



## Módulo 1

# Importância da qualidade dos dados e sua influência nas decisões gerenciais

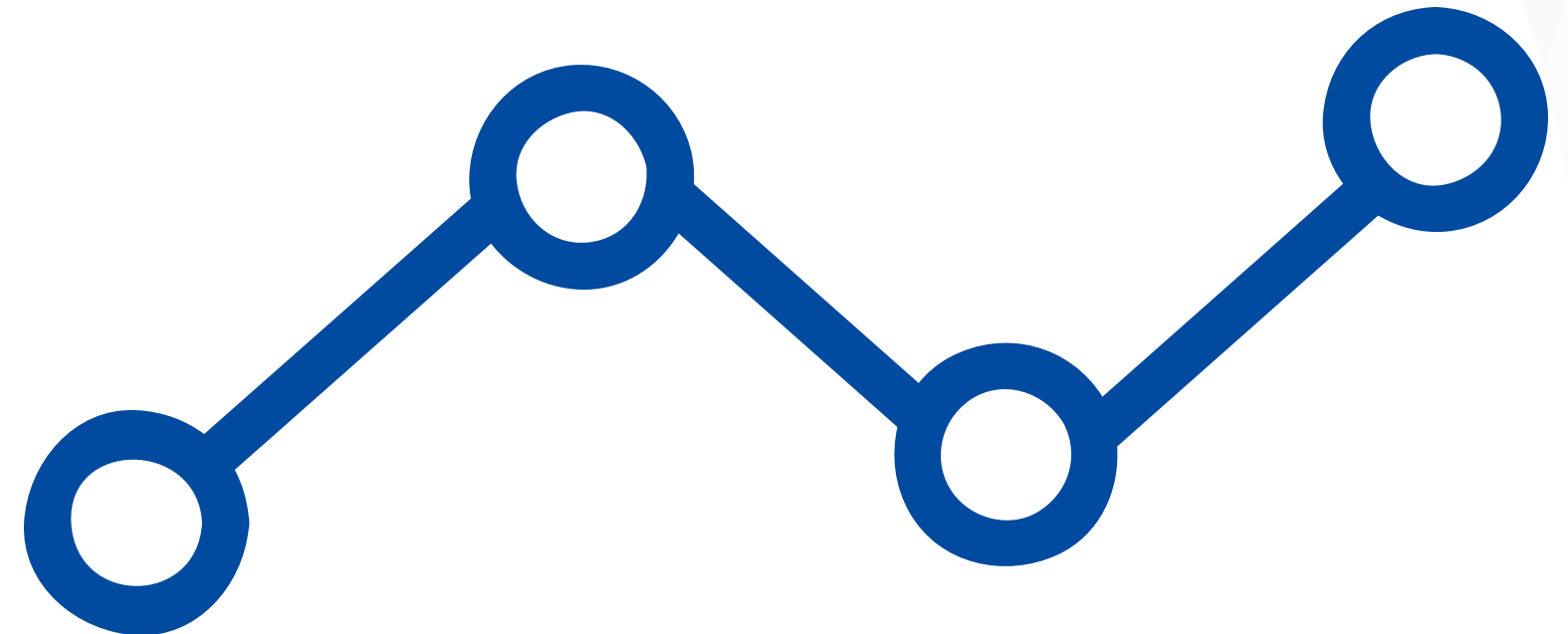


---

**Conforme falamos já nesse módulo, esse tema é fundamental.**

Se a limpeza da base de dados não é feita corretamente, isso pode prejudicar muito as análises futuras.

Tomar decisões baseadas em dados não condizentes com a realidade **pode ser tão ruim ou pior** quanto não tomar decisões baseadas em dados.



É de extrema importância que o trabalho de higienização seja feito de forma muito criteriosa e que se mantenham todos os padrões da base já pré-estabelecidos.

Diante disso, um conceito muito interessante é o da Parametrização.

Assim, o objetivo é criar padrões de classificação dos dados para que a sua base tenha o máximo das informações classificadas e qualificadas.

**Essa é uma técnica para garantir a qualidade dos seus dados no momento da análise.**



## 1. Entender os Dados

- **Conheça a fonte de Dados:**

Compreenda de onde vêm os dados, como foram coletados e como estão organizados.

- **Compreenda o Contexto:**

Entenda o propósito dos dados e como eles serão utilizados. Isso ajudará a determinar quais dados são relevantes e quais podem ser descartados.

## 2. Remoção de Duplicatas

- **Identifique e Elimine Registros Duplicados:**

Use ferramentas de deduplicação para identificar e remover registros duplicados que podem inflar artificialmente os resultados ou análises.





### 3. Corrigir Erros de Formatação

- **Padronize os Formatos:**

Certifique-se de que os dados seguem um formato consistente, como datas no mesmo formato (por exemplo, DD/MM/YYYY), números sem símbolos indesejados, e nomes com capitalização correta.

- **Remova Caracteres Indesejados:**

Elimine espaços em branco extras, símbolos ou caracteres especiais que não devem estar presentes nos dados.

---

### 4. Correção de Dados Incorretos

- **Verifique inconsistências:**

Identifique valores que estão fora do esperado ou que não fazem sentido dentro do contexto dos dados.

- **Remova Caracteres Indesejados:**

Elimine espaços em branco extras, símbolos ou caracteres especiais que não devem estar presentes nos dados.





## 5. Unificação de Dados Preditiva

- **Padronize os Formatos:**  
Consolide informações que possam estar separadas em múltiplos campos, como “Nome” e “Sobrenome”.
- **Agrupe Categorias:**  
Em situações em que existam múltiplas categorias semelhantes, considere agrupá-las para simplificar a análise.

## 6. Remoção de Dados Irrelevantes

- **Identifique Dados Desnecessários:**  
Remova colunas ou registros que não são relevantes para a análise ou o objetivo do projeto.
- **Agrupe Categorias:**  
Em situações em que existam múltiplas categorias semelhantes, considere agrupá-las para simplificar a análise.



## 7. Documentação do Processo

- **Registre as Alterações:**

Documente todas as etapas e as decisões tomadas durante a higienização dos dados.

Isso ajuda a manter a transparência e facilita o trabalho de futuras revisões ou auditorias.

## 8. Automatização e Ferramentas

- **Utilize Ferramentas de Limpeza de Dados:**

Ferramentas como OpenRefine e Trifacta, ou até funções de manipulação de dados em Excel podem facilitar a higienização de dados.

- **Automatize Processos Repetitivos:**

Quando possível, automatize etapas da limpeza de dados para garantir consistência e economia de tempo.

## 9. Revisão e Validação Final

- **Valide os Resultados:**

Após a limpeza, revise os dados para garantir que as correções foram eficazes e que os dados estão prontos para análise.



---

**Dessa forma, você vai conseguir garantir que os dados estarão sendo analisados da forma mais precisa e correta possível.**

Com todas as informações estudadas até aqui, aconselho a especialização cada vez maior no universo dos dados. Estamos vivendo o que está sendo chamado de segunda era digital.

Hoje, não temos mais problemas de falta de dados ou informações. **O grande desafio está em conseguir gerar insights** e tomar decisões com a massificação cada vez maior da informação e a certeza de que aqueles dados correspondem à realidade.

---

