**Analyzing public opinion on the world’s happiest countries**

**Introduction:**

About the video: The video was posted on 2018, although this may seem to be outdated this was the only source to find more than 500 comments and a well moderated comment section. This is key as unmoderated comments usually contain profanity and unfiltered words which could harm the sentiments of the audience.

The video focuses on what are the most important factors in determining a country’s happiness index. The data were compiled from the World Gallup poll, wherein people were asked to value their lives as a whole, with the best possible life being a 10 and the worst possible being a 0. Norway, Sweden, Denmark, Iceland, and Finland, the Nordic countries, are in the top ten. And they've all ranked among the top 10 since the report was first put out. And Switzerland - this year's #5 - has also always hovered near the top also.

There are a few interesting insights gained from this video which can be compared with this year’s world happiness index report to see what the changes and consistencies are.

**Objectives and Research Questions:**

1. **Are the rankings accurate? –** As the reviews are gathered from a survey, there is a chance of sampling error, wherein the sample may not be representative of the entire population. The comments, specifically from people of these countries will help us gather insights regarding this.
2. **What is the general sentiment of the comments? -** The comment section consists of people that agree and do not agree with the topics presented in the video, we can study how many of these are neutral, positive and negative.
3. **What are the general themes/topics? –** This is the most important part of the analysis as it will help us form further research questions and obtain datasets to go further and study the trends in these topics.
4. **What is the correlation between the identified topic?**
5. **What is the degree of centrality and betweenness among the topics?**

**Data Collection and Cleaning**: The comments were collected from a [YouTube video](https://youtu.be/sEKQYYhp3eE?si=LwEUKVx3kEfCgb-k) titled ‘Happiest Countries in the World: Explained’

A total of 1120 comments were collected from the video using the YouTube API key.

The data was then cleaned to exclude, emojis, links, stop words (about 15 words were excluded) and all the comments were converted to lowercase.

The text was then processed to remove non-alphabets, mentions, hyperlinks, pictures and ‘amp’ which is obtained from the ampersand symbol. The cleaned/processed text is then tokenized and stored to perform further analysis.

**Ethical considerations:**

* **Hide Personal Information –** Once the data has been obtained remove all the columns/rows that could contain personal data such as username, email or pictures.
* **Respect Platform Policies –** As the data was obtained using the youtube API, we can ensure that the policies required were followed.
* **Transparency –** The data is being used only for research purposes and not for monetary gains.

**Sentiment Analysis Techniques:** The unique tokens were extracted and a word map was used to analyse the most occurring words/tokens.

A close up of words

AI-generated content may be incorrect.

The next step involved assigning a subjectivity and polarity score to the tokens.

A diagram of different colored dots

AI-generated content may be incorrect. A graph of different colored bars

AI-generated content may be incorrect.

A graph with green and orange squares

AI-generated content may be incorrect.

The topics were generated using the query ‘Analyse the comments of the csv file which contains comments for the video titled ‘Happiest Countries in the World’ and obtain the top 10 topics for the comments’

A graph with colorful bars

AI-generated content may be incorrect.

Polarity scores of each topic(First one is healthcare):

A graph of a number of different colored bars

AI-generated content may be incorrect.

Correlation between the topics:

A chart of different colored squares

AI-generated content may be incorrect.

Dendogram of topics: A graph with different colored lines

AI-generated content may be incorrect.

Degree of centrality and betweeness among the topics:

A diagram of a star with blue dots and black text

AI-generated content may be incorrect.

Future Questions: Analyse the world’s biggest economies, climate, safety, per capita gdp and their happiness ranking,