TEAM RESEARCH AND DEVELOPMENT PROJECT

TEAM NO: NEW\_GROUP 5

TITLE: REPORT OF WORLD HAPPINESS REPORT 2020 ANALYSIS

Under the guidance of,

DR. JOHN NOLL

BY,

HARSHA VADUKKOOT PROCESS

RIDHIT BHASIN

VARSHIL SHUKLA

Z MARIYA

HIMANSHU PATIDAR

**Date: January 8, 2021**

**TABLE OF CONTENTS**

|  |  |
| --- | --- |
| TABLE OF CONTENTS | 2 |
| LIST OF FIGURES | 2 |
| 1. Introduction | 3 |
| 1. Background | 4 |
| 1. Methods | 5 |
| 1. Result | 10 |
| 1. Discussion/Implications | 10 |
| 1. Conclusion | 11 |
| 1. Reference | 12 |

LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| **Fig. No.** | **Figure Name** | **Page No.** |
| FIGURE 3.1 | Test1.png file | 6 |
| FIGURE 3.2 | Test2.png file | 7 |
| FIGURE 3.3 | Plot2.png file | 7 |
| FIGURE 3.4 | Plot1.png file | 8 |

**INTRODUCTION**

Being happy is one of the most basic requirements of a living being. For the sole purpose of attaining happiness, humans have also been trying to establish new technologies, tools and their lifestyle since the beginning of human civilization. Happiness is not touchable entity and that cannot be found in luxuries and richness. The purpose behind the functioning of nations and societies, in general is happiness. Happy people are found to be more effective in their lives and appear to pass on positive vibes to the individuals around them leading us to pick the dataset of World Happiness Report 2020.

It's fascinating that happiness is a non-materialistic thing but it is also one of the essential necessities of life for individuals. While we are working on the World Happiness Report 2020 dataset, a nation's happiness index depends on many factors, such as-

**Ladder score** - The status of individuals lives on a ladder scale ranging from 0 to 10

**Logged GDP per capita** - How much the value of economic production can be attributed to each individual citizen.

**Dystopic** - A world in which misery or injustice exists.

**Healthy Life expectancy** - It is the average life in good health.

**Freedom to make life choices** - Do the individual person have to freedom to make their own choices.

**Generosity** - The virtue of being kind to any person and treat everyone fairly.

**Perceptions of corruption**- Fraudulent actions by the individual who are in power which usually includes bribery.

It is important to understand what the correlation is? before heading to the research question. It is a relationship or association between two or more items which is reciprocal. Our research question is - Is there any correlation between Ladder score and logged GDP per capita? For the research question, the Null hypothesis is that there is no correlation between ladder score and per capita GDP recorded, while the alternative hypothesis is that there is a correlation between ladder score and per capita GDP recorded. Independent variable is a ladder score in our research question and GDP per capita is identified as a dependent variable. There is a wide range of questions that can be posed and the insights gained for study. As stated earlier, we opted for the reported logged GDP per capita and ladder score for our research question from the dataset. With the aid of data analysis tools such as R and its functions such as data visualization and correlation, this literature aims to find insights from World Happiness Data. We attempted data visualization in this literature to figure out our aim for any correlation variables using the plotting of data using two separate types of plots, such as scatterplot and bar plot. Pearson r correlation, Kendall rank correlation, and spearman correlation are 3 techniques to find the correlation coefficient. We chose the Pearson r method to find the correlation coefficient among all three methods.

**BACKGROUND**

As we are working on the dataset of the World Happiness Report-2020 [Mathurin Aché, 2020], There are a plethora of questions that can be raised for analysis and get the insights. From the dataset, we have opted for the logged GDP per capita and Ladder score for our research question as mentioned earlier. This literature aims to focus on finding insights from the world happiness data with the help of data analysis tools like R and its functions such as data visualization and correlation coefficient. On the topic of ours, we have read lots of research papers, articles as well as the Kaggle site to find out the work related to our research question.

The method used by us for getting insights and claims regarding the null hypothesis and the alternative hypothesis is that the same one we have observed from the papers we have gone through [denizardan, 2020] that is data visualization with scatter plot and stated the correlation result of which he has used. The same way [lordadriano, 2020] has used the machine learning model to find out the findings from the same data and got some insights after doing such activity on data. So, basically the approach here for getting the correlation coefficient between two entities which we have selected from the whole dataset and get to know about the state of correlation.

[A Service of zbw, n.d.] have tried to found the correlation between the happiness score and GDP per capita of the developing country. In this literature, we tried data visualization to find out our purpose of any correlation factors using the plotting the data with two different kinds of plots such as scatterplot and bar plot.

There is one more approach to more precise insights is to use the different correlation coefficient methods by which we can summarize for the research question. There are 3 methods of finding the correlation coefficient like Pearson r correlation, Kendall rank correlation and spearman correlation [Statistics Solutions, 2020].

We have tried to find out more on this question but there is a plethora of things we have put on the consideration to narrow down our searches and results. Therefore, what we have found related to our literature background is here with all the citations under the references section.

**METHOD**

The dataset “World Happiness Report 2020” has been downloaded from Kaggle. The dataset has some essential variables which are country, Regional indicator, ladder score, Standard error of ladder score, Upperwhisker, lowerwhisker, Logged GDP per capita, Social support, Healthy life expectancy, Freedom to make life choices, Generosity, Perceptions of corruption, Ladder score in Dystopia, explained by: Log GDP per capita, Explained by: Social support, explained by: Healthy life expectancy, explained by: Freedom to make life choices, explained by: Generosity, Explained by: Perceptions of corruption, Dystopia + residual.

In Rscript we have imported three libraries

* library(ggplot2)
* library(dplyr)
* library(gridExtra)

Then the CSV file of the dataset was imported into a variable known as “whr” in which we have created three png files:

* Test1.png: In which the entire dataset is included without any modification.

**Code in R:**

whr<- read.csv("World Happiness report 2020.csv")

df0<-head(whr)

png("test.png", height = 50\*nrow(df0), width = 200\*ncol(df0))

grid.table(df0)

* Test2.png: the aim of this file was to showcase the cleaned data saved in variable “whr\_data”

**Code in R (preprocessing)**

In which we simplified the names of the column in “whr” dataset.

names(whr)[1]<-"country"

names(whr)[2]<-"region"

names(whr)[3]<-"ladder\_score"

names(whr)[7]<-"GDP"

Then removed the unused column

whr\_data <- whr %>% select(-lowerwhisker,-upperwhisker,

-Social.support,

-Explained.by..Log.GDP.per.capita,

-Explained.by..Social.support,

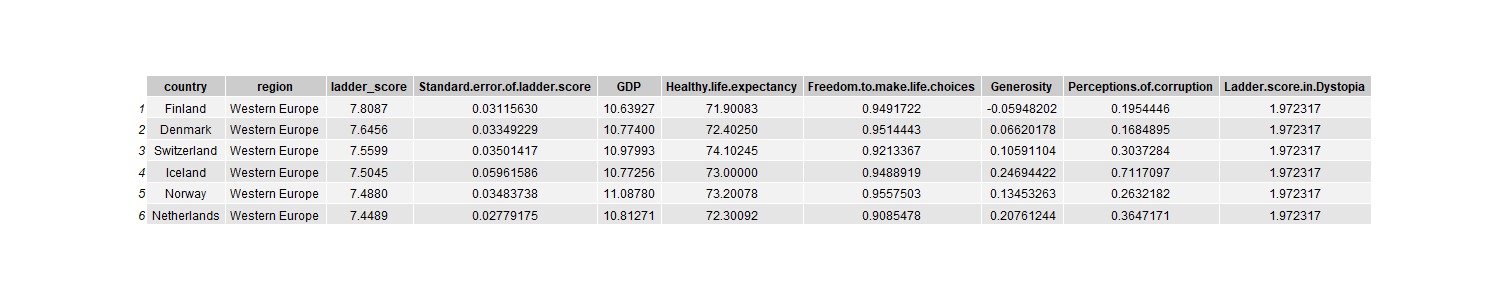
-Explained.by..Healthy.life.expectancy,

-Explained.by..Freedom.to.make.life.choices,

-Explained.by..Generosity,

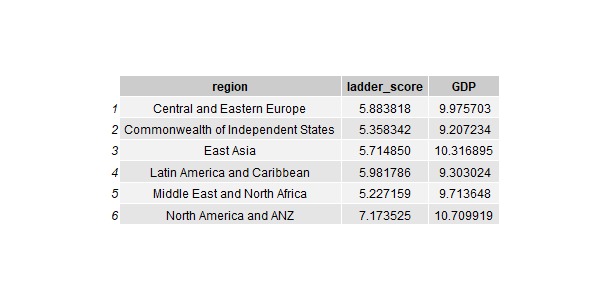
-Explained.by..Perceptions.of.corruption,

-Dystopia...residual)

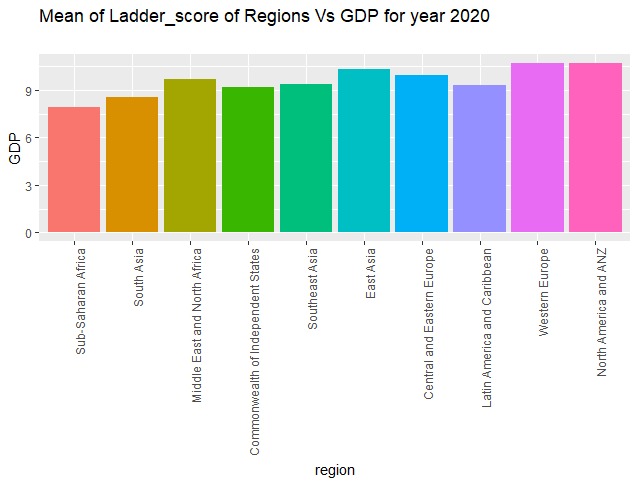


3.1: Test1.png

This data shows the report of the happiness rate of the year 2020 of all the countries in which we mainly focus on the region, ladder\_score and the GDP. In the below graph we have calculated the mean of GDP and Ladder\_score of all the countries which lies in the respective region. After that we have arranged all the region in the ascending order and then plotted accordingly, we have seen the positive correlation between the ladder score and GDP of the regions as the dependent variable GDP increases, the Ladder\_score also increases in most of the cases.

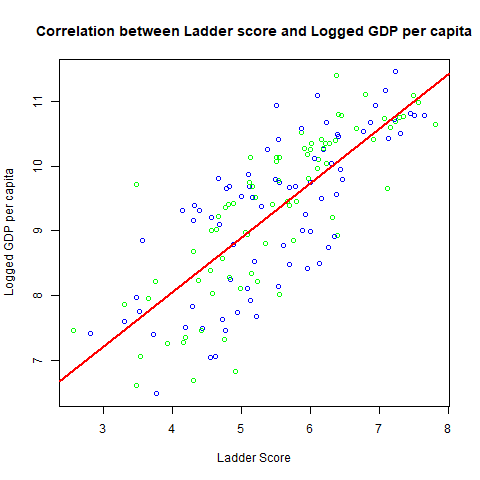


Test2.png



Plot2.png file

* We have plotted a scatter plot with a linear regression line in it to determine the correlation between the ladder\_Score and GDP. In this graph also we have seen a positive correlation between ladder\_Score and GDP.



Plot1.png file

- Analysis using the Pearson method

**Code in R**

whr<- read.csv("World Happiness report 2020.csv")

#Using Pearson's r method to check the correlation

x<- as.numeric(whr$Logged.GDP.per.capita)

y<-as.numeric(whr$Ladder.score)

cor(x[!is.na(x)],y[!is.na(x)],method = "pearson")

#Value is 0.7753744

#To found of the slope and the intercept

lm(formula = y[!is.na(x)]~x[!is.na(x)])

# Performing the test to find out the p-value

cor.test(x[!is.na(x)], y[!is.na(x)], method="pearson")

#p-value < 2.2e-16 which is less than 0.05 so we can reject

#null hypothesis and process towards the alternative hypothesis

The aim of the research question is to find out the correlation between the ladder\_Score and the logged GDP per capita in year 2020. To find out the correlation there are three methods which we can use

* Pearson’s r
* Kendall’s T
* Spearman’s p

In which we have used Pearson’s r because our data is normally distributed, then we have found out the correlation coefficient that is (“0.7753744”) which showcase that it has a positive correlation but with that, we cannot reject our null hypothesis, that’s the reason we have performed correlation test (“cor.test(x[!is.na(x)], y[!is.na(x)], method="pearson")”). The result which we got from this was the data points and the P-value which is 2.2e-16. as we know if the p-value is less than 0.05 then only, we can reject our null hypothesis and we can proceed towards the alternative hypothesis.

**Result**

* Western Europe and north America have the highest happiness rate.
* Correlation coefficient is 0.7753744
* P-value is 2.2e-16
* There is a strong correlation between GDP and ladder\_score

**Discussions/Implications**

During the analysis we conducted the Pearson correlation test and discovered that p-value is less than 0.005, hence we reject the null hypothesis(H0) in favour of the alternative hypothesis(H1) and since the correlation coefficient is greater than 0 it implies that alternative hypothesis (H1) is true and there is a positive correlation between Ladder score and logged per capita.

**Conclusion**

To summarize the report, here we would like mentioned that all the work we have done so far is genuinely authentic and we have attached the references at last to give the credit to our reference and aims to justify our content.

At the last, the data collection for the World Happiness Study 2020 reveals that happiness depends on a number of variables. We tried to answer the question in this research whether there is a link between Ladder Score and Logged GDP per capita. We conducted the Pearson correlation test during the study and found that correlation coefficient is 0.7753744 and p-value is 2.2e-16, so we reject the null hypothesis(H0) in favor of the alternative hypothesis(H1) since the coefficient of correlation is greater than 0, it means that alternative(H1) hypothesis is valid and there is a positive correlation between ladder score and per capita GDP reported.

**References**

[1.] Mathurin Aché (2020). *World Happiness Report up to 2020*. [online] Kaggle.com. Available at: https://www.kaggle.com/mathurinache/world-happiness-report [Accessed 4 Jan. 2021].

[2.] lordadriano (2020). *intro-ml\_Python-linear-regression\_worcap2020*. [online] Kaggle.com. Available at: https://www.kaggle.com/lordadriano/intro-ml-python-linear-regression-worcap2020 [Accessed 3 Jan. 2021].

[3.] denizardan (2020). *World Happiness 2020*. [online] Kaggle.com. Available at: https://www.kaggle.com/denizardan/world-happiness-2020/notebook [Accessed 4 Jan. 2021].

[4.] *A Service of zbw. (n.d.). [online] Available at:* [*https://www.econstor.eu/bitstream/10419/51882/1/665228880.pdf*](https://www.econstor.eu/bitstream/10419/51882/1/665228880.pdf)*.*

[5.] Statistics Solutions. (2020). *Correlation (Pearson, Kendall, Spearman) - Statistics Solutions*. [online] Available at: https://www.statisticssolutions.com/correlation-pearson-kendall-spearman/ [Accessed 4 Jan. 2021].