TERRORISM DATA ANALYSIS

ABOUT DATASET:

Given file "terrorismData.csv"

It is an open-source database including information on terrorist attacks around the world from 1970 through 2017. This dataset includes systematic data on domestic as well as international terrorist incidents that have occurred during this time period

TABLE OF CONTENTS

• Q1 : Print first three lines.

• Q2 : Print all column names.

• Q3 : Print the country names from first 10 rows (in different lines).

• Q4 : Find and Print the total number of Wounded people. Print the count as integer value.

• Q5 : Find and print the total number of wounded people who are from country "India". Print the count as integer value.

• Q6 : Find the number of attack held between day 10 and day 20?(ignoring the year and month) (including both day) Print count of NumberOFAttack as integer value.

• Q7 : Find the number of attack held between 1 Jan 2010 and 31 Jan 2010?(including both date). Note Ignore the case where day is 0 Print count of NumberOFAttack as integer value.

• Q8 : As we knew the Kargil (in Jammu and Kashmir) War that took place between May 1999 and July 1999 (3 Months) ,so there was a huge conflict in Kashmir Valley during this period. In this dataset, there is no information regarding the war between the two countries to find out the casualty during the war. So find out the attack in this period in which maximum casualties happened. Print the count of casualties (as integer), city in which that attack happened and name of attack group.

• Q9 : ind the casualty in the Red Corridor States ? Mainly Red corridor states include Jharkhand, Odisha, Andhra Pradesh, and Chhattisgarh.Print count of Casualty as integer value. Note: Casualty=Killed +Wounded.

• Q10 : Find top 5 Indian Cities which has most number of casualties ? Print top 5 cities along with total casualties in that city. Print count of Casualty as integer value. Note: Ignoring the City which is Unknown.

• Q11 : Find the most frequent day of attack in a terrorismDataset ? Note: Here np.unique can be used. Print count of frequent day and number of attack as Integer value.

• Q12 : The Most Dangerous city in Jammu and Kashmir and the terrorist group which is most active in that city? Print count of number of attacks in that city as integer value. Note:Ignoring the Unknown Terrorist Group.Here Dangerous related with the number of terrorist attacks.

• Q13 : There was formation of new government in India on 26 May 2014. So current government's span is from 26th May 2014 to current. Find out two things from this period- A)

Total number of attacks done in this period in India. Find this count as integer. B) Which Terrorist group was most active in this period in India. Most active means, group which has done maximum number of attacks. Ignore the Unknown group.

• Q14 : Find the frequency of the Casualty in Red Corridor states and in Jammu and Kashmir ? Here Frequency is (Total Casualty/Total Number of a year) Print frequency as integer value. Note:Red Corridor state includes Jharkhand, Odisha, Andhra Pradesh, and Chhattisgarh. Here Casualty=Killed +Wounded.Don't fill the nan value present in Killed and Wounded feature.

• Q15 : Most Deadliest attack in a history of HumanKind? Print count of Killed people as integer value. Note: Here Deadliest attack means, in which the most number of people killed.

• Q16 : Find out the Country with Highest Number of Terror Attack and in which year the most number of terrorist attack happened in that country ? Print count of terror attacks as integer value.

Importing required files and libraries

```
import numpy as np
import pandas as pd
import csv
```

Q1 : Print first three lines

```
with open("terrorismData.csv") as file_obj:
    file_data = csv.reader(file_obj)
    next(file_data)  # Skip the header row

    # Print the first 3 rows
    for _ in range(3):
        row = next(file_data)
        print(" ".join(row))
```

```
1970 7 2 Dominican Republic  Central America & Caribbean Santo Domingo
18.456792 -69.951164 Assassination 1.0 0.0 Julio Guzman  MANO-D
Private Citizens & Property Unknown
1970 0 0 Mexico Federal North America Mexico city 19.371887 -99.086624
Hostage Taking (Kidnapping) 0.0 0.0 Nadine Chaval, daughter  23rd of
September Communist League Government (Diplomatic) Unknown
1970 1 0 Philippines Tarlac Southeast Asia Unknown 15.478598000000002
120.599741 Assassination 1.0 0.0 Employee  Unknown Journalists & Media
Unknown
```

Q2 : Print all column names

```
with open('terrorismData.csv','r') as file_obj:
    file_data = csv.reader(file_obj)
    file_data1 = next(file_data)
    for header in file_data1:
        print(header)
```

```
Year
Month
Day
Country
State
Region
City
Latitude
Longitude
AttackType
Killed
Wounded
Target
Summary
Group
Target_type
Weapon_type
```

Q3 : Print the country names from first 10 rows (in different lines)

```
with open("terrorismData.csv","r")as file_obj:
    file_data= csv.DictReader(file_obj,skipinitialspace=True)
    count = 0
    for row in file_data:
        print(row['Country'])

        count+= 1

        if count >=10:
            break
```

```
Dominican Republic
Mexico
Philippines
Greece
Japan
United States
Uruguay
United States
United States
United States
```

Q4 : Find and Print the total number of Wounded people. Print the count as integer value.

```
with open('terrorismData.csv','r',encoding="utf-8") as file_obj:
    file_data = csv.reader(file_obj,skipinitialspace = True)
    file_list = list(file_data)

wounded = []
```

```
for row in file_list[1:]:
    value = row[10]
    if value !='':
        wounded.append(float(value))

print(int(sum(wounded)))
```

```
411868
```

Q5 : Find and print the total number of wounded people who are from country "India". Print the count as integer value.

```
with open('terrorismData.csv','r',encoding="utf-8") as file_obj:
    file_list = list(csv.DictReader(file_obj, delimiter = ',',
skipinitialspace = True))

country_wounded = {}
for row in file_list:
    key=row['Country']
    value=row['Wounded']
    if value != '':
        value = int(float(value))
    else:
        value = 0
    country_wounded[key] = country_wounded.get(key,0) + value

print(country_wounded['India'])
```

```
28980
```

Q6 : Find the number of attack held between day 10 and day 20?(ignoring the year and month) (including both day) Print count of NumberOFAttack as integer value.

```
File_obj = open('terrorismData.csv','r',encoding="utf-8")
data = csv.DictReader(File_obj,skipinitialspace = True)

day = []
for row in data:
    day.append(row['Day'])
np_day = np.array(day)
np_day = np.array(np_day,dtype = int)
ans = np_day[(np_day >= 10)&(np_day <=20)]
print(len(ans))
```

```
66330
```

Q7 : Find the number of attack held between 1 Jan 2010 and 31 Jan 2010?(including both date). Note Ignore the case where day is 0 Print count of NumberOFAttack as integer value.

```
File_obj = open('terrorismData.csv','r',encoding="utf-8")
data = csv.DictReader(File_obj,skipinitialspace = True)

Month = []
Year = []
Day = []
for row in data:
    Month.append(row['Month'])
    Year.append(row['Year'])
    Day.append(row['Day'])

np_Month = np.array(Month)
np_Year = np.array(Year)
np_Day = np.array(Day, dtype = float)
np_Day = np.array(Day, dtype = int)


np_Day=np_Day[(np_Month =='1') & (np_Year =='2010')]

np_Day = np_Day[(np_Day >=1) & (np_Day <=31)]

print(len(np_Day[np_Day!=0]))
```

```
271
```

Q8 : As we knew the Kargil (in Jammu and Kashmir) War that took place between May 1999 and July 1999 (3 Months) ,so there was a huge conflict in Kashmir Valley during this period. In this dataset, there is no information regarding the war between the two countries to find out the casualty during the war. So find out the attack in this period in which maximum casualties happened. Print the count of casualties (as integer), city in which that attack happened and name of attack group.

```
max_casualty = 0
max_casualty_city = ''
max_casualty_group = ''

with open('terrorismData.csv', 'r', encoding='utf-8') as file:
    file_obj = csv.DictReader(file, skipinitialspace=False)

    killed = []
    wounded = []
    city = []
    group = []

    for row in file_obj:
        if "1999" in row["Year"] and row["Month"] in ["5", "6", "7"]:
            if row["City"] != "Unknown" and row["Group"] != "Unknown"
and "Jammu and Kashmir" in row["State"]:
                killed_val = float(row["Killed"]) if row["Killed"]
else 0.0
```

```
                wounded_val = float(row["Wounded"]) if row["Wounded"]
else 0.0

                killed.append(killed_val)
                wounded.append(wounded_val)
                city.append(row["City"])
                group.append(row["Group"])

    casualties = np.array(killed) + np.array(wounded)

    max_casualty = int(np.max(casualties))
    max_casualty_arg = np.argmax(casualties)

    max_casualty_city = city[max_casualty_arg]
    max_casualty_group = group[max_casualty_arg]

print(max_casualty, max_casualty_city, max_casualty_group)
```

```
22 Kargil District Separatists
```

Q9 : ind the casualty in the Red Corridor States ? Mainly Red corridor states include Jharkhand, Odisha, Andhra Pradesh, and Chhattisgarh.Print count of Casualty as integer value. Note: Casualty=Killed +Wounded

```
with open("terrorismData.csv", encoding="utf8") as file_obj:
    file_data = csv.DictReader(file_obj, skipinitialspace=True)

    killed = []
    wounded = []

    red_corridor_states = ['Jharkhand', 'Odisha', 'Chhattisgarh',
'Andhra Pradesh']

    for row in file_data:
        if row['State'].strip() in red_corridor_states:
            killed_value = row['Killed'].strip()
            wounded_value = row['Wounded'].strip()

            # Replace empty strings with '0'
            if not killed_value:
                killed_value = '0'
            if not wounded_value:
                wounded_value = '0'

            killed.append(float(killed_value))
            wounded.append(float(wounded_value))

    np_killed = np.array(killed, dtype=float)
    np_wounded = np.array(wounded, dtype=float)
```

```
    casualties = np_killed + np_wounded
    total_casualties = int(np.sum(casualties))

    print(total_casualties)
```

```
5628
```

Q10 : Find top 5 Indian Cities which has most number of casualties ? Print top 5 cities along with total casualties in that city. Print count of Casualty as integer value. Note: Ignoring the City which is Unknown.

```
with open('terrorismData.csv', encoding='utf8') as file_obj:
    file_data=csv.DictReader(file_obj, skipinitialspace=True)
    killed=[]
    wounded=[]
    city=[]
    for row in file_data:
        if 'India' in row['Country'] and 'Unknown' not in row['City']:
            city.append(row['City'])
            wounded.append(row['Wounded'])
            killed.append(row['Killed'])
    np_wounded=np.array(wounded)
    np_killed=np.array(killed)
    np_city=np.array(city)

    np_killed[np_killed=='']='0.0'
    np_wounded[np_wounded=='']='0.0'
    np_killed=np.array(np_killed, dtype='float')
    np_wounded=np.array(np_wounded, dtype='float')
    np_casuality=np.array(np_wounded+np_killed, dtype='int')
    city_casualties = zip(np_city, np_casuality)



    citydic = {}

for city, casualties in city_casualties:
    if city in citydic:
        citydic[city] += casualties
    else:
        citydic[city] = casualties

for _ in range(5):
    city, count = max(citydic.items(), key=lambda x: x[1])
    print(city, count)
    del citydic[city]
```

```
Srinagar 3134
New Delhi 2095
Mumbai 2016
```

```
Jammu 1119
Guwahati 822
```

Q11 : Find the most frequent day of attack in a terrorismDataset ? Note: Here np.unique can be used. Print count of frequent day and number of attack as Integer value.

```
file_obj = open('terrorismData.csv',encoding='utf8')
data = csv.DictReader(file_obj, skipinitialspace = True)

days = []


for row in data:
    days.append(row['Day'])

np_day = np.array(days)
# Use np.unique to get unique days and their counts
unique_days, counts = np.unique(np_day, return_counts=True)

# Find the index of the most frequent day
most_frequent_index = np.argmax(counts)

# Get the most frequent day and its count
most_frequent_day = unique_days[most_frequent_index]
number_of_attacks = counts[most_frequent_index]

# Print the result
print(f"{most_frequent_day} {number_of_attacks}")
```

```
15 6500
```

Q12 : The Most Dangerous city in Jammu and Kashmir and the terrorist group which is most active in that city? Print count of number of attacks in that city as integer value. Note:Ignoring the Unknown Terrorist Group.Here Dangerous related with the number of terrorist attacks.

```
path='terrorismData.csv'
df_terrorism=pd.read_csv(path,encoding='ISO-8859-1')
df_terrorism=df_terrorism[df_terrorism['State']=='Jammu and Kashmir']
city_list=df_terrorism['City'].value_counts()
city=city_list.index[0]
attack=city_list.values[0]
df_terrorism=df_terrorism[df_terrorism['City']==city]
group=df_terrorism['Group'].value_counts().index[1]
print(city,attack,group)
```

```
Srinagar 657 Muslim Separatists
```

Q13 : There was formation of new government in India on 26 May 2014. So current government's span is from 26th May 2014 to current. Find out two things from this period-

1. Total number of attacks done in this period in India. Find this count as integer.
2. Which Terrorist group was most active in this period in India. Most active means, group which has done maximum number of attacks. 3.Ignore the Unknown group.

```
td = pd.read_csv("terrorismData.csv")
df = td.copy()

df = df[df.Country == "India"]

a = df[df.Day>=26]
b = a[df.Month == 5]
c = b[df.Year == 2014]

ans1 = c.shape[0]

d = df[df.Month > 5]
e = d[df.Year == 2014]
ans2 = e.shape[0]


f = df[df.Year > 2014]
ans3 = f.shape[0]

print(ans1 + ans2 + ans3,end=" ")

df = df[df.Year>2014]
df = df[df.Group != "Unknown"]
group = df.Group.describe().top

print(group)
```

```
3336 Maoists

C:\Users\Luckshaya Kem\AppData\Local\Temp\
ipykernel_17452\74955651.py:7: UserWarning: Boolean Series key will be
reindexed to match DataFrame index.
  b = a[df.Month == 5]
C:\Users\Luckshaya Kem\AppData\Local\Temp\
ipykernel_17452\74955651.py:8: UserWarning: Boolean Series key will be
reindexed to match DataFrame index.
  c = b[df.Year == 2014]
C:\Users\Luckshaya Kem\AppData\Local\Temp\
ipykernel_17452\74955651.py:13: UserWarning: Boolean Series key will
be reindexed to match DataFrame index.
  e = d[df.Year == 2014]
```

Q14 : Find the frequency of the Casualty in Red Corridor states and in Jammu and Kashmir ?Here Frequency is (Total Casualty/Total Number of a year) Print frequency as integer value. Note:Red Corridor state includes Jharkhand, Odisha, Andhra Pradesh, and Chhattisgarh. Here Casualty=Killed +Wounded.Don't fill the nan value present in Killed and Wounded feature.

```
df = pd.read_csv("terrorismData.csv")

red_corridor_states = ["Jharkhand", "Odisha", "Andhra Pradesh",
"Chhattisgarh"]
df = df[df["State"].isin(red_corridor_states + ["Jammu and Kashmir"])]

df["Casualty"] = df["Killed"] + df["Wounded"]
df_red = df.loc[df["State"].isin(red_corridor_states)]
df_jk = df.loc[df["State"] == "Jammu and Kashmir"]

df_red = df_red.groupby("Year")["Casualty"].sum()
df_jk = df_jk.groupby("Year")["Casualty"].sum()

frequency_red = df_red.sum() / len(df_red)
frequency_jk = df_jk.sum() / len(df_jk)
print(int(frequency_red), int(frequency_jk))
```

```
170 395
```

Q15 : Most Deadliest attack in a history of HumanKind? Print count of Killed people as integer value. Note: Here Deadliest attack means, in which the most number of people killed.

```
df = pd.read_csv("terrorismData.csv")

max_killed = df["Killed"].max()
row = df[df["Killed"] == max_killed]
country = row["Country"].values[0]
group = row["Group"].values[0]

print(int(max_killed), country, group)
```

```
1570 Iraq Islamic State of Iraq and the Levant (ISIL)
```

Q16 : Find out the Country with Highest Number of Terror Attack and in which year the most number of terrorist attack happened in that country ? Print count of terror attacks as integer value.

```
td=pd.read_csv('terrorismData.csv')
df=td.copy()
df=df[df.Country==df.Country.describe().top]
count=df.shape[0]
country=df.Country.describe().top
y={}
for i in df.Year:
    if i in y.keys():
        y[i]+=1
    else:
        y[i]=1
cnt=0
year=0
```

```
for i in y.keys():
    if cnt<y[i]:
        cnt=y[i]
        year=i
print(country, count, year)
```

```
Iraq 24636 2014
```