

# **Formação Livre em Ciência de Dados**

## **Aula 1 | Módulo Básico I**

João Pedro Passos Pereira

# Roteiro Aula 1

## 1. Parte 1:

- a. Introdução à Ciência de Dados;
- b. Conceitos fundamentais em Ciência de Dados;

## 2. Parte 2:

- a. Introdução à Lógica e à Matemática para Ciência de Dados.

# Parte 1

# (1) Introdução à Ciência de Dados

# Ciência de Dados - Processo

## Dados Brutos

(dados antes de serem selecionados quanto ao aproveitamento)



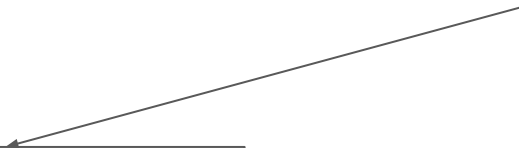
## Dados Líquidos

(dados selecionados com ênfase em aproveitamento)



## Conhecimento

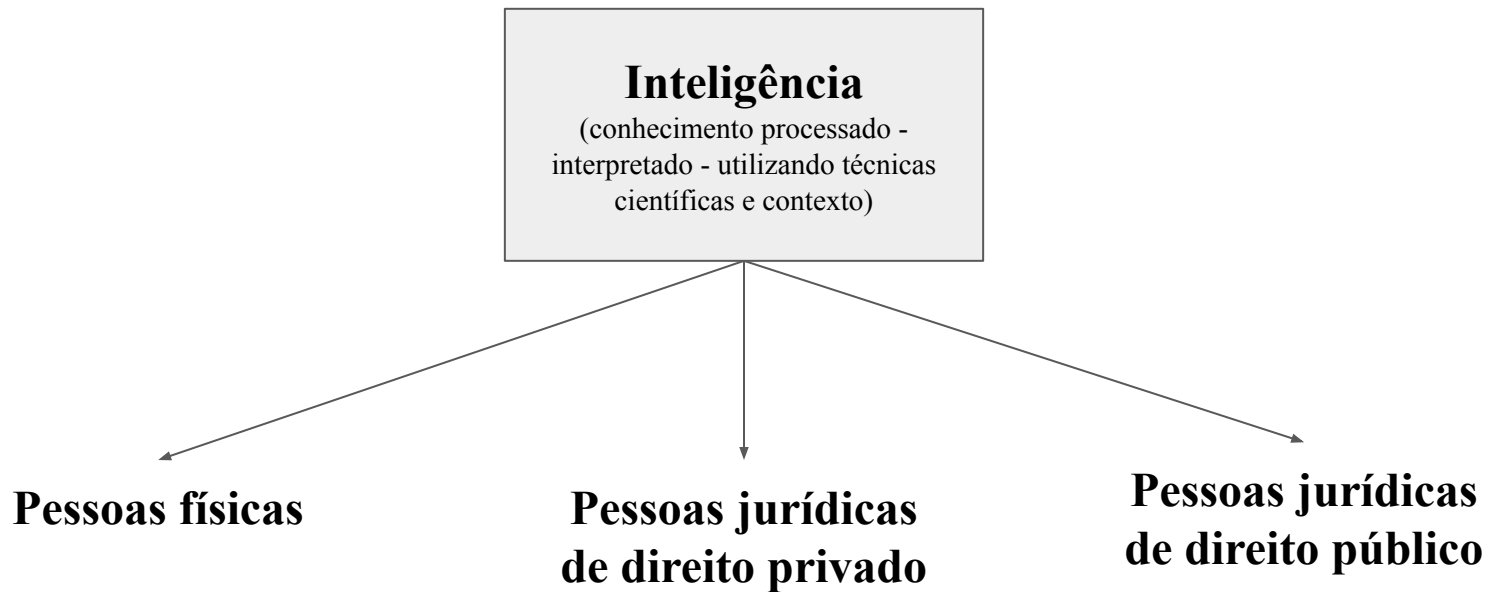
(dados processados por técnicas computacionais e estatísticas)



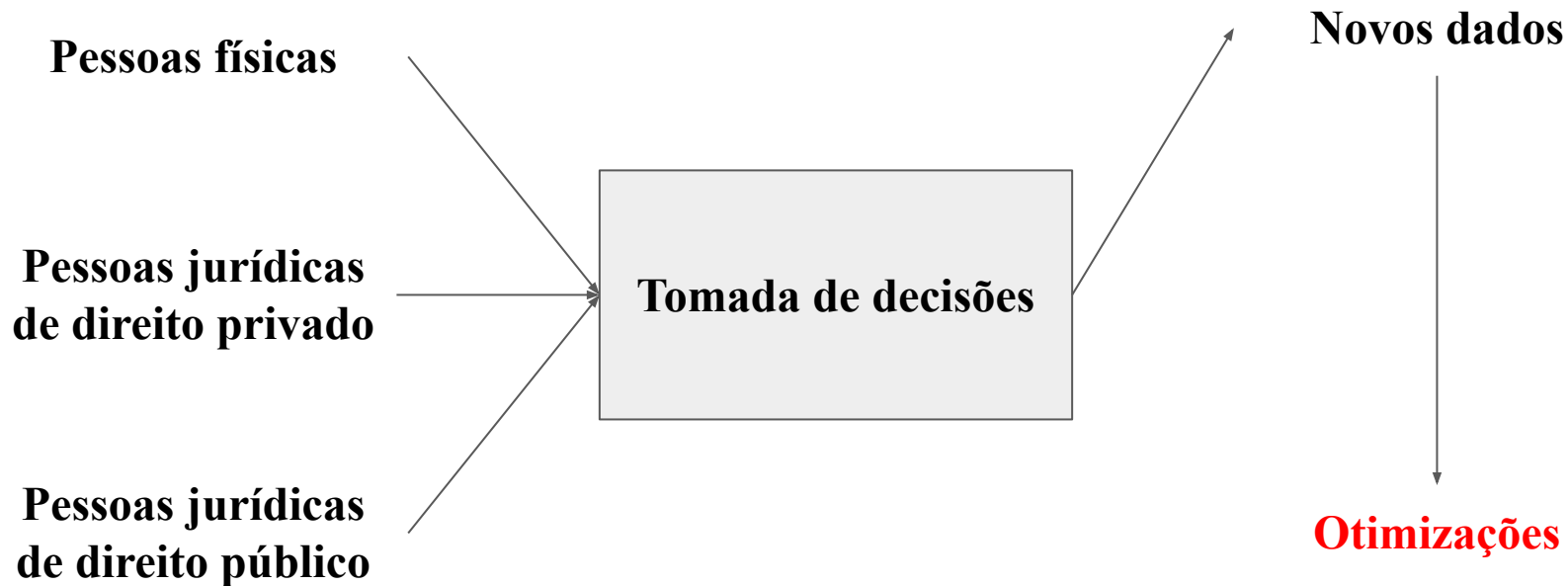
## Inteligência

(conhecimento processado - interpretado - utilizando técnicas científicas e contexto)

# Ciência de Dados - Processo



# Ciência de Dados - Processo



# Ciência de Dados - Competências

**Cientista de dados**

```
graph TD; A[Cientista de dados] --> B[Programação e Exploração com tecnologia]; A --> C[Matemática e Estatística]; A --> D[Raciocínio lógico, Curiosidade, Habilidades diversas];
```

- ❑ **Programação**
- ❑ **Manipulação e Exploração com tecnologia**

- ❑ **Matemática**
- ❑ **Estatística**

- ❑ **Raciocínio lógico**
- ❑ **Curiosidade**
- ❑ **Habilidades diversas**



# Ciência de Dados - Competências

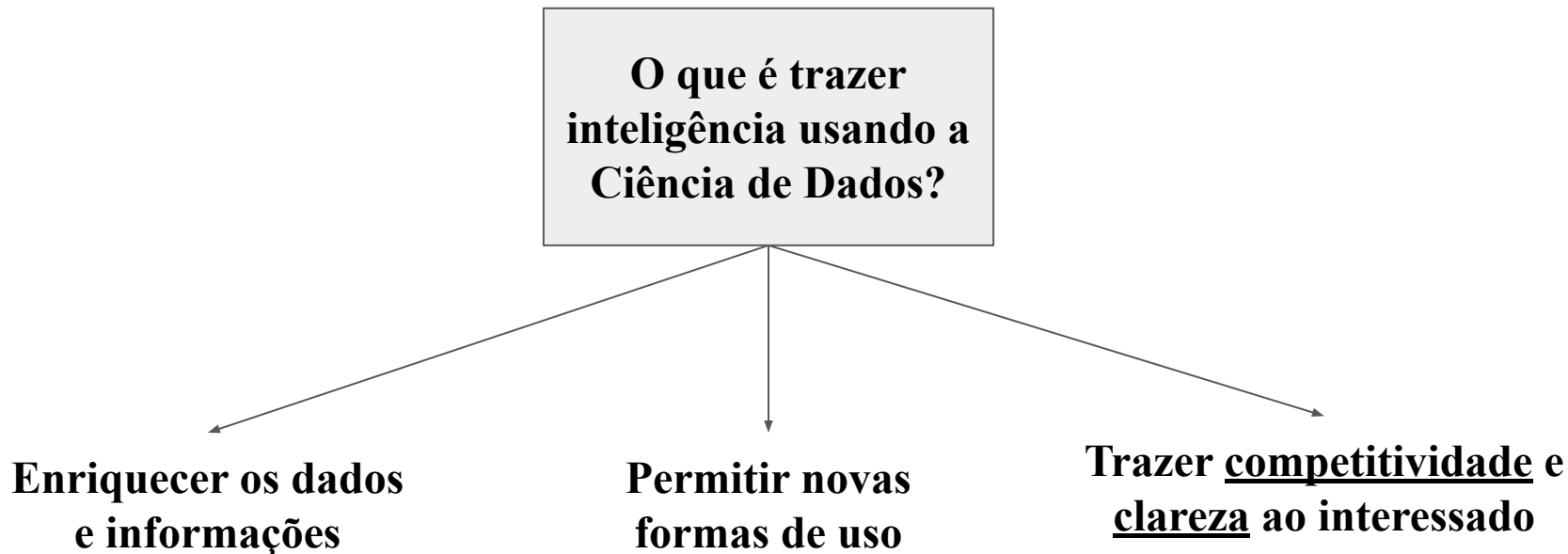


“Cientista de dados é a pessoa que é melhor em estatística do que qualquer engenheiro de software e melhor em engenharia de software do que qualquer estatístico.”

**Josh Wills**

(ex Google, ex Slack, ex Indeed, ex IBM)

# Ciência de Dados - Inteligência



# Ciência de Dados

## Conceito

“Data Science é o termo mais usado para descrever o processo de transformação de dados em conhecimento. Difere-se (pois é mais abrangente) de campos já conhecidos como estatística, análise, busca em bases de dados. É, na verdade, o processo de **exploração e transformação** dos dados em **conhecimento e inteligência** de vários níveis para **organizações e pessoas**.”

# (2) Conceitos fundamentais em Ciência de Dados

# Nomenclaturas iniciais (1)

## Ciência

“Corpo de conhecimentos sistematizados adquiridos via observação, identificação, pesquisa e explicação de determinadas categorias de fenômenos e fatos, e formulados **metódica** e **racionalmente**.”

## Nomenclaturas iniciais (2)

### Pesquisa

“Investigação ou indagação minuciosa.”

É o produto do corpo de conhecimentos científicos, o produto do interesse em conhecer minuciosamente algo.

# Nomenclaturas iniciais (3)

Computadores conectados em rede para aumento do poder computacional e, portanto, de processamento.

Estatística	Matemática	Técnicas computacionais	Lógica
“Ramo da matemática que trata da <u>coleta</u> , da <u>análise</u> , da <u>interpretação</u> e da apresentação de <u>massas de dados numéricos</u> .”	“Ciência que estuda, por método dedutivo, objetos abstratos (números, figuras, <b>funções</b> ) e as relações existentes entre eles.”	Conjunto de conhecimentos em computação e programação de computadores, <b>ou redes de computadores</b> , com a finalidade de resolução de problemas do mundo real.	“Parte da filosofia que trata das formas do pensamento em geral (dedução, indução, hipótese, inferência etc.) e das operações intelectuais que visam à <u>determinação do que é verdadeiro ou não</u> .”



# Nomenclaturas iniciais (4)

<b>Dados Brutos</b>	<b>Dados Líquidos (informação bruta)</b>	<b>Conhecimento (informação líquida)</b>	<b>Inteligência</b>
Dado antes de ser separado quanto à sua relevância.	Dado que ao ser separado quanto à sua relevância, <u>torna-se assim uma informação não processada</u> .	Informação que passou por um pré processamento, extraindo insights iniciais por meio de estatísticas descritivas como exemplo.	Informação transformada em auxiliadora para tomada de decisões, através de método científico, com técnicas estatísticas, computacionais e de contexto.

# Parte 2

# (3) Introdução à Lógica e à Matemática para Ciência de Dados

# Introdução à Lógica - Pressupostos básicos

- Três Princípios (entre outros) regem a Lógica Clássica (mais usada):
  - IDENTIDADE: “tudo é idêntico a si mesmo”;
  - NÃO CONTRADIÇÃO: “duas afirmações contraditórias não podem ser verdadeiras ao mesmo tempo”;
  - TERCEIRO EXCLUÍDO: “para qualquer proposição, ou esta proposição é verdadeira, ou sua negação é verdadeira”:
    - Ex: lógica binária - 0 e 1.

# Introdução à Lógica - Tabela Verdade

## Decorando a Tabela V



$p$	$q$	$\sim p$	$p \vee q$	$p \wedge q$
V	V	F	V	V
V	F	F	V	F
F	V	V	V	F
F	F	V	F	F

Conectivo	Símbolo	Operação Lógica	Valor Lógico
não	$\sim$	negação	Terá valor falso quando a proposição for verdadeira e vice-versa.
e	$\wedge$	conjunção	Será verdadeira somente quando todas as proposições forem verdadeiras.
ou	$\vee$	disjunção	Será verdadeira quando pelo menos uma das proposições for verdadeira.

# Exemplo prático - Uso da tabela verdade

Sendo verdadeiras as proposições:

“Se Marcelo é corinthiano, então André é flamenguista”

“André é vascaíno e Lucca é botafoguense”

“Marcelo é palmeirense ou Lucca é gremista”

A proposição “Lucca é gremista se, e somente se, André é flamenguista” é verdadeira?

Resposta: VERDADE

Conectivo	Símbolo	Operação Lógica	Valor Lógico
não	$\sim$	negação	Terá valor falso quando a proposição for verdadeira e vice-versa.
e	$\wedge$	conjunção	Será verdadeira somente quando todas as proposições forem verdadeiras.
ou	$\vee$	disjunção	Será verdadeira quando pelo menos uma das proposições for verdadeira.

## Decorando a Tabela V



$p$	$q$	$\sim p$	$p \vee q$	$p \wedge q$
V	V	F	V	V
V	F	F	V	F
F	V	V	V	F
F	F	V	F	F

## Exemplo prático - Uso da tabela verdade

A proposição "Lucca é gremista se, e somente se, André é flamenguista" é verdadeira, pois ela está dizendo que se André é flamenguista, então Lucca é gremista e se André não é flamenguista, então Lucca também não é gremista. Como as proposições dadas anteriormente afirmam que André é vascaíno e não flamenguista, então a proposição "Lucca é gremista se, e somente se, André é flamenguista" é verdadeira.

# Introdução à Lógica - Na programação (estruturas)

```
In [1]: a = 10  
b = 20  
if a == b:  
    print('yes')  
else:  
    print('no')
```

O que acontece aqui?

no



# Introdução à Lógica - Na programação (estruturas)

```
1 temperature = 30
2 if temperature > 50:
3     print("Alarm On")
4 else:
5     print("Alarm OFF")
```

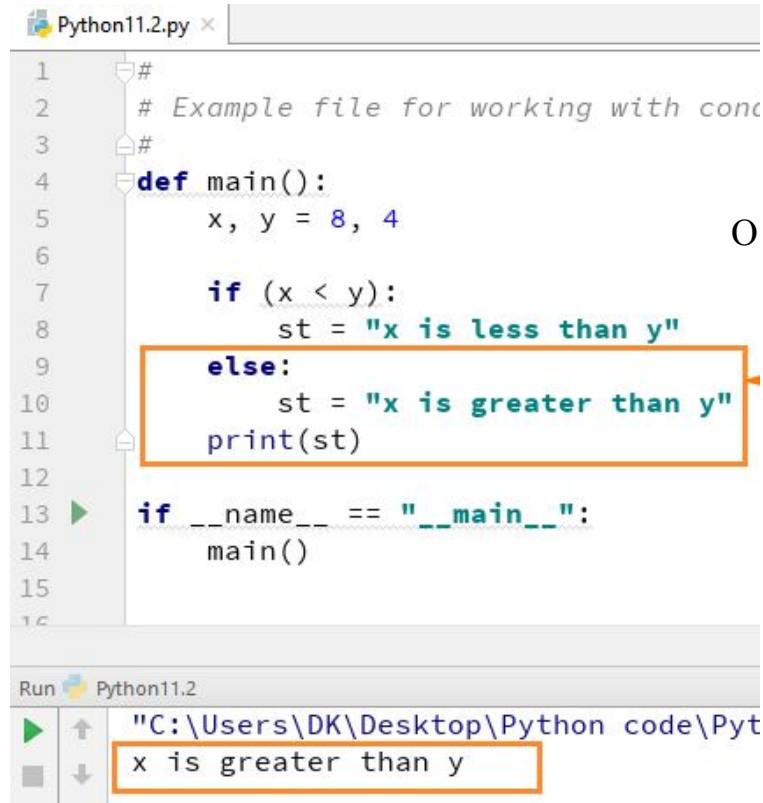
else

Run: hello x

/home/ahmed/PycharmProjects/fui  
Alarm OFF

O que acontece aqui?

# Introdução à Lógica - Na programação (estruturas)



```
Python11.2.py x
1 #
2 # Example file for working with conc
3 #
4 def main():
5     x, y = 8, 4
6
7     if (x < y):
8         st = "x is less than y"
9     else:
10         st = "x is greater than y"
11         print(st)
12
13 if __name__ == "__main__":
14     main()
15
16
```

Run Python11.2

"C:\Users\DK\Desktop\Python code\Pyt  
x is greater than y

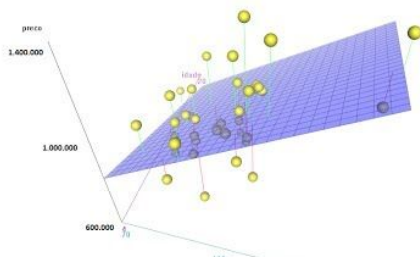
O que acontece aqui?

# Como a matemática será trabalhada na Formação?

- Em duas grandes etapas:
  - Apresentação abrangente;
  - Cálculo:
    - Pré Cálculo;
    - Cálculo em si.

# Introdução à Matemática - O que é preciso ter em mente?

## Álgebra Linear



$$y_1 = \beta_o + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_q x_{1q} + \varepsilon_1$$

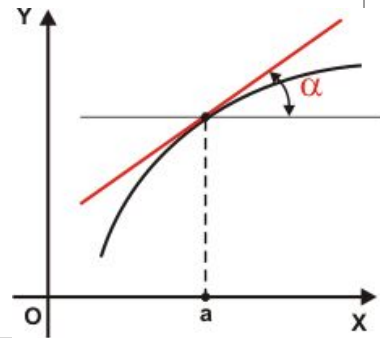
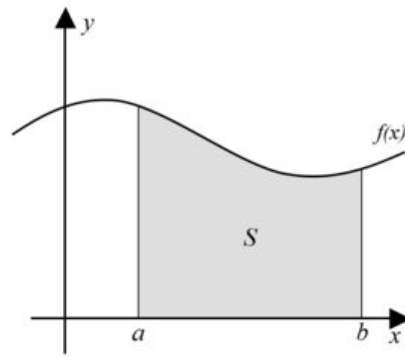
$$y_2 = \beta_o + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_q x_{2q} + \varepsilon_2$$

$\vdots$

$$y_n = \beta_o + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_q x_{nq} + \varepsilon_n$$

## Cálculo diferencial e integral (uma ou mais variáveis)

Aplicável na formulação de modelos e testagem de hipóteses e de significância estatística e matemática.

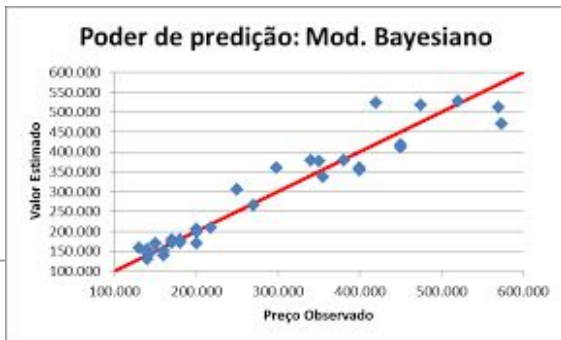
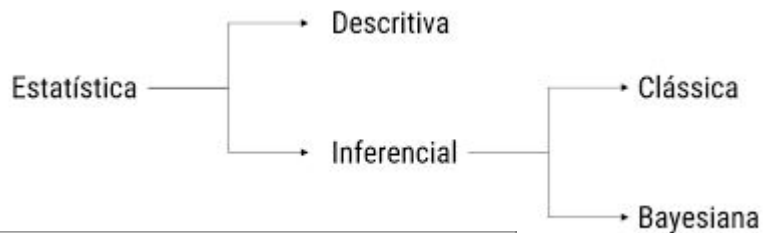


\*Não se preocupe com os termos técnicos, que serão trabalhados em etapas posteriores, apenas tenha em mente a importância de tais ideias

# Introdução à Matemática - O que é preciso ter em mente?

\*Não se preocupe com os termos técnicos, que serão trabalhados em etapas posteriores, apenas tenha em mente a importância de tais ideias

## Estatística



## Probabilidade

Número de casos favoráveis: 28  
Número de casos possíveis: 56

$$p(A) = \frac{\text{Número de casos favoráveis (evento A ocorre)}}{\text{Número total de casos possíveis}}$$

$$p(A) = \frac{28}{56} = \frac{1}{2} = 0,5 = 50\%$$

Esse é o valor da probabilidade clássica calculada pela mãe de José.

nova escola

Probabilidade a posteriori

Probabilidade a priori

Verossimilhança

$$P(A|B) = \frac{P(A) * P(B|A)}{P(B)}$$

Marginalização

Material por: João Pedro Passos Pereira

# Contatos

Meus links úteis

**Instagram:** @j0pewd2

**Site:** <https://joaopedropereira.com.br>

**E-mail:** [contato@joaopedropereira.com.br](mailto:contato@joaopedropereira.com.br)

**Linkedin:** <https://linkedin.com/in/joaopedrowd/>