Hi Arjun,

First of all thank you for giving me an opportunity to work on this incredibly challenging project. This assignment has challenged me on several fronts. The most challenging part of this task was without a doubt working with those large xml files. There were several strategies that I tried & tested and wanted to share them with you.

My first strategy was to try a website at random that converts xml to csv but then quickly realized that almost every converter online has a file size limit of upto 30 MB for xml files while the 4 xml files that I was dealing with were close to 800 MB So the first strategy failed.

My second strategy was to try & use developer environments such as Java & IntelliJ IDEA 2022.2.2 since I learned through some blogs and tutorials that developer tools especially the JAVA xml parser is a powerful tool to upload data from large xml files and transform them into different formats and in much faster way. However the challenge with this approach was that I had to find a specific algorithm that transforms xml data on the developer environment to csv or xlsx. I went online, searched for videos, articles & found an algorithm with a detailed solution. I ran the algorithm on both developer environments and it took a lot of time as I waited until the next day to get the output files. To my disappointment the output files contained only half the data.

At this point I was quite frustrated and doubted whether I would be able to crack this challenge and meet the objective. I spend some more time trying to figure what could have gone wrong & tried to find all kinds of answers through online forums and discussions. As I pursued answers to different questions I came across a code on the one of the forums that converts xml to xlsx and xml to csv. This code uses the xml etree library in python to parse xml files and convert them to xlsx or csv. I was curious and enthusiastic to give it a try. I quickly ran this code on Jupiter notebook & google colab and to my amazement it WORKED ! FINALLY ! Had this been my first strategy I would have surely saved a lot of time and effort. but it was a great learning experience overall.

I’m eager to have your feedback on the work done and really look forward to working further with you on many more challenging endeavours.

Thanks,

Ranjan Surianarayanan – Python Engineer role applicant