





Data Science Unit 4

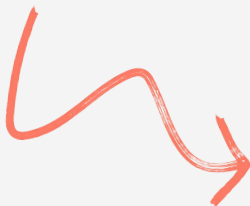
Accessing Data From the Internet



Before we start...

- Make sure you are comfortable
- Have water and maybe a strong coffee handy
- If you need a break... take it!
- If you need a stretch – please go ahead!
- Please mute yourselves if you are not talking
- Have your video on at all times

...and let's get started!





In this session we will...



1. **Define** APIs and Web Scraping
2. **Investigate** APIs and see how they work
3. **Build** a web scraper using BeautifulSoup.



multiverse

Web Services and API





What is an API?





Web API





Famous APIs





Famous APIs





Making API Calls



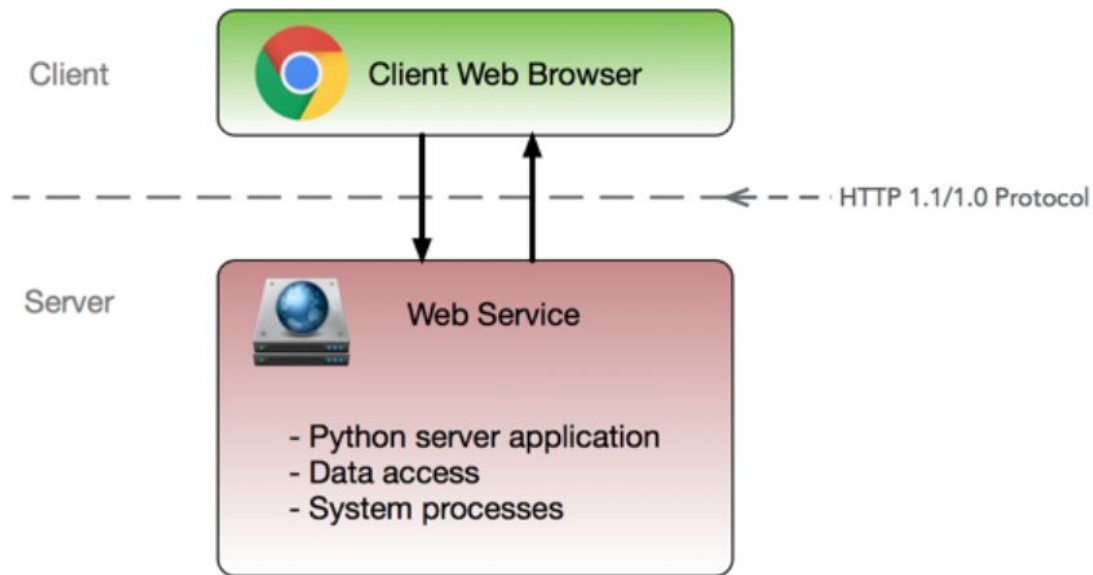


HTTP





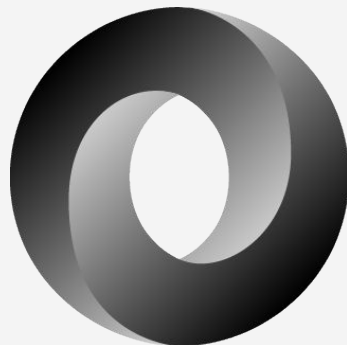
HTTP





JSON

JavaScript



A large, expressive teal brushstroke graphic in the top-left corner, consisting of several overlapping, diagonal strokes that create a sense of movement and energy.

multiverse

Let's Practice



multiverse

Web Scraping





BeautifulSoup





HTML

Appli.ed

powered by WhiteHat

Alastair Tyson

[Dashboard](#)[Objectives](#)[My Apprentices](#)[Progress Reviews](#)[Evidence Upload](#)

[Customise this page](#)[Manage dashboards](#)

WhiteHat Analytics

- Delivery Metrics
- Delivery Attendance
- Delivery Key Dates
- Community Metrics

Administration

- Site administration

Course Content & Assignments

DF 1: Data Analysis in Industry	0%
DF 2: Database Fundamentals	0%
DF 3: Big Data	0%
DF 4: Advanced Analytics	-
DF: BCS Data Analysis Concepts	-
DF: DA Bootcamp	-
DF: DS Bootcamp	-

Assignment Reminders

You have no assignments that require your attention

Elements Console Sources Network Performance

```
<nav role="navigation" class="navbar navbar-default navbar-sit">...</nav>
<!-- Breadcrumb and edit buttons -->
<div class="container-fluid breadcrumb-container">
  ::before
  <div class="row">
    ::before
    <div class="col-sm-12">
      <header id="page-header" class="clearfix">
        ::before
        <div class="clearfix" id="page-navbar">...</div> == $0
      <div id="course-header"></div>
    </div>
  </div>
  ::after
</div>
</div>
<!-- Content -->
<div id="page" class="container-fluid">...</div>
```

html #page-totara-dashboards-8 div div #page-header div#page-na

Styles Computed Event Listeners DOM Breakpoints Properties Acce

Filter

element.style {

* {

box-sizing: border-box;

}

div {

display: block;

}

Inherited from body#page-to...

body {

background-color: #ffffff;

font-family: 'Roboto', sans-serif;

font-size: 13px;



HTML



<div>

<h1>

<a>

<p>



Along came a Spider...

```
# convert this into a soup object  
soup = BeautifulSoup(html.text, 'html.parser')
```

```
# print the restaurant names  
soup.find_all(name='span', attrs={'class': 'rest-row-name-text'})
```

```
# for each element you find, print out the restaurant name  
for entry in soup.find_all(name='span', attrs={'class': 'rest-row-name-text'}):  
    print(entry.text)
```



Warning



A large, expressive teal brushstroke graphic in the top-left corner, consisting of several overlapping, diagonal strokes that create a sense of movement and energy.

multiverse

Let's Practice