**Project Web Scraping**

**Proposed Datasets:**

1st option:

Regarding the **US Police Shootings** we have the following:

* <https://www.kaggle.com/datasets/ahsen1330/us-police-shootings>

This is the Kaggle doc that we have found. It is editable in tableau

These are the following websites that are candidates for webscraping:

* <https://data.census.gov/profile?g=040XX00US48>
* <https://www.bea.gov/>

Proposed framework:

* From the **Kaggle dataset** – determine how many shootings happened, possibly their severity and nature, per state within the USA
* From the **mentioned websites** – determine how economically developed (earnings, unemployment, education) the states are in comparison with the level of shootings

2nd option:

Regarding the **World Happiness** we have the following:

* <https://www.kaggle.com/datasets/unsdsn/world-happiness?select=2016.csv>

This is the Kaggle doc that we have found.

These are the following websites that are candidates for webscraping:

* <https://www.theglobaleconomy.com/rankings/Education_spending/Europe/>

Proposed framework:

* From the **Kaggle dataset** – determine the degree of happiness per country (we may develop with some extra stuff along the way)
* From the **website** - extract indicators that may be correlated with the level of happiness within certain countries (for example spending in education, level of wealth, other stuff)

3rd option:

Regarding the **Netflix Movies and TV Shows** we have the following:

* <https://www.kaggle.com/datasets/shivamb/netflix-shows>
* <https://www.kaggle.com/datasets/prasertk/netflix-daily-top-10-in-us>

This is the Kaggle doc that we have found.

These are the following websites that are candidates for webscraping:

* <https://www.imdb.com/?ref_=nv_home>
* <https://www.rottentomatoes.com/>

Proposed framework:

* From the **Kaggle dataset** – determine Netflix´s movies
* From the **mentioned websites** – determine the movies IMDB ratings

# World Happiness Report – Webscraping Project

Regarding the **World Happiness** we have the following:

* <https://www.kaggle.com/datasets/unsdsn/world-happiness?select=2016.csv>

This is the Kaggle doc that we have found.

These are the following websites that are candidates for webscraping:

* <https://www.theglobaleconomy.com/rankings/Education_spending/Europe/>

Proposed framework:

* From the **Kaggle dataset** – determine the degree of happiness per country (we may develop with some extra stuff along the way)
* From the **website** - extract indicators that may be correlated with the level of happiness within certain countries (for example spending in education, level of wealth, other stuff)

To do:

* Streamlite until 10/04 (3 project sessions)

Work Guide

These are some suggested guidelines for our project and should be seen as such. Doing them might help us achieve the final goal

Regarding **Kaggle** we could do the following:

1. Pick up the 10 happiest countries in the overall rank
2. Pick up the 10 unhappiest countries in the overall rank

(**Streamlite Graph 1** – shows chart graph of the 10 happiest countries versus the 10 unhappiest)

Then we could, for example, try to see which variables are correlated with the happiness of the countries’ populations.

The Kaggle dataset that is presented to us tries to compare Hapiness side by side with the following factors:

* Economic Production
* Social Support
* Life Expectancy
* Freedom
* Absence of Corruption
* Generosity