

**Aula 00 - Profs. Diego
Carvalho e Emannelle
Gouveia**

*SEFAZ-PR (Auditor Fiscal) Análise de
Dados - Conceitos e Aplicações - 2025*

(Pós-Edital)
Autor:

Diego Carvalho

03 de Abril de 2025

Índice

1) Apresentação do Prof. Diego Carvalho - Informática	3
2) Análise de Informações - Dados, Atributos e Métricas - Teoria	5
3) Análise de Informações - Dados, Atributos e Métricas - Questões Comentadas - Multibancas	20
4) Análise de Informações - Dados, Atributos e Métricas - Lista de Questões - Multibancas	25



APRESENTAÇÃO DO PROFESSOR

PROF. DIEGO CARVALHO

FORMADO EM CIÊNCIA DA COMPUTAÇÃO PELA UNIVERSIDADE DE BRASÍLIA (UNB), PÓS-GRADUADO EM GESTÃO DE TECNOLOGIA DA INFORMAÇÃO NA ADMINISTRAÇÃO PÚBLICA E, ATUALMENTE, AUDITOR FEDERAL DE FINANÇAS E CONTROLE DA SECRETARIA DO TESOURO NACIONAL.

ESTRATÉGIA CONCURSOS

 PROFESSOR DIEGO CARVALHO - [WWW.INSTAGRAM.COM/PROFESSORDIEGOCARVALHO](https://www.instagram.com/professordiegovalho)



Sobre o curso: galera, todos os tópicos da aula possuem Faixas de Incidência, que indicam se o assunto cai muito ou pouco em prova. Diego, se cai pouco para que colocar em aula? Cair pouco não significa que não cairá justamente na sua prova! A ideia aqui é: se você está com pouco tempo e precisa ver somente aquilo que cai mais, você pode filtrar pelas incidências média, alta e altíssima; se você tem tempo sobrando e quer ver tudo, vejam também as incidências baixas e baixíssimas. *Fechado?*

INCIDÊNCIA EM PROVA: BAIXÍSSIMA

INCIDÊNCIA EM PROVA: BAIXA

INCIDÊNCIA EM PROVA: MÉDIA

INCIDÊNCIA EM PROVA: ALTA

INCIDÊNCIA EM PROVA: ALTÍSSIMA

Além disso, essas faixas não são por banca – é baseado tanto na quantidade de vezes que caiu em prova independentemente da banca quanto nas minhas próprias avaliações sobre cada assunto.



#ATENÇÃO

Avisos Importantes



O curso abrange todos os níveis de conhecimento...

Esse curso foi desenvolvido para ser acessível a **alunos com diversos níveis de conhecimento diferentes**. Temos alunos mais avançados que têm conhecimento prévio ou têm facilidade com o assunto. Por outro lado, temos alunos iniciantes, que nunca tiveram contato com a matéria ou até mesmo que têm trauma dessa disciplina. A ideia aqui é tentar atingir ambos os públicos - iniciantes e avançados - da melhor maneira possível..



Por que estou enfatizando isso?

O **material completo** é composto de muitas histórias pessoais, exemplos, metáforas, piadas, memes, questões, desafios, esquemas, diagramas, imagens, entre outros. Já o **material simplificado** possui exatamente o mesmo núcleo do material completo, mas ele é menor e mais objetivo. *Professor, eu devo estudar por qual material?* Se você quiser se aprofundar nos assuntos ou tem dificuldade com a matéria, necessitando de um material mais passo-a-passo, utilize o material completo. Se você não quer se aprofundar nos assuntos ou tem facilidade com a matéria, necessitando de um material mais direto ao ponto, utilize o material simplificado.



Por fim...

O curso contém diversas questões espalhadas em meio à teoria. Essas questões possuem um comentário mais simplificado porque **têm o único objetivo de apresentar ao aluno como bancas de concurso cobram o assunto previamente administrado**. A imensa maioria das questões para que o aluno avalie seus conhecimentos sobre a matéria estão dispostas ao final da aula na lista de exercícios e **possuem comentários bem mais abrangentes**.



DADOS

Conceitos Básicos

INCIDÊNCIA EM PROVA: ALTÍSSIMA

DADO

Dados são correspondências de um atributo, característica ou propriedade que, sozinho, não tem significado.

Dados são elementos brutos, sem significado, desvinculados da realidade.

Dados são simples observações sobre o estado do mundo.

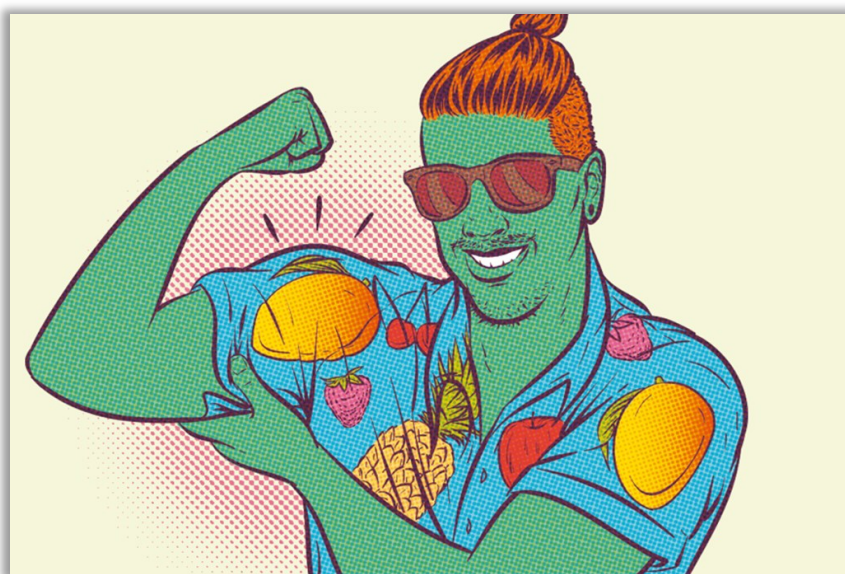
Dados são um conjunto de fatos objetivos e discretos sobre eventos.

Dados são a menor partícula estruturada que compõe uma informação.

O que podemos concluir? **Podemos concluir que dados são fatos ou estatísticas coletadas para servir de referência ou análise.** O grande lance quando se pensa em dados é que eles não têm significado ou semântica quando vistos isoladamente – eles são brutos, simples e desvinculados de contexto/realidade. Professor, ainda não entendi! Cara... imagine que sem nenhum contexto, eu te mostro um relatório com uma tabela escrito a palavra abaixo em todas as linhas e colunas.

manga

Você consegue tirar alguma informação desse relatório? Não, você pode até inferir que se trata – por exemplo – de algum relatório sobre frutas. No entanto, você não pode afirmar com certeza! *Por quê?* Porque isso pode ser tanto um relatório de um hortifruti quanto um relatório de um camiseteria. *E aí, sem nenhum contexto, tem como saber?* Não, é impossível! *Por quê?* Porque isso é apenas um dado: **bruto, simples, objetivo e desvinculado da realidade.**



Dados são sucessões de fatos brutos, que não foram organizados, processados, relacionados, contextualizados, avaliados ou interpretados, representando apenas partes isoladas de eventos, situações ou ocorrências. Eles se constituem de unidades básicas a partir das quais informações poderão ser elaboradas ou obtidas. **Além disso, eles são frequentemente quantificados, facilmente estruturados, facilmente transferíveis e facilmente obtidos por máquinas.**

Quando os dados são organizados, processados, relacionados, contextualizados, avaliados ou interpretados, ele passa a ser uma informação, conforme veremos no tópico seguinte.

(COREN/PA – 2013) Nos Sistemas de Informação, dados são:

- a) informações que possuem significado isoladamente.
- b) representações do conhecimento contextualizado.
- c) definições do real sentido das coisas.
- d) correspondências de um atributo, característica ou propriedade que, sozinho, não tem significado.

Comentários: dados são correspondências de um atributo, característica ou propriedade que, sozinho, não tem significado (Letra D).

(CEFET – 2014) Para os sistemas de informação, definem-se dados como:

- a) melhoria contínua de processos organizacionais.
- b) resultado da análise do ambiente de negócios.
- c) fatos organizados e apresentados de forma útil
- d) sequências de fatos ainda não analisados.
- e) decisões tomadas pelos gestores da organização.

Comentários: dados são sequências de fatos ainda não analisados, isto é, brutos (Letra D).

(CGE/PB – 2008 – Letra B) Dados são elementos brutos, sem significado, desvinculados da realidade.

Comentários: eles realmente são elementos bruto, sem significado e desvinculados da realidade (Correto).

Agora precisamos ver a diferença entre dados estruturados e não estruturados. **Basicamente, dados estruturados são rigidamente organizados e formatados de uma maneira que seja facilmente pesquisável em bases de dados relacionais.** Professor, você falou grego agora! Calma, tudo isso será detalhado futuramente, mas – simplificando – bases de dados relacionais nada mais são que tabelas (linhas e colunas) que armazenam dados.



Logo, se eu consigo organizar um conjunto de dados em um formato que possa ser exibido em uma tabela, então esse é um conjunto de dados estruturados. *Certinho?* Já os dados não-estruturados não possuem um formato ou uma organização predefinida, tornando muito mais difícil sua coleta, processamento e análise. Sim, a maneira como os dados são coletados, processados e analisados depende completamente do formato em que eles estão.



Dados Estruturados

INCIDÊNCIA EM PROVA: ALTÍSSIMA

Dados estruturados são frequentemente categorizados como dados quantitativos. No seu trabalho, provavelmente você está acostumado a trabalhar com eles. Tente se lembrar! *Os dados que você trabalha podem ser organizados em linhas e colunas de bases de dados relacionais?* Se sim, você trabalha com dados estruturados! Exemplos: nomes, datas, endereços, números de cartão de crédito, tabelas do Excel, metadados (que são dados sobre dados), entre outros.

Galera, o bacana dos dados estruturados é que eles podem ser facilmente compreendidos e manipulados por linguagens de máquinas – pode-se manipular dados estruturados com relativa rapidez e essa é uma de suas maiores vantagens: **você pode ordenar, agregar, separar, juntar, inserir, atualizar, deletar e consultar dados e tabelas com facilidade.** A tabela que vimos no exemplo do tópico anterior é um exemplo de dado estruturado.

NOME	SEXO	IDADE	SALÁRIO	NATURALIDADE
Alice	Feminino	18	R\$ 8.500,00	DF
Bernardo	Masculino	26	R\$ 2.000,00	GO
Carolina	Feminino	30	R\$ 9.500,00	TO
Davi	Masculino	22	R\$ 1.500,00	RR
Eduarda	Feminino	32	R\$ 7.500,00	SP
Felipe	Masculino	24	R\$ 3.000,00	RJ
Gabriela	Feminino	28	R\$ 6.500,00	PA
Hugo	Masculino	20	R\$ 1.000,00	RS

Note que, para cada linha dessa tabela, nós temos sempre as mesmas cinco colunas, com os mesmos atributos e os atributos são sempre do mesmo tipo para cada coluna (Ex: Salário é sempre um Número; Sexo é sempre uma Palavra; Naturalidade são sempre duas letras; entre outras). *Viram como é rígido?* E tem mais: eu não consigo inserir um novo atributo na tabela sem antes modificá-la (Ex: CPF) – caso eu queira inserir essa informação, devo adicionar uma coluna!



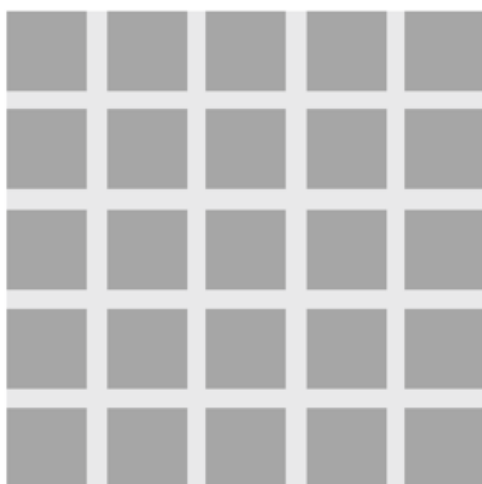
Dados Não-Estruturados

INCIDÊNCIA EM PROVA: ALTÍSSIMA

Dados não-estruturados são dados que não possuem um formato ou organização predefinida. Assim como os dados estruturados são frequentemente categorizados como dados quantitativos, os dados não-estruturados são frequentemente categorizados como qualitativos, e não podem ser processados e analisados utilizando ferramentas e métodos convencionais. Exemplos: texto, vídeo, áudio, atividades de mídias sociais, entre outros.

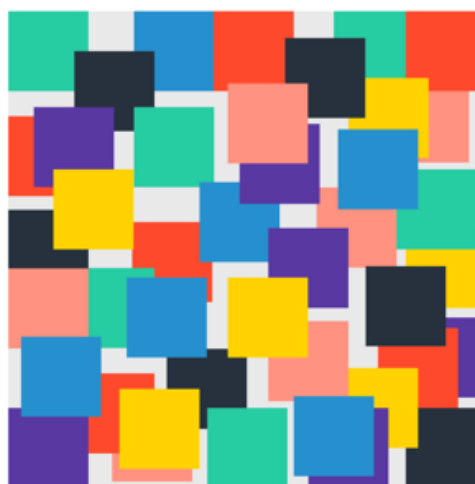
Dados não-estruturados são difíceis de desconstruir porque não têm um formato ou modelo predefinido, significando que não podem ser organizados em uma base de dados relacional. Mais de oitenta por cento de todos os dados gerados atualmente são considerados não-estruturados, e esse número só tende a continuar crescente com o surgimento da Internet das Coisas – tecnologia que permite que vários objetos estejam conectados à internet.

Structured data



Database, CRM, ERP

Unstructured data



Text, audio, videos



Dados Semiestruturados

INCIDÊNCIA EM PROVA: BAIXA

Dados semiestruturados é uma mistura dos dados estruturados com os dados não-estruturados. Eles não estão de acordo com a estrutura formal dos modelos de dados associados com bancos de dados relacionais ou outras formas de tabelas de dados, mas contêm tags ou outros marcadores para separar elementos semânticos e impor hierarquias de registros e campos dentro dos dados. Temos os dados junto com um esquema de representação parcialmente organizados!

Pensem em um dado que não pode ser perfeitamente encaixado em uma base de dados relacional, mas ele também não é completamente sem formato. Um exemplo clássico é o XML, que é uma linguagem que permite autodescrever um dado! Em outras palavras, ela apresenta o dado e, junto com ele, apresenta uma autodescrição. Eu sei que está difícil de imaginar, mas eu vou dar um exemplo...

```
<xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<MUSICAS xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <MUSICA>
    <NOME> THE SCRIPT FOR MY REQUIEM </NOME>
    <BANDA> BLIND GUARDIAN </BANDA>
    <ALBUM> IMAGINATIONS FROM THE OTHER SIDE </ALBUM>
  </MUSICA>
  <MUSICA>
    <NOME> STARLESS </NOME>
    <BANDA> KING CRIMSON </BANDA>
    <ALBUM> MELTDOWN </ALBUM>
  </MUSICA>
  <MUSICA>
    <NOME> CALADVWCH </NOME>
    <BANDA> DARK AVENGER </BANDA>
    <ALBUM> ALIVE IN THE DARK </ALBUM>
  </MUSICA>
</MUSICAS>
```

Esse é um exemplo de dado escrito em linguagem XML! **Notem que os dados em si são os que estão com a cor preta – todo o resto são dados que auxiliam a descrever os dados.** Apesar de, nesse exemplo, termos estruturas idênticas para cada música/banda/álbum, nada impede que haja estruturas diferentes para cada um. Por essa razão, não é possível dizer que se trata de dados completamente estruturados ou dados completamente não-estruturados. *Então, o que ele é?*



XML descreve dados semiestruturados, porque veja que é possível identificar marcações em seu código, portanto você poderia dizer que são dados estruturados. No entanto, essas tags não



possuem uma ordem ou padrão específico. Você pode organizá-las da maneira que achar adequado. **Se meus dados possuem marcações identificáveis, mas sem ordem e padrão fixo, trata-se de dados semiestruturados – assim como HTML e outras linguagens de marcação!**

Por fim, é possível classificar os tipos de dados quanto à capacidade de terem suas estruturas descritas implicitamente ou explicitamente. Vejamos:

- **Dados Estruturados:** possuem uma estrutura predefinida, logo ela pode ser descrita de forma explícita – você bate o olho e identifica um padrão definido;
- **Dados Não-Estruturados:** não possuem uma estrutura predefinida, logo ela não pode ser descrita (implícita ou explicitamente) – você bate o olho e não identifica um padrão definido;
- **Dados Semiestruturados:** possuem parte de sua estrutura predefinida, logo alguns dados podem ser descritos implicitamente – você bate o olho e consegue inferir um padrão nos dados.

(TCE/PB – 2018) Com relação a dados estruturados e não estruturados, assinale a opção correta.

a) Dados não estruturados podem ser caracterizados por arquivos de diversos tipos — textos, imagens, vídeos, entre outros —, cujas estruturas não são descritas implicitamente.

b) Por padrão, documentos do tipo XML (eXtensible Markup Language) são estruturados.

c) Dados não estruturados de um mesmo grupo possuem as mesmas descrições e, conseqüentemente, os mesmos atributos.

d) Por padrão, dados não estruturados são organizados em blocos semânticos.

e) A alta heterogeneidade facilita as consultas aos dados não estruturados, desde que estes estejam ligados por ponteiros.

Comentários: (a) Correto, as estruturas não são descritas implicitamente porque sequer há estrutura; (b) Errado, são semiestruturados; (c) Errado, são os dados estruturados de um mesmo grupo que possuem as mesmas descrições e atributos; (d) Errado, dados estruturados são organizados em blocos semânticos iguais; (e) Errado, ela dificulta as consultas aos dados não-estruturados (Letra A).

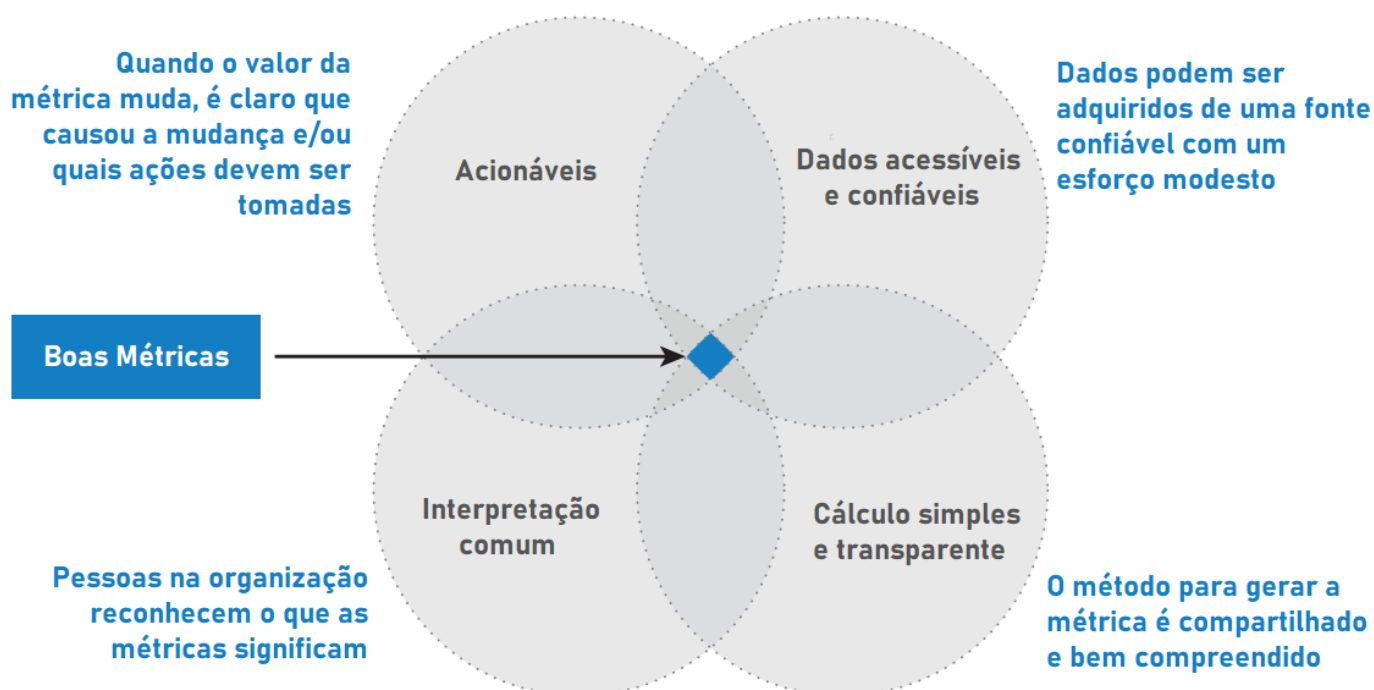


Métricas de Dados

INCIDÊNCIA EM PROVA: BAIXA

Métricas são os valores que você usa para julgar o desempenho. Eles também são o ponto focal de qualquer produto de dados, sendo o reflexo numérico do comportamento do mundo real que sua organização deseja melhorar, evitar ou moldar. As métricas podem criar foco e alinhamento em uma organização, fornecendo clareza sobre como é a melhoria. As métricas também podem criar comportamentos contra-intuitivos ou contrários aos objetivos organizacionais.

Apesar da importância de escolher as métricas certas, é comum não pensar muito sobre o que medir. Erros comuns na escolha de métricas incluem: métricas simples demais, métricas complexas demais, grande quantidade de métricas, métricas enganosas, entre outros. Para evitar essas armadilhas muito comuns e escolher a melhor métrica possível, podemos considerar a seguinte estrutura de métricas que equilibra quatro fatores ou dimensões:



Você consegue imaginar um jogo de basquete sem aro ou uma partida de tênis sem a marcação das linhas? Na ausência de objetivos claros e de uma forma de medir o progresso, os esportes seriam desorganizados. Isso também é verdade para organizações: na ausência de métricas de sucesso organizacional claramente comunicadas, os funcionários não têm como saber se estão no caminho certo, contribuindo de maneira significativa ou apenas perdendo o controle.

Para criar uma cultura de fluência de dados, os líderes devem estabelecer as principais métricas organizacionais que orientam os esforços individuais e da organização, mas o que faz uma boa métrica?



Metas desalinhadas, comportamentos distorcidos e uma sensação equivocada de sucesso são problemas causados pelo uso de métricas ruins em uma organização. Um exemplo comum é o departamento de atendimento ao cliente que rastreia chamadas atendidas sem levar em consideração a qualidade da experiência do chamador. *E se a empresa atendeu a ligação depois de o cliente esperar por uma hora? E se a empresa atendeu a ligação, mas não resolveu o problema?*

O primeiro passo para estabelecer o que constitui uma boa métrica é entender o que você deseja alcançar. As métricas podem servir a dois propósitos: identificar problemas e medir o desempenho. Quando o objetivo é identificar problemas e apontar áreas de ineficiência e ineficácia operacional, definir a métrica certa requer um pouco de trabalho, dado que necessita que se determine quais evidências podem ser encontradas e como exatamente elas aparecem.

Quando o objetivo é medir o desempenho, as métricas de sucesso certas se concentram em medidas que podem ser controladas e onde a melhoria na métrica é uma coisa inequivocamente boa. Depois de ter clareza sobre a finalidade para a qual você usará determinada métrica, o próximo passo é escolher uma boa métrica. Boas métricas incluem várias os fatores ou dimensões vistas na página anterior. *E o que são boas métricas?*

BOAS MÉTRICAS SÃO CRIADAS A PARTIR DE DADOS ACESSÍVEIS E CONFIÁVEIS, FÁCEIS DE COLETAR DE UMA FONTE CONFIÁVEL	Cada métrica leva tempo e esforço para ser construída. Antes de escolher a métrica, vale a pena fazer perguntas à fonte de dados: os dados são coletados de forma a capturar com precisão a realidade? Os dados são difíceis de coletar ou atrasados de forma a prejudicar seu valor? Os consumidores encontrarão motivos para questionar a fonte de dados? Para que os resultados tenham validade e para que uma cultura de engajamento de dados surja em sua organização, os produtores de dados devem ter acesso a dados confiáveis.
BOAS MÉTRICAS SÃO TRANSPARENTES E GERADAS POR CÁLCULOS COMPREENSÍVEIS	Para que as métricas identificadas por uma organização moldem a prática individual e coletiva, elas devem ser facilmente compreendidas, dando transparência de como foram desenvolvidas. Além disso, boas métricas são melhor expressas como um número ou porcentagem. Os números também devem conter pelo menos uma unidade de medida que caracterize a coisa que está sendo contada. As unidades de medida permitem que os analistas de dados expressem consistentemente os resultados de um processo de medição. O uso de duas medidas pode auxiliar nesse processo.
BOAS MÉTRICAS TÊM UMA INTERPRETAÇÃO COMUM E SÃO MEDIDAS DE FORMA CONSISTENTE	É imperativo em uma organização que todas as pessoas reconheçam o que as métricas significam e tenham certeza de que as métricas são geradas de forma consistente. Para facilitar isso, os líderes devem garantir que os processos de medição sejam claramente documentados e que os funcionários compartilhem um conjunto comum de terminologia para entender o significado das apresentações de dados.
BOAS MÉTRICAS SÃO CONTEXTUALMENTE ESPECÍFICAS E AÇIONÁVEIS	Boas métricas têm significado para os membros de uma organização e sinalizam ações específicas que podem ser tomadas. Elas lançam luz sobre uma parte de baixo desempenho de uma organização, registram os resultados dos esforços de melhoria contínua ou demonstram o valor de pessoas e processos específicos dentro da organização.



EXEMPLO DE SETOR	EXEMPLO DE MÉTRICA
TRANSPORTE	Custo de transporte por quilômetro, tal como o total gasto dividido pela quantidade de km.
DEPÓSITOS	Custo por metro quadrado, tal como o custo operacional total dividido pelo tamanho em metro quadrado.
MARKETING	Custo de aquisição de clientes, tal como o custo de vendas/marketing dividido por novos clientes em um determinado período.
E-COMMERCE	Taxa de conversão de um site, tal como a porcentagem de visitantes únicos do site que fazem uma compra.
INTERNET	Receita média por usuário (receita total dividida pelo número total de usuários).

Legal! Agora como nós fazemos para utilizar as métricas para impulsionar a melhoria organizacional como um todo? Vejamos...

- **Escolha algumas poucas métricas-chave em qualquer nível:** muitas organizações acabam com dezenas de métricas-chave. Na verdade, essas organizações normalmente não sabem em que focar. Líderes devem investir tempo e energia para socializar sua organização para um pequeno conjunto de métricas essenciais. O desafio é oferecer uma definição clara do que está sendo medido e demonstrar sua importância. de estabelecer um conjunto de três a cinco métricas, incorpore as métricas em reuniões e outros locais, dado que é saudável que essas métricas se tornem parte do vernáculo organizacional. Aos poucos, as pessoas vão se acostumando com as métricas e suas implicações sobre o negócio.
- **Selecione as principais métricas que se alinham com a missão e a visão:** uma transformação pode ocorrer nas organizações que aprendem no que focar. Com a consciência e a compreensão de um pequeno conjunto de métricas-chave, as pessoas começam a se perguntar quais metas estão associadas a essas métricas – é uma progressão natural. Tanto para as partes interessadas internas quanto externas, há um forte alinhamento entre missão, visão, metas e acompanhamento do progresso da organização. Os esforços de todos podem ser direcionados para essas metas mensuráveis, e as pessoas se concentrarão nos processos que podem impactar essas métricas.
- **Mostre aos funcionários que suas contribuições são essenciais:** líderes verdadeiramente excepcionais não apenas articulam metas organizacionais e identificam métricas para avaliar o progresso operacional, mas também podem ajudar os funcionários a entender como suas contribuições de trabalho são um componente essencial das metas e medidas de progresso mais amplas. As métricas devem ter uma conexão hierárquica que permita que a equipe de linha de frente veja como suas medidas se relacionam com as principais métricas de toda a organização. Na natureza, a forma segue a função e temos inúmeros exemplos de alinhamento entre atributos físicos e condições ambientais. Da mesma forma, quanto mais os indivíduos tiverem uma linha de visão clara entre suas contribuições diárias e as prioridades da organização, mais



capazes (e dispostos) estarão de tomar decisões inteligentes e comprometer sua energia de maneira a sustentar o desempenho organizacional geral.

- **Referenciar métricas-chave e análise de dados ao comunicar metas:** líderes fluentes em dados lideram pelo exemplo. Da mesma forma, os líderes devem modelar a confiança na análise de dados e fazer referência às principais métricas ao comunicar metas e prioridades às partes interessadas. Para construir uma cultura fluente em dados, os líderes devem se comunicar na linguagem de suas métricas, usando dados para apoiar suas decisões sobre o progresso organizacional. Ao fazer isso, eles reforçam a importância das principais métricas selecionadas e demonstram sua alfabetização de dados em fóruns públicos, modelando assim os hábitos mentais desejados.
- **Evitando as armadilhas das métricas:** embora selecionar a métrica certa possa promover o sucesso, a métrica errada pode criar comportamentos inesperados, equivocados e contraproducentes. As métricas erradas podem criar estruturas de incentivo para comportamentos que realmente funcionam contra os objetivos da empresa. A maioria das empresas comete o erro de avaliar o desempenho (estabelecendo métricas de sucesso) ao longo das atividades funcionais. Isso perpetua silos disfuncionais e desconectados dentro de uma organização.

Ainda dentro desse contexto, devem-se considerar mais alguns pontos ao decidir sobre as métricas adequadas:

- **Tendências, distribuições e comparação:** um número único e solitário muitas vezes não revela uma verdadeira compreensão do desempenho de uma organização. É necessário colocar um conceito em torno do valor para permitir que as pessoas tenham uma compreensão mais completa. Ao mostrar tendências, distribuição e comparação, você melhora a compreensão da métrica para o público. A tendência mostra se a métrica se aproxima ou se afasta de uma meta. As comparações com os *benchmarks* do setor mostram como sua organização se destaca em um contexto mais amplo. Por fim, divida o valor por diferentes dimensões (por exemplo, locais, produtos e segmentos de clientes) para revelar se um fator está sustentando ou diminuindo o valor da métrica.
- **Casos extremos:** sempre haverá casos extremos, ou *outliers*, em que uma métrica pode não significar o que você pensa que significa. Vale a pena entender essas situações, mas você não deve permitir que o perfeito seja inimigo do bom. Como líder, você precisa pesar os benefícios de escolher métricas que funcionem para 90%, 95% ou 99% dos casos com os custos desses ganhos incrementais.
- **Teste de responsabilidade:** você poderia responsabilizar alguém por essa métrica sem que ela apresentasse uma dúzia de razões pelas quais isso não faz sentido? Caso contrário, talvez seja necessário reconsiderar a validade e o valor da métrica. Este exercício de pensamento simples é um bom teste do valor da métrica.



- **Autoatendimento:** tenha cuidado para não selecionar métricas simplesmente porque sabe que elas farão você parecer bom. Essas vitórias de curto prazo têm uma maneira de se transformar gradualmente em uma estratégia de longo prazo perdedora para a competitividade e o sucesso organizacional.
- **Deixar ir:** descartar uma métrica, especialmente dentro do contexto de uma organização grande, multifacetada e complexa, é uma coisa difícil de fazer. Existem alguns motivos para isso acontecer: (a) a métrica foi desenvolvida com muito esforço e alto custo; (b) Após o processo de coleta dos dados para as paradas métricas, ele só pode ser reiniciado com muito esforço e alto custo; e (c) pessoas mais altas na organização (que prestaram atenção à métrica pela última vez quando ela foi útil) podem procurá-la se algo der errado. No entanto, devemos deixar as métricas finalmente serem descartadas.



Atributos de Dados

INCIDÊNCIA EM PROVA: BAIXA

Galera, eu escarafunchei exaustivamente a (pouca) literatura sobre fluência de dados e não encontrei absolutamente nada sobre atributos, mas achei uma página que trata do tema:

[HTTPS://WWW.DIGITALFLUENCY.GUIDE/DATA/ATTRIBUTES-OF-DATA](https://www.digitalfluency.guide/data/attributes-of-data)

ATRIBUTOS	DESCRIÇÃO
RELEVÂNCIA	Trata da utilidade de um conjunto de dados para os seus fins, isto é, quão relevantes são os dados para os usos pretendidos.
COBERTURA	Trata da área geográfica que o conjunto de dados descreve, isto é, se os dados se aplicam a apenas uma ou várias regiões geográficas.
AMPLITUDE	Trata de quão abrangente é a amostra do conjunto de dados, isto é, quão amplo é o conjunto de dados, se se aplica a apenas um grupo demográfico, entre outros.
PROFUNDIDADE	Trata da variedade e o número de pontos de dados diferentes no conjunto de dados, isto é, se o conjunto fornece muitos pontos de dados úteis ou apenas alguns.
CUSTO	Trata do custo de acesso e processamento dos dados, isto é, qual é o custo por unidade do conjunto de dados.
HISTÓRIA	Também chamado de comprimento ou duração, trata do alcance histórico do conjunto de dados (se houver), isto é, até quando para trás vai o conjunto de dados.
FREQUÊNCIA	Trata do intervalo de tempo entre os pontos de dados, isto é, da frequência com que os dados são atualizados.
PROCESSAMENTO	Trata da quantidade de correção de erros e rotulagem executada no conjunto de dados antes da distribuição, isto é, quão processados são os dados, se houve correções, etc.
TEMPO DE PUBLICAÇÃO	Trata do período de tempo entre a aquisição dos dados originais e a distribuição do conjunto de dados.
ESTRUTURAÇÃO	Trata do grau em que os atributos dos dados foram definidos e categorizados, isto é, quão estruturado é o conjunto de dados.
CUSTO DE PESQUISA	Trata do custo de aquisição e verificação dos dados em sua forma original, isto é, o custo de aquisição, custo de atualização, entre outros.
QUALIDADE	Trata de quão precisos são os dados originais. Se houver erros, analisa-se se eles são críticos ou apenas uma necessidade de processamento.
TENDÊNCIA	Trata dos vieses replicados nos dados por fatores humanos e/ou de máquina, isto é, se existem vieses intencionais, não intencionais ou inerentes aos dados.
DISPONIBILIDADE	Trata da facilidade de acessar o conjunto de dados, isto é, quão amplamente disponíveis estão os dados, se eles podem ser obtidos facilmente, entre outros.
ORIGINALIDADE	Trata de quão exclusivo o conjunto de dados é comparado a outros conjuntos de dados, isto é, quão original, nova, inovadora é a fonte.
TECNOLOGIA	Trata de quão avançadas são as ferramentas digitais necessárias para acessar e analisar os dados, isto é, se são necessárias tecnologias especiais para acessar ou processar os dados.
TRIAL	Trata do comprimento, a amplitude e a profundidade dos dados de pré-visualização disponíveis para avaliação.
LEGALIDADE/ ÉTICA	Trata da facilidade de acesso e uso dos dados em relação a restrições de lei, regulamentação e padrões éticos.



PORTFOLIO/ ORTOGONALIDADE	Trata da porcentagem de casos de utilização que os dados se aplicam, isto é, se dados são úteis para diversas áreas ou apenas algumas.
ADEQUAÇÃO AO TIPO DE INVESTIMENTO	Trata de quão bem adaptados seus dados estão ao tipo de investimento que você pode fazer, isto é, se os dados são adequados para os tipos de investimento que você está considerando.
FREQUÊNCIA DE TEMPO DA ESTRATÉGIA DE INVESTIMENTO	Trata do grau em que a frequência de dados se alinha com a frequência de investimento pretendida. Se você estiver usando dados para orientar as decisões de investimento em um cronograma específico, os dados são atualizados regularmente o suficiente para serem de valor para você?
CUSTO DE ANÁLISE	Trata do custo de transformar os dados em informações úteis (diferente dos custos de processamento que precedem a análise).

Uma outra fonte traz um segundo conjunto (bem mais enxuto) de atributos de fluência de dados. Vejam na tabela a seguir:

ATRIBUTOS	DESCRIÇÃO
PRECISÃO	A precisão da análise de dados e insights.
PONTUALIDADE	A capacidade de analisar dados rapidamente e obter insights.
CONFIABILIDADE	A capacidade de produzir consistentemente resultados precisos.
RELEVÂNCIA	A capacidade de identificar dados e tendências relevantes.
ESCALABILIDADE	A capacidade de dimensionar processos de análise de dados.
CUSTO/BENEFÍCIO	A capacidade de analisar dados de maneira econômica.
QUALIDADE	A capacidade de produzir insights de dados de alta qualidade.
VALIDADE	A capacidade de validar dados e insights.
SEGURANÇA	A capacidade de proteger dados e proteger a privacidade
ROBUSTEZ	A capacidade de se adaptar às mudanças nas fontes de dados.



Transformação de Dados

INCIDÊNCIA EM PROVA: BAIXA

A **Transformação de Dados** refere-se ao processo de alteração, manipulação ou conversão de dados de uma forma para outra com o objetivo de atender a alguma finalidade específica. Esse processo pode envolver várias etapas e técnicas para preparar os dados para análise, relatórios, armazenamento ou outros fins. Aqui estão algumas das atividades comuns relacionadas à transformação de dados:

TRANSFORMAÇÃO	DESCRIÇÃO
LIMPEZA DE DADOS	Isso inclui a remoção de dados duplicados, a correção de erros e a padronização de formatos, como datas e números. O objetivo é garantir a qualidade e a consistência dos dados.
FILTRAGEM DE DADOS	Isso envolve a seleção de um subconjunto de dados com base em critérios específicos. Por exemplo, filtrar todos os clientes que fizeram uma compra nos últimos seis meses.
AGREGAÇÃO DE DADOS	Agregar dados significa resumir informações detalhadas em níveis mais altos. Isso pode envolver a criação de médias, totais, contagens ou outras estatísticas resumidas.
CONVERSÃO DE FORMATO	Converter dados de um formato para outro, como de texto para números ou vice-versa. Isso é comum ao lidar com dados de diferentes fontes.
NORMALIZAÇÃO DE DADOS	Normalizar dados significa ajustar valores para que eles estejam em uma escala específica ou sigam uma distribuição padrão. Isso é útil em aprendizado de máquina e análises estatísticas.
DERIVAÇÃO DE NOVOS ATRIBUTOS	Às vezes, você precisa criar novos atributos com base nos dados existentes. Por exemplo, calcular a idade com base na data de nascimento.
CONCATENAÇÃO DE DADOS	Combinar dados de várias fontes ou tabelas para criar um conjunto de dados único.
PIVOTEAMENTO DE DADOS	Transformar dados de um formato "longo" para "largo" ou vice-versa. Isso é comum em bancos de dados relacionais.
TRATAMENTO DE VALORES AUSENTES	Decidir o que fazer com valores ausentes, seja preenchendo-os com valores padrão, removendo-os ou imputando valores com base em métodos estatísticos.
CODIFICAÇÃO DE CATEGORIAS	Converter dados categóricos em representações numéricas para uso em algoritmos de aprendizado de máquina.
REDUÇÃO DE DIMENSIONALIDADE	Em análise de dados de alta dimensionalidade, trata-se de reduzir a quantidade de variáveis sem perder muita informação.
PADRONIZAÇÃO DE DATA/HORA	Garantir que todas as datas e horários estejam em um formato consistente e compatível.



QUESTÕES COMENTADAS – DIVERSAS BANCAS

1. (CESPE / DATAPREV – 2023) Arquivos .xml, .csv e .json são exemplos de dados estruturados.

Comentários:

Arquivos .xml (eXtensible Markup Language), .csv (Comma-Separated Values) e .json (JavaScript Object Notation) são exemplos de formatos semiestruturados. Cada um desses formatos tem características específicas que os tornam adequados para armazenar e organizar dados de maneira estruturada:

- XML é um formato flexível que usa tags para definir elementos. É comumente usado para compartilhar dados na internet e em outros ambientes de rede devido à sua capacidade de representar dados complexos e hierárquicos de forma clara.
- CSV é um formato simples que usa vírgulas (ou outro delimitador) para separar valores. Cada linha do arquivo representa uma linha de dados, e cada valor separado por vírgula representa uma coluna. Este formato é amplamente utilizado devido à sua simplicidade e compatibilidade com muitos sistemas.
- JSON é um formato leve de troca de dados, fácil de ler e escrever para humanos, e fácil de analisar e gerar por máquinas. É baseado em um subconjunto da linguagem de programação JavaScript e é muito usado em APIs da web e configurações de aplicativos devido à sua facilidade de uso em ambientes de desenvolvimento web.

Gabarito: Errado

2. (CONSULPLAN / MPE-PA – 2022) Dados são um conjunto de fatos em estado bruto que podem ser utilizados para tomadas de decisão. Os dados podem ser divididos entre estruturados e não-estruturados. Sobre os dados estruturados, assinale a afirmativa correta.

- a) Mensagens de e-mail são dados estruturados.
- b) Dados estruturados são organizados em linhas e colunas.
- c) Dados estruturados são armazenados em arquivos no formato “.TXT”.
- d) Mecanismos de busca possuem dificuldade para pesquisar dados estruturados.

Comentários:

(a) Errado. Mensagens de e-mail são geralmente consideradas dados não estruturados, pois consistem principalmente de texto livre e podem conter uma variedade de formatos e anexos;



- (b) Correto. Dados estruturados são normalmente organizados em linhas e colunas, como em um banco de dados relacional. Esta estrutura permite fácil acesso, manipulação e análise dos dados;
- (c) Errado. Embora seja possível armazenar dados estruturados em arquivos no formato “.txt”, este não é o método predominante ou mais representativo;
- (d) Errado. Mecanismos de busca, especialmente os projetados para bancos de dados, geralmente têm facilidade para pesquisar dados estruturados devido à sua organização clara e previsível.

Gabarito: Letra B

3. (CESPE / Polícia Federal – 2021) Os dados estruturados diferenciam-se dos dados não estruturados pela rigidez em seu formato e pelo fato de poderem ser armazenados em campos de tabelas de um banco de dados relacional.

Comentários:

Perfeito! Dados estruturados são aqueles que são organizados e representados por meio de uma estrutura rígida e formal; já os dados não-estruturados são aqueles que não possuem uma organização rígida e formal, isto é, não seguem uma estrutura padronizada, sendo totalmente flexível e dinâmica.

Gabarito: Correto

4. (CESPE / SERPRO – 2021) Os dados não estruturados constam de bancos de dados relacionais que são eminentemente pesquisáveis tanto por meio de consultas realizadas por humanos quanto por meio daquelas realizadas por algoritmos que usam tipos de dados e nomes de campos, como alfabéticos ou numéricos, moeda ou data.

Comentários:

Na verdade, eles constam de bancos de dados não relacionais, não sendo eminentemente pesquisáveis por consultas realizadas por humanos ou por meio daquelas realizadas por algoritmos que usam tipos de dados e nomes de campos, como alfabéticos ou numéricos, moeda ou data.

Gabarito: Errado

5. (CESPE/ ME – 2020) Embora com características particulares, dados não estruturados podem ser classificados em sua totalidade, assim como os dados estruturados.

Comentários:



Galera, a questão não deixa muito claro o que quis dizer com essa classificação. De todo modo, podemos inferir que essa classificação se trata de padronização. Pois bem, dados não-estruturados não possuem uma estrutura definida e padronizada, logo não podem ser classificados em sua totalidade (como ocorre com dados estruturados).

Gabarito: Errado

6. (FAURGS/TJ-RS – 2018) Os dados _____ são armazenados em _____. Os dados _____ misturam nomes de tipos de dados e valores de dados, mas nem todos eles precisam seguir uma estrutura predefinida fixa. Os dados _____ referem-se à informação exibida na Web, caso em que a informação sobre os tipos dos itens de dados não existe.

Assinale a alternativa que completa, correta e respectivamente, as lacunas do texto acima.

- a) semiestruturados / bancos de dados tradicionais / estruturados / não estruturados
- b) não estruturados / XML / semiestruturados / estruturados
- c) estruturados / HTML / não estruturados / semiestruturados
- d) estruturados / bancos de dados tradicionais / semiestruturados / não estruturados
- e) semiestruturados / Web / estruturados / não estruturados

Comentários:

Os dados estruturados são armazenados em bancos de dados tradicionais. Os dados semiestruturados misturam nomes de tipos de dados e valores de dados, mas nem todos eles precisam seguir uma estrutura predefinida fixa. Os dados não-estruturados referem-se à informação exibida na Web, caso em que a informação sobre os tipos dos itens de dados não existe.

Gabarito: Letra D

7. (CESPE / Polícia Federal – 2018)

CPF
NOME
DATA DE NASCIMENTO
NOME DO PAI
NOME DA MAE
TELEFONE
CEP
NUMERO

As informações anteriormente apresentadas correspondem aos campos de uma tabela de um banco de dados, a qual é acessada por mais de um sistema de informação e também por outras tabelas. Esses dados são utilizados para simples cadastros, desde a consulta até sua alteração,



e também para prevenção à fraude, por meio de verificação dos dados da tabela e de outros dados em diferentes bases de dados ou outros meios de informação.

Considerando essas informações, julgue o item que se segue.

Os dados armazenados na referida tabela são considerados não estruturados.

Comentários:

O enunciado afirma que as informações correspondem aos campos de uma tabela de um banco de dados, logo são dados estruturados. *Por quê?* Porque possuem uma estrutura rígida, formal, homogênea e previamente formatada.

Gabarito: Errado

8. (CESPE / TCE-SC – 2016) Em se tratando de dados estruturados, a informação de esquema está mesclada aos valores dos dados, e cada objeto de dados pode ter atributos diferentes, que não são conhecidos com antecedência. Essa característica os diferencia de dados não estruturados.

Comentários:

Dados mesclados com um esquema de representação parcialmente organizados são características de dados semiestruturados e, não, não estruturados.

Gabarito: Errado

9. (CESPE / TCE-PA – 2016) Em comparação aos dados não estruturados, os dados estruturados demandam mais espaço de armazenamento e um gerenciamento mais cauteloso, uma vez que constituem a maior parte dos dados corporativos.

Comentários:

A maior parte dos dados corporativos são não-estruturados, logo eles necessitam de maior espaço de armazenamento e um gerenciamento mais cauteloso.

Gabarito: Errado

10. (FEPESE / SJC-SC – 2013) Qual software, dentre os listados abaixo, permite a importação (e atualização) de dados estruturados de bancos de dados relacionais para posterior manipulação pelo aplicativo?

- a) Word
- b) Excel



- c) Powerpoint
- d) Internet Explorer
- e) Outlook

Comentários:

Dados estruturados são aqueles que podem ser armazenados em... tabelas. Logo, é o MS-Excel.

Gabarito: Letra B

11. (CESPE/ TRT-ES – 2013) Os dados estruturados que descrevem e permitem encontrar, gerenciar, compreender e (ou) preservar documentos arquivísticos ao longo do tempo são conhecidos como metadados.

Comentários:

Perfeito, os dados estruturados que descrevem e permitem encontrar, gerenciar, compreender e (ou) preservar documentos arquivísticos ao longo do tempo são conhecidos como metadados.

Gabarito: Correto



LISTA DE QUESTÕES– DIVERSAS BANCAS

1. (CESPE / DATAPREV – 2023) Arquivos .xml, .csv e .json são exemplos de dados estruturados.
2. (CONSULPLAN / MPE-PA – 2022) Dados são um conjunto de fatos em estado bruto que podem ser utilizados para tomadas de decisão. Os dados podem ser divididos entre estruturados e não-estruturados. Sobre os dados estruturados, assinale a afirmativa correta.
 - a) Mensagens de e-mail são dados estruturados.
 - b) Dados estruturados são organizados em linhas e colunas.
 - c) Dados estruturados são armazenados em arquivos no formato “.TXT”.
 - d) Mecanismos de busca possuem dificuldade para pesquisar dados estruturados.
3. (CESPE / Polícia Federal – 2021) Os dados estruturados diferenciam-se dos dados não estruturados pela rigidez em seu formato e pelo fato de poderem ser armazenados em campos de tabelas de um banco de dados relacional.
4. (CESPE / SERPRO – 2021) Os dados não estruturados constam de bancos de dados relacionais que são eminentemente pesquisáveis tanto por meio de consultas realizadas por humanos quanto por meio daquelas realizadas por algoritmos que usam tipos de dados e nomes de campos, como alfabéticos ou numéricos, moeda ou data.
5. (CESPE/ ME – 2020) Embora com características particulares, dados não estruturados podem ser classificados em sua totalidade, assim como os dados estruturados.
6. (FAURGS/TJ-RS – 2018) Os dados _____ são armazenados em _____. Os dados _____ misturam nomes de tipos de dados e valores de dados, mas nem todos eles precisam seguir uma estrutura predefinida fixa. Os dados _____ referem-se à informação exibida na Web, caso em que a informação sobre os tipos dos itens de dados não existe.

Assinale a alternativa que completa, correta e respectivamente, as lacunas do texto acima.

- a) semiestruturados / bancos de dados tradicionais / estruturados / não estruturados
- b) não estruturados / XML / semiestruturados / estruturados
- c) estruturados / HTML / não estruturados / semiestruturados
- d) estruturados / bancos de dados tradicionais / semiestruturados / não estruturados
- e) semiestruturados / Web / estruturados / não estruturados

7. (CESPE / Polícia Federal – 2018)

CPF

NOME



DATA DE NASCIMENTO
NOME DO PAI
NOME DA MAE
TELEFONE
CEP
NUMERO

As informações anteriormente apresentadas correspondem aos campos de uma tabela de um banco de dados, a qual é acessada por mais de um sistema de informação e também por outras tabelas. Esses dados são utilizados para simples cadastros, desde a consulta até sua alteração, e também para prevenção à fraude, por meio de verificação dos dados da tabela e de outros dados em diferentes bases de dados ou outros meios de informação.

Considerando essas informações, julgue o item que se segue.

Os dados armazenados na referida tabela são considerados não estruturados.

8. **(CESPE / TCE-SC – 2016)** Em se tratando de dados estruturados, a informação de esquema está mesclada aos valores dos dados, e cada objeto de dados pode ter atributos diferentes, que não são conhecidos com antecedência. Essa característica os diferencia de dados não estruturados.
9. **(CESPE / TCE-PA – 2016)** Em comparação aos dados não estruturados, os dados estruturados demandam mais espaço de armazenamento e um gerenciamento mais cauteloso, uma vez que constituem a maior parte dos dados corporativos.
10. **(FEPESE / SJC-SC – 2013)** Qual software, dentre os listados abaixo, permite a importação (e atualização) de dados estruturados de bancos de dados relacionais para posterior manipulação pelo aplicativo?
- a) Word
 - b) Excel
 - c) Powerpoint
 - d) Internet Explorer
 - e) Outlook
11. **(CESPE/ TRT-ES – 2013)** Os dados estruturados que descrevem e permitem encontrar, gerenciar, compreender e (ou) preservar documentos arquivísticos ao longo do tempo são conhecidos como metadados.



GABARITO

1. ERRADO
2. LETRA B
3. CORRETO
4. ERRADO
5. ERRADO
6. LETRA D
7. ERRADO
8. ERRADO
9. ERRADO
10. LETRA B
11. CORRETO



ESSA LEI TODO MUNDO CONHECE: PIRATARIA É CRIME.

Mas é sempre bom revisar o porquê e como você pode ser prejudicado com essa prática.



1 Professor investe seu tempo para elaborar os cursos e o site os coloca à venda.



2 Pirata divulga ilicitamente (grupos de rateio), utilizando-se do anonimato, nomes falsos ou laranjas (geralmente o pirata se anuncia como formador de "grupos solidários" de rateio que não visam lucro).



3 Pirata cria alunos fake praticando falsidade ideológica, comprando cursos do site em nome de pessoas aleatórias (usando nome, CPF, endereço e telefone de terceiros sem autorização).



4 Pirata compra, muitas vezes, clonando cartões de crédito (por vezes o sistema anti-fraude não consegue identificar o golpe a tempo).



5 Pirata fere os Termos de Uso, adultera as aulas e retira a identificação dos arquivos PDF (justamente porque a atividade é ilegal e ele não quer que seus fakes sejam identificados).



6 Pirata revende as aulas protegidas por direitos autorais, praticando concorrência desleal e em flagrante desrespeito à Lei de Direitos Autorais (Lei 9.610/98).



7 Concurseiro(a) desinformado participa de rateio, achando que nada disso está acontecendo e esperando se tornar servidor público para exigir o cumprimento das leis.



8 O professor que elaborou o curso não ganha nada, o site não recebe nada, e a pessoa que praticou todos os ilícitos anteriores (pirata) fica com o lucro.



Deixando de lado esse mar de sujeira, aproveitamos para agradecer a todos que adquirem os cursos honestamente e permitem que o site continue existindo.