Module 12 Challenge

Deliverable 1: Scrape Titles and Preview Text from Mars News

In [15]:
```
# Import Splinter and BeautifulSoup
from splinter import Browser
from bs4 import BeautifulSoup
```

In [16]:
```
browser = Browser('chrome')
```

Step 1: Visit the Website

1. Use automated browsing to visit the Mars news site. Inspect the page to identify which elements to scrape.

   > **Hint** To identify which elements to scrape, you might want to inspect the page by using Chrome DevTools.

In [17]:
```
# Visit the Mars news site
url = 'https://static.bc-edx.com/data/web/mars_news/index.html'
browser.visit(url)
```

Step 2: Scrape the Website

Create a Beautiful Soup object and use it to extract text elements from the website.

In [18]:
```
# Create a Beautiful Soup object
html = browser.html
soup = BeautifulSoup(html, "html.parser")
```

In [19]:
```
 # Print the parser
print(soup)
```

```
<html><head>
<meta charset="utf-8"/>
<meta content="width=device-width, initial-scale=1" name="viewport"/>
<link href="css/bootstrap.min.5.2.2.css" rel="stylesheet" type="text/cs
s"/>
<link href="css/font.css" rel="stylesheet" type="text/css"/>
<link href="css/app.css" rel="stylesheet" type="text/css"/>
<title>News - Mars Exploration Program</title>
</head>
<body>
<div class="col-md-12">
<div class="row">
<nav class="navbar navbar-expand-lg navbar-light fixed-top">
```

In [49]:
```
# Extract all the text elements
elements = soup.find_all('div', class_='list_text')
```

Step 3: Store the Results

Extract the titles and preview text of the news articles that you scraped. Store the scraping results in Python data structures as follows:

- Store each title-and-preview pair in a Python dictionary. And, give each dictionary two keys: `title` and `preview`. An example is the following:

```
{'title': "NASA's MAVEN Observes Martian Light Show Caused by Majo
r Solar Storm",
 'preview': "For the first time in its eight years orbiting Mars,
NASA's MAVEN mission witnessed two different types of ultraviolet
aurorae simultaneously, the result of solar storms that began on A
ug. 27."
}
```

- Store all the dictionaries in a Python list.
- Print the list in your notebook.

In [54]:
```
print(elements)
```

In [54]:
```
print(elements)
```

```
[<div class="list_text">
<div class="list_date">November 9, 2022</div>
<div class="content_title">NASA's MAVEN Observes Martian Light Show Cause
d by Major Solar Storm</div>
<div class="article_teaser_body">For the first time in its eight years or
biting Mars, NASA’s MAVEN mission witnessed two different types of ultrav
iolet aurorae simultaneously, the result of solar storms that began on Au
g. 27.</div>
</div>, <div class="list_text">
<div class="list_date">November 1, 2022</div>
<div class="content_title">NASA Prepares to Say 'Farewell' to InSight Spa
cecraft</div>
<div class="article_teaser_body">A closer look at what goes into wrapping
up the mission as the spacecraft’s power supply continues to dwindle.</di
v>
</div>, <div class="list_text">
<div class="list_date">October 28, 2022</div>
<div class="content_title">NASA and ESA Agree on Next Steps to Return Mar
s Samples to Earth</div>
<div class="article_teaser_body">The agency’s Perseverance rover will est
ablish the first sample depot on Mars.</div>
</div>, <div class="list_text">
<div class="list_date">October 27, 2022</div>
<div class="content_title">NASA's InSight Lander Detects Stunning Meteoro
id Impact on Mars</div>
<div class="article_teaser_body">The agency’s lander felt the ground shak
e during the impact while cameras aboard the Mars Reconnaissance Orbiter
spotted the yawning new crater from space.</div>
</div>, <div class="list_text">
<div class="list date">October 21, 2022</div>
```

In [55]:
```
# Create an empty list to store the dictionaries
news_item = []
```

In [57]:
```
# Loop through the text elements
# Extract the title and preview text from the elements
# Store each title and preview pair in a dictionary
# Add the dictionary to the list
for element in elements:
    title = element.find('div', class_='content_title').text
    #print(title)

    preview = element.find('div', class_='article_teaser_body').text
    #print(preview)
    news_item = {
        'title': title ,
        'preview': preview
    }
    news_items.append(result)
#print(result)
```

In [58]:
```
# Print the list to confirm success
news_items
```

Out[58]:
```
[{'title': "NASA's MAVEN Observes Martian Light Show Caused by Major So
lar Storm",
  'preview': 'For the first time in its eight years orbiting Mars, NAS
A’s MAVEN mission witnessed two different types of ultraviolet aurorae
simultaneously, the result of solar storms that began on Aug. 27.'},
 {'title': "NASA Prepares to Say 'Farewell' to InSight Spacecraft",
  'preview': 'A closer look at what goes into wrapping up the mission a
s the spacecraft’s power supply continues to dwindle.'},
 {'title': 'NASA and ESA Agree on Next Steps to Return Mars Samples to
```