CASE STUDY

Poverty, life expectancy, and Gross Domestic Product (GDP) are three critical indicators that reflect a country's economic and social conditions.

In Africa, these indicators vary widely between countries, making it challenging to understand the complex relationship between poverty, life expectancy, and GDP.

This case study aims to provide a deep dive into the correlation between these indicators in African countries.

Python -for JUPYTER NOTEBOOK

In [2]:

```python
# Import Libraries for the project
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import missingno as msno
import numpy as np
from collections import Counter
%matplotlib inline
sns.set()
pd.set_option('display.max_columns', None)
from subprocess import check_output
import warnings
warnings.filterwarnings("ignore")
from scipy.stats import skew, kurtosis
```

```python
df = pd.read_excel(r"C:\Users\HENRY OKEOMA\Downloads\life-expectancy 2.xlsx")
df
```

Out[3]:

	Code	Year	Life expectancy at birth (historical)	CountryName	IncomeGroup	Region
0	AGO	1940	26.98	Angola	Lower middle income	Middle Africa
1	AGO	1950	36.30	Angola	Lower middle income	Middle Africa
2	AGO	1951	36.40	Angola	Lower middle income	Middle Africa
3	AGO	1952	36.50	Angola	Lower middle income	Middle Africa
4	AGO	1953	36.70	Angola	Lower middle income	Middle Africa
...	...	...	...	...	...	...
3932	ZWE	2017	60.70	Zimbabwe	Lower middle income	Eastern Africa
3933	ZWE	2018	61.40	Zimbabwe	Lower middle income	Eastern Africa
3934	ZWE	2019	61.30	Zimbabwe	Lower middle income	Eastern Africa
3935	ZWE	2020	61.10	Zimbabwe	Lower middle income	Eastern Africa
3936	ZWE	2021	59.30	Zimbabwe	Lower middle income	Eastern Africa

3937 rows × 6 columns

In [103]:

```python
# We merged the different excel sheets of life expectancy with the country codes after s
```

In [104]:

```python
# Call the columns, so that we can ammend them directly here
df.columns
```

Out[104]:

```
Index(['Code', 'Year', 'Life expectancy at birth (historical)', 'Unnamed: 3',
       'Unnamed: 4', 'Unnamed: 5'],
      dtype='object')
```

In [105]:

```python
# Column Name ammendment

df.columns = ['Code', 'Year', 'Life expectancy at birth (historical)', 'CountryName',
       'IncomeGroup', 'Region']
df.head(2)
```

Out[105]:

	Code	Year	Life expectancy at birth (historical)	CountryName	IncomeGroup	Region
0	AGO	1940	26.98	Angola	Lower middle income	Middle Africa
1	AGO	1950	36.30	Angola	Lower middle income	Middle Africa

In [106]:

```python
df.isnull().sum()
```

Out[106]:

```
Code                                     0
Year                                     0
Life expectancy at birth (historical)    0
CountryName                              0
IncomeGroup                              0
Region                                   0
dtype: int64
```

We have no missing data in our data set

In [8]:

```python
# Check for Duplicates
df.duplicated().sum()
```

Out[8]:

```
0
```

We have no duplicates

```
# Check for the statistical datA of our numerical column
df.describe().astype(int)
```

Out[107]:

	Year	Life expectancy at birth (historical)
count	3937	3937
mean	1984	51
std	21	10
min	1921	12
25%	1967	43
50%	1985	50
75%	2003	58
max	2021	76

Insight: We can see that we have both the mean and median of the Life expectancy almost the same, we shall visaulise this using Histogram

```python
# Histogram of Life Expectancy Across African Regions
sns.histplot(x='Life expectancy at birth (historical)', data=df, bins=10, kde=True);
plt.title('Life expectancy at birth (historical) distribution Across Africa')
plt.show()
```
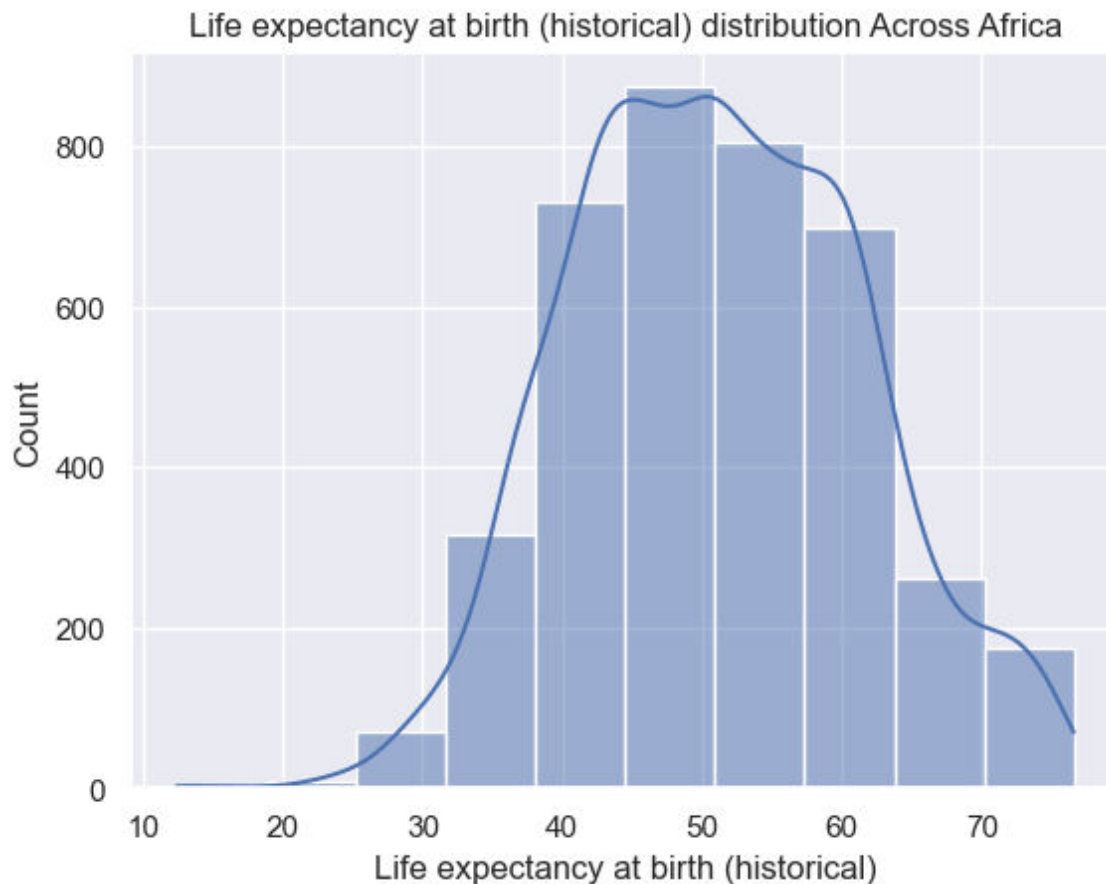
```python
print(df['Life expectancy at birth (historical)'].skew())
print(df['Life expectancy at birth (historical)'].kurtosis())
```

```
0.10413540700310524
-0.37377248673122
```

Insight: We have an almost a normal distribution for age of life expectance with most between 45 - 55yrs. Numbers sharply reduce as Africans get to 60yrs. Skewness of 0.1 is very close for a normal distribution.

Kurtosis is platykurtic, basically showing slight outliers in the left

In [110]:

```python
# Checking All Countries in Africa from Lowest to Hishest in Life Expectancny.
Lowest = df.groupby('CountryName')[['Life expectancy at birth (historical)']].max().sort_
(by='Life expectancy at birth (historical)').astype(int)
Lowest
```

Out[110]:

CountryName	Life expectancy at birth (historical)
Nigeria	52
Chad	53
Central African Republic	55
South Sudan	55
Somalia	57
Côte d'Ivoire	59
Lesotho	59
Mali	59
Guinea	59
Burkina Faso	60
DR Congo	60
Sierra Leone	60
Benin	60
Guinea-Bissau	60
Liberia	61
Mozambique	61
Cameroon	61
Togo	61
Equatorial Guinea	61
Zimbabwe	62
Angola	62
Burundi	62
Zambia	62
Kenya	62
Niger	62
Uganda	63
Namibia	63
Eswatini	63
Djibouti	63
Gambia	63
Congo	64
Malawi	64
Comoros	64
Ghana	64
Mauritania	65
Ethiopia	65

Life expectancy at birth (historical)

CountryName	
Sudan	65
Madagascar	65
South Africa	66
Gabon	66
Botswana	66
Rwanda	66
Tanzania	67
Eritrea	67
Sao Tome & Principe	68
Senegal	68
Egypt	71
Libya	72
Seychelles	74
Morocco	74
Mauritius	75
Tunisia	76
Algeria	76
Cabo Verde	76

Insight: Nigeria, Chad, Central African Republic have the lowest, while Tunisia, Algeria and Cabo Verde (Cape Verde)

BIVARIATE ANALYSIS.

Data Visualisations for better interaction

In [111]:

```python
# The first 15 countries with the least life expectancy
dfx = Lowest.head(10).reset_index()
dfx
```

Out[111]:

|	CountryName	Life expectancy at birth (historical)
0	Nigeria	52
1	Chad	53
2	Central African Republic	55
3	South Sudan	55
4	Somalia	57
5	Côte d'Ivoire	59
6	Lesotho	59
7	Mali	59
8	Guinea	59
9	Burkina Faso	60

```python
# Visualise the top 10 countries with the lowest lif expectancy
plt.figure(figsize=(10,5))

plt.title('Top_10 African Countries with the lowest Life expectancy by Age')

ax = sns.barplot(y='CountryName', x='Life expectancy at birth (historical)', data=dfx, c

# Add labels to each bar
#for index, row in dfx.iterrows():

    #ax.text(index, row.Lifeexpectancyatbirth(historical), row.Lifeexpectancyatbirth(his
```


Top_10 African Countries with the lowest Life expectancy by Age
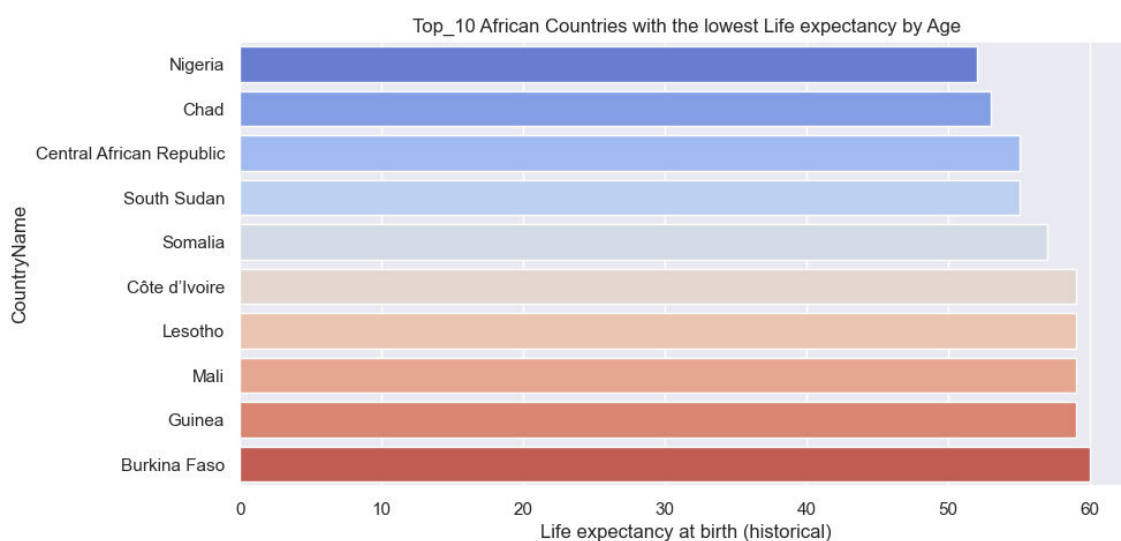
Insight: Nigeria to Burkinafaso are the top 10 countries with the least life expectancy. it is worthy to note that 5 of these countries are in west african region, with 2 in middle african and only 1 in Southern region

In [113]:

```python
# Checking also the top 10 life expectancy by Age
Top_10 = Lowest.tail(10).reset_index()
Top_10
```

Out[113]:

	CountryName	Life expectancy at birth (historical)
0	Sao Tome & Principe	68
1	Senegal	68
2	Egypt	71
3	Libya	72
4	Seychelles	74
5	Morocco	74
6	Mauritius	75
7	Tunisia	76
8	Algeria	76
9	Cabo Verde	76

In [114]:

```python
# Top 10 Countries with the hihghest Life Expectancy
plt.figure(figsize=(8,5))
sns.barplot(y='CountryName', x='Life expectancy at birth (historical)', data=Top_10, pal
plt.title('Top_10 African Countries with the Highest Life expectancy by Age');
```
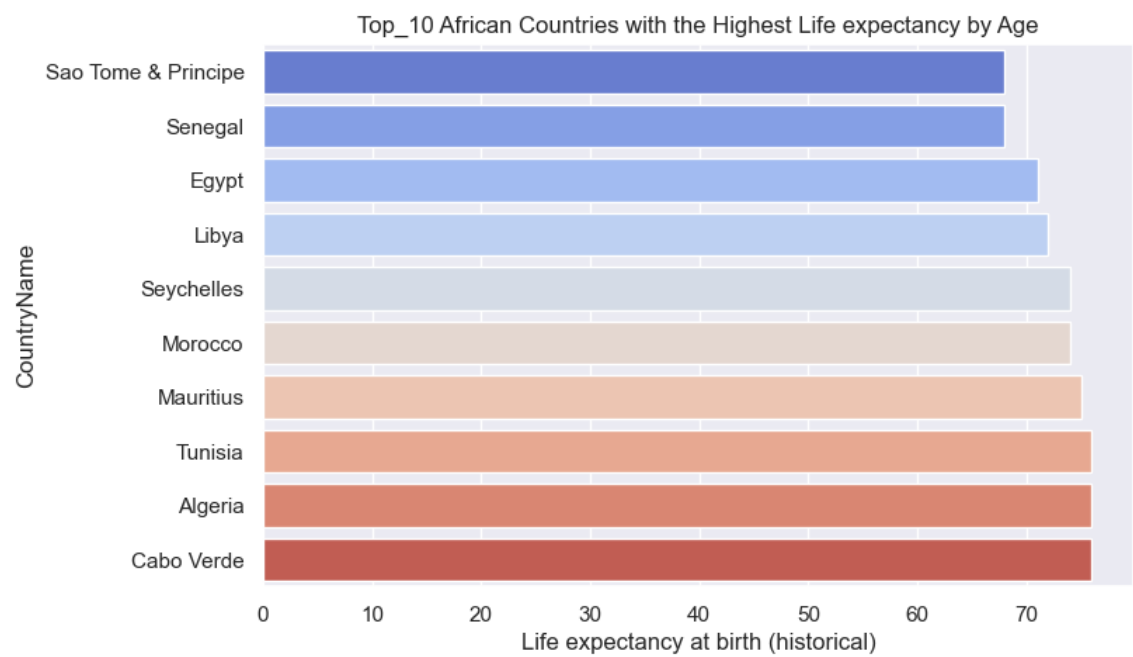
Insight: Cape Verde, Algeria and Tunisia are leading, while STP, Senegal and Egypt are trailing. A good of number (5) of the top_10 are in North Africa while the rest is share in the other region

```
# We Also try to group the years to see if there is a thrend with the life expectancy by
Relationship = round(df.groupby('Year')['Life expectancy at birth (historical)'].mean())
Relationship
```

Out[115]:

	Year	Life expectancy at birth (historical)
0	1921	26
1	1923	29
2	1924	31
3	1927	26
4	1928	31
...	...	...
93	2017	63
94	2018	64
95	2019	64
96	2020	64
97	2021	63

98 rows × 2 columns

```python
# We shall try to visualise the Mean of the life expectancy over the years
plt.figure(figsize=(10,8))
sns.scatterplot(y='Life expectancy at birth (historical)', x='Year', data=Relationship)
plt.title('Relationship between Years and Mean of Life Expectancy in Africa')
plt.show()
```


Relationship between Years and Mean of Life Expectancy in Africa
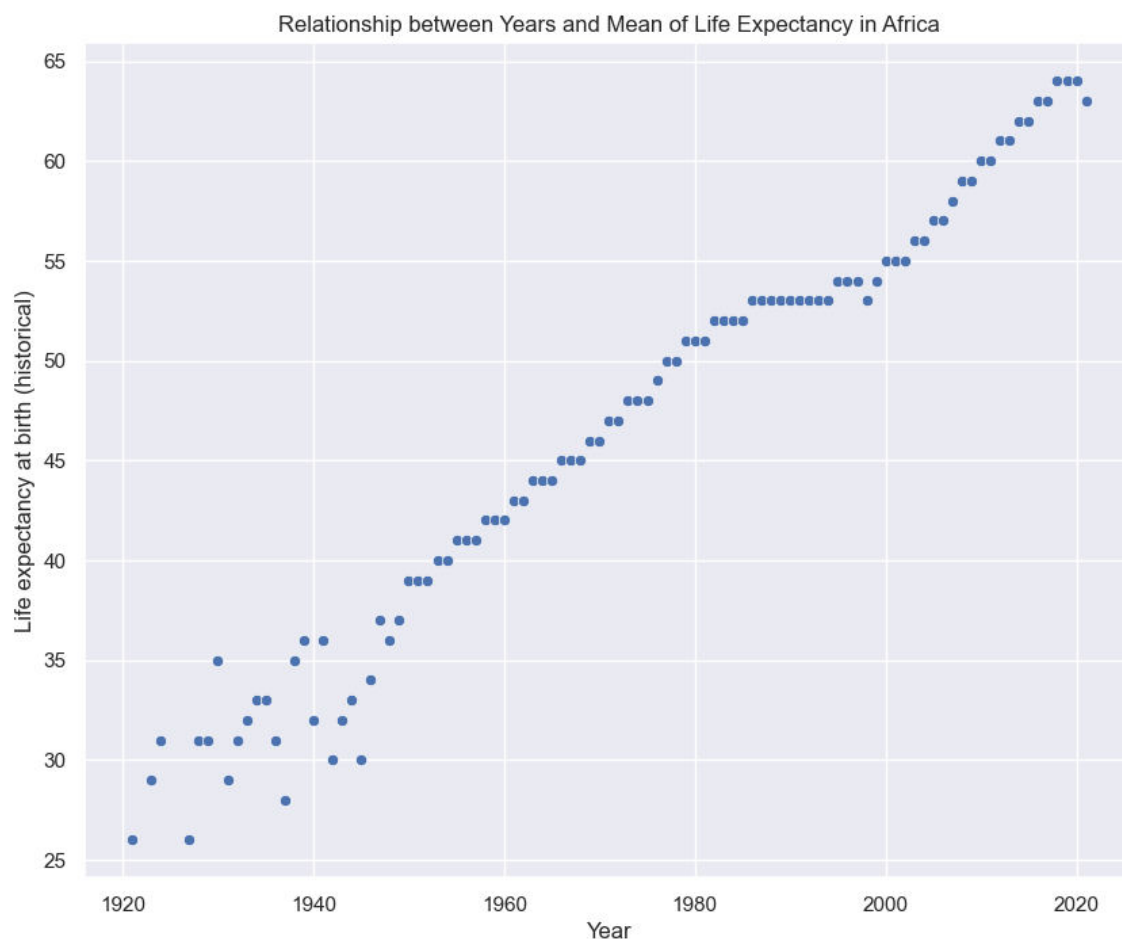
Insight: The Life expectancy in Africa have increased over the years, with goo correlation between the years and ages; looking at the mean values, there has been good and steady improvement

```python
# Checking which incomegroup have the least life expectancy using the group function
df3 = df.groupby('IncomeGroup')['Life expectancy at birth (historical)'].mean().astype(i
df3
```
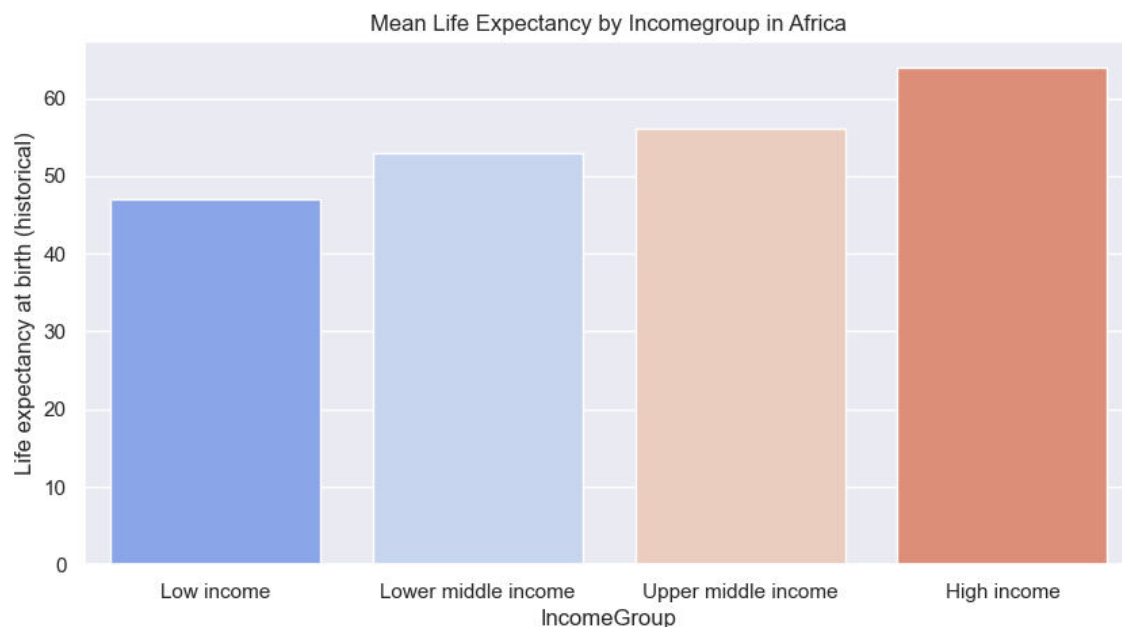
	IncomeGroup	Life expectancy at birth (historical)
0	Low income	47
1	Lower middle income	53
2	Upper middle income	56
3	High income	64

```
# Visualise
plt.figure(figsize=(10,5))
plt.title('Mean Life Expectancy by Incomegroup in Africa')
ax = sns.barplot(y='Life expectancy at birth (historical)', x='IncomeGroup', data=df3, p

plt.show()
```


Insight: Life expectancy increases as the earning capacity increase,

In [120]:

```
# We shall also compare the life expectancy across africa with the Region. to see which
df4 = df.groupby('Region')['Life expectancy at birth (historical)'].mean().astype(int).s
df4
```

Out[120]:

	Region	Life expectancy at birth (historical)
0	Western Africa	48
1	Eastern Africa	50
2	Middle Africa	50
3	Southern Africa	54
4	Northern Africa	56

```python
# We shall also visualise the data above to see where we are
plt.figure(figsize=(10,5))
plt.title('Mean Life Expectancy by Regions in Africa')
ax = sns.barplot(y='Life expectancy at birth (historical)', x='Region', data=df4, palett

plt.show()
```


Mean Life Expectancy by Regions in Africa
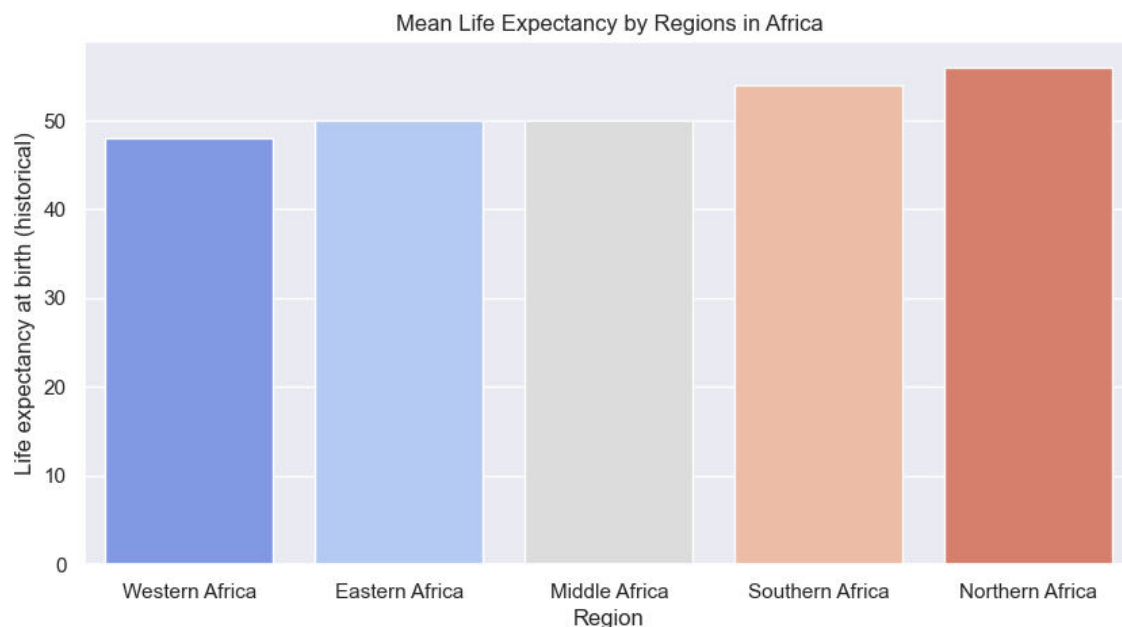
Northern Africa and Southern Africa have the highest Figures with the Western region as lowest mean ages.

MULTIVARIATE ANALYSIS

- 2 Catergories against the Age

In [122]:

```python
# WE SHALL COMPARE THE REGION, INCOME GROUP AND AGAINST LIFE EXPECTANCY
Region2 = df.groupby(['Region', 'IncomeGroup'])['Life expectancy at birth (historical)']
.sort_values().reset_index()
Region2
```

Out[122]:

	Region	IncomeGroup	Life expectancy at birth (historical)
0	Middle Africa	Low income	46
1	Western Africa	Low income	46
2	Eastern Africa	Low income	47
3	Middle Africa	Lower middle income	51
4	Southern Africa	Lower middle income	51
5	Western Africa	Lower middle income	51
6	Eastern Africa	Lower middle income	52
7	Middle Africa	Upper middle income	52
8	Northern Africa	Low income	53
9	Northern Africa	Lower middle income	56
10	Southern Africa	Upper middle income	56
11	Northern Africa	Upper middle income	60
12	Eastern Africa	Upper middle income	63
13	Eastern Africa	High income	64

```python
# We shall go ahead and perform A multivariate analysis to visualise th above
plt.figure(figsize=(15,8))
plt.title('Mean Life Expectancy by Region and IncomeGroup in Africa')
ax = sns.barplot(x='Region', y='Life expectancy at birth (historical)', data=Region2, pa
plt.show()
```
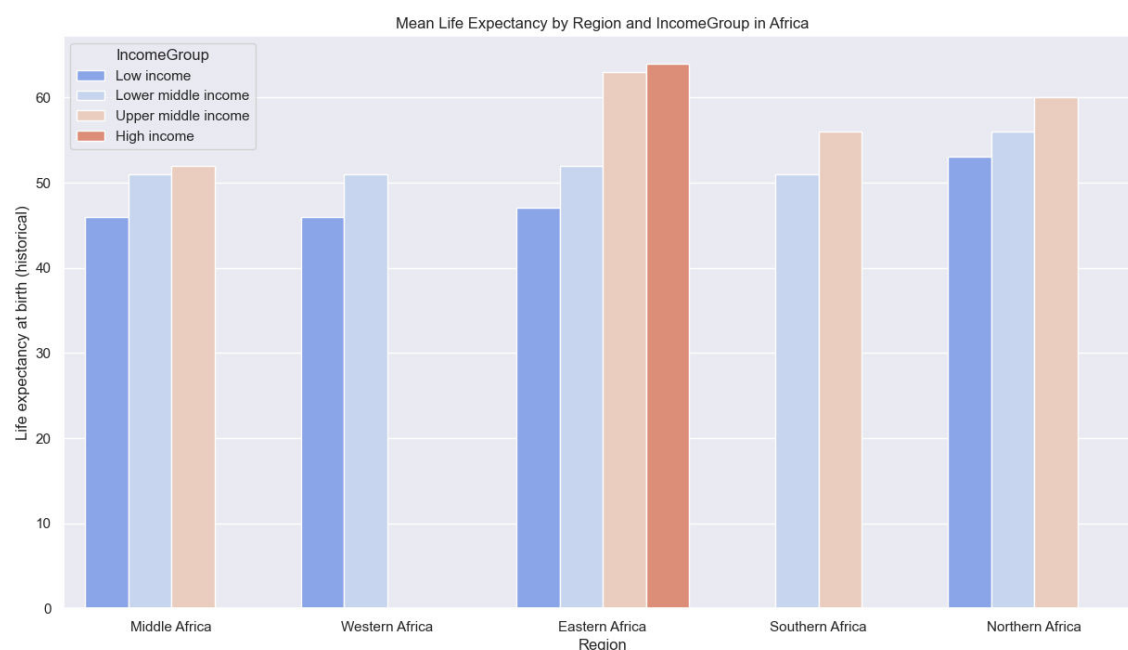
Insights: Only one region in Africa records a 'High Income Class', and the income is also widely distributed which is the Eastern Region.

Worthy to Note that low income class have an average life expactancy of 45years across Middle, western and Eastern region, except the Northern region.

Across all regions, as the income class increase, life expectancy also increase

In [124]:

```
# We need to identity which Country(s) in the Eastern African Region that we have the 'H
df[df['IncomeGroup'] == 'High income'][['CountryName','IncomeGroup', 'Region']]
```

Out[124]:

	CountryName	IncomeGroup	Region
3283	Seychelles	High income	Eastern Africa
3284	Seychelles	High income	Eastern Africa
3285	Seychelles	High income	Eastern Africa
3286	Seychelles	High income	Eastern Africa
3287	Seychelles	High income	Eastern Africa
...	...	...	...
3350	Seychelles	High income	Eastern Africa
3351	Seychelles	High income	Eastern Africa
3352	Seychelles	High income	Eastern Africa
3353	Seychelles	High income	Eastern Africa
3354	Seychelles	High income	Eastern Africa

72 rows × 3 columns

Seychelles in Eastern African remains the only country in Africa with incomegroup as High income

```
# We shall also perform a multivariate analysis on the Mean life expectancy over the yea
#Year and income group
Region3 = df.groupby(['Year', 'IncomeGroup'])['Life expectancy at birth (historical)'].m
.sort_values().reset_index()
Region3
```

Out[125]:

	Year	IncomeGroup	Life expectancy at birth (historical)
0	1927	Low income	23
1	1921	Low income	24
2	1931	Low income	24
3	1927	Lower middle income	26
4	1937	Low income	26
...	...	...	...
319	2013	High income	73
320	2010	High income	73
321	2017	High income	73
322	2018	High income	74
323	2019	High income	74

324 rows × 3 columns

```python
# We shall try to visualise the Mean of the life expectancy over the years
plt.figure(figsize=(10,8))
sns.scatterplot(x='Year', y='Life expectancy at birth (historical)', data=Region3, hue='
plt.title('Relationship between Years, IncomeGroup and Mean of Life Expectancy in Africa
plt.show()
```
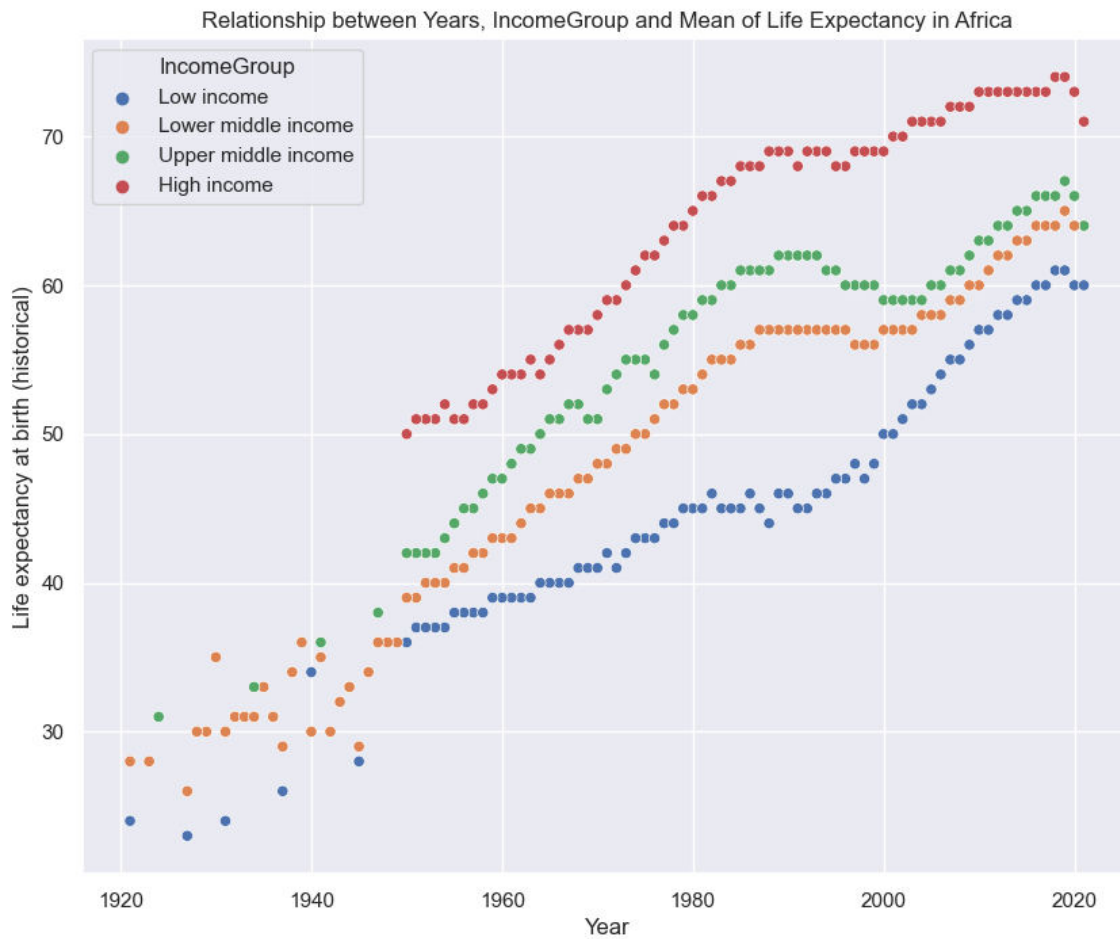
Insight: Life expectancy in general is at its lowest in 1920's and 1940's for both the Low and Lower middle income earners.

From 1950s, the segregation between the various income class became distinct. Although steady progressively, the low income still have the lowest.

In [127]:

```python
# We shall also perform a multivariate analysis on the Mean life expectancy over the yea
#Year and income group
Region3 = df.groupby(['Year', 'Region'])['Life expectancy at birth (historical)'].mean()
.sort_values().reset_index()
Region3
```

Out[127]:

	Year	Region	Life expectancy at birth (historical)
0	1927	Eastern Africa	23
1	1927	Western Africa	23
2	1931	Western Africa	24
3	1937	Western Africa	26
4	1937	Eastern Africa	26
...	...	...	...
395	2016	Northern Africa	72
396	2019	Northern Africa	72
397	2017	Northern Africa	72
398	2018	Northern Africa	72
399	2020	Northern Africa	72

400 rows × 3 columns

```python
# We shall try to visualise the Mean of the life expectancy over the years
plt.figure(figsize=(10,8))
sns.scatterplot(x='Year', y='Life expectancy at birth (historical)', data=Region3, hue='
plt.title('Relationship between Years, Region and Mean of Life Expectancy in Africa')
plt.show()
```
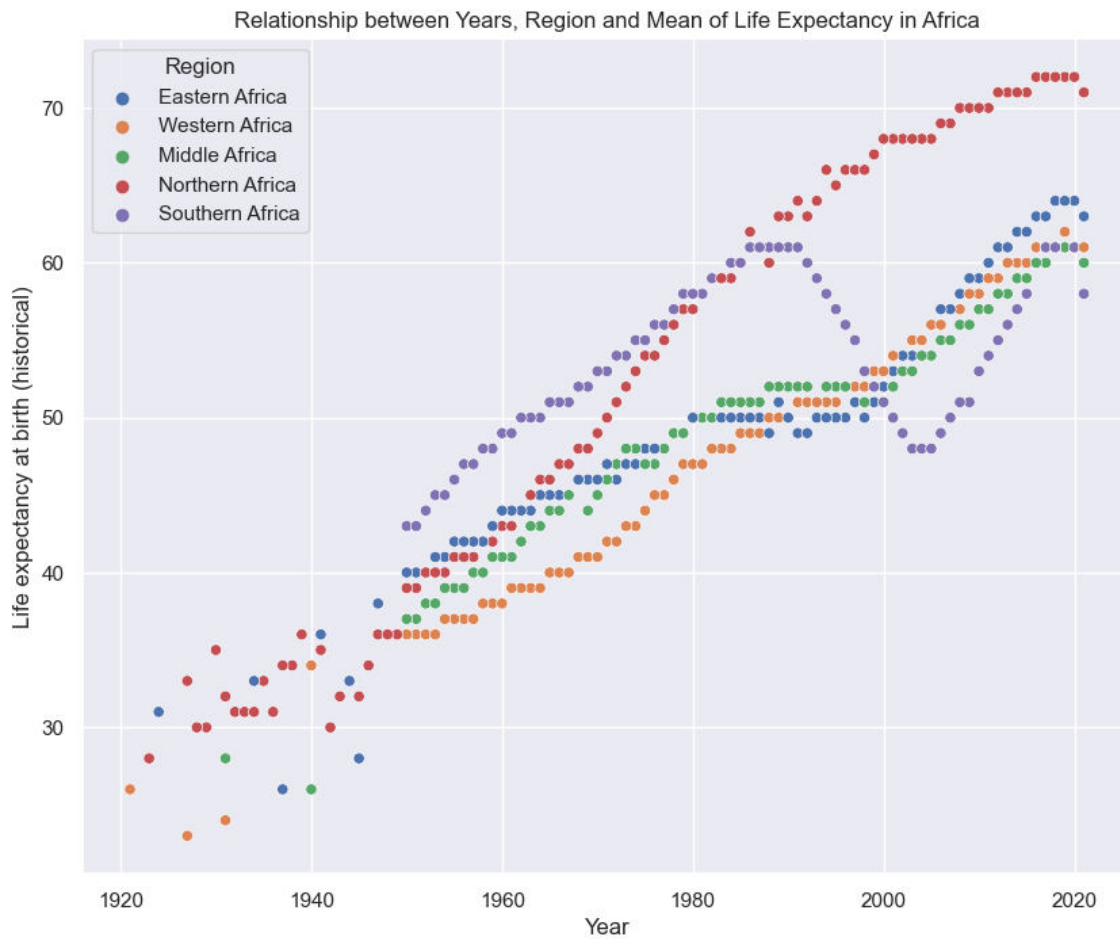
Insight: All regions have experienced a consistent growth in life expectancy with south african region leading before the sharp decline in 1990s up till 2002 (approx).

The Sharp decline is worthy of further investigations, but by research the region experienced a sudden drop beginning after 1995, as the HIV/AIDS epidemic spread throughout the region, beginning in the early 1990s.

In [145]:

```
dfs = df.loc[df['Region'] == 'Southern Africa']
dfs
```

Out[145]:

	Code	Year	Life expectancy at birth (historical)	CountryName	IncomeGroup	Region
289	BWA	1950	44.9	Botswana	Upper middle income	Southern Africa
290	BWA	1951	45.3	Botswana	Upper middle income	Southern Africa
291	BWA	1952	45.9	Botswana	Upper middle income	Southern Africa
292	BWA	1953	46.5	Botswana	Upper middle income	Southern Africa
293	BWA	1954	47.1	Botswana	Upper middle income	Southern Africa
...	...	...	...	...	...	...
3788	ZAF	2017	65.4	South Africa	Upper middle income	Southern Africa
3789	ZAF	2018	65.7	South Africa	Upper middle income	Southern Africa
3790	ZAF	2019	66.2	South Africa	Upper middle income	Southern Africa
3791	ZAF	2020	65.3	South Africa	Upper middle income	Southern Africa
3792	ZAF	2021	62.3	South Africa	Upper middle income	Southern Africa

360 rows × 6 columns

```
dfs[dfs['Year'].between(1990, 2005)]
```

Out[146]:

|	Code	Year	Life expectancy at birth (historical)	CountryName	IncomeGroup	Region
329	BWA	1990	60.5	Botswana	Upper middle income	Southern Africa
330	BWA	1991	59.5	Botswana	Upper middle income	Southern Africa
331	BWA	1992	58.3	Botswana	Upper middle income	Southern Africa
332	BWA	1993	57.7	Botswana	Upper middle income	Southern Africa
333	BWA	1994	56.2	Botswana	Upper middle income	Southern Africa
...	...	...	...	...	...	...
3772	ZAF	2001	57.3	South Africa	Upper middle income	Southern Africa
3773	ZAF	2002	55.7	South Africa	Upper middle income	Southern Africa
3774	ZAF	2003	54.3	South Africa	Upper middle income	Southern Africa
3775	ZAF	2004	54.0	South Africa	Upper middle income	Southern Africa
3776	ZAF	2005	54.0	South Africa	Upper middle income	Southern Africa

80 rows × 6 columns

The Decline in age was remarkable between 1990 - 2005

```python
# We shall also perform a analysis on the count and  income group across the regions of
plt.figure(figsize=(8,5))

plt.title('Count of Income distribution by Region ')
ax = sns.countplot(x='Region', data=df, hue='IncomeGroup')
ax.legend(loc='upper right',bbox_to_anchor=(0.98, 0.9));
```
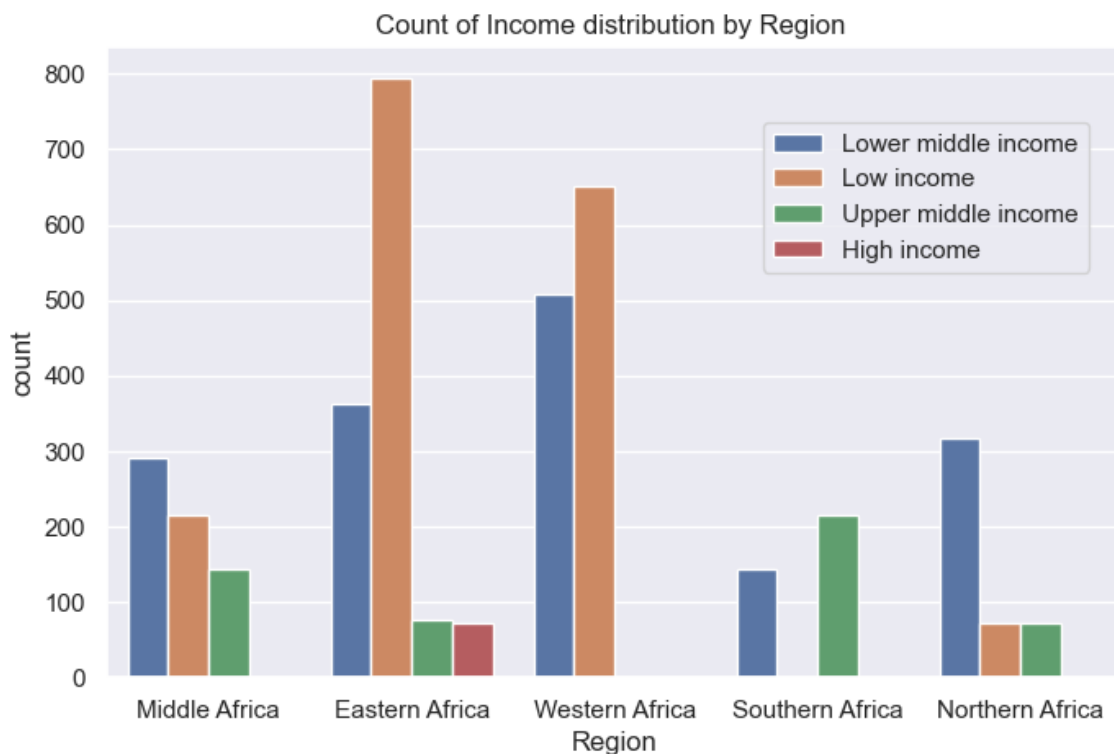

INSIGHT: The Eastern and Western Parts of Africa, have by count most of the Low and lower middle income earnings. Worse even in Western African region with no recorded Upper middle income

South Africa currently do not have low income earners.

Across all zones in Africa have a signifcant population of people with Lower middle income

High income earners are only present in Easter African and no other zone, which is Seychelles as we saw above

Continuation, we shall Import the Dataset for GDP and Population and Merge for Analysis

In [ ]:

```python
# Import Libraries for the project
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import missingno as msno
import numpy as np
from collections import Counter
%matplotlib inline
sns.set()
pd.set_option('display.max_columns', None)
from subprocess import check_output
import warnings
warnings.filterwarnings("ignore")
from scipy.stats import skew, kurtosis
import plotly.express as px
import squarify
```

In [2]:

```python
df = pd.read_excel(r"C:\Users\HENRY OKEOMA\Downloads\African Country GDP Data.xlsx")
df.head(3)
```

Out[2]:

	African Country Code	Country Name	IncomeGroup	Region	Indicator Name	1960	1961
0	AGO	Angola	Lower middle income	Middle Africa	GDP (current US$)	NaN	NaN	
1	BDI	Burundi	Low income	Eastern Africa	GDP (current US$)	1.960000e+08	2.030000e+08	2.135000
2	BEN	Benin	Lower middle income	Western Africa	GDP (current US$)	2.261956e+08	2.356682e+08	2.364349

In [3]:

```python
df.shape
```

Out[3]:

(54, 67)

In [4]:

```
# Checking for Duplicates
df.duplicated().sum()
```

Out[4]:

0

No Duplicates in the DataSet

In [4]:

```
df.isna().sum()
```

Out[4]:

```
African Country Code    0
Country Name            0
IncomeGroup             0
Region                  0
Indicator Name          0
                       ..
2017                    2
2018                    2
2019                    2
2020                    2
2021                    2
Length: 67, dtype: int64
```

We have a lot of missing data in the GDP for over the years, we shall try to describe the data and replace where necessary

```
# We shall try to melt the data set
df_melt = pd.melt(df, id_vars=['African Country Code','Country Name','IncomeGroup', 'Reg
                  var_name= 'Year', value_name='GDP_Current_USD')
df_melt
```

Out[5]:

	African Country Code	Country Name	IncomeGroup	Region	Indicator Name	Year	GDP_Current_USD
0	AGO	Angola	Lower middle income	Middle Africa	GDP (current US$)	1960	NaN
1	BDI	Burundi	Low income	Eastern Africa	GDP (current US$)	1960	1.960000e+08
2	BEN	Benin	Lower middle income	Western Africa	GDP (current US$)	1960	2.261956e+08
3	BFA	Burkina Faso	Low income	Western Africa	GDP (current US$)	1960	3.304428e+08
4	BWA	Botswana	Upper middle income	Southern Africa	GDP (current US$)	1960	3.041231e+07
...	...	...	...	...	...	...	...
3343	TZA	Tanzania	Lower middle income	Eastern Africa	GDP (current US$)	2021	6.784105e+10
3344	UGA	Uganda	Low income	Eastern Africa	GDP (current US$)	2021	4.052979e+10
3345	ZAF	South Africa	Upper middle income	Southern Africa	GDP (current US$)	2021	4.190150e+11
3346	ZMB	Zambia	Low income	Eastern Africa	GDP (current US$)	2021	2.214763e+10
3347	ZWE	Zimbabwe	Lower middle income	Eastern Africa	GDP (current US$)	2021	2.837124e+10

3348 rows × 7 columns

```
df_melt.isna().sum()
```

Out[6]:

```
African Country Code      0
Country Name              0
IncomeGroup               0
Region                    0
Indicator Name            0
Year                      0
GDP_Current_USD         495
dtype: int64
```

We have a TOTAL of 495 data missing in our GDP Figures

In [7]:

```
df_melt.head(5)
```

Out[7]:

	African Country Code	Country Name	IncomeGroup	Region	Indicator Name	Year	GDP_Current_USD
0	AGO	Angola	Lower middle income	Middle Africa	GDP (current US$)	1960	NaN
1	BDI	Burundi	Low income	Eastern Africa	GDP (current US$)	1960	1.960000e+08
2	BEN	Benin	Lower middle income	Western Africa	GDP (current US$)	1960	2.261956e+08
3	BFA	Burkina Faso	Low income	Western Africa	GDP (current US$)	1960	3.304428e+08
4	BWA	Botswana	Upper middle income	Southern Africa	GDP (current US$)	1960	3.041231e+07

We melted the Dataframe for easier analysis, but we still have some missing values on the GDP values

```
# Checking further the median of the Df_melt
df_melt.groupby('IncomeGroup')['GDP_Current_USD'].median()
```

Out[8]:

```
IncomeGroup
High income           3.714722e+08
Low income            2.072350e+09
Lower middle income   5.551653e+09
Upper middle income   6.594022e+09
Name: GDP_Current_USD, dtype: float64
```

In [9]:

```
df_melt.describe()
```

Out[9]:

	GDP_Current_USD
count	2.853000e+03
mean	1.827951e+10
std	5.178105e+10
min	9.122751e+06
25%	9.326486e+08
50%	3.259345e+09
75%	1.184019e+10
max	5.741838e+11

The Mean and the Median are not the same, as a good practice, we shall choose the median which is the 50% quartile to replace the GDP missing values, but this will be grouped by incomegroup. I consider this order necessary, hence using the groupby function

In [10]:

```
# Grouping the incomegroup with GDP datafram
df_melt['GDP_Current_USD'] = df_melt.groupby('IncomeGroup')['GDP_Current_USD'].transform
```

We have replaced all the Missing values using the above and lets call our dataframe to see this

In [11]:

```
df_melt.head(5)
```

Out[11]:

	African Country Code	Country Name	IncomeGroup	Region	Indicator Name	Year	GDP_Current_USD
0	AGO	Angola	Lower middle income	Middle Africa	GDP (current US$)	1960	5.551653e+09
1	BDI	Burundi	Low income	Eastern Africa	GDP (current US$)	1960	1.960000e+08
2	BEN	Benin	Lower middle income	Western Africa	GDP (current US$)	1960	2.261956e+08
3	BFA	Burkina Faso	Low income	Western Africa	GDP (current US$)	1960	3.304428e+08
4	BWA	Botswana	Upper middle income	Southern Africa	GDP (current US$)	1960	3.041231e+07

In [12]:

```
# Get the Max GDP
round(df_melt['GDP_Current_USD'].max())
```

Out[12]:

574183825592

In [17]:

```
df_melt[df_melt['GDP_Current_USD'] == df_melt['GDP_Current_USD'].max()]
```

Out[17]:

	African Country Code	Country Name	IncomeGroup	Region	Indicator Name	Year	GDP_Current_USD
2952	NGA	Nigeria	Lower middle income	Western Africa	GDP (current US$)	2014	5.741838e+11

Nigeria had the highest figure recorded in GDP

```python
# Get Min GDP
round(df_melt['GDP_Current_USD'].min())
```

Out[13]:

9122751

In [14]:

```python
df_melt[df_melt['GDP_Current_USD'] == df_melt['GDP_Current_USD'].min()]
```

Out[14]:

	African Country Code	Country Name	IncomeGroup	Region	Indicator Name	Year	GDP_Current_USD
130	GNQ	Equatorial Guinea	Upper middle income	Middle Africa	GDP (current US$)	1962	9.122751e+06

Equatorial Guinea had the lowest recorded figure in GDP. This was just before President Buhari took over the country

In [15]:

```python
df_melt.head(3)
```

Out[15]:

	African Country Code	Country Name	IncomeGroup	Region	Indicator Name	Year	GDP_Current_USD
0	AGO	Angola	Lower middle income	Middle Africa	GDP (current US$)	1960	5.551653e+09
1	BDI	Burundi	Low income	Eastern Africa	GDP (current US$)	1960	1.960000e+08
2	BEN	Benin	Lower middle income	Western Africa	GDP (current US$)	1960	2.261956e+08

In [ ]:

```python
# We shall bring in the second Data set on Poverty Population and merge with the previou
```

In [16]:

```python
df_pop = pd.read_excel(r"C:\Users\HENRY OKEOMA\Downloads\distribution-of-population-betw
```

```
df_pop.head()
```

	Country Code	Year	$40 - total number of people above poverty line	30−40 - total number of people between poverty lines	20−30 - total number of people between poverty lines	10−20 - total number of people between poverty lines	6.85−10 - total number of people between poverty lines	3.65−6.85 - total number of people between poverty lines	2.15−3.65 - total number of people between poverty lines	1− - t num pe betw pov l
0	DZA	1988	156887	156293	505149	3346519	4721041	9399430	4588806	156ℓ
1	DZA	1995	288	207792	720816	4492444	5764456	10598713	5313792	159′
2	DZA	2011	67878	310554	1157937	10208065	11508101	11935404	1306518	134
3	AGO	2000	281776	135165	503404	2183373	2272061	4537115	2972500	222′
4	AGO	2008	124834	158806	460152	2864525	3103557	6811821	4997330	274ℓ

```
# Checking for missing numbers
df_pop.isna().sum()
```

```
Country Code                                                    0
Year                                                            0
$40 - total number of people above poverty line                 0
$30-$40 - total number of people between poverty lines          0
$20-$30 - total number of people between poverty lines          0
$10-$20 - total number of people between poverty lines          0
$6.85-$10 - total number of people between poverty lines         0
$3.65-$6.85 - total number of people between poverty lines       0
$2.15-$3.65 - total number of people between poverty lines       0
$1-$2.15 - total number of people between poverty lines          0
$1 - total number of people below poverty line                  0
Total No of people Between Poverty Line                          0
dtype: int64
```

In [22]:

```python
# Change Column Name
df_pop.columns = ['African Country Code', 'Year',
        '$40 - total number of people above poverty line',
        '$30-$40 - total number of people between poverty lines',
        '$20-$30 - total number of people between poverty lines',
        '$10-$20 - total number of people between poverty lines',
        '$6.85-$10 - total number of people between poverty lines',
        '$3.65-$6.85 - total number of people between poverty lines',
        '$2.15-$3.65 - total number of people between poverty lines',
        '$1-$2.15 - total number of people between poverty lines',
        '$1 - total number of people below poverty line',
        'Total No of people Between Poverty Line']
df_pop.head(2)
```

Out[22]:

	African Country Code	Year	$40 - total number of people above poverty line	30−40 - total number of people between poverty lines	20−30 - total number of people between poverty lines	10−20 - total number of people between poverty lines	6.85− 10 - total number of people between poverty lines	3.65− 6.85 - total number of people between poverty lines	2.15− 3.65 - total number of people between poverty lines	1−2. - to numb peop betwe pove lin
0	DZA	1988	156887	156293	505149	3346519	4721041	9399430	4588806	15693
1	DZA	1995	288	207792	720816	4492444	5764456	10598713	5313792	15916

In [23]:

```python
# Dimensionality
df_pop.shape
```

Out[23]:

(239, 12)

In [24]:

```python
# Dimensionality
df_melt.shape
```

Out[24]:

(3348, 7)

```
merged_df = pd.merge(df_melt, df_pop, on='African Country Code', how='left')
merged_df.head(5)
```

Out[25]:

	African Country Code	Country Name	IncomeGroup	Region	Indicator Name	Year_x	GDP_Current_USD	Year_y	
0	AGO	Angola	Lower middle income	Middle Africa	GDP (current US$)	1960	5.551653e+09	2000.0	28
1	AGO	Angola	Lower middle income	Middle Africa	GDP (current US$)	1960	5.551653e+09	2008.0	12
2	AGO	Angola	Lower middle income	Middle Africa	GDP (current US$)	1960	5.551653e+09	2018.0	11
3	BDI	Burundi	Low income	Eastern Africa	GDP (current US$)	1960	1.960000e+08	1992.0	
4	BDI	Burundi	Low income	Eastern Africa	GDP (current US$)	1960	1.960000e+08	1998.0	

In [26]:

```
merged_df.shape
```

Out[26]:

(15004, 18)

```python
# Group the Rwgion,Country name with mean GDP
TopSA_NG = merged_df.groupby(['Region','Country Name'])['GDP_Current_USD'].sum()\
.sort_values(ascending=False).head(10).reset_index()
TopSA_NG
```

Out[27]:

	Region	Country Name	GDP_Current_USD
0	Western Africa	Nigeria	7.012178e+13
1	Southern Africa	South Africa	6.039062e+13
2	Northern Africa	Egypt	5.198301e+13
3	Northern Africa	Morocco	1.611594e+13
4	Northern Africa	Algeria	1.298995e+13
5	Western Africa	Côte d'Ivoire	1.221097e+13
6	Northern Africa	Tunisia	8.410399e+12
7	Western Africa	Ghana	6.869761e+12
8	Eastern Africa	Kenya	6.739274e+12
9	Eastern Africa	Ethiopia	5.989339e+12

```python
# Visualise using Treemap
px.treemap(data_frame=TopSA_NG, path=['Region','Country Name'], values='GDP_Current_USD'
 title='Overview of Total GDP of the TOP_10 Countries')
```

Overview of Total GDP of the TOP_10 Countries
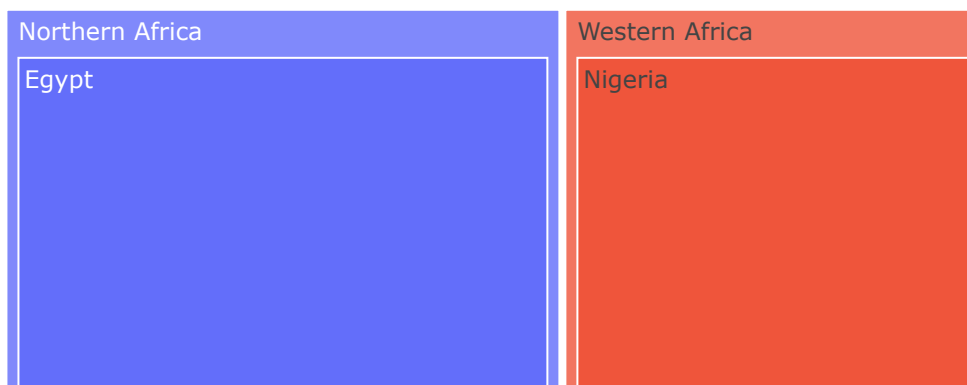
Insight: Nigeria, Egypt and South Africa have an outstanding number, but we shall also look at the population of the people in these countries

In [29]:

```python
# We shall also check the population Top 10 Countries betweeen Poverty Line earning betw
TopSA_NG1 = merged_df.groupby(['Region','Country Name'])['Total No of people Between Pov
.sort_values(ascending=False).head(10).reset_index()
TopSA_NG1
```

Out[29]:

	Region	Country Name	Total No of people Between Poverty Line
0	Western Africa	Nigeria	6.332017e+10
1	Northern Africa	Egypt	4.312263e+10
2	Eastern Africa	Ethiopia	2.252932e+10
3	Southern Africa	South Africa	1.595586e+10
4	Eastern Africa	Uganda	1.511443e+10
5	Eastern Africa	Tanzania	1.060545e+10
6	Northern Africa	Morocco	1.030718e+10
7	Western Africa	Côte d'Ivoire	1.011484e+10
8	Eastern Africa	Kenya	9.722755e+09
9	Western Africa	Ghana	7.328678e+09

```
# Visualise the above
px.treemap(data_frame=TopSA_NG1, path=['Region','Country Name'], values='Total No of peo
 title='Overview of Population of the TOP_10 Countries between Poverty Lines (Daily inco
```

Overview of Population of the TOP_10 Countries between Pov
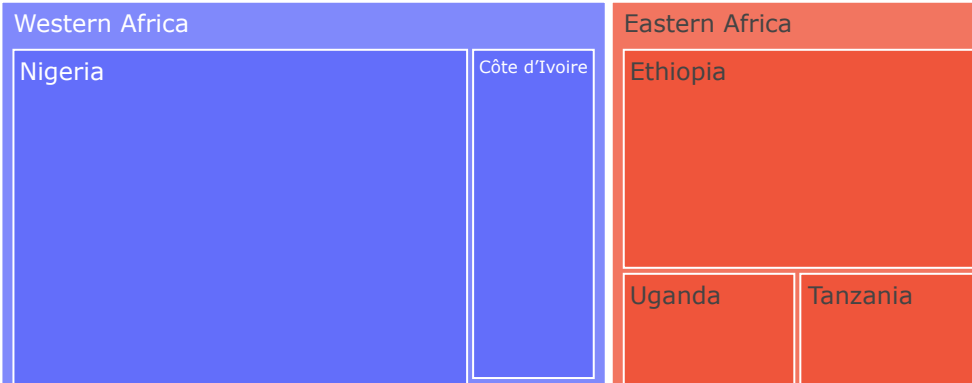
Insight: Nigeria and Egypt have also a good number of people between Poverty Lines and South Africa in comparisonn with their high GDP have minimum number of people between poverty line

We have Data for Countries above poverty Line (daily income >= USD40) and;

below Poverty Line (daily income below USD1)

In [52]:

```
# Total Populationn in different countries above Poverty Line
Top_10_Above = merged_df.groupby('Country Name')['$40 - total number of people above pov
.sum().astype(int).sort_values(ascending=False).head(10).reset_index()
Top_10_Above
```

Out[52]:

|	Country Name	$40 - total number of people above poverty line
0	South Africa	959642572
1	Egypt	134250274
2	Morocco	105308550
3	Tunisia	54632540
4	Kenya	36890682
5	Angola	35963348
6	Côte d'Ivoire	30319922
7	Nigeria	27290912
8	Ethiopia	18621328
9	Namibia	18248646

```
# Visualising the top_10 Countries Population above the Poverty Line
plt.figure(figsize=(8,5))

sns.barplot(y='Country Name', x='$40 - total number of people above poverty line', data=
plt.title('Top_10 Population of Countries Above the Poverty Line >40Usd');
```
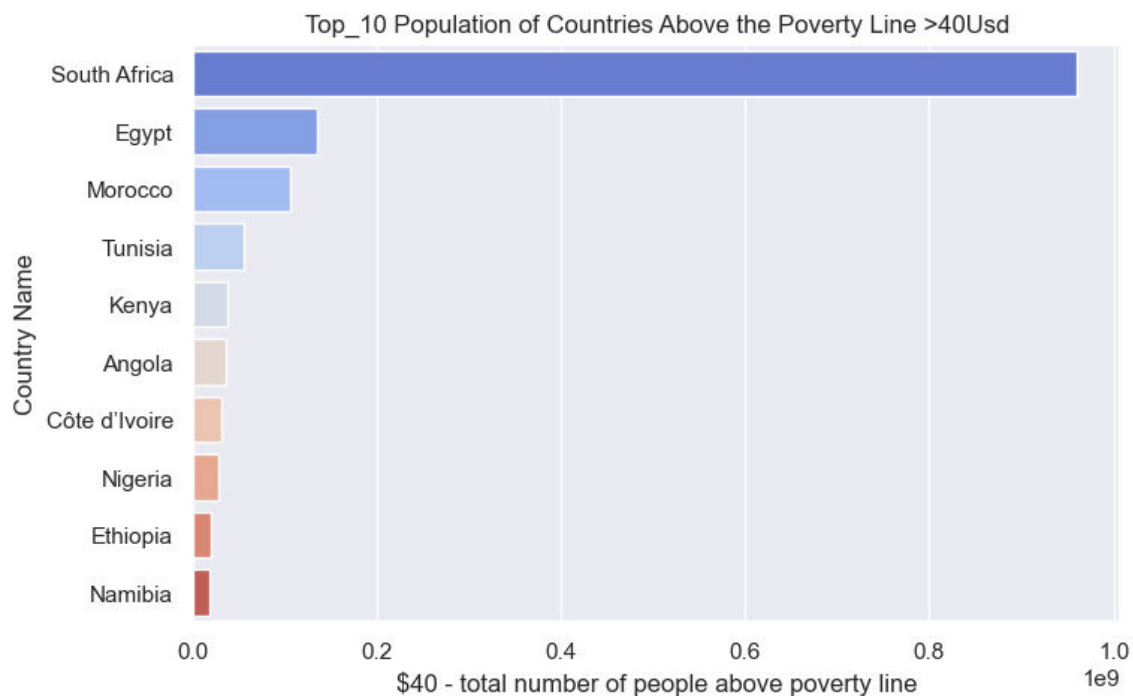
Insight: South African have a very distinct population of people above the poverty line, the following 3 countries are Northern African decent.

In [33]:

```python
# Total population of people in different Countries below the Poverty Line
Top_10_Below = merged_df.groupby('Country Name')['$1 - total number of people below pove
.sum().astype(int).sort_values(ascending=False).head(10).reset_index()
Top_10_Below
```

Out[33]:

	Country Name	$1 - total number of people below poverty line
0	Tanzania	1926253510
1	Mozambique	1868126836
2	Zambia	1759133812
3	Niger	1370009908
4	Ethiopia	1266623606
5	Ghana	1177098644
6	Malawi	946594176
7	South Africa	941903256
8	Burkina Faso	861419072
9	Mali	587256002

```python
# Visualise the above
sns.barplot(y='Country Name', x='$1 - total number of people below poverty line', data=T
plt.title('Top_10 Population of Countries Below the Poverty Line with 1USD daily income'
```

Top_10 Population of Countries Below the Poverty Line with 1USD daily income
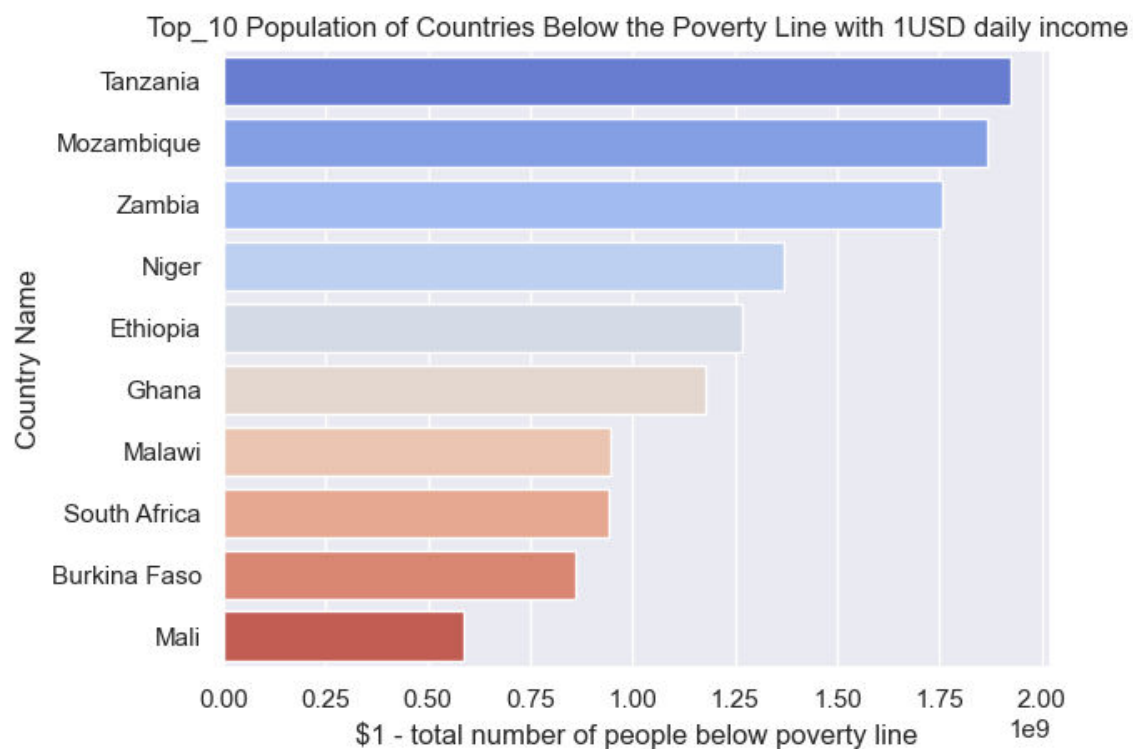
Insight: Top 10 Countries below Poverty line of 1usd per day income, the distributionof the above is between Western and Easter region of Africa. South Africa also is pressent here, but when this is compared with the number above poverty line, this becomes infinitesimal

UNIVARIATE ANALYSIS

In [35]:

```python
# We shall visualise the  Count of Incomegroups
plt.figure(figsize=(10,5))
plt.title('Count of Income Group Across Africa')
ax = sns.countplot(x='IncomeGroup', data=merged_df, palette='coolwarm',order=merged_df['
values = merged_df['IncomeGroup'].value_counts(ascending=False).values
ax.bar_label(container=ax.containers[0], labels=values);
```


Count of Income Group Across Africa
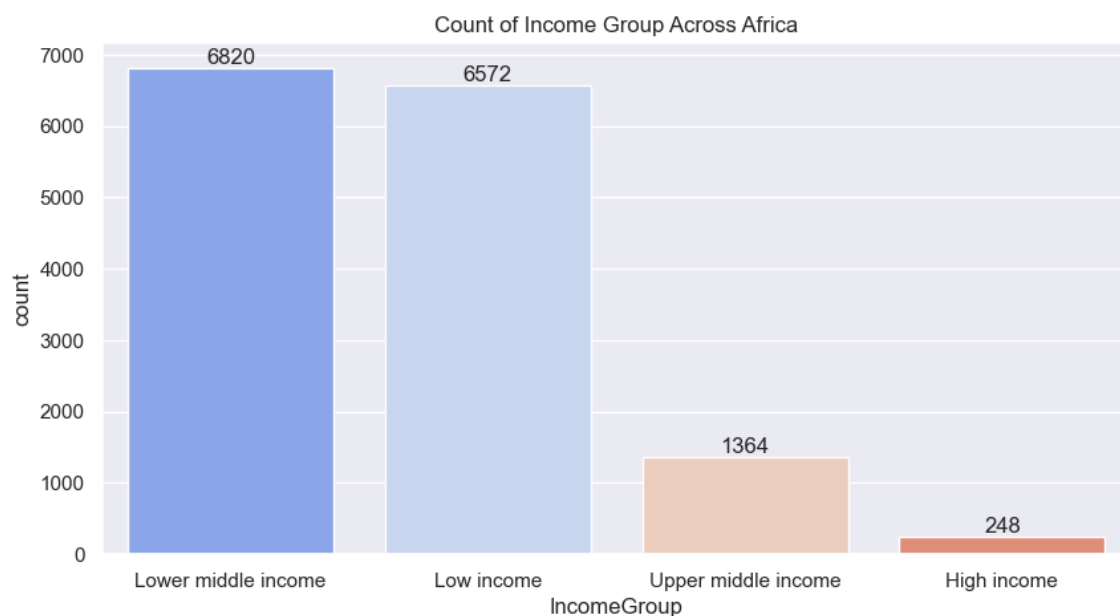
Insight: Majority of the African Population are of Low and Lower middle income as shown above, Very few are of High income in their earnings

BIVARIATE ANALYSIS

In [55]:

```python
# Let us Check the Regions of African and how they regions are performing by GDP
Region = merged_df.groupby('Region')['GDP_Current_USD'].mean().reset_index()
Region
```

Out[55]:

	Region	GDP_Current_USD
0	Eastern Africa	7.797256e+09
1	Middle Africa	9.921481e+09
2	Northern Africa	5.428433e+10
3	Southern Africa	4.508145e+10
4	Western Africa	1.867697e+10

In [56]:

```python
# Visualise the mean GDP per region
plt.figure(figsize=(8,5))
plt.title('Mean GDP by Region in Africa')
ax = sns.barplot(y='Region', x='GDP_Current_USD', data=Region, palette='coolwarm')
plt.show()
```
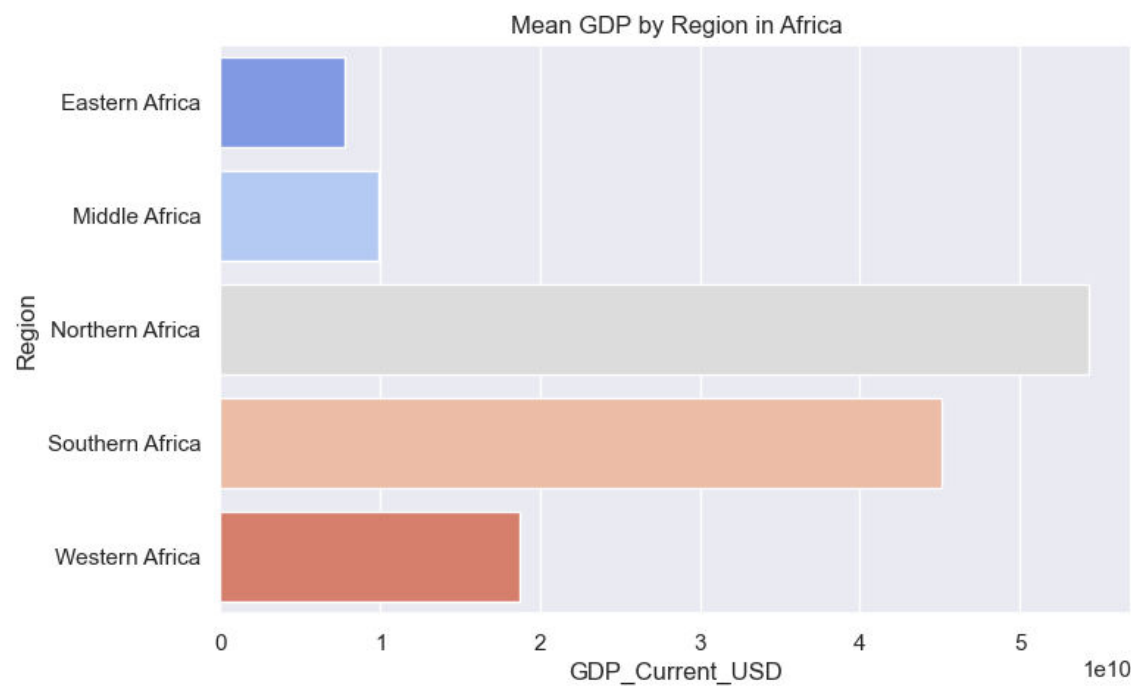
Insight: North African Region and Southern Africa are the top performing Regions by GDP. Eastern Africa and Middle Africa have the list GDP

In [38]:

```python
# But we shall check the GDP of Sourthern African Region to See are they truly doing wel
# or is South Africa as a nation so good that all others are behind the shadows.
dfs = merged_df.loc[merged_df['Region'] == 'Southern Africa']
dfs
```

Out[38]:

	African Country Code	Country Name	IncomeGroup	Region	Indicator Name	Year_x	GDP_Current_USD	Yea
17	BWA	Botswana	Upper middle income	Southern Africa	GDP (current US$)	1960	3.041231e+07	198
18	BWA	Botswana	Upper middle income	Southern Africa	GDP (current US$)	1960	3.041231e+07	199
19	BWA	Botswana	Upper middle income	Southern Africa	GDP (current US$)	1960	3.041231e+07	200
20	BWA	Botswana	Upper middle income	Southern Africa	GDP (current US$)	1960	3.041231e+07	200
21	BWA	Botswana	Upper middle income	Southern Africa	GDP (current US$)	1960	3.041231e+07	201
...	...	...	...	...	...	...	...	
14987	ZAF	South Africa	Upper middle income	Southern Africa	GDP (current US$)	2021	4.190150e+11	200
14988	ZAF	South Africa	Upper middle income	Southern Africa	GDP (current US$)	2021	4.190150e+11	200
14989	ZAF	South Africa	Upper middle income	Southern Africa	GDP (current US$)	2021	4.190150e+11	200
14990	ZAF	South Africa	Upper middle income	Southern Africa	GDP (current US$)	2021	4.190150e+11	201
14991	ZAF	South Africa	Upper middle income	Southern Africa	GDP (current US$)	2021	4.190150e+11	201

1426 rows × 18 columns

```python
xxy = dfs.groupby('Country Name')['GDP_Current_USD'].mean().sort_values().reset_index()
xxy
```

Out[39]:

	Country Name	GDP_Current_USD
0	Lesotho	9.197385e+08
1	Eswatini	1.660253e+09
2	Botswana	5.351144e+09
3	Namibia	6.438873e+09
4	South Africa	1.623404e+11

In [40]:

```python
# We shall visualise the above to see each contries contribution to the GDP of the South
plt.figure(figsize=(8,5))
plt.title('Mean GDP of Southern Africa Region')
ax = sns.barplot(y='Country Name', x='GDP_Current_USD', data=xxy, palette='coolwarm')
plt.show()
```
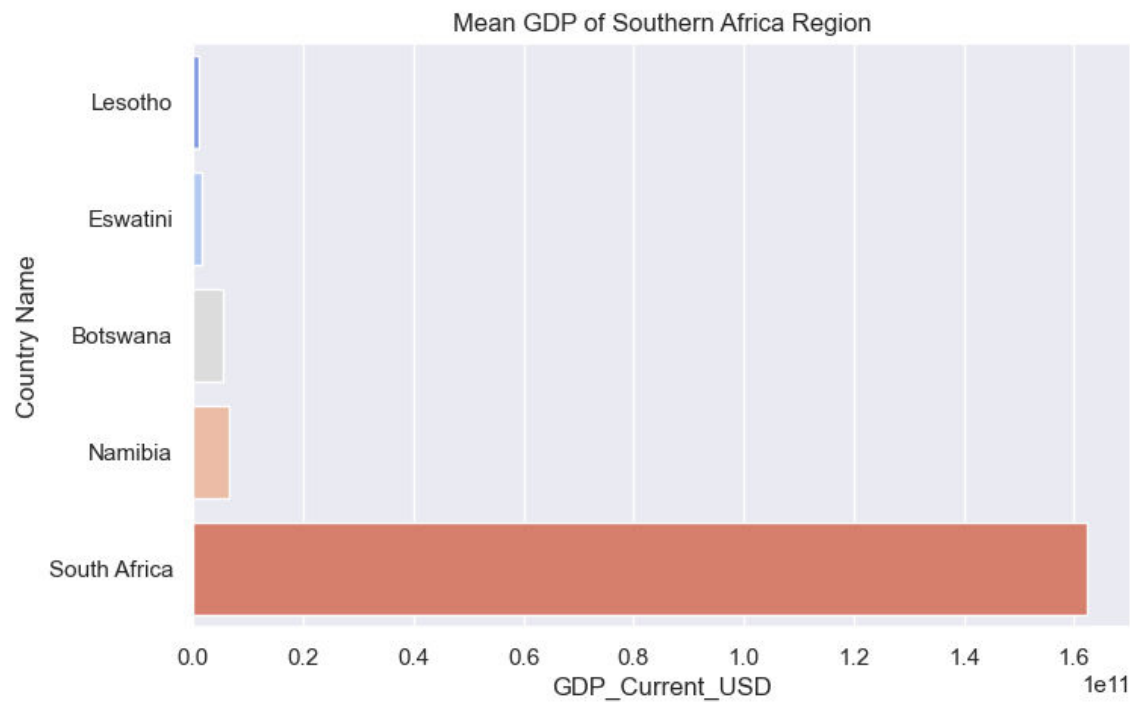
Insight: The contributions of the other countries in the southern regions are infinitesimal, South Africa is the main contributor for the entire region by GDP.

In [41]:

```
# Because Nigeria is Western Africa, with a high GDP we shall also see the contribution
dfn = merged_df.loc[merged_df['Region'] == 'Western Africa']
dfn
```

Out[41]:

	African Country Code	Country Name	IncomeGroup	Region	Indicator Name	Year_x	GDP_Current_USD	Year_
7	BEN	Benin	Lower middle income	Western Africa	GDP (current US$)	1960	2.261956e+08	2003
8	BEN	Benin	Lower middle income	Western Africa	GDP (current US$)	1960	2.261956e+08	2011
9	BEN	Benin	Lower middle income	Western Africa	GDP (current US$)	1960	2.261956e+08	2015
10	BEN	Benin	Lower middle income	Western Africa	GDP (current US$)	1960	2.261956e+08	2018
11	BFA	Burkina Faso	Low income	Western Africa	GDP (current US$)	1960	3.304428e+08	1994
...	...	...	...	...	...	...	...	
14942	SLE	Sierra Leone	Low income	Western Africa	GDP (current US$)	2021	4.042238e+09	2018
14960	TGO	Togo	Low income	Western Africa	GDP (current US$)	2021	8.413201e+09	2006
14961	TGO	Togo	Low income	Western Africa	GDP (current US$)	2021	8.413201e+09	2011
14962	TGO	Togo	Low income	Western Africa	GDP (current US$)	2021	8.413201e+09	2015
14963	TGO	Togo	Low income	Western Africa	GDP (current US$)	2021	8.413201e+09	2018

5518 rows × 18 columns

```python
# Group the west african countries against the mean GDP
yyz = dfn.groupby('Country Name')['GDP_Current_USD'].mean().sort_values().reset_index()
yyz
```

Out[42]:

|	Country Name	GDP_Current_USD
0	Guinea-Bissau	7.437739e+08
1	Gambia	8.355803e+08
2	Sierra Leone	1.470076e+09
3	Togo	1.971759e+09
4	Liberia	2.114177e+09
5	Cabo Verde	2.410252e+09
6	Mauritania	2.625142e+09
7	Niger	3.896524e+09
8	Guinea	4.216966e+09
9	Benin	4.250500e+09
10	Burkina Faso	4.696434e+09
11	Mali	4.952458e+09
12	Senegal	8.157668e+09
13	Ghana	1.582894e+10
14	Côte d'Ivoire	1.790465e+10
15	Nigeria	1.413746e+11

```
# Visualise using the Bar Chart
plt.figure(figsize=(8,5))
plt.title('Mean GDP of Western Africa Region')
ax = sns.barplot(y='Country Name', x='GDP_Current_USD', data=yyz, palette='coolwarm')
plt.show()
```


Mean GDP of Western Africa Region
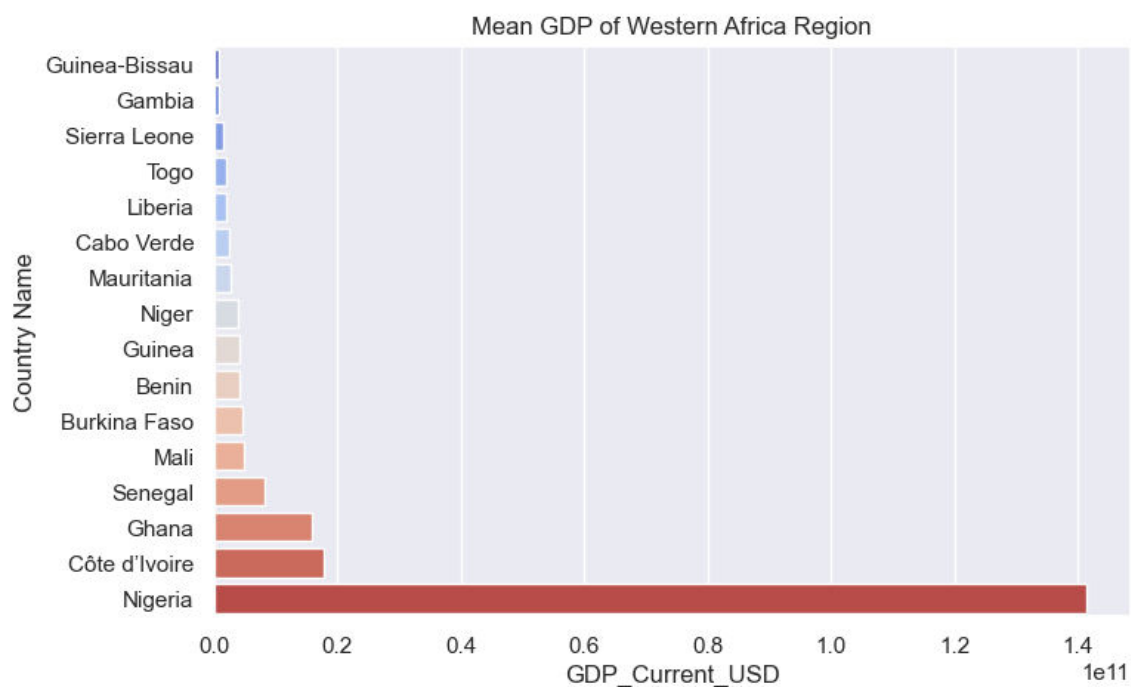
Insight: Nigeria is contributing most of the GDP in the region

In [44]:

```python
# But we shall check the GDP of Sourthern African Region to See are they truly doing wel
# or is South Africa as a nation so good that all others are behind the shadows.
dfN = merged_df.loc[merged_df['Region'] == 'Northern Africa']
dfN
```

Out[44]:

	African Country Code	Country Name	IncomeGroup	Region	Indicator Name	Year_x	GDP_Current_USD	Year_
52	DZA	Algeria	Lower middle income	Northern Africa	GDP (current US$)	1960	2.723593e+09	1988
53	DZA	Algeria	Lower middle income	Northern Africa	GDP (current US$)	1960	2.723593e+09	1995
54	DZA	Algeria	Lower middle income	Northern Africa	GDP (current US$)	1960	2.723593e+09	2011
55	EGY	Egypt	Lower middle income	Northern Africa	GDP (current US$)	1960	5.551653e+09	1990
56	EGY	Egypt	Lower middle income	Northern Africa	GDP (current US$)	1960	5.551653e+09	1995
...	...	...	...	...	...	...	...	
14966	TUN	Tunisia	Lower middle income	Northern Africa	GDP (current US$)	2021	4.668674e+10	1995
14967	TUN	Tunisia	Lower middle income	Northern Africa	GDP (current US$)	2021	4.668674e+10	2000
14968	TUN	Tunisia	Lower middle income	Northern Africa	GDP (current US$)	2021	4.668674e+10	2005
14969	TUN	Tunisia	Lower middle income	Northern Africa	GDP (current US$)	2021	4.668674e+10	2010
14970	TUN	Tunisia	Lower middle income	Northern Africa	GDP (current US$)	2021	4.668674e+10	2015

1736 rows × 18 columns

```python
abc = dfN.groupby('Country Name')['GDP_Current_USD'].mean().sort_values().reset_index()
abc
```
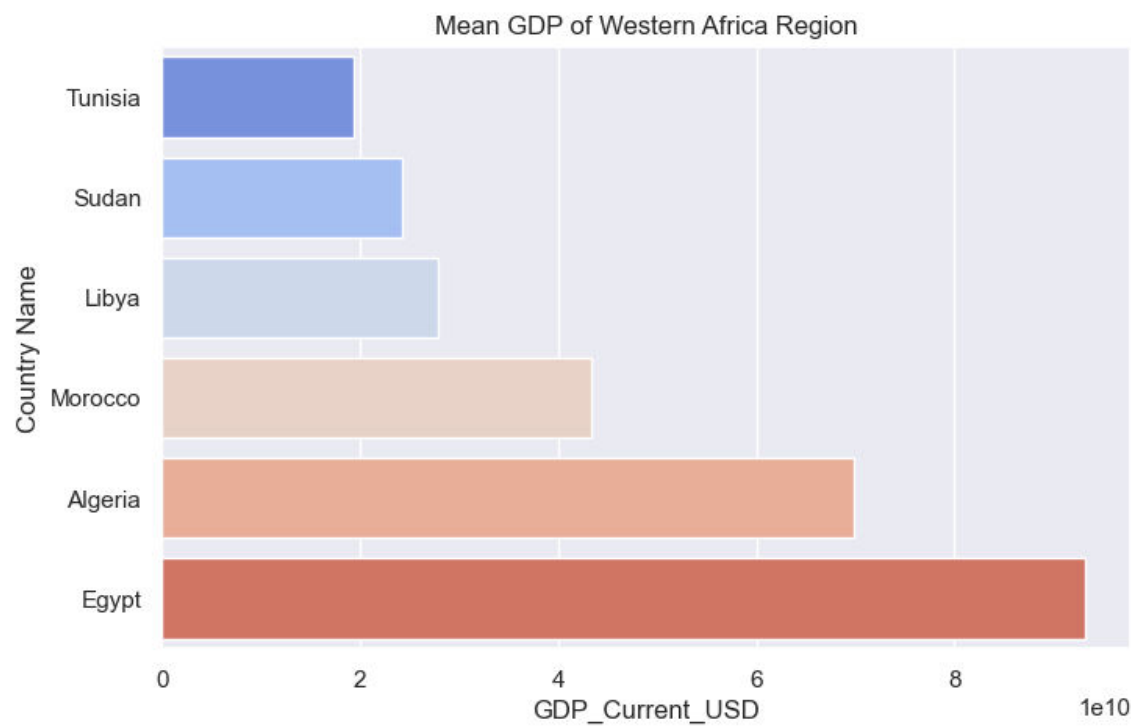
	Country Name	GDP_Current_USD
0	Tunisia	1.937880e+10
1	Sudan	2.427544e+10
2	Libya	2.787348e+10
3	Morocco	4.332241e+10
4	Algeria	6.983843e+10
5	Egypt	9.315951e+10

```python
plt.figure(figsize=(8,5))
plt.title('Mean GDP of Western Africa Region')
ax = sns.barplot(y='Country Name', x='GDP_Current_USD', data=abc, palette='coolwarm')
plt.show()
```


Mean GDP of Western Africa Region

Insight: Egypt is the leader in the Northern Region, but also Algeria, Morocco and the rest of the Northern region are also contributing a gooed measure

◀ ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬ ▶

MULTIVARIATE ANALYSIS

In [47]:

```
# We shall try to view the Relationship between the years and the GDP and see the mean p
# We shall groupby function

Relationship = round(merged_df.groupby(['Year_x','Region'])['GDP_Current_USD'].mean().re
Relationship
```

Out[47]:

|	Year_x	Region	GDP_Current_USD
0	1960	Eastern Africa	1.641097e+09
1	1960	Middle Africa	2.308683e+09
2	1960	Northern Africa	4.229601e+09
3	1960	Southern Africa	3.447754e+09
4	1960	Western Africa	1.831839e+09
...	...	...	...
305	2021	Eastern Africa	3.145379e+10
306	2021	Middle Africa	2.798438e+10
307	2021	Northern Africa	1.936390e+11
308	2021	Southern Africa	1.165376e+11
309	2021	Western Africa	6.337706e+10

310 rows × 3 columns

```python
#We Shall visualise the above relationship with a Lineplot

plt.figure(figsize=(10,8))
sns.lineplot(x='Year_x', y='GDP_Current_USD', data=Relationship, hue='Region')
plt.title('Relationship between Year, Mean GDP & Region ')
plt.show()
```
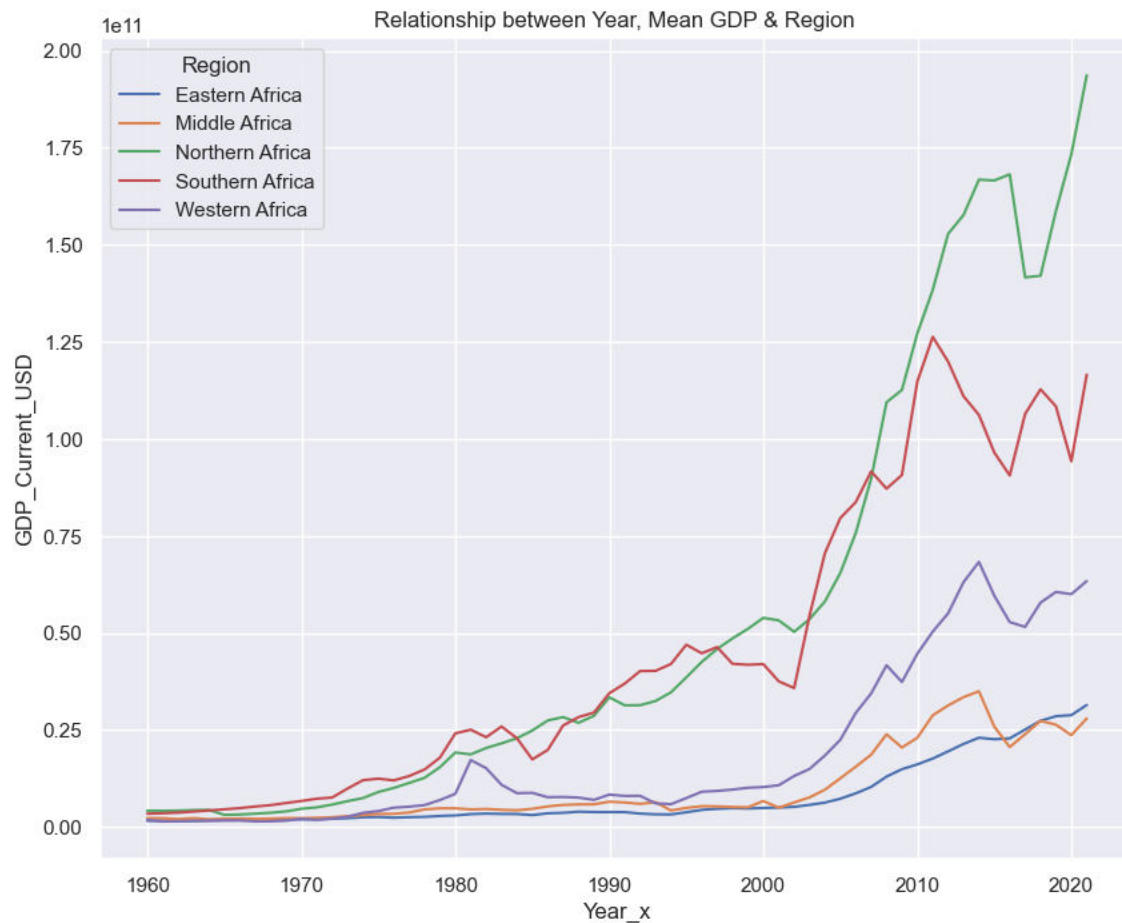
Insight: There has been an upward trend from 1970s for Southern and Northern Africa, which preceeded a sharp rise in the years 2000s. Observe a decline in Southern Africa's GDP during 1995 - 2002 (approx) which also concided during the period of the HIV/Aid Epidemic in those years. Recall, the life expectancy during those years also reduced.

Other regions have experienced crawling growth, however after 2002, they have experience an upward trend

In [49]:

```python
# We shall group the Region andincome  and compare with the mean GDP
Region1 = merged_df.groupby(['Region', 'IncomeGroup'])['GDP_Current_USD'].mean().sort_va
Region1
```

Out[49]:

	Region	IncomeGroup	GDP_Current_USD
0	Eastern Africa	High income	5.188390e+08
1	Southern Africa	Lower middle income	1.289996e+09
2	Western Africa	Low income	3.068504e+09
3	Middle Africa	Upper middle income	5.826812e+09
4	Eastern Africa	Upper middle income	6.072144e+09
5	Middle Africa	Low income	6.080313e+09
6	Eastern Africa	Low income	6.880254e+09
7	Eastern Africa	Lower middle income	1.220816e+10
8	Middle Africa	Lower middle income	1.318583e+10
9	Northern Africa	Low income	2.427544e+10
10	Northern Africa	Upper middle income	2.787348e+10
11	Western Africa	Lower middle income	3.326750e+10
12	Northern Africa	Lower middle income	5.774148e+10
13	Southern Africa	Upper middle income	6.843689e+10

```python
# Visualising the Data Above
plt.figure(figsize=(9,5))
plt.title('Mean GDP by Region and IncomeGroup in Africa')
ax = sns.barplot(x='Region', y='GDP_Current_USD', data=Region1, palette='coolwarm', hue=
plt.show()
```
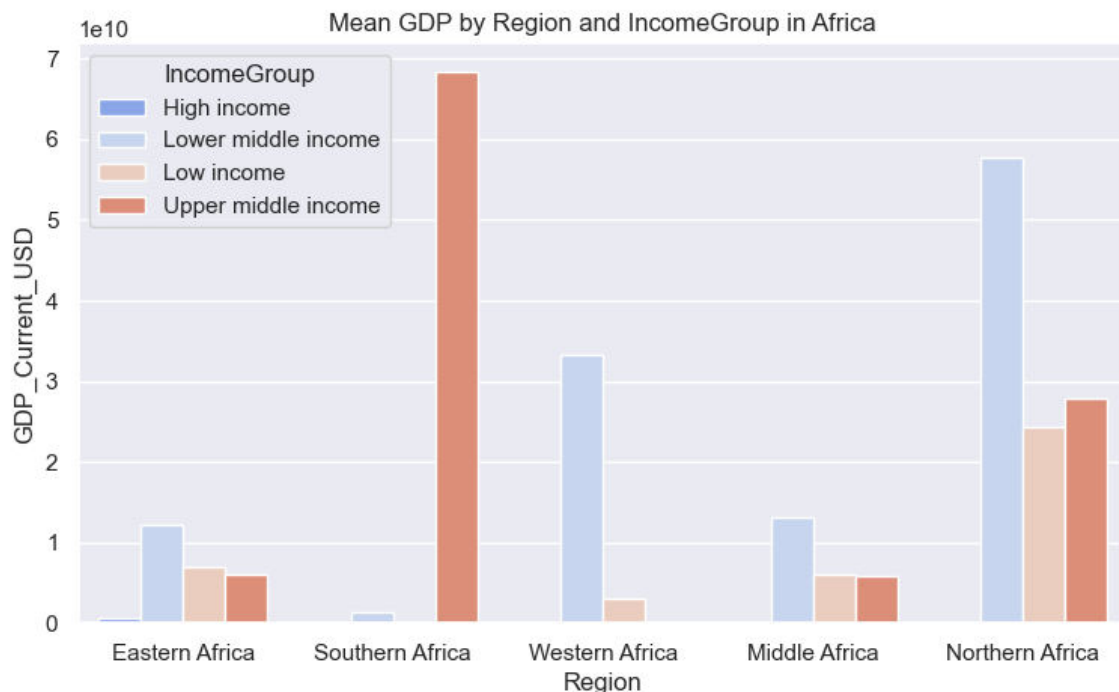

Insight: Southern African have the highest GDP but no distribution in income groups but only the upper middle income.

Northern Africa have a good distribution of incomegroups in the GDP

Western African where Nigeria is the most contributor have only low and lower middle income class.

INSIGHTS on GDP and Income

- Poverty, life expectancy, and Gross Domestic Product (GDP) are three critical indicators that reflect a country's economic and social conditions.

Based on analysis, the below insights were derived:

- 1. GDP of Africa have increased over the years with the least recorded in Equatorial Guinea $9,122,751 (nine million, one hundred twenty-two thousand, seven hundred fifty-one - U$$ 574,183,825,592 (five hundred seventy-four billion, one hundred eighty-three million, eight hundred twenty-five thousand, five hundred ninety-two) . By research, this would appear to be an all-time high figure achieved in 2014 (Figures also online)
- 2. We clearly saw from analytics, that GDP is distributed across, but certain regions recorded the cumulative highest in GDP such as in Northern African and Southern Africa, but breaking the

southern region, we saw South African was just outstanding. in comparison to the rest of the countries in that region. Also, the income group is higher in this regions

- 3. The count of income distribution across Africa, Low-income and lower middle-income earner are majority. The upper income is limited followed by a very scanty High-Income earners. The only country where this exist is SEYCHELLES.
- 4. Income is more distributed in North Africa, Middle Africa and East Africa, but Western and Southern African where we recorded countries with very high GDP (Nigeria and South Africa), income is not distributed. S.A had mostly upper middle income while Nigeria have mostly lower middle income. If we are to exclude these 2 countries from the analysis, then Northern African will be most prominent across the entire Africa.

INSIGHTS on Life Expectancy

- 1. The average life expectancy in Africa is between 45 – 55 (approx.) as we used the histogram to visualise
- 2. According to the Data, Western Africa have most population with the least life expectancy, surprisingly having Nigeria with the least recorded figure even with high GDP as seen above.
- 3. Northern Africa countries recorded the majority in higher life expectancy which was also seen in their cumulative GDP as well as their income distribution among the population.
- 4. Life expectancy have increased over the years from 25yrs to about 76yrs as the maximum recorded figure Cape Verde.
- 5. We also saw that life expectancy increases as the income group increases which is a good insight into our data.
- Also, worthy to note that Life expectancy in general was at its lowest in 1920's and 1940's for both the Low and Lower middle-income earners.
- 1960s saw an increase of life expectancy from all regions
- Further, western Africa is still the least with income distribution and least in life expectancy
- Outlier: Nigeria has the highest GDP, but this was not relatable to their incomegroup of the population as well as the life expectancy, perhaps this needs a deeper study how Nigeria, with a high GDP, but recorded the least Life expectancy, and mostly low-income group in the population (btw poverty) with a Daily income usd30 - usd2.15

CONCLUSION

- GDP have a direct relationship with income of the population (except Nigeria) as well as their life expectancy and from the data we have seen that those regions where we have higher GDP, also recorded high income as well higher life expectancy.
- All countries in Africa basically have a lot to do in terms of increasing their GDP, through Government policies, local contents, and investments as this has shown to affect the overall wellbeing of people/population

RECOMMENDATION

Many regions in African needs help and support to improve their income earnings. Perhaps the good way to start will be from -Countries with low income such as Western Africa can learn what Northern Africa is doing in terms of wealth distribution across the populace. While the government of Africa

In [ ]: