

Bootcamp IGTI: Engenheiro(a) de Dados

Desafio Prático

Módulo 2: Linguagem Python aplicada a Engenharia de Dados

Objetivos de Ensino

Exercitar os seguintes conceitos trabalhados no Módulo:

1. Pandas.
2. Numpy.
3. Classes, Atributos e Métodos.
4. Lógica de programação.

Enunciado

Neste trabalho prático utilizaremos o dataset presente neste [link](#). Baixe o dataset antes de prosseguir.

As atividades abaixo buscam exercitar os conceitos que foram apresentados do módulo a partir do cumprimento de objetivos práticos. Não existe uma forma única de resolver as atividades, mas ao buscar realizá-las, os conceitos serão automaticamente trabalhados.

Atividades

1. Criando uma classe para converter país em continente, a partir do arquivo de usuários: **Arquivo de Usuários: 'worldwide internet users - users.csv'**
 - 1.1. Abra o arquivo de usuários usando pandas. Dica:
 - https://pandas.pydata.org/docs/reference/api/pandas.read_csv.html
 - 1.2. Crie uma classe 'MapaPaísContinente' que possui um atributo do tipo dicionário. Chame esse atributo de 'mapeamento_continente'.
 - 1.3. Crie um método "mapear" nessa classe, que receba um pandas DataFrame e o nome de duas colunas desse dataframe (uma representa

o país e a outra o continente), que percorra duas colunas em conjunto, inserindo 'mapeamento_continente' uma chave para cada país, com valor sendo o continente. Dicas:

- <https://docs.python.org/3.3/library/functions.html#zip>
- <https://pandas.pydata.org/docs/reference/api/pandas.Series.tolist.html>

- 1.4. Crie um método 'retorna_continente' que retorna o continente a partir do nome do país, recebido como um parâmetro que use o atributo 'mapeamento_continente', que use o bloco try/except para lidar com países que não existam dentro do dicionário de mapeamento. Nesse caso, o método deve retornar None e printar a mensagem “Não existe esse país em nossa base”.
2. Adicionando a informação de continente na tabela de velocidades, a partir do mapeamento do arquivo de usuários: Arquivo de velocidades: 'worldwide internet speed in 2022 - avg speed.csv'.
 - 2.1. Abra o arquivo de velocidades usando pandas, armazene em uma variável do tipo Pandas Dataframe.
 - 2.2. Adicione uma nova coluna 'Continente' ao DataFrame. Use o método de mapeamento de continente da classe 'MapaPaisContinente'. Dicas:
 - <https://stackoverflow.com/questions/33518124/how-to-apply-a-function-on-every-row-on-a-dataframe>
 - <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.apply.html>
 - https://pandas.pydata.org/docs/getting_started/intro_tutorials/05_add_columns.html#b
3. Trabalhando com os dados de usuários de Internet: voltando ao arquivo de Usuários: 'worldwide internet users - users.csv'.
 - 3.1. Abra o arquivo de usuários de internet novamente. Dica:
 - https://pandas.pydata.org/docs/reference/api/pandas.read_csv.html

3.2. Crie novas colunas do tipo 'int' para as colunas 'Population' e 'Internet users'. Dicas:

- Existem valores faltantes na base, sendo necessário filtrar antes de fazer a criação das novas colunas.
- <https://stackoverflow.com/questions/20625582/how-to-deal-with-settingwithcopywarning-in-pandas>
- <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.isna.html>
- <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.dropna.html>
- <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.astype.html>
- <https://pandas.pydata.org/docs/reference/api/pandas.Series.str.replace.html>
- <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.describe.html>