

O que torna um país feliz? Uma análise de regressão dos fatores de felicidade global

Luca Ferreira Barboza
Inteligência Artificial e Machine Learning
Caio Borges e Raimundo Silva

RESUMO

A felicidade é algo crucial para o entendimento da sociedade. Este estudo tem como objetivo a identificação e a avaliação dos fatores que mais influenciam na felicidade média de um país. Utilizando a base de dados da World Happiness Report do ano de 2021, que tem dados de 149 países, foi realizada uma análise exploratória e a construção e avaliação de um modelo de regressão linear para prever a felicidade, medida pela Ladder Score. No modelo final foram escolhidas as variáveis PIB per capita, Suporte social e Expectativa de vida saudável, que demonstrou um forte poder preditivo com o R^2 Ajustado de aproximadamente 0.69, ou seja explica 69% da variação de felicidade. Adicionalmente, os testes de diagnóstico confirmaram a robustez e a adequação do modelo às premissas da regressão linear. Por fim, os resultados mostraram que o Suporte social tem um peso maior na felicidade de uma nação, levando em consideração o coeficiente dessa variável no modelo.

Palavras-chave: Regressão, Predição, Felicidade, Machine Learning, Estatística

INTRODUÇÃO

A busca por compreender a felicidade, tornou-se um campo de estudo. Nesse contexto, relatórios como o World Happiness Report são ferramentas muito úteis, oferecendo dados que permitem uma análise dos fatores que contribuem para a qualidade de vida em escala global. Este trabalho se insere nesta área, buscando desvendar as complexas relações entre os indicadores socioeconômicos e a felicidade, utilizando uma abordagem de aprendizado de máquina.

A pergunta central que norteia esta pesquisa é: quais são os fatores determinantes que mais impactam a pontuação de felicidade de um país e é possível construir um modelo preditivo confiável a partir deles? Diante disso, o objetivo geral deste estudo é desenvolver e validar um modelo de regressão linear múltipla para

identificar e quantificar os principais preditores da felicidade nacional, utilizando como base de dados o World Happiness Report de 2021. Busca-se, com isso, aplicar conhecimentos teóricos e práticos em projetos reais de Machine Learning e Inteligência Artificial.

Para atingir este propósito, foram definidos os seguintes objetivos específicos: 1) realizar uma análise exploratória para identificar as correlações entre as variáveis; 2) construir e treinar um modelo de regressão linear para prever a pontuação de felicidade; 3) avaliar a performance e a significância estatística do modelo; e 4) interpretar o impacto de cada variável preditora, como PIB per capita, suporte social e expectativa de vida saudável, no resultado final.

FUNDAMENTAÇÃO TEÓRICA

A análise do desenvolvimento das nações durante algum tempo foi baseada principalmente em termos financeiros, como o Produto Interno Bruto (PIB), por exemplo. Porém, de um tempo para cá, métricas que explicam o bem-estar coletivo, como a felicidade é um campo que vem crescendo para classificar o progresso de um país. Conforme explica Easterlin (2019), um dos pioneiros na área:

Acima de tudo, a felicidade é uma medida muito mais abrangente do que o PIB. Como medida de bem-estar, o PIB, em termos per capita, aproxima-se, na melhor das hipóteses, da mudança média nos rendimentos reais das pessoas, ou seja, na quantidade de bens e serviços produzidos e, na sua maior parte, consumidos pela sociedade. A felicidade, em contraste, registra o efeito no bem-estar não apenas da renda, mas também dos desenvolvimentos em outros aspectos da vida das pessoas, mais importante, a saúde e as circunstâncias familiares. (EASTERLIN, 2019, p. v, tradução nossa)

Essa relação, entre renda e bem estar, foi muito bem explicada pelo próprio Richard Easterlin, no chamado “Paradoxo de Easterlin”. Em sua pesquisa, ele demonstrou uma aparente contradição que desafiou a visão de que o crescimento econômico era o caminho direto para a felicidade. O próprio autor define o paradoxo da seguinte forma: “O Paradoxo afirma que, em um ponto no tempo, a felicidade varia diretamente com a renda, tanto entre nações quanto dentro delas, mas, ao longo do

tempo, a felicidade não tende a aumentar à medida que a renda continua a crescer.” (Easterlin 2016, p. 3, tradução nossa).

Após a constatação do Paradoxo de Easterlin, a literatura sobre o bem-estar voltou-se para fatores não-materiais como principais explicadores da felicidade. Dentre eles, o suporte social consistentemente se destaca como um dos pilares mais robustos do bem-estar de uma nação. O próprio World Happiness Report, em sua edição de 2025, enfatiza o papel crucial das conexões sociais e do cuidado interpessoal na promoção do bem-estar e da felicidade globalmente.

A importância do suporte social se torna ainda mais evidente quando analisada como um mecanismo de proteção contra as adversidades. A literatura demonstra que a presença de laços sociais fortes não apenas melhora a percepção de bem-estar, mas também oferece uma proteção real contra os impactos negativos do estresse na saúde mental. Conforme aponta o relatório: “Os laços sociais também protegem (ou atuam como um amortecedor para) as pessoas dos efeitos tóxicos do estresse, reduzindo o risco de que dificuldades subclínicas se transformem em transtornos de humor.” (HELLIWELL et al., 2025, p. 126, tradução nossa)

Adicionalmente aos fatores sociais, a saúde é consistentemente apontada pela literatura como outro pilar fundamental do bem-estar. Uma abrangente revisão sistemática e meta-análise confirmou existir uma relação positiva e estatisticamente significativa, ainda que de magnitude moderada, entre o estado de saúde geral dos indivíduos e seu nível de bem-estar subjetivo (NGAMABA; PANAGIOTI; ARMITAGE, 2017). Esta evidência reforça a importância da Expectativa de Vida Saudável como uma variável explicativa relevante no modelo preditivo da felicidade nacional.

Apesar da crescente aceitação do bem-estar subjetivo como um indicador de progresso, é imperativo reconhecer as críticas e limitações inerentes à sua mensuração. Autores como Yasha Mounk (2025), alertam para os perigos de uma confiança excessiva em métricas baseadas em autoavaliação e subjetivas.

Para investigar a relação entre os fatores socioeconômicos e a felicidade nacional, a regressão linear múltipla, é um modelo capaz de analisar isso. A escolha deste método é justificada por sua finalidade, “A análise de regressão é elaborada para situações em que se acredita que uma variável esteja relacionada a uma ou mais outras medidas feitas[...]” (SEARLE, 2012, p.75, tradução nossa). Dessa forma, o uso da regressão permite não apenas verificar a existência de uma relação entre a felicidade e

os indicadores socioeconômicos, mas também quantificar o impacto específico de cada um desses indicadores.

A validade das inferências de um modelo de regressão linear depende da satisfação de certas premissas estatísticas, relacionadas à análise dos resíduos, como a normalidade, a independência e a homocedasticidade dos resíduos.

O vetor de erro estimado é costumeiramente referido como o vetor de resíduos. De várias maneiras, seus elementos podem ser plotados e investigados de outras formas para verificar se sugerem que as premissas inerentes ao modelo[...]. (SEARLE, 2012, p.129, tradução nossa)

METODOLOGIA

O presente estudo caracteriza-se como uma pesquisa de natureza quantitativa, com objetivos explicativos e preditivos. Utilizando dados extraídos da base World Happiness Report de 2021, o estudo busca não apenas explicar a relação entre indicadores socioeconômicos e o bem-estar nacional, mas também desenvolver um modelo estatístico capaz de prever a pontuação de felicidade com base nesses fatores. A análise foi conduzida por meio de técnicas de modelagem estatística em ambiente computacional.

A mensuração das variáveis preditivas utilizadas neste estudo foi realizada com base nas perguntas da Pesquisa Mundial Gallup (Gallup World Poll - GWP), detalhada no apêndice do relatório (HELLIWELL et al., 2021).

A variável-alvo, Ladder score (Nota de felicidade), é obtida solicitando aos entrevistados que avaliem suas vidas em uma escala de 0 (pior vida possível) a 10 (melhor vida possível). As variáveis preditivas incluem o PIB per Capita, que mede a prosperidade econômica de cada nação; e o Suporte Social, calculado como a média nacional das respostas à pergunta sobre ter amigos ou parentes com quem contar em momentos de dificuldade. Adicionalmente, o modelo considera a Expectativa de Vida Saudável, com dados da Organização Mundial da Saúde (OMS); a Liberdade para fazer escolhas de vida, que mede a percepção de autonomia dos indivíduos; a Generosidade,

ajustada pela renda para refletir o comportamento altruísta; e a Percepção da Corrupção no governo e nas empresas.

Toda a análise de dados foi conduzida na linguagem de programação R. O processo iniciou-se com a limpeza da base, que incluiu a seleção das variáveis mais relevantes para a pesquisa e a renomeação das mesmas para facilitar o entendimento e a manipulação do código.

Após esta padronização, foi realizada uma análise descritiva das variáveis por meio de tabelas de resumo estatístico. Posteriormente, foi gerada uma matriz de correlação para visualizar a força da relação entre as variáveis, etapa que foi crucial para a seleção dos preditores do modelo final.

Com base nos achados da análise exploratória, foram construídos gráficos de dispersão para as relações de maior interesse e, em seguida, ajustou-se um modelo de regressão linear múltipla. A etapa final e crucial consistiu na validação do modelo. Foram analisadas as suposições básicas da regressão – linearidade, independência, homogeneidade (homocedasticidade) e normalidade dos resíduos – por meio de gráficos de diagnóstico específicos. Essa abordagem metodológica permitiu não apenas descrever quais variáveis influenciam na felicidade, mas também avaliar o nível de confiabilidade do modelo linear.

O modelo preditivo desenvolvido neste estudo é um modelo de regressão linear múltipla, cuja forma geral é expressa pela seguinte equação:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon$$

Onde Y representa a variável dependente (Ladder Score), X_1, X_2, \dots, X_k representam as variáveis preditoras (PIB, Suporte Social, etc.), β_0 é o intercepto, $\beta_1, \beta_2, \dots, \beta_k$ são os coeficientes de regressão que indicam o impacto de cada preditor, e ϵ representa o termo de erro do modelo.

RESULTADOS E CONSIDERAÇÕES FINAIS

Tabela 1: Estatísticas Descritivas dos Fatores de Felicidade

Variável	n	mean	sd	median	min	max
Escala de Felicidade	149.00	5.53	1.07	5.53	2.52	7.84
PIB(log)	149.00	9.43	1.16	9.57	6.63	11.65
Suporte Social	149.00	0.81	0.11	0.83	0.46	0.98
Expectativa de Vida	149.00	64.99	6.76	66.60	48.48	76.95
Liberdade	149.00	0.79	0.11	0.80	0.38	0.97
Generosidade	149.00	-0.02	0.15	-0.04	-0.29	0.54
Percepção da Corrupção	149.00	0.73	0.18	0.78	0.08	0.94

A análise quantitativa inicia-se com a exploração das estatísticas descritivas das variáveis do estudo, apresentadas na Tabela 1. Os dados referem-se a uma amostra de 149 nações. A variável-alvo, Felicidade (Ladder Score), apresenta uma pontuação média e uma mediana de aproximadamente 5,53 com um desvio padrão (DP) de aproximadamente 1,07. A proximidade entre a média e a mediana, aliada à sua posição central na escala de 0 a 10, sugere uma distribuição equilibrada e sem fortes desvios para os extremos de felicidade ou infelicidade na amostra global.

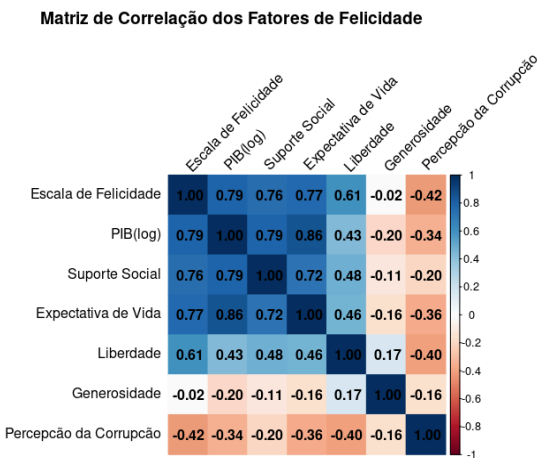


Figura 1: Matriz de correlação das variáveis.

Dando seguimento à análise descritiva, foi gerada uma matriz de correlação para quantificar a força da relação entre todas as variáveis do estudo, conforme apresentado na Figura 1. A análise da matriz revela correlações positivas e fortes entre a variável-alvo, Felicidade, e três fatores principais: PIB ($r = 0.79$), Expectativa de Vida ($r = 0.77$) e Suporte Social ($r = 0.76$). Por isso, elas foram escolhidas como preditoras no modelo de regressão linear múltipla.

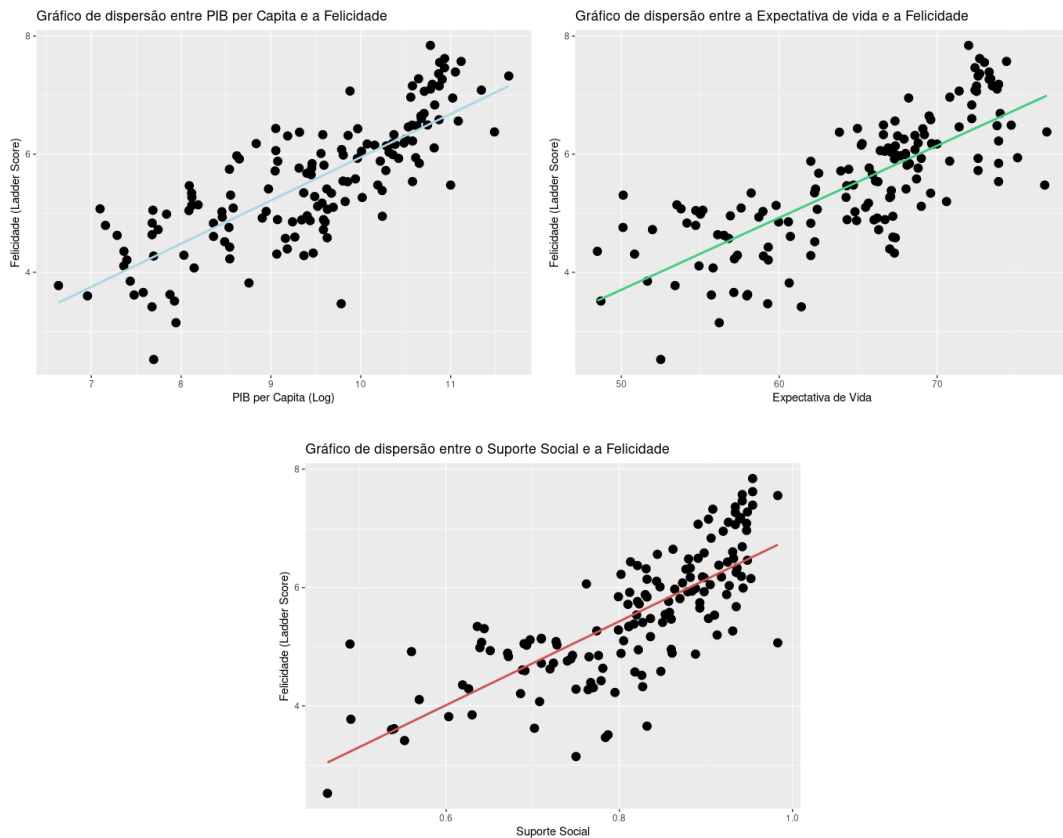


Figura 2: Gráfico de dispersão entre PIB e a Felicidade. **Figura 3:** Gráfico de dispersão entre Expectativa de vida e a Felicidade. **Figura 4:** Gráfico de dispersão entre Suporte Social e a Felicidade.

A análise dos gráficos de dispersão corroborou os resultados da matriz de correlação, evidenciando visualmente a forte relação positiva e linear entre os preditores selecionados e a pontuação de felicidade

Tabela 2: Resultados do Modelo de Regressão Linear Múltipla para a Predição da Felicidade

Variável	Coefficiente (β)	Erro Padrão	Valor-t	p-valor
(Intercepto)	-2.41	0.478	-5.03	< 0.001 ***
PIB(log)	0.273	0.094	2.90	0.004 **
Suporte Social	3.001	0.703	4.27	< 0.001 ***
Expectativa de Vida	0.045	0.014	3.10	0.002 **

R² Ajustado: 0.687

Os resultados do modelo de regressão linear múltipla, ajustado para prever a pontuação de Felicidade, são apresentados na Tabela 2. O valor do R² Ajustado indica que o modelo consegue explicar aproximadamente 68.7% da variação na pontuação de felicidade entre os países.

Analisando os coeficientes individuais, todas as três variáveis preditoras se mostraram estatisticamente significantes. O Suporte Social destacou-se como o preditor de maior impacto positivo ($\beta = 3.00$), sugerindo que um aumento de um ponto na percepção de suporte social está associado a um aumento de três pontos na escala de felicidade. O PIB (logarítmico) ($\beta = 0.27$) e a Expectativa de Vida Saudável ($\beta = 0.045$) também demonstraram uma influência positiva e significativa na determinação do bem-estar nacional.

Com base nos coeficientes, a equação final do modelo preditivo pode ser expressa como:

$$\text{Felicidade} = -2.41 + (0.27 * \text{PIB_log}) + (3.00 * \text{Suporte_Social}) + (0.045 * \text{Expectativa_Vida})$$

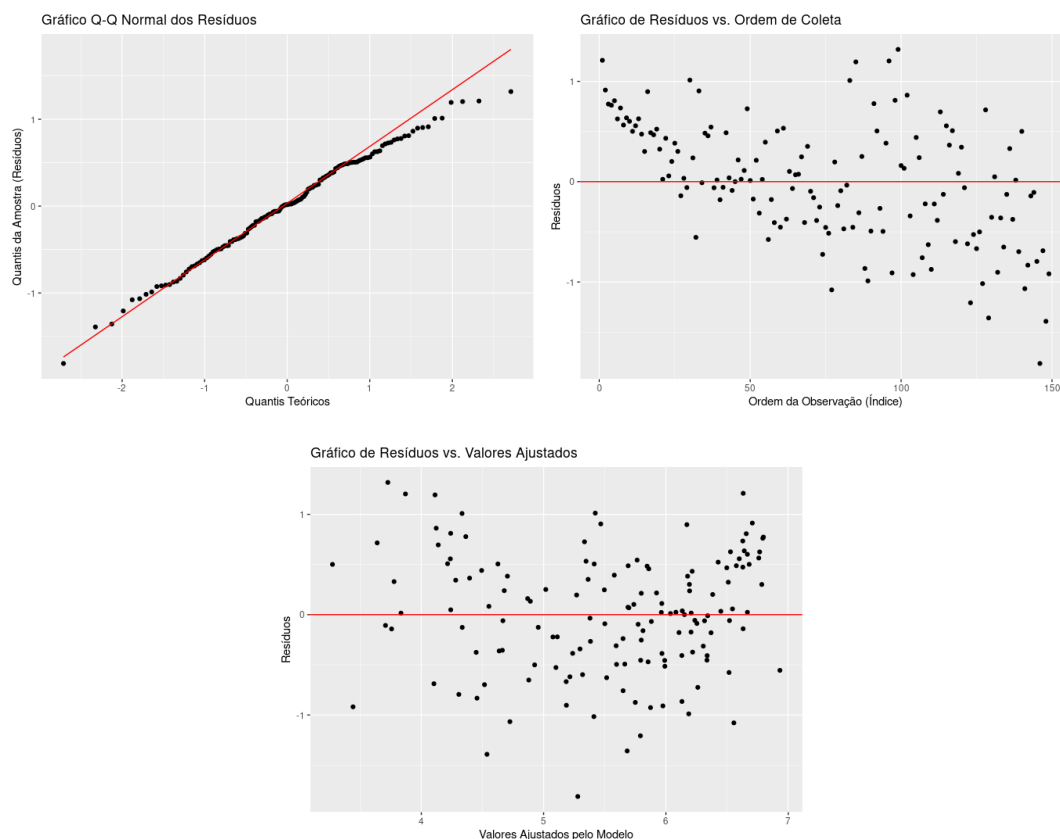


Figura 5: QQ Plot dos Resíduos. **Figura 6:** Gráfico de Resíduos x Ordem. **Figura 7:** Gráfico de Resíduos x valores ajustados.

Por fim, para assegurar a validade estatística do modelo, foi realizada uma análise de diagnóstico dos resíduos. O gráfico de Resíduos vs. Valores Ajustados (Figura 7) demonstrou que os erros estão distribuídos aleatoriamente e sem padrões, atendendo às premissas de linearidade e homocedasticidade. Adicionalmente, o gráfico Q-Q Normal (Figura 5) confirmou que os resíduos seguem uma distribuição normal, com os pontos alinhados à linha de referência teórica. A independência dos erros também foi verificada no gráfico de Resíduos vs. Ordem (Figura 6). Em conjunto, a análise diagnóstica valida o modelo de regressão linear como uma ferramenta apropriada e confiável para a análise dos dados em questão.

Este estudo se propôs a identificar os determinantes da felicidade nacional por meio da construção de um modelo de regressão linear múltipla, utilizando dados do World Happiness Report 2021. A análise resultou em um modelo estatisticamente robusto, validado por testes de diagnóstico, e capaz de explicar aproximadamente 69% da variação no bem-estar das nações. Foi mostrado que o PIB per capita, a Expectativa de Vida Saudável e, principalmente, o Suporte Social são preditores estatisticamente

significativos da felicidade. A principal conclusão extraída é a proeminência do Suporte Social como o fator de maior impacto, sugerindo que a qualidade do tecido social e a força dos laços comunitários são pilares mais críticos para o bem-estar de uma nação do que a prosperidade material de forma isolada.

É fundamental, contudo, reconhecer as limitações inerentes a este estudo. A própria mensuração da felicidade é baseada em auto avaliações subjetivas, uma abordagem que atrai críticas na literatura. Adicionalmente, deve-se ressaltar que os resultados apresentados indicam correlações fortes, mas não estabelecem, por si só, uma relação de causalidade direta entre as variáveis. Apesar dessas ressalvas, os achados reforçam a importância de políticas públicas que, para haja a promoção de sociedades mais felizes e resilientes.

REFERÊNCIAS BIBLIOGRÁFICAS

HELLIWELL, John F; et.al. **World Happiness Report 2021**. New York: Sustainable Development Solutions Network, 2021. Disponível em: <http://worldhappiness.report/>. Acesso em: 23 jun. 2025

EASTERLIN, Richard A. Happiness or GDP? *In*: ROJAS, Mariano (Ed.). **The Economics of Happiness: How the Easterlin Paradox Transformed Our Understanding of Well-Being and Progress**. Cham: Springer International Publishing, 2019.

EASTERLIN, Richard A. **Paradox Lost?**. Bonn: IZA Institute for the Study of Labor, 2016. (IZA Discussion Paper, n. 9676). Disponível em: <https://www.iza.org/publications/dp/9676/paradox-lost>. Acesso em: 23 jun. 2025.

HELLIWELL, John F; et.al. **World Happiness Report 2025**. New York: Sustainable Development Solutions Network, 2025. Disponível em: <http://worldhappiness.report/>. Acesso em: 25 jun. 2025

NGAMABA, Kayonda Hubert; PANAGIOTI, Maria; ARMITAGE, Christopher J. **How strongly related are health status and subjective well-being? Systematic review and meta-analysis**. The European Journal of Public Health, v. 27, n. 5, p. 879–885, 2017. DOI: 10.1093/eurpub/ckx081.

MOUNK, Yascha. The World Happiness Report is a sham. **The Spectator**, 2025. Disponível em: <https://www.spectator.co.uk/article/the-world-happiness-report-is-a-sham/>. Acesso em: 25 jun. 2025.

SEARLE, S. R. **Linear Models**, 1997