DATA SCIENCE PROJECT

World Happiness Report 2023

Efe Altın / Mehmet Hasat Serinkan

Why We Choose This Dataset :

1. Rich Data Source: The World Happiness Report 2023 dataset contains a vast amount of data related to various aspects of happiness, including Logged GDP per Capita, Social Support, Healthy Life Expectancy, Genorosity and Perceptions of Corruption, which can be analyzed and explored through various data science techniques. The dataset also includes data on different countries which providing a broad range of perspectives to explore
2. Relevance: Happiness is a core human goal Therefore, analyzing data from the World Happiness Report can provide valuable insights into the factors that contribute to happiness and inform policies aimed at improving people's quality of life.
3. Practical Applications: this report can also be useful in practice. For example, social welfare organizations can use the findings to design interventions that support vulnerable populations.

Describing the Data:

There are 13 columns in this dataset and these are:

1. Country Name
2. Ladder Score
3. Standart Error of Ladder Score
4. Upperwhisker
5. Lowerwhisker
6. Logged GDP per Capita
7. Social Support
8. Healthy Life Expectancy
9. Freedom to Make Life Choices
10. Generosity
11. Perceptions of Corruption
12. Ladder Score in Dystopia
13. Dystopia + Residual

Why it is complex: The dataset used for the World Happiness Report includes a wide range of variables that are collected from different sources, such as international surveys and official statistics from governments and international organizations. The variables cover various aspects of life, such as social support, freedom to make life choices, and perceptions of corruption. One of the main challenges of this dataset is that it contains both objective and subjective measures of happiness and well-being. For example, income and life expectancy are objective measures, while self-reported measures of happiness and life satisfaction are subjective. Combining these different types of measures and ensuring their validity and reliability can be a complex task. Moreover, the dataset covers a large number of countries. the data may not be consistent or comparable across different countries . Therefore, preprocessing the data to ensure its quality and comparability is an essential step in analyzing this dataset.

Describing the Task:

The task of this project is to perform data analysis on the World Happiness Report 2023 dataset to gain insights and understanding of the factors that contribute to the happiness of different countries.

The project aims to explore the dataset using various data analysis techniques such as data cleaning, data visualization, statistical analysis. The analysis is performed using Python programming language and various libraries such as pandas, seaborn, and matplotlib.

The project also aims to develop a predictive model to predict the happiness score of different countries based on the available features. This is done using a linear regression model and the model performance is evaluated using various metrics such as R-squared and mean absolute error.