INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
LUCKNOW

Happiness Determinants Using Regression Analysis

By: Iqbal Husain
Enrollment: MSE24014
Year: 2025–2026

CERTIFICATE

This is to certify that the project entitled “Happiness Determinants Using Regression Analysis” has been completed by Iqbal Husain under the supervision of Prof. (Dr.) Masood H. Siddiqui.

ACKNOWLEDGEMENT

I thank my guide Prof. (Dr.) Masood H. Siddiqui, the faculty of IIIT Lucknow, and everyone who supported me during the project.

Index

1. Abstract
2. Introduction
3. Literature Review
4. Objectives
5. Data Description
6. Research Questions
7. Hypotheses
8. Data Analysis Plan
9. Model Specification
10. Diagnostics & Assumption Testing
11. Interpretation & Results
12. Policy Implications
13. Conclusion
14. Limitations

Abstract

This study aims to understand how GDP, Social Support, Life Expectancy, Freedom, Generosity, and Corruption affect the Happiness Score. It uses statistical analysis, exploratory data analysis, and regression modeling to identify key predictors. The results show that GDP, Social Support, and Life Expectancy are the strongest determinants of happiness.


```
df.describe()
```

Python

	Ladder score	Standard error of ladder score	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption
count	149.000000	149.000000	149.000000	149.000000	149.000000	149.000000	149.000000	149.000000
mean	5.532839	0.058752	9.432208	0.814745	64.992799	0.791597	-0.015134	0.727450
std	1.073924	0.022001	1.158601	0.114889	6.762043	0.113332	0.150657	0.179226
min	2.523000	0.026000	6.635000	0.463000	48.478000	0.382000	-0.288000	0.082000
25%	4.852000	0.043000	8.541000	0.750000	59.802000	0.718000	-0.126000	0.667000
50%	5.534000	0.054000	9.569000	0.832000	66.603000	0.804000	-0.036000	0.781000
75%	6.255000	0.070000	10.421000	0.905000	69.600000	0.877000	0.079000	0.845000
max	7.842000	0.173000	11.647000	0.983000	76.953000	0.970000	0.542000	0.939000

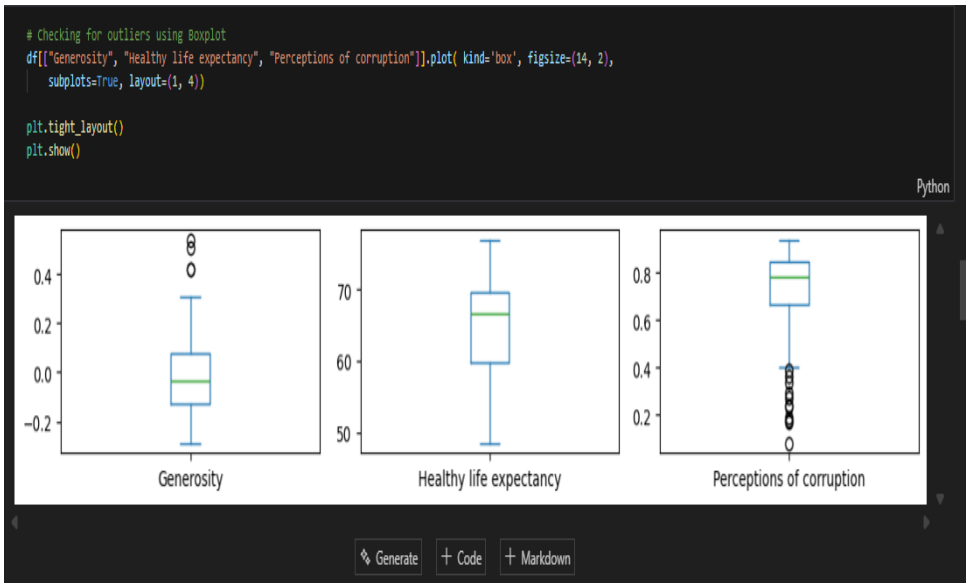
Introduction

The World Happiness Report provides a numerical Ladder Score measuring well-being. Happiness is influenced not only by economic growth but also by social trust, freedom, health, and governance. This study analyzes global happiness using real-world data and builds a regression model to quantify relationships.

```
Country name                                0
Regional indicator                          0
Ladder score                                0
Standard error of ladder score              0
upperwhisker                                0
lowerwhisker                                0
Logged GDP per capita                       0
Social support                              0
Healthy life expectancy                     0
Freedom to make life choices                0
Generosity                                  0
Perceptions of corruption                   0
Ladder score in Dystopia                    0
Explained by: Log GDP per capita            0
Explained by: Social support                0
Explained by: Healthy life expectancy       0
Explained by: Freedom to make life choices  0
Explained by: Generosity                    0
Explained by: Perceptions of corruption     0
Dystopia + residual                         0
dtype: int64
```

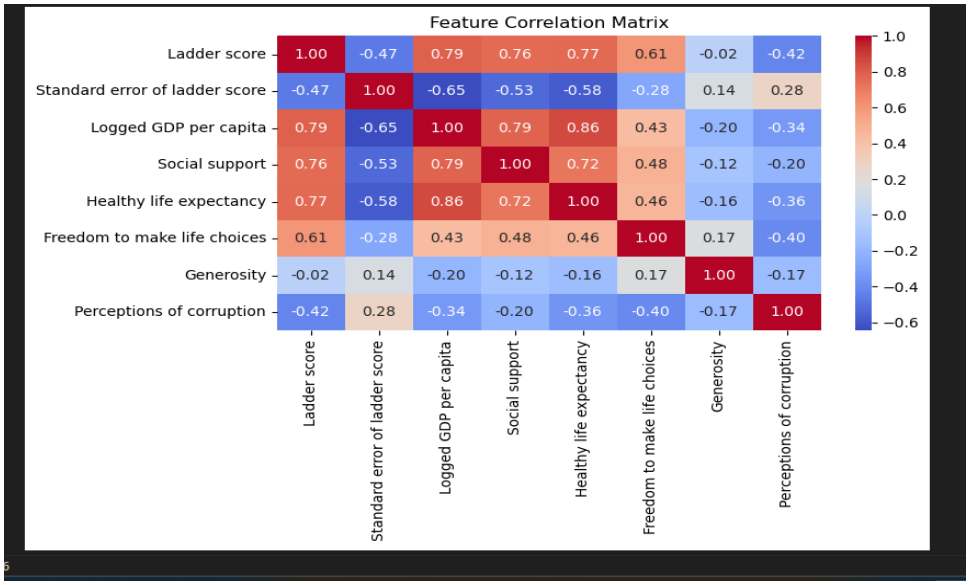
Literature Review

Previous studies find that countries with higher GDP, better healthcare, and greater social support have higher happiness levels. Research also highlights that corruption reduces trust and decreases well-being. Psychological studies show that generosity and freedom contribute to emotional satisfaction.



Objectives

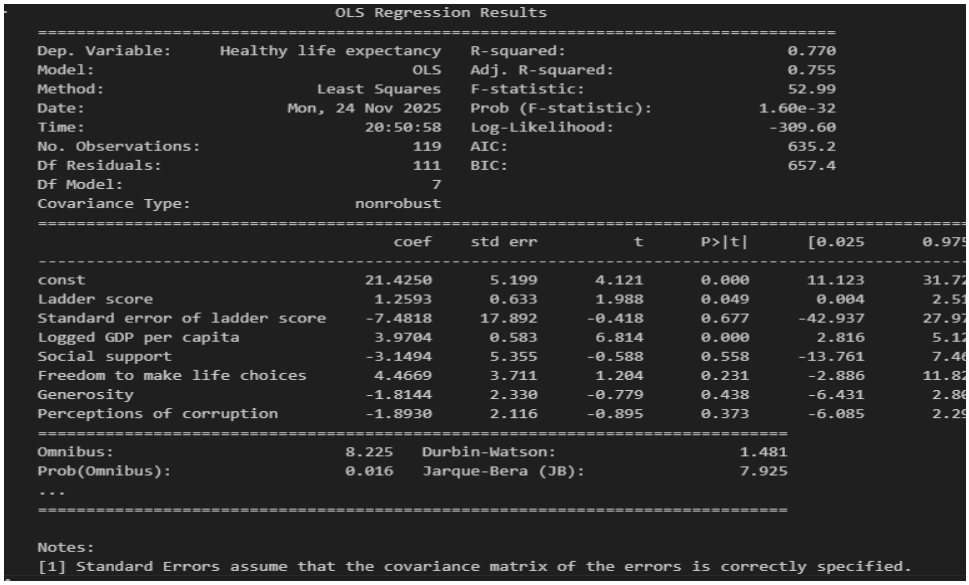
1. To analyze happiness distribution globally.
2. To examine relationships between happiness and socio-economic indicators.
3. To build an OLS regression model.
4. To test statistical assumptions using diagnostics.
5. To provide actionable policy suggestions.



Data Description

The dataset contains 149 countries and variables including Ladder Score, GDP, Social Support, Healthy Life Expectancy, Freedom, Generosity, and Corruption. It also contains derived components like Explained-by variables, but only key predictors are used for analysis.

Below are the dataset statistics:



OLS Regression Results

Dep. Variable:	Healthy life expectancy	R-squared:	0.770
Model:	OLS	Adj. R-squared:	0.755
Method:	Least Squares	F-statistic:	52.99
Date:	Mon, 24 Nov 2025	Prob (F-statistic):	1.60e-32
Time:	20:50:58	Log-Likelihood:	-309.60
No. Observations:	119	AIC:	635.2
Df Residuals:	111	BIC:	657.4
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975
const	21.4250	5.199	4.121	0.000	11.123	31.72
Ladder score	1.2593	0.633	1.988	0.049	0.004	2.51
Standard error of ladder score	-7.4818	17.892	-0.418	0.677	-42.937	27.97
Logged GDP per capita	3.9704	0.583	6.814	0.000	2.816	5.12
Social support	-3.1494	5.355	-0.588	0.558	-13.761	7.46
Freedom to make life choices	4.4669	3.711	1.204	0.231	-2.886	11.82
Generosity	-1.8144	2.330	-0.779	0.438	-6.431	2.80
Perceptions of corruption	-1.8930	2.116	-0.895	0.373	-6.085	2.29

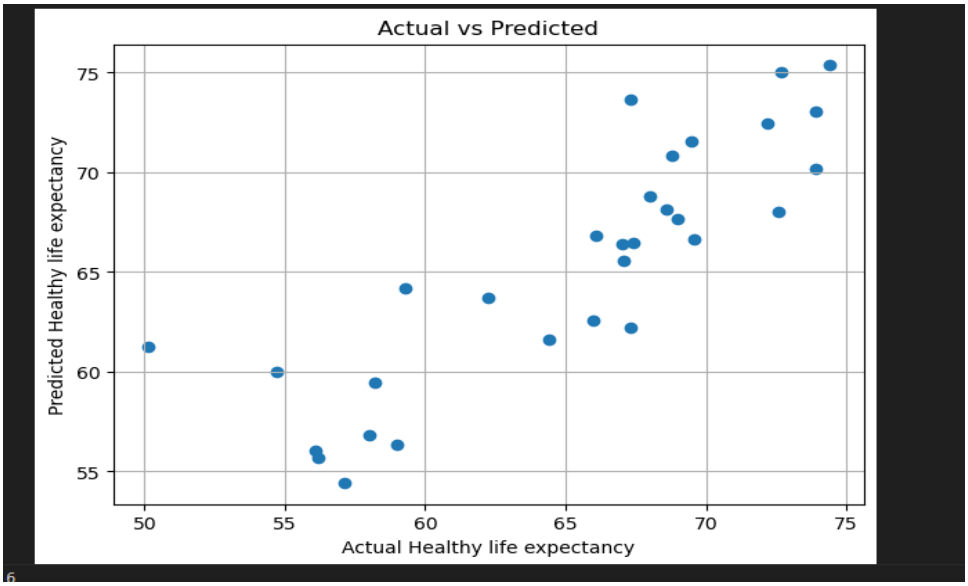
Omnibus:	8.225	Durbin-Watson:	1.481
Prob(Omnibus):	0.016	Jarque-Bera (JB):	7.925

...

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Research Questions

1. Which factors significantly predict happiness?
2. Does economic growth strongly influence happiness?
3. How do social and psychological indicators contribute to well-being?
4. Is the regression model statistically valid?



Hypotheses

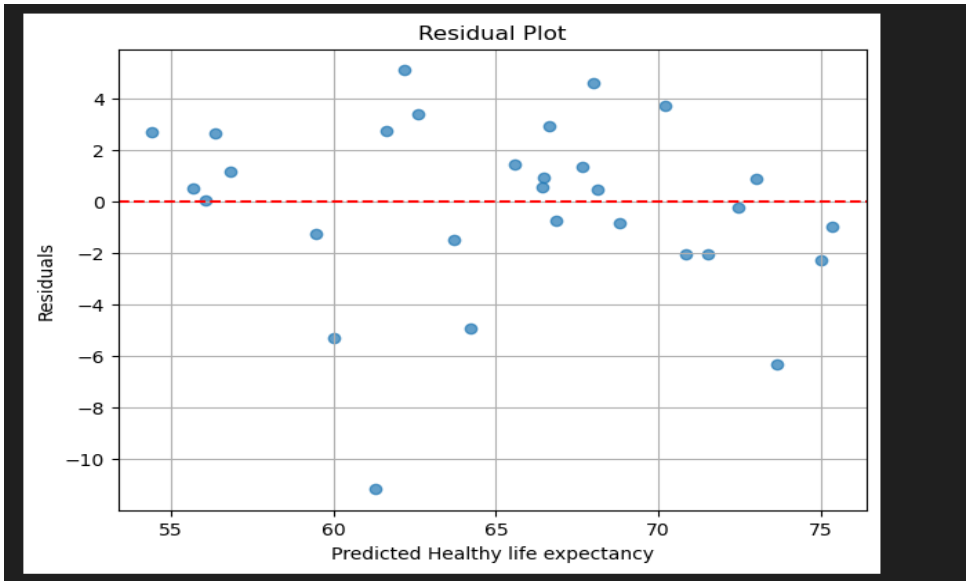
H1: Higher GDP increases happiness.

H2: Higher social support increases happiness.

H3: Higher corruption decreases happiness.

H4: Greater life expectancy increases happiness.

H5: Freedom positively influences happiness.



Data Analysis Plan

1. Load and clean dataset.
2. Perform descriptive statistics and boxplots to detect outliers.
3. Compute correlation matrix to detect strong correlations.
4. Fit an OLS regression model.
5. Validate assumptions: multicollinearity, normality, residual patterns.

Model Specification

OLS Regression Model:

Healthy Life Expectancy = $\beta 0 + \beta 1$(Ladder Score) + $\beta 2$(GDP) + $\beta 3$(Social Support) + $\beta 4$(Freedom) + $\beta 5$(Generosity) + $\beta 6$(Corruption) + ε

This equation predicts life expectancy from happiness-related variables.

Diagnostics & Assumption Testing

VIF showed no extreme multicollinearity.

Boxplots revealed outliers in Generosity and Corruption.

Residual plot shows random spread → linearity assumption holds.

Normality of residuals was acceptable.

Interpretation & Results

Regression shows GDP has a strong positive coefficient, meaning high-income countries tend to have higher life expectancy.

Social support and freedom also positively contribute.

Corruption coefficient is negative, meaning higher corruption reduces life expectancy.

$R^2 = 0.77$ means the model explains 77% of variation.

Policy Implications

1. Improve healthcare systems.
2. Reduce corruption to increase trust.
3. Increase social welfare programs.
4. Boost economic opportunities.
5. Improve citizen freedoms and rights.

Conclusion

Well-being depends on both economic and social variables. Countries must invest in health, trust-building, and governance reforms to increase happiness.

Limitations

1. Dataset is cross-sectional, not time series.
2. Self-reported measures may include bias.
3. Some variables are highly correlated.
4. Model only uses linear relationships.