

Tópicos Especiais em Sistemas de Informação I

ATV A2 - Pandas e Visualização

1. Escolha ou crie 5 exercícios sobre pandas e apresente respostas baseadas em dados originais. Use um dataset da internet e informe a url para download do mesmo. Escolha 5 exercícios abrangendo Series, Dataframe e Index.
2. Apresente visualizações básicas para o dataset escolhido: (a) histograma de pelo menos 1 atributo; (b) heatmap das correlações entre atributos; (c) scatter plot das três melhores correlações encontradas no dataset; (d) plot de barras para um atributo categórico utilizando `value_counts()`.

Entregue código em Jupyter ou link do seu notebook no Google Colab.

Obs: Evite exercícios triviais e evite soluções copiadas de colegas ou da internet. Use sempre dados originais ou modifique de algum modo para que seus exercícios sejam diferentes.

Alguns links que podem ajudar

Pandas https://github.com/guipsamora/pandas_exercises
<https://pynative.com/python-pandas-exercise/>
<https://www.kaggle.com/python10pm/pandas-75-exercises-with-solutions>
<https://www.w3resource.com/python-exercises/pandas/index.php>
<https://www.machinelearningplus.com/python/101-pandas-exercises-python/>

Datasets

<https://dados.gov.br/home>
<https://archive.ics.uci.edu/datasets/> (procure por dataset menos utilizados, evite os comuns: titanic, iris, diabetes, wine, cars, etc)
<https://www.kaggle.com/datasets>

Obs: se você preferir, pode também fazer scrapping com dados originais de algum tema que seja interessante para você.