

Portfolio Assignment 1: Text Processing

1. This script takes a CSV file containing employee records as input and then parses the file for various employee related data points. The program also handles erroneous data by clarifying the incorrect data with the user and modifies data to follow a uniform format. After parsing the file into a series of objects representing individual employee data, the program then dumps the data in a pickle file, unpacks that pickle file and then displays the data in a user-friendly view on the command line.
2. To run the program, type the following command into the command line (from the same directory as the program):

```
python3 HW1_sxr180064.py <relative_path_to_input_CSV>
```

3. Some of the advantages to using Python for text processing are the vast and versatile collection of tools and libraries that exist for this function and the language's relatively simple syntax (and the engineering hours saved from it). On the other side however, Python is dynamically typed and interpreted so it is rather slow thus making large and complicated regex operations very resource heavy and time intensive. Interestingly though, at some point, the speed gains in using a different language like C++ seem to diminish as the task gets increasingly complex.
4. Having coded in Python before, the coding portion was a good review especially since I had not really coded in Python for a while. Regex has always been a topic which I have avoided because of the seeming complexity of the topic. As a result, that was something which I had to learn for this assignment and I'm glad that I did since I also got to realize the convenience of the concept.