Session 7 Labs:

Checklist:

1. Make sure you have your Python install working (ask for help if you don't). I would recommend doing these exercises in a jupyter notebook inside VScode
2. Clone this weeks git repo
3. Make sure you have numpy, pandas and beautiful soup installed

TASK 1:

Install and run though the basic operations of NUMPY. Once you have tried these, look at some of the operations we didn't cover. Try:

Split()
Sort()
Filter()

Check W3 schools for guidance on any of these:
https://www.w3schools.com/python/numpy/default.asp

TASK 2:

1. Choose a .cvs file to work with. There are two in the repo. The demo file from W3 schools contains data around exercise stats. There is also a file containing the collection of a US art gallery.
2. See what data is in the file and think how you might want to view it
3. Clean the data, you can find how to do this here: https://www.w3schools.com/python/pandas/pandas_cleaning.asp
4. Find interesting correlations in the data. Eg. pulse to calorie burn, year to dimensions. Info on this process here: https://www.w3schools.com/python/pandas/pandas_correlations.asp

Optional : Once you've done this, find a new .cvs online and repeat the steps above. Museum exhibit list, sports stats (league tables ect…) are fun. You can find a whole range to play with here: https://www.kaggle.com/datasets?fileType=csv

TASK 3:

Web-Scraping:

Part 1:

1. Take a look at the example notebook. Make sure it will run correctly on your system.
2. Run the code to scrape data from tables on a wiki page of your choice.
3. Save as a .csv
4. Then, use pandas to clean the data (if needed) and try find some interesting correlations

Part 2:

1. Work through the example image scraper notebook step by step.
2. Try to understand the process and then scrape the images from a different webpage to the example. NOTE: make sure you are allowed to scrape the page!