Space X Falcon 9 First Stage Landing Prediction

Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia

Estimated time needed: **40** minutes

In this lab, you will be performing web scraping to collect Falcon 9 historical launch records from a Wikipedia page titled `List of Falcon 9 and Falcon Heavy launches`

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Falcon 9 first stage will land successfully

Several examples of an unsuccessful landing are shown here:

More specifically, the launch records are stored in a HTML table shown below:

Objectives

Web scrape Falcon 9 launch records with `BeautifulSoup`:

- Extract a Falcon 9 launch records HTML table from Wikipedia
- Parse the table and convert it into a Pandas data frame

First let's import required packages for this lab

In [1]:
```
!pip3 install beautifulsoup4
!pip3 install requests
```

```
Requirement already satisfied: beautifulsoup4 in /home/jupyterlab/conda/envs/p
ython/lib/python3.7/site-packages (4.11.1)
Requirement already satisfied: soupsieve>1.2 in /home/jupyterlab/conda/envs/py
thon/lib/python3.7/site-packages (from beautifulsoup4) (2.3.2.post1)
Requirement already satisfied: requests in /home/jupyterlab/conda/envs/python/
lib/python3.7/site-packages (2.28.1)
Requirement already satisfied: charset-normalizer<3,>=2 in /home/jupyterlab/co
nda/envs/python/lib/python3.7/site-packages (from requests) (2.1.0)
Requirement already satisfied: certifi>=2017.4.17 in /home/jupyterlab/conda/en
vs/python/lib/python3.7/site-packages (from requests) (2022.6.15)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /home/jupyterlab/cond
a/envs/python/lib/python3.7/site-packages (from requests) (1.26.11)
Requirement already satisfied: idna<4,>=2.5 in /home/jupyterlab/conda/envs/pyt
hon/lib/python3.7/site-packages (from requests) (3.3)
```

In [2]:
```
import sys
import requests
import unicodedata
import numpy as np
import datetime
import pandas as pd
from bs4 import BeautifulSoup
```

and we will provide some helper functions for you to process web scraped HTML table

In [3]:
```
def date_time(table_cells):
    """
    This function returns the data and time from the HTML  table cell
    Input: the  element of a table data cell extracts extra row
    """
    return [data_time.strip() for data_time in list(table_cells.strings)][0:2]

def booster_version(table_cells):
    """
    This function returns the booster version from the HTML  table cell
    Input: the  element of a table data cell extracts extra row
    """
    out=''.join([booster_version for i,booster_version in enumerate( table_cell
    return out

def landing_status(table_cells):
    """
    This function returns the landing status from the HTML table cell
    Input: the  element of a table data cell extracts extra row
    """
    out=[i for i in table_cells.strings][0]
    return out
```

```
def get_mass(table_cells):
    mass=unicodedata.normalize("NFKD", table_cells.text).strip()
    if mass:
        mass.find("kg")
        new_mass=mass[0:mass.find("kg")+2]
    else:
        new_mass=0
    return new_mass


def extract_column_from_header(row):
    """
    This function returns the landing status from the HTML table cell
    Input: the  element of a table data cell extracts extra row
    """
    if (row.br):
        row.br.extract()
    if row.a:
        row.a.extract()
    if row.sup:
        row.sup.extract()

    colunm_name = ' '.join(row.contents)

    # Filter the digit and empty names
    if not(colunm_name.strip().isdigit()):
        colunm_name = colunm_name.strip()
        return colunm_name
```

To keep the lab tasks consistent, you will be asked to scrape the data from a snapshot of the `List of Falcon 9 and Falcon Heavy launches` Wikipage updated on `9th June 2021`

In [4]:
```
Falcwiki = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Fal
```

Next, request the HTML page from the above URL and get a `response` object

TASK 1: Request the Falcon9 Launch Wiki page from its URL

First, let's perform an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response.

In [5]:
```
# use requests.get() method with the provided static_url
# assign the response to a object
response = requests.get(Falcwiki).text
```

Create a `BeautifulSoup` object from the HTML `response`

In [6]:
```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text co
soup = BeautifulSoup(response,  'html.parser')
```

Print the page title to verify if the `BeautifulSoup` object was created properly

In [7]:
```
print(soup.title)
```

```
<title>List of Falcon 9 and Falcon Heavy launches - Wikipedia</title>
```

TASK 2: Extract all column/variable names from the HTML table header

Next, we want to collect all relevant column names from the HTML table header

Let's try to find all tables on the wiki page first. If you need to refresh your memory about `BeautifulSoup`, please check the external reference link towards the end of this lab

In [8]:
```
# Use the find_all function in the BeautifulSoup object, with element type `tab
# Assign the result to a list called `html_tables`
html_tables = soup.find_all('table')
```

Starting from the third table is our target table contains the actual launch records.

You should able to see the columns names embedded in the table header elements `<th>` as follows:

In [9]:
```
# Let's print the third table and check its content

third_launch_table = html_tables[3]
print (third_launch_table)
```

```
<table class="wikitable plainrowheaders collapsible" style="width: 100%;">
<tbody><tr>
<th scope="col">Flight No.
</th>
<th scope="col">Date and<br/>time (<a href="/wiki/Coordinated_Universal_Time"
title="Coordinated Universal Time">UTC</a>)
</th>
<th scope="col"><a href="/wiki/List_of_Falcon_9_first-stage_boosters" title="L
ist of Falcon 9 first-stage boosters">Version,<br/>Booster</a><sup class="refe
rence" id="cite_ref-booster_11-1"><a href="#cite_note-booster-11">[b]</a></sup
>
</th>
<th scope="col">Launch site
</th>
<th scope="col">Payload<sup class="reference" id="cite_ref-Dragon_12-1"><a hre
f="#cite_note-Dragon-12">[c]</a></sup>
</th>
<th scope="col">Payload mass
</th>
<th scope="col">Orbit
</th>
<th scope="col">Customer
</th>
<th scope="col">Launch<br/>outcome
</th>
<th scope="col"><a href="/wiki/Falcon_9_first-stage_landing_tests" title="Falc
on 9 first-stage landing tests">Booster<br/>landing</a>
</th></tr>
<tr>
<th rowspan="2" scope="row" style="text-align:center;">8
</th>
<td>6 January 2014,<br/>22:06<sup class="reference" id="cite_ref-NASA_Spacefli
ght_48-0"><a href="#cite_note-NASA_Spaceflight-48">[41]</a></sup>
</td>
<td><a href="/wiki/Falcon_9_v1.1" title="Falcon 9 v1.1">F9 v1.1</a>
</td>
<td><a href="/wiki/Cape_Canaveral_Space_Force_Station" title="Cape Canaveral S
pace Force Station">CCAFS</a>,<br/><a href="/wiki/Cape_Canaveral_Space_Launch_
Complex_40" title="Cape Canaveral Space Launch Complex 40">SLC-40</a>
</td>
<td><a href="/wiki/Thaicom_6" title="Thaicom 6">Thaicom 6</a><sup class="refer
ence" id="cite_ref-sxManifest20120925_28-4"><a href="#cite_note-sxManifest2012
0925-28">[22]</a></sup>
</td>
<td>3,325 kg (7,330 lb)
</td>
<td><a href="/wiki/Geostationary_transfer_orbit" title="Geostationary transfer
orbit">GTO</a>
</td>
<td><a href="/wiki/Thaicom" title="Thaicom">Thaicom</a>
</td>
<td class="table-success" style="background: #9EFF9E; vertical-align: middle;
text-align: center;">Success<sup class="reference" id="cite_ref-sn20140106_49-
0"><a href="#cite_note-sn20140106-49">[42]</a></sup>
</td>
<td class="table-noAttempt" style="background: #EEE; vertical-align: middle; w
hite-space: nowrap; text-align: center;">No attempt<br/><sup class="reference"
id="cite_ref-50"><a href="#cite_note-50">[43]</a></sup>
</td></tr>
<tr>
```

```
<td colspan="9">The Thai communication satellite was the second <a href="/wik
i/Geostationary_transfer_orbit" title="Geostationary transfer orbit">GTO</a> l
aunch for Falcon 9. The <a href="/wiki/United_States_Air_Force" title="United
States Air Force">USAF</a> evaluated launch data from this flight as part of a
separate certification program for SpaceX to qualify to fly military payloads,
but found that the launch had "unacceptable fuel reserves at engine cutoff of
the stage 2 second burnoff".<sup class="reference" id="cite_ref-bloomberg20140
722_51-0"><a href="#cite_note-bloomberg20140722-51">[44]</a></sup> Thaicom-6 w
as inserted into a <a href="/wiki/Geostationary_transfer_orbit" title="Geostat
ionary transfer orbit">Super-Synchronous Transfer Orbit</a> of 90,039 km (55,9
48 mi) in <a href="/wiki/Apsis" title="Apsis">apogee</a> with an <a href="/wik
i/Orbital_inclination" title="Orbital inclination">inclination</a> of 22.46° to
the <a href="/wiki/Equator" title="Equator">equator</a>.
</td></tr>
<tr>
<th rowspan="2" scope="row" style="text-align:center;">9
</th>
<td>18 April 2014,<br/>19:25<sup class="reference" id="cite_ref-SFN_LLog_27-
1"><a href="#cite_note-SFN_LLog-27">[21]</a></sup>
</td>
<td><a href="/wiki/Falcon_9_v1.1" title="Falcon 9 v1.1">F9 v1.1</a>
</td>
<td><a class="mw-redirect" href="/wiki/Cape_Canaveral_Air_Force_Station" title
="Cape Canaveral Air Force Station">Cape Canaveral</a>,<br/><a href="/wiki/Cap
e_Canaveral_Space_Launch_Complex_40" title="Cape Canaveral Space Launch Comple
x 40">LC-40</a>
</td>
<td><a href="/wiki/SpaceX_CRS-3" title="SpaceX CRS-3">SpaceX CRS-3</a><sup cla
ss="reference" id="cite_ref-sxManifest20120925_28-5"><a href="#cite_note-sxMan
ifest20120925-28">[22]</a></sup><br/>(Dragon C105)
</td>
<td>2,296 kg (5,062 lb)<sup class="reference" id="cite_ref-52"><a href="#cite_
note-52">[45]</a></sup>
</td>
<td><a href="/wiki/Low_Earth_orbit" title="Low Earth orbit">LEO</a> (<a class
="mw-redirect" href="/wiki/ISS" title="ISS">ISS</a>)
</td>
<td><a href="/wiki/NASA" title="NASA">NASA</a> (<a href="/wiki/Commercial_Resu
pply_Services" title="Commercial Resupply Services">CRS</a>)
</td>
<td class="table-success" style="background: #9EFF9E; vertical-align: middle;
text-align: center;">Success
</td>
<td class="partial table-partial" style="background: #BFE; vertical-align: mid
dle; text-align: center;">Controlled<br/><small>(ocean)</small> <sup class="re
ference" id="cite_ref-ocean_landing_38-1"><a href="#cite_note-ocean_landing-3
8">[d]</a></sup><sup class="reference" id="cite_ref-auto_53-0"><a href="#cite_
note-auto-53">[46]</a></sup>
</td></tr>
<tr>
<td colspan="9">Following second-stage separation, SpaceX conducted a second <
a href="/wiki/Falcon_9_first-stage_landing_tests" title="Falcon 9 first-stage
landing tests">controlled-descent test</a> of the discarded booster vehicle an
d achieved the first successful controlled ocean touchdown of a liquid-rocket-
engine orbital booster.<sup class="reference" id="cite_ref-mit20140422_54-0"><
a href="#cite_note-mit20140422-54">[47]</a></sup><sup class="reference" id="ci
te_ref-aw20140428_55-0"><a href="#cite_note-aw20140428-55">[48]</a></sup> Foll
owing the soft touchdown, the first stage tipped over as expected and was dest
royed. This was the first Falcon 9 booster to fly with extensible landing legs
and the first Dragon mission with the <a href="/wiki/Falcon_9_v1.1" title="Fal
```

```
con 9 v1.1">Falcon 9 v1.1</a> launch vehicle. This flight also launched the <a
href="/wiki/Educational_Launch_of_Nanosatellites" title="Educational Launch of
Nanosatellites">ELaNa 5</a> mission for <a href="/wiki/NASA" title="NASA">NASA
</a> as a secondary payload.<sup class="reference" id="cite_ref-auto2_56-0"><a
href="#cite_note-auto2-56">[49]</a></sup><sup class="reference" id="cite_ref-5
7"><a href="#cite_note-57">[50]</a></sup>
</td></tr>
<tr>
<th rowspan="2" scope="row" style="text-align:center;">10
</th>
<td>14 July 2014,<br/>15:15
</td>
<td><a href="/wiki/Falcon_9_v1.1" title="Falcon 9 v1.1">F9 v1.1</a>
</td>
<td><a class="mw-redirect" href="/wiki/Cape_Canaveral_Air_Force_Station" title
="Cape Canaveral Air Force Station">Cape Canaveral</a>,<br/><a href="/wiki/Cap
e_Canaveral_Space_Launch_Complex_40" title="Cape Canaveral Space Launch Comple
x 40">LC-40</a>
</td>
<td><a class="mw-redirect" href="/wiki/Orbcomm-OG2" title="Orbcomm-OG2">Orbcom
m-OG2</a>-1<br/>(6 satellites)<sup class="reference" id="cite_ref-sxManifest20
120925_28-6"><a href="#cite_note-sxManifest20120925-28">[22]</a></sup>
</td>
<td>1,316 kg (2,901 lb)
</td>
<td><a href="/wiki/Low_Earth_orbit" title="Low Earth orbit">LEO</a>
</td>
<td><a href="/wiki/Orbcomm" title="Orbcomm">Orbcomm</a>
</td>
<td class="table-success" style="background: #9EFF9E; vertical-align: middle;
text-align: center;">Success<sup class="reference" id="cite_ref-og2-01_2014071
4_58-0"><a href="#cite_note-og2-01_20140714-58">[51]</a></sup>
</td>
<td class="partial table-partial" style="background: #BFE; vertical-align: mid
dle; text-align: center;">Controlled<br/><small>(ocean)</small><sup class="ref
erence" id="cite_ref-ocean_landing_38-2"><a href="#cite_note-ocean_landing-3
8">[d]</a></sup><sup class="reference" id="cite_ref-auto_53-1"><a href="#cite_
note-auto-53">[46]</a></sup>
</td></tr>
<tr>
<td colspan="9">Payload included six satellites weighing 172 kg (379 lb) each
and two 142 kg (313 lb) mass simulators.<sup class="reference" id="cite_ref-gu
nter-og2_30-1"><a href="#cite_note-gunter-og2-30">[24]</a></sup><sup class="re
ference" id="cite_ref-gunter-og2-sim_59-0"><a href="#cite_note-gunter-og2-sim-
59">[52]</a></sup> Equipped for the second time with <a class="mw-redirect" hr
ef="/wiki/Launch_vehicle_landing_gear" title="Launch vehicle landing gear">lan
ding legs</a>, the first-stage booster successfully conducted a <a class="mw-r
edirect" href="/wiki/SpaceX_Falcon_9_booster_post-mission,_controlled-descent,
_test_program" title="SpaceX Falcon 9 booster post-mission, controlled-descen
t, test program">controlled-descent</a> test consisting of a burn for decelera
tion from <a class="mw-redirect" href="/wiki/Hypersonic" title="Hypersonic">hy
personic</a> velocity in the upper atmosphere, a <a href="/wiki/Atmospheric_en
try" title="Atmospheric entry">reentry</a> burn, and a final landing burn befo
re soft-landing on the ocean surface.<sup class="reference" id="cite_ref-Space
X22072014_60-0"><a href="#cite_note-SpaceX22072014-60">[53]</a></sup>
</td></tr>
<tr>
<th rowspan="2" scope="row" style="text-align:center;">11
</th>
<td>5 August 2014,<br/>08:00
```

```
</td>
<td><a href="/wiki/Falcon_9_v1.1" title="Falcon 9 v1.1">F9 v1.1</a>
</td>
<td><a class="mw-redirect" href="/wiki/Cape_Canaveral_Air_Force_Station" title
="Cape Canaveral Air Force Station">Cape Canaveral</a>,<br/><a href="/wiki/Cap
e_Canaveral_Space_Launch_Complex_40" title="Cape Canaveral Space Launch Comple
x 40">LC-40</a>
</td>
<td><a href="/wiki/AsiaSat_8" title="AsiaSat 8">AsiaSat 8</a><sup class="refer
ence" id="cite_ref-sxManifest20120925_28-7"><a href="#cite_note-sxManifest2012
0925-28">[22]</a></sup><sup class="reference" id="cite_ref-AsiaSat_SpaceX_61-
0"><a href="#cite_note-AsiaSat_SpaceX-61">[54]</a></sup><sup class="reference"
id="cite_ref-62"><a href="#cite_note-62">[55]</a></sup>
</td>
<td>4,535 kg (9,998 lb)
</td>
<td><a href="/wiki/Geostationary_transfer_orbit" title="Geostationary transfer
orbit">GTO</a>
</td>
<td><a href="/wiki/AsiaSat" title="AsiaSat">AsiaSat</a>
</td>
<td class="table-success" style="background: #9EFF9E; vertical-align: middle;
text-align: center;">Success<sup class="reference" id="cite_ref-as8_20140805_6
3-0"><a href="#cite_note-as8_20140805-63">[56]</a></sup>
</td>
<td class="table-noAttempt" style="background: #EEE; vertical-align: middle; w
hite-space: nowrap; text-align: center;">No attempt<br/><sup class="reference"
id="cite_ref-amspace-20140803_64-0"><a href="#cite_note-amspace-20140803-64">
[57]</a></sup>
</td></tr>
<tr>
<td colspan="9">First time SpaceX managed a launch site turnaround between two
flights of under a month (22 days). GTO launch of the large communication sate
llite from Hong Kong did not allow for propulsive return-over-water and contro
lled splashdown of the first stage.<sup class="reference" id="cite_ref-amspace
-20140803_64-1"><a href="#cite_note-amspace-20140803-64">[57]</a></sup>
</td></tr>
<tr>
<th rowspan="2" scope="row" style="text-align:center;">12
</th>
<td>7 September 2014,<br/>05:00
</td>
<td><a href="/wiki/Falcon_9_v1.1" title="Falcon 9 v1.1">F9 v1.1</a><br/>B1011<
sup class="reference" id="cite_ref-block_numbers_14-6"><a href="#cite_note-blo
ck_numbers-14">[8]</a></sup>
</td>
<td><a class="mw-redirect" href="/wiki/Cape_Canaveral_Air_Force_Station" title
="Cape Canaveral Air Force Station">Cape Canaveral</a>,<br/><a href="/wiki/Cap
e_Canaveral_Space_Launch_Complex_40" title="Cape Canaveral Space Launch Comple
x 40">LC-40</a>
</td>
<td><a href="/wiki/AsiaSat_6" title="AsiaSat 6">AsiaSat 6</a><sup class="refer
ence" id="cite_ref-sxManifest20120925_28-8"><a href="#cite_note-sxManifest2012
0925-28">[22]</a></sup><sup class="reference" id="cite_ref-AsiaSat_SpaceX_61-
1"><a href="#cite_note-AsiaSat_SpaceX-61">[54]</a></sup><sup class="reference"
id="cite_ref-65"><a href="#cite_note-65">[58]</a></sup>
</td>
<td>4,428 kg (9,762 lb)
</td>
<td><a href="/wiki/Geostationary_transfer_orbit" title="Geostationary transfer
```

```
orbit">GTO</a>
</td>
<td><a href="/wiki/AsiaSat" title="AsiaSat">AsiaSat</a>
</td>
<td class="table-success" style="background: #9EFF9E; vertical-align: middle;
text-align: center;">Success<sup class="reference" id="cite_ref-sdc20140907_66
-0"><a href="#cite_note-sdc20140907-66">[59]</a></sup>
</td>
<td class="table-noAttempt" style="background: #EEE; vertical-align: middle; w
hite-space: nowrap; text-align: center;">No attempt
</td></tr>
<tr>
<td colspan="9">Launch was delayed for two weeks for additional verifications
after a malfunction observed in the development of the <a class="mw-redirect"
href="/wiki/F9R_Dev1" title="F9R Dev1">F9R Dev1</a> prototype.<sup class="refe
rence" id="cite_ref-67"><a href="#cite_note-67">[60]</a></sup> GTO launch of t
he heavy payload did not allow for controlled splashdown.<sup class="referenc
e" id="cite_ref-68"><a href="#cite_note-68">[61]</a></sup>
</td></tr>
<tr>
<th rowspan="2" scope="row" style="text-align:center;">13
</th>
<td>21 September 2014,<br/>05:52<sup class="reference" id="cite_ref-SFN_LLog_2
7-2"><a href="#cite_note-SFN_LLog-27">[21]</a></sup>
</td>
<td><a href="/wiki/Falcon_9_v1.1" title="Falcon 9 v1.1">F9 v1.1</a><br/>B1010<
sup class="reference" id="cite_ref-block_numbers_14-7"><a href="#cite_note-blo
ck_numbers-14">[8]</a></sup>
</td>
<td><a class="mw-redirect" href="/wiki/Cape_Canaveral_Air_Force_Station" title
="Cape Canaveral Air Force Station">Cape Canaveral</a>,<br/><a href="/wiki/Cap
e_Canaveral_Space_Launch_Complex_40" title="Cape Canaveral Space Launch Comple
x 40">LC-40</a>
</td>
<td><a href="/wiki/SpaceX_CRS-4" title="SpaceX CRS-4">SpaceX CRS-4</a><sup cla
ss="reference" id="cite_ref-sxManifest20120925_28-9"><a href="#cite_note-sxMan
ifest20120925-28">[22]</a></sup><br/>(Dragon <a href="/wiki/Dragon_C106" title
="Dragon C106">C106</a>.1)
</td>
<td>2,216 kg (4,885 lb)<sup class="reference" id="cite_ref-69"><a href="#cite_
note-69">[62]</a></sup>
</td>
<td><a href="/wiki/Low_Earth_orbit" title="Low Earth orbit">LEO</a> (<a class
="mw-redirect" href="/wiki/ISS" title="ISS">ISS</a>)
</td>
<td><a href="/wiki/NASA" title="NASA">NASA</a> (<a href="/wiki/Commercial_Resu
pply_Services" title="Commercial Resupply Services">CRS</a>)
</td>
<td class="table-success" style="background: #9EFF9E; vertical-align: middle;
text-align: center;">Success<sup class="reference" id="cite_ref-nasacrs4201409
21_70-0"><a href="#cite_note-nasacrs420140921-70">[63]</a></sup>
</td>
<td class="table-no2" style="background: #FFE3E3; color: black; vertical-alig
n: middle; text-align: center;">Uncontrolled<br/><small>(ocean)</small><sup cl
ass="reference" id="cite_ref-ocean_landing_38-3"><a href="#cite_note-ocean_lan
ding-38">[d]</a></sup><sup class="reference" id="cite_ref-fail-13_71-0"><a hre
f="#cite_note-fail-13-71">[64]</a></sup>
</td></tr>
<tr>
<td colspan="9">Fourth attempt of a soft ocean touchdown,<sup class="referenc
```

```
e" id="cite_ref-aw20141016_72-0"><a href="#cite_note-aw20141016-72">[65]</a></
sup> but the booster ran out of liquid oxygen.<sup class="reference" id="cite_
ref-fail-13_71-1"><a href="#cite_note-fail-13-71">[64]</a></sup> Detailed <a c
lass="mw-redirect" href="/wiki/Thermal_imaging" title="Thermal imaging">therma
l imaging</a> infrared sensor data was collected however by NASA, as part of a
joint arrangement with SpaceX as part of research on <a class="mw-redirect" hr
ef="/wiki/Supersonic_retropropulsion" title="Supersonic retropropulsion">retro
propulsive deceleration technologies</a> for developing new approaches to Mart
ian <a href="/wiki/Atmospheric_entry" title="Atmospheric entry">atmospheric en
try</a>.<sup class="reference" id="cite_ref-aw20141016_72-1"><a href="#cite_no
te-aw20141016-72">[65]</a></sup>
</td></tr></tbody></table>
```

```
<tr>
<th scope="col">Flight No.
</th>
<th scope="col">Date and<br/>time (<a
href="/wiki/Coordinated_Universal_Time" title="Coordinated
Universal Time">UTC</a>)
</th>
<th scope="col"><a href="/wiki/List_of_Falcon_9_first-
stage_boosters" title="List of Falcon 9 first-stage
boosters">Version,<br/>Booster</a> <sup class="reference"
id="cite_ref-booster_11-0"><a href="#cite_note-booster-11">[b]
</a></sup>
</th>
<th scope="col">Launch site
</th>
<th scope="col">Payload<sup class="reference" id="cite_ref-
Dragon_12-0"><a href="#cite_note-Dragon-12">[c]</a></sup>
</th>
<th scope="col">Payload mass
</th>
<th scope="col">Orbit
</th>
<th scope="col">Customer
</th>
<th scope="col">Launch<br/>outcome
</th>
<th scope="col"><a href="/wiki/Falcon_9_first-
stage_landing_tests" title="Falcon 9 first-stage landing
tests">Booster<br/>landing</a>
</th></tr>
```

Next, we just need to iterate through the `<th>` elements and apply the provided `extract_column_from_header()` to extract column name one by one

In [10]:
```
column_names = []
for row in third_launch_table.find_all('th'):
name = extract_column_from_header(row):
    if (name != None and len(name) >0):
        column_names.append(name)
# Apply find_all() function with `th` element on first_launch_table
```

```
# Iterate each th element and apply the provided extract_column_from_header() t
# Append the Non-empty column name (`if name is not None and len(name) > 0`) ir
```

```
  File "/tmp/ipykernel_71/384491918.py", line 3
    name = extract_column_from_header(row):
       ^
IndentationError: expected an indented block
```

Check the extracted column names

In [ ]:
```
print(column_names)
```

TASK 3: Create a data frame by parsing the launch HTML tables

We will create an empty dictionary with keys from the extracted column names in the previous task. Later, this dictionary will be converted into a Pandas dataframe

In [ ]:
```
launch_dict= dict.fromkeys(column_names)

# Remove an irrelvant column
del launch_dict['Date and time ( )']

# Let's initial the launch_dict with each value to be an empty list
launch_dict['Flight No.'] = []
launch_dict['Launch site'] = []
launch_dict['Payload'] = []
launch_dict['Payload mass'] = []
launch_dict['Orbit'] = []
launch_dict['Customer'] = []
launch_dict['Launch outcome'] = []
# Added some new columns
launch_dict['Version Booster']=[]
launch_dict['Booster landing']=[]
launch_dict['Date']=[]
launch_dict['Time']=[]
```

Next, we just need to fill up the `launch_dict` with launch records extracted from table rows.

Usually, HTML tables in Wiki pages are likely to contain unexpected annotations and other types of noises, such as reference links `B0004.1[8]`, missing values `N/A [e]`, inconsistent formatting, etc.

To simplify the parsing process, we have provided an incomplete code snippet below to help you to fill up the `launch_dict`. Please complete the following code snippet with TODOs or you can choose to write your own logic to parse all launch tables:

In [ ]:
```
extracted_row = 0
#Extract each table
for table_number,table in enumerate(soup.find_all('table',"wikitable plainrowhe
   # get table row
    for rows in table.find_all("tr"):
```

```
        #check to see if first table heading is as number corresponding to laun
        if rows.th:
            if rows.th.string:
                flight_number=rows.th.string.strip()
                flag=flight_number.isdigit()

        else:
            flag=False
        #get table element
        row=rows.find_all('td')
        #if it is number save cells in a dictonary
        if flag:
            extracted_row += 1

            launch_dict['Flight No.'].append(flight_number)
            datatimelist=date_time(row[0])

            date = datatimelist[0].strip(',')
            launch_dict['Date'].append(date)

            time = datatimelist[1]
            launch_dict['Time'].append(time)

            bv=booster_version(row[1])
            if not (bv):
                bv=row[1].a.string
            launch_dict['Version Booster'].append(bv)

            launch_site = row[2].a.string
            launch_dict['Launch site'].append(launch_site)

            payload = row[3].a.string
            launch_dict['Payload'].append(payload)

            payload_mass = get_mass(row[4])
            launch_dict['Payload mass'].append(payload_mass)

            orbit = row[5].a.string
            launch_dict['Orbit'].append(orbit)

            customer = row[6].text.strip()
            launch_dict['Customer'].append(customer)

            launch_outcome = list(row[7].strings)[0]
            launch_dict['Launch outcome'].append(launch_outcome)

            booster_landing = landing_status(row[8])
            launch_dict['Booster landing'].append(booster_landing)
```

In [ ]:
```
print(len(time), len(bv), len(payload), len(date), len(flight_number), len(laun
```

After you have fill in the parsed launch record values into `launch_dict`, you can create a dataframe from it.

In [ ]:
```
df = pd.DataFrame(launch_dict)
df.head(25)
```

We can now export it to a **CSV** for the next section, but to make the answers consistent and

in case you have difficulties finishing this lab.

Following labs will be using a provided dataset to make each lab independent.

```
df.to_csv('spacex_web_scraped.csv', index=False)
```

Authors

Yan Luo

Nayef Abou Tayoun

Change Log

Date (YYYY-MM-DD)	Version	Changed By	Change Description
2021-06-09	1.0	Yan Luo	Tasks updates
2020-11-10	1.0	Nayef	Created the initial version

In [ ]:

In [ ]: