PROJECT REPORT:

Project Purpose:

Global inequality is as a result of disparities in the distribution of resources among different regions, countries, and citizens in the world. The purpose of this project is to undertake a comparative analysis of Incomes, happiness score, inequality, and Satisfaction.

Dataset Description

Sourced from OpenML, the dataset I used for this project HappySore.csv has n = 111 rows and 11 columns bearing features. These features are numeric analysis like, GDP, Income per Capita, Satisfaction, and happiness score as well as qualitative descriptions like country and region. The following figure (a python snippet) shows the various data parameters:

Reading and Viewing the Data

```
In [3]: income_data = pd.read_csv('happyscore_income.csv')
        income_data
```

Out[3]:

country	adjusted_satisfaction	avg_satisfaction	std_satisfaction	avg_income	median_income	income_inequality	region	happyScore	GDP
Armenia	37	4.9	2.42	2096.76	1731.506667	31.445556	'Central and Eastern Europe'	4.350	0.76821
Angola	26	4.3	3.19	1448.88	1044.240000	42.720000	'Sub-Saharan Africa'	4.033	0.75778
Argentina	60	7.1	1.91	7101.12	5109.400000	45.475556	'Latin America and Caribbean'	6.574	1.05351

Analysis Method Description

In this project, I used K_means clustering as the machine learning algorithm for data clustering so as to come with interesting analytics and visualizations.

Summary of Results (Findings):

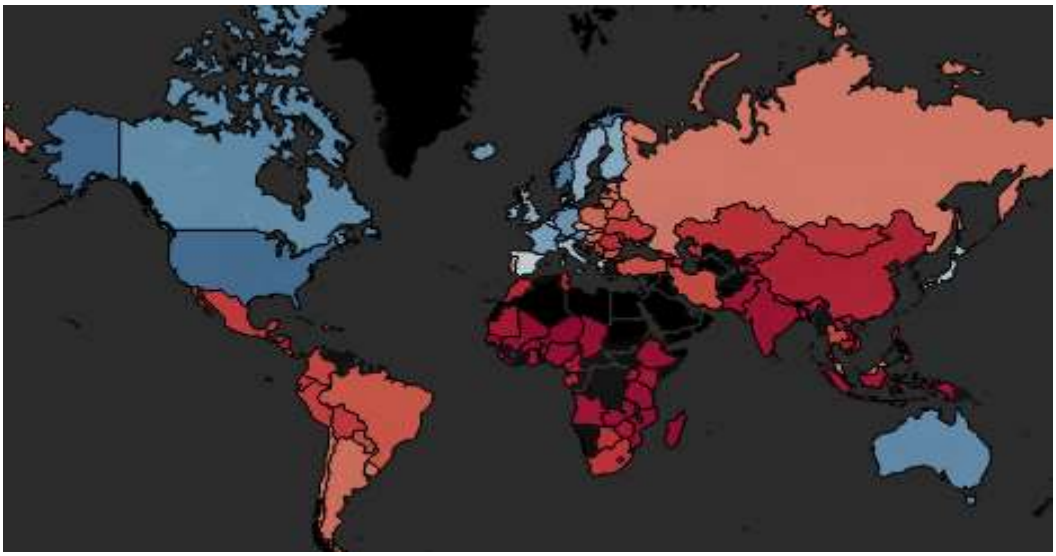
Using the Numpy Python library, I was able to deduce the following from the dataset:

The mean, variance, and standard deviation of the GDP among the countries is:

	GDP	Avg Income	Satisfaction	Happyscore	Inequality
Mean	0.8419090990990992	6442.751416180183	5.936936936936937	5.421909909909911	38.41780505180178
Std_dev	0.3858406511131772	6450.280568446985	1.3502704385730702	1.1754695941020974	8.33774345660215
Variance	0.1488730080514405 5	41606119.41165526	1.8232302572843115	1.3817287666585498	69.51796594811196
Max	1.56391	26182.275	8.5	7.587	63.72666667
Min	0.0153	572.88	2.5	2.839	24.215

Relationship between Average Income and GDP

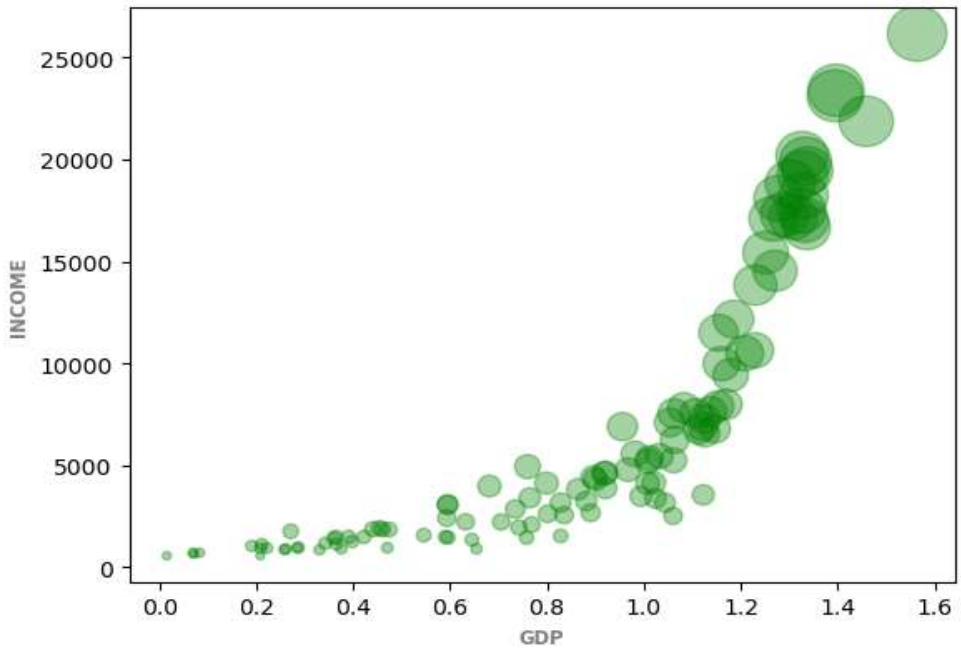
The following is a map generated from Tableau indicating the average income distribution of countries across the globe:



General Trend for the Entire Dataset (with Mean of GDP and Income)

The two variables GDP and Income per Capita are directly proportional. In general, countries with the high GDP had citizens enjoying high incomes. A quick glance in the GDP – Income scatterplot in python confirms this:

```
gdp_mean: 0.8419090990990994 income_mean: 6442.75141618018
Out[91]: Text(0, 0.5, 'INCOME')
```



NB: The sizes of the plots increases with increase in income. Larger plots are notably at the highest points of the scatterplots and also with highest GDP.

For further analysis, the data was categorized into:

Wealthy Category (Highest GDP and Income), Average and High Income Categories, Poor Category

Wealthy Category (With labels of Highest and the Lowest Income Earners in the Wealthy Countries Category)

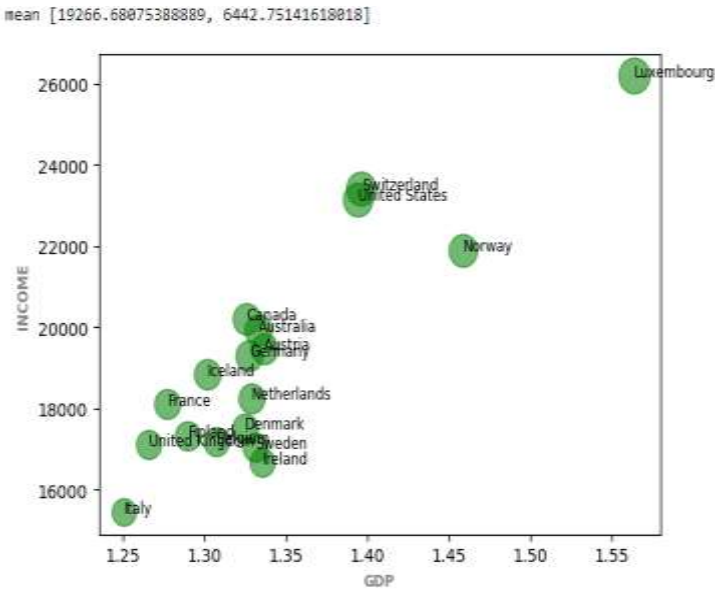
Filter operaton for all the countries with an average_income greater than $15,000

The country with the Highest average income among the Wealthy Countries is Luxembourg, also has the HIGHEST GDP per Capita (largest Greenest).

The country with the Lowest average income among the Wealthy Countries is Italy, also has the Lowest GDP per Capita.

With the mean of the Category Vs the Whole dataset mean.

With mean income of the wealthy category being $19266.70 against the general income mean of 6442.75



Poor Category: With labels of Highest and the Lowest Income Earners in the Poor Countries Category):

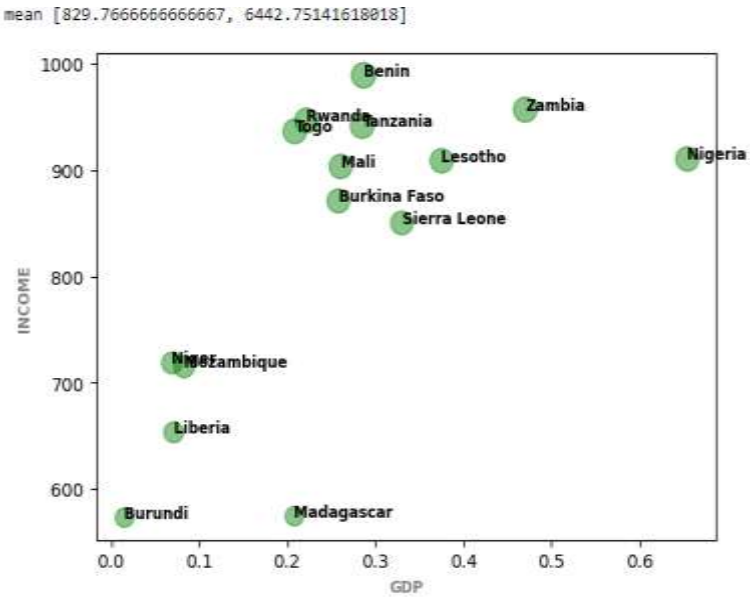
Filter operation for all the countries with an average income less than $1,000

The country with the Highest average income among the Poor Countries is Benin, But not with the HIGHEST GDP per Capita.

The country with the Lowest average income among the Poor Countries is Burundi Also has the Lowest GDP per Capita

Although Nigeria has the Highest GDP it doesn't have the Highest Average Incoem In The Category.

The mean income of the category is $824.8 against the mean income of the entire dataset $6442.80.



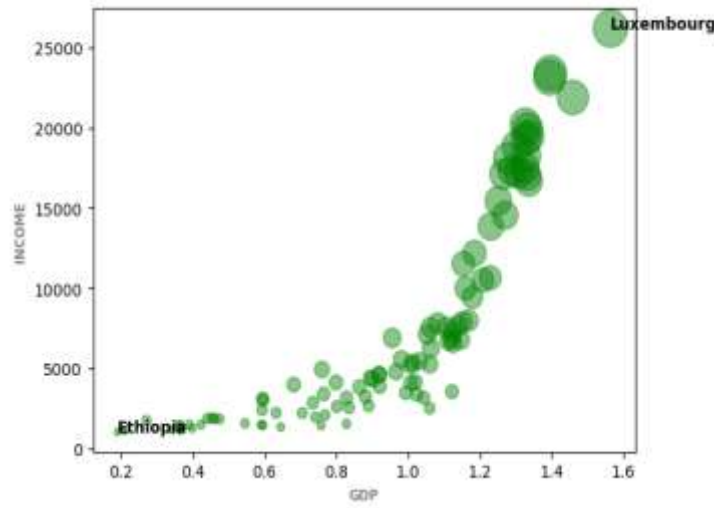
Average and High Income Category (With labels of Highest and the Lowest Income Earners in the Category of Countries)

Filter operation for all the countries with an average income greater than $10,000

The country with the Highest average income among this category is Luxembourg (Also has the HIGHEST GDP per Capita)

The country with the Lowest average income among this category is Ethiopia (Also has the Lowest GDP per Capita)

```
average_mean: 7319.788283291667 data_mean: 6442.75141618018
```



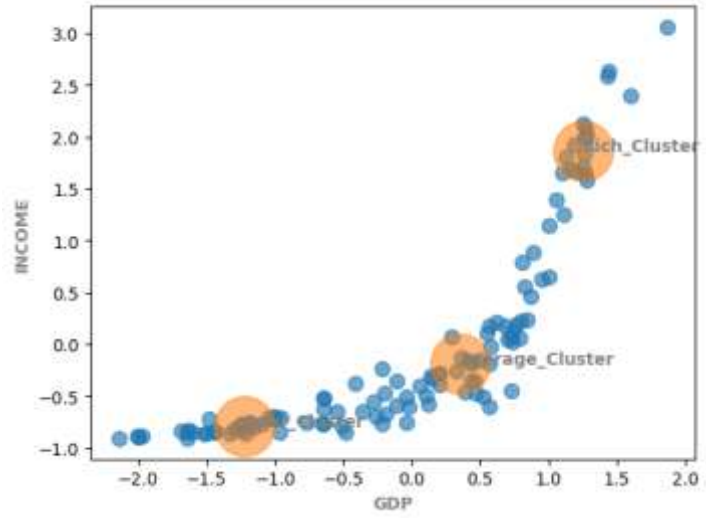
K-Means Clustering

After this quick analysis, I normalized the data in readinss for the K- means clustering. On running the algorithm I was able to come up with 3 clusters of coutries

These three Clusters of Countries are Wealthy , Average, and Poor

From running the K-Means clusteing It would be informed to say that countries with the highest Gross Domestic Product also have the highest Income per capita.

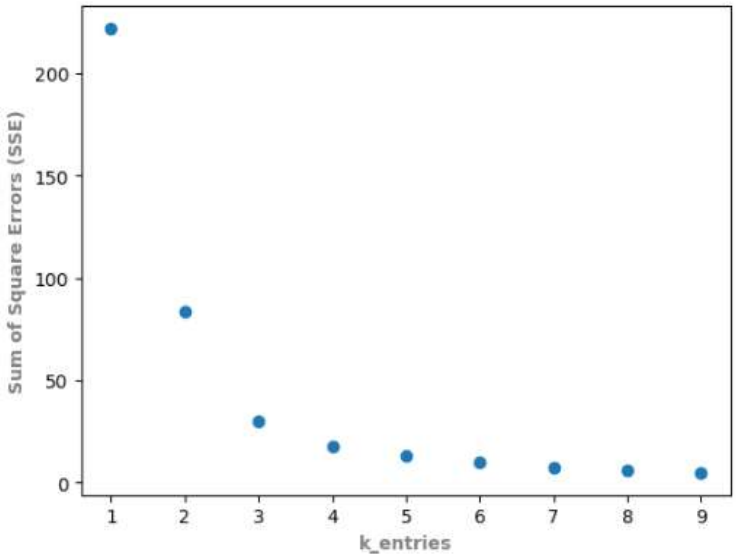
```
Text(-1.21992923, -0.7889247, 'Poor_Cluster')
```



Therefore if you are looking for countries with high paying jobs you can as well look for coutries with the highest GDP in the Rich_Cluster. Since the wealthiest cluster of countries also has the highest paying jobs.

Elbow Plot and Sum of Square Errors

```
[222.0,
 83.35918467257056,
 29.639788497265982,
 17.809988833222576,
 12.956659425019506,
 9.85231214371965,
 7.450517482822795,
 5.999526612679337,
 4.702662226171281]
```



Correlation in Inequality and Satisfaction

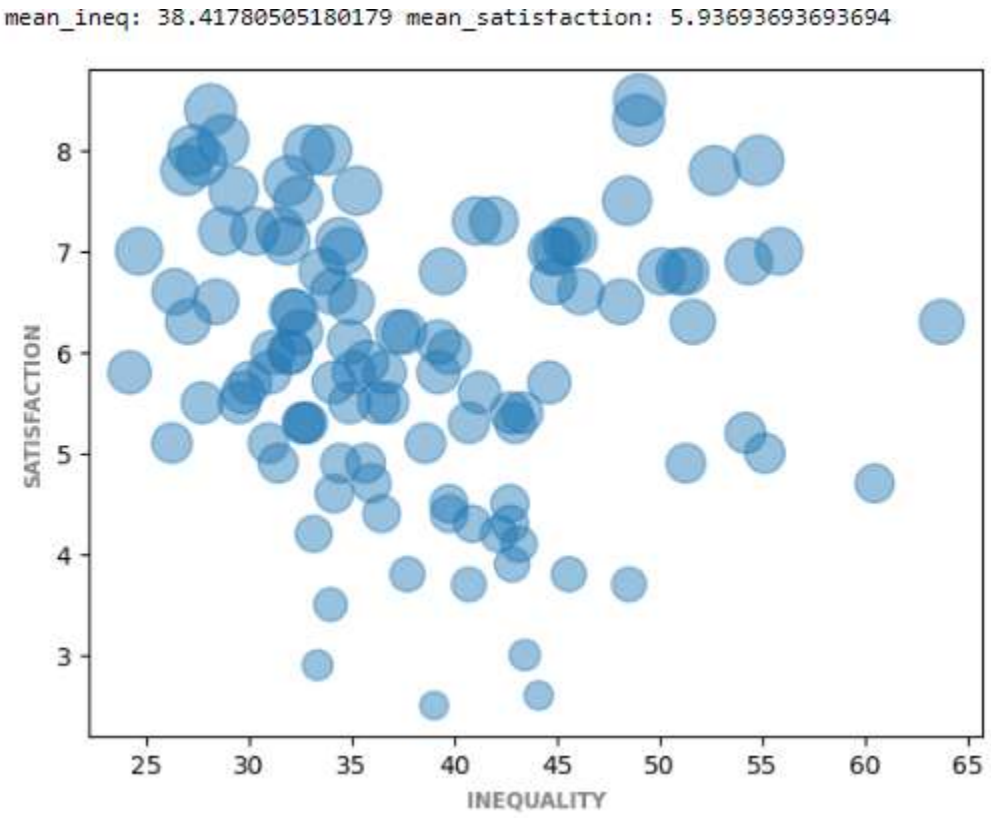
In this section, the analysis looks to answer the question on how Inequality Relates with Satisfaction.

General Correlation

In general, the data seems to agree with the notion that the relationship between variables inequality and satisfaction is an inverse one. This means that the least unequal countries happens to be the most happy and vice versa. The following scatterplot largely confirms this notion:

Please note that the sizes of the plots increases with increase ion satisfaction levels:

The mean inequality and satisfaction of the entire data frame is
Inequality mean: 38.41780505180179, satisfaction mean: 5.93693693693694



Although to a large extent the data agrees with the notion that the more unequal a country is the lower the satisfaction level in the country, there is a category of countries in the extreme upper right of the plot that is generally satisfied despite the high inequality.

This therefore challenges the notion that to be satisfied you need to be in a country with low inequality levels.

Satisfied (Lowest Inequality) Category

Map Analysis

Filter operation for all the countries with an average satisfaction index greater than 7

Here is a general outlook for the ***Inequality in Incomes*** in the satisfied category:

NB: The deeper the color the higher the Inequality against Income Ratio.

Python Scatter Plot

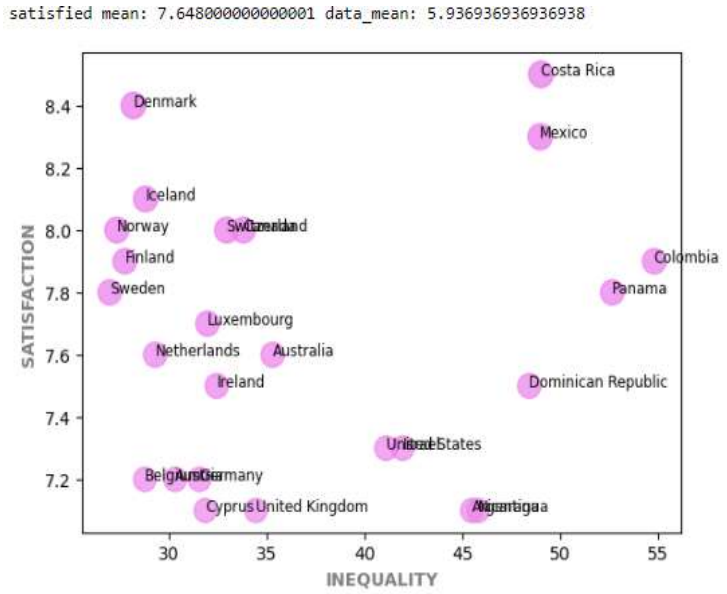
Some Countries have Low Inequity and High Satisfaction Levels (Eg: Denmark, Swiss, Iceland, Norway, Fin, etc.)

Some Countries have Relatively High Inequality with High Satisfaction Levels (Eg: Costa Rica and Mexico)

Some Countries have Low Inequality with Relatively Low Satisfaction (Eg: UK, Cyprus, Belgium, and Germany)

Some countries have Relatively High Inequality with Low Satisfaction Levels (Eg: Argentina, Antigua, US, and Dominican)

Most of the Countries in this Category shows a negative correlation Between Inequality and Satisfaction:



Moderately To Least Satisfied and Unsatisfied Category (With Labels)

Map Analysis

Filter operation for all the countries with an average satisfaction index less than or equal to 7

Within the moderate and low satisfaction category, this is how ***General Inequality in Income*** looks. These are the middle and high income countries of Asia and South America.

Among the low income within this satisfaction category, income inequality is as shown below:

Python Scatter Plot

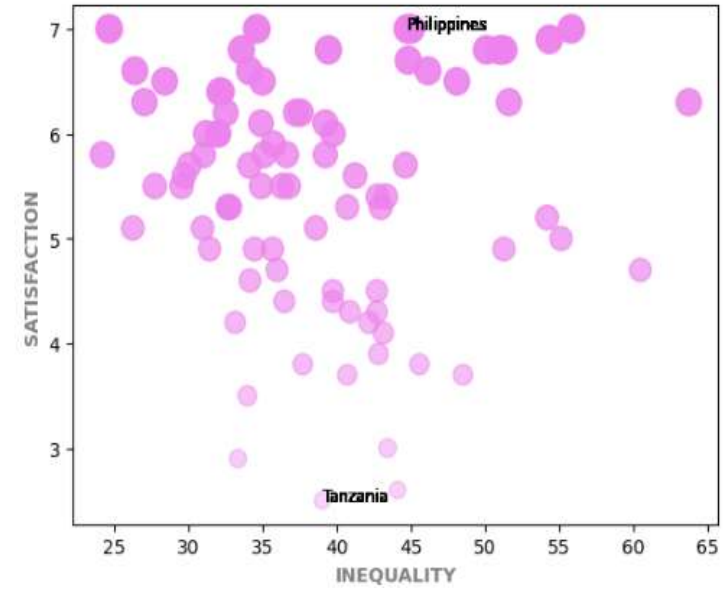
Tanzania Has the Lowest Satisfaction while Philippines has the Highest (Although Inequality is higher than Tanzania).

Most Countries with Low Inequality Have High Satisfaction

Some Countries in this Category have High Inequality and High Satisfaction

The following scatterplot illustrates this. The larger most colored plots indicates maximum satisfaction in the category.

```
moderate mean: 5.439534883720932 data mean: 5.936936936936938
```

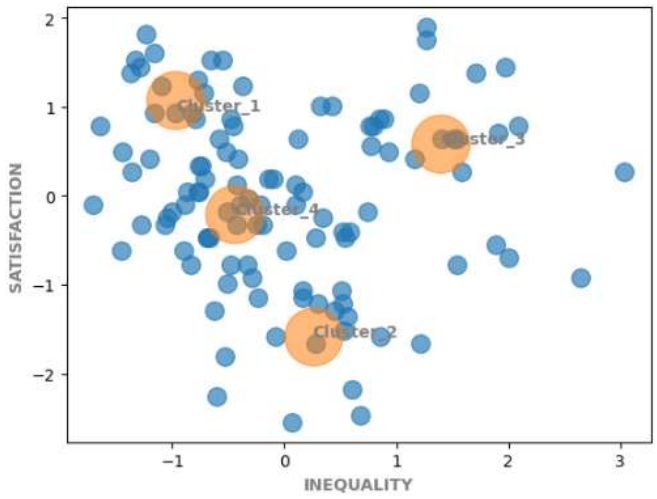


Clustering (Using K-Means)

After this quick analysis, I normalized the data in readinss for the K- means clustering. On running the algorithm I was able to come up with 3 clusters of coutries

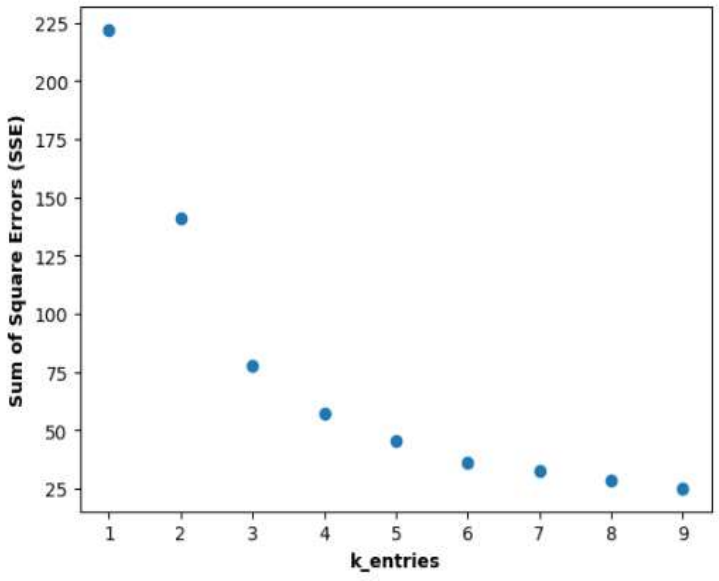
These Four Clusters of Countries are; Cluster_1: Low Inequality with High Satisfaction, Cluster_2: Relatively low Inequality with Very Low Satisfaction, Clutsre_3: High Inequality with High Satisfaction, and Cluster_4: Moderate Inequality with Moderate Satisfaction.

```
Text(-0.4439069, -0.20408749, 'Cluster_4')
```



Elbow Plot and Sum of Squared Errors

```
[222.0,
 141.34649716548395,
 77.99480126125671,
 57.130786453323694,
 45.252012851002675,
 36.186664122441215,
 32.526413411780275,
 28.65773642495513,
 25.042596408275273]
```



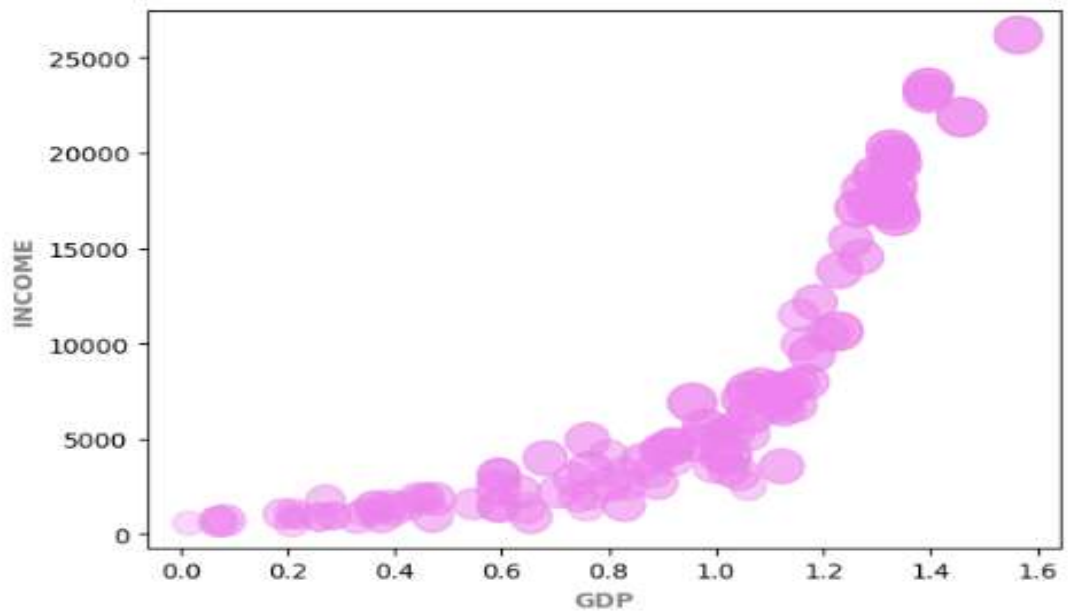
Checking Happiness Levels in GDP vs Income

In this section I sought to establish the relationship between three parameters namely Happiness, GDP and Average Income per capita.

Qn: How Happy are Countries with High (or low) Income and GDP?

To answer this question I created a scatterplot of GDP against income per Capita. To include the happiness parameter in the scatter plot I created a plot whose points become larger with increased happiness.

```
gdp_mean: 0.8419090990990995 income_mean: 6442.751416180181 happiness_mean: 5.421909909909911
Text(0, 0.5, 'INCOME')
```

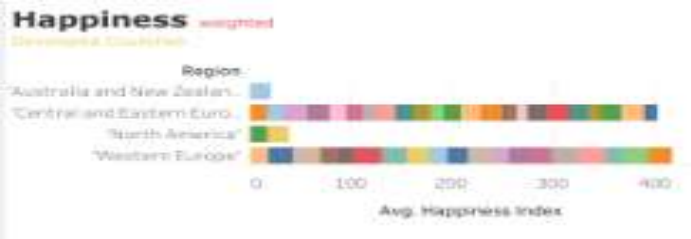


As we had seen the general trend earlier, high gdp countries are also high income countries, the tren is also generally tru from happiness levels. Since most of the largest and most colored plots are highest in the scatterplot, it will be informed to conclude that the most developed and wealthy countries are also the happiest. The reverse is also true.

Happiness in the Average and High Income Category

Regional Happiness (Developed Region)

Filter operation for all the countries with an average income greater than $1400



Python Scatterplot

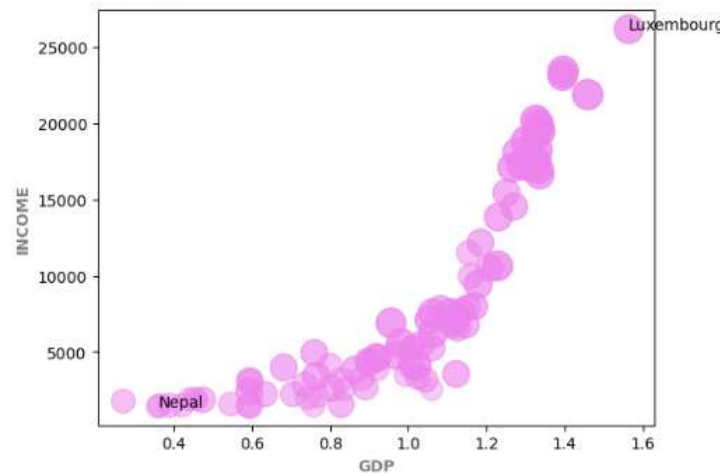
Luxembourg with the Highest GDP and Income per Capita is the LARGEST and DARKEST Plot (Hence the HAPPIEST.

Nepal with a Low GDP and Income per Capita is the SMALLEST and LEAST DARK Plot (Hence the LEAST HAPPY).

Most High-Income and High-GDP Countries have LARGER & DARKER Plots, and Vice-Versa

Hence in this Category Happiness Correlates Positively with GDP and Income.

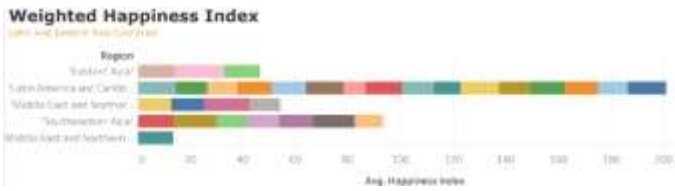
```
gdp_mean: 0.9722402222222223 income_mean: 7728.730657733334 happiness_mean: 5.73085
Text(1.56391, 26182.275, 'Luxembourg')
```



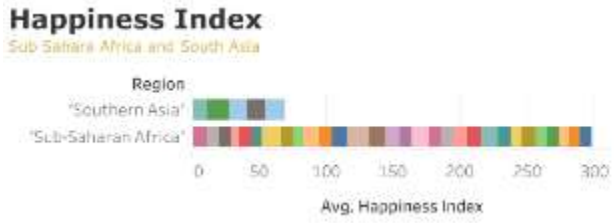
Happiness in the Average and Low Income Category

Regional Happiness (High Income Emerging Region)

Filter operation for all the countries with an average income greater than $1,400



Low Income Emerging Region



Python Scatter Plot

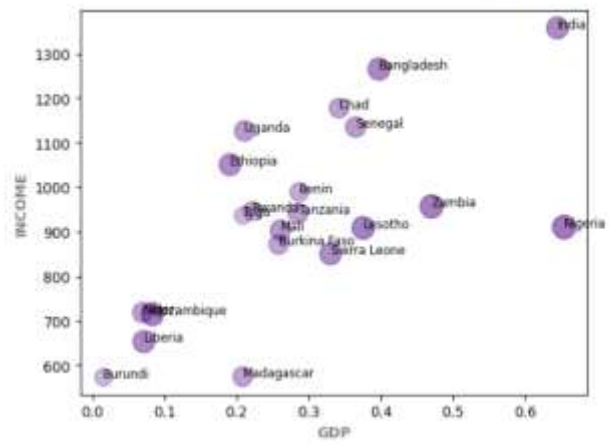
India with the Highest GDP and Income per Capita is the LARGEST and Darkest Plot (Hence the Happiest)

Burundi with the Lowest GDP and Income per Capita is the SMALLET and MOST PALE Plot (Hence the Least Happiest)

Most High-Income and High-GDP Countries in this category have LARGEST and DARKEST Plots, and Vice-Versa.

Hence in this Category Happiness Correlates Positively with GDP and Income. Compare India and Burundi in the scatterplot:

```
income_mean: 931.4118095238096 gdp_mean: 0.28334714285714285 happiness_mean: 0.283347
```



Checking Happiness Levels in Inequality vs Satisfaction

Qn: How Happy are Countries with High (or low) Inequality and Satisfaction Index?

Note that the Sizes and Colors of plots increses with increase in Happiness

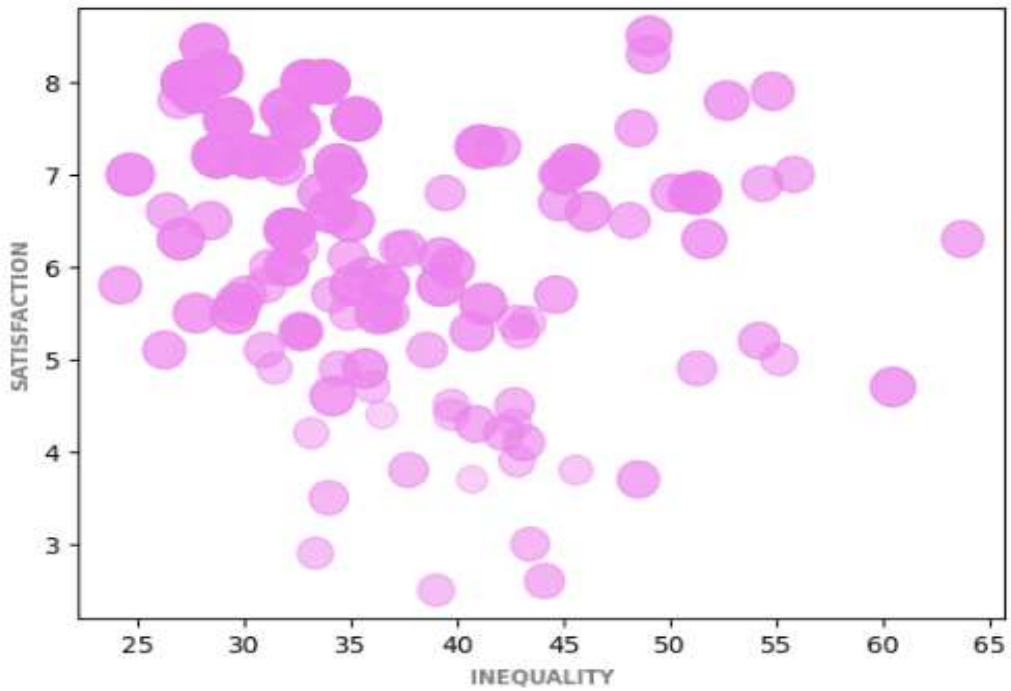
The Least Unequal Countries are the Most Satisfied and Have the LARGEST and DEEPEST COLORED plots (Hence HAPPIEST)

The Least Satisfied Countries have Moderate Inequality and Have the SMALLEST and LIGHTEST COLORED Plots (Hence LEAST HAPPIEST)

Some Countries Have High Satisfaction despite Being Highly Unequal. They are also Quite Happy

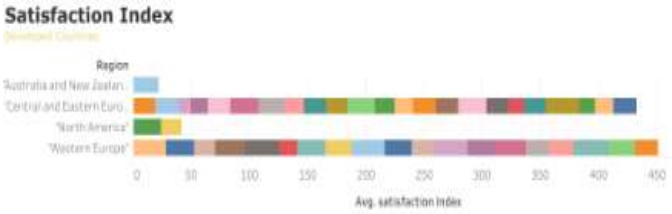
```
Mean_inequality: 38.41780505180179 Mean_satisfaction: 5.93693693693694 Mean_happiness: 5.42196
```

```
Text(0, 0.5, 'SATISFACTION')
```



Happiness in the Most Satisfied Category

Satisfaction in High Income Region



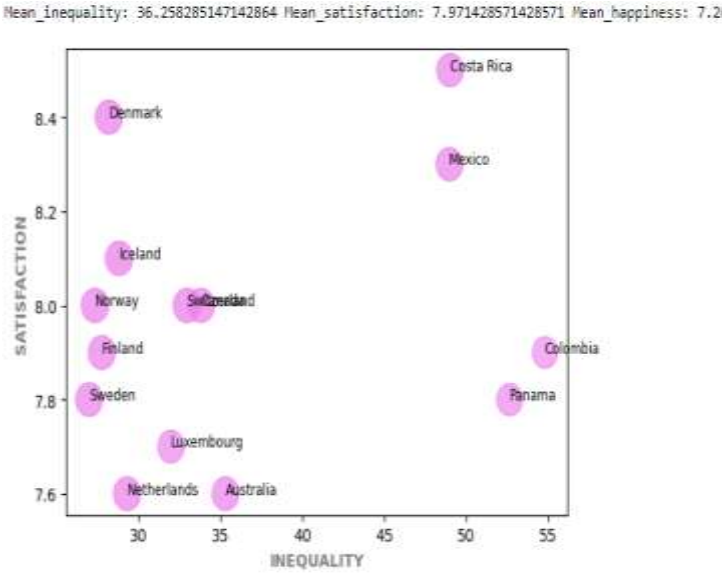
Filter operation for all the countries with an average satisfaction index greater than 7.5

The SIZES and COLORS of the Plots are Almost Equal; Hence Most Low-Inequality High-Satisfaction Courtiers are Quite Happy

Panama and Colombia have Highest Inequality, Least Satisfaction and are the Least Happy in the Category

Denmark Has the Least Inequality and Very High Satisfaction and Appears to be the Happiest in the Category

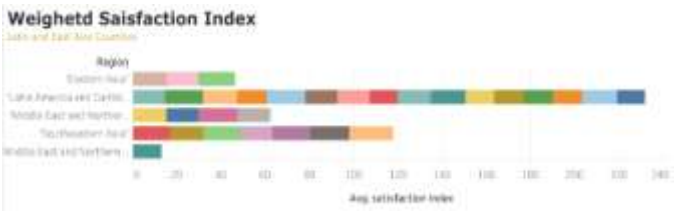
Costa Rica and Mexico Have Relatively High Satisfaction and Inequality; appears to be Very Happy as Well



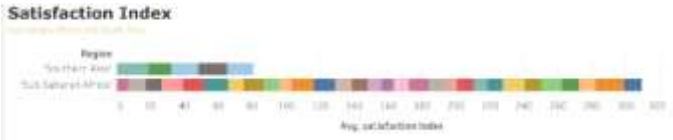
Happiness in the Moderate and Least Satisfied Countries

Satisfaction in:

Middle Income Region



Low Income Region



Filter operaton for all the countries with an average satisfaction index less than or equal to 7.5

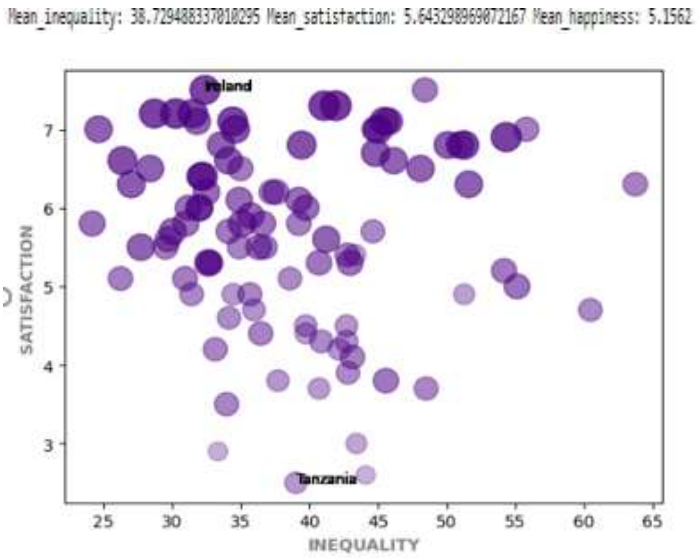
In this Category:

Most Countries with high Satisfaction and low Inequality are the HAPPIEST (Have the LARGEST and DEEPEST COLORED Plots)

Most Countries with low Satisfaction and Moderate Inequality are the LEAST HAPPY (Have the SMALLEST and LIGHTEST COLORED Plots)

Tanzania Has the Lowest Satisfaction with the SMALLEST and LIGHTEST COLORED Plot (Hence RELATIVELY UNHAPPY)

Ireland has the Highest Satisfaction with the LARGEST and DEEPEST COLORED Plots (Hence the HPPIEST)



Get Code on GitHub:

https://github.com/jobojiambo/Global-Satisfaction-Happiness-and Inequality/blob/main/Happyscore%20Portfolio%20Project.ipynb

Tableau Link:

https://public.tableau.com/authoring/GlobalAverageIncomeDistributin/Dashboard1#1