

GLOBAL HAPPINESS MODELING AND LIFESTYLE ANALYSIS

1. Project Goal

The primary objective of this project is to explore the relationships between environmental, economic, and lifestyle factors and the overall happiness levels of countries. By leveraging advanced data science techniques, the analysis aims to:

- Identify the most influential predictors of happiness.
- Quantify their impact using machine learning models.
- Develop accurate predictive models to estimate a country's happiness level based on measurable indicators.

This project demonstrates end-to-end data science skills including data cleaning, exploratory data analysis, feature engineering, predictive modeling, and model evaluation.

2. Problem Statement

Despite extensive research on global happiness, policymakers often lack a quantitative understanding of how diverse factors such as environmental, economic, and lifestyle interact to influence well-being. Key challenges include:

- Missing or inconsistent data across countries and cities.
- Multiple interacting factors influencing happiness simultaneously.
- Difficulty in prioritizing factors for policy or actionable recommendations.

This project addresses these challenges through systematic data cleaning, exploratory analysis, and predictive modeling.

3. Data Overview

The dataset contains **42 countries/cities** with **12 variables** capturing environmental, economic, health, and lifestyle measures, including:

- Sunshine hours (City)
- Cost of a bottle of water (City)
- Obesity levels (Country)
- Life expectancy (years) (Country)
- Pollution (Index score) (City)
- Annual average hours worked

- Outdoor activities (City)
- Number of take-out places (City)
- Cost of a monthly gym membership (City)
- Happiness levels (Country) (*target variable*)

Data preprocessing included handling missing values, standardizing formats, and ensuring numeric columns were properly typed for modeling.

4. Exploratory Data Analysis (EDA)

4.1 Distribution of Features

Histograms were generated to visualize distributions of key variables such as sunshine hours, pollution, cost of gym membership, life expectancy, and happiness levels.

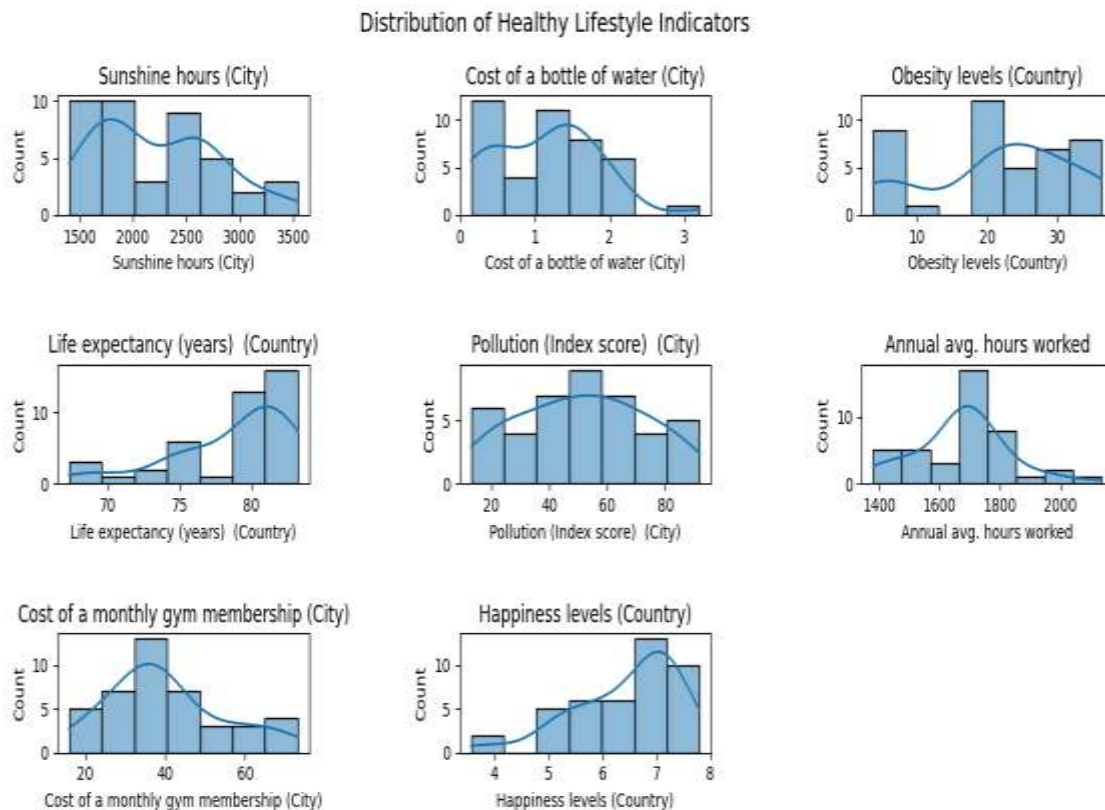


Figure 1: Distribution of Healthy Lifestyle Indicators

Observation Highlights:

- Life expectancy and happiness levels are skewed toward higher values.
- Pollution and obesity levels show greater variability across countries.
- Economic factors, including water and gym costs, vary widely, reflecting differences in affordability.

4.2 Correlation Analysis

A correlation heatmap was generated to visualize how variables interact:

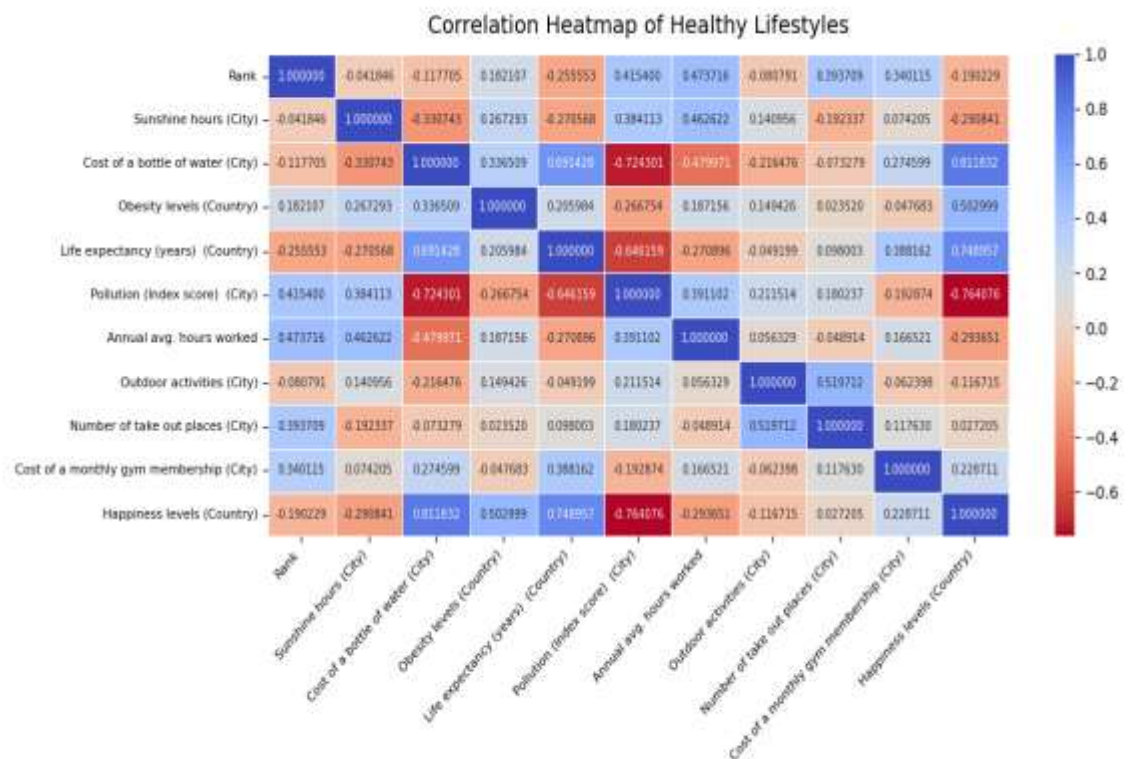


Figure 2: Correlation Heatmap of Healthy Lifestyles

Key findings:

- Happiness levels correlated **positively** with life expectancy and sunshine hours.
- Negative correlations were observed with pollution index and obesity levels.
- Economic factors (e.g., gym membership costs) showed moderate relationships, suggesting affordability impacts happiness indirectly.

4.3 Bivariate Relationships

Scatter plots revealed patterns between happiness and major predictors:

- Higher **pollution** levels were associated with *lower happiness*.
- More **sunshine hours** correlated with *increased happiness*.

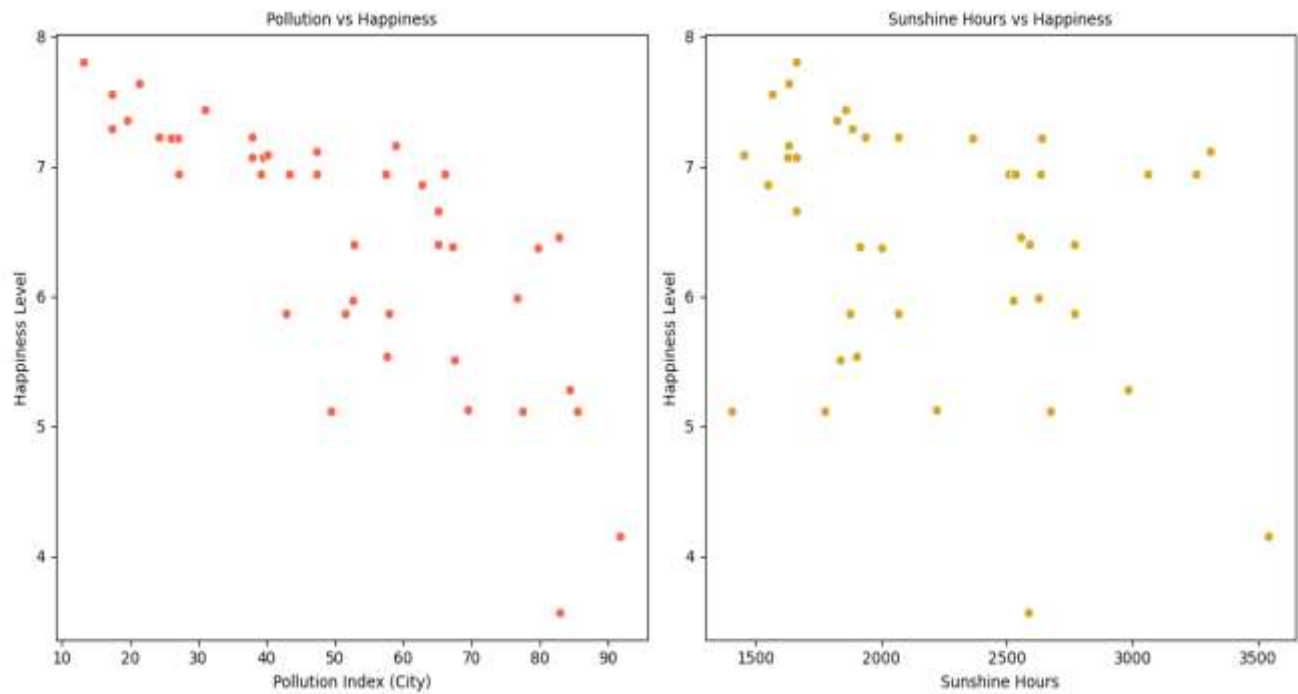


Figure 3: Relationship between Environmental and Lifestyle Factors and Happiness Levels

5. Model Development

Multiple regression and ensemble models were developed to predict country happiness levels:

- Linear Regression
- Ridge Regression
- Lasso Regression
- Random Forest Regressor
- Gradient Boosting Regressor
- Support Vector Regressor (SVR)

5.1 Model Evaluation Metrics

Models were evaluated using standard regression metrics:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- R^2 Score (coefficient of determination)

Summary of Results:

Model	MAE	MSE	R^2 Score
Linear Regression	0.162	0.051	0.918
Ridge Regression	0.151	0.04	0.935
Lasso Regression	0.218	0.057	0.908
Random Forest Regressor	0.198	0.046	0.927
Gradient Boosting Regressor	0.233	0.073	0.882
Support Vector Regressor	0.243	0.068	0.891

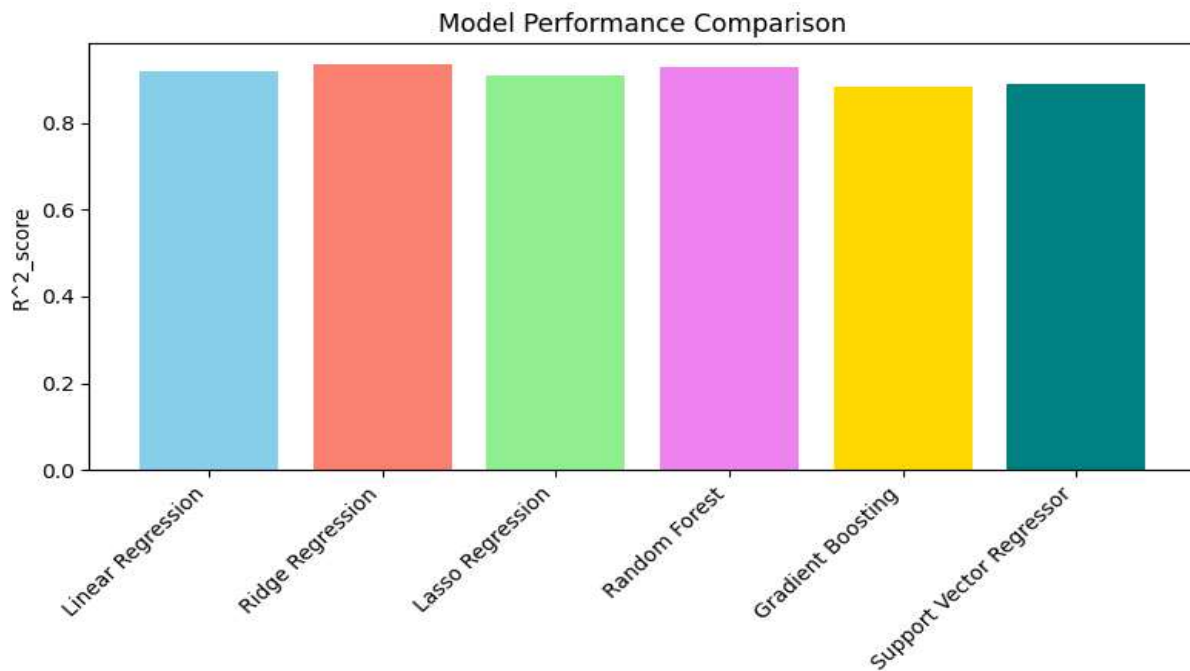


Figure 4: Model Performance Comparison (R^2 Scores Bar Chart)

Ridge Regression and Random Forest models achieved the highest predictive accuracy. Ensemble capture complex interactions, while simpler linear models performed competitively.

5.2 Feature Importance

Feature importance was assessed using the Random Forest model:

Top predictors of happiness:

1. Cost of a bottle of water
2. Life expectancy
3. Pollution index score

Other factors such as obesity levels, number of take-out places, and gym membership costs contributed moderately. Sunshine hours and annual average hours worked were minimally impactful.

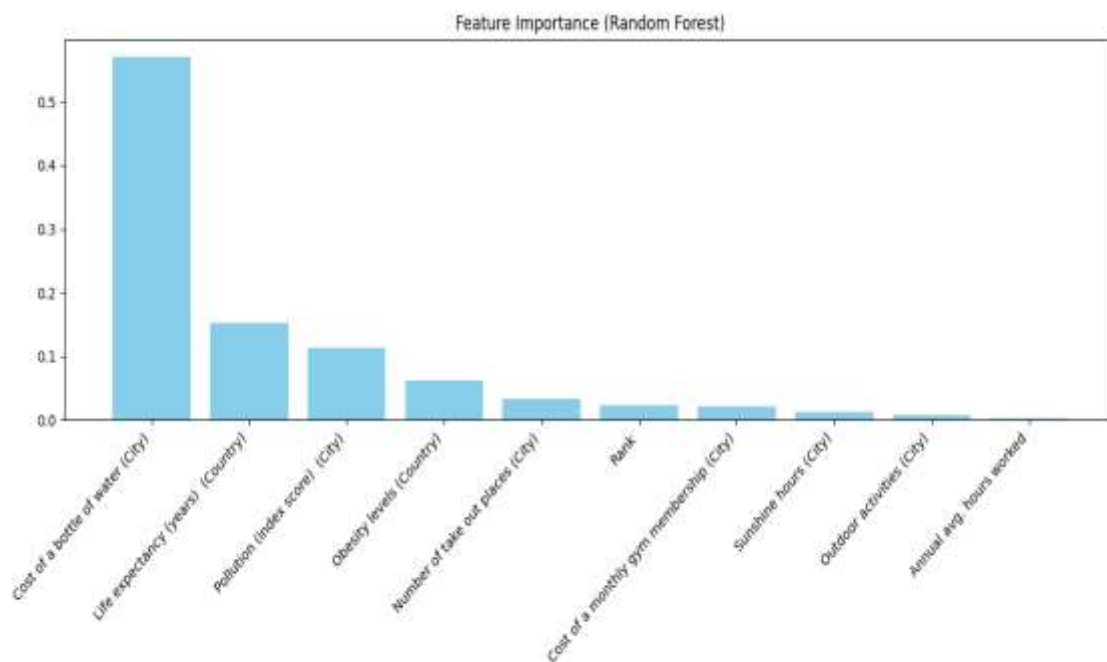


Figure 5: Feature Importance (Random Forest)

Environmental and economic indicators are the primary drivers of happiness in this dataset, providing actionable insights for policymakers.

6. Key Insights and Overall Conclusion

Key Insights:

1. **Economic and environmental factors dominate:** Cost of living (e.g., cost of water) and pollution are strong predictors of happiness.
2. **Health matters:** Life expectancy and obesity levels significantly correlate with well-being.
3. **Lifestyle variables contribute moderately:** Gym memberships, outdoor activities, and take-out availability have smaller but measurable impacts.
4. **Model performance:** Both simple linear models and ensemble methods can accurately predict happiness levels, confirming the dataset's consistency and quality.

Overall Conclusion:

Global happiness is shaped by a complex interplay of factors, yet the analysis reveals that economic affordability, environmental quality, and longevity are the most influential determinants. While lifestyle and work-related metrics play a smaller role, the data demonstrates that societies achieve higher levels of well-being when citizens have access to affordable essentials, live in cleaner environments, and enjoy longer, healthier lives. These insights provide actionable guidance for policymakers and stakeholders: prioritizing initiatives that improve public health, reduce environmental hazards, and ensure economic accessibility can produce measurable gains in societal happiness. By focusing on these critical areas, governments and organizations can implement evidence-based strategies that promote sustainable well-being on both national and local scales.