

## Lista de Exercícios – Agregação de Series/DataFrames

1. Considere os conjuntos de dados contidos no [link 1](#) e no [link 2](#). Construa um novo conjunto de dados a partir da agregação desses dois, usando as três funções diferentes que estudamos. Comente a diferença entre os resultados de cada função.
2. Crie um DataFrame para cada um dos dicionários a seguir, e faça as seguintes operações:

```
raw_data_1 = {
    'subject_id': ['1', '2', '3', '4', '5'],
    'first_name': ['Alex', 'Amy', 'Allen', 'Alice', 'Ayoung'],
    'last_name': ['Anderson', 'Ackerman', 'Ali', 'Aoni', 'Atiches']}

raw_data_2 = {
    'subject_id': ['4', '5', '6', '7', '8'],
    'first_name': ['Billy', 'Brian', 'Bran', 'Bryce', 'Betty'],
    'last_name': ['Bonder', 'Black', 'Balwner', 'Brice', 'Btisan']}

raw_data_3 = {
    'subject_id': ['1', '2', '3', '4', '5', '7', '8', '9', '10', '11'],
    'test_id': [51, 15, 15, 61, 16, 14, 15, 1, 61, 16]}
```

- (a) Una os dois primeiros DataFrames no sentido das linhas e chame o DataFrame resultante de *all\_data* (use *join()*);
  - (b) Repita a operação, fazendo agora no sentido das colunas, e chame o resultado de *all\_data\_col*;
  - (c) Funda *all\_data* e *data3* a partir da coluna *subject\_id* (use *merge()*);
  - (d) Una somente os dados que tem o mesmo valor na coluna *subject\_id*, para os conjuntos *data1* e *data2*;
  - (e) Mescle todos os valores em *data1* e *data2*, com registros correspondentes de ambos os lados, quando disponíveis.
3. Considere o conjunto de dados disponíveis nos links a seguir, referentes a pesquisa [link](#) e extensão, disponíveis nesse [link](#), do Instituto Federal do Norte de Minas – IFNMG.

Construa um novo conjunto de dados, agregando informações dos diversos conjuntos disponíveis nos links acima, que contenha projetos de pesquisa realizados, pesquisadores, projetos de extensão executados e eventos e prestação de serviços executados pela extensão, por *campus* e por ano.

As premissas para o exercício são:

- (a) nenhum arquivo deve ser acessado localmente, ou seja, todos os dados devem ser carregados via url;
  - (b) todas as suas escolhas sobre como executar a tarefa devem ser apresentadas e discutidas antes do código;
  - (c) o nome dos *campi* devem ser transformados nos índices do conjunto de dados;
  - (d) o novo conjunto de dados deve ser gravado no formato CSV.
4. Considere os dados de vendas anuais de etanol hidratado e derivados de petróleo por município – gasolina de aviação, disponíveis no [link](#).

Construa um novo conjunto de dados, a partir dos disponíveis no link acima, que tenha informações de vendas por estado e por ano.

As premissas para o exercício são:

- (a) nenhum arquivo deve ser acessado localmente, ou seja, todos os dados devem ser carregados via url;
- (b) todas as suas escolhas sobre como executar a tarefa devem ser apresentadas e discutidas antes do código;
- (c) o novo conjunto de dados deve ser gravado no formato JSON;
- (d) É necessário responder a pergunta: o que é JSON e quais as vantagens dele sobre o CSV?