

TECNOLOGIA DA INFORMAÇÃO

Fundamentos sobre Análise de Dados



SUMÁRIO

| | |
|--|----|
| Apresentação | 3 |
| Fundamentos sobre Análise de Dados | 4 |
| 1. Análise de Dados no Setor Público | 4 |
| 2. Conceitos Relacionados à Análise de Dados | 5 |
| 2.1. Dado | 5 |
| 2.2. Informação | 5 |
| 2.3. Dados Abertos | 6 |
| 2.4. Dados Estruturados, Semiestruturados e Não Estruturados | 6 |
| 2.5. <i>Big Data</i> | 9 |
| 2.6. Hadoop | 9 |
| 2.7. Aprendizado de Máquina (<i>Machine Learning</i>) | 10 |
| 2.8. Mineração de Dados (<i>Data Mining</i>) | 11 |
| 2.9. Linguagem R | 12 |
| 3. Tipos de Análise de Dados | 13 |
| 4. As Variáveis na Análise Exploratória de Dados | 15 |
| 5. O Processo de <i>Data Science</i> | 20 |
| 6. Arquitetura de BI (<i>Business Intelligence</i> ou Inteligência de Negócios) | 21 |
| Resumo | 26 |
| Questões de Concurso | 28 |
| Gabarito | 35 |
| Referências | 36 |

APRESENTAÇÃO

Olá, querido(a) amigo(a)!

“Seja você quem for, seja qual for a posição social que você tenha na vida, a mais alta ou a mais baixa, tenha sempre como meta muita força, muita determinação e sempre faça tudo com muito amor e com muita fé em Deus, que um dia você chega lá. De alguma maneira você chega lá.” (Ayrton Senna).

O momento perfeito não “surge”. Ele é construído. Construa o seu. Você tem suas próprias dificuldades, problemas, vitórias e soluções. Continue firme e, em breve, estará colhendo os frutos da vitória.

Rumo então à aula sobre os fundamentos da **análise de dados!**

FORÇA e muita DETERMINAÇÃO nos estudos!

Em caso de dúvidas, acesse o fórum do curso ou entre em contato.

Um forte abraço!

Profª Patrícia Quintão

Instagram: @coachpatriciaquintao

WhatsApp: (31) 99442.0615

FUNDAMENTOS SOBRE ANÁLISE DE DADOS

1. ANÁLISE DE DADOS NO SETOR PÚBLICO

Atualmente, uma quantidade significativa de dados é processada diariamente pelas empresas, pelos **órgãos públicos**, etc.

Conforme destaca ENAP (2020), ***“quando bem trabalhados, monitorados e analisados, os dados servem para auxiliar as instituições em muitos aspectos. Cada vez mais os órgãos públicos vêm tomando decisões com base nos dados, seja para **detecção de anomalias**, **monitoramento de indicadores** ou **melhoria de processos**. Muitos órgãos já perceberam a importância de se realizar análise sobre os dados e os ganhos que esta atividade fornece”.***

A seguir, destacamos várias **iniciativas** com foco na **análise de dados públicos** (ENAP,2020):

1. O Ministério da Economia disponibiliza diversos **painéis públicos** em seu portal, com o objetivo de dar mais transparência em suas ações e permitir um melhor acompanhamento de seus trabalhos.
2. O Portal da Transparência, mantido pela CGU, disponibiliza painéis sobre vários temas, permitindo que os dados sejam baixados para análise individual.



Figura. Portal da Transparência – CGU (ENAP, 2020)

3. Análise de dados para localização de vítimas do rompimento da barragem de Brumadinho, pelo Corpo de Bombeiros Militar de Minas Gerais.

4. Análise de vínculos para detecção de fraudes pelo Tribunal de Contas do Estado de São Paulo e Conselho Administrativo de Defesa Econômica.

5. Fiscalização contínua de folhas de pagamento da Administração Pública, pelo TCU.

6. Detecção de anomalias para identificar a prática de conluio em licitações do governo federal, pela CGU.

7. Na área de saúde, centros de controle e prevenção de doenças utilizam os dados para prever surtos de gripe e rastrear padrões da doença.

8. Na área de detecção e prevenção de crimes, a análise de dados permite identificar padrões de atividades criminosas e lugares com alta incidência de criminalidade.

2. CONCEITOS RELACIONADOS À ANÁLISE DE DADOS

2.1. DADO

Um **dado** é um registro de alguma entidade. Um nome é um dado, uma foto é um dado, 134 é um dado, 5 é um dado, etc. **Não há semântica envolvida no dado.** Não há uma interpretação sobre esta sequência de símbolos. É algo “bruto”, como o número 10 ou a letra F. Eles, por si só, não fazem sentido nenhum, ficando impossibilitados de se tirar alguma conclusão do que são, quais os seus propósitos ou a que pertencem.

Obs.: Moresi(2001) destaca que **dados** são fatos ou observações “crus”. Mais especificamente, **os dados são medidas objetivas e quantitativas dos atributos (características) de entidades como pessoas, lugares, coisas e eventos (conjunto de fatos).**

Os **dados** são uma parte pequena da informação, que sozinhos não fazem sentido.

2.2. INFORMAÇÃO

A **informação** é um dado depois de processado, ou uma contextualização de um dado... Como assim? “10” é um dado, mas e se eu disser o seguinte: “No dia 10 não haverá aula!”. Nesse caso, o 10 passou a ter sentido (ou passou a ter “**contexto**”) e agora é uma informação!

Informações são conjuntos de dados significativos e úteis a seres humanos em processos como o de tomada de decisões. “São **dados interpretados, dotados de relevância e propósito**” (DRUCKER, 1999).

Obs.: Conforme destacado por Moresi(2001), **informações**, são dados que foram organizados e ordenados de forma coerente e significativa para fins de compreensão e análise (sendo a base para ações coordenadas).

A transformação de dados em informação é frequentemente realizada através da apresentação dos dados em uma forma compreensível ao usuário. As informações são produzidas pelo processamento de dados. Elas são utilizadas para revelar o significado dos dados.

2.3. DADOS ABERTOS

Obs.: Com o objetivo de dar transparência ao cidadão, diversos países disponibilizam na internet bases de dados governamentais classificadas como **dados abertos**.

Dados são abertos quando qualquer pessoa pode livremente **usá-los**, **reutilizá-los** e **redistribuí-los**, estando sujeito a, no máximo, a exigência de creditar a sua autoria e compartilhar pela mesma licença, é o que destaca a Fundação do Conhecimento Aberto (*Open Knowledge Foundation* – OKF). Isso geralmente é satisfeito pela publicação dos dados em formato aberto e sob uma licença aberta.

Quando os dados são produzidos, coletados ou custodiados por autoridades públicas e disponibilizados em formato aberto, considera-se que são **dados abertos governamentais** (<http://portal3.tcu.gov.br/portal/pls/portal/docs/2689107.PDF>).

No Brasil, a **Lei 12.527/2011 (Lei de Acesso à Informação – LAI)** dispõe que **as informações de interesse coletivo ou geral devem ser obrigatoriamente divulgadas pelos órgãos e entidades públicos em seus sítios oficiais**, os quais devem atender, por exemplo, aos seguintes requisitos:

- possibilitar o acesso automatizado por sistemas externos em formatos abertos;
- serem estruturados e legíveis por máquina;
- estarem acompanhados de detalhes sobre os formatos utilizados para estruturação da informação;
- serem autênticos, íntegros e atualizados.

Obs.: Nem sempre as análises são feitas com **dados abertos**, normalmente devido a questões de sigilo das informações.

2.4. DADOS ESTRUTURADOS, SEMIESTRUTURADOS E NÃO ESTRUTURADOS

Podemos classificar os dados, de acordo com sua estrutura, de três formas: **dados estruturados**, **dados semiestruturados** e **dados não estruturados**.

- **Dados não estruturados:** são aqueles que **existem em seu estado original (bruto)**, ou seja, **no formato em que foram coletados**. Portanto, **estão em um formato que não possibilita o processamento que produz informações**.

Como exemplos de dados desse tipo podemos citar: textos, imagens, vídeos, documentos, entre outros. Ao contrário dos dados estruturados, os dados não estruturados **não possuem estrutura definida**.

- **Dados estruturados:** são o **resultado da obtenção de dados não estruturados e de sua formatação (estruturação)** visando facilitar o armazenamento, a utilização e a geração de informações. A estrutura (formato) é aplicada com base no tipo de processamento que se deseja executar nos dados.

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

Obs.: | **Alguns dados podem não estar prontos (não estruturados) para determinados tipos de processamento, mas podem estar prontos (estruturados) para outros tipos.**

Exemplo: o valor de dados 37890 pode se referir a um CEP, um valor de vendas ou um código de produto. Se representar um CEP ou um código de produto e for armazenado como texto, não será possível executar cálculos matemáticos com ele. Por outro lado, se esse valor representar uma transação de vendas, será necessário formatá-lo como numérico.

Para ilustrar o conceito de **estrutura**, imagine uma pilha de faturas impressas em papel. Caso deseje simplesmente armazená-las como imagens para recuperação e exibição futura, é possível escaneá-las e salvá-las em formato gráfico. Por outro lado, se desejar obter informações como vendas mensais totais e médias, esse armazenamento gráfico não seria útil. Em vez disso, é possível armazenar os dados das faturas em um formato de planilha (estruturado) de modo a permitir a execução dos cálculos necessários. Na verdade, em sua maioria, os dados que encontramos são mais bem classificados como **semiestruturados**.



Figura. Dados Estruturados, SemiEstruturados e Não Estruturados (UNIVERSIDADE DA TECNOLOGIA, 2019)

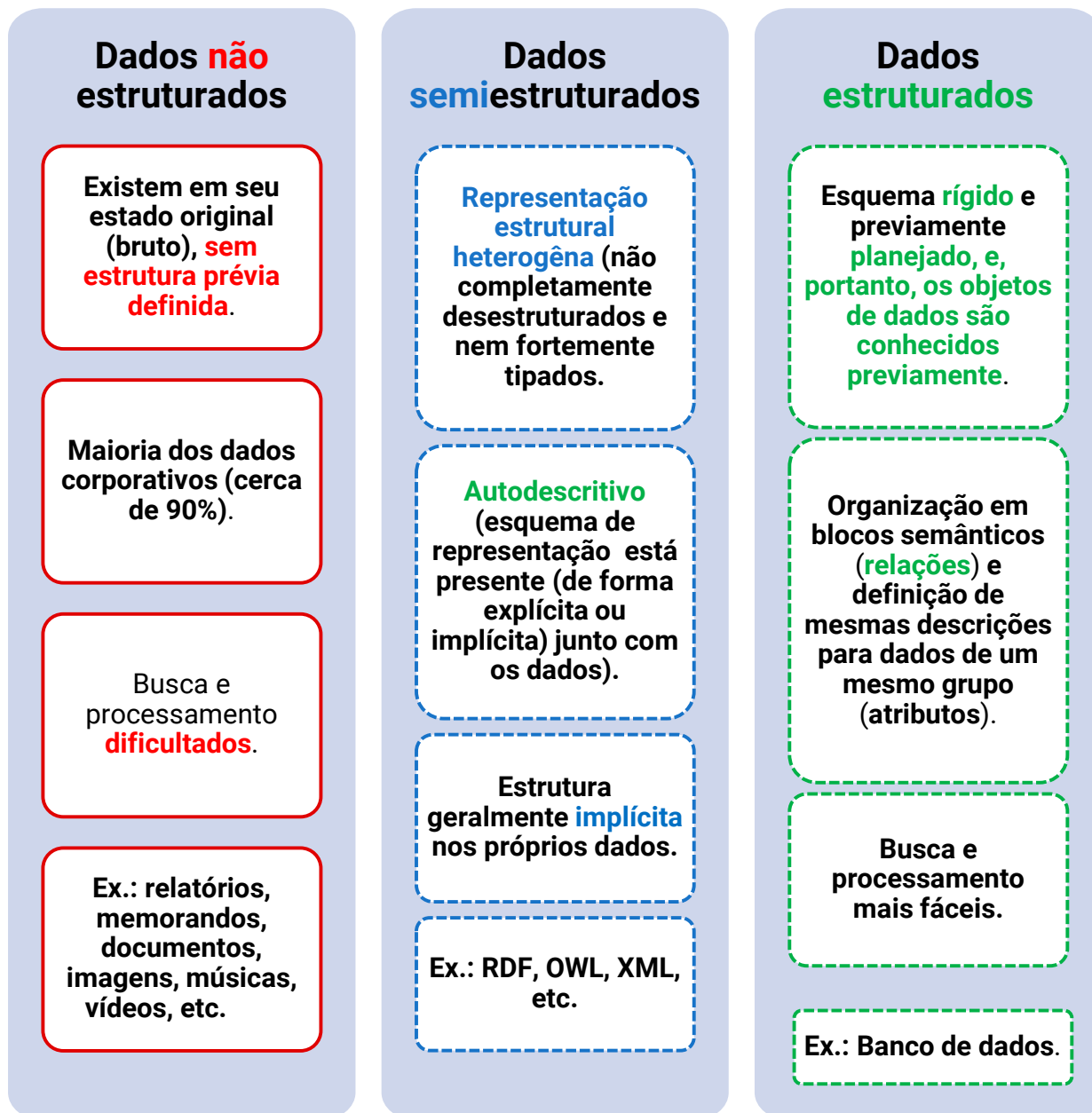
Conforme visto na figura, alguns autores também abordam uma **terceira classificação**, conhecida como **dados semiestruturados**.

- **Dados semiestruturados** são aqueles que **já foram parcialmente processados**. Como o nome pode indicar, são dados que contêm **parte de sua estrutura rígida e outra parte não rígida**. Possuem uma **representação estrutural heterogênea**, não sendo nem completamente não

estruturados e nem estritamente tipados. Por exemplo, olhando-se uma página comum da web, os dados são apresentados em um formato pré-organizado para transmitir alguma informação.

- As corporações **não** se limitam ao uso de dados **estruturados**, também utilizam dados **semiestruturados** e **não** estruturados.

Esquemmatizando!



2.5. BIG DATA

Siewert (2013) destaca que o termo **Big Data** é “definido genericamente como a **captura, gerenciamento** e a **análise de dados que vão além dos dados tipicamente estruturados**, que podem ser consultados e pesquisados através de bancos de dados relacionais.

Frequentemente são **dados obtidos de arquivos não estruturados** como **vídeo digital, imagens, dados de sensores, arquivos de logs e de qualquer tipo de dados não contidos em registros típicos com campos que podem ser pesquisados**”.

Big Data tem dados estruturados e não estruturados!
Big Data é o termo que descreve o imenso volume de dados – estruturados e não estruturados – que impactam os negócios no dia a dia.

De maneira geral, **Big Data** não se refere apenas aos dados, mas também às **soluções tecnológicas criadas para lidar com dados em volume, variedade e velocidade significativos** (CESPE/2018).

Segundo Siewert (2013), **o Big Data tem variadas fontes de dados** como:

- dados gerados pelas máquinas (redes de sensores, logs);
- dispositivos móveis (vídeo, mensagens, fotografias);
- comunicação máquina a máquina, a “Internet das coisas”;
- dados em bancos de dados relacionais oriundos das transações da organização;
- imagens de documentos, etc.

De acordo com Landim (2015), trata-se de um termo usado para descrever **grandes e complexos conjuntos de dados que são muito difíceis de capturar, processar, armazenar, buscar e analisar com os sistemas de base de dados convencionais**.

Obs.: **Big data** refere-se a uma nova geração de tecnologias e arquiteturas projetadas para processar volumes muito grandes e com grande variedade de dados, permitindo alta velocidade de captura, descoberta e análise (Cespe/2018).

2.6. HADOOP

Quando nos referimos a Big Data, apenas um banco de dados do tipo **não** basta. É necessário também contar com **ferramentas** (Ex.: **Hadoop** é a principal referência) que permitam o tratamento correto do volume de dados.

Hadoop é uma plataforma *open source* **desenvolvida especialmente para processamento e análise de grandes volumes de dados**, sejam eles estruturados ou não estruturados. **É utilizado em larga escala por grandes corporações, como Facebook e Twitter, em aplicações Big Data.**

Útil para aplicações que envolvam dados massivos para processamento paralelo (embora seja interessante para processamento de quaisquer dados), geralmente utilizando um cluster de computadores (Devmedia, 2016).

Obs.: O **Apache Hadoop** é um *framework open source* (estrutura de código aberto) para processamento e gerenciamento de grandes volumes de dados (Big Data). Também pode ser definido como um ecossistema de ferramentas e métodos para armazenamento, distribuição e **análise de dados estruturados e não estruturados** (ENAP,2020).

2.7. APRENDIZADO DE MÁQUINA (*MACHINE LEARNING*)

Machine Learning (ou **Aprendizado de Máquina**) é uma área de estudo que busca dar aos computadores a habilidade de aprender sem serem programados explicitamente.

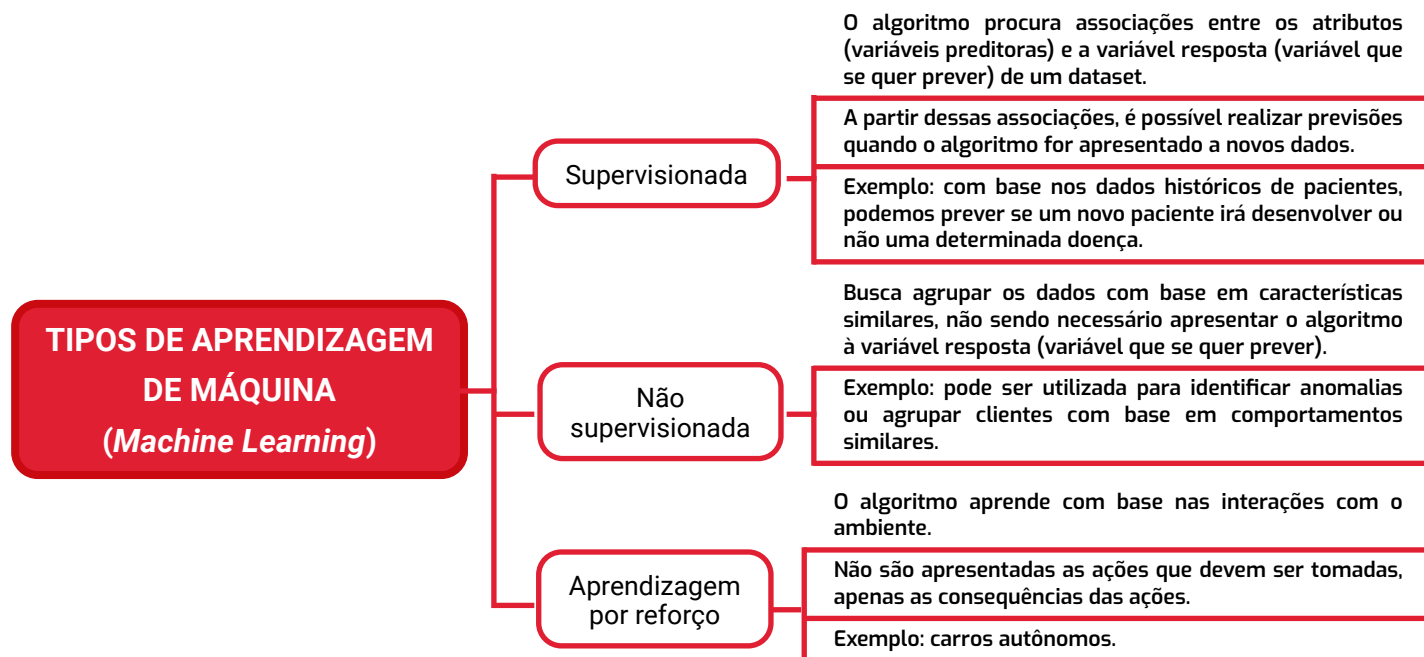
Segundo artigo da Data Science Academy (2018), “a aprendizagem de máquina é um subconjunto da inteligência artificial (IA), o segmento da ciência da computação que se concentra na criação de computadores que pensam da maneira que os humanos”. A máquina aprende com seus erros e acertos e é capaz de fazer previsões e tomar decisões baseadas em sua experiência, que pode ser compartilhada para outras máquinas.

Veja mais: <https://www.youtube.com/watch?v=Z1YHbl0lh88>

ENAP (2020) destaca alguns exemplos da **aplicação do aprendizado de máquina** na atualidade:

- 1. Detecção de anomalias e fraudes;
- 2. Identificação de mensagens *spam* em e-mails;
- 3. Segmentação de clientes;
- 4. Carros e drones autônomos;
- 5. Mecanismos de busca;
- 6. Segurança de tecnologia da informação,
- 7. Logística.

A aprendizagem de máquina é frequentemente dividida em três categorias amplas. São elas (ENAP, 2020):



2.8. MINERAÇÃO DE DADOS (DATA MINING)

Data Mining deve ser entendido como **um passo da descoberta de conhecimento, independentemente se será sobre uma base de dados ou sobre quaisquer outros repositórios de conhecimento.**

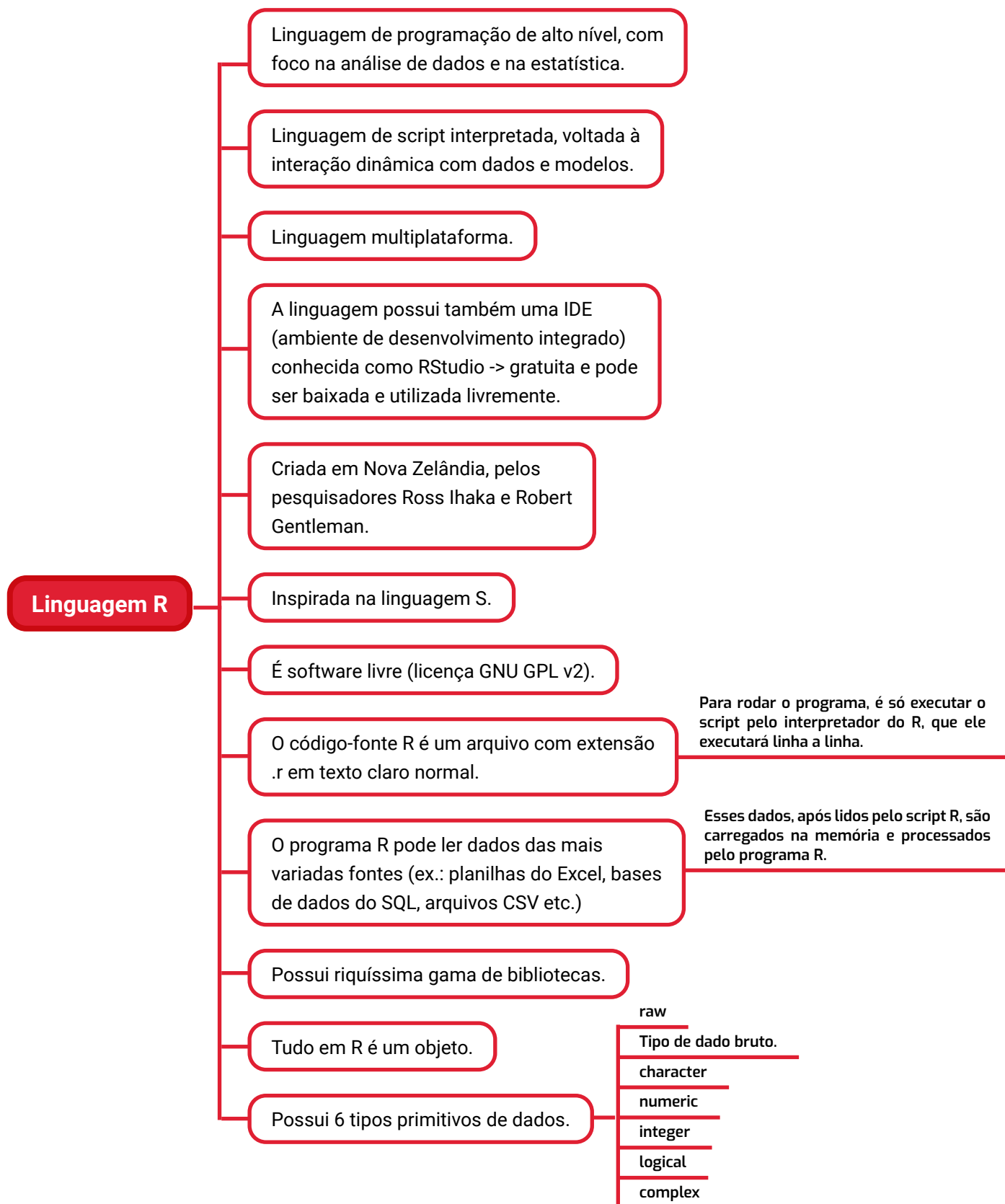
O **Data Mining** é entendido como o **processo de identificar informações relevantes**, tais como padrões, associações, mudanças, anomalias e estruturas, **em grandes conglomerados de dados** que estejam em banco de dados ou outros repositórios de informações.

Fayyad et al. (1996) sintetiza **Data Mining** como “o **processo não trivial de identificar, em dados, padrões válidos, novos, potencialmente úteis e ultimamente compreensíveis**”.

Obs.: **Mineração de Dados (ou Data Mining)**

É o processo de análise de conjuntos de dados que tem por objetivo a **descoberta de padrões interessantes** e que possam **representar informações úteis**.

2.9. LINGUAGEM R



3. TIPOS DE ANÁLISE DE DADOS

A **análise de dados** é frequentemente dividida em alguns **tipos**, baseada nos propósitos gerais de sua realização (ENAP, 2020). Vamos à descrição desses tipos:

| | |
|----------------------------|--|
| Análise Preditiva | Busca prever comportamentos futuros e tendências com base nos dados conhecidos (ENAP, 2020). Visa utilizar dados históricos para aplicar a eventos futuros, desconhecidos . |
| Análise Prescritiva | A análise prescritiva é parecida com a preditiva, mas a ideia é buscar os efeitos dos eventos futuros . Traça, portanto, as possíveis consequências de cada ação . É como se a preditiva fosse capaz de detectar um possível desaquecimento do mercado no futuro, enquanto a prescritiva seria responsável por <u>prever o impacto</u> desse desaquecimento nas vendas da empresa (ENAP, 2020). |
| Análise Descritiva | Foca no presente , visando descrever as características dos dados e eventos correntes para subsidiar decisões de efeitos imediatos (ENAP, 2020). Representa a área de investigação de dados que busca tanto descrever fatos relevantes , não triviais e desconhecidos dos usuários, como analisar a base de dados, principalmente pelo seu aspecto de qualidade, para validar o conhecimento encontrado . |
| Análise Diagnóstica | Visa entender as relações de causa e efeito entre eventos . Por exemplo, pode-se cruzar fatores como dados de vacinação, saneamento básico e higiene para entender as causas da erradicação de determinada doença em parte da população (ENAP, 2020). |

DIRETO DO CONCURSO

001. (CESPE/TCE-RJ/2021) Com relação a noções de mineração de dados e Big Data, julgue o item que se segue. Na mineração de dados preditiva, ocorre a geração de um conhecimento obtido de experiências anteriores para ser aplicado em situações futuras.



A mineração de **dados preditiva** busca **prever comportamentos futuros e tendências com base nos dados conhecidos** (ENAP, 2020).

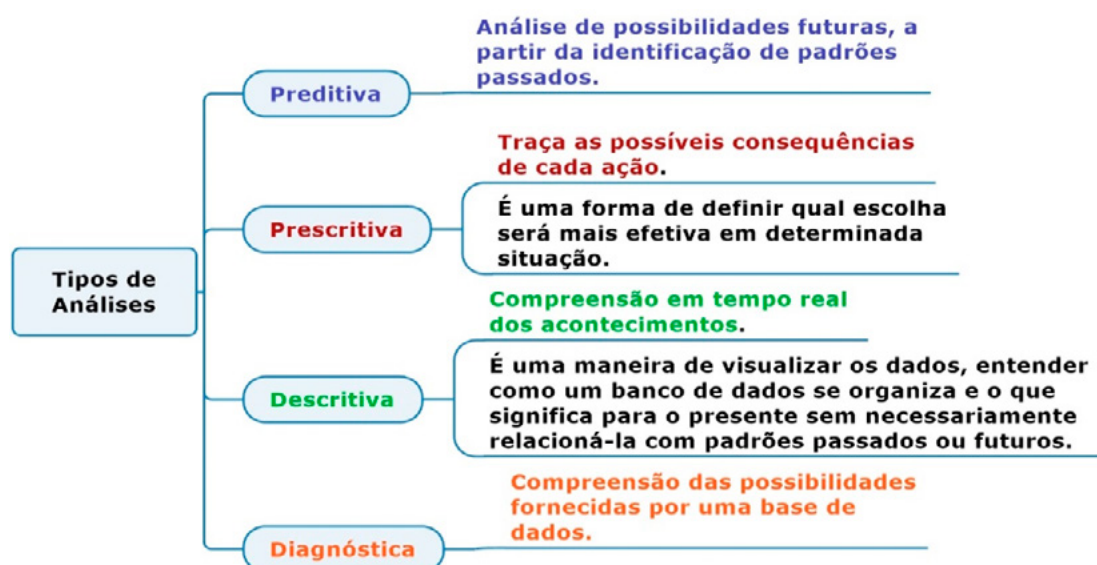


Figura. Tipos de Análises (QUINTÃO, 2021)

Certo.

002. (CESPE/TCE-MG/2018) Uma empresa, ao implementar técnicas e softwares de big data, deu enfoque diferenciado à análise que tem como objetivo mostrar as consequências de determinado evento. Essa análise é do tipo

- a) preemptiva.
- b) perceptiva.
- c) prescritiva.
- d) preditiva.
- e) evolutiva.



A **análise prescritiva** visa prever os efeitos futuros dos eventos. Busca, portanto, **os efeitos dos eventos futuros**.

Letra c.

4. As VARIÁVEIS NA ANÁLISE EXPLORATÓRIA DE DADOS

As características estudadas pela **análise exploratória de dados** são conhecidas como **variáveis**.

Chamamos de **variável** toda **característica que se pretende avaliar estatisticamente em um determinado conjunto de elementos** (**amostra** ou **população**). Cada elemento desse conjunto terá uma medida para essa característica, seja ela expressa em termos numéricos ou não.

Nesse contexto, existem dois **tipos de variáveis** a serem consideradas: a quantitativa e a qualitativa.



Figura. Tipos de Variáveis (QUINTÃO, 2021)

| Variável Quantitativa | |
|---|--|
| <p>Pode ser expressa em termos numéricos (altura, peso, número de ocorrências, dias, tempo, etc.).</p> | <p>Dividimos esse tipo de variável em <u>duas espécies</u>: as variáveis quantitativas discretas, e as quantitativas contínuas.</p> <ul style="list-style-type: none"> • Discretas: expressas em números contáveis. O número de valores possíveis é finito ou "enumerável". Exemplos: número de passageiros por viagem, número de crianças por sala de aula, número de computadores por casa, número de revistas vendidas, quantidade de consultas médicas, quantidade de habitantes das cidades brasileiras etc. • Contínuas: são mensuráveis em um intervalo, ou seja, podem assumir qualquer número (inteiro ou não) dentro de um intervalo de valores. Exemplos: tempo para chegar ao trabalho, distância percorrida, tempo para executar uma tarefa, altura dos alunos de uma escola. |

| Variável Qualitativa | |
|---|--|
| <p>É expressa não por valores, mas por um atributo.</p> <p>Por exemplo: nacionalidade, religião, cor dos olhos, sexo, preferência musical, etc.</p> | <p>Dividimos esse tipo em duas espécies, que são: variáveis qualitativas nominais e ordinais.</p> <ul style="list-style-type: none"> As variáveis qualitativas nominais são aquelas em que não se pode estabelecer uma ordem para elas: intenção de voto, cor dos olhos, time de futebol, marcas de carros, nome de bebidas, sexo - masculino ou feminino etc. As variáveis qualitativas ordinais são aquelas em que é possível estabelecer uma ordem: marco temporal (século XIX, XX, XI), grau de escolaridade (1º, 2º, 3º grau), escala de frequência (pouco, médio, muito), nível de escolaridade - (analfabeto / fundamental / médio / superior) etc. |

Obs.: Os níveis de mensuração podem ser **quantitativos** (discretos ou contínuos) ou **qualitativos** (nominais ou ordinais).

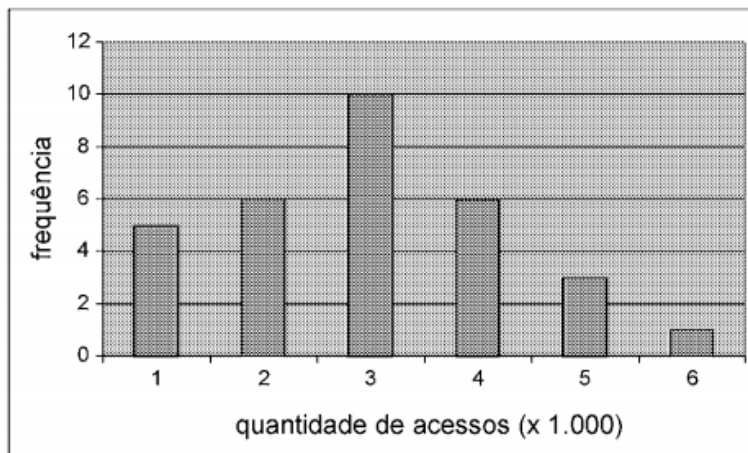
Em um experimento, ainda podemos considerar as variáveis dependentes ou independentes.

| Variável | Descrição |
|------------------------------|--|
| Variável Dependente | Representa uma grandeza cujo valor <i>depende</i> de como a variável independente é manipulada. São apenas medidas ou registradas. |
| Variável Independente | Representa uma grandeza que está sendo manipulada em um experimento. São as variáveis manipuladas. |

Obs.: O nível de manipulação das variáveis se refere a uma variável independente.
A manipulação de uma variável independente pode ser realizada em um ou mais níveis.
O nível mínimo de manipulação é de presença-ausência da variável independente.

DIRETO DO CONCURSO

003. (CESPE/SERPRO/ANALISTA/2010)

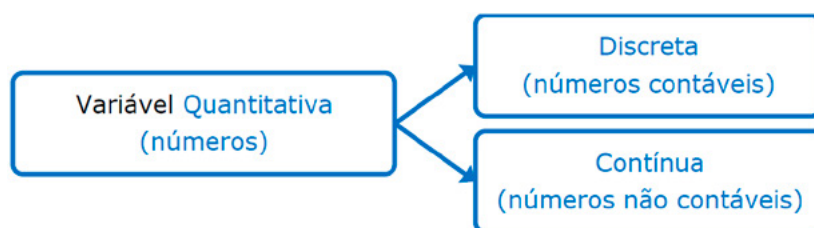


Certa empresa, em determinado mês, realizou levantamento acerca da quantidade diária de acessos simultâneos ao seu sistema, cujo resultado é mostrado na figura acima. A partir das informações apresentadas nessa figura, e considerando que a distribuição da quantidade diária de acessos simultâneos é representada pela variável X , julgue os itens seguintes. É correto classificar a variável X como uma variável quantitativa ordinal.

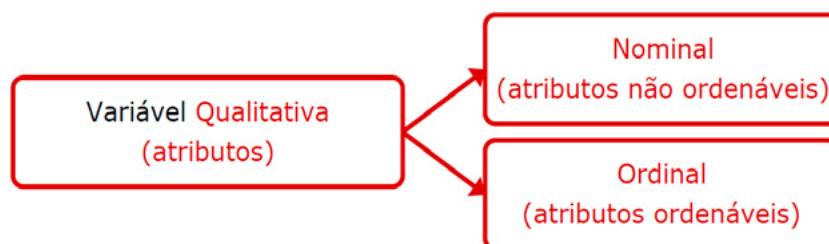


A variável denominada “Quantidade de Acessos” é uma variável **quantitativa**, pois pode ser expressa em termos numéricos.

A partir do indicado no gráfico, podemos afirmar, ainda, que é uma variável **quantitativa discreta**, pois a quantidade de acessos é um número “contável”.



Conforme vimos nesta aula, as variáveis **ordinais** são relacionadas com as variáveis **qualitativas**.



Errado.

004. (CESPE/TCU/ANALISTA DE CONTROLE EXTERNO/2008)

| tipo | 2006 | 2007 |
|------------------------|------------|------------|
| funcionários | 70 | 74 |
| estagiários | 16 | 16 |
| prestadores de serviço | 14 | 10 |
| total | 100 | 100 |

Demonstrações contábeis da CAIXA – Exercício de 2007.
Internet: <www.caixa.gov.br> (com adaptações).

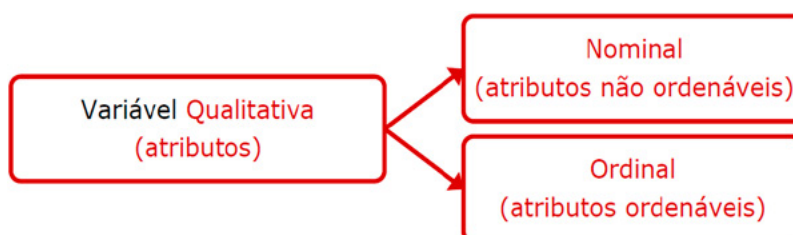
Considerando a tabela anterior, que apresenta a distribuição do quadro de colaboradores da CAIXA, em mil pessoas, no final dos anos de 2006 e 2007, julgue o item seguinte.

Se uma variável X registra a classificação de cada pessoa do quadro como funcionária, estagiária ou prestadora de serviço, então X é uma variável qualitativa.



Conforme vimos nesta aula, a variável qualitativa é aquela **expressa** não por valores, mas **por um atributo**. Por exemplo: nacionalidade, religião, profissão, sexo, cor dos olhos, etc.

Pode-se observar que a variável “classificação” **registra os atributos dos dados**, então, trata-se de uma **variável qualitativa**.



Certo.

005. (CESPE/TCE-PB/AUDITOR DE CONTAS PÚBLICAS/2018) Na análise exploratória de dados, são examinados conjuntos de dados cujas principais características serão resumidas, muitas vezes com métodos visuais. Acerca desse assunto, assinale a opção correta.

- a) Uma variável de nível de manipulação nominal identifica as categorias.
- b) Uma variável de nível de manipulação contínua pode assumir infinitos valores.
- c) Uma variável de nível de mensuração discreta pode assumir apenas alguns valores.
- d) Uma variável de nível de mensuração define como uma variável se relaciona com as outras.
- e) Uma variável de nível de manipulação define a quantidade de informação apresentada por cada variável.



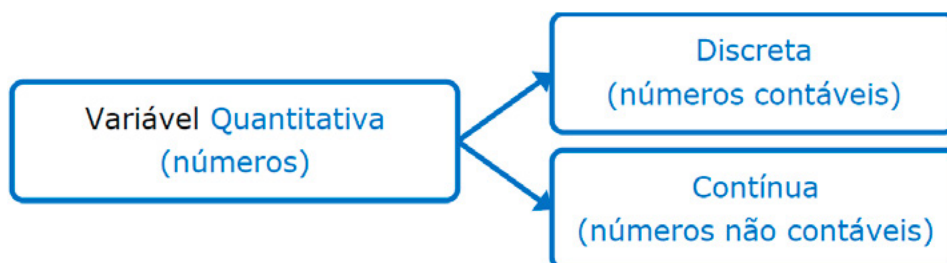
a) Errada. Uma variável de nível de **mensuração** nominal identifica as categorias.

- b) Errada. Uma variável de nível de **mensuração** contínua pode assumir infinitos valores.
- c) Certa. Uma variável de nível de **mensuração** discreta pode assumir apenas alguns valores.
- d) Errada. Uma variável de nível de **manipulação** define como uma variável se relaciona com as outras.
- e) Errada. Uma variável de **mensuração do tipo quantitativa** define a quantidade de informação apresentada por cada variável.

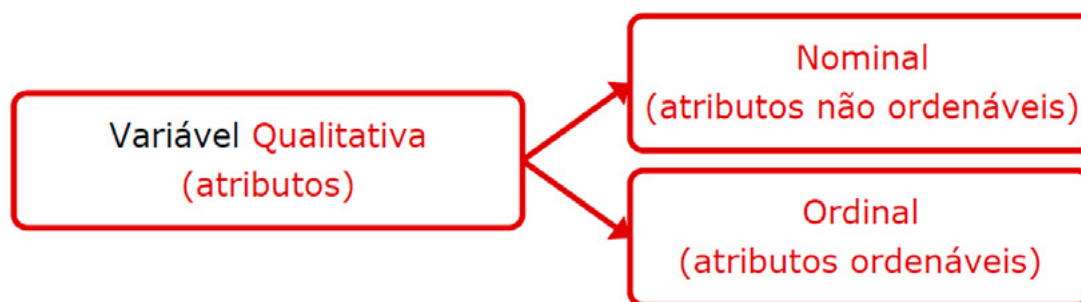
Letra c.

006. (CESPE/CEBRASPE/MINISTÉRIO DA ECONOMIA/TECNOLOGIA DA INFORMAÇÃO/CIÊNCIA DE DADOS/2020) Acerca de conceitos, premissas e aplicações de big data, julgue o item subsequente.

Um atributo é denominado ordinal quando as variáveis podem ser colocadas em ordem, mas não é possível quantificar a diferença entre os resultados.



A **variável qualitativa** é aquela expressa não por valores, mas por um atributo. Elas também podem ser divididas da seguinte maneira:



- As **variáveis qualitativas nominais** são aquelas em que **não se pode estabelecer uma ordem para elas**: intenção de voto, cor dos olhos, time de futebol, marcas de carros, nome de bebidas, sexo - masculino ou feminino etc.
- As **variáveis qualitativas ordinais** são aquelas em que **é possível estabelecer uma ordem**: marco temporal (século XIX, XX, XI), grau de escolaridade (1º, 2º, 3º grau), escala de frequência (pouco, médio, muito), nível de escolaridade – (analfabeto / fundamental / médio / superior) etc.

Certo.

5. O PROCESSO DE *DATA SCIENCE*

Data science ou **ciência de dados**, conforme destaca ENAP (2020) é a arte de **extrair conhecimento por meio dos dados** para se **tomar melhores decisões, realizar previsões e entender o passado**.

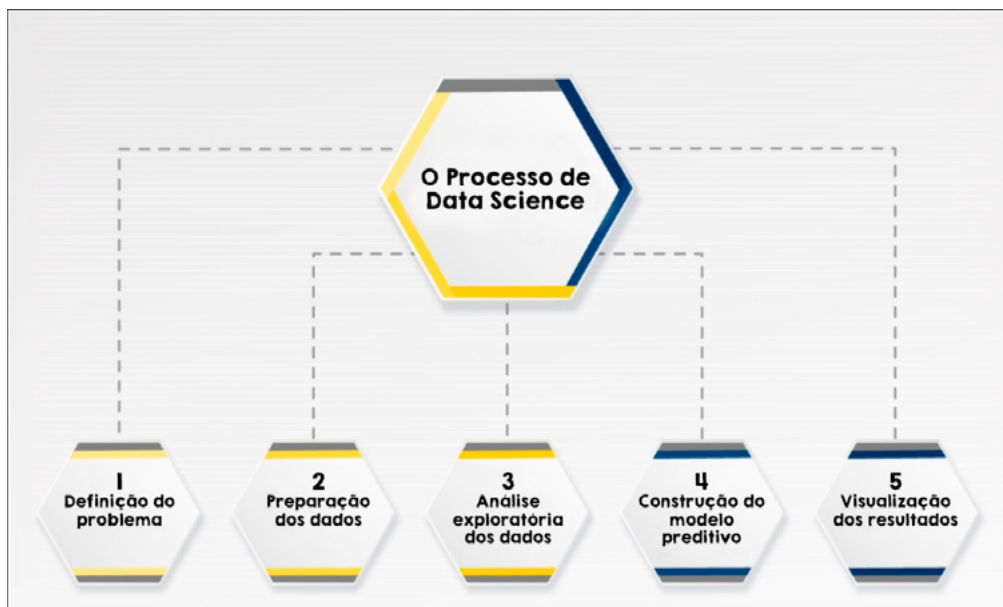
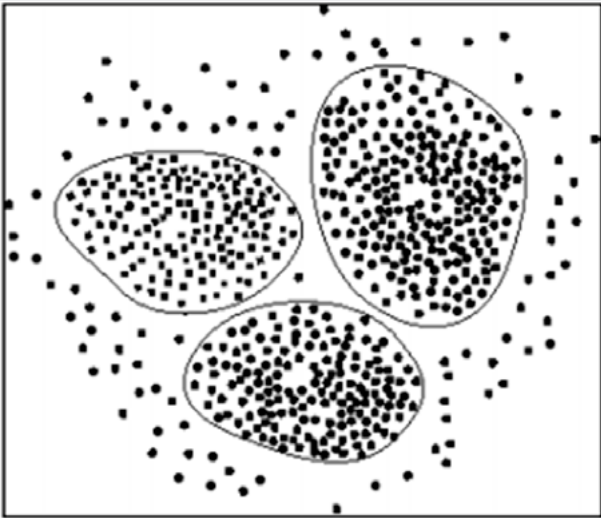


Figura. Processo de Data Science. Fonte: ENAP (2020)

Vamos a uma descrição geral de cada uma das etapas, conforme ENAP (2020).

| | |
|---|---|
| 1 | <p>Definição do problema: nesta etapa busca-se entender o problema e as questões de negócio que devem ser respondidas. Alguns questionamentos:</p> <ul style="list-style-type: none"> • O que se quer resolver com a <u>análise</u>? • <u>Que tipo de análise</u> será feita? Descritiva, diagnóstica, preditiva? • <u>Quais dados</u> são necessários? |
| 2 | <p>Preparação dos dados: está relacionada com a obtenção, limpeza, normalização e transformação dos dados.</p> |
| 3 | <p>Análise exploratória dos dados: busca-se obter um panorama de <u>como os dados estão organizados</u>. A apresentação dos dados é fundamental, pois o objetivo é entender as características e os relacionamentos deles. Algumas questões:</p> <ul style="list-style-type: none"> • Quais são os tipos das variáveis (atributos)? • Como estão as distribuições dos dados? • Existem valores <i>missing</i> (NA/Null)? • Existem variáveis redundantes? • Existem outliers (registros que apresentam grande discrepância em relação à maioria dos registros). A figura seguinte identifica visualmente a presença de <i>outliers</i>, em que os pontos externos aos polígonos são valores fora dos padrões da população observada. |

| | |
|---|---|
| |  <p>Quais variáveis possuem correlação?</p> |
| 4 | <p>Construção do modelo preditivo.</p> <p>A modelagem preditiva permite realizar previsões com base em dados históricos por meio da criação de um modelo. Nem todos os projetos passam por esta etapa, depende da definição do problema de negócio.</p> |
| 5 | <p>Visualização dos resultados: são apresentados os resultados da análise por meio de várias formas, como: <i>dashboard</i> (painel gerencial), relatório, planilha, arquivo csv, etc. Existem diversas ferramentas que podem ajudar nesta parte.</p> |

Obs.: Para cada etapa do processo de *data science*, existem diversas ferramentas que podem ser utilizadas, dentre elas a linguagem R.

6. ARQUITETURA DE BI (*BUSINESS INTELLIGENCE* OU INTELIGÊNCIA DE NEGÓCIOS)

Nesse momento do estudo, você já deve estar familiarizado com o conceito de BI (*Business Intelligence*), que abrange **aplicativos, ferramentas e metodologias** usadas para **coleta, tratamento, armazenamento, recuperação e disseminação** de informações com o objetivo de auxiliar o processo de tomada de decisões organizacionais complexas.

Vamos então apreciar as diversas etapas do processo de BI.

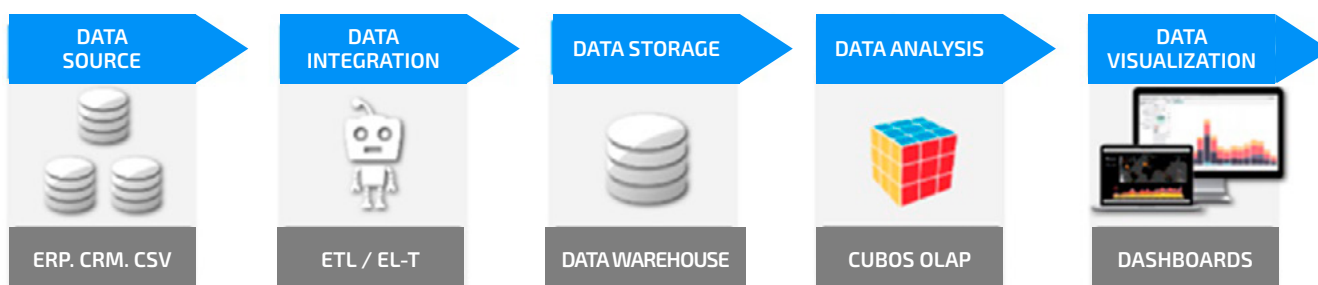


Figura. Arquitetura de BI. Fonte: BI Academy (2020)

- **Data Source (Fonte de Dados)**

Geralmente, o processo de BI começa no *Data Source*. É nesse local em que se encontram os dados que ainda não viraram informação.

Alguns exemplos: arquivos de textos e CSVs, arquivos de Excel, ERP, CRM, etc.

- **Data Integration (Integração de Dados)**

Depois de você ter identificado o *Data Source*, a primeira fase na arquitetura de BI é **Data Integration**.

Agora é o momento de desenvolver e executar o processo de **ETL** (**E** – Extração (Extract); **T** – Transformação (Transform); **L** – Carga (Load)).

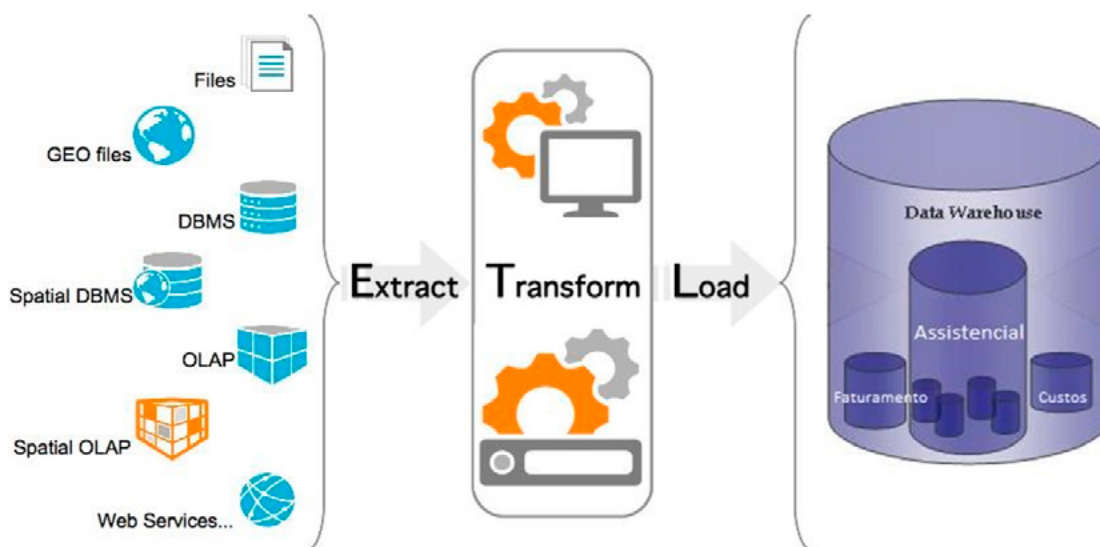


Figura. ETL. Fonte: http://laudosweb.hospitalorione.com.br/manual_html/etl_v4.1.0_painel.png

ETL

Os diferentes processos concentrados no conceito de extração, transformação e carga de dados em um Data Warehouse denominam-se ETL, em inglês Extract – Transform – Load.

Na etapa de **extração**, os dados relevantes ou de interesse são capturados de uma base de dados estruturada ou fontes diversas (exemplo redes sociais).

Em seguida, na etapa de **transformação** os dados são processados ou moldados para compor a informação na base do DW. Este processamento inclui atividades como normalização e geração de chave que identifica o registro na base.

Por fim, a etapa de **carga** consiste na inserção dos dados na base do Data Warehouse ou Data Mart.

As ferramentas de software ETL têm como função a **extração** de dados de diversos sistemas, a **transformação** desses dados de acordo com as regras de negócio e a **carga** dos dados em um Data Mart ou um DW.

Existem diversas ferramentas que podem ser utilizadas para implementar isso, como: Pentaho Data Integration, Power Center, Talend, SSIS (*SQL Server Integration Services*) e ODI (*Oracle Data Integrator*).

- **Data Storage (Armazenamento de Dados).** O **Data Warehouse** (Armazém de Dados) é o repositório de dados centrais, em que fica a informação. Para a criação do *Data Warehouse*, podemos usar bancos de dados relacionais, como PostgreSQL, Oracle Database, SQL Server e Teradata.
- **Data Analysis (Análise de Dados).** Aqui temos os **cubos OLAP** (*Online Analytical Processing*), *Data Mining*, etc.

É nessa etapa, por exemplo, que começamos a cruzar as informações disponibilizadas, bem como podem ser feitas análises iniciais e a identificação de padrões passados, problemas futuros etc.

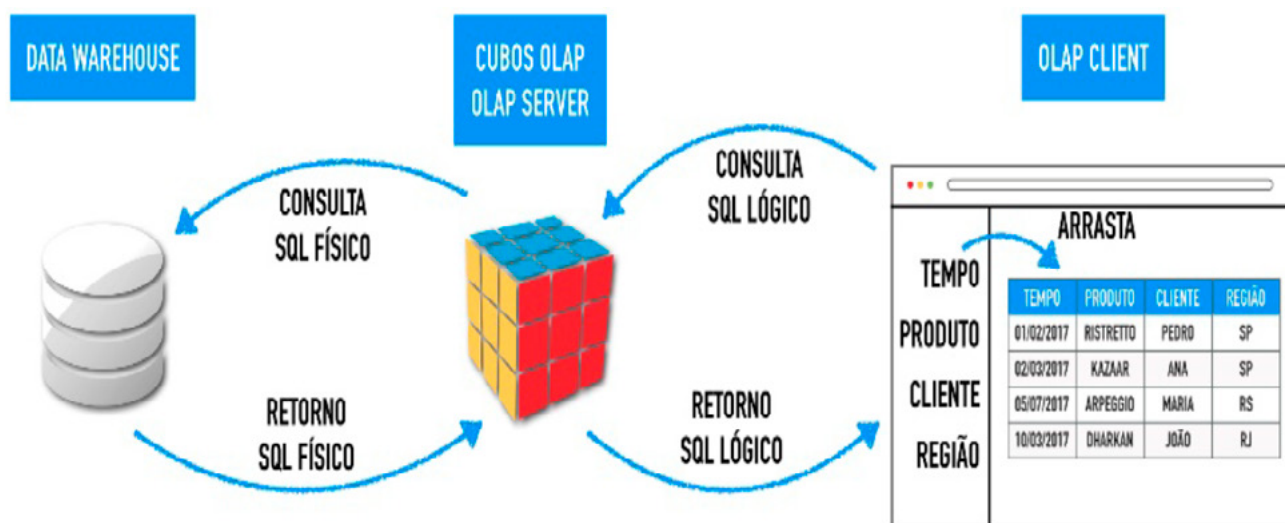


Figura. Data Analysis. Fonte: BI Academy (2020)

Exemplos de ferramentas que se pode utilizar para criar os cubos: Qlikview, Microsoft BI, Microstrategy, OBIEE, Pentaho Mondrian etc.

- **Data Visualization (Visualização de Dados)**

Os usuários podem percorrer as diferentes **dimensões** e seus cruzamentos, **visualizando** os valores resultantes das medidas em cada caso.

Obs.: A **visualização de dados** é mais fácil de implementar quando os dados necessários estão em um **Data Warehouse** ou, melhor ainda, em um banco de dados multidimensional especial ou servidor.

Desde o fim dos anos 90, a **visualização de dados evoluiu tanto na computação convencional**, em que é integrada às ferramentas e aplicações de suporte à decisão, como na **visualização inteligente**, que inclui a interpretação de dados (informação) (TURBAN, 2009).

Veja a seguir alguns tipos de **ferramentas** que podem ser **utilizadas para navegar pelo cubo**:

- **planilhas de cálculo**: podem ser conectadas com a estrutura dimensional e alimentar uma tabela dinâmica com a informação retirada dos cubos;
- **painéis de controle (Dashboards)**: conectam-se com a estrutura dimensional e geram **indicadores** que permitem uma rápida visão do *status* atual das variáveis básicas e sua relação com os objetivos da empresa e/ou negócio. **Apresenta visualmente as informações mais importantes e necessárias para tomada de decisão**. Assim, em uma única tela, por exemplo, é possível organizar análises, gráficos, KPI (*Key Performance Indicator* - é uma métrica associada a um processo) e *reports* de toda a empresa ou de um ponto de vista do negócio ou departamento.

Importante: *Dashboard* bom não é o que tem gráficos bonitos, é o que entrega as informações necessárias para a tomada de decisão.



Figura. Exemplo de Dashboard. Fonte: BI Academy (2020)

- **Soluções ou aplicativos desenvolvidos sob medida** especialmente para a empresa, ou **aplicativos genéricos** ou orientados para algum mercado em particular;
- **Geradores de relatórios** que podem ser conectadas na estrutura dimensional e gerar relatórios com a informação retirada dos cubos. A técnica de *BI reporting* consegue transformar os dados em informações compreensíveis, flexíveis e fáceis de serem analisadas pelo usuário final em relatórios pré-formatados.

Obs.: O que é “Visualização de Dados”?

O termo **visualização de dados** está relacionado às tecnologias que dão suporte à visualização e, algumas vezes, à interpretação de dados e informações em vários pontos ao longo da cadeia de processamento de dados (TURBAN, 2009).

Ela inclui imagens digitais, sistemas geográficos, interfaces gráficas de usuário, gráficos, realidade virtual, representações dimensionais, vídeos e animações (TURBAN, 2009).

As **ferramentas visuais** podem ajudar a identificar relações, como por exemplo, tendências. Ao usar ferramentas visuais de análise, gerentes, engenheiros e outros profissionais podem reconhecer problemas que passaram despercebidos, durante anos, pelos métodos de análise padrão (TURBAN, 2009).

RESUMO

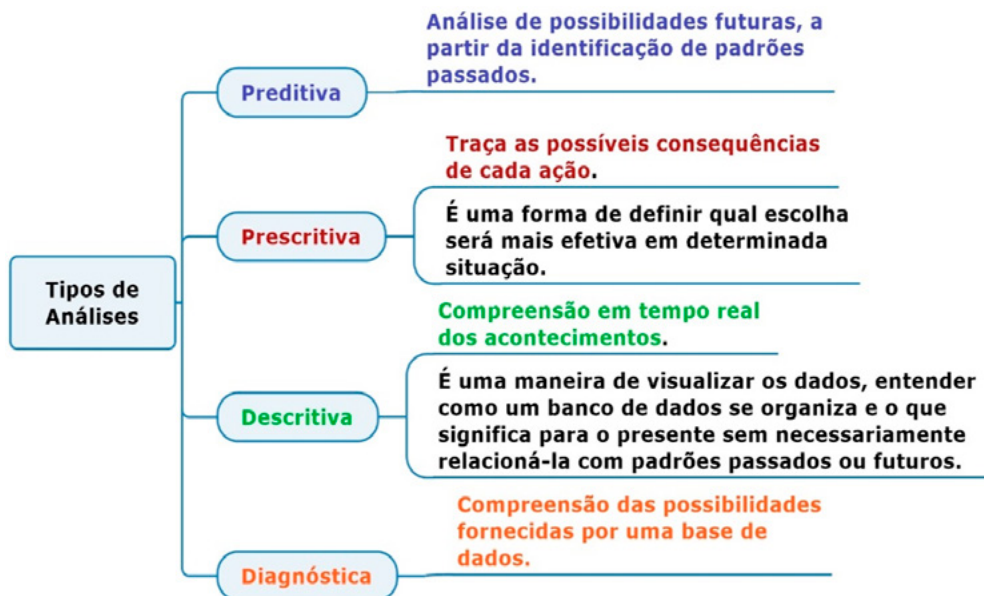
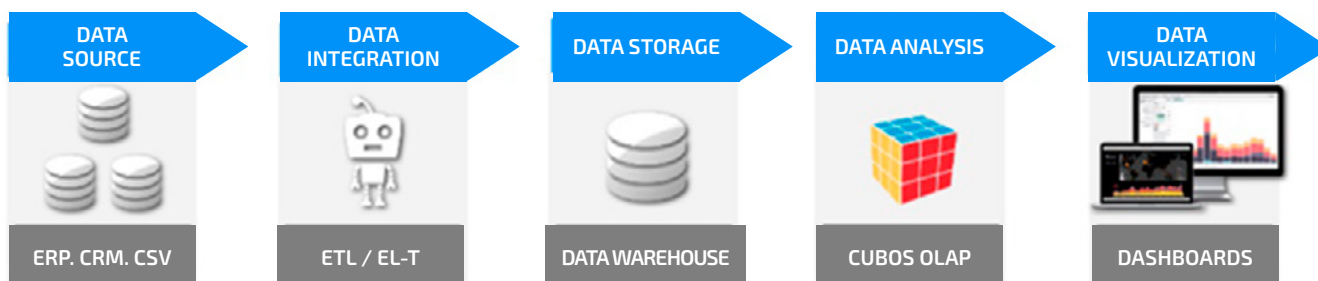


Figura. Tipos de Análises (QUINTÃO, 2021)



Figura. Tipos de Variáveis (QUINTÃO, 2021)

Vamos então apreciar as diversas etapas do processo de BI.



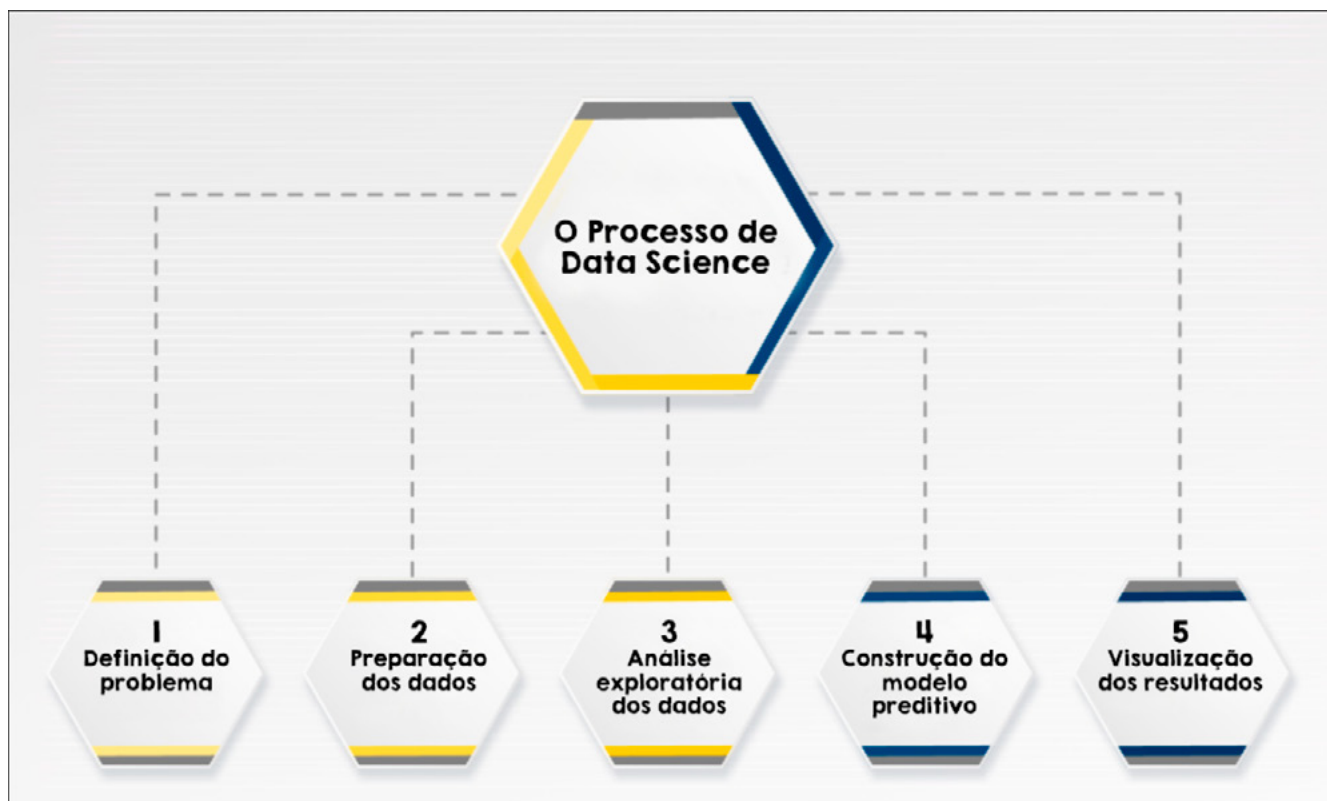
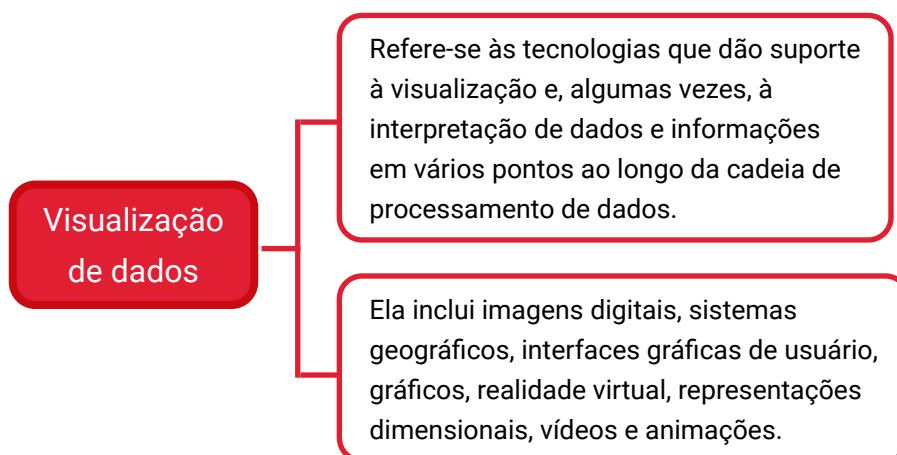


Figura. Processo de Data Science. Fonte: ENAP (2020)

Análise de Dados

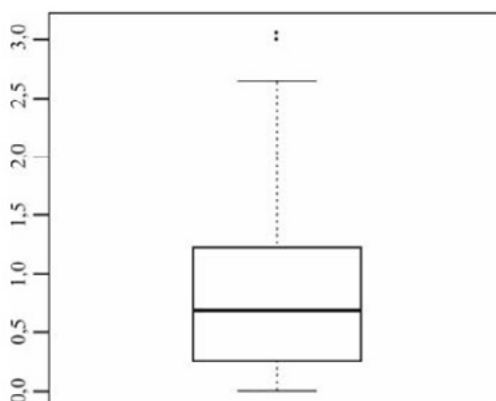
Processo de inspecionar/coletar, limpar, transformar e modelar dados, para que eles possam fornecer informações úteis para a tomada de decisão.

Visualização de Dados



QUESTÕES DE CONCURSO

007. (CESPE/INPI/ANALISTA DE PLANEJAMENTO/ESTATÍSTICA/2013) Um indicador de desempenho **X** permite avaliar a qualidade dos processos de governança de instituições públicas. A figura mostra, esquematicamente, a sua distribuição, obtida mediante estudo amostral feito por determinada agência de pesquisa. A tabela apresenta estatísticas descritivas referentes a essa distribuição.



| | |
|------------------------|------|
| média amostral | 0,80 |
| desvio padrão amostral | 0,70 |
| primeiro quartil | 0,25 |
| mediana | 0,70 |
| terceiro quartil | 1,20 |
| mínimo | 0 |
| máximo | 3,10 |

Com base nessas informações, julgue o item a seguir.

X – representa uma variável qualitativa ordinal.



X – representa uma variável **quantitativa contínua**. As variáveis quantitativas contínuas são **mensuráveis em um intervalo**, ou seja, **podem assumir qualquer número (inteiro ou não) dentro de um intervalo de valores**. Exemplos: tempo para chegar ao trabalho, distância percorrida, tempo para executar uma tarefa, altura dos alunos de uma escola.

Errado.

008. (CESPE/INPI/ANALISTA DE PLANEJAMENTO/ESTATÍSTICA/2013)

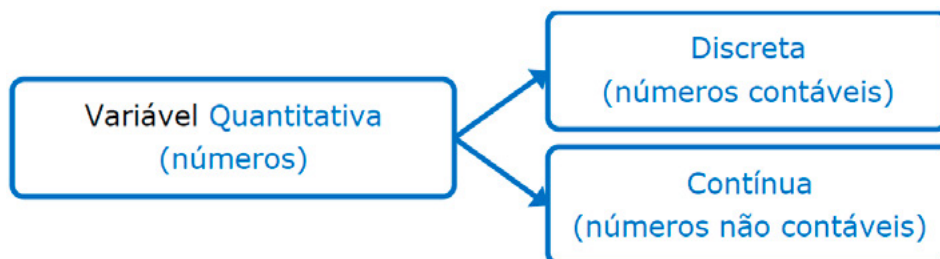
0 0 3 3 3 4 1 2 2 1 1 2 3 3 4 1 2 2 0 4

Considerando os dados acima, que representam a quantidade de doutores presentes em pesquisas que geraram patentes industriais, julgue os próximos itens.

O nível de mensuração da variável é qualitativo ordinal.



O nível de mensuração da variável é **quantitativo discreto**.



Errado.

009. (CEBRASPE/CESPE/AUDITOR ESTADUAL/TCM-BA/CONTROLE EXTERNO/2018) Acerca de visualização e análise exploratória de dados, assinale a opção correta.

- a) As variáveis são características que podem ser observadas ou medidas em cada elemento pesquisado, sendo classificadas em inteiras, textuais ou imagens.
- b) A entrada de dados deve assumir convenções tais como: os dados devem estar no formato de matriz; cada linha da matriz deve ter pelo menos um atributo identificador; e cada coluna da matriz deve ter um nome distinto.
- c) A informação de uma ou mais variáveis que não estiver disponível (por exemplo, falta de resposta, rasura etc.) deve ser marcada (ou indicada) com zero.
- d) Um arquivo csv (comma separated values), no qual as colunas de um conjunto de dados são separadas por vírgula, pode ser aberto em qualquer editor de texto ou planilha eletrônica.
- e) A eliminação completa (casewise deletion) consiste em eliminar os arquivos que apresentem mais de um dado perdido.



Vamos aos comentários das assertivas:

- a) Errada. As variáveis são características que podem ser observadas ou medidas em cada elemento pesquisado, sendo classificadas em quantitativas e qualitativas.



Figura. Tipos de Variáveis (QUINTÃO, 2021)

b) Errada. A **entrada de dados** deve assumir certas convenções, conforme <https://www.inf.ufsc.br/~andre.zibetti/probabilidade/aed.html>

- Os dados devem estar no formato de **matriz**
- Cada linha da matriz corresponde a uma **unidade experimental**.
 - Elemento da população ou amostra no qual observamos as variáveis
- Cada coluna da matriz corresponde a uma **variável**

c) Errada. O zero, nesse caso, indicará uma quantidade que foi medida.

d) Certa. Um **arquivo csv** (*comma separated values*), em que as colunas de um conjunto de dados são separadas por vírgula, pode ser aberto em qualquer editor de texto ou planilha eletrônica.

e) Errada. A eliminação completa ("*casewise deletion*") consiste em eliminar todos os casos (linhas) que apresentem pelo menos um dado perdido.

Letra d.

010. (CESPE/TCE-SC/AUDITOR-FISCAL DE CONTROLE EXTERNO/2016) Para a realização de prognósticos por meio de técnicas de mineração de dados, parte-se de uma série de valores existentes obtidos de dados históricos bem como de suposições controladas a respeito das condições futuras, para prever outros valores e situações que ocorrerão e, assim, planejar e preparar as ações organizacionais.



A mineração de dados pode mostrar como certos atributos dos dados se comportarão no futuro. A análise de prognóstico (ou previsão) representa a área de investigação de dados que busca inferir resultados a partir dos padrões encontrados na análise descritiva, ou seja, prognosticar o comportamento de um novo conjunto de dados. Parte de uma série de valores obtidos de dados históricos bem como de suposições controladas a respeito das condições futuras, para prever outros valores e situações que ocorrerão e, assim, planejar e preparar as ações organizacionais. Por exemplo um prognóstico pode descobrir padrões nos dados que ajudam os gerentes a estimar o valor futuro de variáveis com números de vendas.

Referência: https://pt.wikipedia.org/wiki/Minera%C3%A7%C3%A3o_de_dados.

Certo.

011. (INÉDITA/2020) A grande vantagem dos dados estruturados nos sistemas computacionais é que eles são mais facilmente buscáveis e acessíveis, no entanto, os dados não estruturados não seguem nenhum tipo de modelo ou organização definida.



Isso mesmo! Os **dados estruturados** são **mais facilmente buscáveis e acessíveis**, enquanto os **dados não estruturados** (Exemplos: imagens, textos, vídeos, documentos...) não seguem

nenhum tipo de modelo ou organização predefinida. **Os dados não estruturados são abundantes na Internet e nas organizações, sendo a maior parte dos dados no ambiente corporativo.**
Certo.

012. (CESPE/TCE-PA/AUDITOR DE CONTROLE EXTERNO/ÁREA INFORMÁTICA/ANALISTA DE SEGURANÇA/2016) Julgue o item subsecutivo, referente à gestão do ciclo de vida da informação – ILM (*Information Lifecycle Management*). 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.



Os dados não estruturados são abundantes na internet e nas organizações, sendo **a maior parte dos dados (cerca de 90%) no ambiente corporativo**. Assim, demandam mais espaço de armazenamento e um gerenciamento mais cauteloso.

Errado.

013. (CESPE/TCE-SC/AUDITOR DE TI/2016) A respeito de dados estruturados, não estruturados e abertos, julgue os itens subsequentes. 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.



Os **dados estruturados** são aqueles que **têm um formato específico e são representados em formato estrito**. Os registros de uma tabela de um banco de dados relacional são exemplos de dados estruturados. Com eles, você sabe como são os campos de uma tabela. Por exemplo, uma tabela Aluno pode ser formada por um campo ID que será um número inteiro, um nome que será uma cadeia de caracteres e assim sucessivamente.

Já os **dados semiestruturados** são **auto descritivos**, assim, o esquema de representação está presente (de forma explícita ou implícita) juntamente com o dado. Além disso, este tipo de dados apresenta como característica a **estrutura parcial**, que significa que nem todos dados estão organizados em uma estrutura.

Conforme visto, o item fala justamente dos dados semiestruturados, e não dos estruturados. Por isso, item **errado**.

Errado.

014. (CESPE/TCE-PB/AUDITOR DE CONTAS PÚBLICAS/DEMAIS ÁREAS/2018) Dados não estruturados de um mesmo grupo possuem as mesmas descrições e, conseqüentemente, os mesmos atributos.



Os **dados não estruturados** **não** têm estrutura definida, dessa forma **não** seguem nenhum modelo ou padrão.

Cabe destacar também que a **estrutura dos dados semiestruturados** é irregular em virtude da **ausência de regras rígidas de esquema**, o que permite que elementos semelhantes tenham conjuntos de atributos diferentes.

Errado.

015. (CESPE/TCE-PB/AUDITOR DE CONTAS PÚBLICAS/DEMAIS ÁREAS/2018) Por padrão, dados não estruturados são organizados em blocos semânticos.



Os **dados estruturados** é que estão comumente armazenados em **blocos semânticos**, assim, elementos de natureza similar são agrupados nas mesmas estruturas.

Errado.

016. (CESPE/TCE-PB/AUDITOR DE CONTAS PÚBLICAS/DEMAIS ÁREAS/2018) A alta heterogeneidade facilita as consultas aos dados não estruturados, desde que estes estejam ligados por ponteiros.



A alta heterogeneidade **dificulta** as consultas aos **dados não estruturados**!

Errado.

017. (CESPE/SERPRO/ANALISTA/DESENVOLVIMENTO DE SISTEMAS/2008) No que concerne a tópicos avançados, julgue os itens subsequentes.

Software de inteligência empresarial, como mineração de dados, CRM e datawarehouse, por exemplo, aplicam métodos de inteligência artificial e robótica avançados para a representação e extração da informação em grandes bases de dados.



Mineração de dados não é um software!

A **mineração de dados** é um campo interdisciplinar que reúne técnicas de **aprendizado de máquina**, **reconhecimento de padrões**, **estatísticas**, **banco de dados** e **visualização** para abordar a questão da extração de informações a partir de grandes bases de dados”

(Evangelos Simoudis, citado em Daniel T. Larose, Discovering Knowledge in Data – An Introduction to Data Mining).



Figura. Disciplinas envolvidas com Mineração de Dados (HAN & KAMBER, 2006)

É o processo de análise de conjuntos de dados que tem por objetivo a descoberta de padrões interessantes e que possam representar informações úteis.

Errado.

018. (CESPE/POLÍCIA FEDERAL/PAPILOSCOPISTA POLICIAL FEDERAL/2018) Acerca da definição de dados e informação e sua representação em sistemas de informação, julgue o item que segue.

Dados são fatos que descrevem os objetos de informação, por exemplo, eventos e entidades.

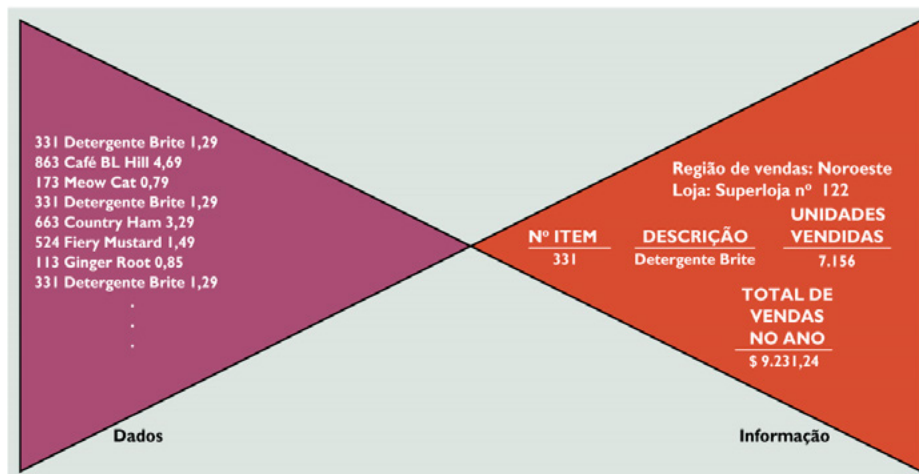


Conforme destaca Baltzan (2016), um **fato** é a **confirmação ou validação de um evento ou objeto**. No passado, as pessoas conheciam os fatos pelos livros, em primeiro lugar. Hoje, basta apertar um botão e elas podem descobrir qualquer coisa, de qualquer lugar, a qualquer momento.

- **Dados** são **fatos brutos** que **descrevem as características de um evento ou objeto**.
- **Informações** são dados convertidos em um contexto significativo e útil por usuários finais específicos.

Informações são um **conjunto de dados** cuja forma e conteúdo são apresentados de maneira útil para uso em um processo de tomada de decisão.

Complementando, **informações** são **conjuntos de dados** significativos e úteis a seres humanos em processos como o de tomada de decisões. A transformação de dados em informação é frequentemente realizada através da apresentação dos dados em uma forma compreensível ao usuário. As informações são produzidas pelo processamento de dados. Elas são utilizadas para revelar o significado dos dados.



Fonte: (O'BRIEN, 2006)

Inteligência de negócios (BI – Business Intelligence) é a informação obtida a partir de múltiplas fontes, como fornecedores, clientes, concorrentes, parceiros e indústrias, que analisa tendências, padrões e relacionamentos para a tomada de decisão estratégica.

Certo.

GABARITO

- 7. E
- 8. E
- 9. d
- 10. C
- 11. C
- 12. E
- 13. E
- 14. E
- 15. E
- 16. E
- 17. E
- 18. C

REFERÊNCIAS

ALECRIM, E. **O que é big data?** 2013. Disponível em: <<http://www.infowester.com/big-data.php>>. Acesso em: 05 jul. 2020.

BIG DATA BUSINESS. **Big Data Analytics: você sabe o que é?** Disponível em: <<http://www.bigdatabusiness.com.br/voce-sabe-o-que-e-big-data-analytics/>> Acesso em: 10 mar. 2019.

_____. **Tipos de análise de Big Data: você conhece todos os 4?** Disponível em: <<http://www.bigdatabusiness.com.br/conheca-os-4-tipos-de-analises-de-big-data-analytics/>>. Acesso em: 10 mar. 2019.

BRASIL. Controladoria-Geral da União. **Portal da Transparência**. Brasília: CGU, c2020. Disponível em: <http://portaltransparencia.gov.br/>. Acesso em: 17 ago. 2020.

BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão. **Painéis**. Brasília: MPDG, [2020]. Ambiente em migração. Disponível em: <http://www2.planejamento.gov.br/planejamento/paineis>. Acesso em: 17 ago. 2020.

BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão. Secretaria de Tecnologia da Informação. **O que são dados abertos?** Brasília: MPDG, [2020]. Disponível em: <http://dados.gov.br/pagina/dados-abertos>. Acesso em: 17 ago. 2020.

BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão. Secretaria de Tecnologia da Informação. **Portal Brasileiro de Dados Abertos**. Brasília: MPDG, [2020]. Disponível em: dados.gov.br. Acesso em: 17 ago. 2020.

BRITO, S. H. B. Afinal, O Que é Big Data? 2013. Disponível em: <<http://labcisico.blogspot.com.br/2013/08/afinal-o-que-e-big-data.html>>.

CÁNEPA, G. **What You Need to Know about Machine Learning**. Birmingham: Packt Publishing, 2016.

CUESTA, H.; KUMAR, S. **Practical Data Analysis**. 2. ed. Birmingham: Packt Publishing, 2016.

ENAP. **Análise de Dados em Linguagem R**. Fundação Escola Nacional de Administração Pública. Brasília, DF. 2020.

FELDMAN, S. A Minute on the Internet in 2019. **Statista**, New York, 29 mar. 2019. Disponível em: <https://www.statista.com/chart/17518/internet-use-one-minute/>. Acesso em: 7 ago. 2020.

FREITAS JUNIOR, J. C. S.; MAÇADA, A. C. G.; OLIVEIRA, M.; BRINKHUES, R. A. Big Data e Gestão do Conhecimento: Definições e Direccionamentos de Pesquisa. **Revista**

Alcance, v. 23, n. 4, p. 529-546, out./dez. 2016. Disponível em: <https://www.redalyc.org/jatsRepo/4777/477749961006/477749961006.pdf>. Acesso em 22 maio 2020.

FUENTES, A. **Hands-On Predictive Analytics with Python**. Birmingham: Packt Publishing, 2018.

GOLLAPUDI, S. **Practical Machine Learning**. Birmingham: Packt Publishing, 2016.

GOMES, G. L. **Análise de dados na prática com R Studio**. DevMedia, Brasília, 2018. Disponível em: <https://www.devmedia.com.br/analise-de-dados-na-pratica-com-r-studio/39279>. Acesso em: 17 ago. 2020.

HADOOP: O que é, conceito e definição. **Cetax**, [s. l.]. Disponível em: <https://www.cetax.com.br/blog/apache-hadoop/>. Acesso em: 25 maio 2020.

JOSEPH, R. **Big Data Analytics in Government: How the Public Sector Leverages Data Insights**. Intellectyx, Denver, 26 jun. 2019. Disponível em: <https://www.intellectyx.com/blog/big-dataanalytics-in-government-how-the-public-sector-leverages-data-insights/>. Acesso em: 7 ago. 2020.

MONTEIRO, L. P. Dados Estruturados e Não Estruturados. **Blog Tecnologia da Informação**. São Paulo: Universidade da Tecnologia, 2019. Disponível em: <https://universidadedatecnologia.com.br/dados-estruturados-e-nao-estruturados/>. Acesso em: 25 jun. 2020.

OZDEMIR, S. **Principles of Data Science**. Birmingham: Packt Publishing, 2016.

PERRICOS, C.; KAPUR, V. Anticipatory government: Preempting problems through predictive analytics. **Deloitte Insights**, New York, 24 jun. 2019. Disponível em: <https://www2.deloitte.com/us/en/insights/industry/public-sector/government-trends/2020/predictive-analytics-ingovernment.html>. Acesso em: 7 ago. 2020.

QUINTÃO, P. L. **Tecnologia da Informação para Concursos**, 2021.

SIX Big Data Use Cases for the Public Sector. **Ingram Micro**, Irvine, 25 jan. 2017. Disponível em: <https://imagine.next.ingrammicro.com/data-center/six-big-data-use-cases-for-the-public-sector>. Acesso em: 17 ago. 2020.

SEMINÁRIO INTERNACIONAL SOBRE ANÁLISE DE DADOS NA ADMINISTRAÇÃO PÚBLICA, 5., 2019, Brasília. **Programa** [...]. Brasília: TCU: Enap, 2019. Disponível em: <http://www.brasildigital.gov.br/brasil-digital/programa/>. Acesso em: 17 ago. 2020.

17 CASOS de uso de machine learning. **Data Science Academy**, [s. l.], 8 ago. 2018. Disponível em: <http://datascienceacademy.com.br/blog/17-casos-de-uso-de-machine-learning/>. Acesso em: 28 junho 2020.

LANTZ, B. **Machine Learning with R**. 2. ed. Birmingham: Packt Publishing, 2015.

LIU, Y. H. **Python Machine Learning By Example**. Birmingham: Packt Publishing, 2017.

REZA, M. **Machine Learning**. Pittsburg, PA: CMUCC, c1995. Disponível em: <http://www.contrib.andrew.cmu.edu/~mndarwis/ML.html>. Acesso em: 7 ago. 2020.

SILVA, M. N. P. da S. **Variáveis na Estatística**. Disponível em: <https://mundoeducacao.uol.com.br/matematica/variaveis-na-estatistica.htm>. Acesso em: 7 abr.2021.

SAS. **MACHINE Learning: o que é e qual sua importância?**, São Paulo, 2019. Disponível em: https://www.sas.com/pt_br/insights/analytics/machine-learning.html. Acesso em: 10 ago. 2020.

VORHIES, W. **Prescriptive versus predictive analytics - a distinction without a difference?** 2014. Disponível em: <<https://www.datasciencecentral.com/profiles/blogs/prescriptive-versus-predictive-analytics-a-distinction-without-a>>. Acesso em: 20 ago. 2020.

WALKOWIAK, S. **Big Data Analytics with R**. Birmingham: Packt Publishing, 2016.

WIKERSON, L. **De que maneira o Big Data melhora nossa vida diária?** 2015. Disponível em: <<http://www.tecmundo.com.br/tecnologia-da-informacao/80027-maneira-big-data-melhora-nossa-vida-diaria-infografico.htm>>. Acesso em: 04 jul. 2020.

Patrícia Quintão



Mestre em Engenharia de Sistemas e computação pela COPPE/UFRJ, Especialista em Gerência de Informática e Bacharel em Informática pela UFV. Atualmente é professora no Gran Cursos Online; Analista Legislativo (Área de Governança de TI), na Assembleia Legislativa de MG; Escritora e Personal & Professional Coach.

Atua como professora de Cursinhos e Faculdades, na área de Tecnologia da Informação, desde 2008. É membro: da Sociedade Brasileira de Coaching, do PMI, da ISACA, da Comissão de Estudo de Técnicas de Segurança (CE-21:027.00) da ABNT, responsável pela elaboração das normas brasileiras sobre gestão da Segurança da Informação.

Autora dos livros: Informática FCC - Questões comentadas e organizadas por assunto, 3ª. edição e 1001 questões comentadas de informática (Cespe/UnB), 2ª. edição, pela Editora Gen/Método.

Foi aprovada nos seguintes concursos: Analista Legislativo, na especialidade de Administração de Rede, na Assembleia Legislativa do Estado de MG; Professora titular do Departamento de Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia; Professora substituta do DCC da UFJF; Analista de TI/Suporte, PRODABEL; Analista do Ministério Público MG; Analista de Sistemas, DATAPREV, Segurança da Informação; Analista de Sistemas, INFRAERO; Analista - TIC, PRODEMGE; Analista de Sistemas, Prefeitura de Juiz de Fora; Analista de Sistemas, SERPRO; Analista Judiciário (Informática), TRF 2ª Região RJ/ES, etc.

@coachpatriciaquintao

/profapatriciaquintao

@plquintao

t.me/coachpatriciaquintao

**NÃO SE ESQUEÇA DE
AVALIAR ESTA AULA!**

**SUA OPINIÃO É MUITO IMPORTANTE
PARA MELHORARMOS AINDA MAIS
NOSSOS MATERIAIS.**

**ESPERAMOS QUE TENHA GOSTADO
DESTA AULA!**

**PARA AVALIAR, BASTA CLICAR EM LER
A AULA E, DEPOIS, EM AVALIAR AULA.**

AVALIAR 