

Educação Profissional Paulista

Técnico em
**Ciência de
Dados**

Visualização de dados

Práticas de design, formatação e apresentação dos dados

Aula 1

Código da aula: [DADOS]ANO1C3B4S29A1

Visualização de
dados

Mapa da Unidade 3 Componente 4

semana

26

Cores e elementos
visuais

semana

29

Você está aqui!

Práticas de design,
formatação e
apresentação dos
dados

semana

30

Relatório

**Visualização de
dados**

**Mapa da
Unidade 3
Componente 4**

Você está aqui!

Práticas de design, formatação
e apresentação dos dados

Aula 1

Código da aula: [DADOS]ANO1C3B4S29A1

29



Objetivos da aula

- Conhecer como projetar, formatar e apresentar visualizações de dados eficazes de modo que sejam facilmente compreensíveis.



Recursos didáticos

- Recurso audiovisual para exibição de vídeos e imagens;
- Acesso ao laboratório de informática e/ou à internet.



Duração da aula

50 minutos.



Competências técnicas

- Criar e entender representações visuais.



Competências socioemocionais

- Desenvolver a capacidade de trabalhar em equipe para criar visualizações colaborativas e promover a comunicação eficaz ao apresentar e discutir os resultados.



Primeiras ideias

Como exibir informações de maneira que sejam fáceis e agradáveis de ler em um gráfico ou em um relatório?"

Você já notou como alguns gráficos são mais fáceis de entender do que outros? O que você acha que faz a diferença na forma como as informações são apresentadas?

Quando você vê um infográfico ou um meme informativo nas redes sociais, o que chama sua atenção primeiro? Como você acha que isso se relaciona com a maneira como os dados são organizados e exibidos?

Se você tivesse que criar um pôster para explicar um jogo de videogame ou um esporte usando apenas números e imagens, como você garantiria que seus amigos o entendessem rapidamente? O que você usaria para destacar as partes mais importantes?

Ponto de partida

Suponha que você e seus colegas façam parte de um grupo de jovens interessados em criar uma campanha para promover hábitos alimentares saudáveis entre adolescentes.

O grupo decidiu, então, coletar dados sobre os hábitos alimentares de adolescentes em diferentes bairros. Os dados incluem tipos de alimentos consumidos, frequência de consumo de fast-food e de alimentos frescos.

O desafio é usar esses dados para desenvolver visualizações que mostrem as tendências alimentares e incentivem escolhas mais saudáveis.

Como você apresentaria os dados coletados? O que você acredita que seja interessante mostrar nessa pesquisa?



Etapas da fase de preparação dos dados do Crisp-DM

1 Seleção de dados

Identificar e selecionar os dados relevantes para a análise, descartando os dados desnecessários.

2 Limpeza de dados

Identificar e corrigir erros, valores ausentes e inconsistências nos dados.

3 Construção de dados

Criar novos atributos ou *features* derivadas dos dados existentes, como agregações, combinações ou transformações.

4 Integração de dados

Combinar dados de múltiplas fontes para criar um único conjunto de dados integrado.

5 Formatação de dados

Garantir que os dados estejam no formato correto para análise, incluindo conversão de tipos de dados e padronização de formatos.

6 Revisão de dados

Revisar os dados preparados para garantir que atendam aos requisitos de qualidade e estejam prontos para a análise.

Construindo
o **conceito**

Introdução

Na visualização de dados, uma apresentação eficaz é tão importante quanto a própria análise, envolvendo não apenas a estética dos gráficos, mas também a clareza e a eficácia na comunicação das informações.

Aspectos como layout, escolha de cores e organização dos dados são essenciais para influenciar a compreensão e a interpretação do público.

A formatação ajuda a destacar informações-chave, usar anotações explicativas e escolher escalas apropriadas nos gráficos.



Tome nota

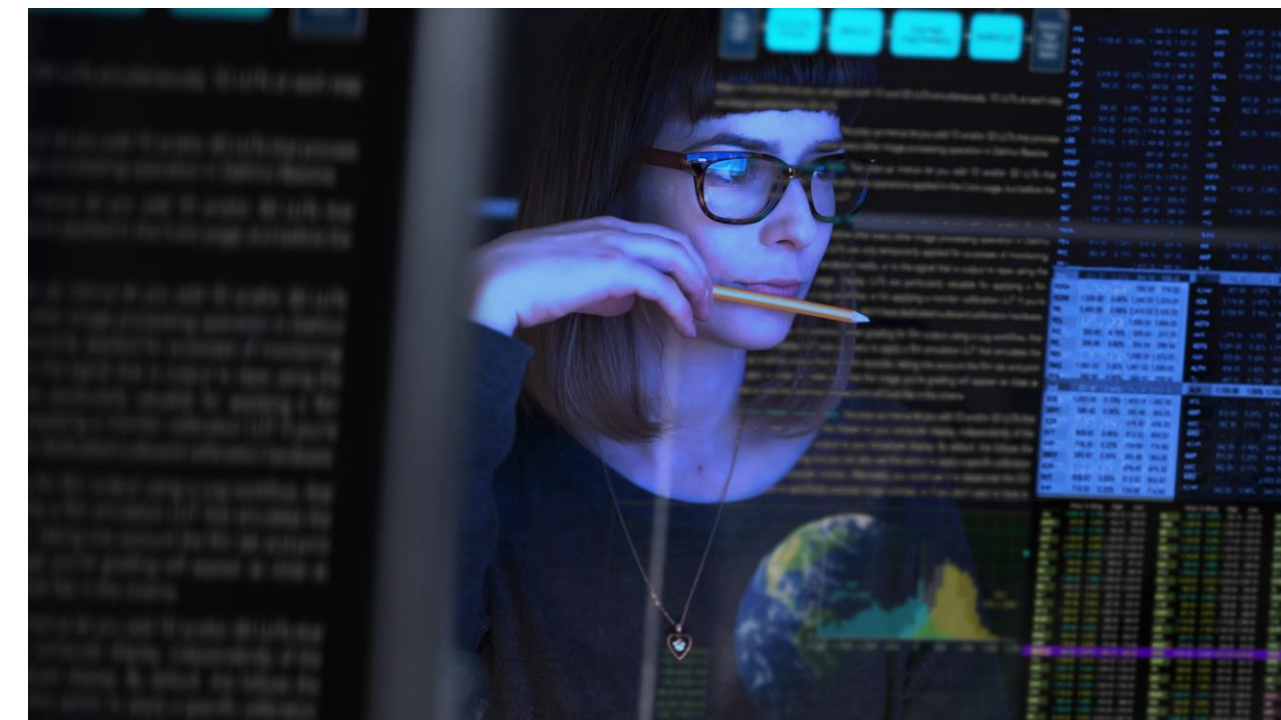
Essas ações não apenas tornam a visualização mais atraente, mas também reforçam a objetividade e a precisão das informações, evitando confusões e sobrecarga cognitiva.

Construindo
o **conceito**

Introdução

É de fundamental importância, antes de tentar criar a visualização da informação, ter realizado as etapas que fazem parte do processo de CRISP-DM (modelo de projeto altamente utilizado em projetos de ciência de dados):

- ▶ Entendimento do negócio;
- ▶ Entendimento dos dados;
- ▶ Preparação dos dados.



© Getty Images

Construindo o conceito

Formatação de dados

É essencial em Ciência de Dados preparar os dados antes de fazer análises profundas ou visualizações. Essa etapa organiza os dados para que sejam facilmente usáveis por humanos e máquinas, corrigindo formatos inadequados, como dados inconsistentes ou números registrados como texto.

1

Padronização de unidades e escalas

Assegurar que todos os dados estejam na mesma unidade de medida ou escala para facilitar comparações e análises.

2

Codificação de categorias

Transformar variáveis categóricas em formatos numéricos (como codificação *one-hot* ou codificação de *labels*) para que possam ser utilizadas em algoritmos de modelagem.

3

Tratamento de datas e horas

Converter *strings* de data/hora para um formato de data/hora consistente ou para características numéricas que possam ser usadas em análises temporais.

4

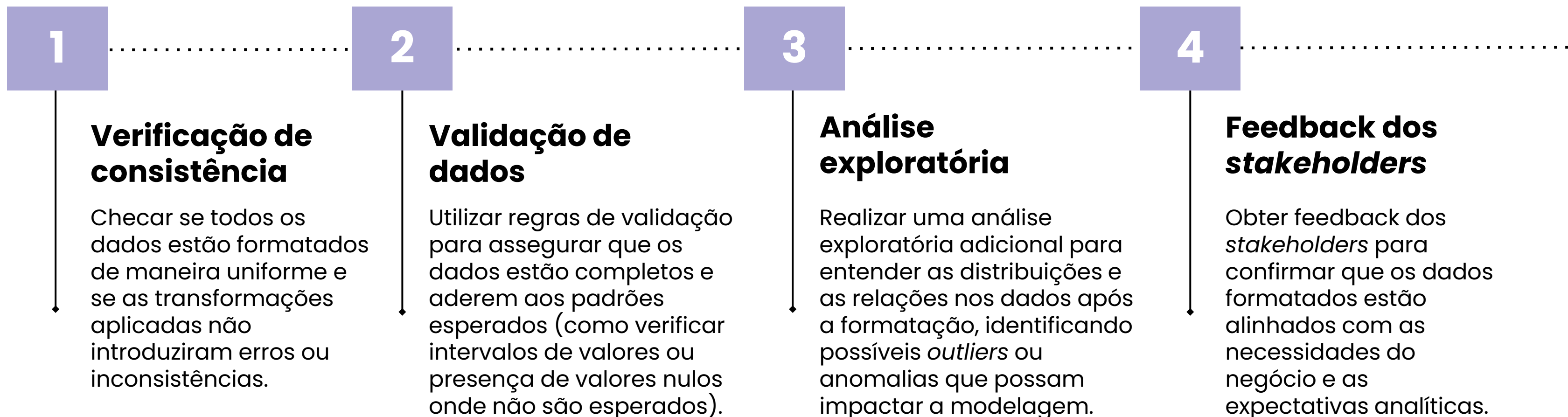
Normalização ou padronização

Aplicar técnicas para normalizar ou padronizar características numéricas, especialmente importantes para modelos que são sensíveis à escala dos dados, como muitos algoritmos de *machine learning*.

Construindo o conceito

Revisão de dados

É o processo de verificar a qualidade dos dados após a formatação e outras transformações, garantindo que estejam prontos para a próxima etapa de criação de relatórios ou modelagem de algoritmos de *machine learning*.





Vamos
fazer um
quiz

Qual é o objetivo principal da formatação de dados em Ciência de Dados?

Transformar todos os dados em texto.

Organizar e estruturar os dados para análise.

Aumentar o volume de dados disponíveis.

Criar gráficos complexos diretamente dos dados brutos.



Vamos
fazer um
quiz

Por que é importante normalizar ou padronizar dados antes de proceder com análises em Ciência de Dados?

Para garantir que todos os gráficos sejam coloridos.

Para facilitar o armazenamento de dados em diferentes formatos.

Para assegurar que a análise seja imparcial e precisa.

Para evitar o uso de softwares de análise de dados.



Vamos
fazer um
quiz

Qual é a importância da revisão de dados na fase de preparação de dados no CRISP-DM?

Para verificar se os dados podem ser publicados na internet.

Para garantir que os dados estão corretamente formatados e prontos para a modelagem.

Para converter todos os dados em formatos gráficos.

Para duplicar os dados e aumentar o tamanho do *dataset*.



© Getty Images

O que nós
**aprendemos
hoje?**

Então ficamos assim...

- 1** Vimos que é essencial formatar os dados de forma que sejam acessíveis e interpretáveis tanto por humanos quanto por máquinas. A normalização de formatos, como datas e números, é essencial para garantir análises eficazes e resultados precisos;
- 2** Compreendemos que revisar os dados após a formatação é fundamental para assegurar que estejam consistentes e prontos para a modelagem. Essa revisão é vital para identificar e corrigir quaisquer inconsistências ou erros introduzidos durante a formatação;
- 3** Vimos que dentro de uma projeto de dados entende-se a importância de todas as etapas do processo do Crisp-DM para seguir com formatação e revisão de dados que são pré-requisitos para a próxima fase de criação de relatórios e modelagem. Entende-se que seleção, limpeza, construção e integração dos dados já foram executadas anteriormente.

Saiba mais

Quer transformar informações complexas em algo fácil de entender e lembrar?

Visual Thinking e Graphic Recording são ferramentas que usam o poder dos visuais para comunicar ideias complexas de forma clara e memorável.

WORKANA. *O que é Visual Thinking?*, [s.d.]. Disponível em: <https://i.workana.com/guias/o-que-e-visual-thinking/>. Acesso em: 26 jul. 2024.

Referências da aula

KNAFLIC, C. N. *Storytelling com dados*: um guia sobre visualização de dados para profissionais de negócios. Rio de Janeiro: Alta Books, 2019.

Identidade visual: imagens © Getty Images

**Educação
Profissional
Paulista**

Técnico em
**Ciência de
Dados**