3) If you were asked to impute null values in a column of a file that was 365 Gigabytes, what would you do? What tools would you use? What tools would you NOT use?

Firstly, I'd check if the data in the column is numerical or categorical because we would have to deal with them differently.
If the data is categorical we can just create another column called "OTHER" to differentiate the null values from the rest. This is usually the most efficient way since missing data is probably important information. We can also fill the nulls with the most common class, statistically speaking - the mode of the column.

Otherwise, if it's numerical, we can handle the null values in various other ways. We can choose to fill the voids with the mean or median - in this way we will avoid data loss at the expense of accuracy. We can also choose to simply replace the nulls with 0. This is essentially like the "OTHER" in the categorical section. Again, missing values in themselves are very informative that's why it is good to deal with them by adding a new label for them.

Now the main part - the dataset is 366GB huge. If we try to directly read the CSV file using the pandas library we would get a memory/storage error. An efficient way to deal with this is to read chunks of this dataset - some 50-100Mbs at a time and keep appending them to a SQL database (preferably). We can choose other databases too. But if we choose SQL we can write SQL queries using pandas in python to read the database.

We can choose to impute missing values after reading a small chunk of the data or to do it all together after everything is in the SQL database. Now the question arises, which one is more preferable:

1. If the database is sorted or maybe has some order, for example, the salaries are listed in sorted order then we would want the imputed values to be somewhat similar to that of the salaries in that chunk of data. What I mean by this is that, if the first 100 data points have salaries ranging from $40000 - $80000 given the mean of the entire dataset to be $300000, we would want to null values to take on a value that's around that range instead of the mean.
2. If there's no order, then imputation can be done in the end.

To answer the question now, I wouldn't use python ONLY. But I'd rather use Python, SQL, and the pandas library.

4) What would you do if you were asked to do the above task every Thursday morning at 2:00 am?

If we can have the PC turned on throughout the night then we can have a scheduler to run the code. Otherwise, we can run it on a server like Google Cloud or AWS which is practically better.

5) Who is your favorite mathematician, statistician, or computer scientist and why?

Computer science is such a vast field with continuous developments and inventions by extraordinary scientists over the years. I have had the privilege of taking a class with one such computer scientist who I consider my role model in the field. Dr. Ramesh Sitaraman has had a splendid career and was part of Akamai's CDN team. However, the reason he is my favorite is more personal. He made me fall deeper in love with something I thought I couldn't be more in love with. I took an algorithms course with him and I never thought I'd find a course like that very interesting but once he began explaining concepts, I found clarity in my thought. I understood algorithms better. He changed my perspective on how I should think about problems. Whenever I encountered a coding problem, I began coding immediately and dealt with errors as they appeared; however, after a few lectures with him, I tried to design solutions to problems before starting the coding process. Now, I begin with penning down all different edge cases that could be possible.

I used to love coding but after meeting Professor Sitaraman, I started loving computer science more than I thought I could.