

WORLD HAPPINESS RANKINGS AND LIFE EVALUATION FACTORS

DS 4002 | Group 10 | Dec. 1, 2023

Project Lead: Zoe Averill
Sujith Panchumarthy
Carson Crenshaw

Data Acquisition

Collecting, cleaning,
finalizing dataset

Analysis

Build graphics,
regression analysis &
model, prediction

Results

Addressing the
hypothesis and future
steps

Motivation for Study



Contemporary society has increasingly prioritized people's quality of life



Governments are interested in which factors influence happiness to guide policy and improve public welfare

Research Questions

- *If a correlation can be established between happiness and life evaluations, can it accurately predict a country's happiness in the near future?*
- *How can a predictive model for happiness be leveraged by policymakers in the implementation of new national programs?*



Hypothesis

Conducting an exploratory data analysis and building a regression model will demonstrate that freedom is one of the most significant factors that influences a country's happiness score.

Using the most significant life evaluation factors, an accurate model will be able to correctly predict the happiness scores of 2022.



Time-Series Regression Analysis & Model

Data Acquisition

Original dataset – United Nation's World Happiness Report

- Begins in early **2000s**.
- Scores for **all** internationally recognized countries.
- **Surveys** of individuals' assessment of their own lives.

Column Name	Description	Data Type
Country	Name of the Country	String
Year	Year for the data provided	Integer
Happiness Score	Average reported happiness score from 0 to 10. Lower score indicates less happiness.	Float
Log GDP per Cap	Logarithmic GDP per capita determined by WHR	Float
Social Support	Social support score from 0 to 1. Lower score indicates less social support as determined by WHR	Float
Life Expectancy	Life expectancy at birth in years	Float
Freedom	Freedom to make life choices score from 0 to 1. Lower score indicates less freedom as determined by WHR	Float
Generosity	Generosity score where a lower score indicates less generosity as determined by WHR	Float
Corruption	Citizens' perception of government corruption score from 0 to 1. Lower score indicates lesser belief of corruption as determined by WHR	Float
Continent	Corresponding continent for the country	String

Cleaned WHR.csv file. Organization of columns/variables. Long format data.

Finalized dataset after cleaning

- **2013-2019**
- **103** countries.
- Complete reference of data for time period.
- **More in tricky decision...**

Analysis Plan

Engineering Component

Exploratory
data analysis

With respect to
countries' happiness
scores and 6 life
evaluation variables

Build
interactive
graphics

Allow users
to investigate

Analytic Component

Regression
analysis of
happiness on life
evaluation metrics
& model

Built from
conclusions derived
from graphical
representations
→ Develop model

Determine
accuracy of
model

Evaluate on
cleaned dataset
and reserved
2022 data

Predict future
happiness
scores

Use most
significant factors
to predict

Life evaluation variables:
GDP, Social Support, Life Expectancy, Freedom,
Generosity, Corruption (Gov. trust)

Tricky Analysis Decision: Data Choices

Initially, our group wanted to prioritize working with as much complete data over the largest timespan as possible. We finally reduced our discussion to two options:



OR



Option 1

2010-2019

84
countries

Option 2

2013-2019

103
countries

Note: 2020 and 2021 were exempted because of COVID-19 potentially skewing data.

Additionally, we could only gather information for 103 countries because of limited information sharing.

Give up 3 years or 19 countries?

OPTION 2

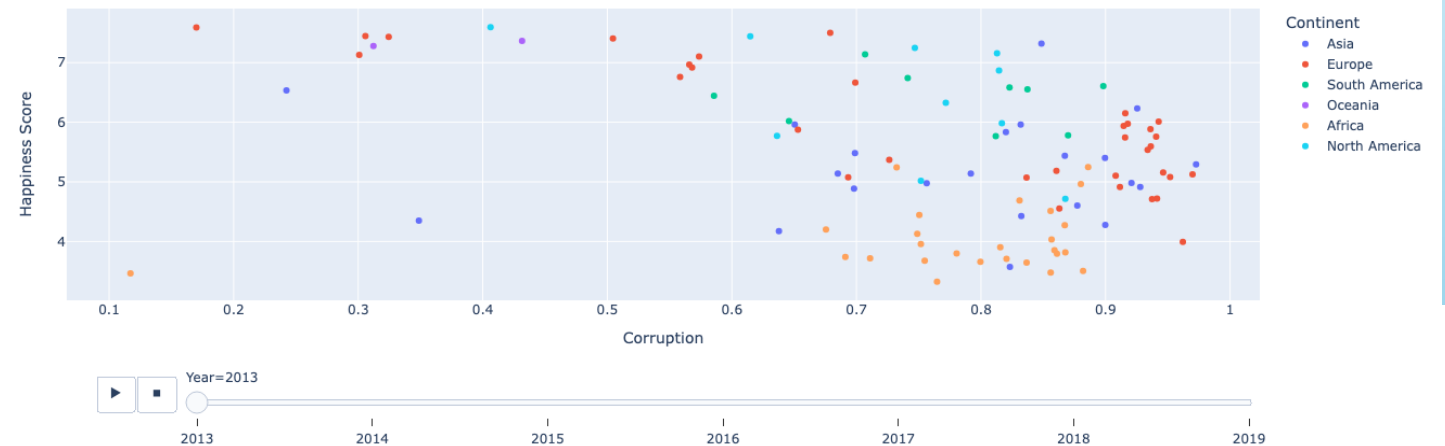
Bias & Uncertainty Validation

- Only data for 7 years and 104 countries
- Omitted countries loosely representative of third world countries → Conclusions are drawn from a dataset skewed towards industrialized countries
- After exploratory analysis, removed generosity and corruption as factors for our model because they were not linear relationships
 - If included, made model perform worse

Happiness Score vs. Generosity



Happiness Score vs. Corruption



Results and Conclusions

$$\text{Happiness Score} = 0.368 * \text{Life Expectancy} + 0.297 * \text{GDP} + 0.276 * \text{Freedom} + 0.247 * \text{Social Support} + 5.507$$

Note: When interpreting the beta values, the factors represent standard deviations (for ease of comparing importance).

Model Interpretations

- Life Expectancy is most important factor while Social Support is least
- Comparatively, Life Expectancy is 49% more impactful than Social Support. GDP & Freedom are 20%, and 12% respectively.
- RMSE = 0.58



2022 Predictions

For the data on 104 countries in 2022, the model's predictions resulted in an RMSE of 0.6

Since the model and 2022 predictions RMSE's are extremely similar, the model can confidently be extrapolated to years beyond the training data

Next Steps

Further Testing

Test prediction accuracy on other years outside of 2013-2019

Improve Data & Model Quality

Collect data for missing countries and devise a method to include Generosity and Corruption in model

Government Policy

Guide policy to improve public welfare by focusing on the factors that make people the happiest



References

- [1] “About | The World Happiness Report.” Available: <https://worldhappiness.report/about/>. [Accessed Nov. 5, 2023].

- [2] “Home | The World Happiness Report,” Jun. 20, 2023. Available: <https://worldhappiness.report/>. [Accessed Nov. 12, 2023].

- [3] C. Crenshaw, “WHR,” *GitHub*, Nov. 12, 2023. [Online]. Available: https://github.com/C-Crenshaw/Project3_DS4002/blob/8865410d04c3de81929f88c549ff6bd3455ceb5c/WHR.csv . [Accessed Nov. 12, 2023].

- [4] “Foreword | The World Happiness Report,” Mar. 18, 2022. Available: <https://worldhappiness.report/ed/2022/foreword>. [Accessed Nov. 5, 2023].

- [5] J.F. Helliwell, R. Layard, and J.D. Sachs, “World Happiness Report 2021,” WellBeing International., Sustainable Development Solutions Network, New York, 2021.

[Link to MI1 Doc](#)

[Link to MI2 Doc](#)

