EXECUTIVE SUMMARY

04 October, 2023

PREPARED FOR

FML Professors

IMT Atlantique

PREPARED BY

Kehan LIU | Nam NGUYEN LE | Nour ABBOUD | Ramisa HEIDARI | Sastélvio MANUEL

IMT Atlantique

PROBLEM STATEMENT

The challenge of understanding and promoting happiness on a global scale is multifaceted, with disparities among countries and regions. Factors like access to resources, economic inequality, mental health, and external crises all play a role. Tailored interventions are needed, as different countries value different factors, such as generosity or economic situation. The World Happiness Report highlights the difficulty in comprehensively assessing and addressing these factors, and the interpretation of happiness scores is debated. A more holistic, culturally sensitive approach is needed to genuinely improve people's well-being globally.

To do so, we come up with the following question: How can a more holistic, culturally sensitive, and actionable tool lead to effectively address multifaceted factors contributing to understanding and influence the happiness among countries?

PROPOSED SOLUTION

Data science's capacity to process vast datasets enables us to gain a comprehensive and nuanced understanding of the happiness landscape. It allows us to uncover patterns, correlations, and hidden relationships within data that might otherwise remain elusive.

To achieve that, we implemented a data science project were we applied the principal steps required to extract meaningful insights:

Data Collection

We conducted an exploratory analysis using the provided dataset to uncover patterns in the happiness ranking and gain insights into the uniqueness of each country. The analysis focused on parameters like Economy (GDP per Capita), Family (Social factor), Health (Life Expectancy), Freedom, Generosity, and Trust (Perception of Government Corruption).

Data Cleaning and Preprocessing

In this step, we dealt with missing values, outliers, and data inconsistencies. We grouped the data by region and used forward fill and backward fill methods to address missing data in two specific variables (region and trust). For numerical data (trust), we calculated the mean of the column, while for categorical data (region), we searched for geographical information to fill in the gaps.

Exploratory Data Analysis (EDA)

This step is to understand the dataset's characteristics, including summary statistics, data distributions, correlations, and visualisations. To do so, we checked the correlation matrix between the columns (features or variables) to understand how they influence each other.

Image 1: Correlation matrix between all the original features

The above information suggests that GDP per Capita, Family and Life Expetancy have higher correction on the happiness.

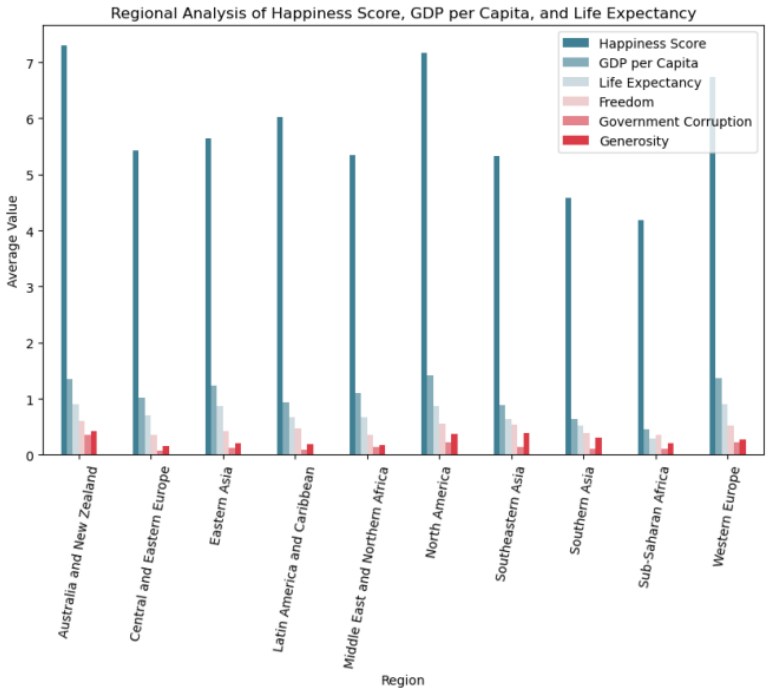

Image 2: Variables/features importance by region

On the previous graph we understand that in some regions such as Middle East and Northen Africa, the Fredom has higher impact into happiness rather than Government Corruption and Generosity, that both have almost the same importance. While Australia and New Zeland has the three aforementioned variables at the same importance.

Feature Selection/Engineering

To enhance the readability of our analysis and visualizations, we employed techniques to reduce the number of features in the dataset. One of these techniques was Principal Component Analysis (PCA), a statistical method that simplifies the dataset. Additionally, we used the TSNE Technique (t-distributed stochastic neighbor embedding) as a non-linear approach to validate our analysis, serving a similar purpose to PCA.

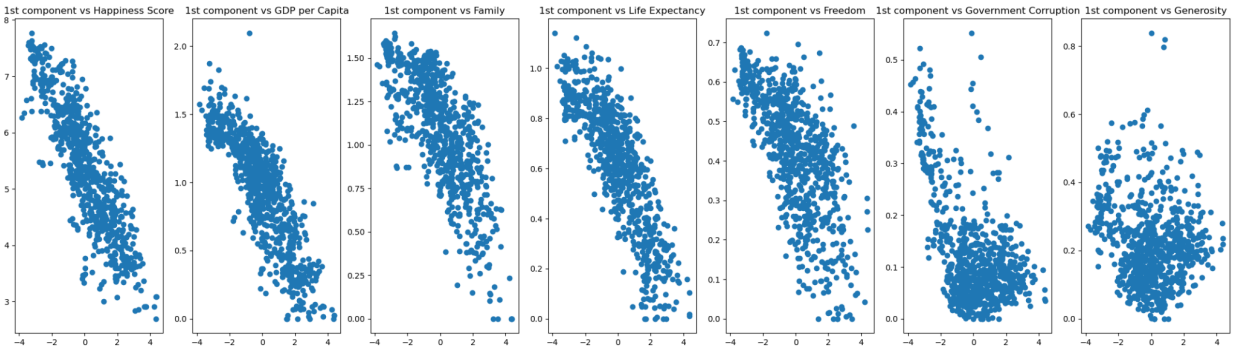

Image 3: Correlation between Component 1 (Poverty) and the original features

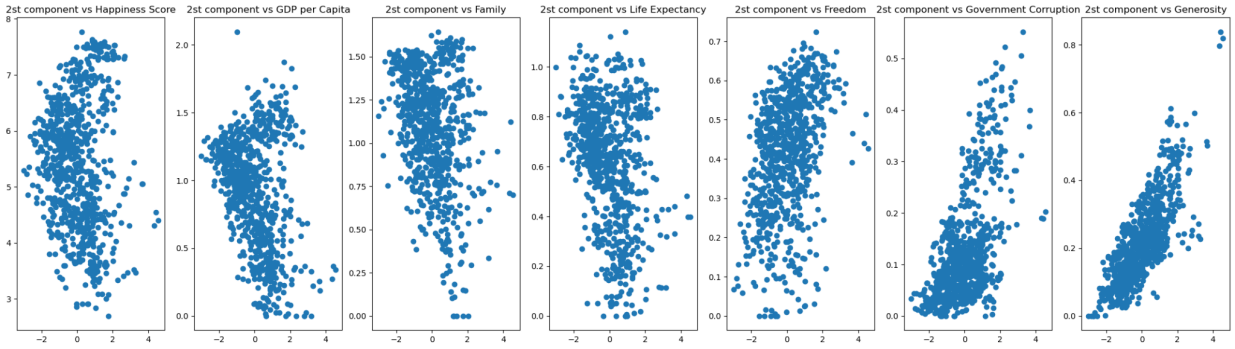

Image 4: Correlation between Component 2 (Generosity) and the original features

From the two previous visualizations, it's evident that Component 1 (Poverty), has a significantly negative correlation with the Happiness Score, GDP, Family, and Life Expectancy. This indicates that poverty plays a crucial role in a country's overall happiness. Conversely, Component 2 (Generosity) tends to have a positive influence on a country's Happiness Score. Additionally, we applied clustering methods (K-means, hierarchical clustering, Gaussian mixture model, mean shift) to group countries based on similarities and patterns within their features. This analysis revealed that: a) Rich and generous countries are often located in North America and Western Europe; b) Poor and generous countries are frequently found in Sub-Saharan Africa; c) The remaining countries, which do not fit into the previous two groups, tend to be less generous.

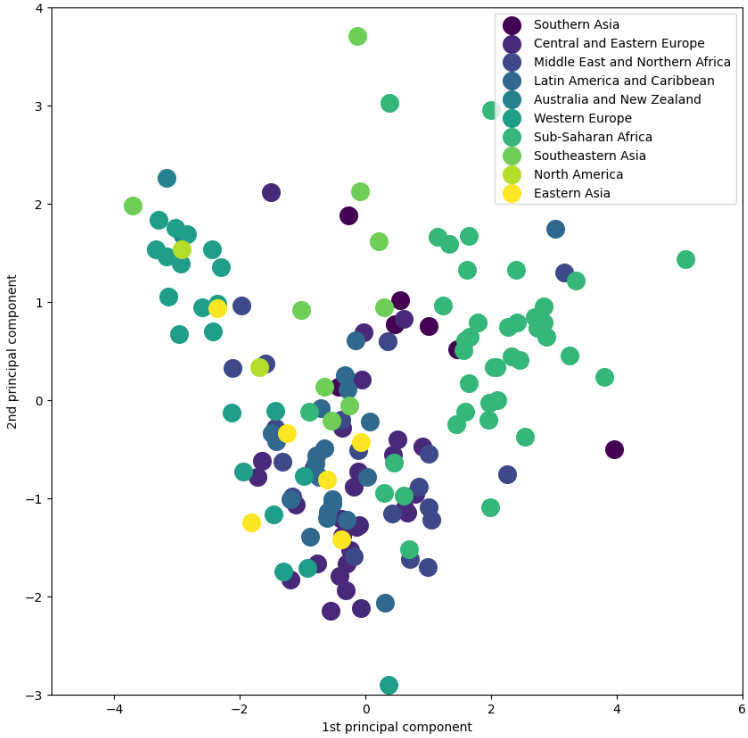

Image 4: Groups of regions based on features

VALUE OF THIS PROJECT

This study addresses the intricate and urgent challenge presented by the state of happiness in the world. With disparities in happiness levels among countries, regions, understanding and tackling this issue is imperative. Data science, with its capacity to process vast datasets, offers a formidable tool for gaining a comprehensive understanding of the happiness landscape. This analysis aimed to reveal the underlying patterns and relationships that define happiness rankings and the specific attributes of each country.

FINAL THOUGHTS AND NEXT STEPS

One notable finding is the correlation between components like "Poverty" and "Generosity" and the Happiness Score. For instance, the negative correlation of the "Poverty" component with the Happiness Score, GDP, Family, and Life Expectancy underscores its significant role in overall happiness. The clustering of countries based on their characteristics further revealed geographical patterns, highlighting the association between wealth, generosity, and geographic location.

As we move forward, several crucial next steps come into focus:

- **Refinement of Data Models**: apply various machine learning algorithms, and assessing predictive accuracy for understanding and predicting happiness scores.
- **Cultural Sensitivity**: incorporate cultural sensitivity. This includes considering cultural nuances, values, and societal norms that influence happiness perceptions.
- **Policy Implications**: the insights derived from our data analysis should be translated into actionable policy recommendations.
- **Data Integration**: integrate additional data sources, such as real-time data on global crises, economic trends, and social developments.
- **Stakeholder Engagement**: engaging with governments, INGOs, is pivotal to ensure that our findings and recommendations are useful to decision making.
- **Longitudinal Studies**: to track changes in happiness levels over time and assess the impact of interventions and policies on it.
- **Ethical Considerations**: As we delve deeper into data analysis, it is essential to uphold ethical standards in data collection, analysis, and reporting, ensuring privacy, fairness, and transparency in our methodologies.