**Q1. What is Web Scraping? Why is it Used? Give three areas where Web Scraping is used to get data.**

Web scraping is the process of extracting data from websites by using automated software tools. It is used to extract structured information from websites and convert it into a structured format such as a spreadsheet or a database.

Web scraping is used in various areas such as:

* E-commerce: Online retailers use web scraping to extract data about their competitors’ products and prices, which they use to adjust their own prices.
* Marketing and Sales: Marketers use web scraping to extract data from social media sites and online forums to monitor customer sentiment and to identify trends.
* Research and Development: Researchers use web scraping to collect data on various topics such as scientific papers, patents, and clinical trials, which they use for analysis and to gain insights.

In general, web scraping is used to gather data that can be used for analysis, research, and decision-making purposes.

# Q2. What are the different methods used for Web Scraping?

There are different methods that can be used for web scraping, including:

* Using programming languages such as Python, PHP, or Ruby to write custom scripts that scrape data from websites.
* Using web scraping software tools such as BeautifulSoup, Scrapy, or Selenium, which automate the process of web scraping.
* Using APIs provided by websites to access and retrieve data in a structured format.
* Using browser extensions such as Web Scraper, Data Miner, or Octoparse to extract data from websites.
* Using specialized web scraping services that offer data extraction and scraping as a service.

Each method has its own advantages and limitations, and the choice of method depends on the complexity of the data to be scraped, the scale of the project, and the technical expertise of the scraper

# Q3. What is Beautiful Soup? Why is it used?

Beautiful Soup is a Python library used for web scraping purposes to pull the data out of HTML and XML files. It provides ways to navigate, search, and modify a parsed HTML or XML document. Beautiful Soup parses the HTML file in a tree-like structure so that we can extract the information in a more hierarchical and more readable manner. Beautiful Soup can be used to extract specific data or to extract data from the entire HTML file. It's commonly used in data mining and data extraction from the web.

# Q4. Why is flask used in this Web Scraping project?

Flask is a lightweight web framework that is commonly used for building web applications, including web scraping applications. It can be used to develop a RESTful API that allows users to interact with the scraped data through HTTP requests. Flask provides a simple and flexible way to create an API, and it can easily integrate with other Python libraries that are commonly used for web scraping, such as Beautiful Soup and Requests. In addition, Flask provides support for a variety of web servers, making it easy to deploy and run a web scraping application on different platforms.

# Q5. Write the names of AWS services used in this project. Also, explain the use of each service.

Here's an explanation of AWS CodePipeline and AWS Elastic Beanstalk:

* AWS CodePipeline: It is a fully managed continuous delivery service that helps to automate the build, test, and deployment process for applications. It supports multiple integration points with a variety of source code repositories, build systems, and deployment services, including AWS Elastic Beanstalk, AWS CodeDeploy, and Amazon S3. In this project, CodePipeline is used to automate the deployment of the Flask app.
* AWS Elastic Beanstalk: It is a fully managed service for deploying and scaling web applications and services. It provides a platform for developers to deploy applications in different languages such as Python, Java, PHP, Ruby, and more. Elastic Beanstalk automatically handles the deployment, scaling, monitoring, and maintenance of the application. In this project, Elastic Beanstalk is used to deploy the Flask app and host it on the web.