

Introdução ao python com data mining e chatbot

Seminário



Rafael
Menezes

Oceanógrafo
Mestre em Biotecnologia Marinha



QUEM SOU



- Prof. Rafael Menezes
 - Carioca, educador, pesquisador, consultor, gestor de hospedagem, pythoner, botafoguense, surfista, jiu-jiteiro, netfixer e gamer.
 - Formação:
 - Graduado em Oceanografia (USP);
 - Mestre em Biotecnologia Marinha (IEAPM/UFF).
 - Linhas de pesquisa:
 - Processos costeiros;
 - Manejo e Conservação;
 - Ciência de Dados;
 - Desenvolvimento de ferramentas para o ensino de ciências exatas e do mar.



RESUMO



idaco

INSTITUTO DE DESENVOLVIMENTO E AÇÃO COMUNITÁRIA

SEMINÁRIO PYTHON



Introdução ao PYTHON



Introdução à Mineração de
Dados



Introdução ao uso de chatbots
com fins educacionais

AULA 1

O QUE É PROGRAMAÇÃO?

**É o processo de escrita, testes e manutenção de
programas de computadores.**

INFORMAÇÃO CODIFICADA



idaco
INSTITUTO DE DESENVOLVIMENTO E AÇÃO COMUNITÁRIA

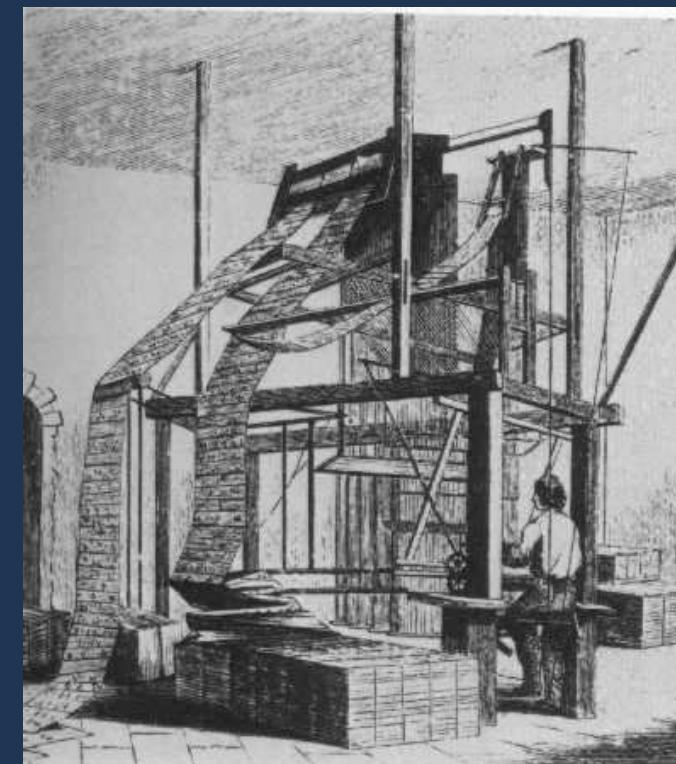
AULA 1

O QUE É PROGRAMAÇÃO?

É o processo de escrita, testes e manutenção de programas de computadores.

INFORMAÇÃO CODIFICADA

- Códigos são mais antigos do que computadores
- 1801
 - Tear de Jacquard



idaco

INSTITUTO DE DESENVOLVIMENTO E AÇÃO COMUNITÁRIA

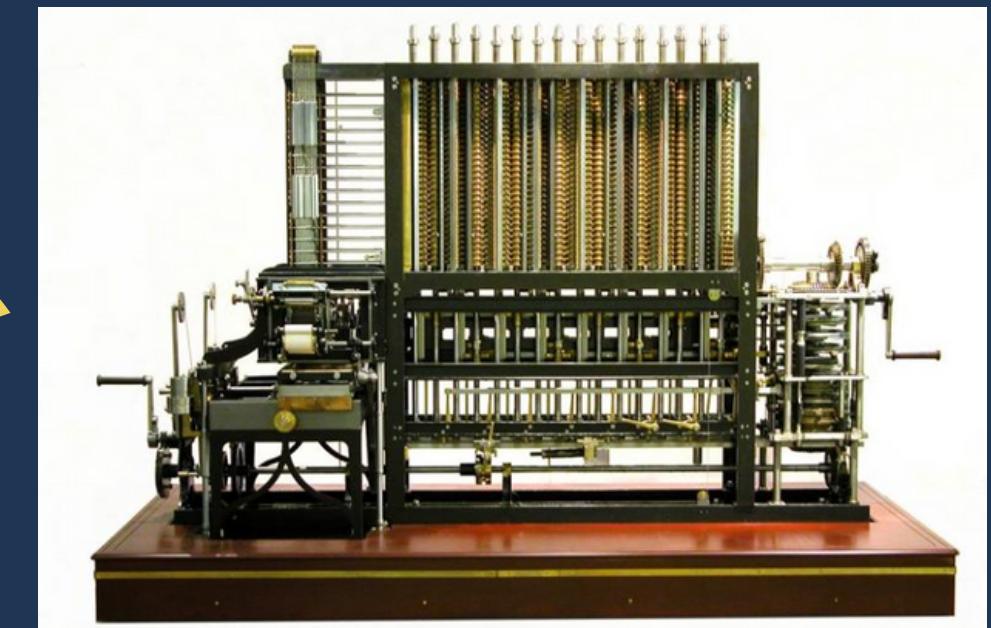
AULA 1

O QUE É PROGRAMAÇÃO?

É o processo de escrita, testes e manutenção de programas de computadores.

INFORMAÇÃO CODIFICADA

- Códigos são mais antigos do que computadores
- 1801
 - Tear de Jacquard
- 1842
 - Ada Lovelace 
 - Máquina analítica de Charles Babbage



AULA 1

O QUE É PROGRAMAÇÃO?

É o processo de escrita, testes e manutenção de programas de computadores.

INFORMAÇÃO CODIFICADA

- Códigos são mais antigos do que computadores
- 1801
 - Tear de Jacquard
- 1842
 - Ada Lovelace 
 - Máquina analítica de Charles Babbage
- 1954
 - FORmula TRANslator (FORTRAN)



```
PROGRAM TRIVIAL
INTEGER I
I=2
IF(I .GE. 2) CALL PRINTIT
STOP
END
SUBROUTINE PRINTIT
PRINT *, 'Hola Mundo'
RETURN
END
```



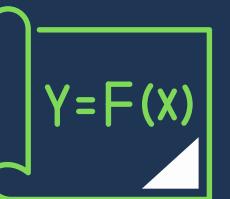
F

AULA 1

COMO FUNCIONA UM CÓDIGO?

São comandos com funções, condições e repetições.

FUNÇÕES



AULA 1

COMO FUNCIONA UM CÓDIGO?

São comandos com funções, condições e repetições.



FUNÇÕES

$$Y = F(X)$$



$$Y = 2 * X$$

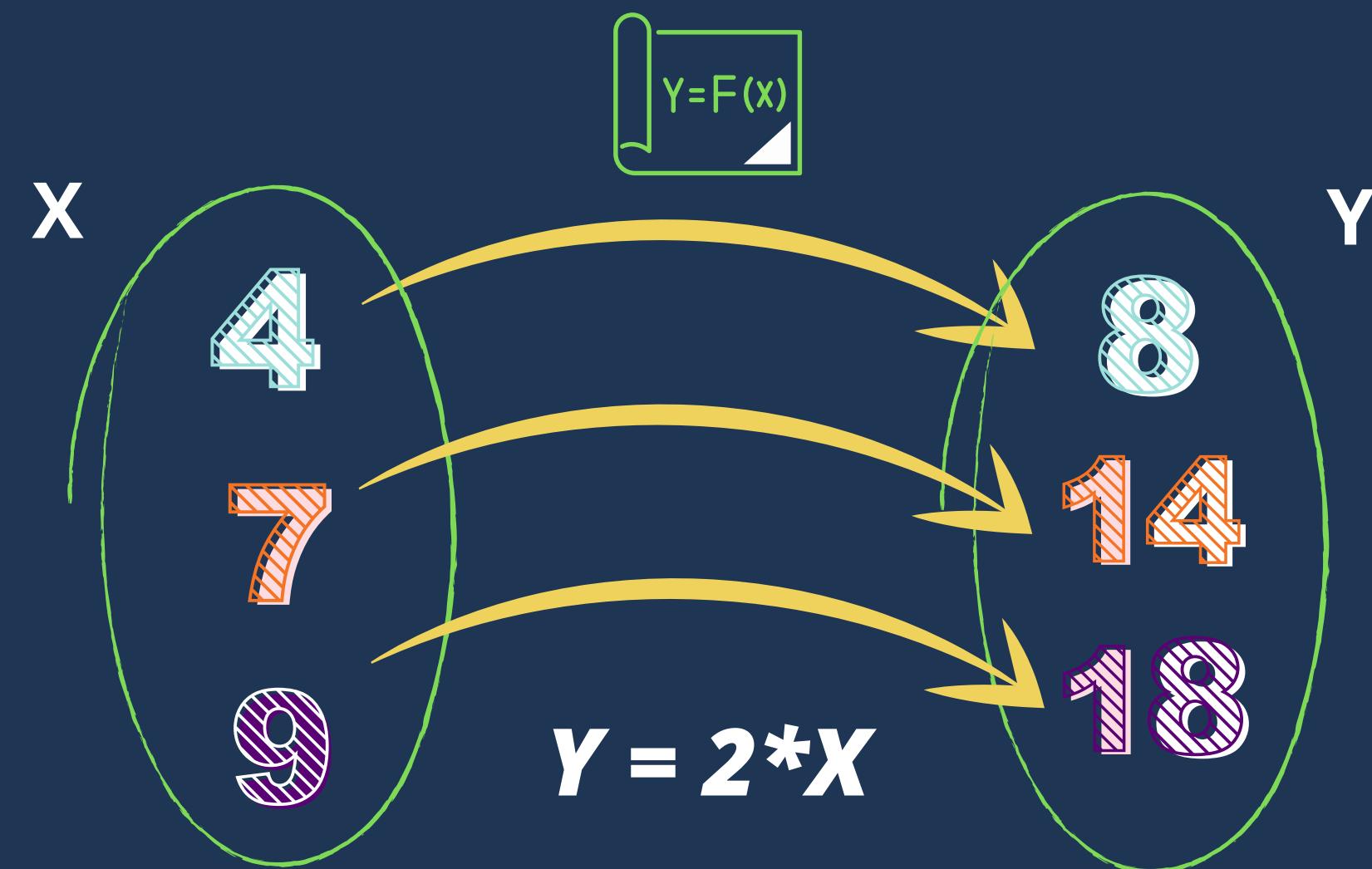
AULA 1

COMO FUNCIONA UM CÓDIGO?

São comandos com funções, condições e repetições.



FUNÇÕES



AULA 1

COMO FUNCIONA UM CÓDIGO?

São comandos com funções, condições e repetições.



AULA 1

COMO FUNCIONA UM CÓDIGO?

São comandos com funções, condições e repetições.



AULA 1

O QUE É O PYTHON?

É uma linguagem de programação livre de uso geral



AULA 1

O QUE É O PYTHON?

É uma linguagem de programação livre de uso geral



idaco

INSTITUTO DE DESENVOLVIMENTO E AÇÃO COMUNITÁRIA



- 80's-91 
- Guido Van Rossum

AULA 1

O QUE É O PYTHON? 

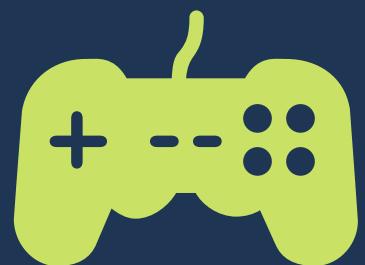
É uma linguagem de programação livre de uso geral



idaco
INSTITUTO DE DESENVOLVIMENTO E AÇÃO COMUNITÁRIA



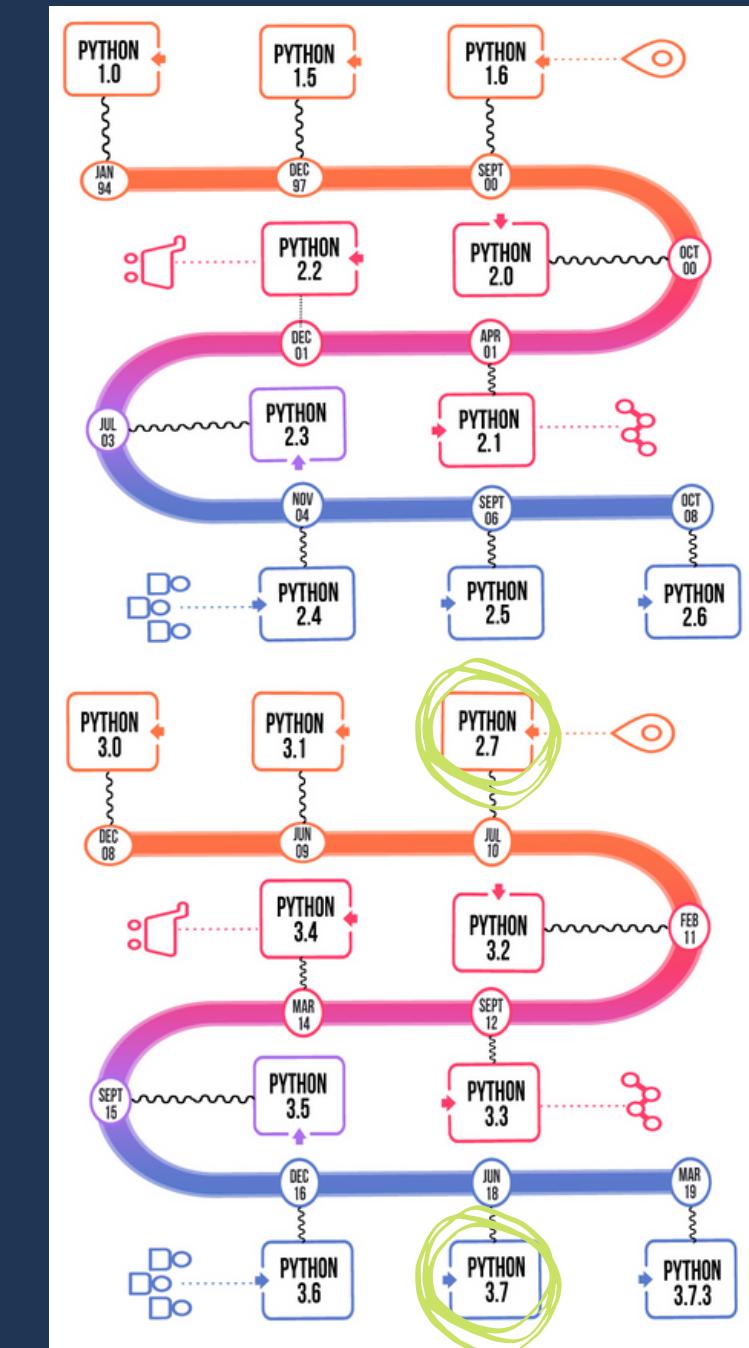
- 80's-91
- Guido Van Rossum



AULA 1

O QUE É O PYTHON?

É uma linguagem de programação livre de uso geral



AULA 1

QUEM USA O PYTHON?



"Instagram currently features the world's largest deployment of the Django web framework, which is written entirely in Python."



"It all got started, I believe, because the very earliest Googlers (Sergey, Larry, Craig, ...) made a good engineering decision: "Python where we can, C++ where we must.""



"Developers at Netflix have the freedom to choose the technologies best suited for the job. More and more, developers turn to Python due to its rich batteries-included standard library, succinct and clean yet expressive syntax, large developer community, and the wealth of third party libraries one can tap into to solve a given problem."



"Spotify's backend consists of many interdependent services, connected by [its] own messaging protocol over ZeroMQ. Around 80% of these services are written in Python."

AULA 1

QUEM USA O PYTHON?



"Instagram currently features the world's largest deployment of the Django web framework, which is written entirely in Python."



"It all got started, I believe, because the very earliest Googlers (Sergey, Larry, Craig, ...) made a good engineering decision: "Python where we can, C++ where we must.""

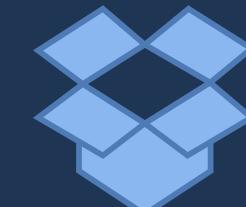


"Developers at Netflix have the freedom to choose the technologies best suited for the job. More and more, developers turn to Python due to its rich batteries-included standard library, succinct and clean yet expressive syntax, large developer community, and the wealth of third party libraries one can tap into to solve a given problem."



"Spotify's backend consists of many interdependent services, connected by [its] own messaging protocol over ZeroMQ. Around 80% of these services are written in Python."

UBER



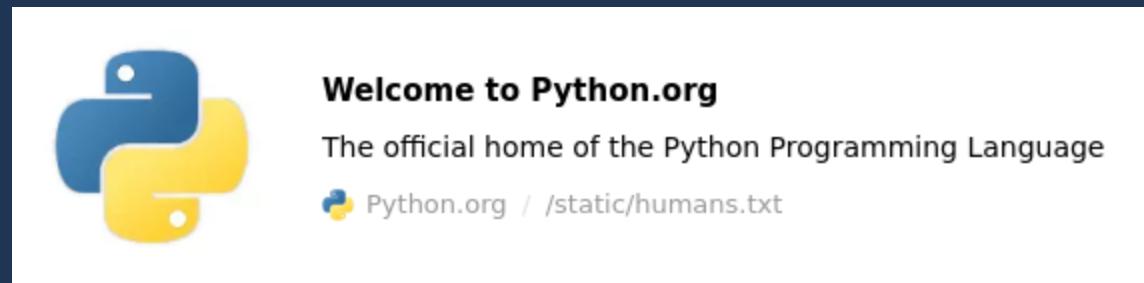
AULA 1

VATANGENS DO PYTHON?

Uso livre.

Comunidade ampla.

Fácil leitura (identação).



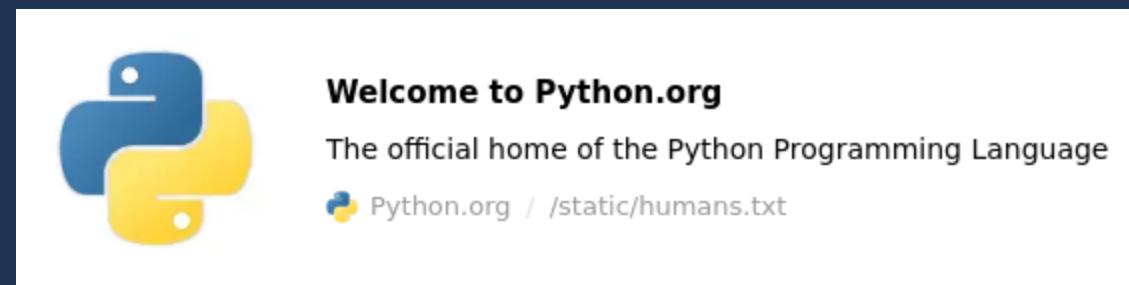
AULA 1

VATANGENS DO PYTHON?

Uso livre.

Comunidade ampla.

Fácil leitura (identação).



- **Repositórios**

- **GitHub** 

- **Conda cloud** 

- **Videoaulas**

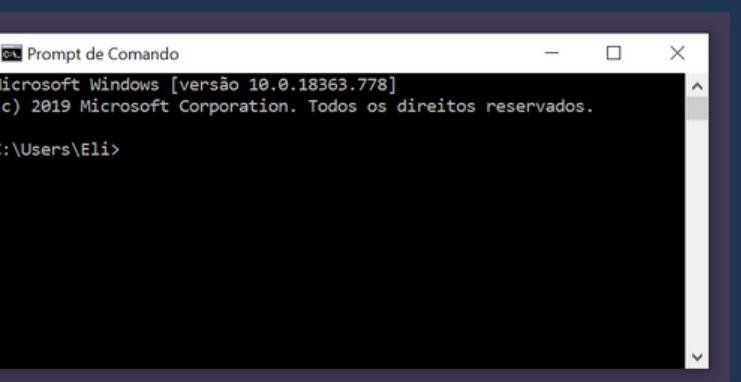
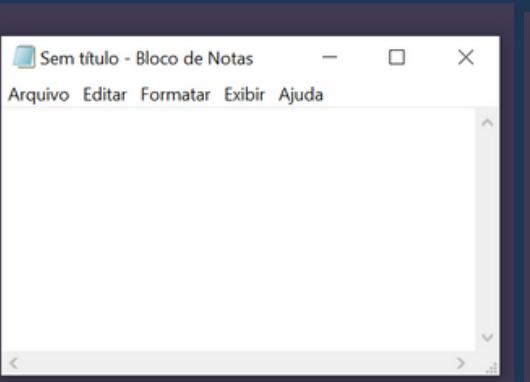
- **Youtube** 
- **Udemy** 
- **DataCamp** 



AULA 1

ONDE PROGRAMAR?

Editores de texto



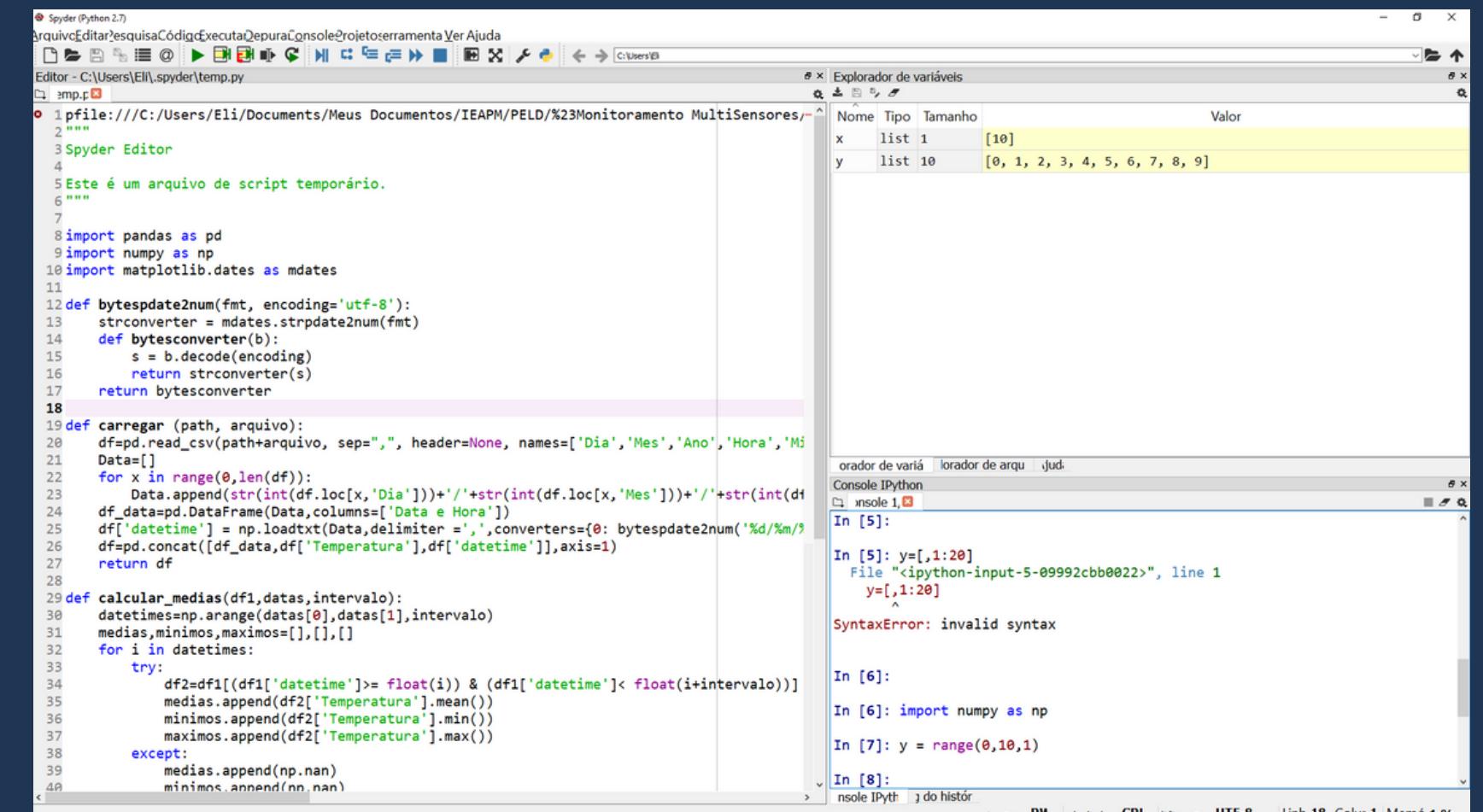
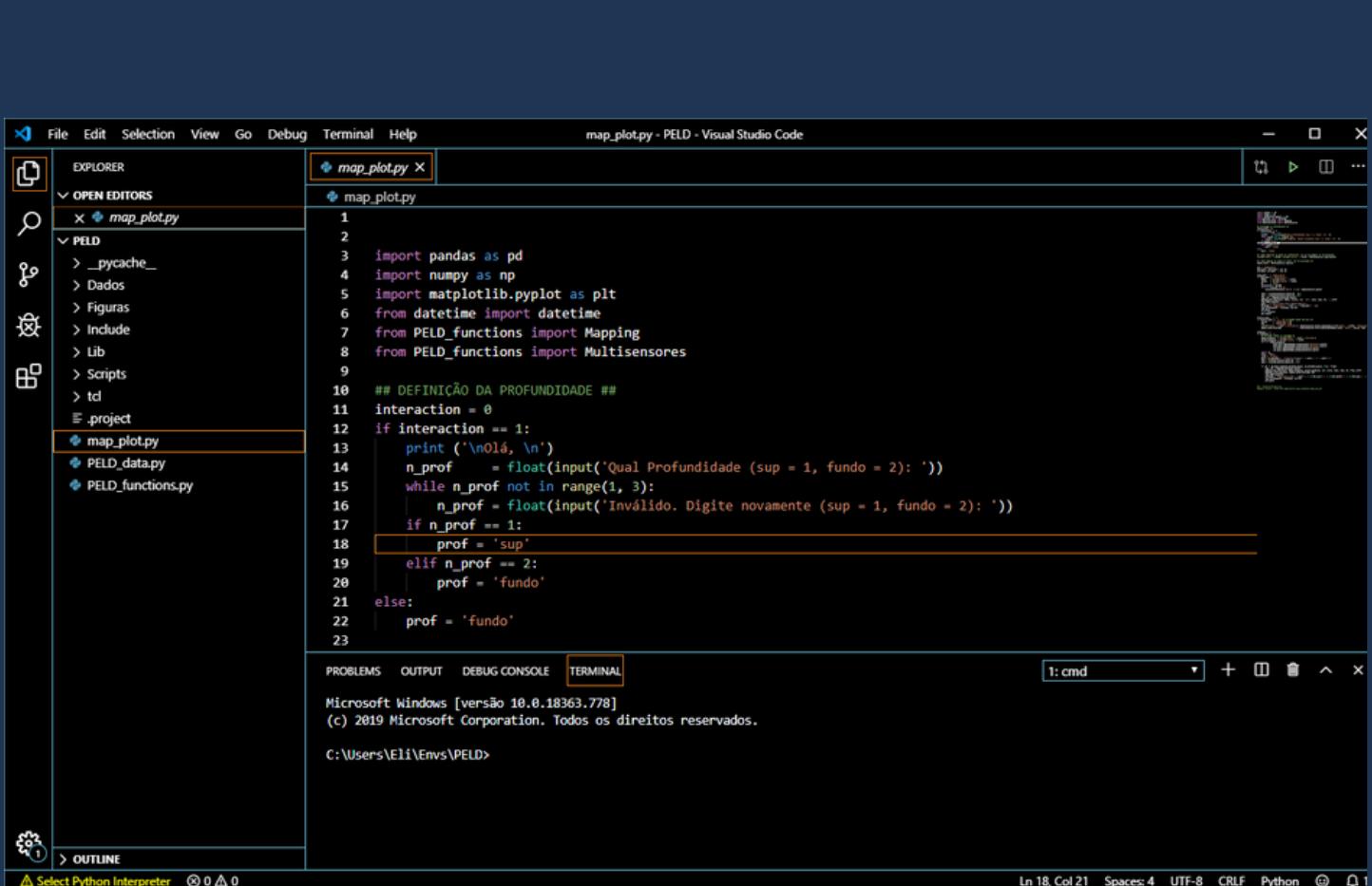
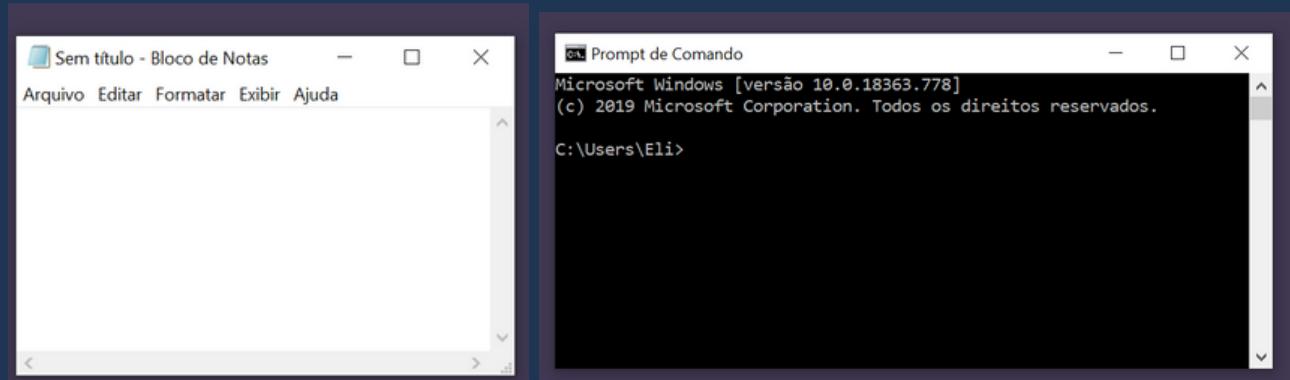
AULA 1

ONDE PROGRAMAR?



Editores de texto

IDE - Integrated Development Environment



AULA 1

ONDE PROGRAMAR?

Editores de texto

IDE - Integrated Development Environments

API - Application Programming Interfaces



back-end x front-end

C:\

APP

The screenshot shows a Visual Studio Code interface with the following details:

- File Explorer:** Shows the project structure with files: `map_plot.py`, `PELD_data.py`, and `PELD_functions.py`.
- Code Editor:** Displays the content of `map_plot.py`. The code imports pandas, numpy, and matplotlib.pyplot, and defines a function to calculate depth based on user input (sup or fundo).
- Terminal:** Shows the command line output from running the script in a Windows cmd window. It includes the Windows version, copyright notice, and the path `C:\Users\Eli\Env\PELD>`.

```
1
2
3 import pandas as pd
4 import numpy as np
5 import matplotlib.pyplot as plt
6 from datetime import datetime
7 from PELD_functions import Mapping
8 from PELD_functions import Multisensores
9
10 ## DEFINIÇÃO DA PROFUNDIDADE ##
11 interaction = 0
12 if interaction == 1:
13     print ('\nOlá, \n')
14     n_prof = float(input('Qual Profundidade (sup = 1, fundo = 2): '))
15     while n_prof not in range(1, 3):
16         n_prof = float(input('Inválido. Digite novamente (sup = 1, fundo = 2): '))
17     if n_prof == 1:
18         prof = 'sup'
19     elif n_prof == 2:
20         prof = 'fundo'
21 else:
22     prof = 'fundo'
23
```

The screenshot shows the Spyder Python IDE interface. The top menu bar includes 'Arquivo', 'Editor', 'esquise', 'Código', 'Executa', 'Console', 'Projeto', 'erramenta', 'Ver', and 'Ajuda'. The toolbar contains icons for file operations like Open, Save, Run, and Stop. The left pane is the code editor with the file 'amp.py' open, containing Python code for reading CSV files and calculating averages. The right pane is the 'Explorador de variáveis' (Variables Explorer) showing two lists: 'x' (list of 1 element [10]) and 'y' (list of 10 elements [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]). Below the code editor is a 'Console IPython' window with several input and output cells. Cell 5 shows an attempt to slice a list with invalid syntax: 'In [5]: y=[,1:20]'. Cell 6 shows the resulting SyntaxError: 'SyntaxError: invalid syntax'. Cells 7 and 8 show imports and a range function.

```
Spyder (Python 2.7)
Arquivo Editar esquise Código Executa Console Projeto erramenta Ver Ajuda
Editor - C:\Users\Eli\spyder\temp.py
amp.py

1 pfile:///C:/Users/Eli/Documents/Meus Documentos/IEAPM/PELD/%23Monitoramento MultiSensores/
2 """
3 Spyder Editor
4
5 Este é um arquivo de script temporário.
6 """
7
8 import pandas as pd
9 import numpy as np
10 import matplotlib.dates as mdates
11
12 def bytespdate2num(fmt, encoding='utf-8'):
13     strconverter = mdates.strpdate2num(fmt)
14     def bytesconverter(b):
15         s = b.decode(encoding)
16         return strconverter(s)
17     return bytesconverter
18
19 def carregar(path, arquivo):
20     df=pd.read_csv(path+arquivo, sep=",", header=None, names=['Dia','Mes','Ano','Hora','Mi'])
21     Data=[]
22     for x in range(0,len(df)):
23         Data.append(str(int(df.loc[x,'Dia']))+'/'+str(int(df.loc[x,'Mes']))+'/'+str(int(df.loc[x,'Ano'])))
24     df_data=pd.DataFrame(Data,columns=['Data e Hora'])
25     df['datetime'] = np.loadtxt(Data,delimiter=',',converters={0: bytespdate2num('%d/%m/%Y')})
26     df=pd.concat([df_data,df['Temperatura'],df['datetime']],axis=1)
27     return df
28
29 def calcular_medias(df1,datas,intervalo):
30     datetimes=np.arange(datas[0],datas[1],intervalo)
31     medias,minimos,maximos,[],[]
32     for i in datetimes:
33         try:
34             df2=df1[(df1['datetime']>= float(i)) & (df1['datetime']< float(i+intervalo))]
35             medias.append(df2['Temperatura'].mean())
36             minimos.append(df2['Temperatura'].min())
37             maximos.append(df2['Temperatura'].max())
38         except:
39             medias.append(np.nan)
40             minimos.append(np.nan)

Explorador de variáveis
Nome Tipo Tamanho Valor
x list 1 [10]
y list 10 [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]

In [5]:
In [5]: y=[,1:20]
File "<ipython-input-5-09992cbb0022>", line 1
      y=[,1:20]
      ^
SyntaxError: invalid syntax

In [6]:
In [6]: import numpy as np

In [7]: y = range(0,10,1)

In [8]:
```

AULA 1

ONDE PROGRAMAR?



<https://www.anaconda.com/distribution/#download-section>

- ✓ ○ Atenção ao sistema operacional
- Add to PATH



idaco
INSTITUTO DE DESENVOLVIMENTO E AÇÃO COMUNITÁRIA

AULA 1

ONDE PROGRAMAR?



<https://www.anaconda.com/distribution/#download-section>

- ✓ ○ Atenção ao sistema operacional
- Add to PATH



S1 junho/2020

CURSO ONLINE

**Introdução
à computação científica
com  python**

Rafael G. de Menezes

Clube do Cientista

SUMÁRIO

- O que é este curso?
- Quem é você no mundo da programação?
- [Introdução à programação](#)
- Conceitos e comandos básicos
- Processando dados
- Visualizando dados

AULA 1

COMO COMEÇAR?

Python.

Hello World.

https://github.com/clubedocientista/IdacoNave_intro_Python/



VAMOS PROGRAMAR?

RESUMO



idaco

INSTITUTO DE DESENVOLVIMENTO E AÇÃO COMUNITÁRIA

SEMINÁRIO PYTHON



Introdução ao PYTHON



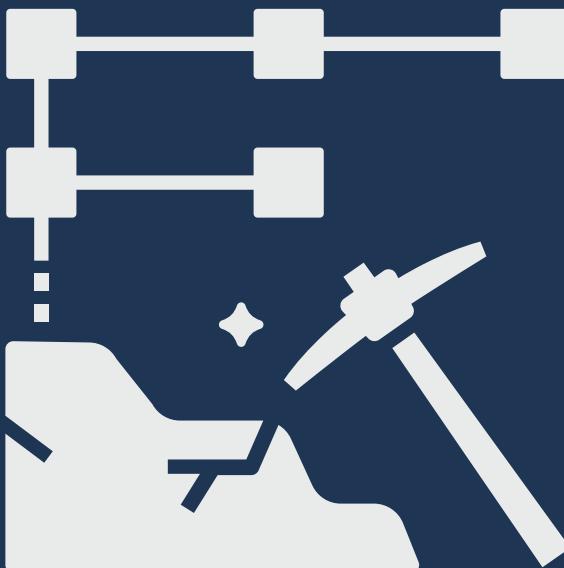
Introdução ao uso de chatbots
com fins educacionais

AULA 2

O QUE É DATA MINING?

Mineração de Dados.

É a extração de padrões consistentes de dados.

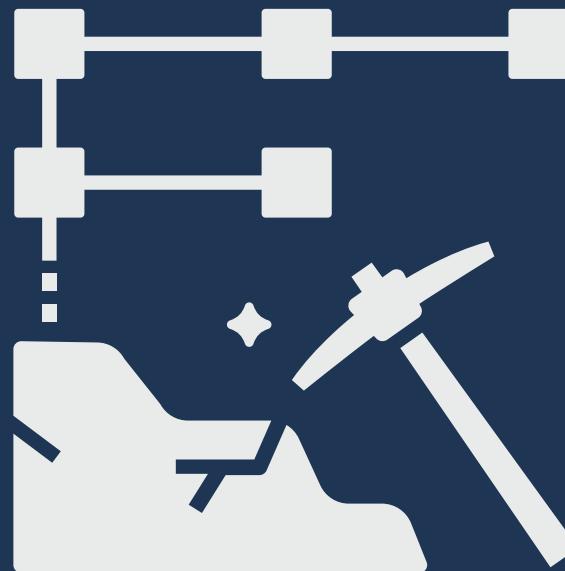


AULA 2

O QUE É DATA MINING?

Mineração de Dados.

É a extração de padrões consistentes de dados.

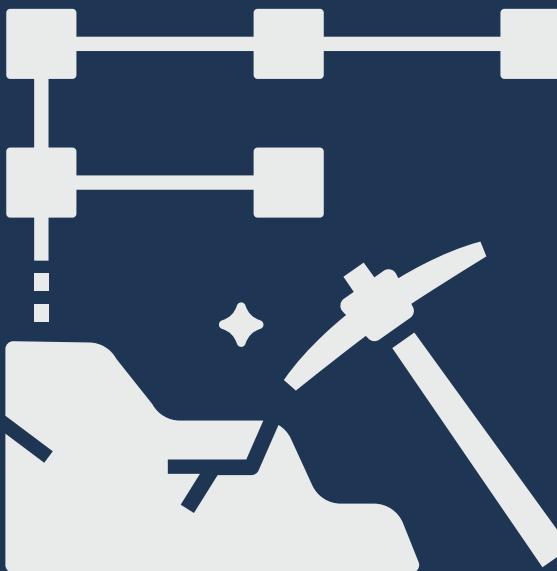


AULA 2

O QUE É DATA MINING?

Mineração de Dados.

É a extração de padrões consistentes de dados.



- **Inteligência artificial**
- **Transformação digital**
- **Marketing com Big Data**

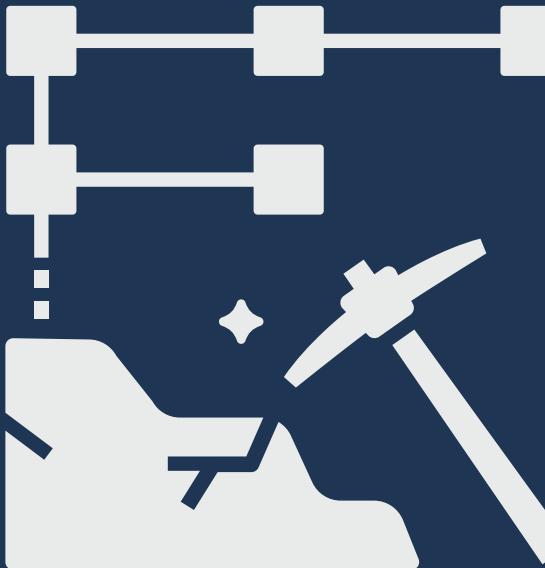


AULA 2

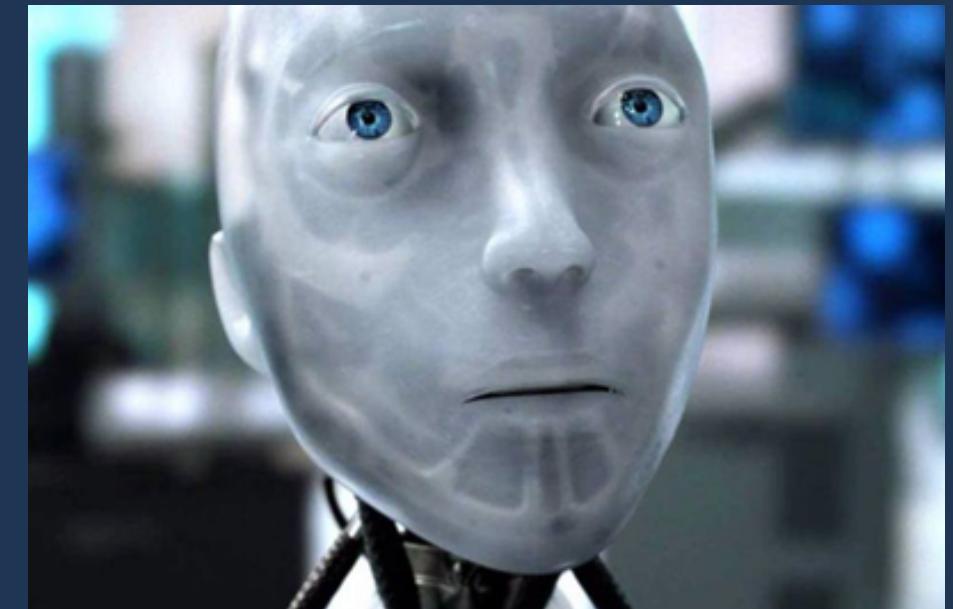
O QUE É DATA MINING?

Mineração de Dados.

É a extração de padrões consistentes de dados.



- **Inteligência artificial**
- **Transformação digital**
- **Marketing com Big Data**

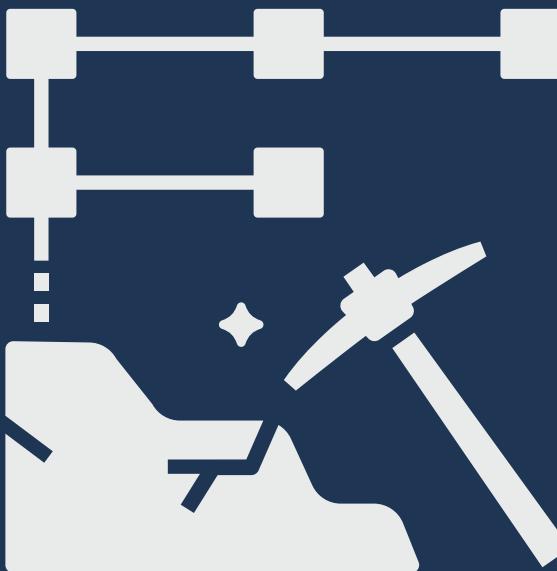


AULA 2

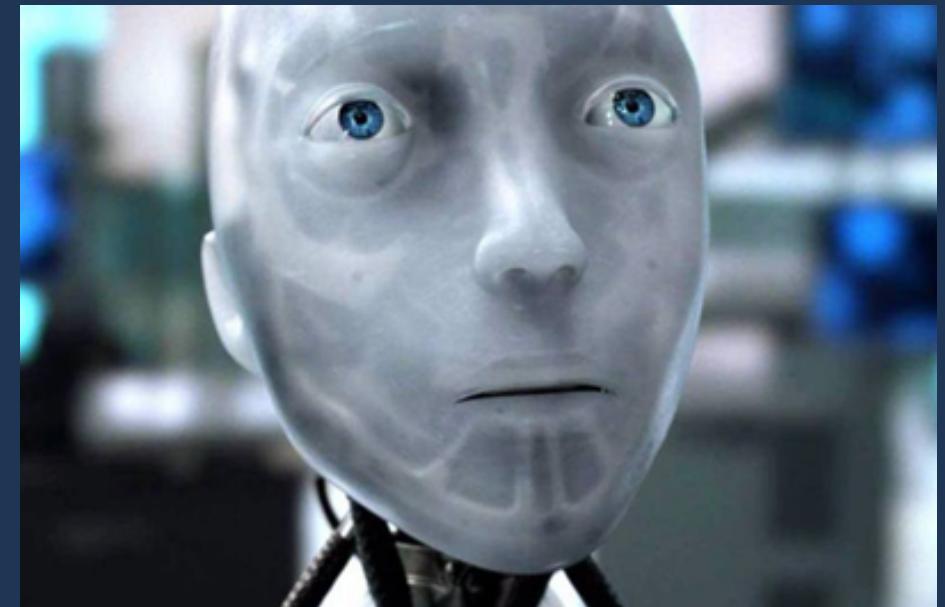
O QUE É DATA MINING?

Mineração de Dados.

É a extração de padrões consistentes de dados.



- **Inteligência artificial**
- **Transformação digital**
- **Marketing com Big Data**

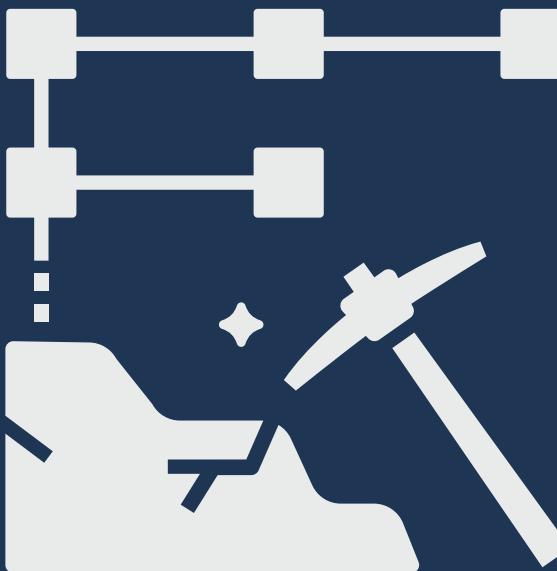


AULA 2

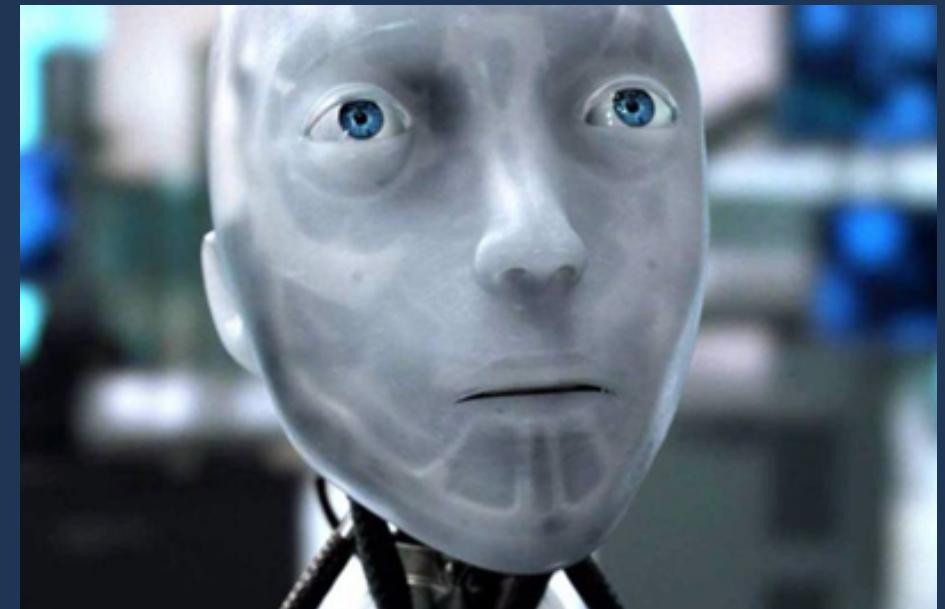
O QUE É DATA MINING?

Mineração de Dados.

É a extração de padrões consistentes de dados.



- **Inteligência artificial**
- **Transformação digital**
- **Marketing com Big Data**



AULA 2

EXEMPLOS IA

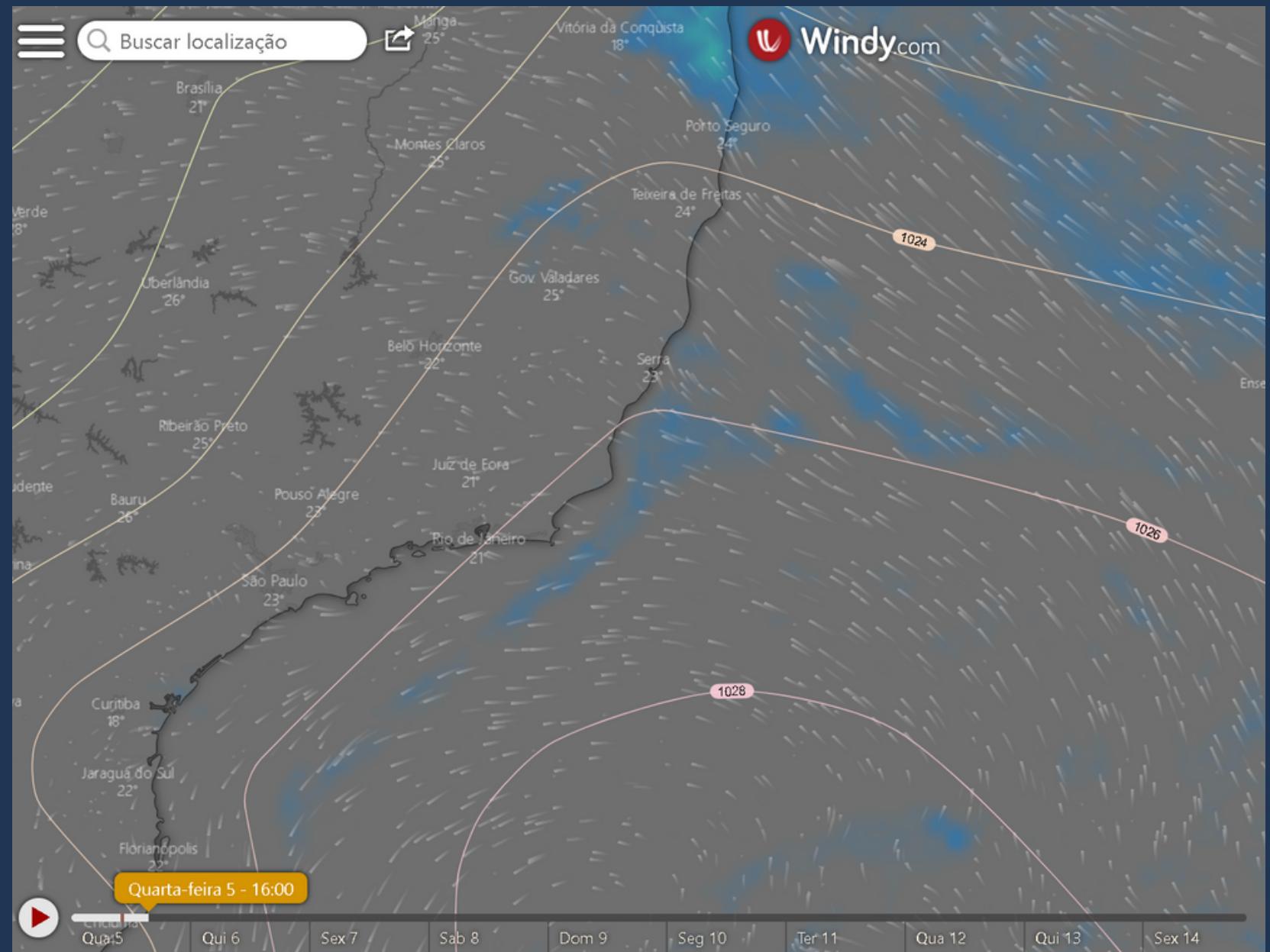


idaco
INSTITUTO DE DESENVOLVIMENTO E AÇÃO COMUNITÁRIA

AULA 2

EXEMPLOS IA

Previsão do tempo

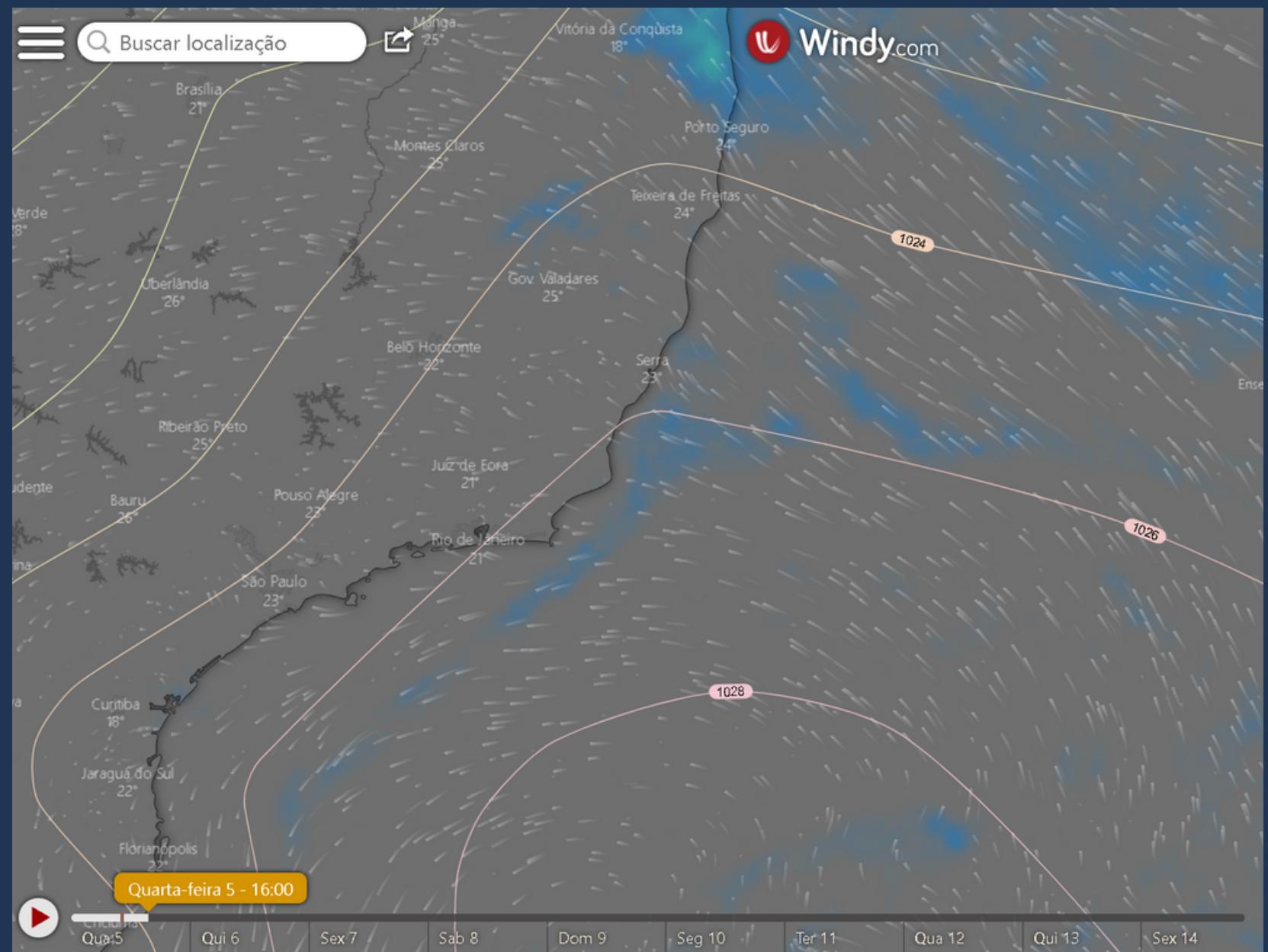


idaco
INSTITUTO DE DESENVOLVIMENTO E AÇÃO COMUNITÁRIA

AULA 2

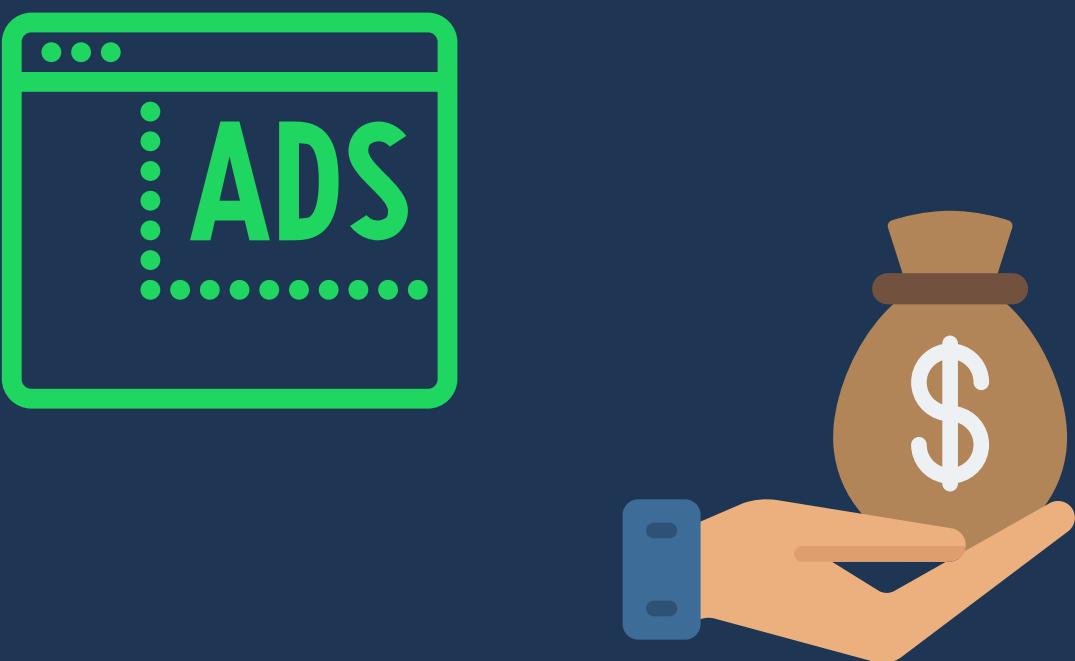
EXEMPLOS IA

Previsão do tempo



idaco

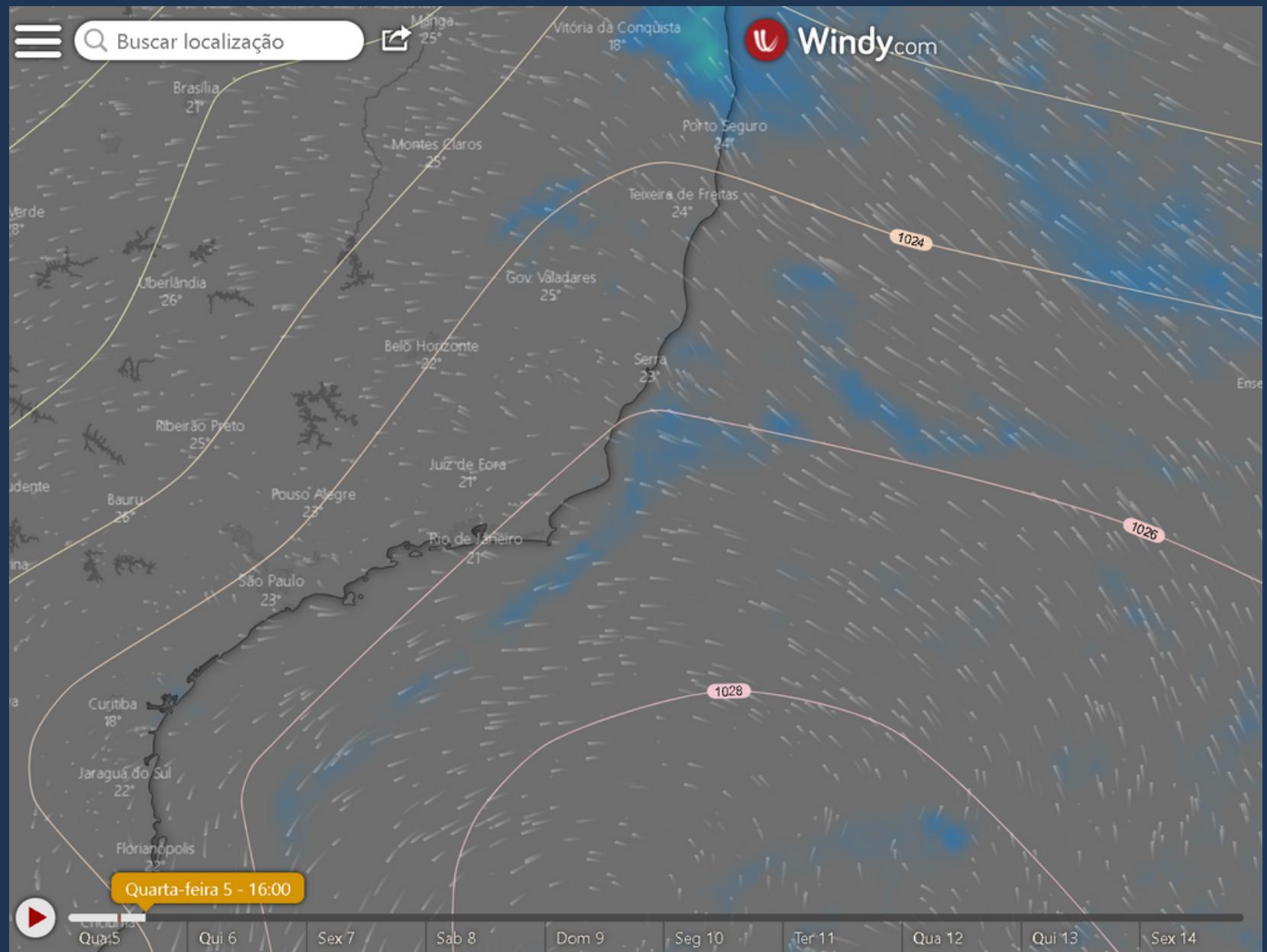
INSTITUTO DE DESENVOLVIMENTO E AÇÃO COMUNITÁRIA



AULA 2

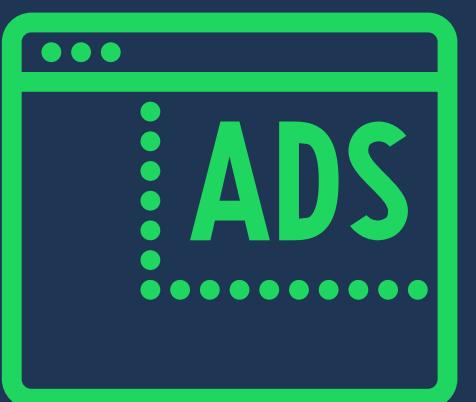
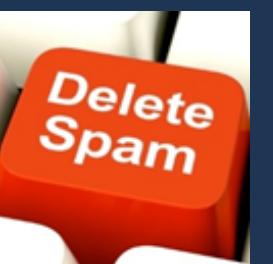
EXEMPLOS IA

Previsão do tempo



idaco

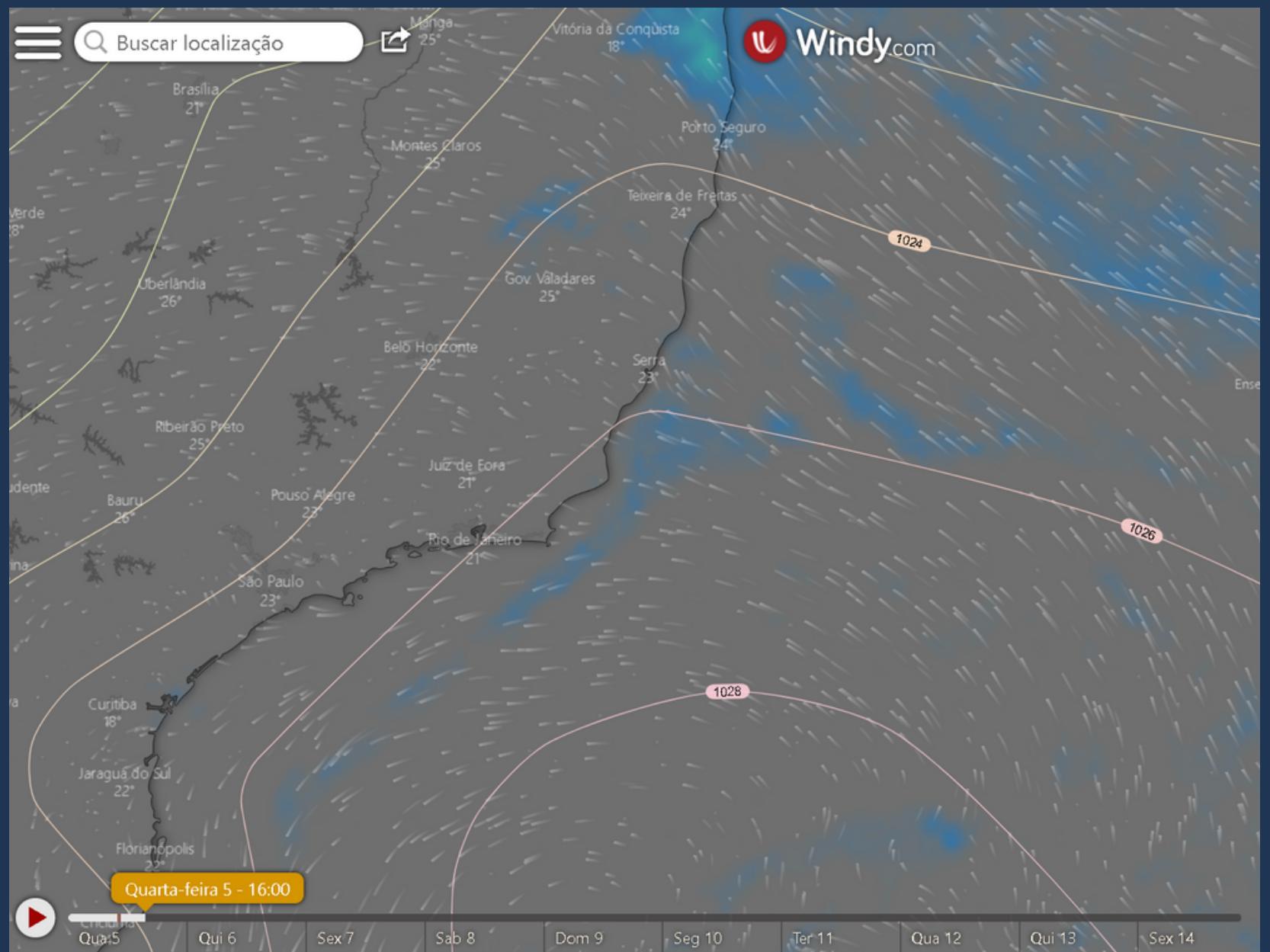
INSTITUTO DE DESENVOLVIMENTO E AÇÃO COMUNITÁRIA



AULA 2

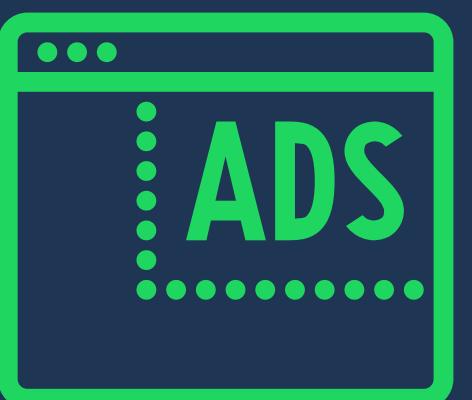
EXEMPLOS IA

Previsão do tempo



idaco

INSTITUTO DE DESENVOLVIMENTO E AÇÃO COMUNITÁRIA



AULA 2

TIPOS DE DADOS

Números x texto



idaco

INSTITUTO DE DESENVOLVIMENTO E AÇÃO COMUNITÁRIA

AULA 2

TIPOS DE DADOS

Números x texto

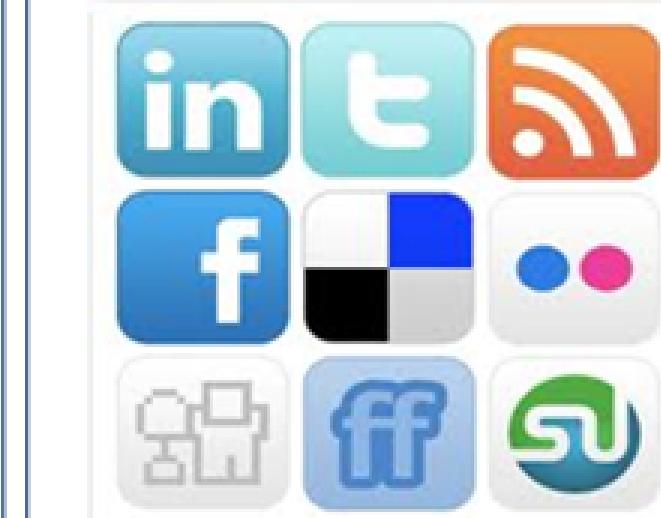


Dados estruturados



0.103	0.176	0.387	0.300	0.379
0.333	0.384	0.564	0.587	0.857
0.421	0.309	0.654	0.729	0.228
0.266	0.750	1.056	0.936	0.911
0.225	0.326	0.643	0.337	0.721
0.187	0.586	0.529	0.340	0.829
0.153	0.485	0.560	0.428	0.628

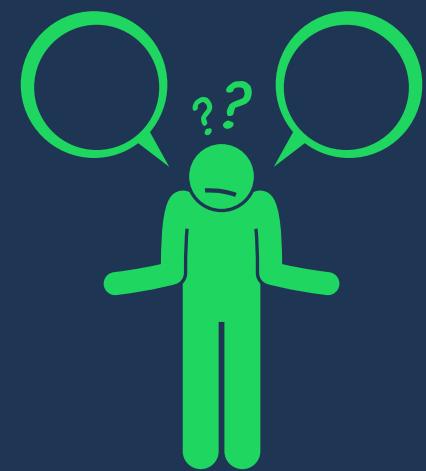
Dados não estruturados



AULA 2

COMO FAZER?

Etapas



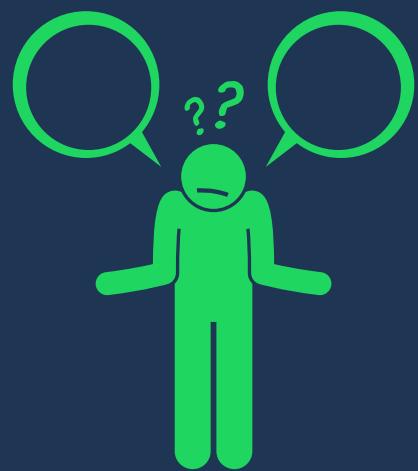
- Definição do problema



AULA 2

COMO FAZER?

Etapas



- Definição do problema



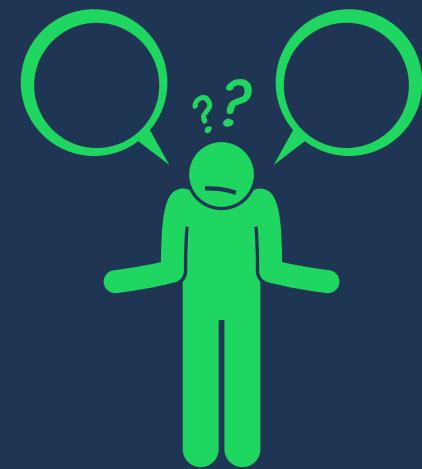
- Limpeza dos dados
 - redundantes/duplicados
 - irrelevantes
 - errôneos



AULA 2

COMO FAZER?

Etapas



- Definição do problema



- Limpeza dos dados
 - redundantes/duplicados
 - irrelevantes
 - errôneos



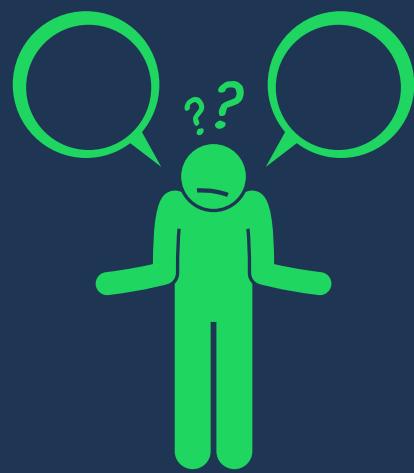
- Filtragem
 - base de dados



AULA 2

COMO FAZER?

Etapas



- **Definição do problema**

- **Limpeza dos dados**
 - redundantes/duplicados
 - irrelevantes
 - errôneos



- **Filtragem**
 - base de dados

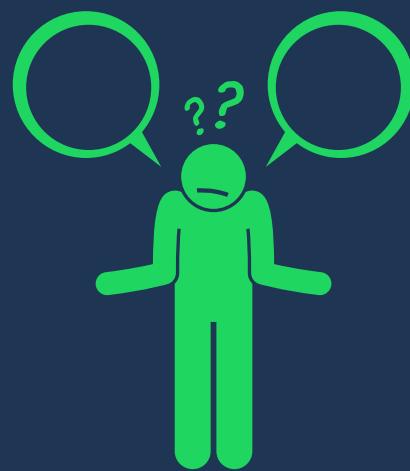
- **Mineração**
 - Redes neurais
 - Árvores de decisão
 - etc



AULA 2

COMO FAZER?

Etapas



- **Definição do problema**

- **Limpeza dos dados**
 - redundantes/duplicados
 - irrelevantes
 - errôneos



- **Filtragem**
 - base de dados



- **Mineração**
 - Redes neurais
 - Árvores de decisão
 - etc



- **Previsão**
 - Modelagem
 - Aprendizado de máquina



AULA 2

MINERAÇÃO DE TEXTOS

Diferentes métodos



idaco

INSTITUTO DE DESENVOLVIMENTO E AÇÃO COMUNITÁRIA

DULUNICAMPIM
EIGALCWPENRG
YNALPALAVRAS
DLZGMCSLTMCNA
HAQPOALPFWHUA
SAFIIONACUCATI

AULA 2

MINERAÇÃO DE TEXTOS

Diferentes métodos

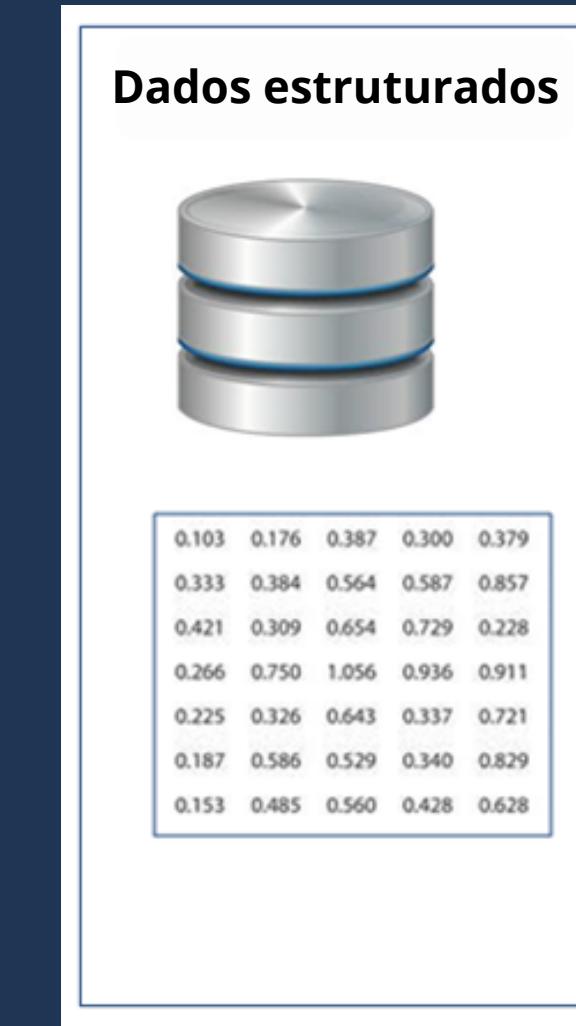


DULUNICAMPIM
EIGALCWPENRG
YNALPALAVRAS
DLZGMCSLTMCNA
HAQPOALPFWHUA
SAFIIONACUCATI

AULA 2

MINERAÇÃO DE TEXTOS

Diferentes métodos



DULUNICAMPIM
EIGALCWPENRG
YNALPALAVRAS
DLZGMCSLTMCNA
HAQPOALPFWHUA
SAFIIONACUCATI

AULA 2

MINERAÇÃO DE TEXTOS

Como estruturar? Pré-processamento

- **Tokenização** ['Quarta-feira','passada','','dia','29','de','julho','','o','Brasil','teve','1595','notificações','de','mortes','por','COVID','.']")
- **Minúsculas e maiúsculas** ['quarta-feira','passada','dia','29','de','julho','o','brasil','teve','1595','notificações','de','mortes','por','covid']
- **Stopwords** ['quarta-feira','passada','dia','29','julho','brasil','1595','notificações','mortes','covid']
- **Plurais e sinônimos** ['quarta-feira','passada','dia','29','julho','brasil','1595','notificação','morte','covid']
- **Stemming** ['quarta-feira','passada','dia','29','julho','brasil','1595','notifica','morte','covid']
- **Bag of Words** {0:'quarta-feira',1:'passada',2:'dia',3:'29',4:'julho',5:'brasil',6:'1595',7:'notifica',8:'morte',9:'covid'}
- **Vetorização (binagem)** [1, 1, 1, 1, 1, 1, 1, 1, 1, 0]



AULA 2

VAMOS PRATICAR?

Python.

https://github.com/clubedocientista/PUD_intro_python



idaco
INSTITUTO DE DESENVOLVIMENTO E AÇÃO COMUNITÁRIA

VAMOS PROGRAMAR?