Sustainable happiness

Abstract—**In this report we conduct a study on sustainable world happiness by combining world happiness report and happy planet index report. It checks the effect of ecological factors on the happiness of countries. It also analyses the relationship between ecological footprint, GDP and overall healthy life expectancy of a country. The study mainly focusses on world's happiest country Finland. It then predicts the happiness score (ladder score) using multilinear regression model.**

I. INTRODUCTION

Investigating about happiness has become a very interesting and exciting research area over many years. In order to help guide public policy, the UN general assembly invited member countries in July 2011 to measure happiness of their people. On April 1 2012 , the first happiness report was released which had great international attention. The ranking for world happiness index is based on ladder score. In 2006 New Economic Foundation introduced happy planet index considering human wellbeing and environmental impact. It gives an idea of how well a country's wellbeing is sustainable .Realising the damaging effects of environmental pollution, organisations and governments are trying to build policy and decisions considering ecological factors too. This report looks into the effect on ecological factors on world happiness by combining world happiness index and happy planet index reports.

II. DATA AND ANALYTICAL QUESTIONS

A. DATA

World happiness report is a publication of the sustainable development solution which is powered by data from the Gallup World Poll and Lloyd's Register Foundation. World happiness score is measured with various attributes such as GDP, Freedom to make life choices, Corruption ,Healthy life expectancy etc . For this project we have used Kaggle datasets between years 2015- 2020 for happiness index[2].

The ranking for world happiness is calculated based on six factors:

Sl.no	World happiness index	
	Factors	*Data Source*
1	GDP per capita	OECD Economic Outlook No 106
2	Healthy Life expectancy	World Health Organization's (WHO) Global Health Observatory data repository
3	Generosity	Gallup World Poll
4	Social support	Gallup World Poll
5	Freedom	Gallup World Poll
6	Corruption	Gallup World Poll

Table 1

Happy planet Index measures sustainable wellbeing of all[1].It is measured considering three main factors: life expectancy(from the UN), subjective wellbeing(from the Gallup world poll) and ecological footprint(from the Global footprint network).The dataset we used is from the official website of Happy planet index [1]. Currently we have data till 2020 for happy planet index. Hence, we merge datasets of world happiness index with dataset Happy planet index from 2015 – 2020

B. Analtical Questions

1. Do ecological footprints affect the ladder score of a country?

2. Does biocapacity per year plays any crucial role in having better ladder score?

3. Does GDP per capita, health life expectancy and ecological footprints have any impact on the ladder score of the world's happiest Countries?

4. Can we predict better happiness with features like GDP, health and ecological footprints alone?
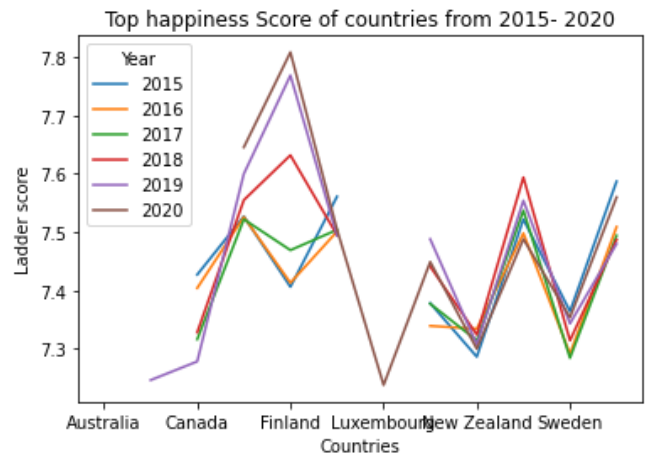

Fig 1 Top Happiness score of countries

III. ANALYSIS

A. Plan

To investigate more on the research question, the following plan was derived. To achieve the best results, several iterations were done and some of the steps were re-visited.

1. Collect dataset between 2015 – 2020. Add additional column year to each dataset for easy analysis.
2. Check if any columns need to be renamed or added extra to merge with happy planet index
3. Merge the world happiness dataset with happy planet index dataset.
4. Clean data: handling missing values and outlier removal
5. Feature engineering

6. Check for correlation between Ladder score, log GDP per capita, Health (life expectancy), Biocapacity for year and Ecological factors.
7. Build a predictive model to check if GDP, health and ecological factors alone can give better ladder score to a country
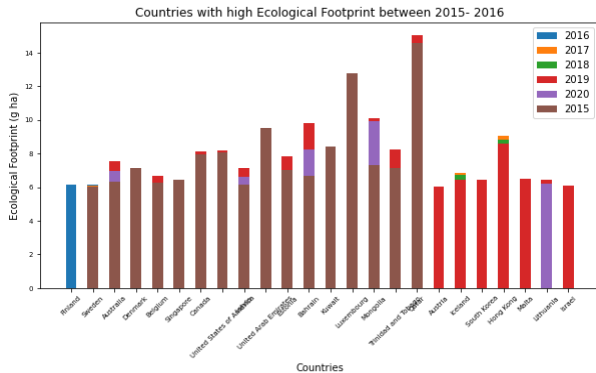

Fig 2 Countries with high ecological footprint

.

B. Data preparation and derivation

There are around 6 datasets (one for each year from 2015 - 2020) to merge for world happiness report. The features need to be sorted, renamed and paired accordingly to avoid issues while concatenating. The extra columns needed are added to each required dataset. Also, value zero is assigned for these columns with missing data. Before merging all the dataset an extra column for year is added to differentiate each data. The final merged dataset for world happiness index contains columns as shown in table 2.

Sl No	World Happiness Index	Happy Planet Index
	Column names	Column names
1	Country	Country
2	Region	Continent
3	Ladder score	HPI
4	Standard error of ladder score	ISO
5	upperwhisker	Population (thousands)
6	lowerwhisker	Life Expectancy (years)
7	Logged GDP per capita	Ladder of life (Wellbeing) (0-10)
8	Social support	Ecological Footprint (g ha)
9	Healthy life expectancy	HPI rank
10	Freedom to make life choices	Biocapacity for year (g ha)
11	Generosity	GDP per capita ($)
12	Perceptions of corruption	Year
13	Ladder score in Dystopia	
14	Log GDP per capita	
15	Health (Life Expectancy)	
16	Freedom	
17	Explained by: Social support	
18	Explained by: Generosity	
19	Government Corruption	
20	Dystopia + residual	
21	Family	
22	Happiness Rank	
23	Year	

Table 2

For Happy planet Index the dataset was extracted from a report [1]. The column 'year' had to be added extra in the dataset. After merging world happiness index and happy planet index datasets, missing values were replaced with either null or mean. Duplicate columns are removed. Outlier removal and feature engineering was done after this for improved prediction.

C. Analysis of data

From figure 1 and 2, we could observe that some of the countries like Finland, Canada etc with high world happiness index have high ecological footprints. We could also observe that Finland is seen to have high ecological footprint only in 2016 whereas Canada is seen to have high ecological footprint for 2015 and 2019.Infact the value has increased for Canada. From fig 2 we can see Canada's happiness score is seen to reduce in 2019 and Finland is shown to have great improvement in happiness score. This clearly signifies the effect of ecological footprints over happiness score or ladder score of a country. This analysis answers our first research question.
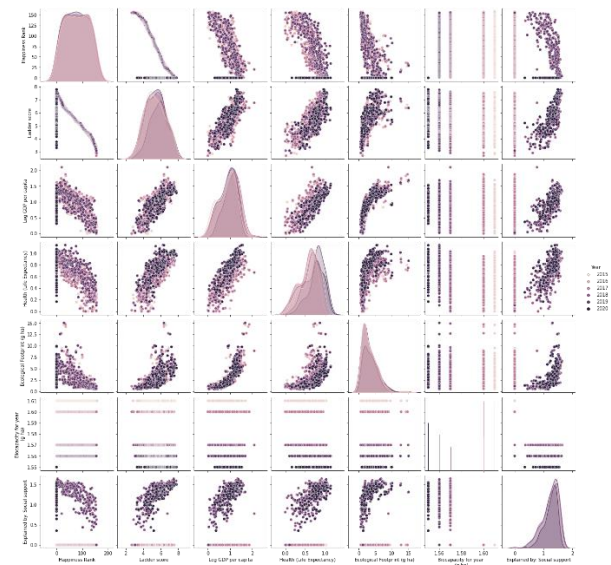

Fig 3 Countries with high ecological footprint

The figure 3 shows the pair wise relationship between the following features: Happiness Rank, Ladder score, Log GDP per capita, Health (Life Expectancy), Ecological Footprint (g ha), Biocapacity per year (g ha), Explained by: Social support with respect to year. It's seen that ladder score has good pair wise relationship with Log GDP per capita, Health (Life Expectancy) and Ecological Footprint (g ha). Whereas no much change in ladder score is observed for Biocapacity per year. To investigate on this further we checked the Pearson correlation coefficient of the features against ladder score to check for the linear relationship between them.

From table 3 it is evident that there is moderate positive correlation between ladder score and ecological footprint. Similarly, both Log GDP per capita and Health (Life Expectancy) shows good correlation with ladder score. This could lead to multicollinearity. Since Log GDP per capita is highly correlated to ecological footprint as well, we drop it and will build predictive model with ladder score, ecological footprint and Health (Life Expectancy). Table 3 and Figure 5

2

illustrates that ladder score has huge correlation towards log GDP per capita, Health (Life Expectancy) and ecological footprint. This analysis answers our third research question.

| Sl.no	Pearson correlation	
	Field	*Correlation coefficient*
1	ecological footprint	0.6612652548907889
	ladderscore	
2	ecological footprint	-0.09219557589288654
	Biocapacity for year	
3	Log GDP per capita	0.7333118938884975
	ladderscore	
4	Health (Life Expectancy)	0.7333118938884975
	ladderscore	
5	ecological footprint	0.854284882830217
	Log GDP per capita	
6	ecological footprint	0.6930312082493724
	Health (Life Expectancy)	

Table 3
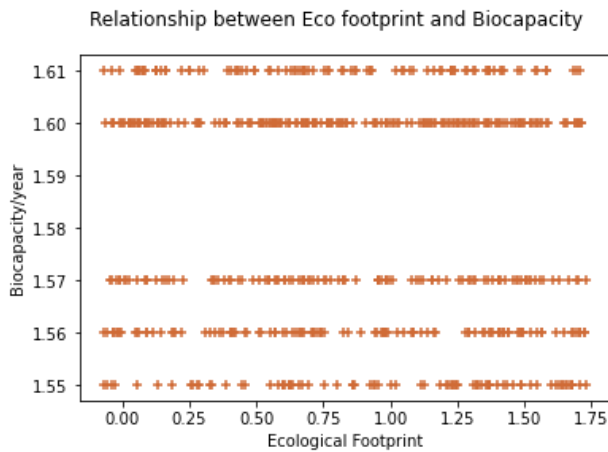
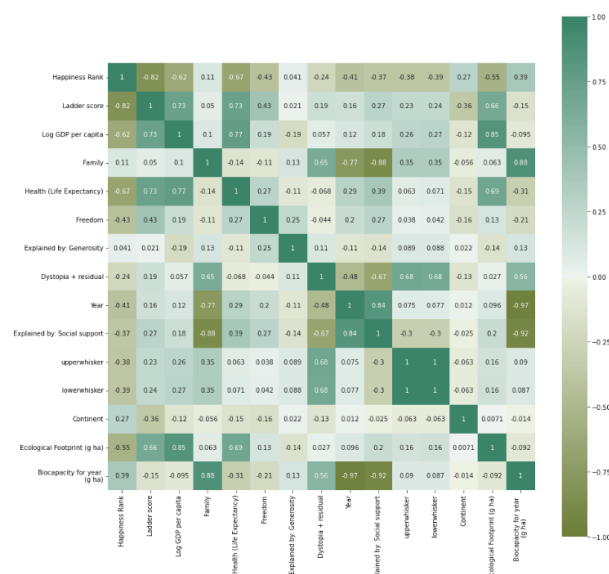Fig 4 Ecological Footprint vs Biocapacity/year

Fig 5 Correlation matrix between all the features

We also observe that ecological footprint and Biocapacity for year is having negligible correlation. This analysis answers our second research question Hence this feature doesn't require any further investigation and is removed from the analysis.

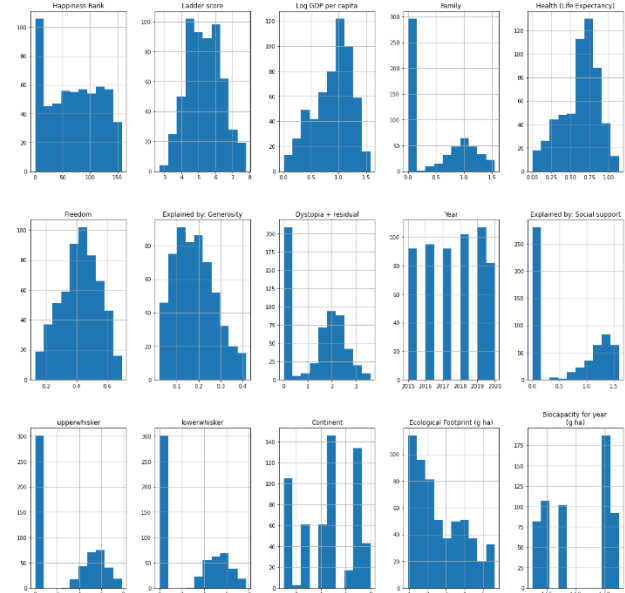Checked if any of the features need a transformation

Fig 6 Checking for Data transformation

It's observed that log transformation needs to be applied on Ecological Footprints to make the analysis more valid.

D. Construction of models,Prediction and validation

As, we could see a linear relationship between the target and predictor variables, multilinear regression will be a good model for prediction. Since Decision tree is good for complex dataset it is also used along with multilinear regression. Since multilinear regression gave better accuracy (89%) when compared with Decision tree (69%), it was chosen as model for prediction. As we are predicting numerical variables, it's important to do scaling on target variable. Hence to obtain this we used TransformedTargetRegressor class for linear regression. The predicted value had slight variation from the original data.

IV. FINDINGS REFLECTION AND FUTURE WORK

We see a huge correlation between Health (Life Expectancy) and ecological footprint in figure 5. Growth of a nation is very much linked to the productivity and wellbeing of its people [7]. The productivity is often affected by poor environmental health condition which leads to various illness. Moreover, the increased medical cost result in loss of people's earnings [7]. Thus, this could in turn affect the economic growth and thereby the happiness score of a country.
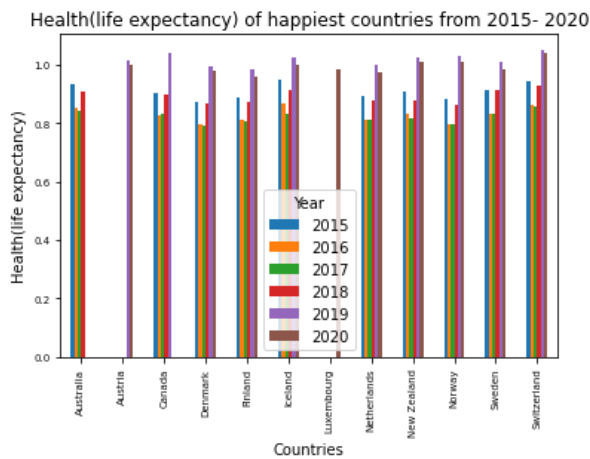
Fig 7 Countries with high Healthy life expectancy

The figure 7 illustrates how well countries perform on health life expectancy between year 2015 – 2020. Switzerland is seen to outperform than the rest of the countries. Most of the them have a small drop in life expectancy for 2020.This could be due to the pandemic. Finland also shows good life expectancy. When we compare between fig 2 and fig 7, we could also observe that Finland is seen to have reduced its ecological footprints after 2016 and slowly raised the life expectancy.
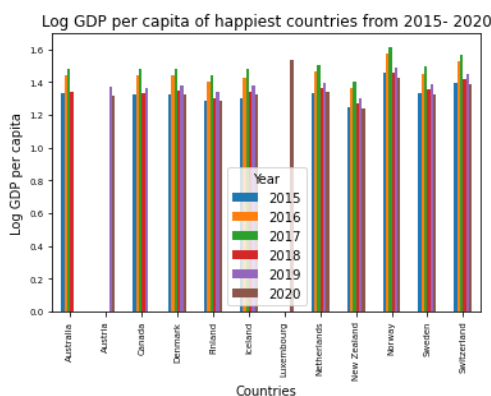

Fig 8 Countries with high GDP per capita

From figure 8 its evident that 2016 and 2017 countries have high GDP. But when compared with healthy life expectancy we can see a decline in these years.

From the analysis we could conclude that ecological footprints affect the country's economy, health and overall happiness. Diseases and poor health hinder economic growth of a country.[7] Finland is seen to perform better each year and is said to be world's happiest country. When considered on terms of sustainable happiness Finland is seen to improve its ladder score each year by reducing its ecological footprints.

It's seen that after 2016 Finland reduced its ecological footprints below 6.

While predicting ladder score for existing value, we could see slight variation. Which implies that there are some factors which affect the ladder score apart from Healthy life expectancy and Ecological footprint. This analysis answers our fourth research question. Since GDP per capita was dropped due to multicollinearity, we couldn't investigate how it affects the prediction. In future we will look into the world happiness considering all the 6 factors (GDP per capita, Healthy life expectancy, Generosity, Social support, Freedom and corruption) from world happiness report along with ecological factors from happy planet index. For further understanding about the factors which leads to a happy nation, we will compare the worlds happiest and least happiest countries. As part of performance improvement, will do kfold cross validation and grid search CV. Also, residual sum of squares will be calculated to explore how well the regression model fits the dataset.

Sl No	Section	Word count
1	Abstract	70
2	Introduction	135
3	Analytical questions and Data	263
4	Analysis	963
5	Finding reflection and future work	434

REFERENCES

[1] https://happyplanetindex.org/

[2] https://www.kaggle.com/mathurinache/world-happiness-report-20152021

[3] W. -j. Hsieh and B. Chen-heng Yang, "The pursuit of happiness: An empirical study of BLI of OECD countries," 2015 International Conference on Orange Technologies (ICOT), 2015, pp. 83-87, doi: 10.1109/ICOT.2015.7498481.

[4] S. F. Hashemi Fesharaki, A. Behrouz, J. Yang, D. Y. Wohn and M. Cha, "Green Space and Happiness of Developed Countries," 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), 2020, pp. 247-250, doi: 10.1109/BigComp48618.2020.00-67.

[5] Graham, C., 2012. *Happiness around the world: The paradox of happy peasants and miserable millionaires*. Oxford University Press.

[6] https://sustainabledevelopment.un.org/index.php?page=view&type=400&nr=617&menu=35

[7] Z. B. Irfan, "Environmental pollution and public health: The socio-economic analysis of the global drivers of change," Recent Advances in Space Technology Services and Climate Change 2010 (RSTS & CC-2010), 2010, pp. 19-24, doi: 10.1109/RSTSCC.2010.5712791.
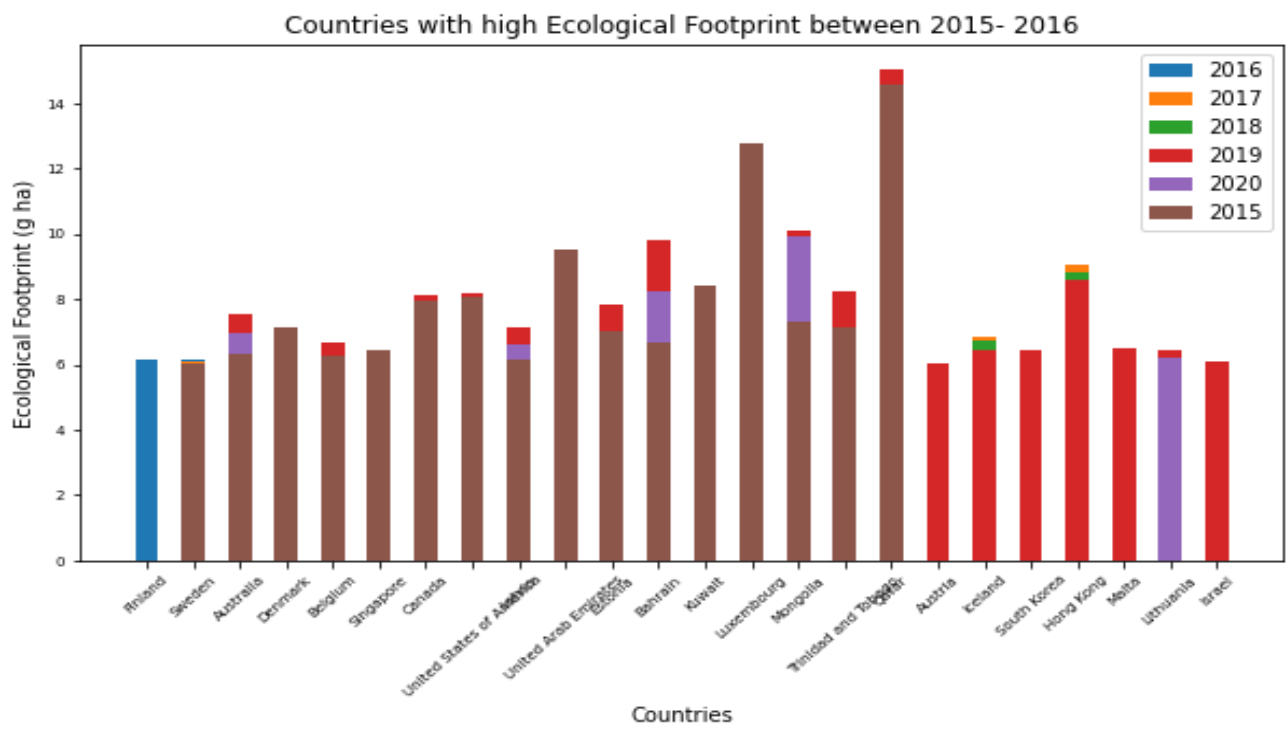
Fig 2 Countries with high ecological footprint