08/03/2021

CURNEU MEDTECH INNOVATIONS PRIVATE LIMITED – SD03Q014

REPORT – 01

Problem Statement:

To understand and perform Exploratory Data Analysis to retrieve deep information and fit the suitable Machine Learning model for the given dataset, and to fetch best accuracy for the fitted model. Here, the "reported\_numbers.csv" for easy use.

Abstract:

The maxim is to find if the total cases and deaths due to malaria increases by each year. Initially Exploratory Data Analysis will be done to explore the data in a deeper way and suitable model will be fitted after crucial analysis of each and every variable. Finally the statement must be concluded whether the cases increase or not.

Dataset Description:

There were initially three datasets and the common attributes among them were "No. of deaths", "No. of cases", "WHO regions", "Country", and "Year". The three datasets were,

* Reported\_numbers.csv 🡪 Reported no. of cases across the world.
* Estimated\_numbers.csv 🡪 Estimated no. of cases across the world.
* Incidenceper1000popat\_risk.csv 🡪 Incidence per 1000 people at risk area.

EDA:

The graphs were plotted using reported cases file.

* The first bar plot was plotted between no. of cases and country to determine the highest number of cases country wise between a time period of 2000-2018.
* Another bar plot was plotted between no. of deaths and country to determine the highest number of malarial deaths between a time period of 2000-2018.
* Again a bar plots were plotted between no. of cases and WHO regions to determine the highest number of cases and deaths for the same time period 2000-2018.
* Finally, it was detected that "Africa" had higher number malarial cases and so a bar plot was plotted between Africa and other regions to find the trend line.
* Extra plots were plotted to find the spread of malaria country wise every year and situation of malaria across the world.

Code:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import plotly.express as px

import plotly.io as pio

import plotly.graph\_objects as go

from plotly.subplots import make\_subplots

import warnings

warnings.filterwarnings("ignore")

%matplotlib inline

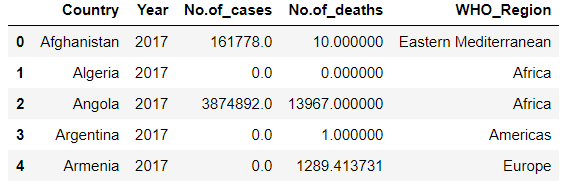
dt=pd.read\_csv("reported\_numbers.csv")

dt.rename(columns={'No. of cases':'No.of\_cases','No. of deaths':'No.of\_deaths','WHO Region':'WHO\_Region'},inplace=True)

dt['No.of\_deaths'].fillna(value=dt['No.of\_deaths'].mean(), inplace=True)

dt['No.of\_cases'].fillna(value=dt['No.of\_cases'].mean(), inplace=True)

dt.head(5)

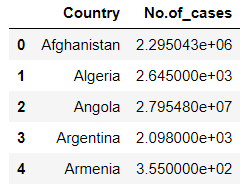


dtgroup=dt.groupby('Country')["No.of\_cases","No.of\_deaths"].sum().reset\_index()

dtgroup.head()

dtcase = dtgroup[["Country","No.of\_cases"]]

dtcase.head()



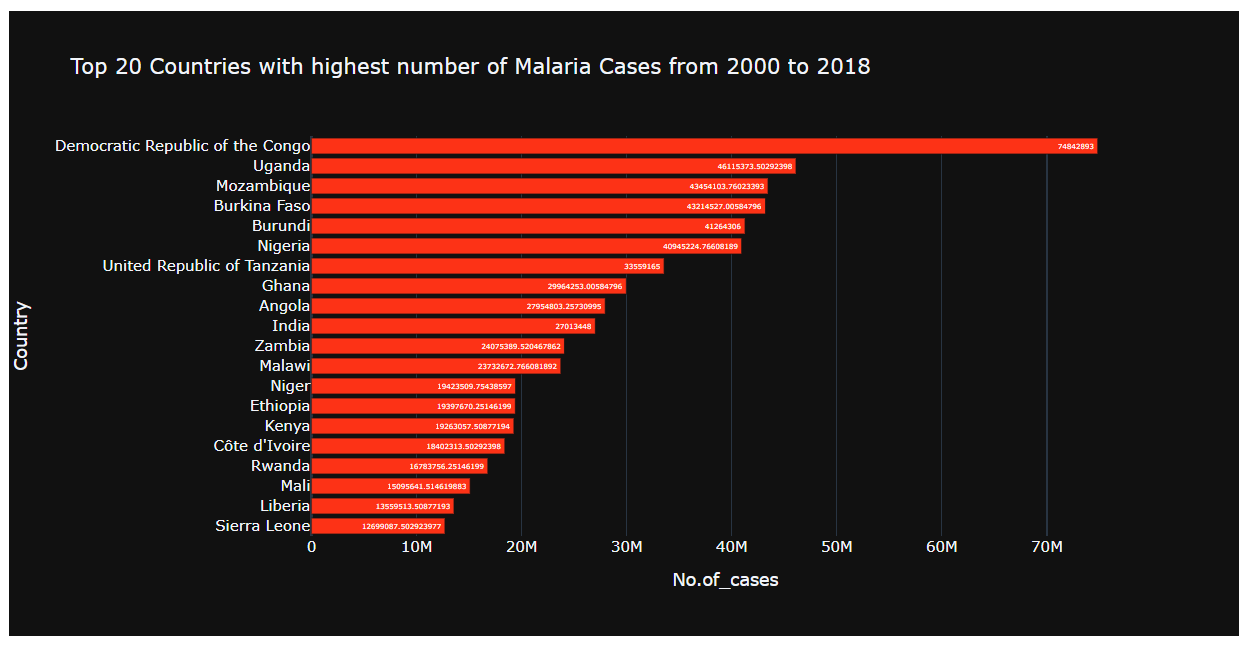
pio.templates.default ='plotly\_dark'

fig = px.bar(dtcase.sort\_values("No.of\_cases",ascending=False)[:20][::-1],x="No.of\_cases",y ="Country",text="No.of\_cases",

title="Top 20 Countries with highest number of Malaria Cases from 2000 to 2018",

color\_discrete\_sequence= px.colors.qualitative.Light24,height=500,orientation="h")

fig.show()



dtdeath = dtgroup[["Country","No.of\_deaths"]]

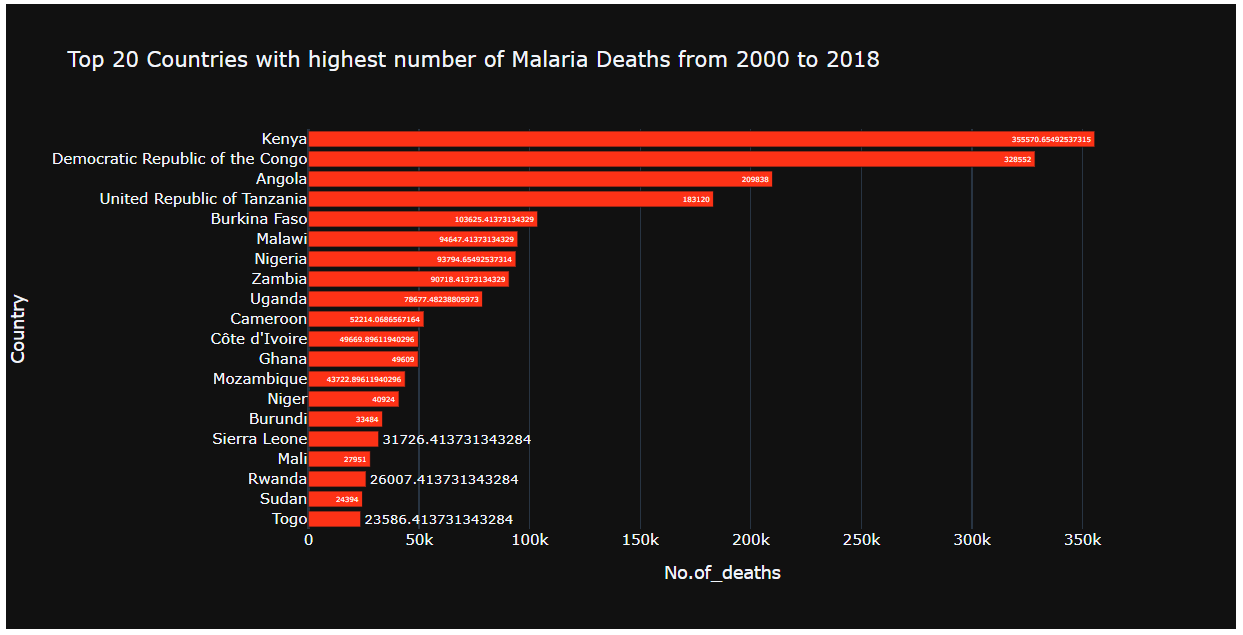
pio.templates.default ='plotly\_dark'

fig = px.bar(dtdeath.sort\_values("No.of\_deaths",ascending=False)[:20][::-1],x="No.of\_deaths",y ="Country",text="No.of\_deaths",

title="Top 20 Countries with highest number of Malaria Deaths from 2000 to 2018",

color\_discrete\_sequence= px.colors.qualitative.Light24,height=500,orientation="h")

fig.show()



who\_group =dt.groupby('WHO\_Region')["No.of\_cases","No.of\_deaths"].sum().reset\_index()

who\_group.head().style.background\_gradient(cmap ='Reds')

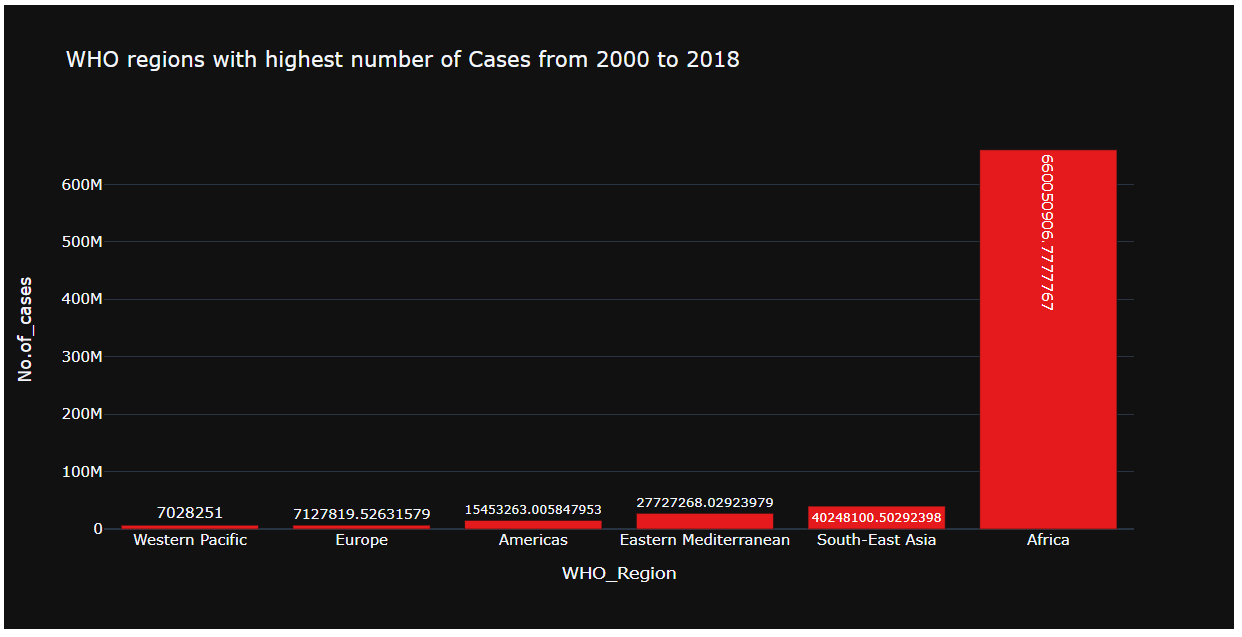
pio.templates.default = "plotly\_dark"

fig = px.bar(who\_group.sort\_values("No.of\_cases",ascending=False)[::-1],y="No.of\_cases",x ="WHO\_Region",text="No.of\_cases",

title="WHO regions with highest number of Cases from 2000 to 2018",

color\_discrete\_sequence= px.colors.qualitative.Set1,height=500,orientation="v")

fig.show()



pio.templates.default = "plotly\_dark"

fig = px.bar(who\_group.sort\_values("No.of\_deaths",ascending=False)[::-1],y="No.of\_deaths",x ="WHO\_Region",text="No.of\_deaths",

title="WHO regions with highest number of Deaths from 2000 to 2018",

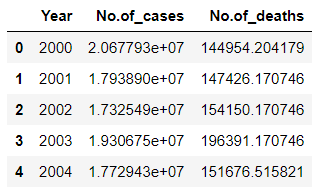
color\_discrete\_sequence= px.colors.qualitative.Set1,height=500,orientation="v")

fig.show()



yrgroup= dt.groupby("Year")[["No.of\_cases","No.of\_deaths"]].sum().reset\_index()

yrgroup.head()



fig\_dims = (20, 5)

fig, axes = plt.subplots(1, 2, figsize=fig\_dims)

sns.barplot(x = 'Year' , y = 'No.of\_cases' , data = dt[dt['WHO\_Region'] == 'Africa'], ax= axes[0]).set\_title("In Africa")

sns.barplot(x = 'Year' , y = 'No.of\_cases' , data = dt[dt['WHO\_Region'] != 'Africa'], ax= axes[1]).set\_title("In other WHO Regions")

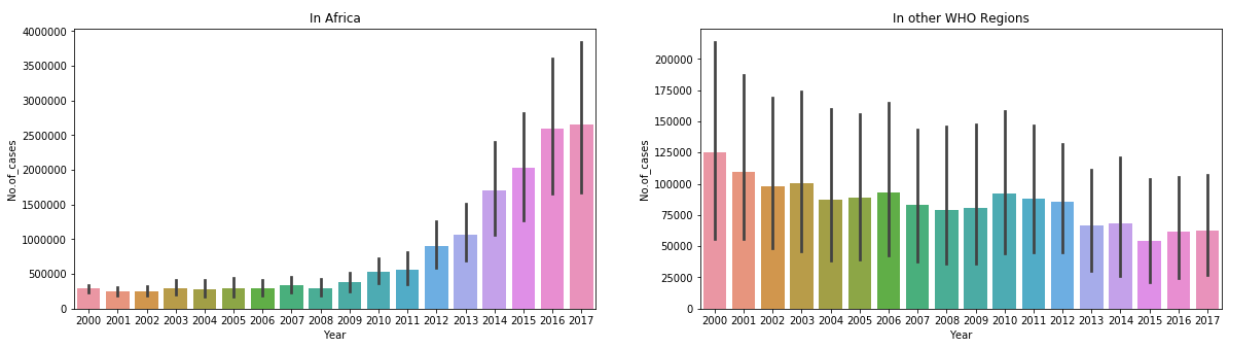


fig = make\_subplots(rows=1,cols=2,column\_titles = ('No. of cases','No. of deaths'))

trace\_1 = go.Scatter(x=yrgroup['Year'],y=yrgroup['No.of\_cases'],name='Cases',opacity=0.9,mode='lines+markers',line\_color='pink')

trace\_2 = go.Scatter(x=yrgroup['Year'],y=yrgroup['No.of\_deaths'],name='Deaths',opacity=0.9,mode='lines+markers',line\_color='violet')

fig.append\_trace(trace\_1,1,1)

fig.append\_trace(trace\_2,1,2)

fig.update\_layout(title\_text="Spread of Malaria according to Year")

fig.show()

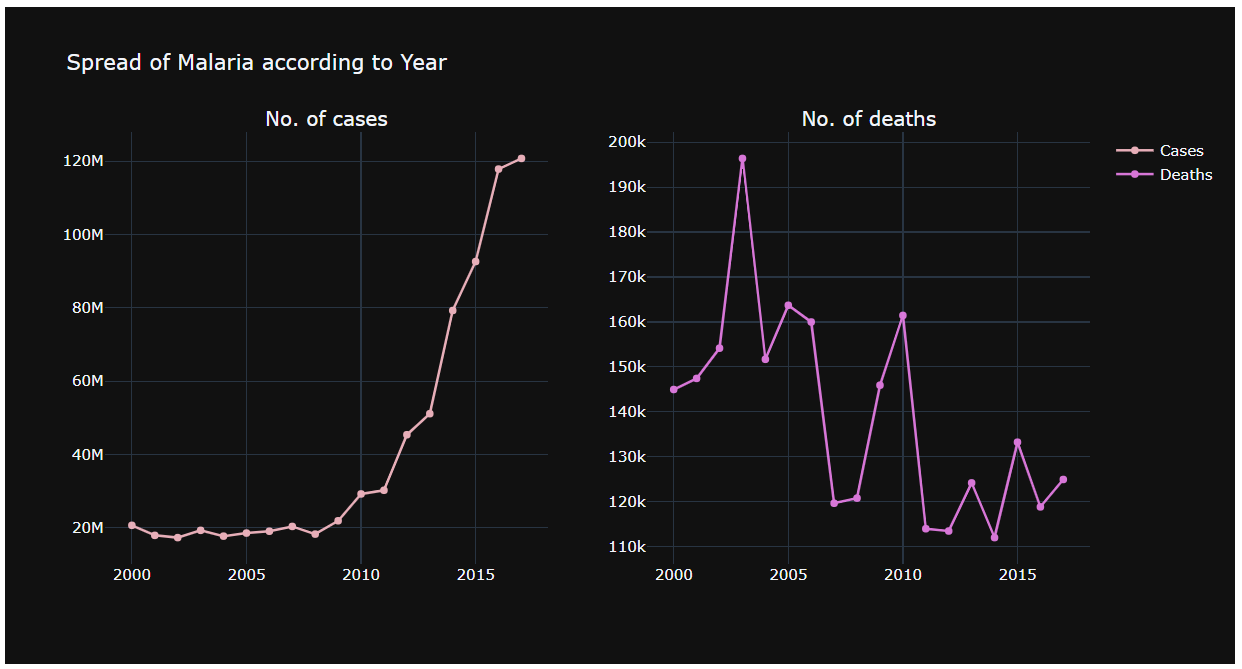
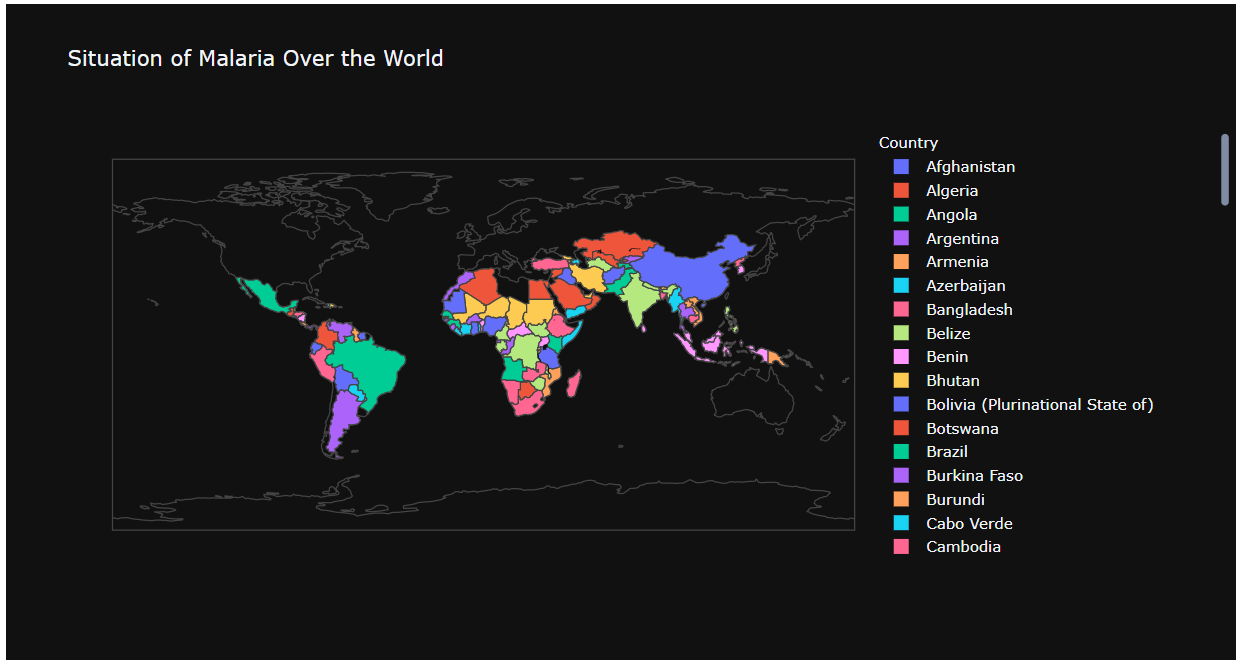


fig = px.choropleth(dtgroup,locationmode="country names",locations ="Country",hover\_data = ["Country","No.of\_cases","No.of\_deaths"],

hover\_name = "Country",color="Country",title="Situation of Malaria Over the World")

fig.show()



Conclusion:

From the graphs, it can be inferred that as the number of cases increases, the number of death decreases. Hence, it depends upon the year. The world map plot depicts the spread of malarial cases across the countries. From the above graphs, it can be inferred that Africa had the most number of infected cases than the other regions. The number of cases in WHO regions has many deviations though. We can be roughly derive that Africa is the most affected country and number of cases may increase there. Even though Democratic republic of Congo is one of the top 20 countries in infected cases, Africa leads it in the latter cases which can be seen in the plots.