```python
# -*- coding: utf-8 -*-
"""
Created on Sat Jun 25 22:32:16 2022

@author: Jordan Matthew Pauley

Note: The purpose of this program is to first, create a dataset containing
the locations of all U.S. Embassy cities around the world; and second, to
get and attach their GPS coordinates.
"""
# Importing libraries
import requests
import pandas as pd
import numpy as np
from geopy.exc import GeocoderTimedOut
from geopy.geocoders import Nominatim

# Setting default dataframe display option to max
pd.set_option("display.max_rows", None, "display.max_columns", None)

#------------------------------------------------------------------------------
# Reading HTML code from state department 'list of posts' webpage into Python
r = requests.get('https://travel.state.gov/content/travel/en/us-visas/visa-information-resources/li
print(r.text[0:1000]) # Results valid

# Parsing HTML using Beautiful Soup library
from bs4 import BeautifulSoup
soup = BeautifulSoup(r.text, 'html.parser')

# Parsing all 'title' key values located within HTML 'a' tags
# storing results within a list
posts = []
for line in soup.find_all('a'):
    data = line.get('title')
    posts.append(data)

# Finding index of the first embassy location in list
i = posts.index('Abidjan, Côte dIvoire') # Index of first site location = 63

# Finding index of the last embassy location in list
j = posts.index('Zagreb, Croatia') + 1 # Index of last site location = 219 + 1

# Slicing and storing Embassy locations only
posts = posts[i:j]

# Filtering null values from list
posts = list(filter(None, posts))

# Sorting list alphabetically
posts.sort() # Beirut, Lebanon has a duplicate at index 19

del posts[19] # Beirut, Lebanon duplicate deleted

# Creating function to update post locations
def updater(x,y):
    k = posts.index(x)
    posts[k] = y
```

```
# Updating post locations
updater('U.S. Embassy Madras', 'Chennai, India')
updater('Havana', 'Havana, Cuba')
updater('Hong Kong and Macau', 'Hong Kong')
updater('Panama City, Panama', 'Panamá City, Panama')
updater('American Institute in Taiwan', 'Taipei City, Taiwan')
updater('Tel Aviv, Isreal', 'Tel Aviv, Israel')

# Removing duplicate locations
posts.remove('Havana, Cuba via Guyana')
posts.remove('Lisbon, Portugal via Paris, France')
posts.remove('Ponta Delgado, Azores Archipelago of Portugal via Paris France')
posts.remove('Tashkent, Uzbekistan - Diversity Visa')

# Reading posts list into pandas dataframe
df_Posts = pd.DataFrame(posts, columns = ['Post Locations'])

# Checking for Null values
df_Posts.isnull().sum() # 0 Null values

#-------------------------------------------------------------------------------
# Creating program to retrieve the GPS coordinates of post locations
# Defining dictionary and inserting post locations
cities = {'City': posts}

# Converting dictionary to dataframe
df_city = pd.DataFrame(cities)

# Creating lists to store the latitude and longitude of embassy post locations
longitude = []
latitude = []

# Creating function to find GPS coordinates for each post location
def findGeocode(city):
    try:
        geolocator = Nominatim(user_agent="posts_scraper.py")
        return geolocator.geocode(city)
    except GeocoderTimedOut:
        return findGeocode(city)

# Finding and storing the GPS coordinates for each post location
for l in (df_city["City"]):
    if findGeocode(l) != None:
        loc = findGeocode(l)
        latitude.append(loc.latitude)
        longitude.append(loc.longitude)
    else:
        latitude.append(np.nan)
        longitude.append(np.nan)

# Creating dataframes from latitude and longitude lists
df_lon = pd.DataFrame(longitude, columns = ['Longitude'])
df_lat = pd.DataFrame(latitude, columns = ['Latitude'])

# Appending post location dataframe with latitude and longitude information
df_Posts = pd.concat([df_Posts, df_lat, df_lon], axis=1)
```

```
# Writing results to .csv file
df_Posts.to_csv('embassy_posts.csv', index=False)
```