**Final Project Midpoint Submission**
MSDS 422, Summer 2024 Diana Rodos

*World Happiness Report: Gaining Insights on Happiness*

## Executive Summary

In this project, we analyzed the World Happiness Report dataset to explore the factors contributing to the happiness of various countries. Our objectives included understanding the relationships between different socioeconomic and health-related variables and the overall happiness (Life Ladder) score and developing visualizations to communicate these insights effectively. We prepared the dataset for advanced analysis and modeling through Exploratory Data Analysis (EDA), data preparation, and feature engineering.

## Problem Statement/Research Objectives

The primary objective of this research was to analyze the World Happiness Report dataset to identify key factors that influence the happiness scores of countries worldwide. Specifically, we aimed to:

1. Understand the distribution and trends of happiness scores over time and across regions.
2. Identify the most influential factors contributing to the happiness scores.
3. Explore the relationships between socioeconomic indicators and happiness.
4. Create insightful visualizations to communicate the findings effectively.

## Exploratory Data Analysis

We began our analysis by loading the dataset and conducting a thorough Exploratory Data Analysis (EDA). Key steps included examining the dataset's structure, summary statistics, and missing values. Missing values were addressed by imputing them with the mean values specific to each country. We calculated Descriptive statistics and performed a correlation analysis to identify relationships between the Life Ladder (happiness score) and other variables.
Key Findings from EDA:

- **Happiness Scores by Region**: We visualized the distribution of happiness scores by region using boxplots and KDE plots.

- **Trends Over Time**: We plotted the happiness scores over time for different regions to observe trends and changes.
- **Correlation Insights**: We found that Log GDP per capita, Social support, and Healthy life expectancy at birth had the highest positive correlations with the Life Ladder score, while Perceptions of corruption and Negative affect had negative correlations.

## Data Preparation/Feature Engineering

We imputed missing values with country-specific means to enhance our analysis and model performance. We created a new feature, the Well-being Index, by combining key socioeconomic indicators such as Log GDP per capita, Social support, and Healthy life expectancy at birth. We also standardized the numerical features to ensure they were comparable. Additionally, we mapped each country to its respective region by adding a regional indicator column.

## Visualizations

We created various visualizations to communicate our findings effectively. A boxplot of happiness scores by region illustrated the distribution of scores across different regions. A KDE plot displayed the density of happiness scores for each region, highlighting regional variations. We also plotted the change in happiness scores over time for different regions, providing insights into temporal trends. A correlation heatmap visualized the relationships between the Life Ladder score and other variables. A scatter plot of the Well-being Index versus Life Ladder score showed the relationship between well-being and happiness, color-coding by region. Lastly, we created a geographical visualization that mapped the Well-being Index across regions, providing a global perspective on happiness and well-being.

## Insights

Our analysis revealed that Western Europe consistently had the highest happiness scores, with countries like Finland, Denmark, and Switzerland leading the rankings. In contrast, Sub-Saharan Africa had the lowest happiness scores, with countries such as South Sudan and Zimbabwe at the bottom. Economic stability, social support, and health are critical factors influencing happiness. Countries with higher GDP per capita, better social support systems, and higher life expectancy tend to have higher happiness scores. Positive emotions and lower perceptions of corruption were also associated with higher happiness levels.

Link to dataset: https://www.kaggle.com/code/jainaru/world-happiness-rankings-complete-eda