https://github.com/someshkr/weather-scrape/blob/main/web_scraping%20weather.ipynb

Web scrapping Fresco play

```
Dates_r = pd.date_range(start = '1/1/2009',end = '11/28/2018',freq = 'M')
dates = [str(i)[:4] + str(i)[5:7] for i in Dates_r]
dates[0:5]

df_list = []
index = []
for k in range(len(dates)):
    url = "http://www.estesparkweather.net/archive_reports.php?date="
    url += dates[k]
    page  = requests.get(url)
    soup =  BeautifulSoup(page.content,'html.parser')

    table = soup.find_all('table')
    #type(table)

    raw_data = [row.text.splitlines() for row in table]
    raw_data = raw_data[:-9]
    raw_data
    for i in range(len(raw_data)):
        raw_data[i] = raw_data[i][2:len(raw_data[i]):3]
    for i in range(len(raw_data)):
```

```
        c = ['.'.join(re.findall("\d+",str(raw_data[i][j].split()[:5])))for j in
range(len(raw_data[i]))]
        df_list.append(c)
        index.append(dates[k] + c[0])
    f_index = [index[i] for i in range(len(index)) if len(index[i]) > 6]
    data = [df_list[i][1:] for i in range(len(df_list)) if len(df_list[i][1:]) == 19]
    #data=np.array(data)


from datetime import datetime
final_index = [datetime.strptime(str(f_index[i]), '%Y%m%d').strftime('%Y-
%m-%d') for i in range(len(f_index))]
final_index=pd.Series(final_index,dtype='datetime64[ns]')


col = ['Average temperature (°F)', 'Average humidity (%)',
'Average dewpoint (°F)', 'Average barometer (in)',
'Average windspeed (mph)', 'Average gustspeed (mph)',
'Average direction (°deg)', 'Rainfall for month (in)',
'Rainfall for year (in)', 'Maximum rain per minute',
'Maximum temperature (°F)', 'Minimum temperature (°F)',
'Maximum humidity (%)', 'Minimum humidity (%)', 'Maximum pressure',
'Minimum pressure', 'Maximum windspeed (mph)',
'Maximum gust speed (mph)', 'Maximum heat index (°F)']
df = pd.DataFrame(data, columns = col, index = final_index, dtype='float64')
df.index = pd.to_datetime(df.index)
```

```
df=df[:-3]
df.head()


import pickle
with open("dataframe.pk", "wb") as file:
   pickle.dump(df, file)
```

The regex r"[-*] ?([^-*].*?) ?[-*]" will look for:

[-*] ? - or * followed by optional space

([^-*].*?) grouping any character different than - or * as few as possible

?[-*] optional space followed by - or *