

```
In [18]: from selenium import webdriver
import pandas as pd
from datetime import datetime
```

```
In [19]: url = 'https://www.livetraffic.com/incident-list/sydney'
driver = webdriver.Chrome(r'C:\Users\gelza\Downloads\chromedriver_win32\chromedriver.exe')
driver.get(url)
```

```
In [20]: events = driver.find_elements_by_class_name('incident-list')
```

```
In [ ]: newLocation, newTitle, newDate, isActive = ([] for i in range(4))

for event in events:
    location = event.find_elements_by_class_name('incident-title')
    title = event.find_elements_by_class_name('bold-subheading')
    date = event.find_elements_by_class_name('heading-date-time')

    for value in location:
        if value is None:
            newLocation.append("No Location")
        else:
            newLocation.append(value.text)

    for value in title:
        if value is None:
            newTitle.append("No Title")
        else:
            newTitle.append(value.text)

    for value in date:
        if value is None:
            newDate.append("No date")
        else:
            newDate.append(value.text)

    for value in title:
        # The naive method, use if below fails

        if value is None:
            isActive.append(0)
        else:
            if("hedule" in value.text or "HEDULE" in value.text):
                isActive.append(0)
            else:
                isActive.append(1)

        for value in date:
            startDate = value.text[4:15].strip() # Only works if format is "Day Month Year"
            startDate = datetime.strptime(startDate, '%d %b %Y')
            if(len(value.text) > 20):
                endDate = value.text[-11:].strip()
                endDate = datetime.strptime(endDate, '%d %b %Y')
            currentDate = datetime.now()
            if(currentDate > startDate and endDate > currentDate):
                isActive.append(1)
            else:
                isActive.append(0)

    # print(newLocation, newTitle, newDate, isActive)
    # Uncomment above to format as [] [] []

    # res = list(zip(newLocation, newTitle, newDate, isActive))
    # res # Print 'res' for format [(locations, titles, dates, isActives)]

df = pd.DataFrame(res, columns = ["Location", "Title", "Date", "Is Active" ])

df #Formatted in dataframe
```

```
In [ ]:
```