Goodmorning!

Mining Instagram + Creating an InstaBot

Katelyn Suhr

ksuhr@ucsc.edu

Instagram Background

About Instagram

- A photo and video-sharing social networking platform
- Owned by Facebook, Inc
- Includes photo filters, tagging, locations, likes, public/private profiles, stories, and much more
- 1 billion users as of May 2019

About Instagram Developer API

- Instagram deprecated their API Platform to protect user's privacy and security
- Introduced Instagram Graph API for accessing data in large and medium sized Instagram Business Accounts
- Must find another way to scrape data!

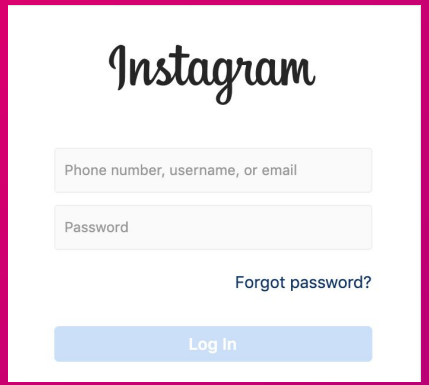
Use Selenium for Automation

- Selenium automates browsers
  - Need to import webdriver object from Selenium to control Chrome
- Desired Functionalities
  - Sign In
  - Turn Off Notifications
  - Follow a user
  - Unfollow a user
  - Get a user’s followers
  - Get data for images

Use Chrome Developer Tools

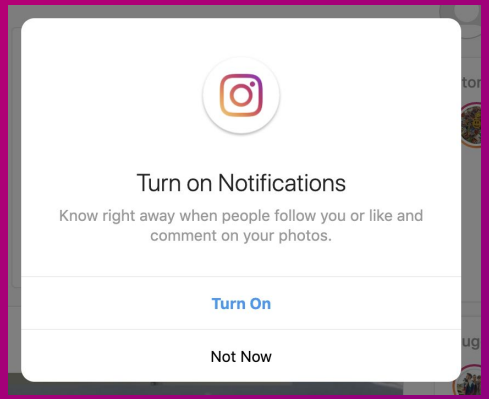
Sign In

- Inspect Sign In Page and find <input class> names for username and password
- Use find_elements_by_css_selector()



Turn Off Notifications

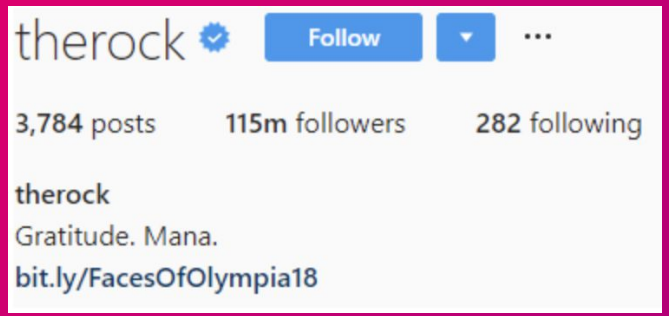
- Inspect page again and find “Not Now” button in HTML
- Click the button



Use Chrome Developer Tools

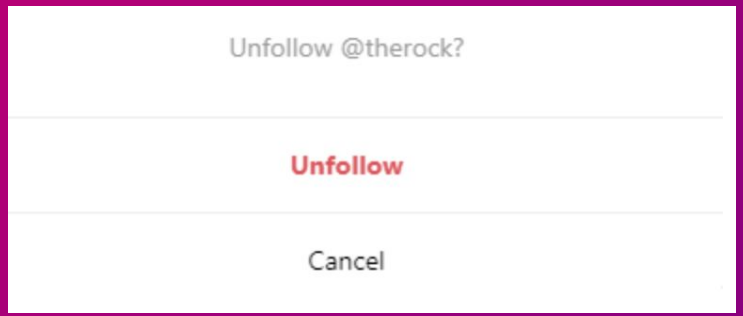
Follow A User

- Check it you are already following
- The first button of the page is “Following”
- Use find_elements_by_css_selector() to find button and click it



Unfollow A User

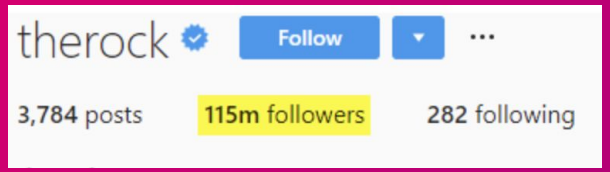
- Find first button of the page and check if we are “Following”
- Click on button & modal will pop
- Find first button using XPath selector and click it



Use Chrome Developer Tools, Selenium, BeautifulSoup, Pandas

Get A User's Followers

- Navigate to "x followers", click it and a modal will pop
- Use selenium to press SPACE bar and keep scrolling down list until we've reached max users
- Use find_elements_by_css_selector() to select each user and grab their url link by finding <a> <href> tag



Get Data For Image

- Go to Profile and grab top 12 image links
- Open source page for each picture and use BeautifulSoup to parse
- Extract the JSON script and parse through it to get photo metadata
- Put it all in a Pandas DataFrame and plot what we find

Challenges

- Parsing through JSON object and extracting data individually
- Working with JSON dictionary keys that change
- Infinite Scrolling to get more photos
- Handling different types of posts: Image, Video, Slider Image

Sources

- https://medium.com/@mottet.dev/lets-create-an-instagram-bot-to-show-you-the-power-of-selenium-349d7a6744f7

- https://medium.com/@srujana.rao2/scraping-instagram-with-python-using-selenium-and-beautiful-soup-8b72c186a058