

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 Universa ([- 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."](https://empresas.alura.com.br/e3t/Ctc/I8+113/d2z6gD04/VWX6ZJ3mHDBKV73CSB78R_5rVZjc4S5BKYd3N7jDJw03l5QzW7Y8-PT6lZ3lXW7QgkrL7kbvv7W7ttZfx3h-rVCW7RBFxv243HzzW2WGcTr86Tk1kw5mx_c124M5VhW5y1p2H1HdHsnW4W4BKb952G3PW1vSQg95bT_VZW8hzt7274WcT1W6t7JV15JRfKnW16D0Vb65cyV_MRXB1NVqs-fW94RTgM3l6gKZN6RHw_Czlh12N5DDPRRwR28qVwsfCQ2d07L8N4h99hvpN1X7W2Qm9GR1zXb4hW6HgZJ11Pf49yW1V3mcn1wz0YDW702gNb3bg900W2X2SP170Ng-cW1F3J7q7N6Rx9W1d8KcF1y5P78VP1_3m2-PMsqVJMfZZ9bNKtpf4vP46x04) l. Esta classificação é decimal, pois varia de acordo com a classe de cada assunto:</p></div><div data-bbox=)

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

(https://empresas.alura.com.br/e3t/Ctc/I8+113/d2z6gD04/VWX6ZJ3mHDBKV73CSB78R_5rVZjc4S5BKYd3N7jDJWj3l5QzW8wLKSr6lZ3q6W1Yfbzq4JrScMW8jG3_64cJcZ7W7GkJMZ4WXkNsw3bb0-q88rzGkw6SLMW92fQMTyW8-jxSy6BQTd2W7bZV-J6L0HQrW3j1S757Ls1g4W4P9rJ51qWs8kV6SV2_856-ZfW2hy-QD6_F71LW7YYGW24s1PyhW4Q5YkX45Kn9dVZcZcz6HR100MDtYph9S68qw2CZgjp4Rhg0cw1hR9ZZ3WJf-HVhCPdp2pw91-N4Qhz7T3Rpv6W26vF-Y7_NG04N6xsPYNwhMcwW6P_Mz12rpwDFW4v7vFb59QH75W7FD54j7zz6XSW8_xMyg4NtbCdW68qWrf7rgV-lW6f_9wt8PYkjVW6rBcX95qCnmZf3NdCYH04) sobre como adicionar elementos em uma lista.

RESPOSTA DO EXERCÍCIO DO DIA 1

- Solução do Dia 1

(https://empresas.alura.com.br/e3t/Ctc/I8+113/d2z6gD04/VWX6ZJ3mHDBKV73CSB78R_5rVZjc4S5BKYd3N7jDJWj3l5QzW95jsWP6lZ3ltVR3HnM91xVf_W5m2gmN40KxFSW6RbQcL1DFx1FW2dLK4P3xvhJZW8tZYZS2rYmspw78NcLX3V78mcw5X4kwx2-tKmMW8fVgWY7gmzyWVK0TQK6RcXx6VZGP5s8pTC3VN4zk5X6fCqkyW3zTNVw5_TKjyW8KLbMh6V_Pp0W8hxF-D4y3krjW9863-m3b3K76W2K0Tf54r095JVTzm-26NntJZW6HKTZN3BdCXRW56MZ2Q4RxwyGW2nVqqt3HQXrSW8rXQZJ77pG3rW6YNf7f3g6TKKW8P2nRK6vvpBNW7JPNl58gQSp2W6FsgLk7w_8p1VtSWSq4_jJS6W29jnm3fVQdHW3VJpfq7V384JW1ncLlM1rdwm8W2hqmZ98z608Wf8HXKY04)

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 (https://empresas.alura.com.br/hs/preferences-center/pt/direct?data=w2nXS-N30h-BrW3P0X_G43zPWQW4fFJhY4kBlTGW4cv7r238kGXkw3_wpgN3jb7Mkw1W-QS2zKQ2hW252x-N3bfh72W4mbNB63dbd9bw2TRd1x2KPcgNW3P8qmQ2WFwwHW2zKqjG4fFf81W2FtX-41_9NBZW2TydXd43DdHMMW2Cvyf-3z7THwW3MbBG-2-LVbjW3ghkwb2v_TfLW3JVLyL3dkb-dW36xdcC3M0GPRW1Q3KF83_pBkwW3yPMHz2KC89CW3NQBc-1Y_9sQW1VjBnL32xgxWW2Rsy6F2WxK-_W20YJ242-cdYRW1VzQR_30r5JMW2nXqs71VzlrXW2-GM2j3f__SqW1VjcVZ2xXTWLW1QktVy41Sw1MW3DM1L_1S41k1W4kcsL14twJnJW1NsL0d1QrKW4W2-MKYn2KFHKcW45PGV02YMZkSW3NXvkJ3LQ2V0W2-Jppw3yRnXlW2-bjmw3XN4q9W1Y_BYc2PwtqvW2z-_mn2sZW2JW4mm67N2PWpc9W3XvDBK2Ks1VDW2WnnC749J_QHW32q2153ZsJ65W23gR112-bRQ_W4t2Mzj3z9d90W41W1qy2YpT7LW3Hc2K02FBnHMMW3XNy913_XTLCW3_Bb053ZV95fW32pHfP2Pmhj4W4rCpDx2MBznDW4rb1mb3z9-570&_hsenc=p2ANqtz-_R2Fi204n8AGk7DbzZE4RKP7_TpTt9R8JE950v8AZcjckNJ9-OHISP10YHi_1ft0ozZy0-buY60B3iHRm5SezA2y8SNP51mXpA-sYWucJ_EKhohBc&_hsmi=270877722)