



TÓPICOS EM CIÊNCIA DE  
DADOS PARA O ESPORTE

# COLETA DE DADOS E SENSORIAMENTO

DIEGO RODRIGUES DSC

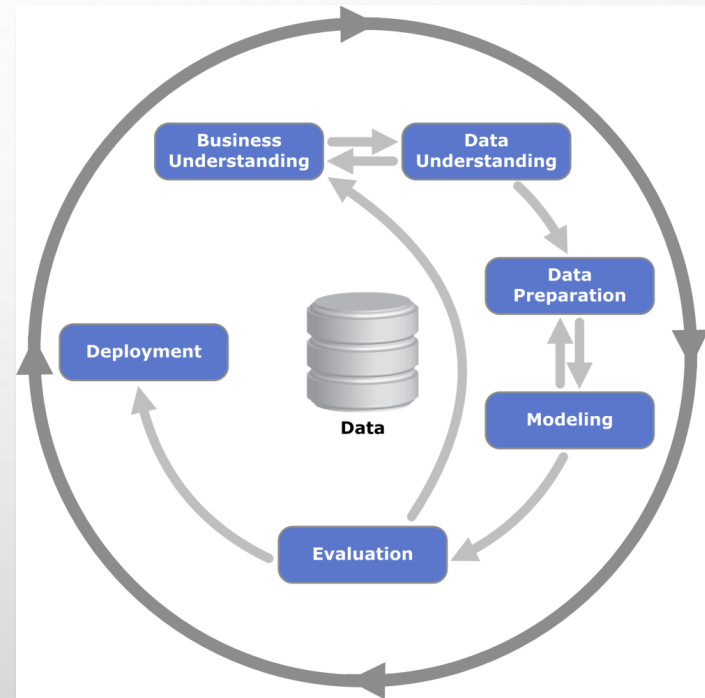
INFNET

# AGENDA

- PARTE 1 : TEORIA
  - MERCADO DE INTELIGÊNCIA ARTIFICIAL
  - MODELO 360
  - CRISP-DM
  - REPRESENTAÇÃO DOS DADOS
  - ADQUIRINDO DADOS
  - FORMATOS PARA DADOS
  - EXEMPLOS
- PARTE 2 : PRÁTICA
  - SETUP INICIAL DO AMBIENTE PYTHON
  - IMPORTAR UMA BASE DE DADOS COM PANDAS

# CRONOGRAMA

DIA	NÚMERO	ÁREA	AULA	TRABALHOS
30/1/2024	1	Intro	Introdução a Disciplina e Organização do Ambiente	
1/2/2024	2	Dados	Coleta de Dados e Sensoriamento	
6/2/2024	3	Estatística	Variáveis Aleatórias	Grupos
8/2/2024	4		Análise Exploratória	
15/2/2024	5		Estatísticas para Ranqueamento	
20/2/2024	6		Ranqueamento Estatístico : ELO	
22/2/2024	7		Ranqueamento Estatístico : Glicko	
27/2/2024	8		Ranqueamento Estatístico : TrueSkill	
29/2/2024	9		Ranqueamento Estatístico : XELO	Base de Dados
5/3/2024	10	ML	Modelos de Aprendizado de Máquina	
7/3/2024	11		Machine Learning: Classificação	
12/3/2024	12		Machine Learning: Regressão	
14/3/2024	13		Machine Learning: Agrupamento	Pesquisa
19/3/2024	14		Machine Learning: Visão Computacional	
21/3/2024	15	Esportes	Aplicações & Artigos: Esportes Independentes	Modelo
26/3/2024	16		Aplicações & Artigos: Esportes de Objeto	
28/3/2024	17		Aplicações & Artigos: Esportes de Combate	
2/4/2024	18		Aplicações & Artigos : Betting	
4/4/2024	19	Workshop	Workshop	
9/4/2024	20		Apresentações de Trabalhos I	Apresentação
11/4/2024	21		Apresentações de Trabalhos II	



CRISP-DM

# MERCADO DE INTELIGÊNCIA ARTIFICIAL

## INOVAR É UMA NECESSIDADE

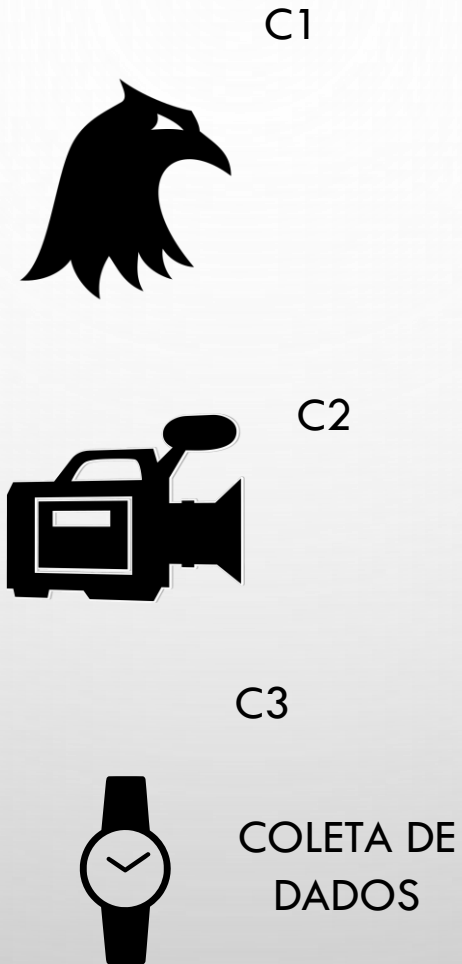
O mundo já está investindo em inteligência artificial e machine learning. E você?

Os *softwares* de IA, são a chave para diagnosticar e resolver os seus principais desafios de inovação e é por isso que cada vez mais empresas fazem investimentos robustos nesse tipo de iniciativa.

De todo o montante investido no setor, cerca de 45% são em aplicações verticais, voltadas para a resolução de problemas específicos das mais variadas naturezas, e assim tornar a operação mais inteligente, com menos desperdício, melhor tomada de decisão, aumento dos lucros e resultados visíveis a médio e longo prazo.



# MERCADO DE INTELIGÊNCIA ARTIFICIAL (ESPORTIVO)



SERVIÇOS  
DIGITAIS





# SCOUTING 360°

ESPORTE POR LAZER  
OU POR COMPETIÇÃO?



## TÉCNICO



Desempenho técnico medido por exercícios recorrentes ou avaliações periódicas.

## FÍSICO



Treinamento físico e dados de saúde dos atletas.

## TÁTICO



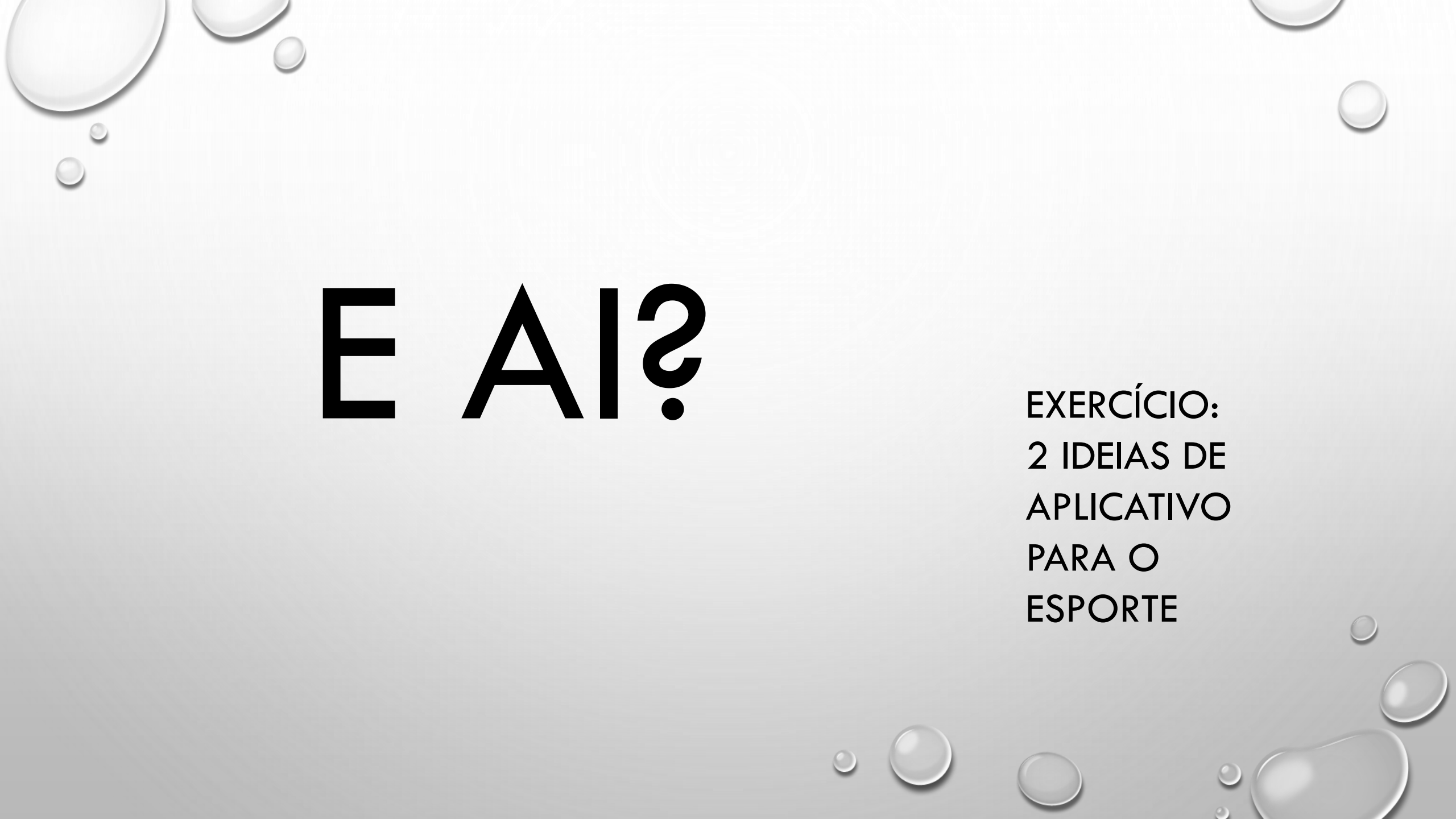
Jogo reduzido ou simulação de combate em treinos semanais.

## ESTRATÉGICO



Integração dos Dados ao Calendário de Competições e seus Resultados.

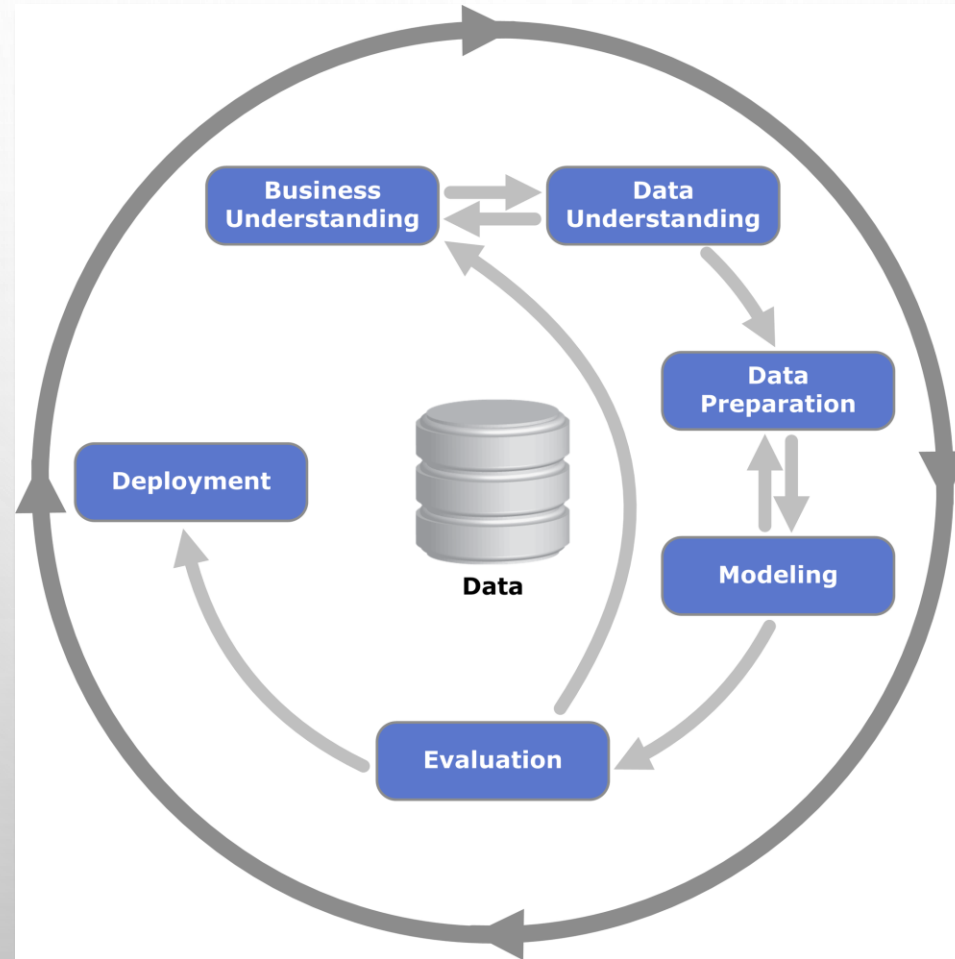




# E AI?

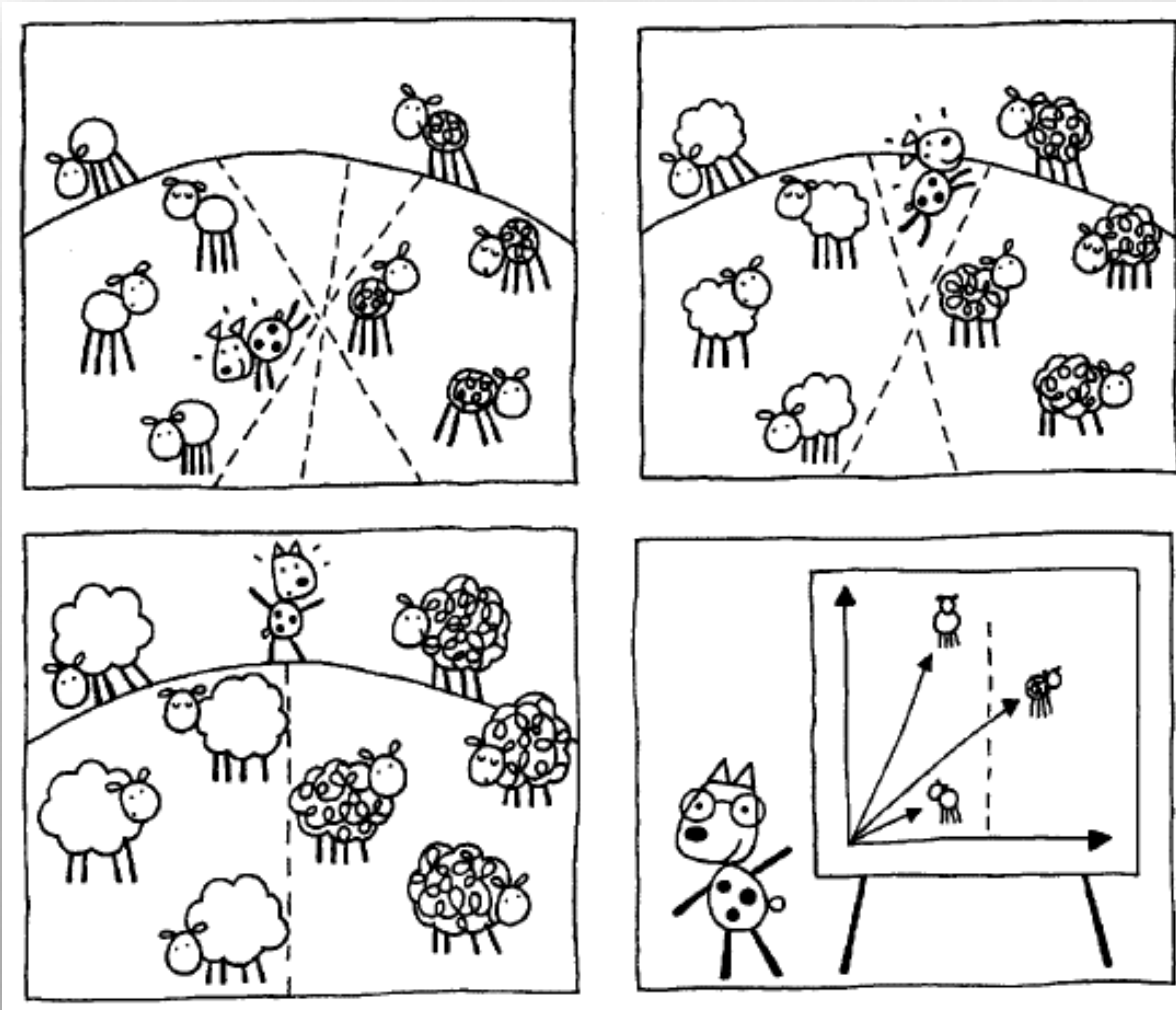
EXERCÍCIO:  
2 IDEIAS DE  
APLICATIVO  
PARA O  
ESPORTE

# CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING





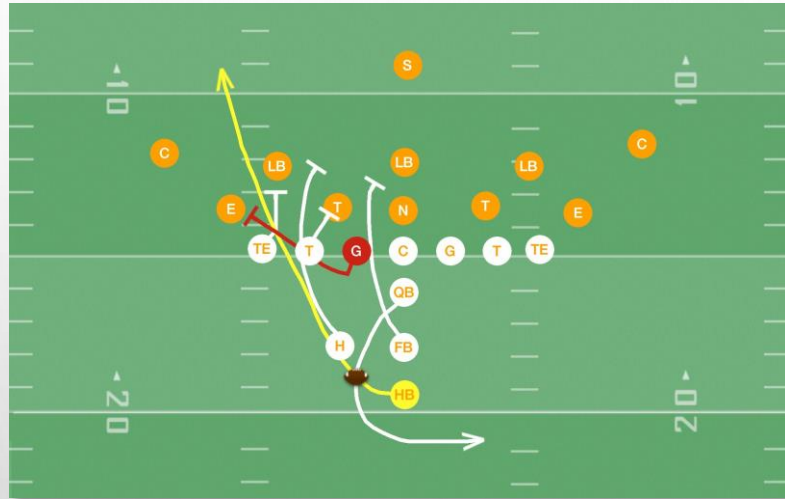
# REPRESENTAÇÃO DOS DADOS



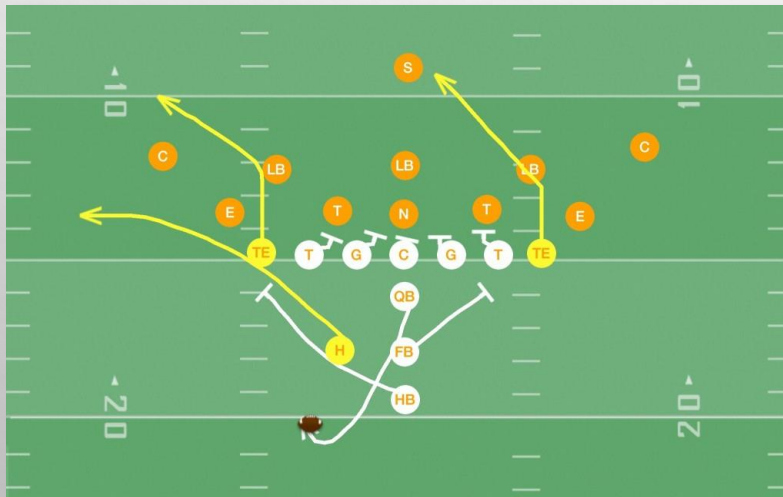
Como quantificar o necessário para transpor a necessidade do mundo físico para o mundo digital?

Como classificar automaticamente entre ratos e elefantes?

# REPRESENTAÇÃO DOS DADOS



Qual a probabilidade do time adversário correr ou passar?



- Contínuo
- Discreto
- Binário
- Categórico
- Classe

N	Atributo	Tipo
1	Distância para o First Down	Contínuo
2	Posição no Campo	Contínuo
3	Diferença no Placar	Contínuo
4	Quarto de Jogo	Discreto
5	Tempo	Contínuo
6	Dois Minutos Finais	Binário
7	Última Jogada - Passe	Binário
8	Última Jogada - Corrida	Binário
9	Última Jogada - Times Especiais	Binário
10	Última Jogada - Jardas	Contínuo
11	Tipo de Jogada (Corrida/Passe)	Classe

# REPRESENTAÇÃO DOS DADOS

game	quarter	time	down	distance	field	score	play
1	1	9	1	10	34	0	2
1	1	42	1	10	47	0	1
1	1	83	2	6	49	0	2
1	1	93	3	6	49	0	2
1	1	119	1	10	58	0	2
1	1	163	2	2	66	0	1
1	1	203	3	1	67	0	1
1	1	239	1	10	69	0	2
1	1	270	2	14	65	0	1
1	1	315	3	13	66	0	2
1	1	364	1	10	80	0	1
1	1	397	2	2	88	0	1
1	1	431	3	5	85	0	2
1	1	476	1	9	91	0	1
1	1	514	2	8	92	0	2
1	1	523	3	8	92	0	2
1	1	529	4	8	92	0	3
1	1	852	1	10	34	3	2
1	1	859	2	10	34	3	1
1	1	891	3	8	36	3	2
1	2	0	1	10	59	3	2
1	2	37	1	10	71	3	2
1	2	46	2	10	71	3	2
1	2	53	3	10	71	3	2
1	2	94	1	14	86	3	1

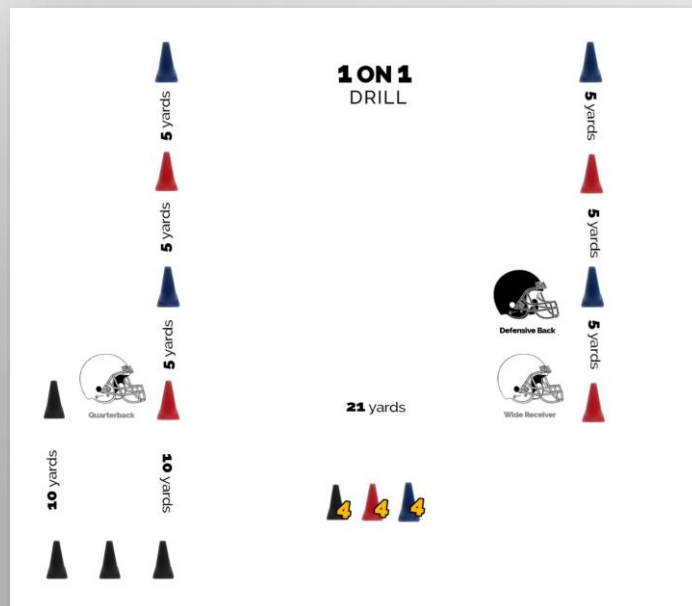
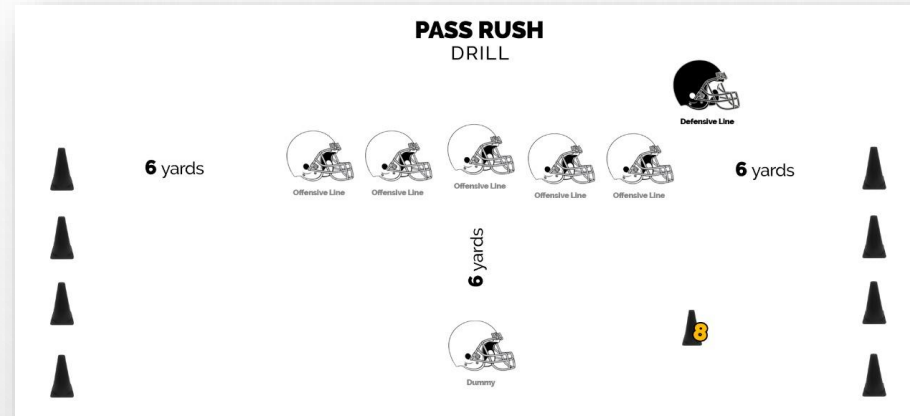
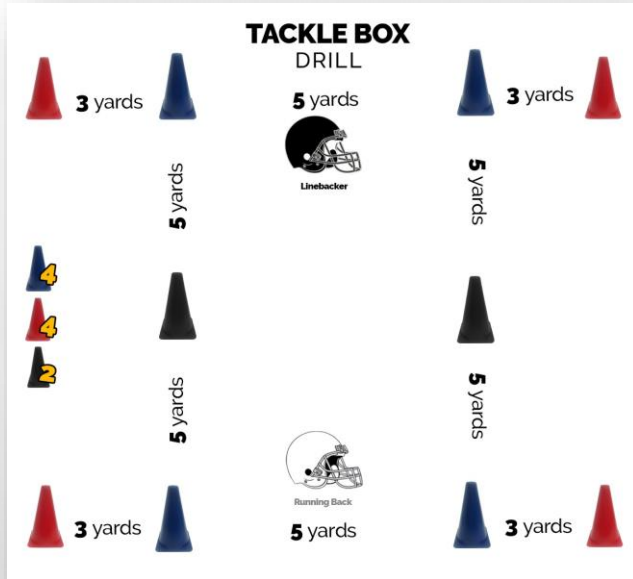
Cada linha é uma observação.

O que define a unicidade da linha é chamado de grão ou chave.

Cada coluna é um atributo ou variável independente.

A planilha inteira é uma amostra.

# ADQUIRINDO DADOS

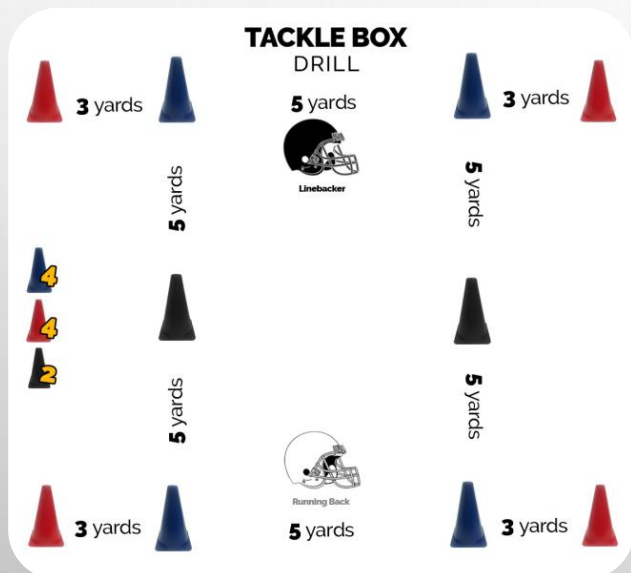


Como garantir a qualidade do dado coletado?

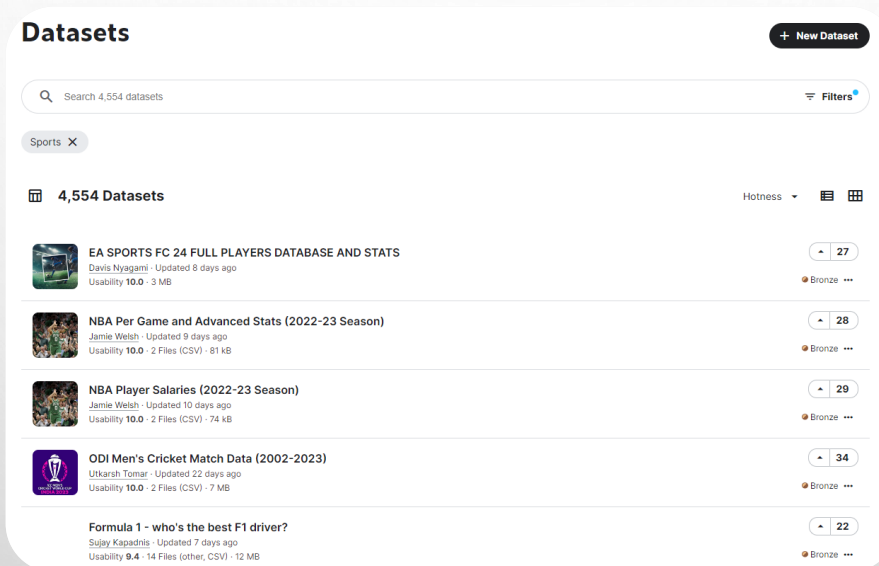
Como garantir que as observações são comparáveis?

Como alinhar observações obtidas por diferentes sensores?

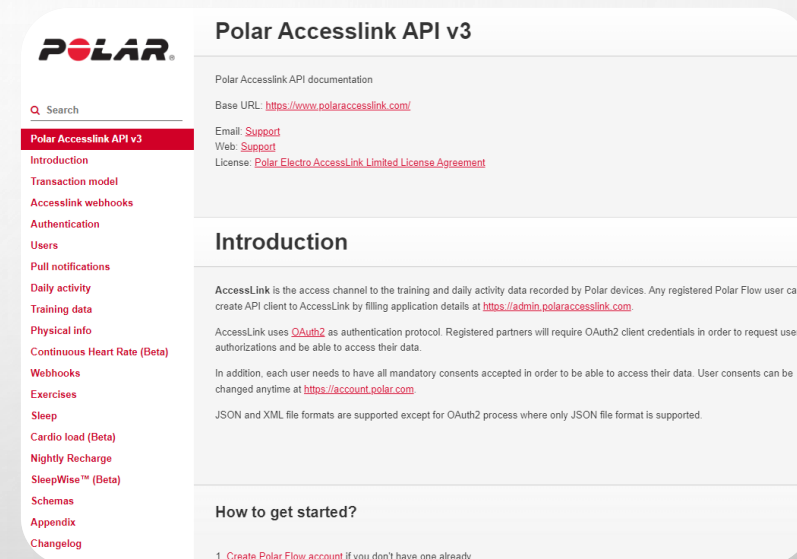
# ADQUIRINDO DADOS



Anotar ou Filmar



Pesquisar Bases Abertas



Coletar Device via API



# FORMATOS PARA DADOS



UTF8  
Separador de coluna  
Separador de Decimal



Mais seguro salvar  
como CSV UTF8 ao  
invés de usar XLSX

{ j s o n }

Estrutura de  
“Dicionário” Python

Padrão na  
comunicação Web



Tensor  
Cada frame com 3  
canais de cores

The background of the slide is a light gray gradient. In the top-left and bottom-right corners, there are several realistic water droplets of various sizes, rendered with soft shadows and highlights to give them a three-dimensional appearance. In the center of the slide, there is a faint, circular watermark. It features a globe with latitude and longitude lines, and the text "EXEMPLOS" is written across it in a light, sans-serif font.

# EXEMPLOS

# SETUP INICIAL DO AMBIENTE PYTHON

4. Bibliotecas  
necessárias para  
o Projeto



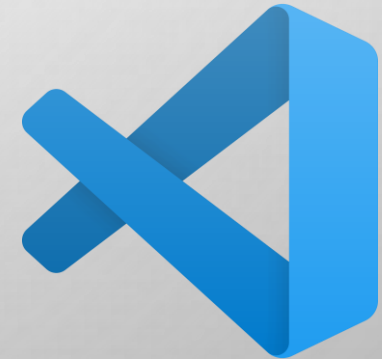
2. Gestor de Ambiente

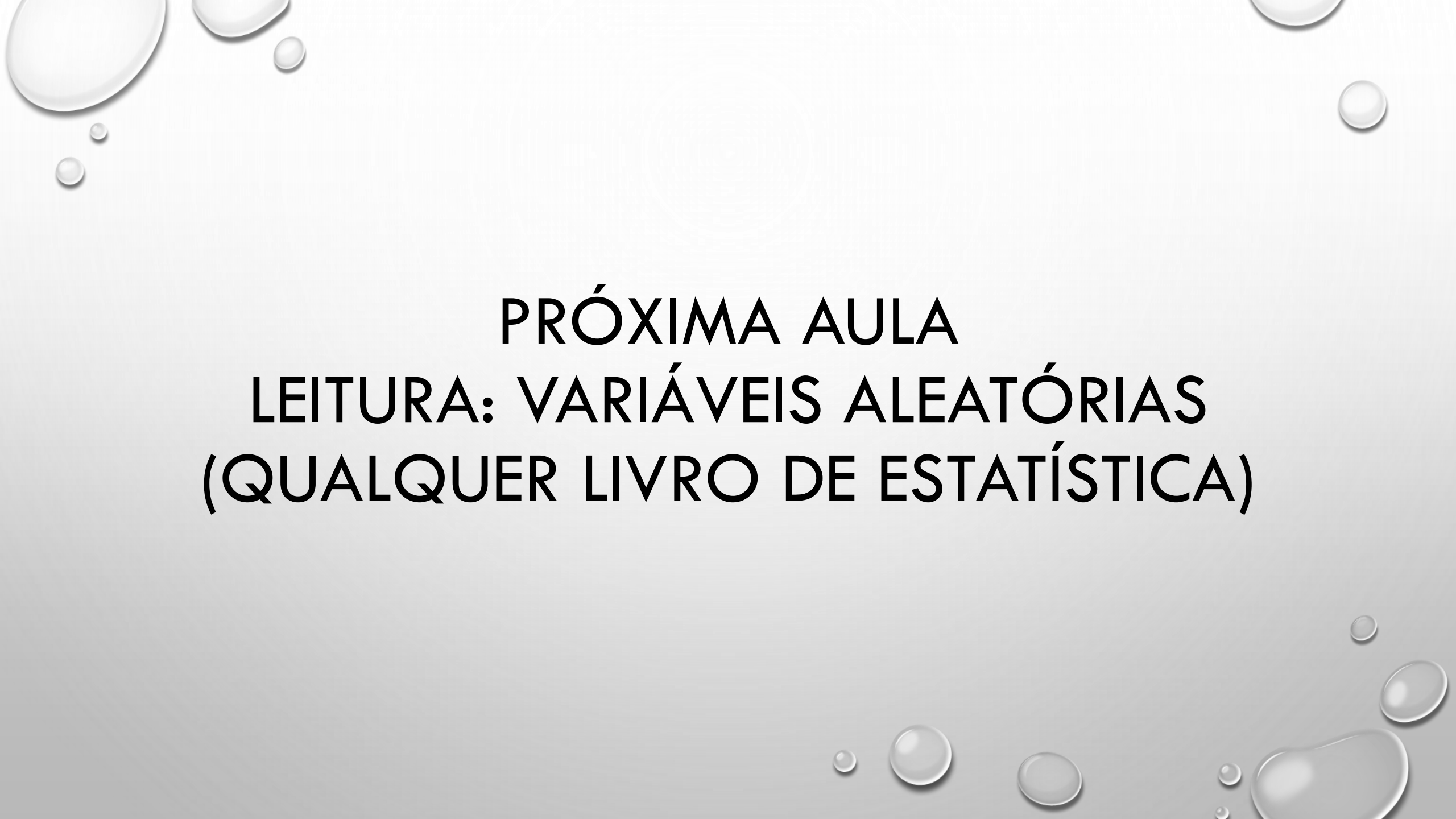


3. Ambiente  
Python do Projeto



1. Editor de Código





**PRÓXIMA AULA**  
**LEITURA: VARIÁVEIS ALEATÓRIAS**  
**(QUALQUER LIVRO DE ESTATÍSTICA)**