



Licap

Formação Cientista de Dados



PUC Minas

Formação do Cientista de Dados

Entendendo os tipos de variáveis – Módulo Básico

Luis Enrique Zárate

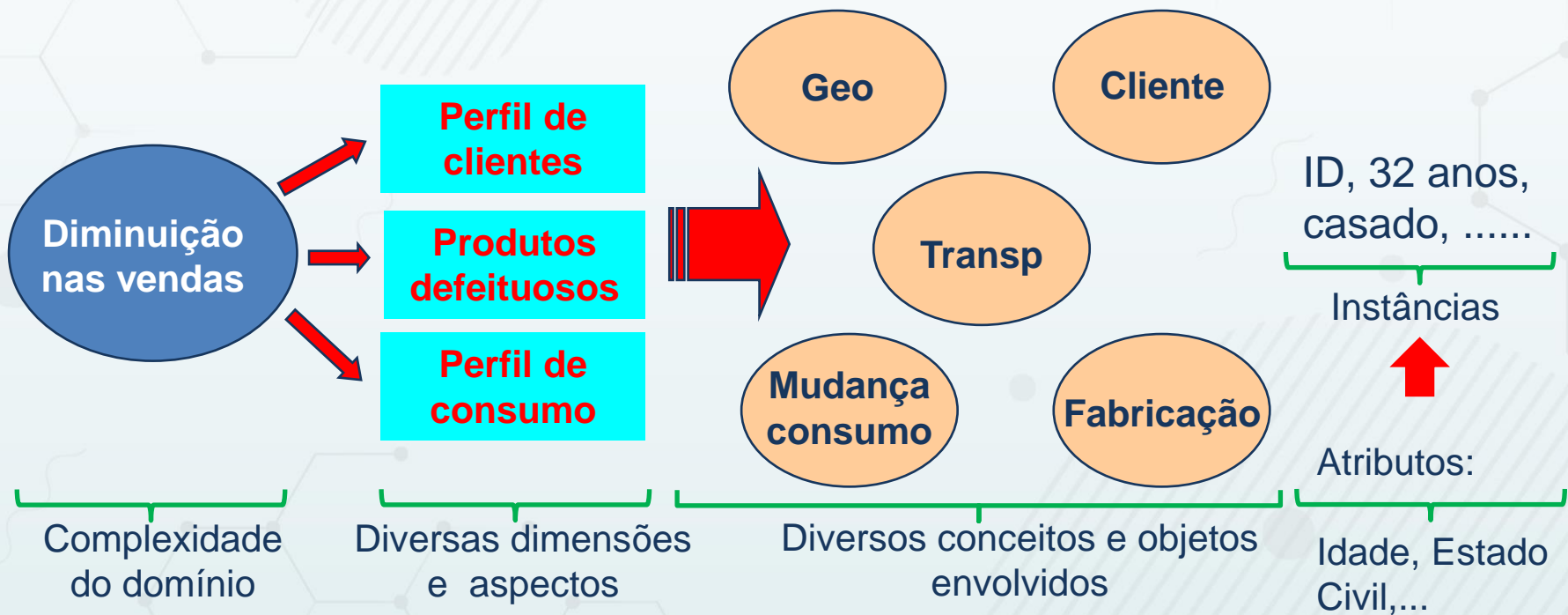
→ Conteúdo do Curso



1. Entendendo a natureza do dados
2. Tipos conceituais de variáveis
3. Variáveis Quantitativas e Qualitativas
4. Variáveis qualitativas Nominais e Ordinais
5. Variáveis quantitativas Intervalar, relacional e relacional nomeada
6. Conversão entre tipos de variáveis

Natureza do dado – conceitos e objetos vinculados

O mundo é complexo e é formado por aspectos, conceitos e objetos, os quais estão vinculados através de suas instâncias individuais e atributos que o caracterizam.



Tipos conceituais de variáveis Licap

Variável Escalar

Valor da medida + Escala da medida

(Exemplo: Velocidade 50 m/s; Idade 30 anos).

Variável Vectorial

Precisa de mais de um valor para ser definida

(Exemplo, Localização [10°latitude, 30°longitude]). Cada valor da variável possui sentido individual e em conjunto.

Tipos de variáveis

Variáveis Quantitativas

Variáveis discretas: Podem assumir apenas um número finito ou infinito contável de valores entre dois valores. Correspondem a dados inteiros.

Exemplos: Número de filhos, Número de vezes que exercita por mês, Número de bactérias no leite, etc.

Variáveis contínuas: Assumem valores infinitos entre dois valores numa escala contínua, onde valores fracionais são relevantes. Correspondem a dados reais.

Exemplo: Peso, Altura, Idade, Tempo, Temperatura, etc.

Tipos de variáveis



Variáveis Qualitativas (ou categóricas) Politómicas

Variáveis nominais: Não existe uma ordem natural entre as categorias.

Exemplo: estado civil, religião, profissão, raça, cor de pele, etc.

Variáveis ordinais: Existe uma ordenação natural entre as categorias.

Exemplo: nível de escolaridade, mês de observação.

Variáveis Qualitativas (ou categóricas) Dicotómicas

Variáveis que podem adotar somente dois valores.

Exemplo: Sexo {M,F}.

→ Variáveis qualitativas nominais



- Os valores dessas variáveis recebem nomes “nominalmente” são impostos por nós.

Exemplo:

Alimentos: café, aceitona, palmito;

Vestimenta: calça, camisa, blusa;

Artigos de Escritório: caneta, papel, lápis, etc.

- Essas medidas não possuem uma ordem inerente podendo ser codificadas de formas arbitrárias como $\{1,2,3...\}$, $\{A, B, C,...\}$ ou pelos próprios nomes atribuídos aos valores.
- Cada medida (nome) possui notórias diferenças na sua identidade, mas essa diferença não pode ser manipulada matematicamente. A sequência dos números é somente uma conveniência, não implicam numa ordem que quantifique a diferença.

→ Variável qualitativa ordinal



“Pense na medição da temperatura do seu café”

- A medição pode estar limitada a: “*quente*” ou “*frio*”
- A medição anterior possui menos informação que: “*fervendo*”, “*muito quente*”, “*quente*”, “*morno*”, “*frio*”, “*congelado*”.
- Dependendo da aplicação pode ser necessário aumentar ou diminuir o detalhe da informação. Menores detalhes proporcionam resultados mais genéricos. Maiores detalhes proporcionam resultados mais específicos.

→ Variável qualitativa ordinal



As variáveis de escala ordinal devem cumprir a condição de *TRANSITIVIDADE*.

Se $(A > B)$ e $(B > C)$ então $(A > C)$

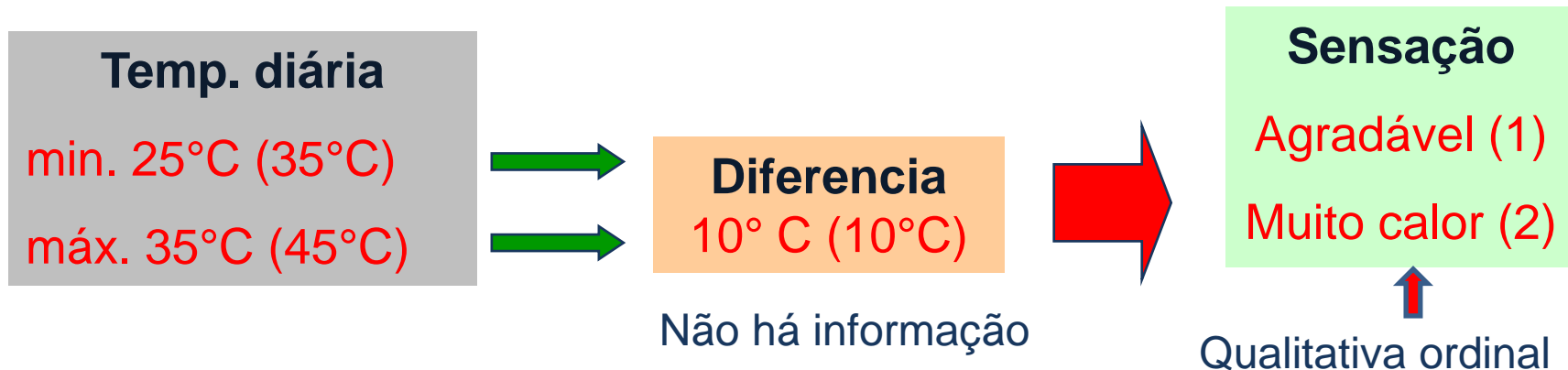
Caso seja necessário criar uma nova codificação para uma variável ordinal é necessário cumprir a condição de *TRANSITIVIDADE*.

→ Por exemplo, para a variável temperatura do café:

“*fervendo*” (1,0), “*muito quente*” (0,8), “*quente*” (0,6), “*morno*” (0,4), “*frio*” (0,2), “*congelado*” (0,0).

→ Variável quantitativa de escala Intervalar

- Existe informação contida na ordem do ranking dos valores medidos.
- A diferença entre valores não traz informação adicional.
- A variável quantitativa que corresponde a uma escala intervalar pode ser transformada em variável qualitativa (categórica) ordinal.



→ Variável quantitativa de escala relacional

- É a escala que leva a maior quantidade de informação.
- A diferença entre valores possui informação e pode ser expressa por meio de uma relação com as propriedades da escala.
- A variável quantitativa que possui uma escala relacional pode ser transformada em variável qualitativa (categórica) ordinal.



Variável quantitativa de escala relacional

- Consideremos o registro do Peso (kg.) de um indivíduo ao longo de 3:

Registro	Atributos				
#	...	Peso_1º mês	Peso_2º mês	Peso_3º mês
1	...	125	100	90	...

Mudança de peso

$[1^{\circ}-2^{\circ}] = - 25 \text{ kg.}$

$[1^{\circ}-3^{\circ}] = - 35 \text{ kg.}$



Redução – opção 1

$[1^{\circ}-2^{\circ}] = 25/125 = 0,20 \text{ (20\%)}$

$[1^{\circ}-3^{\circ}] = 35/125 = 0,28 \text{ (28\%)}$

Redução – opção 2

$[1^{\circ}-3^{\circ}] = 0,28 \text{ (28\%)}$

Registro	Atributos		
#	Redução_em_três_Meses
1	28%

→ Variável quantitativa de escala relacional nomeada

- A escala relacional perde a informação da grandeza original.
- É necessário manter conhecidas as unidades da medida relacional.

Venda atual = 5 *unidades*
Venda passada = 4 *unidades*



Diferencia = 1
Relação = $5/4 = 1,25$



Unidades vendidas = ? milhão de reais = ? milhão de euros = ?

➡ Convertendo variável quantitativa em qualitativa



- Uma variável originariamente quantitativa pode ser coletada, transformada e armazenada de forma qualitativa.
- Essas medidas não possuem uma ordem inerente.

Variável Original

Idade (em anos):
quantitativo discreta



Variável Coletada

Faixa etária:
Adolescente: 13 a 19 anos
Jovem Adulto: 20 a 30 anos
qualitativa ordinal

➡ Convertendo nominal em ordinal



- A escala nominal não possui informação relevante.
- É possível converter uma variável nominal em ordinal e ainda incorporar informação acerca da magnitude da diferença entre valores.

Nome do Vinho (Nominal)	Preferência %
Marca 1 (Vinho Cabernet)	51
Marca 2 (Vinho Merlot)	49
Marca 3 (Vinho Shiraz)	0



Variável ordinal - quantitativa

SE (Cabernet é-pref-a Merlot) **E**

(Merlot é-pref-a Shiraz) **ENTÃO**

(Cabernet é-pref-a Shiraz)

SE (Cabernet não-é-dispon) **ENTÃO** (Merlot é-pref-a Shiraz)

- O principal objetivo da mineração de dados é transformar a informação contida no conjunto de dados de forma que possa ser diretamente utilizada (pelos algoritmos de Aprendizado de máquina) e compreendida.
- O entendimento da informação contida nas variáveis permite melhores previsões do comportamento de alguns aspectos do mundo.
- O entendimento da informação de uma variável usualmente não é perfeita, existindo incertezas, portanto o conhecimento extraído não é perfeito e exige restrições.





Licap

Formação Cientista de Dados

Obrigado!



PUC Minas