

BIG DATA Y PYTHON

MÁSTER EN BIG DATA

201020

GABRIEL MARÍN DÍAZ

EJERCICIOS

Ejercicio

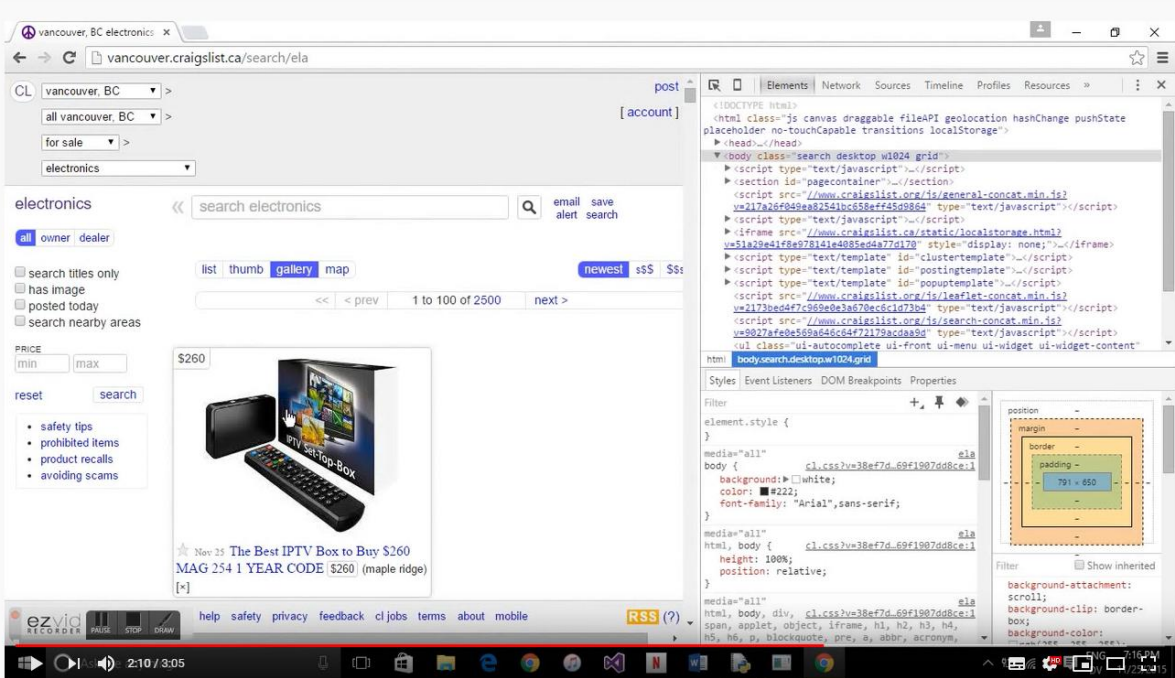
**OBTENER LA SERIE HISTÓRICA DE COVID 19
EN FORMATO CSV, INTENTAR OBTENER EL
INCREMENTO DIARIO DE CASOS
DIAGNOSTICADOS, HOSPITALIZACIONES,
UCI, FALLECIMIENTOS. LLEVAR EL
RESULTADO A UN FORMATO EXCEL Y
REPRESENTAR LOS DATOS EN GRÁFICOS**

Webscrapping

Revisar el siguiente vídeo... hacer lo mismo con una página web de compras en español.

← → ↺ 🔒 <https://www.youtube.com/watch?v=bhYulVzYRng>

≡ YouTube how to web scrape with python (selenium) 🔍



The screenshot shows a web browser window with the address bar displaying <https://www.youtube.com/watch?v=bhYulVzYRng>. The browser is displaying a YouTube video titled "How to Web Scrape with Python (Selenium/ChromeDriver)". The video player shows a browser window with the Craigslist website open, displaying search results for "electronics" in Vancouver, BC. The browser's developer tools are open, showing the HTML structure of the page. The video player controls at the bottom indicate the video is 2:10 / 3:05 long, has 1239 likes, and 18 comments. The video is by Charles Clayton, who has 3310 subscribers. A "SUSCRIBIRSE" button is visible.

How to Web Scrape with Python (Selenium/ChromeDriver)

88.827 visualizaciones · 26 nov. 2015

Charles Clayton
3310 suscriptores

SUSCRIBIRSE

Click me!

Uso de APIs

EJERCICIO 1

Vamos a mejorar el ejercicio de escucha en twitter...

- Introduzcamos como parámetro de escucha “COVID-19”.
- Escuchemos durante un período de tiempo razonable (30 minutos).
- El resultado obtenido lo dirigiremos a un fichero JSON (“escucha_covid.json”).
- Utilizando la librería de Python Pandas, vamos a ir creando un estudio de los tweets generados: ubicación del tweet, texto, fuente del tweet (source), número de respuestas que ha recibido el tweet,... y todo lo que se os ocurra para obtener estadísticas, esta parte la trabajaremos esta semana.

Es importante que reviséis la estructura de un tweet, de ahí podréis obtener toda la información y que trabajéis la biblioteca Pandas para análisis de datos.

!Muchas Gracias!

GABRIEL MARÍN DÍAZ
LCDO. CIENCIAS FÍSICAS UCM

www.linkedin.com/in/gabrielmarindiaz/