



## **What came first, Happiness or Freedom?**

A Data Driven study of the relationship between Human Freedoms and Happiness

Data Viz and Management Course Project

Student's Name: Mashaël Alzaid

Student's ID: 18215041

## Introduction:

Studying the major factors of happiness and human freedom studies have always been a hot area of research for its direct impact on peoples' lives. Ruut Veenhoven studied the relationship between freedoms and happiness in 46 nations in the early 1990's and showed in his research that freedom does not always breed happiness, and suggested that economic freedom deserves priority (link to a paper: <https://personal.eur.nl/veenhoven/Pub2000s/2000a-full.pdf>) This project aims to utilize the available data in recent years and explore the relationship between world happiness and human freedom using Machine Learning and Visualization.

## Dataset:

This project used 3 datasets (Human Freedom Index, World Happiness Dataset (2015) and World Happiness Dataset (2016)).

Human Freedom Index Dataset was conducted by CATO institute and other organizations, The report used 2016 as the latest year because it is the most recent year for which sufficient data are available. There are 37 variables covering 162 countries presented in the Economic Freedom of the World report. **(Data is available on: <https://www.kaggle.com/gsutters/the-human-freedom-index>)**. On the other hands, World Happiness Dataset which was conducted by the united nations, the data is extracted from Gallup World Poll, where a survey that ask people to evaluate certain features in a scale of 10, where 0 is worst and 10 is best **(Data is available on: <https://www.kaggle.com/unsdsn/world-happiness>)**

## Process:

**(Code and Viz are available on: [https://github.com/mashaalzaid/World\\_happines\\_Vs\\_Freedom/blob/master/HF\\_WH.ipynb](https://github.com/mashaalzaid/World_happines_Vs_Freedom/blob/master/HF_WH.ipynb))**

The project is divided into 2 sections. Exploratory Data Analysis and Visualization:

1- EDA (Exploratory Data Analysis)

Exploratory data analysis (EDA) aims to analyze to briefly describe the datasets, it was used for both used datasets.

### 1.a: Human Freedom Dataset

The HF dataset was installed as a CSV file and imported using Panda library to be used in the notebook. The dataset consists of 1272 rows and 8 columns ('Year', 'ISO\_Code', 'Countries', 'PERSONAL FREEDOM (Score)', 'ECONOMIC FREEDOM (Score)', 'HUMAN FREEDOM (Score)', 'HUMAN FREEDOM (Rank)', 'HUMAN FREEDOM (Quartile)') which were renamed for the purpose of conveniently calling them during the process of analysis.

#### 1.a.1 Data Exploratory

##### 1.a.1.1 box plot of Human Freedom Dataset distribution

For the purpose of checking the distribution of the dataset, this plot shows distribution of the dataset numerical columns and as illustrated in the figure below, only the Personal Freedom Score is skewed to the left, however Economic Freedom Score and Human Freedom Score are compressed, while Human Freedom Quartile was evenly distributed.

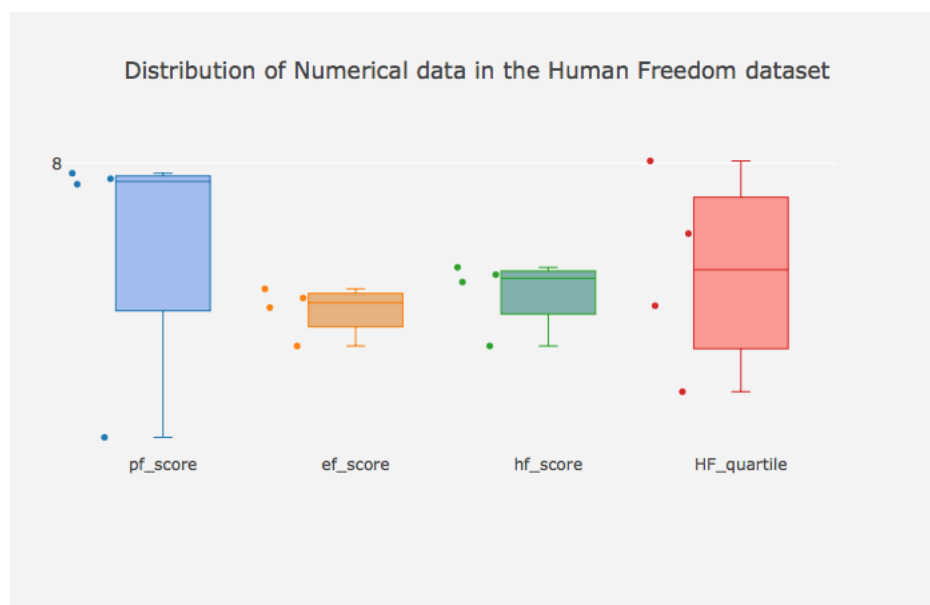


Fig1: box plot of Human Freedom Dataset distribution

### 1.a.2 Preprocessing

In this part, I filled the NAs by the mean value of each feature, however for columns with NAs above 30% I deleted it.

### 1.a.3 Target Feature Extraction

In the part also I divided the countries to 3 classes (High, Medium and Low level of freedoms)

### 1.b World Happiness Dataset

The WH dataset was installed from links and imported using Panda library to be used in the notebook. The dataset consists of 315 rows and 12 columns ('Country', 'Region', 'Happiness Rank', 'Happiness Score', 'Economy (GDP per Capita)', 'Family', 'Health (Life Expectancy)', 'Freedom', 'Trust (Government Corruption)', 'Generosity', 'Dystopia Residual', 'year']) which were renamed for the purpose of conveniently calling them during the process of analysis.

#### 1.b.1 Data Exploratory

##### 1.b.1.1 box plot of dataset distribution

For the purpose of checking the distribution of the dataset, this plot shows distribution of the dataset numerical columns and as illustrated in the figure below, all the features are evenly distributed.

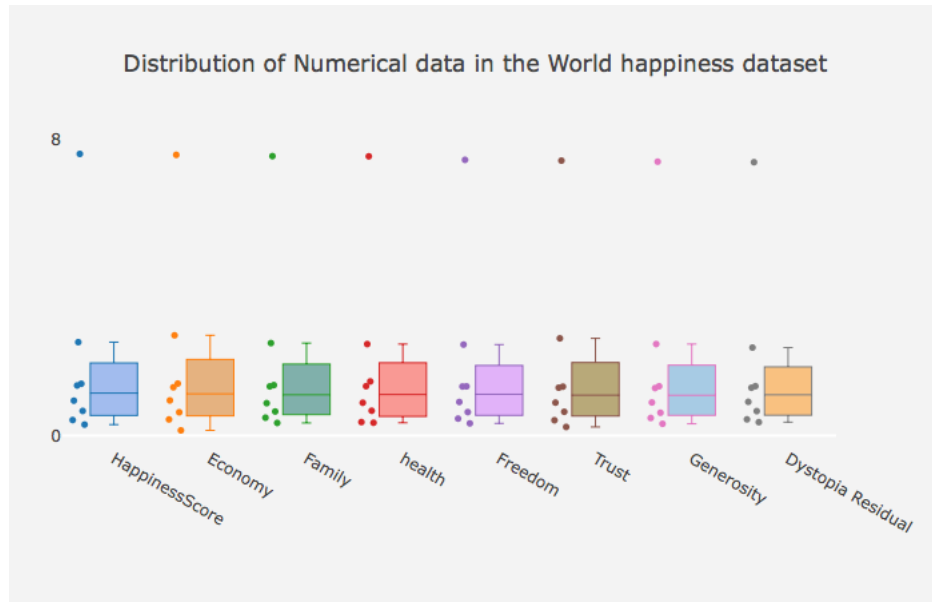


Fig2: box plot of World Happiness Dataset distribution

### 1.b.2 Preprocessing

In this part, I filled the NAs by the mean value of each feature, however for columns with NAs above 30% I deleted it.

### 1.b.3 Target Feature Extraction

In the part also I divided the countries to 3 classes (High, Medium and Low level of Happiness)

## 2- Data Visualization

In this project I used python and plotly to produce interactive visualizations. The graphs aim to answer specific questions, and the type and coloring and settings are decided accordingly. I used 4 types of graphs (Scatter, Bubble Radar plots and a Heatmap)

### 2.a Scatter plot: How Generosity affects World Happiness Score?

This graph aims to understand the relationship between Generosity and Happiness and whether these two factors impact each other.

The x-axis represents Happiness Score where y-axis represents Generosity.

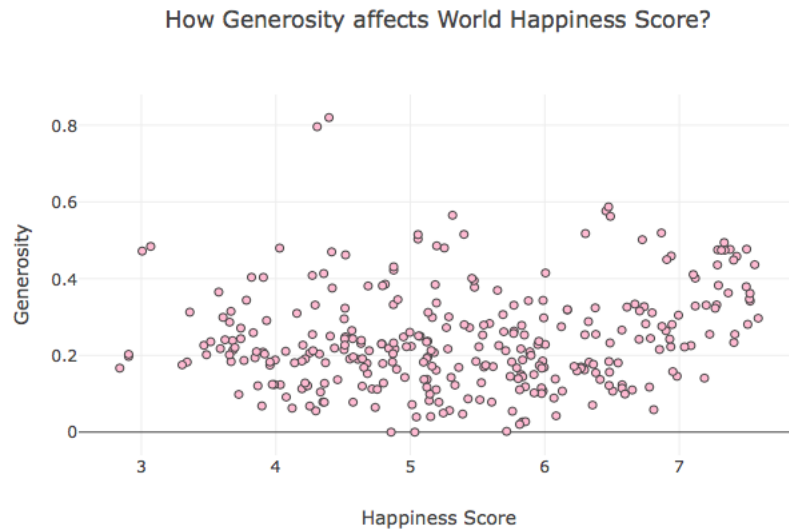


Fig3: Scatter plot: How Generosity affects World Happiness Score?

## 2.b bubble plot: Does Health threaten Happiness?

The second graph is a bubble plot that shows the change in health with respect to happiness score in 2015 and 2016. The countries are colored based on region. The plot allows the user to switch on or off the region and focus on the ones under study.

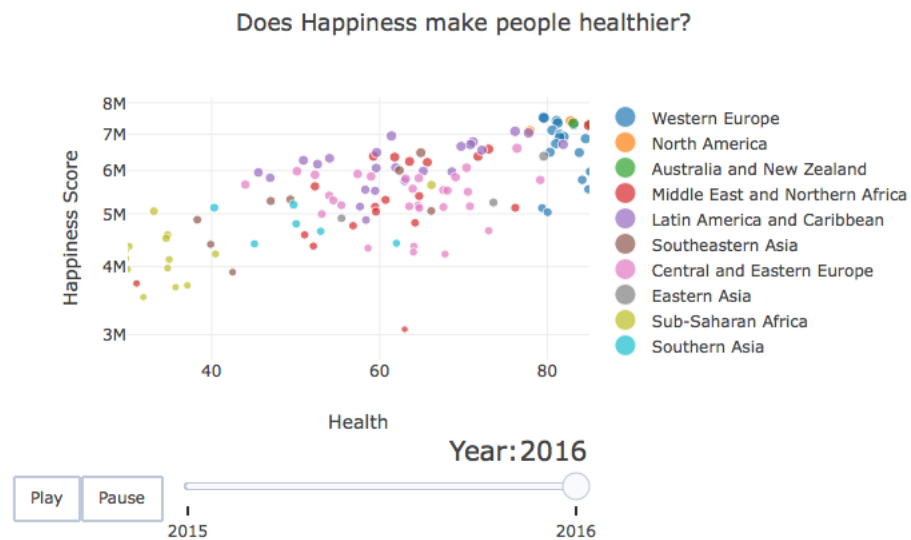


Fig4: bubble plot: Does Health threaten Happiness?

2.c Radar plot: How different scores of freedoms affects world happiness?

In this plot, and as I mentioned in the Target feature section, I divided the countries to high, med and low level of happiness countries. The goal of this graph is to link between these classes of countries and study the limit of freedoms they have.

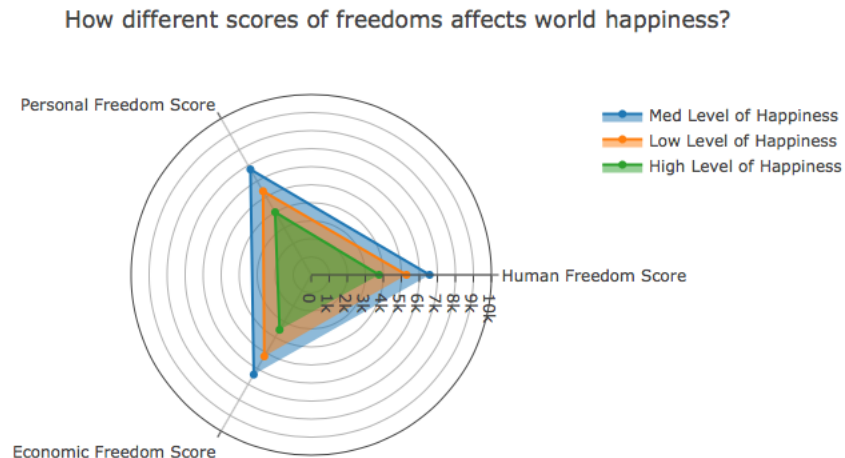


Fig5: Radar plot: How different scores of freedoms affects world happiness?

2.d Heatmap: How world happiness factors and human freedom factors correlate?

The heatmap aims to study the correlation between the human freedom factors and the world happiness factors. The darker the color of the cell the more correlation between the 2 factors that represent that specific cell and vice versa.

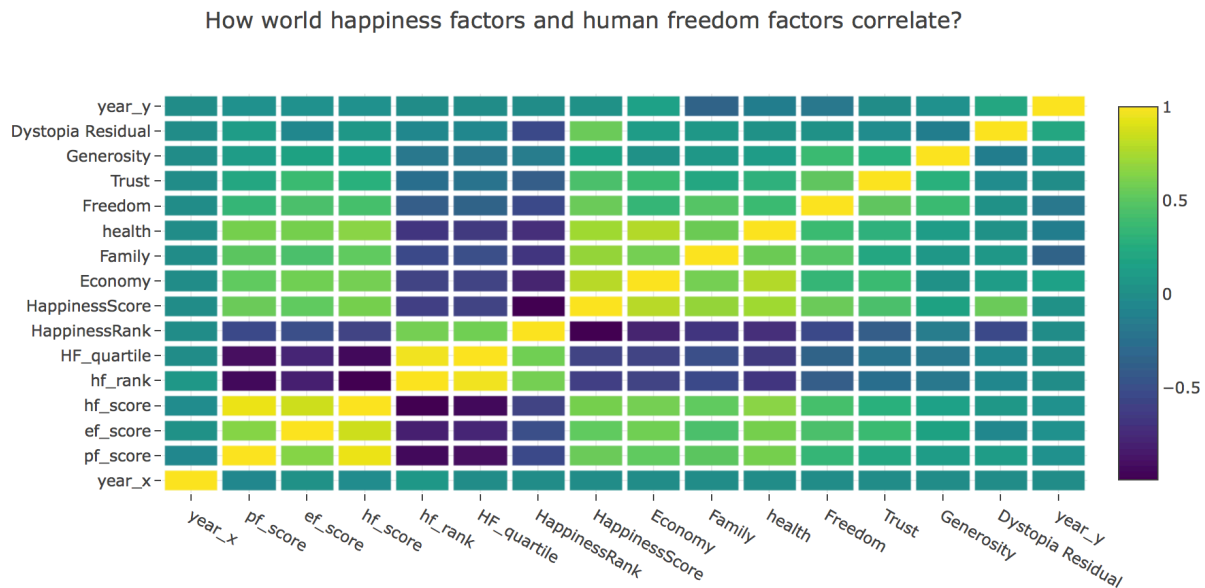


Fig6: Heatmap: How world happiness factors and human freedom factors correlate?

### Result:

(Video is available on: <https://youtu.be/QuNC9iVIP5w>)

This project used plotly and python and a tiny experiment with D3.js. The graphs were chosen carefully based on the data and the desired question. Below I will analyse the graphs, and its intakes as well as recommendations.

First, as can be seen in plot 2.a (fig3), there's a linear relationship between the two features however the strength of the two features appears to be weak meaning that generosity of a country doesn't breed happiness for its nation and vice versa. The graph could be improved by adding more dimensions that describe the situation more precisely. However that could possibly increase the complexity of the graph. This graph uses the Proximity principle from Gestalt Laws

In 2.b plot (fig4) we can see a high positive correlation between health and happiness, meaning that with more happiness countries are more likely to be healthier and vice versa. However, it was noticed that some countries show a decrease in one factor and increase in the other such as Malaysia between 2015 and 2016. The Plot is well-organized nonetheless due to the lack of data the graph shows results from 2 years and adding more years would make it more meaningful. This graph uses the Proximity principle from Gestalt Laws.



The plot in section 2.c (fig5) shows that all the classes of countries have Economic Freedom that equals Personal Freedom that equals Human Freedom. However, what is surprising in this graph is the fact that Medium Happy countries have more freedom than High Happy countries. This encourages more investigation to understand the hidden factors that affect the world's happiness and freedoms. The graph is organized and clear and the legends could be minimized or maximized based on the class of the country. The graph could be presented more clearly with 3D visualization. This graph uses the similarity principle from Gestalt Laws.

The heat map in section 2.d (fig6) shows that 3 factors are dominating all the other factors (Human Freedom rank, Human Freedom quartile and Happiness Score). This is a good plot and a good selection, nonetheless such a graph suggests to come with a case study from the real world. This is a call for sociologists who will read this work to study further these findings in further details. This graph uses the similarity principle from Gestalt Laws.

### **Conclusion:**

We saw in the report a data-driven study that analyzes the relationship between human freedoms and world happiness. The project presented the Preprocessing the visualization part as well as the intake from the visualizations.

The project raises an important question: Does this say that the less freedom people are more likely to be happy?

It is recommended that sociologists utilize this study and conduct further research on the raised questions to deeply understand the human and the societies.