

World Happiness Report

Manojkumar Darshankar
PES1UG20CS662
CSE,PES UNIVERSITY
Bengaluru
mvdsdarshankar@gmail.com

Satvikkumar
PES1UG20CS687
CSE,PES UNIVERSITY
Bengaluru
satvikkumarhangarge@gmail.com

Manjula I Kamannavar
PES1UG21CS818
CSE,PES UNIVERSITY
Bengaluru
manjulakamannavar1976@gmail.com

Abstract— This paper describes our first research experience in the field of data analytics by using open-source software tools Python and jupyter notebook. The project focuses on Exploratory Data Analysis and Data Visualization for the data-sets in picture. This would help us study the behavior of the attributes and their values thus highlighting points of interest for us to work ahead with.

Keywords— EDA,Data Visualization,happiness index,Python

I. INTRODUCTION

Happiness is a multidimensional factor. Analyzing happiness is complicated since it is a subjective matter, and it varies. The underlying challenge here is to quantify happiness and well-being in a single frame to see which strategies are by far the most efficient. The report presents a framework gauging people's qualitative assessment of their lives, progress and well-being through quantitative indicators. Such as GDP per Capita, Social support, Healthy life expectancy etc.. Our aim is to Model the Happiness Index over a time span of 3 years (using 2019 dataset) and compare it with the present scenario i.e. 2022. The major problem we might face is the effect of covid-19 on the happiness index of popular countries.

II. RELATED WORK

There are many published papers related to the world happiness index that considers models to perform prediction of the Happiness Index. Concepts of Artificial intelligence and Machine learning are used to build models to predict Happiness, and other works use other machine learning techniques.

In paper [1] the author specifies about Dystopia, residuals, why only six factors are used to explain life evaluations, role of social media on happiness of people, why Bhutan is not listed in 2022 WHR.

In paper [2] The author discussed the happiness measures currently in use across countries, specifically the Gallup World Poll (GWP), the World Values Survey (WVS), and the European Social Survey (ESS), and asks whether or not these measures can provide valid information about quality of life that can be used to guide policy-making. He discussed the causes of happiness and misery, based on 30 years of research on the topic.

In Paper [3] The Authors pooled results Gallup World Poll Surveys From 2017-2019. Showed how top countries maintained their top spots over the years. Most countries showed significant changes from 2008-2012 to 2017-2019. India, with close to a fifth of global population, saw a

1.2-point decline. The conclusion of the research was social environments are of first-order importance or the quality of life.

PROBLEM STATEMENT

A. Dataset

We obtained our datasets from the Kaggle website. The World Happiness Report is a publication that contains Region and rankings of national happiness, based on respondent ratings of their own lives, which the report also correlates with various (quality of) life factors.. The parameters vary based on the targeted solution. We chose the data that felt relevant to our study of the world happiness report and its ability to use survey data to determine how people evaluate their lives in more than 156 countries. There are no null values, outliers are supposed to be handled.

B. Exploratory Data Analysis

It was found that 7 out of 9 of the attributes were of type – float and others were int and object. It represents that 9 columns in this data have float, integer and object data type. Data inconsistency prevails as long as missing values are not treated properly. From fig 1.1 and scatterplots it is clear that PCA helps visualize data

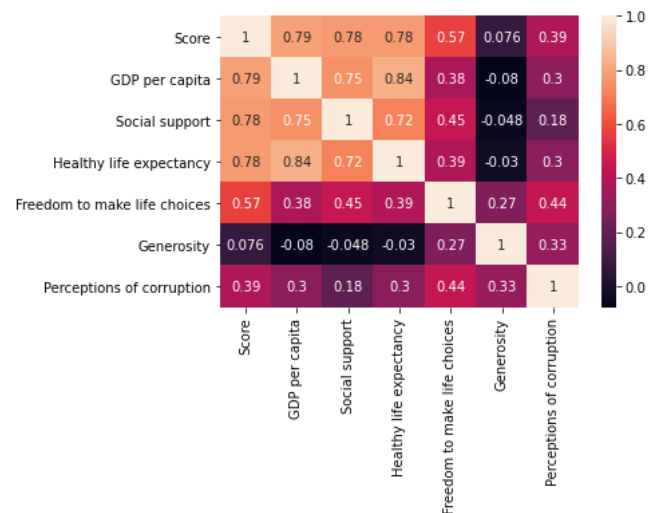


Fig 1.1 Heatmap shows correlation of different variables in the dataset.

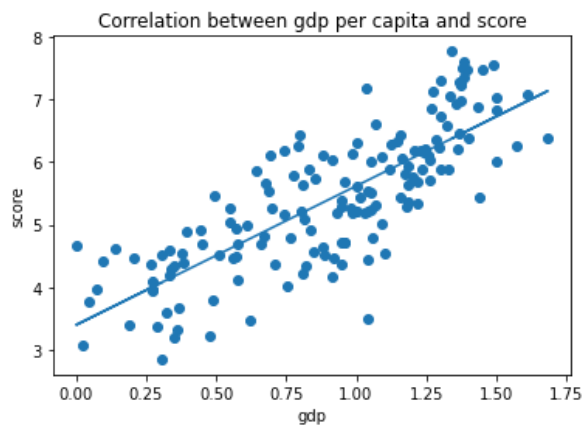


Fig 1.2

We can see that there is a strong correlation between wealth and happiness. meaning wealthier countries are happier.

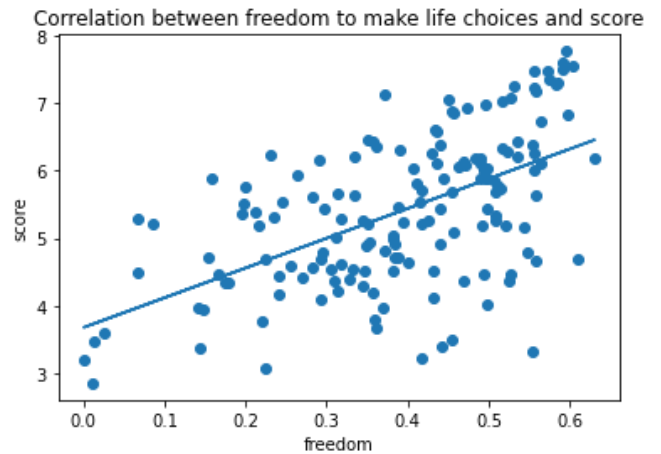


Fig 1.3

Positive correlation between freedom and happiness exists.

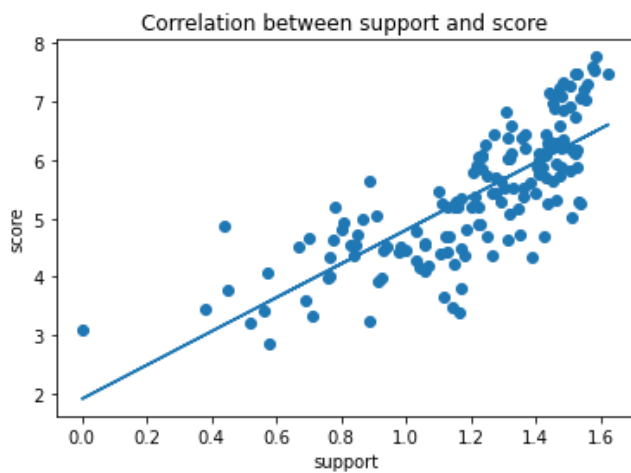


Fig 1.4

Family Support contributes to happiness of people as there is a strong correlation between happiness and support

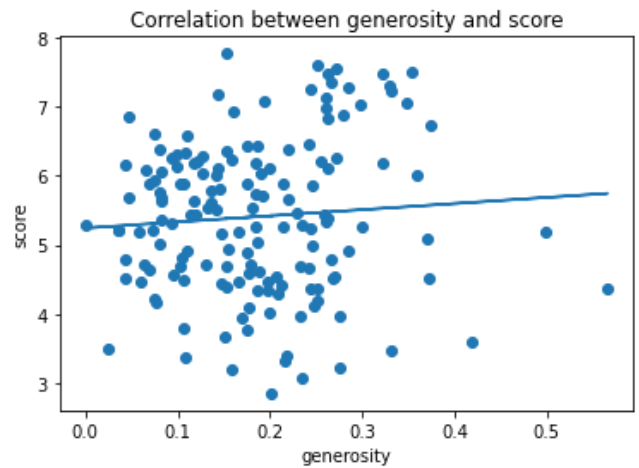


Fig 1.5

Generosity does not affect the happiness of people, but generosity cannot be quantified easily, so its difficult to make inferences on this relation.

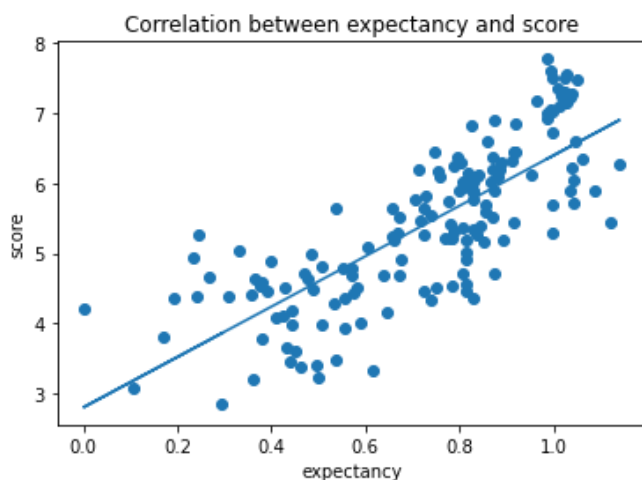


Fig 1.6 There is a strong positive correlation between life expectancy and happiness

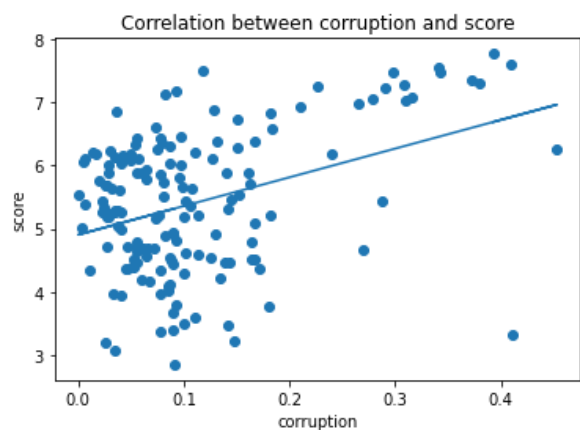
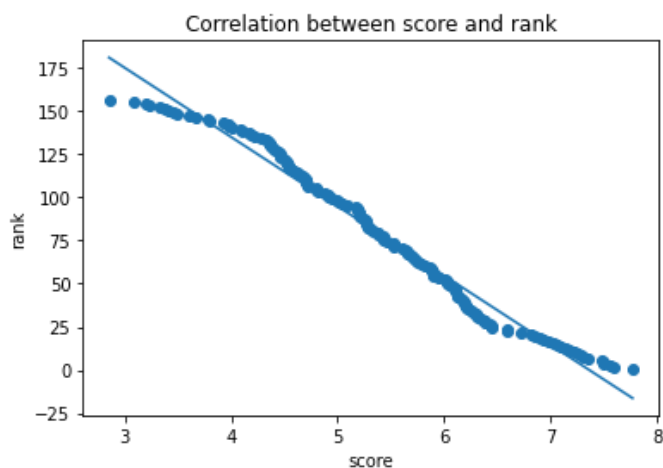


Fig 1.7 Slight positive correlation between happiness of people and corruptio



this it can be inferred that higher the score higher the rank of the country or Region

GDP per capita has a positive correlation with people's happiness

Social Support has a positive correlation on people's happiness

Healthy life expectancy has a positive correlation on people's happiness

Generosity does not affect people's happiness much.

General Freedom has a positive correlation on people's happiness.

Corruption has a weak but positive correlation on people's happiness.

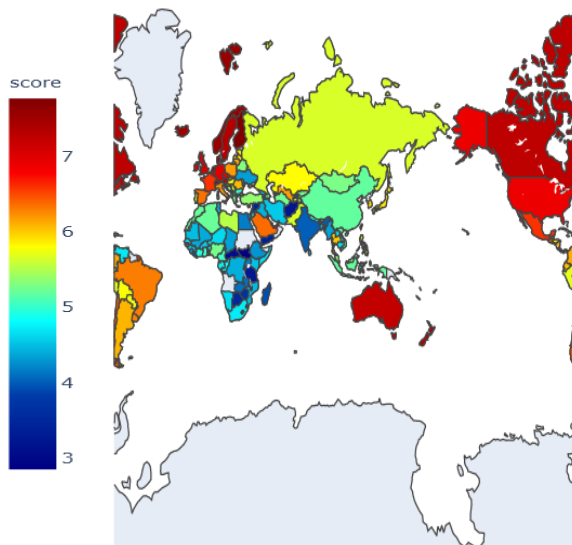


Fig 2.1 The above Plot shows the happiness score chloropleth map around different parts of the world

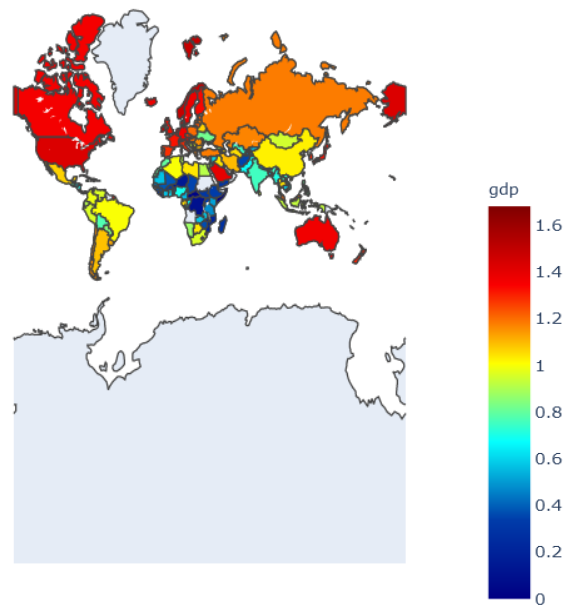


Fig 2.2 the above plot shows the Gdp per capita chloropleth map around different parts of the world.

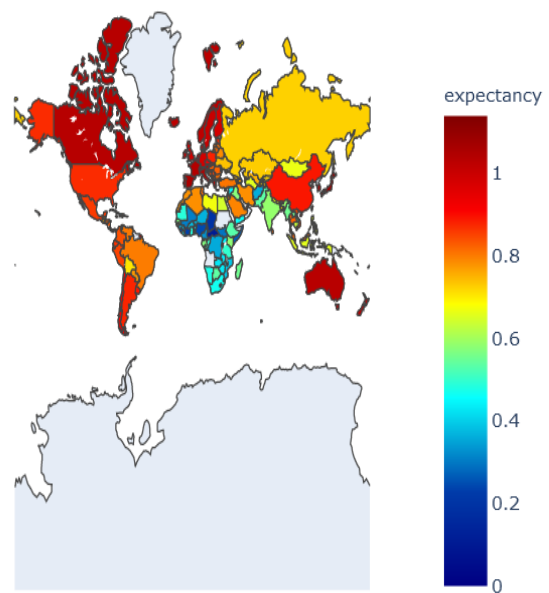


Fig 2.3 The above Plot shows the expectancy score chloropleth map around different parts of the world

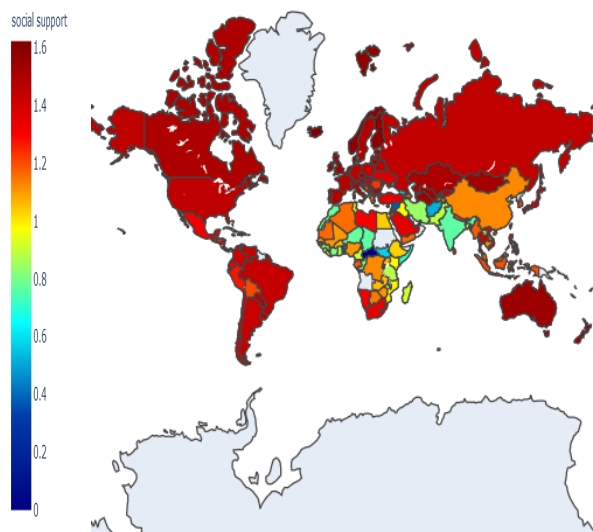


Fig 2.4 The above Plot shows the social support score chloropleth map around different parts of the world

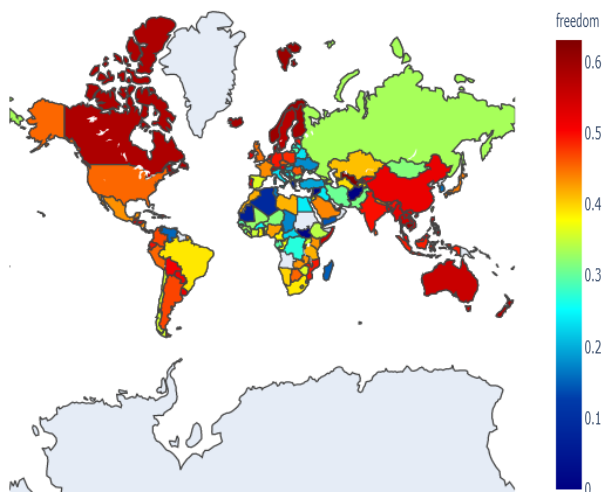


Fig 2.5 The above Plot shows the freedom score chloropleth map around different parts of the world.

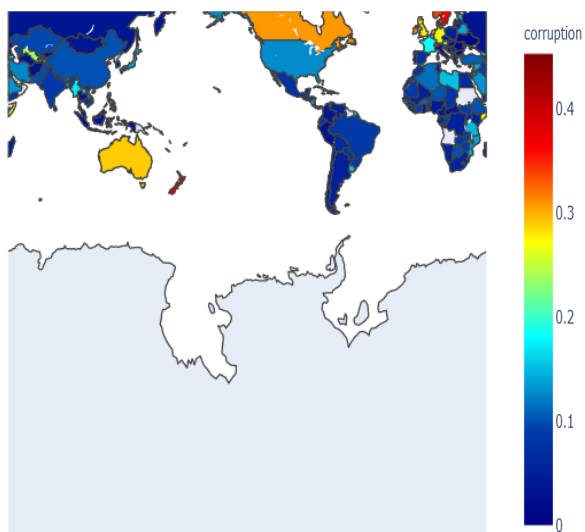


Fig 2.6 The above Plot shows the corruption chloropleth map around different parts of the world

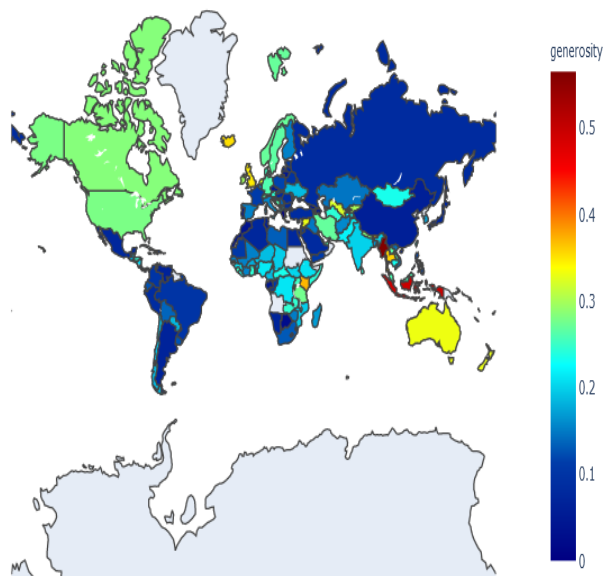


Fig 2.7 The above Plot shows the generosity score chloropleth map around different parts of the world

REFERENCES

1. World happiness Report 2022

<https://worldhappiness.report/>

2. World Happiness Report 2017

Authors : John F. Helliwell, Richard Layard, Jeffrey D. Sachs

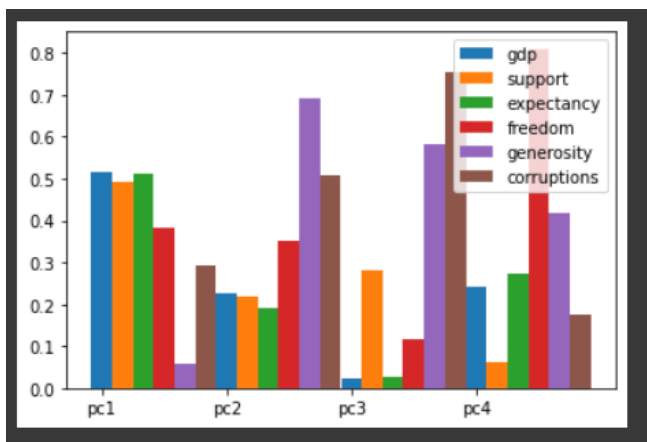
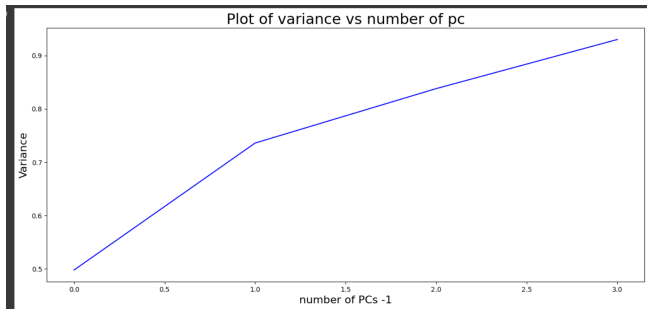
https://www.wellbeingintlstudiesrepository.org/hw_happiness/6/

3. World Happiness, Trust and Deaths under COVID-19

Author : Max Norton, Shun Wang, Haifang Huang

https://www.researchgate.net/profile/Shun-Wang-31/publication/350511691_World_Happiness_Trust_and_Deaths_under_COVID-19/links/6063d19b299bf173677dc90c/World-Happiness-Trust-and-Deaths-under-COVID-19.pdf

C. PCA to determine features that significantly determine the happiness index for that nation and providing reasonable explanation to some of the visible inferences that have come about due to these operations.



Thus the features that affect the happiness index very highly are GDP, life expectancy, and social support since they contribute a lot for the first principal component, which has the highest variation ratio of 50%. This makes sense as Money or income, Healthy life, and support from family and friends contribute to healthy life.

D. Multiple Linear Regression Model to predict the Happiness Score with 2016, 2017, 2018 for training of the model and 2019 for the testing of the model. We found the slope and the **intercepts of the model's equation**, which helps us determine which features affect the Score largely based on the slopes. Used **MSE** to **check accuracy** of model.

E. CONCLUSIONS.

I. FROM EDA

We found the correlation between different features using scatter plots, and heatmaps.

We plotted the choropleth maps for the different features around the globe.

We analysed the correlations of different variables and made conclusions on what variables might affect the Happiness index of a country.

We plotted box plots to see the range of values of the features of the dataset.

II. FROM PCA.

We scaled the independent variables and applied PCA to see the most prominent features.

We found the loading scores of the principal components.

We plotted the relationship between Principal components and loading scores, which helps us determine which feature affects that principal component more.

We found out that GDP, social support and Healthy life expectancy affect the happiness index the highest.

III. FROM MULTIPLE LINEAR REGRESSION

We built the multiple linear regression model to predict the Happiness Score with 2016, 2017, 2018 for training of the model and 2019 for the testing of the model.

We found the slope and the intercepts of the model's equation, which helps us determine which features affect the Score largely based on the slopes.

We found the mean squared error of the model to check its accuracy.

WE then found that the 5 countries with the largest HI depending on the Economy(GDP) and Trust in government.

We found that the 5 countries with the most consistent HI over the span of 3 years and the features that vary the most over the span of 3 years.

