Dhirubhai Ambani Institute of Information and
Communication Technology

CS306 : Data Analysis and Visualisation

Analysis of Sustainable Development Goals (SDGs), human well-being, and world happiness

Authors
Vishwa Shah : 201801036
Riddhi Tanna : 201801427

Guide: Prof. Pankaj Kumar

5th May, 2021

Table of Contents

1 Introduction

The Sustainable Development Goals (SDGs) are a collection of 17 interlinked global goals designed to be a blueprint to achieve a better and more sustainable future for all. The SDGs were set in 2015 by the United Nations General Assembly and are intended to be achieved by the year 2030. The UN Sustainable Development Solutions Network (SDSN) has been issuing the World Happiness Report for the past few years. Recently, they issued the WHR for 2021.

Fulfillment of various development goals can have an impact on the well-being of people. We analyse the correlation of the these two indices and how various goals contribute to the well-being of people of different regions.

The countries have been divided into 7 categories on the basis of the region they belong to, namely: OECD, Eastern Europe and Central Asia, LAC, East and South Asia, MENA, Oceania and Africa.

2 Datasets

2.1 2020 Sustainable Development Goals

The 2020 SDG Index score is calculated as a metric to compare different countries and it gives a cumulative score to all countries on the basis of the scores for various SDGs. The score metric ranges from *0-100* and treats all SDGs with equal weights. Total of 193 countries are included in this dataset.

SDG Number	Goal
SDG 1	End Poverty
SDG 2	Zero Hunger
SDG 3	Good Health and Well-Being
SDG 4	Quality Education
SDG 5	Gender Equality
SDG 6	Clean Water and Sanitation
SDG 7	Affordable and Clean Energy
SDG 8	Decent Work and Economic Growth
SDG 9	Industry , Innovation and Infrastructure
SDG 10	Reduced Inequalities
SDG 11	Sustainable Cities and Communities
SDG 12	Responsible Consumption and Production
SDG 13	Climate Action
SDG 14	Life Below Water
SDG 15	Life on Land
SDG 16	Peace, Justice and Strong Institutions
SDG 17	Partnerships for the Goals

Table 1: 17 SDGs provided by UN

2.2 World Happiness Report - SDSN

World Happiness Report measures the happiness score of different countries on a scale of *0-10* and takes into account 6 factors: GDP, life expectancy, perception of corruption, generosity, freedom to make life choices and social support. The data is from the Gallup World Poll, which continually surveys 160 countries representing about 98% of the world's adult population.

The respondents value their current lives on a scale of 0–10, with the worst possible life as 0 and

the best possible life as 10. The data is from nationally representative samples. In the collection of data, the sentiment and mood of the people during the time of the survey can affect their answers, however, life evaluations are widely recognised as the standard measure of subjective well-being.

The ladder score mentioned in this report hereon will refer to the Happiness Index (*0-10*).

3   Method and Results

The two datasets have a lot of variables which are interdependent on one another. To identify trends and relations between them, we employ methods of univariate Pearson correlation coefficient and ordinary least squares regression in the following sections.

3.1   Regression Analysis

3.1.1   Dealing with missing data

The missing index scores for specific countries were imputed using the regional average (provided in the dataset) to avoid loss of observations. Each country belongs to one of the 7 regions mentioned in the next section.
We had to drop countries belonging to Oceania when we analysed variables between the two datasets since the WHR dataset has missing data for those countries for the year 2020.

3.1.2   Basic EDA

We initially start with the exploratory data analysis of the SDG dataset and the WHR dataset. Comparison of SDG scores of various regions is as shown in Fig. 9. As expected OECD countries (with greater economic development) have the highest scores. Countries of South Asia and other developing countries of the MENA region have a moderate SDG Index score. Countries in Oceania have the lowest SDG Index score of them all.
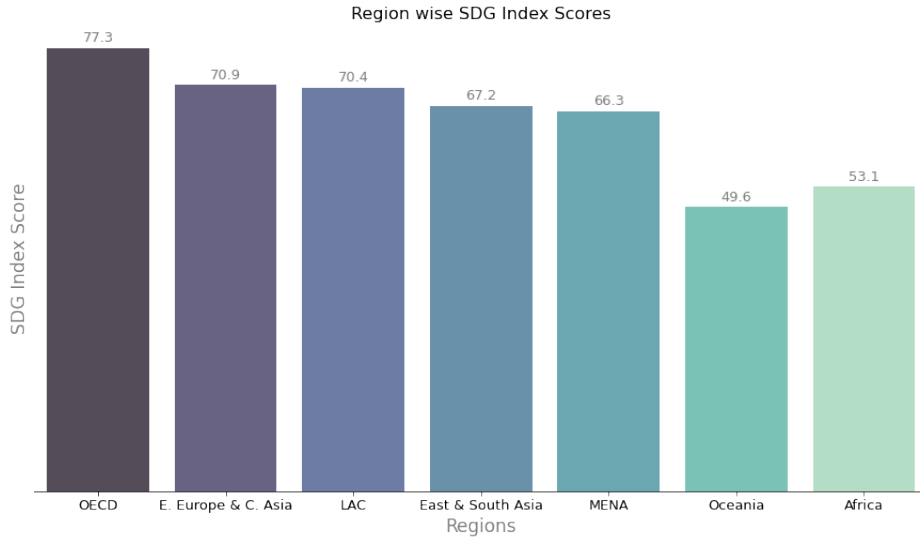

Figure 1: Region wise SDG Index Scores

3.1.3 Linear Regression and Pearson Correlation Coefficient

Fig. 2 shows the scatter plot of the Happiness Index (HI) with SDG Index score. Countries are coded according to the six regions they belong to. Regression analysis was done and the best fit curve was modeled. The SGDI and Happiness Index (ladder score from WHR) have a high correlation coefficient of almost 0.79. The RMSE of the linear and quadratic fit are as shown in Fig. 3. As seen, the RMSE does not decrease significantly for a quadratic fit. The Pearson correlation coefficient is high as well, suggesting that the two variables have a high linear correlation, hence we can say that the indices follow a linear trend. This coincides with our visual observation in Fig. 2.

Figure 2: Linear and quadratic fits for Ladder Score from WHR and the SDG Index Score

```
R-value:  0.7838567988434155
R-squared:  0.6144314810930468
p-value:  2.895198420064518e-28
RMSE - linear:  0.47043931218397445
RMSE - quadratic:  0.43585532260051923
```

Figure 3: R-value, R-squared, and RMSE for the two fits in Fig. 2

As seen in Fig. 2 developed countries of the OECD region have a high SDGI and HI, that is to say these developed countries who do well in their goal fulfillment do better in terms of human well being. Developing and under-developed countries of Africa, have to still work on poverty, healthcare etc goals and concentrate on providing basic necessities to increase the HI. This suggests that Economic activities are really important when considering the countries of lower well being. Countries with higher well-being, need to concentrate more on other factors.

We performed Linear Regression of each of the 17 SDGs with the ladder score. Goals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 15, 16 and 17 have a positive or somewhat neutral slope whereas Goals 12, 13 and 14 have a negative slope. This coincides with the Pearson correlation coefficients obtained in Fig. 6.
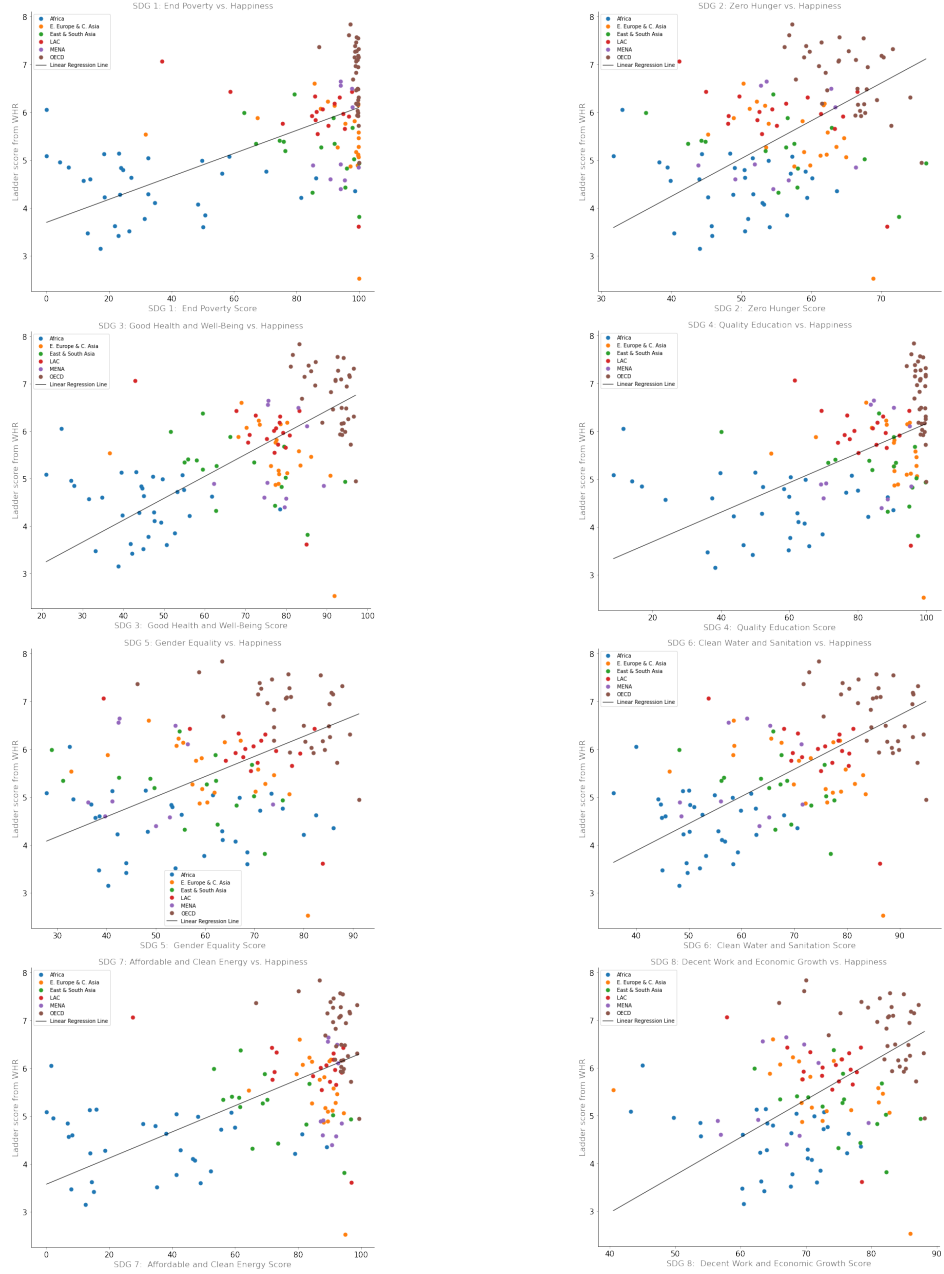
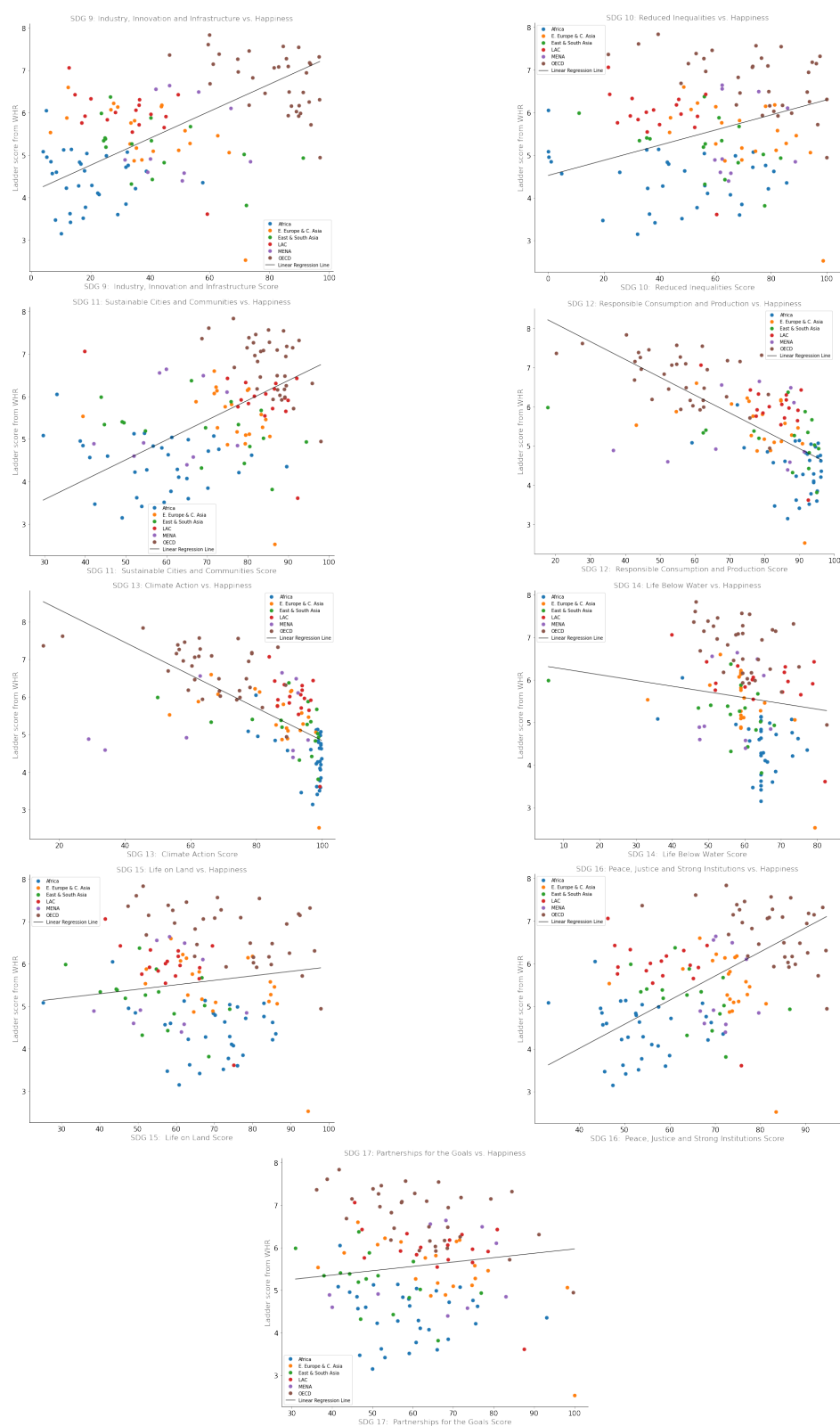Figure 4: Linear Regression of 17 SDGs with Ladder Score (HI)

Figure 5: Linear Regression of 17 SDGs with Ladder Score (HI)

	Ladder score
End Poverty	0.653751
Zero Hunger	0.656955
Good Health and Well-Being	0.819651
Quality Education	0.598823
Gender Equality	0.594611
Clean Water and Sanitation	0.745076
Affordable and Clean Energy	0.687654
Decent Work and Economic Growth	0.686449
Industry, Innovation and Infrastructure	0.786397
Reduced Inequalities	0.376200
Sustainable Cities and Communities	0.661820
Responsible Consumption and Production	-0.739541
Climate Action	-0.701473
Life Below Water	-0.122427
Life on Land	0.138162
Peace, Justice and Strong Institutions	0.703630
Partnerships for the Goals	0.129590
Ladder score	1.000000

Figure 6: Pearson Correlation Coefficients of SDGs with the ladder score

We see that goals like 'Clean Water and Sanitation', 'Industry, Innovation and Infrastructure', 'Good Health and Well-Being' etc. have a high correlation with the HI, whereas, goals that include Climate Change, 'Responsible Consumption and Production' etc. (with long term benefits) are negatively correlated with HI. This shows that there is short-term trade-off between economic development/good well-being and effective consumption/climate action. This suggests that countries fail to take care of the environment while simultaneously keeping in mind the well-being of the people.
We also see that certain goals like 'Life on Land', 'Partnerships for the Goals' and 'Life under Water have negligible impact on the well-being.

These inferences can be better understood by observing three specific countries that lie in different zones of development. We analyse the well-being of Afghanistan, India and Finland in Fig. 7. In the HI as well as the SDGI, Afghanistan lies lower than India which lies lower than Finland (which bags the first place). From Fig. 7, we can see that Finland scores high in goals like 'Industry, Innovation and Infrastructure', 'Good Health and Well-Being' and 'Peace, Justice and Strong Institutions'. In turn, these are the goals that have high positive correlation with the HI. This explains why Finland ranks high on the HI. India seems to score higher on goals such as 'Climate Change' and 'Responsible Consumption and Production' which are negatively correlated with the HI. For India to improve its position on the HI, it will have to hence focus on goals which are highly positively correlated with the HI, such as the ones that Finland seems to focus on.

Another aspect that India can focus on is to find methods such that climate action and responsible consumption go hand in hand with the well-being of the citizens of the country. Their accessibility perhaps may be an issue they would have to tackle, while simultaneously focusing on incentivizing green jobs.

Fig. 8 shows the correlation coefficients of different features used in HI with the 2020 SDGI. We see that GDP has a high correlation coefficient which matches with the results above.
For countries that rank low, we know from **1** that this is largely due to higher perceptions of corruption and lower levels of social support. Social support and freedom have a high correlation and these metrics can prove to be important contributors to the rankings of higher countries. We see the perception of corruption has a negative correlation with the HI. This explains why countries with higher levels of perceptions of corruption rank low even though their GDP per capita is considerably high.

The dataset also contains metrics for positive and negative emotions of people. The survey asks
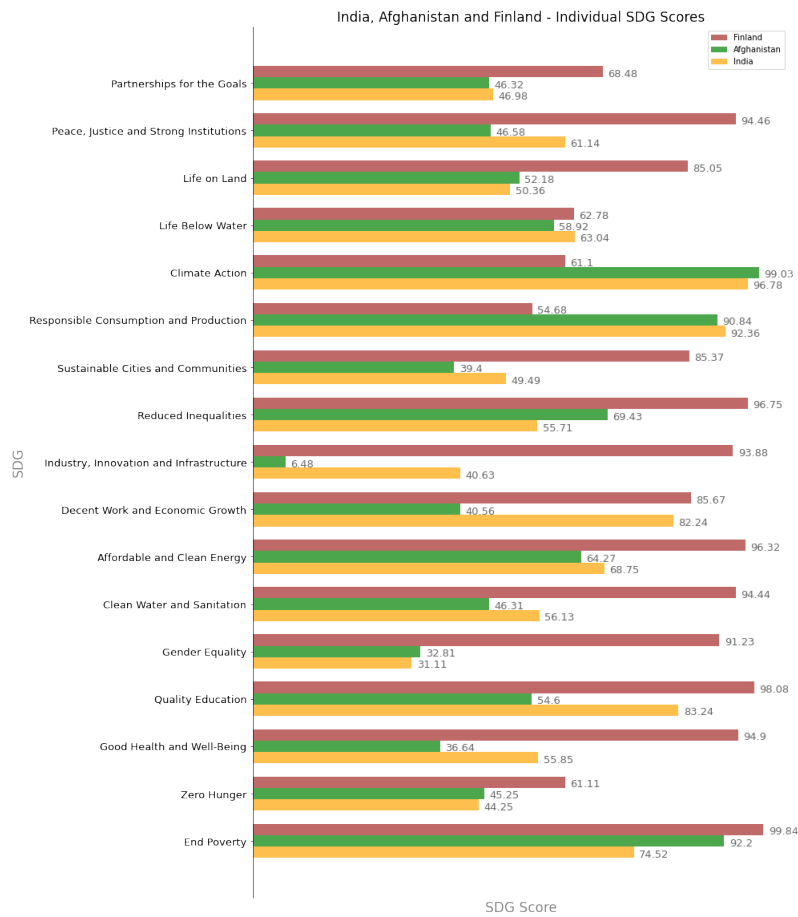
Figure 7: Comparing SDG scores of India, Afghanistan and Finland

the respondents about how they felt the day before they took the survey. Positive emotions include emotions of "love", "trust", "smile" and negative aspects include emotions like "anger", "sadness", "stress" etc.

SDGI and positive affects have a correlation coefficient of 0.23 and SDGI and negative affects have a correlation coefficient of -0.37. This suggests that the SDGI score is affected more by the negative affects than by the positive ones. This could have also been due to the effects of restrictions (such as social distancing and lockdowns) in place due to COVID in 2020.
This also suggests that countries that are not doing well in terms of the SDGs also tend to have populations that are experiencing more negative emotions.
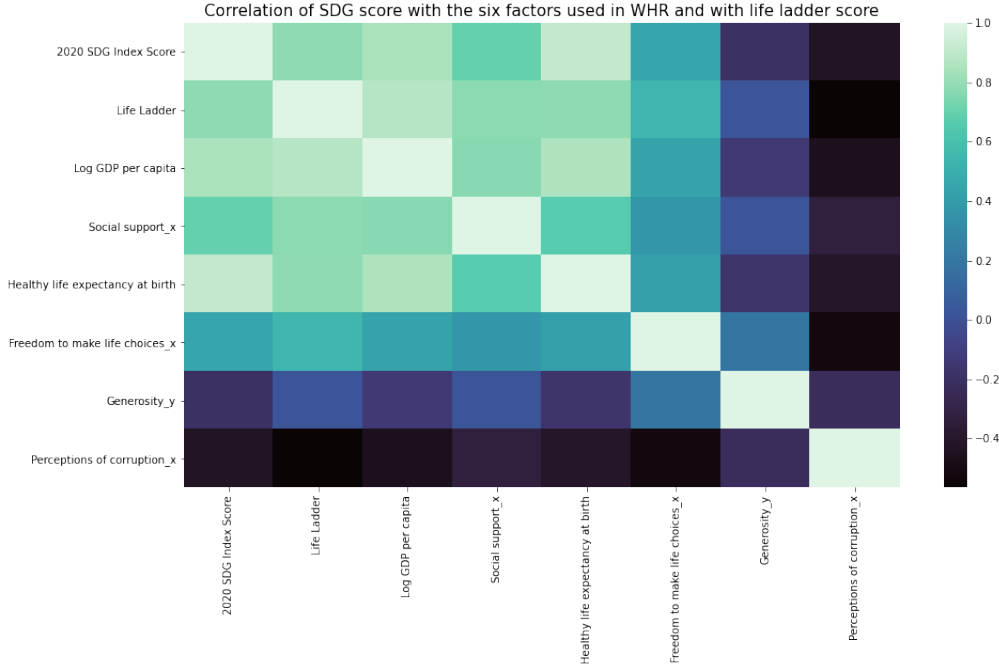
Figure 8: Heatmap of correlation coefficients of factors of HI with 2020 SDGI

	2020 SDG Index Score	Positive affect	Negative affect
2020 SDG Index Score	1.000000	0.228699	-0.370674
Positive affect	0.228699	1.000000	-0.353242
Negative affect	-0.370674	-0.353242	1.000000

Figure 9: Correlation coefficients of positive and negative affects with the 2020 SDGI

3.2   Principal Component Analysis

In the SDG dataset, we have 17 SDGs. We consider these goals as 17 dimensions of the dataset. In general, it is difficult to visualize these many dimensions (or any dimension more than 3). We want to now analyse which goals the countries focus on at large and which ones they ignore at large. Hence, we employ Principal Component Analysis to this dataset to reduce the number of dimensions to 2. When we perform the PCA, we find that the first two components (which we will now refer to as PC1 and PC2 respectively) account for almost 62% of the variance in the dataset. Fig. 10 shows the loading plot after performing PCA. The loading plot shows the weights of each of the 17 dimensions (SDGs) on axis PC1 and axis PC2. PC1 accounts for 53% of the explained variance. From Fig. 10, we can see that goal 14, goal 15 and goal 17 contribute lesser weights to PC1. We can say that these three goals are the least focused on by countries at large. From Table. 1, we know that these goals are 'Life Below Water', 'Life on Land' and 'Partnerships for the Goals'.

From Fig. 5, we can see that the correlation coefficient of:

goal 14 = -0.122

goal 15 = 0.138

goal 17 = 0.129

These values indicate low correlations of these goals with the HI.

From Fig. 10, we can also see that countries have at large focused on goal 3, goal 9 and goal 16. From Table. 1, we know that these goals are 'Good Health and Well-Being', 'Industry, Innovation, and Infrastructure', and 'Peace, Justice and Strong Institutions'.

From Fig. 5, we can see that the correlation coefficient of:

goal 3 = 0.81

goal 9 = 0.78

goal 16 = 0.703
These values indicate high correlations of these goals with the HI.
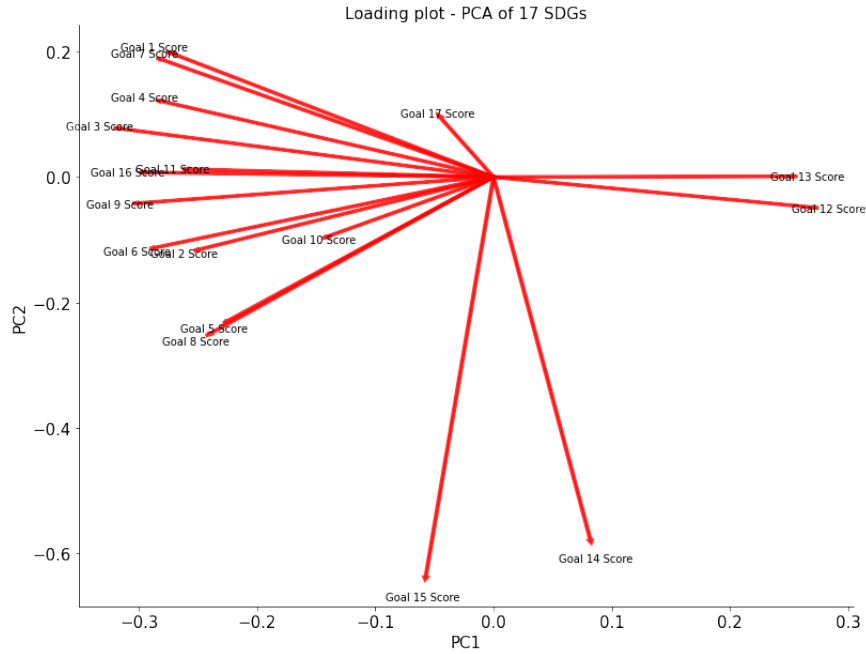

Figure 10: Loading plot after performing PCA on the SGDs

4 Conclusion

We see that there is short-term trade-off between economic development, health and well-being with effective and responsible consumption, and climate action. This suggests that countries fail to take care of the environment while simultaneously keeping in mind the well-being of the people. India seems to score higher on goals such as 'Climate Change' and 'Responsible Consumption and Production' which are negatively correlated with the HI. For India to improve its position on the HI, it will have to hence focus on goals which are highly positively correlated with the HI, such as the ones that Finland seems to focus on.
Another aspect that India can focus on is to find methods such that climate action and responsible consumption go hand in hand with the well-being of the citizens of the country. Their accessibility perhaps may be an issue they would have to tackle, while simultaneously focusing on incentivizing green jobs.

Governments of developing countries should focus more on building infrastructure, innovation and removing poverty.
Governments of all countries at large need to focus on balancing the trade-off between well-being and responsible consumption and climate action since that is the biggest challenge that we need to face head-on in the near future.

As in all indices, these indices also come with flaws. They have faced criticism as well. But on a whole, they are the best indices measured at the moment.

5 Link to the GitHub repository

GitHub Repository for the data analysis code

6   References

https://worldhappiness.report/ed/2021/

https://www.sdgindex.org/reports/sustainable-development-report-2020/

https://medium.com/analytics-vidhya/analyzing-world-happiness-index-report-890aedeb138

https://www.nature.com/articles/s41598-020-71916-9.pdf

https://www.kaggle.com/dgtech/world-happiness-with-basic-visualization-and-eda