Web Scraping Challenge Homework

Terrence Cummings

DOCUMENTATION

Key Github documents:

1. Mission\_to\_mars.ipynb
   1. Jupyter notebook with Mars data scraping code
2. Scrape\_mars.py
   1. Script containing the Python code from the mission\_to\_mars.ipynb.
   2. Called by “/scrape” route of app.py.
3. App.py
   1. Flask app with “/” route that renders index.html.
   2. “/scrape” route calls scrape\_mars.py script.
4. Index.html (in templates directory)
   1. Main web page
   2. Rendered by app.py
5. Mars\_webpage\_screenshot.pdf
   1. A screenshot of the final webpage.

Key features of webpage (index.html)

1. Jumbotron from Bootstrap
   1. Refresh Data
      1. Initiates a new scrape of all the websites.
      2. Results are put in a Python dictionary and stored in a Mongodb document which is read into the webpage by app.py “/” route redirect.
   2. View Featured Image
      1. Jumbotron background is the featured image from the scrape.
      2. Button opens separate window to view the full size feature image.
2. News Articles
   1. Scrape reads in date, headline, summary, and URL of 40 news articles from the first page of the NASA Mars news site.
   2. Use Bootstrap carousel to provide access to the first 5 articles.
   3. “NASA” word is a hyperlink to the NASA website articles page.
   4. Article title is a hyperlink to the full article on the NASA website.
3. Mars hemispheres
   1. Uses a Bootstrap carousel to access the 4 hemispheres
   2. Title of each carousel slide is a hyperlink to the full size image of the hemisphere
4. Current Mars weather
   1. As required, and additionally includes a hyperlink text to the Twitter feed.
5. Mars Facts
   1. As required, and additionally includes a hyperlink text to the source website.