In [1]:

```python
import requests
from bs4 import BeautifulSoup
page = requests.get("https://forecast.weather.gov/MapClick.php?lat=37.7772&lon=-122.416
soup = BeautifulSoup(page.content, 'html.parser')

# display scrapped data
print(soup.prettify())
```

```
<!DOCTYPE html>
<html class="no-js">
 <head>
  <!-- Meta -->
  <meta content="width=device-width" name="viewport"/>
  <link href="http://purl.org/dc/elements/1.1/" rel="schema.DC"/>
  <title>
   National Weather Service
  </title>
  <meta content="National Weather Service" name="DC.title">
   <meta content="NOAA National Weather Service National Weather Service"
name="DC.description"/>
   <meta content="US Department of Commerce, NOAA, National Weather Servi
ce" name="DC.creator"/>
   <meta content="" name="DC.date.created" scheme="ISO8601"/>
   <meta content="EN-US" name="DC.language" scheme="DCTERMS.RFC1766"/>
   <meta content="weather, National Weather Service" name="DC.keywords"/>
   <meta content="NOAA's National Weather Service" name="DC.publisher"/>
   <meta content="National Weather Service" name="DC.contributor"/>
```

In [2]:

```python
seven_day = soup.find(id="seven-day-forecast")
forecast_items = seven_day.find_all(class_="tombstone-container")
tonight = forecast_items[0]
print(tonight.prettify())
```

```
<div class="tombstone-container">
 <p class="period-name">
  Today
  <br/>
  <br/>
 </p>
 <p>
  <img alt="Today: Sunny, with a high near 70. West wind 5 to 14 mph, with g
usts as high as 18 mph. " class="forecast-icon" src="newimages/medium/few.pn
g" title="Today: Sunny, with a high near 70. West wind 5 to 14 mph, with gus
ts as high as 18 mph. "/>
 </p>
 <p class="short-desc">
  Sunny
 </p>
 <p class="temp temp-high">
  High: 70 °F
 </p>
</div>
```

In [3]:

```python
period = tonight.find(class_="period-name").get_text()
short_desc = tonight.find(class_="short-desc").get_text()
temp = tonight.find(class_="temp").get_text()
print(period)
print(short_desc)
print(temp)
```

```
Today
Sunny
High: 70 °F
```

In [4]:

```python
#extract the title attribute from the img tag.
img = tonight.find("img")
desc = img['title']
print(desc)
```

```
Today: Sunny, with a high near 70. West wind 5 to 14 mph, with gusts as high
as 18 mph.
```

In [5]:

```python
#extract all information from the Page
period_tags = seven_day.select(".tombstone-container .period-name")
periods = [pt.get_text() for pt in period_tags]
periods
```

Out[5]:

```
['Today',
 'Tonight',
 'Monday',
 'MondayNight',
 'Tuesday',
 'TuesdayNight',
 'Wednesday',
 'WednesdayNight',
 'Thursday']
```

In [6]:

```
#get other three fields
short_descs = [sd.get_text() for sd in seven_day.select(".tombstone-container .short-de
temps = [t.get_text() for t in seven_day.select(".tombstone-container .temp")]
descs = [d["title"] for d in seven_day.select(".tombstone-container img")]
print(short_descs)
print(temps)
print(descs)
```

```
['Sunny', 'Partly Cloudy', 'Sunny', 'Mostly Clear', 'Sunny', 'Mostly Clear',
'Mostly Sunny', 'Partly Cloudy', 'Mostly Sunny']
['High: 70 °F', 'Low: 55 °F', 'High: 70 °F', 'Low: 56 °F', 'High: 74 °F', 'L
ow: 57 °F', 'High: 69 °F', 'Low: 56 °F', 'High: 67 °F']
['Today: Sunny, with a high near 70. West wind 5 to 14 mph, with gusts as hi
gh as 18 mph. ', 'Tonight: Partly cloudy, with a low around 55. Southwest wi
nd 5 to 10 mph, with gusts as high as 20 mph. ', 'Monday: Sunny, with a high
near 70. Southwest wind 5 to 10 mph. ', 'Monday Night: Mostly clear, with a
low around 56. West southwest wind 9 to 14 mph becoming light southwest  aft
er midnight. Winds could gust as high as 18 mph. ', 'Tuesday: Sunny, with a
high near 74. Light south southwest wind becoming west southwest 5 to 10 mph
in the afternoon. ', 'Tuesday Night: Mostly clear, with a low around 57.',
'Wednesday: Mostly sunny, with a high near 69.', 'Wednesday Night: Partly cl
oudy, with a low around 56.', 'Thursday: Mostly sunny, with a high near 6
7.']
```

In [7]:

```
import pandas as pd
weather = pd.DataFrame({
    "period": periods,
    "short_desc": short_descs,
    "temp": temps,
    "desc":descs
})
weather
```

Out[7]:

	period	short_desc	temp	desc
0	Today	Sunny	High: 70 °F	Today: Sunny, with a high near 70. West wind 5...
1	Tonight	Partly Cloudy	Low: 55 °F	Tonight: Partly cloudy, with a low around 55. ...
2	Monday	Sunny	High: 70 °F	Monday: Sunny, with a high near 70. Southwest ...
3	MondayNight	Mostly Clear	Low: 56 °F	Monday Night: Mostly clear, with a low around ...
4	Tuesday	Sunny	High: 74 °F	Tuesday: Sunny, with a high near 74. Light sou...
5	TuesdayNight	Mostly Clear	Low: 57 °F	Tuesday Night: Mostly clear, with a low around...
6	Wednesday	Mostly Sunny	High: 69 °F	Wednesday: Mostly sunny, with a high near 69.
7	WednesdayNight	Partly Cloudy	Low: 56 °F	Wednesday Night: Partly cloudy, with a low aro...
8	Thursday	Mostly Sunny	High: 67 °F	Thursday: Mostly sunny, with a high near 67.

In [8]:

```python
import re
temp_nums = weather["temp"].str.extract("(?P<temp_num>\d+)", expand=False)
weather["temp_num"] = temp_nums.astype('int')
temp_nums
weather["temp_num"].mean()
```

Out[8]:

63.77777777777778

In [9]:

```python
writer = pd.ExcelWriter('file_name.xlsx', engine='xlsxwriter')
df = pd.DataFrame(weather)
df.to_excel(writer)
writer.save()
```

In [ ]:

```python

```