

I Seminário de Extensão do Programa UFMS Digital

Data Quality aplicada a Tomada de Decisões Baseada em Dados

Trilha Temática IV - Ciência de Dados Data: 31/05/2024





Autoria





Brenik Domingos de Oliveira - <u>brenik.oliveira@ufms.br</u> Acadêmico do Curso Tecnologia em Ciências de Dados

Raul da Costa Vieira - <u>raul.c.vieira@ufms.br</u>
Acadêmico do Curso Tecnologia em Ciências de Dados



Sobre o Minicurso



Sobre o minicurso





1. Conteúdos abordados

- 1.1. Uso de dados
- 1.2. Data Quality
 - 1.2.1. Visão Geral
 - 1.2.2. Padronização
- 1.3. Data-driven
 - 1.3.1. Mapeamento de dados
 - 1.3.2. Tomada de Decisão baseada em dados

Sobre o minicurso





1. Objetivos

- a. Demonstrar o uso de Data Quality para Ciência de Dados
- b. Exemplificar a aplicação de Data Quality no cenário de Tomada de Decisões Baseada em Dados

Organização





- 1. Parte 1 Uso de Dados (30 min.)
- 2. Parte 2 Data Quality (30 min.)
 - 2.1. Visão Geral
 - 2.2. Padronização
- 3. Parte 3 Data-driven (30 min.)
 - 3.1. Mapeamento de dados
 - 3.2. Tomada de Decisão baseada em dados









Gestores ao redor do mundo em 2023

Dizem que tomada de decisão baseada em dados é um Importante objetivo em seus programas de dados

PRECISELY, 2023

Tomada de decisão baseada em dados se tornou consistente no mundo corporativo.







"a construção do código não deve apenas funcionar, precisa estar bem projetado."

² David J. Malan





Na Ciência de Dados o gerenciamento de dados não deve apenas funcionar, ele precisa ser bem projetado.







"Entender os fatores que afetam a performance da tomada de decisões é uma nova contribuição do campo de análise de dados"

³ LIN et al., 2022



Parte 1 Uso de dados







Compreendendo o cenário



OF DATA





"De acordo com a Statista, a quantidade total de dados previstos para serem criados, capturados, copiados e consumidos globalmente em 2023 é de 120 Zettabytes, o número projetado para crescimento é de 181 Zettabytes até 2025."

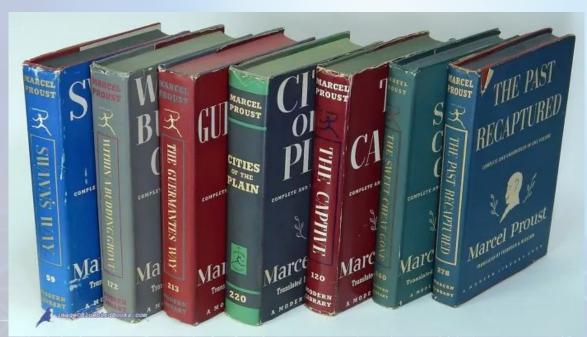
4 DOMO, 2023

181 Zettabytes = 181.10¹² Gigabytes





Curiosidade!!!



1 letra "a" = 1 byte

O Livro "Em busca do tempo perdido" é o maior romance conhecido, com aprox. 4.215 páginas, 1.267.069 palavras na versão em inglês.

Aprox. 7,5 Mi. letras. ≅ 0,0075 Gigabytes.





Curiosidade!!!



Expedição Apollo 11 (1969)

Computador de bordo com 32,768 bits de RAM.

4.096 letras (aprox. 600 palavras).





A indústria passa por um intenso processo de coleta de dados...





Porém, a coleta de dados demanda organização





"(...) até 2022, 70% das organizações acompanharão rigorosamente os níveis de *data quality* via métricas, aprimorando-os em 60% para reduzir significativamente os riscos operacionais e custos."

⁵ GARTNER, 2021







May 20, 2021 Contributor: La

Data and analytics leaders who share times more measurable economic be

How to Improve Your Data Quality Gartner. Newsroom Topics Media Contacts Media Red Media Contacts Media Red

July 14, 2021 Contributor: Mana

D&A leaders must take pragmatic and their enterprise data quality if they war organizations' digital transformation.

LONDON, U.K., May 22, 2023

Gartner Identifies 12 Actions to Improve Data Quality

Gartner Analysts Are Exploring Data Trust, Governance and Privacy at the Gartner Data & Analytics Summit, May 22-24, in London, U.K.





"Anualmente, a má qualidade de dados (poor data quality) custa uma média de U\$12,9 Milhões às organizações."

5 GARTNER, 2021









Visão Geral





- 1) Qual padrão de dados é considerado de alta qualidade?
- 2) Quais fatores influenciam a qualidade de dados?
- 3) Como a qualidade de dados influencia a tomada de decisões?

³ LIN et al., 2022





Qualidade Intrínseca

Relacionada aos aspectos objetivos, imutáveis (não variáveis) dos dados.

Qualidade Contextual

Depende da interação "tarefa-pessoa" com relação a qualidade demandada para aquela tarefa em particular.

⁶ MAHANTI, 2018





Intrínseca

- Acurácia
- Consistência
- Completude
- "Prontidão" (Defasagem, Latência)

Contextual

- Relevância
- Valor Agregado
- Quantidade
- Credibilidade
- Acessibilidade
- Reputação...



⁶ MAHANTI, 2018





PADRONIZAÇÃO

Um exemplo de critério para qualidade de dados!

Mensurando a Qualidade



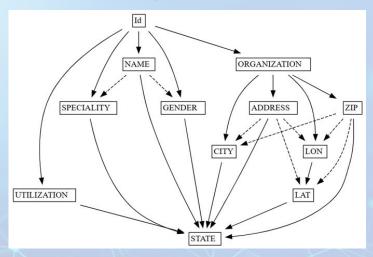


Dependência Funcional

Cliente (ID →Nome, Organização, Endereço, CEP, Cidade..)

- "Cliente" possui um esquema relacional
- Para uma <u>instância</u> qualquer no banco de dados de "Cliente", deve haver um atributo que satisfaz as condições da dependência funcional.

Gráfico de Dependência Funcional



⁷ MICROSOFT, 2024





Dependência Funcional

Functional dependencies

Consider FD: customer(NI# \rightarrow name, AC, phn, street, city, zip)

• NI# is a key: there is a unique record for each distinct NI#.

Consider instance I of R:

NI#	AC	phn	name	street	city	zip
SC1234566	131	1234567	M. Smith	Mayfield	EDI	EH4 8LE
SC1234566	020	1234567	M. Smith	Portland	LDN	W1B 1JL

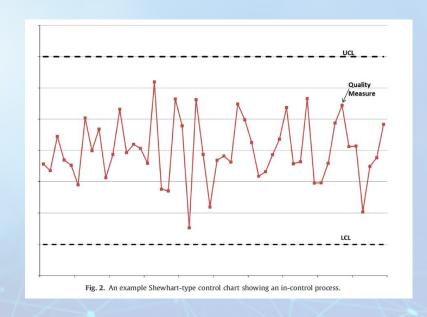
- I does not satisfy (or <u>violates</u>) the FD.
- For SC1234566, at least one of the records must be dirty.





Uso de Carta de Controle para Mensurar o desvio da qualidade de dados.

Exemplo aplicado em Supply Chain utilizando Controle Estatístico de Processo (CEP, SPC em inglês).



⁹ HAZEN et al., 2014



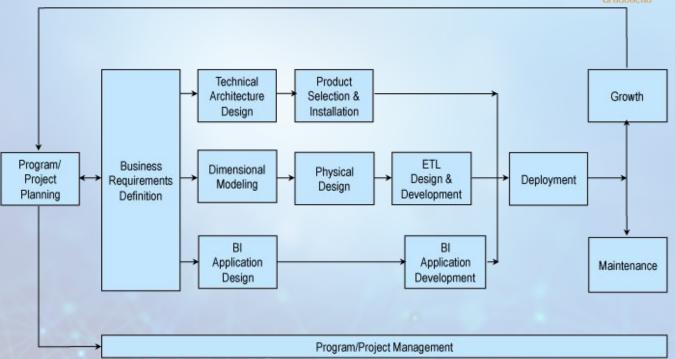












Kimball Lifecycle approach of Data Warehouse/Business Intelligence



Parte 3 Data-driven



Parte 3 - Data-driven







Uma introdução à metodologia de mapeamento adotada em Data Quality

Parte 3 - Data-driven





Mapeamento com uso do ciclo PDCA para direcionar a tomada de decisão



Parte 3 - Data-driven







O ciclo PDCA é contínuo.

Mudanças se tornam visíveis desde o início do processo e transformações macro são perceptíveis a longo prazo.

11 WERKEMA, 1995





- 1 Identificação do Problema
 - 2 Análise do Fenômeno
 - 3 Análise do Processo
 - 4 Plano de Ação

PDCA

Plan (planejamento): Inicia-se com perguntas, tais como:

Objetivo: Qual o problema e como atingir o objetivo?

Métodos: Como realizar as mudanças? Há necessidade de novos colaboradores ou mudança de processo?

Métricas: Como projetar as métricas?







Do (fazer ou executar): Iniciar o processo conforme planejamento (Plano de Ação).

- Capacitação da equipe, para implementação proposta.
- Acompanhamento dos processos e análise das métricas, para entender as mudanças aos dados apresentados no cenário anterior e atual.







Check (Verificação): Junção dos dados coletados após implementação do novo processo.

- Comparação de cenário anterior e atual
- Realização de ajustes necessários, com inovação constante e estratégica, em busca de melhoria das métricas.







Act (Agir): Correções, mudando desde a cultura empresarial.

 Observar mudanças negativas, analisar os pontos que se tornaram os ofensores, realizar a manobra corretiva e implantar a nova cultura no local.





Livros de diferentes métodos, para aprofundar os conceitos apresentados.

Análise SWOT: quando usar e como fazer

Gerenciamento de Projetos com PDCA

Inferência Estatística - Como Estabelecer
Conclusões com Confiança no Giro do
PDCA e DMAIC







Cases de aplicação da qualidade e data-driven em estratégias empresariais





Para ilustrar essa aplicação será abordado o storytelling de dados, ferramentas de visualizações de dados e exemplos práticos de tomadas de decisões, demonstrando resultados de cases reais.







12 Storytelling:

Ao adaptar os dados a um modo de apresentação, teremos diferentes possibilidades de construir uma narrativa, com uso de áudios, vídeos, gráficos e outros.





Como iniciar o storytelling em dados?



- Descrição do problema: com mapeamento e ações de verificação (PDCA).
- Apresentação de forma visual das métricas.
- Uso de ferramentas de visualização como Tableau, Power BI e Excel.



Power Bl





Exemplo de Storytelling aplicado no mundo corporativo:

Case de sucesso Heineken 2016



13 Heineken | The Cliché





Case de Visualização de dados utilizando a ferramenta
Power BI - (Visualização e apresentação de métricas, após todos passos anteriores)



14 XPERIUN, 2023



ANÁLISE DE VENDAS

Ano

2018

2019

Mês

Todos



R\$171 Mi Total Vendido



R\$89 Mi Total Lucro



52,01% % Margem



2017

23.189 Notas Emitidas



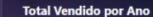
7.655
Cobertura Clientes

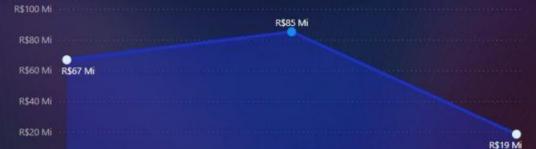












Total por Continente

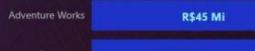
RS0 Mi



Total por País e Continente



Total por Marca

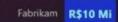








Southridge Video R\$14 Mi





Northwind Trad...



Encerramento



Encerramento





Conclui-se que o conteúdo apresentado, pretendeu introduzir o ouvinte aos fundamentos de Data Quality, de maneira contextual, orientada à tomadas de decisões. Por meio de cases reais, embasados na solução de problemas, demonstrou-se os resultados obtidos e a melhoria contínua de processos de inovação.





¹PRECISELY & DREXEL. **2023 Data Integrity Trends and Insights report**. p. 6. 2023. Disponível em: https://www.precisely.com/resource-center/analystreports/lebow-report-2023>.

²MALAN, David J. **CS50**, **Lecture 1**. Disponível em: < https://video.cs50.io/ywg7cW0Txs4?screen=S5uYVc7bm0o>.

³LIN at al. Overview of Data Quality: Examining the Dimensions, Antecedents, and Impacts of Data Quality. 2022. Disponível em: https://link.springer.com/article/10.1007/s13132-022-01096-6>.





⁴DOMO. **Data never sleeps 11.0**. Disponível em: <<u>https://www.domo.com/learn/infographic/data-never-sleeps-11</u>>

⁵GARTNER. How to Improve Your Data Quality. 2021.

Disponível em:

https://www.gartner.com/smarterwithgartner/how-to-improve-your-data-quality

⁶MAHANTI, Rupa. Data Quality: dimensions, measurement, strategy, management and governance. 2018.





- ⁷ MICROSOFT. **Tutorial: Limpar dados com dependências funcionais**. 2024. Disponível em: https://learn.microsoft.com/pt-br/fabric/data-science/tutorial-data-cleaning-functional-dependencies>
- § GEERTS, Floris. Foundations of Data Quality. Disponível em:

 http://adrem.uantwerpen.be/sites/default/files/tutorial-general-dq.pd

 f
- ⁹ HAZEN, Benjamin T., *et al.* Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. 2014. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0925527314001339>





- 10 KIMBALL GROUP. **Kimball DW/BI Lifecycle Methodology**.
 201_. Disponível em: https://www.kimballgroup.com/data-ware
 house-business-intelligence-resources/kimball-techniques/dw-bi-lifecycle-method/>
- ¹¹ WERKEMA, Maria C. C. **As ferramentas da Qualidade no gerenciamento de processos**. Vol 1, Ferramentas da Qualidade. Escola de Engenharia, UFMG/MG. 1995.
- ¹² MAILCHIMP. Storytelling com dados: Transforme números em histórias cativantes. Disponível em: < https://mailchimp.com/pt-br/resources/data-storytelling/>





13 HEINEKEN. **The Cliché**. Disponível em: < https://youtu.be/w
rmKl_HV4-A?si=91qe6avPp3JusXCc>

14 XPERIUN. Exemplos de Relatórios e Dashboards com Power BI. 2023. Disponível em: https://xperiun.com/blog/exemplos-de-relatorios-com-power-bi/



AGEAD Agência de Educação Digital e a Distância

