

Data science aplicada aos negócios

Prof. Dr. Álvaro Campos
Ferreira

alvaro.ferreira@idp.edu.br

Avaliação Diagnóstica

Avaliação Diagnóstica

Realizem a avaliação diagnóstica. Ela não contará na pontuação final, mas é obrigatória.

- A avaliação está no link:
<https://forms.gle/xXnqqXGccSvcYT8XA>
- A avaliação deve ser realizada até amanhã (16/08/2022)

Tratamento e limpeza de dados

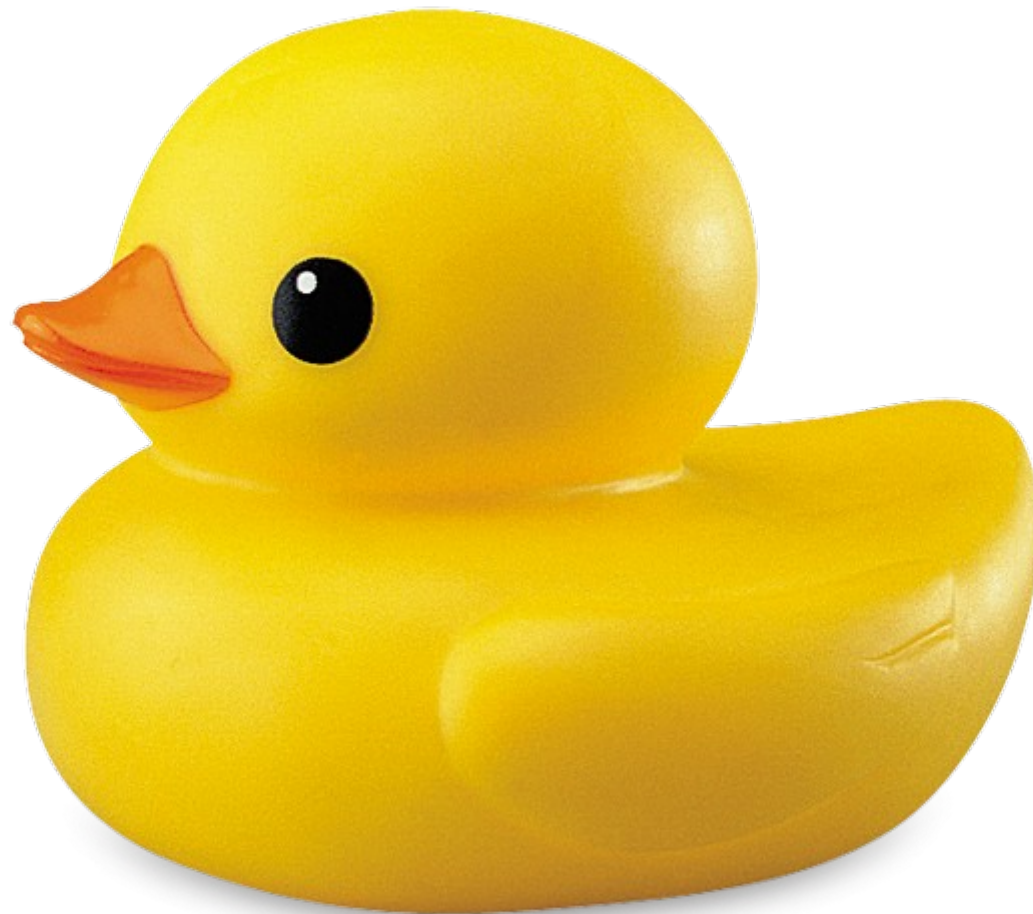
Qualidade de dados

Dimensões da qualidade de dados:

- Completude ou integridade
- Conformidade
- Validade
- Acurácia e precisão

Patinho de borracha

Sempre que possível,
vamos explicar para
o patinho o que o
código faz ou
pretende fazer.



Dados

Categorizados

- Escala Nominal
- Escala Ordinal
- Exemplos:
 - Lista de nomes
 - Lista de menções

Quantitativos

- Escala Intervalar
- Exemplos:
 - Lista de salários
 - Preços diários
 - Número de usuários

Dados

Seção transversal

- Todas as observações são realizadas ao mesmo tempo

Série temporal

- Observações realizadas ao longo do tempo

Tratamento e limpeza de dados

Qualidade de dados

Dimensões da qualidade de dados:

- Completude ou integridade
- Conformidade
- Validade
- Acurácia e precisão

Qualidade de dados

Problemas	Soluções
Erros de dados	Utilizar entrada de dados automatizada, formulários <i>web</i> para entrada de dados individuais, com checagem de integridade de dados, menus de localização e botões de opção
Dados duplicados	Redesenhar o modelo de dados e normalizar o banco de dados relacional
Dados comprometidos	Implementar uma abordagem de defesa profunda à segurança de dados
Dados faltando	Tornar campos obrigatórios nos formulários de entrada de dados

Fonte: Adaptado de Turban *et al.* (2013).

Outliers

Outliers são valores encontrados fora de um intervalo razoável com relação às outras observações. É um ponto diferente dos demais, não necessariamente um erro. Sua identificação é importante e pode possuir informações interessantes sobre o sistema.

Vieses

Vieses são desvios sistemáticos na forma de pensar, coletar ou analisar dados, dificultando ou até impossibilitando a compreensão do fenômeno em análise.

- Podem ocorrer em qualquer etapa da análise
 - ~ Coleta/ Tratamento dos dados
 - ~ Análise e visualização

Business Intelligence (BI)

A inteligência de negócios (BI) é uma metodologia de coleta, análise e interpretação de dados relacionados ao problema de negócio de um empreendimento.

Geralmente se utiliza muito de visualização de dados, especialmente utilizando “dashboards”

Key Performance Indicators (KPIs)

Métricas e pontos de dados principais para monitorar a integridade de um negócio, departamento ou processos específicos.

- Conectado em bases de dados
- Organizado na forma de *Dashboards*
- Monitoramento em tempo real

Key Performance Indicators (KPIs)

Nesse momento, o *storytelling* é uma ferramenta poderosa na tomada de decisão.

- Observe de maneira crítica o conteúdo de seu relatório para narrativas que podem ter enviesado sua análise.
- Procure determinar o contexto, potenciais conflitos e incorpore-os em sua análise.

Histórias de dados

Histórias de dados - Data Storytelling

Representar os dados não é suficiente para comunicar as informações obtidas pela análise realizada.

É preciso contar uma história com os dados.

Esse processo se tornou cada vez mais importante a medida que dados se tornaram mais prevalentes na vida cotidiana.

História de dados

Os elementos de uma história de dados são:

- Dados
- Narrativa
- Visualizações

História de dados

Como estruturar sua narrativa?

- Setup
 - Contexto
- Conflito
 - Pergunta
- Resolução
 - Resposta



INSTITUTO BRASILEIRO DE ENSINO,
DESENVOLVIMENTO E PESQUISA