

ANALYZING THE FACTORS CONTRIBUTING TO GLOBAL HAPPINESS AND CORRUPTION

Dinula Eshan Jayawardana - COADDS 23.1-001

Advanced Diploma in Data Science

National Institute of Business Management (NIC)

Colombo, Sri Lanka

Date of Submission – 28th March 2024

“A project report submitted to the national Institute of Business management for the partial fulfillment of requirement of the Advanced Diploma in Supplies and Material Management.”

DECLARATION

I hereby declare that the work presented in this project report was carried out independently by myself and have cited the work of others and given due reference diligently.

.....

Eshan Jayawardana

.....

Date

I certify that the above student carried out his project under my supervision and guidance.

.....

Supervisor:

Ms. Chamilanka Wanigasekara

.....

Date

ACKNOWLEDGEMENT

First of all, I would like to express my gratitude to all those who have provided support and necessary guidance for the successful completion of this research. Without their guidance I would not have been able to complete this research successfully. I must first express my thanks to the research supervisor Ms. Chamilanka Wanigasekara who provided knowledge, encouragement and active guidance during the research. I also express my gratitude to my friends who have always supported me and shared their knowledge with me. Finally, I would like to thank my family for their support.

ABSTRACT

This research mainly studies how happiness exists in the world and factors like social environment, political climate, freedom and the relationship between these factors and corruption in a country. The main purpose of the research is to identify and focus on the factors that affect happiness. Also, through this study, the factors necessary for the independence of the country's continents are also studied. Due to the inclusion of five years of data in this research, the effect of some factors is noticeable.

Also, the research talks about corruption in poor countries and also talks about the political connection to it. It also discusses the issues of social environment and philanthropy in the society.

This research includes statistical tests to further study the objectives and hypotheses. Also, various analysis methods and various relationships are used to further study these factors. However, some conclusions have to face various limitations in the field and in conducting further studies.

Table of Contents

CHAPTER 1	1
INTRODUCTION	1
1.1 BACKGROUND	1
1.2 RESEARCH PROBLEM	2
1.3 RESEARCH QUESTIONS	2
1.4 OBJECTIVES OF THE PROJECT	3
1.5 SCOPE OF THE RESEARCH	3
1.6 JUSTIFICATION OF THE RESEARCH	3
1.7 EXPECTED LIMITATION	4
1.8 PROPOSED WORK SCHEDULE	5
CHAPTER 2	6
LITERATURE REVIEW	6
2.1 INTRODUCTION OF RESEARCH THEME	6
2.2 THEORETICAL EXPLANATION ABOUT THE KEY WORDS IN THE TOPIC	6
2.3 FINDING BY OTHER RESEARCHERS	8
2.4 THE RESEARCH GAP	10
2.5 TABLE FOR VARIABLES THEIR DEFINITIONS AND SOURCES	10

2.6 CHAPTER CONCLUSION -----	12
 CHAPTER 3-----	13
METHODOLOGY -----	13
3.1 INTRODUCTION-----	13
3.2 POPULATION, SAMPLE AND SAMPLING TECHNIQUE ----	13
3.3 TYPE OF DATA TO BE COLLECTED AND DATA SOURCES	13
3.4 DATA COLLECTION TOOLS AND PLAN -----	14
3.5 CONCEPTUAL FRAMEWORK -----	14
3.6 HYPOTHESIS -----	15
3.7 METHODS OF DATA ANALYSIS-----	16
 CHAPTER 4-----	17
DATA ANALYSIS AND INTERPRETATION -----	17
4.1 DATA ANALYSIS AND INTERPRETATION INTRODUCTION	17
4.2 DATASET UPLOADING TO THE PYTHON -----	17
4.3 DATA PREPROCESSING-----	18
4.4 DESCRIPTIVE AND EXPLORATORY DATA ANALYSIS-----	20
4.5THE INFLUENCE OF FACTORS AFFECTING THE HAPPINESS SCORE OF A COUNTRY -----	26

4.6 EXAMINING THE INTERRELATIONSHIP BETWEEN CONTINENTS AND FREEDOM. -----	30
4.7A STUDY OF THE RELATIONSHIP BETWEEN GENEROSITY AND HEALTH WITH GDP PER CAPITA -----	33
CHAPTER 5-----	36
DISCUSSION AND RECOMMENDATIONS -----	36
5.1 DISCUSSION -----	36
5.2 RECOMMENDATION -----	38
5.3 CONCLUSION -----	38
APPENDICES -----	40
REFERENCES -----	45

Table of Figures

FIGURE 1=CONCEPTUAL FRAMEWORK -----	14
FIGURE 2=NULL VALUES -----	18
FIGURE 3=OUTLIER BOXPLOT 1 -----	19
FIGURE 4=OUTLIER BOXPLOT 2 -----	19
FIGURE 5=DESCRIPTIVE DATA ANALYSIS -----	20
FIGURE 6=HAPPINESS SCORE BY CONTINENT BAR GRAPH -----	21
FIGURE 7=HAPPINESS SCORE BY YEAR LINE CHART-----	22
FIGURE 8=HAPPIEST AND UNHAPPIEST COUNTRIES BAR GRAPH -----	22
FIGURE 9=COUNTRIES WITH THE HEALTH RATES PIE CHART -----	23
FIGURE 10=CONTINENT WISE CONTRIBUTION TO WORLD ECONOMY PIE CHART	24

FIGURE 11= CORRELATION HEAT MAP-----	25
FIGURE 12= SCATTER PLOTS WITH HAPPINESS SCORE -----	26
FIGURE 13= SCATTER PLOTS WITH HAPPINESS SCORE -----	27
FIGURE 14= ACTUAL AND PREDICTED HAPPINESS SCORE SCATTER PLOT-----	29
FIGURE 15= TUKEY’S HSD RESULTS TABLE -----	31
FIGURE 16= GDP PER CAPITA AND HEALTH SCATTER PLOT -----	34
FIGURE 17= GDP PER CAPITA AND GENEROSITY SCATTER PLOT-----	35

CHAPTER 1

Introduction

1.1Background

In recent years, the study of human happiness is an important task that goes beyond various fields and beyond various researches. The reason for that is because nowadays man is paying more attention in terms of technology and economy. Therefore, various factors are used to find out about the happiness and social satisfaction of a country.

Factors that are given more attention in this research are happiness indicators, different social structures, economic conditions, health, government problems and corruption. There, comparative decisions are made between different countries and continents.

This analysis can be further expanded by examining factors such as per capita income, family and social support, freedom and generosity.

Also, because it is possible to get information on a global basis, it is easy to study the changes that occur over time by examining the landscape and the factors that affect it. Also, with the information on continents, the changes that occur due to cultural changes can also be studied here.

The basic need of this research is to identify the strategies to maintain social well-being with the economic and technological development happening around the world. It also looks for new strategies to improve happiness and good attitude skills of human beings.

1.2 Research Problem

- Understanding the complex web of factors that contribute to happiness scores across nations has been problematic.
- Differences in the degree of freedom across continents raise intriguing questions about social and political factors.
- The complexity between economic prosperity and social generosity raises questions about the morality and economic structure of nations.

1.3 Research Questions

- How does family environment and health influence global happiness score variation?
- Differences in the level of freedom across continents and how it affects social well-being?
- How does GDP per capita and a country's economy affect the generosity of its citizens?

1.4 Objectives of the Project

- Analyzing the relative influence of factors affecting different country happiness scores.
- Compare the level of freedom between continents using relevant indicators and investigate the factors affecting it.
- Investigate the correlation between per capita gross domestic income and generosity and examine the economic and social factors that influence it.

1.5 Scope of the Research

In this research, the factors affecting the happiness sign are studied globally. In this research, the factors such as social environment, economic level, political situation, freedom, family environment are studied. And it mainly describes how corruption and freedom of a country affect this happiness score. It also gets an idea of how happiness scores vary. Python is mainly used for analysis. Where data is used for statistical tests such as regression model, correlation analysis and ANOVA. Visual analysis methods are also used.

1.6 Justification of the research

By studying this research, we can gain an understanding of the factors that affect happiness globally. This study can gain an understanding of the difference between happiness and freedom among different nations when considering factors such as per capita domestic production, health, social harmony, and government behavior. This research gives importance

to fields like sociology and economics. Along with these facts, it is necessary to gain an understanding of the moral and economic dynamics between different races. This makes it possible to understand that things like happiness, freedom and corruption are the most neglected things in a nation.

1.7 Expected Limitation

Any research project has to face some limitations. Some of the data in this research data set may not be sufficient to study the entire population. And because this data set is obtained through another website, there is a problem with the accuracy or reliability of the data. Also, because time was limited to conduct further study of the data, a secondary data set had to be used. And because this data set is a secondary data set, there are some limitations for some heavy analysis methods. Therefore, due to the limitation of the number of statistical data analysis methods, it will be difficult to study the research questions.

1.8 Proposed Work Schedule

	Nov	December				January				February				March		
	1	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3
Choosing the Topic																
Find the dataset and Literature Review																
Development of Methodology																
Proposal Writing																
Completion of Project Proposal																
Research and Data Analysis																
Final Report Submission																

CHAPTER 2

Literature Review

2.1 Introduction of Research Theme

Analyzing data and making decisions using statistical data has become prominent in the recent world. In this data set, global human good mood or happiness and the factors that affect it, gives knowledge about how people change due to the corruption of a country.

In this literature review, before getting more knowledge about our research, the data analysis and the published reports are studied. The most prominent in this research is the use of data analysis methods and the use of visualization methods. There, it becomes easy to identify the various relationships between these factors and analyze them. These literature reviews are important to get an important understanding for that.

2.2 Theoretical Explanation about the Key Words in the Topic

Happiness

The term happiness has been given various definitions. Happiness is such a complex emotion that it is difficult to define precisely. Happiness is living in good conditions like material prosperity, freedom and peace. Also, the terms freedom and peace have similarities to happiness.(Ruut Veenhoven). In the past few decades, the word happiness has become a major topic in public policy, environmental nomenclature, and psychology. The Stiglitz-Sen-Fitoussi Commission took the initiative to bring the topic of happiness to public attention to measure social progress and focus more attention on happiness.

(Stiglitz J. 2009). Happiness is an important component in the health literature and various constructs have been defined for it.(Stone AA, Mackie C, eds. 2013). And people seek happiness in different ways. That is because happiness is necessary to live. They try to get happiness by getting more income, a prestigious job, a shiny car. But happiness has been shown not to be.(Sheldon & Lyubomirsky, 2006a). For that, man has found simple behavioural strategies in everyday life to increase happiness.(Sin & Lyubomirsky, 2009). The word happiness has been used as a basic factor to measure the quality of life in developed countries since the distant past. This has focused attention on the word happiness in nations. (Veenhoven 1999). However, no measure of the word happiness has been found so far. But a number of scales have been found to have reasonable validity. (Larsen et al., 1985).

Corruption

Corruption is a problem that mainly arises between the government and the market economy. It can be pointed out that it is mainly the government itself. Hence handling corruption becomes somewhat difficult. Corruption discourages foreign investment. Abuse of public power for personal gain. Due to factors such as uncertainty related to the country's expenditure, it affects the country's standing. Because of that the word happiness fades away as the people have to experience the damage. (Shleifer and Vishny, 1993; Mauro,1995; Wei, 2000a). However, scholars say that more corruption tends to facilitate transactions and investments in cases where there are legal regulations. (Huntington, 1968; Leff,1989). It has been revealed that the government of many corrupt and poor countries is very corrupt. Research on pollution has shown this to be the case. Sections such as per capita local level also arise from this section. (Treisman 2000; Paldam 1999). All the

conclusions show that corruption in poor countries is mainly related to the government. It directly affects the happiness of the people of the country.(Becker, Gary S., and George Stigler J. 1974.).

2.3 Finding by Other Researchers

In this research, (Satrovic, Cetiner, Muslija (2018)) analyzed whether corruption control is important for public happiness using data from 59 countries from 2007 to 2019. This research aims to provide evidence on the impact of pollution control on happiness for social progress.

In the research done by B Akkaya in the year 2022, the effect of corruption and economic freedom on national happiness has been studied. This is global research. Here studies have been done using visualization methods. This analysis has been done using factors such as per capita income of man.

According to a study by Ruut Veenhoven, a global study in the early 1990s explored freedom and happiness. It is mentioned here that the topics of freedom and happiness are linked to each other. In this study it is stated that there is no limit to the words freedom and happiness.

In 2021, Carol Lee Graham examines insights from studies of happiness around the world. There it is shown how happiness changes with the economic level. It has also been discussed how it changes in crisis situations. It also describes how friends, crime and corruption affect this.

A study by Arvin B Mak and Lew Byron analyzes development aid, poverty and the happiness of nations using data from 118 countries from 1996 to 2009. Because the research

has been done by exploring the relationships between various factors between each country, the factors affecting happiness and corruption can be studied more.

Another research examines the global relationship between corruption and well-being and happiness (Terry Anderson). It has been pointed out how political, economic and social factors in the country affect this. And in the analysis of the data, relationships between various factors have been shown.

In 2010, Bo Rothstein conducted a study on corruption, happiness, trust and public welfare. There, the political opportunities affecting this are studied further. The discussion here is mostly about social welfare. It analyzes social welfare at the time of corruption and social welfare at the time of happiness. It further discusses how these factors relate to social trust.

Monica Violeta Achim conducts a study between corruption, economy and happiness in Romania in 2020. In this study, this research has been done by exploring information such as gender, age gaps, learning environments, and family environment. This can study how this situation affects globally.

Andrew Steptoe does an annual review of happiness and health. There the main factors for happiness are analyzed. Factors like learning, family environment, social friendships, mental status, economic level etc.

Accordingly, it is clear that the topic of happiness and corruption is a very important topic. By conducting this study, we hope to further analyze the factors affecting this. Accordingly, the relationship between the important factors and causes is also studied.

In 2015, Simeon Djankov, Elena Nikolova and Jan Zilinsky studied the happiness gap in Eastern Europe. Here, a comprehensive investigation was conducted using data from the early 1990s to 2014. In this study, it is easy to get an understanding because of the research done using statistical data.

2.4 The Research Gap

The research gap is the lack of professional research and the lack of research using statistical data in studying the relationship between global happiness and corruption. In the research, the factors affecting happiness such as freedom, social association, family background etc. have been discussed, but the factors affecting corruption and its causes are unclear. Although social, economic and political climates that affect happiness have been discussed more in research, attention to other general factors has been limited. Although the reasons for corruption are discussed with the government, there is minimal discussion of the issues from the society. To overcome this research gap, more study of various factors is done using statistical data analysis.

2.5 Table for Variables their Definitions and Sources

Variable	Source	Definition
Country	From the Kaggle	Country name
Happiness score	From the Kaggle	Average of responses to evaluation questions from the Gallup World Poll
GDP per capita	(Helliwell and Huang 2008a, 617)	The extent to which GDP contributes to the calculation of the Happiness Score. (Gross Domestic Product)
Family	From the Kaggle	The amount that the family contributes to the calculation of the happiness score

Health	From the Kaggle	A measure of how well health contributes to happiness
Freedom	(Veenhoven 1997)	The numerical value between the value of a country's freedom and happiness
Generosity	From the Kaggle	A statistical value obtained for generosity in the country
Government trust	From the Kaggle	The extent to which trust in a country's government affects happiness scores
Dystopia residual	From the Kaggle	Scores based on a comparison with the world's saddest country
Continent	From the Kaggle	The region where the country is located
Year	From the Kaggle	Year of data collection
Social support	(Uslaner 2002)	Social support or the amount of being happy with each other
Cpi score	(Helliwell and Huang 2008a, 617)	Corruption perception index (the higher the better).

2.6 Chapter Conclusion

This can explain the causes and effects of happiness and corruption. It is the nature of the government that has deprived many poor citizens of their happiness and freedom. It is clear how these conditions affect the economic difficulties and friendships of the society. There are many factors that indicate that lack of happiness leads to the decline of society. Therefore, attention should be paid to matters such as the health of the country. As social corruption increases, social problems increase more and more. Therefore, governments should work to promote the word happiness.

CHAPTER 3

Methodology

3.1 Introduction

This section discusses the tests conducted in the data set and the use of various data analysis to study the research problems and objectives. Using the data to make various hypotheses, regression model, correlation analysis and various data visualization methods are used here. It is to identify the main elements that a country must have for happiness and to explain and analyze the reasons for its global change.

3.2 Population, Sample and Sampling Technique

The research considers the factors that affect happiness in different countries. From 2015 to 2020, published values for happiness in each country are given here. Also, nine factors affecting it are given here and their numerical values are given. All 792 items are shown in the data set. Because a huge amount of data has to be obtained in the countries, a random sampling technique is used for this.

3.3 Type of data to be collected and data Sources

For this research, a secondary data set obtained from Kaggle websites is the data set "Happiness and Corruption 2015-2020". It includes 3 categorical variables and 10 numerical data.

3.4 Data Collection tools and plan

Collected randomly from the kaggle website to analyze how human happiness, freedom and suffering are changing globally, this includes methods of collecting secondary data such as surveys, observations, discussions, and reports.

3.5 Conceptual Framework

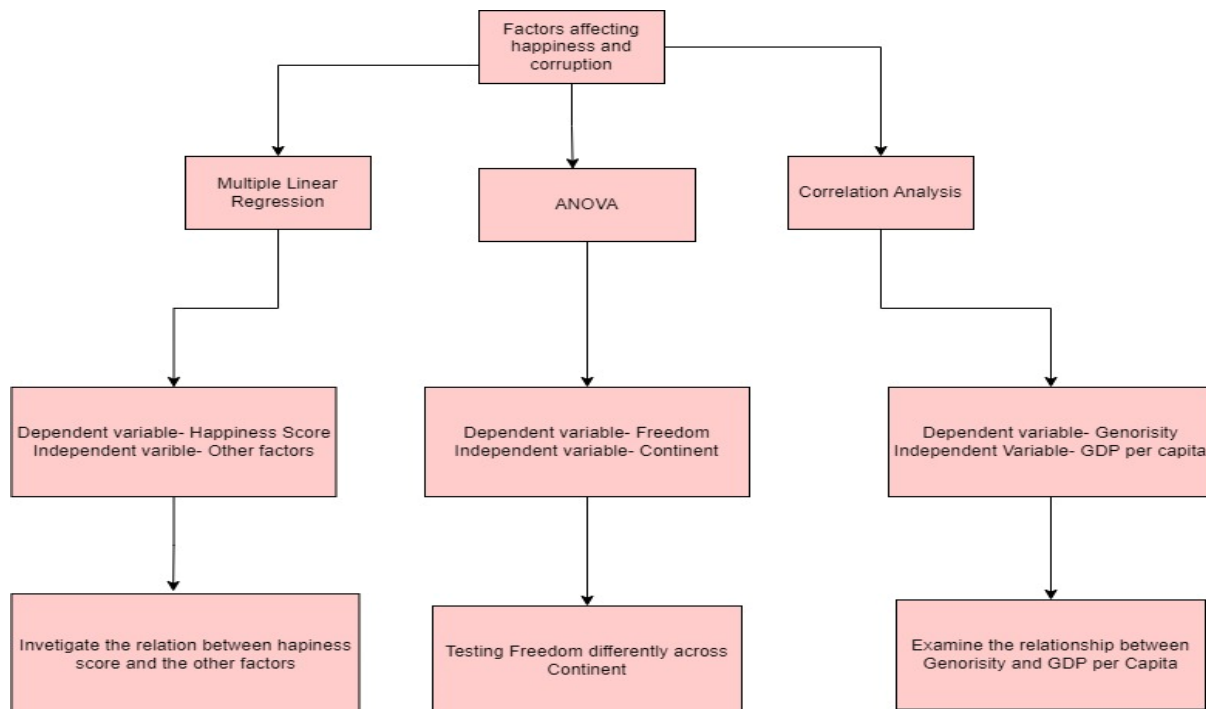


Figure 1=Conceptual framework

A visual framework is used to identify and examine the complex factors that influence global happiness and well-being. Happiness includes variables such as trust in government, health, social support, economy and freedom. This sub-framework outlines the relationships between variables and includes the effects of politics. By systematically studying this picture frame, the main factors that take care of the happiness score, which is the main objective, can be analyzed and the relationships between them can be analysis.

3.6 Hypothesis

Hypotheses in statistical data analysis are statements that can be tested in data analysis.

There are two types of this called null hypothesis and alternative hypothesis. The null hypothesis states that there is no difference between two variables. The alternative hypothesis states that there is a difference between two variables.

H0 = Null Hypothesis

H1 = Alternative Hypothesis

H0 = There is no significant difference between per capita production and generosity analysis for the value of happiness scores within the country.

H1= There is a significant difference between per capita production and generosity analysis for the value of happiness scores within the country.

H0 = There was no association between family environment scores and freedom analysis.

H1 = There is a relationship between family environment scores and freedom analysis.

H0 = There is no relationship between the state of government and the state of health in the country for the corruption rate.

H1 = There is a relationship between the state of the country and the state of health for the rate of corruption.

3.7 Methods of Data Analysis

Python is used to explain the objectives of the research and to analyze the data. Statistical data analysis should be done to gain insight into this data set and the research.

Studying other factors that affect happiness scores in the country and finding out which factors are related to happiness scores.

- ✓ Multiple Linear Regression

Comparison and analysis of degrees of freedom between continents using different factors.

- ✓ ANOVA

Investigate the correlation between per capita gross domestic income and philanthropy and examine the economic and social factors that influence it.

- ✓ Correlation analysis

Using pie charts, scatter plots, box plot, bar chart, histogram to further analyze the relationships between variables.

- ✓ Exploratory Data Analysis

CHAPTER 4

Data analysis and interpretation

4.1 Data analysis and interpretation introduction

This chapter discusses the methods of data analysis used to show the validity of the previously mentioned objectives. The methods of data analysis are shown and the data analysis is described and it is clearly described with the research objectives. The data processing and data wrangling operations are described here for the changes made in the original data set before the data analysis of the research.

4.2 Dataset Uploading to the Python

The obtained data set has been uploaded to Python to prepare the necessary background for data analysis. At the same time, Python libraries required for data analysis are also provided.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
from scipy import stats
from scipy.stats import f_oneway
from statsmodels.stats.multicomp import pairwise_tukeyhsd
```

```
df=pd.read_csv("WorldHappiness_Corruption_2015_2020.csv")
df
```

	Country	happiness_score	gdp_per_capita	family	health	freedom	generosity	government_trust	dystopia_residual	continent	Year	social_support	cpi_score
0	Norway	7.5370	1.616463	1.533524	0.796667	0.635423	0.362012	0.315964	2.277027	Europe	2015	0.000000	88
1	Denmark	7.5220	1.482383	1.551122	0.792566	0.626007	0.355280	0.400770	2.313707	Europe	2015	0.000000	91
2	Iceland	7.5040	1.480633	1.610574	0.833552	0.627163	0.475540	0.153527	2.322715	Europe	2015	0.000000	79
3	Switzerland	7.4940	1.564980	1.516912	0.858131	0.620071	0.290549	0.367007	2.276716	Europe	2015	0.000000	86
4	Finland	7.4690	1.443572	1.540247	0.809158	0.617951	0.245483	0.382612	2.430182	Europe	2015	0.000000	90
...
787	Botswana	3.4789	0.997549	0.000000	0.494102	0.509089	0.033407	0.101786	0.257241	Africa	2020	1.085695	60
788	Tanzania	3.4762	0.457163	0.000000	0.442678	0.509343	0.271541	0.203881	0.718963	Africa	2020	0.872675	38
789	Rwanda	3.3123	0.343243	0.000000	0.572383	0.604088	0.235705	0.485542	0.548445	Africa	2020	0.522876	54

```
df.columns
```

```
Index(['Country', 'happiness_score', 'gdp_per_capita', 'family', 'health',  
      'freedom', 'generosity', 'government_trust', 'dystopia_residual',  
      'continent', 'Year', 'social_support', 'cpi_score'],  
      dtype='object')
```

4.3 Data Preprocessing

The data set used in this research has been uploaded by the author in csv format. There are nearly 790 data in the original data set and the same data was used for the research. Before conducting the research, an understanding of this data set was obtained through Excel. Then the null value is checked using python.

```
<class 'pandas.core.frame.DataFrame'>  
Int64Index: 688 entries, 2 to 791  
Data columns (total 13 columns):  
#   Column                Non-Null Count  Dtype  
---  ---  
0   Country                688 non-null   object  
1   happiness_score        688 non-null   float64  
2   gdp_per_capita         688 non-null   float64  
3   family                 688 non-null   float64  
4   health                 688 non-null   float64  
5   freedom                688 non-null   float64  
6   generosity             688 non-null   float64  
7   government_trust        688 non-null   float64  
8   dystopia_residual       688 non-null   float64  
9   continent               688 non-null   object  
10  Year                   688 non-null   int64  
11  social_support          688 non-null   float64  
12  cpi_score               688 non-null   int64  
dtypes: float64(9), int64(2), object(2)  
memory usage: 75.2+ KB  
None  
Index(['Country', 'happiness_score', 'gdp_per_capita', 'family', 'health',  
      'freedom', 'generosity', 'government_trust', 'dystopia_residual',  
      'continent', 'Year', 'social_support', 'cpi_score'],  
      dtype='object')
```

Figure 2=Null values

When testing for null values, there are no null values in the data set.

Before doing this research, I first need to look for outliers in my data set. Outliers are data points that deviate significantly from the overall pattern of a data set. It deviates more or less from the data. Outliers distort statistical measurements and affect the conclusions drawn from the data. Identifying outliers is important for identifying patterns and relationships within the data set.

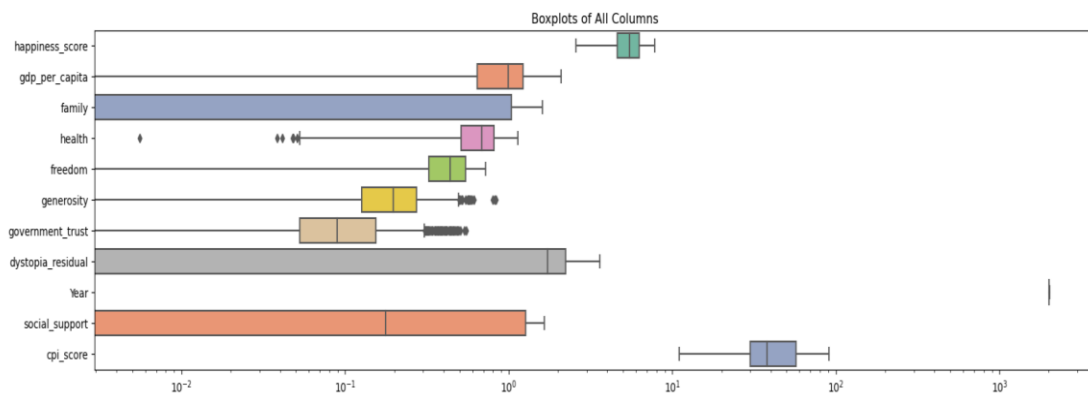


Figure 3=Outlier boxplot 1

A box plot is used to test for those outliers. This makes it easier to check outsiders. There are outliers in some variables. Therefore, those extraneous values have to be removed.

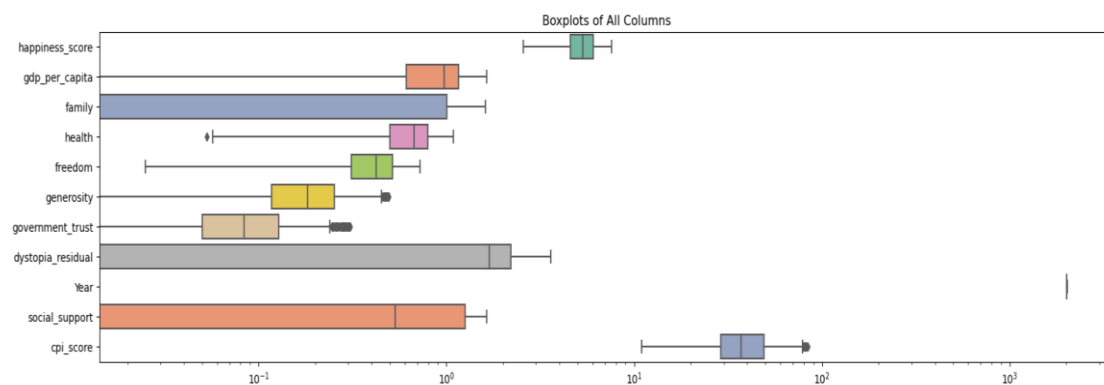


Figure 4=Outlier boxplot 2

After the removal process in the dataset, the dataset was re-indexed by python.

4.4 Descriptive and Exploratory Data Analysis

This section considers data analysis and visualization to investigate the research objectives.

There, first the summaries of the data set are obtained.

[9]:	happiness_score	gdp_per_capita	family	health	freedom	generosity	government_trust	dystopia_residual	Year	social_support	cpi_score
count	688.000000	688.000000	688.000000	688.000000	688.000000	688.000000	688.000000	688.000000	688.000000	688.000000	688.000000
mean	5.336017	0.891672	0.481718	0.633787	0.407443	0.193130	0.097454	1.353302	2017.502907	0.616018	40.840116
std	1.022603	0.364710	0.542082	0.218668	0.143039	0.103417	0.067045	1.081733	1.704517	0.632076	16.290772
min	2.566900	0.000000	0.000000	0.053000	0.000000	0.000000	0.000000	0.000000	2015.000000	0.000000	11.000000
25%	4.560275	0.608691	0.000000	0.500213	0.310303	0.116865	0.050000	0.000000	2016.000000	0.000000	29.000000
50%	5.337500	0.974690	0.000000	0.670130	0.422920	0.182344	0.083587	1.696714	2018.000000	0.533684	37.000000
75%	6.070250	1.162577	1.005370	0.796035	0.516000	0.253070	0.128000	2.209258	2019.000000	1.258272	49.000000
max	7.561000	1.632952	1.610574	1.088000	0.724000	0.492774	0.306000	3.602140	2020.000000	1.644000	84.000000

Figure 5=Descriptive data analysis

Here a statistical summary of the data set is shown. The summary table provides insight into information such as total variable size, minimum value and maximum value. It also shows statistical data such as mean values, percentages of variances, and standard deviations.

The above table of analysis results obtained by python gives a clear understanding of the variables in the data set. Further, data visualization is used to gain a clear understanding of these variables and understand how the data is visualized in the variables. Here data visualization is the easy to understand and visually express complex data and graphical representation of information. It explores patterns in the data set, various changes in the data, etc. Various visualization media such as graphs are used for that. It enables effective understanding of data and systematic communication.

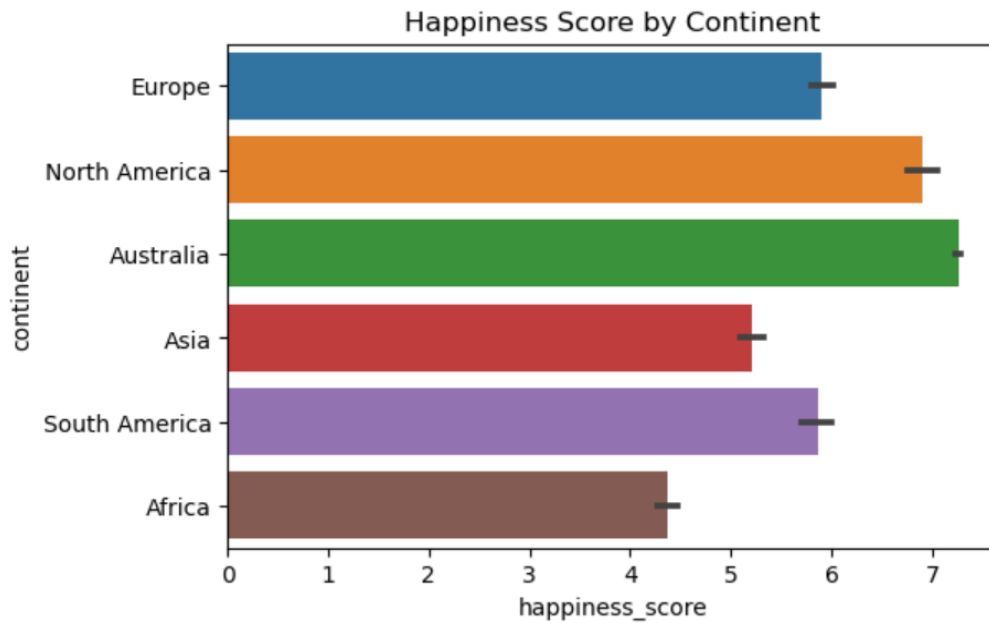


Figure 6=Happiness score by Continent bar graph

This chart gives an insight into how the sign of happiness flows between each continent. Here variable happiness score is used for x-axis and names of continents for y-axis. It includes data from 6 continents. This graph shows the average value of happiness between continents. The chart shows the Australian continent with the highest average value of 7. The African continent shows the lowest average value of nearly 4. Also, this gives an idea of the common value between other continents and can differentiate between happiness scores.

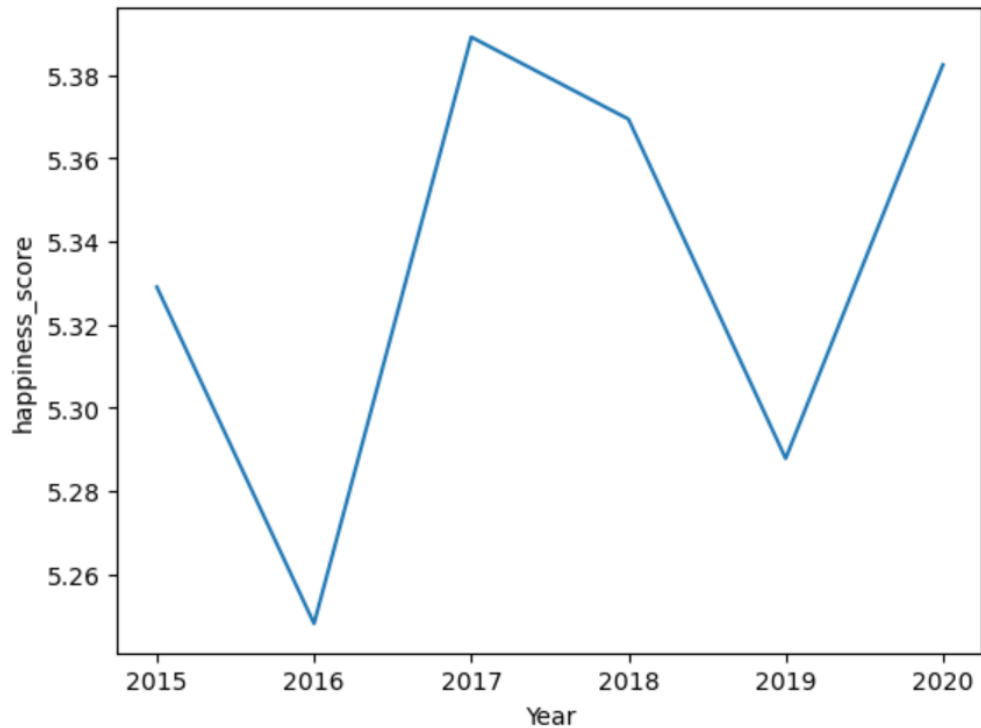


Figure 7=Happiness score by Year Line chart

This line graph shows the change in happiness sign from 2015 to 2020. Here the X-axis shows the year and the Y-axis shows the happiness score. This line graph shows the average happiness score over one year. The lowest average value is shown in the year 2016, which is a gross value of about 5.26. The year 2017 has increased and shows the highest average happiness value in these five years. It is a rough value of about 5.38. This line graph is important to understand the general flow of happiness value during these five years.

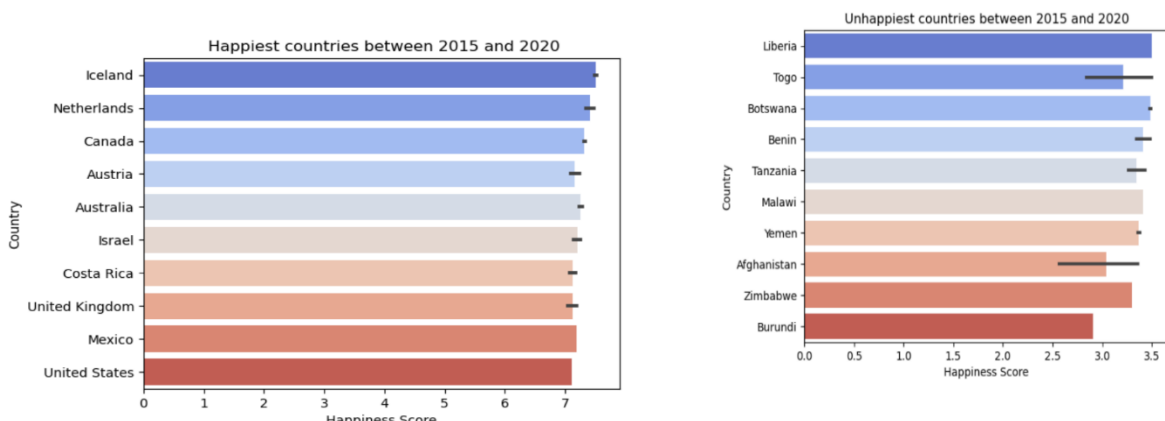


Figure 8=Happiest and Unhappiest Countries bar graph

Charts 1 and 2 provide a comparative analysis of the happiest and unhappiest countries with the happiness score. In both graphs, the X-axis shows the happiness score and the Y-axis shows the country name. From 2015 to 2020, information is shown between the countries with the highest happiness value and the lowest happiness value. Iceland, Netherland and Canada have the highest happiness values. Also, Burundi, Afghanistan and Togo are the countries showing the lowest level of happiness.

Countries With The Highest Health Rates



Figure 9=Countries with the Health Rates Pie chart

The improvement of the existing health level of a country has become a factor affecting the happiness of the country. The health of people in a country has a great impact on the happiness of a country. An example is the decrease in this level of health with the effect of covid and the effect of the sign of happiness on it. This pie chart shows the countries that maintained the highest levels of health from 2015 to 2020. In those years, Japan has maintained the highest level of health. In Japan, the happiness score is almost 6.

Continent Wise Contribution to World Economy

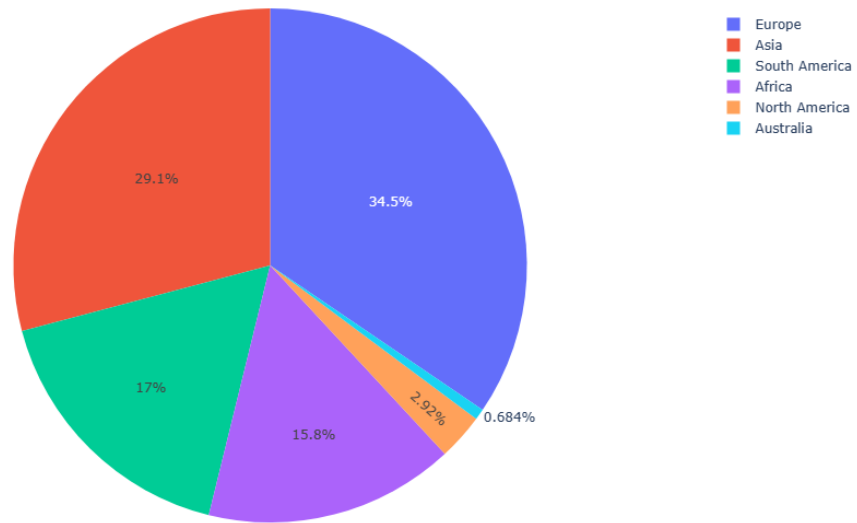


Figure 10=Continent wise contribution to world economy pie chart

This pie chart shows the relative percentage of participation of all continents in the world economy.

The continent of Europe occupies a special place there. It contributes to the world economy with a large percentage of 34.5%. However, European countries contribute directly to tasks such as trade, capital and technological development. Also, the Asian continent also contributes a lot to the world economy. It is a huge percentage of 29.1%. The Asian continent is home to powers like China, Japan and India. Also, industrialization with huge population between China and India provides great support to the world economy due to technological advancement. Also, Brazil, Argentina and Chile, which are the main countries in South America, contribute 17% to this. And the African continent contributes to this with a percentage of 15.8%. They contribute to this with their growing population, vast resources and increasing investment. Also, according to Pie chart, finally, North America and Australia contribute to the world economy with percentages of 2.96% and 0.684% respectively.

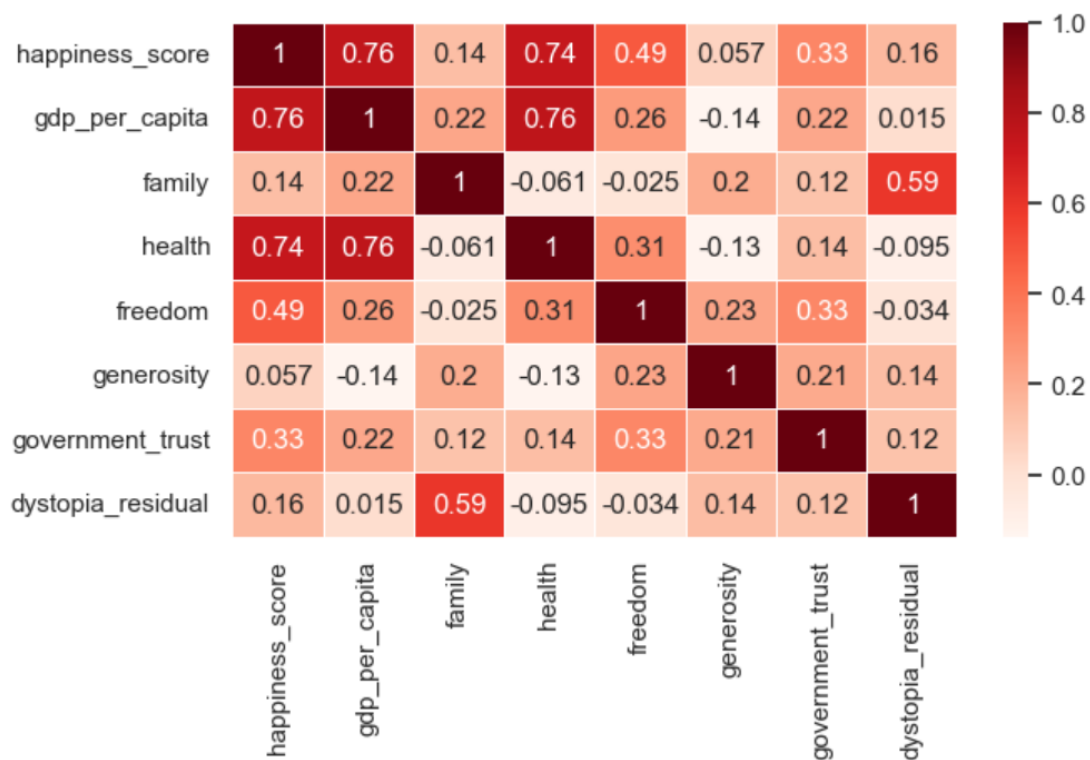


Figure 11=Correlation Heat map

This correlation heatmap provides insight into the relationships between the variables. Here dark red colors show positive correlations and light red colors show negative correlations. For example, the GDP per capita and the happiness score (0.76) show that the increase in one variable affects the happiness score due to the positive value shown here. And this negative relationship shown between family and health (-0.061) means that the increase in one variable affects the decrease in the other variable. In this way, the correlation heat map can clearly identify the relationships between the variables.

4.5 The influence of factors affecting the happiness score of a country

The main purpose of the research is the examination of the relationship between the factors affecting the happiness sign. The multiple linear regression method is used to further examine that objective. The study of the relationship between these factors is very important to get a deep understanding of the flow of the happiness score. In this method, the calculation is done using happiness score as Y variable (Dependent variable) and other factors as X variable (Independent variable). Here the relationship of these factors can be observed by scatter plot before using the multiple linear regression method.

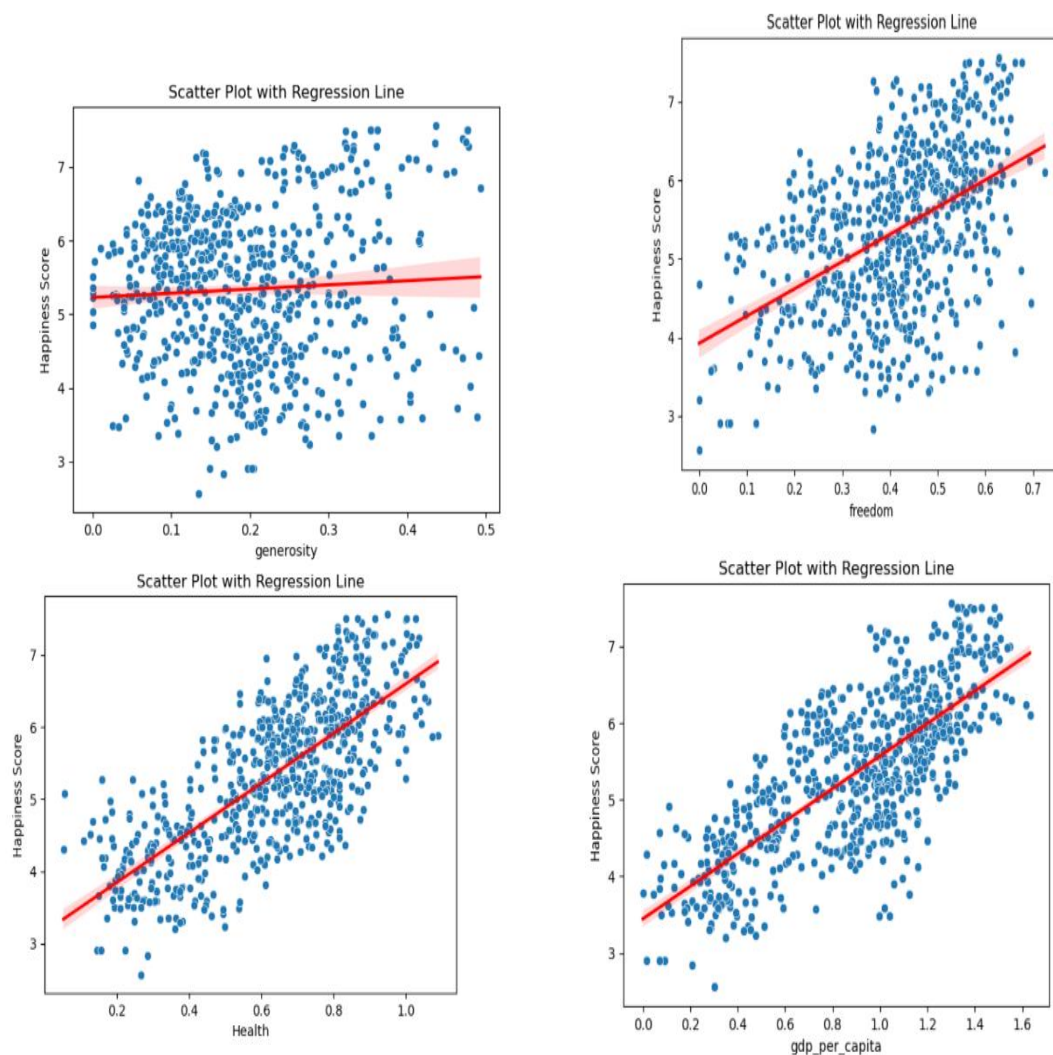


Figure 12=Scatter plots with Happiness score

Here a scatter plot was used to visually represent the relationship between two variables. These scatter plots are mainly created using the happiness sign. Each point in the scatterplot corresponds to a specific pair of data. Different patterns can be studied that show different relationships between variables. Here you can get basic knowledge about the nature of the relationship of the variables and this is more important for the subsequent analysis of the research. Multiple linear regression can be used to further study these relationships.

OLS Regression Results						
=====						
Dep. Variable:	happiness_score	R-squared:	0.721			
Model:	OLS	Adj. R-squared:	0.718			
Method:	Least Squares	F-statistic:	293.0			
Date:	Thu, 29 Feb 2024	Prob (F-statistic):	6.64e-185			
Time:	23:03:57	Log-Likelihood:	-552.23			
No. Observations:	688	AIC:	1118.			
Df Residuals:	681	BIC:	1150.			
Df Model:	6					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	2.2759	0.086	26.452	0.000	2.107	2.445
health	1.6928	0.159	10.619	0.000	1.380	2.006
government_trust	1.5380	0.338	4.556	0.000	0.875	2.201
family	0.0940	0.044	2.160	0.031	0.009	0.179
gdp_per_capita	1.1218	0.099	11.299	0.000	0.927	1.317
generosity	0.7631	0.221	3.449	0.001	0.329	1.198
freedom	1.5816	0.165	9.598	0.000	1.258	1.905
=====						
Omnibus:	12.168	Durbin-Watson:	1.492			
Prob(Omnibus):	0.002	Jarque-Bera (JB):	12.345			
Skew:	-0.305	Prob(JB):	0.00209			
Kurtosis:	3.241	Cond. No.	27.7			
=====						

Figure 13=Scatter plots with Happiness score

This summary gives a clear understanding of the differences in happiness sign. In the model, the R- squared value of 0.721 indicates that approximately 72% of the factors influencing the happiness score can be attributed to the variables included in the model. Here the adjusted R - squared value of 0.718 shows a stable outlook. And here mainly the F-statistic shows a value of 293 with a low p value of $6.64e^{-185}$. This increases the reliability of the happiness sign as a predictive tool. Also, considering the coefficients of these predictor variables separately increases the understanding of this. Here each variable expresses its effect. Where the coefficients of these variables are health (1.6928), government trust (1.5380), family (0.0940), per capita gross domestic product (1.1218), generosity (0.7631), and freedom (1.5816) with their low p-value (<0.05). with shows greater statistical significance. This means that these variables contribute significantly to the happiness score relationship.

There should be a clear understanding of the components while interpreting and analyzing the results of the model. Here the statistical value of the variables explains the significance of the happiness value. It is important to understand that the R- squared value accounts for a significant portion of the variance of the happiness score.

The multiple linear regression method facilitates the understanding and study of complex relationships between predictor variables and happiness scores. A scatter plot is used to further reveal the effect of those variables.

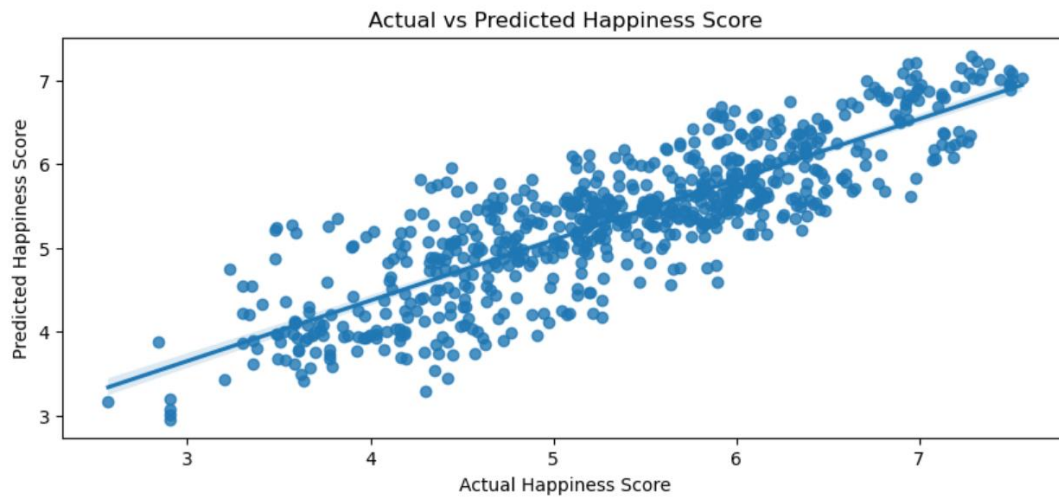


Figure 14=Actual and predicted happiness score scatter plot

The scatter plot shows a positive correlation between actual and predicted happiness scores. That is, when the actual happiness score increases, the predicted happiness score also increases. There are also some outliers where the happiness scores are lower or higher than predicted by the model. According to the model, sometimes happiness scores are different than expected. Also, in some cases, happiness scores vary less than expected. In addition to the obtained variables, it is concluded that other factors influence the happiness scores.

4.6 Examining the interrelationship between continents and freedom.

To study freedom globally, this research explores freedom across continents. Factors such as freedom, political climate, civil and economic freedom vary globally. In order to identify the factors affecting freedom and to study the relationship between them, the focus is on the continents of Asia, Australia, Africa, South America and North America. The main objective here is to examine whether there are statistically significant differences in freedom scores between continents. The primary purpose of this is to conduct an Anova test. Anova is a statistical method that investigates or compares between multiple groups. It tests whether there are significant differences between the means of different groups.

ANOVA Results:

F-statistic: 12.78965919556293

P-value: 6.675840729991192e-12

In the ANOVA test, the F-statistic represents the ratio between the between-group variance and the within-group variance. A higher F-score indicates a greater likelihood of group differences. And the p value measures the precision of the F value. The p value is important for decision making. The null hypothesis is rejected if the p-value is less than the specified value of less than 0.05. In this test, the dependent variable is the freedom score and the independent variable is the continent. According to the results of ANOVA test, F-statistic is 12.79 and p-value is 6.675e-12. The low P value of this result provides evidence to reject the null hypothesis. Therefore, it can be concluded that there are significant differences in at least one group. Here the high F-statistics in our analysis indicate significant variability in the freedom scores. It can also be observed that the observed differences are not due to chance as its p-value shows a significantly lower value.

Post-hoc tests are important to identify specific groups that differed from each other when the ANOVA test yielded a significant result. Tukey's Honestly Significant Difference (HSD) test is a similar post hoc test. It provides detailed insight into pairwise group differences. In this test, some differences were found due to the rejection of the null hypothesis. A post-hoc analysis was performed to identify specific pairs of continents with significantly different freedom scores.

Tukey's HSD Results:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Africa	Asia	0.071	0.0	0.0302	0.1118	True
Africa	Australia	0.245	0.0272	0.0166	0.4735	True
Africa	Europe	0.0314	0.2559	-0.0101	0.0729	False
Africa	North America	0.1482	0.0015	0.0393	0.2571	True
Africa	South America	0.1052	0.0	0.0588	0.1516	True
Asia	Australia	0.174	0.2489	-0.0542	0.4023	False
Asia	Europe	-0.0396	0.0605	-0.0802	0.001	False
Asia	North America	0.0772	0.3252	-0.0314	0.1858	False
Asia	South America	0.0342	0.2666	-0.0114	0.0798	False
Australia	Europe	-0.2136	0.0819	-0.442	0.0147	False
Australia	North America	-0.0968	0.8778	-0.3464	0.1528	False
Australia	South America	-0.1398	0.5039	-0.3691	0.0895	False
Europe	North America	0.1168	0.0271	0.008	0.2256	True
Europe	South America	0.0738	0.0001	0.0276	0.12	True
North America	South America	-0.043	0.8775	-0.1538	0.0678	False

Figure 15=Tukey's HSD Results Table

The results in this table show a significant difference between the African continents and Asia. The low mean value and low p value confirm that there is a significant difference between these continents. A significant difference is also shown between Africa and Australia with a low p value. There is a slight mean difference between Africa and Europe, but a p value greater than 0.05 indicates no difference. It indicates that there is a significant difference between Africa and North America. Africa and South America show significant differences with lower

p-values. A p- value greater than 0.05 indicates no significant difference between Asia and Australia.

There is a slight mean difference between Asia and Europe but the p value does not suggest a significant difference. Also, none of these comparisons show significant differences between Asia and North America, Asia and South America, Australia and South America, North America and South America. This is because the p-value is significantly higher. Also, there are significant differences in freedom scores between Europe and North America, Europe and South America. There is no significant difference between North America and South America as the p-value is greater than 0.05. These relationships provide insight into how independence is marked across continents. This allows for further study of factors influencing these disparities.

4.7 A study of the relationship between generosity and health with GDP per capita

This analysis explores the interrelationships between per capita local production and health and generosity affecting society. Among these factors, we first analyze the correlation between per capita local production and health. It is important to explore the relationship between these factors and identify patterns between them using statistical indicators. A positive relationship or a negative relationship influences certain decisions.

GDP per capita and Health

Pearson Correlation Coefficient: 0.7627193506935926
P-value: 5.709614849475764e-132

Here the Pearson correlation coefficient is approximately 0.763 which indicates a strong positive linear correlation between gross domestic product and health. This shows a tendency for health-related indicators to increase as per capita local production increases. This further highlight that economic well-being and health outcomes are closely linked, with a Pearson correlation coefficient of 0.763. In correlation studies, a value of 0 indicates no correlation and 1 indicates a perfect positive correlation. So, in this analysis the coefficient of 0.763 is close which further highlights the strength of the relationship. A scatter plot is used as a methos of visualization to further clarify this information.

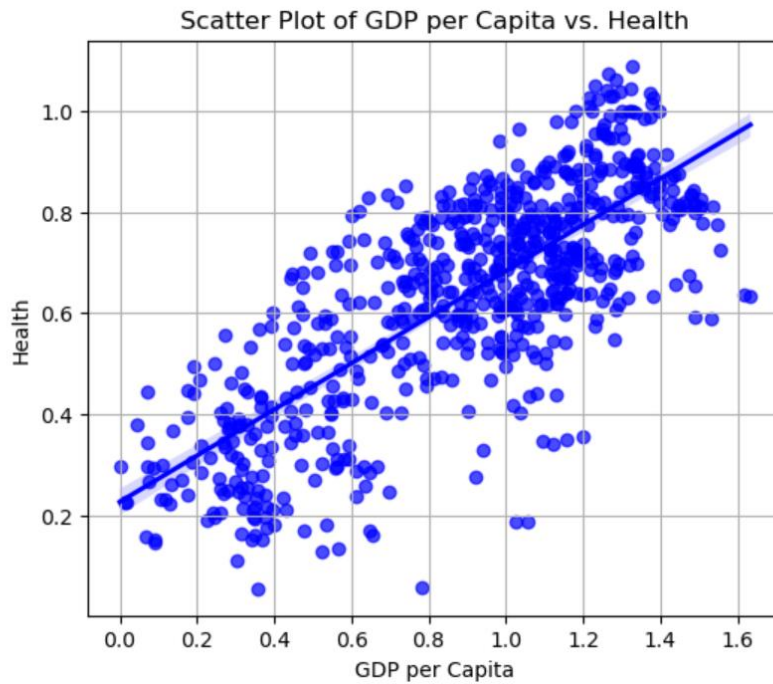


Figure 16=GDP per capita and Health Scatter Plot

Also, in this study a very low p - value of $5.71e-132$ has been obtained. It further explains the stability of the numerical value obtained. This low p-value indicates that the probability of obtaining such a strong association by chance is extremely remote. Therefore, this correlation rejects the null hypothesis that there is a positive relationship between per capita local production and health.

GDP per capita and Generosity

Pearson Correlation Coefficient: -0.13601292016206507
P-value: 0.0003465818985900956

A negative correlation coefficient of -0.136 in this test indicates a weak inverse relationship between GDP per capita and generosity. This negative sign means that generosity decreases as the economic status of the society increases. This Pearson correlation value approaching zero means that the observed correlation is small and is not the only determinant of generosity. A scatter plot is used to further study this situation.

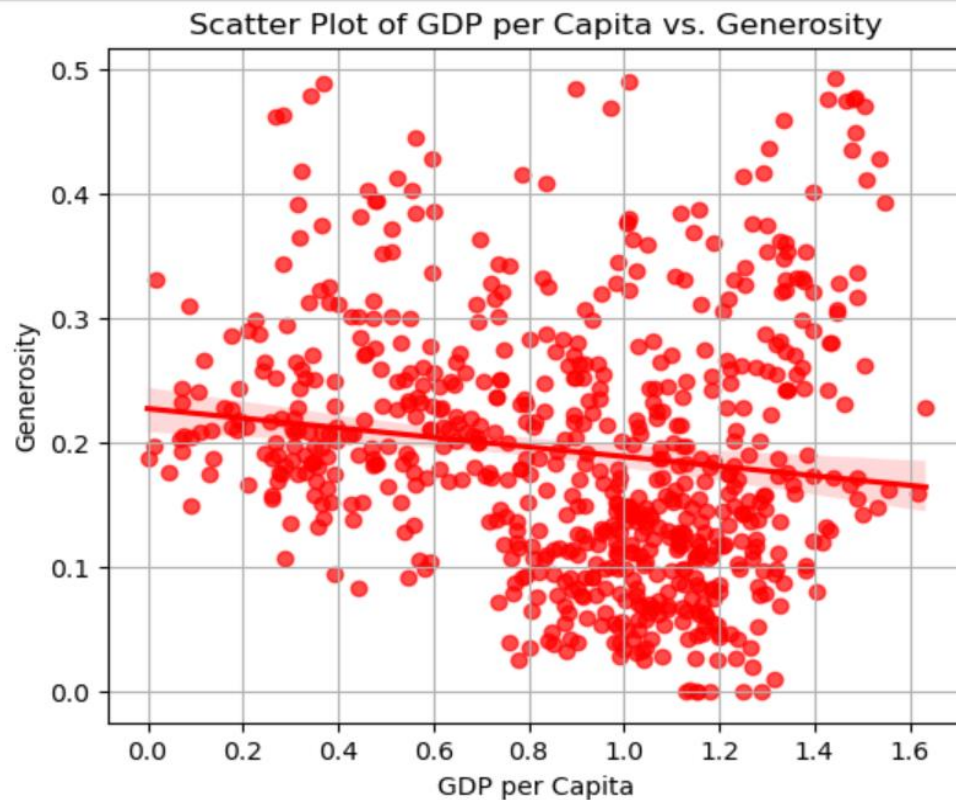


Figure 17=GDP per capita and Generosity Scatter Plot

The low p -value (0.00034) shown by these results is generally less than 0.05 and confirms the statistical significance of the correlation. This confirms that this negative value is not the result of chance. This negative correlation may suggest that wealthier countries tend to be less generous in general. But here it cannot be fully analyzed using only these economic factors.

CHAPTER 5

Discussion and Recommendations

5.1 Discussion

From 2015 to 2020, this research was conducted to find the factors affecting happiness and corruption in a country and to analyze those factors globally. The main purpose of this research is to identify the main factors that affect the happiness of a country and at the same time to identify the effect of those factors on corruption as well as to identify the patterns among those factors. In this research, graphic designer methods were used to observe how this happiness flows between countries and continents. There the factors affecting were observed globally. Also, it was confirmed that there is a relationship between the factors of government trust, family environment, freedom, generosity, health and per capita domestic production for the happiness score which is the main element in the research. Also, according to the way freedom flows between continents, some knowledge of happiness scores can be gained. In doing that test, it was confirmed that there is a variation in freedom scores between the continents. Also, post-hoc test was used to understand the differences in global freedom and identify specific continents with differences. Also, in the further investigation, the per capita GDP was studied. There, the relationship between health and generosity was studied. It shows a good correlation between GDP per capita and health, while showing a slightly weaker correlation with generosity.

It is important for analysts to identify factors that affect happiness scores. There, factors such as health, various economic indicators and the credibility of the government, which mainly contribute to the well-being of the society, are important. In further examination of these factors, the development of infrastructure in the health sector, the promotion of public health sectors, the creation of an economic situation with social protection and a political atmosphere

can be focused on the factors that affect the increase in happiness scores. And here, more than immediate solutions are expected for social security in the future, and creating a suitable environment for human beings to live in. Here, in comparative analysis between freedoms and continents, significant differences and reasonable solutions to identify those differences form the basis for policy planning that builds a geo-global perspective. Here, various disparities between the continents are recognized and efforts are made to improve this freedom globally. Also, considering the positive relationship between GDP per capita and health, it examines how economic development in developed countries affects this. Also, considering the negative correlation between GDP per capita and generosity, it examines the sociocultural factors that influence this in rich countries. While there are obvious reasons for low levels of generosity in countries with the lowest economic levels, the socio-cultural and ethical dimensions that contribute to this pattern need to be carefully scrutinized. In this way, global leaders can develop various strategies to increase the factors of freedom and generosity.

It is important to pay attention to and understand social dynamics in order to explore certain directions in the future. It is important to identify the mechanism between the identified factors and the factors involved. Also, it is necessary to do a special analysis between continents to analyze the social, political and economic factors that cause differences between freedoms between each continent. Different strategies to promote freedom come from this understanding and it is important to identify challenges and opportunities in different regions. Generosity and health in a country does not depend on the economic level of a country. Cultural, social and policy-related factors can be explored to gain insight into various patterns and relationships. Finally, this study examines changes in happiness scores, levels of freedom, economic levels, and health levels, and thus is important for researchers to identify future trends. Through this research, it is important to identify the complex interrelationships of future economic level, happiness and various socio-cultural aspects.

5.2 Recommendation

This study has been conducted with the primary objective of analyzing the main factors that affect the decrease or increase in happiness in a country. The study showed a strong relationship between trust in government, health, GDP per capita and freedom. Also analyzed how freedom flows globally. For that, considering it separated into continents, it shows that there are differences in freedom between the continents. Finally, the relationship between GDP per capita and health and between GDP per capita and generosity was studied.

To improve happiness i.e. to increase happiness scores governments should focus on powerful factors such as health, government trust and economic indicators. Governments should adopt excellent approaches. Cultural elements should be accepted for various criteria of happiness and intervention is required for maximum effectiveness. Improving economic policies for global peace and emphasizing social responsibility can create a more just society. Strong public health strategies and educational programs on happiness are important to create an informed population. And improving cooperation and measuring happiness can help create a harmonious and sustainable global community.

5.3 Conclusion

This research is mainly aimed at analyzing the happiness score of a country from the year 2015 to the year 2020. Also, it was considered how freedom flows and how it changes in continents. Finally, we consider the relationship between health and generosity for GDP per capita.

First, the multiple regression model performed here was performed. There, mainly the factors affecting the happiness score were analyzed. Factors such as health, government trust, freedom, generosity and GDP per capita were found to influence happiness scores. And a multiple linear

regression plot with predicted and actual values was used to further show that this was true. where the model predictions agree well with the actual results. Research uses global freedom to analyze this freedom between continents. Here, the ANOVA test was used to find the continents showing differences between patterns of freedom. Correlation test was used to consider the final objective of the research. There is a positive relationship between GDP per capita and health. That is, when the gross domestic product of man increases, the quality of health in the country increases. But considering per capita GDP and generosity, a negative relationship was found. It means that as the GDP per capita increases, the generosity of man in the country decreases. Finally, in addition to these factors, social, cultural and political influence should also be taken into account in the research.

Appendices

Importing necessary python libraries

```
# import all necessary libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import f_oneway
from statsmodels.stats.multicomp import pairwise_tukeyhsd
from scipy.stats import pearsonr
import plotly.express as px
import statsmodels.api as sm
```

Import CSV file

```
# import a dataset
df=pd.read_csv("WorldHappiness_Corruption_2015_2020.csv")
df
```

Data Preprocessing

```
# check outliers
plt.figure(figsize=(20, 5))
sns.boxplot(data=df, orient='h', palette='Set2')
plt.xscale('log')
plt.title('Boxplots of All Columns')
plt.show()
```

```
# remove outliers with IQR
def remove_outliers(df):
    # Filter only numeric columns
    numeric_cols = df.select_dtypes(include=[np.number]).columns
    numeric_data = df[numeric_cols]

    # Calculate the first and third quartiles (Q1 and Q3)
    Q1 = numeric_data.quantile(0.25)
    Q3 = numeric_data.quantile(0.75)

    # Calculate the interquartile range (IQR)
    IQR = Q3 - Q1

    # Define the Lower and upper bounds for outliers
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR

    outliers = ((numeric_data < lower_bound) | (numeric_data > upper_bound))
    filtered_df = df[~outliers.any(axis=1)]

    return filtered_df
```



```

# new dataset view
filtered_df = remove_outliers(df)

print("Original DataFrame:")
print(df)
print("\nFiltered DataFrame (without outliers):")
print(filtered_df)

# get new dataset boxplot
plt.figure(figsize=(20, 5))
sns.boxplot(data=filtered_df, orient='h', palette='Set2')
plt.xscale('log')
plt.title('Boxplots of All Columns')
plt.show()

# columns names
filtered_df.columns

# check duplicate values
duplicates = filtered_df.duplicated().sum()
duplicates

# check non - values
print(filtered_df.info())
print(filtered_df.columns)

```

Descriptive Data Analysis

```

# exploratory data analysis for selected fields
filtered_df.describe()

```

Exploratory Data Analysis

```

#Happiness score by Continent
plt.figure(figsize=(6,4))
sns.barplot(data=filtered_df, y='continent', x='happiness_score')
plt.title('Happiness Score by Continent')

#Happiness score by Year
sns.lineplot(data=filtered_df.reset_index(), x='Year', y='happiness_score', err_style=None)
plt.show()

# Find happiest country
df_happiest = filtered_df[(filtered_df["happiness_score"] > 7.0) & (filtered_df["Year"] <= 2020)].copy()
df_happiest.sort_values(by='happiness_score', ascending=False, inplace=True)
sns.barplot(x="happiness_score", y="Country", data=df_happiest, palette="coolwarm")
plt.title("Happiest countries between 2015 and 2020")
plt.xlabel("Happiness Score")
plt.ylabel("Country")
plt.show()

```

```
# Find unhappiest country
df_unhappiest = filtered_df[(filtered_df["happiness_score"] < 3.5) & (filtered_df["Year"] <= 2020)].copy()
df_unhappiest.sort_values(by='happiness_score', ascending=False, inplace=True)
sns.barplot(x="happiness_score", y="Country", data=df_unhappiest, palette="coolwarm")
plt.title("Unhappiest countries between 2015 and 2020")
plt.xlabel("Happiness Score")
plt.ylabel("Country")
plt.show()
```

```
# find continet contribution to world economy
fig = px.pie(filtered_df, values='gdp_per_capita', names='continent', title='Continent Wise Contribution to World Economy', height=650)
fig.show()
```

```
# correlation heat map
columns_list = ['happiness_score', 'gdp_per_capita', 'family', 'health', 'freedom', 'generosity', 'government_trust', 'dystopia_residual']

corr_matrix = filtered_df[columns_list].corr()

# Create a heatmap with a red colormap
sns.set(style="white") # Optional, sets the style
plt.figure(figsize=(7, 4))
sns.heatmap(corr_matrix, annot=True, cmap="Reds", linewidths=.5)
plt.show()
```

```
# find best helath rate country
fig = px.pie(filtered_df.nlargest(6, 'health'), values='health', names='Country', title='Countries With The Highest Health Rates',
             color_discrete_sequence=px.colors.sequential.RdBu)
fig.show()
```

Multiple Linear Regression

```
# happiness score and freedom scatter with regression line
sns.scatterplot(x='freedom', y='happiness_score', data=filtered_df)
sns.regplot(x='freedom', y='happiness_score', data=filtered_df, scatter=False, color='red')

plt.title('Scatter Plot with Regression Line')
plt.xlabel('freedom')
plt.ylabel('Happiness Score')
plt.show()

# happiness score and health scatter with regression line
sns.scatterplot(x='health', y='happiness_score', data=filtered_df)
sns.regplot(x='health', y='happiness_score', data=filtered_df, scatter=False, color='red')

plt.title('Scatter Plot with Regression Line')
plt.xlabel('Health')
plt.ylabel('Happiness Score')
plt.show()
```

```

# happiness score and gdp per capita scatter with regression line
sns.scatterplot(x='gdp_per_capita', y='happiness_score', data=filtered_df)
sns.regplot(x='gdp_per_capita', y='happiness_score', data=filtered_df, scatter=False, color='red')

plt.title('Scatter Plot with Regression Line')
plt.xlabel('gdp_per_capita')
plt.ylabel('Happiness Score')
plt.show()

# happiness score and generosity scatter with regression line
sns.scatterplot(x='generosity', y='happiness_score', data=filtered_df)
sns.regplot(x='generosity', y='happiness_score', data=filtered_df, scatter=False, color='red')

plt.title('Scatter Plot with Regression Line')
plt.xlabel('generosity')
plt.ylabel('Happiness Score')
plt.show()

# multiple linear regression result
X = np.column_stack((filtered_df['health'], filtered_df['government_trust'], filtered_df['family'],
                    filtered_df['gdp_per_capita'], filtered_df['generosity'], filtered_df['freedom']))
X_with_intercept = sm.add_constant(X)

# Create the OLS model using the formula
formula = 'happiness_score ~ health + government_trust + family + gdp_per_capita + generosity + freedom'
model = sm.OLS.from_formula(formula, data=filtered_df)

results = model.fit()

print(results.summary())

# actual and predicted happiness score scatter plot
X = np.column_stack((filtered_df['health'], filtered_df['government_trust'],
                    filtered_df['family'], filtered_df['gdp_per_capita'], filtered_df['generosity'], filtered_df['freedom']))
X_with_intercept = sm.add_constant(X)
model = sm.OLS(filtered_df['happiness_score'], X_with_intercept)
results = model.fit()
predicted_values = results.predict(X_with_intercept)

# Scatter plot with regression line
plt.figure(figsize=(10, 4))
sns.regplot(x=filtered_df['happiness_score'], y=predicted_values)
plt.title('Actual vs Predicted Happiness Score')
plt.xlabel('Actual Happiness Score')
plt.ylabel('Predicted Happiness Score')
plt.show()

```

ANOVA Test

```

# continent and freedom anova result
anova_result = f_oneway(
    filtered_df[filtered_df['continent'] == 'Asia']['freedom'],
    filtered_df[filtered_df['continent'] == 'Europe']['freedom'],
    filtered_df[filtered_df['continent'] == 'Africa']['freedom'],
    filtered_df[filtered_df['continent'] == 'North America']['freedom'],
    filtered_df[filtered_df['continent'] == 'South America']['freedom'],
    filtered_df[filtered_df['continent'] == 'Australia']['freedom'])
print("ANOVA Results:")
print("F-statistic:", anova_result.statistic)
print("P-value:", anova_result.pvalue)

```

```
# one way anova result
tukey_result = pairwise_tukeyhsd(filtered_df['freedom'], filtered_df['continent'])
print("\nTukey's HSD Results:")
print(tukey_result)
```

Correlation Analysis

```
# gdp per capita and health correlation value
correlation_coefficient, p_value = pearsonr(filtered_df['gdp_per_capita'], filtered_df['health'])
print("Pearson Correlation Coefficient:", correlation_coefficient)
print("P-value:", p_value)
```

```
# visualize with scatter plot
plt.figure(figsize=(6, 5))
sns.regplot(x='gdp_per_capita', y='health', data=filtered_df, color='blue', scatter_kws={'alpha':0.7})
plt.title('Scatter Plot of GDP per Capita vs. Health')
plt.xlabel('GDP per Capita')
plt.ylabel('Health')
plt.grid(True)
plt.show()
```

```
# gdp per capita and generosity correlation value
correlation_coefficient, p_value = pearsonr(filtered_df['gdp_per_capita'], filtered_df['generosity'])
print("Pearson Correlation Coefficient:", correlation_coefficient)
print("P-value:", p_value)
```

```
# visualize with scatter plot
plt.figure(figsize=(6,5))
sns.regplot(x='gdp_per_capita', y='generosity', data=filtered_df, color='red', scatter_kws={'alpha':0.7})
plt.title('Scatter Plot of GDP per Capita vs. Generosity')
plt.xlabel('GDP per Capita')
plt.ylabel('Generosity')
plt.grid(True)
plt.show()
```

References

Bo Rothstein, 2010, *Corruption, Happiness, Social Trust and the Welfare State: A Causal Mechanisms Approach*

[Corruption, Happiness, Social Trust and the Welfare State: A Causal Mechanisms Approach \(gu.se\)](#)

Terry Anderson, *A Global Examination of the Relationship Between Corruption and Well-Being and Happiness*

[A Global Examination of the Relationship Between Corruption and Well-Being and Happiness | Journal of Politics and Democratization \(gipa.ge\)](#)

Elma Satrovic , Ozge Cetiner , Adnan Muslija, 2018, *WHETHER CONTROL OF CORRUPTION MATTERS FOR HAPPINESS: EVIDENCE FROM PANEL DATA ANALYSIS*

<https://doi.org/10.17261/Pressacademia.2018.921>

Simeon Djankov , Elena Nikolova , Jan Zilinsky , 2016, *The happiness gap in Eastern Europe*
[The happiness gap in Eastern Europe - ScienceDirect](#)

Berke Akkaya, 2022, *The Mediator Role of Economic Freedom in the Effect of Corruption Perception on National Happiness: A Case of World Countries*

<https://doi.org/10.24988/ije.962820>

Carol Graham, 2010, *Adaptation amidst Prosperity and Adversity: Insights from Happiness Studies from around the World*

[Adaptation amidst Prosperity and Adversity: Insights from Happiness Studies from around the World | The World Bank Research Observer | Oxford Academic \(oup.com\)](#)

B. Mak ARVIN Byron LEW, 2012, *Development Aid, Corruption, and the Happiness of Nations: Analysis of 118 countries over the years 1996-2009*, Mak Arvin

[Development Aid, Corruption, and the Happiness of Nations: Analysis of 118 countries over the years 1996-2009 \(repec.org\)](#)

J. C. Ott, 2010, *Government and Happiness in 130 Nations: Good Governance Fosters Higher Level and More Equality of Happiness*

[Government and Happiness in 130 Nations: Good Governance Fosters Higher Level and More Equality of Happiness | Social Indicators Research \(springer.com\)](#)