

# Loading Selenium Page Source into Beautiful Soup

---



**Pratheerth Padman**  
FREELANCE DATA SCIENTIST



# Module Overview



**Introduction to Beautiful Soup**

**Selenium to Beautiful Soup**

**Demo: Scraping info from a dynamic  
webpage**



# Introduction to Beautiful Soup

---

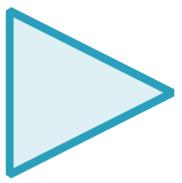


# Beautiful Soup

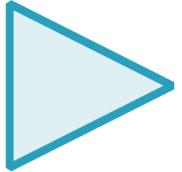
Beautiful Soup is a Python library for pulling data out of HTML and XML files.



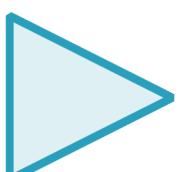
# Why Beautiful Soup?



**Simple and pythonic**



**No need to worry about encodings**



**Multiple parsers**



# Types of HTML Parsers

Python's HTML  
Parser

LXML's HTML  
Parser

HTML5 Lib



# Objects in Beautiful Soup

**Tag**

**NavigableString**

**Comment**

**BeautifulSoup**

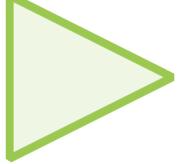


# Transitioning from Selenium to Beautiful Soup

---



# Why Transition?



**Selenium has everything needed to extract info from HTML**



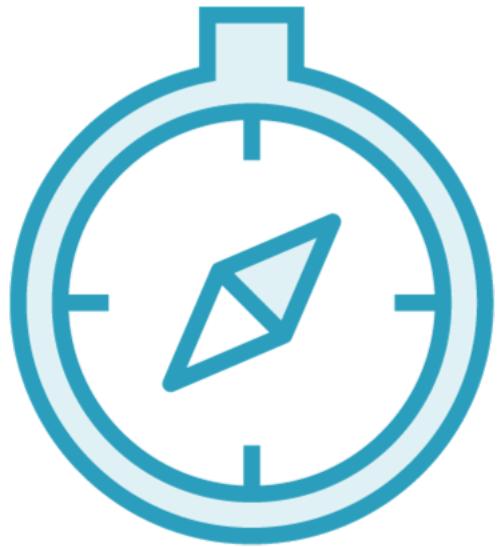
**Problem - it is slow**



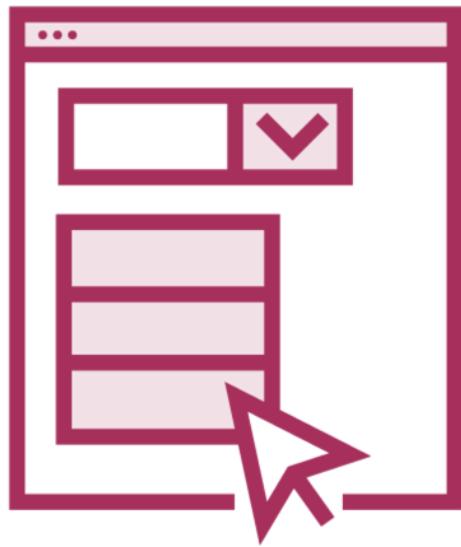
**Faster to let beautiful soup parse page source**



# Transition Process Flow



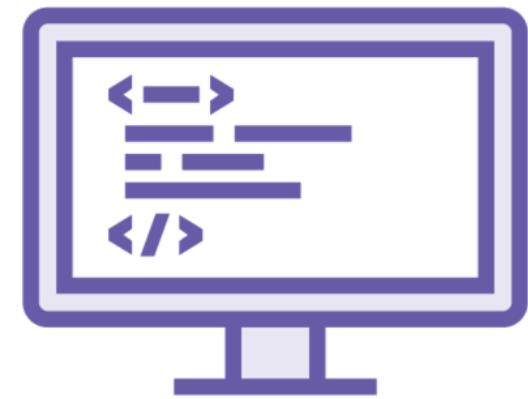
Navigate to  
required  
website



Perform  
necessary  
browser actions



Get page source  
info



Pass page  
source info to  
b.s for parsing



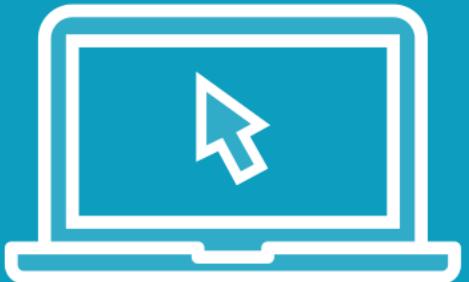
# Demo



**Demo: Combining Python, Selenium and Beautiful Soup to scrape a dynamic webpage - Introduction**



# Demo



**Demo: Combining Python, Selenium and Beautiful Soup to scrape a dynamic webpage - Practical implementation**



# Summary



**Got introduced to Beautiful Soup**

**Reasons for using Beautiful Soup**

**Parsers and objects in Beautiful Soup**

**Extracting data from a dynamic website**

