



Residência em Tecnologia da Informação e Comunicação

INSTRUÇÃO PRÁTICA - IP-P005

PP - PROGRAMAÇÃO EM PYTHON

OBJETIVOS DA ATIVIDADE

Trabalhar com **Pandas** e suas funcionalidades e características básicas. Aplicar os recursos e funcionalidades básicas para operar com tabelas de dados em **Pandas**.

Exercício 1: Criando um conjunto de dados

Vimos em sala de aula como trabalhar com objetos `Series` e `DataFrames` e as vantagens de se trabalhar com conjuntos de dados armazenados em versões aprimoradas de matrizes `ndarrays` de **NumPy** nas quais as linhas e colunas são identificadas com rótulos em vez de simples índices inteiros.

Vamos criar nosso primeiro conjunto de dados para análise:

- Utilizando apenas os membros da equipe com "cobaías" vamos coletar as seguintes informações:
 - **identificador:** Para fazermos uma pesquisa sem expor os dados pessoais cada membro da equipe será identificado por um código de tipo string formado pela `tic18Py` seguido dos três primeiros números do CPF e os dois últimos do ano de nascimento. Exemplo `tic18Py06472`;
 - **Idade:** Um número inteiro com a idade do membro da equipe;
 - **Formação:** Um código formado por números inteiros que significam:
 - 0: Formação técnica.
 - 1: Formação técnica graduação em andamento
 - 2: Graduação em andamento
 - 3: Graduação concluída
 - **Área de formação Geral:** Se a resposta anterior for diferente de zero, um código formado por números inteiros que significam:
 - 0: Engenharia
 - 1: Computação
 - **Área de formação específica:** Se a resposta a pergunta anterior for 0 ou 1 este campo será uma string com o curso específico que está sendo cursado;
 - **Andamento da graduação:** Se a resposta para Formação for 1 ou 2 colocar aqui um valor de ponto flutuante indicando o percentual de curso concluído;
 - **Tempo de formado:** Se a resposta para Formação for 3 colocar aqui um valor inteiro com a quantidade de anos de formado;
 - **Tinha experiência prévia em programação?** (True ou False)

Colete estas informações, pode ser manualmente, e crie listas com as mesmas. Aqueles campos para os que não se aplique resposta no questionário preencha com `None`.

```
In [ ]: identificador = []
        idade = []
        formação = []
        formaçãoGeral = []
        formaçãoEspecífica = []
        andamentoGraduação = []
        tempoFormação = []
        experiênciaPrevia = []
```

Exercício 2: Trabalhando com Series

Vamos criar um objeto `Index` utilizando a lista de identificador. Utilizando este objeto vamos converter cada uma das listas, geradas no exercício anterior, em objetos da classe `Series`. Com base nos objetos criados:

- Determine qual a idade média dos membros da equipe, qual ou quais os membros mais jovens e qual o quais os membros mais velhos.
- Identifique se, pelo perfil de formação dos membros da equipe, os membros são predominantemente técnicos, graduandos ou graduados.
- De acordo com a lista de Formação Geral identifique se e equipe é formada majoritariamente por engenheiros ou por formados em algum curso da área de computação.

```
In [1]: import numpy as np
import pandas as pd
print(np.__version__ )
print(pd.__version__ )
```

```
Intel MKL WARNING: Support of Intel(R) Streaming SIMD Extensions
4.2 (Intel(R) SSE4.2) enabled only processors has been deprecated.
Intel oneAPI Math Kernel Library 2025.0 will require Intel(R) Adva
nced Vector Extensions (Intel(R) AVX) instructions.
Intel MKL WARNING: Support of Intel(R) Streaming SIMD Extensions
4.2 (Intel(R) SSE4.2) enabled only processors has been deprecated.
Intel oneAPI Math Kernel Library 2025.0 will require Intel(R) Adva
nced Vector Extensions (Intel(R) AVX) instructions.
1.26.2
2.1.4
```

Exercício 3: Trabalhando com DataFrames

Crie um `DataFrame` utilizando as informações coletadas. Utilize o mesmo `Index` do exercício anterior e de nomes apropriados às colunas. Você pode criar o objeto a partir das listas coletadas ou dos objetos `Series`. Veja como foi feito o tratamento de dados faltando.

- Extraia as informações do exercício anterior mas agora a partir do `DataFrame`.
- A coluna que armazena a formação específica pode conter diferentes rótulos. Proponha uma estratégia para tratar os dados nesta coluna. Leve em consideração que os dados coletados até aqui são poucos mas que esta tabela pode ser escalada para analisar todos os residentes da trilha **Python** ou, eventualmente, das outras trilhas.

```
In [ ]: # Gerando DataFrame
```

Exercício 3: Trabalhando com `MultiIndex`

Agora vamos criar uma nova dimensão dos dados considerando cada trilha por separado. Agrupe os índices por trilha **Python**, **.Net** e **Java**. Para testar o funcionamento troque com colegas de outras equipes as listas coletadas e utilize elas para preencher os dados das duas novas dimensões.

- Extraia as informações do exercício anterior mas agora a partir do `DataFrame` com `MultiIndex` de forma a caracterizar o perfil dos alunos da residencia.
- Repita o procedimento anterior mas, desta vez, para cada turma por separado buscando caracterizar os alunos de cada trilha.

```
In [ ]: # Gerando MultiIndex
```