

Chegou a hora, Lisandra Fernanda de! Você vai começar a sua jornada no #7DaysOfCode.

Estou muito animado para começar esse desafio! Iremos explorar os dados de empréstimos dos acervos do sistema de bibliotecas da UFRN.

Um dos objetivos de uma biblioteca é garantir que os materiais informacionais estejam sendo utilizados. Os empréstimos realizados podem ser um indicador, mesmo que de forma básica (pois você não consegue garantir que haja uma leitura ou utilização real).

Por este motivo, entender a quantidade de empréstimos se torna importante.

Questões de diferentes perspectivas podem surgir como:

- A quantidade de empréstimos está aumentando ou diminuindo ao decorrer dos últimos anos?
- Em quais bibliotecas do sistema estão a maior quantidade de empréstimos?
- Quais são os temas mais emprestados? E os menos?

Com estas e outras informações será possível entender o cenário e apresentá-lo à diretoria das bibliotecas, para que possam tomar melhores decisões na melhoria da infraestrutura, dos recursos e processos da unidade de informação.

Mas para que tudo isso seja realizado, você precisará começar com a coleta e organização dos dados para que possa trabalhar com eles nas próximas análises.

Borá lá?!

Você trabalhará com dados apenas dos últimos 10 anos disponíveis. Por isso, importe para seu Jupyter Notebook (https://empresas.alura.com.br/e3t/Ctc/I8+113/d2z6gD04/VW1y1M7Qc0PqW1NK6r16812C-W4Rs6st5BHPxQN4Tr4NC3l5QzW7Y8-PT6lZ3p1W557CMd6VFHGNVdcx3S6HHJ6jW99bLG88GMSQ1W8yq2GP3NzdQ6W5CMGP_3P0g4YW6Yz5b-69cN9xV5JKDK54bF29N1zX6hcRxL-_W8rFP1G54WspFW14zgz54sCbxVW5ZQD9R9hXcqHW3l8NhT65npd1W6qGmCN6dxxXxW1-_pH6wwVD_W2STjr63WvXf7N7-bcG_d7JrMVW-MVD4KwpRfW6lcZsW4w9MvN1F7h2yTFgp1W8dxxDQ4J0CbRVZ9Qd85Lq7LHW4d1k8q5Lm8BFW7G8_hf7mFJymW8DHZr1hzt6RVyLVX4k12d_VMZKtz2jPf8Pf7PFtNK04) os dados de:

- Empréstimos dos acervos das bibliotecas de UFRN
(https://empresas.alura.com.br/e3t/Ctc/I8+113/d2z6gD04/VW1y1M7Qc0PqW1NK6r16812C-W4Rs6st5BHPxQN4Tr4Mq5jGjMW5BWr2F6lZ3m3V6jG1b8fxmfHV8_Hz_40vbRMW8MBP-K827S3GVfghMG3-3lkmW8N0rHr1FVQqTN1fcdHgNywBRW84Z0RV3D0rZvW5nm_m16f4Z8GN24qtdh1zd6hW8bxt2l1DB5gsW7LyX4s2qhF1zV4lM2r8H4s3PW4l1fkM5-qG1QV-vPnj8J-8C9W1L4c471FkRnPW55R94Y19BHG2VXPxY94cYDxGN7wgQjsbFNT7W8KhvqZ4CXpN1W8F15NX1pkkG3W8l4Gjc68bG_SW3vtflQ8JyHDJW1WSg4D7VyqNPVVZjfd7kYhZsN4zj-8ljlGg0W5xLpsF4MKqgfw4wqLqJ1rgN1TW1C3RQt6Zf1XQW4HxsDt1z3NZBW95RY4Y3Py7sNN30RPm_7vN7GW4cj0xr4C3nTSW4yVH1Y2zj5txW106c6Z37hGdMf1LDzF-04)

- Exemplares do acervo
(https://empresas.alura.com.br/e3t/Ctc/I8+113/d2z6gD04/VW1y1M7Qc0PqW1NK6r16812C-W4Rs6st5BHPxQN4Tr4Mq5jGjMW5BWr2F6lZ3pPN6sVh5m16vhrW8jPCx01gB8f7W86-YlJ6v5ljXW3l14Qv3f1RTgN7-HQ05M0DLMW3mDpPP4LjgtvW54MW6Z1mcdJZW7TfTtV7_Bv15W32SrwS5VJrtgW8M6hny8vLmY2VLS3L68Wls7_W6RSc1C1Qw5wkW4c62M55R1XcMW6Hr4Dk3WvNdCVVj0X37zH5nLW5r5Y5j93_mCGW5tt6tK3WxgWrW5f5Zwj44rfD5N6dm8k7DTPF2W4yjp7X7VK5JMW7xH6F197sbNYW7PTgJK4R_CPMW1nMDKK4N1_Y2W2xsBnv8WH4_fw2svtZd86kTgzW3YHz0f2h1FG1W2vCx4J6-NvmVW6DJbXm7_69m-VxwVzK4-TNfHN9lgSNxz-BBZW94s0nz3NK0DJW7h5p442ZCG92W3_Mhb55qwwBbW5KckFd92Mjg5f6-YYLg04)

Dados baixados? Ok, mas são diversas tabelas diferentes e isso dificulta o trabalho. Portanto, o seu primeiro passo é unificar em um único Dataframe todos os dados pertinentes para a análise.

Comece pelos empréstimos e você terá os dados das transações. Depois, mescle com os dados do acervo, para que você possa entender, por exemplo, de qual biblioteca era o material emprestado ou a qual tema ele se referia. Elas se relacionam pela coluna de código de barras de cada material.

Lembre-se que é muito comum receber dados nulos ou duplicados, por isso não deixe de fazer a limpeza.

DICA

Você pode importar os dados diretamente do Github para seu notebook apenas passando o endereço do link “Raw” como origem.

Confira a documentação

(https://empresas.alura.com.br/e3t/Ctc/I8+113/d2z6gD04/VW1y1M7Qc0PqW1NK6r16812C-W4Rs6st5BHPxQN4Tr4Nj3l5QzW7lCdLW6lZ3mnW3bhWzB5PHjgrW5FCf1V5qdY23W3RPjn91-1D36W4b4S5Qg6m200nVrBRvt5C36_pN1vhP7M5CwRjW8-D5Nx3TWMDmW7P83cs7XbmrMw8GCsSf5hD3wtN4TM0Xmd18kVW961v445qLF1fw46XP_N3s3V1HW355Mh55LbXc0N96dLw7HgQ7nW783YJN6cd6q4W4hrC132WXNRPW3QglGj36T154W5wrjNV1lJqmvW7tb06Y93R-6HW75R5Vj4JwZ7nW1_XGSb4Rvs93W1-rs2J8pJH26W3q93Mv8hKmt8W58DV7D3FbyY8f8FGr2W04) do Pandas das diferentes formas de entrada de dados.

EXTRA

Deixei este artigo da Alura

(https://empresas.alura.com.br/e3t/Ctc/I8+113/d2z6gD04/VW1y1M7Qc0PqW1NK6r16812C-W4Rs6st5BHPxQN4Tr4Nj3l5QzW7lCdLW6lZ3ksW3l5b0J2VWZzcN1_Z_l6kJDPhN4F1w6th4HR9N5kGg8RFSZDcW3H75Dq4gwsdnW5lKXFn87_qp5V6ymMN5h72FrW2Cp0r27C15_9W3G78712px51bV3Pg928bDLbcVPcL1K7T2gm1VZNfv13kB6ypW50Ynks4bj2jXVwmWfV8QB0vgVJXvjd3yMmmdW321sF01P1Df8W7-D46f4bkP0jW1wj1z65cfpmLW3HVVkl8tMcqJN3pJdbSQchPXW2Yn39W3tQ_JRVpRgK46k4xrrw1t7HJC43TR6lW76F0HM8th0zGf3GmC1Y04) explicando um pouco mais sobre o formato de arquivo Apache Parquet e também esse meu texto (https://empresas.alura.com.br/e3t/Ctc/I8+113/d2z6gD04/VW1y1M7Qc0PqW1NK6r16812C-W4Rs6st5BHPxQN4Tr4NW3l5QzW8wLKSr6lZ3nSW6wKRFY22D05GW1NBvc53vBZ5SW5k1t_42sW8jGW32xpF147-2RLW1RYPgG3J-Sq5W656Jht2wmbftW33bFWl1wLrK8W4mTDqQ1xGQSYW5lWVhN4GDf2PW3RB_1Y6PxQ0YW8PWNNH3PFzgwW4n_jVq906TrjW7VPqYG2c_pyGW2WNNsb7T6R5QW2FgPN81sxjPXVbfyD7501QT2V4F2z-79Sth2W7p1Ltf5zwb_ZN92KfBmCjkz6VWW4GJ9lLRdPW8WN3Q81wkDZ4W6LJFHN57-V_VW2ZJyKQ1Tn40mVchbgc85PG9BW7mhwnk20cqJ3N1hFj5-jdbJXW6H0PTm56wdJyW3dRxCs8PHXXNf5jR7yK04) que explica como você pode jogar fora os dados usando o Pandas.

Esse foi só o primeiro dia e amanhã a gente se encontra aqui, para mais um #7DaysOfCode de Python Pandas.

Após finalizado, não deixe de postar seus resultados nas redes sociais e me marcar.

Ficarei muito feliz de ver o seu trabalho!

Boa jornada!

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-SmW4hq3Sf21kxN_W2Kq_Mh1VrL7FW4mtNLP36pdbNW1Sqj8P255t92W1Xx1D92Hyy6FW2TSxkr49Q4MsW3VS4MH2q_V40W34prxl4tFJkWW4hv6F53MbH8FW2TQQZx2-bjWxw4fsN3W30twcmW2FBq9w4pyNP7W4mGM7V2Ky16jW4thCbY23ry3DW2MSjtZ3VXhfGW3NPR2v2PtYJ_W3ZD2KY232cHTW3jwFtt32lgDtW25mBLx3Fh2w1W2nGgvz3ZTq5dW2B3bhV2sDbKGW3VRDRw3yRSNBW4pDtTp2CKDgWw1QdVft3yWFM9W1_8vZ12YgHW2W4kbZ4h49xY-4W3S_dYL3z98L1W1Sc3gg4t6qxcW4mDFbD3K2tfGW1QdWvd2G0dSpW4hmxWz4htQT4W30ncs12vQtFvW1SwF612C-kDnW3XXYy21BGg_RW1XwGqb2WQK30W30tXkv1V8lNBW3Q_nCf2RCTV5W1X0G0Q3ZTQPXW2MGkL51Qwv2vW213P1B4kdHdhw2FzBkl2sQGv2W3BXg5b41GZ2CW3z2-Hm3j9Qn-W2zKtv12xWc4hW2CD9mn1W_Lq-W3272Bd3yMHcyW3ZttrC3DSgp7W3P0V0s2qCB4xW3z2vW4ptc98W1L95NZ3gpcKSW34vRGf4tcQFS0&_hsenc=p2ANqtz--88aSXBozI80x1vM2TTYv0lgP4yiaY0Mp1-OAgPFaWvo3-A9WMQRtv08frQXYd418PDgI4nMuVe96FLOr5UVQPeh16hgXC3zPQCxQkoLrn449ZCqI&_hsmi=270874190)