

# 06\_\_assignment

November 16, 2022

## 1 End of course assignment

It's time for you to put your newly developed Python skills to use! The assignment is simple: find some data and tell a story. At the very least, this will require you to:

- locate and download an interesting dataset
- load it into Python
- make a chart to visualise some trend in the data
- share your findings with the group in a 3-5 minute presentation

You can complete the task in this notebook using either Jupyter or [Google Colab](#), or by writing a script in Spyder. Some potential sources of interesting data are listed below, but there are many other sources of quality data on the internet and you are free to choose whatever you want. In the final session, you will be able to share your screen via zoom, or you can post an image or pdf document in the group chat.

### 1.1 Potential sources of data

#### 1.1.1 [kaggle](#)

This excellent little website is as much a thriving community of data scientists as it is a source of quality data. After making a free account, you can browse over 178,000 data sets or search using keywords. The kaggle community even post code, analyses and discussion points, which you may look to for inspiration.

Here are some interesting data sets that caught my eye on kaggle:

- [Blood Type Distribution by Country](#)
  - What are the most and least common blood types, and which countries appear to have the most and least of this blood type?
- [Indian Food 101](#)
  - How many unique ingredients are there across all dishes in the data set?
  - What is the most common ingredient?
  - From which state do most dishes originate?
  - Including both preparation and cooking time, what are the fastest and slowest dishes to make?
- [World Happiness Report](#)
  - In the period 2015 - 2019, which countries were the happiest? For which countries did happiness increase / decrease most substantially in this period?

### 1.1.2 [hockeyfights.com](#)

I'm not a hockey fan, let alone a bloodthirsty one. This website makes the list purely because it is a well-curated source of interesting data. I used it some years ago in a similar assignment to explore the relationship between number of fights and games won / lost in a season for the Detroit Redwings and Toronto Maple Leafs.

- Which are the top three most bellicose hockey teams?
- Which teams have the greatest fighting rivalry?
- Is there a relationship between number of fights in a season and the number of games won / lost?

### 1.1.3 [Our World in Data](#)

This brilliant website publishes data-driven reports on poverty, disease, hunger, climate change, war and many other big problems that exist in the World. Most reports include a link to download the raw data, usually in CSV format. Here are a few examples:

- [Global CO<sup>2</sup> Emissions](#)
  - Since records began, which 10 countries are responsible for the most CO<sup>2</sup> emissions resulting from fossil fuel burning? Which 10 are responsible for the least emissions?
  - As above, but for emissions resulting from land use changes
- [Wild Mammals are making a comeback](#)
  - Download the data and recreate the chart on the web page
- [Air Pollution](#)
  - Over the last 10 years in the United Kingdom, which pollutants have increased, and which have decreased?

### 1.1.4 [Google Trends](#)

Google trends allows you to analyze and compare 'interest' in specific search terms over time and by geographical region, and to then download the results in CSV format. Often this reveals perplexing trends that warrant further exploration. For example, when comparing interest in 'marriage' and 'divorce' across the world since 2004, most countries were more interested in marriage, but France showed a strong preference for divorce. What's going on there?

### 1.1.5 [Google Dataset Search](#)

This is Google's search engine for datasets. With a simple keyword search, users can find datasets hosted in thousands of repositories across the Web.

Good luck!

[ ]: