WORLD ANALYSIS OF DISEASE DEATHS

Introduction

We did an exploratory analysis of the world's diseases deaths by countries. Exploratory Data Analysis or (EDA) is the understanding the data sets by summarizing their main characteristics often plotting them visually. This step is very important especially when we arrive at modelling the data in order to apply Machine learning. Plotting in EDA consists of Histograms, Box plot, Scatter plot and many more. It often takes much time to explore the data. Through the process of EDA, we can ask to define the problem statement or definition on our dataset which is very important

What we are going to explore today?

I got a nice data-set of people deaths by different diseases all over the world from DataLeads company for an assessm ent. This data-set contains 6686 rows and 37 columns of different deaths causes eg- 'Year', 'Execution', 'Meningitis (d eaths)', 'Lower respiratory infections (deaths)', 'Intestinal infectious diseases (deaths)', 'Protein-energy malnutrition (deaths)', 'Terrorism (deaths) etc

Let's jump into it!

First and foremost we need to import the important libraries like Pandas, numpy, matplotlib and seaborn.

```
# Importing required libraries.
import pandas as pd
import numpy as np
import seaborn as sns #visualisation
import matplotlib.pyplot as plt #visualisation
%matplotlib inline
sns.set(color_codes=True)
```

Loading the data into data frame using pandas library!

Loading the data into the pandas data frame is certainly one of the most important steps in EDA, as we can see that the value from the data set is comma-separated. So all we have to do is to just read the CSV into a data frame and pandas data frame does the job for us.

```
# reading the dataset
df= pd.read_csv(data)
df.head()

# for displaying the maximum columns
pd.set_option('display.max_columns', None)
df.head(10)

# Checking the information about the dataset types:
df.info()


# sorted the value into general float format!

pd.set_option('display.float_format', lambda x: '%.3f' % x)

# create a new column for total death which sums up all the deaths from different diseases

df['Total_Deaths'] = df.iloc[:, 4:-1].sum(axis= 1)
```

Segregation task for our dataset to extract different sub-datasets from the main file:

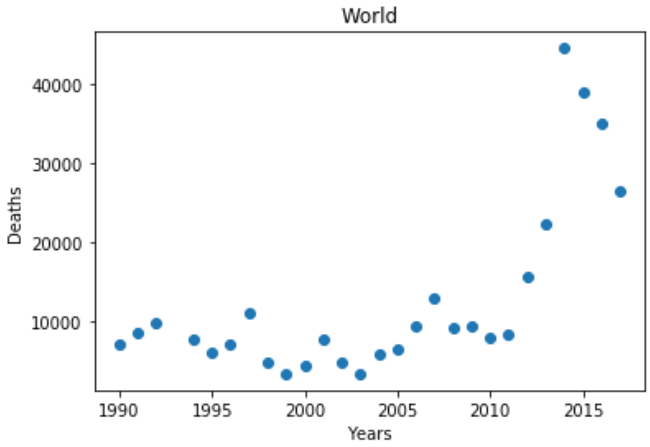
To solve this problem I created a FOR loop to extract files in excel format by countries names('Entity). I first check all the unique values of our column(Entity) and then convert the array into a list so that we can easily use our loop in list.

```
## created a loop to extract different datasets from the same masterfile!!

for i in country_lst:
   grp = df.groupby(df.Entity) ## grouping the columns
   temp = grp.get_group(i)
   print(f'we created dataset from the masterfile: { i}')
   temp.to_excel(i + ".xlsx")
```

Visualization Task to generate 10 infographics using the loop:

To solve this problem I created a FOR loop again to plot first 10 graphs of countries using column suicide rate per year. I group the column with the entity name and then plot each graphs separately using matplotlib library.

```
## creating a loop to plot 10 different coutries graphs based on Year and Suicide deaths
for i in country_lst[:10]:
   grp = df.groupby(df.Entity) ## grouping the columns
   temp = grp.get_group(i)
   print(f'we created graphs from the masterfile based on suicide deaths and year: { i}')
   plt.plot(temp['Year'], temp['Suicide (deaths)'])
   plt.title(i)
   plt.xlabel('Years')
   plt.ylabel('Deaths')
   plt.show()

## creating a loop to plot 10 different countries scatter graphs based on Year and Suicide deaths
for i in country_lst[:10]:
   grp = df.groupby(df.Entity) ## grouping the columns
   temp = grp.get_group(i)
   print(f'we created graphs from the masterfile based on suicide deaths and year: { i}')
   plt.scatter(temp['Year'], temp['Suicide (deaths)'])
   plt.title(i)
   plt.xlabel('Years')
   plt.ylabel('Deaths')
   plt.show()
```



From our given graphs we come into a conclusion that Argentina have thee maximums deaths due to Suicide of around ~5500 from year 2013- 2017
In Argentina Suicide rate is increases from year 1995 -2002 and then they take some measures to reduce it but again there is a surge in suicide deaths during 2013- 2017 years.
In Albania there is a surge increase in suicidal deaths from year 1997 – 2015 significantly which is more than 150 deaths per annum consistently.
In Afghanistan its also increases from year 2000 to 2017 from around 1000 deaths to 2000 deaths per annum

So from here we can also conclude that from year 2000 around there is an increase in deaths numbers of people from suicide in most of the countries.

Then I created a loop to check the variations of deaths due to terrorism in different countries:
I plotted different scatter plots to get insights that which country have the highest deaths due to terrorism!! we get out answer from the plots that country Afghanistan had most deaths from Terrorism of around 6000 people deaths in year 2015-2017

And the country which have least deaths from terrorism is Comoros which is almost 0 since 1994- 2000 and 1 case was happened in year 1992 only!

From the world graph we can also conclude that from 2010 there is an increase in terrorist attack all over the world from 10000 per year to 40000.

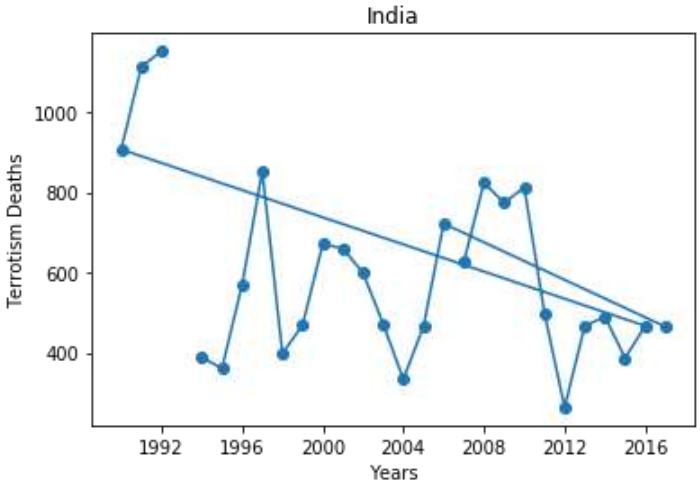
We also figure out the correlations between the alcohol an drug addiction to neural diseases which is almost 100% correlated to each other

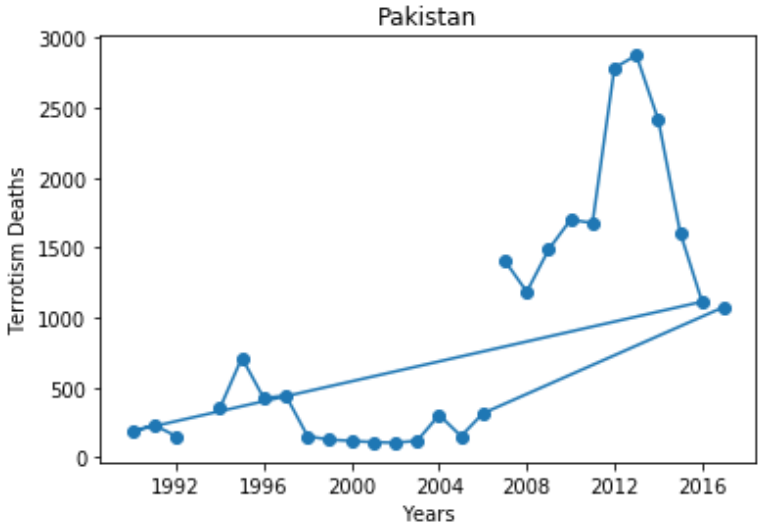
Comparing deaths due to Terrorism and Cardiovascular diseases in India and Pakistan

I created a new data frame for country name India and Pakistan using all death causes columns. Then comparing the terrorism deaths and Cardiovascular Heart diseases deaths.

From our Terrorism and deaths graph we can say that there was terror attack in Pakistan which killed around 2700 people in Pakistan during year 2010 -2013.

And In India there was a surge in terror attack during year 2004 – 2012 which killed around 700 people deaths per annum.

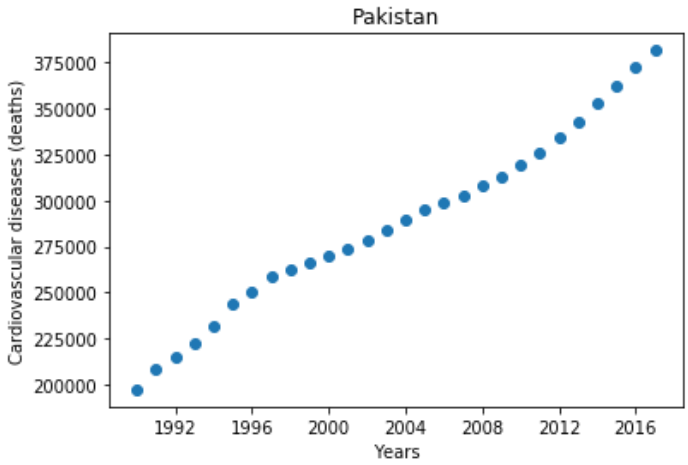


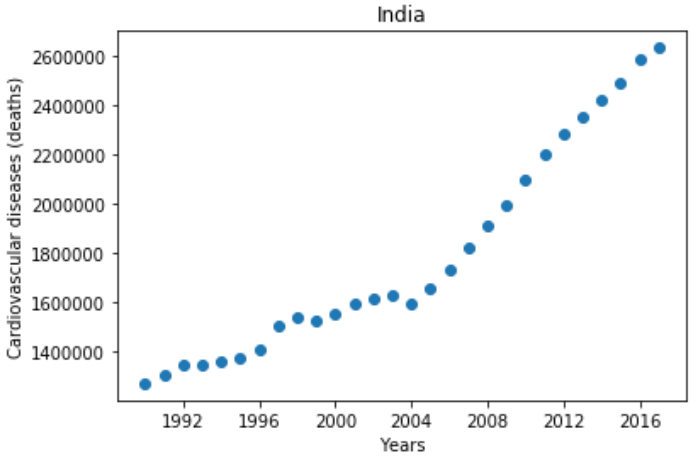


From our Cardiovascular disease deaths and year graph we can say that there is a surge increases in heart diseases in India. From year 2004 there is a spike in deaths due to heart diseases and there are around 2700000 deaths happened in year 2017 which was maximum.

So as in Pakistan there is a surge in deaths due to heart diseases from year 2004. And in Pakistan the maximum deaths happened in year 2017 aswell which was around 375000 which was less than India.

So from here we can conclude that people who died due to cardiovascular diseases are higher in India as compare to Pakistan by around 2300000 people



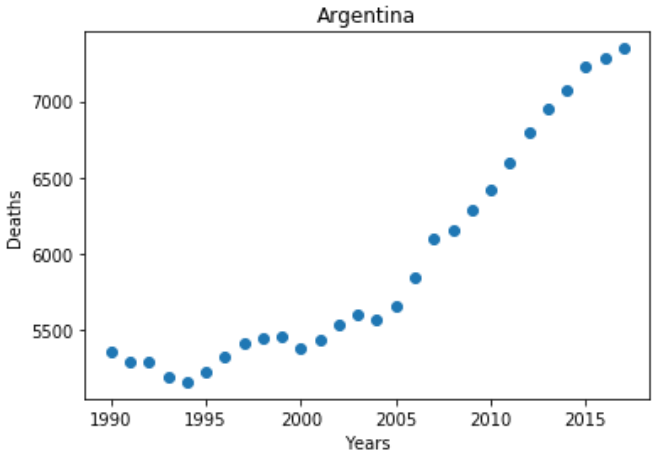


Comparing deaths due to Liver, Cancer, Malaria and TB in India, Pakistan, Canada, US and UK countries

LIVER DEATHS RESULT

Created graphs for liver disease death and year of first 10 countries. We can conclude from the graphs that countries name Argentina and Andean Latin America have the most deaths due to Liver diseases of around ~10000 and ~13000 in year 2017

And we can also conclude that from the year 2000 there is a significant increase in liver diseases in all of the countries!

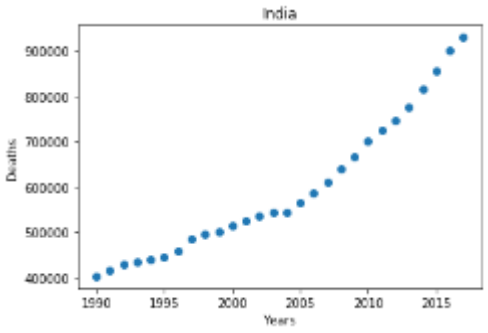


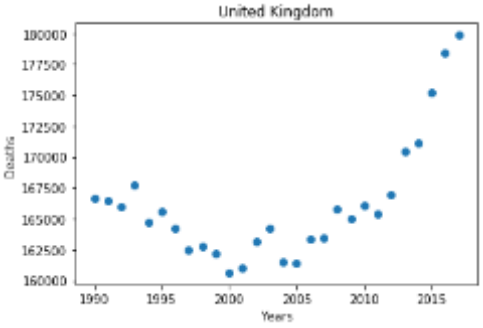
CANCER DEATHS RESULT

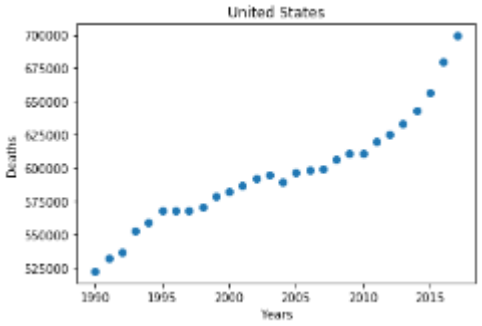
From the cancer death graph of our favourite country list it is clear that in every countries there is an increase in deaths of people from cancer disease.

India has the max deaths rate of around ~900000 in year 2017, and US comes to 2 with around ~700000 deaths in year 2016.

from Our graph of UK country we can say that they had taken some measure to control deaths from Cancer in year between 1997- 2000 but again increase from year 2007.





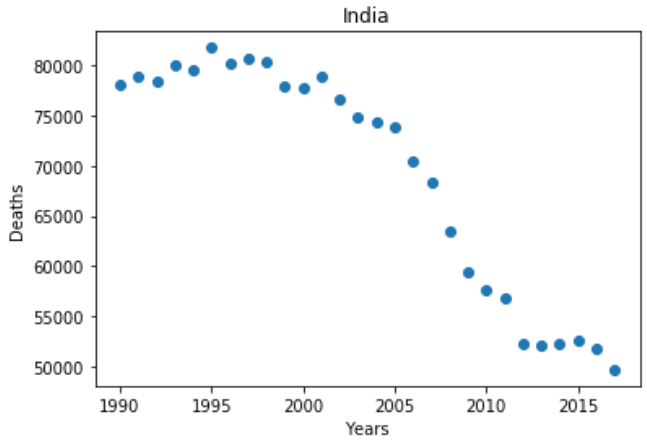


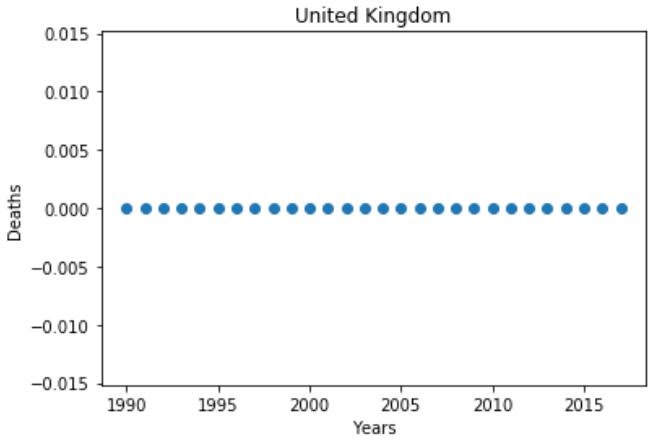
MALARIA DEATHS RESULT

From our favorite list of countries we can conclude that there is almost no death because of Malaria in United States, UK and Canada

But Sadly Malaria deaths in India and Pakistan are significantly high around~ 90000 in year between 1990 to 2000 and around~ 6000 in year 2000- 2005 in Pakistan

Although India is decreases the death rate of malaria from last 20 years, but It needs to be eradicated as US and UK. It shows that India and Pakistan is more unhygienic and lives in dirty environment that leads to Malaria mosquitoes





TUBERCULOSIS DEATHS RESULT

In India year between 1990- 2000 there were around 600000 died because of TB which was reduces to 400000 in year 2017. But it should be reduced more and India needs to work on the TB health issues.

While in US, UK and Canada it already reduces to less than 800 people died due to TB from year 2000- 2017