



Lisandra Godoi <lysfernandagodoi@gmail.com>

#7DaysOfCode - Python Pandas 2/7: 🧑 Limpeza de dados

1 mensagem

Francisco Foz <contato@7daysofcode.io>
Responder a: contato@alura.com.br
Para: lysfernandagodoi@gmail.com

27 de agosto de 2025 às 08:00



Python Pandas



E aí, Lisandra Fernanda de, conseguiu desenvolver o desafio do primeiro dia e já está tudo pronto para o próximo?

Só pra garantir, vou deixar sempre a minha resolução lá no final deste e-mail, beleza? Assim você pode comparar com a sua, mas não esqueça de **compartilhar comigo as suas resoluções, quero ver também!**

Hoje é o segundo dia do seu desafio #7DaysOfCode com Python Pandas, bora praticar?

Ontem você importou e organizou os dados. Hoje, você irá começar a manipulá-los, ou seja, tirar o que não for necessário, agrupar dados, atribuir novas informações, etc.

Claro, os dados precisam fazer sentido e estar preparados para o contexto da análise a ser realizada. Por isso, a “limpeza” dos dados é uma parte essencial em um projeto de ciência de dados.

Você irá iniciar a limpeza e atribuir mais contexto aos seus dados para depois aprofundar-se nas análises.

A Ciência de Dados é uma área interdisciplinar que abrange programação, matemática/estatística e conhecimento do negócio. Seu objetivo é extrair de dados, informações úteis que agregam valor e resolvem problemas. Entender o negócio é fundamental para que não se tenha análises sem explicações ou mesmo sem o foco necessário para a resolução dos problemas.

Você deve ter visto que tem uma coluna identificada como “localização” e diversos números nela, mas você sabe o que significam estes números?

"Os itens do acervo em uma biblioteca são organizados por um sistema de classificação de acordo com o respectivo tema. Existem diversos sistemas, mas este conjunto está de acordo com a CDU - Classificação Decimal Universal. Esta classificação é decimal, pois varia de acordo com a classe de cada assunto:

- 000 a 099: Generalidades. Ciência e conhecimento.
- 100 a 199: Filosofia e psicologia.
- 200 a 299: Religião.
- 300 a 399: Ciências sociais.
- 400 a 499: Classe vaga. Provisoriamente não ocupada.
- 500 a 599: Matemática e ciências naturais.
- 600 a 699: Ciências aplicadas.
- 700 a 799: Belas artes.
- 800 a 899: Linguagem. Língua. Linguística.
- 900 a 999: Geografia. Biografia. História."

Portanto, se um material tiver um código de localização 720, ele está dentro da classe geral de “**Belas Artes**”; ou se tiver um código 028, estará dentro da classe geral de “**Generalidades. Ciência e**

conhecimento”.

Para isso, crie uma nova coluna com os valores da localização, para refletir a respectiva classe geral na CDU.

Você precisará ainda excluir alguns dados e modificar outros.

A coluna *"registro_sistema"*, por exemplo, não está fazendo sentido para essa análise, por isso você pode excluí-la.

Já a coluna da matrícula (*"matricula_ou_siape"*) não está com um formato muito legível. Transforme-a em formato *String*.

DICA

Para mapear os valores, você pode usar uma lógica condicional de acordo com a numeração e armazená-los em uma lista.

EXTRA

Dê uma olhada nesse texto da Alura sobre como adicionar elementos em uma lista.

RESPOSTA DO EXERCÍCIO DO DIA 1

– Solução do Dia 1

Bom desafio, te vejo amanhã nesse mesmo horário!

Ah, lembre-se de compartilhar o seu código no seu GitHub e nas suas redes sociais com a hashtag **#7DaysOfCode**.

Francisco Foz

Bibliotecário e Analista de Dados

Alura, Vila Mariana, São Paulo, SP

[Não quero mais receber os Desafios](#)