Happiness and Unemployment

Nanda Kishore Vallamkondu[1], Sarvathanayan Muralidharan[2],
Shriya Reddy Pulagam[3]
[1,2,3] Indiana University Bloomington, USA

Introduction

The Happiness Index is a measure of well-being that ranks 156 countries by how happy their citizens perceive themselves to be. The index is based on a global survey that asks people to rate their own lives on a scale of 0 to 10. The index is used to understand how happiness varies across the world and to identify the factors that contribute to happiness. The World Happiness Index is published annually by the United Nations Sustainable Development Solutions Network and is an important tool for policymakers who seek to understand and improve the well-being of their citizens.

The global happiness index is calculated using factors such as life expectancy, social support, and freedom to make life choices. These factors would be appropriate for any regular year but since COVID-19 reshaped the lives of people all around the world, we might have to consider a broader variety of factors in explaining the Happiness Index. Unemployment is generally considered to be a negative factor that can contribute to lower levels of happiness and well-being. When people are unemployed, they may experience financial difficulties and a loss of purpose and social connections. This can lead to increased stress, anxiety, and depression, which can negatively impact overall happiness. However, the relationship between happiness and unemployment is complex and can vary depending on the specific circumstances of each person and country. In this project, we would like to explore the impact of the Unemployment Rate on the World Happiness Index as the employment of people was highly impacted during the pandemic. This project tries to answer, "What effect did the unemployment rate, annual GDP, and life expectancy rate have on world happiness during COVID-19?"

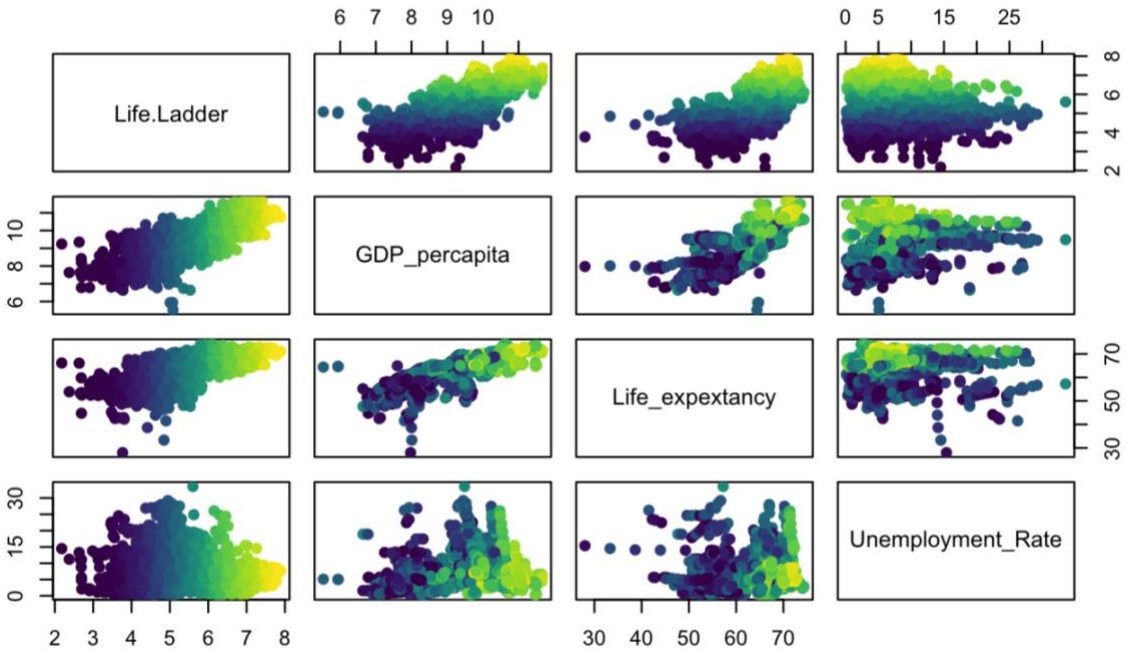
Data description and Correlation

The dataset has been taken from World Happiness Report, it consists of 2090 rows of data with the Happiness index and various factors considered in calculating the index from 2008-2021. We considered data from 2010 to 2021 as there are many missing and Null values before 2010.

We considered GDP and Life Expectancy from WHR data as two of the quantitative variables. Looking at individual columns the Happiness index ranges between 0 to 10. Life expectancy is the avg number of years a person lives, and it ranges from 28(Haiti) to 75(Japan). We have considered unemployment as our third quantitative variable, we took the unemployment data from the world bank, which provides free data about development in all the countries, and merged this data with world happiness data using the country name. Unemployment is calculated using the total percent of the labor force, it ranges from 0 to 34% in our data, Cambodia has the least unemployment rate of 0 and South Africa has the highest with 34%. As we know, the lower the unemployment rate the better. We considered this as our third variable because there was a drastic change in the world employment rate during the pandemic. Finally, we used continents data as most of our analysis will be based on continents since it will be easier to check the trends.

To observe the correlation between Happiness Index and three quantitative variables, we have plotted a pair plot. The points have based on the Happiness Index.

Fig-1: Correlation Plot of Happiness Score and dependent variables

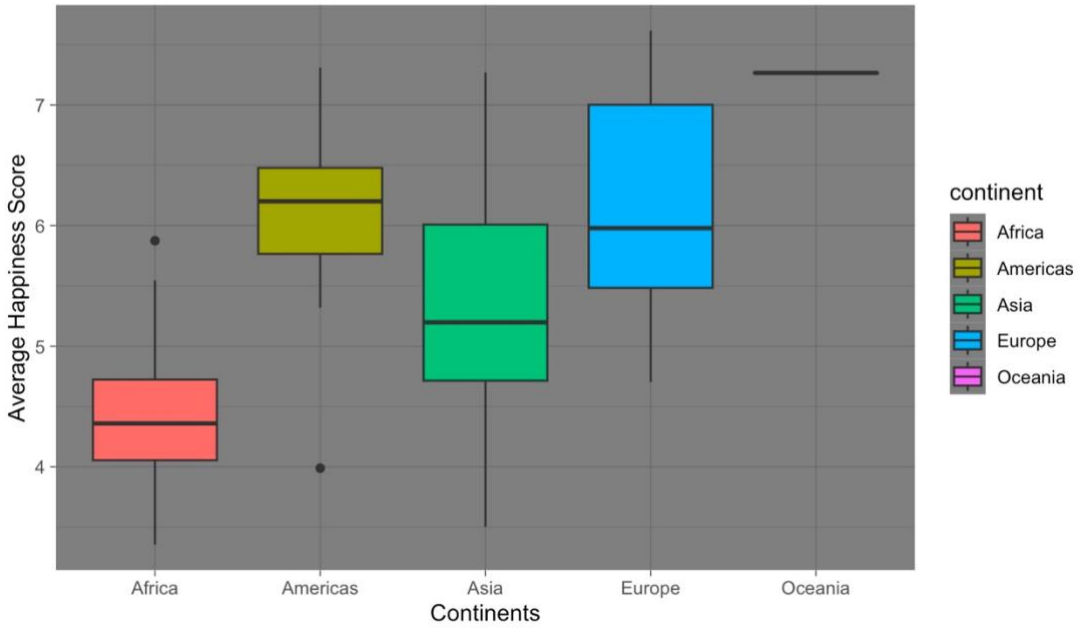


From the pair plot above, we can derive that Happiness Index has a high positive correlation with GDP per capita and Life Expectancy. The relationship appears to be a weaker negative between Happiness Index and Unemployment Rate.

We have also tried to see how the Happiness index varies by continent. For this analysis, we have plotted a box plot by taking mean values of happiness from 2010 to 2021 over countries.

From the box plot below, we can observe the median, ranges of the happiness index for each continent. It can be observed that Africa has the least Happiness Index with a minimum going as low as 3.35. The Americas has some of the highest Happiness indexes, the highest Happiness index corresponds to Canada in the Americas. Asia has the most interesting box plot out of all continents, as the range of Happiness goes from 3.5 to 7.2 with a median of around 5.2. Looking at Europe has the highest Happiness index with 7.61 for Denmark, and 7.60 for Finland. Now, coming to Oceania, it does not form a whisker because it has only two countries.

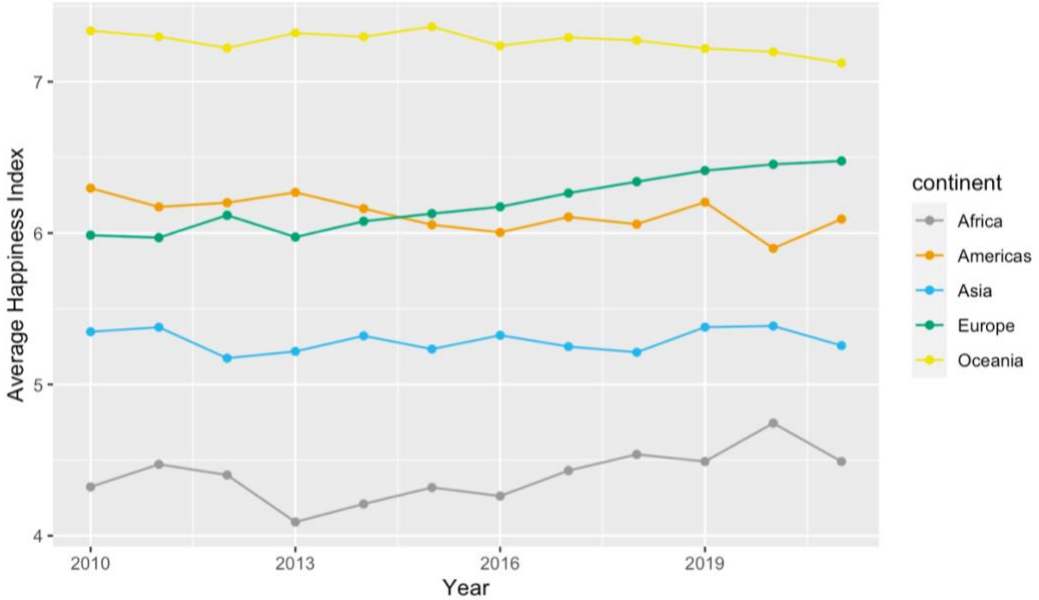
Fig – 2: Happiness over Continents



To understand the variations over years, we have also plotted the average Happiness index when grouped by continent for each year (Fig 2). For this analysis a time series graph has been used, time series graphs are particularly useful as they make trends easy to spot. These trends are important as they can be used to project future data.

The below graph depicts a multi-line time series, where each line represents a continent and its trends during the period 2010-2021.

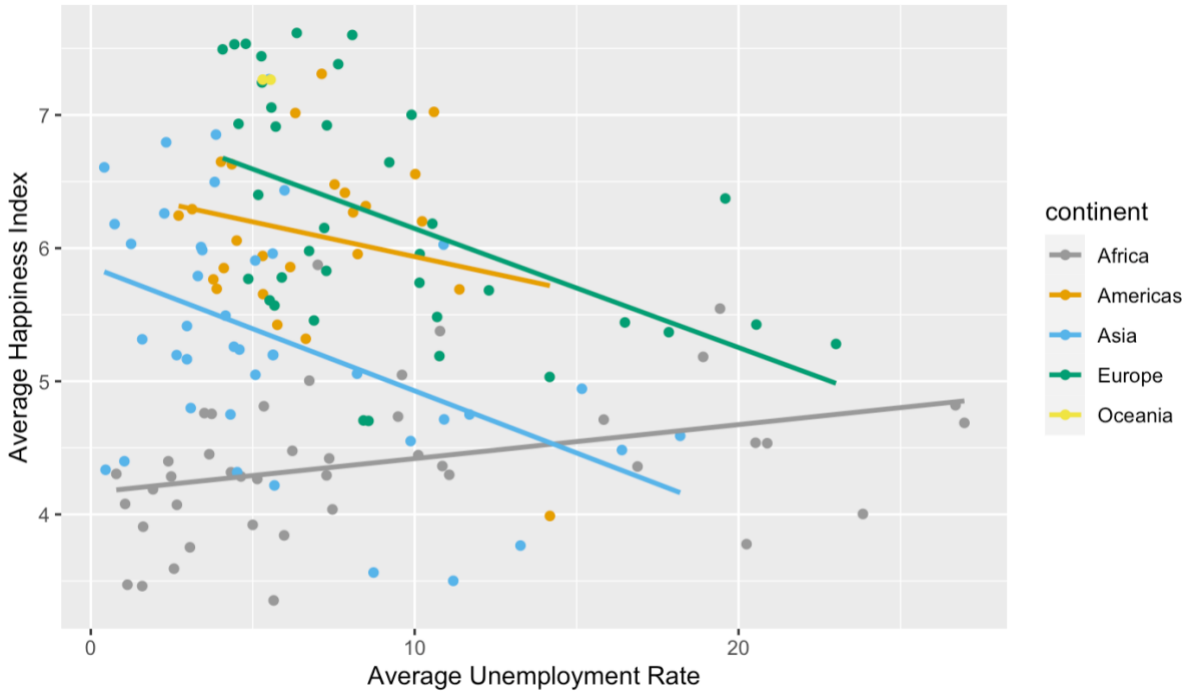
Fig -3: Change of Happiness over the years



The trends in each continent can be clearly observed from the above Fig - 3, Oceania consistently has a higher average Happiness index because there are only two countries Australia and New Zealand with more than 7 indexes every year. African countries are less happy every year compared to other continents. We can rank the continents by Happiness Index as Oceania, Europe, America, Asia, and Africa. One thing to note about this plot is Americas had higher happiness till 2015, but Europe took over the Americas in 2015 and had a higher happiness index throughout the years (excluding Oceania).

Since we know that Happiness Index has a high positive correlation with GDP per capita. Life Expectancy, and a weak negative relation with Unemployment. We want to further analyze how this relationship varies for individual continents.

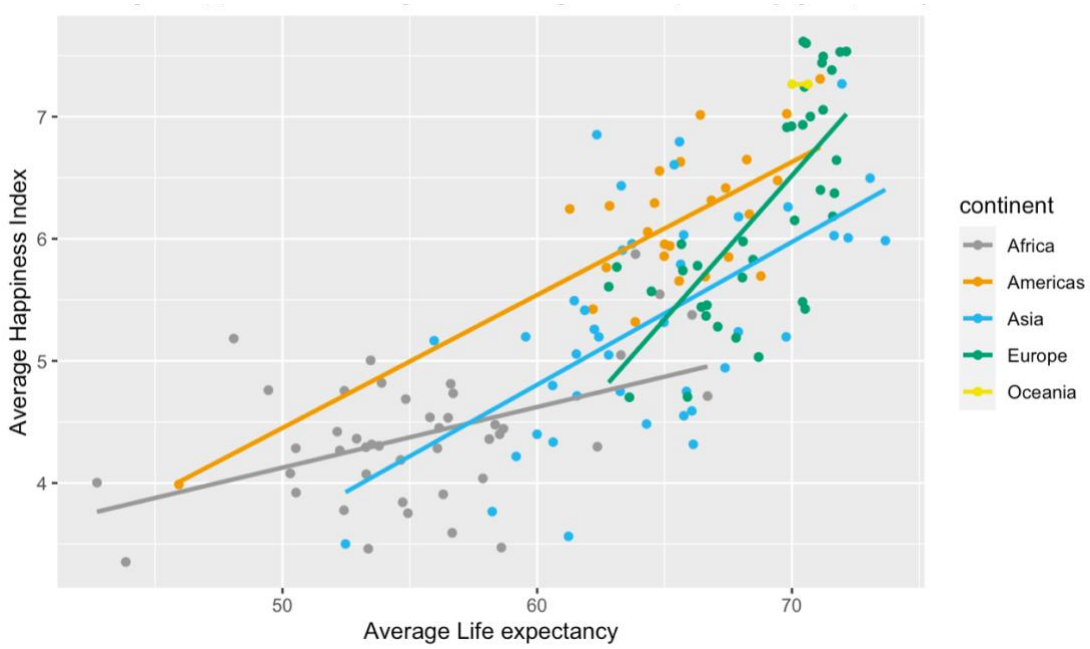
Fig – 4: Avg Happiness Index vs Avg Unemployment grouped by continents



It can be clearly observed from Fig – 4 that the Happiness index of European, Asian, and American countries decreases with increasing Happiness Index, but surprisingly African countries have a slight positive relation. As unemployment rate increases happiness also increases.

For life expectancy we can see from Fig 5, all the continents follow a positive trend with Happiness Index.

Fig – 5: Avg Happiness Index vs Avg Unemployment grouped by continents

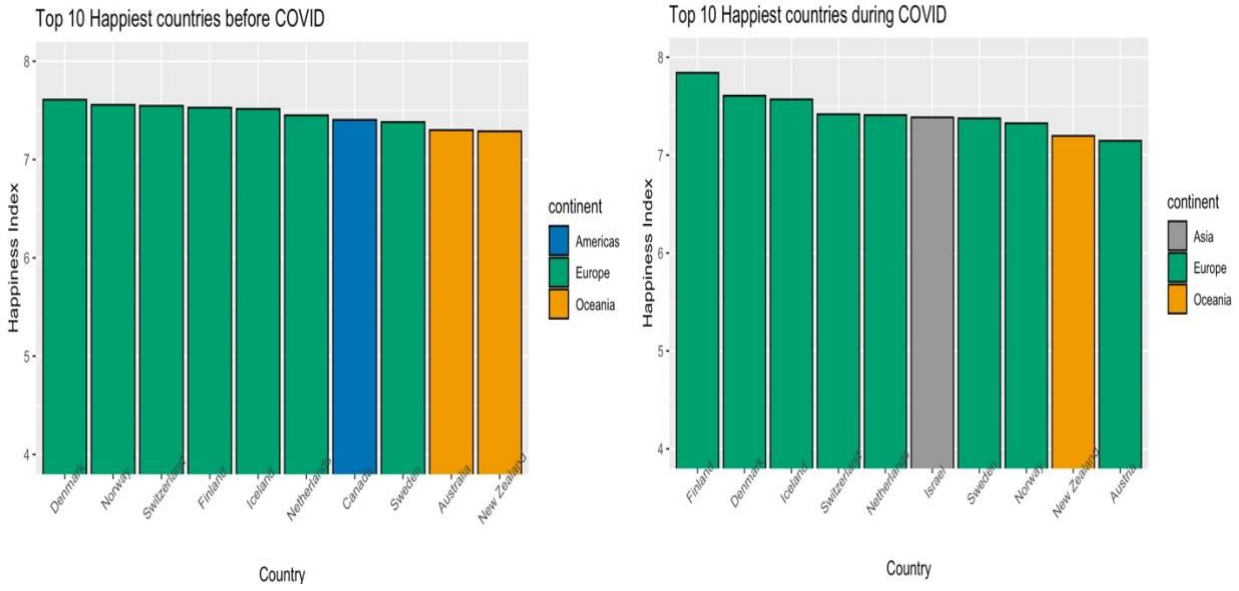


Happiness Index before and during COVID

To address our research question, we have split the data into two data frames. The data from 2010-2018 is considered as before COVID-19 and data from 2019-2021 is taken during COVID-19.

Now, we will focus on how the Happiness index changed before and during the pandemic. First, we have plotted bar graphs showing the difference between the top 10 happiest countries before and during Covid-19.

Fig - 6



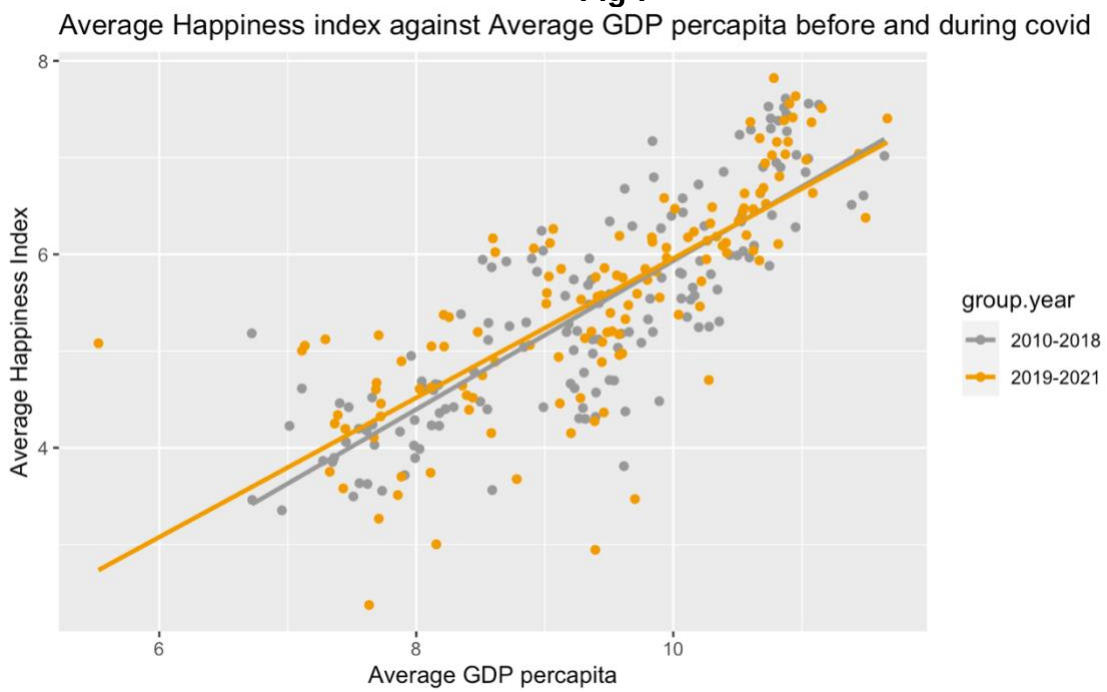
From Fig 6, we observe that there has been a slight variation in the top 10 happiest countries from before the pandemic. Finland has the highest Happiness index during COVID with 7.84, but Finland was in 4th place before the pandemic. Europe has dominated the list of the top 10 happiest countries both before and during COVID-19. Also, we can see that there are no countries from Africa in the top 10 list as observed from the whisker plot before. We can also observe that Canada's happiness index diminished during COVID as it is not present in the top 10 list, but Israel's Happiness has increased during the pandemic.

Change in Effect of Independent variables on Happiness Index

The research question is to identify how the effect of the independent variables on the Happiness index varies before and during covid. To analyze that we have used lm model on three independent variables (GDP per capita, Life expectancy, and Unemployment Rate) individually against our dependent variable which is our Happiness Index. From the intercept and coefficients, we observed that there is not much change before and during covid. This means that the effect covid has on the relationship between dependent and independent variables is negligible.

To visualize it, we plotted them.

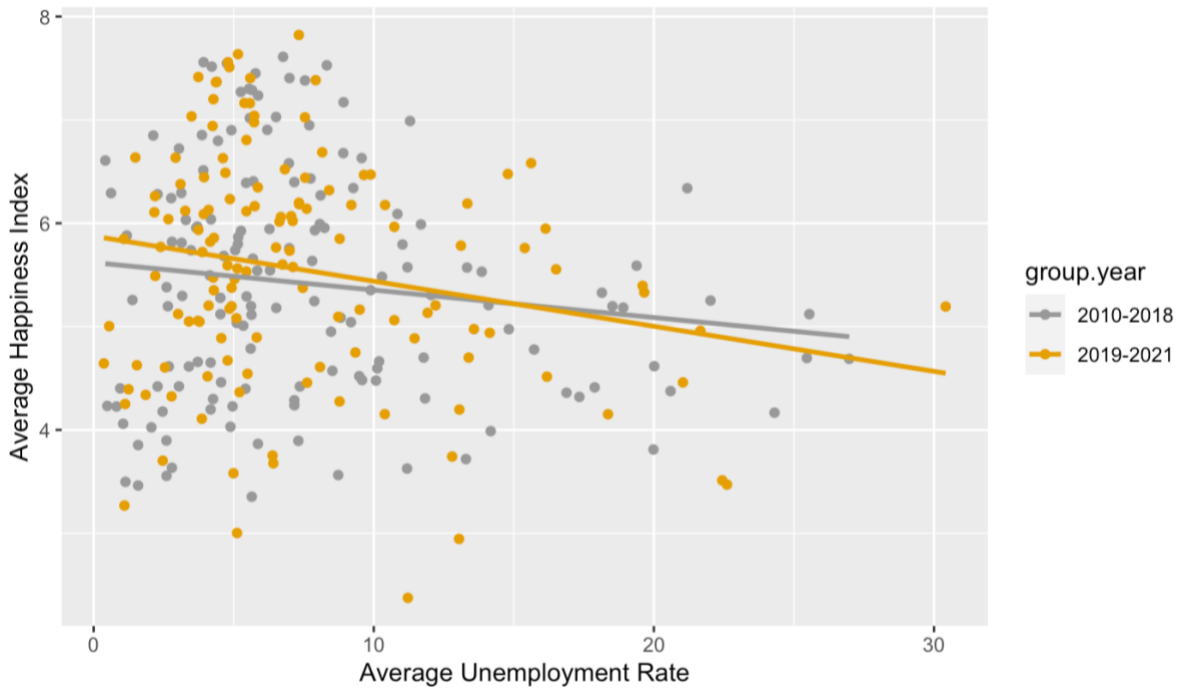
Fig 7



The gray line here indicates the trend of the Average Happiness Index on Average GDP per capita before covid (2010-2018) and the orange line represents the trend during covid. We can clearly see that the lines coincide. This reiterates that the effect covid has on the relationship between dependent and independent variables is negligible. A similar trend can be observed for GDP per capita as well.

Next, we did a similar plot for Unemployment Rate.

Fig 8: Avg Happiness Index against Avg Unemployment Rate before & during covid



From Fig 8, there is a significant difference in slope between the orange and grey lines. The orange line has a little deep slope, this indicates that Unemployment Rate has a higher significance on Happiness Index during the covid period compared to before covid period.

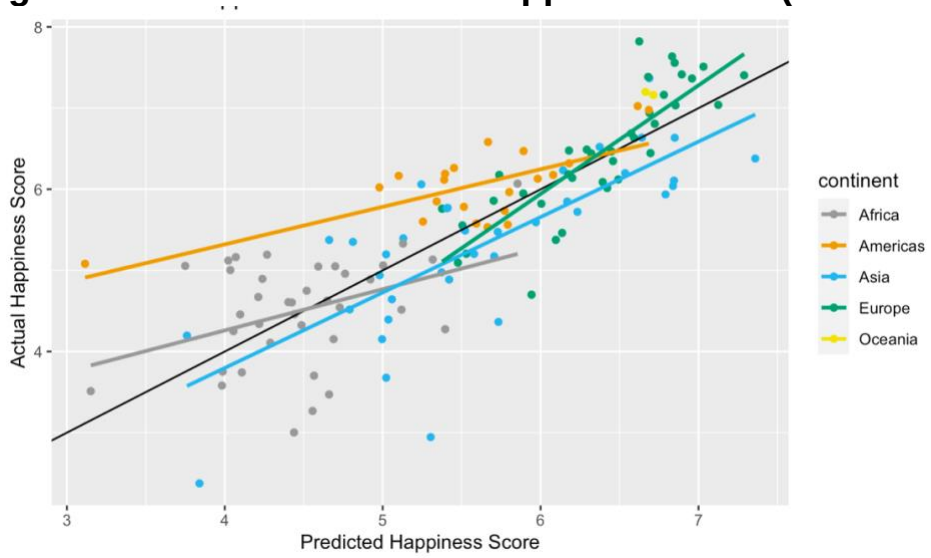
Model predictions

We have tried to build a model and predict the outcomes of the Happiness Index during COVID. For this purpose, we used the before-covid (2010-2018) dataset for our model training and fit. The dataset during covid is for prediction.

Multiple Linear Regression Model

We fit the linear model on all independent variables together against Happiness Index. We then predicted the model on the data from 2019 to 20121. The original values and predicted values are plotted over the continents, to see how well the model predicts the Happiness index based on the three quantitative variables. We have drawn a reference line to determine how well the linear model fits the data.

Fig – 9: Actual vs Predicted Happiness Index (lm model)

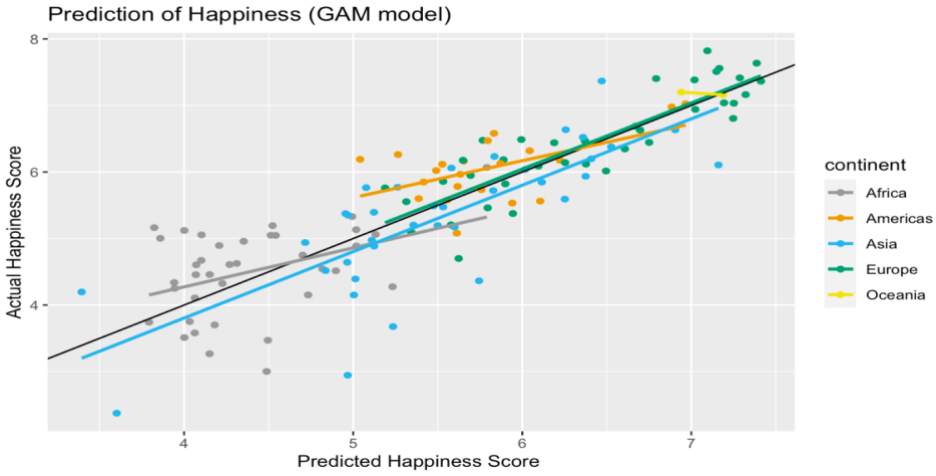


The Fig – 9 graph is the relationship between the actual and predicted happiness index. Each point corresponds to a country with its predicted and original happiness score during covid. The model seems to have been predicted better for Asian countries. American countries have high actual happiness scores than predicted scores and the remaining continents have mixed outcomes.

Generalized Additive Model (GAM)

After fitting the lm model, we have also tried to predict Happiness during COVID with the Gam model. A similar graph has been plotted for gam as well to understand the predictions of the model.

Fig 9



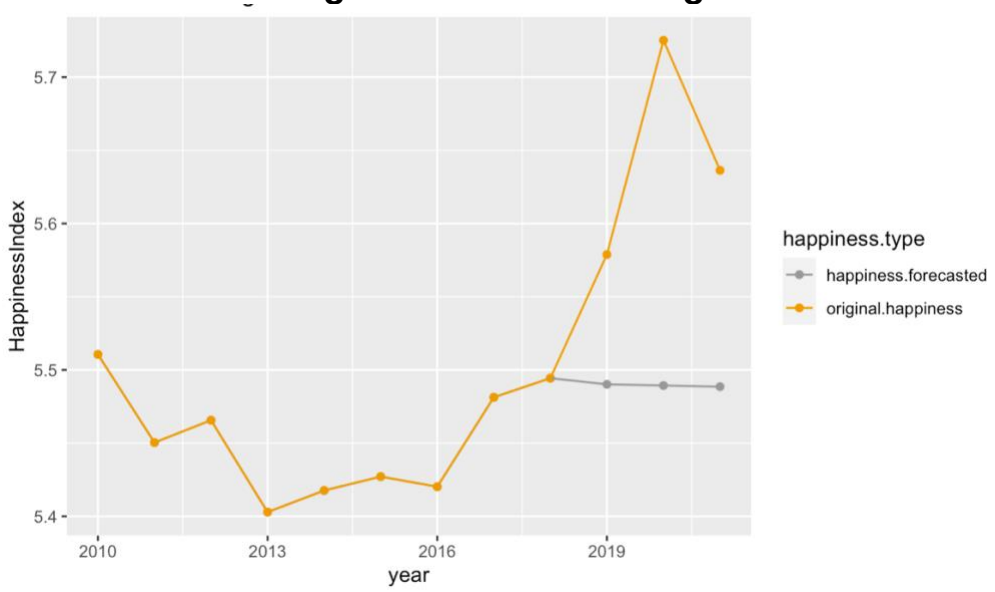
The abline here again determines how well the model predicted. Europe has almost perfect predictions. Asian countries have a similar slope, but different intercepts compared to the

reference line. The Americas predictions have also improved from the lm model as the outliers are not present.

Time Series Forecasting

Time series forecasting means predicting future values over a period without considering any independent variables. It develops a model based on previous data and applies them to make future observations. We have used the Holt trend method for forecasting, we created a data frame with an average happiness index for each year from 2010 to 2018. This data frame has been passed to holt's trend method to forecast the aggregate Happiness index for the next 3 years (during COVID). This forecasting of data determines whether COVID had an impact on Happiness.

Fig 10: Holt forecasting



The orange line represents the original happiness index of all countries for various years. The gray line above is the forecasted values of Happiness during the COVID.

Results

From fig 8, we can say that Unemployment has a slightly high negative correlation with the happiness index during COID than before COVID, whereas there is no significant difference in the relationship between the Happiness index with GDP and Life Expectancy. Now, we will compare multiple linear regression and generalized additive model to see the best fit model. For this, we calculated the mean squared error (MSE) between the original and predicted happiness scores, the lesser the MSE the better the model. MSE for lm is 0.4157 and gam is 0.3523. From these results, we can see that gam is overall a better model.

From fig 10 Holt Forecasting, we can see the increase in the Happiness index from 2018-2020, but there is a slight dip from 2020 to 2021. However, the Holt trend method forecasted that Happiness doesn't increase from 2018-2021. In an expected situation, the actual average

happiness index should decrease due to covid from 2019-2021 but it did not. The forecasted happiness score should ideally increase however it did not.

Conclusions, Limitations & Future Work

To address our research question, we have plotted various plots with lm and made predictions with gam. From our analysis can conclude that due to covid, the only trend between Happiness Index and Unemployment Rate has changed, while the other two variables did not vary.
The effect of annual GDP and life expectancy rate on world happiness during COVID-19 is the same as it was before COVID-19, while the effect the Unemployment Rate has on COVID-19 has slightly increased. Coming to our predictions, by looking at mean squared errors for gam and lm, it can be concluded that gam is a better fit. From the Holt trend forecasting method, we cannot draw conclusions as it gave unexpected results.

One of the main limitations of our analysis was that we had to aggregate the data either by country or by year. Crucial trends for individual data points might not be observed as aggregating the individual trends is neutralized. The average Happiness Index has increased during covid 19, which we thought would decrease. This makes deriving significant conclusions harder. Most of the forecasting methods predicted constant average happiness for 2019 to 2021, only Holt's has given different predictions. In the future, we can work on the data of a particular year instead of aggregating or we can analyze trends in sub-regions instead of searching the trends in each continent.