UC San Diego

Causal Effect of Socio-Economic Factors on Happiness

Presenters:

Abhisheik Jadhav (A69027702)
Anuj Goenka (A69027680)
Reuben Chatterjee (A59026323)

Introduction

Introduction

Objective -

The objective of our study is to identify the key socioeconomic factors influencing happiness and to quantify their causal effects using causal inference techniques.

We use two sophisticated causal inference methods — LiNGAM and DoWhy to achieve the same.

What is Happiness index?

The Happiness Index is a survey that measures a person's happiness in different areas of life, such as psychological well-being, health, time balance, community, social support, education, and arts.

We explore the causal effect of socioeconomic factors, specifically focusing on GDP, Social Support and Health, on happiness.

Dataset

The Economics of Happiness, 2015-2019.

This dataset is the 'World Happiness Report' data from 2015 to 2019. It includes variables such as GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption."

Source: Kaggle datasets.

The Economics of Happiness, 2015-2019.

Overview -

	Country	Happiness rank	Happiness Score	GDP per capita	Social support	Healthy life	Freedom	Generosity	Corruption	Year
0	Switzerland	1	7.587	1.39651	1.34951	0.94143	0.66557	0.29678	0.41978	2015
1	Iceland	2	7.561	1.30232	1.40223	0.94784	0.62877	0.43630	0.14145	2015
2	Denmark	3	7.527	1.32548	1.36058	0.87464	0.64938	0.34139	0.48357	2015
3	Norway	4	7.522	1.45900	1.33095	0.88521	0.66973	0.34699	0.36503	2015
4	Canada	5	7.427	1.32629	1.32261	0.90563	0.63297	0.45811	0.32957	2015

The Economics of Happiness, 2015-2019.

Key Variables:

- Happiness Score: The dependent variable representing the overall happiness of a country.
- GDP per capita: An indicator of economic performance.
- Social support: The perceived social support available to individuals.
- Healthy life expectancy: The expected length of a healthy life.
- Freedom to make life choices: The perceived freedom to make personal decisions."

Data Cleaning and Preparation

Data Preprocessing

- Rows with missing values were dropped.
- Unnecessary columns like country and Happiness rank were dropped as they were not essential to the analysis.
- After cleaning, the dataset includes X rows and Y columns, with no missing values.

EDA

Correlation matrix:

	Happiness Score	GDP per capita	Social support	\
Happiness Score	1.000000	0.794563	0.644818	
GDP per capita	0.794563	1.000000	0.576752	
Social support	0.644818	0.576752	1.000000	
Healthy life	0.745391	0.781259	0.564788	
Freedom	0.558843	0.374387	0.442243	
Generosity	0.135408	-0.003958	-0.036177	
Corruption	0.399636	0.323426	0.127070	
Year	0.009634	0.021506	0.389182	

	Healthy life	Freedom	Generosity	Corruption	Year
Happiness Score	0.745391	0.558843	0.135408	0.399636	0.009634
GDP per capita	0.781259	0.374387	-0.003958	0.323426	0.021506
Social support	0.564788	0.442243	-0.036177	0.127070	0.389182
Healthy life	1.000000	0.367594	0.016499	0.270518	0.139949
Freedom	0.367594	1.000000	0.273832	0.449428	0.014332
Generosity	0.016499	0.273832	1.000000	0.322026	-0.190412
Corruption	0.270518	0.449428	0.322026	1.000000	-0.125700
Year	0.139949	0.014332	-0.190412	-0.125700	1.000000

EDA

Heatmap of pearson coefficients:

UC San Diego

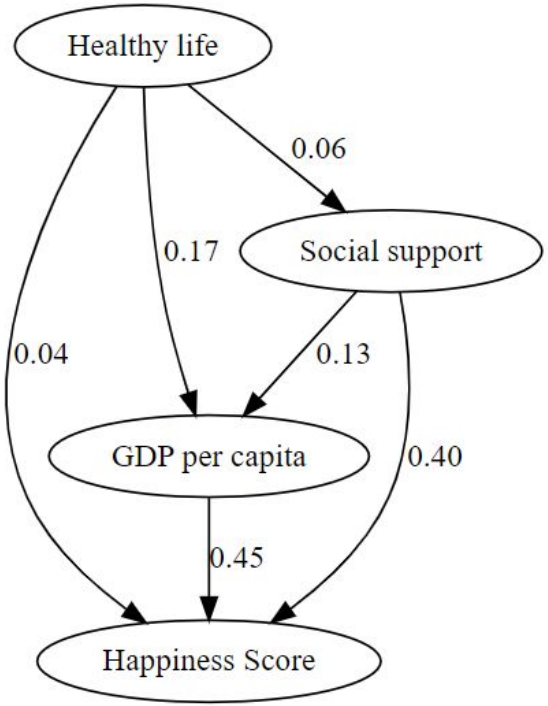
Methodology

Workflow

We used the LiNGAM (Linear Non-Gaussian Acyclic Model) method to estimate causal relationships between variables. This was followed by a detailed causal analysis using the DoWhy framework

- Data was split into training and testing sets.
- LiNGAM was used to estimate the causal structure.
- The causal graph was visualized.
- DoWhy was used to estimate the causal effects.

LiNGAM (Linear Non-Gaussian Acyclic Model)



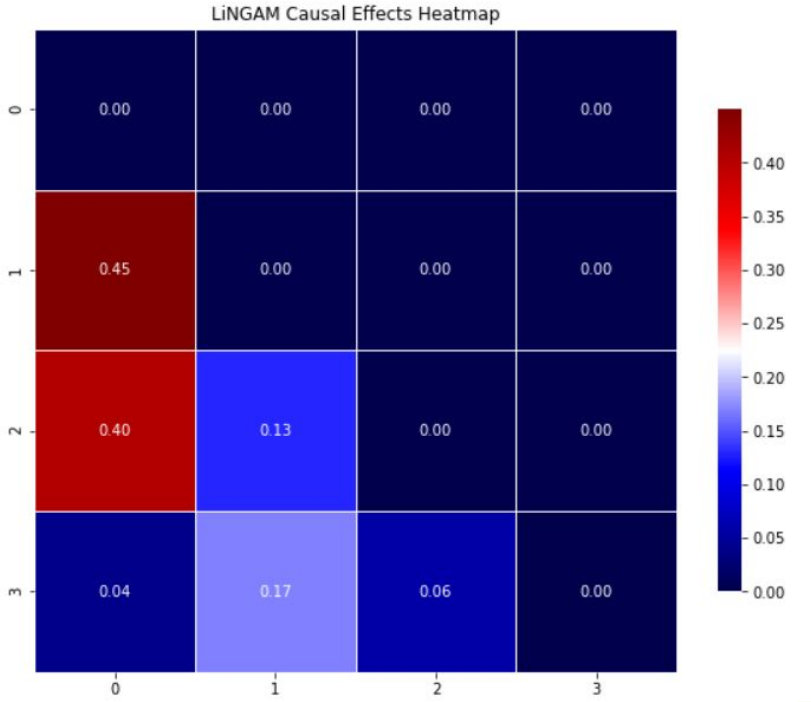
Why LiNGAM:

To discover the causal structure in the data by exploiting non-Gaussian distributions.

The analysis revealed that GDP per capita, social support, and healthy life expectancy have significant causal effects on happiness.

LiNGAM (Linear Non-Gaussian Acyclic Model)

1. **GDP per capita (1) on Happiness Score (0): 0.45**
2. **Social Support (2) on Happiness Score (0): 0.40**
3. **Healthy Life (3) on Happiness Score (0): 0.04**



UC San Diego

DoWhy

Purpose: To estimate causal effects using graphical models and potential outcomes frameworks.

DoWhy helped us to get the **Average Treatment Effect**.

Why DoWhy?

- Explicit identifying assumptions.
- Separation between identification and estimation.
- Automated Validation of Assumption.

DoWhy

- **The estimated Average Treatment Effect (ATE) for GDP per capita is 1.1959**, indicating that a one-unit increase in GDP per capita leads to an increase of approximately 1.196 units in the Happiness Score.

- **The estimated ATE for Social support is 0.5862**, suggesting that a one-unit increase in Social support leads to an increase of approximately 0.586 units in the Happiness Score.

Sr. No.	Causal Inference for each field as treatment variables	Value
1	GDP per capita	1.959
2	Social Support	0.586
3	Healthy life	1.026
4	Freedom	1.987

DoWhy

- **The estimated ATE for Healthy life is 1.0261**, indicating that a one-unit increase in Healthy life expectancy leads to an increase of approximately 1.026 units in the Happiness Score.

- **The estimated ATE for Freedom is 1.9876**, suggesting that a one-unit increase in Freedom leads to an increase of approximately 1.988 units in the Happiness Score.

Sr. No.	Causal Inference for each field as treatment variables	Value
1	GDP per capita	1.959
2	Social Support	0.586
3	Healthy life	1.026
4	Freedom	1.987

Propensity Scores

Propensity score matching aims to create comparable treatment and control groups by matching individuals based on observed characteristics. This helps control for confounding variables.

Steps involved:

1. Predict propensity scores representing the likelihood of receiving the treatment. (linear regression)
2. Propensity Score Scaling using StandardScaler.
3. Nearest Neighbors Matching to create matches between treated and control units.
4. Matched Dataset Creation.
5. Causal Inference using DoWhy

Propensity Scores

Sr. No.	Estimated ATE using Propensity for given fields	Value
1	GDP per capita	1.2736
2	Social Support	0.5605
3	Healthy life	1.2526
4	Freedom	1.7513

Propensity Scores

1. Estimated Average Treatment Effect with Propensity Score Matching for GDP per capita= 1.2736: On average, increasing GDP per capita by one unit is associated with an increase of approximately 1.2736 units in the Happiness Score, according to the estimated causal effect obtained through propensity score matching.

2. Estimated Average Treatment Effect with Propensity Score Matching for Social support= 0.5605: On average, increasing Social support by one unit is associated with an increase of approximately 0.5605 units in the Happiness Score.

3. Estimated Average Treatment Effect with Propensity Score Matching for Healthy life= 1.2526: On average, increasing Healthy life by one unit is associated with an increase of approximately 1.2526 units in the Happiness Score.

4. Estimated Average Treatment Effect with Propensity Score Matching for Freedom= 1.7513:On average, increasing Freedom by one unit is associated with an increase of approximately 1.7513 units in the Happiness Score.

Key Findings from our Analysis

Consistent Positive Effects:

In both methods, GDP , Healthy life, and Freedom consistently show substantial positive effects on Happiness Score.

Social support Influence:

Social support has a positive effect, but the magnitude is smaller compared to the other features.

Methodological Consistency:

While there are differences in the estimated magnitudes between the two methods, the general trend is similar.

Conclusion

Conclusion

Our study aimed to identify the causal effect of socioeconomic factors on happiness using advanced causal inference techniques.

The analysis revealed that economic performance, freedom, health and social support are significant drivers of happiness.

Thus, policymakers should focus on improving economic conditions, social support systems, and healthcare to enhance overall happiness.

References and Links

Dataset:
https://www.kaggle.com/datasets/nikbearbrown/the-economics-of-happiness-simple-data-2015 2019/data

https://academic.oup.com/ejcts/article/53/6/1112/4978231

https://medium.com/data-science-at-microsoft/causal-inference-part-1-of-3-understanding-the-fundamentals-816f4723e54a

https://blog.ml.cmu.edu/2020/08/31/7-causality/

https://github.com/py-why/dowhy

https://lingam.readthedocs.io/en/latest/tutorial/lingam.html

Thank you