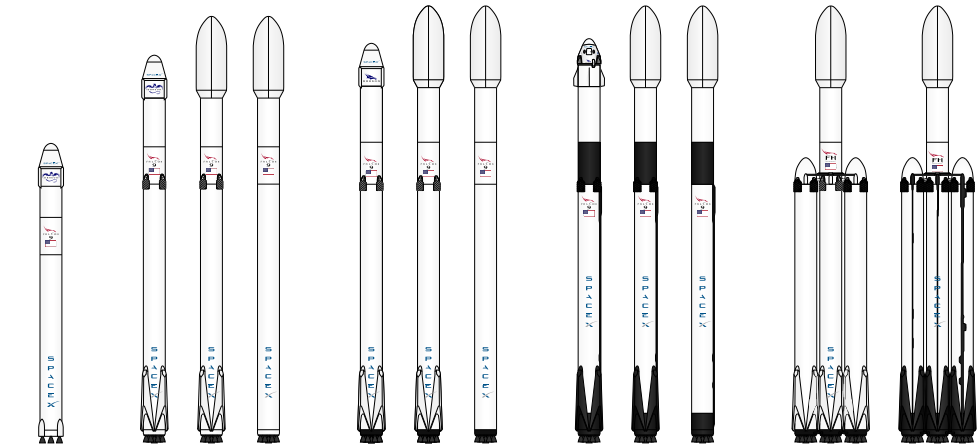
Space X Falcon 9 First Stage Landing Prediction

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:

2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 △ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 △ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 △ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 △ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 △ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 △)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 △ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 △ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

Objectives

Web scrap 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 [6]:
```
!pip3 install beautifulsoup4
!pip3 install requests
```

```
Requirement already satisfied: beautifulsoup4 in /home/jupyterlab/conda/envs/python/
lib/python3.7/site-packages (4.11.1)
Requirement already satisfied: soupsieve>1.2 in /home/jupyterlab/conda/envs/python/l
ib/python3.7/site-packages (from beautifulsoup4) (2.3.2.post1)
Requirement already satisfied: requests in /home/jupyterlab/conda/envs/python/lib/py
thon3.7/site-packages (2.29.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /home/jupyterlab/conda/en
vs/python/lib/python3.7/site-packages (from requests) (3.1.0)
Requirement already satisfied: idna<4,>=2.5 in /home/jupyterlab/conda/envs/python/li
b/python3.7/site-packages (from requests) (3.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /home/jupyterlab/conda/envs/
python/lib/python3.7/site-packages (from requests) (1.26.15)
Requirement already satisfied: certifi>=2017.4.17 in /home/jupyterlab/conda/envs/pyt
hon/lib/python3.7/site-packages (from requests) (2023.5.7)
```

In [30]:
```
import sys

import requests
from bs4 import BeautifulSoup
import re
import unicodedata
import pandas as pd
```

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

In [8]:
```
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_cells.st
    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 [9]:
```
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falco
```

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 [12]:
```
# use requests.get() method with the provided static_url
# assign the response to a object
```

```
data=requests.get(static_url)
data.status_code
```

Out[12]:
```
200
```

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

In [13]:
```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup=BeautifulSoup(data.text)
```

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

In [14]:
```
# Use soup.title attribute
soup.title
```

Out[14]:
```
<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 [15]:
```
# Use the find_all function in the BeautifulSoup object, with element type `table`
# 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.

In [16]:
```
# Let's print the third table and check its content
first_launch_table = html_tables[2]
print(first_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="List of
Falcon 9 first-stage boosters">Version,<br/>Booster</a> <sup class="reference" id="c
ite_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="#ci
te_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 f
irst-stage landing tests">Booster<br/>landing</a>
</th></tr>
<tr>
<th rowspan="2" scope="row" style="text-align:center;">1
</th>
<td>4 June 2010,<br/>18:45
</td>
<td><a href="/wiki/Falcon_9_v1.0" title="Falcon 9 v1.0">F9 v1.0</a><sup class="refer
ence" id="cite_ref-MuskMay2012_13-0"><a href="#cite_note-MuskMay2012-13">[7]</a></su
p><br/>B0003.1<sup class="reference" id="cite_ref-block_numbers_14-0"><a href="#cite
_note-block_numbers-14">[8]</a></sup>
</td>
<td><a href="/wiki/Cape_Canaveral_Space_Force_Station" title="Cape Canaveral Space F
orce 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/Dragon_Spacecraft_Qualification_Unit" title="Dragon Spacecraft Qu
alification Unit">Dragon Spacecraft Qualification Unit</a>
</td>
<td>
</td>
<td><a href="/wiki/Low_Earth_orbit" title="Low Earth orbit">LEO</a>
</td>
<td><a href="/wiki/SpaceX" title="SpaceX">SpaceX</a>
</td>
<td class="table-success" style="background: #9EFF9E; vertical-align: middle; text-a
lign: center;">Success
</td>
<td class="table-failure" style="background: #FFC7C7; vertical-align: middle; text-a
lign: center;">Failure<sup class="reference" id="cite_ref-ns20110930_15-0"><a href
="#cite_note-ns20110930-15">[9]</a></sup><sup class="reference" id="cite_ref-16"><a
```

```
href="#cite_note-16">[10]</a></sup><br/><small>(parachute)</small>
</td></tr>
<tr>
<td colspan="9">First flight of Falcon 9 v1.0.<sup class="reference" id="cite_ref-sf
n20100604_17-0"><a href="#cite_note-sfn20100604-17">[11]</a></sup> Used a boilerplat
e version of Dragon capsule which was not designed to separate from the second stag
e.<small>(<a href="#First_flight_of_Falcon_9">more details below</a>)</small> Attemp
ted to recover the first stage by parachuting it into the ocean, but it burned up on
reentry, before the parachutes even deployed.<sup class="reference" id="cite_ref-par
achute_18-0"><a href="#cite_note-parachute-18">[12]</a></sup>
</td></tr>
<tr>
<th rowspan="2" scope="row" style="text-align:center;">2
</th>
<td>8 December 2010,<br/>15:43<sup class="reference" id="cite_ref-spaceflightnow_Cla
rk_Launch_Report_19-0"><a href="#cite_note-spaceflightnow_Clark_Launch_Report-19">[1
3]</a></sup>
</td>
<td><a href="/wiki/Falcon_9_v1.0" title="Falcon 9 v1.0">F9 v1.0</a><sup class="refer
ence" id="cite_ref-MuskMay2012_13-1"><a href="#cite_note-MuskMay2012-13">[7]</a></su
p><br/>B0004.1<sup class="reference" id="cite_ref-block_numbers_14-1"><a href="#cite
_note-block_numbers-14">[8]</a></sup>
</td>
<td><a href="/wiki/Cape_Canaveral_Space_Force_Station" title="Cape Canaveral Space F
orce 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/SpaceX_Dragon" title="SpaceX Dragon">Dragon</a> <a class="mw-redi
rect" href="/wiki/COTS_Demo_Flight_1" title="COTS Demo Flight 1">demo flight C1</a><
br/>(Dragon C101)
</td>
<td>
</td>
<td><a href="/wiki/Low_Earth_orbit" title="Low Earth orbit">LEO</a> (<a href="/wiki/
International_Space_Station" title="International Space Station">ISS</a>)
</td>
<td><style data-mw-deduplicate="TemplateStyles:r1126788409">.mw-parser-output .plain
list ol,.mw-parser-output .plainlist ul{line-height:inherit;list-style:none;margin:
0;padding:0}.mw-parser-output .plainlist ol li,.mw-parser-output .plainlist ul li{ma
rgin-bottom:0}</style><div class="plainlist">
<ul><li><a href="/wiki/NASA" title="NASA">NASA</a> (<a href="/wiki/Commercial_Orbita
l_Transportation_Services" title="Commercial Orbital Transportation Services">COTS</
a>)</li>
<li><a href="/wiki/National_Reconnaissance_Office" title="National Reconnaissance Of
fice">NRO</a></li></ul>
</div>
</td>
<td class="table-success" style="background: #9EFF9E; vertical-align: middle; text-a
lign: center;">Success<sup class="reference" id="cite_ref-ns20110930_15-1"><a href
="#cite_note-ns20110930-15">[9]</a></sup>
</td>
<td class="table-failure" style="background: #FFC7C7; vertical-align: middle; text-a
lign: center;">Failure<sup class="reference" id="cite_ref-ns20110930_15-2"><a href
="#cite_note-ns20110930-15">[9]</a></sup><sup class="reference" id="cite_ref-20"><a
href="#cite_note-20">[14]</a></sup><br/><small>(parachute)</small>
</td></tr>
```

<tr>
<td colspan="9">Maiden flight of Dragon capsule, consisting of over 3 hours of testing thruster maneuvering and reentry.^{[15]} Attempted to recover the first stage by parachuting it into the ocean, but it disintegrated upon reentry, before the parachutes were deployed.^{[12]} <small>(more details below)</small> It also included two CubeSats,^{[16]} and a wheel of Brouère cheese.
</td></tr>
<tr>
<th rowspan="2" scope="row" style="text-align:center;">3
</th>
<td>22 May 2012,
07:44^{[17]}
</td>
<td>F9 v1.0^{[7]}
B0005.1^{[8]}
</td>
<td>CCAFS,
SLC-40
</td>
<td>Dragon demo flight C2+^{[18]}
(Dragon C102)
</td>
<td>525 kg (1,157 lb)^{[19]}
</td>
<td>LEO (ISS)
</td>
<td>NASA (COTS)
</td>
<td class="table-success" style="background: #9EFF9E; vertical-align: middle; text-align: center;">Success^{[20]}
</td>
<td class="table-noAttempt" style="background: #EEE; vertical-align: middle; white-space: nowrap; text-align: center;">No attempt
</td></tr>
<tr>
<td colspan="9">Dragon spacecraft demonstrated a series of tests before it was allowed to approach the International Space Station. Two days later, it became the first commercial spacecraft to board the ISS.^{[17]} <small>(<a href="#COTS_d

```
emo_missions">more details below</a>)</small>
</td></tr>
<tr>
<th rowspan="3" scope="row" style="text-align:center;">4
</th>
<td rowspan="2">8 October 2012,<br/>00:35<sup class="reference" id="cite_ref-SFN_LLo
g_27-0"><a href="#cite_note-SFN_LLog-27">[21]</a></sup>
</td>
<td rowspan="2"><a href="/wiki/Falcon_9_v1.0" title="Falcon 9 v1.0">F9 v1.0</a><sup
class="reference" id="cite_ref-MuskMay2012_13-3"><a href="#cite_note-MuskMay2012-1
3">[7]</a></sup><br/>B0006.1<sup class="reference" id="cite_ref-block_numbers_14-3">
<a href="#cite_note-block_numbers-14">[8]</a></sup>
</td>
<td rowspan="2"><a href="/wiki/Cape_Canaveral_Space_Force_Station" title="Cape Canav
eral Space 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/SpaceX_CRS-1" title="SpaceX CRS-1">SpaceX CRS-1</a><sup class="re
ference" id="cite_ref-sxManifest20120925_28-0"><a href="#cite_note-sxManifest2012092
5-28">[22]</a></sup><br/>(Dragon C103)
</td>
<td>4,700 kg (10,400 lb)
</td>
<td><a href="/wiki/Low_Earth_orbit" title="Low Earth orbit">LEO</a> (<a href="/wiki/
International_Space_Station" title="International Space Station">ISS</a>)
</td>
<td><a href="/wiki/NASA" title="NASA">NASA</a> (<a href="/wiki/Commercial_Resupply_S
ervices" title="Commercial Resupply Services">CRS</a>)
</td>
<td class="table-success" style="background: #9EFF9E; vertical-align: middle; text-a
lign: center;">Success
</td>
<td rowspan="2" style="background:#ececec; text-align:center;"><span class="nowrap">
No attempt</span>
</td></tr>
<tr>
<td><a href="/wiki/Orbcomm_(satellite)" title="Orbcomm (satellite)">Orbcomm-OG2</a><
sup class="reference" id="cite_ref-Orbcomm_29-0"><a href="#cite_note-Orbcomm-29">[2
3]</a></sup>
</td>
<td>172 kg (379 lb)<sup class="reference" id="cite_ref-gunter-og2_30-0"><a href="#ci
te_note-gunter-og2-30">[24]</a></sup>
</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-partial" style="background: #FE9; vertical-align: middle; text-alig
n: center;">Partial failure<sup class="reference" id="cite_ref-nyt-20121030_31-0"><a
href="#cite_note-nyt-20121030-31">[25]</a></sup>
</td></tr>
<tr>
<td colspan="9">CRS-1 was successful, but the <a href="/wiki/Secondary_payload" titl
e="Secondary payload">secondary payload</a> was inserted into an abnormally low orbi
t and subsequently lost. This was due to one of the nine <a href="/wiki/SpaceX_Merli
n" title="SpaceX Merlin">Merlin engines</a> shutting down during the launch, and NAS
```

A declining a second reignition, as per ISS visiting vehicle safety rules, the primary payload owner is contractually allowed to decline a second reignition. NASA stated that this was because SpaceX could not guarantee a high enough likelihood of the second stage completing the second burn successfully which was required to avoid any risk of secondary payload's collision with the ISS.^{[26]}^{[27]}^{[28]}
</td></tr>
<tr>
<th rowspan="2" scope="row" style="text-align:center;">5
</th>
<td>1 March 2013,
15:10
</td>
<td>F9 v1.0^{[7]}
B0007.1^{[8]}
</td>
<td>CCAFS,
SLC-40
</td>
<td>SpaceX CRS-2^{[22]}
(Dragon C104)
</td>
<td>4,877 kg (10,752 lb)
</td>
<td>LEO (ISS)
</td>
<td>NASA (CRS)
</td>
<td class="table-success" style="background: #9EFF9E; vertical-align: middle; text-align: center;">Success
</td>
<td class="table-noAttempt" style="background: #EEE; vertical-align: middle; white-space: nowrap; text-align: center;">No attempt
</td></tr>
<tr>
<td colspan="9">Last launch of the original Falcon 9 v1.0 launch vehicle, first use of the unpressurized trunk section of Dragon.^{[29]}
</td></tr>
<tr>
<th rowspan="2" scope="row" style="text-align:center;">6
</th>
<td>29 September 2013,
16:00^{[30]}
</td>
<td>F9 v1.1<sup class="refer

ence" id="cite_ref-MuskMay2012_13-5">[7]</sup>
B1003^{[8]}
</td>
<td>VAFB,
SLC-4E
</td>
<td>CASSIOPE^{[22]}^{[31]}
</td>
<td>500 kg (1,100 lb)
</td>
<td>Polar orbit LEO
</td>
<td>MDA
</td>
<td class="table-success" style="background: #9EFF9E; vertical-align: middle; text-align: center;">Success^{[30]}
</td>
<td class="table-no2" style="background: #FFE3E3; color: black; vertical-align: middle; text-align: center;">Uncontrolled
<small>(ocean)</small>^{[d]}
</td></tr>
<tr>
<td colspan="9">First commercial mission with a private customer, first launch from Vandenberg, and demonstration flight of Falcon 9 v1.1 with an improved 13-tonne to LEO capacity.^{[29]} After separation from the second stage carrying Canadian commercial and scientific satellites, the first stage booster performed a controlled reentry,^{[32]} and an ocean touchdown test for the first time. This provided good test data, even though the booster started rolling as it neared the ocean, leading to the shutdown of the central engine as the roll depleted it of fuel, resulting in a hard impact with the ocean.^{[30]} This was the first known attempt of a rocket engine being lit to perform a supersonic retro propulsion, and allowed SpaceX to enter a public-private partnership with NASA and its Mars entry, descent, and landing technologies research projects.^{[33]} <small>(more details below)</small>
</td></tr>
<tr>
<th rowspan="2" scope="row" style="text-align:center;">7
</th>
<td>3 December 2013,
22:41^{[34]}
</td>
<td>F9 v1.1
B1004
</td>

```
<td><a href="/wiki/Cape_Canaveral_Space_Force_Station" title="Cape Canaveral Space F
orce 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/SES-8" title="SES-8">SES-8</a><sup class="reference" id="cite_ref
-sxManifest20120925_28-3"><a href="#cite_note-sxManifest20120925-28">[22]</a></sup><
sup class="reference" id="cite_ref-spx-pr_42-0"><a href="#cite_note-spx-pr-42">[35]
</a></sup><sup class="reference" id="cite_ref-aw20110323_43-0"><a href="#cite_note-a
w20110323-43">[36]</a></sup>
</td>
<td>3,170 kg (6,990 lb)
</td>
<td><a href="/wiki/Geostationary_transfer_orbit" title="Geostationary transfer orbi
t">GTO</a>
</td>
<td><a class="mw-redirect" href="/wiki/SES_S.A." title="SES S.A.">SES</a>
</td>
<td class="table-success" style="background: #9EFF9E; vertical-align: middle; text-a
lign: center;">Success<sup class="reference" id="cite_ref-SNMissionStatus7_44-0"><a
href="#cite_note-SNMissionStatus7-44">[37]</a></sup>
</td>
<td class="table-noAttempt" style="background: #EEE; vertical-align: middle; white-s
pace: nowrap; text-align: center;">No attempt<br/><sup class="reference" id="cite_re
f-sf10120131203_45-0"><a href="#cite_note-sf10120131203-45">[38]</a></sup>
</td></tr>
<tr>
<td colspan="9">First <a href="/wiki/Geostationary_transfer_orbit" title="Geostation
ary transfer orbit">Geostationary transfer orbit</a> (GTO) launch for Falcon 9,<sup
class="reference" id="cite_ref-spx-pr_42-1"><a href="#cite_note-spx-pr-42">[35]</a>
</sup> and first successful reignition of the second stage.<sup class="reference" id
="cite_ref-46"><a href="#cite_note-46">[39]</a></sup> SES-8 was inserted into a <a h
ref="/wiki/Geostationary_transfer_orbit" title="Geostationary transfer orbit">Super-
Synchronous Transfer Orbit</a> of 79,341 km (49,300 mi) in apogee with an <a href="/
wiki/Orbital_inclination" title="Orbital inclination">inclination</a> of 20.55° to t
he <a href="/wiki/Equator" title="Equator">equator</a>.
</td></tr></tbody></table>
```

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

```
<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 [17]:
```
column_names = []

# Apply find_all() function with `th` element on first_launch_table
# Iterate each th element and apply the provided extract_column_from_header() to ge
# Append the Non-empty column name (`if name is not None and len(name) > 0`) into a
for element in first_launch_table.find_all('th'):
    name=extract_column_from_header(element)
    if name is not None and len(name)>0:
        column_names.append(name)
```

Check the extracted column names

In [18]:
```
print(column_names)
```

```
['Flight No.', 'Date and time ( )', 'Launch site', 'Payload', 'Payload mass', 'Orbi
t', 'Customer', 'Launch outcome']
```

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 [19]:
```
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 [25]:
```
extracted_row = 0
#Extract each table
for table_number,table in enumerate(soup.find_all('table',"wikitable plainrowheader
   # get table row
    for rows in table.find_all("tr"):
        #check to see if first table heading is as number corresponding to launch a
        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
            # Flight Number value
            # TODO: Append the flight_number into launch_dict with key `Flight No.`
            launch_dict['Flight No.'].append(flight_number)
            #print(flight_number)
            datatimelist=date_time(row[0])

            # Date value
            # TODO: Append the date into launch_dict with key `Date`
            date = datatimelist[0].strip(',')
            launch_dict['Date'].append(date)
            #print(date)

            # Time value
            # TODO: Append the time into launch_dict with key `Time`
```

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

            # Booster version
            # TODO: Append the bv into launch_dict with key `Version Booster`
            bv=booster_version(row[1])
            if not(bv):
                bv=row[1].a.string
            print(bv)
            launch_dict['Version Booster'].append(bv)

            # Launch Site
            # TODO: Append the bv into launch_dict with key `Launch Site`
            launch_site = row[2].a.string
            launch_dict['Launch site'].append(launch_site)
            #print(launch_site)

            # Payload
            # TODO: Append the payload into launch_dict with key `Payload`
            payload = row[3].a.string
            launch_dict['Payload'].append(payload)
            #print(payload)

            # Payload Mass
            # TODO: Append the payload_mass into launch_dict with key `Payload mass
            payload_mass = get_mass(row[4])
            launch_dict['Payload mass'].append(payload_mass)
            #print(payload)

            # Orbit
            # TODO: Append the orbit into launch_dict with key `Orbit`
            orbit = row[5].a.string
            launch_dict['Orbit'].append(orbit)
            #print(orbit)

            # Customer
            # TODO: Append the customer into launch_dict with key `Customer`
            customer = row[6].text.strip()
            launch_dict['Customer'].append(customer)
            #print(customer)

            # Launch outcome
            # TODO: Append the launch_outcome into launch_dict with key `Launch out
            launch_outcome = list(row[7].strings)[0]
            launch_dict['Launch outcome'].append(launch_outcome)
            #print(launch_outcome)

            # Booster landing
            # TODO: Append the launch_outcome into launch_dict with key `Booster la
            booster_landing = landing_status(row[8])
            launch_dict['Booster landing'].append(booster_landing)
            #print(booster_landing)
```

F9 v1.0B0003.1
F9 v1.0B0004.1
F9 v1.0B0005.1
F9 v1.0B0006.1
F9 v1.0B0007.1
F9 v1.1B1003
F9 v1.1
F9 v1.1
F9 v1.1
F9 v1.1
F9 v1.1
F9 v1.1
F9 v1.1
F9 v1.1
F9 v1.1
F9 v1.1
F9 v1.1
F9 v1.1
F9 v1.1
F9 FT
F9 v1.1
F9 FT
F9 FT
F9 FT
F9 FT
F9 FT
F9 FT
F9 FT
F9 FT
F9 FT
F9 FT
F9 FT♺
F9 FT
F9 FT
F9 FT
F9 FTB1029.2
F9 FT
F9 FT
F9 B4
F9 FT
F9 B4
F9 B4
F9 FTB1031.2
F9 B4
F9 FTB1035.2
F9 FTB1036.2
F9 B4
F9 FTB1032.2
F9 FTB1038.2
F9 B4
F9 B4B1041.2
F9 B4B1039.2
F9 B4
F9 B5B1046.1
F9 B4B1043.2
F9 B4B1040.2

F9 B4B1045.2
F9 B5
F9 B5B1048
F9 B5B1046.2
F9 B5
F9 B5B1048.2
F9 B5B1047.2
F9 B5B1046.3
F9 B5
F9 B5
F9 B5B1049.2
F9 B5B1048.3
F9 B5[268]
F9 B5
F9 B5B1049.3
F9 B5B1051.2
F9 B5B1056.2
F9 B5B1047.3
F9 B5
F9 B5
F9 B5B1056.3
F9 B5
F9 B5
F9 B5
F9 B5
F9 B5
F9 B5
F9 B5
F9 B5
F9 B5
F9 B5
F9 B5B1058.2
F9 B5
F9 B5B1049.6
F9 B5
F9 B5B1060.2
F9 B5B1058.3
F9 B5B1051.6
F9 B5
F9 B5
F9 B5
F9 B5
F9 B5 ♺
F9 B5 ♺
F9 B5 ♺
F9 B5 ♺
F9 B5
F9 B5B1051.8
F9 B5B1058.5
F9 B5 ♺
F9 B5 ♺
F9 B5 ♺
F9 B5 ♺
F9 B5 ♺
F9 B5B1060.6

```
F9 B5 ♺
F9 B5B1061.2
F9 B5B1060.7
F9 B5B1049.9
F9 B5B1051.10
F9 B5B1058.8
F9 B5B1063.2
F9 B5B1067.1
F9 B5
```

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

In [32]:
```
df=pd.DataFrame({ key:pd.Series(value) for key, value in launch_dict.items() })
df.head()
```

Out[32]:

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booste landin
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failur
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failur
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt\r
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attemp
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	No attempt\r

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