CS24420 Scientific Python --- Semester 2 --- Practical Worksheet 1

Pandas DataFrame

The tasks this week are about using the pandas DataFrame class. You'll need to refer to your lecture notes and also the online documentation for pandas.

1. Write a Python program to to visualize the recent share price history of Alphabet Inc. (the parent company of Google)

The data is available on Blackboard as a csv file: `GOOGL_2021-22.csv`. You will need to download this file.

The data has columns '`date`', '`open`', '`high`', '`low`', '`close`' and '`volume`'. This is a time series dataset, so we would like the index (the row labels) to be the date column instead of just successive integers.

a) Read historical data from the CSV file into a Pandas `DataFrame`. with `read_csv`. To make the date column be the index for the data you will need to use the optional parameters `index_col` and `parse_dates`. See the Pandas documentation for `read_csv` to find out how to use these.

b) Print the summary statistics for the dataset.

c) Plot the '`volume`' column as a line plot.

d) Slice the data to create a new `DataFrame`, keeping only columns from open price ('`open`') to close price ('`close`') inclusive (4 columns) and assign it to a new variable.

e) Plot the sliced data as a line plot.

f) Create a new column named '`change`' in the `DataFrame`, consisting of the '`open`' value subtracted from the '`close`' value. Plot the '`change`' column as a line plot.

   Hint: ***`df['newcol'] =`*** *some expression involving* ***`df['somecol1']`*** *and* ***`df['somecol2']`***

g) ***[optional extra]****: Repeat the analysis for Meta Platforms Inc. (Facebook) using the file:* `FB_2021-22.csv`*. Then create a new DataFrame containing just the* `close` *column for each of the two companies and plot them on one graph.*

The data for this exercise was taken from mactrotrends.net

Plotting hints: (1) If you want to plot two or more **separate** plots in the same script, you should call `plt.clf()` between them - otherwise they will add to the same plot. (2) Remember you can save your plot as a graphic file using `plt.savefig("filename.png")`

2. Write a Python program to analyse some data from the brain dataset

The data is available on Blackboard as a comma-separated file: `brain.csv`. You will need to download this file.

a) Read the data from the file into a Pandas `DataFrame`.

b) Print the summary statistics for the dataset.

c) Group the data by the `Gender` column.

d) Produce boxplots of the grouped `Height` and `Weight` columns.

e) Why is this probably not a very useful way of showing the data?

2023-02-03 V2.0 LGT