

Milestone 1 – Project Proposal

The purpose of this document is to define the second project of the DSC 680 Applied Data Science course. The business problem, data set, methodology, ethical considerations, and challenges/issues will be overviewed.

Topic

Project Name: World Happiness Regression Model

This project focuses on generating a regression model to predict levels of happiness for countries around the world.

Business Problem

Since 2012, the Sustainable Development Solutions Network has been publishing the *World Happiness Report* every year around March 20 in recognition of the International Day of Happiness. The *World Happiness Report*, “reflects a worldwide demand for more attention to happiness and well-being criteria for government policy” (2023). A thorough understanding of happiness around the globe is important for several reasons. First, the underlying factors for the happiest or unhappiest countries can be compared. Leaders within each country can review their current policies to understand areas to improve for a happier country. Second, major events such as wars or pandemics may have a significant impact on mental health. Leaders within countries with sharp declines in happiness for their citizens can act when needed. Third, many scientific studies indicate great benefits for individuals that tend to be happier compared to peers. Some of these benefits include improved heart health, handling stress more effectively, better immune systems, pain reduction, healthy lifestyles, and increased life expectancy (Mead, 2019). Although the focus for this project will consider happiness by country, it is also important to recognize the benefits of happiness on an individual level. The primary objective for this project is to create a regression model to predict the happiness level of countries. The research questions related to this analysis are shown below:

- Which countries are the happiest?
- Which countries are the unhappiest?
- Which features are most significant for the happiest countries?
- Which model provides the best accuracy for predicting a country’s happiness level?

Datasets

The data for this analysis is from Gallup World Poll surveys used for the annual *World Happiness Report*. The data is comprised of survey results rolled out to sample populations across more than 100 different countries. The survey has been sent out since 2012, however only data from 2015-2023 will be considered for the analysis. The target feature for the analysis is the Happiness Rating. According to the Sustainable Development Solutions Network, the Happiness Ratings “are based on answers to the main life evaluation question. The Cantril ladder asks respondents to think of a ladder, with the best possible life for them being a 10 and the worst possible life being a 0. They are then asked to rate their own current lives on that 0 to 10 scale. The rankings are from nationally representative samples over three years” (2023). There are six additional variables included in the data for each country: GDP per capita,

freedom, corruption, generosity, social support, and life expectancy. The first link below connects to the World Happiness Reports sent out each year (except for 2014). The data is broken down as separate CSV files for each year and may include additional features. The additional links below reference the CSV files for the World Happiness Report, except from Kaggle.

World Happiness Report - [Read the Reports | The World Happiness Report](#)

World Happiness Report CSV's 2015-2022 (Kaggle) - [World Happiness Report up to 2022 | Kaggle](#)

World Happiness Report CSV's 2023 (Kaggle) - [World Happiness Report 2023 | Kaggle](#)

The details about the main features within the datasets are shown in the table below:

Table 1: Main Feature Overview (as described by Sustainable Development Solutions Network)

Feature	Feature Description by the Sustainable Development Solutions Network (2023)
Happiness Rating (Ladder Score)	Happiness score based on sample of survey responses from a scale of 0 (not happy) to 10 (perfectly happy).
GDP per Capita	Gross Domestic Product (GDP) per capita, or how much each country produces, is divided by the number of people in the country. GDP per capita gives information about the size of the economy and how the economy is performing.
Social Support	Having someone to count on in times of trouble. "If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?"
Healthy Life Expectancy	More than life expectancy, how is your physical and mental health? Mental health is a key component of subjective well-being and is also a risk factor for future physical health and longevity. Mental health influences and drives a number of individual choices, behaviors, and outcomes.
Freedom to make Life Choices	"Are you satisfied or dissatisfied with your freedom to choose what you do with your life?" This also includes Human Rights. Inherent to all human beings, regardless of race, sex, nationality, ethnicity, language, religion, or any other status. Human rights include the right to life and liberty, freedom from slavery and torture, freedom of opinion and expression, the right to work and education, and many more. Everyone is entitled to these rights without discrimination.
Generosity	"Have you donated money to a charity in the past month?" A clear marker for a sense of positive community engagement and a central way that humans connect with each other. Research shows that in all cultures, starting in early childhood, people are drawn to behaviors which benefit other people.
Perception of Corruption	"Is corruption widespread throughout the government or not" and "Is corruption widespread within businesses or not?" Do people trust their governments and have trust in the benevolence of others?

There is another feature called Dystopia that represents the score for the unhappiest fictitious country in the world. This feature may not be included in the analysis, but it is yet to be decided. Additional features such as Year and Rank will be added to each data set if they are not already included. The data will then be merged into one file and used for the remainder of the analysis. Any additional data used for the analysis will be identified and included later in the project.

Methods

The Cross Industry Standard Process for Data Mining (CRISP-DM) will be followed for this project. The high-level phases for this process include Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The Business Understanding phase was considered above in the Business Problem section. To recap, this project aims to create a regression model to predict the happiness level of countries around the world. The data is in many different locations, so consolidating into one location will be the initial focus. A thorough understanding of the data will follow with univariate and bivariate analysis. Additional Exploratory Data Analysis will be performed as needed. The data will then be prepared for the regression model. The initial models chosen for the analysis are Linear Regression, Decision Tree Regressor, and Random Forest Regressor. This may update/change as the analysis progresses. The evaluation metrics that will be used to compare model performance will be R^2 (test and train), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE). A recommendation will then be made whether to deploy the model that performs the best.

Ethical Considerations

Ethics are essential when working on any type of data science task or project. All steps for data consolidation and cleaning will be clearly communicated throughout the project. Data privacy concerns were considered in this analysis. The initial survey data from individuals is not included in this project. The data comes from the published reports from the Gallop World Poll surveys and *World Happiness Report* from the Sustainable Development Solutions Network. The actions generated from this project are intended to show country leaders how to improve happiness in their countries. There is no intent to compare countries from a competitive standpoint, rather offer an opportunity for country leaders to recognize opportunities to make a significant impact on moral around the globe. All resources utilized in this project are outlined in the references section to provide appropriate credit.

Challenges/Issues

The first challenge for this project is that the data is spread out across multiple locations. The data will be consolidated in a single file to simplify the analysis and make it easier for individuals wanting to expand from this project in the future. The second challenge is the risk that there are no useful insights present for this analysis. It is yet to be determined, but a major goal is to find the driving factors to predict happiness for countries around the world. An alternative challenge along the same lines is considering the consequences for the useful actions recommended from this project. Even if happiness is improved from certain actions, will it have an impact on other critical features for the country such as GDP per capita. Another challenge is whether the model that performs the best is suitable for deployment. This will depend on the performance of the model and will need to be reviewed as the analysis progresses. Lastly, buy-in from leaders within countries will most likely be the biggest challenge for this project. Recommendations can be made with support from the data; however it is up to leaders within countries to listen and take the necessary action.

References

Ache, M. (2022). *World Happiness Report up to 2022*. About Dataset. Retrieved July 08, 2023, from <https://www.kaggle.com/datasets/mathurinache/world-happiness-report?select=2022.csv>

DSC 680 Applied Data Science (2237-1)

Bellevue University

Professor Williams

Milestone 2 – Project Proposal

Jake Meyer

07/08/2023

Hunter, M. (2023, March 20). *The world's happiest countries for 2023*. CNN Travel. Retrieved July 07, 2023, from <https://www.cnn.com/travel/article/world-happiest-countries-2023-wellness/index.html>

Global Finance. (2022, April 27). *The Happiest Countries in the World*. Global Data. Retrieved July 07, 2023, from <https://www.gfmag.com/global-data/non-economic-data/happiest-countries#:~:text=High%20GDP%20per%20capita%2C%20social%20support%20in%20times,time%20in%20their%20report%20on%20global%20life%20satisfaction.>

Mead, E. (2019, June 5). *6 Benefits of Happiness According to the Research*. Happiness & SWB. Retrieved July 07, 2023, from <https://positivepsychology.com/benefits-of-happiness/>

Rozzelle, J. (2023, June 28). *Global Unhappiness Levels in 2022 Match All-Time High, Report Finds*. U.S. News World Report. Retrieved July 07, 2023, from <https://www.cnn.com/travel/article/world-happiest-countries-2023-wellness/index.html>

Singh, A. (2023). *World Happiness Report 2023*. About Dataset. Retrieved July 08, 2023, from <https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2023>

Sustainable Development Solutions Network. (2023). *World Happiness Report*. About Dataset. Retrieved July 07, 2023, from <https://www.kaggle.com/datasets/unsdsn/world-happiness?select=2015.csv>

World Happiness Report and Gallup World Poll. (2023). *World Happiness Report*. About. Retrieved July 07, 2023, from <https://worldhappiness.report/about/>

Appendix – Milestone Template Criteria

Each Milestone must follow the following template

Milestone 1 - Proposal

Submit a proposal with the following topics covered as a PDF

- Topic - Describe and name your project in 1-2 sentences max
- Business Problem - Describe the business problem your project is trying to solve and/or the research questions you will explore
- Datasets - where are you getting your data? Describe the data that you will use to solve the problem
- Methods - What analysis methods will you use to complete this project? Note this is just a proposal, your project can adapt as you work on it
- Ethical Considerations - What are some potential ethical concerns of this topic or analyzing the data?
- Challenges/Issues - What are some issues and challenges do you think you might face?
- References - What sources will you use to validate your results and support your project topic?

Milestone 2 - Draft White Paper

Submit a draft of your white paper with the following topics covered as a PDF

- Business Problem
- Background/History
- Data Explanation (Data Prep/Data Dictionary/etc)
- Methods
- Analysis
- Conclusion
- Assumptions
- Limitations
- Challenges
- Future Uses/Additional Applications
- Recommendations
- Implementation Plan
- Ethical Assessment