

## Descrição da base de dados

Ana Flávia Dias Rodrigues - 561288  
Camilla Damasceno da Silva - 571915  
Jorge Luiz Mendes Moreira Junior - 553273  
Luiza Ávila Defranco Gonçalves - 587490  
Stefany Gaspar Xavier França - 582931

Instituto de Ciências Exatas e Informática  
Pontifícia Universidade Católica de Minas Gerais (PUC Minas)

### 1. Qual a base a ser utilizada?

**World Happiness Report**

Link: [www.kaggle.com/unsdsn/world-happiness?select=2019.csv](https://www.kaggle.com/unsdsn/world-happiness?select=2019.csv)

### 2. Quantos atributos?

Contém 9 atributos.

### 3. Qual o tipo de atributos (binário, nominal, etc)? qual o significado de cada atributo?

8 atributos numéricos e 1 nominal.

1. **Overall rank**
2. **Country or region**
3. **Score**
4. **GDP per capita**
5. **Social support**
6. **Healthy life expectancy**
7. **Freedom to make life choices**
8. **Generosity**
9. **Perceptions of corruption**

A coluna **Overall rank** mostra a posição em que o país ficou após a análise. **Country or region** o nome do país ou região, o Score é o total de pontos do país/região. As colunas: **GPD (PIB) per capita**, **Social support**, **Healthy life expectancy**, **Freedom to make life choices**, **Generosity** e **Perceptions of corruption** descrevem até que ponto esses fatores contribuem na avaliação da felicidade em cada país, seus números estimam como cada um dessas escolhas contribuem para o valor dado em uma avaliação tipo-escada, em que cada país teve representantes avaliando suas condições de vida. Por exemplo, se os habitantes acham que de 1 a 10 a qualidade de vida é 7, os números estimam quanto cada um contribui para essa escala.

### 4. Quantas instâncias?

Contém 156 instâncias.

### 5. É um problema de classificação ou não?

É um problema de regressão.

## **6. Fazer uma pequena descrição da base de dados.**

As classificações são baseadas nas respostas à principal pergunta de avaliação de vida feita na enquete. Isso é chamado de escada Cantril: pede aos entrevistados que pensem em uma escada, com a melhor vida possível para eles sendo 10 e a pior vida possível sendo 0. Eles então são solicitados a avaliar suas próprias vidas atuais de 0 a 10. As classificações são de amostras representativas nacionalmente, para os anos 2015-2019. Eles são baseados inteiramente nas pontuações da pesquisa, usando os pesos do Gallup para tornar as estimativas representativas. As subbarras mostram a extensão estimada em que cada um dos seis fatores - níveis de PIB, expectativa de vida, generosidade, apoio social, liberdade e corrupção - contribui para tornar as avaliações de vida mais altas em cada país do que na Distopia, um país hipotético que tem valores iguais às médias nacionais mais baixas do mundo para cada um dos seis fatores. As sub barras não têm impacto na pontuação total relatada para cada país, mas são apenas uma forma de explicar para cada país as implicações do modelo estimado em. As pessoas costumam perguntar por que alguns países têm classificação mais alta do que outros - as sub barras (incluindo os resíduos, que mostram o que não é explicado) são uma tentativa de dar uma resposta a essa pergunta.