Introduction Methodology Results

Sapienza University of Rome

Happiness paradox

When well-being doesn’t mean mental health?

When smiling countries hide suffering.

Mariia Kalianova

Behind the numbers

We tested what the world tells us — and what the data screams beneath it.

What are we really measuring?

Do richer, safer, happier countries actually have fewer suicides?

Or are they just better at pretending?

Is economic prosperity truly protective against mental health collapse?

Or does wealth just make it easier to hide despair?

Are the world's most 'satisfied' populations silently struggling?

Can we trust the smile in the statistics?

What hidden patterns emerge when we stop trusting reported well-being?

Can clustering reveal the countries that don't fit the narrative?

Datasets

2 datasets, which include:

1. Suicide Rates in different countries across the years.
2. World Happiness Report.

source: www.kaggle.com

Building dataset that tells the truth

Final dataset:

- 77 countries
- 6 final variables
- fully aligned, no missing values
- ready for analysis

who_suicide_statistics

country	year	sex	age	suicides_no	population
Albania	1985	female	15-24 years		277900
Albania	1985	female	25-34 years		246800
Albania	1985	female	35-54 years		267500
Albania	1985	female	5-14 years		298300
Albania	1985	female	55-74 years		138700
Albania	1985	female	75+ years		34200
Albania	1985	male	15-24 years		301400
Albania	1985	male	25-34 years		264200
Albania	1985	male	35-54 years		296700
Albania	1985	male	5-14 years		325800
Albania	1985	male	55-74 years		132500
Albania	1985	male	75+ years		21100
Albania	1986	female	15-24 years		283900
Albania	1986	female	25-34 years		252100
Albania	1986	female	35-54 years		273200
Albania	1986	female	5-14 years		304700
Albania	1986	female	55-74 years		141700
Albania	1986	female	75+ years		34900
Albania	1986	male	15-24 years		306700
Albania	1986	male	25-34 years		269000
Albania	1986	male	35-54 years		302000
Albania	1986	male	5-14 years		331600
Albania	1986	male	55-74 years		134800
Albania	1986	male	75+ years		21400

+

Country name	Regional indicator	Ladder score	Standard error of ladder score
Finland	Western Europe	7.842	0.032
Denmark	Western Europe	7.620	0.035
Switzerland	Western Europe	7.571	0.036
Iceland	Western Europe	7.554	0.059
Netherlands	Western Europe	7.464	0.027
Norway	Western Europe	7.392	0.035
Sweden	Western Europe	7.363	0.036
Luxembourg	Western Europe	7.324	0.037
New Zealand	North America and ANZ	7.277	0.040
Austria	Western Europe	7.268	0.036
Australia	North America and ANZ	7.183	0.041
Israel	Middle East and North Africa	7.157	0.034
Germany	Western Europe	7.155	0.040
Canada	North America and ANZ	7.103	0.042
Ireland	Western Europe	7.085	0.040
Costa Rica	Latin America and Caribbean	7.069	0.056
United Kingdom	Western Europe	7.064	0.038
Czech Republic	Central and Eastern Europe	6.965	0.049
United States	North America and ANZ	6.951	0.049
Belgium	Western Europe	6.834	0.034
France	Western Europe	6.690	0.037
Bahrain	Middle East and North Africa	6.647	0.068
Malta	Western Europe	6.602	0.044

=

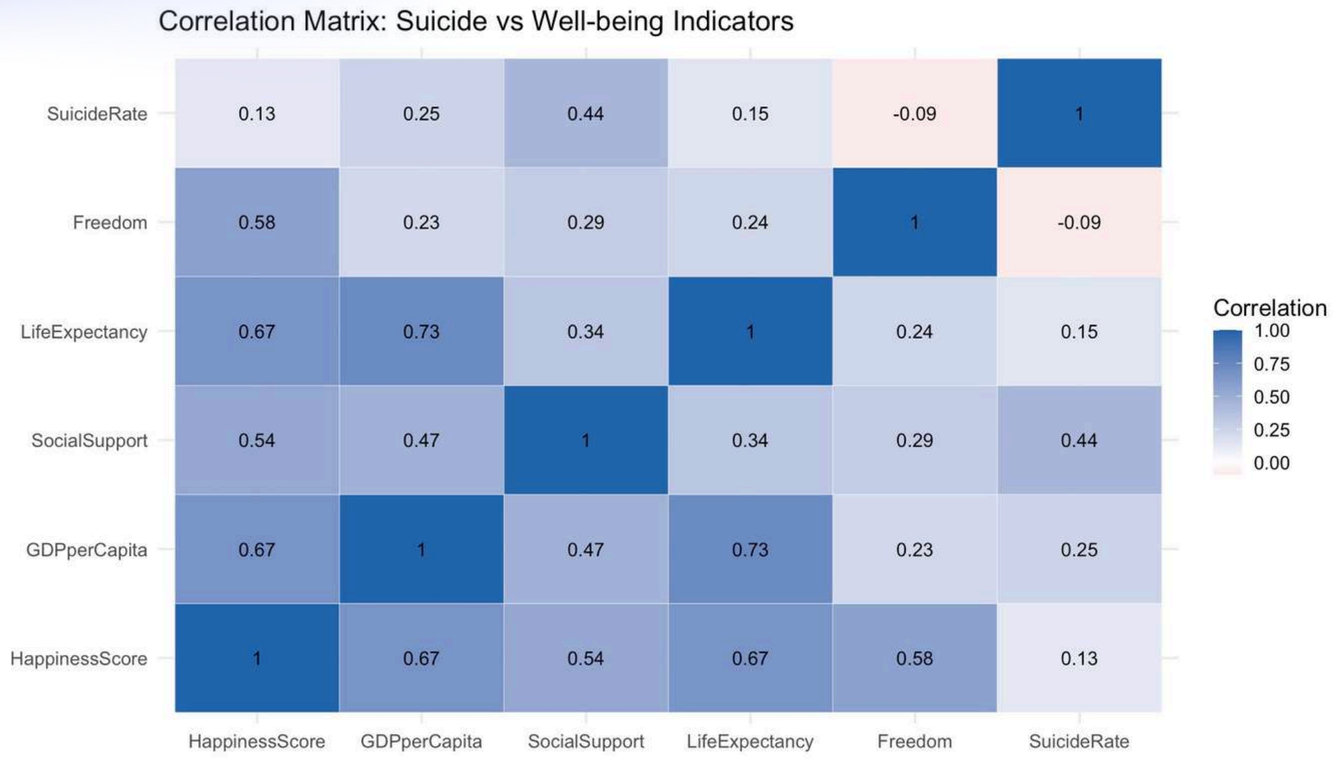
table_works

HappinessScore	GDPperCapita	SocialSupport	LifeExpectancy	Freedom	Country	SuicideRate
7.842	10.775	0.954	72	0.949	Finland	16.3368737997341
7.62	10.933	0.954	72.7	0.946	Denmark	11.5082485437804
7.571	11.117	0.942	74.4	0.919	Switzerland	13.6343957325547
7.554	10.878	0.983	73	0.955	Iceland	13.7838417363103
7.464	10.932	0.942	72.4	0.913	Netherlands	11.5439866256936
7.392	11.053	0.954	73.3	0.96	Norway	11.7752655277297
7.363	10.867	0.934	72.7	0.945	Sweden	12.8708786135854
7.324	11.647	0.908	72.6	0.907	Luxembourg	10.7789853279057
7.277	10.643	0.948	73.4	0.929	New Zealand	12.7563447264443
7.268	10.906	0.934	73.3	0.908	Austria	15.5398415731553
7.183	10.796	0.94	73.9	0.914	Australia	12.4941634566603
7.157	10.575	0.939	73.503	0.8	Israel	5.36222252702211
7.155	10.873	0.903	72.5	0.875	Germany	12.9408716649805
7.103	10.776	0.926	73.8	0.915	Canada	11.9854896315736
7.085	11.342	0.947	72.4	0.879	Ireland	12.0234776204192
7.069	9.88	0.891	71.4	0.934	Costa Rica	7.03434336898244
7.064	10.707	0.934	72.5	0.859	United Kingdom	7.77486454899293
6.951	11.023	0.92	68.2	0.837	United States of America	14.0987225941763
6.834	10.823	0.906	72.199	0.783	Belgium	18.61833725798
6.69	10.704	0.942	74	0.822	France	16.4415367488275
6.647	10.669	0.862	69.495	0.925	Bahrain	1.20663652429911
6.602	10.674	0.931	72.2	0.927	Malta	6.58455428018462
6.561	11.085	0.844	67.333	0.932	United Arab Emirates	1.65256588132993

Exploratory Analysis Techniques

- Correlation Matrix - to identify linear relationship between variables.
- Scatter plots with Regression Lines - to visualize relationships.
- Standardization - to prepare clean input for clustering.
- K-Means - to identify hidden patterns.

Shouldn't happiness protect us?

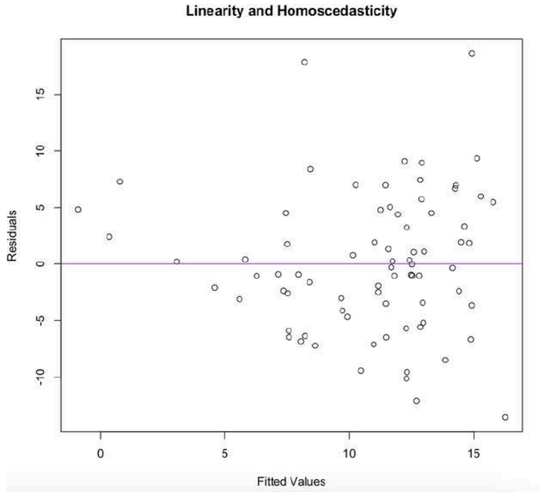


Suicide rate shows weak correlations with all well-being indicators (r = 0.13 with happiness, r = 0.15 with Life expectancy).

The only moderate link is with social support (r = 0.44) — unexpectedly positive.

Overall, traditional well-being metrics do not predict suicide risk.

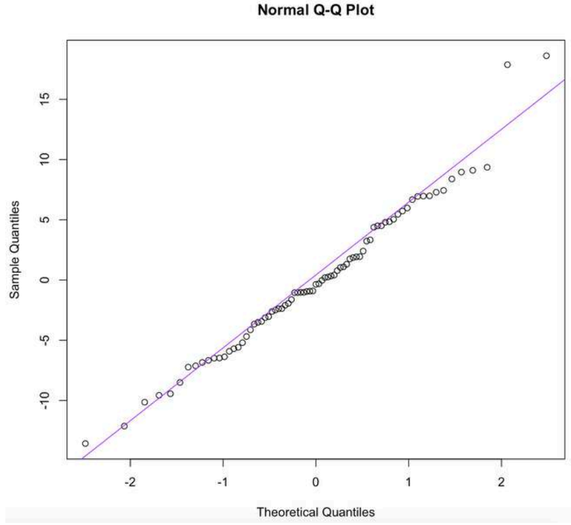
Libraries used: dplyr, ggplot2, reshape2




```
> vif(model)
HappinessScore   GDPperCapita
      3.495202       2.657582
 SocialSupport LifeExpectancy
      1.474254       2.496997
       Freedom
      1.638546
```

```
        Shapiro-Wilk normality test

data:  resid(model)
W = 0.97875, p-value = 0.2246
```




```
t test of coefficients:

                 Estimate Std. Error
(Intercept)    -39.436615  21.151670
HappinessScore  -1.237076   2.733197
GDPperCapita     1.541514   1.522815
SocialSupport   60.112009  16.764925
LifeExpectancy   0.019882   0.389005
Freedom        -15.063783  13.981037
                t value  Pr(>|t|)
(Intercept)    -1.8645 0.0663897 .
HappinessScore -0.4526 0.6522085
GDPperCapita    1.0123 0.3148421
SocialSupport   3.5856 0.0006135 ***
LifeExpectancy  0.0511 0.9593820
Freedom        -1.0774 0.2849292
---
Signif. codes:
  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05
  ‘.’ 0.1 ‘ ’ 1
```

Assumption check

We checked key assumptions before multiple linear regression. The residual plot shows approximate linearity, but there are some problems with homoscedasticity. The QQ plot and Shapiro-Wilk test (p = 0.22) confirm normality of residuals. All VIF values were below 5, indicating no multicollinearity.

Empirical Expectations vs. Statistical Reality

```
Call:
lm(formula = SuicideRate ~ HappinessScore + GDPperCapita + SocialSupport +
    LifeExpectancy + Freedom, data = merged_data)

Residuals:
     Min       1Q   Median       3Q      Max
-13.5739  -3.6608  -0.3588   4.4978  18.6185

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -39.43661   19.07525  -2.067 0.042341 *
HappinessScore  -1.23708    1.74112  -0.711 0.479718
GDPperCapita     1.54151    1.69388   0.910 0.365875
SocialSupport   60.11201   15.21693   3.950 0.000182 ***
LifeExpectancy   0.01988    0.30948   0.064 0.948958
Freedom        -15.06378   10.20320  -1.476 0.144264
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.315 on 71 degrees of freedom
Multiple R-squared:  0.2551,    Adjusted R-squared:  0.2026
F-statistic: 4.862 on 5 and 71 DF,  p-value: 0.0007003
```

A multiple regression model explains 20% of the variance in suicide rate.

Only Social Support emerges as a statistically significant predictor ($p < 0.001$) — and its effect is positive.

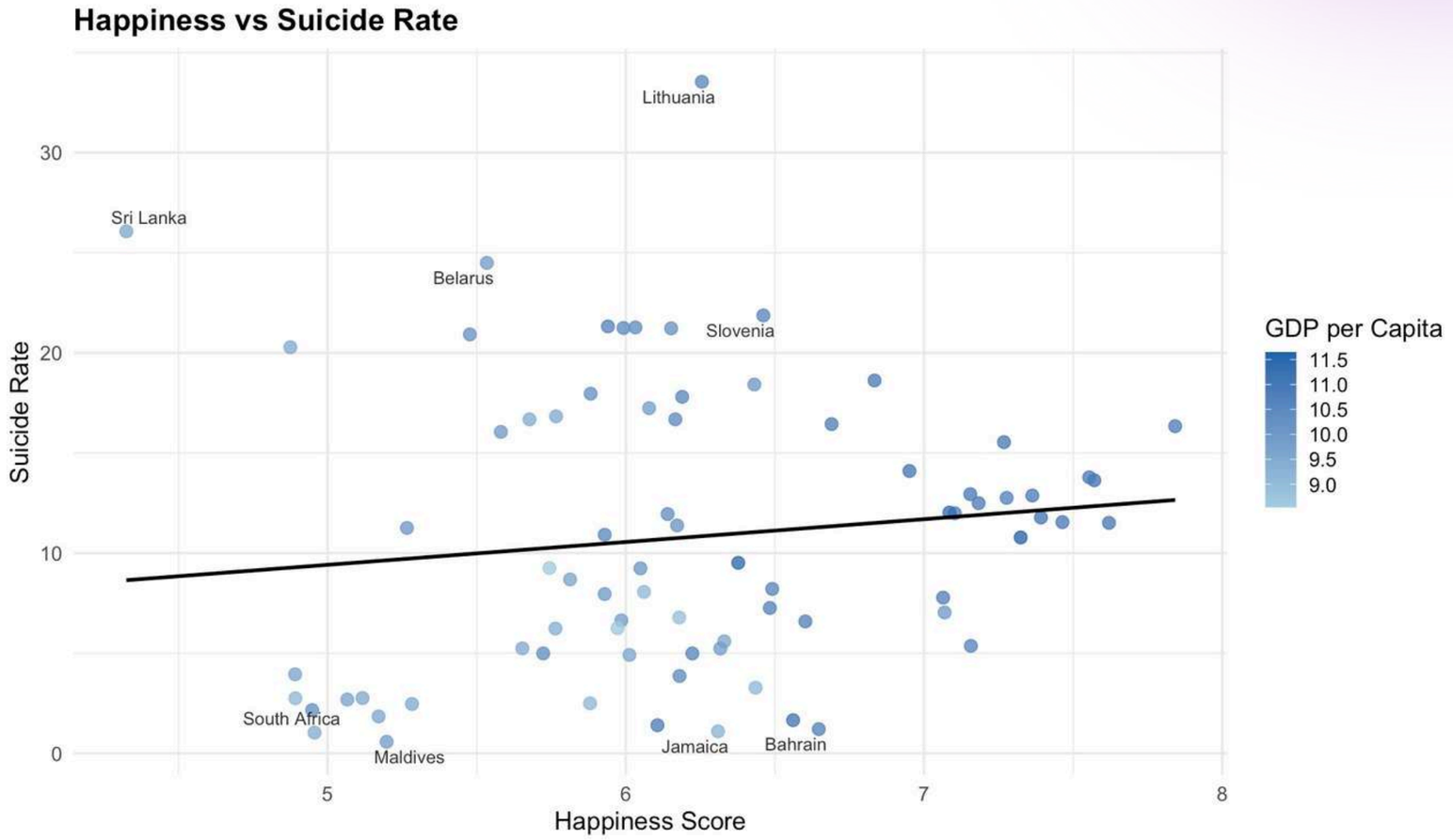
All other indicators — including Happiness, GDP, Freedom, and Life Expectancy — show no significant association.

Surprisingly, the one variable we assume protects mental health is the one most strongly linked to higher suicide.

A Counterintuitive Trend

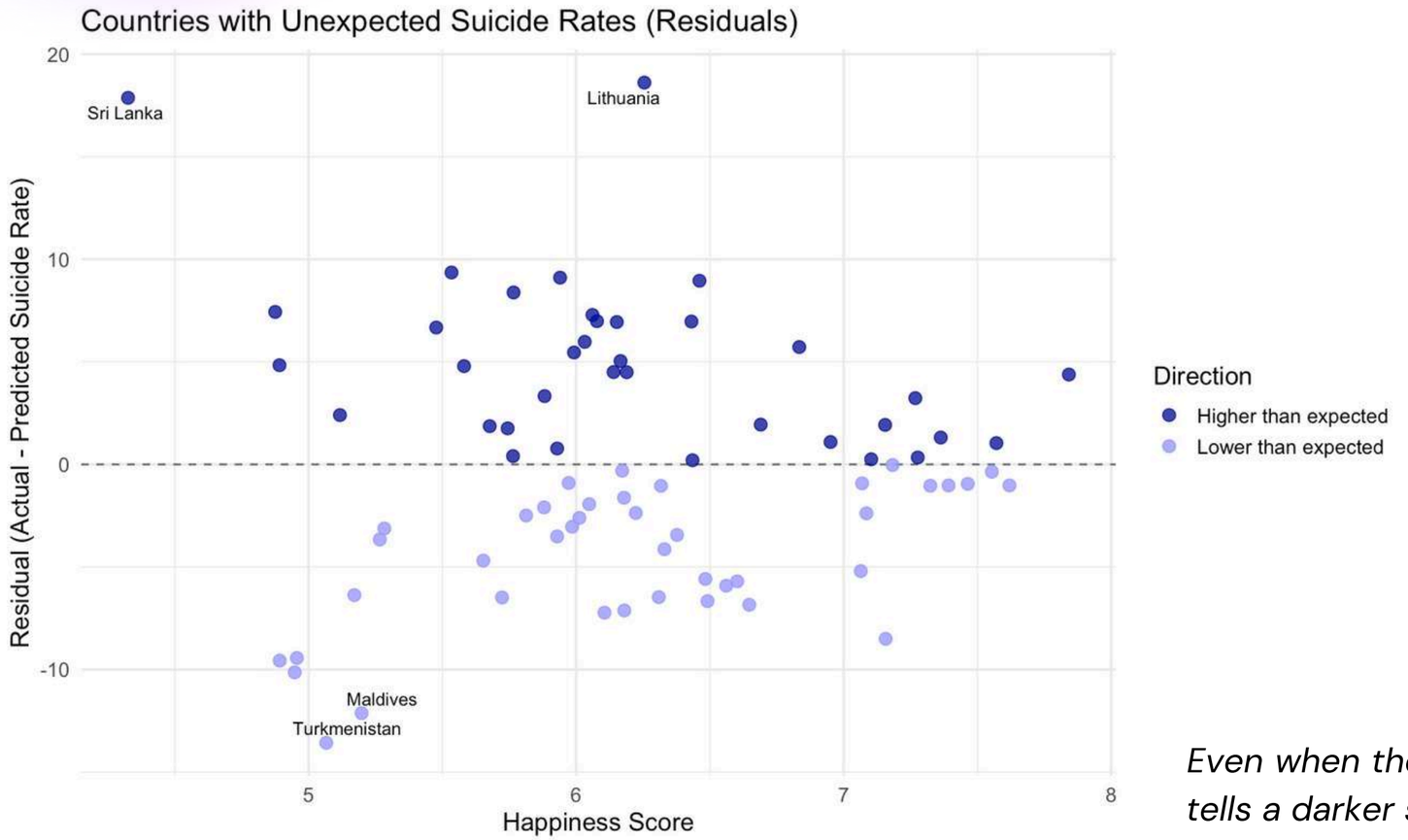
The scatterplot reveals a weak positive relationship between happiness and suicide rate.
While we expect happiness to reduce suicidality, some of the highest suicide rates occur in countries with moderate to high happiness scores (e.g. Lithuania, Slovenia, Belarus).
Countries with lower suicide rates (e.g. Maldives, Jamaica, Bahrain) are not necessarily the happiest.

Happiness does not protect against suicide — and may coexist with hidden psychological stress.



Libraries used: dplyr, ggplot2, ggrepel

The Countries That Break the Model



This plot visualizes residuals from the regression model, showing the gap between actual suicide rate and the rate predicted by well-being indicators.

Countries above the zero line have much higher suicide rates than expected, while those below have lower-than-predicted rates.

These residuals suggest that some countries are statistical outliers — where well-being indicators fail to explain mental health outcomes.

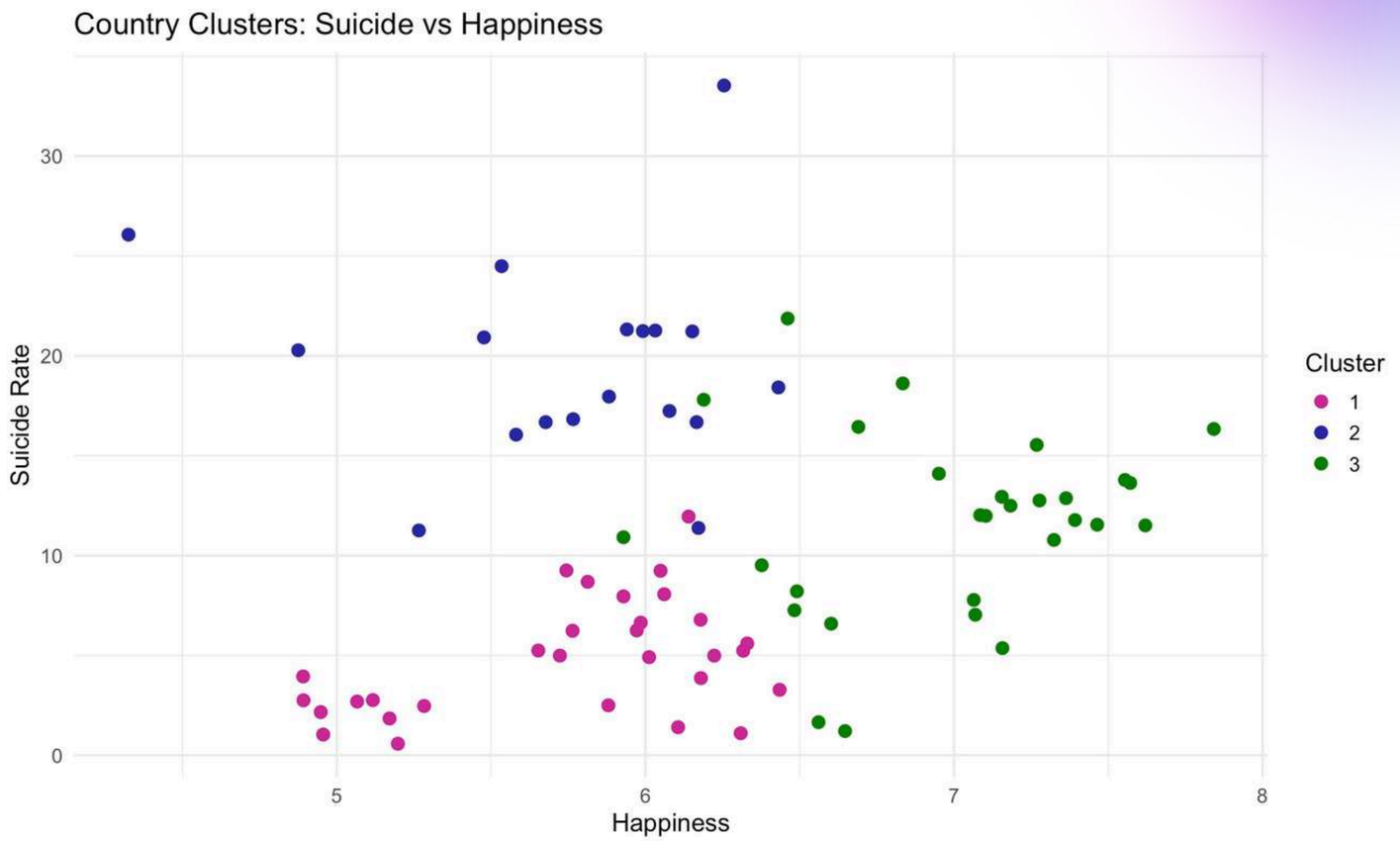
Even when the model says they should be fine — the reality tells a darker story.

Cluster analysis

K-means clustering (K = 3) on suicide rate and happiness score reveals three statistically distinct country groups:

- Cluster 1: Lowest suicide rates, clustered tightly at moderate happiness. These countries show suppressed variation and may reflect underreporting or systemic data suppression.
- Cluster 2: Highest suicide rates despite mid-to-high happiness scores. These are statistical outliers — structurally "well" countries with disproportionate mental health burdens.
- Cluster 3: Distributed around moderate suicide rates and higher happiness. Represents more stable or transparent systems with average reporting.

Clustering uncovers underlying heterogeneity not captured by linear models. The same happiness score leads to radically different suicide outcomes across clusters.



Answers to our questions

Do richer, safer, happier countries actually have fewer suicides?

1

No. Some of the happiest and wealthiest countries show higher-than-expected suicide rates.

Is economic prosperity truly protective against mental health collapse?

2

No. GDP per capita shows no significant predictive power.

Is happiness a good proxy for mental health?

3

No. Happiness scores do not predict actual suicide rates.

Can we trust global well-being rankings to reflect mental health outcomes?

4

Not reliably. Transparency, stigma, reporting accuracy, and cultural context strongly affect suicide data.

Conclusion

In this project, we applied a complete data science pipeline:

- Data cleaning and feature selection from global datasets
- Exploratory analysis through correlation matrices and visual inspection
- Multiple linear regression to test predictive power
- Clustering (K-means) to reveal hidden behavioral patterns
- And residual analysis to identify countries where the model fails

Through this, we showed that statistical well-being indicators like GDP, happiness, and life expectancy do not reliably predict suicide rates — and sometimes contradict our expectations.

This is the power of data analysis: It questions assumptions, quantifies uncertainty, and helps uncover what doesn't fit the dominant narrative.

Thank you for attention!

I hope that you were interested in the project question.

I am now ready to answer any questions you may have.

Mariia Kalianova