

BUTIÁ INVESTIMENTOS  
TESTE DE ESTÁGIO

---

**Exercício 1** Considere os dois arquivos **csv** disponibilizados junto com estes exercícios:

- O primeiro, **marketData.csv**, contém preços de fechamento do índice SP500 e de 10 ações de empresas do mercado Americano. Os preços compreendem ao período de 2007 a 2009.
- O outro arquivo, **sentimentData.csv**, inclui o resultado da análise semântica de notícias relacionadas a cada uma das 10 empresas. Cada linha representa uma notícia textual que foi pré-processada e transformada em três valores numéricos: as probabilidades da notícia ser positiva, negativa ou neutra (a soma das probabilidades é 1). Há também uma coluna adicional indicando a relevância daquela notícia àquela empresa. A relevância é um valor entre 0 e 1 onde 1 indica alta relevância (por exemplo, a notícia é sobre a empresa especificamente). O período das notícias é o mesmo, 2007 a 2009.

Neste exercício, o objetivo é produzir um Notebook (em **Python** ou **R**) onde você deve:

1. Utilizar alguma lógica para gerar séries de sentimento agregado para cada ativo, com a mesma frequência temporal das séries de preços. Para cada notícia individual, você pode combinar as probabilidades para criar um *score* único e opcionalmente utilizar a relevância.
2. Plotar gráficos comparando a média móvel dos sentimentos diários de cada ativo com as respectivas séries de preços.
3. Calcular um sentimento agregado relativo ao índice SP500 combinando as séries de cada ativo, e comparar a média móvel desta série agregada com os preços do SP500.

Você pode também opcionalmente substituir (ou complementar) os passos 2 e 3 acima caso tenha alguma ideia interessante. Pode ser, por exemplo, alguma análise exploratória ou uma simulação simplificada de alguma estratégia baseada nestes dados.

**Exercício 2** Considere que você possua as seguintes tabelas em um banco de dados para representar preços diários de índices financeiros (benchmarks) e ativos:

AssetPrice		BenchmarkPrice	
PK,FK1	asset (varchar)	PK,FK1	benchmark (varchar)
PK	date (date)	PK	date (date)
	open (float) low (float) high (float) close (float) volume (float)		open (float) low (float) high (float) close (float)

Você carregou os preços de fechamento do benchmark ABC entre duas datas e o próximo passo é carregar os preços de fechamento do ativo XYZ. Ao carregar os dados de XYZ, é necessário que as datas carregadas sejam idênticas às do benchmark, independente se há preços faltando ou sobrando para XYZ. No exemplo abaixo, as colunas **Dados carregados** indicam qual deve ser o resultado da busca caso as datas sejam entre 02/08/2021 e 16/08/2021.

Dados no banco				Dados carregados	
ABC		XYZ		XYZ	
date	close	date	close	date	close
02/08/2021	10.23	02/08/2021	21.45	02/08/2021	21.45
03/08/2021	10.35	03/08/2021	21.48	03/08/2021	21.48
05/08/2021	10.09	04/08/2021	21.75	05/08/2021	22.01
06/08/2021	10.58	05/08/2021	22.01	06/08/2021	NaN
09/08/2021	10.61	09/08/2021	21.97	09/08/2021	21.91
10/08/2021	10.54	11/08/2021	21.73	10/08/2021	NaN
11/08/2021	10.43	12/08/2021	21.84	11/08/2021	21.71
12/08/2021	10.50	13/08/2021	21.77	12/08/2021	21.84
13/08/2021	10.61	16/08/2021	21.95	13/08/2021	21.77
16/08/2021	10.64			16/08/2021	21.95

Escreva uma query em SQL que garanta que os preços de fechamento serão carregados corretamente.