

swiggydatascape

August 15, 2024

```
[ ]: !pip install selenium  
!apt-get update  
!apt install chromium-chromedriver  
!cp /usr/lib/chromium-browser/chromedriver /usr/bin
```

Collecting selenium

Downloading selenium-4.23.1-py3-none-any.whl.metadata (7.1 kB)

Requirement already satisfied: urllib3<3,>=1.26 in

/usr/local/lib/python3.10/dist-packages (from urllib3[socks]<3,>=1.26->selenium) (2.0.7)

Collecting trio~=0.17 (from selenium)

Downloading trio-0.26.2-py3-none-any.whl.metadata (8.6 kB)

Collecting trio-websocket~=0.9 (from selenium)

Downloading trio_websocket-0.11.1-py3-none-any.whl.metadata (4.7 kB)

Requirement already satisfied: certifi>=2021.10.8 in

/usr/local/lib/python3.10/dist-packages (from selenium) (2024.7.4)

Requirement already satisfied: typing_extensions~=4.9 in

/usr/local/lib/python3.10/dist-packages (from selenium) (4.12.2)

Requirement already satisfied: websocket-client~=1.8 in

/usr/local/lib/python3.10/dist-packages (from selenium) (1.8.0)

Requirement already satisfied: attrs>=23.2.0 in /usr/local/lib/python3.10/dist-packages (from trio~=0.17->selenium) (24.2.0)

Requirement already satisfied: sortedcontainers in

/usr/local/lib/python3.10/dist-packages (from trio~=0.17->selenium) (2.4.0)

Requirement already satisfied: idna in /usr/local/lib/python3.10/dist-packages (from trio~=0.17->selenium) (3.7)

Collecting outcome (from trio~=0.17->selenium)

Downloading outcome-1.3.0.post0-py2.py3-none-any.whl.metadata (2.6 kB)

Requirement already satisfied: sniffio>=1.3.0 in /usr/local/lib/python3.10/dist-packages (from trio~=0.17->selenium) (1.3.1)

Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-packages (from trio~=0.17->selenium) (1.2.2)

Collecting wsproto>=0.14 (from trio-websocket~=0.9->selenium)

Downloading wsproto-1.2.0-py3-none-any.whl.metadata (5.6 kB)

Requirement already satisfied: pysocks!=1.5.7,<2.0,>=1.5.6 in

/usr/local/lib/python3.10/dist-packages (from urllib3[socks]<3,>=1.26->selenium) (1.7.1)

Collecting h11<1,>=0.9.0 (from wsproto>=0.14->trio-websocket~=0.9->selenium)

```

    Downloading h11-0.14.0-py3-none-any.whl.metadata (8.2 kB)
Downloading selenium-4.23.1-py3-none-any.whl (9.4 MB)
      9.4/9.4 MB
40.5 MB/s eta 0:00:00
Downloading trio-0.26.2-py3-none-any.whl (475 kB)
      476.0/476.0 kB
19.1 MB/s eta 0:00:00
Downloading trio_websocket-0.11.1-py3-none-any.whl (17 kB)
Downloading wsproto-1.2.0-py3-none-any.whl (24 kB)
Downloading outcome-1.3.0.post0-py2.py3-none-any.whl (10 kB)
Downloading h11-0.14.0-py3-none-any.whl (58 kB)
      58.3/58.3 kB
3.3 MB/s eta 0:00:00
Installing collected packages: outcome, h11, wsproto, trio, trio-
websocket, selenium
Successfully installed h11-0.14.0 outcome-1.3.0.post0 selenium-4.23.1
trio-0.26.2 trio-websocket-0.11.1 wsproto-1.2.0
Get:1 https://cloud.r-project.org/bin/linux/ubuntu jammy-cran40/ InRelease
[3,626 B]
Hit:2 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64
InRelease
Get:3 http://security.ubuntu.com/ubuntu jammy-security InRelease [129 kB]
Hit:4 http://archive.ubuntu.com/ubuntu jammy InRelease
Get:5 http://archive.ubuntu.com/ubuntu jammy-updates InRelease [128 kB]
Ign:6 https://r2u.stat.illinois.edu/ubuntu jammy InRelease
Get:7 https://r2u.stat.illinois.edu/ubuntu jammy Release [5,713 B]
Get:8 https://r2u.stat.illinois.edu/ubuntu jammy Release.gpg [793 B]
Get:9 http://archive.ubuntu.com/ubuntu jammy-backports InRelease [127 kB]
Get:10 http://security.ubuntu.com/ubuntu jammy-security/main amd64 Packages
[2,171 kB]
Get:11 http://archive.ubuntu.com/ubuntu jammy-updates/restricted amd64 Packages
[2,963 kB]
Get:12 http://security.ubuntu.com/ubuntu jammy-security/universe amd64 Packages
[1,133 kB]
Get:13 http://security.ubuntu.com/ubuntu jammy-security/restricted amd64
Packages [2,867 kB]
Get:14 https://r2u.stat.illinois.edu/ubuntu jammy/main amd64 Packages [2,553 kB]
Get:15 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 Packages
[1,423 kB]
Get:16 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 Packages [2,447
kB]
Get:17 http://archive.ubuntu.com/ubuntu jammy-backports/universe amd64 Packages
[33.7 kB]
Hit:18 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy InRelease
Hit:19 https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu jammy
InRelease
Hit:20 https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu jammy InRelease
Fetched 16.0 MB in 5s (2,961 kB/s)

```

```

Reading package lists... Done
W: Skipping acquire of configured file 'main/source/Sources' as repository
'https://r2u.stat.illinois.edu/ubuntu jammy InRelease' does not seem to provide
it (sources.list entry misspelt?)
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  apparmor chromium-browser libfuse3-3 liblzo2-2 libudev1 snapd squashfs-tools
  systemd-hwe-hwdb
  udev
Suggested packages:
  apparmor-profiles-extra apparmor-utils fuse3 zenity | kdialog
The following NEW packages will be installed:
  apparmor chromium-browser chromium-chromedriver libfuse3-3 liblzo2-2 snapd
  squashfs-tools
  systemd-hwe-hwdb udev
The following packages will be upgraded:
  libudev1
1 upgraded, 9 newly installed, 0 to remove and 45 not upgraded.
Need to get 28.5 MB of archives.
After this operation, 118 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 apparmor amd64
3.0.4-2ubuntu2.3 [595 kB]
Get:2 http://archive.ubuntu.com/ubuntu jammy/main amd64 liblzo2-2 amd64
2.10-2build3 [53.7 kB]
Get:3 http://archive.ubuntu.com/ubuntu jammy/main amd64 squashfs-tools amd64
1:4.5-3build1 [159 kB]
Get:4 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libudev1 amd64
249.11-0ubuntu3.12 [78.2 kB]
Get:5 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 udev amd64
249.11-0ubuntu3.12 [1,557 kB]
Get:6 http://archive.ubuntu.com/ubuntu jammy/main amd64 libfuse3-3 amd64
3.10.5-1build1 [81.2 kB]
Get:7 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 snapd amd64
2.63+22.04ubuntu0.1 [25.9 MB]
Get:8 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 chromium-
browser amd64 1:85.0.4183.83-0ubuntu2.22.04.1 [49.2 kB]
Get:9 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 chromium-
chromedriver amd64 1:85.0.4183.83-0ubuntu2.22.04.1 [2,308 B]
Get:10 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 systemd-hwe-
hwdb all 249.11.5 [3,228 B]
Fetched 28.5 MB in 1s (20.2 MB/s)
Preconfiguring packages ...
Selecting previously unselected package apparmor.
(Reading database ... 123594 files and directories currently installed.)
Preparing to unpack .../apparmor_3.0.4-2ubuntu2.3_amd64.deb ...
Unpacking apparmor (3.0.4-2ubuntu2.3) ...

```

```

Selecting previously unselected package liblzo2-2:amd64.
Preparing to unpack .../liblzo2-2_2.10-2build3_amd64.deb ...
Unpacking liblzo2-2:amd64 (2.10-2build3) ...
Selecting previously unselected package squashfs-tools.
Preparing to unpack .../squashfs-tools_1%3a4.5-3build1_amd64.deb ...
Unpacking squashfs-tools (1:4.5-3build1) ...
Preparing to unpack .../libudev1_249.11-0ubuntu3.12_amd64.deb ...
Unpacking libudev1:amd64 (249.11-0ubuntu3.12) over (249.11-0ubuntu3.10) ...
Setting up libudev1:amd64 (249.11-0ubuntu3.12) ...
Selecting previously unselected package udev.
(Reading database ... 123802 files and directories currently installed.)
Preparing to unpack .../udev_249.11-0ubuntu3.12_amd64.deb ...
Unpacking udev (249.11-0ubuntu3.12) ...
Selecting previously unselected package libfuse3-3:amd64.
Preparing to unpack .../libfuse3-3_3.10.5-1build1_amd64.deb ...
Unpacking libfuse3-3:amd64 (3.10.5-1build1) ...
Selecting previously unselected package snapd.
Preparing to unpack .../snapd_2.63+22.04ubuntu0.1_amd64.deb ...
Unpacking snapd (2.63+22.04ubuntu0.1) ...
Setting up apparmor (3.0.4-2ubuntu2.3) ...
Created symlink /etc/systemd/system/sysinit.target.wants/apparmor.service →
/lib/systemd/system/apparmor.service.
Setting up liblzo2-2:amd64 (2.10-2build3) ...
Setting up squashfs-tools (1:4.5-3build1) ...
Setting up udev (249.11-0ubuntu3.12) ...
invoke-rc.d: could not determine current runlevel
invoke-rc.d: policy-rc.d denied execution of start.
Setting up libfuse3-3:amd64 (3.10.5-1build1) ...
Setting up snapd (2.63+22.04ubuntu0.1) ...
Created symlink /etc/systemd/system/multi-
user.target.wants/snapd.apparmor.service →
/lib/systemd/system/snapd.apparmor.service.
Created symlink /etc/systemd/system/multi-
user.target.wants/snapd.autoimport.service →
/lib/systemd/system/snapd.autoimport.service.
Created symlink /etc/systemd/system/multi-user.target.wants/snapd.core-
fixup.service → /lib/systemd/system/snapd.core-fixup.service.
Created symlink /etc/systemd/system/multi-user.target.wants/snapd.recovery-
chooser-trigger.service → /lib/systemd/system/snapd.recovery-chooser-
trigger.service.
Created symlink /etc/systemd/system/multi-user.target.wants/snapd.seeded.service
→ /lib/systemd/system/snapd.seeded.service.
Created symlink /etc/systemd/system/cloud-
final.service.wants/snapd.seeded.service →
/lib/systemd/system/snapd.seeded.service.
Unit /lib/systemd/system/snapd.seeded.service is added as a dependency to a non-
existent unit cloud-final.service.
Created symlink /etc/systemd/system/multi-user.target.wants/snapd.service →

```

```

/lib/systemd/system/snapd.service.
Created symlink /etc/systemd/system/timers.target.wants/snapd.snap-repair.timer
→ /lib/systemd/system/snapd.snap-repair.timer.
Created symlink /etc/systemd/system/sockets.target.wants/snapd.socket →
/lib/systemd/system/snapd.socket.
Created symlink /etc/systemd/system/final.target.wants/snapd.system-
shutdown.service → /lib/systemd/system/snapd.system-shutdown.service.
Selecting previously unselected package chromium-browser.
(Reading database ... 124032 files and directories currently installed.)
Preparing to unpack .../chromium-
browser_1%3a85.0.4183.83-0ubuntu2.22.04.1_amd64.deb ...
=> Installing the chromium snap
==> Checking connectivity with the snap store
===> System doesn't have a working snapd, skipping
Unpacking chromium-browser (1:85.0.4183.83-0ubuntu2.22.04.1) ...
Selecting previously unselected package chromium-chromedriver.
Preparing to unpack .../chromium-
chromedriver_1%3a85.0.4183.83-0ubuntu2.22.04.1_amd64.deb ...
Unpacking chromium-chromedriver (1:85.0.4183.83-0ubuntu2.22.04.1) ...
Selecting previously unselected package systemd-hwe-hwdb.
Preparing to unpack .../systemd-hwe-hwdb_249.11.5_all.deb ...
Unpacking systemd-hwe-hwdb (249.11.5) ...
Setting up systemd-hwe-hwdb (249.11.5) ...
Setting up chromium-browser (1:85.0.4183.83-0ubuntu2.22.04.1) ...
update-alternatives: using /usr/bin/chromium-browser to provide /usr/bin/x-www-
browser (x-www-browser) in auto mode
update-alternatives: using /usr/bin/chromium-browser to provide /usr/bin/gnome-
www-browser (gnome-www-browser) in auto mode
Setting up chromium-chromedriver (1:85.0.4183.83-0ubuntu2.22.04.1) ...
Processing triggers for udev (249.11-0ubuntu3.12) ...
Processing triggers for hicolor-icon-theme (0.17-2) ...
Processing triggers for libc-bin (2.35-0ubuntu3.4) ...
/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_5.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc_proxy.so.2 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_opencl.so.0 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libur_loader.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_level_zero.so.0 is not a
symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc.so.2 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbb.so.12 is not a symbolic link

```

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_0.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind.so.3 is not a symbolic link

Processing triggers for man-db (2.10.2-1) ...

Processing triggers for dbus (1.12.20-2ubuntu4.1) ...

cp: '/usr/lib/chromium-browser/chromedriver' and '/usr/bin/chromedriver' are the same file

```
[ ]: import csv
from bs4 import BeautifulSoup
import time
from selenium import webdriver
from selenium.webdriver.chrome.service import Service
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.chrome.options import Options

# Set up Chrome options
chrome_options = Options()
chrome_options.add_argument('--headless')
chrome_options.add_argument('--no-sandbox')
chrome_options.add_argument('--disable-dev-shm-usage')

[ ]: # Function to scrape listing elements from the Swiggy website
def scrape_listings(soup):
    return soup.select('a.RestaurantList__RestaurantAnchor-sc-1d3nl43-3')

# Function to scrape the restaurant name from a listing
def scrape_restaurant_name(listing):
    name_element = listing.select_one('div.sc-beySbM.iKLEMo')
    return name_element.text.strip() if name_element else None

# Function to scrape the image URL from a listing
def scrape_image_url(listing):
    img_element = listing.select_one('div.sc-jsEegt img')
    return img_element['src'] if img_element else None

# Function to scrape the rating from a listing
def scrape_rating(listing):
    rating_element = listing.select_one('span.sc-beySbM.jAtOQO')
    return rating_element.text.strip().split('.')[0].strip() if rating_element
    ↪ else None

# Function to scrape the cuisine type from a listing
```

```

def scrape_cuisine(listing):
    cuisine_element = listing.select('div.
↳sw-restaurant-card-descriptions-container div.sc-beySbM.bRTXBF')
    return cuisine_element[0].text.strip() if cuisine_element else None

# Function to scrape the location from a listing
def scrape_location(listing):
    location_element = listing.select('div.
↳sw-restaurant-card-descriptions-container div.sc-beySbM.bRTXBF')
    return location_element[1].text.strip() if len(location_element) > 1 else
↳None

# Function to scrape the delivery time from a listing
def scrape_delivery_time(listing):
    delivery_time_element = listing.select_one('div.sc-beySbM.brneVe')
    if delivery_time_element:
        split_text = delivery_time_element.text.strip().split('.')
        if len(split_text) > 1:
            return split_text[1].strip()
        else:
            return split_text[0].strip() # Return the first element if no
↳separator is found
    return None

```

```

[ ]: !apt-get update
!apt install firefox-geckodriver
!pip install selenium webdriver_manager

import csv
from bs4 import BeautifulSoup
import time
from selenium import webdriver
from selenium.webdriver.firefox.options import Options as FirefoxOptions
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC

# Set up Firefox options
firefox_options = FirefoxOptions()
firefox_options.add_argument("--headless")

# Define your scraping functions here (scrape_listings, scrape_restaurant_name,
↳etc.)

def main():
    # Set up Selenium WebDriver for Firefox
    driver = webdriver.Firefox(options=firefox_options)

```

```

# Navigate to the Swiggy website URL
url = 'https://www.swiggy.com/city/bangalore'
driver.get(url)

# Open a CSV file to write the restaurant data
with open('swiggy_data.csv', 'w', newline='', encoding='utf-8') as csv_file:
    fieldnames = ['restaurant_name', 'image_url', 'rating', 'cuisine', 'location', 'delivery_time']
    writer = csv.DictWriter(csv_file, fieldnames=fieldnames)

    # Write the header
    writer.writeheader()

    while True:
        # Wait for the page to load
        time.sleep(5)

        # Get the page source and create a BeautifulSoup object
        soup = BeautifulSoup(driver.page_source, 'html.parser')

        # Scrape restaurant listings
        listings = scrape_listings(soup)

        # Iterate through each listing and extract restaurant information
        for listing in listings:
            restaurant_name = scrape_restaurant_name(listing)
            image_url = scrape_image_url(listing)
            rating = scrape_rating(listing)
            cuisine = scrape_cuisine(listing)
            location = scrape_location(listing)
            delivery_time = scrape_delivery_time(listing)

            # Write the restaurant information to the CSV file
            writer.writerow({
                'restaurant_name': restaurant_name,
                'image_url': image_url,
                'rating': rating,
                'cuisine': cuisine,
                'location': location,
                'delivery_time': delivery_time
            })

        # Check if there's a "Show More" button
        try:
            show_more_button = WebDriverWait(driver, 10).until(

```



```

        EC.element_to_be_clickable((By.CSS_SELECTOR, "div.
↪RestaurantList__ShowMoreContainer-sc-1d3nl43-0"))
    )
    driver.execute_script("arguments[0].click();", show_more_button)
except:
    # If there's no "Show More" button, we've reached the end
    break

    # Close the browser
    driver.quit()

    print("Data extraction complete. Check the 'swiggy_data.csv' file.")

# Run the main function
if __name__ == "__main__":
    main()

# Download the CSV file
from google.colab import files
files.download('swiggy_data.csv')

```

```

Hit:1 https://cloud.r-project.org/bin/linux/ubuntu jammy-cran40/ InRelease
Hit:2 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64
InRelease
Hit:3 http://archive.ubuntu.com/ubuntu jammy InRelease
Ign:4 https://r2u.stat.illinois.edu/ubuntu jammy InRelease
Hit:5 https://r2u.stat.illinois.edu/ubuntu jammy Release
Hit:6 http://security.ubuntu.com/ubuntu jammy-security InRelease
Hit:7 http://archive.ubuntu.com/ubuntu jammy-updates InRelease
Hit:9 http://archive.ubuntu.com/ubuntu jammy-backports InRelease
Hit:10 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy InRelease
Hit:11 https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu jammy
InRelease
Hit:12 https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu jammy InRelease
Reading package lists... Done
W: Skipping acquire of configured file 'main/source/Sources' as repository
'https://r2u.stat.illinois.edu/ubuntu jammy InRelease' does not seem to provide
it (sources.list entry misspelt?)
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
Package firefox-geckodriver is not available, but is referred to by another
package.
This may mean that the package is missing, has been obsoleted, or
is only available from another source
However the following packages replace it:
firefox

```

E: Package 'firefox-geckodriver' has no installation candidate

Requirement already satisfied: selenium in /usr/local/lib/python3.10/dist-packages (4.23.1)

Requirement already satisfied: webdriver_manager in /usr/local/lib/python3.10/dist-packages (4.0.2)

Requirement already satisfied: urllib3<3,>=1.26 in /usr/local/lib/python3.10/dist-packages (from urllib3[socks]<3,>=1.26->selenium) (2.0.7)

Requirement already satisfied: trio~=0.17 in /usr/local/lib/python3.10/dist-packages (from selenium) (0.26.2)

Requirement already satisfied: trio-websocket~=0.9 in /usr/local/lib/python3.10/dist-packages (from selenium) (0.11.1)

Requirement already satisfied: certifi>=2021.10.8 in /usr/local/lib/python3.10/dist-packages (from selenium) (2024.7.4)

Requirement already satisfied: typing_extensions~=4.9 in /usr/local/lib/python3.10/dist-packages (from selenium) (4.12.2)

Requirement already satisfied: websocket-client~=1.8 in /usr/local/lib/python3.10/dist-packages (from selenium) (1.8.0)

Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from webdriver_manager) (2.32.3)

Requirement already satisfied: python-dotenv in /usr/local/lib/python3.10/dist-packages (from webdriver_manager) (1.0.1)

Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from webdriver_manager) (24.1)

Requirement already satisfied: attrs>=23.2.0 in /usr/local/lib/python3.10/dist-packages (from trio~=0.17->selenium) (24.2.0)

Requirement already satisfied: sortedcontainers in /usr/local/lib/python3.10/dist-packages (from trio~=0.17->selenium) (2.4.0)

Requirement already satisfied: idna in /usr/local/lib/python3.10/dist-packages (from trio~=0.17->selenium) (3.7)

Requirement already satisfied: outcome in /usr/local/lib/python3.10/dist-packages (from trio~=0.17->selenium) (1.3.0.post0)

Requirement already satisfied: sniffio>=1.3.0 in /usr/local/lib/python3.10/dist-packages (from trio~=0.17->selenium) (1.3.1)

Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-packages (from trio~=0.17->selenium) (1.2.2)

Requirement already satisfied: wsproto>=0.14 in /usr/local/lib/python3.10/dist-packages (from trio-websocket~=0.9->selenium) (1.2.0)

Requirement already satisfied: pysocks!=1.5.7,<2.0,>=1.5.6 in /usr/local/lib/python3.10/dist-packages (from urllib3[socks]<3,>=1.26->selenium) (1.7.1)

Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->webdriver_manager) (3.3.2)

Requirement already satisfied: h11<1,>=0.9.0 in /usr/local/lib/python3.10/dist-packages (from wsproto>=0.14->trio-websocket~=0.9->selenium) (0.14.0)

Data extraction complete. Check the 'swiggy_data.csv' file.

<IPython.core.display.Javascript object>

<IPython.core.display.Javascript object>