Vewpoint Interview: Questions 4-6

4. I would probably use something like sklearn's impute function within python to accomplish this. It works with many data formats, and parameters allow for specifying how the null data is handled. I am not sure if this is optimal though for extremely large file sizes.

5. If I was using python I would most likely set up a process on a server using something such as schedule (linked below). It provides a very simple API for scheduling python tasks at specific times. I would make sure this process is running at the end of the day each Wednesday so it would complete on time that night.

https://pypi.org/project/schedule/

6. Im a big fan of what Bill Gates has been doing with his foundation. They use a lot of big data and follow the scientific process while addressing humanitarian issues, such as eradicating polio in Africa. I would like to find a way to use data science and learning in a positive way such as this.