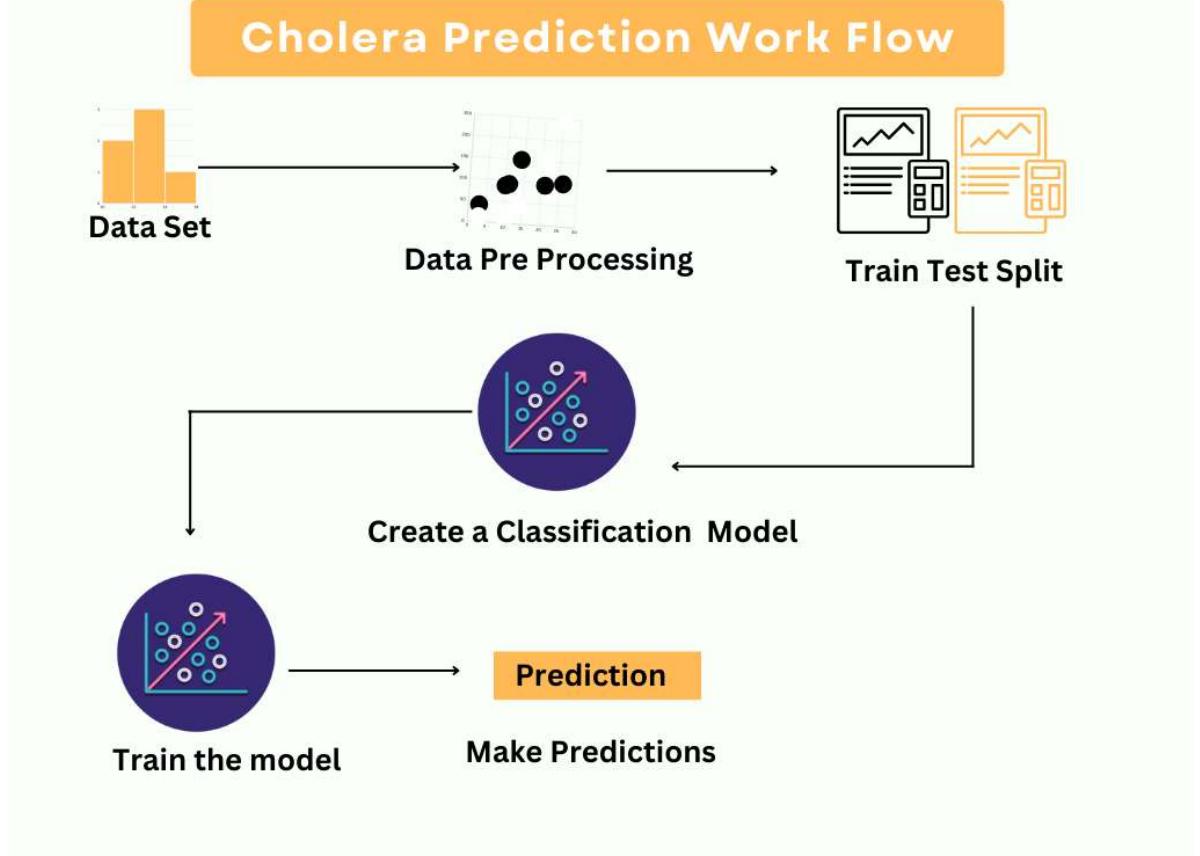


```
In [105]: 1 from IPython.display import Image
2
3 # Provide the file path to your image
4 Image(filename="C:/Users/Yusuf Nasir/OneDrive/Desktop/Zakiya/Cholera Pred:
5
```

Out[105]:



Importing Dependencies

```
In [1]: 1 import numpy as np # Linear algebra
2 import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
3
4 # Input data files are available in the read-only "../input/" directory
5 # For example, running this (by clicking run or pressing Shift+Enter) will
6
7 import os
8 for dirname, _, filenames in os.walk('/kaggle/input'):
9     for filename in filenames:
10         print(os.path.join(dirname, filename))
```

1

Reading the cholera Dataset as input

```
In [5]: 1 file = pd.read_csv(r'C:\Users\Yusuf Nasir\OneDrive\Desktop\Zakiya\data.csv')
2 file.head()
```

Out[5]:

	Country	Year	Number of reported cases of cholera	Number of reported deaths from cholera	Cholera case fatality rate	WHO Region
0	Afghanistan	2016	677	5	0.7	Eastern Mediterranean
1	Afghanistan	2015	58064	8	0.01	Eastern Mediterranean
2	Afghanistan	2014	45481	4	0	Eastern Mediterranean
3	Afghanistan	2013	3957	14	0.35	Eastern Mediterranean
4	Afghanistan	2012	12	0	0.1	Eastern Mediterranean

IDENTIFYING THE DATA TYPE

```
In [6]: 1 file.dtypes
```

Out[6]:

Country	object
Year	int64
Number of reported cases of cholera	object
Number of reported deaths from cholera	object
Cholera case fatality rate	object
WHO Region	object
dtype: object	

IDENTIFYING DATA GAPS

```
In [7]: 1 file.isnull().sum()
```

Out[7]:

Country	0
Year	0
Number of reported cases of cholera	22
Number of reported deaths from cholera	117
Cholera case fatality rate	127
WHO Region	0
dtype: int64	

DATA CLEANING

```
In [8]: 1 file.replace(np.nan,0,regex=True,inplace=True)
```

DATA CLEAN CHECK

In [9]: 1 file.isnull().sum()

Out[9]:

Country	0
Year	0
Number of reported cases of cholera	0
Number of reported deaths from cholera	0
Cholera case fatality rate	0
WHO Region	0
dtype: int64	

CHECKING IF UNKNOWN VALUES EXISTS

In [10]: 1 file[(file['Number of reported deaths from cholera']=='Unknown') | (file[

Out[10]:

Country	Year	Number of reported cases of cholera	Number of reported deaths from cholera	Cholera case fatality rate	WHO Region
761	Germany	2016	1	Unknown	Unknown
					Europe

CORRECTING THE UNKNOWN VALUES

In [11]: 1 file.replace('Unknown',0,regex=True,inplace=True)

IDENTIFYING THE TOTAL NUMBER OF COUNTRIES IN THE DATA SET

In [12]: 1 country_list=file.Country.unique()
2 len(file.Country.unique())

Out[12]: 162

Upon checking the data, I observed an issue with this particular entry

In [13]: 1 file[file['Number of reported cases of cholera']=="3 5"]

Out[13]:

Country	Year	Number of reported cases of cholera	Number of reported deaths from cholera	Cholera case fatality rate	WHO Region
1059	Iraq	2016	3 5	0 0	0.0 0.0
					Eastern Mediterranean

This entry needs to be cleaned. Since the fatality rate is given as 0.0 and 0.0, we assume, that all the entries here to be 0 and assign the variables accordingly.

In [14]:

```
1 file['Number of reported cases of cholera'] = file['Number of reported cas  
2 file['Number of reported deaths from cholera'] = file['Number of reported  
3 file['Cholera case fatality rate'] = file['Cholera case fatality rate'].s
```

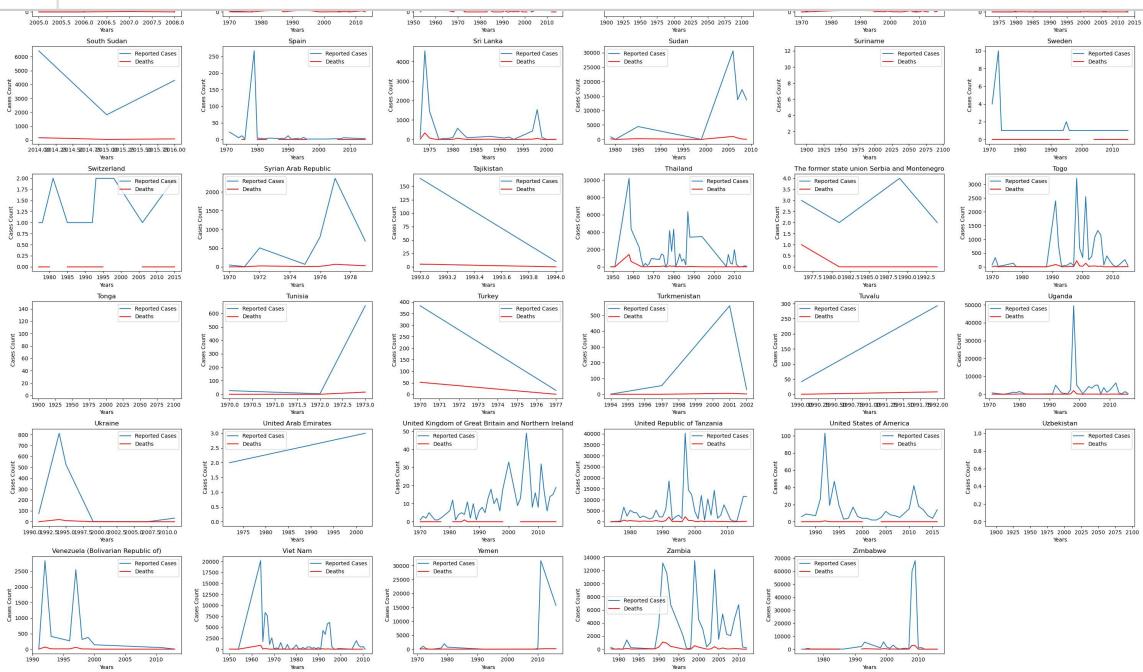
Country wise visualization of Total Number of Reported Cases and Total Number of Deaths over the Years

In [15]:

```

1 import seaborn as sns
2 import matplotlib.pyplot as plt
3 fig = plt.figure(figsize=(30,90))
4 for c,num in zip(file.Country.unique(), np.arange(1,len(file.Country.unique())):
5     file_req=file[file['Country']==c]
6     ax = fig.add_subplot(27,6,num)
7     x=file_req['Year']
8     y1=pd.to_numeric(file_req['Number of reported cases of cholera'])
9     ax.plot(x,y1)
10    y2=pd.to_numeric(file_req['Number of reported deaths from cholera'])
11    ax.plot(x,y2,color='r')
12    ax.legend(['Reported Cases','Deaths'])
13    ax.set_title(c)
14    ax.set_xlabel('Years')
15    ax.set_ylabel('Cases Count')
16
17 plt.tight_layout()
18 plt.show()

```



In []:

1

In [55]:

```

1 import warnings
2 warnings.filterwarnings('ignore', category=DeprecationWarning)
3
4 # Your existing code follows
5 y11 = file_req.groupby('Year').apply(lambda x: np.sum(pd.to_numeric(x['Number of reported cases of cholera'])))
6 y21 = file_req.groupby('Year').apply(lambda x: np.sum(pd.to_numeric(x['Number of reported deaths from cholera'])))
7

```

WHO Region wise visualization of Total Number of Reported Cases and Total Number of Deaths

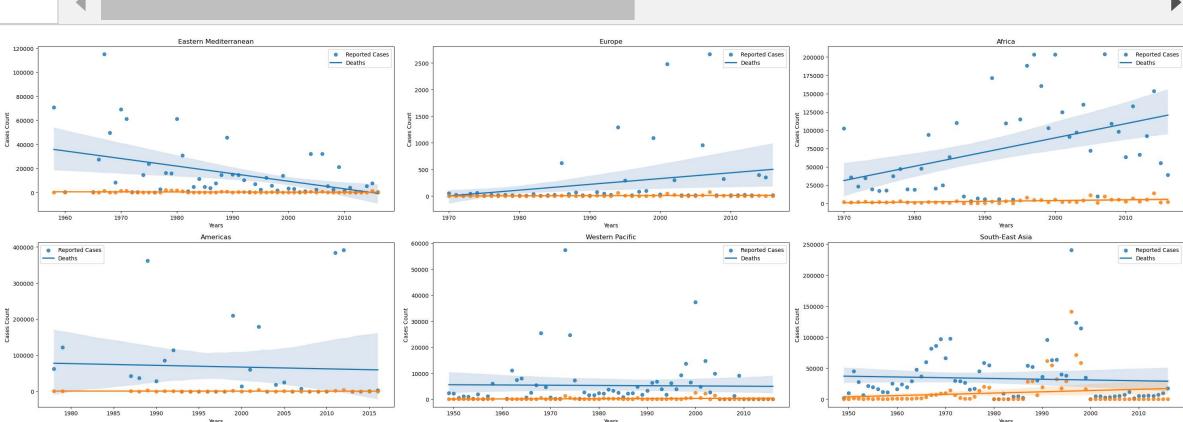
over the Years

In [56]:

```

1 fig = plt.figure(figsize=(30,10))
2 for c, num in zip(file['WHO Region'].unique(), np.arange(1, 1 + len(file[
3     file_req = file[file['WHO Region'] == c]
4     ax = fig.add_subplot(2, 3, num)
5     x = file_req['Year'].unique()
6
7     # Summing reported cases and plotting
8     y11 = file_req.groupby('Year').apply(lambda x: np.sum(pd.to_numeric(x
9         y1 = y11['Counts']
10        sns.regplot(x=x, y=y1) # Fixed here
11
12     # Summing reported deaths and plotting
13     y21 = file_req.groupby('Year').apply(lambda x: np.sum(pd.to_numeric(x
14         y2 = y21['Counts']
15        sns.regplot(x=x, y=y2) # Fixed here
16
17     ax.legend(['Reported Cases', 'Deaths'])
18     ax.set_title(c)
19     ax.set_xlabel('Years')
20     ax.set_ylabel('Cases Count')
21
22 plt.tight_layout()
23 plt.show()
24

```



To identify the number of Cases Registered in the World (All time data)

In [57]:

```

1 file_ctry=file.groupby(['Country']).apply(lambda x:np.sum(pd.to_numeric(x
2 print('The Number of Cholera Cases for the counties:\n',file_ctry)
3 plt.figure(figsize=(30,10))
4 plt.scatter(file_ctry['Country'],file_ctry['Count of Cholera Cases'])
5 plt.xlabel('Countries',size=20)
6 plt.ylabel('Cholera Cases Count',size=20)
7 plt.title('Cholera Cases Distribution- All Time',size=40)
8 plt.xticks(rotation=90)
9 plt.show()

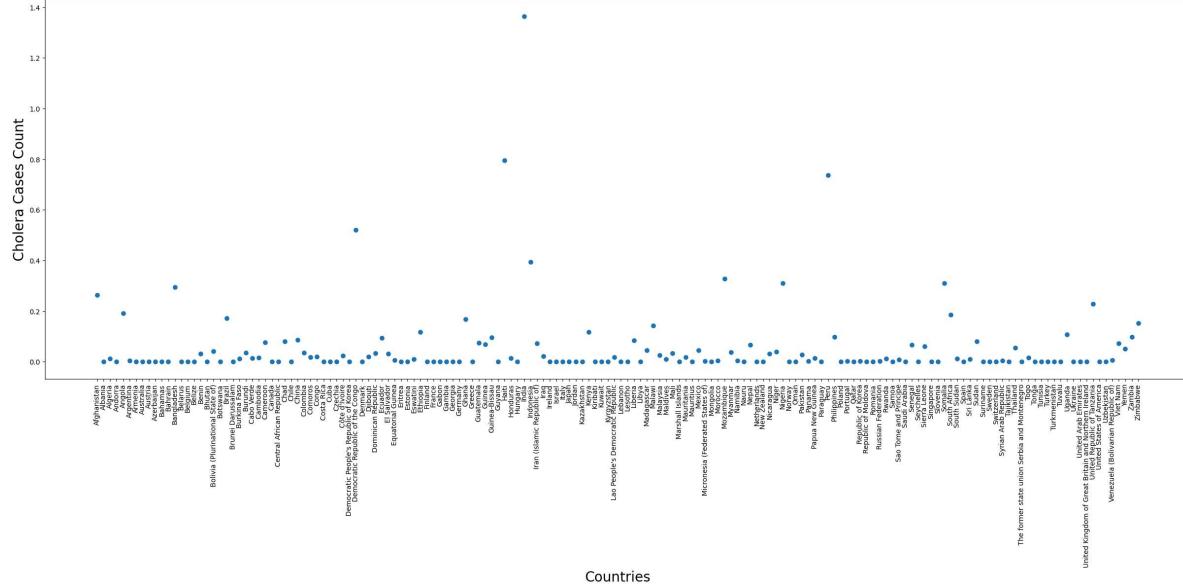
```

The Number of Cholera Cases for the counties:

	Country	Count of Cholera Cases
0	Afghanistan	263843.0
1	Albania	626.0
2	Algeria	12729.0
3	Andorra	0.0
4	Angola	191036.0
..
157	Venezuela (Bolivarian Republic of)	6969.0
158	Viet Nam	73128.0
159	Yemen	52462.0
160	Zambia	97606.0
161	Zimbabwe	153428.0

[162 rows x 2 columns]

Cholera Cases Distribution- All Time



Importing a shape file, that will be required for the plotting of the Global Heatmap

In [33]:

```

1 import geopandas as gpd
2
3 # Correct file path (using raw string Literal)
4 fp = r"C:\Users\Yusuf Nasir\OneDrive\Desktop\Zakiya\TM_WORLD_BORDERS-0.3.shp"
5
6 # Read the shapefile
7 map_df = gpd.read_file(fp)
8
9 # Display the GeoDataFrame
10 print(map_df.head())
11

```

	FIPS	ISO2	ISO3	UN	NAME	AREA	POP2005	REGION	\
0	AC	AG	ATG	28	Antigua and Barbuda	44	83039	19	
1	AG	DZ	DZA	12	Algeria	238174	32854159	2	
2	AJ	AZ	AZE	31	Azerbaijan	8260	8352021	142	
3	AL	AL	ALB	8	Albania	2740	3153731	150	
4	AM	AM	ARM	51	Armenia	2820	3017661	142	

	SUBREGION	LON	LAT	\
0	29	-61.783	17.078	
1	15	2.632	28.163	
2	145	47.395	40.430	
3	39	20.068	41.143	
4	145	44.563	40.534	

geometry

0	MULTIPOLYGON (((-61.68667 17.02444, -61.73806 ...
1	POLYGON ((2.96361 36.80222, 2.98139 36.80694, ...
2	MULTIPOLYGON (((45.08332 39.76804, 45.26639 39...)
3	POLYGON ((19.43621 41.02106, 19.45055 41.06, 1...
4	MULTIPOLYGON (((45.57305 40.63249, 45.52888 40...))

The Countries in the .shp file needs to be the same as the Country name in the data file. Identifying those countries and replacing their names.

In [35]:

```

1 print(map_df.columns)
2

```

```

Index(['FIPS', 'ISO2', 'ISO3', 'UN', 'NAME', 'AREA', 'POP2005', 'REGION',
       'SUBREGION', 'LON', 'LAT', 'geometry'],
      dtype='object')

```

In [36]:

```

1 map_df.NAME.replace({'Burma': 'Myanmar'}, regex=True, inplace=True)
2 map_df.NAME.replace({'Korea, Republic of': 'Republic of Korea'}, regex=True, inplace=True)
3 map_df.NAME.replace({'Russia': 'Russian Federation'}, regex=True, inplace=True)
4 map_df.NAME.replace({'United Kingdom': 'United Kingdom of Great Britain and Northern Ireland'}, regex=True, inplace=True)
5 map_df.NAME.replace({'United States': 'United States of America'}, regex=True, inplace=True)
6 map_df.NAME.replace({'Venezuela': 'Venezuela (Bolivarian Republic of)'}, regex=True, inplace=True)

```

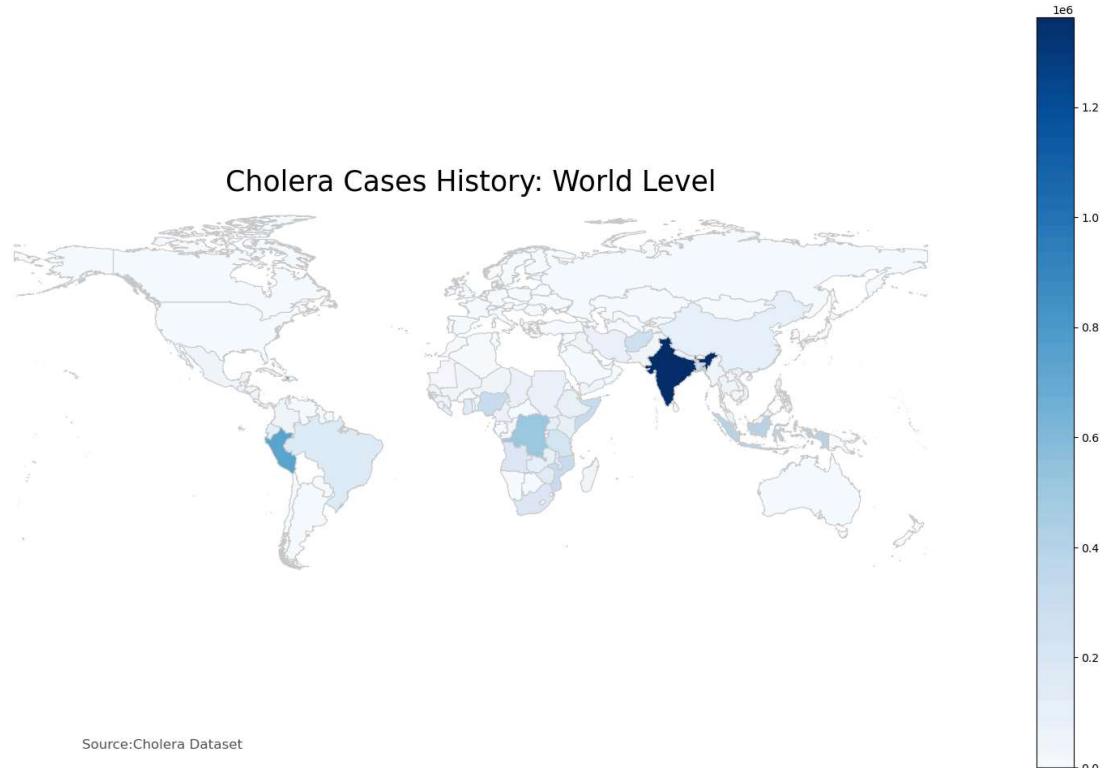
Creating a World Heat Map to display the total number of cases

In [58]:

```

1 merged = map_df.set_index('NAME').join(file_ctry.set_index('Country'))
2 variable = 'Count of Cholera Cases'
3 vmin, vmax = np.min(merged['Count of Cholera Cases']), np.max(merged['Count of Cholera Cases'])
4
5 fig, ax = plt.subplots(1, figsize=(20, 12))
6 merged.plot(column=variable, cmap='Blues', linewidth=0.8, ax=ax, edgecolor='black')
7 ax.axis('off')
8 ax.set_title('Cholera Cases History: World Level', fontdict={'fontsize': 16})
9 ax.annotate('Source:Cholera Dataset', xy=(0.1, .08), xycoords='figure fraction')
10
11 sm = plt.cm.ScalarMappable(cmap='Blues', norm=plt.Normalize(vmin=vmin, vmax=vmax))
12 sm._A = []
13 cbar = fig.colorbar(sm)
14

```



To identify the span of last 10 years from the dataset

```
In [39]: 1 file_yrs=file.Year.unique()
2 file_yrs.sort()
3 final_10_yrs=file_yrs[-10:]
4 final_10_yrs
```

```
Out[39]: array([2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016],
dtype=int64)
```

Countries with the Least number of Cases in the past 10 years

```
In [59]: 1 ten_yr_case=file[file['Year'].isin(final_10_yrs)]
2 file_yr_ctry=ten_yr_case.groupby(['Country']).apply(lambda x:np.sum(pd.to
3 best_cases_country=file_yr_ctry.sort_values(by='Count of Cholera Cases',a:
4 print('Countries having the least number of Cholera Cases in the last 10 y
    <   >
```

Countries having the least number of Cholera Cases in the last 10 years:

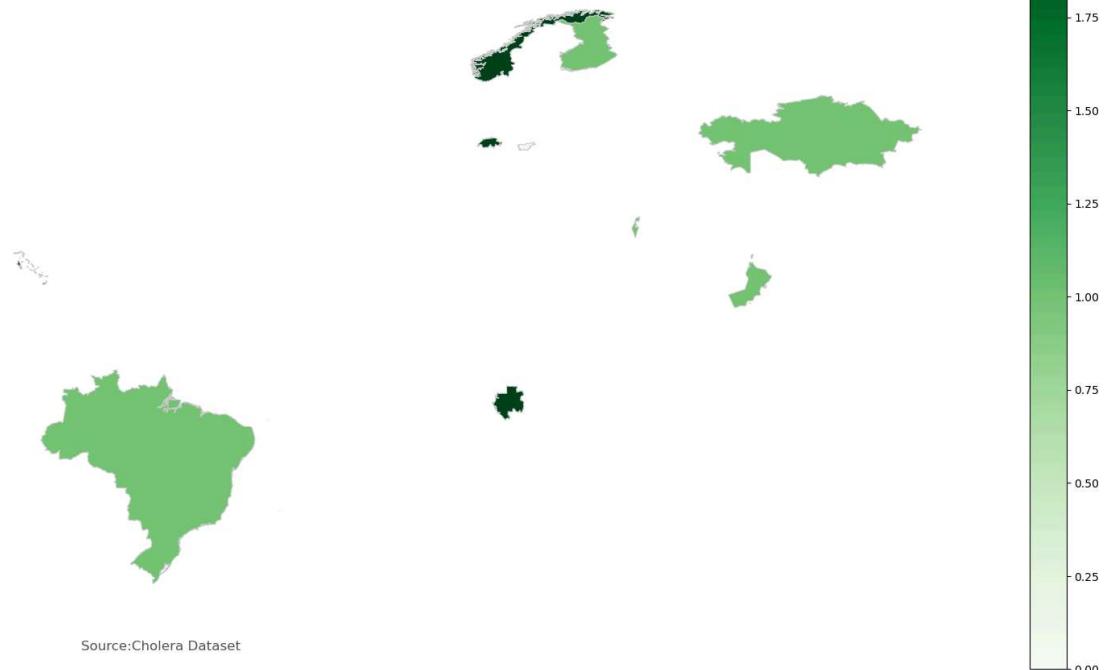
	Country	Count of Cholera Cases
76	Slovenia	0.0
45	Kazakhstan	1.0
42	Israel	1.0
7	Brazil	1.0
63	Oman	1.0
29	Finland	1.0
31	Gabon	2.0
3	Bahamas	2.0
62	Norway	2.0
83	Switzerland	2.0

Plotting the Countries with the Least Number of Cases in the Last 10 Years

In [60]:

```
1 merged = map_df.set_index('NAME').join(best_cases_country.set_index('Country'))
2 variable = 'Count of Cholera Cases'
3 vmin, vmax = np.min(merged['Count of Cholera Cases']), np.max(merged['Count of Cholera Cases'])
4
5 fig, ax = plt.subplots(1, figsize=(20, 12))
6 merged.plot(column=variable, cmap='Greens', linewidth=0.8, ax=ax, edgecolor='black')
7 ax.axis('off')
8 ax.set_title('Countries with the Least Count of Cholera (Last 10 Years)', fontweight='bold')
9 ax.annotate('Source:Cholera Dataset', xy=(0.1, .08), xycoords='figure fraction')
10
11 sm = plt.cm.ScalarMappable(cmap='Greens', norm=plt.Normalize(vmin=vmin, vmax=vmax))
12 sm._A = []
13 cbar = fig.colorbar(sm)
14
```

Countries with the Least Count of Cholera (Last 10 Years)



List of Countries with the Maximum number of cases in the past 10 Years

```
In [42]: 1 worst_cases_country=file_yr_ctry.sort_values(by='Count of Cholera Cases', ascending=False)
2 print('Countries having the most number of Cholera Cases in the last 10 years:')
```

Countries having the most number of Cholera Cases in the last 10 years:

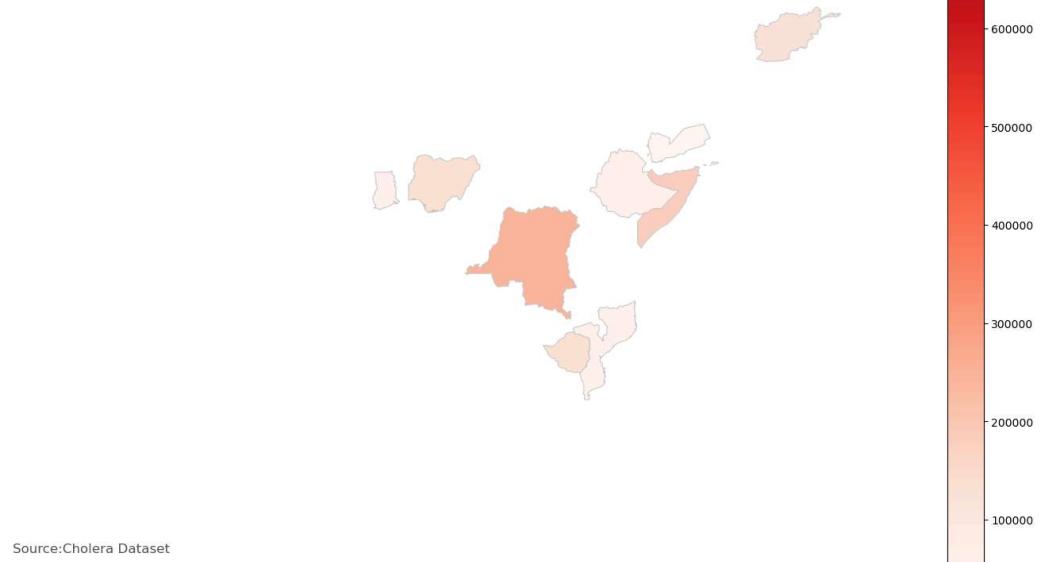
	Country	Count of Cholera Cases
37	Haiti	795794.0
22	Democratic Republic of the Congo	246985.0
77	Somalia	179693.0
61	Nigeria	137846.0
95	Zimbabwe	130537.0
0	Afghanistan	119339.0
28	Ethiopia	61174.0
34	Ghana	53171.0
55	Mozambique	52715.0
93	Yemen	47895.0

Plotting the Countries with the Maximum number of Cases in the last 10 Years

In [61]:

```
1 merged = map_df.set_index('NAME').join(worst_cases_country.set_index('Country'))
2 variable = 'Count of Cholera Cases'
3 vmin, vmax = np.min(merged['Count of Cholera Cases']), np.max(merged['Count of Cholera Cases'])
4
5 fig, ax = plt.subplots(1, figsize=(20, 12))
6 merged.plot(column=variable, cmap='Reds', linewidth=0.8, ax=ax, edgecolor='black')
7 ax.axis('off')
8 ax.set_title('Countries with the Highest Count of Cholera (Last 10 Years)')
9 ax.annotate('Source:Cholera Dataset', xy=(0.1, .08), xycoords='figure fraction')
10
11 sm = plt.cm.ScalarMappable(cmap='Reds', norm=plt.Normalize(vmin=vmin, vmax=vmax))
12 sm._A = []
13 cbar = fig.colorbar(sm)
14
```

Countries with the Highest Count of Cholera (Last 10 Years)



Countries Level Analysis- Deaths from Cholera (All Time)

In [62]:

```

1 file_ctry_deaths = file.groupby(['Country']).apply(lambda x: (pd.to_numeric(x['Count of Deaths from Cholera']), len(x)))
2 print('The Number of Cholera Cases for the counties:\n', file_ctry_deaths)
3
4 plt.figure(figsize=(30, 10))
5 plt.bar(file_ctry_deaths['Country'], file_ctry_deaths['Count of Deaths from Cholera'])
6 plt.xlabel('Countries', size=30)
7 plt.ylabel('Cholera Death Cases Count', size=30)
8 plt.title('Cholera Death Cases Distribution- All Time', size=40)
9 plt.xticks(rotation=90)
10 plt.show()
11
12 merged = map_df.set_index('NAME').join(file_ctry_deaths.set_index('Country'))
13 variable = 'Count of Deaths from Cholera'
14 vmin, vmax = np.min(merged['Count of Deaths from Cholera']), np.max(merged['Count of Deaths from Cholera'])
15
16 fig, ax = plt.subplots(1, figsize=(20, 12))
17 merged.plot(column=variable, cmap='PuBu', linewidth=0.8, ax=ax, edgecolor='black')
18 ax.axis('off')
19 ax.set_title('Countries with the Highest Count of Death from Cholera', fontweight='bold')
20 ax.annotate('Source:Cholera Dataset', xy=(0.1, .08), xycoords='figure fraction')
21
22 sm = plt.cm.ScalarMappable(cmap='PuBu', norm=plt.Normalize(vmin=vmin, vmax=vmax))
23 sm._A = []
24 cbar = fig.colorbar(sm)
25

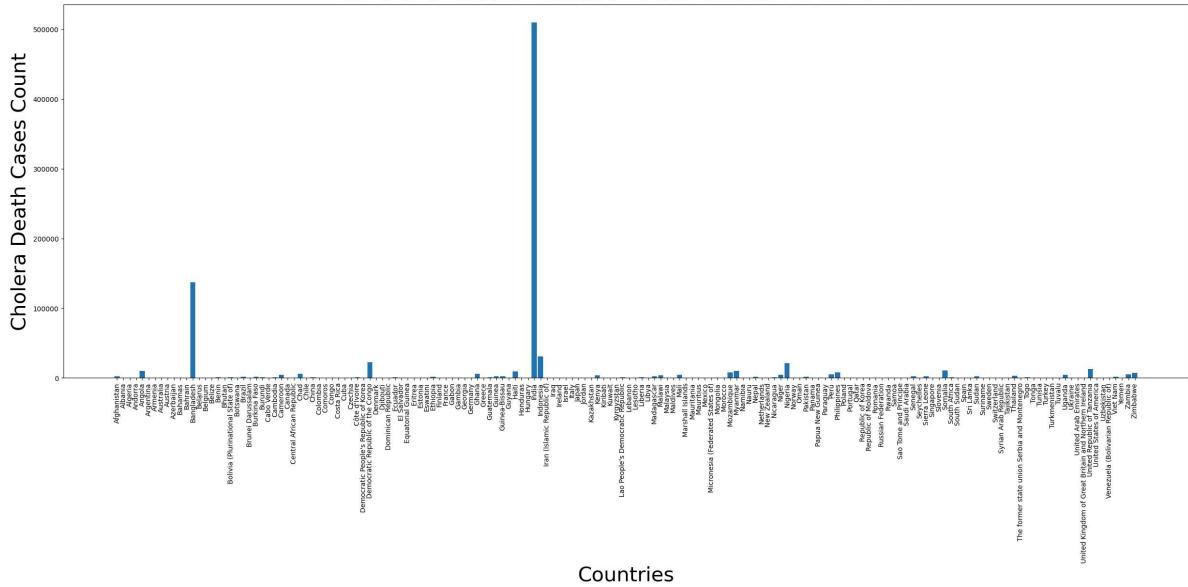
```

The Number of Cholera Cases for the counties:

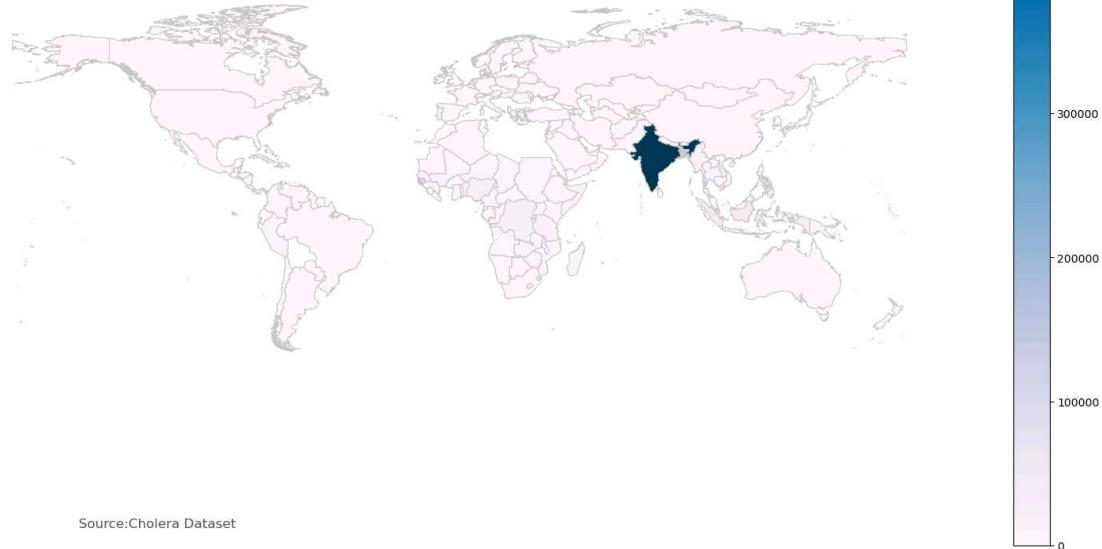
	Country	Count of Deaths from Cholera
0	Afghanistan	2641.0
1	Albania	25.0
2	Algeria	650.0
3	Andorra	0.0
4	Angola	9920.0
..
157	Venezuela (Bolivarian Republic of)	163.0
158	Viet Nam	1592.0
159	Yemen	572.0
160	Zambia	4877.0
161	Zimbabwe	7137.0

[162 rows x 2 columns]

Cholera Death Cases Distribution- All Time



Countries with the Highest Count of Death from Cholera



Creating a "WHO Region vs Year Heatmap" for the Total Counts of Cases- Analysis of the Last 10 years

In [63]:

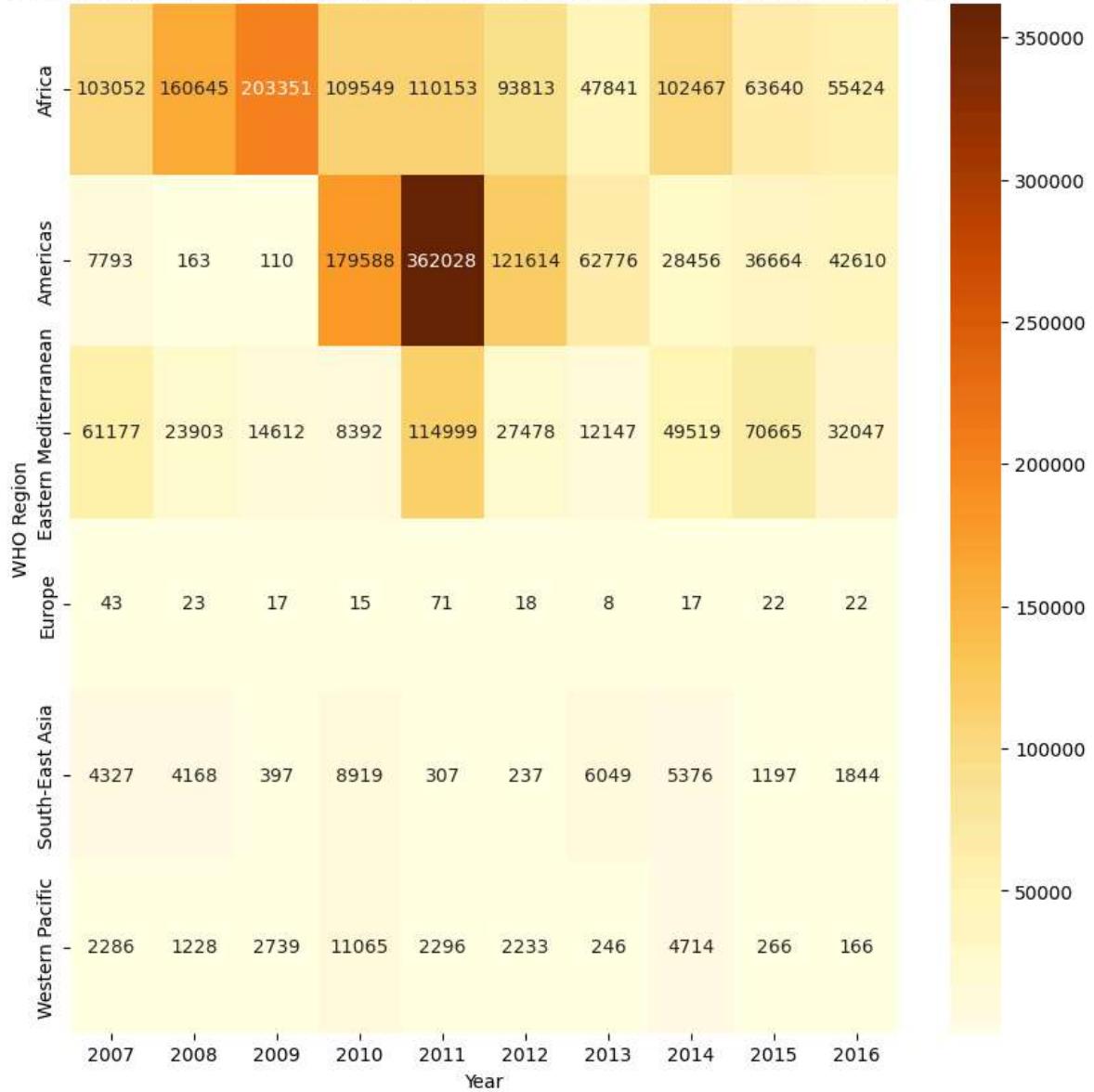
```

1 file_yr_region=ten_yr_case.groupby(['WHO Region','Year']).apply(lambda x: x['Count of Cholera Cases'].sum())
2 heatmap1_data = pd.pivot_table(file_yr_region, values='Count of Cholera Cases', index=['WHO Region'],columns='Year')
3
4 plt.figure(figsize=(10,10))
5 sns.heatmap(heatmap1_data,annot=True,fmt='.0f', cmap='YlOrBr').set_title('WHO Region vs Year Heatmap of Cholera Cases- in the last 10 years')

```

Out[63]: Text(0.5, 1.0, 'WHO Region vs Year Heatmap of Cholera Cases- in the last 10 years')

WHO Region vs Year Heatmap of Cholera Cases- in the last 10 years



Plotting the average fatality rate in the Nations over the years

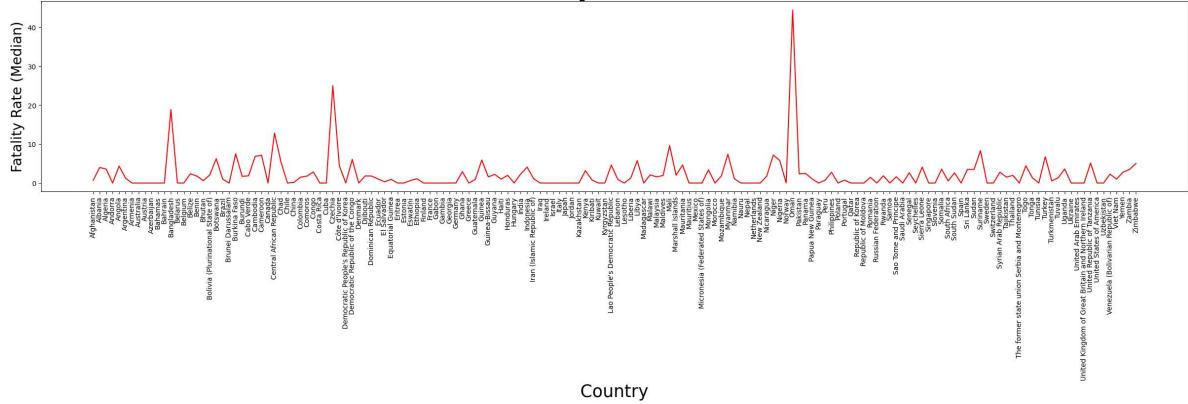
In [64]:

```

1 fat_ctry = file.groupby('Country').apply(lambda x: pd.to_numeric(x['Choler
2 fat_ctry.replace(np.NaN, 0, inplace=True)
3
4 plt.figure(figsize=(30, 5))
5 plt.plot(fat_ctry['Country'], fat_ctry['Median Fatality'], color='r')
6 plt.xlabel('Country', size=25)
7 plt.ylabel('Fatality Rate (Median)', size=20)
8 plt.title('Median Fatality Rate- Nation wise', size=40)
9 plt.xticks(rotation=90)
10
11 merged = map_df.set_index('NAME').join(fat_ctry.set_index('Country'))
12 variable = 'Median Fatality'
13 vmin, vmax = np.min(merged['Median Fatality']), np.max(merged['Median Fat
14
15 fig, ax = plt.subplots(1, figsize=(20, 12))
16 merged.plot(column=variable, cmap='OrRd', linewidth=0.8, ax=ax, edgecolor:
17 ax.axis('off')
18 ax.set_title('Countries with the Highest Fatality Rates from Cholera', fo
19 ax.annotate('Source:Cholera Dataset', xy=(0.1, .08), xycoords='figure fra
20
21 sm = plt.cm.ScalarMappable(cmap='PuBuGn', norm=plt.Normalize(vmin=vmin, v
22 sm._A = []
23 cbar = fig.colorbar(sm)
24

```

Median Fatality Rate- Nation wise





We will now be summarizing the overall data for 2 nations- One which has the highest counts in all time data, and the other having the highest counts in the past 10 years

We are defining a function from where we shall be receiving the Counts Analysis for the particular country

In [69]:

```

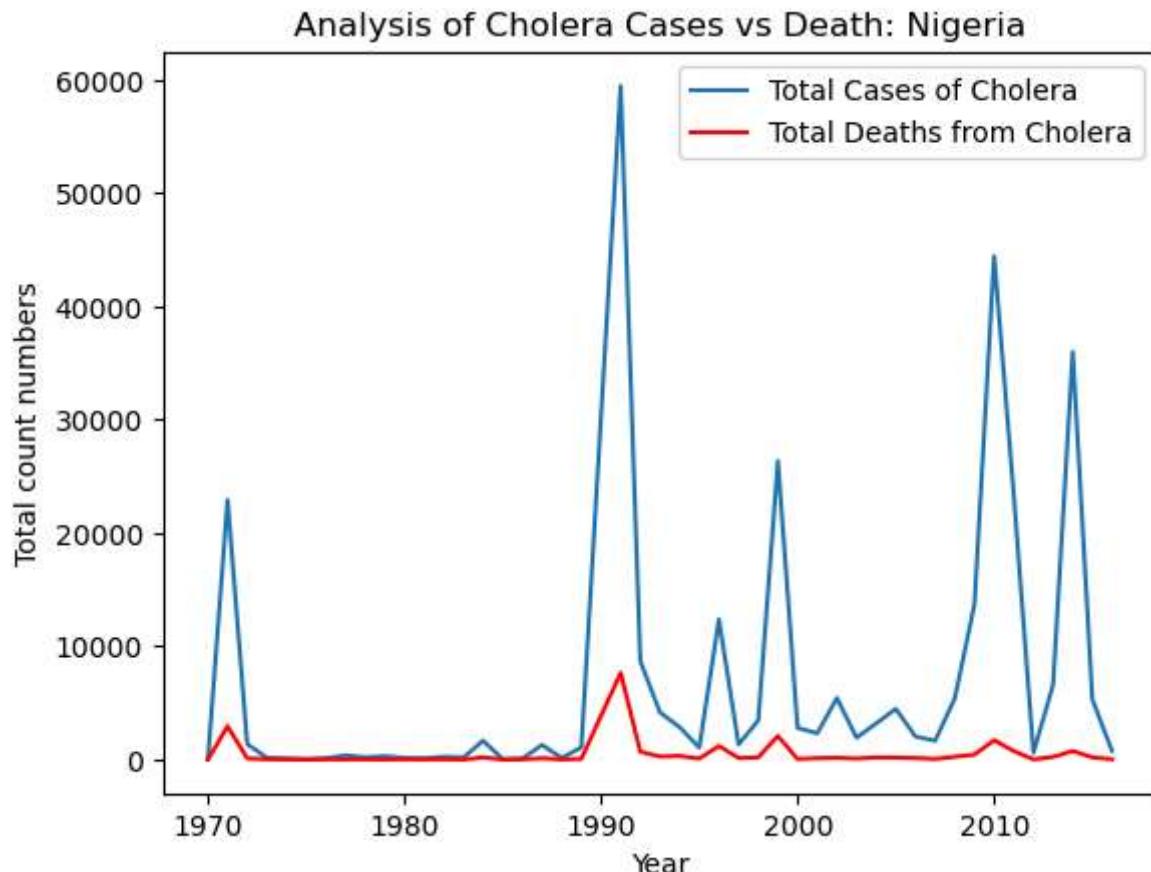
1 def country_summary(country):
2     country_file = file[file['Country'] == country]
3     country_case_counts = country_file.groupby('Year').apply(lambda x: pd
4     country_death_counts = country_file.groupby('Year').apply(lambda x: pd
5
6     plt.plot(country_case_counts['Year'], country_case_counts['Total Cases'])
7     plt.plot(country_death_counts['Year'], country_death_counts['Total Deaths'])
8     plt.xlabel('Year')
9     plt.ylabel('Total count numbers')
10    plt.title('Analysis of Cholera Cases vs Death: {}'.format(country))
11    plt.legend(['Total Cases of Cholera', 'Total Deaths from Cholera'])
12
13    print('The Year {:.0f} has seen the maximum number of Cholera Cases: {}'
14          .format(np.argmax(country_case_counts['Total Cases']),
15                  np.max(country_case_counts['Total Cases of Cholera'])))
16
17    print('The Year {:.0f} has seen the maximum number of deaths due to Cholera: {}'
18          .format(np.argmax(country_death_counts['Total Deaths']),
19                  round(np.max(country_death_counts['Total Deaths from Cholera']))))
20

```

In [70]:

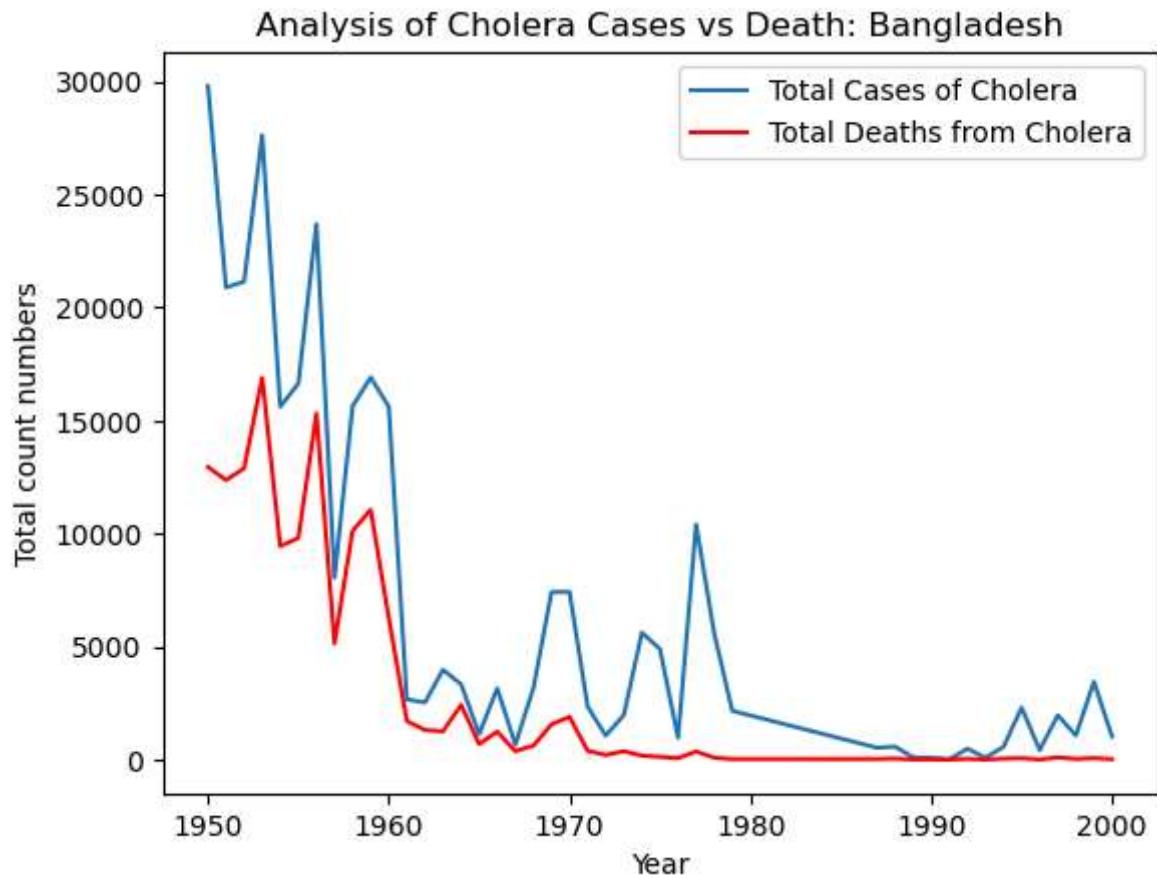
1 Nigerian_Data=country_summary('Nigeria')

The Year 1991 has seen the maximum number of Cholera Cases: 59478
 The Year 1991 has seen the maximum number of deaths due to Cholera: 7654



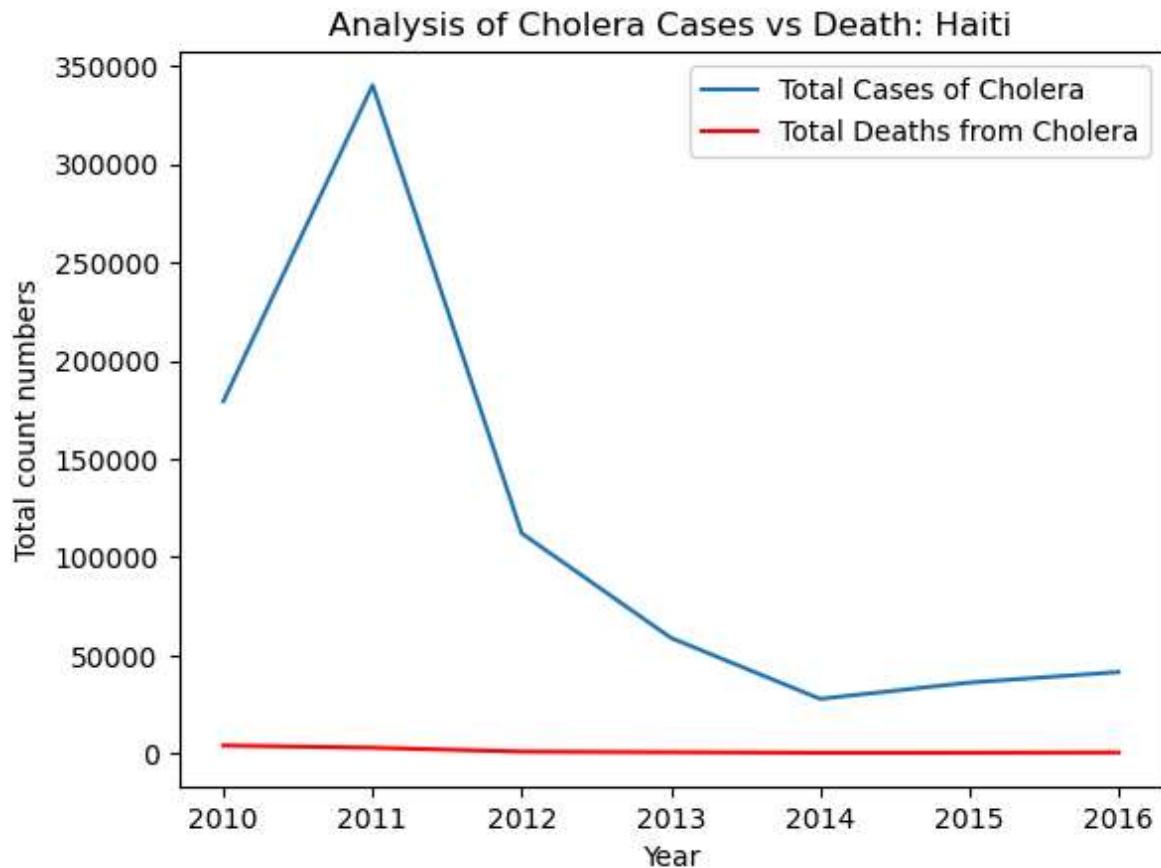
```
In [71]: 1 Bangladesh_Data=country_summary('Bangladesh')
```

The Year 1950 has seen the maximum number of Cholera Cases: 29809
The Year 1953 has seen the maximum number of deaths due to Cholera: 16904



In [72]: 1 Haiti_Data=country_summary('Haiti')

The Year 2011 has seen the maximum number of Cholera Cases: 340311
The Year 2010 has seen the maximum number of deaths due to Cholera: 3990



Machine Learning Model

In [106]: 1 `from sklearn.model_selection import train_test_split`
2 `from sklearn.ensemble import RandomForestClassifier`
3 `from sklearn.metrics import classification_report, accuracy_score`
4 `from sklearn.preprocessing import LabelEncoder`
5 `from sklearn.model_selection import cross_val_score`

In [74]: 1 `# Loading dataset`
2 `file = pd.read_csv(r'C:\Users\Yusuf Nasir\OneDrive\Desktop\Zakiya\data.csv')`

In [75]:

```

1 # Data Preprocessing
2 # Replace 'Unknown' values with 0
3 file.replace('Unknown', 0, regex=True, inplace=True)
4
5 # Clean specific columns (example for "Reported Cases" and "Reported Deaths")
6 file['Number of reported cases of cholera'] = pd.to_numeric(file['Number of reported cases of cholera'])
7 file['Number of reported deaths from cholera'] = pd.to_numeric(file['Number of reported deaths from cholera'])
8 file['Cholera case fatality rate'] = pd.to_numeric(file['Cholera case fatality rate'])
9
10

```



1

In [114]:

```

1 # Encode categorical data like 'Country' and 'WHO Region' using LabelEncoder()
2 le_country = LabelEncoder()
3 file['Country_encoded'] = le_country.fit_transform(file['Country'])
4
5 le_region = LabelEncoder()
6 file['Region_encoded'] = le_region.fit_transform(file['WHO Region'])
7
8 # Feature Engineering: Create some new features based on the existing data
9 file['Year'] = pd.to_datetime(file['Year'], format='%Y')
10 file['Month'] = file['Year'].dt.month
11 file['Case Fatality Rate'] = file['Number of reported deaths from cholera'] / file['Number of reported cases of cholera']
12
13

```



Create a binary target variable (1 = Potential outbreak, 0 = No outbreak)

In [77]:

```

1 file['Potential_Outbreak'] = (file['Number of reported cases of cholera'] > 0) * 1
2
3 # Select Features and Target
4 features = ['Year', 'Month', 'Country_encoded', 'Region_encoded', 'Case Fatality Rate']
5 target = 'Potential_Outbreak'
6
7
8

```



In [112]:

```

1 import pandas as pd
2 from sklearn.ensemble import RandomForestClassifier
3 from sklearn.model_selection import train_test_split
4 from sklearn.metrics import accuracy_score
5 from sklearn.preprocessing import LabelEncoder
6

```

Extract useful date-related features from 'Month'

```
In [109]: 1 file['Month'] = pd.to_datetime(file['Month'], errors='coerce')
2 file['Year'] = file['Month'].dt.year
3 file['Month'] = file['Month'].dt.month
4
```

Encode categorical variables

```
In [ ]: 1
2 label_encoder = LabelEncoder()
3 file['Country_encoded'] = label_encoder.fit_transform(file['Country'])
4 file['Region_encoded'] = label_encoder.fit_transform(file['WHO Region'])
5
```

Select features (X) and target (y)

```
In [ ]: 1 X = file[['Year', 'Country_encoded', 'Region_encoded', 'Month', 'Case Fatality_Rate']]
2 y = file['Potential_Outbreak'] # Assuming this is the target variable
3
```

Train-Test Split

```
In [ ]: 1
2 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
3
```

Initialize and train the Random Forest model

```
In [111]: 1 rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
2 rf_model.fit(X_train, y_train)
3 y_pred = rf_model.predict(X_test)
4
```

Model Accuracy

In [113]:

```

1
2 accuracy = accuracy_score(y_test, y_pred)
3 print(f'Accuracy: {accuracy}')

```

Accuracy: 0.7975951903807615

In [108]:

```
1 rf_model.fit(X_train, y_train)
```

Out[108]:

```

v      RandomForestClassifier
RandomForestClassifier(random_state=42)

```

In []:

```
1
```

In [91]:

```

1 predictions = rf_model.predict(X_test)
2 scores = accuracy_score(y_test, predictions)
3 scores

```

Out[91]: 0.7975951903807615

In [92]:

```
1 predictions
```

```

Out[92]: array([1, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1,
    0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0,
    1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0,
    0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0,
    1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0,
    1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0,
    1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0,
    0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1,
    0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0,
    0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0,
    0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1,
    0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0,
    1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0,
    1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0,
    1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1,
    0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
    0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
    0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0,
    0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1,
    0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
    0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
    0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
    0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
    0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
    0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
    0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0])

```

In []:

```
1
```

