, Transporte e Educação possuem variados casos de uso de análise de dados no para promover eficiência e transparência.

Use métodos de análise apropriados para extrair insights dos dados.



### Dica



*Tome decisões  
baseadas nos insights  
obtidos e implemente  
ações estratégicas.*

- »» **Saúde Pública:** uso de dados para prever surtos de doenças e alocar recursos de saúde de forma eficiente, com análise avançada de dados de saúde é possível realizar previsão de tendências.
- »» **Segurança Pública:** análise de dados de criminalidade para otimizar a distribuição de forças policiais e reduzir a criminalidade, ferramenta como o **SAS Visual Analytics** para análise de dados de incidentes criminais e identificação de padrões.
- »» **Transporte Público:** utilização de dados de tráfego e transporte para melhorar a eficiência do transporte público e reduzir congestionamentos, ferramenta como o **Google BigQuery** para análise de grandes volumes de dados de tráfego em tempo real.
- »» **Educação Pública:** análise dos dados educacionais para identificar áreas de melhoria no sistema escolar e personalizar programas educacionais, o **Power BI** pode ser utilizado para visualização de dados educacionais e identificação de tendências e lacunas no desempenho estudantil.

# Conclusão

*Relembrando os processos que são fundamentais para análise de dados*

**Definir Objetivos:** Estabeleça claramente o que deseja alcançar com a análise de dados.

**Coletar Dados:** Garanta a coleta de dados relevantes e de alta qualidade.

**Escolher Ferramentas:** Selecione as ferramentas de análise de dados que melhor se adequam às suas necessidades.

**Analizar Dados:** Use métodos de análise apropriados para extrair insights dos dados.

**Interpretar Resultados:** Transforme os resultados da análise em informações açãoáveis.

**Implementar Ações:** Tome decisões baseadas nos insights obtidos e implemente ações estratégicas.

**Monitorar e Ajustar:** Monitore os resultados das ações implementadas e faça ajustes conforme necessário.

A análise de dados é uma **competência crucial no setor público moderno**. Ao utilizar **métodos e ferramentas adequadas**, as organizações públicas podem transformar dados brutos em insights valiosos que orientam a tomada de decisão e promovem a eficiência e a transparência.



# Referências

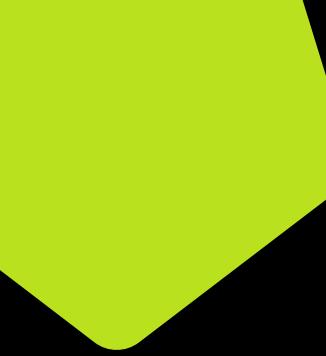


## Livros

- "Data Science for Business" - Foster Provost e Tom Fawcett.
- "Python for Data Analysis" - Wes McKinney.

## Links interessantes

- <https://www.domo.com/learn/infographic/data-never-sleeps-11>  
(Imagens: Crescimento populacional da internet global e Os dados nunca dormem)
- Competições: <https://www.kaggle.com/competitions>
- Competições: <https://semeval.github.io/SemEval2024/>
- Benchmark: <https://gluebenchmark.com/tasks>
- Paper sobre Bert treinado no Domínio Legal/Legislativo em: Português:  
[https://www.researchgate.net/publication/378909440\\_RoBERTaLexPT\\_A\\_Legal\\_RoBERTa\\_Model\\_pretrained\\_with\\_eduplication\\_for\\_Portuguese](https://www.researchgate.net/publication/378909440_RoBERTaLexPT_A_Legal_RoBERTa_Model_pretrained_with_eduplication_for_Portuguese)
- <https://huggingface.co/>  
(Disponibiliza ferramentas de machine learning e processamento de linguagem natural - PLN)
- <https://scikit-learn.org/stable/>  
(Biblioteca de código aberto de machine learning para Python)
- <https://www.tensorflow.org/?hl=pt-br>  
(Biblioteca de código aberto para machine learning em grande escala, deep learning)



*Nós confiamos nos dados*

In data we trust