MDSC-102

Final Lab Assignment

Dataset Description:

The dataset consists of various economic indicators of six major countries in the world, i.e., USA, UK, China, Germany, India and Japan. These economic indicators like:

GDP per capita, GDP growth(annual), unemployment, inflation, Import and export of goods, Govt debt, Population growth, Life expectancy and poverty are used to observe the growth and economic trend lines of the countries over the period of 1990-2020.

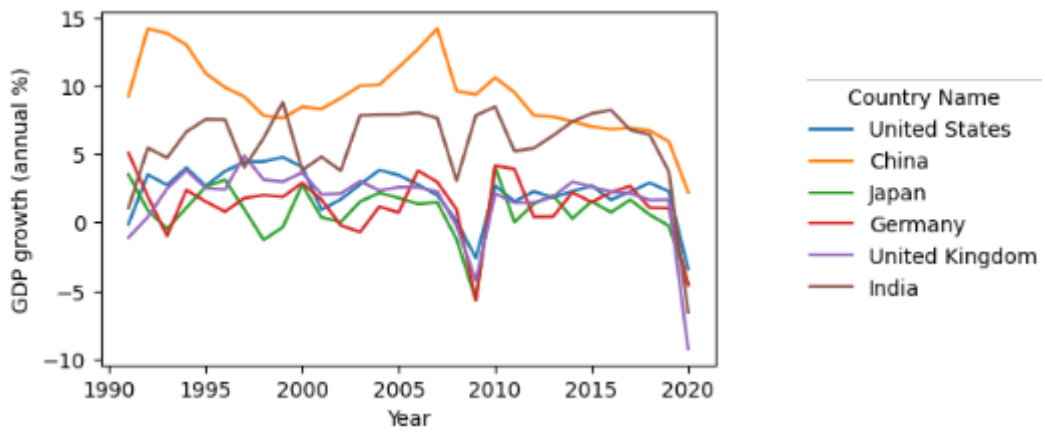
The dataset has null values only in Central government debt and poverty headcount columns. So, these null values are handled by filling them with 0.0 in those columns as shown below:

```
df.fillna(0.0,inplace=True)
df.isnull().sum()
```

```
Central government debt, total (% of GDP)                    0
Poverty headcount ratio at $1.90 a day (2011 PPP) (% of population)    0
```

Now, let's observe the trend lines for various indicators over the period of 1990-2020.

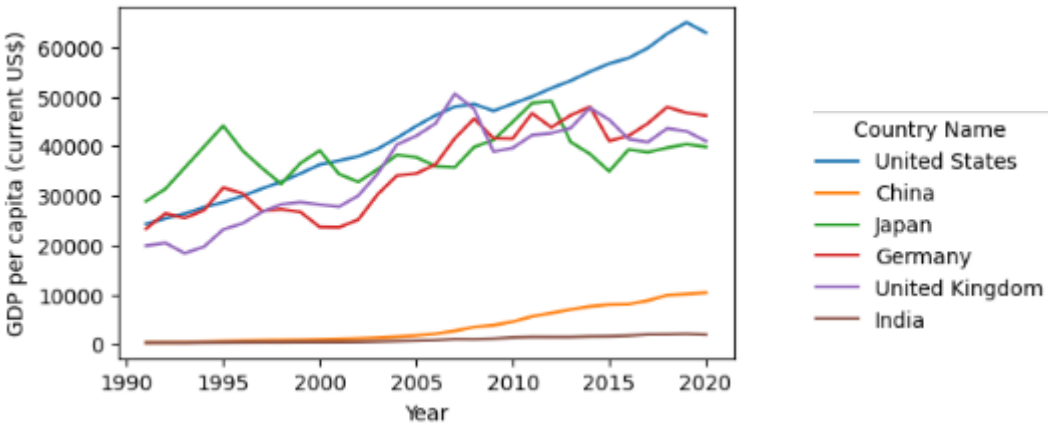
1. GDP growth (annual %)



All the countries GDP's growth rate ranges over a scale of (-5 to +5) % while China's GDP growth rate was b/w (10-15) %. Therefore, China has emerged to be the most powerful in terms of GDP growth over this period.

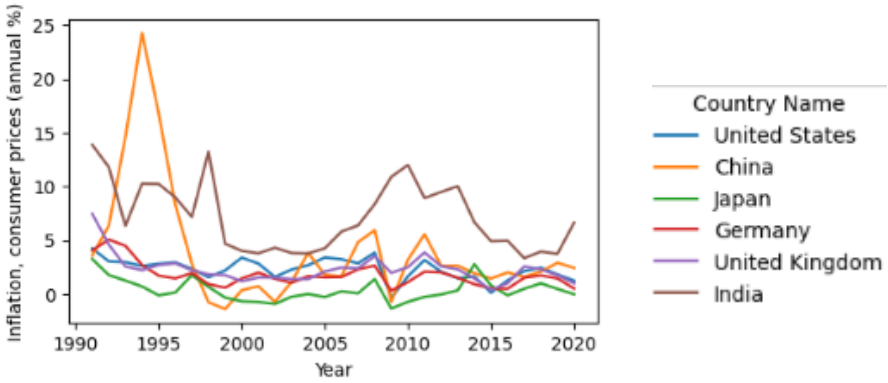
Also, in 2008 and 2020 all the countries' economies observed a sharp decrease which says that they had suffered from economic crisis, may be due to some epidemic like COVID in 2020.

2. GDP per Capita **– GDP / total population of the country**



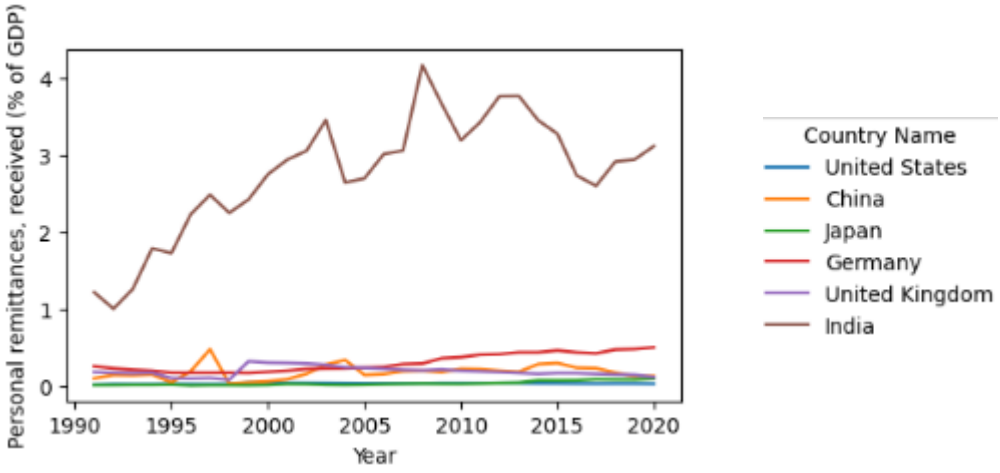
Even though China's GDP growth was largest among all the countries, GDP per capita of US is the largest among all the countries. This means US population is much smaller than China's and China's population is largest among all other countries. This means average income of a person in US is largest and least in India.

3. **Inflation in the country – how prices very increasing per product**



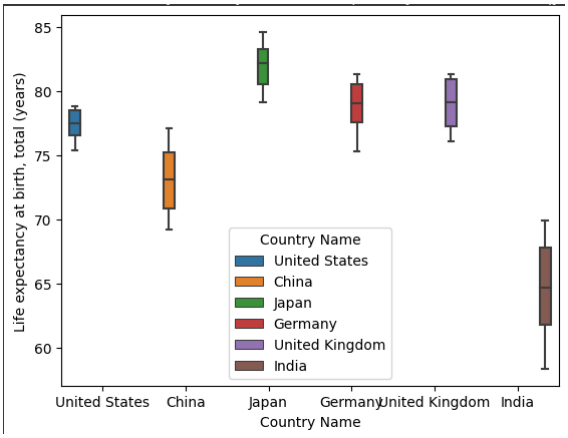
China had faced a huge inflation in 1993 which means there was a financial crisis in that year. Later on, it learnt to control it while Japan had faced the least inflation. And India was always the top one in terms of inflation which is a bad sign for its social condition.

4. **Personal Remittance – which means citizens of one country living in abroad for occupation and run livelihood in their home country by financial aid.**



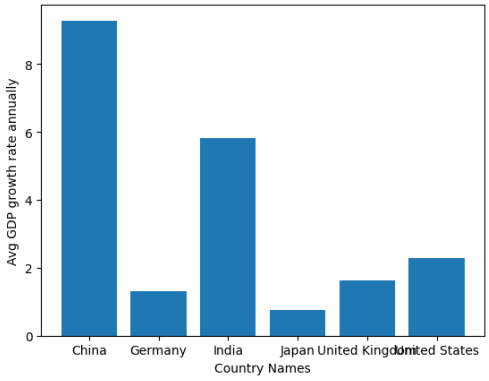
It can be observed that Indian citizens are living in abroad in large numbers to run their families in their home country due to which India's economic growth rate may be much smaller and facing inflation and unemployment.

5. **Life expectancy – average life span of a person in each country**



Japan always had the highest average lifespan while India has the least and is spread over a wide range of 63-67 years. While US has the least range from 76-79 years.

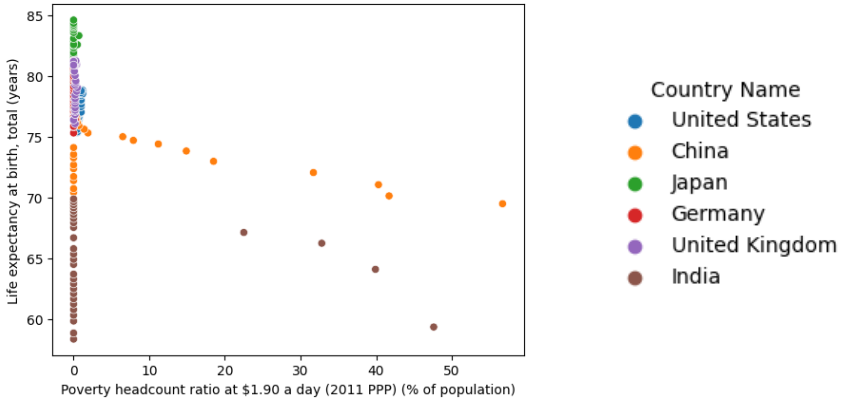
Average GDP growth rate over all the years from 1990-2020



China topped the list with: 10% / year

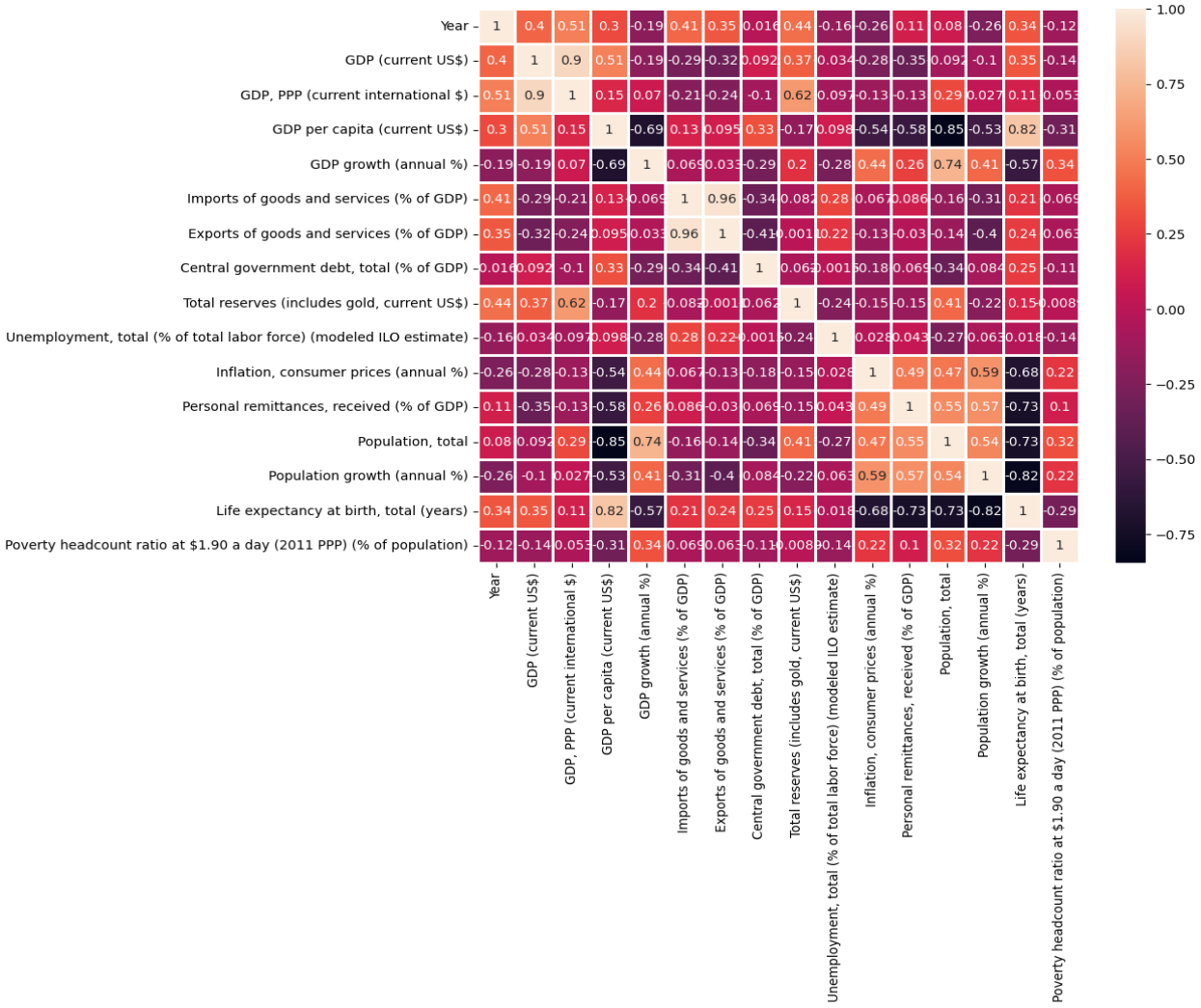
India with: 6% / year

US with: 3% / year, followed by UK, Germany and Japan.



Poverty vs Life expectancy – As poverty increases, the Life expectancy decreases. And over the years, poverty occurred only in India and China and hence the life expectancy, i.e., average life span of a person decreased.

Heat map depicting the co-relation b/w various economic indicators



Based on the Heat map of co-relation b/w various features: -

a) **GDP per capita vs personal remittance: -0.58**

That means, if more citizens of one country live/work in abroad, their own country's GDP growth will be less and the living standards of their country will degrade.

b) **Import and export of goods: 0.96**

As imports of goods in a country increases, their exports also increase.

c) **Inflation vs population growth: 0.59**

As population of a country increases, cost/product, access to resources also become expensive and therefore increases the cost of living in that country provided its GDP growth is smaller than its population growth.

d) **Total reserves vs year: 0.44**

As time progressed, countries started to preserve their resources, earnings and their wealth in order to live through any future financial crisis to prevent any social anarchy.

e) **GDP PPP vs GDP annual growth: 0.9**

GDP Purchasing power parity, i.e., the living standards of a country increases as its GDP annual growth increases.

f) **GDP per capita vs Life Expectancy: 0.82**

As GDP per capita, i.e., if average income of a person in a country increases they can live long by accessing any available resources and can lead healthy life and hence their average life span increases.

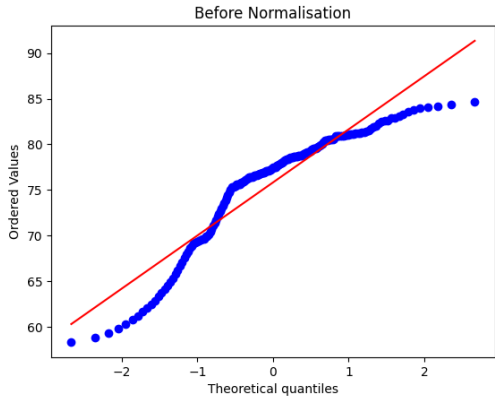
Hypothesis Testing for mean (Average life expectancy of a person):

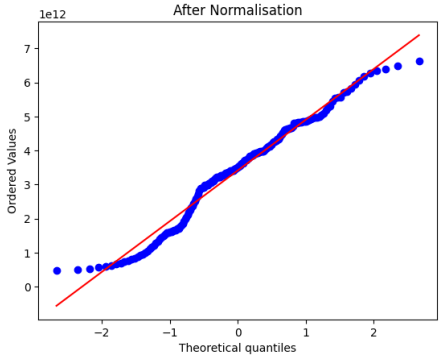
According to the records, the average life expectancy of a person in the world is 72.72 years as of 2020. I'm performing hypothesis testing on this based on the 180 samples that I have in my data frame.

H_0: $\mu = 72.72$ VS H_1: $\mu \neq 72.72$

$\alpha = 0.05$ (level of significance)

Data before and after normalisation






```
# loading the data sample from df
sample = df['Life expectancy at birth, total (years)']

# setting level of significance and null hypothesis
alpha = 0.05
mu_0 = 72.72

# for normalising the data using boxcox
temp , lamda = stats.boxcox(sample)
old_skew = sample.skew()
print('lambda = ',lamda)

# printing the skweness before and after normalisation
print('skewness before normalisation = ',old_skew)

# sample = (temp - np.mean(temp))/np.std(temp)
# sample = pd.Series(sample)

print('skewness = ', pd.Series(temp).skew())
print('distribution of the data is standard normal with parameters : ',temp.mean(),temp.var())

# variance is unknown but sample size is large , hence we use z-test
z_calc = (np.mean(temp) - (mu_0**lamda - 1)/lamda)/((np.var(temp)/(len(sample)))**0.5)
```

```
lambda =  7.0930666786769505
skewness before normalisation =  -1.112432785369464
skewness =  -0.19484222635952386
distribution of the data is standard normal with parameters :  3414764220801.7456 2.2158095045358187e+24
z_calc =  10.412944918747307
```

```python
# calculating the p-value
p = 2 * (1 - stats.norm.cdf(abs(z_calc)))

print('p_value = ',p)

# if p value is less than level of significance , we reject our H0
if p < alpha:
    print('Reject - Avg life span of a person != 72.72')

else:
    print('Do not reject - Avg life span of a person = 72.72')
```

```
p_value =  0.0
Reject - Avg life span of a person != 72.72
```