April 27th, 2023

CS555 – Term Project

Zuowen Tang

Data Science Research

On Humanity and Happiness

Based on World Happiness Report from 2017-2019

# Introduction

Every individual might have different goals, ambitions, and purposes in their life, yet one thing is what we all want: happiness. No matter who you are, an intelligent young scientist who wishes to work at NASA, a beautiful dancer who wants to be on a big stage, or an old gentleman who only wishes to spend more time with his family, happiness is always the thing that attracts us. Today we live in a society where everything seems able to be quantified, knowledge, feelings, personal information, and maybe happiness as well. Starting in 2011, the World Happiness Report began to gather nationwide data based on scientific facts and *subjective ratings from respondents based on their own life[[1]](#footnote-1),* with various life factors, to see which country is the happiest. On the other hand, this data set could also answer these questions: what determines people's happiness? Under what circumstances can we live the best life? Also, based on the data set we already have, can we predict happiness in the future?

# Data Set Description

This data science research is based on the dataset of the World Happiness Report from 2017-2019, which can be accessed on the public data set website *Kaggle* and *WHR's webpage*. There are six main variables that are considered important to people's life. *They include GDP per capita, social support, healthy life expectancy, freedom, generosity, and corruption[[2]](#footnote-2).*

This study will mainly consider variables such as **GDP per Capita**, **Social Support**, and **Healthy Life Expectancy** as the dominant factors since they strongly correlate with the overall score. **GDP per Capita** is *an economic metric that breaks down a country's economic output per person[[3]](#footnote-3).* **Social Support**, also known as family support in the past data set report, is more personal and subjective. It means how much emotional support a person can receive from their family/friends. This variable is rated by respondents based on their experience. **Healthy Life Expectancy** means *the average number of years that a person can expect to live in full health[[4]](#footnote-4)*. The other three variables, **Freedom** *(whether a person can make a decision freely),* **Generosity** *(how much/often people will donate to charity),* and **Corruption** *(the political corruption or the abuse of power based on people's experience),* will only be mentioned in very few specific tests.

# Methods Introduction

The following statistical methods will be used in this report to help the writer draw connections between factors and form a conclusion: two-sample mean tests, correlation tests, simple and multiple linear regression, and ANOVA tests. Diagrams, such as histograms, box plots, scatter, and three-dimensional plots, will also be used in pursuance of virtualizing the dataset. Furthermore, machine learning algorithms, for instance, Linear Discriminant Analysis, Quadratic Discriminant Analysis, and Naive Bayes Classifier, will be used for the predictive analysis.

1. *FAQ | The World Happiness Report*. (n.d.). Home | The World Happiness Report. [↑](#footnote-ref-1)
2. *About | The World Happiness Report*. (n.d.). Home | The World Happiness Report. [↑](#footnote-ref-2)
3. The Investopedia Team. (2009, July 8). *GDP Per Capita Defined: Applications and Highest Per Country*. [↑](#footnote-ref-3)
4. PhD, M. S. (2007, May 30). *Healthy Life Expectancy and How It’s Calculated*. Verywell Health; Verywell Health. [↑](#footnote-ref-4)