Data Science Questions

March 4, 2021

All data files are in `/u1/junk/Data`. Your program must read directly from those files. Do not copy the data files.

1 Elementary Question

The data file `before.csv` contains temperature data for Terre Haute airport in the year 2019. There are three fields on each line: the date, the high temperature (in °F), and the low temperature. Unfortunately, some of the temperature data is missing. Missing data is indicated by the letter 'M'. For example, the following might be a few of the lines from the file:

```
01-03,42,25
01-04,47,22
01-05,54,M
01-06,53,28
01-07,57,48
```

This would be the data for January 3 through January 7. As you can see the low temperature for January 5 is missing. Your job is to use *linear interpolation* to repair the data. In this case, you would replace the M value by the average of the low temperatures for the day before and the day after. The repaired data would look like this:

```
01-03,42,25
01-04,47,22
01-05,54,25
01-06,53,28
01-07,57,48
```

Repair all missing data, and save your results in a file named `after.csv` in the current directory. Note: There may be missing data for consecutive days, but the data for January 1 and December 31 will be there.

2 Intermediate Question

Figure 1 shows the arrangement of a group of childrens blocks. These blocks were recovered from a mini-blackhole and cannot be used for educational purposes. The blocks need to be repaired. Your job is to repair and rearrange them so they look like Figure 2. Write a Python program that does this. Note the 'Z' is omitted.

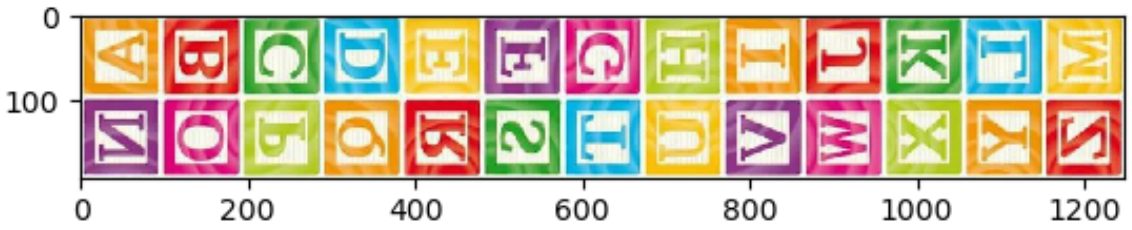

Figure 1: Before

The original picture is stored in `before.npy` (a Numpy binary data file created with `save`). Write a program that creates `npy` and `png` files containing the modified picture. Use the file names `after.npy` and `after.png`. One point will be deducted for every `for-loop` you use.

Note: After you load the data file, you will find that the image is a three dimensional array, with the third dimension being color. You do not need to change any colors. The pictures of the individual letters are all `96x96` pixels. For each pixel, three 8-bit unsigned integers are given, specifying the red, green, and blue coordinates of the pixel.

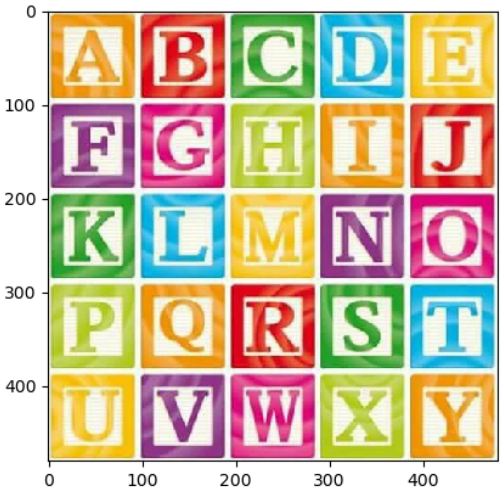

Figure 2: After