1. Escolha 1 site de sua preferência e realize web scraping com a biblioteca BeautifulSoup, formando uma base de dados salva em arquivo csv.

Sugestão de problemas:

Site IMDB

Preços de produtos

Sites de notícias

Meteorologia, etc...

* Anexar arquivo python (.py) e base de dados formada.
* Atividade individual

Formar base de dados com mais de uma característica, por exemplo:

(PREÇO, DESCRIÇÃO, PRODUTO),

(NOME DO FILME, DIRETOR, AVALIAÇÃO)

(TEMPERATURA MAX, TEMPERATURA MIN)

**Resposta do 1º quesito:**

**O arquivo do código utilizado e o arquivo gerado estão em anexo desta atividade, mas posso adiantar que foquei em capturar dados de preços de produtos de cadeiras de escritórios. Para isso, eu utilizei o site da Amazon e coletei vários dados a respeito no meu arquivo gerado, como o nome da cadeira, preço completo, juros, e a quantidade de estrelas das avaliações.**

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

2. Selecione uma base de dados em um dos Sites **Repositórios** indicados a seguir.

Definir:

* Cenário de Dados (Saúde, Educação, Transporte, Segurança...)
* Um problema ou uma pergunta/hipótese relevantes que pode ser respondida a partir dos dados:

Ex: "A partir dos dados podemos saber quais bairros registram maior número de multas por excesso de velocidade"

* Tipo e escala das variáveis:

Qualitativo (ordinal, nominal), Quantitativo (contínuo, discreto)

* Link para a base de dados

**Resposta do 2º quesito:**

**2.1 Cenário de dados: Saúde (World Health Organization)**

**2.2 Um problema ou uma pergunta/hipótese relevantes que pode ser respondida a partir dos dados: A partir dos dados podemos saber quais são os países com maiores expectativas de vida para ambos ou sexos ou um sexo específico**

**2.3 Tipos e escala das variáveis: Quantitativo com variáveis contínuas ( Ex: 75.9)**

**2.4: Link para a base de dados:** [**https://www.who.int/data/gho/data/indicators/indicator-details/GHO/life-expectancy-at-birth-(years)**](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/life-expectancy-at-birth-(years))

**( segundo link:** [**https://data.who.int/countries/076**](https://data.who.int/countries/076)**)**

Tabela

Descrição gerada automaticamente

Código usaddo para o web scraping:

import csv

import requests

from bs4 import BeautifulSoup

# URL da página que você deseja fazer scraping

url = "https://www.amazon.com.br/s?k=cadeira+de+escritorio&crid=399EITZ668JMX&sprefix=cadeira+%2Caps%2C186&ref=nb\_sb\_ss\_ts-doa-p\_1\_8"

# Conexão: Enviar uma solicitação GET para a URL

response = requests.get(url)

# Verifique se a solicitação foi bem-sucedida (status 200)

if response.status\_code == 200:

    # Parse a página com o BeautifulSoup

    soup = BeautifulSoup(response.text, "html.parser")

    # Encontre os elementos HTML que contêm os títulos de produtos e preços

    products = soup.find\_all("span", class\_="a-size-base-plus a-color-base a-text-normal")

    prices = soup.find\_all("span", class\_="a-price-whole")

    prices\_fraction = soup.find\_all("span", class\_="a-price-fraction")

    juros = soup.find\_all("span", class\_="a-size-base a-color-secondary")

    estrelas = soup.find\_all("span", class\_="a-size-base s-underline-text")

    # Cria o arquivo CSV com codificação UTF-8

    file = open('precos\_cadeiraEscritorio.csv', 'w', newline='')

    writer = csv.writer(file)

    headers = ['Produto', 'Preço', 'Juros', 'Estrelas']  # Adicionando uma coluna para "Preço"

    writer.writerow(headers)

    # Loop pelos elementos e imprimir os títulos e preços

    for product, price, fraction, juro, estrela in zip(products, prices, prices\_fraction, juros, estrelas):

        product\_text = product.text.strip()

        juros\_text = juro.text.strip()

        price\_text = price.text.strip() + fraction.text.strip()  # Concatena preço e fração

        estrelas = estrela.text.strip()

        print("Produto:", product\_text)

        print("Preco:", price\_text)

        print("Juros:", juros\_text)

        print("Estrelas:", estrelas)

        # Cada produto e preço

        row = [product\_text, price\_text, juros\_text, estrelas]

        # Salva os dados no arquivo CSV

        writer.writerow(row)

    file.close()

else:

    print("Falha ao acessar a página:", response.status\_code)