**Web Crawler**

**Overview**

This web crawler was developed to crawl https://vrbo.com/ for valid rooms and scrape prices of random rooms using python2's BeautifulSoup library. BeautifulSoup can be installed via the following command:
pip install bs4

The zip file turned in contains the following files:

1. outputFile.txt containing the mean, variance and distribution of prices
2. page_#.html representing the html documents saved, where # represents a number between 0 and 9
3. cookieFile.txt containing the valid cookies obtained from the server
4. .py python scripts

To run the crawler, execute the following command:
python Crawler.py

**Methodology**

The crawler randomly chooses amongst a list of 7 good proxies, obtained via the CIAO chrome plugin, and 2 user-agents (Macintosh's Safari and Linux Firefox) to send valid requests to the target website.
A random integer between 111111 and 911111 is generated with python2's random module and a request for that room is sent to the target url using the following url format:
https://vrbo.com/{roomNumber}
The result of the request is checked for two things:

1. A redirect to the search page
2. A 404 status code

Upon passing these checks, the html document obtained is considered valid and BeautifulSoup is used to scrape the prices off the html document.

**Results**

The statistics from scraping 10 pages off the target site are contained in the file outputFile.txt
The contents are as follows:
Cost: [190.0, 140.0, 195.0, 157.0, 146.0, 1750.0, 177.0, 159.0, 196.0, 212.0]
Variance: 223859.16
Mean: 332.20

The results may not be exactly reproduced since a room to scrape is chosen at random