

```
In [ ]: ▶ # Ask:  
# Build a web application that scrapes various websites for data related  
# the information in a single HTML page. The following outlines what you
```

```
In [1]: ▶ # !pip install splinter  
# !pip install flask_pymongo  
# !pip install pymongo
```

```
In [2]: ▶ # import libraries and requirements  
import time  
import requests  
import pymongo  
import pandas as pd  
from splinter import Browser  
from bs4 import BeautifulSoup  
from selenium import webdriver
```

Nasa Mars News

```
In [ ]: ▶ # Step 1 - Scraping  
# Complete your initial scraping using Jupyter Notebook, BeautifulSoup, Flask  
# Create a Jupyter Notebook file called mission_to_mars.ipynb and use the  
# analysis tasks. The following outlines what you need to scrape.
```

```
In [3]: ▶ # executable_path = {'executable_path': 'chromedriver.exe'}  
# browser = Browser('chrome', **executable_path, headless=False)  
  
executable_path = {'executable_path': '/usr/bin/chromedriver'}  
browser = Browser('chrome', **executable_path, headless=False)
```

```
In [6]: ▶ mars_data = {}  
hemisphere_image_urls = []  
  
news_url = 'https://mars.nasa.gov/news/'  
browser.visit(news_url)  
time.sleep(1)  
  
html = browser.html  
soup = BeautifulSoup(html, "html.parser")
```

In [5]: `soup.head()`

```
Out[5]: [<script src="//api-public.addthis.com/url/shares.json?url=http%3A%2F%2Fmars.nasa.gov%2Fnews%2F%3Fpage%3D0%26per_page%3D40%26order%3Dpublish_date%2Bdesc%252Ccreated_at%2Bdesc%26search%3D%26category%3D19%252C165%252C184%252C204%26blank_scope%3DLatest&callback=_ate.cbs.rcb_7zn10" type="text/javascript"></script>,
  <script src="//www.reddit.com/api/info.json?url=http%3A%2F%2Fmars.nasa.gov%2Fnews%2F%3Fpage%3D0%26per_page%3D40%26order%3Dpublish_date%2Bdesc%252Ccreated_at%2Bdesc%26search%3D%26category%3D19%252C165%252C184%252C204%26blank_scope%3DLatest&jsonp=_ate.cbs.rcb_clkx0" type="text/javascript"></script>,
  <script src="//graph.facebook.com/?id=http%3A%2F%2Fmars.nasa.gov%2Fnews%2F%3Fpage%3D0%26per_page%3D40%26order%3Dpublish_date%2Bdesc%252Ccreated_at%2Bdesc%26search%3D%26category%3D19%252C165%252C184%252C204%26blank_scope%3DLatest&fields=og_object%7Bengagement%7D&callback=_ate.cbs.rcb_azym0" type="text/javascript"></script>,
  <script src="//api-public.addthis.com/url/shares.json?url=https%3A%2F%2Fmars.nasa.gov%2Fnews%2F%3Fpage%3D0%26per_page%3D40%26order%3Dpublish_date%2Bdesc%252Ccreated_at%2Bdesc%26search%3D%26category%3D19%252C165%252C184%252C204%26blank_scope%3DLatest&callback=_ate.cbs.rcb_9okk0" type="text/javascript"></script>]
```

```
In [6]: # Scrape the NASA Mars News Site and collect the latest News Title and Paragraph
# Assign the text to variables that you can reference later.
news_soup = BeautifulSoup(html, 'html.parser')
result = news_soup.find('div', class_='content_title')
news_title = result.next_element.get_text()
result1 = news_soup.find('div', class_='article_teaser_body')
news_p = result1.get_text()

mars_data["news_title"] = news_title
mars_data["news_p"] = news_p
mars_data
```

```
Out[6]: {'news_title': 'Things Are Stacking up for NASA's Mars 2020 Spacecraft',
'news_p': 'As the July 2020 launch date inches closer, the next spacecraft headed to the Red Planet is assembled for more testing.'}
```

```
In [ ]: # JPL Mars Space Images - Featured Image
# Visit the url for JPL Featured Space Image here.
# Use splinter to navigate the site and find the image url for the current image
# and assign the url string to a variable called featured_image_url.
# Make sure to find the image url to the full size .jpg image.
# Make sure to save a complete url string for this image.
```

```
In [7]: image_url = 'https://www.jpl.nasa.gov/spaceimages/?search=&category=Mars'
        browser.visit(image_url)
        time.sleep(1)

        html = browser.html
        image_soup = BeautifulSoup(html, "html.parser")
        image = image_soup.find('div', class_='carousel_items')
        image_url = image.article['style']
        url = image_url.split('/s')[-1].split('.')[0]
        featured_image_url = 'https://www.jpl.nasa.gov' + '/s' + url + '.jpg'

        mars_data["featured_image_url"] = featured_image_url
        mars_data
```

```
Out[7]: {'news_title': 'Things Are Stacking up for NASA's Mars 2020 Spacecraft',
         'news_p': 'As the July 2020 launch date inches closer, the next spacecraft headed to the Red Planet is assembled for more testing.',
         'featured_image_url': 'https://www.jpl.nasa.gov/spaceimages/images/wallpaper/PIA17470-1920x1200.jpg'}
```

```
In [ ]: # Mars Weather
        # Visit the Mars Weather twitter account here and scrape the latest Mars
        # Save the tweet text for the weather report as a variable called mars_weather
```

```
In [8]: weather_url = 'https://twitter.com/marswxreport?lang=en'
        browser.visit(weather_url)
        time.sleep(1)
        html = browser.html

        weather_soup = BeautifulSoup(html, 'html.parser')
        weather = weather_soup.find('div', class_='js-tweet-text-container')

        mars_weather = weather.p.text
        mars_data["mars_weather"] = mars_weather
        mars_data
```

```
Out[8]: {'news_title': 'Things Are Stacking up for NASA's Mars 2020 Spacecraft',
         'news_p': 'As the July 2020 launch date inches closer, the next spacecraft headed to the Red Planet is assembled for more testing.',
         'featured_image_url': 'https://www.jpl.nasa.gov/spaceimages/images/wallpaper/PIA17470-1920x1200.jpg',
         'mars_weather': 'InSight sol 141 (2019-04-20) low -98.3°C (-144.9°F) high -19.7°C (-3.5°F)\nwinds from the SW at 4.7 m/s (10.6 mph) gusting to 12.9 m/s (28.8 mph)\npressure at 7.40 hPa pic.twitter.com/CQr1QQt3cM'}
```

```
In [ ]: # Mars Facts
        # Visit the Mars Facts webpage here (https://space-facts.com/mars/) and
        # including Diameter, Mass, etc.
        # Use Pandas to convert the data to a HTML table string.
```

```
In [9]: facts_url = 'http://space-facts.com/mars/'  
tables = pd.read_html(facts_url)  
tables
```

```
Out[9]: [ 0 1  
0 Equatorial Diameter: 6,792 km  
1 Polar Diameter: 6,752 km  
2 Mass: 6.42 x 10^23 kg (10.7% Earth)  
3 Moons: 2 (Phobos & Deimos)  
4 Orbit Distance: 227,943,824 km (1.52 AU)  
5 Orbit Period: 687 days (1.9 years)  
6 Surface Temperature: -153 to 20 °C  
7 First Record: 2nd millennium BC  
8 Recorded By: Egyptian astronomers]
```

```
In [10]: df = tables[0]  
df.columns = ['Mars_planet_profile', 'Value']  
df
```

```
Out[10]:
```

	Mars_planet_profile	Value
0	Equatorial Diameter:	6,792 km
1	Polar Diameter:	6,752 km
2	Mass:	6.42 x 10 ²³ kg (10.7% Earth)
3	Moons:	2 (Phobos & Deimos)
4	Orbit Distance:	227,943,824 km (1.52 AU)
5	Orbit Period:	687 days (1.9 years)
6	Surface Temperature:	-153 to 20 °C
7	First Record:	2nd millennium BC
8	Recorded By:	Egyptian astronomers

```
In [11]: ▶ mars_facts = df.to_dict('records')
Table = []
for i in range(0, len(mars_facts)):
    temp = list(mars_facts[i].values())
    Table.append(temp)
mars_data["mars_facts"] = Table
mars_data
```

```
Out[11]: {'news_title': 'Things Are Stacking up for NASA's Mars 2020 Spacecraft',
'news_p': 'As the July 2020 launch date inches closer, the next spacecraft headed to the Red Planet is assembled for more testing.',
'featured_image_url': 'https://www.jpl.nasa.gov/spaceimages/images/wallpaper/PIA17470-1920x1200.jpg',
'mars_weather': 'InSight sol 141 (2019-04-20) low -98.3°C (-144.9°F) high -19.7°C (-3.5°F)\nwinds from the SW at 4.7 m/s (10.6 mph) gusting to 12.9 m/s (28.8 mph)\npresure at 7.40 hPa',
'mars_facts': [['Equatorial Diameter:', '6,792 km'],
['Polar Diameter:', '6,752 km'],
['Mass:', '6.42 x 10^23 kg (10.7% Earth)'],
['Moons:', '2 (Phobos & Deimos)'],
['Orbit Distance:', '227,943,824 km (1.52 AU)'],
['Orbit Period:', '687 days (1.9 years)'],
['Surface Temperature:', '-153 to 20 °C'],
['First Record:', '2nd millennium BC'],
['Recorded By:', 'Egyptian astronomers']]}
```

```
In [ ]: ▶ # Mars Hemispheres
# Visit the USGS Astrogeology site here to obtain high resolution images
# You will need to click each of the links to the hemispheres in order to
# resolution image.
# Save both the image url string for the full resolution hemisphere image
# the hemisphere name. Use a Python dictionary to store the data using the
# Append the dictionary with the image url string and the hemisphere title
# one dictionary for each hemisphere.
```

```
In [12]: ▶ executable_path = {"executable_path": "chromedriver"}
browser = Browser("chrome", **executable_path, headless=False)

url = "https://astrogeology.usgs.gov/search/results?q=hemisphere+enhanced"
browser.visit(url)

html = browser.html
soup = BeautifulSoup(html, "html.parser")

h3s = soup.find_all("h3")
```

```
In [13]: titles = []
for h3 in h3s:
    h3 = str(h3)
    h3 = h3[4:-14]
    titles.append(h3)
titles
```

```
Out[13]: ['Cerberus Hemisphere',
'Schiaparelli Hemisphere',
'Syrtis Major Hemisphere',
'Valles Marineris Hemisphere']
```

```
In [14]: img_urls = []
for title in titles:
    browser.click_link_by_partial_text(title)

    html = browser.html
    soup = BeautifulSoup(html, "html.parser")

    img_urls.append(soup.find("div", class_="downloads").find("a")["href"])
img_urls
```

```
Out[14]: ['http://astropedia.astrogeology.usgs.gov/download/Mars/Viking/cerberus_
enhanced.tif/full.jpg',
'http://astropedia.astrogeology.usgs.gov/download/Mars/Viking/schiapare
lli_unenhanced.tif/full.jpg',
'http://astropedia.astrogeology.usgs.gov/download/Mars/Viking/syrtis_ma
jor_unenhanced.tif/full.jpg',
'http://astropedia.astrogeology.usgs.gov/download/Mars/Viking/valles_ma
rineris_unenhanced.tif/full.jpg']
```

```
In [ ]: # 'http://astropedia.astrogeology.usgs.gov/download/Mars/Viking/cerberus_
# 'http://astropedia.astrogeology.usgs.gov/download/Mars/Viking/schiapare
# 'http://astropedia.astrogeology.usgs.gov/download/Mars/Viking/syrtis_m
# 'http://astropedia.astrogeology.usgs.gov/download/Mars/Viking/valles_m
```

```
In [15]: hemisphere_image_urls = []
for title, img_url in zip(titles, img_urls):
    hemisphere_image_urls.append({"title": title, "img_url":img_url})

hemisphere_image_urls
```

```
Out[15]: [{'title': 'Cerberus Hemisphere',
'img_url': 'http://astropedia.astrogeology.usgs.gov/download/Mars/Viki
ng/cerberus_enhanced.tif/full.jpg'},
{'title': 'Schiaparelli Hemisphere',
'img_url': 'http://astropedia.astrogeology.usgs.gov/download/Mars/Viki
ng/schiaparelli_unenhanced.tif/full.jpg'},
{'title': 'Syrtis Major Hemisphere',
'img_url': 'http://astropedia.astrogeology.usgs.gov/download/Mars/Viki
ng/syrtis_major_unenhanced.tif/full.jpg'},
{'title': 'Valles Marineris Hemisphere',
'img_url': 'http://astropedia.astrogeology.usgs.gov/download/Mars/Viki
ng/valles_marineris_unenhanced.tif/full.jpg'}]
```

```
In [18]: ➤ mars_data['hemi_urls'] = hemisphere_image_urls
mars_data
```

```
Out[18]: {'news_title': 'Things Are Stacking up for NASA's Mars 2020 Spacecraft',
'news_p': 'As the July 2020 launch date inches closer, the next spacecr
aft headed to the Red Planet is assembled for more testing.',
'featured_image_url': 'https://www.jpl.nasa.gov/spaceimages/images/wall
paper/PIA17470-1920x1200.jpg',
'mars_weather': 'InSight sol 141 (2019-04-20) low -98.3°C (-144.9°F) hi
gh -19.7°C (-3.5°F)\nwinds from the SW at 4.7 m/s (10.6 mph) gusting to
12.9 m/s (28.8 mph)\npresure at 7.40 hPa',
'mars_facts': [['Equatorial Diameter:', '6,792 km'],
['Polar Diameter:', '6,752 km'],
['Mass:', '6.42 x 10^23 kg (10.7% Earth)'],
['Moons:', '2 (Phobos & Deimos)'],
['Orbit Distance:', '227,943,824 km (1.52 AU)'],
['Orbit Period:', '687 days (1.9 years)'],
['Surface Temperature:', '-153 to 20 °C'],
['First Record:', '2nd millennium BC'],
['Recorded By:', 'Egyptian astronomers']],
'hemi_urls': [{'title': 'Cerberus Hemisphere',
'img_url': 'http://astropedia.astrogeology.usgs.gov/download/Mars/Vik
ing/cerberus_enhanced.tif/full.jpg'},
{'title': 'Schiaparelli Hemisphere',
'img_url': 'http://astropedia.astrogeology.usgs.gov/download/Mars/Vik
ing/schiaparelli_unenhanced.tif/full.jpg'},
{'title': 'Syrtis Major Hemisphere',
'img_url': 'http://astropedia.astrogeology.usgs.gov/download/Mars/Vik
ing/syrtis_major_unenhanced.tif/full.jpg'},
{'title': 'Valles Marineris Hemisphere',
'img_url': 'http://astropedia.astrogeology.usgs.gov/download/Mars/Vik
ing/valles_marineris_unenhanced.tif/full.jpg'}]}
```

```
In [ ]: ➤
```