Python Lab 8 Exercises

Exploratory Data Analysis (Pandas)

Question 1.

Consider the dataset in the file "fruit.csv" on the harvest of different categories of fruit in Spain and the UK (1961–2019). *Download the CSV file and save it in the same location as your Python code for this question.*

The data comes from the Food and Agriculture Organization of the United Nations (FAO) statistics http://www.fao.org/faostat/. The variable *area* is the harvested area (units of hectares) and *yield* is kilograms of fruit per hectare.

(a) Write Python/Pandas code to find the number of unique values of fruit categories (the variable *item*).

(b) Write Python/Pandas code to determine which fruit categories are harvested in both Spain and the UK. *Hint: consider a set intersection.*

(c) Consider only the year 2019. Find the total area harvested in each country. Also construct a summary table giving the total area harvested for each category of fruit, given in decreasing order of area.

(d) We wish to plot the *yield* of Strawberries each year on a line graph showing one line for Spain and one line for the UK (together with a legend). Complete the missing pieces of code below to carry out this task.

```
AAA = fruit.query( box A )
BBB = AAA.pivot(index= box B , columns= box C , values= box D )
CCC = BBB.reset_index()
plt.figure()
CCC.plot.line(x='Year', y= box E )
plt.show()
```

(e) Consider only fruit harvested in the UK. Write Python/Pandas code to plot side-by-side boxplots giving the *yield* of each category of fruit. *Make sure the plot has a suitable title and fruit category labels are clearly readable.*