

Python Lab 7 Exercises

Pandas

Question 1.

Give two advantages of Pandas over Microsoft Excel for manipulation or analysis of data.

Question 2.

Consider the dataset in the file "msleep.csv" on the sleep times and weights of mammals. You can download this dataset from the Aula site for 7143CEM. *Download the CSV file and save it in the same location as your Python code for this question.*

```
import pandas as pd
import matplotlib.pyplot as plt
msleep = pd.read_csv('msleep.csv')
```

- (a) What Pandas code would you use to find the number of mammals (rows) and variables (columns) in this dataset?
- (b) The taxonomic order of each mammal is given in the variable *order*. Write two different ways in Pandas to find the number of unique values of this variable.
- (c) Write Python code to construct a summary table giving the frequency of each value of the variable *vore* (carnivore, herbivore, omnivore, etc) and construct a horizontal barchart of these frequencies.
- (d) Consider only mammals of *order* Rodentia (rodents). Write Python code to plot side-by-side boxplots of the variables *sleep_cycle* and *sleep_rem*, and to sort the rodents in decreasing order of brain weight (the variable *brainwt*).