

Report on  
**World Happiness Exploratory Data Analysis**

For  
**ETP (VIVA) of Course Code : INT353**

**Submitted by: Nitesh**

**Reg. no: 12016394**

**Roll no: RK20MDB51**

**Section: K20MD**

On date: 25-11-2022

# INTRODUCTION

## **Why this Project ?**

My purpose behind choosing the World Happiness Report is to focus on the impact on the structure and quality of people's lives, and to evaluate how governments around the world are doing. As it can be used effectively to assess the progress of nations. I want to estimate the extent to which each of six factors – economic production, social support, life expectancy, freedom, absence of corruption, and generosity plays role in living conditions in particular country. Also we will get to know effect personal and national variations in happiness.

## **Project Domain**

The World Happiness Report is a landmark survey of the state of global happiness . The report continues to gain global recognition as governments, organizations and civil society increasingly use happiness indicators to inform their policy-making decisions. Leading experts across fields – economics, psychology, survey analysis, national statistics, health, public policy and more – describe how measurements of well-being can be used effectively to assess the progress of nations. The reports review the state of happiness in the world today and show how the new science of happiness explains personal and national variations in happiness.

## **About dataset**

The columns following the happiness score estimate the extent to which each of six factors – economic production, social support, life expectancy, freedom, absence of corruption, and generosity – contribute to making life evaluations higher in each country than they are in Dystopia, a hypothetical country that has values equal to the world's lowest national averages for each of the six

- **Ladder score:** Happiness score or subjective well-being. This is the national average response to the question of life evaluations.
- **Logged GDP per capita:** The GDP-per-capita time series from 2019 to 2020 using countryspecific forecasts of real GDP growth in 2020.
- **Social support:** Social support refers to assistance or support provided by members of social networks to an individual.
- **Healthy life expectancy:** Healthy life expectancy is the average life in good health - that is to say without irreversible limitation of activity in daily life or incapacities - of a fictitious generation subject to the conditions of mortality and morbidity prevailing that year.
- **Freedom to make life choices:** Freedom to make life choices is the national average of binary responses to the GWP question 'Are you satisfied or dissatisfied with your freedom to choose what you do with your life?' ... It is defined as the average of laughter and enjoyment for other waves where the happiness question was not asked
- **Generosity:** Generosity is the residual of regressing national average of response to the GWP: 'Have you donated money to a charity in the past month?' on GDP per capita
- **Perceptions of corruption:** The measure is the national average of the survey responses to two questions in the GWP: 'Is corruption widespread throughout the government or not' and 'Is corruption widespread within business or not?'
- **Ladder score in Dystopia:** It has values equal to the world's lowest national averages. Dystopia as a benchmark against which to compare contributions from each of the six factors. Dystopia is an imaginary country that has the world's least-happy people... Since life would be very unpleasant in a country with the world's lowest incomes, lowest life expectancy, lowest generosity, most corruption, least freedom, and least social support, it is referred to as 'Dystopia.' in contrast to Utopia.

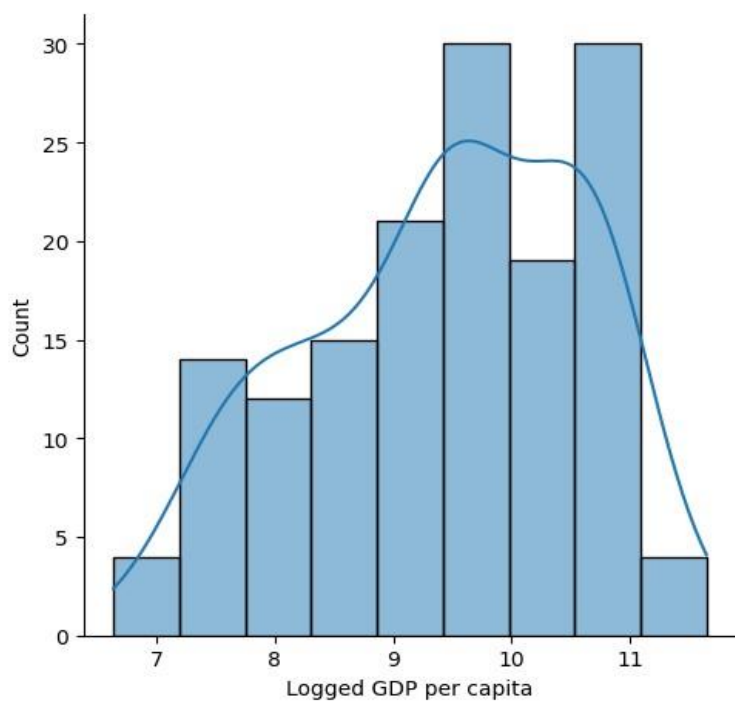
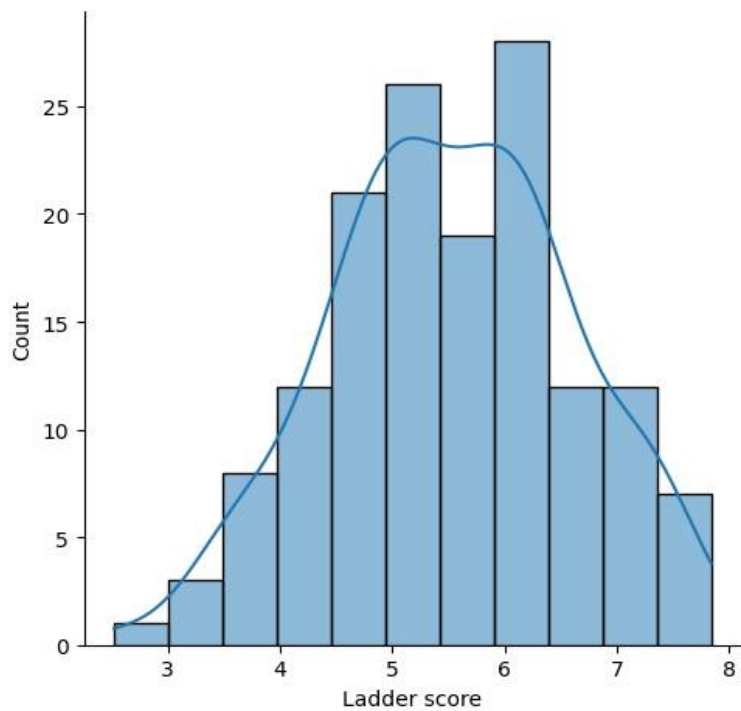
## APPROACH

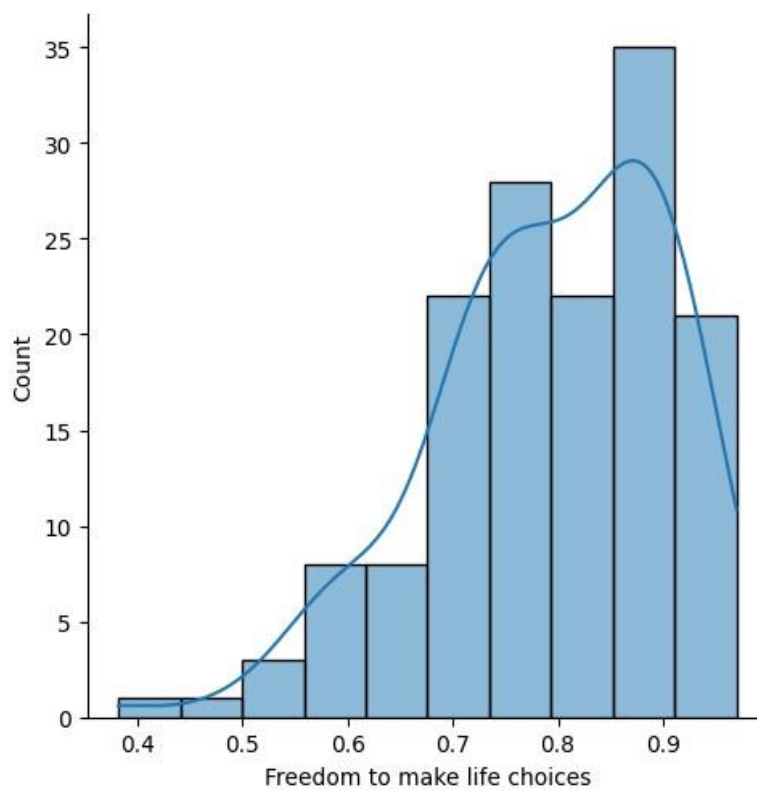
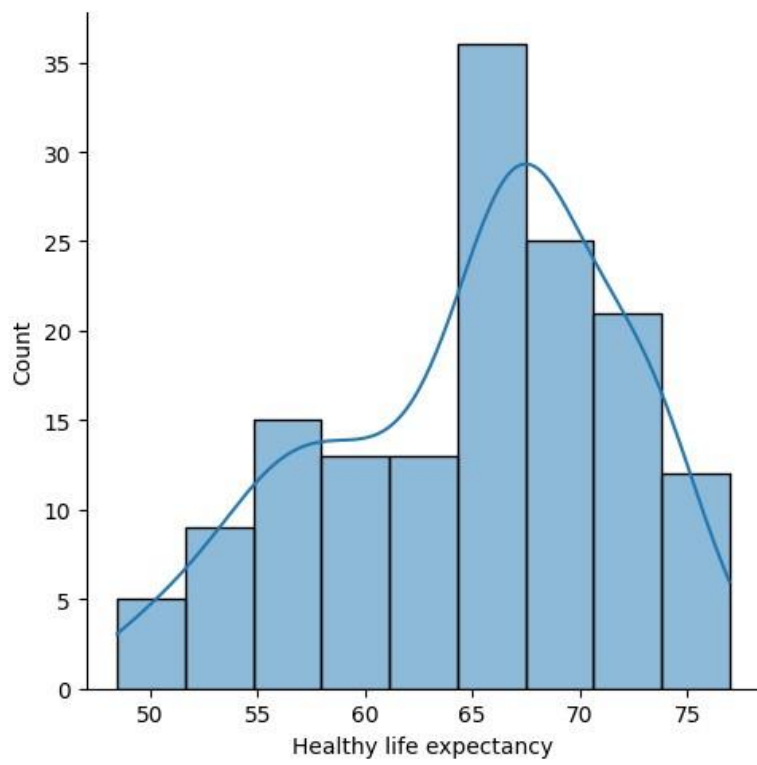
- Firstly I clean the data by handling missing values, unwanted columns and rows and by handling outliers.
- data preparation I will start to understand the relationships in dataset.
- Then I will do univariate and multivariate analysis and EDA to get the data visualization.
- After all these things I will write conclusion that what type of factors are actually responsible for happiness index of particular country and why some countries usually rank higher than others

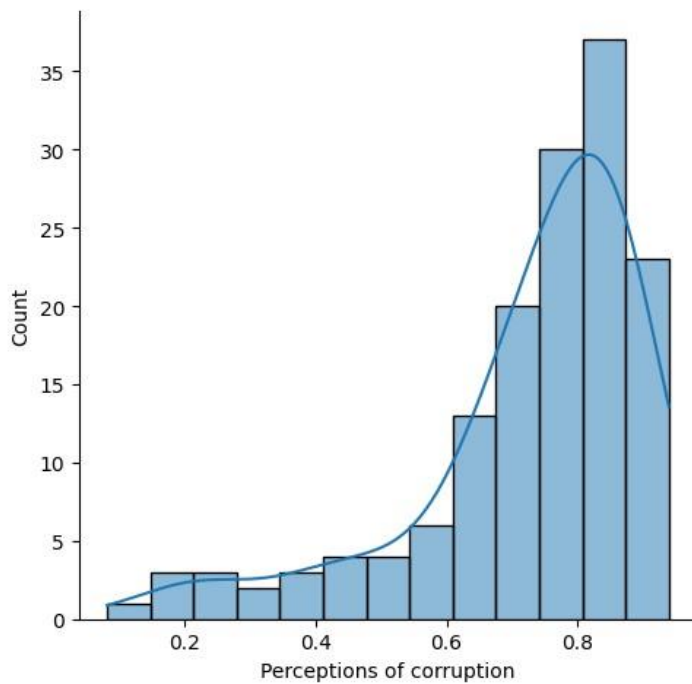
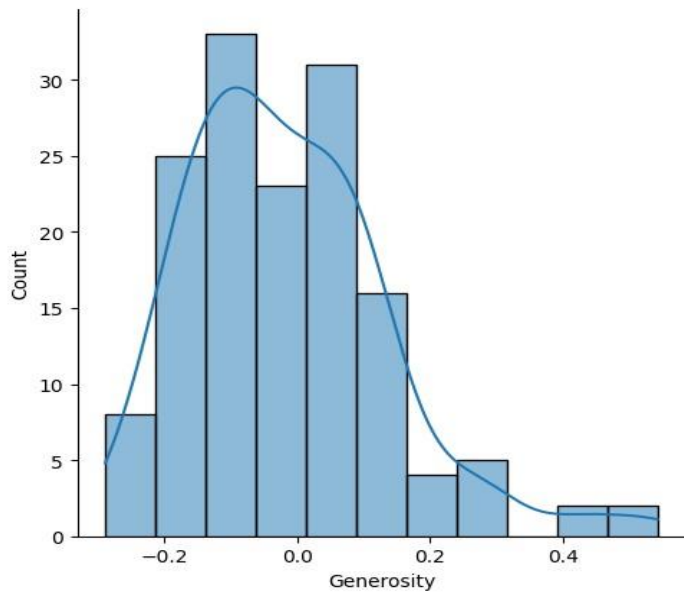
# DATA DISTRIBUTION

After cleaning & preparation of data, we go through deep inside every data parameters and identified many trends and interesting insights.

Some of the concluded results are given below:







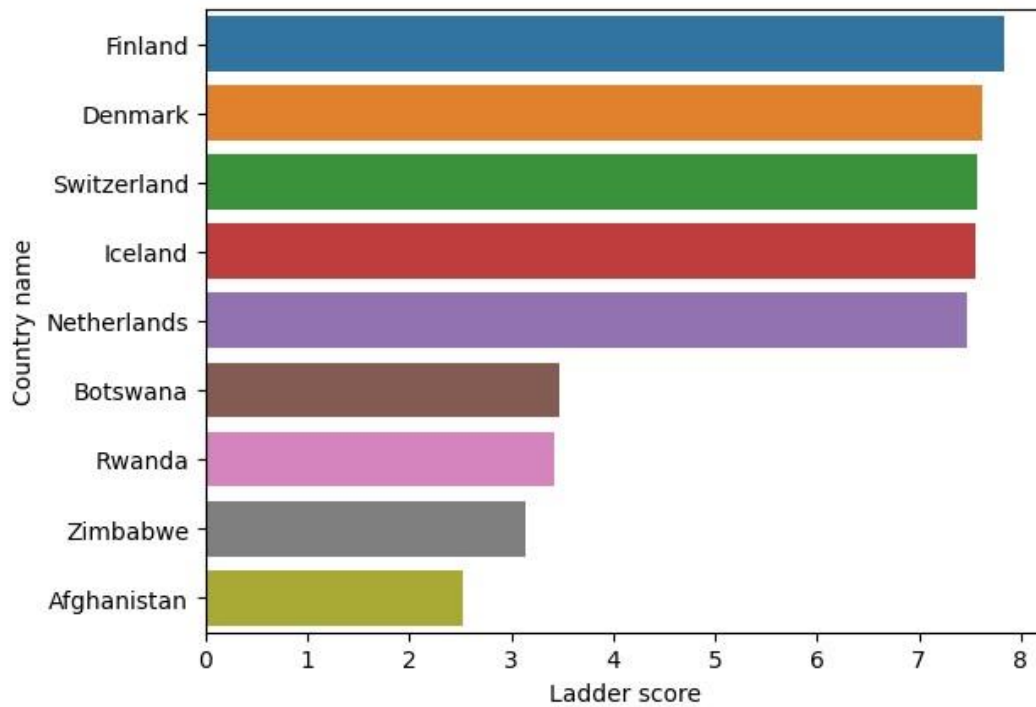
## Insights:

From the above statistical analysis we can conclude that mostly countries have average ladder score (not so high and not so low), social support and health conditions are good in most of the part of the world, but in term of corruption and generosity we still need improvement.

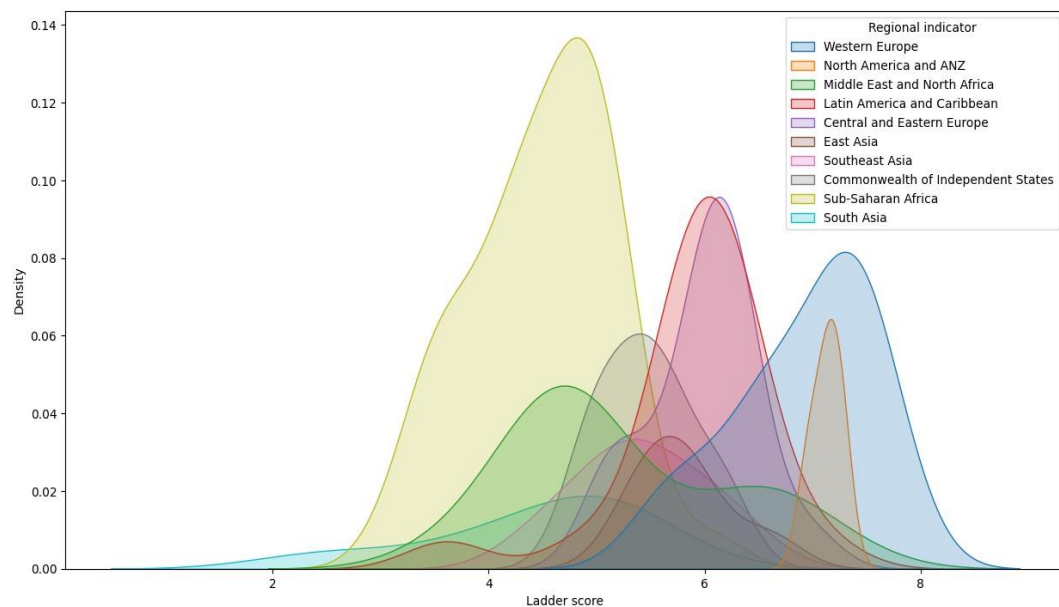
# Univariate analysis

After performing univariate data analysis we get following results and Insights that is given below:

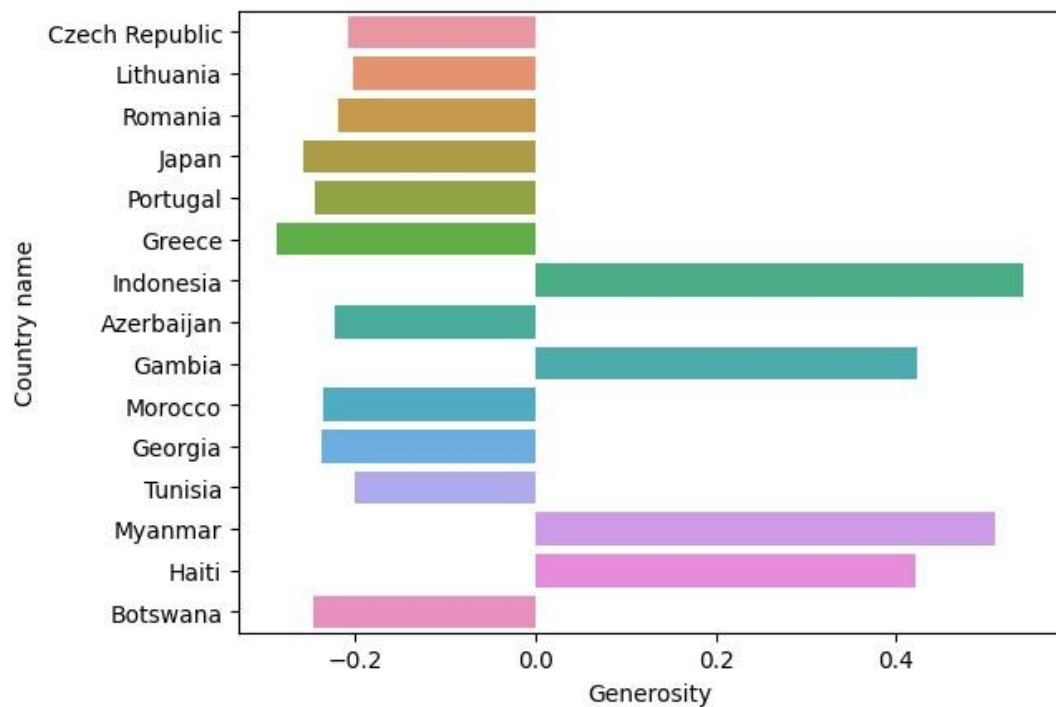
## 1. Most happiest and unhappiest Countries



## 2. Ladder Score Distribution by Regional Indicator



### 3. Most Generous and Most Ungenerous Countries



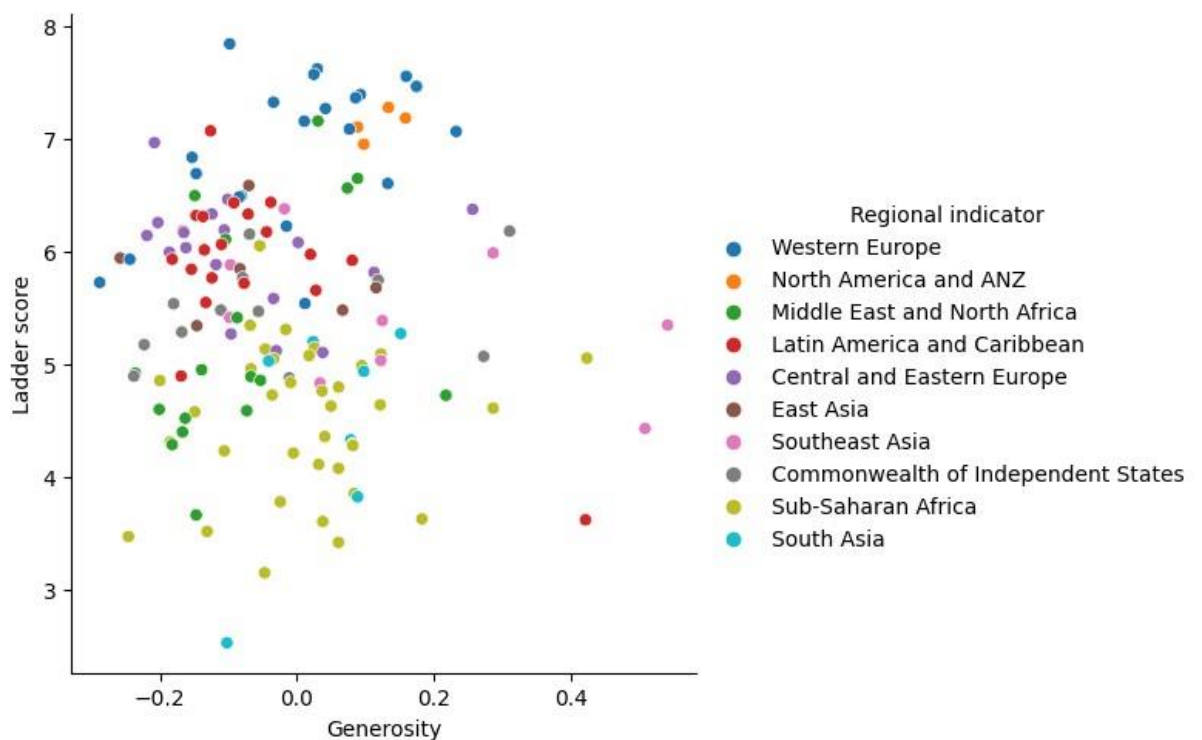
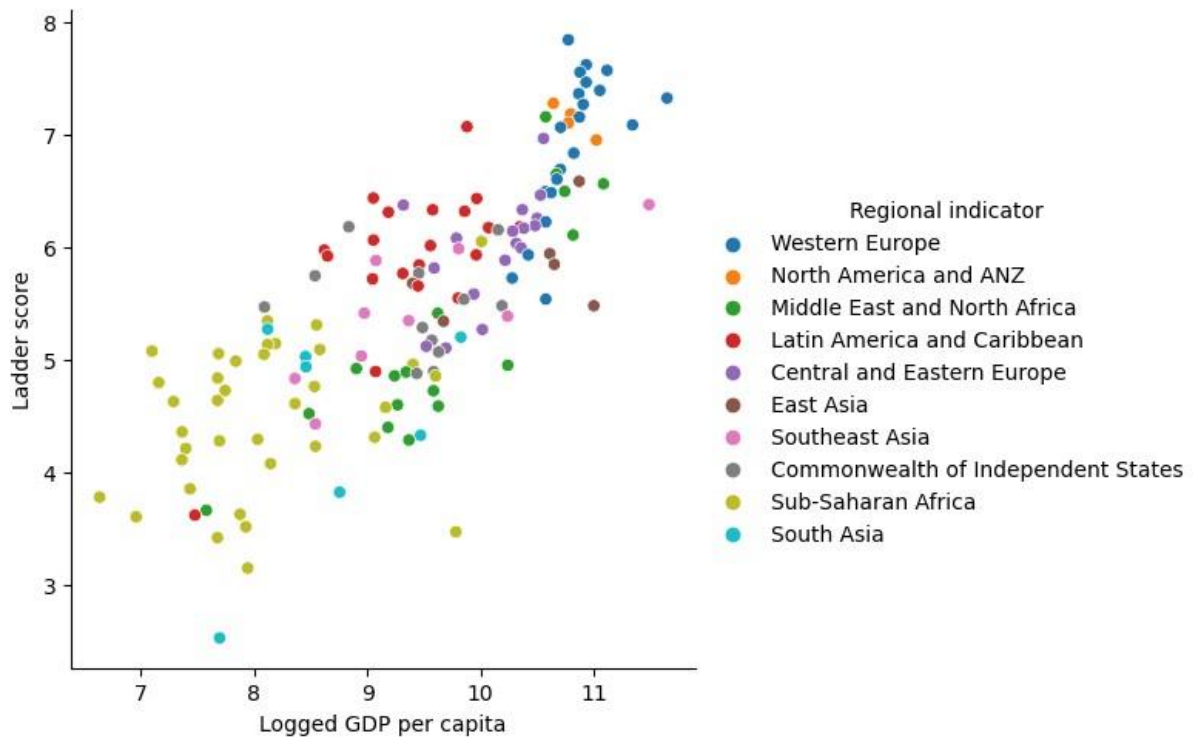
#### Insights:

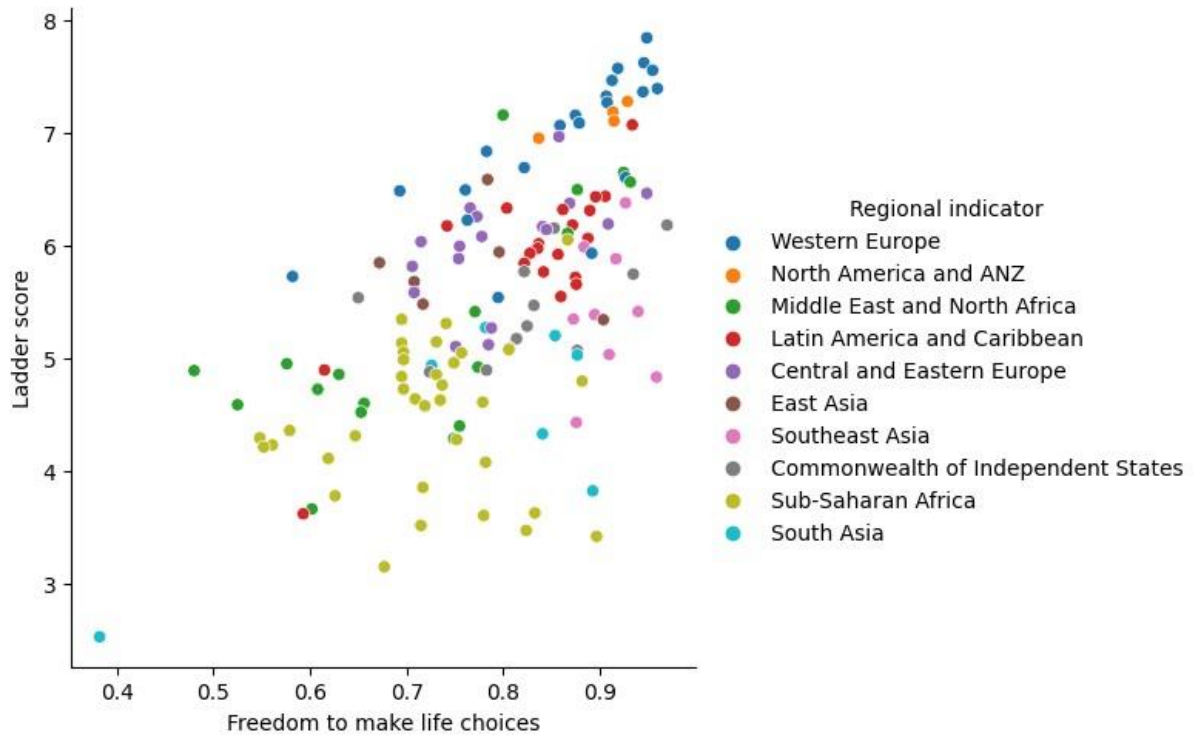
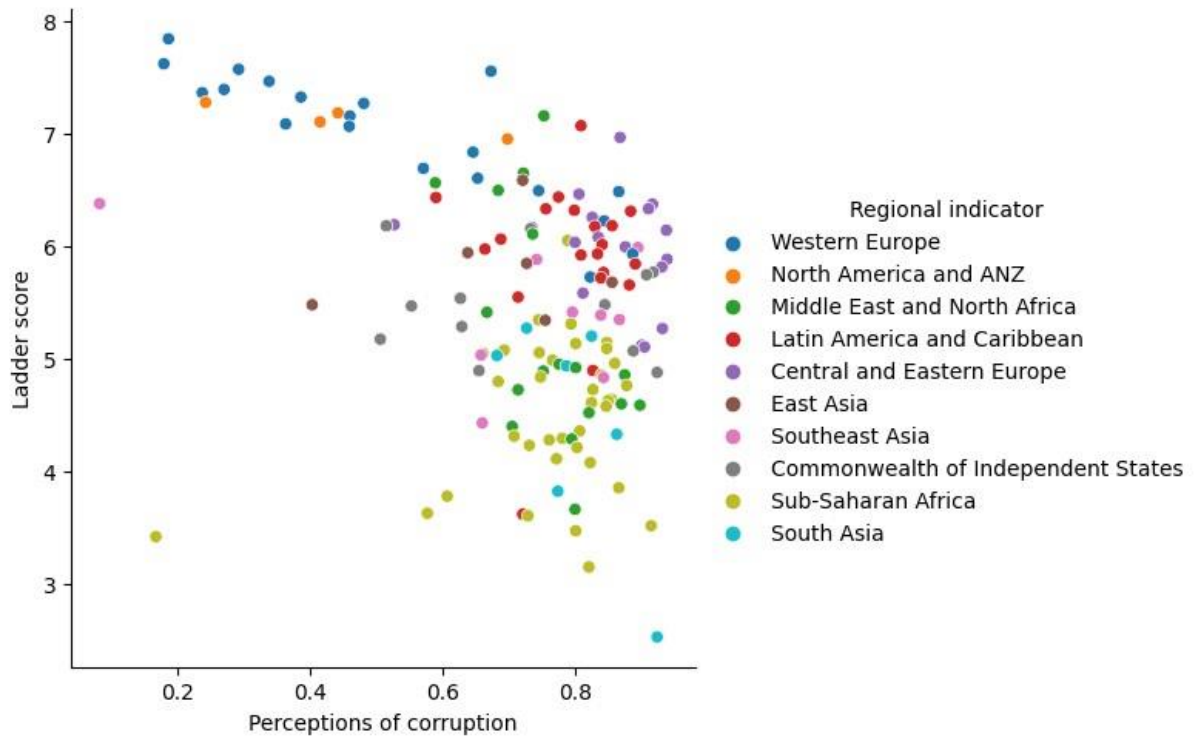
Univariate analysis shows that mostly european and american countries have high ranking in term of world happiness, followed by east asian, south asian and then african countries. interstingly developing countries have more generosity rate in comparison to some developed countries

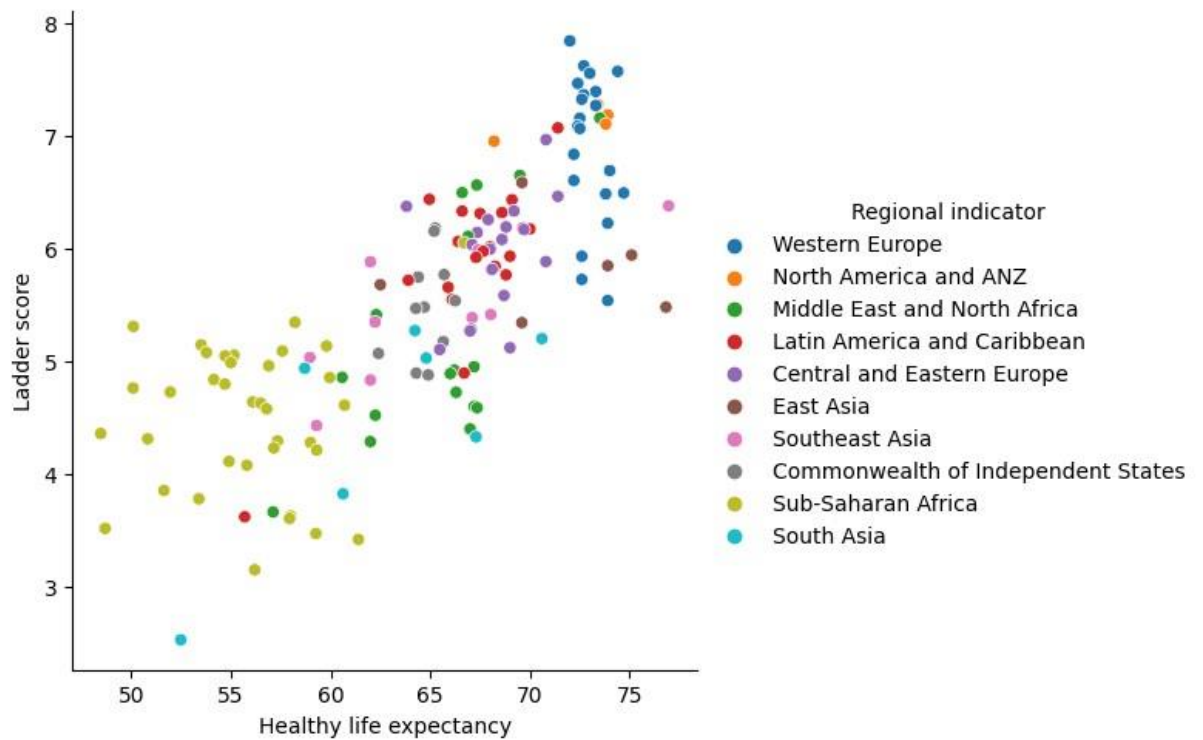


# Bivariate analysis

We have done Bivariate analysis on Relationship between Ladder Score and various feature according to Different Regions. Below is gained results and derived insights:





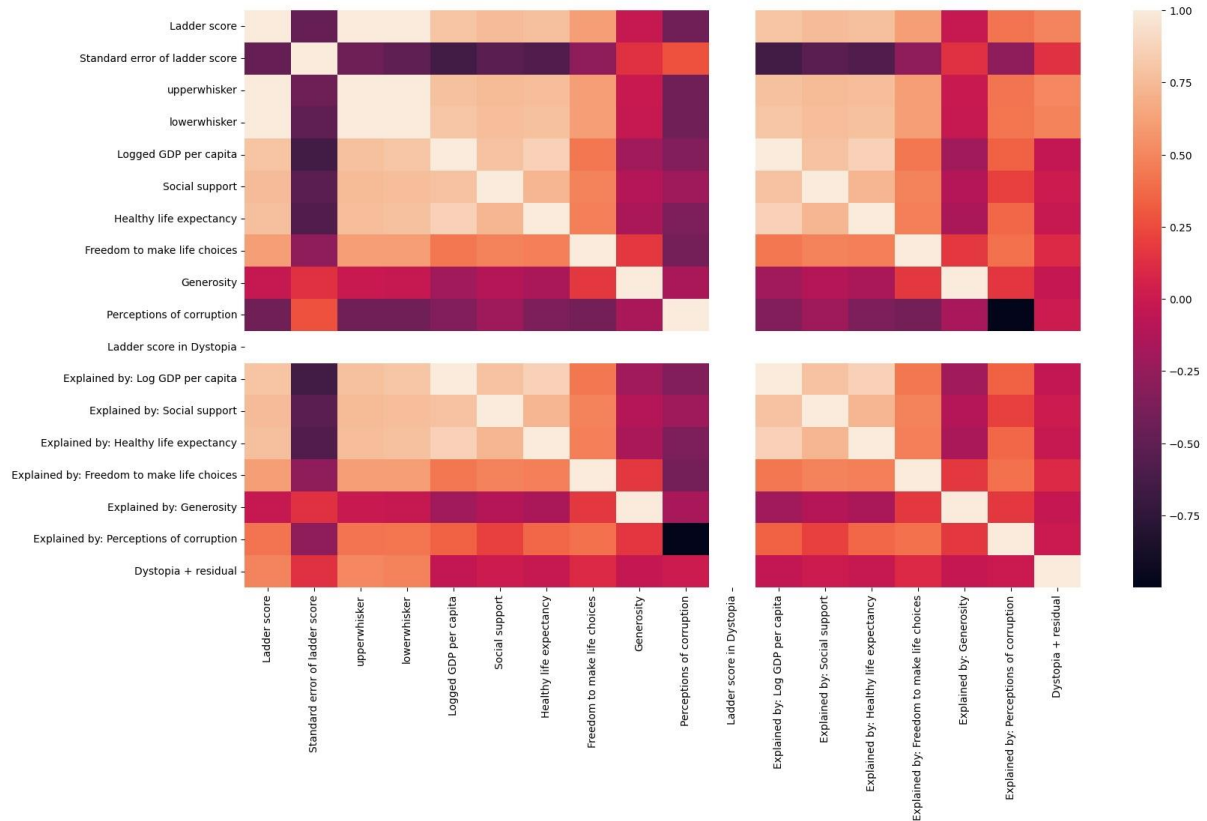


### Insights:

From bivariate analysis we can conclude that the countries having high GDP, health condition, and freedom is having high ranking in term of happiness, secondly corruption is major problem of the world except some European countries.

# Multivariate analysis

In multivariate analysis we have used treemap to determine effect of various factors and have determined most influencing factors. Below is our derived observation:



## Insights:

From the above treemap we can conclude that corruption is having high influence on various socio economic factors, secondly generosity in society is also a factor.

## Conclusion

From the above data distribution, univariate, bivariate and multivariate analysis of Dataset of world happiness report. We have concluded that to improve any country's happiness ranking. We must have to reform social structure of that region. As we have seen corruption, freedom and generosity are the main influencing factors in deciding happiness ranking of that country. These all factors have direct or indirect relation with social ethics and moral values. So government should also focus on quality and betterment of social structure apart from income perspective. And somehow GDP and income like factors also dependable as all these things have affect on an individual efficiency of earning capacity.