Merging Historical Satellite Launches with Weather

In [1]:
```
import pandas as pd
from splinter import Browser
from bs4 import BeautifulSoup
import time
import re
import lxml
```

Scrape the heavens-above.com for satellite information

In [2]:
```
url1 = "https://heavens-above.com/SatInfo.aspx?satid="
url2 = "&lat=0&lng=0&loc=Unspecified&alt=0&tz=UCT"

url = url1 + str(satid_latest) + url2
url
```

Out[2]:
```
'https://heavens-above.com/SatInfo.aspx?satid=44914&lat=0&lng=0&loc=Unspecified&alt=0&tz=UCT'
```

In [3]:
```
def init_browser():
    # @NOTE: Replace the path with your actual path to the chromedriver
    executable_path = {"executable_path": "C:\chromedriver\chromedriver.exe"}
    return Browser("chrome", **executable_path, headless=False)

browser = init_browser()
launches = {}
```

In [4]:
```
browser.visit(url)
html = browser.html
soup = BeautifulSoup(html, "html.parser")
```

In [6]:
```
soup.find("span", class_= "pagehead").get_text()
```

Out[6]:
```
'STARLINK-1073 - Satellite Information'
```

In [7]:
```
soup.find("span", id="ctl00_cph1_lblLaunchDate").get_text()
```

Out[7]:
```
'07 January 2020 02:19'
```

In [9]:
```
for x in range(44800, 44914):
    try:
        url = url1 + str(x) + url2
        browser.visit(url)
        html = browser.html

        time.sleep(3)

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

        header = soup.find("span", class_="pagehead").get_text()
        date = soup.find("span", id="ctl00_cph1_lblLaunchDate").get_text()
        table = soup.find_all('table')[6]
        site = table.find_all('td')[3].get_text()

        current_launch = pd.DataFrame({"Satellite Name": header,
                                       "Launch Date": date,
                                       "Launch Site": site}, index=[x])
        launches = launches.append(current_launch)

    except:
        print("Page Not Found")
```

In [35]:
```
launches
```

Out[35]:

	Launch Site	Satellite Name	Launch Date
44913	Plesetsk, Russia	COSMOS 2491 DEB - Satellite Information	25 December 2013 00:31

Find the date and launch site from tables

In [11]:
```
table = soup.find_all('table')[6]
table
```

Out[11]:
```
<table>
<tbody><tr>
<td>
                                                        Date
                                                        (UTC)
                                                    </td>
<td>
<span id="ctl00_cph1_lblLaunchDate">25 December 2013 00:31</span>
</td>
</tr>
<tr><td valign="top">Launch site</td><td>Plesetsk, <br/>Russia</td></tr>
<tr><td>Launch vehicle </td><td>Rokot</td></tr>
</tbody></table>
```

In [12]:
```
site = table.find_all('td')[3].get_text()
site
```

Out[12]:
```
'Plesetsk, Russia'
```

Write launch data to CSV

In [13]:
```
launches_output_file = "launches.csv"
launches.to_csv(launches_output_file)
```

In [30]:
```
launches_read = pd.read_csv(launches_output_file)
launches_read
```

Out[30]:

	Unnamed: 0	Satellite Name	Launch Date	Launch Site	lat	lon
0	0	COSMOS 2491 DEB	25 December 2013 00:31	Plesetsk, Russia	62.927545	40.575023
1	1	COSMOS 2491 DEB	25 December 2013 00:31	Plesetsk, Russia	62.927545	40.575023
2	2	COSMOS 2543	25 November 2019 17:52	Plesetsk, Russia	62.927545	40.575023
3	3	COSMOS 2543 (GLONASS	11 December 2019 08:54	Plesetsk, Russia	62.927545	40.575023
4	4	FREGAT R/B	11 December 2019 08:54	Plesetsk, Russia	62.927545	40.575023
...	...	...	...	...	...	...
120	120	SJ-7 DEB	05 July 2005 22:40	Jiuquan Satellite Launch Center, China	40.984523	100.191185
121	121	SJ-7 DEB	05 July 2005 22:40	Jiuquan Satellite Launch Center, China	40.984523	100.191185
122	122	SJ-7 DEB	05 July 2005 22:40	Jiuquan Satellite Launch Center, China	40.984523	100.191185
123	123	RS-44	26 December 2019 23:11	Amateur radio	NaN	NaN
124	124	RS-44	26 December 2019 23:11	Amateur radio	NaN	NaN

125 rows × 6 columns

Add coordinates to launch locations

In [31]:
```
launches_read["Launch Site"].unique()
```

Out[31]:
```
array(['Plesetsk, Russia', 'Satish Dhawan Space Centre (SHAR), India',
       'Taiyuan Space Launch Center, China',
       'Cape Canaveral Air Force Station, United States', 'Electron',
       'Baikonur Cosmodrome, Kazakhstan',
       'Xichang Satellite Launch Center, China',
       'Centre Spatial Guyanais , French Guiana',
       'Jiuquan Satellite Launch Center, China', 'Amateur radio '],
      dtype=object)
```

In [18]:
```
# To retrieve weather information, we also needed the coordinates of the site
s.

coords = [{"Launch Site": 'Plesetsk, Russia', "lat": "62.927545", "lon": "40.5
75023"},
          {"Launch Site": 'Satish Dhawan Space Centre (SHAR), India', "lat":
"13.733271", "lon": "80.234446"},
          {"Launch Site": 'Taiyuan Space Launch Center, China', "lat": "38.848
830" , "lon": "111.608180"},
          {"Launch Site": 'Baikonur Cosmodrome, Kazakhstan', "lat": "45.96428
7", "lon": "63.305522"},
          {"Launch Site": 'Cape Canaveral Air Force Station, United States',
"lat": "28.491981", "lon": "-80.580114"},
          {"Launch Site": "Baikonur Cosmodrome, Kazakhstan","lat" : "45.964585
1","lon": "63.3030541"},
          {"Launch Site": 'Electron', "lat": "-39.261579", "lon": "177.864987"
},
          {"Launch Site": 'Xichang Satellite Launch Center, China', "lat": "2
7.8907315", "lon": "102.2434799"},
          {"Launch Site": 'Centre Spatial Guyanais , French Guiana', "lat":
"4.8862848", "lon": "-53.0689692"},
          {"Launch Site": 'Jiuquan Satellite Launch Center, China', "lat": "4
0.9845227", "lon": "100.1911854"}
         ]
```

In [19]:
```
coords_df = pd.DataFrame(coords)
coords_df
```

Out[19]:

	Launch Site	lat	lon
0	Plesetsk, Russia	62.927545	40.575023
1	Satish Dhawan Space Centre (SHAR), India	13.733271	80.234446
2	Taiyuan Space Launch Center, China	38.848830	111.608180
3	Baikonur Cosmodrome, Kazakhstan	45.964287	63.305522
4	Cape Canaveral Air Force Station, United States	28.491981	-80.580114
5	Baikonur Cosmodrome, Kazakhstan	45.9645851	63.3030541
6	Electron	-39.261579	177.864987
7	Xichang Satellite Launch Center, China	27.8907315	102.2434799
8	Centre Spatial Guyanais , French Guiana	4.8862848	-53.0689692
9	Jiuquan Satellite Launch Center, China	40.9845227	100.1911854

Merge coordinates into launches table

In [20]:
```
#The Coordinates dataframe could be merged with the launch dataframe to provid
e coordinates of each launch.
merge_df = pd.merge(launches_read, coords_df, on="Launch Site", how='outer')
```

In [22]:
```
#Save the scraped data into a CSV file.

launches_output_file = "launches.csv"
merge_df.to_csv(launches_output_file)
```

In [25]:
```
launches_read = pd.read_csv(launches_output_file)
launches_read
```

Out[25]:

	Unnamed: 0	Satellite Name	Launch Date	Launch Site	lat	lon
0	0	COSMOS 2491 DEB	25 December 2013 00:31	Plesetsk, Russia	62.927545	40.575023
1	1	COSMOS 2491 DEB	25 December 2013 00:31	Plesetsk, Russia	62.927545	40.575023
2	2	COSMOS 2543	25 November 2019 17:52	Plesetsk, Russia	62.927545	40.575023
3	3	COSMOS 2543 (GLONASS	11 December 2019 08:54	Plesetsk, Russia	62.927545	40.575023
4	4	FREGAT R/B	11 December 2019 08:54	Plesetsk, Russia	62.927545	40.575023
...	...	...	...	...	...	...
120	120	SJ-7 DEB	05 July 2005 22:40	Jiuquan Satellite Launch Center, China	40.984523	100.191185
121	121	SJ-7 DEB	05 July 2005 22:40	Jiuquan Satellite Launch Center, China	40.984523	100.191185
122	122	SJ-7 DEB	05 July 2005 22:40	Jiuquan Satellite Launch Center, China	40.984523	100.191185
123	123	RS-44	26 December 2019 23:11	Amateur radio	NaN	NaN
124	124	RS-44	26 December 2019 23:11	Amateur radio	NaN	NaN

125 rows × 6 columns

In [26]:
```
#Clean up column names
launches_read = launches_read[["Satellite Name",
                               "Launch Date",
                               "Launch Site",
                               "lat", "lon"]]
```

In [27]:
```
launches_read
```

Out[27]:

	Satellite Name	Launch Date	Launch Site	lat	lon
0	COSMOS 2491 DEB	25 December 2013 00:31	Plesetsk, Russia	62.927545	40.575023
1	COSMOS 2491 DEB	25 December 2013 00:31	Plesetsk, Russia	62.927545	40.575023
2	COSMOS 2543	25 November 2019 17:52	Plesetsk, Russia	62.927545	40.575023
3	COSMOS 2543 (GLONASS	11 December 2019 08:54	Plesetsk, Russia	62.927545	40.575023
4	FREGAT R/B	11 December 2019 08:54	Plesetsk, Russia	62.927545	40.575023
...	...	...	...	...	...
120	SJ-7 DEB	05 July 2005 22:40	Jiuquan Satellite Launch Center, China	40.984523	100.191185
121	SJ-7 DEB	05 July 2005 22:40	Jiuquan Satellite Launch Center, China	40.984523	100.191185
122	SJ-7 DEB	05 July 2005 22:40	Jiuquan Satellite Launch Center, China	40.984523	100.191185
123	RS-44	26 December 2019 23:11	Amateur radio	NaN	NaN
124	RS-44	26 December 2019 23:11	Amateur radio	NaN	NaN

125 rows × 5 columns

Save again to CSV

In [28]:
```
#Save again with cleaned up column names
launches_output_file = "launches.csv"
launches_read.to_csv(launches_output_file)
```