

## SD212 12-Week Exam Practice

1. Suppose you have two files `cities.csv` and `precip.csv` that look like this:

```
city,state,population,mayor
Phoenix,AZ,1680992,Kate Gallego
Dansville,NY,4433,Dennis Mahus
Ithaca,NY,32108,Laura Lewis
Frederick,MD,78171,Michael O'Connor
Annapolis,MD,40812,Gavin Buckley
```

```
city,yearly precipitation
Ithaca,37.3
Annapolis,47.27
Phoenix,8.03
Frederick,40.49
```

Write a Python program that reads in these two files to create a new file `cityrain.csv` that looks like below. (Ordering of the rows and number of decimal digits doesn't matter, but notice that Dansville isn't there!)

```
city,state,population,monthly precipitation
Phoenix,AZ,1680992,0.67
Ithaca,NY,32108,3.11
Frederick,MD,78171,3.37
Annapolis,MD,40812,3.94
```

2. Consider a file `badcities.csv`, similar to the last problem but with some issues:

```
city,state,population,mayor
Dansville,NY,4433,Dennis Mahus
San Marino,,33745,Luca Beccari
Ithaca,NY,32108,Laura Lewis
New York,NY,,Eric Adams
Annapolis,MD,40812,Gavin Buckley
```

Write a Python program **or** a bash script that:

- Removes any lines where the state name is blank
- Changes any missing population values to 10000 (not a actually good idea in practice!)
- Sorts the rows by population, increasing
- Saves to a new file `goodcities.txt`

For the example above, the resulting file should be

```
city,state,population,mayor
Dansville,NY,4433,Dennis Mahus
New York,NY,10000,Eric Adams
Ithaca,NY,32108,Laura Lewis
Annapolis,MD,40812,Gavin Buckley
```

(Extra challenge: try doing this separately in Python *and* in bash)

3. Suppose there are 10 text files called `file1.txt`, `file2.txt`, ..., `file10.txt`.

Write a **bash script** that displays each filename along with the number of lines in that file, in any order, like

```
390 file6.txt
3053 file10.txt
6124 file1.txt
1602 file9.txt
19416 file8.txt
2761 file5.txt
84 file2.txt
8511 file3.txt
17632 file4.txt
8675 file7.txt
```

Your script should work **in parallel** so the lines in each file are counted simultaneously.

4. Now write a **Python program** that counts the number of lines in each of these 10 files `file1.txt`, `file2.txt`, etc. **in parallel** and prints only the name of the file that has the most lines.

So for the files in the previous example, your program would just have one line of output:

```
file8.txt
```

5. Suppose there is a file `english.txt` that contains all English words in lowercase, like

```
english.txt
able
constable
brave
rave
revealing
stable
travesty
veal
```

Write a Python program that counts how many words appear within some other word.

In the case of the small dictionary above, the count would be 4, for able (**constable** or **stable**), rave (**brave** or **travesty**), stable (**constable**), and veal (**revealing**).

Assume that the file has exactly 100,000 lines. For full credit, you should design your program run efficiently **in parallel**.



11. Consider the following scenario:

A big AI company has developed a popular chat bot which is used interactively by people around the world to answer their questions.

The U.S. military learns that insurgent terrorists may be asking questions about how to find weaknesses in federal government installations and cause harm to military personnel.

The DoD asks the AI company to secretly add code that detects when a user starts asking questions along these lines, and in such cases collects their IP address and user data to be immediately shared with military intelligence.

Write a brief few paragraphs to the head of the AI company outlining the ethical considerations that should be made in deciding whether and in what manner to honor the DoD's request.

(We are not looking for a certain answer, but rather for clear reasoning that demonstrates understanding of the technical and ethical issues involved.)