

DATA science com PYTHON

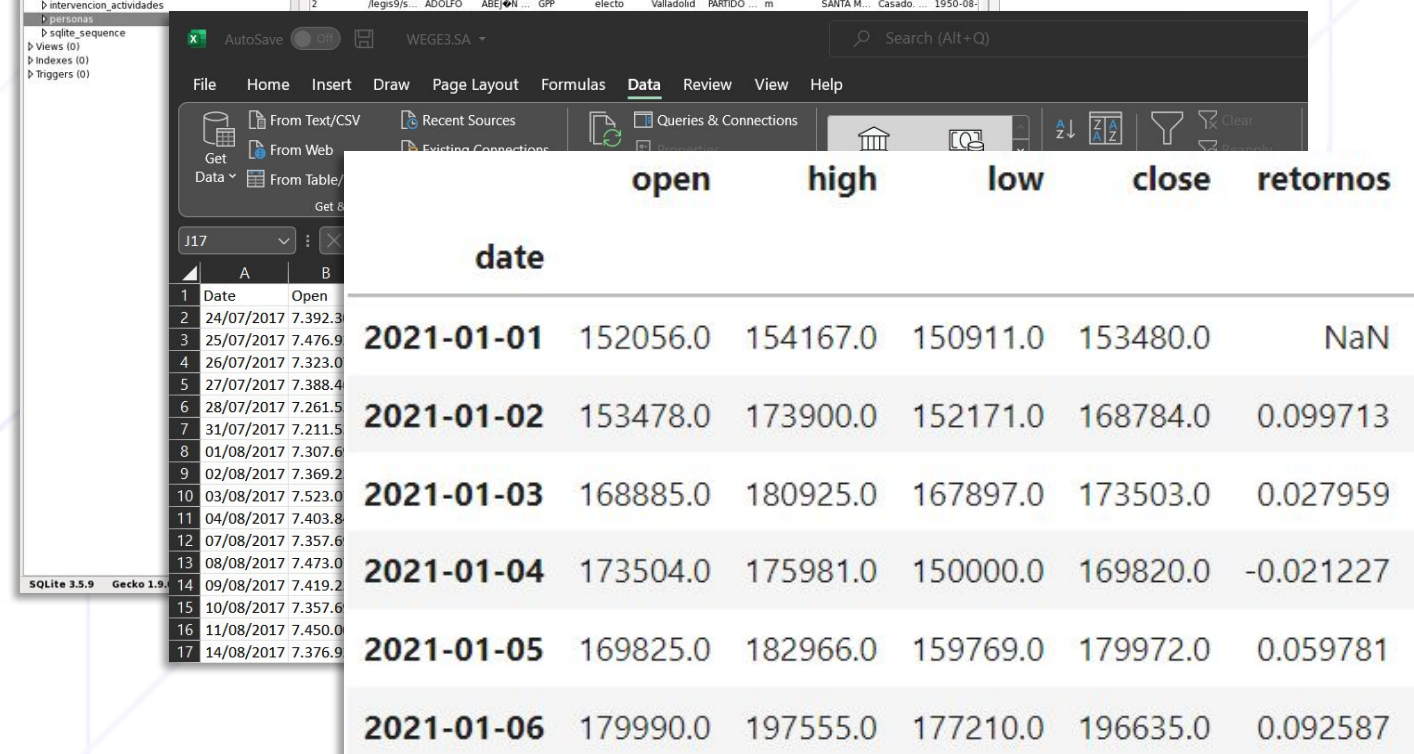

mÓDULO 2

Obtenção e armazenamento de dados



Trabalhando com Dados

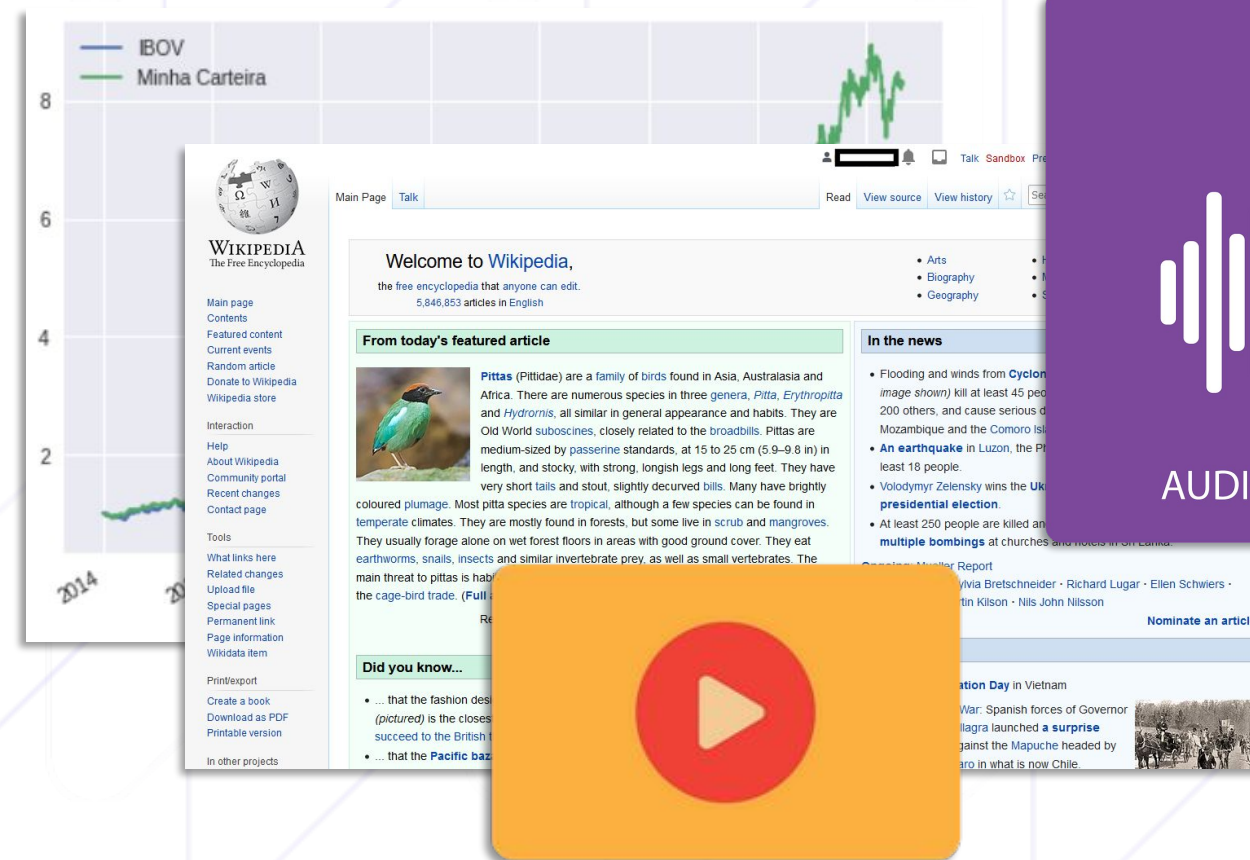
TIPOS DE DADOS



	date	open	high	low	close	retornos
1	2021-01-01	152056.0	154167.0	150911.0	153480.0	NaN
2	2021-01-02	153478.0	173900.0	152171.0	168784.0	0.099713
3	2021-01-03	168885.0	180925.0	167897.0	173503.0	0.027959
4	2021-01-04	173504.0	175981.0	150000.0	169820.0	-0.021227
5	2021-01-05	169825.0	182966.0	159769.0	179972.0	0.059781
6	2021-01-06	179990.0	197555.0	177210.0	196635.0	0.092587

DADOS ESTRUTURADOS

Todas as informações e perspectivas estão incorporadas no preço



DADOS DESESTRUTURADOS

Padrões de movimentos do passado influenciam o que irá acontecer com o preço

E DE ONDE VÊM ESSES DADOS?



Banco de dados de arquivos SIMPLES

*Arquivos como .xlsx, .csv, .txt, etc.
É a forma mais simples de armazenamento e
compartilhamento de dados. Normalmente utilizado
para dados de pequeno volume.*

Armazenamento "On-Premises"

*Servidores estão localizados dentro da infraestrutura
da própria empresa. Os dados são compartilhados
entre os membros da empresa através da rede local.*

Armazenamento "CLOUD"

*Armazenamento feito em servidores remotos, longe
do cliente final e disponível em qualquer lugar do
mundo. O acesso é feito através da internet.*

DADOS VIA WEB

*Os dados podem ser extraídos de sites através de
'webscraping', ou podem também ser obtidos a
partir de fornecedores externos, através de APIs.*

CLASSIFICAÇÃO

Dentro de um banco de dados, normalmente as variáveis assume os seguintes formatos

Codigo do Produto	Símbolos (coleções)	Preço Produto	Quantidade Vendida	Estado Civil	Idade	Região	Aspectos E-commerce
347 148623749	Trevo da Sorte	206	2	Em um relacionamento estável	16	Sul	Guia de tamanhos e de medidas
348 148623751	Trevo da Sorte	205	2	Em um relacionamento estável	16	Sudeste	Ter a medida das peças explicada
349 148623753	Trevo da Sorte	206	3	Em um relacionamento estável	16	Sudeste	Ter a medida das peças explicada
350 148623755	Trevo da Sorte	201	1	Em um relacionamento estável	17	Sul	Guia de tamanhos e de medidas

variáveis quantitativas

Medida numérica de algum aspecto dos dados. Expressam alguma característica mensurável. Também chamadas de variáveis numéricas.

variáveis qualitativas

Qualificam o indivíduo ou observação em categorias. Expressam características que não são numéricas. Também chamadas de variáveis categóricas.

PROTEÇÃO DE DADOS

LGPD

IMPORTÂNCIA

Passou a definir de forma jurídica o que são dados pessoais

PADRONIZAÇÃO

Visa regulamentar e padronizar normas e práticas orientadas ao manuseio e armazenamento de dados pessoais

FISCALIZAÇÃO

Falhas no cumprimento das normas estabelecidas pela lei podem gerar multas pesadas

CONSENTIMENTO

Os cidadãos precisam consentir que seus dados possam ser utilizados para tratamento pelas empresas. Passam a ter o direito de não autorizar, revogar seu consentimento, pedir que seus dados sejam deletados, dentre outros.



OBRIGADO